



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DISEÑO DE PROPUESTA DE ARQUITECTURA DE GESTIÓN DE DATOS CIENTÍFICOS
PARA EL CENTRO DE INVESTIGACIÓN *WEB INTELLIGENCE CENTRE*

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE ALEJANDRO MALDONADO BRAVO

PROFESOR GUÍA:
JUAN DOMINGO VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:
PEDRO MALDONADO ARBOGAST
ALBERTO CABEZAS BULLEMORE
ROCÍO RUIZ MORENO

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR AL
TITULO DE: Ingeniero Civil Industrial
POR: Felipe Alejandro Maldonado Bravo
FECHA: 07/12/2016
PROFESOR GUIA: Juan Domingo Velásquez Silva

DISEÑO DE PROPUESTA DE ARQUITECTURA DE GESTIÓN DE DATOS CIENTÍFICOS PARA EL CENTRO DE INVESTIGACIÓN *WEB INTELLIGENCE CENTRE*

El presente trabajo de título busca diseñar una propuesta de arquitectura de gestión de datos de investigación enfocada en el acceso abierto a los datos para el Web Intelligence Centre (WIC, centro perteneciente a la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile), con el fin de prevenir la falta de disponibilidad de datos generados en el centro para trabajos futuros y así evitar el tener que llevar a cabo los experimentos por segunda vez.

El primer pilar de la arquitectura son las políticas de gestión de datos científicos. Para su diseño se utilizó una serie de pasos ofrecida por el Data Curation Centre (UK). El primer paso fue realizar una evaluación del estado actual del WIC respecto a la gestión de datos, posterior a ello se efectuó una revisión y comparación de las recomendaciones sobre los contenidos de las políticas de un número de entidades que ha desarrollado la materia en los últimos años. De este ejercicio se extrajeron los contenidos mínimos que debe contener este documento. A continuación, se analizó quiénes serán los encargados de aprobar la propuesta para poder realizar una versión borrador del documento sujeto a cambios, los cuales se vieron reflejados en el diseño de la versión final.

El segundo pilar de la arquitectura es la infraestructura tecnológica (TI) que soportará la gestión de los datos. Para esto se realizó un levantamiento de requerimientos con el fin de realizar un benchmark de los repositorios y software especializados en la creación de repositorios científicos existentes. De esta fase se consideró a la plataforma DSpace como la mejor solución para diseñar y crear finalmente la propuesta de repositorio de datos del centro.

El tercer y último pilar tiene relación al soporte necesario para realizar la gestión de datos en el centro, para esto se diseñaron guías, documentos y plantillas de documentos. Además se esquematizaron los procesos principales que se desprenden de las políticas diseñadas.

Después se llevó a cabo un proceso de validación de la arquitectura creada respecto a los términos técnicos y requisitos para poder ejecutar la gestión buscada. Donde se concluye que la propuesta permitirá realizar una gestión de datos científicos en el Web Intelligence Centre y con esto traer beneficios, que a pesar de no ser directamente cuantificables en contraste a los costos, se verán reflejados en el corto y largo plazo tanto en el funcionamiento interno del centro como en la investigación que se realizará en el futuro, los cuales ayudarán a cumplir los objetivos del centro. También se presentaron propuestas de mejoras a la arquitectura desprendidas de la validación realizada.

Por último, se realizó una propuesta relacionada a la gestión del cambio que se deberá llevar a cabo en el centro con el fin de implementar la arquitectura propuesta de manera correcta para poder generar finalmente el valor que se busca.

Agradecimientos

Quiero agradecer en primer lugar a mi familia. A mis papás, por todo el apoyo que siempre me han dado desde que tengo memoria y por todo el esfuerzo que han puesto para que pueda ir cumpliendo mis metas una a una. A mi hermana, que ha sido siempre un apoyo constante, el cual se intensificó desde que me vine a Santiago a estudiar, gracias por aguantarme, por ayudarme y siempre estar ahí, gracias por leerte la memoria completa más de una vez (que es un poco extensa), sin ti claramente no hubiera podido hacer todo esto.

Quiero agradecer también a mis amigos que vienen desde el colegio, los "chichas", en especial a Pepo, Mati y Pablito, por estar también en todo momento, por todos los momentos compartidos, por la amistad incondicional, porque cada triunfo de alguno lo celebramos como si fuera propio y cada derrota la hacemos olvidar entre todos.

Quiero agradecer a todas las personas que me han acompañado en mi camino desde que llegué a Santiago en especial a mis compañeros de la universidad con los que he compartido en los años (mención honrosa al equipo Waria) y a todas las demás personas que fueron parte de mi vida de una manera u otra e hicieron mi vida más amena estos años universitarios.

Quiero agradecer a la Feria Empresarial, el grupo de personas que me hizo vivir con más ganas la universidad, que me enseñó cosas que van más allá de lo que podré hacer como profesional algún día, por todos los momentos tantos buenos o malos que pasamos, por todos los amigos que quedarán y por todo el apoyo demostrado este último tiempo.

Quiero agradecer a mi profesor guía y a los profesores del taller de la memoria, por guiarme, aconsejarme y principalmente por presionarme a hacer un trabajo mejor.

Y finalmente quiero agradecer a los integrantes del WIC, que a pesar de yo no haber estado tan presente, siempre supe que ellos si lo estaban, siempre dándome una mano cuando lo necesité.

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes Generales	1
1.1.1. Web Intelligence Centre	1
1.1.2. Financiamiento a la investigación en Chile	5
1.1.3. Gestión de Datos Científicos	7
1.2. Descripción y Justificación del Proyecto	9
1.3. Objetivos	11
1.3.1. Objetivo General	11
1.3.2. Objetivos Específicos	12
1.4. Metodología	12
1.5. Alcances	13
1.6. Resultados Esperados	14
1.7. Estructura de Memoria	14
2. Marco Teórico	16
2.1. Datos de Investigación	16
2.2. Ciclo de Vida de los Datos de Investigación	16
2.3. Acceso Abierto	17
2.4. Ciencia Abierta	18
2.5. Gestión de Datos generados en investigación	19
2.6. Arquitectura de Gestión de Datos Científicos	19
2.7. Plan de gestión de datos	20
2.8. Políticas de Gestión de Datos de Investigación	21
2.8.1. CODATA	21
2.8.2. RECODE	21
2.8.3. ANDS	22
2.8.4. DCC UK	22
2.8.5. Diseño de Políticas de Gestión de Datos	23
2.9. Metadatos	23
2.10. Estándares de Metadatos	24
2.11. Digital Object Identifier	24
2.12. Creative Commons	24
2.13. Open Data Commons	26
2.14. Test de Usabilidad	26
2.14.1. Componentes de la usabilidad	26
2.14.2. Procedimiento	27

2.14.3. Tipos de test de usabilidad	28
3. Políticas de Gestión de Datos Científicos	29
3.1. Estado Actual	29
3.2. Tabla de Contenidos	31
3.2.1. Comparación de guías para la formulación de políticas	31
3.2.2. Contenidos	38
3.3. Ruta de Aprobación	41
3.4. Borrador de Políticas de Gestión de Datos	42
3.4.1. Encabezado y Pie de página	42
3.4.2. Principios y Objetivos	42
3.4.3. Alcances	43
3.4.4. Definiciones	43
3.4.5. Texto de las Políticas	44
3.4.6. Roles y Responsabilidades	53
3.4.7. Aprobación y revisión periódica	56
3.5. Versión Final de Políticas de Gestión de Datos	57
4. Almacenamiento	60
4.1. Levantamiento de Requerimientos y Funcionalidades	60
4.1.1. Subida de Archivos	60
4.1.2. Re-uso y descubrimiento	61
4.1.3. Accesos	62
4.1.4. Otros	62
4.2. Repositorios Existentes	63
4.3. Software para Repositorios de Datos Científicos	65
4.4. DSpace	67
4.4.1. Carpeta Dspace	67
4.4.2. Carpeta Webapps/JSPUI	68
4.4.3. Aplicación	69
5. Soporte	72
5.1. Guías	72
5.1.1. Licencias	72
5.1.2. Metadatos	75
5.1.3. Repositorio	79
5.2. Documentos y templates	82
5.2.1. Compromiso de cumplimiento de políticas	82
5.2.2. Permiso de Acceso a Datos del centro	82
5.2.3. Reporte por Proyecto	84
5.2.4. Plan de Manejo de Datos	87
5.3. Procesos	89
5.3.1. Proceso general por integrante	89
5.3.2. Subir Archivos al repositorio	90
5.3.3. Permiso de Acceso a Datos	90
6. Validación Arquitectura	92

6.1.	Evaluación de Usabilidad	92
6.1.1.	Elaboración	92
6.1.2.	Resultados	97
6.1.3.	Resultados para las preguntas de investigación del test	101
6.1.4.	Cambios y Mejoras	103
6.1.5.	Repositorio	103
6.1.6.	Guía de Licencias	104
6.1.7.	Guía para el uso del Repositorio	105
6.1.8.	Guía de Metadatos	105
6.2.	Validación de Usabilidad	106
6.2.1.	Resultados	106
6.2.2.	Preguntas de Investigación	109
6.2.3.	Cambios y Propuestas de Mejora	111
6.3.	Arquitectura de Gestión de Datos Científicos en el WIC	113
6.3.1.	Gestión del Cambio	117
7.	Conclusiones	120
7.1.	Conclusiones Generales	120
7.2.	Recomendaciones y Trabajo Futuro	122
	Bibliografía	124
A.	Antecedentes	132
A.1.	Investigación en Chile	132
A.2.	Web Intelligence Centre	133
A.3.	Entrevistas	134
B.	Marco Teórico	138
B.1.	Creative Commons: Condiciones	138
C.	Políticas de Gestión de Datos	139
C.1.	Logos	139
C.2.	Formatos de Archivos	140
C.3.	Borrador Políticas de Gestión de Datos	143
C.4.	Versión Final de Políticas de Gestión de Datos	150
D.	Guías y Documentos	158
D.1.	Guía de licencias	159
D.2.	Guía de Metadatos	170
D.3.	Guía para el uso del Repositorio	204
D.4.	Compromiso de cumplimiento de Políticas	225
D.5.	Permiso de Acceso a datos del centros	226
D.6.	Template: Reporte por proyecto	228
D.7.	Template: Primera versión DMP	230
D.8.	Template: Segunda versión DMP	232
D.9.	Ejemplo: Primera versión DMP	234
D.10.	Ejemplo: Segunda versión DMP	237

E. Validación Arquitectura	244
E.1. Test de Usabilidad	244
E.1.1. Formulario inicial	244
E.1.2. Guión de Orientación	244
E.1.3. Entrevista post pruebas	245

Índice de tablas

3.1. Definiciones	32
3.2. Alcances	32
3.3. Responsabilidades	33
3.4. Criterios	33
3.5. Infraestructura	34
3.6. Data Management Plan	35
3.7. Re-uso / Descubrimiento	35
3.8. Reconocimiento	36
3.9. Reportes, monitoreo y sanciones	36
3.10. Ética	37
3.11. Otros tópicos	37
6.1. Evaluación: Dificultad de tareas asignadas	98
6.2. Validación: Dificultad de tareas asignadas	107
C.1. Formatos aceptables para la preservación de datos	142

Índice de figuras

1.1. Gasto en Investigación y Desarrollo / Porcentaje del PIB	5
2.1. Ciclo de Vida	17
2.2. Arquitectura de RDM	20
2.3. Espectro licencias CC	25
4.1. Repositorios Indexados	63
4.2. Página de Inicio	70
5.1. Ejemplo de condiciones de Guía de Licencias	74
5.2. Ejemplo de licencia de Guía de Licencias	74
5.3. Ejemplo de contenido de Guía de Metadatos	78
5.4. Ejemplo de contenido de Guía para el uso del Repositorio	81
5.5. Template reporte de proyecto 1	86
5.6. Template reporte de proyecto 2	86
5.7. Portada Plan de Gestión de Datos v1	87
5.8. Contenido Template Plan de Gestión de Datos v1	88
5.9. Contenido Template Plan de Gestión de Datos v2	88
5.10. Proceso: Proceso general por integrante	89
5.11. Proceso: Subir Archivos al repositorio	90
5.12. Proceso: Permiso de Acceso a Datos	91
6.1. Evaluación: Tiempos por tarea	98
6.2. Guía de Licencias: Tercera Etapa	104
6.3. Guía de Licencias: Sexta Etapa	104
6.4. Validación: Tiempos por tarea	106
A.1. RATIO GASTO I+D/PIB Y EVOLUCIÓN DEL GASTO EN I+D.	132
A.2. GASTO EN I+D SEGÚN FUENTE DE FINANCIAMIENTO (Porcentaje).	133
A.3. Estructura Organizacional WIC.	134
B.1. Tipos de condiciones.	138
C.1. Logo Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.	139
C.2. Logo Web Intelligence Centre.	139

Capítulo 1

Introducción

1.1. Antecedentes Generales

Durante años los investigadores han sido tanto consumidores como productores de datos científicos [1]. El actual desarrollo de tecnologías de almacenamiento, comunicación y procesamiento de ellos han cambiado la forma en que se lleva la ciencia [2], donde se crean y utilizan grandes *sets* de datos con el fin de examinarlos para descubrir nuevos comportamientos o patrones en las personas, la web, las sociedades, el planeta e incluso el universo.

La importancia de los datos generados en investigaciones cumple el rol de ser el método para validar las distintas hipótesis y modelos en las distintas áreas de investigación que existen. Históricamente los resultados se han presentado a través de artículos, reportes, libros y otros recursos de información científica, dejando en segundo plano a los datos utilizados [3].

El libre acceso a estos ha sido el foco a seguir por distintos gobiernos y organizaciones mundiales generando una revolución de ciencia libre a nivel global [4]. Es por esto que la gestión de los datos científicos poco a poco ha tomado relevancia a nivel mundial, con el fin de preservarlos a largo plazo, aumentando su disponibilidad y permitiendo no sólo re-utilizarlos para nuevas investigaciones, sino que también para revalidar o corregir su objetivo original [5].

Por otro lado, una de las áreas más nuevas en la ciencia es el *Web Intelligence*, termino introducido por primera vez en el año 2010 [6] y que se define como una nueva dirección para la investigación científica que busca analizar tanto los roles como los impactos de la Inteligencia Artificial (IA) y las tecnologías de información avanzadas [7].

1.1.1. Web Intelligence Centre

Con la misión de desarrollar investigación pionera en el campo de las tecnologías de información, generando nuevas formas de solucionar problemas complejos de ingeniería a través de herramientas basadas en la Web de las Cosas [8] se creó en el Departamento de Ingeniería Indus-

trial (DII) de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile el centro de investigación “Web Intelligence Centre”(WIC). Desde sus inicios ha sido dirigido por el profesor Juan Velasquez, Phd.

La visión del centro es: “Ser un líder a nivel internacional en la investigación de tecnologías de información y comunicaciones aplicadas a la resolución de problemas del mundo real.”

Los objetivos que busca cumplir el WIC son:

- Publicar en las principales revistas, conferencias y editoriales relacionadas con Web Intelligence.
- Proveer un servicio profesional, excelente y rápido para todos nuestros clientes.
- Dictar cursos de orientación práctica acerca de las Tecnologías de Información y su aplicación en los negocios.

A través de los años ha realizado diversas publicaciones en revistas científicas internacionales y ha producido numerosas memorias de pregrado de estudiantes de la Universidad de Chile, lo cual se sigue haciendo hasta el día de hoy, trabajando principalmente con ingenieros o estudiantes de ingeniería. El WIC también es parte del Web Intelligence Consortium, el cual agrupa centros relacionados a la investigación y desarrollo de Inteligencia Web en el mundo. La estructura organizacional del centro se encuentra en el Anexo A.2

En el WIC se encuentran distintos proyectos relacionados al Web Intelligence, los principales son:

AKORI

Durante los últimos años el WIC en conjunto a la Facultad de Medicina de la Universidad de Chile ha desarrollado un proyecto financiado por el Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) llamado “Plataforma informática basada en Web-Intelligence y herramientas de Análisis de exploración visual para la mejora de la estructura y contenido de sitios web” (AKORI: Advanced Kernel for Ocular Research and web Intelligence).

Este proyecto propone crear una plataforma que permita mejorar la estructura y los contenidos en las páginas web, utilizando técnicas de *eye-tracking*, análisis pupilas y encefalografía (EEG), que han generado distintas bases de datos con el objetivo de analizarlas para lograr el propósito del proyecto.

Con esto se busca cumplir cuatro objetivos: [9]

- Repositorio de datos en base a estímulos visuales desde sitios web.
- Desarrollo de plataforma prototipo que permita extrapolar las preferencias y el comportamiento de potenciales usuarios web.
- Definición de la estrategia de empaquetamiento y transferencia tecnológica, y valorización de la Propiedad Intelectual y del mercado.
- Adaptación de algoritmos de minería de datos masivos.

Dentro de este proyecto se ha generado una gran cantidad de datos en los experimentos realizados en los distintos trabajos enmarcados en él. La gran mayoría de estos datos deben tener un cuidado especial en el cómo se almacenan y quién tiene acceso a ellos debido a que fueron tomados directamente de personas que participaron en los experimentos. Algunos *sets* de datos se han perdido en el tiempo, otros son guardados por el centro para realizar futuras investigaciones.

OpinionZoom

OpinionZoom busca ser una aplicación para realizar seguimiento de tópicos de negocios en Twitter con un sistema de detección de alarmas para cuando existan tanto caídas como mejoras a la popularidad de productos de ciertos clientes [10].

A partir de este proyecto han salido nuevas investigaciones, como por ejemplo el diseño e implementación de un sistema para caracterizar el consumo de marihuana en Twitter y el diseño, desarrollo e implementación de una aplicación de *Web Opinion Mining* para identificar el sentimiento de los usuarios de Twitter con respecto a la industria del *retail*, entre otros.

DOCODE

DOCODE nace como una solución a la problemática del plagio académico. Es un software que facilita el proceso de análisis de originalidad de proyectos digitales [11].

Durante los años 2010 y 2011 participó en la competencia internacional de detección de plagio (PAN) donde obtuvo el quinto y primer lugar respectivamente.

El objetivo general del proyecto DOCODE es “Empaquetar, transferir y comercializar el servicio para la determinación de originalidad de documentos DOCODE”. Y sus objetivos específicos son:

- Desarrollar un prototipo tecnológico comercializable de DOCODE que opere en condiciones reales mediante la prestación de un servicio en modalidad SaaS.
- Desarrollar y aplicar una estrategia de protección y explotación de la propiedad intelectual de DOCODE.
- Generar un paquete tecnológico comercializable del servicio de análisis de originalidad de documentos DOCODE.
- Desarrollar y ejecutar una estrategia de transferencia y comercialización mediante la creación de una unidad de negocios interna y la venta, licencia o representación internacional.

KOKORO

El proyecto KOKORO se define como una plataforma informática para el análisis de factores de riesgo cardiovascular en el monitoreo a distancia de pacientes en la Atención Primaria de Salud: “Una aplicación del *data mining* en *Health Informatics* a grandes volúmenes de datos biomédicos (*Big Data*)”

Se realiza en conjunto con el Instituto Nacional del Tórax y propone monitorear variables fisiológicas de pacientes crónicos cardíacos, desarrollar una plataforma informática que pueda generar indicadores y alertas tempranas con el fin de mejorar la gestión de la salud de pacientes ayudando a prevenir potenciales infartos.

Por otro lado, busca crear una base de datos para analizar la correlación entre variables propias del índice de Framingham y de los trastornos del sueño. Con esto se generarán algoritmos de minería de datos masivos (*Big Data*) para encontrar patrones que permitan caracterizar el comportamiento de un paciente con riesgos cardiovasculares y se valorizará el impacto de las tecnologías de información desarrollada en el sistema público de salud en Chile.

Patagonia Los Lagos

Este proyecto nace de la dificultad para caracterizar la demanda y en configurar y promover la oferta de la región de las Lagos respecto a su sector turístico.

WHALE en su primera etapa es una plataforma informática que busca caracterizar la demanda y la comunicación de ofertas con mayor valor agregado de productos y servicios de la industria del turismo de la X Región de Los lagos de Chile.

SONAMA

SONAMA es un proyecto de plataforma para la realización de seguimientos o monitoreo *online* sobre la opinión y consumo de marihuana y alcohol, utilizando usuarios de Twitter del país.

Este proyecto está en sus inicios y nació como complemento de la encuesta nacional de drogas. Actualmente se está buscando la mejora de sus resultados y el desarrollo de la plataforma que pueda entregar esta información lo más cercano al tiempo real.

Financiamiento

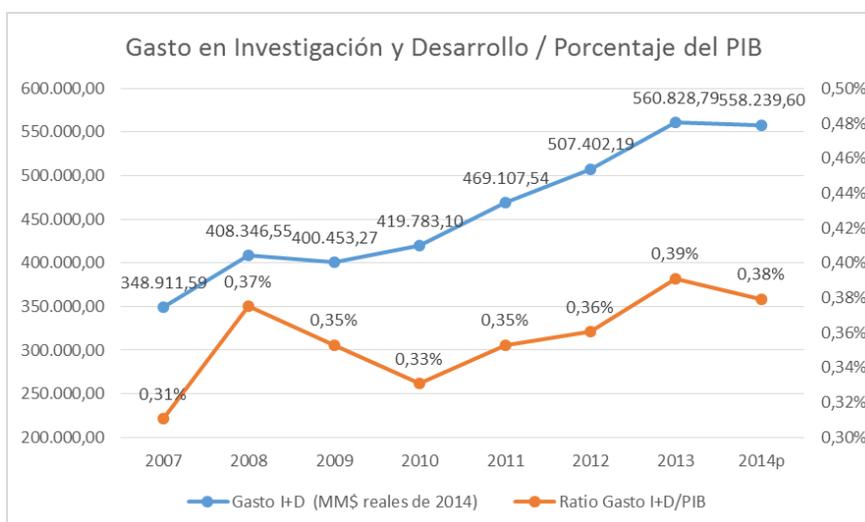
Para el financiamiento de los proyectos del centro se han utilizado fondos públicos enfocados en investigación realizada en universidades del país.

El proyecto AKORI ha sido principalmente financiado por fondos Fondef IDEA, FONDECYT y actualmente por InnovaChile de CORFO. Por otro lado, OpinionZoom ha recibido fondos de InnovaChile (Proyecto de I+D Aplicada) y DOCODE tanto de FONDEF (Investigación y Desarrollo (I+D)) como InnovaChile al igual que los anteriores. El proyecto WHALE fue financiado por fondos FONDEF y finalmente el proyecto KOKORO cuenta con financiamiento propio del centro de investigación.

1.1.2. Financiamiento a la investigación en Chile

De acuerdo a datos otorgados por la V Encuesta Nacional sobre Gasto y Personal en Investigación y Desarrollo (año de referencia 2014) [12], ha existido un alza en el gasto en I+D en Chile durante los últimos años llegando a casi un 0.4 % del PIB, con una leve baja el año 2014, el cual se puede observar en la figura 1.1. A pesar de esta alza, el gasto es bastante bajo en comparación a los otros países pertenecientes a la Organización para la Cooperación y el Desarrollo Económico (OCDE), del cual Chile es parte y donde junto con México se encuentran bastante alejados del 2.36 % promedio de los otros países que la conforman (de acuerdo a los resultados del año 2013) [13]. Por otro lado, el año 2014, el 44 % de la investigación fue financiada por fondos otorgados por el Estado, el cual representa el esfuerzo del Ministerio de Economía por aumentar el gasto total en el país. Los datos se pueden observar en el Anexo A.1.

Figura 1.1: Gasto en Investigación y Desarrollo / Porcentaje del PIB



Fuente: Primera, Segunda, Tercera, Cuarta y Quinta Encuesta Nacional sobre Gasto y Personal en I+D. Se utiliza el deflactor del IPC desde el INE y datos de cuentas nacionales del Banco Central de Chile para PIB. P: Preliminar

Respecto al contexto de la Universidad de Chile, la mayor parte de la investigación se realiza gracias a fondos públicos y es una de las instituciones de educación superior líderes a nivel latinoamericano en esta. Sólo el año 2015 se adjudicaron más de 200 fondos otorgados por el Estado y se tenían en total más de 800 proyectos vigentes, además de contar también con fondos propios de la universidad para el desarrollo de más investigaciones [14].

Los principales fondos a nivel país que se otorgan a distintas entidades con fines de desarrollar investigación en ellas vienen principalmente de dos instituciones; CONYCYT y CORFO y estos son:

CONICYT: FONDEF

El Fondo de Fomento al Desarrollo Científico y Tecnológico es un programa de la Comisión Nacional de Investigación Científica y Tecnológica de Chile (CONICYT) que tiene como propósito el contribuir al aumento de la competitividad de la economía nacional y el mejoramiento de la calidad de vida de los ciudadanos, buscando vincular las instituciones de investigación, empresas y otras entidades para el desarrollo de proyectos de investigación aplicada y desarrollo tecnológico.

Cuenta con distintos tipos de financiamiento [15]. El primero de ellos es “Investigación y Desarrollo en Acción” o IDeA, el cual financia proyectos de I+D aplicada con alto contenido científico, los cuales en un periodo relativamente breve puedan convertirse en nuevos productos, procesos o servicios que impacten al país en ámbitos económicos y sociales.

El segundo es el Programa Valorización de la Investigación en la Universidad o VIU, que promueve el emprendimiento apoyando proyectos nacidos a partir de investigaciones realizadas por estudiantes de pre y postgrado en las universidades del país.

Finalmente, el Programa Fondef Regional, que busca contribuir al desarrollo científico, tecnológico y de innovación en regiones.

CONICYT: FONDECYT

El Fondo Nacional de Desarrollo Científico y Tecnológico es un programa de CONICYT que busca estimular y promover el desarrollo de investigación científica y tecnología básica, siendo el principal fondo de este tipo en Chile. Cuenta con tres instrumentos de apoyo [16].

Con orientación a investigaciones con trayectoria “Proyectos Regulares” contempla financiamiento y apoyo de personal e infraestructura para los distintos proyectos, los cuales deben tener una duración entre dos y cuatro años.

El segundo instrumento son los “Proyectos de Iniciación en Investigación”, el cual busca investigadores jóvenes que hayan obtenido el grado de doctor en los últimos años, obteniendo financiamiento para personal técnico y tesis.

Finalmente se encuentra “Proyectos de Postdoctorado”, que al igual que el anterior busca investigadores que hayan obtenido el grado de doctor en los últimos años, pero en este caso se debe contar con un investigador patrocinante y una institución patrocinante.

Un punto importante de estos instrumentos es que los proyectos adjudicados deben ser realizados en el país.

CORFO: InnovaChile

El Comité Innova Chile es un organismo dependiente de la Corporación de Fomento de la Producción del Gobierno de Chile (CORFO), el cual busca apoyar la productividad del país utilizando

un *set* de instrumentos para promover el proceso de creación de ideas para agregar valor a las empresas en el futuro. Algunas de las herramientas que utiliza son: “Empaquetamiento Tecnológico para Nuevos Negocios”, Concurso de Empaquetamiento y Transferencia de I+D”, “Gestión de la innovación”, “Concurso de Proyecto de I+D Aplicada”, “Innovación en Productos y Procesos Basados en I+D (Prototipo)”, entre otros, para los cuales otorga financiamientos a las distintas empresas, instituciones, centros, universidades o emprendedores.

Sumados a estos se encuentran otros programas de CONICYT y CORFO como FONDAP, FONDECYT, PAI y PCI entre otros.

1.1.3. Gestión de Datos Científicos

A medida que la gestión de datos científicos fue tomando fuerza e importancia en la comunidad científica, se han realizado distintos proyectos y creado instituciones encargadas de promover esta actividad que se enmarca en las buenas prácticas de la investigación, algunos casos ejemplares son los siguientes:

Data Curation Centre UK

El Data Curation Centre o DCC fue creado gracias a la recomendación de la estrategia de acceso continuo y preservación digital de la compañía inglesa sin fines de lucro *Joint Information Systems Committee* (JISC) el año 2002, con el fin de establecer un centro nacional para resolver los desafíos que traía la curación de datos.

Gracias a esto, el año 2004 y liderado por la Universidad de Edimburgo del Reino Unido y otras instituciones y universidades del mismo país fue lanzado el DCC. Desde sus inicios hasta la actualidad ha incentivado el uso de buenas prácticas tanto a nivel país como internacionalmente, siendo un ejemplo para otros proyectos relacionados a la gestión de datos de investigación.

Su plataforma web cuenta con material de apoyo y servicios para todo aquel que desee usarlo y su lema es “Porque una buena investigación requiere de buenos datos”, haciendo referencia a la importancia que tiene la gestión de datos durante todo el ciclo de vida de estos tanto para las universidades como para el mundo científico.

Australian National Data Service

El Australian National Data Service o ANDS fue fundado por el gobierno de Australia a través de la Estrategia de Colaboración Nacional de Infraestructura de Investigación (NCRIS en inglés) el 2008 y gracias a los Fondos de Inversión Educativa (EIF) del gobierno desde mediados del año 2009. [17]

Busca apoyar los desafíos de almacenamiento y gestión de los datos de investigación generados en Australia, además de hacerla disponible y accesible para su revalidación y posterior re-uso [18].

Para esto presta apoyo para el desarrollo de políticas, trabaja en proyectos relacionados a la gestión de datos con distintas universidades y cuenta con un servicio para visibilizar de mejor manera las investigaciones y los datos generados en su país a través de la plataforma *Research Data Australia*.

Unión Europea

En la Unión Europea (EU) se han desarrollado diversos programas relacionados a la investigación y con influencia de la gestión de datos científicos, dos de ellos son el proyecto RECODE y HORIZON2020.

Horizonte 2020 es el más grande de los programas de investigación e innovación de la Unión Europea, es un instrumento de financiamiento con el fin de asegurar la competitividad global de Europa. Cuenta con el respaldo político de líderes europeos, los cuales estuvieron de acuerdo en que la investigación es una inversión para el futuro [19]. Cuenta con guías referentes a la gestión de datos científicos y uno de sus proyectos más grandes y conocidos es OpenAire.

Por otro lado, HORIZON2020 presenta una guía sobre qué debe contener un plan de gestión de datos en las investigaciones para que los datos sigan los principios FAIR (por sus siglas en inglés), que significa que sean hallables, accesibles, interoperables y re-utilizables [20]. Los principios "FAIR" fueron desarrollados por miembros de la comunidad FORCE11, estos cuatro fundamentos tienen el objetivo de guiar tanto a los productores como editores de datos para maximizar el valor agregado que se obtiene al realizar publicaciones académicas digitales [21].

RECODE o *Policy RECommendations for Open Access to Research Data in Europe* utilizando las redes, comunidades y proyectos existentes hace frente a los desafíos relacionados al acceso abierto a los datos referentes a la disponibilidad, difusión y preservación de estos. Este proyecto generó un documento de recomendaciones para el diseño de políticas de gestión de datos enfocada en el acceso abierto a estos basándose en las buenas prácticas para ello [22].

Proyecto LEARN

El proyecto LEARN o *LEaders Activating Research Networks* al igual que RECODE y HORIZON2020 es un proyecto de la Unión Europea, pero algo a destacar de él es que una de sus organizaciones colaboradoras es la Comisión Económica para América Latina y el Caribe (CEPAL), la cual tiene su sede en Chile.

Este proyecto, con el fin de crear una infraestructura tecnológica que una Europa y otros continentes, busca tomar el *roadmap* para los datos científicos, creada por la Liga de Universidades de Investigación Europea (LERU) para entregar un modelo de políticas de gestión de datos científicos y herramientas para su implementación junto instrucciones en cinco idiomas para asegurar una mayor cobertura [23].

En Octubre del presente año se desarrolló un Congreso en Chile donde expusieron expertos en la generación de conocimiento acerca de la gestión de datos o RDM por sus siglas en inglés, con el fin de buscar estrategias y alianzas para su desarrollo.

Chile

Durante el año 2010, la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) en conjunto con la Universidad de la Frontera (UFRO) realizaron un estado del arte respecto a la gestión de datos de investigación en el país.

A través del Sistema de Información Científica de CONICYT, y en conjunto con distintos profesionales, se generó un documento a partir de este estado del arte que buscó generar directrices para políticas de datos con un enfoque de acceso abierto, a pesar de que hubo un gran acercamiento la iniciativa no continuó trabajando y quedó en espera a un mayor desarrollo del área en el país.

Por otro lado, en la Universidad de Chile se está realizando un proyecto enfocado en el acceso abierto a los datos e información científica, enfocado a la información bibliográfica relacionada con la universidad, para esto, la Dirección de Servicios de Información y Bibliotecas de la universidad (SISIB) lanzó la plataforma web "datos.uchile.cl" el presente año (2016) [24]. Por otro lado y siguiendo el foco del acceso abierto a los datos, actualmente la SISIB está pensando en construir políticas de datos y un repositorio, proyecto que está recién iniciando y aún no existe información oficial sobre los enfoques que seguirá.

Además de todos los proyectos, iniciativas e instituciones nombradas, existen otros más por todo el mundo. La gestión de datos científicos ha tomado importancia, la adopción de estas buenas prácticas en instituciones y países ha ido en alza y se están llegando a acuerdos entre todos sobre cómo debería llevarse a cabo, permitiendo una mayor colaboración entre todos para lograr una gestión de datos efectiva, la cual es sólo un parte de la de meta de asegurar o garantizar el acceso continuo a los resultados de la investigación académica [25].

1.2. Descripción y Justificación del Proyecto

Actualmente el Web Intelligence Centre (WIC) perteneciente a la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile se encuentra con una baja disponibilidad de sus bases de datos generadas en los experimentos realizados dentro de uno de sus proyectos, AKORI.

El principal problema de esto es que al querer continuar alguna de las investigaciones ya realizadas, se debe replicar el experimento que se desarrolló en ellas para poder generar de nuevo los datos utilizados, los cual en general son imposibles de realizar con exactitud, llevando costos tanto monetarios, por las herramientas tecnológicas utilizadas, como de horas persona para el centro y en caso de que no se busque una manera de que esta baja disponibilidad de los datos no ocurra para nuevas investigaciones, los costos seguirán aumentando.

La pérdida de los datos de investigación es algo común y se da por distintas razones [26] y en todas las áreas de la investigación, un claro ejemplo es el área de la psicología, en donde existe el problema de baja disponibilidad de los datos utilizados en sus estudios [27]. Por otro lado, existe una correlación negativa entre los años desde que una investigación fue publicada y la probabilidad de encontrar los datos que se utilizaron para realizarla si sólo los investigadores son los encargados de guardarlos [28].

Para llevar a cabo un análisis de lo que sucede en el centro referente a los datos se entrevistó a los estudiantes que realizaron sus trabajos de título en el proyecto AKORI (Anexo A.3: Entrevistas) y se obtuvieron los siguientes resultados: no existen protocolos sobre qué hacer con los datos generados, no se realiza entrenamiento o no hay conocimiento sobre planes de manejo de datos ni mucho menos se tuvo la obligación o recomendación de hacerlos. Finalmente, la mayoría utiliza espacios personales para almacenar sus datos, los cuales no son respaldados en algún repositorio de datos otorgado por el centro, lo más cercano es una copia de seguridad que realizó uno de los alumnos de manera personal, para su propio uso, no un requerimiento del WIC.

Por otro lado, la gestión de datos científicos o *research data management* ha sido un foco importante para la comunidad científica e instituciones de educación respecto a la investigación, un caso ejemplar es el de Australia en donde el año 2007 se desarrolló el Código Australiano [29] para una “Conducta Responsable de Investigación con un foco especial en la gestión de los datos generados, el cual fue firmado por distintas instituciones educacionales del país para comprometerse con las buenas prácticas de investigación. Los beneficios que trae consigo la gestión de datos producidos en investigación son diversos, como aumentar la disponibilidad y reusabilidad de los datos, compartir los resultados, cumplir los requerimientos de las instituciones, financiadores y sobre temas éticos a la hora de realizar experimentos, mitigando riesgos y aumentando la reputación de la institución en donde exista una gestión clara de datos científicos [30], además de poder organizar y agrupar la información obtenida al final de cada investigación de manera más simple y ordenada [5].

Con esto último en mente, se busca proponer el diseño de una arquitectura de gestión de datos de investigación para el centro que permita ofrecer una solución efectiva a la baja disponibilidad de los datos generados en futuras investigaciones generadas dentro de él, teniendo toda la información recolectada y centralizada en un sólo lugar. Además, permitiría solucionar el problema de la poca estandarización en procesos realizados, algo que actualmente se busca en el WIC.

Por otro lado, uno de los objetivos del centro es publicar en las principales revistas relacionadas a su área de investigación. Revistas como JETWI mencionan dentro de sus declaraciones sobre la ética y malas conductas en las publicaciones, que los autores deben hacer disponible los datos primarios utilizados hasta 10 años después de haber publicado su trabajo [31], además de otras revistas como PLOS ONE, Nature y Science que dentro de sus políticas requieren también tener los datos necesarios para las validaciones de las publicaciones disponibles [32], esto hace ver que las revistas importantes están adoptando de a poco la cultura de requerir los datos relacionados a las publicaciones que realizan, lo que con una arquitectura de gestión de datos se podría cubrir, pues es uno de sus objetivos.

Volviendo a la poca estandarización de procesos, algunos de estos pueden no solo traer problemas de pérdidas de datos, si no también temas legales que pueden implicar hasta la pérdida de fondos por no cumplir algunos requerimientos relacionados a experimentos realizados en investigaciones, principalmente con humanos.

Para esto se pretende diseñar Políticas de Gestión de Datos Científicos para el Web Intelligence Centre, en donde queden plasmados los requerimientos del centro para que no vuelva a ocurrir lo sucedido con las investigaciones anteriores. Por otro lado, se desarrolla una propuesta de infraestructura tecnológica para el almacenamiento a corto y largo plazo de los nuevos datos generados en el centro, propuesta que debe cumplir las necesidades y exigencias de las políticas para finalmente diseñar el soporte necesario para que los integrantes del centro puedan tanto seguir las políticas

diseñadas como utilizar la infraestructura propuesta, entregando guías y *templates* con este fin. Con la arquitectura ya diseñada, se realiza una validación de ésta, enfocándose en los dos últimos puntos, con el fin de estudiar la factibilidad de que la organización la utilice y dar propuestas para su implementación de la arquitectura en el WIC.

Como es la primera vez que se realizaría una gestión de datos de manera estructurada en el centro, se entrega también una propuesta para realizar una evaluación de esta gestión en el centro, con el fin de analizar su funcionamiento y verificar que se esté utilizando de manera adecuada el repositorio y cumpliendo todas las reglas enmarcadas en las políticas diseñadas en caso de ser implementada.

Sumado a esto y tomando en cuenta el llamado de la OECD de abrir los datos creados en investigaciones financiadas por fondos públicos [33], se busca enfocar este trabajo al acceso abierto a los datos de investigación debido al carácter público tanto de la Universidad de Chile como de los fondos asignados a todos los proyectos realizados en el centro, en donde existe el problema mencionado.

Al abrir los datos, las investigaciones se hacen reproducibles, generando nuevas interrogantes y produciendo con esto un avance en la investigación e innovación de la ciencia, además de permitir una mayor verificación por parte de otros científicos [34], algo que permitiría reducir la cantidad de investigaciones que han sido anuladas por errores cometidos dentro de estas [35]. Por otra parte también se ha demostrado con los años que dejar abiertos los datos aumenta la tasa de citación de las investigaciones realizadas [36].

Finalmente se considera que siguiendo la visión del centro, sobre ser líder a nivel mundial en investigación, es necesaria la adquisición de las buenas prácticas reconocidas a nivel mundial sobre ella, permitiendo no sólo mejorar internamente el cómo se lleva la investigación, disminuyendo los problemas que trae consigo no tenerlas, sino que también su visibilidad al exterior, permitiendo abrir nuevas oportunidades de cooperación para nuevas investigaciones con la comunidad científica, además del reconocimiento por llevar a cabo nuevos descubrimientos de una manera comprometida con la investigación e innovación tanto dentro de la universidad como del país.

1.3. Objetivos

Este trabajo de título tiene un objetivo general y cuatro objetivos específicos, los cuales son:

1.3.1. Objetivo General

Diseñar una propuesta de arquitectura de gestión de datos de investigación para el Web Intelligence Centre enfocada en el acceso abierto a los datos, con el fin de prevenir la falta de disponibilidad de datos generados en el centro en trabajos futuros".

1.3.2. Objetivos Específicos

1. Analizar las buenas prácticas referentes a los datos científicos de distintas organizaciones a nivel mundial con y sin enfoque en el acceso abierto.
2. Diseñar Políticas de Gestión de Datos Científicos para el WIC.
3. Definir propuesta para el almacenamiento de datos generados en el futuro en el WIC.
4. Diseñar protocolos y el soporte necesario para el cumplimiento de las políticas de gestión de datos científicos del WIC.

1.4. Metodología

Los pasos que se siguieron para llegar a los objetivos planteados fueron:

Primero se diagnosticó el problema tanto del proyecto AKORI como del centro en su totalidad sobre la baja disponibilidad de las bases de datos generadas en experimentos anteriores, para esto se realizaron entrevistas a las personas que desarrollaron algún experimento durante su estadía en el proyecto. Luego se revisaron las respuestas recogidas para este diagnóstico.

Posterior a esto se estudió el área de gestión de datos de investigación para buscar un enfoque a largo plazo en donde, para las nuevas investigaciones, esto no vuelva a suceder. Como no existen metodologías claras creadas para crear una arquitectura de gestión de datos científicos, pero sí existe un énfasis en la creación de políticas para las instituciones, se procedió de la siguiente manera:

Se examinaron los puntos importantes que presentan las distintas organizaciones científicas a nivel mundial respecto a cómo crear políticas que sustenten un manejo responsable de los datos, contando también las que contienen un enfoque de acceso abierto a estos. Una vez revisadas y comparadas estas guías de contenidos se revisaron distintas políticas formuladas en base a ellas en distintas universidades del mundo con el fin de poder comparar su estructura y tomar algunas como base para la formulación de estas en el WIC.

Como tampoco existen metodologías claras de cómo diseñarlas y con el fin de crear finalmente el documento que contenga las políticas de gestión de datos científicos para el centro, se tomó como guía una serie de pasos diseñados por el DCC del Reino Unido, los cuales son simples y fueron creados para servir de guía en las instituciones que están realizando esta acción por primera vez, estos cinco pasos son mencionados por el proyecto RECODE dentro de sus recomendaciones [37] y el proyecto LEARN, entre otros.

Debido a que es primera vez que en el centro se busca crear una arquitectura de gestión de datos de investigación, por lo que se requiere algo tanto simple de seguir como de implementar, se utilizaron estos cinco pasos entregados por el DCC UK.

El primer paso fue realizar un consolidado del estado actual sobre la gestión de datos científicos en el centro, para luego continuar con la creación de la tabla de contenidos, donde se utilizó la com-

paración mencionada anteriormente. El siguiente paso fue revisar la ruta de aprobación de políticas en el centro y luego continuar con el diseño del primer borrador del documento, su validación y finalmente de la versión final.

El segundo pilar de la arquitectura de gestión de datos de investigación tiene que ver con la infraestructura TI, la que debe poder cumplir todos los requisitos impuestos por las políticas que se desea diseñar. Entonces, una vez diseñadas las políticas, se continuó con un levantamiento de requerimientos técnicos del lugar para almacenar los datos en el futuro, ya sea por términos de capacidad de espacio o de cumplir las políticas recién diseñadas, esto con el fin de realizar un benchmarking de los sitios, *software* y herramientas disponibles en el mercado principalmente promovidas por las iniciativas y/o proyectos a nivel mundial sobre la gestión de datos científicos.

Una vez realizado el benchmarking, dentro de las soluciones factibles a nivel técnico encontradas se eligió la que contenía más ventajas a nivel técnico como solución, la cual fue un software especializado en el diseño de repositorios de datos científicos y con este se diseñó el repositorio que satisface los requisitos de manera interna, es decir, en un servidor local (no online), siendo esta la propuesta final de almacenamiento. En primera instancia se buscó una solución ya existente, es decir un repositorio ya creado, al cual se le hubiesen tenido que aplicar todas las configuraciones que permitan el cumplimiento de los requerimientos, lo cual no sucedió. Por otro lado, en el caso de no haber existido soluciones factibles, se hubiese procedido a construir un prototipo no funcional del repositorio de datos.

Finalmente y como se menciona anteriormente, es primera vez que se busca realizar una acción de este tipo en el centro de investigación. El soporte a los miembros de las instituciones es necesario, ya que si no existe, puede hacer fallar los intentos por mejorar las prácticas de investigación de enfoque abierto en los distintos lugares que lo intentan [38]. Con el fin de fomentar el buen cumplimiento de las políticas y el uso de la infraestructura TI, se diseñaron todas las guías y pautas necesarias para ello, contando con distintos documentos y *templates* con este objetivo. Sumado a esto, también se hace necesario la creación de un mecanismo de evaluación del cumplimiento de toda la arquitectura diseñada para realizar un seguimiento efectivo de su utilización, a pesar de que las políticas en general cuentan con un seguimiento y monitoreo propio, no abarcan el conjunto de los pilares mencionados. Como se busca la pronta implementación de la gestión de datos científicos en el WIC, se diseñó una pauta de evaluación enfocada tanto en el cumplimiento de las políticas como de la correcta utilización de la infraestructura TI elegida.

Una vez diseñada la arquitectura, se procedió a realizar una validación de esta, realizando un test de usabilidad para ver qué tan útil y fácil de usar es la propuesta de almacenamiento escogida junto a las guías diseñadas, para terminar con propuestas de mejoras a ella y el cómo llevar a cabo su implementación en el centro.

1.5. Alcances

El trabajo se realiza específicamente para el Web Intelligence Centre (WIC), tanto las Políticas de Gestión de Datos como el Prototipo no funcional del Repositorio de datos en caso de ser necesario y las guías y documentos que se generen para su cumplimiento.

Respecto a los datos, se trabaja específicamente con los datos primarios de investigación, es decir generados en el mismo centro de investigación.

Respecto al almacenamiento de los datos se desarrolla la configuración necesaria para el cumplimiento de los requerimientos de las políticas de gestión de datos diseñadas en este trabajo a la plataforma escogida como solución, en caso de que no encontrar una, el alcance será un prototipo no funcional de repositorio de datos para el centro.

Finalmente, las guías, documentos y evaluaciones que se generan con el fin de poder seguir las políticas diseñadas serán exclusivas para ellas y el lugar de almacenamiento seleccionado.

1.6. Resultados Esperados

Los resultados esperados de este trabajo de título tienen una directa relación con los objetivos anteriormente nombrados.

Para el primer objetivo específico se pretende obtener el análisis de las guías para la formulación de Políticas de Gestión de Datos Científicos de una institución.

Para el segundo objetivo se busca consolidar el estado actual del centro de investigación respecto a la gestión de los datos generados en él, el diseño del borrador de las Políticas de Gestión de Datos para su validación y finalmente la versión o documento final de estas.

Para el tercer objetivo específico se espera obtener los requerimientos técnicos para un repositorio de datos científicos para el centro que permita cumplir las necesidades otorgadas por las políticas y el benchmark de las soluciones existentes en el mercado. Una vez obtenida la solución escogida se realizará la configuración del repositorio escogido, sea un repositorio existente, un software para el diseño de repositorios de datos científicos o un prototipo no funcional en caso de que no existan soluciones factibles encontradas.

Finalmente, para el último objetivo específico se desea diseñar distintas guías y *templates* de distintos documentos con el fin de que los integrantes del centro puedan seguir las políticas y desarrollar una gestión de datos de investigación efectiva.

Con esto entonces se espera cumplir el objetivo general de esta memoria; realizar una propuesta de arquitectura de gestión de datos científicos en el WIC para prevenir la baja disponibilidad de los datos generados en sus experimentos.

1.7. Estructura de Memoria

La estructura del presente trabajo de título es la siguiente:

En el presente y primer capítulo se presenta una introducción de la memoria, mostrando el contexto y antecedentes relacionados a la investigación, el centro en donde se realiza la memoria

y la gestión de datos científicos. Por otro lado, se expone el objetivo general de este, junto a los objetivos específicos y la metodología utilizada para sus cumplimientos. Finalmente se especifica el alcance dentro del centro de investigación donde se trabaja junto a los resultados esperados.

El segundo capítulo detalla el marco teórico del trabajo con el fin de proveer al lector la información necesaria para el entendimiento del trabajo que se desarrolla en los siguientes capítulos de este informe, principalmente detallando los conceptos relacionados a la gestión de datos como datos de investigación, plan de gestión de datos, metadatos y políticas de gestión de datos científicos entre otros.

El tercer capítulo muestra el diseño de las Políticas de Gestión de Datos Científicos del centro de investigación, mostrando los pasos utilizados con este fin, los cuales son la presentación de un estado actual del centro respecto a la gestión de datos científicos, un desarrollo de la tabla de contenido del documento, revisión de la ruta de aprobación de estas políticas, el diseño del borrador del documento, para terminar con la versión final de las políticas, explicando en detalle cada uno de ellos.

El cuarto capítulo presenta la solución propuesta de almacenamiento para el centro, junto a los pasos que se utilizan para llegar a esta, los cuales son; el levantamiento de requerimientos para el repositorio, el análisis de las posibles soluciones y finalmente la elección.

En el quinto capítulo se desarrollan los documentos que sirven de guía a los integrantes del centro para cumplir los requerimientos de las políticas diseñadas. Se desarrollan tres guías principales, sobre el uso de los metadatos requeridos, las licencias a asignar y sobre el uso del repositorio. Además se entregan las plantillas para los documentos requeridos de acuerdo a las políticas diseñadas y una esquematización de los procesos que serán necesarios de llevar a cabo en el centro.

En el sexto capítulo se desarrolla una validación para la arquitectura de gestión de datos científicos diseñada en los capítulos anteriores, enfocada en la propuesta de almacenamiento y las guías desarrolladas. También se analiza el cómo llevar a cabo la implementación de esta en el centro.

En el séptimo y último capítulo de este informe, se realizan las conclusiones del presente trabajo, agregando una discusión y propuestas de trabajo futuro.

Capítulo 2

Marco Teórico

2.1. Datos de Investigación

Existen distintas definiciones para datos de investigación (*Research Data*), o datos científicos, de acuerdo a las políticas de gestión de datos de investigación de la Universidad de Oxford [39] una de ellas es “la información recolectada (sin importar la forma o medios en que pueda existir) necesaria para validar las observaciones, conclusiones o resultados de un proyecto de investigación”. Estas validaciones en general se hacen utilizando software y herramientas estadísticas dependiendo del tipo de investigación.

Los datos científicos pueden ser generados para distintos propósitos y a través de distintos procesos, por ejemplo puede ser de observación (generalmente datos capturados en tiempo real), experimental (datos generados por equipos de laboratorios), simulaciones, entre otros [40].

De acuerdo a la OECD, los datos de investigación se definen como registros de hechos (puntajes numéricos, registros textuales, imágenes, sonidos) usados como fuentes primarias para la investigación científica, donde son aceptados en la comunidad científica como necesarios para validar los hallazgos en las investigaciones [33]. Principalmente estos hallazgos son presentados a través de publicaciones en revistas científicas en forma de *papers*.

2.2. Ciclo de Vida de los Datos de Investigación

De acuerdo al centro nacional especializado en el archivado de datos (UK Data Archive) presenta el siguiente diagrama de ciclo de vida de los datos producidos por investigaciones [41].

1. Creando los datos: Diseño de la investigación, plan de manejo de datos, plan para compartir datos, localizar datos existentes, coleccionar los datos (experimentos, simulaciones, mediciones, etc.), capturar y crear la metadata.
2. Procesando los datos: Analizar los datos, digitalizarla, transcribirla, trasladarla, validarla, lim-

piarla, anonimización si es necesario, descripción de los datos, gestión y almacenamiento de ella.

3. Analizando los datos: Interpretarlos, producir resultados, publicaciones, preparar los datos para preservarlos.
4. Preservando los datos: Transformar al mejor formato, transportarla a un medio que se ajuste, respaldar y almacenar los datos, creación de metadatos y documentación, archivar los datos.
5. Dando acceso a los datos: Distribución de los datos, compartir los datos, controlar los accesos, establecer derechos de autor y promocionar los datos.
6. Re-usando los datos: Realizar un seguimiento a la investigación, nuevas investigaciones, encargarse de revisiones de la investigación, examinar los resultados, aprender y enseñar.



Figura 2.1: Ciclo de Vida

Fuente: UK Data Archive

2.3. Acceso Abierto

Existen distintos intentos de cómo definir formalmente el acceso abierto (Open Access (OA)), Alma Swan muestra las tres declaraciones en donde se define OA [42].

La primera es la realizada por Budapest Open Access Initiative (BOAI). En esta iniciativa, el acceso abierto se define como una forma de accesibilidad a la literatura científica en revistas de manera gratuita, on-line y sin restricciones para todos los científicos, estudiantes, maestros, escolares y otras mentes curiosas. Para que esto sea posible, BOAI describe dos formas de cómo hacerlo: La primera es utilizando el autoarchivado (depositar copias de los artículos en archivos de acceso abierto o la ruta verde como es conocida) y la segunda es mediante publicaciones en revistas

de acceso abierto (donde estas hacen que sus contenidos sean libremente accedidos en la web, a esto se le conoce como la ruta dorada").

La segunda es la Declaración de Bethesda, la cual fue construida en base a BOAI. En esta se define que una publicación de acceso abierto satisface dos condiciones: La primera es que los autores y propietarios de los derechos de autor garantizan a todos los usuarios el acceso y uso de la publicación sujeto a una atribución a estos. La segunda es que una versión completa del trabajo (junto a todo el material suplementario incluyendo el permiso para su utilización) en un repositorio respaldado por alguna organización bien establecida que busque habilitar el acceso abierto a largo plazo.

La tercera declaración, es la Declaración de Berlin Sobre el Acceso Abierto al conocimiento en Ciencias y Humanidades, la cual es en esencia define el acceso abierto igual que la de Bethesda, pero además recomienda a las instituciones implementar políticas para exigir a los investigadores a depositar una copia de sus artículos publicados en un repositorio de acceso abierto y además para estimular a estos a publicar en revistas con este foco.

Estas tres definiciones conjuntas se conocen como “la definición BBB del Acceso Abierto” es una de las más aceptadas a pesar de que han existido más intentos de cómo definirlo.

Por otro lado también Peter Suber dice que la literatura que cuenta con acceso abierto es la que es digital, online, libre de costos y de la mayoría del *copyright* o restricciones por licencias. También menciona que lo que hace esto posible es el internet y el consentimiento de los autores y/o poseedor de los derechos de autor. Definición que también es influenciada por la definición BBB. Suber también menciona los dos caminos del acceso abierto, la ruta dorada y la ruta verde peter [43].

2.4. Ciencia Abierta

Open Science o “Ciencia Abierta” de acuerdo a la Organización para la Cooperación y Desarrollo (OECD) [44] se refiere a un enfoque de la investigación basada en un mayor acceso a los datos de la investigación pública, ayudado por herramientas y plataformas tecnológicas de información y comunicación, además de una colaboración más amplia en la ciencia incluyendo a los no científicos y por otro lado el uso de herramientas alternativas de derechos de autor para difundir los resultados de la investigación. Tiene el potencial de mejorar la eficiencia y calidad de la investigación reduciendo los costos de la recolección de datos, facilitando la explotación de los datos que no están siendo utilizados o son poco accesibles a bajo costo y también aumentando las oportunidades de colaboración tanto en investigación como en innovación. Esta cooperación y acceso a datos puede ayudar a solucionar problemas globales como el del cambio global y también puede motivar el desarrollo de capacidades en los países en vías de desarrollo a través de las oportunidades que generaría la colaboración entre países.

Existen distintos proyectos relacionados al movimiento de la ciencia abierta, uno de los grandes proyectos relacionados a la ciencia abierta es NOVA en Argentina y ALMA en Chile.

NOVA o Nuevo Observatorio Virtual Argentino: En el cual se busca coordinar las iniciativas de

distintas instituciones a lo largo de Argentina con el fin de lograr una mayor eficiencia y productividad en el acceso, manejo y análisis de las observaciones astronómicas capturadas. Esta información almacenada es pública y puede ser utilizada tanto por científicos del área como por el público en general [45].

ALMA o Atacama Large Millimeter/ submillimeter Array: es una asociación internacional entre el Observatorio Europeo Austral (ESO), la Fundación Nacional de Ciencia de EE.UU (NSF) y los Institutos Nacionales de Ciencias Naturales de Japón (NINS), junto con NRC (Canadá), NSC y ASIAA (Taiwán), y KASI (República de Corea), en cooperación con la República de Chile. Es el mayor proyecto astronómico que existe en el mundo [46].

2.5. Gestión de Datos generados en investigación

De acuerdo a la Universidad de Boston la gestión de datos o RDM (*Research Data Management*) es un área esencial de una investigación responsable. Con gestionar los datos se puede asegurar que los datos de investigación sean precisos, completos, auténticos y confiables, incrementar la eficiencia de la investigación, asegurar la seguridad de los datos y minimizar el riesgo de perderla, ahorrar tiempo y recursos a largo plazo respecto a los procesos científicos y estar al día respecto a las prácticas de los demás investigadores en cuanto al manejo de datos en la actualidad [47]. Está insertado en las buenas prácticas de investigación referente a los datos generados.

2.6. Arquitectura de Gestión de Datos Científicos

Existen pocas entidades que definen una arquitectura de gestión de datos (*Data Management Framework*), el Servicio Nacional de Datos Australiano (ANDS) lo explica como los elementos básicos en una institución para apoyar una eficaz gestión de datos científicos, el cual cuenta con cuatro categorías o pilares dentro de ella: Políticas institucionales y procedimientos, Infraestructura TI, Servicios de soporte y gestión de metadatos [48].

A pesar de que para ANDS son cuatro los pilares fundamentales, para el caso de esta memoria se considerarán solo tres, en los cuales estarán englobados los cuatro anteriormente mencionados. Esto se hace con el fin de simplificar su explicación, además de que se considera que el cuarto elemento, sobre metadatos, puede ser cubierto por las políticas de gestión y el soporte necesario, ya que está enfocado en generar metadatos con el fin de que los datos puedan ser utilizados en el futuro tanto por gente interna como externa de la situación, acción que debe estar declarada en las políticas ya que es una de sus misiones, por otro lado el cómo hacerlo debe ser explicado en el elemento de soporte de la arquitectura.

En la figura 2.2 quedan plasmados los tres pilares o elementos que se consideran necesarios para realizar una gestión de datos efectiva.

Como se puede apreciar, los tres se conectan entre sí y van de la mano para cumplir su misión. De las políticas salen los requerimientos de infraestructura TI, la cual debe permitir ejecutar estas

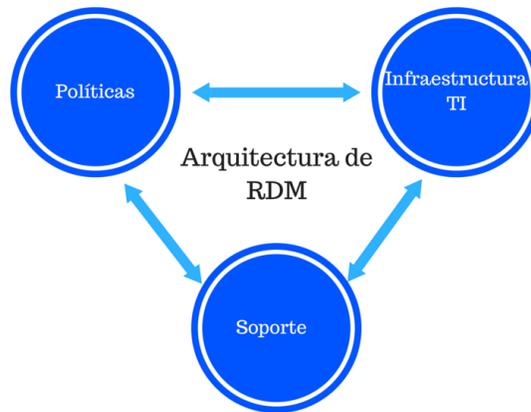


Figura 2.2: Arquitectura de RDM

Fuente: Elaboración propia

políticas, para lo cual es necesario de un apoyo para los integrantes de la institución tanto para utilizar esta infraestructura tecnológica como para seguir los demás puntos que se consideran dentro de las políticas.

2.7. Plan de gestión de datos

En el programa de la Unión Europea (EU) Horizonte 2020 definen al plan de manejo de datos (DMP) como un documento en donde se muestra cómo serán manejados los datos científicos durante y después de cada proyecto. Debe describir el ciclo de vida para todos los conjuntos de datos que serán recolectados, procesados o generados por el proyecto de investigación [49]. Este documento puede ser pedido antes, durante o después de una investigación, lo cual va a depender de los requerimientos de los financiadores o de la institución en donde se desarrolle. Horizon2020 también presenta una guía sobre qué debe contener un Plan de Gestión de Datos para que se cumplan los principios FAIR, además de un template en los anexos de esa guía [20].

Por otro lado, en Argentina, de acuerdo a la ley N° 26.899, todos los datos depositados en repositorios científicos debe ser acompañado de un plan de gestión de datos, el cual debe contener los siguientes elementos como mínimo [50]: Creadores; Identificación del Proyecto de investigación; Identificación de la agencia u organismo de financiamiento que financia la investigación; Tipología de los datos que se generarán y recopilarán durante el proyecto; Referencia, nombre del conjunto de datos e identificador único; Estándares que se utilizarán en su colecta, manipulación y registro; Descripción general del conjunto de datos y de los datos que serán generados o recopilados; Cómo serán explotados y/o compartidos/accesibles los datos para su verificación, reutilización, redistribución, etc. En cuál/cuáles repositorio/s se alojarán los conjuntos de datos generados y recopilados, períodos de embargo, software necesario y otras herramientas que permitan su reutilización; Condiciones de acceso (licencias de uso); En línea con las posibilidades de excepción establecidas por la Ley y explicitadas en este Reglamento, si los datos no pudieran difundirse, será necesario declarar el motivo; Esquemas de metadatos con que se describirán los conjuntos de datos (de acuerdo al esquema que utilicen los repositorios donde los deposite, respetando siempre las Directrices SNRD);

Medidas de conservación y preservación que se tomarán durante el proyecto y previamente a su depósito en el repositorio; Costos asociados.

2.8. Políticas de Gestión de Datos de Investigación

Para que exista una gestión eficiente de los datos se necesitan políticas en las instituciones de investigación con tal de integrarlos a las comunidades y proyectos internacionales relacionados a sus áreas de investigación. Estas políticas representan las reglas y requerimientos que entrega la institución respecto al manejo y preservación de los datos producidos en investigaciones en ellas.

Existen distintos enfoques y recomendaciones a la hora de diseñarlas, a continuación se presentan algunas de ciertos consejos reconocidos a nivel mundial respecto a ciencia y desarrollo.

2.8.1. CODATA

De acuerdo al Comité de Información para Ciencia y Tecnología (CODATA) del Consejo Internacional para la Ciencia (ICSU) existen distintos puntos importantes que deben estar presentes en una Política de Gestión de Datos, los cuales son los siguientes: [51]

1. Una definición de Datos de Investigación (Research Data).
2. Una visión general de los datos dentro de los alcances de las políticas.
3. Una indicación de criterios generales para la selección de datos de investigación.
4. Un resumen de responsabilidades de cada entidad relacionada.
5. La indicación de la infraestructura disponible y responsabilidad de los costos.
6. Una visión de los requerimientos de los planes de manejo de datos de los investigadores (DPM).
7. Recomendaciones para permitir descubrimiento y re-uso de datos.
8. Estipulaciones para fomentar el reconocimiento y la recompensa para los proveedores de datos.
9. Un resumen de los requerimientos de información, control de cumplimiento y posibles sanciones.

2.8.2. RECODE

Por otro lado, el proyecto RECODE financiado por la Unión Europea, que busca entregar recomendaciones para facilitar el acceso libre a datos de investigación, presenta una guía práctica para

el diseño de políticas para las instituciones de investigación, los puntos importantes a la hora de crearlas son los siguientes [37]:

1. Conocimiento de políticas internacionales.
2. Participación en dialogo y colaboración.
3. Evaluación de la infraestructura existente y requerida.
4. Evaluación de los costos de implementación relacionados a las políticas.
5. El desarrollo del contenido de las políticas.
6. Planes de gestión de datos (DPMs).
7. Orientación a los investigadores.
8. Recompensa a investigadores.
9. Políticas para mecanismos de monitoreo.
10. Además respecto a su contenido debe ser dirigido a los siguientes problemas: Acceso libre como predeterminado, responsabilidades, lugar para depositar datos, tiempo en que serán depositados los datos, licencias.

2.8.3. ANDS

El Servicio Nacional de Datos Australiano (ANDS) también cuenta con una guía para la formulación de políticas de gestión de datos científicos para instituciones, algunos de los puntos importantes que toma en cuenta son los siguientes: Definiciones, excepciones, responsabilidades, periodos de retención de registros y datos científicos, infraestructura de almacenamiento, seguridad y protección, acceso a datos, movimiento y destrucción de datos, propiedad y requerimientos especiales [52].

2.8.4. DCC UK

Finalmente, “The Data Curation Centre” (DCC) del Reino Unido ha realizado ciertas comparaciones de las políticas de manejo de datos en las distintas instituciones del país, los puntos evaluados son los siguientes: Definiciones, apoyo, PGD, alcances, propiedad, financiamiento externo, criterios, tiempos que serán depositados los datos, ética, accesos, costos y revisión de políticas [53].

2.8.5. Diseño de Políticas de Gestión de Datos

Por otro lado, el DCC también presenta una metodología para ayudar al diseño de las Políticas de Manejo de Datos, estos 5 pasos son los siguientes [54]:

1. Revisar y mapear la infraestructura regulatoria existente: Va a depender de las circunstancias, algunos puntos críticos son las leyes regulatorias del país, requerimiento de los financiadores, políticas existentes, revisar los estándares entre las sociedades científicas.
2. Desarrollar una tabla de contenidos: Revisar las instituciones relacionadas con políticas ya creadas, seleccionar los puntos principales que irán dentro de la política (template).
3. Obtener respaldo: Investigar el proceso para aprobar nuevas políticas dentro de la institución.
4. Consultar, escribir borrador, revisar: Averiguar requerimientos antes de empezar a escribir, si no se conocen consultar a los stakeholders.
5. Aprobación e Implementación.

2.9. Metadatos

La definición más común para metadatos es “datos acerca de datos”. A pesar de esto un metadato es en sí un dato, puede ser almacenado y manejado en una base de datos. Un metadato es un dato que es usado para describir a otro, es decir, su uso lo hace un metadato [55].

En el contexto de los datos científicos, los metadatos son la documentación para el conjunto de datos, expresados a través de una sintaxis formal, guardan toda la información necesaria para su uso posterior. Es adjuntado a los datos de investigación usualmente en dos archivos separados, algunos de los puntos que debe contener son los siguientes [56]:

1. Una descripción de los datos.
2. Nombres, etiquetadas y descripciones para las variables y sus valores.
3. Explicaciones de códigos y esquemas de clasificación usados.
4. Códigos y razones para valores perdidos.
5. Descripciones de los formatos de los archivos utilizados para almacenar los datos.
6. Documentaciones de los protocolos experimentales.
7. Documentación del código escrito para análisis realizados.

2.10. Estándares de Metadatos

Según “The Data Curation Centre” (DCC) [57] debido a la importancia de un buen metadato para el acceso y re-uso de los datos de investigación se han realizado iniciativas para formalizar las especificaciones mínimas de los metadatos para las distintas disciplinas en específico.

Por eso, existen distintos estándares para cada área en específico, algunos de los esquemas más conocidos son: DDI - Data Documentation Initiative, para las ciencias sociales, Darwin Core, para la disciplina de la biología, DataCite Metadata Schema, Dublin Core y RDF Data Cube Vocabulary que fue desarrollado sin tener un área en específico. La “Research Data Alliance” (RDA), organización que busca que existan facilidades para compartir las distintas investigaciones a través del mundo [58], cuenta con una lista de algunos de los estándares más utilizados y actuales para distintas áreas de investigación [59].

2.11. Digital Object Identifier

De acuerdo a la Universidad de Chile, Digital Object Identifier (DOI) es una forma de identificar un objeto digital el cual, aunque cambie de URL seguirá teniendo la misma identificación. [60]

Se utiliza principalmente para citar, publicar, buscar y localizar documentos científicos electrónicos.

2.12. Creative Commons

Creative Commons (CC) es una organización sin fines de lucro que permite el reuso y que se comparta (*sharing*) la creatividad y conocimiento a través de entregar herramientas legales de manera gratuita. Desde su versión 4.0 permite aplicar una licencia a conjuntos de datos [61].

Existen 5 tipos de condiciones que se aplican a los conjuntos de datos, los cuales son representados por imágenes que se encuentran en el Anexo B.1, estas condiciones son:

1. BY: Reconocimiento (*Attribution*): la cual obliga a dar atribución o el crédito apropiado enlazando a la página de la licencia.
2. NC: No Comercial (*Non Commercial*): que prohíbe el uso del material para fines comerciales.
3. ND: Sin Derivados (*No Derivatives*): que prohíbe compartir adaptaciones del material.
4. SA: Compartir Igual (*Share Alike*): que requiere que adaptaciones del material sean compartidos con el mismo tipo de licencia de este.
5. CC0: Sin Derechos de Autor: que elimina todo derecho o limitaciones para potenciales usua-

rios del material.

Con estas condiciones, existen siete tipos de licencia que resultan de sus combinaciones [62], que van desde lo más abierto (color verde) a lo más restringido, sumado a esto está el caso extremo en donde están todos los derechos reservados (rojo) el cual no es considerado dentro de las licencias de Creative Commons, el espectro se encuentra en la figura 2.3.

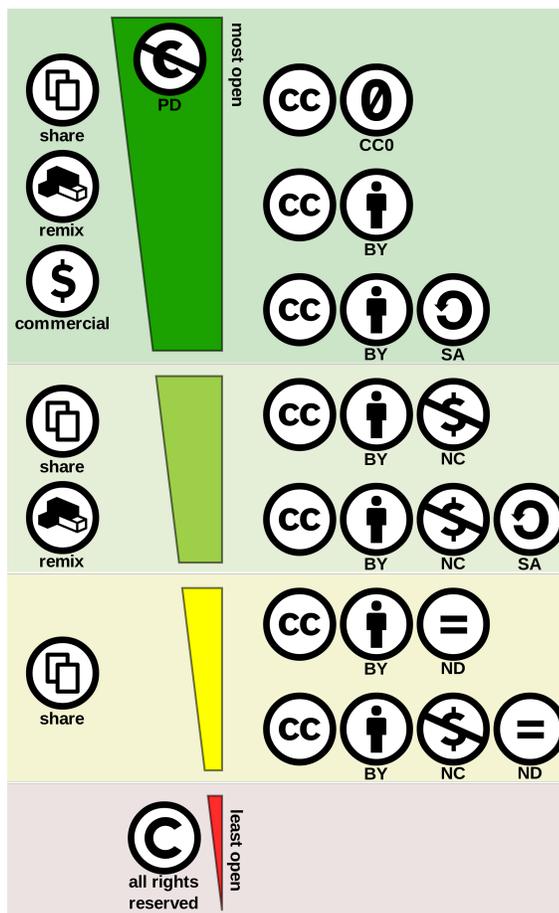


Figura 2.3: Espectro licencias CC

Fuente: Wikipedia

Cada una de estas licencias obliga a los futuros usuarios a seguir distintas condiciones para usar los *sets* de datos. Comenzando desde lo menos restrictivo, los dos primeros tipos de licencia: CC0 y CC - BY implican las condiciones recién explicadas sobre la eliminación de derechos y *copyright* de los datos y la atribución al autor de estos, respectivamente.

La licencia CC-BY SA permite generar nuevas obras a partir del material original sin restricciones, siempre y cuando estas creaciones estén bajo la misma licencia y se atribuya al autor del material utilizado.

La licencia CC BY-ND, mientras se reconozca la autoría de los datos, posibilita su redistribución de cualquier tipo siempre y cuando la obra no se modifique y sea distribuida en su totalidad.

La licencia CC BY-NC permite crear nuevos conjuntos de datos o darle distintos usos siempre

y cuando sea sin fines comerciales, además de reconocer la autoría de estos.

La licencia CC BY-NC-SA es igual a la anterior solo que las nuevas creaciones deben estar bajo la misma licencia que el *set* original.

Finalmente, la licencia CC BY-NC-ND, que es la más restrictiva de las anteriores y sólo permite descargar y compartir el material con otras personas, siempre que se atribuya el autor, pero sin poder manipularlas o usarlas con fines comerciales.

2.13. Open Data Commons

ODC es un proyecto sin fines de lucro que busca entregar soluciones legales a los datos abiertos [63]. Para esto propone tres tipos de licencias [64].

La primera es PDDL (*Public Domain Dedication and License*) que dedica la base de datos y su contenido al dominio público, libre para que todos la puedan usar como deseen.

La segunda es ODC-BY (*Attribution License*) donde los usuarios son libres de usar el material como quieran pero con la obligación de proveer atribución a la fuente de los datos y/o bases de datos.

La tercera y última es (ODC-ODbL) (*Open Database License*), que es la más restrictiva de las tres, en donde se debe dar atribución a la fuente de los datos, pero si existe una nueva versión o uso del material, ésta debe ser accesible y distribuida utilizando el mismo tipo de licencia.

2.14. Test de Usabilidad

Un test de usabilidad es una técnica para el diseño de interacciones centrado en el usuario, que permite a través de ciertas pruebas a usuarios evaluar un producto [65].

2.14.1. Componentes de la usabilidad

La usabilidad, de acuerdo a Nielsen es un atributo de calidad que mide lo fácil de utilizar una interfaz de usuario. Es definida por 5 componentes [66]:

1. Facilidad de aprendizaje (*Learnability*): Qué tan fácil para un usuario es realizar ciertas tareas básicas la primera vez que utilizan la interfaz.
2. Eficiencia: Una vez que los usuarios hayan aprendido el diseño, qué tan rápido pueden realizar ciertas tareas.
3. Facilidad de memorizar (*Memorability*): Cuando los usuarios vuelen a utilizar la interfaz luego

de un periodo de no usarlo, que tan fácil es esta acción.

4. Errores: Cuántos errores el usuario comete, qué tan graves son y qué tan fácil pueden recuperarse de estos.
5. Satisfacción: Qué tan agradable es utilizar la interfaz.

Por otro lado, otro atributo de calidad es la utilidad, que la define como el ofrecimiento de las características que se necesitan, o para lo que está hecha la interfaz de usuario que se desea analizar.

Estos dos atributos son igual de importantes ya que, se puede tener algo fácil de usar, pero no útil o por otro lado algo que ofrece lo que se necesita pero que no es fácil de utilizar, resultando en ambos casos con el abandono de la interfaz.

2.14.2. Procedimiento

Para realizar un test de usabilidad se recomienda la realización de un estudio básico, siguiendo cinco pasos [67]:

Primero, decidir qué es lo que desea aprender del test, para con esto determinar las metas y tareas que se asignarán al usuario.

Segundo, escoger a los participantes y gestionar los horarios, los participantes deben ser el público objetivo de la interfaz a evaluar. Generalmente se necesitan entre 4 y 6 personas para realizar el test. Se deben gastar entre 20 minutos y 2 horas por cada participante.

Tercero, realizar las sesiones, se debe realizar en un ambiente en donde se pueda escuchar lo que están diciendo los participantes. Es importante recalcarles que es el sitio el que está siendo evaluado y no ellos, además de pedirles que no pidan ayuda. Se debe explicar qué se requiere de ellos en un momento previo al test. Este paso tiene relación con *Thinking Aloud*, que es una de las herramientas más utilizadas para la realización de pruebas de usabilidad, en donde se les pide a los participantes pensar en voz alta, lo cual permite descubrir qué es lo que realmente el usuario piensa acerca de la interfaz, además de ser una herramienta barata, robusta, flexible, convincente y fácil de aprender [68].

Cuarto, evaluar los datos. Se debe tomar atención a las áreas donde los usuarios se sintieron frustrados, se demoraron mucho o simplemente no pudieron completar las tareas.

Quinto, se pueden hacer muchas más actividades. A pesar de que estos estudios básicos ayudan, se pueden realizar más acciones.

Se debe tener en cuenta que observar a los participantes utilizando la interfaz da más información que un *focus group* o una encuesta.

Por otro lado, de acuerdo a Rubin y Chisnell, para realizar un test de usabilidad, primero se debe crear un plan de prueba, el cual es la base del test y dirige al cómo, cuándo, dónde, a quién, por qué y qué del test de usabilidad. Las partes de un plan son [69]:

1. Propósito, metas y objetivos del test.
2. Preguntas de investigación.
3. Características de los participantes.
4. Diseño del test.
5. Lista de tareas.
6. Ambiente de la prueba, equipamiento y logística.
7. Rol del moderador del test.
8. Datos a coleccionar y medidas de evaluación.
9. Contenidos del reporte y presentación.

Respecto a la cantidad de participantes, Nielsen considera que se debe utilizar a 5 participantes, con los cuales se pueden encontrar cerca del 85 % [70] de los problemas en la interfaz que se quiera probar, por otro lado, utilizando a 15 participantes se podría llegar a todos los problemas de esta interfaz. Cree además que la utilización de más participantes es un gasto de recursos y tiempo ya que la curva entre 5 y 15 participantes crece de manera muy lenta y poco significativa. A pesar de esto sí cree que para ciertos tests, donde se busca un análisis cuantitativo, categorizar contenidos (*Card Sorting*) o utilizar eye-trackers, es necesario una mayor cantidad de participantes [71].

2.14.3. Tipos de test de usabilidad

Existen distintos tipos de *tests* o pruebas de usabilidad, entre ellos están los *tests* exploratorios, cuyo objetivo es explorar lo que piensan los usuarios de la tarea que se está tratando de facilitar y evaluar si existe intuición en la interfaz, los *tests* de evaluación, que tienen de objetivo determinar la efectividad de la interfaz, los *tests* de validación, que se hacen en la etapa final y evaluar el rendimiento respecto al mercado de la usabilidad del producto, finalmente los *tests* de comparación que sirven para evaluar cualquiera de los dos tests mencionados anteriormente.

Capítulo 3

Políticas de Gestión de Datos Científicos

A pesar del auge en la importancia de la creación de políticas de gestión de datos de investigación en las distintas instituciones existen muy pocas metodologías diseñadas con este fin, aunque sí existen recomendaciones sobre qué puntos claves deberían estar presentes en ellas. A pesar de esto, el Data Curation Centre del Reino Unido ofrece un conjunto de pasos con el fin de facilitar la creación de estas, ya sea para un nivel institucional, de facultad o de departamento, estos pasos se encuentran explicados en el capítulo 2.

Estos cinco pasos son simples y sirven como una guía para las instituciones que están diseñando por primera vez sus políticas de datos científicos, es mencionada por el proyecto RECODE dentro de sus recomendaciones [37] y el proyecto LEARN, entre otros.

Debido a que es primera vez que en el centro se busca crear una arquitectura de gestión de datos de investigación, por ello se espera algo tanto simple de seguir como de implementar, se procede a seguir esta guía entregada por el DCC UK.

En este capítulo se muestra el desarrollo de estos pasos en el diseño de Políticas de Gestión de Datos Científicos para el Web Intelligence Centre de la Universidad de Chile.

3.1. Estado Actual

El primer paso es referente al estado actual en que se encuentra el centro de investigación poniendo énfasis en las regulaciones que existen o se deben cumplir por parte de investigadores del WIC.

Dentro de los proyectos del centro, AKORI y OpinionZoom son los dos con más investigaciones vigentes dentro del WIC, la principal diferencia de estos dos radica en que el segundo utiliza Tweets extraídos de una API de Twitter para sus investigaciones y mejoras a los algoritmos utilizados en su software, mientras que en el primero se han realizado experimentos con personas para extraer datos y utilizarlos en los distintos trabajos, es decir son datos en los que se requiere más cuidado en el cómo tratarlos y almacenarlos que los de OpinionZoom.

En el presente año no se han realizado experimentos en AKORI. La mayoría de los memoristas que sí lo hicieron, ya no se encuentran en el centro, por esta razón se les realizó una ronda de entrevistas (Anexo A.3), en donde se pudo observar que respecto a la gestión de datos que se realiza en el WIC, no existen pautas o protocolos a seguir al momento de generar datos a través de experimentos de una investigación. En general no se tiene un conocimiento de lo que es un plan de manejo de datos, ni se da entrenamiento ni guías para ello, lo cual se mantiene hasta el día de hoy.

Por otro lado, en OpinionZoom a pesar de la naturaleza de sus datos, también se necesitan almacenar los datos extraídos de Twitter. Consultando a los actuales integrantes de OpinionZoom se obtienen respuestas parecidas al caso de AKORI, es decir, no se realiza una documentación, plan de gestión de datos, utilización de metadatos ni nada por el estilo dentro de este proceso.

Como ya se ha mencionado la gestión de datos es uno de los grandes focos de las buenas prácticas en la investigación e involucra no solo el hecho de respaldar los datos al final de estas, si no también temas referentes a la documentación, la manera de explicarlos usando metadatos y además el trato ético que se les debe realizar, entre otros.

Un punto importante referente al proyecto AKORI es que se desarrolló en conjunto a la Facultad de Medicina de la Universidad de Chile, los cuales a pesar de no tener tan estructurada la gestión de datos sí contaban con algunos puntos en sus reglas, como el tema ético (obligaban a los alumnos que trabajaban en AKORI a hacerles firmar un consentimiento informado (CI) a cada persona que realizaba el experimento). Por otro lado, esto era un protocolo de la Facultad de Medicina y no del centro de investigación en sí, por lo que no se podría contar como algo perteneciente al centro.

En el WIC no existen documentos, reglas o políticas sobre el cómo tratar los datos durante y después de la investigación, lo que ha llevado al problema mencionado de la baja disponibilidad futura de los datos recolectados. Además de esto no existe entrenamiento otorgado a los trabajadores sobre la gestión de datos, lo cual en general es tratado en las políticas de las instituciones en el mundo.

Por otro lado, existen ciertos reglamentos en algunos de los fondos que se utilizan para financiar los proyectos del WIC. Uno de los requerimientos para algunos proyectos es la aprobación por parte del Comité de Ética de la institución a la cual depende, en este caso el de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, el cual busca la protección de los derechos, la seguridad y el bienestar de los seres vivos relacionados con alguna investigación realizada por la Facultad [72]

En conversaciones con uno de los integrantes de este Comité, se llegó a la conclusión de que no existen protocolos especiales creados en la facultad y estos se rigen principalmente por los requisitos éticos que entregan los mismos fondos de cada proyecto que analizan. Uno de los puntos sensibles para el centro son los relacionados al proyecto AKORI en el cual se realizan experimentos utilizando personas.

Es por esto que se deben cumplir ciertas leyes relacionados tanto a los experimentos como con los datos que se obtienen, dos de estas son la Ley 19.628 Sobre Protección de la Vida Privada, donde principalmente se prohíbe difundir los datos personales obtenidos en los experimentos (y también de otras formas, pero para términos de esta memoria no aplican) [73]. Otra de las leyes involucradas es la Ley 20.120 Sobre la Investigación Científica en el ser humano, su genoma, y

prohíbe la clonación humana. En esta ley se muestran ciertas restricciones que se deben tomar en cuenta, como la obligatoriedad de realizar un consentimiento previo, expreso, libre e informado, o, en su defecto, el de aquel que deba suplir su voluntad en conformidad con la ley a las personas que participen de ciertos experimentos [74], los cuales también deben ser firmados por el encargado o director del proyecto. Este último punto es importante ya que bajo los requerimientos de algunos fondos, por ejemplo Fondecyt, el comité de ética de la facultad debe realizar una auditoría una vez realizada la investigación en donde se podrían requerir los consentimientos firmados, que en caso de que no existan pueden llevar a la suspensión del proyecto [75].

Finalmente, respecto a la infraestructura tecnológica para el almacenamiento y seguridad de estos datos, el centro no cuenta con un repositorio especializado para esto. Cuenta con un servidor de gran capacidad (4 Tb) otorgado por la facultad y algunos computadores de menor capacidad utilizados como servidores por parte del centro, los cuales se utilizan principalmente para albergar a las distintas aplicaciones web y software que se han creado en distintas investigaciones y proyectos. Uno de los alumnos mencionó una carpeta de DropBox la cual era para el AKORI, pero esta carpeta estaba alojada en su cuenta personal y tenía el fin de almacenar otro tipo de archivos, como lecturas relacionadas a los temas de AKORI, códigos de algoritmos e imágenes, no los datos obtenidos en los experimentos. Los demás alumnos han utilizado al igual que él, un espacio de almacenamiento personal. Con esto se puede decir que el centro de investigación no cuenta actualmente con una arquitectura tecnológica disponible para el almacenamiento de datos científicos generados en él.

3.2. Tabla de Contenidos

Para desarrollar la tabla de contenidos que tendrán las Políticas de Gestión de Datos Científicos del WIC se utilizará una comparación realizada de las guías y recomendaciones de cuatro entidades internacionales, las cuales ofrecen una lista de contenidos para las políticas y sobre qué trata cada uno de ellos.

3.2.1. Comparación de guías para la formulación de políticas

En esta sección se realizará una comparación por tópicos o puntos importantes a considerar según cuatro organismos internacionales que cuentan con guías sobre el contenido que deben tener las políticas de gestión de datos científicos de las instituciones. Estas entidades son CODATA, Proyecto RECODE, ANDS y DCC (UK).

Se mostrará una tabla que explicará qué recomienda cada una de ellas respecto a cada punto junto a un análisis comparativo de estos. Cabe destacar que algunos de los tópicos que se explicarán no necesariamente son un título o sección de la tabla de contenidos real de las políticas, si no temas a tratar en alguno o todos ellos.

CODATA GUIDELESS	Definir Datos de Investigación (<i>Research Data</i>).
DCC UK	Definir palabras claves para la gestión de datos, como "datos", "documentación", <i>research data</i> , etc.
RECODE PROJECT (OPEN AS DEFAULT)	No menciona.
AUSTRALIAN NATIONAL DATA SERVICE	Definir todo lo que se está usando o se menciona en el documento, tales como Archivos, Repositorio Institucional, etc.

Tabla 3.1: Definiciones

Definiciones

Para este primer tópico, las principales diferencias entre las recomendaciones de cada guía tienen relación con el tipo de palabras que es necesario definir para hacerlas efectivas dentro de las políticas de gestión de datos de la institución. CODATA por ejemplo menciona sólo la definición principal de *research data*, la cual será utilizada en todo el documento, por otro lado ANDS es más general y recomienda definir todas las palabras claves que se utilicen (no necesariamente relacionadas a la gestión de datos (RDM) con el fin de eliminar la confusión y generar la menor incertidumbre posible entre los usuarios de las políticas. Un punto a destacar es que en las recomendaciones de RECODE no se hace mención a este punto. La tabla comparación se puede observar en la tabla 3.1.

Alcances

CODATA GUIDELESS	Definir los alcances (qué datos exclusivamente compartir, son afectados por las políticas, etc.). Diferencia entre datos que sirven para verificar resultados (que deben estar disponibles) y todos los datos generados (disponibles por un periodo).
DCC UK	Especificar a quién cubre las políticas (<i>staff/estudiantes</i>).
RECODE PROJECT (OPEN AS DEFAULT)	No menciona.
AUSTRALIAN NATIONAL DATA SERVICE	Especificar a quién cubre. Menciona especificar las excepciones donde la política no aplicará.

Tabla 3.2: Alcances

Para el caso de CODATA, los alcances van relacionados a qué tipo de datos de datos se aplicarán las políticas (separándolas en dos grupos). Por otro lado, el DCC y ANDS hablan de las entidades que serán afectadas por estas políticas, este último recomendando explicitar las excepciones de estas. Al igual que el caso anterior en las recomendaciones de RECODE no se especifica este punto. La tabla comparación se puede observar en la tabla 3.2.

CODATA GUIDELESS	Especificar quién es el encargado para cada actividad: crear políticas, dar soporte, preservación de datos, velar por que se ejecuten las políticas, etc. (Financiador, Investigador, Organizaciones, Institución, Servicios para RDM).
DCC UK	Definir los roles institucionales en apoyar la gestión junto a las responsabilidades de los investigadores.
RECODE PROJECT (OPEN AS DEFAULT)	Mencionar específicamente los roles y responsabilidades de todos los <i>stakeholders</i> . Quién se hace cargo de dar apoyo a los investigadores, guías, etc.
AUSTRALIAN NATIONAL DATA SERVICE	Definir responsabilidades para <i>staff</i> , estudiantes, supervisores e institución.

Tabla 3.3: Responsabilidades

Responsabilidades

En este caso, los cuatro concuerdan en especificar las responsabilidades y obligaciones de las personas o stakeholders involucrados, ya sean estudiantes (tesistas principalmente para el caso de este trabajo), supervisores, académicos e instituciones entre otros. Un punto importante para el DCC es el quién se hace responsable de brindar apoyo en cada paso de la gestión de los datos a los investigadores, que en general recae en la institución, lo cual es primordial para el cumplimiento y seguimiento de las políticas de gestión de datos. La tabla comparación se puede observar en la tabla 3.3.

Criterios

CODATA GUIDELESS	Especificar cuáles datos son más importantes de compartir (además de software y herramientas disponibles o recomendadas).
DCC UK	Especificar qué datos y documentación es requerida para ser respaldada, además del tiempo que será guardada.
RECODE PROJECT (OPEN AS DEFAULT)	Mencionar qué datos respaldar, por cuánto tiempo y dónde.
AUSTRALIAN NATIONAL DATA SERVICE	No menciona.

Tabla 3.4: Criterios

Este tema se mezcla para CODATA con el punto anterior. Sin contar el servicio australiano, concuerdan en que se debe especificar qué datos son los realmente necesarios de compartir y respaldar, además de los tiempos en que deberán ser guardados y el lugar específico para esto. Algunos recomiendan repositorios nacionales o internacionales especializados en el área de investigación, otros espacios locales. La tabla comparación se puede observar en la tabla 3.4.

Infraestructura

CODATA GUIDELESS	Mencionar qué existe actualmente para respaldar datos, en caso de que no se cuente con algo para esto se debe indicar quién se hará cargo de esto. Costos: quién se hace cargo de los costos de preservación o qué parte del total paga la institución, dar recomendaciones sobre repositorios (si es que existen centros nacionales o internacionales especializados en el área).
DCC UK	Definir quién es el responsable de los costos de RDM. Mencionar si se usaran repositorios externos o internos, servicios o centros especializados.
RECODE PROJECT (OPEN AS DEFAULT)	Especificar quién se hará cargo de los costos para la implementación de la política respecto a RDM, mencionar la infraestructura existente y necesaria para que exista una colaboración de datos abiertos.
AUSTRALIAN NATIONAL DATA SERVICE	Detallar cómo y cuándo serán guardados los datos. Si son digitales o físicos. Quien proveerá la seguridad y cuales datos serán seguros.

Tabla 3.5: Infraestructura

En general todos hablan sobre especificar lo que existe en la institución para respaldar datos científicos. En algunos casos se recomienda utilizar colecciones o repositorios nacionales o internacionales específicos del área en caso de que se quieran abrir los datos. ANDS menciona también nombrar a los encargados de la seguridad de los datos seleccionados, pero esto se puede especificar en los puntos anteriores de criterio y responsabilidades. Por otro lado, se habla de los costos relacionados a la gestión de los datos, quién se hará cargo de cada costo y por cuánto tiempo. La tabla comparación se puede observar en la tabla 3.5.

Data Management Plan

Se pide como requerimiento al investigador la creación de un plan de manejo de datos, este puede ser pedido por la institución o algún financiador externo. CODATA menciona escribir los requerimientos que deben incluirse explícitamente en este plan de datos. Por otro lado, también se recomienda dar opciones de software y herramientas disponibles para hacer más fácil el proceso, en la sección de responsabilidades debe ir un punto en que la institución se comprometa a apoyar al investigador en este punto. La tabla comparación se puede observar en la tabla 3.6.

Re-uso / Descubrimiento

Sobre el descubrimiento se habla del cómo, dónde y para quién estarán disponibles los datos. Sobre el re-uso el cómo se llevará a cabo (citaciones, licencias, tiempo de privilegios, etc.). También

CODATA GUIDELESS	Detallar los requerimientos del DMP: Extensión, cuándo se debe hacer (y cuántos), mínimos, qué herramienta usar, etc. Especificar cuánto tiempo debe estar preservada la data. Formatos, metadatos etc. También para algunas áreas existen aproximaciones técnicas sobre formatos o estándares de metadatos, mencionarlas de manera directa o indirecta (no necesariamente adoptarlas).
DCC UK	Especificar como requerimiento la realización de un DPM ya sea por requerimiento institucional o del financiador.
RECODE PROJECT (OPEN AS DEFAULT)	Requerir que los investigadores realicen un DMP.
AUSTRALIAN NATIONAL DATA SERVICE	No menciona.

Tabla 3.6: Data Management Plan

CODATA GUIDELESS	Recomendaciones o requerimientos: Definir cómo acceder a los datos (ej: links en las publicaciones), recomendaciones sobre metadatos, identificadores y licencias para permitir su re-uso.
DCC UK	Está en el tópico “Ética”, mencionar en lo que potencialmente afectará el re-uso de los datos. Quién tiene acceso y cómo, si es abierto y qué estará disponible.
RECODE PROJECT (OPEN AS DEFAULT)	Mencionar licencias.
AUSTRALIAN NATIONAL DATA SERVICE	Especificar quién tiene acceso y de qué manera, dependiendo de la sensibilidad de la data. Los datos deberán estar disponibles para otros investigadores excepto existan temas éticos, de privacidad o confidencialidad.

Tabla 3.7: Re-uso / Descubrimiento

existen restricciones por temas éticos que también son mencionados en los siguientes tópicos. La tabla comparación se puede observar en la tabla 3.7.

Reconocimiento

CODATA GUIDELESS	Abrir los datos requiere ciertos cambios en la cultura de la investigación: Aparte de requerir abrir los datos se pueden agregar tópicos sobre la necesidad para el reconocimiento y ganancias para los investigadores que abren los datos. Se requiere el reconocimiento de los datos y la citación de esta. Se menciona los accesos privilegiados por cierto periodo a investigadores.
DCC UK	No menciona.
RECODE PROJECT (OPEN AS DEFAULT)	Mencionar los reconocimientos para los investigadores de manera formal.
AUSTRALIAN NATIONAL DATA SERVICE	No menciona.

Tabla 3.8: Reconocimiento

Las recomendaciones dadas por CODATA y RECODE que promueven abrir los datos, mencionan el otorgar reconocimiento (cuál será este reconocimiento) a los investigadores que abran sus datos (si es que la institución lo permite). Igual que el punto anterior se menciona la citación tanto para las publicaciones como para los datos utilizados además de los tiempos de privilegios para algunos casos. La tabla comparación se puede observar en la tabla 3.8.

Reportes, monitoreo y sanciones

CODATA GUIDELESS	Mencionar cómo serán monitoreadas las políticas, qué reportes y sanciones serán realizados, como reportes anuales de planes de manejo de datos o reportes finales luego de cierto periodo.
DCC UK	Recomienda especificar que exista una revisión periódica.
RECODE PROJECT (OPEN AS DEFAULT)	Explica que donde haya creación de datos, debe existir recursos, revisiones y monitoreo de las practicas.
AUSTRALIAN NATIONAL DATA SERVICE	Hablar del cuándo serán revisadas las políticas (para analizar si han sido efectivas).

Tabla 3.9: Reportes, monitoreo y sanciones

Los cuatro mencionan especificar una revisión del cumplimiento de estas políticas, en algunos casos y si es que existirán sanciones, cuáles serán. Definir si serán periódicas, al principio o al final, si serán revisiones a los planes de datos propuestos por los investigadores o en general, etc. En general se considera un punto importante para la mejora constante en el largo plazo. La tabla comparación se puede observar en la tabla 3.9.

Ética

CODATA GUIDELESS	Para algunas áreas, mostrar los requerimientos en temas legales y éticos sobre datos personales por ejemplo.
DCC UK	Para el reuso, propiedad y temas de datos personales.
RECODE PROJECT (OPEN AS DEFAULT)	No se menciona explícitamente.
AUSTRALIAN NATIONAL DATA SERVICE	Menciona para ciertos casos de datos sensibles especificar quiénes y cómo tendrán acceso.

Tabla 3.10: Ética

En este punto se habla de los temas para datos sensibles (como datos personales), quiénes tendrán acceso, cómo será el re-uso de estos, etc. También se deben agregar otros temas legales y de confidencialidad.

Este tópico es crítico para centro, ya que en el WIC se realizan experimentos en donde se toman muestras de datos sensibles. La tabla comparación se puede observar en la tabla 3.10.

Otros tópicos

	Propiedad	Requerimientos externos	Eliminación, destrucción y movimiento de registros.
CODATA GUIDELESS	No menciona	No menciona	No menciona.
DCC UK	Menciona quien es el dueño de los datos científicos (Institución).	Mencionar que los requerimientos de financiamiento externo que son priorizados a los de la institución.	No menciona.
RECODE PROJECT (OPEN AS DEFAULT)	No menciona	No menciona	No menciona
AUSTRALIAN NATIONAL DATA SERVICE	Mencionar quién será el dueño de los datos	No menciona	Quien será el responsable, cómo se hará y cuándo.

Tabla 3.11: Otros tópicos

Existen algunos puntos, los cuales no son mencionados como tópicos principales dentro de las guías, pero si los recomiendan considerar a la hora de formular nuevas políticas para una institución, el primero es el tema de la propiedad de los datos u *ownership*, es decir quién será el dueño de los datos y las investigaciones, es importante para el centro debido a que se busca crear aplicaciones comerciales a partir de estos experimentos. Cabe destacar que estos puntos son mencionados en

las guías que no están involucradas en el acceso abierto. CODATA y RECODE tienen un enfoque importante que es el “Open as Default” (Abiertos por defecto) lo que hace orientar todos los tópicos mencionados para que esto se cumpla. El segundo punto es el tema de los requerimientos externos, el cuándo estos serán priorizados a las políticas del centro. En el caso del WIC, al ser financiado por fondos públicos se debe analizar cuáles de estos requerimientos se debe estar seguro de agregar a las políticas que se busca diseñar con el fin de que cumplan, además de los del Comité de Ética para la Investigación de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, de la que es parte y tiene ciertos requerimientos para aprobar las investigaciones realizadas referentes al tema ético de esta. Finalmente, el último punto es sobre la destrucción o movimiento de los datos almacenados en los repositorios locales, el cual debe ser estructurado y bajo la responsabilidad de un encargado perteneciente al centro. La tabla comparación se puede observar en la tabla 3.11.

3.2.2. Contenidos

El segundo paso de la guía de 5 pasos para el diseño de políticas de gestión de datos científicos otorgada por el Data Curation Centre consiste en realizar un benchmark de distintas políticas de distintas instituciones y estructurar el contenido de las políticas [54].

También recomienda tener en cuenta las siguientes diez secciones: Información de Cabecera (Título del documento, logo institucional, etc.), Título de la política, Objetivos, Principios, Alcances, Texto de las políticas, Roles y Responsabilidades, Aprobación y Revisión Periódica, Información de pie de página (versión, estado, etc.) y finalmente Anexos.

Revisando las políticas mostradas en [53], donde se encuentran los documentos para cerca de treinta universidades del Reino Unido UK (entre las cuales se encuentran la Universidad de Oxford, Universidad de Edinburgo, Universidad de Nottingham, Universidad de Manchester y Birmingham, entre otros) y las publicadas en [76], donde están estos archivos de aproximadamente cuarenta universidades australianas (entre las que se encuentran la Universidad de Melbourne, Universidad de Sidney, Universidad Nacional Australiana, Universidad La Trobe y Universidad de Griffith, entre otras) se puede observar que existen ciertas similitudes entre las políticas pertenecientes a cada país en cuanto a estructura, pero en general las estas diferencias entre ellas son principalmente que algunos tópicos son considerados como secciones dentro de las políticas, a diferencia de la guía del DCC donde son subsecciones de la sección “Texto de las políticas”, como por ejemplo propiedad de los datos, plan de manejo de datos, entre otros. Un ejemplo de esto es la Universidad de Birmingham del Reino Unido, que tiene una sección de Principios y Alcances en vez de dos separadas como propone la guía del DCC [77].

Considerando esto último, que las diferencias de estructura en general son bastante pequeñas y que lo más importante es el contenido final de estas, se tomó como base la recomendación otorgada por la guía de 5 pasos del DCC exceptuando los "Principios" de las políticas que se abarcan en los objetivos de esta. A continuación, se explica qué contiene cada sección.

Encabezado y Pie de página

Tal como lo recomienda la guía, en estas dos partes del documento se encuentra el título del documento, el logo institucional y del centro, la versión de las políticas y el número de la página actual, entre otros.

Principios y Objetivos

En esta sección se presentarán los principios en que se basan las políticas y los objetivos principales de las políticas referentes a la gestión de los datos científicos generados en las distintas investigaciones del Web Intelligence Centre.

Alcances

En este punto se mencionarán las entidades que serán afectadas directa o indirectamente por estas políticas, utilizando las recomendaciones comparadas anteriormente. Respecto a lo que CODATA consideraba con alcances, se decidió tomar ese punto en donde se hable de los criterios para realizar la separación entre los datos que son necesarios guardar y los que no.

Texto de las Políticas

En esta sección se tratarán todos los tópicos, exceptuando Reconocimientos, comparados en el punto anterior, cada uno se tomará como una subsección de esta y los párrafos estarán enumerados.

Los tópicos son los siguientes:

- Criterios: En esta subsección se especificará el criterio para elegir qué datos y documentación serán los almacenados, además de en qué lugar y por cuánto tiempo se hará.

- Infraestructura: En esta subsección se especificará la infraestructura con la que contará el centro al momento de ser efectivas estas políticas. También se definirá si se deben respaldar los datos en algún repositorio externo que puede ser institucional a nivel facultad o universidad, un repositorio nacional o internacional especializado en el área científica de cada investigación. Para el borrador este punto se dejará en blanco debido a que no se tiene certeza sobre la solución para el almacenamiento de datos.

- *Data Management Plan*: Se especificará la obligatoriedad de realizar un plan de manejo de datos, además de en qué momento debe hacerse. Se tomará en cuenta la recomendación dada por CODATA, por lo que también se determinarán los requisitos mínimos que deban contener estos documentos y en caso de existir recomendar ciertas herramientas para producirlos.

- Propiedad: Al ser un centro de investigación que busca diseñar nuevas propuestas comerciales, es necesario especificar quién será el dueño de los datos generados en el WIC en sus distintos

proyectos.

- Re-uso y Descubrimiento: Se especificará quiénes tendrán acceso a los datos guardados en el repositorio del centro, si los datos serán abiertos o tendrán restricciones de acceso y las restricciones para abrirlos.

- Eliminación, destrucción y movimiento de registros: Se detallará quiénes serán los encargados para su realización.

- Ética: En esta sección irán los detalles sobre los requerimientos legales sobre los datos que puedan ser sensibles y que deban tratarse de manera especial.

- Requerimientos Externos: La totalidad de los proyectos del WIC han sido financiados por fondos públicos tanto de CORFO como de CONICYT, es por esto que es necesario especificar que en caso de que existan requerimientos relacionados a la gestión de los datos producidos en investigaciones financiadas por alguno de estos fondos se les deban dar prioridad ante las políticas que se están diseñando.

Roles y Responsabilidades

Respecto a los roles y responsabilidades se tomará en cuenta que las cuatro guías que se analizaron anteriormente recomiendan fuertemente especificar qué debe hacer cada stakeholder y qué rol cumplirán respecto a la gestión de datos científicos.

Definiciones

En esta sección se definirán todas las palabras que puedan causar confusión respecto a su significado y que serán las utilizadas explícitamente en las políticas de gestión. Al ser primera vez que se hace una aproximación en el centro de investigación donde se planea introducir estas políticas no se puede asumir que todos los integrantes del centro tienen los conocimientos de las palabras utilizadas dentro de estas políticas, por esto se utilizará la recomendación dada por el Servicio de Datos Australiano (ANDS) donde se busca definir todas las palabras necesarias para cumplir las políticas que generen alguna clase de incertidumbre respecto a su significado.

Aprobación y revisión periódica

Tal como se recomienda en las guías analizadas, debe existir una revisión periódica del contenido de las políticas de gestión de datos científicos, principalmente con el fin de que no queden obsoletas en el tiempo. En esta sección se declararán los tiempos mínimos en que se deberán revisar y analizar si es necesario algún cambio, además de quién deberá aprobar estos cambios.

Anexos

En caso de ser necesario se adjuntarán documentos u otro tipo de anexos que sean utilizados por algún punto dentro de las políticas diseñadas.

3.3. Ruta de Aprobación

El tercer paso consiste en revisar la ruta para que se aprueben estas políticas. Al ser un centro de investigación pequeño la estructura organizacional de este es bastante pequeña (Anexo A.2) y en general todas las decisiones pasan finalmente por el encargado del centro, el profesor Juan Velásquez, phd.

Las políticas son un documento que deben ser firmados por cada integrante que realice algún tipo de investigación en el centro. Al ser proyectos financiados por fondos públicos y en donde el dinero se le entrega a la universidad y no al centro en sí, los miembros que participen en ellos deben firmar un convenio con la universidad para recibir los pagos correspondientes a su trabajo, ya sean alumnos o profesionales.

Debido a esto existen dos alternativas, la primera es adjuntar las políticas al convenio que se firma al momento de entrar al centro y la segunda, es realizar un documento aparte que sea firmado por los integrantes del WIC y sea entregado por el mismo centro, en este caso por parte del encargado de este.

Los convenios que se entregan primero pasan por los encargados de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile para finalmente ser firmados por el decano de esta, luego se dirigen a la Casa Central de la misma universidad para ser aprobados y hacerse efectivo los pagos relacionados a ellos. Esto quiere decir que finalmente los investigadores del centro pasan legalmente a realizar un trabajo para la universidad y no para el centro como institución. La Universidad de Chile al ser una institución pública cuenta con una normativa bastante grande y complicada, por lo que es muy difícil que se consiga la aprobación para adjuntar las políticas de gestión de datos a los convenios firmados con esta, para que en caso de no cumplimiento se puedan tomar las medidas necesarias respecto al punto legal.

Debido a esto, la segunda opción es la más viable para el caso de esta memoria. La desventaja de esta alternativa es que al no tener repercusiones legales se podrían dejar de lado en algunos casos, por otro lado, al ser un centro pequeño en donde es más fácil controlar el cumplimiento de las políticas podría no ser un problema a largo plazo. Sumado a esto, el fin de las políticas es evitar problemas relacionados a los datos al finalizar las investigaciones, algo que se busca como centro de investigación por parte de todos sus integrantes, por lo que se espera un cumplimiento por parte de ellos.

Finalmente, la segunda opción será la utilizada dentro del centro de investigación para hacer efectivo el compromiso de los integrantes a cumplir las políticas que se contarán en el futuro. Debido a esto y como será algo interno que no involucrará otras entidades como la misma facultad o la universidad, la ruta de aprobación queda directamente ligada al encargado del centro, el cual

será el que finalmente las aprobará en su versión inicial y futuras, antes de esto deberán aprobarlas también los encargados de cada proyecto que se esté realizando en el momento que se quieran aprobar las políticas.

3.4. Borrador de Políticas de Gestión de Datos

En la presente sección se desarrolla el cuarto paso donde se diseña el borrador de las políticas de gestión para el centro. Se toman en cuenta los puntos mencionados en las secciones anteriores sobre su estructura y contenido. Se presentará el contenido que se integra al borrador junto a comentarios relacionados a estos. Por otro lado, en algunos puntos se mencionan requisitos que en caso de no haber realizado o estudiado la gestión de datos científicos sería muy difícil de cumplir, por lo que se necesitarían guías y material de apoyo si se quiere que los integrantes del WIC puedan realizarlas, en estos casos se mencionan las guías que son diseñadas en los próximos capítulos de este informe.

3.4.1. Encabezado y Pie de página

En el encabezado del documento de las políticas se muestra el título del documento, el cuál es: "Políticas de Gestión de Datos Científicos, WIC, FCFM, Universidad de Chile."

Los logos utilizados son tanto el de la facultad como el del centro, ambos adjuntos en el anexo C.1.

En el pie de página va la versión del documento, que en este caso es la número 1.0, que será la entregada al término de este capítulo. Esta versión irá cambiando a medida que se vayan modificando las políticas en el futuro. También se incluye el enumerado de páginas del documento.

3.4.2. Principios y Objetivos

Las políticas se basan bajo el principio de que los datos primarios son esenciales para los resultados u *outputs* de las investigaciones realizadas en el WIC. Debido a esto la gestión de los datos científicos se debe hacer primero respetando todos los acuerdos relacionados a la privacidad, propiedad u otros requisitos impuestos a los datos, asegurando siempre su integridad, teniendo que ser promovido por todos los integrantes del centro.

Los objetivos de estas políticas son por una parte fomentar las buenas prácticas relacionadas a la investigación en los integrantes del centro, asegurar la preservación de los datos que se generan dentro de él (que es el principal problema a resolver), mitigar los riesgos relacionados a un mal uso de los datos y asegurar que todos puedan realizar una buena gestión de los datos de investigación cumpliendo estas políticas.

El texto que contiene el documento es el siguiente:

“Estas políticas se basan bajo los principios de que los datos primarios son esenciales para los resultados de la investigación realizada en el centro, es por esto que la gestión de estos datos debe realizarse asegurando su integridad para facilitar y dar paso a nuevas investigaciones respetando todos los acuerdos de privacidad, propiedad u otros requerimientos. Esto debe ser promovido en todos los integrantes del centro de investigación.

Este documento establecerá todos los requerimientos relacionados a la gestión de datos de investigación que sean generados en alguna investigación realizada en el Web Intelligence Centre (WIC) de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Los objetivos de esta política son los siguientes:

- Fomentar y establecer una cultura de buenas prácticas de investigación referentes a la gestión de datos científicos generados en el Web Intelligence Centre.
- Asegurar la preservación de los datos generados en el WIC, almacenándola y haciéndola accesible para su uso y re-uso con el fin de continuar las investigaciones realizadas.
- Disminuir los riesgos relacionados a una mala gestión de datos en el WIC.
- Asegurar que los integrantes del centro sean capaces de cumplir con todas las políticas o reglamentos en donde la buena gestión de datos contribuya a la investigación."

Con estos objetivos plasmados en el documento de las políticas se espera cumplir el objetivo de la gestión de datos de investigación en el centro.

3.4.3. Alcances

Como se mencionó en los alcances de este trabajo de título, los alcances de estas políticas están principalmente enmarcados dentro de los integrantes del centro, pero también considera a personas externas en el caso de que estén involucrados en alguna de las investigaciones de algún proyecto del WIC.

Para el caso de las instituciones, en este punto se menciona a todos los miembros de la universidad que realicen alguna investigación en la institución y en algunos casos a los alumnos que puedan estar realizando trabajos en alguna de ellas, al ser en general todos los miembros del centro parte de algún proyecto se considera necesario que aplique a cada uno de ellos el documento de las políticas.

El texto que contiene el documento es el siguiente:

“Las Políticas de Gestión de Datos Científicos se aplicarán a todos los integrantes del Web Intelligence Centre o externos que realicen o estén involucrados en alguna investigación o sean parte de algún proyecto del centro."

3.4.4. Definiciones

En esta sección se utilizan las definiciones dadas en el capítulo 2 de este informe.

Las palabras que se describen son: Datos de Investigación, Plan de Manejo de Datos, Datos Sensibles, Digital Object Identifier

Las descripción y definiciones que están en esta sección del documento serán:

“Los siguientes términos y sus definiciones serán las utilizadas por estas políticas en todas sus secciones.

1. Datos de Investigación

O datos científicos, es la información recolectada (sin importar la forma o medios en que pueda existir) necesaria para validar las observaciones, conclusiones o resultados de un proyecto de investigación.

2. Plan de Manejo de Datos

Documento en donde se muestra cómo serán manejados los datos científicos durante y después de cada proyecto. Debe describir el ciclo de vida para todos los conjuntos de datos que serán recolectados, procesados o generados por el proyecto de investigación.

3. Datos Sensibles

Son todos los datos o conjuntos de datos que puedan tener problemas legales, confidencialidad o éticos. Principalmente cuando son datos generados en experimentos realizados con humanos o animales.

4. Digital Object Identifier

Es una forma de identificar un objeto digital el cual, aunque cambie de URL seguirá teniendo la misma identificación.

3.4.5. Texto de las Políticas

En esta sección se detallan los puntos que contienen las políticas referentes a los distintos contenidos mencionados en este capítulo, exceptuando la infraestructura, que es desarrollada en el siguiente capítulo de este informe.

Propiedad

Al igual que en la mayoría de las políticas revisadas, el *ownership* o propiedad de los datos que se generan en las instituciones recae en ellos mismos, exceptuando los casos en que existan requisitos de los financiadores, que en caso de Chile no existen a pesar de ser fondos públicos. La única diferencia para este caso es que el WIC está dentro de la Universidad de Chile, por lo se debe agregar la excepción si es que existe un requisito por parte de esta para el centro.

Un punto importante es que los datos generados en experimentos con personas son exclusivamente de ellos mismos, es decir, no del centro mismo, pero estos permiten utilizarlos por parte del WIC para sus investigaciones.

El texto que contiene el documento en esta subsección es el siguiente:

- a. Los derechos de todos los datos de investigación generados dentro del marco de algún proyecto o investigación realizada en el centro serán exclusivamente del WIC, a excepción de que existan requerimiento de los financiadores o de la universidad.
- b. En caso de ser datos generados en experimentos con seres humanos, los datos serán de propiedad de ellos, pero los derechos a usar serán exclusivos del centro de investigación. Estos derechos de uso deben ser descritos en los consentimientos informados que firmen al momento de realizar estos experimentos.

Criterios

Como el objetivo del trabajo es prevenir la falta de disponibilidad de los *sets* de datos creados en los experimentos realizados en el centro, estos deberán guardarse sin excepción en el repositorio del centro, además se deben almacenar todos los datos que sean considerados de interés para futuras investigaciones.

Un punto importante mencionado al comienzo de este capítulo es el tema legal relacionado a algunos datos generados, especialmente cuando se trata de experimentos realizados con/en humanos. De acuerdo a la ley 20.120 de Chile se debe guardar un consentimiento informado (CI) de cada persona que participó en los experimentos, además por parte de la facultad y el comité de ética estos pueden ser requeridos al centro para una revisión, es por esto que se deberán guardar en versión digital todos estos CI, con el fin de evitar problemas futuros.

El texto que contiene el documento en esta subsección es el siguiente:

- a. Los datos científicos que sean creados a partir de cualquier clase de experimento realizado bajo alguna investigación ejecutada en el centro, deberá ser depositados en el repositorio entregado por el centro sin excepción y siguiendo las guías entregadas por el centro.
- b. En caso de que existan temas legales relacionados a datos generados en experimentos, se deberá guardar de manera digital toda la documentación necesaria para respaldar la toma de los datos, por ejemplo, consentimientos informados para el caso de experimentos realizados con seres humanos, que queda descrito bajo la ley 20.120 de Chile, u otros documentos requeridos por el Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas.
- c. Se debe respaldar todos los datos que sean considerados de interés para futuras investigaciones, siguiendo los mismos protocolos que los datos mencionados en el punto a.

Plan de Manejo de Datos

Debido a la importancia de contar con planes de gestión de datos o *Data Management Plan* (DMP), antes de realizar cualquier tipo de experimento referente a alguna investigación de alguno de los proyectos, los integrantes deberán entregar dos versiones del documento al jefe de proyecto,

el cual tendrá que aprobarlos para poder continuar. La primera versión deberá ser entregada antes de realizar el trabajo y uno al finalizarlo.

En un comienzo se había pensado en requerir sólo una versión al comienzo de la investigación como se hace normalmente en otras instituciones, pero para el caso del centro de investigación, en que la mayor parte de los integrantes que realizan experimentos nuevos son memoristas, esto puede ser un poco complicado. A través de experiencias anteriores al momento de hacer estos experimentos y tomar los datos no se tiene claro que se va a hacer más adelante, qué algoritmos se utilizarán, metodologías, entre otros. Esto hace difícil poder completar un plan de manejo de datos como se espera obtener al simplemente no contar con esa información. Por otro lado, uno de los usos que se les da a los planes de gestión de datos es justamente asegurar que el investigador tiene un conjunto de pasos definidos a desarrollar, lo que puede ayudar a los memoristas a realizar sus experimentos.

Para continuar se deben analizar tres puntos importantes que involucran al plan de manejo de datos y a otras secciones posteriores, estos son: licencias, formatos de archivos y metadatos.

Licencias: En Chile, la ley N° 17.336 de Propiedad Intelectual (Derechos de Autor) [78] incluye a los datos y las bases de datos, pero sólo cuando cumplen el estándar de originalidad, es decir que por razones de la selección o disposición de sus contenidos puedan considerarse como originales. Cabe destacar que abrir los datos no implica la renuncia a los derechos de propiedad o titularidad de los datos, solo permite a terceros su acceso y manipulación a alguno de ellos [3].

Las licencias permiten especificar los distintos términos de uso que se les quiera imponer a los terceros que deseen reusar los datos, y a pesar de que el acceso abierto a los datos implica una licencia lo menos restrictiva posible, existen distintos tipos de licencia que se pueden utilizar.

El mecanismo más recomendado en las guías y políticas revisadas es Creative Commons (CC), que desde su versión 4.0 permite aplicar una licencia a conjuntos de datos [61], a diferencia de su versión anterior esta es internacional y no de cada país. Por otro lado, siguiendo el enfoque de datos abiertos se encuentra Open Data Commons. La principal diferencia entre estas dos es que la segunda no permite condicionar a que los nuevos usos no sean de uso comercial o más restrictivamente a que no se pueda manipular de ninguna manera, sumado a esto CC es utilizada y reconocida a nivel mundial, además de ser recomendada por las instituciones revisadas, por lo que será utilizado como obligación para el caso de estas políticas. En Chile, CC Chile está liderada por la Universidad de Chile a través del Sistema de Servicios de Información y Bibliotecas, SISIB y la organización no gubernamental ONG Derechos Digitales [79].

Como cada proyecto y experimento, junto a los datos generados en ellos, son distintos entre sí, no se considera apropiado restringir a solo uno de los 7 tipos de licencia que entrega Creative Commons. Por esto, para cada uno de los conjuntos de datos generado se deberá, por parte de la persona que realice el experimento o la situación que genere los datos, escoger entre las posibles licencias de CC en su versión 4.0 o el copyright en caso de ser estrictamente necesario o requerido por el centro. A pesar de que esto permite restringir al máximo los datos, es decir, todos los derechos reservados, se recomienda no utilizar este tipo de restricción y aplicar una licencia lo más permisiva posible, con el fin de fomentar el re-uso de los datos generados en el centro. Cabe destacar que aplicar una licencia abierta a archivos de datos que contienen información sensible no significa que

es posible su publicación.

Para que sea más fácil su utilización se desarrolla una guía de cómo aplicar licencias a los datos generados en los próximos capítulos.

Formatos de Archivos: Como se busca la preservación de largo plazo se necesita asegurar que los formatos en que se guarden los archivos en el repositorio puedan seguir utilizándose. Existen muchos formatos que se puede creer tendrán una perduración razonable como las producidas por Microsoft Office, pero a pesar de esto, los formatos son de propiedad de alguna entidad, por lo que no necesariamente existirán para siempre. Es por esto que las políticas se enfocan en utilizar formatos abiertos.

Para este punto existen distintas instituciones que proponen una lista de formatos aceptables para la preservación a largo plazo, tanto el DDC del Reino Unido como ANDS de Australia recomiendan utilizar los otorgados por UK Data Archive, el cual es uno de los repositorios más grandes de su país [80], los cuales están optimizados para el fin que se busca.

Como la investigación realizada en el centro varía con los años, se pedirá el uso de estos formatos mencionados. UK Data archive entrega dos listados de formatos por cada tipo de dato, unos más aceptables que otros. Se pedirá utilizar alguno de los formatos en estos dos listados, pero se recomienda el uso de los "aceptables para compartir, reutilizar y preservar datos." antes que los "aceptables para la preservación de datos". La tabla de formatos se puede observar en el Anexo C.2 En caso de que se quiera utilizar otro, este deberá cumplir con que sea un formato abierto para asegurar su uso y prevalencia en el tiempo, además de que deberá ser aprobado por los jefes de proyectos.

Metadatos: Para el caso del esquema de metadatos para datos científicos, existen distintos estándares creados con el fin de asegurar su re-uso y fácil entendimiento. Uno de los esquemas más utilizados es el Dublin Core, el problema de este modelo es que no fue hecho para datos de investigación, que es lo que se busca describir con los metadatos en este trabajo. Cuatro de los servicios actuales especializados en datos de investigación más utilizados [81] son DataCite, Dataverse, Dryad y Figshare, servicios creados tanto en Estados Unidos como en Reino Unido.

Mientras figshare en su plataforma utiliza un esquema de metadatos que contiene muy pocos elementos (cerca de 12) para describir de buena manera datos científicos, la plataforma Dataverse, utiliza demasiados (un esquema de más de 80 propiedades). Este último está especializado para investigaciones del área de las ciencias sociales.

Por otro lado, DataCite actualmente cuenta con uno de los esquemas más recomendados por distintas instituciones, como el DCC. Cuenta con un mayor soporte y a diferencia de los otros, se ha ido actualizando a lo largo de los años llegando a la versión 4.0 al año 2016 en donde se han ido agregando nuevos metadatos al esquema con el fin de hacer más fácil el descubrimiento, identificación y citación de los datos a través de la utilización de DOIs, el cual es uno de los mecanismos más utilizados en la actualidad para este fin y en donde DataCite ha otorgado más de cuatro millones de DOIs a distintos *sets* de datos de investigación [82]. Pensando en este objetivo principal de DataCite y el soporte con el que cuenta actualmente, además de que tanto el DCC como ANDS, entidades que fueron utilizados como guías para este trabajo, son miembros de DataCite,

es el seleccionado a utilizar para la arquitectura de gestión de datos del WIC.

Existen distintas guías para la creación de metadatos, especialmente para Dublin Core, por ejemplo una creada por la Biblioteca Nacional de Chile [83] y la realizada por el repositorio institucional del Consejo Superior de Investigaciones Científicas (CSIC) [84]. A pesar de esto se necesita de una guía para el caso específico del centro, la cual es desarrollada en capítulos posteriores.

Contenido del DMP: Como se menciona anteriormente se requieren dos versiones, luego de una revisión de distintos ejemplos de planes de manejo de datos, alguno de los cuales se encuentran en [85]. Las dos versiones deberán contener lo siguiente:

La primera versión debe contener los siguientes puntos: nombre del trabajo que se está realizando (por ejemplo si es una memoria, el nombre del trabajo de título), nombre del proyecto del centro a cuál pertenece (por ejemplo AKORI u OpinionZoom), fecha de creación del documento, nombre del encargado del proyecto en el WIC, código y fuente de financiamiento del trabajo (por ejemplo FONDEF CA12i10061), descripción del proyecto y de la investigación o experimentos a realizar, datos del investigador (nombre y contacto), políticas que deban cumplir en el centro, descripción de los datos que se buscará generar en el proyecto (incluyendo formatos, tamaños esperados de archivos y cómo serán adquiridos y procesados en caso de contar con esta última información, junto a las fechas estimadas de los experimentos que se realicen o extraigan datos).

La segunda versión debe contener los puntos mencionados en la primera versión, diferenciándose en la descripción de los datos generados, en donde en la primera versión eran los esperados (formatos, tamaños, fechas de experimentos) y en la segunda deberá ser los resultados y fechas reales de los experimentos en caso de realizarse. Además, deberá incluir las metodologías y algoritmos aplicados a los datos generados, las notas de laboratorio si es que se tienen, dónde y en qué formatos serán almacenados junto a los metadatos que contendrán. En caso de que se generen datos sensibles, qué documentos se guardarán referentes a temas legales o requerimientos de la universidad o financiadores y definir quién será el dueño de los datos mencionados en el documento.

Las dos versiones deberán ser almacenadas de manera digital en el repositorio del proyecto otorgado por el centro.

Con el fin facilitar el diseño de este documento, se desarrolla en los próximos capítulos un ejemplo utilizando un trabajo de título desarrollado anteriormente en el centro junto al *template* o plantilla del DMP para su posterior futuro. Existen dos herramientas *online* que permiten una fácil creación de estos documentos, estos son DMPOnline y DPMTool, pero tienen dos problemas, el primero es que no tienen soporte en español, en el centro el inglés no es algo obligatorio para los integrantes, el segundo problema es que son restrictivos respecto a las plantillas a utilizar, es decir, estas no pueden ser editadas para agregar o cambiar algunos puntos de los *templates* impuestos, donde por ejemplo se podrían haber cambiado los títulos de las secciones del plan de los *templates* a español para ser utilizado de mejor manera para el centro, por otro lado tampoco permiten agregar nuevas plantillas y sólo se pueden utilizar las registradas por ellos de algunos financiadores e instituciones científicas.

El texto que contiene el borrador referente al plan de manejo de datos es el siguiente:

- a. Cualquier integrante del centro de investigación que desee realizar una investigación en él, deberá entregar dos versiones de un plan de manejo de datos al jefe de proyecto, uno antes de iniciar cualquier tipo de experimento referente a estas investigaciones y otro una vez finalizada esta investigación.
- b. La primera versión deberá contener los siguientes puntos: nombre del trabajo que se está realizando, nombre del proyecto del centro a cuál pertenece, fecha de creación del documento, nombre del encargado del proyecto en el WIC, código y fuente de financiamiento del trabajo, descripción del proyecto y de la investigación o experimentos a realizar, datos del investigador (nombre y contacto), políticas que deban cumplirse en el centro, descripción de los datos que se espera generar en el proyecto (incluyendo formatos, tamaños esperados de archivos y cómo serán adquiridos y procesados en caso de contar con esta última información, junto a las fechas estimadas de los experimentos que se realicen o extraigan datos).
- c. La segunda versión del plan de manejo de datos deberá contener los puntos mencionados en a) cambiando la información de los datos esperados a generar y fechas estimativas de experimentos en caso de realizarse, por los resultados y fechas reales de estos, además de las metodologías y algoritmos aplicados a los datos generados, las notas de laboratorio si es que se tienen, dónde y en qué formatos serán almacenados junto a los metadatos que contendrán. En caso de que se generen datos sensibles qué documentos se guardarán referentes a temas legales o requerimientos de la universidad o financiadores y definir quién será el dueño de los datos mencionados en el documento.
- d. Para su diseño se usarán los estándares acordados en las guías y plantillas otorgados por el centro.
- e. Las dos versiones una vez aprobadas deberán ser almacenadas en el repositorio otorgado por el centro, de manera digital.
- f. Debe ser aprobado por el Jefe del Proyecto en donde se esté realizando la investigación.

Para ver el esquema del proceso que deben seguir los integrantes del centro, en donde se incluye el diseño de las dos versiones, se presenta el proceso en el capítulo 5.3.1.

Un punto importante respecto a los planes de gestión de datos es que busquen mostrar que los datos entregados cumplen los principios FAIR, a pesar de que no se utiliza el *template* entregado por Horizon2020 con este fin, sí se debe recalcar que con los requisitos impuestos sobre su contenido sí se pueden cumplir estos principios. Primero para que los datos sean "hallables" (*Findability*) se pide que se entreguen los metadatos del esquema de DataCite necesarios para su descubrimiento utilizando por identificador único (DOI), además de la versión, keywords entre otras propiedades que a pesar de no ser obligatorias se recomienda agregar (Los metadatos pedidos se encuentran en mayor detalle en el Capítulo 5 del presente informe). Respecto a que los datos sean "accesibles" (*Accessibility*), se especifica dónde serán alojados los datos y si serán abiertos o cerrados, a pesar de que en el plan no se pide el cómo será el acceso a estos datos junto a sus restricciones, esto sí se muestra en las políticas de gestión de datos del centro (que en el plan se debe agregar las regulaciones que se deban cumplir para generar y guardar los datos). En cuanto a la "interoperabilidad" (*Interoperability*), se debe agregar qué formatos, software y herramientas se utilizaron para el desarrollo del proyecto. Finalmente respecto al "reuso" (*Reusability*), en los metadatos se debe agre-

gar la licencia que se asignará a los datos, que como recomendación debe permitir su reutilización y no ser restrictiva (aunque esto puede ocurrir), en caso de existir un período de embargo también se debe agregar, además de los años en que estará en el repositorio los datos a almacenar.

Requerimientos Externos

Como las investigaciones son realizadas gracias a financiadores públicos y en el marco de la Universidad de Chile, en caso de que existan requerimientos por parte de ellos, es decir “externos.”^{al} centro, se deben respetar con prioridad a lo estipulado en estas políticas. Si esto sucede, se deberán tomar en cuenta a la hora de realizar la revisión de las políticas (que será mencionado más adelante) para analizar su inclusión en el documento, principalmente debido a que puedan ser reglas que serán aplicadas desde ese momento en adelante por parte de una de las dos entidades mencionadas y no dejar obsoletas las políticas.

El texto del borrador para esta subsección es el siguiente:

- a. En caso de que exista algún requerimiento externo, ya sea a nivel institucional o de parte del financiador, se deben respetar estos requisitos con prioridad a las políticas de gestión de datos.
- b. En caso de que esto suceda, al momento de realizar la revisión periódica deberá ser analizado si es necesario agregar estos requerimientos a las políticas para futuras versiones.

Re-uso y Descubrimiento

Siguiendo lo mencionado tanto en las guías de recomendaciones para la creación de políticas y la revisión hecha de algunas existentes, los datos deberán ser almacenados por un período mínimo de 10 años, utilizando los formatos y metadatos especificados en el plan de manejo de datos.

Como los proyectos del centro tienen fines de convertirse en distintas aplicaciones, algunas comerciales, durante las investigaciones en curso solo los integrantes del centro tendrán acceso a los datos generados en ellos, al igual que si una vez finalizadas estos datos no se declaren como abiertos. En caso de que alguna entidad externa requiera acceso a los datos, el jefe de proyecto o la directora ejecutiva deberán aprobarlo.

A pesar de que se busca un enfoque abierto, se debe tener cuidado en el cuándo declarar los datos como abiertos. En caso de que exista la posibilidad de tener problemas legales, de privacidad, confidencialidad o éticos los datos no podrán ser considerados como abiertos.

Con el fin de proteger la investigación realizada en el centro, se debe especificar claramente si existirá un tiempo de privilegio o embargo para el acceso a los datos además de la forma en que se citará al centro y autor de los datos. Por otro lado, si existe un requisito por parte de publicadores comerciales sobre los datos, los derechos de acceso y re-uso de ellos no podrán ser otorgados sin retener los derechos de abrir estos datos.

Como se menciona anteriormente, los DOIs son uno de los mecanismos más utilizados para

la identificación y descubrimiento de los datos científicos. Con esto en mente, en caso de que se publique alguna investigación realizada en el centro y necesite de datos generados en ella para su validación, se deberá adjuntar el DOI asignado a estos datos para que sea más fácil su acceso y en caso de ser necesario, su citación.

El texto del borrador para esta subsección es el siguiente:

- a. Los datos de investigación generados deberán ser almacenados por un período mínimo de 10 años, utilizando los formatos y metadatos especificados en el plan de manejo de datos.
- b. Los derechos de acceso y re-uso no se podrán otorgar a publicadores comerciales sin retener el derecho a abrir los datos de investigación.
- c. El acceso a los datos de investigación para investigaciones que están en curso sólo será disponible para integrantes actuales del centro.
- d. El acceso a los datos de investigación para investigaciones finalizadas, si es que no se han declarado como abiertos, deberá ser exclusivo para integrantes actuales del centro de investigación y podrá ser realizado desde el momento en que se almacenan por primera vez hasta la fecha estipulada de retención.
- e. Solo se podrán abrir los datos en donde no puedan existir problemas legales, de privacidad, confidencialidad o éticos.
- f. En caso de abrir los datos se deberá especificar la existencia de tiempos de privilegio de acceso ellos, el cual deberá ser definido por el jefe del proyecto.
- g. En caso de abrir los datos de investigación se deberá especificar la forma en que se citará al centro y al autor de estos datos en los lugares que se publiquen.
- h. Si se desarrollan publicaciones referentes a cierta investigación que generó datos científicos y estos son necesarios para su validación, se deberá adjuntar información para acceder a los datos (por ejemplo, un link o en caso de que exista un *Digital Object Identifier* (DOI)) en la publicación para acceder a los datos utilizados en ella.
- i. Si alguna entidad externa requiere acceso a los datos generados, se deberá aprobar por parte de la directora ejecutiva su posible acceso. No se dará acceso cuando se puedan tener problemas legales o éticos.

Ética

En la primera sección de este capítulo se mostraron las distintas leyes y requerimientos a los que está sujeto la investigación dentro del centro, es importante recalcar que todos los integrantes del WIC deberán comprometerse a cumplir cada uno de ellos con el fin de minimizar cualquier tipo de problema que pueda llevar no hacerlo.

Con este mismo fin, se recalca que no se podrán abrir datos que sean considerables como sen-

sibles y además se deberá digitalizar y almacenar toda la documentación exigida por las leyes de Chile (ley 20.120 específicamente), el Comité de ética de la facultad o financiadores, para evitar este no cumplimiento mencionado.

El texto del borrador para esta subsección es el siguiente:

- a. Todos los integrantes del centro de investigación deberán realizar sus investigaciones cumpliendo todas las leyes, requerimientos del centro, universidad y financiadores, relacionados a la generación de datos sensibles, especialmente los referentes a experimentos en donde se utilizaron seres humanos.
- b. Toda investigación realizada en humanos deberá cumplir lo estipulado en las leyes 19.628 y 20.120 de Chile, sumado a esto, las exigencias del Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- c. No se podrán abrir conjunto de datos sensibles para evitar el riesgo al no cumplimiento de alguno de los puntos a. o b.
- d. Toda documentación generada para la realización de experimentos con humanos, exigidos por la ley, Comité de Ética o financiadores, deberá ser digitalizada y almacenada junto a los datos en el repositorio del centro.
- e. Los datos generados en experimentos con humanos son de propiedad de estos, el centro de investigación sólo cuenta con los derechos a utilizarlos en sus investigaciones, esto debe quedar plasmado en los documentos exigidos por las entidades mencionadas en los puntos anteriores.

Eliminación, destrucción y movimiento de registros

Una vez pasado el periodo de retención, que en este caso es de 10 años, se debe analizar si los datos continuarán siendo guardados por un nuevo periodo o simplemente se eliminarán. Los jefes de proyectos al estar más conscientes del potencial uso de los datos generados en cierto proyecto serán los encargados de realizar este análisis y llevar a cabo las tareas necesarias para cumplirlas, como cambiar el tiempo de expiración al nuevo seleccionado o eliminar los datos del repositorio, cualquier acción deberá ser aprobado por la directora ejecutiva.

En caso de que se considere que algunos datos almacenados no sean de ayuda o interés futuro, pero no han concluido los años del periodo de su retención, la aprobación de su eliminación al igual que los casos anteriores será por parte de la directora ejecutiva y llevada a cabo por los jefes de proyecto al que pertenezcan los datos.

Como va a existir la posibilidad de que se abran los datos, en caso de que se quieran mover registros o dejarlos abiertos en algún repositorio externo, esto también se deberá aprobar por parte de la directora ejecutiva pero llevado a cabo por los jefes de proyectos involucrados.

Finalmente, en un caso más extremo, que se cambie la plataforma de almacenamiento, el procedimiento necesario para llevarlo a cabo deberá ser realizado por todos los jefes de proyectos en conjunto a la directora ejecutiva, asegurando que se mantengan todos los estándares acordados

tanto para los formatos como los metadatos.

El texto del borrador para esta subsección es el siguiente:

- a. Una vez pasado el período de retención de algún conjunto de datos, se deberá reevaluar su almacenamiento por un nuevo período.
- b. En caso de que se desee eliminar registros del repositorio del centro, los jefes de proyectos serán los encargados de realizar esta acción.
- c. En caso de que se deseen mover registros o abrir los datos en un repositorio externo, esto deberá ser aprobado por parte de la directora ejecutiva y tendrá que ser llevado a cabo por el jefe de proyecto, siguiendo todas las normas de formatos y metadatos necesarios para su re-uso y descubrimiento.
- d. En caso de realizarse una migración a otro repositorio otorgado por el centro, el traslado al nuevo lugar de almacenamiento deberá ser realizado en conjunto por todos los jefes de proyectos y la directora ejecutiva, asegurando los estándares acordados de formatos y metadatos.

3.4.6. Roles y Responsabilidades

Utilizando la estructura organizacional del centro (Anexo A.2) se describen los roles y responsabilidades de cada uno.

Primero y por parte del centro, se debe proporcionar de toda la infraestructura de almacenamiento para los datos, entregar facilidades para su acceso y disponibilidad incentivando su re-uso tanto dentro como fuera de él. Por otro lado, dejar claro quién será el dueño de los datos generados en él, dar a conocer a todos los integrantes del centro las políticas junto a el soporte necesario para que puedan cumplir los requerimientos dados en ellas. Finalmente se debe comprometer a realizar una revisión periódica del documento diseñado y encargarse de todos los costos que puede traer consigo la gestión de datos científicos en él.

En general, cuando se trata de una institución como una universidad, estas responsabilidades están separadas en los distintos departamentos y unidades académicas que la conforman, al ser el centro tratado en esta memoria como una institución y al no existir distintas unidades como en el caso mencionado, las responsabilidades caen en el WIC. Con este trabajo se espera crear todos los mecanismos para que el centro pueda cumplir estas responsabilidades, algunos puntos quedarán plasmados en las políticas, otros en la infraestructura que se desea proponer y el apoyo que se generará para el cumplimiento y uso de estas dos partes.

El encargado o director del centro será el encargado de aprobar cualquier cambio en las políticas diseñadas y también de los documentos de compromiso acerca del cumplimiento de estas políticas por parte de los integrantes del WIC. Este último documento es diseñado en los próximos capítulos.

La directora ejecutiva, que aparte de estar involucrada en algunos proyectos está encargada de algunas labores más administrativas del centro, por esto al igual que el director deberá pre-aprobar los cambios que puedan realizarse en políticas.

Un punto importante es que para los *sets* de datos generados es que luego de un período de retención, se debe evaluar si se mantendrán o se eliminarán del repositorio del centro, en caso de que se quiera sigan almacenados se debe especificar por cuánto tiempo mínimo más hacerlo, para una vez transcurrido ese tiempo evaluarlo nuevamente. A pesar de que este periodo en este caso es de 10 años y no se sabe cómo irá desarrollándose el centro de investigación respecto a su estructura organizacional, para el actual organigrama, la directora ejecutiva será la encargada de aprobar este análisis, además de aprobar los traslados o eliminación de datos que no hayan cumplido aún este período de 10 años en caso de ser necesario, procedimientos que serán llevado a cabo por parte de los jefes de proyecto.

El siguiente escalón del organigrama son los jefes de cada proyecto en el centro, en general son personas que hicieron sus trabajos de título en ellos años anteriores o profesionales que tienen más antigüedad en el WIC. Debido a su experiencia deberán también aprobar los cambios en las políticas, en parte serán los primeros en revisarlas antes de enviarlas a la directora ejecutiva.

Los jefes de proyecto son los encargados de supervisar los trabajos que realizan los memoristas en el centro en cada proyecto, es por esto que también su labor será supervisarlos respecto a la buena gestión de los datos científicos que generen sus integrantes, deberán revisar que todos sus supervisados hayan firmado el acuerdo de cumplimiento de las políticas y en caso de estar realizando investigaciones en el proyecto, también deberán firmarlos cumpliendo ambos roles. Cuando los proyectos terminen, tendrán que realizar un reporte explicitando si se cumplieron o no los puntos de las políticas en el proyecto. Finalmente y siguiendo el rol de supervisor, deberán aprobar los planes de manejos de datos que hagan los integrantes de proyecto con el fin.

Por último, los integrantes de cada proyecto deben primero firmar el acuerdo de las políticas, identificar qué datos y documentos serán los que se deban respaldar para su re-uso, asegurar la integridad de estos durante todo su ciclo de vida, asegurar también que los datos queden respaldados en los formatos especificados en las políticas del centro y cumplir todas las acciones referidas a los protocolos sobre temas éticos y legislativos de los datos.

Algunas de estas obligaciones son respaldadas en la sección anterior, por lo que no se encuentran explícitas en esta, solo se agregaron responsabilidades en aspectos generales para el cumplimiento de las políticas.

El texto del borrador será el siguiente:

”Los roles y responsabilidades serán los siguientes:

1. El Web Intelligence Centre tiene la responsabilidad de:
 - a. Proporcionar infraestructura de almacenamiento segura para los datos de investigación generados en el mismo centro.
 - b. Entregar facilidades para el almacenamiento, disponibilidad y accesibilidad a conjuntos de datos generados en él para incentivar su re-uso dentro y fuera de este.
 - c. Especificar quién es dueño de los datos científicos generados en investigaciones realizadas en el centro.

- d. Dar soporte referente a la gestión de datos científicos a todos los integrantes del centro.
 - e. Dar a conocer las políticas del centro a todos sus integrantes.
 - f. Realizar una revisión periódica de las políticas con el fin de que no queden obsoletas en el tiempo.
 - g. Ser el responsable por los costos relacionados a la gestión de datos científicos en el centro.
2. El director del centro tiene el deber de:
- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
 - b. Aprobar los documentos de compromiso por parte de los integrantes del centro acerca del cumplimiento de estas políticas.
3. La directora ejecutiva deberá:
- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
 - b. Aprobar traslados o destrucción de datos de investigación guardados en el repositorio del centro en donde no haya terminado el periodo de retención de estos.
 - c. Analizar, una vez terminado el periodo de retención de ciertos datos, si se mantendrán o se eliminarán los sets de datos del repositorio utilizado por el centro, en el primer caso además deberá especificar por cuánto tiempo más seguirán almacenados, en el segundo caso deberá ser la encargada de eliminarlos.
 - d. Agregar en caso de aceptar nuevos formatos de archivos permitidos por el centro para subir al repositorio a la guía de metadatos en la sección formatos de los Anexos de ese documento.
4. Los jefes de proyecto deberán:
- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
 - b. Supervisar a los miembros de los proyectos sobre el cumplimiento de las políticas.
 - c. Asegurar que todos los integrantes del proyecto hayan firmado el acuerdo de cumplimiento de las políticas.
 - d. En caso de realizar algún tipo de investigación también deberán cumplir los deberes de los demás integrantes del proyecto en donde la realice.
 - e. Ser los encargados de eliminar los registros relacionados a su proyecto del repositorio, en caso de que se considere necesario.
 - f. Una vez terminado el proyecto entregar un reporte de cumplimiento de las políticas dentro del proyecto.
 - g. Aprobar los planes de manejo de datos entregados por los integrantes del proyecto que quie-

ran realizar una investigación en él.

5. Los integrantes de los proyectos tienen la responsabilidad de:
 - a. Firmar un acuerdo en donde se especifica que se tiene conocimiento y se realizará el cumplimiento de las políticas del centro.
 - b. Identificar qué datos y documentos serán los que se deban respaldar para su posterior re-uso.
 - c. Resguardar los datos que generen en sus investigaciones bajo los requerimientos de las políticas del centro.
 - d. Asegurar la integridad de los datos durante todo el ciclo de vida de estos.
 - e. Almacenar los datos de investigación en formatos y con lo metadatos suficientes para asegurar su fácil re-uso y descubrimiento.
 - f. Realizar las acciones contenidas en el documento de acuerdo a los protocolos éticos y legislativos respectivos."

3.4.7. Aprobación y revisión periódica

Como se menciona anteriormente, la revisión periódica de las políticas es esencial debido al acelerado paso que está tomando el tema de la gestión de datos de investigación en el mundo y también de a poco en el país.

Es por esto que el documento tendrá que tener una revisión periódica de al menos una vez al año, que es lo que duran los trabajos de título en promedio y lo que hacen en general las distintas instituciones revisadas que cuentan con políticas de gestión de datos, con el fin de ir mejorando cada vez las políticas y evitar su obsolescencia con el paso del tiempo. En caso de querer realizar los cambios se considera que dos meses es un tiempo prudente para reescribir el documento, el cual deberá ser firmado nuevamente por todos los miembros del centro, principalmente por dos razones: la primera es que conozcan los cambios que se realizaron y el segundo es que no existan diferencias de requerimientos entre los mismos integrantes del centro.

Utilizando el organigrama del centro, los primeros que analizarán la aprobación del nuevo documento en caso de existir, serán los jefes de proyectos, seguidos por la directora ejecutiva y finalmente por el encargado del centro.

El texto presente en el borrador es el siguiente:

"Este documento será sujeto a una revisión periódica de al menos una vez al año, con el fin de ir mejorando las políticas y evitar su obsolescencia con el paso del tiempo. Una vez hecha la revisión se deberán actualizar en menos de 2 meses una vez pasada la revisión.

Todos los cambios que se produzcan deberán ser primero revisados y aprobados por los jefes de cada proyecto y la directora ejecutiva para finalmente ser aprobadas por el director o encargado del

centro.

Si se realizan cambios se deberá firmar un nuevo acuerdo de cumplimiento por parte de los integrantes del centro con el fin de que todos sean conscientes de estos cambios, y se comprometan a seguirlos."

Anexos

No se consideraron herramientas online debido a sus problemas mencionados anteriormente, por otro lado, se contará de manera externa con todas las guías y templates necesarios para hacer efectivas las políticas. Por esto, el borrador no contiene anexos.

El borrador que se puede revisar en el Anexo C.3 del presente informe es el presentado a los encargados del centro para su revisión, con el fin de mejorarlas para su pronta aprobación. Una vez ya aprobadas se procede con el levantamiento de requerimientos en el siguiente capítulo. La sección relacionada a la infraestructura de almacenamiento se rellenará una vez determinado cuál será la solución final para esto, esta sección debe ser aprobada también por las mismas entidades recién mencionadas.

3.5. Versión Final de Políticas de Gestión de Datos

Después de presentar el documento diseñado en la sección anterior del presente informe a los encargados del centro, se considera necesario agregar los siguientes puntos:

Documento de Acceso a datos

Para algunos datos que se consideren sensibles, el acceso a ellos debe ser restringido aún para algunos miembros del centro o hasta personas del mismo proyecto al que están relacionados. Para esto debe existir un documento especial de confidencialidad para estos conjuntos de datos, que deberán firmar cada una de las personas que requieran acceso a ellos. Esto es agregado al tópico de Ética del texto de las políticas. La persona que deberá analizar si se debe o no utilizar este documento será el jefe de proyecto, lo cual también será agregado en los roles y responsabilidades de cada entidad afectada por las políticas. Este documento al igual que los otros es diseñado en capítulos posteriores.

El texto agregado en el tópico de Ética es el siguiente: " En caso de considerarlo necesario, algunos conjuntos de datos necesitarán de un documento de permiso de acceso a estos que será firmado por cada persona que desee acceso a ellos y contendrá las reglas específicas que se deberán cumplir para ellos. Este documento será entregado por el centro y deberá ser aprobado por los jefes de proyecto, además su versión digital deberá ser almacenada en el repositorio del proyecto. "

Para la sección de los roles y responsabilidades de las políticas, se agrega al punto de los jefes de proyecto lo siguiente: "Evaluar los conjuntos de datos generados en el proyecto respecto a si se

necesita un acceso especial debido a su naturaleza sensible o no, en caso de ser necesario deberá aprobar todos los documentos de requerimiento de permiso a ellos que se necesiten."

Infraestructura

Respecto a la infraestructura, ésta es seleccionada en el Capítulo 4 del presente informe, en este tópico se muestran los requerimientos a utilizar para el repositorio entregado por el centro, uno de ellos es el esquema de metadatos que se deberá a utilizar al momento de subir archivos, el cual será el creado por DataCite en su última versión, que actualmente es la 4.0, serán obligatorios para cualquier conjunto de datos que se almacene en el repositorio y que haya sido generado por algún integrante del centro de investigación. Además, en caso de que los datos generados no estén en alguno de los formatos aceptados, se deberán convertir a alguno de estos, en estos casos es obligación subir los archivos convertidos con todos los metadatos especificados.

Las guías que serán necesarias para utilizar cada una de las funcionalidades mínimas para el cumplimiento de todas las políticas son diseñadas en capítulos posteriores. El texto para la versión final sobre la infraestructura es el siguiente:

- a. Todos los datos y documentos que se generen en el marco de alguna investigación realizada en el centro, deberán ser almacenados en el repositorio de cada proyecto entregado por este.
- b. Al momento de subir algún conjunto de datos al repositorio, se deberán rellenar todos los metadatos considerados como mínimos del esquema de metadatos de la guía de metadatos entregada por el centro.
- c. Cuando se desee subir un archivo de datos científicos al repositorio, se debe asegurar que esté en uno de los formatos permitidos por el centro, en caso de no estarlo se debe convertir antes de subirlo.
- d. En caso de querer agregar formatos o metadatos que no han sido considerados a la fecha, estos deberán ser agregados y aprobados por la directora ejecutiva a la configuración del repositorio, siguiendo las guías entregadas por el centro.
- e. En caso de que un integrante del centro requiera acceder a un repositorio de otro proyecto, deberá pedir aprobación por parte del jefe del proyecto al que desea acceder.
- f. La gestión de cuentas del repositorio de integrantes y externos al centro será realizada por la directora ejecutiva.
- g. Se deberán utilizar las guías entregadas por el centro para su utilización.

Con esto agregado, se sumaron las siguientes responsabilidades:

–Para los jefes de proyecto: Aprobar y generar los accesos a los integrantes del centro que requieran permisos a su proyecto. –Para la directora ejecutiva: Agregar y aprobar nuevos formatos y metadatos a la configuración del repositorio en caso de ser requerido, gestionar las cuentas del repositorio para los integrantes y externos del centro que requieran acceso al este. –Para todos los integrantes

del centro: Leer y utilizar las guías entregadas por el centro referentes al uso del repositorio del centro.

Definiciones

Se considera que algunas palabras utilizadas dentro de las políticas necesitaban contar con una definición más explícita para el mejor cumplimiento de los puntos, estas son:

Metadatos Se utiliza la definición presentada en el capítulo dos, la cual es "la documentación para el conjunto de datos, expresada a través de una sintaxis normal, que guarda toda la información necesaria para su uso posterior. Son adjuntados a los datos de investigación."

Esquema de Metadatos Se usa la definición entregada por DataCite [86], la cual es una lista de las propiedades de metadatos básicos seleccionados para una identificación precisa y coherente de un recurso para su citación y recuperación".

Consentimiento Informado Se utiliza la definición entregada por el Comité Asesor de Bioética FONDECYT/CONICYT [87], la cual es un proceso cuyo fundamento es una conversación entre investigador y probando (voluntario). Los puntos esenciales de esta interacción personal son recogidos en un documento firmado por el investigador para ratificar el proceso de información, y por el probando para confirmar que otorga consentimiento para participar en el estudio."

Datos Primarios Se usa la siguiente definición: "Datos que son recaudados por primera vez en una investigación en el centro de investigación, es decir, aquellos que fueron recolectados en otra investigación no se consideran como primarios". Esta definición fue utilizada en el Capítulo 1 del presente informe.

Estas definiciones fueron agregadas a la versión anterior en la sección de definiciones.

La versión final de las políticas se encuentra en el Anexo C.4. Este documento será la versión 1.0 de las Políticas de Gestión de Datos Científicos del Web Intelligence Centre. Para generar este documento se siguieron todas las recomendaciones dadas por las instituciones dadas al comienzo de este capítulo, las cuales fueron extrapoladas al contexto del centro de investigación, por lo que cumplen todos los requerimientos para realizar la gestión buscada. Los principales procesos que salen de estas políticas se pueden observar de mejor manera en el Capítulo 5.3.

En el próximo capítulo se busca proponer una infraestructura tecnológica que permita almacenar los datos generados en el centro de manera que se cumplan todas las reglas y protocolos impuestas en las políticas recién diseñadas.

Capítulo 4

Almacenamiento

En el presente capítulo se buscan las posibles soluciones de almacenamiento para los datos científicos generados en el Web Intelligence Centre. Para esto primero se realiza un levantamiento de requerimientos técnicos que deban cumplir los posibles recursos a utilizar, una diferenciación entre los tipos de soluciones existentes, para luego hacer el benchmark mencionado y seleccionar las mejores soluciones respecto al nivel técnico.

4.1. Levantamiento de Requerimientos y Funcionalidades

Con el fin de buscar una solución que pueda cumplir todos los requerimientos técnicos necesarios para el almacenamiento de los datos de investigación generados en el centro cumpliendo a la vez todos los puntos presentes en las políticas diseñadas en el capítulo anterior se realiza un levantamiento de requerimientos.

Existen distintos requisitos expresados en las políticas que se traducen a alguna capacidad del software que se utiliza con el fin de cumplirlos.

4.1.1. Subida de Archivos

Primero, respecto a la subida de archivos en el centro se han generado distintos conjuntos de datos que actualmente se encuentran en distintos computadores o lugares de almacenamiento, primero se debe juntar todo en un lugar y después ir agregando los datos generados en nuevas investigaciones. Estos archivos deben ser guardados en los formatos que son permitidos por el centro, algunos pueden ser archivos pesados, por lo que el repositorio debe permitir la ingesta de estos datos.

Estos requerimientos se traducen en los siguientes puntos:

1. Debe aceptar archivos de gran tamaño (mayores a 2 GB).

2. Debe permitir subir más de un archivo a la vez.
3. Debe aceptar múltiples tipos de archivos (como mínimo los considerados como aceptables para la preservación a largo plazo por el UK Data Archive)
4. Debe tener capacidad suficiente para almacenar los datos y documentos desarrollados en el Web Intelligence Centre.
5. Debe ser escalable en capacidad de espacio de almacenamiento en el tiempo.

4.1.2. Re-uso y descubrimiento

Hay ciertos puntos importantes que permiten fomentar el reuso de los datos, el primero tiene que ver con la capacidad de usar y asignar DOIs a los datos, lo cual es requisito de las políticas. Por otro lado, en caso de que un externo o integrante del centro ingrese a algún archivo y quiera utilizarlo, debe saber cómo se quiere que se citen estos datos, además de bajo qué licencia se encuentran actualmente junto al formato y tamaño y qué versión de estos se está revisando.

Finalmente, al momento de subir los archivos, se pedirán ciertos metadatos, acá hay dos casos que funcionarían. El primero y más recomendable es que el esquema interno del repositorio sea DataCite en su versión 4.0. La segunda opción es que el repositorio permita subir archivos XML al momento de subir los archivos (donde se podría agregar los metadatos del esquema requerido por el centro), por otro lado, en caso de que funcione con un esquema distinto a DataCite, debe contar con propiedades homólogas a las obligatorias por DataCite y por el centro o permitir agregarlas al sistema del repositorio. Estas propiedades obligatorias tienen relación a la citación de los archivos por parte de DataCite, para lo que se necesitan un mínimo de datos como título, autores y año de publicación de los datos. Los requerimientos se traducen en los siguientes puntos:

1. Debe ser capaz de habilitar la recolección de metadatos por parte de proveedores de servicios externos, con el fin de hacer más visibles los datos. Un ejemplo de esto es soportar OAI-PMH.
2. Debe permitir importar/exportar metadatos en algún formato abierto como XML o CSV (en caso de que se quiera migrar a otra plataforma por ejemplo).
3. Debe permitir asignar DOIs.
4. Debe permitir asignar licencias de Creative Commons.
5. Mostrar información acerca del formato, tamaño, citación sugerida, licencia asignada y versión de los archivos.
6. Debe contar con un esquema de metadatos flexible, que al momento de subir archivos requiera como mínimo las propiedades obligatorias del esquema de DataCite en su versión 4.0 u homólogas a estas.

4.1.3. Accesos

Sobre el acceso a los datos, el cual es un requerimiento importante para el caso del centro que cuenta con datos sensibles, los cuales necesitan accesos restringidos. Por eso, estos accesos deben ser totalmente configurables por parte de las personas que gestionarán el repositorio. Esto se describe en los siguientes requerimientos:

1. Habilitar la especificación de derechos de acceso a los archivos a cualquier nivel de la colección de datos.
2. Permitir bloquear vistas públicas a datos según se desee.
3. Permitir búsquedas en el repositorio según *keywords*, como títulos, autores entre otros.

4.1.4. Otros

Por otro lado, debe permitir crear y manejar cuentas de usuario, que en este caso serán para los integrantes del centro de investigación. Respecto a la disciplina en que se enfoca el repositorio, ésta debe no ser específica de un área o ciencia ya que el centro realiza investigaciones que apuntan a distintas disciplinas, además se busca utilizar sólo un repositorio y no uno para cada tipo de investigación.

Otro punto que se considera necesario para realizar seguimiento a los datos, es tener estadísticas de cuantas descargas se realizan dentro del repositorio a través del tiempo, lo que permite analizar de mejor manera si los datos se están dando a conocer de una manera eficiente a la comunidad científica.

En algunos casos, debido a temas legales y preferencias del centro de investigación, el repositorio a utilizar debe permitir almacenar ciertos datos en servidores internos (*hosting*) y no externos (proveídos por el repositorio)

Finalmente, se necesita que el repositorio cuente con la documentación y guías de uso necesarias para realizar los cambios que se necesiten además de soporte por parte del proveedor del software o aplicación en caso de tener problemas. En caso de pertenecer a alguna organización debe estar claro cuáles son sus políticas de uso y propiedad de los archivos subidos.

Estos últimos puntos mencionados se traducen en los siguientes requisitos:

1. Gestionar cuentas de usuarios del repositorio.
2. Repositorio de datos científicos multi-disciplinario.
3. Contar con estadísticas de uso de los archivos subidos.
4. Permitir almacenar internamente (WIC) algunos datos (o todos).
5. Debe contar con la documentación y soporte de la versión que se desee utilizar o esté disponible.

El repositorio del centro debe cumplir como mínimo estos requisitos técnicos y de funcionalidades, a continuación, se describe la búsqueda de una solución que cumpla con estos requerimientos mencionados para ser utilizado como el repositorio de datos científicos del Web Intelligence Centre.

4.2. Repositorios Existentes

Durante los últimos años, la cantidad de publicaciones que referencian a un repositorio ha ido en alza [88]. Para buscar un repositorio de datos, tanto como para obtener datos como para almacenarlos, existen distintos listados de repositorios. Los dos más grandes son Open Access Directory, el cual se enfoca en repositorios para datos abiertos y Re3data.org o Registry of Research Data Repositories.

En el año 2014, existían registros de repositorios creados tanto por Re3Data.org, DataBib y DataCite, los cuales contenían una gran cantidad de repositorios indexados en sus páginas o servicios. El año 2015 se decidió unir a Re3Data.org los repositorios que se encontraban en Databib y a finales del mismo año este nuevo registro pasó a ser un servicio otorgado por DataCite [89]. Con esto DataCite busca cumplir su objetivo de hacer más confiable y fácil el buscar repositorio de datos científicos y con esto aumentar la tasa de citas a datos científicos. Actualmente existen más de 1500 repositorios indexados por parte de Re3Data.org, el crecimiento de este número se puede observar en la Figura 4.1.

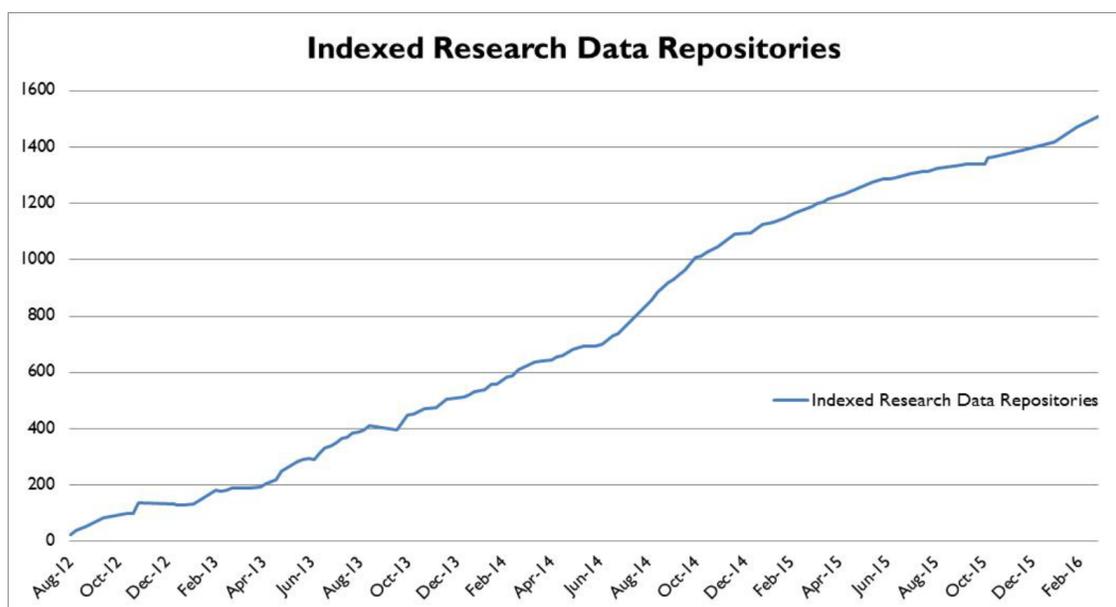


Figura 4.1: Repositorios Indexados

Fuente: Re3data.org

Para analizar un posible repositorio a utilizar para el centro de investigación se utiliza Re3data, principalmente debido a la gran cantidad de repositorios indexados en él, lo cual permite una búsqueda más rápida y efectiva. El primer paso es filtrar esta gran cantidad de repositorios con algunos

de los requisitos que debe cumplir el repositorio, utilizando las mismas herramientas que otorga la página de Re3data.org

El centro cuenta con distintos proyectos, los cuales generan datos de distintos tipos de ciencias, por ejemplo AKORI genera datos neurológicos y OPINIONZOOM utiliza tweets de Twitter para sus investigaciones. Es por esto que se utiliza como primer filtro a repositorios que estén enfocados en tópicos generales o multidisciplinarios y no de una sola rama de la ciencia. Los grandes temas para filtrar en la aplicación son Humanidades y Ciencias Sociales, Ciencias Naturales, Ciencias de la Ingeniería y Ciencias de la Vida (*Life Science*), no existe un filtro especializado en “temas generales.” sinónimos de esta frase.

Respecto Humanidades y Ciencias Sociales se considera que debe ser agregado al filtro debido a que una de las áreas que abarca es la conducta humana, en donde los tweets utilizados en las investigaciones deberían ser enmarcados. Ciencias de la Vida dentro de sus sub-áreas en Re3Data.org contiene repositorios que almacenen datos neurológicos y médicos, por lo que también debe ser considerado en este filtro.

Por otro lado, Ciencias Naturales está enfocada en repositorio de datos relacionados a Química, Física y Geociencia (incluyendo geografía), por lo que no se considera como filtro al no ser parte de los tópicos desarrollados en las investigaciones del centro. Finalmente, Ciencias de la Ingeniería si es utilizado como filtro, ya que los temas que lo conforman tienen relación directa a ciencias de la computación e inteligencia artificial, tópicos vinculados con la investigación del WIC. Con esto hecho, la cifra baja dramáticamente a 106 repositorios que contienen datos relacionados a las tres áreas utilizadas.

El segundo filtro que se utiliza tiene que ver con el tipo de repositorio que se busca, el cual para el caso de Re3data.org se separan en disciplinarios, institucionales y “otros”, que en general son repositorios en donde al momento del registro no se rellenó bien este campo. Para el caso de los institucionales, en general pertenecen a universidades de todo el mundo (principalmente de Estados Unidos, Reino Unido y Australia), el problema de estos repositorios es que son para miembros de cada institución, por lo que quedan eliminados como posibles opciones para el WIC.

De los 30 repositorios etiquetados como “otros”, 7 también son institucionales por lo que también son eliminados de la posible lista de repositorios a utilizar. Para el caso de los etiquetados como disciplinarios 4 también son eliminados por la misma razón. De los restantes, algunos están incorrectamente etiquetados e igual son institucionales o dirigidos a ciertos países o instituciones, por lo que también se eliminan de las posibles soluciones. Con esto quedan 14 repositorios en total para continuar con el benchmarking entre ellos, los cuales son: Amazon Web Services, Dataverse (Harvard) figshare, Zenodo, Dryad, ICSU World Data System, Junar, Mendeley Data, myExperiment, Open Science Framework, RunMyCode, The Data Hub, FLOSSmole, Scholarly Database y Scientific Data Repository.

A pesar de que se promueven los datos abiertos en este trabajo, el repositorio a utilizar debe poder ser capaz de no abrir ciertos datos, que pueden causar problemas éticos o legales al centro y a sus integrantes en caso de se hiciera. Al revisar las posibles opciones existen algunas que no cumplen lo mencionado, estos son: Scientific Data Repository, FLOSSmole, Dataverse (Harvard), ICSU World Data System y RunMyCode, Zenodo.

En segundo lugar, están los derechos que se les da a los datos, para el caso de esta memoria se trabaja con Creative Commons (CC), por lo que el repositorio debe ser capaz de agregar este tipo de licencia a los datos que se suban en él. Los únicos repositorios que operan con este tipo de licencia son: figshare, Dataverse (Harvard) Zenodo, Mendeley Data, MyExperiment y Open Science Framework.

Finalmente, estos repositorios no permiten administrar localmente archivos para el repositorio, es decir, hay que subir a su servidor todos los datos que se quieran almacenar con excepción de MyExperiment. Este último permite crear colecciones en las que en vez de subir un archivo, estos pueden ser redirigidos de algún link (es decir, pueden estar almacenados en otro servidor, el cual podría ser interno por ejemplo), pero para este caso el centro debería contar con un repositorio online para hacerlo, lo cual se está buscando crear. Con esto, se concluye que no existe una solución factible para el centro de los repositorios existentes.

4.3. Software para Repositorios de Datos Científicos

Gracias al movimiento que se ha generado con la gestión de datos científicos, se han creado distintos software y plataformas especializadas para generar repositorios a distintas instituciones. Una parte de los repositorios que se encuentran registrados en Re3data.org están construidos en base a ellos. Los software principales son ArchivesSpace, DataVerse, CKAN, CONTENTdm, DSpace, EPrints y Fedora.

ArchiveSpace es una aplicación web de código abierto de gestión de información de archivos, su objetivo es apoyar a la gestión y proveer acceso web a archivos, manuscritos y objetos digitales. Su lema es “construida para archivistas por archivistas”.

Dataverse es un es una aplicación web de código abierto para compartir, preservar, citar, explorar y analizar datos científicos. Facilita la disponibilidad de los datos a los usuarios y permite replicar los trabajos de otros de manera más simple, en donde para los investigadores, autores de datos, publicadores, distribuidores de datos e instituciones afiliadas reciben crédito y visibilidad en la web [90].

CKAN es un software gratuito y de código abierto. Es una solución para hacer los datos accesibles, proveyendo herramientas para publicar, compartir, buscar y usar estos datos. Está enfocada en publicadores de datos como gobiernos, empresas y organizaciones nacionales o cercanas al Reino Unido [91].

CONTENTdm es un sistema de gestión de colecciones digitales. Es un servicio de OCLC que es una cooperativa mundial de bibliotecas que se enfocan en dar servicios tecnológicos para sus miembros. Cuenta con una versión de prueba de 60 días para su uso.

DSPACE es un software para organizaciones académicas, comerciales o sin fines de lucro para la construcción de repositorios digitales. Es gratis y permite un acceso abierto y fácil a todo tipo de contenido digital. Su foco principal son los repositorios institucionales. Además trabaja en conjunto con el Instituto Tecnológico de Massachusetts (MIT) y es respaldado por DuraSpace [92].

EPrints es una plataforma de repositorios digitales de código abierto, fue desarrollado por la Universidad de Southampton (UK). Lleva más de 15 años proveyendo estabilidad y servicios innovadores de repositorios para el sector académico [93].

Fedora, al igual que DSpace está administrado por DuraSpace y se define como un sistema para repositorios robusto, modular y de código abierto, para la gestión y difusión de contenidos digitales. Se utiliza para bibliotecas y archivos digitales, además para dar un acceso especializado a colecciones digitales grandes y complejas, además de a datos científicos [94].

Dentro de estos seis posibles software para la construcción de un repositorio, sólo dos están enfocados en repositorios institucionales o para universidades, estos son Dspace y EPrints.

Por otro lado, CKAN se enfoca más en publicadores de datos del Reino Unido y Fedora, a pesar de ser administrado por DuraSpace y tener algunas universidades que cuentan con repositorios construidos en su plataforma, está destinado a archivos y bibliotecas de mayor magnitud y tipos de datos que a los de una universidad, siendo ejemplos casos como de Librerías Nacionales de Eslovaquia, Suecia, Portugal, Estonia, Francia y Escocia entre otras. Para el caso de CONTENTdm, éste es un Saas (Software as Services) y sólo permite configurar el sitio web, pero ellos almacenan los datos, por lo que queda descartada como posible opción.

En cuanto a ArchiveSpace, para acceder a la documentación completa con el fin de realizar el análisis técnico, se necesita ser miembro de su comunidad, lo cual es pagado, por esta razón se elimina de la comparación debido a que no se puede acceder por completo a las funcionalidades y saber realmente si cumple o no los requisitos buscados.

Con esto sólo quedan tres posibles soluciones: Dspace, EPrints y Dataverse, las cuales son las más populares a la hora de diseñar repositorios institucionales a nivel mundial [95]. Actualmente DSpace se encuentra trabajando en el lanzamiento de la versión 6 de su plataforma y su versión actual es la 5.6, en cuanto a EPrints en la 3.3 y Dataverse la 4.5.1. Los tres cuentan con un sitio web con la documentación de sus respectivas documentaciones [96] [97] [98].

Al realizar el análisis técnico de estas dos plataformas, utilizando las documentaciones, las versiones Demo de cada plataforma y algunas comparaciones realizadas por distintas instituciones como UNESCO [99], se observa que Dspace y Eprints cumplen todos los requerimientos, algunos en distintos niveles, pero a modo general si lo hacen. de los requerimientos que debe tener el sistema. Dataverse por otro lado, tiene sólo soporta el esquema de DataCite versión 3.1 y no la 4.0 que es la que se utiliza en este trabajo.

Una de las principales desventajas de DSpace es que es un poco más complicado de instalar que EPrints [100], a pesar de esto DSpace cuenta con uno de los mejores soportes y apoyo profesional en comparación a otras plataformas, en donde EPrints es bastante más inferior. Por otro lado, el sistema de su estructura interna del repositorio permite tener una granularidad mucho mayor y simple a la hora de administrar los accesos a los datos que se deseen subir que EPrints [101], además de permitir búsquedas internas más avanzadas gracias a su soporte de Dublin Core Qualified [102].

Como las dos soluciones por defecto no vienen configuradas para cumplir todos los requerimientos, es necesario modificar en algunos casos las aplicaciones internas, o en caso de que se quiera agregar algo nuevo, para esto cada una de las plataformas está construida en distintos len-

guajes de programación, EPrints utiliza Pearl y DSpace Java. En la Facultad de Ciencias Físicas y Matemáticas dentro de sus ramos para las generaciones anteriores al 2011, en uno de sus ramos se enseñaba el uso del lenguaje Java para todos los nuevos alumnos de esta facultad, actualmente todos los integrantes del centro, exceptuando los memoristas que ingresaron este año son de estas generaciones. Al consultar a los integrantes que están encargados del desarrollo de software del centro, los cuales serán los encargados en el futuro de la mantención de este, el 100 % mencionó que tiene más conocimientos en Java y prefiere este lenguaje por sobre Pearl, lo cual se puede explicar por lo recién comentado.

Debido a las razones recién presentadas, se utiliza la plataforma de DSpace para el diseño del repositorio del WIC. La construcción de este repositorio se detalla en la siguiente sección.

4.4. DSpace

Para diseñar el repositorio del centro se utiliza la versión 5.6 de DSpace, que es la más actual lanzada a la fecha, partiendo por la instalación y configuración inicial de la plataforma utilizando la documentación [96] que ofrece DuraSpace, organización que administra DSpace. Se realiza el montaje del repositorio en un computador personal con el fin de entregar una propuesta de almacenamiento para el centro.

Para realizar las configuraciones necesarias al repositorio, se deben editar archivos de dos carpetas, la principal llamada Dspace y otra llamada jspui que se encuentra en la carpeta webapps alojada en el servidor, que en este caso es local. A continuación se describen los cambios y configuraciones realizadas en las dos carpetas mencionadas y en la aplicación ya montada.

4.4.1. Carpeta Dspace

Primero se cambiaron algunos parámetros del archivo dspace.cfg (el cual se encuentra en Dspace/config) con el fin de responder a los requerimientos del centro, además de realizar algunas personalizaciones, los cambios fueron los siguientes:

–Se cambió la ruta por defecto de xmlui a jspui, los cuales son las dos interfaces que ofrece DSpace, se eligió JSPUI debido a que tiene un mejor diseño, experiencia de usuario y facilidad de aprendizaje (pensando en que es primera vez que se va a utilizar este tipo de repositorios en el centro), a pesar de que es un poco más complicado realizar cambios de *layout* que XMLUI, pero éste último consume y necesita mucho más hardware para tener un rendimiento considerable. Para esto se cambió la línea 39 del archivo.

–También se configuraron las líneas 54 y 57 para configurar la base de datos creada en PostgreSQL, se prefirió este sistema de administración, ya que Oracle está más enfocado a bases de datos más grandes y se busca diseñar el repositorio para el centro y no para la universidad o para un archivo nacional donde se podría justificar mejor. La base de dato en DSpace sólo guardan en su mayoría metadatos no los archivos que se suben. Además consume más recursos y es más difícil de administrar que PostgreSQL.

–Se realizaron las configuraciones de email, utilizando por el momento un email gmail personal. Para esto se cambiaron las líneas 98, 101, 102, 108, 116, 119, 122, 125 y 137 del archivo dspace.cfg. En la última línea mencionada se quitó dejó como no comentado el código de esa línea con el fin de poder utilizar gmail, en caso contrario existirían problemas.

–También se configuró la API de Creative Commons para poder otorgar las licencias a los archivos subidos. Para esto se dejó como comentado el código de esta línea, ya que si uno selecciona una jurisdicción específica de la licencia dspace otorgaría licencias de versión 3.0, las cuales no son las requeridas por el centro.

–Por otro lado, para activar la capacidad de dar un periodo de embargo a los archivos subidos se cambió la línea 914 del archivo, en donde se dejó como no comentada (se eliminó el signo #)

–Finalmente, la configuración por defecto hace permitir archivos de máximo 512 MB, para dejarlo como indefinido se cambió el valor del parámetro de la línea 1734 por -1.

Dentro de la carpeta Dspace también se cambiaron los archivos que se encuentran dentro de la carpeta email, los cuales son los templates para los emails que se envían durante el uso de la plataforma, por ejemplo los de registro, asignación de tareas y alertas enviadas al administrador entre otros. Se cambiaron las frases en donde decía “Dspace” por “Repositorio Web Intelligence Centre”

Para aplicar las licencias Creative Commons y periodos de embargo, se editó el archivo item-submission.xml, el cual indica las distintas etapas que se realizarán al momento de subir un archivo al repositorio. En él hay dos etapas cuatro, una comentada (línea 230) y otra que es la que viene por defecto (línea 220), se procedió a cambiar la no comentada por la por defecto. Para el caso de las licencias, existe un paso 6 en el archivo el cual se encuentra comentado (si una etapa está como comentario del archivo no se aplicará en el proceso), ya que por defecto no viene activada la aplicación de licencias Creative Commons, para poder utilizarlas se les quitaron los tags usados para insertar comentarios en archivos XML los cuales son iguales a los de HTML. (Línea 250)

En la página inicial de la aplicación se muestra el contenido que se encuentra en news-top.html. Se cambió el contenido que viene por defecto a “Bienvenidos al Repositorio del Web Intelligence Centre”.

Finalmente, para subir archivos estos deben ser descritos, es decir, agregar metadatos de los items que se están subiendo. La plataforma interna funciona con el esquema Dublin Core, la cual no puede ser cambiada por otra de acuerdo a su documentación, pero permite agregar en la misma interfaz nuevos esquemas y propiedades. Por esto, se agregaron al archivo input-forms.xml algunas propiedades del esquema de DataCite para que estas sean requeridas en las primeras etapas de la creación de un ítem en el repositorio..

4.4.2. Carpeta Webapps/JSPUI

Por otro lado, para editar el *layout* de la página, se cambiaron los siguientes archivos:

En la carpeta discovery, en el archivo static-sidebar-facet.jsp se editó la línea 64 para cambiar el

mensaje por defecto a “Descubrimiento”

En la carpeta image se agregaron los logos del centro de investigación que irán en la página en vez de los por defecto de Dspace.

En la carpeta layout se cambiaron cinco archivos:

–footer-default.jsp, en donde se cambió el footer de la página, se cambió lo existente y se agregó el logo del centro y la declaración “Web Intelligence Centre, Universidad de Chile. Algunos derechos reservados. 2016”

–header-default.jsp, en donde se cambió el título por “Repositorio WIC”, además se agregó el logo y la frase “Bienvenidos al Repositorio de Datos Científicos. El WIC pertenece a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.” al inicio de la página.

–navbar-default.jsp, en donde se cambió el logo de DSpace por el del WIC en el menú superior de la página. También se cambió la palabra Communities por Proyectos, ya que las comunidades serán los proyectos del centro, esto se explicará en la guía para el uso del repositorio que se diseñará en el capítulo 5 del presente informe. También se cambió el término “My Dspace” por “Mi Repositorio”.

–navbar-admin.jsp, en donde también se cambió el logo de DSpace por el del WIC en el menú superior de la página.

–navbar-minimal.jsp, en donde nuevamente se cambió el logo de DSpace por el del WIC en el menú superior de la página.

Finalmente, para los archivos que se encuentran directamente en la carpeta jsps se realizaron los siguientes cambios:

En el archivo community-home.jsp se eliminó la línea 89 que mostraba una frase por defecto.

En el archivo community-list.jsp se cambiaron los mensajes por defecto a Proyectos del WIC y Revisa los distintos proyectos del centro de las líneas 149 y 150 respectivamente.

Finalmente, en el archivo home.jsp se realizaron los mismos cambios que en community-list en las líneas 166 y 167. También se eliminaron las noticias (Topnews y SideNews) para que no se muestren en la página, ya que el centro cuenta con su propia página en donde la información del centro es presentada.

4.4.3. Aplicación

Con los cambios realizados, se procedió a entrar a la página en donde el inicio de la página, que es lo que ven los usuarios al ingresar al repositorio, como se observa en la figura 4.2



Figura 4.2: Página de Inicio

Fuente: Elaboración propia

Una vez dentro se creó una cuenta de administrador para realizar las configuraciones pertinentes para cumplir todos los requerimientos y se realizaron ciertas pruebas con el fin de aprender el cómo funciona el repositorio internamente.

La primera acción fue agregar los formatos requeridos por el centro al registro de formatos del repositorio, después se agregó algunas propiedades de metadatos que se necesitarán completar al momento de subir archivos, es decir los utilizados por el esquema de DataCite, aparte de los obligatorios por parte de Dspace.

Con esto terminado, se procedió a crear comunidades, sub-comunidades y colecciones a modo de prueba, donde también se agregaron archivos de prueba con el fin de analizar si se cumplían todos los requisitos técnicos. Se crearon cuentas de prueba para revisar el funcionamiento de los accesos y la asignación de las licencias. Después de realizar pruebas para cada uno de los requerimientos se concluye que el repositorio diseñado cumple con todas las necesidades impuestas por las políticas de gestión de datos diseñadas en el capítulo anterior exceptuando por uno, el cual es el poder asignar DOIs.

Para asignar DOIs a archivos o publicaciones es necesario ser miembro o socio de alguna agencia de registro de DOIs como DataCite o CrossRef. Como esto actualmente no es posible de realizar, ya que existen requisitos monetarios por parte de ellos, no se realizaron pruebas para validar el cumplimiento de esta funcionalidad requerida. A pesar de esto, DSpace es capaz de integrar la asignación de DOIs, lo cual se puede revisar en su documentación [103]. En caso de que el centro decida implementar la arquitectura, será necesario convertirse miembro de alguna de estas agencias de registro de DOIs, después de esto se necesitaría montar el repositorio construido de manera local en alguno de los servidores con los que cuenta el centro y configurar las asignaciones de DOIs para el futuro. Con esto también se podrá utilizar una función de conversión de metadatos (de los que utiliza Dspace a Datacite) para lo cual se deberá editar el archivo DIM2DataCite.xml de la carpeta config/crosswalks del lugar donde esté alojada la carpeta principal de dspace.

La propuesta final de almacenamiento para los datos científicos del centro de investigación es entonces el repositorio construido en la plataforma DSpace y configurado para que cumpla todos los requerimientos y funcionalidades necesarias para cumplir las políticas de gestión de datos cien-

tíficos, menos la mencionada anteriormente. En el siguiente capítulo se desarrolla todo el soporte necesario para este mismo fin, contando una guía para el uso del repositorio, pensando en que como cualquier software nuevo, éste se debe aprender a utilizar y por lo que su correcta utilización es vital si se quiere realizar una gestión de datos científicos efectiva en el centro.

Capítulo 5

Soporte

En el presente capítulo se diseñan todos los documentos o protocolos que se deberán seguir con el fin de que cada integrante del centro sea capaz de cumplir los requerimientos introducidos en las políticas de gestión de datos de investigación del centro.

5.1. Guías

A continuación, se diseñan las guías de licencias, metadatos y repositorio, necesarias para el cumplimiento de las políticas diseñadas.

5.1.1. Licencias

Esta guía servirá para dar a conocer las licencias de Creative Commons en su versión 4.0 que utilizará el centro de investigación para los datos generados dentro de él. La licencia escogida será la que se ingresará a la hora de rellenar los metadatos cuando se agregue un archivo al repositorio del centro.

La elección del tipo de licencia va a depender de cada conjunto de datos que se genere. La licencia permite especificar las condiciones que deben cumplir los terceros que quieran reutilizar los datos cuando estos se publiquen. Esta publicación no implica que sean abiertos o que no requieran un acceso a estos por parte de los terceros, cuando se habla de publicación de datos se habla del momento en que los datos están disponibles para descargar por parte de la comunidad científica, pero esta disponibilidad no significa que no necesiten un permiso especial o algo parecido.

Respecto a la elección, no se va a imponer un tipo de licencia específico a excepción de los datos que sean considerados como sensibles, para los cuales se deberán guardar todos los derechos y nunca podrán ser disponibles para su descarga para externos. La razón es principalmente debido a que aplicar una licencia a este tipo de datos no da el derecho a hacerlos disponibles, sea la licencia permisiva o no, debido a temas legales y de confidencialidad de los datos.

La estructura del documento será la siguiente:

Se empieza con una introducción, dando a conocer la definición en el contexto de los datos de las licencias y se presenta a Creative Commons, que es la organización que entrega estas herramientas legales para la correcta reutilización de los datos. Se continua dando una declaración del objetivo del documento, el cual fue mencionado al principio de esta subsección y se prosigue a explicar los pasos para asignar las licencias junto una explicación de las cinco condiciones que se pueden aplicar a los datos usando CC para finalmente describir las siete licencias de Creative Commons y como se utiliza cada una.

El texto que se encuentra en la guía es el siguiente:

Introducción

“En el contexto de los datos científicos, las licencias permiten especificar los términos de usos que se quiera imponer a los terceros que quieran reutilizar los datos.

Para facilitar esto, existen licencias estandarizadas a nivel internacional. Una de estas es Creative Commons (CC), la cual es la utilizada por el Web Intelligence Centre.

CC es una organización sin fines de lucro que permite el reuso y que se comparta (sharing) la creatividad y conocimiento a través de la entrega de herramientas legales de manera gratuita.

Generalmente estas licencias se utilizan en las distintas publicaciones de investigaciones o libros, para esto utilizan la declaración de la forma presentada y en ciertas ocasiones la imagen de la licencia, lo cual es recomendado por CC. Desde su versión 4.0 permite aplicar una licencia a conjuntos de datos."

Objetivo

“El objetivo de este documento es dar a conocer las posibles licencias a aplicar a los datos generados en el Web Intelligence Centre. La licencia escogida será la asignada a los metadatos de los archivos que serán agregados al repositorio del centro en la propiedad “Rights”."

Contenido

“Para asignar una licencia de Creative Commons se deben realizar dos pasos:

El primer paso es al momento de subir un archivo, en una de las etapas se debe asignar el tipo de Licencia de Creative Commons que se crea conveniente utilizar (o no), las combinaciones serán presentadas en la sección licencias de este documento.

El segundo paso es al momento de crear el archivo XML de metadatos del archivo a subir al repositorio del centro, se deben agregar los metadatos de la propiedad de Rights y su subpropiedad

rightsURI, esta última es importante ya que redirige al código legal de la licencia. Los valores que se deben agregar en estos puntos se detallarán en la sección licencias de este documento.

Otra opción es agregar una declaración en la página donde se descargue el archivo de esta forma:

This data is available under the “nombre de la licencia” whose full text can be found at: “URI de la licencia”

Para el caso del centro, se mostrará en estas páginas la licencia que se agregue a los datos, por lo que sólo es necesario agregar la licencia a la hora de subir un archivo al repositorio en la etapa de metadatos. "

Las condiciones y licencias se presentan como se puede observar en las figuras 5.1 y 5.2 respectivamente.

Condiciones

Existen 5 tipo de condiciones que se aplican a los conjuntos de datos, estas son:

BY: Reconocimiento (Attribution)



Obliga a dar atribución o el crédito apropiado enlazando a la página de la licencia.

Figura 5.1: Ejemplo de condiciones de Guía de Licencias

Fuente: Elaboración Propia



Obliga a dar atribución o el crédito apropiado enlazando a la página de la licencia.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution 4.0 International License (CC BY 4.0)

rightsURI: <https://creativecommons.org/licenses/by/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”: **Creative Commons**.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **Yes**
- Allow modifications of your work? A lo cual se debe responder **Yes**

Con esto hecho, se asignará la licencia CC-BY al archivo en el repositorio.

Figura 5.2: Ejemplo de licencia de Guía de Licencias

Fuente: Elaboración Propia

Se debe considerar agregar al final de documento, todos los cambios agregados en las versiones posteriores a la 1.0 para contar con un registro de todas las variaciones realizados a este.

Finalmente, al igual que el documento de las políticas en el encabezado se muestran los logos del centro y de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Por otro lado, se agrega una tabla de contenidos o índice donde los números de página irán en el pie de estas. La versión que es la 1.0 irá en la portada de la guía. La guía se puede observar en el Anexo D.1

5.1.2. Metadatos

Este documento sirve de guía para revisar qué significa cada propiedad con la que cuenta el esquema de metadatos de DataCite seleccionado.

En el esquema existen propiedades que son obligatorias y otras opcionales, a pesar de que siempre se recomienda rellenar la mayor cantidad posible de metadatos, se aplicarán las mismas restricciones para el caso del centro para la mayoría de los casos, exceptuando por el momento Identifier.

En la actualidad el centro de investigación no puede asignar DOIs ya que no es un miembro activo de alguna organización que permita hacerlo. Debido a esto, hasta que esto no se pueda realizar la propiedad de Identifier no será obligatoria y cuando esto suceda, este documento deberá ser actualizado con el fin de explicar este proceso.

Algunas de las propiedades opcionales pueden ser útiles para el centro a la hora de encontrar los distintos *sets* de datos dentro del repositorio. Es por esto que las propiedades Rights (junto a su subpropiedad righthsURI) y FundingReference (junto a sus subpropiedades funderName, funderIdentifier, funderIdentifierType, awardNumber, awardURI y awardTitle) se consideran obligatorias en caso de contar con toda la información necesaria (no todos los financiadores tienen un identificador único como un DOI, por lo que las subpropiedades funderIdentifier y funderIdentifier serán opcionales). Para el caso de los documentos como planes de manejo de datos, consentimientos informados y otros que sean utilizados sólo para uso interno del centro y no se publicarán a externos, no es obligación rellenar los metadatos aparte del título, la fecha de publicación (que para estos casos será la de creación) y autor de estos. Respecto al formato de estos últimos, al ser principalmente texto será necesario subirlos en formato PDF.

La propiedad Rights se define como las licencias aplicadas a los recursos y FundingReference como información acerca de los financiadores del recurso. La primera es importante ya que es necesario y obligatorio asignar una licencia a los datos, sea restrictiva o permisiva dependiendo de la sensibilidad de los datos, así se hace más fácil verificar cuales datos están con cada licencia y asignar las correctas en caso de existir errores. Para la segunda propiedad, como el centro de investigación funciona principalmente con fondos públicos como FONDEF y FONDECYT, es importante saber a cuáles datos fueron generados por cada concurso ganado por el WIC y así tener un registro de esto para trabajos futuros.

Por otro lado, el esquema de DataCite no cuenta con una propiedad de citación, como sí lo hace el esquema utilizado por DSpace. Como se quiere fomentar el reuso de los datos, además de que se realice de manera correcta, se deberá también rellenar el metadato de citación al momento de subir

archivos en el repositorio del centro, con el formato presentado en la siguiente subsección.

La estructura de este documento es la siguiente:

Primero se presenta una introducción a los metadatos, utilizando la definición descrita en el capítulo 2 del presente informe y el objetivo del documento, el cual es ayudar a la comprensión de las propiedades del esquema con el fin de que sean finalmente utilizados por los integrantes del centro. En el esquema existen 19 propiedades junto a sus derivadas subpropiedades. Se muestra una tabla por cada propiedad definiéndolas y describiendo sus restricciones (si es obligatorio, opcional, si puede ser repetido en más de una ocasión o debe ser único), al comienzo de cada tabla también se adjuntan ejemplos en lenguaje XML sacados principalmente de la página de DataCite y otros inventados bajo el contexto del centro para su mejor entendimiento.

Después se presentan las descripciones y definiciones de cada propiedad, las cuales son traducciones de la documentación del esquema [104] junto a pequeños ajustes al igual que en algunos ejemplos donde se utilizan ejemplos más cercanos al WIC.

El texto que irá al comienzo del documento es el siguiente:

Introducción

“En el contexto de los datos científicos, los metadatos son la documentación para el conjunto de datos que guardan toda la información necesaria para su uso posterior.

La definición más común para metadatos es “datos acerca de datos”. A pesar de esto un metadato es en sí un dato, puede ser almacenado y manejado en una base de datos. Un metadato es un dato que es usado para describir a otro, es decir, su uso lo hace un metadato.

En este documento se presentará el listado metadatos o propiedades del esquema de DataCite en su versión 4.0, los cuales tienen el fin de proveer una identificación precisa de un recurso para propósitos de citación y recuperación o descubrimiento.

Algunas propiedades tienen subpropiedades y algunas de estas subsubpropiedades, las cuales pueden ser opcionales u obligatorias dependiendo cada caso.

Existen algunas que son opcionales pero altamente recomendadas. Para el caso del Web Intelligence Centre será requisito utilizar como mínimo las propiedades que sean consideradas como obligatorias por DataCite y las además “Rights” y “FunderReference”. Por otro lado, se recomienda rellenar la mayor cantidad de metadatos posibles, sean opcionales o no.”

Objetivo

“El objetivo de este documento es dar a conocer las propiedades del esquema de DataCite en su versión 4.0 para facilitar su comprensión y utilización por parte de los integrantes del Web Intelligence Centre a la hora de agregar archivos al repositorio del centro. Se presentará cada propiedad

y a sus derivados junto a ejemplos en lenguaje XLM, definiciones y en algunos casos los valores posibles para la propiedad."

Posterior a esto se presentarán todas las propiedades como se explica en el objetivo del documento, son 19 propiedades junto a sus derivados, dejando un total de 68 definiciones. También se mostrarán los pasos para agregar los metadatos a los archivos, estos pasos deberán ser seguidos por todos los integrantes del centro a la hora de subir datos científicos al repositorio."

Propiedades

"Para las propiedades que tienen una lista de valores aceptados, estos valores son explicados con mayor detalle en los apéndices de la documentación del esquema de DataCite (página 27 en adelante), disponible en su página web al igual que alguno de los ejemplos y las traducciones realizadas para las definiciones de cada propiedad.

Se deberán agregar metadatos en dos partes:

La primera es al momento de subir un archivo, donde el repositorio requiere de algunos datos para su realización. A pesar de no utilizar directamente el esquema de DataCite, algunas de las propiedades son homólogas, solo que a veces con distintos nombres.

Por esto, se deberán rellenar los siguientes metadatos en la primera etapa de subir archivos al repositorio: Authors (que sería Creator en el esquema DataCite), Title, Publisher (que tienen el mismo nombre), Date of Issue (que sería PublicationYear en DataCite) y Citation, el cual tendrá que tener el siguiente formato: Creator (PublicationYear): Tittle. Version. Publisher. ResourceType. (esto utilizando las definiciones del esquema de DataCite, las cuales se encuentran en este documento).

Luego en la tercera etapa se deberán rellenar todas las propiedades que se encuentran en ella, es decir "CC License" (Rights), CC License URI (rightsURI), Funding Information (FunderReference), Funder Name (funderName), Funder Identifier (funderIdentifier), Project Name (awardTitle) y Project Code (awardURI), donde los nombres entre paréntesis representan las propiedades o subpropiedades del esquema de DataCite.

A pesar de que sólo las mencionadas son obligatorias, se recomienda también rellenar los demás campos en las tres primeras etapas de agregar archivos del repositorio. Estos metadatos serán mostrados en la información del archivo en el repositorio, pero no son necesariamente iguales a los pedidos por el centro, los cuales se deberán rellenar en la segunda parte de agregar metadatos:

Esta segunda parte, se puede realizar al momento de subir los archivos o posterior a esto. Esta parte consiste en crear un archivo XML siguiendo como ejemplo el Anexo Ejemplo completo archivo XML de este documento. Aquí se tendrán que rellenar como mínimo las propiedades Creator, Title, Publisher, PublicationYear, ResourceType, Rights y FunderReference. Se recomienda rellenar la mayor cantidad de metadatos en este archivo, recordar que no son necesariamente las mismas propiedades que se muestran en el repositorio a la hora de subir archivos, estas son exclusivamente utilizando el esquema de DataCite en su versión 4.0.

Una vez creado este archivo, se puede adjuntar junto al archivo que se desea agregar al reposi-

torio en la etapa de seleccionar archivos a subir. La segunda opción es una vez subidos los datos, ingresar a la página (ítem) que se generó y seleccionar EDIT en el menú izquierdo, al final de la página saldrá un botón verde con la opción Add Bitstream, donde se seleccionará este archivo, el cual debe tener el nombre metadata-datacite.xml (formato XML).

A continuación, se presentarán las propiedades del esquema de metadatos de Datacite en su versión 4.0, junto a su descripción. Cuando se muestra una subpropiedad obligatoria, es sólo en el caso de que la propiedad principal sea declarada."

Un ejemplo de como se ven las tablas por propiedad se puede observar en la figura 5.3.

<p>Identifier</p>
<p>Ejemplos</p> <pre><identifier identifierType="DOI">10.5072/example-full</identifier></pre>
<p>1. Identifier</p> <p>Restricción: Es una propiedad opcional (por ahora) y debe ser única. Definición: Es un string único que identifica el recurso.</p> <p>Al utilizar el esquema de DataCite, el identificador debe ser un DOI (Digital Object Identifier) registrado.</p> <p>El formato debe ser "10.1234/foo". Donde foo es una combinación de letras.</p>
<p>1.1. IdentifierType</p> <p>Restricción: Es una propiedad obligatoria y debe ser única. Definición: Es el tipo del identificador.</p> <p>En este caso el único valor aceptado por DataCite es DOI. Existen otros tipos de identificadores, pero al utilizar el esquema de DataCite se hace obligatorio utilizar los DOIs.</p>

Figura 5.3: Ejemplo de contenido de Guía de Metadatos

Fuente: Elaboración Propia

En caso de agregar nuevos formatos permitidos por parte del centro, se deberá editar la sección formatos de los anexos de este documento y agregar una declaración de que estos formatos ahora serán permitidos. En la versión 1.0 no se agregaron nuevos, por lo que la declaración es la siguiente:

"En la presente versión de este documento, no se han agregado nuevos formatos permitidos por parte del Web Intelligence Centre, por lo que los formatos permitidos son los considerados como aceptables por parte del UK Data Archive, los cuales se muestran a continuación."

Donde después de esta declaración se muestra la tabla de formatos considerados como aceptables por parte del UK Data Archive, la cual se encuentra en el Anexo C.2

Se debe considerar agregar al final de documento, todos los cambios agregados en las versiones posteriores a la 1.0 para contar con un registro de todas las variaciones realizados a este.

Finalmente, al igual que el documento de las políticas en el encabezado se muestran los logos del centro y de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Por otro lado, se agrega una tabla de contenidos o índice donde los números de página irán en el pie de estas. La versión que es la 1.0 irá en la portada de la guía. La guía se puede ver en el Anexo D.2

Citación

La utilización de ciertas propiedades de manera obligatoria (*mandatory*) tiene distintas razones, una de estas es el formato de citación que recomienda DataCite utilizando sus cinco primeras propiedades: Creator, PublicationYear, Title, Publisher e Identifier. Con esto la citación quedaría de la siguiente manera:

Creator (PublicationYear): Tittle. Publisher. Identifier

Utilizando otras propiedades que son recomendadas, como Version y ResourceType la citación quedaría de la siguiente manera:

Creator (PublicationYear): Tittle. Version. Publisher. ResourceType. Identifier

Un ejemplo completo otorgado por la documentación del esquema es:

Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V. 2.1. Geological Institute, University of Tokyo. <http://doi.org/10.1594/PANGAEA.726855>

En caso de que algún externo encuentre datos a través de DOIs o Metadatos en el buscador de DataCite, éste recomendará citar de la manera recién mostrada siempre y cuando estén los metadatos mínimos requeridos.

5.1.3. Repositorio

La correcta utilización del repositorio es necesaria si se desea realizar una buena gestión de datos científicos en el centro. Este documento sirve de guía principalmente para los nuevos usuarios de la infraestructura creada a la hora de la utilización de esta.

Para su diseño, se revisaron todos los requerimientos y funcionalidades analizadas en el capítulo anterior, realizando cada una de las actividades por separado.

A pesar de que DSpace cuenta con más funcionalidades de las requeridas, sólo se enfoca el diseño de la guía a las consideradas como mínimas para el cumplimiento de todos los requerimientos, en versiones posteriores a este documento se podrían agregar nuevas funcionalidades que se consideren necesarias o se quieran utilizar.

Un punto importante es la estructura para el almacenamiento de archivos en el repositorio. Primero se tienen las comunidades, que para el caso de una universidad serían las distintas facultades de ésta, después las sub-comunidades, que serían los departamentos de cada facultad, después las

colecciones que serían las entidades que conforman cada sub-comunidad, como centros de investigación de cierto departamento, cursos, sub-departamentos, entre otros. Finalmente se tienen los items, que serían las distintas investigaciones o conjuntos de archivos que se deseen subir en cada colección. Una comunidad puede tener múltiples sub-comunidades, cada sub-comunidad puede tener dentro más sub-comunidades (hasta el nivel que se desee) y cada una de ellas múltiples colecciones. Cada colección puede tener múltiples items, y cada item múltiples archivos, para este caso se debe tener cuidado, ya que los metadatos son a nivel de item, no de cada archivo que contiene cada item, por lo que si se quiere tener mayor granularidad se debe crear un item por cada archivo que se desea describir a través de los metadatos requeridos.

Para el caso del centro, las comunidades se traducen en los proyectos del centro, como AKORI y DOCODE entre otros. Las sub-comunidades se descartaron debido a la estructura organizacional del centro, donde se tienen por cada proyecto sólo integrantes de este por separado y no sub-grupos. A pesar de esto se deja decisión del centro la utilización de sub-comunidades en el futuro. Siguiendo con la estructura de DSpace, las colecciones serán las investigaciones de cada integrante del proyecto y los items los archivos que ellos deban ir subiendo.

La estructura de esta guía es parecida a las dos anteriores, primero se parte con una introducción donde se presenta el documento y su incidencia en el centro, luego se muestra el objetivo de este, el cual es ayudar a la comprensión y utilización del repositorio, para finalmente describir cada una de las funcionalidades, las cuales son: creación de cuentas y grupos, creación de comunidades, sub-comunidades, colecciones e items (junto a la explicación de cada una de ellas en el contexto del centro), la edición de cada item (eliminar, autorizaciones de acceso, agregar nuevos archivos, agregar o editar metadatos) y finalmente agregar o editar los formatos conocidos por el centro y los metadatos del repositorio.

El texto del documento es el siguiente:

Introducción

“En el presente documento se describirán algunas de las funcionalidades permitidas en el repositorio del centro de investigación. Estas funcionalidades son las consideradas como mínimas para el cumplimiento de las políticas de gestión de datos científicos del centro.

Para la realización de ciertos procedimientos, es necesario utilizar las guías de licencias y metadatos del centro.

Se presentará en detalle el paso a paso para cada funcionalidad junto a imágenes sacadas del repositorio del centro para servir como guía para las personas que utilizarán por primera vez el sistema.”

Objetivo

“El objetivo de este documento es facilitar la comprensión del funcionamiento del repositorio, junto con ser una ayuda para los integrantes del Web Intelligence Centre a la hora de realizar las

distintas acciones necesarias para el cumplimiento de las Políticas de Gestión de Datos del centro."

Contenido

“A continuación, se presentarán los procedimientos a seguir para las actividades claves para el uso del repositorio del centro:

Donde un ejemplo de cómo están explicados los procedimientos se puede observar en la figura 5.4."

The screenshot shows the 'Describe the Collection' form in the repository interface. The form is titled 'Describe the Collection' with a question mark icon. Below the title, it says 'Please check the boxes next to the statements that apply to the collection.' There are seven checkboxes with corresponding statements:

- New Items should be publicly readable
- Some users will be able to submit to this collection
- The submission workflow will include an accept/reject step
- The submission workflow will include an accept/reject/edit metadata step
- The submission workflow will include an edit metadata step
- This collection will have delegated collection administrators
- New submissions will have some metadata already filled out with defaults

At the bottom right of the form, there is a blue button labeled 'Next >'.

Donde:

- 1. New items should be publicly readable:** Si se selecciona, los archivos subidos serán públicos para cualquier persona externa. En caso de haberla seleccionado, en cada colección se pueden editar estos permisos.
- 2. Some users will be able to submit to this collection:** Si se selecciona, en los siguientes pasos de la creación de la colección, saldrá la opción de agregar personas (con cuenta) o grupos de personas para que tengan permiso de agregar archivos a la colección.
- 3. The submission workflow will include an accept/reject step:** Si se selecciona, en los siguientes pasos se deberá seleccionar a la persona o grupo encargado de aceptar/rechazar un nuevo *submit*.
- 4. The submission workflow will include an accept/reject/edit metadata step:** Si se selecciona, en los siguientes pasos se deberá seleccionar a las personas o grupos encargados de aceptar/rechazar/editar los metadatos agregados por el integrante al momento de subir un ítem.

Figura 5.4: Ejemplo de contenido de Guía para el uso del Repositorio

Fuente: Elaboración Propia

Finalmente, al igual que el documento de las políticas en el encabezado se muestran los logos del centro y de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Por otro lado, se agrega una tabla de contenidos o índice donde los números de página irán en el pie de estas. La versión que es la 1.0 irá en la portada de la guía. La guía se puede ver en el Anexo D.3

5.2. Documentos y templates

A continuación se detallan los objetivos y contenidos de los distintos documentos y templates necesarios para hacer efectiva la gestión de datos de investigación en el centro.

5.2.1. Compromiso de cumplimiento de políticas

Este documento es el que será firmado por todos los integrantes del centro y aprobados por el encargado del WIC. Su estructura es parecida al de las políticas respecto al encabezado y pie de página, al igual que todos los documentos generados en este trabajo.

Se debe ser claro sobre qué versión de las políticas se están firmando, por lo que debe declararse en el documento, la versión entregada en este trabajo es la 1.0.

El principal fin de este documento es que cada integrante declare haber leído y entendido las políticas de gestión diseñadas, para poder cumplir los requerimientos insertados en ellas. Finalmente se agrega la fecha en que se firmó el acuerdo en caso de ser necesario en el futuro.

El texto del documento es el siguiente:

“Yo, _____ (RUT: _____), perteneciente al proyecto _____ del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, declaro haber leído y aceptado las Políticas de Gestión de Datos Científicos del centro en su versión ___.

Con esto, me comprometo a cumplir las reglas estipuladas en las Políticas del centro mientras sea parte de él, siguiendo los protocolos y teniendo claro cuáles son mis responsabilidades y cuáles son las de los demás integrantes del centro. ”

Al final del documento se encuentran los espacios para las firmas del director del centro y el integrante, además de la fecha en que se realizó este compromiso.

La versión 1.0 de este documento se encuentra en el Anexo D.4

5.2.2. Permiso de Acceso a Datos del centro

En algunos casos, algún externo puede requerir el acceso a datos generados en el centro por alguna investigación de alguno de sus proyectos. En otros casos, los datos pueden estar fuertemente restringidos legalmente incluso para integrantes del centro. En ambos casos se deberá firmar un documento el cual deberá ser aprobado y firmado por los jefes de proyectos a los cual pertenecen los datos generados. Para esto se genera un documento que contiene dos partes: primero se explican las condiciones para solicitar el acceso y segundo la solicitud que deberá firmarse.

En caso de que la persona solicitante no cumpla algún requisito, el centro tendrá el poder de

eliminar cualquier tipo de acceso a estos datos, eliminando su cuenta del repositorio o quitándole los accesos de esta cuenta. Para dar accesos se puede revisar la guía de uso del repositorio.

El texto de condiciones y solicitud será el siguiente:

Condiciones

“Con el fin de respaldar la seguridad e integridad de algunos datos científicos generados en el centro, el Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas debe aprobar cualquier tipo de acceso a algunos datos sensibles y puede guardarse el derecho de revocar este permiso si lo considera necesario.

Cualquier persona externa o interna al centro que desee acceder a los datos deberá firmar una solicitud de acceso, la cual se encuentra en la segunda hoja de este documento, comprometiéndose a cumplir todos los requisitos impuestos por el centro, expresados en sus Políticas de Gestión de Datos Científicos y seguir sin excepción los pasos incluidos en las guías de metadatos, licencias y uso del repositorio del centro.

En caso de ser aceptada, la persona podrá optar al acceso a los datos guardados en el repositorio de datos científicos del centro, en donde se le creará una cuenta y darán los permisos necesarios para esto.

En caso de que no se cumpla algunos de los requerimientos durante el período que cuente con acceso a los datos, este permiso será eliminado. El centro se guarda el derecho a aceptar o eliminar permisos cuando lo desee o considere necesario.

En la siguiente hoja se encuentra el permiso que deberán firmar los solicitantes de acceso."

Permiso de Acceso

“Yo, _____ (RUT: _____), solicito acceso a los datos generados en la investigación _____ del proyecto _____ del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Para esto me comprometo a cumplir todos los requerimientos y protocolos impuestos por el centro en relación a los datos científicos y su manejo. Conociendo cuáles son mis responsabilidades y consecuencias en caso de no cumplirlas."

Al final de este documento hay un espacio para las firmas y la fecha en que se realizará esta solicitud.

La versión de este documento es la 1.0, cuenta con los logos del centro y la facultad al igual que los documentos y guías anteriormente diseñadas. El documento se encuentra en el Anexo D.5

5.2.3. Reporte por Proyecto

Este reporte debe realizarlo el jefe de proyecto vigente una vez al año, desde el momento en que se implemente la arquitectura completa en el centro.

El objetivo de este reporte es documentar todos los problemas que se tuvieron a la hora de cumplir las políticas por parte de cada proyecto, esto para que cuando llegue el momento de la revisión periódica de las políticas se tenga en cuenta qué puntos fueron los más difíciles de cumplir. Además de ser un método evaluador de la arquitectura propuesta en caso de ser implementada.

Con estos documentos se podrá observar si el cumplimiento de las políticas varía de manera significativa entre cada proyecto del centro, para lo cual se puede analizar las causas de esto, por ejemplo si es referente al tipo de proyecto o trabajos que se realizan en ellos, ya que algunos pueden generar más datos que otros, o simplemente casos personales.

Para esto se tomará la lista de todos los integrantes que participaron en el proyecto, que firmaron también el compromiso de cumplimiento de las políticas, donde se analizará el cumplimiento del documento de manera individual. Se deberá especificar el trabajo que realizaron, que para el caso del centro en general son los trabajos de título de los estudiantes que integran el centro.

Para el análisis individual, existen distintas actividades que pueden servir de guía a la hora de analizar de manera numérica el cumplimiento de las políticas por parte de los integrantes de los proyectos. Estas acciones que involucran en algunos casos más de algún punto escrito en el documento son las siguientes: la digitalización de todos los documentos generados para la realización de experimentos con humanos (en caso de que existan estos experimentos), la entrega de los planes de manejo de datos (en sus dos versiones), la correcta creación de metadatos (en el cual quedan respaldados varios tópicos de las políticas, como los tiempos de privilegio, los derechos de los datos, licencias, formatos, entre otros) y el almacenamiento de los datos generados en el repositorio propuesto. Cada una de estas acciones tiene un % de cumplimiento total.

Para el caso de la digitalización, se tomará la cantidad total de documentos que se crearon o firmaron (como consentimientos informados y planes de manejo de datos) y sacar la proporción de documentos digitalizados y almacenados por el integrante. Es decir:

$C_d = DG/T_d$, Con C_d = Cumplimiento Digitalizaciones, DG = Documentos digitalizados por el integrante, T_d = Total de documentos que debía digitalizar.

Respecto al plan de manejo de datos, existen dos versiones, cada una representa el 50% de cumplimiento de este punto. En cada una de estas dos versiones se tienen requerimientos sobre su contenido, en caso de haber entregado todos en cada uno se tendrá el 100% de cumplimiento, en caso contrario se obtendrá el proporcional entre contenidos entregados y el total de estos. Numéricamente:

$C_p = 50\% * RC_1/TR_1 + 50\% * RC_2/TR_2$, Con C_p = Cumplimiento Plan de Manejo de Datos, RC_1 y RC_2 el número de requerimientos cumplidos para las versiones 1 y 2 respectivamente, TR_1 y TR_2 el total de requerimientos para las versiones 1 y 2 respectivamente.

Para la creación de los metadatos, se tienen un mínimo de propiedades que deben ser especifi-

cados en ellos. Además, hay dos partes en donde se deben crear estos metadatos, en el repositorio y en el archivo XML que se debe adjuntar al archivo subido. Por esto cada una de estas dos partes tendrá un 50 % del cumplimiento de este punto. En caso de que el archivo XML no se encuentre en el repositorio se considerará como totalmente no cumplido el 50 % de esta parte.

Para calcular el porcentaje de cumplimiento de esta acción simplemente se tomará la proporción entre el total de propiedades entregados y los demandados. Las cuales actualmente son Creator, Title, PublicationYear Publisher, Rights, FunderReference, ResourceType, Identifier y Citation (es decir 9 para el caso del repositorio y 8 para el archivo XML ya que la citación en caso de ser tomado por DataCite se hará automática). Recordar que actualmente no se encuentra la asignación de DOIs implementada, por lo que esta propiedad se puede sacar del reporte del primer año ocurrido luego de la implementación final de la arquitectura. Esto se traduce a:

$C_m = 50\% * RM_1 / TM_1 + 50\% * RM_2 / TM_2$, Con C_m = Cumplimiento Metadatos, RM_1 y RM_2 el número de requerimientos de metadatos cumplidos para las partes 1 y 2 de asignación de metadatos respectivamente, TM_1 y TM_2 el total de requerimientos de metadatos para las partes 1 y 2 de sus creaciones.

Finalmente, para el caso de los datos generados, se tendrá un total de conjunto de datos que serán subidos por el alumno al repositorio del centro y al igual que el caso anterior se tomará la proporción entre los subidos y los generados por el alumno. Se debe analizar los datos generados y los criterios de las políticas sobre si se deberían haber subido o no. Con esto el valor del cumplimiento sería el siguiente:

$C_r = DSC / DR$, Con C_r = Cumplimiento del uso del repositorio, DSC = Datos subidos correctamente, es decir que cumplan con criterio estipulado, DR = Datos que deberían haber sido agregados al servidor de acuerdo a los mismos criterios.

En cada una de las acciones, en caso de no ser de cumplimiento total se deberán agregar los puntos que no se realizaron correctamente como observación, para mantener este registro.

Con estos cuatro porcentajes de cumplimientos de cada acción se podrá calcular el desempeño promedio de cada uno de los integrantes del centro. Teniendo esto se podrá calcular el promedio final de cada proyecto, con el fin de ir viendo la evolución de estos a lo largo de los años.

$CIP_i = 25\% * C_{di} + 25\% * C_{pi} + 25\% * C_{mi} + 25\% * C_{ri}$, con CIP_i el cumplimiento del integrante número i del proyecto, y C_{di} , C_{pi} , C_{mi} y C_{ri} , los cumplimientos del integrante i en cada uno de los puntos mencionados anteriormente.

Finalmente se saca el promedio total del porcentaje de cumplimiento obtenido en cada proyecto. El formato del template se puede ver en las figuras 5.5 y 5.6. Al igual que todas las guías y documentos generados se agregan en el encabezado el nombre del documento y los logos del centro y de la facultad. La versión 1.0 del documento se encuentra en el Anexo D.6

Reporte final Proyecto X

Jefe de Proyecto	
Mail de contacto	
Fecha	
Financiamiento	

Integrantes

Integrante 1

Nombre:	
Mail de contacto:	
Trabajo:	

	Cumplimiento (%)	Observaciones
Digitalización documentos		
Entrega PGDs		
Metadatos		
Uso repositorio		
TOTAL		

Integrante 2

Nombre:	
----------------	--

Figura 5.5: Template reporte de proyecto 1

Total

El cumplimiento promedio del proyecto X es el siguiente:

	Cumplimiento (%)
Digitalización documentos	
Entrega PGDs	
Metadatos	
Uso repositorio	
TOTAL	

Observaciones:

Figura 5.6: Template reporte de proyecto 2

5.2.4. Plan de Manejo de Datos

Cada integrante del centro que realice investigación dentro de él deberá entregar dos versiones del documento, tal como se especifica en las políticas diseñadas, y deben ser aprobado por el jefe de proyecto del centro. Como se busca que todos los integrantes realicen estos planes de gestión de datos, se entrega un *template* de las dos versiones, con el fin de estandarizar estos documentos en todo el centro. Algunas secciones del documento se pueden observar en las figuras 5.7, 5.8 y 5.9

Como se ha mencionado durante el informe, es primera vez que se realizará una gestión de datos en el centro, incluyendo el plan de gestión de datos (DMP), del cual no existía conocimiento por parte de los integrantes antes de iniciar este trabajo. Por esta razón se entrega un ejemplo de plan de gestión de datos para ambas versiones utilizando un trabajo de título realizado en el centro en años anteriores, al cual no se le realizó una gestión de datos de investigación al momento de realizarlo. El trabajo escogido es el de Gino Slanzi, el cual realizó su trabajo durante el año 2014 y ya no es parte del centro de investigación. Aparte de sacar información, se cambiaron algunos nombres e inventaron algunos datos con el fin de mostrar lo más completo posible el ejemplo, ya que algunos puntos no necesariamente estaban escritos en el informe de memoria del cual se basó el ejemplo de DMP mostrando que con el tiempo se va perdiendo información valiosa para posteriores usos de los datos obtenidos en algunos trabajos.



Plan de Gestión de Datos



Plan de Gestión de Datos

Nombre del proyecto de investigación

Autor _____
Fecha _____
Versión _____
URI _____
IPR _____

Figura 5.7: Portada Plan de Gestión de Datos v1

Datos Administrativos:

NOMBRE INVESTIGADOR	
TRABAJO A REALIZAR	
PROYECTO WIC	
CÓDIGO DEL PROYECTO	
FINANCIAMIENTO	
ENCARGADO DE PROYECTO	
MAIL DE CONTACTO	
TELÉFONO DE CONTACTO	

Requerimientos a cumplir

Descripción del Proyecto

Outputs

Se espera obtener los siguientes outputs:

#	Output	Tipo	Formato / Duración / Tamaño
1			
2			
3			
4			

Figura 5.8: Contenido Template Plan de Gestión de Datos v1

Requerimientos a cumplir

Descripción del trabajo

Outputs

Con el experimento se obtuvieron los siguientes OUTPUTS:

#	Output	Tipo	Formato / Duración / Tamaño
1			
2			
3			

Metodologías y Algoritmos utilizados

Datos a Almacenar

Metadatos

Documentos

Propiedad

Figura 5.9: Contenido Template Plan de Gestión de Datos v2

Finalmente, al igual que todos los documentos diseñados, se agregan los logos y número de versión al encabezado y pie de página de los templates. Las versiones 1.0 de los templates la

primera y segunda versión del plan de gestión de datos se encuentran en los Anexos D.7 y D.8 junto a los ejemplos creados que se encuentran en D.9 y D.10 respectivamente.

5.3. Procesos

En base a todo lo diseñado, algunos requerimientos se traducen en procesos que deben ser realizados en ciertos órdenes establecidos. Por esta razón se diseñan los procesos utilizando la notación BPMN, la cual es entendida por la mayoría de los integrantes del WIC, con el propósito de estandarizar el cómo se realizan estos procesos relacionados a la gestión de datos dentro del centro. Para esto se utiliza el software Bizagi Modeler en su versión 3.

A continuación se describen los procesos:

5.3.1. Proceso general por integrante

Durante la estadía de cada integrante en el centro, en caso de generar datos en alguna investigación deberá cumplir todos los requerimientos de las políticas del centro. En cada una de estas investigaciones que genere datos científicos deberá realizar el siguiente proceso:

Primero, el integrante debe haber firmado el documento de compromiso de cumplimiento de las políticas del centro, el cual debe ser aprobado. Una vez hecho esto se realizarán los experimentos para los cuales antes de ejecutarlos debe entregar un plan de gestión de datos (en su primera versión) el cual debe ser aprobado por el jefe de proyecto. Una vez terminada la investigación deberá entregar la segunda versión del plan de gestión de datos que también debe ser aprobada para recién ahí subir los datos al repositorio de manera definitiva, esta última tarea se encuentra definida en el siguiente punto del capítulo. El proceso descrito se puede observar en la figura 5.10

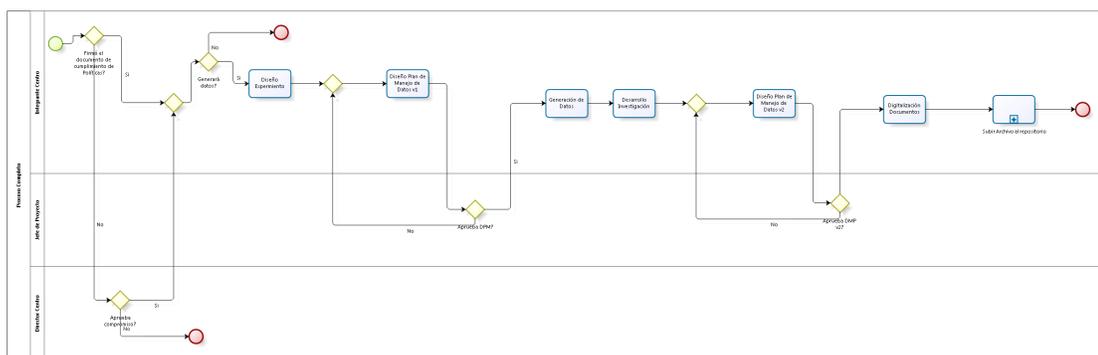
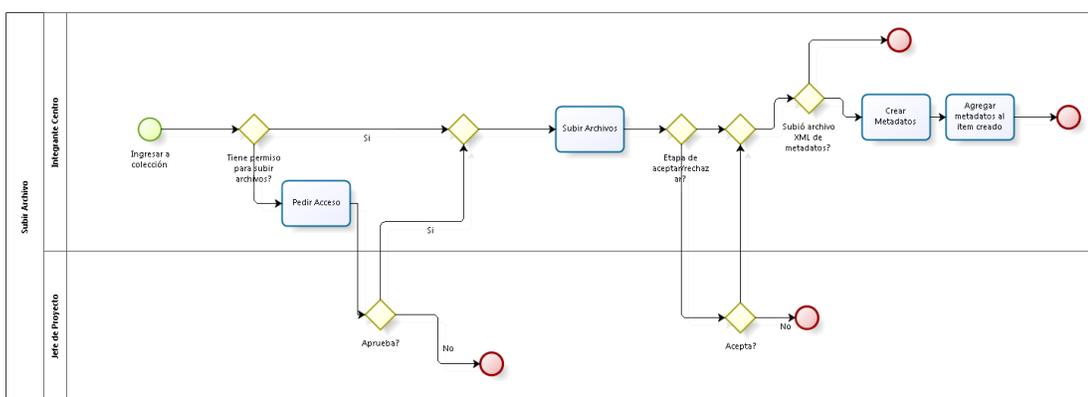


Figura 5.10: Proceso: Proceso general por integrante

5.3.2. Subir Archivos al repositorio

A pesar de que esta acción está definida en la guía para el uso del repositorio, se necesita ser claro en los pasos que deben realizar antes las personas que deseen realizarla puesto que es una de las más importantes dentro de la gestión de los datos en el centro y debe ser hecha de manera correcta sin excepciones con tal de asegurar la preservación de los datos generados.

El primer paso es revisar si se tiene acceso a la colección en donde desea subir los nuevos datos, en caso de no tenerlo debe solicitar este acceso para después realizar los pasos explicados en la guía del repositorio en la sección "Crear Item". Una vez hecho esto debe subir el archivo XML en donde se encuentran los metadatos en el esquema de DataCite en caso de no haberlo hecho al momento de crear el item. Este proceso queda descrito en la figura 5.11



Powered by
bizagi
Modeler

Figura 5.11: Proceso: Subir Archivos al repositorio

5.3.3. Permiso de Acceso a Datos

Para los casos en que se necesite un acceso permitido por parte del centro a ciertos datos, lo primero es que el solicitante del permiso requiera el documento al centro, a través de un email de contacto o de manera presencial. Una vez recibido este documento deberá firmarlo y entregárselo al Jefe del Proyecto en donde fueron generados estos datos para ser aprobados o rechazados, donde se les responderá la decisión final al solicitante.

En caso de ser aprobada la solicitud se procederá a la creación de una cuenta en el repositorio del centro y se le asignarán los permisos para que pueda ver los ítems a los que requirió permiso. Cuando finalice su investigación para la cual necesitaba los datos se procederá a eliminar su cuenta o retirar los accesos al ítem en cuestión, esta decisión dependerá de la preferencia del centro. Este proceso se describe en la figura 5.12

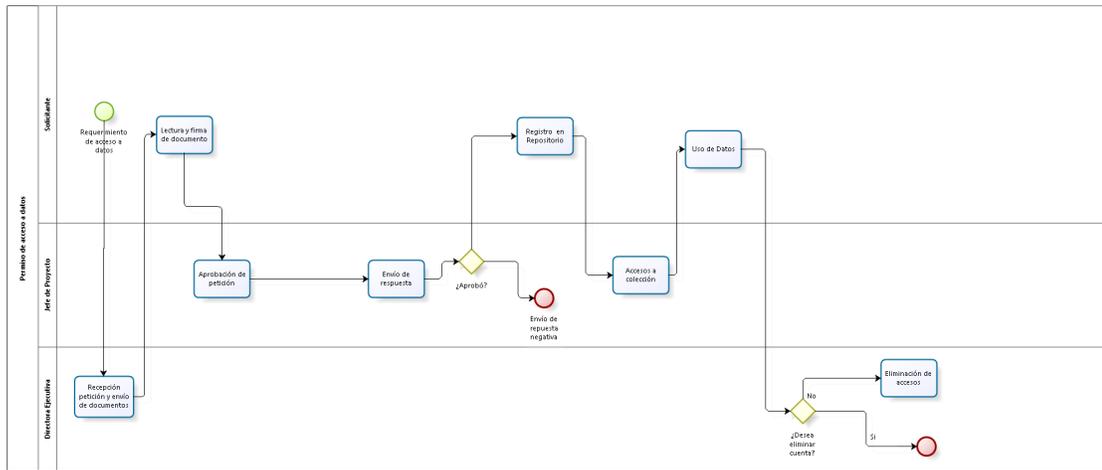


Figura 5.12: Proceso: Permiso de Acceso a Datos

Estos procesos diseñados con notación BPMN son entregados al centro para que puedan utilizarlos en caso de querer implementar la propuesta. En el siguiente capítulo se desarrolla una validación de la arquitectura enfocada en los últimos dos pilares de esta diseñadas, con miras a generar mejoras y propuestas de implementación en el centro.

Capítulo 6

Validación Arquitectura

Para analizar la factibilidad de implementación de la arquitectura de gestión de datos científicos diseñada en los capítulos anteriores se realiza primero un test de usabilidad de evaluación al repositorio, en donde se analiza que tan fácil es utilizarlo por parte de los integrantes del WIC, para este test, tendrán que utilizar también las guías de licencias, metadatos y repositorio, con el fin de obtener información tanto de la infraestructura TI como del soporte generado. Con esto se busca mejorar tanto el soporte como la infraestructura diseñada con el fin de realizar un test de usabilidad de validación con los cambios generados a través de los primeros resultados.

En segundo lugar, se realiza una evaluación económica de la arquitectura.

6.1. Evaluación de Usabilidad

Para realizar este test de usabilidad primero se crea un plan de prueba, con algunos de los puntos mencionados en el capítulo 2.

6.1.1. Elaboración

Para idear un plan de la prueba, se deben analizar los siguientes puntos:

Propósito, metas y objetivos del test

El objetivo del test de usabilidad es primero probar la usabilidad del repositorio, analizando si los participantes pueden o no utilizarlo de manera efectiva. Por otro lado, se va a analizar el cumplimiento de las guías diseñadas a la hora de utilizar el sistema.

Para esto se medirá la facilidad de aprendizaje del repositorio, se analizará que tan eficiente es, se contarán los errores cometidos (junto a qué tan graves son y qué tan fácil son de corregirlos) y

qué tan agradable es el sitio del repositorio para cada uno de los participantes.

Preguntas de investigación

Las preguntas que se quieren contestar con este test son las siguientes:

1. ¿Qué tan fácil es registrarse?
2. ¿Qué tan fácil es subir un archivo al repositorio utilizando las guías diseñadas?
3. ¿Qué tan fácil es agregar o editar un metadato en un ítem ya creado?
4. ¿Qué tan fácil es agregar un archivo a un ítem ya creado?
5. ¿Está satisfecho con el diseño del repositorio?
6. ¿Está satisfecho con las guías diseñadas?
7. ¿Las guías son suficientes para realizar las acciones necesarias?
8. ¿Es muy largo el test?

Para el caso de los jefes de proyectos, se les preguntará además:

1. ¿Qué tan fácil es realizar una tarea de aceptar/rechazar un ítem subido?
2. ¿Qué tan fácil es crear una colección en el repositorio?

Para el caso de la directora ejecutiva, además de todas las preguntas anteriores, se preguntará:

1. ¿Qué tan fácil es crear o editar una cuenta?
2. ¿Qué tan fácil es crear o editar un grupo?
3. ¿Qué tan fácil es crear una comunidad?
4. ¿Qué tan fácil es agregar un nuevo formato conocido al repositorio?
5. ¿Qué tan fácil es agregar o editar un metadato del registro de metadatos en el repositorio?

Características de los participantes

Se buscará realizar el test a 5 integrantes del centro que sean memoristas, jefes de proyecto o la directora ejecutiva del centro, los cuales son suficientes para realizar este tipo test [70].

Diseño del test

A cada participante, se le pidió que revisara con anterioridad las guías diseñadas. Se espera que dure entre 60 y 90 minutos cada test.

Preguntas previas al test (2 minutos): Se consultarán datos del participante relacionados al centro, específicamente a qué proyecto pertenece y cuál es su rol dentro de éste. Además, se preguntará si efectivamente realizaron o no la revisión de las guías. Estas preguntas se encuentran en el Anexo E.1.1

Introducción (2-4 minutos) Se explicará la importancia de su colaboración en el test de cada participante además de introducir al experimento que se va a realizar, para esto se leerá el guión de orientación, el cual se encuentra en el Anexo E.1.2

Realización de tareas (60-90 minutos): Los participantes realizarán 2 veces cada tarea asignada, menos la tarea de registrarse, la cual se realiza sólo una vez. Se irá describiendo y realizando cada tarea por separado.

Entrevista post tareas (5 minutos): Se realizará una entrevista una vez finalizadas las tareas por segunda vez. Las preguntas a realizar se encuentran en el Anexo E.1.3

Listado de Tareas

Las tareas a realizar son las siguientes:

Para todos los participantes:

1. Registrarse e iniciar sesión en el repositorio. Se considera exitosa esta tarea al momento de ingresar al perfil del usuario, la cual es la página a la que redirige el login del repositorio. (2 minutos)
2. Crear un ítem en la colección "Colección Test" de la comunidad "Proyecto Test" que fueron creadas para el test, siguiendo todos los protocolos de las guías entregadas. Esta tarea tiene dos métricas de éxito, una es que efectivamente se cree un ítem y la segunda es que estén todos los metadatos requeridos y licencias asignadas. Para esta prueba se les dará un contexto hipotético de un *set* de datos generado, con el objetivo que rellenen los metadatos y asignen licencias respecto a ese escenario, esto para simular que el participante generó los datos con anterioridad y conoce el contexto de estos. Por temas de tiempo, se eliminará temporalmente la restricción de subir el archivo XML con el esquema de DataCite. (15 minutos)
3. Agregar un archivo a un archivo al ítem creado. Esta acción se considera completada si al ingresar al ítem aparece el nuevo archivo subido. (5 minutos)

4. Agregar un metadato en el ítem creado (con cualquier valor). Esta tarea se considera completada si al revisar todos los metadatos del ítem, aparece el nuevo metadato creado con el valor asignado. (5 minutos)
5. Editar un metadato en el ítem creado (con cualquier valor). Esta tarea se considera completada si al revisar todos los metadatos del ítem, aparece el metadato con el nuevo valor. (5 minutos)

Contexto primera realización : Suponga usted que generó un set de datos en un experimento, de los cuales tiene completo derecho a hacer lo que se le dé la gana y desea subirlos al repositorio, estos datos fueron generados bajo el contexto del proyecto "Test de usabilidad del repositorio", el cual es financiado por CORFO a través de un INNOVA código TFTFTF324234. Por otro lado, fueron tomados en el costanera center el mes pasado. Los datos obtenidos se encuentran en español y son los datos de comienzos de atención de un local de comida rápida. El archivo es horaempie-zaatencion.xlsx (asumir que es un archivo correcto)

Contexto segunda realización : Suponga usted que generó un set de datos que desea subir al repositorio, estos datos fueron generados bajo el contexto del proyecto "Test de usabilidad del repositorio", el cual es financiado por CONICYT a través de un FONDEF código ABCDEFG. Estos datos fueron obtenidos a través de la realización de un experimento utilizando eye-trackers y fueron tomados en el centro de investigación. El archivo se llama test1.json (asumir que es un archivo correcto)

Para los jefes de proyecto:

1. Crear una colección en una comunidad asignada. Esta tarea se considera exitosa al momento de ingresar a la nueva colección creada. (5 minutos)
2. Realizar la tarea de aceptar o rechazar un ítem creado por algún integrante. Esta tarea se considera completada cuando al momento de aceptar un ítem creado, éste aparece en la colección a donde pertenece, o cuando se seleccione rechazar. (5 minutos). Para este punto se subirán dos archivos a la colección "test" de la comunidad "Proyecto Test aceptadas para el test en el repositorio.

Primer archivo: Será un archivo xlsx junto al archivo XML con las propiedades de DataCite, con los metadatos ingresados al repositorio correctamente junto a la licencia bien asignada. Se espera que sea aceptado.

Segundo archivo: Será el mismo archivo con todos los metadatos menos Creator y Publication-Year. Se espera que sea rechazada.

Para la directora ejecutiva:

1. Crear una cuenta de usuario. Esta tarea se considera completada cuando el nuevo usuario aparece en el listado de EPersons del repositorio. (3 minutos)

2. Crear un grupo. Esta tarea se considera exitosa cuando el nuevo grupo aparece en el listado. (3 minutos)
3. Editar un grupo. Esta tarea se considera exitosa cuando se agregue o elimine un usuario o grupo al grupo en cuestión (3 minutos).
4. Crear una comunidad. Esta tarea se considera exitosa cuando esta nueva comunidad sale en el listado de comunidades. (4 minutos)
5. Agregar un nuevo formato al registro de formatos. Esta tarea se considera exitosa cuando al ingresar se muestre el nuevo formato con todos los datos pedidos correctos. (5 minutos)
6. Agregar un elemento a un esquema de metadatos cualquiera. Esta tarea se completa cuando al ingresa al esquema salga el nuevo elemento creado. (3 minutos)
7. Editar un elemento en un esquema de metadatos cualquiera. Esta tarea se completa cuando al ingresa al esquema salga el elemento editado con el nuevo valor asignado. (3 minutos)

Ambiente de la prueba, equipamiento y logística

La prueba se realizará utilizando un computador personal en donde montó de manera local el repositorio diseñado en el capítulo 4.

Además se entregarán las guías de licencias, metadatos y uso del repositorio junto a los contextos para los datos que se deben subir en una de las tareas de manera impresa.

Las respuestas serán tomadas en otro computador.

Datos a coleccionar y medidas de evaluación

Para responder las preguntas de investigación se tomarán los siguientes datos:

1. Número de errores cometidos la primera y la segunda vez que realizan cada tarea.
2. Tiempos de realización de cada tarea.
3. Facilidad de uso general del repositorio.
4. Utilidad de las guías entregadas.
5. Impresiones sobre el repositorio y el test.

Moderadores

El test se realizará utilizando un único moderador, que a la vez será también el observador. Éste será el encargado de presentar el experimento, indicar las instrucciones, observar el experimento y anotar las observaciones y respuestas a las preguntas realizadas.

6.1.2. Resultados

Se realizó el test a un integrante del proyecto DOCODE, uno de SAIDEMA y PSYODEMA (es parte de los dos) y a tres de OPINIONZOOM, entre los cuales había dos memoristas, dos jefes de proyecto y un desarrollador. En esta oportunidad no se realizaron pruebas para las tareas de administrador debido a problemas de tiempo, por lo que se realiza en el test de validación posterior. De los participantes sólo uno había subido alguna vez datos de investigación a un repositorio especializado y ninguno revisó de manera completa las guías entregadas.

Los resultados obtenidos están relacionados a los 5 componentes que definen a la usabilidad: facilidad de aprender, eficiencia, capacidad de recordar, cantidad de errores y satisfacción con el diseño del usuario.

Facilidad de Aprender

Para analizar este componente, se utiliza la información del tiempo en que se demoraron en realizar cada una de las tareas asignadas y la primera pregunta de la entrevista post-pruebas. Para esto se pidió que realicen cada actividad dos veces, con el fin de comparar estos tiempos. En la Figura 6.1 se pueden observar los promedios tiempos en segundos de cada tarea, donde las primeras 5 tareas son las que todos realizaron y las tareas 6 y 7 fueron para los participantes que eran jefes de proyectos.

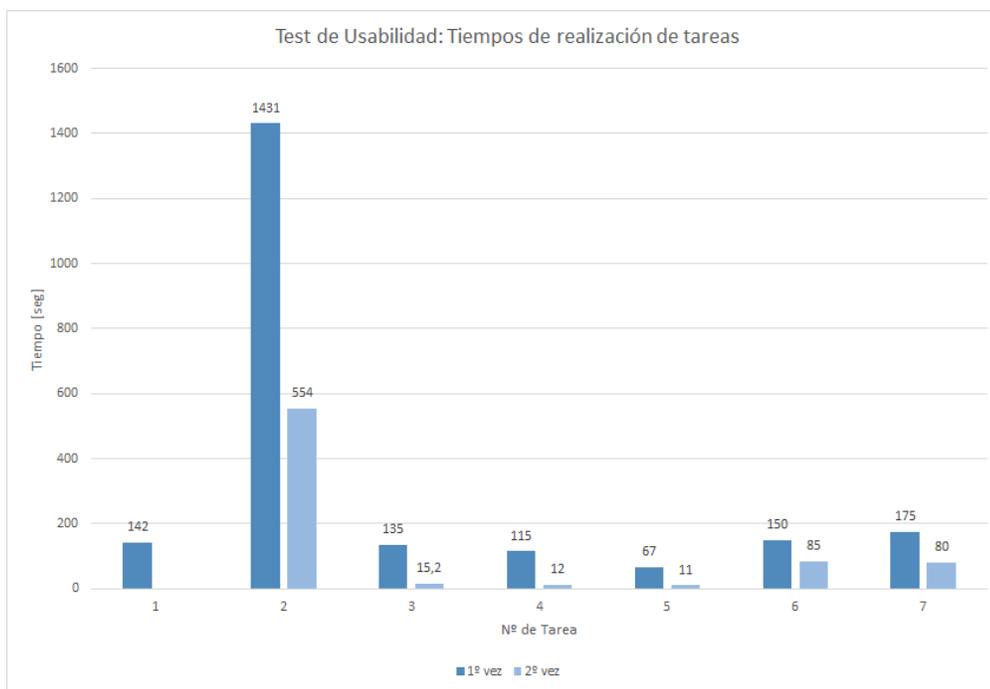


Figura 6.1: Evaluación: Tiempos por tarea

Viendo las diferencias promedio entre los tiempos, se puede apreciar que en todas las tareas el tiempo de realización de estas disminuyó considerablemente, en algunos casos casi 9 veces. Por otro lado, en la tarea número 2 que consistía en crear un ítem en el repositorio y era la actividad más larga, el tiempo utilizado disminuyó a la mitad. Con esto dicho, se puede observar que hubo un aprendizaje del cómo utilizar el sistema por parte de los usuarios.

Respecto a la primera pregunta de la entrevista que se realizó después del test, en donde se consultó la percepción de la dificultad de las tareas. La tabla 6.1 muestra el porcentaje de la cantidad de participantes que seleccionaron cada nivel de dificultad por tarea.

Nº Tarea	Muy fácil	Fácil	Medio	Difícil	Muy difícil
1	60 %	40 %	0 %	0 %	0 %
2	0 %	40 %	40 %	20 %	0 %
3	40 %	60 %	0 %	0 %	0 %
4	60 %	40 %	0 %	%	%
5	80 %	20 %	0 %	0 %	0 %
6	100 %	0 %	0 %	0 %	0 %
7	50 %	50 %	0 %	0 %	0 %

Tabla 6.1: Evaluación: Dificultad de tareas asignadas

Donde se puede apreciar que sin contar la tarea de crear un ítem, las demás tareas se consideran como fáciles o muy fáciles para los usuarios. Respecto a crear un ítem, al momento de contestar esta pregunta se reiteró el comentario de que se considera intermedio o difícil debido a los requerimientos de las guías más que el software en sí, otros mencionaron que era fácil, solo había que

aprender bien los requerimientos de las guías antes.

Eficiencia

La eficiencia en el contexto de un test de usabilidad tiene relación al qué tan rápido los participantes pueden realizar ciertas tareas una vez aprendido el diseño. Utilizando en análisis anterior, se concluye que el uso del repositorio con las guías es eficiente, ya que los tiempos utilizados en cada tarea en cada iteración disminuyen, en algunos casos drásticamente.

Errores

Respecto a la cantidad de errores cometidas por los usuarios, en la primera tarea sólo hubo un error que fue un intento de registro a través de ingresar a “Olvidé contraseña”, el cual no es un error grave y se puede solucionar fácilmente. Por otro lado este paso no estaba incluido en la guía del repositorio.

En la segunda tarea, todos los usuarios pudieron subir en términos de utilización del software de manera correcta. Por otro lado, respecto a los metadatos sí existieron algunos errores cometidos por los usuarios, principalmente de definiciones estos fueron:

1. Dos participantes pusieron el nombre completo (en el formato correcto Apellido, Nombre) en el campo de Last Name en vez de hacerlo por separado, uno de estos dos participantes en la segunda iteración siguió cometiendo el mismo error, el otro se dio cuenta al leer el *placeholder* del *input*.
2. Un participante no agregó el año de publicación en la citación en ambas oportunidades.
3. Un participante en la tercera etapa relleno mal el campo de licencia en ambas iteraciones, donde sólo puso una versión resumida del nombre de la licencia, por ejemplo en vez de Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0) utilizó CC BY-NC-SA.
4. Un participante puso en el campo de citación al publicador como creador. En la segunda iteración se dio cuenta del error y lo hizo correctamente.
5. Un participante confundió en la primera iteración ResourceType con formato al momento de rellenar la citación, pero se dio cuenta de su error en la tercera etapa donde ResourceType estaba definido y debía ser rellenado.

Existieron 5 errores distintos cometidos en esta tarea sienta, en la primera iteración se realizaron en total 6 errores y en la segunda 3.

De estos, el primer error no se encuentra grave debido a que en términos técnicos el nombre quedará registrado correctamente, a pesar de esto se podría entregar en la guía un ejemplo con una imagen de esta etapa en donde quedaría visualizado esto para solucionar este error.

El segundo error se considera de gravedad media, ya que en general en todos los formatos de citación se requiere el año de publicación. A pesar de esto, en ningún lugar existe un ejemplo de cómo debería rellenarse este campo, por otro lado, rellenó bien el campo del año, por lo que el error no fue por no saber qué debía poner o qué se entendía por año de publicación.

El tercer error tampoco se considera grave debido a que es el nombre abreviado del que se debía poner, además puso bien la URL a la cual dirige la licencia, por otro lado, el participante no leyó la guía en donde salía el cómo debía rellenarse este campo.

El cuarto error podría considerarse como más grave que los anteriores porque si alguien utiliza esa citación le estaría quitando crédito al autor de los datos, a pesar de esto el participante se dio cuenta del error. Por otro lado, no existe un lugar donde salga un ejemplo completo de este formato de citación, por lo que se considerará agregarlo.

El quinto error no se considera grave debido a que se dio cuenta en el momento del error, por otro lado, se considera en el futuro dar la definición como *hint* en la primera etapa sobre el qué significa ResourceType para que esto no ocurra nuevamente.

En las tareas número 3, 4, 5 y 6 no se cometieron errores, exceptuando en una oportunidad en la tarea 4 en la primera iteración, donde se debía agregar un metadato y selección el botón *update* en vez del botón que se utiliza para agregar, este hecho no es considerado como grave principalmente debido a que en el momento en que no se agregó el metadato, el participante revisó la guía y encontró su error, haciéndolo bien en el segundo intento.

Para la séptima tarea, se cometió el mismo error en ambos casos, que fue que no realizaron la evaluación (correcta o incorrecta) del ítem subido, este error es grave debido a que es uno de sus deberes como jefes de proyectos y ellos son encargados de analizar si los datos han sido subidos correctamente a la comunidad de su proyecto. Por otro lado, terminado el test se les mencionó este hecho y ambos respondieron

Facilidad de memorizar

Por problemas de tiempo, no se realizará una prueba para analizar este componente. A pesar de esto se puede estimar pensando en dos componentes anteriores en donde los tiempos y cantidad de errores se redujeron a más de la mitad en algunos casos, por lo que se estima que la facilidad de memorizar es medianamente alta.

Satisfacción

Para saber la satisfacción de la interfaz y guías entregadas se consultó una vez terminadas las tareas la percepción del diseño del repositorio y de las guías.

Respecto al repositorio la mayoría se mostró satisfecho, sólo un integrante encontró que la usabilidad de la página no era tan buena, a pesar de esto el 100% de los participantes consideró que permite realizar todo lo que se necesita hacer para los datos por lo que la consideró como

útil. En cuanto al diseño, sólo un participante mencionó que en algunos lados estaba demasiado sobrecargado de información, que se podría eliminar o poner en otro lado.

Por otro lado, al consultar por las guías, sí estaban satisfechos respecto a si estaba toda la información necesaria para realizar las tareas, pero respecto al cómo se presentaban algunos mostraron quejas, principalmente respecto a la guía de metadatos, donde la molestia principal era que en el repositorio las cosas aparecían con un nombre y en la guía con otro, que es algo que se puede cambiar para mejorarla, además mencionaron que podría haber un ejemplo con imágenes como en la guía del repositorio para explicar las primeras dos etapas en donde se deben rellenar los metadatos. Finalmente consideraban un poco innecesario rellenar algunas cosas dos veces.

Para las otras dos estaban satisfechos, sólo que habían algunas tareas que no salían en la guía del repositorio, por lo que se podría considerar agregarlas.

Con estos resultados por cada componente se considera que el repositorio con las guías cumplen las expectativas de facilidad de uso. A pesar de esto hay campos que se pueden mejorar para que se cometan menos errores, sobre todo en las guías y en el formulario para crear un ítem.

6.1.3. Resultados para las preguntas de investigación del test

A continuación se responden de manera más específica las preguntas de investigación, las cuales eran las que se querían responder con el test realizado.

¿Qué tan fácil es registrarse?

El 100% consideró esta tarea como fácil o muy fácil, sólo se recibió una recomendación de agregar un botón de registro al lado de Log In en la página.

¿Qué tan fácil es subir un archivo al repositorio utilizando las guías diseñadas?

En promedio la dificultad fue considerada como media. Esto era esperable debido a la cantidad de requerimientos especialmente sobre los metadatos, además en algunas ocasiones se consideraba tedioso o innecesario rellenar algunos dos veces. Por otro lado, subir el archivo hablando del repositorio, sin requerimientos, fue considerada como fácil, lo que la hacía más difícil eran los metadatos que había que rellenar.

¿Qué tan fácil es agregar o editar un metadato en un ítem ya creado?

El 100% de los participantes consideró que era fácil o muy fácil agregar o editar un metadato.

¿Qué tan fácil es agregar un archivo a un ítem ya creado?

El 100 % de los participantes consideró que era fácil o muy fácil agregar o editar un metadato.

¿Las guías son suficientes para realizar las acciones necesarias?

El 80 % considera que sí son necesarias, el otro participante cree que está toda la información pero no mostrada de la manera correcta por lo que en términos prácticos no lo era.

¿Es muy largo el test?

A pesar de que en general los tests duraron en promedio casi 1 hora, no se encontró largo, principalmente debido a que por no leer antes las guías algunos tuvieron que realizarlas en el momento, por lo que consideraron que el test en sí no lo era.

¿Qué tan fácil es realizar una tarea de aceptar/rechazar un ítem subido?

El 100 % de los jefes de proyecto consideró que la tarea era muy fácil de realizar

¿Qué tan fácil es crear una colección en el repositorio?

El 100 % de los jefes de proyecto consideró que la tarea era muy fácil de realizar

¿Está satisfecho con el diseño del repositorio?

Sólo un participante indicó quejas sobre la usabilidad, los demás se encontraron satisfechos y encontraban que era bastante útil para el centro, también se recibieron comentarios sobre el botón "Submit to this collection", el cual podría tener un mensaje más corto.

¿Está satisfecho con las guías diseñadas?

No se encontraron totalmente satisfechos con la guía de metadatos, respecto a las otras dos el 100 % indicó que encontraban que estaban bien. Sobre la guía de metadatos se considera que está toda la información pero su formato no es el correcto. Faltan imágenes y las propiedades están con distintos nombres y orden de los que salen en la página. Esto se considerará cambiar para la validación.

Sobre el propósito del test

Respecto al objetivo principal del test, el cual era probar la usabilidad del repositorio, se considera como cumplido y teniendo un resultado positivo, a pesar de esto se cree necesario realizar mejoras con el fin de validar de manera final tanto el repositorio como las guías. En la siguiente sección se mencionarán estos cambios y mejoras que se desarrollarán.

6.1.4. Cambios y Mejoras

En base a los resultados y comentarios recibidos luego de la ejecución del test de usabilidad de evaluación del repositorio y guías se realizan los siguientes cambios y mejoras en cada una de ellas:

6.1.5. Repositorio

De acuerdo a los comentarios obtenidos mientras se realizaban las pruebas de usabilidad, algunos puntos fueron escuchados más de una vez por parte de los participantes, por esto se decide realizar los siguientes cambios:

1. Se agrega el botón Sign Up, el cual fue una recomendación obtenida del test de evaluación.
2. Se elimina de las páginas distintas al Inicio el contenido donde sale el logo del centro que se puede ver en la figura 4.2. Esto debido a que de acuerdo a los participantes sobrecargaba demasiado las páginas de las colecciones que de por sí ya muestran harta información.
3. En el formulario para ingresar un ítem se realizan varios cambios:

Se cambian los *labels* de cada campo por los nombres de las propiedades en DataCite, durante los tests más de una vez los participantes buscaron el campo del repositorio en la guía de metadatos y no podían encontrarlas porque tenían distintos nombres.

Se mueven Resource Type y Resource Type General a la primera etapa, principalmente porque al momento de agregar la citación se requiere Resource Type y no estaba la definición debido a que se encontraba en la tercera etapa. Además se deja por defecto Dataset y no Audiovisual para Resource Type General.

Se agrega un ejemplo de citación para que sirva de guía. Fue uno de los campos en donde se cometieron errores, pero esto se puede deber a que no existían ejemplos ni en la guía ni en el repositorio.

Se cambia la palabra “fundador” por “financiador.”^{en} el campo awardNumber.

Se elimina el campo Sponsors debido a que se repetía la información en la siguiente etapa con funderReference. Se dejan como campos obligatorios todos los elementos de la primera etapa. Este fue un comentario recibido en más de una ocasión, que algunos campos la guía decía que eran obligatorios pero en el repositorio no lo eran.

Se eliminan algunos campos que eran opcionales y no tenían una propiedad homóloga en el esquema de DataCite.

6.1.6. Guía de Licencias

A pesar de tener una buena recepción se encontraron algunos detalles simples de corregir:

1. Se arregla un error ortográfico en donde en vez de “Rights.” estaba escrito “Righths”
2. Se cambian de lugar los dos últimos párrafos de la guía ya que la gente en general los pasaba por alto por estar al final, se colocaron antes de la sección Condiciones.
3. Se destacan las frases “Para asignar una licencia de Creative Commons se deben realizar dos pasos:”, “El primer paso”, “El segundo paso.” al comienzo de la sección Contenidos.
4. Como hubo un buen recibiendo de poner imágenes del proceso en las guías, se decide agregar dos imágenes que muestran un ejemplo de las etapas 3 y 6 de crear un ítem en el repositorio. Además de agregar que se debe escribir el nombre completo y no la abreviación de la licencia, además se arregla la redacción de los dos primeros párrafos de la sección Contenido para ser más explícito en el qué se debe hacer. Las imágenes agregadas se pueden observar en las figuras 6.2 y 6.3.



Describe Describe Describe Upload Verify License License Complete

Submit: Describe this Item ?

Please fill further information about this submission below.

Debe escribir el nombre completo. Ejemplo: Creative Commons Attribution 4.0 International License (CC BY 4.0). Para el centro se utilizarán solo licencias de Creative Commons y copyright, para más información de estas revisar la Guía de Licencias del centro.

Rights

El URI de la licencia, <https://creativecommons.org/licenses/by/4.0/>

rightsUri

Figura 6.2: Guía de Licencias: Tercera Etapa



Describe Describe Describe Upload Verify License License Complete

Submit: Use a Creative Commons License

If you wish, you may add a Creative Commons License to your item. Creative Commons licenses govern what people who read your work may then do with it.

License Type

< Previous Cancel/Save Next >

Figura 6.3: Guía de Licencias: Sexta Etapa

6.1.7. Guía para el uso del Repositorio

A pesar de tener una alta aceptación por los participantes, existían algunas tareas que no salían en este documento, por lo que se agregan junto al arreglo de redacción de algunas tareas:

1. Se agrega la guía de registro en la página que no estaba en la versión anterior con el nombre de Creación de Cuentas.
2. Se cambia nombre de "Cuentas de Usuario.^a "Cuentas de usuario desde el admin" para no confundir el registro con las funciones de administrador.
3. En la tarea de "Crear Item", se deja en claro que las etapas 1, 2 y 3 están especificadas en la guía de metadatos, igual que para las otras etapas se agregarán imágenes, que fue algo comentado como positivo por los participantes.
4. Se agrega la guía para las tareas de revisión de ítem, para el caso de los jefes de proyectos. No existían estos puntos en la guía, en el test ambos participantes no realizaron bien esta tarea. Igual que en los casos anteriores se agregan imágenes del procedimiento.

6.1.8. Guía de Metadatos

Esta guía fue la que recibió más críticas negativas debido a la cantidad de información que tiene. Por esto se realizan los siguientes cambios pensando en el test de validación que se realiza después.

1. Se agrega un ejemplo de citación, que fue uno de los puntos en donde se cometieron errores a la hora de crear ítems y que no estaba en ningún lado.
2. Se cambia de lugar el primer párrafo de la sección Propiedades debido a que se malinterpretó la frase "página 27 en adelante" que se refería a la documentación del esquema y no de la guía. Se ubicó antes del párrafo previo a las definiciones de cada propiedad.
3. Al igual que en la guía de licencias se destacan las frases "La primera parte", "La segunda parte" "Se deberán agregar metadatos en dos partes:." al inicio de la sección Propiedades.
4. Se corrige un error ortográfico donde decía "Tittle" por "Title".
5. Se agregan imágenes de las primeras etapas en la sección propiedades, especificando la obligatoriedad de rellenar algunos campos. Además de cambiar la redacción de los párrafos que mencionaban a estas etapas debido a que se cambiaron los órdenes de los campos en el repositorio.

Estos documentos son entregados previamente a los integrantes del centro para que puedan revisarla antes de realizar el test, con el fin de disminuir los tiempos de realización de las tareas. Con estos cambios ya realizados se realiza la validación de usabilidad del repositorio y las guías. El cual es detallado en la siguiente sección.

6.2. Validación de Usabilidad

Para realizar el test de validación se realiza la misma prueba de evaluación hecha anteriormente pero con distintos integrantes del centro, integrando esta vez las tareas relacionadas a la administración que en el test anterior no se analizaron. La única diferencia es que las guías entregadas a los participantes y el repositorio a utilizar son las versiones mejoradas gracias a los resultados de la primera prueba.

6.2.1. Resultados

En este caso se realizó el test a tres integrantes del centro, donde uno realizó todas las tareas incluidas las relacionadas a la administración, el cual es jefe del proyecto AKORI y además en algunas ocasiones realizará estas acciones de administrador. Otro participante integrante del proyecto SONAMA sólo realizó las tareas que todos debían hacer y el tercer *tester*, el cual realizó la prueba la vez anterior pero en esta ocasión ejecutó las labores de administrador. Por otro lado, en todos los casos la revisión de la guía no fue completa o no se realizó.

Se revisan las cinco componentes del atributo usabilidad al igual que en el test pasado, pero comparando los resultados obtenidos con los anteriores.

Facilidad de Aprender

Los promedios de tiempo entre la primera y segunda etapa son los siguientes:

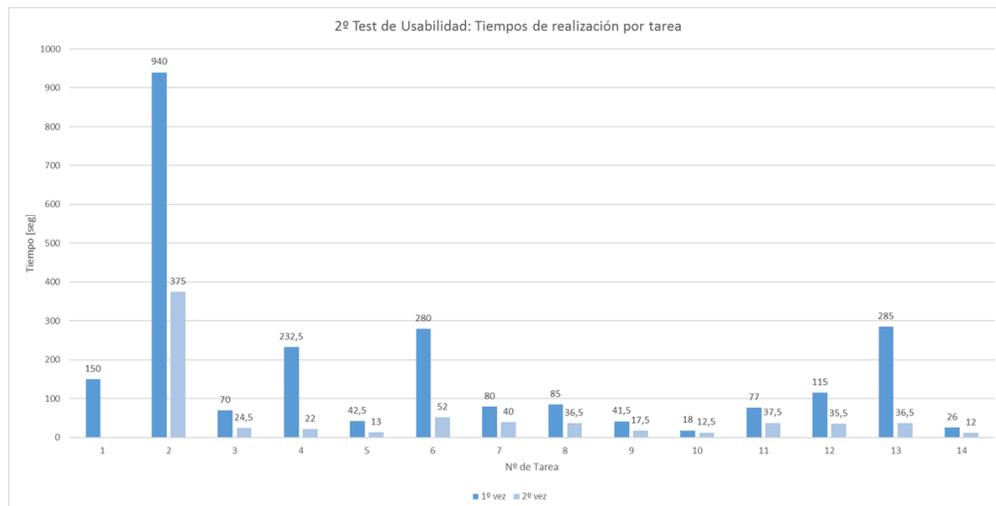


Figura 6.4: Validación: Tiempos por tarea

Donde de la tarea 1 a la 5 son las relacionadas a los integrantes de proyectos, 6 y 7 al jefe de proyecto y las siguientes al administrado.

El nivel de dificultad que percibieron los participantes por tarea es la siguiente:

Nº Tarea	Muy fácil	Fácil	Medio	Difícil	Muy difícil
1	100 %	0 %	0 %	0 %	0 %
2	50 %	0 %	0 %	50 %	0 %
3	100 %	0 %	0 %	0 %	0 %
4	0 %	50 %	50 %	%	%
5	100 %	0 %	0 %	0 %	0 %
8	50 %	50 %	0 %	0 %	0 %
9	50 %	0 %	50 %	0 %	0 %
10	50 %	50 %	0 %	0 %	0 %
11	50 %	0 %	50 %	0 %	0 %
12	0 %	100 %	0 %	0 %	0 %
13	0 %	0 %	50 %	50 %	0 %
14	50 %	50 %	0 %	0 %	0 %

Tabla 6.2: Validación: Dificultad de tareas asignadas

Y donde el jefe de proyecto declaró que encontró las tareas 6 y 7 (relacionadas a su cargo) de nivel de dificultad muy fácil y fácil respectivamente.

Como se puede observar en las figuras 6.1 y 6.4 el tiempo promedio decrece considerablemente en casi todas las primeras actividades.

Para la primera actividad el promedio se mantuvo, lo que hace inferir que es el tiempo que se necesita para el registro del repositorio, por otro lado, esta actividad se considera como una actividad de nivel fácil o muy fácil

En la segunda tarea, que era la más complicada, los tiempos promedios bajaron considerablemente respecto al test de evaluación, esto se cree es producto de los cambios realizados en las guías y en el repositorio. Por otro lado, la diferencia entre la primera y segunda instancia entre ambos test es muy parecida, llegando a ser menos de la mitad para la segunda instancia. La percepción entre los dos participantes fue bastante distinta, uno lo consideró muy fácil de realizar, mientras el otro difícil.

Las tareas número 3, 4 y 5 se consideraron en general como de dificultad muy baja exceptuando por el de crear un metadato, por otro lado, sus tiempos disminuyeron a menos de la mitad.

Para el caso del jefe de proyecto y las tareas 6 y 7, se observa que al igual que en la evaluación, los tiempos bajaron a la mitad, además de considerarlas como actividades de fáciles de realizar.

Para las tareas de administración se observa que en todos los casos el tiempo de realización de las tareas disminuyó a más de la mitad en muchos casos. El caso particular de la tarea número 13, en donde el tiempo bajó casi 7 veces, se debió a que a los participantes les costó encontrar el lugar para ingresar al esquema de metadatos, lo que influyó en la respuesta de considerarla de una dificultad media o difícil.

Con estos resultados se concluye que respecto al repositorio y las guías hubo un aprendizaje por parte de los participantes que se traduce en las bajas considerables de tiempo utilizado en ejecutar las tareas, además considerando las percepciones en general se consideraron de una dificultad baja, permitiendo un aprendizaje más rápido.

Eficiencia

Los tiempos entre las iteraciones bajaron considerablemente, en algunos casos drásticamente igual que en el test de evaluación, por lo que se considera que el uso del repositorio y las guías es eficiente.

Errores

Los errores cometidos en este test fueron los siguientes:

1. Un participante confundió publicador con creador en la citación en la primera iteración, en la segunda no cometió el error.
2. Los dos participantes no agregaron en ambas oportunidades el ResourceType en el campo de citación.
3. Un participante en una ocasión se equivocó de licencia y en vez de Copyright seleccionó una de Creative Commons.

El primer error se repitió respecto al primer test. Se considera como grave debido a que si alguien utiliza esta citación se le estaría quitando crédito al autor de los datos, a pesar de esto al igual que en el caso anterior el participante se dio cuenta de su error.

El segundo error no se considera como grave debido a que respecto al formato que entrega DataCite este parámetro es opcional. En esta instancia se agregó un ejemplo de citación en donde no se encontraba tampoco el ResourceType, por lo que se atribuye el error a que siguieron el ejemplo en vez del formato, se considerará agregar esta propiedad al ejemplo en el futuro.

El tercer error se considera como grave a pesar de que no tiene implicancias debido a que estos datos no serán compartidos, a pesar de esto no se siguieron los contenidos de la guía. Por otro lado, se observó que el participante al momento de revisar la guía de licencia en el test, pasaba por alto y creía que la última licencia era la penúltima de la guía sin observarla.

Facilidad de memorizar

Al igual que en el caso anterior, por problemas de tiempo no se realizará una prueba para analizar este punto. Siguiendo el procedimiento para el test de evaluación, observando los dos componentes anteriores, los tiempos y errores se redujeron respecto una iteración a otra, por lo que se estima que la facilidad de memorizar es medianamente alta.

Satisfacción

Al analizar las respuestas de la entrevista post-test, se recibieron comentarios positivos acerca del repositorio y las guías, donde no se contabilizaron esta vez malos comentarios sobre las guías. Respecto al repositorio, existieron algunas quejas sobre ciertos puntos: al momento de realizar ciertas acciones al inicio de la página aparecen mensajes que parecen alertas, lo que hace parecer que algo se hizo mal, lo cual al leerlas se observa que es sólo un mensaje de información (excepto cuando realmente informa sobre errores realizados). Por otro lado, el nombre de algunos botones se considera cambiar, como Add Bitstream para añadir archivos por ejemplo.

A pesar de estos comentarios, se considera por parte de los participantes como una aplicación que cumple todas las necesidades que se buscan y en general es bastante intuitiva encontrándose satisfechos con esta, excepto para algunas funcionalidades de administrador, en donde hay que conocer un poco su estructura para saber utilizarla.

Con estos resultados, se considera que el repositorio con las guías al igual que para el primer test cumplen los estándares buscados, aunque existen posibilidades de mejoras para hacer más cómoda su navegación aún más. Por otro lado, los tiempos promedios en general bajaron, exceptuando la tarea de agregar un metadato en donde en primera instancia el promedio de realización fue casi el doble que en el primer test, un punto a destacar es la segunda asignación en donde el promedio entre las dos iteraciones bajó cerca de 5 minutos, avalando los cambios realizados para mejorar la arquitectura.

6.2.2. Preguntas de Investigación

Respecto a las preguntas de investigación, en comparación con el test, las respuestas que cambiaron fueron las siguientes:

¿Qué tan fácil es agregar o editar un metadato en un ítem ya creado?

En esta ocasión un participante consideró esta tarea como regular, principalmente se debió a que al revisar la lista de todos los metadatos disponibles consideró que era necesario saber muchas definiciones para cumplir esta tarea.

¿Está satisfecho con las guías diseñadas?

En esta oportunidad sólo se recibió una queja relacionada a los metadatos que aparecen a la hora de querer agregar uno nuevo a un ítem ya creado. Esto se debió a que las propiedades que se encuentran en ese lugar son del esquema de Dublin Core y no DataCite como se explicó anteriormente, por lo que no se encontraban en la guía. Aparte de esto los usuarios se encontraron satisfechos con las guías en general y consideraron que servían y eran necesarias para poder realizar las tareas requeridas.

Por otro lado, en el test anterior no se realizaron tareas de administración, a pesar de que sólo dos personas realizaron estas pruebas, la cantidad de personas que las ejecutarían serían 3 personas, es decir se abarcó a más de la mitad de los posibles usuarios. A continuación se contestan las preguntas relacionadas a ellas:

¿Qué tan fácil es crear o editar una cuenta?

Los dos participantes consideraron que era una tarea fácil o muy fácil

¿Qué tan fácil es crear o editar un grupo?

Los dos participantes consideraron que era una tarea fácil o muy fácil

¿Qué tan fácil es crear una comunidad?

Un participante consideró que era de dificultad fácil y otro mediana.

¿Qué tan fácil es agregar un nuevo formato conocido al repositorio?

Los dos participantes consideraron que era una tarea fácil o muy fácil

¿Qué tan fácil es agregar o editar un metadato del registro de metadatos en el repositorio?

Los dos participantes consideraron que era una tarea fácil o muy fácil editar un elemento de algún esquema de los metadatos, pero de dificultad moderada para el caso de agregar una nueva.

Sobre el propósito del test

Respecto al test anterior, los promedios de tiempos bajaron considerablemente en general validando los cambios realizados en primera instancia. Se considera bajo los resultados en los componentes del atributo usabilidad que el repositorio y las guías son positivos en cuanto a usabilidad de estos, además de considerarse útil por parte de los participantes de ambos tests. Por otro lado, a raíz de los tres test realizados salieron a relucir posibles cambios y mejoras a realizar con el fin de cometer menos errores y hacer más fácil de utilizar la arquitectura.

6.2.3. Cambios y Propuestas de Mejora

Cambios Realizados

Se realizan algunos cambios menores en base a los resultados obtenidos en el test. Por otro lado, algunos cambios que necesitan de más recursos, como tiempo y conocimientos técnicos se dejan propuestos como mejoras a realizar en caso de que se implemente finalmente la arquitectura.

Se agrega al ejemplo de citación el Resource Type en el repositorio y en la guía de metadatos, el cual se encontraba en el formato solicitado, pero no en el ejemplo, al observar los errores se intuye que los participantes se guiaron por el ejemplo priorizándolo al formato.

En el repositorio, en la etapa tres de crear un ítem, se agregan las opciones de licencias bajo un menú desplegable con el fin de hacer más rápido la elección y asegurar que no existan errores de tipografía o abreviaciones.

En el formulario de creación de colecciones y comunidades se presentan definiciones en español en los mensajes de ayuda de cada campo. En algunos casos hubo confusión sobre el significado de estos.

Se agrega un ejemplo de Publisher en la primera etapa de crear ítem.

Se cambian los botones para añadir metadatos y archivos a ítems creados. En varias ocasiones se observó que algunos participantes perdían tiempo buscando estos botones, además Bitstream causaba confusión en algunos casos por no saber su significado.

A pesar de que no fue algo cuestionado durante los tests, respecto a la propiedad funderIdentifier, del cual existen distintos tipos de identificadores para los financiadores, se acota a usar sólo el tipo "Crossref Funder", principalmente debido a que las posibles opciones para asignación de DOIs será Crossref, además la universidad los obtiene a través de esta organización. Entonces por ejemplo para algún proyecto financiado por CONICYT, funderIdentifier sería: <http://dx.doi.org/10.13039/501100002848> y la propiedad funderType debería ser rellenado con "Crossref Funder".

Finalmente, para la guía de licencias se coloca una imagen en donde se explica la definición y el cómo debe rellenar los campos de los metadatos. Esto principalmente debido a que, al momento de revisar la guía, todas las otras licencias contaba con una imagen, entonces al momento de revisar de manera rápida la guía y al pasar por Copyright, como no existía una imagen parecía que era la continuación de la anterior.

Propuestas de Mejora

Algunas posibles mejoras encontradas necesitan de más tiempo y conocimiento técnico para ser realizadas, por lo que se dejan como propuestas de mejora en caso de que se implemente finalmente la arquitectura, estas son:

En algunos casos, al seleccionar el botón "atrás" del navegador aparecía un error. Es un problema

recurrente en muchas páginas web, para solucionarlo se necesita un análisis más profundo del código de Dspace.

El funcionamiento de las alertas y mensajes de información en ciertas ocasiones generaba confusión, haciendo pensar que se había cometido un error. Se propone revisar el sistema de mensajes en los flujos para que esto no suceda.

Una queja que se repitió fue que se debía seleccionar la licencia dos veces, lo que podía generar errores de tipeo o selección de la licencia en la sexta etapa de crear un ítem. Se propone realizar los cambios necesarios para que sólo se tenga que agregar en una de las dos ocasiones (en la tercera o en la sexta etapa del proceso).

Para que no existan errores de citación, se podría generar un *javascript* que autorellene ese campo con lo que se vaya relleno en los campos anteriores (Creator, Publicationyear, Title, etc.)

A la hora de agregar metadatos a ítems creados, se puede seleccionar de una larga lista de propiedades, principalmente del esquema de Dublin Core de Dspace, el problema de esto es que no necesariamente tienen el mismo nombre que las propiedades de DataCite, además de que al momento de crear ítems no se muestran todas estas, por lo que se podría considerar agregarlos a la guía de metadatos para su mejor entendimiento.

También se podría considerar realizar el test nuevamente con los integrantes que ya realizaron el test con el fin de analizar la componente de Facilidad de memorizar del atributo de usabilidad, donde se pueden generar más propuestas de mejoras en base a este.

Un punto que no quedó tan claro es sobre la creación de un archivo XML para los metadatos del esquema de DataCite de cada ítem creado, el cual debe hacerse en paralelo al momento de crear los ítems en el repositorio, para el caso de los tests por temas de tiempo no se requirió que se haga pero se mencionó en algunos casos. Existen dos opciones que pueden ser evaluadas para simplificar este proceso. El primero es la utilización de un generador de archivos xml con el esquema de DataCite, actualmente existe uno, pero no ha sido actualizado a la versión 4.0 de DataCite [105]. Para esto se puede esperar que se actualice o hacerlo internamente en el centro. La segunda opción se debe realizar al momento de configurar la asignación de DOIs en Dspace [103] cuando se seleccione la opción a utilizar para esto. Dspace cuenta con un plug in conversor de metadatos (de Dublin Core a DataCite) dentro de su configuración, la cual viene descargada por defecto al momento de instalar Dspace en un servidor, el problema que tiene es que utiliza la versión 2.2 de DataCite, para que funcione de mejor manera para el caso de este trabajo se podría editar este archivo (DIM2DataCite.xml), el cual se encuentra en la carpeta dspace/config/crosswalks y configurarlo para ser utilizado con la versión 4.0 de DataCite. Esta última opción es la más recomendada ya que ahorraría tiempo a los integrantes del centro a la hora de agregar archivos al repositorio (ya que no tendrían que crear el archivo XML y rellenar todos los datos dos veces, lo que puede resultar tedioso al momento de realizarlo), por otro lado la primera opción puede ser utilizada hasta antes del momento de agregar la configuración de la asignación de DOIs del repositorio.

Finalmente, DSpace cuenta con muchas funcionalidades no mencionadas o configuradas en este trabajo, las cuales son explicadas en su documentación [96], como el servidor de handles, los cuales son un tipo de identificador más simple que los DOIs pero que se podría agregar como mínimo en caso de no contar con el DOI en cierto momento. Por esto se recomienda revisar completamente

estas funcionalidades y analizar si pueden agregar valor o facilitar aún más el cumplimiento de todas las políticas en el centro de investigación.

6.3. Arquitectura de Gestión de Datos Científicos en el WIC

Realizar una gestión de datos de investigación en cualquier lugar tiene ciertos requisitos y protocolos que se deben cumplir con el fin de que sea efectiva. Estos requerimientos son descritos en el diseño de las políticas del centro, creando reglas dentro de ellas con el fin de que todas estén cubiertas en el documento generado sin excepción.

Respecto a la arquitectura tecnológica para el almacenamiento futuro de los datos, se necesitó primero realizar un listado de todos los requisitos técnicos que debía poder realizar en orden de cumplir las reglas de las políticas, es decir, permitir una gestión de datos efectiva. En cuanto al software, a pesar de que cumple los requerimientos técnicos, como cualquier software debe ser considerado como “usable” por parte de los futuros usuarios, sino aunque sea práctico en cuanto a funcionalidades, no será utilizado. Gracias a la realización de los tests de usabilidad, en donde los participantes se encontraron satisfechos con este, a pesar de que existen pequeñas mejoras que se pueden realizar sobre su diseño, se concluye que es fácil de aprender a utilizar y es eficiente, por otro lado, para que no existan errores en su uso, los cuales ocurrieron, se proponen ciertas mejoras.

Uno de los factores claves es el soporte a los investigadores para que puedan realizar esta gestión a través del cumplimiento de las políticas y el correcto uso del repositorio. Para analizar esto se realizaron los tests mencionados anteriormente donde los participantes pudieron realizar las tareas asignadas que son parte de la gestión de datos, utilizando estas guías y el repositorio diseñado, encontrándolas útiles y suficientes para poder hacerlas, por otro lado y al igual que para el repositorio, existen oportunidades de mejora con el fin de que las tareas sean realizadas de manera correcta y sin cometer errores. El poder realizar las tareas en el repositorio, que es donde más se traduce la gestión de datos científicos en términos prácticos ya que todas las cosas que se hagan en él deben ser revisadas o aprobadas para asegurar la integridad de los datos, se estarían cumpliendo todos los requisitos de las políticas diseñadas, haciendo posible realizar una gestión de datos.

Con esto se concluye que la arquitectura diseñada permitirá realizar una gestión de datos científicos en el centro de investigación respecto a los términos técnicos y requisitos para poder ejecutarla.

Costos

A pesar de esto, existen ciertos costos extras relacionados a la utilización de la arquitectura en los que tendría que incurrir el centro.

Para efectos de este trabajo, se enfoca siempre en las investigaciones futuras del centro, por lo que se estima el número de sets de datos que se deberán subir al repositorio para estas investigaciones.

Dentro del centro, los proyectos que generan datos primarios que deberían ser subidos al repo-

sitorio son AKORI, OpinionZoom, SONAMA y se espera que también lo haga el proyecto KOKORO. En general entran 5 nuevos memoristas por año actualmente, los cuales pueden ser 1 o 2 para cada proyecto, para este análisis se sobreestimaré el número total y se considerarán dos memoristas por semestre por proyecto ya que dependiendo de cada semestre algunos obtendrán uno, dos o ningún integrante nuevo.

En AKORI, durante las investigaciones realizadas con anterioridad en promedio se realizaba un experimento por cada memorista que entraba. Cada uno de estos generaba una o dos colecciones de datos, de eye-trackers y de encefalogramas. Por otro lado, la cantidad de personas que participaban en estos experimentos variaba entre 20 y 30 personas. Los datos generados por los eye-trackers en promedio pesaban 40 MB cada uno, mientras que los de EEG 350 MB. Esto hace un aproximado de 11 700 MB en espacio requerido considerando 30 participantes, además de los documentos generados como consentimientos informados y para el futuro los planes de gestión de datos, lo que sumaría entre 25 y 35 MB más al total pensando en que cada documento pesa 1 MB en formato PDF, es decir, aproximadamente 11.5 GB por experimento, lo que daría un total de 46 GB de espacio requerido para 4 experimentos anuales realizados.

En cuanto a SONAMA y OpinionZoom, la diferencia con el anterior es que en este caso los datos primarios pesan entre 1 y 10 MB ya que son principalmente archivos de texto. Lo que dejaría 160 MB de espacio total requerido en caso de que cada una pese 10 MB.

Finalmente, para el proyecto KOKORO no se tiene certeza de cuánto pesarán los datos en caso de que se generasen ya que actualmente no se crean datos primarios en este proyecto, además no se sabe que instrumentalización se ocupará si es que se hicieron, por lo que no se puede estimar cuánto espacio se necesitará.

Para utilizar la arquitectura diseñada, el servidor interno donde será alojado el repositorio deberá contar con espacio suficiente para almacenar todos los datos del centro, actualmente el centro cuenta con servidores sin utilización por lo que se aprovecharía uno de estos. Como se esperarían 46.2 GB por año sin contar los generados por KOKORO y en donde en locales comerciales como PC Factory, los precios de discos duros de 1 TB de capacidad para servidores ronda los CLP \$150.000 [106], por lo se necesitaría uno nuevo cada 20 años aproximadamente. Esta cantidad de años puede variar según los tamaños de los datos generados en KOKORO en caso de existir.

Por otro lado, el centro deberá invertir en horas persona con el fin de asegurar que todos los integrantes realicen una gestión de datos. Para los integrantes del centro existe una reunión obligatoria una vez por semana que se podría aprovechar para mostrar la arquitectura y enseñárselas a los nuevos memoristas, la cantidad de sesiones necesarias va a depender del aprendizaje de estos. A pesar de esto, se puede hacer un estimativo de cuánto saldría realizar una sesión de la siguiente manera:

Actualmente se tienen 9 profesionales y 10 memoristas como integrantes del centro, considerando un sueldo aproximado de CLP \$1.100.000 para los profesionales y CLP \$200.000 para los memoristas, que los profesionales trabajan 44 horas a la semana y los memoristas 35, que las reuniones duran 2 horas en promedio y que un mes tiene cuatro semanas, el costo sería de CLP \$141.100 por sesión.

El último de los costos relacionados a la arquitectura propuesta tiene que ver con la posibilidad

de asignar DOIs. Para poder hacer esto existen dos opciones, una es ser miembro de alguna agencia registradora, de las cuales existen 10 en total. Estos son: Airiti Inc, Crossref, China National Knowledge Infrastructure (CNKI), DataCite, Entertainment Identifier Registry (EIDR), The Institute of Scientific and Technical Information of China (ISTIC), Japan Link Center (JaLC), Korea Institute of Science and Technology Information (KISTI), Multilingual European DOI Registration Agency (mEDRA) y Publications Office of the European Union (OP).

De estos Airiti Inc, CNKI e ISTIC están enfocados para China, KISTI para Corea del Sur, JaLC para Japón, EIDR para películas, series de televisión y otros recursos audiovisuales y los últimos dos y OP para Europa.

Dejando a Crossref, DataCite y mEDRA como posibles opciones.

La cantidad estimada de DOIs necesarios por el centro según la cantidad de colecciones que se estima generarán en el futuro es de 2 DOIs por cada experimento de AKORI (uno por los de eye-trackers y otro por los de EEG), 1 por cada investigación realizada por OpinionZoom y SONAMA y se asumirá uno por una hipotética generación de datos primarios en el proyecto KOKORO.

Esto hace que sea necesario por año cerca de 20 DOIs a asignar de manera estimada, considerando 2 memoristas por proyecto al semestre.

Por otro lado, si se desea se puede asignar un DOI a cada set de dato obtenido en cada experimento de AKORI o KOKORO, es decir por cada sujeto participante y por cada instrumento utilizado. La granularidad de esto depende del cómo se quiere que sean citados los datos, lo que en este caso sería por la colección de todas ya que el conjunto sería lo necesario para validar la investigación y no la de un sujeto en particular.

Para ser miembro de DataCite y poder asignar DOIs se debe pagar una cuota fija de €8.500 [107] (utilizando una tasa de conversión de €1 = CLP \$750, que es el promedio del año 2016 [108], serían aproximadamente CLP \$6.375.000), dejando asignar la cantidad de DOIs que se quieran y además pertenecer a la comunidad de DataCite teniendo voto a la hora de tomar decisiones en DataCite.

mEDRA tiene un precio según la cantidad de DOIs que se requieran. De acuerdo a lo estimado para el centro la tarifa sería de €150 (que serían aproximadamente CLP \$112.500) [109]. En caso de necesitar más de 25 DOIs el precio serían €400 (que serían aproximadamente CLP \$300.000).

Finalmente, en Crossref el precio para el caso del centro sería de US \$275 más US \$0.06 por DOI para datos (utilizando una tasa de conversión de USD \$1 = CLP \$658, que es el promedio del año 2016 [110], serían aproximadamente CLP \$181.000 más CLP \$40 por DOI) [111].

La segunda opción es pagar la asignación de DOIs a algún miembro de estas agencias de registro de DOIs. En caso de querer utilizar a DataCite o mEDRA, estos dos no cuentan con miembros en Sudamérica o con soporte a este, por lo que la única posibilidad sería ser miembro directo de estas agencias.

En cuanto a Crossref, esta sí cuenta con distintos miembros en América y en particular tiene en Chile. Uno de estos miembros es la Universidad de Chile, cuenta que es gestionada por la Dirección de Servicios de Información y Bibliotecas de la universidad. Para poder obtener DOIs a través de la universidad se podría negociar con el SISIB y analizar la posibilidad de hacer esto posible sin

incurrir en mayores gastos extra para ellos, el costo seguiría siendo CLP \$40 o lo que se acuerde por DOI requerido, pero la suma será despreciable en comparación a la tarifa de hacerse socio. Por otro lado, esto podría restringir la cantidad de DOIs que el centro pueda asignar y generar problemas en caso de que quisieran más, todo dependería de la negociación entre las dos entidades.

A pesar de esto último y con lo mencionado al inicio de este informe, en donde desde el SISIB se está comenzando a armar un proyecto para generar políticas de gestión de datos en la universidad, la mejor opción sería elegir la alternativa del SISIB principalmente debido a que los DOIs estarían registrados a nombre de la universidad, igual que todas las publicaciones de los integrantes de la institución (actualmente las tesis de los alumnos no cuentan con DOIs asignados, pero las publicaciones de académicos de la universidad sí), los costos serían menores que utilizando Crossref y finalmente, en caso de que las políticas que se generen se hagan efectivas, la transición sería más simple de llevar a cabo debido a que se estaría trabajando desde antes con este sistema, que también puede marcar la pauta de cómo se podría llevar con los distintos centros de investigación de la universidad.

Beneficios

En contraste a esto y como se menciona en el capítulo 1 de este informe, la realización de gestión de datos científicos dentro de una institución trae diversos beneficios, entre los cuales se encuentran aumentar la visibilidad de los datos generados lo que se traduce en aumento de citaciones, permite mayores oportunidades de colaboración con otros investigadores de las disciplinas trabajadas, facilita el reuso de los datos en el futuro tanto interna como externamente, reduce el riesgo de perder datos generados y permite cumplir los requerimientos de los financiadores y universidades respecto a los datos.

Para el caso del centro, utilizar efectivamente la arquitectura de gestión de datos científicos diseñada generará valor de distintas maneras, la primera y relacionada al problema de la descentralización de los datos que se encuentra, donde los datos de cada proyecto están separados entre sí y en algunos casos dentro del mismo proyecto no se sabe dónde se encuentran los datos o si aún siquiera existen. La gestión de los datos sería clave y efectiva para que esto no suceda en el corto y largo plazo, permitiendo reutilizar de mejor manera los datos que fueron financiados por investigaciones pasadas y aprovechar de mejor manera los nuevos fondos ganados para realizar mejoras o nuevas investigaciones.

Por otro lado, uno de los focos del centro es publicar en distintas revistas científicas, las cuales cada vez tienen más requisitos sobre los datos relacionados a las publicaciones los cuales serían cumplidos en caso de implementar la arquitectura.

Como el centro trabaja con fondos públicos concursables, que en algunos casos pueden contar al igual que las revistas científicas con requisitos relacionados a los datos que se generen en los proyectos ganados. Con la arquitectura se permitiría poder cumplir cualquier tipo de requerimiento impuestos por los financiadores y también de la universidad. Por otro lado en Chile pocas instituciones realizan una gestión de datos estructurada como lo haría el centro de investigación, lo que podría considerarse como una ventaja al momento de concursar debido a que se podrían cumplir cualquier tipo de necesidad, además de mostrarse comprometido con la investigación y el fomento

al reuso de datos generados por fondos públicos, lo cual a nivel mundial se está promoviendo por entidades como la OECD hace años.

También dentro de los proyectos que se realizan en el centro, se generan datos que pueden ser considerados como sensibles, ya que pueden repercutir en temas legales en caso de existir algún tipo de problema, la arquitectura diseñada permitiría en primer lugar asegurar que se cumplan todos los protocolos y requerimientos del cómo se deben cuidar y almacenar estos datos, además de los documentos legales que permitieron tomar esos datos, los cuales pueden ser requeridos por los distintos comité de ética de la universidad o del financiador y que en caso de no que se encuentren disponibles puede significar la pérdida de los fondos, lo que podría repercutir negativamente al momento de postular a un nuevo concurso. Esto quiere decir que la arquitectura ayuda a mitigar los riesgos relacionados a la mala gestión de los datos, la cual en este caso puede significar pérdida de prestigio, tiempo y dinero, además de oportunidades para continuar la investigación.

La arquitectura diseñada permite también promover el acceso abierto en caso de que se desee, lo que podría traer más beneficios como el aumento de citaciones a los trabajos realizados ya que posibilita una mayor visibilidad del centro, lo cual aparte de aumentar las probabilidades de colaboración con otras comunidades científicas en el país o en el mundo y aumentar redes de contacto, lo hace mejorando el prestigio de la institución frente a ellos y la universidad.

Finalmente, en el centro se están realizando procesos para estandarizar ciertos procedimientos que realizan los integrantes de este, que van desde plantillas para reportes o presentaciones hasta software a utilizar. Para el caso de los datos no existen estas estandarizaciones por lo que la arquitectura permitiría normalizar esta área que en el centro no ha sido solucionada. Con lo que se concluye que la arquitectura diseñada permitirá realizar una gestión de datos científicos en el Web Intelligence Centre y con esto traer los beneficios nombrados para este, que a pesar de no ser directamente cuantificables en contraste a los costos mencionados, se verán reflejados en el corto y largo plazo tanto en el funcionamiento interno del centro como en la investigación que se realizará en el futuro, los cuales ayudarán a cumplir los objetivos del centro.

6.3.1. Gestión del Cambio

A pesar de que la solución propuesta permitiría realizar una gestión de datos científicos, la cual es necesaria en el centro, para poder implementarla de una manera efectiva, en donde los integrantes realmente realicen una gestión de datos utilizando la arquitectura, es necesario realizar una gestión del cambio en el centro de investigación, la cual se define como el proceso u acción sistemática necesaria para la transformación de la organización desde un estado actual a otro considerado como más deseable [112]. Donde el estado actual es un centro en donde no existen protocolos o procesos relacionados a los datos que generan en su investigación realizadas y la visión futura es un centro en donde exista una correcta gestión de los datos científicos que asegure la integridad y preservación de estos datos de una manera comprometida con la investigación y las buenas prácticas de ella, fomentando la reutilización y el acceso a estos.

En caso de no manejar bien los cambios, se puede utilizar más tiempo y recursos en poder generar efectivamente el cambio, algo que claramente no se busca para el centro de investigación.

Para realizar este cambio en el centro, en donde se pasaría a realizar una gestión de datos por parte de los integrantes, que implica en ciertas ocasiones realizar trabajos que podrían considerarse como extras, innecesarios o hasta a veces aburridos, se necesita primero partir por reconocer que es algo que va a llevar tiempo y no se va a poder realizar de un día para otro, si se desea hacer de la manera correcta.

Lo segundo e igual de importante, es darle a conocer el por qué se está implementando la arquitectura, qué beneficios trae consigo hacer esto, no sólo para el centro los puntos mencionados anteriormente, sino que también personalmente adquirir buenas prácticas de investigación para trabajos futuros, los cuales pueden o no ser en el centro.

Como se menciona anteriormente, existe una instancia obligatoria para los integrantes del centro en donde una vez por semana se reúnen con el fin de ir presentando los avances en sus trabajos e investigaciones de cada integrante y además para dar información que puede influir en todos o en el funcionamiento del centro. Esta instancia debe ser aprovechada para presentar la arquitectura en primera instancia, conversar sobre la gestión de datos científicos, su contexto, qué significa, qué hay que hacer y qué beneficios trae, a modo de inducción para los integrantes.

Uno de los más grandes errores de acuerdo a Kane [112], es el subestimar los tiempos que se necesita para capacitar a la gente a realizar los nuevos procesos, que en este caso no es sólo subir un archivo al repositorio de datos, es estructurar el proceso de los datos desde antes que se tomen hasta que se vuelvan a utilizar. Por eso además de la inducción es necesaria una sesión para mostrar el cómo se hace y en qué orden, siguiendo los procesos presentes en el capítulo 5. Actualmente en el centro se están estandarizando los procesos que se realizan internamente y se está dando foco a los nuevos memoristas que ingresan, para mostrarle estos requisitos que van desde una plantilla para las presentaciones internas a software que utilizar. En esta sesión se debería también agregar la arquitectura.

Es importante señalar que con que exista una sesión, no significa que todos van a haber aprendido todo lo necesario. Comprendiendo que el soporte para realizar una gestión de datos es uno de los pilares fundamentales de la gestión, se deben entregar todas las guías y templates diseñados con el fin de que puedan seguirlas, utilizarlas y aprender de manera más rápida y efectiva, haciendo más fácil esta transición. Por otro lado, el centro debe asegurarse de que todos sigan los pasos exitosamente, por lo que se debería también hacer pequeños tests con el fin de evaluar su desempeño, que en caso de que sea bajo se debe enfocar tiempo en que este sea el máximo posible, ya que es necesario que todos lo hagan bien y no solo algunos. Además, se debe enfatizar en que se están haciendo bien las cosas a cada integrante, ya que no es poco común que se pueda sentir miedo de estar haciendo algo mal y que podrían perjudicar al centro en este caso, generando resistencia al cambio dentro de ellos, perjudicando el proceso en la organización.

Para llevar a cabo estos cambios, un factor crítico es que los jefes de proyectos estén comprometidos con la gestión de datos científicos, ya que ellos serán finalmente los que validen y aprueben los procesos realizados por los integrantes de sus proyectos, deben ser los líderes de este cambio en cada uno de sus grupos junto a la directora ejecutiva, la cual debe velar por que los jefes de proyectos realicen sus labores relacionadas a la gestión de datos en el WIC.

Por otro lado, se considera que para que el repositorio funcione correctamente, se debería asignar a un miembro del centro con experiencia en Java, para que sea el encargado de su mantención y

desarrollador de las mejoras que se quieran realizar en el futuro. También se puede considerar asignar a una persona para los cambios y mejoras que deban realizarse a los documentos como las guías y plantillas entregadas, con el fin de integrar los posibles futuros cambios en toda la arquitectura y no sólo el lugar de almacenamiento.

Finalmente, la gestión de datos científicos es un movimiento a nivel mundial que ha ido creciendo a lo largo de los años, esto implica que también puede ir cambiando en el tiempo encontrando nuevas formas de realizar mejor las cosas. Con esto en mente, en el centro de investigación se debe realizar una revisión continua tanto de las políticas como del repositorio y también de todos los documentos entregados con el fin de que no queden obsoletos en el tiempo, realizando los reportes por proyectos anuales, ver las fallas y solucionarlas para evitar que vuelvan a suceder y asegurándose que todos estos cambios y mejoras que se realicen se traspasen a todos los integrantes del centro.

Capítulo 7

Conclusiones

El presente y último capítulo resume las principales conclusiones obtenidas del trabajo de título realizado. Esto permite también agregar recomendaciones para etapas futuras relacionadas con la implementación y mejora de la propuesta creada.

7.1. Conclusiones Generales

Este trabajo tenía el propósito de diseñar una propuesta de arquitectura de gestión de datos científicos en el Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. El presente capítulo resume las principales conclusiones obtenidas tanto en el estudio realizado para la creación de la arquitectura como en su diseño y validación.

Si bien en la actualidad la gestión de datos científicos ha sido el foco a seguir y fomentar por las grandes entidades relacionados a la investigación a nivel mundial, en Chile se está bastante atrasado y se encuentra lejos de ser un país comprometido con estas buenas prácticas. Implementar la arquitectura diseñada marcaría pautas de cómo se realizan ciertos procesos en la investigación, las cuales podrían ser no sólo en el centro, si no a nivel de facultad o hasta universidad, en donde se está pensando en diseñar políticas de datos para la universidad y esta arquitectura podría servir de apoyo del cómo realizarla. Por parte del centro, en caso de ya contar con una arquitectura armada y funcional, en caso de que se diseñen y sea obligatorio utilizarlas, la transición a estas políticas sería mucho más simple y rápida debido a que los integrantes del centro ya conocerían sobre la gestión de datos, además de ya estar acostumbrados a realizar ciertos procedimientos relacionados a él.

La falta de protocolos y reglas definidas para el cómo deben ser tratados los datos desde antes que se generan hasta que se reutilizan en el futuro, es decir durante todo su ciclo de vida, traen problemas como la pérdida de los datos generados, el no cumplimiento de requerimientos por parte de financiadores, publicadores o los comité de éticas que analizan las investigaciones, la mala toma de datos a la hora de realizar experimentos por no haber sido aprobado su diseño y también disminuir las probabilidades de poder reutilizar los datos o replicar investigaciones pasadas. Al analizar la situación del centro, en donde hasta el día de hoy no existe una disponibilidad completa de los datos generados en el proyecto AKORI, lo que genera problemas para reutilizar los datos, analizar inves-

tigaciones pasadas y gastos asociados al tener que diseñar y realizar todo un experimento de nuevo, se concluye que se necesita una gestión de datos científicos en el centro para evitar este problema en investigaciones futuras y para además aprovechar los beneficios que trae consigo realizarla.

Para realizar una gestión de datos de investigación se necesita de una arquitectura que sustente tanto las reglas y protocolos que se necesitan para hacerla, una arquitectura tecnológica que permita almacenar los datos de manera correcta siguiendo estas reglas y finalmente el soporte necesario para que los investigadores las cumplan y se mejore continuamente.

Respecto a las políticas de gestión de datos científicos, en su diseño se desprenden las siguientes conclusiones: los archivos deben ser respaldados en formatos que sean considerados como aceptables para su preservación a largo plazo, las licencias Creative Commons son las que más se adaptan a la situación, se debe utilizar las licencias menos restrictivas para facilitar el reuso de los datos y el esquema de metadatos de DataCite es el más adecuado para utilizar en la arquitectura. Finalmente se concluye que las políticas cumplen todas las recomendaciones y requisitos que de los grandes impulsores de este movimiento global bajo el contexto del Web Intelligence Centre.

Sobre la arquitectura tecnológica propuesta, primero se concluye que no existen soluciones factibles en los repositorios ya existentes para las políticas diseñadas. Por otro lado, la plataforma DSpace es la mejor solución de software para la construcción de repositorios entre los analizados para el caso del centro de investigación. En cuanto al repositorio construido, se concluye que cumple todas las funcionalidades requeridas por las políticas diseñadas.

En la última etapa, se realizaron pruebas de usabilidad del repositorio y las guías diseñadas como soporte de la arquitectura, en donde se puede concluir que estas últimas son suficientes para realizar las tareas en el repositorio, además de ser fácil de aprender y eficiente en cuanto al tiempo que se necesita para esto. Con esto se concluye que la arquitectura propuesta cumple con todos los requerimientos técnicos y permitirá realizar una gestión de datos de investigación dentro del centro en caso de ser implementada.

Hacer esta gestión trae costos pero a su vez grandes beneficios, se concluye que implementar la arquitectura diseñada generará valor para el centro no sólo evitando el problema de la pérdida de datos que se tiene, que era lo buscado en un principio, si no que permitirá cumplir de mejor manera sus objetivos, dará ventajas a la hora de competir por fondos concursables, posibilitará satisfacer los requerimientos de los financiadores y comités de ética y también ayudará a estandarizar los procesos internos del centro.

Para obtener mayores beneficios es necesario abrir los datos generados en el centro, siempre dentro de lo permitido por la ley. Se debe promover esta práctica dentro del centro en el futuro utilizando la arquitectura diseñada, la cual puede hacer disponible este libre acceso a los datos gestionados en ella.

Como la gestión de datos de investigación es un movimiento que ha ido creciendo a lo largo de los años, tomando fuerza en la comunidad científica se debe realizar una revisión periódica de toda la arquitectura para ir adaptándose a los nuevos estándares que se irán desarrollando en el tiempo.

Por último, se concluye que para que esta propuesta sea implementada en el centro, es necesario realizar una gestión de este cambio, donde se deberán invertir horas persona con todos los integran-

tes del centro para poder ocupar la arquitectura propuesta de manera correcta y generar finalmente el valor que se busca.

7.2. Recomendaciones y Trabajo Futuro

Con el fin de implementar la arquitectura propuesta en este trabajo, una de las primeras acciones que se debe negociar con la Dirección de Servicios de Información y Bibliotecas de la Universidad de Chile (SISIB) para esto. Se recomienda esto, antes que convertirse miembro de alguna agencia registradora de DOIs, debido a que estos cuentan con una cuenta que asigna estos identificadores a nombre de la universidad, como se hace actualmente con todas las publicaciones realizadas por algún investigador de esta, permitiendo una mayor visibilidad de los datos. Pensando en conversaciones con el SISIB, se debería analizar la situación actual del proyecto de diseño de políticas y revisar en conjunto las desarrolladas en este trabajo para servir de apoyo al SISIB. Una vez esto hecho se debe configurar el repositorio para que pueda asignar DOIs a los ítems creados en el sistema y editar el conversor de metadatos de los utilizados en el repositorio a DataCite para que funcione con el esquema actual de este (4.0).

Si bien se realizaron ciertos tests de usabilidad en el repositorio, sólo se contó con la participación de algunos integrantes del centro, se recomienda realizar nuevos tests con los nuevos integrantes del centro con el fin de evaluar de mejor manera tanto el repositorio como las guías entregadas, además de poder analizar de mejor manera el componente de facilidad de memorizar del atributo usabilidad

Como primer paso para la mejora de la arquitectura, se deben tomar en cuenta las propuestas de mejora para estas entregadas en el capítulo anterior, a las cuales hay que analizar su factibilidad técnica y práctica antes de desarrollarlas. Respecto a esto último, se recomienda designar a una persona encargada de la mantención y desarrollo del repositorio para hacer efectivas las propuestas y para que exista un seguimiento continuo a nivel técnico del sitio. Estas mejoras deben ser iterativas, sobre todo en las redacciones de las guías entregadas para que cada vez sea más simple para un nuevo integrante utilizarla. Una de las propuestas es analizar las funcionalidades de Dspace como el servidor de handles, el cual se recomienda utilizar en el futuro como un sustituto del DOI si es que en el momento no se tiene asignado uno y se desea subir un archivo al servidor.

Respecto a los planes de gestión de datos que se deben entregar, se podría pensar en diseñar una herramienta online simple del estilo de DMPOnline y DMPTool utilizando las plantillas entregadas en este trabajo, en donde se puedan ir guardando y editando los planes que se vayan generando, esta aplicación debería ser integrada al repositorio. También permitiría facilitar la aprobación por parte de los Jefes de Proyecto en el caso de utilizar cuentas en esta aplicación.

Dspace continuamente lanza nuevas versiones de su plataforma, actualmente lanzó una versión nueva (6.0) para cual se recomienda analizar su posible uso respecto al propuesto en este trabajo, la cual está hecha en su versión 5.6, que era la más actual al momento de desarrollar esta memoria. También al momento de subir al servidor oficial el repositorio, este debe ser registrado en re3data.org para su mayor visibilidad en la comunidad científica. Por otro lado, existen más alternativas que las revisadas en este trabajo, como Zenodo y DKAN para la construcción de repositorios, a

pesar de que la plataforma escogida cumple con todas las necesidades requeridas, se puede analizar la utilización de estas dos alternativas en el futuro.

Además, se recomienda seguir las propuestas relacionadas a la gestión del cambio en el centro, las cuales fueron mencionadas en el capítulo anterior, esto para poder implementar de manera efectiva la arquitectura diseñada y realizar efectivamente una gestión de datos de investigación. Como se mencionó en las propuestas, se debe realizar un especial énfasis en los jefes de proyecto, debido a que éstos serán los que deban aprobar o rechazar los planes de gestión de datos y en algunos casos los archivos subidos al repositorio, por lo que deben tener los conocimientos necesarios para poder hacerlo de la manera correcta e incentivar a los integrantes de su proyecto a cumplir los requisitos de las políticas.

Finalmente se propone analizar una posible oportunidad de negocios respecto a la arquitectura diseñada, ya que se cuentan con todo el proceso de creación de esta y además una ya diseñada. Se puede estudiar la posible venta del servicio de gestión de datos en otros centros de investigación o instituciones, lo que permitiría estandarizar la gestión de datos científicos en la misma universidad o en el país. Para esto, se debe considerar ser miembro de alguna agencia registradora de DOIs para contar con una mayor cantidad de identificadores a asignar, lo que podría depender de un futuro modelo de negocios que se pueda diseñar.

Bibliografía

- [1] J. Wood, T. Andersson, A. Bachem, C. Best, F. Genova, D. Lopez, W. Los, M. Marinucci, L. Romary, H. Van de Sompel *et al.*, “Riding the wave: How europe can gain from the rising tide of scientific data. final report of the high level expert group on scientific data; a submission to the european commission, october 2010,” 2010.
- [2] O. Publishing, *OECD Science, Technology and Industry Outlook 2012*. OECD Pub., 2012.
- [3] CONICYT, *Datos Científicos Abiertos, La Ciencia la hacemos entre Todos*. CONICYT, 2014.
- [4] G. Boulton, P. Campbell, B. Collins, P. Elias, W. Hall, G. Laurie, O. O’Neill, M. Rawlins, J. Thornton, P. Vallance *et al.*, “Science as an open enterprise,” *Royal Society, London*, vol. 104, 2012.
- [5] C. on Library and I. Resources, *Research Data Management: Principles, Practices and Prospects*. Council on Library and Information Resources, 2013.
- [6] N. Zhong, J. Liu, Y. Y. Yao, and S. Ohsuga, “Web intelligence (wi),” in *Computer software and applications conference, 2000. COMPSAC 2000. The 24th annual international*. IEEE, 2000, pp. 469–470.
- [7] Y. Y. Ning Zhong, Jiming Liu, *Web Intelligence*. Springer-Verlag Berlin Heidelberg, 2003.
- [8] W. I. Centre. Web intelligence centre – quiénes somos. Consultado: Octubre 2016. [Online]. Available: <http://wic.uchile.cl/quienes-somos/>
- [9] ——. Web intelligence centre – quiénes somos. Consultado: Octubre 2016. [Online]. Available: <http://wic.uchile.cl/akori/>
- [10] ——. Web intelligence centre – opinionzoom. Consultado: Octubre 2016. [Online]. Available: <http://wic.uchile.cl/opinionzooms/>
- [11] ——. Web intelligence centre – docode. Consultado: Octubre 2016. [Online]. Available: <http://wic.uchile.cl/docode/>
- [12] G. d. C. Ministerio de Economía, Fomento y Turismo. Cuarta encuesta nacional sobre gasto y personal en investigación y desarrollo (año de referencia 2013). Marzo: Octubre 2016. [Online]. Available:

<http://www.economia.gob.cl/estudios-y-encuestas/encuestas/encuestas-de-innovacion-e-id/cuarta-encuesta-nacional-sobre-gasto-y-personal-en-investigacion-y-desarrollo-ano-de-referencia-2013>

- [13] C. Diario La Tercera. Chile sube gasto en i+d, pero aún está lejos de la meta. Consultado: Marzo 2016. [Online]. Available: <http://www.latercera.com/noticia/negocios/lt-emprende/2016/01/3131-665614-9-chile-sube-gasto-en-id-pero-aun-esta-lejos-de-la-meta.shtml>
- [14] U. de Chile. Investigación - hechos y cifras. Consultado: Septiembre 2016. [Online]. Available: <http://www.uchile.cl/portal/presentacion/hechos-y-cifras/categorias/54564/investigacion>
- [15] C. FONDEF. ¿qué es fondef? Consultado: Octubre 2016. [Online]. Available: <http://www.conicyt.cl/fondef/sobre-fondef/que-es-fondef/>
- [16] C. FONDECYT. ¿qué es fondecyt? Consultado: Octubre 2016. [Online]. Available: <http://www.conicyt.cl/fondecyt/sobre-fondecyt/que-es-fondecyt/>
- [17] A. N. D. Service. Ands - governance. Consultado: Octubre 2016. [Online]. Available: <http://www.ands.org.au/about-us/governance>
- [18] ——. Ands - about us. Consultado: Octubre 2016. [Online]. Available: <http://www.ands.org.au/about-us>
- [19] E. Commision. H2020 - what is horizon 2020? Consultado: Octubre 2015. [Online]. Available: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>
- [20] E. COMMISSION, “Guidelines on fair data management in horizon 2020,” *HORIZON 2020*, 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [21] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific data*, vol. 3, 2016.
- [22] R. Project. Recode - home. Consultado: Octubre 2015. [Online]. Available: <http://recodeproject.eu/>
- [23] L. Project. Learn - project summary. Consultado: Octubre 2016. [Online]. Available: <http://learn-rdm.eu/about/project-summary/>
- [24] D. U. de Chile. ¿qué es datos.uchile.cl? Consultado: Noviembre 2016. [Online]. Available: http://datos.uchile.cl/pagina/quienes_somos
- [25] R. Erway, *Starting the Conversation: University-Wide Research Data Management Policy*. ERIC, 2013.
- [26] DataONE. (2012) Dataone education module: Data management. Consultado: Octubre 2016. [Online]. Available: http://www.dataone.org/sites/all/documents/L01_DataManagement.

pptx

- [27] J. M. Wicherts, D. Borsboom, J. Kats, and D. Molenaar, “The poor availability of psychological research data for reanalysis.” *American Psychologist*, vol. 61, no. 7, p. 726, 2006.
- [28] T. H. Vines, A. Y. Albert, R. L. Andrew, F. Débarre, D. G. Bock, M. T. Franklin, K. J. Gilbert, J.-S. Moore, S. Renaut, and D. J. Rennison, “The availability of research data declines rapidly with article age,” *Current biology*, vol. 24, no. 1, pp. 94–97, 2014.
- [29] A. N. D. Service. Research data policy and the Australian code for the responsible conduct of research. Consultado: Octubre 2015. [Online]. Available: <http://ands.org.au/guides/code-awareness>
- [30] T. J. Whyte, A., “Making the case for research data,” *DCC Briefing Papers. Edinburgh: Digital Curation*, 2011. [Online]. Available: <http://www.dcc.ac.uk/resources/briefing-papers/>
- [31] JETWI. Publication ethics. Consultado: Octubre 2016. [Online]. Available: <http://www.jetwi.us/index.php?m=content&c=index&a=lists&catid=167>
- [32] A. Victoria University. Publications requiring data deposit. Consultado: Octubre 2016. [Online]. Available: <http://guides.library.vu.edu.au/content.php?pid=489543&sid=4015042>
- [33] O. Publishing, *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Pub., 2007.
- [34] C. L. Borgman, “The conundrum of sharing research data,” *Journal of the American Society for Information Science and Technology*, vol. 63, no. 6, pp. 1059–1078, 2012.
- [35] F. C. Fang, R. G. Steen, and A. Casadevall, “Misconduct accounts for the majority of retracted scientific publications,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 42, pp. 17 028–17 033, 2012.
- [36] H. A. Piwowar, R. S. Day, and D. B. Fridsma, “Sharing detailed research data is associated with increased citation rate,” *PloS one*, vol. 2, no. 3, p. e308, 2007.
- [37] R. Proyect, *RECODE Policy Recommendations for Open Access to Research Data*. RECODE Proyect, 2014. [Online]. Available: <http://policy.recodeproject.eu/assets/recode-handbook.pdf>
- [38] N. Pontika, P. Knoth, M. Cancellieri, and S. Pearce, “Fostering open science to research using a taxonomy and an elearning portal,” in *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 2015, p. 11.
- [39] U. of Oxford. Policy on the management of research data and records. Consultado: Octubre 2015. [Online]. Available: <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/>
- [40] B. U. Libraries. What is research data? Consultado: Noviembre 2015. [Online]. Available:

<http://www.bu.edu/datamanagement/background/whatisdata/>

- [41] U. D. Archive. Research data lifecycle. Consultado: Octubre 2015. [Online]. Available: <http://www.data-archive.ac.uk/create-manage/life-cycle>
- [42] A. Swan, *Policy guidelines for the development and promotion of open access*. UNESCO, 2012.
- [43] P. Suber, “Open access overview,” *Retrieved from Peter Suber’s website: <http://www.earlham.edu/~peters/fos/overview.htm>*, 2007.
- [44] OECD. Open science. Consultado: Octubre 2015. [Online]. Available: <https://www.oecd.org/sti/outlook/e-outlook/stipolicyprofiles/interactionsforinnovation/openscience.htm>
- [45] C. Argentina. Nuevo observatorio virtual argentino. Consultado: Octubre 2015. [Online]. Available: <http://nova.conicet.gov.ar/index.html>
- [46] A. L. M. A. (ALMA). Sobre alma. Consultado: Octubre 2015. [Online]. Available: <http://www.almaobservatory.org/es/sobre-alma>
- [47] B. University. Importance of data management. Consultado: Octubre 2016. [Online]. Available: <http://www.bu.edu/datamanagement/background/importance/>
- [48] ANDS. Creating a data management framework. Marzo: Octubre 2016. [Online]. Available: <http://www.ands.org.au/guides/creating-a-data-management-framework>
- [49] D.-G. f. R. . I. European Commission, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. European Commission, 2016.
- [50] B. O. de la República Argentina. Resolución 753 - e/2016. Consultado: Diciembre 2016. [Online]. Available: <https://www.boletinoficial.gob.ar/DetalleNormativa/1120337>
- [51] C. on Data for Science and T. (CODATA), *Current Best Practice for Research Data Management Policies*. CODATA, 2014.
- [52] A. N. D. S. (ANDS), *Outline of a Research Data Management Policy for Australian Universities / Institutions*. ANDS, 2010.
- [53] L. Horton and D. (2014), “Overview of uk institution rdm policies, digital curation centre,” *Digital Curation Centre UK*, 2014. [Online]. Available: <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>
- [54] M. Donnelly, “Five steps to developing a research data policy’. dcc ‘quickstart’ leaflets,” *Digital Curation Centre UK*, 2014. [Online]. Available: <http://www.dcc.ac.uk/resources/policy-and-legal/five-steps-developing-research-data-policy/five-steps-developing-research>
- [55] B. E. Bargmeyer and D. W. Gillman, “Metadata standards and metadata registries: An overview,” in *International Conference on Establishment Surveys II*, vol. 19, 2000.

- [56] B. University. Metadata. Consultado: Agosto 2016. [Online]. Available: <http://www.bu.edu/datamanagement/outline/elements/metadata/>
- [57] U. D. C. C. (DCC). Metadata standars. Consultado: Agosto 2016. [Online]. Available: <http://www.dcc.ac.uk/resources/metadata-standards>
- [58] R. D. Alliance. About. Consultado: Agosto 2016. [Online]. Available: <https://rd-alliance.org/about.html>
- [59] ——. Metadata directory. Consultado: Agosto 2016. [Online]. Available: <http://rd-alliance.github.io/metadata-directory/standards/>
- [60] U. de Chile. Información y bibliotecas - doi. Consultado: Octubre 2016. [Online]. Available: <http://www.uchile.cl/portal/informacion-y-bibliotecas/ayudas-y-tutoriales/100282/doi-digital-object-identifier>
- [61] C. Commons. Creative commons - frequently asked questions. Consultado: Octubre 2016. [Online]. Available: <https://creativecommons.org/faq>
- [62] ——. Sobre las licencias. Consultado: Octubre 2016. [Online]. Available: <https://creativecommons.org/licenses/>
- [63] O. D. Commons. Open data commons - about. Consultado: Octubre 2016. [Online]. Available: <http://opendatacommons.org/about/>
- [64] C. University. Introduction to intellectual property rights in data management. Consultado: Octubre 2016. [Online]. Available: <http://data.research.cornell.edu/content/intellectual-property>
- [65] J. Nielsen, *Usability engineering*. Elsevier, 1994.
- [66] ——, “Usability 101: Introduction to usability (2012),” *URL: http://www.nngroup.com/articles/usability-101-introduction-to-usability/[accessed 2016-10-15][WebCite Cache]*, 2014.
- [67] J. Nielsen, K. P. Coyne, and M. Tahir, “Make it usable,” *PC Magazine*, vol. 20, no. 3, pp. 24–28, 2001.
- [68] J. Nielsen, “Thinking aloud: The n°1 usability tool,” *URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool//[accessed 2016-10-15][WebCite Cache]*, 2012.
- [69] J. Rubin and D. Chisnell, *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons, 2008.
- [70] N. Jakob, “Why you only need to test with 5 users,” *Jakob Nielsen’s Alterbox*. [URL: www.useit.com/alterbox/20000319.html], vol. 19, 2000.
- [71] J. Nielsen, “How many test users in a usability study,” *Nielsen Norman Group*, vol. 4, 2012.

- [72] F. de Ciencias Físicas y Matemáticas Uchile. Investigación - comité de Ética. Consultado: Octubre 2016. [Online]. Available: <http://ingenieria.uchile.cl/investigacion/presentacion/88397/comite-de-etica>
- [73] B. del Congreso Nacional de Chile. Sobre proteccion de la vida privada - ley 19.628. Consultado: Octubre 2016. [Online]. Available: <http://www.leychile.cl/Navegar?idNorma=141599>
- [74] ——. Sobre la investigacion cientifica en el ser humano, su genoma, y prohíbe la clonacion humana - ley 20.120. Consultado: Mayo 2016. [Online]. Available: <http://www.leychile.cl/Navegar?idNorma=253478>
- [75] FONDECYT, “Requerimientos Ético/bioético en instituciones patrocinantes, propuesta de pauta de seguimiento de proyectos fondecyt en ejecución,” *CONICYT*, 2012. [Online]. Available: http://www.conicyt.cl/fondecyt/files/2012/10/requerimientos_etico_bioetico.pdf
- [76] ANDS. Ands - policies. Consultado: Mayo 2016. [Online]. Available: <https://projects.ands.org.au/policy.php>
- [77] U. of Birmingham. University of birmingham - code of practice for research. Consultado: Mayo 2016. [Online]. Available: <http://www.birmingham.ac.uk/Documents/university/legal/research.pdf>
- [78] B. del Congreso Nacional de Chile. Propiedad intelectual - ley n° 17.336. Consultado: Octubre 2016. [Online]. Available: <https://www.leychile.cl/Navegar?idNorma=28933>
- [79] C. Commons. Quienes somos. Consultado: Octubre 2016. [Online]. Available: <https://creativecommons.cl/quienes-somos/>
- [80] U. D. Archive. About us. Consultado: Octubre 2015. [Online]. Available: <http://www.data-archive.ac.uk/about>
- [81] S. Farnel and A. Shiri, “Metadata for research data: current practices and trends,” in *Proceedings of the 2014 International Conference on Dublin Core and Metadata Applications, Austin, Texas, USA*. Citeseer, 2014, pp. 74–82.
- [82] J. Brase, M. Lautenschlager, and I. Sens, “The tenth anniversary of assigning doi names to scientific data and a five year history of datacite,” *D-Lib magazine*, vol. 21, no. 1, p. 10, 2015.
- [83] C. de Metadatos de la Biblioteca Nacional de Chile, “Guía para la creación de metadatos usando dublin core,” *Biblioteca Nacional de Chile*, 2009.
- [84] M. DUBLIN and C. CARÁCTER, “Datasets: plantilla normalizada para la descripción de registros en digital. csic,” *DIGITAL.CSIC*, 2014.
- [85] U. D. C. C. (DCC). Data plan guidance and examples. Consultado: Agosto 2016. [Online]. Available: <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>
- [86] DataCite. Datacite metadata schema. Consultado: Octubre 2016. [Online]. Available:

<https://schema.datacite.org/>

- [87] C. A. de Bioética FONDECYT/CONICYT. Sugerencias para escribir un consentimiento informado en estudios con personas. Consultado: Septiembre 2016. [Online]. Available: <http://www.med.uchile.cl/images/pdf/consentimiento%20informado%20sugerencias.pdf>
- [88] R. Trends. A quick look at references to research data repositories. Consultado: Octubre 2016. [Online]. Available: <https://www.researchtrends.com/issue-38-september-2014/a-quick-look-at-references-to-research-data-repositories/>
- [89] Re3data.org. Datacite, re3data.org, and databib announce collaboration. Consultado: Octubre 2016. [Online]. Available: <http://www.re3data.org/2014/03/datacite-re3data-org-databib-collaboration/>
- [90] T. D. Project. About. Consultado: Noviembre 2016. [Online]. Available: <http://dataverse.org/about>
- [91] CKAN. About. Consultado: Octubre 2016. [Online]. Available: <http://ckan.org/about/>
- [92] DSpace. About dspace. Consultado: Octubre 2016. [Online]. Available: <http://www.dspace.org/introducing>
- [93] EPrints. About us. Consultado: Octubre 2016. [Online]. Available: <http://www.eprints.org/uk/index.php/about/>
- [94] Fedora. About fedora. Consultado: Octubre 2016. [Online]. Available: <http://fedorarepository.org/about>
- [95] D. I. Valencia Velázquez and D. Ortiz Ancona, “Software para construir repositorios digitales,” *Universidad Nacional Autónoma de México (Mexico)*, 2013.
- [96] DuraSpace. Dspace 5.x documentation. Consultado: Octubre 2016. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC5x/DSpace+5.x+Documentation>
- [97] ——. Dspace 5.x documentation. Consultado: Octubre 2016. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC5x/DSpace+5.x+Documentation>
- [98] T. D. Project. Dataverse 4.5.1 guides. Consultado: Noviembre 2016. [Online]. Available: <http://guides.dataverse.org/en/latest/>
- [99] J. G. Bankier and K. Gleason, “Institutional repository software comparison,” 2014.
- [100] W. Nixon, “Daedalus: Initial experiences with eprints and dspace at the university of glasgow,” *Ariadne*, no. 37, 2003.
- [101] G. Pyrounakis and M. Nikolaidou, “Comparing open source digital library software.” 2009.
- [102] O. A. Forum. Comparison of dspace and eprints. Consultado: Octubre 2016. [Online]. Available: <https://www.oaforum.org/resources/tvtoolscomp.php>

- [103] D. Dspace. Doi digital object identifier. Consultado: Noviembre 2016. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC5x/DOI+Digital+Object+Identifier>
- [104] DataCite. Datacite metadata schema documentation for the publication and citation of research data. Consultado: Octubre 2016. [Online]. Available: https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf
- [105] GitHub. datacite-metadata-generator. Consultado: Noviembre 2016. [Online]. Available: <https://github.com/mpaluch/datacite-metadata-generator>
- [106] PCFactory. Categoría discos servers. Consultado: Octubre 2016. [Online]. Available: <https://www.pcfactory.cl/?categoria=411&papa=407>
- [107] DataCite. Become a member. Consultado: Octubre 2016. [Online]. Available: <https://www.datacite.org/become.html>
- [108] investing.com. Eur - clp. Consultado: Octubre 2016. [Online]. Available: <http://es.investing.com/currencies/eur-clp-historical-data>
- [109] mEdra. Terms and conditions. Consultado: Octubre 2016. [Online]. Available: <https://www.medra.org/en/terms.htm>
- [110] investing.com. Usd - clp. Consultado: Octubre 2016. [Online]. Available: <http://es.investing.com/currencies/usd-clp-historical-data>
- [111] CrossRef. Publisher fees. Consultado: Octubre 2016. [Online]. Available: http://www.crossref.org/02publishers/20pub_fees.html
- [112] W. S. Kane, *The truth about thriving in change*. FT Press, 2008.
- [113] figshare. About. Consultado: Octubre 2016. [Online]. Available: <https://figshare.com/about>
- [114] J. Jeng, "What is usability in the context of the digital library and how can it be measured?" *Information technology and libraries*, vol. 24, no. 2, 2005.
- [115] R. C. Amorim, J. A. Castro, J. R. da Silva, and C. Ribeiro, "A comparison of research data management platforms: architecture, flexible metadata and interoperability," *Universal Access in the Information Society*, pp. 1–12, 2016.
- [116] Dialnet. El acceso abierto a la literatura científica. Consultado: Octubre 2015. [Online]. Available: <http://dialnet.unirioja.es/info/ayuda/oai>

Anexos A

Antecedentes

A.1. Investigación en Chile

Figura A.1: RATIO GASTO I+D/PIB Y EVOLUCIÓN DEL GASTO EN I+D.

RATIO GASTO I+D/PIB Y EVOLUCIÓN DEL GASTO EN I+D								
	2007	2008	2009	2010	2011	2012	2013	2014p
Producto Interno Bruto (MM\$ corrientes)	90.428.771	93.847.932	96.443.761	110.998.729	121.319.462	129.027.553	137.028.983	147.184.925
Var. PIB Nominal		3,80%	2,80%	15,10%	9,30%	6,40%	6,20%	7,40%
Producto Interno Bruto (MM\$ encadenados)	90.856.522	93.847.932	92.875.262	98.219.034	103.954.673	109.627.615	114.260.687	116.424.841
Var. PIB Real		3,30%	-1,00%	5,80%	5,80%	5,50%	4,20%	1,90%
Gasto interno I+D (MM\$ corrientes)	280.785	351.923	340.357	367.393	428.781	470.675	535.924	558.240
Var. Gasto I+D Nominal		25,30%	-3,30%	7,90%	16,70%	9,80%	13,90%	4,20%
Ratio Gasto I+D/PIB	0,31%	0,37%	0,35%	0,33%	0,35%	0,36%	0,39%	0,38%
Gasto I+D (MM\$ reales de 2014)	348.912	408.347	400.453	419.783	469.108	507.402	560.829	558.240
Var. Gasto I+D Real		17,00%	-1,90%	4,80%	11,70%	8,20%	10,50%	-0,50%

Fuente: Primera, Segunda, Tercera, Cuarta y Quinta Encuesta Nacional sobre Gasto y Personal en I+D. Se utilizó el deflactor del IPC desde el INE y datos de cuentas nacionales del Banco Central de Chile para PIB. P: Preliminar

Figura A.2: GASTO EN I+D SEGÚN FUENTE DE FINANCIAMIENTO (Porcentaje).

GASTO EN I+D SEGÚN FUENTE DE FINANCIAMIENTO (Porcentaje).								
	2007	2008	2009	2010	2011	2012	2013	2014p
Empresas	39%	44%	27%	25%	34%	35%	34%	32%
Estado	36%	34%	38%	40%	34%	36%	38%	44%
Ed. Superior	19%	17%	14%	13%	10%	9%	12%	9%
IPSFL	3%	2%	2%	2%	2%	2%	1%	1%
Fondos Internacionales	4%	3%	19%	20%	21%	18%	15%	14%
Total	100%							

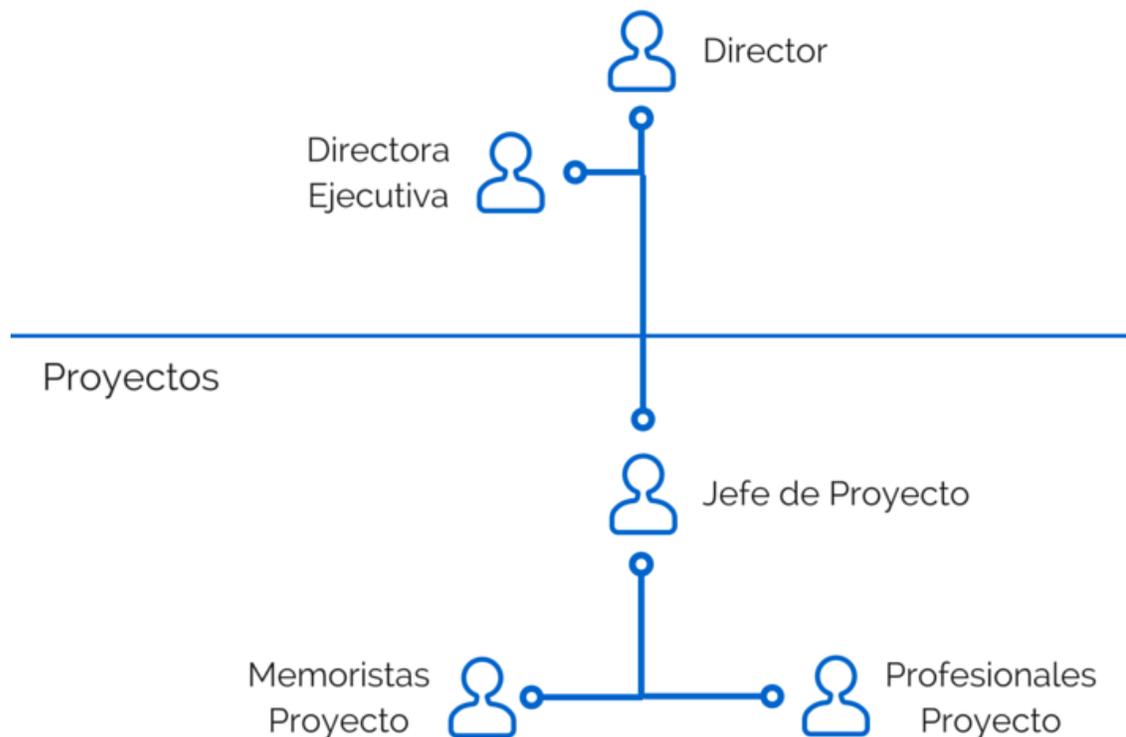
Fuente: Primera, Segunda, Tercera, Cuarta y Quinta Encuesta Nacional sobre Gasto y Personal en I+D. Se utilizó el deflactor del IPC desde el INE y datos de cuentas nacionales del Banco Central de Chile para PIB. P: Preliminar

A.2. Web Intelligence Centre

Estructura Organizacional

La Estructura Organizacional del Web Intelligence Centre se puede encontrar en la figura A.3.

Figura A.3: Estructura Organizacional WIC.



Fuente: Elaboración Propia.

A.3. Entrevistas

Claudio Aracena

- Año en que realizó la memoria: 2013.
- Breve reseña del experimento y las bases de datos que generó:

Era un experimento que juntaba eye tracking, dilatación pupilar y EEG. Se mostraba una imagen con alguna connotación emocional y las personas exploraban por 4 segundos la imagen. Se calificaban las imágenes y las valencias. Eran 180 imágenes, duraba aproximadamente media hora por persona (eran 10)

- ¿Qué se hizo con estas bases de datos?

Las guardaba en un Dropbox que teníamos en el AKORI, no sé si lo siguen usando.

- ¿Cómo están documentadas esas bases de datos?

No, no se realizó nada de eso.

- ¿Existe una descripción formal de qué contiene y cómo están organizadas?

No, tampoco.

- ¿Se utilizaron metadatos específicos para describirlas?

No, tampoco.

- ¿Sabes lo que es un plan de datos? ¿El WIC, en su proyecto AKORI le proporcionó ayuda para realizar su plan de manejo de datos?

No, pero yo estaba bajo la tutela de otro doctor (Enzo Brunetti), yo solo hacia el experimento, el quizás tenía algo que ver con eso.

- ¿El WIC en su proyecto AKORI tiene políticas para esto (quién es dueño de los datos, donde debe almacenarlas al terminar su investigación, etc.?)

No, por lo menos cuando lo hice yo no existía nada de eso.

Joaquín Jadue

- Año en que realizó la memoria: 2013

- Breve reseña del experimento y las bases de datos que generó:

Primero se invitó a personas a participar, la idea era obtener la mayor cantidad de personas posible, aproximadamente 20. Tenían que firmar un consentimiento, el experimento era: la persona se sentaba frente al computador, primero calibraban y después se daban instrucciones de que se trataba el experimento y objetivos, para que sea lo más espontaneo posible, la persona elegía imágenes de una pantalla que tenía 9 imágenes (3 x 3) y la persona que estaba conectada al eye tracker podía elegir la imagen que más le llamaba la atención. 90 pantallas distintas, 90 categorías.

- ¿Qué se hizo con estas bases de datos?

Las respaldé yo mismo, usábamos un computador de medicina que fallaba un poco y nadie aseguraba que iban a estar la próxima vez que fuéramos, la respaldó Claudio también porque podían borrarlas del computador, el cual no se podía manipular mucho, de hecho teníamos problemas para usar pendrives.

- ¿Cómo están documentadas esas bases de datos?

No, nada de eso.

- ¿Existe una descripción formal de qué contiene y cómo están organizadas?

En la memoria, lo que entregaba el eye tracker no más.

- ¿Se utilizaron metadatos específicos para describirlas?

No, tampoco.

- ¿Al momento de realizar la memoria en algún momento tuvo que firmar, leer o informarse sobre el plan de manejo de datos?

Yo no, solo a las personas un consentimiento.

- ¿El WIC, en su proyecto AKORI le proporcionó ayuda para realizar su plan de manejo de datos?

No, tampoco.

- ¿El WIC en su proyecto AKORI tiene políticas para esto (quién es dueño de los datos, donde debe almacenarlas al terminar su investigación, etc.?)

Deben existir pero conmigo fueron más relajados en ese sentido.

Luis Salas

- Año en que realizó la memoria: 2014

- Breve reseña del experimento y las bases de datos que generó:

Se realizó un experimento parecido a uno anterior (de Joaquín Jadue), el donde se simulaba un sitio web muy simple con una página de inicio para presentar el experimento y después cada persona veía una pantalla en negro con una grilla de 3x3 con imágenes de una misma categoría, eran 90. Al final aparecía una página de agradecimiento. Como Joaquín no las compartió, se perdieron por eso se tuvo que replicar el experimento. La diferencia era que ahora también se medía la actividad cerebral con un encefalograma (EEG).

- ¿Qué se hizo con estas bases de datos?

Me quedé con algunas copias, y el Claudio Aracena me las pidió una vez para continuar las investigaciones.

-¿Cómo están documentadas esas bases de datos?

No estaban documentadas.

- ¿Existe una descripción formal de qué contiene y cómo están organizadas?

No, la verdad no.

- ¿Se utilizaron metadatos específicos para describirlas? No, tampoco.

- ¿Al momento de realizar la memoria en algún momento tuvo que firmar, leer o informarse sobre el plan de manejo de datos?

En medicina les pidieron firmar un papel pero solo por el tipo de experimento que estábamos haciendo (por ser temas médicos)

- ¿El WIC, en su proyecto AKORI le proporcionó ayuda para realizar su plan de manejo de datos?

No, no se realizan.

- ¿El WIC en su proyecto AKORI tiene políticas para esto (quién es dueño de los datos, donde debe almacenarlas al terminar su investigación, etc.?)

No existe nada de eso.

Anexos B

Marco Teórico

B.1. Creative Commons: Condiciones

Figura B.1: Tipos de condiciones.



Fuente: Elaboración propia

Anexos C

Políticas de Gestión de Datos

C.1. Logos

Figura C.1: Logo Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.



Figura C.2: Logo Web Intelligence Centre.



C.2. Formatos de Archivos

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
<p>Quantitative tabular data with extensive metadata.</p> <p>a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data.</p>	<p>SPSS portable format (.por)</p> <p>delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information</p> <p>some structured text or mark-up file containing metadata information, e.g. DDI XML file.</p>	<p>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta).</p> <p>MS Access (.mdb/.accdb)</p>
<p>Quantitative tabular data with minimal metadata</p> <p>a matrix of data with or without column headings or variable names, but no other metadata or labelling.</p>	<p>comma-separated values (CSV) file (.csv)</p> <p>tab-delimited file (.tab)</p> <p>including delimited text of given character set with SQL data definition statements where appropriate.</p>	<p>delimited text of given character set - only characters not present in the data should be used as delimiters (.txt)</p> <p>widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods).</p>
<p>Geospatial data</p> <p>vector and raster data</p>	<p>ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn)</p> <p>geo-referenced TIFF (.tif, .tfw)</p> <p>CAD data (.dwg)</p> <p>tabular GIS attribute data</p>	<p>ESRI Geodatabase format (.mdb)</p> <p>MapInfo Interchange Format (.mif) for vector data</p> <p>Keyhole Mark-up Language (KML) (.kml)</p> <p>Adobe Illustrator (.ai), CAD data (.dxf or .svg)</p> <p>binary formats of GIS and CAD packages</p>

<p>Qualitative data</p> <p>textual</p>	<p>eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml)</p> <p>Rich Text Format (.rtf)</p> <p>plain text data, ASCII (.txt)</p>	<p>Hypertext Mark-up Language (HTML) (.html)</p> <p>widely-used proprietary formats, e.g. MS Word (.doc/.docx)</p> <p>some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti</p>
<p>Digital image data</p>	<p>TIFF version 6 uncompressed (.tif)</p>	<p>JPEG (.jpeg, .jpg) but only if created in this format</p> <p>TIFF (other versions) (.tif, .tiff)</p> <p>Adobe Portable Document Format (PDF/A, PDF) (.pdf)</p> <p>standard applicable RAW image format (.raw)</p> <p>Photoshop files (.psd)</p>
<p>Digital audio data</p>	<p>Free Lossless Audio Codec (FLAC) (.flac)</p>	<p>MPEG-1 Audio Layer 3 (.mp3) but only if created in this format</p> <p>Audio Interchange File Format (AIFF) (.aif)</p> <p>Waveform Audio Format (WAV) (.wav)</p>
<p>Digital video data</p>	<p>MPEG-4 (.mp4)</p> <p>motion JPEG 2000 (.mj2)</p>	

Documentation and scripts	Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)	plain text (.txt) some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0
---------------------------	---	--

Tabla C.1: Formatos aceptables para la preservación de datos

Fuente: UK Data Archive

Políticas de Gestión de Datos Científicos

1. Principios y Objetivos

Estas políticas se basan bajo los principios de que los datos primarios son esenciales para los resultados de la investigación realizada en el centro, es por esto que la gestión de estos datos debe realizarse asegurando su integridad para facilitar y dar paso a nuevas investigaciones respetando todos los acuerdos de privacidad, propiedad u otros requerimientos. Esto debe ser promovido en todos los integrantes del centro de investigación.

Este documento establecerá todos los requerimientos relacionados a la gestión de datos de investigación que sean generados en alguna investigación realizada en el Web Intelligence Centre (WIC) de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Los objetivos de esta política son los siguientes:

- Fomentar y establecer una cultura de buenas prácticas de investigación referentes a la gestión de datos científicos generados en el Web Intelligence Centre.
- Asegurar la preservación de los datos generados en el WIC, almacenándola y haciéndola accesible para su uso y re-uso con el fin de continuar las investigaciones realizadas.
- Disminuir los riesgos relacionados a una mala gestión de datos en el WIC.
- Asegurar que los integrantes del centro sean capaces de cumplir con todas las políticas o reglamentos en donde la buena gestión de datos contribuya a la investigación.

2. Alcances

Las Políticas de Gestión de Datos Científicos se aplicarán a todos los integrantes del Web Intelligence Centre o externos que realicen o estén involucrados en alguna investigación o sean parte de algún proyecto del centro.

3. Definiciones

Los siguientes términos y sus definiciones serán las utilizadas por estas políticas en todas sus secciones.

3.1. Datos de Investigación

O datos científicos, es la información recolectada (sin importar la forma o medios en que pueda existir) necesaria para validar las observaciones, conclusiones o resultados de un proyecto de investigación¹.

3.2. Plan de Manejo de Datos

Documento en donde se muestra cómo serán manejados los datos científicos durante y después de cada proyecto. Debe describir el ciclo de vida para todos los conjuntos de datos que serán recolectados, procesados o generados por el proyecto de investigación².

3.3. Datos Sensibles

Son todos los datos o conjuntos de datos que puedan tener problemas legales, confidencialidad o éticos. Principalmente cuando son datos generados en experimentos realizados con humanos o animales.

3.4. Digital Object Identifier

Es una forma de identificar un objeto digital el cual, aunque cambie de URL seguirá teniendo la misma identificación³.

¹ <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/>

² D.-G. f. R. I. European Commission, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. European Commission, 2016

³ <http://www.uchile.cl/portal/informacion-y-bibliotecas/ayudas-y-tutoriales/100282/doi-digital-object-identifier>

4. Políticas

4.1. Propiedad

- a. Los derechos de todos los datos de investigación generados dentro del marco de algún proyecto o investigación realizada en el centro serán exclusivamente del WIC, a excepción de que existan requerimiento de los financiadores o de la universidad.
- b. En caso de ser datos generados en experimentos con seres humanos, los datos serán de propiedad de ellos, pero los derechos a usar serán exclusivos del centro de investigación. Estos derechos de uso deben ser descritos en los consentimientos informados que firmen al momento de realizar estos experimentos.

4.2. Criterios

- a. Los datos científicos que sean creados a partir de cualquier clase de experimento realizado bajo alguna investigación ejecutada en el centro, deberá ser depositados en el repositorio entregado por el centro sin excepción y siguiendo las guías entregadas por el centro.
- b. En caso de que existan temas legales relacionados a datos generados en experimentos, se deberá guardar de manera digital toda la documentación necesaria para respaldar la toma de los datos, por ejemplo, consentimientos informados para el caso de experimentos realizados con seres humanos, que queda descrito bajo la ley 20.120 de Chile, u otros documentos requeridos por el Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas.
- c. Se debe respaldar todos los datos que sean considerados de interés para futuras investigaciones, siguiendo los mismos protocolos que los datos mencionados en el punto a.

4.3. Plan de Manejo de Datos

- a. Cualquier integrante del centro de investigación que desee realizar una investigación en él, deberá entregar dos versiones de un plan de manejo de datos al jefe de proyecto, uno antes de iniciar cualquier tipo de experimento referente a estas investigaciones y otro una vez finalizada esta investigación.
- b. La primera versión deberá contener los siguientes puntos: nombre del trabajo que se está realizando, nombre del proyecto del centro a cuál pertenece, fecha de creación del documento, nombre del encargado del proyecto en el WIC, código y fuente de financiamiento del trabajo, descripción del proyecto y de la investigación o experimentos a realizar, datos del investigador (nombre y contacto), políticas que deban cumplir en el centro, descripción de los datos que se espera generar en el proyecto (incluyendo formatos,

- tamaños esperados de archivos y cómo serán adquiridos y procesados en caso de contar con esta última información, junto a las fechas estimadas de los experimentos que se realicen o extraigan datos). Para su diseño se usarán los estándares acordados en el anexo A de este documento.
- c. La segunda versión del plan de manejo de datos deberá contener los puntos mencionados en a) cambiando la información de los datos esperados a generar y fechas estimativas de experimentos en caso de realizarse, por los resultados y fechas reales de estos, además de las metodologías y algoritmos aplicados a los datos generados, las notas de laboratorio si es que se tienen, dónde y en qué formatos serán almacenados junto a los metadatos que contendrán. En caso de que se generen datos sensibles qué documentos se guardarán referentes a temas legales o requerimientos de la universidad o financiadores y definir quién será el dueño de los datos mencionados en el documento.
 - d. Para su diseño se usarán los estándares acordados en las guías y *templates* otorgados por el centro.
 - e. Las dos versiones una vez aprobadas deberán ser almacenadas en el repositorio otorgado por el centro, de manera digital.
 - f. Debe ser aprobado por el Jefe del Proyecto en donde se esté realizando la investigación.

4.4. Requerimientos Externos

- a. En caso de que exista algún requerimiento externo, ya sea a nivel institucional o de parte del financiador, se deben respetar estos requisitos con prioridad a las políticas de gestión de datos.
- b. En caso de que esto suceda, al momento de realizar la revisión periódica deberá ser analizado si es necesario agregar estos requerimientos a las políticas para futuras versiones.

4.5. Re-uso y Descubrimiento

- a. Los datos de investigación generados deberán ser almacenados por un período mínimo de 10 años, utilizando los formatos y metadatos especificados en el plan de manejo de datos.
- b. Los derechos de acceso y re-uso no se podrán otorgar a publicadores comerciales sin retener el derecho a abrir los datos de investigación.
- c. El acceso a los datos de investigación para investigaciones que están en curso sólo será disponible para integrantes actuales del centro.
- d. El acceso a los datos de investigación para investigaciones finalizadas, si es que no se han declarado como abiertos, deberá ser exclusivo para integrantes actuales del centro de investigación y podrá ser realizado desde el momento en que se almacenan por primera vez hasta la fecha estipulada de retención.
- e. Solo se podrán abrir los datos en donde no puedan existir problemas legales, de privacidad, confidencialidad o éticos.
- f. En caso de abrir los datos se deberá especificar la existencia de tiempos de privilegio de acceso ellos, el cual deberá ser definido por el jefe del proyecto.

- g. En caso de abrir los datos de investigación se deberá especificar la forma en que se citará al centro y al autor de estos datos en los lugares que se publiquen.
- h. Si se desarrollan publicaciones referentes a cierta investigación que generó datos científicos y estos son necesarios para su validación, se deberá adjuntar información para acceder a los datos (por ejemplo, un link o en caso de que exista un Digital Object Identifier (DOI)) en la publicación para acceder a los datos utilizados en ella.
- i. Si alguna entidad externa requiere acceso a los datos generados, se deberá aprobar por parte de la directora ejecutiva su posible acceso. No se dará acceso cuando se puedan tener problemas legales o éticos.

4.6. Ética

- a. Todos los integrantes del centro de investigación deberán realizar sus investigaciones cumpliendo todas las leyes, requerimientos del centro, universidad y financiadores, relacionados a la generación de datos sensibles, especialmente los referentes a experimentos en donde se utilizaron seres humanos.
- b. Toda investigación realizada en humanos deberá cumplir lo estipulado en las leyes 19.628 y 20.120 de Chile, sumado a esto, las exigencias del Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- c. No se podrán abrir conjunto de datos sensibles para evitar el riesgo al no cumplimiento de alguno de los puntos a. o b.
- d. Toda documentación generada para la realización de experimentos con humanos, exigidos por la ley, Comité de Ética o financiadores, deberá ser digitalizada y almacenada junto a los datos en el repositorio del centro.
- e. Los datos generados en experimentos con humanos son de propiedad de estos, el centro de investigación sólo cuenta con los derechos a utilizarlos en sus investigaciones, esto debe quedar plasmado en los documentos exigidos por las entidades mencionadas en los puntos anteriores.

4.7. Eliminación, destrucción y movimiento de registros.

- a. Una vez pasado el período de retención de algún conjunto de datos, se deberá reevaluar su almacenamiento por un nuevo período.
- b. En caso de que se desee eliminar registros del repositorio del centro, los jefes de proyectos serán los encargados de realizar esta acción.
- c. En caso de que se deseen mover registros o abrir los datos en un repositorio externo, esto deberá ser aprobado por parte de la directora ejecutiva y tendrá que ser llevado a cabo por el jefe de proyecto, siguiendo todas las normas de formatos y metadatos necesarios para su re-uso y descubrimiento.
- d. En caso de realizarse una migración a otro repositorio otorgado por el centro, el traslado al nuevo lugar de almacenamiento deberá ser realizado en conjunto por todos los jefes de

proyectos y la directora ejecutiva, asegurando los estándares acordados de formatos y metadatos.

5. Roles y Responsabilidades

Los roles y responsabilidades serán los siguientes:

5.1. El Web Intelligence Centre tiene la responsabilidad de:

- a. Proporcionar infraestructura de almacenamiento segura para los datos de investigación generados en el mismo centro.
- b. Entregar facilidades para el almacenamiento, disponibilidad y accesibilidad a conjuntos de datos generados en él para incentivar su re-uso dentro y fuera de este.
- c. Especificar quién es dueño de los datos científicos generados en investigaciones realizadas en el centro.
- d. Dar soporte referente a la gestión de datos científicos a todos los integrantes del centro.
- e. Dar a conocer las políticas del centro a todos sus integrantes.
- f. Realizar una revisión periódica de las políticas con el fin de que no queden obsoletas en el tiempo.
- g. Ser el responsable por los costos relacionados a la gestión de datos científicos en el centro.

5.2. El director del centro tiene el deber de:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Aprobar los documentos de compromiso por parte de los integrantes del centro acerca del cumplimiento de estas políticas.

5.3. La directora ejecutiva deberá:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Aprobar traslados o destrucción de datos de investigación guardados en el repositorio del centro en donde no haya terminado el periodo de retención de estos.
- c. Analizar, una vez terminado el periodo de retención de ciertos datos, si se mantendrán o se eliminarán los sets de datos del repositorio utilizado por el centro, en el primer caso además deberá especificar por cuánto tiempo más seguirán almacenados, en el segundo caso deberá ser la encargada de eliminarlos.
- d. Agregar en caso de aceptar nuevos formatos de archivos permitidos por el centro para subir al repositorio a la guía de metadatos en la sección formatos de los Anexos de ese documento.

5.4. Los jefes de proyecto deberán:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Supervisar a los miembros de los proyectos sobre el cumplimiento de las políticas.
- c. Asegurar que todos los integrantes del proyecto hayan firmado el acuerdo de cumplimiento de las políticas.
- d. En caso de realizar algún tipo de investigación también deberán cumplir los deberes de los demás integrantes del proyecto en donde la realice.
- e. Ser los encargados de eliminar los registros relacionados a su proyecto del repositorio, en caso de que se considere necesario.
- f. Una vez terminado el proyecto entregar un reporte de cumplimiento de las políticas dentro del proyecto.
- g. Aprobar los planes de manejo de datos entregados por los integrantes del proyecto que quieran realizar una investigación en él.

5.5. Los integrantes de los proyectos tienen la responsabilidad de:

- a. Firmar un acuerdo en donde se especifica que se tiene conocimiento y se realizará el cumplimiento de las políticas del centro.
- b. Identificar qué datos y documentos serán los que se deban respaldar para su posterior re-uso.
- c. Resguardar los datos que generen en sus investigaciones bajo los requerimientos de las políticas del centro.
- d. Asegurar la integridad de los datos durante todo el ciclo de vida de estos.
- e. Almacenar los datos de investigación en formatos y con lo metadatos suficientes para asegurar su fácil re-uso y descubrimiento.
- f. Realizar las acciones contenidas en el documento de acuerdo a los protocolos éticos y legislativos respectivos.

6. Aprobación y Revisión Periódica

Este documento será sujeto a una revisión periódica de al menos una vez al año, con el fin de ir mejorando las políticas y evitar su obsolescencia con el paso del tiempo. Una vez hecha la revisión se deberán actualizar en menos de 2 meses una vez pasada la revisión.

Todos los cambios que se produzcan deberán ser primero revisados y aprobados por los jefes de cada proyecto y la directora ejecutiva para finalmente ser aprobadas por el director o encargado del centro.

Si se realizan cambios se deberá firmar un nuevo acuerdo de cumplimiento por parte de los integrantes del centro con el fin de que todos sean conscientes de estos cambios, y se comprometan a seguirlos.

Políticas de Gestión de Datos Científicos

1. Principios y Objetivos

Estas políticas se basan bajo los principios de que los datos primarios son esenciales para los resultados de la investigación realizada en el centro, es por esto que la gestión de estos datos debe realizarse asegurando su integridad para facilitar y dar paso a nuevas investigaciones respetando todos los acuerdos de privacidad, propiedad u otros requerimientos. Esto debe ser promovido en todos los integrantes del centro de investigación.

Este documento establecerá todos los requerimientos relacionados a la gestión de datos de investigación que sean generados en alguna investigación realizada en el Web Intelligence Centre (WIC) de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Los objetivos de esta política son los siguientes:

- Fomentar y establecer una cultura de buenas prácticas de investigación referentes a la gestión de datos científicos generados en el Web Intelligence Centre.
- Asegurar la preservación de los datos generados en el WIC, almacenándola y haciéndola accesible para su uso y re-uso con el fin de continuar las investigaciones realizadas.
- Disminuir los riesgos relacionados a una mala gestión de datos en el WIC.
- Asegurar que los integrantes del centro sean capaces de cumplir con todas las políticas o reglamentos en donde la buena gestión de datos contribuya a la investigación.

2. Alcances

Las Políticas de Gestión de Datos Científicos se aplicarán a todos los integrantes del Web Intelligence Centre o externos que realicen o estén involucrados en alguna investigación o sean parte de algún proyecto del centro.

3. Definiciones

Los siguientes términos y sus definiciones serán las utilizadas por estas políticas en todas sus secciones.

3.1. Datos de Investigación

O datos científicos, es la información recolectada (sin importar la forma o medios en que pueda existir) necesaria para validar las observaciones, conclusiones o resultados de un proyecto de investigación¹.

3.2. Datos Primarios

Datos que son recaudados por primera vez en una investigación en el centro de investigación, es decir, aquellos que fueron recolectados en otra investigación no se consideran como primarios

3.3. Metadatos

Documentación para el conjunto de datos, expresada a través de una sintaxis formal, que guarda toda la información necesaria para su uso posterior. Son adjuntados a los datos de investigación².

3.4. Esquema de Metadatos

Lista de las propiedades de metadatos básicos seleccionados para una identificación precisa y coherente de un recurso para su citación y recuperación³.

3.5. Plan de Manejo de Datos

Documento en donde se muestra cómo serán manejados los datos científicos durante y después de cada proyecto. Debe describir el ciclo de vida para todos los conjuntos de datos que serán recolectados, procesados o generados por el proyecto de investigación⁴.

3.6. Datos Sensibles

Son todos los datos o conjuntos de datos que puedan tener problemas legales, confidencialidad o éticos. Principalmente cuando son datos generados en experimentos realizados con humanos o animales.

3.7. Digital Object Identifier

Es una forma de identificar un objeto digital el cual, aunque cambie de URL seguirá teniendo la misma identificación⁵.

3.8. Consentimiento Informado

Proceso cuyo fundamento es una conversación entre investigador y probando (voluntario). Los puntos esenciales de esta interacción personal son recogidos en un documento firmado por el investigador para ratificar el proceso de información, y por el probando para confirmar que otorga consentimiento para participar en el estudio⁶.

¹ <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/>

² <http://www.bu.edu/datamanagement/outline/elements/metadata/>

³ <https://schema.datacite.org/>

⁴ D.-G. f. R. I. European Commission, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. European Commission, 2016

⁵ <http://www.uchile.cl/portal/informacion-y-bibliotecas/ayudas-y-tutoriales/100282/doi-digital-object-identifier>

⁶ Sugerencias para escribir un consentimiento informado en estudios con personas.

4. Políticas

4.1. Propiedad

- a. Los derechos de todos los datos de investigación generados dentro del marco de algún proyecto o investigación realizada en el centro serán exclusivamente del WIC, a excepción de que existan requerimiento de los financiadores o de la universidad.
- b. En caso de ser datos generados en experimentos con seres humanos, los datos serán de propiedad de ellos, pero los derechos a usar serán exclusivos del centro de investigación. Estos derechos de uso deben ser descritos en los consentimientos informados que firmen al momento de realizar estos experimentos.

4.2. Criterios

- a. Los datos científicos que sean creados a partir de cualquier clase de experimento realizado bajo alguna investigación ejecutada en el centro, deberá ser depositados en el repositorio entregado por el centro sin excepción y siguiendo las guías entregadas por el centro.
- b. En caso de que existan temas legales relacionados a datos generados en experimentos, se deberá guardar de manera digital toda la documentación necesaria para respaldar la toma de los datos, por ejemplo, consentimientos informados para el caso de experimentos realizados con seres humanos, que queda descrito bajo la ley 20.120 de Chile, u otros documentos requeridos por el Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas.
- c. Se debe respaldar todos los datos que sean considerados de interés para futuras investigaciones, siguiendo los mismos protocolos que los datos mencionados en el punto a.

4.3. Infraestructura TI

- a. Todos los datos y documentos que se generen en el marco de alguna investigación realizada en el centro, deberán ser almacenados en el repositorio de cada proyecto entregado por este.
- b. Al momento de subir algún conjunto de datos al repositorio, se deberán rellenar todos los metadatos considerados como mínimos del esquema de metadatos de la guía de metadatos entregada por el centro.
- c. Cuando se desee subir un archivo de datos científicos al repositorio, se debe asegurar que esté en uno de los formatos permitidos por el centro, en caso de no estarlo se debe convertir antes de subirlo.
- d. En caso de querer agregar formatos o metadatos que no han sido considerados a la fecha, estos deberán ser agregados y aprobados por la directora ejecutiva a la configuración del repositorio, siguiendo las guías entregadas por el centro.

- e. En caso de que un integrante del centro requiera acceder a un repositorio de otro proyecto, deberá pedir aprobación por parte del jefe del proyecto al que desea acceder.
- f. La creación de cuentas de integrantes y externos del centro, para el repositorio deberá ser requerida a la directora ejecutiva.
- g. Se deberán utilizar las guías entregadas por el centro para su utilización.

4.4. Plan de Manejo de Datos

- a. Cualquier integrante del centro de investigación que desee realizar una investigación en él, deberá entregar dos versiones de un plan de manejo de datos al jefe de proyecto, uno antes de iniciar cualquier tipo de experimento referente a estas investigaciones y otro una vez finalizada esta investigación.
- b. La primera versión deberá contener los siguientes puntos: nombre del trabajo que se está realizando, nombre del proyecto del centro a cuál pertenece, fecha de creación del documento, nombre del encargado del proyecto en el WIC, código y fuente de financiamiento del trabajo, descripción del proyecto y de la investigación o experimentos a realizar, datos del investigador (nombre y contacto), políticas que deban cumplir en el centro, descripción de los datos que se espera generar en el proyecto (incluyendo formatos, tamaños esperados de archivos y cómo serán adquiridos y procesados en caso de contar con esta última información, junto a las fechas estimadas de los experimentos que se realicen o extraigan datos). Para su diseño se usarán los estándares acordados en el anexo A de este documento.
- c. La segunda versión del plan de manejo de datos deberá contener los puntos mencionados en a) cambiando la información de los datos esperados a generar y fechas estimativas de experimentos en caso de realizarse, por los resultados y fechas reales de estos, además de las metodologías y algoritmos aplicados a los datos generados, las notas de laboratorio si es que se tienen, dónde y en qué formatos serán almacenados junto a los metadatos que contendrán. En caso de que se generen datos sensibles qué documentos se guardarán referentes a temas legales o requerimientos de la universidad o financiadores y definir quién será el dueño de los datos mencionados en el documento.
- d. Para su diseño se usarán los estándares acordados en las guías y *templates* otorgados por el centro.
- e. Las dos versiones una vez aprobadas deberán ser almacenadas en el repositorio otorgado por el centro, de manera digital.
- f. Debe ser aprobado por el Jefe del Proyecto en donde se esté realizando la investigación.

4.5. Requerimientos Externos

- a. En caso de que exista algún requerimiento externo, ya sea a nivel institucional o de parte del financiador, se deben respetar estos requisitos con prioridad a las políticas de gestión de datos.
- b. En caso de que esto suceda, al momento de realizar la revisión periódica deberá ser analizado si es necesario agregar estos requerimientos a las políticas para futuras versiones.

4.6. Re-uso y Descubrimiento

- a. Los datos de investigación generados deberán ser almacenados por un período mínimo de 10 años, utilizando los formatos y metadatos especificados en el plan de manejo de datos, para los metadatos ver Anexo B.
- b. Los derechos de acceso y re-uso no se podrán otorgar a publicadores comerciales sin retener el derecho a abrir los datos de investigación.
- c. El acceso a los datos de investigación para investigaciones que están en curso sólo será disponible para integrantes actuales del centro.
- d. El acceso a los datos de investigación para investigaciones finalizadas, si es que no se han declarado como abiertos, deberá ser exclusivo para integrantes actuales del centro de investigación y podrá ser realizado desde el momento en que se almacenan por primera vez hasta la fecha estipulada de retención.
- e. Solo se podrán abrir los datos en donde no puedan existir problemas legales, de privacidad, confidencialidad o éticos.
- f. En caso de abrir los datos se deberá especificar la existencia de tiempos de privilegio de acceso ellos, el cual deberá ser definido por el jefe del proyecto.
- g. En caso de abrir los datos de investigación se deberá especificar la forma en que se citará al centro y al autor de estos datos en los lugares que se publiquen.
- h. Si se desarrollan publicaciones referentes a cierta investigación que generó datos científicos y estos son necesarios para su validación, se deberá adjuntar información para acceder a los datos (por ejemplo, un link o en caso de que exista un Digital Object Identifier (DOI)) en la publicación para acceder a los datos utilizados en ella.
- i. Si alguna entidad externa requiere acceso a los datos generados, se deberá aprobar por parte de la directora ejecutiva su posible acceso. No se dará acceso cuando se puedan tener problemas legales o éticos.

4.7. Ética

- a. Todos los integrantes del centro de investigación deberán realizar sus investigaciones cumpliendo todas las leyes, requerimientos del centro, universidad y financiadores, relacionados a la generación de datos sensibles, especialmente los referentes
- b. Toda investigación realizada en humanos deberá cumplir lo estipulado en las leyes 19.628 y 20.120 de Chile, sumado a esto, las exigencias del Comité de Ética de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- c. No se podrán abrir conjunto de datos sensibles para evitar el riesgo al no cumplimiento de alguno de los puntos a. o b.
- d. Toda documentación generada para la realización de experimentos con humanos, exigidos por la ley, Comité de Ética o financiadores, deberá ser digitalizada y almacenada junto a los datos en el repositorio del centro.
- e. Los datos generados en experimentos con humanos son de propiedad de estos, el centro de investigación sólo cuenta con los derechos a utilizarlos en sus investigaciones, esto debe quedar plasmado en los documentos exigidos por las entidades mencionadas en los puntos anteriores.
- f. En caso de considerarlo necesario, algunos conjuntos de datos necesitarán de un documento de permiso de acceso a estos que será firmado por cada persona que desee acceso a ellos y contendrá las reglas específicas que se deberán cumplir para ellos. Este documento será entregado por el centro y deberá ser aprobado por los jefes de proyecto, además su versión digital deberá ser almacenada en el repositorio del proyecto.

4.8. Eliminación, destrucción y movimiento de registros.

- a. Una vez pasado el período de retención de algún conjunto de datos, se deberá reevaluar su almacenamiento por un nuevo período.
- b. En caso de que se desee eliminar registros del repositorio del centro, los jefes de proyectos deberán realizarlo.
- c. En caso de que se deseen mover registros o abrir los datos en un repositorio externo, se deberá aprobar por parte de la directora ejecutiva y se tendrá que ser llevado a cabo por el jefe de proyecto, siguiendo todas las normas de formatos y metadatos necesarios para su re-uso y descubrimiento.
- d. En caso de un cambio de repositorio, el traslado al nuevo lugar de almacenamiento deberá ser realizado en conjunto por todos los jefes de proyectos y la directora ejecutiva, asegurando los estándares acordados de formatos y metadatos.

5. Roles y Responsabilidades

Los roles y responsabilidades serán los siguientes:

5.1. El Web Intelligence Centre tiene la responsabilidad de:

- a. Proporcionar infraestructura de almacenamiento segura para los datos de investigación generados en el mismo centro.
- b. Entregar facilidades para el almacenamiento, disponibilidad y accesibilidad a conjuntos de datos generados en él para incentivar su re-uso dentro y fuera de este.
- c. Especificar quién es dueño de los datos científicos generados en investigaciones realizadas en el centro.
- d. Dar soporte referente a la gestión de datos científicos a todos los integrantes del centro.
- e. Dar a conocer las políticas del centro a todos sus integrantes.
- f. Realizar una revisión periódica de las políticas con el fin de que no queden obsoletas en el tiempo.
- g. Ser el responsable por los costos relacionados a la gestión de datos científicos en el centro.

5.2. El director del centro tiene el deber de:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Aprobar los documentos de compromiso por parte de los integrantes del centro acerca del cumplimiento de estas políticas.

5.3. La directora ejecutiva deberá:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Aprobar traslados o destrucción de datos de investigación guardados en el repositorio del centro en donde no haya terminado el periodo de retención de estos.
- c. Analizar, una vez terminado el periodo de retención de ciertos datos, si se mantendrán o se eliminarán los sets de datos del repositorio utilizado por el centro, en el primer caso además deberá especificar por cuánto tiempo más seguirán almacenados, en el segundo caso deberá ser la encargada de eliminarlos.
- d. Agregar y aprobar nuevos formatos y metadatos a la configuración del repositorio en caso de ser requerido, crear las cuentas del repositorio para los nuevos integrantes y externos que requieran acceso al este.

5.4. Los jefes de proyecto deberán:

- a. Aprobar cualquier cambio en Políticas de Gestión de Datos Científicos.
- b. Supervisar a los miembros de los proyectos sobre el cumplimiento de las políticas.

- c. Asegurar que todos los integrantes del proyecto hayan firmado el acuerdo de cumplimiento de las políticas.
- d. En caso de realizar algún tipo de investigación también deberán cumplir los deberes de los demás integrantes del proyecto en donde la realice.
- e. Una vez terminado el proyecto entregar un reporte de cumplimiento de las políticas dentro del proyecto.
- f. Aprobar los planes de manejo de datos entregados por los integrantes del proyecto que quieran realizar una investigación en él.
- g. Evaluar los conjuntos de datos generados en el proyecto respecto a si se necesita un acceso especial debido a su naturaleza sensible o no, en caso de ser necesario deberá aprobar todos los documentos de requerimiento de permiso a ellos que se necesiten.
- h. Aprobar y generar los accesos a los integrantes del centro que requieran permisos a su proyecto.

5.5. Los integrantes de los proyectos tienen la responsabilidad de:

- a. Firmar un acuerdo en donde se especifica que se tiene conocimiento y se realizará el cumplimiento de las políticas del centro.
- b. Identificar qué datos y documentos serán los que se deban respaldar para su posterior re-uso.
- c. Resguardar los datos que generen en sus investigaciones bajo los requerimientos de las políticas del centro.
- d. Asegurar la integridad de los datos durante todo el ciclo de vida de estos.
- e. Almacenar los datos de investigación en formatos y con lo metadatos suficientes para asegurar su fácil re-uso y descubrimiento.
- f. Realizar las acciones contenidas en el documento de acuerdo a los protocolos éticos y legislativos respectivos.
- g. Leer y utilizar las guías entregadas por el centro referentes al uso del repositorio del centro.

6. Aprobación y Revisión Periódica

Este documento será sujeto a una revisión periódica de al menos una vez al año, con el fin de ir mejorando las políticas y evitar su obsolescencia con el paso del tiempo. Una vez hecha la revisión se deberán actualizar en menos de 2 meses una vez pasada la revisión.

Todos los cambios que se produzcan deberán ser primero revisados y aprobados por los jefes de cada proyecto y la directora ejecutiva para finalmente ser aprobadas por el director o encargado del centro.

Si se realizan cambios se deberá firmar un nuevo acuerdo de cumplimiento por parte de los integrantes del centro con el fin de que todos sean conscientes de estos cambios, y se comprometan a seguirlos

Anexos D

Guías y Documentos

A continuación se adjuntan las guías y documentos creados.

Web Intelligence Centre, Universidad de Chile

Guía de Licencias

Versión 1.0

Web Intelligence Centre
15-10-2016

Índice

Introducción	1
Objetivo	1
Contenido	2
Condiciones	3
BY: Reconocimiento (Attribution)	3
NC: No Comercial (Non Commercial)	3
ND: Sin Derivados (No Derivatives)	3
SA: Compartir Igual (Share Alike)	4
CC0: Sin derechos de autor	4
Licencias.....	4
Licencia CC0	4
Licencia CC – BY	5
Licencia CC-BY SA.....	5
La licencia CC BY-ND	6
Licencia CC BY-NC	6
Licencia CC BY-NC-SA.....	7
Licencia CC BY-NC-ND.....	7
Copyright	8

Introducción

En el contexto de los datos científicos, las licencias permiten especificar los términos de usos que se quiera imponer a los terceros que quieran reutilizar los datos.

Para facilitar esto, existen licencias estandarizadas a nivel internacional. Una de estas es Creative Commons (CC), la cual es la utilizada por el Web Intelligence Centre.

CC es una organización sin fines de lucro que permite el reuso y que se comparta (*sharing*) la creatividad y conocimiento a través de la entrega de herramientas legales de manera gratuita.

Generalmente estas licencias se utilizan en las distintas publicaciones de investigaciones o libros, para esto utilizan la declaración de la forma presentada y en ciertas ocasiones la imagen de la licencia, lo cual es recomendado por CC. Desde su versión 4.0 permite aplicar una licencia a conjuntos de datos¹.

Objetivo

El objetivo de este documento es dar a conocer las posibles licencias a aplicar a los datos generados en el Web Intelligence Centre. La licencia escogida será la asignada a los metadatos de los archivos que serán agregados al repositorio del centro en la propiedad "Rights".

¹ <https://creativecommons.org/faq>

Contenido

Para asignar una licencia de Creative Commons se deben realizar dos pasos:

El primer paso es al momento de subir un nuevo archivo al repositorio, en la tercera de esta acción se deberán rellenar dos campos como se muestra en la siguiente imagen:



Submit: Describe this Item ?

Please fill further information about this submission below.

Debe escribir el nombre completo. Ejemplo: Creative Commons Attribution 4.0 International License (CC BY 4.0). Para el centro se utilizarán solo licencias de Creative Commons y copyright, para más información de estas revisar la Guía de Licencias del centro.

Rights

El URI de la licencia. <https://creativecommons.org/licenses/by/4.0/>

rightsUri

El contenido con que debe rellenar estos dos campos (Rights y rightsURI) se encuentran en este documento detallado por cada una de las 8 opciones disponibles por el centro, debe asignar el tipo que se crea conveniente utilizar. Debe rellenar el nombre completo de la licencia y no su abreviación. En segundo lugar, en la sexta etapa de la misma acción se debe seleccionar la licencia con la que se rellenó en la etapa 3. Para esto saldrá un menú desplegable en "License Type" con algunas opciones como se muestra en la siguiente imagen:



Submit: Use a Creative Commons License

If you wish, you may add a [Creative Commons License](#) to your item. Creative Commons licenses govern what people who read your work may then do with it.

License Type

< Previous

Cancel/Save

Next >

La opción que debe seleccionar en "License Type" está descrita en este documento en la sección de licencias

El segundo paso es al momento de crear el archivo XML de metadatos que debe adjuntar durante o después de subir el archivo de datos. Se deben agregar los metadatos de la propiedad de **Rights** y su subpropiedad **rightsURI**, esta última es importante ya que redirige al código legal de la licencia. Los valores que se deben agregar son los mismos que utilizó en el primer paso.

Otra opción es agregar una declaración en la página donde se descargue el archivo de esta forma:

This data is available under the "nombre de la licencia" whose full text can be found at: "URI de la licencia"

Para el caso del centro, se mostrará en estas páginas la licencia que se agregue a los datos, por lo que sólo es necesario agregar la licencia a la hora de subir un archivo al repositorio en la etapa de metadatos.

Los datos sensibles para el caso de las licencias son un caso especial. Se habla de datos sensibles cuando contienen información privada que permita identificar a personas o porque debido a temas legales estos no pueden ser publicados, en general esto pasa cuando se realizan experimentos con seres humanos. Asignar una licencia permisiva o de acceso abierto no significa que estos puedan ser compartidos, es por esto que para cuando existan estos casos en el centro de investigación, la propiedad Rights deberá rellenarse con Copyright.

Para los otros casos, se recomienda asignar licencias permisivas para fomentar su re-uso. Esta decisión debe ser aprobada por los jefes de proyecto al momento de revisar las segundas versiones de los planes de gestión de datos de los integrantes, en donde se deben describir los metadatos que contendrán los datos subidos al repositorio.

En caso de querer utilizar algún recurso externo que cuente con alguna licencia de Creative Commons, la organización cuenta con una sección de preguntas frecuentes² en donde estas interrogantes son respondidas, además de entregar más información sobre las licencias y sus posibles usos.

Condiciones

Existen 5 tipo de condiciones que se aplican a los conjuntos de datos, estas son:

BY: Reconocimiento (Attribution)



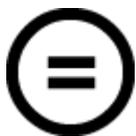
Obliga a dar atribución o el crédito apropiado enlazando a la página de la licencia.

NC: No Comercial (Non Commercial)



Prohíbe el uso del material para fines comerciales.

ND: Sin Derivados (No Derivatives)



Prohíbe compartir adaptaciones del material.

² <https://creativecommons.org/faq>

SA: Compartir Igual (Share Alike)



Requiere que adaptaciones del material sean compartidos con el mismo tipo de licencia de este.

CC0: Sin derechos de autor



Elimina todo derecho o limitaciones para potenciales usuarios del material.

Licencias

Con estas condiciones, existen siete tipos de licencia que resultan de sus combinaciones³, que van desde lo más abierto (color verde) a lo más restringido, sumado a esto está el caso extremo en donde están todos los derechos reservados (rojo) el cual no es considerado dentro de las licencias de Creative Commons, el espectro se encuentra en la figura 1.

Cada una de estas licencias obliga a los futuros usuarios a seguir distintas condiciones para usar los datos. Comenzando desde lo menos restrictivo, los dos primeros tipos de licencia:

Licencia CC0



Elimina todos los derechos y copyright de los datos, es decir se pueden usar libremente por cualquier tercero.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons CC0 1.0 Universal License

rightsURI: <https://creativecommons.org/publicdomain/zero/1.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”: **CC0**.

³ <https://creativecommons.org/licenses/>

Licencia CC – BY



Obliga a dar atribución o el crédito apropiado enlazando a la página de la licencia.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution 4.0 International License (CC BY 4.0)

rightsURI: <https://creativecommons.org/licenses/by/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”:
Creative Commons.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **Yes**
- Allow modifications of your work? A lo cual se debe responder **Yes**

Con esto hecho, se asignará la licencia CC-BY al archivo en el repositorio.

Licencia CC-BY SA



Permite generar nuevas obras a partir del material original sin restricciones, siempre y cuando estas creaciones estén bajo la misma licencia y se atribuya al autor del material utilizado.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

rightsURI: <https://creativecommons.org/licenses/by-sa/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”:
Creative Commons.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **Yes**
- Allow modifications of your work? A lo cual se debe responder **ShareAlike**

Con esto hecho, se asignará la licencia CC-BY-SA al archivo en el repositorio.

La licencia CC BY-ND



Mientras se reconozca la autoría de los datos, posibilita su redistribución de cualquier tipo siempre y cuando la obra no se modifique y sea distribuida en su totalidad.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution-NoDerivates 4.0 International License (CC BY-ND 4.0)

rightsURI: <https://creativecommons.org/licenses/by-nd/4.0/>

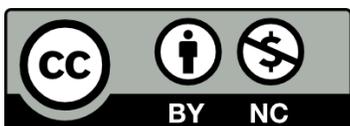
Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type” Creative Commons.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder Yes
- Allow modifications of your work? A lo cual se debe responder No

Con esto hecho, se asignará la licencia CC-BY-ND al archivo en el repositorio.

Licencia CC BY-NC



Permite crear nuevos conjuntos de datos o darle distintos usos siempre y cuando sea sin fines comerciales, además de reconocer la autoría de estos.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

rightsURI: <https://creativecommons.org/licenses/by-nc/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”: **Creative Commons.**

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **No**
- Allow modifications of your work? A lo cual se debe responder **Yes**

Con esto hecho, se asignará la licencia CC-BY-NC al archivo en el repositorio.

Licencia CC BY-NC-SA



Es igual a la anterior solo que las nuevas creaciones deben estar bajo la misma licencia que el set original.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)

rightsURI: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”:
Creative Commons.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **No**
- Allow modifications of your work? A lo cual se debe responder **ShareAlike**

Con esto hecho, se asignará la licencia CC-BY-NC-SA al archivo en el repositorio.

Licencia CC BY-NC-ND



Es la más restrictiva de las anteriores y sólo permite descargar y compartir el material con otras personas, siempre que se atribuya el autor, pero sin poder manipularlas o usarlas con fines comerciales.

Las propiedades Rights y rightsURI son las siguientes:

Rights: Creative Commons Attribution-NonCommercial-NoDerivates 4.0 International License (CC BY-NC-ND 4.0)

rightsURI: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”:
Creative Commons.

En ese momento aparecerán dos preguntas:

- Allow comercial uses of your work? A lo cual se debe responder **No**
- Allow modifications of your work? A lo cual se debe responder **No**

Con esto hecho, se asignará la licencia CC-BY-NC-SA al archivo en el repositorio.

Copyright

Mantiene todos los derechos de los datos, que en general serán todos los casos en que los datos nunca serán publicados (disponibles para externos) debido a temas legales o de confidencialidad.

Las propiedades Rights y rightsURI son las siguientes:

Rights: © Web Intelligence Centre “año actual”, Todos los derechos reservados.

No se debe rellenar la propiedad rightsURI.

Al momento de subir un archivo al repositorio, en la etapa “License” se debe seleccionar en “License Type”: No **Creative Commons License**. En caso de no asignar ningún tipo de licencia, automáticamente se reservan todos los derechos, se muestra el siguiente mensaje en la página de cada ítem:

Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated.

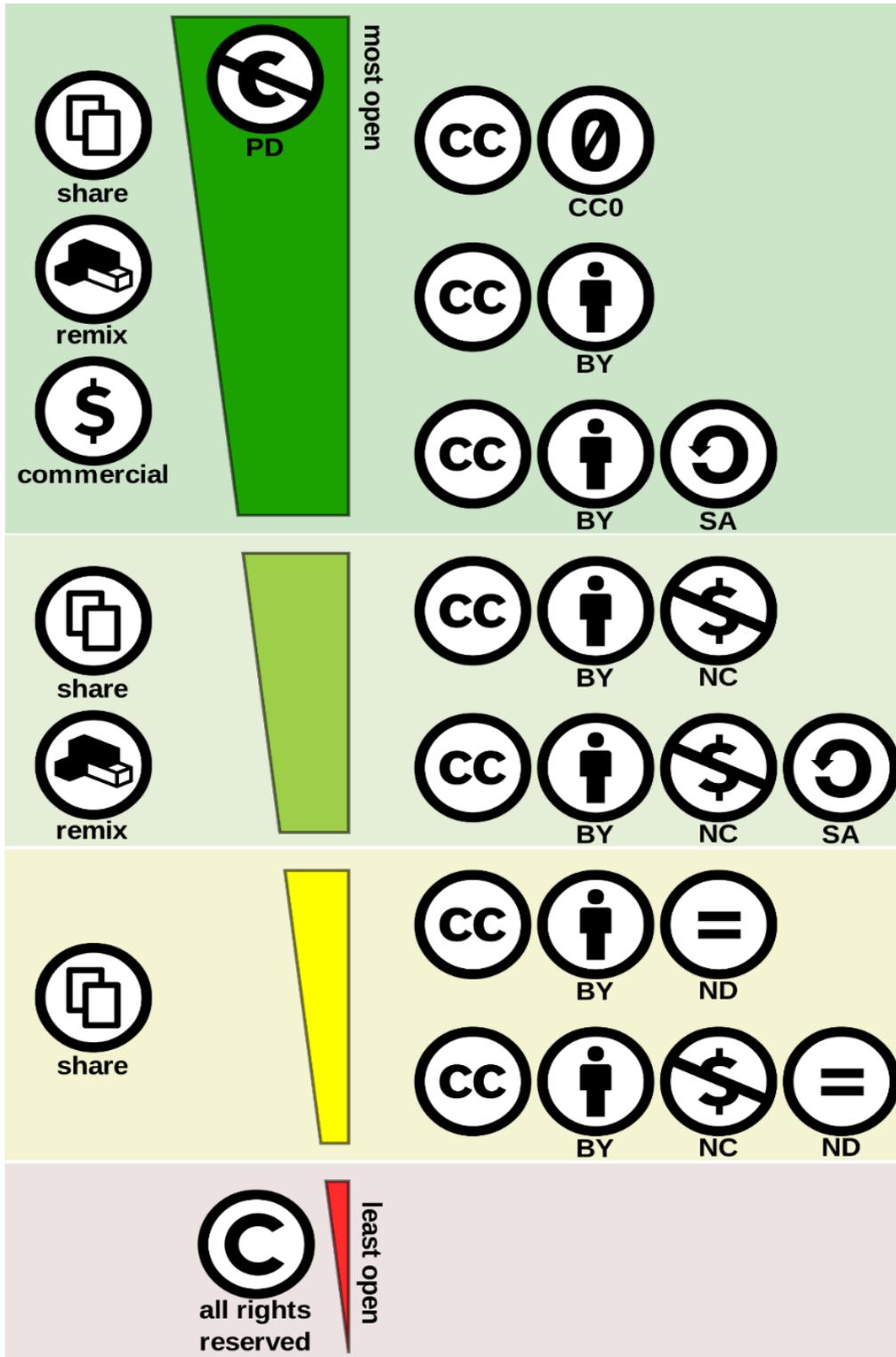


Figura 1: Espectro licencias CC

Web Intelligence Centre, Universidad de Chile

Guía de Metadatos

Versión 1.0

Web Intelligence Centre
15-10-2016

Índice

Introducción	1
Objetivo	1
Propiedades.....	2
Identifier	3
Ejemplos	3
1. Identifier	3
1.1. IdentifierType	3
Creator.....	4
Ejemplos	4
2. Creator.....	4
2.1. creatorName.....	4
2.1.1. familyName	5
2.1.2. givenName.....	5
2.2. nameIdentifier	5
2.2.1. nameIdentifierScheme	5
2.2.2. schemeURI.....	5
2.3. affiliation.....	5
Title.....	6
Ejemplos	6
3. Title.....	6
3.1. titleType.....	6
Publisher.....	7
Ejemplos	7
4. Publisher	7
PublicationYear.....	7
Ejemplos	7
5. PublicationYear.....	7
Subject	8
Ejemplos	8
6. Subject.....	8
6.1. subjectScheme.....	8

6.2.	schemeURI.....	9
6.3.	valueURI.....	9
Contributor		9
Ejemplos		9
7.	Contributor	9
7.1.	contributorType.....	10
7.2.	contributorName.....	10
7.2.1.	familyName	10
7.2.2.	givenName.....	10
7.3.	nameIdentifier	10
7.3.1.	nameIdentifierScheme	10
7.3.2.	schemeUri.....	11
7.4.	affiliation.....	11
Language.....		11
Ejemplos		11
8.	Language.....	11
Date		12
Ejemplos		12
9.	Date	12
9.1.	dateType.....	12
ResourceType		13
Ejemplos		13
10.	ResourceType	13
10.1.	ResourceTypeGeneral	13
AlternativeIdentifier		14
Ejemplos		14
11.	AlternativeIdentifier	14
11.1.	alternativeIdentifierType.....	14
RelatedIdentifier.....		15
Ejemplos		15
12.	RelatedIdentifier.....	15
12.1.	relatedIdentifierType.....	15
12.2.	relationType.....	16

12.3.	relatedMetadataScheme.....	16
12.4.	schemeURI.....	16
12.5.	schemeType.....	16
	Size.....	17
	Ejemplos.....	17
13.	Size.....	17
	Format.....	17
	Ejemplos.....	17
14.	Format.....	17
	Version.....	18
	Ejemplos.....	18
15.	Version.....	18
	Rights.....	18
	Ejemplos.....	18
16.	Rights.....	18
16.1.	rightsURI.....	19
	Description.....	19
	Ejemplos.....	19
17.	Description.....	19
17.1.	descriptionType.....	19
	GeoLocation.....	20
	Ejemplos.....	20
18.	GeoLocation.....	21
18.1.	geoLocationPoint.....	21
18.1.1.	pointLongitude.....	21
18.1.2.	pointLatitude.....	21
18.2.	geoLocationBox.....	21
18.2.1.	westBoundLongitude.....	21
18.2.2.	eastBoundLongitude.....	22
18.2.3.	southBoundLatitude.....	22
18.2.4.	northBoundLatitude.....	22
18.3.	geoLocationPlace.....	22
18.4.	geoLocationPolygon.....	22

18.4.1. polygonPoint.....	23
18.4.1.1. pointLongitude	23
18.4.1.2. pointLatitude	23
FundingReference	24
Ejemplos	24
19. FundingReference.....	24
19.1. funderName.....	24
19.2. funderIdentifier	24
19.2.1. funderIdentifierType	25
19.3. awardNumber.....	25
19.3.1. awardURI	25
19.4. awardTitle.....	25
Anexos	26
Ejemplo completo archivo XML.....	26
Formatos.....	28

Introducción

En el contexto de los datos científicos, los metadatos son la documentación para el conjunto de datos que guardan toda la información necesaria para su uso posterior.

La definición más común para metadatos es “datos acerca de datos”. A pesar de esto un metadato es en sí un dato, puede ser almacenado y manejado en una base de datos. Un metadato es un dato que es usado para describir a otro, es decir, su uso lo hace un metadato¹.

En este documento se presentarán el listado metadatos o propiedades del esquema de DataCite en su versión 4.0, los cuales tienen el fin de proveer una identificación precisa de un recurso para propósitos de citación y recuperación o descubrimiento.

Algunas propiedades tienen subpropiedades y algunas de estas subsubpropiedades, las cuales pueden ser opcionales u obligatorias dependiendo cada caso.

Existen algunas que son opcionales, pero altamente recomendadas. Para el caso del Web Intelligence Centre será requisito utilizar como mínimo las propiedades que sean consideradas como obligatorias por DataCite (por el momento exceptuando Identifier) y además “Rights” y “FunderReference”. Por otro lado, se recomienda rellenar la mayor cantidad de metadatos posibles, sean opcionales o no.

Objetivo

El objetivo de este documento es dar a conocer las propiedades del esquema de DataCite en su versión 4.0 para facilitar su comprensión y utilización por parte de los integrantes del Web Intelligence Centre. Se presentará cada propiedad y a sus derivados junto a ejemplos en lenguaje XLM, definiciones y en algunos casos los valores posibles para la propiedad.

¹ B. E. Bargmeyer and D. W. Gillman, “Metadata standards and metadata registries: An over-view,” in International Conference on Establishment Surveys II, vol. 19, 2000

Propiedades

Para las propiedades que tienen una lista de valores aceptados, estos valores son explicados con mayor detalle en los apéndices de la documentación del esquema de DataCite² (página 27 en adelante), disponible en su página web³ al igual que alguno de los ejemplos y las traducciones realizadas para las definiciones de cada propiedad.

Se deberán agregar metadatos en dos partes:

La primera es al momento de subir un archivo, donde el repositorio requiere de algunos datos para su realización. A pesar de no utilizar directamente el esquema de DataCite, algunas de las propiedades son homólogas, solo que a veces con distintos nombres.

Por esto, se deberán rellenar los siguientes metadatos en la primera etapa de subir archivos al repositorio: Authors (que sería Creator en el esquema DataCite), Title, Publisher (que tienen el mismo nombre), Date of Issue (que sería PublicationYear en DataCite) y Citation, el cual tendrá que tener el siguiente formato: Creator (PublicationYear): Tittle. Version. Publisher. ResourceType. (esto utilizando las definiciones del esquema de DataCite, las cuales se encuentran en este documento).

Luego en la tercera etapa se deberán rellenar todas las propiedades que se encuentran en ella, es decir “CC License” (Rights), CC License URI (rightsURI), Funding Information (FunderReference), Funder Name (funderName), Funder Identifier (funderIdentifier), Proyect Name (awardTitle) y Proyect Code (awardURI), donde los nombres entre paréntesis representan las propiedades o subpropiedades del esquema de DataCite.

A pesar de que sólo las mencionadas son obligatorias, se recomienda también rellenar los demás campos en las tres primeras etapas de agregar archivos del repositorio. Estos metadatos serán mostrados en la información del archivo en el repositorio, pero no son necesariamente iguales a los pedidos por el centro, los cuales se deberán rellenar en la segunda parte de agregar metadatos:

Esta segunda parte, se puede realizar al momento de subir los archivos o posterior a esto. Esta parte consiste en crear un archivo XML siguiendo como ejemplo el Anexo Ejemplo completo archivo XML de este documento. Aquí se tendrán que rellenar como mínimo las propiedades Creator, Title, Publisher, PublicationYear, ResourceType, Rights y FunderReference. Se recomienda rellenar la mayor cantidad de metadatos en este archivo, recordar que no son necesariamente las mismas propiedades que se muestran en el repositorio a la hora de subir archivos, estas son exclusivamente utilizando el esquema de DataCite en su versión 4.0.

Una vez creado este archivo, se puede adjuntar junto al archivo que se desea agregar al repositorio en la etapa de seleccionar archivos a subir. La segunda opción es una vez subidos los datos, ingresar a la página (ítem) que se generó y seleccionar EDIT en el menú izquierdo, al final de la página saldrá un botón verde con la opción Add

² DataCite Metadata Working Group. (2016). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. DataCite e.V.
<http://doi.org/10.5438/0012>.

³ <https://schema.datacite.org/meta/kernel-4.0/>

Bitstream, donde se seleccionará este archivo, el cual debe tener el nombre **metadata-datacite.xml (formato XML)**

A continuación, se presentarán las propiedades del esquema de metadatos de Datacite en su versión 4.0, junto a su descripción. Cuando se muestra una subpropiedad obligatoria, es sólo en el caso de que la propiedad principal sea declarada.

Identifier

Ejemplos

```
<identifier identifierType="DOI">10.5072/example-full</identifier>
```

1. Identifier

Restricción: Es una propiedad **opcional (por ahora)** y debe ser única.

Definición: Es un string único que identifica el recurso.

Al utilizar el esquema de DataCite, el identificador debe ser un DOI (Digital Object Identifier) registrado.

El formato debe ser "10.1234/foo". Donde foo es una combinación de letras.

1.1. IdentifierType

Restricción: Es una propiedad **obligatoria** y debe ser única.

Definición: Es el tipo del identificador.

En este caso el único valor aceptado por DataCite es DOI. Existen otros tipos de identificadores, pero al utilizar el esquema de DataCite se hace obligatorio utilizar los DOIs.

Creator

Ejemplos

```
<creators>
  <creator>
    <creatorName>Maldonado, Felipe</creatorName>
  </creator>
</creators>
```

```
<creators>
  <creator>
    <creatorName>Sarmiento, Matias</creatorName>
  </creator>
  <creator>
    <creatorName>Webber, John</creatorName>
  </creator>
  <creator>
    <creatorName>Díaz, Sebastian</creatorName>
  </creator>
</creators>
```

```
<creators>
  <creator>
    <creatorName>Miller, Elizabeth</creatorName>
    <givenName>Elizabeth</givenName>
    <familyName>Miller</familyName>
    <nameIdentifier schemeURI=http://orcid.org/
      0000-0001-5000-007 nameIdentifierScheme="ORCID">
    </nameIdentifier>
    <affiliation>DataCite</affiliation>
  </creator>
</creators>
```

2. Creator

Restricción: Es una propiedad **obligatoria**, puede haber múltiples registros.

Definición: Investigadores involucrados en la producción de los datos, o en caso de ser una publicación, los autores de esta en orden de prioridad.

Pueden ser nombres de corporaciones, instituciones o personas.

2.1. creatorName

Restricción: Es una propiedad **obligatoria**, única para cada "creator".

Definición: Nombre del creador.

El formato es apellido, nombre (Ejemplo: Maldonado, Felipe)

2.1.1. familyName

Restricción: Es una propiedad **opcional pero recomendada**, única para cada “creator”.

Definición: Apellido del creador.

Ejemplo: Maldonado

2.1.2. givenName

Restricción: Es una propiedad **opcional pero recomendada**, única para cada “Creator”.

Definición: Nombre del creador.

Ejemplo: Felipe

2.2. nameIdentifier

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples nameIdentifier para cada “Creator”.

Definición: Identificador único de una entidad legal o individual, de acuerdo a varios esquemas.

El formato depende del esquema

2.2.1. nameIdentifierScheme

Restricción: Es una propiedad **obligatoria** y única para cada nameIdentifier agregado (por cada “Creator”).

Definición: El nombre del esquema de identificador de nombres.

Ejemplo: ORCID, ISNI

2.2.2. schemeURI

Restricción: Es una propiedad opcional y única para cada nameIdentifier agregado (por cada “Creator”).

Definición: El URI del esquema de identificador de nombres.

Ejemplo: <http://www.isni.org>

<http://orcid.org>

2.3. affiliation

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples registros por cada “Creator”.

Definición: La organización o institución a la cual el creador está afiliado.

Texto Libre.

Ejemplo: DataCite; Universidad de Chile;

Title

Ejemplos

```
<titles>
  <title>Właściwości rzutowań podprzestrzeniowych</title>
  <title titleType="TranslatedTitle">
    Translation of Polish titles
  </title>
</titles>
```

```
<titles>
  <title>Full DataCite XML Example</title>
  <title titleType="Subtitle">
    Demonstration of DataCite Properties.
  </title>
</titles>
```

3. Title

Restricción: Es una propiedad **obligatoria**, pueden agregarse uno o múltiples registros.

Definición: Nombre o título para el cual el recurso es conocido.

Texto Libre.

3.1. titleType

Restricción: Es una propiedad opcional y única por cada "Title" que se agregue.

Definición: El tipo del "Title".

Lista de valores aceptados:

- AlternativeTittle
- Subtitle
- TranslatedTitle
- Other

Publisher

Ejemplos

```
<publisher>Purdue University Research Repository (PURR)</publisher>
```

4. Publisher

Restricción: Es una propiedad **obligatoria** y única.

Definición: El nombre de la entidad que retiene, archiva, publica impresiones, distribuye, libera o produce el recurso. Esta propiedad es usada para formular la citación del set del recurso.

Ejemplos:

World Data Center for Climate (WDCC);

GeoForschungsZentrum Potsdam (GFZ);

Geologica Institute, University of Tokyo;

PublicationYear

Ejemplos

```
<publicationYear>2014</publicationYear>
```

5. PublicationYear

Restricción: Es una propiedad **obligatoria** y única.

Definición: El año en que los datos fueron o serán hechos públicamente disponibles.

Formato: YYYY

Si hay un periodo de embargo, usar la fecha cuando el embargo termine.

Para el caso de los datasets, "publish" se refiere a cuando se harán disponibles los datos a la comunidad de investigadores.

Si no existe un valor del año de publicación estándar, usar la fecha que se prefiera de una perspectiva para la citación.

Si la fecha de publicación no puede ser determinada, utilizar la fecha de registro del set de datos.

Subject

Ejemplos

```
<subjects>
  <subject subjectScheme="DDC">
    830 German & related literatures
  </subject>
  <subject>Polish Literature</subject>
</subjects>
```

```
<subjects>
  <subject>Assessment</subject>
  <subject>Information Literacy</subject>
  <subject>Engineering</subject>
  <subject>Undergraduate Students</subject>
  <subject>CELT</subject>
  <subject>Purdue University</subject>
</subjects>
```

```
<subjects>
  <subject xml:lang="en-us" schemeURI=http://dewey.info/
  subjectScheme="dewey">
    000 computer science
  </subject>
</subjects>
```

6. Subject

Restricción: Es una propiedad **opcional pero recomendada**. Puede haber múltiples entradas.

Definición: Tema, keywords, código de clasificación, frase clave que describe el recurso.

Texto Libre.

6.1. subjectScheme

Restricción: Es una propiedad opcional y única por cada "Subject" que se agregue.

Definición: El nombre del esquema del subject o código de clasificación o autoridad si alguno es utilizado.

Texto Libre.

6.2. schemeURI

Restricción: Es una propiedad opcional y única por cada "Subject" que se agregue.

Definición: El URI del esquema de identificador de tema.

Ejemplos:

<http://id.loc.gov/authorities/subjects>

<http://udcdata.info/>

6.3. valueURI

Restricción: Es una propiedad opcional y única por cada "Subject" que se agregue.

Definición: el URI del término del subject.

Ejemplos:

<http://id.loc.gov/authorities/subjects/sh85026196>

<http://udcdata.info/037278>

Contributor

Ejemplos

```
<contributors>
  <contributor contributorType="ProjectLeader">
    <contributorName>Starr, Joan</contributorName>
    <nameIdentifier schemeURI=http://orcid.org/
      nameIdentifierScheme="ORCID">
      0000-0002-7285-027X
    </nameIdentifier>
    <affiliation>California Digital Library</affiliation>
  </contributor>
</contributors>

</contributors>
  <contributor contributorType="DataCollector">
    <contributorName>Doe, John</contributorName>
    <nameIdentifier nameIdentifierScheme="ORCID"
      schemeURI="http://orcid.org/">
      0000-0001-5393-1421
    </nameIdentifier>
  </contributor>
</contributors>
```

7. Contributor

Restricción: Es una propiedad **opcional pero recomendada**, pueden existir múltiples registros.

Definición: La institución o persona responsable de coleccionar, gestionar, distribuir o contribuir de alguna otra manera al desarrollo del recurso.

7.1. contributorType

Restricción: Es una propiedad **obligatoria**, única para cada “Contributor”.

Definición: Tipo del contribuidor.

Lista de valores aceptados:

ContactPerson, DataCollector, DataCurator, DataManager, Distributor, Editor, HostingInstitution, Producer, ProjectLeader, ProjectManager, ProjectMember, RegistrationAgency, RegistrationAuthority, RelatedPerson, Researcher, ResearchGroup, RightsHolder, Sponsor, Supervisor, WorkPackageLeader, Other

7.2. contributorName

Restricción: Es una propiedad **obligatoria**, única para cada “Contributor”.

Definición: Nombre del contribuidor.

El formato es apellido, nombre (Ej: Maldonado, Felipe)

7.2.1. familyName

Restricción: Es una propiedad **opcional pero recomendada**, única para cada “Contributor”.

Definición: Apellido del contribuidor.

Ejemplo: Maldonado

7.2.2. givenName

Restricción: Es una propiedad **opcional pero recomendada**, única para cada “Contributor”.

Definición: Nombre del contribuidor.

Ejemplo: Felipe

7.3. nameIdentifier

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples nameIdentifier para cada “Contributor”.

Definición: Identificador único de una entidad legal o individual, de acuerdo a varios esquemas.

El formato depende del esquema.

7.3.1. nameIdentifierScheme

Restricción: Es una propiedad **obligatoria** y única para cada nameIdentifier agregado (por cada “Contributor”).

Definición: El nombre del esquema de identificador de nombres.

Ejemplo: ORCID, ISNI

7.3.2. schemeUri

Restricción: Es una propiedad opcional y única para cada nameIdentifier agregado (por cada “Contributor”).

Definición: El URI del esquema de identificador de nombres.

Ejemplo: <http://www.isni.org>

<http://orcid.org>

7.4. affiliation

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples registros por cada “Contributor”.

Definición: La organización o institución a la cual el contribuidor está afiliado.

Texto Libre.

Ejemplos: DataCite; Universidad de Chile;

Language

Ejemplos

```
<language>en-us</language>
```

```
<language>ESP</language>
```

8. Language

Restricción: Es una propiedad opcional y única.

Definición: El lenguaje principal del recurso.

Ejemplos: en, esp, de, fr.

O cualquiera aceptado por IETF BCP 47 o ISO 6391 (https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes)

Date

Ejemplos

```
<dates>  
  <date dateType="Created">2016-10-30</date>  
</dates>
```

9. Date

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples registros

Definición: Fechas diferentes relevantes para el trabajo.

Formatos: YYYY, YYYY-MM-DD, o cualquier formato descrito en <https://www.w3.org/TR/NOTE-datetime>

9.1. dateType

Restricción: Es una propiedad **obligatoria** y única por cada Date.

Definición: El tipo de fecha.

Lista de valores aceptados:

- Accepted
- Available
- Copyrighted
- Collected
- Created
- Issued
- Submitted
- Updated
- Valid

ResourceType

Ejemplos

```
<resourceType resourceTypeGeneral="Text">Plan de Gestión de Datos</resourceType>
```

```
<resourceType resourceTypeGeneral="Image">Avatar</resourceType>
```

```
<resourceType resourceTypeGeneral="Dataset">Sujeto 1</resourceType>
```

10. ResourceType

Restricción: Es una propiedad **obligatoria** y única.

Definición: Descripción del recurso.

El formato es abierto, pero se recomienda un término singular

Los que sean de tipo general text, se pueden ser texto libre o los terminos de la lista de CASRAI (http://dictionary.casrai.org/Output_Types)

Ejemplos:

Dataset/Census Data, donde dataset es el tipo general y CensusData el ResourceType

Text/Conference Abstract, donde Text es el tipo general y Conference Abstract el ResourceType

10.1. ResourceTypeGeneral

Restricción: Es una propiedad **obligatoria** y única.

Definición: El tipo general del recurso.

Lista de valores aceptados:

Audiovisual, Collection, Dataset, Event, Image, InteractiveResource, Model, PhysicalObject, Service, Software, Sound, Text, Workflow, Other

AlternativIdentifier

Ejemplos

```
<alternateIdentifiers>
  <alternateIdentifier alternateIdentifierType="URL">
    https://schema.datacite.org/meta/kernel-4.0/example/datacite-example-full-v4.0.xml
  </alternateIdentifier>
</alternateIdentifiers>
```

11. AlternativIdentifier

Restricción: Es una propiedad opcional, puede haber múltiples registros.

Definición: Un identificador o identificadores distintos al principal (DOI) aplicados al recurso, debe ser un string alfanumérico que sea único en su dominio. Puede ser usado para identificadores locales.

Esta propiedad debería ser utilizada para otro identificador de la misma instancia (misma localización, mismo archivo)

Texto libre.

Ejemplo: E-GEOD-34814

11.1. alternativIdentifierType

Restricción: Es una propiedad **obligatoria** y única por cada AlternativIdentifier.

Definición: El tipo de AlternativIdentifier.

Texto Libre

Un ejemplo puede ser: "Número de adhesión local"

RelatedIdentifier

Ejemplos

```
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="DOI" relationType="Is
    PartOf">
    10.5272/oldertestpub
  </relatedIdentifier>
</relatedIdentifiers>
```

```
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="URL" relationType="Has
    Metadata" relatedMetadataScheme="citeproc+json" schemeURI="htt
    ps://github.com/citation-style-language/schema/raw/master/csl-
    data.json">http://data.datacite.org/application/citeproc+json/1
    0.5072/example-full
  </relatedIdentifier>
  <relatedIdentifier relatedIdentifierType="arXiv" relationType="
    IsReviewedBy">
    arXiv:0706.0001
  </relatedIdentifier>
</relatedIdentifiers>
```

12. RelatedIdentifier

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples registros.

Definición: Identificadores de recursos relacionados. Deben ser identificadores únicos globales.

Texto Libre. Usar esta propiedad para indicar subconjuntos de propiedades

12.1. relatedIdentifierType

Restricción: Es una propiedad **obligatoria** y única por cada RelatedIdentifier.

Definición: El tipo de RelatedIdentifier.

Lista de valores aceptados:

ARK, arXiv, bibcode, DOI, EAN13, EISSN, Handle, IGSN, ISBN, ISSN, ISTD, LISSN, LSID, PMID, PURL, UPC, URL, URN

12.2. [relationType](#)

Restricción: Es una propiedad **obligatoria** y única por cada RelatedIdentifier.

Definición: Descripción de la relación del recurso registrado (A) y el recurso relacionado (B).

Lista de valores aceptados:

IsCitedBy, Cites, IsSupplementTo, IsSupplementedBy, IsContinuedBy, Continues, HasMetadata, IsMetadataFor, IsNewVersionOf, IsPreviousVersionOf, IsPartOf, HasPart, IsReferencedBy, References, IsDocumentedBy, Documents, IsCompiledBy, Compiles, IsVariantFormOf, IsOriginalFormOf, IsIdenticalTo, IsReviewedBy, Reviews, IsDerivedFrom, IsSourceOf

12.3. [relatedMetadataScheme](#)

Restricción: Es una propiedad opcional y única.

Definición: El nombre del esquema.

Usar sólo para las relaciones HasMetadata e IsMetadataFor

12.4. [schemeURI](#)

Restricción: Es una propiedad opcional y única.

Definición: Uri del relatedMetadataScheme.

Usar sólo para las relaciones HasMetadata e IsMetadataFor

12.5. [schemeType](#)

Restricción: Es una propiedad opcional y única.

Definición: El tipo del relatedMetadataScheme, linkeado con el schemeURI.

Usar sólo para las relaciones HasMetadata e IsMetadataFor

Ej: XDS, DDT, Turtle

Size

Ejemplos

```
<sizes>
  <size>10KB</size>
  <size>2 pages</size>
</sizes>

<sizes>
  <size>256 pages</size>
</sizes>
```

13. Size

Restricción: Es una propiedad opcional, puede haber múltiples entradas.

Definición: Información no estructurada acerca del tamaño del recurso.

Texto libre.

Ejemplos: "6 MB", "15 pages", "20 paginas"

Format

Ejemplos

```
<formats>
  <format>pdf</format>
</formats>

<formats>
  <format> application/pdf</format>
</formats>
```

14. Format

Restricción: Es una propiedad opcional, puede haber múltiples entradas.

Definición: Formato del recurso.

Texto libre. Usar prioritariamente la extensión del archivo o MIME type.

Ej: PDF, XML, MPG o application/pdf, text/xml, video/mpeg

Los formatos aceptados se encuentran en la sección formatos de los Anexos de este documento.

Version

Ejemplos

```
<version>5.5</version>
```

```
<version>1</version>
```

15. Version

Restricción: Es una propiedad opcional y única.

Definición: El número de la versión del recurso.

Sugerido: major_version.minor_version

Registrar un nuevo identificador (DOI) para cambios mayores

Debe ser utilizado con las propiedades 11 y 12 para indicar varias actualizaciones de información

Puede ser utilizado en conjunto a la propiedad 17 para indicar la naturaleza y el rango de la versión de archivo/registro

Rights

Ejemplos

```
<rightsList>  
  <rights rightsURI="http://creativecommons.org/licenses/by-sa/4.0">Creative Commons Attribution-ShareAlike 4.0  
  International (CC BY-SA 4.0)</rights>  
</rightsList>
```

16. Rights

Restricción: Es una propiedad **obligatoria**, puede haber múltiples entradas.

Definición: Cualquier licencia (rights information) para el recurso.

Texto libre:

Ejemplo: Creative Commons Attribution 3.0 Germany License

Para el centro se utilizarán solo licencias de Creative Commons, para más información de estas revisar la Guía de Licencias del centro.

16.1. rightsURI

Restricción: Es una propiedad **obligatoria** y única por cada Rights creado.

Definición: El URI de la licencia.

Ejemplo: <http://creativecommons.org/licenses/by/3.0/de/deed.en>

Description

Ejemplos

```
<descriptions>
  <description descriptionType="Abstract">
    Políticas de Gestión de Datos Científicos.
  </description>
</descriptions>
```

17. Description

Restricción: Es una propiedad **opcional pero recomendada**, puede haber múltiples entradas.

Definición: Toda la información adicional que no califique para alguna de las propiedades, puede ser usada para información técnica.

El formato es abierto.

Es la mejor practica para suplir una descripción.

17.1. descriptionType

Restricción: Es una propiedad **obligatoria** y única por cada Description agregado.

Definición: El tipo de la descripción.

Lista de Valores Aceptados:

- Abstract
- Methods
- SeriesInformation
- TableOfContents
- TechnicalInfo
- Other

GeoLocation

Ejemplos

```
<geoLocations>
  <geoLocation>
    <geoLocationPlace>
      Santiago, Chile
    </geoLocationPlace>
  </geoLocation>
</geoLocations>
```

```
<geoLocations>
  <geoLocation>
    <geoLocationPlace>
      Facultad de Ciencias Físicas y Matemáticas,
      Universidad de Chile
    </geoLocationPlace>
    <geoLocationPoint>
      <pointLongitude>-70.663944</pointLongitude>
      <pointLatitude>-33.457776 </pointLatitude>
    </geoLocationPoint>
  </geoLocation>
</geoLocations>
```

```
<geoLocations>
  <geoLocation>
    <geoLocationPlace>
      Facultad de Ciencias Físicas y Matemáticas,
      Universidad de Chile
    </geoLocationPlace>
    <geoLocationBox>
      <westBoundLongitude>
        -70.664995
      </westBoundLongitude>
      <eastBoundLongitude>
        -70.660974
      </eastBoundLongitude>
      <southBoundLatitude>
        -33.458582
      </southBoundLatitude>
      <northBoundLatitude>
        -33.456887
      </northBoundLatitude>
    </geoLocationBox>
  </geoLocation>
</geoLocations>
```

18. GeoLocation

Restricción: Es una propiedad **opcional pero recomendada**, puede haber más de un registro.

Definición: Lugar o nombre del lugar donde los datos fueron obtenidos o a donde los datos están enfocados.

Repetir esta propiedad para indicar diferentes localizaciones

18.1. geoLocationPoint

Restricción: Es una propiedad **opcional pero recomendada** y única.

Definición: Punto en espacio de una localización.

18.1.1. pointLongitude

Restricción: Es una propiedad **obligatoria** y única para si se utiliza geoLocationPoint.

Definición: Dimensión longitudinal del punto.

Ejemplo: -67,302, Dominio [-180,180]

18.1.2. pointLatitude

Restricción: Es una propiedad **obligatoria** y única para si se utiliza geoLocationPoint.

Definición: Dimensión latitudinal del punto.

Ejemplo: 31.233, Dominio [-90,90]

18.2. geoLocationBox

Restricción: Es una propiedad **opcional** y única.

Definición: Limites espaciales de un box.

Un box está definido por dos puntos geográficos, esquina izquierda inferior y esquina derecha superior. Cada punto está definido por su longitud y latitud.

18.2.1. westBoundLongitude

Restricción: Es una propiedad **obligatoria** y única en caso de utilizar geoLocationBox.

Definición: Longitud más al oeste del box.

Ejemplo: -67,302, Dominio [-180,180]

18.2.2. eastBoundLongitude

Restricción: Es una propiedad **obligatoria** y única en caso de utilizar geoLocationBox.

Definición: Longitud más al este del box.

Ejemplo: -67,302, Dominio [-180,180]

18.2.3. southBoundLatitude

Restricción: Es una propiedad **obligatoria** y única en caso de utilizar geoLocationBox.

Definición: Latitud más al sur del box.

Ejemplo: 31.233, Dominio [-90,90]

18.2.4. northBoundLatitude

Restricción: Es una propiedad **obligatoria** y única en caso de utilizar geoLocationBox.

Definición: Latitud más al norte del box.

Ejemplo: 31.233, Dominio [-90,90]

18.3. geoLocationPlace

Restricción: Es una propiedad **opcional**.

Definición: Descripción de una localización geográfica.

Texto Libre.

18.4. geoLocationPolygon

Restricción: Es una propiedad **opcional** y única.

Definición: Set de puntos y líneas que conectan puntos en una cadena cerrada.

Un polígono está delimitado por puntos geográficos. Cada punto es definido por un par de longitud y latitud. El último punto debe ser igual al primero.

18.4.1. polygonPoint

Restricción: Es una propiedad **obligatoria** en caso de utilizar geoLocationPolygon, deben ser 4 puntos.

Definición: Un punto en el polígono.

Deben ser 4.

Puntos que hacen una curva cerrada.

18.4.1.1. pointLongitude

Restricción: Es una propiedad **obligatoria** y única para cada polygonPoint.

Definición: Longitud de un punto.

Ejemplo: -67,302, Dominio [-180,180]

18.4.1.2. pointLatitude

Restricción: Es una propiedad **obligatoria** y única para cada polygonPoint.

Definición: Latitud de un punto.

Ejemplo: 31.233, Dominio [-90,90]

FundingReference

Ejemplos

```
<fundingReferences>
  <fundingReference>
    <funderName>INNOVA, CORFO CHILE</funderName>
    <awardNumber>
      15TET-40582
    </awardNumber >
    <awardTitle>
      Empaquetamiento y Commercialization Sistema DOCODE.
    </awardTitle>
  </fundingReference>
</fundingReferences>
```

19. FundingReference

Restricción: Es una propiedad **obligatoria**, puede haber múltiples entradas.

Definición: Información acerca del soporte de financiamiento (financiador) para el recurso.

Es la mejor practica para suplir la falta de información de financiadores cuando ésta a sido recibida.

19.1. funderName

Restricción: Es una propiedad **obligatoria** y única para cada FundingReference agregado.

Definición: Nombre del financiador.

Ejemplos: Gordon and Betty Moore Foundation; CONICYT;

19.2. funderIdentifier

Restricción: Es una propiedad **opcional** y única para cada FundingReference agregado.

Definición: Identificador del financiador.

Ejemplo: <http://dx.doi.org/10.13039/100000936>

19.2.1.funderIdentifierType

Restricción: Es una propiedad **opcional** y única para cada FundingReference agregado.

Definición: Nombre del financiador.

Lista de valores aceptados:

ISNI

GRID

Crossref Funder

Other

19.3. awardNumber

Restricción: Es una propiedad **obligatoria** y única para cada FundingReference agregado.

Definición: Código asignado por el fundador.

Ejemplos: GBMF3859.01

15TET-40582

19.3.1.awardURI

Restricción: Es una propiedad **opcional** y única.

Definición: URI que redirige a la página proveída por el financiador para más información acerca del financiamiento (grant).

URI que redirige a la página proveída por el financiador para más información acerca del financiamiento (grant).

19.4. awardTitle

Restricción: Es una propiedad **obligatoria** y única para cada FundingReference agregado.

Definición: El título del financiamiento (human readable).

Ejemplos: Gordon and Betty Moore Foundation

Empaquetamiento y Commercialization Sistema DOCODE.

Anexos

Ejemplo completo archivo XML

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://datacite.org/schema/kernel-4"
xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.5072/example-full</identifier>
  <creators>
    <creator>
      <creatorName>Miller, Elizabeth</creatorName>
      <givenName>Elizabeth</givenName>
      <familyName>Miller</familyName>
      <nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0000-0001-5000-0007</nameIdentifier>
      <affiliation>DataCite</affiliation>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en-us">Full DataCite XML Example</title>
    <title xml:lang="en-us" titleType="Subtitle">Demonstration of DataCite
Properties.</title>
  </titles>
  <publisher>DataCite</publisher>
  <publicationYear>2014</publicationYear>
  <subjects>
    <subject xml:lang="en-us" schemeURI="http://dewey.info/"
subjectScheme="dewey">000 computer science</subject>
  </subjects>
  <contributors>
    <contributor contributorType="ProjectLeader">
      <contributorName>Starr, Joan</contributorName>
      <nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0000-0002-7285-027X</nameIdentifier>
      <affiliation>California Digital Library</affiliation>
    </contributor>
  </contributors>
  <dates>
    <date dateType="Updated">2014-10-17</date>
  </dates>
  <language>en-us</language>
  <resourceType resourceTypeGeneral="Software">XML</resourceType>
  <alternateIdentifiers>
    <alternateIdentifier
alternateIdentifierType="URL">http://schema.datacite.org/schema/meta/kernel-
3.1/example/datacite-example-full-v3.1.xml</alternateIdentifier>
  </alternateIdentifiers>
  <relatedIdentifiers>
    <relatedIdentifier relatedIdentifierType="URL" relationType="HasMetadata"
relatedMetadataScheme="citeproc+json" schemeURI="https://github.com/citation-style-
language/schema/raw/master/csl-
```

```

data.json">http://data.datacite.org/application/citeproc+json/10.5072/example-
full</relatedIdentifier>
  <relatedIdentifier relatedIdentifierType="arXiv"
relationType="IsReviewedBy">arXiv:0706.0001</relatedIdentifier>
  </relatedIdentifiers>
  <sizes>
    <size>3KB</size>
  </sizes>
  <formats>
    <format>application/xml</format>
  </formats>
  <version>3.1</version>
  <rightsList>
    <rights rightsURI="http://creativecommons.org/publicdomain/zero/1.0/">CC0 1.0
Universal</rights>
  </rightsList>
  <descriptions>
    <description xml:lang="en-us" descriptionType="Abstract">
      XML example of all DataCite Metadata Schema v4.0 properties.
    </description>
  </descriptions>
  <geoLocations>
    <geoLocation>
      <geoLocationPlace>Atlantic Ocean</geoLocationPlace>
      <geoLocationPoint>
        <pointLongitude>-67.302</pointLongitude>
        <pointLatitude>31.233</pointLatitude>
      </geoLocationPoint>
      <geoLocationBox>
        <westBoundLongitude>-71.032</westBoundLongitude>
        <eastBoundLongitude>-68.211</eastBoundLongitude>
        <southBoundLatitude>41.090</southBoundLatitude>
        <northBoundLatitude>42.893</northBoundLatitude>
      </geoLocationBox>
    </geoLocation>
  </geoLocations>
</resource>

```

Fuente: DataCite Metadata Schema 4.0⁴

⁴ <https://schema.datacite.org/meta/kernel-4.0/example/datacite-example-full-v4.0.xml>

Formatos

En la presente versión documento, no se han agregado nuevos formatos permitidos por parte del Web Intelligence Centre, por lo que los formatos permitidos son los considerados como aceptables por parte del UK Data Archive, los cuales se muestran a continuación.

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
<p>Quantitative tabular data with extensive metadata</p> <p>a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data</p>	<p>SPSS portable format (.por)</p> <p>delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information</p> <p>some structured text or mark-up file containing metadata information, e.g. DDI XML file</p>	<p>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta)</p> <p>MS Access (.mdb/.accdb)</p>
<p>Quantitative tabular data with minimal metadata</p> <p>a matrix of data with or without column headings or variable names, but no other metadata or labelling</p>	<p>comma-separated values (CSV) file (.csv)</p> <p>tab-delimited file (.tab)</p> <p>including delimited text of given character set with SQL data definition statements where appropriate</p>	<p>delimited text of given character set - only characters not present in the data should be used as delimiters (.txt)</p> <p>widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)</p>
<p>Geospatial data</p> <p>vector and raster data</p>	<p>ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn)</p> <p>geo-referenced TIFF (.tif, .tiff)</p> <p>CAD data (.dwg)</p> <p>tabular GIS attribute data</p>	<p>ESRI Geodatabase format (.mdb)</p> <p>MapInfo Interchange Format (.mif) for vector data</p> <p>Keyhole Mark-up Language (KML) (.kml)</p> <p>Adobe Illustrator (.ai), CAD data (.dxf or .svg)</p> <p>binary formats of GIS and CAD packages</p>
<p>Qualitative data</p> <p>textual</p>	<p>eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml)</p> <p>Rich Text Format (.rtf)</p> <p>plain text data, ASCII (.txt)</p>	<p>Hypertext Mark-up Language (HTML) (.html)</p> <p>widely-used proprietary formats, e.g. MS Word (.doc/.docx)</p> <p>some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti</p>

Digital image data	TIFF version 6 uncompressed (.tif)	JPEG (.jpeg, .jpg) but only if created in this format TIFF (other versions) (.tif, .tiff) Adobe Portable Document Format (PDF/A, PDF) (.pdf) standard applicable RAW image format (.raw) Photoshop files (.psd)
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) but only if created in this format Audio Interchange File Format (AIFF) (.aif) Waveform Audio Format (WAV) (.wav)
Digital video data	MPEG-4 (.mp4) motion JPEG 2000 (.mj2)	
Documentation and scripts	Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)	plain text (.txt) some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHTML 1.0

Fuente: UK Data Archive⁵

⁵ <http://data-archive.ac.uk/create-manage/format/formats-table>

D.3. Guía para el uso del Repositorio

Web Intelligence Centre, Universidad de Chile

Guía para el uso del Repositorio

Versión 1.0

Web Intelligence Centre
15-10-2016

Índice

Introducción	1
Objetivo	1
Contenido	1
Creación de cuentas	1
Cuentas de usuario desde el admin	2
Grupos	4
Estructura del repositorio para almacenamiento de archivos.....	4
Crear Comunidad.....	5
Crear subcomunidades.....	6
Crear Colecciones	7
Permitir crear ítems.....	8
Crear ítems	8
Autorizaciones en Ítems.....	11
Eliminar ítems.....	13
Agregar nuevos archivos a un ítem.	14
Agregar o editar metadatos en ítem creado	14
Formatos.....	15
Metadatos	15
Aprobar o Rechazar ítems (Jefes de Proyecto).....	16

Introducción

En el presente documento se describirán algunas de las funcionalidades permitidas en el repositorio del centro de investigación. Estas funcionalidades son las consideradas como mínimas para el cumplimiento de las políticas de gestión de datos científicos del centro.

Para la realización de ciertos procedimientos, es necesario utilizar las guías de licencias y metadatos del centro.

Se presentará en detalle el paso a paso para cada funcionalidad junto a imágenes sacadas del repositorio del centro para servir como guía para las personas que utilizarán por primera vez el sistema.

Objetivo

El objetivo de este documento es facilitar la comprensión del funcionamiento del repositorio, junto con ser una ayuda para los integrantes del Web Intelligence Centre a la hora de realizar las distintas acciones necesarias para el cumplimiento de las Políticas de Gestión de Datos del centro.

Contenido

A continuación, se presentarán los procedimientos a seguir para las actividades claves para el uso del repositorio del centro:

Creación de cuentas

Para crearse una cuenta en el repositorio se debe entrar a la página y a la derecha del menú superior debe seleccionar Sign In, en donde debe seleccionar su email con que desea registrarse. Una vez hecho esto debe seleccionar el botón “Register”. Después de unos minutos le llegará un mail con un link como se muestra en la siguiente imagen:

Para completar el registro, ingresa a:

<http://localhost:8080/jspui/register?token=8ca11b30f1642d0487242c23e608b8f6>

Si necesitas ayuda con tu cuenta, por favor envía un mail a felipe.maldonado.91@gmail.com.

Repositorio Web Intelligence Centre, Universidad de Chile

Email not readable? [Click here](#) to view raw email data.

Al seleccionar el link se redirigirá a un formulario donde deberá rellenar algunos campos, como se muestra en la imagen:

Registration Information

Please enter the following information. The fields marked with a * are required.

First name*:

Last name*:

Contact telephone:

Language:

Please choose a password and enter it into the box below, and confirm it by typing it again into the second box. It should be at least six characters long.

Password:

Again to Confirm:

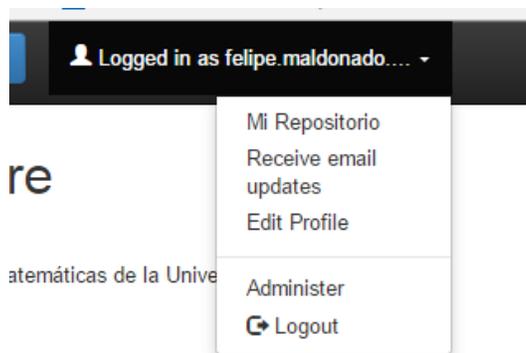
Complete Registration

Una vez completado el registro deberá logearse con sus datos seleccionando el botón “Log in” en el menú superior. Al ingresar con sus credenciales se redirigirá a una vista igual a la de la imagen anterior, sólo que dirá Edit Profile en vez de Registration Information. Esta es la página a la cual se redirige cada vez que se logea en la página.

Cuentas de usuario desde el admin

Para crear cuentas existen dos formas:

La primera, desde la cuenta administradora del repositorio se pueden agregar usuarios o EPerson como es el nombre que DSpace les da a los usuarios. Para esto hay que ingresar a la sección **administer**, en el lado derecho del encabezado de la página, como se muestra a continuación:



Luego seleccionar en el menú superior Access Control -> E-people. En esta página se puede agregar usuarios (botón verde, Add EPerson) o editar existentes.

Al seleccionar Add Person se crea automáticamente un usuario con todos sus valores vacíos y se ingresa finalmente a la página para editar EPersons, la cual se ve de la siguiente manera:

Edit EPerson newuser11: ?

Email:	<input type="text" value="newuser11"/>
Last Name	<input type="text"/>
First Name	<input type="text"/>
Phone:	<input type="text"/>
Language:	<input type="text" value="cl"/>
Can Log In:	<input type="checkbox"/>
Require Certificate:	<input type="checkbox"/>

En este punto no se puede asignar una contraseña, por lo que se le deberá avisar a la persona que se le creó la cuenta que debe ingresar al login de la página y seleccionar "Have you forgotten your password?" para crear una nueva.

En esta página también se puede dar la opción de permitir/denegar el ingreso a la cuenta creada u editada. Para eso se debe seleccionar/quitar selección de la opción **Can Log In**.

La segunda forma es simplemente una persona cualquiera registrarse al repositorio. Al momento de registrarse le llegará un mail al administrador, el cual puede editar este usuario (por ejemplo, quitarle el permiso de logearse al repositorio) o eliminarlo.

Grupos

Los grupos (*groups*) en Dspace son un conjunto de EPersons (usuarios) a los cuales se les puede asignar permisos o tareas según se desee. Cada grupo puede contener a otros grupos y EPersons según se necesite.

Para crear o editar un grupo se debe ir a la sección **administer** y seleccionar en el menú de arriba Access Control -> Groups

En esta página se mostrarán todos los grupos creados hasta el momento. Cada vez que se crea una colección o comunidad se crea un grupo nuevo. Existen dos por defecto: Anonymous, el cual no se puede eliminar y no se le puede asignar usuarios ni grupos, este grupo representa a las personas sin cuentas que ingresen a la página. El otro grupo por defecto es Administrator, que tampoco puede ser eliminado, pero si editado y representa a él/los administradores del repositorio, que tienen total acceso a todas las funcionalidades del sistema.

Para crear un grupo se debe seleccionar el botón verde con el texto: "Create New Group". En esta página se ve de la siguiente forma:

Edit Group : new group20 (id: 20) ?

Name:

After making changes to this group you must click the "Update Group" button below.

EPerson Members

Remove Selected

Select E-people...

Group Members

Remove Selected

Select Groups...

Update Group

Donde se debe seleccionar un nombre al grupo (por ejemplo, integrantes AKORI) y seleccionar a las personas y grupos que quiera pertenezcan a este grupo (pueden seleccionarse sólo EPersons, sólo grupos o ambos a la vez).

Estructura del repositorio para almacenamiento de archivos

La estructura del centro para el almacenamiento es la siguiente: existen comunidad, sub-comunidades, colecciones e ítems.

Las **comunidades** en el contexto de una universidad, serían las distintas facultades dentro de esta. Una comunidad puede tener múltiples sub-comunidades.

Las **sub-comunidades** (que son opcionales) serían los departamentos de cada facultad. Las cuales pueden contener sub-sub comunidades (hasta el nivel que se quiera) y cada una de ellas múltiples colecciones.

Las **colecciones** podrían ser los sub-departamentos, centros de investigación, cursos o cualquier entidad parte del departamento. En cada colección se pueden crear múltiples ítems.

Los **ítems** son los archivos que se suben en cada colección. Un ítem puede contener múltiples archivos, pero hay que considerar que **los metadatos definen al ítem, no a cada archivo dentro del ítem**.

Para el caso del centro, la estructura sería sin utilizar sub-comunidades (aunque esto queda de manera opcional)

Las comunidades serán los proyectos en los que se encuentra trabajando el centro de investigación, como AKORI, DOCODE y OpinionZoom entre otros.

Las colecciones serían los trabajos de cada integrante de cada proyecto (en general sería el trabajo de título o memoria de cada uno, en la cual se pueden generar distintos datos que se traducirían en ítems dentro de cada colección).

Crear Comunidad

Para crear una comunidad nueva se debe estar logeado con una cuenta administradora e ingresar a la sección **administer**. Luego seleccionar Content en el menú superior e ir a “Communities & Collections”.

En esta página se mostrará el listado de comunidades creadas, a la derecha se debe seleccionar el botón “Create Top-Level Community”. Aparecerá el siguiente formulario:

The screenshot shows the 'Edit Community' interface for a community named 'OpinionZoom'. The form is divided into two main sections: 'Community's metadata' and 'Community's settings'.
Community's metadata:
- **Name:** OpinionZoom
- **Short Description:** Descripción corta ejemplo
- **Introductory text (HTML):** A text area containing HTML tags: <p>Ejemplo</p> and <p>Texto en html</p>
- **Copyright text (plain text):** Todos los derechos reservados.
- **Side bar text (HTML):** A text area containing HTML tags: <p>Side Bar ejemplo </p>
- **Logo:** A preview of the logo 'OPINIONZOOM' with buttons for 'Upload new logo...' and 'Delete (no logo)'.
Community's settings:
- A red button at the top: 'Delete this Community...'
- A list of settings with corresponding buttons: 'Community Administrators: Create', 'Community's Authorizations: Edit...', and 'Community's Curations: Curate...'.
At the bottom of the form, there are two large buttons: 'Update' (green) and 'Cancel' (orange).

El nombre será el nombre del proyecto, Short Description es la descripción que se muestra en la página inicial como se ve en la siguiente imagen:



Las informaciones rellenas en los otros campos se mostrarán en la página del proyecto en el repositorio de la siguiente manera:



Al seleccionar Edit, se mostrará el mismo formulario que se ve al crear la comunidad. A la derecha se pueden agregar personas o grupos como administradores de estas comunidades, que en este caso será como mínimo el jefe del proyecto. Además, se pueden editar las autorizaciones de acceso.

Crear subcomunidades

Si una persona es administradora de una comunidad, esta puede crear sub-comunidades a gusto seleccionando el botón "Create Sub-community" en el menú derecho de la vista de la comunidad. El formulario y los pasos son iguales que los de crear una comunidad.

Crear Colecciones

Si una persona es administradora de la comunidad (o sub-comunidad en la que se está viendo), le saldrá la opción de crear una colección dentro de esta (En el menú derecho). Al seleccionarla saldrá un cuadro con las siguientes opciones:

Describe the Collection ?

Please check the boxes next to the statements that apply to the collection.

<input type="checkbox"/>	New items should be publicly readable
<input type="checkbox"/>	Some users will be able to submit to this collection
<input type="checkbox"/>	The submission workflow will include an <i>accept/reject</i> step
<input type="checkbox"/>	The submission workflow will include an <i>accept/reject/edit metadata</i> step
<input type="checkbox"/>	The submission workflow will include an <i>edit metadata</i> step
<input type="checkbox"/>	This collection will have delegated collection administrators
<input type="checkbox"/>	New submissions will have some metadata already filled out with defaults

[Next >](#)

Donde:

- 1. New items should be publicly readable:** Si se selecciona, los archivos subidos serán públicos para cualquier persona externa. En caso de haberla seleccionado, en cada colección se pueden editar estos permisos.
- 2. Some users will be able to submit to this collection:** Si se selecciona, en los siguientes pasos de la creación de la colección, saldrá la opción de agregar personas (con cuenta) o grupos de personas para que tengan permiso de agregar archivos a la colección.
- 3. The submission workflow will include an accept/reject step:** Si se selecciona, en los siguientes pasos se deberá seleccionar a la persona o grupo encargado de aceptar/rechazar un nuevo *submit*.
- 4. The submission workflow will include an accept/reject/edit metadata step:** Si se selecciona, en los siguientes pasos se deberá seleccionar a las personas o grupos encargados de aceptar/rechazar/editar los metadatos agregados por el integrante al momento de subir un ítem.
- 5. The submission workflow will include an edit metadata step:** Si se selecciona, en los siguientes pasos se deberá seleccionar a las personas o grupos encargados de editar los metadatos agregados por el integrante al momento de subir un ítem.
- 6. This collection will have delegated collection administrators:** Si se selecciona, se deberán asignar las personas o grupos que serán administradoras de la colección, en caso de no seleccionarlo, al ingresar con una cuenta administradora se podrán agregar después.
- 7. New submission will have some metadata already filled out with defaults:** Si se selecciona, en la etapa final de la creación de colecciones se podrán seleccionar propiedades a gusto, y darles un valor por default, que en el momento de subir un archivo estos podrían ser cambiados por el integrante que esté realizando esa acción.

Respecto a los puntos 3, 4 y 5. No es recomendable seleccionar los 3 juntos. Una opción es seleccionar 3 y 5 y seleccionar a dos personas/grupos distintos entre sí (si no, al momento de aceptar un archivo, le saldrá la tarea de editar metadatos a continuación). La otra opción es seleccionar solo 4, para el cual habrá sólo una

persona/grupo encargado de la tarea. Los archivos no estarán disponibles en el repositorio hasta que estos sean aceptados en caso de haber seleccionado una de estas opciones.

Luego de seleccionar las opciones que se deseen, se mostrará un formulario con información a rellenar sobre la colección (datos como el nombre, descripción y logo de la colección entre otros). Los siguientes pasos dependerán de las opciones elegidas en el paso anterior, los cuales corresponderán a asignar personas y grupos a cada tarea (*task*) y/o asignar valores por default a metadatos.

Si no se selecciona ninguna de las 7 opciones, sólo saldrá el formulario de información de la colección y una opción para seleccionar a las personas/grupos que podrán revisar los ítems creados dentro de esta colección. Cualquier persona puede ingresar a la colección, pero si no tiene autorización no podrá ver los ítems dentro de estas (mucho menos descárgalos), se recomienda crear un grupo de personas, que serán los que tendrán acceso a ellos y asignarlos en este paso.

El encargado de las tareas y administrador de la colección debe ser el jefe de proyecto.

Permitir crear ítems

Si no se seleccionan las personas que tienen permiso para subir archivos a esta colección (*submitters*), para agregar a una nueva EPerson o grupo se debe entrar con una cuenta administradora de la comunidad (o sub-comunidad) y en el menú derecho aparecerá la opción Edit Submitters, donde podrán ser agregados.

Crear ítems

Para crear un ítem o “submit to this collection”, se debe ingresar a la página de la colección donde se deseen crear ítems.

Primero debe contar con autorización para esto (ver Crear Colecciones – Permitir crear ítems). En caso de que sí la tenga, le saldrá el siguiente botón en la página de la colección:

The screenshot shows a user interface for a digital collection. At the top, there is a blue header with the word "Browse". Below this, there are four search filter buttons: "Issue Date", "Author", "Title", and "Subject". A prominent green button labeled "Submit to This Collection" is centered below the filters. At the bottom of the interface, there is a grey box containing the text "Subscribe to this collection to receive daily e-mail notification of new additions" followed by a blue "Subscribe" button. To the right of this text are three RSS feed icons labeled "RSS 1.0", "RSS 2.0", and "RSS".

Collection's Items (Sorted by Submit Date in Descending order): 1 to 7 of 7

Este procedimiento cuenta con 7 etapas.

Las tres primeras son para describir el ítem que se desea crear, en la primera etapa deberá rellenar datos como el creador, título y año de publicación entre otros, como se muestra en la siguiente imagen:

Describe Describe Describe Upload Verify License License Complete

Submit: Describe this Item ?

Please fill in the requested information about this submission below. In most browsers, you can use the tab key to move the cursor to the next input box or button, to save you having to use the mouse each time.

Apellido (Last Name) y Nombre (First Name)

Creators *

Maldonado Felipe + Add More

Título principal del ítem

Title *

Archivos de Prueba

Si el ítem tiene títulos alternativos, agrégelos acá

Other Titles

+ Add More

Fecha de publicación de los archivos.

PublicationYear *

Month: (No Month) Day: Year: 2016

Nombre del publicador, en general es una institución.

Publisher *

Web Intelligence Centre, Universidad de Chile

El tipo general del recurso. Lista de valores aceptados: Audiovisual, Collection, Dataset, Event, Image, InteractiveResource, Model, PhysicalObject, Service, Software, Sound, Text, Workflow, Other

ResourceTypeGeneral *

Dataset

Descripción corta del recurso. El formato es abierto, pero se recomienda un término singular. Los que sean de tipo general text, se pueden ser texto libre o los terminos de la lista

En esta etapa todos los campos son obligatorios. Si tiene alguna duda sobre qué va en cada campo puede consultar la guía de metadatos del centro.

En la segunda etapa, los campos son opcionales.

En la tercera etapa se mostrarán campos relacionados a los derechos que se otorgarán a los datos e información del financiador del proyecto en el cual están enmarcados los datos, como se muestra en la siguiente imagen:

Describe Describe Describe Upload Verify License License Complete

Submit: Describe this Item ?

Please fill further information about this submission below.

Debe escribir el nombre completo. Ejemplo: Creative Commons Attribution 4.0 International License (CC BY 4.0). Para el centro se utilizarán solo licencias de Creative Commons y copyright, para más información de estas revisar la Guía de Licencias del centro.

Rights *

El URI de la licencia. <https://creativecommons.org/licenses/by/4.0/>

rightsUri

Información acerca del soporte de financiamiento (financiador) para el recurso. Es la mejor practica para suplir la falta de información de financiadores cuando ésta a sido recibida. Ejemplo: El proyecto fue financiado a través de un proyecto FONDEF de CONICYT, código 12356. El proyecto tiene el titulo de "Empaquetamiento y Commercialization Sistema DOCODE."

FundingReference

Nombre del financiador principal. Ejemplos: Gordon and Betty Moore Foundation; CONICYT; FONDEF - CONICYT; CORFO; INNOVA-CORFO

funderName

OPCIONAL: Identificador del financiador, en caso de que lo tenga. Ejemplo: <http://dx.doi.org/10.13039/100000936>

funderidentifier

El título del financiamiento (human readable). Ejemplo: Gordon and Betty Moore Foundation Empaquetamiento y Commercialization Sistema DOCODE

Para los campos Rights y rightsURI puede consultar la guía de licencias. Para los relacionados al financiador la guía de metadatos.

En la cuarta etapa se deberán agregar los archivos que se deseen subir en el nuevo ítem a crear, recordar que los metadatos agregados en los tres pasos anteriores describen al ítem y no a cada archivo que se sube.

Describe Describe Describe Upload Verify License License Complete

Submit: Upload a File ?

Please enter the name of the file on your local hard drive corresponding to your item. If you click "Browse...", a new window will appear in which you can locate and select the file on your local hard drive.

Please also note that the DSpace system is able to preserve the content of certain types of files better than other types. [Information about file types and levels of support for each are available.](#)


Select a file or drag & drop files ...

< Previous Cancel/Save Next >

En la segunda parte de esta etapa se mostrará la siguiente vista:

Submit: File Uploaded Successfully ?

Your file was successfully uploaded.

The table below shows the files which have uploaded for this item.

Primary bitstream	File	Size	Description	File Format	Access Settings
	procedimientos.docx	15176 bytes	None	Microsoft Word XML (known)	 

Add Another File

You can verify that the file(s) have been uploaded correctly by:

- Clicking on the filenames above. This will download the file in a new browser window, so that you can check the contents.
- The system can calculate a checksum you can verify. [Click here for more information.](#) [Show checksums](#)

< Previous

Cancel/Save

Next >

En donde se verán los archivos subidos, los formatos de los archivos (en caso de aparecer unknown, significa que el formato no está agregado al listado de formatos conocidos del repositorio), en este caso se puede asignar manualmente el formato del archivo subido.

En caso de querer agregar una fecha de embargo a algún archivo, se puede seleccionar en Access Settings el botón Change para asignarla.

La quinta etapa es de verificación, en donde se pueden editar todos los metadatos y archivos recién agregados.

La sexta etapa es asignar la licencia Creative Commons que se desee utilizar, para esto se debe utilizar la Guía de Licencias del centro.

La séptima etapa es aceptar una licencia estándar del repositorio.

I Do Not Grant the License

I Grant the License

En caso de seleccionar “I Do Not Grant the License” todo el proceso se borrará y se devolverá a la página de la colección, en caso contrario saldrá un mensaje de error o éxito según sea el caso.

Autorizaciones en Ítems

Las autorizaciones de acceso, lectura, escritura, eliminación son completamente editables para cada uno de los ítems de cada colección.

Para editarlos se debe ingresar al ítem con alguna cuenta con permisos para editar el ítem e ingresar a “Edit” en el menú derecho. Al ingresar se encontrará otro menú a la derecha y se debe seleccionar “Authorizations”.

Saldrán opciones para editar las autorizaciones (Policies) para el ítem completo y para cada archivo que se encuentra en el ítem (se pueden subir varios archivos al mismo ítem)

La edición es igual para cada una de las opciones, se mostrará el ejemplo para las políticas del ítem:

Para agregar una nueva autorización se debe entrar al botón Add Policy y para editar/eliminar alguna existente al botón edit. Si se eliminan todas las políticas (incluyendo el Anonymous – Read), ninguna persona (con o sin cuenta) podrá ingresar o ver este ítem a menos que sea administrador de la colección.

you will notice an extra bundle and bundle for each item, and those contain the license text for the item.

Item Policies			
Add New Policy			
ID	Action	EPerson	Group
369	READ	...	Anonymous
			Edit Delete

Si se desea agregar una política nueva, hay que tener en consideración que se trabaja con grupos y no con usuarios (EPersons) por separado, por lo que si se quiere ingresar a sólo una persona es necesario crear un grupo nuevo (ver Cuentas - Grupo).

El formulario para agregar una nueva política es el siguiente:

Edit Policy for Item 18: ?

Group:

Anonymous
 Administrator
 COLLECTION_2_SUBMIT
 COLLECTION_2_ADMIN
 COMMUNITY_5_ADMIN
 COLLECTION_7_DEFAULT_ITEM_READ
 COLLECTION_9_DEFAULT_ITEM_READ
 COLLECTION_10_DEFAULT_ITEM_READ
 new_group20
 Grupo test

Action:

Start Date:

End Date:

[Save](#) [Cancel](#)

Donde se mostrarán todos los grupos (se puede seleccionar sólo uno a la vez) que se quiera dar autorizaciones personalizadas.

En Action, existen las siguientes opciones: Read, Write, Add, Remove y Admin (Withdrawn_read no se utiliza para asignar políticas, cuando uno selecciona Withdraw un ítem todas las políticas se cambian a withdrawn_read).

Read es para permitir ver el archivo y poder descargarlo.

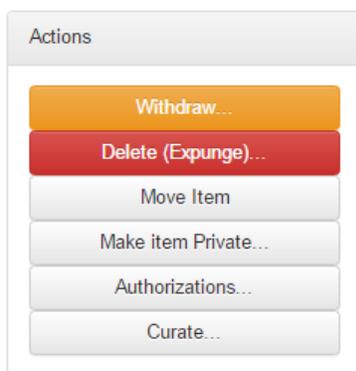
Admin para dar todas las autorizaciones al grupo seleccionado, es decir, podrá agregar metadatos, agregar archivos al ítem y eliminarlos (los archivos que contiene el ítem, no el ítem completo)

Las otras opciones para los tipos de autorizaciones y las fechas de termino e inicio de la política, se encuentran sin uso actualmente por parte de DSpace, por lo que no se utilizarán en el centro.

Al seleccionar el botón Save, se creará la nueva política para el ítem.

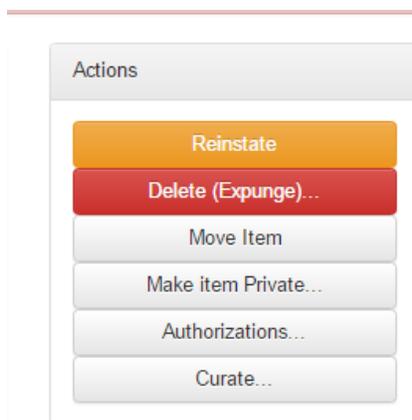
Eliminar ítems

Para eliminar el ítem por completo se debe ser administrador del ítem o colección. Para esto se debe seleccionar el botón Edit del menú de la derecha y seleccionar Delete como muestra en la siguiente imagen:



Saldrá una página para verificar esta acción, para lo cual hay que seleccionar Delete al final de esta en caso de querer continuar.

Una segunda opción es seleccionar Withdraw, lo que hace esta acción es sacar temporalmente el ítem del repositorio y sólo dejarlo disponible para los administradores. Para revisar cuales son los archivos que están en este estado se debe ingresar a la sección **administer** y en el menú superior seleccionar Content -> Withdrawn ítems. Donde saldrá el listado con la opción de editar cada uno de los ítems (dando la opción de devolver el ítem al repositorio seleccionando Reinstale) como se muestra en la siguiente imagen:



Agregar nuevos archivos a un ítem.

Puede ocurrir que una vez creado un ítem se quieran subir más archivos a este, hay que recordar que los metadatos definen al ítem y no a cada archivo, en caso de que se quiera definir a cada archivo se deberá agregar cada uno por separado en forma de ítem.

Como se describe en la guía de metadatos del centro, se debe subir el archivo de metadatos usando el esquema de DataCite y en formato XML en conjunto a los archivos del ítem. Esto puede realizarse en dos formas, una al momento de crear un ítem nuevo o después, que sería agregar un nuevo archivo a un ítem existente.

Para eso se debe tener una autorización de administrador del ítem (ver Autorizaciones en ítems)

Se debe ingresar a la página del ítem y seleccionar Edit. Al final de la página que se cargará saldrá la opción Add Bitstream, en donde se podrá subir un archivo nuevo.

Agregar o editar metadatos en ítem creado

No todos los metadatos existentes en el sistema del repositorio se consultan a la hora de agregar un nuevo ítem a una colección.

Para agregar o editar metadatos para cierto ítem, se debe ingresar a "Edit", primero se mostrarán los metadatos existentes y al final de estos estará la opción de agregar uno nuevo, como se muestra en la siguiente imagen:

dc	title
	Ejemplo
	es
dc.contributor.advisor	

Donde se debe seleccionar el metadato a agregar (que puede ser de cualquier esquema creado, no necesariamente del esquema dc que es interno de DSpace), agregar en el textarea el valor de este, además del lenguaje si se desea (ej: es, en, en_US) y seleccionar el botón +.

Para editar metadatos, se puede editar uno o todos a la vez, simplemente cambiando los valores y yendo al final de la página (debajo de los archivos) y seleccionar Update.

View license.txt Written by org.dspace.cc 2 (License) 1

Add Bitstream Cancel Update

Formatos

Esta acción debe ser realizada por la directora ejecutiva o algún administrador del repositorio:

Dentro de la sección **administer**, se pueden describir los formatos que serán reconocidos por el sistema, es decir, al momento de subir un archivo el repositorio reconocerá el formato del archivo, en caso contrario se mostrará la palabra unknown en vez del formato.

DSpace no permite bloquear archivos de algún formato en específico, se pueden subir archivos en cualquiera que se desee, por eso es recomendado agregar una etapa de aceptar/rechazar en la creación de ítems por parte de integrantes del centro.

Para agregar los formatos que serán conocidos se debe ir al menú superior y seleccionar General Settings -> Bitstream Format Registry. En esta página se encontrarán todos los formatos seleccionados por defecto por el centro, con la opción de editarlos o eliminarlos.

Para agregar uno nuevo se debe ir al final de la página y seleccionar Add New, al hacer click en el botón se actualizará la página y aparecerá un nuevo registro vacío como se ve en la siguiente imagen:

79	MIME Type	Name	Long Description	Unknown ▾	Extensions	Update	Delete...
----	-----------	------	------------------	-----------	------------	--------	-----------

En donde se deben rellenar los datos pedidos y seleccionar Update para guardarlo. En el penúltimo *input* se debe seleccionar la opción Known para que el repositorio finalmente lo reconozca. Se pueden agregar los formatos que se estimen conveniente, se recomienda tener como mínimo los formatos permitidos por el centro de investigación, los cuales pueden ir cambiando en el tiempo.

Metadatos

El sistema interno de Dspace funciona con el esquema de Dublin Core, esto no se puede cambiar. A pesar de esto se pueden crear nuevos esquemas como se desee. En caso de querer agregar una nueva propiedad o esquema se debe ir al menú superior de la sección **administer** y seleccionar General Settings -> Metadata Schema Registry. En la parte inferior existe un formulario para agregar nuevos esquemas, siguiendo el formato de los que están más arriba.

Para agregar o editar las propiedades de uno existente, se debe seleccionar el esquema haciendo click en el nombre del espacio (el cual se muestra como una URL)

ID	Namespace	Name
1	http://dublincore.org/documents/dcmi-terms/	dc
2	http://purl.org/dc/terms/	dcterms

Update Delete...

E ir agregando las propiedades (elements) y subpropiedades (qualifiers) que se deesen. Actualmente DSpace no tiene soporte para sub-subpropiedades.

Al agregar un nuevo registro a cualquier esquema, este estará disponible para ser agregado en cualquier ítem creado (ver Agregar o editar metadatos en ítem creado).

Si se desea que la nueva propiedad se muestre al momento de crear un ítem (ver Crear Item), ésta propiedad debe ser creada en el esquema “dc” sin excepción. Luego para agregarlo al formulario se deben seguir las instrucciones de la documentación de DSpace¹. Este paso es más complicado ya que implica editar archivos directamente de donde esté alojado el repositorio.

Aprobar o Rechazar ítems (Jefes de Proyecto)

En caso de que alguna colección tenga etapas de aceptar/rechazar o editar metadatos, al momento en que alguien cree un ítem en esa colección se le asignará una tarea (*task*) al jefe del proyecto de esa colección.

Se le avisará vía mail, como muestra en la siguiente imagen:

para mí ▾

A new item has been submitted:

Title: Archivos de Prueba
Collection: test
Submitted by: Felipe Maldonado (felipe.maldonado@feriaempresarial.cl)

Requiere revisión

To claim this task, please visit your "My Dspace" page: <http://localhost:8080/jspui/mydspace>

Many thanks!

Repositorio Web Intelligence Centre, Universidad de Chile

Donde al ingresar al link le saldrá el siguiente contenido:

Tasks in the Pool

Below are tasks in the task pool that have been assigned to you.

Task	Item	Submitted To	Submitted By
Review Submission	Archivos de Prueba	test	Felipe Maldonado

Saldrá el listado de tareas asignadas al momento (que no se hayan realizado aún). Al seleccionar “Take Task” saldrá una información de resumen como se muestra en la siguiente imagen:

¹ <https://wiki.duraspace.org/display/DSPACE/Alter+submission+input+forms>

Preview Task

The following item has been submitted to collection **test**. In order to accept the task of reviewing this item, please click "Accept This Task" below.

Title:	Archivos de Prueba
Authors:	Maldonado, Felipe
Issue Date:	2016
Publisher:	Web Intelligence Centre, Universidad de Chile
Citation:	Maldonado, Felipe (2016) : Archivos de Prueba. V.1.0 . eb Intelligence Centre, Universidad de Chile. Dataset/Test Files

Files in This Item:			
File	Description	Size	Format
TEST.xlsx		17,42 kB	Microsoft Excel XML

[View/Open](#)

[Cancel](#) [Accept This Task](#)

Finalmente, al aceptar la tarea seleccionando "Accept This Task" se redirigirá a la siguiente vista:

Perform Task

The following item has been submitted to collection **test**. Please review the item, check that it meets the criteria for entry into the collection. After reviewing the item, please approve or reject the item using the controls at the bottom of the page.

Title:	Archivos de Prueba
Authors:	Maldonado, Felipe
Issue Date:	2016
Publisher:	Web Intelligence Centre, Universidad de Chile
Citation:	Maldonado, Felipe (2016) : Archivos de Prueba. V.1.0 . eb Intelligence Centre, Universidad de Chile. Dataset/Test Files

Files in This Item:			
File	Description	Size	Format
TEST.xlsx		17,42 kB	Microsoft Excel XML

[View/Open](#)

Approve	If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".
Reject	If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and re-submit.
Do Later	If you wish to leave this task for now, and return to your "My DSpace", use this option.
Return Task to Pool	To return the task to the pool so that another user can perform the task, use this option.

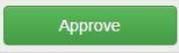
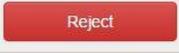
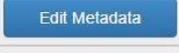
Donde el menú inferior puede tener tres formas. Si la colección contaba con una etapa de Accept/Reject será como en la imagen anterior. Donde debe seleccionar Approve o Reject. La tarea consiste en revisar el archivo que se subió (Seleccionando View/Open se puede descargar el archivo) además de que la información que se muestra en la imagen esté correcta en cuanto a contenido de los metadatos.

En caso de que tenga una etapa de Edit Metadata saldrá el siguiente menú:

	Once you've edited the item, use this option to commit the item to the archive.
	Select this option to correct, amend or otherwise edit the item's metadata.

Donde al seleccionar Edit Metadata se podrán cambiar los contenidos de los metadatos rellenos por el integrante en el momento de crear el ítem. Si no desea editar nada más debe seleccionar Commit to Archive, esta tarea no tiene la opción de rechazar algún ítem.

Finalmente, si la colección tiene una etapa de Accept/Reject/Edit Metadata, el menú será el siguiente:

	If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".
	If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and re-submit.
	Select this option to correct, amend or otherwise edit the item's metadata.

En donde al igual que para el caso anterior, podrá editar los metadatos agregados al momento de crear el ítem, pero a diferencia del caso anterior podrá rechazarlo en caso de considerarlo necesario.

Estas etapas se definen cuando se crea la colección en alguna comunidad (proyecto). Los jefes de proyectos pueden crear estas etapas cuando lo deseen en caso de que la colección ya estaba creada de antes. Para esto debe ingresar a la colección y en el menú izquierdo debe seleccionar Edit. Le aparecerá el siguiente menú:

Delete this Collection...

Submission Workflow

Submitters: Edit... Delete

Accept/Reject Step: Create...

Accept/Reject/Edit Metadata Step: Create...

Edit Metadata Step: Create...

Collection Administrators: Edit... Delete

Collection's settings

Item template: Edit... Delete

Collection's Authorizations: Edit...

Collection's Curations: Curate...

En donde en caso de no existir las etapas podrá crearlas o editarlas.

D.4. Compromiso de cumplimiento de Políticas

Documento de Compromiso

Yo, _____ (RUT: _____), perteneciente al proyecto _____ del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, declaro haber leído y aceptado las Políticas de Gestión de Datos Científicos del centro en su versión __.

Con esto, me comprometo a cumplir las reglas estipuladas en las Políticas del centro mientras sea parte de él, siguiendo los protocolos y teniendo claro cuáles son mis responsabilidades y cuáles son las de los demás integrantes del centro.

Juan Velásquez

Director

FIRMA

Santiago, __ de ____ de 20__.

D.5. Permiso de Acceso a datos del centros

Permiso de Acceso a Datos

Condiciones

Con el fin de respaldar la seguridad e integridad de algunos datos científicos generados en el centro, el Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas debe aprobar cualquier tipo de acceso a algunos datos sensibles y puede guardarse el derecho de revocar este permiso si lo considera necesario.

Cualquier persona externa o interna al centro que desee acceder a los datos deberá firmar una solicitud de acceso, la cual se encuentra en la segunda hoja de este documento, comprometiéndose a cumplir todos los requisitos impuestos por el centro, expresados en sus Políticas de Gestión de Datos Científicos y seguir sin excepción los pasos incluidos en las guías de metadatos, licencias y uso del repositorio del centro.

En caso de ser aceptada, la persona podrá optar al acceso a los datos guardados en el repositorio de datos científicos del centro, en donde se le creará una cuenta y darán los permisos necesarios para esto.

En caso de que no se cumpla algunos de los requerimientos durante el período que cuente con acceso a los datos, este permiso será eliminado. El centro se guarda el derecho a aceptar o eliminar permisos cuando lo desee o considere necesario.

En la siguiente hoja se encuentra el permiso que deberán firmar los solicitantes de acceso.

Permiso de Acceso

Yo, _____ (RUT: _____), solicito acceso a los datos generados en la investigación _____ del proyecto _____ del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

Para esto me comprometo a cumplir todos los requerimientos y protocolos impuestos por el centro en relación a los datos científicos y su manejo. Conociendo cuáles son mis responsabilidades y consecuencias en caso de no cumplirlas.

Jefe de Proyecto

FIRMA SOLICITANTE

Santiago, ____ de ____ de 20 ____.

D.6. Template: Reporte por proyecto

Reporte final Proyecto X

Jefe de Proyecto	
Mail de contacto	
Fecha	
Financiamiento	

Integrantes

Integrante 1

Nombre:	
Mail de contacto:	
Trabajo:	

	Cumplimiento (%)	Observaciones
Digitalización documentos		
Entrega PGDs		
Metadatos		
Uso repositorio		
TOTAL		

Integrante 2

Nombre:	
Mail de contacto:	
Trabajo:	

	Cumplimiento (%)	Observaciones
Digitalización documentos		
Entrega PGDs		
Metadatos		
Uso repositorio		
TOTAL		

Total

El cumplimiento promedio del proyecto X es el siguiente:

	Cumplimiento (%)
Digitalización documentos	
Entrega PGDs	
Metadatos	
Uso repositorio	
TOTAL	

Observaciones:

Plan de Gestión de Datos
D.7. Template: Primera versión DMP

Plan de Gestión de Datos

Nombre del proyecto de investigación

Autor

Fecha

Versión

URI

IPR

Datos Administrativos:

NOMBRE INVESTIGADOR	
TRABAJO A REALIZAR	
PROYECTO WIC	
CÓDIGO DEL PROYECTO	
FINANCIAMIENTO	
ENCARGADO DE PROYECTO	
MAIL DE CONTACTO	
TELÉFONO DE CONTACTO	

Requerimientos a cumplir

Descripción del Proyecto

Outputs

Se espera obtener los siguientes outputs:

<i>#</i>	<i>Output</i>	<i>Tipo</i>	<i>Formato / Duración / Tamaño</i>
1			
2			
3			
4			

Plan de Gestión de Datos

Nombre del proyecto de Investigación

Autor

Fecha

Versión

URI

IPR

Datos Administrativos:

NOMBRE INVESTIGADOR	
TRABAJO A REALIZAR	
PROYECTO WIC	
CÓDIGO DEL PROYECTO	
FINANCIAMIENTO	
ENCARGADO DE PROYECTO	
MAIL DE CONTACTO	
TELÉFONO DE CONTACTO	

Requerimientos a cumplir

Descripción del trabajo

Outputs

Con el experimento se obtuvieron los siguientes OUTPUTS:

<i>#</i>	<i>Output</i>	<i>Tipo</i>	<i>Formato / Duración / Tamaño</i>
1			
2			
3			

Metodologías y Algoritmos utilizados

Datos a Almacenar

Metadatos

Documentos

Propiedad

Plan de Gestión de Datos

Estudio del Impacto del uso de electroencefalograma en la identificación de
Website Keyobjects

Autor	Juan Alberto Pérez Olivares
Fecha	26/05/2016
Versión	1
URI	Wic.uchile.cl
IPR	Web Intelligence Centre @ 2016. Algunos derechos reservados

Datos Administrativos:

(Los datos a continuación no son reales)

NOMBRE INVESTIGADOR	Juan Alberto Pérez Olivares
TRABAJO A REALIZAR	Estudio del Impacto del uso de electroencefalograma en la identificación de Website Keyobjects
PROYECTO WIC	AKORI
CÓDIGO DEL PROYECTO	CA12i10061 C
FINANCIAMIENTO	FONDEF
ENCARGADO DE PROYECTO	Gino Slanzi R.
MAIL DE CONTACTO	Ja.perez.O@mail.com
TELÉFONO DE CONTACTO	+6598456456

Requerimientos a cumplir

Esta investigación deberá cumplir con todos los requisitos dados por las Políticas de Gestión de Datos Científicos del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas Universidad de Chile.

Descripción del Proyecto

El proyecto AKORI busca generar un servicio tecnológico de apoyo a la toma de decisiones de estructura, contenido y diseño de sitios web, destinado a las organizaciones.

La investigación realizada por el presente autor enmarcada en el proyecto AKORI tiene el fin de buscar obtener conocimiento empírico del tipo de emociones que sienten los usuarios al verse enfrentados a los sitios Web, de manera de contar con un respaldo para la toma de decisiones que optimicen el diseño y la efectividad de los sitios Web.

Para esto se realizará un experimento utilizando el equipo de Eye Tracker EyeLink1000 y un equipo de electroencefalograma Neuroheadset EPOC. Por otro lado se utilizará el sitio web del programa MBA de Ingeniería Industrial de la Universidad de Chile a la fecha y se utilizarán 20 sujetos para tomar los datos. El experimento consiste básicamente en registrar el posicionamiento ocular, dilatación pupilar y la actividad cerebral de los sujetos al ser expuestos a estímulos visuales presentados en un sitio web, mostrando las distintas páginas del sitio web de manera secuencial y cambiándolas cuando la persona considere que no hay nada más que les llame la atención o les interese, sin tiempos mínimos o máximos por página.

Este experimento se espera realizar entre el 30/05/2016 y el 10/06/2016 en las instalaciones del Web Intelligence Centre.

Outputs

Con el experimento se espera obtener los siguientes OUTPUTS:

#	Output	Tipo	Formato / Duración / Tamaño
1	20 sets Eye Tracker	Datos Cualitativos	EDF / 30 min promedio / 30 MB promedio
2	20 sets EEG	Datos Cualitativos	BDF / 30 min promedio/ 400 MB promedio
3	Información de sujetos	Datos Cualitativos	XLS / 1 MB

Plan de Gestión de Datos

Estudio del Impacto del uso de electroencefalograma en la identificación de
Website Keyobjects

Autor	Juan Alberto Pérez Olivares
Fecha	26/12/2016
Versión	2
URI	Wic.uchile.cl
IPR	Web Intelligence Centre @ 2016. Algunos derechos reservados

Datos Administrativos:

(Los datos a continuación no son reales)

NOMBRE INVESTIGADOR	Juan Alberto Pérez Olivares
TRABAJO A REALIZAR	Estudio del Impacto del uso de electroencefalograma en la identificación de Website Keyobjects
PROYECTO WIC	AKORI
CÓDIGO DEL PROYECTO	CA12i10061 C
FINANCIAMIENTO	FONDEF
ENCARGADO DE PROYECTO	Gino Slanzi R.
MAIL DE CONTACTO	Ja.perez.O@mail.com
TELÉFONO DE CONTACTO	+6598456456

Requerimientos a cumplir

Esta investigación deberá cumplir con todos los requisitos dados por las Políticas de Gestión de Datos Científicos del Web Intelligence Centre de la Facultad de Ciencias Físicas y Matemáticas Universidad de Chile.

Descripción del trabajo

El proyecto AKORI busca generar un servicio tecnológico de apoyo a la toma de decisiones de estructura, contenido y diseño de sitios web, destinado a las organizaciones.

La investigación realizada por el presente autor enmarcada en el proyecto AKORI tiene el fin de buscar obtener conocimiento empírico del tipo de emociones que sienten los usuarios al verse enfrentados a los sitios Web, de manera de contar con un respaldo para la toma de decisiones que optimicen el diseño y la efectividad de los sitios Web.

Para esto se realizó un experimento entre el 10/06/2016 y el 15/06/2016, utilizando el equipo de Eye Tracker EyeLink1000 y un equipo de electroencefalograma Active Two System de Biosemi, el cual fue prestado por el Laboratorio de Neurosistemas de la Facultad de Medicina de la Universidad de Chile, lugar donde se realizó el experimento. En un principio se iba a utilizar el equipo Neuroheadset EPOC, pero por problemas técnicos y de software se optó por cambiarlo.

Por otro lado, se utilizó el sitio web del programa MBA de Ingeniería Industrial de la Universidad de Chile a la fecha (www.mbauchile.cl) y se utilizaron 20 sujetos para tomar los datos. El experimento consistió básicamente en registrar el posicionamiento ocular, dilatación pupilar y la actividad cerebral de los sujetos al ser expuestos a estímulos visuales presentados en un sitio web, mostrando las distintas páginas del sitio web de manera secuencial y cambiándolas cuando la persona considere que no hay nada más que les llame la atención o les interese, sin tiempos mínimos o máximos por página.

El grupo elegido constó de 17 hombres y 3 mujeres, con una edad promedio de 24,2 años (con una varianza de 1,64 años). Personas principalmente universitarios y profesionales de diversas áreas como medicina, ingeniería, fotografía, entre otros.

El protocolo utilizado para cada toma de muestras fue el siguiente:

1. Preparar equipos: tener equipos y dispositivos encendidos y listos para ser utilizados apenas llegaba el sujeto de prueba.

2. Bienvenida y relleno de documentos: Al llegar a la sala de registros, se le dio la bienvenida y agradecimientos a los sujetos por su participación en el experimento. Además, se explicó brevemente lo que se hará y se les pidió que lean y firmen un consentimiento informado.

3. Configurar EEG: Los pasos son los siguientes:

- Medición de las dimensiones de la cabeza del sujeto, siendo tres las distancias principales: nasion-inion (desde la frente a la nuca); trago-trago (desde una oreja a la otra); circunferencia. Con estas dimensiones se eligió una gorra adecuada y se instaló de manera correcta en la cabeza del sujeto.
- Limpieza con alcohol las zonas donde se colocan los electrodos externos (alrededor del ojo y zona mastoidea), para reducir la impedancia de la piel y obtener señales más precisas y limpias.
- Colocar los 8 electrodos externos alrededor de los ojos y en las zonas mastoideas con adhesivos para electrodos y gel conductor que conecta el metal del electrodo a la piel.
- Escoger la gorra de EEG de 32 orificios de acuerdo al tamaño de la circunferencia de la cabellera del sujeto.
- Acomodar la gorra de EEG ajustando la posición Cz de la misma a la mitad de las distancias entre nasion-inion y trago-trago.
- Suministrar gel conductor en cada orificio cuidando que este en contacto con la piel de la cabellera.
- Ubicar la banda de 32 electrodos en la gorra de EEG respetando la posición correspondiente a cada uno.
- Conectar los electrodos al equipo para traspasar la información a un computador.
- Revisar en el programa ActiView en el computador de destino de las señales para comprobar la configuración ajustada.

4. Configurar Eye Tracker: Para cada usuario se debe configurar el dispositivo de acuerdo a sus propias características. El proceso es el siguiente:

- Ajustar la cabeza del sujeto, según su altura, al soporte dedicado para este propósito.
- Ajustar la altura de la pantalla de estímulo a la altura de los ojos del sujeto. La parte central de la pantalla debe quedar justo frente a los ojos del sujeto.
- Ajustar la cámara infrarroja a la posición de la cabeza considerando que deben detectarse tanto el reflejo corneal y la pupila. Además, se debe fijar el enfoque de la cámara.

- Calibrar el Eye Tracker al ajuste realizado en la CPU exclusiva del EyeLink 1000.
- Validar la calibración realizada.

5. Registro: Se lleva a cabo el experimento, la instrucción declarada a cada sujeto fue la siguiente: "Navegue libremente por el sitio, sin límites de tiempo (mínimo ni máximo) por página. Utilice las flechas abajo y arriba para bajar o subir en la página. Cuando termine de explorar cada página presione la flecha derecha en el teclado para pasar a página siguiente.

6. Fin del experimento: Al finalizar el experimento se realizaron los siguientes pasos:

- Detener el registro de los equipos y retirar los instrumentos del sujeto.
- Lavar el cabello del sujeto para remover los residuos de gel conductor.
- Finalmente se lavar los instrumentos utilizados, limpiar las zonas ocupadas y dejar en perfecto orden la sala de registro.

Esto fue repetido por cada sujeto que se sometió al experimento.

NOTA 1: Al momento de realizar el experimento al sujeto 5, se cortó la luz por unos segundos, pero todo volvió a la normalidad después de esto.

NOTA 2: Para el sujeto 15 se tomó una muestra más larga debido a que en la mitad del experimento tuvo que atender una llamada, continuando el experimento después esto.

NOTA 3: La muestra del sujeto 20 tuvo que ser eliminada debido a una falla técnica en el computador del EEG.

Outputs

Con el experimento se obtuvieron los siguientes OUTPUTS:

#	Output	Tipo	Formato / Duración / Tamaño
1	20 sets Eye Tracker	Datos Cualitativos	EDF / 30 min promedio / 27 MB promedio
2	20 sets EEG	Datos Cualitativos	BDF / 30 min promedio/ 356 MB promedio
3	Información de sujetos	Datos Cualitativos	XLS / 1 MB

Para los archivos EDF las columnas significan lo siguiente:

Variable	Definición
<time>	tiempo en milisegundos.
<xpl>	posición ojo izquierdo en eje x.
<ypl>	posición ojo izquierdo en eje y.
<psl>	tamaño pupila izquierda (área or diámetro).
<xpr>	posición ojo derecho en eje x.
<ypr>	posición ojo derecho en eje y.
<psr>	tamaño pupila derecha (área or diámetro).
<xvl>	velocidad instantánea ojo izquierdo en eje x (grados/sec).
<yvl>	velocidad instantánea ojo izquierdo en eje y (grados/sec).
<xvr>	velocidad instantánea ojo derecho en eje x (grados/sec).
<yvr>	velocidad instantánea ojo derecho en eje y (grados/sec).
<xr>	resolución en eje x (unidades de posición/grados).
<yr>	resolución en eje y (unidades de posición/grados).

Metodologías y Algoritmos utilizados

Para el análisis de los datos se utilizó el software Matlab.

Para los datos del EyeTracker, se realiza un proceso KDD:

Selección: Se eligieron los 19 de 20 archivos EDF, los cuales fueron convertidos al formato ASCII, que pesan en promedio 39 MB

Preprocesamiento: se eliminaron los elementos referentes a los Parpadeos (se realizó una interpolación para arreglar las pérdidas de señales producidas por el parpadeo), Sacadas (para arreglar este fenómeno se lleva la señal desde el punto donde termina una sacada al punto donde comienza y generar una interpolación lineal entre los extremos de ese intervalo), Altas Frecuencias (se eliminaron todas las frecuencias mayores a 2 Hz ya que se consideran como ruido).

Transformación: Se separaron cada una de las matrices en intervalos, generando una nueva matriz que indicaba la página en que se encontraba cada sujeto, el objeto que estaba mirando, el índice inicial y el final. Se extrajeron las siguientes propiedades de cada sujeto: Tiempo promedio por cada objeto, indicador delta (diferencia entre máxima y mínima contracción de la pupila del sujeto) y número de vistas al objeto.

Minería de Datos: Se utilizó el agrupamiento K-Means considerando como variables de clasificación el Tiempo promedio y el indicador delta. Después se creó un nuevo indicador definido como el producto entre el indicador delta y el tiempo, se ordenó de mayor a menor y se eliminó los que tienen valor igual a cero. Se seleccionaron los 20 primeros objetos para analizar los resultados.

Para los datos del encefalograma, también se realiza un proceso KDD

Selección: Se eligieron los 19 de 20 archivos BDF, que pesan en promedio 356 MB. Se utilizó el plug-in Eeglab de Matlab y Eye-Eeg también de Matlab para sincronizar las grabaciones de EEG y EyeTrackers. Se trabajó con las matrices formadas por los valores de las señales de electrodos en función del tiempo.

Preprocesamiento: Se bajó la tasa de muestreo de 2000Hz a 500Hz, se filtró las frecuencias dejando los datos entre 1 y 60Hz, se sincronizó la medición de ambos artefactos usando el plug-in Eye-Eeg. Después de esto se realizó un análisis de componentes principales (ACP) distinguiendo 32 componentes en donde se eliminan los ruidos realizando interpolaciones.

Transformación: Primero se graficó los ERP para comparar la actividad cerebral de los Website Keyobjects, se utilizaron las señales promediadas de los 20 objetos obtenidos en el proceso KDD de los datos generados por el Eyetracker, y el promedio de las señales de otros objetos vistos por los sujetos. Para entender las señales, se separan por intervalos obteniendo matrices que indican el tiempo, el objeto, los intervalos de tiempo y las señales de todos los canales del EEG. Se graficaron después los ERP para todos los canales y para electrodos por separados, en este trabajo se utilizaron todos los canales, el electrodo PZ y el electrodo FC6 para la medición de bioactividad cerebral.

Se realizó una descomposición de las señales en bandas de frecuencias que coincidían con las ondas típicas del EEG (Alpha, Beta, Gamma, Delta, Theta) a través de la transformación de Wavelet. Para esto se crean tres bases de datos, una con las señales de cada objeto caracterizada por las 14 variables para todos los canales, y las otras dos con contienen las mismas variables para el electrodo PZ y FC6 respectivamente.

Las catorce variables son: Tiempo, Indicador Delta, Vistas, Indicador Delta * Tiempo, media, varianza, desviación estándar y varianza de cada señal. Las características: Energía, porcentaje de energía y RMS de las bandas Delta y Alpha(Para elegir las se descompuso usando la transformación de Wavelet en 6 niveles de detalle y uno final de aproximación junto a la función Wavelet Daubechies de orden 5 y eligiendo estas dos que eran las que más aportaban información) y finalmente la variable Dimensión Fractal (usando algoritmo Higuchi)

Minería de datos: Se realizó un algoritmo de clustering K-means usando los siguientes atributos como clasificadores: [Tiempo y energía banda delta], [Tiempo y % de Energía Banda Delta], [Tiempos y Estadísticos típicos], [Tiempo y RMS], [Tiempo y Dimensión Fractal], [Tiempo * Indicador delta y Energía] y [Tiempo * Indicador delta y % de Energía]

Datos a Almacenar

Los datos a almacenar en el repositorio del WIC se consideran de carácter sensible debido a su naturaleza y son los archivos obtenidos en el experimento en los formatos requeridos por el centro, en este caso los 40 sets de datos fueron convertidos a formato ASCII.

Metadatos

Los metadatos cambiarán los conjuntos de datos de eye-trackers y EEG, son los siguientes:

Creator: Pérez Olivares, Juan Alberto

Title: Datos Eye Tracker (EEG) de experiment para la identificación de Website Keyobjects

PublicationYear: 2016

Publisher: Web Intelligence Centre, Universidad de Chile

ResourceTypeGeneral: Dataset

ResourceType: Dataset/Eye Tracker (EEG Data)

Citation: Pérez Olivares, Juan Alberto (2016): Datos Eye Tracker (EEG) de experiment para la identificación de Website Keyobjects. V.1.0. Web Intelligence Centre, Universidad de Chile. Dataset/Eye Tracker (EEG Data)

Rights: © Web Intelligence Centre “año actual”, Todos los derechos reservados.

funderName: FONDEF – CONICYT

awardTitle: Estudio del Impacto del uso de electroencefalograma en la identificación de Website Keyobjects

awardNumber: CA12i10061 C

Documentos

Estos datos serán guardados en el repositorio del proyecto AKORI junto a todos los documentos generados: consentimientos informados digitalizados, información de los sujetos en formato CSV y el presente documento en formato PDF. También se adjuntará el archivo de metadatos en el esquema de DataCite en formato XML

Propiedad

Los datos generados son de exclusiva propiedad de los sujetos utilizados en el experimento, los cuales en el consentimiento informado aceptaron que estos sean utilizados para esta investigación

Anexos E

Validación Arquitectura

E.1. Test de Usabilidad

E.1.1. Formulario inicial

1. ¿A qué proyecto pertenece dentro del Web Intelligence Centre?
2. ¿Cuál es su rol dentro de ese proyecto y del centro?
3. ¿Revisó las guías de licencias, metadatos y de uso del repositorio? (Sí / No)
4. ¿Ha subido datos científicos a algún repositorio de datos alguna vez?

E.1.2. Guión de Orientación

Gracias por participar en este test. Estoy aquí para analizar la usabilidad del repositorio y las guías que diseñé para el centro y conocer sus fortalezas y debilidades.

El test consiste en que realices una cierta cantidad de tareas que dependerá de tu cargo en el centro, utilizando el computador que se encuentra al frente tuyo. La idea es que realices todas estas tareas pensando en voz alta, mencionando todas las observaciones que se te vayan pasando por tu cabeza. Se te va a describir una tarea y luego debes realizarla, en algunos casos deberé ingresar al computador para realizar ciertas configuraciones del repositorio para que puedas ir cumpliendo las tareas. Cada tarea deberás realizarla dos veces.

Espero que seas totalmente sincero/a, ya que esto me va a permitir encontrar de mejor manera los posibles errores o problemas que puedan ocurrir. Recuerda que se está probando el repositorio y no a ti o a los otros usuarios de este.

¿Tienes alguna pregunta antes de comenzar?

E.1.3. Entrevista post pruebas

1. ¿Cuál es la dificultad de las tareas asignadas? (preguntar por cada tarea) - Muy fácil - Fácil - Normal - Difícil - Muy difícil

Comentarios:

2. ¿Las guías entregadas, son suficientes para realizar las tareas? -Si, son suficientes. -No, no lo son.

Comentarios:

3. ¿Considera que fue muy largo el test?

Comentarios:

4. A modo general, ¿qué le parece el repositorio?

Comentarios:

5. ¿Le cambiarías algo al repositorio?

Comentarios: