



ESTADO DEL ARTE NACIONAL E INTERNACIONAL EN MATERIA DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA Y TECNOLOGICA Y RECOMENDACIONES DE BUENAS PRACTICAS



Julio, 2010

INFORME FINAL

Documento Ejecutivo que resume el Benchmarking Internacional, el Análisis de Capacidades Chilenas, aprendizajes internacionales y Recomendaciones de Buenas Prácticas.

EQUIPO DEL ESTUDIO

CONICYT, Mandante

Patricia Muñoz, Directora Depto. Información Científica

Paula González, Directora de Estudios y Planificación Estratégica

Paula Castro, Depto. De Estudios y Planificación Estratégica

UNIVERSIDAD DE LA FRONTERA-INSTITUTO DE DESARROLLO REGIONAL Y LOCAL, Ejecutor Estudio

Ronald Cancino, Jefe de Proyecto.

Patricio Padilla, Coordinador Metodológico

Oswaldo Curaqueo, Investigador

Mario Gatica, Investigador

Jorge Petit-Breuilh, Investigador

Ignacio Rodríguez y Paulina Sanhueza, Análisis Marco Normativo Nacional

Cristian Alister, Investigador Asistente

Felipe Mellado, Investigador Asistente

Fabiola Ramos, Comunicaciones

Hilda Hernández, Análisis Cualitativo

Dasten Julián, Análisis Cualitativo

Fidel Inostroza, Álvaro Álamos, Jorge Sandoval, Cesar Santibáñez, Call Center

INDICE

| | |
|---|----|
| INDICE | 2 |
| INTRODUCCION | 3 |
| REFLEXIONES TEORICAS. COMUNIDADES EPISTEMICAS, REDES Y MARCOS INSTITUCIONALES EN EL ACCESO A DATOS DE INVESTIGACION E INFORMACION CIENTIFICA..... | 6 |
| II. METODOLOGIA GENERAL DEL ESTUDIO | 9 |
| II.1. Metodología para el Benchmarking Internacional..... | 9 |
| II.2 Metodología para el Análisis y Visualización de las Capacidades Chilenas en materia de Gestión de Datos de Investigación e Información Científica y Tecnológica..... | 11 |
| II.2.1 Metodología para el análisis de capacidades chilenas en gestión de datos de investigación e información científica y tecnológica | 12 |
| II.2.2. Encuesta a investigadores y académicos..... | 14 |
| III. ESTADO DEL ARTE INTERNACIONAL EN MATERIAS DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA Y TECNOLOGICA..... | 16 |
| III.1. La gestión de datos de investigación e información científica desde una óptica histórico-evolutiva..... | 17 |
| III.2. Especificidades disciplinarias en gestión de datos de investigación e información científica | 23 |
| IV. CAPACIDADES CHILENAS DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA Y TECNOLOGICA | 38 |
| IV.1. Capacidades de instituciones e investigadores chilenos en gestión de datos de investigación e información científica..... | 38 |
| Resultados y conclusiones sobre institucionalidad | 39 |
| Resultados y conclusiones sobre capital humano..... | 41 |
| Resultados y conclusiones sobre infraestructura tecnológica..... | 43 |
| Resultados y conclusiones sobre patrimonio | 45 |
| Resultados y conclusiones sobre vinculación y redes | 49 |
| Resultados y conclusiones sobre política de gestión | 52 |
| IV.2. La experiencia chilena en oceanografía y sismología..... | 60 |
| IV.3. Percepciones de los agentes que realizan gestión de datos de investigación e información científica ... | 80 |
| V. APRENDIZAJES INTERNACIONALES Y BUENAS PRÁCTICAS..... | 98 |

INTRODUCCION

En el año 2002, John Brockman editó un libro, “Los próximos 50 años. El conocimiento humano en la primera mitad del siglo XXI”, en el cual 25 grandes científicos del mundo, se aventuran a identificar prospectivamente el estado de situación de sus disciplinas científicas, y el estado de desarrollo tecnológico por ellas generadas. La confluencia de la “mimética”, la inteligencia artificial, la biotecnología, los avances en física, cosmología, etc., indican un escenario de futuro de altísima complejidad sociotécnica, con transformaciones incluso en la comprensión de la “naturaleza humana”, y de la “vida”. En la base de estas transformaciones, se encuentra el cambio en la naturaleza de generación del conocimiento científico, y allí, la gestión de datos de investigación e información científica.

Si inmediatamente posterior a la II Guerra Mundial se produce el fenómeno conocido como la Big Science, aquel estallido de producción científica y tecnológica anidada a la industria y el desarrollo militar, hoy estamos en presencia de una segunda gran transformación, generada por el impacto de las tecnologías de la información. Si la Big Science transforma la organización de la ciencia, escala su producción y define campos de investigación científica nuevos, hoy se enfrenta lo que Ulrich Beck llama la “cientificación reflexiva”. La ciencia, comienza a operar sobre las bases que ella misma había generado, construyéndose un nuevo entorno de actuación. Así, la nueva “Revolución Científica” (Interagency Working Group on Digital Data, 2009) viene a actuar sobre los impactos al menos de las imágenes digitales generadas por la Astronomía, la secuenciación del DNA en genómica, las tecnologías inalámbricas y los sensores digitales y satelitales en geociencias y la modelación computacional en meteorología (op.cit., pág.3).

Pues bien, todo pareciera indicar que la profunda transformación del quehacer científico afecta hoy no sólo a la organización de la ciencia (marcada por la emergencia de redes mundiales de organización y acceso privado al conocimiento), sino también al modo mismo de producción de conocimiento. **Los datos generados por los investigadores cobran una trascendencia inimaginada en décadas pasadas.** Si anteriormente, los investigadores producían conocimiento apelando a estados del arte ubicados en repositorios – impresos primero, digitales después- de publicaciones, hoy son los datos de investigación los requeridos para avanzar en la producción de ciencia. Esto es, la interdependencia y actuación en redes no es ya un dispositivo de organización, sino también y cada vez más, un mecanismo de producción de conocimiento.

Este proceso mundial, generador de redes internacionales de comunidades de ciencia, tecnología e innovación (OECD, 2007) viene a reconfigurar de manera compleja todo el proceso de creación de conocimiento científico (e-ciencia), de distribución del mismo (revistas científicas), de producción de relaciones jerárquicas en organizaciones científicas (sociedades y congresos internacionales), así como un profundo llamado a reorientar las preguntas científicas hacia requerimientos públicos locales (prioridades nacionales, federales, regionales) o mundiales.

En este contexto, Chile se encuentra haciendo profundos esfuerzos para transitar hacia la sociedad del conocimiento. Ello implica un conjunto de complejos matices de transformación. Desde el diseño de la arquitectura institucional hasta la cultura científica hay un conjunto de procesos emergentes y transformaciones en proceso y requeridas. De un lado, cambios en la arquitectura del Sistema Nacional de

Innovación y en los Sistemas Regionales de Innovación, redefinición de instrumentos de fomento a la ciencia, tecnología e innovación, búsqueda de mecanismos que aseguren el escalamiento en la producción de innovaciones tecnológicas de alto impacto y el desarrollo de ajustes institucionales en el ámbito educacional y productivo. Así, nos encontramos en Chile en medio de un proceso de profunda transformación del paradigma científico y tecnológico que incluye un cambio en las bases mismas de producción de conocimiento (la naturaleza del conocimiento, los mecanismos de producción del mismo y las tecnologías que inciden en su producción y difusión) y un cambio en el entorno institucional (cambio en la arquitectura institucional, en el marco normativo, en la cultura científica y en las prioridades de ciencia, tecnología, innovación y desarrollo productivo). El modo como se articulen estos dos componentes de transformación, definirá de manera importante el futuro de la ciencia en Chile.

Es en este marco donde es posible comprender las palabras de Ángel Gurría, en el discurso de Adhesión de Chile a la OECD:

“La determinación de construir una economía más fuerte y de enfrentar los retos domésticos e internacionales le ha servido a Chile en su caminar hacia los altos estándares de la membresía de la OCDE. Chile ha hecho avances significativos en un número importante de áreas, incluyendo cambios legislativos en respuesta a los requerimientos de los Comités de la OCDE. ... Estos son pasos importantes, que requirieron dedicación y una fuerte voluntad política. Ellos apuntan claramente hacia la implementación de mejores políticas.”¹

Nos encontramos así en la actualidad, entre otras, en la “travesía de las políticas de gestión de datos de investigación e información científica”. Esta, se encuentra cruzada por las transformaciones de la ciencia, de los tipos y estilos de política pública, de los requerimientos internacionales, pero también de las capacidades y demandas productivas, científicas y de innovación existentes al interior del país. Por ello, dadas las transformaciones en proceso, es posible establecer una **hipótesis prospectiva**:

El Acceso a Datos de Investigación se articulará a desarrollos emergentes ya instalados en Chile en algunos campos disciplinarios, así como encontrará una vía en el desarrollo de Plataformas Emergentes de Ciencia y Tecnología en el marco de Programas Nacionales de Ciencia y Tecnología disciplinarios y/o sectoriales de soporte de los Sectores productivos con potencial de crecimiento, definidos en la Estrategia Nacional de Innovación.

¹ Discurso de Ángel Gurría, en http://www.oecd.org/document/36/0,3343,en_2649_34487_44280356_1_1_1_1,00.html. Accesado el 15 de Diciembre, 2009.

Ahora bien, desde el punto de vista internacional, Chile debe avanzar hacia la instalación de estándares de gestión de datos de investigación e información científica, indicados por la propia OECD:

En este marco, el presente documento, sintetiza los resultados del estudio *“Consultoría Estado del Arte Nacional e Internacional sobre Manejo y Políticas de Acceso a Datos de Investigación e Información Científica y Tecnológica Financiada con Fondos y Recomendaciones de Buenas Prácticas”*.

El Objetivo central del estudio es *Conocer el estado del arte nacional e internacional sobre manejo y políticas de acceso a datos de investigación e información científica financiado con fondos públicos, que permita elaborar una política nacional de datos de investigación e información científica y tecnológica*.

El informe, pone a disposición los siguientes elementos:

Reflexiones teóricas, comunidades científicas,

redes e información: se propone una breve reflexión, relevante a la hora de comprender el impacto en la conformación y evolución de las comunidades científicas nacionales que genera y generarán los datos de investigación e información científica.

Metodología para el análisis de capacidades de Gestión de Datos de Investigación e Información Científica:

se incorpora la metodología del Benchmarking Internacional y el levantamiento de información de las capacidades chilenas.

Benchmarking Internacional en Gestión de Datos de Investigación e Información Científica: se analiza la evolución de la temática y los principales aprendizajes disciplinarios e interdisciplinarios

Capacidades Chilenas de Gestión de Datos de Investigación e Información Científica: se identifican y describen las capacidades chilenas a partir de encuestas nacionales y entrevistas realizadas.

Aprendizajes Internacionales y Buenas Prácticas: se sintetizan las principales recomendaciones para el diseño e implementación en Chile de una Política de Gestión de Datos de Investigación e Información Científica y Tecnológica.

ESTANDARES OECD EN GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

- **Apertura:** conciliar el interés de un acceso abierto a los datos permitiendo acrecentar la calidad y la eficacia de la investigación y la innovación, y la necesidad de restringir el acceso en ciertos casos, con el objeto de proteger el interés social, científico o comercial.
- **Transparencia:** hacer disponibles y accesibles a nivel internacional las informaciones sobre las organizaciones que producen los datos, la documentación sobre los datos que ellas producen, y las especificaciones de las condiciones inherentes a la utilización de estos datos.
- **Responsabilidad Formal:** proveer reglas institucionales formales explícitas sobre las responsabilidades de las diversas partes que intervienen en las actividades ligadas a los datos, que involucran: la autoridad, la mención de los productores, la propiedad, las restricciones con respecto al uso, las modalidades financieras, las reglas éticas, las condiciones de licencias y responsabilidad civil.
- **Profesionalismo:** desarrollar reglas institucionales para la gestión de datos digitales producto de la investigación basada en normas profesionales aplicables y sobre valores inscritos en los códigos de conducta de las comunidades científicas implicadas
- **Protección de la propiedad intelectual:** describir los medios para obtener un acceso abierto en el contexto de los diferentes regímenes jurídicos del copyright u otras legislaciones sobre la propiedad aplicables a las bases de datos, así como aquellos que están en relación con la protección del secreto de fabricación.
- **Interoperabilidad:** en cooperación con otras organizaciones internacionales, prestar debidamente atención a la necesidad de disponer de normas internacionales pertinentes y de uso común.

REFLEXIONES TEÓRICAS. COMUNIDADES EPISTEMICAS, REDES Y MARCOS INSTITUCIONALES EN EL ACCESO A DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

El acceso a datos de investigación e información científica, como un fenómeno propio del impacto de las tecnologías de información en la organización y producción misma del conocimiento científico y tecnológico, ponen en el centro el problema de las redes. En los estudios sociales de la ciencia y la tecnología, en la propia teoría de redes sociales y en el neoinstitucionalismo económico, viene desarrollándose una profunda reflexión sobre el rol de las redes en la construcción del conocimiento, en la difusión de información-conocimiento (e innovación) y en la tipología de redes. Esto implica también un campo de reflexiones sobre el carácter de bien público de los datos de investigación e información científica. Por ello, la presente sección, propone una síntesis de estas aproximaciones.

La propiedad del conocimiento es un aspecto de no fácil solución debido a la misma complejidad de la generación del conocimiento. La taxonomía entre conocimiento básico y aplicado y su posterior categorización entre bien público, sujeto a libertad académica, y el conocimiento patentable o apropiado por el sector privado no es divisible totalmente al igual que la distinción descubrimiento/invento, la cual en la actualidad es mucho más compleja ya que el definir *descubrimiento* como el avance del conocimiento en el funcionamiento en los sistemas naturales y sociales, e *invento* como un producto o proceso orientado a un determinado fin no genera en todos los casos la posibilidad de separar rígidamente ambas dimensiones. Un ejemplo de esta situación es la ocurrida el año 2.000 cuando se dieron a conocer 355.000 secuencias de ADN protegidas por patentes lo cual para algunos es una mala aplicación de la propiedad intelectual y para otros correcta ya que de ella deriva conocimiento aplicado velozmente como la ingeniería genética, lo cual la dejaría en el carácter de invento².

Una red se entiende como un conjunto de nodos vinculados entre sí donde los nodos corresponden a actores individuales y/o corporativos que movilizan recursos en torno a ciertos objetivos determinados. Bajo este supuesto se generan interacciones desde las cuales surgen estructuras reticulares que pueden tener efectos positivos o negativos. Estos efectos o externalidades de las redes, son destacadas por su importancia para la gestión de políticas de I+D+I, y el diseño de metodologías de networking. A continuación, se identifican estas externalidades:

- **Formación de confianza como base para la cooperación:** En las redes uno de los recursos más significativos que circula es la información. Si la estructura de una red presenta la suficiente densidad (y la información que circula no es redundante) y/o existen actores que tienen un papel de puentes o conectores que vinculan a subgrupos o cliques no directamente conectados entre sí, es probable que esta

Desde el punto de vista de la Gestión de Datos de Investigación e Información Científica, la identificación de redes tecnocientíficas será relevante como Proxy para la comprensión de la estructura que organiza ciertos campos científicos y será a través de esta estructura que circulará la información y los datos de investigación. Al existir redes científicas y tecnológicas de investigación, estas facilitarán la circulación y por ende, al existir mecanismos de confianza instalados o en proceso en estas redes, entonces existirán nodos con capacidades distribuidas para la captura, almacenamiento, procesamiento y diseminación de datos de investigación e información científica.

² Para una mayor profundización de esta materia en el campo de la genómica véase: Bien privado, bien colectivo y bien público en la era de la genómica. Maurice Cassier. Organización de los Estados Iberoamericanos.

estructura facilite la circulación de información acerca de las preferencias y competencias de los actores. Si esta información muestra que los actores tienen suficientes credenciales y/o que cumplen sus compromisos, se les considerará confiables y se puede producir un depósito multilateral de confianza (entendida como una expectativa sobre las preferencias y comportamientos -Herreros, 2002) entre los actores que integran la red.

Esta confianza puede reforzarse si se generan interacciones iteradas con resultados eficientes, produciéndose así una base para la cooperación condicional y estable, pues, en términos de teoría de juegos, la mutua convergencia de expectativas reducen el riesgo en los intercambios y genera equilibrios en que las partes involucradas obtienen ganancias. En este escenario es posible la generación compartida de innovaciones de procesos y productos y la difusión de innovaciones. Esta lógica de situación favorecida por la estructura de las redes, está a la base del compartir know-how en *joint venture* (Porter y Fuller, 1986) y de la gestión eficiente en los *clúster* (Gordon y McCan, 2000).

- **Eficiencia colectiva.** Las redes pueden concebirse como estructuras de gobernanza que, utilizando el lenguaje de la nueva economía institucional (Williamson, 1975), están entre el mercado y la jerarquía. Estas redes, al tratarse de vínculos con historia y continuidad, pueden generar procesos de intercambios que, en términos de la nueva sociología económica (Granovetter, 1985; Nee, 2005), están incrustados en las relaciones sociales. Esto significa que las redes *multiplex*, es decir con diversos tipos contenidos relacionales, favorecen la creación de confianza, la que permite reducir los costes de transacción derivados de hacer cumplir los compromisos. Es posible que esta reducción de costes por el lado de la confianza, sea acompañada por adecuados arreglos institucionales formales como contratos e incentivos, que al generar ganancias mutuas repetidas, reducen la incertidumbre de los actores implicados en las redes (MacCann, Tomokazu y Gordon, 2003; Smitka, 1991), lo que favorece intercambios de recursos (por ejemplo el intercambio intensivo de información). Esta reducción de costes de transacción sumada a las economías de escala logradas gracias a la localización en el caso de las empresas aglomeradas (que incluye además, asumir en conjunto, por ejemplo, costes logísticos de transporte, telecomunicaciones, etc.) hacen que se produzca en forma intencionada (eficiencia colectiva activa) una gestión eficiente en toda la red, es decir se produce eficiencia colectiva.

Es necesario contar con arreglos institucionales formales, capaces de generar confianza entre los agentes que comparten datos de investigación. Así mismo, estos mecanismos, variarán según la estructura de redes por las que circulan esos datos e información y con ello, los incentivos que se generen para potenciar estos procesos.

- **Aprendizaje colectivo.** El aprendizaje colectivo o interactivo es posible gracias a los vínculos y la historia de las relaciones que sostienen diversos actores implicados en procesos de innovación. Además de los marcos institucionales que faciliten los intercambios (Lundvall, 1992; Jonson, 1992) y de la existencia de una “*cultura local*” (Pilon y DeBresson, 2003), la información que circula a través de redes densas pero no jerárquicas y con nodos que permiten el acceso a nuevo conocimiento, permiten que la información circule adecuadamente y se generen trayectorias que favorezcan la acumulación de conocimiento tácito y codificado común (Dupuy y Gilly, 1996). Las condiciones propuestas por Lawson y Lorenz (1999) como necesarias para el aprendizaje organizacional, esto es, compartir conocimiento y generar nuevo conocimiento combinando diversos conocimientos, también son requeridas para el aprendizaje colectivo como insumo para la innovación en las redes.

Las redes que cuentan con trayectorias de trabajos colaborativos, serán plataformas más fecundas para la circulación de datos e información científica. Así mismo, serán fundamentales las capacidades de las redes para generar aprendizaje, esto es, producir conocimiento científico con los datos e información que por la red

- **Fortalecimiento de *comunidades epistémicas* como creadoras de nuevo conocimiento codificado.** Las *comunidades epistémicas*, comunidades o redes de práctica, subgrupos cognitivos o colegios invisibles, pueden concebirse como redes semi-cerradas de especialistas (profesionales) en un área de conocimiento y/o producción, que, en base a vínculos cara a cara, de co-presencia y co-localización y teniendo como base un status similar y un reconocimiento mutuo, comparten información estratégica (recomendaciones, oportunidades de acceso a recursos) y conocimiento tácito y codificado (Brown y Duguid, 1991; Schrader, 1991; Wenger, 1998). La investigación empírica ha destacado dos propiedades estructurales de redes que favorecen la creación de nuevo conocimiento codificado entre comunidades epistémicas. Estas propiedades son alta densidad en los cliques (subgrupos) que conforman los especialistas y además, la existencia de nodos que tengan un alto valor de intermediación, es decir que conecten al subgrupo con otros subgrupos o comunidades (Giulliani, 2005; Morrison, 2005).

Los campos científicos cuentan con estructuras construidas que facilitan la circulación de conocimiento, y esas estructuras son las aprovechables para el despliegue de plataformas de acceso a datos e información científica, en la medida que el carácter semicerrado de la comunidad, permite potenciar procesos ya generados de confianza, aprendizajes y especializaciones de conocimiento apropiado que requieren datos.

Así, si las redes y comunidades epistémicas constituyen la estructura social disponible en la ciencia, es en y a través donde la información, y los datos, circulan y es posible hacer producir, almacenar, procesar y difundir datos de investigación e información científica.

II. METODOLOGÍA GENERAL DEL ESTUDIO

II.1. METODOLOGÍA PARA EL BENCHMARKING INTERNACIONAL

La realización del Benchmarking Internacional comprende distintas etapas y procesos, los cuales se realizaron a lo largo del estudio, efectuando lo siguientes análisis:

- **Construcción de una base bibliográfica:** recopilación y selección de literatura acorde a los objetivos del estudio. En éste punto se contó con la colaboración del experto internacional (Paul Uhlir), principalmente en la recomendación de documentación actualizada y relevante en el campo de tendencias globales a partir del análisis del impacto de la Globalización, las TICs, y la emergencia de redes internacionales de Ciencia, Tecnología e Innovación. además de la identificación de Políticas y experiencias Internacionales de Gestión de Datos de Investigación e Información Científica.
- **Análisis evolutivo de las Políticas y experiencias internacionales.** Se elaboró un modelo que identifica la evolución en el tiempo de las principales temáticas asociadas a la Gestión de Datos de Investigación e Información Científica.
- **Identificación de Políticas y experiencias Internacionales de Gestión de Datos de Investigación e Información Científica.** Se orientó a detectar buenas prácticas en experiencias disciplinarias y transdisciplinarias.

El principal producto del Benchmarking Internacional corresponde a un reporte sobre el estado del arte internacional en la materia del estudio, señalando buenas prácticas y políticas recomendadas por instituciones u organizaciones con credibilidad internacional. Este análisis incluyó una revisión de las instituciones, políticas, normas y mejores prácticas internacionales en el tema, así como las prácticas y recomendaciones de organismos de reconocido expertos en políticas públicas y el tema de este estudio como OECD, RICYT, ICSU, CoData, etc.

Además se distinguieron disciplinas, áreas del conocimiento, sectores de aplicación y descripción del marco legal e institucional internacional en cuanto a las prácticas de acceso a información primaria desde los diversos puntos de vista involucrados (derechos de propiedad intelectual, resguardos de privacidad y confidencialidad para los sujetos de investigación, secreto estadístico, etc.).

La principal aproximación corresponde a la identificación de modelos de política pública y experiencias de proyectos a nivel internacional en Gestión de Datos de Investigación e Información Científica, lo que resulta fundamental para generar tanto aprendizajes institucionales, como para identificar estándares internacionales que deberán ser implementados en el caso de Chile.

Para avanzar en este sentido, se ha sistematizado literatura internacional sugerida tanto por la contraparte técnica como por el asesor internacional del estudio Paul Uhlir. La Bibliografía se organizó en dos campos:

- Identificación de experiencias internacionales.
- Análisis evolutivo de experiencias internacionales

Respecto de la identificación de experiencias internacionales, el procedimiento para el ordenamiento y posterior análisis fue el siguiente: se construyó una matriz que permitió identificar cada una de las experiencias identificando respectivamente en cada columna:

- a) **NOMBRE:** nombre de la experiencia o texto trabajado.
- b) **FOCO TEMATICO:** eslabón de la cadena de gestión de datos e información al cual se focaliza la experiencia o la discusión del texto: captura, almacenamiento, procesamiento.
- c) **ACTORES:** instituciones participantes en la experiencia y el rol que cumplen.
- d) **AREA CIENTIFICA:** categorías de conocimiento en las cuales se desarrolla la experiencia.
- e) **RESUMEN:** descripción sintetizada del proceso.

Respecto del modelo evolutivo desarrollado, se construyó una visualización gráfica del proceso considerando los siguientes elementos:

i) LINEA DE TIEMPO: construcción de una línea de tiempo para comprender la dinámica de los distintos procesos y experiencias la cual comprende desde fines de los 80' hasta la actualidad debido a que, en dicho período, es donde se encuentra la mayor cantidad de referencias especializadas en la temática.

ii) PROCESOS: aspectos relacionados con el operar de las políticas de gestión de datos e información científica.

ii.1. Componentes de Gestión: se refiere al o los ámbitos de la cadena para la gestión de datos e información: captura, almacenamiento, procesamiento y diseminación, que han estado mayormente presente en las políticas y/o experiencias.

ii.2. Tecnologías: se refiere al tipo de tecnologías e infraestructura utilizada mayormente para la Gestión de Datos de Investigación e Información Científica.

ii.3. Problemáticas: se refiere a los desafíos y dificultades que han emergido en los procesos de captura, almacenamiento, procesamiento y diseminación de datos de investigación e información científica.

iii) Experiencias: casos y experiencias características y relevantes y pertinentes de gestión de datos e información científica.

A partir de los procedimientos y fuentes de información arriba definidos, se dio cuenta del estado del Arte Internacional de la siguiente forma:

- Análisis de tendencias globales de las tecnologías de la información y conocimiento en el contexto de la globalización de un modo general. Realizando un análisis de tendencias internacionales globales de la ciencia (de la big science a las redes de investigación a las redes de acceso a datos),

de la tecnología (impactos del desarrollo tecnológico en las TICs). Finalmente, se analizó el origen de los procesos de acceso a datos de investigación.

- Posteriormente se realizó la primera revisión del estado del arte mediante la descripción y análisis de las principales experiencias de gestión de datos e información científica. Esta literatura versa tanto sobre experiencias concretas de gestión de datos e información como reflexiones de la temática a nivel de políticas públicas, tecnologías, levantamiento de plataformas virtuales, rol de los actores ligados a la gestión de datos e información.
- Explicación, análisis y gráfica de la evolución de las experiencias de gestión de datos e información científica revisadas en la sección anterior mediante un modelo que sintetiza procesos, tecnologías y experiencias en una perspectiva evolutiva.
- Finalmente se produce la identificación de aprendizajes emanados de la revisión del estado del arte de acuerdo a los componentes claves del estudio: Política, Capital Humano, Patrimonio, Infraestructura Tecnológica y Vinculaciones.

II.2 METODOLOGÍA PARA EL ANÁLISIS Y VISUALIZACIÓN DE LAS CAPACIDADES CHILENAS EN MATERIA DE GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA Y TECNOLÓGICA.

A continuación se realiza una breve descripción de la metodología para el levantamiento de información nacional. En un primer momento corresponde a la metodología para el análisis de las capacidades chilenas en materia de Gestión de Datos e Información Científica y Tecnológica, desde el levantamiento de información, la sistematización y el análisis, mediante la utilización de niveles de información relevante: estadísticas globales, análisis según disciplinas científicas, y análisis según la cadena de valor de gestión de datos de investigación e información científica y tecnológica. Además del protocolo de aplicación utilizado para los distintos instrumentos utilizados en la recolección de información, los cuales corresponde a:

- Encuesta a investigadores
- Encuesta a instituciones (información científica)
- Encuesta a instituciones (datos de investigación)
- Entrevista a actores relevantes

Mientras que en un segundo momento se indica la metodología utilizada en el diseño de la Base de Datos y el Modelo de Visualización desarrollado.

II.2.1 Metodología para el análisis de capacidades chilenas en gestión de datos de investigación e información científica y tecnológica

La identificación de capacidades en la Gestión de Datos de Investigación e Información Científica y Tecnológica en Chile implica un proceso complejo que incluye tres ejes de análisis que se retroalimentan pero poseen un margen de diferenciación. El primero enfocado a la descripción de los principales procedimientos y protocolos para la recolección de información. El segundo, orientado a relacionar el estado de capacidades de acuerdo a las áreas científico-tecnológicas en que se expresan y un tercero encargado de analizar la información desde el punto de vista de los eslabones de la cadena de valor de la gestión de datos de investigación e información científica³.

En la primera parte se exponen los principales ámbitos del estudio. En la segunda se detallan las distintas etapas y tareas de la metodología de análisis, desde la identificación de los contactos hasta la forma de visualizar y analizar los resultados. En la tercera parte, se expone la arquitectura del proceso metodológico, las variables de estudio y la explicación de los distintos niveles de análisis junto a sus indicadores, incluyendo el protocolo de aplicación de los distintos instrumentos. Finalmente en la cuarta parte, se detalla el proceso de aplicación de los distintos instrumentos.

Ámbitos de análisis

La recolección de información contempla variables comprendidas en los ámbitos claves del estudio:

- **Institucionalidad:** nivel de institucionalidad de las unidades o departamentos con labores de gestión de datos de investigación y/o información científica, junto al nivel de exclusividad que poseen estas labores en relación con otras tareas y funciones.
- **Capital humano:** características de los recursos humanos responsables de la gestión de datos de investigación y/o información científica, pertinencia de profesionales, capacidades, habilidades técnicas, etc.
- **Infraestructura tecnológica:** características y pertinencia del equipamiento asociado a la gestión de datos de investigación y/o información científica, computadores, servidores, PDA, software, etc.
- **Patrimonio:** características del patrimonio generado con fondos públicos y las prácticas de diseminación asociado a éste.
- **Vinculación:** características de los proyectos, programas y/o convenios de cooperación con otras instituciones ligadas a la gestión de datos de investigación y/o información científica, además de la percepción respecto a los beneficios que otorgan las vinculaciones interinstitucionales para estas labores.
- **Políticas de gestión:** existencia y características de políticas o procedimientos internos que regulen y estandaricen la gestión de datos de investigación y/o información científica.

³ La cadena de valor de la gestión corresponde a los siguientes eslabones: planificación, captura, almacenamiento, procesamiento, validación y diseminación

Etapas de análisis

El proceso de análisis de los resultados de las encuestas y entrevistas comprendió seis etapas:

Etapa 1: Identificación de contacto y de instituciones.

Etapa 2: Contacto y selección de instrumentos a aplicar.

Etapa 3: Aplicación de instrumentos.

Etapa 4: Sistematización.

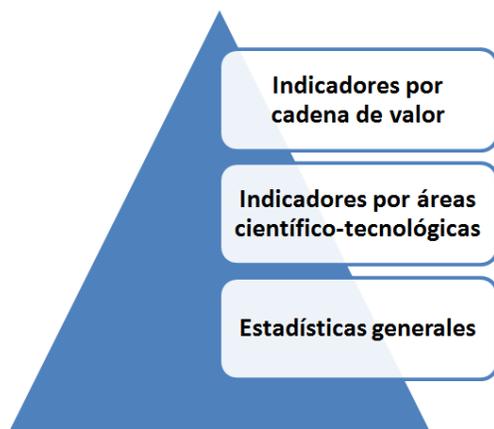
Etapa 5: análisis de capacidades.

Etapa 6: Diseño y herramientas de visualización de información.

PLAN DE ANÁLISIS: SET DE INDICADORES PARA LA IDENTIFICACIÓN DE CAPACIDADES

El plan de análisis para los resultados arrojados por las encuestas y entrevistas posee tres niveles que profundizan en distintas capas de la información recolectada. El primer nivel está enfocado a un rastreo de variables claves por cada uno de los seis ámbitos de estudio. El segundo, a la construcción de indicadores que permiten el análisis de éstas pero en relación a las áreas científicas desarrolladas por las instituciones chilenas. Finalmente, en el tercer nivel, igualmente se analizó la relación entre las variables más importantes pero ahora en relación con los eslabones de la cadena del valor de la gestión de datos de investigación e información científica.

Niveles de Análisis de capacidades chilenas



Fuente: elaboración propia

A continuación se detallarán cada uno de estos niveles junto a sus medidas de análisis:

1°.- Estadísticas generales: En este nivel se recogen las principales variables por ámbito de forma desagregada, lo que corresponde a una primera aproximación al estado de las capacidades de las instituciones chilenas en su gestión de datos de investigación e información científica.

2°.- Indicadores por áreas científico-tecnológicas: en este nivel se construyen indicadores que permiten el análisis de las principales variables por ámbito pero en su relación con las áreas científicas OCDE para identificar potencialidades focalizadas en disciplinas científicas extrayendo así fortalezas y debilidades.

3°.- Indicadores por cadena de valor: al igual que en el anterior, en este nivel se analizará la relación entre las principales variables y su anidación a uno o más eslabones de la gestión de datos de investigación y/o información científica: planificación, captura, almacenamiento, procesamiento, validación y diseminación

II.2.2. Encuesta a investigadores y académicos

Ámbitos de Análisis

Para conocer el estado de las capacidades de la gestión de datos de investigación e información científica en los académicos e investigadores chilenos se optó por una vía diferente en la forma de recolectar los datos. Se privilegió la cantidad de investigadores respecto al detalle del instrumento de recolección de información. Se aplicó una encuesta de aproximadamente 15 preguntas a casi cinco mil investigadores y académicos en Chile donde los principales ejes a trabajar fueron los siguientes:

- **Caracterización:** datos del investigador junto a su propia clasificación respecto a las disciplinas OCDE de acuerdo a las labores que desempeña.
- **Tipo de gestión:** identificación de la existencia y modalidad de gestión, ya sea en datos de investigación, en información científica, en ambas o ninguna.
- **Capital Humano:** grado de conocimiento de conceptos fundamentales que sustentan las actuales políticas internacionales de gestión de datos de investigación y/o información científica: Open Access, Derecho de Autor y Ley de Transparencia
- **Patrimonio:** características en el acceso y prácticas de diseminación del patrimonio recolectado por los investigadores en sus instituciones gracias al financiamiento público.
- **Vinculación:** identificar la participación de los investigadores en redes, proyectos, programas o convenios de cooperación que permitan compartir datos de investigación y/o información científica.
- **Políticas de gestión:** identificar la existencia de políticas o procedimientos estandarizados en las instituciones de los investigadores respecto a la gestión de datos y/o información científica y conocer su parecer respecto a la existencia de políticas nacionales que aborden estas temáticas.

Plan de análisis para encuesta a investigadores.

Set de Indicadores

El análisis que para identificar las capacidades en la gestión de datos de investigación y/o información científica en Chile posee los dos niveles explicados anteriormente en el caso de las instituciones:

Niveles de análisis encuesta a investigadores



Fuente: elaboración propia

1°.- Estadísticas generales: En este nivel se recogen las principales variables por ámbito de forma desagregada, lo que corresponde a una primera aproximación al estado de las capacidades de los investigadores/académicos en su gestión de datos de investigación e información científica.

2°.- Indicadores por áreas científico-tecnológicas: en este nivel se construyen indicadores que permiten el análisis de las principales variables por ámbito pero en su relación con las áreas científicas OCDE para identificar potencialidades focalizadas en disciplinas científicas extrayendo así fortalezas y debilidades.

NOTA: EL DETALLE DE INDICADORES Y SU ANALISIS, SE ENCUENTRA DISPONIBLE EN EL DOCUMENTO CAPACIDADES CHILENAS DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA. ENCUESTA A INSTITUCIONES E INVESTIGADORES.

III. ESTADO DEL ARTE INTERNACIONAL EN MATERIAS DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA Y TECNOLOGICA

La Gestión de Datos de Investigación e Información Científica, ha venido desarrollándose tanto por requerimientos propios de la naturaleza de la investigación científica y tecnológica actual, al impacto de las TICs, y también en el direccionamiento y regulación pública de la actividad. Si bien las tendencias globales empujan a la actividad científica de avanzada a compartir datos e información, existen una serie de problemas emergentes que contienen su desarrollo (indicadas en la Tabla Nº 1). Estas, indican que el desarrollo tecnológico potencia tanto el desarrollo del conocimiento científico (cambiando incluso la naturaleza de creación del conocimiento), como la estructuración de comunidades epistémicas, ahora de naturaleza global. El escalamiento del conocimiento así, gatilla el propio desarrollo tecnológico de la infraestructura de redes y la disposición de tecnologías para la captura, procesamiento y diseminación del mismo, pero en este proceso, emergen problemas que ponen en el centro la capacidad de planificación, tanto en la selección de tecnologías para la mejor gestión de la cadena de valor, como para la indexación y documentación de datos e información, así como en torno a la anticipación de requerimientos científicos y acceso a datos en tiempo real (junto por ejemplo, a las decisiones de que almacenar).

| Tendencias y problemas internacionales en la gestión de datos de investigación e información científica | |
|---|---|
| Tendencias globales | Problemas |
| <ul style="list-style-type: none"> • Disminución en el costo de los equipos computacionales y de comunicación • Aumento de las capacidades para la obtención de datos científicos • Instauración de las comunicaciones digitales inalámbricas • Aumento en la explotación de las redes de banda ancha y de las capacidades de transmisión de datos de video • Cambios en el dominio de las redes de datos • Aumento en el soporte técnico para el trabajo colaborativo • Aumento en las capacidades para el procesamiento de lenguaje natural • Reconocimiento creciente de la importancia de los estándares • Cooperación en el monitoreo y control de la actividad de las redes • Aumento en el uso de redes internas (intranets) | <ul style="list-style-type: none"> • Aumento en la congestión de la internet • Inadecuada descripción e indexación de datos • Rápida obsolescencia de los medios electrónicos de almacenamiento • Vulnerabilidad de las redes electrónicas de datos • Potencialidad de que los requerimientos científicos de hardware no sean alcanzados • Falta de suficientes redes de datos internacionales en tiempo real |

Fuente: Elaboración propia IDER/UFRO, 2009

Todo ello pone en el centro dos series de temáticas relevantes. En primer lugar, las experiencias disciplinarias indican que las tendencias y problemas cobran particularidades anidadas a la naturaleza del conocimiento en cuestión de las disciplinas, pero también (en segundo lugar), las capacidades de regulación, incentivos y resguardos de los procesos de gestión de datos de investigación e información científica.

En este contexto, el presente capítulo analiza el estado del arte internacional en tres momentos sucesivos:

Primero. El análisis histórico-evolutivo. Es relevante identificar las trayectorias tecnológicas, científicas y de regulación que hacen comprensible el estado presente y las tendencias de futuro.

Segundo. El examen de experiencias disciplinarias, permite tanto identificar experiencias relevantes, como caracterizar sus tendencias y problemas.

Tercero. El aprendizaje internacional en la óptica de gestión de cadenas de valor del acceso a datos de investigación e información científica, permite poner el acento en las materias regulatorias de los diferentes eslabones de la cadena.

A ello se dedican las siguientes secciones del capítulo.

III.1. LA GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA DESDE UNA ÓPTICA HISTÓRICO-EVOLUTIVA.

En la presente sección se describe y analiza de manera temporal y detallada los aspectos más importantes de las experiencias internacionales. El análisis comienza con el comportamiento de la gestión de datos e información durante la década de los 90' y continúa con la última década, para lo cual se propone un modelo gráfico que comprende los principales aspectos del análisis como guía de lectura.

i.- Finales de los 80' y la década de los 90'

Desde finales de los 80' y durante toda la década de los 90' los principales componentes de la cadena de Gestión de Datos de Investigación e Información Científica es el almacenamiento y procesamiento. Se debe considerar que las características de aquel entonces distan mucho de la enorme velocidad y configuración global del Internet actual, lo mismo que con las tecnologías de procesamiento de información.

El caso del **Laboratorio de Fermilab**⁴ durante los 90' es muy interesante para conocer el estado del procesamiento y disseminación utilizado para los datos ya que todo el proceso prácticamente dependía de estructuras físicas como las cintas de 8mm las que permanecían en el mismo laboratorio para su análisis. Sin embargo, como en los proyectos participaban varias instituciones situadas en distintos países, era necesario transportar los datos en maletines para poder analizarlos en otras dependencias. Esto además de ser lento, es riesgoso por la posibilidad de extravío o deterioro de los archivos. La manera de corregir este problema fue el uso de la videoconferencia pero en la década de los 90' ésta aún era demasiado costosa debido al estado del Internet de aquel entonces. En este mismo caso, una forma efectiva de notificación a las demás instituciones e investigadores sobre los avances realizados fue mediante el uso de LaTeX⁵ con postscripts outputs⁶ y código ACSII⁷. Gracias a estos complementos tecnológicos fue posible acceder a los más recientes análisis y sus resultados a los pocos minutos de su ingreso al sistema.

⁴ National Research Council (1997). *Bits of Power: Issues in Global Access to Scientific Data*. Committee on Issues in the Transborder Flow of Scientific Data. Washington, D.C., National Academy Press.

⁵ **LaTeX** es una potente herramienta de edición de textos científicos. Es un lenguaje de marcado para documentos, y un sistema de preparación de documentos, formado por un gran conjunto de macros de TeX, escritas inicialmente por Leslie Lamport (LamportTeX) en 1984, con la intención de facilitar el uso del lenguaje de composición tipográfica. Es muy utilizado para la composición de artículos

Otro aspecto de la época, más relacionado con la gestión de información científica, es la aparición de los archivos Eprint⁸ los cuales ofrecen la posibilidad de ser corregidos de varias fuentes antes de ser publicados. Esto flexibiliza y refina el trabajo realizado por los investigadores en la materialización de sus estudios al poder interactuar con pares antes de publicar de manera definitiva. Este sistema fue posible gracias al gran desarrollo y conectividad global de Internet.

Pensando en la colaboración público-privada se encuentra el caso de intercambio de información ligado a las ciencias químicas denominado: **Secuencias Genómicas**⁹, en el cual instituciones de distintos países, mediante financiamiento estatal, se hacen responsable de recolectar datos generados tanto por instituciones académicas, gubernamentales como laboratorios industriales, para luego traspasarla a un formato estándar y finalmente distribuirla diariamente de tal forma de mantener una sola base de datos internacional, común y asequible en Internet en cualquier momento.

En esta década, específicamente en el componente técnico, la emergencia de motores de búsqueda es uno de los avances más importantes debido a la gran e imparable proliferación de información. Estos buscadores se pueden clasificar en Índices temáticos¹⁰ y Motores de búsqueda¹¹. El primero en inventarse fue *Wandex* en 1993 por *la World Wide Web Wanderer* el cual actualmente no se encuentra vigente. Otros de aquella época fueron *Aliweb*¹², *Lycos*¹³ y *WebCrawler*¹⁴, este último como el precursor de texto completo.

Otra experiencia clave en la diseminación de datos en la década de los 90' es el **Archivo del Telescopio Espacial Hubble**¹⁵, en la cual los datos generados por el Telescopio Hubble son almacenados y diseminados a través del Instituto de Ciencias de Laboratorios Espaciales los que quedan disponibles para toda la comunidad de astrónomos y el público en general mediante Internet.

Como se puede apreciar, el aspecto fundamental a trabajar por estas experiencias, es el almacenamiento, procesamiento y diseminación de datos e información científica. Si bien la literatura no hace un hincapié en la captura no es equivalente a señalar que no fue un componente tratado, sino más bien, y como lo

académicos, tesis y libros técnicos, dado que la calidad tipográfica de los documentos realizados con LaTeX es comparable a la de una editorial científica de primera línea.

⁶ **PostScript** es lenguaje de programación que contiene un rico conjunto de operadores gráficos, fue diseñado y desarrollado por Adobe Systems Incorporated y fue utilizado por el público en general a mediados de 1980.

⁷ **ASCII**: American Standard Code for Information Interchange. Código de caracteres basado en el alfabeto latino. Fue creado en 1963 por el Comité Estadounidense de Estándares (ASA, conocido desde 1969 como el Instituto Estadounidense de Estándares Nacionales, o ANSI) como una refundición o evolución de los conjuntos de códigos utilizados entonces en telegrafía.

⁸ **EPrint** es una versión digital de un documento de investigación (generalmente un artículo de revista, pero también podría ser una tesis, conferencias papel, libro capítulo, o un libro) que es accesible en línea, ya sea en un repositorio local, o un (a reserva o disciplina basada) repositorio digital.

⁹ National Research Council (1997).

¹⁰ Sistemas de búsqueda por temas o categorías jerarquizados (aunque también suelen incluir sistemas de búsqueda por palabras clave). Se trata de bases de datos de direcciones Web elaboradas "manualmente", es decir, hay personas que se encargan de asignar cada página web a una categoría o tema determinado (wikipedia).

¹¹ Son sistemas de búsqueda por palabras clave. Son bases de datos que incorporan automáticamente páginas web mediante "robots" de búsqueda en la red (wikipedia).

¹² Aliweb: <http://www.aliweb.com/>

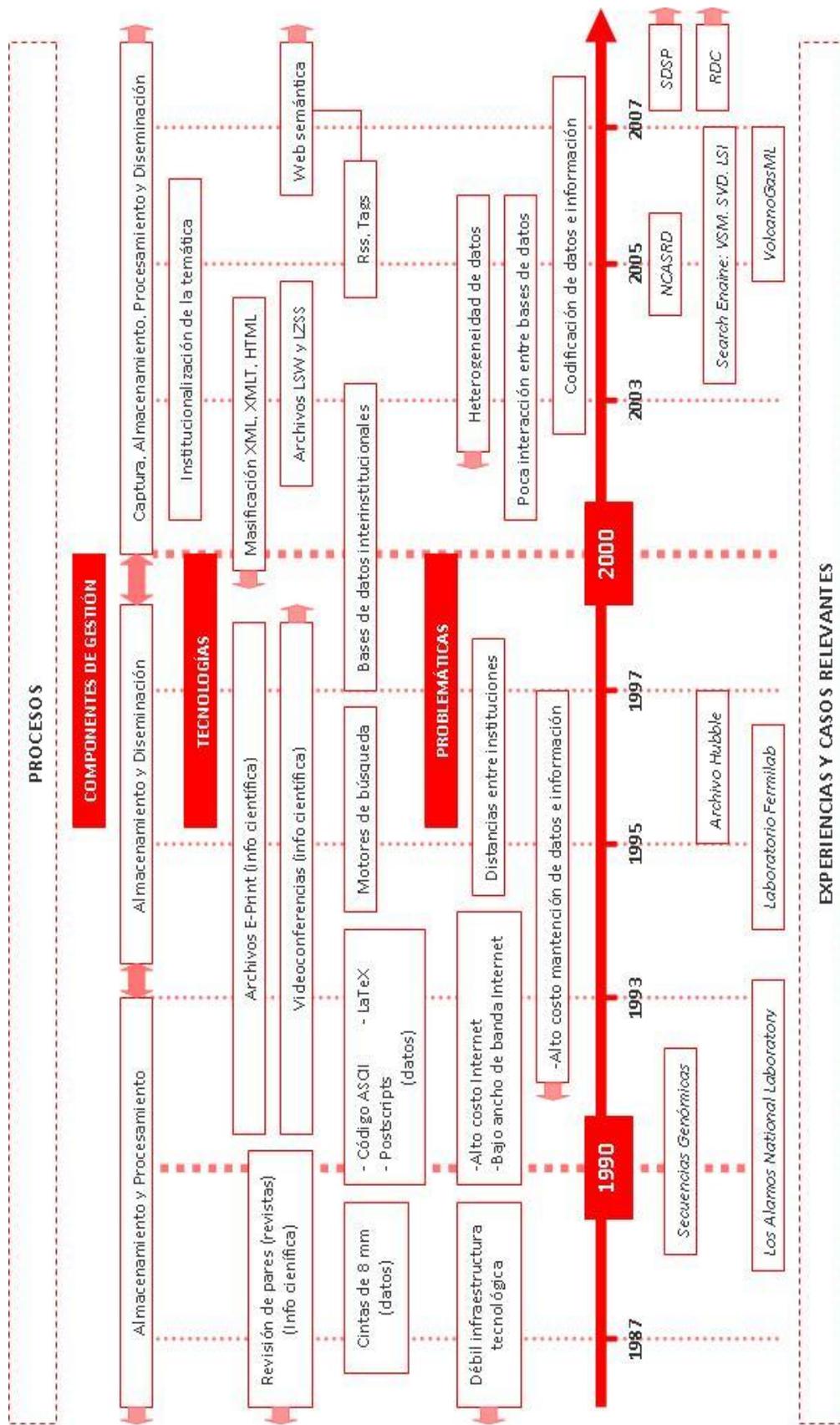
¹³ Lycos: <http://www.lycos.com/>

¹⁴ WebCrawler : <http://www.webcrawler.com/>

¹⁵ National Research Council (1997). Citado anteriormente.

demuestran las experiencias analizadas, los desafíos de la década de los 90' fueron esencialmente mejorar la velocidad y conectividad de entrega de información y datos para facilitar el análisis y el trabajo de diversas instituciones que tienen relación temática. El caso del archivo Hubble y las secuencias genómicas refiere a como comunidades de científicos que comparten un área temática interactúan en tiempo real con los mismos datos pero en diferentes lugares.

Otro aspecto importante a señalar es la participación **del sector público mediante el financiamiento** y sobre todo la capacidad de las **distintas instituciones partícipes de actuar sincronizadamente para la administración de los datos**. Esto permite que las instituciones puedan capturar, almacenar, procesar y diseminar datos para los cuales poseen mayores capacidades en su administración, transformando a cada institución en un engranaje con características particulares que ayudan a la conformación de una gran base de datos común, homogénea y de mayores facilidades y herramientas para su administración y análisis.



ii. Del 2000 hasta la actualidad

De los muchos fenómenos que ocurren en esta década respecto a lo que implica la Gestión de Datos de Investigación e Información Científica, uno de los más interesantes es la divulgación e institucionalización de la temática, es decir, el paso que se da desde la detección del problema de la gestión de datos e información hasta que se reconoce y se toman medidas para su solución. Se comienza progresivamente a considerar que para solucionar un problema de carácter sistémico no basta sólo con identificar el problema, sino además hacer reconocible el problema para todos los actores que son parte de él y su solución.

Es por esta razón que países como Canadá han abordado el problema inicialmente desde la sensibilización de los actores involucrados en la generación de conocimiento científico como son los investigadores, centros de investigación y los actores públicos y privados que financian o son partícipes de políticas de ciencia, tecnología e innovación. Para abordar este desafío, Canadá junto a todos estos actores configuró una política de gestión de datos e información científica los cuales se reunieron en un foro denominado: *National Consultation on Access to Scientific Research Data (NCASRD)*¹⁶, en el cual se propuso desarrollar una estrategia para garantizar el libre acceso y la preservación de la información financiada con fondos públicos y ejecutar una batería de iniciativas las cuales se han clasificado de la siguiente manera:

-Institucionalidad: crear un organismo que oriente y supervise la formación de una estructura permanente de la organización de acceso a datos en Canadá el cual debe velar por tareas como a) Institucionalización y sensibilización de la temática a nivel científico y político, b) otorgar financiamiento a iniciativas que aborden la gestión de datos e información, c) velar por la participación y distribución de los datos en redes internacionales y defender los intereses para Canadá, así como también los aspectos legales al interior del país. d) velar por la calidad y actualización de los datos.

-Política: otros aspectos sugeridos de igual importancia son a) financiar la formación y especialización de investigadores dedicados a la gestión de datos, b) incorporación de componentes que aborden el acceso a datos en las bases o términos de referencias de iniciativas de financiamiento para las instituciones científicas ejecutoras. c) revisión por pares de metadatos.

-Formación: a) otorgar incentivos a investigadores que contribuyan al mejoramiento de la gestión de datos e información científica y b) el desarrollo y énfasis a carreras universitarias que se relacionen con la gestión de datos e información científica como bibliotecología.

Al igual que Canadá, China también realizó una planificación al 2020 con sus expertos en gestión de datos e información pero mucho más focalizada y específica. El objetivo del Ministerio de Ciencia chino es desarrollar un gran sistema en red para la gestión de datos de investigación el cual incorpore la mayor cantidad de áreas científicas desarrolladas en ese país. Esta plataforma es denominada *SDPS: Scientific Data Sharing Program*¹⁷, y además de configurarse como un componente tecnológico, incorpora todo un

¹⁶ Sabourin, M. & Dumouchel, B. (2007). Canadian National Consultation on Access To Scientific Research Data. *Data Science Journal*, Vol. 6, pp.OD26-OD35.

¹⁷ Guan-Hua Xu (2007). Open Access to Scientific Data: Promoting Science and Innovation. *Data Science Journal*, Volume 6, Vol. 6 pp.OD21-OD25.

espectro de normativas que tiendan a regular y administrar el rol de cada actor participante. El sistema se propone construir con cuarenta centros de datos de investigación y trescientas bases de datos que se gestionaran y distribuirán mediante un gran portal web maestro.

Respecto a los componentes técnicos, en los últimos años se observa una gran interacción entre distintos tipos de archivos como XML¹⁸ y su interacción con el lenguaje HTML¹⁹, RSS²⁰, entre otros. Las bases de datos tienden a interactuar entre sí por lo que emerge la necesidad de definir criterios comunes en la indexación de datos, además del requerimiento de lenguajes y formatos que sean de uso masivo y permitan la colaboración de muchos investigadores. Ejemplo de esta situación es el uso extendido de los archivos XML en compatibilidad con otro tipo de archivos y lenguajes en la vulcanología desarrollada en Europa mediante el uso del sistema **VolcanoGasML**²¹. Este permite la captura y almacenamiento de datos geológicos centrándose en lo operativo y masivo de los archivos XML. En este sentido VolcanoGasML no es una gran base de datos en XML ya que es compatible con otros archivos y exportable a otros sistemas, pero su uso mayoritario con este tipo de archivos fue justamente por la enorme extensión y usabilidad que posee el XML por los investigadores ligados a la geología y esencialmente la vulcanología.

Especial atención merece el actual desarrollo técnico e informático de la Internet y las formas de capturar, almacenar, procesar y diseminar datos e información. Dentro de las experiencias que se describirán más adelante se puede encontrar una enfocada a la captura de datos con un alto componente tecnológico como el cada vez más especializado **Sistema GUI**²², un interfaz de reconocimiento gráfico que reconocen escrituras y las procesan digitalmente. También el gran avance y relevancia que poseen los motores de búsqueda en plataformas web los cuales van desde el reconocimiento textual de la búsqueda, agrupación y selección de ítems en base al contenido solicitado, hasta la identificación mediante técnicas semánticas latentes de la indexación de direcciones.

¹⁸ XML: Extensible Markup Language, es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C). Es una simplificación y adaptación del SGML y permite definir la gramática de lenguajes específicos (de la misma manera que HTML es a su vez un lenguaje definido por SGML). Por lo tanto XML no es realmente un lenguaje en particular, sino una manera de definir lenguajes para diferentes necesidades. Algunos de estos lenguajes que usan XML para su definición son XHTML, SVG, MathML (wikipedia).

¹⁹ HTML: HyperText Markup Language, lenguaje de marcado predominante para la construcción de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes (wikipedia).

²⁰ RSS: es una familia de formatos de fuentes web codificados en XML. Se utiliza para suministrar a suscriptores de información actualizada frecuentemente. El formato permite distribuir contenido sin necesidad de un navegador, utilizando un software diseñado para leer estos contenidos RSS (agregador) (wikipedia).

²¹ Reiter, E. (2007). VolcanoGasML: a format to exchange geochemical Volcanic gases data. *Data Science Journal*. Vol. 6 , pp.7-18.

²² Chen – Yuan Liu, Huang – Cuang Lin, and Kojima, M. (2007). A Character Recognition on object oriented design for Tibetan Buddhist Texts. *Data Science Journal*, Vol 6, PP. 137-144.

III.2. ESPECIFICIDADES DISCIPLINARIAS EN GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

Oceanografía

En el caso de la oceanografía existen al menos cinco experiencias significativas en gestión de datos e información científica, las cuales corresponden al *Palmer Station Long-Term Ecological Research program (Palmer LTER)*, *United - States Joint Global Ocean Flux Study (US JGOFS)*, desarrollados por el *National Center for Health statistics (USA)*, algunos programas destacados para la custodia de datos oceanográficos, como el Comisión Oceanográfica Intergubernamental (COI), el SCOR junto a su capítulo español: *SCOR-Es, Scientific committee on oceanic research*, y por último el International Oceanographic Data and Information Exchange (IODE), dependiente del Intergovernmental Oceanographic Commission dependiente de UNESCO.

En el primero de estos (Palmer LTER) se centra principalmente en tareas referidas tanto a la gestión de datos, como a su descripción y acceso. Es así como ya en el año 2001, este proyecto culminó con la aprobación oficial de la Lengua de Metadatos Ecológicos (EML), mientras que en la actualidad el sistema se sustenta en una base de datos relacional con una API orientada a objetos que soportan la capa de datos basada en una plataforma Web de consulta.

El segundo proyecto de integración de datos oceanográficos corresponde al US JGOFS, el que se creó como un programa para comprender el ciclo global del carbono y el conjunto de los elementos, analizando desde una visión interdisciplinaria la forma en que los océanos intercambian estos elementos con la atmósfera, el suelo marino y las fronteras del continente. Este es un sistema de datos personalizados que puede generar conjuntos de datos que coinciden con los intereses del investigador mediante la combinación de múltiples fuentes de datos, bajo una plataforma que contiene bases de datos relacionales.

1.1.- Instancias internacionales en la gestión de datos oceanográficos

Comisión Oceanográfica Intergubernamental (COI)

Comisión creada en 1960, perteneciente a la UNESCO, promueve la cooperación internacional y coordina programas de investigación marina, servicios, sistemas de observación, la mitigación de desastres y el desarrollo de la capacidad para aprender más y gestionar mejor la naturaleza y los recursos de los océanos y zonas costeras. De esta forma, COI contribuye a la mejora de prácticas y la toma de decisiones de los estados que lo conforman, fomentando el desarrollo sostenible²³.

²³ Fuente: COI web (http://www.ioc-unesco.org/index.php?option=com_frontpage&Itemid=1). Accesado el 13 de enero de 2010.

Dentro de las principales acciones del COI se encuentran:

a) Prevención y reducción de los impactos de los peligros naturales

-Promover la gestión integrada y sostenible de vigilancia y sistemas de alerta de los riesgos naturales costeros y oceánicos, en coordinación con otras organizaciones intergubernamentales.

-Capacitar a la comunidad en riesgo sobre la prevención de riesgos naturales, impacto, medidas de preparación y mitigación.

b) Mitigación de impactos y adaptación al cambio climático y la variabilidad

-Aumentar la comprensión de la función del océano en la variabilidad climática y el cambio climático mediante capacitaciones.

-Contribuir a la mejor predicción del clima mediante la observación de los océanos a escala regional y mundial.

-Aumentar la comprensión de los impactos del cambio climático y la variabilidad en los ecosistemas marinos y sus recursos vivos.

c) Salvaguardar la salud de los ecosistemas oceánicos

-Evaluación constante del estado del medio marino, incluidos los aspectos socioeconómicos.

-Investigación y la vigilancia que contribuya a prevenir el hábitat marino y el mantenimiento de su biodiversidad.

-Contribuir al mantenimiento de los ecosistemas oceánicos saludables centrándose en las necesidades regionales.

d) Desarrollar procedimientos en políticas de gestión que conduzcan a la sostenibilidad de los recursos costeros y el medio ambiente oceánico

-Mejorar la cooperación regional y la participación de los Estados miembros mediante la creación y fortalecimiento de capacidades y transferencia tecnológica.

-Facilitar la ciencia relacionada con la gestión de los recursos oceánicos y costeros.

-Desarrollo y aplicación de instrumentos de apoyo que mejoren la gestión integrada de océanos y Zonas Costeras.

Scientific Committee on Oceanic Research (SCOR)

SCOR es la principal organización no gubernamental internacional para la promoción y coordinación de las actividades internacionales de oceanografía. SCOR no tiene los recursos para financiar la investigación directa, por lo que las actividades científicas de SCOR se centran en promover la cooperación internacional en la planificación y la realización de investigaciones oceanográficas, y la solución de los problemas metodológicos y conceptuales que dificultan la investigación. Científicos de las treinta y seis naciones participan en SCOR, se establecen en grupos de trabajo y comités de dirección científica de grandes proyectos de investigación oceanográfica.

El principal objetivo del SCOR es avanzar en el bienestar de la población humana, reducción de las pérdidas de vida y propiedades debidas a riesgos naturales, y permitan una mejor conservación de la biodiversidad y del ambiente global, mediante éste conocimiento es necesaria una coordinación global del acceso a datos e información científica de alta calidad. Este enfoque, reconocido en numerosas propuestas internacionales, requiere colaboración a través de fronteras estatales y la existencia de una gestión adecuada de los datos oceanográficos en cada país, es en éstos lineamientos es donde se centra el SCOR.

Destaca la experiencia SCOR-Es, representación española del SCOR, el que tiene como objetivo mejorar la resolución de problemas conceptuales y metodológicos que entorpezcan la investigación oceanográfica²⁴. El SCOR realiza varias actividades para la mejora de la gestión de datos oceanográficos, entre ellas se encuentran:

1. Patrocinio a grupos de trabajo para tratar temas oceanográficos de interés internacional.
2. Patrocina la organización de programas de investigación oceanográfica de gran escala los que abordan temáticas como el papel del océano en el cambio climático, la creciente importancia de las proliferaciones algales nocivas.²⁵
3. Trabaja asociadamente con otras instituciones para la observación del océano y la gestión de datos para programas internacionales de investigación.
4. Fomentar el conocimiento oceanográfico en países en vías de desarrollo mediante escuelas de postgrado y el patrocinio de becas a investigadores.

En España el SCOR-Es patrocinó a un grupo de trabajo para generar el plan de mejoramiento en la gestión de datos oceanográficos acorde con los estándares internacionales. Para esto el equipo desarrolló un proyecto que contempló las siguientes etapas: 1) Análisis de la normativa internacional, 2) Identificación

²⁴ De acuerdo a su página web: los objetivos de SCOR son la promoción de la cooperación internacional en la planificación y realización de investigación oceanográfica y en la resolución de problemas metodológicos y conceptuales que dificulten esta investigación. SCOR es, además, el órgano consultivo de la Comisión Oceanográfica Intergubernamental (COI) de UNESCO. La COI proporciona a los miembros de las Naciones Unidas un foro de cooperación global para estudios y problemas relacionados con los océanos. Fuente: <http://www.scor-int.org>. Accesado el 13 de Enero de 2010.

²⁵ En su página web se señalan los siguientes programas: World Ocean Circulation Experiment (WOCE), Joint Global Carbon Flux Study (JGOFS), Global Ocean Ecosystem Dynamics (GLOBEC), Global Ecology and Oceanography of Harmful Algal Blooms (GEOHAB), Surface Ocean Atmosphere Study (SOLAS) y Marine Biogeochemistry and Ecosystem Research (IMBER). Fuente: <http://www.scor-int.org>. Accesado el 13 de Enero de 2010.

de brechas, 3) elaboración de recomendaciones, 4) Elaboración de Informe final. A continuación se detallarán cada una de estas etapas:

Etapa 1. Análisis normativa internacional: en esta etapa se revisaron las principales iniciativas sobre archivo y gestión de bases de datos/metadatos oceanográficos disponibles a nivel español e internacional. Las instancias identificadas pertenecientes a la Unión Europea fueron:

Global Monitoring for Environment and Security (GMES): generada en la Unión Europea para apoyar los objetivos europeos de desarrollo sostenible y proporcionar datos de alta calidad, información y servicios. Además proporciona acceso a información geoespacial para gestores y autoridades.²⁶

Blue Book: libro publicado en 2007, se relaciona con la iniciativa anterior ya que se basa en sus lineamientos y además ayuda al GMES en los procedimientos para la recolección de datos propone el desarrollo de un Red Europea de Observación e Información del Mar (European Marine Observation and Data Network, EMODNET).²⁷

Convención de Aarhus: en la que sus accionar se orienta a la necesidad de compartir información por los organismos públicos de los países signatarios.

La directiva INSPIRE: creada en 2007, permite una normativa para la generación de infraestructura y equipamiento de datos que proporcione servicios integrados de información a las autoridades públicas de los países miembros.²⁸

Etapa 2. Identificación de brechas: en esta etapa se realizaron talleres de trabajo para conocer las brechas existentes entre las iniciativas disponibles y las necesidades de la comunidad oceanográfica española en materia de gestión de datos oceanográficos.

En la matriz siguiente se expondrán los resultados en base a escenarios actuales y deseados que el proyecto arrojó respecto a los principales aspectos que España debe trabajar para mejorar su gestión en datos oceanográficos. Este análisis incluye los siguientes ejes:

- Recopilación, control de calidad y almacenamiento de datos de manera que queden disponibles para el futuro.
- Distribución de datos a científicos, gestores, industria y público.
- Establecimiento de protocolos para la adquisición y tratamiento de datos.
- Adopción de una Política de datos.
- Desarrollo de productos de datos adecuados.

²⁶ Más información de los procedimientos de GMES: http://www.bmvbs.de/Anlage/original_1006811/The-Way-to-European-Earth-Observation-Services-GMES-The-Munich-Roadmap-accessible.pdf . Accesado el 13 de Enero de 2010.

²⁷ Más detalles sobre el plan de acción del Blue Book: <http://ec.europa.eu/maritimeaffairs/> . Accesado el 13 de Enero de 2010.

²⁸ Más información sobre la directiva INSPIRE: www.ec-gis.org/inspire . Accesado el 13 de Enero de 2010.

| RECOPIACIÓN, CONTROL DE CALIDAD Y ALMACENAMIENTO DE DATOS DE MANERA QUE QUEDEN DISPONIBLES PARA EL FUTURO. | | |
|---|---|---|
| ÁMBITO | ESCENARIO ACTUAL | ESCENARIO DESEADO |
| - Facilitar el acceso en tiempo presente a los datos de las estaciones meteorológicas y oceanográficas operativas en aguas españolas. | - Algunos portales mantenidos por una o varias instituciones, con datos en tiempo presente. - Participación en portales europeos en desarrollo. | - Continuidad de iniciativas existentes. Existe el compromiso político de dar continuidad a ESEOO (Establecimiento de un Sistema Español de Oceanografía Operacional) a través de una Oficina Española de Oceanografía Operacional, que debería estar estrechamente coordinada con el Centro Coordinador. |
| - Facilitar el acceso a datos y metadatos archivados. | - Diversos centros de datos con segmentos de información y con niveles heterogéneos en cuanto a organización y prestaciones. - Conexiones a sistemas de datos internacionales a través de algunos proyectos y centros de datos. | - Coordinación e integración de las bases de datos existentes, de manera que los usuarios puedan localizar la información necesaria a través del Centro Coordinador. |
| ESTABLECIMIENTO DE PROTOCOLOS PARA LA ADQUISICIÓN Y TRATAMIENTO DE DATOS | | |
| ÁMBITO | ESCENARIO ACTUAL | ESCENARIO DESEADO |
| Sin ámbitos específicos | - Protocolo común para series temporales de medidas puntuales (ESEOO) - Diversos protocolos estandarizados para otros tipos de datos oceanográficos - Existen guías de buenas prácticas para los distintos tipos de datos, elaboradas en el marco de grupos de trabajo internacionales. | - Seleccionar los protocolos adecuados y diseminar la información pertinente a todos los investigadores interesados. - Apoyo a las iniciativas existentes como SEADATANET. (Este proyecto tiene como objetivo desarrollar una infraestructura paneuropea para la gestión de todos estos diversos conjuntos de datos) |
| DESARROLLO DE PRODUCTOS DE DATOS ADECUADOS | | |
| ÁMBITO | ESCENARIO ACTUAL | ESCENARIO DESEADO |
| Sin ámbitos específicos | - Las instituciones existentes ofrecen datos y productos de datos de diversos niveles (regional, nacional, internacional). - Esfuerzo conjunto de diversas instituciones en relación con las series temporales puntuales (ESEOO). Entre otros productos, este proyecto ha realizado un atlas institucional de información marina homogéneamente analizada, que está disponible en su página web. | - Nuevos productos de datos adecuados a la evolución de la demanda. |
| ADOPCIÓN DE UNA POLÍTICA DE DATOS | | |
| ÁMBITO | ESCENARIO ACTUAL | ESCENARIO DESEADO |
| Sin ámbitos específicos | -Cada Organismo tiene una política de datos o reglamentación particular. -Política de datos Antárticos a nivel estatal. -Compromisos con políticas internacionales. | Elaborar una política nacional de datos de obligado cumplimiento para todos los datos y productos obtenidos y elaborados con financiación pública. |

Etapa 3. Recomendaciones: en esta etapa se elaboraron una serie de recomendaciones para generar políticas de gestión de datos oceanográficos que atiendan a las brechas detectadas. Las recomendaciones fueron las siguientes:

- 1. Generar un centro coordinador:** cabeza de la estructura planteada para liderar la gestión de datos oceanográficos. Este centro tiene como sus funciones orientadoras: a) ofrecer servicios propios de gestión de datos, b) actuar como integrador y foco de referencia, a nivel estatal e internacional, de la información administrada por los Centros Acreditados. Este centro debe poseer un director, un vicedirector científico y uno técnico además de un comité de dirección.
- 2. Generar centros acreditados:** centros que ofrecen prestaciones enfocadas a un determinado tipo de datos, una institución o grupo de instituciones, o una región geográfica. Estos deben cumplir ciertos requerimientos de calidad para poder operar.
- 3. Financiamiento:** se propone un financiamiento a toda esta estructura entre el Estado y los propios centros acreditados, además de incluir en las normativas de los fondos la obligatoriedad de suministrar los datos obtenidos mediante esta vía. Las formas que se proponen para generar esto último son: a) financiar los centros de datos para que puedan realizar siempre gestión de datos o b) incluir los gastos generados por la gestión de datos en el presupuesto del proyecto, de modo que éste pudiese costear la contribución del centro de datos adecuado.
- 4. Priorización del tipo de datos:** necesidad de establecer prioridades respecto al tipo de datos que se gestionarán. El sugerido por el estudio en una primera etapa fueron los datos oceanográficos básicos: temperatura, salinidad, concentración de clorofila y de nutrientes y series temporales de datos obtenidos en continuo, como nivel del mar, corrientes y oleaje.
- 5. Políticas de datos:** necesidad de formalizar los procesos en la gestión de datos, necesidad de reconocer la importancia para el desarrollo de la ciencia el abordar esta temática desde el Estado.

Etapa 4. Informe final: Preparar y publicar un informe con las recomendaciones extraídas. El extracto y principales conclusiones de este informe son las que se han descrito en esta sección²⁹.

²⁹ El informe completo se encuentra en la web de SCOR-Es: <http://www.scor-es.org/>

Oceanographic Data and Information Exchange (OIDE)

Por último se destaca lo realizado por el IODE, el cual tiene dentro de sus líneas principales de acción la Gestión de Datos Marítimos, en donde el principal objetivo radica en ser una guía a través de algunas iniciativas relacionadas con la gestión de datos marítimos y disposición de los principales actores del área, centrándose, fundamentalmente, en los datos oceanográficos físicos y biológicos, no así en hidrográficos, químicos y/o geológicos. Además es una instancia de discusión de nuevas tendencias y desarrollos que determinan el futuro de este campo a nivel mundial.

Sus áreas de desarrollo corresponden a:

- **Meteorología y defensa de las cosas:** determinación de la influencia climática, mediante el análisis de patrones climáticos diarios y con indicadores en “tiempo real”.
- **Predicción de “El Niño”:** monitoreo constante de la temperatura de la costa con una gran cantidad de boyas, lo cual permite la obtención y posterior gestión de gran cantidad de datos oceanográficos que se componen de temperatura y salinidad de los últimos 8 años, en base a los datos de los últimos 100 años con los que se cuenta.
- **Gestión de la vida y recursos no vivos:** conocimiento acabado y manejo de datos profesionales asociados a los recursos pesqueros existentes para el Atlántico Nororiental.
- **Centros de Datos en Evolución:** generación de mecanismos para la “el descubrimiento de datos”, vinculado a un sistema de distribución de los mismos, haciendo uso masivo de internet y las tecnologías de base de datos existentes.
- **Toma de decisiones:** permitir que los marítimos de evolución deben ser impulsados por la tecnología, resolviendo las siguientes desafíos: constante aumento del volumen de datos; gran diversidad en los tipos de datos e integración de los mismos; y gran discrepancia en la escala de los datos e información de relevancia dentro del área.

ii.- Recursos Naturales, Medio Ambiente, Agricultura, Población y la salud básica

Esta área científica que agrupa un conjunto de disciplinas, a partir de las características de los datos que producen, muestra un caso destacado y significativo para la gestión de datos e información, el cual corresponde al programa de Intercambio de Datos Científicos (SDSP), coordinado por el Ministerio de Ciencia y tecnología de China, y que en la actualidad agrupa un conjunto de instituciones recolectoras y administradoras de datos, como lo son el Centro de Datos Científicos Meteorológicos, el Centro de Datos Científicos de Desarrollo Rural, el Centro de Datos Científicos Agrícolas, el Centro de Datos Científicos de Control de la Población, el centro de Datos Científicos de Medicina Básica, el Centro de Datos Científicos del Sistema Tierra y el Centro de Datos Científicos del Ambiente Espacial.

El inicio de este programa se remonta al año 2001, en donde da comienzo el periodo experimental del Programa de Intercambio de Datos Científicos (SDSP), cuyo objetivo principal era establecer una arquitectura de intercambio de datos que facilitara el uso de datos científicos. Esto, en primer lugar, a través del establecimiento de leyes, políticas y estándares y, en segundo lugar, por medio de la utilización

de las tecnologías de información y comunicación que sirviesen de soporte para la preservación e intercambio de datos científicos.

En la primera etapa, de experimentación, (2001 – 2005) se abordaron los procesos de diseño y planificación general del sistema tales como: análisis de políticas y marco legal, redacción de políticas y regulaciones relevantes, desarrollo de tecnologías y estándares, establecimiento de centros y redes de datos científicos, iniciación del proyecto piloto de transferencia de datos, identificación de mecanismos óptimos para consolidación e intercambio de datos ya existentes, selección de los 25 centros de datos participantes del proyecto piloto y otros centros candidatos para etapas más avanzadas, y acumulación de experiencias del período experimental.

En la segunda etapa, de implementación (2006 – 2010), se ha continuado con el establecimiento tanto de las tecnologías de intercambio, como de las políticas y leyes, mientras que se ha extendido la cobertura del programa a un mayor número de centros y redes, mejorando las tecnologías y estándares utilizados, con un fortalecimiento de la cooperación entre centros de datos en diferentes áreas de investigación, y mejorando la capacidad de desarrollo de productos de datos de alto nivel, asegurando la calidad de estos últimos.

A partir de esta experiencias, para el año 2020, China pretende aplicar el SDSP (Scientific Data Sharing Program) con el objetivo de establecer una red de gestión de datos científicos y un mecanismo para compartir datos con un sistema de servicio con una eficaz estructura y de amplia cobertura de la mayoría de las ciencias básicas y de bienestar público, además de las plataformas web asociadas al quehacer científico. Con ello se prevé la construcción de un Gestor de Datos y el intercambio de servicios con un sistema de tres niveles: 1) estructura de 40 centros de datos científicos o de redes (los cuales proveen su información al portal), 2) 300 bases de datos y 3) un Portal Maestro.

iii.- Epidemiología

En el caso de la epidemiología se encuentra la experiencia significativa del programa del *Research Data center* (RDC). El Research Data Center (RDC) es un repositorio de datos que se encuentra ubicado en la sede NCHS en Hyattsville, MD. Dicho repositorio funciona como una "caja fuerte" de datos a modo de seguro, en donde se protegen los datos y existe un acceso restringido de los analistas invitados que responden a determinados criterios: se permite el ingreso, bajo una estricta supervisión, para generación de estadísticas (restringidas) en ficheros de microdatos. Para obtener acceso al RDC, los investigadores deben presentar una propuesta para su revisión y aprobación.

Tal como está concebido, el NCHS-RDC tiene seis áreas de trabajo para usuarios. Además, hay espacio para la oficina RDC personal y NCHS Analista Invitado. Las áreas de trabajo RDC son "únicas" y no tienen ningún vínculo con la red NCHS, la CDC-NCHS central, o la Internet. No hay suficiente espacio de almacenamiento en el área de trabajo y en el servidor para los datos confidenciales. PC-SAS, SUDAAN y STATA ya están instalados en los puestos de trabajo y programación adicional / analítica, por lo tanto se pueden añadir distintos lenguajes de programación cuando sea necesario.

Las estaciones de trabajo para los medios extraíbles, tales como los puertos USB están configurados para ser inaccesibles a los usuarios. Las estaciones de trabajo se han configurado de tal manera que los usuarios reciben acceso de sólo lectura a los datos y archivos solicitados y pueden escribir sólo en la estación de trabajo local del disco duro. Estas restricciones garantizan que los usuarios no pueden eliminar información que no ha sido sometida a una revisión de confidencialidad.

iv.- Matemáticas y Estadística.

El desarrollo de capacidades de almacenamiento y procesamiento en red de datos de investigación ha venido transformándose en un aspecto relevante en Matemáticas y estadísticas, dada la creciente complejidad de cálculo (desarrollo de algoritmos) que aquí se enfrenta, la simulación y predicción. En específico, es relevante el análisis del programa llevado a cabo por *School of Computer Science and Engineering, Xidian University, Xian (China)*, y específicamente su *Dept of Computer Science, Xian Institute of Post-telecommunications, (China)*, titulado *correlación geométrica de datos*.

En éste caso, el desarrollo de las técnicas de regresión lineal (LR) y la regresión de soporte vectorial (RVS) que son ampliamente utilizadas en el análisis multivariante de datos, específicamente en tareas de proyección y predicción estadística, se ha propuesto, recientemente, la utilización de la correlación geométrica del aprendizaje (GcLearn) para mejorar la capacidad predictiva de LR y SVR, a través de la minería y el uso de correlaciones entre los datos de una variable (correlación interna). Así ha generado un desafío específico y una necesidad impostergable sobre el almacenamiento y acceso de datos para operaciones y análisis multivariante.

El rendimiento de predicción del método GcLearn y demuestra que LR y SVR utilizando GcLearn tendrán un mejor rendimiento para las tareas de predicción cuando se obtiene una buena correlación interna, demostrando con resultados experimentales que se reducen los errores de predicción entre el 4% - 46% en comparación con LR y ϵ -SVR tradicionales. Esto ayuda a validar la utilización de GcLearn para optimizar los procesos de proyección y predicción estadística.

Las políticas de almacenamiento y acceso a datos permite, en este caso, el desarrollo de algoritmos que permitan una mejor predicción y rendimiento que los métodos tradicionales.

v.- Computación y Ciencias de la Información.

En ésta disciplina se agrupan un conjunto diverso de experiencias en el marco del almacenamiento de datos, siendo las más destacadas en el campo internacional la desarrollada por *School of Law, University College Dublin (Irlanda)*, y, *Arthur Cox, (Dublin and London)* en cuanto a protección de Derechos de autor y resguardo de propiedad intelectual; *Computer Science Department, Bar-Ilan University, (Israel)*, con dos experiencias relevantes, como el desarrollo de una técnica de almacenamiento en comprensión de imágenes y desarrollo de software, y por otro, una técnica de eficiencia relativa en comprensión y procesamiento de datos a través de Lzw y Lzss; el caso del *Center for Spatial Information Science, The University of Tokyo*, con el desarrollo de Metadatos de Imagen Basado En Normas ISO / Tc 211; el caso del

Department of Computer Science, University of West Florida (USA) con la utilización de la minería de datos para estudiar los patrones de votación en Estados Unidos; y el caso del *Department of Information Technology, Thiagarajar collage of Engineering, Madurai (India)*, con una clasificación de cluster para la descripción de textos.

En el caso de *School of Law, University College Dublin (Irlanda)*, se ha constatado que en la Unión Europea la generación de bases de datos cuenta con una normativa (Directiva) orientada a la mediación de los intereses de los distintos actores involucrados, resguardando los derechos de autor y el uso no autorizado y/o reutilización de los contenidos de la(s) base(s) de datos. Sin embargo, en los últimos años esta normativa está recibiendo cuestionamientos relativos a las limitantes que impone sobre la industria.

Esta normativa Directiva es analizada y cuestionada en la actualidad a la luz de dos informes recientes del Reino Unido donde se propone un régimen más liberal de derechos de autor. Esto sugiere la formulación de un nuevo marco normativo en el acceso de datos e información, en donde la competencia desleal debe ser abordada por las nuevas leyes de competencia desleal para Irlanda y el Reino Unido y no a través de la revisión de la Directiva.

School of Law, University College Dublin (Irlanda) finaliza su análisis identificando una destacada práctica en la estandarización de normativas, por lo cual recomienda un sistema integrado de metadatos de imagen, que responda de mejor forma a necesidades específicas.

Las experiencias identificadas en *Computer Science Department, Bar-Ilan University, (Israel)*, muestran, como el desarrollo de una técnica de almacenamiento en comprensión de imágenes y desarrollo de software, y por otro, una técnica de eficiencia relativa en comprensión y procesamiento de datos a través de Lzw y Lzss.

En el primero de estos casos, se ha iniciado el desarrollo de esta técnica a partir de la constatación de que los formatos de imagen más conocidos utilizan técnicas de compresión para reducir el peso de las mismas. La importancia de mejores performances de compresión radica en la reducción de imágenes, lo que es un procedimiento importante para la transmisión de archivos a través de redes y archivos bibliotecarios. Así es como se desarrolla un método de compresión Burrows-Wheeler Transform (BWT) en imágenes, por sobre las técnicas utilizadas en JPEG. Las ventajas de BWT con respecto a JPEG en el procedimiento de compresión, muestran los distintos pros y contras de ambas técnicas. Por ejemplo, el formato JPEG tradicional utiliza el algoritmo Huffman o codificación aritmética, el que a pesar de su capacidad, tiene la desventaja fundamental de perder información en el proceso, lo que repercute en la calidad de la imagen final. Por otro lado, JPEG basado en Burrows-Wheeler reduce significativamente una menor cantidad de datos.

La conclusión obtenida en éste caso por *Computer Science Department, Bar-Ilan University, (Israel)*, es proponer la utilización del formato JPEG utilizando la técnica de compresión Burrows-Wheeler, ya que las pruebas entre ambas alternativas otorgan ventaja en compresión y calidad al algoritmo Burrows-Wheeler, aunque su tiempo de ejecución es mayor.

En el segundo de las experiencias llevadas a cabo por el *Computer Science Department, Bar-Ilan University*, (Israel), encontramos el desarrollo de una técnica de eficiencia relativa en comprensión y procesamiento de datos a través de Lzw y Lzss.

Luego de un cuidadoso análisis del uso y comparación de eficiencia de métodos de comprensión de datos con archivos LZW y LZSS, el equipo de trabajo concluyó que generalmente LZSS puede comprimir datos de una mejor manera que LZW. Sin embargo, los archivos LZSS son más costosos, y dependiendo de la longitud de la base de datos, LZW puede ser la mejor opción de compresión, ya que en longitudes de tamaño mediano obtiene resultados de buena manera, por lo tanto, la regularidad se expresa en el archivo LZW como la mejor opción.

En el caso del *Center for Spatial Information Science, The University of Tokyo* se analiza el *Desarrollo de Metadatos de Imagen Basado En Normas ISO / Tc 211*, a partir de la heterogeneidad de normas ISO existentes para metadatos.

Los metadatos de imagen son datos anexos que ayudan a la descripción del contenido, la calidad, condición y características de un conjunto de datos de imagen (datos gráficos). La importancia de éstos se focaliza particularmente en la integración de los datos de imagen, en donde se debe reconstruir de modo preciso las condiciones geométricas en las que se realizó la observación.

La problemática central en los estándares para la creación y difusión de metadatos de imagen, radica en una variedad de normas ISO existentes que sólo cubren las necesidades de ciertas temáticas disciplinares específicas. Es así como existen tres normas, de las cuales la primera (ISO 19115) exige la magnitud temporal y espacial, la segunda (ISO 19115, parte 2) cerciora la calidad de los datos, la representación espacial, el contenido y la adquisición de la información, mientras que la tercera (ISO 19130) precisa la relación entre información del sensor y la información de geolocalización.

Como conclusión y solución a la problemática, el *Computer Science Department, Bar-Ilan University*, (Israel), propone la elaboración de un modelo integrado de metadatos de imagen basado en las ISO existente y en necesidades específicas. Para lo anterior se hace necesario revisar los tipos de metadatos y los elementos de metadatos existentes, ejemplificados en modelos UML.

Así es como se logra desarrollar un caso concreto de modelo integrado, utilizado actualmente para la búsqueda e integración de datos de imágenes satelitales, que corresponde a la experiencia CEOP, un sistema global de observación de ciclos de agua y energía para necesidades científicas y sociales que requiere la integración de imágenes de diversos satélites.

La experiencia desarrollada por *Department of Computer Science, University of West Florida (USA)* requirió de un conjunto de técnicas para el análisis de datos con el fin de estudiar los patrones de votación de representantes en las cámaras de Estados Unidos. Para ello se procesaron los datos en bruto que se encuentran en la página Web <http://clerk.house.gov>. Para ello se analizaron factores tales como la importancia cualitativa, reglas de asociación, análisis de árboles de decisión e interpretación de resultados interesantes. Los datos de entrada para los análisis pueden encontrarse en formato de datos XML, planillas de cálculo, tablas anidadas, etc.

Por último, analizamos la experiencia del *Department of Information Technology, Thiagarajar collage of Engineering, Madurai* (India), en donde a partir de una identificación de los parámetros en los cuales se desarrollan los procesos de búsqueda de información mediante el uso de plataformas Web, se presentan distintos sistemas para la búsqueda descriptivas de documentos en Web, donde los principales son:

1. *VSM, modelo del espacio de vector*: es un método de recuperación de la información que utilice operaciones algebraicas – lineales para comparar datos textuales. Utiliza Heurísticos para la identificación de documentos en el proceso de búsqueda.
2. *SVD, descomposición del valor singular*: utiliza frases de los documentos que pueden aproximarse desde una derivación del significado verbal (es el más complejo).
3. *LSI, técnica semántica latente de la indexación de direcciones*: utiliza los datos de entrada usando los conceptos, en términos literales, encontrados en los documentos.

Se enfatiza en la práctica del *Department of Information Technology, Thiagarajar collage of Engineering, Madurai* (India) el auge que cobran en la actualidad las búsquedas especializadas en la Web. Así su investigación en esta área está orientada a desarrollar algoritmos más rápidos y más exactos para el agrupamiento de documentos.

vi.- Ingeniería Mecatrónica; Tecnologías de la Información y Comunicación

En ésta área se la experiencia desarrollada en conjunto entre *Department of Mechatronic Engineering, Huafan University* (Taipei, Taiwán), y *Department of Information Technology and Communication, TNIT, Huafan University*, (Taipei, Taiwán), en donde se ha desarrollado un modelo cuantitativo de análisis de patentes tecnológicas, que permite describir la trayectoria de una tecnología, aplicado al área de robótica humanoide.

Para el análisis se utilizaron documentos de Japón, el país que mayor desarrollo presenta en robótica humanoide, específicamente en un área clave del desarrollo de esta tecnología, que es el control de medios y el control de efectos del movimiento caminar de los androides. Lo cual sirve tanto para el seguimiento de una trayectoria tecnológica, como también para conocer los principales desarrollos futuros. Se consideraron todas las patentes, un total de 257, otorgadas a distintas entidades desde 1987 a 2005.

El modelo consta de una matriz que confronta dos clases de factores en cuestión: el control de medios y control de efectos del movimiento caminar. Los patrones encontrados permiten determinar las tecnologías enfatizadas, que han pasado a ser consideradas soluciones, tecnologías maduras o descartadas. De manera complementaria se expone un análisis descriptivo que da a conocer las empresas que participan en la creación de la tecnología, en este caso las patentes de robótica androide, además de los años en que irrumpe el desarrollo de la misma y las tendencias desde dicho momento

El modelo permite ordenar y sintetizar una importante cantidad de información de patentes. Lo anterior puede facilitar el análisis y planificación en modelos de inteligencia competitiva y vigilancia tecnológica.

vii.- Ciencias Físicas de Laboratorio

Dentro de esta área se encuentran dos experiencias que están vinculadas al almacenamiento, procesamiento y diseminación de datos e información científica. La primera corresponde a la desarrollada por Fermilab en el campo de la Física electro débil, eventos quark top, eventos quark b, eventos de prueba en cromodinámica cuántica, y partículas exóticas, para el desarrollo de un detector de colisiones en el almacenamiento, procesamiento y diseminación de datos.

Esta experiencia se refiere al proceso del laboratorio *Fermi National Accelerator Laboratory (Fermilab, USA)* el cual desde el año 1967 ha contribuido al modelo estándar de partículas y fuerzas fundamentales. Lo esencial de esta experiencia es cómo se debió enfrentar la administración de la gran cantidad de datos, la cual en un principio se almacenaba en cintas de 8 mm. y se trasladaban a otras instituciones en maletines. Posteriormente se utilizó el código ASCII en un directorio accesible central para luego enviar notificaciones a las instituciones colaboradoras

El segundo caso que destacamos es la elaborada por *Los Alamos National Laboratory* (institución administradora de datos), en el campo de la fenomenología de la Física de partículas, para el procesamiento y diseminación de información científica.

Esta experiencia se destaca en la utilización en la década de los 90' de los archivos E-Print ya que en la década de los 70' los investigadores del área de Ciencias Físicas de Laboratorio priorizaban el rápido acceso a información científica por sobre el filtro que representa el proceso convencional de evaluación de pares y referencias. Los archivos E-Print son borradores (artículos pre-impresos) los que se almacenan y distribuyen en *Los Alamos Nacional Laboratory*. Este proceso flexibiliza la finalización de artículos pasando por una estela de correcciones hasta llegar a su estado final.

viii.- Genómica.

Se debe destacar la experiencia en el campo específico de las Secuencias de ADN, Genomas completos de Organismos y Productos Genómicos Macromoleculares desarrollada en conjunto por el *Centro Nacional para Información Biotecnológica, la Base de Datos del Genoma (USA), la Base de Datos de ADN (Japón), Organización Europea de Biología Molecular* y el *Instituto Bio-informático Europeo*. (Instituciones administradoras de Datos), para el desarrollo de secuencias genómicas.

Así como este caso de almacenamiento y diseminación de datos, se refiere al intercambio de información en ciencias químicas mediante el almacenamiento a través de una serie de bases de datos alojadas en distintas instituciones financiadas por fondos de investigación gubernamentales. Estas organizaciones se preocupan de recolectar los datos generados tanto por instituciones académicas, gubernamentales como

laboratorios industriales, para luego traspasarla a un formato estándar y finalmente distribuirla diariamente de tal forma de mantener una sola base de datos internacional común disponible en Internet.

ix.- Ciencias Astronómicas

Se destaca en las ciencias astronómicas la experiencia del Archivo del Telescopio Espacial Hubble, desarrollado por *National Aeronautic and Space Administration* (institución ejecutora principal) y la Agencia Espacial Europea (institución ejecutora secundaria), en materia de procesamiento y diseminación de datos.

El telescopio espacial Hubble (TEH) fue desarrollado por la NASA con una participación del 15% de la Agencia Espacial Europea (AEE), institución que también participa en su operación. Fue puesto en órbita el 24 de abril de 1990 inaugurando el programa de Grandes Observatorios. El TEH se encuentra disponible para el uso de la Comunidad Internacional de Astronomía, todos los datos científicos obtenidos son archivados y mantenidos para el uso prioritario del astrónomo que propuso la observación durante 1 año; luego de este periodo se dejan los datos a disposición de todos los astrónomos a través de un sistema de revisión de pares normal y con acceso al público en general vía Internet.

x.- Geología, Vulcanología y Sismología

Aquí se identifica la experiencia de captura y procesamiento de datos numéricos, realizada por J. H. Obenholzner y *Museum of Natural History of Vienna* (Austria), aportando información técnica, documentación y bibliografía, con el fin de generar un formato para el intercambio de datos geoquímicos de los gases volcánicos con el sistema VolcanoGasML, concentrándose en los procesos de levantamiento y almacenamiento de datos geológicos, y centrándose en lo operativo y masivo del uso de sistemas en base a archivos XML, se ha desarrollado un sistema y estándares, el ya mencionado “*VolcanoGasML*”, ofreciendo un sistema compuesto por dos capas de información:

Capa 1: información del volcán en estudio.

Capa 2: contiene datos químicos.

El sistema VolcanoGasML, posee diversas ventajas, donde se pueden mencionar como principales: a) Permite almacenaje y búsquedas históricas de datos y b) El tipo de archivo XML, XMLT, permite exportarlo fácilmente a HTML para publicación en web, a un nuevo archivo de XML, ASCII para importar los datos en otros software, SVG para crear gráficos, RSS para noticias web o el pdf.

Por su parte en el área de Sismología se encuentra el USGS Earthquake Hazards Program en Estados Unidos, el cual corresponde a un sitio web especializado en la entrega de información referente a la ocurrencia de sismos en Estados Unidos, contando con un Programa de Riesgo de Terremoto, informando respecto a la ocurrencia de movimientos telúricos con la finalidad de supervisión y notificación, por medio

del Advanced National Seismic System (ANSS). Los servicios asociados al anejo de datos e información referente a sismos son los siguientes:

- **Terremotos recientes:** información online de eventos a nivel nacional y mundial en sólo minutos.
- **ShakeMap:** información entregada mediante medición instrumental de la intensidad del movimiento en una región determinada.
- **ShakeCast:** aplicación automatizada de distribución para usuarios sobre movimientos sísmicos en un lugar determinado y las probabilidades de nueva ocurrencia.
- **Mapas de peligros:** indicación de zonas del país y su probabilidad de ocurrencia de movimientos telúricos.
- **Notificaciones de terremoto:** notificaciones automatizadas de los terremotos a través del correo electrónico, buscapersonas o teléfonos móviles.
- **Catalogo de terremotos y datos:** catalogo online que permite al usuario descargar información y datos técnicos.
- **Formas de onda en tiempo real:** visualización de formas de onda en tiempo real en base a la medición de 60 estaciones que están disponibles en línea las 24 horas del día.
- **Información de terremotos por Estado:** información acerca de los peligros de terremotos, sismicidad histórica, fallos, además de enlaces adicionales disponibles por cada Estado.
- **Películas de agitación de estructuras:** películas en formato Quicktime creadas a partir de las grabaciones de estructuras completamente instrumentadas durante los terremotos.
- **Mapa de pronóstico Afetshock:** mapa generado en función del tiempo y la probabilidad de movimientos fuertes en cualquier lugar de California en las próximas 24 horas.

Además del Advanced National Seismic System (ANSS), el USGS Earthquake Hazards Program se cuenta y se vincula con: The Global Seismographic Network, el cual es una red permanente de estado de la técnica sismológica y geofísica con sensores conectados por una red de telecomunicaciones, que actúa como multi-uso científico y de las instalaciones de un recurso de control, investigación y educación para la sociedad; USGS National Strong-Motion Project; y The Seismic Network Operations.

IV. CAPACIDADES CHILENAS DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA Y TECNOLOGICA

El presente capítulo, expone los principales antecedentes respecto de las capacidades chilenas en gestión de datos de investigación e información científica. En específico:

- Encuesta a instituciones e investigadores: se exponen aquí de manera sucinta, los antecedentes aportados por la encuesta realizada a instituciones chilenas que realizan investigación con fondos públicos y a investigadores nacionales.
- La experiencia chilena en oceanografía y sismología: dada la actualidad de este campo disciplinario, se exponen antecedentes sobre capacidades chilenas de investigación científica y tecnológica a este respecto.
- Percepciones de los agentes: se exponen los resultados del trabajo de campo realizado con instituciones e investigadores que realizan gestión de datos, recogándose sus percepciones y propuestas.

IV.1. CAPACIDADES DE INSTITUCIONES E INVESTIGADORES CHILENOS EN GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

A continuación se presenta una síntesis de los resultados de la encuesta nacional a instituciones e investigadores. Se procede sintetizando para cada uno de los 6 ejes de análisis, cuales son:

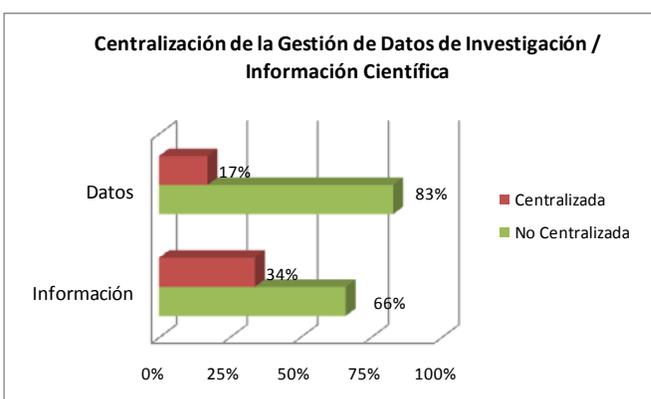
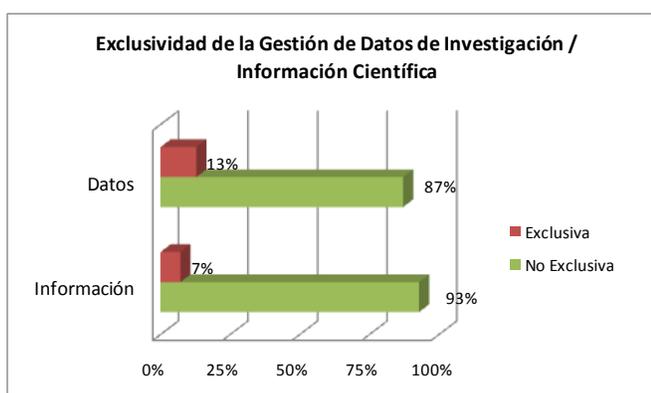
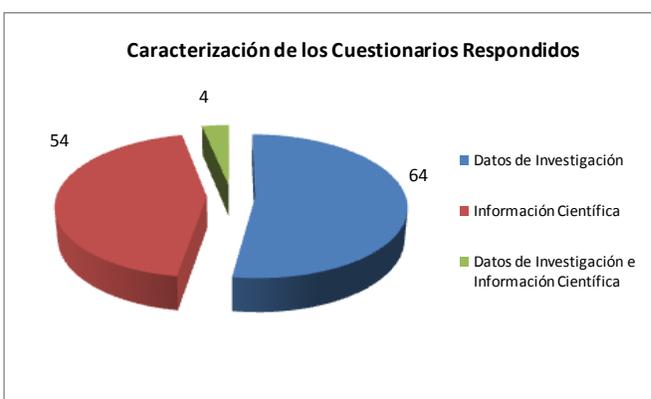
- **Institucionalidad**
- **Capital Humano**
- **Infraestructura Tecnológica**
- **Patrimonio**
- **Redes y Vinculaciones**
- **Políticas de Gestión**

Resultados y conclusiones sobre institucionalidad

Este corresponde al primer nivel de análisis, en donde se analiza la institucionalidad de las unidades o departamentos con labores de gestión de datos de investigación y/o información científica, además del nivel de exclusividad que posee estas labores en relación con otras tareas y funciones, considerando los distintos niveles propuestos: estadísticas generales, indicadores por área científico-tecnológica y finalmente indicadores por cadena de valor. A continuación se presentan las principales conclusiones extraídas del análisis efectuado, enfocado principalmente al foco de la gestión, centralidad/exclusividad y percepción respecto del cumplimiento de metas.

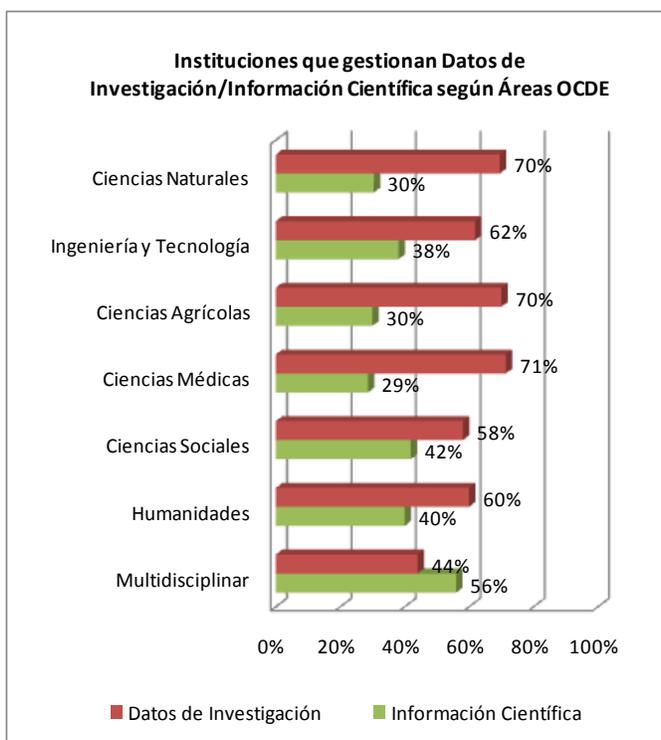
El análisis reflejó que en Chile existe una concentración de instituciones que gestionan mayormente datos de investigación que información científica, las que corresponden, también en su mayoría, a Centros de Investigación. Universidades e Institutos.

Un porcentaje altísimo de instituciones chilenas que gestionan datos de investigación e información científica lo hacen de forma desorganizada, ya que no existen unidades abocadas y especializadas sólo en la gestión, traspasando esas responsabilidades a los propios investigadores. Cada uno de estos resuelve sus problemas en la gestión según sus propios conocimientos y necesidades, lo que evidencia una ausencia de protocolos, estándares y la imposibilidad de que las propia instituciones posean planes de mejoramiento, detección de debilidades, requerimientos y cumplimiento de metas, pensando en la gestión de datos de investigación e información científica, en síntesis, una ausencia de formalización en este tipo de labores en las instituciones chilenas lo que impide planes de mejoramiento al interior de las unidades y el tipo de avance de las mismas áreas científicas en Chile.

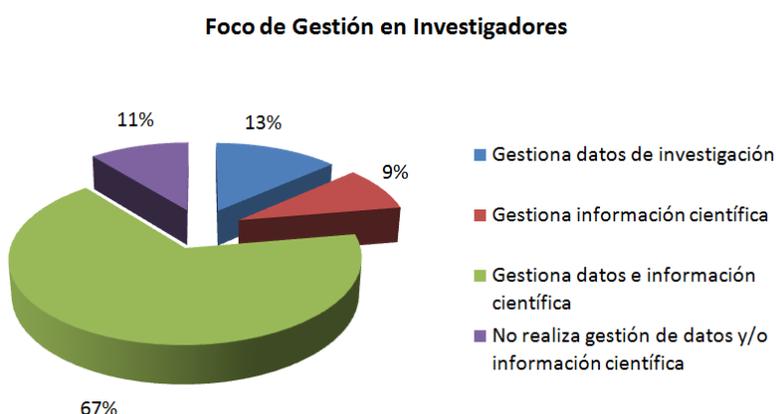


Para la gestión de datos de investigación, las áreas más trabajadas son 1) ciencias médicas, 2) ciencias naturales y ciencias agrícolas, y 4) ingeniería y tecnología. Para la gestión de información científica los resultados arrojan que las áreas más trabajadas son: 1) multidisciplinarias, 2) ciencias sociales y 3) humanidades.

Respecto a las etapas de la gestión se observó una ausencia de especialización por parte de las instituciones en alguna de ellas: planificación, captura/obtención, almacenamiento, procesamiento, validación y diseminación, ya que éstas declaran poseer el mismo nivel de especialidad en todas ellas. Esto podría generar la sensación de que existen buenas prácticas en la gestión de datos de investigación e información científica en las instituciones chilenas en el sentido de mantener un desarrollo armónico en cada una de las etapas de la gestión. Sin embargo, los resultados de este estudio dan cuenta que existen sólo desarrollo focalizados en algunas etapas donde, por ejemplo, la etapa de diseminación se confunde sólo con el mantener el patrimonio preferentemente en páginas web sin la promoción de éste a actores o instituciones que potencialmente podrían darle uso. Más grave es la situación de etapas de validación, tanto para datos de investigación como información científica, las que se encuentran ausentes en la mayoría de las instituciones.



Por último se evidencia una clara falta de especialización de capital humano pertinente a las etapas de la gestión, así como el apoyo de equipamientos, tecnologías y protocolos que puedan sostener el desarrollo de las etapas de la gestión de datos de investigación e información científica.



Respecto a los investigadores en Chile se observó una mayor proporción de quienes gestionan tanto datos de investigación como información científica. Esto refiere a que de forma independiente los investigadores deben asumir y enfrentar por ellos mismos las

complejidades que presenta la gestión de este tipo de patrimonio, ya que en Chile aún se carece de una institucionalización de estas prácticas.

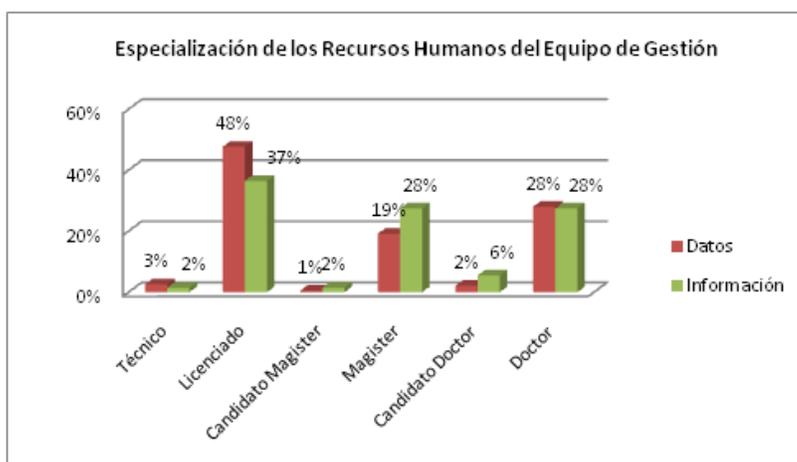
Es necesario señalar que el hecho de que en Chile se logre un proceso de formalización en la gestión de datos de investigación e información científica, se traduciría en la creación de unidades y profesionales dedicados específicamente a éste ámbito, lo cual facilitaría la labor a los científicos e investigadores.

Para el caso de las disciplinas científico-tecnológicas, si bien existen diferencias en el número de casos, las proporciones de distribución de tipos de gestión mantienen la tendencia indicada anteriormente, es decir, mucha mayor gestión de datos de investigación e información científica de forma simultánea, seguida de gestión exclusiva de datos de investigación, y finalmente gestión exclusiva de información científica.

Resultados y conclusiones sobre capital humano

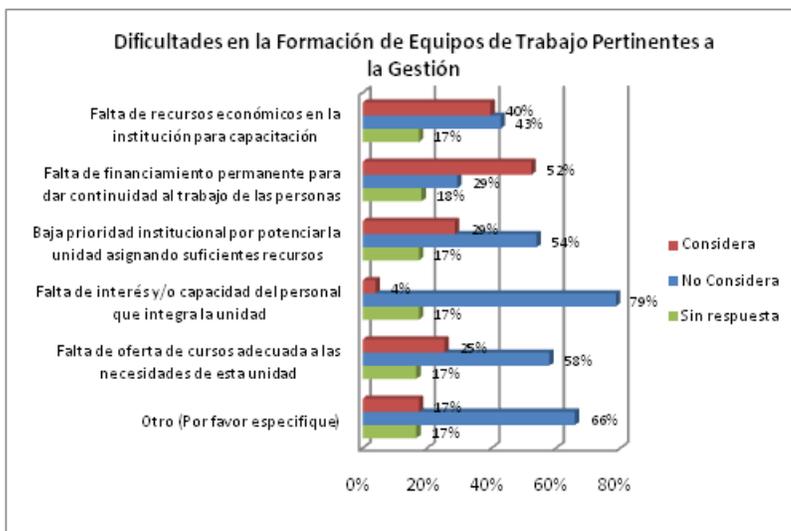
La gestión de datos de investigación e información científica de una disciplina o área científica no es una labor que se adquiere automáticamente al conocer dicha área o disciplina. Esto no quiere decir, por ejemplo, que un oceanógrafo sea incapaz de poder realizar gestión de datos oceanográficos, sino que para realizar dicha gestión son necesarios otros componentes que van más allá de la oceanografía en sí. Estos componentes corresponden a la documentación, el conocimiento de protocolos para la obtención de datos e información, manejo tecnologías para la gestión de bases de datos, procesos de diseminación, entre otros. Estas competencias formalmente se obtienen en programas de magister y doctorado específicos y la situación que ocurre en Chile es que los mismos profesionales, sin poseer necesariamente estas habilidades, son los encargados de realizar la gestión. Esto evidencia una necesidad de formar, atraer y retener capital humano avanzado para la gestión de datos de investigación e información científica en Chile, debido a la carencia de profesionales con un alto nivel de competencia en este campo.

De las razones aludidas por las propias instituciones para explicar lo bajos estándares en la gestión de datos e información, es la no disposición de recursos para continuar el trabajo de personas y unidades que aborden este tipo de gestión. En este sentido, se percibe una carencia en la propia normativa, en los fondos y en las prioridades emanadas del Estado en materia de ciencia y tecnología para crear y fortalecer iniciativas que vayan en beneficio de la gestión de datos de investigación e información científica en Chile.



La percepción que las propias instituciones poseen de lo que implica la gestión de datos de investigación e información científica refiere además al actual escenario que prácticamente no contempla el desarrollo de actividades ligadas al fortalecimiento de estas tareas. Esto es un aspecto más que demuestra lo poco institucionalizado que se encuentran estas labores en la actividad científica, ya que dificultan el fortalecimiento de equipos especializados para realizar de mejor manera esta labor y si bien no ponen en

duda la buena percepción que mayormente se hace de sus equipos de trabajo, invita a pensar de que existe un alto desconocimiento de estándares, metodologías procedimientos y competencias necesarias para la gestión de datos de investigación e información científica.



Concordante con estos resultados en capital humano, las habilidades dominadas por los gestores actuales de datos de investigación e información científica en Chile, distan de las fundamentales y los estándares internacionales. Existe un desconocimiento en protocolos, manejo de conceptos como la ley de propiedad intelectual, bajo manejo de gestores de contenidos, gestores de bases de datos entre otros aspectos. Esto al margen, como se señaló anteriormente, de la relativa buena percepción respecto a las habilidades de los equipos de trabajo.

Respecto a las habilidades por área científica tecnológica se puede señalar que no existe una tendencia general ni homogénea. Los resultados varían dependiendo del tipo de habilidades y su pertinencia y usabilidad en las distintas áreas, dependen del tipo de institución y por último dependen si el tipo de gestión es en datos de investigación y/o información científica. Una de las mayores consideraciones detectadas, según área científico-tecnológica y por tipo de institución, corresponde al manejo de tecnología aplicada a datos, tecnologías y aplicaciones web, derechos de autor y Open Source. Mientras que las consideraciones menos significativas se producen en las habilidades técnicas en el manejo de e-repositories y protocolos para la transmisión de contenidos en Internet (OAI-PMH).

Los eslabones donde se detectan mayores deficiencias y dificultades, en los actuales equipos que tienen el rol de gestionar datos de investigación y/o información científica en Chile, son los de obtención/captura, manifestadas mayormente por instituciones que gestionan datos de investigación y la etapa de validación, señalada mayormente por las que gestionan información científica. Si bien no existe un estado sobresaliente en ninguna de las etapas de la gestión por parte de las instituciones chilenas, las etapas de obtención/captura y validación son las mayormente identificadas como deficientes.

Respecto a las capacidades de los investigadores que llevan a cabo las labores de gestión de datos de investigación e información científica se observó que el manejo conceptual y teórico de elementos claves para estas labores es deficiente. Se analizó el grado de conocimiento de los investigadores chilenos con los conceptos de Open Access, Derechos de Autor y Ley de Transparencia y si bien existen diferenciaciones al interior de las áreas científicas, se puede señalar que existe un alto nivel de desconocimiento e ignorancia respecto a lo que estos conceptos implican.

En Open Access se visualiza una tendencia negativa a la familiaridad de los encuestados; mientras en lo concerniente a Derechos de Autor y Ley de Transparencia donde se reconoce mayoritariamente una familiaridad en términos medios.

Esta situación da cuenta de posibles focalizaciones o requerimientos de capacitación en el conocimiento, manejo y puesta en práctica de elementos asociados a los conceptos necesarios para un buen manejo de políticas de gestión de datos de investigación e información científica en Chile.

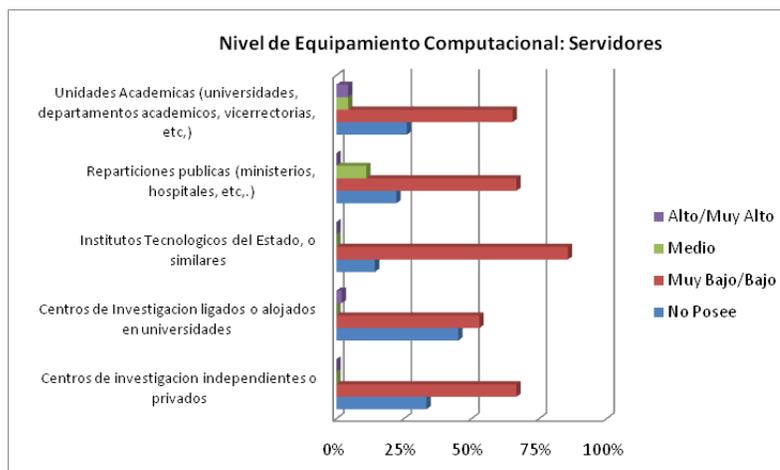
Resultados y conclusiones sobre infraestructura tecnológica

La disponibilidad de Infraestructura es muy diferenciada según los tipos de instituciones pero el escenario general es la poca disposición de equipamiento pertinente a las labores de gestión de datos de investigación e información científica.

Los resultados arrojaron además que mayoritariamente se observa un bajo número de administradores de bases de datos, una carencia de protocolos para los procedimientos de respaldo, formatos manuales y semiautomáticos para el tratamiento de datos de investigación e información científica, escaseando métodos automáticos o más avanzados.

Como se pudo observar en los resultados, existe una clara ausencia de procesos formales, protocolos y normativas internas para definir, estandarizar y ejecutar procesos de respaldo de datos de investigación e información científica lo que demuestra que la mayoría de estos procesos quedan a merced de los investigadores. Esta situación genera inseguridad pensando en cambios institucionales, los cambios de investigadores entre instituciones, la inserción de nuevos investigadores quienes cada uno a su manera impregnan nuevos procedimientos.

El poder ahondar en detalle en los requerimientos en equipamiento e infraestructura, es un estudio por sí solo, pero estos primeros resultados dan cuenta de una enorme debilidad en materia de equipamiento lo que va acompañado de dos aspectos fundamentales: la incapacidad de las mismas instituciones de poder detectar cuáles son sus requerimientos pensando en los estándares internacionales y el no disponer de



profesionales competentes para el uso de estas tecnologías. Ambos aspectos van en un detrimento y estanco en materia de adquisición y uso de equipos pertinentes para la gestión de datos de investigación e información científica.

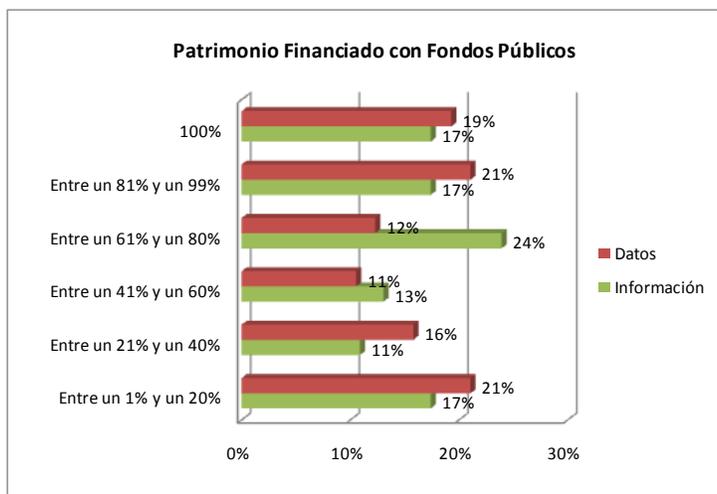
Respecto a las áreas científicas, a pesar de que ingeniería y tecnología, ciencias médicas, ciencias sociales y las multidisciplinarias, en algunos aspectos han destacado con mayores capacidades en infraestructura y equipamiento, los resultados son claros y categóricos ya que no existen áreas que posean buenos resultados en todos los ámbitos identificados en este estudio en lo que respecta a infraestructura. El poder disminuir esa brecha es importante debido a que poseer alta capacidad técnica en el uso de la infraestructura tecnológica facilita el proceso investigativo y en consecuencia, la permisibilidad en la transferencia de datos de investigación e información científica.

Las áreas científico-tecnológicas trabajadas en Chile tienen un alto potencial de intercambio, debido a que las instituciones donde éstas son trabajadas, poseen alto nivel de transferencia para sus datos de investigación e información científica. Los aspectos técnicos deficitarios son los referentes a la capacidad de cálculo y los mejor evaluados son los concernientes a la seguridad de las redes y mantención y respaldo.

A pesar de disponer de infraestructura y equipamiento en muchos casos, se observa una mayor falta de protocolos para respaldar datos de investigación e información científica por la mayoría de las áreas científicas. En este sentido, se carecen de normativas y políticas internas las que están totalmente supeditadas a las prácticas de los propios investigadores. Esto demuestra que las áreas no tienen un cuerpo o consistencia interna, es decir, que las áreas científico-tecnológicas trabajadas en Chile, tienen un nivel de gestión heterogéneo lo que dificulta el progreso y avance de las disciplinas y áreas científicas en su totalidad, esto tiene como síntoma el total desequilibrio y comportamientos asimétrico de una misma área científica en el país, dificultando lo que a su vez dificulta aplicar políticas de mejoramiento a nivel global debido a que la realidad de las áreas queda supeditada a las vicisitudes de los cientos de instituciones e investigadores que la gestionan y trabajan.

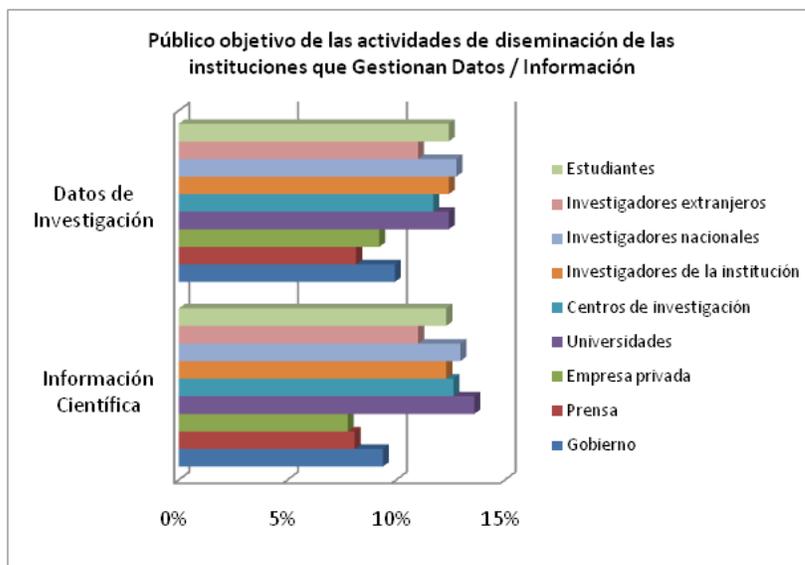
Resultados y conclusiones sobre patrimonio

Los datos de investigación e información científica generada en las instituciones chilenas son mayormente financiados con fondos públicos. Esta situación, si bien no es equivalente a pensar en un Estado con facultades de limitar el control que las mismas instituciones puedan hacer de su patrimonio, sí le confiere una mayor responsabilidad respecto al acceso de estos datos de investigación e información científica generados con fondos públicos debido a que adquieren un carácter patrimonial.



Los eslabones de la cadena de gestión de datos de investigación e información científica, donde son considerados la obtención, almacenamiento, procesamiento, validación, respaldo y diseminación, son realizados por las instituciones mediante procedimientos manuales en mayor medida; sin descartar la posibilidad que ciertos datos e información deban ser procesados de forma menos automatizada, se puede evidenciar un gran atraso en métodos más avanzados y automáticos para realizar la gestión de datos de investigación e información científica, así como los actuales motores de bases de datos utilizados y los bajos niveles de integración que poseen las bases de datos.

Respecto al tratamiento del patrimonio con el entorno de las instituciones, los resultados arrojaron que no existen preferencias a un sector en específico para la entrega o restricción de datos de investigación e información científica, ni así el uso de transacciones en dinero para la adquisición de este patrimonio. Sin embargo, algunas instituciones señalan que el cobro en dinero, por ciertos datos de investigación e información científica, responde a las características del patrimonio solicitado, es decir, cierto margen de patrimonio de carácter estratégico puede ser entregado a otros actores mediante la firma previa de convenios que resguarden el buen uso del patrimonio requerido.



Las razones más señaladas como merecedoras para restringir el acceso de otros actores a los datos de investigación e información científica generados por las instituciones son:

Protección a la privacidad de personas o instituciones: instituciones que señalan no difundir ni diseminar su patrimonio arguyendo a la protección de derecho de autor y propiedad intelectual.

Falta de normativa que regule y garantice buen manejo de nuestro patrimonio: estas instituciones señalan que en Chile no se ha desarrollado una normativa que permita discernir de buena manera qué patrimonio compartir y cuál no. En este sentido se prefiere restringir el acceso para evitar el mal uso del patrimonio difundido/diseminado.

Mantener ventaja competitiva: las instituciones señalan que al mantener el patrimonio en reserva se evita que otros actores puedan aprovechar este material y posicionarse sobre la institución aprovechándose del trabajo realizado y no entregando nada a cambio.

Restricciones en las bases de los fondos que financian nuestro patrimonio: refiere a que las mismas normativas de los fondos con los que desarrollan y gestionan el patrimonio impide, de acuerdo a las bases y sus normativas, la libre disposición y entrega del patrimonio.

Obtener más control sobre el uso de nuestro patrimonio: refiere a la mayor autonomía que puedan tener las instituciones respecto al patrimonio que ellas mismas generan. Esta razón es señalada mayormente sólo por instituciones que gestionan datos de investigación.

Falta de recursos para asegurar un acceso seguro y expedito: refiere a la imposibilidad de presentar de mejor manera, de forma accesible y con estándares mínimos, el patrimonio a diseminar debido a la carencia de recursos económicos y de capital humano.

La diseminación es percibida por las instituciones con ventajas y desventajas, en el siguiente cuadro se detallan las recogidas en el estudio:

| Ventajas | Desventajas |
|---|---|
| <p>a) Posicionamiento y/o reconocimiento institucional: la <i>diseminación</i> otorgaría un mayor prestigio científico a las instituciones debido a que se convierten en referentes para otros actores al depender de los datos que diseminan.</p> | <p>a) Limitaciones por el mal uso de información: el diseminar información científica es considerado un riesgo por las instituciones por el mal uso que otros actores puedan hacer de dicho patrimonio.</p> |
| <p>b) Contribución al desarrollo del campo científico tecnológico: las instituciones consideran beneficioso para el avance científico-tecnológico en su conjunto, el incorporar prácticas de <i>diseminación</i> del patrimonio.</p> | <p>b) Limitaciones a la competitividad y/o patrimonio: las instituciones señalan que al diseminar su patrimonio se limitan su competitividad, en términos científicos, ya que permitirían que otras instituciones se adjudiquen concursos, publiquen y/o reciban financiamiento a partir de su patrimonio.</p> |
| <p>c) Generación de redes de intercambio y procesos de retroalimentación: contribuye a generar vínculos con otras instituciones</p> | <p>c) Limitaciones presupuestarias y de financiamiento: las instituciones aluden una falta de recursos económicos para realizar estas</p> |

| | |
|---|---|
| <p>mediante el intercambio de patrimonio.</p> <p>d) Permite la transparencia efectiva del gasto público: la <i>diseminación</i> se percibe como señal transparente debido a que el patrimonio generado con fondos públicos es trabajado por las instituciones con fines científicos y es un bien para el país.</p> <p>e) Generación de ventajas a sectores productivos: la <i>diseminación</i> de los datos de investigación se considera pertinente para el desarrollo de sectores productivos y el aumento de su competitividad.</p> <p>f) Vinculación del medio con la institución: la <i>diseminación</i> es considerada una forma de vinculación con actores e instituciones regionales, nacionales e internacionales.</p> <p>g) Mayores exigencias a la calidad de la información³⁰ diseminada: al diseminar el patrimonio las instituciones consideran que el esfuerzo por un mejor procesamiento y presentación de los datos debe ser mayor, y por ende, una mejor práctica institucional.</p> <p>h) Conocimiento al servicio de la sociedad: la <i>diseminación</i> es considerada como una forma de contribuir al desarrollo ya que los datos pueden ser utilizados por estudiantes de pre y postgrado, servicios públicos y el sector privado.</p> | <p>labores.</p> <p>d) Limitaciones temporales: las instituciones aluden una falta de tiempo para realizar estas labores.</p> <p>e) Limitaciones en el interés de la difusión: las instituciones señalan que sus investigadores y encargados no tienen mayor interés en divulgar su patrimonio.</p> <p>f) Limitaciones en la interpretación de la información: las instituciones aluden que otros actores, usuarios del patrimonio diseminado, puedan interpretar los resultados de mala forma debido a la gran especialización de las características del patrimonio que disponen.</p> |
|---|---|

Existe muy poca diferenciación en las áreas científicas respecto a la proporción de patrimonio financiado con fondos públicos, hecho que ocurre tanto para las instituciones que gestionan datos de investigación como información científica. Bajo una mirada global, sólo es posible distinguir áreas científico-tecnológicas que cuentan con patrimonio financiado mayoritariamente con fondos públicos y áreas científicas con un similar financiamiento del sector público y no público.

³⁰ Por "información" en este caso se entiende como patrimonio, como el acervo y conjunto de bases de datos.

No existe una correspondencia entre áreas científicas y el público objetivo de las actividades de diseminación, en este sentido destaca una amplia gama de entidades objeto de diseminación por cada una de las áreas científicas desarrolladas en Chile y con distribución uniforme en cuanto a los porcentajes para cada una. Como foco de diseminación es necesario subrayar la mayoritaria importancia del público nacional especializado, tales como Investigadores pertenecientes a su institución e Investigadores nacionales por sobre el público que no pertenece al mundo científico (prensa y empresa privada).

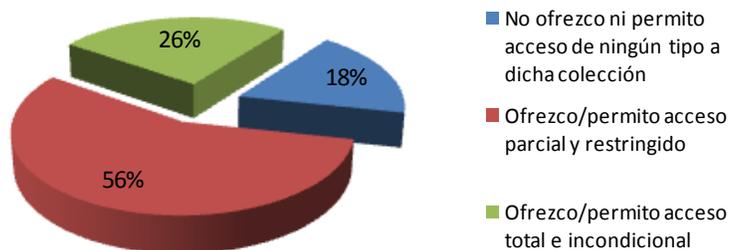
Respecto a los investigadores los resultados arrojaron que existe una tendencia a compartir el patrimonio de datos de investigación e información científica de forma gratuita. Sin embargo, esta situación se complejiza en la medida que no existen garantías para una adecuada diseminación y entrega del patrimonio que atiendan la propiedad intelectual, el derecho de autor y que no repercutan de forma negativa en la competitividad científica de los investigadores que entregan su patrimonio pudiendo ser mal utilizado por otros actores.

Al diferenciar a los investigadores según el foco de gestión, quienes gestionan sólo datos de investigación manifiestan ser bastante más restrictivos que quienes sólo gestionan información científica, mientras que el grupo que gestiona tanto datos como información presenta un mayor equilibrio en este punto.

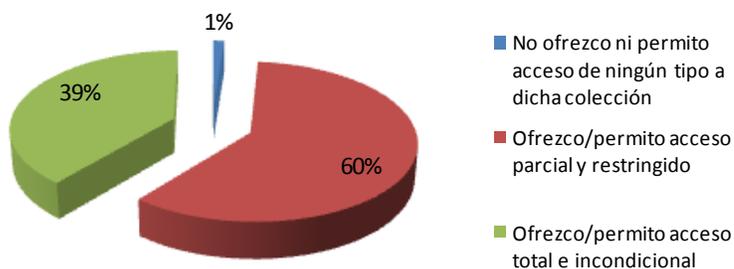
Especificando el análisis a áreas científicas no se encontraron grandes diferencias en el formato de acceso al patrimonio. Las áreas relacionadas con la gestión datos de investigación en su mayoría restringen el acceso al patrimonio a excepción de humanidades, ingeniería y tecnología, áreas donde se observa una mayor accesibilidad al patrimonio.

Para el grupo que gestiona sólo información científica, todas las áreas científicas presentan niveles de accesibilidad parcial, pero con importante proporción de investigadores que permiten el acceso total donde destacando ciencias sociales, ciencias naturales, ingeniería y tecnología y humanidades.

Accesibilidad de Entorno al Patrimonio de Datos



Accesibilidad del Entorno al Patrimonio de Información

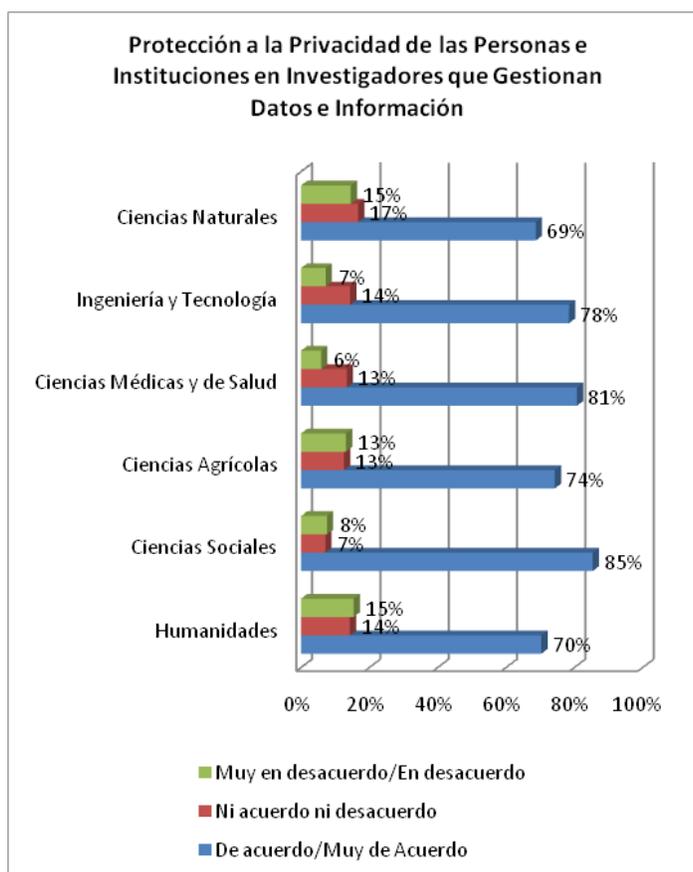


El grupo que gestionan tanto datos de investigadores como información científica presenta un mayor balance que los anteriormente descritos, con una mayoría de investigadores que señala restringir y entregar acceso parcial a su patrimonio, además de proporciones similares de investigadores que señalan no ofrecer acceso y ofrecer acceso total a éste.

De acuerdo a las áreas científicas relacionadas con la gestión de datos de investigación e información científica de forma simultánea, serían ciencias naturales, ciencias sociales y humanidades las que ofrecen accesibilidad total y gratuita a su patrimonio.

Por su parte, entre los motivos que mayor importancia tienen los investigadores para restringir el acceso al patrimonio es la protección a la privacidad de las personas e instituciones involucradas en la generación de dicho conocimiento. Esto se presenta de mayor manera en ciencias médicas y sociales y para aquellos investigadores que gestionan sólo datos de investigación y los que gestionan datos de investigación e información científica de forma simultánea.

En términos generales no se aprecian grandes diferencias entre los investigadores que gestionan datos de investigación e información científica. No obstante, en el grupo que señala gestionar ambos, existe evidencia que permite señalar que poseen mejores capacidades de gestión ya que poseen mayor familiaridad con las variables críticas de la gestión: conceptos claves, accesibilidad a patrimonio, etc.

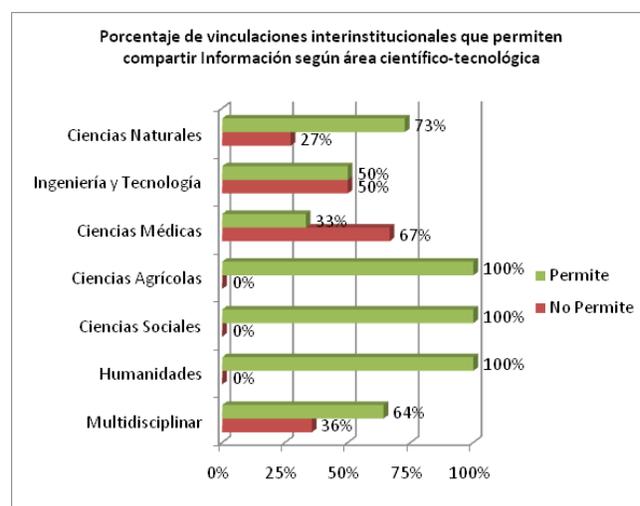
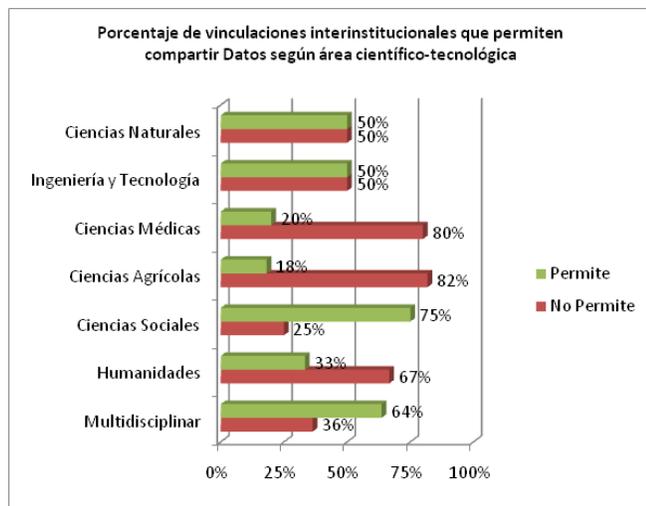


Resultados y conclusiones sobre vinculación y redes

Si se considera la vinculación de forma escalonada se pueden distinguir distintas etapas o momentos, en el más básico se encuentra la existencia de vínculos informales que nacen de la propia interacción de investigadores y/o funcionarios con pares de otras entidades, la relación en este caso se produce de forma natural y obedece a las características personales de cada investigador y no a una estrategia o política institucional por establecer vínculos. El segundo momento es cuando esos, u otros vínculos, se formalizan pactando modalidades de interacción y cooperación que quedan en un marco legal asignando responsabilidades, derechos y deberes para cada institución. El carácter de estos vínculos no posee mayor efecto en las actividades diarias de cada una de las instituciones ya que generalmente pasan por apoyo institucional, intercambios de investigadores y pasantías. El Tercer momento es el que interesa para la

buen gestión de datos de investigación e información científica y refiere a convenios firmados, que incluyan en sus cláusulas la posibilidad de intercambiar el patrimonio de datos de investigación e información científica generado por las instituciones, teniendo directa incidencia en el quehacer de las entidades.

El caso de las vinculaciones de las instituciones chilenas que gestionan datos de investigación e información científica mostró un alto nivel de presencia en redes internacionales de carácter mayormente académico y público. Además se evidenció una valoración positiva de estas redes para la actividad científica por parte de las instituciones.



La mayoría de las instituciones que gestionan datos de investigación plantea que no participa de redes, proyectos o programas de cooperación debido a:

- a) **Unidad de reciente creación:** señalan que su unidad o institución no posee el tiempo suficiente de existencia para establecer vínculos.
- b) **Inexistencia de unidades para interactuar:** señalan no poseer conocimiento sobre redes, proyectos, programas o convenios de cooperación donde puedan compartir su patrimonio.
- c) **Existe otro canal destinado a esto:** muchas instituciones señalaron que ellas directamente no estaban conectadas a redes ya que la universidad, mediante direcciones de investigación y vicerectorías es la que genera estas vinculaciones.
- d) **Desconocimiento:** instituciones que señalan no conocer algún tipo de programa de cooperación o red para vincularse.
- e) **Ausencia de sistematización:** instituciones que señalan poseer un nivel de sistematización muy bajo y precario de sus bases de datos como para compartirlas en redes, proyectos o programas de cooperación.

A pesar de que la mayoría de las áreas científicas desarrolladas en Chile tienen altos niveles de interacción y permisibilidad para compartir tanto datos de investigación como información científica, ya

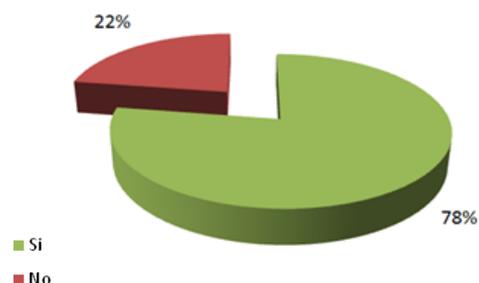
sea con instancias nacionales e internacionales, existen áreas como ciencias médicas, ciencias agrícolas y humanidades donde las posibilidades de compartir datos e información son mucho menores y escasas.

Por otra parte, el nivel de vínculos con posibilidad de compartir datos de investigación es mucho menor al que ocurre con la información científica en Chile, es por esto que preliminarmente se puede señalar que las diferencias no obedecen al nivel de las áreas científicas, sino al componente de lo que esté en cuestión, es decir, si a lo que se comparte o restringe son datos de investigación o información científica. De esta forma un área científica cualquiera comparte su información científica pero tiende a restringir sus datos de investigación.

Respecto a los análisis de acuerdo a las distintas etapas de la gestión de datos de investigación e información científica, los resultados no reflejaron especificidades en este nivel, es decir, no existe en Chile una especialización notoria en alguna(s) etapa(s) de la gestión, que permita atribuir algún beneficio en el desarrollo de estas por las actividades de vinculación.

Los investigadores, en su mayoría, declaran efectivamente participar en redes o convenios de cooperación que permiten compartir el patrimonio. De los restantes, es probable que una proporción menor participe en redes, pero que no permiten compartir el patrimonio. Se debe recordar que en el capítulo orientado al análisis de instituciones en materia de gestión de datos de investigación e información científica, un porcentaje muy menor de las redes de colaboración impide compartir el patrimonio.

Porcentaje de Investigadores que Participan en Redes o Programas de Diseminación de Datos y/o Información Científica

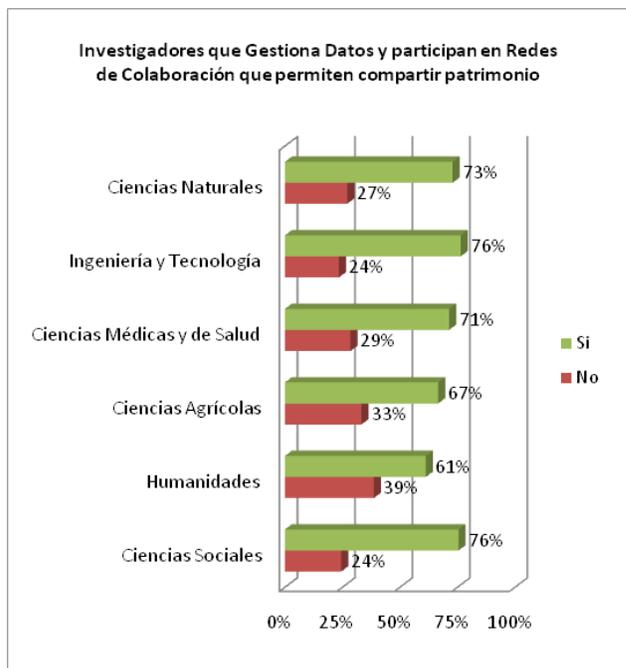


Según la diferenciación por foco de gestión, a pesar de presentar una mayoría de investigadores que participan en redes de colaboración que permiten compartir el patrimonio, los investigadores que sólo gestionan información científica presentan menor proporción de participación en redes.

Las áreas científicas con una mayor proporción de investigadores que participan en redes de colaboración que permiten compartir el patrimonio son ingeniería y tecnología y ciencias sociales, ciencias naturales y ciencias médicas.

A partir de estos antecedentes es posible afirmar que existe una importante proporción de investigadores chilenos que participan en redes de colaboración que permiten compartir el patrimonio, no obstante es importante considerar que muchas de estas redes se establecen en el marco de determinados proyectos y por lo tanto finalizan con ellos.

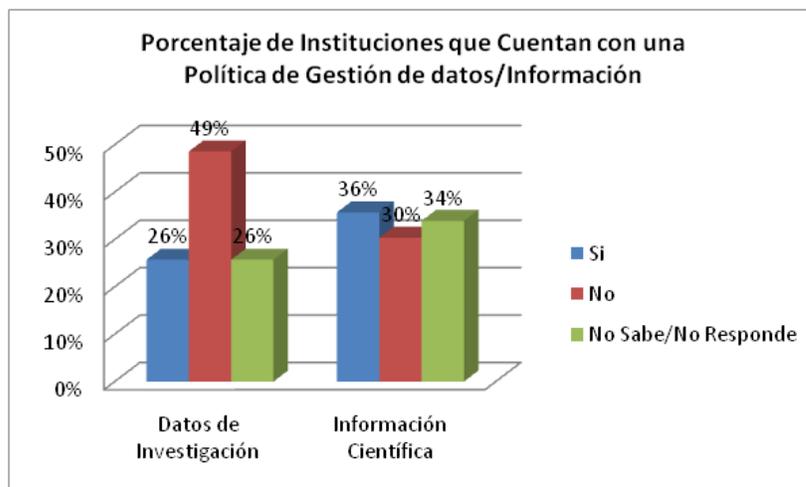
Además, para que la situación ideal manifestada por los investigadores, en el sentido de dejar a disposición su patrimonio a otras entidades e investigadores, requiere el diseño de instrumentos y normativas que resguarden la actividad científica en lo referente a propiedad intelectual y derechos de autor, estableciendo parámetros que permitan distinguir, entre lo que es el legítimo derecho de protección de determinado conocimiento de carácter estratégico y por otra parte el deber del Estado de velar por un uso adecuado y fructífero de la ciencia que se financia con fondos estatales la que tiene un carácter patrimonial y en pro del bien común.



Resultados y conclusiones sobre política de gestión

El disponer de políticas de gestión de datos e información resulta fundamental para definir patrones y protocolos para el uso y gestión del patrimonio, además de ser una forma de coordinación de los esfuerzos de la institución por una gestión eficiente y eficaz de los procesos y resultados de la actividad científica

Los resultados evidenciaron una clara brecha entre la realidad nacional y las buenas prácticas ya que la mayor parte de las instituciones que gestionan datos de investigación e información científica no poseen políticas de gestión y en otras ocurre que algunos de sus funcionarios ni siquiera saben de su existencia. Respecto a las instituciones que sí poseen políticas de gestión para regular sus procedimientos, el nivel



de aplicación que poseen es sólo por algunas personas y además son sólo indicativas, por lo que existe una ausencia de protocolos, manuales de operaciones y niveles de formalización básicos con que gestores y profesionales, al interior de las instituciones, puedan operar en relación a estas materias. En los casos

donde sí existen políticas de gestión, lo que ocurre mayormente es que éstas no se aplican debido a que su implementación no es viable debido a la carencia de recursos y la escasa institucionalización que poseen por la incapacidad de modificar prácticas e incorporar protocolos.

Respecto a las políticas de gestión en las áreas científicas existe una menor disponibilidad de políticas, y nivel de implementación de las mismas, en aquellas áreas científicas donde se gestionan mayormente datos de investigación en relación a aquellas donde se gestiona información científica. Las áreas científicas donde mayormente se detectó la disposición de políticas de gestión son ciencias naturales, ciencias médicas y humanidades. Las tres pertenecientes a instituciones que gestionan información científica.

Se detectó además que los investigadores no poseen información clara sobre el alcance y detalle de las políticas institucionales cuando existen en sus instituciones.

En lo concerniente al comportamiento de las políticas de gestión, pensando en las distintas etapas de la gestión: *planificación, captura/obtención, almacenamiento, procesamiento, validación y diseminación*, se puede observar que los resultados no evidencian algún tipo de relación entre la realización de alguna etapa de la gestión de datos de investigación e información científica y la existencia de una política, el nivel de implementación de ésta o la percepción respecto a una política nacional de gestión. Esto evidencia lo precario y poco especializado escenario de las instituciones chilenas respecto a la diferenciación y especialización de sus etapas de la gestión.

| Beneficios | Perjuicios |
|--|---|
| <p>a) Mejoramiento del rendimiento y eficiencia de los recursos: Permitiría economizar recursos de todo orden: ahorro del presupuesto público, recursos humanos dentro de la institución y tiempo de gestión</p> <p>b) Facilitación de conformación de redes de intercambio y difusión de resultados: Permitiría establecer redes formales de difusión de resultados y trabajo colaborativo interinstitucional al estandarizar indicadores que facilitarían la comparación e interacción, generando ventajas para la institución al darse a conocer mediante la visualización de la investigación realizada.</p> <p>c) Facilitación de acceso a datos e información: Garantizaría el acceso público a datos/ información estableciendo con claridad la obligación de las instituciones de</p> | <p>a) Aumento de la burocracia: Que la política nacional se constituya en una traba burocrática que entorpezca en vez de facilitar la gestión que realizan las instituciones</p> <p>b) Imposible homogeneizar los datos: Imposibilidad de homogeneizar datos/información por tratarse de campos epistemológicos muy diferentes, con lógicas de trabajo y objetivos esperados diferentes</p> <p>c) Limitaciones a partir del investigador: El investigador debe decidir hasta cuándo deben guardarse los datos/investigación, pues es él quien estima el momento en que su divulgación es importante y/o necesaria</p> <p>d) Inexistencia de interés institucional: No existe, por parte de la institución</p> |

| | |
|---|--|
| <p>ponerla a disposición.</p> <p>d) Potenciaría el desarrollo de la investigación: Facilitaría la redistribución de información, permitiendo mejorar el desarrollo y desempeño de las investigaciones futuras eliminando la duplicidad de investigaciones respecto a un mismo tema.</p> <p>e) Propensión a un modelo de estandarización y mejoramiento del modelo de gestión: Permitiría sistematizar una labor que hoy es un tanto informal transparentando parámetros en la entrega y medición de indicadores. Existiría un marco que regule y centralice, de alguna manera, los datos/información generados.</p> | <p>motivaciones ni interés en tal política</p> |
|---|--|

Algunas instituciones señalaron que la existencia de beneficios y/o perjuicios dependería de los siguientes factores:

a) Depende si constituye una desventaja comparativa: Que tal política no constituya una desventaja comparativa al poner a disposición de otros investigadores (nacionales o extranjeros) los resultados del trabajo de los investigadores de la institución y que puedan ser utilizados para generar publicaciones, patentes, etc.

b) Dependería de los costos, financiamiento y estructura del diseño: Dependería de quién asuma los costos, la modalidad de financiamiento y la estructura general del diseño.

c) Protocolos y política de compartición- protección de datos/información: Se deberían establecer protocolos de confidencialidad y resguardos jurídicos, pues no todas las instituciones están dispuestas a dejar en manos de otras instituciones sus datos, en algunos casos la información es estratégica y puede ser relevante para la formación de una empresa externa que comercialice la información.

d) Depende si potencia el desarrollo de la investigación: Depende de si facilita la redistribución de información, permite mejorar el desarrollo y desempeño de las investigaciones futuras eliminando la duplicidad de investigaciones respecto del mismo tema y una lectura más provechosos que la que actualmente es posible hacer. Normalizando los datos/información y facilitando análisis posteriores.

e) Depende si tiene una propensión a un modelo de estandarización y mejoramiento del modelo de gestión de datos e información: Depende de si permite sistematizar la labor de gestión, transparentando

parámetros en la entrega y medición de indicadores. De si crea un marco que regule y centralice, de alguna manera, los datos/información generados.

Por último, las sugerencias que las instituciones realizan para la construcción de una política son:

a) Subcontratar una institución especializada: instituciones que señalan la necesidad de contratar asesorías a entidades especializadas en estas temáticas para que pueda formular normas, instalación de software, diseño de base datos, realización de capacitaciones, capture la información, etc. Esta institución debe estar acreditada para esta función (con certificado de INN) que facilite la tarea a los investigadores y a los administradores públicos

b) Formar una red disciplinaria: instituciones que señalan la necesidad que en primer lugar se genere una red de acuerdo a las diferentes disciplinas de investigación, es decir, que existan políticas nacionales diferenciadas en áreas científicas según corresponda debido a las diferencias en la estructura de los campos científicos.

c) Plataforma centralizada: instituciones que solicitan que la disseminación del patrimonio sea de carácter obligatorio, de libre acceso, fácil y útil. Manteniendo la confidencialidad de los involucrados en los casos que lo amerite. Además señalan que para esto debe existir un ente centralizado que se encargue de llevar a cabo estas tareas

d) Financiamiento permanente para la gestión y difusión: instituciones que señalan que el financiamiento de la o las unidades a cargo de la gestión del patrimonio no dependa de proyectos concursables, sino disponga de un presupuesto estable para crear sistemas de gestión y contratar personas capacitadas para implementarlos y ejecutarlos.

e) Eliminar las cláusulas de confidencialidad: instituciones que señalan la necesidad de que se modifique el marco legal eliminando las restricciones que impiden la libre difusión de los datos de investigación e información científica.

f) Elaboración participativa de la política de gestión: instituciones que señalan la necesidad que la política se elabore e implemente de forma participativa, que no sea impuesta, sino que fuera un proceso en el que los investigadores y directores de instituciones participen desde su nacimiento.

g) Necesidades previas a la elaboración de la política de gestión: instituciones que señalan que antes de construir una política se debe definir con anterioridad la finalidad de una política, los beneficios y costos que tendría para su implementación en Chile.

h) Alinear estándares internacionales: instituciones que señalan la necesidad de detectar experiencias internacionales para poder ajustar una política nacional ajustándose a normativas y parámetros universales para tener compatibilidad y competitividad con el desarrollo científico mundial.

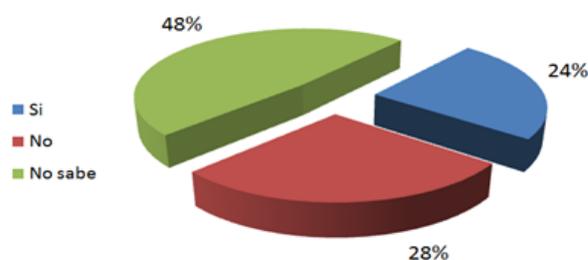
Para el análisis aplicado a investigadores fue posible observar importantes deficiencias en cuanto a la posesión e implementación de políticas y/o normativas que regulen la gestión de datos de investigación e

Información científica. En primer lugar destaca el desconocimiento de una importante proporción de los investigadores y académicos acerca de la existencia de esta en las instituciones donde se desempeñan. La misma situación de desconocimiento o no existencia de una política interna de gestión se extrapola con escasas variaciones entre los grupos de investigadores que gestionan datos o información. A este respecto, sólo el grupo que gestiona datos e información presenta niveles levemente más homogéneos de conocimiento y existencia de política de gestión, pero también con una mayoría de investigadores que no sabe acerca de la existencia de esta.

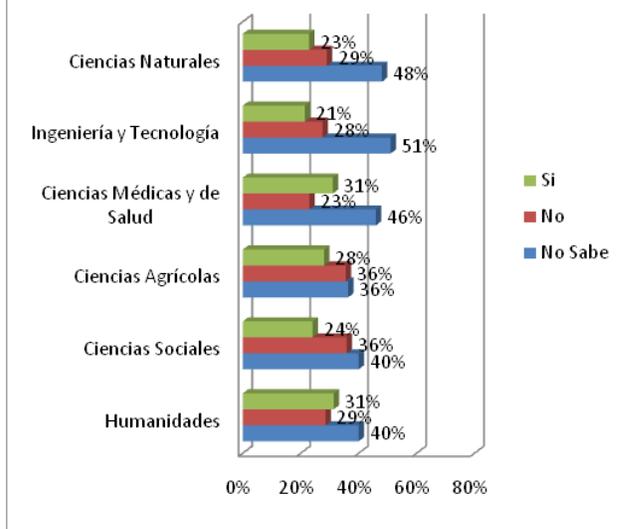
Dentro de este marco y tratando de diferenciar las áreas OCDE, se puede mencionar que para el caso de aquellos investigadores que mencionan sólo gestionar datos de investigación o sólo gestionan Información científica, destaca ingeniería y Tecnología y Ciencias médicas como aquellas que mejor situación presentan. Mientras que para aquellos investigadores que gestionan ambos, destacan ciencias médicas y humanidades como aquellas áreas que mejor conocimiento de política poseen.

Como especificación de lo anteriormente dicho, la aplicación de la política en aquel minoritario grupo de investigadores que señala poseer una política de gestión en su institución, destaca principalmente como aplicada de una forma irregular y poco sistemática, hecho que para los investigadores que sólo gestionan datos de investigación es ampliamente mayoritaria. Como diferencia importante puede señalarse lo ocurrido en los casos que gestionan tanto datos de investigación como información científica, quienes presentan bastante rigurosidad en la aplicación de la política.

Existencia de Política de Gestión de Datos de investigación y/o Información Científica

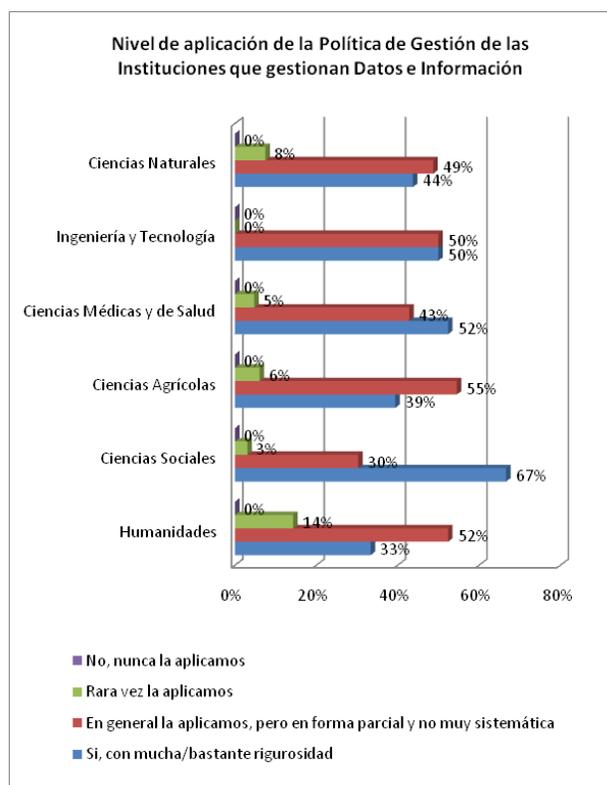


Existencia de una política de gestión en la institución de los investigadores que gestionan datos e Información



En cuanto a las áreas OCDE, es posible señalar que para quienes gestionan sólo datos de investigación destacan las ciencias sociales como aquellas que aplican más rigurosamente la política, seguidas por ingeniería y tecnología y ciencias médicas. Para el caso de aquellos investigadores que gestionan sólo información científica, destacan las ciencias naturales y las como aquellas que aplican más rigurosamente la política. En tercer lugar, para aquellos investigadores que gestionan datos e información destacan ciencias sociales y ciencias médicas.

No obstante todo lo anteriormente descrito, se destaca las expectativas positivas hacia la existencia de una política nacional de gestiona de datos y/o información científica, en cuanto a la consideración de la existencia de ventajas derivadas de su existencia, hecho que ocurre de igual forma en los distintos focos de gestión y las áreas OCDE.



Entre las ventajas y desventajas identificadas por los investigadores si existiese una política nacional de gestión de datos de investigación e información científica se encuentran las siguientes:

| Ventajas | Desventajas |
|--|--|
| <p>Mayor competitividad: una política puede ayudar a mejorar la competitividad científica entre investigadores e instituciones.</p> <p>Fortalecimiento de vínculos: una política de de gestión puede ayudar a generar mayor cooperación, interacción y potenciación de la investigación entre investigadores.</p> <p>Visibilización de la ciencia nacional: una política tendría incidencia directa en la mayor visibilización y difusión de la actividad científica chilena.</p> <p>Mayor eficiencia y eficacia: una política de gestión a nivel nacional permitiría conocer,</p> | <p>Exceso de burocracia: la política puede transformarse en un exceso de burocracia para el tratamiento de datos de investigación e información científica entorpeciendo así la actividad científica.</p> <p>Reducción de competitividad científica: entregar acceso a los datos de investigación e información científica puede provocar una limitación en la competitividad científica de las instituciones que diseminan dicho patrimonio.</p> <p>Tensión en los derechos de autor: se perjudicarían los derechos de los autores ya que cada uno tendría la obligación de diseminar el patrimonio que ha desarrollado independiente de que ellos así</p> |

| | |
|--|--|
| <p>de forma más inmediata, el tipo de actividades que desarrollan otras instituciones en una misma área científica evitando la replicación de éstas.</p> | <p>lo estimen conveniente.</p> <p>Dificultades en la implementación: por aspectos de normativas, recursos económicos, físicos y cultura científica, la instauración de una política de este tipo sería muy difícil de ejecutarse correctamente.</p> <p>Usos inadecuados del patrimonio: algunos investigadores consideran que dejar a disposición patrimonio de datos de investigación y/o información científica puede ser utilizado para malos propósitos por otros actores.</p> |
|--|--|

Además, otros investigadores señalan que la política podría traer **ventajas o desventajas dependiendo** de los siguientes factores:

Transparencia: depende si la política en su nacimiento, desarrollo y aplicación, incorpora la participación de los actores que involucrará su ejecución (científicos, funcionarios, etc.).

Eficacia: la política puede resultar muy eficiente o ineficiente dependiendo del nivel de trabajo que involucre desarrollarla y que no responda a tiempos que sean irreales para poder abarcar todo lo que implicaría un proceso como éste.

Voluntad de investigadores: a pesar de que una política esté bien ideada e implementada, si los investigadores e instituciones no son capaces de respetarla y participar para su buen funcionamiento, ésta no tendrá ninguna incidencia.

Nivel de acceso a patrimonio v/s derechos de autor: la política debe tratar de concebir de la mejor manera el límite que existe entre la entrega/disenminación de patrimonio financiado con fondos públicos y el resguardo a los derechos de autor y propiedad intelectual de los científicos que lo desarrollan.

Áreas científicas: la política debe limitar parámetros y procedimientos dependiendo de las áreas científicas ya que un diseño homogéneo, sin pensar en las diferencias disciplinarias puede desembocar en un fracaso.

Por último, las sugerencias que realizaron los investigadores para pensar una futura política de gestión de datos de investigación e información científica, son las siguientes:

- a) Acceso restringido:** los investigadores señalan que el acceso a datos de investigación e información científica debe quedar reducido sólo a los investigadores e instituciones que participan de aquel intercambio de información no dejándolo a disposición para cualquier tipo de actor que lo solicite.
- b) Acceso total con restricciones:** se sugiere que el acceso al patrimonio debe ser custodiado mediante simples filtros donde la persona o institución que ingrese a un repositorio online, se identifique dejando claro su nombre, institución y motivación para el uso de los datos de investigación y/o información científica capturada.
- c) Alinear con estándares internacionales:** el diseño de la política debe incorporar parámetros y niveles internacionales en las áreas científicas que correspondan para tener compatibilidad con otros países y poder generar más vínculos y redes científicas.
- d) Necesidad de hacer capacitaciones:** se solicita la realización de una amplia difusión de los beneficios de una política de gestión, tanto al interior de la comunidad científica como de la ciudadanía en general y/o capacitación de los investigadores que participarán de ésta debido a que estos procedimientos pueden ser muy distintos a las prácticas llevadas hasta ahora por los investigadores.
- e) Elaboración participativa de la política de gestión:** investigadores que señalan la necesidad que la política se elabore e implemente de forma participativa, que no sea impuesta, sino que fuera un proceso en el que los investigadores y directores de instituciones participen desde su nacimiento.
- f) Evitar trabas burocráticas:** el desarrollo de esta política debe contemplar que su ejecución no se convierta en un entramado burocrático que termine por entorpecer, en vez de facilitar, el desarrollo científico del país.
- g) Gestión descentralizada:** los investigadores sugieren que no debe existir un manejo centralizado de la gestión de datos de investigación e información científica, sino que se permita la gestión institucional de acuerdo a protocolos comunes.
- h) Implementar marcha blanca:** se sugiere disponer de un período de prueba, en la ejecución de la política, que permita diagnosticar problemas, solucionarlos y familiarizar a los investigadores con ella.
- i) Incentivar trabajo en red:** se sugiere desarrollar mecanismos que promuevan la participación y colaboración entre investigadores, fomentando el trabajo en equipos interdisciplinarios/interinstitucionales.
- j) Reglas claras y uniformes:** los investigadores sugieren que se deben establecer reglas claras, convertidas en ley, que garanticen la aplicación de la política de gestión de manera igualitaria a todas las instituciones e investigadores del país, estableciendo sanciones que penalicen el mal uso del patrimonio compartido.

IV.2. LA EXPERIENCIA CHILENA EN OCEANOGRAFÍA Y SISMOLOGÍA

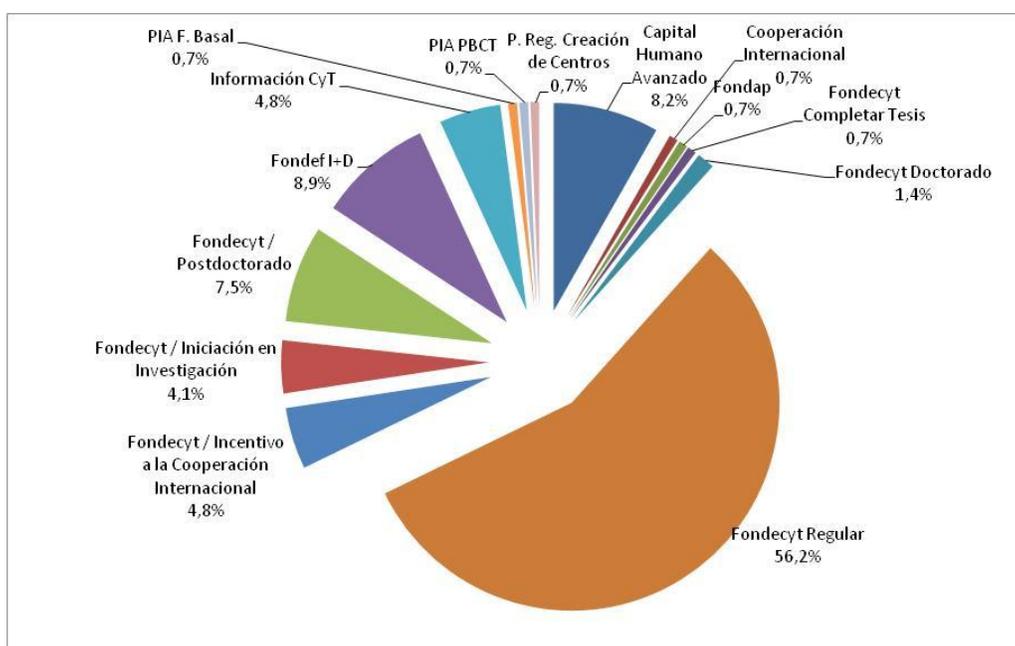
Capacidades Científico-Tecnológicas de la Oceanografía en Chile

A continuación se presenta una caracterización de las capacidades científicas y tecnológicas del área de Oceanografía. Comprende una descripción del conjunto de proyectos que se encuentran en el Repositorio Institucional de Conicyt en los siguientes ámbitos: instrumentos, años, instituciones, regiones, disciplinas y redes de instituciones e investigadores. Éste comprende un total de 147 proyectos y/o iniciativas seleccionadas bajo la categoría Oceanografía, la que corresponde a una sub disciplina de Ciencias de la Tierra dentro del repositorio.

Instrumentos

Es posible identificar una serie de instrumentos de financiamiento en el área de la Oceanografía. Si bien en la gráfica siguiente es posible visualizar un total de 14, éstos se pueden agrupar en 8, los que corresponden a Capital Humano Avanzado, Cooperación Internacional, Fondap, Fondecyt, Fondef, Información Científica y Tecnológica, Programa de Investigación Asociativa (PIA) y Programa Regional.

Instrumentos de Financiamiento Oceanografía 1984-2010

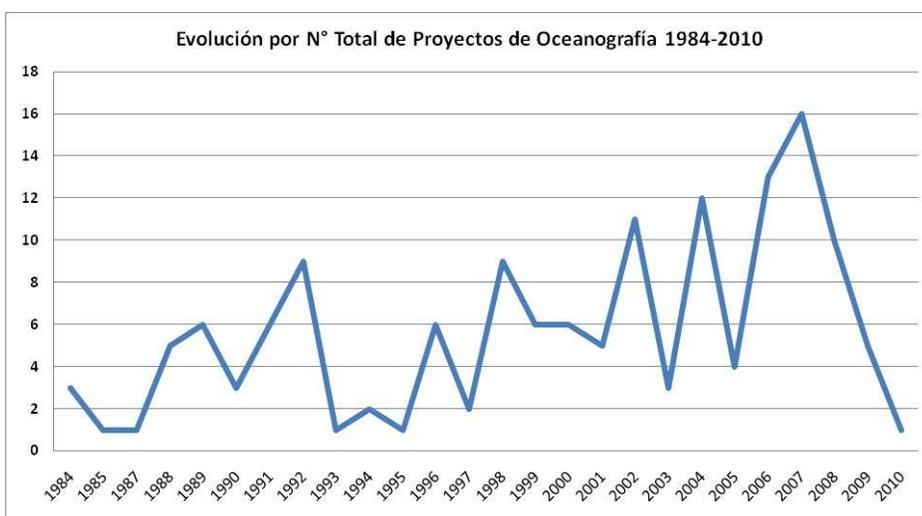


Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Se destaca la gran predominancia de Fondecyt, quien concentra más del 74% del total de proyectos, siendo Fondecyt Regular el más significativo, seguido muy por debajo por Fondef y el programa de Capital Humano Avanzado.

Años

El comportamiento del área de oceanografía a lo largo de los años, en relación al número de proyectos, es muy disperso y poco constante, presentando una gran diferenciación del número total de proyectos en los diferentes años de los cuales se disponen datos, inclusive triplicándose los proyectos en el transcurso de dos años y luego presentando un descenso importante. Sin embargo la explicación radica en la naturaleza propia del principal tipo de instrumento que financia los proyectos, ya que la duración estimada de proyectos Fondecyt es de 2 a 3 años, período en el cual los investigadores muchas veces se ven impedidos de contar con nuevos proyectos Fondecyt, ya que aun se encuentran en ejecución de los mismos, esto explicaría el comportamiento inestable del número total de proyectos, además da cuenta de la posible existencia de un número reducido de investigadores que ejecutan proyectos Fondecyt. Así mismo, la carencia de un programa nacional de investigación, que oriente este campo disciplinario, genera una inversión poco constante en los proyectos.



Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Pese a lo anteriormente expresado es posible establecer un aumento en los mínimos y máximos alcanzados en los distintos rangos de tiempo (2-3 años) existentes. Cabe destacar que no es posible efectuar aseveraciones con respecto al año 2010, ya que aun no se encuentran la totalidad de proyectos de Oceanografía en los datos disponibles.

Instituciones

En el presente apartado se expone la totalidad de instituciones principales ejecutoras de proyectos analizados, las cuales corresponden a un total de sólo 13. Tal como se puede apreciar en la siguiente gráfica existe una clara predominancia de instituciones de tipo académica, principalmente universidades. Además existe una clara superioridad en el número total de proyectos por parte de la Universidad de Concepción con cerca del 60%.



Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

A diferencia de la Universidad de Concepción, las instituciones presentan un comportamiento relativamente homogéneo en el número de proyectos, existiendo dos grandes grupos. El primero agrupa a 6 instituciones que se encuentran sobre los 5 proyectos en total. Mientras que el segundo lo componen igual número de instituciones, pero 3 de éstas presentan sólo un proyecto, las cuales corresponden a la U. de Magallanes, SERNAGEOMIN e I. Profesional Osorno.

Región

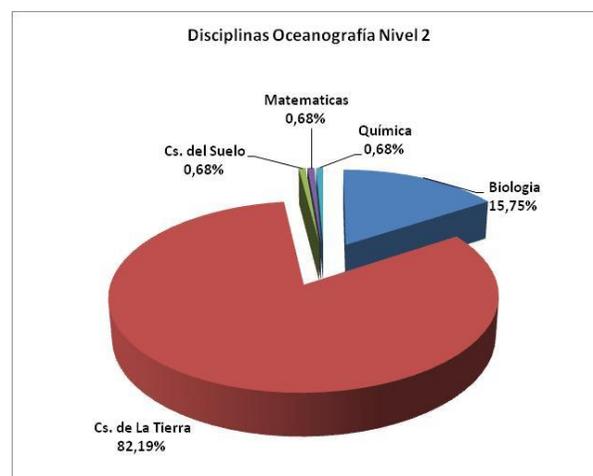
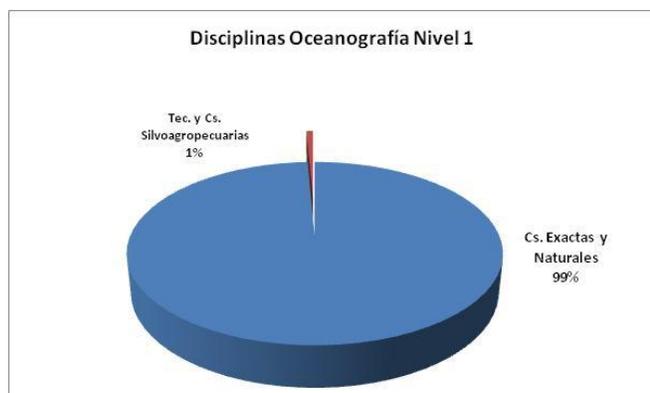
En lo que respecta a la distribución regional de los proyectos de investigación es posible afirmar que existe una directa relación con la ubicación de la institución responsable, es por ello que la Región del Bío-Bío es la que concentra la mayoría de los proyectos debido a la Universidad de Concepción. Muy por debajo se encuentran las regiones de Valparaíso, Los Ríos y Metropolitana.



Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Por su parte las regiones de Antofagasta, Coquimbo, Los Lagos y Magallanes y la Antártica Chilena presentan un número muy reducido de proyectos, inclusive con tan sólo un proyecto las dos últimas.

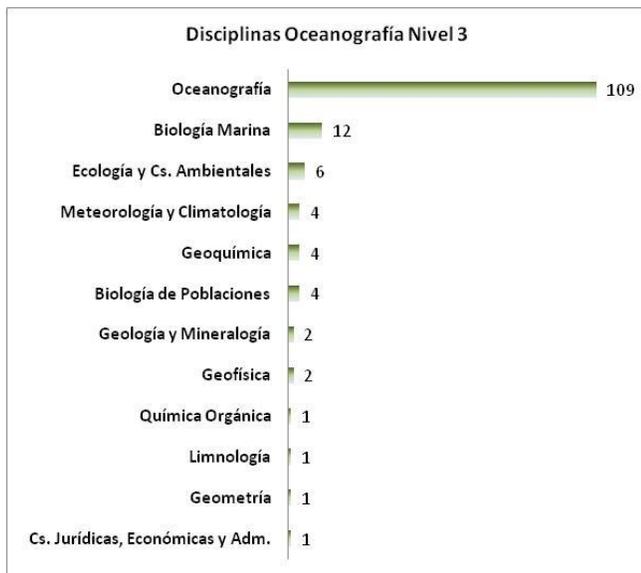
Disciplinas



Tal como se indicó anteriormente el análisis de las disciplinas se muestra en tres niveles diferentes, debido a la categorización que utiliza Conicyt. Para el caso de Oceanografía se muestran las dos primeras (más generales) en porcentajes totales en relación a la totalidad de proyectos de Oceanografía (gráfico de torta), mientras que en el tercer nivel se muestra en gráfico de barras las disciplinas ordenadas por la importancia que presentan en relación al número total.

En el primer nivel es posible identificar la presencia de sólo dos disciplinas principales, las cuales corresponden a Tecnología y Ciencias Silvoagropecuarias y a Ciencias Exactas y Naturales, siendo ésta última muy superior, ya que alcanza al 99% del total de proyectos. De dichas disciplinas se desprenden, en

un segundo nivel, cinco sub-disciplinas. En éste punto no existe una gran diversificación de disciplinas, existiendo una clara predominancia de Ciencias de la Tierra (82%), seguido de Biología (15%), mientras que con tan sólo el 0,68% cada una se encuentran Ciencias del Suelo, Matemáticas y Química.



Mientras que en el tercer nivel es posible encontrar un número superior de disciplinas (12) que en los niveles antes descritos. Sin embargo existe una considerable superioridad de oceanografía, por sobre las restantes disciplinas. Cabe destacar el alcance del área Oceanografía en su relación con diversas disciplinas de áreas como las meteorología, biología, química, física, geometría e inclusive con una relacionada con las ciencias sociales.

Redes

En el presente apartado se presenta un breve análisis de redes del área de la Oceanografía, el cual integra una muestra gráfica de la situación de interacciones de las instituciones y los investigadores, considerando a los principales, así como a los asociados en ambos casos.

Instituciones

La red de instituciones de Oceanografía está compuesta por un total de 100 instituciones, presentando niveles de vinculación muy dispares, ya que existen claras diferenciaciones en el número de instituciones y relaciones dependiendo el tipo de éstas, ya sean públicas, privadas y/o académicas. A continuación se presenta, a modo de resumen, una tabla que integra las principales instituciones, ordenadas por nivel de importancia, siendo ésta el número total de interacciones.

| Centralidad de Red Instituciones Oceanografía | | |
|---|-------------------------------|------------------|
| N° Ranking | Institución | N° Vinculaciones |
| 1 | U. de Concepción | 67.000 |
| 2 | Instituto de Fomento Pesquero | 26.000 |
| 3 | U. Austral de Chile | 22.000 |
| 4 | U. Cat. del Norte | 19.000 |
| 5 | Agromarina Huildad Ltda. | 16.000 |

| | | |
|----|--|--------|
| 6 | Armada de Chile | 15.000 |
| 7 | Cultivos Marinos Cenculmavique | 14.000 |
| 8 | Cultivos Martin Rojas Rojas | 14.000 |
| 9 | U. Cat. de la Santísima Concepción | 14.000 |
| 10 | Asoc. de Miticultores | 14.000 |
| 11 | Cultivos Marinos Mario Cerna Rosales | 14.000 |
| 12 | Automar Chiloe | 14.000 |
| 13 | Cultivos Marinos Vilupulli Ltda. | 14.000 |
| 14 | Granja Marina Cultivos Pleamar | 14.000 |
| 15 | Cultivos Osvaldo Galindo | 14.000 |
| 16 | Cultivos Eduardo Mondaca | 14.000 |
| 17 | Cultivos Marinos Doña Ana Ltda. | 14.000 |
| 18 | Marine Garden S.a. | 14.000 |
| 19 | Federación Gremial de las Industrias Pesqueras de la Macrozona X, XI y XII, (FIPES F.G) | 13.000 |
| 20 | Intesal | 13.000 |
| 21 | Comision Nacional de Investigación Científica Y Tecnológica | 13.000 |
| 22 | Asoc. de la Industria del Salmón (Salmón Chile A.G) | 13.000 |
| 23 | Dirección General de Aguas Región de Aysén | 13.000 |
| 24 | Gobierno Regional de Aysén | 13.000 |
| 25 | Instituto Chileno Campos de Hielo | 13.000 |
| 26 | Asoc. Cámara de Turismo de Coyhaique A.G. | 13.000 |
| 27 | Instituto de Investigaciones Agropecuarias | 13.000 |
| 28 | U. Of Montana | 13.000 |
| 29 | U. de Siena | 13.000 |
| 30 | Asoc. Gremial de Buzos Y Pescadores Y Ramas Similares Artesanales Independientes de Tongoy | 12.000 |
| 31 | U. de Chile | 10.000 |

Para efectuar una aproximación a la existencia de “grupos” al interior de la red, se procedió a determinar los Cliques existentes, lo que corresponde a una determinación de la totalidad de grupos que poseen una

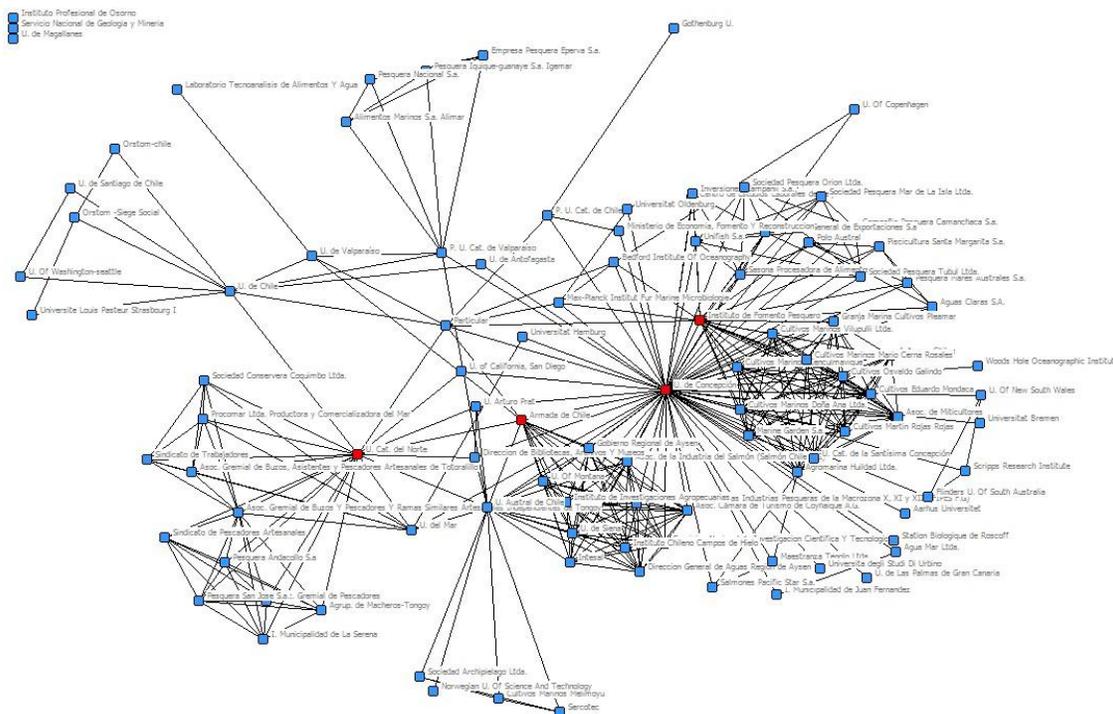
relación con 3 o más instituciones. Cabe destacar la existencia de 3 instituciones que no presentan vinculación alguna. La red de Oceanografía contiene un total de 32 grupos, los cuales se detallan en la siguiente tabla.

| | |
|--|--|
| 1: Agromarina Huilddad Ltda. Asoc. de Miticultores Automar Chiloe Cultivos Eduardo Mondaca Cultivos Marinos Cenculmavique Cultivos Marinos Doña Ana Ltda. Cultivos Marinos Mario Cerna Rosales Cultivos Marinos Vilupulli Ltda. Cultivos Martin Rojas Rojas Cultivos Osvaldo Galindo Granja Marina Cultivos Pleamar Instituto de Fomento Pesquero Marine Garden S.a. U. Cat. de la Santísima Concepción U. de Concepción | 2: Aguas Claras S.A. Instituto de Fomento Pesquero Pesquera Mares Australes S.a. Piscicultura Santa Margarita S.a. Polo Austral U. de Concepción |
| 3: Bedford Institute Of Oceanography Instituto de Fomento Pesquero Max-Planck Institut Fur Marine Microbiologie Particular U. de Concepción | 4: Consorcio General de Exportaciones S.a Instituto de Fomento Pesquero Inversiones Campanil S.a. Sasona Procesadora de Alimentos S.a. U. de Concepción Unifish S.a. |
| 5: Agua Mar Ltda. I. Municipalidad de Juan Fernandez U. de Concepción | 6: Armada de Chile Asoc. Cámara de Turismo de Coyhaique A.G. Asoc. de la Industria del Salmón (Salmón Chile A.G) Comision Nacional de Investigacion Cientifica Y Tecnológica Dirección General de Aguas Región de Aysén Federación Gremial de las Industrias Pesqueras de la Macrozona X, XI y XII, (FIPES F.G) Gobierno Regional de Aysén Instituto Chileno Campos de Hielo Instituto de Investigaciones Agropecuarias Intesal U. Austral de Chile U. de Concepción U. de Siena U. Of Montana |
| 7: Armada de Chile Particular U. de Concepción | 8: Centro de Estudios Laborales de Valparaíso P. U. Cat. de Chile U. de Concepción |
| 9: Compañía Pesquera Camanchaca S.a. Consorcio General de Exportaciones S.a Sasona Procesadora de Alimentos S.a. Sociedad Pesquera Mar de La Isla Ltda. Sociedad Pesquera Orión Ltda. Sociedad Pesquera Tubul Ltda. U. de Concepción Unifish S.a. | 10: Dirección de Bibliotecas, Archivos Y Museos U. Arturo Prat U. Austral de Chile U. Cat. del Norte U. de Concepción |
| 11: Flinders U. Of South Australia Scripps Research Institute U. de Concepción U. Of New South Wales | 12: Agromarina Huilddad Ltda. Maestranza Tenglo Ltda. Salmones Pacific Star S.a. U. de Concepción |
| 13: Ministerio de Economía, Fomento Y Reconstrucción P. U. Cat. de Chile U. de Concepción | 14: P. U. Cat. de Chile Particular U. de Concepción |
| 15: P. U. Cat. de Valparaíso U. Austral de Chile U. de Concepción | 16: U. Austral de Chile U. Cat. del Norte U. de Concepción U. del Mar |
| 17: U. Cat. del Norte U. de Concepción U. of California, San Diego | 18: Particular U. Cat. del Norte U. de Concepción |
| 19: U. de Concepción U. del Mar Universitat Hamburg | 20: Max-Planck Institut Fur Marine Microbiologie U. de Concepción Universitat Oldenburg |
| 21: Agrup. de Macheros-Tongoy Asoc. Gremial de Buzos Y Pescadores Y Ramas Similares Artesanales Independientes de Tongoy Asoc. Gremial de Pescadores I. Municipalidad de La Serena Pesquera Andacollo S.a Pesquera San José S.a. Sindicato de Pescadores Artesanales U. Cat. del Norte | 22: Alimentos Marinos S.a. Alimar Empresa Pesquera Eperva S.a. P. U. Cat. de Valparaíso Pesquera Iquique-guanaye S.a. Igemar Pesquera Nacional S.a. |
| 23: Asoc. Gremial de Buzos Y Pescadores Y Ramas Similares Artesanales Independientes de Tongoy Asoc. Gremial de Buzos, Asistentes y Pescadores Artesanales de Totoralillo Procomar Ltda. Productora y Comercializadora del Mar Sindicato de Trabajadores Sociedad Conservera Coquimbo Ltda. U. Cat. del | 24: Cultivos Marinos Melimoyu Sercotec Sociedad Archipiélago Ltda. U. Austral de Chile |

| | |
|---|--|
| Norte U. del Mar | |
| 25: Orstom -Siege Social Orstom-chile U. de Chile | 26: Orstom -Siege Social U. de Chile Universite Louis Pasteur Strasbourg I |
| 27: Particular U. Cat. del Norte U. de Chile | 28: Particular U. de Chile U. de Valparaíso |
| 29: P. U. Cat. de Valparaíso U. de Chile U. de Valparaíso | 30: U. de Chile U. de Santiago de Chile U. Of Washington-seattle |
| 31: Armada de Chile Particular U. de Valparaíso | 32: Bedford Institute Of Oceanography Instituto de Fomento Pesquero U. Of Copenhagen |

A continuación se presenta el grafo de red del área de Oceanografía, en dicho grafo se destacan en rojo aquellas instituciones detectadas de mayor importancia en relación al número total de vinculaciones.

Red de Instituciones Oceanografía



Fuente: elaboración propia IDER, 2010

Investigadores

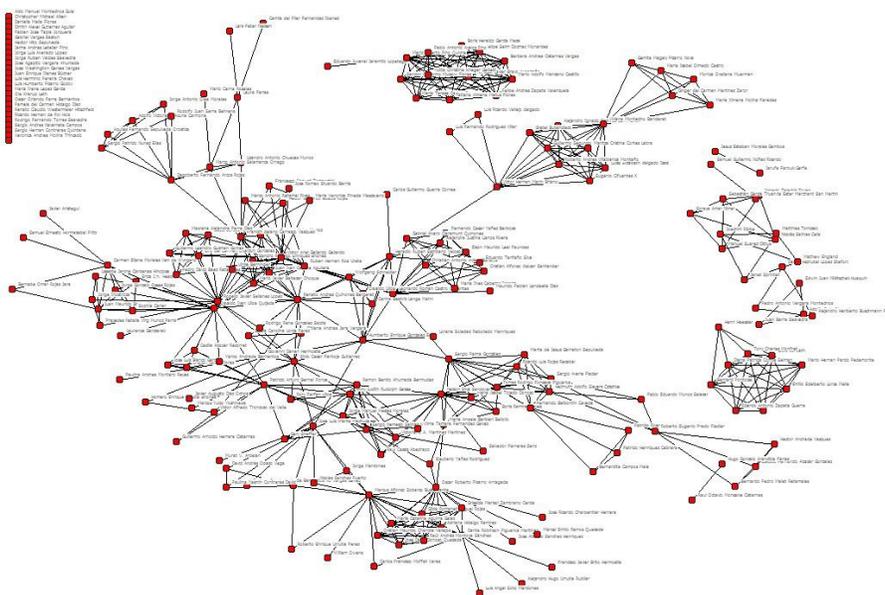
A diferencia del caso de instituciones el presente apartado muestra los principales investigadores del área de la sismología y la determinación del número total de cliques existentes, dado que la cantidad total de nodos existentes es muy alta.

219 son los investigadores que componen la red de Oceanografía, incluyendo tanto a los responsables, así como a los asociados. 25 de éstos no poseen vinculación alguna con otro investigador, mientras que los restantes se diferencian en gran medida unos de otros. A continuación se presenta un ranking de los 10 investigadores con mayor nivel de vinculación al interior de la red de Oceanografía.

| Centralidad de Red Investigadores Oceanografía | | |
|---|-----------------------------------|-------------------------|
| Ranking | Investigador (a) | N° Vinculaciones |
| 1 | Víctor Ariel Gallardo Gallardo | 22.000 |
| 2 | Renato Andrés Quinones Bergeret | 19.000 |
| 3 | Oswaldo Ivan Ulloa Quijada | 19.000 |
| 4 | Nelson Silva Sandoval | 18.000 |
| 5 | Franklin delano Carrasco Vásquez | 17.000 |
| 6 | Marcus Alfonso Sobarzo Bustamante | 16.000 |
| 7 | Leonardo Román Castro Cifuentes | 15.000 |
| 8 | Oscar Roberto Pizarro Arriagada | 15.000 |
| 9 | Rogelio Javier Sellanes López | 13.000 |
| 10 | Rubén Hernán Roa Ureta | 13.000 |

En lo que respecta a los grupos existentes al interior de la red es posible afirmar que existen un total de 57 cliques, que agrupan a todos aquellos que se encuentran en un grupo que contiene 3 o más investigadores. Cabe destacar que esta cantidad de cliques es muy significativo, lo cual no necesariamente supone una vinculación entre estos mismos grupos generados.

Red de Investigadores Oceanografía



Fuente: elaboración propia IDER, 2010

Lo expresado anteriormente es posible de apreciar en el grafo de red de investigadores de Oceanografía, en donde existe un gran número de grupos muy distantes y aislados entre sí, destacando sólo aquellos investigadores que cumplen el rol de intermediarios entre alguno de éstos grupos.

Capacidades Científico-Tecnológicas de la Sismología en Chile

Para el desarrollo del presente análisis se utiliza como de datos el Repositorio Institucional de Conicyt, en dónde la selección de los datos de sismología se estableció un proceso diferente al de Oceanografía, el cual contempló los siguientes pasos:

- 1° Selección de un total de 666 proyectos que corresponden a Ciencias de la Tierra, ya que ésta es la disciplina principal afín de posibles proyectos asociados a sismología.
- 2° Búsqueda de los principales conceptos afines a la Sismología.
- 3° Listado de conceptos afines, los cuales corresponden a:

| | | |
|---------------------|-------------------|---------------|
| - Sismología | - Sismo | - Seísmo |
| - Terremoto | - Temblor | - Telúrico |
| - Cataclismo | - Maremoto | - Marejada |
| - Tectónica (a) | - Tectónicos (as) | - Sísmico (a) |
| - Placa | - Nazca | - Pacífica |
| - Sudamericana | - Epicentro | - Hipocentro |
| - Corteza Terrestre | - Richter | - Mercalli |
| - Seismology | - Seismic | - Earthquake |
| - Earthshock | - Tsunami | - Sunami |
| - Tectonic (s) | - Sismic | - Telluric |
| - Cataclysm | - Epicentre | - Hypocenter |

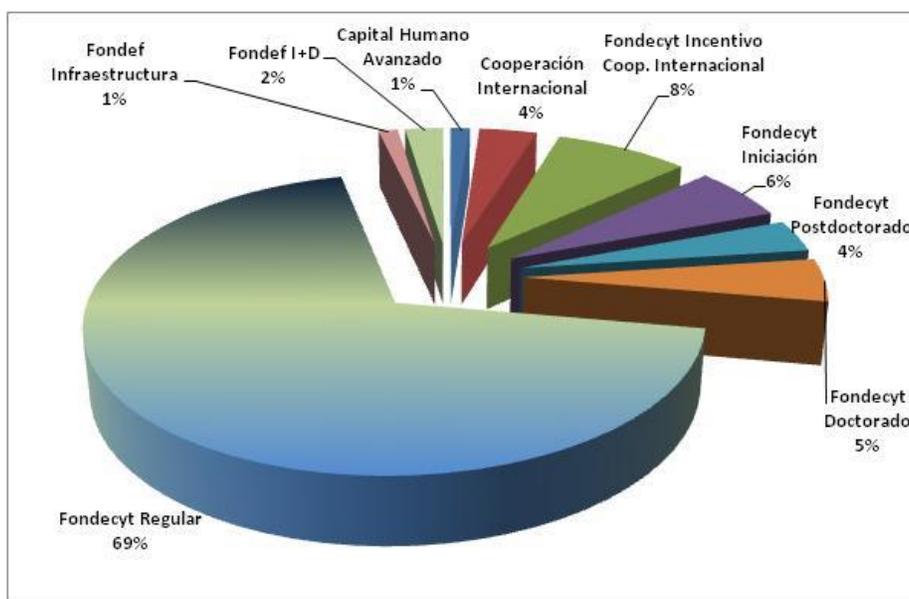
4° Búsqueda y selección de proyectos que contiene alguna de los conceptos clave seleccionados en los siguientes campos de la base de datos: título, cuerpo, descripción, institución y/o sector de aplicación.

Lo anterior se efectuó debido a que no existe una disciplina, o bien alguna categorización dentro del Repositorio Institucional de Conicyt que permita establecer proyectos y/o iniciativas asociados al área de la Sismología.

A continuación se presenta el análisis descriptivo de las capacidades científicas y tecnológicas en el área de la Sismología, el cual comprende los siguientes ámbitos de análisis: instrumentos, años, instituciones, región, disciplinas y red de instituciones e investigadores. Éste análisis comprende un total de 83 proyectos y/o iniciativas seleccionadas bajo los criterios expuestos anteriormente.

Instrumentos

Es posible identificar una serie de instrumentos de financiamiento en el área de la Sismología, los cuales alcanzan a 9. Sin embargo, dichos instrumentos es posible agruparlos en cuatro: Fondecyt, Cooperación Internacional, Capital Humano Avanzado y Fondef. El primero de éstos es el que aglutina mayor cantidad de proyectos con más del 90% del total de éstos. Mientras que los restantes corresponden a Cooperación Internacional (4%), Fondef (3%) y Capital Humano Avanzado (1%).

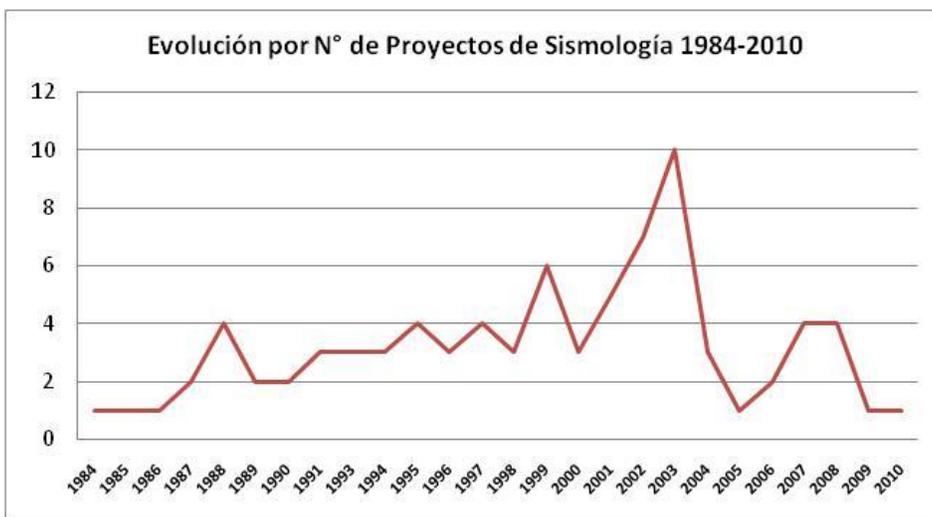


Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Cabe destacar la gran recurrencia de instrumentos asociados a investigación básica más que aplicada, como lo podría ser el Fondef.

Años

El comportamiento del área de Sismología a lo largo de los años, en relación al número de proyectos es muy disgregado por distintos periodos de tiempo, ya que presenta una gran diferenciación en el número total de proyectos en los diferentes años de los cuales se disponen datos. Desde 1988 a 1998 es posible encontrar un comportamiento homogéneo con leves disparidades entre los años, sin embargo en el año 1999 se produce un incremento de más del doble que al año anterior, produciéndose el año siguiente una baja significativa, para luego alcanzar en el año 2003 el máximo número de proyectos de todo el periodo en análisis, mientras que el año 2005 se cuenta con un proyecto, a diferencia del 2003 con diez. Lo que se puede apreciar en la gráfica en el último periodo de tiempo se explica por la naturaleza propia del principal tipo de instrumento que financia los proyectos, ya que la duración estimada de proyectos Fondecyt (sobre el 90% de total de proyectos) es de 2 a 3 años.



Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Destaca el reducido número de proyectos que existen en ésta área, ya que además del repunte en el 2004-2005 no existe un número significativo, a diferencia de otras áreas, como ejemplo de esto la Oceanografía, dando cuenta del reducido número de investigaciones del área y con un muy bajo nivel de incremento en el tiempo.

Instituciones

En lo que respecta a las instituciones principales ejecutoras de proyectos del área sismología es posible afirmar lo siguiente: existe una gran disparidad entre instituciones ejecutoras; destaca la existencia de instituciones principalmente de tipo académico; a nivel nacional es sólo una institución la que concentra más del 66% del total de proyectos ejecutados. Ésta última corresponde a la Universidad de Chile, que tal como se aprecia en la siguiente gráfica posee la mayor cantidad de proyectos del área a nivel nacional.

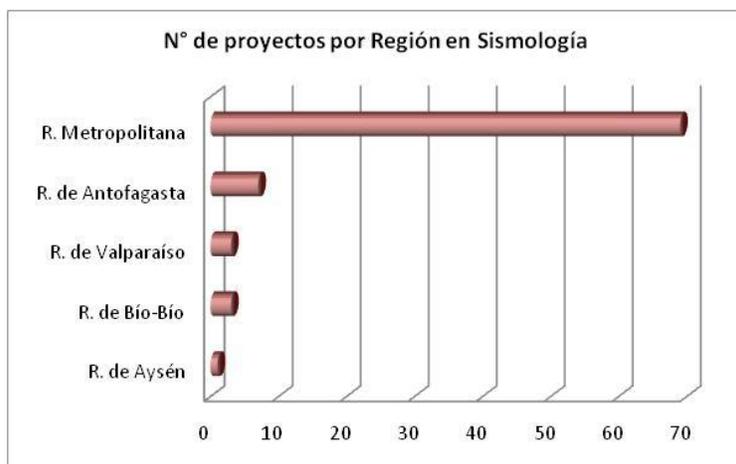


Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Destaca el bajo número de proyectos que presentan la mayoría de las instituciones, ya que de las 9 que existen, 7 se encuentran bajo los 10 proyectos, inclusive 3 de éstas poseen tan sólo 1 proyecto cada una.

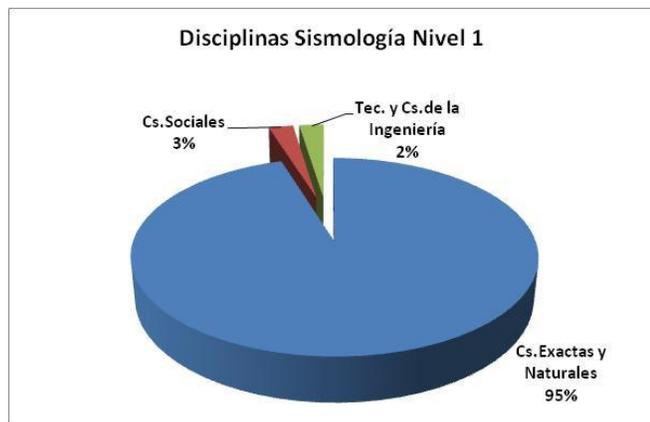
Regiones

Situación similar a la distribución por institución principal es la que presenta la distribución del número de proyectos por región, ya que es tan sólo una región la que concentra la casi totalidad de proyectos de un total de cinco. Además cabe destacar el bajo número de regiones en el país que han ejecutado proyectos en ésta área. Sin embargo no es posible afirmar categóricamente lo anterior, ya que la región asociada a la proyecto corresponde al domicilio de la institución principal ejecutora y no necesariamente se condice con la aplicabilidad del proyecto. Siendo así es la Universidad de Chile quien concentra mayor cantidad de proyectos, lo cual se condice plenamente con una mayor recurrencia de la Región Metropolitana, ya que la primera tiene domicilio en dicha región.



Fuente: Elaboración propia IDER-UFRO 2010, en base a Repositorio Institucional CONICYT

Disciplinas

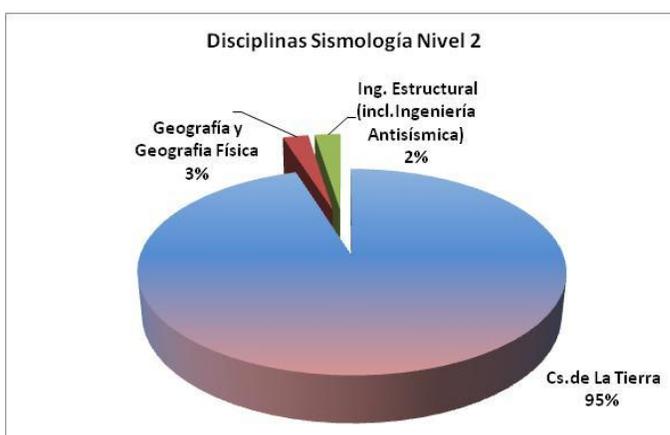


En lo que respecta a las disciplinas, es posible afirmar que existe una característica particular en términos del número de éstas en los distintos niveles, ya que son tan sólo tres disciplinas por cada uno de los niveles. Cabe recordar la segmentación de niveles indicada anteriormente, lo cual explica la realización del análisis en 3 niveles diferentes.

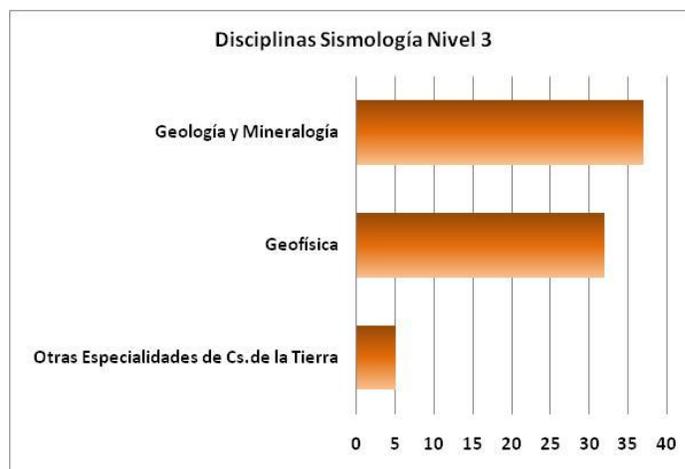
En el primer de los niveles la mayor concentración (95%) de proyectos corresponde a Ciencias Exactas y Naturales, quedando tan sólo un 2% para

Tecnología y Ciencias de la Ingeniería. Cabe destacar la existencia de Ciencias Sociales como disciplina principal dentro de sismología, la cual ocupa un 3% del total de proyectos.

Para el segundo nivel de disciplinas es posible encontrar disciplinas que se aproximan más a la sismología en una mayor profundidad. Sin embargo el número de disciplinas, tal como se indicó anteriormente, es sólo de 3, además concentrando un 95% del total en Ciencias de la Tierra, seguido muy por debajo por Geografía y Geografía Física (se asocia a Ciencias Sociales como sub-disciplina) y por Ingeniería Estructural, ésta última incluye Ingeniería Antisísmica.



Ya en el tercer nivel de análisis existe una distribución no tan pronunciada como en los niveles anteriores, ya que la mayoría de proyectos no se concentra en una sola disciplina. En éste nivel la disciplina más recurrente corresponde a Geología y Mineralogía, seguida muy de cerca por Geofísica, mientras que Otras especialidades de Ciencias de la Tierra es la que agrupa un menor número de proyectos. Cabe destacar que en la gráfica correspondiente al tercer nivel es posible visualizar el número total de proyectos por disciplina, mientras que en los niveles anteriores se realiza en porcentajes.



REDES

A continuación se presenta un breve análisis de redes, el cual integra una muestra gráfica de la situación de interacciones de las instituciones y los investigadores, considerando a los principales, así como a los asociados en ambos casos.

Instituciones

Para el caso de sismología existen un total de 39 instituciones que interactúan, las cuales tienen un nivel de relaciones bastante diferenciadas entre, ya que existen instituciones que poseen sólo una vinculación, mientras que existen otras con altos niveles de interacciones. A continuación se presenta una tabla que resume el número total de interacciones de cada una de las instituciones en orden de importancia.

| Centralidad de Red Instituciones Sismología | |
|---|---------------------|
| Institución | N° de Interacciones |
| U. de Chile | 24.000 |
| Serv. Nacional de Geología y Minería | 11.000 |
| Particular | 10.000 |
| Western WashingtonU. | 7.000 |
| U. Católica del Norte | 7.000 |
| U. de Buenos Aires | 6.000 |
| British Antarctic Survey | 5.000 |
| Australian NationalU. | 5.000 |
| U.e de Sao Paulo | 5.000 |
| U. de Concepción | 5.000 |
| Orstom-chile | 4.000 |
| Universite Louis Pasteur Strasbourg I | 4.000 |
| British Geological Survey | 3.000 |
| Cheltenham And Gloucester College Of Higher Education | 3.000 |
| Sociedad Internacional Petrolera S.a. | 3.000 |
| Geological Survey Of Japan | 3.000 |
| Ejército de Chile | 3.000 |

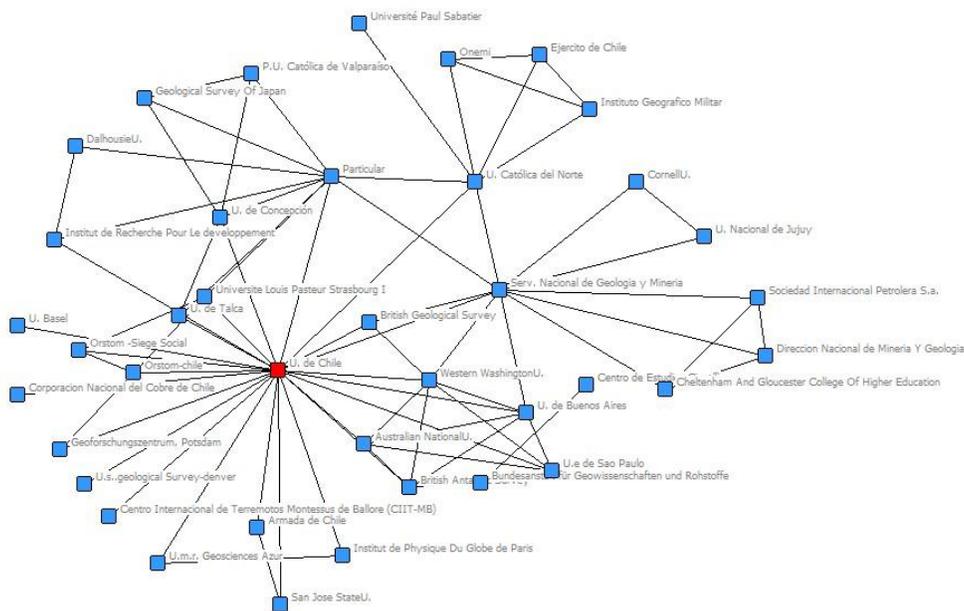
| | |
|---|-------|
| U. de Talca | 3.000 |
| Onemi | 3.000 |
| Orstom -Siege Social | 3.000 |
| Institut de Recherche Pour Le developpement | 3.000 |
| P.U. Católica de Valparaíso | 3.000 |
| Dirección Nacional de Minería Y Geología | 3.000 |
| Instituto Geográfico Militar | 3.000 |
| Institut de Physique Du Globe de Paris | 2.000 |
| U.m.r. Geosciences Azur | 2.000 |
| San José StateU. | 2.000 |
| U. Nacional de Jujuy | 2.000 |
| CornellU. | 2.000 |
| DalhousieU. | 2.000 |
| Armada de Chile | 2.000 |
| Geoforschungszentrum, Potsdam | 2.000 |
| Corporación Nacional del Cobre de Chile | 1.000 |
| Centro de Estudios Científicos | 1.000 |
| U. Basel | 1.000 |
| U.s...geological Survey-denver | 1.000 |
| Bundesanstalt für Geowissenschaften und Rohstoffe | 1.000 |
| Université Paul Sabatier | 1.000 |
| Centro Internacional de Terremotos Montessus de Ballore (CIIT-MB) | 1.000 |

En análisis de redes una aproximación a determinar la conformación de “grupos de investigación” es por medio de la detección de grupos que aglutinan a más de 3 nodos. Para el presente análisis corresponde al número total de grupos (cliques) que se presentan en la red de 3 o más instituciones. A continuación se presenta una tabla resumen que contiene a los 16 cliques existentes con la identificación de cada una de las instituciones que componen la relación.

| Número de Cliques Instituciones Sismología | |
|--|---|
| 1: Particular Serv. Nacional de Geología y Minería U. Católica del Norte U. de Chile | 2: Institut de Recherche Pour Le developpement Particular U. de Chile |
| 3: Particular U. de Chile U. de Concepción U. de Talca | 4: Particular U. de Chile Universite Louis Pasteur Strasbourg I |
| 5: Australian NationalU. British Antarctic Survey U. de Buenos Aires U. de Chile U.e de Sao Paulo Western WashingtonU. | 6: British Geological Survey Serv. Nacional de Geología y Minería U. de Chile Western WashingtonU. |
| 7: Geoforschungszentrum, Potsdam Orstom-chile U. de Chile | 8: Institut de Physique Du Globe de Paris U. de Chile U.m.r. Geosciences Azur |
| 9: Orstom -Siege Social Orstom-chile U. de Chile Universite Louis Pasteur Strasbourg I | 10: Armada de Chile San José StateU. U. de Chile |
| 11: Serv. Nacional de Geología y Minería U. de Buenos Aires U. de Chile Western WashingtonU. | 12: Cheltenham And Gloucester College Of Higher Education Dirección Nacional de Minería Y Geología Serv. Nacional de Geología y Minería Sociedad Internacional Petrolera S.a. |
| 13: CornellU. Serv. Nacional de Geología y Minería U. Nacional de Jujuy | 14: DalhousieU. Institut de Recherche Pour Le developpement Particular |
| 15: Ejercito de Chile Instituto Geográfico Militar Onemi U. Católica del Norte | 16: Geological Survey Of Japan P.U. Católica de Valparaíso Particular U. de Concepción |

Lo anteriormente expresado es posible visualizarlo gráficamente en el siguiente grafo de red de instituciones que corresponde a Sismología

Red de Instituciones Sismología



Fuente: Elaboración propia IDER-UFRO 2010.

Investigadores

A diferencia del caso de instituciones el presente apartado muestra las principales investigadores del área de la sismología y la determinación del número total de cliques existentes, dado que la cantidad total de nodos existentes muy alta.

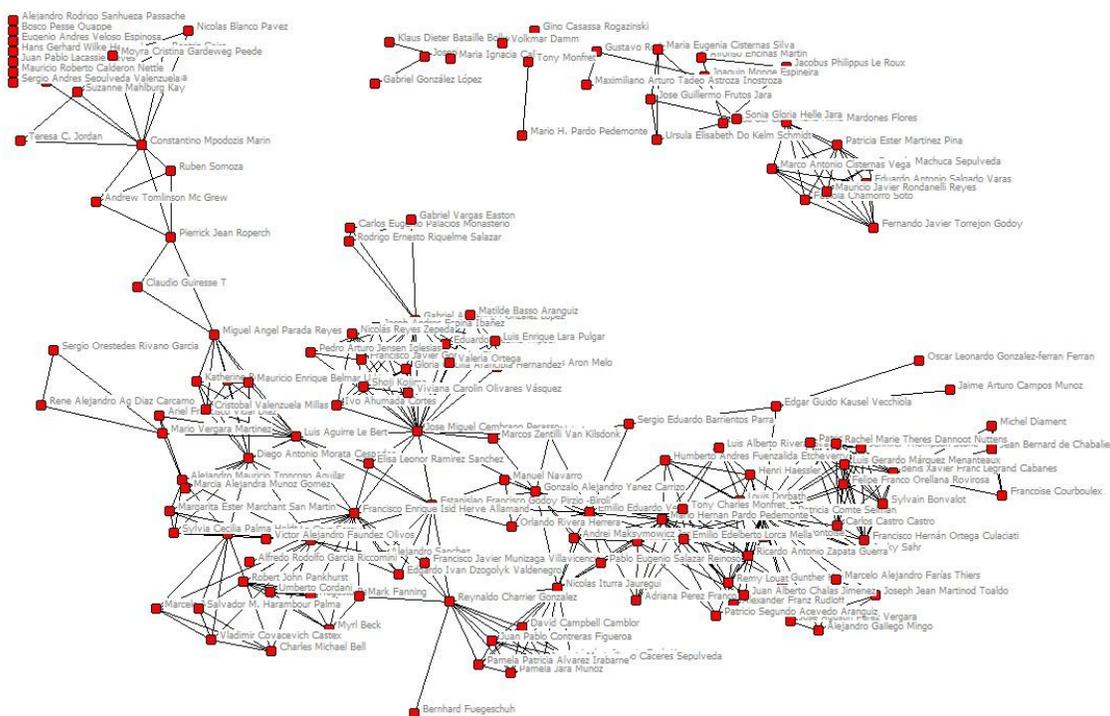
151 investigadores componen la red de sismología, considerando a investigadores principales y asociados. Cabe destacar la gran diferenciación que existe entre investigadores y el número de vinculaciones, existiendo 7 investigadores que no presentan vinculación alguna. A continuación se presenta un ranking de los 10 investigadores con mayor nivel de vinculación al interior de la red.

| Centralidad de Red Investigadores Sismología | | |
|--|---------------------------------------|------------------|
| Ranking | Investigador (a) | Nº Vinculaciones |
| 1 | Diana Patricia Comte Selman | 31.000 |
| 2 | José Miguel Cembrano Perasso | 21.000 |
| 3 | Mario Hernán Pardo Pedemonte | 21.000 |
| 4 | Francisco Enrique Isid Herve Allamand | 18.000 |

| | | |
|----|---|--------|
| 5 | Gabriel Armando González López | 17.000 |
| 6 | Emilio Eduardo Vera Sommer | 16.000 |
| 7 | Tony Charles Monfret | 15.000 |
| 8 | Luis Aguirre Le Bert | 13.000 |
| 9 | Estanislao Francisco Godoy Pirzio -Biroli | 13.000 |
| 10 | Manuel Suarez Dittus | 13.000 |

Un total de 45 cliques son los que existen al interior de la red de investigadores de sismología, lo cual da cuenta de buenos niveles de cohesión, pero no total, ya que como es posible de apreciar en el grafo de de red existen una serie de “grupos de investigación” no vinculados entre sí, además de la importancia de un número reducido de investigadores que son los nodos centrales y condicionan las relaciones entre los distintos grupos.

Red de Investigadores Sismología



Fuente: Elaboración propia IDER-UFRO 2010.

Dado lo anterior es posible hablar de una red de investigadores deficiente en lo que respecta a la relación entre grupos, ya que presenta buenos niveles de cohesión sólo en determinados grupos entre sí, no así entre éstos, los cuales no suman más de 8 o 9.

IV.3. PERCEPCIONES DE LOS AGENTES QUE REALIZAN GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

A continuación se presenta el análisis de las entrevistas realizadas a instituciones que gestionan datos de investigación y Unidades de administración y gestión de información científica, con la intención de profundizar sobre los resultados obtenidos del cuestionario y entrevista telefónica realizada en el marco del estudio denominado “Estado del Arte Nacional e Internacional sobre Gestión de Datos e Información Científica Financiados con Fondos Públicos”, teniendo como criterio de selectividad en la aplicación de las entrevistas la amplia trayectoria de estas instituciones en la materia.

Las instituciones y actores entrevistados, son los siguientes:

| Institución | Responsable |
|--|--|
| Centro de Microdatos, Universidad de Chile | Patricia Medrano, Profesora Facultad de Economía y Negocios e Investigadora Centro Microdatos. |
| Dirección de Investigación y Doctorado, Pontificia Universidad Católica de Chile | Rosario Del Villar, Sub Directora de Investigación y Doctorado |
| Comisión Chilena de Energía Nuclear | Eduardo Cortés, Jefe Oficina Asesora de Cooperación técnica y Relaciones Internacionales |
| Centro de Excelencia en Astrofísica | Guido Garay, Director |
| Vicerrectoría de Investigación y Desarrollo Departamento de Investigación, Universidad de Chile | Leonardo Reyes, Coordinador SEGI |
| Vicerrectoría de Investigación y Desarrollo, Universidad de Santiago | Mauricio Escudey, Vicerrector de Investigación y Desarrollo |
| Servicio Hidrográfico y Oceanográfico de la Armada de Chile | Gonzalo Vallejos, Jefe de Unidad de Gestión |
| Instituto de Fomento Pesquero | Sergio Pino, Director Departamento Planificación y Gestión Estratégica |

En primer lugar se da cuenta del modelo de análisis a utilizar. En segundo lugar se proponen algunas Hipótesis generales para cada eje. En tercer lugar se exponen los Resultados emergidos a partir de las entrevistas. Y en cuarto lugar, junto al resumen general de resultados, se incluye una matriz que sintetiza, por eje de análisis, las Percepciones de los entrevistados y asocia a ellas algunas Recomendaciones que se consideran pertinentes.

Modelo de análisis

De acuerdo a la información entregada por los entrevistados, es posible levantar cuatro ejes de análisis, relacionados entre sí:

- a) **Política nacional de gestión de datos/información** donde se encuentran las percepciones, sugerencias y aportes que las instituciones consideran podrían hacer a esta
- b) **Gestión institucional de datos/información**, que refiere a la descripción de la gestión realizada, las dificultades que enfrentan, protocolos propios, infraestructura, capital humano y financiamiento
- c) **Marco normativo** donde se abordan los marcos normativos asociados a los datos/información que manejan: estándares, propiedad, acceso y regulaciones de acceso
- d) **Redes y vinculaciones** que describe las redes de las que participan y los roles que desempeñan en ellas.

1. Hipótesis

A continuación, se presentan un conjunto de hipótesis generales organizadas en torno a los 4 ejes del análisis cualitativo:

1.1 Sobre una Política nacional de gestión de datos/información

- La Política nacional de gestión de datos/información no constituye un concepto o idea que se encuentre incorporada dentro del imaginario de los investigadores nacionales
 - Como concepto articulado, la comunidad científica no se encuentra muy familiarizada con su existencia y alcances
 - En este primer acercamiento, es posible encontrar nociones distintas sobre que implicaría
- La necesidad de la normativa y oportunidades que una política nacional de gestión de datos/información entregaría al campo científico nacional se percibe aunque no de manera organizada
 - Los investigadores reconocen carencias en cuanto a estándares, regulaciones, normativas, etc. pero no las asocian necesariamente a la ausencia de una política nacional
- El financiamiento de esta política, modalidad y dependencia de este, es la principal preocupación técnica que podría surgir al plantearles la posibilidad de su desarrollo a los investigadores que gestionan datos/información

- El actual mecanismo de financiamiento público de investigación científica implica dificultades para la realización de esta tarea, su continuidad y proyección
- Las instituciones e investigadores que gestionan datos/información deben realizar esta tarea con recursos limitados y, a su entender, insuficientes
- Se espera que la política nacional no requiera utilización de recursos institucionales para su implementación y aplicación
- Que la política nacional de gestión se convierta en una barrera burocrática o una traba administrativa que entorpezca el desarrollo de la investigación científica en vez de facilitarla es una aprehensión fundamental respecto de ésta.

1.2 Sobre el estado actual de la Gestión institucional de datos/información

- No existen protocolos institucionales para la gestión de datos/información que se refieran a la totalidad del proceso
 - El manejo de los datos/información queda a criterio de los investigadores responsables y no existen normas permanentes para su tratamiento
 - Las normas o criterios comunes que es posible observar se refieren a etapas específicas del proceso de gestión
- No se entiende, consensuadamente, a que refiere un protocolo institucional de gestión
 - No existe una definición común en todas las instituciones acerca de qué es un protocolo de gestión de datos/información
 - Existen etapas de gestión que son dejadas fuera del alcance del protocolo, al referirse a él o a la posibilidad de su implementación
- Existen carencias institucionales en capital humano e infraestructura que dificultan una eficiente gestión de los datos/información de manera permanente es precaria
 - La infraestructura tecnológica con que cuentan las instituciones que gestionan datos/información es insuficiente
- A las carencias de capital humano e infraestructura se suman las deficiencias de planificación de la gestión
 - El escaso capital humano es distribuido de manera ineficiente, sobrecargándole de labores o desaprovechando sus capacidades
 - La infraestructura es subutilizada y no se considera el desarrollo o adquisición de herramientas tecnológicas de bajo costo (incluso gratuitas), como *open source* o software libre

1.3 Sobre el Marco Normativo existente en relación a la gestión de datos/información

- El marco normativo está íntimamente ligado al financiamiento recibido para realizar la investigación
 - Las instituciones deben someterse a las disposiciones de la fuente de financiamiento respecto de acceso/disponibilidad de datos/información para instituciones o investigadores externos

- No entrega claridad respecto de las posibilidades y/u obligaciones en cuanto a disponibilidad de datos/información para otros investigadores, ajenos a la institución
 - A pesar de que son las fuentes de financiamiento quienes norman respecto del acceso/disponibilidad de datos/información generados con sus recursos, no entregan claridad respecto de a quién, bajo que modalidad o restricciones se puede compartir
- No establece incentivos al trabajo en red
- No indica ni sugiere modalidades de acceso a datos/información
 - El marco normativo establecido por las fuentes de financiamiento al que las instituciones debiesen adherir no ofrece mecanismos de acceso a datos/información

1.4 Sobre las Redes y Vinculaciones

- Las redes establecidas por las instituciones tienden a tener por objetivo aumentar sus fuentes de financiamiento

2. Resultados

2.1 Política nacional de gestión de datos/información

De acuerdo a las entrevistas realizadas, a partir de las percepciones de una política nacional de gestión de datos/información emergen las siguientes proposiciones para su diseño e implementación:

- ❖ Necesidad de contar con un protocolo general para la gestión de datos/información, que garantice que esta gestión no dependerá de criterios individuales
- ❖ Normativa clara respecto de la propiedad de los datos/información
- ❖ Financiamiento para la implementación de la política nacional de gestión, a nivel de capacitación, capital humano e infraestructura
- ❖ Validación de estándares comunes de calidad de datos/información, que aseguren su calidad y aplicabilidad por parte de quienes accedan a ellos
- ❖ Se sugieren medidas, principalmente, en torno a las regulaciones de acceso a datos/información tendientes a la conformación de una consejería que evalúe a quienes se registren en una página web para obtener estos datos/información
- ❖ La política nacional de gestión de datos/información debe tender a facilitar el trabajo de investigadores

Por política nacional de gestión de datos/información se entiende una política de alcance nacional que establezca procedimientos estandarizados a la gestión de datos/investigación científica, que comprende varias etapas que van desde la Planificación, Obtención, Almacenamiento, Procesamiento, Análisis y eventual Diseminación.

En relación a la elaboración y aplicación de una política de estas características, los entrevistados organizan su discurso en torno a dos dimensiones: **1) la explicitación de los cambios necesarios para que sea posible este tipo de política y 2) sugerencias para su diseño e implementación.**

En cuanto a los **Cambios** que los entrevistados identifican como necesarios para que una política de gestión de datos/información científica nacional resulte positiva y permita potenciar el desarrollo científico del país, los plantean de dos maneras.

Como una crítica a la institucionalidad vigente y la exclusiva administración de recursos que ha hecho en relación a la ciencia, desde su perspectiva, sin dirección política ni planificación estratégica.

“...no sé si conicyt es la estructura adecuada...”

“Yo siento que en todo eso falta una estructura de una política real. Solo se está haciendo una administración de platas y no de políticas”.

Y, por otra parte, identificando los **requerimientos económicos y normativos** necesarios para su realización.

“...la gente tiene nuevas herramientas y éstas tienes que enseñárselas a los astrónomos y aquí es difícil, eso no se hace”.

“Tiene que comprender la cantidad tremenda de discos ponte tu, los recursos computacionales, eso es lo que se necesita y segundo, los recursos humanos, porque esos son clave para hacer más eficiente la parte en que te digan que es lo que tienes que hacer o guiar”.

“... hay centros justamente que tienen una ayuda computacional tremenda, eso falta aquí”.

Tal como muestran las citas, se mencionan como requerimientos: Contemplar financiamiento suficiente para la capacitación de los actores que participan de la gestión de datos/información, aseguramiento de capital humano calificado suficiente para estas tareas e inversión en infraestructura adecuada para las instituciones a cargo de esta gestión.

Además, se requiere Establecer normas definidas y uniformes en cuanto a propiedad y regulaciones de acceso/disponibilidad de datos/información.

“Hay reglas generales, pero si hablamos de accesos a datos debería venir en parte de los fondos que los que otorgan estos financiamientos y en el fondo fijen el rayado de la cancha, porque en general tener información pública también implica costos, o sea, en que lo va a usar, en que servidores y bajo qué condiciones, tecnologías y estándares. Sobre todo si son datos de personas”.

Este requerimiento de claridad respecto de la propiedad de los datos por parte de las fuentes de financiamiento tiene que ver, también, con las críticas a la institucionalidad vigente que causan dudas respecto de las normativas antiguas y su convivencia con la posible nueva política nacional de gestión, como muestra la cita a continuación:

“Se trata de compartir lo que venga de aquí para adelante. Todo proyecto funciona bajo contrato, se tendría q modificar los contratos antiguos y eso no creo que lo vayan a hacer. Ahora, ¿qué pasa con la cantidad de datos que tengo que no están inscritos a ese sistema? No sé si va a haber una voluntad de decir cómo se va a solucionar o si los compartimos”.

Estos requerimientos se encuentran asociados a las sugerencias que los entrevistados plantean para el diseño de la política nacional de gestión. En este sentido, Validar los datos/información a compartir es identificado como una necesidad urgente:

“... una cosa es si queremos tener una base de datos que podamos intercambiar, que podamos estudiar, que produce ángulos que se pueda estudiar en la Universidad de Chile, en Rancagua o en Argentina tenemos que entregar datos buenos, que estén validados, que estén respaldados por un proceso de control de calidad impecable y cualquiera que sean los datos, siempre tiene que hacerse una medición en alguna parte y esa medición tiene que hacerse”

Además, se propone que los protocolos de gestión y/o validación tengan carácter interinstitucional, que aseguren que la gestión de datos/información, incluida la calidad de los datos/información gestionados, no dependa del criterio individual del personal a cargo de esta labor.

“es cuestionable si es que se va a seguir con el mismo mecanismo o no, y la única manera de garantizar eso, es con un protocolo escrito y de alguna manera validado por ese tipo de normas ISO 900 alguna variación externa que te obligue a seguirlo”

“si se cambiara de unidad la persona a cargo, o no sé las 4 o 5 personas claves de la encuesta a lo mejor no se podría seguir con el mismo protocolo, entonces la única manera de asegurar que esto trascienda a las personas es que primero en la institución este escrito y entre instituciones este validado por algún organismo externo, que puede ser nacional, internacional, da lo mismo pero que en el fondo sea parte de los términos de referencia para cualquier manejo de base de datos”

Se sugiere, también, que los sistemas utilizados por cada institución estén vinculados entre sí para facilitar el trabajo de los investigadores en las distintas etapas de gestión permitiéndoles compartir sus datos/información a través de una Plataforma web que regule el libre acceso a ellos.

“tener vinculación con nuestros sistemas propios, que existieran comunicaciones...”

“Yo creo que hay que trabajar de una forma de consejería, en donde una web que puede acceder a todos los datos, y como alguna manera fue entrevistado, (...) con el (...) también tiene una ficha, son mas quisquillosos en cuáles son las razones para obtener los datos o no, pero debería ser algo así, pero debería ser así, para todas las encuestas una ficha que alguien valide si te da acceso a los datos o no, una clave”

Esta sugerencia da cuenta de la disposición a compartir y colaborar de los entrevistados, disposición que se ve frustrada tanto por la ambigüedad de la normativa referida a propiedad de datos/información, como a la ausencia de un protocolo de gestión de datos y validación de éstos que permita la estandarización de su calidad y la interrelación entre instituciones.

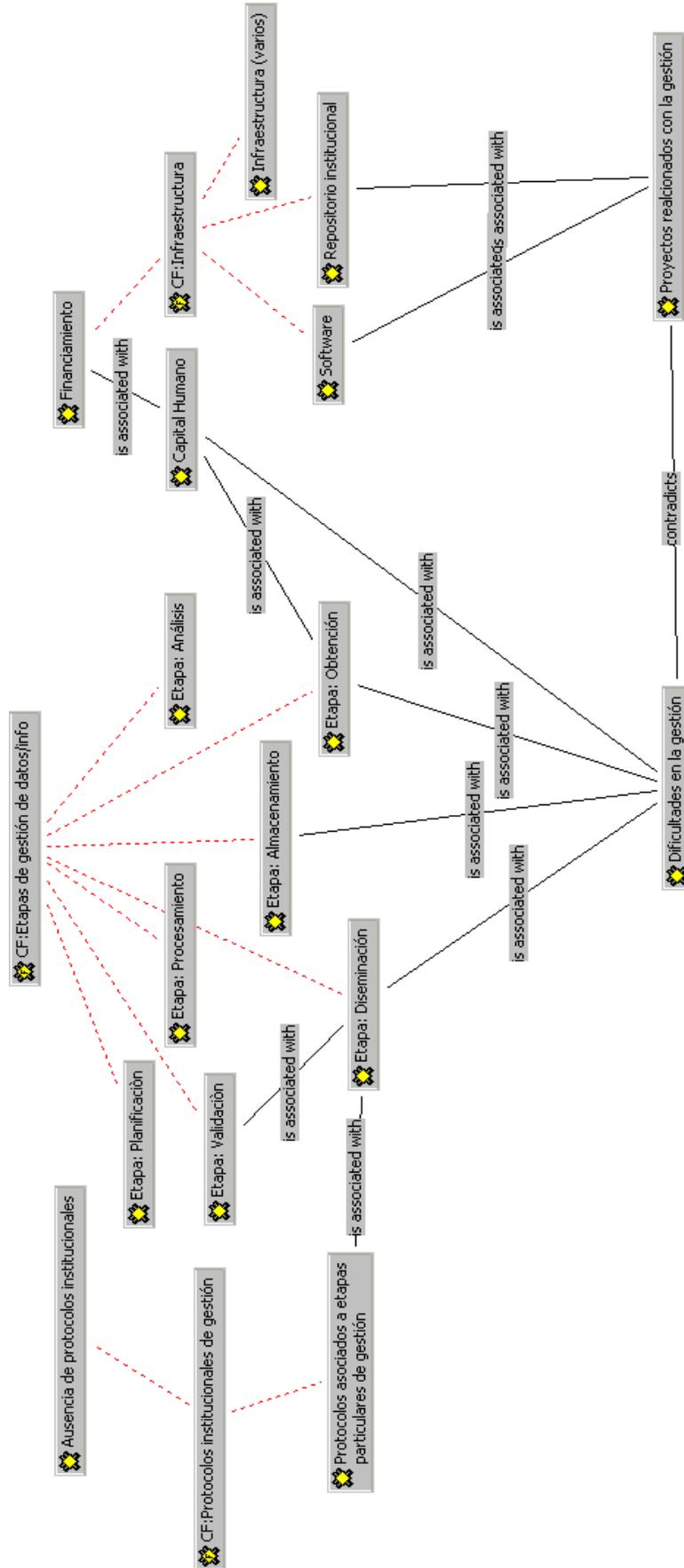
2.2 Gestión institucional de datos/información

Sobre la gestión institucional de datos/información, a partir de la información entregada por los entrevistados, se observa que:

- ❖ La existencia de protocolos de gestión de datos/información a nivel institucional no es una generalidad y suelen referirse a etapas puntuales de esta gestión, no al proceso completo
- ❖ No existe, entre estas instituciones, claridad sobre cuáles son las etapas de gestión de las cuales participa
- ❖ De lo anterior se desprende que se identifican dificultades de gestión en etapas puntuales, principalmente a la etapa de Obtención y Diseminación por la ausencia de claridad respecto de la propiedad de los datos/información
- ❖ Se intenta dar solución a las dificultades en la gestión a partir de mejoras a la infraestructura computacional

Tal como se expuso en el apartado 2.1 referido a Política nacional de gestión de datos/información científica, esta gestión comprende varias etapas que van desde la Planificación, Obtención, Almacenamiento, Procesamiento, Análisis hasta su eventual Diseminación. Y corresponde a un conjunto de directrices, estándares, criterios y procedimientos que asegure una gestión eficiente del patrimonio, un resguardo adecuado y un aprovechamiento apropiado de los recursos.

GESTIÓN INSTITUCIONAL DE DATOS/INFORMACIÓN CIENTÍFICA



La descripción del proceso de gestión institucional de datos/información científica que realizaron los entrevistados puede leerse en torno a 4 ejes: los Protocolos institucionales de gestión, las Etapas de gestión, las Dificultades asociadas a la gestión de datos/información y los recursos.

En cuanto a los protocolos institucionales, se desprende de las entrevistas que o no existen para nada o están asociados a etapas puntuales de la gestión:

“En general dentro de esta, lo que ocurre es como hay autonomía dentro de las facultades y a nivel de lo académico, no existe ningún tipo de política”

“No, en este minuto yo diría que no existe, en este minuto lo que hace cada investigador cuando publica algo, cuando produce alguna información la maneja el, el saca sus proyectos y todos sus informes son un servicio solicita un servicio si va a publicación la escribe y la presenta a un congreso, etc. Lamentablemente no tenemos una política de centralización”

“...hay distintos protocolos, el cómo más importante es el de confidencialidad de los datos”

“Todas las bases las maneja una unidad de encuesta que esta acá en el piso 7 y a los investigadores y a las contrapartes se les entre en nominación y control bien estricto en respecto a la confidencialidad de los datos”

Se tiende a identificar como el “*más importante*” aquel protocolo de confidencialidad que resguarda los datos personales, sin considerar aparentemente que un protocolo también puede considerar el intercambio de datos/información bajo ciertas normas.

En cuanto a las etapas de gestión, es posible observar que no existe mucha claridad respecto de la diferencias entre una y otra, como cuando se afirma por ejemplo:

“Todos, es que estamos en mayor o menor grado metidos”

También es posible notar que las etapas de gestión que parecen estar más asociadas a dificultades en la gestión son las de Obtención y Disseminación, precisamente por la ausencia de protocolos y criterios de acceso a datos/información entre instituciones y entre los propios miembros de la institución.

“Es difícil obtener información de otras universidades”

“A mí me encantaría que nosotros no recibiéramos una cuestión un poquito más general que estas cosas, que los informes finales que se están entregando en Fondecyt en las publicaciones que ellos hacen no recibimos la información”

“Cada vez que uno va partiendo con un proyecto lo está haciendo de cero, como si en Chile no se hubiese hecho nada porque desconoce que está absolutamente imposibilitado de acceder”

“...todo lo que tiene que ver con investigación te diría yo que es información propia de los investigadores y ellos ningún mecanismo y yo creo que tampoco está dentro de nuestras facultades para obtener estos datos, es lo que ellos hacen”

En cuanto a los Recursos, que incluyen capital humano e infraestructura (Software, repositorios institucionales, varios), son parte importante de las dificultades para la gestión:

“Bueno, siempre los recursos es un tema, hemos buscado en el fondo las asociaciones que hemos hecho para tener con financiamiento”

Esto porque repercuten directamente sobre la infraestructura que las instituciones son capaces de adquirir o desarrollar, además de la cantidad de capital humano que logran mantener destinado a las distintas labores de gestión.

“En esta vicerrectoría nosotros hemos desarrollado software ad hoc a todas las necesidades que tenemos. Tenemos todo el sistema por ejemplo de evaluación de investigadores en línea que fue desarrollado acá, también tenemos todo un sistema de rendición en línea de proyectos internos que también tiene CONYCIT que también está hecho aquí.

También tenemos un software que tiene que ver con todo lo que es documentación, otro para la gestión de pagos”

“...es difícil mantener la comisión, porque nosotros mismos nos vamos metiendo a las páginas web de ahí vamos entrando a las publicaciones que va generando la universidad, y eso es un mecanismo...”

“...no es que la te diría yo es una asistente administrativa, que tiene otras funciones porque es secretaria de una autoridad del ... investigaciones , en este momento es secretario del vicerrector de investigación del trabajo , vicerrector adjunto, pero lo que hace esta persona a su vez, es lo que hace el día a día en este premio”

Y, finalmente, también en relación con la infraestructura y las dificultades asociadas a la gestión, aparecen proyectos orientados fundamentalmente a la mejora de infraestructura computacional. Curiosamente, no considerando el factor capital humano como un aspecto a reforzar.

“...nosotros tenemos un proyecto que no está implementado, o sea que está pensado en el área de biblioteca que tiene que ver con un repositorio de la universidad, dentro de la universidad, que está pensado en que se considere como un elemento importante, o sea como poder alojar ahí todas las publicaciones que generan los académicos de la universidad”

2.3 Marco Normativo

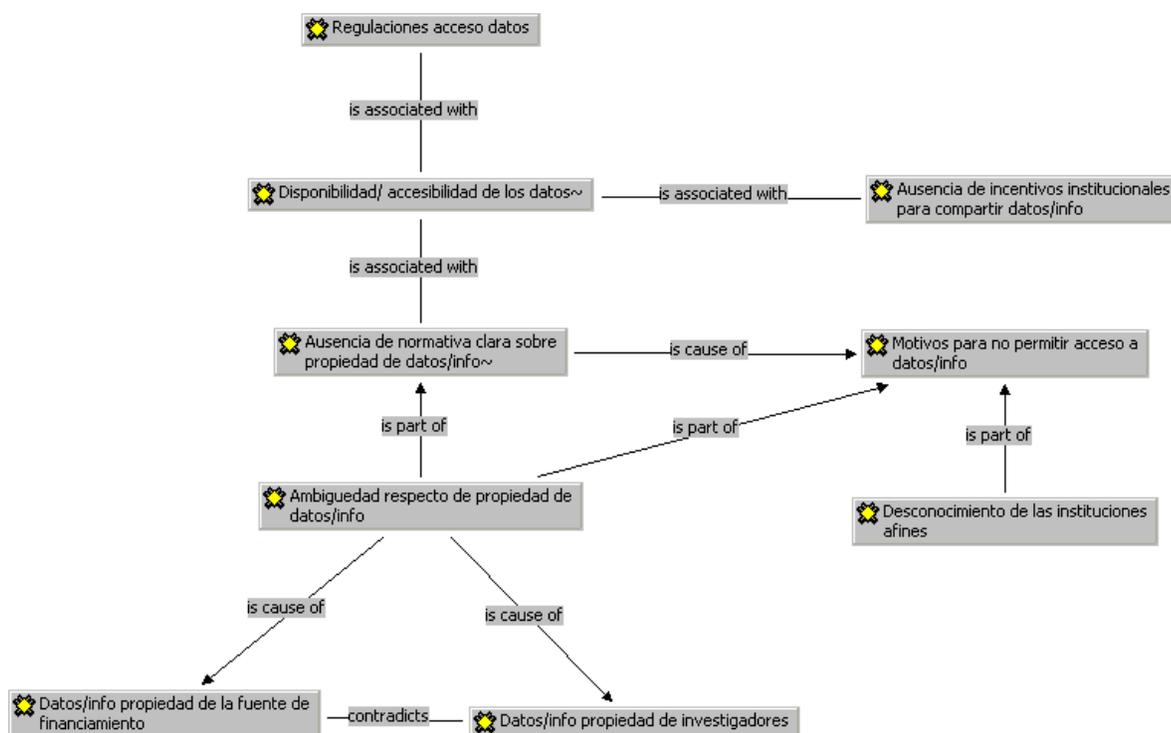
En relación al Marco Normativo que existe sobre la disponibilidad/acceso a datos de investigación y/o información científica, los entrevistados entregan las siguientes consideraciones:

- ❖ La regulaciones de acceso a datos/información que repercuten sobre la disponibilidad y acceso a estos por parte de otras instituciones y personas tiende, generalmente, a entregar datos/información previa solicitud y evaluación de los objetivos y los solicitantes, quedando a criterio de cada institución en ausencia de normativas claras al respecto por parte de la fuente de financiamiento

- ❖ La ambigüedad respecto de la propiedad de datos/información conduce a que algunas instituciones no entreguen acceso a su patrimonio por considerarlo propiedad del mandante, mientras que otros lo consideran propio y si entregan acceso a él
- ❖ La disponibilidad de datos/información, en el caso de las instituciones que permiten el acceso a estos esta, en algunos casos, disponible a través de una plataforma electrónica que valida a quienes acceden a estos

El Marco normativo refiere a disposiciones explícitas o implícitas que regulen el acceso a los datos de investigación y/o la información científica que generan los proyectos.

MARCO NORMATIVO DE ACCESO A DATOS/INFORMACIÓN CIENTÍFICA



Las regulaciones del acceso a datos están relacionadas con la Disponibilidad/accesibilidad que se entregue a ellos y esta disponibilidad, a la vez, está marcada por la ambigüedad respecto de la propiedad de los datos/información manejados.

De acuerdo a esto, en cuanto a las regulaciones concretas de acceso a datos, no existe una única norma que se aplique a todas las instituciones entrevistadas, ni siquiera se utilizan los mismos criterios de disponibilidad de datos/información. Tal como se ve en las siguientes citas, existe controversia y ambigüedad respecto de cómo regular.

“... hay una controversia metodológica, que como te decía al comienzo nosotros somos muy guardianas de la confidencialidad de las bases de datos y aquí la muestra la seleccionaba el INE, la aplica la institución ejecutora, y MIDEPLAN quería tener los dato y los nombres de las personas a las que se encuesta y nosotros creemos que eso no va con nuestra metodología y yo creo que es obligación”

“Tienes que hacer una solicitud específicamente para poder acceder a ello”

“Nosotros evaluamos si podemos o no entregar esa información”

“...tendrían que llenar solicitud y nosotros evaluar la situación y si es competente o no entregar esa información”

Se entiende, a partir de estas citas, que en los casos en que el patrimonio de datos/información se encuentra a disposición de la comunidad, son los investigadores a cargo de este (ni siquiera se mencionan protocolos institucionales establecidos para discernir quién accede o no al patrimonio mencionado) quienes deciden, de acuerdo a sus propios criterios el acceso a los datos/información generados con financiamiento público.

Las únicas disposiciones que se interpretan claramente son las referidas a la confidencialidad de datos, tal como se observa a continuación:

“Está en la hoja de las encuestas está escrito en una ley, si porque en el fondo tu entregas información de ingresos, de deudas y conformación del hogar y que tu sabes que la declaración de ingresos, deudas o conflictos de interés que tú puedas tener, con nombre y apellido... o respondes otra cosa, entonces se pierde la confiabilidad de datos que se entregan, por eso que está en un tema extremadamente sensible... mantener este secreto”

Para todo el resto de los datos/información parece existir un vacío normativo que genera dudas e incertidumbre en las instituciones que gestionan y no les permite asegurar la disponibilidad de su patrimonio para otras personas y/o instituciones:

“... una vez que se termina el estudio depende de quién hay que pedirle la autorización de la contraparte o ver si están disponibles los datos”

De acuerdo a las interpretaciones institucionales o personales respecto de este vacío/ambigüedad normativa, cuando los datos/información están disponibles se observan dos modalidades de acceso: 1) a pedido y 2) para los miembros de la red.

La primera modalidad se encuentra explicitada en las entrevistas en citas como:

“Hay distintos, depende de la destitución demandante, por ejemplo las tasas de empleo están disponibles, peor mandando un correo, hay un correo de contacto en la página si quiere solicitar los datos llenas una ficha y una persona te va a mandar un correo con los datos. La encuesta de protección social, ahí está más estructurada, pero también tiene una ficha tú dices porque quieres los datos, que se yo y hay una revisión valida, pero en el fondo es para que quieres usar los datos y crear una clave de acceso y se descargan los datos del servidor”

“... que si yo te lo mando por correo, ese tipo de cosas...”

“Esa información cuando se tiene se comparte, o sea a pedido...”

“... tampoco hay una política interna en ese sentido, nosotros desde acá en general siempre nos piden información de las publicaciones de una unidad, nos piden información de, la, la, la cuando la tenemos la entregamos, por número, por unidad, esa, esa información la entregamos”

La segunda modalidad, tiene que ver con que los datos/información están a disposición de las instituciones y/o investigadores que participan de la red o comunidad científica es indicada por los entrevistados en las siguientes citas:

“Esa planilla que tenemos nosotros la pasamos a quien quiera seguir con eso dentro de ese grupo, si quiere comparar sus datos con los nuestros”

“Eso lo hicimos, lo hicimos en el IAEA en, no me acuerdo en dos o tres proyectos lo hicimos, no me acuerdo bien, porque fue hace un tiempito, en un inter regional precisamente y ¡ah! Esa era otra cosa era de materiales de referencia para ver químicos y ensayos de aptitud ... Entonces ahí en esa plataforma poníamos todo lo que hacíamos y se produce un material de referencia, lo poníamos ahí, escribíamos sus características que estaban certificadas como habíamos hecho este plan, esa era la forma”

“... esa información está disponible para los países participantes”

“... a nivel en que se termina un proyecto, produce un documento final que está disponible a todos los miembros del OIEA”

“... lo que se hace es que los datos mismos los obtienen los expertos de los observatorio s y ellos te mandan a ti estos”

“... típicamente en todos los observatorios del mundo esos datos son accesibles para el resto de la comunidad en un periodo de tiempo de un año”

“Es por ciencia no más, la ciencia es universal; por eso como te digo, no solo en Chile tenemos acceso a los telescopios (...) y tengo que colaborar en todas partes del mundo, entonces cuando tu trabajas y te mandan los datos, bueno en ese caso no es en archivos Fitz, que tu puedes estar procesando en cuanto a paquetes en (...) y al final tu escribes las cosas en Fitz como para poder enviárselas a otro como distintos requisitos de contribución a la ciencia, pero sí, todo es a utilizar y no hay cosas pecuniarias, todos nos cambiamos los datos, en aras de entender mejor la ciencia y cada uno tiene un área de experiencia diferente, entonces, juntamos gente con diferentes habilidades y yo creo q así se logra una mejor apreciación de los datos”

Es posible, además, dar cuenta de una asimetría en cuanto a la utilización de tecnologías para compartir datos/información entre las instituciones entrevistadas. Pues algunos utilizan servidores y formatos validados entre instituciones, mientras otros simplemente entregan los datos/información de acuerdo a alguna de las dos modalidades anteriormente descritas.

La ambigüedad existente sobre la propiedad de datos/información tiene que ver con la idea de propiedad sobre los datos/información que tienen los investigadores a cargo de la gestión.

Algunos investigadores consideran tener propiedad sobre estos datos/información:

“... todo lo que tiene que ver con investigación te diría yo que es información propia de los investigadores y ellos ningún mecanismo y yo creo que tampoco está dentro de nuestras facultades para obtener estos datos, es lo que ellos hacen”

“... estos datos que tú obtienes son para ti”

Mientras que otros los consideran patrimonio de la fuente de financiamiento. Así se observa en los siguientes extractos de entrevistas realizadas:

“... en ese sentido que alguien pueda decir que esto es público por que recibió en una parte financiamiento público, para ellos en el fondo no es tan así, porque parten muchos proyectos con financiamientos compartidos. Entonces el dato que genera o la batería de datos que genera finalmente son públicos – privado”

“Porque quien le pide el informe final es Conicyt y se lo entrega a Conicyt, entonces quien le da la plata, entonces te fijas”

Así esta ambigüedad y ausencia de normativa se convierte en la principal causa de no permitir acceso a datos/información. Aunque, además, existe desconocimiento sobre instituciones afines con las cuales compartir:

“... cuando aparece un proyecto nuevo ahí sí que nos vamos un poco de a cabezazo porque un tema nuevo, porque no sabemos a quién enviarle”

Que se suma a la ausencia de incentivos institucionales para esta potenciar esta dinámica entre investigadores.

Consultado sobre esta materia, uno de los entrevistados es claro al decir que en la institución existe:

“Absolutamente nada”

En cuanto a incentivos para poner a disposición o entregar acceso al patrimonio de datos/información.

Finalmente, se teme estar violando algún código ético al entregar datos/información:

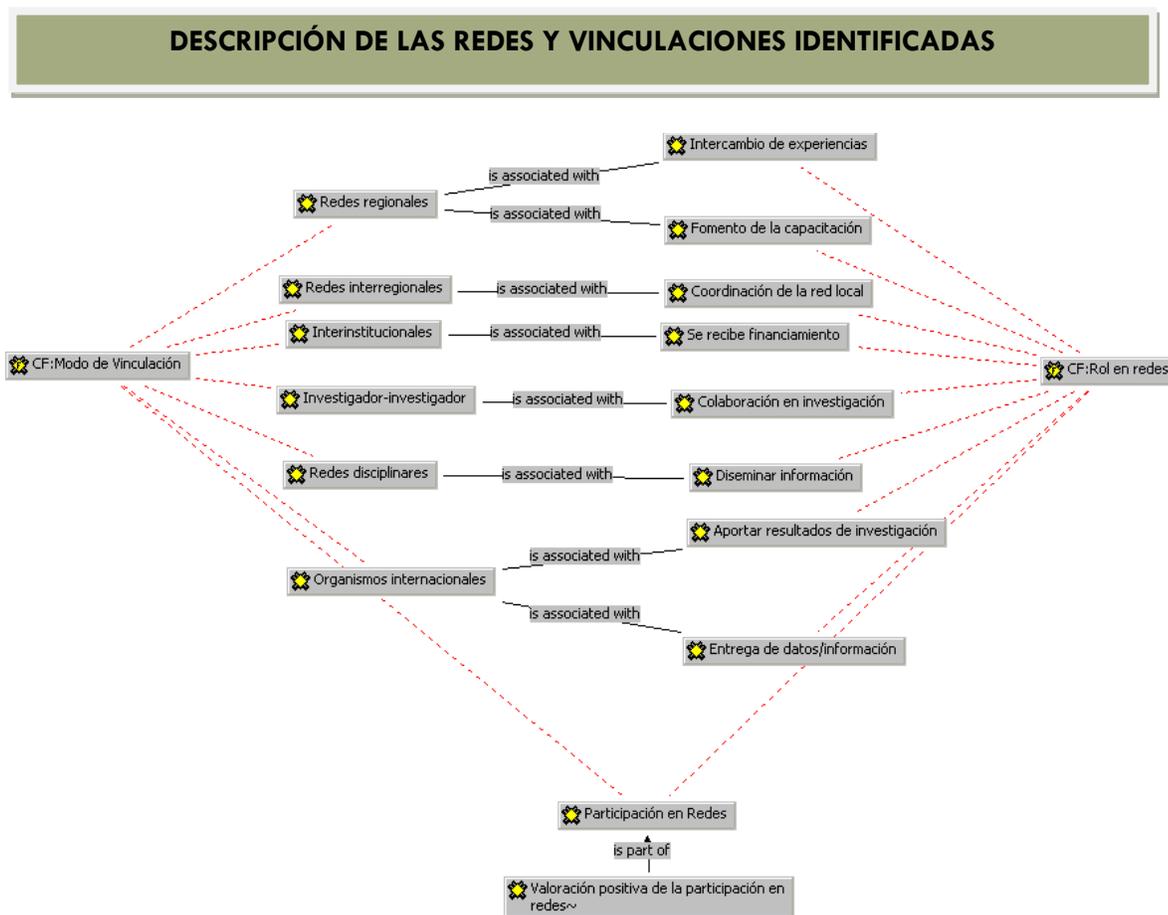
“Dicen que por razones fundamentales, que la transparencia, que preguntarles a las personas si pueden entregar la información, que si se las entregamos a ustedes que pueden decir los otros...”

2.4 Redes y vinculaciones

La descripción de las redes y vinculaciones identificadas por los entrevistados, muestra que:

- ❖ Se identifican 6 tipos de redes en las cuales las instituciones/investigadores entrevistados se encuentran participando: Redes Regionales, Interregionales, Interinstitucionales, Investigador-Investigador, Disciplinarias y de Organismos Internacionales.
- ❖ Se identifican 8 roles posibles de desempeñar: Intercambio de experiencias, Fomento de la capacitación, Coordinación de la red local, Recepción de financiamiento, Colaboración en investigación, Disseminación de información, Aporte de resultados de investigación y Entrega de datos/información.
- ❖ La participación en redes es valorada y evaluada positivamente por los entrevistados

En este apartado se describen las redes y vinculaciones que indican tener los investigadores entrevistados, así como el rol que desempeñan en ellas.



Tal como se observa en el esquema, los entrevistados entregaron información sobre 6 tipos de Redes.

Las redes regionales que son integradas por instituciones que investigan sobre tópicos similares en Latino América y el Caribe. Y los principales roles identificados dentro de estas son: el intercambio de experiencias y el fomento de la capacitación vía postulación a fondos y diseño de cursos.

Las redes Interregionales son el vínculo de las anteriores con Redes de otras regiones, en este tipo de redes los investigadores entrevistados indicaron que sus labores son fundamentalmente de coordinación y supervisión del funcionamiento de la red local.

Las redes Interinstitucionales, son aquellas que se establecen directamente entre una y otra institución, estas redes son también internacionales y fundamentalmente implican recibir financiamiento para la investigación. En relación con estas se encuentran las Redes Investigador-Investigador, que no tienen carácter institucional ni de financiamiento y donde el rol es específicamente colaborar con la investigación del otro.

Las redes disciplinares, en tanto, refieren a grupos de investigadores e instituciones que se agrupan de acuerdo a temas comunes como indica uno de los entrevistados:

“...en varios en, medicina nuclear, radioterapia, diagnostico, tenemos otro en agricultura...”

Y el rol en ellas tiene que ver con diseminar la información y los resultados obtenidos al resto de la comunidad científica del área.

“... nosotros lo que hacemos es ponemos la información en una página web o se la enviamos a quienes ya sabemos que están trabajando en eso”

Y, finalmente, se identificó la participación en una Red de Organismo Internacional, que agrupa todas las instituciones que trabajan sobre la materia en cuestión alrededor del mundo. En el caso mencionado se hace alusión a:

“...gran organización mundial que es el Organismo Internacional de Energía Atómica, las OIEA”

Y el rol desempeñado dentro de esta red tiene que ver con entregar datos/información para ser utilizada en la investigación por los demás miembros y en aportar resultados de investigación de alta calidad, con el fin de aumentar y mejorar el conocimiento en el área.

3. Conclusiones

El análisis de las entrevistas realizadas arroja algunas conclusiones importantes de relevar:

PRIMERO: Se da cuenta de un estado de la gestión de datos/información deficiente en algunos aspectos y bastante perfectible en otros. En general, la ausencia de protocolos estandarizados para la gestión de los datos/información perjudica una labor de excelencia al impedir el manejo de criterios de excelencia comunes, conocidos y respetados, por (al menos) todas las instituciones que trabajan con el mismo tipo de datos/información.

SEGUNDO: El actual marco normativo que debería indicar la disponibilidad de los datos/información aparece, en algunos casos, como desconocido o ambiguo, incitando a la aplicación de criterios arbitrarios

por parte de instituciones e investigadores en relación a la entrega de acceso al patrimonio de datos/información obtenido con financiamiento público.

TERCERO: La participación en redes parece encontrarse sometida a la riqueza de los vínculos de investigadores o al desarrollo en investigación alcanzado por áreas del conocimiento en particular. No se percibe una dirección estatal que organice estratégicamente la potenciación de ciertas redes y vinculaciones.

CUARTO: A pesar de no estar familiarizados con una idea común respecto de que implica e involucra una política nacional de gestión de datos/información, los investigadores entrevistados concuerdan en la necesidad urgente de regular la gestión existente y de trazar lineamientos generales que organicen con fines claros la asignación, distribución y administración de los recursos destinados por el Estado a la investigación científica del país.

V. APRENDIZAJES INTERNACIONALES Y BUENAS PRÁCTICAS

En esta sección, se presenta una síntesis de los aprendizajes internacionales detectados, entendidos ahora como buenas prácticas, de los que se extraen aprendizajes estratégicos para Chile.

SOBRE LA POLITICA DE GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA.

1º.- Las experiencias internacionales indican que el primer paso de las políticas de gestión de datos e información es identificar los problemas que existen para cada área científica en esta temática pero luego, y muy importante, el dar a conocer e institucionalizar la problemática en todos los actores involucrados. Ello, dado el contexto nacional, constituye un aspecto clave en el siguiente entendido: deberá definirse en Chile las problemáticas en torno a las cuales priorizar el desarrollo de la política. Así, la definición de Sectores con Potencial de Crecimiento y/o los Ejes de Vocación Productiva Regional pueden ser un eje articulador. Así también, la definición de ciencia de avanzada, en torno a centros y redes de investigadores en temáticas de alta productividad, pueden ser una vía relevante también.

2º.- Creación de organismos que resguarden los datos de algún área científica determinada. Por lo general se alimentan de los resultados emanados de proyectos, investigaciones generadas con fondos públicos y puestas a disposición de la comunidad científica. Esta experiencia internacional, en el contexto chileno, implicará definir el rol de las agencias públicas y sobre todo los institutos tecnológicos y/o la creación de agencias de Gestión de Datos de Investigación e Información Científica. Ello, dependerá del marco institucional que regula las instituciones y del modelo de interacciones de governance y red de política pública que se diseñe.

3º.- Los fondos que permiten el apoyo, o la total realización de proyectos científicos, incluyen dentro de sus aspectos normativos no sólo lo relacionado con los objetivos globales del proyecto sino además lo concerniente al trabajo y entrega de datos en específico. Ello, en el caso chileno, requiere ser desarrollado. El examen de la normativa de los fondos y las propuestas de modificación del marco normativo, serán fundamentales.

Lo anterior, implica al menos 3 aprendizajes relevantes para el caso chileno:

i.- Necesidad de hacer patente el carácter público de los datos de investigación e información científica financiados con fondos estatales y su importancia para el desarrollo y la competitividad: ninguna política de gestión de datos de investigación e información científica, ya sea en una institución o en el país, puede ser desarrollada de manera eficiente y enriquecedora si es que la idea de privacidad exclusiva de ésta persiste. Existen matices en los procesos de patentamientos, información de trato especial y propiedad intelectual, pero la información financiada con fondos estatales se transforma en un bien público. En el caso de las Políticas de Intercambio de Datos Científicos en China evidenció la dificultad de desarrollar una política nacional de gestión cuando en la cultura científica local persistió la idea de propiedad personal de los investigadores de los datos de investigación e información científica financiados con fondos públicos. Para esto se realizó una campaña que puso en valor la necesidad de desarrollar políticas de gestión de datos como soporte para el desarrollo del país y su competitividad.

Atendiendo el caso de la política China, se hace necesario instalar la discusión respecto a cómo entender lo público y cómo los distintos fondos, además de ayudar a promover el desarrollo científico, están pensados y orientados al desarrollo de los países, lo que implica no sólo cumplir con los objetivos planteados en la fase de postulación de los proyectos de investigación y estudios aplicados, sino además, incorporar en el acervo de productos comprometidos la parte intermedia relacionada con los procesos como lo son las bases de datos y el levantamiento de información, ya que éstos son vitales para el avance de iniciativas posteriores.

ii.- Necesidad de resguardo de datos de investigación e información científica de trato especial: existe cierta información que si bien no puede ser diseminada de forma abierta a todas las instituciones por el mal uso que puede hacerse de ella, es necesario construir mecanismos que permitan darle algún tipo de diseminación que sea beneficioso y útil. Un caso donde esta situación ocurre son con los datos epidemiológicos del National Center for Health Statistics en Estados Unidos. Debido a lo delicado de la información manipulada, se construyó un “seguro de investigación” llamado RDC, en éste el almacenamiento responde a determinados criterios y se permite el ingreso bajo una estricta supervisión. La única forma de acceder al RDC es mediante una propuesta de los investigadores que permita aceptarlos y hacer uso de este material. Ejemplos de restricción del RDC son la configuración especial de los puertos USB y todo tipo de mecanismos de transferencia de datos que lo hacen inaccesible. Las estaciones de trabajo se han configurado de tal manera que los usuarios reciben acceso de sólo la lectura a los datos y archivos solicitados y pueden escribir sólo en la estación de trabajo local del disco duro. Estas restricciones garantizan que los usuarios no pueden eliminar información que no ha sido sometido a una revisión a la confidencialidad.

En este sentido, la información de difícil gestión, con características que la hacen diferentes y merecedoras de protocolos especiales, deben ser identificadas en las áreas científicas, establecer los motivos de resguardo y generar el tipo de restricciones y condiciones de acceso para su uso.

iii.- Necesidad de promover la transdisciplina en la gestión de datos de investigación e información científica: la gestión de datos de investigación e información científica no posee una disciplina o campo científico por excelencia de su propiedad, en ella interceden gestores de información, ingenieros informáticos, lingüistas, bibliotecólogos, entre otros. En consecuencia, para establecer políticas nacionales de gestión es necesario disponer además de este tipo de profesionales, investigadores y académicos que entiendan y conozcan en detalle la estructura del campo científico que se trabajará en una política; esto debido a que cierta información merece una decodificación distinta, tiene una composición más cualitativa la que podría requerir tanto de habilidades informáticas y técnicas para su procesamiento en grandes bases de datos que interactúen, pero además, un conocimiento acabado de cómo esta información es procesada, relacionada y tipificada para que en el proceso de su almacenamiento y diseminación tenga una utilidad en los investigadores que la trabajarán.

SOBRE EL CAPITAL HUMANO DEDICADO A LA DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA.

1º.- A nivel internacional, se ha puesto énfasis en la generación y desarrollo de carreras universitarias ligadas a la administración de información como bibliotecología. A nivel nacional,

este campo deberá ser analizado desde el punto de vista de las competencias a desarrollar en esos campos disciplinarios, pero también en el diseño de incentivos a la formación mediante postítulos y/o postgrados, utilizando los recursos disponibles en el sistema de formación de capital humano del país.

2º.- Incentivos a investigadores que aborden proyectos, sistemas, plataformas o medios de gestión de datos de investigación y/o información científica. A nivel internacional, este aspecto relevante implica que en Chile, deberán generarse incentivos vía la definición de requisitos en los marcos normativos que regulan la financiación pública y mediante la conformación de redes y/o consorcios científico-tecnológicos. El rol de las Universidades y/o centros de investigación será también fundamental.

Lo anterior implica así dos desafíos estratégicos:

i.- Necesidad de introducir las áreas científicas en su contexto internacional en los planes de mejoramiento de gestión de datos de investigación e información científica: El caso de las políticas de intercambio de datos científicos en China demuestra lo perjudicial que es para un avance en las políticas nacionales de gestión de datos, no disponer de grupos de investigación de clase mundial que permitan conectar el estado de la gestión de datos de un área científica con la discusión y características que tiene ésta en el contexto internacional. Por ejemplo, el desarrollar un plan que mejore la gestión de datos vulcanológicos en Chile, debe coordinarse con otras latitudes, manteniendo estándares internacionales, posibilidad de interactuar con repositorios de otros países lo que permite beneficiarse de otras iniciativas, generar convenios de cooperación y proyectos transnacionales.

ii.- Necesidad de generar planes de atracción, formación y retención de capital humano pertinente a la gestión de datos de investigación e información científica: Como se señalaba anteriormente, un desarrollo de capacidades en la gestión de datos de investigación e información científica, contempla aspectos tecnológicos, institucionales y normativos, pero también, y de forma transversal, de capital humano. Existen profesiones y habilidades que tienen mayor relación con este tipo de labor lo que debe ser intensificado en planes de pre y post grado, así como también el traer expertos internacionales en la materia para ayudar a desarrollar y fortalecer una política nacional de gestión.

SOBRE LA INFRAESTRUCTURA TECNOLÓGICA REQUERIDA PARA LA GESTIÓN DE DATOS DE INVESTIGACIÓN E INFORMACIÓN CIENTÍFICA

1º.- El avance y mayor desarrollo de Internet es la plataforma que permite a su vez el desarrollo de sistemas que posibilita gestionar datos e información a través de este canal cada vez con mayor especialización y velocidad.

2º.- Comunicación e interacción entre distintos sistemas de información. El desarrollo de tecnologías para la interoperabilidad, es condición fundamental para el desarrollo de capacidades.

3º.- Progresivo traspaso de estructuras físicas para el almacenamiento de datos e información a estructuras digitales (bases de datos online). La distinción entre almacenamiento on-line y off-line por los usuarios, es clave a la hora de discriminar el desarrollo de capacidades de hardware y software para todos los eslabones de la cadena.

4º.- Posibilidad de edición de la información antes de su aparición oficial y reenvío a diversos actores e instituciones relacionadas a la temática. El desarrollo de tecnologías de captura de datos e información científica y tecnológica será un tema relevante, así como el procesamiento del mismo.

5º.- Especialización de los buscadores debido a la gran proliferación de información. Los investigadores, en la medida que progresan en el uso de TICs, especializan las búsquedas, requiriendo buscadores especializados, lo que incidirá favorablemente en el desarrollo de software en el país.

Estos 5 ejes, generan desafíos, cuales son:

i.- Necesidad de identificar requerimientos tecnológicos diferenciados por área científica-tecnológica: La complejidad del conocimiento se ha traducido en una diferenciación en distintas áreas y disciplinas de acuerdo al nivel de especificidad tanto en su propio operar como en el equipamiento e infraestructura que acompaña la labor científica. De la misma forma que los médicos, físicos y antropólogos deben apoyarse en distintos elementos, software y equipamientos para desarrollar su labor, la gestión de datos e información también tiene un nivel de especialización que hace necesario el poder detectar tipos de software, almacenes de datos y repositorios con ciertas particularidades para cada área científico-tecnológica. En este sentido, al construir una política nacional de gestión de datos de investigación e información científica, se requiere identificar el estado actual de la disposición de infraestructura y equipamiento además de detectar brechas para adquirir gradualmente los elementos faltantes.

ii.- Necesidad de generar repositorios en web por área científica: En casos como la oceanografía en España y Estados Unidos, la institucionalidad responsable de coordinar, a nivel macro los datos levantados con fondos públicos, son centros que se encargan de los procesos de validación, procesamiento, almacenamiento y disseminación en repositorios comunes que permiten una claridad a la comunidad científica de disponer de portales de acceso y a la vez de trabajo para estos mismos datos, dejando espacios mediante foros, wikis y servicios 2.0 para la participación de investigadores en red.

iii.- Considerando las implicancias para política científica en materia de **requerimientos de soporte científico-tecnológico para el desarrollo de los sectores competitivos**, y el **requerimiento de desarrollo de plataformas de Ciencia, Tecnología e Innovación y Programas Nacionales de Ciencia, tecnología e Innovación**, se hace necesario el *Mejoramiento y mayor desarrollo de Internet* como plataforma que permite a su vez el desarrollo de sistemas que posibilitan gestionar datos e información a través de este canal cada vez con mayor especialización y velocidad.

iv.- A la vez a partir de las transformaciones del diseño institucional, y el **rediseño de la arquitectura institucional del Sistema Nacional de Innovación** se requiere el *Mejoramiento de la comunicación e interacción entre distintos sistemas de información*.

v.- De acuerdo a los cambios en la **creación de nuevos tejidos institucionales** en definitiva deben apuntar al *Traspaso progresivo de estructuras físicas para el almacenamiento de datos e información a estructuras digitales* (bases de datos online), y a la (*Generación de un sistema de edición de la información* antes de su aparición oficial y reenvío a diversos actores e instituciones relacionadas a las temáticas específicas, referido especialmente al almacenamiento, procesamiento, validación y diseminación de datos de investigación e información científica).

vi.- Se hace a la vez necesario desarrollar una **Especialización de los buscadores** debido a la gran proliferación de información, que permita una eficiente captura y diseminación de datos de investigación e investigación científica.

SOBRE EL PATRIMONIO EN GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA

1º.- -El patrimonio se homologa. La estructura de los datos de un área científica se hace común para poder almacenarlas con los mismos criterios y permitir ser diseminada con utilidad al resto de la comunidad científica.

2º.- El patrimonio se distribuye en sistemas madres. Grandes plataformas que funcionan en red y operadas por una unidad especial, además de dedicarse a determinadas áreas científicas con proximidad temática, almacenan y diseminan datos e información. En experiencias más sofisticadas incluso se pueden realizar análisis y simulaciones online al poseer una estructura de datos similar.

Estos dos ejes, generan dos desafíos:

i.- Necesidad de desarrollar metodologías que incorporen datos y metadatos no accesibles con las tecnologías presentes: En el caso de la oceanografía española se identificó la necesidad de rescatar datos y metadatos no accesible actualmente para lo que se debió desarrollar metodologías y mecanismos de acceso a ellos para su posterior estandarización. En este sentido, los estándares internacionales deben ser incorporados, pero a la vez resguardando las particularidades nacionales en las prácticas de datos precedentes permitiendo así tener mecanismos adicionales que se encarguen de procesar, almacenar y validar información y datos de distinta composición.

Ejemplo de la situación anterior, pero relacionado mayormente con la información patrimonial (identitaria, cultural) es el caso del reconocimiento de textos tibetanos y budistas. En esta experiencia se presenta un método para cifrar y compilar textos budistas de escrituras tibetanas originales que se encuentran en formato Romanizad. Para ello se utiliza un sistema GUI, interfaz con utilización de gráficos (reconocimiento de gráficas). En las experiencias presentadas se logró un resultado de reconocimiento del 99,4% para 28.954 caracteres.

En Chile no existe una base de datos, o fuente de información que almacene e intérprete material patrimonial, cultural e identitario, por lo mismo, una práctica que atienda esta situación puede perfectamente apoyarse en la caso de los textos tibetanos y budistas.

ii.- Necesidad de diferenciar procesos de homologación y heterogeneidad en los datos: si bien es frecuente en la literatura internacional los ejemplos que tienden a defender y promover los procesos de homologación de los datos en pos de su interacción y comparabilidad, también se apunta a los requerimientos de incorporar otro tipo de medidas que permitan a determinados datos ser heterogéneos en vista de su excesiva especialización pero a la vez poseer buscadores semánticos y promover la colaboración transdisciplinaria para gestionarlos.

SOBRE LAS VINCULACIONES EN Y PARA LA GESTION DE DATOS DE INVESTIGACION E INFORMACION CIENTIFICA

1º.- El sector público contribuye fuertemente para generar plataformas que permitan la comunicación e interacción entre distintas instituciones ligadas a un área científica. Estas vinculaciones inicialmente son mediante el intercambio y disposición de información y datos de investigación los cuales pueden ayudar al desarrollo del conjunto de las instituciones que utilizan esta información y/o mediante la interacción en proyectos interinstitucionales. Para el caso de Chile, será clave la definición del rol de los agentes del Sistema Nacional y los Sistemas Regionales de Innovación. Así mismo, la decisión de creación de Agencias de Coordinación y/o almacenamiento de Datos de Investigación e Información Científica será un tema relevante.

2º.- Las experiencias demuestran que el poseer una homogeneidad y formatos estándar en la información y datos generados por instituciones de un área científica determinada, provoca la emergencia de proyectos entre dos o más instituciones (ver los casos de Secuencias genómicas).

Ello implica, que se generan mecanismos internos entre comunidades epistémicas que potencian su desarrollo mediante datos de investigación e información científica. Por ello, la detección de las redes y sus niveles de interacción, será relevante para analizar la existencia de capacidades.

Estos dos ejes, implican:

i.- Considerando los de **requerimientos de soporte científico-tecnológico para el desarrollo de los sectores competitivos**, y el **requerimiento de del desarrollo de plataformas de Ciencia, Tecnología e Innovación y Programas Nacionales de Ciencia, tecnología e Innovación**, se hace imprescindible la (i) *Generación de plataformas que permitan la comunicación e interacción entre distintas instituciones ligadas a un área científica*, inicialmente mediante el intercambio y disposición de información y datos de investigación, los cuales pueden ayudar al desarrollo del conjunto de las instituciones que utilizan esta información, y/o mediante la interacción en proyectos interinstitucionales que apunten a la planificación y diseminación de datos de investigación e información científica.

ii.- *De forma de responder a los aprendizajes de las experiencias internacionales* revisadas y, por ello, al **cumplimiento de estándares internacionales**, se debe promover el (ii) *Desarrollo de* una homogeneidad y formatos estándar en la información y datos generados por instituciones de un área científica determinada, con el objetivo de provocar la emergencia de proyectos entre dos o más instituciones dirigidos al fortalecimiento del procesamiento, validación y diseminación de datos de investigación e información científica

FUENTES BIBLIOGRÁFICAS

- Bartolo, L. (2006). "International Scientific Data, Standards, and Digital Libraries: An NSF NSDL (U.S.) and CODATA Workshop". Kent State University, Materials Informatics Lab, United States
- Benavent, R. (2009) "Factor de impacto, competencia comercial entre Thomson Reuters y Elsevier, y crisis económica". Revista electrónica Think Epi, Grupo de Análisis sobre Estrategia y Prospectiva de la Información. <http://www.thinkepi.net/>
- Brown, J.S., P. Duguid (1991). "Organisational Learning and Communities of Practice: Toward a Unified View of Working, Learning and Innovation", *Organization Science*, 2 (1)
- Cash, Patrick and Mink, Dustin (2007). *Data Mining Techniques to Study Voting Patterns in the US*". Sikha Bagui, Data Science Journal. Volúmen 6, Department of Computer Science, University of West Florida, Pensacola.
- Chen – Yuan Liu, Huang – Cuang Lin, and Kojima, M. (2007) *A Character Recognition on object oriented design for Tibetan Buddhist Text*. Data Science Journal, Volumen 6, Department of Information Technology and Communication, Tungham University, Taipei, Taiwan.
- Dumouchel, B., Sabourin, M. (2007) *Canadian National Consultation On Access To Scientific Research Data*.
- Dupuy, C., Gilly, P. (1996). "Collective Learning and Territorial Dynamics: A New Approach to the Relations Between Industrial Group and Territories", *Environment and Planning*, 28: 1603-1616.
- Elster, J. (2006). "No queme sus puentes antes de llegar a ellos. Algunas ambigüedades y complejidades de los compromisos previos", en A. Mockus y F. Cante (comps.), *Acción Colectiva, racionalidad y compromisos previos*. Bogotá: Universidad Nacional de Colombia.
- Herreros, F. (2002). *¿Por qué confiar? El Problema de la creación de capital social*, Madrid: Centro de Estudios Avanzados en Ciencias Sociales del Instituto Juan March de Estudios e Investigaciones.
- García M., Cancino R. (2008). "Diseño institucional para el fortalecimiento de los sistemas regionales de innovación". Ponencia presentada al V Encuentro de Estudios Regionales.
- Giuliani, E. (2005). "The Structure of Cluster Knowledge Networks: Uneven and Selective, not Pervasive and Collective", *DRUID Working Paper*, Nº 05-11.
- Giuliani, E., M. Bell (2004). *When micro shapes the meso: Learning networks in a Chilean wine cluster*. Paper Nº 15, SPRU Electronic Working Paper Series.
- Gordon, I., P. MacCan (2000). Industrial Cluster: Complexes, Agglomeration and/or Social Network?, *Urban Studies*, 37 (3): 513-532
- Granovetter, M. (1985). "Economic action and social structure: the problem of embeddedness", *American Journal of Sociology*, 91: 481-510.

- IDER-UFRO (2008). Informe final del “Estudio de Capacidades instaladas en regiones para la gestión de la ciencia, la tecnología y la innovación”. SUBDERE-IDER.
- IDER-UFRO (2008). Informe de Avance del Estudio “Plataformas emergentes de ciencia y tecnología para los sectores con potencial de crecimiento en Chile”.
- Lawson, C., Lorenz, E. (1999). “Collective learning, tacit knowledge and regional innovative capacity”, *Regional Studies*, 33 (4): 305-317.
- Lundvall, B. (1992). *National System of Innovation. Towards Theory of Innovation and Interactive Learning*. London: Pinter
- MacCann, P., Tomokazu, A., Gordon, I. (2002). “Industrial Cluster, transactions cost and the institutional determinants of MNE location behaviour”, *International Business Review*, 11: 647-663.
- Martínez, C. (2004). “La capacidad innovadora de las redes de desarrollo regional: el valor añadido de la colaboración, la competitividad y la difusión de conocimiento”, *Información Comercial Española*, 812: 55-69
- Morrison, A., R. Rabelotti (1995). “Knowledge dissemination and informal contacts in an Italian wine local system”. *Paper presentado en Summer Conference Dynamics of Industry and Innovation: Organization, Network and System*. Copenhagen, 27-Junio 2005.
- National Center for Health Statistics. (2008). *Guidelines for Proposal Submission*. Last revised: November 12.
- National Research Council (1997) *Bits of Power: Issues in Global Access to Scientific Data*. Committee on Issues in the Transborder Flow of Scientific Data.
- OECD (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding*.
- Pilon, S., DeBresson, C. (2003). “Local Culture and Regional Innovation Network: Some Propositions”, en D. Fornhal y T. Brenner (eds.), *Cooperation, Network and Institutions in Regional Innovation Systems*. Cheltenham: Edward Elgar.
- Porter, M., Fuller, M. (1986). “Coalitions and global strategy”, en M. Porter (ed.), *Competition in Global Industries*. Boston: Harvard Business Press.
- Ramaraj, N. and Senthamarai, S. (2007) *Concept Based Clustering for Descriptive Document Classification*. *Data Science Journal*, Volumen 6, página 1 – 8. India
- Reichman, J.H. and Paul F. Uhlir, (2003) “A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment”, *66 Law & Contemporary Problems*.
- Reiter, Erick (2007). *VolcanoGasML: a format to exchange geochemical Volcanic gases data*. *Data Science Journal*, Volúmen 6, pág 7 – 18, *Francia*.

Shepsle, K. (1998). "Discrecionalidad, instituciones y el problema del compromiso del gobierno", en S. Saiegh y M. Tomáis (comps.). *La Nueva Economía Política: Racionalidad e Instituciones*. Buenos Aires: Eudeba

Varios autores. (1997). "Bits of Power Issues in Global Access to Scientific Data". *Committe on Issues in the Transborder Flow of Scientific Data, National Research Council*. NAS.

Varios autores. (1995). "Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources". Steering Committee for the Study on the Long-Term Retention of Selected Scientific and Technical Records of the Federal Government, National Research Council.

Varios autores (2009). "Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council". *Interagency Working Group on Digital Data HARNESSING THE POWER of DIGITAL DATA for SCIENCE AND SOCIETY*.

Uhlir & Schröder, (2007). "Open Data for Global Science", *Data Science Journal*, CODATA

Uhlir, et al, "Toward Implementation of the GEOSS Data Sharing Principles", (2009). *Journal of Space Law*

Williamson, O. (1975). *Markets and Hierarchies. Analysis and Antitrust Implications*. New York: The Free Press.

Wiseman, Yair. (2007) *The Relative Efficiency of Data Compression By LZW and LZSS*. *Data Science Journal*, Volúmen 6, páginas 1 – 6, Israel.

Fuentes en Internet:

Fundación Española para la Ciencia y la Tecnología: <http://www.fecyt.es/>

Organización de Los Estados Iberoamericanos: <http://www.oei.es/>