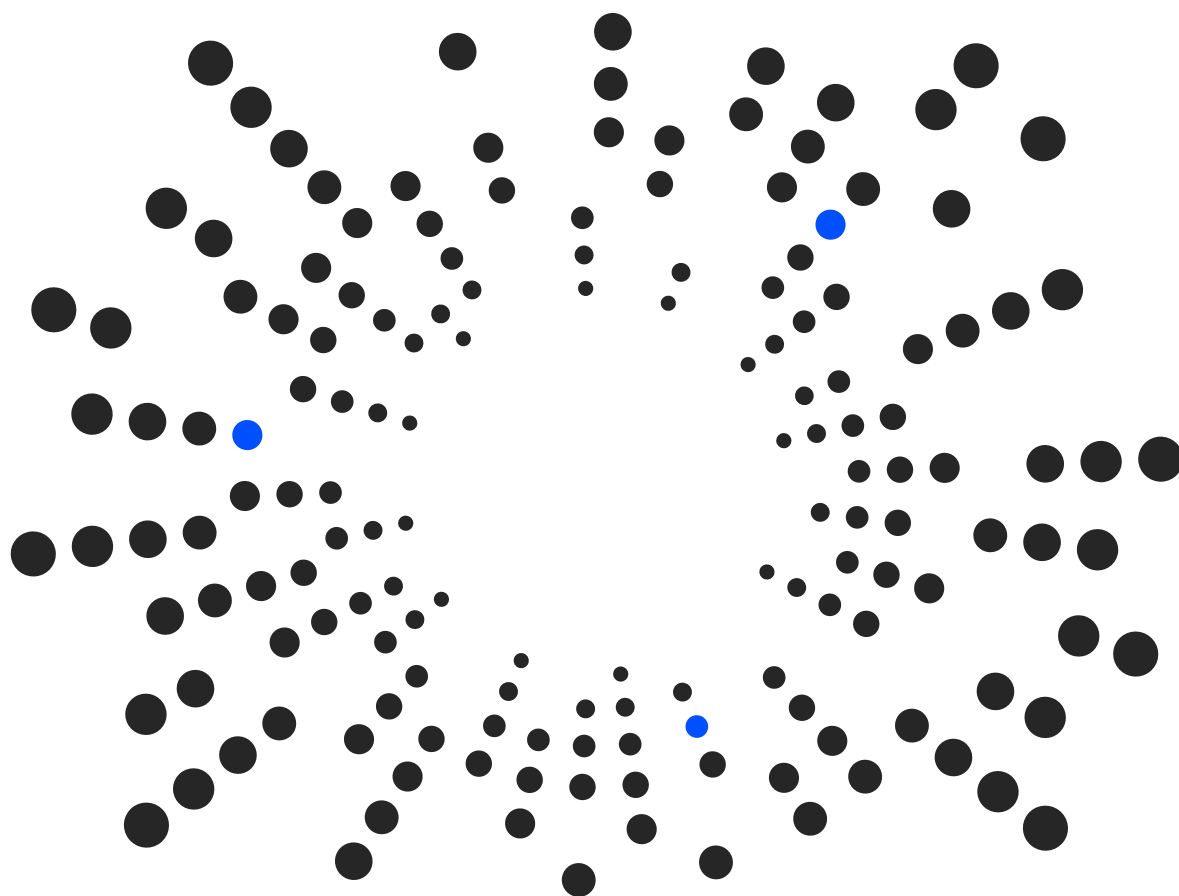


Elaboración de un Mapa de Fuentes de Información para Caracterizar el Ecosistema de CTCI



CHILE, ENERO 2021

AUTORES

Isónoma Consultorías Sociales

Rodrigo Fernández

Pablo Cárcamo

Cecilia Robayo

José Manuel Pino

CONTRAPARTE CONSEJO CTCI

Katherine Villarroel Gatica,

Secretaria Ejecutiva

Paulina Peña Romero

Este estudio presenta un mapa para identificar y caracterizar las fuentes de información de las agencias públicas y algunos privados que tienen relación directa e indirecta con temáticas de CTCI. Además, propone preguntas de investigación desafiantes que permitirán mejorar la caracterización del ecosistema CTCI, ampliando la visión sobre indicadores.

ETIQUETAS

Documento técnico, Estudios / CTCI / Evaluación y Monitoreo

Tabla de Contenidos

Presentación—p.6	
1. Antecedentes —p.8	
2. Objetivos del estudio —p.10	
3. Marco conceptual —p.12	
3.1. Datos e indicadores en CTCI ¿por qué y para qué? —p.13	
3.2. Línea de base y brechas para fuentes de datos sobre el ecosistema de CTCI —p.16	
3.3. Elementos de diagnóstico para las fuentes de datos —p.20	
4. Metodología —p.26	
4.1. Levantamiento, sistematización y análisis de información secundaria —p.27	
4.2. Levantamiento, sistematización y análisis de información primaria —p.28	
5. Casos internacionales y nacionales de uso de datos —p.31	
5.1. Levantamiento, sistematización y análisis de información secundaria —p.32	
5.2. Análisis de información secundaria —p.35	
5.3. Benchmarking —p.46	
6. Mapa de fuentes de información vinculadas a CTCI —p.49	
6.1. Metodología —p.50	
6.2. Diagnóstico de fuentes que componen al mapa —p.54	
6.3. Diseño y representación gráfica del mapa —p.62	
7. Preguntas de investigación que requieren de otras fuentes de información —p.77	
7.1. Potenciales preguntas de investigación —p.78	
7.2. Análisis de la información construida —p.87	
8. Conclusiones —p.89	
9. Referencias bibliográficas —p.94	
10. Anexos —p.98	
1. Levantamiento de información secundaria —p.99	
2. Levantamiento de información primaria —p.105	

Índice de Tablas

- | | |
|--|---|
| Tabla 1. Dimensiones y niveles para la caracterización del ecosistema —p.18 | Tabla 12. Actores relevantes a entrevistar, perfil tradicional —p.102 |
| Tabla 2. Línea de base y brechas en fuentes de datos para caracterización del ecosistema —p.19 | Tabla 13. Actores relevantes a entrevistar, perfil no tradicional —p.103 |
| Tabla 3. Cruce entre perfiles de actores a entrevistar y categorías analíticas —p.29 | Tabla 14. Pauta de entrevista para actores del perfil “Análisis, promoción o ejecución de las políticas públicas” —p.105 |
| Tabla 4. Análisis de la experiencia en investigación —p.35 | Tabla 15. Pauta de entrevista para actores del perfil “Generación, almacenamiento, procesamiento y disposición de datos” —p.109 |
| Tabla 5. Análisis de repositorios —p.40 | |
| Tabla 6. Diagnóstico de fuentes para establecimientos formales —p.54 | |
| Tabla 7. Diagnóstico de fuentes para empresas —p.56 | |
| Tabla 8. Diagnóstico de fuentes para instituciones de educación superior —p.58 | |
| Tabla 9. Diagnóstico de fuentes para emprendimientos —p.59 | |
| Tabla 10. Análisis de dimensionamiento general de la red para mapa desagregado de fuentes de información —p.73 | |
| Tabla 11. Métricas de centralidad para cada una de las tablas/fuentes del mapa desagregado —p.74 | |

Índice de Figura

- | | | | |
|-----------|--|------------|--|
| Figura 1. | Data Wheel, modelo DMBOK —p.22 | Figura 10. | Primera representación gráfica mapa lógico de fuentes para ecosistema CTCI —p.69 |
| Figura 2. | Hexágono de factores ambientales, modelo DMBOK —p.25 | Figura 11. | Segunda representación gráfica mapa lógico de fuentes para ecosistema CTCI —p.70 |
| Figura 3. | Diagrama de Contexto para el diagnóstico de fuentes de datos —p.53 | Figura 12. | Tercera representación gráfica mapa lógico de fuentes para ecosistema CTCI —p.71 |
| Figura 4. | Resumen de resultados para diagrama de contexto —p.60 | Figura 13. | Cuarta representación gráfica mapa lógico de fuentes para ecosistema CTCI —p.72 |
| Figura 5. | Primera representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI —p.63 | | |
| Figura 6. | Primera representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI —p.64 | | |
| Figura 7. | Segunda representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI —p.65 | | |
| Figura 8. | Tercera representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI —p.66 | | |
| Figura 9. | Cuarta representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI —p.67 | | |

Presentación

A continuación, se presenta el Informe final del estudio: “Elaboración de un mapa de fuentes de información para caracterizar el ecosistema de ciencia, tecnología, conocimiento e innovación”, solicitado por el Consejo Nacional de Innovación para el Desarrollo.

En el capítulo 1, se presentan los antecedentes que dan origen al estudio.

En el capítulo 2, se indican los objetivos del estudio.

En el capítulo 3, se revisa el marco conceptual, donde se desarrollan brevemente las nociones fundamentales para el desarrollo del presente estudio.

En el capítulo 4, se da a conocer el diseño metodológico, a través del cual se planificaron y elaboraron estrategias y procedimientos que permitieron la recogida de datos, su procesamiento, análisis e interpretación con el propósito de dar respuesta a los problemas planteados en los objetivos de este estudio.

En el capítulo 5, se revisa el análisis del levantamiento de información secundaria, a través de la cual se revisaron experiencias de investigación (papers) en otros países y en Chile donde se utilicen datos del sistema público, distintos a los típicamente utilizados en estudios de ciencia, tecnología, conocimiento e innovación (CTCI).

En el capítulo 6, se presenta el Mapa Conceptual de Fuentes de Información vinculadas directamente a CTCI.

En el capítulo 7, se desarrolla el análisis de la información primaria, con la cual se apoya la elaboración de preguntas de investigación que requieren de otras fuentes de información.

En el capítulo 8, se dan a conocer las principales conclusiones del estudio.

1. Antecedentes

1. Antecedentes

Los avances relacionados a los procesos de transformación digital han permitido disponer de amplios volúmenes de datos referidos a los distintos actores que componen el ecosistema de Ciencia, Tecnología, Conocimiento e Innovación (CTCI). Esta situación ha sido aprovechada por un número creciente de países, poniendo a disposición de la comunidad científica distintas fuentes de información relativa a la actividad de reparticiones y agencias públicas. Lo anterior ha impactado positivamente en las posibilidades de implementar nuevas metodologías que constituyan insumos para el mejoramiento en el diseño y evaluación de políticas públicas (JISC, 2011).

Dado que el sector público no sólo es un importante usuario o consumidor de información, sino también uno de sus principales productores, si la comunidad científica tiene un mayor acceso para un uso más efectivo de la información generada por agencias públicas (OECD, 2008), se esperan mayores beneficios tanto en términos del volumen y calidad de la investigación científica, como en la calidad de la evidencia generada para el diseño y evaluación de políticas públicas.

En Chile existen fuentes de datos con un gran nivel de detalle y periodicidad para la caracterización del ecosistema de CTCI, tales como las encuestas nacionales de innovación y de gasto y personal en I+D, o la información referente a los fondos públicos referidos a la oferta programática de CORFO y ANID, los cuales requieren de un acto administrativo en el que se ingresa información detallada¹ sobre las entidades postulantes, asimismo como en relación a los proyectos concursados. Por otra parte, una ventaja del contexto nacional es contar con un Rol Único Tributario (RUT) para las personas y establecimientos, lo que permite la trazabilidad detallada de la información para diversas fuentes de datos.

El Consejo Nacional de Innovación para el Desarrollo considera de una alta relevancia hoy contar con un mapa que permita identificar y caracterizar las fuentes de información de las agencias públicas que tengan relación directa e indirecta con temáticas de Ciencia, Tecnología, Conocimientos e Innovación, para luego proponer estrategias para su uso, de manera tal de potenciar su valiosa existencia al ponerla a disposición de la comunidad científica, lo que permitirá mejorar la caracterización del ecosistema para el desarrollo de lineamientos estratégicos y políticas públicas basadas en evidencia en torno a estas temáticas.

1. Información detallada sobre cada beneficiario (RUT, actividad económica, ventas, número de trabajadores), información del proyecto adjudicado, información de línea base que permite medir los resultados esperados, entre otros.

2. Objetivos del estudio

2. Objetivos del estudio

Objetivo General

Elaborar un mapa general que muestre las fuentes de información que permitirían caracterizar el ecosistema de CTCI, incluyendo en este mapa su ubicación, disponibilidad, protocolos de seguridad asociados, formato, calidad de los datos contenidos, periodicidad de actualización, permisos de acceso, disposición a compartir, entre otros.

Objetivos Específicos

- 1. Mostrar casos internacionales y nacionales de uso de datos:** Encontrar experiencias de investigación (papers) en otros países y en Chile donde se utilicen datos del sistema público, distintos a los típicamente utilizados en estudios de CTCI, como número de investigadores, patentes, publicaciones, gasto en I+D, tasa de innovación en empresas. etc.
- 2. Dibujar una primera versión del mapa:** Identificar en las entidades públicas (Ministerios, Agencias, Institutos, entre otros) qué información tradicional existe en sus repositorios de datos, los cuales permitan caracterizar el ecosistema CTCI (considerando todas las fuentes de información disponibles y sus características), como por ejemplo: Encuestas de Innovación en Empresas, Encuesta de Personal y Gasto en I+D, Presupuesto Públicos para I+D (GBARD), GEM, Encuesta Nacional de Percepción Social de la Ciencia y la Tecnología, etc.
- 3. Definir preguntas de investigación desafiantes junto a sus implicancias en el mapa definido en este estudio:** Sobre la base de la revisión de casos, entrevistas y los requerimientos de la contraparte definir preguntas de investigación que requieren de otras fuentes de información, por ejemplo, trayectorias laborales de PhD, crecimiento en ventas de empresas con base tecnológica. Con esto analizar la disponibilidad de información cruzada.
- 4. Generar una segunda versión del mapa con nuevas fuentes de información y propuestas de posibles conexiones:** Identificar otras fuentes menos tradicionales que sirvan a responder las preguntas de investigación identificadas en el punto anterior, y explicitar qué desafíos y ventajas existen para mejorar la coordinación de las distintas fuentes de información. Realizar propuestas de principios básicos que aporten a la disponibilidad y utilización de los datos dada la situación actual descrita en este estudio, considerando recomendaciones de organizaciones internacionales como la OCDE y ONU, entre otros.

3. Marco conceptual

3. Marco conceptual

3.1. Datos e indicadores en CTCI ¿Por qué y para qué?

La creación de indicadores para caracterizar los ecosistemas de CTCI en el mundo han sufrido cambios que se condicen con las distintas definiciones y roles establecidos para la ciencia y la tecnología por parte de las políticas públicas. Mientras en un comienzo los datos sobre gasto en I+D, capital humano y otras inversiones asociadas con la innovación eran suficientes para la gestión de la CTCI, las demandas actuales promueven nuevas formas de pensar el fenómeno.

La esperanza de que la ciencia y tecnología permitan responder a los desafíos derivados del cambio climático y entregue respuestas para el desarrollo económico ha provocado una discusión a nivel global sobre la necesidad de robustecer los indicadores actuales y crear nuevas fuentes de información.

Las reflexiones en este sentido han estado lideradas por académicos e instituciones globales, quienes abogan por una gestión con mayor reflexividad en los ecosistemas de CTCI. Comenzando en 2012 con la declaración de San Francisco sobre la evaluación científica², la búsqueda de nuevas formas de medición de la actividad científica y tecnológica ha estado orientada por distintas inspiraciones, pero que aun así comparten el espíritu por la identificación de los impactos que las métricas tradicionales no son capaces de capturar.

En 2015 el manifiesto de Leiden (Hicks et. Al. 2015) y la revisión sobre el rol de las métricas en la gestión y evaluación de la investigación coinciden en el avance de las sugerencias planteadas por la declaración de San Francisco. El reporte “The metric tide”³, encargado por el gobierno del Reino Unido identificó un amplio listado de problemas con los sistemas de evaluación de la ciencia y la tecnología en la actualidad, enfatizando problemas de incentivos a nivel individual, y de interoperabilidad de los datos en términos de gestión. El manifiesto de Leiden propone 10 principios de evaluación basados en el trabajo de los académicos que forman parte del manifiesto, que buscan aumentar la transparencia, mejorar la construcción de indicadores y buscar complementos para la interpretación de resultados.

2. Disponible en <https://sfedora.org/read/>

3. Disponible en <https://responsiblemetrics.org/the-metric-tide/>

3. Marco conceptual

Por otra parte, en 2017 el manual de Valencia, también llamado Manual de Indicadores de Vinculación de la Universidad con el Entorno Socioeconómico⁴, busca aportar a la discusión con indicadores de transferencia tecnológica y otras formas de interacción entre la universidad y la sociedad. Posteriormente en 2018 la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo publica la documentación⁵ con la cual considerará la revisión de las políticas públicas en el área de la ciencia y la tecnología, prestando especial atención a los efectos en los objetivos de desarrollo sostenible.

Estas publicaciones demuestran la preocupación global por la búsqueda de nuevos indicadores para las políticas en el área de la CTCI, pero son pocos los países que efectivamente han sido capaces de promover la experimentación y construcción de nuevos datos en estos ámbitos. Salir de los patrones tradicionales de medición es un esfuerzo que requiere capacidades, intenciones y creatividad, una combinación difícil para instituciones y actores orientados a la generación de información en base a estándares internacionales.

En relación con el proceso de construcción de indicadores, es importante tener en consideración la lógica de su conformación y los objetivos que se persiguen. Según estas características el valor de los indicadores, su orientación y formato de construcción puede presentar alta variabilidad.

Inspirado en el ejercicio reflexivo sobre la evaluación las políticas públicas en ciencia y tecnología realizado por Jordi Molas-Gallart (2012), es posible clasificar los objetivos en al menos tres categorías:

- Objetivos distributivos, que buscan apoyar la decisión para la gestión y transferencia de recursos públicos.
- Objetivos de mejora y aprendizaje, cuya característica principal se basa en la posibilidad de entregar retroalimentación detallada para el perfeccionamiento de los estándares medidos.
- Objetivos de control, que buscan la constitución de auditorías para la evaluación del uso efectivo de recursos y tareas comprometidas.

4. Disponible en <http://www.ricyt.org/2017/06/manual-de-indicadores-de-vinculacion-de-la-universidad-con-el-entorno-socioeconomico-manual-de-valencia/>

5. UNCTAD (2018). Effectively harnessing science, technology and innovation to achieve the Sustainable Development Goals. Disponible en https://unctad.org/system/files/official-document/ciimem4d17_en.pdf

3. Marco conceptual

A este podemos agregar indicadores de gestión del cambio/direccionalidad, derivados de las nuevas tendencias por la búsqueda de políticas públicas en CTCI capaces de responder a desafíos climáticos y sociales.⁶

Al mismo tiempo es relevante tomar en cuenta el grado de alcance que los indicadores consideran. De acuerdo con el nivel de alcance esperado por los indicadores, podemos distinguir un continuo dependiente del grado de Granularidad o Globalidad que los indicadores pueden alcanzar. A modo de ejemplo los desafíos a medir podrían clasificarse

- Desafíos Globales, como por ejemplo los Objetivos de Desarrollo Sostenible.
- Desafíos sectoriales, como por ejemplo la búsqueda de encadenamientos productivos en base a actividades altas en conocimiento.
- Desafíos disciplinares, como por ejemplo la medición efectiva del impacto de distintas disciplinas en la sociedad.
- Desafíos locales, como por ejemplo la búsqueda de indicadores y mecanismos de conservación ambiental en localidades específicas.

Esta clasificación de indicadores es relevante para la búsqueda de nuevas formas de medición y la construcción de nuevas métricas para evaluación de políticas y acciones en el ecosistema de CTCI, uno de los objetivos específicos del presente informe. En particular, la búsqueda de nuevas fuentes de información, preguntas de investigación y potenciales indicadores debe estar guiada por una orientación que sea capaz de dilucidar la razón del indicador y su uso posterior. Esto necesariamente debe estar vinculado con orientaciones estratégicas del ecosistema, que permitan entregar luces en espacios donde falta de información, pero al mismo tiempo permitan tomar acciones que vayan más allá de los esfuerzos distributivos y de control.

6. Por ejemplo, en los nuevos marcos de la política pública en CTCI promovidos por el Transformative Innovation Policy Consortium. Más información disponible en <http://www.tipconsortium.net/>.

3.2. Línea de base y brechas para fuentes de datos sobre el ecosistema de CTCI

Los datos que no son típicamente utilizados en el estado del arte en estudios de CTCI son los que permiten el abordaje de temas emergentes. A su vez, un tema puede ser considerado emergente cuando se orienta a aquellos fenómenos del ecosistema que en la actualidad no pueden ser contrastados empíricamente. Estos fenómenos son los que permitirían caracterizar de mejor manera a la estructura y dinámica del ecosistema y sus agentes, para de esa manera avanzar en dotar de mayor evidencia al diseño y evaluación de políticas.

Lo anterior, es especialmente sensible en el caso de los ecosistemas de CTCI, ya que el elemento común a sus agentes es el estar involucrados en actividades que son intensivas en conocimiento, el que a su vez es un tipo de bien/servicio de alta complejidad teórica y empírica.

Desde el punto de vista teórico, el conocimiento es un bien cuyo beneficio social es intrínsecamente mayor que el privado, dada su condición de bien público. A su vez, lo que define económicamente a un bien público es la ausencia de rivalidad en su acceso y de congestión para su consumo, es decir, que una idea no se acaba cuando más agentes la consumen y no es necesario hacer una fila para acceder a ella en base a sus características intrínsecas. De este modo, copiar una idea siempre será menos costoso que producirla, razón por lo cual se han desarrollado distintos mecanismos de compensación para enfrentar este problema en la alineación de incentivos a innovar por parte de los agentes.

Por otro lado, el conocimiento puede ser entendido como un bien rico en colaboración - y no sólo en competitividad - dependiendo del nivel de análisis. En efecto, si bien la competitividad entre agentes por el desarrollo e introducción exitosos de una idea es un elemento importante de la dinámica del conocimiento, también es cierto que los competidores necesitan colaborar con otros agentes en distintos niveles para competir. Estos agentes pueden ser otras empresas, instituciones proveedoras de conocimiento, emprendimientos de base científica y/o tecnológica, los hubs de innovación y emprendimiento, y las distintas instancias del sector público tanto a nivel local como central.

3. Marco conceptual

Por otro lado, las capacidades involucradas en la captura de datos para los agentes del ecosistema sólo permitían capturar fenómenos contables al interior de empresas, instituciones proveedoras de conocimiento y el sector público, gracias al desarrollo de los sistemas de registros administrativos en primer lugar, y luego gracias al desarrollo de encuestas a establecimientos que permitieron caracterizar al personal, unidades organizacionales y actividades colaborativas primarias involucradas en procesos de innovación. El caso de los emprendimientos permaneció inexistente para las fuentes de información, salvo aproximaciones mediante encuestas de hogares.

Para el trazado de una línea de base y las brechas de información respecto del estado de las fuentes de datos, se consideró la visión CNID 2020 para la caracterización del ecosistema, la que considera por un lado un conjunto de cinco dimensiones relevantes basadas en el modelo de Uzi de Han (2008), estas son:

- **Capital de Conocimiento:** instituciones proveedoras, desarrollos, mecanismos de protección y vigilancia tecnológica y todo tipo de actividades orientadas a la I+D o la I+D+i.
- **Capital Financiero:** financiamiento de proyectos, personal, infraestructura, equipamiento, etc.
- **Capital Humano:** capacidades desarrolladas a nivel del personal involucrado, tanto desde el punto de vista académico como de la experiencia en el sector.
- **Capital Social:** actividades colaborativas, redes de financiamiento y mecanismos para el establecimiento de vínculos (confianza y reciprocidad en redes).
- **Capital de Emprendimiento:** dinámica de la actividad y valoración de emprendimientos de base científica y/o tecnológica.

3. Marco conceptual

Por otro lado, se contemplan tres niveles de análisis (Olivari & Sturbin, 2016; CNID, 2018), los cuales son:

- **Stock:** situación actual de las variables consideradas en cada dimensión.
- **Conexiones:** vinculaciones entre agentes y proyectos en cada dimensión.
- **Evolución:** trazabilidad de unidades de análisis, agentes y variables consideradas en cada dimensión.

En la siguiente Tabla se expone el modelo considerado por la contraparte para la caracterización del ecosistema.

Tabla 1. Dimensiones y Niveles para la caracterización del ecosistema

DIMENSIÓN	NIVEL DE ANÁLISIS		
	Stock	Conexiones	Evolución
Capital de Conocimiento			
Capital Financiero			
Capital Humano			
Capital Social			
Capital de Emprendimiento			

Fuente: Elaboración propia en base a CNID (2018).

3. Marco conceptual

Luego, se consideraron los resultados del trabajo realizado por Price Elton respecto del análisis de las fuentes datos a nivel de empresas, tales como el Servicio de Impuestos Internos (SII), el Servicio Nacional de Aduanas (SNA) y las encuestas relacionadas del Instituto Nacional de Estadísticas (INE) y el Ministerio de Economía (MINECON). En particular, se destaca que las fuentes consultadas permiten establecer un seguimiento y trazabilidad de las empresas, lo que permite cubrir algunos elementos relacionados con las brechas en el nivel de análisis referido a la evolución.

En la siguiente tabla se resume la línea de base (celdas en blanco) y la brecha relativa las fuentes de datos (celdas en celeste), tomando en consideración las consideraciones de la visión CNID 2020 y el análisis de las fuentes en el trabajo de Price Elton. Esta línea de base orientó la búsqueda y selección de los *papers*, del mismo modo en que orientó el diseño del mapa de fuentes para el ecosistema de CTCI a nivel nacional para focalizar aquellos aspectos relacionados a las brechas.

Tabla 2. Línea de base y brechas en fuentes de datos para caracterización del ecosistema

DIMENSIÓN	NIVEL DE ANÁLISIS		
	Stock	Conexiones	Evolución
Capital de conocimiento	Cobertura actual de las fuentes de datos, visión CNID 2020		Cobertura en base al trabajo de Price Elton
Capital financiero			
Capital Humano			
Capital Social			
Capital de Emprendimiento			

Fuente: Elaboración propia en base a CNID (2018) y Price Elton (2016).

Sin duda, el avance de las tecnologías de información y comunicación ha generado condiciones que permiten el manejo integrado de grandes volúmenes de datos, abriendo así grandes posibilidades, tanto para mejorar el acceso a información relevante de las partes interesadas, como para incrementar la calidad y la pertinencia de las intervenciones de política pública.

3.2. Elementos de diagnóstico para las fuentes de datos

En efecto, esta idea concebida alrededor de las “3V” del Big Data (Volumen, Variedad, Velocidad) permite reconocer elementos de tipo cuantitativo para la definición de un sistema de datos masivos. Por ejemplo, pueden desarrollarse criterios sobre los cuales el Volumen de los datos presiona por la exploración de nuevas tecnologías de almacenamiento que sean más eficaces y ágiles, del mismo modo que su Variedad instala la necesidad de considerar distintas arquitecturas que permitan interoperar fuentes estructuradas, semiestructuradas y no estructuradas. Por último, el modo en que la Velocidad del flujo de datos instala la necesidad de contar con estrategias de cómputo que permitan optimizar la capacidad de filtro sobre aquellas variaciones que son relevantes para responder de manera oportuna a las preguntas de valor mediante la extracción de los *insights* del proceso de análisis y visualización de la información.

Con posterioridad, esta idea de las 3V ha sido enriquecida con otras propiedades que son deseables para el diseño, desarrollo y mantención de un sistema de datos masivos, como por ejemplo la Veracidad de los datos, su capacidad de Viralización en aplicaciones y otras plataformas, su capacidad de Visualización y, por sobre todo, su capacidad para agregar Valor a la gestión.

No obstante, la situación anteriormente descrita corresponde a elementos que se ubican en la superficie de los actuales sistemas de datos masivos. En este sentido, la transición más importante al respecto es de índole cualitativa y se relaciona con el carácter mismo que adquieren los datos en el contexto de la cuarta revolución industrial o 4.0, caracterizada por la progresiva separación entre activos físicos y digitales (IMD, 2016-2019). Esta separación progresiva da pie a la apertura de las posibilidades de cambio, tanto a nivel incremental como a nivel disruptivo en todas las esferas de la actividad humana, desde la forma de concebir procesos productivos, la administración del Estado, el desarrollo de la actividad científica y tecnológica y las actividades de la sociedad civil en general.

3. Marco conceptual

Este potencial disruptivo de los activos digitales se relaciona con el proceso en que los datos han dejado de ser considerados como antecedentes, para pasar ser activos estratégicos en distintas esferas y ámbitos de actividad. Lo anterior implica que los datos no pueden ser considerados solamente como abstracciones y representaciones relativas a los fenómenos, ya que de alguna manera podemos sostener que los datos cada vez se acercan más a constituir los fenómenos mismos.

Esta transición de los datos a constituirse en activos estratégicos da cuenta de desafíos que no son solamente técnicos y tecnológicos en relación a toda la cadena de trabajo involucrada en su adquisición, almacenamiento, limpieza, análisis y visualización. En efecto, las tareas preliminares a aquellas de índole técnica y tecnológica están asociadas a cuáles son las preguntas de valor que se pretenden responder mediante el trabajo con datos, y en qué condiciones se encuentra la organización para el mejor aprovechamiento de las instancias de valor que estos representan.

De lo anterior, se desprenden las principales motivaciones que empujan al desarrollo de modelos para el gobierno de datos (*data governance*) y para la gestión de datos (*data management*), lo que será desarrollado en la siguiente sección.

La amplitud e intensidad con la cual los sistemas de datos masivos han adquirido un rol central en diversas esferas de actividad, tanto en el sector público, el sector privado y la academia, ha ido generando una progresiva masa crítica relativa a modelos de organización de operaciones que son intensivas y se encuentra “orientadas” por datos. Estas consideraciones no sólo corren para la organización y gestión de este tipo de operaciones, sino también para efectos del diseño y desarrollo de los aspectos organizacionales que permiten un mejor aprovechamiento del valor que es posible extraer de este tipo de activos, que hoy podemos considerar como estratégicos.

En este sentido, el diagnóstico para obtener un mapa general de fuentes de información que permitan una caracterización del ecosistema de CTCI, debe ser capaz de relacionar los aspectos estrictamente técnicos de los datos involucrados, con aquellos aspectos asociados a las estructuras organizacionales involucradas en su producción y uso.

Lo anterior, define la transición de la discusión sobre modelos de organización para el descubrimiento de conocimiento tales como *Knowledge Discovery Database* - KDD y posteriores (SEMMA o CRISP-DM), a los modelos de lo que hoy se denomina como “*Data Governance*”. En

3. Marco conceptual

efecto, la primera familia de modelos es heredera de un tiempo en que los datos aún no alcanzaban el nivel de masividad actual, por lo que los aspectos que relacionan a los datos con la estructura de la organización son parciales y, en algunos casos, algo opacos.

Para efectos del presente estudio, el enfoque conceptual que orientará el diagnóstico de las fuentes de información para la caracterización del ecosistema de CTCI será el estándar desarrollado por la *Data Management Association* – DAMA. Este estándar es conocido por su sigla DMBOK (*Data Management Body of Knowledge*), el cual ha permitido dar un paso importante en la base conceptual y metodológica relacionada a la consideración de los datos como activos estratégicos.

El DMBOK-2 (segunda edición) define al *Data Governance* como “*el ejercicio de autoridad, control y toma de decisiones compartidas (planificación, seguimiento y ejecución) sobre la gestión de los activos de datos*” (DMBOK, 2017). Dentro de esta definición, el modelo considera un primer marco de trabajo que consiste en la identificación de las áreas de conocimiento estratégico de las operaciones relacionadas a los activos de datos dentro de una organización. Estas áreas estratégicas constituyen la “Rueda DAMA” o “DAMA Wheel”, la cual es representada en la siguiente figura:

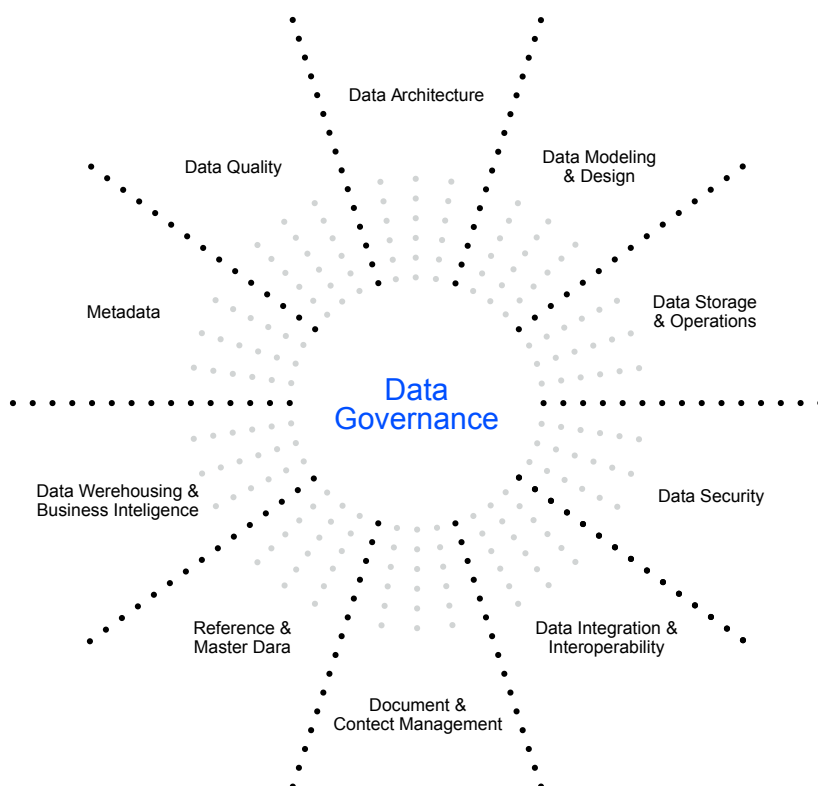


Figura 1. Data Wheel, modelo DMBOK

Fuente: DMBOK, 2017.

3. Marco conceptual

A su vez, la rotación de las áreas de conocimiento que son estratégicas para los activos de datos son:

- **Arquitectura:** estructura y funcionamiento de los sistemas de registro y gestión de datos para los agentes del ecosistema de CTCI. Lo anterior considera sistemas de bases de datos para la gestión de proyectos y programas de la oferta programática en CTCI, bases de datos de registros más amplios y permanentes tales como el Servicio de Impuestos Internos (SII) y el Servicio Nacional de Aduanas. En este nivel se consideran las tecnologías involucradas en la adquisición y levantamiento de sus datos, las que van desde encuestas por muestreo, censos y registros administrativos, los que en cada caso reportan distintos atributos para su caracterización tales como la periodicidad, continuidad, marcos muestrales, procedimientos de muestreo, criterios de empadronamiento, etc.
- **Diseño y modelos:** el diseño alude a la organización del esquema que permite identificar los requerimientos de representación de los datos, en función de las características de su arquitectura. Por su parte, los modelos refieren a la forma en que se conciben los datos como una representación de los conceptos de dominio relevantes (por ejemplo, objetivos y beneficiarios de la oferta programática) tanto en un nivel físico, lógico y conceptual. En este nivel se consideran elementos tales como el uso de distintos esquemas para motores de bases de datos, tales como el esquema relacional (SQL), orientado a objetos, basados en tiempo, basados en hechos, o relacionales (NoSQL), etc. También se consideran los distintos formatos de exportación y descarga para usuarios internos y externos.
- **Almacenamiento y operaciones:** tecnologías y capital humano involucrado en la administración y gestión de las operaciones relacionadas con el mantenimiento del ciclo de vida de los datos, asimismo como de aquellas relacionadas a su disponibilidad para clientes internos y externos. Lo anterior está sujeto al tipo de diseño y modelo de representación de los datos, y recorre distintas rúbricas de evaluación tales como el modelo ACID, modelo ACP y BASE.
- **Seguridad:** normas y procedimientos relacionados con el acceso a los datos en distintos niveles, lo que considera características de los usuarios de los activos de datos, tanto dentro (clientes internos) como fuera de la organización (clientes externos). También se consideran elementos relacionados con la protección de datos personas e industriales.

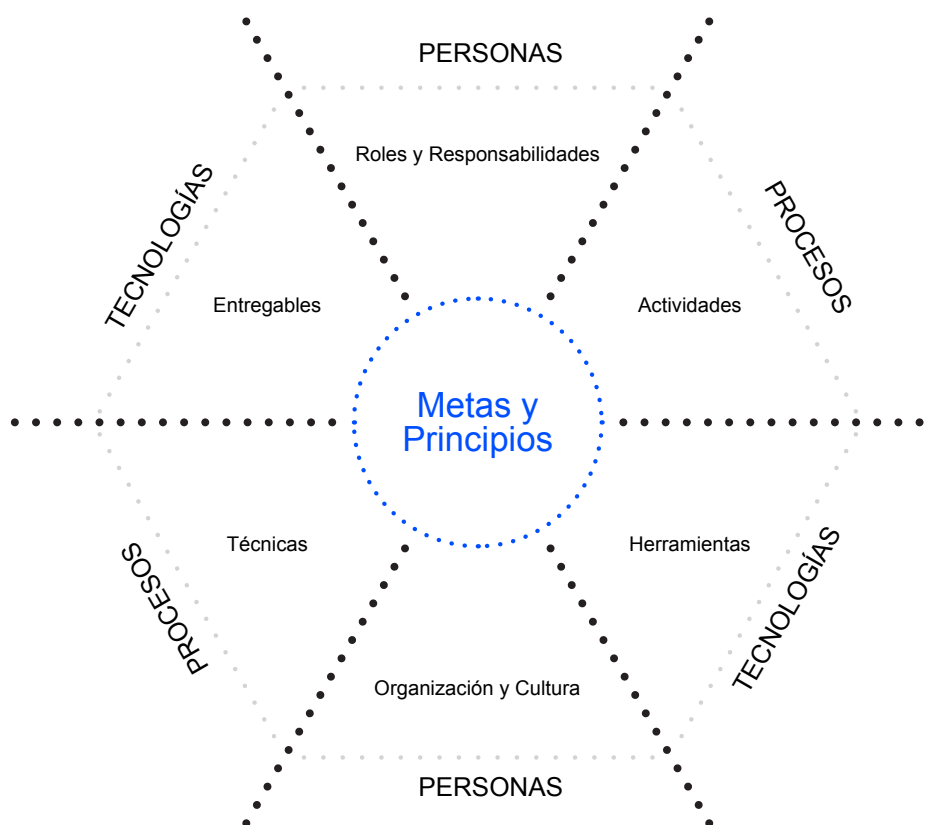
3. Marco conceptual

- **Integración e interoperabilidad:** recursos y procedimientos involucrados en la consolidación de datos, tanto desde el punto de vista de su conversión a distintos formatos y su capacidad de migración entre distintas aplicaciones y organizaciones.
- **Gestión documental y de contenido:** procesos relacionados con fuentes de datos no estructurados, vale decir, que no se ciñen al esquema relacional. Lo anterior implica la identificación de campos de normalización para fuentes de datos documentales o de libre selección, por ejemplo, aquellos campos en la gestión de proyectos en los que solicitan descripciones y justificaciones de proyecto. En este sentido, se hace énfasis en modelos de tratamiento para información de texto para distintas fuentes que no forman el núcleo de negocios de un programa o acción, pero que permiten complementar la comprensión de procesos y agentes involucrados (por ejemplo, redes sociales, solicitudes en OIRS, etc.).
- **Datos maestros y de referencia:** en la producción de datos converge la acción de distintas unidades y departamentos, los que a su vez tienen su propia planificación y estructura. De este modo, la gestión de datos maestros y de referencia permite identificar eventuales redundancias de información, asimismo como aquellos datos cuyo potencial de valor esté siendo subutilizado.
- **Data warehousing e inteligencia de negocios:** forma en que el funcionamiento general del stock y flujo de datos dentro de la organización permite el cumplimiento de metas comprometidas, tanto a nivel de procesos como de resultados e impactos.
- **Metadatos:** nivel de integración y documentación de la descripción de los datos producidos y utilizados. Tal y como se suele señalar, los metadatos son “datos que describen datos”, vale decir, el conjunto de diccionarios relacionados a los distintos datos producidos, asimismo como su nivel de coherencia e integración para distintas herramientas (API’S, ETL, etc.).
- **Calidad de los datos:** normas y procedimientos asociados a estándares de integridad, cobertura, continuidad y otras propiedades asociadas a la calidad de los datos.

3. Marco conceptual

Por otro lado, el modelo DMBOK-2 considera un modelo de diagnóstico genérico de factores ambientales, el que consiste en la intersección entre personas, procesos y tecnologías involucrados en cada área. Esta intersección se articula en torno a las metas y principios, tanto a nivel de la organización como a nivel de cada área en particular. De este modo, los factores ambientales respecto del trabajo con datos en el contexto de una organización pueden ser entendidos como un hexágono articulado en torno a las metas y principios de la organización, área o equipo, tal y como puede apreciarse en la siguiente figura:

Figura 2. Hexágono de factores ambientales, modelo DMBOK



Fuente: DMBOK, 2017

Este modelo conceptual para el diagnóstico de la organización es la base para el diseño del diagrama de contexto, el cual consiste en una operacionalización de estos criterios para efectos de identificar las personas, procesos y tecnologías críticas involucradas en el trabajo con datos, lo cual será detallado en la metodología relativa al diseño e implementación del mapa.

4. Metodología

4. Metodología

A continuación, se presenta el resumen del diseño metodológico para abordar cada uno de los objetivos específicos del presente trabajo.

4.1. Levantamiento, sistematización y análisis de información secundaria

El presente trabajo contempló un intensivo tratamiento de información secundaria, tanto para efectos de la caracterización de la experiencia nacional e internacional en investigación, como para el diseño e implementación de la primera versión del mapa.

En el primer caso, se trató de la selección de un conjunto de papers nacionales e internacionales⁷ en base a la matriz de línea de base descrita en el capítulo anterior, de manera tal que el análisis permitiera cubrir la mayor combinación entre dimensiones (capitales dentro del ecosistema) y niveles. Luego, se confeccionó una matriz de registro⁸ y sistematización del contenido de los papers, para de ese modo realizar el ejercicio de benchmarking que permitió obtener una visión panorámica de hasta qué punto la experiencia en investigación se hace cargo o no de las preguntas de frontera relacionadas a las brechas identificadas en la matriz conceptual.

Por otro lado, para el análisis de los repositorios de datos abiertos, fueron seleccionadas distintas experiencias internacionales tanto en Europa, Estados Unidos, Asia y América Latina⁹. Del mismo modo que en el caso de la experiencia en investigación, se diseñó e implementó una matriz de registro, sistematización¹⁰ y análisis de contenido referido a la documentación de estos repositorios, para de ese modo realizar el ejercicio de benchmarking que permitió comparar estas experiencias, asimismo como identificar las buenas prácticas en cada uno de los campos de caracterización de estos repositorios.

Por último, en relación al tratamiento de información secundaria para el diseño e implementación del mapa, se construyó una matriz de diagnóstico basada en las áreas de la Data Wheel que pudieron ser cubiertas por el contenido de la documentación de las fuentes seleccionadas. De esta manera, la implementación de estas matrices de diagnóstico permitió comprender de mejor manera las implicancias del mapa, asimismo como identificar el potencial de interoperabilidad de estas fuentes para la caracterización del ecosistema.

7. Revisar en el documento, Anexo 1 papers seleccionados.

8. Anexo 2 Matriz sistematización papers

9. Revisar en el documento, Anexo 3 Repositorios seleccionados.

10. Anexo 4 Matriz sistematización repositorios.

4. Metodología

4.2. Levantamiento, sistematización y análisis de información primaria

Para el diseño e implementación del mapa por un lado, y para definir preguntas de investigación desafiantes junto a sus implicancias en el mapa definido en este estudio, se levantó información clave de distintos actores que pertenecen a este ecosistema (CNID, 2019)¹¹, entre ellos: entidades públicas, empresas, emprendimientos de base científica y/o tecnológica, universidades acreditadas, centros de excelencia nacional e internacional, hubs de transferencia tecnológica, incubadoras apoyadas por CORFO, centros I+D ANID, centros financiados por el Programa Iniciativa Científica Milenio (ICM), Instituto tecnológico de la producción (ITP), Oficina de Transferencia Tecnológica y Licenciamiento (OTL).

Se entrevistó a una muestra de 26 actores clave del ecosistema, quienes entregaron información relevante en relación al objetivo de fortalecer el ecosistema CTCI.

Las entrevistas tuvieron dos objetivos fundamentales: i) Identificar las características de los datos en los repositorios de información del ecosistema de Ciencia, Tecnología, Conocimiento e Innovación y ii) entregar información útil para definir preguntas de investigación que requieren de otras fuentes de información que estén vinculadas indirectamente con CTCI.

Para lograr los propósitos planteados, se utilizó una muestra diferenciada según los perfiles de los potenciales entrevistados y las categorías analíticas relevantes para la investigación.

Se definió como “perfiles de los entrevistados”, el vínculo y conocimiento del potencial entrevistado con el área de la generación, almacenamiento, procesamiento y disposición de información relativa al ecosistema de CTCI en Chile, o bien con el área del análisis, promoción o ejecución de las políticas públicas de CTCI en Chile. En este sentido, se consideró que quienes trabajan directamente con datos, no necesariamente tienen conocimiento detallado del funcionamiento del ecosistema, mientras que quienes trabajan en el área de las políticas públicas, no necesariamente conocen en detalle las bases de datos con las que se relacionan.

11. CNID (2019). Ciencia, Tecnología, Conocimiento e Innovación Para Chile. Disponible en: http://www.cnid.cl/wp-content/uploads/2019/12/CTCI-para-Chile-y-Contexto-para-la-reflexion_web.pdf

4. Metodología

Por su parte, las “categorías analíticas” se refieren a una característica relacional del potencial entrevistado con el ecosistema de CTCI. Esta relación está definida por un vínculo tradicional y no tradicional con la ejecución y evaluación de políticas públicas en materia de CTCI. El límite entre lo tradicional y no tradicional no es necesariamente fijo, pero se entenderán por tradicional las disciplinas y análisis típicamente vinculados al ecosistema de CTCI, como por ejemplo STEM en el caso de disciplinas tradicionales, y las artes y humanidades en el caso de disciplinas no tradicionales. También a modo de ejemplo, los análisis econométricos vinculados a patentes, presupuesto de I+D y otros pueden ser definidos en el espacio de lo tradicional, mientras que observaciones provenientes de la sociología, antropología o estudios sociales de la ciencia pueden ser identificados como no tradicionales.

Para objetivos de la distinción de la muestra y las pautas de entrevista se realizó el cruce de categorías y perfiles con tal de identificar con claridad los espacios de generación de información, y la cantidad de entrevistas que se realizarán por cada subsegmento (Tabla 3).

Tabla 3. Cruce entre perfiles de actores a entrevistar y categorías analíticas

Perfil Tradicional	Perfil No Tradicional
Generación, almacenamiento, procesamiento y disposición de datos	Generación, almacenamiento, procesamiento y disposición de datos
Análisis, promoción o ejecución de las políticas públicas	Análisis, promoción o ejecución de las políticas públicas

Fuente: Elaboración propia.

4. Metodología

Para proceder al análisis de la información generada a partir de la aplicación de las entrevistas semiestructuradas, se utilizará la técnica de análisis de contenido, que consiste en “la descripción, interpretación y análisis de los patrones observados en la información cualitativa, así como los mecanismos y relaciones de causalidad que este tipo de información permite identificar” (Ezeminari, Rudqvist y Subbarao, 2002). Esta técnica refiere a la modalidad de análisis latente (implícito) en un texto o documento. Esto es “el proceso de identificar, codificar y categorizar patrones primarios en los datos” (Mayan, 2001, p.23), buscando el significado de palabras específicas, pero también de frase o párrafos completos, tomando en cuenta el contexto dentro del cual dichas palabras o frases fueron utilizadas y generando categorías analíticas que permitan incorporar el contexto, en el análisis del contenido mismo más allá del significado tradicional o coloquial de una palabra.

El análisis de contenido latente se inicia con un proceso de codificación o “proceso de identificar palabras, frases, temas o conceptos dentro de los datos de manera tal que los patrones subyacentes puedan ser identificados y analizados” (Morse y Field, 1995 en Mayan, 2001, p.24), lo cual se realizará, a través del programa de análisis de datos cualitativos asistidos por computadora, Atlas Ti V7, que permite analizar datos cualitativos. Al respecto, el software permitió agrupar códigos en base a dimensiones que se consideraron fundamentales a lo largo del estudio¹².

12. Para más información, consultar Pauta de Entrevistas en Anexo 5 de este informe.

5. Casos internacionales y nacionales de uso de datos

5. Casos internacionales y nacionales de uso de datos

5.1. Levantamiento, sistematización y análisis de información secundaria

En relación al objetivo específico vinculado a los casos nacionales e internacionales de uso de datos, se realizó una exhaustiva búsqueda, revisión y análisis de investigaciones (papers) internacionales y nacionales que utilizaron datos del sistema público y privado, incluyendo bases de datos públicas como las de CORFO, ANID, bases de datos de encuestas, bases de datos privadas, entre otras, para luego categorizarlas y describirlas por medio de dos metodologías: la sistematización de la información y el análisis bibliográfico y documental, y benchmarking. Además de lo anterior, se contempló la revisión de distintos repositorios académicos y de producción científica para la selección de trabajos relevantes, tales como JSTOR, WOS, Scielo, etc.

En segundo lugar, se analizaron diversas experiencias internacionales relacionadas, donde se tuvo como primera referencia a aquellas experiencias analizadas en el estudio “Disponibilidad de información con origen en instituciones públicas sobre productividad, desarrollo económico e innovación” (Price Elton, 2016). En este trabajo se consideraron las siguientes experiencias de buenas prácticas a nivel internacional:

- UK Data Service, Reino Unido.
- Centre for European Economic Research (ZEW).
- Research Data Centre de Alemania.
- The Boston Census Research Data Center (BRDC), Estados Unidos.
- Research Data Centres Program (RDC), Canadá.

Se complementaron estos casos de base por medio del análisis de casos de buenas prácticas en aquellos países del Asia Pacífico que durante las últimas décadas han emergido como experiencias líderes a nivel internacional. Entre otros casos, se analizarán las siguientes experiencias:

- Korean Social Science Data Archive (KOSSDA), Corea.
- Korea Open Research Data (KORD Repository), Corea.
- China Data Institute.
- China Data Lab.
- National Bureau of Statistics, China.
- Japan Data Repository Network (JDARN).

5. Casos internacionales y nacionales de uso de datos

También se incluyó la experiencia en América Latina al respecto, en particular las desarrolladas desde la Comisión Económica para América Latina (CEPAL), el Banco Interamericano de Desarrollo (BID), la Iniciativa Latinoamericana de Datos Abiertos (ILDA) y distintas experiencias nacionales, como es el caso del portal de datos abiertos de la República del Perú (datosabiertos.gov.pe), Argentina (datos.gov.ar), Colombia (datos.gov.co), México (datos.gov.mx) y Brasil (datos.gov.br).

Respecto de las experiencias nacionales, se consideraron los siguientes repositorios de acceso público:

- **Servicio de Impuestos Internos:** registro histórico de personalidades jurídicas, serie 2005-2019.
- **Servicio Nacional de Aduanas:** registro histórico de estadísticas de comercio internacional (Importaciones y Exportaciones).
- **Instituto Nacional de Propiedad Intelectual (INAPI):** registro histórico de patentes.
- **Datos.gob:** depósito de datos públicos generales.
- **Energía Abierta:** depósito de datos referidos a energía y medio ambiente.
- **Economía Abierta:** depósito de datos referidos a dimensiones del desarrollo y la actividad económica.
- **ANID:** datos sobre productividad científica, tecnológica y becas relacionadas con la oferta programática de la Agencia Nacional de Investigación y Desarrollo.
- **Data Ciencia:** dashboard de consulta de información referida a productividad científica.
- **Data Emprendedor:** portal para el acceso a información sobre actividad emprendedora para beneficiarios de Startup Chile.

5. Casos internacionales y nacionales de uso de datos

Luego del levantamiento de información, se revisó el marco conceptual, documentación y materiales disponibles respecto al uso de bases de datos, identificando elementos claves para la categorización y descripción del material revisado, a través de la sistematización de información documental existente. La información fue sistematizada a partir de una matriz de registro del contenido que permitió identificar características claves de cada una de estas, además de poder agruparlas según características específicas¹³.

Finalmente, se llevó a cabo un método de comparación, mejoramiento y definición de mejores prácticas –benchmarking- (Meade, 1998), por medio del cual se identificaron las experiencias más exitosas e innovadoras en cuanto al uso de datos públicos, las cuales se compararon con el caso chileno.

13. Para mayor información, consultar Anexo 4 de este informe.

5.2. Análisis de información secundaria¹⁴

En lo que concierne a la experiencia en investigación. A continuación, se expone el resumen del análisis de la ficha de registro y sistematización.

Tabla 4. Análisis de la experiencia en investigación

Año

- Los trabajos datan del año 2011 en promedio, concentrados en los tramos 2002-2009 y 2013-2019, lo que corresponde al lapso considerado en la línea de base.

Clasificación Temática

- Los temas recurrentes se asocian a los factores que determinan el desarrollo de actividades de I+D a nivel de empresas, asimismo como sus efectos sobre la productividad. Consideraciones respecto de las variables que permiten medir la capacidad exportadora de las empresas.
- Entre los factores determinantes se identifica la estructura de costos de las empresas y otras características relacionadas con el gasto en personal y capital humano.
- Entre las consecuencias de las actividades de I+D se consideran los beneficios sociales, los efectos de derrame, efectos sobre economías de escala y ámbito, efectos sobre la capacidad exportadora y otros efectos secundarios.
- Seis de estos trabajos abordan temas relacionados a la colaboración en actividades de I+D desde una perspectiva clásica. Emergen enfoques basados en teoría de grafos a nivel teórico. Lo anterior es visto desde el punto de vista de los incentivos a colaborar con distintos agentes públicos y/o privados, como desde el punto de vista de los efectos de la colaboración sobre variables de resultado asociadas a actividades de I+D.

14. Cabe señalar que la sistematización completa de la información secundaria levantada fue entregada a contraparte en dos archivos en formato Excel.

5. Casos internacionales y nacionales de uso de datos

- Evaluación de impacto de subsidios a actividades intensivas en conocimiento, tanto a nivel de empresas como de emprendimientos.
- Caracterización de emprendimientos según sus dimensiones organizacionales, sectoriales y de acceso a redes dentro del ecosistema.
- En temas relacionados a financiamiento, emerge el tratamiento de los factores que determinan el acceso a instrumentos corporativos tales como el CVC.
- Ausencia de temas relacionados al tratamiento empírico de la estructura organizacional y confirmación de redes al interior de las instituciones proveedoras de conocimiento y al interior del sector público.

Pregunta de Investigación e hipótesis

- En general, los trabajos se orientan hacia la identificación de efectos causales, ya sea desde variables de innovación sobre la productividad, o desde la productividad hacia variables de innovación. Esto permite dotar a los trabajos de un cuerpo de hipótesis en el marco de un enfoque de análisis confirmatorio, es decir, donde se testean supuestos empíricamente para identificar no sólo la magnitud y sentido con el que las variables independientes ejercen su efecto sobre la dependiente, sino también para la identificación de la dirección de estos efectos.
- También hay un grupo de cuatro trabajos de orientación exploratoria, es decir, no están orientados por hipótesis como es el caso del enfoque propio del análisis confirmatorio, por tanto, dichas investigaciones buscan entregar elementos que aporten a la discusión desde una perspectiva teórica sin especificar la obtención de las fuentes de datos obtenidas ni tampoco hacer inferencias sobre algún país en particular o empresas determinadas.

5. Casos internacionales y nacionales de uso de datos

Métodos

- El 40% de los paper analizados utilizan métodos de regresión como su técnica principal, aunque con distintos énfasis, dependiendo de la presencia de hipótesis.
- Entre los métodos de regresión más utilizados se encuentra el modelo lineal y los modelos de elección discreta (logit y probit). También se consideran modelos para datos de panel con un enfoque cuasi experimental, en particular para el testeo de efectos causales, donde se consideran modelos de diferencias en diferencias o modelos de efectos fijos/aleatorios.
- Algunos trabajos implementan modelos de Propensity Score Matching para selección de muestras comparables.
- Existencia de modelos para el análisis de redes basados en teoría de grafos, que permiten estimar indicadores de estructura y morfología de las redes de colaboración.
- Presencia marginal de metodologías cualitativas (1 paper de 2018, donde el investigador principal es José Guimoná), donde mediante entrevistas con actores claves se busca medir la importancia de las políticas públicas para atraer IED relacionado con la I+D en países emergentes.
- Ausencia total de metodologías relacionadas con aprendizaje de máquina (machine learning), mostrando una tendencia a utilizar mecanismos de análisis tradicionales en los entornos de innovación, generándose una brecha respecto a las nuevas corrientes de investigación que utilizan diversas disciplinas al analizar rendimientos o segmentación del ecosistema según sus distintos actores.
- Ausencia total de tratamiento de datos no estructurados, en particular datos de texto.
- Metodologías emergentes en el análisis de redes, aunque estos trabajos aún tienen un énfasis teórico que empírico. Las fuentes de datos de los trabajos empíricos están asociadas a datos privados de baja escala, es decir, se trata de fuentes que no son de acceso público, independientemente de que hayan sido generadas en instancias de investigación fomentadas y financiadas por agencias públicas.

Unidades de Análisis

- Fundamentalmente empresas, seguidas de emprendimientos, instituciones proveedoras de conocimiento como universidades y centros de investigación.

5. Casos internacionales y nacionales de uso de datos

Variables Consideradas

- Generalmente se trata de variables de stock tales como gasto y personal en I+D, medidas convencionales de productividad a nivel de empresas, tipos de innovación según clasificadores propuestos por el Manual de Oslo y tipos de actividades de I+D según clasificadores propuestos en el Manual de Frascati.
- También se incluyen variables de segmentación para el tamaño de la empresa, sector de actividad económica y localización geográfica.
- Variables emergentes relacionadas a co-laboración en actividades de I+D.
- Variables relativas a las características de la estructura organizacional a nivel de empresas y emprendimientos.

Fuentes de Datos | Fuentes Públicas

- Predominan el uso de encuestas longitudinales y registros administrativos a nivel de empresas.
- Predominan las encuestas en el caso del análisis de los emprendimientos.
- Sólo 4 papers empíricos no utilizan fuentes públicas (hay dos papers teóricos), el resto acude a censos industriales, encuestas de innovación e I+D, registros administrativos y datos derivados de instrumentos públicos.

Fuente: Elaboración propia.

La información recogida de los *papers* analizados nos presenta que la técnica de análisis más utilizada para investigar el entorno de I+D corresponde a modelos de regresión con un 40% de los casos, principalmente para ver la relación entre variables cuantitativas con el impacto estas tienen sobre variables relacionadas al rendimiento empresarial en I+D. Por otro lado, no se registran trabajos que ocupen técnicas de *machine learning* que permitan trabajar con estrategias de modelamiento más sofisticadas, como es el caso de la detección de patrones cuando se consideran una gran cantidad de variables para describir a las industrias analizadas, o la utilización de redes neuronales para observar la capacidad que tiene una industria de enfrentar de buena manera el cambio tecnológico.

Hay un conjunto de tres investigaciones que busca analizar la importancia de las redes sociales en los procesos de emprendimiento para determinar si mediante conexiones que tengan dichas empresas obtienen un beneficio adicional en su posterior desarrollo, o si su desempeño está asociado a otros factores de naturaleza endógena.

5. Casos internacionales y nacionales de uso de datos

Las investigaciones nacionales están concentradas principalmente en el análisis del rendimiento de las políticas públicas locales, sobre los resultados a posteriori sobre los procesos de innovación en empresas de todos los tamaños. Mientras que los *papers* extranjeros no se concentran en analizar políticas públicas específicas, más bien investigan elementos que podrían ser transversales a cualquier país, como el caso de las redes de apoyo de los emprendimientos y las redes de vínculos de los emprendedores, señalando que muchas veces se comparten dichos lazos, pero no siempre es el caso. También se destaca el rol que juegan las universidades y otras instituciones proveedoras de conocimiento en los procesos de innovación en las empresas. De igual manera la literatura internacional se orienta al análisis de procesos basados en cambios tecnológicos, efectos derrame, departamentalización de las empresas o la relación entre entes privados y públicos.

El 80% de los *papers* considera como unidad de análisis a las empresas de manera aislada, mientras que el restante 20% analiza interacciones de empresas (Grandes, Pequeñas, Medianas) con otros actores que podrían intervenir en el proceso de desarrollo como lo son las universidades en el caso francés, o los organismos intermedios (públicos-privados) como lo fue en el caso mexicano.

El 60% de las investigaciones analizadas trabaja con datos de acceso público, ya sean de alguna institución pública o que estén en un formato de datos abiertos (para la descarga de cualquier persona). En el caso de los datos de CORFO, la información que se utiliza es generada por una agencia pública, pero reservada ya que no todos los datos obtenidos son de acceso público siendo necesario llegar a un acuerdo con la institución para su uso. El otro 40% de las investigaciones se divide en una revisión teórica, conceptual o levantan sus propios datos para poder realizar los análisis que desarrollan.

Las fuentes de datos extranjeras se concentran en el Ministerio de Economía y en el Ministerio de Estadística, mientras que para el caso chileno CORFO es la principal fuente de información junto al Servicio de Impuestos Internos (SII) y el Instituto Nacional de Estadística (INE).

Por su parte, el análisis de los repositorios de datos abiertos fue de mayor complejidad dada su naturaleza heterogénea. En efecto, en cada uno de ellos se encontró una variedad importante de contenidos, formas de organización de recursos, sitios relacionados, etc.

Tabla 5. Análisis de Repositorios

Objetivos

- El objetivo principal que comparten los repositorios analizados consiste en apoyar las investigaciones con el uso de datos para el avance de la ciencia local y con ello incentivar a la masificación de su uso.
- Existen repositorios con usos específicos o que detallan quienes los pueden utilizar, uno de estos repositorios específicos (zew) señala claramente que el uso de los datos que la institución puede facilitar es de uso investigativo/académico y no pueden ser utilizados con fines comerciales bajo ningún punto de vista.
- Mientras que el repositorio de UK y Energía Abierta (Chile), señalan abiertamente que el uso de sus datos puede ser para proyectos de investigación, para el uso ciudadano o para proyectos privados que tengan las empresas que requieran contar con información oficial para realizar operaciones en el país.
- Esta aclaración cobra importancia porque la mayoría los repositorios de datos abiertos declaran que su uso puede ser para cualquier persona, sin hacer especificaciones.

Instituciones Participantes

- Todos los repositorios analizados incluyen información proveniente de datos públicos, ya sea a nivel de ministerios o de países miembros como es el caso de CEPAL y Eurostat.
- No todos los repositorios incluyen datos a nivel de empresa (recolectados por instituciones públicas), dentro de los que si incluyen destacamos al Servicios de Impuestos Internos (SII) y EL ZEW de Alemania ya que permiten descargar datos del comportamiento económico a nivel empresarial

Acceso

- La mayoría de los repositorios cuentan con acceso a Datos Abiertos, los repositorios latinoamericanos tienen la particularidad de que no es necesario inscribirse para poder acceder a los datos que tienen en línea, mientras que en el caso de los repositorios europeos y asiáticos es necesario registrarse para descargar la información que es requerida.
- En el caso de China, es necesaria una acreditación de alguna universidad local o alguna que tenga convenio con el Gobierno Chino, como es el caso de la Universidad de Harvard.

5. Casos internacionales y nacionales de uso de datos

Seguridad

- La mayoría de los repositorios entregan los datos de manera agregada y en caso de que la unidad de análisis sea una empresa o persona, la información viene anonimizada para que no sea posible identificar dicho individuo, incluso si se cruzan e integran distintas bases de datos.
- Un caso que no tiene los datos anonimizados corresponde al Servicio de Impuestos Internos de Chile (SII) que entrega elementos que nos permiten identificar de manera individual a las empresas que están incluidas en la base de datos mediante el RUT.

Formato de Descarga

- La mayoría de los repositorios (menos los chinos y dataciencia) permiten descargar sus datos en CSV o formato Excel (xlsx), también hay repositorios que adicional a dichos formatos incluyen la descarga de PDF, MP3, JPG según el formato original de la información.
- Se destacan los repositorios que contienen información de data no estructurada como es el repositorio canadiense y el chino (Data Institute).

Fuentes estructuradas – Encuestas

- Las principales encuestas que se utilizan en los repositorios de datos corresponden a empresas y encuestas a nivel de hogares para la caracterización social de la población.

Fuentes estructuradas – Registros administrativos

- Los repositorios analizados incluyen datos de las empresas gubernamentales e instituciones públicas como ministerios, siendo la principal fuente de datos.

Fuentes estructuradas – Censos

- No todos los repositorios incluyen información censal, los repositorios orientados a las Ciencias Sociales incluyen dicha información como también quienes son el repositorio oficial del país como data Colombia.
- Los restantes repositorios (37%), no incluyen información censal o no es informada en sus tipos de datos disponibles.

Fuentes no estructuradas

- El 40% de los repositorios analizados incluye información de fuentes no estructuradas como audios, mapas, texto y fotografías.

Fuente: Elaboración propia.

5. Casos internacionales y nacionales de uso de datos

Los repositorios consultados tienen en común la característica de que están orientados al trabajo de investigadores, estudiantes y profesores (universitarios y escolares). Son organismos de consulta académica que buscan incentivar el trabajo con datos para la investigación, principalmente enfocados al ámbito de ciencias sociales y económicas que ayuden a responder temáticas de interés nacional, tanto desde el punto de vista de la productividad, como desde el punto de vista de la equidad y sustentabilidad ambiental.

Repositorios tales como el de UK y Chile (Energía Abierta), señalan abiertamente que el uso de datos abiertos genera un valor agregado al mundo empresarial, si bien existen otros repositorios de datos que contienen información de carácter empresarial, solamente estos repositorios (UK y Energía Abierta) declaran públicamente que la contenida es de utilidad para la toma de decisiones en contextos de negocio. En este caso, los datos dispuestos al acceso abierto son entendidos como activos estratégicos, cuyo valor es puesto para orientar la toma de decisiones en ambientes empresariales, mostrando con esto la versatilidad de los datos liberados sin restringirlos únicamente al sector académico.

En general los repositorios analizados contienen datos de diversas fuentes, que abarcan las principales encuestas y censos tanto a nivel de hogares, personas como de establecimientos. Entre los establecimientos considerados, se trata generalmente de encuestas y censos a empresas tanto a nivel general como sectorial, establecimientos educativos y de salud. También es posible encontrar registros administrativos para empresas, tanto en temas relacionados a la actividad exportadora, variables de caracterización de personal y de actividad económica.

Emergen repositorios que incluyen datos no estructurados tales como entrevistas, *focus group*, archivos multimedia. Este es un hallazgo relevante, debido a que podemos obtener la información sin procesar, para que de ese modo el usuario pueda hacer sus propios análisis, como por ejemplo con los archivos de audio de las entrevistas realizadas o grabaciones de los *focus group*, entregando la posibilidad de llegar a otros resultados con los mismos datos, ya que será el criterio del investigador qué elementos se considerarán como importantes para el respectivo análisis, tal y como es el caso del repositorio KOSSDA (Corea del Sur), que en los datos que almacena se consideran entrevistas. Generalmente los repositorios están orientados hacia la gestión y disposición de datos estructurados, lo que

5. Casos internacionales y nacionales de uso de datos

permite abrir una línea de exploración para el caso nacional en relación con aquellas fuentes de datos no estructurados relacionados con el contenido de llenado en formularios, asimismo como es el caso de los productos y anexos generados en el contexto de investigaciones académicas.

Por otro lado, una forma de clasificar los distintos tipos de datos que se encuentran en los repositorios anteriormente mencionados y según su accesibilidad permite definir tres categorías principales.

- **Datos Abiertos:** Accesible para cualquier usuario, no es necesario registrarse
- **Datos Protegidos:** Solamente para usuarios debidamente registrados.
- **Datos de acceso cerrado:** Requiere una autorización especial del dueño de los datos o de la institución almacenadora de la información.

La información que poseen estos servicios de datos proviene principalmente de universidades (centros de estudios) y organismos públicos siendo estos últimos la red que proporciona la información de carácter fiable y estable.

Instituciones como el *Datacenter* de UK contribuyen de manera recurrente en la edición del “libro blanco”, documento que detalla los trabajos realizados por agencias e instituciones públicas, y que tiene la finalidad dar a conocer a la opinión pública los avances en determinadas materias. Por ejemplo, en Chile cada 10 años el Ministerio de Defensa publica su “libro blanco” con la información de lo que se ha realizado en los últimos 10 años de gestión. En el caso británico, este documento es utilizado para conocer el estado de avance de las investigaciones académicas que se están realizando. Por otro lado, el servicio alemán ZEW está orientado al análisis para la evaluación de las instituciones económicas europeas.

El repositorio de datos de Corea del Sur (*Dataon*) señala que la institución responsable de dicho proyecto (kisti), está enfocada tanto en el desarrollo de la I+D a nivel nacional, como en dinamizar el ecosistema científico tecnológico, el que además considera en su misión la creación de un ecosistema nacional de supercomputación.

5. Casos internacionales y nacionales de uso de datos

Respecto a la seguridad y privacidad de los datos, los repositorios dependientes de instituciones europeas señalan de manera clara el carácter confidencial de los datos que son proporcionados, por lo que los datos están debidamente anonimizados tanto a nivel personal como industrial, es decir, para que no sea posible identificar a una persona o institución o establecimiento a través de la intersección de variables que permiten la individualización de estas unidades de análisis. Lo anterior se ajusta a las normativas europeas de protección de datos personales e industriales, para aquellas instancias en las que la infraestructura de almacenamiento se encuentre en territorio de la UE.

La información referida a la seguridad de los datos no es lo suficientemente clara en los repositorios de Asia y Estados Unidos. Para el caso de los repositorios nacionales consultados, podemos señalar que la información disponible para la caracterización del ecosistema de CTCI del país está distribuida en distintos centros de datos. Si bien datos.gob es el principal depósito de datos públicos, gran parte de la información requerida para una mejor caracterización del ecosistema no se encuentra en dicho sitio web o, en caso de estar disponible, la información se encuentra desactualizada.

Los repositorios de datos chilenos y latinoamericanos responden a las propiedades establecidas para estándares de gobierno abierto, vale decir, para facilitar y democratizar su acceso y uso. Estos repositorios, además de proporcionar el acceso a datos sobre ciencia e innovación, también incluyen amplios sets de datos sociales en encuestas tales como la CASEN o la ENE, asimismo como a nivel del Censo de Población y Vivienda. En relación a los estándares de seguridad, cada país se rige por su ley de protección de datos o de estadística, no logrando tener un criterio común respecto a anonimización de los datos a nivel de empresas y personas.

En los repositorios nacionales es posible acceder a datos provenientes del Servicio de Impuestos Internos, Aduanas y ANID, los que contemplan información relevante para investigaciones temáticas, tanto a nivel de empresas como de emprendimientos, pero el acceso a dichas bases de datos se debe realizar por cada organismo de manera independiente, por tanto, no están en un repositorio común que permita obtener esa información desde una fuente de datos.

Siguiendo este mismo ejemplo, en Chile existe un repositorio orientado exclusivamente a emprendimientos (DataEmprendedor), el que pertenece a CORFO y cuenta con la colaboración de la Pontificia Universidad Católica

5. Casos internacionales y nacionales de uso de datos

de Chile. Dicho repositorio no está integrado a ninguna de las otras bases de datos señaladas anteriormente, o vinculada a alguna de las páginas web de dichos organismos, generando que no sea una base de datos que cuente con interoperabilidad para poder cruzar los datos con otras fuentes de información, tales como Aduanas o el Servicio de Impuestos Internos, pero que permite explorar el capital de emprendimiento nacional, y como podría vincularse con aspectos relacionados con – por ejemplo - la oferta programática de ANID, ya que se podría analizar si estos emprendimientos han recibido alguna ayuda estatal por parte de la institución, como por ejemplo, si alguno de sus impulsores ha estudiado con Becas Chile o si tiene en su portafolio alguna adjudicación de fondos para la innovación e investigación local. Esta base de datos eventualmente podría ser interoperada con la información de patentes de INAPI, o la actividad comercial registrada por el Servicio de Impuestos Internos, siempre y cuando sea CORFO la instancia que guíe este proceso para que dichas bases de datos puedan ser interoperables mediante un proceso de integración y con ello medir el impacto de determinados programas o políticas públicas que buscan incentivar la innovación en Chile.

5.3. Benchmarking

Considerando los criterios para el establecimiento de la línea de base y primeras brechas de las fuentes de información consideradas en el trabajo de Price Elton, la situación relativa al uso nacional de fuentes de datos puede ser entendida tanto desde el análisis de la experiencia en investigación, como del análisis de los distintos repositorios considerados.

En relación a la experiencia nacional en investigación, en términos generales el estado del arte está alineado con los estándares internacionales tanto a nivel de las clasificaciones temáticas y el uso de fuentes públicas para la implementación de metodologías de alto nivel. Lo anterior está alineado con la línea de base propuesta, ya que se trata de variables e indicadores de stock a nivel de empresas con distinto nivel de detalle, donde destaca el uso de las encuestas del INE y MINECON.

Por otro lado, se consideran indicadores de evolución para aquellas variables de caracterización de las empresas que provienen de los datos del SII, tal y como es el caso de la investigación de Arraíz (2012), en el cual se analiza el impacto del programa de Desarrollo de Proveedores en Chile, concluyendo que la participación en dicho programa implica una mejora integral en las empresas participantes, donde según el tamaño de los proveedores-empresas, se reportan beneficios en un lapso que va entre uno y dos años. Las bases de datos del SII pueden ser consideradas como la principal y más importante fuente de datos públicos a nivel de empresas, ya que tal y como se señala en el trabajo de Price Elton, permite la identificación de cada empresa a nivel de RUT y año comercial. No obstante, los datos de acceso público sólo consideran variables agregadas, y no el detalle de los formularios relacionados al proceso de tributación.

Cabe señalar que, para el caso nacional, no se encontraron trabajos orientados a enfoques emergentes como el análisis de redes aplicado sobre los agentes del ecosistema, lo que es coherente con la brecha identificada a nivel de indicadores de conexión. La única investigación en tratar este tema, sobre los agentes implicados en el proceso de innovación lo hace mediante el análisis tradicional de contenido en entrevistas a agentes clave dentro del ecosistema (Gimona et. al. 2018).

5. Casos internacionales y nacionales de uso de datos

Otro hallazgo de interés se relaciona con que las características de las instituciones proveedoras de conocimiento, los emprendimientos y el sector público son de escasa presencia en los trabajos seleccionados. En particular, no se consideran aspectos relacionados a la dotación de personal en este tipo de agentes, mientras que existen importantes fuentes de datos de acceso abierto para ello. Del mismo modo, estas consideraciones permitirían eventualmente contar con datos relativos a la estructura organizacional para estos agentes, y de esa manera abrir el espectro del análisis y caracterización del ecosistema.

Para el caso de las instituciones proveedoras de conocimiento, tanto el Consejo Nacional de Educación (CNED) como el Sistema de Información de Educación Superior (SIES) disponen de datos detallados para la caracterización de la dotación de docentes según su régimen de vinculación, asimismo como de los aspectos relacionados con infraestructura y equipamiento. Este será un aspecto para profundizar en las etapas que siguen del presente trabajo.

Respecto del análisis de los repositorios nacionales, si bien en el trabajo de Price Elton se definen temáticas y preguntas de frontera que pueden ser abordadas por las fuentes de datos analizadas en ese trabajo, se trata de la profundización en el ámbito de fenómenos contables (comercio internacional y tributación) por un lado, y de los elementos primarios de la actividad innovativa de las empresas, sin considerar a los demás agentes dentro del ecosistema. Por otro lado, las fuentes consultadas están referidas únicamente a las empresas, sin considerar al sector público, instituciones proveedoras de conocimiento, emprendimientos y *hubs* de innovación.

En relación con las políticas de acceso, se constata que los repositorios nacionales cuentan con datos y microdatos abiertos, y que permiten la implementación de aplicaciones ya que cuentan con servicios de API, los que requiere de la obtención de credenciales de acceso de manera expedita para desarrolladores.

A su vez, se identifican repositorios que no son utilizados en las experiencias de investigación analizadas en el presente trabajo, como es el caso de ChileCompra. En efecto, dicho repositorio permite la exploración de elementos relativos a proyectos y licitaciones relacionadas con el ecosistema, tanto a nivel de datos estructurados (fechas, montos, etc.) como de datos no estructurados. Lo anterior es relevante ya que los campos de registro de texto constituyen un potencial de información a explorar para la caracterización temática de proyectos y licitaciones, para de esa manera contar con posibilidades de implementación de indicadores de conexión que permitan abrir el abanico de las dimensiones consideradas para la caracterización del ecosistema.

5. Casos internacionales y nacionales de uso de datos

Otro elemento crítico se relaciona con el repositorio de la ANID, ya que no contempla la posibilidad de exportar los datos históricos de cada uno de los proyectos y programas que constituyen su oferta, con lo cual se identifica una brecha importante de acceso una de las principales fuentes de datos para la caracterización del ecosistema.

De este modo, en términos generales, el uso de datos derivados de los sistemas de registro de la oferta programática del Estado constituye la principal brecha relativa a las fuentes de datos para la caracterización del ecosistema. Estas fuentes son fundamentales para medir temas emergentes tales como la colaboración en actividades y la conformación de redes, ya que los formularios de postulación y los datos de adjudicación registran entidades ejecutoras y coejecutoras. Del mismo modo, estas fuentes contienen datos no estructurados relacionados a la descripción de iniciativas que permiten la exploración de indicadores no convencionales para el desarrollo de actividades de investigación orientadas a un mejor diagnóstico y evaluación de la política pública en la materia.

Es así como en el presente trabajo se explorarán fuentes relacionadas con los sistemas de registro de la oferta programática de las entidades públicas relacionadas. En efecto, la diversificación y cobertura crecientes de la oferta de programas públicos para el fomento de actividades intensivas en conocimiento presenta una enorme oportunidad de exploración de fenómenos emergentes a nivel de empresas, emprendimientos e instituciones proveedoras de conocimiento.

6. Mapa de fuentes de información vinculadas a CTCI

6. Mapa de fuentes de información vinculadas a CTCI

6.1. Metodología

El objetivo del mapa de fuentes de información vinculadas a CTCI es identificar en las entidades públicas (Ministerios, Agencias, Institutos, entre otros) qué información tradicional existe en sus repositorios de datos, los cuales permitan caracterizar el ecosistema CTCI (considerando todas las fuentes de información disponibles y sus características), como por ejemplo: Encuestas de Innovación en Empresas, Encuesta de Personal y Gasto en I+D, Presupuesto Públicos para I+D (GBARD), GEM, Encuesta Nacional de Percepción Social de la Ciencia y la Tecnología, etc.

El proceso de diseño e implementación del mapa contempló dos etapas

En la primera etapa, se tomó como referencia de base los resultados del estudio de Price Elton (2016)¹⁵. En efecto, la sistematización realizada entre las fuentes relacionadas al Servicio de Impuestos Internos, el Servicio Nacional de Aduanas y algunos productos estadísticos basados en encuestas por muestreo (Encuesta Longitudinal de Empresas, Encuesta de Personal y Gasto en Innovación, Encuesta Industrial Anual) permitió contar con elementos de integración de datos a nivel de empresas, sobre los cuales es posible comenzar a desarrollar las vinculaciones con el resto de los datos generados en y por los agentes del ecosistema de CTCI.

Por otro lado, esta primera versión del mapa consideró fuentes de datos tradicionales complementarias, en cuyas plataformas era posible el acceso a archivos descargables, ya que esta opción permite acceder a mayor información para la caracterización de la fuente.

15. Estudio "Disponibilidad de información con origen en instituciones públicas sobre productividad, desarrollo económico e innovación" desarrollado por la consultora Price Elton para el CNID durante el año 2016.

6. Mapa de fuentes de información vinculadas a CTCI

Las fuentes complementarias que corresponden a sistemas de registro para esta primera versión del mapa son:

- Instituto Nacional de Propiedad Intelectual (INAPI): tabla de marcas y patentes.
- Sistema de Información de Educación Superior (MINEDUC-SIES): caracterización de instituciones de Educación Superior.
- Base de Datos Índices, Consejo Nacional de Educación (CNED-Indices): caracterización de instituciones de Educación Superior.
- Web of Science (WoS): repositorio de trabajos científicos.
- SCielo: repositorio de trabajos científicos.
- Scopus: repositorio de trabajos científicos.
- DataEmprendimiento: datos abiertos de CORFO para beneficiarios del programa Startup Chile.
- Encuesta de trayectoria de profesionales con el grado de doctor (Ministerio de Economía).
- Encuesta Nacional de innovación en empresas (Ministerio de Economía).
- Encuesta Nacional de Tecnologías de la Información y Comunicación (Instituto Nacional de Estadísticas y Ministerio de Economía).
- Encuesta Nacional sobre Gasto y Personal en Investigación y Desarrollo (Instituto Nacional de Estadísticas y Ministerio de Economía).
- Créditos Presupuestarios Públicos para I+D (Instituto Nacional de Estadísticas y Ministerio de Economía).

Por otro lado, también se han considerado aquellas fuentes que corresponden a distintos tipos de encuestas relacionadas a la caracterización del ecosistema, tales como:

- Encuesta Nacional Industrial Anual (Instituto Nacional de Estadísticas).
- Global Entrepreneurship Monitor (GEM-Universidad del Desarrollo).
- Encuesta de Microemprendimiento (Instituto Nacional de Estadísticas y Ministerio de Economía).
- Encuesta Nacional de Percepción Social de la Ciencia y la Tecnología en Chile (Ministerio de Ciencia, Tecnología, Conocimiento e Innovación).
- Encuesta Longitudinal de Empresas (Ministerio de Economía).

6. Mapa de fuentes de información vinculadas a CTCI

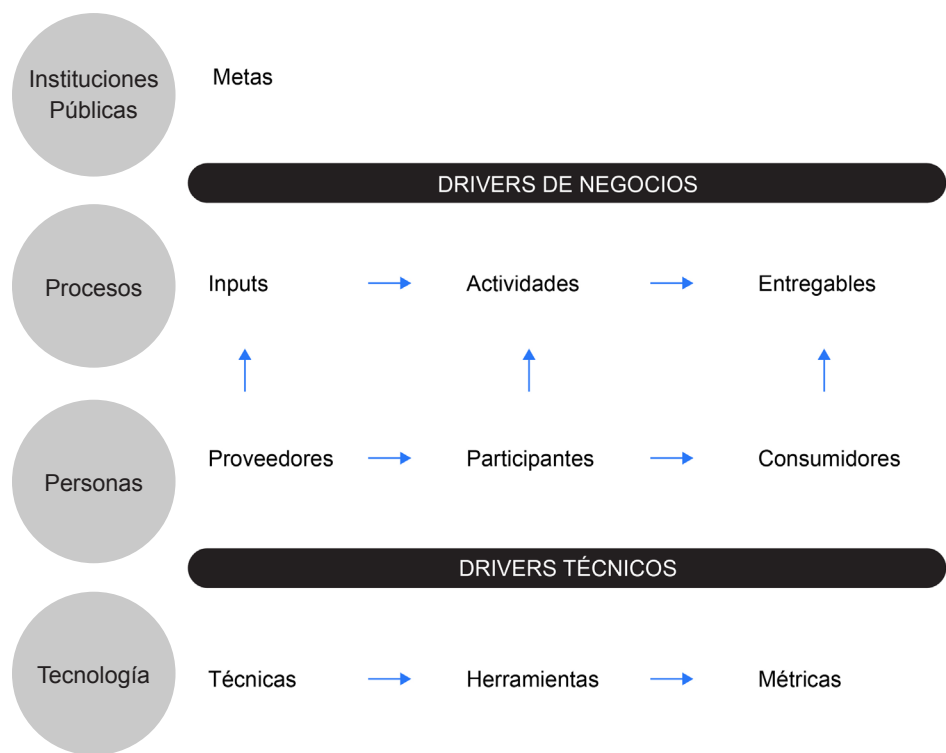
Las fuentes que no cuentan con esta posibilidad de acceder a productos descargables fueron objeto de diagnóstico y representación para la segunda versión del mapa, la que consideró la información cualitativa levantada a través del diagrama de contexto. En particular, estas fuentes se relacionan con sistemas de registros administrativos, distintos niveles de acceso parcial a usuarios externos tales como el Banco de Proyectos de la Fundación para la Innovación Agraria (FIA), el Sistema de gestión de proyectos de CORFO (SGP), el Sistema de gestión de información estadística de la Comisión de mercados financieros (CMF), entre otros.

Un primer elemento a considerar en el diagrama de contexto son los drivers, vale decir, elementos de contexto mayor que definen y acotan metas y objetivos dentro de la organización. Los drivers pueden distinguirse, por un lado, entre aquellos relacionados al dominio o negocio (en este caso, los relacionados a los objetivos de política y programas) donde el énfasis está puesto en las definiciones y metas de la organización (parte superior de la Figura 4). Por otro lado, podemos distinguir a aquellos drivers que se relacionan con elementos de orden técnico y tecnológico, donde el énfasis es en las técnicas y herramientas involucradas, asimismo como las métricas asociadas a cómo estas permiten el cumplimiento de metas y objetivos (parte inferior de la Figura 4).

Dentro de este marco, luego se consideran los distintos inputs de proceso relacionados a la fuente de información dentro de la organización, donde el rol de las personas es concebido como el de proveedores de distintos servicios relacionados para el desarrollo de actividades. A su vez, las actividades son clasificadas entre aquellas que son de planificación, desarrollo, operación y control de los procesos relacionados a cada fuente de información. En este sentido, el rol de las personas es el de participantes en estos procesos, sobre los cuales son relevantes las consideraciones propias de la estructura administrativa de los agentes considerados dentro del ecosistema.

6. Mapa de fuentes de información vinculadas a CTCI

Figura 3. Diagrama de Contexto para el diagnóstico de fuentes de datos



Fuente: DMBOK, 2017.

Se consideran los entregables de los procesos, los cuales deben ser definidos en virtud de las necesidades de las áreas/personas involucradas como consumidores (pueden ser clientes internos o externos a la organización).

Para homologar la información comprendida en el diagrama de contexto, se elaboró una ficha tipo que identifique a cada actor la presencia o ausencia de los elementos que se espera encontrar e identificar. Entre los elementos a incluir en la ficha se incluyen los siguientes: ubicación, disponibilidad, protocolos de seguridad asociados, formato, calidad de los datos contenidos, periodicidad de actualización, permisos de acceso, disposición a compartir. Además, la ficha incluirá descripciones que posteriormente permitirán realizar adicionalmente perfiles por actores y otras características destacadas que den paso a la caracterización del ecosistema CTCI.

6. Mapa de fuentes de información vinculadas a CTCI

6.2. Diagnóstico de fuentes que componen al mapa

Las fuentes fueron organizadas en función de los actores del ecosistema que están incluidos en ellas, lo que para esta versión implica los siguientes segmentos:

Establecimientos formales: fuentes que incluyen a todo tipo de actor formalmente establecido, lo que incluye a empresas, emprendimientos, agencias del sector público e instituciones proveedoras y difusoras de conocimiento.

Tabla 6. Diagnóstico de fuentes para establecimientos formales.

Institución Encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
SCielo	Metadatos	<p>Tipo: Registros administrativos.</p> <p>Adquisición: ingreso de documentos relativos a trabajos científicos/solicitudes de inscripción de propiedad intelectual en repositorio relativo cada sistema de indexación.</p>	<p>El conjunto de tablas relacionadas refieren a:</p> <p>Papers: identificación y contenido de los trabajos publicados.</p> <p>Autores: personas involucradas en la autoría de la publicación.</p>	<p>Las tablas pueden relacionarse mediante el nombre del autor, nombre de la institución, y clasificación del paper.</p>	<p>Diccionarios disponibles con la descripción de cada tabla.</p> <p>Esta fuente considera contenido no estructurado, lo que implica procedimientos de estructuración de contenidos para la generación de metadatos dinámicos.</p>	<p>Cobertura:</p> <p>Transversal: universo de publicaciones para cada sistema de indexación.</p> <p>Longitudinal: el inicio de la serie depende de la antigüedad de cada journal, asimismo como de su digitalización.</p>
Scopus	Metadatos	<p>Periodicidad: tiempo real ya que la actualización sincroniza las agendas del total de Journals, mientras que para INAPI la periodicidad es anual.</p>	<p>Nombre institución: afiliación institucional de cada autor involucrado</p> <p>Área temática: clasificación de la publicación.</p>			
WoS	Metadatos	<p>Formato: permite obtener distintos formatos, desde datos de panel hasta redes.</p>	<p>Cada repositorio permite la representación de datos organizados en redes (grafos) ya que cada trabajo permite el registro de más de un agente, lo que permite obtener indicadores de conexión relacionados a la colaboración entre agentes a nivel de autores e instituciones.</p>			
INAPI	ION		<p>Conjunto de tablas relacionadas que refieren a:</p> <ul style="list-style-type: none"> • Marcas • Patentes 	<p>Las tablas pueden relacionarse con el número de inscripción de la solicitud, el nombre de la institución y nombre de los aplicantes.</p>		<p>Cobertura:</p> <p>Transversal: universo de solicitudes de inscripción de propiedad intelectual.</p> <p>Longitudinal: la serie inicia el año 2014.</p>

6. Mapa de fuentes de información vinculadas a CTCI

Institución Encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
SII	Estadísticas de Empresa	<p>Tipo: Registros administrativos.</p> <p>Adquisición: Llenado obligatorio de formularios para personalidades jurídicas.</p> <p>Periodicidad: Anual o mensual dependiendo del formulario</p>	<p>Conjunto de tablas relacionadas mediante esquemas relacionales, referidas a:</p> <ul style="list-style-type: none">Formulario 22Formulario 29Libros de comprasLibros de ventasDeclaración jurada de renta No. 1887Declaración jurada de renta No. 1879	Todas las tablas pueden relacionarse mediante el RUT de la empresa	Formularios disponibles para la exploración de campos.	<p>En ambos casos la cobertura transversal considera al universo de establecimientos formalmente establecidos, pero para SNA se considera sólo la actividad importadora/exportadora</p> <p>La cobertura longitudinal permite una trazabilidad de largo plazo, para SNA desde 1991 y para SII desde 1993</p>
SNA	Estadísticas COMEX		<p>Conjunto de tablas relacionadas mediante esquemas relacionales, referidas a:</p> <ul style="list-style-type: none">ImportacionesExportaciones		Esta fuente está estructurada a nivel de cada partida de exportación/importación, por lo que el catálogo de metadatos considera a la empresa como un atributo y no como una unidad de análisis	

Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Empresas: fuentes que sólo operan a nivel de empresas (la única excepción incluida es el Catastro de Créditos Presupuestarios Públicos para la I+D).

Tabla 7. Diagnóstico de fuentes para Empresas.

Institución encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
MINECON	Encuesta-Trayectoria de profesionales con el grado de doctor	<p>Tipo: levantamiento secundario administrativo</p> <p>Adquisición: actualización del directorio de doctores chilenos y extranjeros residentes en Chile</p> <p>Periodicidad: no informada</p>	Tabla unificada que permite caracterizar a los doctores residentes en Chile, tanto a nivel académico como trayectoria laboral	La tabla puede relacionarse mediante el RUT de los doctores (RUT no disponible para usuarios finales)	Diccionarios no disponibles con la información total de los metadatos obtenidos	Se realiza una encuesta online una vez que se ha limpiado la base de datos generada, eliminado los casos sin RUT, mayores de 70 años y repetidos
INE	Encuesta Nacional Industrial Anual (ENIA)	<p>Tipo: Encuesta por muestreo</p> <p>Adquisición: envío de encuesta por correo físico para llenado por parte de la empresa</p> <p>Periodicidad: anual</p>	Tabla unificada con los resultados de la encuesta para cada uno de los módulos y secciones del respectivo cuestionario	Las tablas pueden relacionarse mediante el RUT de la empresa con otras bases de datos. (RUT no disponible para usuarios finales)	Diccionarios disponibles con la descripción de los campos de cada una de las tablas	Es representativa a nivel nacional por sector económico, región y tamaño de empresa.
INE-MINECON	Encuesta TIC	<p>Tipo: Encuesta probabilística por muestreo estratificado.</p> <p>Adquisición: envío de encuesta por internet para llenado por parte de la empresa</p> <p>Periodicidad: no informada</p>	Tabla unificada con los resultados de la encuesta para cada uno de los módulos y secciones del respectivo cuestionario	La encuesta cuenta con un campo de RUT que permite identificar a la empresa con los registros del SII (RUT no disponible para usuarios finales)	Diccionarios disponibles con la descripción de los campos de cada una de las tablas	Las muestras tienen representatividad estratificada

6. Mapa de fuentes de información vinculadas a CTCI

Institución encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
INE-MINECON	Créditos Presupuestarios Públicos para I+D	<p>Tipo: Catastro</p> <p>Adquisición: catastro de Instituciones públicas y programas, en caso de no encontrar información envían encuesta a los encargados de la unidad</p>	Tabla unificada con la información desagregada de los montos que tienen las instituciones públicas para I+D	<p>Las tablas pueden relacionarse mediante el RUT de la institución pública.</p> <p>(RUT no disponible para usuarios finales)</p>	Diccionarios disponibles con la descripción de los campos de cada una de las tablas.	Es un catastro que tienes dimensiones censales, por lo que permitiría realizar estimaciones a nivel nacional y sectorial dentro de las distintas reparticiones dentro del sector público.
INE	Encuesta Nacional Industrial Anual (ENIA)	<p>Tipo: Encuesta por muestreo</p> <p>Adquisición: envío de encuesta por correo físico para llenado por parte de la empresa</p> <p>Periodicidad: anual</p>	Tabla unificada con los resultados de la encuesta para cada uno de los módulos y secciones del respectivo cuestionario	Todas las tablas pueden relacionarse mediante el RUT de la empresa	Diccionarios disponibles con la descripción de los campos de cada una de las tablas	<p>Las muestras tienen representatividad nacional, lo que no permite realizar estimaciones a nivel sectorial.</p> <p>Por otro lado, las muestras se han visto sujetas a cambios por la modificación de los directorios industriales.</p>
MINECON	Encuesta Nacional sobre Gasto y Personal en Investigación y Desarrollo	<p>Tipo: Catastro</p> <p>Adquisición: catastro de Universidades, Empresas, IPSFL (Ins sin fines de lucro), Instituciones estatales</p> <p>Periodicidad: no informada</p>				
MINECON	Encuesta Longitudinal de Empresas	<p>Tipo: Encuesta por muestreo</p> <p>Adquisición: envío de encuesta por correo físico para llenado por parte de la empresa</p> <p>Periodicidad: sin frecuencia definida</p>				

Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Instituciones de Educación Superior: se incluyen Universidades, Centros de Formación Técnica, Institutos Profesionales y FF.AA. En estas fuentes no se incluyen Institutos y/o Centros de Investigación adscritos a universidades.

Tabla 8. Diagnóstico de fuentes para Instituciones de Educación Superior

Institución encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
MINE-DUC	SIES	<p>Tipo: Registros administrativos.</p> <p>Adquisición: llenado de formulario por parte de unidades de análisis institucional y similares dependientes de las instituciones de educación superior (Universidades, Institutos Profesionales, Centros de Formación Técnica y FF.AA).</p> <p>Periodicidad: Anual.</p> <p>Formato: datos de panel.</p>	<p>Modelo de tablas relacionadas referidas a:</p> <ul style="list-style-type: none"> • Titulados de la Educación Superior. • Oferta Académica: composición y distribución de carreras y programas. • Matrícula: caracterización de la demanda por carreras y programas. • Personal académico: composición y distribución por persona y Jornada Completa Equivalente (JCEA). • Buscadores: empleabilidad e ingresos. 	<p>Variables de llave para relacionar tablas:</p> <p>Primaria: Código de la institución.</p> <p>Secundaria: código de sede, código del campus, código del programa/carrera.</p> <p>No se considera el RUT, lo que dificulta el proceso de integración con otras fuentes tales como SII, SNA, INE y MINECON.</p>	<p>Diccionarios disponibles con la descripción de cada campo en cada tabla.</p> <p>Campos comunes:</p> <ul style="list-style-type: none"> • Nombre de la institución. • Tipo de institución: Universidad, Instituto Profesional, Centro de Formación Técnica, FF.AA. • Clasificación institucional: distintas clasificaciones tales como CRUCH/Privadas; Acreditadas/No acreditadas, etc. • Año de registro. 	<p>Cobertura:</p> <p>Transversal: universo de instituciones de educación superior.</p> <p>Longitudinal: serie SIES comienza el 2007 y serie CNED comienza el 2005.</p>
CNED	Índices		<p>Conjunto de tablas relacionadas mediante esquemas relacionales referidas a:</p> <ul style="list-style-type: none"> • Institucional: caracterización de infraestructura, equipamiento y estructura de personal • Oferta Académica Pregrado. • Matrícula. 			

Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Emprendimientos: fuentes relacionadas con la actividad emprendedora.

Tabla 9. Diagnóstico de fuentes para Emprendimientos

Institución encargada	Fuente	Arquitecturas	Diseño y modelos	Integración e Interoperabilidad	Metadatos	Calidad
INE-MINECON	Encuesta de Microemprendimiento	<p>Tipo: muestreo bifásico con sobremuestreo.</p> <p>Adquisición: hogares que contienen al menos un microemprendimiento declarado en la Encuesta Nacional de Empleo.</p> <p>Periodicidad: Bi-anual</p>	Tabla unificada con los resultados de la encuesta para cada uno de los módulos y secciones del respectivo cuestionario	Las tablas no pueden relacionarse con otra base de datos debido a que no tiene un identificador "llave".	Diccionarios disponibles con la descripción de los campos de cada una de las tablas.	El catastro tiene representatividad nacional, permite realizar estimaciones a nivel sectorial.
CORFO	DataEmprendimiento	<p>Tipo: Registros administrativos.</p> <p>Adquisición: llenado para usuarios del programa Startup Chile.</p> <p>Periodicidad: anual</p>	Tabla unificada con las características de los emprendimientos que son o han sido beneficiarios del programa Startup Chile.	Las variables de llave son el nombre del emprendimiento y un ID ficticio generado por el sistema. No se considera el RUT ya que no todos los beneficiarios cuentan con iniciación de actividades en el SII.	Diccionarios disponibles con la descripción de los campos de cada una de las tablas	<p>Las muestras tienen representatividad nacional, lo que no permite realizar estimaciones a nivel sectorial.</p> <p>Por otro lado, los criterios de selección no son claros ya que no se exponen los marcos muestrales para la selección de las respectivas unidades</p>
GEM UDD	Encuesta de actitudes y contexto de la actividad emprendedora	<p>Tipo: Encuesta por muestreo.</p> <p>Adquisición: levantamiento presencial de cuestionario sobre muestra seleccionada a nivel de personas.</p> <p>Periodicidad: anual</p>	Tabla unificada con los resultados de la encuesta para cada uno de los módulos y secciones del respectivo cuestionario	No existen variables de llave, lo que sólo permite la integración a nivel de atributos agregados (región, tipo de actividad, etc.)		

Fuente: Elaboración propia.

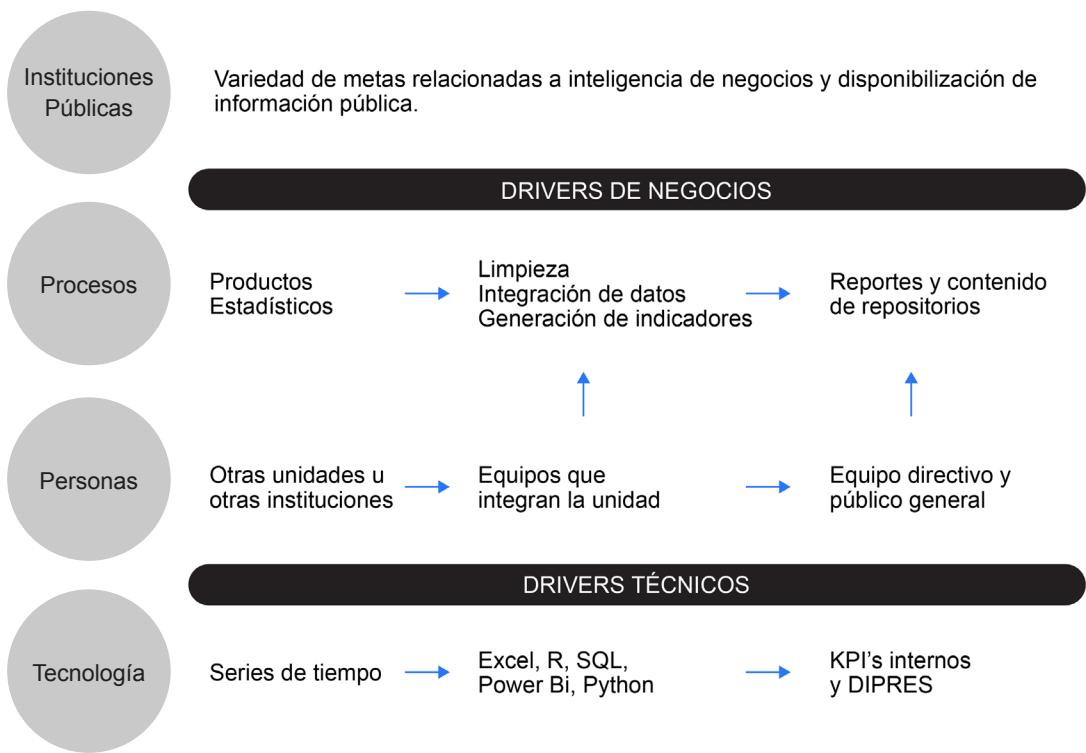
6. Mapa de fuentes de información vinculadas a CTCI

Por otro lado, en el caso de aquellas fuentes que o contaron con la documentación para el diagnóstico, la siguiente figura resume los principales contenidos relacionados al diagrama de contexto.

En términos generales¹⁶, los drivers de negocio se traducen en metas relacionadas a inteligencia de negocios en las organizaciones entrevistadas, mientras que por el lado de los drivers técnicos, no se registran mayores desarrollos desde el punto de vista de las técnicas, con excepción de la CMF que declaró un especial énfasis en el tratamiento de series de tiempo para sus productos estadísticos. Desde el punto de vista de las herramientas utilizadas, destaca el uso de Excel y Power BI, mientras que el uso de herramientas de programación estadística de código abierto se ha convertido en un estándar al respecto, destacando el uso de R y Python. El último componente de los drivers técnicos se relaciona con las métricas para la evaluación del logro de procesos, donde destacan distintos KPI internos y aquellos indicadores exigidos por la Dirección de Prupuestos (DIPRES) respecto de la oferta programática de las instituciones levantadas en el proceso de entrevistas.

16. Para consultar el detalle de los resultados para cada entrevista levantada, ver Anexo 6 Diagrama de Contexto.

Figura 4. Resumen de resultados para diagrama de contexto



Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Al observar los resultados relacionados a la intersección entre procesos y personas, los *inputs* son preferentemente productos estadísticos internos, pero también se consideran inputs de otras instituciones en el contexto de colaboraciones para la triangulación de datos. En particular destaca el caso del INAPI, cuya Unidad de Inteligencia de Negocios realiza procesos de integración de los datos referidos a solicitudes de patentes con datos del SII y de CORFO para efectos de implementar indicadores compuestos a nivel de empresa que permitan una mejor prospección de escenarios relativos a la demanda por instrumentos de propiedad intelectual. Estas colaboraciones presentan un buen precedente para efectos de avanzar en los acuerdos necesarios para la implementación de estándares de interoperabilidad con un mayor potencial de escalamiento.

En relación a las actividades involucradas para la transformación de los inputs en entregables, son los mismos equipos los encargados de realizarlas, los que en algunos casos cuentan con la colaboración de otros equipos asociados a las áreas de comunicaciones cuando se trata de productos estadísticos orientados a consumidores externos. También se identifica la colaboración de equipos asociados a las correspondientes direcciones para efectos de contar con una retroalimentación efectiva respecto de eventuales resultados fuera de rango, o para evaluar la eventual incorporación de mejores métricas.

Respecto de las actividades, estas se relacionan con etapas primarias relacionadas al trabajo con datos, es decir, labores de preprocesamiento, limpieza, normalización, integración y formateo de datos. Por otro lado, también destaca el cálculo de indicadores relacionados a las métricas de evaluación de logro (KPI internos y los exigidos por la DIPRES) y su correspondiente incorporación en plataformas de visualización para inteligencia de negocios (Tableau y Power BI). En este sentido, se destaca la ausencia de ejercicios de analítica que incorporen técnicas relacionadas a modelos de segmentación y predicción, para efectos de identificar potenciales áreas de mejoramiento de la gestión o una mejor focalización de recursos relacionados a la oferta programática de las instituciones.

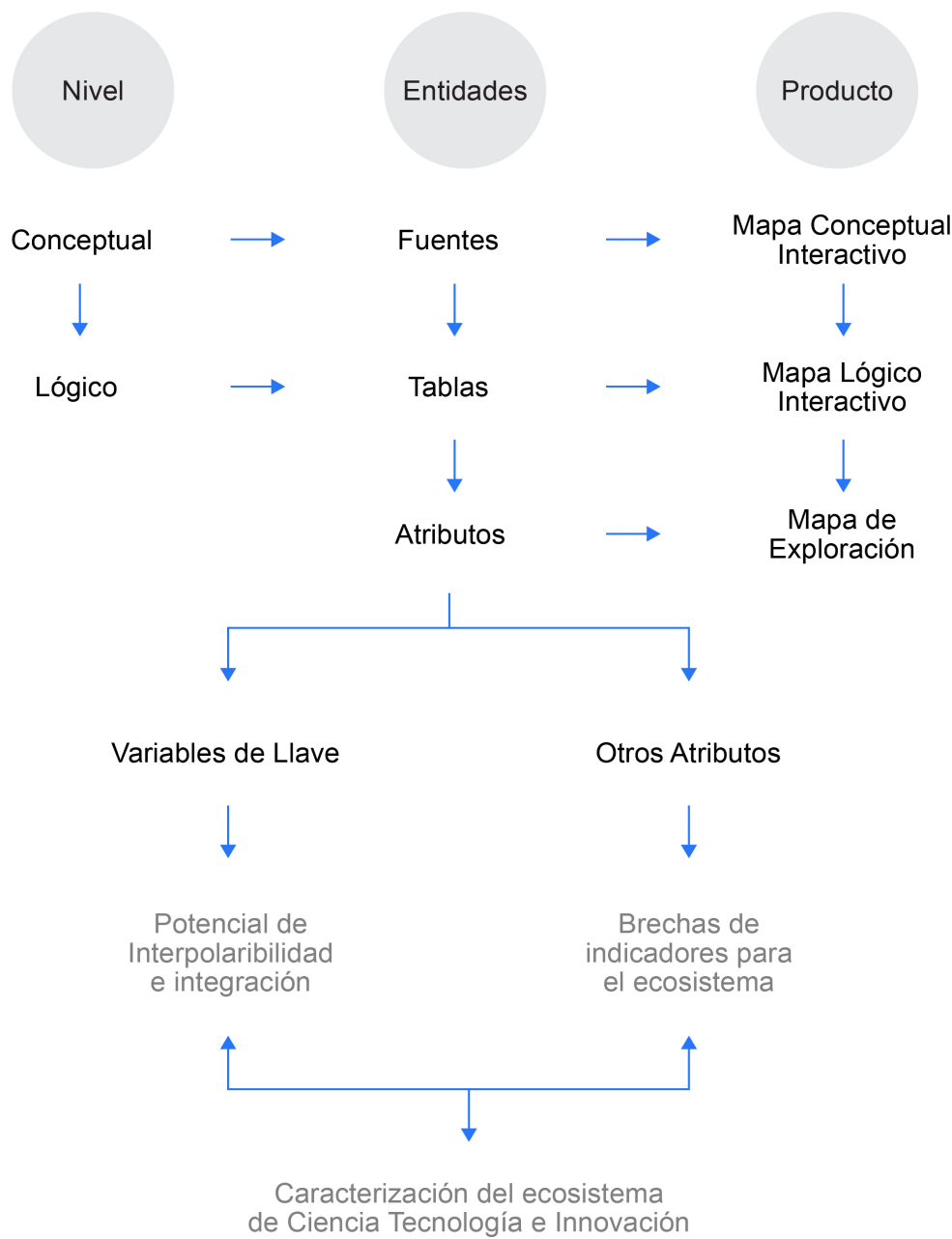
Por último, los entregables corresponden a dos grandes grupos de productos. Por un lado aquellos que constituyen los reportes de uso interno para el monitoreo y evaluación de la gestión (directamente vinculados a los KPI descritos), donde los consumidores son fundamentalmente los distintos equipos directivos dentro de la organización. Por otro lado, otro de los principales entregables se relaciona con la implementación y mantención de los repositorios de datos abiertos, cuyos principales consumidores son investigadores y público general.

6.3. Diseño y representación gráfica del mapa

El diseño del mapa consideró dos niveles de abstracción. En primer lugar, un nivel conceptual en el cual se representan las fuentes y sus posibles conexiones en función de los tipos de capitales que cubren en su conjunto de variables, para de esa forma obtener una visión panorámica del mapa. En segundo lugar, se consideró un nivel lógico en el que las fuentes en detalle según el conjunto de tablas que las componen, para de esa manera establecer las relaciones en un nivel más detallado mediante la identificación de variables de llave primaria que permiten identificar el potencial de interoperabilidad e integración de las fuentes en este nivel, para lo cual se procedió a representar este nivel del mapa como un grafo o red, lo cual además de facilitar la representación gráfica de este nivel del mapa, también permite la obtención de las métricas que permiten caracterizar su estructura y morfología. Lo anterior alude a caracterizar indicadores de centralidad e intermediación cada una de las tablas consideradas.

6. Mapa de fuentes de información vinculadas a CTCI

Figura 5. Primera representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI



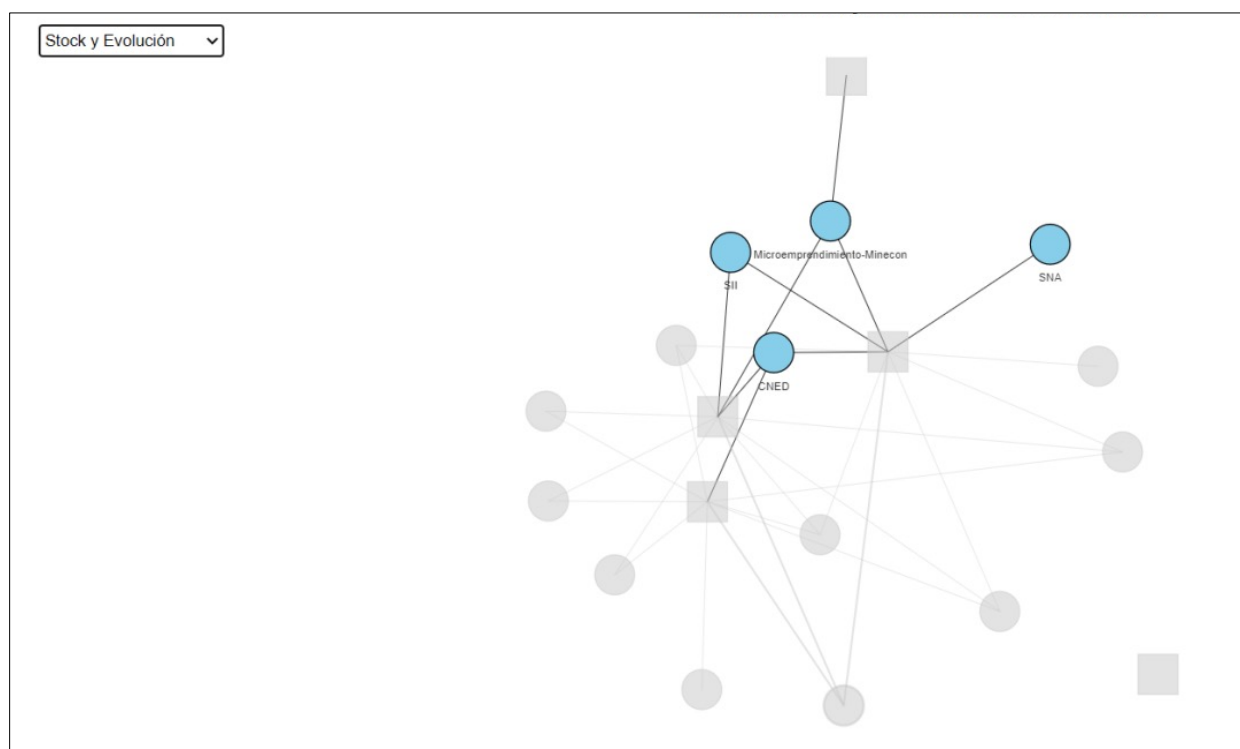
Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Por otro lado, hacia el sector superior izquierdo de la vista se encuentra un selector de niveles de análisis. Tal y como se aprecia en la Figura 6 se despliega cada una de las posibles combinaciones relativas a los niveles de análisis a los que estas fuentes permiten acceder.

De este modo, cuando seleccionamos a aquellas fuentes que permiten acceder a datos relativos al Stock y Evolución, el mapa interactivo deja en la superficie visual sólo a aquellas fuentes que permiten acceder a estos tres niveles de análisis, ocultando a aquellas que no lo permiten, tal y como puede observarse en la siguiente figura.

Figura 7. Segunda representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI

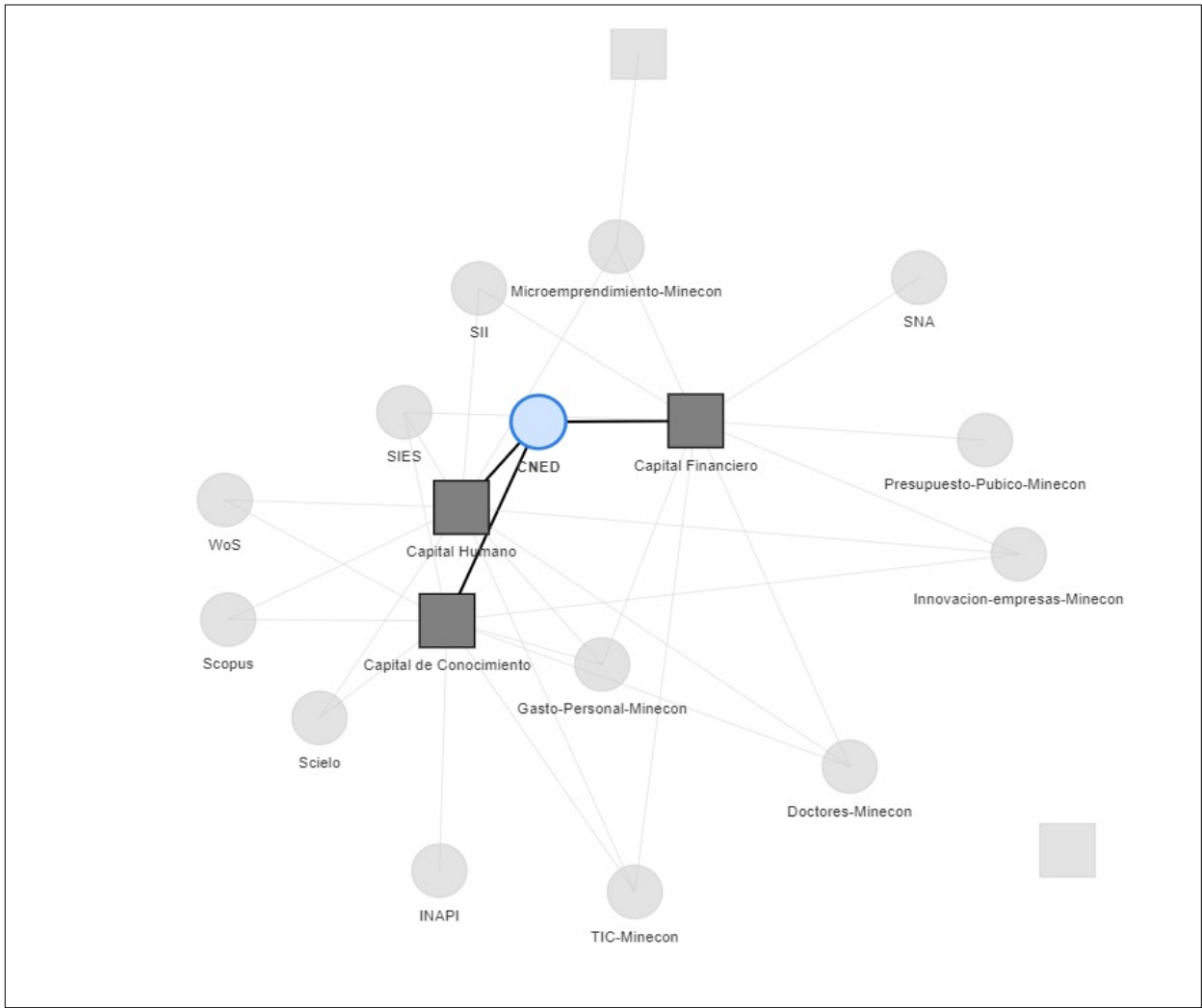


Fuente: Elaboración propia.

Por otro lado, cuando hacemos click en alguna de las fuentes, se despliegan aquellos capitales que pueden ser caracterizados mediante sus datos. En la siguiente figura, hemos seleccionado la fuente CNED (Consejo Nacional de Educación), la que permite acceder a la caracterización de elementos relativos al Capital Humano, el Capital Financiero y el Capital de Conocimiento dentro del ecosistema, tanto a nivel de su stock como de sus evoluciones.

6. Mapa de fuentes de información vinculadas a CTCI

Figura 8. Tercera representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI

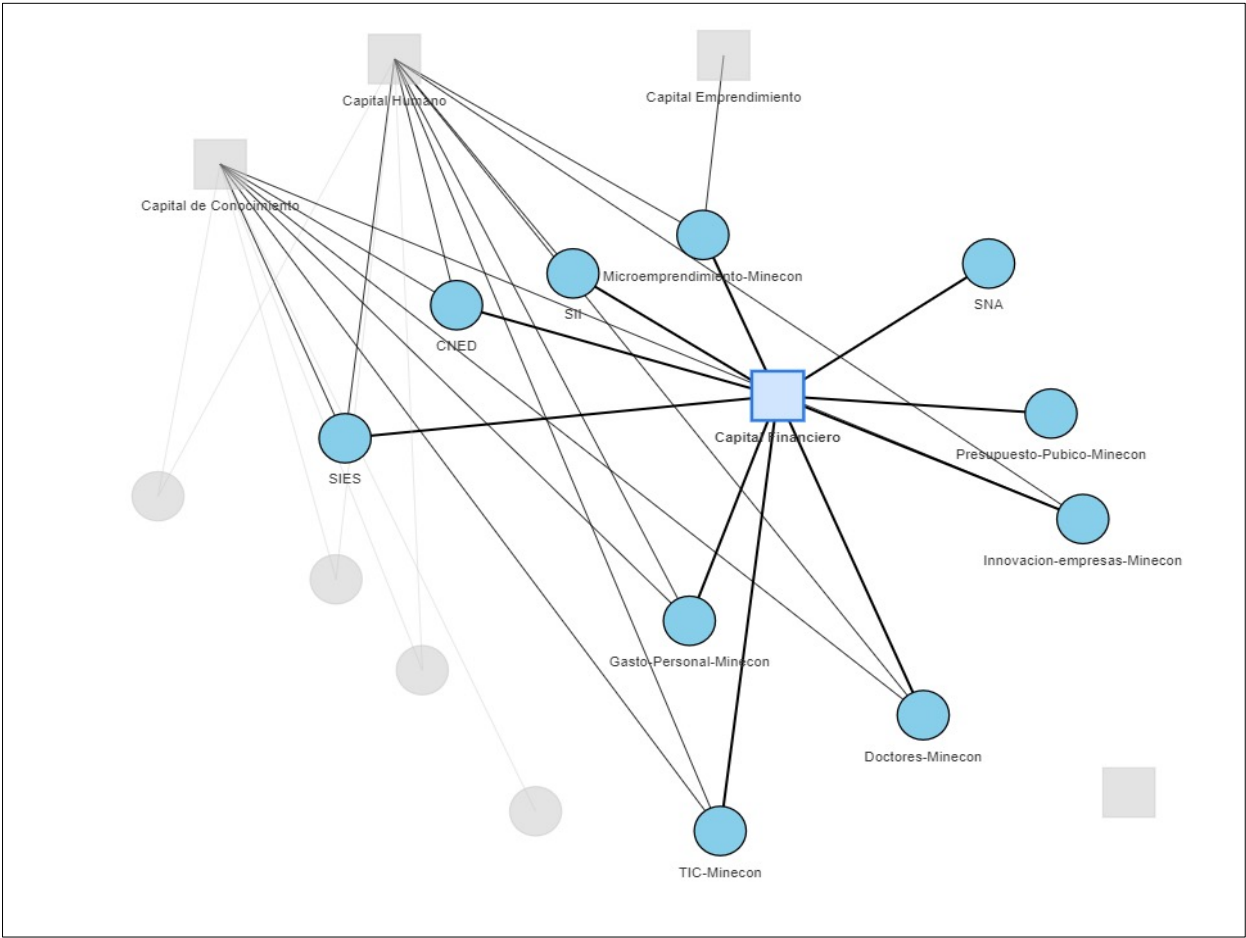


Fuente: Elaboración propia.

También podemos explorar el mapa tomando como referencia a cada uno de los capitales, vale decir, si hacemos click en alguna de las cajas grises se desplegarán aquellas fuentes que permiten su caracterización. Por ejemplo, en la siguiente Figura se muestra una captura de pantalla en la que se ha seleccionado el Capital Financiero, y se despliegan aquellas fuentes que permiten caracterizarlo, tales como el CNED, el SII, el SNA, etc.

6. Mapa de fuentes de información vinculadas a CTCI

Figura 9. Cuarta representación gráfica mapa conceptual interactivo de fuentes para ecosistema CTCI



Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Como observación general respecto de la estructura de esta primera versión del mapa, destaca el que las fuentes relativas a emprendimientos no permiten su integración ya que no cuentan con variables de llave bien definidas. En el caso de DataEmprendimiento se cuenta con el nombre del emprendimiento y con identificador ficticio, lo cual no es suficiente para la eventual integración con otras fuentes. En este sentido, la posibilidad de integración de esta fuente debe considerar estándares de seguridad para el uso de datos personales, lo que en efecto permitiría la integración con los datos del SII a nivel de personas naturales y con otras fuentes a considerar para la segunda versión del mapa, por ejemplo, fuentes relacionadas a datos de seguridad social.

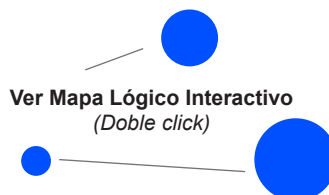
Por otro lado, también se diseñó e implementó un **mapa lógico** de fuentes de información para la caracterización del ecosistema de CTCI, la que consistió en la desagregación de cada una de estas según sus distintas tablas, para de esa forma obtener un panorama más detallado de los datos que contempla cada fuente, asimismo como para dar un mayor énfasis a las posibilidades de interoperabilidad e integración de estas. Si el mapa conceptual permite la vincular capitales y niveles a una o más fuentes, el mapa lógico permite observar las vinculaciones de las distintas tablas que componen a cada fuente. Un elemento importante a considerar es que el potencial de integración de estas fuentes no considera los productos estadísticos que son de acceso abiertos a los usuarios, ya que no todas estas fuentes consideran variables de llave de acceso público. En este sentido, se hace un énfasis en el potencial de integración e interoperabilidad para efectos de la administración de las fuentes, mientras que la disponibilidad de estos datos, será abordada en las conclusiones del estudio en relación a aquellos aspectos que definen los estándares de seguridad en el acceso a distintos productos estadísticos por parte de usuarios externos.

Respecto de la estructura del mapa lógico, no se han considerado todos aquellos atributos que relacionan a cada una de las tablas, sino sólo aquellos que operan como variable de llave primaria, es decir, aquellas variables que permiten la conjunción de tablas a nivel de identificadores individuales (RUT, Código o Nombre de la institución/ establecimiento) para simplificar la representación gráfica y así facilitar la navegación interactiva¹⁷.

17. Para mayores detalles sobre cuáles son estas variables de llave primaria, consultar Anexo 7 Matriz mapa lógico

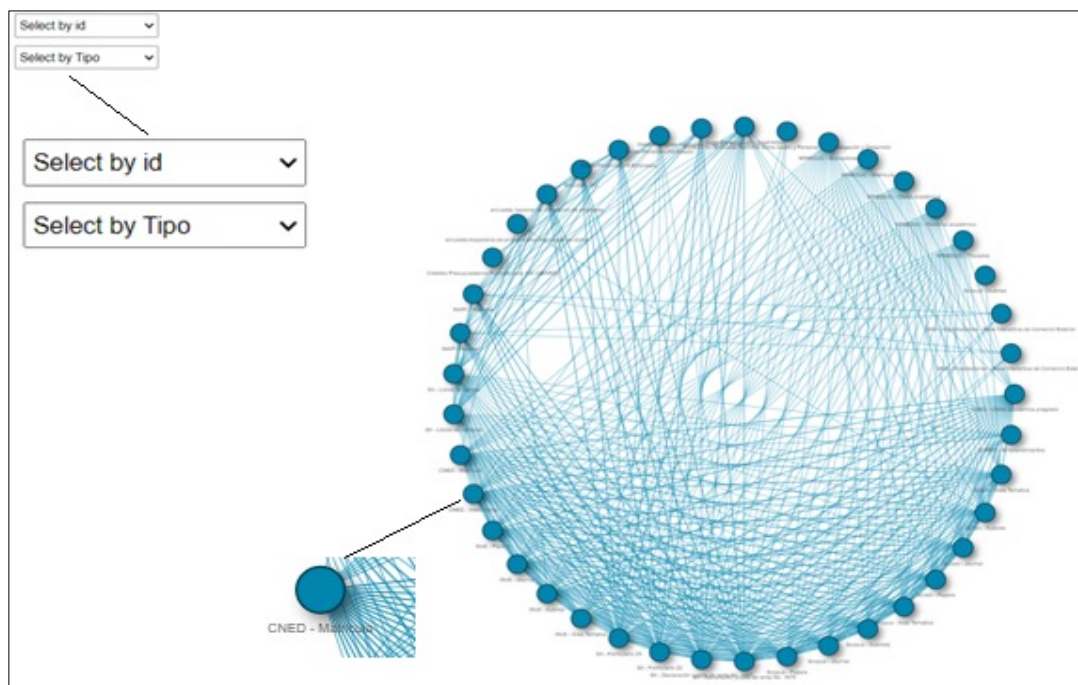
6. Mapa de fuentes de información vinculadas a CTCI

En el siguiente link es posible acceder al mapa lógico interactivo de fuentes de información para la caracterización del ecosistema de CTCI:



La primera vista del mapa interactivo se presenta en la Figura 10, donde cada uno de los círculos de color azul representa una tabla. A su vez cada uno de estos nodos viene etiquetado de manera tal que sea posible identificar la fuente a la que pertenece la tabla, tal y como se puede observar en el acercamiento de la parte inferior de la Figura 10.

Figura 10. Primera representación gráfica mapa lógico de fuentes para ecosistema CTCI



Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Por otro lado, al observar el acercamiento a la parte superior izquierda de la Figura 8, se aprecian dos selectores para la visualización del mapa. El primero de estos selectores permite seleccionar una tabla dentro del total que constituye al listado elaborado, mientras que el segundo selector permite acotar la búsqueda a la combinación de capitales para los cuales es posible encontrar datos en una tabla.

Por ejemplo, en la Figura 11 hemos seleccionado a todas las tablas que permiten acceder a datos para la caracterización del capital de conocimiento, el capital financiero y el capital humano dentro del ecosistema.

Figura 11. Segunda representación gráfica mapa lógico de fuentes para ecosistema CTCI

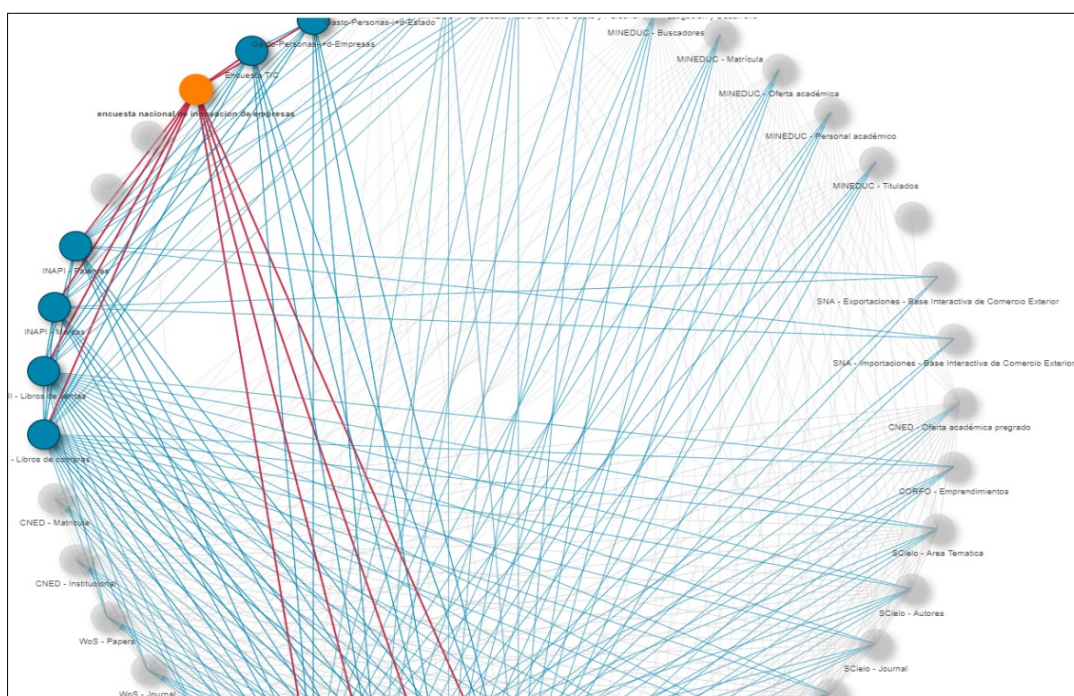


Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Podemos seguir explorando nuestro resultado mediante un acercamiento a estos nodos con el scroll del mouse, y al hacer click en cualquiera de ellos, se activará un color naranja para indicar la tabla seleccionada, y un color rojo que indicará qué tablas pueden ser interoperadas e integradas a la tabla seleccionada, tal y como observarse en la siguiente Figura.

Figura 13. Cuarta representación gráfica mapa lógico de fuentes para ecosistema CTCI



Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

Un aspecto relevante a considerar en ambos niveles del mapa es que las fuentes del SII y la SNA son las que tienen el mayor potencial de interoperabilidad e integración en un alto estándar de calidad, ya que cuentan con una variable de llave estructurada (RUT) y una no estructurada (nombre del establecimiento/institución). En el caso del resto de las fuentes, se cuenta con códigos internos como es el caso de MINEDUC-SIES y CNED-Índices, o con campos no estructurados tales como el nombre de la institución o de los autores como es el caso de los repositorios para publicaciones indexadas.

Por último, en la siguiente tabla se sintetiza el análisis de las métricas que permiten caracterizar esta red de fuentes detalladas para la caracterización del ecosistema:

Tabla 10. Análisis de dimensionamiento general de la red para mapa desagregado de fuentes de información

Indicador	Valor	Interpretación
Densidad de la red	0,89	La red tiene una capacidad de conexión equivalente al 89% de sus nodos
Promedio distancia entre nodos	1,38	Entre nodo y nodo, hay una distancia promedio de 1,38 nodos lo que implica una baja intermediación (conexiones suelen ser directas)
Transitividad de la red	0,86	El 86% de las relaciones entre fuentes son transitivas, es decir, existe un potencial de integración cruzada
Modularidad	16,00	El conjunto de fuentes puede ser agrupado en 16 comunidades emergentes distintas

6. Mapa de fuentes de información vinculadas a CTCI

Por otro lado, la Tabla 11 da cuenta del porcentaje de relaciones concentradas para cada nodo, es decir, el potencial de interoperabilidad e integración de cada tabla correspondiente a su fuente. De este modo, los valores de centralidad más altos indican el porcentaje de relaciones de integración posibles de establecer con el resto de las fuentes, donde destaca la tabla de Oferta Académica del pregrado de la base de datos Índices del CNED, ya que mediante el código de institución puede relacionarse con el resto de las tablas referidas a Educación Superior, mientras que con el nombre de la institución es posible de ser relacionada con fuentes tales como el SII, INAPI, SNA y los repositorios de publicaciones científicas.

Tabla 11. Métricas de centralidad para cada una de las tablas/fuentes del mapa desagregado.

Centralidad (capacidad de interoperabilidad con otras fuentes)	Nodo
1,000	CNED - Oferta académica pregrado
0,976	CNED - Matricula
0,964	SII - Declaración jurada de renta No. 1879
0,943	SCielo - Autores
0,943	SCielo - Journal
0,943	SCielo - Papers
0,943	Scopus - Area Tematica
0,943	Scopus - Journal
0,943	Scopus - Papers
0,937	WoS - Autores
0,937	WoS - Journal
0,937	WoS - Papers

6. Mapa de fuentes de información vinculadas a CTCI

0,919	SCielo - Area Tematica
0,918	Scopus - Autores
0,885	CNED - Institucional
0,770	SII - Declaración jurada de renta No. 1887
0,768	SII - Formulario 22
0,767	SII - Formulario 29
0,734	WoS - Area Tematica
0,568	SII - Libros de ventas
0,568	SII - Libros de compras
0,491	MINEDUC - Buscadores
0,491	MINEDUC - Matrícula
0,491	MINEDUC - Oferta académica
0,491	MINEDUC - Personal académico
0,491	MINEDUC - Titulados
0,255	INAPI - Patentes
0,228	INAPI - Marcas
0,208	CORFO - Emprendimientos
0,071	MINECON - Encuesta Nacional sobre Gasto y Personal en Investigación y Desarrollo
0,058	SNA - Exportaciones - Base Interactiva de Comercio Exterior
0,058	SNA - Importaciones - Base Interactiva de Comercio Exterior
0,026	Scopus - Autores

Fuente: Elaboración propia.

6. Mapa de fuentes de información vinculadas a CTCI

De este modo, la presentación del mapa conceptual interactivo permite un conocimiento detallado de las conexiones entre las fuentes con la línea de base, tanto en relación con los capitales considerados para la caracterización del ecosistema, como en relación a los distintos niveles de análisis que permiten profundizar esta caracterización. Por su parte, el mapa lógico interactivo permite profundizar a nivel de cada una de las tablas, cuáles son las posibilidades de interoperabilidad e integración entre distintas fuentes, tanto a nivel de las variables que operan como llaves primarias para el enlazamiento de entidades, como a nivel de qué capitales son capturados por los datos de cada una de las tablas seleccionadas.

Por otro lado, la presentación interactiva de ambos mapas permite la realización de consultas particulares para cada capital o nivel de análisis, sin necesidad de recorrer el mapa completo. Este es un elemento de valor agregado en el presente trabajo ya que aporta elementos que pueden ser utilizados para un repositorio integrado de datos para la caracterización del ecosistema de CTCI.

Por último, en el presente estudio se incluye un Anexo¹⁸ que permite la exploración de ejemplos representativos de cómo cada una de estas fuentes contiene variables que cubren los distintos capitales definidos en la matriz conceptual. De este modo, se definen criterios de usabilidad del mapa que permite a quienes lo consulten, dar con aquellas fuentes, tablas y variables que permitan la implementación de indicadores tanto tradicionales, como no tradicionales para abordar preguntas y objetivos de investigación relacionados a la estructura y dinámica del ecosistema.

18. Ver anexo 8 Matriz de exploración

7. Preguntas de investigación que requieren de otras fuentes de información

7. Preguntas de investigación que requieren de otras fuentes de información

7.1. Potenciales preguntas de investigación

Uno de los elementos más importantes que permite el tratamiento empírico de preguntas de investigación es la trazabilidad de los datos utilizados. En efecto, la posibilidad de contar con conjuntos de datos trazables para distintas unidades de análisis, ha permitido la apertura de un amplio abanico para el uso de métodos de investigación más exhaustivos, que permitan no sólo responder de mejor manera las preguntas que orientan el trabajo de investigación, sino también sugerir otras líneas de formulación de problemas que hasta hace algún tiempo eran considerados de tipo más teórico.

Uno de los objetivos del estudio es definir preguntas de investigación desafiantes junto a sus implicancias en el mapa definido en este estudio. De esta forma, sobre la base de la revisión de casos, entrevistas y los requerimientos de la contraparte se definirán preguntas de investigación que requieren de otras fuentes de información y con esto analizar la disponibilidad de información cruzada.

En esta sección, se triangularon, por un lado, los resultados de la sistematización de la bibliografía seleccionada y, por otro, los resultados del proceso de entrevista con los actores seleccionados para el ecosistema. Considerando lo anterior, la presentación de los siguientes resultados y su análisis se organizó en función de la identificación de preguntas desafiantes para el monitoreo y evaluación del ecosistema de ciencia, tecnología, conocimiento e innovación.

7. Preguntas de investigación que requieren de otras fuentes de información

1. ¿Cómo medir la direccionalidad y orientación práctica de la ciencia en Chile?

1.1. ¿Cómo puede calcularse el nivel de vínculo existente entre ciencia y empresa en el país?

Se identificó un discurso reiterado asociado a la falta de vinculación entre el desarrollo científico y su uso práctico, coincidiendo además en la imposibilidad de cuantificar el fenómeno. Al mismo tiempo que se valora el capital científico chileno como una de las fortalezas principales del ecosistema, se identifica que los mecanismos tradicionales de medición del trabajo académico no son capaces de dar cuenta del nivel de impacto de la ciencia en Chile. En palabras de los entrevistados:

“el ecosistema, como tú dices, tiene estas dos líneas súper marcadas y no son muy integradores entre sí, es la parte de la academia y la parte privada” (Entrevistado Número 7 del perfil Análisis, promoción o ejecución de las políticas públicas).

“Yo sé que históricamente ese es un problema, que las empresas hacen muy poco, con excepción de algunas grandes. Se han hecho muchos esfuerzos en América Latina, en general, pero no hay una mayor cultura en esa dirección. Hay distintas razones. El mercado, la demanda, no es muy grande en estos países, la clase media es un gran sector que tiene muy poca capacidad de adquirir, de adquisición. Entonces, hay empresarios, pero son empresarios de sobrevivencia, que están tratando de sobrevivir no más. Entonces, es muy limitado el nivel, y cuando es, tiende a ser hacia fuera” (Entrevistado Número 11 del perfil Análisis, promoción o ejecución de las políticas públicas).

“A nivel de la academia hay soluciones preciosas a problemas inexistentes. En la academia se desarrollan soluciones [...], y avanzan y avanzan, pero nunca han salido a preguntarle a la industria: “Bueno, ¿esto es lo que tú realmente necesitas?”. O le preguntan a Google, pero no les preguntan a los usuarios, no le preguntan al potencial usuario, no le preguntan al que podrían desplazar, no le preguntan a quién le podrías mejorar sus procesos, a esos no les preguntan, hasta muy tarde. Eso inevitablemente gatilla una desconexión en estas soluciones” (Entrevistado Número 1 del perfil Análisis, promoción o ejecución de las políticas públicas).

7. Preguntas de investigación que requieren de otras fuentes de información

Esto, hace pensar que podría ser valiosa la creación de un indicador que mida el estado de vinculación del sector científico con el sector productivo, de forma tal de poder conocer si existen grados de avance en la materia, más allá de los casos de éxito puntuales que puedan utilizarse para difundir y promover el ecosistema nacional. Esto además podría permitir la identificación de la agregación de valor sectorial en términos productivos de la ciencia en Chile.

1.2. ¿Cómo cuantificar y disminuir la brecha de financiamiento para empresas de base científico-tecnológica?

Otro tema común en algunas entrevistas se refiere a la ausencia de financiamiento para el emprendimiento basado en ciencia, haciendo referencia a la ausencia de suficientes capitales de riesgo o a la creación de formas de financiamiento novedosas capaces de soportar el peso del capital inicial en este tipo de iniciativas. Los entrevistados lo señalan de la siguiente forma:

“Creo que el principal problema es ese, que los bancos no te pescan dando plata a tus proyectos y lo otro es que no hay firmas de capital semilla o digamos de estas firmas de inversión de alto riesgo donde el tipo tiene a lo mejor, no sé, 500 mil dólares y los reparte en 20 proyectos sabiendo que 9 de los proyectos van a fallar, pero que uno va a permitir que recupere las inversiones perdidas, ese tipo de cosas no existen, digamos.” (Entrevistado Número 3 del perfil Análisis, promoción o ejecución de las políticas públicas).

“Entonces de lo que se ha hablado mucho es del Venture capital, por ejemplo, que, si bien ahí Chile está haciendo cosas interesantes, está empujando fuerte, digamos, yo creo que ahí falta aún, creo que todavía no tiene la fuerza que debería. Yo sé que tiene varios desafíos, digamos, pa formarse, pa crearse, pa sustentarse” (Entrevistado Número 6 del perfil Análisis, promoción o ejecución de las políticas públicas).

7. Preguntas de investigación que requieren de otras fuentes de información

La impresión de la falta de capitales para la innovación en Chile podría constituirse en una realidad si es que existiera una forma de medir la cantidad de capital para la innovación en Chile en términos proporcionales, utilizando benchmarks que consideren como variables de control el tamaño de la población, el gasto fiscal en innovación y otros datos relevantes. Una vez cuantificada con exactitud, la brecha podría indicar si se requieren esfuerzos de una magnitud relevante para el logro de los objetivos principales del ecosistema.

Uno de los aspectos más complejos involucrados en la posibilidad de proporcionar datos para responder esta pregunta, es la dificultad de obtener una trazabilidad efectiva de los emprendimientos. En efecto, los emprendedores son agentes del ecosistema que operan fundamentalmente en una esfera individual, donde las dimensiones organizacionales son de escaso o nulo desarrollo, lo que impide su trazabilidad mediante sistemas de registro que operan a nivel de empresas.

En este sentido, uno de los desafíos más importantes para efectos del enriquecimiento de las fuentes de información para la caracterización del ecosistema, se relaciona con la posibilidad de integrar datos a nivel de personas con los datos a nivel de establecimientos. Si bien las dificultades técnicas no son menores en la implementación de proyectos de interoperabilidad, los principales desafíos son de orden normativo institucional, ya que se trata de garantizar estándares de protección de datos personales.

Una estrategia para la disponibilidad de datos personales puede consistir en la agregación de entidades y atributos para los datos descargables. Mientras que la agregación de entidades se traduce en que cada fila de una tabla se constituye de una cantidad de entidades de una misma categoría (por ejemplo: 5 emprendedores de base científica y tecnológica), la agregación de atributos se traduce en la presentación de datos en tramos que no permitan la individualización de las entidades).

2. ¿Cómo medir la coordinación en y del ecosistema CTCI en Chile?

2.1 ¿Cuál es el nivel de conexión e integración general del ecosistema de CTCI en Chile?

Un comentario recogido a partir las entrevistas tiene relación con la falta de conexión del ecosistema de CTCI en Chile, cuestionando incluso el uso de la palabra ‘ecosistema’ para referirse a las políticas públicas de CTCI en el país. Las siguientes citas son un ejemplo de este tipo de opiniones.

“yo no estoy tan seguro que exista ese ecosistema. Sí creo que hay algunas redes que en general, entre comillas, se podría decir que profitan de los fondos públicos; no sé qué tanto éxito tengan en términos de vincular las posibilidades de los conocimientos y las capacidades científico-tecnológicas de la academia o de la sociedad civil con la empresa” (Entrevistado Número 3 del perfil Análisis, promoción o ejecución de las políticas públicas).

“Como hay un diagnóstico súper consensuado, de que, en el fondo, el sistema estaba fragmentado, de que no había un encadenamiento claro de cómo el I+D se conectaba con la innovación, la transferencia. Y, hoy en día, el hecho de que, en el fondo, se esté de alguna manera reforzando a través de una nueva y primera política que hace la avanzada” (Entrevistado Número 8 del perfil Análisis, promoción o ejecución de las políticas públicas).

La creación de indicadores globales de integración del ecosistema de CTCI en Chile puede ser valiosa para tener certeza de los avances o retrocesos en términos de vinculación general. En este sentido, puede ser posible la creación de indicadores de integración utilizando algunos datos ya generados de manera pública en Chile, y promover la creación de indicadores de conexión específicos entre algunos de los actores más relevantes.

Este tipo de indicadores requiere también de un cambio en la forma en la que se entiende la medición misma, ya que en términos generales asumimos que la medición está orientada hacia propiedades de los agentes por separado, las que después agregamos mediante el procesamiento de los datos. Lo anterior también repercute sobre la forma en que representamos los datos que alimentan estos indicadores, ya que, en términos generales, los formatos tradicionales de datos analizables se organizan en tablas donde las filas representan a las entidades y las columnas a sus distintas características.

7. Preguntas de investigación que requieren de otras fuentes de información

Lo anterior puede ser reenfocado mediante la incorporación de formatos de datos que permitan su representación como redes de nodos vinculados por aristas. Sin necesidad de entrar en detalle sobre los formatos que permiten la representación de datos en red, lo importante reside en la posibilidad de acceder a enfoques que permitan la implementación efectiva de indicadores de mayor complejidad, como lo es el caso de los fenómenos relacionados a la integración y la coordinación al interior del ecosistema CTCI.

La noción de complejidad se relaciona estrechamente con la posibilidad de caracterizar vinculaciones entre agentes, para de esa forma comprender mejor sus características y propiedades. Por su parte, la capacidad de registrar estas vinculaciones se relaciona directamente con explotar los datos de los que actualmente se dispone, tanto en los repositorios académicos (WoS, Scopus, SCielo), como en los sistemas de registro del INAPI, CORFO, FIA y ANID. En efecto, estos sistemas son los que permiten el registro colaborativo de entidades para sus respectivas materias institucionales, es decir, se trate del registro de marcas/patentes por un grupo de aplicantes, la postulación a fondos de investigación o fomento por grupos de empresas e instituciones proveedoras de conocimiento o la elaboración conjunta de un paper de divulgación de resultados de proyectos.

2.2. ¿Cómo se puede calcular el costo y beneficio de la coordinación en los distintos ámbitos de la práctica de la CTCI en Chile?

A pesar de no ser un discurso dominante, a partir de algunas entrevistas puede considerarse relevante el análisis de la estructura de incentivos y apoyo para la coordinación en el ecosistema de CTCI en Chile. Tomando en cuenta que la coordinación requiere de esfuerzos y recursos que no necesariamente están disponibles para los actores, y que muchas veces los indicadores actuales no miden ni promueven la colaboración, su cálculo en términos específicos para cada parte del ecosistema puede ser un primer paso para el monitoreo, asignación de recursos y valoración del trabajo en conjunto.

En términos prácticos esto debería implicar el incrementar recursos para los proyectos que requieren la colaboración entre actores e instituciones, pero en referencia a un cálculo de los costos y beneficios del trabajo común.

7. Preguntas de investigación que requieren de otras fuentes de información

3. ¿Cómo medir el impacto científico y tecnológico más allá de su relación con la economía?

En relación con la identificación de espacios novedosos y desafiantes para la creación de indicadores de impacto de la CTCI en Chile, la construcción de indicadores distintos al impacto económico o de mercado fueron parte importante de algunas de las conversaciones sostenidas. A continuación, se presentan algunas citas de los entrevistados para contextualizar esta pregunta de investigación:

“yo creo que son espacios de innovación que no se ven porque el sistema está tan concentrado en la innovación productiva como económica, de producción de un objeto comercializable que no ven estos otros espacios de innovación que impactan en otras dimensiones de lo social, o sea, a nivel, por ejemplo, de fortalecer comunidades, de fortalecer prácticas culturales, de fomentar la creatividad para elaborar, por ejemplo, no sé, traumas en alguna comunidad específica ¿ya? o para visibilizar ciertos temas de esa comunidad” (Entrevistado Número 13 del perfil Análisis, promoción o ejecución de las políticas públicas).

“licencié una nueva tecnología y pasan 3 años hasta que yo no logro vender algo asociado a esa tecnología, o nunca la vendo, probablemente yo nunca vendo algo asociado a esa tecnología. Por lo tanto, no puedo asignarle un royalty para pagarle a la universidad. Pero, resulta que optimicé mi planta, contraté a 6 personas más, y eso no se mide hoy. Entonces, creo que hay tremendos esfuerzos, pero, también, no los estamos mirando de la manera correcta” (Entrevistado Número 1 del perfil Análisis, promoción o ejecución de las políticas públicas).

En este ámbito, sería relevante la creación de indicadores de señalización o de demanda no económica que orienten la búsqueda de soluciones novedosas por parte de los actores científicos y tecnológicos en Chile.

Un ejemplo de este tipo puede observarse en la reciente publicación de un estudio realizado en Colombia que vincula las demandas sociales con las orientaciones de la ciencia básica y aplicada¹⁹. En este mismo sentido, también puede mencionarse el esfuerzo de investigadores para vincular las necesidades sociales de distintos países con el output de su producción científica (Ciarli & Rafols, 2019). Ambas investigaciones son bastante

19. Ramírez, M., Estévez, J. H. G., Goyeneche, O. Y. R., & Rodríguez, C. E. O. (2020). Fostering place-based coalitions between social movements and science for sustainable urban environments: A case of embedded agency. *Environment and Planning C: Politics and Space*, 38(7-8), 1386-1411.

7. Preguntas de investigación que requieren de otras fuentes de información

recientes, por lo que resulta relevante promover la creación de indicadores similares para el caso chileno, que permitan establecer el vínculo entre las prioridades y el financiamiento del sector científico-tecnológico, y su vinculación con problemas no económicos.

Otra forma de observar el fenómeno tiene que ver con la medición y promoción de las empresas de “triple impacto”, donde se busca incentivar la creación de empresas que permitan solucionar problemas de tipo económico, social y ambiental. Si bien existen indicadores de medición de triple impacto en algunas organizaciones²⁰, su masividad y uso en escenarios de política pública es limitada. El cálculo o la generación de datos públicos asociados al triple impacto potencial de la CTCI en Chile es escaso, y puede ser un espacio interesante para la creación de nuevos indicadores.

Finalmente, la expansión del concepto de innovación para poder considerar y medir conceptos como “innovación social”, “innovación desde la base de la pirámide”, “*grassroot innovation*” y otros similares permitirá el cálculo de nuevos impactos de la CTCI en Chile. Si bien la encuesta nacional de innovación en empresas ha integrado una pregunta sobre innovación social, los esfuerzos por medir o cuantificar los impactos de la ciencia y la tecnología en espacios de innovación social no han sido desarrollados en profundidad.

4. ¿Cómo medir e incentivar la diversidad en los desarrollos científico-tecnológicos?

Los equipos de investigación que involucran equipos multidisciplinarios generalmente abordan temas complejos y relevantes para la sociedad. Además, las miradas multidisciplinarias permiten la captura de problemas de características diversas, que promueven soluciones innovadoras y novedosas. En relación con el valor cada vez más patente de la conversación entre disciplinas diversas, los indicadores y los incentivos para su realización no permiten identificar la importancia real de este tipo de investigaciones en Chile.

La literatura académica en este ámbito ha identificado algunas barreras generales para el trabajo multidisciplinario²¹, y entre ellas se encuentran: menor probabilidad de publicar, problemas para llevar a cabo visiones epistémicas integradoras sobre el tema de investigación y dificultades de las evaluaciones por pares. La mayoría de estas barreras son consecuencia de la complejidad de lograr un alto grado de capacidades multidisciplinarias del tema de investigación.

20. Por ejemplo, el estudio “Impacto de las empresas B en Latinoamérica” considera la metodología de la organización internacional Sistema B. Estudio disponible en: <http://biblioteca.academiab.org/items/show/106>

21. Algunos trabajos relevantes en esta área son: Schuitema, G., & Sintov, N. D. (2017). Should we quit our jobs? Challenges, barriers and recommendations for interdisciplinary energy research. *Energy Policy*, 101, 246-250; Gleed, A., & Marchant, D. (2016). *Interdisciplinarity: Survey report for the Global Research Council 2016*

7. Preguntas de investigación que requieren de otras fuentes de información

La creación de indicadores en este ámbito permitiría conocer el nivel de desarrollo del trabajo multi e interdisciplinar en Chile, su ausencia o relevancia en la aproximación a problemas complejos, y la necesidad de facilitarla o promoverla en contextos específicos.

Las posibilidades de implementación de este tipo de indicadores se relacionan con la explotación de datos de texto, los que muchas veces están involucrados en la descripción de proyectos postulados a instrumentos públicos, en el contenido de los *papers* para la divulgación de resultados, u otros canales digitales relacionados con el ingreso de datos de texto por parte de los agentes del ecosistema. El diseño de algoritmos para el análisis de texto permitiría construir una categorización del contenido relacionado a la actividad en distintos campos dentro del ecosistema, de manera tal que podrían construirse métricas de diversificación que operen por dentro de las categorizaciones tradicionales (sistemas de clasificación en los *journals* y distintos estándares de categorización de áreas usados a nivel oficial, en particular los recomendados por la OECD).

Para finalizar esta sección, es importante mencionar que es posible identificar algunas potenciales preguntas que no tienen un correlato directo con palabras de los entrevistados, pero que emergen de la conexión de información y proceso de las conversaciones. Entre ellas puede mencionarse las siguientes:

- ¿Cuáles son las condiciones laborales de los científicos chilenos y como se comparan con otros países?
- ¿Cuál es el nivel de vinculación entre desafíos de sustentabilidad, inversión en CTCI y producción científica?
- ¿Cómo medir el impacto del conocimiento en el ecosistema, por ejemplo, en las industrias culturales?
- ¿Cómo medir los niveles de transferencia tecnológica, según tipo de tecnología, sector y objetivos?

Recordando el comienzo del presente documento, la relevancia y pertinencia de las preguntas presentadas en esta sección debe estar vinculada con los objetivos y alcances de la estrategia y acciones de política pública en CTCI. Para esto es necesario la conformación de espacios de reflexividad que permitan a quienes son evaluados entregar sus comentarios no sólo sobre la constitución del indicador, si no que también sobre las razones para su conformación.

7.2. Análisis de la información construida

En base a los intereses de la investigación, la definición de un espacio no tradicional fue considerado como parte de la estrategia metodológica, definiendo categorías analíticas para la distribución de la muestra, principalmente en base a su negación o contradicción con lo claramente definido como tradicional.

Como se mencionó anteriormente, el proceso de entrevistas para este estudio puede caracterizarse como de requerimientos creativos, altos. Muchas veces los entrevistados no fueron capaces de responder directamente las preguntas, porque el proceso de propuesta de nuevos indicadores es algo que requiere tiempo y vínculo de ideas complejas. De cierta manera, se espera facilitar un proceso de innovación en la construcción de indicadores de CTCI.

Considerando lo anterior, y en base a las entrevistas realizadas, el espacio de lo “no tradicional” puede ser diferenciado en al menos dos categorías: lo incremental y lo radical o disruptivo. El espacio incremental identifica deficiencias en los métodos tradicionales, e intenta complementar, modificar y mejorar lo que se mide actualmente. El espacio radical identifica formas distintas de entender la CTCI, que no necesariamente conversa con la forma tradicional de operar.

En este sentido, preguntas de investigación como la 1, 2 se vinculan con la mirada incremental, mientras que las preguntas 3 y 4 tienen una mirada mucho más radical sobre el proceso de identificación o creación de nuevos indicadores para la CTCI en Chile. Comprender la naturaleza de la pregunta de investigación permitirá definir los esfuerzos e inversiones requeridas para su creación y difusión en los espacios de análisis, promoción o ejecución de las políticas públicas en Chile.

En relación con un análisis global de las entrevistas realizadas, existen dos grandes preguntas que debiesen responderse para garantizar un proceso de diversificación de indicadores para la CTCI en Chile.

7. Preguntas de investigación que requieren de otras fuentes de información

En primer lugar, es necesario identificar qué capacidades y recursos existen para la creación de indicadores de nicho en los espacios donde se requieren. Tal y como se identificó en la conformación del mapa de fuentes de información, en Chile las grandes agencias públicas generan datos relevantes para el ecosistema de CTCI en Chile que cumplen con estándares internacionales y poseen alta rigurosidad metodológica. Pero para la construcción de respuestas a las preguntas de investigación que buscan respuestas radicales a los problemas del ecosistema se requerirán recursos y capacidades que no necesariamente están presentes en las instituciones productoras de datos. En términos prácticos, no se requieren las mismas habilidades para medir el gasto en I+D en las empresas que para medir el impacto social de las tecnologías de regadío en el campo en Chile. La construcción de espacios que promuevan miradas diversas sobre la CTCI y que cuenten con recursos para construir nuevos datos son condiciones necesarias para la materialización de la generación de datos no tradicionales sobre el ecosistema de CTCI en Chile.

Finalmente, es importante considerar las razones para promover la creación de nuevos indicadores, los objetivos esperados y los niveles de participación de los involucrados. Esto implica tener en cuenta el factor performativo de los indicadores y su efecto en lo que se espera medir. La recomendación final en este ámbito es recoger las sugerencias del Manifiesto de Leiden²² en términos de la valoración de los contextos específicos en donde se realiza la actividad científica y tecnológica, el involucramiento de quienes serán medidos en la creación de los indicadores, la valoración de los datos cualitativos en los procesos de monitoreo y evaluación y la aceptación de que una evaluación y monitoreo robustos requieren de múltiples indicadores que permitan una visión completa de lo que se espera medir o evaluar.

22. Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431.

8. Conclusiones

8. Conclusiones

Una primera conclusión del presente trabajo apunta a que actualmente existen diversas fuentes de datos que permiten cubrir las brechas identificadas, tanto en lo que respecta al stock para los distintos capitales considerados para la caracterización del ecosistema, como también en relación a datos que permitirían implementar indicadores de conexión y evolución. Lo anterior cuenta particularmente para el caso de los agentes que se constituyen de manera formal, tal y como es el caso de las empresas, las instituciones proveedoras de conocimiento y las distintas agencias del sector público. No obstante, el caso de los emprendimientos es el más complejo dado que se trata de un agente difícil de trazar dada su naturaleza volátil respecto de los distintos sistemas de registro.

Desde el punto de vista del stock de capitales, las fuentes analizadas permiten contar con datos para caracterizar de manera precisa distintas variables relativas al ámbito financiero tanto en el sector público como en las empresas, incluso es posible caracterizar la infraestructura tanto para las empresas (distintas encuestas de innovación) como para instituciones proveedoras de conocimiento (universidades y centros de investigación en los sistemas de registro tanto del MINEDUC, el CNED y la ANID).

En cuanto al *capital humano*, las fuentes diagnosticadas permiten contar con una gran diversidad de datos para la caracterización de la dotación de recursos humanos de manera especial en las instituciones proveedoras de conocimiento, ya que se cuenta con información para caracterizar al personal académico según distintas categorías, tanto a nivel del régimen de vinculación (planta jornada completa, media jornada u honorarios), según su nivel de formación académica, según sexo. También se cuentan con datos para la caracterización del personal no académico, al menos a nivel agregado.

Estas consideraciones también pueden ser tomadas en cuenta para las empresas, donde si bien no se cuenta con el mismo nivel de detalle, el sistema de registro del SII permite la caracterización según el número de trabajadores, mientras que en el caso de las distintas encuestas consideradas es posible acceder a datos valiosos relativos a la estructura organizacional de las empresas, en particular en las áreas de apoyo a los procesos de innovación. En este sentido, uno de los temas a explorar en futuras investigaciones se relaciona con la posibilidad de diseñar e implementar indicadores que permitan la comparación de distintas estructuras organizacionales, tanto a nivel de la empresas como del Estado, para comprender aquellos elementos del capital humano que no forman parte de la agenda tradicional de investigación, orientada preferentemente a las características formales de las trayectorias formativas de los trabajadores.

8. Conclusiones

Respecto del *capital de emprendimiento*, tal y como se ha sostenido a lo largo del presente trabajo, las actuales fuentes de datos no permiten generar un cuadro de correcta trazabilidad entre los emprendimientos. No obstante, actualmente existen sistemas de registro de datos personales que sí permitirían explorar la posibilidad de integración e interoperabilidad de la actividad emprendedora, tales como el registro civil y los distintos sistemas de información del Ministerio del Trabajo y Previsión Social. En esta línea, las principales dificultades técnicas se relacionan con la posibilidad de enlazar el RUN de la persona a la actividad de un emprendimiento que haya sido efectivamente registrado por esa persona, lo cual sólo cuenta para los programas públicos relacionados. Una alternativa para la integración de datos que no sean exclusivos del sector público, es establecer un marco de alianzas y trabajo en conjunto con distintos *Hubs* de Innovación y emprendimiento, para efectos de consolidar un sistema único de registro de la actividad emprendedora. En este tipo de esfuerzos puede ser relevante contar con la experiencia de algunas plataformas que proveen servicios de evaluación y monitoreo de emprendimientos, tal y como es el caso de *charly.io* apoyada por Startup Chile.

Otro elemento relevante sobre el capital emprendedor se relaciona con el emprendimiento corporativo, vale decir, las actividades emprendedoras que surgen y se desarrollan al interior de organizaciones tales como las empresas, el Estado y las instituciones proveedoras de conocimiento. En este sentido, las distintas encuestas de innovación e I+D consideradas cuentan con datos que al menos permiten aproximar la presencia de una cultura emprendedora a nivel de la empresa.

Para el caso del *capital de conocimiento*, las fuentes diagnosticadas dan cuenta de una amplia posibilidad de contar con datos para indicadores tradicionales tales como el gasto en actividades de I+D, el uso de distintos mecanismos protección de nuevos desarrollos, el conocimiento científico involucrado en la actividad innovadora, entre otros. Un elemento importante al respecto es la posibilidad de analizar el contenido de los distintos productos de divulgación científica, en particular los trabajos almacenados en los distintos repositorios académicos considerados en el presente trabajo. En efecto, el análisis del contenido de papers mediante algoritmos de tratamiento de datos de texto permitiría precisar de mejor manera el uso de distintas metodologías y técnicas en aquellos trabajos científicos con participación de investigadores e investigadoras nacionales.

En cuanto al capital social, se observan fuentes emergentes que permitirían contar con datos para la implementación de indicadores relativos a la confianza en las instituciones relacionadas directa e indirectamente

8. Conclusiones

con el ecosistema. En particular, la Encuesta Nacional de Percepción Social de la Ciencia y la Tecnología que ya cuenta con su segunda versión para el acceso público. Por otro lado, cuando se considera la dimensión asociativa del capital social, en el presente trabajo se ha decidido tomarla como un nivel de análisis relativo a las conexiones entre los agentes para los distintos tipos de capitales.

Sobre esto último, actualmente se dispone de fuentes que permitirían medir los vínculos entre distintos agentes en el desarrollo de actividades de distinta naturaleza. En particular, los repositorios académicos permiten la representación de datos en red para el análisis de la colaboración en la producción científica, tanto a nivel de autores como de instituciones. Al mismo tiempo, las instituciones pueden ser enlazadas a otras bases de datos que permiten una mejor caracterización de estos vínculos a nivel del Estado, empresa y la academia. Del mismo modo, los distintos sistemas de registro asociados a agencias públicas proveedoras de instrumentos de fomento permiten esta posibilidad de representación de datos en red, tanto en el caso de aquellos programas colaborativos de CORFO, ANID, FIA, FONDART, etc.

Desde el punto de vista de la evolución, gran parte de estas fuentes permite acceder a una trazabilidad en el tiempo, tanto para las empresas, las instituciones proveedoras de conocimiento y el Estado. No obstante, este potencial de trazabilidad en el tiempo es relativo dependiendo de la fuente, ya que en el caso de las distintas encuestas revisadas, al no tratarse de datos de panel, la integración de distintas entidades a lo largo del tiempo tiene como resultado probable una merma no menor en la muestra consolidada. Lo anterior quiere decir que son pocas las empresas que pueden ser integradas en series de tiempo para la exploración de la evolución de aquellas variables más detalladas sobre sus capitales.

En relación con la identificación de preguntas de investigación desafiantes en el ámbito de la CTCI, es importante destacar que se trata de un proceso creativo similar a la generación de productos innovadores, donde se pueden distinguir esfuerzos incrementales y radicales a la hora de proponer nuevas miradas sobre el ecosistema, sus componentes y sus consecuencias. En la propuesta de nuevas preguntas se ha logrado un equilibrio entre ambos espacios, permitiendo el desarrollo de un “portafolio de preguntas” que mantiene un balance entre la búsqueda por mejorar lo que ya existe, pero también identificar oportunidades de hacer cosas de forma diferente. Un desafío considerable para el futuro de la CTCI en Chile es lograr este equilibrio entre, por una parte, mantener datos que permiten la comparabilidad internacional y miden el impacto en la economía de la

8. Conclusiones

ciencia y la tecnología y, por otro lado, promover la creación de datos de nicho con objetivos distintos a lo tradicional.

Las preguntas de investigación identificadas dan cuenta de que la mayoría de los temas tradicionales en CTCI en Chile ya están relativamente resueltos, y el próximo paso implica preguntarse por la direccionalidad de la tecnología, la identificación de impactos no económicos y la identificación de los niveles de cohesión e integración de actores particulares y a nivel general del ecosistema. Para dar respuesta a estas interrogantes, es necesario la promoción de miradas integradoras, diversas, multidisciplinarias y transdisciplinarias, cuyas características son fundamentales considerando la definición de ecosistema de CTCI establecida por el CNID.

El fomento en la creación de nuevos indicadores requiere capacidades de contextualización, focalización y equilibrio entre los portafolios de nuevas formas de medir y las formas de medición actual.

Una interrogante interesante para la política pública en CTCI en los próximos años tiene que ver con la capacidad efectiva de las instituciones locales y las organizaciones de la sociedad civil para crear, promover e interactuar con datos de manera efectiva. Las preguntas de investigación identificadas dan cuenta de la necesidad de crear nuevos datos cuya construcción caerá muchas veces en organizaciones como municipalidades, organismos intermediarios y organizaciones de la sociedad civil, quienes no necesariamente cuentan con el capital requerido para crear, mantener y estandarizar información relevante para el ecosistema. En este sentido, puede ser necesario considerar estructuras de incentivos y financiamiento para la creación de capacidades y recopilación de información en “espacios no tradicionales” para la creación de datos “no tradicionales”, lo que implicaría integrar las capacidades y conocimiento de disciplinas como la ciencia de datos en espacios locales y regionales.

Considerando el total de la información presentada, un ámbito donde podría ser relevante profundizar en estudios posteriores tiene que ver con el rol del análisis de redes y la construcción de datos relacionales para la promoción de políticas públicas en CTCI con sentido para el país. La construcción de sentido a través de datos solo puede fortalecerse a través de la construcción de conexiones, y las conexiones no se construyen en el vacío. El rol de la contextualización de la información, la identificación de conexiones y ausencias de vinculación, y el fortalecimiento del trabajo colaborativo puede apoyar la generación de políticas e instrumentos que promuevan la Ciencia, Tecnología, Conocimiento e Innovación con intencionalidad y direccionalidad en el país.

9. Referencias bibliográficas

9. Referencias bibliográficas

- Ial, C. y Menéndez, M. (2018). Desafíos de monitorear la contribución de la CTCI a grandes retos: Aplicación al desafío país de sostenibilidad del recurso hídrico en Chile. Documento de Trabajo N°8. Consejo Nacional de Innovación para el Desarrollo (CNID). Recuperado de: <http://www.cnid.cl/portfolio-items/desafios-de-monitorear-la-contribucion-de-la-ctci-a-grandes-retos-aplicacion-al-desafio-pais-de-sostenibilidad-del-recurso-hidrico-en-chile/>
- Arocena, R. y Sutz, J. (2016). “Innovación y sistemas nacionales de innovación en procesos de desarrollo,” en *Repensando el desarrollo latinoamericano: una discusión desde los sistemas de innovación*, 1st ed., A. Erbes and D. Suárez, Eds. Los Polvorines: Universidad Nacional de General Sarmiento, pp. 69–102. Recuperado de: https://www.researchgate.net/publication/305720547_Repensando_el_desarrollo_latinoamericano_Una_discusion_desde_los_sistemas_de_innovacion
- Arraiz, I, Henriquez, F & Stucchi, R (2012): “Supplier development programs and firm performance: evidence from Chile” en *Small Business Economics*, vol. 41, 277–293.
- Bravo-Ortega, C., Farren, A.I., Herrera, H., Price, J.J (2016): “Disponibilidad de información con origen en instituciones públicas para estudios estadísticos sobre productividad, desarrollo económico e innovación. Informe Final”, Consejo Nacional de Innovación para el Desarrollo (CNID).
- Banco Mundial. (2010). *La Formulación de Políticas en la OECD: Ideas para América Latina*. Banco Mundial. Recuperado de: <http://documents1.worldbank.org/curated/pt/929251468017387470/pdf/592070WP0SPANII101public10BOX358364B.pdf>
- Berlinski, S. y Schady, N. (2015). *Los primeros años: el bienestar infantil y el papel de las políticas públicas*. Washington: BID. Recuperado de: <https://publications.iadb.org/es/los-primeros-anos-el-bienestar-infantil-y-el-papel-de-las-politicas-publicas>
- Bracho, T. (2010). Políticas basadas en evidencia: la política pública como acción informada y objeto de investigación. En M. Merino, y G. Cejudo, (comps.) *Problemas, decisiones, soluciones. Enfoques de política pública* (pp. 291-319). México: CIDE. Recuperado de: https://www.researchgate.net/publication/272815357_Problemas_decisiones_y_soluciones_Enfoques_de_politica_publica
- Cárdenas, J.P. et al. (2015). *Cartografía del conocimiento en Chile*. Consejo Nacional de Innovación para el Desarrollo (CNID). Recuperado de: <http://www.cnid.cl/portfolio-items/cartografia-del-conocimiento-en-chile/>
- Ciarli, T., & Ràfols, I. (2019). The relation between research priorities and societal demands: The case of rice. *Research Policy*, 48(4), 949-967.
- Consejo Nacional de Ciencia, Tecnología, Conocimiento e Innovación para el Desarrollo (2019). *Ciencia, tecnología, conocimiento e innovación para Chile*. Santiago de Chile. Recuperado de: http://www.cnid.cl/wp-content/uploads/2019/12/CTCI-para-Chile-y-Contexto-para-la-reflexion_web.pdf
- DAMA International (2017). “DAMA-DMBOK: Data Management Body of Knowledge. Second Edition”, Technics Publications, Basking Ridge, New Jersey
- Davies, P. (2004) ‘Is evidence-based government possible?’ Jerry Lee Lecture, presented at the 4th Annual Campbell Collaboration Colloquium, Washington DC. Recuperado de: <https://web.archive.nationalarchives.gov.uk/20091013084422/http://www.nationalschool.gov.uk/policyhub/downloads/JerryLeeLecture1202041.pdf>
- Estrada, E. (2012). *The Structure of complex networks*. Oxford University Press. New York, EUA. Recuperado de: <https://global.oup.com/academic/product/the-structure-of-complex-networks-9780198783800?cc=us&lang=en&>
- Ezeminari, K., Rudqvist, A., y Subbarao, K. (2002). *Impact evaluation concepts and methods*. En *Evaluation and Poverty Reduction*. World Bank.

9. Referencias bibliográficas

- Flores-Crespo, P. (2013). El enfoque de la política basada en evidencia: análisis de su utilidad para la educación de México. *Revista Mexicana de Investigación Educativa*, 265-290. Recuperado de: <http://www.scielo.org.mx/pdf/rmie/v18n56/v18n56a12.pdf>
- Guimona, J., Chaminadeb, C., Maggi, C. & Salazar-Elena, J. (2018): "Policies to attract R&R-related FDI in small emerging countries: aligning incentives with local linkages and absorptive capacities in Chile", *Journal of International Management*, Volume 24, Issue 2, June 2018, 165-178.
- Gouvernement du Québec (2009). *L'école, j'y tiens! Tous ensemble pour la réussite scolaire*. Québec: Ministère de l'Éducation
- JISC (2011) "Data centres: their use, value and impact". UK Research Information Network, noviembre. Recuperado de: http://www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature News*, 520(7548), 429.
- Merino, M. (2010). Problemas, decisiones y soluciones: Enfoques de política pública. México: CIDE. Recuperado de: https://www.researchgate.net/publication/272815357_Problemas_decisiones_y_soluciones_Enfoques_de_politica_publica
- Metcalf, S. and Ramlogan, R. (2008). "Innovation systems and the competitive process in developing economies," *Q. Rev. Econ. Financ.*, vol. 48, no. 2, pp. 433-446. Recuperado de: https://www.researchgate.net/publication/223313103_Innovation_Systems_and_the_Competitive_Process_in_Developing_Countries
- Molas-Gallart, J. (2012). Research governance and the role of evaluation: A comparative study. *American Journal of Evaluation*, 33(4), 583-598.
- OECD (2008). Recommendation on Public Sector Information (PSI). Recuperado de: <http://www.oecd.org/sti/ieconomy/oecdrecommendationonpublicsectorinformationpsi.htm#:~:text=The%20OECD%20Recommendation%20on%20public,-Council%20on%2030%20April%202008.>
- Olivari, J. y Stubrin, L. (2016). "Reflexiones sobre aproximaciones metodológicas y empíricas para el estudio de los sistemas nacionales de innovación," en *Repensando el desarrollo latinoamericano: una discusión desde los sistemas de innovación*, Los Polvorines: Universidad Nacional de General Sarmiento, pp. 143-184. Recuperado de: https://www.researchgate.net/publication/305490282_Reflexiones_sobre_aproximaciones_metodologicas_y_empiricas_para_el_estudio_de_los_sistemas_nacionales_de_innovacion
- Pinilla, R. y Fontcuberta, X. (2010): Evaluación de políticas públicas e innovación social. Un marco para la política basada en la evidencia. *Encuentros multidisciplinares*, 12(35), 21-30. Recuperado de: <http://www.encuentros-multidisciplinares.org/Revistan%BA35/Rafael%20Pinilla%20Palleja%20y%20Xavier%20Fontcuberta%20Estrada.pdf>
- Price Elton (2016). Disponibilidad de información con origen en instituciones públicas para estudios estadísticos sobre productividad, desarrollo económico e innovación. Estudio mandado por CNID. Informe final. Recuperado de: <https://www.cnid.cl/wp-content/uploads/2018/04/disponibilidad-de-informacio%c-c%81n-con-origen-en-instituciones-pu%cc%-81blicas-para-estudios-estadi%cc%81sticos-sobre-productividad-desarrollo-econo%cc%81mico-e-innovacio%cc%81n.pdf>
- Rip, A. (2012). "The Context of Innovation Journeys," *Creat. Innov. Manag.*, vol. 21, no. 2, pp. 158-170. Recuperado de: https://www.researchgate.net/publication/239768583_The_Context_of_Innovation_Journeys

9. Referencias bibliográficas

- Schneider, A. e Ingram, H. (2007). "Public Policy and Democratic Citizenship: What Kinds of Citizenship Does Policy Promote?" En: Fischer, F.; Miller, G. y Sidney, M. *Handbook of Public Policy Analysis. Theory, Politics and Methods*. Boca de Ratón: CRC Press. Recuperado de: https://www.researchgate.net/publication/289692376_Public_Policy_and_Democratic_Citizenship
- Stone, D. (1997). *Policy Paradox, The Art of Political Decision Making*. New York: W.W. Norton and Company. Recuperado de: https://www.researchgate.net/publication/40219337_Deborah_Stone_Policy_Paradox_The_Art_of_Political_Decision_Making
- Susskind, L. (2006). "Arguing, Bargaining, and Getting Agreement". En: Moran, Michael; Rein, Martin; Goodin, Robert. *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press. Recuperado de: https://www.researchgate.net/publication/242538774_Arguing_Bargaining_and_Getting_Agreement
- Sutcliffe, S. y Court, J. (2006) *Herramientas para diseñadores de políticas públicas en países en desarrollo*. Recuperado de: <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3520.pdf>
- Sutcliffe, S. y Court, J. (2005) *Evidence-Based Policy-making: What is it? How does it work? What relevance for developing countries?* Overseas Development Institute. Recuperado de <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3683.pdf>
- Vieytes, R. (2004). *Metodología de la Investigación en Organizaciones, Mercado Sociedad. Epistemología y Técnicas*. Argentina. Editorial de las Ciencias.
- Weber, K. y Truffer, B. (2017). "Moving innovation systems research to the next level: towards an integrative agenda," *Oxford Rev. Econ. Policy*, vol. 33, no. 1, pp. 101–121. Recuperado de: https://www.researchgate.net/publication/316037512_Moving_innovation_systems_research_to_the_next_level_Towards_an_integrative_agenda
- Wessner, C. (2005). "Entrepreneurship and the innovation ecosystem policy lessons from the United States," en *Local Heroes in the Global Village*, Springer, pp. 67–89. Recuperado de: https://link.springer.com/chapter/10.1007/0-387-23475-6_5

10. Anexos

1. Levantamiento de información secundaria

Anexo 1. Papers Seleccionados

a) Fuentes secundarias revisadas, sistematizadas y analizadas

La bibliografía mínima incluida es la propuesta por las bases técnicas de esta licitación:

- Acemoglu, D., P. Aghion, R. Griffith & F. Zilibotti (2010) “Vertical Integration and Technology: Theory and Evidence” Journal of the European Economic Association Vol 8, Issue 5, 989-1033.
- Álvarez, R., Crespi, G., and Cuevas, G. (2012). Public Programs, Firm Performance and Employment: Evidence from Chile. Inter-American Development Bank Technical Notes, IDB-TN-375, 1-46.
- Arráiz, I., F. Henríquez & R. Stucchi (2013). “Supplier development programs and firm performance: evidence from Chile”. Small Bus Econ 41, 277-293.
- Navarro, L. (2018). “Entrepreneurship Policy and Firm Performance: Chile’s CORFO Seed Capital Program”. Estudios de Economía, 45(2).
- Smith, V., M. Dilling-Hansen, T. Eriksson & E. Strojér Madsen (2004) R&D and productivity in Danish firms: some empirical evidence, Applied Economics.
- Roberto Álvarez & Claudia Bravo-Ortega & Lucas Navarro, 2016. “Product mix changes and performance in Chilean plants” Industrial and Corporate Change, Oxford University Press, vol. 25(6), pages 1001-1017.

Paralelamente, la revisión bibliográfica se complementó con los siguientes textos:

- Alghamdi, M. (2019): “Individual and social incentives versus R&D network restriction”, *Journal Korean Society of Applied Mathematics*, Vol.23, No.4, 329–350.
- Amir R., Yin J. & Troege M. (2008): “On additive spillovers and returns to scale in R&D”, *International Journal of Industrial Organization*, 26 (2008) 695–703
- Bravo-Ortega, C., Benavente, J. M., & González, Á. (2014). Innovation, exports, and productivity: Learning and self-selection in Chile. *Emerging Markets Finance and Trade*, 50(sup1), 68-95.
- Caloffi, A., Mariani, M., Rossi, F., & Russo, M. (2018). A comparative evaluation of regional subsidies for collaborative and individual R&D in small and medium-sized enterprises. *Research Policy*, 47(8), 1437-1447.
- Cassiman B & Veugelers R, (2002): “R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium” *The American Economic Review*, Vol. 92, No. 4 (Sep., 2002), pp. 1169-1184.
- Guimon, J., Chaminadeb, C., Maggi, C. & Salazar-Elena, J. (2018): “Policies to attract R&R-related FDI in small emerging countries: aligning incentives with local linkages and absorptive capacities in Chile”, *Journal of International Management*, Volume 24, Issue 2, June 2018, Pages 165-178.
- Mañez, J., Rochina, M.E., Sanchis, A., Sanchis, J (2009): “The Role of Sunk Costs in the Decision to Invest in R&D”, *Journal of Industrial Economics* 57(4):712-735
- Meijaard, J., Brand, M. J., Mosselman, M. (2005), “Organizational Structure and Performance in Dutch Small Firms”. *Small Business Economics*, Vol. 25, 83:96

10. Anexos

- Monjon, S., & Waelbroeck, P. (2003). Assessing spillovers from universities to firms: evidence from French firm-level data. *International Journal of Industrial Organization*, 21(9), 1255-1270.
- Rojas, M. G. A., Solis, E. R. R., & Zhu, J. J. (2018). Innovation and network multiplexity: R&D and the concurrent effects of two collaboration networks in an emerging economy. *Research Policy*, 47(6), 1111-1124.
- Sahaym A, Steensma HK & Barden JQ, (2010): "The influence of R&D investment on the use of corporate venture capital: An industry-level analysis", *Journal of Business Venturing*, Vol 25, Issue 4, 376-388.
- Schiavone, F. & Simoni, M. (2011): "An experience-based view of co-opetition in R&D networks", *European Journal of Innovation Management*, Vol. 14 No. 2, 136-154.
- Stuart T & Sorendon O, (2005): "Strategic networks and entrepreneurial ventures", *Strategic Entrepreneurship Journal*, Vol. 1, Issue 3-4, 211:227.

b) Sistematización de información secundaria

Para el levantamiento de información secundaria, se consideraron dos unidades de análisis. En primer lugar, la experiencia en investigación que consideró la sistematización de la lectura de los papers establecidos en las bases de licitación, más aquellos sugeridos en la propuesta que buscan responder y caracterizar el ecosistema a nivel nacional e internacional, priorizando una pluralidad en las metodologías utilizadas y los aspectos en los que se definen tanto la línea de base como las brechas identificadas, particularmente en torno a los indicadores de conexión. Otro aspecto considerado para la selección de los papers corresponde al marco temporal en cual fueron escritos para analizar si existen preguntas de fronteras que eran más recurrentes en algunos años y no en otros, para así determinar si el entorno académico influye en la formulación de preguntas y temas analizados.

Para el registro sistematizado de las dimensiones relevantes de la experiencia en investigación, se diseñó e implementó una matriz de vaciado de contenido documental, la que consideró los siguientes campos:

10. Anexos

Tabla 12. Campos de registro ficha de sistematización bibliográfica

Campo	Observaciones
Título	Título del paper
Año	Año de publicación
Autores	Investigadores
Journal	Revista o Journal de publicación
Palabras Claves	Permiten implementar la clasificación temática
Clasificación Temática	No se utilizó la clasificación del JET (Journal of Economic Theory) para dotar de mayor amplitud a la categorización
Pregunta de Investigación	Permite conocer el nivel de profundidad de los objetivos de investigación
Hipótesis	
Método	permite conocer la profundidad en el análisis
Unidades de Análisis	Agentes del ecosistema considerados
Variables Consideradas	Según su variedad para abordar temas emergentes
Fuentes de Datos	Encuestas, registros, etc.
Fuente Pública	En caso de que aplique, detallar cuáles
Repositorio	
Resultados	Verificación de hipótesis y recomendaciones de política

Fuente: Elaboración propia.

Anexo 2. Repositorios seleccionados

Se sistematizaron las características de distintos repositorios de datos abiertos, tanto a nivel nacional como internacional. Respecto del análisis de los repositorios de datos abiertos, se elaboró una ficha de registro para la sistematización con los siguientes campos:

Tabla 13. Campos de registro ficha de sistematización de repositorios

Campo	Observaciones
Nombre Repositorio	Nombre del repositorio consultado
Objetivos	Propósito que se busca cumplir con tener un repositorio de datos
Instituciones Participantes	Instituciones que colaboran para la realización del repositorio, ya sea con bases de datos o con una colaboración gubernamental
Acceso	Formas de acceder a la información
Seguridad	Datos que no permitan identificar personas o entidades de manera individual
Formato de Datos	Tipo de dato que se encuentra en el repositorio, ejemplo: Series de Tiempo
Formato de descarga	Tipo de archivos que se encuentran disponibles para descargar
Fuente Estructurada- Encuestas	Datos pertenecientes a encuestas
Fuente Estructurada- Registros Administrativos	Datos pertenecientes a registros administrativos
Fuente no Estructurada	Datos no estructurados como imágenes, audio, archivos multimedia

Fuente: Elaboración propia.

10. Anexos

Los repositorios analizados para esta investigación corresponden a:

- UK Data Service , Reino Unido.
- Centre for European Economic Research (ZEW).
- Research Data Centre de Alemania.
- The Boston Census Research Data Center (BRDC), Estados Unidos.
- Research Data Centres Program (RDC), Canadá.
- Korean Social Science Data Archive (KOSSDA), Corea.
- Korea Open Research Data (KORD Repository), Corea.
- China Data Institute.
- China Data Lab.
- National Bureau of Statistics, China.
- Japan Data Repository Network (JDARN).
- CEPAL
- BID
- ILDA
- Datos abiertos de la República del Perú
- Datos abiertos Argentina
- Datos abiertos Colombia
- Datos abiertos México
- Datos abiertos Brasil
- SII
- ADUANAS
- INAPI
- Datos.gob
- Energía Abierta
- Economía Abierta
- ANID
- Data Ciencia
- Data Emprendedor

2. Levantamiento de información primaria

Anexo 3. Trabajo de campo

Se realizaron 19 entrevistas entre los días 10 de diciembre de 2020 a 15 de enero 2021. La mayoría de las entrevistas logradas (19) pertenece al perfil de “Análisis, promoción o ejecución de las políticas públicas”, debido principalmente a que el acceso al perfil de “Generación, almacenamiento, procesamiento y disposición de datos” (7 entrevistas) resultó ser más difícil de acceder que lo presupuestado inicialmente, ya que la diversidad organizacional de las instituciones involucradas no permitió identificar de manera directa a los informantes clave relacionados con los sistemas de información.

c) Muestra de actores entrevistados/as: Perfil tradicional

Tabla 14. Actores relevantes a entrevistar, perfil tradicional

N°	Perfil	Tipo	Institución/Organización	Nombre	Cargo
1	Análisis, promoción o ejecución de las políticas públicas	Entidad pública	Ministerio de Ciencia Tecnología Conocimiento e Innovación	María José Bravo	Jefa de División de Estudios y Estadísticas
2		Asociación de investigadores	Red Chilena de investigadores en innovación	Daniel Goya	Investigador
3		Experto/a en Big Data	GobLab UAI	María Paz Hermosilla	Directora de GobLab UAI
4		Empresa	Deloitte	Nicolás Tagle	Manager en OMNIA-AI
5		Emprendimiento de base científica y/o tecnológica	Metric Arts	Patricio Cofré	Co-Fundador
6		Hub de transferencia tecnológica	Hub Tec Chile	Bernardita Araya	Directora Ejecutiva

10. Anexos

N°	Perfil	Tipo	Institución/Organización	Nombre	Cargo
7		Centro I+D ANID	Centro de Modelamiento Matemático (CMM)	Roxana Trujillo	Investigadora encargada del Área Geoespacial
8		Centros financiados por ICM	Instituto Fundamento de los Datos	Pablo Barceló	Director Alterno
9		ITP	Fundación Chile	Ricardo Morgado	Director de estrategia y Desarrollo-Expande Fundación Chile
10		Entidad pública	Fundación para la Innovación agraria (FIA)	Dominique Chaveau	Jefa Unidad de Desarrollo Estratégico
11		Entidad pública	Departamento de Computación (DCC) de la Universidad de Chile	Bárbara Poblete	Profesora Asociada
12		Incubadora apoyadas por Corporación de Fomento de la Producción (Corfo)	Incuba UC	Francisca Rojas	Subdirectora de Emprendimiento
13	Generación, almacenamiento, procesamiento y disposición de datos	Entidad pública	Comisión de Mercados Financieros	Walter Illanes	Jefe Unidad de Estadística de la Dirección de Estudios
14		Entidad pública	Superintendencia de Seguridad Social	José Inostroza	Experto en datos
15		Entidad pública	Fundación para la Innovación Agraria (FIA)	Constanza Mantelli	Jefa de Unidad de Información y Difusión
16		Entidad pública	Instituto Nacional de Propiedad Industrial (INAPI)	Álvaro González	Encargado Unidad Inteligencia de Negocios
17		Entidad pública	CORFO	Brenda Rain	Encargada de Unidad de Inteligencia de Negocios

10. Anexos

N°	Perfil	Tipo	Institución/Organización	Nombre	Cargo
18		Entidad pública	CORFO	Leonor Saravia	Ejecutiva Gerencia de Emprendimiento
19		Entidad pública	ANID	Bastían Girardi	Jefe de Gabinete ANID
20	Análisis, promoción o ejecución de las políticas públicas	Experto internacional	Leiden University, Centre for Science and Technology Studies	Ismael Rafols	Investigador Senior Leiden University, Centre for Science and Technology Studies
21		Experto internacional	Matías Ramírez (SPRU)	Matías Ramírez (SPRU)	Líder Hub Latinoamericano y Caribeño de Innovación Transformativa
22		Asociación de investigadores	Asociación de investigadores en Artes y Humanidades	Carolina Gainza	Fundadora Asociación de Investigadores en Artes y Humanidades
23		Asociación de investigadores	Red de estudios sociales de Ciencia y tecnología	Jorge Gibert	Profesor Titular, Escuela de Ingeniería Comercial, Facultad de Ciencias Económicas y Administrativas, Universidad de Valparaíso
24		Centro de estudios	Capital Biodiversidad	Alejandra Figueroa	Directora organización sin fines de lucro
25		Centro de estudios	ITREND (Instituto para la resiliencia ante desastres)	Catalina Undurraga	Directora de ITREND
26	Generación, almacenamiento, procesamiento y disposición de datos	Asociación de empresarios/as	Sistema B Chile	Zdenka Astudillo	Directora Ejecutiva

Fuente: Elaboración propia.

10. Anexos

En términos de la ejecución de la pauta de entrevista acordada inicialmente, la aplicación de las entrevistas no provocó mayores cambios en los contenidos tratados. El principal apoyo metodológico desarrollado durante el proceso de aplicación de entrevistas consistió en la sistematización preliminar de temas recurrentes, lo que permitió facilitar el proceso de entrevista a través de ejemplos concretos que surgieron en conversaciones previas.

Una de las principales dificultades durante el proceso de construcción de información fue la generación de respuestas directas frente a las preguntas que requerían un esfuerzo creativo mayor. En particular, debido a que uno de los objetivos de las entrevistas era recopilar ideas sobre nuevos datos o indicadores relevantes, las respuestas de las entrevistas permiten identificar brechas o escenarios ideales más que preguntas de investigación o propuestas directas.

Debido a las expectativas altamente creativas del proceso de entrevistas, es decir, la necesidad de co-crear con los entrevistados nuevos indicadores o datos para el ecosistema de CTCI, el proceso de análisis se enfoca principalmente en la búsqueda de diversidad por sobre saturación del discurso. Aun así, aparecen temas comunes, que permiten entregar relevancia a algunos de las preguntas de investigación sistematizadas con posteridad.

Anexo 4. Instrumentos para el levantamiento de información cualitativa

La diferencia identificada entre los distintos perfiles de potenciales entrevistados requiere una aproximación diferenciada a la hora de definir las preguntas que serán aplicadas a cada uno de ellos.

Las pautas de entrevista fueron construidas en base a información de la propuesta original, los resultados del informe 1 entregado con anterioridad, documentación del Consejo Nacional de Ciencia, Tecnología, Conocimiento e Innovación para el Desarrollo y la experiencia del equipo de investigación en temas relacionados.

La estructura general de las pautas de entrevista sigue una distinción tradicional en metodologías de investigación cualitativa en términos de la identificación de dimensiones y subdimensiones de análisis, que conllevan posteriormente preguntas específicas a realizar con los entrevistados.

Las dimensiones y subdimensiones de la pauta de entrevista para el perfil de entrevistados correspondiente a “análisis, promoción o ejecución de las políticas públicas” buscan identificar las percepciones de los entrevistados

10. Anexos

respecto al funcionamiento y disponibilidad de información del ecosistema de CTCI en Chile. También existen secciones orientadas a identificar espacios no tradicionales donde podría existir ausencia de información o datos no considerados en la actualidad.

La sección final está dedicada a la profundización de los aspectos generales considerados al inicio de la entrevista. Las dimensiones y subdimensiones de esta sección están inspiradas en el documento “Desafíos de monitorear la contribución de la CTCI a grandes retos: Aplicación al desafío país de sostenibilidad del recurso hídrico en Chile”, publicado por CNID. En particular, la caracterización del ecosistema de CTCI en base a las dimensiones de estructura, funcionamiento, desempeño y evaluación permiten especificar percepciones y evaluaciones generales sobre el ecosistema para lograr detalles relevantes para la investigación.

Tabla 16. Pauta de entrevista para actores del perfil “Análisis, promoción o ejecución de las políticas públicas”

Dimensión	Sub Dimensión	Pregunta
Percepciones del ecosistema de CTCI chileno	Evaluación general del ecosistema de CTCI en Chile	¿Cuál es su evaluación general del ecosistema de CTCI en Chile? ¿Cree que hay algunas ausencias respecto, por ejemplo, algunos actores, organizaciones o instituciones que deberían estar y no están?
	Fortalezas del ecosistema de CTCI en Chile	¿Cuáles son las fortalezas del ecosistema de CTCI en Chile?
	Debilidades del ecosistema de CTCI en Chile	¿Cuáles son las debilidades del ecosistema de CTCI en Chile?
	Existencia y disponibilidad de información sobre el ecosistema de CTCI en Chile	¿Cuál es su opinión respecto a los datos disponibles para evaluar, promover o analizar el sistema de CTCI en Chile? ¿Son suficientes? ¿Su disponibilidad es adecuada?
		En su experiencia ¿Se ha encontrado en la posición en que le gustaría que un dato específico estuviera disponible o existiera?
	Recomendaciones de mejoras para el ecosistema de CTCI en Chile	¿Qué recomendaciones de mejoras puede dar para el ecosistema de CTCI en Chile?

10. Anexos

Dimensión	Sub Dimensión	Pregunta
Percepciones sobre indicadores del ecosistema de CTCI chileno	Identificación de ausencia de información sobre indicadores tradicionales del ecosistema de CTCI en Chile	¿Crees que existen datos o indicadores sobre el ecosistema de CTCI que en Chile no se producen o construyen?
	Recomendación de indicadores no tradicionales para la caracterización del ecosistema de CTCI chileno	¿Crees que existe o debería existir algún tipo de dato o indicador para evaluar, promover o analizar el sistema de CTCI en Chile distinto a los tradicionales (número de patentes, gasto en I+D, etc.?)
	Identificación de fuentes de información existentes en Chile que podrían ser incorporadas en la caracterización del ecosistema de CTCI chileno	¿Podrías recomendar alguna fuente de información o base de datos disponible en Chile o en otros países para evaluar, promover o analizar el sistema de CTCI?
Evaluación del ecosistema de innovación (Elementos de profundización)	Evaluación de la estructura del ecosistema (nodos y relaciones)	¿Cuál es su evaluación de la instituciones, organizaciones, individuos y vínculos presentes en el ecosistema? ¿Crees que existen ausencias o esfuerzos replicados en algún área?
	Evaluación del funcionamiento del ecosistema (organización)	¿Cuál es su evaluación del funcionamiento del ecosistema, en términos de su organización?
	Evaluación del desempeño del ecosistema	¿Cuál es su evaluación del desempeño del ecosistema en la actualidad? ¿Qué crees que explica este desempeño general?
	Evaluación de la evolución del ecosistema	¿Cuál es su evaluación de la evolución del ecosistema en los últimos 20 años?

Fuente: Elaboración propia.

10. Anexos

Por su parte, las dimensiones y subdimensiones de la pauta de entrevista para el perfil de entrevistados correspondiente a “generación, almacenamiento, procesamiento y disposición de datos” están basadas en los criterios desarrollados para las herramientas de diagnóstico de gobierno de datos derivadas del DMBOK. Estas dimensiones aluden al hexágono de factores ambientales, en el que convergen tecnologías, personas y procesos articulados en torno a las metas y principios de cada área de gestión del conocimiento.

Tabla 17. Pauta de entrevista para actores del perfil “Generación, almacenamiento, procesamiento y disposición de datos”

Dimensión	Subdimensión	Pregunta
Drivers de Negocio	Metas	¿Cuáles son las metas que debe cumplir su unidad?
	Principios	¿Cuáles son los principios que inspiran el quehacer de su unidad?
Procesos	Inputs con los que trabaja la unidad/división	¿Cuáles son los inputs con los que trabaja la unidad/división?
	Actividades de Planificación	¿Cuáles son las principales actividades de planificación de su unidad?
	Actividades de Control	¿Cuáles son las principales actividades de control de su unidad?
	Actividades de Desarrollo	¿Cuáles son las principales actividades de desarrollo de su unidad?
	Actividades de Operacionales	¿Cuáles son las principales actividades operacionales de su unidad?
	Entregables que proporciona la unidad/división	Qué tipo de entregables proporciona su unidad/división?

10. Anexos

Dimensión	Subdimensión	Pregunta
Personas	Proveedores de los inputs con los que trabaja la unidad/división	¿Quiénes son los proveedores de los inputs con los que trabaja su unidad/división?
	Participantes en las distintas Actividades	¿Qué actores participan en las distintas actividades que desarrolla su unidad?
	Consumidores relacionados a los entregables de la unidad/división	¿Quiénes son los consumidores relacionados a los entregables de su unidad/división?
Drivers Técnicos	Técnicas: procedimientos para la ejecución de actividades	¿Qué técnicas (procedimientos) se utilizan para la ejecución las actividades de su unidad/división?
	Herramientas: software y servicios similares involucrados en la ejecución de actividades	¿Qué herramientas (software y servicios similares) están involucrados en la ejecución de las actividades de su unidad/división?
	Métricas que orientan el trabajo de la unidad/división	¿Cuáles son las métricas que orientan el trabajo de su unidad/división?

Fuente: Elaboración propia.



CONSEJO NACIONAL
DE CIENCIA, TECNOLOGÍA,
CONOCIMIENTO E INNOVACIÓN
PARA EL DESARROLLO