

## PROGRAMA DE FORMACIÓN COMPETITIVIDAD

### FORMACIÓN DE CAPITAL HUMANO EN DATA SCIENCE EN EL CAMPO DE LA ASTROINFORMÁTICA

Roberto González<sup>1</sup>, María Inés Álvarez<sup>1</sup>, Paola Bordon<sup>2</sup>, Andrés Escala<sup>3</sup>, Francisco Förster<sup>3</sup> y Joaquín Prieto<sup>3</sup>

*Estudio Encargado y Financiado por CORFO, Iniciativa Desarrollada en Conjunto con el Comité de Transformación Digital y su Proyecto Astro Data. Agente Operador: Asociación de Exportadores de Frutas de Chile (ASOEX).*

#### Resumen

Este documento *White Paper* resume el primer diagnóstico y caracterización de la situación actual de formación de capital humano en ciencia de datos y Astroinformática en Chile, identificando los perfiles de competencia que componen la cadena de valor de la Astroinformática, para generar planes formativos junto a recomendaciones que garanticen la adecuada articulación con el sector productivo.

Chile es un polo astronómico de nivel mundial. Esto, gracias a la calidad de sus cielos, lo que ha permitido la existencia de una gran cantidad de observatorios astronómicos internacionales e instituciones que forman estudiantes de pre y posgrado, en donde el astrónomo, y en particular el astroinformático, se han visto beneficiados técnicamente, adquiriendo amplia experiencia en el manejo y análisis de grandes volúmenes de datos generados por observatorios y simulaciones computacionales. Dichas habilidades, presentes naturalmente en nuestro país, son fundamentales para la práctica de la ciencia de datos, y deben ser consideradas como una ventaja comparativa enorme, a nivel regional, pues disponibiliza conocimiento y capital humano especializado difícil de encontrar en otros países. En el escenario en que Chile albergará sobre el 55% del área colectora de telescopios (reflectores) a nivel mundial en los próximos años, la Astroinformática es llamada a ser un motor diferenciador en aspectos de formación tecnológica, relativos a transformación digital y ciencia de datos, con los observatorios jugando el papel de laboratorios de especialización en manejo de grandes volúmenes de datos.

El presente estudio, enfocado en la región de Coquimbo, se refiere al contexto nacional en torno a la Astroinformática, e identifica las capacidades y oportunidades que posee la IV región en relación a este tema. En particular, se resaltan aquellas que puedan convertir a la región en un polo de desarrollo en el área de la ciencia de datos, como el hecho de que esta posee el 25% de la capacidad astronómica de Chile y junto al LSST producirá más datos que cualquier otro telescopio antes construido en el mundo. El estudio incluye levantamiento de información propia mediante una encuesta, entrevistas y una mesa de trabajo; Presenta perfiles ocupacionales, oferta/demanda formativa y laboral, benchmarking internacional, planes formativos, estudio de potencial regional, y un plan de acción y difusión.

---

<sup>1</sup> Pontificia Universidad Católica

<sup>2</sup> Universidad Alberto Hurtado

<sup>3</sup> Universidad de Chile

Las condiciones naturales del norte de Chile, han permitido la instalación de grandes observatorios astronómicos convirtiendo al país en la capital mundial de la astronomía. Esto se ve reflejado no sólo en la gran productividad científica asociada, sino que también en el surgimiento de nuevas áreas de investigación con potencial de transferencia de conocimiento hacia otros sectores de la sociedad. Una de ellas, la **Astroinformática**, ha sido pionera en la aplicación de técnicas de ciencia de datos a grandes volúmenes, así como también en la formación de capital humano avanzado con conocimiento práctico en ciencia de datos dentro de nuestro país. La relevancia de esta disciplina seguirá en aumento en los próximos años ya que Chile albergará sobre el 55% del área colectora (reflectores) mundial, con la puesta en marcha de LSST, GMT y E-ELT.

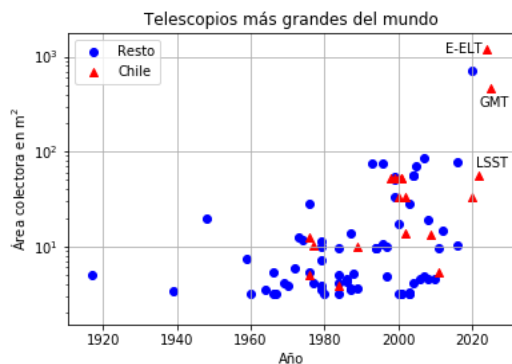


Fig. 1: Área colectora nominal de los telescopios reflectores con diámetro mayor a 2 metros, en función del tiempo del comienzo de su operación.

La **transformación digital**, la digitalización y automatización transversal de procesos, permite el aumento de la eficiencia y productividad del sector industrial, comercial y público. Para que esto ocurra se necesita la incorporación de nuevas tecnologías asociadas a ciencia de datos, lo cual requiere de capital humano avanzado con nuevas capacidades técnicas y organizacionales, escasamente reflejadas en los planes de

formación educacionales del país. Por esta razón, la astronomía y la Astroinformática ofrecen una ventaja estratégica, que no se puede desaprovechar, teniendo como objetivo la aceleración de la transformación digital del país[1].

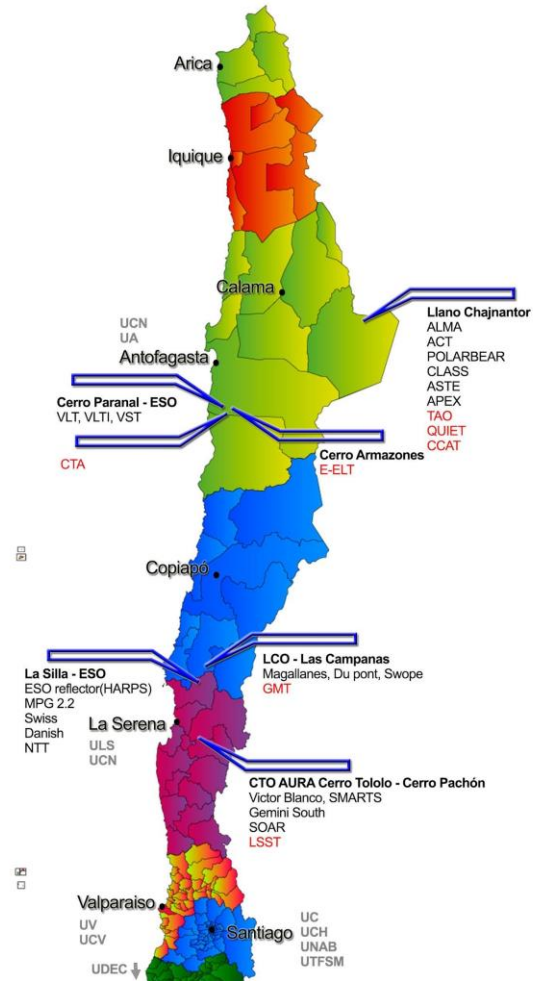


Fig. 2: Observatorios y telescopios instalados en Chile. En rojo se indican los telescopios en construcción y en gris las Universidades que imparten carrera de pregrado en Astronomía.

Este estudio corresponde al primer trabajo realizado que aborda y reporta la situación actual de formación de capital humano en ciencia de datos en Chile y se construye realizando un diagnóstico y caracterización de la formación en ciencia de datos y Astroinformática, logrando identificar las

competencias y conocimientos en común que poseen estas disciplinas. Para esto, se utilizó como estructura la **cadena de valor de la Astroinformática**[2], estructura que involucra cuatro componentes o eslabones: adquisición y generación de datos, acceso y gobernanza de datos, análisis de datos, y visualización y exploración de datos.



Fig. 3: Cadena de Valor de la Astroinformática.

Si bien el presente estudio se refiere al contexto nacional, este se enfoca en la IV región de Coquimbo, identificando las capacidades y oportunidades existentes en ella y, en particular, aquellas que puedan convertirla en un polo de desarrollo en el área de ciencia de datos, como el hecho de que la región posee el 25% de la capacidad astronómica de Chile y que junto al LSST producirá más datos que cualquier otro telescopio antes construido. Concretamente, se hace un catastro del capital humano formado en ciencia de datos, se identifican perfiles de formación y la oferta existente, se caracteriza la demanda en distintas áreas: industria, academia, centros de investigación y observatorios, para, finalmente, realizar las recomendaciones necesarias que garanticen la articulación entre el capital humano y el sector productivo.

La cantidad y complejidad de los datos producidos tanto en la industria como en los observatorios se incrementará dramáticamente los próximos años, generando la necesidad de

uso de nuevas tecnologías y de la ciencia de datos. La demanda laboral de personas con conocimientos en ciencia de datos, a su vez, ha crecido un factor cinco, aproximadamente, en los últimos cinco años, a nivel mundial. Esto ha derivado en la proliferación de programas formativos y de especialización en ciencia de datos desde hace 6-8 años en diferentes partes del mundo. En este sentido, Chile no se ha quedado atrás, sin embargo la oferta formativa, en programas cortos, como diplomados y cursos, ha comenzado muy recientemente hace no más de tres años, y la demanda desde la industria se ha ido adaptando lentamente a estos nuevos roles y requerimientos, producto de la transformación digital.

Al ser este uno de los primeros estudios sobre el tema, la información disponible es limitada y se requiere un **Levantamiento de información** adicional a través de encuestas, entrevistas a actores clave y mesas de trabajo.

---

**Encuesta nacional:** Al no existir fuentes sobre los científicos de datos en el país, se hizo necesario diseñar y realizar una encuesta exploratoria sobre los grupos más visibles en el área de ciencia de datos y la comunidad en torno a la Astroinformática en Chile. El objetivo central de ello, fue recopilar información sobre la formación académica, situación laboral, salarios y competencias. El universo total de la encuesta es de aprox. 5 mil personas, incluyendo titulados en ciencia de datos, miembros de la comunidad en línea *Big Data & Data Science en Chile* (grupo más numeroso y activo encontrado), SOCHIAS, SOCHIFI, observatorios, profesionales de departamentos de astronomía en Chile y profesionales de la industria relacionados con ciencia de datos en Chile. A pesar de lo valiosa que es esta encuesta, se hace notar que es de carácter exploratorio y no pretende ser un censo

de científicos de datos en el país, desafío que escapa de los objetivos de este estudio.

En la encuesta se registraron 577 respuestas y se encontró una disparidad de género similar a la que se observa en las comunidades de datascience internacionales (kaggle 2018). El 38% proviene del área de computación e informática. El 18% proviene del área de física o astronomía. La edad promedio de los encuestados es de 34 años. El 75% de los encuestados se concentraron en la región metropolitana y sólo un 3.7%, en la región de Coquimbo.

El 84% estudió en una universidad chilena. El 21% tiene doctorado y el 27%, magíster. El 65% de los encuestados ha tomado cursos de formación en línea. Sólo un 22% considera que adquirió más del 60% de los conocimientos útiles en su trabajo en el pregrado. Un 55% trabaja en el sector privado, 29% en academia, 9% en el sector público, 3% en observatorios, 2.5% en centros de investigación privados y 2% en otras instituciones. En relación a la cadena de valor de la Astroinformática, el 48% trabaja en análisis de datos, 17% en adquisición y generación de datos, 14% en visualización y exploración, 7% en acceso y gobernabilidad de datos y 15% en otras actividades.

De forma paralela se describe el grupo de encuestados que no provienen del grupo de la red social Facebook. Este sub-conjunto, que corresponde al 24% del universo total de encuestados, está dominado por investigadores y académicos. Posee una proporción de mujeres mayor a la del universo total de encuestados y tiene por cierto, un perfil internacional. La mayoría de sus integrantes (85%) posee estudios de posgrado y el área de posgrado más común (70%) es en física o astronomía. Dada la

formación de este sub-conjunto de encuestados, la gran mayoría se dedica a el análisis de datos.

**Entrevistas:** Se realizaron entrevistas a actores clave en el área de ciencia de datos y Astroinformática. Las entrevistas permitieron comprender las experiencias formativas y laborales cruciales de cada entrevistado, y revelaron la brecha existente entre el mundo formativo y laboral, así como la demanda de profesionales en ciencia de datos y Astroinformática. Cuando fue posible se identificaron las capacidades y oportunidades de la región de Coquimbo, con un foco en la definición de perfiles ocupacionales especializados que requiere la región.

Los entrevistados fueron seleccionados de una lista de 75 actores clave, de construcción propia, divididos en distintas áreas: actores claves nacionales de la industria y de la academia; y actores claves internacionales de los observatorios, de la industria, y de la academia. Se entrevistó a 14 actores clave. Se destacan frases claves y conclusiones principales, y se analiza la composición de las entrevistas en términos de frecuencias de palabras e ideas clave, las cuales son de utilidad especial para el benchmarking, diseño de perfiles, planes formativos, y el plan de acción.

**Mesas de trabajo:** Se organizaron mesas de trabajo para generar una discusión con actores relevantes en la región de Coquimbo. Se abordaron temas relacionados con la formación académica en ciencia de datos y Astroinformática, transferencia tecnológica e inserción de capital humano avanzado en la industria. Las mesas de trabajo lograron generar una discusión en torno a las temáticas planteadas, destacando el tema relacionado con qué directrices seguir para generar cambios en los programas existentes o bien crear

programas nuevos que incluyan ciencia de datos en un contexto compatible con la demanda regional y nacional. Otro tema, relevante para los participantes fue el relacionado con las dificultades de compatibilizar y articular iniciativas entre la industria y la academia, evidenciando las barreras que existen actualmente para generar ese vínculo.

**Perfiles ocupacionales:** Se presentan los perfiles ocupacionales técnicos y profesionales relacionados al área de ciencia de datos y Astroinformática. En primer lugar, se detallan los perfiles reportados para 19 carreras relacionadas dictadas por diferentes centros de formación técnicos y profesionales, y su relación con cada uno de los componentes de la cadena de valor de la Astroinformática.

En segundo lugar, se identifican las características ocupacionales para cada una de las componentes de la Astroinformática. En base a los resultados de la encuesta nacional se analizaron las tareas más realizadas en el trabajo u ocupación, en función del más alto nivel educacional alcanzado por el trabajador, de acuerdo a la cadena de valor de la Astroinformática. Se encontró que la tarea más demandada de la cadena de valor es el análisis de datos. Los titulados de carreras de más de 6 años (generalmente ingenierías civiles), o trabajadores con grado de magíster y doctorado son quienes realizan más a menudo este tipo de tareas. Tanto en pregrado como en postgrado se observa un continuo desde un extremo donde expertos en computación e informática distribuyen su trabajo de forma relativamente uniforme en la cadena de valor de la Astroinformática, hasta otro extremo donde estadísticos, físicos y astrónomos concentran

alrededor del 70% de su tiempo en el análisis de datos.

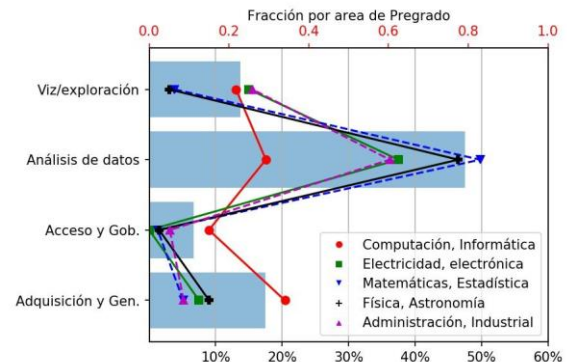


Fig 4. Tareas realizadas en el trabajo según el área de estudio de pregrado

También se analizaron las competencias en el área de ciencia de datos. Los encuestados con postítulo, diplomado o MBA y magíster son los que en mayor proporción presentan competencia en habilidades relacionadas a la nube, reportería y *Business Intelligence*, y bases de datos. Los magísteres y doctorados destacan en competencias en lenguajes de bajo, medio y alto nivel, sistemas paralelos y big data, y machine learning. Se observa una distribución uniforme de habilidades requeridas para la adquisición y generación de datos, pero una concentración de habilidades relacionadas a lenguajes de alto nivel, computación paralela y big data, y machine learning para el análisis de datos. En general, una fracción muy pequeña de los encuestados dice haber recibido más del 60% de los conocimientos requeridos para su trabajo en su formación de pregrado, desde un 8% en ciencias sociales o 10% en electricidad y electrónica hasta un 30% en física y astronomía.

Posteriormente, y en base a los resultados obtenidos de los ejercicios descritos anteriormente, se presentan los perfiles ocupacionales basados en la metodología ChileValora para técnicos y profesionales en las

cuatro áreas de la Astroinformática. Estos perfiles desde su concepción están pensados en formar profesionales cuya formación y orientación se construye entorno a los datos, capaces de generar soluciones y valor agregado a los datos en cualquier escenario. Esto lo diferencia de los perfiles actuales de ciencia de datos, cuya formación está orientada a utilizar y analizar los datos hacia un área específica del conocimiento o negocio. (i.e. Analista de datos médicos)

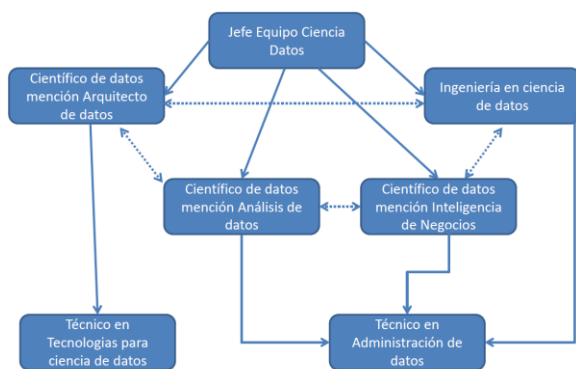


Fig 5: Perfiles laborales profesionales y técnicos relacionados a la Astroinformática y data science. Diagrama de articulación de los perfiles en un equipo de ciencia de datos.

Para el nivel profesional se describen 4 perfiles del científico de datos: (1) **Científico de datos mención Análisis de datos**: orientado a las técnicas de análisis de datos, en donde el objetivo es extraer la mayor información y valor desde los datos. Es el núcleo de todo grupo de ciencia de datos; (2) **Científico de datos mención Arquitectura de datos**: orientado a diseñar la infraestructura y montar las plataformas asociadas a todos los aspectos de la cadena de la Astroinformática: la adquisición, generación, manejo y acceso de los datos, herramientas para análisis y exploración, herramientas de visualización; (3) **Ingeniería en ciencia de datos**: Orientado al desarrollo de soluciones productivas y escalables asociadas a la cadena de

la Astroinformática. Especialista en el manejo correcto, estandarización, e interoperabilidad de los datos (gobernanza); y (4) **Científico de datos mención Inteligencia de negocios**: orientado al análisis de datos en el ámbito específico de los negocios, con fuerte énfasis en la exploración de datos, en la comunicación de resultados, reportería, y visualización.

Para el nivel técnico se describen 2 perfiles del científico de datos: (1) **Técnico en Administración de datos**: orientado a técnicas operacionales para administrar, gestionar, mantener sistemas de datos que son utilizados en la cadena de data science. Su labor es acceder, administrar, curar, disponibilizar, estandarizar y dar seguridad a los datos; (2) **Técnico en tecnologías para ciencia de datos**: posee conocimiento técnico en las tecnologías utilizadas en la cadena de data science. Su labor es ensamblar, operar y administrar la infraestructura computacional, y las plataformas orientadas a manejo de grandes volúmenes de datos tanto de forma local como en la nube, que suelen ser usadas en el ámbito de la Astroinformática, pero de uso y acceso limitado en la informática tradicional.

Los perfiles descritos anteriormente, permitirían no sólo satisfacer la demanda de capital humano en data science, sino que también permitirían una mejor asignación de recursos humanos en términos de competencias y tareas para cada una de las componentes de la cadena de valor de la Astroinformática, lo que impactaría directamente en la productividad de las empresas, academia y observatorios. Por otra parte, se destaca que estos perfiles están en línea con la diversificación de la ciencia de datos en sub-ramas especializadas. Esta diversificación se observa en países desarrollados, donde los roles de ciencia de datos tienen mayor madurez en el mercado [4]. De esta forma, la

especialización observada en países desarrollados se ha traspasado al esquema educativo nacional, donde se diferencian particularmente las carreras técnicas y profesionales.

Finalmente, se presentan los marcos de cualificaciones complementarios a los perfiles y se realiza una caracterización de los perfiles diseñados bajo la metodología de ChileValora.

**Oferta y demanda formativa:** Se analiza la oferta y demanda formativa asociada a la Astroinformática en Chile.

*Oferta formativa:* Se analizó el número de programas, vacantes ofrecidas y el perfil formativo tanto en pregrado como en posgrado de cada una de las carreras potencialmente asociadas a la Astroinformática, en el país. Las carreras elegidas como carreras asociadas a la Astroinformática son carreras que ofrecen cursos que permiten desarrollar una o varias de las competencias requeridas en los 4 componentes de la cadena de valor de la Astroinformática.

Para el año 2018, se observa que las universidades concentran la mayor matrícula en las carreras relacionadas a la Astroinformática, seguidas por los Institutos Profesionales y, finalmente, por las carreras impartidas por los Centros de Formación Técnica. Se aprecia una brecha de género importante en todas ellas.

Las carreras con mayor oferta en términos de alumnos matriculados son Ingeniería Civil Industrial, Ingeniería en Computación e Informática, Técnico en Electricidad y Electricidad Industrial, Técnico en Computación e Informática e Ingeniería Civil en Computación e Informática. Es decir, las carreras con mayor oferta incluye ingenierías civiles (6 años),

carreras técnicas (3 años) y carreras de 4-5 años (licenciaturas e ingenierías).

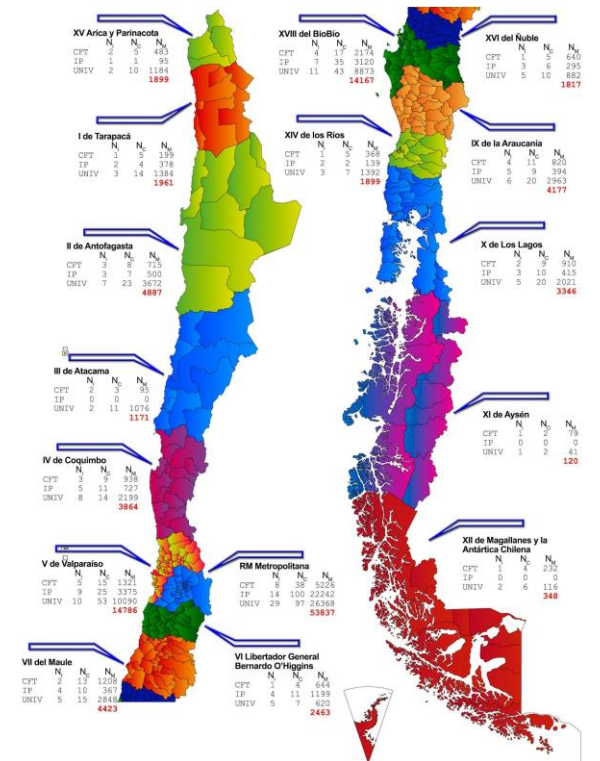


Fig 6: Distribución de la oferta y demanda formativa por región y tipo de institución. Ni es el número de instituciones, Nc el número de carreras impartidas, y Nm es la matrícula total para el año 2018.

A nivel de posgrado nacional, la carrera de posgrado con mayor número de alumnos matriculados es el Doctorado en Matemáticas, seguido por los Magíster en Ingeniería Eléctrica, Magíster en Ingeniería Informática, Magíster en Física. El diplomado en Big Data aparece como uno de los nuevos programas de diplomado con la mayor matrícula, lo que muestra que la oferta está respondiendo a la demanda por capital humano especializado en esa área.

Considerando la oferta formativa a nivel nacional, se observa una gran concentración en la región metropolitana, seguida de las regiones de Valparaíso y BíoBío, tanto en pregrado como

postgrado. A pesar de tener un gran desarrollo en astronomía, la región de Coquimbo no tiene una gran oferta formativa en carreras relacionadas a la Astroinformática. En particular, los programas de postgrado son exclusivamente en física y astronomía.

Matrícula	Número	%
Carreras de postgrado relacionadas al data science	1984	2.7%
Carreras de postgrado data science	284	0.4%
Total carreras de postgrado	71701	97.3%

*Tabla 1: Porcentajes de la matrícula de las carreras de data science y afines para el año 2018. Elaboración propia en base a los datos de Matrícula del Ministerio de Educación.*

Adicionalmente, se considera la oferta online para realizar cursos y especializaciones en el campo del ciencia de datos. Existen numerosos sitios online especializados que entregan conocimientos relacionados a todos los eslabones de la cadena de valor de la Astroinformática. Además de esto, también entregan competencias en varios lenguajes de programación (python, R, Ruby, Java, etc.), tipos de bases de datos (SQL, Oracle, Mongo DB, cassandra, etc.) y técnicas avanzadas de análisis de datos (machine learning) y visualización (Power BI, Tableau, pyplot, etc.). Los sitios más populares incluyen Coursera, edX, Udemy, Datacamp, Udacity y Codeacademy, entre otros. La demanda por este tipo de formación sería relativamente alta, ya que según la encuesta llevada a cabo en este estudio, más del 50% de los encuestados declaró haber cursado estudios online. A nivel nacional existe oferta de cursos online relativo a ciencia de datos, tanto desde la academia con los MOOC (cursos online masivos y abiertos) y educación continúa, como desde el sector privado con cursos y capacitaciones en plataformas/herramientas específicas.

Una iniciativa interesante a destacar es “*La Serena School for Data Science*”, llevada a cabo por la ULS (universidad de La Serena) en conjunto con AURA, en la IV región, recibiendo apoyo de universidades, observatorios y del gobierno. Esta escuela se ha convertido en uno de los eventos de ciencia de datos más importantes del mundo académico y astronómico. La escuela de ciencia de datos hace interactuar a los estudiantes con técnicas y set de datos reales para dotarlos de conocimiento práctico aplicable tanto en la academia como en la industria.

Por último, se realiza una proyección de la demanda formativa en carreras relacionadas a la Astroinformática para la región de Coquimbo, utilizando las proyecciones de población en dos escenarios: 3% y 5% de la población demanda formación en Astroinformática y data science. Se estima una demanda entre 15 mil y 25 mil estudiantes de todos los niveles formativos.

**Oferta y demanda laboral:** Se estudia la demanda laboral en materia de adopción y utilización de los 4 componentes de Astroinformática y data science, en general.

La demanda por capital humano se refiere a los requerimientos de las empresas, academia, observatorios, u otras instituciones, de personal altamente calificado para que realice una serie de tareas que conforman una ocupación o un puesto de trabajo en el campo de ciencia de datos y la Astroinformática. Estas tareas, incluyen la adopción y utilización de las 4 componentes de la cadena de valor de la Astroinformática.

De acuerdo al Ministerio del Trabajo de Estados Unidos, la ocupación laboral relacionada a ciencia de datos involucra todos los trabajos que incluyen sufijos/prefijos de *Data*, *Data Science* y *Data Scientist* como *Analyst*, *Manager*,



*Solutions, Specialist, Big Data, Machine Learning, Mining*, entre otros. Se proyecta un crecimiento entre 2016 al 2026 en el número de empleos de un 28%. Adicionalmente, otras ocupaciones cercanas a ciencia de datos como estadística, desarrollo de software (aplicaciones), y seguridad informática, también presentan un crecimiento en el número de empleos cercano al 30% en el periodo 2016-2026.

Palabras Clave	Chile		Argentina		Brasil		Usa	
	N	%	N	%	N	%	N	%
Data Science	16509	0.47	22657	0.33	100958	0.29	5851954	3.90
Data Analytics	4329	0.12	7754	0.11	31626	0.09	1608414	1.07
Big Data	5097	0.15	7321	0.11	42186	0.12	729402	0.49
Machine Learning	2721	0.08	3408	0.05	21876	0.06	521134	0.35
Data Scientist	442	0.01	494	0.01	2159	0.01	56983	0.04
Ciencia de datos	3	0.00	7	0.00	1	0.00	490	0.00
Usuarios(MM)	3.5	19.11	6.8	15.10	35	16.50	150	45.67

*Tabla 2: Recurrencia de profesiones y habilidades para diferentes palabras claves o "keywords" en búsqueda de personas en LinkedIn Marzo 2019, según país y la fracción respecto al número de usuario. Los valores en naranja indican la fracción de usuarios respecto a la población del país.*

Para analizar las competencias y profesiones relacionadas a las cuatro componentes de la cadena de valor de la Astroinformática se realizó una búsqueda de palabras clave de la red LinkedIn, donde casi el 80% de los chilenos económicamente activos participa. La fracción de usuarios que presenta las palabras clave *data science* o ciencia de datos en Chile y EE.UU es de un 0.47% y 3.9%, respectivamente. Por otro lado, las ofertas laborales en bolsas de trabajo en línea que presentan las palabras clave *data science* o ciencia de datos en Chile y EE.UU. representan el 3.3% y el 7.5% del total, respectivamente.

También se analizaron bolsas de trabajo de universidades chilenas, donde se observó una concentración de ofertas laborales con las palabras clave *data science* o ciencia de datos en las dos universidades líderes en los rankings de

calidad: U. Chile y U. Católica de Chile con 8.6% y 6%, respectivamente.

País	Data Science	Relativas	Total	% DS/Total	%DS/Rel
Chile	1472	12212	44582	3.3	12.1
USA	330588	796544	4457539	7.4	41.5
Australia	9181	26798	144265	6.4	34.3
Holanda	9724	49901	272351	3.6	19.5
Japon	2769	10135	55543	5.0	27.3
Francia	14562	28571	567777	2.6	51.0
Canada	19299	52374	335807	5.7	36.8
Italia	2817	22197	122083	2.3	12.7
China	10549	51903	3809422	0.3	20.3
Alemania	17510	81808	684348	2.6	21.4
UK	40498	136682	536021	7.6	29.6

*Tabla 3: Ofertas de trabajo en el sitio Indeed.com en data science y para carreras tradicionales asociadas a data science, para diferentes países del Benchmarking. Las búsquedas de palabras claves es en inglés a excepción de Chile que se realiza en español e inglés. Los países en verde son de habla inglesa y donde el sitio tiene mayor penetración.*

Para estimar la expansión de la demanda, se analiza la demanda actual y luego se asumen dos escenarios: optimista y pesimista, con crecimiento anual de 25% y 10%, respectivamente. En un escenario optimista se proyectan 4100 y 8000 puestos de trabajo en 3 y 6 años, respectivamente. En un escenario pesimista se proyectan 1870 y 2486 puestos de trabajo en 3 y 6 años, respectivamente.

Para estimar la proyección de la demanda formativa se utilizaron las entrevistas a los actores clave y la encuesta realizada para este informe. En base a esta información se determinó que las habilidades y competencias más importantes y con mayor proyección a futuro son: 1) capacidad de trabajo en equipo y comunicación, 2) competencias en visualización y exploración de datos y, 3) competencias en inteligencia artificial y machine learning. Al mismo tiempo, las características de la oferta que tienen mayor retorno corresponde a estudios de doctorado, formación de pregrado

en matemáticas, estadística, administración, industrial y economía; y competencias en manejo en la nube, reportería y *business intelligence*, y *machine learning*.

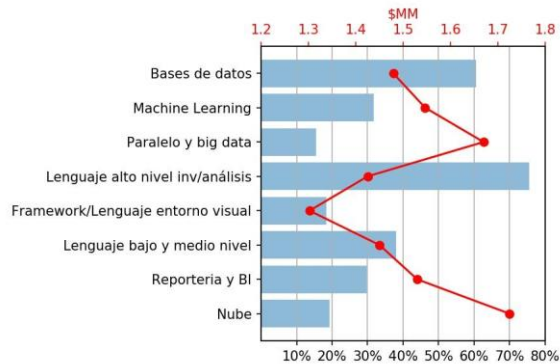


Figura 7: Porcentaje de encuestados que dice dominar alguna herramienta/conocimiento computacional (barras) y salarios mensuales promedios de los encuestados que dice dominar ese campo (línea roja).

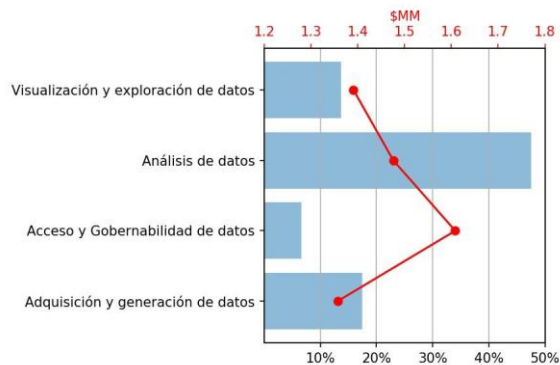


Fig 8: Porcentaje de encuestados que responde realizar tareas relacionadas a las 4 áreas de la cadena de valor de la Astroinformática (barras) y el promedio del salario líquido mensual(rojo).

**Benchmarking:** Se realiza un estudio de benchmarking con el objetivo de generar un diagnóstico respecto de las brechas más importantes que tiene el país en comparación a los países líderes en data science y Astroinformática a nivel mundial.

Se identificaron los países líderes en el desarrollo de la Astroinformática: Estados Unidos,

Australia, Holanda, Japón, Francia, Canadá, Italia, China, Alemania, y Reino Unido.

Se revisaron ciertos indicadores agregados en temas de investigación, innovación y desarrollo, tales como gasto en Investigación y Desarrollo (I+D) como porcentaje de Producto Interno Bruto (PIB), porcentaje de doctorados e investigadores respecto de la población en edad de trabajar, y carreras de educación superior en data science.

Los indicadores son desalentadores. Chile se encuentra lejos de los países líderes, tanto en financiamiento como en formación de capital humano. Los países líderes gastan en promedio 2.2% del PIB en I+D, 6 veces lo que gasta Chile. Los países líderes tienen un alto número de capital humano altamente calificado. En promedio, tienen 10.4 PhDs por cada 1000 habitantes entre 25 y 65 años, alrededor de 100 veces más que Chile [3].

En relación al número de investigadores con jornada completa respecto del empleo total, en promedio, los países líderes cuentan con 9 veces más investigadores que Chile por cada 1000 trabajadores.

De las entrevistas a los actores clave internacionales se destaca el avance que ha tenido Alemania respecto a la inserción de científicos en la industria. Esto, en el ámbito de ciencia de datos, donde es natural y frecuente que graduados en Astronomía/Astrofísica sean contratados por empresas de primer nivel situadas en las mismas ciudades, ya que saben que los buenos científicos poseen habilidades y competencias que son fácilmente transferibles a la industria.

Un punto destacable a notar es que dos actores internacionales mencionan que el aprendizaje de tecnologías debe centrarse en resolver

problemas concretos para la industria o área de especialización y que estudiar sólo en base a un marco teórico es obsoleto y poco efectivo.

En general, varios entrevistados coinciden en las ventajas que tiene Chile gracias a los grandes telescopios que posee y, en especial, a los que entrarán en operación en los próximos años. Telescopios que incrementarán enormemente el volumen de datos generados hasta ahora. Es en este punto en el que el gobierno debe tomar un rol protagónico y financiar iniciativas e infraestructura relacionadas a la Astroinformática, que no sólo beneficiará a los astrónomos, sino a toda una generación de personas con habilidades en torno al *big data* y la ciencia de datos.

Coinciden también los entrevistados en el gran crecimiento que se proyecta para la ciencia de datos. Crecimiento que se observa y manifiesta en países como USA, donde un gran número de estudiantes de pregrado en física y astronomía optan por sacar grados en *data science*.

Las entrevistas, confirman una necesidad y urgencia de generar nuevos programas de formación en ciencia de datos para nuestro país, dándole a la astronomía un rol protagónico y fundamental en la articulación de dichos programas con la industria.

Desde el punto de vista educativo, planteamos el desafío de unificar y consolidar competencias por área y perfiles. Esto último es crucial, pues las competencias son definidas en el sistema educativo actual desde los centros de formación y no desde una política nacional que defina y estructure los requerimientos de los diferentes espacios de desarrollo profesional.

**Planes formativos:** Se diseñan planes formativos que contengan competencias transversales y de especialización requeridas en el mercado laboral, de acuerdo a lo señalado por los actores clave y el análisis realizado en el

presente estudio para establecer una trayectoria educativa. Así, se escogen los 2 perfiles ocupacionales que tendrían mayor demanda, el perfil profesional de Científico de datos mención Análisis de datos y el perfil de Técnico en Administración de Datos.

La carrera de Científico de datos mención Análisis de datos se estructura desde un perfil formativo profesional y con articulación laboral de manera transversal durante la formación. Se plantea una carrera de 4 años más una práctica profesional, sin embargo, se podría ajustar el plan formativo en algún tipo de especialización, minor, o diplomado de 1 año, enfocado en alumnos que homologan gran parte de las competencias o módulos curriculares en matemáticas, estadísticas y bases de programación.

La carrera de Técnico en Administración de Datos busca generar competencias de trabajo en equipos multidisciplinarios bajo un lenguaje común, y que pueda desempeñar labores de preparación o curaduría de datos entre otros. Se propone una carrera de hasta 3 años, pero se podría ajustar el plan formativo en algún tipo de especialización, minor, o diplomado de 1 año o menos, enfocado en alumnos que homologan gran parte de las competencias o módulos curriculares en fundamentos básicos de matemáticas, e informática como es el caso de alumnos que han cursado carreras técnicas de informática o computación. Esta carrera busca desarrollar con mayor énfasis la articulación laboral.

Se destacan también las competencias clave que debe tener todo egresado de una carrera de *data science*: pensamiento computacional y estadístico, base matemática, construcción y evaluación de modelos, base de algoritmos y software, curación de datos, y transferencia de conocimiento (comunicación y responsabilidad). Todas estas competencias se conectan y

potencian en la cadena de valor de la Astroinformática.

**Estudio de Potencial Regional:** Se analiza el potencial de la IV región para formar capital humano en ciencia de datos y Astroinformática. Se identifican capacidades, brechas y oportunidades en torno a esta área.

La cuarta región posee capacidades claras para desarrollar la ciencia de datos. La región se encuentra en una posición ventajosa respecto a otras regiones del país al albergar algunos de los telescopios más poderosos del mundo. La capacidad astronómica instalada la convierte en un laboratorio natural para el desarrollo de la ciencia de datos. Además, en la región se observa que muchas de las ramas que posee la ciencia de datos a nivel formativo, perteneciente a la cadena de valor de la Astroinformática, ya se encuentran disponibles en las carreras impartidas por los establecimientos educacionales existentes, por lo que la implementación de esta disciplina no significa comenzar desde cero, sino que la creación parcial de nuevos ramos entorno a una reestructuración de los ya existentes, localmente.

La región sí posee brechas importantes que deben ser salvadas para ser capaz de desarrollar la ciencia de datos. Estas brechas están asociadas, principalmente al ecosistema productivo local llamado a absorber el capital humano generado en esta nueva área del saber. La región posee un bajísimo nivel de digitalización en la industria, lo que cierra las puertas a cualquier esfuerzo por insertar capital humano formado en ciencia de datos, localmente. Ciertamente, el desarrollo de planes formativos en data science debe ir de la mano con una transformación digital de la industria

local para aumentar las opciones laborales del capital humano formado.

Además, en la región se instalará el LSST, llamado a revolucionar la generación de datos astronómicos a nivel mundial. Este telescopio tendrá su centro de datos a un costado de la ULS, por lo que en principio no deberían existir barreras geográficas para entablar una relación de colaboración entre el centro de datos del LSST y la ULS. Por último, la celebración de la escuela de ciencia de datos de la ULS, cada año mejor posicionada, aparece como un potencial nodo de desarrollo de la disciplina que debe ser aprovechado por el gobierno regional para conectar la ciencia de datos proveniente de la astronomía con la industria local.

Se observa que articulando y potenciando el ecosistema formado por Academia, Industria y Observatorios, la región puede aprovechar al máximo las oportunidades de desarrollo que se le presentan. En efecto, cambio en las políticas de las instituciones académicas en relación a la evaluación de sus académicos, mayor coordinación y diálogo entre academia, observatorios e industria, y mayor incentivo económico y políticas facilitadoras de la interacción academia-industria son claves para crear un ecosistema capaz de crear capital humano en torno a la ciencia de datos y absorberlo de forma local, sustentablemente.

**Plan de Acción y Difusión:** Se presenta un plan de acción y difusión para, por un lado, implementar programas formativos específicos en el área de la Astroinformática y, por otro lado, desarrollar el ecosistema productivo en torno a la ciencia de datos a nivel regional. Se plantea una metodología de Espiral de Nesta para asegurar un alto grado de escalabilidad de las acciones propuestas.

Las acciones están apuntadas a generar un diálogo constructivo entre los actores relevantes, esto es: Academia, Observatorios e Industria, y los organismos impulsores de políticas públicas, Ministerio de Educación, del Trabajo y Corfo, por ejemplo. En este cuadro, el gobierno debe tomar el rol de articulador de las diferentes medidas que se tomen para asegurar el desarrollo adecuado de la ciencia de datos en la región de Coquimbo.

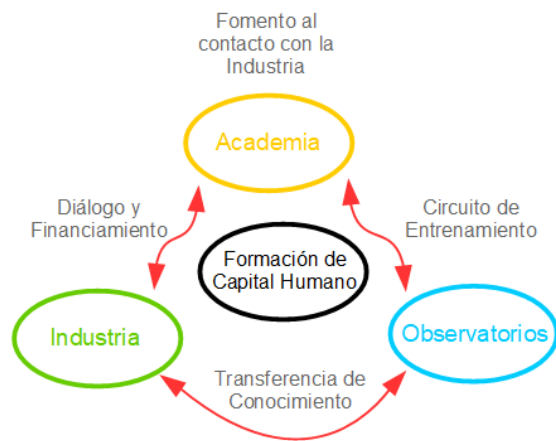


Fig 9: Esquema de necesidades a satisfacer para asegurar un desarrollo de la ciencia de datos.

Las necesidades identificadas se resumen en i) la creación de instancias para un diálogo entre la academia y la industria, ii) la creación de instrumentos de financiamiento para la aplicación de soluciones desde la ciencia de datos a la industria, iii) la modificación de las métricas de desempeño académicos para fomentar la interacción con la industria, y iv) el fomento de la transferencia de conocimiento bidireccional entre industria y observatorios. Los puntos enumerados, anteriormente, deben interpretarse como medidas a tomar para el desarrollo de la ciencia de datos en la región. Todas ellas en conjunto levantarán un escenario propicio para generar capital humano especializado en ciencia de datos de nivel técnico y profesional, siguiendo planes

formativos adecuados a la realidad de la región y, ayudando al desarrollo económico local.

Sumado al plan de acción asociado a las necesidades enumeradas anteriormente, se sugiere además la creación de un centro interdisciplinario de ciencia de datos. La creación de dicho centro aparece como una oportunidad inigualable teniendo en cuenta que la ULS, a través de la organización de la ya tradicional escuela de ciencia de datos ha demostrado que tiene la capacidad técnica y humana para liderar un centro de esas características. Si a lo anterior se le suma la capacidad astronómica instalada existente y (como el LSST-Observatorio Vera Rubin) en la región, la creación de este potencial centro interdisciplinario para la ciencia de datos aplicada a todo ámbito de la sociedad se ve más factible aún.

Se plantea un plan de difusión que busca dar a conocer los resultados de este estudio y evidenciar la importancia que posee el desarrollo de la Astroinformática como impulsor de la ciencia de datos en la región. Se sugiere difundir la información en establecimientos educacionales superiores y escolares, empresas, el sector público y observatorios. Todo esto con el apoyo de Corfo como actor difusor principal.

## Referencias

1. "Astroinformatics Business Opportunities (actual offer, future demand, and opportunities for Chile)", E&Y, 2018.
2. "Astroinformatics Program Progress Report", Arancibia D. et al., 2018 Chilean Economic Development Agency, Digital Transformation Agency.
3. OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris.
4. "Curriculum Guidelines for Undergraduate Programs in Data Science", De Veaux et al, *Annu. Rev. Stat. Appl.* 2017. 4:15–30