



Universidad
de Navarra

DATAI
INSTITUTO DE CIENCIA DE LOS
DATOS E INTELIGENCIA ARTIFICIAL

PLAN ESTRATÉGICO 2020/2025



1. DESCRIPCIÓN DEL CENTRO

El 11 de diciembre de 2019 se puso en marcha el Instituto de Ciencia de los Datos e Inteligencia Artificial (en adelante DATAI). Es un centro de investigación transversal, de innovación y formación de la Universidad de Navarra que busca un trabajo colaborativo de diferentes grupos y personas de la Universidad con un peso importante en Ciencia de los Datos e Inteligencia Artificial.

El Instituto, con **dependencia orgánica de TECNUN** y sede en Pamplona, pretende actuar como órgano transversal que aglutina proyectos de análisis estadístico y procesamiento de datos que ya se están realizando en distintas facultades, grupos y servicios de la Universidad. Nace con una clara vocación de colaboración e integración de la actividad investigadora en Big Data e Inteligencia Artificial (IA) y de formación.

La **Junta directiva** del Instituto está formada por su Director Don Jesús López Fidalgo, tres subdirectores: Dña. Elisabeth Viles (**Tecnun-Escuela de Ingenieros**), Doña. Stella Maris Salvatierra (**facultad de Económicas**) y D. Sergio Ardanza-Trevijano Moras (**Facultad de Ciencias**); el Gerente D. Fernando de la Puente García-Ganges y el secretario D. José María González Gullón.

DATAI tiene **miembros** en todos los campus de la universidad: Pamplona, Madrid, San Sebastián y Barcelona. Además de las distintas Facultades y Escuelas de la Universidad de Navarra, el Instituto trabaja con personal de los centros de investigación y asistencia como son el CEIT, el CIMA y la CUN, entre otros.

Muchos de los investigadores externos de empresas como BBVA y HBC Banca Internacional también participan en las labores de formación y docencia del Instituto.

	2020	2021
Miembros Adscritos	2 investigadores, 2 doctorandos, 2 apoyos a la investigación (1 a TP), 2 PAS, 1 administrativa, 1 Gerente (TP)	2 investigadores, 2 posdocs, 2 doctorandos, 2 apoyos a la investigación, 1 apoyo docente, 2 PAS, 1 administrativa, 1 Gerente (TP)
Miembros Asociados	7 Económicas, 7 Tecnun, 2 Educación y Psicología, 3 Ciencias, 8 IESE, 4 CIMA, 1 Derecho, 1 Comunicación, 2 Medicina, 3 CUN, 4 CEIT	7 Económicas, 7 Tecnun, 2 Educación y Psicología, 3 Ciencias, 8 IESE, 4 CIMA, 1 Derecho, 1 Comunicación, 2 Medicina, 3 CUN, 4 CEIT
Miembros Invitados	12 de BBVA y otros 6	12 de BBVA y otros 6

Tabla 1 – Evolución de los miembros de DATAI en 2020 Y 2021

DATAI cuenta actualmente con 72 miembros distribuidos en miembros adscritos (12), asociados (42) e invitados (18); organizados en 8 áreas de investigación que conforman DATAI. A su vez, cada área dispone de sus propias líneas de investigación.



Tiene como objetivo tener una masa crítica suficiente de miembros a incrementar durante la duración del plan hasta llegar a 20 miembros adscritos, 50 asociados y 30 invitados

	2020	2021	2022	2023	2024	2025
Adscritos	10	12	14	16	18	20
Asociados	42	42	44	46	48	50
Invitados	18	18	21	24	27	30
TOTAL	70	72	79	86	93	100

Tabla 1 – Proyección de miembros de DATAI 2020-2025.

DATAI cuenta con un **Consejo Asesor** formado por representantes de todas las facultades y escuelas de la universidad: *Arquitectura, Ciencias, Comunicación, Derecho, Económicas, Educación y Psicología, Enfermería, Farmacia y Nutrición, Filosofía y Letras, Medicina, Teología*. También tiene representantes de los institutos *CEIT, CIMA y CUN*, así como de instituciones externas como *BBVA*.

DATAI cuenta con un **Comité Científico** formado por 5 destacados miembros en el ámbito de la Ciencia de los Datos:

- Prof. Amparo Alonso Vetanzos, miembro del consejo superior de España de inteligencia artificial y ex-presidenta de la AEPIA (Spanish Association for Artificial Intelligence).
- Prof. Enrique del Castillo, distinguished professor de la Universidad de Pensilvania.
- Dr. Nuria Oliver, Directora de Ciencia de Datos de Vodafone,
- Prof. John Stufken, Bank of America Excellence Professor de la Universidad North Carolina Greensboro.
- Prof. Trevor Hastie, miembro honorífico, Departamento de Estadística y Departamento de Ciencia de Datos Biomédicos de la Universidad de Stanford.

Entre sus objetivos el más importante es la **investigación**, así como la **innovación y transferencia** de conocimiento al ámbito industrial, empresarial y social. También existe una fuerte implicación en la **formación** de investigadores, estudiantes y profesionales. Por esta razón, el Instituto dirige un Máster oficial en Big Data Science destinado principalmente a jóvenes profesionales, así como estudiantes de doctorado, en diferentes áreas con necesidades importantes de análisis de datos. El siguiente gráfico representa bien estas tres patas en las que se asienta el instituto.

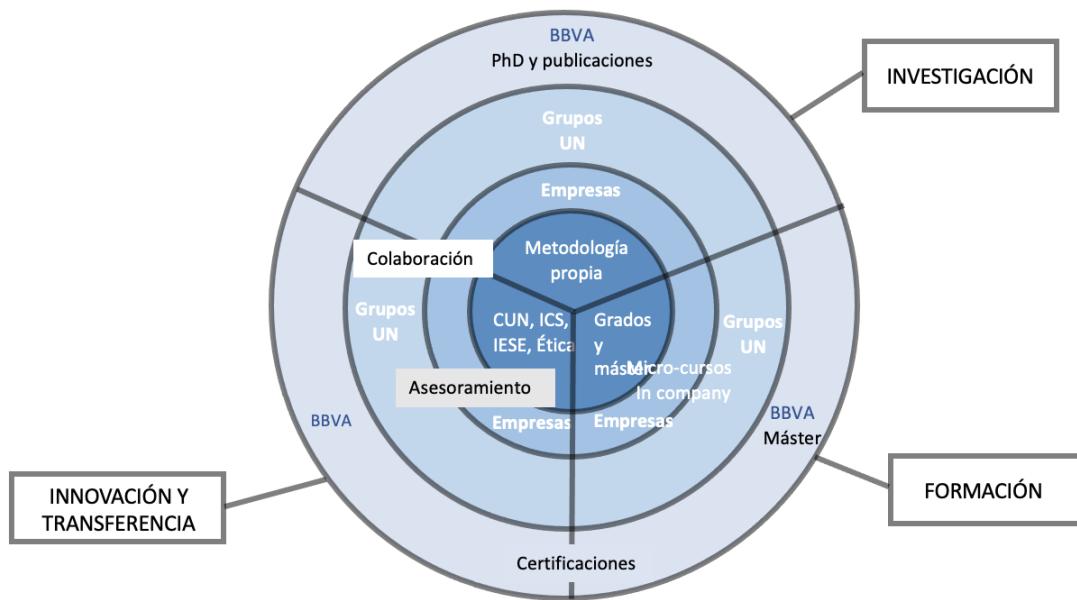


Figura 1 – Esquema de la estructura de DATAI.

2. ANÁLISIS DAFO

A continuación se presenta el análisis de DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades) de DATAI.

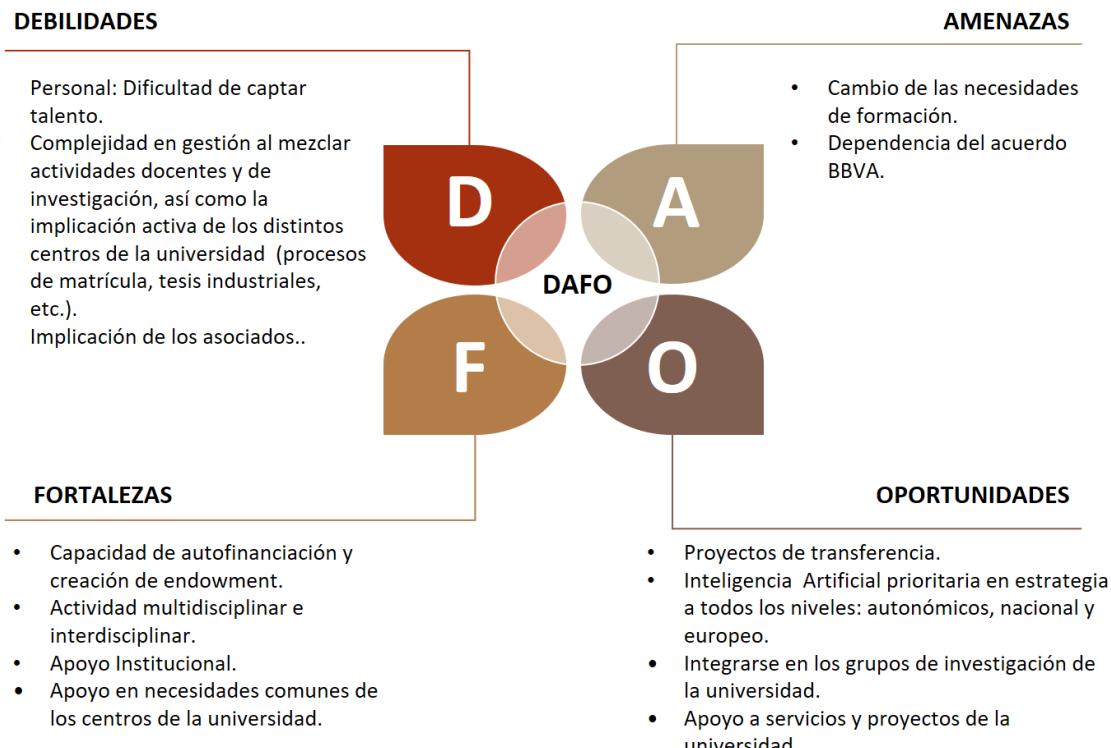


Figura 2 – Análisis DAFO de DATAI.



3. INVESTIGACIÓN

3.1 Líneas de investigación

3.1.1 Líneas de investigación de Miembros Adscritos

3.1.1.1 - Diseño Óptimo de experimentos

Modelización en diversas áreas, tales como crecimiento de tumores, ensayos clínicos en las tres primeras fases, retención de la radiación en el cuerpo humano o farmacocinética. Esta es una de las líneas del instituto, que lleva en marcha más de 30 años y en la que se han obtenido resultados relevantes. Siendo línea estratégica, necesita en estos momentos acabar de formar un equipo sólido de trabajo en DATAI. El responsable de esta línea tiene formado un grupo consolidado con las universidades de Salamanca y Castilla La-Mancha donde colaboran también investigadores de la Universidad Pública de Navarra, Almería o La Laguna. Más recientemente, dentro de esta línea se está haciendo un esfuerzo por aplicar las técnicas de diseño óptimo de experimento a los problemas esenciales que están surgiendo en Ciencia de Datos, Big Data e Inteligencia artificial. En concreto:

1. **Active learning** (Submuestreo inteligente) en big data. Uno de los problemas importantes del tratamiento de grandes cantidades de datos es el coste computacional de los análisis y ajustes de modelos, que en algunos aspectos crece exponencialmente convirtiéndolos en problemas NP. Una técnica habitual, basada en el teorema central del límite o en las leyes de los grandes números, consiste en tomar una submuestra aleatoria y hacer los cálculos con una cantidad razonable de datos. Esta línea de investigación busca mejorar el método de muestreo, de modo que se capten los datos más informativos dentro de la muestra grande. Esto adquiere especial relevancia cuando la distribución de las variables explicativas es altamente asimétrica o con curtosis extrema. Este último es un tema apenas estudiado en la literatura de modo formal.
2. **Algoritmos verdes**. El incremento de la capacidad computacional, junto con la posibilidad de recoger grandes cantidades de datos, nos ha llevado a desarrollar e implementar algoritmos que requieren mucho tiempo de computación, con el consiguiente gasto de energía. En parte, los llamados algoritmos verdes buscan optimizar ese coste energético mediante algoritmos más eficientes que requieran menos recursos de ejecución. Las técnicas de diseño óptimo pueden aplicarse a la elección óptima de los parámetros iniciales de un algoritmo, de modo que su velocidad de convergencia sea mayor y así reducir el tiempo de computación.



3. **Experimentos in silico.** También llamados computer experiments, son cada vez más frecuentes a través de los llamados modelos ***digital twins***. Son simuladores construidos con complejos sistemas de ecuaciones en derivadas parciales, que habitualmente requieren un gran coste computacional. La teoría de ***computer experiments*** busca simplificar esos modelos de modo que con una mínima pérdida de precisión, puedan conseguirse resultados en un tiempo mucho menor, a la vez que se añade una mayor ***explicabilidad*** del modelo.
4. **Selección de modelos.** Típicamente para un mismo fenómeno existen muchos modelos rivales posibles, a lo que se añade una gran cantidad de variables, entre las que hay que seleccionar unas pocas que sean suficientes para explicar la respuesta y hacer predicciones. En este aspecto la discriminación de models juega un papel esencial en el que el equipo lleva trabajando durante años con algunas aportaciones muy relevantes.
5. **Medicina personalizada.** Se han desarrollado diseños experimentales personalizados, de modo que dependiendo de las circunstancias del sujeto se realiza un experimento u otro, en lugar de la aleatorización plena para todos los sujetos. Esto permite obtener mejores ajustes del modelo con menos experimentos. Se busca usar esta técnica para proponer un tratamiento ajustado a cada paciente, teniendo en cuenta sus circunstancias particulares.

3.1.1.2 - Redes complejas y análisis de redes sociales

Los fundamentos de la ciencia de redes (Network Science) se aplican cada vez más para estudiar una variedad de fenómenos del mundo real, muchos de los cuales constituyen sistemas complejos por su naturaleza intrínseca. Investigadores de diversos campos consideran la formulación y estudio de sistemas complejos como un tema crucial. Estos sistemas complejos pueden ser modelados mediante redes (denominadas también grafos), que están representadas por nodos y enlaces (componentes de la red), entre los cuales se establecen interacciones que definen patrones y estructuras útiles complejas de identificar. De manera general, los nodos representan las entidades de la red y los enlaces las conexiones que se establecen entre ellas. Entre los ejemplos de redes complejas podemos mencionar: las estructuras cerebrales, las relaciones sociales, las comunicaciones en la telefonía móvil, las interacciones biológicas moleculares, las transacciones bancarias, las redes de transporte, etc.

En muchos de estos dominios de aplicación, los datos de la red resultan complejos, tanto en el gran volumen de información a procesar como en la variedad de estos, con lo cual, constituye un desafío su análisis. Por tanto, dados los niveles de información, estudiar las



estructuras e influencias de los diferentes componentes de una red (nodos y enlaces), induce el estudio de nuevos modelos y técnicas eficientes.

En este contexto, el Análisis de Redes Complejas (Complex Network Analysis) constituye el área de investigación relacionada con el diseño, desarrollo y aplicación de modelos matemáticos-computacionales para la detección y el análisis de las estructuras y patrones subyacentes de interés presentes en estas redes.

Entre las tareas objetivo de la línea de investigación cabe mencionar:

1. La detección de grupos de nodos (entidades) con un comportamiento similar (ej. grupos de personas con intereses comunes en una red social, conjunto de proteínas similares en una red biológica).
2. La identificación de nodos influyentes en la red (ej. usuarios en una red social con alto número de seguidores, cuentas con alto número de operaciones en una red de transacciones bancarias).
3. El análisis de la dispersión de “información” (ej. identificación de las entidades de la red que maximizan la dispersión en un escenario epidemiológico).
4. La predicción de futuras relaciones entre los nodos (ej. identificar futuras relaciones de “conveniencia” en una red de empresarios de negocios).
5. La identificación de nodos anómalos (ej. detección de sensores con mediciones atípicas respecto a la norma en unas redes de sensores inalámbricos), etc.

La amplia gama de escenarios en los cuales resulta pertinente la modelación mediante redes, hace el análisis de redes complejas una línea de investigación transversal, resultando de gran aplicabilidad en el contexto actual.

3.1.1.3 - Machine Learning en Bioinformática y Neurociencia Computacional

La medicina traslacional trata de combinar diversas disciplinas científicas para mejorar la prevención, el diagnóstico y el tratamiento de las afecciones clínicas. Es, por definición, un campo altamente interdisciplinar cuyo objetivo principal es mejorar el sistema sanitario mundial. Aunque la mayoría de las disciplinas biomédicas han evolucionado para incorporar dispositivos digitales, los algoritmos que recogen, procesan y analizan los datos son abrumadoramente desconocidos y/o inaccesibles para la mayoría de los profesionales.

Lo que hoy se conoce como la revolución de la medicina digital no es más que la evolución natural de las tecnologías de la información que ayudan a las evaluaciones clínicas y médicas. El aprendizaje automático ya está marcando la diferencia en este contexto.

Un ejemplo destacado es el uso de la neurociencia básica para personalizar las evaluaciones médicas de los pacientes con enfermedades cerebrales. Más concretamente,



un programa que lleve la investigación neurocientífica básica hasta los protocolos hospitalarios, pasando por validaciones clínicas podría lograr el desarrollo de tratamientos personalizados.

Línea enfocada en el aprendizaje automático en Bioinformática y Neurociencia Computacional. Concretamente se utilizará el tratamiento de gran número de variables que generan los análisis genéticos, habitualmente con pocos sujetos. En particular se buscará detectar interacciones fiables de los genes utilizando clasificadores basados en redes jerárquicas bayesianas. Expresión micro RNA en pacientes con diversas patologías. Identificación de biomarcadores en cáncer colorectal. Clasificación automática de neuronas.

3.1.1.4 - Fair Learning

Fairness (equidad en castellano) en el contexto de la IA se refiere a los diversos intentos de corregir el sesgo algorítmico. Aunque las definiciones de imparcialidad son siempre controvertidas, los resultados de un algoritmo pueden considerarse justos si son independientes de una variable determinada, especialmente las consideradas sensibles, como los rasgos de los individuos que no deberían correlacionarse con el resultado (es decir, el género, la etnia, la orientación sexual, la discapacidad, etc.). En el aprendizaje automático, el problema del sesgo algorítmico es bien conocido y estudiado.

Los debates sobre la equidad en el aprendizaje automático tienden a centrarse en el impacto de los diferentes modelos en los grupos socialmente sensibles sin embargo, la imparcialidad es un tema amplio, y es relevante para casi cualquier contexto. ¿A qué usuarios se les muestran determinados anuncios, se les ofrecen determinados precios, obtienen recompensas, obtienen menores tiempos de espera en las llamadas y se les identifica en función de distintos modelos de propensión?.

En los últimos años se han propuesto varios criterios de equidad, pero predominan dos enfoques: la **equidad demográfica** (equivalente a la eliminación del impacto dispar) y la **igualdad de oportunidades**.

La equidad demográfica garantiza que cualquier decisión que tome un modelo de IA no esté relacionada con un atributo sensible (por ejemplo, la raza, el sexo o la edad). En otras palabras, ser hombre o mujer no debería determinar si una persona cometerá un crimen.

La igualdad de oportunidades es un poco más sutil: requiere que los individuos que reúnen los requisitos para un buen resultado obtengan ese resultado con la misma probabilidad, independientemente de que sean miembros del grupo sensible. Por ejemplo, el porcentaje



de individuos que reúnen los requisitos para obtener un préstamo y que acaban recibiéndolo no debería diferir entre grupos raciales.

La investigación sobre la equidad en el aprendizaje automático es un tema relativamente reciente pero que está cobrando gran relevancia debido a la expansión de la IA y la utilización de los datos. Es un campo transversal que tiene implicaciones éticas y legales y sobre el que la comisión europea ya ha puesto el foco en su estrategia acerca de IA destacando como uno de los siete elementos principales la equidad a la hora de desarrollar modelos de IA. España, dentro de su estrategia de Inteligencia Artificial también remarca la importancia del desarrollo sostenible de la IA.

En este contexto DATAI planea desarrollar una línea de investigación de *fairness* centrada en:

- El estudio de la equidad algorítmica, medición y mitigación de sesgos así como el desarrollo de nuevas herramientas y modelos matemáticos para mejorar la equidad.
- La aplicación del *fairness* para el desarrollo de una IA sostenible y ética para la mejora de la sociedad.

3.1.1.5 - Procesamiento del Lenguaje Natural

La Procesamiento del Lenguaje Natural (PLN) es un subcampo de la IA que se ocupa principalmente del procesamiento y la comprensión del lenguaje humano por parte de las máquinas. Al permitir que las máquinas entiendan el lenguaje humano, el PNL mejora la precisión y la eficacia de los procesos. Esto se consigue automatizando diversas tareas repetitivas. Algunos ejemplos de aplicaciones de procesamiento del lenguaje natural son la clasificación de entradas, la traducción automática, la comprobación de la ortografía y el resumen.

Es un área de estudio que comenzó por los años 50s y que está recobrando relevancia debido a la gran cantidad de texto que se genera hoy en día y la capacidad para poder procesarlo. De forma genérica se puede dividir análisis de texto en dos grandes partes:

- **Análisis sintáctico:** se utiliza para establecer el significado observando la gramática de una frase. Es el proceso de estructurar el texto utilizando las convenciones gramaticales del lenguaje. Esencialmente, consiste en el análisis de oraciones dividiéndolas en grupos de palabras y frases que crean una oración correcta. Esto no tiene en cuenta el hecho de que las frases pueden carecer de sentido, que es el punto en el que el análisis semántico viene a echar una mano.
- **Análisis semántico:** Nuestra comprensión del lenguaje se basa en los años que llevamos escuchando y conociendo el contexto y el significado. El lenguaje humano es dinámico. Con la invención de los algoritmos de aprendizaje automático, los



ordenadores son capaces de entender el significado y la lógica que hay detrás de nuestras expresiones al menos hasta cierto punto.

Con el desarrollo del PNL, hoy en día, se es capaz de realizar **análisis de sentimiento** para el lenguaje humano. Esto le permite detectar emociones en el texto, lo que constituye una de las aplicaciones de PNL. De esta forma un algoritmo puede leer un texto y etiquetar la emoción subyacente (si es positiva, negativa o neutra). Esto, junto con la proliferación de las redes sociales permite realizar investigaciones que hace pocos años hubieran sido imposibles. Por ejemplo, uno puede medir de forma objetiva cuál es la opinión que la población tiene respecto a un evento concreto.

El PNL es un campo que se ha desarrollado principalmente para el análisis de texto en inglés y solo es en estos últimos años que se ha comenzado a extender al castellano. Su aplicación por tanto a nuestra lengua es una de las líneas más emergentes a desarrollar en el futuro próximo en la investigación relacionada con IA. Esto ha quedado reforzado por la estrategia del gobierno de España que está en vías de publicar un Proyecto Estratégico para la Recuperación y Transformación Económico (PERTE) asociado a la lengua española y la IA. Línea de investigación cuyos objetivos son:

- El análisis del sentimiento a través del PNL dado su alto potencial de desarrollo e impacto que puede tener en el futuro en la sociedad especialmente en la de habla hispana.
- Desarrollar colaboraciones en proyectos que puedan beneficiarse del PNL con los diferentes centros de la universidad para que hagan de DATAI una entidad de excelencia que promueva investigación transversal a toda la universidad.

3.1.2 Líneas de investigación lideradas por miembros asociados e invitados:

3.1.2.1 - Advanced Analytics – BBVA

BBVA cree que el conocimiento derivado de los datos financieros puede transformar la industria bancaria y su papel en el mundo. Por ello, el área de Advanced Analytics investiga en la aplicación de la IA al sector bancario. Entre las líneas de investigación destacan:

- Técnicas de optimización y aprendizaje automático para portafolios bancarios.
- Métodos avanzados de *pricing* dinámico..
- Motores de recomendación.

3.1.2.2 - AI and Management – IESE

La nueva Iniciativa sobre Inteligencia Artificial y el Futuro de la Dirección es un proyecto multidisciplinar que analiza el impacto de la IA en la dirección de empresas y que



contribuye a formar a los directivos en el uso ético y socialmente responsable de la IA en sus compañías.

La inteligencia artificial afectará en los próximos años a todas las esferas de la actividad económica, por lo que los directivos deben aprender a adaptarse a la evolución de este sector. También deben ser capaces de transformar a sus empresas y asegurarse de que sus empleados, y ellos mismos, poseen las habilidades necesarias para seguir aportando valor en este nuevo contexto. En consecuencia, el propósito de la nueva iniciativa del IESE sobre IA es responder a estas necesidades de generación de conocimiento y formación directiva. Las áreas de investigación son:

- Las habilidades en IA y el mercado laboral.
- El impacto de la automatización industrial.
- El desarrollo y el uso de la IA en las empresas, especialmente en Europa.
- El futuro del trabajo y las organizaciones.

3.1.2.3 - Biología computacional

CIMA

La biología molecular ha sufrido una revolución debido a la capacidad de estudiar simultáneamente el funcionamiento y la expresión de miles de genes y proteínas en el cuerpo del paciente.

Gracias al uso de la tecnología informática, las bases de datos y el análisis estadístico podemos analizar con precisión y rapidez grandes cantidades de datos que nos permiten comprender la complejidad de los mecanismos que causan las enfermedades.

El Programa de Biología Computacional del CIMA - Universidad de Navarra tiene actualmente dos líneas principales de investigación:

- Análisis de datos transcriptómicos, tanto a gran escala como a resolución de célula única.
- Desarrollo de nuevos formatos de archivo para el almacenamiento y acceso a los datos ómicos.
- Métodos de Aprendizaje Automático para resolver problemas biomédicos y su traslación a la clínica.

TECNUN

El Grupo de Biología Computacional de TECNUN es un equipo multidisciplinar con una amplia experiencia en el desarrollo de algoritmos de optimización, análisis estadísticos, así como métodos de aprendizaje automático y aprendizaje profundo, aplicados a diferentes problemas biológicos. Nos centramos principalmente en la salud humana mediante la integración de datos de alto rendimiento (genómica, transcriptómica, proteómica,



metabolómica, etc.) y bases de datos biológicas (genómica, farmacológica, metabólica, etc.). Las principales líneas de investigación son:

- Oncología de precisión.
- Modelos predictivos de toxicidad de fármacos basados en características estructurales de las moléculas.
- Esquemas de compresión adaptados a diferentes datos ómicos.
- Análisis de secuenciación de ADN.

3.1.2.4 - Ciencias de la computación

Las ciencias de la computación abarcan un amplio abanico, desde sus fundamentos teóricos y algorítmicos hasta los desarrollos de vanguardia en visión por ordenador, sistemas inteligentes, bioinformática y otras áreas apasionantes. El papel de los informáticos puede agruparse en las siguientes categorías: diseñar e implementar software, desarrollar formas eficaces de resolver problemas informáticos e idear nuevas formas de utilizar los ordenadores.

La informática es esencialmente el estudio de cómo "mapear" el lenguaje de las máquinas en un lenguaje más comprensible y utilizable para que las personas puedan resolver problemas complejos. Este campo utiliza la teoría de la computación en el cálculo de la información. Las líneas de investigación son:

- Cambio climático.
- Análisis de imagen.
- Modelado y simulación.
- Análisis topológico de datos.
- Blockchain y estructuras de datos autenticadas.
- Soft Computing y resiliencia.

3.1.2.5 - Ética y derecho

El reto del área de Ética y Derecho, se centra en trabajar para que las nuevas tecnologías, al afectar al conjunto de la sociedad, busquen no sólo una mejora en las previsiones, sino que éstas tengan claridad en los factores que las definen. Que se promueva una legislación, sin detener el avance en el desarrollo de dichas tecnologías, que permita entender las funciones sociales o biológicas que intervienen en ella. Que en definitiva se tienda a obtener el conocimiento de los fenómenos sociales antes de su predicción. Las líneas de investigación son:

- Ética e IA.
- Regulación legal de Big Data, IA y protección de datos.
- Ética de la virtud e IA.



3.1.2.6 - Health Analytics

Health Analytics se refiere al uso de grandes cantidades de datos recopilados para proporcionar a las organizaciones información útil y mejorar la atención al paciente. Estos conocimientos se desarrollan en disciplinas analíticas para impulsar la toma de decisiones basada en datos. A su vez, estas decisiones mejoran la planificación, la gestión, la medición, el aprendizaje y, en última instancia, la calidad de la atención al paciente.

A medida que las organizaciones sanitarias de todo el mundo se esfuerzan en mejorar la atención al paciente y hacer más con menos, la analítica cobra relevancia. El desarrollo de las competencias analíticas puede ayudar a las organizaciones sanitarias a aprovechar el Big Data para mejorar sus resultados y ofrecer un valor añadido a los pacientes. Las líneas de investigación asociadas a esta rama son:

- Análisis de señales biomédicas.
- Neurociencia de sistemas.
- Imagen médica.
- Inferencia causal en biomedicina.
- Epidemiología.
- Investigación clínica.

3.1.2.7 - Industrial Data Analysis & Information Management

La investigación del grupo se centra principalmente en la medición del impacto del Big Data en la industria de alta tecnología. Dos son las principales áreas de especialización de nuestra investigación para el desarrollo de procesos industriales, la analítica de la ciberseguridad y el Big Data. Las principales áreas de investigación son:

- Protocolos de comunicaciones y seguridad.
- Identificación de anomalías y parámetros significativos en procesos.
- Modelos de predicción y mantenimiento preventivo.

3.2 Objetivos de investigación

Para el periodo 2020-2025 y conforme al plan estratégico de investigación de la Universidad, DATAI pretende llegar a ser un Centro de Excelencia en Ciencia de los Datos e IA, generando conocimiento y difundiéndolo a través de la publicación de trabajos de investigación en revistas de impacto internacional, transfiriendo conocimientos al ámbito industrial, empresarial y social, y con una fuerte implicación en la formación de investigadores, estudiantes y profesionales

Los objetivos de investigación son consolidar las líneas de investigación propias del Instituto e integrar y fomentar la colaboración con las líneas de investigación de los miembros asociados. Los objetivos para los próximos tres años se concretan en:



1. Desarrollo y consolidación de las líneas del plan Científico
2. Renovación del contrato programa con BBVA y búsqueda de nuevas alianzas.
3. Obtención de nuevos proyectos de investigación (ver Anexo I):
 - a. Proyectos europeos.
 - b. Proyectos Nacionales.
 - c. Proyectos del Gobierno de Navarra.
 - d. Contratos de investigación.
4. Mantener e incrementar la producción científica, en particular el impacto de las publicaciones. Durante los años 2020 y 2021 la actividad científica de DATAI ha dado lugar a 27 publicaciones (13 Q1, 10 Q2 y 4 Q3). Los números actuales se refieren solamente a los investigadores que son miembros adscritos o forman parte de la junta directiva (en total 5 investigadores). No se incluyen las publicaciones de los miembros asociados puesto que el nombre de DATAI, debido a su corta existencia, no aparece todavía consolidado en las bases de datos de WoS, lo que hace que el número real de artículos publicados bajo su afiliación no sea fácil de detectar. Por otro lado, los miembros asociados e invitados están comenzando a firmar sus artículos con la afiliación del instituto, lo que hace que la mayor parte de sus artículos no estén todavía contabilizados.
5. Incrementar el reconocimiento y prestigio de DATAI a través de la asistencia y organización de actividades y seminarios científicos (ver Anexo I).
6. Implicación de los miembros asociados en la investigación impulsada por DATAI.
7. Organización de premios a los mejores artículos del curso 2021-22 para aumentar la implicación del comité científico y el reconocimiento de DATAI.



4. FORMACIÓN.

4.1 Tesis Doctorales.

El principal objetivo del instituto es alcanzar la excelencia investigadora en el campo del Big Data y la IA. Una parte fundamental de la investigación son las tesis doctorales que permiten formar nuevos investigadores que se integrarán en las líneas estratégicas y continuar expandiendo los equipos de investigación liderados por investigadores senior.

Actualmente DATA tiene en marcha 14 tesis doctorales y se planea continuar con ellas y seguir atrayendo candidatos para iniciar algunas más.

	2020	2021	2022	2023	2024	2025
Tesis industriales (activas)	7	11	12	12	12	12
Tesis no industriales (activas)	3	3	4	5	6	6
TOTAL	10	14	16	17	18	18

Tabla 2 – Tesis doctorales activas anualmente en el periodo 2020 - 2025. Se estima que la finalización de una tesis se produce a los cuatro años de su comienzo.

4.1 Máster en Big Data Science.

En los cursos 2017-18 y 2018-19 se impartió un programa de especialistas en Big Data en la sede de Pamplona (en el Aula Siemens Gamesa del ICS). En 2019-2020 el programa pasó a ser un máster oficial y, desde entonces, se desarrolla íntegramente en la Sede de Posgrado de Madrid (Edificio Alumni).

El Máster Oficial en Big Data Science se dirige a perfiles orientados al análisis de datos, con interés por la estadística y la computación. Es un programa que proporciona nuevas oportunidades de cualificación a quienes buscan especializarse en una disciplina tan demandada. Se imparte en formato Executive que facilita compaginar el desempeño profesional con la asistencia a clase. Está pensado para un público con diferentes perfiles de formación universitaria que cuenten con conocimientos de estadística, econometría, investigación operativa o análisis cuantitativo. No se requieren conocimientos avanzados



de programación para realizarlo, pero sí una predisposición a trabajar en entornos informáticos, utilización de algoritmos y análisis de datos.

El máster pretende acercar la realidad tecnológica a los alumnos, con el fin de que puedan conocer de primera mano qué se está haciendo en las empresas más punteras y prestigiosas a nivel internacional. A lo largo del curso, intervienen profesionales de compañías e instituciones de prestigio, con amplia experiencia en la materia. Provienen de sectores como el biosanitario, financiero, consultoría, energía, telecomunicaciones, retail o investigación. En sus sesiones presentan casos de uso reales, así como ejemplos prácticos de utilización de herramientas y técnicas de Big Data en proyectos reales.

El máster consta de 60 ECTS, repartidos en cuatro bloques:

1. **Análisis de datos:** Se pretende que el alumno desarrolle sentido crítico estadístico, necesario en el ámbito empresarial. Se utilizan métodos estadísticos tradicionales como modelos de regresión, árboles de decisión o reducción de la dimensionalidad. También se imparten metodologías más recientes como Machine Learning y Deep Learning. Entre otros, se estudian Random Forest, K-Means, Procesamiento de Lenguaje Natural o Redes Neuronales.
2. **Programación y Computación:** Partiendo de un nivel básico, se pretende que el alumno adquiera conocimientos medios-avanzados y sea capaz de seguir el resto de materias, así como desarrollar cierta autonomía para la fase de aprendizaje personal que precede al máster. Los lenguajes de programación impartidos son R y Python, por ser los más populares y demandados en el ámbito profesional. En este módulo se incluye la fase de extracción de datos, donde el alumno adquiere la capacitación de desenvolverse con bases de datos tradicionales, que por ahora son las más habituales en empresas. También se incluye el manejo de entornos propiamente entendidos como Big Data, como es Hadoop o Spark, y técnicas de recogida de datos en redes sociales como Twitter o Facebook, scraping de web o recogida de imágenes.
3. **Proyectos:** El Máster pretende dar una formación sólida en cuanto a conocimientos técnicos, pero también una visión de negocio, de modo que una vez que concluya el Máster los alumnos puedan actuar como un puente entre los niveles ejecutivo y técnico de un proyecto. De este modo se abordarán, de la mano de profesionales de empresas y multinacionales punteras, casos prácticos y de éxito, buscando aplicar conceptos adquiridos en los dos primeros módulos.
4. **Trabajo Fin de Máster (TFM):** Tiene un peso destacado en el programa. Se busca un enfoque práctico que a la vez dé solución a problemas y proyectos reales propuestos por empresas con las que existen acuerdos de colaboración. Puede ser codirigido tanto por éstas como por académicos de la Universidad de Navarra, y supone una excelente oportunidad para que los alumnos lideren la puesta en marcha de proyectos de impacto en su entorno profesional.

Aunque el programa está coordinado por DATAI, los profesores y académicos de la Universidad provienen de centros y facultades multidisciplinares, entre los que se encuentran la Escuela de Ingenieros, Económicas, Ciencias, Medicina o CIMA. A través de



esta diversidad, se busca enriquecer la experiencia de aprendizaje de los alumnos, proporcionando una docencia que abarca diferentes ámbitos de actuación.

En el curso 2020-2021 contamos con 2 grupos:

- 39 alumnos en grupo abierto
 - Edad media: 26,7
 - 30% internacional
- 25 alumnos en grupo BBVA
 - Edad media: 33
 - 20% Internacional

El Máster se someterá a una revisión continua de sus contenidos por parte de la dirección académica y la junta directiva de DATAI con el objetivo de:

- Mantener y continuar con el crecimiento del número de alumnos que participan en el máster mejorando su prestigio y reconocimiento.
- Realizar una evaluación continua de contenidos y su actualización puesto que es previsible que en unos 5 años el perfil del científico de datos cambiará. Es importante continuar evaluando el máster y adaptando el contenido a los cambios en el mercado.
- Desarrollar nuevas alianzas estratégicas con empresas y actores del sector Big Data e IA en España.

4.3 Cursos

4.3.1 Internos

4.3.1.1 Microcursos

Se ofrece este programa para permitir que profesionales de todas las titulaciones, incluso los que provienen de áreas claramente relacionadas como matemáticas y estadística, se actualicen y ganen algún grado en el uso de las principales metodologías y herramientas disponibles en el mercado para resolver problemas asociados a la gestión y análisis de datos.

Estos cursos se enmarcan en la normativa de enseñanzas propias de la Universidad de Navarra de 4 de septiembre de 2017. Por la cual estos programas se realizan en la modalidad online síncrona bajo la tipología de Curso de Formación Superior. Existen un total de 7 cursos que definen dos itinerarios de formación diferenciados, aunque cada uno de los cursos se puede realizar de forma independiente. Todos ellos constan de 2 créditos ETCS.

- Herramientas de Gestión de Datos (3 cursos).



- o Gestión de datos con Python (2 ETCS).
 - o Gestión de datos con R (2 ETCS).
 - o Hojas de cálculo para el análisis de datos (2 ETCS).
- Técnicas de Ciencia de Datos (4 cursos)
 - o Estadística de la Ciencia de los Datos (2 ETCS).
 - o Machine Learning (2 ETCS).
 - o Visualización (2 ETCS).
 - o Análisis de Redes Sociales (2 ETCS)

Los objetivos de los microcursos son los siguientes:

- Realizar al menos un microcurso al año que de servicio en la Universidad, a demanda a investigadores o empleados.
- Realizar si existe suficiente demanda los microcursos para Pymes a través de [IRIS](#), el Polo de Innovación Digital de Navarra cuyo objetivo es acelerar la transformación digital de las empresas y otros interesados.

4.3.2 Externos

4.3.2.1 Emeritus

Emeritus es una compañía que desarrolla curso online en formato semisíncrono con las universidades más prestigiosas del planeta. Para ello, colabora con más de 50 universidades de primer nivel en Estados Unidos, Europa, América Latina, el Sudeste Asiático, India y China. La empresa se encarga de proporcionar la plataforma tecnológica sobre la que se imparten los cursos, su digitalización y grabación, y las universidades de desarrollar el contenido de estos y poner el profesorado de reconocido prestigio.

La universidad ha llegado a un acuerdo para el desarrollo de un curso en colaboración con Emeritus titulado “Big Data, de los datos a las decisiones” enfocado a enseñar a ejecutivos de compañías de todo el mundo a sacar partido de los datos. DATAI desarrollará el contenido del curso dentro del acuerdo marco con Emeritus para su puesta en marcha.

El acuerdo con Emeritus es de interés para DATAI y a la universidad ya permitirá aprender de la metodología y forma de trabajar seguida por una empresa líder en el diseño de cursos *online* síncronos y por otro lado ayudará en la difusión global de la marca DATAI-UNAV. Los objetivos concretos son:

- Evaluar la experiencia Emeritus y su impacto económico para decidir si DATAI continúa ofreciendo cursos en la plataforma.
- En caso de decidir continuar con Emeritus, realizar un curso adicional en la plataforma alineado con el .



4.3.2.2 MOOCs

El curso en línea masivo y abierto o MOOC (acrónimo en inglés de Massive Open Online Course) es una modalidad de aprendizaje en línea dirigido a un número ilimitado de participantes a través de Internet y según el principio de educación abierta y masiva. En este sentido los MOOC contribuyen a la interacción de los estudiantes con los modelos implantados en internet, a través de las redes sociales. Se caracterizan por ser cursos abiertos, gratuitos (o con una matrícula de un precio muy reducido, por ejemplo 20 euros) y de fácil acceso.

DATAI se ha comprometido a desarrollar junto a BBVA un curso MOOC que se alojará en la plataforma [Coursera](#) (la más popular del mundo) destinado a perfiles de analistas de datos que quieren iniciarse en el mundo de la analítica. El desarrollo de este curso ayudará a consolidar el acuerdo estratégico con BBVA y permitirá recoger experiencia en la creación de este tipo de cursos que cada vez son más comunes.

El objetivo principal de la realización de MOOCs es:

- Reforzar la marca DATAI a través de la realización de un curso MOOC en colaboración con BBVA.

4.4 Cursos In Company y programas directivos

Los programas In Company son programas intensivos y de duración corta con el objetivo de ayudar a las empresas en temas específicos que estén relacionados con el Big Data y la ciencia de datos y están dirigidos a directivos.

Un ejemplo sería un curso de Big Data e Inteligencia Artificial en la empresa para la toma de decisiones: Este curso estaría orientado a que los directivos de grandes empresas entiendan como el Big Data y la IA pueden transformar sus organizaciones y a hacerlas más competitivas. Aprovechando su carácter multidisciplinar, DATAI está en disposición de desarrollar cursos de diferentes temáticas en el ámbito del Big Data.

Además, se pretende utilizar este tipo de cursos como una potencial herramienta de internacionalización. Dada la incipiente demanda de los cursos de Big Data e IA, un curso especializado para directivos podría ofrecerse en Miami enfocado a empresas de Latino América aprovechando el prestigio de la universidad en esa zona geográfica. Todo esto con el objetivo a largo plazo de establecer relaciones internacionales con nuevas instituciones y empresas que podrían traer financiación a través de proyectos.

El objetivo de DATAI con los cursos In Company es:



- Realizar al menos un curso In Company que permita entablar relaciones y difundir la marca con ejecutivos de grandes empresas.



5. TRANSFERENCIA

5.1 Acuerdos estratégicos

El instituto tiene un **acuerdo estratégico** de carácter anual y renovable con **BBVA** que se divide en tres partes bien diferenciadas.

1. **Máster en Big Data Science:** Se ha llegado a un acuerdo para que el Máster exclusivamente para los trabajadores de la Entidad esté siendo impartido de forma específica y *ad hoc* para sus trabajadores. El número mínimo de alumnos es de 15 para la formación del grupo específico.
2. **Certificación profesional en ciencia de datos:** La Certificación Profesional de Científico de Datos es una prueba que permite a los profesionales de la industria de la ciencia de datos:
 - Validar su conocimiento y habilidades en la disciplina de la ciencia de los datos.
 - Diferenciarse de otros profesionales a través de una certificación profesional de prestigio.

La Certificación tiene como misión definir un estándar en el campo de la ciencia de datos a través del cual las organizaciones puedan identificar profesionales con dominio del campo de la analítica avanzada. Además, la Certificación refuerza la credibilidad y visibilidad de la profesión de científico de datos. Consta de dos partes claramente diferenciadas que se abordarán de manera secuencial:

- Parte I - Revisión de casos de uso (duración estimada de 2 meses y medio desde la primera sesión).
- Parte II - Examen (2 semanas después de la finalización del caso de uso).

DATAI ha creado la primera certificación profesional en ciencia de datos de España. Actualmente solo se ofrece a empleados de BBVA en el marco del acuerdo con el banco, pero se espera poder abrirla al público general en el futuro.

3. **Tesis doctorales industriales:** Se plantean un máximo de 7 tesis anuales que serán dirigidas por personal de DATAI y codirigidas por personal doctor del Banco con

	2020	2021	2022	2023	2024	2025
Adscritos	10	12	14	16	18	20
Asociados	42	42	44	46	48	50
Invitados	18	18	21	24	27	30
TOTAL	70	72	79	86	93	100

Tabla 3 – Proyección de tesis, alumnos de máster y certificaciones del acuerdo DATAI-BBVA

DATAI trabajará durante los próximos años para consolidar el acuerdo y fomentar la colaboración en publicaciones científicas con profesores y doctores de la Universidad de



Navarra y profesionales del BBVA. Los doctorandos y los empleados de BBVA que ya son doctores realizarán tareas de investigación conjuntas de interés para el Banco y para los investigadores de la Universidad. Dichas tareas se podrán presentar en congresos internacionales y que tengan como fruto la publicación de artículos en revistas de impacto. Además BBVA participa activamente en la docencia del máster.

DATAI tiene como objetivo en los acuerdos estratégicos:

- Mantener el acuerdo estratégico con BBVA y extenderlo con la firma de proyectos de investigación asociados al mismo.
- Establecer un acuerdo programa con otra organización.

5.2 SofIA

DATAI está desarrollando SofIA, un programa de innovación abierta dedicado exclusivamente a Inteligencia Artificial. Pretende unir soluciones de startups con foco en IA con las necesidades de grandes empresas del sector de la banca, farma, salud, seguros y distribución.

El programa se vertebral a través de tres figuras, los *Advocates*, *Startups* y *Mentores*.

- **Advocates:** Son grandes empresas que tienen dificultades para innovar debido a su estructura y tamaño y buscan comenzar a utilizar la IA como elemento diferenciador en sus negocios.
- **Startups:** Pequeñas empresas que han desarrollado un producto basado en IA y que buscan probarlo y mejorarlo.
- **Mentores:** Directivos con dilatada experiencia que ayudan a las *Startups* a sacar el mayor partido posible a la experiencia.

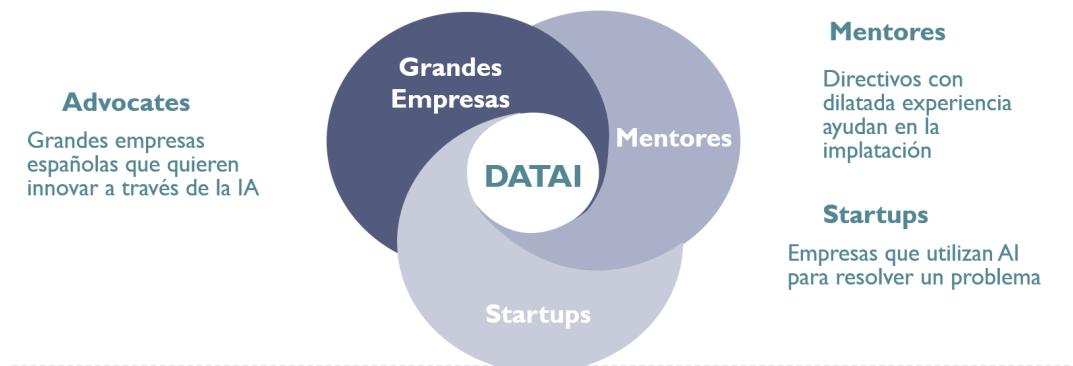


Figura 3 – Estructura del programa SofIA que organizará DATAI.

El programa se desarrollará conjuntamente con Innovation Factory de la universidad que aportará su experiencia en innovación y emprendimiento. Además, se establecerán alianzas con las grandes tecnológicas del mercado (Amazon, Microsoft, Accenture) para que



colaboren con el programa incrementando su difusión aumentando su prestigio. El programa tiene una duración de 6 meses y se realizará en la sede de Madrid y se espera que tenga una gran repercusión significativa en los medios.

El desarrollo de este programa tiene como objetivos:

- Contar con al menos 4 grandes empresas que actuarán como Advocates dentro del programa.
- Posicionar a DATAI y la universidad como organizaciones pioneras en el desarrollo de la IA y su implantación en las empresas españolas. La IA ha sido definida como una de las áreas estratégicas por el gobierno de España dentro de su estrategia digital 2020-2025.
- Fomentar un ecosistema de intercambio de conocimiento entre la universidad y las empresas en el entorno de la IA que a medio plazo permita desarrollar alianzas con grandes empresas para la investigación y el desarrollo de proyectos de Big Data.

5.1 Colaboraciones internas

El Instituto promoverá que las colaboraciones con Centros y grupos de la Universidad sean siempre colaboraciones centradas en la investigación.

5.1.1 De investigación.

DATAI colabora activamente con diferentes centros e institutos de la universidad. Cabe destacar las establecidas hasta la fecha con el Instituto de Cultura y Sociedad (ICS) y los grupos Atlantes en el que una persona de DATAI participa dedicando un 20% de su tiempo.

Dada su naturaleza interdisciplinar el instituto busca desarrollar colaboraciones con diferentes facultades y consolidar estas en acuerdos que permitan desarrollar investigaciones a largo plazo. Prueba de ello son las experiencias iniciadas recientemente con enfermería, nutrición y medio ambiente de las que se espera publicar. El objetivo del instituto en cuanto a colaboraciones internas es convertirse en la entidad de referencia para la analítica avanzada y colaborar en la medida con el resto de investigadores de la universidad para elevar la excelencia investigadora aportando su profunda experiencia en el análisis de datos. Los objetivos de las colaboraciones de investigación son:

- Aparecer como autor en las publicaciones que se produzcan durante la colaboración.
- Pedir proyectos competitivos que permitan financiar al instituto.



5.1.2 De servicios.

Hasta la fecha en el año 2020 se han llevado un total de 10 colaboraciones con enfermería, comunicación, psicología, CUN, admisión y alumni.

DATAI apoyará en forma de servicio que serán llevados a cabo prioritariamente por personal técnico y el Instituto no destinará más del 20% de su capacidad a esta tarea. El objetivo de las colaboraciones de servicios es:

- Prestar servicio a la universidad en el contexto del análisis de datos y en la medida de lo posible convertir las colaboraciones en publicaciones científicas.
- Conseguir contratos de transferencia con empresas que ayuden a posicionar al instituto y financiarlo.

5.3 Proyecto GAIA-X

[Gaia-X](#) es un proyecto iniciado por Europa en el que representantes de la empresa, la política y la ciencia de Europa y de todo el mundo trabajan juntos, mano a mano, para crear una infraestructura de datos federada y segura. Las empresas y los ciudadanos cotejarán y compartirán datos, de manera que mantengan el control sobre ellos. Deberán decidir qué ocurre con sus datos, dónde se almacenan, y conservar siempre la soberanía de los datos.

La estructura organizativa de Gaia-X se basa en tres pilares: la Asociación Gaia-X, los centros nacionales Gaia-X y la Comunidad Gaia-X. Dentro de éstos, hay varios grupos de trabajo y comités. El intercambio dentro, entre y más allá de estos pilares hacia otras partes interesadas (por ejemplo, la Comisión de la UE, las iniciativas internacionales) está garantizado. Existen diferentes Grupos de Trabajo (GT) en el proyecto para crear espacios de datos agrupados por temáticas:

- GT de Espacio de datos de Turismo (incluyendo Cultura y Deportes)
- GT de Espacio de datos de Salud
- GT de Espacio de datos de Industria 4.0
- GT de Espacio de datos de Movilidad (incluyendo tráfico e infraestructuras de transporte)
- GT de Espacio de datos agroalimentario
- GT de Espacio de datos Sociales
- GT de Espacio de datos de Cambio climático (incluyendo medioambiente, economía circular y energías renovables)
- GT de Espacio de datos de Ingeniería y Construcción (incluyendo SmartCities, arquitectura y redes de agua)
- GT de Tecnologías habilitadoras digitales
- GT de Gestión del dato



- GT de Ética, legislación y responsabilidad

La universidad a través de DATAI fue seleccionada por el hub español de GAIA (creado a finales de 2021) para coliderar el desarrollo del espacio de datos de salud en España. Esto representa una oportunidad excelente para posicionar a ayudar a la universidad a erigirse como líder en el desarrollo del futuro de los datos en salud, sector estratégico y con amplia experiencia para la universidad.

GAIA a lo largo de su desarrollo está previsto que se financie con fondos europeos (en las iniciativas Horizon Europe y Digital Europe) y a nivel nacional la SEDIA (Secretaría de Estado de Digitalización e Inteligencia Artificial) también sacará convocatorias. Cabe destacar también que el Gobierno de Navarra es redactor y fundador de GAIA España pudiendo generar a medio plazo colaboraciones que permitan hacer de Navarra un territorio más avanzado y competitivo en tema de datos y sobre todo en salud. El objetivo de DATAI en el contexto de GAIA-X es:

- Trabajar con la CUN, la Facultad de Medicina y el CIMA para definir una estrategia y sacar el mayor rendimiento a la posición de liderazgo en el grupo de datos de salud de GAIA implementando casos de uso que permitan su desarrollo y pidiendo financiación para llevarlos a cabo.



4. COMUNICACIÓN Y DIVULGACIÓN

Se coordina con el departamento de comunicación del vicerrectorado de investigación. Su actividad consiste en los siguientes apartados.

Webs

Existen dos sitios web que permiten dar visibilidad en los medios digitales al instituto y al Máster:

- [DATAI](#)
- [MEBS](#)

DATAI trabajará en la mejora continua de las webs y su posicionamiento para incrementar su reconocimiento e impacto en las redes.

Memoria de actividades

Recientemente DATAI completó la redacción de una memoria de actividades que lista de forma detallada todas las actividades realizadas por el instituto en su corta existencia.

Esta memoria se utilizará para dar difusión a DATAI y como posible herramienta de marketing para presentar el instituto a otras organizaciones.

Concurso de Incubadora de Sondeos y Experimentos

El compromiso de DATAI para la difusión de la estadística en la sociedad lleva al instituto a liderar la organización de la incubadora de sondeos y experimentos. Para ello DATAI colabora con el Departamento de Estadística, Informática y Matemáticas de la Universidad Pública de Navarra (UPNA), con la colaboración de la Sociedad Navarra de Profesores de Matemáticas “TORNAMIRA” Matematika Irakasleen Nafar Elkartea; del Instituto de Estadística de Navarra; y del Departamento de Educación del Gobierno de Navarra y la delegación provincial del Instituto Nacional de Estadística.

El concurso se realiza anualmente, y está dirigido a alumnos de E.S.O., Bachillerato y Ciclos Formativos de grado básico o medio en Navarra con interés por la estadística y la ciencia de los datos. Los participantes deben realizar un proyecto de estadística, probabilidad sobre un tema de interés elegido por el equipo: sociedad, salud, ecología, economía y consumo, medio ambiente.



ANEXO I - Investigación

A - Proyectos de investigación

Proyectos Oficiales Concedidos:

MINECO 2020-2021: Diseños Experimentales para Modelos en Ingeniería Salud y Ciencia de los Datos.

Beca FPU: Álvaro Cía, ya contratado y haciendo la tesis.

Talento senior 2021 ANDIA (GN): Supone la contratación de un senior y dos ayudantes para que forme un equipo.

Proyectos oficiales solicitados:

Proyecto INE: Construcción y análisis de grafo de conocimiento para el análisis de datos estadísticos abiertos enlazados

Proyecto de Transición ecológica y digital de la AEI conjuntamente con BIOMA.

Contratos de investigación con empresas:

Procesamiento del lenguaje natural (firmado).

Parches de glucosa, etc (en negociación).

B - Artículos científicos 2020 - 2021

- De La Calle-Arroyo, Carlos; López-Fidalgo, Jesús; Rodríguez-Aragon, Licesio J. **Optimal designs for Antoine Equation.** CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS. 214, ELSEVIER, 15/07/2021. ISSN 0169-7439, ISSN 1873-3239. Q1.
- Belar, Alazne; Arantzamendi, María; Santesteban, Yolanda; López-Fidalgo, Jesús; Martínez, Marina; Lama, Marcos; Rullan, María; Olza, Ines; Breeze, Ruth; Centeno, Carlos. **Cross-sectional survey of the wish to die among palliative patients in Spain: one phenomenon, different experiences.** BMJ SUPPORTIVE & PALLIATIVE CARE. 11- 2, pp. 156 - 162. BMJ PUBLISHING GROUP, 01/06/2021. ISSN 2045-435X, ISSN 2045-4368. Q2
- Pallotti, María Caterina; López-Fidalgo, Jesús; Centeno, Carlos; Celin, Daniela; Biasco, Guido; Giovannini, Maddalena; Maltoni, Marco; Noguera, Antonio. **Does Delirium Phenomenology in Persons with Advanced Cancer follow a Specific Pattern?** JOURNAL OF PALLIATIVE MEDICINE. 24 - 7, pp. 1061 - 1066. MARY ANN LIEBERT, INC, 16/04/2021. ISSN 1096-6218, ISSN 1557-7740. Q2



- Gamero-Salinas, Juan; Kishnani, Nirmal; Monge-Barrio, Aurora; López-Fidalgo, Jesús; Sanchez-Ostiz, Ana. **The influence of building form variables on the environmental performance of semi-outdoor spaces. A study in mid-rise and high-rise buildings of Singapore.** ENERGY AND BUILDINGS. 230, ELSEVIER SCIENCE SA, 01/01/2021. Q1
- Landecho, Manuel F.; Alegria-Murillo, Leire; López-Fidalgo, Jesús; Colina, Inmaculada; Santesteban, Virginia; Garcia-Unciti, Marisol; Beloqui, Oscar; Frühbeck, Gema; Cuervo, Marta. **Unravelling gender-specific factors that link obesity to albuminuria.** EUROPEAN JOURNAL OF CLINICAL INVESTIGATION. 50 - 11, WILEY, 01/11/2020. ISSN 0014-2972, ISSN 1365-2362. Q1.
- López-Fidalgo, Jesús; Amo-Salas, Mariano. **Optimal dose calibration in radiotherapy.** RADIATION PHYSICS AND CHEMISTRY. 174, PERGAMON-ELSEVIER SCIENCE LTD, 01/09/2020. ISSN 0969-806X. Q1.
- Higuera, Manuel; Howes, Adam; López-Fidalgo, Jesús. **Optimal experimental design for cytogenetic dose-response calibration curves.** INTERNATIONAL JOURNAL OF RADIATION BIOLOGY. 96 - 7, pp. 894 - 902. TAYLOR & FRANCIS LTD, 02/07/2020. ISSN 0955-3002, ISSN 1362-3095. Q1.
- Gamero-Salinas, Juan; Monge-Barrio, Aurora; Kishnani, Nirmal; Lopez-Fidalgo, Jesus; Sanchez-Ostiz, Ana. **Passive cooling design strategies as adaptation measures for lowering the indoor overheating risk in tropical climates.** ENERGY AND BUILDINGS. 252, ELSEVIER SCIENCE SA, 01/12/2021. ISSN 0378-7788, ISSN 1872-6178. Q1.
- Gamero-Salinas, Juan; Kishnani, Nirmal; Monge-Barrio, Aurora; Lopez-Fidalgo, Jesus; Sanchez-Ostiz, Ana. **Evaluation of thermal comfort and building form attributes in different semi-outdoor environments in a high-density tropical setting.** BUILDING AND ENVIRONMENT. 205, PERGAMON-ELSEVIER SCIENCE LTD, 01/11/2021. ISSN 0360-1323, ISSN 1873-684X. Q1.
- Belar, Alazne; Martinez, Marina; Centeno, Carlos; Lopez-Fidalgo, Jesus; Santesteban, Yolanda; Lama, Marcos; Arantzamendi, Maria. **Wish to die and hasten death in palliative care: a cross-sectional study factor analysis.** BMJ SUPPORTIVE & PALLIATIVE CARE. BMJ PUBLISHING GROUP, 14/10/2021. ISSN 2045-435X, ISSN 2045-4368. Q2.
- Berni, Rossella; Lopez Fidalgo, Jesus Fernando; Vining, Geoff G. **Foreword to the Special Issue on "Statistics, Statistical Engineering, and Innovation for Industry 4.0"** FOREWORD. APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY. 37 - 5, pp. 837 - 838. 23/09/2021. ISSN 1524-1904, ISSN 1526-4025. Q3.



- Arias-Casais, Natalia; Lopez-Fidalgo, Jesus; Garralda, Eduardo; Jose Pons, Juan; Rhee, John Y.; Lukas, Radbruch; de Lima, Liliana; Centeno, Carlos. **Trends analysis of specialized palliative care services in 51 countries of the WHO European region in the last 14 years.** PALLIATIVE MEDICINE. 34 - 8, pp. 1044 - 1056. SAGE PUBLICATIONS LTD, 10/06/2020. ISSN 0269-2163, ISSN 1477-030X. Q1.
- Clark, David; Baur, Nicole; Clelland, David; Garralda, Eduardo; Lopez-Fidalgo, Jesus; Connor, Stephen; Centeno, Carlos. **Mapping Levels of Palliative Care Development in 198 Countries: The Situation in 2017.** JOURNAL OF PAIN AND SYMPTOM MANAGEMENT. 59 - 4, pp. 794 - +. ELSEVIER SCIENCE INC, 01/04/2020. ISSN 0885-3924, ISSN 1873-6513. Q2.
- Garcia-Rodenas, Ricardo; Carlos Garcia-Garcia, Jose; Lopez-Fidalgo, Jesus; Angel Martin-Baos, Jose; Wong, Weng Kee. **A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs.** COMPUTATIONAL STATISTICS & DATA ANALYSIS. 144, ELSEVIER, 01/04/2020. ISSN 0167-9473, ISSN 1872-7352. Q2.
- Moreno-Ajona, David; Irimia, Pablo; Rodriguez, Jose Antonio; Garcia-Veloso, Maria Jose; Lopez-Fidalgo, Jesus; Fernandez-Alonso, Leopoldo; Grochowitz, Lukasz; Munoz, Roberto; Dominguez, Pablo; Gallego-Cullere, Jaime; Martinez-Vila, Eduardo. **Elevated circulating metalloproteinase 7 predicts recurrent cardiovascular events in patients with carotid stenosis: a prospective cohort study.** BMC CARDIOVASCULAR DISORDERS. 20 - 1, BMC, 26/02/2020. ISSN 1471-2261. Q3.
- Sánchez-Cárdenas, M. A., Garralda, E., Arias-Casais, N. S., Benítez, E., Van Steijn, D., Moine, S., ... & Centeno, C. (2021). **Palliative care integration indicators: An European regional analysis.** BMJ Supportive & Palliative Care. Q2.
- Aguilar, C., Serna-Jiménez, J., Benítez, E., Valencia, V., Ochoa, O., & Sotelo, L. I. (2021). **Influence of high power ultrasound on natural microflora, pathogen and lactic acid bacteria in a raw meat emulsion.** Ultrasonics Sonochemistry, 72, 105415. Q1.
- Balaguer, Á., Benítez, E., De la Fuente, J., & Osorio, A. (2021). **Maternal and paternal parenting styles as a whole: validation of the simple form of the parenting style evaluation scale.** Anales de Psicología/Annals of Psychology, 37(1), 77-87. Q3.
- Balaguer, Á., Benítez, E., Albertos, A., & Lara, S. (2020). **Not everything helps the same for everyone: relevance of extracurricular activities for academic achievement.** Humanities and Social Sciences Communications, 7(1), 1-8. Q2.

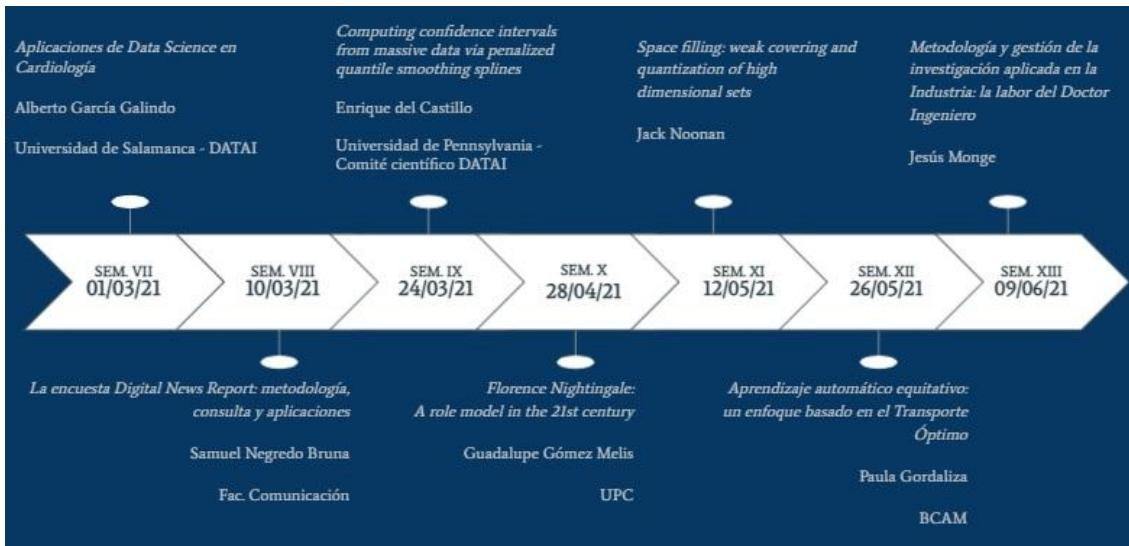
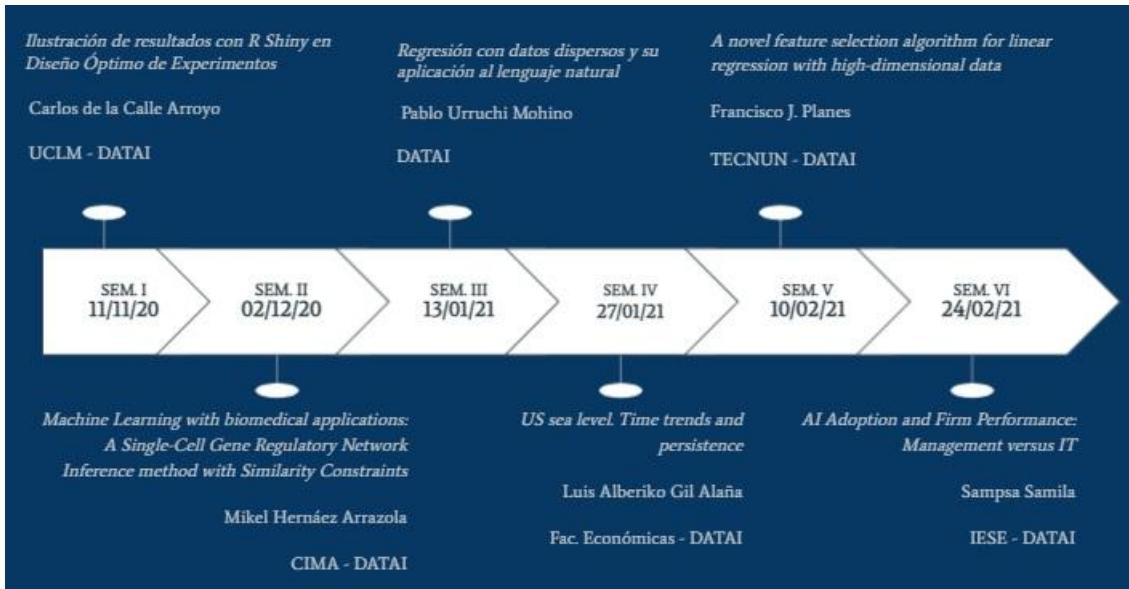


- Sánchez-Cárdenas MA, Garralda E, Benítez E, Arias-Casais N, van Steijn D, Centeno C. **Palliative Care Coverage across European National Health Systems: Proposal of a Synthetic Indicator.** International Journal of Environmental Research and Public Health. 2021; 18(20):10753. Q2.
- Balaguer, Á., Benitez, E., de la Fuente, J., & Osorio, A. (2021). **Structural empirical model of personal positive youth development, parenting, and school climate.** Psychology in the Schools. Q3.
- Calatrava, M., de Irala, J., Osorio, A., Benitez, E., & Lopez-del Burgo, C. (2021). **Matched and Fully Private? A New Self-Generated Identification Code for School-Based Cohort Studies to Increase Perceived Anonymity.** Educational and Psychological Measurement, 00131644211035436. Q2.
- Albertos, A., Koning, I., Benítez, E., & De Irala, J. (2021). **Adolescents' Alcohol Use: Does the Type of Leisure Activity Matter? A Cross-National Study.** International journal of environmental research and public health, 18(21), 11477. Q1.
- Martinez, I., Viles, E., & Olaizola, I. G. (2021). **Data Science Methodologies: Current Challenges and Future Approaches.** Big Data Research, 24, 100183. Q1.
- Otamendi, U., Martinez, I., Quartulli, M., Olaizola, I. G., Viles, E., & Cambarau, W. (2021). **Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules.** Solar Energy, 220, 914-926. Q2.
- Bejines, C.; Aranza-Trevijano, S.; Elorza, J. **On Self-Aggregations of Min-Subgroups.** Axioms 2021, 10, 201. Q1.
- Bejines, C; Aranza-Trevijano, S; Chasco, MJ; Elorza, J. **Aggregation of indistinguishability operators.** Fuzzy Sets and Systems, 2021. Q1.

C - Actividades científicas

- **I Jornada Científica:** 16 de diciembre de 2020. La Vicerrectora de investigación, Iciar Astiasarán, fue la encargada de la apertura de esta jornada. A continuación varios responsables de centros y facultades, como CIMA, Tecnun, CUN, Derecho, Económicas, DATAI y CEIT hicieron sus presentaciones.

Además DATAI organiza seminarios científicos con el objetivo de difundir la investigación en Big Data y la ciencia de los datos. A continuación se presentan los seminarios que tuvieron lugar en el curso 2020 - 2021.



Incrementar la producción científica, incrementando las publicaciones en Q1.

Otras actividades Científicas:

- a. Organización de jornadas y congresos. Hay dos previstos:
 - a) Jornadas científicas del 9 al 13 de mayo de 2022.
 - b) JEDE: Encuentro internacional de diseño óptimo de experimentos y bioestadística en 2023.
- b. Presentación de trabajos en congresos nacionales e internacionales (conferencias plenarias, sesiones invitadas y contribuciones)



Implementar plan de organización interna entre miembros adscritos, asociados e invitados.

C - Otras actividades

- **I Reunión Consejo Asesor:** 14 de diciembre de 2020. Asistieron representantes de prácticamente todos los centros. El Director del instituto, Jesús López Fidalgo, hizo una presentación del instituto, y, posteriormente, se abordaron propuestas para impulsar el crecimiento de DATAI.
- **Visita de Dña. Carme Artigas,** Secretaria de Estado de Digitalización e Inteligencia Artificial y la directora general de IA, **Laura Flores:** 23 de marzo de 2021.
- **Reunión con Guzman Garmendia, Director general de Telecomunicaciones y Digitalización** (DGTT) del Gobierno de Navarra: 24 Junio de 2021.

DATAI realizará actividades y reuniones divulgativas con el objetivo de dar a conocer el instituto y establecer así nuevas alianzas, reclutar talento y conseguir nuevos proyectos que ayuden a financiar al instituto.