



SYNERATECH

**Servicios de Consultoría en Business Intelligence
para el Sector de Investigación Sanitaria**

-Donde los descubrimientos se convierten en soluciones-

MÁSTER EN BUSINESS INTELLIGENCE Y BIG DATA

2022/2023

Adrián Turiel Charro

Cristina Jiménez Aguirre

Francisco Javier García Campos

Gabriela Vartic Olaru

Marta García Álvarez

Rafa Vera Marañón



SYNERATECH

1.	INTRODUCCIÓN.....	5
1.1.	Contexto	5
1.2.	Oportunidad de negocio y objetivo	9
1.3.	Motivación y viabilidad de la propuesta.....	10
1.4.	Identificación del Problema	11
1.5.	Propuesta de Solución	14
1.6.	Hipótesis a validar	15
2.	INVESTIGACIÓN Y VALIDACIÓN DE LOS DATOS	16
2.1.	Proceso de Validación.....	16
2.1.1.	Métodos de Investigación Observacional	16
1)	Encuestas Online	16
2)	Entrevistas Personales	16
2.1.2.	Métodos de Investigación Experimental	16
2.1.3.	Validación de las hipótesis	17
2.2.	Conclusiones	18
2.3.	Análisis de Competidores	19
2.3.1.	Futuver	20
2.3.2.	Clarcat	21
2.3.3.	NorayBio	22
2.3.4.	Conclusiones.....	23
3.	ANÁLISIS Y DIAGNÓSTICO: DAFO	24
3.1.	Debilidades	24
3.2.	Amenazas	24
3.3.	Fortalezas	25
3.4.	Oportunidades	25
4.	PLAN ESTRATÉGICO	27
4.1.	Modelo de negocio	27
4.1.1.	Propuesta de valor.....	27
4.1.2.	Segmento de clientes	27
4.1.3.	Relación con los clientes.....	28
4.1.4.	Canales de distribución	28
4.1.5.	Actividades clave	29
4.1.6.	Recursos clave	29

4.1.7.	Socios/Proveedores clave	29
4.1.8.	Fuente de ingresos	30
4.1.9.	Estructura de costes.....	30
4.2.	Misión, Visión y Objetivos.....	30
5.	PLAN DE ACCIÓN.....	31
5.1.	Alcance del proyecto	31
5.2.	Análisis de recursos	31
5.2.1.	Recursos humanos	31
5.2.2.	Estructura organizativa.....	31
5.3.	Mapa de procesos	32
5.4.	Procesos clave	34
5.5.	Cronograma	36
6.	SOLUCIÓN TECNOLÓGICA	37
6.1.	Producto o Servicio Mínimo Viable.....	37
6.2.	Definición de la solución.....	37
6.3.	Base de datos en la nube.....	37
6.4.	Máquina Virtual.....	38
6.5.	Tecnologías, herramientas y recursos utilizados.....	38
6.5.1.	Bases de datos	38
6.5.2.	Google Colab	39
6.5.3.	Librerías Python.....	39
6.5.4.	Power BI.....	40
6.5.5.	Neo4j.....	40
6.5.6.	GitHub.....	40
6.5.7.	Gestión y Comunicación	41
6.5.8.	Google Cloud Platform (GCP)	41
6.5.9.	Docker	42
6.6.	Ánalisis de las fuentes de datos	43
6.7.	Modelo Lógico	44
6.7.1.	Modelo lógico relacional.....	44
6.7.2.	Modelo lógico no relacional	45
6.8.	Arquitectura.....	45
6.9.	Solución propuesta	46
6.9.1.	Capa de servicio.....	46
6.9.2.	Capa de aplicación.....	49
6.9.3.	Visualización de datos.....	50
A.	PowerBI	50
B.	Neo4j.....	58

C. Procesamiento de Lenguaje Natural a través de Interfaz Gráfica de Usuario (GUI)	61
6.9.4. Prospectiva	62
7. INDICADORES CLAVE DE DESEMPEÑO (KPIs) PARA SYNERATECH	64
7.1. Indicadores de Plataforma.....	64
7.2. Indicadores Financieros	64
7.3. Indicadores de Relación.....	65
7.4. Simulación de los Indicadores.....	65
8. ASPECTOS FINANCIEROS	69
8.1. Rentabilidad del proyecto	69
8.2. Ingresos.....	70
8.3. Gastos.....	71
8.4. Conclusiones	72
9. CONCLUSIONES.....	75
10. BIBLIOGRAFÍA.....	77
11. ANEXO I: Glosario de Campos de la BBDD	78
12. ANEXO II: MODELO LÓGICO RELACIONAL	92
13. ANEXO III: ENCUESTA VALIDACIÓN DE HIPÓTESIS	93
14. ANEXO IV: RESUMEN EJECUTIVO	94

1. INTRODUCCIÓN

1.1. Contexto

El **cáncer es la principal causa de muerte en todo el mundo**, y se espera que sea la barrera más importante para aumentar la esperanza de vida en todos los países del mundo. Según la Organización Mundial de la Salud (OMS), en 2019, el cáncer fue la primera o segunda causa de muerte antes de los 70 años en 112 de 183 países, y ocupó el tercer o cuarto lugar en otros 23 países (1) (2). La **incidencia y la mortalidad por cáncer están creciendo rápidamente en todo el mundo**, específicamente ha crecido del 14,9% en 2007 al 17,1% en 2017 del total de muertes en todo el mundo (3). Esta tendencia de crecimiento que se sigue observando a lo largo de los años se ha reflejado en las predicciones del volumen de mercado mundial de la oncología, mostrando una tasa de crecimiento a 10 años del 8,8% como recoge la consultora Precedence Research en su informe de 2023 (**Figura 1.1**). En dicho informe se recoge que en 2022 el diagnóstico del cáncer ocupó el 44% del mercado y el tratamiento el 56%, mientras que el mercado que más crecimiento experimentó fue el europeo (4).



Figura 1.1 Proyección del tamaño de mercado de la oncología desde 2022 a 2032 (4).

Sin embargo, el cáncer no ha sido el único desafío contra la salud que hemos sufrido en los últimos años. En 2019, la **pandemia de COVID-19** azotó al mundo entero y **desafió nuestro sistema de salud y economía**, y en consecuencia, ha reconfigurado nuestras prioridades y enfoques en muchos ámbitos, incluyendo el empresarial (5).

Los mercados no son ajenos a estas circunstancias y en los últimos años se apreció un **incremento sin precedentes de las inversiones de capital de riesgo en el sector salud** en Estados Unidos según el Silicon Valley Bank (**Figura 1.2**). En 2022 se produjo un descenso y un posterior estancamiento en las inversiones debido las políticas de aumento de tipo de interés que tomó la Reserva Federal en ese período, que lastran las rentabilidades a largo plazo debido al encarecimiento de los gastos generales de las empresas (6).

En este escenario se ha evidenciado la importancia crucial de la inversión en ciencia y tecnología, y más específicamente, en sectores como el biotecnológico y farmacéutico. La necesidad urgente de desarrollar soluciones eficaces y rápidas, desde pruebas de diagnóstico hasta tratamientos y vacunas, ha demostrado el enorme

potencial de la biotecnología y farmacología para dar respuestas a la crisis sanitaria. Es por esta razón que **el aumento de la inversión privada en estos campos es no solo justificable, sino también estratégicamente esencial.**

US Healthcare Venture Capital Fundraising
2012-H1 2023



Figura 1.2. Inversión de capital riesgo en el sector salud en Estados Unidos (6).

El **desarrollo de una nueva tecnología es un camino muy largo (Figura 1.3)** que comienza en la investigación básica. La **investigación básica** se centra en entender los mecanismos y entidades biológicas que intervienen en una enfermedad o en una patología concreta. En base a este conocimiento, la **investigación aplicada** se centra en interaccionar e intervenir sobre estos mecanismos para lograr un resultado específico. Hasta este punto, **todo este conocimiento se suele desarrollar en centros de investigación públicos y está financiados principalmente por el Estado**, al menos en España. Los resultados de la investigación suelen desembocar en la publicación de los resultados en revistas científicas o en la protección de estos mediante patentes, lo que garantiza el monopolio de explotación de esos resultados a sus inventores y propietarios.

Cuando una tecnología o un proyecto llega a este punto se enfrenta **a la etapa conocida como “valle de la muerte”**. Esta etapa del desarrollo de una tecnología se denomina así porque se encuentra con importantes **dificultades para obtener la financiación necesaria para continuar con su desarrollo**. En ese punto **los costes de la investigación son mucho mayores y la incertidumbre es muy alta** en cuanto al éxito de la tecnología. Por un lado, el **Estado no ofrece suficiente financiación** para desarrollar todos los proyectos que han llegado a este nivel de maduración y por otro lado los inversores del **sector privado consideran muy arriesgadas este tipo de inversiones**. Por esta razón, este es uno de los puntos más delicados de la innovación y que dificultan la llegada de nuevas soluciones a los pacientes.

En este punto, una de las estrategias más comunes consiste en **crear una start-up como vehículo que posibilita la inversión de capital privado** y permite seguir desarrollando la tecnología en el laboratorio. En este escenario, se produce la **transferencia de tecnología** desde los centros públicos a la nueva empresa, es decir, licencian la explotación de los resultados a la nueva start-up y se comienza a buscar financiación. En muchas ocasiones, estas empresas logran la financiación necesaria para **comenzar los ensayos clínicos en pacientes**, que es la fase más cara en todo el proceso de creación de una nueva tecnología. Si tienen éxito, la empresa suele ser **comprada por una farmacéutica más grande** capaz de producir y distribuir el fármaco para que llegue finalmente a los pacientes.

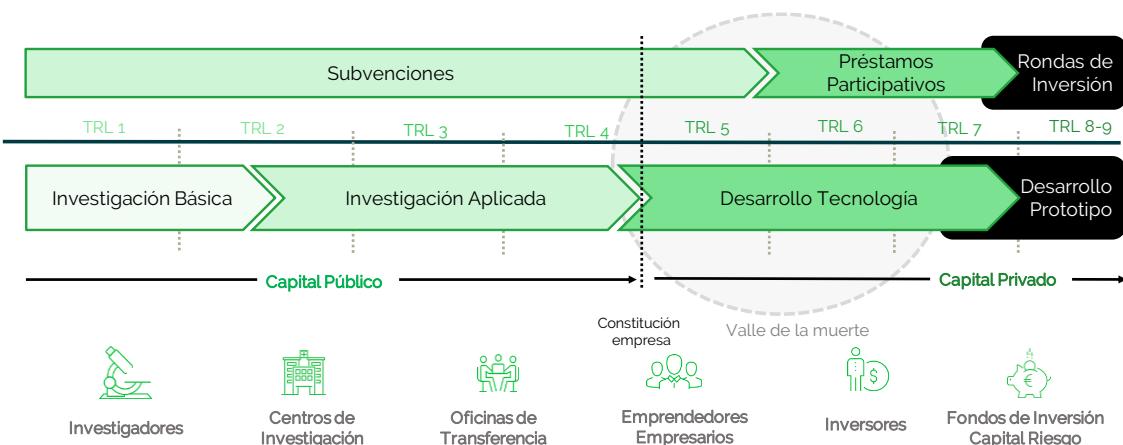


Figura 1.3. Esquema del desarrollo de una nueva tecnología hasta el desarrollo de un prototipo.

Como hemos descrito, para que la investigación e innovación llegue al paciente, es **necesario que se involucren y coordinen diversos agentes de la sociedad** durante el proceso de generación de una nueva tecnología además de los propios investigadores. Desde **agentes estatales y comunitarios** que crean el marco legislativo, la infraestructura necesaria, y los programas de financiación pública para impulsar la investigación. Incluyendo **agentes del sector privado** como pueden ser las farmacéuticas y biotecnológicas que posibilitan la producción, distribución y en muchos casos la propia creación de estas soluciones, e inversores privados como fondos de inversión, business angels, fondos de capital riesgo, etc. que aportan su gran capacidad de financiación a proyectos que el sector público no tiene capacidad para financiar debido a sus limitados recursos. Las **entidades filantrópicas han ido cobrando más importancia en los últimos años** al proporcionar fuentes de financiación sin retorno, con el objetivo de garantizar el impacto social de muchas investigaciones.

En este sentido, el ecosistema innovador español fue analizado y sus carencias fueron descritas detalladamente en el **informe de la OCDE de 2021** ("Mejorar la Transferencia de Conocimiento y la Colaboración entre Ciencia y Empresa en España"), donde se destacaba el **papel necesario de la colaboración entre el sector privado y el público**, y que la falta actual de esta colaboración se debía, entre otras causas, a la falta de cultura empresarial en el sector de la investigación y a la falta de conocimiento técnico, experiencia y confianza en el sector biotecnológico y farmacéutico de los inversores españoles (7). En otras palabras, el **desconocimiento reciproco del papel del investigador y del inversor reducía el desarrollo tecnológico y el impulso de la innovación en España**.

Con la intención de unificar criterios y acciones entorno al desarrollo de nuevas tecnologías que tuvieran impacto en las sociedades europeas, se elaboró la **"Nueva Agenda de Innovación de la UE" en 2022** (8). Este documento recogía una serie de recomendaciones que fueron adoptadas por España en varias iniciativas legislativas y programáticas: *Plan de Transferencia y Colaboración*, *La Ley de Fomento del Ecosistema de las Empresas Emergentes (startups)*, *El Plan de Recuperación, Transformación y Resilencia (PRTR)*, *La Ley de la Ciencia, la Tecnología y la Innovación (LCTI)*, *El proyecto de Ley Orgánica del Sistema Universitario (LOSU)*.

El *Plan de Transferencia y Colaboración* del Ministerio de Ciencia e Innovación **recoge una serie de iniciativas que pretenden unificar criterios y sistematizar la transferencia de conocimiento y las relaciones entre el sector público y privado** en materia de

innovación, entre las cuales se incluyen iniciativas de capacitación de los agentes del ecosistema innovador (9).

Por otro lado, la **Ley de Start-ups** pretende abordar las debilidades en la creación de nuevas empresas a través de diferentes iniciativas como: **beneficios fiscales para compañías e inversores, reducir burocracia para inscripción de nuevas start-ups, estandarizar el concepto de start-up, facilitar permisos de residencia para emprendedores y facilitar la concesión de licencias para pruebas temporales** (10).

No es casualidad que el desarrollo de nuevas tecnologías en la actualidad se vincule al fomento y al desarrollo ágil de nuevas empresas o start-ups. En los últimos años cerca del 60% de los nuevos fármacos aprobados por la FDA (responsable de la regulación de medicamentos en Estados Unidos) se han originado en pequeñas empresas (**Figura 1.4**) (11). Las estadísticas en fusiones y adquisiciones (M&A) en el sector farmacéutico y biotecnológico hasta 2021 indicaban que este tipo de transacciones había aumentado en los últimos años y que las grandes biofármacos estaban apostando cada vez más en adquirir pequeñas entidades o start-ups que desarrollan prototipos hasta la fase preclínica, en lugar de desarrollar ellos mismos la investigación hasta ese punto (6).

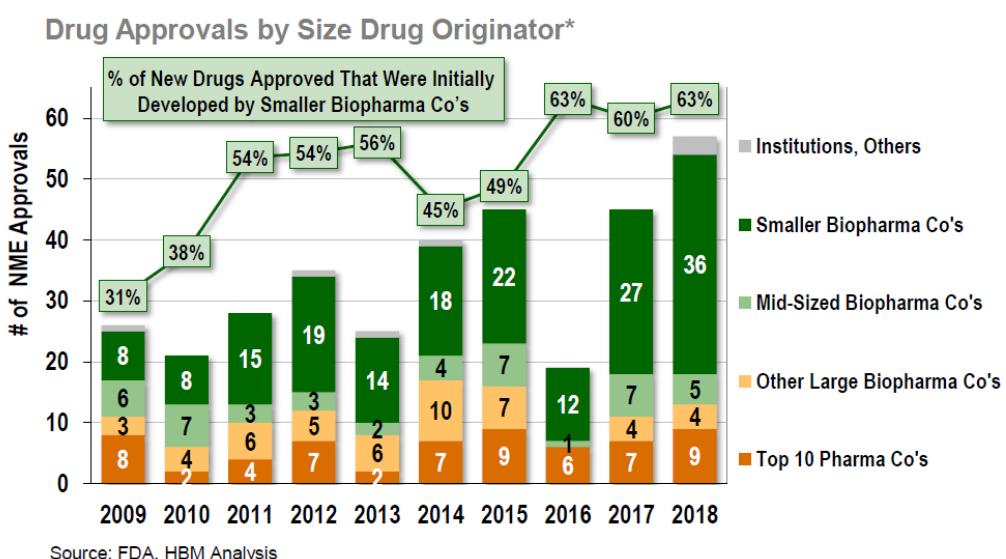


Figura 1.4. Número de fármacos aprobados por la FDA en función del tamaño de la empresa que lo originó (11).

Esto supone una gran oportunidad para los inversores de fondos de capital riesgo, ya que la **inversión en start-ups puede reportarles grandes rendimientos** si esta acaba siendo absorbida por una gran biofarma durante el desarrollo de su tecnología. Por esa razón la inversión de capital privado en este sector se ha multiplicado y concretamente se ha centrado en la financiación de start-ups que surgen en el seno de centros de investigación públicos.

Sin embargo, como mencionábamos anteriormente, el desconocimiento técnico de los inversores y por lo tanto la desconfianza hacia este tipo de inversión lastran la innovación en España. En este sentido, se abre la puerta al uso de herramientas digitales de explotación del dato, al uso de principios de analítica avanzada propios de otros sectores y a la implantación de inteligencia artificial a diversos procesos para solucionar las carencias del ecosistema y mejorar la toma de decisiones. En otras palabras, como ocurre en muchos otros sectores estratégicos en la actualidad europea, se hace

necesaria una transformación digital basada en las nuevas tecnologías de explotación del dato.

El plan de recuperación de la **Comisión Europea Next Generation EU** incluye un nuevo Fondo de Reconstrucción y Resiliencia que también considera una de sus **prioridades el financiar inversiones relacionadas con la Transformación Digital** con miras a impulsar una recuperación económica. **Los datos están en el corazón de esta transformación digital de la UE** que influye en todos los aspectos de la sociedad y la economía y sobre todo pone el foco sobre la salud y las tecnologías verdes (12). En este sentido, estas ayudas públicas a entidades privadas y públicas se articulan en España a través del Plan de Recuperación, Transformación y Resilencia (PRTR) mencionado anteriormente (13), y suponen una oportunidad importante para modernizar el sector biotecnológico y farmacéutico español.

Por otro lado, la **Asociación Española de Empresas Biotecnológicas (AseBio)** en su informe anual 2022 recalcó que la **mayor inversión en I+D** se realizaba entorno a las actividades económicas relacionadas con **servicios de I+D**, que incluyen servicios de **consultoría y transformación digital**, y en el **sector farmacéutico**. También se incluía el sector biotecnológico entre los sectores que más invertían en I+D, solo por detrás de los dos mencionados y por detrás de los vehículos a motor y la educación (14).

1.2. Oportunidad de negocio y objetivo

En base a toda la información expuesta consideramos que:

1. El mercado sanitario entorno a las soluciones para pacientes del cáncer y el sector sanitario en general ha crecido en los últimos años y se prevé que crezca en los siguientes.
2. La inversión global en el sector sanitario ha aumentado en los últimos años.
3. A pesar de esta apuesta por los mercados, se observan grandes carencias en el ecosistema innovador español que dificultan el desarrollo de nuevas tecnologías que puedan llegar al paciente.
4. Se han sembrado importantes iniciativas legislativas en España y Europa para modernizar y potenciar la innovación en el sector sanitario.
5. La Comisión Europea ha apostado por la Transformación Digital enfocada en la explotación del dato como uno de los pilares claves de la recuperación económica entorno a los fondos Next Generation EU.
6. Los servicios de consultoría y transformación digital se encuentran entre los principales servicios de I+D que contrata el sector empresarial español.

Por lo tanto, consideramos que una **empresa conocedora del sector farmacéutico y biotecnológico que ofrezca servicios de consultoría y transformación digital** supondría una piedra angular sobre la que pivotarían las necesidades y soluciones del ecosistema innovador español. La **adecuada gestión de los datos usando principios de Business Intelligence** podrían ayudar a ordenar y mejorar la eficiencia en la gestión de los proyectos de investigación, abordando las principales carencias del sector y promoviendo mejorar su atractivo para **fomentar la inversión privada**, estrechando la relación entre inversores y entidades investigadoras.

El objetivo es ofrecer servicios de consultoría y transformación digital personalizados a diferentes agentes del ecosistema innovador europeo con el objetivo de fomentar la inversión privada, mejorar la gestión de los proyectos de investigación y en último término acelerar el desarrollo de nuevas tecnologías que lleguen al paciente.

1.3. Motivación y viabilidad de la propuesta

Los **miembros fundadores de Syneratech** cuentan con una **amplia experiencia en el ecosistema investigador español y estrecha relación con agentes clave del mismo**, desde investigadores en el área de la oncología, gestores de centros de investigación públicos, directivos de empresas farmacéuticas y biotecnológicas, hasta gestores de fondos de capital riesgo, inversores en general y gestores de entidades filantrópicas.

Concretamente, Syneratech cuenta con una estrecha relación estratégica con la **Fundación Científica de la Asociación Española Contra el Cáncer**. La **Asociación Española Contra el Cáncer**, (en adelante AECC), es una ONG de carácter benéfico asistencial y sin fin de lucro. Se constituyó el 5 de marzo de 1953 con el propósito de luchar contra el cáncer. En la Asociación Española Contra el Cáncer se aborda el cáncer desde un enfoque integral, desde todos los frentes que afectan a la Sociedad. Se realizan actividades de **prevención** (mediante organización de eventos divulgativos, formación, campañas de detección temprana, etc.), **apoyo y acompañamiento a los pacientes de cáncer**, y por último **apoyo a la investigación contra el cáncer**.

La **Fundación Científica AECC** se creó en 1971 en el seno de la AECC y tiene como misión **financiar desde la sociedad civil una investigación de calidad contra el cáncer que mejore la vida de las personas**, en un entorno globalizado, involucrando a todos los sectores de la sociedad en la obtención de resultados. Promover la consolidación y estabilización del sistema de investigación e innovación contra el cáncer en España.

La Fundación **ha destinado más de 100 millones de € en proyectos de investigación desde 2014**, 24,5M€ solamente en 2022. Ha financiado **más de 565 proyectos**, ha financiado contratos de más de **1.000 investigadores** y más de **150 centros de investigación** diferentes han estado involucrados en sus subvenciones.

Por lo tanto, la Fundación cuenta con una amplia experiencia sobre las necesidades y problemas del sector investigador contra el cáncer y dispone de mucha información interesante en relación al desarrollo y financiación de proyectos de investigación contra el cáncer.

Los **valores e intereses de Syneratech están alineados con los de la Fundación Científica**, en el sentido de que **buscamos mejorar el ecosistema investigador español y acelerar el desarrollo de nuevas tecnologías para que lleguen antes a los pacientes**.

Gracias a esta **colaboración estratégica entre Syneratech y la Fundación Científica**, podremos **validar nuestro modelo de negocio en un entorno real** y con **datos** provenientes del **propio sector**.

La viabilidad de la propuesta se valorará a lo largo de esta memoria a través de:

1. Identificación de problemas del sector y propuestas de soluciones.
2. Identificación y validación de hipótesis del mercado.
3. Análisis de potenciales competidores en el mercado.
4. Análisis de debilidades, amenazas, fortalezas y oportunidades.
5. Elaboración de un modelo de negocio detallado.
6. Creación de un producto o servicio mínimo viable (PMV) que sea de utilidad para la Fundación Científica y que permita validar las soluciones tecnológicas que pretendemos ofrecer entre nuestros servicios.

1.4. Identificación del Problema

Entre las carencias del sector innovador en España descritas detalladamente en el informe de la OCDE de 2021 ("Mejorar la Transferencia de Conocimiento y la Colaboración entre Ciencia y Empresa en España") (7), se destacaban:

- **Cultura de innovación poco integrada** en estrategias empresariales que derivaba del desconocimiento técnico de los inversores y una baja cultura del emprendimiento;
- **Baja capacidad de absorción del sistema productivo;**
- **Reducido nivel de confianza** entre empresas e instituciones de investigación;
- **Escaso conocimiento de la investigación** llevada a cabo en el sector público de investigación por parte de los propios investigadores y los inversores;
- **Barreras normativas y burocráticas a la colaboración en I+D+I;**
- **Complejidad de las medidas de apoyo a las actividades de I+D+I colaborativa;**

Tras realizar varias entrevistas con directoras y gestores de la Fundación Científica de la AECC, oficinas de transferencia de centros de investigación públicos, investigadores, expertos en transferencia de tecnología, miembros de fondos capital riesgo e inversores en general, se han identificado varios problemas y necesidades del sector que recogemos a continuación y que coinciden con las conclusiones del informe de la OCDE de 2021:

	Problema	Causas (Contexto)	Solución Propuesta
INVESTIGADORES	Falta de financiación	Existe una alta demanda de fondos para financiar proyectos de investigación oncológica en España. Los investigadores necesitan acceder a recursos financieros adecuados para llevar a cabo estudios clínicos, adquirir equipamiento y contratar personal especializado. Además, se requiere una mayor diversificación de las fuentes de financiamiento para garantizar la sostenibilidad de estos proyectos.	Atracción de inversión privada hacia el sector de la investigación en salud. Desarrollo de herramientas digitales que faciliten la toma de decisiones a inversores en este sector.
	Falta de colaboración interdisciplinaria	Los proyectos de investigación oncológica requieren la participación de expertos de diferentes disciplinas, como oncólogos, biólogos, bioinformáticos, estadísticos, entre otros. Existe una necesidad de fomentar la colaboración y el intercambio de conocimientos entre estos profesionales para impulsar la innovación y avanzar en la lucha contra el cáncer.	Herramienta de visualización que permita la rápida identificación de proyectos y resultados relacionados con la misma enfermedad (o tipo de cáncer en concreto), tipo de investigación o los mismos investigadores
	Dificultad para identificar sinergias	Debido a la complejidad de los proyectos y a que su clasificación no está indexada, es difícil establecer colaboraciones entre investigadores. Ellos mismos se buscan en función de relaciones personales o afinidades particulares.	Algoritmo de clasificación automática de proyectos y herramienta de visualización que permita identificación rápida de proyectos similares

	Problema	Causas (Contexto)	Solución Propuesta
GESTORES INVESTIGACIÓN	Falta de acceso a datos y muestras	Los investigadores necesitan acceder a bases de datos y muestras biológicas de calidad para llevar a cabo sus investigaciones. Esto implica la necesidad de establecer acuerdos y mecanismos de colaboración con hospitales, centros de investigación y bancos de muestras para asegurar el acceso a este tipo de recursos.	Construcción de bases de datos en la nube con estándares comunes y compartidos por varias entidades
	Falta de formación y capacitación en transferencia tecnológica	La transferencia tecnológica es una asignatura pendiente constante en la investigación contra el cáncer. La carencia de profesionales en transferencia tecnológica, valorización de proyectos y elaboración de estrategias de mercado lastra la llegada de nuevas tecnologías al paciente. Es necesario atraer este tipo de perfiles al sector.	Acceso a cursos de capacitación ofrecidos por los mejores expertos del sector. Servicios de capacitación para investigadores.
	Dificultad para optimizar recursos	Debido a la especificidad técnica de los proyectos se deja a criterio exclusivo del investigador los suministros que utiliza. Sería posible optimizar los gastos de los proyectos si los gestores supieran cuales son las técnicas y los equipos más usados en sus centros o empresas.	Herramientas de gestión de resultados que sugieran acciones para ayudar en la toma de decisiones en base a modelos de predicción
EXPLOTACIÓN DEL DATO	Bases de datos limitadas y rígidas	Muchos centros de investigación no cuentan con expertos en el análisis o ingeniería de datos que diseñen bases de datos escalables y por lo tanto, al cabo de los años, las bases son ineficientes o quedan obsoletas para ser explotadas por nuevas herramientas	Servicios de asesoría y desarrollo tecnológico para elaborar bases de datos escalables, eficientes y personalizadas a las necesidades del cliente
	La mayoría de la información de la base de datos no puede ser explotada	Por falta de planificación y conocimiento de los centros de investigación, mucha información queda recogida en archivos poco explotables como documentos PDFs	Servicios de desarrollo de modelos para recuperar información de documentos digitales no explotables como PDFs
	Escasa o nula política de calidad del dato	Muchos centros de investigación no cuentan con expertos en gobernanza del dato por lo que no han elaborado una guía de criterios ni han establecido protocolos para revisar y garantizar la calidad de los datos durante la recolección y la explotación.	Servicios de análisis y asesoría sobre gobernanza del dato y estrategia de explotación
	Dificultad para extraer estadísticas y conclusiones de su base de datos	En ocasiones los centros carecen de una estrategia adecuada de recolección de datos, lo que provoca que muchas veces no se recolecte la información necesaria para establecer métricas útiles o fiables	Servicio de asesoría de creación de métricas y recopilación de datos para medir productividad y rendimiento de los procesos

	Problema	Causas (Contexto)	Solución Propuesta
INVERSORES	Falta de criterios unificados de evaluación y valorización	La valoración de los proyectos suele ser una tarea delegada en expertos del sector con amplia experiencia en investigación y conocimiento de los mercados, pero muchas veces se rigen por criterios subjetivos y poco contrastables. Muchas veces la autoridad del experto cuenta más que los datos alrededor del proyecto.	Elaboración de una guía de valorización de proyectos con criterios definidos Servicio de asesoría para establecer métricas fácilmente recopilables y cuantificables que reflejen adecuadamente los criterios de valorización Diseñar modelos de machine learning y NLU (Natural Language Understanding) que permitan valorizar automáticamente los proyectos
	Dificultad para valorar los proyectos en función del mercado	Muchas veces los criterios de valorización de los proyectos de investigación no son compatibles o comparables con las métricas que se utilizan en el mundo empresarial, lo que dificulta que los inversores entiendan o confíen en las valorizaciones que acceden de los proyectos	Servicios de asesoría para establecer métricas útiles
	Falta de conocimiento técnico de inversores	Normalmente los inversores no tienen el conocimiento técnico adecuado para apreciar si un proyecto de investigación supone una oportunidad de inversión o de valorar los riesgos a los que se enfrenta, lo que les disuade muchas veces de realizar la inversión	Herramientas de visualización y modelos cuantitativos que proporcionen puntuaciones racionales y entendibles por los inversores
	Falta de difusión y transferencia de resultados	Los resultados de los proyectos de investigación oncológica deben ser difundidos y transferidos a la comunidad científica, médica y a la sociedad en general. Una mayor difusión didáctica de los resultados contextualizados	Herramientas de visualización que permitan apreciar la calidad y producción científica, junto a un resumen didáctico del proyecto

Todo este paradigma nos deja con una gran cantidad de datos sin explotar que pueden ayudar a la toma de decisiones en proyectos relacionados con el ámbito de la salud, ya sea por parte de inversores o por los propios gestores de fundaciones, universidades o entidades relacionadas con la investigación.

El análisis de la problemática concreta de este sector se recoge en el Diagrama de Ishikawa de la **Figura 1.5**.



Figura 1.5. Diagrama de Ishikawa que recoge la problemática concreta del sector investigador español y que Syneratech podría solucionar.

Priorizando las necesidades que pueda resolver Syneratech en un Producto o Servicio o Mínimo Viable para la Fundación Científica y más **en relación con los datos con los que dispone**, se podría abordar el problema de disponer de **bases de datos limitadas y rígidas**, la **dificultad para optimizar recursos** y la **falta de colaboración interdisciplinar o dificultad para identificar sinergias**.

Estos problemas pueden ser extrapolados a otros sectores y, por lo tanto, nos permite crear una solución que responda a la oportunidad de negocio que demandan otros tipos de entidades.

1.5. Propuesta de Solución

En Syneratech buscamos resolver la falta de conocimiento y experiencia en el campo de la explotación de datos del sector de la investigación sanitaria así como la dificultad para optimizar recursos e identificar sinergias en el sector, ofreciendo asistencia personalizada, mediante la aplicación de herramientas de Business Intelligence.

En nuestro PMV elaborado para la Fundación Científica de la AECC pretendemos abordar varios análisis y soluciones que consideramos fundamentales y necesarios para la Fundación y el sector de la investigación en general:

- Construcción de una base de datos escalable y que permita alta flexibilidad en la explotación de sus datos.
- Análisis de gastos y recursos de cada proyecto: Extracción de información de facturas y gastos asociados a cada proyecto para identificar patrones de gastos y optimizar recursos.
- Análisis de sinergias entre investigadores y proyectos: Identificación de colaboraciones potenciales entre investigadores y la identificación de áreas de interés comunes.
- Análisis de estadísticas de la base de datos: Extracción y análisis de estadísticas relevantes de la base de datos para obtener información útil y facilitar la toma de decisiones.
- Análisis de la calidad del dato y establecer protocolos que garanticen la veracidad del dato.

Ofrece una solución tecnológica personalizada que analiza los datos de los proyectos de investigación para conocer la situación actual de las organizaciones y poder proceder a definir KPIs y crear una estrategia de gobierno del dato, con el fin de tomar

mejores decisiones en la elección de los proyectos de investigación, asegurando la viabilidad financiera.

Los resultados se someterán a la validación por parte de las organizaciones y, posteriormente, se desarrollarán mejoras en colaboración con estas. Esto nos posibilitará la obtención de retroalimentación directa de los usuarios, permitiendo evaluar oportunidades de mejora del servicio y la optimización de otros procesos relacionados.

1.6. Hipótesis a validar

Asimismo, con las premisas anteriores y la situación del mercado, se procede a plantear las hipótesis de negocio a validar.

Hipótesis 1: Al menos un 20% de los centros de investigación públicos estarían interesados en contratar servicios de consultoría externa para analizar sus bases de datos e integrar soluciones basadas en la ciencia de datos y Business Intelligence con el objetivo de mejorar la eficacia de su trabajo y sus procesos.

Hipótesis 2: Si ofrecemos servicios de implementación de políticas de gobernanza y explotación de datos a inversores, centros de investigación o empresas, al menos el 35% de ellos reportará una mejora significativa en su toma de decisiones.

Hipótesis 3: Al menos un 30% de inversores en el campo de la salud y la tecnología estarían dispuestos a aumentar su capital invertido y considerar nuevas oportunidades de inversión en este sector si contarán con herramientas que mejoraran su entendimiento sobre los proyectos de investigación y les permitiera realizar un seguimiento del avance de los mismos.

Hipótesis 4: Los centros de investigación públicos estarían interesados en la estandarización de sus datos y su integración en una red que facilitara la identificación sinergias entre proyectos de investigación con el objetivo de aumentar la colaboración entre centros de investigación, lo que resultará en investigaciones más efectivas y con mayor impacto.

Concretamente, un mejor análisis de datos conduciría a mejorar la eficiencia en la gestión, seguimiento, evaluación y valorización de los proyectos, identificando tendencias y sinergias, mejorando su clasificación, gestionando mejor sus recursos, reduciendo la incertidumbre, incrementando tasas de éxito y, definitivamente, mejorando la toma de decisiones en general y aportando valor añadido a los datos.

Finalmente se presenta una hipótesis social que una vez introducido nuestro servicio en el mercado sería interesante de comprobar.

Hipótesis Social: Una mejora en la gestión de datos en el campo de la investigación aumentará la financiación de proyectos y contribuirá a la retención del talento científico.

2. INVESTIGACIÓN Y VALIDACIÓN DE LOS DATOS

2.1. Proceso de Validación

2.1.1. Métodos de Investigación Observacional

Como metodología de validación de hipótesis observacional se escogieron dos técnicas diferentes: **encuestas online** y **entrevistas personales** dirigidas a diferentes profesionales del sector de la investigación, salud y financiero.

Además, con el fin poder construir un producto mínimo viable para la Fundación Científica de la AECC se han realizado **entrevistas personales** con los profesionales encargados de los departamentos clave de la Fundación. Con ello pudimos conocer sus dudas, sugerencias y los problemas que se encuentran en su trabajo referente a los datos que manejan.

También se ha realizado un estudio de mercado para conocer los impedimentos que diferentes profesionales del sector de la investigación, salud y financiero se encuentran a la hora de escoger donde realizar su inversión, conocer los proyectos innovadores que existen en ese momento en el mercado o las ayudas que se ofrecen, entre otras.

1) Encuestas Online

Como herramienta se utilizó Google Forms para crear una encuesta de investigación de mercado (Anexo III) dirigida a tipos de profesionales diferentes:

- Inversores
- Gestores de proyectos de centros de investigación o entidades de financiación sin ánimo de lucro
- Responsables científicos o comerciales de empresas biotecnológicas o farmacéuticas
- Investigadores

Las preguntas de los formularios diseñados se recogen en el ANEXO.

2) Entrevistas Personales

Se realizaron tres entrevistas personales a trabajadoras de la Fundación Científica de la AECC:

- Project Manager y Data Analyst
- Directora de Calidad y Procesos
- Directora Científica/Subdirectora Fundación Científica

2.1.2. Métodos de Investigación Experimental

El desarrollo de soluciones basadas en la explotación de datos para la Fundación Científica de la AECC se usará como experiencia piloto para la validación de hipótesis de manera empírica.

2.1.3. Validación de las hipótesis

La validación de hipótesis basándose en las encuestas online, refuerzan la información recogida en el análisis de mercado y las entrevistas personales.

Hipótesis 1: El 81.3% de los expertos han considerado alguna vez contratar servicios externos para analizar y mejorar la eficacia de su centro de investigación.

Algunos de los retos en la correcta gestión de datos a los que se enfrentan estos profesionales en sus centros de investigación son:

- Dificultad en la interpretación de los datos económicos, ya que no están recogidos en una herramienta que los centralice y estandarice.
- Bases de datos poco elaboradas, desconocimiento de gestión del dato y sus herramientas.
- Falta de financiación, desconocimiento de la metodología de la recopilación de datos.

Debido a estos el **67.8% estarían dispuesto a invertir** en servicios de consultoría para mejorar sus procesos y eficacia.

Por ello, consideramos **validada la hipótesis 1**, más del 20% de los centros de investigación públicos estarían interesados en contratar servicios de consultoría externa para analizar sus bases de datos e integrar soluciones basadas en la ciencia de datos y Business Intelligence con el objetivo de mejorar la eficacia de su trabajo y sus procesos.

Hipótesis 2: El 37.5% de los encuestados no realizan un análisis de datos para la toma de decisiones en su empresa, centro de investigación o a la hora de decidir su próxima inversión. Asimismo, el 93.8 % de estos cree que una mejor gestión y explotación de datos podría influenciar de forma positiva esta toma de decisiones.

Solamente el 12.5% de las instituciones, en las que los expertos trabajan, tienen implementado políticas de gobernanza de datos.

En consecuencia, **existe la posibilidad que el 35%** de ellos, reporten una mejora significativa en su toma de decisiones al crear servicios de implementación de políticas de gobernanza personalizados y explotación de datos.

Hipótesis 3: Para la validación de esta hipótesis se han tenido en cuenta la opinión de los inversores privados o públicos que están en busca de proyectos innovadores. El **45.5% tienen mucho interés** en buscar nuevos proyectos que les permita aumentar el capital invertido, siempre que tuvieran herramientas que mejoren el seguimiento y compresión de dichos proyectos y sus inversiones.

Hipótesis 4: El **75% de los investigadores** que han contestado la encuesta online, consideran que la colaboración entre centros de investigación es importante para el avance de la ciencia.

Solamente el 23.1% opina que los problemas que ha experimentado en la colaboración con otros centros de investigación no son debido a la falta de estandarización de datos.

Asimismo, ellos creen que la colaboración y la identificación de sinergias entre proyectos de investigación mejorarían la efectividad de las investigaciones en su centro.

Hipótesis Social: Su validación constaría en tres pasos:

Paso 1: Medición de la financiación antes y después de la mejora en la gestión de datos.

- Recopila datos sobre la financiación de proyectos de investigación antes de implementar las mejoras en la gestión de datos.
- Implementa las mejoras en la gestión de datos, como la estandarización de informes, la implementación de sistemas de seguimiento y la creación de paneles de control.
- Lleva a cabo un seguimiento continuo de la financiación de proyectos después de la implementación de las mejoras.
- Compara los datos de financiación antes y después de las mejoras utilizando análisis estadísticos para determinar si hay un aumento significativo.

Paso 2: Evaluación de la retención del talento científico.

- Realiza encuestas o entrevistas a científicos y personal de investigación para evaluar su satisfacción laboral y su percepción de la mejora en la gestión de datos.
- Recopila datos sobre la retención de científicos y personal de investigación en el centro antes y después de las mejoras en la gestión de datos.
- Analiza las respuestas de las encuestas y compara las tasas de retención antes y después de las mejoras.

Paso 3: Análisis de correlación.

- Realiza un análisis de correlación para determinar si existe una relación estadística significativa entre la implementación de mejoras en la gestión de datos, el aumento de la financiación de proyectos y la retención del talento científico.
- Utiliza herramientas estadísticas para evaluar la fuerza y la dirección de estas relaciones.

2.2. Conclusiones

Después de la realización de las encuestas online contestadas por profesionales del sector de la salud, las entrevistas personales efectuadas a trabajadores de la Fundación Científica de la AECC y una investigación de mercado, con el fin de conocer la situación actual y las oportunidades de negocio, podemos decir que:

- 1- Proporcionar una plataforma de seguimiento y transparencia a los inversores aumentaría la inversión y mejoraría la toma de decisiones.
- 2- Necesitan **crear informes de agregaciones** de gastos por tipo de proyecto, centros de investigación, investigadores, proveedores, etc. Con el fin de encontrar sinergias y tener un mayor control del proyecto.
- 3- Necesitan ser capaces de **identificar relaciones entre grupos de investigación (por proyectos, investigadores u otros atributos) de manera eficiente.**
- 4- **Análisis de gastos** en fungibles, salarios medios y tarifas de servicios internos.
- 5- Necesitan implementar metodologías y políticas relacionadas con la **calidad del dato.**
- 6- Necesitan **establecer métricas de calidad de los proyectos** a partir de los datos de seguimiento que nos proporcionan.

En definitiva, es necesaria en el mercado una solución de Business Intelligence, pensada tanto para los centros de investigación como los inversores, que aporte transparencia a los datos y permita tener todos los KPI necesarios para conocer en todo momento en qué situación se encuentran los proyectos. Con ello, mejoraría la toma de decisiones y las sinergias entre centros e investigadores.

2.3. Análisis de Competidores

Una vez identificadas las necesidades del mercado y los principales servicios que demanda y que Syneratech les podría ofrecer, procedemos a analizar los potenciales competidores que ya se encuentran proporcionando servicios o herramientas similares a los que quiere desarrollar nuestra empresa. La información recopilada se ha obtenido mediante prospección en su página web, perfiles profesionales en Linkedin, publicaciones en prensa digital y entrevistas personales con comerciales de las empresas.

2.3.1. Futuver



Descripción	
El Grupo Futuver se define como una Compañía Global de Servicios Tecnológicos y Consultoría de Gestión, establecida en 1997, que opera tanto en mercados nacionales como internacionales. Formada por expertos en Tecnología, la Consultoría estratégica y de Procesos, y Desarrollo de Negocio. Su enfoque se centra en la integración de metodologías y sistemas de información para llevar a cabo proyectos de Modernización Administrativa y gestión digital en las áreas de Personas, Procesos y Tecnología.	
Web	https://www.futuver.com/
Ámbito	Nacional e Internacional (Europa y Latinoamérica)
Sede	Gijón, Madrid (España); México, Panamá y Ecuador
Oferta	
<p>Principales servicios</p> <ul style="list-style-type: none"> • Servicios de desarrollo de tecnología • Ofrecen servicio de asesoramiento de la calidad del dato • Ofrecen servicio de ingeniería de procesos • TRAMITARE: Plataforma de Tramitación Electrónica de Expedientes que permite la gestión y el seguimiento de expedientes o trámites 100% de forma electrónica integrando en todos sus procesos la Firma Electrónica Avanzada y Certificado Digital. Es una plataforma altamente configurable y que permite la Gestión completa de todas las Ayudas emitidas por su entidad, trabajar en todas sus Fases, desde la emisión de la Convocatoria hasta su Cierre, durante todo su ciclo de vida, de una forma organizada, centralizada y sencilla. La plataforma se comporta como una base de datos SQL exportable a Excel. 	
<p>Formas de Comercialización</p> <ul style="list-style-type: none"> • Integración de la plataforma en los servidores del cliente. Esta opción requiere del pago de una licencia vitalicia inicial que cubre la integración. Adicionalmente, es necesario pagar una cuota de mantenimiento anual para obtener soporte por parte del equipo de Futuver para resolver incidencias o necesidades posteriores. • Software as a Service (SaaS). Pago de una licencia anual por usuario y la plataforma se aloja en los servidores del proveedor. En la cuota anual está incluido el mantenimiento. 	
<p>Fortalezas:</p> <ul style="list-style-type: none"> • Partners oficiales de Microsoft, lo que les proporciona soporte nativo en Azure • Han sido proveedores de una plataforma de gestión al Centro Nacional de Investigaciones Oncológicas (CNIO) • Han desarrollado la plataforma de gestión de convocatorias de la Sociedad Pública de Gestión Ambiental del Gobierno Vasco (Ihobe) 	
<p>Debilidades:</p> <ul style="list-style-type: none"> • Sus plataformas no tienen integración con PowerBI. • Esquema rígido de la plataforma que afecta a su flexibilidad y explotabilidad • Requiere contrato de mantenimiento para obtener soporte en la migración y exportación de datos entre herramientas. • No ofrecen asesoría en la estrategia de explotación del dato. • No integran servicios de inteligencia artificial 	
Tarifas	Comercialización como SaaS: Licencia anual 400€/usuario + coste primer año configuración, parametrización y puesta en marcha según necesidades.

2.3.2. Clarcat



Descripción	
Clarcat se define como una consultora tecnológica de confianza para el desarrollo de negocio con 15 años de experiencia. Especializada en la transformación digital, analítica avanzada e inteligencia artificial.	
Web	https://www.clarcat.com/
Ámbito	Nacional e Internacional (Europa)
Sede	Gijón, Madrid, Barcelona, A Coruña (España)
Oferta	
<p>Principales servicios</p> <ul style="list-style-type: none"> • Servicios de desarrollo de tecnología • Ofrecen servicio de asesoramiento de la calidad del dato • Ofrecen soluciones de analítica avanzada para todo tipo de empresas • Ofrecen servicios de gestión de clientes avanzado • Desarrollo de aplicaciones para todo tipo de negocios • Servicios de asesoramiento en la estrategia de explotación del dato 	
<p>Formas de Comercialización</p> <ul style="list-style-type: none"> • Tarifa variable según proyectos. Ofrecen al cliente comenzar con un proyecto pequeño para tener resultados rápidamente y demostrar su eficiencia y utilidad, generando confianza desde el primer momento. 	
<p>Fortalezas:</p> <ul style="list-style-type: none"> • Proyectos y servicios a medida del cliente. Gran diversidad de servicios. • Han trabajado con muchos clientes y proyectos de diversa naturaleza (Mahou, Generalitat de Catalunya, ABANCA, AirEuropa). • Trabajan ofreciendo modelo predictivos y prescriptivos de inteligencia artificial. • Trabajan con fuentes internas y externas de datos, según necesidad del cliente. • Capacidad de incluir imágenes y vídeos en las bases de datos y modelos. 	
<p>Debilidades:</p> <ul style="list-style-type: none"> • No cuentan con una plataforma o herramienta diseñada por ellos, ni ofrecen licencias para herramientas concretas. • Sus clientes no tienen ningún perfil asociado a la investigación en cáncer o en salud en general. 	
Tarifas	<p>Tarifa variable según proyectos.</p> <p>Ejemplos reales:</p> <ul style="list-style-type: none"> • Estudio de agregación para BBDD con formularios con los mismos campos - 23.000€ (3 meses de trabajo) • Estudio de agregación para BBDD incluyendo servicio de estandarización de datos - 39.000€ (5 meses de trabajo)

2.3.3. NorayBio



Descripción	
NorayBio se define como una empresa bioinformática, cuyo objetivo es dotar a las instituciones y empresas del ámbito de las biociencias, de herramientas de software que les ayuden a gestionar la gran cantidad de datos que derivan de sus investigaciones.	
En NorayBio diseñan, desarrollan e implementan software específico para el sector de la biotecnología, para facilitar al usuario el análisis y la interpretación de los datos obtenidos.	
Web	https://www.noraybio.com/
Ámbito	Nacional e Internacional (Europa, Lationamérica, Canadá, Australia)
Sede	Vizcaya (España)
Oferta	
<p>Principales servicios</p> <ul style="list-style-type: none"> • Ofrecen diferentes herramientas digitales para la gestión de muestras biológicas, datos clínicos y animales de un laboratorio (ANIBIO y NORAYBANKS). • Ofrecen herramienta para la gestión de proyectos de investigación y la gestión de documentación de Comités Éticos (NORAYDOCS). • En sus inicios ofrecían servicios de desarrollo de software personalizado, pero es una línea de negocio que han abandonado y ahora solo ofrecen licencias para sus herramientas. • Ofrecen servicios de formación, integración e instalación de sus herramientas en los procesos de los laboratorios. 	
<p>Formas de Comercialización</p> <ul style="list-style-type: none"> • Vitalicia: Pago único con acceso por tiempo ilimitado a la aplicación. • Suscripción: Acceso a la aplicación durante el tiempo contratado. • Contratos de Mantenimiento: Pago anual por soporte para resolución de incidencias. 	
<p>Fortalezas:</p> <ul style="list-style-type: none"> • Ofrecen servicios personalizados de adaptación e integración de su software a las necesidades del cliente dentro de un límite. • Sus principales clientes son grandes biobancos y plataformas como el Biobanco del CNIO o el CIBER-bbn. • Disponen de más de 100 clientes diferentes (muchos de ellos en Latinoamérica). • Varios años de experiencia (empresa fundada en 2002). • Garantizan trazabilidad de los procesos gestionados por sus herramientas. 	
<p>Debilidades:</p> <ul style="list-style-type: none"> • Rigidez de sus herramientas, restringida adaptabilidad. • Cobran el mantenimiento de sus licencias al margen para ofrecer soporte técnico. • No ofrecen servicios de consultoría ni asesoría en la gestión del dato más allá de las capacidades de sus herramientas. 	
Tarifas	<p>Licencia Vitalicia:</p> <ul style="list-style-type: none"> • 15.000€ Licencia vitalicia • 23.000€ Licencia vitalicia + Instalación + 1 año de soporte • 2.250€-3.000€ (15-20% del precio de la licencia vitalicia) precio del mantenimiento anual

2.3.4. Conclusiones

Tras el análisis de los potenciales competidores de Syneratech hemos observado que existe un nicho de mercado no cubierto que podríamos aprovechar y por lo tanto confiamos en la viabilidad de nuestra propuesta en base a las siguientes conclusiones:

1. Ningún competidor de mercado ofrece servicios especializados dirigidos tanto a inversores como investigadores.
2. Ningún competidor de mercado ofrece soluciones integrales de consultoría y desarrollo de herramientas dirigidos al sector sanitario de la investigación.
3. La mayoría de los competidores no disponen de experiencia en el mundo de las inversiones de capital riesgo o investigación sanitaria.
4. La mayoría de los competidores ofrecen licencias para herramientas propias demasiado rígidas y poco personalizables.
5. Las licencias de estos competidores resultan demasiado caras y requieren un pago anual elevado para recibir soporte o asistencia.
6. Aquellos que no ofrecen herramientas propias, no incluyen la licencia de las herramientas utilizadas en sus tarifas.
7. Los servicios de consultoría que ofrecen los competidores son muy vagos o extremadamente caros.
8. La mayoría de los competidores no ofrecen servicio de asesoría en la estrategia de explotación del dato.

3. ANÁLISIS Y DIAGNÓSTICO: DAFO

Una vez analizado el entorno y nuestros competidores llevamos a cabo el análisis DAFO. Este se realiza para proporcionar una visión integral de la situación actual de Syneratech en el mercado, ayudando a identificar qué áreas deben ser mejoradas (debilidades) y cómo se pueden aprovechar los factores externos positivos (oportunidades), así como prepararse para mitigar los posibles riesgos (amenazas) y capitalizar las ventajas internas (fortalezas). Toda la información detallada a continuación nos permite decidir qué modelo de negocio será el más eficaz y valido, para nuestra Start-up.



3.1. Debilidades

- Dificultad en encontrar las herramientas correctas, para explotar los datos debido a su diversidad y diferentes fuentes y formatos.
- Falta de estructura y criterios objetivos para valorar los proyectos.
- Ausencia o escasa política del gobierno del dato.
- Dificultad para crear KPI transversales y de referencia en el mercado, con el fin de poder comparar proyectos.
- Escasez de apoyo metodológico para elaborar propuestas y ejecutar proyectos.
- Carencia de Políticas de investigación definidas.
- Inconvenientes para establecer redes de colaboración.
- Sistemas de información no integrados ni orientados para facilitar la investigación.
- Falta de interés de la sociedad en conseguir desarrollar proyectos de investigación.
- Escaso mecenazgo y participación social en la investigación.
- Poca actividad investigadora debido a la precariedad salarial en este ámbito.
- Envejecimiento de las plantillas de investigadores.

3.2. Amenazas

- Falta de interés político en la investigación.
- Cambios políticos frecuentes con cambios de prioridades y objetivos.

- Pérdida de convenios con fuentes externas de financiación.
- Empresas existentes que valoricen proyectos en el sector salud y supongan una competencia directa.
- Poca difusión de avances en la gestión de proyectos de salud, ya que la falta de impulso mediático puede dejar morir soluciones tecnológicas de BI validas y útiles para la industria.
- Bajo grado de éxito en la obtención de financiación externa.
- Ausencia de Planes de Investigación dotados económicamente, para bordar las necesidades de equipamiento y personal.
- Falta de convocatorias anuales para el fomento de la investigación (proyectos de investigación, becas, contratos pre y postdoctorales, estancias en el extranjero, adquisición de infraestructura, etc.)
- Reducido número de empresas privadas colaborando en la investigación contra el cáncer.
- Burocratización y rigidez en la gestión de la investigación.
- Expectativas externas de obtención de resultados inmediatos de investigación e innovación.
- Conflictos de interés en los proyectos financiados por la industria.

3.3. Fortalezas

- Acceso a una gran cantidad de datos que con un correcto tratamiento pueden aportar información muy valiosa que ayude a la toma de decisiones.
- Conocimiento del sector y de sus necesidades actuales.
- Creación de iniciativas para el crecimiento de nuevas soluciones de IT.
- Capacidad de crear sinergias y obtener conclusiones necesarias para la valorización de proyectos.
- Experiencia en el desarrollo de recursos e informes útiles que faciliten la toma de decisiones en recursos clave.
- Servicio valido para extrapolarlo a otro sectores o áreas de salud.
- Crecimiento en soluciones técnicas que sustenten de forma objetiva y metódica las decisiones de inversión.
- Gracias a nuestra solución tecnológica, existen mayores posibilidades de apostar por proyectos de innovación.
- Proporcionar una visión clara y detallada de todos los aspectos claves de un proyecto de innovación.
- Actualización de una correcta política del dato, bajo los nuevos parámetros definidos.

3.4. Oportunidades

- Actualmente, debido a la situación vivida por le Covid-19, el sector salud está cogiendo protagonismo y por ello podría conseguir más inversión.
- Diferentes agentes como fundaciones, universidades, inversores privados o centros de investigación se convierten en posibles clientes de nuestro servicio, dado que les podemos ayudar en la toma de decisiones.
- Desarrollo de una solución tecnológica que permita la creación de los KPI necesarios para la comparación de los proyectos de investigación. Crear nuevas redes de colaboración con otras instituciones.
- Fomento del mecenazgo gracias a poder consolidar mediante datos la inversión.
- Acudir a Fondos europeos y FEDER (Interreg).

- Estimular la capacidad de contribución económica de la sociedad, mediante acciones de crowdfunding.
- Transición hacia sistemas de información más eficientes, accesibles y explotables.
- Desarrollo de la TICs en todas las áreas de salud.
- Existencia de recursos científico-tecnológicos suficientes para poder crecer y mejorar.
- Apoyo de la investigación orientada hacia las necesidades y áreas clave dentro del sector.
- Sinergias existentes entre distintos servicios y grupos de investigación.
- Existencia de equipos de investigación consolidados dinámicos y competitivos, que desarrollan proyectos interesantes para el sector.

4. PLAN ESTRATÉGICO

4.1. Modelo de negocio



4.1.1. Propuesta de valor

Nuestro modelo de negocio se basa en ofrecer soluciones de Business Intelligence dirigidas a organizaciones que gestionan o buscan valorizar proyectos de innovación en cualquier rama del sector salud. Nuestro valor diferencial radica en la integración de nuestra amplia experiencia en el ecosistema innovador y en la explotación de datos, lo que nos permite ayudar a nuestros clientes a:

- Maximizar el potencial de sus proyectos innovadores, optimizando su gestión financiera y promoviendo la colaboración estratégica.
- Alcanzar una visión clara y detallada de las oportunidades de inversión, así como la capacidad de identificar posibles colaboraciones y sinergias que impulsen el éxito de sus iniciativas.

Nuestra propuesta de valor generaría varios beneficios a nuestros clientes:

- Establecer relaciones entre proyectos, trabajadores y centros de trabajo, poder hilar y dar sentido a una gran cantidad de datos y como consecuencia optimizar procesos y facilitar la toma de decisiones.
- Crear procesos que faciliten el análisis y enfoque a diferentes áreas, definiendo KPIs y métricas automáticas.
- Realizar una Estrategia de Gobierno de dato generalista y protocolos para implantar una mínima política de calidad del dato y asegurar la veracidad del dato.



4.1.2. Segmento de clientes

Cualquier entidad que requiera realizar valorizaciones de proyectos innovadores de investigación son potenciales clientes. Los principales agentes que podrían estar interesados en estas soluciones o en implantar una solución de Business Intelligence de este tipo o la explotación de los datos contenidos en ellas serían:

- **Centros Públicos de Investigación** – El servicio sería útil para gestionar los activos y los proyectos de los diferentes centros de investigación, favoreciendo la colaboración y la optimización de recursos. Ej: Instituto de Salud Carlos III
- **Fundaciones** que invierten en investigación para diferentes enfermedades. (Fundación LaCaixa, Fundación Kaertor)
- **Consultoras de base Biotecnológica** – Las consultoras dispondrán de herramientas para valorizar las tecnologías, analizar tendencias en el mercado y en los esfuerzos de los investigadores.
- **Fondos de Inversión de Capital Riesgo** – El servicio sería útil para valorizar las tecnologías que se están desarrollando (Capital Cell, InverBio, INVIVO CAPITAL, INVERREADY, Columbus Venture)
- **Proveedores de suministros de laboratorio** – Conocerán mejor las necesidades de los laboratorios.
- **Universidades** – Se conocería con mayor detalle el perfil de trabajadores que requiere el ecosistema innovador para adaptar los estudios que ofrece.

- **Start-ups** - Pequeñas empresas que quieren valorizar sus proyectos emergentes podrían estar interesadas en comparar el valor de sus tecnologías con los contenidos en nuestra base de datos.
- **Empresas Farmacéuticas o biotecnológicas** - Grandes empresas pueden estar interesadas en valorizar sus proyectos emergentes comparándolos con los contenidos de nuestra base de datos. También podríamos ofrecerles implementar soluciones de Business Intelligence a su pipeline (conjunto de proyectos, productos o servicios) existente.

4.1.3. Relación con los clientes



Es una relación B2B, donde mediante canales bilaterales fomenten la comunicación fluida con el cliente ante cualquier duda y otros plenamente informativos. Por ejemplo, canales habituales y necesarios:

- Email y teléfono enfocados a dudas y problemas técnicos.
- Relación directa con los clientes, para abordar los proyectos de forma personalizada, tanto de forma presencial, como por teléfono, videoconferencia o email. Nuestra idea es vender un servicio al cliente, una solución global donde el equipo de consultoría gestiona la implementación de la solución tecnológica y realiza el seguimiento, para garantizar la correcta aplicación de las métricas.

Buscamos seguir una estrategia omnicanal, donde todos los canales estén conectados y el cliente sea el centro de todas nuestras acciones. Cualquier comunicación sea conocida en el momento por toda la organización, para trabajar de forma sincronizada y poder ofrecer un trato personalizado.

Queremos desarrollar proyectos a largo plazo, donde el cliente se sienta acompañado en todo el proyecto, guiado y por supuesto que gracias a la creación de Reporting pueda llegar a las conclusiones necesarias para la toma de decisiones.

4.1.4. Canales de distribución



Canales enfocados a clientes potenciales:

- Congresos de investigación, de innovación (Congreso TRANSIERE celebrado todos los años).
- Eventos de inversores.
- Organizar formaciones de valorización para visibilizar nuestros casos de éxito.
- Llamadas personales a gerentes de centros de investigación.

Canales para clientes:

Al inicio se considera que la mejor opción es que solo haya canales propios que garanticen el mejor servicio y den el mejor conocimiento y desarrollo. Según hemos comentado en la relación con los clientes, los principales canales serían:

- Teléfono
- Email
- Visita física y reuniones



4.1.5. Actividades clave

Nuestras principales actividades:

- Captación de clientes para mantener un crecimiento estable y poder mantener activo el proyecto, dando lugar a nuevas vías de desarrollo y explotación de los datos ya que pueden ser tratados para abordar diferentes campos o necesidades.
- Actualización constante de la tecnología implantada, para proveer al cliente siempre de los últimos desarrollos y así poder facilitar los análisis pertinentes y decisiones estratégicas. Abordar las soluciones que salen al mercado desde un punto de vista personalizado para cada cliente y enfoque.
- Estudio de las bases de datos y tratamiento adaptado a las necesidades. Así como planteamiento de nuestra propuesta adicional en aquello que consideremos que se puede mejorar. Buscamos crear una relación sostenida en el tiempo y no gestionar proyectos a corto plazo.
- Gestión personalizada del cliente para abordar e implementar la solución tecnológica en su empresa/proyecto.



4.1.6. Recursos clave

- Capital humano, equipo de consultores especializados en la actividad a desarrollar.
- Equipo informático, soporte de IT físico, para el desarrollo del trabajo, además de conexión de alta velocidad a internet y entornos de trabajo seguros en la nube.
- Herramientas específicas que se requieren:
 - Google Cloud Platform (GCP)
 - PostgreSQL
 - Python
 - Neo4J
 - API OpenAI (Text-Davinci-003)
 - PowerBI



4.1.7. Socios/Proveedores clave

Los principales asociados claves:

- Proveedor de software para la recepción del dato y posterior tratamiento. Todos los propietarios de las licencias de aquellos programas necesarios para la explotación, procesamiento y visualización del dato.
- Proveedor de los servidores en la nube, para el correcto almacenamiento del dato en las bases.
- Proveedor de hardware y productos tecnológicos asociados y necesarios para la actividad.

Con todos nuestros socios o proveedores buscamos crear relaciones comerciales que nos beneficien a la hora de renovar planes de pago, o actualización de productos, así como con los contratos de mantenimiento.

4.1.8. Fuente de ingresos



- Servicios de consultoría para implementar y personalizar la solución en las entidades interesadas.
- Contratos de mantenimiento y actualización del sistema.
- Las ventas de un SaaS (Software as a Service) por la comercialización y el uso del software.

4.1.9. Estructura de costes



- Coste del personal de ejecución del proyecto.
- Costes de desarrollo y adquisición de tecnologías y herramientas. (Pago de licencias)
- Costes operativos y de mantenimiento.
- Posibles costes asociados a la obtención de datos e implementación de políticas de calidad del dato.
- Costes en áreas como marketing y otros derivados de externalizar actividades ajenas a nuestro core como servicios de asesoría jurídica o laboral.

4.2. Misión, Visión y Objetivos

Misión

Proveer servicios de consultoría mediante herramientas de Business Intelligence al sector de la salud, para proporcionarles soluciones en la valorizar proyectos.

Visión

Ser la solución referente para la clasificación de proyectos en el sector salud, para la toma de decisión en la financiación de proyectos de investigación e innovación.

Objetivos de negocio

1. Crear una cartera de clientes, basada en una red de colaboración, mediante la asistencia a los principales congresos y ferias, para captar a nuestros clientes potenciales.
2. Fidelizar a nuestros clientes mediante acuerdos de colaboración a largo plazo y convertirlos en un servicio esencial en su gestión.
3. Conseguir un crecimiento sostenido a partir del 2º año.
4. Gestionar acuerdos de colaboración con las principales entidades del sector.
5. Control de los gastos estructurales.
6. Ofertar servicios de alto valor añadido.
7. Utilizar estándares de calidad en prestación de servicios como la ISO 9001 y gestión de proyectos como PMBOK o ITIL.

5. PLAN DE ACCIÓN

5.1. Alcance del proyecto

El alcance de este proyecto comprende el desarrollo de una herramienta de gestión y análisis de datos, que permita a cualquier empresa, y en este caso particular, a la AECC, sacar el máximo provecho a los datos de los que disponen, optimizando su gestión financiera, y detectando posibles oportunidades estratégicas.

Está herramienta debe permitir la centralización de los datos en una infraestructura escalable, y debe permitir el análisis y explotación de los datos para que los clientes sean capaces de extraer el valor de los mismos.

En el alcance del proyecto, se pretende cubrir una serie de objetivos que se describirán a continuación.

- Migración de los datos de la AECC desde su plataforma a una base de datos relacional
- Limpieza y standarización de datos
- Creación de dashboards en PowerBI
- Creación de un modelo no relacional en Neo4j
- Implementación de queries en lenguaje natural (NL) sobre la base de datos

5.2. Análisis de recursos

5.2.1. Recursos humanos

El equipo de Syneratech está formado por sus miembros fundadores. A medida que aumente la carga de trabajo se considerará la posibilidad de incorporar nuevo personal.

5.2.2. Estructura organizativa

El propósito de nuestra estructura organizativa es definir roles que los miembros deben desempeñar para colaborar de manera eficiente, facilitando así el logro de nuestros objetivos mediante una mejora y una colaboración continua. Se pretende sacar el máximo partido de los perfiles de los integrantes del equipo.

Se han diferenciado los siguientes roles:

- Líder del proyecto: Responsable de:
 - Comunicación con la AECC
 - Fijación y seguimiento de las reuniones
 - Fijar hitos para el desarrollo del proyecto
 - Asignar tareas al resto de miembros del equipo
 - Aportar el know-how del sector y del cliente
- Perfiles técnicos:
 - Perfiles financieros/estratégicos: Encargados de desarrollo del modelo de negocio, plan estratégico y aspectos financieros.
 - Perfiles técnicos: Encargados de la solución tecnológica: Extracción y limpieza de datos, despliegue de la infraestructura, creación del modelo de datos y explotación de los mismos.

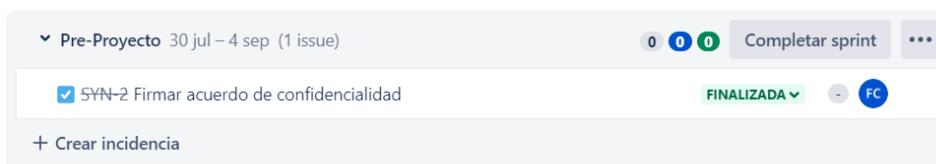
5.3. Mapa de procesos

En base a los perfiles definidos anteriormente, para la consecución de los objetivos del proyecto se han fijado diferentes hitos, desglosados en subtareas, cada una de las cuales se le ha asignado a un miembro u otro del equipo en función de la naturaleza de la tarea y el perfil requerido para ella.

Para la planificación de hitos y asignación de tareas se ha hecho uso de Jira. Por otro lado, para el seguimiento del trabajo realizado por cada miembro del equipo y el avance en el proceso, se han hecho reuniones semanales a través de Google Meet.

También se ha llevado a cabo un proceso de seguimiento de incidencias, que ha resultado fundamental para ir corrigiendo errores que se han ido descubriendo a lo largo de todo el proceso.

A continuación se muestran los diferentes hitos, y su desglose en tareas específicas asignadas a los miembros del equipo:



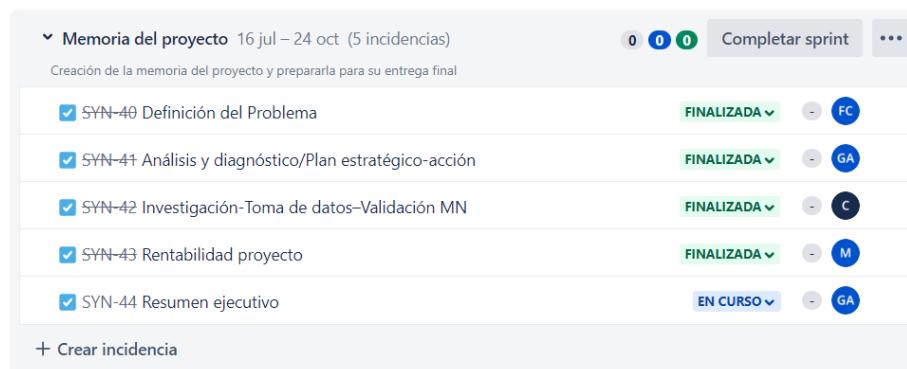
Pre-Proyecto 30 jul – 4 sep (1 issue)

0 0 0 Completar sprint ...

SYN-2 Firmar acuerdo de confidencialidad FINALIZADA ✓ - FC

+ Crear incidencia

Figura 5.1.Definición del hito “Pre-Proyecto”



Memoria del proyecto 16 jul – 24 oct (5 incidencias)

Creación de la memoria del proyecto y prepararla para su entrega final

0 0 0 Completar sprint ...

<input checked="" type="checkbox"/> SYN-40 Definición del Problema	FINALIZADA ✓ - FC
<input checked="" type="checkbox"/> SYN-41 Análisis y diagnóstico/Plan estratégico-acción	FINALIZADA ✓ - GA
<input checked="" type="checkbox"/> SYN-42 Investigación-Toma de datos–Validación MN	FINALIZADA ✓ - C
<input checked="" type="checkbox"/> SYN-43 Rentabilidad proyecto	FINALIZADA ✓ - M
<input checked="" type="checkbox"/> SYN-44 Resumen ejecutivo	EN CURSO ✓ - GA

+ Crear incidencia

Figura 5.2.Definición del hito “Memoria del proyecto”



Hito 0 -ETL 4 sep – 15 sep (5 incidencias)

0 0 0 Completar sprint ...

<input checked="" type="checkbox"/> SYN-11 Establecer los campos que se van a utilizar	FINALIZADA ✓ - FC
<input checked="" type="checkbox"/> SYN-12 Crear la estructura de la base de datos necesaria pa...	FINALIZADA ✓ - AT
<input checked="" type="checkbox"/> SYN-38 Crear un glosario de campos	FINALIZADA ✓ - FC
<input checked="" type="checkbox"/> SYN-14 Extracción de los datos de la plataforma de la AECC	FINALIZADA ✓ - FC
<input checked="" type="checkbox"/> SYN-13 Determinar los campos de la base de datos que ne...	FINALIZADA ✓ - RV

+ Crear incidencia

Figura 5.3.Definición del hito “ETL”

▼ Hito 1 - Despliegue 4 sep – 2 oct (5 incidencias)

	0 0 0	Completar sprint	...
<input checked="" type="checkbox"/> SYN-15 Análisis de diferentes soluciones de despliegue	FINALIZADA	- RV	
<input checked="" type="checkbox"/> SYN-17 Despliegue de la infraestructura	FINALIZADA	- AT	
<input checked="" type="checkbox"/> SYN-16 Carga de datos	FINALIZADA	- AT	
<input checked="" type="checkbox"/> SYN-18 Creación del modelo relacional	FINALIZADA	- RV	
<input checked="" type="checkbox"/> SYN-19 Limpieza y tratamiento de datos	FINALIZADA	- AT	

+ Crear incidencia

Figura 5.4.Definición del hito “Despliegue”

▼ Hito 2 - Explotación - PowerBI 2 oct – 20 oct (5 incidencias)

	0 0 0	Completar sprint	...
<input checked="" type="checkbox"/> SYN-20 Conexión de PowerBI con la BBDD	FINALIZADA	- P	
<input checked="" type="checkbox"/> SYN-21 ETL	FINALIZADA	- AT	
<input checked="" type="checkbox"/> SYN-22 Dashboard 1: Overview	FINALIZADA	- C	
<input checked="" type="checkbox"/> SYN-23 Dashboard 2: Financiero y territorial	EN CURSO	- GA	
<input checked="" type="checkbox"/> SYN-24 Dashboard 3: Análisis detallado por proyecto	EN CURSO	- FC	

+ Crear incidencia

Figura 5.5.Definición del hito “PowerBI”

▼ Hito 3 -Análisis Perfil- Neo4j 2 oct – 20 oct (2 incidencias)

	0 0 0	Completar sprint	...
<input checked="" type="checkbox"/> SYN-26 Creación modelo no relacional en Neo4j	FINALIZADA	- RV	
<input checked="" type="checkbox"/> SYN-25 Análisis no relacional de trabajadores-proyectos-organizaciones	EN CURSO	- AT	

+ Crear incidencia

Figura 5.6. Definición del hito “Análisis Perfil-Neo4j”

▼ Hito 4 -Implementar NL Queries 8 oct – 14 oct (3 incidencias)

	0 0 0	Completar sprint	...
<input checked="" type="checkbox"/> SYN-46 Investigación de la tecnología	FINALIZADA	- RV	
<input checked="" type="checkbox"/> SYN-47 Implementación de NL Queries	FINALIZADA	- AT	
<input checked="" type="checkbox"/> SYN-48 Pruebas	FINALIZADA	- RV	

+ Crear incidencia

Figura 5.7.Definición del hito “Implementar NL Queries”

▼ Hito 5 -Mejoras/Trabajo futuro 20 oct – 24 oct (3 incidencias)

0 1 1 Iniciar sprint ...

<input checked="" type="checkbox"/> SYN-29 Revisar la BBDD y la política de dato de la empresa	TAREAS POR HACER	- C
<input checked="" type="checkbox"/> SYN-10 Análisis de consecución de los hitos	TAREAS POR HACER	- C
<input checked="" type="checkbox"/> SYN-39 Posibles mejoras y trabajo futuro	TAREAS POR HACER	- GA

+ Crear incidencia

Figura 5.8. Definición del hito “Mejoras/Trabajo futuro”

5.4. Procesos clave

ANÁLISIS

- **Necesidades del cliente**

Resulta fundamental explorar las necesidades y puntos de mejora a los que el cliente se enfrenta antes de la implementación de una solución. Una comprensión detallada de las necesidades del cliente facilitará el desarrollo de una solución perfectamente adaptada a sus requerimientos. Las tareas a llevar a cabo en esta fase son las siguientes:

- Entender los procesos y forma de trabajo actual del cliente.
- Familiarizarse con las herramientas y sus funciones que utiliza el cliente.
- Identificación de los objetivos, misión, valores, etc. del cliente
- Investigar la influencia que puede tener la solución propuesta.
- Analizar soluciones similares en el mercado (competidores)

- **Compresión de los de datos y su finalidad**

Antes de proponerse cualquier solución, es fundamental entender de los datos de los que cliente dispone, y como este cliente utiliza los mismos. Esto es necesario para adaptar la solución al uso que el cliente le quiere dar.

También es importante ser consciente de la calidad de los datos antes de proponer la solución, ya que esta calidad podría condicionar/limitar su uso.

- **Análisis de técnicas**

Es importante conocer el estado del arte de la tecnología antes de proponer una solución, así como investigar las potenciales estrategias aplicables a los datos recolectados con el fin de obtener resultados óptimos y representativos. Un amplio conocimiento sobre las diferentes técnicas y modelos permitirá determinar qué solución se adapta mejor al caso concreto que se está trabajando y asegurar la confiabilidad en los resultados obtenidos.

DISEÑO

- **Selección de la infraestructura**

Conocidas las diferentes alternativas tecnológicas, y las necesidades del cliente (previsión de escalabilidad, requisitos de privacidad, frecuencia de actualización, criticidad de la disponibilidad de los datos, etc), se seleccionar la infraestructura óptima para cada caso.

- **Proceso ETL:**
Diseño de un proceso automático mediante el cual la herramienta es capaz de transformar los datos brutos a datos interpretables. Es un proceso costoso, en el que influye de manera muy significativa la calidad de los datos. En esta etapa encontramos diferentes fases:
 - Obtención de datos.
 - Transformación y carga de datos.
 - Aplicación de modelos.
 - Exploración y visualización.

○ **Modelo de datos**

Es de gran importancia tener en mente el uso que se le va a dar a los datos para diseñar un modelo relacional que permita obtener el máximo valor de los datos. En este punto es fundamental conocer, tanto los datos disponibles, como el uso que se quiere hacer de ellos, para establecer las conexiones necesarias.

DESARROLLO E IMPLEMENTACIÓN

- **Desarrollo de la solución técnica:** Esta etapa del plan de acción es de especial importancia, ya que implica la creación y puesta en marcha de la solución. Consiste en el despliegue de la infraestructura seleccionada, con el dimensionamiento adecuado para el rendimiento óptimo de la herramienta. Supone el arranque definitivo de la solución.
- **Implementación de los modelos elegidos:** Se procede con la configuración, ejecución, validación y evaluación de los modelos seleccionados.
- **Visualización:** Se introduce una herramienta de visualización diseñada para el cliente, donde se mostrarán de manera muy visual todos los datos que resultan de interés para el cliente, el cual ya no se tendrá que preocupar de interpretar los datos en bruto, si no que dispondrá de diferentes gráficas diseñadas para mostrar concretamente lo que el cliente ha manifestado que resulta de interés. El objetivo es proporcionar la máxima información de manera simple, legible e intuitiva.
- **Documentación técnica**
Esta documentación consiste en:
 - Guía de uso
 - Glosario de campos de la BBDD
 - Gestión de incidencias

VALIDACIÓN Y DESPLIGUE

- **Fase piloto:** Ejecución de diferentes pruebas en la herramienta, para comprobar su correcto funcionamiento y detectar posibles errores:
- **Soporte temprano:** Durante la primera versión de la herramienta, se habilitará un sistema de gestión de incidencias, de manera que puedan reportarse los diferentes errores y malfuncionamientos detectados en la herramienta piloto.

SOPORTE Y MANTENIMIENTO

- **Formación:** Se hará una demo de la herramienta al cliente, para enseñarle a este todas las funcionalidades, y guiarles en el proceso que deben seguir para sacar el máximo provecho a la herramienta. Además, se le proveerá con tutoriales y guías de funcionamiento, así como con cursos de formación bajo demanda.
- **Soporte:** Se proveerá de un servicio al cliente que tratará de resolver dudas técnicas del funcionamiento de la herramienta.
- **Mantenimiento:** se establecerá un mantenimiento periódico tanto de las infraestructuras como de la base de datos, para garantizar su correcto funcionamiento en todo momento.

5.5. Cronograma

A continuación, se muestra el cronograma del proyecto mediante un Gantt Chart:

SYNERATECH GANTT CHART

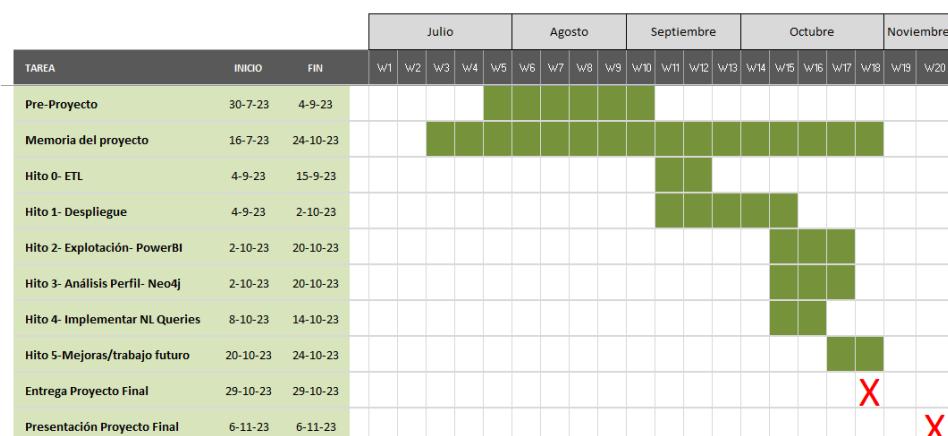


Figura 5.9.Cronograma

6. SOLUCIÓN TECNOLÓGICA

La solución tecnológica propuesta y desarrollada para nuestro proyecto es una combinación de diversas herramientas y tecnologías, seleccionadas para ofrecer una respuesta óptima y escalable a las necesidades planteadas.

6.1. Producto o Servicio Mínimo Viable

Como prueba de concepto de los servicios que nuestra empresa podría ofrecer, hemos decidido establecer una colaboración con la Fundación Científica de la Asociación Española Contra el Cáncer, partiendo de la entrega de un Producto o Servicios Mínimo Viable (PMV).

La Fundación Científica, a través de su labor de gestión de ayudas y financiación, **cuenta con muchos datos explotables sobre proyectos de investigación contra el cáncer que pueden ser muy valiosos para la Fundación y para otros agentes relacionados con la innovación en cáncer.**

La complejidad del cáncer, la diversidad de sus formas, los altos requerimientos técnicos y costos de investigación, así como las tasas de éxito reducidas, dificultan la inversión privada. Esto resulta en insuficiente investigación e innovación en cáncer. La Fundación Científica de la AECC necesita optimizar la explotación de sus bases de datos, que están en una plataforma comercial bastante rígida. Esto los lleva a depender de herramientas comunes de ofimática como Microsoft Excel para realizar análisis cotidianos.

Un mejor análisis de datos conduciría a mejorar la eficiencia en la gestión, seguimiento, evaluación y valorización de los proyectos, identificando tendencias y sinergias, mejorando su clasificación, gestionando mejor sus recursos, reduciendo la incertidumbre, incrementando tasas de éxito y, definitivamente, mejorando la toma de decisiones en general y aportando valor añadido a los datos.

6.2. Definición de la solución

La solución tecnológica propuesta se basa en la implementación de un sistema integrado que permite la recolección, procesamiento, almacenamiento y visualización de datos.

En colaboración con la AECC, el objetivo del proyecto es ofrecer un análisis detallado y una visión clara de la información relacionada con proyectos de investigación contra el cáncer, centros de investigación e investigadores. Para ello, se ha optado por una arquitectura modular que facilita la escalabilidad y adaptabilidad a futuros requerimientos.

6.3. Base de datos en la nube

Los despliegues alojados en la nube eliminan la necesidad de adquirir y mantener un hardware, mientras que siguen ofreciendo un cierto grado de control sobre la base de datos. Esta opción es para aquellos que buscan un término medio entre los servicios en local y los servicios gestionados en la nube. Se describen a continuación sus principales características:

- Permite la escalabilidad sin necesidad de infraestructura adicional.
- Ofrece la posibilidad de utilizar licencias de bases de datos existentes en muchos casos.
- Se integra mejor con los sistemas locales que las bases de datos gestionadas en la nube.

- Ofrece una optimización limitada ya que los datos no se almacenan en hardware propio.
- Proporciona un acceso y control más limitado sobre el ancho de banda que los despliegues en local.

6.4. Máquina Virtual

Una máquina virtual es un entorno de computación virtualizado que se ejecuta dentro de un servidor físico y emula un sistema de hardware completo, lo que incluye una CPU, memoria, almacenamiento y sistema operativo.

- Aislamiento y Control: Las máquinas virtuales proporcionan un alto grado de aislamiento, ya que cada VM funciona como una entidad independiente. Esto significa que puedes instalar y ejecutar una base de datos en una VM sin preocuparte por las interferencias con otros servicios o aplicaciones en el mismo servidor físico. También tienes un control completo sobre la configuración de la VM y del sistema operativo que ejecuta, lo que permite personalizar la configuración para las necesidades de la base de datos.
- Compatibilidad: Las VMs son versátiles y pueden alojar una variedad de sistemas de gestión de bases de datos, como MySQL, PostgreSQL, SQL Server, Oracle, entre otros. Esto permite implementar bases de datos relacionales o NoSQL según tus necesidades.
- Escalabilidad Vertical: Si necesitas aumentar los recursos de tu base de datos, puedes hacerlo mediante la escalabilidad vertical. Esto implica agregar más recursos de hardware (como CPU o memoria) a la VM existente para mejorar el rendimiento de la base de datos. Sin embargo, la escalabilidad vertical tiene límites físicos en términos de cuántos recursos puede proporcionar un servidor físico.
- Aislamiento de Cargas de Trabajo: Si tienes diferentes bases de datos o aplicaciones que requieren diferentes configuraciones o versiones de software, puedes alojar cada una en su propia VM para evitar conflictos y problemas de compatibilidad.

Sin embargo, es importante tener en cuenta que las máquinas virtuales también tienen algunas desventajas, como el consumo de recursos significativos, el tiempo de arranque y el tamaño de las imágenes de VM. Además, la gestión de múltiples VM puede ser más compleja en comparación con enfoques más modernos como el uso de contenedores o servicios de bases de datos administrados en la nube.

6.5. Tecnologías, herramientas y recursos utilizados

A lo largo de este apartado se van a exponer aquellas herramientas implicadas en el desarrollo de este proyecto, explicando sus principales ventajas.

6.5.1. Bases de datos

Por un lado, encontramos **PostgreSQL**, un sistema de gestión de bases de datos relacional orientado a objetos. Es compatible con diversos modelos de datos, es muy robusto, potente y escalable.

Dos características clave de PostgreSQL es que posee data types (tipos de datos) avanzados y permite ejecutar optimizaciones de rendimiento avanzadas, que son características que por lo general solo se ven en sistemas de bases de datos comerciales, como por ejemplo SQL Server de Microsoft u Oracle.

Sus principales ventajas:

- Gratis.
- Código abierto: Existe una gran comunidad de desarrolladores que respaldan y mejoran continuamente el código, gracias al cual se ha convertido en uno de los mejores gestores de bases de datos a nivel mundial.
- Multiplataforma: Esto quiere decir que se trata de un software que puede correr bajo distintos entornos y sistemas operativos, y es compatible con muchos de los servidores web más populares como Apache, Nginx y LiteSpeed.
- Facilidad de uso: Una de las razones de la popularidad de esta herramienta es su fácil administración a través de paneles con PgAdmin. Esta interfaz gráfica amigable permite a los administradores y desarrolladores gestionar bases de datos, tablas, consultas SQL, y otras funciones de manera sencilla, lo que ha llevado a que PgAdmin sea ampliamente utilizado.
- PostgreSQL tiene una gran capacidad para manejar grandes volúmenes de datos. Las bases de datos de gran tamaño pueden hacer pleno uso del MVCC (Multi-Version Concurrency Control), un método de control que nos permite realizar tareas de escritura y lectura simultáneamente, resultando en un gran rendimiento.
- Soporte total de ACID: ACID (atomicity, consistency, isolation y durability) hace referencia a las características de atomicidad, consistencia, aislamiento y durabilidad de las transacciones que se realizan en una base de datos, lo que garantiza que, en caso de producirse un fallo durante una transacción, los datos no se perderán ni terminarán donde no deban.

6.5.2. Google Colab

Esta herramienta permite ejecutar notebooks de Jupyter en la nube de forma gratuita y colaborativa en tiempo real, donde usamos scripts de Python para el desarrollo de tareas ETL (Extraction, Transformation, Load).

6.5.3. Librerías Python

En esta sección, destacaremos las bibliotecas clave:

- **Pandas**

Ha sido esencial en nuestro proyecto para gestionar y analizar los datos de forma eficiente. Principalmente utilizada para el análisis de datos y la manipulación de

estructuras de datos, como DataFrames. Proporciona una amplia gama de funciones y métodos para cargar, limpiar y transformar datos, lo que la hace esencial en el ámbito de la ciencia de datos y la ingeniería de datos.

- **GraphDatabase (Neo4j)**
Se emplea principalmente para trabajar con bases de datos de grafos, como Neo4j. Permite crear, consultar y manipular datos almacenados en un formato de grafo.
- **Psycopg2**
Biblioteca de Python que facilita la conexión y manipulación de bases de datos PostgreSQL en aplicaciones Python, permitiendo consultas y actualizaciones de datos de manera eficiente.
- **Openai**
La librería "openai" en Python permite la interacción con modelos de lenguaje de OpenAI, como GPT-3.5, para tareas de procesamiento de lenguaje natural. Esta librería facilita la generación de texto, traducción de idiomas, respuesta a preguntas y muchas otras tareas de procesamiento de lenguaje natural. Los desarrolladores pueden integrar fácilmente los modelos de OpenAI en sus aplicaciones y proyectos, aprovechando las capacidades avanzadas de IA para mejorar la automatización y la inteligencia en el procesamiento de texto.

Concretamente se ha empleado el modelo "text-davinci-003" para la traducción de lenguaje natural a consultas SQL.

- **TKINTER**
Librería gráfica de Python utilizada para crear interfaces de usuario (GUI) de forma sencilla y eficaz. En nuestro caso, se emplea para proporcionar una interfaz de usuario sencilla para interactuar con la base de datos mediante procesamiento de lenguaje natural gracias a la librería de OpenAI.

6.5.4. Power BI

Herramienta de visualización de datos desarrollada por Microsoft que permite crear informes interactivos y paneles de control a partir de diversas fuentes de datos. Es ampliamente utilizado para analizar y presentar datos de manera visual y efectiva.

6.5.5. Neo4j

Es una base de datos de grafo de código abierto especializada en el eficiente almacenamiento y consulta de datos relacionales. Utiliza una estructura de grafo para representar y consultar datos, lo que la hace especialmente adecuada para modelar y analizar relaciones complejas entre entidades.

6.5.6. GitHub

Es una plataforma en línea que se utiliza para el desarrollo colaborativo de proyectos de software. Permite a los desarrolladores gestionar el control de versiones de su código, realizar seguimiento de problemas, y facilita la colaboración en equipos distribuidos.

En nuestro caso, al emplear herramientas cloud para el desarrollo, su principal función ha sido alojar la versión final del código. Se incluye el [enlace al repositorio](#) utilizado. Por motivos de privacidad y protección de datos, es necesario solicitar

acceso al repositorio. Por favor, ponerse en contacto con cualquier miembro del equipo antes de proceder.

6.5.7. Gestión y Comunicación

- **Jira**

Plataforma de gestión de proyectos y seguimiento de problemas utilizada para planificar, rastrear y gestionar el trabajo en equipos de desarrollo de software. Proporciona herramientas para la gestión ágil de proyectos, seguimiento de problemas, creación de hojas de ruta y colaboración en equipos.

- **Meet**

Herramienta de videoconferencia desarrollada por Google. Permite realizar reuniones en línea, videoconferencias y colaboración en tiempo real con otras personas, ya sea en el trabajo o en situaciones personales. Meet es ampliamente utilizado para comunicarse y colaborar a distancia.

6.5.8. Google Cloud Platform (GCP)

Se trata de una plataforma de servicios en la nube proporcionada por Google. Ofrece una amplia gama de servicios, incluidos servicios de computación, almacenamiento, bases de datos, aprendizaje automático, análisis y más.

Para nuestro proyecto, nos hemos apoyado en la infraestructura y la tecnología de Google para levantar una instancia de máquina virtual (VM) en la cual instalamos Docker y creamos una imagen de PostgreSQL donde persistir los datos e interactuar de manera deslocalizada.

Son varios los motivos que han hecho inclinarnos por esta decisión:

- Mayor privacidad de los datos (ofrece conectividad SSL y reglas de firewall para garantizar la seguridad).
- Mayor accesibilidad: Accesibilidad a los datos por todos los miembros del equipo (alojar la BBDD en un equipo local de uno de los miembros del equipo suponía que el resto de los miembros del equipo debían acceder en remoto a dicho ordenador). Al estar disponible en la nube, se puede acceder desde cualquier lugar si se dispone de una conexión a internet.
- Accesibilidad simultánea.
- No requiere inversiones en hardware: a diferencia de una máquina física, no requiere componentes físicos y evita gastos de mantenimiento.
- Incluye la posibilidad de pago por uso: De esta manera, se ahorrarían costes al no tener que invertir en un equipo que puede que no se utilice periódicamente.
- Aporta flexibilidad y escalabilidad: una máquina virtual permite que se pueda aumentar o reducir la potencia de su infraestructura de manera rápida y sencilla según las necesidades del usuario. En el caso de utilizar alojamiento local, puede ocurrir que el almacenamiento en disco se llene, que el usuario necesite un equipo más potente para lidiar con pico de

volumen de trabajo o que el servidor se quede pequeño. En estas ocasiones, una máquina virtual en la nube es la opción requerida.

- Ofrece mayor seguridad: la rápida programación de los backups facilita la realización de las copias de seguridad de las máquinas virtuales.
- Backups y actualizaciones automáticas.

En la actualidad, hay 3 empresas que dominan el mercado del Cloud Computing, y son Amazon Web Services, Microsoft Azure y Google Cloud Platform, y a pesar de que todas estas comparten las ventajas clave, GCP destaca por su equilibrio entre precios competitivos y rendimiento escalable, lo que lo convierte en nuestra elección preferida para este proyecto.

6.5.9. Docker

Docker es una plataforma de contenedores que simplifica la creación, distribución y ejecución de aplicaciones en entornos aislados y portátiles. Permite empaquetar software en unidades estandarizadas llamadas contenedores, que incluyen todo lo necesario para desplegar el software en cualquier entorno (bibliotecas, herramientas de sistema, código y tiempo de ejecución).

Con Docker, puedes crear y gestionar estos compartimentos de manera sencilla, permitiendo que varias aplicaciones comparten los recursos de la computadora sin problemas.

A diferencia de las máquinas virtuales, los contenedores comparten el sistema operativo principal del servidor en lugar de crear virtualizaciones individuales, lo que optimiza la eficiencia de recursos.

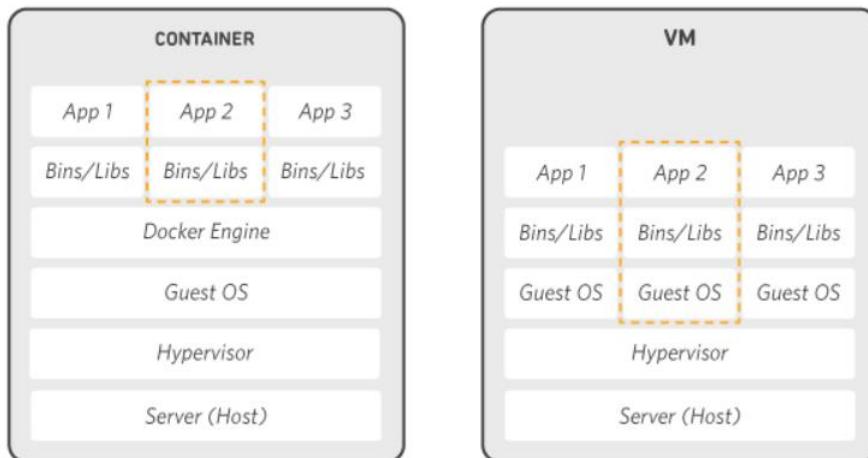


Figura 6.1. Diferencias entre un contenedor y una VM

Los contenedores ofrecen varias ventajas clave. Permiten la ejecución de aplicaciones en diversos entornos sin complicaciones, garantizan un aislamiento seguro, facilitan la escalabilidad al desplegar múltiples instancias, agilizan la gestión con herramientas como Docker, posibilitan implementaciones rápidas y permiten el versionamiento y control de cambios, lo que simplifica la administración de bases de datos y aplicaciones relacionadas.

Aunque Docker es una herramienta muy popular para la creación y gestión de contenedores, también presenta ciertos inconvenientes de seguridad:

- Aislamiento insuficiente: Los contenedores de Docker comparten el mismo kernel del sistema operativo subyacente, lo que puede resultar en un aislamiento insuficiente entre contenedores y el sistema operativo anfitrión. Un contenedor comprometido podría potencialmente afectar a otros contenedores o al sistema anfitrión.
- Configuración incorrecta: Una configuración incorrecta del entorno Docker puede dejar expuestas vulnerabilidades de seguridad. Esto incluye la configuración de permisos inadecuada, la exposición de puertos innecesarios, y el uso de imágenes de contenedor obsoletas o no seguras.
- Imágenes no seguras: Las imágenes de Docker pueden contener software obsoleto o vulnerabilidades de seguridad no parcheadas. Es crucial utilizar imágenes seguras y mantenerlas actualizadas para mitigar estos riesgos.

En lugar de Docker, algunas organizaciones optan por combinar **Jenkins** y **Kubernetes** para gestionar sus pipelines de CI/CD (Integración Continua y Despliegue Continuo) y orquestar sus contenedores respectivamente, lo que puede ofrecer una mayor seguridad y control.

6.6. Análisis de las fuentes de datos

En primer lugar, es necesario abordar previamente las carencias y limitaciones de los datos que maneja la Fundación Científica de la AECC y que podrían ayudar a solventar las necesidades del sector de la investigación e innovación contra el cáncer.

La **Fundación Científica** gestiona todo lo relacionados con sus actividades a través de la **plataforma SmartSimple**, donde está contenida toda la información de sus bases de datos. **SmartSimple Grants Management** es una plataforma web en la que confían numerosas entidades para gestionar y controlar sus procesos. Muchas de estas entidades se dedican a la adjudicación de subvenciones o financiación.

SmartSimple es un **CRM avanzado** que ofrece una **interfaz sencilla y adaptable para el usuario** y que permite la fácil recolección y almacenamiento de datos (15). Su principal desventaja reside en que **sacrifica la libre explotación de los datos** que contiene en favor de ofrecer una interfaz sencilla y segura para el usuario.

La **Base de Datos de la Fundación Científica de la AECC** se comporta como una base de datos NoSQL. Los distintos niveles de cada aplicación o UTA dentro de la BBDD contienen expedientes. Estos expedientes (o documentos) contienen diferentes campos a voluntad del usuario. Estos campos son compartidos por las diferentes UTAs y pueden acceder a ellos libremente para construir expedientes personalizados.

SmartSimple permite definir **expedientes** (o vistas) que pueden ser **exportados en forma de tablas** (Excel, csv, etc.), de esta forma extrajimos los datos más relevantes de la base de datos de la Fundación Científica para crear relaciones entre las diferentes tablas, **dando lugar a una base de datos relacional más explotable**.

El proceso ETL (Extracción, Transformación y Carga) ha sido fundamental para garantizar que estos datos se integren de manera coherente y estructurada en la base de datos. Para ello, se ha utilizado Google Collab, que permite la manipulación y procesamiento de datos en un entorno cloud con lenguaje de programación Python.

Los principales problemas de la gestión de proyectos de la Fundación Científica de la AECC son comunes a los de los centros públicos de investigación descritos anteriormente:

- **Bases de datos limitadas y rígidas:** Hacen uso de CRM de gestión de proyectos simples donde se recogen títulos de proyectos, trabajadores implicados y presupuestos.
- **Información que no puede ser explotada:** La información está almacenada en memorias anuales o facturas contenidas en documentos PDF.
- **Dificultad para optimizar recursos:** No se realizan análisis de agregación de gastos entre los diferentes proyectos y convocatorias que financian.
- No disponen de mecanismos ni protocolos establecidos para identificar o determinar la calidad de los datos que recogen.
- **Dificultad para identificar sinergias:** Debido a la complejidad de los proyectos y a que su clasificación no está indexada, es difícil establecer colaboraciones entre investigadores. Ellos mismos se buscan en función de relaciones personales o afinidades particulares.
- **Dificultad para extraer cualquier estadística de su base de datos:** Sus datos no están indexados ni tienen una estructura definida, lo que dificulta la explotación de sus datos, requiriendo tediosos tratamientos para obtener valor de los datos.
- **Dificultad para valorar los proyectos:** Debido a la diversidad de los proyectos y a la especificidad de las tecnologías, es difícil identificar los proyectos más prometedores y con mayor proyección. La clasificación y valoración de los proyectos se realiza uno a uno y de forma manual. Existen criterios de clasificación y valorización, pero no están unificados.

6.7. Modelo Lógico

Hemos seguido una estrategia de almacenamiento propia para los datos de la Fundación, ofreciendo dos posibles modelos lógicos, el relacional y no relacional.

6.7.1. Modelo lógico relacional

El modelo lógico se ha diseñado teniendo en cuenta la naturaleza de los datos y las relaciones entre ellos. Se ha utilizado PostgreSQL como sistema de gestión de bases de datos, debido a su robustez y capacidad para manejar grandes volúmenes de información. A través del framework PgAdmin, se ha administrado y gestionado la base de datos, garantizando su integridad y coherencia.

Paralelamente, se ha procesado una porción de datos para desplegar en la plataforma cloud propia de Neo4j (Neo4j Aura), una base de datos no relacional basada en grafos que permitirá realizar análisis de relaciones de los datos objetivo.

A continuación, se muestra el diagrama de entidad relación que describe cómo se relacionan y organizan los datos en la base de datos PostgreSQL (para ver esta imagen en más detalle, ir a ANEXO II: MODELO LÓGICO RELACIONAL):

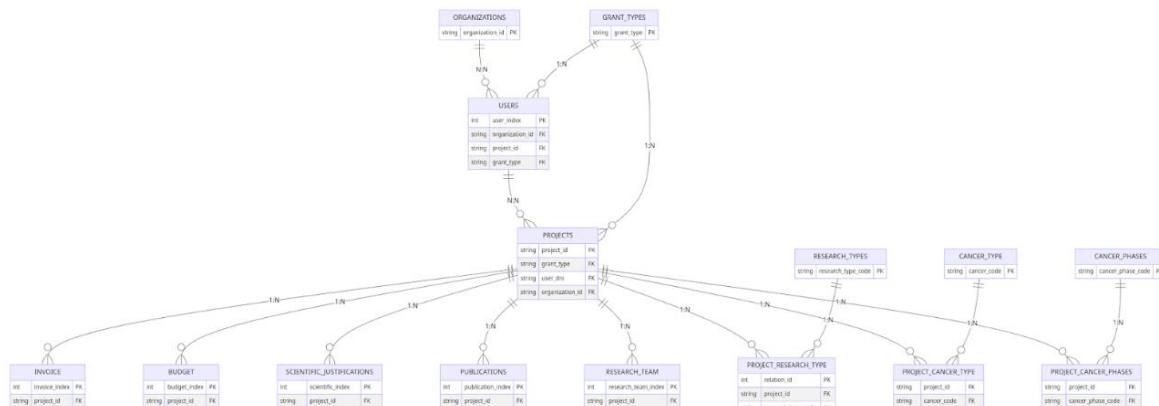


Figura 6.2. Esquema Relacional en PostgreSQL

6.7.2. Modelo lógico no relacional

Paralelamente, se procesó una muestra del total de datos para desplegar en la plataforma cloud propia de Neo4j (Neo4j Aura), una base de datos no relacional basada en grafos que permitirá realizar análisis de relaciones de los datos objetivo.

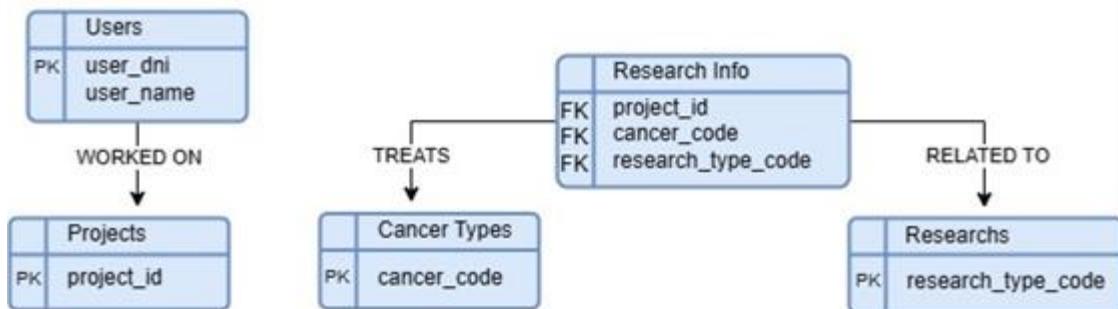


Figura 6.3. Esquema No Relacional en Neo4j

6.8. Arquitectura

Para este proyecto hemos optado por una base de datos no supervisada **PostgreSQL**, como una imagen en Docker y desplegada en una **Máquina Virtual (VM)** instanciada en **Google Cloud Platform (GCP)**.

Esta arquitectura garantiza un entorno controlado y reproducible. Además, nos ha permitido partir de un despliegue modesto en términos de coste, y evaluar su funcionamiento antes de comprometerse con una solución más costosa, como puede ser una base de datos supervisada en AWS o Azure.

El proceso de ETL se realiza mediante Google Colab desde el almacenamiento temporal en Onedrive (cliente) y Google Drive (Syneratech), para cargar los datos limpios a

PostgreSQL en la VM por un lado y la base de datos no relacional Neo4J Aura por otro.

Las distintas herramientas de visualización (PowerBI) y consulta (NLP-SQL) atacan directamente a la máquina virtual.

6.9. Solución propuesta

Nuestro proyecto aborda una solución técnica centrada en la gestión y visualización de datos mediante la infraestructura de Google Cloud Platform (GCP) y herramientas como Docker, Power BI y Neo4j. A continuación, se detallan los aspectos más relevantes.

6.9.1. Capa de servicio

La capa de servicio está compuesta por la infraestructura que soporta la solución, incluyendo la máquina virtual en GCP y Docker, que proporciona un entorno estandarizado para el despliegue de aplicaciones.

La VM utilizada posee características específicas que la hacen adecuada para este despliegue.

Instancia de vm serie **e2-micro**:
vcpu: **2 vuds**
memory: **1gb**
sistema operativo: **debian**
version: **debian gnu/linux 11 (bullseye)**
boot disk type: **standart persistent disk**
size: **20gb**

Por otro lado, como se mencionó anteriormente, Docker, proporciona un entorno de desarrollo controlado y reproducible. Durante el proceso de implementación, se utilizaron comandos de Linux para crear y administrar contenedores Docker en nuestra instancia de MV.

A continuación, describimos los pasos más relevantes y comandos Linux utilizados para configurar el acceso a la MV desde cualquier IP, instalar Docker, instalar PgAdmin y crear imagen PostgreSQL:

- En el ambiente GCP, se creó una **regla de firewall** que permite acceso a la imagen PostgreSQL en el puerto 5432.
- **Conexión al bash** VM-GCP mediante ssh key (tecnología gcp) con regla de firewall establecida para conexión de cualquier ip.
- **Conexión VM-PgAdmin:**
Una nueva ip de VM se genera con cada arranque. Esa ip se debe introducir en propiedades de conexión de PgAdmin en el host e introducir la contraseña establecida.

- **Comandos linux** necesarios instalar Docker en MV, instalar PgAdmin y crear imagen Docker de PostgreSQL :

- Actualización del Servidor

```
~$ sudo apt-get update
```

- Instalación de Paquetes Necesarios. Estos paquetes incluyen "ca-certificates" para certificados, "curl" para transferir datos, "gnupg" para cifrar y "lsb-release" para información del sistema.

```
~$ sudo apt-get - install ca-certificates curl gnupg lsb-release
```

- Creación de un Directorio para Llaves GPG de APT:

```
~$ sudo mkdir -p /etc/apt/keyrings
curl -fsSL https://download.docker.com/linux/debian/gpg | 
sudo gpg --dearmor -o /etc/apt/keyrings/docker.gpg
```

- Descarga y Configuración de la Llave GPG para un Repositorio Docker:

```
~$ curl -fsSL https://download.docker.com/linux/debian/gpg | 
sudo gpg --dearmor -o /etc/apt/keyrings/docker.gpg
```

- Añadir un Repositorio de Docker al Sistema de Gestión de Paquetes:

```
~$ echo deb [arch=$(dpkg --print-architecture) signed-by=/etc/apt/keyrings/docker.gpg]
https://download.docker.com/linux/debian ${lsb_release -cs} stable | sudo tee /etc/apt/sources.list.d/docker.list > /dev/null
```

- Actualizar Lista de Paquetes Disponibles:

```
~$ sudo apt-get update
```

- Instalación de Paquetes Relacionados con Docker:

```
~$ sudo apt-get install docker-ce docker-ce-cli containerd.io
docker-compose-plugin
```

- Verificación de la Instalación de Docker:

```
~$ sudo docker run hello-world
```

- Descargar Imagen Oficial de Docker para PostgreSQL desde Docker Hub:

```
~$ sudo docker pull postgres
```

- Iniciar un Nuevo Contenedor PostgreSQL:

```
sudo docker run --name some-postgres -e
POSTGRES_PASSWORD=BESTPASSWORDEVER -v
POSTGRES:/var/lib/postgresql/data -p 5432:5432 -d postgres
```

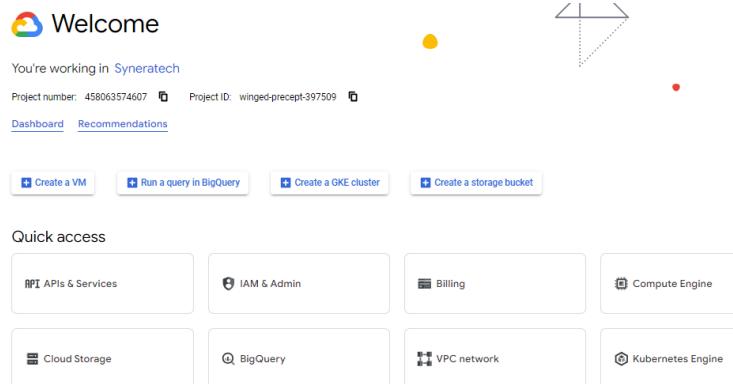


Figura 6.4. Panel del workspace para el proyecto Syneratech en Google Cloud Platform

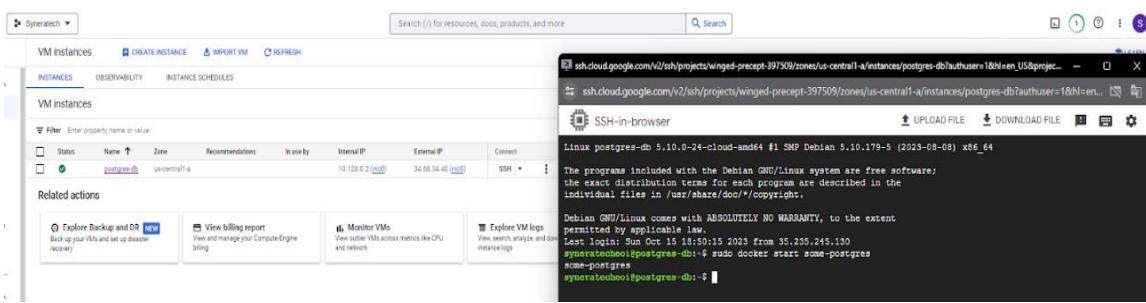


Figura 6.5. Instancia de la máquina virtual y consola bash arrancando imagen docker de postgresSQL en GCP

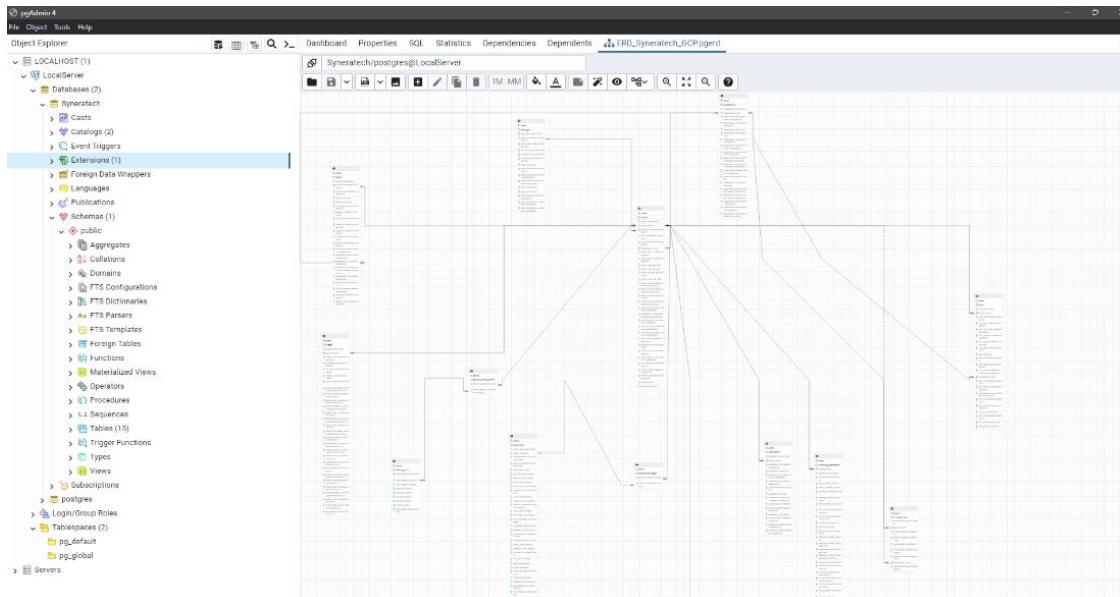


Figura 6.6. Interfaz PgAdmin

Siendo conscientes de que este proyecto tiene un alcance académico y propósito de MVP (Minimum Viable Product), algunas tecnologías utilizadas como Docker pueden no ser muy recomendables en entornos de producción por aspectos de seguridad.

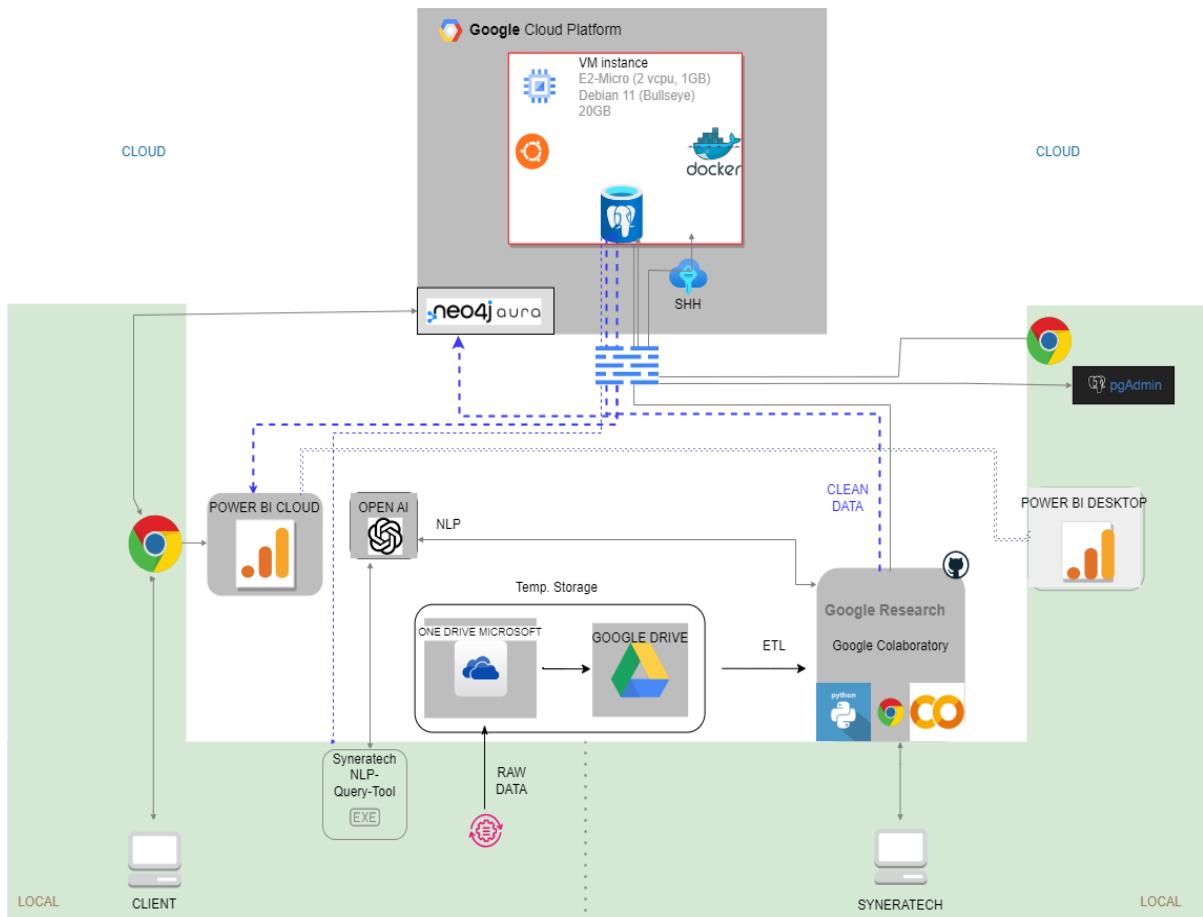


Figura 6.7. Arquitectura del proyecto

6.9.2. Capa de aplicación

La visualización de datos es un componente clave para la interpretación y comprensión de la información procesada. Se planea integrar herramientas como Power BI para presentar los datos de manera gráfica e interactiva, permitiendo así una mejor toma de decisiones basada en evidencia, en conjunción con visualizaciones de grafos obtenidas de Neo4J.



Figura 6.8. Interfaz de uso de Neo4j

6.9.3. Visualización de datos

La visualización de datos es un componente clave para la interpretación y comprensión de la información procesada. Se planea integrar herramientas como Power BI para presentar los datos de manera gráfica e interactiva, permitiendo así una mejor toma de decisiones basada en evidencia, en conjunción con visualizaciones de grafos obtenidas de Neo4j.

A. PowerBI

La adecuada visualización de un conjunto de datos permite identificar tendencias, desviaciones o situaciones clave importantes que permiten mejorar la toma de decisiones y tener mayor control sobre los factores que interfieren en la actividad de cualquier entidad.

Se proporcionó a la Fundación Científica de la AECC un **dashboard avanzado** elaborado con PowerBI que contenía una amplia gama de visualizaciones basadas en el conjunto de datos contenidos en su base de datos y que Syneratech volcó en Google Cloud Platform (GCP) como se ha descrito previamente. El dashboard se configuró de forma que estuviera conectada a la base de datos en la nube de GCP y estuviera en todo momento actualizada. Adicionalmente, el dashboard fue diseñado de forma que la Fundación pudiera configurarlo para que se conectaría a su base en Smart Simple a través de una conexión que sigue el protocolo Open Data Protocol (OData). OData se trata de un **protocolo de transferencia de datos accesible por RESTful API**, lo que permite **intercambiar datos a través de la red**.

Por otro lado, el dashboard avanzado fue diseñado con una interfaz sencilla y ágil para facilitar la consulta y visualización de los datos, mejorando notablemente la experiencia de usuario respecto al acceso a la información contenida en la base de datos.

A continuación, se muestra brevemente las diferentes secciones del dashboard:

1- Pantalla de Inicio con acceso directo a cada una de las secciones:



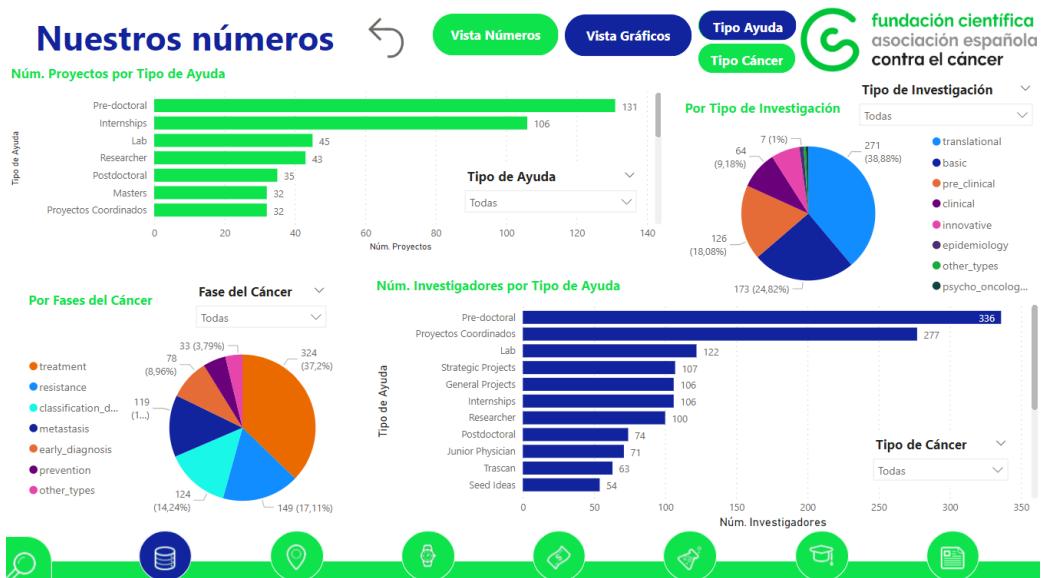
The screenshot shows the main interface of a PowerBI dashboard. At the top left is the Syneratech logo. Below it is the AECC logo. A search bar labeled "Búsqueda" is located at the bottom left. To the right is a vertical column of seven green rounded rectangular cards, each representing a different data section: "Nuestros Números", "Mapa de Proyectos", "Tendencia Temporal", "Análisis de Gastos", "Análisis de Proveedores", "Impacto Social", and "Detalles de un Proyecto". Each card has a small icon next to its title.

2- Dashboard interactivo que muestra de forma agregada todos los proyectos activos de la Fundación, número de centros financiados, número de investigadores implicados en los proyectos, financiación destinada, número de convocatorias y número de solicitudes desestimadas. También se aprecia que

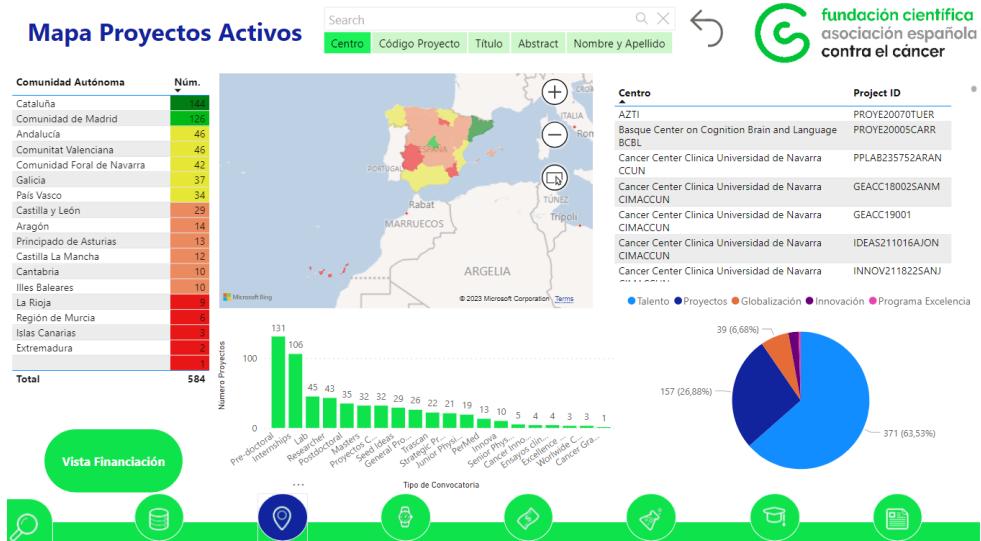
cada página del informe contiene una **barra de navegación inferior** que permite saltar rápidamente a cada una de las secciones. El logo institucional de la Fundación de la esquina superior derecha permite **volver a la pantalla de inicio** en cualquier momento.



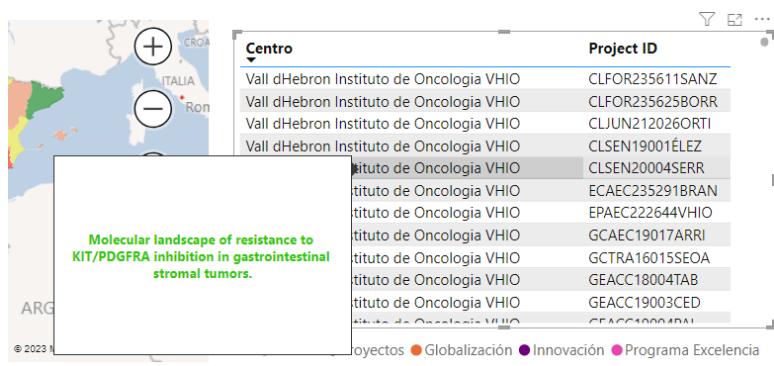
El dashboard contiene una variante de los números mostrados en **forma de gráficos con varios segmentadores** que permite ver aquellas convocatorias, aquellos tipos de investigación o tipos de cáncer donde se destinan más ayudas. Se accede a todas estas versiones a través de botones interactivos.



- 3- **Mapa territorial** para conocer la distribución de las ayudas e identificar los territorios más competitivos a nivel de investigación en España:



Dispone de una vista alternativa (Vista Financiera) que, en lugar de mostrar número de ayudas activas, muestra el dinero destinado a cada territorio. La tabla de la derecha que indica los Centros de Investigación y los códigos de los proyectos dispone de una ventana personalizada para mostrar información sobre el objeto que indica el título del proyecto:



- 4- **Tendencia temporal** de diferentes métricas que permite ver la evolución con el de tiempo del número de solicitudes recibidas por la Fundación, el número de convocatorias que abre cada año, el número de investigadores que trabajan en los proyectos financiados, la financiación asignada a cada año y cada tipo de convocatoria, y el número de publicaciones científicas publicadas cada año relacionadas con los proyectos que financian.



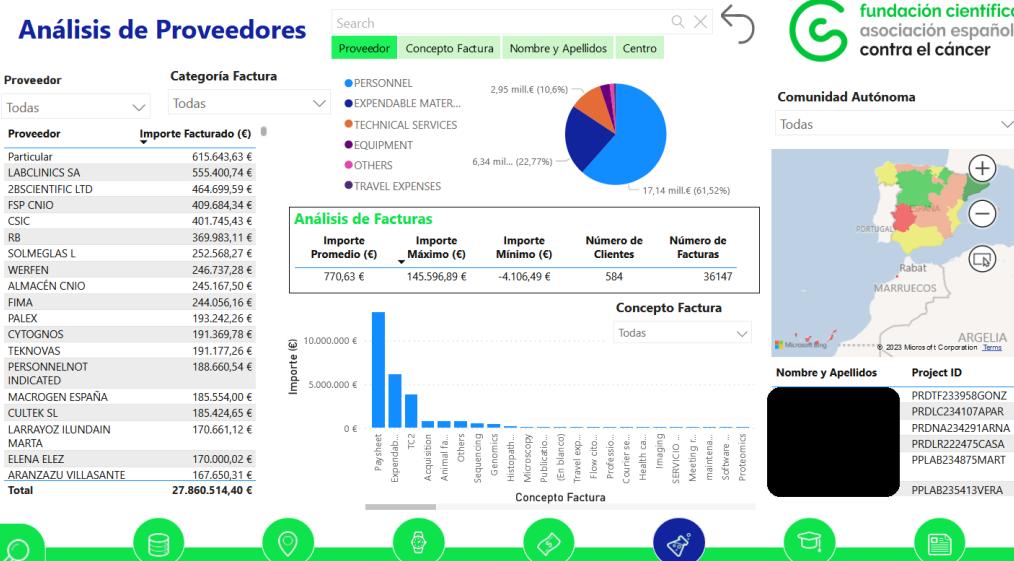
5- **Análisis de Gastos**, que permite ver qué tipos de ayudas, tipos de cáncer, fases del cáncer, tipos de investigación o proveedores reciben más dinero proveniente de las ayudas de la Fundación.



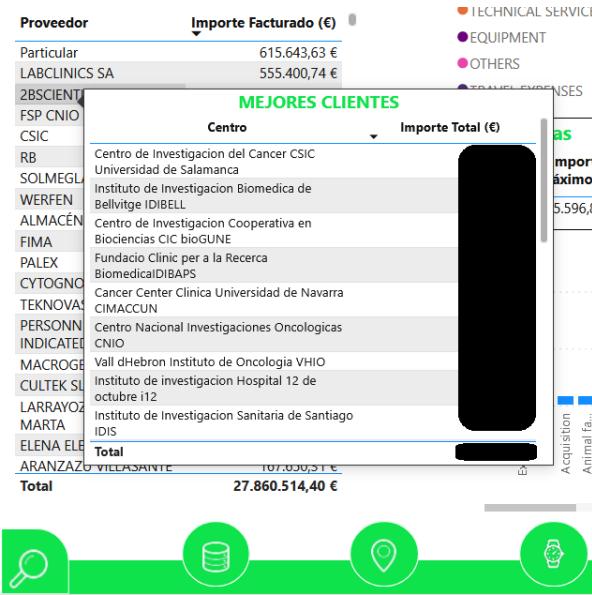
También permite la visualización por categoría de gasto de las facturas o por conceptos de la propia factura:



- 6- **Análisis de Proveedores**, que permite buscar y visualizar los diferentes proveedores que proporcionan productos o servicios a los proyectos que financia la Fundación. También permite segmentar por categoría o concepto de factura a los diferentes proveedores. Permite analizar las facturas de un proveedor a través de la visualización del importe promedio, máximo o mínimo de factura, así como su número de clientes y facturas que emite a los proyectos financiados por la Fundación. Por otro lado, permite visualizar en un mapa qué territorios trabajan con un proveedor en concreto.



También se ha incluido una ventana que muestre los principales clientes de un proveedor cuando te desplazas con el cursor por encima de su nombre:



- 7- **Impacto Social**, que permite visualizar los diferentes números relacionados con el impacto en la sociedad de la investigación financiada por la Fundación, como por ejemplo el impacto en pacientes, la atracción de fondos adicionales al proyecto, impacto en el número de contrataciones o el impacto en el número de publicaciones científicas.



- 8- **Búsqueda Avanzada**. El dashboard avanzado incluye una página de búsqueda avanzada que permite usar múltiples segmentadores que tienen efecto inmediato en el resto de páginas. También permite la búsqueda de texto en el contenido del proyecto para buscar conceptos claves, personas o centros específicos. La página de búsqueda tiene utilidad inmediata en la búsqueda de proyectos concretos para visualizarlos en la página de "Detalles del Proyecto".

Búsqueda por Texto:

Search

Código Proyecto	Título	Abstract	Centro	Nombre y Apellido
Resultados Búsqueda Proyectos: 434				
<input type="checkbox"/> AIO14142199GARC <input type="checkbox"/> AIO1512774GOME <input type="checkbox"/> AIO1512776GARC <input type="checkbox"/> AIO1512858MONT <input type="checkbox"/> AIO1512929FRR <input type="checkbox"/> AIO16163609MART <input type="checkbox"/> AIO183816GAVI				
<input type="checkbox"/> Detection of markers in infiltrate... <input type="checkbox"/> A transdisciplinary approach to L... <input type="checkbox"/> Advanced Therapies Accelerator ... <input type="checkbox"/> AECC-IDIBAPS Excellence Progra... <input type="checkbox"/> Early detection and intervention ... <input type="checkbox"/> FEASIBLE (FederAted Synthetic ... <input type="checkbox"/> Genomic instability as the basis f...				
<input type="checkbox"/> (En blanco) <input type="checkbox"/> - Introduction: Chemotherapy-induced ...				
<input type="checkbox"/> - Resumen del proyecto: Debido a la sit...				
<input type="checkbox"/> #6461623; Background: Despite the larg...				
<input type="checkbox"/> (CAR)-T cell-based therapies have achie...				
<input type="checkbox"/> 1 in 4 people will die of cancer and the ...				
<input type="checkbox"/> 15-20% of human cancers are associate...				
<input type="checkbox"/> A diagnosis of cancer is one of the mos...				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> - Achucarro Basque Center for Neuro...				
<input type="checkbox"/> Agencia Aragonesa para la Investigaci...				
<input type="checkbox"/> Agrupación Centros de Investigaci...				
<input type="checkbox"/> AKRIBEA 6174 SL				
<input type="checkbox"/> ALTUM Sequencing				
<input type="checkbox"/> Amadix				
<input type="checkbox"/> Resultados Búsqueda Abstract: 3489				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> - Introduction: Chemotherapy-induced ...				
<input type="checkbox"/> - Resumen del proyecto: Debido a la sit...				
<input type="checkbox"/> #6461623; Background: Despite the larg...				
<input type="checkbox"/> (CAR)-T cell-based therapies have achie...				
<input type="checkbox"/> 1 in 4 people will die of cancer and the ...				
<input type="checkbox"/> 15-20% of human cancers are associate...				
<input type="checkbox"/> A diagnosis of cancer is one of the mos...				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> - Achucarro Basque Center for Neuro...				
<input type="checkbox"/> Agencia Aragonesa para la Investigaci...				
<input type="checkbox"/> Agrupación Centros de Investigaci...				
<input type="checkbox"/> AKRIBEA 6174 SL				
<input type="checkbox"/> ALTUM Sequencing				
<input type="checkbox"/> Amadix				
<input type="checkbox"/> Resultados Búsqueda Centros: 434				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> - Introduction: Chemotherapy-induced ...				
<input type="checkbox"/> - Resumen del proyecto: Debido a la sit...				
<input type="checkbox"/> #6461623; Background: Despite the larg...				
<input type="checkbox"/> (CAR)-T cell-based therapies have achie...				
<input type="checkbox"/> 1 in 4 people will die of cancer and the ...				
<input type="checkbox"/> 15-20% of human cancers are associate...				
<input type="checkbox"/> A diagnosis of cancer is one of the mos...				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> - Achucarro Basque Center for Neuro...				
<input type="checkbox"/> Agencia Aragonesa para la Investigaci...				
<input type="checkbox"/> Agrupación Centros de Investigaci...				
<input type="checkbox"/> AKRIBEA 6174 SL				
<input type="checkbox"/> ALTUM Sequencing				
<input type="checkbox"/> Amadix				
<input type="checkbox"/> Resultados Búsqueda Personas: 3489				
<input type="checkbox"/> (En blanco)				
<input type="checkbox"/> PERSONAS				

Filtros:

- Tipo de Ayuda: (En blanco), Acknowledgment, Cancer Grand Challenges, Cancer Innova
- Tipo de Convocatoria: (En blanco), Globalización, Innovación, Programa Excelencia
- Estado del Proyecto: Approved, Approved - Pending Grant, Approved - Pre-Award, Closed
- Fase del Proyecto: (En blanco), classification_diagn..., early_diagnosis, metastasis, other_types
- Fase del Proyecto: (En blanco), Ajudicado, Finalizado, Prórroga
- Tipo de Investigación: basic, clinical, epidemiology, innovative, other_types, pre_clinical, psycho_oncological, translational
- Género: (En blanco), Female, Male, No answer
- Comunidad Autónoma: (En blanco), Andalucía, Aragón, Cantabria, Castilla La Mancha, Castilla y León, Cataluña, Comunidad de Madrid, Comunidad Foral de Navarra
- Año Convocatoria: (En blanco), 2014.0, 2015.0, 2016.0

Icons: Search, Database, Location, Watch, Wallet, Alert, Graduation, Document.

9- **Detalles del Proyecto.** Este último dashboard permite acceder rápidamente a toda la información contenida en la base de datos correspondiente a un único proyecto. Dispone de varias vistas alternativas que permite conocer el resumen del proyecto, el equipo de trabajo, el número de publicaciones y más métricas interesantes para evaluar el progreso y los resultados del proyecto.

Search

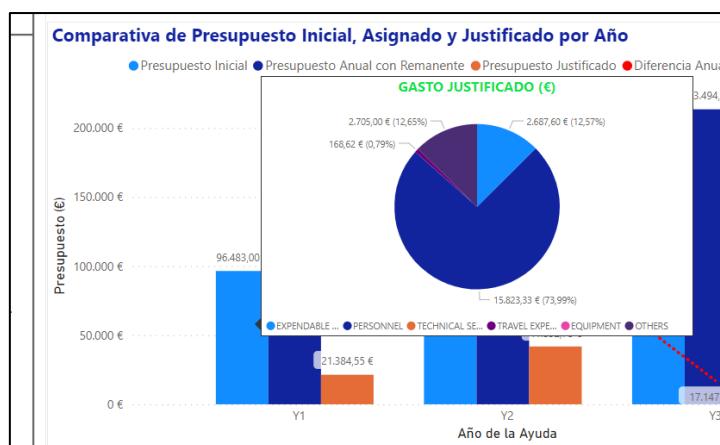
Código Proyecto	Título	Abstract	Centro
Resultados Bús			
<input type="checkbox"/> AIO14142199GARC <input type="checkbox"/> AIO1512774GOME <input type="checkbox"/> AIO1512776GARC <input type="checkbox"/> AIO1512858MONT			
Convocatoria: Proyectos AECC 2018 Tipo Convocatoria: Proyectos Fase del Proyecto: Prórroga Estado del Proyecto: Closed Inicio Proyecto: 11/01/2018 Fin Proyecto: 31/10/2022 Duración Proyecto (años): 3 Tipo Investigación: colon_cancer Tipos de Cáncer: classification_dia... Fases del Cáncer: metastasis treatment			
Project ID: [REDACTED] Centro: Vall d'Hebron Institut de Oncologia VHIR Info Económica Abstract Despite the major advances in the treatment of metastatic colorectal cancer (mCRC), the survival rate remains very poor. Reliable prognostic biomarkers are			
Puntuación AEI: 88 Investigador Responsable: [REDACTED] Equipo de Investigación: 13 Nombre Apellido Rol [REDACTED]			
Publicaciones: (En blanco) Título Revista Factor Impacto			
Proyecto Ejecutado (%): 80 Personas Contratadas: (En blanco) Impacto en Pacientes: (En blanco)			

Icons: Search, Database, Location, Watch, Wallet, Alert, Graduation, Document.

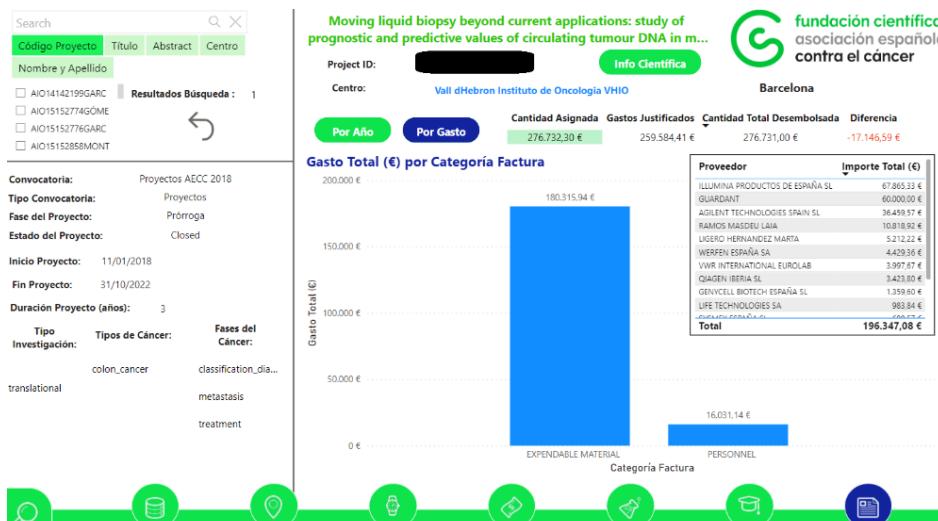
Dispone también de una vista alternativa que permite visualizar y controlar los gastos en función del presupuesto inicial y los gastos justificados en los diferentes años de la ayuda, lo que permite identificar rápidamente desviaciones o deficiencias en el seguimiento.



Esta vista económica también cuenta con una ventana personalizada que muestra información por conceptos de gasto una vez que te desplazas por encima de la gráfica:



La vista económica también muestra una variante que permite observar los gastos de los proyectos clasificados por diferentes categorías, y a la vez mostrar sus principales proveedores en una tabla.



En definitiva, hemos analizado el contenido de la base de datos de la Fundación y hemos elaborado un dashboard avanzado con una interfaz sencilla e intuitiva que permite visualizar rápida y fácilmente los principales aspectos y detalles de la actividad investigadora financiada por la Fundación, mostrando métricas claves para el seguimiento de los proyectos y que pueden ser entendidas y transmitidas fácilmente a un amplio público, incluyendo inversores y otros agentes del ecosistema investigador.

B. Neo4j

- **Ejemplo 1**

Objetivo:

Representar una red de profesionales, donde se muestra su grado de cercanía para establecer contacto con un usuario específico a través de su participación en proyectos.

Queries:

Crea relaciones en la base de datos Neo4j entre nodos de usuario (Users), proyectos (Projects), y una tabla intermedia (Projects Info).

```
MATCH (p:Projects),(f:Projects Info), (u:Users)
WHERE p.project_id = f.project_id AND f.user_dni = u.user_dni
CREATE (u)-[:WORKED_ON]->(p)
```

Consulta que recupera las conexiones entre usuarios y proyectos que se han establecido mediante la primera query. La query ha sido limitada a 35 nodos para poder apreciar mejor el ejemplo.

```
MATCH p=()-[:WORKED_ON]->() RETURN p LIMIT 35;
```

Resultado:

En este ejemplo, los nodos naranjas representan a los usuarios, y están vinculados a los nodos azules, que corresponden a los proyectos en los que han participado.

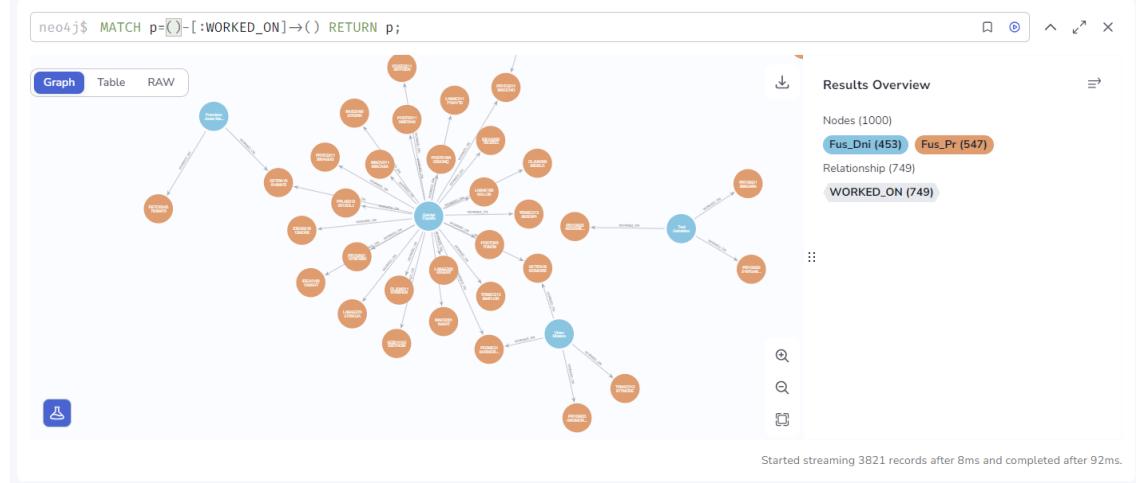


Figura 6.9.Ejemplo 1 de consulta en Neo4j

- **Ejemplo 2**

Objetivo:

El objetivo es proporcionar una visión general de las relaciones entre la entidad de investigación, los tipos de cáncer y los tipos de investigación.

Queries:

Esta consulta tiene como objetivo crear relaciones entre nodos de tipo *Research Info* y nodos de tipo *Cancer Types* basándose en la coincidencia de un código de cáncer entre ellos.

```
MATCH (p:Research_Info)
MATCH (c:Cancer_Types {cancer_code: p.cancer_code})
CREATE (p)-[:TREATS]->(c)
```

Esta consulta busca relacionar nodos de tipo *Research Info* con nodos de tipo *Research Types* si hay coincidencia en los códigos de tipo de investigación.

```
MATCH (r:Research_Types), (p:Research_Info)
WHERE ANY(item IN r.research_type_code WHERE item IN
p.research_type_code)
MERGE (p)-[:RELATED_TO]->(r)
```

Esta consulta busca crear dos vistas de patrones. La primera vista (p1) relaciona nodos de tipo *Research Info* con nodos de tipo *Cancer Types* a través de la relación *TREATS*. La segunda vista (p2) verifica si existe una relación “*RELATED TO*” con nodos de tipo *Research Types*. La consulta está limitado a 10

```
MATCH p1=(p:Research_Info)-[:TREATS]->(c:Cancer_Types)
OPTIONAL MATCH p2=(p)-[:RELATED_TO]->(r:Research_Types)
RETURN p1, p2 LIMIT 10;
```

Resultado:

En este ejemplo, el nodo naranja representa un tipo de cáncer. Observamos que está circundado por nodos verdes que simbolizan proyectos de investigación exclusivamente relacionados con ese cáncer en particular. A su vez, estos proyectos se conectan mediante la relación “*RELATED TO*” con nodos azules que representan los diferentes tipos de investigación que aborda cada proyecto, tales como “basic,” “clinical,” “epidemiology,” “innovative,” y otros.

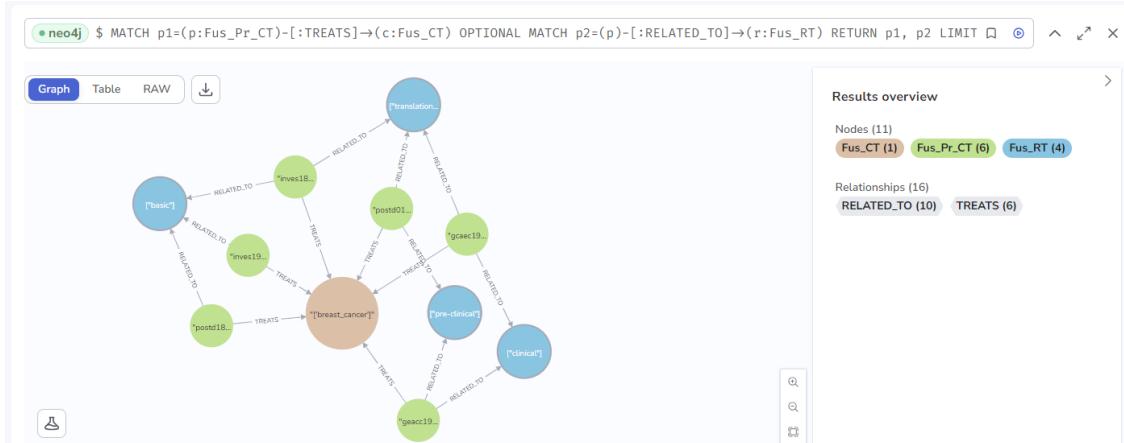


Figura 6.10.Ejemplo 2.1 de consulta en Neo4j

Al ajustar los colores y tamaños de los nodos, optimizando sus posiciones y eliminando las restricciones de límites, es posible lograr resultados similares a los que se muestran a continuación:

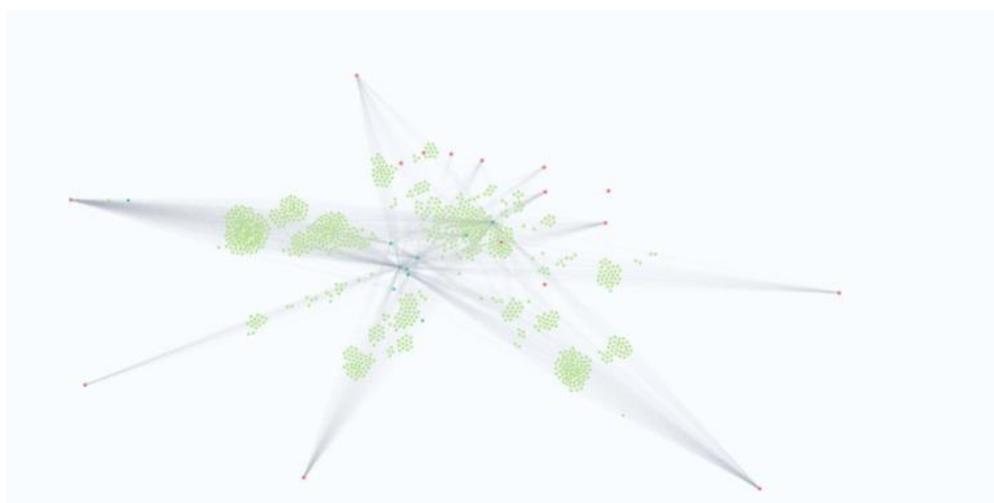


Figura 6.11.Ejemplo 2.2 de consulta en Neo4j

C. Procesamiento de Lenguaje Natural a través de Interfaz Gráfica de Usuario (GUI)

- **Ejemplo 1**

A continuación, se muestra un ejemplo sencillo de interacción mediante botones, en el que la consulta ya está predefinida. Esto se ha diseñado para consultas frecuentes y para ofrecer una interacción más eficiente con el usuario.

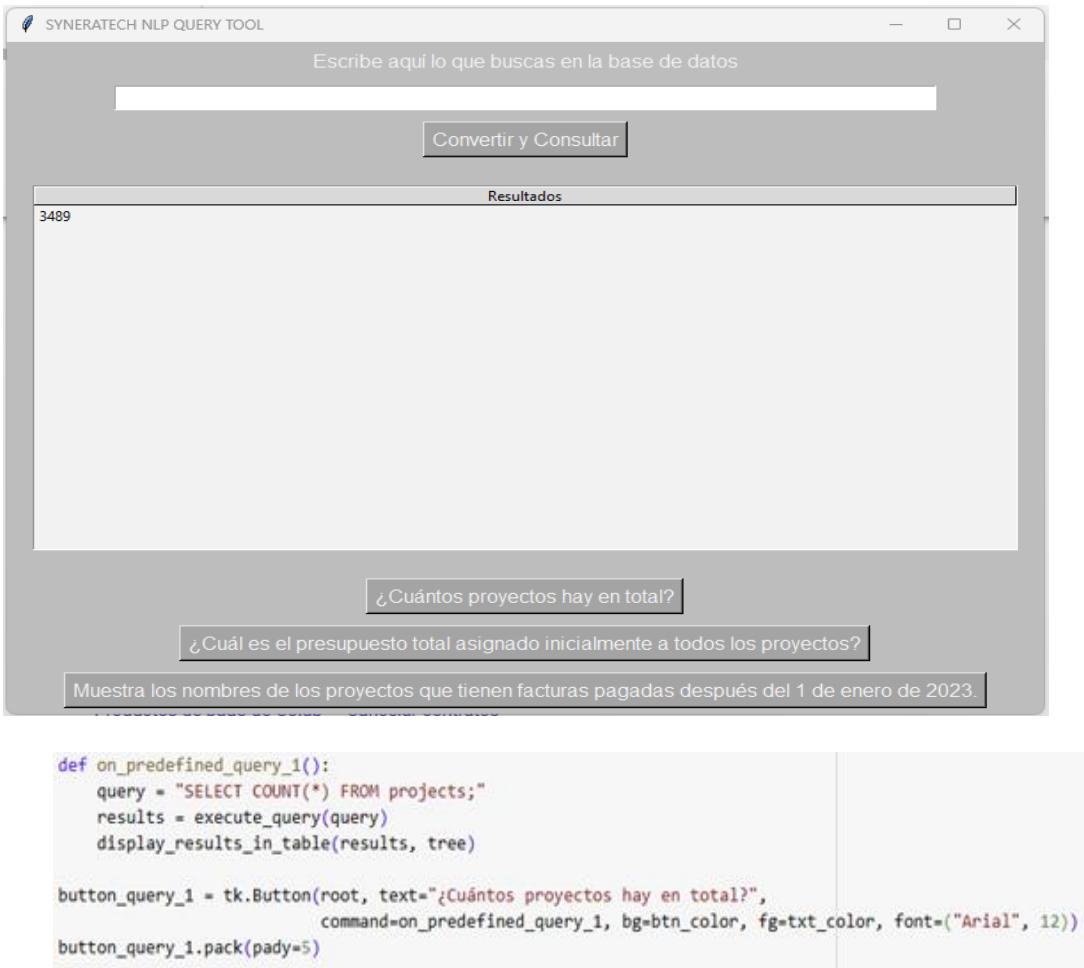


Figura 6.12.Ejemplo 1. Arriba interfaz gráfica de usuario mostrando el resultado de haber pulsado primer botón. Abajo código python definiendo la query hardocodada para dicho botón.

- **Ejemplo 2**

Se muestra un ejemplo más complejo con interpretación del lenguaje natural. El usuario puede expresar sus necesidades o consultas en texto, y al presionar el botón, se utiliza el Procesamiento de Lenguaje Natural (NLP) para traducir ese texto en una consulta SQL compleja y eficiente. Esto es aplicable tanto en español como en inglés.

SYNERATECH NLP QUERY TOOL

Escribe aquí lo que buscas en la base de datos

Which organization has the most researchers?

Convertir y Consultar

Resultados

FUNDACIO PRIVADA INSTITUT DE INVESTIGACIO ONCOLÓGICA VALL D'HEBRON - VHIO

¿Cuántos proyectos hay en total?

¿Cuál es el presupuesto total asignado inicialmente a todos los proyectos?

Muestra los nombres de los proyectos que tienen facturas pagadas después del 1 de enero de 2023.

```
*** Generated SQL Query: SELECT organization_legal_name, COUNT(*) total_researchers
FROM organizations o
    INNER JOIN users u ON o.organization_id = u.organization_id
    INNER JOIN research_team rt ON u.user_dni = rt.user_dni
GROUP BY organization_legal_name
ORDER BY total_researchers DESC
LIMIT 1 ;
```

Figura 6.13. Ejemplo 2. Arriba interfaz de usuario mostrando el texto con la solicitud deseada a la base de datos. Esta solicitud, será traducida a SQL y la query recibida, será enviada contra la base de datos. Abajo se muestra el output obtenido de traducir el texto a una consulta SQL.

Este es el resultado generado al aplicar el modelo de Procesamiento de Lenguaje Natural (NLP) para interpretar la consulta previa.

El propósito de incluir esta salida en el código es permitirnos analizar la consulta que se está ejecutando en la base de datos, verificar su precisión y comprender su funcionamiento.

6.9.4. Prospectiva

A continuación, se detallan las principales conclusiones extraídas y posibles mejoras para trabajos futuros.

En el proceso de desarrollo, consideramos la posibilidad de utilizar tecnologías de vanguardia, como **Django**, siguiendo el patrón de diseño Modelo-Vista-Controlador

(MVC). Esto nos permitirá potenciar la creación de una interfaz de usuario atractiva y eficiente.

Esta implementación se convertirá en un pilar esencial para la integración de nuestros programas en un único lugar, proporcionando la capacidad de visualizar todas las representaciones gráficas y acceder a datos de manera conveniente. Asimismo, esa API REST garantizará una interacción sencilla para los usuarios y una visualización eficiente de la información.

Tenemos en agenda la integración de modelos predictivos que harán uso de técnicas avanzadas de análisis de datos.

Se ha considerado la extracción de campos específicos de documentos PDF, como los CV de los profesionales, con el objetivo de obtener información útil para facilitar la creación de sinergias entre los miembros del equipo y fomentar la colaboración en nuevos proyectos. También queríamos extraer los resúmenes de las memorias de proyectos y procesar esa información mediante Procesamiento del Lenguaje Natural (PLN) para facilitar la identificación de temas específicos y fomentar una mejor gestión del conocimiento. Estos modelos se incorporarán a la solución para ofrecer información predictiva valiosa basada en los datos recopilados.

7. INDICADORES CLAVE DE DESEMPEÑO (KPIs) PARA SYNERATECH

La monitorización y análisis del desempeño de SYNERATECH se realizará mediante una serie de KPIs que nos permitirán entender el estado actual y evolución de la plataforma. Estos indicadores nos ayudarán a tomar decisiones basadas en datos y a ajustar nuestras estrategias según sea necesario.

A continuación, se presentan los KPIs propuestos para SYNERATECH:

7.1. Indicadores de Plataforma

Objetivo	KPI	Periodicidad	Fórmula de Cálculo
Seguimiento de la participación activa en la plataforma	Inversores Activos	Mensual	Número total de inversores activos en el mes
Seguimiento de la cantidad de proyectos supervisados	Proyectos Monitorizados	Mensual	Número total de proyectos monitorizados en el mes
Evaluar la eficiencia del proceso de generación de informes	Tiempo Promedio Informes	Mensual	Sumatoria de tiempos de generación de informes / Número total de informes

7.2. Indicadores Financieros

Objetivo	KPI	Periodicidad	Fórmula de Cálculo
Evaluar ingresos mensuales	Ingresos	Mensual	Ingresos totales del mes (crecimiento del 5% mensual supuesto)
Analizar la rentabilidad	ROI	Mensual	$ROI = \text{Ingresos} - \text{Costo} / \text{costo} \times 100\%$ (Con un costo constante de 15,000 por mes)
Seguimiento de los gastos asociados a los proyectos	Gastos por Proyecto/Centro/Investigador/Proveedor	Mensual	Gastos totales incurridos por proyecto, centro de investigación, investigador y proveedor

7.3. Indicadores de Relación

Objetivo	KPI	Periodicidad	Fórmula de Cálculo
Seguimiento de la colaboración de centros de investigación	Centros Asociados	Mensual	Número total de centros de investigación asociados en el mes
Evaluar la participación de investigadores	Investigadores Activos	Mensual	Número total de investigadores activos en el mes
Seguimiento de la colaboración entre grupos de investigación	Relaciones Grupos	Mensual	Número total de relaciones identificadas entre grupos de investigación en el mes

Estos indicadores se han seleccionado con base en su relevancia para evaluar el desempeño y crecimiento de SYNERATECH en sus áreas clave. La implementación y seguimiento de estos KPIs garantizarán una gestión eficaz y basada en datos de la plataforma.

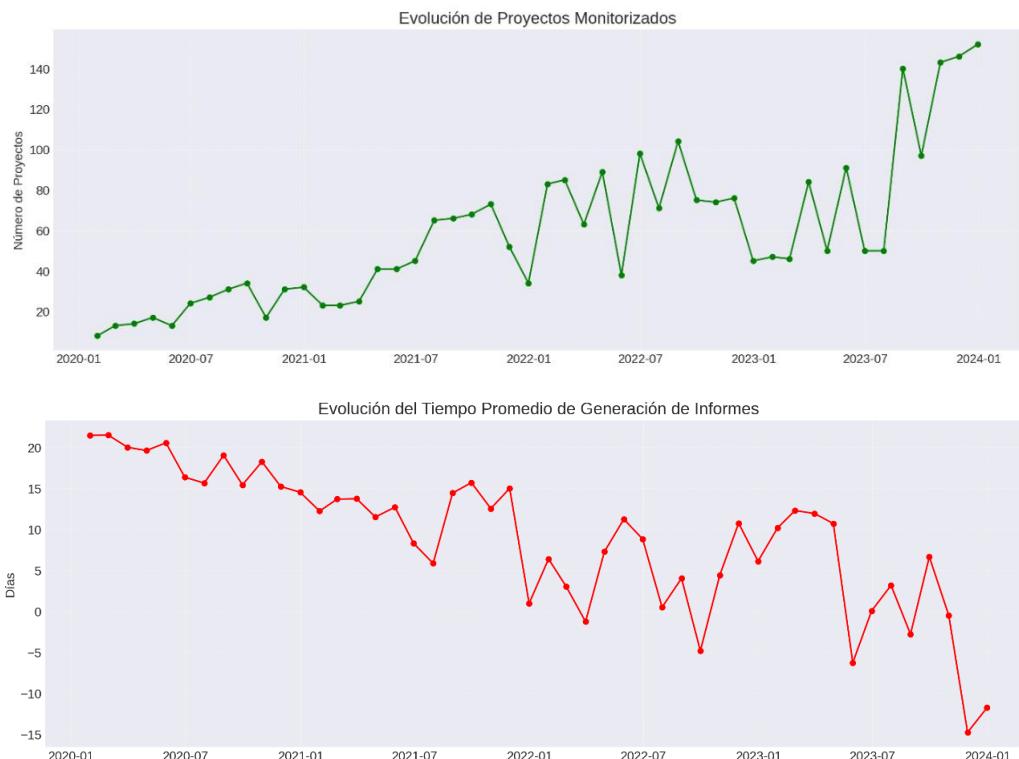
7.4. Simulación de los Indicadores

A continuación, se muestra una representación simulada de los KPI con datos ficticios para facilitar la comprensión.

Abajo e muestran las gráficas para los indicadores de plataforma:

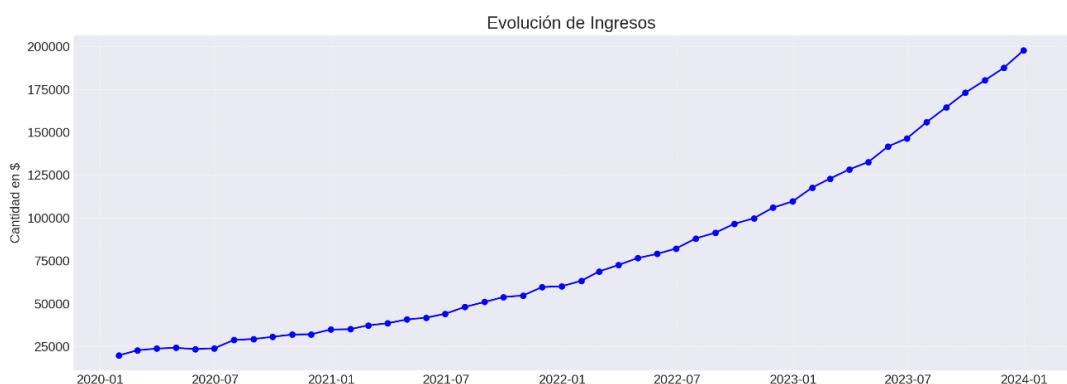
- Evolución de Inversores Activos:** Esta gráfica muestra cómo ha cambiado el número de inversores activos a lo largo del tiempo.
- Evolución de Proyectos Monitorizados:** Aquí se observa el número de proyectos que se han monitorizado mes a mes.
- Evolución del Tiempo Promedio de Generación de Informes:** Esta gráfica muestra cómo ha variado el tiempo promedio que toma generar informes a lo largo de los meses.

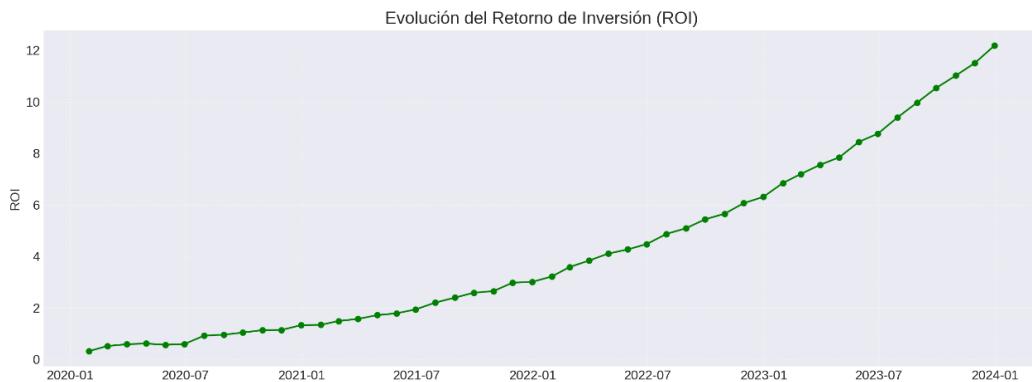




Para las gráficas para los indicadores financieros:

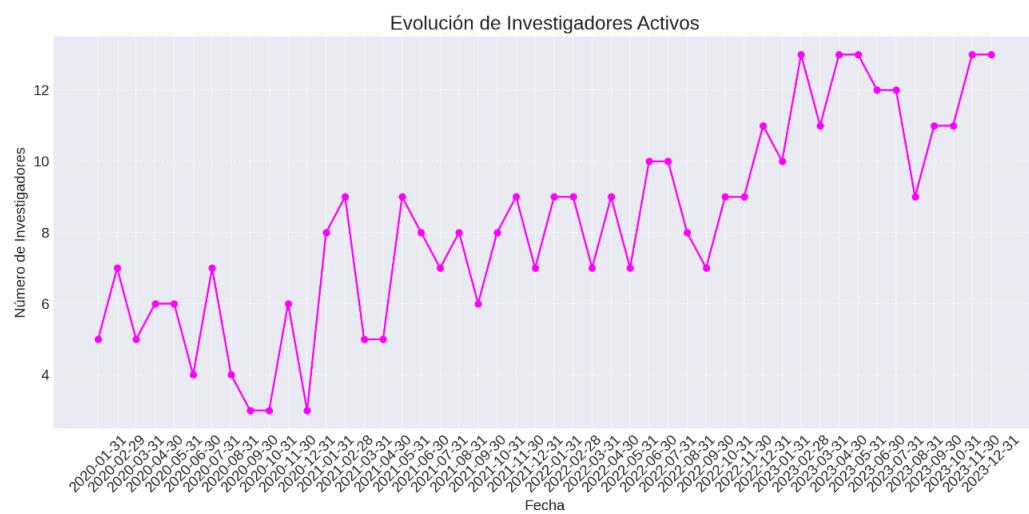
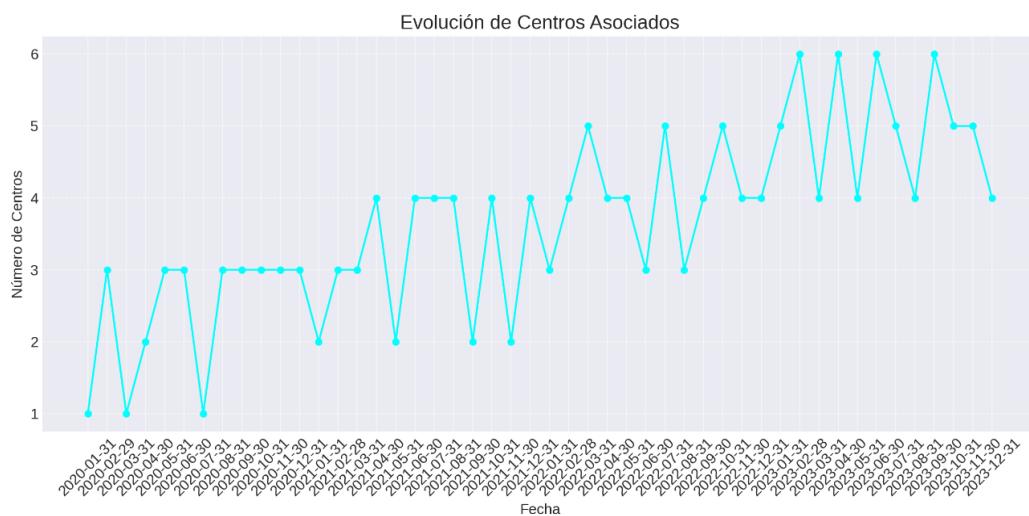
1. **Evolución de Ingresos:** Muestra cómo han cambiado los ingresos a lo largo del tiempo.
2. **Evolución del Retorno de Inversión (ROI):** Representa cómo ha variado el ROI mes tras mes.

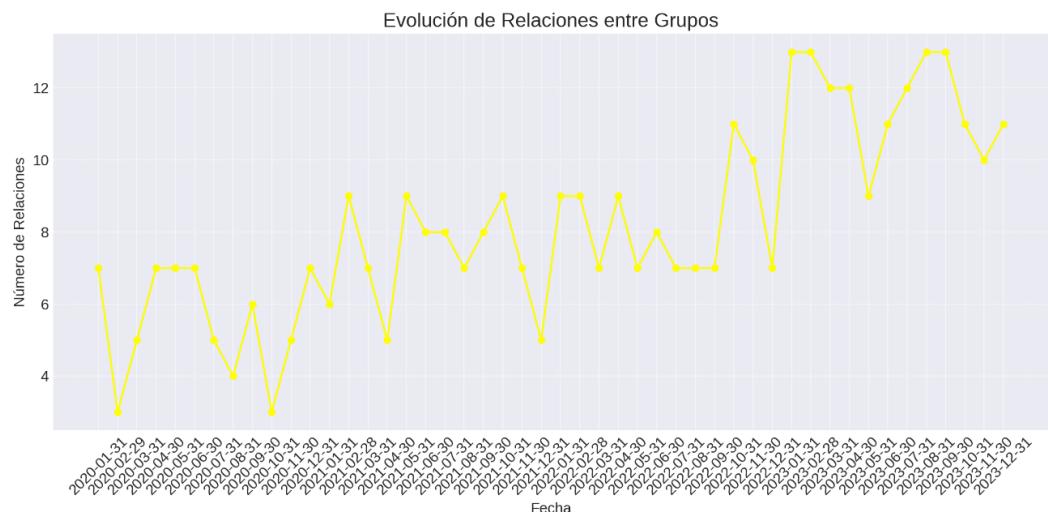




Y por último, las gráficas para los "Indicadores de Relación":

1. **Evolución de Centros Asociados:** Muestra el número de centros de investigación asociados en la plataforma mes a mes.
2. **Evolución de Investigadores Activos:** Presenta el número de investigadores activos en la plataforma cada mes.
3. **Evolución de Relaciones entre Grupos:** Representa el número de relaciones identificadas entre grupos de investigación en la plataforma cada mes.





8. ASPECTOS FINANCIEROS

8.1. Rentabilidad del proyecto

Como punto clave de este proyecto, damos paso al análisis de viabilidad, mediante un estudio económico financiero, gracias al cual podremos medir la rentabilidad del proyecto y calcular varios indicadores decisivos, para valorar la viabilidad del proyecto.

El primer punto que trataremos será la inversión inicial, en Syneratech somos 6 miembros que confiamos en este proyecto y la solución tecnológica que planteamos, por lo realizaremos una aportación de capital propio, para cubrir las primeras diligencias, como el registro de la empresa y primeros gastos administrativos derivados de la puesta en marcha. A continuación, detallamos como se distribuyen estos primeros gastos. Abordaremos las necesidades de la empresa gracias a los diferentes perfiles que tenemos, de manera que no todos asumiremos tareas de gestión de proyecto, también cubriremos necesidades de marketing, finanzas o fiscales.

Las cantidades aportadas inicialmente irán destinadas a la partida de Capital Social con las que se cubrirán la constitución de la empresa:

- Registro de la Sociedad Limitada, la cantidad asciende a 3000 euros, para consolidar la empresa bajo el nombre de Syneratech S.L.
- Solicitud de certificado de denominación social ante el Registro Mercantil Central.
- Gastos de notaría
- Inscripción en el Registro Mercantil Provincial y publicación en BORME
- Alta de los administradores de la sociedad en Seguridad Social

Otros gastos relacionados con el equipo de trabajo necesario para el desarrollo de la actividad se distribuyen en el espacio físico y componentes tecnológicos que se desglosan principalmente en hardware y software, entrando en detalle:

- Alquiler de oficina y gastos comunes asociados a la actividad, luz, agua, etc.
- Hardware: Ordenadores y material de trabajo como impresoras.
- Software: Licencias para el uso de aplicaciones y herramientas de trabajo, así como pago por entornos de la nube, para obtener la capacidad necesaria para albergar de forma segura todo el proyecto. Así como una web corporativa.
- Capital humano: Inicialmente solo participaremos los miembros del proyecto. En los años siguientes si tenemos contemplada la contratación de personal.

Recopilando y cuantificando la inversión inicial sería:

Inversión inicial				
	AÑO 0	AÑO 1	AÑO 2	AÑO 3
Gastos de constitución (Alta y gastos administrativos relacionados)	3.600 €			
Otros gastos:				
Hardware	20.000 €			
TOTAL INVERSIONES	23.600,00 €	- €	- €	- €

A dichos gastos se añadirán gastos adicionales y necesarios según comience la actividad que no son de aplicación al inicio, pero si en el transcurso del primer año. Para hacer frente a dichos gastos solicitaremos un préstamo por importe de 50.000 euros a un 3% de tipo de interés fijo a 4 años. Esto generará unos intereses de 3.122 euros que apuntaremos en la partida de gastos financieros y que más adelante veremos en detalle.

8.2. Ingresos

Nuestra principal actividad es el desarrollo de una herramienta capaz de explotar gran cantidad de datos que en el sector salud donde nos hemos especializado, es complicado el tratamiento de datos por la falta de automatización, complejidad de bases de datos e información disponible en diversos formatos y sin estructura alguna en muchos casos. Mediante la creación de esta herramienta podemos limpiar y extraer gran cantidad de datos, crear las correctas relaciones y poder analizar información de forma automática, la cual se muestra visualmente mediante Power BI. Por ello nuestra forma de conseguir ingresos sería mediante dos vías:

- Licencia de uso de la herramienta
- Servicio de consultoría que gestione el uso de la herramienta

La diferenciación es que, si el cliente adquiere el servicio completo, la explotación de la herramienta se hará de forma correcta y optimizando al máximo su uso. Si el cliente solo adquiere la opción de licencia de la herramienta, puede que debido a la complejidad y falta de "know how" no consiga poder analizar todos los datos de forma correcta.

Para la captación inicial de clientes, dejaríamos durante un mes el acceso a la herramienta a clientes interesados, así como un pequeño estudio piloto por la parte de consultoría donde mostraríamos el potencial del proyecto. La parte de la herramienta no tendría coste en este paquete y la de consultoría solo 5.000 euros, que en caso de contratación del servicio se restaría de la cuota final.

El paquete de uso de la herramienta solo incluye una guía básica de uso y acceso por un año, con opción de renovar el acceso. El total anual de este paquete asciende a 1.000 euros.

El precio del paquete completo que incluye servicio de consultoría tendría un precio adicional por persona requerida al proyecto de 10.000 euros mensuales. Con este paquete el cliente estaría cubierto de forma global desde la extracción de la información, el correcto tratamiento del dato y los informes de visualización, así como un detallado Reporting con las principales conclusiones. Depende de la empresa y la cantidad de datos a tratar se haría una propuesta por las personas que se considerasen necesarias, para la viabilidad del proyecto. El tiempo mínimo que necesita consultoría para llegar a la fase final es de 6 meses, ya que previamente se necesita un correcto estudio del dato y una adaptación de la herramienta a las necesidades del cliente. Pasado ese tiempo, el trabajo se centra en tareas de Reporting y mejoras constantes, así como propuestas de valor añadido.

Ingresos y beneficios	
Proyecto piloto	5.000 €
Licencia anual herramienta	1.000 €
Servicios de consultoría + Licencia herramienta (importe anual)	121.000 €

Proyección clientes a 4 años				
	AÑO 0	AÑO 1	AÑO 2	AÑO 3
Captación clientes	10	12	15	20
Licencia anual herramienta	3	5	7	9
Servicios de consultoría + Licencia herramienta (importe anual)	2	4	7	11

Ingresos y beneficios				
	AÑO 0	AÑO 1	AÑO 2	AÑO 3
Captación clientes	50.000 €	60.000 €	75.000 €	100.000 €
Licencia anual herramienta	3.000 €	5.000 €	7.000 €	9.000 €
Servicios de consultoría + Licencia herramienta (importe anual)	232.000 €	464.000 €	812.000 €	1.276.000 €
TOTAL INGRESOS/BENEFICIOS	285.000,00 €	529.000,00 €	894.000,00 €	1.385.000,00 €

8.3. Gastos

Tener una proyección lo más exacta posible de los gastos nos ayudara al correcto desarrollo del proyecto y a facilitarnos las decisiones en el transcurso de estos 4 años que inicialmente hemos planteado como tiempo necesario para crecer y estabilizar la empresa con unos ingresos constantes y recurrentes.

Los principales gastos a tener en cuenta son:

- Gastos iniciales de creación de la sociedad.
- Gastos de alquiler y asociados al funcionamiento de la oficina. Valorando una subida del 2% anual.
- Licencias de software y programas necesarios para el desarrollo de la actividad. Debido a la proyección planteada, pronosticamos un 5% de aumento en pago de licencias.
- Material de oficina y hardware. Al igual que con la parte de licencias, debido al crecimiento proyectado de la empresa que conllevara contrataciones, valoramos un 20% anual de incremento, el aumento es mayor que en la parte de licencias, ya que estas muchas veces son paquetes que agrupan varias licencias.
- Web, mantenimiento y actualización.

- Entorno seguro en la nube, donde residirá nuestro servidor. Debido a que esperamos crecer en clientes cada año, debemos contemplar un incremento de capacidad cada año y tener en cuenta que no todos los clientes tratar la misma cantidad de datos ni de la misma forma. Inicialmente marcaremos un 50% de incremento cada año.
- Acciones de marketing, enfocadas en ferias y eventos del sector donde darnos a conocer. Nuestra idea es tener una inversión mayor los primeros años para darnos a conocer y después mantener unos gastos estables en esta partida, para participar en eventos y actos puntuales.
- Contratación de capital humano, durante el 1º año, los socios seremos los componentes de la empresa y después del primer año se contempla la entrada de dos personas y mantener ese incremento hasta el año 4.

A continuación, mostramos el detalle económico de los gastos teniendo en cuenta las variaciones comentadas:

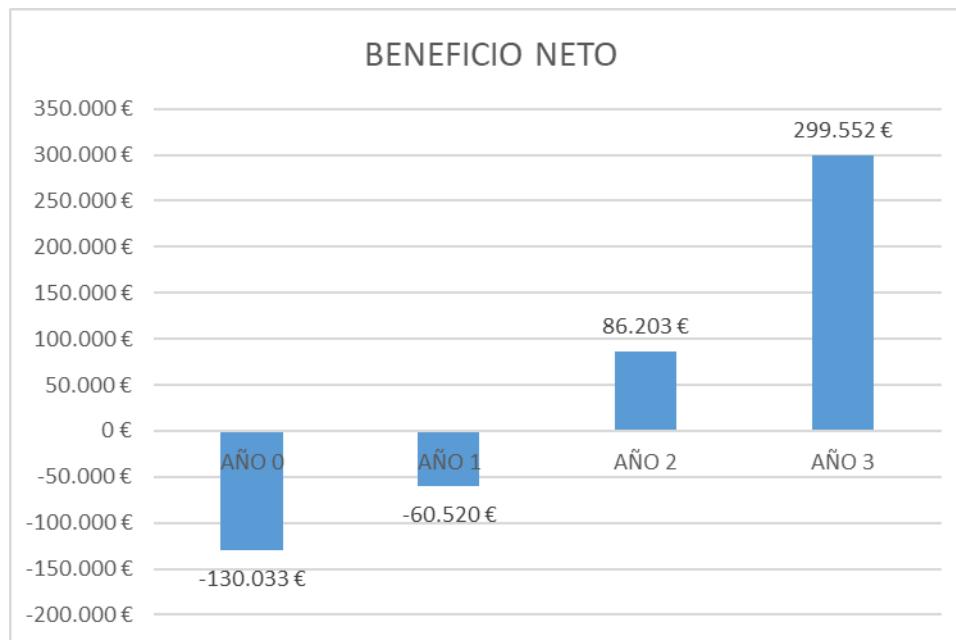
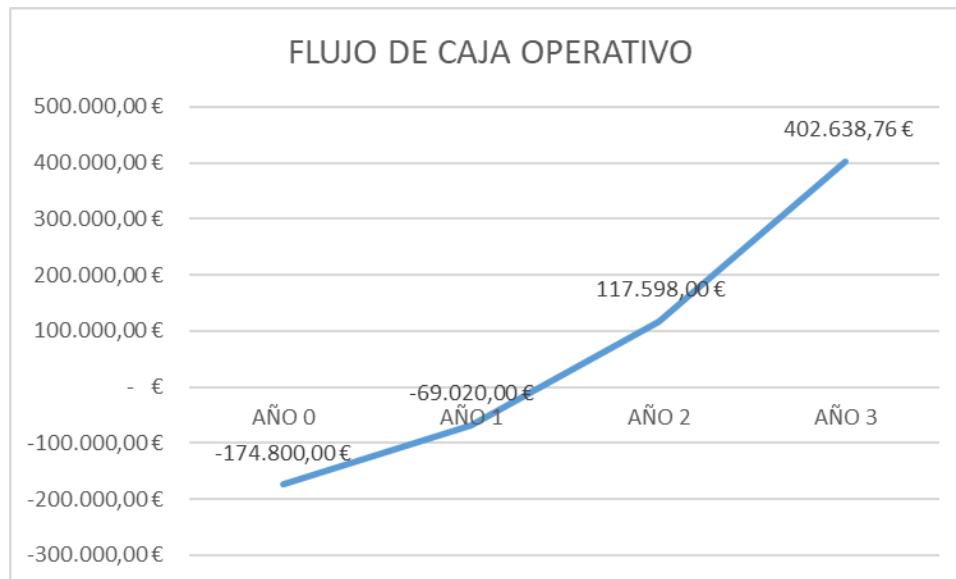
Gastos				
	AÑO 0	AÑO 1	AÑO 2	AÑO 3
Gastos iniciales de constitución	3.600 €	-	-	-
Gastos corrientes del desarrollo de la actividad:				
Alquiler de oficina y gastos asociados al funcionamiento de la actividad	30.000 €	30.600 €	31.212 €	31.836 €
Software y licencias	10.000 €	10.500 €	15.750 €	23.625 €
Material de oficina y hardware	-	4.000 €	4.800 €	5.760 €
Web y mantenimiento	5.000 €	5.000 €	5.000 €	5.000 €
Servidor	4.000 €	6.000 €	9.000 €	13.500 €
Marketing				
Acciones de marketing	20.000 €	18.000 €	12.000 €	12.000 €
Gastos de personal				
Sueldos y salarios(+S.S)	360.000 €	519.120 €	692.640 €	883.440 €
Gastos de gestoria laboral	3.600 €	4.800 €	6.000 €	7.200 €
TOTAL GASTOS	436.200,00 €	598.020,00 €	776.402,00 €	982.361,24 €

8.4. Conclusiones

Según las previsiones en el 1º Semestre del 3º Año recuperaremos la inversión inicial.

La evolución del flujo de caja es creciente lo que también denota la viabilidad del proyecto. Las tasas que hemos aplicado son las sugeridas en las fórmulas facilitadas, un 15%.

	AÑO 0	AÑO 1	AÑO 2	AÑO 3
FLUJO DE CAJA OPERATIVO	-174.800,00 €	- 69.020,00 €	117.598,00 €	402.638,76 €
VALOR ACTUAL	-174.800,00 €	- 62.745,45 €	97.188,43 €	302.508,46 €
ACUMULADO	-174.800,00 €	-237.545,45 €	-140.357,02 €	162.151,44 €



Por último, los valores del VAN, TIR, TIRM y PAYBACK arrojan unos valores adecuados que respaldan la viabilidad de la propuesta:

VAN:	468.445,11	
TIR:	35,54%	
TIRM	32,66%	
PAYBACK	2,08	Durante el 1º semestre del 3º año

A principio del 3º año consideramos que habremos recuperado la inversión inicial. La combinación con un VAN positivo sugiere una alta viabilidad financiera del proyecto. La inversión parece ser rentable y podría generar rendimientos significativos.

La TIR del 35,54% indica que este proyecto ofrece un rendimiento significativamente superior a las tasas de interés convencionales. Asimismo, la TIRM del 32,66% sugiere que, si los costos de financiamiento aumentan a lo largo del tiempo, el proyecto aún debería ser rentable, aunque a un ritmo ligeramente más moderado.

Después de valorar desde el punto de vista económico proyecto, consideramos viable la puesta en marcha del proyecto, **todos los indicadores y ratios analizados nos muestran una proyección positiva.**

9. CONCLUSIONES

Gracias al camino recorrido en la creación de este proyecto hemos podido desarrollar muchas de las tecnologías y conocimientos de los que dispone el equipo de Syneratech, demostrando su capacidad de proporcionar los servicios que ofrece usando datos y escenarios reales. Los principales hitos que hemos logrado en la elaboración del producto o servicio mínimo viable para la Fundación Científica de la Asociación Española Contra el Cáncer han sido:

1. Desplegar una base de datos.
2. Administrar y gestionar la base de datos desplegada.
3. Utilizar un contenedor y una máquina virtual para alojar la base.
4. Gestionar procesos ETL mediante scripts en Notebooks de Google Collab.
5. Poder analizar relaciones gracias a Neo4j (Neo4j Aura).
6. Visualizar todos nuestros análisis, mediante la creación de dashboards en PowerBI.

Todo esto sin perder del punto de mira el aspecto económico y valorando el ahorro en costes, sin verse perjudicado el uso de la tecnología y los requerimientos mínimos de capacidades y desarrollos necesarios.

Podemos concluir que hemos demostrado la viabilidad de Syneratech como consultora de Business Intelligence para el sector de investigación sanitaria en base a los siguientes puntos:

1. Existe demanda de servicios de consultoría y transformación digital en el sector de investigación sanitaria.
2. Syneratech ofrece servicios y tarifas que sus principales competidores no ofrecen actualmente.
3. Hemos elaborado un modelo de negocio identificando nuestra propuesta de valor diferencial, nuestros principales potenciales clientes, nuestros socios y proveedores, nuestras actividades estratégicas, así como los canales de distribución y recursos necesarios para nuestra empresa.
4. Hemos ofrecido una solución tecnológica basada en herramientas de vanguardia a un cliente real con datos reales, demostrando las capacidades de Syneratech.
5. Hemos identificado indicadores de plataforma de inversores, indicadores financieros y de relaciones con los agentes del ecosistema que permiten medir y monitorizar el rendimiento y desempeño de nuestra empresa.
6. Hemos realizado un estudio de rentabilidad de la empresa en un escenario no del todo favorable que ofrece una recuperación de la inversión a principios del tercer año de vida de la empresa.

Por todo esto, queremos reivindicar la viabilidad de Syneratech y de las capacidades de su equipo. Juntos hemos crecido no solo en aspectos relacionados con la tecnología, si no también en otros aspectos más centrados en el mundo empresarial como:

- Detección de oportunidad de negocio.
- Análisis de viabilidad.
- Elección de tecnología a implantar.
- Cálculo de rentabilidad económica.

Siempre valorando cada aspecto a tratar, sin perder la visión global del proyecto. Para ello hemos tenido en cuenta en todo momento los aspectos operativos y estratégicos para dar solución a la necesidad detectada.

Además, para nosotros la ejecución de este proyecto ha significado afrontar un trabajo real para un cliente que tiene una necesidad real, la cual esperamos haber cubierto.

Dentro de los seis miembros que componen este proyecto, los perfiles son muy diversos, lo que nos ha permitido aprender los unos de los otros y poder cubrir las diferentes cuestiones que el proyecto requería. Con la entrega de este proyecto finalizamos este año de aprendizaje y es el resultado de todo lo aprendido.

10. BIBLIOGRAFÍA

1. Organización Mundial de la Salud (OMS). [En línea] 2 de Febrero de 2022. <https://www.who.int/es/news-room/fact-sheets/detail/cancer>.
2. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 71, 2021, CA Cancer J Clin., Vol. 3, págs. 209-249.
3. Saloni Dattani, Fiona Spooner, Hannah Ritchie and Max Roser. <https://ourworldindata.org/causes-of-death>. [En línea] 2023. [Citado el: 10 de 10 de 2023.] <https://ourworldindata.org/causes-of-death>.
4. Research, Precedence. Oncology Market - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023 – 2032. 2023.
5. The impact of COVID-19 on globalization. Nistha Shrestha, Muhammad Yousaf Shad, Osman Ulvi, Modasser Hossain Khan, Ajlina Karamehic-Muratovic, Uyen-Sa D.T. Nguyen, Mahdi Baghbanzadeh, Robert Wardrup, Nasrin Aghamohammadi, Diana Cervantes, Kh. Md Nahiduzzaman, Rafdzah Ahmad Zaki, Ubydul Haque. 2020, One Health, Vol. 11, pág. 100180.
6. Jackie Spencer, Kale Frank, Raya Bousleiman. Healthcare Industry Trends Mid-Year 2023 Update. s.l. : Silicon Valley Bank, 2023.
7. Mejorar la Transferencia de Conocimiento y la Colaboración entre Ciencia y Empresa en España. s.l. : OECD publishing, 2021.
8. Nueva Agenda Europea de Innovación. s.l. : Comisión Europea, 2022.
9. Plan de Transferencia y Colaboración del Ministerio de Ciencia e Innovación. s.l. : Ministerio de Ciencia e Innovación. Secretaría General Técnica. Centro de Publicaciones., 2022.
10. Ley 28/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes (Ley de Start-Ups). s.l. : Boletín Oficial del Estado (BOE), 2022. núm. 306, de 22/12/2022..
11. HBM New Drug Approval Report. s.l. : HBM Partners, 2018.
12. The EU's 2021-2027 long-term budget and NextGenerationEU. s.l. : Comisión Europea, 2020.
13. Plan de Recuperación, Transformación y Resilencia. [En línea] [Citado el: 17 de Octubre de 2023.] <https://planderecuperacion.gob.es/>.
14. Informe AseBio 2022: Deep Tech para Transformar el Mundo. s.l. : Asociación Española de Bioempresas (AseBio), 2023.
15. SmartSimple Wiki. [En línea] [Citado el: 26 de Octubre de 2023.] https://wiki.smartsimple.com/wiki/Main_Page.

11. ANEXO I: Glosario de Campos de la BBDD

Tabla	Nombre del Campo	Descripción Contenido	Tipo de Campo	Valores Permitidos
USERS	project_id	Código identificativo del proyecto caracterizado por 5 letras correspondientes al tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante	Texto Libre	Todos
USERS	grant_type	Nombre de la convocatoria de la beca junto al año de la convocatoria. El año indica el año natural en el que se producirá la adjudicación y donde se liberarán los primeros fondos	Texto Libre	Todos
USERS	project_status	Estado de la solicitud del proyecto. Indica el estado de la solicitud, si ha sido aprobada, es decir, adjudicataria, declinada o se encuentra en evaluación	Lista (único valor permitido)	Eligibility; Draft EOI; Submitted EOI; Revisions Requested EOI; Under Review EOI; Declined EOI; Draft Application; Submitted Application; Revisions Requested Application; Under Review Application - Evaluation; Draft Scientific Committee; Submitted Scientific Committee; Under Review Application - Committee; Approved - Pre-Award; Approved - Pending Grant Agreement; Initial Draft Project; Approved; Declined; Closed; Withdrawn; On-Hold Application; Out of time;
USERS	user_name	Nombre y primer apellido del solicitante de la ayuda	Texto Libre	Todos
USERS	user_short_name	Nombre del solicitante de la ayuda	Texto Libre	Todos
USERS	user_surname	Primer y segundo apellido del solicitante de la ayuda	Texto Libre	Todos
USERS	user_id_type	Tipo de documento identificativo	Lista (único valor permitido)	DNI;NIE;Passport;(vacío)
USERS	dni	Número de documento identificativo. Se usará como clave primaria (primary key) para unir proyectos y usuarios	Texto Libre	Todos
USERS	date_birth	Fecha de nacimiento del solicitante de la ayuda	Fecha	Todos
USERS	gender	Género del solicitante de la ayuda	Lista (único valor permitido)	Male; Female; No answer;(vacío)
USERS	place_birth	Lugar de nacimiento del solicitante de la ayuda	Texto Libre	Todos
USERS	user_current_organization	Centro de trabajo actual del solicitante de la ayuda. A veces no coincide con el centro de investigación asociado a la ayuda porque la administración del centro recae en una institución superior que gestiona varios centros	Texto Libre	Todos
USERS	organization_id	Institución o centro de investigación que gestiona la ayuda del solicitante y donde se suele desarrollar la investigación. Sirve como	Texto Libre	Todos

		clave secundaria para unir con el resto de tablas		
USERS	user_role	Rol que el usuario desempeña en la plataforma. Puede ser investigador (Researcher), evaluador (Reviewer) o puede ser el gestor de un centro de investigación (Grant Manager) que presenta la solicitud de proyecto en nombre de otra persona, o presenta la documentación en nombre de otro responsable de una ayuda activa	Lista (único valor permitido)	Grant Manager; Researcher;(vacío)
USERS	user_province	Provincia donde trabaja el usuario o solicitante. A veces no coincide con la provincia de la ayuda	Texto Libre	Todos
USERS	organization_province	Provincia del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
USERS	user_city	Ciudad donde trabaja el usuario o solicitante	Texto Libre	Todos
USERS	user_country	País donde trabaja el usuario o solicitante	Texto Libre	Todos
USERS	user_cv_summary	Resumen del Curriculum Vitae elaborado por el propio solicitante. Pocos solicitantes lo tienen completado.	Texto Libre	Todos
USERS	user_orcid	Código ORCID del investigador con el que se relaciona con las publicaciones científicas. Pocos usuarios la han incorporado a sus perfiles	Numéricos	Todos
USERS	user_creation_date	Fecha de creación del perfil del usuario	Fecha	Todos
USERS	user_update_date	Fecha de última modificación del perfil del usuario	Fecha	Todos
INVOICE	project_id	Código identificativo del proyecto caracterizado por 5 letras correspondientes al tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante	Texto Libre	Todos
INVOICE	invoice_status	Estado de la solicitud del proyecto. Indica el estado de la solicitud, si ha sido aprobada, es decir, adjudicataria, declinada o se encuentra en evaluación	Lista (único valor permitido)	Approved;Cancel;Declined;Draft;Submitted
INVOICE	invoice_name	Nombre libre de la factura que le asigna el usuario	Texto Libre	Todos
INVOICE	invoice_date	Fecha de emisión de la factura	Fecha	Todos
INVOICE	payment_date	Fecha del pago de la factura (suele coincidir con la fecha de la factura, aunque en ocasiones tiene cierta demora)	Fecha	Todos
INVOICE	reporting_year	Año de la ayuda en el que se reporta la factura. Ejemplo: Si se reporta la factura durante la justificación económica pasado el primer año, este valor será "Y1"	Texto Libre	Todos
INVOICE	starting_evaluation_period	Fecha de inicio del período que abarca la justificación económica	Fecha	Todos
INVOICE	end_evaluation_period	Fecha del final del período que abarca la justificación económica	Fecha	Todos
INVOICE	supplier_id	Nombre del proveedor o emisor de la factura. Puede contener nombres ligeramente diferentes del mismo proveedor	Texto Libre	Todos
INVOICE	supplier_vat	Número VAT (identificación fiscal) del proveedor o emisor de la factura. Puede contener errores	Texto Libre	Todos
INVOICE	concept	Categorías generales a las que pertenece la factura	Lista (único valor permitido)	TRAVEL EXPENSES;PERSONNEL;EXPENDABLE MATERIAL;TECHNICAL SERVICES;OTHERS;EQUIPMENT
INVOICE	subconcept	Subcategorías más específicas a las que pertenece la factura	Lista (único valor permitido)	Expendable laboratory material;Health care material;Cleaning material;Computer consumables;Office equipment;TC2;Paysheet;Acquisition;rental;maintenance or repair of

				equipment;Software applications;Animal facility;Sequencing;Proteomics;Genomics ;Histopathology;Flow cytometry;Microscopy;Imaging;Travel expenses;Subsistence allowances;Meeting registration;Stays of researchers;Publication/Dissemination;Professional services;Books/Scientific journals;Training activities;Courier services;Outsourcing activities;Industrial/Intellectual property rights;Patents/Licensing results;Technical/Legal/Business advice;Innovation and transfer of results;Commercial/Market analysis;Regulatory development;Production/Design/Prototyping of proof of concept;Publication/Dissemination;Professional services;Books/Scientific journals;Training activities;Courier services;Others
INVOICE	concept_description	Descripción libre del contenido de la factura. Muchas veces incluye los artículos que se compran, o el nombre del trabajador al que le pertenece la nómina, etc.	Texto Libre	Todos
INVOICE	organization_claim	Institución o centro de investigación que aparece como cliente en la factura, muchas veces coincide con el organization_id pero otras no porque es texto libre y a veces se equivocan. En otras ocasiones se emite la factura a la institución superior que gestiona los pagos del centro de investigación	Texto Libre	Todos
INVOICE	organization_id	Institución o centro de investigación que gestiona la ayuda del solicitante y donde se suele desarrollar la investigación. Sirve como clave secundaria para unir con el resto de tablas	Texto Libre	Todos
INVOICE	organization_economic_justification	Institución o centro de investigación que declara la factura en una justificación económica, muchas veces coincide con el organization_id	Texto Libre	Todos
INVOICE	organization_vat	Número VAT (identificación fiscal) del centro de investigación que aparece como cliente en la factura. Puede contener errores	Texto Libre	Todos
INVOICE	invoice_amount	Importe total en euros de la factura	Numéricos	Todos
INVOICE	invoice_manager	Persona responsable del centro o institución encargada de presentar las facturas en la justificación económica	Texto Libre	Todos
INVOICE	user_name	Nombre y primer apellido del solicitante de la ayuda	Texto Libre	Todos
INVOICE	expense_type	Tipo de gasto de la factura, es decir, si se ha producido el gasto fuera (external) o dentro (internal) del propio centro. Los gastos externos suelen ser compras a proveedores nacionales o internacionales, y los gastos internos son compras al almacén o servicios del propio centro de investigación	Lista (único valor permitido)	External;Internal
PROJECTS	project_id	Código identificativo del proyecto caracterizado por 5 letras correspondientes al	Texto Libre	Todos

		tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante		
PROJECTS	grant_type	Nombre de la convocatoria de la beca junto al año de la convocatoria. El año indica el año natural en el que se producirá la adjudicación y donde se liberarán los primeros fondos	Texto Libre	Todos
PROJECTS	project_status	Estado de la solicitud del proyecto. Indica el estado de la solicitud, si ha sido aprobada, es decir, adjudicataria, declinada o se encuentra en evaluación	Lista (único valor permitido)	Approved - Pending Grant Agreement;Approved - Pre-Award;Approved;Closed;Declined EOI;Declined;Initial Draft Project;On-Hold Application;Out of time;Under Review Application - Evaluation;Under Review EOI
PROJECTS	project_phase	Estado de la solicitud del proyecto. Indica el estado de la solicitud, si ha sido aprobada, es decir, adjudicataria, declinada o se encuentra en evaluación	Lista (único valor permitido)	Eligibility; Draft EOI; Submitted EOI; Revisions Requested EOI; Under Review EOI; Declined EOI; Draft Application; Submitted Application; Revisions Requested Application; Under Review Application - Evaluation; Draft Scientific Committee; Submitted Scientific Committee; Under Review Application - Committee; Approved - Pre-Award; Approved - Pending Grant Agreement; Initial Draft Project; Approved; Declined; Closed; Withdrawn; On-Hold Application; Out of time;
PROJECTS	user_name	Nombre y primer apellido del solicitante de la ayuda	Texto Libre	Todos
PROJECTS	dni	Número de documento identificativo. Se usará como clave primaria (primary key) para unir proyectos y usuarios	Texto Libre	Todos
PROJECTS	organization_id	Institución o centro de investigación que gestiona la ayuda del solicitante y donde se suele desarrollar la investigación. Sirve como clave secundaria para unir con el resto de tablas	Texto Libre	Todos
PROJECTS	project_start_date	Fecha de inicio del proyecto cubierto por una ayuda económica concreta	Fecha	Todos
PROJECTS	project_end_date	Fecha del final del proyecto cubierto por una ayuda económica concreta	Fecha	Todos
PROJECTS	project_update_d_end_date	Fecha actualizada del final del proyecto cubierto por una ayuda económica concreta. Esto ocurre cuando se conceden prórrogas de tiempo para gastarse el dinero asignado si no lo han gastado en el tiempo asignado, también debido a bajas médicas, por maternidad o paternidad, etc.	Fecha	Todos

PROJECTS	project_close_date	Fecha de cierre del proyecto cubierto por una ayuda económica concreta. Esto ocurre cuando el beneficiario ya ha presentado toda la documentación relacionada con las justificaciones de la ayuda y se da por cerrada la aportación de datos de este proyecto	Fecha	Todos
PROJECTS	project_initial_assigned_amount	Cantidad inicial de dinero asignada al proyecto antes de que comience el mismo y que se refleja en la firma del convenio	Numéricos	Todos
PROJECTS	project_current_assigned_amount	Cantidad actual de dinero firmada en el convenio asignada al proyecto junto a la cantidad aprobada en las extensiones de presupuesto. Esto ocurre cuando se dan bajas de maternidad o paternidad, por situaciones excepcionales como durante el confinamiento de la pandemia de COVID-19. También se conceden extensiones si lo solicitan los investigadores de manera justificada y el patronato de la Fundación aprueba conceder estos fondos de manera excepcional	Numéricos	Todos
PROJECTS	project_paid_amount	Cantidad total de dinero desembolsada hasta el momento para un proyecto en concreto	Numéricos	Todos
PROJECTS	project_total_extensions_amount	Cantidad total de dinero aprobada como extensión a la cantidad inicial asignada a un proyecto. Ejemplo: años de prórroga en contratos predoctorales o posdoctorales, o extensiones excepcionales	Numéricos	Todos
PROJECTS	organization_province	Provincia del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
PROJECTS	organization_community_aut	Comunidad autónoma del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
PROJECTS	research_type	Clasificación de los proyectos de investigación por tipo de investigación. Clasificación usada en todos las convocatorias salvo PPL (prácticas de laboratorio), CLFOR (clínico en formación), INNOV (AECC Innova) e IMPAC (AECC Impacto)	Atributos en Lista (varios valores por celda)	Basic;Pre-Clinical;Clinical;Epidemiology;Psycho-oncological;Translational;Others;Innovative
PROJECTS	research_type_innova	Clasificación de los proyectos de investigación por tipo de investigación. Clasificación usada solo en la convocatoria AECC Innova	Atributos en Lista (varios valores por celda)	Therapeutics;Diagnostic;Medical Devices;Drug discovery;Preclinical, clinical - validation assays
PROJECTS	research_type_startup	Clasificación de los proyectos de investigación por tipo de investigación. Clasificación usada en todos las convocatorias salvo PPL (prácticas de laboratorio), CLFOR (clínico en formación), INNOV (AECC Innova) e IMPAC (AECC Impacto)	Atributos en Lista (varios valores por celda)	Therapeutics;Diagnostic;Medical Devices;Drug discovery;Preclinical, clinical - validation assays;Digital health
PROJECTS	project_name	Nombre del proyecto	Texto Libre	Todos
PROJECTS	abstract	Resumen de la propuesta de proyecto. Suele contener brevemente la necesidad médica no resuelta, la solución propuesta y los objetivos que se quieren abarcar con la financiación solicitada	Texto Libre	Todos
PROJECTS	type_cancer	Clasificación de los proyectos de investigación por tipo de cáncer donde tendrían impacto. Clasificación usada en todos las convocatorias salvo INNOV (AECC Innova)	Atributos en Lista (varios valores por celda)	Acute Lymphoblastic Leukemia;Acute Lymphoblastic Leukemia (ALL);Acute Myeloblastic Leukemia;Anal Cancer;Bile Duct Cancer;Bladder Cancer;Breast Cancer;Childhood Cancer;Central Nervous System;Cervix Cancer;Chronic myelogenous leukemia;Colon

		Cancer;Colorectal Cancer;Cronic Lymphatic Leukemia;Esophagu s Cancer;Gallblader Cancer;Hematologi cal Malignancies;Hodgk in Lymphoma;Kidney Cancer;Laringeal Cancer;Liver Cancer;Lung Cancer;Melanoma Skin Cancer;Mesotheli oma;Multiple Myeloma;Nasophar yngeal Carcinoma or Cavum;Nasopharyn geal Carcinoma / Cavum;Non- hodgkin Lymphoma;Osteosar coma;Others;Ovaria n Cancer;Pancreatic Cancer;Prostate Cancer;Rare Type of Cancer;Rectal Cancer;Skin Carcinoma;Small Intestine Cancer;Soft Tissue Sarcoma;Stomach Cancer;Testicular Cancer;Thyroid Cancer;Uterine and Endometrial Cancer;Vaginal Cancer;All Cancer Types		
PROJECTS	phase_cancer	Clasificación de los proyectos de investigación por fase del cáncer donde tendrían impacto (diagnóstico, prevención, tratamiento, metástasis, etc). Clasificación usada en todos las convocatorias salvo INNOV (AECC Innova)	Atributos en Lista (varios valores por celda)	Clasification and Diagnosis;Prevention ;Early Diagnosis;Classificati on and Diagnosis;Treatment; Resistance;Metastasi s;Others
PROJECTS	aei_evaluation_s core	Puntuación asignada por la Agencia Estatal de Investigación (AEI) tras la evaluación del proyecto. No todos los proyectos son evaluados por la AEI	Numéricos	Todos
PROJECTS	sponsor_name	Nombre del patrocinador del proyecto. Algunas veces, una empresa aporta capital para patrocinar el proyecto. En la mayoría de las ocasiones el patrocinador es alguna de las sedes provinciales de la AECC donde se desarrolla el proyecto	Texto Libre	Todos
PROJECTS	granted_year	Año de adjudicación de la ayuda. Ejemplo: las convocatorias adjudicadas en 2023 se han convocado en noviembre de 2022. En este campo aparecería "Adjudicados en 2023"	Texto Libre	Todos
PROJECTS	research_team_ table	Campo que contiene una tabla compleja sobre los investigadores que trabajan en un proyecto concreto	Texto Libre	Todos
PROJECTS	keywords	Palabras clave que describen el proyecto y que han sido definidas por los gestores de proyecto de la Fundación Científica. Apenas se han usado y pocos proyectos las contienen	Atributos en Lista (varios valores por celda)	Patient advocacy;ex holder AECC grant;Leukemia;Sarc oma;Prostate;Colan giocarcinoma;Wom en;Recurrence;Men; Nanotechnology;Che motherapy;Immunot herapy ;PD1/PD L1;TILs;CAR-T;Others

				CARs;B cells;T cells;Nanobody ;Personalized medicine?;Big data;OMICS;Liquid biopsy;Metastasis;Metabolism;Immune response;Side effects;Long survival ;Toxicity ;Microbiology ;Cell therapy ;BRAF;BRCA;p53;PTEN;RAD51;PARP;EGFR;VAV;HIF;STAT3;TET2;HER2;YES;mTORC;KRAS;DNA;RNA;miRNA;RNA;Protein;Circulating tumor DNA ;Circulating tumor cells;Cancer Stem Cells ;Dormant cells;DNA damage;Cell proliferation ;Exosomes;Tumor growth;Tumor microenvironment;Immune System;Quiescence;Infiltrated;Epigenetics;Heterogeneity;Genetic engineering;Oncogene;Tumor suppressor gene;Pre-malignant state;Screening;CRISPR;Fecal Occult Blood Test;Sequencing ;SC-Seq;Cell line;Mice;Zebrafish;3D;Cultures;Organoids;PDX;Patient sample;In vitro;In vivo;Clinical trial ;Triple negative;Hereditary cancer;Candidate Innovation;Academic trial;Other treatments;Vaccines ;Patents;ER;Neuroblastoma;Medulloblastoma
PROJECTS	protectable_ip	Si el proyecto ha generado o espera generar resultados susceptibles de ser protegidos mediante patente	Lista (único valor permitido)	Yes;No
GRANT_TYPES	grant_type	Nombre de la convocatoria de la beca junto al año de la convocatoria. El año indica el año natural en el que se producirá la adjudicación y donde de liberarán los primeros fondos	Texto Libre	Todos
GRANT_TYPES	grant_status	Estado de la convocatoria. Si la convocatoria se abrirá próximamente estará en estado "Draft", se encuentra ya abierta y acepta solicitudes estará en estado "Active", si ya está cerrada y no acepta solicitudes estará en estado "Closed".	Lista (único valor permitido)	Draft;Closed;Active; No Status;Rescheduled
GRANT_TYPES	submission_start_date	Fecha en la que se abre la convocatoria y a partir de la cual ya se podrán presentar solicitudes	Fecha	Todos
GRANT_TYPES	submission_end_date	Fecha en la que se cierra la convocatoria y a partir de la cual ya no se podrán presentar solicitudes	Fecha	Todos
GRANT_TYPES	resolution_date	Trimestre del año o fecha aproximada en la que se espera publicar la resolución o tener la lista de adjudicatarios	Texto Libre	Todos

GRANT_TYPES	grant_year	Año de la convocatoria al que está asociada la ayuda. Por ejemplo, la convocatoria de las ayudas del año 2023 se publican en septiembre-noviembre de 2022, y se resuelven a lo largo del 2023. La fecha límite de firma de contratos o convenios de las ayudas 2023 es el 1 de diciembre de 2023	Numéricos	Todos
GRANT_TYPES	grant_eoi	Si la convocatoria cuenta con una fase de evaluación previa a través de la presentación de una "Expression Of Interest (EOI)". Que consiste en una propuesta preliminar de proyecto donde se hace un primer cribado entre todas las propuestas para quedarse con las mejores antes de enviarlas a la Agencia Estatal de Investigación	Lista (único valor permitido)	Yes;No
GRANT_TYPES	grant_classification	Clasificación de las convocatorias en los grandes programas de apoyo a la investigación de la Fundación: Talento (destinadas a personas y para mejorar su formación y capacitación), Proyectos (destinadas a proyectos de investigación de diversa naturaleza); Globalización (diseñadas para que puedan solicitarlas grupos o personas de fuera de España y que suelen tener particularidades ya que se diseñan y se adaptan sobre plantillas europeas junto a otras organizaciones o instituciones), Innovación (destinadas a proyectos innovadores y con el objetivo de apoyar el desarrollo de nuevas tecnologías y acelerar su llegada a los pacientes, también promueven la trasferencia tecnológica de los centros de investigación al mercado), Programa Excelencia (se trata de ayudas de gran cuantía destinadas a centros de investigación para que desarrollen una línea de investigación amplia y coordinada con varios grupos del propio centro y varios subproyectos diferentes), Reconocimientos (premio o menciones que organiza la Fundación para premiar la labor investigadora de los investigadores)	Lista (único valor permitido)	Talento;Proyectos;Globalización;Innovación;Programa Excelencia;Reconocimientos
GRANT_TYPES	grant_subclassification	Nombre de cada una de las convocatorias de apoyo a la investigación de la Fundación Científica de la AECC.	Lista (único valor permitido)	Seed Ideas;Innova;Lab;Projects;Coordinated Groups;Internships;Pre-doctoral;Masters;Postdoctoral;Researcher;Junior Physician;Senior Physician;Trascan;Worldwide Cancer Research;PerMed;Accelerator Awards;Ensayos clínicos;Programs;Proyectos Coordinados;Strategic Projects;General Projects;Excellence program;Cancer Grand Challenges;Cancer Innova;Talent;Impact;Survivorship;Catedras;Acknowledgment;Sinergias
GRANT_TYPES	grant_description	Descripción de la convocatoria, objeto y posibles beneficiarios	Texto Libre	Todos
GRANT_TYPES	grant_duration	Duración en años de la ayuda	Numéricos	Todos
GRANT_TYPES	grant_amount	Cuantía de dinero destinada a cada una de las ayudas adjudicadas	Numéricos	Todos
GRANT_TYPES	grant_recipient	Requisitos del solicitante para poder solicitar la ayuda. Ejemplos: titulación universitaria, expediente o qué tipo de centro, etc.	Texto Libre	Todos
ORGANIZATIONS	organization_id	Institución o centro de investigación que gestiona la ayuda del solicitante y donde se	Texto Libre	Todos

		suele desarrollar la investigación. Sirve como clave secundaria para unir con el resto de tablas		
ORGANIZATIONS	organization_legal_name	Nombre fiscal de la institución o centro de investigación que gestiona la ayuda del solicitante y donde se suele desarrollar la investigación	Texto Libre	Todos
ORGANIZATIONS	organization_vat	Número VAT (identificación fiscal) del centro de investigación que aparece como cliente en la factura. Puede contener errores	Texto Libre	Todos
ORGANIZATIONS	organization_address	Dirección postal del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_fiscal_address	Dirección fiscal del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_city	Ciudad del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_province	Provincia del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_fiscal_province	Provincia fiscal del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_community_aut	Comunidad autónoma del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_country	País del centro de investigación donde se desarrolla la ayuda	Texto Libre	Todos
ORGANIZATIONS	organization_postcode	Código postal del centro de investigación donde se desarrolla la ayuda	Numéricos	Todos
ORGANIZATIONS	organization_primary_contact	Contacto suministrado por el centro como canal de comunicación	Texto Libre	Todos
ORGANIZATIONS	organization_parent_company	Consorcio o institución al que pertenece el centro de investigación donde se desarrolla el proyecto.	Texto Libre	Todos
ORGANIZATIONS	organization_type	Tipo de organización o centro de investigación donde se desarrolla la ayuda	Lista (único valor permitido)	Mixed Center;Research Center;University;Hospital;Other
ORGANIZATIONS	organization_status	Tipo de organización o centro de investigación donde se desarrolla la ayuda en cuanto al expediente que tenemos de ese centro en la Fundación. Si el expediente no está validado el estado será "Pending Validation"	Lista (único valor permitido)	Pending Validation;Grant Research Center;Homologated Research Center - NOT VIEW EXTERNAL;Centros Empresas Evaluadores - NOT VIEW EXTERNAL;Center Program Research Center;INTERNAL;NOT VALID
ORGANIZATIONS	organization_legal_status	Tipo de entidad donde se desarrolla el proyecto de investigación, si es una entidad pública o privada y si tiene o no ánimo de lucro	Lista (único valor permitido)	Not for Profit;For Profit;Private Entity;Public Entity
BUDGET	project_id	Código identificativo del proyecto caracterizado por 5 letras correspondientes al tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante	Texto Libre	Todos
BUDGET	budget_status	Estado de la justificación económica en la que se presenta el presupuesto para el año de la ayuda que empieza. Ejemplo: "Active" significa que la actividad de justificación está activa, es decir, está programada; "Completed" significa que la actividad está rellenada y validada por la fundación; "Submitted" es que se ha presentado; "Revisions Requested" significa que, una vez presentada, la Fundación ha detectado fallos que deben corregirse; "Under Review" significa que la Fundación lo está revisando	Lista (único valor permitido)	Active;Completed;Revisions Requested;Submitted;Under Review
BUDGET	reporting_year	Año de la ayuda en el que se reporta la factura. Ejemplo: Si se reporta la factura	Texto Libre	Todos

		durante la justificación económica pasado el primer año, este valor será "Y1"		
BUDGET	user_name	Nombre y primer apellido del solicitante de la ayuda	Texto Libre	Todos
BUDGET	starting_evaluation_period	Fecha de inicio del período que abarca la justificación económica	Fecha	Todos
BUDGET	end_evaluation_period	Fecha del final del período que abarca la justificación económica	Fecha	Todos
BUDGET	budget_personnel_proposal	Presupuesto destinado a personal propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_expendable_material_proposal	Presupuesto destinado a fungibles propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_equipment_proposal	Presupuesto destinado a equipamientos propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_travel_expenses_proposal	Presupuesto destinado a viajes propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_technical_services_proposal	Presupuesto destinado a servicios técnicos propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_others_proposal	Presupuesto destinado a otros conceptos propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	total_budget_proposal	Presupuesto total propuesto para el año de la ayuda que empieza. Esta propuesta se hace en la solicitud de la ayuda, es decir, se hace antes de la adjudicación o desembolso de ninguna cantía de dinero.	Numéricos	Todos
BUDGET	budget_personnel_current	Presupuesto destinado a personal propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	budget_expendable_material_current	Presupuesto destinado a fungibles propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	budget_equipment_current	Presupuesto destinado a equipamiento propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	budget_travel_expenses_current	Presupuesto destinado a viajes propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	budget_technical_services_current	Presupuesto destinado a servicios técnicos propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	budget_others_current	Presupuesto destinado a otros conceptos propuesto para el año de la ayuda que empieza sumado al remanente de la misma partida del año anterior	Numéricos	Todos
BUDGET	total_budget_current	Presupuesto total propuesto para el año de la ayuda que empieza sumado al remanente total del año anterior	Numéricos	Todos
BUDGET	budget_personnel_justified	Presupuesto destinado a personal justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos

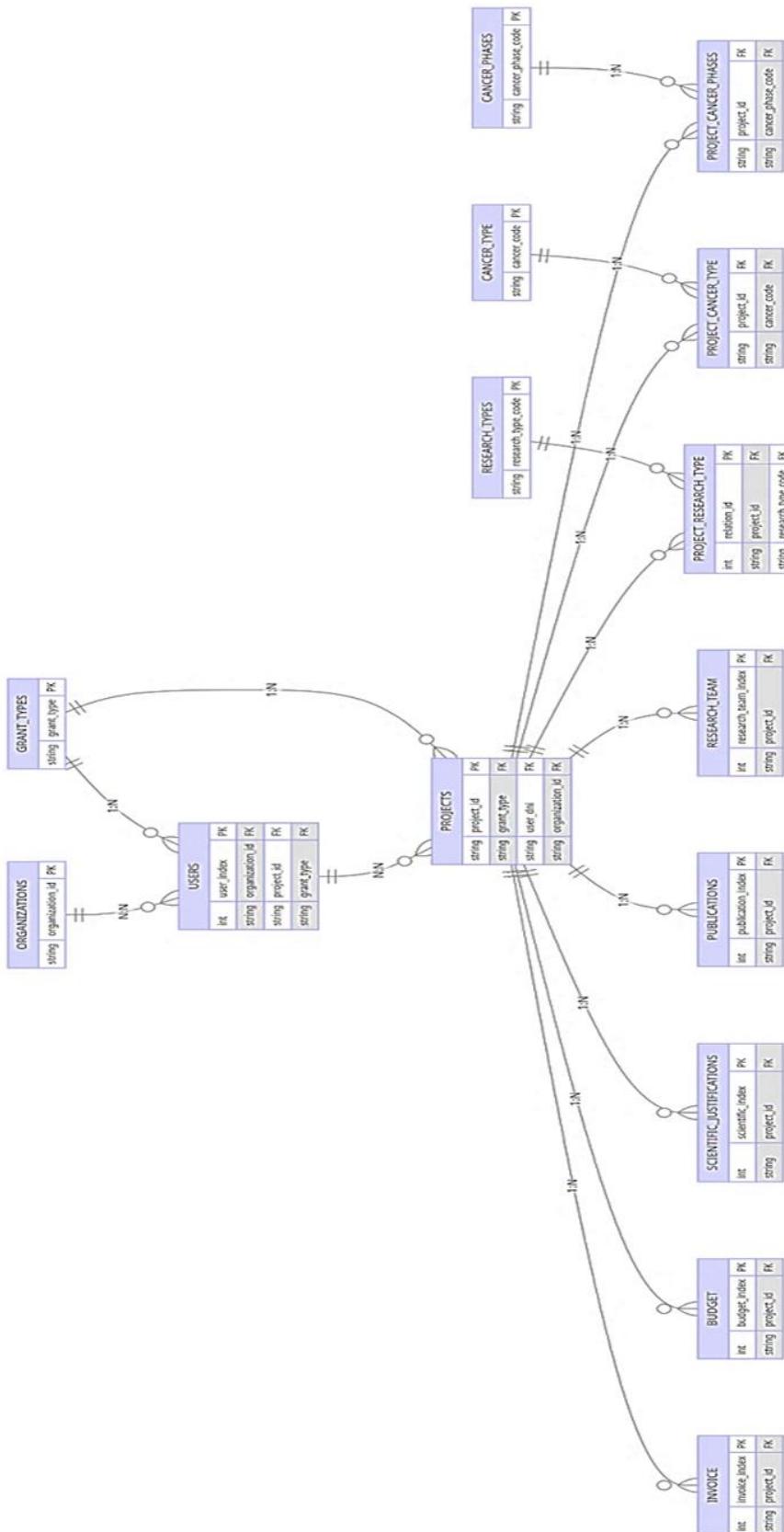
BUDGET	<code>budget_expendable_material_justified</code>	Presupuesto destinado a fungibles justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
BUDGET	<code>budget_equipment_justified</code>	Presupuesto destinado a equipamiento justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
BUDGET	<code>budget_travel_expenses_justified</code>	Presupuesto destinado a viajes justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
BUDGET	<code>budget_technical_services_justified</code>	Presupuesto destinado a servicios técnicos justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
BUDGET	<code>budget_others_justified</code>	Presupuesto destinado a otros conceptos justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
BUDGET	<code>total_budget_justified</code>	Presupuesto total justificado a través de facturas para el año de la ayuda que acaba de terminar y para el que se realiza la justificación económica	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	<code>project_id</code>	Código identificativo del proyecto caracterizado por 5 letras correspondientes al tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	<code>description_scientific_just</code>	Descripción del seguimiento científico. Suele indicar a que mensualidad, anualidad y si se trata de un seguimiento intermedio o final	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	<code>type_scientific_just</code>	Tipo de justificación científica. Hace referencia a si es una justificación anual ("Annual Report"), es decir, si es un seguimiento tras un año concreto de la ayuda. También hay seguimientos de menor tiempo como semestrales ("Progress Report"). También hay seguimientos externos para ayudas de grupos coordinados ("External Report"). Por último, los seguimientos que se hacen al final del período que abarca la ayuda y que suelen ser los más completos y que más información contienen ("Final Report").	Lista (único valor permitido)	Annual Report;Progress Report;External Report;Final Report
SCIENTIFIC JUSTIFICATIONS	<code>status_scientific_just</code>	Estado de la justificación científica en la que se presenta los resultados para el año que acaba de transcurrir. Ejemplo: "In Progress" significa que la actividad de justificación está en progreso, es decir, está programada; "Completed" significa que la actividad está rellenada y validada por la fundación; "Submitted" es que se ha presentado; "Revisions Requested" significa que, una vez presentada, la Fundación ha detectado fallos que deben corregirse; "Under Review" significa que la Fundación lo está revisando	Lista (único valor permitido)	In Progress;Completed ;Revisions Requested;Submitted;Under Review
SCIENTIFIC JUSTIFICATIONS	<code>deadline_scientific_just</code>	Fecha límite que tiene el solicitante para presentar la documentación requerida para la justificación científica	Fecha	Todos
SCIENTIFIC JUSTIFICATIONS	<code>starting_evaluation_period</code>	Fecha de inicio del período que abarca la justificación económica	Fecha	Todos
SCIENTIFIC JUSTIFICATIONS	<code>end_evaluation_period</code>	Fecha del final del período que abarca la justificación económica	Fecha	Todos
SCIENTIFIC JUSTIFICATIONS	<code>user_observation_scientific_just</code>	Observaciones aportadas por el usuario en cuanto a la justificación científica presentada	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	<code>scientific_just_evaluation_comments</code>	Comentarios incluidos por los gestores de la Fundación valorando los resultados y la justificación científica presentada	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	<code>scientific_desviacion</code>	Valoración que hacen los gestores de la Fundación donde se indica el grado de desviación de los resultados del proyecto respecto al plan de trabajo original, de acuerdo con la información aportada en la justificación científica	Lista (único valor permitido)	Desviación Alta;Desviación Media;Sin desviación

SCIENTIFIC JUSTIFICATIONS	patient_impact	Respuesta a la pregunta ¿El proyecto ha tenido impacto directo en pacientes? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual.	Lista (único valor permitido)	Yes;No;Not applicable
SCIENTIFIC JUSTIFICATIONS	patient_impact_description	Respuesta a la pregunta ¿El proyecto ha tenido impacto directo en pacientes? Pero especificando de qué tipo de una lista. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual.	Lista (único valor permitido)	Diagnosis;Survival;Treatment;Life quality
SCIENTIFIC JUSTIFICATIONS	additional_funding_impact	Respuesta a la pregunta ¿Has obtenido financiación adicional gracias a los resultados de este proyecto? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. Esta pregunta también se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Lista (único valor permitido)	Yes;No
SCIENTIFIC JUSTIFICATIONS	additional_funding_impact_origin	Respuesta a la pregunta ¿Has obtenido financiación adicional gracias a los resultados de este proyecto? Pero indicando el origen de la financiación adicional, siendo Nacional, o Internacional o ambos. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. Esta pregunta también se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Atributos en Lista (varios valores por celda)	International;National
SCIENTIFIC JUSTIFICATIONS	additional_funding_impact_grant	Respuesta a la pregunta ¿Has obtenido financiación adicional gracias a los resultados de este proyecto? Pero indicando el nombre de las ayudas, becas o financiación conseguida. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. Esta pregunta también se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	additional_funding_impact_amount	Respuesta a la pregunta ¿Has obtenido financiación adicional gracias a los resultados de este proyecto? Pero indicando la cantidad de financiación adicional conseguida. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. Esta pregunta también se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	people_hired	Número de personas contratadas durante el período de la ayuda. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. En algunos casos, en años diferentes siguen indicando que han contratado a un trabajador pero se refieren al mismo trabajador que el año anterior, por lo que a veces la suma de este campo puede proporcionar valores erróneos de personas contratadas.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	women_hired	Respecto al número de personas contratadas (ver campo people hired), respuesta a la pregunta ¿Cuántas de ellas son mujeres? Este campo se utiliza como encabezado en GMS para separar los apartados, no contiene ninguna información.	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	women_posdoc_hired	Respecto al número de personas contratadas (ver campo people hired), respuesta a la pregunta ¿Cuántas de ellas son mujeres con perfil posdoctoral? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. En algunos casos, en años diferentes siguen indicando que han contratado a un trabajador pero se refieren al mismo trabajador que el año anterior, por lo que a veces la suma de este campo puede	Numéricos	Todos

		proporcionar valores erróneos de personas contratadas.		
SCIENTIFIC JUSTIFICATIONS	women_predoc_toral_hired	Respecto al número de personas contratadas (ver campo people hired), respuesta a la pregunta ¿Cuántas de ellas son mujeres con perfil predoctoral? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. En algunos casos, en años diferentes siguen indicando que han contratado a un trabajador pero se refieren al mismo trabajador que el año anterior, por lo que a veces la suma de este campo puede proporcionar valores erróneos de personas contratadas.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	women_senior_posdoc_hired	Respecto al número de personas contratadas (ver campo people hired), respuesta a la pregunta ¿Cuántas de ellas son mujeres con perfil senior posdoctoral (investigadora de plantilla)? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. En algunos casos, en años diferentes siguen indicando que han contratado a un trabajador pero se refieren al mismo trabajador que el año anterior, por lo que a veces la suma de este campo puede proporcionar valores erróneos de personas contratadas.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	women_technician_hired	Respecto al número de personas contratadas (ver campo people hired), respuesta a la pregunta ¿Cuántas de ellas son mujeres con perfil de técnico? Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. En algunos casos, en años diferentes siguen indicando que han contratado a un trabajador pero se refieren al mismo trabajador que el año anterior, por lo que a veces la suma de este campo puede proporcionar valores erróneos de personas contratadas.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	people_project	Número de personas trabajando directa o indirectamente en el proyecto. Esta pregunta se les hace cada año que dura la ayuda, es decir, cada vez que presentan una justificación científica anual. La suma de este campo puede proporcionar valores erróneos de personas que participan en el proyecto, ya que se suman varias veces el mismo número de personas	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	clinical_impact	Respecto al impacto en pacientes, respuesta a la pregunta ¿Su proyecto ha influido en la elaboración de políticas, ha mejorado protocolos o ha generado orientaciones clínicas sobre el cáncer? Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Atributos en Lista (varios valores por celda)	Policy making; Protocols (basic/translational science); Cancer clinical guidances; Any of them; Other
SCIENTIFIC JUSTIFICATIONS	clinical_impact_others	Respecto al impacto en pacientes, respuesta a la pregunta ¿Su proyecto ha influido en la elaboración de políticas, ha mejorado protocolos o ha generado orientaciones clínicas sobre el cáncer? Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Texto Libre	Todos
SCIENTIFIC JUSTIFICATIONS	innovation_impact	Respecto al impacto en innovación, respuesta a la pregunta "Gracias a la financiación por parte de la FC AECC: Se ha creado una empresa, se han establecido colaboraciones con entidades privadas, se han establecido colaboraciones de investigación nacionales o internacionales". Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Atributos en Lista (varios valores por celda)	A technology-based company has been set up; Other collaboration networks with private entities have been established; Other national or international research networks have been

				Established; Not applicable
SCIENTIFIC JUSTIFICATIONS	project_executed	Respuesta a la pregunta: Dentro del marco de su proyecto ¿Qué porcentaje de ejecución de los objetivos del proyecto se ha alcanzado? Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Numéricos	Todos
SCIENTIFIC JUSTIFICATIONS	award_impact	Respuesta a la pregunta: ¿Ha recibido algún premio o reconocimiento tras la obtención de la beca FC AECC? Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Lista (único valor permitido)	Yes;No
SCIENTIFIC JUSTIFICATIONS	number_patients_enrolled	Número de pacientes inscritos en el ensayo clínico. Esta pregunta se les hace tras el último año de la ayuda, es decir, cuando presentan la justificación científica final.	Numéricos	Todos
PUBLICATIONS	project_id	Código identificativo del proyecto caracterizado por 5 letras correspondientes al tipo de ayuda, seguido de dos números indicando el año de la convocatoria (Ej: 22 para 2022), y seguido de 4 números identificativos de la solicitud y las primeras 4 letras del primer apellido del solicitante	Texto Libre	Todos
PUBLICATIONS	publication_title	Título de la publicación científica	Texto Libre	Todos
PUBLICATIONS	publication_abstract	Abstract (resumen) de la publicación científica	Texto Libre	Todos
PUBLICATIONS	publication_first_author	Primer autor de la publicación científica (No tiene porqué coincidir con el beneficiario o el investigador principal del proyecto)	Texto Libre	Todos
PUBLICATIONS	publication_type	Tipo de publicación científica (review, artículo de investigación, libro, capítulo de libro, contribución a un trabajo colectivo, etc.)	Lista (único valor permitido)	--Select-- ;Review;Research Article;Book;Book Chapter;Contribution to Collective Work
PUBLICATIONS	journal	Nombre de la revista donde se ha publicado el artículo científico	Texto Libre	Todos
PUBLICATIONS	publication_date	Fecha de publicación del artículo científico	Fecha	Todos
PUBLICATIONS	publication_internal_status	Estado interno de la actividad en la plataforma. Indica si el gestor de la Fundación ha aprobado el registro de esta publicación (Approved), si el investigador simplemente ha presentado el registro pero está pendiente de aprobación (Submitted) o si por el contrario el investigador solo ha creado el registro pero no lo ha presentado (Draft)	Lista (único valor permitido)	Approved;Submitted;Draft
PUBLICATIONS	journal_impact_factor	Factor de Impacto de la revista donde se publica el artículo científico. A mayor valor, indica mayor relevancia de la revista científica, y por lo tanto el artículo se considera que tendrá mayor calidad e interés	Texto Libre	Todos
PUBLICATIONS	journal_quartile	Indica el cuartil al que pertenece la revista científica. Es una medida de posición de una revista en relación con todas las de su área. Si dividimos en 4 partes iguales un listado de revistas ordenadas de mayor a menor índice de impacto, cada una de estas partes será un cuartil. Está relacionada con la relevancia y la importancia de la revista en su sector.	Atributos en Lista (varios valores por celda)	1;2;3;4
PUBLICATIONS	publication_year	Año de publicación del artículo científico	Numéricos	Todos
PUBLICATIONS	publication_doi	Es el código de identificación digital (DOI) o de identificación de una publicación impresa (ISBN/ISSN) que ayuda a identificar inequivocamente una publicación científica.	Texto Libre	Todos
PUBLICATIONS	publication_status	Estado de publicación del artículo científico, es decir, si se encuentra publicado, aceptado (pero no publicado aún), In press (pendiente de impresión), en revisión, o en preparación.	Atributos en Lista (varios valores por celda)	--Select-- ;Published;Accepted;In Press;In Review;In Preparation
PUBLICATIONS	publication_pubmed	Es el código de una publicación científica dentro de la base de datos PubMed que ayuda a identificar inequivocadamente una publicación científica.	Numéricos	Todos

12. ANEXO II: MODELO LÓGICO RELACIONAL



13. ANEXO III: ENCUESTA VALIDACIÓN DE HIPÓTESIS

¡Queremos conocer tu opinión!

* Indica que la pregunta es obligatoria

<https://docs.google.com/forms/d/1YifnlFfu30t3DE-BbLjAK8tNx29scma79qz8G2lilio/edit>



SYNERATECH

1. Nombre: *

2. Profesión *

3. Empresa/ Organización: *

4. ¿Ha considerado alguna vez contratar servicios externos para analizar y mejorar * la eficacia de su centro de investigación?

Marca solo un óvalo.

Sí

No

5. ¿Qué retos enfrenta en su centro de investigación que podría resolver mediante * la correcta gestión de sus datos?

6. ¿Estaría dispuesto a invertir en servicios de consultoría para mejorar sus procesos y eficacia? *

Marca solo un óvalo.

 Sí No

7. ¿Realiza un análisis de datos para la toma de decisiones en su empresa, su centro de investigación o a la hora de decidir su próxima inversión? *

Marca solo un óvalo.

 Sí No

8. ¿Cree que una mejor gestión y explotación de datos podría mejorar sus decisiones? *

Marca solo un óvalo.

 Sí No

9. ¿Estaría dispuesto a adoptar servicios de gobernanza de datos para mejorar sus * procesos de toma de decisiones?

Marca solo un óvalo.

Sí

No

Ya tenemos implementado políticas de gobernanza de datos en el centro de investigación / empresa

Si eres un inversor, por favor, contesta las siguientes preguntas.

En caso contrario puedes pasar a la siguiente pagina.

10. ¿Cuál es su nivel de interés en invertir en proyectos de investigación en salud y tecnología?

Marca solo un óvalo.

Sin interés

Poco interés

Bastante interés

Mucho interés

11. ¿Qué herramientas o información adicional cree que necesitaría para tomar decisiones de inversión más informadas en estos campos?

12. ¿Estaría dispuesto a aumentar su capital invertido si tuviera acceso a herramientas que mejoraran su comprensión de los proyectos de investigación y su seguimiento?

Marca solo un óvalo.

- Sí
 No
 Tal vez

Si trabajas en un centro de investigación público, por favor, contesta las siguientes preguntas.

En caso contrario puedes pasar a la siguiente pagina.

13. ¿Considera que la colaboración entre centros de investigación es importante para el avance de la ciencia?

Marca solo un óvalo.

- Sí
 No
 Tal vez

14. ¿Ha experimentado algún problema en la colaboración con otros centros de investigación debido a la falta de estandarización de datos?

Marca solo un óvalo.

- Sí
 No

15. ¿Estaría dispuesto a participar en una red que facilite la colaboración y la estandarización de datos?

Marca solo un óvalo.

Sí

No

16. ¿En qué medida cree que la colaboración y la identificación de sinergias entre proyectos de investigación mejorarían la efectividad de las investigaciones en su centro?

Marca solo un óvalo.

0 1 2 3 4 5

Nad Mucho

¡Muchas gracias por su colaboración!

17. ¿Desea ser contactado por el equipo de Syneratech para conocer en más detalle los servicios que ofrecemos? *

Marca solo un óvalo.

Sí

No

18. Correo electrónico de contacto:

19. Teléfono de contacto:



SYNERATECH



SYNERATECH

Resumen Ejecutivo

En Syneratech buscamos resolver la falta de conocimiento y experiencia en el campo de la explotación de datos del sector de la investigación sanitaria, así como la dificultad para optimizar recursos e identificar sinergias en el sector, ofreciendo asistencia personalizada, mediante la aplicación de herramientas de Business Intelligence.

En 2019 la pandemia de COVID-19 azotó al mundo entero y desafió nuestro sistema de salud y economía, y en consecuencia, ha reconfigurado nuestras prioridades y enfoques en muchos ámbitos, incluyendo el empresarial.

Los mercados no son ajenos a estas circunstancias y en los últimos años se apreció un incremento sin precedentes de las inversiones de capital de riesgo en el sector salud.

Oportunidad de negocio

Nosotros te ofrecemos la seguridad en tu toma de decisiones.



Actualmente nos encontramos un sector de la salud donde la inversión está intervenida por varios agentes de índole privada en mayor medida y también pública.

A nivel tecnología y desarrollo de sistemas, es un sector con muchas carencias que necesita de un gran componente innovador que facilite el desarrollo de nuevas tecnologías.

Por todo ello, nace la solución de Syneratech con el fin de aportar valor al sector y conseguir que las inversiones lleguen a los proyectos de investigación y a paliar la fuga de talentos.

Situación del mercado

El sector salud está en crecimiento continuo.

Gracias a soluciones de Business Intelligence, se pueden tratar los datos para obtener análisis prescriptivos y predictivos que ayudan en la toma de decisiones estratégicas.

Estamos en un momento de transformación digital, donde muchas empresas y sectores ven necesario el cambio hacia la digitalización.

Visión

Convertirnos en referente para la clasificación de proyectos en el sector salud, para la toma de decisión en la asignación de recursos clave de proyectos de investigación e innovación.

Misión

Proveer servicios de consultoría mediante herramientas de Business Intelligence al sector de la salud, para proporcionarles soluciones en la valorización de los proyectos.

Objetivos

Gestionar acuerdos de colaboración con las principales entidades del sector.

Fidelizar a nuestros clientes mediante acuerdos de colaboración a largo plazo y convertirnos en un servicio esencial en su gestión.

Conseguir un crecimiento sostenido a partir del 2º año.

Crear una cartera de clientes, basada en una red de colaboración, mediante la asistencia a los principales congresos y ferias, para captar a nuestros clientes potenciales.

Ofertar servicios de alto valor añadido.

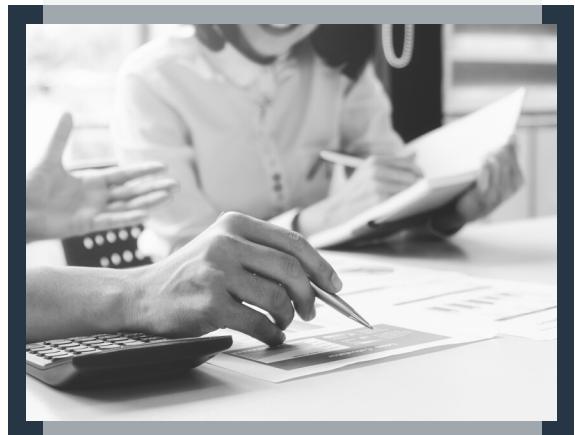
Utilizar estándares de calidad en prestación de servicios como la ISO 9001 y gestión de proyectos como PMBOK o ITIL.

Control de los gastos estructurales.

Nuestros Servicios

Análisis de la calidad del dato

Establecer protocolos que garanticen la veracidad del dato.

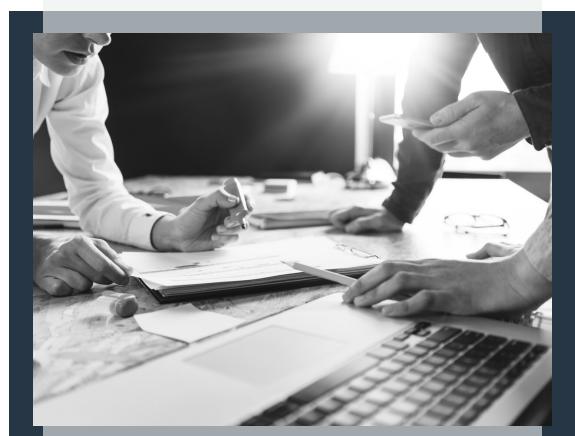


Construcción de base de datos

Escalables y con alta flexibilidad en la explotación de sus datos.

Análisis financieros de los proyectos

Extracción de información de facturas y gastos asociados a cada proyecto para identificar patrones de gastos y optimizar recursos.



Análisis de sinergias

Identificación de colaboraciones potenciales entre investigadores y la identificación de áreas de interés comunes.

Análisis de estadísticas de la base de datos

Extracción y análisis de estadísticas relevantes de la base de datos para obtener información útil y facilitar la toma de decisiones.



Propuesta de valor

Syneratech ayuda en la valorización de proyectos de investigación enfocados en el sector salud, es una solución tecnológica basada en Business Intelligence, que va de la mano de servicios de consultoría totalmente personalizados.

Ofrecer servicios de consultoría relacionados con la gobernanza y explotación de datos surgidos en el ámbito de proyectos innovadores.

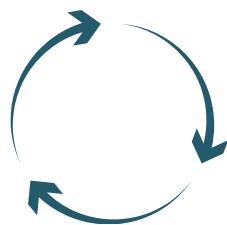
Nuestro valor diferencial radica en la integración de nuestra amplia experiencia en el ecosistema innovador y en la explotación de datos en dicho sector, lo que nos permite ayudar a nuestros clientes a:

Maximizar el potencial de sus proyectos innovadores, optimizando su gestión financiera y promoviendo la colaboración estratégica.

Alcanzar una visión clara y detallada de las oportunidades de inversión, así como la capacidad de identificar posibles colaboraciones y sinergias que impulsen el éxito de sus iniciativas.

Beneficios

Realizar una Estrategia de Gobierno de dato generalista y protocolos para implantar una mínima política de calidad del dato y asegurar la veracidad del dato.



Establecer relaciones entre proyectos, trabajadores y centros de trabajo, poder hilar y dar sentido a una gran cantidad de datos y como consecuencia optimizar procesos y facilitar la toma de decisiones.

Crear procesos que faciliten el análisis y enfoque a diferentes áreas, definiendo KPIs y métricas automáticas.

Producto mínimo viable

Para el lanzamiento de nuestro proyecto, hemos sido contratados por la Fundación Científica AECC para ayudarles en la gestión de recursos, para facilitar la toma de decisión de forma objetiva y metódica, creando estándares comunes para valorar con los mismos criterios los proyectos de investigación.

Se han integrado las siguientes tecnologías para ofrecer diversas soluciones a la Fundación:



Instancia de una máquina virtual en Google Cloud con imagen Docker de PostgreSQL.



Procesamiento del lenguaje natural (NLP) traducido a SQL.



Para los procesos ETL se han desarrollado scripts en Notebooks de Google Collab.



Representación gráfica mediante Neo4j para representar relaciones entre investigaciones por medio de los proyectos que han colaborado, clasificados por tipo de cáncer.

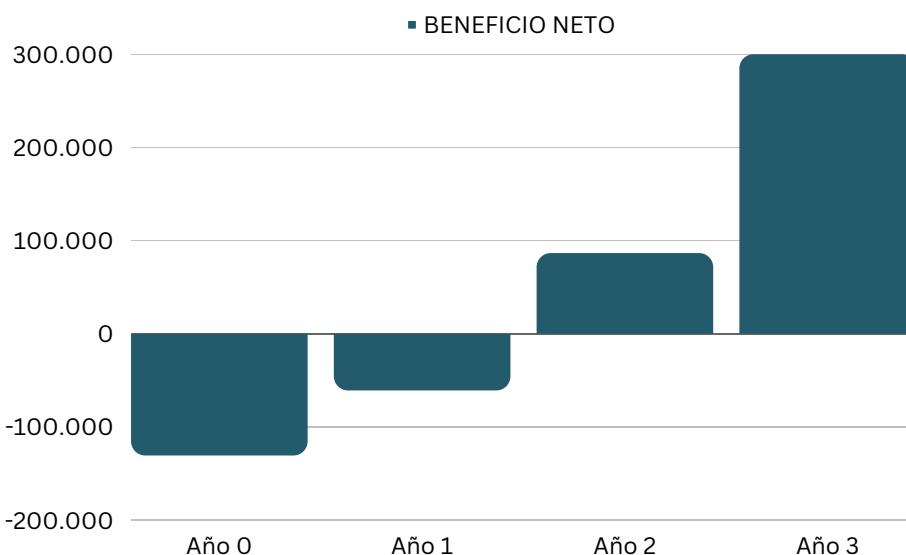
Mediante Power BI se han creado los Dashboard necesarios para facilitar a la AECC la toma de decisiones, identificación de problemas y creación de soluciones.



Rentabilidad

Para el inicio del proyecto se cuenta con una aportación inicial de los socios y se ha solicitado una financiación externa de 50.000 euros durante el primer año.

El PAYBACK de 2,08 años señala que la inversión inicial se recuperará en un plazo relativamente corto. La combinación con un VAN positivo fortalece la viabilidad financiera del proyecto. La inversión parece ser rentable y podría generar rendimientos significativos.



PAY-BACK

A partir del primer semestre del tercer año.



TIR

35,54%



TIRM

32,66%

La TIR del 35,54% indica que este proyecto ofrece un rendimiento significativamente superior a las tasas de interés convencionales.

Asimismo, la TIRM del 32,66% sugiere que, si los costos de financiamiento aumentan a lo largo del tiempo, el proyecto aún debería ser rentable, aunque a un ritmo ligeramente más moderado.

Flujo de Caja operativo

