



UNIVERSIDAD DE EXTREMADURA

CONCURSO ENTRE ACREDITADOS
ACCESO AL CUERPO DE PROFESORES
TITULARES DE UNIVERSIDAD

—

PROYECTO INVESTIGADOR

DF3614 - Profesor Titular de Universidad

—

Marino Linaje Trigueros

Cáceres, 4 de agosto de 2020



UNIVERSIDAD DE EXTREMADURA

CONCURSO ENTRE ACREDITADOS
ACCESO AL CUERPO DE PROFESORES
TITULARES DE UNIVERSIDAD

—

BOE nº 126, 6 de mayo de 2020

Código del concurso: 2019/B/045

—

Departamento: Tecnología Computadores y de las Comunicaciones

Área: Arquitectura y Tecnología de Computadores

Escuela Politécnica

Presidente titular: Juan Antonio Gómez Pulido

Secretario titular: Miguel Ángel Vega Rodríguez

Vocal 1º titular: Pilar Bachiller Burgos

Vocal 2º titular: Rosa María Pérez Utrero

Vocal 3º titular: Pablo Bustos García de Castro

—

Presidente suplente: Francisco Fernández de Vega

Secretario suplente: Juan Carlos Díaz Martín

Vocal 1º suplente: Antonio Plaza Miguel

Vocal 2º suplente: Javier Plaza Miguel

Vocal 3º suplente: Julio Ballesteros Rubio

ÍNDICE DE SECCIONES

Índice de figuras	5
Índice de tablas.....	6
Índice de siglas y acrónimos.....	7
1 Introducción.....	8
1.1 Organización.....	9
2 Contexto.....	10
2.1 Organización, marco legal y normativo.....	10
2.2 Fuentes de financiación.....	16
2.3 Personal y del desarrollo propuesto	28
3 Desarrollo del proyecto de investigación	32
3.1 Resumen.....	32
3.2 Introducción.....	35
3.3 Objetivos	42
3.4 Antecedentes.....	44
3.5 Experiencia previa	50
3.6 Metodología y actividades	53
3.7 Cronograma.....	56
3.8 Presupuesto	58
3.9 Transferencia y difusión de resultados	59
4 Resumen, conclusiones y agradecimientos	61
Bibliografía	63

ÍNDICE DE FIGURAS

Ilustración 1. Propuestas presentadas respecto a las financiadas por área temática en España	20
Ilustración 2. Propuestas presentadas respecto a las financiadas por área temática en Extremadura	20
Ilustración 3. Arriba - Clasificación de organizaciones extremeñas por captación de fondos en el programa H2020. Abajo - Clasificación de organizaciones extremeñas por partición en el programa H2020. Fuente: H2020 dashboard	21
Ilustración 4. Estadística sobre gasto en actividades de I+D en el sector empresas porcentual por Comunidades autónomas. Fuente: INE 2020 (datos de 2017).....	25
Ilustración 5. Gasto en I+D por comunidades incluyendo gastos de personal (2017). Fuente: INE	26
Ilustración 6. Estructura preliminar del programa Horizonte Europa. Fuente: UE	27
Ilustración 7. Distribución en capas de computación	33
Ilustración 8. Relación de costes entre el almacenamiento en Cloud y el almacenamiento Mist/Edge propuesto	40
Ilustración 9. Relación entre probabilidad de fallo (hardware) y número de réplicas	41
Ilustración 10. Cronograma propuesto	57

ÍNDICE DE TABLAS

Tabla 1. Distribución de la financiación planificada en el Plan estatal de investigación científica y técnica y de innovación (2013-2016). Fuente: Ministerio de Ciencia e Innovación.....	23
Tabla 2. Previsión de financiación de la I+D+i en Extremadura para el periodo 2017-2020. Fuente:	24
Tabla 3. Tiempo requerido para llenar una memoria de 8GB	39
Tabla 4. Requisitos del framework objetivo.....	42
Tabla 5. Resumen comparativo del estado del arte con los requisitos fijados para este proyecto	48
Tabla 6. Presupuesto desglosado en partidas para la fase A	59
Tabla 7. Presupuesto desglosado en partidas para la fase B.....	59

ÍNDICE DE SIGLAS Y ACRÓNIMOS

Se usan siglas y acrónimos en este documento para mejorar la velocidad de lectura de este. Las siglas y acrónimos usados ordenados por orden alfabético de la sigla o acrónimo son los siguientes:

BOE: Boletín Oficial del Estado (España)

DOE: Diario Oficial de Extremadura

EPCC: Escuela Politécnica (situada en Cáceres)

KPI: Indicador clave de rendimiento (*Key Performance Indicator*)

RTOS: Sistema operativo en tiempo real (*Real Time Operating System*)

SGTRI: Servicio de Gestión y Transferencia de Resultados de la Investigación

UE: Unión Europea

UEx: Universidad de Extremadura

UbiCom: Computación Ubicua, término equivalente a Inteligencia Ambiental y Sistemas Pervasivos entre otros.

1 INTRODUCCIÓN

El presente documento tiene como objetivo dar cumplimiento a la normativa legal vigente que regula el acceso a los cuerpos docentes universitarios. En particular, con lo dispuesto en la resolución publicada en el BOE con número 126 (el día 6 de mayo de 2020 en su sección II.B., página 31432). En ella, se recogen los datos referentes al concurso con código 2019/B/045, incluyendo el número de plaza del concurso (DF3614), la contextualización de esta y la composición del tribunal.

Dicha plaza se contextualiza dentro del Departamento de Tecnología de los Computadores y las Comunicaciones de la Universidad de Extremadura (UEX) y está asociada al Área de Arquitectura y Tecnología de los Computadores. Concretamente la resolución especifica:

Docencia en fundamentos de computadores. Investigación relacionada con sistemas ubicuos (adquisición de datos, procesamiento de información, transmisión, almacenamiento, interacción y relacionados)

Este documento, centrado en el proyecto investigador, versa de este modo, sobre investigación en *Sistemas Ubicuos* (UbiCom). En concreto, sobre la adquisición de datos, procesamiento de información, transmisión, almacenamiento, interacción y relacionados en ellos.

Este proyecto docente se presenta en cumplimiento con la Ley Orgánica 4/2007 de 12 de abril, BOE de 13 de abril, por la que se modifica la Ley Orgánica 6/2001 de 21 de diciembre, BOE de 24 de diciembre, en su Artículo 62.4. Este proyecto, sigue la normativa de concursos para el acceso a plazas de los cuerpos docentes universitarios entre acreditados de la Universidad de Extremadura (17 de diciembre de 2008), modificada por Resolución del Rector de 10 de marzo de 2011 (DOE núm. 59 de 25 de marzo de 2011).

Debido a la delicada situación epidemiológica que estamos viviendo, el Rectorado de la Universidad de Extremadura también ha sacado distintas instrucciones que afectan, entre otras, a la presentación de este concurso. Este se realizará en la medida de lo posible presencialmente, aunque podría realizarse de manera no presencial en caso de necesidad.

La normativa existente no contempla cómo debe organizarse ni el documento referente al proyecto docente ni al investigador. Esto facilitan la reflexión libre sobre el contexto de la investigación, incluyendo sus estructuras, sistemas de financiación, etc.

La presentación y debate sobre este documento proporcionan también un marco que no siempre es fácil de conseguir como es que varios profesores, con larga experiencia docente, puedan aportar su visión sobre un tema muy concreto, como es la propuesta de investigación presentada. Así, se espera que, de estas discusiones, se

pueda incorporar la retroalimentación que el tribunal pueda hacer sobre la misma, para mejorarla con vistas a futuras convocatorias de financiación.

1.1 ORGANIZACIÓN

El proyecto investigador se ha organizado como sigue: en el capítulo 2 se contextualiza la propuesta con relación al marco legal y a las fuentes de financiación, internacionales, nacionales y regionales. El capítulo 3 se centra en el desarrollo del proyecto de investigación propuesto y que incluye los detalles más habituales en muchas convocatorias. Finalmente, el capítulo 4 intenta servir de resumen e intenta plantear conclusiones reflexivas.

2 CONTEXTO

Investigar supone aplicar la razón a la comprensión de la realidad externa, con el fin de identificarla, conocerla, comprenderla, adaptarla y/o mejorarla, entre otras. Se completa la labor del investigador mejorando el estado del conocimiento/arte, incrementando los conocimientos de la sociedad, publicando y difundiendo adecuadamente el conocimiento adquirido.

Así, la investigación se puede considerar como el medio básico para el progreso de la sociedad y proporciona las bases para la transferencia social del conocimiento. Por ello, es lógico que constituya una función esencial de las universidades.

Como uno de sus objetivos esenciales, la universidad asume el desarrollo de la investigación, la formación de investigadores y la difusión del conocimiento adquirido a la sociedad, incluidos otros investigadores en las diferentes ramas del conocimiento: artes y humanidades, ciencias y tecnologías. Esto debe hacerlo en las tres vertientes de la investigación, la básica, la aplicada y la de desarrollo, que podemos especificar como:

- La investigación básica (o fundamental), que consiste en una investigación original, novedosa, cuya finalidad es el progreso del conocimiento científico, sin necesidad de tener objetivos comerciales específicos fijados inicialmente.
- La investigación aplicada, que consiste en descubrir o aplicar conocimientos científicos que puedan realizarse/ejecutarse en productos, procesos, servicios... que son lo novedoso (no la investigación en sí).
- La investigación de desarrollo, que tiene como fin la aplicación de la investigación a productos, procesos, servicios... pero siempre dentro de hipótesis no corrientes, muchas veces contrarias al estado de la técnica.

El presente proyecto de investigación se enmarca entre el desarrollo y la aplicabilidad, principalmente la primera.

En esta sección principalmente, se contextualiza el marco actual de la investigación, desde la perspectiva de pertenencia a la Universidad de Extremadura. Estar en esta comunidad facilita el acceso a ciertas fuentes de financiación, a la vez que restringe el acceso a otras. Del mismo modo, establece un marco legal y sobre todo normativo característico.

2.1 ORGANIZACIÓN, MARCO LEGAL Y NORMATIVO

La universidad no fija la investigación que realiza su personal docente e investigador, siendo la libre investigación es un derecho y un deber de este personal. Los límites los marca el investigador aplicando la razón, e incluye temas como la seguridad, el alineamiento con los fines generales de cada universidad, la disponibilidad técnica y económica y el aprovechamiento de los medios y recursos disponibles o accesibles. Con lo que cuentan las universidades es con un marco

normativo para diferentes aspectos relacionados con la investigación, como veremos más adelante en esta sección.

La actual ley en vigor (LOMLOU), incide en el aspecto investigador de la universidad en España. En ella se especifica que la labor investigadora en la universidad puede ser integrada por grupos o institutos de investigación, así como los fines que debe perseguir la investigación en la universidad:

La universidad desarrollará una investigación de calidad y una gestión eficaz de la transferencia del conocimiento y la tecnología, con los objetivos de contribuir al avance del conocimiento y del desarrollo tecnológico, la innovación y la competitividad de las empresas, la mejora de la calidad de vida de la ciudadanía, el progreso económico y social y un desarrollo responsable equitativo y sostenible, así como garantizar el fomento y la consecución de la igualdad.

Así principalmente, la organización de la investigación en la universidad gira en torno a los grupos de y a los institutos universitarios de investigación. Dando otros entes soporte a la universidad como p.ej., los departamentos que se ocupan de la aprobación de la participación de sus miembros en contratos y convenios de investigación.

Por un lado, los grupos de investigación son agrupaciones de investigadores con una serie de líneas de investigación comunes o relacionadas. Cada investigador universitario en España solo puede pertenecer a un grupo de investigación.

Cada grupo puede definir sus líneas de investigación de manera libre y estas pueden ser genéricas o más específicas o cambiando junto al estado del arte o de la técnica. Las fronteras de la investigación no son inamovibles y las contribuciones científicas ayudan a desplazar esta frontera, hasta que es asumida como estado del arte o de la técnica.

Algunos grupos, debido al carácter básico y al largo recorrido requerido por sus líneas de investigación, tienen líneas estáticas o al menos muy constantes en el tiempo. Sin embargo, otros, debido a que su investigación es principalmente aplicativa y/o al estado muy cambiante del estado del arte/técnica en el contexto de sus líneas, varían estas de forma habitual, valiéndose de los conocimientos que van adquiriendo sus investigadores, mediante la participación en otros proyectos y el desarrollo de trabajos dirigidos entre otros.

Por otro lado, los institutos universitarios de investigación son agrupaciones de mayor tamaño. Podríamos definirlos como centros dentro de la organización universitaria dedicados a la investigación interdisciplinar, creados para reunir a investigadores y grupos de investigación. Su orientación es agrupar grupos e investigadores inmersos en investigaciones afines y/o complementarias, a fin de aunar esfuerzos y facilitar la cooperación entre los mismos, generando sinergias.

Estos centros están pensados para generar estructuras de investigación coherentes, y posibilitando infraestructuras de carácter científico difícilmente alcanzables por los grupos de investigación de manera independiente p.ej., la creación y/o acceso a infraestructuras de carácter científico difícilmente alcanzables por su coste de creación, uso o mantenimiento.

Además, tienen otros objetivos entre los que destacan: favorecer la cooperación con otros centros de I+D+i, tanto públicos como privados, y favorecer la atracción de talento investigador de otras entidades nacionales o supranacionales.

Es cada universidad la que se autogestiona y organiza internamente, para llevar el conjunto de labores relacionadas con la investigación de la manera que estima más correcta. La investigación conlleva labores dispares de entre las que listamos a continuación las más habituales:

- Búsqueda de potenciales fondos de investigación.
- Redacción científica y económica de proyectos.
- Creación de consorcios de investigación.
- Gestión de aspectos legales, entre los que se incluye la protección de datos y de resultados.
- Gestión del personal que trabaja en los proyectos, de los cuales existe un amplio número de figuras contractuales.
- Difusión de resultados en contextos científicos.
- Justificación de las ayudas recibidas.
- Gestión de cátedras de patrocinio.
- Difusión social de resultados y fomento de vocaciones científico-técnicas.

Dado que cada universidad se autoorganiza, en el resto del documento nos centraremos únicamente en la Universidad de Extremadura (UEx). Concretamente, la Junta de Extremadura ofrece un listado¹ de todos los grupos de investigación de la región extremeña, con un total de 231 grupos. La UEx aporta, con diferencia, el más amplio catálogo de grupos de investigación, con un total de 195.

La UEx tiene independencia organizativa con respecto al gobierno regional, aunque este último aporta fondos de investigación, que asigna directamente a los grupos de investigación en base a ciertos criterios competitivos.

Dentro de los órganos de gobierno de la UEx, el que tiene un papel investigador más predominante, es el Vicerrectorado de Investigación, Transferencia e Innovación, que cuenta con portal web propio². Este vicerrectorado tiene un amplio número de competencias fijadas en el DOE (*Resolución de 17 de diciembre de 2014*, núm. 1, 2/enero/2015). Por no nombrar todas, las agrupamos resumidamente a continuación en 4 grandes grupos:

¹ <https://secti.gobex.es/SECTI/publico/catalogo/catalogo.jsf>

² <http://investigalia.unex.es>

- Apoyo a la actividad investigadora
- Infraestructura científica
- Gestión de la investigación
- Valorización de la investigación

La UEx apuesta por el fomento de la investigación y, para ello, además de proporcionar servicios a los investigadores, se ocupa de activar las vocaciones entre jóvenes investigadores. Para ello cuenta con el *Plan de iniciación a la investigación, desarrollo tecnológico e innovación*³ que renueva, generalmente, anualmente y que permite a los jóvenes investigadores, entre otros, ser por primera vez investigadores responsables de un proyecto, aprendiendo a gestionar el mismo, así como a familiarizarse con la estructura de investigación de la UEx. Este plan anual suele incluir otras acciones adicionalmente p.ej., para fomentar las actividades formativas entre futuros investigadores, como becas y contratos, entre otros.

Uno de los servicios que depende de este Vicerrectorado, es el Servicio de Gestión y Transferencia de Resultados de la Investigación (SGTRI) denominado Oficina de Transferencia de Resultados de Investigación (OTRI) en otras universidades. En España, estas oficinas/servicios universitarios se agrupan y comparten inquietudes a través de la *Red de unidades de gestión de la investigación*⁴.

SGTRI es un servicio para la gestión de la actividad investigadora y la administración de los fondos generados por la UEx, que ejecuta la política definida por los órganos de gobierno competentes. Las funciones que le corresponden, según los Estatutos de la UEx, son principalmente las siguientes:

- Identificar y difundir la oferta científica y técnica de la Universidad.
- Establecer, facilitar y desarrollar las relaciones entre la Universidad y cualquier demandante o promotor de investigación científico-técnica, sea público o privado.
- Facilitar y gestionar la transferencia de los resultados de la investigación científico-técnica, contratando en nombre de la Universidad los correspondientes trabajos y efectuando por cuenta de los investigadores cuantos actos y gestiones fueran precisos.
- Gestionar los convenios, contratos y proyectos de investigación.
- Establecer y gestionar la base de datos de investigadores, Grupos de investigadores e investigación de la Universidad.
- Informar a los investigadores y Grupos de Investigación de las convocatorias públicas de financiación de proyectos, becas, infraestructuras y cuantas otras lleguen a su conocimiento.
- Gestionar los derechos de propiedad industrial procedentes de los resultados de la investigación desarrollada por la Universidad.

³ https://www.unex.es/investigacion/secciones/plan_iniciacion

⁴ <http://redugi.es/>

Desde la creación de estos estatutos han surgido nuevas necesidades relacionadas con la investigación, no contempladas inicialmente y, que han sido asumidas por este servicio. Entre otros, se ocupa de aspectos relacionados con la creación de empresas de base tecnológica surgidas en el seno de la UEx, también conocidas como spin-offs, y su relación con la UEx.

Dada la gran diversidad de grupos de investigación de UEx y quizás por el esfuerzo que requieren las diversas labores que realiza este servicio en relación con los recursos, principalmente humanos, de que dispone, hace que sea difícil que adapten la información a las necesidades de los investigadores. También que puedan prestar ayuda con la cada vez mayor cantidad de trámites administrativos que conlleva la aceptación de un proyecto de investigación.

Por ello, se ha visto la necesidad de crear figuras de enlace, denominadas gestores de proyecto, como puente entre los grupos de investigación y el SGTRI. Estas figuras existen en otras universidades también, aunque no siempre organizadas del mismo modo. Si bien el nombre de gestor de proyecto, parece que conlleva que es un personal que ayuda a la gestión de proyectos específicos, esta no es su única labor.

Además de facilitar a los investigadores la gestión administrativa de los proyectos, mediante el conocimiento de la documentación y su formalización dentro de la universidad, también colaboran en la búsqueda de fuentes de financiación para nuevos proyectos y la formalización de estos, así como la gestión de recursos humanos, procesos de concurso público relacionados con ofertas de trabajo de investigación.

Se entiende que, con todas estas labores, facilitan enormemente no solo la labor del investigador, sino también la propia labor del SGTRI, ya que, al estar muy familiarizados con los trámites administrativos, requieren menos ayuda por parte del SGTRI para su consecución.

Dando servicios tan diversos, y no disponiendo de personal propio especializado para todos ellos (legal, económico...), en muchas ocasiones SGTRI se apoya en otros órganos de la UEx, a los que consulta o deriva la realización de ciertas tareas a otros servicios transversales a toda la estructura organizativa de la UEx. Los investigadores también pueden solicitar el uso de los servicios de manera individual. Para todo ello, la UEx cuenta con servicios y unidades transversal a todos los objetivos de la universidad y que incluye muchos relativos a investigación. Entre estos servicios, está el de recursos humanos o el servicio jurídico entre muchos otros.

Otro órgano de la UEx con especial orientación hacia la investigación, y sobre todo hacia la difusión de conocimiento tanto dentro de la propia universidad como hacia la sociedad en general, es el servicio de difusión de la cultura científica de la UEx. Este servicio también cuenta con portal web⁵ propio. Es dependiente del gabinete de información y comunicación, y junto a otros servicios/oficinas con objetivos similares de otras universidades, se agrupan en la red de Unidades de Cultura

⁵ <http://culturacientifica.unex.es>

Científica y de la Innovación de la Fundación Española para la Ciencia y la Tecnología (FECYT).

Cabe destacar que este servicio es muy cercano, tanto con los investigadores como con los medios de comunicación, sirviendo de puente de unión entre estos. Entre sus funciones destacan:

- Comunicar los resultados de las actividades de investigación científica y desarrollo tecnológico que genera la UEx. P.ej., a través de notas de prensa, la revista *viceversa* o el programa de radio *tubo de ensayo*.
- Promover y organizar actividades de divulgación científica dirigidas a la sociedad en general y jóvenes de educación primaria, secundaria bachillerato y formación profesional, tales como la *Noche europea de los investigadores* o *Desayuna con la ciencia*.
- Asesorar y apoyar al investigador al comunicar temas relacionados con la ciencia y la tecnología.

En relación con las empresas, la UEx dispone de incubadoras donde acoger las de nueva creación asociadas a la universidad, además de un *Parque científico y tecnológico*⁶ donde se pueden consolidar, en un entorno favorable para su desarrollo. Este parque facilita además la cercanía física entre universidad y empresa, con el fin de que surjan sinergias. Esto fomenta la petición de proyectos en común a convocatorias de diferente índole, la contratación de servicios y otras interacciones como son las cátedras de patrocinio.

Así, podemos concluir que la UEx cuenta con una compleja estructura organizativa orientada a la investigación, tanto para investigar propiamente, como para gestionarla en muy diferentes vertientes. Parte de esta información se encuentra recogida en el portal de datos abiertos⁷ de la UEx. Creo que incluir más información de este modo, siendo fácilmente procesable por sistemas de computación, permitiría crear y encontrar más fácilmente sinergias entre grupos de investigación.

Los proyectos de investigación de planes o acciones competitivas internacionales, nacionales y regionales siguen sus propias pautas, y cuentan con normativa propia de la convocatoria o del organismo que las convoca.

Otro tipo de proyectos, los no competitivos, suponen la contratación de servicios de la UEx mediante convenios y contratos de investigación, aunque en ocasiones, pueden ser también de formación u otros tipos de transferencia de conocimiento. Este segundo tipo de proyectos se articula, en la mayoría de los casos, en torno a normativa estatal, siguiendo la normativa fijada por el artículo 83 de la Ley Orgánica de Universidades (6/2001, de 21 de diciembre).

Para otros aspectos de la investigación, la UEx cuenta con normativa propia, que se apoya en leyes, órdenes, decretos y normativa de otras entidades y organismos de

⁶ <http://fundecyt.es/>

⁷ <http://opendata.unex.es>

diferente ámbito. Por ejemplo, la UEx tiene normativa propia sobre la protección de resultados de investigación, que se apoya en leyes como la de patentes (BOE, Ley 11/1986, de 20 de marzo, núm. 73, Art. 6, de Patentes de Invención y Modelos de Utilidad) a la que añade características propias, tanto sobre el proceso de petición de patentes y modelos de utilidad por parte de los investigadores, como sobre la autoría y propiedad intelectual de estos.

De modo similar, la normativa de creación de empresas de base tecnológica establece las pautas para su creación, así como el marco de la relación entre la Universidad y la spin-off creada. Lógicamente, estas empresas tienen que constituirse lógicamente en base a la normativa autonómica y estatal vigente en cada momento, además de cumplir con la normativa económica en relación con sus actividades, dependiendo de la sede de la empresa.

Listamos, a continuación, la normativa principal de la UEx para regular ciertos aspectos relacionados con la investigación, a la que se puede acceder en detalle junto a otras normativas relacionadas en la web *investigalia*⁸:

- Normativas sobre gestión de personal de investigación:
 - Reglamento de régimen general de becarios de investigación.
 - Normativa para contratación o nombramiento de personal técnico de apoyo.
 - Normativa sobre becarios de investigación.
 - Normativas del personal investigador.
 - Acuerdo normativo para la contratación de personal científico e investigador.
 - Acuerdo normativo sobre personal con actividad investigadora.
- Normativa para celebrar contratos y convenios.
- Normativa sobre la protección de resultados de la investigación.
- Normativa creación de empresas de base tecnológica.
- Normativa de grupos de investigación.
- Normativa sobre la evaluación de grupos de investigación.
- Normativa sobre trabajo de grado.
- Normativa de creación de cátedras y aulas de patrocinio.

2.2 FUENTES DE FINANCIACIÓN

La financiación de la investigación en la universidad puede provenir de fuentes públicas, privadas o de una combinación de ambas.

La financiación privada está sujeta a menos normas que la pública y suele ser no competitiva. Aunque eso depende de la entidad o de la convocatoria concreta de

8

<http://investigalia.unex.es/#!/page36.do?link=oln266.redirect&acond12=es&rcond3.att2=74&kcond92.att3=258>

proyecto que la entidad financiadora privada quiera realizar. En ocasiones, no es extraño que se realicen contrataciones de servicios ofertados por la universidad que, para algunos servicios, dispone de precios públicos que publicita anualmente en los presupuestos. En estos casos, el proyecto puede no estar asociado a ningún tipo de convocatoria.

Sin embargo, las fuentes públicas de financiación sí se encuentran sujetas a una serie de normas más estrictas, con el fin de asegurar que el dinero público se gasta de la forma más eficientemente posible. Para ello, incluyen un régimen de concurrencia competitiva a los fondos económicos disponibles.

Atendiendo a las fuentes de financiación de la investigación, tanto en la UEx como en el resto de las universidades españolas, los fondos económicos provienen fundamentalmente, de programas de ayudas en convocatorias competitivas. Estos fondos pueden tener origen europeo, nacional, regional o, de manera puntual, provenir de la propia universidad, generalmente por un motivo muy específico.

Los investigadores presentan proyectos a estos programas de financiación, que incluyen, no sólo la idea de investigación sobre la que se desea trabajar y el objetivo perseguido, sino muchos otros aspectos para la consecución del proyecto. Entre los habituales, podemos destacar: el estado del arte/técnica existente en la actualidad, el equipo investigador que llevará a cabo el proyecto y la experiencia curricular de este sobre la materia tratada en el proyecto, una planificación y responsabilidad de la ejecución de éste y el presupuesto económico necesario para llevar a cabo el proyecto.

Las convocatorias de proyecto difieren unas de otras, pues son convocadas por organismos diversos, en ocasiones poco conexiones o sin directrices comunes. Esto dificulta a los investigadores la petición de proyectos, que no usan las mismas secciones, los mismos campos, ni las restricciones de tamaño en estos y que ni siquiera tienen un punto de acceso único para enviarlos, localizar convocatorias abiertas, etc. Así, no es extraño que pidan al equipo investigador incluir más datos respecto a los anteriormente mencionados, como publicación y difusión de resultados esperados, la metodología de trabajo que se piensa seguir para la gestión de éste, o la patentes y modelos de utilidad que se pretenden obtener del proyecto, por citar algunos de los requisitos.

Como ya hemos avanzado, los proyectos de investigación se consiguen financiar, generalmente, en un contexto de concurrencia competitiva. Así, las propuestas de proyecto no tienen asegurada a priori su financiación, y no la consiguen de manera inmediata tras la presentación de éste a una determinada convocatoria. Habitualmente, los proyectos son evaluados por expertos o comisiones de expertos que establecen la elegibilidad de la propuesta para ser financiados totalmente, parcialmente o ser completamente rechazados. La dependencia de la cantidad de fondos dotada para la convocatoria es determinante, por lo que incluso propuestas bien valoradas, pueden ser rechazadas si el número de propuestas válidas es alto en relación con los fondos disponibles para la convocatoria.

En la opinión del autor de este proyecto investigador, el apoyo de entes públicos a la investigación es importante y es el único que puede dar lugar a resultados y

beneficios previamente insospechados o que logren un avance significativo en el conocimiento, que permita evolucionar otro tipo de proyectos. Me refiero especialmente a los proyectos de investigación básica y en menor medida a los de desarrollo, para los que se hace muy difícil encontrar entidades privadas que los financien.

Esto se debe a que la investigación básica requiere grandes aportaciones de capital para la obtención de una salida, que no es aplicable directamente a mercado. La financiación, en este caso, se asume por parte de los estados y permite que los proyectos relacionados con p.ej., el colisionador de partículas, obtengan financiación. Esto es importante, ya que estos avances permiten que surjan nuevos proyectos, antes inviables y algunos con aplicabilidad industrial hacia nuevas estructuras en materiales o cualquier otro campo. Respecto a los proyectos de desarrollo, pocas empresas los financian, en este caso porque los perciben como ideas contrarias a la lógica y a lo que conocen.

Esto no ocurre con los proyectos de tipo investigación aplicada. Además de poder ser financiados por empresas privadas, interesadas en la aplicación directa de conocimientos ya adquiridos por la universidad, también se pueden financiar dentro de un sistema mixto público-privado. Esto permite a las empresas “invertir” en la aplicación de conocimiento adquirido por las universidades en productos, servicios y procesos, de una forma bastante inmediata. Este sistema también resulta interesante para las universidades, pese a que no obtienen conocimiento científico básico, ya que además de conseguir financiación, les permite conseguir experiencia que pueden aplicar en futuros proyectos.

Actualmente, la financiación estatal y regional carece de los medios económicos disponibles en la Unión Europea (UE). España no destaca dentro de los países más industrializados por el porcentaje de su producto interior bruto destinado a investigación. Es por esto, por lo que las convocatorias de proyectos de investigación más potentes económicamente son, mayoritariamente y con diferencia, las gestionadas por la Unión Europea. La UE organiza las convocatorias de investigación habitualmente, enmarcadas en programas plurianuales, buscando objetivos tales como mejorar el intercambio de conocimiento entre sus estados miembros, favorecer la excelencia científica, la competitividad y la innovación en la UE respecto al resto del mundo. Pese a que la UE ofrece numerosas convocatorias de financiación y se puede revisar el conjunto de convocatorias atendiendo a los objetivos en la web oficial⁹, en este documento solo describiremos dos de las más interesantes para nuestro objetivo.

En este momento, el programa en vigor con mayor dotación dentro de la UE es el *8º Programa marco*, denominado *Horizonte 2020*¹⁰, que abarca los años 2014-2020. Tiene una dotación cercana a los 100 billones de euros y se estructura en tres líneas básicas, que todo proyecto que desee competir debería conocer:

⁹ https://ec.europa.eu/info/funding-tenders/funding-opportunities/funding-programmes/overview-funding-programmes_es

¹⁰ <https://ec.europa.eu/programmes/horizon2020>

1. Ciencia excelente, para reforzar la excelencia científica de la UE a nivel mundial, principalmente mediante iniciativas de temática abierta y en general, en proyectos individuales.
2. Liderazgo industrial, para acelerar el desarrollo de las tecnologías. Principalmente tecnologías de la información y la comunicación, nanotecnología, materiales avanzados, biotecnología, fabricación y transformación avanzadas, y tecnología espacial. También para ayudar, a las PYME innovadoras europeas, a convertirse en empresas líderes y para facilitar la financiación de riesgo en actividades de investigación e innovación en su llegada al mercado.
3. Retos sociales, para aportar una respuesta directa a las prioridades políticas y los retos identificados en la estrategia Europa 2020. Tales como: la seguridad, la energía, el transporte, el cambio climático y el uso eficaz de los recursos, la salud y el envejecimiento, los métodos de producción respetuosos del medio ambiente y la gestión del territorio. Básicamente es un eje de I+D+i que permite investigar e innovar sobre las grandes cuestiones que afectan a los ciudadanos europeos y que son de muy distintos ámbitos.

El programa H2020 está teniendo muy buenos resultados para los investigadores españoles, lo que nos debe animar a intentar conseguir el acceso a este tipo de financiación. En el informe sobre resultados parciales de participación en el marco de H2020¹¹, España se sitúa en el cuarto puesto atendiendo a la financiación total conseguida, tan solo por detrás de Alemania, Reino Unido y Francia. En cuanto a participación, somos los segundos de los 28. Avanzamos puestos respecto a la financiación conseguida porque tenemos proyectos de menor importe, seguramente derivado de la diferencia salarial con los otros países comentados en este párrafo. Se puede consultar el conjunto de detalles y otros datos de interés en la web de resumen del H2020 por países¹².

En el período 2014-2019, las entidades españolas han obtenido una financiación total de 2.816 millones de euros. Esto supone algo más del 10% del total de la financiación repartida por el programa entre los 28 miembros (UE antes de la separación de Reino Unido). Es de esperar que, pese a que la financiación se mantenga en el futuro programa (incluso descendiendo con relación al incremento en el coste de vida), el hueco que deja Reino Unido, lo completen otros países. Este es otro dato que nos invita a la participación en este programa en el futuro.

El programa H2020 es sumamente transparente y pueden verse en su web informes y estadísticas muy variados. Entre ellos, podemos destacar que aparecen

¹¹

http://eshorizonte2020.cdti.es/recursos/doc/Programas/Cooperacion_internacional/HORIZONTE%202020/31845_107107202011211.pdf

¹²

<https://webgate.ec.europa.eu/dashboard/extensions/CountryProfile/CountryProfile.html?Country=Spain>

datos relativos al porcentaje de propuestas con participación española aceptadas, sobre el total de propuestas presentadas. Han sido en torno al 22%, siendo más de 4.000 los proyectos financiados con participación nacional. Este dato es sumamente positivo, ya que la media de aceptación es tan solo del 13,03%. Desgraciadamente, solo 550 de ellos fueron coordinados por entidades españolas.

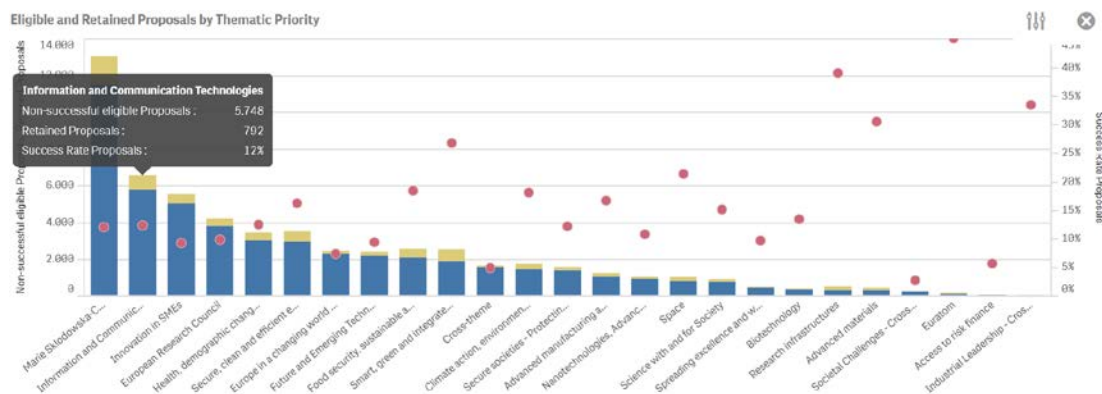


Ilustración 1. Propuestas presentadas respecto a las financiadas por área temática en España

La situación para los proyectos de investigación presentados dentro del área de tecnologías de la información y la comunicación es diferente, tal y como se detalla en la Ilustración 1, con 792 propuestas financiadas, respecto a las 5.748 no financiadas (12%).

La figura también presenta un dato interesante sobre las áreas con más posibilidades de ser financiadas. Esto, nos puede llevar a incluir en el proyecto una orientación hacia casos de uso concretos, dentro de las áreas con mayor financiación.

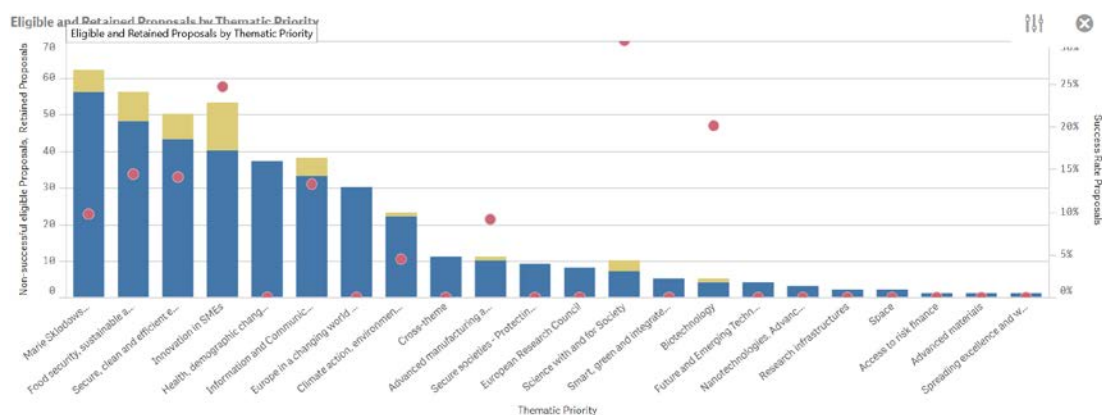


Ilustración 2. Propuestas presentadas respecto a las financiadas por área temática en Extremadura

Otro dato destacable es que, entre los beneficiarios españoles, las empresas son las organizaciones que más contribuyen, reuniendo el 38,3% de la financiación obtenida por nuestro país. Las universidades españolas ocupan un segundo lugar con

el 20,1% sobre la financiación a entidades nacionales. Ya muy por debajo de estos números se encuentran otro tipo de entidades, como los centros públicos de investigación (11,6%), asociaciones de investigación (10,6%), centros tecnológicos (10%), Administraciones Públicas (5,5%) y asociaciones (3,9%).

Desgraciadamente, no nos encontramos al mismo nivel en todas las comunidades autónomas españolas. Extremadura es la última comunidad autónoma por volumen de financiación proveniente del programa H2020. En la Ilustración 2, se recoge la información referente al grado de aceptación de proyectos en relación con nuestra comunidad autónoma.

Atendiendo directamente a proyectos relacionados con la propuesta aquí presentada¹³, IoT y términos similares, por un lado, se observa que no hay muchas propuestas que se hayan financiado en H2020 con estos descriptores o categorías, tan solo 14. Por otro lado, se observa que son relativamente grandes (180 participantes) y con importes altos de financiación (49,18M €), con participantes del tamaño y trascendencia de Siemens entre muchos otros. Esta situación refuerza quizás todavía más la propuesta que hacíamos anteriormente, de orientarlo a casos de uso en sectores/áreas con mayores posibilidades de obtener financiación.

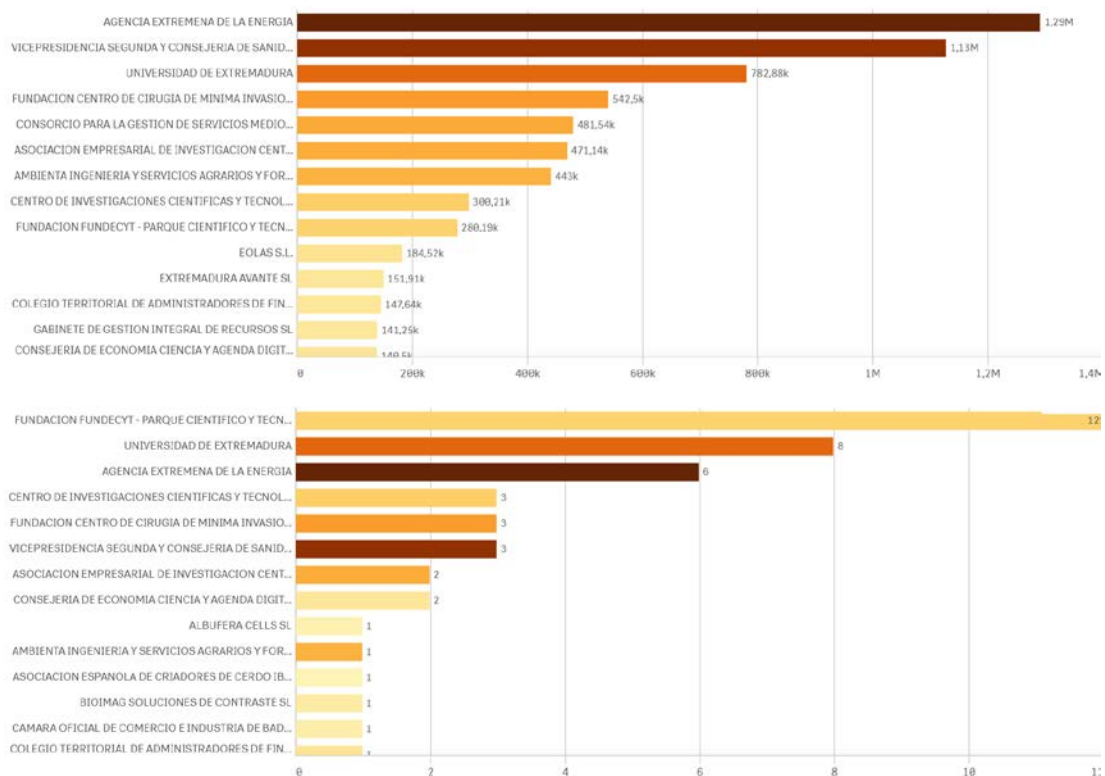


Ilustración 3. Arriba - Clasificación de organizaciones extremeñas por captación de fondos en el programa H2020. Abajo - Clasificación de organizaciones extremeñas por participación en el programa H2020. Fuente: H2020 dashboard

¹³ <https://webgate.ec.europa.eu/dashboard/sense/app/93297a69-09fd-4ef5-889f-b83c4e21d33e/sheet/a879124b-bfc3-493f-93a9-34foe7fba124/state/analysis>

En Extremadura, la financiación referente al programa H2020 se reparte, principalmente, entre la Agencia extremeña de la energía, la Vicepresidencia segunda y Consejería de sanidad y la Universidad de Extremadura en ese orden, en cuanto a captación de fondos. La Ilustración 3 recoge estas clasificaciones.

Sin embargo, nuevamente la participación recoloca la clasificación en base al número de proyectos. Es interesante destacar de la Ilustración 3, como el *FUNDECYT – Parque científico y tecnológico de Extremadura* está impulsando los proyectos europeos en la región. Pese a participar en muchos, en base a las cantidades de financiación captadas, lo más probable es que lo haga como apoyo a otras entidades de la región.

Creo que el Parque científico y tecnológico debe servir de punto de unión entre la Universidad y las empresas, PYME principalmente, para aumentar considerablemente las posibilidades de aceptación de las propuestas de proyecto presentadas. Esto debería hacerse, viendo las áreas con mayores posibilidades y contactando con los grupos de la Universidad expertos en cada área.

Dentro de la UE, existen otros programas de financiación, como la actual convocatoria *Interreg V¹⁴*, que cubre el periodo 2014-2020. Esta convocatoria busca la cohesión de territorios y que surjan acciones y políticas comunes y de cohesión entre diferentes miembros de la UE. Es una ayuda diferente del H2020 en cuanto a beneficiarios potenciales de financiación, ya que se focaliza en agentes nacionales, regionales y locales.

El objetivo de estas convocatorias es muy ambicioso, intentando buscar la armonía económica, social y de desarrollo territorial entre los miembros de la UE. Son bien conocidas las grandes diferencias en todos estos aspectos, entre los miembros de la UE. Además, también existen diferencias dentro de las regiones de los estados miembros. El programa *Interreg-POCTEP¹⁵* se especializa, concretamente, en promover el desarrollo y la cooperación transfronteriza de índole pública y privada entre España y Portugal, no dejando atrás temas como el entorno rural y la despoblación.

Los planes nacionales de investigación, menos cuantiosos que los europeos, son difíciles de conseguir por la alta competencia debido, en gran medida, al bajo importe de fondos puestos en valor. En este caso, también tienen un marco plurianual común, en concordancia con el marco europeo. Este marco estatal se denomina *Plan estatal de investigación científica y técnica y de innovación¹⁶*, abarca los años 2013-2020 y sirve de marco estratégico de referencia para el conjunto del país, en materia de investigación e innovación.

¹⁴ <https://www.interregeurope.eu/>

¹⁵ <https://www.poctep.eu/>

¹⁶

<https://www.ciencia.gob.es/portal/site/MICINN/menuitem.29451c2ac1391f1febebed1001432eao/?vgnextoid=bcd9c5e1876f5610VgnVCM1000001d04140aRCRD>

La Tabla 1 recoge la financiación estimada para los diferentes programas contemplados en el plan estatal durante el periodo 2013-2016. No ha sido posible encontrar información oficial para el periodo actual, correspondiente al periodo 2017-2020.

Los objetivos del plan estatal citado son, en la práctica y a mi entender, claros y se orientan principalmente a contribuir a la consolidación del sistema español de ciencia, tecnología e innovación. Estos están en gran medida sustentados por grupos de investigación ya consolidados por la calidad de su investigación y su impacto científico, técnico, social y/o económico. Así mismo, también intentan incentivar la participación de entes privados en actividades de I+D+i y fomentar la interacción universidad-empresa, haciendo que muchas convocatorias concretas requieran la participación de ambos tipos de entidades.

Tabla 1. Distribución de la financiación planificada en el Plan estatal de investigación científica y técnica y de innovación (2013-2016). Fuente: Ministerio de Ciencia e Innovación¹⁷

	Presupuesto total	Subvención	Préstamo	
Promoción del Talento y su Empleabilidad en I+D+i	1.394.472.158 €	1.194.472.158 €	200.000.000 €	13%
Formación	698.126.061 €	698.126.061 €		
Incorporación	633.370.000 €	433.370.000 €	200.000.000 €	
Movilidad	62.976.097 €	62.976.097 €		
Fomento de la Investigación Científica y Técnica de Excelencia	977.457.000 €	753.457.000 €	224.000.000 €	9%
Generación de conocimiento	528.457.000 €	528.457.000 €		
Fortalecimiento institucional	192.000.000 €	152.000.000 €	40.000.000 €	
Infraestructuras científicas y equipamiento	257.000.000 €	257.000.000 €		
Liderazgo Empresarial en I+D+i	2.312.600.000 €	208.600.000 €	2.104.000.000 €	21%
I+D+i Empresarial	1.014.150.000 €	75.100.000 €	939.050.000 €	
Tecnologías facilitadoras esenciales	789.950.000 €		789.950.000 €	
I+D+i colaborativa orientada a las demandas del tejido productivo	508.500.000 €	133.500.000 €	375.000.000 €	
I+D+i Orientada a los Retos de la Sociedad	6.137.889.402 €	2.269.579.802 €	3.868.309.600 €	57%
Subprograma Retos	4.192.964.940 €	1.573.364.940 €	2.619.600.000 €	
Acción estratégica en Salud	447.364.114 €	447.364.114 €		
Acción estratégica en Economía y Sociedad Digital	1.497.560.348 €	248.850.748 €	1.248.709.600 €	
Total	10.822.418.560 €	4.426.108.960 €	6.396.309.600 €	41% 59%

A nivel regional, Extremadura vertebró la financiación de la investigación también mediante un sistema plurianual, actualmente cubriendo el periodo 2014-2020 y que se denomina *Estrategia regional de Investigación e Innovación para la Especialización Inteligente de Extremadura (Estrategia RIS3)*¹⁸.

Es uno de los instrumentos principales de la región para articular y poner en marcha la transformación económica necesaria para hacer de Extremadura una región más competitiva en un contexto global y de manera sostenible en el tiempo. Dado el estado económico regional y el de abandono rural, que no solo afecta a nuestra región, éste es un objetivo muy ambicioso. Se trata generalmente de proyectos con menos requisitos que cumplir a la hora de redactarlos (no requieren p.ej., consorcio). Pero

¹⁷ <https://www.ciencia.gob.es/stfls/MICINN/Prensa/FICHEROS/2018/PlanEstatalIDI.pdf>

¹⁸ <http://www.ris3extremadura.es/>

disponen de menor importe, su memoria económica es más simple, y su impacto es más limitado, por lo que no deben detallarse tantos aspectos.

En la Tabla 2 se recoge la previsión de financiación total de I+D+i en Extremadura para el periodo 2017-2020, incluyendo aquella asociada al VI Plan regional.

La Tabla 2 da una idea de la importancia del plan regional en Extremadura, ya que acapara aproximadamente el 50% de la financiación de la investigación en la comunidad, año tras año.

Extremadura no es una región que haya destacado en su historia por la inversión en I+D. Consultando datos del Instituto Nacional de Estadística (INE) sobre su encuesta de innovación en 2017¹⁹, la última hasta la fecha, en Extremadura la inversión no llegó a los 30.000.000€, la más baja con diferencia de todas las comunidades autónomas españolas.

Tabla 2. Previsión de financiación de la I+D+i en Extremadura para el periodo 2017-2020. Fuente:

FINANCIACIÓN I+D+i Extremadura (2017-2020)					
ÓRGANO	PROGRAMAS	2017	2018	2019	2020
ACTUACIONES SECRETARÍA GENERAL CIENCIA, TECNOLOGÍA E INNOVACIÓN	P. PARA EL DESARROLLO DE LA CARRERA INVESTIGADORA E INCORPORACIÓN DE TALENTO Y SU EMPLEABILIDAD	12.632.054 €	13.263.657 €	13.926.840 €	14.623.182 €
	P. PARA EL FOMENTO DE LA INVESTIGACIÓN CIENTÍFICA Y TÉCNICA	7.225.283 €	7.586.547 €	7.965.875 €	8.364.168 €
	P. DESARROLLO I+D+i EN TEJIDO EMPRESARIAL	20.447.989 €	21.470.388 €	22.543.908 €	23.671.103 €
	P. DESARROLLO Y ORGANIZACIÓN CENTROS I+D+i SECTI Y SUS INFRAESTRUCTURAS CIENTÍFICAS	18.233.431 €	19.145.103 €	20.102.358 €	21.107.476 €
	P. PARA LA TRANSFERENCIA Y DIFUSIÓN DE CULTURA CIENTÍFICA	2.177.225 €	2.286.086 €	2.400.391 €	2.520.410 €
(1) TOTAL		60.715.982 €	63.751.781 €	66.939.370 €	70.286.339 €
ACTUACIONES DE OTRAS CONSEJERÍAS EN I+D+i	P. PARA EL DESARROLLO DE LA CARRERA INVESTIGADORA E INCORPORACIÓN DE TALENTO Y SU EMPLEABILIDAD	3.411.942 €	3.480.181 €	3.549.784 €	3.620.780 €
	P. PARA EL FOMENTO DE LA INVESTIGACIÓN CIENTÍFICA Y TÉCNICA	7.222.570 €	7.367.022 €	7.514.362 €	7.664.649 €
	P. DESARROLLO I+D+i EN TEJIDO EMPRESARIAL	2.678.446 €	2.732.015 €	2.786.655 €	2.842.388 €
	P. DESARROLLO Y ORGANIZACIÓN CENTROS I+D+i SECTI Y SUS INFRAESTRUCTURAS CIENTÍFICAS	13.210.096 €	13.474.298 €	13.743.784 €	14.018.660 €
	P. PARA LA TRANSFERENCIA Y DIFUSIÓN DE CULTURA CIENTÍFICA	4.145.841 €	4.228.758 €	4.313.333 €	4.399.600 €
TOTAL		30.668.896 €	31.282.273 €	31.907.919 €	32.546.077 €
ACTUACIONES OTROS ORGANISMOS	P. DESARROLLO I+D+i EN TEJIDO EMPRESARIAL	4.500.000 €	6.500.000 €	9.500.000 €	11.500.000 €
(3) TOTAL		4.500.000 €	6.500.000 €	9.500.000 €	11.500.000 €
FINANCIACIÓN TOTAL VI PLAN REGIONAL DE I+D+i		95.884.878 €	101.534.055 €	108.347.289 €	114.332.416 €
OTRAS APORTACIONES A LA INVESTIGACIÓN	UEX (33% gastos personales I+D+i)	30.646.424 €	30.646.424 €	30.646.424 €	30.646.424 €
	SES (5% gastos generales en I+D+i)	75.372.011 €	75.372.011 €	75.372.011 €	75.372.011 €
	TOTAL	106.018.436 €	106.018.436 €	106.018.436 €	106.018.436 €
FINANCIACIÓN TOTAL I+D+i		201.903.313 €	207.552.490 €	214.365.725 €	220.350.851 €

No se ha encontrado información de fuentes oficiales sobre el grado de consecución de las partidas designadas por los planes ni nacionales ni extremeños, relacionados en la Tabla 1 y Tabla 2. Así que no podemos saber realmente el importe destinado a I+D+i en estos últimos años. Sería un ejercicio de transparencia que estos datos fueran públicos, ya que ayudarían a entender que se dedicasen más esfuerzos a unos programas o convocatorias o a otras. No se han encontrado datos globales sobre

¹⁹ <https://www.ine.es/jaxi/Datos.htm?path=/t14/po61/a2017/lo/&file=03001a.px#!tabs-grafico>

la demanda de estos planes de manera centralizada, por lo que no es posible hacer tampoco un análisis detallado de manera directa.

Es extraño comprobar cómo, siendo el INE una interesante herramienta para la orientación y refuerzo de los planes de investigación, ni los *Indicadores del Sistema Español de Ciencia y Tecnología*, ni los indicadores referentes a la *Evaluación de la I+D* figuren en el *Plan estadístico nacional 2017-2020* de este organismo.

Pese a que las propuestas competitivas provenientes de la comunidad autónoma extremeña se pueden ver favorecidas en algunas convocatorias, por ser región objetivo de la UE, resulta difícil en esta región conseguir financiación privada, dado el bajo número de empresas. En este sentido, si analizamos en detalle la situación de la I+D en Extremadura, según los datos del INE, se observa que la inversión en investigación en Extremadura es la segunda más baja de todo el territorio nacional, junto con Baleares y Cantabria. La Ilustración 4 destaca este hecho y puede observarse como las regiones con mayor inversión desde el sector privado son también las que mayor número de empresas concentran en su territorio.

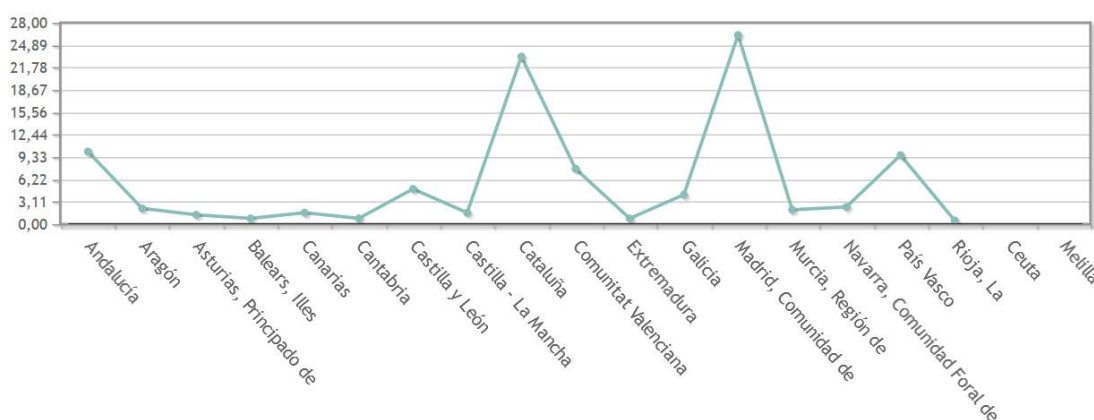


Ilustración 4. Estadística sobre gasto en actividades de I+D en el sector empresas porcentual por Comunidades autónomas. Fuente: INE 2020 (datos de 2017)

Adicionalmente, cabe señalar que, del total de gasto en I+D en la comunidad de Extremadura según el INE, tan solo el 6,5% procede de empresas de consideradas como alta tecnología (44 empresas extremeñas, 11 de ellas manufactureras y 33 del sector servicios). Mientras, en las comunidades destacables positivamente en base a los valores de la Ilustración 4 (como son la Comunidad de Madrid o Cataluña) el porcentaje de inversión en I+D por estas mismas empresas está entre el 35% y el 40% de la inversión total.

Según los últimos datos disponibles (2017) del INE, el sector servicios acapara en la región con diferencia, al igual que la media nacional, las actividades en I+D con una inversión de 5.544.000 €, frente a los 1.863.000 € del sector manufacturero. Nuevamente, con datos muy lejanos a la media nacional y, en especial, con relación a las comunidades autónomas que más actividad I+D en entorno empresarial poseen

(principalmente Comunidad de Madrid, Cataluña y País Vasco en ese orden, según las estadísticas consultadas relativas a diferentes partidas de inversión).

La Ilustración 5 da una idea clara de la concentración de gasto en I+D por comunidades. Ya que el principal gasto son costes de personal, se entiende que una comunidad con mayor gasto, o tiene salarios para el personal de I+D más altos y/o tiene una mayor cantidad de personal. En cualquiera de los dos casos, lo coloca como potencia con relación al resto de comunidades.

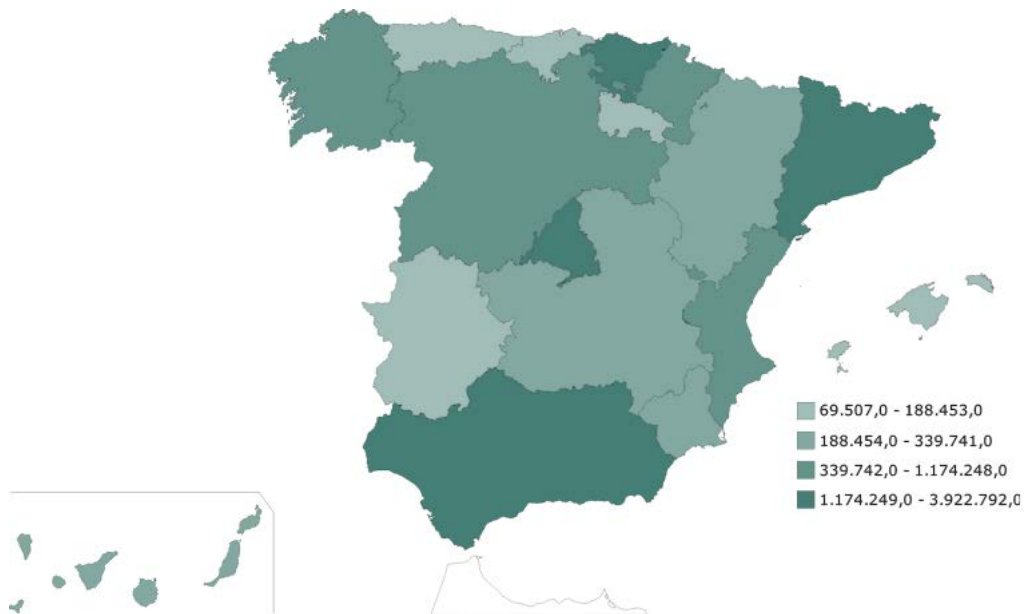


Ilustración 5. Gasto en I+D por comunidades incluyendo gastos de personal (2017). Fuente: INE²⁰

Según datos recabados del VI Plan regional, el número de personas dedicadas a la I+D en la región, se ha visto reducido de las 2.402 personas en 2010, a las 1.978 personas en 2015, lo que supone una disminución del 17,7%. No hemos podido encontrar datos oficiales de periodos más recientes, pero este fuerte descenso lo vemos a diario en la Universidad, un entorno propicio para la contratación de personal desarrollando labores en I+D. Entiendo que este hecho pone de manifiesto la debilidad del sistema y la necesidad de implementar instrumentos específicos que restablezcan e incrementen el potencial investigador de la región.

En el año 2015, el personal investigador en Extremadura no alcanzaba 1,20 investigadores por cada 1.000 habitantes, mientras la media nacional era de 2,63 investigadores por cada 1.000 habitantes. Este hecho incrementa todavía más la distancia ya existente en investigación entre comunidades autónomas.

No se han encontrado datos tan detallados del INE, ni de otras fuentes oficiales, para periodos posteriores a los comentados. Pero sí deben existir, ya que en la página

²⁰ <https://www.ine.es/jaxi/Datos.htm?path=/t14/p057/a2018/lo/&file=07001.px#!tabs-mapa>

web del INE, en la sección de estadísticas territoriales referentes a ciencia y tecnología²¹, aparecen mapas (pero no datos concretos) para el año 2018 por comunidades autónomas. En cualquier caso, vistos los mapas de calor por comunidades, estos datos corroboran afirmaciones hechas a lo largo de esta sección.

A nivel de justificación documental de los gastos realizados y consecución de los objetivos del proyecto, esta debe ser prácticamente igual de cuidadosa en las convocatorias regionales, nacionales y europeas. Esto se debe a que finalmente los fondos puestos en valor mediante estas convocatorias provienen en su mayoría de fondos europeos (FEDER...). Por lo que están sujetos a las reglas de gasto que marca la UE.

Dadas las dotaciones económicas y la competencia de las convocatorias nacionales, el autor de este documento considera que estas pueden ser útiles en dos contextos principalmente. Uno, como un añadido sobre otras convocatorias, cuando las consiguen grupos de investigación consolidados que ya cuentan con otro tipo de financiación (sobre todo supranacional). Otro, para mantener líneas de investigación no tan punteras, redes de trabajo y contacto con otros investigadores de otras universidades del ámbito nacional, o bien para abrir nuevas líneas de trabajo sobre investigaciones incipientes en convocatorias concretas.

En general, los fondos regionales y nacionales captados, si se usan correctamente, pueden obtener resultados previos que abran diferentes líneas de investigación, principalmente dentro del ámbito de la UE. Además, estos resultados preliminares o parciales quizás puedan asegurar, con mayor probabilidad, que una potencial inversión privada vaya a tener un retorno.

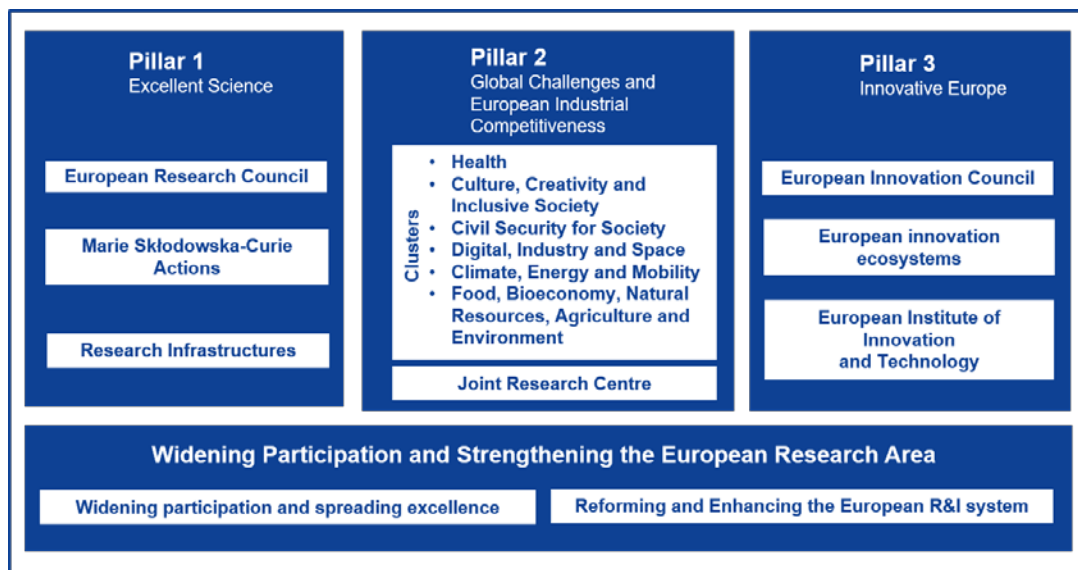


Ilustración 6. Estructura preliminar del programa Horizonte Europa. Fuente: UE²²

²¹ <https://www.ine.es/dynInfo/Infografia/Territoriales/galeriaCapitulo.html?capitulo=4333>

²² https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

Como último punto de esta sección de financiación y, una vez concluido que las convocatorias europeas deben ser objetivo prioritario de los proyectos solicitados, quizás no las propias de nuestra rama de conocimiento por su bajo ratio de aceptación, finalizamos esta sección con los avances ya oficiales sobre la continuación del programa H2020, que termina este año.

Al nuevo programa se le denomina actualmente Horizonte Europa²³ y se basa en tres pilares, similares a los del H2020, pero con matices. La Ilustración 6 muestra estos tres pilares.

Este nuevo programa marco incluye 5 misiones específicas de gran importancia para la UE:

- Adaptación al cambio climático, incluyendo transformaciones sociales
- Cáncer
- Ciudades inteligentes y climáticamente neutras
- Océanos, mares y aguas costeras e interiores saludables
- Salud del suelo y alimentación

Sería conveniente incluir aspectos relacionados en el proyecto solicitado, con vistas a futuras convocatorias. En concreto, UbiCom entra dentro de ciudades inteligentes, por lo que potenciar la orientación hacia el bajo consumo energético sería deseable.

2.3 PERSONAL Y DEL DESARROLLO PROPUESTO

El investigador autor de este documento pertenece actualmente al grupo de ingeniería del software Quercus²⁴. El grupo nació en 1996 en el seno de la Universidad de Extremadura y actualmente lo constituyen 26 investigadores. No son habituales los grupos formados por tantos investigadores. Esto conlleva que las líneas e inquietudes de investigación, pese a estar relacionadas con el ámbito de la ingeniería de software, sean muy diversas. En relación con estas líneas de investigación el grupo, dirigido por Juan Hernández Núñez desde su creación, ofrece servicios de investigación, formación y consultoría.

Las líneas de investigación del grupo actualmente son: las arquitecturas orientadas a servicios, los datos abiertos semánticos relacionados, el desarrollo de software dirigido por modelos, la computación en la nube y la computación móvil, la ingeniería web, el aprendizaje máquina/automático, los sistemas pervasivos/ubicuos, internet de las cosas, ciudades inteligentes y relacionados.

Como puede observarse, muchas de estas líneas están interrelacionadas entre sí y han propiciado y siguen propiciando, en muchas ocasiones, interacciones entre los

²³ https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

²⁴ <https://quercusseg.unex.es/>

miembros del grupo, que se concretan en publicaciones, proyectos y contratos. En captación de fondos por transferencia de conocimiento, el grupo Quercus ocupa un lugar relevante dentro de la UEx.

Algunas de estas líneas de investigación, han dado lugar a la creación de varias empresas por algunos integrantes del grupo de investigación. En concreto *Gloin S.L.*²⁵ y *Homeria Open Solutions S.L.*²⁶, siendo socio fundador de esta última el autor de este documento.

También formo parte del *Instituto de investigación en tecnologías informáticas aplicadas de Extremadura* (INTIA), que actualmente y, desde su creación, es coordinado también por Juan Hernández Núñez. El equipo investigador que lo conforma es interdisciplinar, pero con temática afín, tal y como se ha comentado al introducir este tipo de organizaciones de investigación universitarias.

En concreto, el equipo lo conformamos más de 25 investigadores relacionados con las Tecnologías Informáticas, específicamente ingeniería del software, arquitectura de computadores e ingeniería de comunicaciones. Por ello, INTIA es un instituto capaz de afrontar algunos de los retos más importantes en materia TIC.

Algunos retos potenciales específicos por resolver que se plantean por parte del instituto son, entre otros: la Computación paralela y distribuida de altas prestaciones, la ingeniería del software interrelacionada con el desarrollo de arquitecturas de computadores, el desarrollo de sistemas de comunicación especializados mediante hardware y software.

Este instituto pretende servir de referente en el conocimiento científico y técnico de las TIC, para la mejora del estado del arte y de la técnica y su transferencia al sector privado y público. Para ello, la pretensión es abordar proyectos de investigación, desarrollo e innovación que coincidan con los retos regionales, nacionales y europeos, contribuyendo con ello, al desarrollo de los diferentes temas objetivo de los programas de investigación.

A medio y largo plazo, el instituto pretende acceder a proyectos de investigación básicos, que le permitan aportar nuevo conocimiento relacionado con sus objetivos de investigación principales.

Mediante la captación de fondos para investigación aplicada, el instituto pretende responder a algunos de los retos fijados por las estrategias y programas marco de Extremadura, España y Europa. En estos, se pueden aplicar los conocimientos de sus miembros y requieren una aproximación interdisciplinar desde el punto de vista de las comunicaciones, las arquitecturas de computadores y la ingeniería del software. Como vemos, algunos de estos están claramente alineados con los sistemas ubicuos, objetivo de este documento.

²⁵ <http://www.gloin.es/>

²⁶ <https://www.homeria.com/>

Como se refleja de lo anterior, tanto el grupo como el instituto de investigación a los que pertenezco, tienen los sistemas ubicuos entre sus líneas de investigación o de posible aplicación.

El departamento de tecnología de los computadores y las comunicaciones, en el que actualmente imparto docencia, también cuenta con asignaturas más allá de las de formación básica, donde se tratan temas de entrada a estos retos de investigación.

A nivel personal, cuento con diversas publicaciones relacionadas directa y estrechamente con sistemas ubicuos. En estas se incluye una reciente muy concreta [18] en la línea que se presenta como propuesta de proyecto de investigación en este documento, en la sección 3.

Sistemas ubicuos, pervasivos o la inteligencia ambiental se pueden considerar como términos homónimos, así como relacionados íntimamente con muchos términos como domótica, industria 4.0, etc.

Los sistemas ubicuos tienen su base principal en las tecnologías de la información y la comunicación. La novedad que incluyen es la integración de la computación en todos los objetos que nos rodean en todo momento, de manera que las personas al usarlos no los perciban como objetos de computación. Se persigue en UbiCom que interactuemos con ellos de manera transparente (interacción implícita o explícita, pero usando medios naturales para la persona). La computación ubicua (UbiCom) busca adaptar los sistemas a las personas.

Los sistemas UbiCom sienten el entorno y actúan sobre él mediante la adición a los sistemas de computación de sensores y actuadores microcontrolados. Así, todos los sensores y actuadores cuentan con un pequeño computador, con capacidad computacional limitada y muy orientada a controlar otros dispositivos. Estos dispositivos que son una parte básica en este tipo de sistemas.

La independencia del objeto de cableado, tanto para corriente como para las comunicaciones puede ser, en muchos casos, un requisito para la propia interacción transparente con el dispositivo.

Tras el gran auge vivido, y todavía en constante evolución, por las tecnologías de computación y las de comunicación inalámbrica, podemos decir que la humanidad se encuentra en un momento en que los sistemas que integran sensores y actuadores se están multiplicando a nuestro alrededor, muchas veces de manera embebida, usando estas tecnologías. Esto se conoce como IoT (*Internet of Things*, Internet de las Cosas).

IoT es un término comúnmente adoptado hoy día, tanto en el ámbito de la investigación, donde es un tema candente, como en la propia sociedad, que está viviendo la integración y cómo afecta a su día a día. Generalmente, esta tecnología, se asocia al envío continuado de datos desde los objetos (nodos) y este es uno de los temas que más nos interesan en relación con este proyecto investigador.

Para este proyecto, es interesante comentar que en UbiCom es habitual que los nodos formen redes de sensores y actuadores y tengan capacidades de interconexión muchas veces entre ellos, aunque no sea de manera directa punto-a-punto, sino a través de otros nodos de la red.

Entre los muchos retos que plantea la adopción masiva de IoT están la saturación de redes de comunicaciones por el envío continuado de información, en ocasiones masivos, al tratarse de fuentes continuas multimedia, como audio o vídeo.

Hoy día, y de manera creciente, vemos una adopción importante de asistentes de voz, sistemas de televigilancia, etc. Sin embargo y con diferencia, el caso más generalizado de dispositivos IoT para el ámbito de UbiCom es el de dispositivos dotados de sensores o actuadores que hacen uso de datos discretos, como medidas físicas unidimensionales individuales o agrupados: como la temperatura, la humedad o el ritmo cardíaco entre otras.

El almacenamiento de los datos que producen esta gran cantidad de sistemas de manera autónoma y continuada en el tiempo, sobre todo pensando que se encontrarán a miles a nuestro alrededor, es uno de los problemas directamente relacionados con lo propuesto en este proyecto de investigación.

Se asume que es interesante, o puede serlo para el futuro, tener un histórico de los datos. No solo para su consulta puntual sobre históricos, sino para servir de fuente de datos para procesamiento automático. Lo que puede traer ventajas adicionales, como la detección de datos anómalos, la optimización de envíos de información y con ello, de autonomía de los sistemas basados en baterías, etc.

Las capacidades computacionales y de almacenamiento de los objetos/nodos a nuestro alrededor siguen evolucionando y estamos en un momento en que, muchas veces, están siendo infrautilizadas. Al mismo tiempo, necesitamos adquirir nuevo hardware para almacenar los datos que produce la propia red de sensores y/o para realizar labores de procesamiento, muchas veces no excesivamente complejas computacionalmente.

3 DESARROLLO DEL PROYECTO DE INVESTIGACIÓN

El presente proyecto de investigación se orienta como una propuesta de investigación no unida a ninguna convocatoria concreta actual. Se planea utilizar este documento como base para la presentación de varios proyectos dependiendo de diversos factores, incluyendo las posibilidades de dedicación del personal investigador del grupo (Quercus SEG) y del instituto de investigación al que pertenezco (INTIA).

Se plantea como un proyecto plurianual, a 6 años vista, que podría materializarse en varios proyectos de 3+3 años, 4+2, etc. atendiendo al contexto de financiación y personal disponible y entidades interesadas en el momento de la petición.

Como breve introducción al desarrollo del proyecto propuesto, este versa sobre aspectos relacionados con la gestión del almacenamiento y el procesamiento de datos generados por una red de sensores de los sistemas ubicuos, tal y como propone la convocatoria del concurso en el que se enmarca. Esto incluye otros aspectos relacionados de manera bastante directa, como es la adquisición de los datos una vez almacenados o la transmisión de estos en la propia red, con el objetivo de conseguir sistemas confiables y óptimos.

3.1 RESUMEN

Los dispositivos sensores / actuadores se están adoptando actualmente de forma masiva, a menudo como nodos de redes de sensores más grandes. Estas redes de sensores suelen estar dedicadas a la adquisición de contexto (por ejemplo, obtener la temperatura), así como a proporcionar servicios de actuación (por ejemplo, abrir las persianas).

Sin embargo, con respecto a su propio almacenamiento de datos, la tendencia tanto a nivel de investigación como en prácticamente todos los despliegues es que los datos generalmente se envíen a servidores *fog/cloud* para su almacenamiento. En este sentido, usar almacenamiento en la propia red que forman los sensores que generan los datos, no es una tendencia y mucho menos una opción de despliegue habitual.

En la Ilustración 7 se muestran las diferentes capas en las que se suele dividir la computación atendiendo a diversos factores, principalmente, la cercanía al nodo generador de información, aunque también las necesidades computacionales.

Por supuesto, las soluciones de almacenamiento en *fog* o *cloud* ofrecen varias ventajas sobre las soluciones de almacenamiento en red de sensores, pero también ciertos inconvenientes. P.ej., en entornos *cloud*, pueden aparecer problemas de privacidad y legales, que desaparecen al almacenar los datos en servidores *fog*, más cercanos a la producción del dato. Sin embargo, en entornos *fog* se debe comprar y mantener hardware adicional, lo que aumenta los costes. Coste que incluso se hace más relevante cuando se requiere que el almacenamiento sea redundante y distribuido entre varios computadores, para aumentar la seguridad ante fallos hardware.

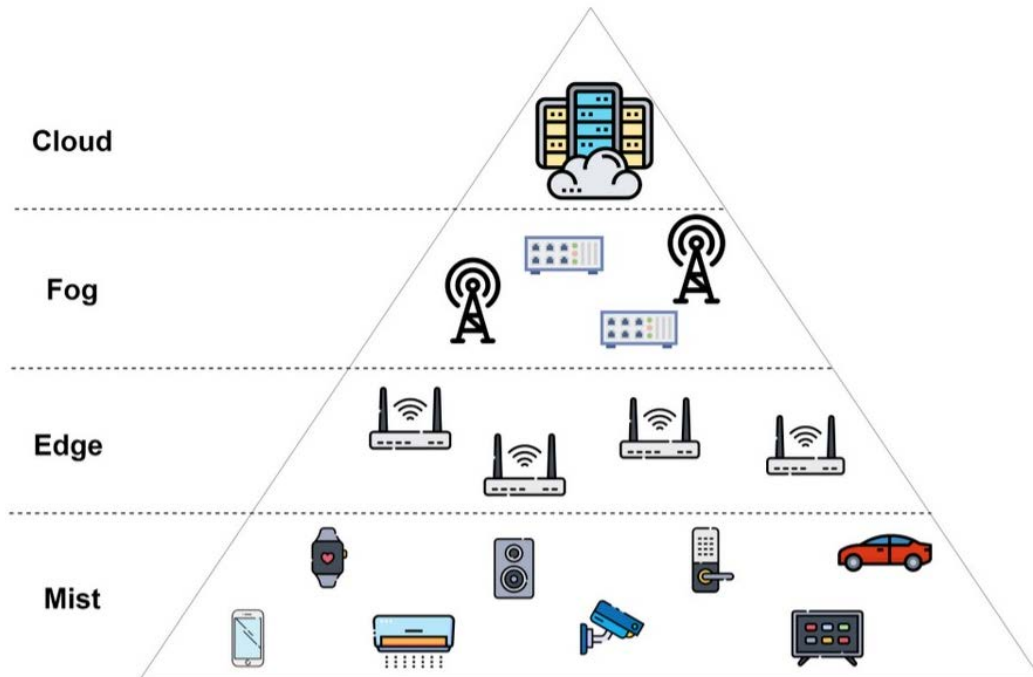


Ilustración 7. Distribución en capas de computación

Una alternativa más reciente es el almacenamiento en entornos *edge*. Más cercanos a donde se produce el dato y, por definición, cerca del usuario (usando PCs, smartphones, routers...). Estos dispositivos ya los tiene el usuario, por lo que su adquisición no requiere un sobrecoste adicional. Pese a contar con procesadores muy potentes con relación a las tareas de almacenamiento, no son muchos en número, lo que no permitiría asegurar algunas características de gran interés para el almacenamiento, como la replicación.

Hoy en día, los nodos de sensores ya nos rodean en nuestras casas, oficinas e incluso algunos de ellos se encuentran ya en edificios inteligentes, así como en nuestro propio cuerpo en forma de computación vestible. Se contarán en un futuro por miles a nuestro alrededor y han aumentado significativamente sus capacidades computacionales y de almacenamiento en los últimos años.

Los antiguos microcontroladores de 8-bits, todavía en uso, van dando paso a otros que pueden disponer de las capacidades (*stack*) de comunicación requeridas en el propio microcontrolador para abaratar costes, facilitar el desarrollo de soluciones y minimizar su tamaño. Esto debido a que reutilizan parte de su hardware para varios objetivos.

Así, la mayor parte de los nuevos microcontroladores que pretenden hacerse un hueco en este mercado, generalmente tienen arquitecturas de 32 bits y exhiben mayores capacidades de computación y memoria que los anteriores. Esto está posibilitando que el uso de sistemas operativos en tiempo real (RTOS) también se esté generalizando en este ambiente con capacidades computacionales y de almacenamiento tan limitadas.

Siguiendo los pasos de investigaciones previas y nuestro propio trabajo, pretendemos que los sistemas de almacenamiento tradicionales en *fog/cloud* se puedan combinar o incluso reemplazarse por soluciones de almacenamiento *mist/edge* para muchos casos de uso. Entendemos que una nueva solución de almacenamiento específico para este tipo de datos debería contar con varias de las ventajas habituales en el estado de la técnica de almacenamiento actuales, como son la replicación y distribución de los datos almacenados.

Cuanto más bajamos en las capas de despliegue *cloud/fog/edge/mist*, más heterogeneidad encontramos en los nodos de computación de cada capa, tanto a nivel de capacidades de computación como de almacenamiento. Pero también en cuanto a conectividad, tipo de energía para mantener su funcionalidad (p.ej., batería o A/C) o tiempo de funcionamiento (p.ej., muchos nodos sensores/actuadores exhiben un comportamiento dormir-despertar, para optimizar consumos).

Lo que proponemos es, hasta donde alcanza nuestro conocimiento, la primera solución de almacenamiento distribuido en considerar la heterogeneidad del nodo sensor para encontrar la replicación de almacenamiento óptima, según las capacidades de los nodos de las capas *edge/mist*. Ya tenemos desarrollado un algoritmo inicial y una implementación prototipo [18], que nos sirve de antecedente y comienzo del proyecto.

La implementación prototipo creada, pese a no estar completamente optimizada, ya nos ha servido para comparar diversos parámetros. Especialmente, entre el almacenamiento *mist/edge* propuesto y el almacenamiento *cloud*, que usan una gran parte de las soluciones de mercado. En concreto, ya hemos comparado nuestro prototipo con Amazon S3, una de las alternativas más económicas, demostrando ciertas ventajas, incluido el coste de almacenamiento en el tiempo.

Denominamos al proyecto propuesto FairSP4U (*Fair Storage and Processing for Ubiquitous systems*). El objetivo principal de este proyecto es llegar a una solución de almacenamiento distribuido replicado en nodos, aprovechando las capacidades de computación de las capas *mist* y *edge* y para ello, usando los propios nodos y teniendo en cuenta la heterogeneidad de estos. Sobre esta capa de almacenamiento distribuido se propone la creación, como segundo objetivo, de un sistema de procesamiento distribuido que permita que los nodos puedan mejorar sus capacidades, usando las de otros recursos de la red de forma justa.

El proyecto propone para ello el desarrollo y testeo de un framework, para lograr que sea aplicable en despliegues reales. Se espera que permita decrementar el coste recursivo y el consumo de los sistemas usados actualmente para estas labores. Esto requiere tanto las pruebas en entornos reales, que planteamos a modo de casos de estudio, como el estudio del sistema mediante simulaciones, que permitan generar y replicar situaciones para la recogida de datos, análisis y toma de decisiones.

Para el ámbito de este proyecto, acotamos los sistemas de computación distribuidos a aquellos que usan paso de mensajes para la comunicación, careciendo de relojes de sincronización. Así, podemos ver los nodos también como sistemas de computación paralela pero siempre débilmente acoplada [10]. Es decir, el framework

que se construirá deberá lograr que los nodos colaboren entre sí para llegar a una solución única como composición de varias soluciones parciales. Pero también que compitan para llegar a una solución de la manera más eficiente o a una solución de consenso, por citar algunas posibilidades habituales citadas en la bibliografía.

3.2 INTRODUCCIÓN

Durante la última década, hemos estado observando una enorme adopción de dispositivos sensores/actuadores. Su tamaño y costo se han reducido significativamente, mientras que su almacenamiento y las capacidades computacionales han crecido, como predijo Moore [23], pero no siguiendo estrictamente sus predicciones.

Esta adopción masiva incluye al sector público y privado, y se relaciona con términos comúnmente usados como Internet de las cosas (IoT), *SmartX*, Industria 4.0 o ciudades inteligentes, por citar algunos. De manera más genérica podemos hablar de inteligencia ambiental, computación pervasiva y UbiCom por citar algunos de los términos más habituales. Todos ellos tienen una base común de dispositivos con capacidades de sensorización/actuación e interconectividad, que les permiten comunicarse para trabajar juntos hacia un objetivo común. Ciertos términos conllevan ciertas características, como el de UbiCom, cercano también a otros como el de computación tranquila (*calm computing*), orientado este último, también hacia la transparencia con los sistemas de computación desde el punto de vista del usuario, al que deben favorecer.

Hoy en día, estas redes de sensores están formadas, principalmente, por dispositivos con capacidades de sensorización. Estos nodos se usan para detectar y obtener datos contextuales (p.ej., la temperatura actual en una sala de reuniones). Otros nodos se usan para cambiar las condiciones del ambiente (p.ej., encender/apagar el A/C mediante dispositivos actuadores), y en muchas ocasiones, un punto central (i.e., servidor). Este punto central se usa para el almacenamiento de los datos generados, su procesamiento y por medio de los nodos actuadores, el lanzamiento de órdenes de actuación, automáticas preferiblemente ante ciertas condiciones. Dichas condiciones pueden incluir el contexto adquirido por los nodos sensores de la propia red, servicios externos, etc.

El servidor central también se utiliza para recuperar ciertos datos históricos, cuando sea necesario. Este servidor se ubica en la niebla (*fog*) o en la nube (*cloud*), en las soluciones de almacenamiento tradicionales.

Los entornos en niebla (*fog*) están ganando atractivo para mejorar el comportamiento del sistema al reducir el procesamiento, la comunicación y los requisitos de energía [30]. Sin embargo, centrándonos en el almacenamiento, cuando se requiere que sea redundante y distribuido, debe existir más de un servidor central, al menos tantos como réplicas, si se quiere evitar pérdida por daño físico. Esto aumenta los costes de adquisición y/o mantenimiento cuando somos propietarios de los servidores (*fog*).

Así, en muchos casos, los despliegues se basan en servidores en las nubes de terceros (*cloud*), usándose como servicios para reducir complejidad para el equipo de desarrolladores y el coste de adquisición de servidores. Además, esto mejora generalmente la tolerancia a fallos, que requiere de un mayor número de equipos para alcanzar alta fiabilidad de almacenamiento, algo que se suele ofrecer de manera nativa en estas nubes, de por sí distribuidas.

Así, por un lado, las soluciones de computación en la nube facilitan la tolerancia a fallos, aunque por otro lado, cada capa intermedia añadida reduce esta confiabilidad, tal y como muestran ciertos estudios [20,27], y parece lógico pensar.

Para contrarrestar los efectos que puede tener una incorrecta gestión del almacenamiento de datos, diferentes algoritmos y frameworks se han propuesto por la comunidad y empresas en lo que se conoce como *clustering storage*. Estos pueden funcionar a diferentes niveles de granularidad (fichero, bloque, tupla, etc.). Entre ellos quizás sea Hadoop HDFS [32,38] uno de los más conocidos en sus inicios. Estos frameworks se encargan de almacenar la información en diferentes servidores, estableciendo réplicas distribuidas de los mismos y aumentando la tolerancia a fallos del sistema de almacenamiento. Algunas de las limitaciones de este tipo de frameworks es que están pensados para correr en microprocesadores, no microcontroladores, que son los que proporcionan la capacidad computacional a los objetos en IoT, UbiCom, etc.

Otro problema relacionado con el almacenamiento en la nube es la privacidad, que afecta especialmente a los datos almacenados fuera de las instalaciones del propietario de estos (usuario, organización, etc.). Por razones de privacidad, muchas organizaciones y gobiernos tienen regulaciones específicas sobre los datos que se pueden almacenar externamente. Por lo que el almacenamiento debería ubicarse y mantenerse dentro las instalaciones de la organización, en algunos casos de manera obligatoria y en muchos casos de manera preferente.

Además, HDFS, y muchos otros sistemas de almacenamiento en la nube, se diseñaron teniendo en cuenta grandes volúmenes de datos (archivos, bases de datos, etc.) y no están optimizados para las necesidades de almacenamiento de IoT (p.ej., la información de temperatura necesita solo unos pocos bytes de datos, pero se almacenará muchas veces por hora, no como un vídeo publicado en la web, que ocupará un mayor tamaño). Además, mientras que los entornos en la nube proporcionan una gran capacidad de computación y almacenamiento, el aumento de las distancias implicadas en la comunicación significan empeorar la capacidad de respuesta en tiempo real, perder el control sobre dónde se encuentran geográficamente ubicados los datos, sobrecargar la red de datos y un alto coste [3,8,20].

Podemos concluir entonces, que las soluciones basadas en entornos *cloud* imposibilitan que se cumplan los requisitos de algunas aplicaciones IoT actuales, tal y como ha sido estudiado y demostrado [1]. Otros investigadores también han reportado problemas relacionados con costes, escalabilidad, latencia y tiempos de respuesta [34], que pueden llevar a categorizar ciertos casos de uso, como inviables. Estos estudios no difieren de los estudios previos que nosotros mismos hemos observado, donde la

latencia de red en entornos *edge/mist* combinados es entre un 40% y un 750% mejor que en alternativas *cloud* habituales [18].

El paradigma de la computación en la niebla (*fog computing*) [25] supera parcialmente este desperdicio de recursos, al agregar los datos al plano de red, reemplazando así el enfoque común actualmente de computación en la nube. La localización de estos medios de computación evita que los datos salgan del entorno del área local donde se producen los datos.

Hay varios enfoques centrados en diferentes aspectos de este plano de datos: en la definición de topología de la red para el correcto almacenamiento de datos y replicación [17,34], en referencia al equilibrio de carga para hacer frente a la sobrecarga de los nodos y la red [17,34], o en proporcionar un almacenamiento distribuido robusto [7,26].

Se han implementado diferentes técnicas de gestión de datos para apoyar el paradigma de *fog computing*, con relación a la propuesta de este proyecto. Entre ellos, podemos destacar la agregación de datos jerárquicos [37]. La computación y el almacenamiento en entornos *fog* puede incluso facilitarse mediante el uso de tecnologías de virtualización (p.ej. [42]), que permitirían mejorar ciertos aspectos, como el coste de mantenimiento de estos sistemas o los tiempos de adquisición de datos al usar almacenamiento en computadores con altas capacidades.

Sin embargo, se siguen sin poder tener en cuenta los despliegues *fog* en este tipo de entornos *UbiCom*, ya que generalmente dificultan la viabilidad económica del proyecto por el crecimiento de costes que suponen.

Por un lado, como vemos existen problemas con el uso de almacenamiento en *cloud* o *fog*, que sería necesario mejorar. Por otro lado, la mayor parte de capacidades de computación y almacenamiento de las capas *edge* y *mist* se desperdician actualmente [13].

Edge computing es un paradigma diseñado para aprovechar las capacidades de computación y almacenamiento de los nodos, *hubs*, *gateways* disponibles en las capas intermedias, entre los entornos *fog/cloud* y *mist* [9]. Del mismo modo, *mist computing* es un paradigma prometedor para explotar las capacidades computacionales de los dispositivos de gama baja. Y con ello, mejorar el cumplimiento de los requisitos de las aplicaciones que requieren de un procesamiento muy cercano de los datos.

Fijándonos en el procesamiento de los datos sensorizados, cuando este se lleva a entornos *cloud* o *fog*, también aparecen varios de los problemas ya comentados para el almacenamiento. Al fin y al cabo, el procesamiento requiere de datos y si estos están alejados de la red, aumentan ciertas variables asociadas a la distancia, como la latencia, el consumo de energía y hace que aparezcan problemas relacionados con la privacidad, como ya comentábamos, para el caso de almacenamiento de datos en entorno *cloud*.

Algunos de los problemas antes mencionados, de computación y almacenamiento en la niebla y en la nube podrían superarse, o al menos mejorarse, utilizando las capacidades de almacenamiento y computación de *edge* y *mist*. Con ello

se mejoraría el tiempo de respuesta, el conocimiento de la ubicación, la sobrecarga general de la red y los costes de implementación [36].

Actualmente, cada despliegue de una red de sensores en un entorno UbiCom supone el estudio de muchas alternativas diferentes y posibles soluciones durante las fases de análisis y diseño. Desde algunas de bajo nivel, que pueden conllevar restricciones en niveles superiores, como p.ej., qué microcontroladores, módulos de comunicación y protocolos o topologías de red son viables o más adecuados para un proyecto en particular. Hasta otras con un mayor nivel de abstracción, como cuando se despliega una red para toda una ciudad (SmartCity), en el campo (SmartAgriculture) y que puede tener implicaciones a más bajo nivel, como que los nodos funcionen con baterías, debido a la falta de electricidad en el área de despliegue o por simple comodidad durante este.

Estas y otras condiciones establecen los requisitos e imponen limitaciones a la WSN, que el diseñador debe proponer para el proyecto en particular. Incluso cuando los diseñadores pueden utilizar componentes estándar y/o de última generación, la forma en que se mezclan todos estos componentes y las restricciones impuestas por el contexto de implementación conduce, muchas veces, a soluciones WSN a medida / ad-hoc.

En concreto, y en relación directa con este proyecto, hay muchos casos de uso diferentes que requieren diferentes tipos de almacenamiento, según los requisitos concretos de cada despliegue. Lo mismo ocurre con el procesamiento, que puede tener ventajas al ser distribuido y resuelto por varias unidades computacionales de forma colaborativa o no. Este proyecto no pretende proveer una solución única para todos los casos de uso en UbiCom, sino proporcionar una solución novedosa, actualmente no disponible, que permita mejorar ciertos aspectos que lo conviertan en la más idónea para ciertos casos de uso.

En este proyecto proponemos una solución novedosa para el almacenamiento de datos persistentes y redundantes dentro de la propia red de sensores (entornos *mist* y *edge*). Esta solución aprovecha la reducida funcionalidad que generalmente se da a los nodos de la red (detección / actuación), siendo computacionalmente capaces de mucho más, y aprovecha algunas de estas capacidades agregándoles la posibilidad de almacenar datos de su almacenamiento masivo (p.ej., su propio almacenamiento flash o una tarjeta de memoria SD que se puede añadir adicionalmente, a un coste muy reducido).

La solución que proponemos requiere la compra de hardware adicional, en el peor de los escenarios, cuando el nodo no dispone de almacenamiento secundario o se desea aumentar las capacidades de este. Solo en esos casos se requiere añadir un lector y una memoria de tarjeta SD, pudiendo utilizarse siempre el hardware de almacenamiento nativo si ya estaba disponible. Aun así y en base a nuestros trabajos previos [18], este hardware es de uso masivo y muy asequible de adquirir, pudiendo subsistir la información del nodo durante varios años en este tipo de almacenamiento. En ocasiones puede que más allá de la vida útil planeada para el propio objeto IoT.

Tabla 3. Tiempo requerido para llenar una memoria de 8GB

Número de réplicas almacenadas (con entrada de datos desde todos ellos)	Tiempo requerido para llenar la memoria	
	1 entrada por minuto	1 entrada por segundo
3	605 años	10 años
5	363 años	6 años
7	259 años	4 años

La Tabla 3 muestra nuestras estimaciones de almacenamiento con réplica en una memoria de 8GB, extrapolables a otros tamaños de almacenamiento. Teniendo en cuenta que cada dispositivo de sensorización mandará la información en intervalos distintos, es difícil calcular el tiempo de llenado de memoria sin conocer el caso de uso específico al que se aplicará. Pero siendo una memoria no excesivamente grande, muy barata y vistos los datos de la tabla atendiendo al número de réplicas, parece un tipo de almacenamiento viable a estudiar.

En cuanto a procesamiento, no se requieren capacidades adicionales. Los algoritmos se ejecutan sobre los microcontroladores disponibles, de manera independiente o coordinada, usando más o menos tiempo para su ejecución. En cualquier caso, siempre se pueden añadir nuevos nodos con alta capacidades computacionales a una red de nodos existente si se considera necesario, aunque no es la visión habitual que proponemos.

Definimos nuestra propuesta como “justa” porque los nodos seleccionados para almacenar la información y procesarla se elegirán en función de sus características y capacidades, no de modo predeterminado ni aleatorio, sino en base a los nodos disponibles en la red en cada momento. Por lo tanto, entre dos dispositivos idénticos, tanto en comportamiento como en hardware, si dotamos a uno de ellos de mayor capacidad de almacenamiento que al otro dispositivo, el primero tendría más probabilidades de ser seleccionado para almacenar información que el segundo según nuestra propuesta.

Del mismo modo, un nodo que use baterías de poca capacidad debería ser menos probable que fuera elegido, ya que aumentar su consumo podría llevar a que dejase de estar disponible en un menor espacio de tiempo. Además, las capacidades de los nodos son dinámicas y pueden cambiar durante la vida útil del nodo, siendo la salida del algoritmo de selección de almacenamiento y procesamiento propuesto afectado por estos cambios.

En base a nuestros estudios previos, nuestra propuesta reducirá el costo de almacenamiento en comparación con las soluciones típicas en la nube, tal y como mostramos en la Ilustración 8. Además, no sobrecarga la red de mensajes y agrega reconocimiento de ubicación de forma implícita. La figura corresponde a un despliegue de un caso de uso de UbiCom concreto con 5 réplicas, para conseguir una tolerancia a fallos superior a la ofertada por Amazon.

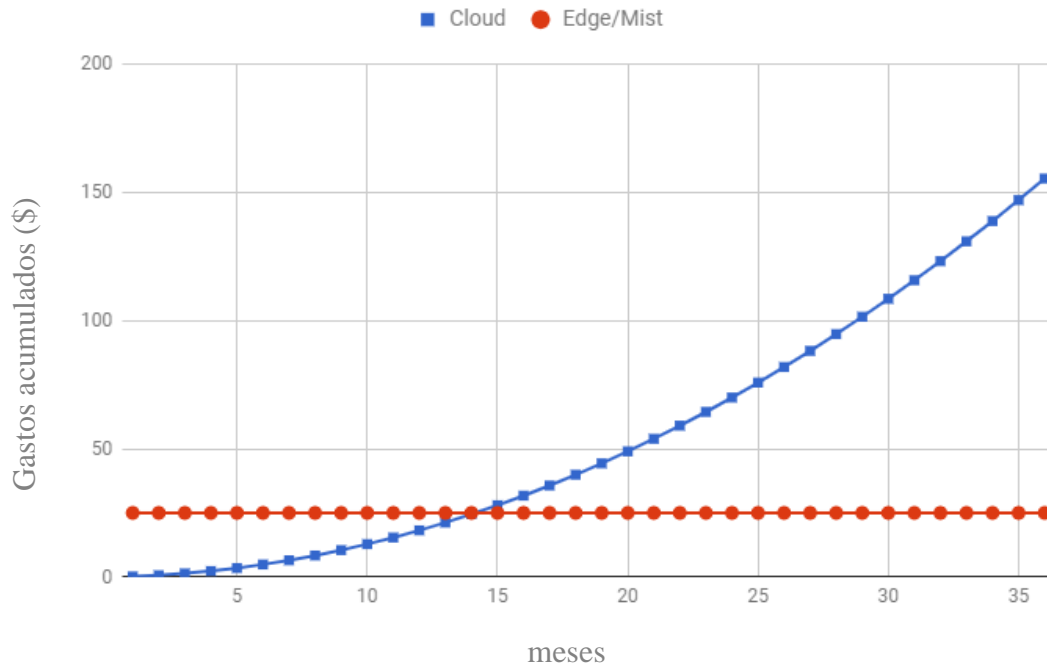


Ilustración 8. Relación de costes entre el almacenamiento en Cloud y el almacenamiento Mist/Edge propuesto

Como puede observarse, pese a haber un coste inicial mayor en la solución propuesta, esta tiene un comportamiento lineal constante (durante toda la vida útil del hardware), mientras que el crecimiento de coste de almacenamiento crece de manera exponencial. Esto hace que, a partir de los 14 meses, algunos casos de uso habituales se vean recompensados desde ese momento por un decremento del coste del sistema ubicuo en funcionamiento. Lógicamente, cada despliegue tendrá sus necesidades de hardware que pueden llevar a que se alcance el momento de ahorro de costes de ejecución, antes o después.

Otra ventaja del proyecto propuesto es que proporciona un alto nivel de privacidad. Debemos tener en cuenta que los problemas de privacidad de datos de sensores pueden llevar a que la información de un sensor se puede utilizar para inferir p.ej., cuándo no hay nadie en una casa o en la oficina, etc.

La última ventaja que queremos destacar del proyecto es que permite establecer niveles personalizables de redundancia y tolerancia a fallos. Esta característica, no se puede personalizar, al menos en grano fino ni de manera directa, en entornos *cloud*. Esto puede hacer que estos niveles sean excesivos para cierto conjunto de casos de uso e insuficientes para otros. En ocasiones pueden incluso conllevar un coste implícito que sobre el que no tenemos margen de actuación.

En la Ilustración 9 se detalla el comportamiento en cuanto a tolerancia a fallos del sistema que proponemos y el número de réplicas, que no podemos controlar en entornos *cloud* y por ello es constante (el dato mostrado corresponde al servicio concreto de almacenamiento *cloud* ofertado por Amazon S3). Como vemos, a partir

de los datos de la tabla, no es necesaria una gran cantidad de réplicas para mejorar la tolerancia a fallos hardware ofertada por Amazon.

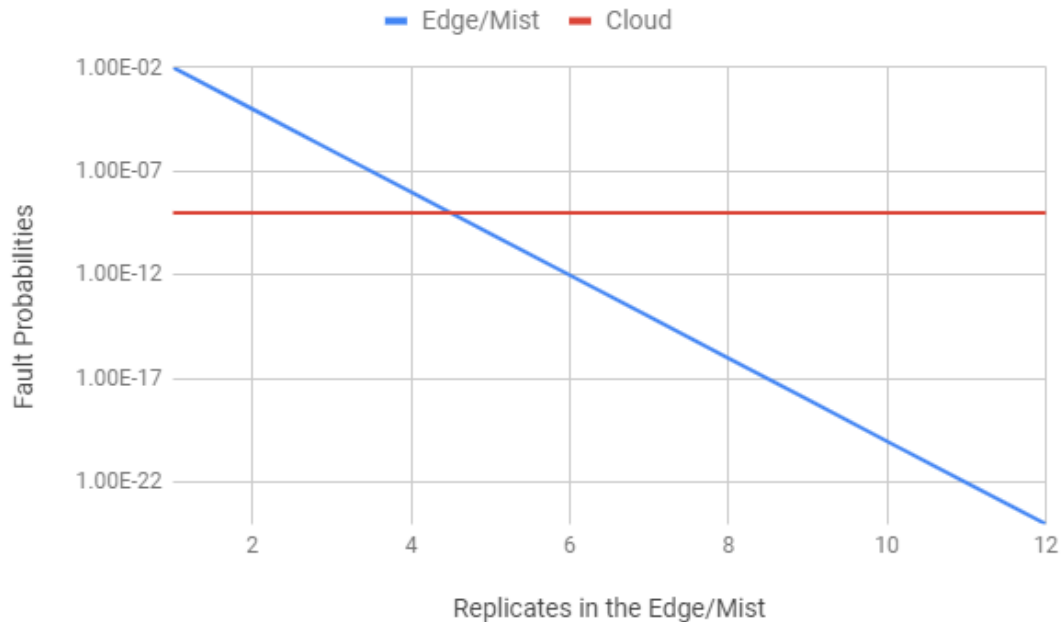


Ilustración 9. Relación entre probabilidad de fallo (hardware) y número de réplicas

Sin embargo, los sistemas pueden verse afectados por las condiciones de su entorno de funcionamiento y lógicamente tiene sentido pensar que el hardware situado en la cima del cráter de un volcán activo, por poner un caso extremo, tiene más posibilidades de fallar que uno desplegado en una zona tranquila. Distribuir datos es una buena práctica para la seguridad de estos, y la propuesta actual es compatible con este tipo de soluciones, pues nada impide combinar la solución propuesta con otros modos de almacenamiento en *fog* o *cloud* para copias de seguridad, etc.

En estudios previos, hemos fijado la formulación matemática tanto del prototipo de sistema de almacenamiento que hemos diseñado e implementado [18], como de distintos parámetros que podrían llevar a diseñar e implementar un sistema similar para la distribución de computación en redes de sensores [29].

En sistemas distribuidos débilmente acoplados, el paso de mensajes es una de las piezas clave. En este proyecto se sigue la tendencia de usar sistemas de mensajería actuales que se ensamblen correctamente con las tendencias actuales de comunicación en UbiCom [21].

En este sentido, los sistemas de publicación/subscripción son, sin duda, pieza clave por lo bien que se adaptan al modo de funcionamiento de los sistemas computacionales empotrados, que exhiben comportamientos de despertar- dormir de forma habitual. Además, estos sistemas de comunicación son parametrizables en varios aspectos. Por ello, hemos decidido asentar este proyecto sobre ellos, ya que actualmente son sistemas de mensajería robustos, tanto a nivel de diseño como de implementación.

Entre las ventajas de usar este tipo de protocolos de comunicación ya establecidos y, pese a que cada protocolo y sus implementaciones puede tener características ligeramente diferentes, podemos nombrar con relación a este proyecto varias de gran interés.

En primer lugar, tolerancia a fallos con relación a errores en la comunicación proveedor-consumidor, que permite a los clientes volver a conectarse automáticamente. Muchas implementaciones cuentan con persistencia de mensajes, lo que hace que, incluso antes de un reseteo del sistema inesperado, los mensajes puedan ser enviados al restaurar el sistema. Como los microcontroladores se caracterizan por unas restauraciones del sistema tras reseteo muy baja, generalmente menor a 1 segundo en los más complejos, esto no debería ser un problema.

En segundo lugar, diferentes niveles de calidad de servicio, lo que permite adecuarse a casos de estudio con nodos que necesiten garantizar la entrega de mensaje y otros que no.

En tercer lugar, ordenamiento, es decir, los mensajes siempre llegan en el orden en que se producen. Esto es importante para p.ej., detección de anomalías en eventos periódicos, etc.

3.3 OBJETIVOS

El objetivo principal es mejorar los costes y consumo energético existentes en el estado del arte y de la técnica en relación con el almacenamiento y computación de/con datos provenientes de redes de sensores, en el contexto de la computación ubicua.

Para ello, fijamos un primer subobjetivo de este proyecto, que es proveer una solución de almacenamiento distribuido replicado en nodos, aprovechando las capacidades de almacenamiento de las capas *mist* y *edge* (teniendo en cuenta la heterogeneidad de los nodos y que sea aplicable en despliegues reales).

Sobre estos datos distribuidos se propone como segundo subobjetivo poder ejecutar algoritmos computados dentro de la propia red que permitan distinta variabilidad en la distribución de la computación, atendiendo a las capacidades computacionales y de comunicación disponibles en la propia red. Estos algoritmos pueden requerir procesamiento paralelo y distribuido si es necesario.

Estos dos objetivos se especifican en una serie de requisitos definidos, que el proyecto debe conseguir completar. Los requisitos fijados inicialmente para el proyecto se recogen en la Tabla 4. Estos requisitos sirven para conocer las características más genéricas que se espera conseguir del sistema de almacenamiento y deben servir, también, para comparar las propuestas existentes dentro del estado de la técnica.

Tabla 4. Requisitos del framework objetivo

ID	Descripción corta	Descripción larga
----	-------------------	-------------------

R1	Computacionalmente simple	Funcional en dispositivos heterogéneos con muy bajas capacidades computacionales y de almacenamiento
R2	Redundante y distribuido	Los datos se deben almacenar de forma distribuida y redundante. El procesamiento debe poder distribuirse colaborativa o competitivamente.
R3	Autogestionado	Capaz de gestionarse a sí mismo sin intervención humana (más allá de la configuración inicial)
R4	Disponible	Al menos un nodo debe estar disponible para almacenamiento y procesamiento de datos
R5	Escalable	Capaz de añadir nuevos nodos sin reconfigurar otros nodos
R6	Adaptativo	Capaz de añadir nuevos sensores y, de este modo, nuevos modelos de datos de forma transparente. También capaz de añadir niveles de QoS para facilitar su adecuación a casos de uso diversos
R7	Mantenible	Capaz de conservar las acciones a realizar y/o aumentar estas de formas sencilla, incluyendo una gestión de red simple
R8	Fiable	Número de posibles puntos de fallo reducido
R9	Seguro	Encriptación de almacenamiento y comunicaciones. Debe prevenir, en la medida de lo posible, ataques para robar información
R10	En tiempo real	Capaz de recoger datos de dispositivos sensores en cualquier momento y almacenarlos
R11	Privacidad	Los datos deben almacenarse en dispositivos confiables pertenecientes al ámbito del nodo que produce el dato
R12	Eficiente	Consumo reducido de energía
R13	Compatible	Capaz de integrarse con sistemas de sensorización y/o actuación existentes
R14	Extensible	Capaz de adaptarse a diferentes tecnologías de implementación y despliegue
R15	Sensible al tráfico	Capaz de priorizar ciertos tipos de tráfico, para no sobrecargar la red con el nuevo tráfico requerido
R16	Localidad	Intentar mantener las réplicas de los datos producidos por un nodo en los mismos nodos si es posible, para mejorar la adquisición de datos en consultas orientadas a computación en el nodo

Otros objetivos que se derivan de los requisitos y que son de interés para el proyecto son: el aumento y/o control del nivel de fiabilidad, privacidad y seguridad en base a los requerimientos de cada caso de uso (despliegue) concreto.

Potencialmente, la localización de datos almacenados de manera distribuida podría empeorar los resultados de la fase 2 y limitarla en alcance y viabilidad. Por ello, durante el desarrollo de la fase 1, se tendrá en cuenta que los datos de un mismo sensor se almacenarán, en la medida de lo posible, siempre en los mismos nodos. Esto evitará que, si los nodos A-H tienen réplicas de un dato sensorizado por el nodo A y se quiere seleccionar un nodo para ejecutar un algoritmo fuera del nodo A, origen de los datos, que no se tengan que enviar los datos necesarios a otros nodos, sino que alguno de estos ya mantenga copia de todos los datos, economizando las transmisiones, sobrecarga de red y consumo energético por las mismas. Esto se recoge en el requisito R16.

3.4 ANTECEDENTES

En esta sección, nos centraremos en revisar el estado del arte sobre almacenamiento y procesamiento distribuido en redes de sensores usando los recursos computacionales y de almacenamiento de las capas *edge* y/o *mist* del sistema.

Tal y como hemos avanzado en la sección 3.2, ninguno de los enfoques *cloud/fog* considera los recursos y capacidades disponibles en los nodos sensores de la red para distribuir los datos sobre ellos de manera justa. Sin embargo, la combinación de ideas y soluciones proporcionadas por varios de ellos pueden ayudar a concretar una aproximación exitosa para solucionar el problema.

Queremos destacar que la bibliografía, al tratar los dispositivos Edge, incluye una amplia gama de dispositivos. Algunos con capacidades computacionales muy bajas (conocidos en la bibliografía, mayoritariamente en inglés como *low-end devices*) y que son los dispositivos objetivo prioritario de este proyecto, hasta servidores informáticos situados en esta capa y que no son de interés para este proyecto, por no tratarse de dispositivos base (requeridos al tener ya una función en la red, como es la de sensorizar, actuar o enrutar comunicaciones) de una red de nodos en UbiCom.

Generalmente, los dispositivos objetivo que tenemos en este proyecto usan microcontroladores en lugar de microprocesadores, con una clara tendencia, como el propio nombre indica, hacia el control más que hacia el proceso. Sin embargo, cuentan con unidades aritmético-lógicas, memoria principal y el resto de las unidades funcionales presentes en los microprocesadores, lo que posibilita que realicen cálculos.

Las capacidades de los microcontroladores evolucionan continuamente y lo hacen en todos sus componentes basados en transistores. Esto hace que mejoren continuamente, tanto a nivel de computación como de almacenamiento, siendo en relación cada vez más económicos y capaces que sus versiones predecesoras y con un consumo energético menor.

El mismo comportamiento lo podemos observar en el más cercano para la mayoría entorno de los smartphones. Por un lado, existe un aumento continuado de la capacidad de cómputo, permitiendo ejecutar aplicaciones más complejas cada año. Por otro lado, las memorias de almacenamiento cada vez tienen mayor capacidad a un precio más reducido, es decir, hay un decremento continuado en el coste/GB.

Tanto para el caso de microcontroladores o microprocesadores como de memorias (especialmente cuando carecen de elementos mecánicos), esto se debe a la escala de integración de los transistores que las conforman, que cada vez son más por área de superficie.

El almacenamiento en redes de sensores se puede categorizar como centralizado o distribuido [19,24]. En las soluciones centralizadas, los datos se envían a un único nodo responsable de almacenar los datos o se envían fuera de la red. En las soluciones que proponen un almacenamiento distribuido, el nodo sensor mantiene generalmente el dato después de haberlo recogido del ambiente y lo envía a otro nodo o a un conjunto de ellos para replicar la información.

Descartamos revisar en esta sección las propuestas con enfoques de almacenamiento centralizado, pues no permitirían conseguir varios de los requisitos fijados. Tras revisar el trabajo relacionado, vemos dos tendencias, la de almacenamiento de datos completamente distribuido y parcialmente distribuido, siendo este último el que nos resulta más útil.

Respecto al almacenamiento de datos completamente distribuido, incluimos en esta categoría las propuestas en las que todos los nodos contribuyen de manera equitativa a generar y almacenar información. Muchas de estas optan almacenar los valores tomados de los sensores localmente, en el propio dispositivo, y si su memoria está llena, delegan esta función en otro nodo. Pese a que eso es un tipo de almacenamiento distribuido, queda lejos de cumplir los requisitos fijados en este documento para la solución buscada.

ProFlex [34] y SUPPLE [5] requieren topologías de red de árbol o malla para funcionar. ProFlex soporta redes heterogéneas con un nodo móvil central, usando un esquema de almacenamiento flexible que crea múltiples estructuras para la replicación de datos. En primer lugar, crea varias estructuras de almacenamiento en forma de árbol, donde los nodos maestros son los que tienen mejores características. En segundo lugar, envía al nodo central de cada árbol el factor de importancia de cada uno de ellos, para el almacenamiento de datos. Finalmente, el nodo central recibe la información del nodo y lo envía a los nodos de almacenamiento. Su principal desventaja es su problema para garantizar la seguridad de los datos.

Por su parte, SUPPLE consta de tres etapas. En la primera, crea una estructura lógica en forma de árbol, donde el nodo central es responsable de recibir datos y replicarlos en la red. En la segunda, asigna pesos a los nodos que representan su probabilidad de almacenar datos. Y en la tercera, los datos se envían al nodo central y la alta sobrecarga del mensaje. Sus principales inconvenientes son el alto consumo de energía del nodo central.

Ninguna de las dos propuestas incluye la característica de peso calculable en tiempo de ejecución, por lo que no se trata de sistemas dinámicos/adaptativos en tiempo de ejecución.

C&R-DS [16] y S&D-DS [28] son enfoques con un énfasis en la seguridad y privacidad de los datos.

El objetivo de C&R-DS es prevenir ataques de robo de información producida por los sensores en redes que constan de tres tipos de nodos: sensores, almacenamiento y maestros. Los sensores envían datos periódicamente a los nodos de almacenamiento mediante comunicaciones encriptadas. Creemos que, por su enfoque, deja muchas otras cuestiones de lado y esta aproximación tendría problemas relacionados, principalmente, con la pérdida de datos y la tolerancia a fallos en los nodos de almacenamiento al intentar integrarla como parte del proyecto que proponemos.

En S&D-DS la seguridad se basa en el uso de una clave secreta compartida y código Reed-Solomon [39]. Este enfoque es resistente a la existencia de nodos comprometidos y utiliza una técnica para verificar la integridad de los datos distribuidos. Los inconvenientes que detectamos en esta investigación son referentes al almacenamiento en sí y a las comunicaciones principalmente, aunque supone una aproximación interesante a nivel de integridad de datos almacenados.

C-Storage [34] y DSforIOT [12] se centran en el balanceo de carga, mejorando el problema de la baja capacidad de memoria en nodos sensores. Actualmente, en nuestro prototipo de almacenamiento que hemos probado, no hemos notado este tipo de problemas, pero más pruebas son necesarias.

C-Storage utiliza técnicas de compresión de datos para almacenar información y transmitirla. Su principal desventaja está relacionada con problemas de seguridad cuando un nodo está comprometido.

La contribución principal de DSforIOT es proporcionar un mecanismo de baja complejidad computacional para replicación de datos de forma distribuida. Sin embargo, el nulo control sobre el número de réplicas y el algoritmo en sí no dan lugar a una alta disponibilidad de datos.

TinyDSM [26] y DSforCDA [40] son propuestas que agrupamos por centrarse ambas en la fiabilidad, proporcionando un almacenamiento distribuido robusto para que los datos siempre se puedan recuperar, a pesar de posibles fallos en el nodo/los nodos de almacenamiento.

TinyDSM garantiza la disponibilidad de datos al introducir datos relativos a la redundancia. Sus principales desventajas son los problemas de seguridad y el equilibrio de carga.

DSforCDA presenta suficiente redundancia para hacer posible la recuperación de datos después de un fallo, añadiendo según los autores una muy pequeña sobrecarga de red. El sistema de gestión de réplicas que propone se orienta a conseguir una alta disponibilidad de datos. Su principal desventaja es que necesita un procesamiento computacional intensivo, lo que dificulta su uso en nodos integrados restringidos y en entornos que requieren resultados instantáneos.

DDAS4AN [15] es un enfoque que define un esquema para seleccionar los datos que deben almacenarse localmente en los nodos de IoT. Con este esquema, los nodos recopilan datos de su entorno en forma de flujos multidimensionales y deciden, en tiempo real, qué datos deben almacenarse localmente para su posterior procesamiento. Para almacenar los datos de manera eficiente, en primer lugar, este enfoque identifica

si el dato detectado es un valor atípico mediante la aplicación de un esquema de consenso. Si el dato no es un valor atípico, el enfoque identifica los nodos donde se debe replicar mediante el uso de un modelo heurístico para identificar los nodos top-k, donde se debe replicar el dato teniendo en cuenta la correlación con los datos ya almacenados por cada nodo. El mayor problema de esta aproximación son nuevamente las capacidades computacionales requeridas en este caos para identificar los valores atípicos y los nodos correlacionados para almacenar la información detectada.

RBCNS [41] define una diseminación local y un algoritmo de reconstrucción basado en regiones. Este enfoque se centra en aquellas situaciones en las que los usuarios solo pueden necesitar información almacenada en regiones donde se detectan los eventos de monitoreo y la recuperación del campo de datos globales no es realmente necesario. Para ello, los autores utilizan una estrategia de difusión local *t-hop* en la que cada lectura del sensor solo se transmite por *t* saltos y no tiene que difundirse por toda la red. Para aumentar la eficiencia del algoritmo, se propone un algoritmo de codificación diferida para explotar la correlación entre las lecturas del sensor, evitando la codificación de las lecturas de regiones irrelevantes. Este enfoque limita la sobrecarga de la red y la codificación de datos. Aunque se propone un algoritmo de recuperación de datos entre regiones, la complejidad de recuperar esta información aumenta. Además, no tiene en cuenta la heterogeneidad de los nodos a nivel de capacidades, tratándolos a todos por igual.

ElfStore [22] es un almacenamiento distribuido / federado de baja sobrecarga para las capas *edge* y *fog*. A modo de comparación con nuestro enfoque, no almacena mediciones únicas de un sensor, sino flujos de bloques de datos en lugar de mediciones individuales. También selecciona los nodos apropiados para el almacenamiento de réplicas automáticamente de acuerdo con su confiabilidad. Se han realizado dos pruebas virtuales para ver el comportamiento del framework, pero ninguna real. Para los autores de estas, un microcontrolador con 4 núcleos es un nodo *edge*. Ningún microcontrolador de gama baja dentro de la mayoría de los productos de IoT es el foco de esta propuesta, que requiere más capacidades computacionales.

Un resumen comparativo en base a los requisitos fijados por este proyecto se incluye en la Tabla 5 donde se observan en muchas ocasiones el texto ND (No Disponible), ya que la implementación no se ha podido probar, al no proporcionarse en el estudio citado o por no estar ya disponible principalmente. En ciertos casos y sin ese código fuente o al menos ejecutable, damos por ciertas las afirmaciones que hacen los autores en los artículos de investigación correspondientes, aunque es imposible replicar sus resultados para contrastarlos.

Tabla 5. Resumen comparativo del estado del arte con los requisitos fijados para este proyecto

REQUISITOS	ProFex	SUPPLE	C&R-DS	S&D-DS	S-Storage	DS for IoT	TinyDSM	DS for CDA	DDAS4AN	RBCNS	ElfStore
R1	Si	ND	ND	ND	Si	Si	ND	No	No	Si	No
R2	No	Si	No	Si	Si	Si	Si	Si	Si	Si	Si
R3	No	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
R4	Si	Si	No	No	No	Si	Si	Si	Si	Si	Si
R5	Si	Si	No	Si	Si	Si	Si	No	ND	Si	ND
R6	No	No	No	No	Si	Si	Si	No	ND	ND	ND
R7	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
R8	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si	Si
R9	No	No	Si	Si	No	No	No	No	ND	ND	ND
R10	No	No	Si	No	Si	No	Si	No	Si	Si	Si
R11	No	No	Si	Si	No	No	No	No	ND	ND	ND
R12	Si	No	No	Si	Si	No	No	Si	ND	ND	ND
R13	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
R14	Si	No	Si	No	Si	Si	Si	Si	Si	Si	Si
R15	No	No	No	No	No	No	No	No	No	No	No
R16	Si	Si	ND	ND	No	Si	ND	ND	No	Si	ND

De la tabla podemos sacar varias conclusiones. La primera, que el trabajo relacionado parece al menos estar correctamente correlacionado con los objetivos perseguidos en este proyecto, pues la suma de varias propuestas permite alcanzar los requisitos fijados en este proyecto, excepto R13 y R15. Además, y como así puede observarse, todas las propuestas cubren, al menos parcialmente, muchos de los requisitos. La segunda, queda claro que ninguna solución existente cubre todos los requisitos que hemos fijado en este proyecto.

Existe más trabajo relacionado con este proyecto que se detallamos en [18], pero que entendemos que pese a ser relacionado, no lo está tan íntimamente como el seleccionado para esta sección.

Respecto a la distribución de carga computacional en la propia de red de nodos, la transmisión de información dentro de esta impone restricciones que algunos algoritmos no pueden salvar para mejorar ciertos aspectos de calidad de servicio. Ya en [33] se identifica la transmisión como uno de los potenciales problemas de la distribución computacional y se apuntan problemas como el disparo en el consumo energético, que tiene un papel muy relevante como objetivo de nuestra propuesta.

En [2] se analiza la mejora potencial del rendimiento obtenida al balancear la carga computacional en una red de sensores. En el estudio se muestra que enviar los datos generados por cada nodo sensor a través de múltiples rutas, en lugar de una sola ruta, permite una conservación de energía significativa. Propone un nuevo modelo analítico para sistemas de balanceo de carga, que se complementa con la simulación para evaluar cuantitativamente los beneficios de la técnica propuesta.

En [4] se evalúa la utilidad de distribuir un algoritmo muy simple y habitual como son la adición de los códigos de corrección de errores. Si bien esto es un caso muy concreto de lo que pretendemos en este proyecto, lo interesante de la aportación de esta propuesta va más allá, al incluir una perspectiva energética, estudiando la energía consumida en la codificación-decodificación y la transmisión de datos en relación con la energía ahorrada. Los resultados muestran que diferentes configuraciones permiten ahorrar entre un 15% y un 60% de la energía del nodo distribuyendo la carga computacional.

En esta misma línea en [31] se presenta un algoritmo distribuido para encontrar la ubicación de un vehículo. Demuestran que distribuyendo el cálculo se puede mejorar enormemente el consumo de energía. Para ello, estudian dos enfoques, uno en el que el procesamiento se realiza donde se producen los datos y luego transmite el resultado al sumidero/concentrador y otro, en el que se transmite una mayor cantidad de datos y se procesan en el concentrador/sumidero. El trabajo muestra que el consumo de energía depende no solo de la cantidad de datos y de carga del proceso, sino también y de manera significativa de la calidad de los recursos de comunicación del sensor. Encontramos bibliografía similar, en la que el consumo de energía de la comunicación parece ser un parámetro determinante en las transmisiones en redes de sensores. No se proporcionan frameworks para el soporte genérico de distribución de carga computacional, quedando limitado el estudio al caso de uso presentado.

El trabajo relacionado encontrado cuenta también con el estudio de ciertos algoritmos que, por su alto potencial de paralelización, se ven beneficiados al ser distribuidos. En [6] se propone un sistema de computación distribuida colaborativa en redes de sensores con energía limitada. Computacionalmente, los nodos tienen baja capacidad de procesamiento digital de señales. Los autores detectan que el tiempo de cálculo disminuye con la distribución, pero llega a un punto en el que, al aumentar el número de nodos, el tiempo de cálculo empieza a aumentar, perdiendo eficiencia. Los autores detallan que, en cuanto al consumo de energía, cuanto mayor es el número de nodos computacionales y las comunicaciones requeridas, mayor es el consumo de energía para cada nodo. En este trabajo de investigación, el estudio se limita intencionalmente al algoritmo de Transformada de Fourier y no es extensible a otros algoritmos porque no toma en cuenta diferentes grados de distribución.

En [14] se presenta un trabajo más genérico, donde los autores analizaron algunos algoritmos específicos, obteniendo el comportamiento escalable para el tiempo de cálculo de cada algoritmo y el consumo de energía asociado. Se asume que la topología de la red es fija y conocida, ya que la variabilidad en otro caso haría crecer enormemente la combinatoria existente.

Otras propuestas, más que distribuir el procesamiento en sí, realizan procesamiento en el nodo o fuera de él orientado a la mejora del tráfico de datos y consecuentemente del consumo en muchos casos. Los autores proponen en [35] un esquema de comunicación de datos que utiliza el filtro adaptativo de mínimos cuadrados jerárquicos (HLMS). Las técnicas de predicción HLMS que predicen los valores medidos tanto en el nodo sensor como en el concentrador se analizan y luego se requiere que los nodos sensores solo envíen aquellas lecturas que se desvían de la

predicción por un umbral de error. Esta computación puede llevarse a cabo en el propio nodo generador de datos o en su nodo sumidero, logrando importantes ahorros de transmisión y energía al reducir la cantidad de datos enviados por cada nodo. Para el proyecto propuesto, esto llevaría también a una disminución del espacio de almacenamiento requerido.

Quizás uno de los trabajos con más ideas interesantes sea [11], donde se estudia la relación entre computación y comunicación entre el nodo productor del dato y el nodo sumidero. El trabajo se basa en el hecho de que, en muchas situaciones, un agente no está interesado en obtener todos los datos de todos los nodos sensores, sino en recopilar una función relevante de las mediciones realizadas por el nodo sensor (i.e., tipo agregación de datos, como la media, el máximo o el mínimo). El trabajo se centra en funciones donde solo los datos de un sensor son importantes. Su análisis proporciona información interesante sobre la complejidad de los algoritmos que podrían realizar los nodos sensores o el concentrador en relación con las comunicaciones necesarias.

Al igual que para el almacenamiento, incluso cuando ninguno de los estudios anteriores resuelve los problemas fijados por los requisitos del proyecto en cuanto a distribución de la computación, muchos de ellos brindan información y conocimientos valiosos, que se usarán en el proyecto como en momentos puntuales o para respaldar las decisiones.

3.5 EXPERIENCIA PREVIA

El investigador responsable de este proyecto cuenta con experiencia previa para la gestión del proyecto, habiendo sido investigador responsable en otros proyectos previos y habiendo organizado recursos humanos y espacios tanto en el sector privado como en el público. Se destacan a continuación los proyectos actualmente en curso en los que el investigador actúa como responsable, para subrayar la experiencia en gestión en la actualidad:

- AMALVE – Desarrollo de algoritmos *machine learning* para el estudio y predicción de los efectos del uso del vehículo eléctrico sobre la producción energética y el despliegue inteligente de infraestructura de recarga
Entidad financiadora: GAMMA SOLUTIONS S.L.
Anualidades: 2019-2020
Identificador: 030/20
Cuantía: 12.250€
- WIOT – Diseño de un sistema IoT para la monitorización del comportamiento de los vehículos eléctricos para mejorar la eficiencia energética
Entidad financiadora: WENEA MOBILE ENERGY S.L.
Anualidades: 2019-2020
Identificador: 031/20
Cuantía: 6.000€

- VOCINA: desarrollo de VOcaciones Científicas y técnicas en Nativos digitales
Ámbito: Estatal (Entidad financiadora: Fundación española para la ciencia y la tecnología)
Identificador: FCT-18-13562
Cuantía: 54.335€(subvencionado 18.000€)

En los tres proyectos citados anteriormente, existe relación con el uso de microcontroladores en entornos relacionados con UbiCom. En AMALVE colaboro en el desarrollo de algoritmos que faciliten la toma de decisiones. Este sistema de toma de decisiones recibe datos provenientes, entre otros, de los sensores embebidos en las estaciones de recarga.

En WIOT colaboro en el diseño y desarrollo de un dispositivo que permita recoger ciertos parámetros de un coche eléctrico en funcionamiento. Estos datos se envían al móvil y de ahí a Internet, para proporcionar consejos al conductor. Principalmente, los consejos se orientan a la recomendación sobre modificación de rutas hacia estaciones de recarga teniendo en cuenta la optimización de consumo de energía. Para ello, se hace necesario conocer el estado actual de energía en el vehículo y la previsión de distancia máxima segura a recorrer antes de consumirla.

En VOCINA hemos estamos creando un juego ciber-físico que permite a la gente aprender a programar sistemas ciber-físicos, dotados de una dimensión virtual y otra física, muy en consonancia con UbiCom. Para ello, el juego permite resolver retos en un mundo virtual a través del móvil y, en ciertos momentos, se debe interactuar también sensores y actuadores que proporcionamos en el mundo físico conectados con ciertos elementos electrónicos a un microcontrolador.

También considero interesante resaltar la experiencia en base a los proyectos en los que formo parte del equipo investigador. Solo se muestran aquellos en curso y que tienen relación con la temática del proyecto presentado:

- Contexto situacional arquitectura tecnológica para automatizar la conexión de las personas a los dispositivos inteligentes
Ámbito: Proyecto plan regional
Anualidades: 2019-2022
Identificador: RTI2018-094591-B-I00
Cuantía: 146.859€
- Contexto situacional: una arquitectura de gestión de la información personal para una mejor integración persona-tecnología
Ámbito: Proyecto plan nacional
Anualidades: 2019-2021
Identificador: IB18030
Cuantía: 107.932€
- QoS metrics -in-the-loop for better robot navigation
Ámbito: Proyecto plan europeo
Anualidades: 2019-2020
Identificador: 732410-RobMoSys-MIRoN
Cuantía: 18.000€

Los dos primeros se centran justamente en entornos de UbiCom, donde se intenta resolver de manera transparente al usuario la interconexión y la interacción con dispositivos computacionales. En el tercero, europeo, tratamos con parámetros de calidad de servicio (QoS) para sistemas robóticos, que cuentan habitualmente con sensores y actuadores además de procesadores y microcontroladores. QoS es una dimensión que queremos añadir a este proyecto, ya que nos permitirá ser viables en un mayor número de casos de uso.

Además, he participado en varios proyectos y contratos con empresas relacionados con microcontroladores, que se encuentran ya finalizados, en algunos de ellos como investigador responsable. Todos ellos aparecen recogidos en el CVN para su consulta.

Así mismo, soy miembro del Smart Open Lab²⁷, uno de los FabLab de la UEx. Está situado en la Escuela Politécnica, lo que me facilita el acceso a herramientas de prototipado en cualquier momento. Proyectos como VOCINA se desarrollan físicamente en sus instalaciones.

En cuanto a la temática objetivo del proyecto, el investigador responsable ha participado en diversas publicaciones relacionadas con la temática descrita en este proyecto. En concreto, y a modo de resumen, se proporcionan dos de las más recientes y que tienen relación muy directa, siendo antecedentes, con el proyecto presentado. Ambas se encuentran publicadas en revistas indexadas en JCR, en el primer cuartil (Q1):

- [18] *Mist and edge storage: Fair storage distribution in sensor networks*
M Linaje, J Berrocal, A Galan-Benitez
IEEE Access 7, 123860-123876, 2019
Indicios de calidad: factor de impacto JCR-Web of Science 3.745
(posición 35 de 156, Q1), factor de impacto 0.775 (Q1) SJR-Scopus
- [29] *A framework to design the computational load distribution of wireless sensor networks in power consumption constrained environments*
D Sánchez-Álvarez, M Linaje, FJ Rodríguez-Pérez
Sensors 18 (4), 954, 2018
Indicios de calidad: factor de impacto JCR-Web of Science 3.031
(posición 15 de 61, Q1), factor de impacto 0.592 SJR-Scopus

No se detalla más información en esta sección sobre ellas, pues han sido referencias ya a lo largo de este documento en varias ocasiones.

Además, el investigador responsable tiene experiencia demostrable por trabajos recientemente publicados sobre generación de código para microcontroladores. Es algo que seguramente exploremos en este proyecto, con el objetivo para facilitar su uso y hacerlo transparente a los desarrolladores de firmware. En este caso la

²⁷ <https://www.smartopenlab.com/>

publicación se ha realizado en un prestigioso congreso, donde además se hizo una demo que dio lugar a otra publicación adicional:

- *Artifact Abstract: Deployment of APIs on Android Mobile Devices and Microcontrollers*
S Laso, M Linaje, J Garcia-Alonso, JM Murillo, J Berrocal
2020 IEEE International Conference on Pervasive Computing and Communications (PERCOM)
Indicios de calidad: clase 1 y rating A+ en *GII-GRIN-SCIE (GGS)*
Conference Rating. También catalogada como A++ en *CORE*.

A esta experiencia me gustaría añadir que, actualmente, dirijo 2 tesis doctorales estando una de ellas muy relacionada con la temática objetivo de este proyecto. Se trata de la tesis de David Sánchez, que como primera contribución dio lugar al artículo “*A framework to design the computational load distribution of wireless sensor networks in power consumption constrained environments*” que es una de las bases para la caracterización de distribución de computación y consumos de este proyecto.

El proyecto parte además de un prototipo diseñado e implementado de almacenamiento en redes de sensores que hemos licenciado como código abierto y que se encuentra disponible online²⁸ para que la comunidad de investigadores y las propias empresas interesadas puedan aportar mejoras sobre él. Este diseño ya ha sido evaluado y mejorado tras la revisión por terceros y se espera obtener más críticas durante el desarrollo del proyecto actual a medida que otros investigadores vayan usando ese trabajo previo.

3.6 METODOLOGÍA Y ACTIVIDADES

Se prevé utilizar una metodología de desarrollo en 2 fases, que en caso de tratarse de un proyecto de mayores dimensiones podrían tomarse por paquetes de trabajo. Una primera fase (fase A) para la resolución del problema de almacenamiento distribuido y la recuperación de estos datos a posteriori en redes de sensores en base a los objetivos y requisitos fijados en la sección 3.3 (Objetivos). En una segunda fase (fase B) se resolverá el problema de procesamiento distribuido en base a estos mismos requisitos sobre el sistema de almacenamiento distribuido.

La fase B es más larga que la fase A por tres motivos. En primer lugar, no se parte de una implementación previa, como es el caso del almacenamiento que cuenta con una versión prototipo ya realizada. En segundo lugar, el aumento de la complejidad del sistema, ya que se pretende en la distribución computacional se realice sobre el sistema de almacenamiento de la fase A.

Como consecuencia de esto y de la propia naturaleza de los sistemas informáticos, entendemos que esto nos llevará a modificar, esperamos que levemente, el sistema de almacenamiento al encontrar bugs, optimizaciones, etc. que no habían

²⁸ <https://github.com/algalanb/EM-Project>

surgido en la fase A. Por eso especialmente las actividades de implementación son más largas. Esto equivale a la fase de mantenimiento que no estaba contemplada dentro del desarrollo específicamente y que es inherente a cualquier sistema.

Cada fase tiene hitos, que se alcanzan al terminar el desarrollo de cada caso de uso. No se especifican los casos de usos concretos en este momento, pues se considera hacer de dos modos: si alguna empresa está interesada en la tecnología, seguramente lo haga para intentar decrementar sus costes en un despliegue real y entonces ese sería uno de los casos de uso a contemplar.

En base a su complejidad, se podría dividir en 2 o 3 casos de uso con complejidad incremental. En caso de que no haya empresas interesadas o haya casos de uso sin cubrir, se buscarán foros de publicación (congresos y revistas) sobre temas concretos que se puedan usar en los casos de uso para maximizar la difusión del proyecto.

Se propone una planificación mensual, con alto nivel de abstracción, para un proyecto de 6 años (72 meses), sin saber el equipo que trabajará en el proyecto y por lo tanto su formación y rendimiento, por lo que es una planificación aproximada. La planificación de desarrollo del sistema sigue la metodología de desarrollo en cascada con sus típicas actividades de desarrollo (análisis, diseño...). En este caso se propone que sea además iterativa incremental en relación con cada caso de uso. Por debajo del mensual, se propone usar metodologías de desarrollo ágiles que usen *sprints* semanales, tipo *scrum* o similares, que se acordarán con el equipo de trabajo en base a su experiencia previa y preferencias.

Durante las fases del desarrollo iterativo incremental se recogerán medidas tanto de pruebas en entornos reales como mediante simulaciones que permitan generar y replicar situaciones para la recogida de datos, análisis y toma de decisiones. En concreto, proponemos en este proyecto medir 5 aspectos (KPI, *Key Performance Indicator*) que detectamos como especialmente relevantes en este tipo de contexto:

- KPI1: uso de memoria de almacenamiento y coste de este en consecuencia
- KPI2: tolerancia a fallos
- KPI3: sobrecarga de red
- KPI4: consumo de energía
- KPI5: latencia

Entendemos que estos 5 aspectos/KPI son claves para la valoración por parte de personal técnicos a la hora de realizar despliegues en entornos reales.

Actividades contempladas para el correcto desarrollo y gestión del proyecto propuesto:

- Gestión de proyecto (A01 y B01): existen dos actividades, una para cada fase del desarrollo, A y B. La gestión será encargo del investigador responsable.
- Formación (A02 y B02): para la formación específica del equipo en materia de desarrollo. Muy especulativa a priori a falta de conocer el CV del personal contratado, aunque con confianza pues en EPCC se sigue

enseñando C/C++ en los Grado de Informática y este es un lenguaje básico para el manejo de microcontroladores. Adicionalmente necesitarán otros lenguajes. La formación la guiará el equipo investigador.

- Estado de la técnica (A03 y B03): para la formación del equipo en el estado de la técnica actual. Se propone que sean ellos quienes realicen la búsqueda y la comparación en base a los requisitos fijados y sea el equipo investigador del proyecto quien guíe este proceso. Incluye el estado del arte en relación con los simuladores de redes objetivo, algo que no se ha realizado en este documento.
- Elección CEs (A08y B08): Se pospone esta tarea lo más posible en la planificación con el objetivo de encontrar los casos de estudio con mayor impacto a nivel de difusión, siguiendo los criterios de elección fijados en esta sección.
- Análisis (A04, A10, A15, B04, B10, B15, B20): Tras el estudio de la técnica, se fijarán los requisitos finales de los productos que de obtendrán en cada fase. Pese a que partimos de unos requisitos iniciales ya fijados a modo de propuesta inicial, derivados del trabajo previo ya realizado, conviene actualizarlos teniendo en cuenta también el aumento de las capacidades computacionales de los nodos en el tiempo (al terminar el proyecto, en el más cercano de los escenarios estaríamos en 2027).
- Diseño (A05, A11, A16, B05, B11, B16, B21): incluye propuestas de diferentes niveles de abstracción, desde el nivel de arquitecturas hardware y software a usar, modularidad, algoritmos, reusabilidad de proyectos o partes de proyectos disponibles en repositorios públicos o proyectos de investigación de terceros, etc.
- Implementación (A06, A12, A17, B06, B12, B17, B22): Es la fase de codificación en sí, también contempla el despliegue en dispositivos hardware, así como la codificación para las simulaciones.
- Testeo (A07, A13, A18, B07, B13, B18, B23): pruebas a realizar sobre el sistema, algunas se podrán automatizar y esta actividad incluye este trabajo a realizar.
- Desarrollo CEx (A09, A14, A19): Desarrollo de casos de uso específicos seleccionados, incluyendo como salida la toma de medidas de los KPI seleccionados. Es decir, incluye la codificación, el despliegue, la toma de valores de los KPI y su análisis. Estos valores deben servir para la su publicación, además de como entrada a la siguiente iteración de desarrollo del proyecto.
- Publicación (A20 y B25): Se prevé poder publicar desde que las primeras implementaciones estén disponibles tras la primera vuelta en el desarrollo iterativo del sistema y hasta el final del proyecto, aunque no se descarta lógicamente hacerlo más allá en el tiempo.

3.7 CRONOGRAMA

Se muestra a continuación el cronograma propuesto en relación con las actividades e hitos descritos con anterioridad. Es una aproximación que contempla todas las actividades requeridas y también los periodos vacacionales, que se contemplan como 1 mes por año trabajado. Concretamente, durante el mes 8 de cada año.

No obstante, tanto por normativa de EPCC y UEx sobre cierre de instalaciones, como legislativa sobre el derecho a elegir los periodos de vacaciones por parte del trabajador, así como por la imposibilidad de conocer el mes y año de comienzo del proyecto a priori, harán que esta planificación requiera los ajustes necesarios. En cualquier caso, estos ajustes no alteran el número de horas trabajadas por año.

No se contemplan directamente actividades o periodos de ajuste relacionados con incidencias, como bajas por enfermedad, etc., pero estas están tenidas en cuenta de manera global en las actividades a las que se han sobredimensionado temporalmente para poder absorber estos riesgos.

La Ilustración 10 recoge visualmente el cronograma propuesto para la realización de este proyecto.

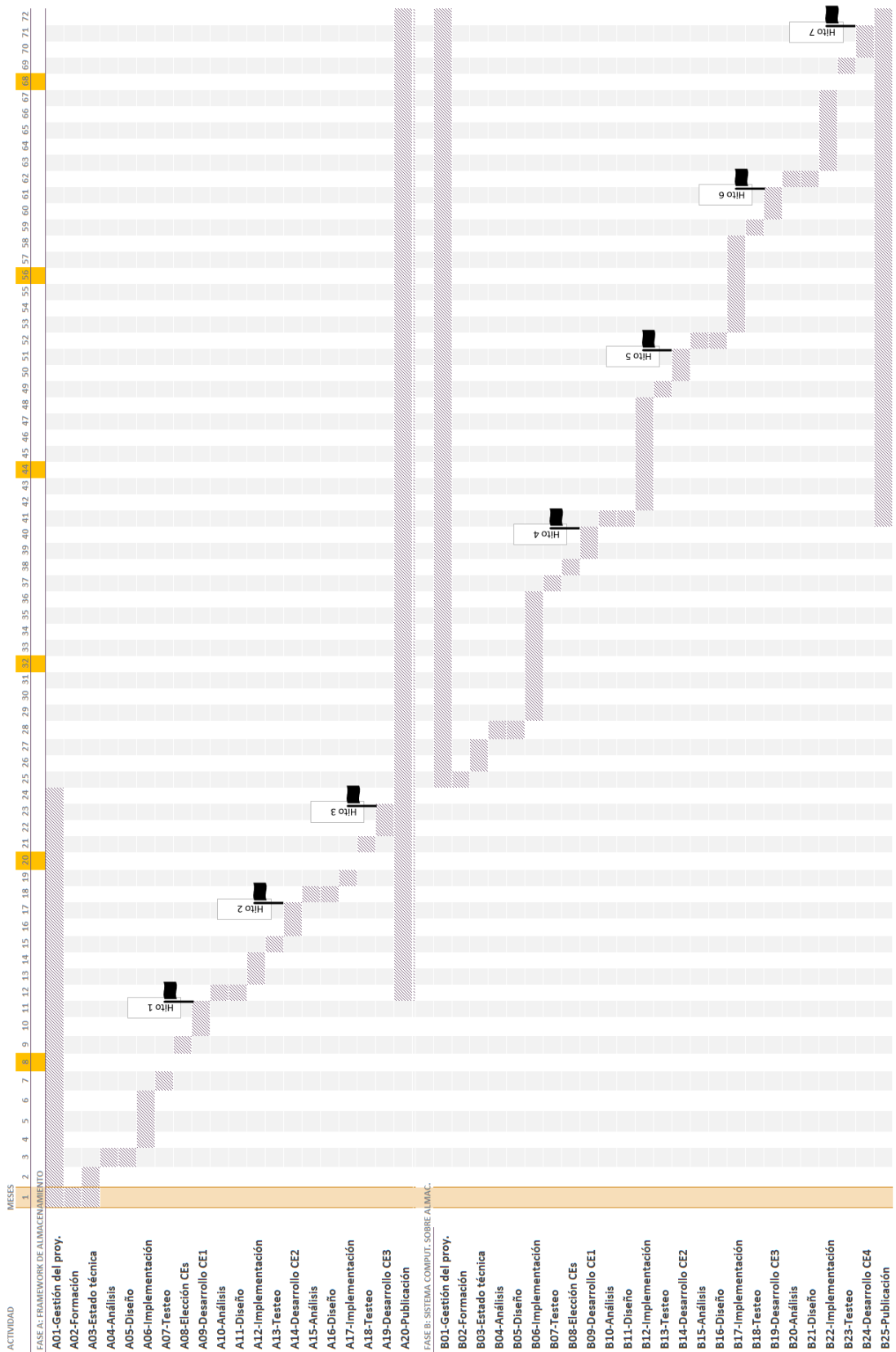


Ilustración 10. Cronograma propuesto

3.8 PRESUPUESTO

Dividimos en esta sección el presupuesto por fases, A y B y partidas: personal, material fungible, material inventariable, publicación, dietas y desplazamientos.

Ciertos gastos, como la adquisición de hardware de sensorización, hubs, etc. para las pruebas y desarrollo de los casos de uso, se incluye distribuido entre las fases A y B atendiendo a la duración cada fase, calculando la amortización de este, no siendo material reutilizable más allá del proyecto.

Tenemos previsto comprar como dispositivos *edge* algunos dispositivos móviles (*smartphones*) que permitan hacer pruebas incluyendo dentro del proyecto propuesto sistemas en movimiento y estáticos. Este tipo de hardware tiene por normativa una amortización de 4 años. Este plazo de amortización coincide con el de otros gastos relacionados con hardware requeridos en este proyecto, como son los equipos de desarrollo (PCs, portátiles, pantallas, teclados, ratones...).

El plazo de 4 años supera el tiempo de proyecto, si se contemplan las fases A y B como un solo proyecto, con lo que podríamos imputar en ese caso el 100% del coste al proyecto. Si cada fase se presentase a dos convocatorias distintas o inicialmente solo se presentase la fase A, si fuera necesario prorratear los gastos de amortización más allá del periodo de ejecución del proyecto para la fase A. En este documento, se hará de esta forma, estimándose el coste para proyectos separados.

Por coherencia en base a como se ha representado el cronograma de la Ilustración 10 se ha prorrateado su coste entre las fases A y B estimando un tiempo de vida útil de 4 años, los 2 primeros sobre la fase A y los dos siguientes sobre la fase B. Se contemplan equipos no solo para el personal contratado en el proyecto, sino también para el personal investigador que los pusiese necesitar.

Los gastos de publicación incluyen los importes por publicar en revistas de acceso abierto (*open access*), así como las cuotas de inscripción en congresos. También los servicios de revisión y traducción requeridos en cada caso. La asistencia a congresos requiere de gastos de desplazamiento y dietas y se contemplan de manera separada en la partida correspondiente.

Los gastos de personal, la partida más cuantiosa, contempla 3 personas año a razón de 40.000€brutos por persona, para 2 años en la fase A y para 3 años en la fase B. Este dato se podría ajustar a medida que se conociese el perfil de candidatos interesados para afinarlo, pues es el coste redondeado de un Técnico de Apoyo según las tablas salariales de la UEx, sin contemplar trienios ni otros gastos similares. Existen proyectos que establecen límites de coste de personal, como los proyectos del plan regional, que requerirían también ajustes en esta partida en caso de concurrir a esas convocatorias.

Tabla 6. Presupuesto desglosado en partidas para la fase A

FASE A	
Partida presupuestaria	Importe
Personal	240.000 €
Material fungible	2.000 €
Material inventariable	15.000 €
Publicación	6.000 €
Dietas y desplazamientos	9.000 €
TOTAL	272.000 €

La Tabla 6 contempla la distribución de las partidas presupuestarias de manera agrupada para la fase A, mientras que las mismas quedan recogidas en la Tabla 7 para la fase B.

Tabla 7. Presupuesto desglosado en partidas para la fase B

Fase B	
Partida presupuestaria	Importe
Personal	360.000 €
Material fungible	2.000 €
Material inventariable	15.000 €
Publicación	8.000 €
Dietas y desplazamientos	12.000 €
TOTAL	397.000 €

3.9 TRANSFERENCIA Y DIFUSIÓN DE RESULTADOS

Se ve difícil que las empresas relacionadas con el almacenamiento de datos de sensores se sientan interesadas en promover el proyecto, pues este es competencia directa hacia los ingresos que obtienen por el almacenamiento recurrente de estos datos como servicio.

Entendemos que pueden estar interesadas en promover esta tecnología empresas que actualmente estén gastando o prevén gastar grandes cantidades por el almacenamiento de datos en la nube (*cloud*) y/o en hardware propio en sus instalaciones (*fog*). Al ser el mayor problema el gasto recurrente de datos históricos almacenados, entendemos que el proyecto será especialmente atractivo para empresas que ya están teniendo estos gastos, así como problemas relacionados con la privacidad en relación con la geolocalización del almacenamiento de estos datos.

Es objetivo del proyecto detectar los sectores con más probabilidades de uso y tenerlo en cuenta, tanto en la búsqueda de potenciales colaboradores. como en la elección de los casos de uso para que les resulte más cercana la idea.

El proyecto se pretende liberar como código fuente abierto y no se descarta que una vez creado el equipo considere la posibilidad de crear una spin-off como soporte al framework. Esta empresa daría soporte, tanto para el desarrollo de modificaciones a

medida para ciertos clientes, como para el asesoramiento en despliegues de redes de sensores y el propio despliegue de estos.

La comunidad de investigación ya ha recibido con buenas críticas varios trabajos del investigador responsable de este proyecto y otros compañeros que probablemente formen parte del equipo investigador del mismo en relación con partes que se prevé usar como punto de partida en el desarrollo propuesto, en especial en referencia al almacenamiento, que es la base del proyecto.

4 RESUMEN, CONCLUSIONES Y AGRADECIMIENTOS

A modo de resumen, podemos decir a grandes rasgos que se ha revisado en este documento el contexto legal y de financiación internacional, nacional y regional para el desarrollo de un proyecto de investigación referente al concurso con código 2019/B/045.

Tener que dedicar tiempo a meditar un proyecto de investigación a varios años vista, como el que se propone, me ha supuesto un interesante reto y me servirá sin duda para orientar más claramente mi investigación hacia un objetivo ambicioso durante los próximos años. En ocasiones, el día a día no nos facilita tiempo para pararnos y pensar en el largo plazo, y he intentado que este proyecto sirva para ello.

Aun sabiendo que se propone una configuración anual y de recursos poco habitual y no se corresponde con una convocatoria de proyecto concreta, se ha realizado atendiendo a la sostenibilidad de la propuesta de investigación en el tiempo. Esto incluye las capacidades reales que podemos tener de contratación, con relación a las posibilidades de captación de personal a largo plazo pensando en tesis doctorales. Este es el tipo de personal que buscamos en mayor medida, pues tienen previsión de tener un recorrido más largo. No se pueden tener muchos perfiles tampoco de este tipo, ya que requieren un alto nivel de supervisión. No se espera que los 6 recursos humanos previstos en la planificación para las fases A y B realicen la tesis, pero si al menos entre 2-3 de ellos.

Mi gratitud a todas las compañeras y compañeros con los que he compartido investigación e inquietudes durante mis casi 20 años de carrera. También, a las personas que me han impartido clases en estos años, tanto técnicas como docentes. Gracias a todas ellas ha sido posible llegar a un momento en el que ha sido posible redactar este documento con el nivel de madurez profesional que he alcanzado.

Dado que no se proporcionan guías concretas para la creación de los proyectos de investigación que se deben redactar, agradezco a Antonio Gordillo y Pedro Núñez su generosidad, al compartir conmigo sus documentos que han sido consultados como ejemplos, a modo de guías.

*Dedicado a todas aquellas personas que me quieren.
Gracias a vosotrxs, soy mejor persona.*

BIBLIOGRAFÍA

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. 2015. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun. Surv. Tutor.* 17, 4 (2015), 2347–2376. DOI:<https://doi.org/10.1109/COMST.2015.2444095>
- [2] J.N. Al-Karaki, R. Ul-Mustafa, and A.E. Kamal. 2004. Data aggregation in wireless sensor networks - exact and approximate algorithms. In *2004 Workshop on High Performance Switching and Routing, 2004. HPSR.*, 241–245. DOI:<https://doi.org/10.1109/HPSR.2004.1303478>
- [3] Paolo Bellavista, Javier Berrocal, Antonio Corradi, Sajal K. Das, Luca Foschini, and Alessandro Zanni. 2019. A survey on fog computing for the Internet of Things. *Pervasive Mob. Comput.* 52, (January 2019), 71–99. DOI:<https://doi.org/10.1016/j.pmcj.2018.12.007>
- [4] Fatma Bouabdallah, Nizar Bouabdallah, and Raouf Boutaba. 2009. On Balancing Energy Consumption in Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* 58, 6 (July 2009), 2909–2924. DOI:<https://doi.org/10.1109/TVT.2008.2008715>
- [5] Aline Carneiro Viana, Thomas Herault, Thomas Largillier, Sylvain Peyronnet, and Fatiha Zaïdi. 2010. Supple: A Flexible Probabilistic Data Dissemination Protocol for Wireless Sensor Networks. In *Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWIM '10)*, ACM, New York, NY, USA, 385–392. DOI:<https://doi.org/10.1145/1868521.1868586>
- [6] C.F. Chiasserini. 2002. On the concept of distributed digital signal processing in wireless sensor networks. In *MILCOM 2002. Proceedings*, 260–264 vol.1. DOI:<https://doi.org/10.1109/MILCOM.2002.1180450>
- [7] Ángel Cuevas, Manuel Urueña, Gustavo de Veciana, Rubén Cuevas, and Noël Crespi. 2014. Dynamic Data-Centric Storage for long-term storage in Wireless Sensor and Actor Networks. *Wirel. Netw.* 20, 1 (January 2014), 141–153. DOI:<https://doi.org/10.1007/s11276-013-0598-5>
- [8] A. V. Dastjerdi and R. Buyya. 2016. Fog Computing: Helping the Internet of Things Realize Its Potential. *Computer* 49, 8 (August 2016), 112–116. DOI:<https://doi.org/10.1109/MC.2016.245>
- [9] Pedro Garcia Lopez, Alberto Montresor, Dick Epema, Anwitaman Datta, Teruo Higashino, Adriana Iamnitchi, Marinho Barcellos, Pascal Felber, and Etienne Riviere. 2015. Edge-centric Computing: Vision and Challenges. *SIGCOMM Comput Commun Rev* 45, 5 (September 2015), 37–42. DOI:<https://doi.org/10.1145/2831347.2831354>
- [10] Sukumar Ghosh. 2014. *Distributed Systems: An Algorithmic Approach, Second Edition*. CRC Press.
- [11] A. Giridhar and P.R. Kumar. 2005. Computing and communicating functions over sensor networks. *IEEE J. Sel. Areas Commun.* 23, 4 (April 2005), 755–764. DOI:<https://doi.org/10.1109/JSAC.2005.843543>
- [12] Pietro Gonizzi, Gianluigi Ferrari, Vincent Gay, and Jérémie Leguay. 2015. Data dissemination scheme for distributed storage for IoT observation systems at large scale. *Inf. Fusion* 22, (March 2015), 16–25. DOI:<https://doi.org/10.1016/j.inffus.2013.04.003>
- [13] Pengfei Hu, Sahraoui Dhelim, Huansheng Ning, and Tie Qiu. 2017. Survey on fog computing: architecture, key technologies, applications and open issues. *J.*

- Netw. Comput. Appl.* 98, (November 2017), 27–42. DOI:<https://doi.org/10.1016/j.jnca.2017.09.002>
- [14] N. Khude, A. Kumar, and A. Karnik. 2005. Time and energy complexity of distributed computation in wireless sensor networks. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, 2625–2637 vol. 4. DOI:<https://doi.org/10.1109/INFCOM.2005.1498546>
 - [15] Kostas Kolomvatsos, Panagiotis Oikonomou, Maria G. Koziri, and Thanasis Loukopoulos. 2018. A Distributed Data Allocation Scheme for Autonomous Nodes. *2018 IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorldSCALCOMUICATCCBDComIOPSCI* (2018), 1651–1658.
 - [16] Kusuma K V and Prasad M R. 2013. Confidential and Reliable Data Storage in WSN. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 3, 4 (April 2013), 148–151.
 - [17] Wen-Hwa Liao, Kuei-Ping Shih, and Wan-Chi Wu. 2010. A grid-based dynamic load balancing approach for data-centric storage in wireless sensor networks. *Comput. Electr. Eng.* 36, 1 (January 2010), 19–30. DOI:<https://doi.org/10.1016/j.compeleceng.2009.04.003>
 - [18] Marino Linaje, Javier Berrocal, and Alfonso Galan-Benitez. 2019. Mist and Edge Storage: Fair Storage Distribution in Sensor Networks. *IEEE Access* 7, (2019), 123860–123876. DOI:<https://doi.org/10.1109/ACCESS.2019.2938443>
 - [19] Xingpo Ma, Junbin Liang, Renping Liu, Wei Ni, Yin Li, Ran Li, Wenpeng Ma, and Chuanda Qi. 2018. A Survey on Data Storage and Information Discovery in the WSNs-Based Edge Computing Systems. *Sensors* 18, 2 (February 2018), 546. DOI:<https://doi.org/10.3390/s18020546>
 - [20] H. Madsen, B. Burtschy, G. Albeanu, and F. Popentiu-Vladicescu. 2013. Reliability in the utility computing era: Towards reliable Fog computing. In *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*, 43–46. DOI:<https://doi.org/10.1109/IWSSIP.2013.6623445>
 - [21] L. Magnoni. 2015. Modern Messaging for Distributed Sytems. *J. Phys. Conf. Ser.* 608, (May 2015), 012038. DOI:<https://doi.org/10.1088/1742-6596/608/1/012038>
 - [22] Sumit Kumar Monga, Yogesh Simmhan, and others. 2019. ElfStore: A Resilient Data Storage Service for Federated Edge and Fog Resources. *ArXiv Prepr. ArXiv190508932* (2019).
 - [23] G. E. Moore. 2006. Cramming more components onto integrated circuits, Reprinted from *Electronics*, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Soc. Newsl.* 11, 3 (September 2006), 33–35. DOI:<https://doi.org/10.1109/N-SSC.2006.4785860>
 - [24] Neenu M. Nair. 2013. Survey on Distributed Data Storage Schemes in Wireless Sensor Networks. *Indian J Comput Sci Eng* 4, (2013).
 - [25] OFC. 2019. *OpenFog Consortium*. Retrieved from <https://www.openfogconsortium.org/>
 - [26] Krzysztof Piotrowski, Peter Langendoerfer, and Steffen Peter. 2009. tinyDSM: A highly reliable cooperative data storage for Wireless Sensor Networks. In *2009 International Symposium on Collaborative Technologies and Systems*, 225–232. DOI:<https://doi.org/10.1109/CTS.2009.5067485>
 - [27] Amir Rahmani, Pasi Liljeberg, Jürjo-Sören Preden, and Axel Jantsch (Eds.). 2018. *Fog Computing in the Internet of Things: Intelligence at the Edge*. Springer International Publishing. Retrieved July 17, 2018 from <http://www.springer.com/gp/book/9783319576381>

- [28] Wei Ren, Yi Ren, and Hui Zhang. 2010. Secure, dependable and publicly verifiable distributed data storage in unattended wireless sensor networks. *Sci. China Inf. Sci.* 53, 5 (May 2010), 964–979. DOI:<https://doi.org/10.1007/s11432-010-0096-7>
- [29] David Sánchez-Álvarez, Marino Linaje, and Francisco-Javier Rodríguez-Pérez. 2018. A Framework to Design the Computational Load Distribution of Wireless Sensor Networks in Power Consumption Constrained Environments. *Sensors* 18, 4 (March 2018). DOI:<https://doi.org/10.3390/s18040954>
- [30] H. Shen, L. Zhao, and Z. Li. 2011. A Distributed Spatial-Temporal Similarity Data Storage Scheme in Wireless Sensor Networks. *IEEE Trans. Mob. Comput.* 10, 7 (July 2011), 982–996. DOI:<https://doi.org/10.1109/TMC.2010.214>
- [31] Eugene Shih, Seong-Hwan Cho, Nathan Ickes, Rex Min, Amit Sinha, Alice Wang, and Anantha Chandrakasan. 2001. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. In *Proceedings of the 7th annual international conference on Mobile computing and networking (MobiCom '01)*, Association for Computing Machinery, New York, NY, USA, 272–287. DOI:<https://doi.org/10.1145/381677.381703>
- [32] K. Shvachko, H. Kuang, S. Radia, and R. Chansler. 2010. The Hadoop Distributed File System. In *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 1–10. DOI:<https://doi.org/10.1109/MSST.2010.5496972>
- [33] Kazem Sohraby, Daniel Minoli, and Taieb Znati. 2007. *Wireless Sensor Networks: Technology, Protocols, and Applications*. John Wiley & Sons.
- [34] A. Talari and N. Rahnavard. 2011. CStorage: Distributed Data Storage in Wireless Sensor Networks Employing Compressive Sensing. In *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 1–5. DOI:<https://doi.org/10.1109/GLOCOM.2011.6134318>
- [35] Liansheng Tan and Mou Wu. 2016. Data Reduction in Wireless Sensor Networks: A Hierarchical LMS Prediction Approach. *IEEE Sens. J.* 16, 6 (March 2016), 1708–1715. DOI:<https://doi.org/10.1109/JSEN.2015.2504106>
- [36] Minoru Uehara. 2018. Mist Computing: Linking Cloudlet to Fogs. In *Computational Science/Intelligence and Applied Informatics*, Roger Lee (ed.). Springer International Publishing, Cham, 201–213. DOI:https://doi.org/10.1007/978-3-319-63618-4_15
- [37] Korosh Vatanparvar and Mohammad Abdullah Al Faruque. 2018. Control-as-a-Service in Cyber-Physical Energy Systems over Fog Computing. In *Fog Computing in the Internet of Things*. Springer, Cham, 123–144. DOI:https://doi.org/10.1007/978-3-319-57639-8_7
- [38] Tom White. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc. Retrieved June 15, 2020 from <http://shop.oreilly.com/product/0636920033448.do>
- [39] Stephen B. Wicker and Vijay K. Bhargava. 1999. *Reed-Solomon Codes and Their Applications*. John Wiley & Sons.
- [40] S. Yulong, X. Ning, P. Qingqi, M. Jianfeng, X. Qijian, and W. Zuoshun. 2011. Distributed Storage Schemes for Controlling Data Availability in Wireless Sensor Networks. In *2011 Seventh International Conference on Computational Intelligence and Security*, 545–549. DOI:<https://doi.org/10.1109/CIS.2011.126>
- [41] s zhou, Y. He, S. Xiang, K. Li, and Y. Liu. 2019. Region-Based Compressive Networked Storage with Lazy Encoding. *IEEE Trans. Parallel Distrib. Syst.* 30, 6 (June 2019), 1390–1402. DOI:<https://doi.org/10.1109/TPDS.2018.2883550>

- [42] OpenStack Open Source Cloud Computing Software. *OpenStack*. Retrieved August 26, 2020 from <https://www.openstack.org/>