

# Propuesta de Proyecto de la Dirección General de Sistemas de Información

## **1 RESUMEN DE LA PROPUESTA**

**DIRECTORA GENERAL DE SISTEMAS DE INFORMACIÓN:** María Inmaculada Medina Bulo.

**TÍTULO DEL PROYECTO:** Implantación de un sistema de inteligencia empresarial en la Universidad de Cádiz y mejora de la calidad de los datos institucionales.

## RESUMEN

Con las crecientes exigencias de distintos tipos de información provenientes de las administraciones autonómica y central, y el despliegue de los nuevos títulos de grado y máster, la Universidad de Cádiz necesita disponer de información contrastada, consistente y actualizada para poder tomar decisiones, cumplir la normativa vigente y elaborar los informes que se le solicitan. Un error en los datos implicados en estos procesos, puede acarrear un gran coste, no solo económicamente, sino porque puede perjudicar a la imagen de nuestra institución y retrasar la implantación de otras medidas de carácter estratégico.

En ese sentido, hace algunos años se diseñó una primera versión de almacén de datos centralizado a partir de las bases de datos de las aplicaciones de Universitas XXI (UXXI) que la Universidad de Cádiz tiene contratadas, con la colaboración de las universidades públicas andaluzas, la Junta de Andalucía y el concurso de una empresa externa (Fujitsu). Para ello se empleó la misma infraestructura tecnológica requerida por UXXI, es decir, el sistema de gestión de bases de datos Oracle. A partir del almacén de datos inicial, se diseñaron una serie de informes mediante Oracle Discoverer, que se utilizan actualmente para dar respuesta a algunas de las necesidades de información más acuciantes de la UCA.

Sin embargo, existen diversos inconvenientes. En primer lugar, se trata de una solución parcial, pues existe una ingente cantidad de datos institucionales ajenos a UXXI, al almacén de datos centralizado e incluso a las propias bases de datos Oracle disponibles, pues existen otras muchas fuentes de información repartidas a lo largo de la Universidad de Cádiz. Algunas de las fuentes externas mencionadas no son bases de datos propiamente dichas, sino meras hojas de cálculo o documentos de otros tipos. En otros casos, los datos que se incorporan están incompletos, son incorrectos o inconsistentes entre sí, por lo que dificultan la extracción de información útil. En segundo lugar, Oracle Discoverer no es la herramienta más apropiada para personas sin conocimientos técnicos y tampoco es lo suficientemente flexible para dar respuesta a la demanda de informes cada vez más complejos. En tercer lugar, el almacén de datos nunca se llegó a implantar en su totalidad, ni se completó la validación de los datos con los que se alimentaba, ni de las consultas a partir de las que se generaban sus informes. Finalmente, no se llegó a desarrollar un sistema de información integrado, ni con el apoyo del almacén de datos existente, ni con el de otros componentes.

Este proyecto describe una serie de acciones que, si se ejecuta adecuadamente con la colaboración de todos los agentes implicados, volverá a poner en funcionamiento la idea de un almacén de datos centralizado, reducirá la dependencia de datos externos, incrementará su calidad y permitirá que se utilice como componente en la construcción de un sistema de información unificado para la Universidad de Cádiz. Las acciones que lo componen se dividen en varios frentes:

- Recogida de datos, con mayor automatización y calidad, y procesos mejorados de filtrado e inserción en las bases de datos institucionales.
- Integración de fuentes de datos ajenas a las bases de datos de Universitas XXI.
- Explotación de datos a través de informes y cuadros de mando sencillos de interpretar.
- Publicación de los datos requeridos por los organismos competentes.

## 2 INTRODUCCIÓN

### 2.1 Antecedentes

En los últimos años, la Universidad de Cádiz (UCA) ha visto cómo sus necesidades de información han aumentado más que nunca. La Junta de Andalucía y el Ministerio de Educación exigen cada vez más informes detallados, con multitud de indicadores de calidad y rendimiento, en los que un error puede acarrear graves perjuicios para nuestra institución. Por otro lado, el despliegue de los nuevos títulos en la situación económica actual exige, si cabe, un mejor uso de los recursos disponibles, lo que requiere a su vez disponer de información útil que permita decidir su distribución más efectiva.

Parte de los datos requeridos trascienden las necesidades de la propia UCA, tal es el objetivo de iniciativas como las del proyecto *Aporta* de la entidad pública *Red.es* y su portal *datos.gob.es*. Este portal representa una de las principales iniciativas nacionales para la reutilización de la información del sector público (RISP) y mantiene un catálogo de diversas fuentes de información de todo tipo procedentes de organismos públicos, como el Ministerio de Educación y Deporte, la CRUE o la Universidad Pablo de Olavide [1], por citar algunos.

Siguiendo estas tendencias, el Plan Estratégico de la UCA (PEUCA) inicialmente elaborado en 2005 recogió en su objetivo 6 la necesidad de «conseguir un sistema de información que permita la toma de decisiones de los órganos de gestión de la UCA y mejora de los procesos de comunicación interna». El proyecto que se propone atiende en particular al apartado 6.1 «analizar los procesos actuales de generación de información y comunicación e implantar un sistema de información de acuerdo con las demandas informativas» y estos subapartados:

- 6.1.1. Potenciar canales de distribución de la información en formato digital.
- 6.1.2. Crear y potenciar puntos informativos cercanos al usuario de las actividades de los diferentes servicios de la UCA.
- 6.1.3. Poner en marcha, validar y mejorar el Sistema de Información para la Dirección (Data-Warehouse).

En una etapa previa, la UCA intentó desarrollar en colaboración con otras universidades andaluzas, la Junta de Andalucía y una empresa externa (Fujitsu) contratada al efecto, un almacén de datos centralizado con tecnologías de Oracle. La idea era desarrollar este almacén a partir de la información disponible en las bases de datos gestionadas por distintos módulos contratados dentro de la plataforma de gestión universitaria *Universitas XXI (UXXI)*, e integrar así los datos disponibles para crear fuentes de datos de nivel superior que explotaran eficazmente.

Sin embargo, tras varios años de desarrollo y un esfuerzo considerable, este proyecto nunca se completó de forma satisfactoria por distintas razones de muy diversa índole. Entre las causas

técnicas, cabe destacar que, salvo contadas excepciones, los datos no habían sido validados por las unidades implicadas, tampoco se incluían fuentes de datos externas a UXXI de vital importancia, y las tecnologías elegidas dificultaban tanto la integración de fuentes de datos heterogéneas como su exportación a catálogos públicos como el de *Red.es*, además de la elaboración de informes detallados y flexibles.

La UCA intentó también mejorar sus sistemas de información mediante desarrollos propios o en colaboración con la Fundación Universidad-Empresa de Cádiz (FUECA). Una de las primeras iniciativas que se llevó a cabo fue SIUCA (Sistema de Información de la UCA, [venus.uca.es/calidad/siuca](http://venus.uca.es/calidad/siuca)). Este sistema implementa una serie de informes mediante una lógica codificada a mano, por lo que cada nuevo informe requiere un esfuerzo considerable. Otra carencia de SIUCA es que solo opera con datos locales, y no se conecta directamente a las bases de datos institucionales, por lo que la actualización de los datos disponibles requiere realizar una serie de tareas manuales.

Por otro lado, SIDIR se desarrolla originalmente utilizando la herramienta empleada en la Oficina de Posgrado ([posgrado.uca.es](http://posgrado.uca.es)) y posteriormente se reutiliza para ofrecer información general a distinto personal de la UCA ([sidir.uca.es](http://sidir.uca.es)). A diferencia de SIUCA, SIDIR puede conectarse a bases de datos y extraer datos de ellas mediante consultas SQL que se mantienen separadas del resto del código. Sin embargo, crear un nuevo modelo de informe requiere conocimientos técnicos y, dependiendo del tipo de informe, las consultas SQL pueden resultar muy complejas y costosas de modificar en caso de que cambien las necesidades de información.

Un problema común encontrado en la mayoría de desarrollos promovidos por la UCA es la ausencia de un proceso de ingeniería en el desarrollo de las aplicaciones informáticas. Este desarrollo artesanal produce como consecuencia la falta de documentación precisa, completa y actualizada, incertidumbres en cuanto a la seguridad de la aplicación y los propios datos<sup>1</sup> y, en ocasiones, incluso la ausencia del propio código fuente y modelos de datos de la aplicación.

Más allá de las aplicaciones desarrolladas internamente en la UCA, y desde hace ya bastantes años, se utiliza la herramienta Oracle Discoverer 10g para generar informes a partir de las bases de datos operacionales. En el último año también se ha utilizado la herramienta para generar informes a partir del almacén de datos de la UCA. Como ya se ha comentado, este almacén fue creado originalmente como resultado de un contrato entre varias universidades y la Junta de Andalucía con Fujitsu. Sin embargo, el proceso de extracción y carga, en lugar de emplear una solución especializada como Oracle Warehouse Builder, consiste en un conjunto de programas codificados a mano en PL/SQL.

Aunque Oracle 10g incorpora diversas funciones útiles para inteligencia empresarial, está algo anticuado en términos de calidad de la interfaz y funciones básicas, ya que es un producto de 2005. Actualmente se está en proceso de migrar los informes basados en Oracle Discoverer 10g a Oracle BI 11g, que es el programa para informes de la Oracle BI Standard Edition. Oracle BI Answers 11g (de la Oracle BI Enterprise Edition) surge de la adquisición de Siebel Analytics por parte de Oracle Corporation, por lo que es un producto completamente nuevo y puede exigir cambios importantes antes de su adopción. Sin embargo, es importante remarcar que su disponibilidad puede depender de si la Junta de Andalucía renueva la licencia para Oracle BI Enterprise Edition 11g.

---

<sup>1</sup>Que, en muchas ocasiones, incluyen datos de carácter personal o especial protección.

Un problema recurrente era, y sigue siendo, la baja calidad de los datos disponibles. La necesidad de mejorar su definición, recogida y proceso es clara si se ha de evitar que falten datos clave, que sean incorrectos o, simplemente, inconsistentes entre sí. Sin un proyecto que considere íntegramente la calidad de los datos como uno de sus objetivos fundamentales, será difícil garantizar la credibilidad de la información proporcionada por nuestra institución y que las decisiones estratégicas y de gestión que se tomen en su seno estén bien fundamentadas.

Esto exige el concurso de todos los agentes implicados en la gestión de los datos institucionales y su compromiso con las políticas establecidas al efecto, aunque la tecnología puede jugar un papel muy importante en la solución de los problemas descritos, como se mostrará a continuación.

## 2.2 Calidad de los datos

El conocido dicho «garbage in, garbage out»<sup>2</sup> resume la importancia de disponer de datos de calidad si se desea obtener información útil y fiable: si a un sistema se le suministran datos pobres, la información que se extraerá de él será, a buen seguro, también pobre. Este problema fue identificado ya en los inicios de la Informática: de hecho, la frase apareció por primera vez en prensa en un artículo del 1 de abril de 1963 de la Indiana Gazette.

El incremento de la frecuencia con la que es necesario interconectar y reutilizar los datos hoy día, ha acentuado la importancia de disponer de datos precisos y fiables. Wang, Strong y Guarascio [2] definieron y clasificaron jerárquicamente qué atributos debían poseer los datos de calidad. Esta clasificación, representada en la figura 1, ha servido como punto de partida de numerosos trabajos. Para ser de calidad, los datos no solo deben ser fieles a la realidad, sino que deben cumplir otras propiedades, como ser fiables respecto a su origen, relevantes en el contexto en el que se recogen, comprensibles, fáciles de procesar y seguros, entre otras.

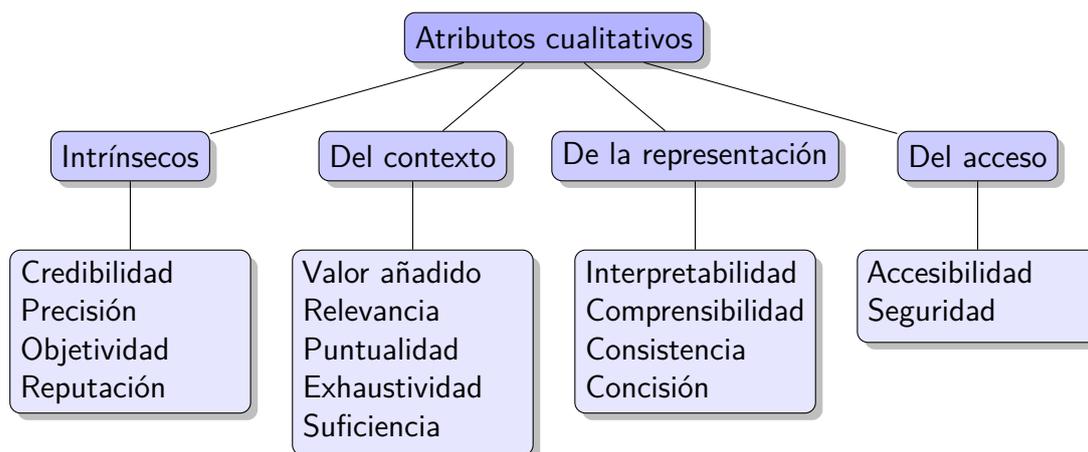


Figura 1: Atributos de calidad de los datos, según Wang, Strong y Guarascio [2]

Won y col. [3] presentaron una taxonomía exhaustiva de los diversos problemas que pueden venir asociados a los datos. Resumiendo, los datos «sucios» se pueden dividir en:

<sup>2</sup>Literalmente, «basura entra, basura sale».

1. Ausentes (exigidos o no por el sistema)
2. Presentes
  - a) Incorrectos:
    - Por problemas *detectables* dentro del propio sistema: datos de tipo incorrecto, referencias a datos que ya no existen, duplicidades, inconsistencias, resultados de transacciones incompletas, etc.
    - Por problemas *indetectables* dentro del propio sistema: datos caducados, direcciones postales incorrectas, números de teléfono inexistentes, etc.
    - Por error humano: de recogida de datos, de formato, de sincronización, etc.
  - b) Inútiles:
    - Datos contradictorios.
    - Datos ambiguos (homónimos, incompletos).
    - Datos que no respetan los estándares establecidos.

En la literatura existen numerosos estudios que relacionan directamente la calidad de los datos con el rendimiento de las organizaciones: la tesis de Slone [4] al respecto es un ejemplo, y Madnick y col. [5] plantean una categorización de los trabajos más relevantes de cada área. Otro trabajo interesante es el de Fisher y Kingma [6], que analiza el impacto que tuvo la baja calidad de los datos en los desastres de la lanzadera Challenger y el derribo por error de un avión de pasajeros por el navío de guerra USS Vincennes.

Todo esto ha impulsado la creación de una variedad de herramientas diseñadas específicamente para la limpieza y consolidación de los datos, tanto privativas como de código abierto. En el lado privativo, grandes empresas como Oracle o SAP integran funciones de limpieza de datos en sus productos de nivel superior. Por otro lado, algunas de las herramientas de código abierto más conocidas actualmente son Data Cleaner [7], Data Wrangler [8], Google OpenRefine [9] o Talend Open Studio for Data Quality [10].

**Propuesta** Proponemos el establecimiento de políticas eficaces sobre la recogida, manipulación y seguridad de los datos, con una definición expresa de quiénes son los responsables de qué datos y de qué mecanismos han de emplearse para asegurar la calidad de los datos. El empleo de las herramientas descritas no puede en ningún caso sustituir a estas políticas, aunque reconocemos su utilidad para detectar problemas comunes.

### 2.3 Inteligencia empresarial

Chaudhuri, Dayal y Narasya [11] definen «inteligencia empresarial»<sup>3</sup> como un conjunto de tecnologías de ayuda a la toma de decisiones que permiten a los gestores tomar mejores decisiones de forma más rápida. Estos autores atribuyen el auge de la inteligencia empresarial al coste cada vez menor que tiene para las organizaciones el guardar cantidades ingentes de datos procedentes de todo tipo de fuentes.

---

<sup>3</sup>En inglés, «business intelligence».

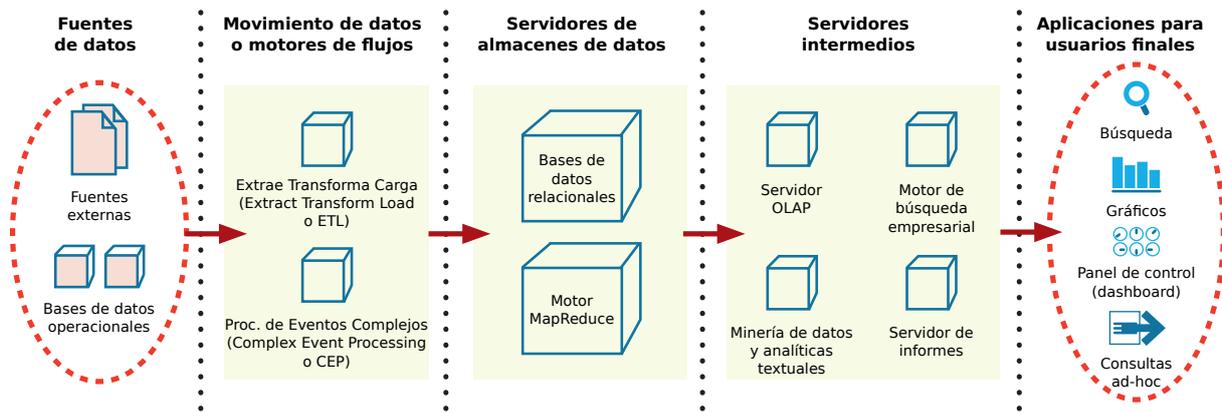


Figura 2: Componentes de una solución de inteligencia empresarial completa

Cuando se dispone de todos estos datos almacenados convenientemente, las empresas pueden aplicarles distintas técnicas de análisis: desde su consulta por distintos criterios para la elaboración de informes, a la realización de análisis multidimensionales o la ejecución de algoritmos avanzados de minería de datos.

**Componentes** La figura 2 resume los componentes principales de un sistema completo de inteligencia empresarial, que se describen a continuación.

- *Fuentes de datos:* dentro de una organización puede haber diversos tipos de fuentes de información. Por ejemplo, las propias bases de datos operacionales o transaccionales, hojas de cálculo, documentos de texto, correos electrónicos o registros de acceso, entre otras. Igualmente, es posible acceder a datos de fuentes externas, como los disponibles en el portal [datos.gob.es](http://datos.gob.es) o en servicios externos como OpenStreetMap o Twitter.

No obstante, cuanto menos estructurados se encuentren los datos, más difícil y costoso es su uso. Las hojas de cálculo, documentos de texto o correos electrónicos no son el lugar adecuado para almacenar los datos importantes que una organización necesita para su trabajo.

- *Movimiento de datos o motores de flujos:* normalmente, los datos en las fuentes originales están organizados para facilitar el trabajo diario (registrar una matrícula, gestionar un pago, etc.) y no su análisis. Si el análisis de los datos se hiciera directamente sobre las fuentes originales, se podría perder el marco de referencia temporal y, además, entorpecer el trabajo de los usuarios y los técnicos. En términos técnicos, el procesamiento directo de los datos que se emplean en el trabajo diario se conoce como *On-Line Transactional Processing* u OLTP, y el procesamiento de datos para su análisis se conoce como *On-Line Analytical Processing* u OLAP.

Para ello, la opción más habitual es aprovechar las horas de menor actividad, por ejemplo, las nocturnas, para lanzar procesos automatizados de Extracción, Transformación y Carga (*Extract Transform Load* o ETL) que tomen los datos de las diversas fuentes y los vuelquen en un almacén de datos pensado específicamente para tareas de análisis. Una vez los datos están en el almacén, se puede trabajar con ellos libremente sin molestar a los demás usuarios y con una referencia temporal clara: el momento en el que se extrajeron.

Las soluciones de inteligencia empresarial más populares incluyen herramientas específicas de ETL, como Warehouse Builder de Oracle [12], Kettle de Pentaho [13] o Talend Open Studio for Data Integration [14].

Si se necesitan datos más recientes en el almacén, se puede aumentar la frecuencia de carga, realizando cargas pequeñas frecuentemente en lugar de grandes cargas diarias o semanales. Si se llega al punto de tener que responder en tiempo real a situaciones puntuales, una posibilidad es enviar cada cambio que se realice en las fuentes originales a un motor de procesamiento de eventos complejos (*Complex Event Processing* o CEP) que reconozca ciertos patrones de eventos y reaccione a ellos. En la comunidad de código abierto, el más conocido es Esper [15].

- *Servidores de almacenes de datos*: alojan los almacenes en que se volcará la información extraída en el paso anterior.

En casos de pequeña o mediana envergadura lo habitual es emplear un sistema de gestión de bases de datos relacionales (basados en tablas) tradicional para alojar los contenidos del almacén, como Oracle, MySQL o PostgreSQL. Sin embargo, las bases de datos relacionales analíticas (Relational OLAP o ROLAP) se diseñan de forma distinta a las bases de datos relacionales transaccionales que se usan para el trabajo cotidiano: utilizan esquemas lógicos de *estrella* o *copo de nieve*, como describen Vassiliadis y Sellis [16].

Un ejemplo de un esquema lógico de estrella puede verse en la figura 3. En este tipo de esquemas hay una tabla central de hechos («Notas») con los datos que queremos analizar (las notas), y luego una serie de tablas con las distintas dimensiones sobre las que queremos hacer los análisis (por alumnos, por asignaturas, por fechas o por profesores, en este caso). Cada una de las dimensiones puede describirse a varios niveles: por ejemplo, en el caso de los profesores se podrían agrupar los datos por departamento o área y analizar medias, desviaciones típicas o cualquier otro tipo de cálculo.

Esta posibilidad de ver los datos proyectados sobre diversas dimensiones se conoce como «cubo OLAP». En este caso, el cubo resultante tendría 4 dimensiones: si por ejemplo fijásemos el alumno y el semestre, nos saldría una tabla de 2 dimensiones con las notas del alumno en dicho semestre. Sin embargo, si fijásemos la asignatura, el profesor y el semestre, nos saldría una lista unidimensional con las notas de una determinada convocatoria.

En casos de mayor envergadura, puede ser necesario emplear herramientas más especializadas, como bases de datos con implementaciones directamente multidimensionales (Multidimensional OLAP o MOLAP). Las MOLAP pueden dar más rendimiento, a costa de complicar considerablemente su mantenimiento y operación.

Si hay tantos datos que no basta con un servidor, se puede dividir el almacenamiento y el procesamiento entre varios. Un esquema muy utilizado actualmente y popularizado por Google es MapReduce, que divide un problema en muchos subproblemas, los resuelve por separado y luego los reúne en una respuesta final. La implementación más conocida de código abierto de MapReduce es Apache Hadoop [17], que además permite tratar el espacio en disco de varios ordenadores como un solo disco de gran capacidad.

- *Servidores intermedios*: una vez que los datos necesarios han sido integrados en el almacén, se puede instalar una serie de servidores que aprovechen los datos que hay en él y faciliten su explotación.

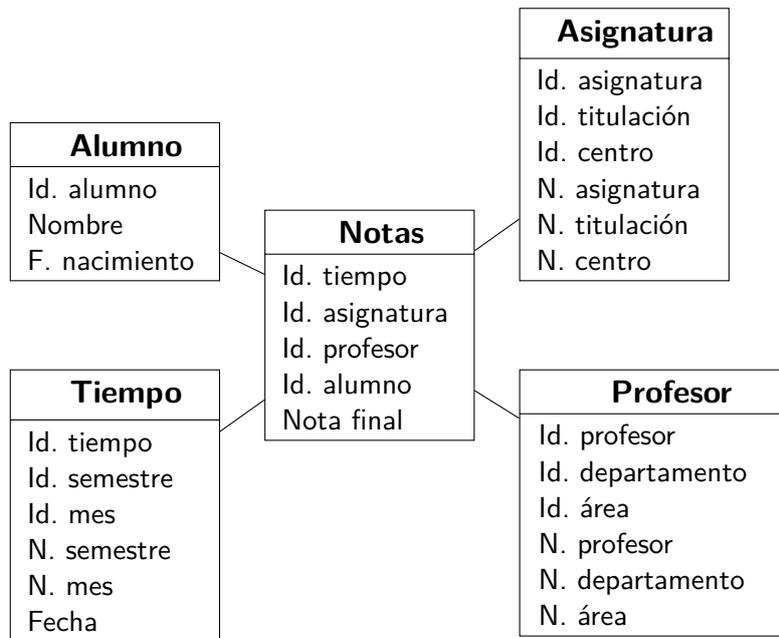


Figura 3: Ejemplo de esquema ROLAP en estrella

Los servidores de uso más inmediato son los de informes, que generan documentos y gráficos a partir de consultas realizadas sobre las tablas del almacén de datos. En el lado privativo existen soluciones como Oracle Discoverer 10g u Oracle Business Intelligence Publisher 11g, y también se pueden encontrar soluciones de código abierto como el Pentaho Reporting Engine [18] o el JasperReports Server [19].

Los servidores OLAP permiten visualizar los cubos OLAP antes mencionados y ofrecen muchas formas de agregar los datos recolectados en el almacén a lo largo de las dimensiones establecidas. Frente a soluciones privativas como Oracle Essbase, Mondrian [20] es la principal opción de código abierto. Mondrian puede usarse como un motor sin interfaz, o a través de interfaces web como las ofrecidas por SpagoBI [21], JasperSoft OLAP [22] o por Pentaho Business Analytics [23] (que también disponen de informes normales).

Por otro lado, los motores de búsqueda empresarial crean un índice a partir de todos los documentos de la organización, que pueden, por ejemplo, estar disponibles en su *intranet*, y permiten buscar en él. Aparte de opciones privativas como Oracle Secure Enterprise Search o Google Search Appliance [24], uno de los proyectos relacionados de código abierto de mayor calibre es Apache Solr [25].

Por último, existe la posibilidad de aplicar técnicas más avanzadas de análisis sobre los datos, generando no solo modelos descriptivos, sino incluso predictivos. Para ello, se pueden utilizar técnicas de minería de datos como las disponibles en el Oracle Data Miner o las implementadas en las herramientas de código abierto Weka [26] o R [27]. En particular, Pentaho integra a Weka en algunas de sus ediciones, y SpagoBI integra tanto a Weka como a R.

- *Aplicaciones para usuarios finales*: existe la posibilidad de aprovechar los servidores anteriores para ofrecer funciones no solo a los administradores, sino a usuarios normales sin conocimientos técnicos. Dependiendo de la plataforma escogida, se pueden plantear búsquedas de texto, gráficas personalizadas o crear consultas *ad hoc* (para generar nuevos informes sin necesidad de que lo haga un técnico).

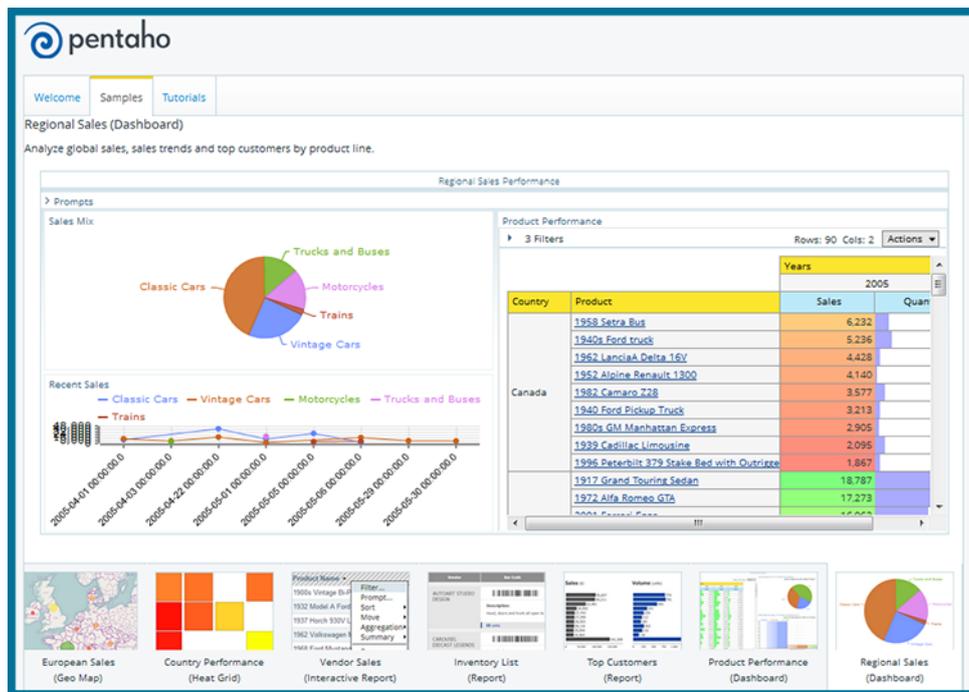


Figura 4: Ejemplo de panel de control elaborado con Pentaho

Una posibilidad especialmente interesante es la de crear paneles de control o «dashboards» en los que se recoja visualmente toda la información útil para usuarios de un determinado perfil. La figura 4 muestra un ejemplo de un panel de control elaborado con Pentaho. Estos paneles de control pueden emplearse como cuadros de mando.

**Opciones disponibles** Se observa que para cada uno de los componentes del esquema de la figura 2 existen opciones tanto privativas como de código abierto. En muchos casos las opciones de código abierto ofrecen dos posibilidades: una versión gratuita disponible para el público en general, y una versión de pago con soporte y funciones extra. Descartando los motores de búsqueda empresarial (la UCA cubre sus necesidades en esta área mediante Google y la búsqueda interna disponible en su propia web), una comparativa preliminar produce los resultados de la tabla 1. Golfarelli [28] realizó un estudio parecido en 2009, concluyendo que las soluciones de código abierto habían avanzado lo suficiente como para ser buenas sustitutas de las privativas en muchas ocasiones.

A juzgar por los resultados observados, Pentaho parece ser el motor con mayor comunidad de usuarios y funciones más avanzadas. Algunas de las carencias de la versión de código abierto han sido suplidas por la propia comunidad en cierta medida, como el motor Saiku (con soporte opcional de pago) de informes interactivos y el componente de Community Dashboard para reemplazar a los paneles de control que solo están disponibles en la edición de pago.

La solución de BI en su versión 10g de Oracle parece encontrarse un tanto limitada en la situación actual de licencias de la UCA, ya que no se disponen de licencias para OLAP o para el Warehouse Builder. Además, con la siguiente versión, la 11g, la situación ha cambiado sustancialmente, ya que Oracle ha reemplazado sus productos por los que obtuvo al comprar a la empresa Siebel.

Criterio	Oracle BI 10g	Oracle BI Suite EE+ 11g <sup>a</sup>	Pentaho Community	Pentaho Enterprise	Jaspersoft BI Community	Jaspersoft BI Enterprise	SpagoBI
Código abierto	No	No	Sí	Parcial	Sí	Parcial	Sí
Coste licencia	Alto	Alto	Gratis	Medio	Gratis	Medio	Gratis
Nivel de integración	Alto	Alto	Medio	Alto	Medio	Alto	Bajo
Formación en España	Sí	Sí	Sí (StrateBI, REDK)	Sí (StrateBI, REDK)	Sí (GMV)	Sí (StrateBI)	Sí (StrateBI)
Entorno ETL gráfico	Warehouse Builder 10g	Warehouse Builder 11g	Mondrian con Kettle (intérprete)	Mondrian con Kettle (intérprete)	Talend Open Studio (generador Java)		
Lectura	BD, CSV, Excel, Access	BD Oracle	BD, CSV, Excel, Access, JSON, XML, FTP, POP...	BD, CSV, Excel, Access, JSON, XML, FTP, POP...			
Escritura	8-10, CSV, XML	8-11 y otras (OHS), Excel, Access, CSV, XML	BD, CSV, Excel, Access, JSON, XML, FTP, POP...	BD, CSV, Excel, Access, JSON, XML, FTP, POP...			
Motor de informes	Discoverer 10g	Publisher 11g	Pentaho Reporting	Pentaho Reporting	JasperReports	JasperReports	JasperReports, BIRT, BO, Accessible
OLAP	Requiere ODBEE + OLAP	Requiere ODBEE + OLAP	Sí	Sí	Sí	Sí	Sí
Informes interactivos	Limitado	Sí	Sí (Saiku)	Sí	Sí	Sí	Sí
Consultas ad-hoc	Limitado	Sí	Sí (Saiku)	Sí	No	Sí	Sí
Paneles de control	Sí (Oracle Portal)	Sí (Interactive Dashboard)	Sí (Community Dashboard)	Completo	No	Sí	Sí
Acceso móvil	No	Sí (Mobile, \$360/usuario y año)	Solo lectura	Sí	Solo lectura	Solo lectura	Solo lectura

Cuadro 1: Comparativa de las plataformas disponibles para inteligencia empresarial

<sup>a</sup>En esta versión muchos de los componentes vienen de la adquisición de Siebel por parte de Oracle y pueden haber cambiado bastante.

**Propuesta** Se escogerá la herramienta técnicamente más apropiada en función de las disponibilidades presupuestarias y de personal, así como por los requisitos técnicos impuestos por las tecnologías que resulten indispensables en la situación actual para la UCA. En particular, el empleo de herramientas avanzadas de Oracle queda supeditado a sus costes finales, pendientes de la renovación de licencias por parte de la Junta de Andalucía, y a las versiones exigidas por los módulos de OCU UXXI ya en funcionamiento.

## 2.4 Reutilización de la información del sector público

El auge de Internet ha implicado una reducción considerable del coste de almacenar y transmitir grandes cantidades de información, y los ciudadanos cada vez exigen una mayor transparencia acerca del funcionamiento de las administraciones públicas. Además de facilitar un mejor control y ajuste de las políticas públicas, estos datos abiertos u «open data» se emplean para crear nuevas oportunidades de negocio en el sector privado. Según un informe del ONTSI [29], el sector en 2012 daba empleo a entre 3 600 y 4 400 profesionales, y su actividad económica en España rondaba entre los 330 y 550 millones de euros. Según el informe MEPSIR de 2006 [30], el mercado europeo de reutilización de información pública ascendería a unos 27 000 millones de euros.

Estos beneficios sociales y económicos potenciales hicieron que la Unión Europea publicara en 1998 el *Libro Verde sobre la Información del Sector Público en la Sociedad de la Información* por el cual llamaba la atención sobre la necesidad de aumentar la apertura de la información de las administraciones públicas. Esta iniciativa se materializó en la Directiva 2003/98/CE, que obligaba a los estados miembro a ofrecer la información a precios razonables, en condiciones claras e igualitarias y prohibiendo acuerdos exclusivos en general. Esta directiva fue posteriormente revisada en junio de 2013 exigiendo el derecho fundamental de acceso a toda la información pública de acuerdo a las legislaciones nacionales, reduciendo los precios máximos exigibles a consumidores, incluyendo a museos y otras entidades culturales en su ámbito, insistiendo en la obligación de ser transparentes en el precio y condiciones de uso e invitando a los estados miembros a utilizar formatos abiertos y legibles mediante computador (una demanda muy común de la comunidad «open data»).

En España, la Directiva 2003/98/CE se trasladó mediante la Ley 37/2007 sobre reutilización de la información del sector público. El Real Decreto 1495/2011 desarrolla esta ley para el ámbito del sector público estatal, obligando a cada uno de sus órganos a desarrollar un *Plan de Reutilización de la Información del Sector Público*. Entre algunas de las exigencias, un plan RISP requiere la creación y mantenimiento de un portal de datos abiertos que facilite la búsqueda y adquisición de la información por parte de ciudadanos y empresas. Tal es el caso del portal [datos.gob.es](http://datos.gob.es) creado en el proyecto *Aporta*, que agrega información en diversos formatos (PDF, CSV, Excel, etc.) de organismos como el Instituto Nacional de Estadística, el Ministerio de Hacienda, la Junta de Castilla y León o el Ayuntamiento de Zaragoza.

En lo que respecta al sector universitario español, uno de los referentes actuales es la Universidad Pablo de Olavide de Sevilla, con su primer portal <http://www.upo.es/datos-abiertos> basado en la plataforma de *blogging* Wordpress y su posterior portal [datos.upo.gob.es](http://datos.upo.gob.es), esta vez basado en la solución CKAN [31] de código abierto. En estos portales de datos abiertos

se incluye información estadística acerca de estudiantes de nuevo ingreso, presupuestos de ingresos y gastos de la UPO, y calificaciones en las asignaturas de grado.

La ventaja de usar una plataforma como CKAN es su mayor flexibilidad respecto a una solución genérica como Wordpress, y la posibilidad de hacer búsquedas sobre el catálogo completo de forma cómoda. En general, la calidad de la publicación de los datos abiertos es un factor importante que incide en su utilización real. Berners-Lee [32] definió una escala de 5 estrellas para calificar una implementación concreta de una iniciativa de datos abiertos:

- 1★ Disponible en una página web bajo una licencia abierta (Creative Commons Attribution [33], por ejemplo).
- 2★ Disponible en un formato estructurado legible por un programa (como una hoja de cálculo Excel).
- 3★ Además, que el formato sea abierto (como un fichero CSV, en lugar de una hoja de cálculo Excel).
- 4★ Además, que se empleen estándares de W3C relacionados con Web Semántica como RDF [34] o SPARQL [35] para identificar a las entidades.
- 5★ Además, que se enlacen los datos a los de otras entidades.

Según la experiencia del proyecto *Aporta*, las empresas interesadas en reutilizar los datos abiertos de la administración generalmente se contentaban con una implementación 3★. Podría decirse que este nivel de implementación tendría actualmente el mejor equilibrio entre valor aportado y coste de implementación.

Dentro del ámbito de las administraciones públicas españolas, sería necesario tener en cuenta además los requisitos impuestos por la *Norma Técnica de Interoperabilidad de la Secretaría de Estado de Administraciones Públicas* [36] (Resolución 19/2/2013, BOE 4/3). En esta norma técnica se imponen diversas restricciones más detalladas acerca de cómo hacer disponible el recurso y otros detalles técnicos.

**Propuesta** Se escogerá la herramienta técnicamente más apropiada en función de las disponibilidades presupuestarias y de personal, así como por los requisitos técnicos impuestos por las tecnologías que resulten indispensables en la situación actual para la UCA. En particular, el empleo de una plataforma como CKAN de código abierto frente a una solución genérica como Wordpress aportaría una mayor flexibilidad.

## 2.5 Aplicaciones dirigidas por dominios

El proceso de desarrollo de una aplicación de gestión exige una serie de pasos. De manera muy resumida, en una arquitectura tradicional basada en capas se requiere diseñar un modelo de datos e implementar las capas de persistencia, lógica de negocio y presentación. Cada una de estas etapas conlleva un tiempo y esfuerzo considerables, por lo que para ciertos casos sencillos

puede que en vez de desarrollar una aplicación integrada con las bases de datos institucionales, se haya desarrollado una solución *ad hoc* mediante una o varias hojas de cálculo (por ejemplo, en Excel) o una base de datos de escritorio (por ejemplo, en Access). Esto genera no pocos problemas y conduce a la fragmentación de los datos que manejan las distintas unidades.

Sin embargo, en la mayoría de estos casos sencillos la lógica a implementar es muy parecida: crear, leer, actualizar y eliminar entradas de diversos tipos y poco más. A este conjunto común de operaciones se le conoce como «Create, Read, Update and Delete» (CRUD) y su implementación tiende a ser bastante repetitiva.

La naturaleza mecánica de este tipo de código ha impulsado el desarrollo de marcos de trabajo que pueden acelerar considerablemente el desarrollo de aplicaciones CRUD al generar parte de la aplicación a partir de una descripción de los tipos de entradas que se van a gestionar o del propio modelo de datos (el *dominio* del sistema). Tres de los marcos más conocidos para el lenguaje de programación Java son:

- OpenXava [37] es, probablemente, el más sencillo de usar de los tres: a partir de un par de clases Java con anotaciones puede generar las funciones CRUD básicas. Para implementar validaciones y lógica de negocio se integra con otras tecnologías, como ABL o Hibernate Validation. La seguridad y la navegación se pueden implementar mediante el portal Liferay. Es un proyecto relativamente maduro, desarrollado desde 2009.
- Spring Roo [38] es un proyecto iniciado en 2009 por SpringSource, la empresa detrás del muy extendido marco de desarrollo Spring. Se trata de una interfaz para la generación rápida de aplicaciones Java basadas en las prácticas recomendadas oficialmente para Spring. Una de las facilidades que incorpora es la posibilidad de generar aplicaciones CRUD, pudiendo incluso generarlas a partir de una base de datos ya existente.
- Por otro lado, Apache Isis [39] es un proyecto más joven y con menos documentación, habiendo salido de la Apache Incubator en diciembre de 2012. Sin embargo, tiene una arquitectura más refinada, genera interfaces más amigables y permite un mayor grado de control. A cambio, tiene una curva de aprendizaje algo más pronunciada.

**Propuesta** Proponemos que se emplee uno de estos marcos de trabajo para reemplazar paulatinamente, en el orden de prioridad que se estime más oportuno, las soluciones *ad hoc* presentes en algunas unidades por aplicaciones integradas con el almacén de datos centralizado y que incluyan las comprobaciones necesarias para asegurar una mejor recogida de los datos de cara a la posibilidad de incorporar tecnologías de inteligencia empresarial a la UCA.

Una vez seleccionado el marco más adecuado, también proponemos acelerar el proceso de creación de nuevas aplicaciones mediante el desarrollo de un generador de nivel superior, que produciría la entrada necesaria para el marco a partir de un modelo o especificación de muy alto nivel. Existen muchas herramientas existentes para definir, validar y transformar modelos que podrían utilizarse en este contexto, como el generador de lenguajes específicos de dominio Xtext [40] o la familia de lenguajes de gestión de modelos Epsilon [41].

### 3 OBJETIVOS DEL PROYECTO

La finalidad del presente proyecto es incrementar la accesibilidad y visibilidad de la información dentro y fuera de la UCA. Dentro de la UCA, la información permitirá hacer un mejor uso de los recursos de la Universidad y tomar decisiones de manera más informada. Fuera de la UCA, la información será de utilidad para aumentar el grado de transparencia de la Universidad de Cádiz ante el público en general.

Los objetivos concretos de este proyecto son:

- O1. *Mejorar la calidad de los datos almacenados.* La calidad de los datos se mejorará de acuerdo a la caracterización de Wang, Strong y Guarascio [2] y la taxonomía de fallos de Won y col. [3]. Se emplearán herramientas especializadas para detectar e implementar posibilidades de mejora, haciendo los procesos de extracción más mantenibles y reemplazando soluciones *ad hoc* basadas en aplicaciones ofimáticas por aplicaciones web de desarrollo rápido de bajo coste.
- O2. *Facilitar la explotación de los datos.* Se integrarán herramientas que permitan a usuarios finales crear informes sencillos y que los usuarios técnicos puedan crear informes complejos más fácilmente.
- O3. *Hacer más visible la información dentro de la UCA.* Se creará un portal interno con los recursos repartidos a lo largo de la organización, detalles técnicos de interés general y paneles de control que resuman la información de interés para ciertos sectores de la UCA.
- O4. *Mejorar la transparencia de la UCA.* Se creará un portal de cara al exterior para liberar datos que puedan ser de interés general más allá de la Universidad de Cádiz.
- O5. *Aumentar la accesibilidad de la información.* Se reducirán los pasos necesarios para consultar los informes disponibles, aprovechando el directorio LDAP de la UCA.

Este proyecto servirá como punto de partida para la unificación y potenciación de los sistemas de información de la UCA alrededor de una plataforma especializada de inteligencia empresarial. En futuros proyectos se podrían integrar nuevos aspectos, como la federación con portales de datos abiertos (como `datos.gob.es`) de acuerdo a la Norma Técnica de Interoperabilidad o la integración de técnicas de minería de datos para estudios más avanzados, entre otros aspectos.

#### 3.1 Alcance del proyecto

##### Aspectos incluidos

- Introducir una plataforma colaborativa en la que se documentarán las fuentes de datos disponibles, mejorando su visibilidad.
- Mejorar los procesos de recogida reemplazando soluciones *ad hoc* actuales por aplicaciones web integradas en el sistema de información de la UCA.
- Migrar los informes repartidos actualmente entre varios sistemas a un portal unificado basado en herramientas actuales e integrado con la infraestructura de la UCA.
- Generar nuevos informes y paneles de control para responder a las nuevas necesidades de la UCA.
- Introducir un portal de datos abiertos para potenciar la reutilización de la información de la UCA por terceras partes para la generación de nuevas oportunidades de información y negocio.

### **Aspectos excluidos o limitados**

- La plataforma colaborativa de documentación se limitará a contener descripciones textuales y gráficas, y no se integrará directamente con la plataforma de inteligencia empresarial.
- Las soluciones *ad hoc* a reemplazar en este primer proyecto serán de naturaleza más sencilla, como hojas de cálculo Excel, bases de datos Access o documentos OpenDocument.
- El portal unificado se basará en una solución existente para portales, en vez de un desarrollo a medida.
- Los informes y paneles de control se basarán estrictamente en los componentes incluidos en la plataforma de inteligencia empresarial escogida. No se implementarán nuevos componentes a medida en este proyecto.
- El portal de datos abiertos utilizará formatos abiertos para publicar los datos, pero no incluirá información de tipo semántico, ya que es muy costosa de incluir. En este proyecto no se federará el portal con otros portales CKAN.

## **4 METODOLOGÍA Y PLAN DE TRABAJO**

El proyecto será llevado a cabo por cinco técnicos y un funcional por cada área que deba tratar el Sistema de Información. Además de poseer conocimientos generales de inteligencia empresarial, cada técnico se especializará en un área concreta:

- Un técnico de informes ("TI") con conocimientos avanzados de diseñadores de informes avanzados y *ad hoc*, como el Report Designer de Pentaho BI, los Report Services de Oracle BI Standard Edition o el Publisher de la Oracle BI Enterprise Edition.

- Un técnico de paneles de control (“TC”) con conocimientos avanzados del diseñador de paneles de control de la plataforma escogida. Algunos diseñadores de paneles de control incluyen el Community Dashboard Framework de Pentaho BI Community Edition, el Dashboard Designer de Pentaho BI Enterprise Edition, el Discoverer de la Oracle BI Standard Edition o los Interactive Dashboards de la Oracle BI Enterprise Edition.
- Un técnico de bases de datos (“TBD”) con conocimientos de gestión de bases de datos Oracle, herramientas de procesos de Extracción, Transformación y Carga («Extract Transform Load» o ETL) y herramientas de evaluación de calidad de datos. Algunos ejemplos de herramientas ETL incluyen Oracle Data Integrator (que reemplazará a Oracle Warehouse Builder en un futuro), Pentaho Kettle o Talend Open Studio for Data Integration. Algunos ejemplos de herramientas de evaluación de calidad de datos incluyen Talend Open Studio for Data Quality o Google OpenRefine.
- Un técnico de portales de sistemas de información (“TP”) con conocimientos avanzados de administración de plataformas de integración empresarial como Pentaho BI u Oracle BI y la capacidad de integrar el directorio LDAP de la UCA en la plataforma BI escogida. Para el portal de datos abiertos basado en CKAN, necesitará tener conocimientos de despliegue y mantenimiento de aplicaciones web basadas en Python.
- Un técnico de ingeniería dirigida por modelos (“TM”) con los conocimientos necesarios para implementar una herramienta para la generación rápida de aplicaciones web de gestión que reemplacen a soluciones *ad hoc* existentes basadas en productos ofimáticos. El técnico deberá tener conocimientos de generación de editores de modelos textuales y gráficos y de tecnologías para validarlos, transformarlos y generar código a partir de ellos. Igualmente, debería ser capaz de empaquetar las herramientas resultantes para su uso por el resto de los técnicos.

Por lo tanto, el técnico debería tener familiaridad con tecnologías de cada una de las siguientes categorías:

1. Generación y personalización de editores gráficos de modelos mediante las metodologías dirigidas por modelos de Eugenia (generando código basado en el Eclipse Graphical Modelling Framework) o Spray (generando código basado en Eclipse Graphiti).
2. Generación de editores textuales de lenguajes específicos de dominio mediante Xtext, con destacado de sintaxis, vista de estructura y validación automática.
3. Validación de modelos mediante restricciones OMG OCL empleando el EMF Validation Framework, o mediante lenguajes específicos como el Epsilon Validation Language.
4. Refactorización de modelos con Epsilon Wizard Language, permitiendo realizar pequeñas transformaciones sobre un mismo modelo.
5. Migración de modelos entre versiones de un metamodelo con Epsilon Flock.
6. Transformación de un modelo de un metamodelo a otro modelo de un metamodelo distinto mediante una de las muchas tecnologías disponibles: Epsilon Transformation Language, ATL, OMG QVT o Henshin.
7. Generación de código a partir de modelos: JET, Xpand, Acceleo o Epsilon Generation Language.

Dado que muchas de estas tecnologías están muy vinculadas a Eclipse, el técnico deberá tener conocimientos de desarrollo y distribución de *plug-ins* Eclipse. Se valorarán conocimientos sobre creación de aplicaciones RCP que integren de serie las nuevas herramientas desarrolladas.

En cuanto a tecnologías de desarrollo rápido de aplicaciones web, se valorarán conocimientos sobre alguno de los marcos actualmente existentes en el mercado, como OpenXava, Apache Isis, Spring Roo o Grails. El técnico deberá ser capaz de seleccionar uno de estos marcos como destino para los procesos finales de generación de código.

Por otro lado, este técnico debería ser capaz de crear o integrar otras herramientas que faciliten el trabajo de los demás técnicos. En particular, el técnico se ocuparía de desplegar y mantener un entorno de soporte al desarrollo de software de acuerdo con el estado del arte, incluyendo sistemas de control de versiones (Subversion o Git), entornos de integración continua (Jenkins) y gestores de proyectos (Redmine o Trac). Se valorarán conocimientos sobre lenguajes de scripting para automatización de tareas repetitivas (Bash o Python), herramientas de construcción de software (Maven), entornos de evaluación continua (SonarQube) y repositorios de artefactos binarios (Nexus).

- Un jefe de proyecto (“JP”) que realice la dirección técnica del proyecto y coordine al resto de participantes. El jefe de proyecto deberá asegurar que la planificación del proyecto se cumple, estableciendo hitos intermedios para asegurarse de que los entregables exigidos se realicen bajo los plazos indicados. Igualmente, deberá controlar el alcance del proyecto en caso de que haya desviaciones del calendario de ejecución respecto a la planificación. Adicionalmente, el jefe de proyecto se ocupará de planificar y organizar las reuniones entre los técnicos y los funcionales, asegurándose de que los técnicos reciban la suficiente información y los funcionales no se vean impedidos en sus labores cotidianas.

Deberán participar funcionales por cada una de las siguientes áreas funcionales: Alumnado, Economía, Personal, Investigación y Transferencia, Ordenación Académica, Planificación, Calidad, Posgrado, Innovación y Tecnologías Docentes, Dirección General Universidad y Empresa y Relaciones Internacionales, sin perjuicio de que se considere conveniente la participación de alguna otra.

Los cometidos de los funcionales serán los siguientes:

- Colaborar con los técnicos en la mejora de la calidad de los datos.
- Colaborar con los técnicos en la mejora de los procesos de recogida de información, fundamentalmente los funcionales de las áreas que utilizan soluciones ad hoc basadas en productos de ofimática para gestionar la información. Deberán colaborar para el efectiva migración de la información a la aplicación web dirigida por dominio.
- Proporcionar información acerca de los informes y paneles de control que necesitan, cumplimentando fichas técnicas incluyendo los siguientes campos, entre otros:
  - Código, denominación, destinatarios, fechas en que se necesita e indicadores de rendimiento relacionados.
  - Áreas funcionales responsables del informe y de carga de datos.

- Bases de datos y tablas a usar.
  - Características y condiciones generales a cumplir.
  - Campos del informe con sus descripciones y valores admisibles.
- Atender a reuniones con los técnicos para aclarar dudas sobre las fichas y validar los informes, paneles de control y portales desarrollados. Estas reuniones deberán limitarse para reducir su impacto sobre sus labores cotidianas. Los técnicos asistirán a los funcionales en la migración a los nuevos procesos implementados.

Además, la Dirección General de Sistemas de Información contará con el apoyo de la Unidad de Planificación en cuanto a la gestión de las solicitudes de información, así como el apoyo y asesoramiento técnico de la Unidad de Calidad y Evaluación, y de la Unidad de Planificación para el desarrollo de informes e indicadores.

#### 4.1 Tareas a realizar

Este borrador de planificación toma como fecha de inicio el 21 de octubre, en que se iniciará la formación de los técnicos y los primeros análisis de necesidades. El proyecto concluiría en un año y medio, abril de 2015. El resto de la planificación se encuentra en los cronogramas de la sección 4.2.

#### Tarea T1. Perfilado de datos

En esta tarea se crearán *perfiles* de calidad de los datos, indicando las condiciones que deberían cumplir para ser utilizables de cara a los procesos posteriores de Inteligencia Empresarial.

##### *Subtarea T1.1. Documentación de la información*

Se creará un catálogo editable de forma colaborativa de las bases de datos más importantes de la Universidad de Cádiz, que sea de utilidad para nuevos técnicos y para la posterior creación de informes y perfiles de calidad de datos.

Este catálogo operará a nivel de fuente de datos completa, dando los detalles de contacto del dueño de la fuente, el responsable técnico y una descripción general de información contenida y detalles técnicos a tener en cuenta en su uso. El catálogo se basará en un *wiki* o sistema de gestión de contenidos («Content Management System» o CMS) que permita un nivel fino de control de acceso, como TWiki [42] o Alfresco [43] respectivamente.

*Responsable(s)*: Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Catálogo de las fuentes de datos a considerar, con descripciones a alto nivel de los tipos de información disponibles y su contenido.
- E2. Plataforma de edición colaborativa que alojará el catálogo.

### *Subtarea T1.2. Selección de herramienta de perfilado de datos*

Se seleccionará una herramienta que permita definir las características de calidad que deben cumplir los datos y posteriormente comprobarlas, de forma repetible. Para ello, se considerarán algunas de las opciones de la sección 2.2, como Talend Open Studio for Data Quality o Google OpenRefine.

*Responsable(s):* Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

E1. Informe comparativo de las herramientas disponibles, con una tabla comparativa de características y una justificación de la herramienta finalmente escogida.

### *Subtarea T1.3. Evaluación de calidad*

Se utilizará la herramienta anteriormente seleccionada para definir los perfiles de calidad de datos de las bases de datos más importantes y detectar datos incompletos, erróneos o inconsistentes.

*Responsable(s):* Jefe de Proyecto, Téc. Bases de Datos y Funcionales.

*Entregable(s):*

E1. Listado justificado de las fuentes de datos de mayor interés.

E2. Especificaciones en la herramienta escogida de los requisitos de calidad de las fuentes de datos seleccionadas.

### *Subtarea T1.4. Correcciones según resultados*

A ser posible, se contactará con las áreas funcionales afectadas por los problemas detectados para que corrijan los problemas de calidad de datos. En caso de no ser factible su corrección, se documentará el proceso que sería necesario para limpiarlos si se necesita usarlos en un informe.

*Responsable(s):* Jefe de Proyecto, Téc. Bases de Datos y Funcionales.

*Entregable(s):*

E1. Informe sobre los problemas detectados en las fuentes de datos.

E2. Informe sobre la respuesta a los problemas detectados.

## **Tarea T2. Mejora de procesos de recogida**

En esta tarea se modernizarán algunos de los procesos de gestión realizados actualmente mediante soluciones *ad hoc* basadas en productos de ofimática, convirtiéndolos en aplicaciones web.

Para acelerar el proceso, se empleará un marco de desarrollo web dirigido por dominio, que tras un caso piloto para familiarizarse con él y comprobar su viabilidad práctica, se agilizará con una nueva herramienta que reduzca el trabajo necesario para iniciar un nuevo proyecto y aprender el entorno.

### *Subtarea T2.1. Prototipado comparativo con marcos dirigidos por dominio*

Se crearán prototipos de un mismo caso de laboratorio de gestión en varios de los marcos de desarrollo de aplicaciones web dirigidos por dominio existentes. Como punto de partida, se podrían comparar las opciones de la sección 2.5.

*Responsable(s):* Jefe de Proyecto y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

- E1. Documento con descripción y justificación del caso de laboratorio.
- E2. Código fuente de los prototipos desarrollados con cada marco de desarrollo seleccionado.
- E3. Informe que compare los marcos seleccionados a partir de las experiencias recogidas durante la elaboración de los prototipos.

*Subtarea T2.2. Estudio de procesos mejorables*

A través de varias reuniones con los funcionales, se creará un listado priorizado de los procesos *ad hoc* que se beneficiarían de migrar a una aplicación web dirigida por dominio, con las muestras y especificaciones necesarias para su posterior implementación.

*Responsable(s)*: Jefe de Proyecto, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s)*:

- E1. Listado priorizado de procesos mejorables.

*Subtarea T2.3. Selección de marco dirigido por dominio*

Se escogerá uno de los marcos probados en la tarea T2.1, evaluando su funcionalidad, usabilidad, extensibilidad y capacidad de automatización a través de los prototipos desarrollados.

*Responsable(s)*: Jefe de Proyecto y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Informe justificado acerca del marco finalmente escogido.

*Subtarea T2.4. Caso piloto*

Como caso piloto, se elaborará manualmente la aplicación de mayor prioridad de la lista utilizando el marco seleccionado. Se tomará nota de los comentarios de la unidad a la que vaya dirigida y de las correcciones que sea necesario hacer sobre la aplicación inicialmente aportada por el marco.

*Responsable(s)*: Jefe de Proyecto, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s)*:

- E1. Código fuente de la aplicación de mayor prioridad.
- E2. Informe acerca de las experiencias obtenidas, haciendo especial énfasis en su aptitud como objetivo para la generación automática de código.

*Subtarea T2.5. Generalización por MDE*

Una vez se haya completado el caso piloto satisfactoriamente, se utilizará la experiencia conseguida para elaborar una primera versión de una herramienta de ingeniería dirigida por modelos («Model-Driven Engineering» o MDE) que acelere el desarrollo de aplicaciones similares para otras unidades.

*Responsable(s)*: Jefe de Proyecto y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Código fuente de la primera versión de la herramienta MDE.
- E2. Distribución binaria de la primera versión de la herramienta MDE.
- E3. Descripción del enfoque seguido y arquitectura de la herramienta.

*Subtarea T2.6. Aplicación 1 de la herramienta MDE*

Se utilizará la herramienta MDE anteriormente desarrollada sobre la siguiente aplicación en la lista, evaluando el grado en que se acelera el proceso y la calidad de la

aplicación generada. El técnico de portales ayudará a que la aplicación se integre dentro de la infraestructura de la UCA, particularmente la autenticación por LDAP.

*Responsable(s):* Jefe de Proyecto, Téc. Ing. Dirigida por Modelos, Téc. Portales y Funcionales.

*Entregable(s):*

- E1. Código fuente de la primera aplicación desarrollada con la herramienta.
- E2. Documentación de la primera aplicación desarrollada.
- E3. Despliegue y demostración de la primera aplicación a los funcionales afectados.
- E4. Informe con las impresiones obtenidas por investigador, técnico de portales y funcionales afectados.

#### *Subtarea T2.7. Revisión de la herramienta MDE*

Se revisará la herramienta MDE en base a las experiencias adquiridas durante su primera aplicación del marco MDE seleccionado. Es posible que en esta fase se deban cambiar los modelos o las transformaciones de la herramienta.

*Responsable(s):* Jefe de Proyecto, Téc. Ing. Dirigida por Modelos y Téc. Portales.

*Entregable(s):*

- E1. Código fuente de la segunda versión de la herramienta MDE.
- E2. Distribución binaria de la segunda versión de la herramienta MDE.
- E3. Descripción justificada de los cambios introducidos.

#### *Subtarea T2.8. Aplicación 2 de la herramienta MDE*

Se aplicará la herramienta MDE sobre la tercera aplicación en la lista, comparando la experiencia con la de la primera versión.

*Responsable(s):* Jefe de Proyecto, Téc. Ing. Dirigida por Modelos, Téc. Portales y Funcionales.

*Entregable(s):*

- E1. Código fuente de la segunda aplicación desarrollada con la herramienta.
- E2. Documentación de la segunda aplicación desarrollada.
- E3. Despliegue y demostración de la segunda aplicación a los funcionales afectados.
- E4. Informe con las impresiones obtenidas por investigador, técnico de portales y funcionales afectados.

#### *Subtarea T2.9. Difusión de resultados*

Se difundirán los resultados obtenidos dentro y fuera de la UCA. Dentro de la UCA, se darán charlas a funcionales sobre los resultados obtenidos, y a alumnos sobre los detalles técnicos que puedan ser de su interés. Fuera de la UCA, se publicará la herramienta MDE como código abierto y se realizarán presentaciones en eventos de interés.

*Responsable(s):* Jefe de Proyecto, Téc. Ing. Dirigida por Modelos, Téc. Portales y Funcionales.

*Entregable(s):*

- E1. Memoria con los eventos realizados.

### **Tarea T3. Preparativos de BI**

En esta primera fase de las iniciativas de inteligencia empresarial («Business Intelligence» o BI), se seleccionará una plataforma de las existentes y se formará a los técnicos en ella. A continuación, se instalará la plataforma seleccionada y se creará un portal que la unifique con la documentación de análisis y las otras iniciativas de sistemas de información de la Universidad. Tanto la plataforma de inteligencia empresarial como el portal deberán integrarse con el directorio LDAP de usuarios de la UCA.

#### *Subtarea T3.1. Prototipado comparativo con plataformas BI*

De cara a formarse sobre las diversas plataformas existentes de inteligencia empresarial y tomar una decisión más informada, en esta etapa previa los técnicos realizarán diversas pruebas de concepto con varias de las plataformas de inteligencia empresarial existentes. Algunas de las posibles opciones se encuentran listadas en la sección 2.3.

*Responsable(s)*: Téc. Informes, Téc. Paneles de Control, Téc. Portales y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

E1. Documento con las impresiones generales obtenidas sobre las plataformas BI existentes.

#### *Subtarea T3.2. Selección de plataforma BI*

Se examinarán las experiencias de las pruebas de concepto previas para decidir qué plataforma concreta utilizar. La plataforma deberá permitir la creación de informes lo suficientemente avanzados y paneles de control y su integración en portales más amigables para usuarios no técnicos.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Paneles de Control, Téc. Portales y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

E1. Informe justificado de la elección de plataforma BI realizada.

#### *Subtarea T3.3. Curso de plataforma BI*

Se solicitará y atenderá a un curso que sienta unas bases sólidas para los técnicos implicados en las iniciativas BI en la plataforma seleccionada. Este curso podría solicitarse a una empresa externa o a un experto dentro de la UCA, si lo hubiere.

*Responsable(s)*: Téc. Ing. Dirigida por Modelos, Téc. Portales, Téc. Paneles de Control, Téc. Informes y Téc. Bases de Datos.

*Entregable(s)*:

E1. Certificados de aprovechamiento de los cursos recibidos.

#### *Subtarea T3.4. Instalación de plataforma BI*

Tras recibir el curso en la plataforma, se desplegará la plataforma BI en la infraestructura de la Universidad de Cádiz, estableciendo los mecanismos necesarios para asegurar su continuidad (monitorización, copias de seguridad, etc.).

*Responsable(s)*: Téc. Portales.

*Entregable(s)*:

E1. Instancia de la plataforma BI instalada.

E2. Plan de mantenimiento de la plataforma BI que describa las tareas a realizar, quién las realizará y cuándo.

#### *Subtarea T3.5. Integración LDAP con BI*

Para evitar la necesidad de gestionar un conjunto distinto de credenciales en la plataforma BI y simplificar el acceso a los usuarios, se configurará la plataforma BI para que admita autenticación y control de acceso por el directorio LDAP de usuarios de la UCA.

*Responsable(s):* Téc. Portales.

*Entregable(s):*

- E1. Integración de la plataforma BI con el directorio LDAP de la UCA.
- E2. Informe con los cambios necesarios en la configuración de la plataforma BI para posterior referencia.

#### *Subtarea T3.6. Creación del portal del SI*

Se instalará un portal de contenidos que permita integrar informes y paneles de control de la plataforma BI escogida, pudiendo visualizarlos en modo de sólo lectura. Este portal debe permitir editar su información de forma colaborativa entre los miembros del proyecto, y debería estar también integrado con el directorio LDAP de la UCA.

*Responsable(s):* Jefe de Proyecto y Téc. Portales.

*Entregable(s):*

- E1. Portal de contenidos instalado y corriendo en la UCA, con enlace desde la sección oportuna de la página principal.
- E2. Instrucciones de acceso y uso básicas para usuarios finales.

### **Tarea T4. Migración SIUCA a BI**

Tras completar el despliegue de la infraestructura de inteligencia empresarial, la primera aplicación a migrar es SIUCA. SIUCA implementa una serie de informes útiles, pero solo opera con datos locales, y no se conecta directamente a las bases de datos institucionales, por lo que la actualización de los datos disponibles requiere realizar una serie de tareas manuales.

#### *Subtarea T4.1. Análisis de informes a migrar*

En esta primera fase se realizarán reuniones con los funcionales para ver cuáles de los informes de SIUCA deberían migrarse a la nueva plataforma y qué mejoras se podrían hacer sobre ellos. Tras las reuniones, se analizará la información recogida para obtener un conjunto uniforme de especificaciones.

*Responsable(s):* Jefe de Proyecto, Téc. Informes, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s):*

- E1. Listado justificado de informes a migrar.

#### *Subtarea T4.2. Revisión del almacén de datos*

En función de los informes a migrar, se revisará si el almacén de datos tiene todo lo necesario y si los procesos de extracción aseguran una calidad suficiente para los datos. En caso de no ser así, se revisará el diseño de la base de datos y se crearán nuevos procesos de extracción utilizando la herramienta ETL de la plataforma BI escogida.

*Responsable(s)*: Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Informe comparando los informes solicitados con la información disponible.
- E2. En caso de necesitar mejoras, informe justificado sobre las mejoras realizadas (nuevos procesos ETL y ampliación de los almacenes).

#### *Subtarea T4.3. Implementación de informes*

Tras haber revisado el almacén, se empleará la herramienta de elaboración de informes de la plataforma BI escogida para implementar los informes de interés y hacerlos accesibles a través de la plataforma BI.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Instrucciones de acceso y descripciones de los informes implementados.

#### *Subtarea T4.4. Integración en portal*

Para simplificar el acceso a los artefactos elaborados, éstos se integrarán en modo de sólo lectura al portal del Sistema de Información.

*Responsable(s)*: Téc. Portales.

*Entregable(s)*:

- E1. Instrucciones de acceso a través del portal de los informes implementados.

#### *Subtarea T4.5. Formación*

Se formará a los funcionales de las áreas funcionales afectadas en el uso del portal y los nuevos elementos añadidos en esta tarea.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Portales y Funcionales.

*Entregable(s)*:

- E1. Manual que recoja las instrucciones de acceso y sus contenidos actuales, dividido por áreas y con explicaciones acerca de la interpretación de los elementos añadidos.
- E2. Cuestionario de evaluación por los funcionales de los informes obtenidos.
- E3. Resultados del cuestionario de evaluación por los funcionales.

### **Tarea T5. Migración SIDIR a BI**

Habiendo adquirido experiencia con la plataforma a través de la migración a SIUCA, se migrarán los informes de SIDIR a la plataforma escogida. En general, el esquema de trabajo será similar al de la migración de SIUCA.

#### *Subtarea T5.1. Análisis de informes a migrar*

Véase la tarea T4.1.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s)*:

- E1. Listado justificado de informes a migrar.

*Subtarea T5.2. Revisión del almacén de datos*

Véase la tarea T4.2.

*Responsable(s)*: Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Informe comparando los informes solicitados con la información disponible.
- E2. En caso de necesitar mejoras, informe justificado sobre las mejoras realizadas (nuevos procesos ETL y ampliación de los almacenes).

*Subtarea T5.3. Implementación de informes*

Véase la tarea T4.3.

*Responsable(s)*: Téc. Informes.

*Entregable(s)*:

- E1. Instrucciones de acceso y descripciones de los informes implementados.

*Subtarea T5.4. Integración en portal*

Véase la tarea T4.4.

*Responsable(s)*: Téc. Portales.

*Entregable(s)*:

- E1. Instrucciones de acceso a través del portal de los informes implementados.

*Subtarea T5.5. Formación*

Véase la tarea T4.5.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Portales y Funcionales.

*Entregable(s)*:

- E1. Manual que recoja las instrucciones de acceso y sus contenidos actuales, dividido por áreas y con explicaciones acerca de la interpretación de los elementos añadidos.
- E2. Cuestionario de evaluación por los funcionales de los informes obtenidos.
- E3. Resultados del cuestionario de evaluación por los funcionales.

**Tarea T6. Paneles de control**

Además de unificar los sistemas existentes, en esta tarea se elaborarían *paneles de control* como el de la figura 4 de la página 10. Estos paneles de control permiten resumir mediante gráficas una gran cantidad de información en una sola pantalla.

*Subtarea T6.1. Análisis de necesidades*

Véase la tarea T4.1.

*Responsable(s)*: Jefe de Proyecto, Téc. Paneles de Control, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s)*:

- E1. Listado justificado de informes a migrar.

*Subtarea T6.2. Revisión del almacén de datos*

Véase la tarea T4.2, cambiando informes por paneles de control.

*Responsable(s)*: Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

- E1. Informe comparando los informes solicitados con la información disponible.
- E2. En caso de necesitar mejoras, informe justificado sobre las mejoras realizadas (nuevos procesos ETL y ampliación de los almacenes).

*Subtarea T6.3. Creación de paneles de control*

Véase la tarea T4.3, cambiando informes por paneles de control.

*Responsable(s):* Jefe de Proyecto, Téc. Paneles de Control y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

- E1. Instrucciones de acceso y descripciones de los paneles de control implementados.

*Subtarea T6.4. Integración en portal*

Véase la tarea T4.4.

*Responsable(s):* Téc. Portales.

*Entregable(s):*

- E1. Instrucciones de acceso a través del portal de los informes implementados.

*Subtarea T6.5. Formación*

Véase la tarea T4.5.

*Responsable(s):* Jefe de Proyecto, Téc. Paneles de Control, Téc. Portales y Funcionales. *Entregable(s):*

- E1. Manual que recoja las instrucciones de acceso y sus contenidos actuales, dividido por áreas y con explicaciones acerca de la interpretación de los elementos añadidos.
- E2. Cuestionario de evaluación por los funcionales de los informes obtenidos.
- E3. Resultados del cuestionario de evaluación por los funcionales.

**Tarea T7. Otros informes desde fuentes existentes**

Además de los informes de SIDIR y SIUCA y los paneles de control, los funcionales u otros miembros de la comunidad UCA necesitarán otros informes para sus tareas diarias. Estos informes se implementarán en esta tarea.

*Subtarea T7.1. Análisis de necesidades*

Véase la tarea T4.1.

*Responsable(s):* Jefe de Proyecto, Téc. Informes, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s):*

- E1. Listado justificado de informes a migrar.

*Subtarea T7.2. Revisión del almacén de datos*

Véase la tarea T4.2.

*Responsable(s):* Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

- E1. Informe comparando los informes solicitados con la información disponible.
- E2. En caso de necesitar mejoras, informe justificado sobre las mejoras realizadas (nuevos procesos ETL y ampliación de los almacenes).

#### *Subtarea T7.3. Implementación de informes*

Véase la tarea T4.3.

*Responsable(s)*: Jefe de Proyecto y Téc. Informes.

*Entregable(s)*:

E1. Instrucciones de acceso y descripciones de los informes implementados.

#### *Subtarea T7.4. Integración en portal*

Véase la tarea T4.4.

*Responsable(s)*: Téc. Portales.

*Entregable(s)*:

E1. Instrucciones de acceso a través del portal de los informes implementados.

#### *Subtarea T7.5. Formación*

Véase la tarea T4.5.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Portales y Funcionales.

*Entregable(s)*:

E1. Manual que recoja las instrucciones de acceso y sus contenidos actuales, dividido por áreas y con explicaciones acerca de la interpretación de los elementos añadidos.

E2. Cuestionario de evaluación por los funcionales de los informes obtenidos.

E3. Resultados del cuestionario de evaluación por los funcionales.

### **Tarea T8. Otros informes y paneles de control desde nuevas fuentes**

Tras haber modernizado varias de las soluciones *ad hoc* a aplicaciones web, podrán servir de fuente de datos para nuevos informes y paneles de control dentro de la UCA. En esta tarea se implementarían dichos informes.

#### *Subtarea T8.1. Análisis de necesidades*

Véase la tarea T4.1.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Paneles de Control, Téc. Ing. Dirigida por Modelos y Funcionales.

*Entregable(s)*:

E1. Listado justificado de informes a migrar.

#### *Subtarea T8.2. Revisión del almacén de datos*

Véase la tarea T4.2.

*Responsable(s)*: Jefe de Proyecto, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s)*:

E1. Informe comparando los informes solicitados con la información disponible.

E2. En caso de necesitar mejoras, informe justificado sobre las mejoras realizadas (nuevos procesos ETL y ampliación de los almacenes).

#### *Subtarea T8.3. Implementación de informes y paneles de control*

Véase la tarea T4.3.

*Responsable(s)*: Téc. Informes y Téc. Paneles de Control.

*Entregable(s)*:

E1. Instrucciones de acceso y descripciones de los informes y paneles de control implementados.

*Subtarea T8.4. Integración en portal*

Véase la tarea T4.4.

*Responsable(s)*: Téc. Portales.

*Entregable(s)*:

E1. Instrucciones de acceso a través del portal de los informes implementados.

*Subtarea T8.5. Formación*

Véase la tarea T4.5.

*Responsable(s)*: Jefe de Proyecto, Téc. Informes, Téc. Paneles de Control, Téc. Portales y Funcionales.

*Entregable(s)*:

E1. Manual que recoja las instrucciones de acceso y sus contenidos actuales, dividido por áreas y con explicaciones acerca de la interpretación de los elementos añadidos.

E2. Cuestionario de evaluación por los funcionales de los informes obtenidos.

E3. Resultados del cuestionario de evaluación por los funcionales.

**Tarea T9. Difusión de iniciativas BI**

Tras la migración de SIUCA y SIDIR y la implementación de nuevos informes y paneles de control, se iniciarán una serie de acciones para difundir los resultados obtenidos dentro y fuera de la UCA. Esto incluye la formación de los funcionales en el uso del portal del SI y la creación de informes *ad hoc*, la realización de charlas o talleres para alumnos y la exposición de las iniciativas en foros nacionales e internacionales.

*Responsable(s)*: Jefe de Proyecto, Téc. Ing. Dirigida por Modelos, Téc. Portales y Funcionales.

*Entregable(s)*:

E1. Memoria con las actividades realizadas para la difusión de las iniciativas BI.

**Tarea T10. Datos abiertos**

Tras la migración de SIDIR y SIUCA, se podría plantear la creación de un portal de datos abiertos para ofrecer a la sociedad el subconjunto considerado de interés general. Esto mejoraría la transparencia de la UCA y permitiría unirse a iniciativas europeas y nacionales como las discutidas en la sección 2.4.

*Subtarea T10.1. Instalación y formación CKAN*

CKAN [31] es un software de código abierto muy utilizado para implantar portales de datos abiertos, y a nivel nacional ya ha sido adoptado por la Universidad Pablo de Olavide, reemplazando al gestor de contenidos Wordpress que usaban anteriormente. En esta tarea, el técnico de portales instalará CKAN en la infraestructura de la Universidad de Cádiz y se formará sobre él.

*Responsable(s)*: Jefe de Proyecto y Téc. Portales.

*Entregable(s)*:

E1. Instancia de CKAN instalada y corriendo en la UCA.

E2. Instrucciones de instalación, puesta punto y mantenimiento de la instancia de CKAN para referencia posterior.

#### *Subtarea T10.2. Selección de datos a publicar*

Dado el marco legal en que se mueve la Universidad, en esta tarea los técnicos y los funcionales se reunirán para revisar la información de que dispone la Universidad y determinar qué información es de interés general y puede publicarse bajo el marco legal vigente.

*Responsable(s):* Jefe de Proyecto, Téc. Portales, Téc. Bases de Datos y Funcionales.

*Entregable(s):*

E1. Listado justificado de fuentes de datos consideradas de interés general y descripción de aquellos contenidos que puedan ser publicados como datos abiertos.

#### *Subtarea T10.3. Filtrado de datos*

Tras seleccionar la información a publicar, se implementarán los procesos de extracción necesarios para preparar los ficheros de entrada que se enviarán a la instalación de CKAN de la UCA.

*Responsable(s):* Jefe de Proyecto, Téc. Portales, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

E1. Especificaciones, implementaciones y manuales de los procesos de extracción y almacén de datos revisado para atender a las fuentes de datos seleccionadas.

#### *Subtarea T10.4. Automatización de actualizaciones*

Se automatizará el envío de la información de CKAN, de manera que los datos del portal se mantengan actualizados día tras día.

*Responsable(s):* Jefe de Proyecto, Téc. Bases de Datos y Téc. Portales.

*Entregable(s):*

E1. Planificación justificada de las actualizaciones.

E2. Implementación de la planificación en los sistemas que alojan a CKAN y la plataforma BI.

#### *Subtarea T10.5. Integración de visualizaciones*

Además de los datos en crudo, se integrarán visualizaciones en el portal para mejorar su estudio por parte de los visitantes.

*Responsable(s):* Jefe de Proyecto, Téc. Portales y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

E1. Especificación de las visualizaciones a añadir al sistema.

E2. Implementación de las visualizaciones en el portal.

E3. Informe describiendo las capacidades de visualización de serie y potenciales de CKAN.

#### *Subtarea T10.6. Difusión del portal CKAN*

Habiendo integrado las visualizaciones, se difundirán los resultados obtenidos y la labor realizada dentro (PAS, PDI y alumnos) y fuera de la UCA (foros científicos y redes sociales).

*Responsable(s):* Jefe de Proyecto, Téc. Portales, Téc. Bases de Datos y Téc. Ing. Dirigida por Modelos.

*Entregable(s):*

E1. Memoria de las actividades realizadas para la difusión de la instancia CKAN en la UCA.

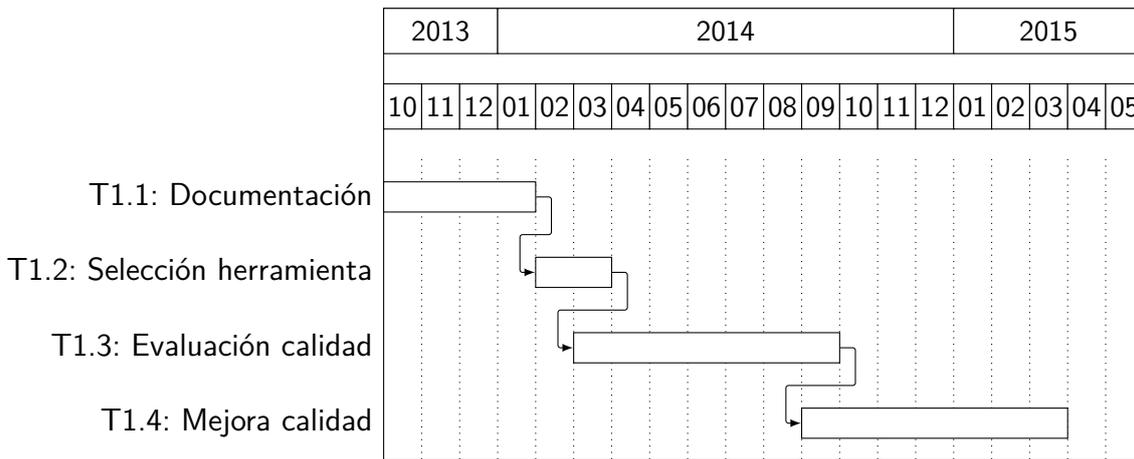


Figura 5: Cronograma por meses para T1 (Perfilado de datos)

## 4.2 Cronogramas

La planificación temporal del proyecto viene recogida en los cronogramas de las figuras 5 a 9.

El perfilado de datos (figura 5) y la mejora de la recogida de información (figura 6) es en su mayor parte independiente del resto, permitiendo que el técnico de bases de datos, el técnico de aplicaciones Web y el técnico de ingeniería dirigida por modelos adelanten trabajo.

La parte más compleja es la de BI (figuras 7 y 8), ya que exige la cooperación simultánea de todos los miembros del equipo, aunque en distintas fases: el técnico de informes se ocupará del análisis y la elaboración de informes, el técnico de bases de datos revisará el diseño de almacenes y procesos de extracción, el técnico de portales integrará los informes y paneles de control en un portal centralizado y el técnico de ingeniería dirigida por modelos aportará tecnologías para acelerar el proceso.

Cuando la migración de SIUCA y SIDIR al nuevo portal se complete, se iniciará en paralelo con las otras tareas el desarrollo del portal de datos abiertos tal y como se muestra en la figura 9.

Al inicio de la ejecución de cada fase de análisis, se planificarán reuniones con los funcionales de las distintas áreas para obtener las especificaciones de los informes y paneles de control a elaborar.

## 4.3 Matriz de trazabilidad tarea/objetivo

En el cuadro 2 se recogen las dependencias entre las tareas antes descritas y los objetivos concretos señalados en la sección 3. Cada marca indica que la tarea en cuestión asiste en la consecución del objetivo correspondiente.

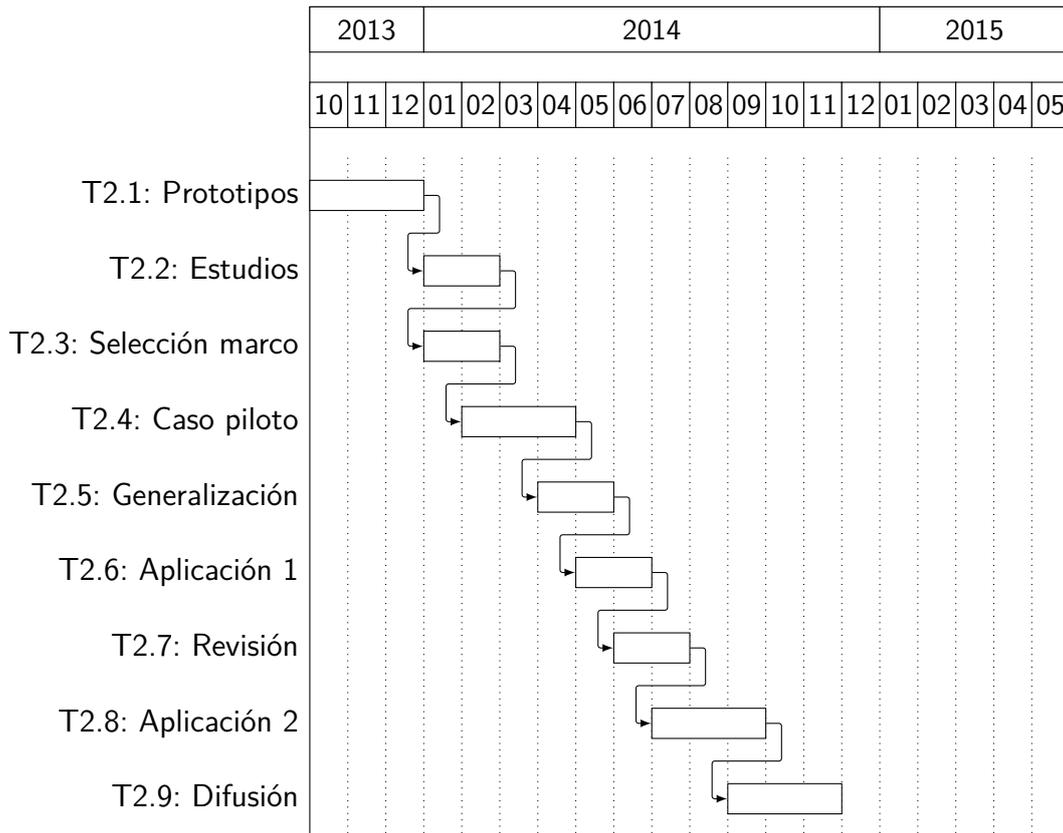


Figura 6: Cronograma por meses para T2 (Mejora de procesos de recogida)

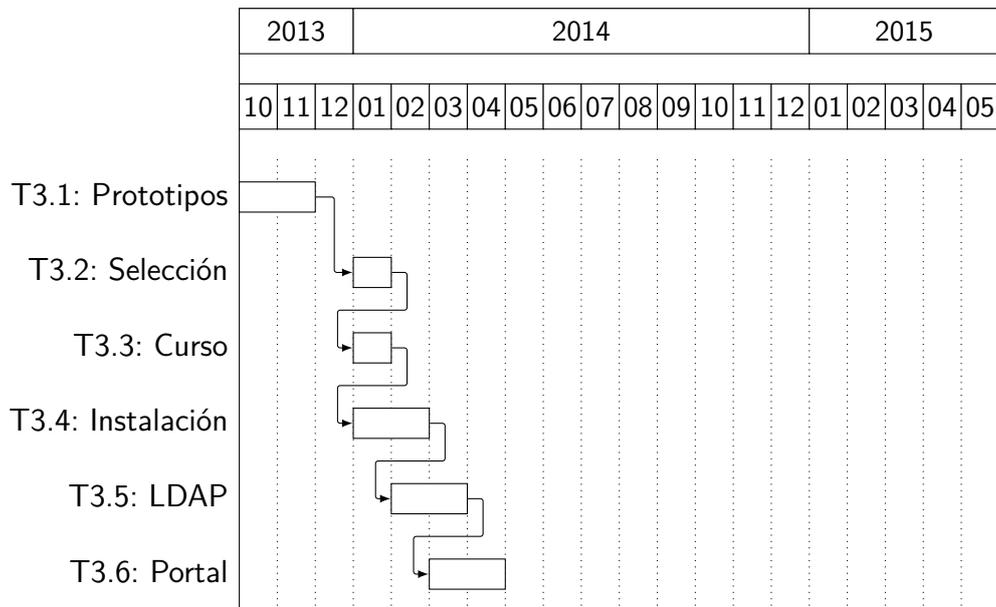


Figura 7: Cronograma por meses para formación y despliegue de infraestructura de Inteligencia Empresarial



Tareas	Objetivos				
	O1	O2	O3	O4	O5
T1.1	■				
T1.2	■				
T1.3	■				
T1.4	■				
T2.1	■				
T2.2	■				
T2.3	■				
T2.4	■	■			
T2.5	■				
T2.6	■	■			
T2.7	■				
T2.8	■	■			
T2.9				■	■
T3.1		■			
T3.2		■			
T3.3		■			
T3.4		■			
T3.5		■			■
T3.6		■			■
T4.1			■		
T4.2			■		
T4.3			■		
T4.4			■		■
T4.5			■		■
T5.1			■		
T5.2			■		
T5.3			■		
T5.4			■		■
T5.5			■		■
T6.1			■		
T6.2			■		
T6.3			■		
T6.4			■		■
T6.5			■		■
T7.1			■		
T7.2			■		
T7.3			■		
T7.4			■		■
T7.5			■		■
T8.1			■		
T8.2			■		
T8.3			■		
T8.4			■		■
T8.5			■		■
T9				■	■
T10.1				■	
T10.2				■	
T10.3				■	
T10.4				■	
T10.5				■	
T10.6				■	

Cuadro 2: Matriz de trazabilidad tarea/objetivo

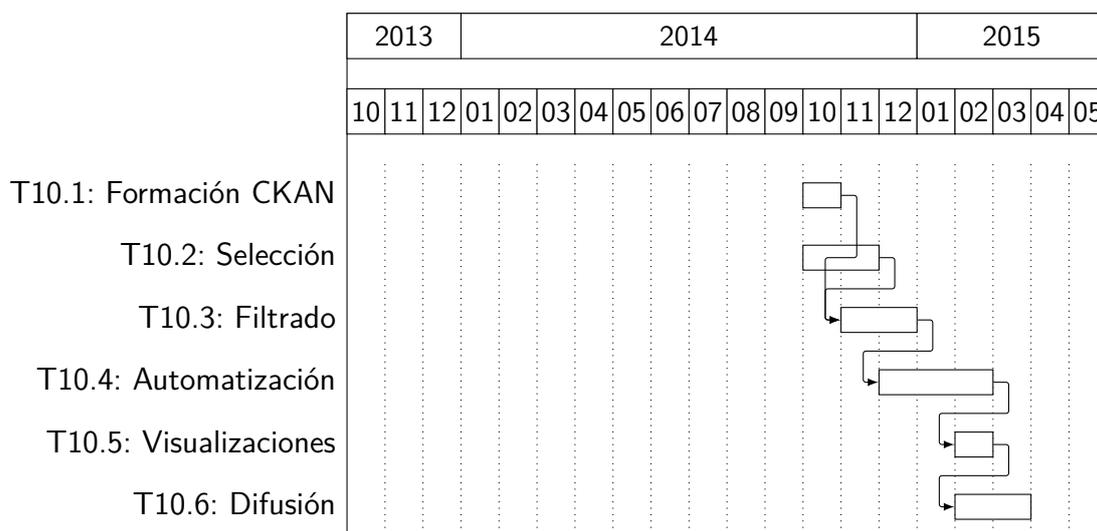


Figura 9: Cronograma por meses para despliegue de portal de datos abiertos

#### 4.4 Indicadores de éxito

Para monitorizar el progreso intermedio del proyecto y evaluar los resultados finales obtenidos, se utilizarán los siguientes indicadores en cada tarea de primer nivel:

- T1, perfilado de datos:
  - Porcentaje de las fuentes de datos usadas en los informes que se documentan en el catálogo colaborativo.
  - Porcentaje de las fuentes de datos que disponen de perfiles de calidad en la herramienta seleccionada.
  - Incidencias de calidad detectadas y resueltas en los perfiles.
- T2, mejora de recogida:
  - Encuesta de satisfacción del funcional destino del caso piloto, basándose en criterios de funcionalidad, usabilidad y calidad de la integración con el portal del SI.
  - Encuesta de satisfacción del funcional destino de la primera aplicación de la herramienta MDE, e informe sobre la calidad técnica de la herramienta y el resultado por parte del técnico de portales.
  - Encuesta de satisfacción del funcional destino de la segunda aplicación de la herramienta MDE, e informe sobre la calidad técnica de la herramienta y el resultado por parte del técnico de portales.
- T4, migración SIUCA:
  - Porcentaje de informes migrados a la plataforma BI integrada.
  - Encuesta de satisfacción de los funcionales destino de los informes migrados.
- T5, migración SIDIR:

- Porcentaje de informes migrados a la plataforma BI integrada.
  - Encuestas de satisfacción de los funcionales destino de los informes migrados.
- T6, paneles de control:
  - Porcentaje de paneles de control implementados de los inicialmente solicitados.
  - Encuestas de satisfacción de los funcionales destino de los nuevos paneles de control.
- T7, nuevos informes desde fuentes existentes:
  - Porcentaje de informes de control implementados de los inicialmente solicitados.
  - Encuestas de satisfacción de los funcionales destino de los nuevos informes.
- T8, nuevos informes y paneles de control desde nuevas fuentes:
  - Porcentaje de informes de control implementados de los inicialmente solicitados.
  - Encuestas de satisfacción de los funcionales destino de los nuevos informes y paneles de control.
- T9, difusión de iniciativas BI:
  - Listado de ponencias y publicaciones realizadas sobre las iniciativas realizadas.
  - Métricas de difusión en medios sociales.
- T10, datos abiertos:
  - Publicación del portal de la UCA en el portal de datos abiertos `datos.gob.es`, cumpliendo sus requisitos de calidad.
  - Grado de automatización de las actualizaciones del portal de datos.
  - Análíticas web (p. ej. Google Analytics) y reutilizaciones por parte de terceros de la información disponible en el portal de datos.

#### 4.5 Análisis de riesgos y medidas correctivas

A continuación se listan los riesgos actualmente identificados en el proyecto, junto con una serie de medidas correctivas dirigidas a reducir su probabilidad o su impacto:

- El mayor riesgo es el de implementar herramientas de recogida, informes o paneles de control que no se ajusten a las necesidades de los funcionales destino y que por lo tanto no se aprovechen.

Para evitar que esto ocurra, todo artefacto deberá basarse en una especificación consensuada por el técnico y el funcional destino en la que se listen los datos a utilizar, el procesamiento deseado y su presentación. Durante su desarrollo, se realizarán reuniones periódicas con los funcionales en las que se demostrará el artefacto producido y se pedirá una valoración acerca de su adecuación. Finalmente, se solicitará el cumplimiento de una encuesta de satisfacción sobre el producto terminado, utilizando la información recogida para los futuros artefactos.

- El siguiente riesgo es que el tiempo exigido para los informes o paneles de control sea superior al necesario, bien sea por motivos técnicos o por cambios en los requisitos.

Para reducir el impacto de este riesgo, se priorizarán los artefactos en función de la inmediatez de la necesidad de información que lo ha generado. En particular, la migración de los informes de SIUCA y SIDIR tomará prioridad sobre la creación de nuevos informes, excepto cuando se necesiten para satisfacer necesidades inmediatas.

En un caso extremo, se considerará posponer otros informes o el portal de datos abiertos a un proyecto posterior. Los perfiles de calidad en T1 podrían reducirse a las fuentes más utilizadas si falta tiempo. Por otro lado, en las herramientas de recogida de T2 se podría limitar la funcionalidad a un conjunto esencial para el trabajo día a día.

- Otro riesgo es que haya falta de coordinación entre técnicos y funcionales, retrasando el proyecto y exigiendo rehacer ciertas tareas.

Para conseguir una mayor coordinación, se habilitará una zona editable colaborativamente por técnicos y funcionales en el portal en la que se guardarán el calendario de reuniones y las memorias, las especificaciones de los artefactos a producir y enlaces a los artefactos que se vayan produciendo (informes, paneles de control, herramientas de recogida, código fuente, etc.).

- A nivel técnico, existe el riesgo de que se escoja un marco de desarrollo dirigido por modelos o una plataforma BI que no sea ideal. Las tareas T2.3 y T3.1 procuran dar el tiempo suficiente a los técnicos para desarrollar prototipos y tomar una decisión más informada. Además, existe también el riesgo de que la formación tanto de los técnicos como de los funcionales no sea la adecuada. A esto último se intenta hacer frente con las tareas T3.3, T4.5, T5.5, T6.5, T7.5, T8.5, T9 y T10.1.

#### 4.6 Matriz de responsabilidades

En las siguientes matrices RACI se recoge el reparto de responsabilidades entre los miembros del proyecto para cada uno de los entregables. La responsabilidad asignada puede ser de cuatro tipos:

**R («responsable»)** Realiza el trabajo y es responsable de su realización.

**A («aprobador»)** Aprueba el trabajo finalizado y a partir de ese momento, se vuelve responsable de él.

**C («consultado»)** Posee alguna información o capacidad necesaria para terminar el trabajo.

**I («informado»)** Debe ser informado sobre el progreso y los resultados del trabajo.

Entregable	Jefe de Proyecto	Téc. Ing. Dirigida por Modelos	Téc. Portales	Téc. Bases de Datos	Téc. Informes	Téc. Paneles de Control	Funcionales
T1.1.E1	A	C		R			
T1.1.E2	A	R		R			
T1.2.E1	A	C		R			
T1.3.E1	A			R			C
T1.3.E2	A			R			
T1.4.E1	A			R			I
T1.4.E2	A			R			I
T2.1.E1	A	R					
T2.1.E2	A	R					
T2.1.E3	A	R					
T2.2.E1	A	R					C
T2.3.E1	A	R					
T2.4.E1	A	R					
T2.4.E2	A	R					C
T2.5.E1	A	R					
T2.5.E2	A	R					
T2.5.E3	A	R					
T2.6.E1	A	R					
T2.6.E2	A	R					I
T2.6.E3	A	R	I				I
T2.6.E4	A	R	C				C
T2.7.E1	A	R	C				
T2.7.E2	A	R	C				
T2.7.E3	A	R	C				
T2.8.E1	A	R	C				
T2.8.E2	A	R	I				I
T2.8.E3	A	R	I				I
T2.8.E4	A	R	C				C
T2.9.E1	A	R	R				C

Entregable	Jefe de Proyecto	Téc. Ing. Dirigida por Modelos	Téc. Portales	Téc. Bases de Datos	Téc. Informes	Téc. Paneles de Control	Funcionales
T3.1.E1		R	R		R	R	
T3.2.E1	A	C	R		R	R	
T3.3.E1	I	R	R	R	R	R	
T3.4.E1			R				
T3.4.E2	A		R				
T3.5.E1	A		R				
T3.5.E2	A		R				
T3.6.E1	A		R				
T3.6.E2	A	I	R	I	I	I	I
T4.1.E1	A	R			R		C
T4.2.E1	A	C		R			
T4.2.E2	A	C		R			
T4.3.E1	A	C			R		
T4.4.E1	A		R		C		I
T4.5.E1	A	I	C	I	R	I	I
T4.5.E2	A				R		I
T4.5.E3	A				C		R
T5.1.E1	A	R			R		C
T5.2.E1	A	C		R			
T5.2.E2	A	C		R			
T5.3.E1	A	C			R		
T5.4.E1	A		R		C		I
T5.5.E1	A	I	C	I	R	I	I
T5.5.E2	A				R		I
T5.5.E3	A				C		R
T6.1.E1	A	R				R	C
T6.2.E1	A	C		R			
T6.2.E2	A	C		R			
T6.3.E1	A	C				R	I
T6.4.E1	A		R			C	I
T6.5.E1	A	I	C	I	I	R	I
T6.5.E2	A					R	I
T6.5.E3	A					C	R

Entregable	Jefe de Proyecto	Téc. Ing. Dirigida por Modelos	Téc. Portales	Téc. Bases de Datos	Téc. Informes	Téc. Paneles de Control	Funcionales
T7.1.E1	A	R			R		C
T7.2.E1	A	C		R			
T7.2.E2	A	C		R			
T7.3.E1	A	C			R		I
T7.4.E1	A		R		C		I
T7.5.E1	A	I	C	I	R	I	I
T7.5.E2	A				R		I
T7.5.E3	A				C		R
T8.1.E1	A	R			R	R	C
T8.2.E1	A	C		R			
T8.2.E2	A	C		R			
T8.3.E1	A	C			R	R	I
T8.4.E1	A		R		C	C	I
T8.5.E1	A	I	C	I	R	R	I
T8.5.E2	A				R	R	I
T8.5.E3	A				C	C	R
T9.E1	A	R	R				C

Entregable	Jefe de Proyecto	Téc. Ing. Dirigida por Modelos	Téc. Portales	Téc. Bases de Datos	Téc. Informes	Téc. Paneles de Control	Funcionales
T10.1.E1	A			R			
T10.1.E2	A	I		R			
T10.2.E1	A	C	R	C			C
T10.3.E1	A	C	R	C			
T10.4.E1	A		R				
T10.4.E2	A		C	R			
T10.5.E1	A		R				
T10.5.E2	A		R				
T10.5.E3	A	I	R				I
T10.6.E1	A	R	R	R			

## Referencias

- [1] Universidad Pablo de Olavide. *Portal de Datos Abiertos*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://datos.upo.gob.es/>.
- [2] Richard Y. Wang, Diane Strong y Lisa M. Guarascio. *Beyond accuracy: what data quality means to data consumers*. Inf. téc. TDQM-94-10. Fecha de última comprobación: 26 de septiembre de 2013. Cambridge, MA, EEUU: Sloan School of Management, MIT, octubre de 1994. URL: <http://web.mit.edu/tdqm/www/papers/94/94-10.html>.
- [3] Kim Won y col. «A Taxonomy of Dirty Data». En: *Data Mining and Knowledge Discovery* 7 (2003), págs. 81-99. ISSN: 1573-756X.
- [4] John P. Slone. «Information quality strategy: an empirical investigation of the relationship between information quality improvements and organizational outcomes». Tesis doct. Capella University, octubre de 2006.

- [5] Stuart E. Madnick y col. «Overview and Framework for Data and Information Quality Research». En: *ACM Journal of Data and Information Quality* 1.1 (junio de 2009).
- [6] Craig W. Fisher y Bruce R. Kingma. «Criticality of data quality as exemplified in two disasters». En: *Information and Management* 39 (enero de 2001), págs. 109-116. ISSN: 0378-7206.
- [7] Human Inference. *Data Cleaner*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://datacleaner.org/>.
- [8] Stanford Visualization Group. *Data Wrangler*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://vis.stanford.edu/wrangler/>.
- [9] Google. *OpenRefine*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://openrefine.org/>.
- [10] Talend. *Talend Open Studio for Data Quality*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.talend.com/products/data-quality>.
- [11] Surajit Chaudhuri, Umeshwar Dayal y Vivek Narasya. «An overview of business intelligence technology». En: *Communications of the ACM* 54.8 (agosto de 2011), págs. 88-98.
- [12] Oracle Corporation. *Oracle Warehouse Builder*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.oracle.com/technetwork/developer-tools/warehouse/overview/introduction/index.html>.
- [13] Pentaho. *Kettle*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://kettle.pentaho.com/>.
- [14] Talend. *Talend Open Studio for Data Integration*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.talend.com/products/data-integration>.
- [15] EsperTech. *Esper*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://esper.codehaus.org/>.
- [16] Panos Vassiliadis y Timos Sellis. «A survey of logical models for OLAP databases». En: *ACM SIGMOD Records* 28.4 (diciembre de 1999), págs. 64-69. ISSN: 0163-5808. DOI: 10.1145/344816.344869. URL: <http://doi.acm.org/10.1145/344816.344869>.
- [17] Apache Software Foundation. *Hadoop*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://hadoop.apache.org/>.
- [18] Pentaho. *Reporting Engine*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://reporting.pentaho.com/>.
- [19] JasperSoft Community. *JasperReports Server*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://community.jaspersoft.com/project/jasperreports-server>.
- [20] Pentaho. *Mondrian*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://mondrian.pentaho.com/>.
- [21] SpagoWorld. *SpagoBI*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/>.
- [22] JasperSoft. *Jasper OLAP 4.5 Data Sheet*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.jaspersoft.com/download/jasperanalysis-datasheet>.

- [23] Pentaho. *Business Analytics*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://pentaho.com/product/business-visualization-analytics>.
- [24] Google. *Search Appliance*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://www.google.com/enterprise/search/products/gsa.html>.
- [25] Apache Software Foundation. *Solr*. Fecha de última comprobación: 26 de septiembre de 2013. 2013. URL: <http://lucene.apache.org/solr/>.
- [26] Mark Hall y col. «The WEKA Data Mining Software: An Update». En: *SIGKDD Explorations* 11.1 (2009). Fecha de última comprobación: 26 de septiembre de 2013. URL: <http://www.kdd.org/newsletter/explorations-july-2009-11-1>.
- [27] WU Institute for Statistics and Mathematics. *The R Project for Statistical Computing*. Fecha de última comprobación: 26 de septiembre de 2013. Septiembre de 2013. URL: <http://www.r-project.org/>.
- [28] Matteo Golfarelli. «Open Source BI Platforms: A Functional and Architectural Comparison». En: *Proceedings of DaWaK 2009*. Ed. por T. B. Pedersen, M.K. Mohania y A. M. Tjoa. Vol. 5691. Lecture Notes in Computer Science. 2009, págs. 287-297.
- [29] Varios. *Estudio de caracterización del Sector Infomediario en España*. Inf. téc. 1207272032398. Fecha de última comprobación: 30 de septiembre de 2013. Observatorio nacional de las telecomunicaciones y la SI (ONTSI), julio de 2012. URL: <http://www.ontsi.red.es/ontsi/es/estudios-informes/estudio-de-caracterizaci%C3%B3n-del-sector-infomediario-en-esp%C3%B1a-edici%C3%B3n-2012>.
- [30] Makx Dekkers y col. *Measuring European Public Sector Information Resources (MEPSIR)*. Inf. téc. Fecha de última comprobación: 30 de septiembre de 2013. HELM Group, junio de 2006. URL: [http://ec.europa.eu/information\\_society/newsroom/cf/document.cfm?doc\\_id=1198](http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=1198).
- [31] The Open Knowledge Foundation. *CKAN*. Fecha de última comprobación: 30 de septiembre de 2013. Septiembre de 2013. URL: <http://ckan.org/>.
- [32] Tim Berners-Lee. *Linked Data: Design Issues*. Fecha de última comprobación: 30 de septiembre de 2013. 2010. URL: <http://www.w3.org/DesignIssues/LinkedData.html>.
- [33] Creative Commons. *Atribución 3.0 Unported*. Fecha de última comprobación: 30 de septiembre de 2013. 2013. URL: <http://creativecommons.org/licenses/by/3.0/deed.es>.
- [34] World Wide Web Consortium. *Resource Description Framework (RDF)*. Fecha de última comprobación: 30 de septiembre de 2013. Marzo de 2013. URL: <http://www.w3.org/RDF/>.
- [35] World Wide Web Consortium. *SPARQL Query Language for RDF*. Fecha de última comprobación: 30 de septiembre de 2013. Enero de 2008. URL: <http://www.w3.org/TR/rdf-sparql-query/>.
- [36] Ministerio de Hacienda y Administraciones Públicas. *Portal de Administración Electrónica — Normas Técnicas de Interoperabilidad: Reutilización de Recursos de Información*. Fecha de última comprobación: 30 de septiembre de 2013. Febrero de 2013. URL: [http://administracionelectronica.gob.es/pae\\_Home/pae\\_Estrategias/pae\\_Interoperabilidad\\_Inicio/pae\\_Normas\\_tecnicas\\_de\\_interoperabilidad.html#REUTILIZACIONRECURSOS](http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Interoperabilidad_Inicio/pae_Normas_tecnicas_de_interoperabilidad.html#REUTILIZACIONRECURSOS).

- [37] OpenXava.org. *Página principal de OpenXava*. Fecha de última comprobación: 1 de octubre de 2013. Octubre de 2013. URL: <http://www.openxava.org/es/web/guest/credits>.
- [38] SpringSource. *Spring Roo*. Fecha de última comprobación: 1 de octubre de 2013. Septiembre de 2013. URL: <https://github.com/spring-projects/spring-roo>.
- [39] Apache Software Foundation. *Apache Isis*. Fecha de última comprobación: 1 de octubre de 2013. 2013. URL: <http://isis.apache.org/>.
- [40] Eclipse Software Foundation. *Xtext*. Fecha de última comprobación: 1 de octubre de 2013. Septiembre de 2013. URL: <http://www.eclipse.org/Xtext/>.
- [41] Dimitrios S. Kolovos y col. *The Epsilon Book*. Fecha de última comprobación: 1 de octubre de 2013. 2013. URL: <http://dev.eclipse.org/svnroot/modeling/org.eclipse.epsilon/trunk/doc/org.eclipse.epsilon.book/EpsilonBook.pdf>.
- [42] Peter Thoeny. *TWiki*. Fecha de última comprobación: 19 de octubre de 2013. Octubre de 2013. URL: <http://twiki.org/>.
- [43] Alfresco Software. *Alfresco*. Fecha de última comprobación: 19 de octubre de 2013. Octubre de 2013. URL: <http://www.alfresco.com/es>.