

# Hacia una universidad abierta

Recomendaciones para el **SUE**

# Hacia una universidad abierta:

## Recomendaciones para el SUE

### Coordinación y edición:

**Enrique Teruel Doñate**

*Delegado del Rector para los Servicios de Informática y Comunicaciones – Universidad de Zaragoza*

*Coordinador del **subgrupo de Gobierno Abierto***

**José Pascual Gumbau Mezquita**

*Director del Gabinete de Planificación y Prospectiva Tecnológica – Universitat Jaume I*

*Coordinador del **grupo de Análisis, Planificación y Gobierno de las TI de la Sectorial TIC de la CRUE***

### Autores:

Este documento es fruto del trabajo del subgrupo de Gobierno Abierto del grupo de Análisis, Planificación y Gobierno de las TI de la Sectorial TIC de la CRUE. Naturalmente, el trabajo, y el grupo, está abierto a la participación y colaboración.

- Lluís Alfons Ariño Martín (Universitat Rovira i Virgili)
- Ricardo Borillo Doménech (Universitat Jaume I)
- Jordi Campos Díaz (Universitat Pompeu Fabra)
- Óscar Corcho (Universidad Politécnica de Madrid)
- Sonia Castro Muñoz (Red.es)
- María José García (Universidad Autónoma de Madrid)
- Irene Garrigós Fernández (Universidad de Alicante)
- José Pascual Gumbau Mezquita (Universitat Jaume I)
- Mónica López del Consuelo (Universidad de Granada)
- Francisco Javier López Pellicer (Universidad de Zaragoza)
- Vicente Matellán Olivera (Universidad de León)
- José Norberto Mazón López (Universidad de Alicante)
- Jorge Pantoja (Universitat Pompeu Fabra)
- Miguel Ángel Sicilia (Universidad de Alcalá de Henares)
- Enrique Teruel Doñate (Universidad de Zaragoza)
- Alicia Troncoso Lora (Universidad Pablo de Olavide)
- José Jacobo Zubcoff Vallejo (Universidad de Alicante)

### Naturaleza y licencia de este documento (open & linked):

Este documento, es la versión 20140605, probablemente definitiva, de un documento vivo de trabajo. El [documento vivo](#) está radicalmente abierto a nuevas participaciones, especialmente universitarias

**Ha sido deliberadamente preparado para prevenir su impresión en papel, las fuentes están enlazadas en lugar de citadas, y algunas de ellas son audiovisuales, no textuales.**



Hacia una universidad abierta por CRUE-TIC / APGTI / Gobierno Abierto se distribuye bajo una [Licencia Creative Commons Atribución-Compartir Igual 4.0 Internacional](#)

# 1. Introducción

El objeto de este documento, fruto del trabajo conjunto del grupo constituido en el marco CRUE-TIC, es promover y facilitar la aplicación de los principios de gobierno abierto en las universidades, sirviendo como guía que describa los puntos que deben ser abordados en la puesta en marcha de una iniciativa así, con el propósito de contribuir a la construcción de una universidad abierta del siglo XXI.

Dado que el primer principio del gobierno abierto es la transparencia, y ésta se practica de forma idónea mediante la apertura de datos, los primeros pasos deberán sentar las bases para que las universidades empiecen a abrir sus datos institucionales de forma coherente, desarrollando políticas de transparencia que conviene adoptar, conjuntamente con políticas de acceso abierto al conocimiento, tanto docente como investigador.

## 1.1. ¿Qué es gobierno abierto?

El gobierno abierto surge como un nuevo modelo de relación entre los gobernantes, las administraciones y la sociedad, del que las universidades no deben permanecer al margen.

El concepto de gobierno abierto considera a los ciudadanos co-responsables de la evaluación, definición, diseño y desarrollo de los servicios que las administraciones públicas les prestan, y les da el poder de solucionar los problemas.

La postura habitual de los gobiernos ha sido la de salvaguardar los datos, las informaciones y el conocimiento que les sirven de base para la toma de decisiones. En este celo protector se han volcado grandes esfuerzos, y en muchos casos hasta hace poco tiempo, no se había planteado qué ocurriría si en lugar de ocultar los abriesen, los hiciesen de dominio público, y dejasen interactuar con ellos a los ciudadanos.

Las [primeras experiencias han sido tan alentadoras](#) que se ha producido un movimiento global de apertura de datos y gobiernos, cuyo nacimiento todavía estamos presenciando. Por ejemplo, la reciente [declaración del G8 en Lough Erne](#) (18 de Junio de 2013), que es un conciso decálogo de políticas sobre la responsabilidad de los gobiernos, incluye un punto sobre apertura y reutilización de información.

Los tres principios fundamentales de este movimiento, recogidos en el celebrado "[memorando de Obama](#)" (2009) son:

- Transparencia: oferta de información clara y actualizada, accesible y reutilizable.
- Participación: intervención de la ciudadanía en todas las actividades del gobierno.
- Colaboración: entre instituciones y llamando a los ciudadanos a compartir lo que saben y a generar soluciones en las áreas donde tienen conocimientos.

El mejor exponente de la cultura de gobierno abierto es la [Declaración de Gobierno Abierto](#) (Open Government Declaration, 2011) firmada por 54 países comprometidos a:

- Incrementar la disponibilidad de información sobre actividades gubernamentales.
- Apoyar la participación cívica en la formulación de políticas y en la toma de decisiones públicas.
- Implementar los más altos niveles de integridad profesional en las administraciones.
- Incrementar el acceso a nuevas tecnologías para la apertura y la rendición de cuentas públicas.

De forma más visual, y en tres minutos, los principios de gobierno abierto se ilustran en [este video promocional](#) de la [Open Government Partnership](#).

Precisamente tras la primera conferencia de la OGP, la sociedad civil, encabezada por [Access Info Europe](#), puso en marcha la iniciativa Open Government Standards mediante la cual se consensuaron una serie de criterios con los que evaluar si una iniciativa, programa o política puede ser calificada de "Gobierno Abierto".

De forma sintética se han fijado los siguientes [estándares](#):

- **Transparencia:** La información se crea y se hace disponible para el público, con excepciones limitadas, actualizada, en formatos abiertos y sin restricciones de uso.
- **Participación:** El público puede participar directamente en la consideración de opciones y toma de decisiones, y puede contribuir con sus ideas y evidencias para servir a la sociedad y los intereses democráticos.
- **Rendición de cuentas:** El gobierno se pone a disposición del público para explicar y responsabilizarse de sus decisiones y acciones, defendiendo normas de conducta e integridad.

## 1.2. ¿Qué es transparencia y datos abiertos?

La transparencia permite a todos los ciudadanos conocer y vigilar el empleo de los recursos públicos y estimula a las instituciones a funcionar de modo más eficiente. Para éstas supone además una herramienta imprescindible para recuperar la confianza del ciudadano.

La administración es una gran fuente de información que cuenta con cantidad de datos de carácter público: mapas, meteorología, tráfico, datos financieros, subvenciones, planes urbanísticos, acuerdos políticos, informes de investigación financiados públicamente en base a los cuales se aprueban normas... Toda esta información pertenece a la ciudadanía.

Los datos serán la materia prima para que nuevos actores elaboren nuevos productos y servicios, creando valor y riqueza. Es una forma de capacitar a los ciudadanos y hacerlos más responsables, regenerando la conexión entre políticos y ciudadanos, facilitando la participación de estos últimos en la vida pública. La Reutilización de la Información del Sector Público (RISP), [avalada oficialmente por directivas europeas](#) y sus transposiciones locales, es un objetivo principal de cualquier iniciativa Open Data, así que conviene elaborar un [plan de reutilización](#). El [W3C](#) también ha elaborado un [documento de recomendaciones](#), aparte de trabajar activamente en tecnologías de interés (formatos, web semántica, etc).

A su vez el proceso de limpiar y preparar los datos para su publicación es en sí mismo beneficioso para los miembros de las administraciones que necesitan acceder a los datos. En otras palabras, abrir la información a los ciudadanos significa hacer la información más accesible dentro de la propia administración. De hecho, hoy por hoy, según un [estudio del sector infomediario \(en 2012\)](#), uno de los reutilizadores finales más señalados de la información es la propia administración.

Dado que la accesibilidad de los datos, su estructuración, su disponibilidad en formatos abiertos y su interconexión favorecen la reutilización, los catálogos de datos abiertos se clasifican por estos criterios, siguiendo la propuesta de [“cinco estrellas” de Tim Berners-Lee](#). Las ventajas de la apertura de datos tecnológicamente avanzada se ilustran en [este video promocional](#) de la [fundación CTIC](#).

La transparencia es demandada de forma cada vez más clamorosa a todas las administraciones y otras entidades a resultas de escándalos reiterados. Es una exigencia que también afecta a las universidades, incluidas las privadas, pues nos debemos a la sociedad a la que servimos, y es inexcusable rendirle cuentas. La [Fundación Compromiso y Transparencia](#) publicó en septiembre de 2012 un [informe sobre la transparencia de las universidades](#) revisando sus portales web, y lo ha [actualizado en 2013](#). En dichos informes se valoran positivamente los esfuerzos realizados en favor de la transparencia, especialmente por parte de las universidades públicas, pero todavía se detectan déficits importantes en áreas como la demanda, resultados y satisfacción de los estudiantes, o el perfil del claustro.

## 1.3. ¿Qué es acceso abierto?

La necesidad de promover la difusión de la investigación y la docencia en el contexto de la sociedad de la información ha favorecido que surgiera el movimiento del acceso abierto, conocido con las siglas OA (Open Access), que promueve que el conocimiento sufragado públicamente sea de dominio público y postula un

acceso permanente, gratuito y libre de restricciones a los contenidos científicos y académicos.

Declaraciones internacionales sobre acceso abierto:

- [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#) (22 de octubre de 2003).
- [Bethesda Statement on Open Access Publishing](#) (20 de Junio de 2003).
- [Budapest Open Access Initiative](#) (14 de febrero de 2002).
- [Declaración de CRUE-REBIUN en apoyo del modelo de acceso abierto](#) (2004).

Recomendaciones e iniciativas europeas:

- Recomendación de la Comisión, de 17 de julio de 2012, relativa al acceso a la información científica y a su preservación (2012/417/UE)
- Iniciativa OpenAIRE (Open Access Infrastructure for Research in Europe, Infraestructura de libre acceso a la investigación en Europa. <http://www.openaire.eu/>)

Legislación española:

- Ley 14/2011, de 1 de junio, de Ciencia, Tecnología e Innovación,
- Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado

Varias universidades españolas tienen políticas institucionales de acceso abierto, a pesar de que se excluyen de las obligaciones y recomendaciones RISP “*los documentos conservados por instituciones educativas y de investigación, tales como centros escolares, universidades, archivos, bibliotecas y centros de investigación, con inclusión de organizaciones creadas para la transferencia de los resultados de la investigación (art. 3.3. Ley 37/2007)*”, es decir, se desaprovechó la ocasión para impulsar la reutilización de los contenidos científicos y académicos, aunque afortunadamente ésta se impulsa a menudo por propia iniciativa, dado que son principios que están en sus genes.

Los datos a que se refieren estas iniciativas son típicamente de tres clases:

- Material docente: [OpenCourseWare](#) (OCW) y [Open Educational Resources](#) (OER); y cursos abiertos [Massive Open Online Course](#) (MOOC)
- Las publicaciones y otros documentos más o menos formales que recogen los resultados de la investigación o informan sobre ella, recogidos en [repositorios institucionales](#), frecuentes en nuestras universidades, con un [impulso desde REBIUN](#), la red de bibliotecas universitarias de la CRUE.
- Datos experimentales o técnicos sobre los que se realiza una investigación, para que pueda ser replicada por otros, o para que otros puedan usarlos para sus propias investigaciones. Existe un importante movimiento internacional en favor de la compartición eficaz de datos, vertebrado en la [Research Data Alliance](#) (RDA).

## 1.4. Iniciativas de apertura

A nivel nacional, desde 2001, el proceso de apertura es impulsado ministerialmente mediante el proyecto interministerial Aporta, y su [portal datos.gob.es](#). Las políticas de gobierno abierto se encuentran en una fase de auténtica eclosión por lo que es difícil mantener un catálogo actualizado de iniciativas, así que mucho mejor enlazar [el catálogo CTIC](#), aunque sea razonablemente incompleto. Varias de estas iniciativas recogen en sus portales interesante información introductoria y de referencia: definiciones, principios, catálogos de buenas prácticas, licencias, ejemplos de uso, etc.

Por ahora la única iniciativa de apertura de datos universitaria española recogida en el catálogo CTIC es [la de la Universidad Pablo Olavide de Sevilla](#), aunque recientemente se han celebrado días de datos abiertos también en la Universidad de Granada y en la de Deusto, y nos consta que otras universidades están preparando sus pilotos de apertura de datos, las más recientes, a la hora de redactar este texto: [Universidad](#)

de [Alicante](#), [Universidad Pompeu Fabra](#), [Universidad de León](#) y [Universitat Jaume I de Castellón](#). En cualquier caso, en general, no abundan iniciativas en las universidades del mundo que puedan servirnos de ejemplo, hay un puñado en [linkeduniversities](#), siendo destacable [la de la Universidad de Southampton](#), la de la [Universidad de Ege](#). Así mismo se empiezan a recoger colecciones de datos abiertos, como en [linkededucation](#), que vienen a sumarse a otros recursos en abierto tradicionalmente ofrecidos de forma más o menos dispersa, y que hoy en día las tecnologías de la web semántica nos permiten enlazar, como expone el informe [Linked Data for Open and Distance Learning](#).

Parece evidente que la apertura está en los “genes universitarios”, por lo que debieran ser la norma, no la excepción, políticas explícitas como la de la Universidad Jaume I de Castellón, que dispone en sus Estatutos que *“fomentará el uso de formatos informáticos abiertos en la comunicación interna y externa, promoverá el desarrollo y el uso de software libre y favorecerá la libre difusión del conocimiento creado por la comunidad universitaria”*. Desde luego, existen unos cuantos núcleos de interés en nuestras universidades, así como grupos académicos con perfiles distintos (legal, político, de comunicación, técnico, semántico...). En el recientemente creado capítulo español de la [Open Knowledge Foundation](#) se ha creado un [subgrupo para trabajar hacia la universidad abierta](#). La Open Knowledge Foundation tiene un grupo de trabajo específicamente dedicado a educación, el [Open Education Working Group](#).

La apertura en la universidad se manifestará además en otros ámbitos, complementarios, como el acceso abierto a los resultados científicos, la educación en abierto, o mediante procesos participativos/abiertos en sus acciones de gobierno, de lo que el reciente proceso participativo integrado en la elaboración del [Plan Estratégico de la Universidad Pablo de Olavide](#) es un ejemplo a imitar.

Como muestra de la complementariedad de las aproximaciones a la apertura, un botón: el [curso sobre datos abiertos en la plataforma de cursos online masivos en abierto de la UNED](#).

## 2. Políticas de apertura

Recogiendo las experiencias previas, y entendiendo que el primer paso para avanzar hacia la apertura es la adopción institucional de políticas claras que fijen la orientación y alcance de las iniciativas, en esta sección se proponen formulaciones básicas de estas políticas, consensuadas dentro del grupo de trabajo de CRUE-TIC, que se propone que sean revisadas y debatidas por un conjunto amplio de universidades interesadas. Así, este trabajo servirá a todas las universidades, que adoptarían, o adaptarían, estas formulaciones básicas. Habitualmente las políticas se acompañan de un despliegue de medidas y planes de puesta en marcha (nombramiento de responsabilidades, planes RISP, etc), pero entendemos que ese paso es posterior, y probablemente más dependiente de la idiosincrasia de cada organización, aunque también convenga tomar ejemplo de otros.

### 2.1. Política de transparencia

La *Universidad* <...> se compromete con el principio de transparencia, y declara su disposición a publicar datos abiertos de todos sus ámbitos de actividad (académico, investigación y transferencia, recursos humanos y materiales, y economía) para que esta información pueda ser reutilizada, con las únicas limitaciones que imponga la legislación, en particular la protección de datos personales o los acuerdos de confidencialidad.

### 2.2. Política de gobierno abierto

La *Universidad* <...> se compromete con los principios de gobierno abierto, pues los estima imprescindibles para un buen gobierno y los reconoce plenamente alineados con el espíritu universitario:

- Transparencia: publicación activa de información clara y actualizada, accesible y reutilizable, mediante la que se rinden cuentas de la actividad y se justifican las acciones de gobierno, así como facilidad de acceso a la información cuyo contenido sea solicitado, con los únicos límites que la legislación expresamente determine para los datos protegidos.

- Participación: establecimiento de canales apropiados para que la ciudadanía, informada gracias a la transparencia, pueda intervenir en los procesos de toma de decisiones.
- Colaboración: promoción de la colaboración tanto interna como externa, entre el personal de la universidad, estudiantes, ciudadanía, empresarios y emprendedores, así como otras instituciones tanto públicas como privadas y la sociedad en general, para la creación de mejores servicios de educación, investigación y transferencia tecnológica.

### 2.3. Política de acceso abierto

La *Universidad* <...> se compromete con los principios en favor del acceso abierto, pues los reconoce plenamente alineados con el espíritu universitario, y en consecuencia promoverá las siguientes acciones:

- Dictar mandatos y recomendaciones dirigidos a la comunidad universitaria para que publiquen en abierto los trabajos financiados total o parcialmente por la Universidad.
- Dotar al repositorio institucional de los medios y recursos necesarios para desarrollar su función de archivo, difusión y preservación de los documentos y otros objetos digitales para su acceso en abierto por la comunidad global, garantizando el cumplimiento de la legislación relativa a los derechos de autor en los materiales depositados.
- Difundir el acceso abierto y las licencias abiertas.

## 3. Buenas prácticas de apertura

Aprovechando que este es un camino que ya ha sido emprendido por otras administraciones, lo que se propone, en un espíritu de reutilización, es adoptar o adaptar el [decálogo general elaborado a partir del día open data en Euskadi](#):

0. Armonización entre Administraciones.
1. Publicar datos en formatos abiertos y estándares.
2. Usar esquemas y vocabularios consensuados y utilizar metadatos abiertos.
3. Inventario en un catálogo de datos estructurado.
4. Datos accesibles desde direcciones web persistentes y amigables.
5. Exponer un mínimo conjunto de datos relativos al nivel de competencias del organismo y su estrategia de exposición de datos
6. Compromiso de servicio, actualización y calidad del dato, manteniendo un canal eficiente de comunicación reutilizador <-> AAPP.
7. Monitorizar y evaluar el uso y servicio mediante métricas.
8. Datos bajo condiciones de uso no restrictivas y comunes.
9. Evangelizar y educar en el uso de datos.
10. Recopilar aplicaciones, herramientas y manuales para motivar y facilitar la reutilización.

De forma complementaria, las iniciativas de datos abiertos reciben a menudo [críticas](#), por sus carencias y limitaciones, por estar demasiado tecnificadas y no ser poco más que páginas webs repletas de enlaces y archivos de utilidad dudosa, y sin garantías. Las recomendaciones de buenas prácticas, como estas, pretenden prevenir justamente esos defectos.

Conviene hacer hincapié en que lo primero es la apertura y la transparencia, sólidamente puestas en marcha por administraciones responsables. Más adelante, si la reutilización lo requiere, serán bienvenidas las sofisticaciones técnicas propias de las últimas (de las cinco) estrellas, que no sirven de nada si la información que se publica no es relevante y de calidad.

Por último, apuntar que la tendencia será seguramente hacia iniciativas de transparencia igualmente transparentes: los actuales portales dedicados dejarán de tener sentido cuando toda la información disponible en los propios sitios web de las administraciones esté disponible de forma abierta, en el propio sitio.

### 3.0. Armonización entre Administraciones

Todos los puntos del decálogo se basan en la premisa de que debe existir una armonización entre todas las Administraciones. Todas las iniciativas Open Data deben compartir los mismos principios y definiciones que se listan en el decálogo. Este punto 0 es básico para la interoperabilidad y aprovechamiento eficiente de las sinergias llevadas a cabo por todos los actores Open Data - RISP. En el caso de las universidades, entendidas como distintas administraciones, naturalmente el primer nivel de armonización es entre ellas.

#### 3.1. Publicar datos en formatos abiertos y estándares

Cualquier iniciativa Open Data debería publicar sus conjuntos de datos, así como la documentación, en formatos abiertos (no-propietarios) y que sean adecuados para permitir la reutilización de los mismos por parte del colectivo reutilizador destinatario, para lo que está disponible un [catálogo de estándares](#). Hay diversos perfiles de destinatarios, los orientados a cuestiones de transparencia, que suelen preferir formatos tipo tabla, y los desarrolladores infomediarios, que pueden apreciar los formatos más sofisticados.

#### 3.2. Usar esquemas y vocabularios consensuados y utilizar metadatos abiertos

Además de los formatos abiertos y estándar, la estructura de los datos debería seguir un convenio o unos esquemas definidos, en particular la [norma técnica de interoperabilidad de RISP](#) (NTI RISP). Si se crean vocabularios o esquemas de representación de la información específicos, éstos se deberían exponer públicamente para que el colectivo reutilizador pueda interpretar correctamente la información. Utilizar metadatos documentados y disponibles como "Metadatos Abiertos" preparados para su reutilización (Nivel 3 en el documento de ISA "[Towards Open Government Metadata](#)").

#### 3.3. Inventario en un catálogo de datos estructurado

Cualquier iniciativa Open Data debe tener un punto de consulta donde se incluya un inventario con información descriptiva y técnica sobre los conjuntos de datos que se exponen. Los metadatos que informan sobre cada conjunto de datos debería seguir una estructura común y estándar, y en el caso de datos específicamente universitarios sería necesario consensuarla. Asimismo, se deberían compartir las taxonomías de temáticas u otras necesarias -p.e., toponimia- para clasificar los conjuntos de datos dentro de los catálogos.

#### 3.4. Datos accesibles desde direcciones web persistentes y amigables

Tanto las fichas de los conjuntos de datos, como la distribución de la propia información (volcado en un archivo, API de consulta, RSS, etc.) deberían de estar accesibles desde URLs (direcciones web) que persistan en el tiempo y así evitar que se pierdan las referencias en el futuro. Además deben seguir una estructura homogénea y bien definida, con información legible para que los reutilizadores conozcan o "intuyan" el contenido referido por dichas direcciones web.

#### 3.5. Exponer un mínimo conjunto de datos relativos al nivel de competencias del organismo y su estrategia de exposición de datos

Cada Administración que impulse una iniciativa Open Data debería crear una hoja de ruta donde especifique la estrategia de exposición de los conjuntos de datos y sus prioridades. Inicialmente, debería publicar los conjuntos de mayor interés según las competencias del propio organismo. En el caso de las universidades convendría mucho que las hojas de ruta estuviesen alineadas para promover la combinación/comparación de datos y la eficiencia en el desarrollo de aplicaciones y servicios específicos.

#### 3.6. Compromiso de servicio, actualización y calidad del dato, manteniendo un



## canal eficiente de comunicación entre reutilizador y AAPP

La Administración debe mantener un mínimo de calidad y servicio en su iniciativa Open Data, manteniendo lo expuesto en la estrategia de publicación y comprometiéndose con su colectivo reutilizador. Debe establecer un canal eficiente de comunicación que permita la interacción bidireccional organismo público - reutilizadores.

### 3.7. Monitorizar y evaluar el uso y servicio mediante métricas

La Administración debe crear métricas y evaluar sus indicadores de uso y servicio de la iniciativa Open Data. De esta forma puede monitorizar el funcionamiento y uso, y así analizar si se está cumpliendo el compromiso con la comunidad de reutilizadores y cuales son las potenciales carencias del sistema o de la estrategia.

### 3.8. Datos bajo condiciones de uso no restrictivas y comunes

Las condiciones de uso deberían ser lo menos restrictivas posible y permitir la reutilización libre, incluso para fines comerciales. Se recomienda la creación y uso de licencias tipo, autodocumentadas y que sean comunes entre distintas administraciones.

### 3.9. Evangelizar y educar en el uso de datos

Es necesario educar en el uso de los datos, tanto a los colectivos de reutilización específicos (sector TIC, periodismo, investigación, etc.) como a la sociedad en general y así fomentar el conocimiento y la inquietud por procesar información de una forma autónoma. Contar con una estrategia de comunicación para dar a conocer los datos que se vayan abriendo y potenciar su reutilización. Por ejemplo, a través de redes sociales y medios de comunicación "tradicionales". Evitar el "disgusto" por los datos.

### 3.10. Recopilar aplicaciones, herramientas y manuales para motivar y facilitar la reutilización

Cualquier iniciativa Open Data debería recopilar ejemplos de uso y herramientas que faciliten y motiven la reutilización de los datos que se publican.

En el caso universitario parece obvio que los primeros reutilizadores serán los propios universitarios (investigadores, tecnólogos y estudiantes, montar iniciativas hacktivistas... partiendo de lo local pero sabiendo que es posible - y muy estimulante - [competir globalmente](#)), y es fundamental que se difundan las aplicaciones que se desarrollen (en [datos.gob.es](#), por ejemplo, hay un apartado de Aplicaciones, además de los datos). Por supuesto esto tiene una importante derivada de costes: las aplicaciones interesantes que desarrolle la universidad (o reutilizador) basadas en datos abiertos consensuados le sirven automáticamente a todas, esto incluso cuando se trate de datos "no comparables", es decir, además del beneficio de poder elaborar servicios combinando los datos de todos (los estudios ofrecidos por todos, la investigación realizada por todos, etc).

## 4. Datos universitarios susceptibles de abrirse

Es importante que las universidades muestren, desde un primer momento, el compromiso que tienen como institución con la iniciativa de datos abiertos y quizá la mejor forma de demostrarlo es poniendo a prueba su resistencia a la liberación de información sensible.

Si nuestra universidad nos pregunta ¿qué datos abrimos? deberíamos responderle con otra pregunta ¿qué datos podemos (poder como sinónimo de capacidad técnica y legal de hacerlo) abrir?

Convendría empezar por aquellos que combinen el ser el "más útil-más barato-más sensible". Liberando información de todas las áreas de negocio de nuestra institución, de aquellas áreas de las que ya dispongamos de datos estructurados y que sean fiables. Debemos liberar datos lo más desagregados posible. Debemos mantenerlos actualizados con una frecuencia adecuada a cada caso. Y también datos que permitan

generar servicios, posiblemente los datos georreferenciados de recursos y actividades en un campus sean de los más aprovechables, véase el [ejemplo de Southampton](#).

Un resultado importante del grupo, lo que toca trabajar ahora más intensamente, será un censo inicial consensado de catálogos de datos de todos los ámbitos (acceso, matrícula, calificaciones y graduación de estudiantes, docencia impartida, investigación realizada, recursos humanos, economía, y recursos materiales). Alcanzar este consenso aumentará el valor de los datos liberados, al facilitar su combinación y comparación, y permitirá enlazarlos técnicamente (dando calidad “cinco estrellas” a su publicación), y por otra parte permitirá que las aplicaciones y servicios que se construyan sobre los datos sean compartidas (eficiencia).

La intención es mantener este censo (inicial y evoluciones siguiendo un procedimiento ordenado y ágil) en el marco del ENI, donde seguramente habrá también información sobre las codificaciones generales (empezando por la de las propias universidades, e incluyendo titulaciones, municipios, países, etc).

Con respecto a las descripciones semánticas, deben utilizarse categorías o metadatos “genéricos”, de vocabularios [ampliamente utilizados](#) como el [Data Catalog Vocabulary \(DCAT\)](#), estándar de la W3C, del que hay un perfil adoptado por la Comisión Europea, y si el organismo tiene algunos más específicos una buena práctica es incluirlos como etiquetas. Estos datos de catalogación también es muy importante que se usen de forma coordinada (punto 2 del decálogo).

Una revisión de los portales de datos abiertos de universidades del mundo permite identificar qué tipo de conjuntos de datos son los habitualmente más publicados (aparte de publicaciones o contenidos educativos, ligados al acceso abierto):

## Información organizativa

- Organigrama. Organizaciones y unidades organizativas que forman parte de la estructura de la universidad.
- Perfiles del personal de la universidad (profesores, investigadores, administrativos, etc.)
- Directorio del personal de la universidad.
- Información de ofertas de trabajo en la universidad.

## Docencia

- Catálogo de estudios, diferenciando oficiales y propios.
- Calendarios y horarios de los cursos.
- Programas de movilidad.
- Catálogos de contenidos educativos.

## Alumnado

- Perfil de los alumnos matriculados.
- Programas de becas
- Rendimiento académico (calificaciones de los estudiantes matriculados, porcentaje de alumnos aprobados, tasa de duración de estudios, tasa de abandonos,...).
- Índices de satisfacción de los alumnos.

## Economía

- Estado de cuentas público. Informe financiero anual.
- Presupuesto de ingresos.
- Presupuesto de gastos.
- Distribuciones del presupuesto.
- Ayudas y subvenciones.
- Licitaciones

## Infraestructuras y servicios

- Edificios y otras entidades geoespaciales presentes en la universidad.
- Información geoespacial de los edificios de la universidad (rectorado, facultades, departamentos, aulas,...).
- Imágenes de los edificios e instalaciones de la universidad.
- Información de accesibilidad de los edificios y servicios de la universidad.
- Equipamiento de las aulas, tanto su localización como sus características (capacidad, plazas, accesibilidad, etc.).
- Catering. Menús de la cafetería.
- Instalaciones y equipos de la universidad.
- Accesos.
- Catálogo de servicios informáticos.
- Catálogo de servicios bibliotecarios y catálogo bibliográfico de la biblioteca de la universidad.
- Servicios a estudiantes: alojamientos (número de plazas/tipo, ubicación, precio...), cafeterías, máquinas expendedoras, instalaciones deportivas, actividades culturales.

## Comunicación

- Eventos que se realizan en la universidad.
- Noticias relacionadas con la universidad.

Dentro de las numerosas opciones disponibles para la selección de unos conjuntos de datos iniciales en el ámbito universitario, resulta especialmente interesante la que ofrece el SIIU ([Sistema Integral de Información Universitaria](#)). Estos datos, con las adaptaciones necesarias para hacerlos susceptibles de abrirse, podrían reutilizarse, por ejemplo, para crear aplicaciones que ayuden a elegir universidad, elaborar rankings, o para rendir cuentas.

Otro conjunto de datos relativamente consolidado que cubre otros ámbitos de la actividad (investigación, docencia, RRHH, economía) son las tablas solicitadas por el Tribunal de Cuentas, aceptando la exigencia de transparencia que este proceso de fiscalización supone, y llevándola más allá, entendiendo que es la sociedad entera a quien hay que rendir cuentas, mediante dicho tribunal, o/y directamente. Entre estos datos consideramos que los más interesantes en primera instancia son los datos sobre actividad investigadora, dado que las dos principales actividades de la universidad al servicio de la sociedad son la educación y la investigación y transferencia. Estos datos podrían ayudar a mejorar el conocimiento sobre la financiación de la investigación, o a establecer contactos entre los investigadores, entre sí y con entidades potencialmente interesadas en aplicar o apoyar su trabajo.

Por su potencial como servicio, y su potencial para originar una fértil compartición de recursos desarrollados sobre el mismo, cabe destacar los recursos materiales (edificios, aulas, laboratorios,...) que pueden georeferenciarse tomando como base la información que proporciona [Openstreetmap.org](#) (es espectacular el detalle de la información de los Campus) y herramientas similares a las que ha construido [GeoSpatiumLab](#) (spin-off UZ) para el ayuntamiento de Zaragoza. Es más, es factible organizar hacktivismo de datos para que los alumnos (y PDI y PAS) participen, con lo que se daría difusión a la iniciativa de apertura. Una vez creados los datos se puede dar difusión continuada mediante aplicaciones que "consuman" esta información.

### 4.1. Datasets sobre enseñanzas: oferta y rendimiento, basados en el SIIU

El principal objetivo del SIIU es disponer de unos indicadores del Sistema Universitario Español que sean de calidad, fiables, que reflejen fielmente la realidad, y que permitan la comparabilidad. Además, deben aportar la información necesaria a todos los agentes del sistema (Ministerio, CCAA, Universidades, Estudiantes, Profesores e Investigadores,...).

Para el desarrollo del SIIU, las Universidades ya realizan un importante esfuerzo de normalización y extracción de datos en un formato predefinido y con semántica común. Basta con aplicarle a los datos una pequeña transformación de consenso para conseguir una publicación de datos muy digna y útil con un esfuerzo

razonable. Debe tenerse en cuenta además, que los ficheros del SIIU están estructurados en formato XML, por lo que sería relativamente fácil llegar a alcanzar un nivel de apertura técnica [conforme al esquema 4\\*](#) de datos enlazados

La promoción de la reutilización de la información pública por parte de terceros es otro de los motivos que nos mueve a seleccionar un catálogo de conjuntos de datos derivados de los ficheros que se entregan al SIIU; si cada universidad seleccionara conjuntos de datos distintos, sin acuerdo previo de formatos y semántica, y aún tratándose de información similar en esencia, la comparación no sería posible. Sin embargo, seleccionando conjuntos de datos como los que se propondrán, con formatos y semántica común ya conocidos, incluso con un número pequeño de universidades que se unan a la iniciativa, pueden realizarse comparaciones entre ellas, y promover la creación de aplicaciones o servicios que utilicen esos datos.

Por último, seleccionar el SIIU como origen del catálogo parece razonable dentro del entorno de restricción de recursos en el que nos encontramos. Desarrollar “desde cero” nuevos procedimientos de extracción para nuevos catálogos sería un esfuerzo que no toda universidad podrá acometer, lo que puede constituir un obstáculo para el desarrollo uniforme de la iniciativa en el tejido universitario.

Como punto de partida para elaborar ese censo consensuado, y que resulte económico y aceptado, se revisan los ficheros del SIIU, y las eventuales necesidades de disociación. De entre todos los ficheros del SIIU, se analizan los ficheros que forman parte del área académica por ser actualmente los datos más depurados en todas las universidades españolas, y los que menos resistencia a su apertura pueden ofrecer por parte de las universidades. De entre todos los ficheros analizados, que han sido un total de 31 ficheros, se propone que 18 ficheros formen parte del citado catálogo común y se recomienda también su apertura inicial en todas las universidades.

Estos 18 ficheros contienen los siguientes datos<sup>1</sup>:

- **Fichero 01.01. AVANCE DE MATRÍCULA EN ESTUDIOS DE GRADO (CENTROS PROPIOS).** En este fichero se encuentran las características generales de los estudiantes matriculados en estudios de grado que se imparten en centros propios de la universidad. Contiene la titulación en la que se encuentran matriculados los estudiantes.
- **Fichero 01.05. AVANCE DE MATRÍCULA EN ESTUDIOS DE MÁSTER (CENTROS PROPIOS).** En este fichero se encuentran las características generales de los estudiantes matriculados en estudios másteres oficiales que se imparten en centros propios de la universidad. Contiene la identificación y ubicación de másteres que cursan los estudiantes.
- **Fichero 01.08. AVANCE DE FORMACIÓN EN DOCTORADO RD 56/2005, 1393/2007 Y 99/2011 (CENTROS PROPIOS).** En este fichero se encuentran las características generales de los estudiantes matriculados en el periodo de formación de estudios de doctorado en los centros propios de la universidad. Contiene la identificación y ubicación de los programas de doctorado que cursan los estudiantes.
- **Fichero 02.01. MATRICULA DE ACCESO AL ESTUDIO DE GRADO (CENTROS PROPIOS).** En este fichero se encuentran datos de estudiantes que se matriculan por primera vez en el grado en el curso académico actual en centros propios de la universidad. Están incluidos los estudiantes procedentes de otro grado así como los estudiantes que trasladen expediente desde planes antiguos. Contiene las titulaciones que cursan los estudiantes, datos estadísticos socio-económicos familiares, estudios con los que acceden y nota de admisión.
- **Fichero 03.01. RENDIMIENTO DE ESTUDIANTES MATRICULADOS EN ESTUDIOS DE GRADO (CENTROS PROPIOS).** En este fichero se encuentran datos del rendimiento de todos los estudiantes matriculados en el curso de referencia en estudios de grado impartidos en centros propios de la universidad. Contiene créditos matriculados, superados, presentados, reconocidos y transferidos para

---

<sup>1</sup> NOTA GENERAL: la “identidad” de los estudiantes ha de ser anonimizada, es decir, cada estudiante es representado por un código disociado de su identidad personal.

cada estudiante. Asimismo contiene también la nota media del expediente académico si el estudiante es titulado en el curso de referencia.

- **Fichero 03.05. RENDIMIENTO DE ESTUDIANTES MATRICULADOS EN ESTUDIOS DE MÁSTER (CENTROS PROPIOS).** En este fichero se encuentran datos del rendimiento de todos los estudiantes matriculados en el curso de referencia en estudios de Máster impartidos en los centros propios de la universidad. Contiene datos de residencia familiar y del estudiante, identificación de los máster e información sobre créditos matriculados, superados, presentados, reconocidos y transferidos. Asimismo contiene también la nota media del expediente académico si el o la estudiante es titulado/a en el curso de referencia.
- **Fichero 03.08. FORMACIÓN EN DOCTORADO 99/2011, 56/2005 y 1393/2007 (CENTROS PROPIOS).** En este fichero se encuentran datos del rendimiento de los estudiantes matriculados en el periodo de formación de estudios de doctorado en los centros propios de la universidad. Contiene datos de residencia familiar y del estudiante, acceso al doctorado, identificación de los doctorados e información sobre créditos matriculados y superados.
- **Fichero 03.10. LECTURA DE TESIS (CENTROS PROPIOS).** En este fichero se encuentran datos sobre los doctorandos que han leído la tesis doctoral en el curso de referencia en los centros propios de la universidad. Se incluirán los doctorandos procedentes de los doctorados 99/2011, 56/2005, Y 1393/2007. Contiene información sobre residencia familiar y personal del estudiante, identificación de la universidad que otorgó la suficiencia y calificación de la tesis.
- **Fichero 04.01. ENTRADA EN EL SISTEMA UNIVERSITARIO ESPAÑOL CON PROGRAMA DE MOVILIDAD.** En este fichero se encuentran datos de todos los estudiantes procedentes de instituciones de educación superior extranjeras que están matriculados en el curso actual en la universidad con algún programa de movilidad. Contiene, entre otros, datos de identificación de los estudiantes, universidad de procedencia, tipo de programa de movilidad y fechas de inicio y finalización. Este fichero debe ser anonimizado para su uso por entidades externas a la universidad.
- **Fichero 04.02. SALIDA A UNA UNIVERSIDAD EXTRANJERA CON PROGRAMA DE MOVILIDAD.** En este fichero se encuentran datos de todos los estudiantes de cualquier ciclo formativo (1er y 2º ciclo, grado, máster y doctorado) que están matriculados/as en la universidad y en el curso actual cursan parte de sus estudios en una institución de educación superior extranjera con algún programa de movilidad. Contiene universidad de destino, tipo de programa de movilidad y fechas de inicio y finalización.
- **Fichero 04.03. SALIDA A OTRA UNIVERSIDAD ESPAÑOLA CON PROGRAMA DE MOVILIDAD.** En este fichero se encuentran datos de todos los estudiantes de cualquier ciclo formativo (1er y 2º ciclo, grado, máster y doctorado) que están matriculadas/os en la universidad y en el curso actual cursan parte de sus estudios en otra universidad española con algún programa de movilidad. Contiene la universidad de destino, tipo de programa de movilidad y fechas de inicio y finalización.

También se propone la apertura inicial de ficheros auxiliares del SIU como aquellos ficheros que contienen datos relacionados con la estructura como los centros, departamentos u otras unidades como escuelas de doctorado, centros de investigación, etc. que forman parte de la universidad y estudios de Grado, Máster, Doctorado y programaciones conjuntas que se imparten en la universidad:

- **Fichero 01.01. LISTADO DE CENTROS.** En este fichero se encuentran datos sobre las características de los centros asociados a la universidad, en este fichero sólo se recogen las escuelas y facultades. Contiene, entre otros, datos de los nombres y ubicación de los centros.
- **Fichero 01.02. LISTADO DE DEPARTAMENTOS.** En este fichero se encuentran datos sobre las características de los departamentos que forman parte de la universidad. Contiene, entre otros, datos de los nombres y ubicación de los departamentos.
- **Fichero 01.03. LISTADO DE OTRAS UNIDADES.** En este fichero se encuentran datos sobre nombre, ubicación y tipo de aquellas otras unidades que forman parte de la universidad. Contiene, entre otros, datos de los nombres, ubicación de las unidades y CIF.
- **Fichero 02.01. ESTUDIOS DE GRADO.** En este fichero se encuentran datos de los estudios de grado regulados por el R.D. 1393/2007 que oferta la universidad. Contiene, entre otros, datos de los nombres de las titulaciones, centros de impartición, ramas de enseñanzas, número de créditos necesarios y

número de créditos ofertados.

- **Fichero 02.02. ESTUDIOS DE MÁSTER.** En este fichero se encuentran datos de los estudios de máster regulados por el R.D. 1393/2007 y por el R.D. 56/2005 que oferta la universidad. Contiene, entre otros, datos de los nombres de los másteres, unidades responsables, ramas de enseñanzas, número de créditos necesarios y número de créditos ofertados.
- **Fichero 02.03. ESTUDIOS DE DOCTORADO.** En este fichero se encuentran datos de los estudios de doctorado regulados por el por el R.D. 99/2011, R.D. 56/2005 y por el R.D. 1393/2007 que oferta la universidad. Contiene, entre otros, datos de los nombres de los programas de doctorado, unidades responsables y ramas de enseñanzas.
- **Fichero 02.04. PROGRAMACIÓN CONJUNTA DE ESTUDIOS OFICIALES DE GRADO.** En este fichero se encuentran datos de las programaciones conjuntas de estudios oficiales de grado que oferta la universidad. Contiene, entre otros, datos de los nombres de los programas conjuntos, centros responsables, ramas de enseñanzas asociadas, número de créditos necesarios y número de créditos ofertados.

En una primera fase, para la puesta en marcha inicial, sería conveniente seleccionar un número reducido de estos ficheros por los cuales comenzar el proceso de apertura y que permitan hacer algún tipo de análisis útil. Por tanto, se proponen los 6 conjuntos de datos siguientes:

- **Fichero 02.01. ESTUDIOS DE GRADO.**
- **Fichero 02.02. ESTUDIOS DE MÁSTER.**
- **Fichero 01.01. AVANCE DE MATRÍCULA EN ESTUDIOS DE GRADO (CENTROS PROPIOS).**
- **Fichero 01.05. AVANCE DE MATRÍCULA EN ESTUDIOS DE MÁSTER (CENTROS PROPIOS).**
- **Fichero 03.01. RENDIMIENTO DE ESTUDIANTES MATRICULADOS EN ESTUDIOS DE GRADO (CENTROS PROPIOS).**
- **Fichero 03.05. RENDIMIENTO DE ESTUDIANTES MATRICULADOS EN ESTUDIOS DE MÁSTER (CENTROS PROPIOS).**

## 4.2. Datasets sobre actividad investigadora

Alguna consideración inicial:

- Habrá datos importantes que no abriremos (nosotros) porque no nos pertenecen, pero que es altamente recomendable combinar con los datos que abrimos: índices de impacto, posición relativa de las revistas en las categorías, citas a los artículos, entidades financiadoras...
- Posiblemente no se incluyan en un dataset parte de los datos disponibles, cuando no sean públicos. Por ejemplo no incluir la cuantía de contratos que no sean públicos, o incluirla de forma agregada en un "contrato" sin investigadores.
- Se identificarán nominalmente los investigadores (son autores de PUBLiCaciOneS, y son trabajadores públicos), preferentemente mediante su [ORCID](#), pero no exclusivamente porque (todavía) no es universal (usar un identificador propio prefijado por el código de la universidad, por ejemplo). Se han hecho esfuerzos de estandarización tanto europeos ([EuroCRIS](#), [CERIF](#)) como nacionales ([CVN](#) de la [FECYT](#)).

Conceptualmente, básicamente se incluyen cuatro entidades:

### Investigador

Un *Investigador* es toda persona que realice una labor investigadora reconocida por alguna entidad. Este concepto incluye desde personal investigador permanente, hasta investigadores en formación. Todo investigador tiene un identificador único (atributo *uid*) y opcionalmente un nombre (atributo *nombre*).

### Entidad

Una *Entidad* representa a una colección de personas organizadas formando una comunidad u otra estructura social. Toda entidad tiene un identificador único (atributo *uid*) y opcionalmente un nombre (atributo *nombre*). El

término es amplio ya que los investigadores pueden formar parte de comunidades cuyo propósito no es la investigación, sino, por ejemplo, la docencia. Normalmente se utilizarán especializaciones de este concepto como por ejemplo Universidad, Departamento, Instituto Universitario de Investigación o Grupo de Investigación. La relación *miembro* se utiliza para relacionar investigadores con entidades. Un investigador puede ser miembro de varias entidades de forma simultánea. Se puede representar de forma explícita la duración de dicha relación (atributos *desde* y *hasta*) y se podría incluir el papel en dicha organización (atributo *rol*).

### Proyecto

Un *Proyecto* es la representación de las ayudas públicas, contratos y convenios para el desarrollo de líneas de investigación, servicio, asesorías y otros que los investigadores pueden participar. Todo proyecto tiene un identificador único (atributo *uid*) y opcionalmente un título descriptivo (atributo *título*). Como atributos (según sea el tipo del proyecto) pueden incluirse la financiación, duración, entidad financiadora (y su ámbito o categoría, y eventualmente relacionarla con el exterior).

La relación *participa* se utiliza para relacionar investigadores con proyectos. Un investigador puede participar en varios proyectos de forma simultánea. Se puede representar de forma explícita la duración de dicha relación (atributos *desde* y *hasta*) y el papel en dicho proyecto (atributo *rol*, en particular, si es investigador principal del proyecto).

### Resultado

Un *Resultado* es la representación de cualquier resultado de investigación disponible a partir de un momento determinado (atributo *fecha*). Todo resultado tiene un identificador único (atributo *uid*) y opcionalmente un título descriptivo (atributo *título*). Normalmente se utilizarán especializaciones de este concepto como por ejemplo Artículo y Patente.

Como atributos (según sea el tipo de resultado) interesan por ejemplo la fuente (e.g., revista, y de esta su factor de impacto y su mejor posición relativa en la lista de categorías - esto podría ser un dato externo, o no incluirse ya que es un dato licenciado, en cuyo caso de querer usarse habría de combinarse con la fuente autorizada) y el número de citas (nuevamente es un dato externo, que idealmente se ha de obtener independientemente, por ejemplo de google scholar, y combinarse si se quiere utilizar).

Los resultados de investigación son generados por uno o varios investigadores.

## 4.3 Datasets sobre equipamientos, infraestructuras y mapas

Dentro de la tipología de datos susceptibles a ser abiertos en una universidad, existe una amplia casuística de datos de naturaleza espacial que va desde datos acerca de las propias estancias de la universidad (coordenadas, áreas, uso dado, etc.) hasta consumos energéticos y de recursos (electricidad, gas, agua, etc.) pasando por datos sobre el material inventariable.

Una de las claves del acceso a la información geoespacial reside en el formato elegido para su representación. Los formatos espaciales elegidos deben permitir maximizar la reutilización de los datos. Esta reutilización vendrá condicionada por el nivel de conocimiento de bases de datos espaciales que tenga el potencial reutilizador, no teniendo las mismas necesidades un reutilizador que simplemente quiera dibujar algo sencillo en un mapa de un campus universitario que aquel que quiera elaborar una aplicación para el cálculo de rutas entre edificios de un campus. Por ello, es conveniente ofrecer formatos en función del perfil del usuario. Es decir, la elección de los formatos viene dada por la audiencia que tendrá el portal de datos abiertos. Otro aspecto a considerar es la diferencia entre un servicio que ofrece acceso a datos espaciales y la descarga de un archivo de datos con la información geográfica.

Teniendo en cuenta las anteriores directrices, las buenas prácticas del mundo de la información geográfica proponen como formatos de descarga los estándares internacionales:

- [GML](#) (Geography Markup Language): lenguaje basado en XML para modelar sistemas geográficos y para el intercambio de datos geográficos.

- **KML** (Keyhole Markup Language): lenguaje de basado en GML para representar datos geográficos. Es muy popular por su uso en Google Earth.

Además, proponen que si se quiere ofrecer un servicio de acceso a información geográfica se utilice el interfaz estándar **WFS** (Web Feature Service). Este permite la realización de peticiones de información geográfica en diversos formatos y es útil para reutilizadores "avanzados".

A la hora de hablar sobre datos abiertos de naturaleza espacial o geográfica como son los datasets sobre equipamientos, infraestructuras y mapas se debe comentar que la comunidad GIS tiene una amplia tradición en la liberación de datos mediante mediante portales especializados. Ejemplos de ello son los esfuerzos en estandarización de formatos y servicios para dichos portales realizados por OGC ([Open Geospatial Consortium](#)) o el portal de acceso a la [Infraestructura de Datos Espaciales de España](#) (IDEE) patrocinada por el Gobierno de España. Por ello es razonable definir de forma separada los objetivos de un portal de datos abiertos para una universidad y de un geoportal universitario con respecto a la información espacial, con independencia de que ambos formen parte de una infraestructura común. El portal de datos actúa como un "one-stop point" para acceder a datos (y APIs) públicos (y de carácter comercial) de las diversas comunidades que forman la Universidad, incluidos aquellos de naturaleza espacial. El geoportal de datos abiertos de la Universidad sería uno de los principales reutilizadores de los datos suministrados por el portal de datos abiertos con una alta componente geográfica. También serviría para dar servicio a usuarios especializados del mundo geográfico. Además, como uno de sus objetivos es situar los diversos datos universitarios como capas en un mapa requiere infraestructura dedicada (e.g. servicios de mapas [WMS](#) y [WMTS](#)) y acceso a bases de datos espaciales destinadas a la creación de mapas.

## 5. Ejemplos de reutilizaciones de estos datos

En este apartado se describen a modo de ejemplos o sugerencias algunas aplicaciones/reutilizaciones vistas en universidades:

- **Simplificadores.** Aplicaciones y soluciones que facilitan el acceso a información compleja resumiéndola en una forma más simple. *Ejemplos:* [College Affordability and Transparency Center \(EE.UU.\)](#), [studentaid.gov \(EE.UU.\)](#), [Análisis de alumnos de nuevo ingreso \(U. Pablo de Olavide\)](#), [Vacancy TreeMap \(U. Oxford\)](#)
- **Directorios y agendas.** Aplicaciones y soluciones que permiten localizar ofertas y demandas de recursos y servicios (personas, eventos, edificios, salas, equipamiento, etc.). Pueden presentar la información sobre un mapa para facilitar la localización. *Ejemplos:* [iSoton \(U. Southampton\)](#), [Campusplan \(U. Münster\)](#), [University Science Area Map \(U. Oxford\)](#), [Portal de datos \(Open University\)](#)
- **Herramientas de gestión.** Aplicaciones y soluciones equivalentes a los (pequeños) aplicativos de gestión corporativos, pero construidas sobre los datos abiertos, en lugar de directamente sobre los sistemas corporativos. *Ejemplos:* [Room Finder \(U. Southampton\)](#).

En todo caso, la reutilización no tiene límites. Cabe imaginar, por ejemplo, la combinación de datos universitarios con datos de ciudades "inteligentes" a la hora de hacer la combinación atractiva como destino turístico-educativo, como en <http://www.studyinspain.info/>

## 6. Recomendaciones técnicas

En el presente apartado trataremos de fijar las bases técnicas iniciales que nos permitan ofrecer un soporte adecuado a la hora de publicar nuestros conjuntos de datos (datasets). Esencialmente, el propósito de esta sección es orientar y facilitar el análisis técnico a las Universidades que emprendan el camino hacia la publicación de sus datos en abierto. Se pretende que este esfuerzo de análisis y orientación pueda beneficiar a todos, sin tener que hacerlo desde cero en cada servicio o departamento TI de cada Universidad. A partir del



momento en que varias Universidades pongan en marcha iniciativas, la colaboración entre ellas a través de foros como la CRUE-TIC, y la publicación de indicadores dentro de UNIVERSITIC, deberán servir para conseguir el mejor servicio, la mejor calidad de publicación de datos, con el menor esfuerzo.

Antes de entrar en las tecnologías y herramientas, organizadas de forma temática, es importante detenerse a reflexionar sobre la proporcionalidad entre los medios y los fines, o entre las soluciones y los problemas. En el caso que nos ocupa, la “dimensión del problema” viene determinada por dos medidas:

- Número de colecciones de datos (datasets)
- Frecuencia de actualización o refresco

De forma orientativa, tentativa, y obligadamente difusa (especialmente en sus fronteras) podríamos hablar de tres niveles de complejidad creciente para el desarrollo de un portal de datos abiertos:

1. Hasta 100 datasets, hasta 10 datasets requieren actualización semanal.
2. Hasta 1000 datasets, hasta 100 datasets requieren actualización semanal.
3. Por encima de 1000 datasets o de 100 que requieran actualización semanal.

El Nivel 1 puede describir las necesidades de un prototipo de portal de Datos Abiertos, o las necesidades “en explotación” de una Universidad de tamaño pequeño, mientras que el Nivel 3 puede ser el que corresponde a las grandes Universidades, grupos de Universidades (e.g., todas las catalanas bajo UNEIX, el Campus Iberus), o a un nodo central (e.g., CRUE).

A la hora de analizar las diferentes soluciones tecnológicas en cada escenario que se presentan hay que tener en cuenta que son de carácter orientativo, es decir, no son las únicas posibles, y los sistemas indicados son sólo ejemplos para comprender mejor cómo es la solución propuesta. Además, lo que se describe a continuación es una infraestructura orientativa básica para cada escenario que puede ser extendida. Por ejemplo el nivel más básico (Nivel 1) puede ser complementado con la creación de Web APIs que usan los datos publicados y cuyas URL se indican en los metadatos de dichos datos .

**Nivel 1: menos de 100 colecciones con menos de 10 colecciones actualizadas cada semana.** En este caso una posible solución es la construcción de un portal Web convencional utilizando un servidor Web de uso habitual (Apache, Nginx, IIS). Este portal se caracterizaría por:

- **Sitio web.** Un sitio web convencional sobre un servidor Web de uso general.
- **Datos publicados.** Estarían almacenados en formatos estructurados abiertos o de uso generalizado dentro del espacio del servidor Web, en una nube pública (S3), o en portales temáticos (geoportal, bibliotecas). En este último caso sólo se permite si hay compromiso de actualización por parte de un tercero.
- **Metadatos publicados.** Estarían como fragmentos enriquecidos (microdatos, microformatos y RDFa) de las páginas HTML estáticas que enlazan a los datos publicados. Debe verificarse son sintácticamente correctos desde el punto de vista de los motores de búsqueda comerciales como Google y Bing. También hay que comprobar que en los fragmentos se utilizan esquemas de marcado reconocidos por dichos buscadores (por ejemplo [schema.org](http://schema.org)). De esta forma se garantiza la visibilidad del dato publicado en dichos buscadores.
- **Catálogo de datos.** Formado por las páginas HTML anteriores. Debe asegurarse que el catálogo puede ser recorrido completamente vía enlaces para facilitar su indexación por motores de búsqueda. La búsqueda en el catálogo se implementa mediante herramientas como Apache Sori y Google Site Search.
- **Datos accesibles.** Los datos tienen una URL accesible y persistente con independencia de dónde estén almacenados. Las páginas HTML con sus metadatos tienen una URL accesible y persistente. Los ficheros recuperados están en formatos estructurados abiertos o de uso generalizado.
- **Compromiso de actualización.** Actualización manual de los datos, metadatos y catálogo de datos por un equipo del Servicio de Informática de la Universidad.
- **Realimentación.** Las páginas HTML incluyen la posibilidad de enviar comentarios a los

administradores.

- **Monitorización.** Ya sea a nivel de servidor Web y a nivel de las páginas HTML se ha habilitado algún mecanismo básico de monitorización de acceso y descarga, como los logs de acceso del servidor Web o Google Analytics.

**Nivel 2: 100-1000 colecciones o con 10-100 colecciones actualizadas cada semana.** En este caso la solución puede pasar por la construcción de un portal Web basado en algún sistema de gestión de contenidos que de soporte a la publicación de las entradas del catálogo de datos. Las diferencias con el Nivel 2 se indican en color verde. Este portal se caracterizaría por:

- **Sitio web.** Un sitio web basado en un sistema de gestión de contenidos.
- **Datos publicados.** Estarían almacenados en formatos estructurados abiertos o de uso generalizado dentro del espacio del servidor Web, en una nube pública (por ejemplo, S3), o en portales temáticos (geoportal, bibliotecas). En este último caso sólo se permite si hay compromiso de actualización por parte de un tercero.
- **Metadatos publicados.** Estarían almacenados en una base de datos, siendo servidos dinámicamente como datos, datos enlazados o fragmentos enriquecidos (microdatos, microformatos y RDFa) de las páginas HTML que enlazan a los datos publicados. Debe verificarse son sintácticamente correctos desde el punto de vista de los motores de búsqueda comerciales como Google y Bing. También hay que comprobar que en los fragmentos se utilizan esquemas de marcado reconocidos por dichos buscadores (por ejemplo [schema.org](http://schema.org)). De esta forma se garantiza la visibilidad del dato publicado en dichos buscadores. Considerar el almacenamiento de dichas páginas HTML en una caché para aumentar el rendimiento del sistema.
- **Catálogo de datos.** Formado por las páginas HTML anteriores. Debe asegurarse que el catálogo puede ser recorrido completamente vía enlaces para facilitar su indexación por motores de búsqueda. La búsqueda en el catálogo se implementa mediante una búsqueda sobre la base de datos donde están los metadatos y/o herramientas como Apache Solr o Google Site Search.
- **Datos accesibles.** Los datos tienen una URL accesible y persistente con independencia de dónde estén almacenados. Las páginas HTML con sus metadatos tienen una URL accesible y persistente. Los ficheros recuperados están en formatos estructurados abiertos o de uso generalizado. Se verifica de forma automática que no hay enlaces rotos a los datos.
- **Compromiso de actualización.** Actualización manual de los datos y la base de datos de metadatos por un equipo del Servicio de Informática de la Universidad.
- **Realimentación.** Hay disponibles herramientas básicas de gestión de la comunidad de usuarios.
- **Monitorización.** Hay disponibles herramientas de monitorización de acceso y descarga de datos.

**Nivel 3: Más de 1000 colecciones o con más de 100 colecciones actualizadas cada semana.** En este caso identificamos como una solución adecuada la construcción de un portal Web basado en algún tipo de gestor de contenidos avanzado y/o un sistema especializado en servir datos (e.g., CKAN, Virtuoso Universal Server). Las diferencias con el Nivel 2 se indican en color verde. Este portal se caracterizaría por:

- **Sitio web.** Un sitio web basado en un gestor de contenidos que ofrece diferentes Web APIs sobre los datos y metadatos publicados (acceso, visualización, integración).
- **Datos publicados.** Estarían almacenados en formatos estructurados abiertos o de uso generalizado dentro del espacio del servidor Web, en una nube pública (S3), en un sistema especializado en servir datos o en portales temáticos (geoportal, bibliotecas). En este último caso sólo se permite si hay compromiso de actualización por parte de un tercero.
- **Metadatos publicados.** Estarían almacenados en una base de datos especializada (p.e., la del propio sistema de gestión de contenidos, Virtuoso, Neo4j), siendo servidos dinámicamente ya sea como datos, como datos enlazados o como fragmentos enriquecidos (microdatos, microformatos y RDFa) de las páginas HTML que enlazan a los datos publicados. Debe verificarse son sintácticamente correctos desde el punto de vista de los motores de búsqueda comerciales como Google y Bing. También hay que comprobar que en los fragmentos se utilizan esquemas de marcado reconocidos por dichos buscadores (por ejemplo [schema.org](http://schema.org)). De esta forma se garantiza la visibilidad del dato

publicado en dichos buscadores. Considerar el almacenamiento de dichas páginas HTML en una caché para aumentar el rendimiento del sistema.

- **Catálogo de datos.** Formado por las páginas HTML anteriores. Debe asegurarse que el catálogo puede ser recorrido completamente vía enlaces para facilitar su indexación por motores de búsqueda. La búsqueda en el catálogo se implementa mediante una búsqueda sobre la base de datos donde están los metadatos y/o herramientas como Apache Solr o Google Site Search.
- **Datos accesibles.** Los datos tienen una URL accesible y persistente con independencia de dónde estén almacenados. Las páginas HTML con sus metadatos tienen una URL accesible y persistente. Los ficheros recuperados están en formatos estructurados abiertos o de uso generalizado. Se verifica de forma automática que no hay enlaces rotos a los datos.
- **Compromiso de actualización.** Delegación de la incorporación y actualización de los datos al sistema mediante formularios, flujos de trabajo, validación automática y acceso controlado basado en roles. Los metadatos pueden ser adquiridos y actualizados automáticamente a partir de la información publicada en otros catálogos de datos remotos.
- **Realimentación.** Herramientas avanzadas de gestión de la comunidad de usuarios que incluyen la posibilidad de comunicarse con el propietario real del dato.
- **Monitorización.** Hay disponibles herramientas de monitorización de acceso y descarga de datos.

## 6.1. Estructura de los metadatos de las colecciones de datos

La información disponible sobre cada una de las colecciones de datos debería de tener la siguiente estructura con carácter general. Esta estructura se basa en los estándares DCAT (Dataset Catalog, prefijo dcat:) y DCMI (Dublin Core, prefijo dct:) que han sido adoptados por la NTI RISP citada anteriormente (ver [guía de aplicación](#)) para describir conjuntos de recursos de información.

- **Nombre** (*dct:title*). Nombre o título del documento o recurso de información; preferentemente en varios idiomas.
- **Descripción** (*dct:description*). Descripción detallada de la colección; preferentemente en varios idiomas.
- **Temática(s)** (*dcat:theme*). Las temáticas controladas a las que pertenece la colección de datos. Los términos utilizados deben hacer referencia vía un identificador a un concepto controlado que pertenece a algún sistema de conocimiento (lista controlada, vocabulario, ontología, etc.).
- **Etiqueta(s)** (*dcat:keyword*). Etiquetas, marcas o tags que permiten categorizar libremente la colección de datos.
- **Identificador** (*dct:identifier*). La URI que identifica a la colección de datos.
- **Fecha de creación** (*dct:issued*). Fecha de la expedición oficial (por ejemplo, la fecha de la primera publicación) de la colección de datos.
- **Fecha(s) de última actualización** (*dct:modified*). Última fecha conocida en la que se modificó o se actualizó el contenido de la colección de datos. Si fuera conveniente se puede mantener el histórico de las modificaciones.
- **Frecuencia de actualización** (*dct:accrualPeriodicity*). Periodo de tiempo aproximado entre actualizaciones de la colección de datos si hubiera.
- **Idioma(s)** (*dct:language*). Idioma(s) en el(los) que se encuentra el contenido de la colección de datos.
- **Organismo que publica la información** (*dct:publisher*). Organismo responsable de publicar el dato que puede ser la propia Universidad o uno de sus organismos.
- **Condiciones de uso** (*dct:license*). Condiciones de uso o licencia aplicable a la colección de datos. Se debería especificar una URI que identifique el recurso que define las condiciones de uso.
- **Cobertura geográfica** (*dct:spatial*). Ámbito geográfico cubierto por la colección de datos. Por ejemplo, las calificaciones de una titulación pueden tener como ámbito geográfico el campus donde se imparte.
- **Cobertura temporal** (*dct:temporal*). Fecha de inicio, fin y duración del periodo cubierto por el recurso. Por ejemplo las calificaciones del curso académico 2012-2013 tendrían como inicio el primer día del

curso académico en 2012 y una duración anual.

- **Vigencia de la colección** (*dct:valid*). Fecha en la que se estima que la colección dejará de ser válida, será modificada o será actualizada.
- **Recursos relacionados** (*dct:references*). Enlaces a otras colecciones de datos y a recursos relacionados con la colección de datos. La colección de datos debería relacionarse con esta propiedad a toda la documentación que se considere relevante y que sirva para interpretar la colección de datos de la forma menos ambigua posible. Un ejemplo de recurso relacionado es el diccionario de datos de la colección de datos.
- **Normativa** (*dct:conformsTo*). Enlaces a normativas relacionadas con la colección de datos cuando fueran relevantes. Por ejemplo, la descripción de una colección de datos sobre cursos de doctorado puede estar enlazada con los R.D. 99/2011, 56/2005 y 1393/2007.
- **Distribución(es)** (*dcat:distribution*). Referencias a los volcados de la colección de datos en sus posibles formatos o a documentos que describen dichos volcados. Estos últimos pueden incluir información sobre el tratamiento que ha recibido la colección para generar el volcado.

La descripción puede ser extendida con otras propiedades. Por ejemplo:

- **Resumen** (*dct:abstract*). Una frase corta y comprensible por cualquiera que describe la colección; preferentemente en varios idiomas.
- **Creador de la información** (*dct:creator*). El responsable último de la información, que puede ser un organismo, una unidad, un servicio o un investigador.
- **Contribuidor de la información** (*dct:contributor*). Las personas, organismos, o servicios que han contribuido a la generación de ese dataset.
- **Versión de** (*dct:isVersionOf*). Si la colección de datos es una nueva versión de una colección de datos anterior.

## 6.2. Formatos

El propósito de este apartado es describir los formatos abiertos estructurados que pueden ser utilizados en la publicación de datos abiertos. El uso del término “abierto” en el formato es relevante y no redundante. Un formato “abierto” es aquel en el que la especificación es públicamente disponible y para el que no existe ninguna limitación en cuanto a su reutilización. El uso de formatos sin especificación pública o con limitaciones para su reutilización puede forzar en el peor caso la compra de software prohibitivamente caro o llevar a la pérdida del acceso al dato por obsolescencia técnica. Por ello, la información debe ser publicada codificada en un formato abierto estructurado.

Analizaremos en este punto diferentes formatos abiertos agrupándolos en las siguientes tres categorías:

- Formatos Estructurados (p.e. CSV, XML, JSON), que son los formatos preferidos por los infomediarios.
- Formatos Semánticos (p.e. RDF/XML, Turtle, JSON-LD) que permiten especificar con mayor precisión la semántica de la información
- Fragmentos enriquecidos<sup>2</sup> (p.e. RDFa, [Schema.org](http://Schema.org)) que facilitan que se pueda embeber datos en páginas HTML.

Un punto de vista complementario al presentado son los formatos recogidos en la [Norma Técnica de Interoperabilidad de Catálogo de estándares](#). Esta norma define un catálogo que entre otros recoge formatos estándar abiertos o de uso generalizado que se consideran necesarios para asegurar los aspectos más prácticos y operativos de la interoperabilidad entre las administraciones públicas y con el ciudadano. Extender este uso al ámbito de la apertura de la Universidad hacia la sociedad puede ser razonable. En los puntos siguientes indicaremos qué formatos ya vienen recogidos en dicha norma.

---

<sup>2</sup> El término “fragmento enriquecido” o [rich snippet](#) fue introducido por Google en 2009. Un término alternativo es [HTML semántico](#).

## 6.2.1. Formatos Estructurados

Los formatos abiertos estructurados son los formatos preferidos por los infomediarios y están ampliamente soportados por todo tipo de aplicaciones. Los más populares son CSV, XML y JSON. Al ser formatos abiertos estructurados no semánticos son formatos de [nivel 3\\* de datos enlazados](#).

- **CSV** (texto separado por comas) es un formato simple utilizado habitualmente para intercambiar datos tabulares entre programas que trabajan habitualmente con formatos incompatibles.
- **XML** es un lenguaje de marcas estandarizado por W3C. Los documentos XML pueden validarse contra un lenguaje o gramática XML de dominio definida, por ejemplo, con los lenguajes DTD o XML Schema.
- **JSON** es un formato de intercambio de datos muy ligero. Es muy popular como alternativa a XML en las APIs de Web.

El catálogo de estándares recoge de forma explícita los formatos CSV, XML y perfiles de XML como SVG, GML ([ISO 19136: XML para Cartografía vectorial y Sistemas de Información Geográfica](#)) y OpenDocument ([ISO 26300: XML para documentos](#)).

## 6.2.2. Formatos semánticos

Los formatos semánticos identifican a una familia de especificaciones del W3C para datos estructurados que permite representar no sólo datos sino también conocimiento. Son los formatos adecuados para publicar información conforme [conforme al esquema 4\\* y 5\\*](#) de datos enlazados.

Los formatos semánticos se pueden organizar por propósito:

- Facilidad para proceso con herramientas XML
  - **RDF/XML**: El primer formato de intercambio de RDF en XML. Formato de referencia para Tim Berners-Lee.
  - **OWL/XML**: Formato de serialización en XML de ontologías codificadas en OWL 2.
- Facilidad para proceso de grandes volúmenes de datos
  - **N-Triples**: Formato de líneas de texto plano para el intercambio de grandes colecciones de datos (e.g. [dbpedia](#)).
  - **N-Quads**: Formato de líneas de texto plano que permite identificar la pertenencia de una sentencia RDF a una determinada colección. Utilizado para el intercambio de grandes colecciones de datos (e.g. [dbpedia](#)).
- Facilidad para leer/escribir por humanos
  - **Turtle**: Formato de intercambio RDF en texto plano pensado para ser leído por humanos. Este formato es compatible con N-Triples.
  - **TriG** (recomendación candidata): Extensión de Turtle que permite representar el concepto de un dataset en RDF.
  - **Functional Syntax** : Formato de serialización en texto plano que facilita ver la estructura forma de una ontología codificada en OWL 2.
  - **Manchester Syntax** : Formato de serialización en texto plano que facilita leer/escribir una ontología DL codificada en OWL 2.
- Facilidad para proceso en clientes Web
  - **JSON-LD**: Perfil de JSON capaz de intercambiar RDF. De muy reciente creación.

Los siguientes formatos se pueden utilizar para exportar datos enlazados y sus vocabularios relacionados, pero todavía no han sido estandarizados o solo son propuestas de recomendación: [N3](#) (sometido), [Binary RDF](#) (sometido) y [Trix](#).

El catálogo de estándares recoge de forma explícita los formatos [Turtle](#) y [N3](#), y de forma implícita los formatos asociados a los estándares RDF ([RDF/XML](#)) y OWL2 ([OWL/XML](#), [Functional Syntax](#) y [Manchester Syntax](#)) como formatos admitidos.

Un repositorio de información semántica puede ser en lugar de publicado ser accedido mediante preguntas a un punto SPARQL. Los formatos estándar en los que devuelve la respuesta un servicio SPARQL son también conformes al esquema 4\* y 5\* y también vienen recogidos en el catálogo de estándares.

### 6.2.3. Fragmentos enriquecidos

Otro mecanismo a considerar son los fragmentos enriquecidos. Estos consisten en añadir cierto marcado al HTML que ya existe en los portales de información actuales con el fin de darle semántica y enriquecer su contenido. Esto permite su procesamiento por terceras partes o por alguno de los buscadores existentes. Las dos aproximaciones principales que existen son:

- [RDFa](#): Enriquecimiento de documentos con marcado RDF previamente definido (incluido en la NTI de Catálogo de estándares).
- Tags HTML: Aplicación de ciertos marcadores semánticos ([microdatos](#), [microformatos](#)) previamente definidos (por ejemplo en [Schema.org](#)) y que permiten el mejor tratamiento de datos personales, eventos, etc.

Finalmente, W3C ha definido un mecanismo estándar denominado [GRDDL](#) que permite introducir marcado semántico en un documento XML (incluyendo páginas XHTML) y transformarlo en RDF mediante XSLT.

El catálogo de estándares recoge de forma explícita el formato embebido [RDFa](#) como admitido.

## 6.3. Almacenes de información enlazada

### Almacenamiento de tripletas: Triple Store

Inicialmente se han evaluado las siguientes soluciones de código abierto para el almacenamiento de datos enlazados:

	Version	SPARQL	OWL	Licencia
OWLIM-Lite	5.3	1.1	Sí	Gratuita, no libre
AllegroGraph	4.1	1.1	Parcial	Gratuita para < 5M tuplas, no libre
4store	1.1.5	1.1	No	GPLv3
Virtuoso	6.1.6	1.1	Sí	GPLv2

Otras opciones a valorar: [Sesame Store](#), [Apache Jena TDB](#), [BigData](#) o [Mulgara](#).

Referencias y comparativa entre los sistemas de almacenamiento propuestos:

- <http://en.wikipedia.org/wiki/Triplestore>
- <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1046&context=techmasters>
- <http://www.w3.org/wiki/RdfStoreBenchmarking>

Según la evaluación realizada y haciendo prevalecer como criterios la flexibilidad, [rendimiento](#), soporte de estándares y licencia abierta, recomendamos el uso de Virtuoso como soporte para la persistencia de tripletas RDF.

## Bases de datos de grafo

Las bases de datos de grafo ofrecen un mecanismo mucho más general a la hora de relacionar información que los almacenes de tripletas, los cuales están diseñados para trabajar exclusivamente con el metamodelo de RDF. Existe la posibilidad de dar el mismo soporte con una base de datos de grafo, que con un almacén de tripletas, con la ventaja de poder relacionar información y publicarla en distintos formatos de una forma más general y abierta. [Neo4j](#) es un ejemplo de base de datos de grafo que tiene integración con RDF mediante [Sail](#).

## Bases de datos relacionales

Dado que la mayor parte de los sistemas de las Universidades utilizan a día de hoy bases de datos relacionales, una opción que compatibiliza los actuales sistemas con la exposición de Linked Open Data es utilizar software de mapping de estas bases de datos con esquemas RDF. Un ejemplo es el software abierto [D2RQ](#), que permite mediante configuración de correspondencias exponer datos enlazados y un endpoint SPARQL sobre una base de datos relacional. En cualquier caso y al igual que en las otras soluciones, puede ser recomendable utilizar este enfoque sobre una réplica de la base de datos original, para no comprometer el rendimiento y estabilidad de la base de datos operativa desde la que se toman los datos.

## 6.4. Lenguajes de consulta

Posiblemente, [SPARQL](#) es a día de hoy el estándar de facto a la hora de realizar consultas sobre conjuntos de datos enlazados. Es por esto que adicionalmente a la posibilidad de descarga directa de los datasets, es muy recomendable ofrecer un endpoint de consulta sobre los propios datos.

Por otra parte, también es interesante ofrecer un servicio de búsqueda full-text que permita buscar sobre el contenido en sí. Las búsquedas full-text pueden de igual forma resultar interesantes a la hora de navegar por la información siempre que se defina un facetado adecuado de la información. Motores de búsqueda como SOLR o ElasticSearch, los cuales se basan en Apache Lucene, pueden ayudarnos a la hora de definir las facilidades de búsquedas full-text, siendo además los más utilizados en la mayoría de los entornos documentales.

También es de gran utilidad la posibilidad de recuperar toda la información relacionada a un identificador cualesquiera. Este servicio puede ofrecerse de manera conjunta al soporte a la desreferenciación en datos enlazados de todos los identificadores HTTP publicados por la organización.

Finalmente, con el objeto de facilitar el desarrollo de soluciones de integración, sería deseable ofrecer un servicio de conversión de etiquetas en identificadores. Un ejemplo de estas consultas lo ofrece el [Reconciliation Service API](#) desarrollado para [OpenRefine](#) (anteriormente conocido como GoogleRefine).

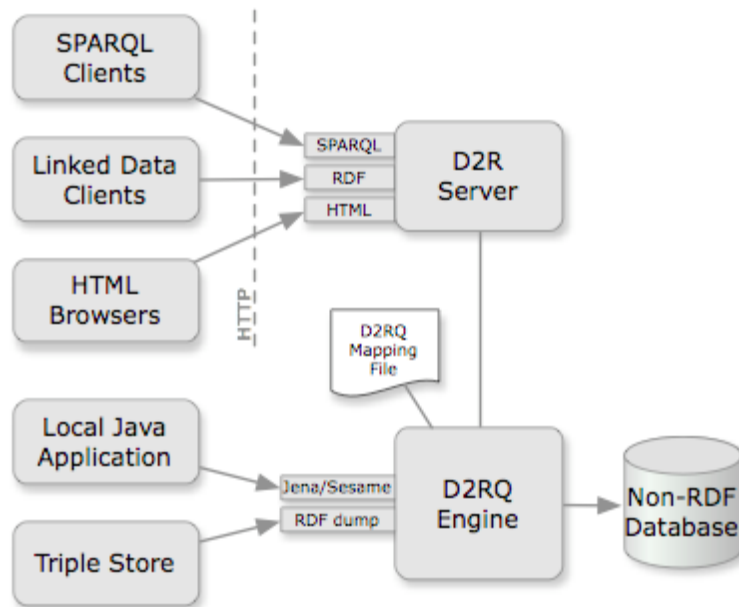
Un buen ejemplo de todas las anteriores aproximaciones es [el portal de datos de Ordnance Survey](#) (UK).

## 6.5. Frameworks

Existen en la actualidad varios frameworks que podemos considerar a la hora de desplegar nuestra plataforma de publicación de datos abiertos. De entre ellos, cabe destacar:

- [D2R Server](#): A tool for publishing relational databases as Linked Data

<http://d2rq.org/>



- Callimachus: Framework de trabajo integrado con Sesame Storage, edición visual de recursos publicados y soporte para búsquedas con SPARQL.

<http://callimachusproject.org/index.xhtml?view>

- CKAN: (<http://ckan.org/>) es un framework completo para la publicación, edición y gestión de datos abiertos desarrollado por la Open Knowledge Foundation con soporte para Linked Data, SPARQL y RDFa. Destaca su interoperabilidad con catálogos de información estándar (Z39.50, CSW) y catálogos Web legados, las herramientas de búsqueda que proporciona, el soporte a la información geoespacial (tanto en búsqueda como en visualización), la inclusión de herramientas sociales, y la posibilidad de actuar como almacén de los datos publicados. [publicdata.eu](http://publicdata.eu) y [data.gov.uk](http://data.gov.uk) son casos de éxito de CKAN.

Como complemento a estos entornos o frameworks, existe la posibilidad de integrar ciertas librerías de utilidad en nuestros desarrollos, con el fin de trabajar de una forma más cómoda con los distintos formatos, conectar con almacenes de datos o realizar búsquedas. Algunos ejemplos destacables son:

- Tinkerpop blueprints: Interfaz genérica para el acceso a distintas implementaciones de base de datos de grafo. Incorpora un storage propio por defecto llamado blueprints, pero puede utilizar el de cualquier tipo de base de datos de grafo existente.

<http://www.tinkerpop.com/>

- Spring Data REST HATEOAS: Sencilla exportación de entidades de datos vía REST.

<http://www.springsource.org/spring-data/rest>

## 6.6. Referencias

Algunos sitios de referencia que pueden ser tomados como ejemplo a la hora de construir el punto de entrada al portal de publicación de datos:

<http://lodum.de/>



<http://www.upo.es/datos-abiertos/#.USpiJx3gmAg>

<http://data.southampton.ac.uk/>

<http://datos.gob.es/datos/>

<http://opendata.aragon.es/>

<http://datos.fundacionctic.org/sandbox/catalog/faceted/>

<http://www.datosabiertos.jcyl.es/>

<http://opendata.euskadi.net/w79-home/es/>

Algunos ejemplos de portales universitarios, en niveles muy dispares de desarrollo, no necesariamente para servir como ejemplo (a seguir) todos ellos:

[http://data.upf.edu/similar\\_data\\_portals](http://data.upf.edu/similar_data_portals)

## 6.7. Integración con los sistemas de información existentes

Existen dos formas para la publicación como dato abierto de datos almacenados en los sistemas de información existentes:

- A través de ficheros en un formato estructurado abierto (CSV, XML) que son exportados directamente desde los sistemas de información que cada una de las universidades tiene disponibles. La ventaja de este enfoque es la simplicidad en la generación de los ficheros y en su gestión, mientras que la principal dificultad se encuentra en la actualización periódica que se hace necesaria en este contexto.
- A través de APIs que permiten acceder a datos generados mediante procesos conocidos como ETL (extracción, transformación y carga) a partir de los sistemas de información. La ventaja de este enfoque es la facilidad de acceso a los datos correspondientes, y su posible transformación al vuelo en formatos abiertos ligeros como JSON, JSON-LD, etc. La dificultad de esta aproximación radica en el hecho de que estas APIs son solo lectura.

Un ejemplo de la primera aproximación es [el portal de datos abiertos desarrollado por la UPO](#) que exporta sus datos en formato CSV y utiliza [CKAN para su difusión](#). Una descripción detallada de este portal se encuentra en la Sección 7.1. [El portal de datos abiertos desarrollado por la UJI](#) es un ejemplo de la segunda aproximación. Este portal tiene una capa abstracta intermedia que realiza un ETL desde los datos de los sistemas de información de la UJI. Los datos extraídos se ofrecen ya sea directamente en el portal Open Data o a través de un API rest solo lectura. Una descripción detallada de este portal se encuentra en la Sección 7.4.

Una vez publicada la información surge la necesidad de federar los datos publicados por las Universidades. La federación de fuentes de datos diversas se puede facilitar normalizando el acceso a las fuentes de datos. De esta manera los datos pueden ser obtenidos con facilidad por aquellos que quieran federar esta información. Si se elige esta vía hay actualmente dos aproximaciones ampliamente utilizadas. La primera aproximación es la exposición del catálogo de datos de acuerdo con estándares como DCAT. La segunda aproximación consiste en la exposición de los datos utilizando APIs de acceso ampliamente utilizadas como la de CKAN. De esta manera, iniciativas de colaboración como la existente entre la CRUE-TIC y [Red.es](#) tendrían fácil la federación de datos universitarios en portales de datos como [datos.gob.es](#).

## 7. Casos

Unas pocas universidades han puesto en marcha, o están desarrollando, iniciativas de datos abiertos. En esta sección se describen algunas de estas iniciativas, su contexto y su orientación.

## 7.1. Universidad Pablo de Olavide

La iniciativa opendata de la Universidad Pablo de Olavide de Sevilla tiene su origen en las II Jornadas de Gobierno Abierto y Opendata Sevilla celebradas en noviembre de 2012 en las que se presenta una versión beta del portal Opendata de la UPO. Desde entonces la UPO ha estado comprometida con el gobierno abierto llevando a cabo experiencias muy innovadoras de procesos participativos para la elaboración del Plan Estratégico de la Universidad como la organización de cursos de verano centrados en la reutilización de la información en el ámbito local. En febrero de 2014 en las III Jornadas de Gobierno Abierto y Opendata se presenta la versión estable del [portal opendata de la UPO](#) que usa CKAN como framework para la gestión de datos en abierto. Los datos publicados se encuentran en formato estructurado csv directamente exportados del sistema de información de la UPO, aunque no se descarta la publicación en otros formatos. Los metadatos de las colecciones de datos son los siguientes:

Metadato	Descripción	Inclusión
Identificador {dct: identifier}	URI que identifica al documento o recurso de información.	Opcional
Nombre {dct: title}	Título o nombre de cada conjunto de dato.	Obligatorio, Múltiple
Descripción {dct: description}	Descripción detallada de cada conjunto de datos.	Obligatorio, Múltiple
Fecha de creación {dct: issued}	Fecha de creación del conjunto de datos. Es la fecha inicial de publicación.	Opcional
Fecha de última actualización {dct: identifier}	Fecha de actualización de datos del conjunto de datos, la fecha en la que se modificaron sus datos.	Opcional
Frecuencia de actualización {dct: modified}	Periodo de tiempo aproximado transcurrido entre actualizaciones del conjunto de datos.	Opcional
Temáticas {dct: theme}	Materia del recurso. Se debe utilizar los términos definidos en el Anexo II, Anexo II.d de la NTI.	Obligatorio, Múltiple

Etiquetas {dct: keyword}	Palabras clave asociadas al conjunto de datos.	Obligatorio, Múltiple
Organismo publicador {dct: publisher}	URL del organismo que publica el dataset, debe seguir el formato establecido por la NTI (ver Anexo II, Anexo II.b)	Obligatorio
Idioma(s) {dct: language}	Idioma(s) en la que se encuentra la información del conjunto de datos.	Obligatorio, Múltiple
Condiciones de uso {dct: license}	Enlace (URL) a una página web con la información detallada acerca de la licencia.	Opcional
Normativa {dct: conformsT}	Enlaces (URLs) a documentos normativos con información relativa a los datos del dataset	Opcional
Cobertura geográfica {dct: spatial}	Cobertura geográfica que abarca los datos del dataset. Se deberá indicar mediante identificadores a nivel nacional, regional o provincial (ver Anexo II, Anexo II.f de la NTI)	Opcional
Cobertura temporal {dct: temporal}	Cobertura temporal que abarcan los datos del dataset (ver Anexo II, Anexo II.c de la NTI)	Opcional
Recurso(s) Relacionado(s) {dct: references}	Enlaces a recursos relacionados con el conjunto de datos.	Obligatorio, Múltiple
Distribución(es) {dct: distribution}	Referencia a los recursos que identifican los valores volcados del dataset.	Obligatorio, Múltiple

Actualmente hay publicados once conjuntos de datos: “Presupuesto de la Universidad Pablo de Olavide. Ejercicio ...”, para los ejercicios 2008 a 2014, “Calificaciones en asignaturas de Grado” y “Estudiantes de

nuevo ingreso en Máster Oficial”.

Estos datos han sido reutilizados a través de simplificadores para ofrecer diferentes visualizaciones de los conjuntos de datos publicados. Los simplificadores usados han sido herramientas de visualización propias del portal, así como a través de la plataforma [OpenSpending](#) de la OKFN y el software gratuito [Tableau Public](#). En concreto, el portal muestra visualizaciones de la población y país de origen de los estudiantes de nuevo ingreso de Máster Oficial del curso 2012-2013, la provincia de origen de los estudiantes de grado del curso 2012-2013, la edad media de los estudiantes de Máster Oficial, la modalidad de acceso de los estudiantes de Grado del curso 2012-2013, las calificaciones de los estudiantes de grado en el curso 2011-2012, entre otras.

Actualmente, la UPO está trabajando en la definición del Plan de Reutilización de la Información Pública (RISP) en colaboración con red.es y en la implementación de un punto SPARQL para pasar a un grado de apertura de 4 o 5 estrellas y favorecer el desarrollo de aplicaciones por parte de usuarios avanzados y empresas del sector infomediario. En particular, para la puesta en marcha de este punto de consulta se ha optado por el formato semántico RDF, la elección de una URI de acuerdo con la Guía de utilización de la NTI y la solución de código abierto Apache Marmotta para el almacenamiento de tripletas.

Por último, la UPO se ha incorporado al grupo de trabajo creado en el seno de Open Data España, la comunidad que reúne a las administraciones públicas españolas que están avanzando en las políticas de apertura de datos y a entidades y profesionales expertos en la materia, como la Fundación CTIC, partner español del W3C. El primer proyecto en el que se está trabajando es en la elaboración de un directorio de la estructura organizativa que se publicará en open linked data.

## 7.2. Universitat Pompeu Fabra

Desde su fundación en 1990, la Universidad Pompeu Fabra ha adoptado un modelo basado en la transparencia, la calidad, la investigación y la innovación.

En términos de transparencia y rendición de cuentas, la Universidad ha destacado por publicar datos institucionales incluyendo sus programas de estudio, estructura organizativa y logros en proyectos de investigación. El ejemplo más representativo de estos esfuerzos es el "UPF en xifres", publicado casi desde su fundación en los años 90.

La UPF quiere seguir mejorando su rendición de cuentas y transparencia aprovechando la oportunidad que ofrece el desarrollo del linked open data como el siguiente paso natural por parte del sistema de educación superior pública en España. Según esto, el desarrollo de iniciativas de datos abiertos enlazados debe considerarse como una prioridad dentro de la estrategia global de la Universidad, apoyada con las herramientas tecnológicas necesarias que pueden fomentar la creación de una plataforma que puede convertirse en la columna vertebral del actual modelo de transparencia.

El Vicerrectorado de Calidad e Innovación institucional promovió una iniciativa institucional de datos abiertos para la UPF. Esta iniciativa fue presentada a la Comisión de calidad de la Universidad y oficialmente se inició a principios de 2013. Actualmente, la ejecución del proyecto se engloba en la estrategia recogida en las líneas estratégicas 2013-2015 de la universidad

Teniendo en cuenta el movimiento de datos enlazados, parte de este proyecto consistía en vincular la mayoría de los datos publicados con recursos similares en otras universidades, con el propósito de apoyar la creación de una red de datos universitarios. El movimiento de datos abiertos enlazados va más allá del beneficio individual de cada institución, puesto que los datos universitarios deben ser agregados, integrados y comparados. Esto significa que sólo publicar información en formatos de datos abiertos no es suficiente, y el uso de formatos de datos enlazados debe promoverse como una buena práctica dentro de los esfuerzos de publicación de cada universidad.

Para desarrollar el proyecto de la UPF, se consideraron las siguientes actividades, basadas principalmente en los pasos recomendados de la guía "[Linked Data for Open and Distance Learning](#)", el "[Open Data Handbook](#)" y las directrices del libro "[Linked Data: Evolving the Web into a Global Data Space](#)"

No.	Actividad	Descripción
1	Estudiar el movimiento de datos abiertos enlazados a través de las universidades europeas.	Este paso fue incluido al principio del proyecto para entender las mejores prácticas y tendencias a tomar en consideración para la iniciativa de la UPF.
2	Aprender y descubrir las fuentes de datos diferentes de la UPF a nivel institucional.	Identificar todos los datos conocidos y desconocidos que la UPF tiene a nivel institucional. Este proceso es necesario para saber qué datos pueden generar un valor agregado para la Universidad, considerándolo como un recurso sin explotar.
3	Elegir los primeros conjuntos de datos que serán publicados.	Elegir la lista de conjuntos de datos que se publicarán en el comienzo del proyecto.
4	Determinar la licencia correcta para cada conjunto de datos.	Elegir una licencia para cada conjunto de datos, para especificar cómo se deben utilizar los datos y los límites.
5	Determinar los metadatos comunes que se utilizarán para todos los conjuntos de datos.	Elegir los metadatos que se utilizará para publicar los datos. Los metadatos son parámetros que describen los datos, como nombre, autor, mantenedor, origen de datos o calidad.
6	Transformar y publicar las bases de datos en formatos abiertos y subirlas a un sistema de gestión de datos.	Transformar los datos (es decir, la normalización, discretizing, etc.) a formatos raw y publicarlos en un sistema de gestión de datos (es decir, CKAN).
7	Seleccionar los conjuntos de datos que deben ser transformados a los formatos de datos enlazados.	Escoger la lista de conjuntos de datos que posteriormente se transformará a datos vinculados.
8	Modelo de los datos para los datos enlazados.	Conjuntos de datos que son modelados según ontologías y vocabularios específicos.
9	Transformar los datos a los estándares RDF.	Los conjuntos de datos se transforman a los formatos de datos enlazados.
10	Vincular datos relacionados con los datos de otras universidades.	Vincular los datos con los conjuntos de datos similares de otras universidades.
11	Que los datos disponibles y detectable mediante un SPARQL Endpoint.	Los conjuntos de datos haciendo visible a través de un punto de Sparql con gráficos RDF.

12	Especificar una estrategia de datos abiertos enlazados por la UPF.	Se define la estrategia del proyecto.
13	Documentar el proceso interno y flujo de trabajo para la apertura de datos futuros.	Se documentan los procesos formales para datos de editorial.
14	Crear una lista de recomendaciones para un plan de comunicación y promoción interna para el proyecto.	Las recomendaciones para el desarrollo de un plan de comunicación y promoción están documentadas. Este paso se centra en promover el uso del portal y comunicar los beneficios del proyecto.
15	Extender la en UPF en xifres, información a publicar utilizando la plataforma de datos abiertos enlazados.	La UPF en xifres se transforma en formatos de datos vinculados con el fin de crear un portal dinámico a través de la ayuda de la plataforma de datos abiertos enlazados.

En términos de datos enlazados, un aspecto importante a considerar es la modelización de los datos a través de ontologías o vocabularios que permitirán representar relaciones y conceptos abstractos. En la actualidad, existen algunos vocabularios u ontologías que pueden utilizarse para representar diferentes tipos de datos. Sin embargo, no están estandarizados para el dominio de la educación y esto es importante para una correcta agregación, integración y comparación de datos universitarios. El último punto se debe considerar como una prioridad a nivel europeo. Esto puede hacerse a través de proyectos internacionales financiados por la Comisión Europea o por los esfuerzos locales de cada país (es importante entender que las normas internacionales deben crearse aprovechando las normas locales que ya se utilizan en algunos países y ciudades para representar sus datos, haciendo que la relación lógica entre ellos).

### 7.3. Universidad de Alicante

La iniciativa de Gobierno Abierto de la Universidad de Alicante se ha erigido en un valor fundamental dentro de las políticas TI trazadas en su nuevo plan estratégico 2014-2019 ([PEUA40](#)). Una de las líneas prioritarias de esta iniciativa es el desarrollo de un proyecto de Reutilización de la Información de la Universidad de Alicante (denominado OpenData4U). Uno de los primeros hitos remarcables de esta línea ha sido la puesta en marcha del Portal de Datos Abiertos de la Universidad de Alicante: [datos.ua.es](http://datos.ua.es).

La Universidad de Alicante viene impulsando desde hace tiempo varias iniciativas relacionadas con el Gobierno Abierto que contribuyen a potenciar la convicción de que la universidad debe ser transparente, participativa y colaborativa. La base de estas iniciativas es la potenciación del Gobierno TI en la Universidad de Alicante, cuya clave radica en la responsabilidad en la toma de decisiones estratégicas a través de la puesta en marcha de su Cartera de Proyectos TI Estratégicos. Gracias a esta iniciativa, el equipo de gobierno de la Universidad de Alicante cuenta con la herramienta perfecta que facilita su plena participación en la toma de decisiones referente a los proyectos TI que se desarrollarán cada año en la universidad, y la comunidad universitaria está informada de los criterios, decisiones y estado actual de cada uno de ellos. Existen varios proyectos TI en la Universidad de Alicante directamente relacionados con su portal de datos abiertos, complementándose de manera estratégica, por ejemplo, [Universidad en Cifras](#) donde se muestran análisis de diversos indicadores en pro de la mejora continua de la calidad en el ámbito universitario; Sistema de Información Geográfica de la Universidad de Alicante ([SIGUA](#)), [UAedf](#) que condensa en un único portal todo el conocimiento docente y divulgativo que posee la Universidad de Alicante, con independencia de la naturaleza de la plataforma sobre la que se ofrece ([OCW UA](#) o MOOCs) o el Repositorio de la Universidad de Alicante ([RUA](#)) que ofrece acceso abierto a documentos generados por los miembros de la Universidad de Alicante en su labor de docencia e investigación.

El principal objetivo del proyecto OpenData4U es crear una metodología para la implantación de un modelo de reutilización de información en la Universidad de Alicante que favorezca el desarrollo de un mercado de infomediarios entre estudiantes de últimos años y recién egresados para mejorar su inserción laboral y desarrollar un tejido productivo sostenible mediante la creación de servicios y su posterior comercialización. En concreto en el proyecto OpenData4U se han planteado una serie de acciones:

- Alineamiento con iniciativas de ámbito nacional e internacional como el grupo de trabajo de datos abiertos de la CRUE-TIC, el grupo de trabajo de universidades de la OKFN o el [portal de datos abiertos a nivel nacional](#).
- Creación de un mapa de datos: implantación de un procedimiento de identificación y catalogación de datos susceptibles de apertura.
- Puesta en marcha del portal [datos.ua.es](#): un portal en el que se recoja toda la información del proyecto, datos liberados, usuarios y aplicaciones infomediarias, etc.
- Dinamización del mercado de infomediarios: desarrollo de un conjunto de acciones orientadas a fomentar aptitudes de reutilización, desde concursos de ideas, hasta hackathones o desconferencias, pasando por jornadas y cursos sobre gobierno abierto y la apertura de datos en la universidad.

El portal [datos.ua.es](#) tiene como objetivo impulsar la reutilización de la información pública generada por una institución como es la Universidad de Alicante con la doble finalidad de:

- Generar un valor social para la ciudadanía, aportando transparencia a través de la exposición de los datos que la Universidad de Alicante genera. Al facilitar el acceso, se fomenta la rendición de cuentas.
- Impulsar un impacto económico positivo en el entorno, mediante nuevos modelos de negocio innovadores basados en la reutilización de los datos abiertos de la Universidad de Alicante, agregando valor a los datos para convertirlos en servicios útiles a la sociedad. Por tanto, la apertura de datos permite la generación de oportunidades para el emprendimiento y la innovación, entre la comunidad universitaria, así como entre la ciudadanía en general (incluyendo emprendedores o empresas del sector infomediario).

La tipología de los datos expuestos es amplia y está alineada con las recomendaciones que se exponen anteriormente en este documento. Con el fin de facilitar la reutilización, además del conocido [esquema cinco estrellas](#), se ha aplicado un modelo de calidad para portales de datos abiertos<sup>3</sup>, desarrollado en el seno de la Universidad de Alicante, que establece y mide diversos criterios de calidad de importancia para los datos abiertos. La intención del portal de datos abiertos de la Universidad de Alicante es cumplir con el máximo nivel de calidad llegando a tener gradualmente una infraestructura semántica que permita la gestión de datos enlazados abiertos (*Linked Open Data*) en todos los ámbitos universitarios.

El portal de datos abiertos de la Universidad de Alicante contiene también un catálogo de aplicaciones cuyo objetivo es facilitar la búsqueda de aplicaciones desarrolladas usando datos abiertos de la propia universidad. Como ejemplo de aplicación, se puede destacar una aplicación para teléfonos móviles desarrollada por el [Servicio de Informática](#) de la Universidad de Alicante: [UAlojamientos](#), que reutiliza los datos de la bolsa de alojamiento y alquiler de pisos de la Oficina de Información al Alumno de la Universidad de Alicante para ayudar en la búsqueda del mejor alojamiento a estudiantes.

## 7.4. Universitat Jaume I

La Universitat Jaume I de Castellón apuesta por un enfoque pragmático y enfocado a facilitar lo máximo posible la labor tanto de las partes interesadas en la información publicada como de los integradores y los desarrolladores de soluciones de reutilización de nuestros datos abiertos.

Desde el 2002 con la primera aprobación de los estatutos de la Universidad se incluye en su disposición adicional Séptima que *"La Universitat Jaume I fomentará el uso de formatos informáticos abiertos en la comunicación interna y externa, promoverá el desarrollo y el uso de software libre y favorecerá la libre difusión del conocimiento creado por la comunidad universitaria. La Universitat promoverá, incentivará y facilitará la*

---

<sup>3</sup> Edgar Oviedo, Jose-Norberto Mazón, José Jacobo Zubcoff: Towards a data quality model for open data portals. CLEI 2013: 1-8

*preservación y el libre acceso a las publicaciones y producción intelectual de su personal en el repositorio documental institucional".*

Para ello, disponemos desde el 2002 del programa "[Conocimiento Abierto](#)": con un responsable académico que se encarga de coordinar las iniciativas tanto estratégicas como operativas que se quieran desarrollar para la libre difusión del conocimiento abierto de la universidad. Como resultado tenemos en marcha el portal [Libertas](#) "Portal de Conocimiento Abierto de la UJI", desde el que se puede acceder:

- Al [repositorio institucional](#): Repositorio documental de la UJI. Es donde se archiva, etiqueta, organiza y difunde los documentos electrónicos, de cualquier tipo, creados por los miembros de la comunidad universitaria como resultado de su actividad académica, docente, investigadora e institucional.
- A la colección [Sapientia](#) (con ISBN): Esta colección pública manuales docentes en línea y en abierto, correspondientes a las materias que se imparten en la Universidad Jaume I para que cualquier docente, estudiante o autodidacta los pueda utilizar.
- Al portal del [OCW](#) (OpenCourseWare del MIT) de la UJI: Contiene materiales docentes creados por profesores para la formación superior bajo licencia Creative Commons. La producción de estos materiales se incentiva anualmente con una convocatoria dirigida al profesorado.
- Desde el 2013, una [convocatoria](#) dirigida al profesorado para incentivar la producción de MOOC.
- Un portal de [MOOC](#): Cursos en línea abiertos y masivos, es decir, cursos de acceso abierto que se realizan a través del web y están enfocados a la participación interactiva a gran escala. Además de materiales de estudio escritos, gráficos o vídeos, típicamente incluye foros de debate y otros mecanismos que contribuyen a crear comunidades de aprendizaje.
- Proyecto de [Videoclases](#) del profesorado: Es portal complemente al anterior. Permite al profesorado que ha grabado en vídeo alguna de sus clases en las aulas docentes, dotadas de mesas multimedia, gestionar sus vídeos teniendo la posibilidad de editar, bajar, borrar y validar los mismos para poder publicarlos voluntariamente en el Aula virtual o en otros canales que la universidad ofrezca.

Trás la publicación de la DIRECTIVA 2013/37/UE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 26 de junio de 2013 por la que se modifica la Directiva 2003/98/CE relativa a la *reutilización de la información del sector público*, y trás la aprobación de la Ley 19/2013, de 9 de diciembre, llamada [Ley de Transparencia, Acceso a la Información Pública y Buen Gobierno](#), recientemente hemos puesto en marcha las siguientes líneas de trabajo:

- La aprobación de la [Licencia UJI](#) (con resolución del rector) que se aplica *al software desarrollado por la Universidad Jaume I y las condiciones en que se puede distribuir y la reutilización de la información pública de la Universidad (open data)*.
- Un portal para el [software Libre desarrollado en la UJI](#) basado en el framework de desarrollo actual de CRUE-TIC, bajo la licencia UJI y que recoge los resultados de la Cartera de Proyectos anual de Gobierno TI de la UJI.
- Un portal para el [Open Data](#): Para fomentar la transparencia, la reutilización, la participación y la colaboración entre todos los actores que pueden estar interesados en utilizar los datos de la UJI en formato abierto.

El portal de OpenData desarrollado da cobertura a dos ámbitos distintos. Por una parte, ha servido para dar soporte a la publicación abierta de información institucional, y por otra ha permitido concentrar todas las APIs y servicios de provisión de datos implementados hasta el momento alrededor del ERP universitario (exportaciones, sincronizaciones, consultas, WebServices SOAP, servicios REST, etc).

Con este planteamiento, se abre la puerta a dos escenarios distintos:

- La pública de acceso abierto y procesamiento de [datasets](#) enlazados.
- La provisión privada de conjuntos de datos a aplicaciones internas e integradores.



Aunque la forma de acceso en ambos escenarios sea distinta, la infraestructura desarrollada para dar soporte a ambas es totalmente reutilizable y proporciona un paraguas idóneo para la gestión integral de todos los orígenes de datos provistos por la institución.

Al margen del seguimiento de la mayoría de las recomendaciones detalladas en el presente documento en cuanto a formatos y estándares se refiere, se ha hecho énfasis en algunos aspectos adicionales como son:

- **Arquitectura REST.** Cualquier conjunto de datos expuesto por la plataforma, se ofrecerá al cliente siguiendo al máximo los principios definidos por HTTP, [REST](#) y [Hypermedia](#).
- **Soporte para múltiples formatos de representación.** Cualquier conjunto de información extraído, transformado y publicado por la plataforma, se almacena internamente en un formato abstracto que permite su posterior exportación a RDF/XML, Turtle, JSON, CSV o cualquier formato que queramos añadir.
- **Documentación de calidad.** Documentación de los servicios expuestos y del acceso a los mismos lo más completa y flexible. Para ello, al margen de la implantación de la ficha de cada dataset con la información detallada en este documento, se ha implantado a nivel de todas las APIs expuestas [Swagger](#) como herramienta de generación de documentación “viva”. Este tipo de herramientas hacen de los servicios expuestos, una representación interactiva, autodocumentada y ejecutable de los servicios.

En definitiva, se ha desarrollado una capa de intermediación que permite la publicación y persistencia automática de definiciones de datasets, independientemente de su origen y formato de publicación elegido.

## 7.5. Universidad de León

La Universidad de León (ULe) puso en marcha su [portal de datos abiertos](#) en septiembre de 2013 a través de un convenio con Telefónica S.A.:

Con él la ULe quería mostrar su compromiso con la apertura de los datos. En la ULe entendemos que como administración pública que somos, tenemos que rendir cuentas ante la sociedad a la que servimos y que nos financia. Además, estamos convencidos de que los diferentes catálogos de datos que iremos publicando pueden servir para la generación de nuevos servicios. Es decir, creemos que la política de transparencia no debe servir sólo como herramienta de fiscalización de la actividad de las organizaciones, creemos que también puede servir para mejorar la calidad de los servicios que presta.

Además de los datos referidos a la propia gestión, es también intención de la ULe publicar datos generados en las actividades de investigación que complementen a nuestro repositorio institucional de documentación, [BULERIA](#), que recoge los resultados elaborados, publicaciones fundamentalmente.

Históricamente los diferentes órganos de la ULe han venido publicando información diversa: la gerencia tiene publicados en la web los presupuestos desde 2005; la oficina de calidad tiene publicado el catálogo completo de procedimientos administrativos, al igual que las memorias de los títulos incluidos informes detallados por titulación de datos estadísticos de alumnos de nuevo ingreso, de rendimiento, etc.; los diferentes vicerrectorados publican igualmente sus memorias referidas a investigación, medioambiente, etc. Toda esta información se viene publicando en forma de memorias elaboradas.

Con este portal la ULe pretendía incorporarse a los proyectos de transparencia y gobierno abierto de la comunidad autónoma a la que pertenece ([portal de datos abiertos de la Junta de Castilla y León](#)) y a la estrategia regional en este ámbito publicando datos no procesados.

Ante la falta de marcos de referencia claros para el sistema universitarios en el 2013, la ULe decidió que en lo referente a los datos de gestión lo apropiado era publicar los mismos datos que se envían a otras administraciones. En particular, se ha comenzado por publicar información enviada al SIIU en el área académica, después de un proceso de anonimización de los datos.

En lo referente a catálogos de datos generados por grupos de investigación se decidió ofertar el portal de forma universal y gratuita para todos los grupos de investigación de la universidad.

La ULe tiene pendiente establecer una normativa de publicación activa de datos a través del portal de datos abiertos, donde se establecerán los periodos de publicación, los datos a publicar y los responsables de dicha publicación. Igualmente está pendiente establecer los procedimientos de publicación pasiva, es decir, la petición de datos por miembros de la comunidad universitaria o terceros interesados en los mismos términos: plazos de respuesta (atendiendo a lo dispuesto en la Ley 19/2013), responsables, canales de solicitud y respuesta, etc.

## 7.6. Universidad de Granada

La Universidad de Granada ha impulsado la creación de una página web de acceso abierto con la publicación de los datos relativos a dicha universidad, llamada *UGR transparente*. Desde la Oficina de Software Libre (OSL) se ha contratado un grupo de trabajo formado por personal de informática y documentación, cuya tarea específica es el desarrollo de este portal de datos abiertos. Las pautas que se han seguido para la estructuración de la información, además de este mismo documento, han sido los criterios que la [Fundación Compromiso y Transparencia](#) recomienda en su informe [Examen de Transparencia](#), así como la Ley 19/2013, de 9 de diciembre, llamada [Ley de Transparencia, Acceso a la Información Pública y Buen Gobierno](#). Como modelo a seguir, se ha tomado el de la [Universidad de Málaga](#) y el de la [Universidad de Alcalá](#).

Sin embargo, aunque la finalidad original de la puesta en marcha de este proyecto fue la página web *UGR transparente*, pronto se advirtió que esta iniciativa no sería suficiente y se comenzó a crear la base de datos abiertos de la Universidad de Granada, *Opendata UGR*, a modo de inventario de datos estructurado. La iniciativa *Opendata UGR* está gestionada con CKAN, de manera que cada conjunto de datos es accesible desde su propia URL. De esta manera, el proceso de trabajo transcurriría de la siguiente manera:

- Búsqueda y recopilación de datos.
- Extracción y transformación de datos a un formato libre y abierto.
- Comprobación e inserción de datos en la base de datos abiertos *Opendata UGR*.
- Traspaso y publicación de datos desde *Opendata UGR* hacia la web *UGR Transparente*.

La búsqueda y recopilación de datos se ha realizado a través de la propia página web de la universidad ([www.ugr.es](http://www.ugr.es)) donde se encuentran publicados la mayoría de los datos en diversos formatos que dificultan su tratamiento (.pdf, .jpg, etc.) y en áreas dispersas de la universidad, lo que complica el manejo de la información por parte del usuario. En este caso, la misión del grupo de trabajo era extraer los datos y transformarlos a un formato abierto y estándar (se eligió el formato .csv). Por otra parte, muchos datos no estaban disponibles en la web, por lo que se han tenido que pedir directamente a las instituciones de la propia universidad; en este caso, Rectorado, Gerencia, Oficina Económica y la Sección de Estadística se han mostrado especialmente colaborativas. De este modo, se solicitó un correo electrónico institucional especialmente para el proyecto de datos abiertos, pues se consideró necesario para la gestión y la comunicación entre el grupo de trabajo y las diferentes secciones de la universidad: [opendata@ugr.es](mailto:opendata@ugr.es).

Por ahora, se han establecido como datasets prioritarios los que se corresponden con los criterios de la Fundación Compromiso y Transparencia, para después continuar con las recomendaciones OKFN. Hasta el momento, los únicos datos que no se han podido encontrar han sido los relativos a los siguientes criterios del informe de la Fundación Compromiso y Transparencia:

- (5.3) Evolución (demanda de plazas con respecto al año anterior): no existen dichos datos
- (9.2) Estados financieros (balance de situación y cuenta de pérdidas y ganancias): empresa privada

Por otra parte, un problema que se repite es el hecho de que la información no está actualizada, por lo que se debe contactar con la sección que corresponda para advertir sobre ello y solicitar datos recientes, además de notificar e informar sobre la Ley de Transparencia, las medidas que la Universidad de Granada ha tomado en

cuanto a ello y los criterios que se deben cumplir para la correcta ejecución de dicha ley. Así, se pretende fomentar el procesamiento de la información de forma autónoma.

La Universidad de Granada, por el momento, no tiene habilitadas *Opendata UGR* ni *UGR Transparente* para su consulta pública (aún en construcción), pero por el momento se encuentran publicados los siguientes conjuntos de datos: “Ley de Transparencia” (requisitos a cumplir), “Fundación Compromiso y Transparencia” (criterios a cumplir), “Justificantes de gasto UGR 2014”, “Misión”, “Plan estratégico”, “Gobierno”, “Estatutos”, “Perfil del profesorado”, “Canales de comunicación”, “Becas y ayudas”, “Investigación (resultados)”, “Satisfacción de alumnos”, “La UGR en los *rankings* internacionales”, “Tasas de rendimiento académico”, “Aulas de informática UGR”, “Matrículas” (2010/2011/2012), “Presupuesto 2013: desglose de ingresos” y “Presupuesto 2013: desglose de gastos”. Según el plan de trabajo, a finales del presente curso estarán publicados todos los datos relativos a los criterios de la Fundación Compromiso y Transparencia y ambas páginas web (*Opendata UGR* y *UGR Transparente*) habilitadas.

Por último, se deben resaltar las actividades que la Oficina de Software Libre (OSL) de la Universidad de Granada ha desarrollado a partir del desarrollo del proyecto de datos abiertos, como varios *hackathon*, las jornadas de datos abiertos *Open Data Day*, colaboración en las “II Jornadas de Periodismo de Datos y Opendata” (Granada y Almería), además de futuros cursos y encuentros que se encuentran en proceso.

## 8. Conclusiones

Nos encontramos en los inicios de un probable camino hacia el gobierno abierto, empezando por la transparencia, que se practica idealmente mediante la apertura de datos. Esta nueva forma de relación entre las instituciones, al menos las públicas, y la ciudadanía está siendo poco a poco demandado por ésta, pero sobre todo ha sido identificada como necesaria, y beneficiosa, en algunas instancias políticas supranacionales o por algunas administraciones nacionales o locales pioneras.

Las universidades, como administraciones públicas que también son, no debieran quedar rezagadas en este camino, complementario a la apertura del conocimiento que está en su naturaleza, y que también está adquiriendo nuevas formas (acceso abierto, cursos abiertos en línea, software de fuentes abiertas, etc).

La autonomía universitaria no debería perjudicar este importante avance, y para ello es necesario un esfuerzo de coordinación que está en el corazón de todas las iniciativas de apertura en las administraciones, en las que se requiere armonización y cumplimiento de normas de interoperabilidad. Esta armonización debiera ser especialmente estrecha en el caso de las universidades, pues daría más valor a la reutilización de sus datos y redundaría en mayor eficiencia.

La coordinación, que se quiere promover desde la sectorial TIC de la CRUE, alcanza también a las soluciones técnicas, adaptadas a cada caso, de forma que no se tengan que replicar los esfuerzos de análisis y diseño de soluciones, y se favorezca la colaboración.