

Máster de Gestión de Contenidos Digitales

Marcado de objetos digitales en repositorios *open access* con licencias *Creative Commons:*

Desarrollo de una guía y prueba piloto en un
repositorio.

Alumno:
Sebastiano Giorgi Scalari

Tutores:
Ignasi Labastida i Juan
Cristóbal Urbano Salido

06/2012

Agradecimientos:

A mis tutores: Cristóbal e Ignasi.

A Cristóbal, que considero el Profesor que todos querrían ser y tener a su lado, por su profesionalidad y capacidades extremas, pero también por la humanidad y la empatía que todos hombres querríamos ostentar.

A Ignasi, por unas tutorías que han sido momentos de placer intelectual intenso y por introducirme a una realidad mejor gracias a nuestras propias ideas.

A mi familia:

A Mireia, cuyo amor hace de mi vida algo especial.

A Pau y Pep, por las risas y los juegos que recuperaremos.

A Benedetta, Allegra, Matilde, Tommaso, Olivia, Paolo e Filippo “peressercisempre”.

A mis padres y mis suegros, ellos saben...



ÍNDICE

| | |
|--|----|
| 1. Resumen ejecutivo | 4 |
| 2. Presentación..... | 6 |
| 2.1 La elección de los derechos de autor aplicados a los objetos digitales | 8 |
| 2.2. Contextualización del trabajo | 10 |
| 2.3. Interés del trabajo | 12 |
| 2.4. Grado de ejecución de proyecto | 13 |
| 3. Definición y objetivos del proyecto..... | 15 |
| 3.1. El “DIPÒSIT DIGITAL” UB” y la política OA de la universidad | 15 |
| 3.2. Especificaciones y definición del proyecto..... | 19 |
| 3.3. Objetivos específicos | 21 |
| 4. Análisis del problema | 22 |
| 4.1 Análisis de la organización | 22 |
| 4.2 Naturaleza de los repositorios digitales | 24 |
| 4.2.1. Tipología de licencias en la gestión de objetos digitales | 30 |
| 4.2.2. Historia reciente de licencias abiertas en la UB | 32 |
| 4.2.3. Creative Commons | 33 |
| 4.3. Métodos de representación de la relación entre los derechos y la licencia | 35 |
| 4.3 La gestión de derechos en repositorios digitales: estado de la cuestión | 53 |
| 4.4. Necesidad de normalización en el mercado | 56 |
| 4.5 Antecedentes y resultados previos que avalan la propuesta | 57 |
| 5. Metodología propuesta..... | 60 |
| 5.1 Desarrollo de una guía de ingesta | 61 |
| 5.2 Realización de una prueba piloto..... | 62 |
| 6. Recursos necesarios | 65 |
| 7. Presupuesto..... | 66 |
| 8. Ejecución y control | 67 |
| 9. Resultados esperados..... | 69 |
| 10. Política de difusión de la guía..... | 70 |
| 11. Referencias | 71 |
| 12. Anexos | 74 |
| 12.1 Guía de ingesta | 74 |

1. Resumen ejecutivo

El objetivo principal de este proyecto es formular una propuesta para identificar y marcar los derechos de autor en objetos digitales alojados en un repositorio institucional, gestionado con el programa Dspace. En concreto, por un lado se analiza el estado del arte del mercado digital con licencias *Creative Commons*, y por otro se identifican otros metadatos a añadir al conjunto ya implementado en el repositorio de la Universitat de Barcelona, el "DIPÒSIT DIGITAL", para conseguir la declaración automática de las políticas de derechos de autor seguidas por las revistas donde se publican las obras de investigadores y docentes de la misma universidad.

En el presente proyecto se plantea el desarrollo de las siguientes acciones.

- La habilitación de los metadatos dedicados a la representación de los derechos de autor (como por ejemplo los elementos *dc.rights* y *dc.rights.uri*) en el momento inicial de envío de un objeto al repositorio, de modo que se habilite la generación y declaración de las políticas de derechos de autor de los objetos en la fase de su recuperación y búsqueda.
- La personalización de un aplicativo de código libre que permita el marcado con licencias de derecho de autor de ficheros en formato PDF.
- La preparación de un plan de acción que detalla las fases, los costes estimados y la implementación de una prueba piloto para la validación del éxito de las acciones de habilitación de nuevos metadatos, tanto en los objetos como en el repositorio). En la prueba piloto, a través de la recuperación de artículos y publicaciones existentes en bases de datos externas de control y gestión de la investigación (como por ejemplo desde el GREC, el gestor de los *curricula* de los profesores de la UB), se hace posible el marcado incrustado los derechos de autor y se adjuntan informaciones relativas al tipo de políticas de derechos de autor adoptadas por las revistas.

El panorama de los sistemas de gestión del conocimiento se presenta cada vez más estructurado y maduro, y muchos recursos están fluyendo en estos repositorios institucionales abiertos, en busca de la sincronía perfecta ente ellos. Las políticas hacia el libre flujo del conocimiento pueden mediar en el potencial de desarrollo de una sociedad (Friend, 2012). A nivel social, esta situación representa un momento de aparición de nuevas formas de producción cultural, que no se sustenta en los modelos

de negocios habituales, y que tampoco comparte las mismas motivaciones socio-culturales como base de su producción. Es una oportunidad para un nuevo contrato social que se negociará entre los autores (*peer producers*) y la extendida comunidad de usuarios, distribuidora y, a la vez, consumidora.

El efecto estimado es la mejora de la gestión de los derechos de autor en su globalidad, gracias a la capacidad de representación semántica del mercado digital de los derechos.

2. Presentación

El contexto digital actual se caracteriza por la proliferación de materiales sin identificación de derechos de autor, a pesar de *que existe* la tecnología capaz de marcarlos.

De todos los materiales que se generan en el entorno académico, este proyecto sólo considera un tipo de resultados de investigación: los artículos científicos.

Cualquier edición, científica o no, se apoya en la protección legal otorgada por los derechos de autor. Paralelamente al uso del clásico “© *all rights reserved*”, que prevé una cesión total de los derechos de explotación del autor a favor del editor, existen *alternativas de protección* y uso de las obras intelectuales para su publicación y difusión.

El editor y el autor pueden elegir, y pactar, licencias que permitan modular los derechos de explotación de manera distinta al “todo excluido” previsto en el tradicional y restrictivo “todo reservado”. Estas licencias, que se encuentran agrupadas bajo el concepto de *Open Content Licensing*¹, se han confeccionado para facilitar la libre difusión de los contenidos en red. El denominado movimiento de acceso abierto a la ciencia (Abadal, 2012), que persigue la libre accesibilidad de los contenidos científicos, se concreta en los llamados *open access journals*, un modelo de revista de libre consulta y difusión.

Las revistas científicas en formato electrónico que eligen este modelo son aquéllas que permiten el acceso abierto y gratuito, así como la reutilización.

Esta situación crea una ecología de la publicación académica mixta, en la cual conviven hoy distintas soluciones para el acceso abierto:

- Con la ruta verde se indica la vía que el autor utiliza cuando autoarchiva su contenido, después de la revisión por pares, directamente en un repositorio institucional o temático. En el flujo editorial clásico esto corresponde a la publicación de una versión *pre-print*, o artículo en fase dos. El autoarchivo refleja la voluntad de potenciación del RI como herramienta alternativa de difusión.
- Con la ruta dorada se identifica una solución que prevé que las revistas sean de acceso abierto, gratuito y con permiso de reutilización. El modelo de negocio que

¹ Estas licencias derivan de las licencias originadas para la difusión del software libre a final de los años 90 del siglo pasado. El contexto europeo de aplicación de las licencias abiertas es estudiado en (Guibault & Angelopoulos, 2011). En esta publicación también se encuentra un artículo muy ilustrativo de este movimiento: Grassmuck, V. “Toward a new social contract: free-licensing into the knowledge commons” (páginas 21-50), en el cual se explica el origen i el valor de los bienes comunes, los “*commons*”, bajo una perspectiva sociológica.

sustenta estas revistas puede ser, o no, el pago por parte del autor (o de su institución financiadora) por la publicación del artículo.

- Con el modelo híbrido se indican las revistas tradicionales que ofrecen abrir aquellos artículos que se pagan, pero mantienen el resto de contenidos cerrado. Este modelo suele permitir la reutilización pero, no siendo una condición imprescindible, no siempre se realiza.

Por todo ello, los RI se configuran como una solución para que los artículos de revistas, así como otros *outputs* de la investigación, puedan ser accesibles en abierto. El autoarchivo de estas versiones por parte del autor puede ser realizado de forma voluntaria, respondiendo a motivaciones personales de incrementar la visibilidad de su obra y, por lo tanto, su prestigio académico-científico, o bien ser consecuencia de los mandatos de las instituciones y agencias financiadoras. Otra consecuencia de estas disposiciones es la alimentación de los RI por parte de algunos editores, en los términos del acuerdo específico que se adopte y, normalmente, respetando un periodo de embargo.

Un objeto digital, como obra intelectual, está legalmente protegido como en cualquier otro contexto de uso y difusión.

La falta de indicación de los derechos de autor asociados a la obra corresponde, en el marco legal contemporáneo, a la atribución de “todos derechos reservados”. Por ello, en el contexto OA resulta esencial explicitar los derechos asociados al objeto digital, indicando la licencia que se atribuye. La disponibilidad y visibilidad de la licencia es imprescindible para comunicar las libertades de uso en ella permitidas.

Sin embargo, en el medio digital se necesita también que estos objetos, con sus relativos derechos, puedan ser identificados y manejados de manera eficaz por las máquinas. Esta necesidad semántica es la razón por la que se han ido creando sistemas de marcado invisibles al ojo humano.

Este trabajo revisa el estado del arte del mercado digital de los derechos de autor con una panorámica de los actuales métodos disponibles.

2.1 La elección de los derechos de autor aplicados a los objetos digitales

La correcta gestión de los derechos de explotación (de reproducción, distribución, comunicación pública y transformación de una obra) es una cuestión muy importante, y mucho más si es para favorecer el acceso abierto, puesto que el acceso incondicional a una obra es el resultado de las decisiones que el autor (normalmente) toma sobre los aspectos que su obra puede generar.

Los científicos están acostumbrados a tener que firmar los CTA (*copyright transfer agreement*), que consisten en acuerdos de cesión total de los derechos sobre la obra a publicar, incluyendo ciertas renunciaciones sobre los usos del trabajo. Sin embargo, hoy es posible publicar sin la necesidad de abandonar todos los derechos, gracias a la alternativa del acceso abierto. Esto se puede conseguir, por ejemplo, otorgando al editor una limitada licencia de publicación y manteniendo así el resto de derechos de explotación. O también eligiendo una de las varias licencias de acceso abierto disponibles y redactadas por instituciones que hayan optado por no utilizar el tradicional copyright restrictivo.

La elección de las licencias abiertas para la difusión del trabajo científico es todavía una realidad minoritaria, pero en constante crecimiento. El incremento existente en la aplicación de estas licencias, combinado con la aprobación de políticas institucionales de acceso abierto a la ciencia (como la disposición aprobada en junio de 2011 por el Consell de Govern de la Universitat de Barcelona), pone de manifiesto el interés del tema en el que se enmarca el presente proyecto.

En general, las instituciones de investigación de financiación, están llevando a cabo políticas para difundir su producción mediante la vía verde y la vía dorada, y dado que los artículos científicos son el medio principal de promoción del trabajo de investigación académica, las editoriales comerciales se ven afectadas por presiones, cuyos efectos están facilitando la aparición de modelos de negocio alternativos y en directa competición con el sector comercial. En esta situación el volumen total de artículos disponible en abierto se vuelve un elemento crítico de atracción.

La capacidad de aumentar la “masa crítica” de los contenidos de los repositorios digitales de acceso abierto, depende de elementos culturales, económicos, institucionales,

tecnológicos y legales; pero también del tiempo, de la motivación para publicar, y del valor añadido de publicar en una revista con cierto índice de impacto. En definitiva, la capacidad de crecimiento del volumen de las publicaciones en acceso abierto depende de unas condiciones que tienen que configurar un entorno que apoye al autor en el momento de divulgar su trabajo.

Con la voluntad de intervenir en algunos de los aspectos mencionados, en particular los tecnológicos, se propone la implementación de técnicas de marcado de derechos a fin de conseguir una gestión más eficaz y eficiente de los derechos de autor asociados a los artículos científicos y divulgados a través del canal del repositorio institucional (en adelante, RI). Con éstas técnicas se pueden incrustar los derechos de autor en el mismo objeto digital depositado, de manera que cuando se difundan puedan reconocerse las libertades y obligaciones asociadas a la obra aún cuando circulen fuera de la plataforma editorial o del RI como en el caso que abordamos en el presente proyecto. Con las buenas prácticas de marcado, además de formalizar visualmente en el documento los derechos asociados, se consigue que también las máquinas puedan manejarlos y entenderlos.

2.2. Contextualización del trabajo

Desde la perspectiva del RI, la herramienta creada para sustentar la vía verde al acceso abierto requiere una sistematización y automatización de la información referente a la gestión de derechos, puesto que la actual es deficitaria bajo varias perspectivas. La propuesta de desarrollo de una función más efectiva de marcado de derechos de autor para los objetos digitales depositados en el RI, se deriva de una necesidad expresada por el Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI) de la Universitat de Barcelona que, en calidad de gestor del "Dipòsit Digital" (el RI de la Universitat de Barcelona, en adelante, DD), ha evidenciado una problemática común a todos los repositorios. La redacción de una guía de ingesta con buenas prácticas, calibrada sobre la necesidad del "mercado incrustado"², quiere ser la base y el sustrato para el desarrollo de nuevos elementos tecnológicos de gestión del repositorio.

El presente trabajo es el resultado de la aplicación de los conocimientos adquiridos durante el curso de Máster de Gestión de Contenidos Digitales de la Universitat de Barcelona del bienio 2010-2012. La finalidad del trabajo es la demostración de las capacidades asumidas gracias al estudio de las doce temáticas cursadas. Se adopta la estructura típica del *project management*, cuya metodología permite una descomposición jerárquica de niveles múltiples que, de hecho, facilita una vista general y, al mismo tiempo, la unidad del proyecto. A través la descomposición y la reducción progresiva de la amplitud temática, se identificarán los elementos realmente manejables, para los que será más fácil presentar una serie de intervenciones posteriores (atribución de responsabilidades, asignación de recursos, estimación de tiempo y costes) que representan los pasos básicos en los que se articula el proceso de planificación y control.

² Con esta definición se hace referencia al valor semántico (marcado) de los metadatos de derechos de autor desarrollados por *Creative Commons*, y al hecho que no se acompañan en un fichero asociado al objeto original (por ejemplo, el artículo científico), sino que son insertados (incrustados) en el mismo fichero. Esto permite que licencia y contenido sean indisolubles.

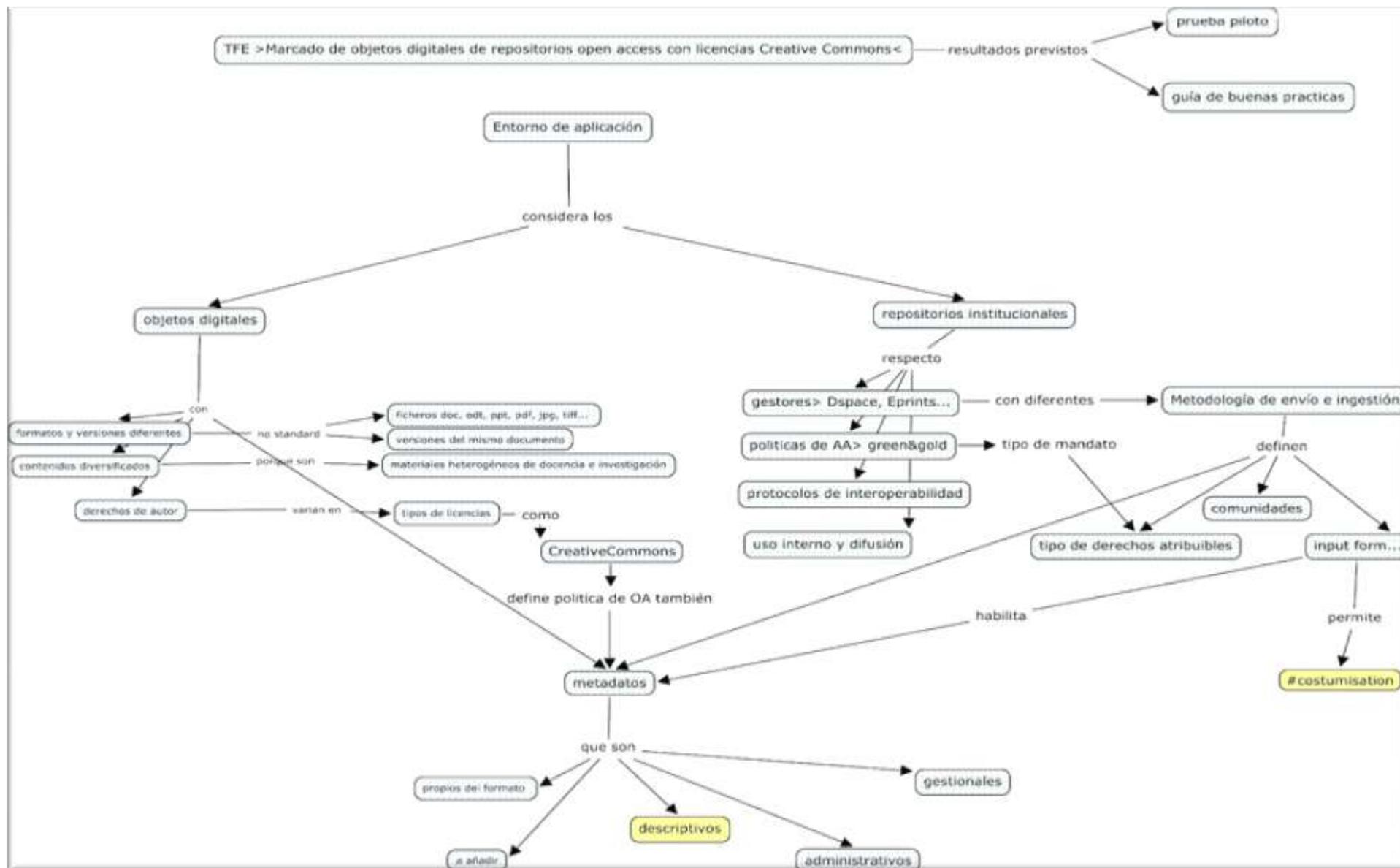


Figura 1: mapa conceptual

2.3. Interés del trabajo

La importancia del acceso abierto como canal de difusión, y de su extrema eficacia en el contexto académico, es un tema controvertido, que todavía no ha encontrado una comprobación científica generalizada. De hecho, el uso cotidiano del marco legal de publicación, reinterpretado a la luz de las licencias abiertas, todavía no es mayoritario, y resulta limitado en contextos culturales que sólo tímidamente van incluyendo medidas concretas para su aplicación. Afortunadamente, se dan casos de éxito que pueden servir de ejemplo y de motor impulsor en experiencias similares.

En este sentido, la introducción en 2011 de un mandato aprobado por el *Consell de Govern* de la Universitat de Barcelona, por el que se solicita al personal docente e investigador a depositar todas sus publicaciones en el RI de acceso abierto de la UB (el “Dipòsit Digital”), marca un punto de inflexión que obliga a reflexionar sobre una posible y mejor sistematización de la gestión de los derechos de autor, por parte de los responsables del Dipòsit.

En el mismo año, a nivel de gobierno central español, se aprobó la Ley de la Ciencia, la Tecnología y la Innovación³, cuyo artículo 37 define las modalidades de difusión en acceso abierto de la producción científica financiada (total o parcialmente) con fondos públicos: tiene que realizarse no más tarde de 12 meses después de su publicación oficial, y en repositorios reconocidos de investigación o en repositorios institucionales, ambos de acceso abierto. Estas nuevas condiciones ponen a las instituciones universitarias y de I+D en un contexto de nuevas alternativas a modelos históricamente consolidados de gestión de las publicaciones científicas.

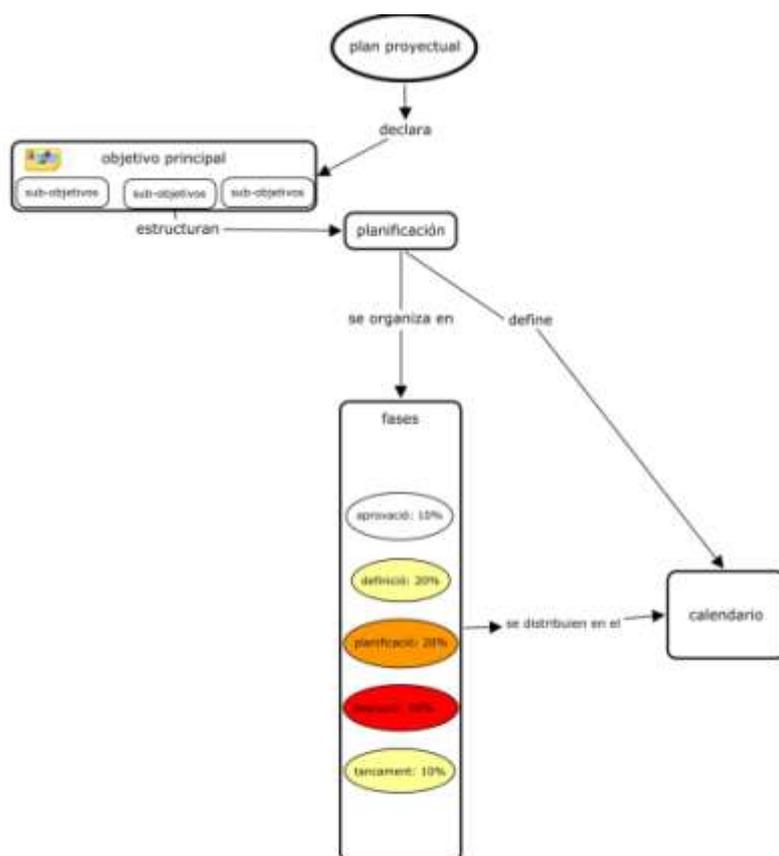
El interés en el tema de este trabajo reside en una necesidad real que surge a partir del contexto de circulación de las obras intelectuales en internet y, naturalmente, también en el caso de los RI. Cuando un usuario descarga un fichero, por ejemplo un PDF, como en el caso un artículo científico, se pierden, si no están incrustados correctamente, los metadatos de derechos de autor que se han generado para su depósito (en la revista o en el repositorio). La referencia a los derechos que se atribuyen a una obra, es fundamental para su posterior uso y difusión.

³ España. Ley 14/2011 de 1 de junio, de la Ciencia, la Tecnología y la Innovación (BOE 131, de 2 de Junio). Versión electrónica: <http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>

2.4. Grado de ejecución de proyecto

El plan de proyecto es completo en los detalles y elementos necesarios para la ejecución del mismo, reservándose hasta el cierre de las pruebas, el análisis que permita evaluar el cumplimiento de los objetivos específicos. La producción de una guía de buenas prácticas, anexa y parte integrante de este trabajo es la concreción del estudio realizado sobre la temática de marcado de objetos digitales en repositorios institucionales con licencias Creative Commons. Respecto a las tareas identificadas para la adopción de las buenas prácticas de ingestión en el repositorio de la Universitat de Barcelona (UB), en este trabajo se detallan tiempos, recursos humanos, y costes directos previstos; sin embargo, en la fecha de depósito administrativo en la Secretaria de la Facultat, no será posible incluir los resultados de las acciones previstas, debido a que se realizarán con posterioridad a dicha fecha.

El plan proyectual aquí presentado no incluye las últimas dos fases del mapa conceptual que se representa a continuación, por las razones que se han explicado anteriormente.



Anàlitzar el estat de la qüestió sobre la tecnologia i els protocols de incrustació de dades amb llicències Creative Commons en objectes digitals.
Poner de manifest el béll uso de la incrustació de dades amb llicències Creative Commons en repositoris open access.
Establecer una rutina de flujo de trabajo para un repositorio digital tipo que contemple de forma sistemática la incrustación de datos de la licencia en el objeto digital.
Realizar una prueba piloto para fundamentar empíricamente la guía de buenas prácticas.

Figura 2: mapa conceptual de un plan de proyecto

Tal como se esquematiza en este mapa conceptual, una vez definidos el objetivo principal y los objetivos secundarios, hay que planificar las actividades, subdivididas en varias tareas, en un conjunto de momentos ligados cronológicamente pero con distintas funciones. Los porcentajes representan el tiempo de dedicación previsto, respecto del total, en cada una de las fases proyectuales. La programación de las actividades y de los razonamientos que retroalimentan el proyecto, se encuentran descritos en el cronograma.

La propuesta de recuperación y reutilización de datos bibliográficos desde fuentes no homogéneas no es nueva y, de hecho, la implementación que aquí se plantea prevé el valor añadido de aumentar la visibilidad entre miembros de una misma comunidad. Disponer de fuentes de información y de herramientas de recuperación masiva de los metadatos de los trabajos publicados por parte de los investigadores y docentes de una universidad y depositarlos en el RI, es una práctica aconsejada por el bien del Acceso Abierto de la Ciencia: en el caso de la UB se parte de la opción de reutilizar como fuente el GREC, la base de datos de *Curricula Vitarum* del personal de la universidad en la que figura la noticia bibliográfica de las publicaciones). También se aconseja que los materiales de investigación sean depositados en los RI y no en repositorios externos a las instituciones, como en el caso de los artículos científicos que se publican en revistas con embargo, después del cual a menudo no se recupera la copia para su depósito en el propio repositorio. Las disposiciones que obligan a la publicación en OA son una solución a estos problemas, a pesar de que estos mandatos todavía no hayan sido impuestos por la mayoría de las partes interesadas de la investigación.

3. Definición y objetivos del proyecto

3.1. El “DIPÒSIT DIGITAL” UB” y la política OA de la universidad

El personal docente e investigador, así como el personal administrativo de soporte, en sus funciones producen materiales, también definidos como objetos digitales, que necesitan de conservación y difusión. En junio de 2011 se aprobó en *Consell de Govern* la declaración sobre la política de Acceso Abierto de la UB (UB, 2008), que proporciona reglas genéricas (quién, qué, cuando) para la ingestión en el repositorio digital. En la disposición se lee:

[...] los miembros de la comunidad académica de la Universitat de Barcelona depositarán una copia electrónica de cualquier publicación académica (artículos de revista, documentos de trabajo, ponencias, comunicaciones, documentos científico técnicos, libros, etc.) en el **RI** de la Universitat de Barcelona, inmediateamente después de su publicación, en un plazo no superior a los seis meses.[...]

Se identifica aquí la voluntad de abrir los contenidos académicos digitales al alcance de toda la sociedad, sin restricciones ni barreras para su acceso. Paralelamente, la UB fomenta una política de comunicación científica en acceso abierto recomendando:

[...] a la comunidad académica la publicación, cuando sea posible, **en revistas de acceso abierto**, y se compromete a facilitar este tipo de publicación, ya sea mediante la aportación de recursos o bien tratando con las editoriales.[...]

- En resumen, este documento define algunas tendencias generales que la UB asume del movimiento OA:
- existe un nuevo modelo de comunicación científica, y se reafirma con la “obligación” al depósito del material elaborado y creado en la comunidad académica
- es necesario el desarrollo de nuevos canales de transmisión de la información, que permitan también simplificar los procesos de acreditación/evaluación del trabajo académico
- ganar terreno al poder editorial y devolverlo a los “*commons*”, a la *res publica*⁴;

⁴ En este sentido, la [comunicación](#) con fecha 17 de abril de 2012 del *Faculty Advisory Council* a todas las Facultades de la Universidad de Harvard, resulta clarificadora: declara que la biblioteca ya no puede asumir el gasto actual (\$3.75M en 2011) en suscripciones a editoriales y distribuidores que mantienen políticas de precios inaceptables, e insta al cuerpo académico a seguir algunos consejos para resolver la insostenibilidad del actual sistema de difusión. Lo que se quiere subrayar de esta comunicación, es el énfasis puesto en la capacidad que

- insistir con la importancia de la conservación y preservación de los resultados de la investigación en cualquier formato (artículos, presentaciones a congresos, cursos, etc.), con la correcta atribución de todas las capas necesarias de metadatos.

Se infiere, por lo dicho, que sólo mediante una plena coparticipación y adhesión de la toda la comunidad científica, se podrán conseguir los beneficios que el movimiento de los contenidos abiertos aboga. El DD es el depósito digital creado para el fomento de las políticas citadas. Se trata de una aplicación *server-side* en código abierto, el Dspace, que permite un servicio de información científica en un ambiente digital e interoperable dedicado a la producción científico-académica de la propia institución. Contempla, por lo tanto, la reunión, el almacenamiento, la organización, la preservación y, sobre todo, la amplia diseminación de la información científica producida en su interior.

La presencia de repositorios digitales en acceso abierto es una realidad en constante incremento. Como se demuestra en la visualización de datos consultable a través de la URL: <http://maps.repository66.org/>, es posible observar la presencia de los repositorios institucionales a nivel mundial, con datos recolectados desde los directorios *Registry of OpenAccess Repository (ROAR)* y *The Directory of Open Access Repository (OpenDOAR)*, en un mapa de google. Resulta interesante la visualización de los repositorios españoles en una vista temporal desde principio de 2005, con dos instalaciones del programa Dspace presentes hasta mitad de 2006, para seguir el crecimiento en todo el territorio hasta final de mayo 2011.

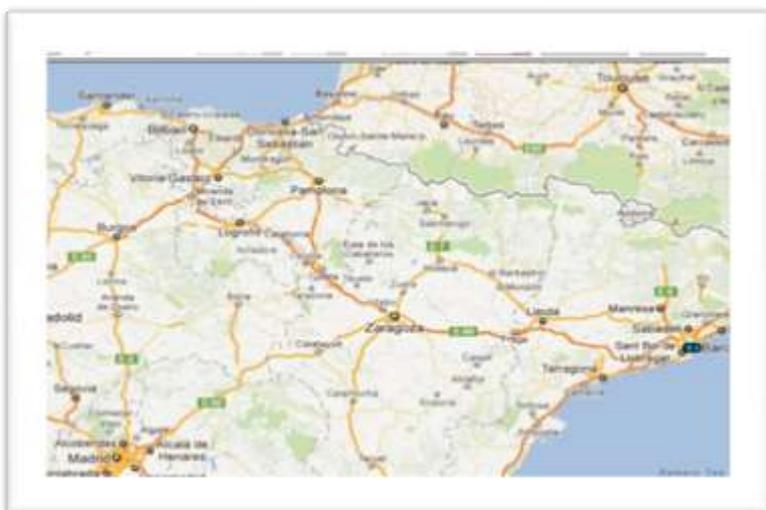


Figura 3: presencia de repositorios basados en Dspace a final de 2005

tienen, facultad y estudiantes por un lado, y biblioteca por el otro, de cambiar el sistema con una actitud explícita de adhesión a los criterios del movimiento open access.

<http://isites.harvard.edu/icb/icb.do?keyword=k77982&tabgroupid=icb.tabgroup143448> [consulta: 02.05.2012]

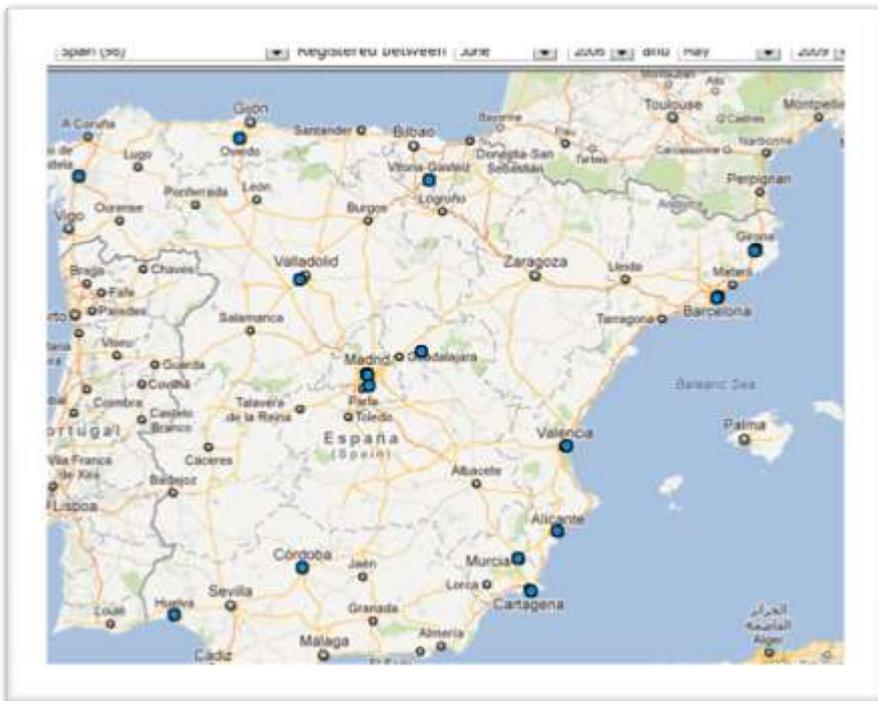


Figura 4: difusión repositorios en la primera mitad del 2006



Figura 5: difusión Dspace hasta mayo 2011

Para una visión actualizada, aunque no muy precisa, de los contenidos presentes en el DD se pueden consultar unas estadísticas desde el ROAR y confrontar el nivel de desarrollo del repositorio de la Universitat de Barcelona con otros repositorios del mundo.

En este gráfico se representa la evolución, desde su creación, de los contenidos depositados en el DD:



Figura 6: evolución anual del depósito de contenidos en el DD

Como han demostrado muchos estudios⁵, los beneficios de los contenidos abiertos en el campo académico son muchos: se incrementa el uso y el impacto de una publicación, así como la cantidad y la calidad de los trabajos de investigación, y se reducen los costes, facilitando así la transferencia directa de los descubrimientos científicos, reduciendo las distancias entre el mundo rico y el pobre y, finalmente, dando una mayor visibilidad a los resultados de la financiación pública. La reducción de costes se realiza también respecto a la posibilidad de reutilización de las obras abiertas⁶.

⁵Como ya se ha mencionado con anterioridad en campo editorial académico los efectos analizados por las investigaciones científicas del acceso abiertos resultan positivos, pero no se dispone todavía de unas evidencia empíricas compartidas por la comunidad entera.

⁶Este aspecto se tratará con mayor atención en relación a las distintas licencias de Creative Commons, ya que presenta problemas todavía no resueltos. En el mes de marzo y abril 2012, en la lista de discusión cc-licenses@lists.ibiblio.org han aparecido consideraciones muy interesantes respecto a las obras derivadas. El mensaje más relevante es de Christopher Allan Webber, en el cual se dan razones de las elecciones hechas para la última versión de las Licencias de Creative Commons, la versión 4.0, todavía en *draft* a la fecha actual, respecto a los temas de las obras derivadas.[Consulta:17.04.2012]

3.2. Especificaciones y definición del proyecto

Actualmente, para el envío de un artículo al repositorio, el autor entra con las credenciales del sistema informático académico y selecciona la colección donde depositar su documento⁷. Sin embargo el proceso actual y las modalidades de envío son objeto de revisión y de propuesta de cambio de algunos elementos, y por ello se proponen unas modificaciones que se concretan en la guía de buenas prácticas, anexa al final, y preparada con el fin de identificar líneas de mejora constantes en la gestión del marcado de derechos de autor, en el contexto digital de un repositorio abierto.

En este proyecto se propone utilizar los datos contenidos en el gestor de los *curricula* del personal docente e investigador de la UB (el sistema de gestión de la investigación, denominado GREC⁸) para crear información sobre objetos digitales, que son los artículos ya publicados, para su depósito en el DD. Ya ha sido desarrollada una pasarela digital para recolectar los datos relativos a las publicaciones que constan en el CV de cada profesor, y transferirlas mediante la pasarela al repositorio. De esta manera se consigue la optimización de la búsqueda sobre estos artículos, gracias a interoperabilidad con GREC garantizada por el software gestor del repositorio, el Dspace. Los metadatos sobre los artículos publicados, que se copian desde el GREC, no incluyen, en la mayoría de los casos, la indicación de los derechos de autor. Esta información tendrá que ser recuperada, para que desde el repositorio se puedan visualizar las opciones que caracterizan los derechos del material recuperado. De esta manera, por un lado se facilita la política de implementación de la UB de acceso abierto, mientras por otro lado, el correcto marcado de los objetos digitales, en un repositorio institucional abierto, facilita la circulación de los materiales en los distintos agentes automáticos presentes en la red (como arañas de buscadores, otros repositorios, etc.). Todo esto puede entenderse tanto como medida de interoperabilidad para sistemas informáticos (el ecosistema de la red evoluciona hacia una interconexión de nivel superior, el 3.0?), pero también como medida práctica de apoyo a la política de liberación del conocimiento mismo, a través del crecimiento de la masa crítica de las colecciones del repositorio.

⁷ Dspace permite el proceso de autenticación de los usuarios con un server LDAP. Esta función hace de conexión a un directorio de usuarios con permisos diferenciados y relativos a funciones.

⁸ La aplicación de gestión de los *curricula* del PDI se encuentra disponible en abierto en la dirección: <https://webgrec.ub.edu/> [Consulta 10.05.2012]

A través del uso de una semántica adecuadamente representativa (tanto al nivel de licencia aquí tratado, como de otras definiciones de metadatos), se consigue un entorno más adaptado a la ecología de la documentación científica hacia la que caminan las universidades. La integración en el entorno de agentes web inteligentes, capaces de entender y “manipular” eficazmente el objeto digital, se enmarca también en la necesidad de conformidad a unos estándares de intercambio de información, en particular las directivas DRIVER⁹ y OpenAire¹⁰. Con la convergencia hacia estos protocolos, se configura un entorno de intercambio, que permite la divulgación de los objetos alojados en un repositorio de manera eficiente, y que transforma cada repositorio en un nodo de una red compleja y comunicante.

La rutina de envío/adquisición de las obras destinadas a la difusión en el repositorio se estima que pueda beneficiarse de las ventajas ofrecidas por una perspectiva que busque aprovechar fuentes de datos ya existentes, pero todavía no relacionadas entre sí. La inclusión de los artículos presentes en los CV de cada investigador en el DD responde a una lógica de crecimiento del material depositado, pero también resulta ser una medida de simplificación del trabajo de difusión del personal docente e investigador (PDI). Es el mismo repositorio que, cumpliendo los estándares de interoperabilidad, genera la difusión de las obras de manera automática con la *recolección* de los repositorios que aplican los mismos estándares.

Toda la nueva información recogida desde el GREC se filtra con una plantilla para la normalización de los metadatos obtenidos, se adjuntan las políticas de derechos de autor y se empaquetan según la norma METS¹¹, para después enviarlos al DD.

⁹ Las directrices se centran en cinco cuestiones: colecciones, metadatos, implementación del protocolo OAI-PMH, prácticas recomendadas y esquemas, y semántica. DRIVER, *Digital Repository Infrastructure Vision for European Research*, es un proyecto realizado por un consorcio financiado por la Unión Europea que ha creado un marco de trabajo tecnológico y organizativo para implementar una capa paneuropea de datos, en el ámbito de la investigación y la educación superior. La página principal del consorcio se encuentra en <http://www.driver-support.eu/> [consulta 02.04.2012]

¹⁰ Las líneas directrices contenidas en la guía openAIRE se refieren a los métodos que los gestores de los repositorios tienen que implementar en las políticas de gestión de datos de conformidad con los requerimientos de acceso abierto de la Comisión Europea. Las normas se pueden consultar o descargar desde <http://www.openaire.eu/> [Consulta 23.05.12]

¹¹ El esquema METS es un estándar para codificar metadatos descriptivos, administrativos y estructurales en relación a los objetos de una biblioteca digital, expresado mediante el lenguaje de esquema XML del World Wide Web Consortium. La norma es mantenida en la Red para el Desarrollo y la Oficina de Normas MARC de la Biblioteca del Congreso de los Estados Unidos de América. La versión actualizada de la norma se puede recuperar en la dirección <http://www.loc.gov/standards/mets/>. [Consulta: 25.05.12]

3.3. Objetivos específicos

Los objetivos específicos del proyecto son los siguientes:

- Dibujo del flujo teórico para la identificación de las políticas de derechos de autor de las revistas donde se hayan publicado artículos académicos, posterior ingestión en el DD de la UB, con la inclusión del marcado de los derechos de autor con licencia CC, y depósito de ficheros a través del módulo SWORD¹² (*Simple Web-service Offering Repository Deposit*).
- Confección de una guía, anexa al final, que recopile las buenas prácticas para las políticas de marcado en los repositorios institucionales.
- Una prueba piloto en el DD, que se efectuará después de la aprobación y validación de las sugerencias contenidas en la guía anexa

¹² SWORD es una aplicación/protocolo diseñada para permitir el depósito masivo de trabajos académicos en los repositorios. Está habilitado para gestionar el ciclo de vida completo del depósito, incluyendo la actualización y supresión de los recursos depositados. La página de difusión de la herramienta es <http://swordapp.org/>, y es mantenida por uno de los creadores del mismo Dspace.

4. Análisis del problema

4.1 Análisis de la organización

El análisis DAFO es fundamental, ya que el proyecto se adapta a la realidad específica de la Universitat de Barcelona. En el [Plan director de la Universidad](#)¹³, se encuentra la siguiente esquematización en forma de matriz, que resume las fuerzas internas y las externas que impactan sobre la UB.

| | Debilidades | Fortalezas |
|------------------|--|--|
| Fuerzas internas | <ul style="list-style-type: none">• Envejecimiento y baja flexibilidad en la configuración de las plantillas de personal• Gran heterogeneidad docente y de investigación en necesidades, oferta y demanda• Número insuficiente de PDI activo en actividades de investigación y transferencia• Largo déficit presupuestario• Déficit en sistemas de información fiables y ágiles• Convivencia de las vías de relación funcional y laboral en PAS y PDI, y baja movilidad• Gran número de proyectos de inversión iniciados | <ul style="list-style-type: none">• Elevado número de demanda de acceso de estudiantes y profesores• Grupo UB: PCB, FBG y otras fundaciones• Tradición, calidad, prestigio y alta competitividad de las enseñanzas y de la investigación• Oferta de grados, másteres y doctorados adaptada al EEES y aprobada en éste• Buena posición en las clasificaciones internacionales• Situación geográfica• Capacidad de mejora en gestión, investigación y docencia• Existencia de espacios adecuados para la creación de proyectos de base tecnológica• Integración de los campus en la ciudad |
| Fuerzas externas | Amenazas | Oportunidades |
| | <ul style="list-style-type: none">• Disminución de la población en edad académica y aumento de la competencia entre universidades• Sistema de financiación deficiente, que no incluye la actividad de investigación ni el mantenimiento ni las particularidades de la UB (antigüedad, dispersión, etc.) | <ul style="list-style-type: none">• Demandas crecientes de la formación de la sociedad, con una oferta heterogénea y generalista• Acceso a figuras y herramientas para potenciar la incorporación de investigadores (ICREA, CIBER)• Posibilidad de financiación y otros ingresos adicionales |

Con las referencias al entorno más inmediato y a los factores internos, que afectan al funcionamiento de la institución, se identifican las características endógenas que repercuten

¹³ <http://www.ub.edu/pladirector/es/index.html> : El espíritu de servicio de la Universitat de Barcelona y el ejercicio de responsabilidad y transparencia informativa, han llevado a elaborar un Plan director que recoge el programa electoral, la experiencia de los primeros meses de gobierno, y el trabajo realizado de evaluación y reflexión internas.

tanto en la consecución de los objetivos de este proyecto, como en el desarrollo de sus actividades. Entre los factores externos, las variables exógenas como el entorno cultural, socioeconómico y político, condicionan las posibilidades y repercuten en la eficacia de las estrategias generales.

Según lo indicado en la matriz DAFO, se identifican las acciones estratégicas a seguir:

- ***Mejora de indicadores docentes***
- ***Altos estándares homogéneos de investigación***
- ***Autonomía y suficiencia universitarias***
- ***Modernización de sistemas de gestión y funcionamiento***
- ***Proyección ciudadana e internacional***

El proyecto propuesto se integra en la línea estratégica de actualización/modernización de los sistemas de gestión y funcionamiento y, en particular, quiere influir en la organización de una porción de la estructura informática dedicada a la conservación y difusión de los productos (*outputs*) de la actividad universitaria, es decir, en la estructura lógica del “Dipòsit Digital”. La misma propuesta se estima que pueda obtener algunos efectos también en otras áreas estratégicas, como en la proyección a nivel social e internacional, o disponer de otras fuentes para los indicadores docentes.

4.2 Naturaleza de los repositorios digitales

A finales de abril de 2012, la UB (CRAI-UB, 2011) participa como colaboradora¹⁴ en los siguientes proyectos de colecciones digitales o repositorios cooperativos:



Figura 7: repositorios con participación de la UB



Fig. 8: logo DD

Sin embargo, en este proyecto sólo se considera el repositorio institucional, el DD, que recoge las publicaciones digitales en acceso abierto derivadas de la actividad del personal docente e investigador de la comunidad universitaria UB, con el objetivo de preservarlas y difundirlas.

Los documentos publicados están organizados por colecciones que pertenecen a cinco grandes categorías: la docente, la institucional, la de investigación, los trabajos de los estudiantes y la mediateca.

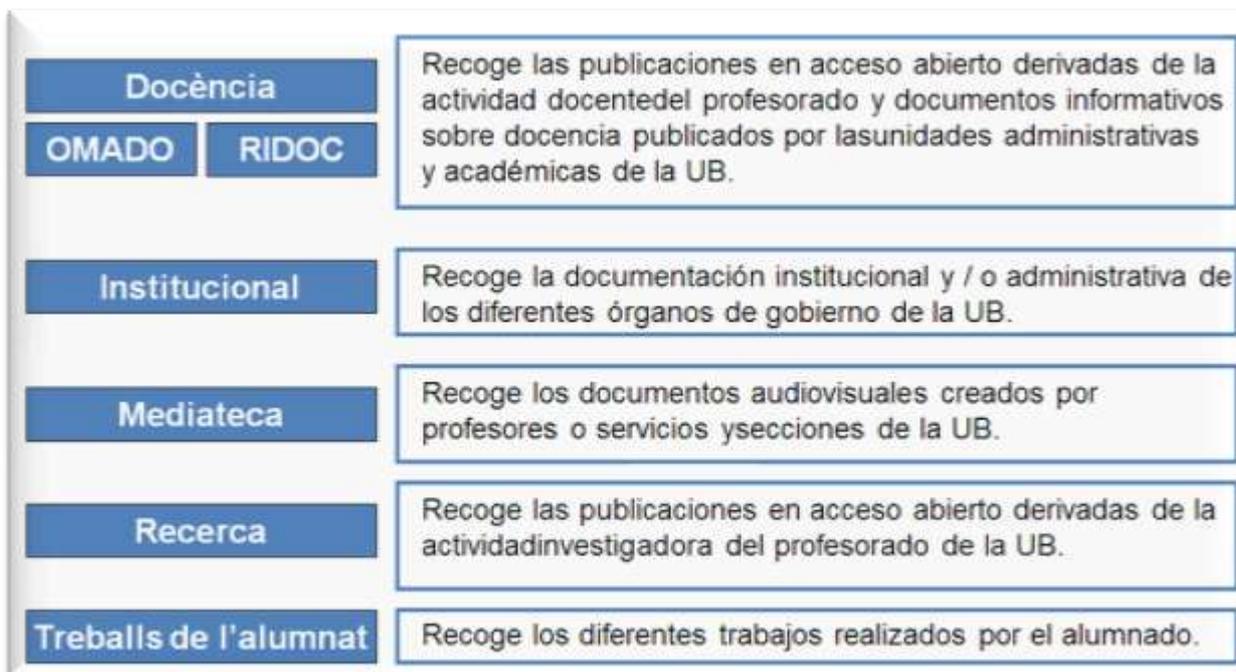


Figura 9: organización colecciones "Dipòsit Digital" UB

El Dipòsit Digital se apoya en Dspace, un programa en *open source*, del cual se presenta un esquema del funcionamiento:

¹⁴ Páginas de referencia para el tema: <http://www.bib.ub.edu/recursos-informacio/e-diposits/> [consulta: 04/04/2012]

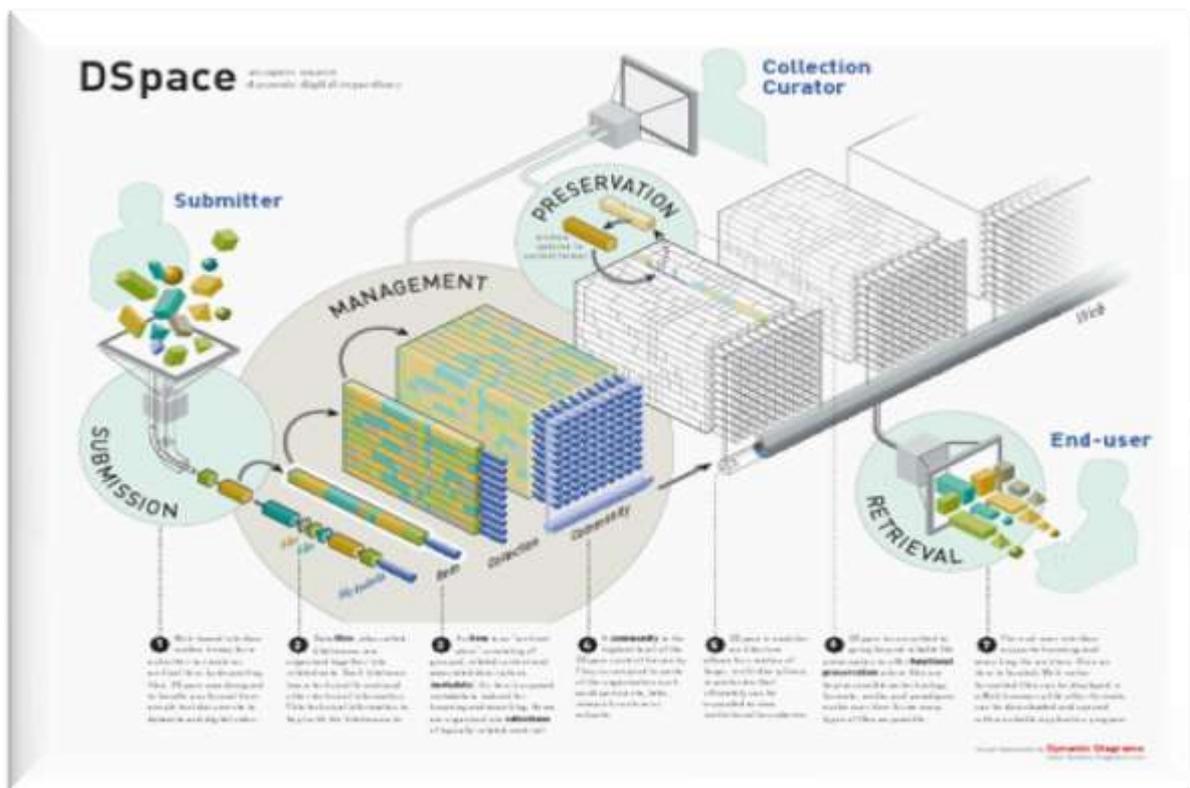


Figura 10: Organización proceso de Dspace. Fuente <http://www.dspace.org/images/stories/dspace-diagram.pdf>

En este diagrama se representan las fases que el software gestiona: una primera fase (*submission*), que consiste en el envío de objetos por parte del depositario; una segunda fase (*management*), de gestión de los objetos; una tercera fase en la que el CRAI, como administrador de las colecciones, tiene que supervisar el proceso global y encargarse de la preservación, que consiste en la actividad de monitorización de la compatibilidad de los formatos de los ficheros; y una última fase de recuperación de los objetos para el uso por parte de un usuario final.

Dspace es una plataforma que permite capturar ítems en cualquier formato, como texto, video, audio y datos, y distribuirlos en la red. La plataforma indexa los objetos depositados, de modo que los usuarios puedan buscar y recuperar los ítems. Dspace es utilizado típicamente como repositorio institucional y, como tal, tiene tres roles principales:

- Facilitar la captura y el envío de materiales científico-educativos, incluyendo metadatos sobre los mismos.
- Facilitar un acceso simple a los materiales, sea a través de listas o de búsqueda.
- Facilitar la preservación a largo plazo.

A continuación se detallan las modalidades de envío relativas al proceso que el objeto digital pasa para ser depositado. Por supuesto, sólo los sujetos autorizados para el depósito en el repositorio de la UB pueden acceder a este proceso.

El flujo de acciones (*workflow*) previstas se resume como sigue:

1. Preparar el documento a depositar en uno de los formatos admitidos y decidir de antemano qué tipo de licencia se le quiere atribuir.

- Entrar en Dspace con las credenciales de acceso a la infraestructura informática de la UB y seleccionar la colección en la que se quiere depositar (si se han recibido credenciales para distintas colecciones).
- Empezar el proceso de envío propiamente dicho.

Este proceso consta de:

- Descripción del objeto (3 pasos en los cuales se crean los metadatos). (Figuras 11-13)
- Envío y verificación envío del documento. (Figuras 14 y 15)
- Selección de la licencia Creative Commons (2 pasos). (Figuras 16-18)
- Fin del proceso (Figura 19)

En las capturas presentadas a continuación se puede ver en detalle el proceso.

Figura 11: paso 1 del envío de un objeto a Dspace (v. default)

Figura 12: paso 2 del envío de un objeto a Dspace (v. de fabrica)

Descriu Descriu Descriu **Carrega 3** Verifica Llicència Llicència Finalitza

Envia: descripció del document

Si us plau, ompliu la informació que es demana a continuació sobre el document. [Més ajuda](#)

Enter appropriate subject keywords or phrases below:

Subject Keywords test cc
 [Obtenir més](#)

Enter the abstract of the item below:

Abstract

Enter the names of any abstract and/or funding codes in the box below:

Sponsors

Enter any other description or comments in this box:

Description

[Anterior](#) [Següent >](#) [Cancel·la/Guarda](#)

Figura 13: paso 3 del envío de un objeto a Dspace (v. default)

Descriu Descriu Descriu **Carrega 3** Verifica Llicència Llicència Finalitza

Lliurament: fitxer enviat

Aquestes són les dades del fitxer que heu enviat. Si us plau, comproveu-les abans de continuar i fer el pas següent. [Més ajuda](#)

| Fitxer | Dimensions | Format del fitxer |
|--|--------------|--|
| FORMULARI TRIA CENTRES CURS 2012_2013 UR.doc | 59.904 bytes | Microsoft Word (conegut) |

[Cliqueu aquí si el format no és el correcte](#)
[Cliqueu aquí si el fitxer no és el correcte](#)

Per comprovar si el fitxer s'ha carregat correctament:

- Cliqueu els noms de fitxer que es mostren més amunt. El fitxer s'obrirà en una pantalla nova, de manera que en pugueu comprovar el contingut.
- El sistema calcula un coeficient de verificació que podeu verificar. [Cliqueu aquí per a més informació](#).
[Mostrar els coeficients de verificació](#)

[Anterior](#) [Següent >](#) [Cancel·la/Guarda](#)

Figura 14: paso 4 del envío de un objeto a Dspace (v. de fabrica)

Descriu Descriu Descriu **Carrega 3** Verifica Llicència Llicència Finalitza

Lliura: verificació el lliurament

Encara no, però quasi!

Si us plau, reviseu les dades del document que acabeu d'enviar. Si hi ha alguna dada incorrecta, torneu enrere i corregiu-la utilitzant els botons situats al costat de l'error, o bé cliqueu la barra d'estat de la part superior de la pàgina. [Més ajuda](#)

Si tot és correcte, si us plau, cliqueu el botó "Següent" del final de la pàgina.

S'obrirà una finestra perquè pugueu verificar els fitxers que heu enviat.

| | |
|--|--|
| El document té més d'un títol: No | Corregir alguna d'aquestes dades |
| Document publicat amb anterioritat: No | Corregir alguna d'aquestes dades |
| El document té més d'un fitxer: No | |
| Authors se, sgs | |
| Title prova metadati | |
| Series/Report No. 0000-0001; | |
| Identifiers Cap | Corregir alguna d'aquestes dades |
| Type Article | |
| Language (Other) | |
| Subject Keywords test cc | |
| Abstract PROVA21 | Corregir alguna d'aquestes dades |
| Sponsors MOI | |
| Description FILE FILE SENZA LIC | |
| Fitxer enviat: FORMULARI TRIA CENTRES CURS 2012_2013 UR.doc - Microsoft Word (Conegut) | Envia un fitxer diferent |

[Anterior](#) [Següent >](#) [Cancel·la/Guarda](#)

Figura 15: paso 5 del envío de un objeto a Dspace (v. default)

[Descriu](#) [Descriu](#) [Descriu](#) [Carrega fitxer](#) [Verifica](#) **Llicència** [Llicència](#) [Finalitza](#)

Lliurament: utilitzant una llicència Creative Commons

Per atorgar al vostre document una llicència Creative Commons, seguïu les instruccions següents. Podreu revisar la tria. Utilitzeu l'enllaç 'Trieu una llicència' per afegir la llicència. Si voleu ometre la llicència Creative Commons, premeu el botó 'Omet Creative Commons', que apareix al final de la pàgina.



Las licencias de Creative Commons le ayudan a compartir su obra manteniendo sus derechos de autor. La gente puede copiar y distribuir su obra siempre que le [reconozcan su autoría](#) -- y sólo en las condiciones que usted especifique aquí. Esta página le ayuda a escoger estas condiciones.

Si desea compartir sin ninguna condición una obra que has creado, escoge [CC0](#). Si estás compartiendo una obra que no está cubierta por derechos de autor o en la que han expirado esos derechos, escoge la [Etiqueta de Dominio Público](#).

¿Permitir el uso comercial de su obra? ([más información](#))

Sí
 No

¿Permitir modificaciones de su obra? ([más información](#))

Sí
 Sí, siempre que se comparta de la misma manera ([más información](#))
 No

[Escoja una licencia](#)

Aviso: Para licenciar una obra, usted debe ser el titular de los derechos o tener una autorización expresa del titular para hacerlo.

Creative Commons no ofrece ninguna asesoría o servicio legal. Ofrecemos documentos legales prediseñados; el resto depende de usted.

[< Anterior](#) [Omet Creative Commons >](#) [Cancelar la Guarda](#)

Figura 16: fase 6: preguntas para licenciar con CC

[Descriu](#) [Descriu](#) [Descriu](#) [Carrega fitxer](#) [Verifica](#) **Llicència** [Llicència](#) [Finalitza](#)

Lliurament: utilitzant una llicència Creative Commons

Per atorgar al vostre document una llicència Creative Commons, seguïu les instruccions següents. Podreu revisar la tria. Utilitzeu l'enllaç 'Trieu una llicència' per afegir la llicència. Si voleu ometre la llicència Creative Commons, premeu el botó 'Omet Creative Commons', que apareix al final de la pàgina.

provided by [Creative Commons](#)

Ha seleccionado la **Licencia Atribución 3.0 Unported**
 Esta licencia está permanentemente ubicada en
<http://creativecommons.org/licenses/by/3.0/>.

Ahora puede [proceder](#)

Figura 17: confirmación licencia elegida

[Descriu](#) [Descriu](#) [Descriu](#) [Carrega fitxer](#) [Verifica](#) [Licència](#) [Licència](#) [Finalitza](#)

Heu de fer un darrer pas: perquè el Dipòsit digital pugui reproduir i distribuir a escala mundial el document lliurat, cal que accepteu els termes següents. Preneu-vos un moment per llegir la llicència, i cliqueu un dels botons situats al final de la pàgina. En pitjar el botó "Accepta la llicència", estareu indicant que accepteu els termes de la llicència. [Més ajuda](#)

No acceptar la llicència no farà que s'esborri el lliurament. El document romandrà a la pàgina "El meu dipòsit". Podeu, o bé eliminar el lliurament, o bé acceptar la llicència més tard, un cop s'hagin resolt els dubtes que pugueu tenir.

NOTE: PLACE YOUR OWN LICENSE HERE
This sample license is provided for informational purposes only.

NON-EXCLUSIVE DISTRIBUTION LICENSE

By signing and submitting this license, you (the author(s) or copyright owner) grants to DSpace University (DSU) the non-exclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that DSU may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that DSU may keep more than one copy of this submission for purposes of security, back-up and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant DSU the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN DSU, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

DSU will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Figura 18: texto de la licencia de fabrica en Dspace para publicar en el depósito, independiente de las CC (ver. de fábrica)

[Descriu](#) [Descriu](#) [Descriu](#) [Carrega fitxer](#) [Verifica](#) [Licència](#) [Licència](#) [Finalitza](#)

Lliurament completat!

El document entrarà ara dins el procés de treball establert per a la col·lecció a la qual l'heu lliurat. Rebreu un avís per correu electrònic quan el document hagi entrat a formar part de la col·lecció, o bé si per alguna raó hi ha problemes amb el lliurament. També podeu comprovar l'estat del vostre lliurament anant a la pàgina "El meu dipòsit".

[Vés al meu dipòsit](#)

[Comunitats i col·leccions](#)

Figura 19: fin del proceso

4.2.1. Tipología de licencias en la gestión de objetos digitales

A partir del momento en que empezaron a circular por la red contenidos digitales, la cuestión de cómo hacer frente al tratamiento del material protegido por derechos de autor en el mundo digital se ha convertido en un tema clave del mundo cultural, económico y jurídico contemporáneo. Ante semejante reto se han encontrado diversas soluciones.

Por una parte, la industria ha creado un modelo de marcado que permite un control total del objeto que el usuario final utiliza: es el *Digital Right Management* (DRM). A través de esta tecnología de identificación de Copyright, los editores no permiten al usuario final ninguna manipulación del objeto digital (en concreto, se ha utilizado para proteger los *e-books* distribuidos por los mayores representantes mundiales de publicaciones electrónicas, como por ejemplo *Amazon* con *Kindle* o *Apple* desde el *Appstore*), y además se mantiene un control después de la venta, limitando al comprador las descargas del objeto y la instalación en distintos aparatos reproductores. Esta política industrial de protección de las obras, con el DRM, ha generado un movimiento de protesta que pide su eliminación¹⁵.

Sin embargo, más allá del sector de la industria editorial plenamente comercial, en otros ámbitos de difusión de material digital se han investigado soluciones para la representación de los derechos de autor. El caso del software -entendido como código fuente - libre (*Open Source*), es un hito en la historia del movimiento *Open Access*. El desarrollo en abierto de cualquier aplicativo se ha concretado al final de los años 80 con la creación de la primera *GNU General Public License* (GPL)¹⁶, creada por Richard Stallman. Esta licencia pretendía, y así sigue, poner el código escrito a disposición gratuita y libre de cualquier otro programador, el cual a su vez tiene que aplicar los mismos derechos a la obra derivada. Esta solución, con la cual se garantiza mantener un producto en continuo desarrollo bajo la misma licencia inicial, ha sido descrita y estudiada desde muchas perspectivas, y en general se le atribuyen muchos logros en la creación de un substrato filosófico y comercial que sustente el acceso libre a los recursos de la red.

La historia de la evolución de los métodos de marcado de objetos digitales es apasionante, dado que refleja posiciones ideológicas de grupos sociales heterogéneos, y enfrentados en el

¹⁵ <http://www.defectivebydesign.org/> [consulta 02/04/2012]

¹⁶ la GPL se creó en 1989 (http://en.wikipedia.org/wiki/GNU_General_Public_License#History)

ámbito de la connotación legal de una obra del intelecto humano. Dado que *Creative Commons* está experimentando desde hace años el mercado de los objetos con distintos sistemas es necesario dar una descripción de los procesos de mercado, desde la cual se contextualizan las debilidades actuales y los desafíos¹⁷ que este sistema de protección de derechos conlleva, hoy y en un futuro próximo, en relación a los repositorios institucionales.

Los movimientos “*open*”, y entre ellos el de *Open Access*, han sido los promotores de una larga discusión sobre los bienes comunes (los “*Commons*” en inglés), que ha llevado a la introducción de licencias abiertas integradas en los distintos sistemas legales. En la actualidad, la difusión y adopción de las licencias CC es muy extendida y sigue un andamio en continuo crecimiento.

El ámbito de aplicación de estas licencias alcanza los contenidos siguientes: ficheros textuales flotantes, audio, video, imágenes y, por supuesto, sitios enteros en red. Las características específicas de las licencias se explican en el apartado 4.2.3. Pueden ser tanto para uso privado (individual) como para uso público y comercial.

¹⁷ Cómo son la variedad original de los formatos de los objetos a marcar, los problemas de uso de las licencias CC-NC o las dificultades en la gestión de los derechos colectivos, etc.

4.2.2. Historia reciente de licencias abiertas en la UB

En el contexto de la UB, la temática sobre licencias abiertas tiene unas raíces bastante profundas. Sin embargo, aquí se presenta sólo una esquemática panorámica con la selección de algunos documentos oficiales. En los últimos años, la Universidad de Barcelona ha visto la progresiva afirmación de los paradigmas del acceso abierto. En la tabla siguiente se resumen los momentos y documentos más importantes. Hay que mencionar también que la UB ha sido la primera sede oficial de difusión e investigación de las licencias Creative Commons en suelo español, en el año 2004.

| | |
|-----------------|---|
| Febrero 2003 | Primer contacto con Creative Commons para emular la licencia CC – MIT (Massachusetts institute of technology) desde la UB |
| Diciembre 2003 | La UB es reconocida como institución española afiliada de Creative Commons |
| Octubre 2004 | La UB formula una licencia CC-UB e implementa su uso al lado de las otras CC |
| Febrero 2008 | La UB y la libre difusión del conocimiento |
| Marzo 2008 | Recommendations from the EUA Working Group on Open Access adopted by the EUA Council on 26th of March 2008 (University of Barcelona, Spain) |
| Junio 2009 | Consejo interuniversitario de Cataluña |
| Octubre 2009 | Firma de la Declaración de Berlín |
| Febrero 2011 | Real Decreto de las Enseñanzas de Doctorado |
| Junio 2011 | Ley de la Ciencia, tecnología e innovación |
| Junio 2011 | Política de acceso abierto en la Universidad de Barcelona |
| Septiembre 2011 | LERU: Ruta hacia el acceso abierto con ejemplos de la UB¹⁸ |

Es de gran importancia la aprobación de la política de acceso abierto en junio de 2011, que resulta una acción de sustento para el crecimiento de los contenidos en el repositorio institucional. Los acuerdos adoptados prevén que el personal docente e investigador siempre considere la oportunidad de publicar en el “Dipòsit Digital de la UB” (enviando y conservando sus trabajos), para favorecer el acceso incondicionado a los resultados de la investigación y la docencia procedentes del ámbito académico.

¹⁸ http://www.leru.org/files/publications/LERU_AP8_Open_Access.pdf [consulta 2 Abril 2012]: en este documento se citan las políticas de acceso abierto de la UB como ejemplo de buenas prácticas.

4.2.3. Creative Commons

Creative Commons (CC) es una organización global que cuenta con una red de instituciones afiliadas en muchos países, que se han encargado de difundir y adaptar la filosofía de las licencias creadas en los Estados Unidos de América. En cada país se ha hecho un trabajo de adaptación al marco legal específico para que las licencias CC pudieran representar una alternativa totalmente legal al tradicional “© todo reservado”.

La oferta de licencias de *Creative Commons* se articula sobre cuatro condiciones que se pueden combinar entre ellas. Se presentan en la siguiente imagen los elementos básicos:

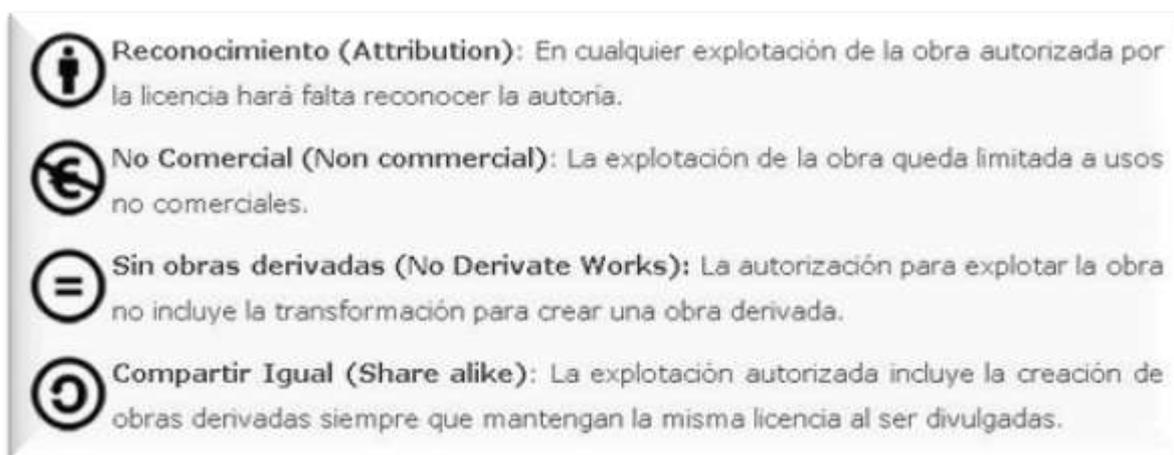


Figura 20: condiciones de las licencias CC

Con estos conceptos básicos se pueden componer 6 licencias distintas:

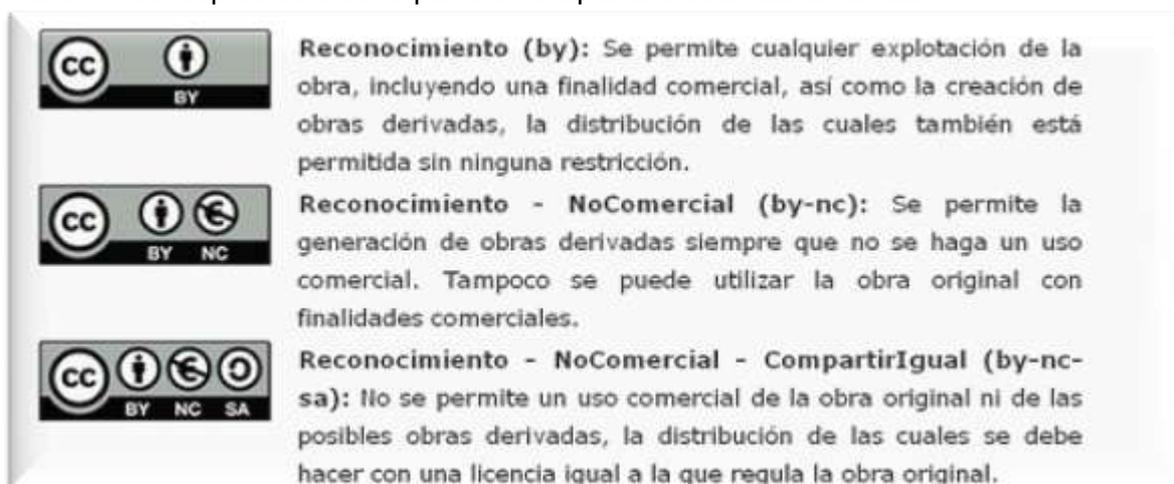


Figura 21: licencias Creative Commons



Reconocimiento - NoComercial - SinObraDerivada (by-nc-nd): No se permite un uso comercial de la obra original ni la generación de obras derivadas.



Reconocimiento - CompartirIgual (by-sa): Se permite el uso comercial de la obra y de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.



Reconocimiento - SinObraDerivada (by-nd): Se permite el uso comercial de la obra pero no la generación de obras derivadas.

Figura 22: licencias Creative Commons

4.3. Métodos de representación de la relación entre los derechos y la licencia



Fig. 23: capas de marcado

La imagen adyacente representa los niveles de marcado en los objetos digitales con licencias *Creative Commons*. Se trata de 3 capas. El nivel más alto es el marcado visible por los usuarios: se incluye en el documento visualizado el logo de la licencia, el autor (Figura 24), y un link al texto del *Commons Deed*.

Commons Deed es una versión simplificada del texto legal tradicional (Figura 25), que constituye el segundo nivel. La última capa del diseño de la licencia juega un papel muy importante en la creación, la copia, el descubrimiento y distribución de las obras. Con este nivel se proporciona una "lectura mecánica" de la versión de la licencia - un resumen de las libertades fundamentales y obligaciones - en un formato que los sistemas de software, motores de búsqueda, y otro tipo de tecnología puedan entender (Figura 26). *Creative Commons* ha desarrollado una forma estandarizada para describir las licencias inteligibles por los agentes en la red. Este lenguaje se llama *CC Rights Expression Language (CC REL)*. A continuación se presenta un ejemplo práctico de cada capa (Figura 23).



Figura 24: primera capa de marcado de derechos de autor. Logo, autor, y licencia atribuida



Figura 25: el *Commons Deed*, la versión legal simplificada de una licencia. La CC BY 3.0¹⁹

¹⁹ <http://wiki.creativecommons.org/Marking> [consulta: 10.06.2012]

```
<div xmlns:dc="http://purl.org/dc/terms/" xmlns:cc="http://creativecommons.org/#ns"
about="http://wiki.creativecommons.org/Marking/Creators"><span
property="dct:title">Marking/Creators - CC Wiki</span><a rel="cc:attributionName" href=""
target="_blank">this site</a><a rel="license" target="_blank"
href="http://creativecommons.org/licenses/by/3.0/"> Attribution 3.0 Unported / CC BY
3.0</a></div>
```

Figura 26: el código del tercer nivel de marcado de un objeto digital

Para entender adecuadamente este código son necesarios algunos antecedentes. Desde su comienzo, la corporación americana sin ánimo de lucro *Creative Commons* ha elaborado distintas modalidades de marcado que se utilizan para garantizar la presencia de los elementos distintivos de una licencia de derechos de autor. Esta necesidad se refleja en distintos niveles, pues hay un usuario humano que necesita de unas referencias a la licencia (el texto de la licencia, el logo,...), visibles en el objeto utilizado, y también hay un agente automático que tiene que manejar el objeto reconociendo los mismos derechos representados en la licencia textual. En el sitio web de *Creative Commons*²⁰ es posible encontrar toda la historia y el desarrollo de las distintas soluciones técnicas realizadas, y están disponibles para su libre modificación.

Una rápida panorámica sobre estas técnicas permite evidenciar la variedad y la heterogeneidad de situaciones que se han estudiado para la mejor representación de las distintas capas de la licencia.

En la *home page* de CC se lee: *“Licensing a work is as simple as selecting which of the six licenses best meets your goals, and then marking your work in some way so that others know that you have chosen to release the work under the terms of that license”*. Y de hecho, hoy en realidad es muy simple seleccionar entre las licencias disponibles, y obtener el logo, el link al texto y los metadatos de la licencia que se ha elegido; lo que resulta menos amigable es la fase del marcado propiamente dicha, que presenta una gran heterogeneidad de soluciones con respecto al objeto a marcar. La estructura y actualización de muchas referencias en el mismo sitio de CC sobre técnicas de marcado causan cierta confusión por la coexistencia de versiones distintas (2009 y 2012) en el mismo sitio web.

Para el entorno profesional, *Creative Commons* ha creado el lenguaje CC REL, anteriormente mencionado, del cual se ha publicado una nueva recomendación que anula y substituye a la anterior (Dulong & De Martin, 2012). Desde 2004, Creative Commons ha trabajado con el

²⁰ <http://creativecommons.org/> [consulta: 15.05.12]

W3C²¹ (*World Wide Web Consortium*) para crear métodos de incrustación (*embedding*) de código RDF (*Resource*

Description Framework) en documentos HTML, y con estos procesos se han ido creando nuevos modelos. Conforme a estos, CC ya no recomienda el uso de RDF/XML en los comentarios HTML para la especificación de la licencia, sino la adopción del lenguaje CC REL.

A continuación se describe la nueva versión del lenguaje. El modelo **abstracto de CC REL** distingue dos clases de propiedades: las **propiedades de trabajo** (*work properties*), que describen aspectos específicos del trabajo, incluyendo los derechos de autor bajo los cuales la obra es distribuida; y las **propiedades de licencia**, que describen la licencia en sí. Los editores se ven afectados sólo en el uso de las propiedades del trabajo, puesto que las propiedades de licencia son utilizadas por *Creative Commons* para definir las especificaciones de las licencias que ella ofrece. Un editor que quiera utilizar las licencias CC tiene que declarar al menos un triple RDF que especifique el valor de las propiedades de la licencia de la obra. Por ejemplo:

<<http://johndoe.com/blog>> **xhtml:license** <<http://creativecommons.org/licenses/by/3.0/>>.

Pero CC sugiere también añadir más información sobre el trabajo, como el título, el nombre y la URL, para atribuir la obra y el tipo de documento. Así, hay que declarar:

- Título: <<http://johndoe.com/blog/>> dc:title "doe's blog"
- Nombre del autor: <<http://johndoe.com/blog/>> cc:attributionName "John Doe"
- URL de atribución²²: <<http://johndoe.com/blog/>>cc:attributionURL <<http://johndoe.com/>>
- Tipo de documento: <<http://johndoe.com/blog/>> dc:type dcmitype:Text.

El editor dispone además de otras dos propiedades del trabajo: **dc:source**, que indica la dirección URL de obra original de un trabajo modificado, y **cc:morePermissions**, que indica una URL donde se encuentran permisos adicionales además de los contemplados en la licencia CC. Éstas, a día de hoy, son las propiedades que *Creative Commons* sugiere utilizar en el marcado de objetos digitales. Otras propiedades podrán ser añadidas en el futuro, gracias al hecho que CC REL hereda la extensibilidad de RDF.

A diferencia de la recomendación inicial que preveía la gestión de las propiedades de licencia por parte del editor, en el nuevo modelo CC REL estas propiedades son competencia exclusiva de *Creative Commons*, evitando así datos redundantes y con posibles errores. La búsqueda de las propiedades de un trabajo se realiza mediante un formato como el de una dirección web,

²¹ <http://www.w3.org/>

²² La propiedad cc:attributionURL es la manera preferida para proveer este URL en un formato legible por una máquina.

con la ventaja de que la herramienta de interpretación de la licencia reconoce la propiedad *xhtml:license*, y sigue la dirección donde encontrará la página del *CC deeds*. En esta misma página están incrustadas las propiedades de las licencias, expresadas en la sintaxis CC recomendada (la RDFa). Las propiedades de licencia definidas en CC REL son: ***cc:permits***, ***cc:prohibits***, ***cc:requires***, ***cc:jurisdiction***, ***cc:deprecatesOn*** y ***cc:legalCode***. Estas propiedades definidas en las licencias CC existentes no admiten modificaciones. Los diferentes atributos de *cc:permits* definen los posibles usos de una obra, con *cc:Reproduction*, de copia en distintos formatos; con *cc:Distribution*, de redistribución; y con *cc:DerivateWorks*, de derivación de otros trabajos. A su vez, para la propiedad *cc:prohibits*, el valor del atributo *cc:CommercialUse* define los posibles usos comerciales de una obra. Los atributos contemplados en *cc:requires* son: *cc:Notice*, *cc:Attribution*, *cc:ShareAlike* y *cc:SourceCode*.

Haciendo una revisión de las herramientas de marcado disponibles se encuentra un primer ejemplo *on-line*²³, en la web de *Creative Commons*. Es un formulario que permite la elección de la licencia y la generación de los elementos para su aplicación en un objeto digital. No se trata de un registro, en el sentido que CC sólo pone a disposición la herramienta de selección, sin registrar los datos que con ella se generen.

- Para los usuarios individuales es posible utilizar el *License choosing tool* vía web, que se presenta en la imagen siguiente:

²³ <http://creativecommons.org/choose/> [Consulta 10.05.2012]

¿Quiere permitir usos comerciales de su obra?

Sí [?]

No [?]

¿Quiere permitir modificaciones de su obra?

Sí [?]

Sí, mientras se comparta de la misma manera [?]

No [?]

Jurisdicción de su licencia [?]

Internacional ▼

Additional Information

The additional fields are optional, but will be embedded in the HTML generated for your license. This allows users of your work to determine how to attribute it or where to go for more information about the work. [?]

¿Cuál es el formato de su obra? Otro ▼

Título de la obra [?]

Attribute work to name [?]

Attribute work to URL [?]

URL fuente de la obra [?]

More permissions URL [?]

[Escoja una licencia](#)

Figura 27: license on-line choosing tool

El código html para el objeto a marcar se obtiene en función:

- del objeto digital a marcar en la casilla “¿Cuál es el formato de su obra?”, que puede ser un fichero audio, video, imagen, texto o un conjunto de datos,
- de las respuestas a las preguntas sobre qué tipo de derechos se quieren conceder a la obra,
- de los datos sobre el autor (“Attribute work to name”),
- de la URL de atribución del trabajo (“Attribute work to URL”) que enlaza con el objeto,
- de una eventual fuente original de la obra (“URL fuente de la obra”) y
- de otros eventuales permisos no contemplados en la licencia CC,

El ejemplo de la siguiente imagen es el resultado de seguir los pasos anteriores. En el marco 1 se puede apreciar el logo; en el segundo las referencias al autor, el título y el link al texto de la licencia; y en el tercer marco se encuentra el código relativo.

De este modo, la herramienta permite copiar y pegar en línea, y en lenguaje HTML, las características generadas por la licencia, que se han determinado en el proceso de elección de la misma.

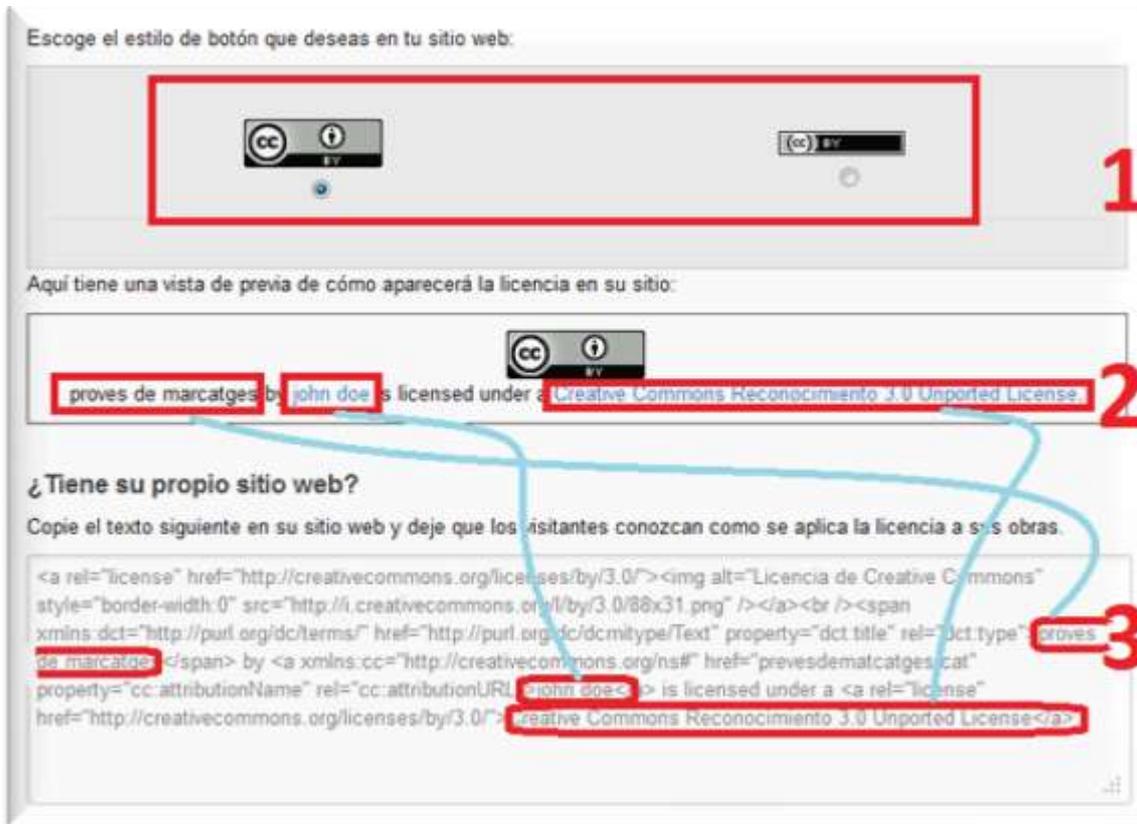


Figura 28. license on-line choosing tool, resultado para texto

Las líneas azules que relacionan los marcos 2 y 3, identifican las propiedades que visualizan los usuarios del marco 2 en el texto del código.

Esta herramienta permite, como se ha mencionado con anterioridad, la generación de HTML para los usuarios que sepan utilizarlo. Y es de gran utilidad debido a que permite, a individuos con conocimientos básicos, incrustar el código en sus páginas web sin tener que escribir, ni tampoco conocer el lenguaje creado para definir los derechos de autor. El hecho de que con pocos *clicks*, y sin tener que conocer el *background* de las licencias sea posible obtener protección legal sobre la propia obra de manera totalmente gratuita, hacen del *Licensing choosing tool* una herramienta muy fácil y rápida de difusión de las licencias CC en entorno web.

A continuación se representa el código para su mejor explicación.

```
<a rel="license" href="http://creativecommons.org/licenses/by/3.0/"></a><br /><span xmlns:dct="http://purl.org/dc/terms/" href="http://purl.org/dc/dcmitype/Text" property="dct:title" rel="dct:type">proves de marcatges</span> by <a xmlns:cc="http://creativecommons.org/ns#" href="http://provesdematcatges.cat" property="cc:attributionName" rel="cc:attributionURL">john doe</a> is licensed under a <a rel="license" href="http://creativecommons.org/licenses/by/3.0/">Creative Commons Reconocimiento 3.0 Unported License</a>
```

licensed under a [Creative Commons Reconocimiento 3.0 Unported License](http://creativecommons.org/licenses/by/3.0/).

Este código es el resultado del uso de distintos lenguajes y esquemas. Además del HTML, se utiliza el RDF en la versión RDFa, que permite la interpretación semántica de los objetos digitales. RDF es un modelo estándar para el intercambio de datos en la Web. RDF tiene características que facilitan la fusión de los datos incluso si los esquemas subyacentes son diferentes, y apoya específicamente la evolución de los esquemas. RDF extiende la estructura de enlaces de la Web con los URI²⁴ mediante los cuales se identifica una relación entre dos objetos a los dos extremos del enlace (esto generalmente se conoce como "triple"). Esta estructura forma un grafo dirigido y etiquetado, donde los extremos representan el enlace entre dos recursos, representados por los nodos del grafo. Se presenta una esquematización de una arquitectura basada en estos lenguajes:

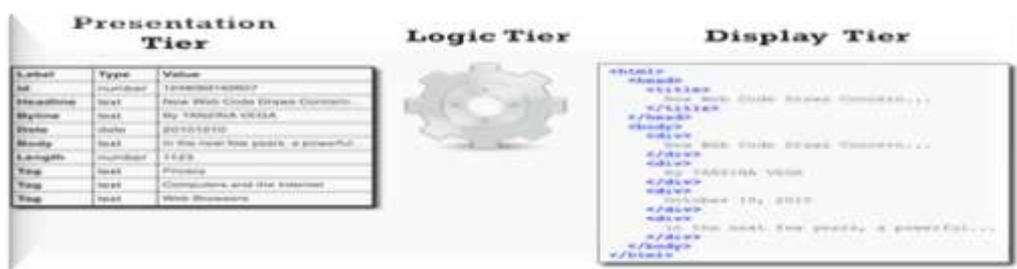


Figura 29: arquitectura a tres niveles de una cualquiera aplicación web actual. Fuente: <http://dev.iptc.org/rNews-Why-rNews>.

La presencia de distintos niveles, o capas, es un elemento clave en el desarrollo de la red semántica, pues se concreta en la capacidad de vehicular contenidos con propiedades descriptivas incrustadas, muy útiles a los usuarios.

En el ejemplo del código generado con el *License choosing tool* se encuentra la inclusión de los metadatos de licencia en un documento HTML.

El atributo `rel="license"` proporciona un solo metadato a través del tag `rel=`, que genera un triple, con predicado `xhtml:license`. El triple relaciona la URI del documento corriente con la URI de la licencia CC-BY y con la URI para el predicado `xhtml:license`. Esto sucede gracias a las características reservadas al prefijo `rel` y su atributo `xhtml:license`, según el cual no hay que definir ningún parámetro más para su uso (como definir un *namespace*), dado que es una característica propia (*built-in*) del tag `rel`. Y el objeto del atributo `rel="license"` se define simplemente a través de la URL en el atributo `href`, que será aplicable a todos los atributos `rel`. Dicho de otra manera, se atribuye una relación (`rel`) con el ítem (el objeto digital a marcar), en

²⁴ Un identificador uniforme de recurso o URI es una cadena de caracteres corta que identifica unívocamente un recurso digital, normalmente accesible en una red o sistema.

el cual se incrusta el argumento de la relación, que es información sobre la licencia. Ésta, a su vez, se encuentra en la dirección `http` precedida del signo `=` y entre comillas: `"http://creativecommons.org/licenses/by/3.0/"`.

De hecho se puede asociar `rel=license` a cualquier tag con un atributo `href` válido; esto significa que para asociar los links a la licencia, cualquier elemento puede estar entre los tags `<a>`, incluyendo, como en este ejemplo, la imagen del logo de la licencia.

Hasta el punto descrito anteriormente se ha utilizado el lenguaje HTML, en su versión XHTML²⁵. Pero en la tercera línea se introduce, con `xmlns:dct`, la asociación del prefijo `dct` con la URL `http://purl.org/dc/terms`, que resulta RDFa. Con `xmlns:dct` se introduce la referencia a un contexto de uso específico, un *namespace*, definido Dublin Core (`dct`), y referido a un dominio específico de términos (`terms`) y tipos de ficheros (`dc:type`) que se encuentra en las direcciones `http://purl.org/dc/terms/` y `"http://purl.org/dc/dcmitype/Text"`.

Explicitar de tal manera un esquema (*namespace*) permite invocar propiedades descritas en el mismo, de manera que cualquier intérprete²⁶ entienda el sentido de la declaración. Ésta, al fin y al cabo, no es más que la potencia que irradian los metalenguajes, capaces de recuperar o describir recursos digitales de naturaleza variada. En el ejemplo aquí tratado, el código generado con el *License choosing tool* representa una licencia de derechos intelectuales CC-BY (Creative Commons ... License) por un texto (Text), un título (`dct:title` y [proves de marcatges](#)) y un autor (`cc:attributionName` y `john doe`), que se referencia con una URL (`cc:attributionURL` y `http://provesdematcatges.cat`). El uso de los esquemas `dct` (Dublin Core) y `cc` (el esquema de *Creative Commons* declarado en la dirección <http://creativecommons.org/ns#>) permiten, por lo tanto, la sincronía de entendimiento entre hombre y máquina.

El mismo resultado lógico se obtiene con la representación de un grafo en el cual el punto de origen se conecta a través de las propiedades de una flecha, a un punto de referencia:



Figura 30: representació de un triple

²⁵ XHTML es HTML expresado como XML válido. El objetivo básico es conseguir la separación del contenido (la información vehiculada) de la forma de su presentación.

²⁶ Un intérprete puede ser tanto el *navegador* de un usuario como el sistema de *recolección* de un repositorio digital; en este sentido, el intérprete de un objeto digital bien marcado con metadatos descriptivos y administrativos incrustados, puede ser a la vez un sistema automático y un ser humano que maneja y extrae desde el mismo lo que le sea más útil.

Cada triple representa una declaración de la relación entre los objetos denotados por los nodos que enlaza. Sus tres componentes son:

- Un sujeto que es una referencia URI de RDF o un nodo.
- Un predicado que es una referencia URI de RDF (una propiedad) que denota una relación.
- Un objeto que es una referencia URI de RDF, un literal o un nodo.

Una triple RDF se escribe convencionalmente en el orden: sujeto, predicado y objeto. La dirección del arco es significativo: siempre apunta hacia el objeto. Los nodos de un grafo RDF son sus sujetos y objetos.

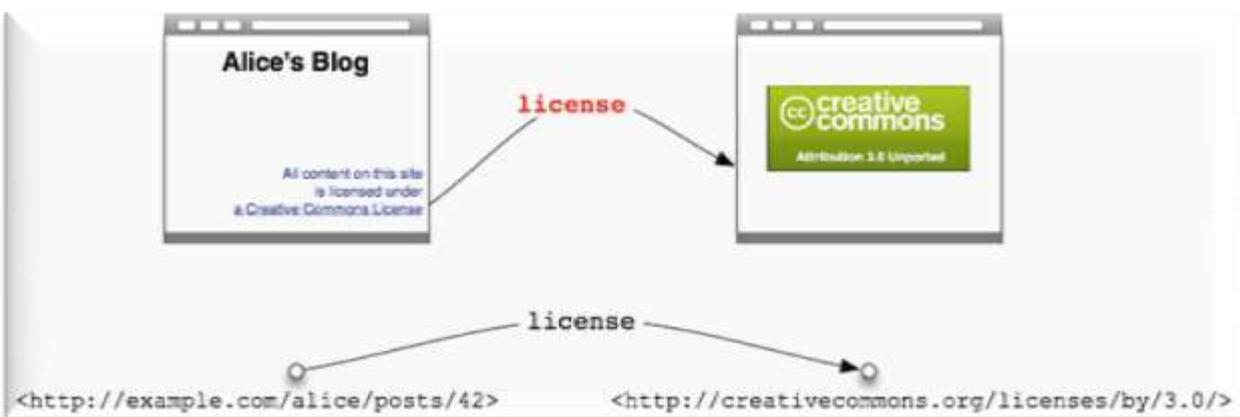


Figura 31: explicación grafica de un triple, o grafo

Dspace, mediante el módulo de licencias *Creative Commons*, permite también atribuir a los objetos depositados una licencia de dominio público, con la cual el detentor de los derechos renuncia a todas sus prerrogativas, dejando el objeto libre y sin limitaciones de uso.

A continuación se muestra una transcripción literal del código relativo a un objeto con licencia de dominio público, para facilitar su comparación con el código de una licencia BY:

```
<p xmlns:dct="http://purl.org/dc/terms/" xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#">
  <a rel="license" href="http://creativecommons.org/publicdomain/zero/1.0/">
    
  </a> <br />
  To the extent possible under law,<a rel="dct:publisher"
  href="http://provesdemarcatge.cat"><span property="dct:title">John Doe</span></a>has waived all copyright
  and related or neighboring rights to <span property="dct:title">proves de marcatge</span>.
  This work is published from:
  <span property="vcard:Country" datatype="dct:ISO3166"
  content="ES" about="http://provesdemarcatge.cat">
  España</span>.</p>
```

No hay gran diferencia en las declaraciones: se utiliza también aquí el esquema `xmlns:dct`, y se invoca también el `xmlns:vcard`, sin necesidad de utilizar el `xmlns:cc` del anterior ejemplo. Es suficiente declarar el editor (`dct:publisher`) que se encuentra en la dirección

(<http://provesdemarcatge.cat>), a cuyo nombre ([john doe](#)) se asocia la renuncia de los derechos sobre la obra con title ([proves de marcatge](#)). Con la propiedad vcard:Country se asocian características relativas a los datos contenidos (datatype="dct:ISO3166), a la lengua (content="ES") y al origen (<http://provesdemarcatge.cat> y España).

A continuación se adjunta una captura de un objeto digital correctamente marcado en un Dspace que implementa la mayoría de los elementos de definición del esquema Dublin Core:

The screenshot shows a DSpace item record with the following metadata:

| | | |
|-----------------------|---|-------|
| dc.contributor.author | Stachniss, Cyril | en_US |
| dc.date.accessioned | 2011-04-20T22:40:54Z | |
| dc.date.available | 2011-04-20T22:40:54Z | |
| dc.date.issued | 2010-12-07 | en_US |
| dc.identifier.uri | http://hdl.handle.net/1721.1/62290 | |
| dc.description | Date : 2003-11-18 Where : Univ of Freiburg, building 079, AIS-Lab Robot-type : Pioneer2 with 1 LMS-Laser What : 1 robot (magnum with stayton board) was driving around in building 079 File format : carmen logger format (POS, FLASER), only laser and odometry data recorded, and a scanmatched rec & carmen log file File(s) : fr079-sm.log.gz fr079-map.tgz | en_US |
| dc.rights | CC0 1.0 Universal | en_US |
| dc.rights.uri | http://creativecommons.org/publicdomain/zero/1.0/ | en_US |
| dc.subject | Indoor sensor data | en_US |
| dc.subject | 2D sensor data | en_US |
| dc.subject | SICK LMS laser range finder | en_US |
| dc.title | fr079 | en_US |

Files in this item

| Files | Size | Format | View | Description |
|----------------------------|---------|-----------|---------------------------|-------------|
| data file | 1.417Mb | Unknown | View/Open | data file |
| data file | 966.3Kb | Unknown | View/Open | data file |
| data file | 1.428Mb | Unknown | View/Open | data file |
| image file | 334.0Kb | GIF image | | image file |

[Download all files as a METS zip archive.](#)

Figura 32: captura de los metadatos de un objeto depositado en un Dspace que implementa todos los elementos DC

Siguiendo la revisión de las técnicas de marcado, se presentan seguidamente otras herramientas. Puesto que uno de los formatos más difundidos en la red es el PDF (*Portable Document Format*), se han desarrollado distintos sistemas para su marcado con referencia a las licencias de autor. El mayor problema es la heterogeneidad de los aplicativos y lenguajes utilizados para la codificación de distintos modelos de metadatos.

Por ello resulta interesante hacer una una revisión de los aplicativos y lenguajes que se han predispuesto para el marcado de los PDF encontrados en la página de *Creative Commons*. Un ejemplo importante, para el objetivo de este trabajo, es un marcador escrito en java que consigue acoplar metadatos con indicaciones sobre derechos de autor a un fichero PDF. El nombre del programa es *PDFLicenseManager* y han contribuido a su implementación Enrico Masala en 2006 y Asheesh Laroia en 2008. Se reproducen unas capturas para demostrar su funcionalidad:

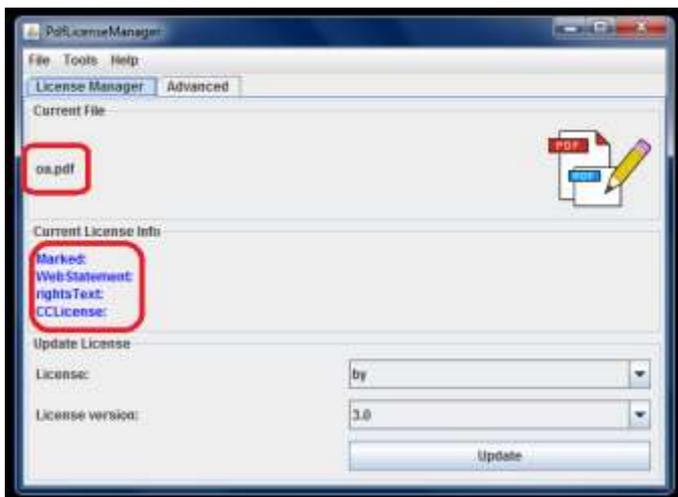


Figura 33: pdflicensemanager: sin metadatos

El programa se ha cargado y ha abierto el fichero oa.pdf: no resultan incrustados metadatos

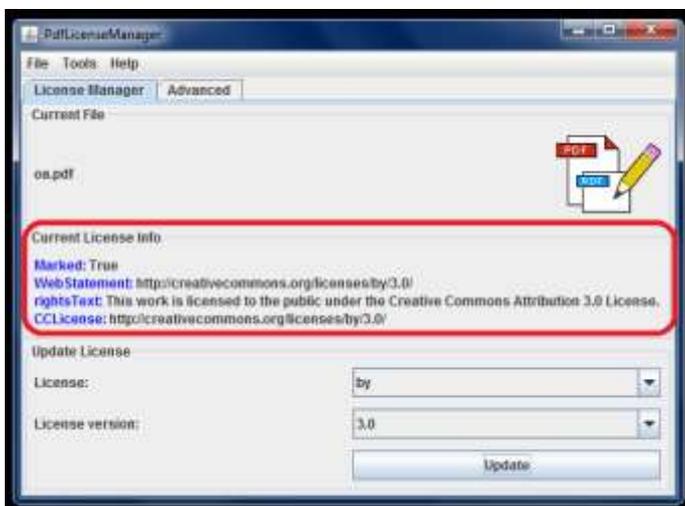


Figura 34: pdflicensemanager: con metadatos

Después de haber elegido el tipo de licencia con los selectores y haber pulsado *Update*, se pueden ver los metadatos añadidos e incrustados en el mismo fichero.

Esta tecnología utiliza el lenguaje XMP (*Extensible Metadata Platform*) desarrollado por su propietario, la empresa *Adobe Systems Incorporated*, y difundido con una patente de uso público (*Public Patent License*), permitiendo así la libre utilización del lenguaje²⁷. En la página de productos de *Adobe* (www.adobe.com/products/xmp) se puede leer: “*Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself*”. De hecho, es posible marcar muchos formatos de contenido (audio, vídeo, texto, imágenes estáticas o dinámicas, etc.). Sin embargo, en este trabajo sólo se consideran los PDF, ya que es el formato de la mayoría de los objetos (artículos de revista) depositados en el repositorio institucional. Los colectivos que utilizan este lenguaje

²⁷ A final de abril de este año (2012) salieron las nuevas especificaciones (Adobe Systems Inc., a), (Adobe Systems Inc., 2012b), (Adobe Systems Inc., 2012c).

son muchos y de gran dimensión, como *Digital Image Submission Criteria* (DISC), *Dublin Core Metadata Initiative* (DCMI), *International Press Telecommunications Council* (IPTC), *Metadata Working Group*, *Picture Licensing Universal System* (PLUS) *Coalition*, *Publishing Requirements for Industry Standard Metadata* (PRISM), *World Wide Web Consortium* (W3C) y, naturalmente, *Creative Commons*. Con una aplicación habilitada para XMP, la información sobre un fichero puede ser capturada durante el proceso de creación de contenidos e incrustada en el archivo, y se transfiere a un sistema de gestión de contenidos. Descripciones significativas y títulos, palabras clave de búsqueda, el autor, y hasta la información de derechos de autor, pueden ser capturados en un formato que sea fácilmente entendido por humanos y por las aplicaciones de software, dispositivos de hardware, o por cualquier sistema *parser* que filtre la información contenida en los metadatos.

Otro interesante ejemplo de desarrollo de APIs para el marcado es representado por el *CCtagger*, un *plugin* para la plataforma de blog *Wordpress*. En las figuras 35-37 se muestran las capturas de las opciones que ofrece esta API.



Figura 35: Ilustración 3: resultado del marcado con CCtagger en Wordpress

CC-Tagger Options

This section describes how the Creative Commons License settings work: The default settings are applied to any media that has not yet been tagged or has its settings set to 'default'.

Jurisdiction of your license Information:

Image Size:

Display a text link to the license: Link No Link

By Default allowed use of untagged media:

By Default allowed modification of untagged media:

add search options to the search form:

Figura 36: opciones del CCTagger

Run CC-Tag Batch Index

[Find more plugins by this author](#)

when you submit this form, the posts between the starting and ending IDs will be tagged with the settings below. Media that is already tagged will not be changed unless the 'overwrite existing license option' is checked.

batch starting post-id
(leaving this blank will select the first media)

batch ending post-id
(leaving this blank will select the latest uploaded media)

Allowed modification of untagged media:

Allowed use of untagged media:

Should these settings overwrite existing settings on the media don't overwrite existing license data overwrite existing license data

Figura 37: proceso de batch en CCTagger

Respecto al *PDFLicenseManager*, descrito con anterioridad, el resultado final en el objeto marcado no presenta diferencias. Sin embargo, aquí no se trata de PDF, sino de distintos formatos de ficheros que pueden ser marcados con los derechos de autor. Una diferencia sustancial es que el primer aplicativo trabaja en local y es independiente (*stand-alone*), mientras que *CCTagger* puede funcionar sólo como *plugin* dentro de otro programa, como por ejemplo en *Wordpress*. Además está escrito en php, a diferencia de *PDFLicenseManager*, que es en java.

Siguiendo en la panorámica sobre los desarrollos de sistemas de marcado con CC, existe un *plug-in* para *MicrosoftWord*, y otro para *OpenOffice*, que permiten atribuir licencias de autor

como si se utilizara el *driver* de una impresora. El resultado es un fichero PDF con metadatos incrustados, además de poder contar con la aplicación, dentro del mismo documento, del logo CC y del link al texto legal de la licencia. En la figura 38 se aclara el contexto de uso.

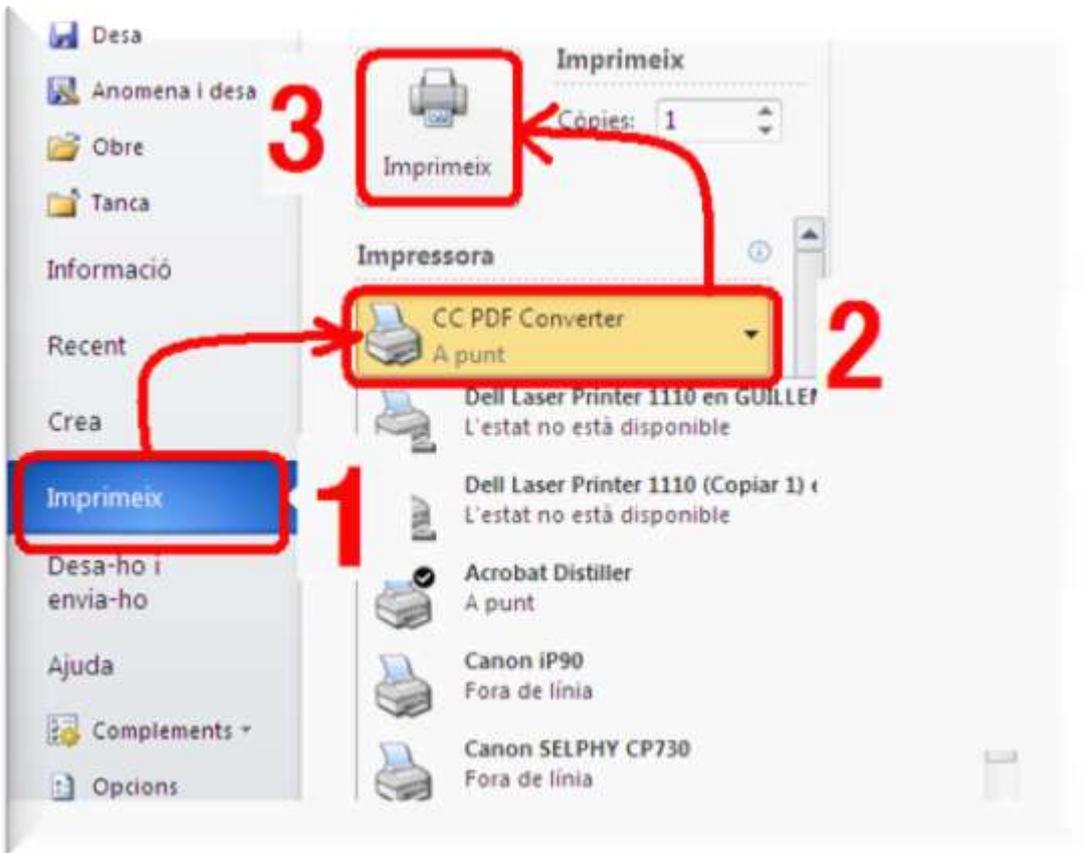


Figura 38: menú principal de MS Word, con CCPDFconverter

En la figura 39 se muestra el resultado de un fichero marcado con este sistema integrado en el mismo procesador *MSWord*.



Figura 39: resultado del Driver CCPDFconverter

A continuación (figuras 40-42), se muestran las capturas del proceso de marcado con *CC PDF Converter*.

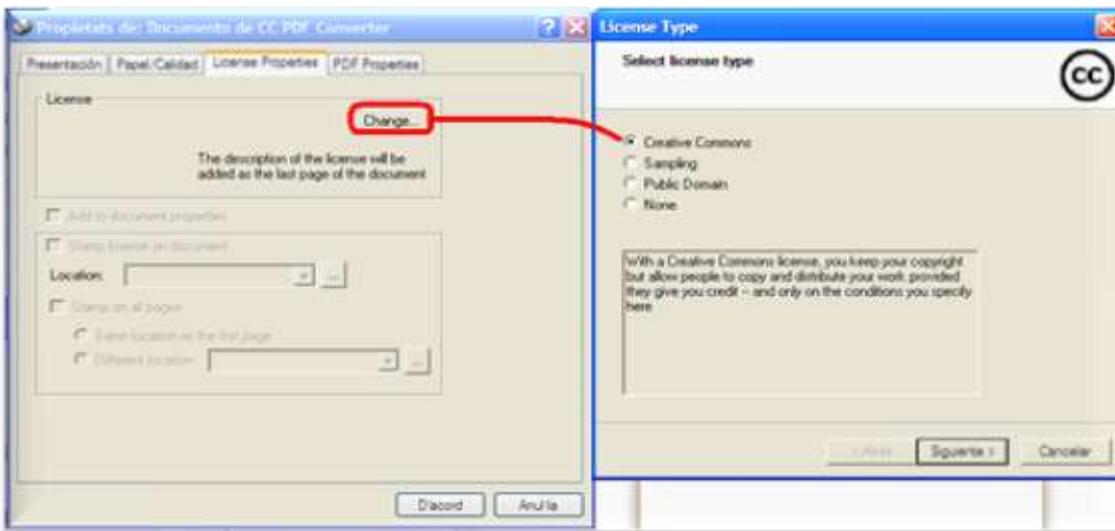


Figura 40: visualización de la ventana de propiedades de CCPDFConverter



Figura 41: proceso de recuperación de esquema en CCPDFconv.



Figura 1: preguntas para obtener la licencia



Figuras 43 y 44: últimas fases de elección de la licencia con CCPDFConverter

Los ejemplos ilustrados pertenecen a la esfera de herramientas que permiten el marcado de ficheros cuyos formatos son mayoritariamente utilizados en el contexto de la red: PDF, JPEG y

HTML. Facilitar el reconocimiento de derechos de autor en una página HTML es hoy algo extremadamente simple: se instala un *plug-in* en un navegador²⁸ y, si el *navegador* encuentra triples en RDFa, muestra al usuario, en distintos formatos, los datos de licencia de autor. Se muestra la visualización de los datos obtenidos con el *plug-in* instalado en el apartado siguiente. Sin embargo, el marcado de los ficheros PDF y JPEG no es tan simple.

El formato JPEG es un ejemplo común de fichero de imagen. Por su amplia difusión, *Creative Commons* ha podido experimentar, a través de la ayuda de desarrolladores voluntarios, la creación de herramientas para el marcado de imágenes. El ejemplo que se presenta es el llamado *PHP JPEG Metadata ToolKit*, escrito por Evan Ozhiker, y se trata de un *script* en PHP que permite obtener la asociación de metadatos descriptivos con un fichero de imagen. El manejo de este software²⁹ no es para principiantes o usuarios finales, pues los conocimientos necesarios para su manejo implican el conocimiento de la consola de mandos en línea en entorno PHP, pero permite la incrustación de metadatos definidos por el usuario y por las cámaras de fotografía que hayan capturado la imagen. El desarrollo de este *script* no es reciente, y se puede considerar superado por las técnicas y los *plug-ins* más actuales. Se presenta una captura de los resultados de marcado con este aplicativo:



Figura 45: Uso del PHP JPEG Metadata ToolKit

Se consideran ahora otros formatos de ficheros.

Para los ficheros audio y video es disponible una aplicación compatible con Windows denominada ccPublisher, del cual se adjuntan algunas capturas descriptivas.

²⁸ Al momento de compilación de este escrito existe un plugin disponible para chrome, firefox y opera a través de <http://openattribute.com/> [consulta 23.05.2012]

²⁹ Se ha descargado el código desde la dirección <http://www.ozhiker.com/electronics/pjmt/> en la fecha 01.04.2012.



Figura 26: CC publisher



Figura 47: selección target y marcado del fichero



Figura 48: selección licencia y opción subida en IA

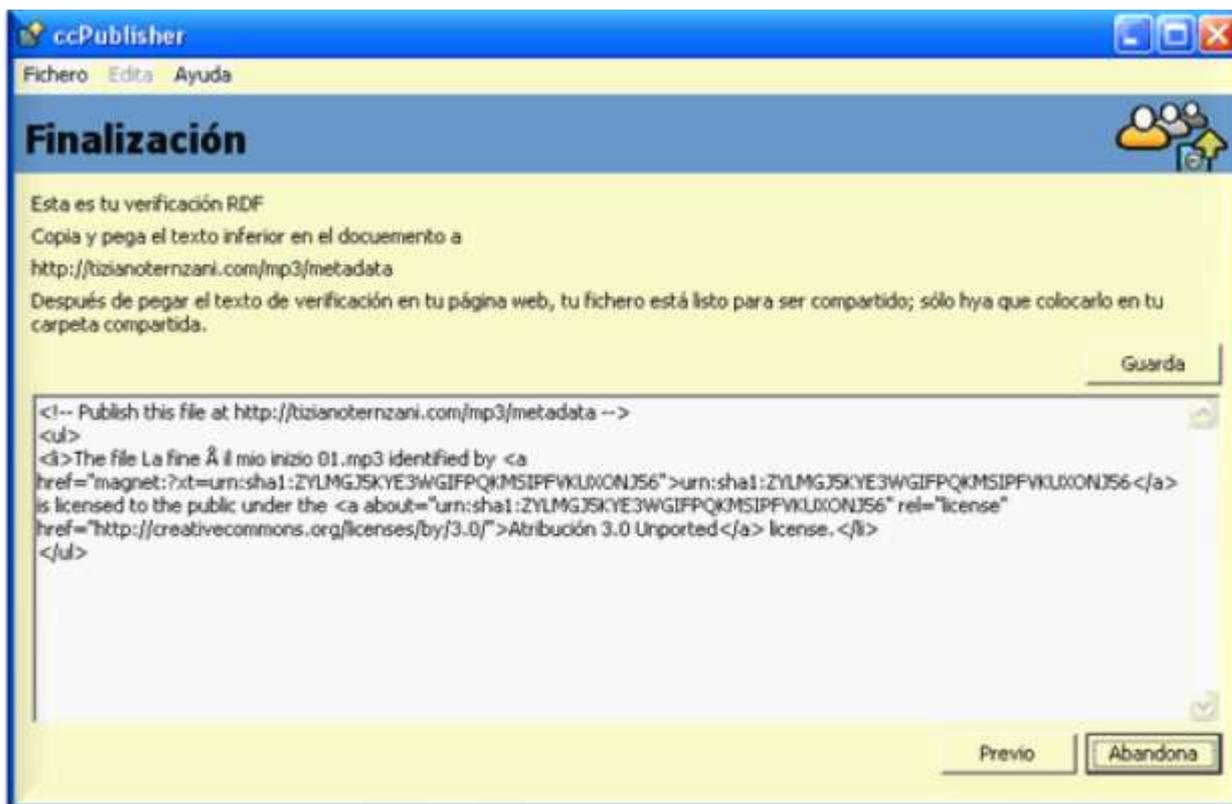


Figura 49: resultado código a copiar y pegar

Como se puede ver en esta última captura el programa genera un código HTML que contiene el título de la obra, una clave de verificación SHA-1 y un enlace a unas licencias relacionado con la clave. Si se incrusta este código en la página indicada en el comentario inicial entre `<!--` y `-->`, que se ha indicado en la fase de marcado, el usuario que utilizará esta obra dispondrá de un mecanismo de verificación, de la autenticidad de los derechos representados en la licencia, a través de la correspondencia de la clave contenida en el objeto utilizado y la clave presente en la dirección indicada para la atribución de los derechos. Esta solución es utilizable y aconsejada al autor, así como al editor, quieran facilitar esta herramienta de incremento de la confianza durante la manipulación del objeto digital descargado.

4.3 La gestión de derechos en repositorios digitales: estado de la cuestión

El conjunto de metadatos que describe un objeto en Dspace se puede distinguir en metadatos de descripción y metadatos de administración. Los metadatos descriptivos describen atributos del objeto, como el nombre del fichero, su autor, el tamaño, etc., mientras que los administrativos se refieren a elementos para la administración del objeto, como la localización del objeto, el usuario que lo ha creado, etc.

- Dspace permite el uso de cualquier esquema de metadatos, incluso los de derechos de autor. En la versión de fábrica, el software se distribuye con el esquema *Dublin Core* desarrollado por la *Dublin Core Metadata Initiative*³⁰. En este conjunto de metadatos se encuentra el elemento `<dc:right>`, el dominio `dc`³¹ predefinido para atribuir los derechos de autor. Pero, entre los campos previstos en la versión “de fábrica”, este metadato no resulta habilitado, aunque esté presente en el esquema. Para su activación y recolección es necesario modificar un fichero de configuración³², que hace referencia a la organización de toda la fase de envío. Una situación parecida se encuentra también en la versión de fábrica de SWORD, la aplicación que permite el envío masivo de objetos y metadatos al repositorio, donde se implementa también el *namespace* Dublin Core, pero que tiene que ser personalizado para que sean habilitados elementos como el *dc.rights.uri*.
- Como se aprecia en las capturas del apartado 4.2, de una versión de Dspace no personalizada (figuras 12-20, excepto fig. 19), en ninguna ilustración aparece un campo de metadatos para la inserción de los derechos de autor (el campo `dc:rights` de *Dublin Core*). Si no se habilita desde el principio de la fase de envío, es mucho más complicado poder añadir este elemento al objeto, pues hay que realizar un trabajo manual de posterior recuperación de los derechos, lo que implica un incremento de las actividades y del personal administrador.

³⁰ En la actualidad Dspace suele integrar otros esquemas de metadatos, como en el caso del DD que integra formatos de metadatos cuales METS, ORE, RDF y el esquema OAI-DC.

³¹ Para el detalle de la personalización de los formularios utilizados durante el envío de un objeto vía la interfaz usuario, se puede consultar (Lewis & Yates, 2008). En particular el modulo sobre configuración desde la página 13 hasta página 24 <http://www.dspace.org/training-grid/configurable-submission-system-for-dspace.html> [consulta 01.03.2012]

³² El fichero de configuración a modificar es el `input-form.xml`, en el cual se habilitan los campos de los esquemas que se cargan en Dspace. Este fichero es parte integrante del proceso de envío, (*submission process*).

- En dichas capturas también se muestra el paso de atribución de licencia de autor, integrado en Dspace, cuya funcionalidad es la de asociar un texto con los derechos de autor al fichero tramitado. En la versión de fábrica se presenta un texto genérico de licencia (figura 19). Sin embargo, Dspace integra también el módulo de licencias *Creative Commons*, que hay que activar (figuras 17-18), personalizando la versión de fábrica. A través de este módulo se presentan al depositario (el sujeto que envía un objeto digital al repositorio) las páginas de selección del tipo de licencia CC que se quiera aplicar. Las combinaciones de licencias disponibles se detallan en el párrafo 4.2.1.
- Durante el proceso de envío de un ítem hay que elegir el tipo de licencia con la que se difundirá el objeto depositado. El módulo de CC presente en Dspace permite elegir entre seis diferentes licencias y, una vez elegida la más apropiada, continua el proceso de envío hasta la fase final, donde se presenta al depositario un resumen de los datos identificativos del objeto.
- Actualmente, en el DD se encuentran objetos correctamente etiquetados, fruto de las pruebas de implementación de esquemas añadidos y de la recuperación posterior de las políticas de derechos de los objetos. Seguidamente, el equipo de administración del repositorio valida el conjunto de datos para la publicación definitiva en abierto a través de interfaz web. Este momento se denomina técnicamente *workflow*, y se coloca idealmente entre la fase de envío (*submission*) y la de archivo (*archiving*): aquí, el equipo de administración del repositorio es el responsable de cumplir con las tareas de optimización, a través de *inputs* iterativos o de un progresivo refinamiento de los metadatos creados hasta el momento.
- La propuesta del proyecto es modificar el actual *workflow*, es decir, el proceso completo de envío de un objeto. Se propone añadir o, mejor dicho, incrustar a los ficheros depositados una capa de información más, la de metadatos dedicados a la connotación legal de una obra, necesaria a las máquinas para conseguir la información sobre los derechos de autor de manera automática. Y, por supuesto, útil también al usuario final que dispondrá de herramientas de reconocimiento y manejo de los derechos³³. Esta integración quiere favorecer los acuerdos necesarios entre la UB y los editores, para que éstos puedan autorizar el depósito de un artículo directamente en

³³ Se hacen ejemplos de estos tipos de herramientas de “descubierta” de los derechos de autor después en el apartado sobre resultados de visualización de marcado.

DD, gracias al respeto de los acuerdos de explotación de la obra, garantizados con la implementación de los nuevos metadatos consultables directamente desde el DD.

La capacidad del DD de dialogar autónomamente, es decir, sin la intervención humana, con los sistemas de gestión de los editores, es el motivo que sustenta la propuesta de proyecto. De este modo, se propone solucionar el problema de la automatización del depósito de artículos en acceso abierto bajo una condición de embargo. Gracias a la capacidad del DD de abrir los artículos sólo transcurrido el periodo previsto, se tutelan los intereses del editor que tiene así más confianza y más motivos para permitir la ingesta de las obras ya publicadas en sus revistas.

4.4. Necesidad de normalización en el mercado

Un ejemplo de normalización es representado por el caso de convergencia de las definiciones *dc:rights* y *xmp:RightsWebStatement*, la primera de ámbito de aplicación *Dublin Core* y la segunda de XMP. El valor actualmente aconsejado es hoy el mismo, el elemento de DC. El desarrollo paralelo de los dos esquemas ha creado elementos de etiquetado idénticos: el valor a relacionar con el elemento es la referencia a los derechos de autor. La duplicación de etiquetas semánticas dificulta la interoperabilidad y, por lo tanto, se ha conseguido un acuerdo de normalización del metadato (y de la referencia a su esquema, por supuesto). La unificación es posible gracias a las propiedades de lenguajes como HTML5, XHTML, XML, RDFa y otros, que permiten una gestión semántica de los metadatos de los objetos, también debido a los acuerdos que se consiguen entre los *players* en la arena de la red. Entre las instituciones más activas se encuentra el consorcio W3C, desde siempre promotora de la difusión de estándares desarrollados en sus grupos de trabajo.

Durante la preparación de este trabajo (en abril de 2012), ha visto luz una publicación académica (Guibault & Angelopoulos, 2011) en acceso abierto en la que se discute y presenta la nueva versión del lenguaje de *Creative Commons* para el mercado, el CC REL, que actualiza la versión de 2008. Aunque se hayan podido leer las nuevas formulaciones propuestas del nuevo dominio <http://creativecommons.org/#ns/ccrel>, deben pasar algunos meses (o años) para que se pueda apreciar su implementación y difusión. La creciente adopción de licencias CC augura una solución definitiva al mercado digital de los objetos digitales.

4.5 Antecedentes y resultados previos que avalan la propuesta

En un contexto como el actual, no estar en línea con los cambios impulsados por los nuevos modelos de comunicación e información significa perder en parte la capacidad de difusión del mensaje vehiculado por cierto medio. Tanto a nivel individual, como de organización, ser un nodo en una red de interconexiones es vital. Los conocimientos sobre la web semántica son, en el contexto digital de la gestión de contenidos, una pre-condición necesaria en una arquitectura informática eficiente y eficaz de soporte a cualquier actividad.

Durante la preparación y redacción de este trabajo se ha podido experimentar en primera persona una situación de gran fermento. Se ha podido seguir la continua actualización de las páginas de *Creative Commons* dedicadas al mercado digital: mucha de la información consultada se ha ido actualizando durante estos meses en espacios comunitarios tales como wikis, listas de distribución entre grupos de expertos, o páginas web. La misma sensación de cambio hacia el acceso abierto se puede extrapolar de las decisiones de varios organismos internacionales, como la UNESCO, el Banco Mundial (*World Bank*), la Casa Blanca, la Unión Europea, y otras muchas, que públicamente optan por políticas de acceso abierto (por ejemplo, se espera para el 20 de junio de 2012 la publicación de la Declaración de París relativa a las políticas de promoción de los Recursos Educativos Abiertos, o la inauguración del Portal Global del Acceso Abierto de la UNESCO para el 1 de noviembre de 2012.). Todo ello implica consecuencias como el estímulo a la creación de repositorios abiertos, así como la adopción de políticas editoriales que dictan inmediata disponibilidad de los documentos producidos por la propia organización (Swan, 2012). Sin referirse a contextos globales, la historia de las licencias abiertas en la UB es otro ejemplo de la creciente y dinámica adaptación al nuevo contexto de libre circulación de la ciencia.

Estos ejemplos no son más que un escenario que sirve para justificar este proyecto, pues la necesidad a la cual concretamente quiere responder la propuesta aquí formulada, es la mejora de la capacidad de difusión y recuperación de los artículos de los investigadores de la UB. La adopción de “puentes” entre sistemas de gestión de la información es justamente un tema de interoperabilidad y de semántica. Y la solución aquí propuesta de aprovechamiento de datos de un sistema para la alimentación de otro no es nada nuevo, pero si podría serlo la integración de la función de marcado de derechos de autor.

La pasarela construida para la recuperación de los datos desde el GREC procura las referencias a las obras publicadas, pero el GREC no se ocupa específicamente de datos relativos a las políticas de derechos de autor adoptadas en las revistas. Cuando no se indiquen en el GREC (se acepta cierta heterogeneidad en los datos introducidos para la identificación de los artículos publicados), estos datos deben añadirse manualmente. Como se ha mencionado con anterioridad, el DD posee la capacidad de expresar las políticas de derechos de autor adoptadas por las revistas, pero tiene que poder contar con mecanismos automáticos de recuperación, que se consiguen mediante la adopción de las nuevas funcionalidades descritas en este trabajo.

Un proyecto actualmente en fase de desarrollo, el SCOAP3³⁴ - Sponsoring Consortium for Open Access Publishing in Particle Physics – está experimentando algo similar a este tipo de enfoque: con una infraestructura informática capaz de hacer dialogar un repositorio institucional con los editores que hayan optado por depositar en él, se quiere conseguir el despliegue de un nuevo modelo de OA. En este nuevo modelo, el editor ya no cuenta con los ingresos que generan las suscripciones de múltiples instituciones, siendo el consorcio su único interlocutor y financiador. Cada socio SCOAP3 recupera su contribución de los fondos que genera la cancelación de las suscripciones a las revistas. Este modelo evita la desventaja obvia de modelos OA en el cual los autores pagan directamente. El presupuesto anual para el funcionamiento del SCOAP3 se establecerá a través de un procedimiento de licitación. De este modo se persigue ahorrar millones de euros, pasando de un modelo de pago por suscripción a uno de pago por publicación.

Otra propuesta previa concreta es el *Woods Hole Open Access Server (WHOAS)*³⁵, que presenta características más sofisticadas, ya que se trata de un proyecto para la interconexión de un repositorio de datos de investigación (no de artículos) para las instituciones de investigación que lo necesitan.

La creación de formatos específicos de ficheros, como el caso de RIS³⁶, de la empresa *Research Information System Incorporated*, para el intercambio de datos bibliográficos, también constituye otro antecedente que demuestra la extrema importancia de la interoperabilidad entre sistemas de gestión de contenidos académicos.

³⁴ <http://scoap3.org/>

³⁵ <https://darchive.mblwhoilib.org/>

³⁶ [http://en.wikipedia.org/wiki/RIS_\(file_format\)](http://en.wikipedia.org/wiki/RIS_(file_format))

Las políticas de financiación pública para la investigación están empezando a conectar el flujo financiero a mecanismos de inmediata disponibilidad (en tiempo real) de los datos de la investigación financiada. Un ejemplo de esta tensión, que afecta a los investigadores financiados con fondos públicos (como responsables de la puesta en abierto de sus resultados), serán los financiamientos del programa Horizon2020³⁷ de la Unión Europea: una vez concedidas las ayudas, para poder hacer efectivo el financiamiento, los investigadores deberán poner a disposición de la comunidad, en abierto, tanto las publicaciones como los datos de la investigación financiada, una vez publicados los resultados y, en su caso, transcurrido el periodo de embargo. Se ha creado un interesante sistema de control (basado en el protocolo OpenAire, que a su vez invoca OAI-PMH) de las publicaciones generadas a partir de proyectos financiados con este modelo: un repositorio central dispone de la lista de los proyectos financiados, a los cuales atribuyen un número identificativo. El investigador, cuando publica sus resultados, que podrán ser bases de datos a las cuales se podrá acceder en tiempo real, tiene que añadir el ID del proyecto para que su repositorio institucional y el del financiador, comuniquen y validen la publicación.

³⁷ http://ec.europa.eu/research/horizon2020/index_en.cfm

5. Metodología propuesta

Se propone el uso de la metodología de estudio de casos, tomando como objeto el repositorio institucional de la Universidad de Barcelona. Se consideran las publicaciones académicas y los repositorios en acceso abierto, con una perspectiva desde el particular al general. Esta configuración metodológica se replica desde un estudio³⁸ que considera y asocia los costes de las actividades puntuales llevadas a cabo por los actores principales involucrados en el registro de la investigación, su certificación, su disseminación y su gestión digital. Contrariamente a la mayoría de los estudios realizados, en esta investigación se han evidenciado los contextos específicos de cada organización considerada, y sus específicas elecciones en términos de escala y de alcance. El resultado del análisis es una valoración (estimación) de los costes de certificación del contenido, de publicación (incluyendo los metadatos) y de implantación/gestión de la estructura informática que sustenta el sistema de divulgación. Estas evidencias empíricas permiten aquí calcular los costes aproximados de realización del proyecto respecto a algunas partidas de gasto, mientras que otras no se podrán evidenciar dada la escasez de detalles en los datos disponibles.

El análisis, o mejor, su dibujo, se complementa con la síntesis de unos consejos sobre buenas prácticas, con los que se quiere aportar una visión actualizada de las experiencias y los consejos de organismos públicos involucrados en el acceso abierto. Estos consejos ponen un mayor énfasis en los aspectos relativos al autor que en aquellos concernientes a los otros actores involucrados en el proceso. Por esta razón, se ha desarrollado una guía de ingesta, que será evaluada y probada (implementación) por los administradores del repositorio institucional de la UB.

³⁸ (Centro ASK, 2011). En este estudio se ha investigado el comportamiento de los autores con respecto al auto-archivo, el uso de los repositorios y por fin, y más importante respecto la propuesta aquí contenida, los costes asociados en generar a los caminos alternativos al acceso a los productos de la investigación, en particular los artículos científicos. Desde el estudio, sobre los “economics” de la publicación en OA, se obtienen “[drivers for] the discussion on the ecology of scholarly publishing around the conditions for sustainability of different business models and on the interplay between different types of organizations]”. La coparticipación de distintos modelos económicos ayuda la libre competición respecto la financiación, con un evidente beneficio en la difusión.

5.1 Desarrollo de una guía de ingesta

Los artículos científicos son el resultado de un largo proceso temporal y lógico: la fase anterior a la revisión del artículo por parte del sistema de evaluación de la revista a la que se envía el artículo (*preprint*), la de evaluación, la de aceptación (*postprint*) y, en último término, el estadio de publicación. Las revistas gestionan todo el proceso, y con las tecnologías disponibles se ha habilitado la oportunidad para los autores de poner en circulación sus obras ya al principio de todo el proceso. Además, un creciente número de agencias de financiación requiere que a la conclusión del proceso de investigación estén disponibles en abierto los resultados, de modo que los autores sufren presiones para publicar en abierto.

A cada fase del proceso de publicación/producción de un artículo se le asocia un cierto valor: el *pre-print* representa una ocasión de primer contacto con la comunidad de investigadores, en búsqueda de *feed-back* y de cooperación, y también para prevenir la competición de grupos de investigadores. En la fase siguiente, ya puede difundirse un contenido de calidad (ha pasado la revisión) que potencia la velocidad con la que se difunden las ideas. La fase de publicación es la definitiva, que certifica la versión que se cita, archiva y conserva. Son las dos últimas fases las que asocian inequívocamente el artículo y el autor con una revista y un editor. Los derechos de autor, como se ha comentado, se transfieren tradicionalmente en la fase de *peer-review* y aceptación. Los editores declaran sus políticas hacia el autoarchivo, y definen bajo qué circunstancias el autor, o terceros, pueden difundir el contenido de la segunda o tercera fase entre la comunidad científica.

La necesidad del desarrollo de una guía de ingesta al DD, en la cual se recojan las buenas prácticas indicadas por las entidades con experiencia y práctica de éxito en los repositorios *Open Access*, se justifica en base a los nuevos requerimientos técnico-administrativos que afrontará el DD en la implementación del correcto marcado de las políticas de derechos de autor usadas por las revistas.

5.2 Realización de una prueba piloto

La aplicación de un nuevo sistema de marcado en fase de envío de un artículo será posible con el desarrollo de una modificación al actual *add-on* de licencias de autor de Dspace. El módulo que permite actualmente marcar el contenido no incrusta la licencia en el fichero depositado, de modo que al descargarlo desde el repositorio, las etiquetas de derechos de autor presentes en los metadatos dentro el repositorio no siguen al objeto bajado. La idea de marcar el fichero no es nueva en el ámbito *Creative Commons*, pero la implementación aquí propuesta es el resultado de un enfoque nuevo y original que quiere, por un lado sustentar la creciente confianza en el acceso abierto a las ciencias, y por otro otorgar mejores herramientas de marcado de los derechos de autor a nivel de repositorio institucional, donde ya están implementados esquemas de recuperación de los derechos, pero donde todavía no se han resuelto, de manera estandarizada, las cuestiones de los tres niveles de marcado de derechos de autor en los ficheros depositados. La prueba podrá considerarse satisfactoria si cumple con los requisitos de funcionalidad del módulo *add-on* modificado, es decir, con la capacidad de acoplar al artículo a distribuir, los metadatos necesarios para el reconocimiento automático y “humano” de los derechos de autor incrustados con las licencias *Creative Commons*.

A continuación se esquematizan las fases de la prueba piloto.

Esquematización de las fases de la prueba piloto



Indicadores de éxito

- *Stickiness* del mercado: la capacidad de garantizar la circulación de los derechos de autor incrustados en los PDF que se extraigan de un repositorio institucional, una vez que el documento circule entre *peers*.
- Incremento de la visibilidad-popularidad de la investigación UB desde el DD.
- *Recolección* desde GoogleScholar y principales portales científicos mundiales.

La prueba piloto de marcado de derechos de autor se realiza con objetos que dispongan de licencias CC. Inicialmente hay que realizar el trabajo de personalización de la API de “marcado incrustado”, para que se puedan tratar los artículos depositados en el repositorio. Se utiliza un sistema de control para el desarrollo colaborativo de software, como puede ser *git*³⁹. Una opción, si no se quieren utilizar los servidores institucionales, es en la nube, y podría apoyarse en *gitHub.com*, que ofrece las funcionalidades y características suficientes para el seguimiento del desarrollo necesario. Este servicio es gratuito siempre y cuando los nuevos aplicativos desarrollados estén disponibles en la plataforma en abierto, además de ofrecer la ventaja de poder contar con la colaboración de otros programadores interesados en participar en el proyecto.

En paralelo, para los artículos del PDI de los cuales se desconoce la política de derechos de autor, se recopilan estos datos manualmente. Esta información se añade en los campos activados para su representación, como el citado elemento *dc:rights* del dominio *Dublin Core*, o los más importantes (en este contexto) elementos de CC REL.

La fase 1 concretiza la aplicación del marcado incrustado y de los campos de políticas de derechos de las revistas.

En la fase 2 se llevan a cabo todas las pruebas necesarias para la validación de los elementos semánticos añadidos: con el reconocimiento de los artículos de prueba marcados por parte de los motores de búsqueda; a través de la recolección por parte de otros repositorios con los que la UB tenga acuerdo de cooperación; y con el reconocimiento por parte de los *plugins* existentes en las distintas aplicaciones, de los derechos relativos a los objetos y sus usos permitidos.

La fase tres prevé el marcado de la totalidad de los objetos depositados con la indicación de sus derechos y, en consecuencia, promueve la *recolección* completa del DD.

Con la consolidación de las prácticas de marcado incrustado y la obtención automática de las políticas de derechos de las revistas, se pasa a la fase 4, donde se buscan acuerdos directos con los editores dispuestos a aprovechar el ulterior canal de difusión que constituirá el DD con depósitos de *preprints*, *postprints* y versiones finales que cuenten con una correcta y transparente gestión de los derechos de autor.

³⁹ **Git** es un software de [control de versiones](#) diseñado pensando en la eficiencia y la confiabilidad del mantenimiento de versiones de aplicaciones, cuando estas tienen un gran número de archivos de [código fuente](#), como el caso de software a personalizar. Fuente: <http://es.wikipedia.org/wiki/Git> [consulta 20.05.2012]

6. Recursos necesarios

El personal previsto para el desarrollo del proyecto relativo a las nuevas funcionalidades del repositorio, es un equipo de cuatro personas con niveles jerárquicos de responsabilidades:

- Un coordinador de proyecto, que asume el control de todas las fases y tareas previstas: se encarga de dirigir el trabajo del informático y de los becarios, planifica acciones de comunicación sobre las nuevas posibilidades habilitadas con el nuevo sistema de representación de derechos, y produce material de formación sobre su uso.
- Un técnico informático con conocimiento de los lenguajes PHP, XML, Python y Javascript, para poder manipular el código del *addon* de *Creative Commons*; configurar los nuevos metadatos en el DD; y integrar el mecanismo automático de validación de las políticas de las revistas, relativas a las publicaciones de los investigadores de la UB.
- Dos becarios, principalmente de soporte al técnico informático, que realizan la verificación manual de las políticas de las revistas y la aplican a los objetos; colaboran en la elaboración e implementación de los distintos test previstos; y además, en la fase de comunicación y difusión de los resultados, ayudan al coordinador en la preparación de material divulgativo.

El programa requerido para la implementación de la nueva API de marcado es de código abierto, así como el código fuente de la aplicación original a partir de la cual se realiza la personalización. La activación de los metadatos de derechos de autor es una actividad cubierta por el informático del CRAI, y no suma costes.

En relación a la infraestructura y arquitectura necesarias para la implementación del proyecto, no se requiere modificación alguna del actual sistema de RI de la UB. En las fases finales del proyecto, habiendo validado los test de marcado y de recogida de datos, debe modificarse DSpace, sin que ello repercuta en el resto de la arquitectura de soporte de la UB.

7. Presupuesto

Para la identificación de los elementos que afectan a los costes de publicación en un repositorio institucional, hay que analizar aspectos como el coste medio asociado a la selección de un artículo (*peer-review*), los costes asociados a su publicación (incluyendo edición, maquetado, marcado de metadatos, referencias cruzadas y sistema de envío/ingesta), así como los costes asociados a la gestión de la estructura digital de soporte. El repositorio necesita también una fase de instalación y prueba antes de considerarse optimizado respecto a las necesidades específicas, así como una fase de gestión ordinaria. Finalmente, hay que considerar los costes de divulgación y formación sobre el uso y las ventajas del repositorio.

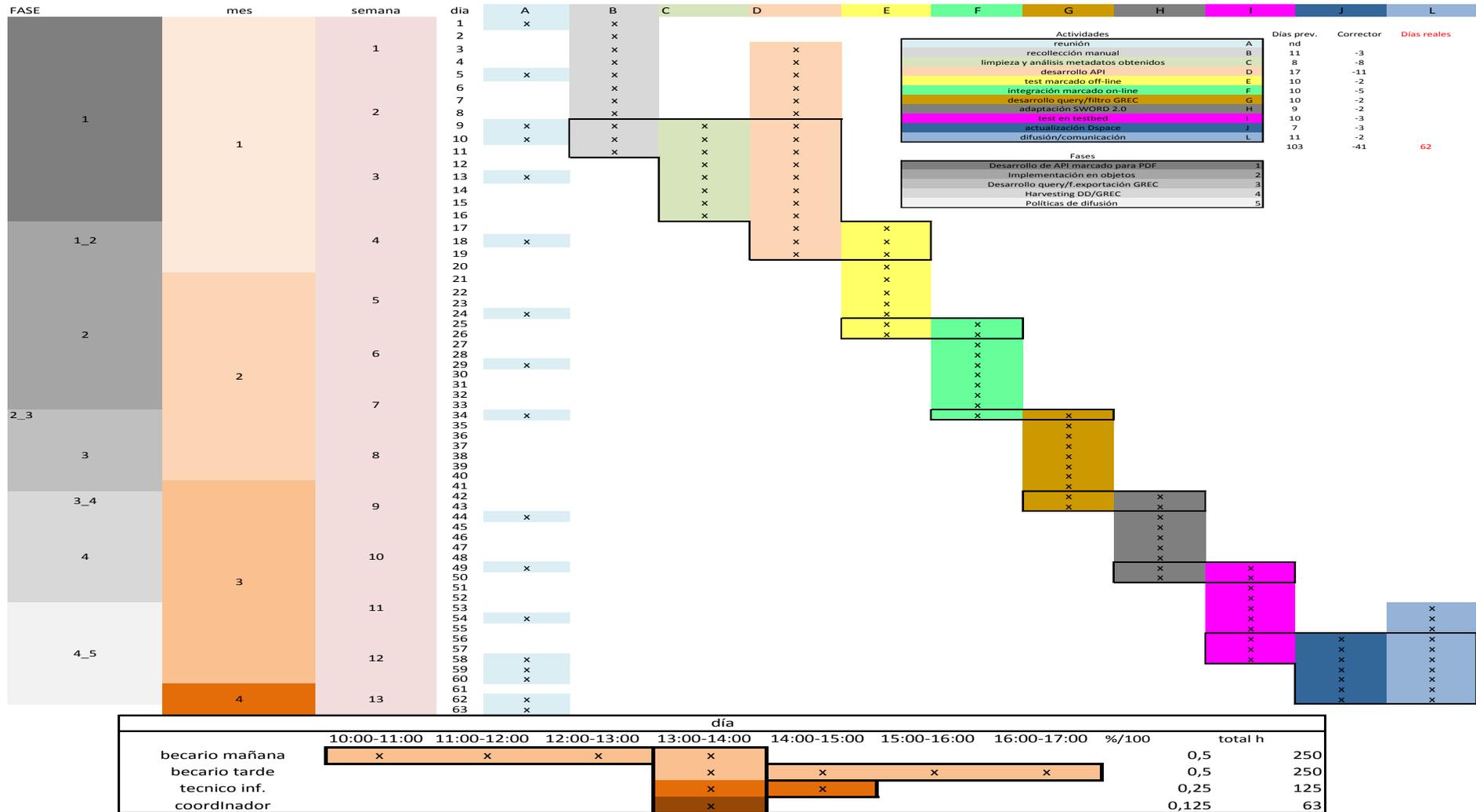
Sin embargo, estos detalles de costes no se consideran en este proyecto, debido a que no afectan directamente a la realización de la prueba piloto y la puesta en marcha del mercado de derechos de autor en el RI.

Tal como se ha explicado en el apartado anterior, los costes marginales del proyecto se refieren al capítulo de personal contratado con cargo al proyecto y, en cuanto a los costes de ejecución, debido a que se usa únicamente programario de libre distribución, sólo se presupuestan gastos para la difusión y comunicación de los resultados.

De acuerdo con el CRAI, se ha previsto exclusivamente la incorporación de los dos becarios de 20h semanales, ya que el coordinador del proyecto y el técnico informático serán designados por el CRAI dentro de su plantilla actual.

| | |
|---|-------|
| Sueldo bruto total estimado 2 becarios 20h x 14 semanas | 4200€ |
| Coste campaña de comunicación y difusión | 2500€ |
| Coste total del proyecto (IVA incluido) | 6700€ |

8. Ejecución y control



En este cronograma se presenta la organización de las fases identificadas para el proyecto y se detallan las distintas actividades. Cada casilla del cronograma representa un día. La superposición de actividades distintas en los mismos días (un factor de corrección en el cálculo de los días totales de duración) está prevista para entrelazar las actividades entre ellas. Como se puede ver, cada cambio de actividad coincide con una reunión de coordinación del equipo, que sirve para evaluar los elementos de la actividad que concluye, y preparar la actividad siguiente. Las horas totales del proyecto consideradas son 500, y se distribuyen entre el doble turno de los becarios en una jornada de siete horas. Con el primer turno, desde las 10 de la mañana hasta las 14 de la tarde, se resuelven los encargos previstos para el día, y en el segundo turno, que se superpone con la última hora del primero, empieza a las 13 y acaba a las 17. La coincidencia de los becarios se ha programado para preparar un entorno que permita el flujo constante de comunicación/colaboración entre los becarios. La organización programada para un día tipo pone de manifiesto que, a pesar de la escasa dedicación prevista para los miembros del CRAI (120 horas para el técnico y poco más de 60 para el coordinador del proyecto), el equipo tiene oportunidades diarias de trabajo conjunto, permitiendo la constante evaluación del desarrollo del proyecto tanto a nivel global, así como respecto al progreso en las distintas actividades previstas.

La capacidad de replicación del marcado incrustado y de las ventajas añadidas de su uso en un repositorio institucional, se probarán con estadísticas de acceso y uso, disponibles desde DSpace mismo, o creadas *ad hoc* para una evaluación específica del impacto obtenido.

La confrontación con datos históricos permitirá dimensionar la variación y el impacto obtenido con la aplicación de las buenas prácticas.

9. Resultados esperados

En las guías que recopilan las buenas prácticas sobre políticas de transición al acceso abierto, las vías verde y dorada representan una solución concreta, viable y efectiva, a pesar de ser varios los modelos que coexisten en la actualidad.

Las instituciones científicas que quieran implementar modelos alternativos de difusión del conocimiento y de los resultados de la investigación, tienen que sustentar todas esas acciones que amplifiquen la identificación de las obras respecto a los derechos de autor y a las políticas de gestión de los derechos de las revistas científicas.

Las personalizaciones propuestas en este proyecto influyen en la imagen que el DD proyecta en la red: poder contar con un sistema de gestión de los derechos de autor que garantice a los editores la conveniencia de permitir depositar en el DD, aumenta la masa crítica de objetos depositados, y hace del repositorio un nodo de divulgación más atractivo, tanto para los investigadores como para el público general.

La magnitud del beneficio que proporciona el acceso abierto, es también función del nivel de optimización que el depósito digital consiga en la divulgación de las libertades de uso concedidas por las licencias abiertas.

La optimización propuesta se compone de varios elementos:

- Si el número total de objetos depositados determina el coste per ítem, este valor disminuye al aumentar su volumen. Por lo tanto, se quiere aumentar el número total de objetos depositados, facilitando una nueva herramienta para la recuperación de artículos ya publicados.
- La introducción del “marcado incrustado” responde a una política de gestión de los derechos más eficaz, tanto para los usuarios humanos como para las máquinas. El beneficio para los autores es una mayor capacidad de difusión y recuperación de sus publicaciones desde una única central de confianza.
- La Universitat de Barcelona cuenta con la optimización de los resultados en los motores de búsqueda internacionales y con un repositorio representativo de todos los trabajos que se llevan a cabo en ella.

Con este sistema, se ofrece la posibilidad de tener un mecanismo a compartir con otras instituciones para marcar e indicar los derechos y, además, sería un buen ejemplo de respeto entre todas las partes implicadas: universidades, investigadores, agencias de financiación y editores.

10. Política de difusión de la guía

Los últimos días de proyecto se dedican a la preparación del material de divulgación sobre los resultados conseguidos.

La transferibilidad de los resultados es función del entorno de desarrollo informático elegido; para este proyecto se ha decidido utilizar código fuente libre, garantizando una condición que a su vez habilita la solución propuesta (se reutiliza el código existente) y garantiza la interoperabilidad de la solución encontrada (se puede copiar/modificar el código obtenido).

Los relativamente escasos días previstos en el plan de trabajo para la producción de material didáctico y divulgativo, se pueden justificar con el hecho de que será el mismo DD la fuente de representación del éxito proyectual.

Se prevé la realización de cursos de formación, tanto para el personal administrativo y de soporte, como para el personal docente e investigador.

Se preparan presentaciones de los resultados para su difusión desde la home page del DD y otras páginas de la web de la UB.

Además, la guía se pondrá a disposición en el mismo DD para el libre uso de la misma.

11. Referencias

- Adobe Systems Inc. XMP Specification. Public Patent License. Adobe Systems Incorporated. Recuperado de http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/xmp_public_patent_license.pdf
- Adobe Systems Inc. XMP Toolkit SDK Overview. Recuperado de <http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMP-Toolkit-SDK-Overview.pdf>
- Adobe Systems Inc. XMP™ Adding Intelligence to Media. (2012). XMP specification. Part 2. Additional properties. Adobe Systems Incorporated. Recuperado de <http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/cs6/XMPSpecificationPart2.pdf>
- Adobe Systems Inc. XMP™ Adding Intelligence to Media. (2012). XMP specification. Part 3 Storage in files. Adobe Systems Incorporated. Recuperado de <http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/cs6/XMPSpecificationPart3.pdf>
- CRAI-UB. (2012). La Universitat de Barcelona i l'accés obert: cap a una nova via de difusió. Universitat de Barcelona. CRAI. Recuperado de <http://hdl.handle.net/2445/22135>
- Creative Commons Corporation. The Power of Open. (2011). Recuperado de <http://thepowerofopen.org>
- Abadal, E. (2012). Acceso abierto a la ciencia. Barcelona: Editorial UOC. Recuperado de <http://eprints.rclis.org/bitstream/10760/16863/1/2012-acceso-abierto-epi-uoc-vfinal-autor.pdf>
- Abelson, H., Adida, B., Linksvayes, M., Yergler N. (2008). ccREL: The Creative Commons Rights Expression Language. Creative Commons . Recuperado de <http://www.w3.org/Submission/ccREL/>
- Adobe Systems Inc. XMP™ Adding Intelligence to Media. (2012). XMP specification. Part 1. Data model, serialization, and core properties. Adobe Systems Incorporated. Recuperado de <http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/cs6/XMPSpecificationPart1.pdf>
- Adobe Systems Incorporated (2008). Pdf 32000-1:2008. Document management — Portable document format — Part 1: PDF 1.7. 2008.
- Aliprandi, S., (2011). *Creative commons: a user guide*. Milano: Ledizioni .
- Allinson, J., François, S. & Lewis, S. (2008). SWORD: Simple Web-service Offering Repository Deposit. Ariadne: Web Magazine for Information Professionals, 54. Recuperado de <http://www.ariadne.ac.uk/issue54/allinson-et-al>

- Burgos Aguilar, J.V., Ramírez Montoya, M. S. (2011). Movilización de recursos educativos abiertos (REA): Enriqueciendo la práctica educativa. *La educación*, (146). Recuperado de http://educoas.org/portal/la_educacion_digital/146/pdf/EXPR_vladimirburgos_ES.pdf
- CAI-UB. Pla de millores dels serveis de biblioteca de la UB. Document aprovat en la reunió del CAI de 13 de març de 2007.
- Cassella, M., & Morando, M. (2012). Un quadro di competenze per i repository manager in Italia. *Biblioteche oggi*, 30(1), pp. 9-23. Recuperado de <http://www.bibliotecheoggi.it/content/n20120100901.pdf>
- Centre de recursos per a l'aprenentatge i la investigació (CRAI-UB). (2010). Memòria d'activitats de 2009. Universitat de Barcelona. Documento accesible en Dipòsit Digital de la UB : <http://hdl.handle.net/2445/15022>
- Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI-UB). (2011). Projecte ORACLE. Pla estratègic 2011-2014. Universitat de Barcelona. Recuperado de <http://www.bib.ub.edu/crai/pla-estrategic/>
- Centro ASK, Università Bocconi. (November 2011). *PEER Economics Report*. Università Commerciale Luigi Bocconi. Milano.
- De Filippo, D., Sanz-Casado, E., Urbano Salido, C., Ardanuy, J., & Gómez-Caridad, I. (2011). El papel de las bases de datos institucionales en el análisis de la actividad científica de las universidades. *Revista Española de Documentación Científica*, 34 (2), pp.165-189, ISSN: 0210-0614, doi: 10.3989/redc.2011.2.797
- Dulong de Rosnay, M., De Martin, J.C. (2012). *The digital public domain: Foundations for an Open Culture*. Open Book Publishers CIC Ltd. ISBN Digital (pdf): 978-1-906924-47-8
- Fitzgerald, A., Hooper, N. & Foong, (2011). *CC & Government Guide Using Creative Commons 3.0 Australia Licences on Government Copyright Materials*. Australia: Creative Commons.
- Frederick Friend, F.(2012). *IPR policy and scientific research knowledge exchange report for scientific policy makers. Knowledge Exchange*. Danish Agency for Culture.
- Guédon, J.C. (2009). It's a repository, it's a depository, it's an archive...: open access, digital collections and value. *ARBOR Ciencia, Pensamiento y Cultura*, CLXXXV 737 pp. 581-595. doi: 1039/arbor.2009.185.
- Guibault, L., Angelopoulos, C. (2011). *Open Content Licensing From Theory to Practice*. Amsterdam: Amsterdam University Press. e-ISBN 978 90 4851 408 3
- JISC ed. (2011). Clustering and Sustaining Digital Resources: the JISC eContent Programme 2009-11. JISC. Recuperado de <http://www.jisc.ac.uk/media/documents/publications/general/2011/JISCeContentClusteringAndSustainingDigitalResources.pdf>
- Lewis, S., Yates, C. (2008). Dspace Configuration de la Colección The Dspace Course. The Dspace Course.. DURASPACE. Recuperado de <http://cadair.aber.ac.uk/dspace/handle/2160/615>

Madden, M., Zickuhr, K. (2011). 65% of online adults use social networking. Pew Internet & American Life Project. Recuperado de <http://pewinternet.org/Reports/2011/Social-Networking-Sites.aspx>

Nichols, D. M., Paynter, G. W., Chan, C., Bainbridge, D., McKay, D., Twidale, M. B., & Blandford, A. (2009). Experiences in Deploying Metadata Analysis Tools for Institutional Repositories. *Cataloging & Classification Quarterly*, 47(3-4), pp. 229-248 Recuperado de <http://dx.doi.org/10.1080/01639370902737281>

Open Access Infrastructure for research in Europe.(2012) *How to make your journal OpenAIRE compliant*. Online workshops – May 7 and 8.

Raymond, L., Chandler, C., Lowry, R., Urban, E., Moncoiffe, G., Pissierssens, P...Miller,H. (2012). Emerging Role of Librarians in Data Publication. E-Science Symposium. Recuperado de http://escholarship.umassmed.edu/escience_symposium/2012/posters/5

Swan, A. (2012). *Policy Guidelines for the development and promotion of open access. The United Nations Educational, Scientific and Cultural Organization*. UNESCO. ISBN 978-92-3-001052-2

UB, Consell de Govern (18/02/2008). La Universitat de Barcelona i la lliure difusió del coneixement. Recuperado de http://www.bib.ub.edu/fileadmin/arxiu/diposits/liliure_difusio_coneixement.pdf

12. Anexos

12.1 Guía de ingesta

De seguida se indican las buenas prácticas identificada en relación al marcado de derechos de autor en repositorios institucionales:

1. Indicar siempre el `dc:rights`, ya sea en registro completo o el simplificado.
2. Si hay el campo `dc:rights`, tiene que aparecer siempre el logo en la página de metadatos del objeto (se encuentra en los tres ficheros que se incluyen en el proceso de envío).
3. Durante la colección de datos a través del *input-form* se compila el campo de los derechos por parte el autor o se completa por parte del validador encargado de la revisión.
4. En el *input-form* se proponen distintos campos para la atribución de la titularidad de los derechos de autor (autor, universidad, etc.), indicados con *dc:rights* y con *dc:rights:uri*. Este elemento permite la conexión directa (enlace) a las declaraciones sobre las políticas de derechos de autor.
5. Habilitar la colección de objetos a depositar a través de la ingesta automática con el software SWORD. Con este fin hay que incluir en el fichero METS y en el fichero de "interpretación" del Dspace la referencia al proceso de *swap-ingest*.
6. Utilizar siempre el marcado incrustado para ficheros PDF.
7. Incluir siempre el elemento `dc:rights.uri` que permite la referencia a la licencia con un enlace.
8. Implementar constantemente las revisiones actualizadas de las directivas DRIVER y OPENAire, con el fin de gestionar la accesibilidad de los objetos (con embargo o no).
9. Prestar constante atención a la colección de investigación con el fin de monitoriza.