

# LA GESTIÓN DE LA INFORMACIÓN EN EL PROYECTO ARQUEOLÓGICO EL CAÑO (PAEC)

ALFREDO FERNÁNDEZ-VALMAYOR

ANA F PAMPILLÓN

JULIA MAYO

MERCEDES GUINEA

CARLOS MAYO

MIGUEL A HERVÁS

JESÚS HERRERÍN

# LA GESTIÓN DE LA INFORMACIÓN EN EL PROYECTO ARQUEOLÓGICO EL CAÑO (PAEC)

## I. INTRODUCCIÓN

El objetivo de esta ponencia es presentar el sistema de gestión de la información utilizado en el Proyecto Arqueológico El Caño (PAEC) y compararlo con algunos sistemas similares. Técnicamente, el sistema que utilizamos en el PAEC es una aplicación web desarrollada con software abierto y tecnologías estándar. En este sistema se define la abstracción “objeto digital” como el conjunto de documentos que de alguna forma documentan un “objeto real” de interés para el proyecto. Además, un “objeto digital” se caracteriza mediante un conjunto de atributos nombrados explícitamente. El “objeto digital” y sus atributos conforman la unidad de información de la base de datos, o repositorio, que constituye el núcleo del sistema de gestión del PAEC. Este sistema recibe el nombre de ODA (Objetos Digitales de Aprendizaje) porque en su origen estuvo vinculado al desarrollo de entornos informatizados para la enseñanza (Fernández-Valmayor *et al.* 2011). La versión que aquí se presenta, ODA 2.5, fue desarrollada por los investigadores del PAEC, en colaboración con el grupo de investigación *Implementation of Language-Driven Software and Applications* (ILSA) de la Universidad Complutense de Madrid (UCM), y con la empresa *Varadero Software Factory*, esta última *spin-off* del citado grupo de investigación de la UCM. Desde el año 2014, la información y materiales del PAEC son gestionados utilizando ODA. En ODA se almacena, clasificada siguiendo criterios arqueológicos, toda la documentación generada por el proyecto.

El sistema de gestión del PAEC se basa en la construcción y mantenimiento de un repositorio de datos y documentos. Este repositorio almacena, clasificándola bajo múltiples perspectivas, toda la información relevante obtenida, o elaborada, por los investigadores. Su construcción implica tres pasos: 1) crear grupos con los documentos multimedia que describen, o estudian, algún aspecto de los “objetos reales” de nuestro interés y crear “objetos digitales” para cada uno de estos grupos, 2) asignar atributos a los “objetos digitales” para caracterizarlos, tanto informal como formalmente, y 3) construir de forma incremental el “modelo de datos” o “esquema de clasificación” del repositorio. Para ello, hay que establecer relaciones entre los atributos y el dominio y rango de cada atributo. Este proceso es incremental y, como veremos, se desarrolla al tiempo que se desarrolla la investigación. Este proceso tiene como objetivo llegar, mediante aproximaciones sucesivas, a obtener el modelo del dominio que es la referencia fundamental para analizar los datos de la sociedad que ocupó el sitio objeto de nuestro estudio. A continuación, detallamos estos tres pasos.

1) Objetos digitales. En ODA, un “objeto digital” representa y describe nuestro conocimiento sobre un “objeto real”. Inicialmente, un “objeto digital” es el conjunto de documentos digitales -archivos multimedia- generados por el estudio del “objeto real”. Si ponemos por caso una vasija arqueológica, el objeto real es el “objeto físico”, la vasija, y el objeto digital el conjunto de informes, fotografías y cualquier otro archivo multimedia que se haya generado a partir del estudio de esa vasija. Los objetos digitales también pueden representar y describir “objetos conceptuales”. Los objetos conceptuales son otro tipo de objetos reales y su caracterización requiere un mayor esfuerzo de abstracción. Por ejemplo, pueden ser objetos conceptuales las unidades estratigráficas identificadas en la excavación. En este caso el objeto digital estaría compuesto por el conjunto de croquis, fotografías y textos, que identifican dicha

unidad estratigráfica. Otros objetos conceptuales pueden ser, un conjunto de referencias bibliográficas, un recorrido docente o los episodios del ritual funerario. Como tantas veces ocurre con los elementos de los sistemas basados en el conocimiento entre “objeto físico” y “objeto conceptual” hay un continuo que tiene aspectos circunstanciales y subjetivos.

2) Atributos de los objetos digitales. Cada objeto digital se describe asignándole un conjunto de pares, “nombre de atributo: valor”. Al nominar los atributos ingresamos en el sistema la lista con las características más relevantes de los objetos. El conjunto de valores de estos atributos podemos verlos como las coordenadas que sitúan al objeto en el espacio conceptual de nuestra investigación. La posterior navegación en este espacio, las rutas del conocimiento, dependerán de la forma en que se relacionen entre sí estos atributos, o lo que es lo mismo, dependerán del “modelo de datos” del repositorio, que tratamos en el apartado siguiente.

Por ejemplo, para situar una vasija de cerámica en el espacio físico podemos usar estos tres pares: “alto mm: 180”, “diámetro mm: 520”, “cota del hallazgo m: -3.25”. Para situarla en un espacio conceptual (tipología cerámica) podemos usar los siguientes cuatro pares: “estilo cerámico: conte temprano”, “vajilla: conte rojo”, “forma: cuenco”, “variedad: hemisférico”.

Asignar a un objeto los pares atributo-valor, necesarios para describir sus características físicas suele ser inmediato, pero asignar los pares necesarios para describir adecuadamente sus características conceptuales (por ejemplo, los episodios de los ritos funerarios) requiere un esfuerzo de abstracción importante.

3) Modelo de datos o esquema de clasificación. ODA, como otros sistemas avanzados de gestión de la información, tiene un formalismo de representación que permite construir modelos basados en las características conceptuales del dominio (Rodríguez-Castro *et al.* 2010; Hearst 2006a, 2006b). Utilizando ODA podemos formalizar mediante atributos las características más importantes de los objetos digitales y construir con ellos una estructura que modele el dominio de conocimiento al que pertenecen los objetos. Esta estructura se construye tomando como base los atributos formalizados, las relaciones entre ellos, las propiedades de estas relaciones y finalmente los vocabularios controlados, es decir, el rango de valores que puede tomar cada atributo.

Es importante señalar que, con estos elementos, construir el modelo de datos de un dominio de conocimiento moderadamente complejo no es en absoluto una tarea sencilla “*The specification of a classification scheme is a difficult task, requiring considerable expertise as well as intellectual and manual effort*” (Giess *et al.* 2007: 3). No obstante, el esfuerzo es rentable ya que un modelo de datos apropiado es la base para hacer más intuitiva la navegación y el análisis de la información almacenada en el repositorio. Además, los usuarios pueden aprender sobre los temas y conceptos del dominio de conocimiento, mientras navegan por él, lo que hace del repositorio una herramienta para la educación muy importante.

ODA permite modificar la estructura del modelo de datos. En los sistemas de gestión de la información, las tareas de diseño y construcción del modelo de datos preceden a las tareas de poblar con datos el repositorio (Schwarzbach *et al.* 2014). Introducir posteriormente cambios en el modelo no suele estar previsto, o no es fácil. “*Of course, it is possible to change the database as the excavation progresses, but even quite minor changes can be time consuming and naturally restrictive. It is unlikely that the field archaeologists will want to contact the database technician every time they want to add a new criterion or field to the database.*” (Eve y Hunt 2008: 1).

Con ODA la situación es diferente, en cualquier etapa de la construcción del repositorio ODA es sencillo modificar y reconfigurar el modelo de datos. Es decir, a medida que la excavación progresa, se pueden ensayar distintos atributos y distintas relaciones entre atributos –distintas configuraciones del modelo– para adecuarlo mejor a nuestra investigación sobre el sitio arqueológico. Dicho de forma más general, para adecuarlo mejor al dominio de conocimiento y explicar mejor lo que representa el sitio arqueológico. El PAEC se ha beneficiado de estas características de ODA. La primera etapa de la construcción del repositorio del PAEC comenzó con un modelo de datos muy sencillo con unos pocos atributos relativos a las características físicas de los hallazgos. Después, como puede comprobarse accediendo al repositorio, el modelo ha evolucionado aun modelo más completo y complejo, y está previsto que esta evolución continúe para el modelo pueda mostrar nuevos aspectos de la cultura Coclé (Mayo, J. y Mayo, C. 2017).

## II. LA GESTIÓN DE LA INFORMACIÓN EN PROYECTOS ARQUEOLÓGICOS

En los proyectos arqueológicos, las operaciones de campo producen una gran cantidad de material y por ende de datos y documentos. El imprescindible trabajo de laboratorio multiplica el número de documentos y datos que los investigadores deben considerar en su trabajo. En el caso del trabajo de campo, esta documentación consiste en el registro de artefactos, restos humanos y contextos. En la actualidad en un proyecto arqueológico toda la documentación generada es digital, o digitalizable, lo que en principio debería significar que es manejable y accesible.

Entonces cabe preguntarse, como lo hicieron Hochstetter *et al.* (2011), porqué en general los datos y documentos de los proyectos arqueológicos no están disponibles en sistemas de información públicamente accesibles, o no lo están en la medida que sería posible, ya que al estar en soporte digital su difusión debería de ser relativamente sencilla. Estos autores señalaron tres factores que frenaban la difusión de los resultados de los proyectos arqueológicos, y también señalaron factores que deberían hacer cambiar la situación:

First [...] advocates of keeping data confidential and not generally available cite issues related to the preservation of the archaeological record [...] One alternative strategy is to make information about the record more public since knowledge and awareness contributes to stewardship.

The second factor [...] follows from the tradition of intellectual property [...] Open access to primary information do not preclude appropriate citation [...] Creative Commons provide an excellent means of releasing data that can be shared as long as proper attributions are made. [...]

A third consideration about the sharing of data is the technological problems associated with information dissemination [...] A number of recent developments have made the sharing of archeological information vastly simpler. [...] Technologically, we are now in an excellent position to make the archaeological record [...] freely available for study to [...] researches Worldwide, and to the general public. (Hochstetter et al. 2011: 388-389)

Ciertamente, esta ha sido la evolución de muchos proyectos arqueológicos (Eve y Hunt 2008; Schulze *et al.* 2012; Geser, y Gonzalez-Perez 2017), aunque también debemos señalar que persiste una relativa escasez de repositorios públicamente accesibles, o lo que sería mas exacto, una relativa escasez de sitios web que den acceso a bases de datos debidamente mantenidas.

En nuestra opinión, la situación presente se debe a motivos que van más allá de los enumerados por Hochstetter *et al.* en 2011. Estos motivos están relacionados con la dificultad y el coste del trabajo de gestión y difusión de la información arqueológica en sí mismos. En efecto, los proyectos arqueológicos son proyectos complejos. Proyectos en los que a veces se subestima la dificultad de gestionar información heterogénea y difundirla en un sitio web bien organizado. A esto hay que añadir que no todos los proyectos arqueológicos destinan recursos suficientes para crear y mantener una buena base de datos. La consecuencia es que la calidad y accesibilidad de la información disponible *online* puede no ser la óptima.

Finalmente, es importante señalar que no hay que confundir los objetivos de un sistema de gestión de la información y acceso *online* a los resultados inmediatos de un proyecto, con las publicaciones científicas en revistas especializadas y congresos. Éstas últimas tienen una escala de tiempo completamente diferente y también un contenido y un objetivo distintos.

### III. REQUISITOS Y DISEÑO DE LA APLICACIÓN ODA

Para diseñar ODA tuvimos en cuenta los requisitos identificados en el Proyecto Arqueológico El Caño y las propuestas formuladas en los proyectos: EISP, *Easter Island Statue Project* (Hochstetter *et al.* 2011); ARK, *Archaeological Recording Kit* (Eve y Hunt 2008); OpenInfRA, *Open Information System for Research in Archeology* (Schulze *et al.* 2012) y ARIADNE, *Advanced Research Infrastructure for Archaeological Dataset Networking in Europe* (Geser, y Gonzalez-Perez 2017).

Los requisitos señalados por estos proyectos coinciden básicamente en lo que se refiere a la recogida de datos en el campo, en el laboratorio y a la gestión de la información arqueológica. Pero, en temas de interoperabilidad, integración de aplicaciones y servicios web especializados, las exigencias son muy distintas. Esta variabilidad depende de los recursos materiales disponibles en cada proyecto. Siendo la razón para excluir todos, o algunos, de los objetivos su elevado coste de desarrollo y mantenimiento. Esto está ligado en general a complejas arquitecturas y estándares, y al esfuerzo requerido para lograr la interoperabilidad e integración efectiva de aplicaciones (Henze *et al.* 2013).

Para construir ODA se utilizó como punto de partida *Chasqui* (Guinea 2004). Una aplicación basada en tecnologías estándar y software libre, que fue desarrollada para informatizar varios museos docentes de la Universidad Complutense de Madrid (UCM). El objetivo era incorporar los museos a la actividad docente, en conexión con el campus virtual de la UCM (Sierra *et al.* 2006; Guinea *et al.* 2009).

La decisión más importante del diseño del sistema ODA es la de no modelar una base de datos para un tipo de proyecto específico (este fue el problema de *Chasqui*), si no, una base de datos genérica que almacena tanto los datos de la excavación como el modelo de datos del dominio (podríamos decir que el modelo del dominio es en sí mismo un resultado de la investigación). Además, se hace del “objeto digital” la metáfora básica para organizar y agrupar los documentos del proyecto. Es decir, la actividad básica trata de identificar la documentación que corresponde a los objetos, físicos o conceptuales, que son objetivo del proyecto y las relaciones entre ellos. En el diagrama UML (*Universal Modelling Language*) (Fig.1) puede verse como entorno a esta abstracción central se sitúan cuatro componentes:

1) Base de datos documental. En una primera aproximación, ODA es simplemente un repositorio de “objetos digitales” y cada objeto digital es un “contenedor” con todos los documentos multimedia que informan sobre un objeto real previamente identificado. Cada uno de los archivos digitales que contiene un objeto digital se considera un “documento-propio” del

objeto si éste se refiere directamente al objeto. Pero los documentos que son parte de un objeto digital también pueden ser “una referencia” a otra documentación no propiamente del objeto pero que, por alguna razón, es importante incorporar al objeto digital. Estas referencias deben considerarse tanto en el plano semántico como en el plano sintáctico. En el primero, y dependiendo de los objetos y documentos relacionados, pueden ser referencias *part-of*, *is-a*, *instance-of*, etc., según la semántica que queramos atribuir al modelo de datos del repositorio. En general los objetos con características más conceptuales se relacionarán con los objetos de los que obtienen su significado. En el segundo, estas referencias pueden ser de tres tipos: a) a un “documento-propio” de otro objeto del repositorio; b) a otro objeto digital del repositorio (el contenedor y todos sus documentos) y c) al contenido de una *url* externa a ODA.

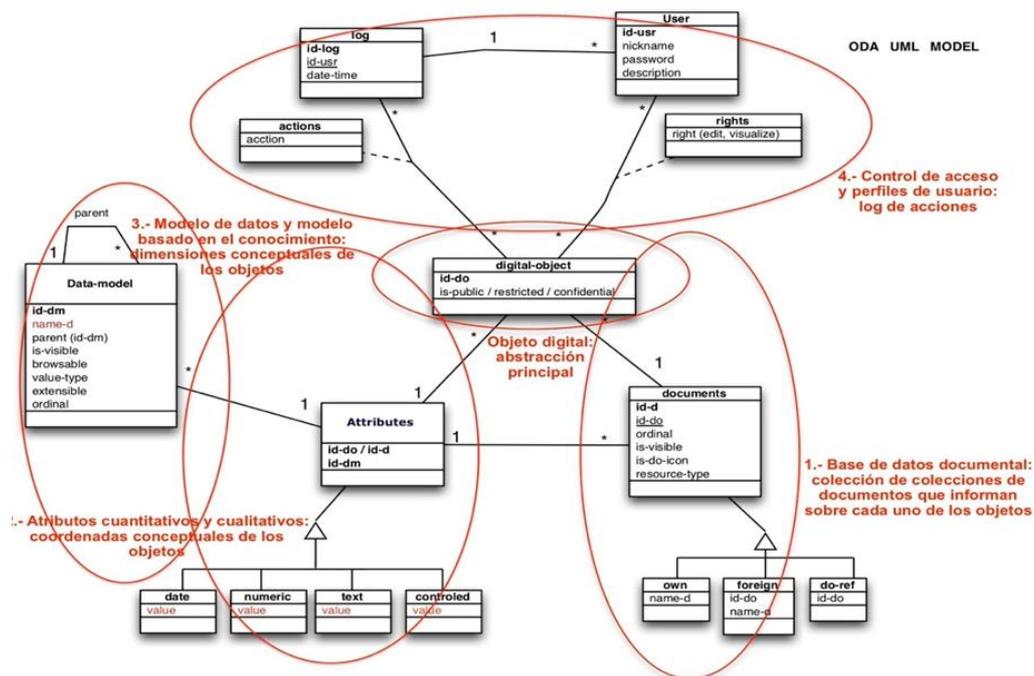


Fig. 1: Diseño de ODA en UML (Universal Modelling Language)

2) Atributos. Además de ser un conjunto de documentos, un objeto digital se caracteriza mediante un conjunto de pares atributo-valor. En un primer nivel semántico, ODA particiona el conjunto de atributos de un objeto digital en dos clases: “datos”, aquellos atributos que definen una propiedad del objeto y “metadatos” (datos sobre los datos), aquellos que expresan las circunstancias en las que se obtienen los datos.

Desde el punto de vista del tipo de valor que puede tomar un atributo éste puede ser:

- Texto libre. Permite describir informalmente las características de un objeto y cualquier otro tipo de circunstancia especial del mismo.
- Numérico o fecha. Son datos objetivos del objeto.
- Controlado. En ODA, la construcción de un modelo de datos formal se basa fundamentalmente en este tipo de atributos. El rango de valores que pueden tomar estos atributos se elige entre los elementos de un “vocabulario controlado”, normalmente una

lista de términos, o coordenadas cualitativas, que permiten situar al objeto con respecto a una dimensión (una propiedad significativa) de nuestro dominio de conocimiento. En palabras de la Oficina Nacional de Estándares de la Información USA:

[...] controlled vocabularies that are used for the representation of content objects in knowledge organization systems including lists, synonym rings, taxonomies, and thesauri. [...]

The primary purpose of vocabulary control is to achieve consistency in the description of content objects and to facilitate retrieval” (ANSI/NISO Z39.19-2005, R2010: i - 1).

En la literatura los atributos reciben el nombre de “facetas” y sus valores “términos”, si son mutuamente excluyentes y exhaustivos respecto al conjunto de objetos que queremos registrar en el repositorio (Denton 2003; Vickery 2008). En este caso podemos decir que cada objeto digital ocupará una “posición” en un espacio conceptual definido por facetas y tendrá como coordenadas los términos de esas facetas.

3) Modelo de datos. El modelo de datos, o esquema de clasificación, es la estructura que soporta la organización de los datos del repositorio y que, por reproducir la estructura del dominio, debe facilitar una gestión más intuitiva y eficiente de la información almacenada en él (Giess *et al.* 2007). Normalmente, el diseño de un repositorio de datos comienza por la construcción del modelo de datos (Schwarzbach *et al.* 2014). Por ejemplo, cuando se diseña una base de datos siguiendo el esquema “entidad-relación”, todas las tablas, sus campos, tipos y las relaciones entre tablas que componen el modelo, deben estar completamente definidos antes de comenzar a poblar la base de datos.

Utilizando ODA el planteamiento es distinto. La construcción del modelo de datos puede realizarse de forma incremental. En el esquema UML de la Fig.1 puede verse que esto es posible porque ODA utiliza una base de datos genérica. ODA emplea el formalismo relacional para construir un sistema genérico en el que los investigadores pueden añadir elementos o cambiar la estructura del modelo de datos del dominio igual que añaden o cambian datos arqueológicos. Así pueden obtenerse, en una aproximación por etapas, resultados similares a los diseños de modelos facetados basados en el conocimiento propuestos por Broughton (2013), o a las *faceted lightweight ontologies* propuestas por Giunchiglia *et al.* (2009).

El tipo de modelo a construir utilizando ODA es una decisión de los investigadores del proyecto. La aproximación seguida en el PAEC combina el modelo propuesto por Denton, basado en facetas, con el propuesto por Hearst que incorpora a las facetas múltiples jerarquías.

Según Denton los modelos de clasificación facetada se pueden entender como

“[...] a set of mutually exclusive and jointly exhaustive categories, each made by isolating one perspective on the items (a facet), that combine to completely describe all the objects in question, and which users can use, by searching and browsing, to find what they need” (Denton 2003).

Hearst *et al.* propusieron ampliar estos modelos de clasificación facetada definiendo múltiples jerarquías de categorías “*Hierarchical Faceted Categories*” (HFC), cada una de las cuales desarrolla una faceta en sucesivas sub-facetas. En esta aproximación podemos considerar que cada jerarquía que describe una perspectiva, o subespacio conceptual, dentro del espacio total. “*Rather than creating one large category hierarchy, build a set of category hierarchies each of which corresponds to a different facet (dimension or feature type) relevant to the collection to be navigated.*” (Hearst *et al.* 2006: 86).

Giunchiglia propone construir ontologías descriptivas estableciendo relaciones semánticas entre facetas *“facets and terms have real world semantics, namely, they are descriptive ontologies which are formed using part-of, is-a and instance-of relations”* (Giunchiglia et al. 2009: 10).

El modelo de datos construido para el repositorio del PAEC se ha desarrollado siguiendo la propuesta HFC en etapas sucesivas. Añadiendo en cada etapa nuevas jerarquías-taxonomías representando nuevos enfoques o perspectivas de la investigación sobre la sociedad Coclé. En la Fig.2 se puede seguir la idea general de este desarrollo en tres etapas.

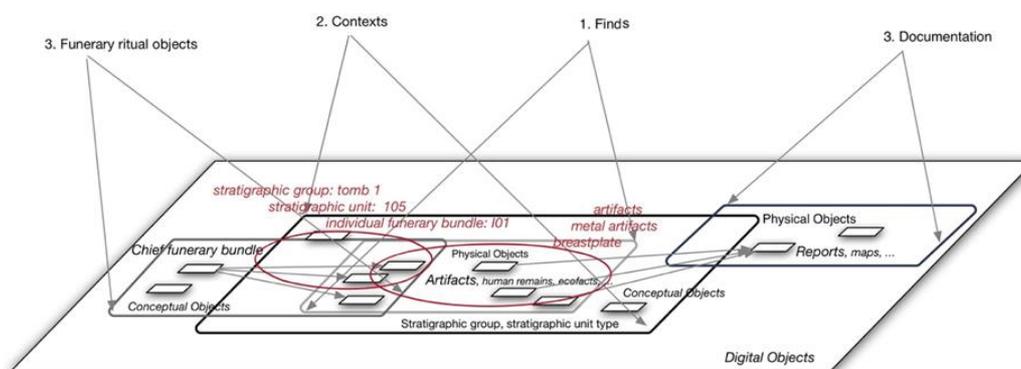


Fig. 2 : Jerarquías desarrolladas hasta el momento en el repositorio del PAEC

En la Fig.2 se puede ver gráficamente que el conjunto de objetos digitales indexados en la etapa 2 *Contexts* subsume al conjunto indexado en la etapa 1 *Finds*. Un mismo objeto, por ejemplo, un pectoral de oro se encuentra en la intersección de dos conjuntos, definidos, el primero siguiendo la taxonomía de hallazgos (un pectoral es un elemento de una clase de hallazgos) y el segundo siguiendo la taxonomía de contextos (un pectoral es parte del ajuar funerario de un individuo). En la Fig.3a, b se puede ver el mismo ejemplo, pero en este caso se navega hasta el objeto utilizando el interfaz de ODA, en el caso 3a) navegando desde hallazgos, y en el caso 3b) desde contextos. También se puede ver como un objeto conceptual *Chief funerary bundle* adquiere su significado mediante su relación con objetos físicos en *Finds* y *Contexts*.

4) Control de acceso y perfiles de usuario. ODA permite establecer diversos niveles de control en el acceso a objetos, documentos y atributos. En el diseño UML de la Fig.1 puede distinguirse, por un lado, el registro de usuarios y sus derechos de acceso, y por otro, la bitácora que registra las acciones realizadas por los usuarios con permiso para crear/modificar/borrar objetos (autoría). Así mismo se indican los tres niveles de acceso de los objetos: públicos, restringidos y confidenciales.

Estos niveles de acceso dan lugar a cuatro perfiles de usuario: 1) Visitante o usuario-final, que puede acceder a los objetos públicos y dentro de éstos a sus documentos visibles. 2) Usuario-registrado. Los usuarios registrados en el sistema pueden acceder a un conjunto adicional de objetos que no es accesible a los visitantes. 3) Administradores. Tienen los mismos privilegios de acceso que los usuarios registrados, pero además pueden crear nuevos objetos y editarlos. 4) Superadministradores. Pueden autorizar o dar de baja a cualquier tipo de usuario del sistema. Pueden crear y editar cualquier objeto del repositorio incluyendo un tipo de objeto privado y un

tipo de atributo que solo es accesible para los superadministradores. Por encima de estos cuatro niveles de usuario estaría el superusuario del servidor anfitrión que tendría acceso directo, y por tanto todo el control, de la base de datos de ODA (Fig.6).

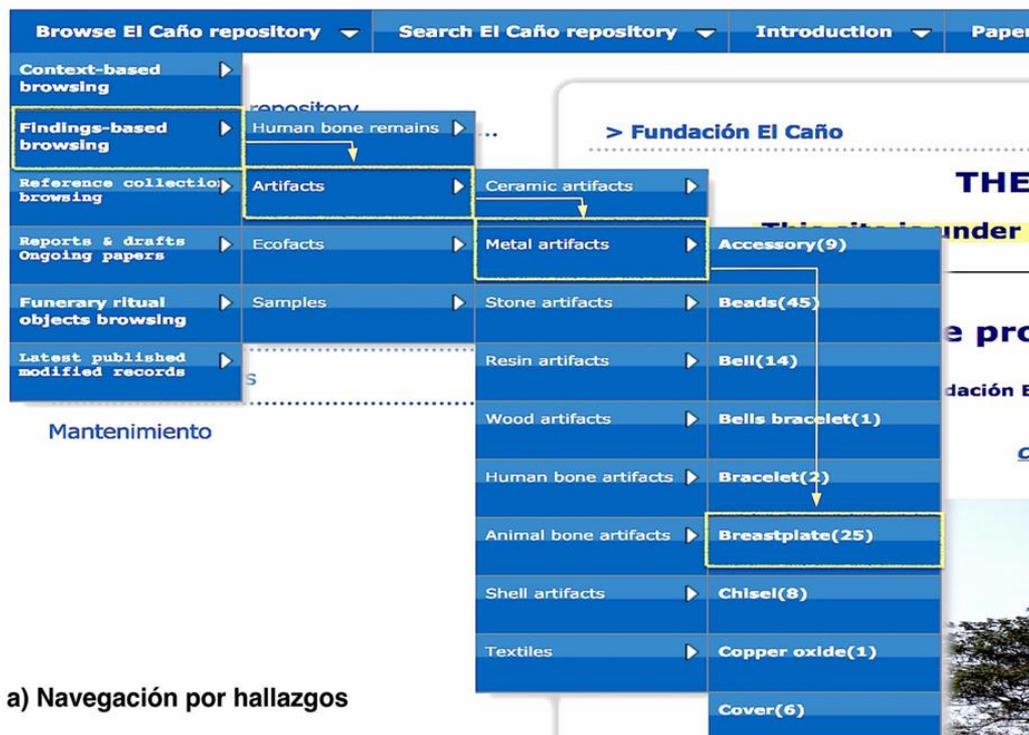


Fig. 3a: Navegación hasta un pectoral, a) desde hallazgos, como un elemento del conjunto de 25 pectorales registrados en el repositorio

#### IV. ARQUITECTURA E IMPLEMENTACIÓN

La arquitectura de ODA tiene como núcleo una base de datos relacional estándar con la que materializamos el diseño UML visto en el apartado anterior. Básicamente, cada elemento del diseño UML se corresponde con una tabla en el modelo relacional (Fig.4). Para obtener el modelo relacional, se simplifican algunos elementos del diseño UML, como por ejemplo los tipos de documentos que se incluyen en la tabla *documents*.

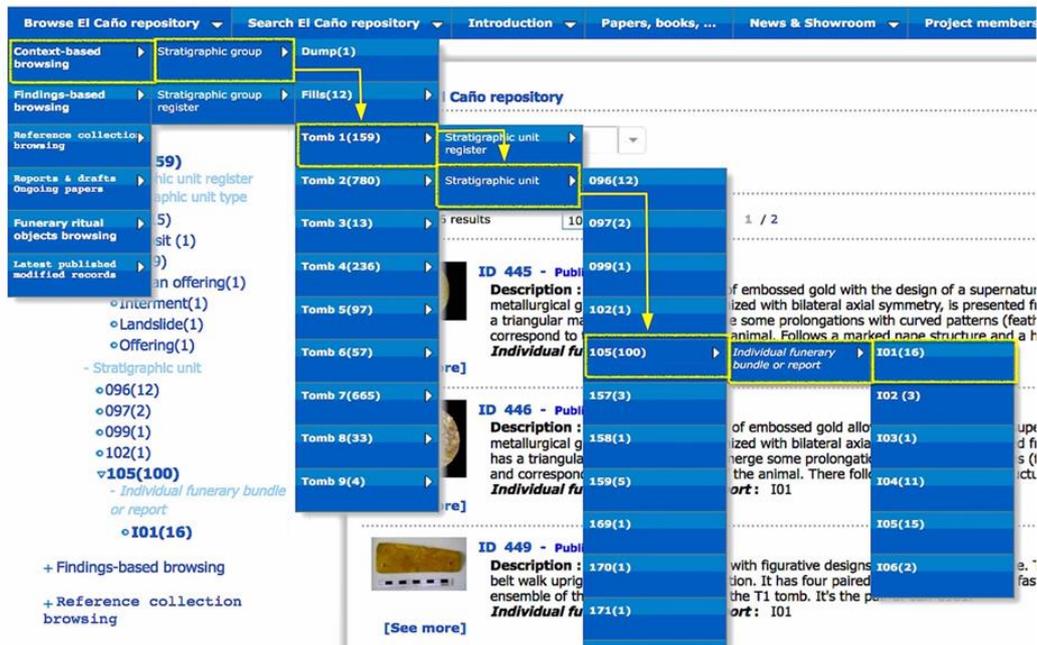
En la Fig.5 se detalla el diseño E-R y las restricciones que las claves externas imponen sobre las tablas. En la implementación del modelo, no se impone la integridad referencial en las relaciones para facilitar el mantenimiento de las tablas. La programación del interfaz verifica las simplificaciones introducidas. Todo este núcleo está dentro de una aplicación web que sirve de interfaz con la base de datos y que no se detalla en este documento.

La implementación de ODA tiene tres componentes principales (Fig.6)

- La base de datos relacional con los datos del dominio y los datos del modelo de datos (Fig.5). Esta base de datos se gestiona utilizando el interfaz ODA, pero es importante

destacar que, como se señala en la Fig.6, el superusuario del servidor anfitrión puede gestionarla directamente y sin restricciones utilizando el servidor de base de datos.

- La carpeta del sistema de archivos del sistema operativo con las carpetas cuyo nombre coincide con el identificador de un objeto digital y que contienen los archivos propios de cada objeto. Como en el caso anterior, los objetos se gestionan utilizando ODA, pero el superusuario del servidor puede acceder a ellos directamente.



b) Navegación por contextos

Fig. 3b: Navegación hasta un pectoral, b) desde contextos. El pectoral es un objeto del conjunto de 16 objetos del ajuar funerario del individuo I01

- La carpeta del sistema operativo con todos los documentos de la aplicación web que sirve como interfaz del sistema.

En nuestra implementación utilizamos un servidor de base de datos relacional (MySQL), un servidor web (Apache) con extensión PHP para interpretar el interfaz ODA con la base de datos relacional y con el sistema de archivos del OS (UNIX). Se han probado otros servicios (Java *plugins*) para exportar/importar la información en otros formatos (Gayoso-Cabada *et al.* 2016).

La implementación de ODA está disponible en la plataforma GitHub: <https://github.com/ILSA-UCM/OdA>.

## V. ODA EN EL CONTEXTO DE SISTEMAS SIMILARES. TRABAJO FUTURO.

Es frecuente que los sitios web, que muestran colecciones arqueológicas importantes, utilicen, como hace ODA, jerarquías de facetas. Con estos sistemas compartimos una estrategia que hace más intuitiva la navegación y facilita que el visitante pueda explorar las colecciones siguiendo el hilo conceptual marcado por alguna de las jerarquías de facetas (Hearst 2006a). Por

ejemplo, esta técnica se utiliza en las colecciones *online* del *Penn Museum*, del *American Museum of Natural History*, o del *Metropolitan Museum of Art* (accedidos en mayo 2018).

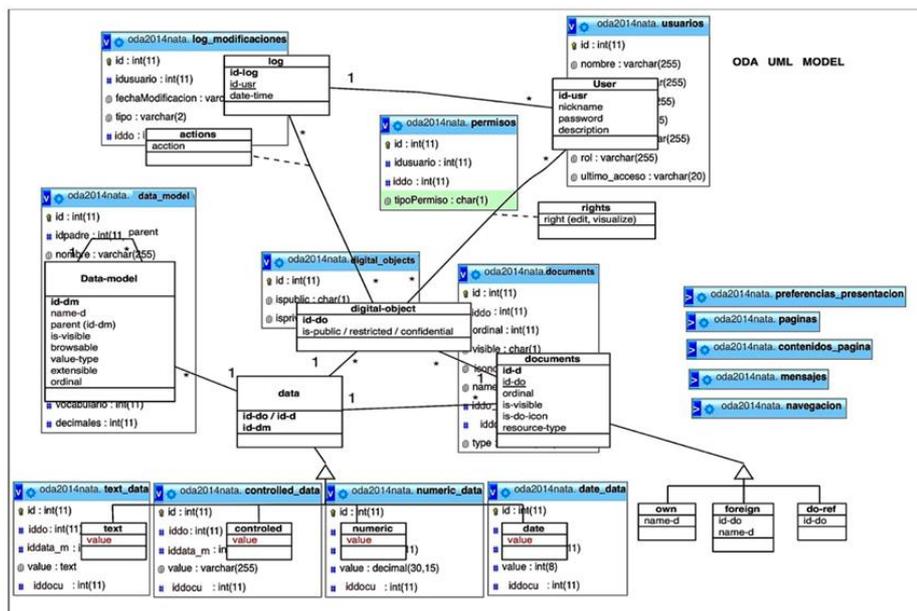


Fig. 4. Del modelo UML al modelo Entidad-Relación (E-R)

Los sitios web que muestran el trabajo de un proyecto de excavación son pocos, pero un caso paradigmático es la empresa *DigVentures* del Reino Unido. En sus excavaciones *DigVentures* utiliza el sistema *Digital Dig Team* que permite poner *online* de forma inmediata los datos obtenidos en la excavación. El visitante puede navegar por los registros de los hallazgos más recientes. La flexibilidad para crear registros es comparable a la de ODA. El motor responsable de esta funcionalidad es el sistema ARK, que luego veremos, y que actúa como *back-end* del sistema *Digital Dig Team*. Un ejemplo del trabajo de esta empresa es el proyecto *Borrowed time* (accedido en mayo 2018): la excavación de un cementerio de la Edad del Bronce en el Noroeste de Inglaterra.

Para valorar ODA en profundidad, la referencia más importante no son los sitios web sino los proyectos que incluyen entre sus objetivos el desarrollo de sistemas de gestión arqueológicos. Entre los proyectos más avanzados, destacan los dos siguientes:

- *Archaeological Recording Kit* (ARK 2018), desarrollado por *L-P: Archaeology*, una empresa de iniciativa privada del Reino Unido pero que colabora con numerosas universidades e instituciones públicas (Eve y Hunt 2008).
- *Open Information System for Research in Archeology* (OpenInfRA 2018) desarrollado por un consorcio dirigido por el *Deutsches Archäologisches Institut* (DAI) y en el que participan la *Brandenburg University of Technology Cottbus* y la *Dresden University of Applied Sciences* (Schulze et al. 2012; Henze et al. 2013; Schwarzbach et al. 2014)

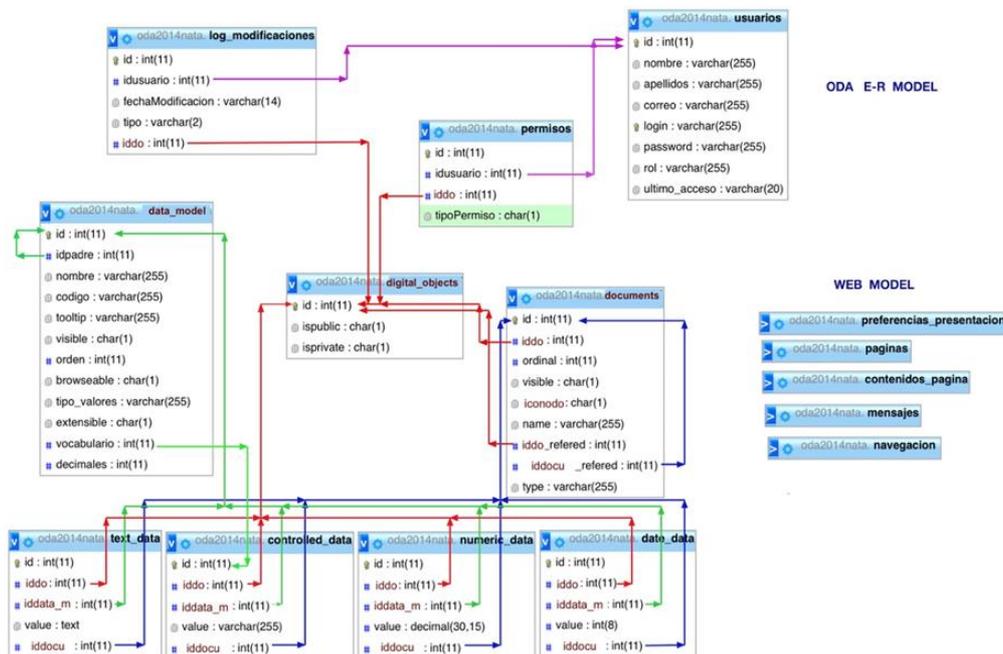


Fig 5 El modelo E-R de ODA

Los sistemas de gestión desarrollados en ambos proyectos optan, como ODA, por utilizar una base de datos relacional con un modelo genérico. Modelo genérico basado en el concepto de “hipertexto” en el caso de ARK, o bien en la abstracción “tema arqueológico” y sus “atributos”, en el caso de OpenInfRA.

Comparando los tres proyectos, podemos concluir que:

- ODA está en línea con los proyectos más avanzados, pero además ofrece ventajas derivadas de la simplicidad y expresividad del modelo genérico que utiliza. Un modelo basado en la abstracción “objeto digital” físico o conceptual y en la capacidad descriptiva de las jerarquías de facetas.
- ODA organiza la navegación mediante jerarquías de facetas como parte de una estrategia más amplia: construir un modelo del dominio que refleje nuestra interpretación de la sociedad que ocupó el sitio. Algunos de los museos más importantes utilizan también jerarquías de facetas para exponer sus colecciones de objetos arqueológicos, pero al no tener explícito un modelo del dominio, la navegación está más restringida.
- ODA tiene un objetivo central que es hacer explícito el modelo del dominio y utilizarlo para explorar y analizar el contenido del repositorio. En principio, esta aproximación, también sería posible en los otros sistemas, aunque más costoso para los investigadores, dado que el modelo genérico que utilizan es más complejo, y está diseñado para satisfacer otros objetivos: producción de textos a partir del hipertexto en el caso de ARK e interoperabilidad de bases de datos en el caso de OpenInfRA.

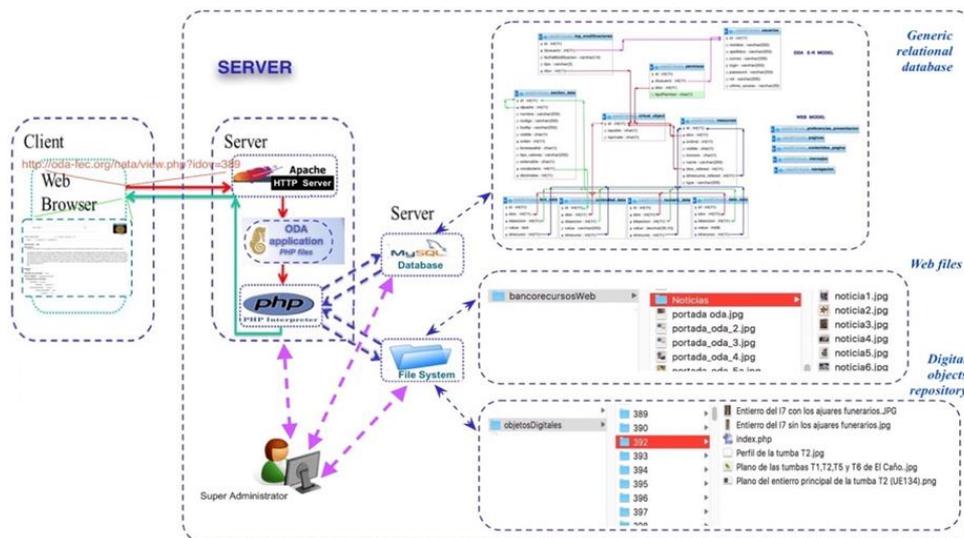


Fig. 6 Esquema de la arquitectura e implementación de ODA

Finalmente enumeramos algunas de las líneas de trabajo pendientes de desarrollar (con ODA y en ODA):

- Trabajos que afectan a la información que contiene el sistema.
  1. Crear recorridos docentes y de apoyo al museo del sitio.
  2. Versión en inglés. Protocolo de sincronización.
  3. Integrar información de otros yacimientos de la cuenca de Río Grande de Coclé, especialmente de Sitio Conte. Llegar a un acuerdo de colaboración con Los museos que tienen esas piezas en sus colecciones.
- Trabajos de refactorización y nuevo desarrollo.
  1. Protocolos de mantenimiento, gestión directa del servidor SQL.
  2. Mejoras en seguridad, usabilidad y mantenimiento del interfaz ODA
  3. Funciones de exportación e importación de datos.
  4. Integrar algunos servicios de aplicaciones externas.

## VI. BIBLIOGRAFÍA

American Museum of Natural History (2018): Collections. En <https://www.amnh.org/our-research/anthropology/collections>. Último acceso mayo 2018

ARIADNE (2018): Infrastructure. En <http://www.ariadne-infrastructure.eu>. Último acceso abril 2018

ARK (2018): An open source solution to project recording. En <https://ark.lparcology.com>. Último acceso abril 2018

DigVentures (2018): Finds at Borrowed Time. En <https://digventures.com/borrowed-time/ddt/browser.php>. Último acceso mayo 2018

Broughton, Vanda (2013): “Faceted Classification as a General Theory for Knowledge Organization”. En: *SRELS Journal of Information Management*, Volume 50, Issue 6, December 2013

Denton, W. (2003): “How to make a faceted classification and put it on the web”. En <http://www.miskatonic.org/library/facet-web-howto.html>. Último acceso enero 2018

Eve, S./ Hunt G. (2008): “ARK: A Development Framework for Archaeological Recording”. En: Posluschny, A./ K. Lambers / I. Herzog (eds.), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology. CAA 2007*, Berlin, Germany, April 2–6, 2007

Fernández-Valmayor, Alfredo/ Fernández-Pampillón, Ana María/ Fernández-Chamizo, Carmen/ Navarro, Antonio/ Cristóbal, Jesús. (2011): “Implantación de un Campus Virtual de Grandes Dimensiones: el Campus Virtual de la UCM”. En: *IEEE RITA*, 6(4), pp. 167-174.

Gayoso-Cabada, Joaquín/ Daniel Rodríguez-Cerezo/ José Luis Sierra. (2016): “Browsing Digital Collections with Reconfigurable Faceted Thesauri”. En: *25th. International Conference on Information Systems Development. ISD2016*. Poland.

Geser, G./ Gonzalez-Perez, C. (2017): “Report on E-Archaeology Frameworks and Experiments”. En <http://ariadne-infrastructure.eu/ger/Media/Files/D17.1-Report-on-e-Archaeology-Framework-and-Experiments>. Último acceso marzo 2018

Giess, M.D./ Wild, P.J./ McMahon C.A. (2007): “The Use of Faceted Classification in the Organisation of Engineering Design Documents”. En: *International Conference on Engineering Design, ICED'07*. 28 - 31 August 2007, Cite des Sciences et de L'industrie, Paris, France.

Giunchiglia, F./ Dutta, B./ Maltese, V. (2009): “Faceted lightweight ontologies”. Technical Report # DISI-09-022 también en: Borgida, A./ Chaudhri, V.K./ Giorgini, P./Yu, E.S.K., (eds.): *Conceptual Modeling: Foundations and Applications*. Volume 5600 of Lecture Notes in Computer Science., Springer (2009) 36–51

Guinea, M. (2004): “El proyecto Chasqui”. En: Alfredo Fernández-Valmayor Crespo/ Ana Fernández-Pampillón Cesteros/ Jorge Merino Granizo (eds.). *En apoyo del aprendizaje en la universidad: hacia el espacio europeo de educación superior*. Madrid. Editorial Complutense. ISBN 84-7491-774-3, págs. 228-233

Guinea, M./ Avila, D./ García, N. (2009): “Construcción de objetos de aprendizaje y desarrollo de experiencias docentes”. En: Alfredo Fernández-Valmayor Crespo/ Amelia Sanz Cabrerizo/ Jorge Merino Granizo (eds.). *Buenas prácticas e indicios de calidad*. Madrid. Editorial Complutense. ISBN 978-84-7491-968-4, págs. 8-13

Hearst/ Smalley/ Chandler (2006): Hierarchical Faceted Categories. En: *CHI 2006 85*

Hearst, M. A. (2006a): Design recommendations for hierarchical faceted search interfaces. En: *Proc. SIGIR 2006, Workshop on Faceted Search*. pp. 26--30.

Hearst, Marti A. (2006b): Clustering versus Faceted Categories for Information Exploration. En: *Communications of the ACM*, April 2006

Henze, F./ Magdalinski N./ Schwarzbach, F./ Schulze, A./ Gerth, Ph./ Schäfer, F. (2013): Concepts and Technologies for a Comprehensive Information System for Historical

Research and Heritage Documentation. En: *XXIV International CIPA Symposium*, 2 – 6 September 2013, Strasbourg, France.

Hochstetter, F./ Haoa, S./ Lipo, C./ Hunt, T. (2011): A Public Database of Archaeological Resources on Easter Island (Rapa Nui) Using Google Earth. En: *Latin American Antiquity*, 22(3), 385-397. Society for American Archaeology.

Lawson, A./ Eklund, P. W./ Goodall, P./ Wray, T./ Daniel, V./ Van Olffen, M. (2010): Designing a digital ecosystem for the new museum environment: The Virtual Museum of the Pacific. En: H. Yeatman (Eds.). *The SInet 2010 eBook* (pp. 227-239). Wollongong: SInet UOW.

Martini, R.G./ Araujo, C./ Libelotto, G.R./ Henriques, P.R. (2016): A Reduced CRM-Compatible Form Ontology for the Virtual Emigration Museum. En: Rocha Á./ Correia A., Adeli H./ Reis L./ Mendonça Teixeira M. (eds.) *New Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing*. Vol. 444. Springer, Cham

Mayo, J./ Mayo, C. (2017): Repositorio de datos del Proyecto Arqueológico El Caño. Centro de Investigaciones Arqueológicas del Istmo. En <http://oda-fec.org/nata>. Último acceso mayo 2018

Metropolitan Museum of Art (2018): Collection. En: <https://www.metmuseum.org/art/collection/search>. Último acceso abril 2018

NISO Z39.19-2005 (R2010) (2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. *National Information Standards Organization NISO (U.S.) 172nd ed.*, (Bethesda, Md): NISO Press.

OpenInfRA (2018): En <https://www.b-tu.de/openinfra/>. Último acceso abril 2018

En <https://fallback.dainst.org/projekt/-/project-display/113484>. Último acceso abril 2018

Penn Museum, University of Pennsylvania Museum of Archaeology and Anthropology (2018). En: <https://www.penn.museum>. Último acceso abril 2018

Rodríguez-Castro, B./ Glaser, H./ Carr, L. (2010): How to reuse a faceted classification and put it on the semantic web. En: Peter F./ Patel-Schneider/ Yue Pan/ Pascal Hitzler/ Peter Mika/ Lei Zhang (Eds.), *Proceedings of the 9th international semantic web conference on the semantic web*. Volume Part I (ISWC'10), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 663-678.

Schulze, A./ Schäfer, F.F./ Gehrt, P./ Henze, F./ Schwarzbach, F.: (2012): OpeninfRA – Storing and Retrieving Information in a Heterogeneous Documentation System. En: *Proceedings of the CAA conference 2012*, University of Southampton.

Schwarzbach, F./ Schäfer, F./ Schulze, A. (2014): Find the Balance - Modelling Aspects in Archaeological Information Systems. En: F. Giligny/ F. Djindjian/ L. Costa/ P. Moscati/ S. Robert (eds.). *Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology CAA 2014 - 21st Century Archaeology*

Sierra, J./ Fernández-Valmayor, A./ Guinea, M./ Hernanz, H. (2006): From Research Resources to Learning Objects: Process Model and Virtualization Experiences. *Journal of Educational Technology & Society*, Vol. 9, Pg. 56-68

Vickery, B. (2008): Faceted classification for the web. *Axiomathes* 18(2) (June 2008) 145–160.