The Almirall Project: a Portal of 19th-century Culture and Thinking

Lluís Vicente, Ateneu Barcelonès

Sergi Montes, Ateneu Barcelonès

Ferran Burguillos, Ateneu Barcelonès

Pablo Casas, Ateneu Barcelonès

Rubén Alcaraz, Ateneu Barcelonès

1. Scenario and presentation

The Ateneu Barcelonès is a civil, private cultural association that has been a part of Spanish society and culture since 1860, having been founded with the aim of promoting dialogue and disseminating knowledge about the country's artistic, historical and intellectual heritage. Though the association was somewhat in decline by the end of 2003, the board of directors elected at that time, chaired by the architect Oriol Bohigas, managed to kick-start its full reorganisation, establishing a momentum for change, especially with regard to its cultural programming and the activities of the Writing School and the Library. The latter is one of Spain's most relevant repositories of heritage collections, specialising in the study of the 19th century and early 20th century, as well as in the social sciences, humanities and literature.

Various scholars and specialists (Jordi Cassasas, 2006; Joaquim Coll, 2006; Manuel Pérez Nespereira, 2007) have also highlighted the relevance of the Library's catalogue, most notably because of the heritage collection specialised in the 19th century and more particularly because of its role as vehicle for introducing European cultural trends into Spain. In fact, 30% of the books published prior to 1901 are found in no other Spanish library.

The wealth of this library lies in the different roles it held between the 19th and 20th centuries. On the one hand it played substitute for the Biblioteca Nacional de Cataluña *(Catalonia National Library)* until 1913, but it also served as a specialised social sciences library throughout the last third of the 19th century and the first third of the 20th century, in the absence of quality specialised university libraries.

As a result of its being used in these capacities, the library became the first recipient in Spain of works and authors on romanticism, positivism, nihilism and Darwinism. In a century replete with revolutions and historic figures such as Napoleon, Maeterlink and Bismark,

Hegel, Renant, Darwin and Nietzsche, one cannot study how they entered our country without taking into account the journals and monographs listed in the Ateneu Library catalogue.

So, if we hope to understand 20th-century Spain, it would appear advisable to know something about the cultural and social framework to which the 19th-century intellectuals and politicians had access. The Ateneu Library seems like a privileged place to start investigating.

As a result of this uniqueness, the Ateneu Library has, since 2005, been taking part in various digitisation projects, including particularly the *Google Library Project,* the *Hemeroteca digital de l'Aula Màrius Torres (Aula Màrius Torres digital newspaper archive)*, the Biblioteca Virtual de Prensa Histórica *(Virtual Library of Historical Press)*, the Biblioteca de Cataluña's ARCA *(Catalonia Library's ARCA)*, the Memoria Digital de Cataluña *(Digital Memory of Catalonia)* and Europeana.

This participation has led to the dispersal of its collections among different repositories, thereby making them more widely available to researchers and remedying the lack of an institutional repository.
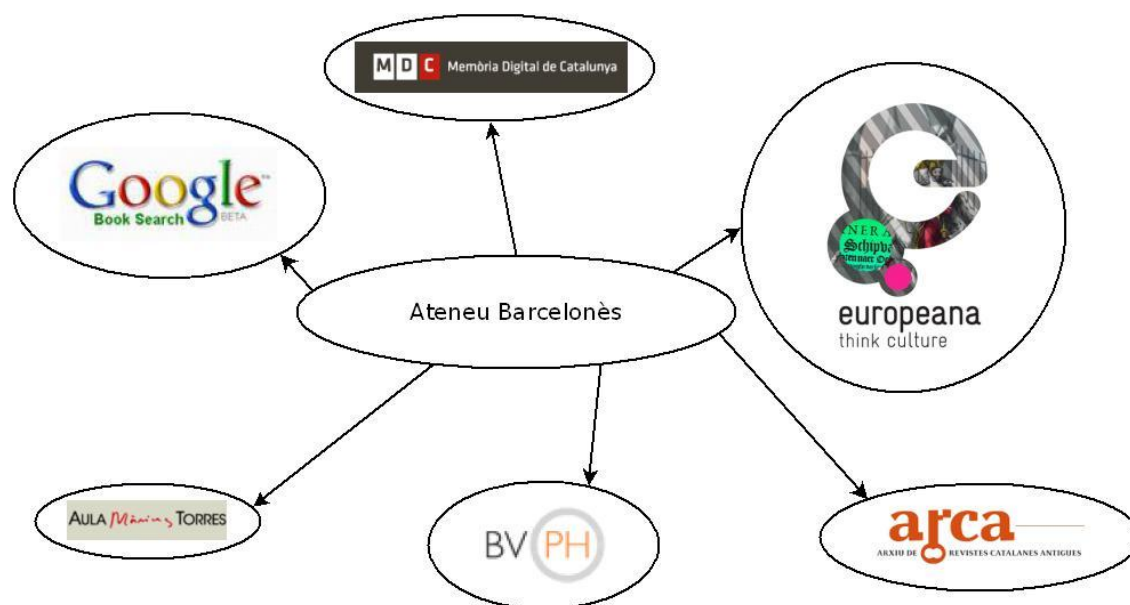
Fig.1. Synergies and digitisations

That is how, with such a formidable task ahead (the dispersal of digital copies and interest in researching culture and ideas in Spain in general, and Catalonia in particular) the Almirall Project was born.

2. The Almirall Project

2.1. What is it?

The Almirall Project aims to be a portal that explains, contextualises and inter-relates 19th-century works, thinkers, artists and schools of thought at Barcelona's Ateneu Library, providing access to digital copies lodged at other portals. In addition, its aim is to progressively integrate, through cooperative agreements with other institutions, other catalogues pertaining to specialised social sciences libraries with published heritage. The project therefore intends to be a channel for providing information and documentation

resources for specialised research in the study of 19th-century thinking, politics and culture. This mission hinges on the following objectives:

- To structure historical and bibliographic discourse on the culture, society and ideas in 19th-century Catalonia and Spain.

- To explain the works that are, or will be digitised, through the study of their impact in Catalonia and Spain, not forgetting the authors and cultural movements.

- To promote the study of sources of information on the 19th century by the research community.

- To provide the academic and research community with the information resources available on the cultural history, the intellectuals and ideas, and the arts, science and society in 19th-century Spain.

- To promote cooperation between universities and research centres in the study of the 19th century.

- To lead the participation of other centres through the digitisation of collections and incorporate the participation of researchers from any institution or university.

The idea is to build a repository structured on the basis of digital copies in other repositories and articles by researchers commenting on and explaining those digital copies. Therefore, one of the characteristics that sets this repository apart from the rest is its added value. It affords not only a digital copy but also comments from an expert. As one can see, this goes further than the typical institutional or subject repository. The digital object is the study of another digitised object.

2.2. Theoretical model:

On the basis of this premise we see that the digital object of the Almirall project can be:

- A 1,500-word scientific article consisting of the introduction of a work, author or cultural movement and its reception in 19th-century Spain. This article has to be revised by a scientific coordinator for final approval. Lastly, to reach a wider spectrum of users, the content is translated into English and Spanish.

- Other classical objects: digitised books, videos and audio files.

All these objects must be able to inter-relate with each other in multiple, limitless ways. Any one content can redirect to other objects that are either internal or dispersed within the ensemble of projects in which the institution is participating.

Finally, as it is a research portal, the model includes the interaction and contact between researchers specialised in the History of Ideas and of Society. It is a matter of generating synergies that promote the development of knowledge.

2.3. Selection of the software

A complex repository implies special, scalable, highly flexible technical requirements, particularly if the proposed model is based on researchers introducing texts that discuss works, cultural movements, authors, receivers and intellectuals. A commented catalogue needs an almost bespoke repository.

For this purpose, in 2008 different alternatives were tested, including:

ContentDM: commercial software developed principally by the University of Washington and OCLC.

- Advantages include rapid implementation, ease of use and a low learning curve. This application was known to the team as a result of their participation in entering digital copies in the Memoria Digital de Cataluña *(Digital Memory of Catalonia)*.

- On the other hand, the somewhat inactive community, the structure centred on the use of graphic documents, an inflexible metadata schema and the cost of a commercial solution, made it a rather unattractive option.

Dspace: free software most used worldwide at the time, from the Massachusetts Institute of Technology.

- Broad, dynamic community, easy to install and use, and with a low start-off cost. What stood out, among other things, was the workflow between the investigator, reviser and publisher profiles. Lastly, the user-orientated services were highly positive.

- On the other hand, we are talking about a repository focused on self-filing, sponsored by the Open Access movement. Also, being somewhat insensitive to changes, it did not adapt well to the structure of inter-object relationships. As it was not flexible and was aimed more at a different type of objects it did not meet the project's specifications.

Greenstone: multilingual platform, sponsored by UNESCO:

- Very easy for creating digital collections, and for learning; also offers the possibility of incorporating various metadata schemas. Large, very active community.

- Poor graphic versatility, cannot be scaled to the needs of the project and, moreover, not very sensitive to matters concerning long-term preservation.

The applications were tested by the Library team, visits were made to people who had some experience using the applications, and individual studies were carried out based on a matrix that re-examined the critical processes of a repository (digital preservation, publication process, access to information, interoperability and administration) and the agents involved

(collection, digital object, metadatum and user). The result of the said schema can be seen

applied to that prepared for Fedora in figure 2.

	Preservation	Publication	Access to information	Interoperability	Administration
Collection	Complete collections can be preserved through FOXML.	Such a versatile system allows publication as and when required by the repository.	Information can be accessed by different search means or by navigation.	Possibility of migration of all the collections, or one by one.	As required by the repository.
Digital objects	Allows different versions of one and the same object.	Orientated towards XML objects, image, video and audio. By creating a specific workflow.	Access by search, navigation and random presentation of content.	SOAP	By creating a specific workflow.
Metadata	The metadata formats tried are: DC, MODS and MARC21	Can be visualised in MODS, DC or METS.	Consulted based on the description of each digital object.	OAIPMH: both for being harvested and for harvesting. Possibilities of MARC data migration to the repository.	
Users	User actions can be preserved through METS and the selection of PREMIS metadata	Different publication-user profiles can be created.	From user interaction with platform content.		The system allows external- and internal-user control.

Fig.2. Table of indicators used for selecting the repository[i]

So, after analysing the different options, the most flexible version was chosen, which was

the one that included the positive points of the other repositories as well as the ensemble of

requirements inherent to the project.

Flexible Extensible Digital Object Repository Architecture is a repository sponsored by the W. Mellon Foundation and Moore Foundation. Worth noting among its main features are:

- The possibility of creating digital objects in any format.

- Flexibility when it comes to selecting metadata schemas, inter-object relationships or preservation schemas.

- The possibility of adding any kind of new preservation strategy, new formats, limitless relationships. We are looking at an extensive repository.

3. Construction and architecture

3.1. Definition

In deciding on such a flexible, as well as powerful, option, a period of reflection and benchmarking was necessary. After all, we wanted to implement a file ingestion system that would allow different types of internal users (content creators), librarians, and external users (network of researchers, potential content creators) to work with it.

Similarly, the system had to be simple enough to enable all the agents involved (creators, scientific coordinators, translators and librarians) to contribute content without any problem.

Likewise, not only did it involve the construction of a technological tool, but many of the Library processes were affected and had to be reformulated:

- Attention to researchers: access to the original documents, query response times, equipment for carrying out their tasks.

- Digitisation policies: selection of monographs, search for works already digitised to avoid duplicating effort, resources, formats and preservation metadata.

- Acquisitions: purchase of works that discuss the heritage works, selection of relevant, exceptional authors in the library.

To this must be added the creation and management of a network of 11 researchers who were going to take part in the front-line research.
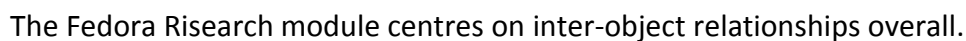
Lastly, both for the sake of the prestige of the future work and out of interest in the creation of a portal of these characteristics, we considered creating a platform in several languages (Catalan, Spanish and English).

3.2. Production and architecture

The basic architecture of the project consists of a repository server, Fedora (Fedora Commons™), where all queries are made using the search engine Lucene, and a web platform capable of administering the data provided by the server. This platform acts as a bridge between the server and the end user of this program, showing only what is of use to him/her. This ensures that the user is limited when it comes to manipulating the server (which is an advantage in terms of security) to the functions provided by the web, although these functions are wide-ranging and more than sufficient for developing the tasks offered, as well as having other user-friendly features.

Just as the user interacts with Almirall, so Almirall interacts with Fedora via the SOAP protocol (or sometimes also with HTTP) and this, in turn, with its internal modules, namely, Generic Search, Lucene and Risearch. A database is also needed which will mainly be used by Fedora but also partly by Almirall.

The figure below shows an abstract view of the general architecture.

The Fedora Generic Search (GSearch) module, along with Lucene, consists of a search and indexation system for faster, easier records recovery. This system allows manual reindexation in the event of something going wrong.

The Fedora Risearch module centres on inter-object relationships overall.

In addition, the metadata on the files in XML files (specifically in FOXML format) are kept on the Fedora guest server. Pictures, audios and videos are all stored physically on the server, along with their corresponding XML.

Object management and structure

The architecture of the repository allows it to hold vastly diverse types of documents, as the model of object defined is very generic. Added to this is its semantically enhanced capacity for including different types of relationships, thanks to the use of the RDF language.

Each of the repository objects has a unique identifier (PID), a properties file (date created, modified, etc.) and different datastreams. These latter contain the basic information that allows management of the digital objects. One of Fedora's main strong points is that it

allows the addition of one or several content items for each object, and these may be housed inside or outside the repository.
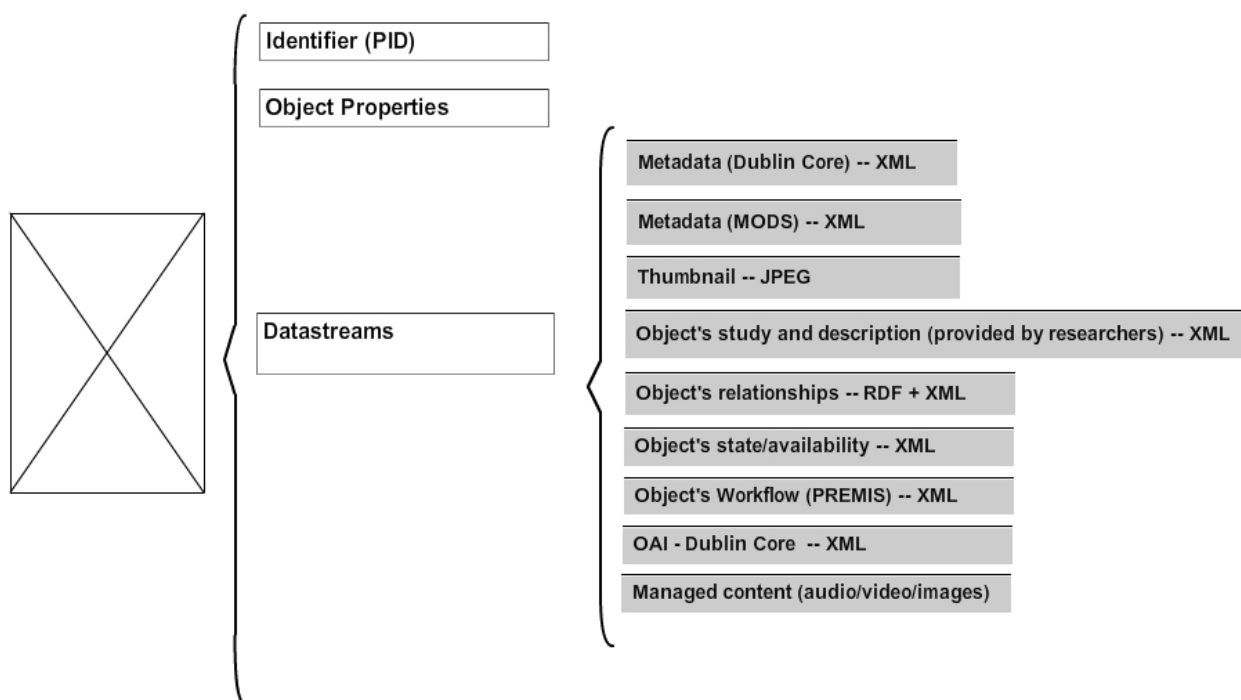

Fig. 2 Structure of a digital object in the repository

The list of datastreams is as follows:

- **DC:** required by Fedora; incorporates an identifier and a title

- **Image:** refers to the miniature representing the object (Thumbnail)

- **MODS:** includes all the metadata describing the object

- **RELS-EXT:** incorporates the relationships with other objects in the repository

- **Comentari** *(Comment)***:** includes the researchers' comments on the work

- **Managed-Content:** present in image-, audio- or video-type objects and reference to multimedia contents housed inside or outside the server

- **Control:** informs about the object's workflow

- Premis: collects the licence, object integrity check, and the events and agents that have intervened, etc.

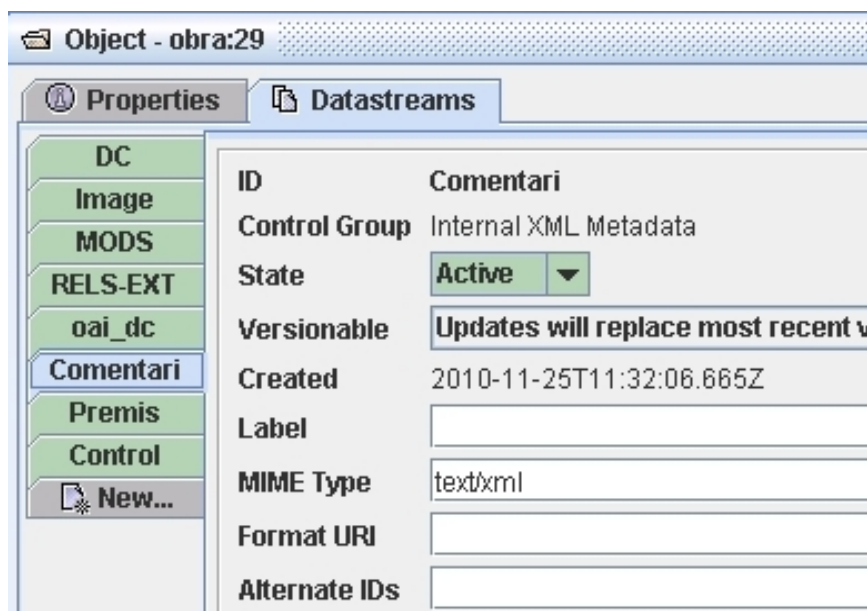- **OAI-DC**: datastream required for an OAI-PMH service provider environment.

Fig. 3 Fedora client's datastreams display

To ensure the integrity of the repository contents a checksum procedure has been developed that checks the selected datastreams and is capable of generating reports on possible mismatches. Specifically, this function acts on the datastream that contains the researchers' contributions (Comentari – *Comment*) and also on the multimedia content datastream.

3.3. Result: first prototype

With the repository constructed, it is worth highlighting some of its main strong points:

Workflow:

This is based on the need for different user profiles that contribute content and value in the object construction process:

- Researcher: the principal creator of content that is ingested via the fields created in the repository.

- Scientific coordinator: responsible for validating object content and proposing inter-object relationships.

- Translator: responsible for controlling appropriate content style and subsequent translation.

- Librarian/Administrator: responsible for creating the digital object so that the researcher can start work on it. Once it has been checked by the scientific coordinator, he/she publishes it and validates aspects such as the bibliography, inter-object relationships and maintenance of the platform navigation tools.
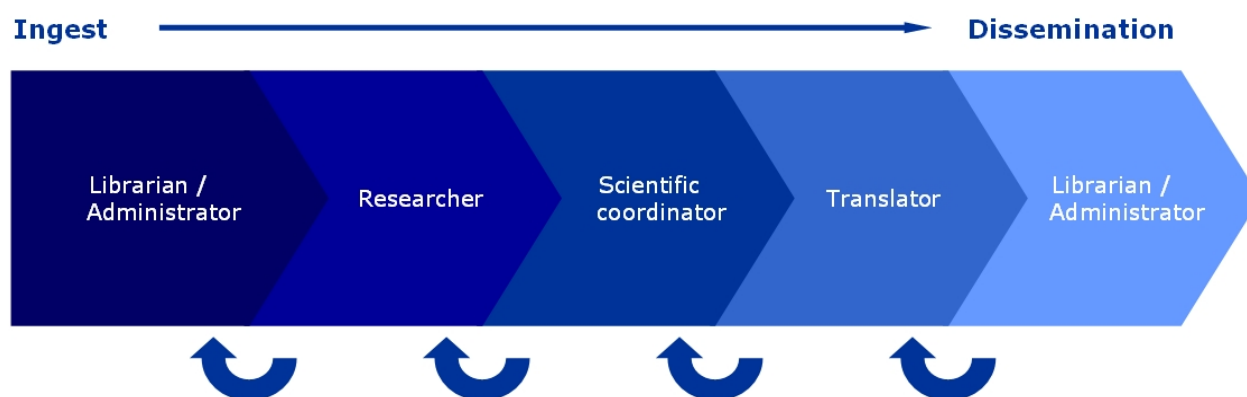
Ingest → **Dissemination**

Librarian / Administrator → Researcher → Scientific coordinator → Translator → Librarian / Administrator

Fig. 4. Workflow diagram.

The above diagram shows that at any stage in the procedure the digital object can be sent back to the previous stage for any revision or correction necessary.

All the profiles approve their interventions under the Creative Commons licence (by-nc-sa/3.0)[1]

---

[1] <http://creativecommons.org/licenses/by-nc-sa/3.0/es/>. [Query: 14/01/2010].

Objects

There are two types of objects:

- The usual images, videos and audios that allow preservation of electronic materials that have not been reposited elsewhere.

- Work, Authors and analytical concepts objects. XMLs created from the contributions of the researchers, scientific coordinators, translators and librarians. These objects can have *n* relationships with the rest of the objects in the repository or with objects lodged in others. They have different datastreams that allow a flexible structure and preservation of the versions.

Preservation Strategies:

Different strategies have been used to ensure the digital object context, origin, integrity and identification:

- The repository's own objects are the FOXMLs[2]. This allows control of the versions and relationships and assures the identification of each type of object.

- Metadata from the PREMIS preservation dictionary have been adopted.

- Mechanisms and strategies for individual and massive file downloading have been created for future migrations.

- The M5 file integrity checking system has been adopted.

- An automatic control has been implemented to ensure link validity.

PREMIS metadata collect a great deal of information automatically. This is the case with all information concerning the traceability of each of the digital objects. The semantic units of the "Event Entity" (*Event*) are a clear example of this. These provide information on the

---

[2] *Introduction to Fedora Object XML (FOXML).* Fedora Project, cop. 2005. <http://fedora-commons.org/download/2.0/userdocs/digitalobjects/introFOXML.html>. [Query: 14/01/2011].

actions affecting each of the objects in the repository, such that maintenance of the digital

origin of the objects (essential for their authentication) is guaranteed.

```xml
- <event>
    - <eventIdentifier>
        <eventIdentifierType>Almirall event ID</eventIdentifierType>
        <eventIdentifierValue>obra:43_1</eventIdentifierValue>
    </eventIdentifier>
    <eventType>Ingesta</eventType>
    <eventDateTime>2010-04-26</eventDateTime>
    <eventDetail/>
    - <linkingAgentIdentifier>
        <linkingAgentIdentifierType>bibliotecari</linkingAgentIdentifierType>
        <linkingAgentIdentifierValue>40998148B</linkingAgentIdentifierValue>
        <linkingAgentRole>bibliotecari</linkingAgentRole>
    </linkingAgentIdentifier>
  </event>
- <event>
    - <eventIdentifier>
        <eventIdentifierType>Almirall event ID</eventIdentifierType>
        <eventIdentifierValue>obra:43_2</eventIdentifierValue>
    </eventIdentifier>
    <eventType>redaccio</eventType>
    <eventDateTime>2010-05-25</eventDateTime>
    <eventDetail/>
    - <linkingAgentIdentifier>
        <linkingAgentIdentifierType>redactor</linkingAgentIdentifierType>
        <linkingAgentIdentifierValue>33955356L</linkingAgentIdentifierValue>
        <linkingAgentRole>redactor</linkingAgentRole>
    </linkingAgentIdentifier>
  </event>
- <event>
    - <eventIdentifier>
        <eventIdentifierType>Almirall event ID</eventIdentifierType>
        <eventIdentifierValue>obra:43_3</eventIdentifierValue>
    </eventIdentifier>
    <eventType>enviat a redactor</eventType>
    <eventDateTime>2010-06-22</eventDateTime>
    <eventDetail/>
    - <linkingAgentIdentifier>
        <linkingAgentIdentifierType>corrector</linkingAgentIdentifierType>
        <linkingAgentIdentifierValue>45543107C</linkingAgentIdentifierValue>
        <linkingAgentRole>corrector</linkingAgentRole>
    </linkingAgentIdentifier>
```

Fig. 4. Fragment of the PREMIS file associated with a repository object.

Interoperability

Besides its own module for allowing harvesters to access the repository, there are others:

- *OAI Harvester*: this module allows harvesting from other repositories, thus permitting the recovery of works in other projects.

- A MODS metadata ingest module has been generated in order to take full advantage of the description given in the institution's bibliography manager.

Services 2.0.

Eventually, it is hoped that the repository will not only contain objects but will also give rise to a network of researchers as a result of:

- the creation of a private space where users can store their searches and objects, share them with other users and also label them

- the possibility of commenting on objects or contributing external links to enhance the content of the objects

- the syndication and automatic citation of the portal objects.

John Stuart Mill 1806-1873

Redactat per l'investigador David Cao, Ramon Alcoberro

Publicat el 29/12/2010.

[+] Cerqueu documents relacionats

Versió en PDF
Guardeu al vostre perfil
Obtingueu citació bibliogràfica
Compartiu |

Los usuarios pueden guardar objetos en su perfil

Biografia i obra

(Londres, 1806 – Avinyó, 1873)
Filòsof anglès. Educat pel seu pare, el també filòsof escocès James Mill (1773-1836), va rebre una vasta formació enciclopèdica sota una disciplina ultra exigent.
Format intel·lectualment en la influència directa de l'utilitarisme i l'escola d'economia política imperant -Jeremy Bentham (1748-1832) i David Ricardo (1772-1823) eren ...

Llegir [+]

Influència a Catalunya i Espanya

Un dels casos més enigmàtics del pensament català de finals del segle XIX i principis del XX és la total i absoluta manca de petja del pensament de Mill en els pensadors i sociòlegs del període. L'utilitarisme victorià fou quasi totalment desconegut a Catalunya, a diferència de l'idealisme anglès, representat ...

Llegir [+]

Comentaris	Bibliografia

Encara no s'han escrit comentaris per aquest document.

Per afegir un comentari heu d'anar als documents guardats del vostre perfil i afegir-lo des d'allà.

Posibilidad de comentar los objetos, así como de añadir nueva información

Fig. 5. Services 2.0 in a digital object.

Comentaris: 0 / Objectes: 5 / Cerques: 0 / Etiquetes: 0 // Heu entrat com **Rubén** / Tanqueu sessió // Administració

Beta 0.9

Almirall
Portal del pensament i cultura del segle XIX

Cerca ràpida   Utilitzeu la **Cerca Avançada** per agefir criteris múltiples
☐ cerca exacte
Escriviu una paraula o frase...

Qui som? **Altres projectes** **Col·laboradors** **Polítiques del dipòsit**
Inici // Perfil

Objectes	Cerques	Comentaris	Etiquetes	Modifiqueu dades

Ordenar per: data de guardat / títol   Elimineu sel·lecció   Genereu Bibliografia

☐ **La civilización**
*Guardat el 14/01/2011. Document guardat per **1 altre usuari**.*
[Afegiu etiquetes]

☐ **Histoire de la littérature anglaise**
Guardat el 14/01/2011.
[Afegiu etiquetes]

☐ **John Stuart Mill 1806-1873**
Guardat el 14/01/2011.
[Afegiu etiquetes]

☐ **Literary studies**
Guardat el 14/01/2011.
[Afegiu etiquetes]

☐ **Walter Bagehot 1826-1877**
*Guardat el 14/01/2011. Document guardat per **1 altre usuari**.*
[Afegiu etiquetes]

Dades personals

Nom: Ruben

Cognoms: Alcaraz

Àlies: Rubén

Correu: ralcaraz@ateneubcn.org

Modifiqueu les dades [+]

Les vostres etiquetes

[   ] OK!

Usuaris relacionats (2)
Nicolas
xanta

Fig. 6. User's private space.

4. Results and future

After 9 months' operation at the production stage (that is, not yet accessible for the public at large), and with the first trial period completed, some of the initial results are worth noting:

- Nearly 200 digital objects were created by researchers, centring on the cultural reception of positivism, technological progress in Spain with the arrival of English works, and the creation of knowledge through the study of journals of that era.

- Participation in all the digitisation projects has been maintained, and the outreach of these digital objects has been reinforced with the explanations and descriptions contributed by researchers.

- Researchers from 4 Spanish universities have taken part, generating a common discourse that guarantees a fairly complete overview in each of the lines.

- The digital preservation of works that have not been deposited in any of the aforementioned digitisation projects has started.

- The policy of keeping an appropriate collection for supporting the researchers is to be maintained, thereby enhancing certain aspects of the Library and making for a more efficient, effective service. The Library helps to generate knowledge and is reinforcing its specialisation in the social sciences and its uniqueness.

After one month of publication for the whole community on the Internet:

- There were some 25 external users and possible contributors.

- There were some 1,300 visits by users who spent, on average, 12 minutes in the repository.

- Three heritage libraries have shown an interest in taking part in the project.

Future plans include:

- Introducing the participation of business partners who can develop technology for improving the search function and introducing clusters.

- Establishing participation with other heritage libraries that can contribute their digitised collections or specialist researchers.

- Translating all the digitised text objects into Spanish and English.

- Introducing other researchers from other universities or research centres.

- Technological revision of the repository by studying user interaction.

- Creating a monograph digitisation plan for 2011-2013.

- Cataloguing as-yet-unprocessed materials of interest for the development of the project.

- Implementing OCR search.

- Using Fedora architecture to create a repository of digitisations not included in the Almirall project.

Bibliography

CAPLAN, Priscilla. *Understanding PREMIS* [on line]. [Washington]: Library of Congress, 2009.
<http://www.loc.gov/standards/premis/understanding-premis.pdf>. [Query: 20/01/2011].

CASSASAS, Jordi (dir.). *L'Ateneu i Barcelona: un segle i mig d'acció cultural*. Barcelona: Diputació de Barcelona: RBA-La Magrana, 2006.

CROW, Raym. *A guide to institutional repository software* [on line]. 2nd ed. New York: Open Society Institute, 2004.<http://www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf>. [Query: 20/01/2011].

GARRIDO PICAZO, Piedad; TRAMULLAS, Jesús. "Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones". *El profesional de la información*. Vol. 15, nº 3 (May-June 2006), p. 171-181.

LAGOZE, Carl [et al.]. "Fedora: an architecture for complex objects and their relationships" [on line]. *International journal on digital libraries*. Vol. 6, no. 2 (2006), p. 124-138. <http://arxiv.org/abs/cs.DL/0501012>. [Query: 20/01/2011].

STAPLES, Thornton; WAYLAND, Ross; PAYETTE, Sandra. "The Fedora Project: an open-source digital object repository management system" [on line]. *D-Lib magazine*. Vol. 9, no. 4 (2003). <http://www.dlib.org/dlib/april03/staples/04staples.html>. [Query: 20/01/2011].

[i] The matrix is taken from the model presented by Jesús Tramullas and Piedad Garrido Picazo in their article on free software: "Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones" *in El Profesional de la Información*, Volume 15, Number 3/May-June 2006