eSpacio-UNED - a Complete Service Repository for Digital Contents

Luis Zorita, Alicia López Medina, Miguel Latorre, Manuel Castro, and Manuel Blázquez

Abstract —**The institutional repository created at UNED (Universidad Nacional de Educación a Distancia, Spain) is a project, originating, from the UNED Library computer services, that aims to provide coverage to members and academic staff at the University. The project, currently in full operation, offers a set of services for the management, preservation and access over the Internet of digital materials created by its members. Derived from these services, the purpose of the repository is the spread of the documentary funds as a tool of consultation and divulgation to researchers, scientists and educators in general.**

This document describes the functional and technical features of the repository, whose implementation keeps Fedora as backend application as well as Fez as Web interface. Both are freeware distributed under GPL license, making it suitable for academic environments, due to reduced costs for computer equipment and constant upgrading resulting in a permanent immunity to obsolescence.

Among the features presented in this paper, there are: the visualization of multiple presentations, the dynamic transformation of the contents with subsequent adaptation to environments outside the repository and the definition of relationships between objects.

Keywords —**digital object, Fedora, integration of services, repository, reusability.**

I. Introduction

THE documentary collections, in respect of works developed by the members of any educational community, have always been part of the legacy and heritage of it. Throughout centuries of teaching and research, the University, as a center of knowledge, has been preserving and using these materials for the development and continuation of new works.

The authors would like to acknowledge the Spanish Science and Innovation Ministry and the CYTED-508AC0341 "SOLITE – SOFTWARE LIBRE EN TELEFORMACIÓN" Project support, as well as the support of the Complimentary Action TSI2007-31091-E "Objetos educativos reutilizables (para el EEES en las especialidades en las especialidades de las tecnologías de la información y las comunicaciones)".

Luis Zorita is with Library of the Spanish University for Distance Education, Madrid, Spain (e-mail: lzorita@pas.uned.es).

Alicia López Medina is with Library of the Spanish University for Distance Education, Madrid, Spain (e-mail: alopezm@pas.uned.es).

Miguel Latorre is with Electronics and Computer Engineering Department Spanish University for Distance Education, Madrid, Spain (e-mail: pelaga@gmail.com).

Manuel Castro is with Electronics and Computer Engineering Department Spanish University for Distance Education, Madrid, Spain (e-mail: mcastro@ieec.uned.es).

Manuel Blázquez is with Electronics and Computer Engineering Department Spanish University for Distance Education, Madrid, Spain (e-mail: manuel.blazquez.merino@gmail.com).

Because of the amount and physical dimensions of materials, institutions held large buildings for storing them, with just a small fraction being accessible to the public. The means of cataloguing and conservation were intended to serve the physical object. Nowadays, the situation has changed due to digital media and the existence of data networks. The physical quantity of materials, that are to be retained, tends to be diminishing, so that institutional protocols are mainly focused on the conservation and management of digital documentation by means of educational platforms and repositories.

In such a way, an institutional digital repository is a set of services that the university offers to its community for the management, preservation, access and dissemination via the Internet of digital materials created by the institution and its members. It is essentially an institutional commitment to control these digital materials [1]. Therefore, an institutional repository is a recognition that intellectual activity is increasingly represented in a digital form and that the main responsibility of an institution is to exert control over its intellectual productions, making them accessible, easily retrievable and ensuring their existence through time. A great projection case in science and academic environment is the Massachusetts Institute of Technology (MIT) [2]. Their Library collections are being promoted, not only in terms of conservation and maintenance but also in the generation of open courses, i.e., the interaction of different materials to focus and to specialize certain specific lectures.

II. Background of the Introduction of the System

Universidad Nacional de Educación a Distancia (UNED) is a global Spanish institution which operates all over the country, having delegations in every region. This special structure makes the University to produce and create digital contents and systems in a different way. The extension through Spain of agencies and staff working for UNED, given the national scope of its reach, with the consequent dispersion of UNED students and professors, is not exactly a feature shared with other universities and institutions. This particular situation, with large numbers of registrations and collaborations, causes the generation of a growing, complex and heterogeneous digital content. At first instance, the UNED library has conducted an analysis approach to digital content accessible through the virtual campus, and a huge number of locations on department websites were recognized. The analysis has given two main evidences. First, it was recognized that

free accessibility of contents is related to departmental personnel's interest in its academic and research dissemination. These materials include journal articles, book chapters, papers presented at various conferences and seminars, etc. Even with such a great academic and educational value, this material has low output through traditional channels of scientific publication. The second conclusion comes from this first evidence and is based on the need to standardize the location of these materials. That is, it is necessary to create a large core to access this documentation in a consolidated and linked way.

III. Characteristics of Managed Digital Objects in a Repository

The main functional properties a digital object is characterized by are based in the conditions of heterogeneity and complexity. The first one consists of the definition of the diversity of objects that a digital system can hold. This diversity is defined by MIME types and can be texts, images, films, sound recordings, tables, datasets, etc. Even an object sustained in a digital system can be composed by a set of any of them, which gives to the object the additional property of complex. Therefore, this second defined feature, complexity, suggests that when generating an object, it is composed of different parts. Therefore, it would be inaccurate to talk about documents when defining the component to integrate into a repository. As a reference example to a specific object, a doctoral thesis, the complexity lies in its composition, being constituted the subject as a pdf document, data tables, and even related graphics software, used to perform certain calculations.

Given the global situation of information and data worldwide, there's an increasing need to make visible productions on the web. The way to make a production visible is to link a heterogeneous and complex object to a persistent URI at the time of its creation. Moreover, taking into account the growing importance of the Semantic Web, it is desirable for every digital object to have a semantic representation of the concept over a local repository. This provision is indicative of the creation of a semantic link that allows each and every one of the objects in each repository.

IV. Design of eSpacio-UNED

The eSpacio -UNED institutional repository is based on Fedora [3], SOA architecture backend for digital content. Fedora has, therefore, defined a set of Web services published by WSDL, where customers can be connected, created by third parties. Consequently, Fedora is a generic management middleware, capable of containing any type of digital object, providing more services than just an application, to which types of content to be managed have to be adapted. This model gives rise to a number of architectural features, of which the main ones are: definition of a new structure for digital objects, that can be serialized in XML by means of transformations (fedora.xsd); ability to manage assets locally or remotely; establishing relationships between digital objects and among the components of a digital object, expressed in RDF [4]; access control using XACML standard with different levels of granularity [repository, object, components of an object (datastreams)]; authentication: login, sign on, LDAP and Shibboleth; control and maintenance of different versions of a digital object; audit and Data Curation. Consequently, Fedora is based on a concept of digital object container that holds the data structure. Each digital object is, therefore, an aggregation of elements that can be used individually or as part of a set. As a result, both can be heterogeneous and complex, equipped with a component that accounts for relations among various objects, both inside and outside repository.

About the functionality of Fedora, free software with GPL license it is a noteworthy characteristic, leading to immediate benefits, such as, fitness for academic environments to be more reliable, robust and secure; very low cost that allows locating financial resources in other areas that would greatly improve the public utility of the universities (infrastructure, scholarships, research support, etc.); and type of software that requires less computational resources and extends the life of computer equipment from both universities and students, avoiding rapid cycles of obsolescence and thereby maximizing the investment.

V. How the Repository is Integrated in the Architecture of the University

SOA architecture is used by UNED to offer its services in order to generate and implement specific applications and modules. As an integrated system, the architecture of the repository is placed in a central situation on a general diagram of system and services. A diagram is shown in Fig. 1, which shows this scenario in the blocks located at the top, which highlights the portal eSpacio, management University site, the OCW open courseware site, Tele-UNED site and other specific webs from Faculties and Departments. The idea of integration is based on FEDORA situation as a bottom layer with functions based on the provision of management services to these sites.

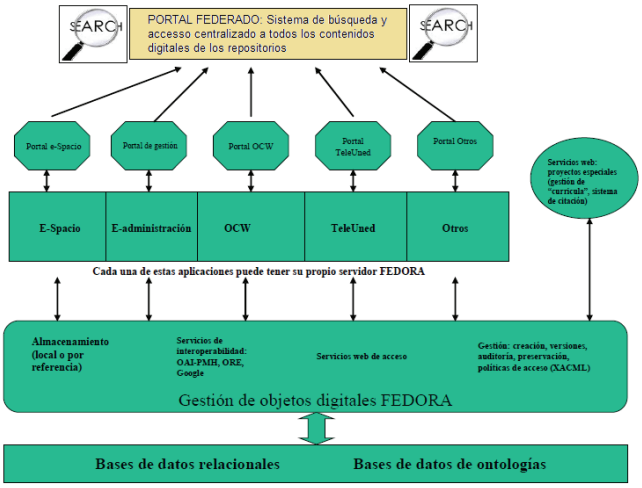

Fig. 1. Diagram of systems integration through FEDORA.

The services and digital objects treatment and management, by means of Fedora, have been made generically, relying on relational databases and ontologies

databases. This integration is easily implemented and meets the requirements described in previous sections.

VI. SERVICES OFFERED BY ESPACIO-UNED

The repository is able to interoperate with other systems, both within the institution as external systems (administrative systems, portals, other repositories) so that their contents are easily removable and reusable [5]. Therefore, the digital repository eSpacio-UNED works with protocols and open standards. It can potentially be integrated as a central element in the information management network of the institution, enabling it to compete and collaborate more effectively and efficiently, as shown in Fig. 2. The standards and protocols provide protection against obsolescence and inaccessibility. They represent a safe way to store contents for future rebuilding, because all of them are universally understood, documented and published. There are no associated costs in relation with use licensing and distribution, given that the used formats are free. The use of open protocols is also a prerequisite to ensure interoperability and therefore the exchange and transfer of information. In this sense, the digital repository of UNED is consistent with the OAI-PMH [6] which means that their contents can be incorporated into networks of similar collections around the world and can be searched by agents from the Internet, including Google, Google Scholar, or OAI services like OAIster.

Ultimately, the most important services provided by the institutional repository can be described as a complete management service of any type of digital objects (CRUD) derived from the exhibition site (WSDL) from several management APIs, a configurable search service for any metadata or text, a data providing service under OAI-PMH, a service for the preservation of digital assets with automated creation of metadata PREMIS [7], and a service of creating on-demand Web rendering subsets of content objects (eg, electronic journals, collections of old books, etc.).

VII. A LOAD EXAMPLE OF LEARNING OBJECTS

The Electronics and Computer Department of the Engineering Technical School at UNED has recently made a number of projects based on the recovery, reuse and organization of digital objects [8]. Those objects have been generated in the teaching and scientific fields. The nature and the origin of these objects is diverse, having objects removed from lessons, practices, departmental records, and resolutions exercises, exams, presentations by department members and documents from conferences in which the department is actively involved.

The amount of digital objects is high, so they revealed the need to unify all in the same context for reuse. The environment and services offered by the institutional repository perfectly fit to the objectives of the department in relation to the dissemination and access to digital objects in the academic field. Once reusable digital objects are individually organized with an additional structure of metadata defining each element, they have been proceded to its incorporation into the repository. As a guide for the load of the digital objects, a description in the standard IEEE-LOM metadata of each one of them has been made. To access, they were stored in an open source database eXist XML (native) and a query is created through a single input file.

As Fedora schema objects are known, the creation of an XSLT transformation was made to begin a process of sequential reading of the file. The result was the incorporation of digital objects and their XML file descriptors in IEEE-LOM standard. The integration of the resulting objects has generated, in the environment of the repository, a number of elements making up the structure of each digital object. So, the components associated with the digital object are LOM metadata component; Dublin Core metadata component arising from the LOM description; component linked to the URL of asset; component of the relationship of that object to other objects in the department and finally, FOXML representation. The burden of these digital objects has been performed at various stages related to the different natures and origins of them. Thus, in a first stage related to the digital objects derived from the teaching of the department, the charge resulted in the incorporation of 1743 objects exploiting the import function to digital formats LOM, METS or ATOM. In a second phase, the incorporation of 5405 objects derived from the presentations and congressional work documents from the department has been performed.

As an example, Fig. 2 shows a general aspect of the hierarchical structure "*isMemberOf*" in which objects and their components are organized, which are available from all elements of a learning object. In the right block, the arrangement of the components is shown, connected to a digital object, known as "*power supply*" which in turn belongs to the group known as "*Analog Electronics*" and that also, in turn is incorporated into all "*DIEEC Learning Objects*" represented to the left of Figure 4. As the component "*Reference1*" was redirecting the assets on the server, a FTP server from all of them was implemented in Apache server which supports *eSpacio*.

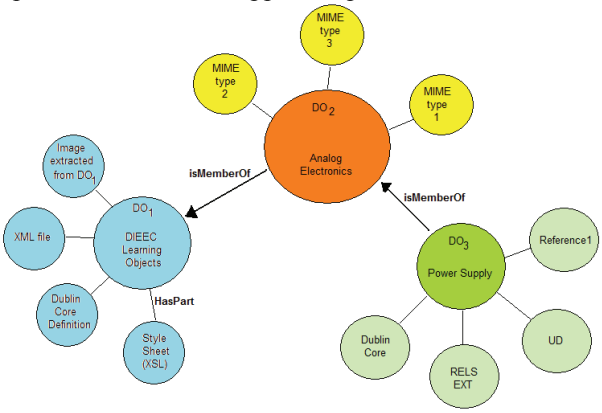

Fig. 2. A digital object and its components organization example.

In conclusion, from a LOM description, digital objects have been generated, available to be displayed on the

Web, with persistent identifiers and all the features added due to its integration in the repository space, and search within that subset, that OAI_PMH protocol incorporates in specialized portals, accessible through Google, for export in XML files, etc.

VIII. A Particular View of the Web Site Interface

The institutional repository, eSpacio-UNED, provides access through the web page address http://e-spacio.uned.es/. The homepage access of the repository has a functional aspect, as shown in Fig. 3. On this page, there are several areas highlighted as external links. These links permit the access to related documentation as well as to other author documentation. It also allows, through additional links, access to other types of searches including those advanced by multiple search fields.

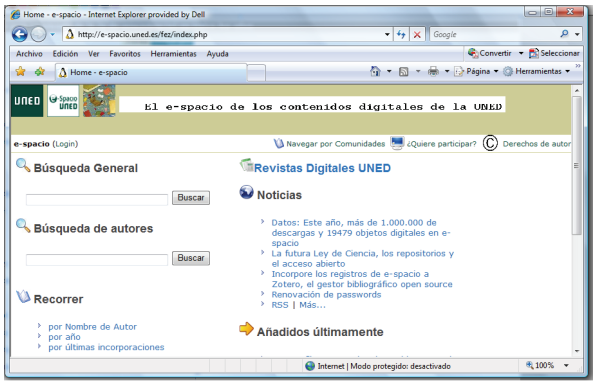

Fig. 3. User interface in eSpacio-UNED – Access homepage.

Given the widespread use of UNED digital journals, they take a leading role with access through direct links to their own page, where every one of them can be introduced, being able to visit the various issues published. Outstanding news and additional features in recent days can also be viewed. If a query is made, the answer is returned in the same dynamic environment, providing a list of matches according to exposed criteria, as shown in Fig. 4. Search terms are marked in yellow in the list box providing the information of document loading date, as well as authoring and description of the digital object type. Two more fields shall also be included, which respectively provide the link to download the document and the value of the source or collection to such digital object belongs to.

Fig. 4. Results page to a query.

The eSpacio-UNED repository website offers some added services such as the ability to view statistics associated with the most downloaded articles per month and year, the most popular authors, as well as general operating data repository on the quantity of available documents and items and countries of origin of the IP's from which consultations have been made and downloaded.

IX. Conclusion

Throughout this paper, the features and characteristics of the institutional repository space-UNED have been outlined. The launch of the repository involves the creation of a common digital space in which productions created by its members are hosted. Due to the functionality of the repository based on the determination of collections that allows a rational classification of digital objects, they can be reused as learning objects. The repository offers a range of services designed to develop a coherent and coordinated approach for the capture, identification, storage and retrieval of digital content. By system integration, an environment has been created to offer a managed treatment of contents, increasing opportunities for more effective use of the results of the activity of the institution and encouraging collaboration across disciplines and units.

From the standpoint of authors, the creator meets a centralized and assisted place to locate their work. Each of them has a persistent URI associated, which enables it to be visible on the Internet and therefore, be found by search agents and users, leading to increased screening and dissemination of its work. In short, the repository eSpacio-UNED aims to stimulate innovation, facilitate a qualitative analysis of the work of its members, sustain the work of teaching and learning, serve as a system for recording ideas and offer a catalogue of intellectual heritage of the institution.

Acknowledgment

In memoriam of Tomas Pollan, a good friend and a best colleague.

References

[1] K. Hahn, C. Lowry, C. Lynch, and D. Shulenberger, "The University's Role in the Dissemination of Research and Scholarship — A Call to Action" 2009: http://www.arl.org/bm~doc/disseminating-research-feb09.pdf

[2] Documental Funds and OpenCourseWare. MIT- Massachusetts Insitute of Technology. http://ocw.mit.edu/OcwWeb/web/home/home/index.htm

[3] Fedora Commons Repository Software: http://www.fedora-commons.org/

[4] Resource Description Framework (RDF): http://www.w3.org/RDF/

[5] C. Bizer, T. Heath, K. Idehen, T. Berners-Lee, "Linked Data on the Web," Proceedings *WWW2008*, Beijing, China: http://www2008.org/papers/pdf/p1265-bizer.pdf

[6] Open Archives Initiative-Protocol for Metadata Harvesting - v.2.0: http://www.openarchives.org/OAI/openarchivesprotocol.html

[7] PREMIS: PREservation Metadata Implementation Strategies: http://www.oclc.org/research/projects/pmwg/

[8] M. Latorre et al., "An experience on Learning Objects Reutilization based on Educational Resources Developed", *Congreso ASEE* – Austin, TX. 2009.