Desarrollo de una red virtual de investigación y educación para la información científico en Cuba

The development of a virtual research and educational network for scientific information in Cuba

Marc Goovaerts 1, Febe Angel Ciudad Ricardo 2, Didiosky Benitez Erice 3

1 Hasselt University Library, Bélgica, mcgoovaerts@gmail.com
2 Universidad de las Ciencias Informáticas (UCI), Cuba, fciudad@uci.cu
3 Universidad Central de Las Villas (UCLV) “Martha Abreu”, Cuba, didiosky@uclv.edu.cu

Resumen

i. INTRODUCCIÓN
El desarrollo del red virtual de investigación y educación es conducido por seis universidades (Camagüey, Holguín, Pinar del Rio, UCI, UCLV y Oriente) en el marco del proyecto ELINF, parte del programa, VLIR-UOS programa ‘NETWORK University Cooperation: Strengthening of the ICT role in Cuban Universities for the development of the society’. Es apoyado por el Ministerio de Educación Superior de Cuba y otras partes interesadas locales.

ii. OBJETIVO
El objetivo final es una red donde los actores en la educación y en la investigación cubana puedan buscar los servicios virtuales necesarios para sus tareas profesionales, acceder a la información para la investigación científica, intercambiar publicaciones y datos por trabajar juntos en un ambiente integrado que soporta la ciencia abierta.

iii. METODOLOGÍA
El proyecto elige solamente soluciones de software de código abierto, usando plataformas específicas para la gestión de la información, de los resultados de la investigación científica y de los sistemas de biblioteca en el ámbito universitario: Moodle, ABCD, DSpace y VIVO. Crea un ambiente integrado.
iv. RESULTADOS
Las universidades asociadas se encuentran en diferentes estadios de la aplicación de las plataformas. Desarrollan una estrategia y política de interoperabilidad. Están desarrollando nuevo módulos y herramientas para realizar la interoperabilidad entre las plataformas, por ejemplo ABCD 3.0, Darkaiv y Auctoritas.

v. CONCLUSIONES
Es un trabajo en progreso. El estadio actual es descrito en este artículo.

Palabras clave: interoperabilidad; ambiente de educación; ambiente de investigación; sistema de biblioteca; repositorio institucional; control de autoridad; identificador único

Abstract

i. INTRODUCTION
The development of a virtual research and educational network is conducted by six universities (Camagüey, Holguín, Pinar del Rio, UCI, UCLV and Oriente) in the framework of the ELINF project which is a part of the 6+6 year VLIR-UOS program ‘NETWORK University Cooperation: Strengthening of the ICT role in Cuban Universities for the development of the society’. It is supported by the Ministry of Higher Education of Cuba and other local stakeholders.

ii. PURPOSE
The final goal is to create a network where the actors in Cuban education and research can find the virtual services for their tasks in education and research, access the necessary information for their research, communicate and work together and exchange publications, data and other forms of scientific information in an environment which supports open science.

iii. METHODOLOGY
The project has chosen to use only open source software solutions. It works with specific platforms for educational, library and information management and research output, respectively, Moodle, ABCD, DSpace and VIVO and create an integrated environment.

iv. RESULTS
The partner universities are in different stadia of implementing the platforms. A strategy and policy of interoperability is developed. The project is developing new tools and modules to realize interoperability between the different products: e.g. ABCD 3.0, Darkaiv and Auctoritas.

v. CONCLUSIONS:
It is a work in progress where this paper presents the actual status.

Keywords: interoperability; learning environment; research environment; library and information systems; institutional repositories; authority control; unique identifiers

Introduction
The VLIR-UOS programme ‘NETWORK University Cooperation: Strengthening of the ICT role in Cuban Universities for the development of the society’ is a 6+6 year program focusing on strengthening of the role of ICT in Cuban Universities for the development of the society (Network cooperation in Cuba: ICT, 2016).

VLIR-UOS (http://www.vliruos.be/) is the organization of the Flemish University Board that supports partnerships between universities and university colleges in Flanders (Belgium) and the South looking for innovative responses to global and local challenges. One of the main goals is to strengthen higher education in the South by setting up long term cooperation projects with universities and networks of universities in the South. (About us, 2016)
The development of the VLIR-UOS ‘Network Cuba’ is conducted by five universities (Camagüey, Holguín, Pinar del Rio, UCI and UCLV) in close cooperation with the VLIR-UOS project of the University of Oriente and the Ministry of Higher Education of Cuba and other local stakeholders.

The following external stakeholders have been identified:

- Ministry of Science, Technology & Environment
- Ministry of Health
- University of Havana
- CUJAE University

The VLIR-UOS ‘Network Cuba’ includes projects focusing on research and the enhancement of the infrastructure. The ELINF project (ICT supporting the educational processes and the knowledge management in higher education) is creating a virtual research and educational environment in Cuba.

The project has chosen to use only open source software solutions. It started with specific platforms for educational, library and information management and research output, resp. Moodle, ABCD, DSpace and VIVO.

The final goal is to create a network where the actors in Cuban education and research can:

- find the virtual services for their tasks in education and research,
- access the necessary information for their research,
- communicate and work together,
- exchange publications, data and other forms of scientific information in an environment which support open science.
The Professional English Training with ICT, which is another part of the ELINF project, will not be discussed in this paper. The paper is focusing on the network development for research, education and information.

**Methodology**

The development of the educational, research and information network is considered on three levels:

- Introduction of the platforms for the educational, the library & information and the university output management at university level. All partners install the same version of the software packages, respectively Moodle as learning environment (About Moodle, 2016), ABCD as library management system (De Smet & Dhamdhere, 2012, pp. 183-191) and DSpace as institutional repository(About DSpace, 2016). The project works preferentially with Open Source Software.

- Development of interoperability between the platforms for the educational, the library & information and university output management:
  - Integration of the local content in network products, e.g. a federated library catalogue or a repository harvester
  - Creation of tools that can access data from other platforms, e.g. integrating library search possibilities in a Moodle
  - Definition and implementation of common standards over the different platforms

- Development of services on the network level to expose and to manage the educational and research output of the universities and to enhance the cooperation and communication in the network.

**Results**

The VLIR-UOS project is in its fourth year of its first phase. Intermediate results have been achieved on the first two levels and can be divided in three groups:
• Implementation of the chosen platforms in the university and the network
• Development of a the ELINF Ecosystem, a Reference Model for ICT supporting educational process and knowledge management
• Development of new tools to support the network strategy

Implementation of the chosen platforms in the university and the network

In the period 2013-2016 the basic tools are installed in the network universities: ABCD (version 1.2) for the library, Moodle (version 2.7-3.0) for the educational network and DSpace (version 5.5) for the university publication output.

The introduction of Moodle went rather smoothly. The main problem was the result of the merging process of Cuban universities in the last years. The participating universities have now a presence of up to 72% of their courses in Moodle. (MOODLE-DSPACE integration at the University of Pinar del Río, 2016).

The implementation of ABCD was confronted with the same challenge of integrating library systems of different universities as a result of the merging process. For example the University of Camaguey announced a 50% coverage of its collection in the catalogue. A major step was taken by the development of a JAVA version of ABCD by the network coordinated at UCI. The new version 3.0 is now in prerelease. A release candidate will be available on September 23 2016. The official version will be available for the Cuban universities before the end of the year.

DSpace is now available in all the universities of the network. A lot of time was needed to realize a global Open Access policy that has to be accepted by all the university boards of the network. Another issue was the standardization of the metadata. The document ‘Procedimiento para la recolección de metadatos en el Sistema de Repositorios Digitales Institucionales de las universidades de la TIC VLIR Cuba’ is a reference book for the repository metadata used in
Cuba. Not all the repositories are yet available on the internet. By October 1 all will be available. A Cuban harvester will be tested out from that period and will be available in the first half of 2017. DSpace 5.5, version JSP has been customized for Cuban institutes by an UCI-team.

![Image](image.png)

Figure 1. RediCuba Interface using the JSP version of DSpace

**An integrated network on research, education and information – the interoperability issue**

To realize the integrated research and educational network the different environments (Learning, Information and Research) have to be further developed in a way that they can interconnect and that contents can be reused. This means the implementation of common standards in the different tools and the implementation and development of extra tools.

The different platforms are using different metadata formats, MARC21 for ABCD, Dublin Core Qualified or DCTerms for DSpace and LOM for Moodle. These are mostly traditional text-based formats which have their limitations. For example, the granularity of DC Qualified or DCTerms is rather limited, surely in the DSpace implementation.
Interoperability between different platforms or data silos needs quality metadata and a standardized approach as described by I. Subirats et al. (2012). FAO developed a set of standards and recommendations for developing quality metadata in the framework of Linked Open Data. ‘These recommendations are referred to as Linked Open Data Enabled Bibliographic Metadata [17] (LOBE-BD) version 2.0. LOBE-BD assists repository managers in four key questions: (a) What kinds of entities and relationships are involved in bibliographic resource descriptions? (b) What properties should be considered for publishing meaningful/useful Linked Open Data-ready bibliographic data? (c) What metadata standards should be used for preparing Linked Open Data-ready bibliographic data? (d) What metadata terms are appropriate in any given property for producing Linked Open Data-ready bibliographic data from a local database? (Subirats, Malapela, Dister, & Zeng, 2012, pp. 161-162). These questions have to be solved to implement the interoperability problem between the different platforms in the network context as well.

The VOA3R project (Stracke, Manouselis, & Sicilia, 2013) proposed the use of a specific application profile, VOA3R-AP, to support interoperability over different systems. The experience of the VOA3R project shows that it is very difficult to impose on a community such a high-level standard of metadata.

In the context of the project it is not relevant to impose on the different platforms the use of a single metadata standard. ABCD, DSpace and Moodle have used MARC, DC or LOM because of the culture and necessities of the library, repository and educational communities respectively.

Therefore the project has to work out a minimal consensus of what can be standardized over the different metadata formats. This has to be realized on the level of the metadata terms. In the
Semantic Web philosophy it also means that items should be defined with a URI. The concept of unique identifiers is related to this approach.

Traditional identifiers in library cataloguing are ISSN for journals or ISBN for books. DOI’s (for electronic content) and ORCID and Researcher ID (for authors) are modern identifiers created for the internet. In the linked (open) data context every concept gets a URI as unique identifier. All the existing identifiers can easily be translated to URI’s.

The authors believe that for interoperability the metadata formats can be different as long as they have common elements. The minimal elements that should be approached in a standard way are the ones that define persons, organizations and content. The next step is to define every content uniquely using unique identifiers. These concepts have to be managed through authority systems.

Moodle, ABCD and DSpace have metadata formats with elements for authors and keywords. But these elements are text-based mostly without authority control systems and surely without a use of unique identifiers. Also the authority tools are not available. Authority tools for the network project needs to support the unique identification of persons, organizations and content.

The universities uses systems for personnel and student management. UPR uses for example SINEGU, while the University of Camaguey is developing SIGEP 2.0. The ELINF project gives the priority to Open Source Software. Open Source Software for managing people and organizations are limited.

VIVO is an open-source semantic publishing platform for making data about research activities visible and accessible. It is based on semantic technologies initially developed at Cornell University and now an incubator project under DuraSpace.org. Its organization of data is based on a bundle of ontologies and data are stored in a triple store. When installed and populated
with researcher interests, activities, and accomplishments, it enables the discovery of research across disciplines at that institution and beyond (About VIVO, 2016).

VIVO is as a CRIS (Current Research Information System) more than an authority tool. It is based on a semantic model. Research entities from person to institute level are already defined in an ontology which can be adapted to the local needs. As a semantic web solution it defines all the entities with a URI. It can also be queried through a SparQL end point. In that way VIVO can be used as an authority system.

The identification of scientific authors is an important issue in scientific communication. ORCID (Open Researcher & Contributor ID) is an international, interdisciplinary, open and not-for-profit organization created to solve the researcher name ambiguity problem for the benefit of all stakeholders, including research institutions, funding organizations, publishers, and researchers themselves. The core mission of ORCID is to provide a registry of persistent unique identifiers for researchers and scholars (Haak, Fenne, Paglione, Pentz, & Ratner, 2012, p. 259). Even while ORCID is proposing to resolve the research name ambiguity, persons will have many ‘unique identifiers’ even in the academic world. Web of Knowledge has its Researcher ID. Elsevier uses a Scopus Author ID and people will get an identification in all the platforms they register in: LinkedIn, Researchgate, Academia.edu, Mendeley, a university identification etc. Further on researchers have a university and the challenge for the ELINF project is to define a standard policy on the use of unique identifiers. VIVO can support the use of different identifiers in its ontology. If necessary the ontology can be extended.

Another challenge is to build up the VIVO content. Manual upload of all the metadata is not really an option. Every university has most of the information available as administrative information in tools like SIGED and SINGENU. VIVO has standard tools to ingest this information. The ingest tools have to be adapted to the format of original data made available by the partners. But if the data is managed in the local administrative tools and VIVO is used
to present the data on the web, then the use of unique identifiers should also be implemented in the local administrative systems. Test versions of VIVO are now installed at the different universities.

To describe content, the authors propose to use vocabularies. In the frame of the project, the following vocabularies are relevant: CCS for Computer Science, MESH for Medicine and life Science and AGROVOC for agriculture. Other interesting vocabularies can be added. Most of these vocabularies are available as SKOS, some are linked data compatible. Some are accessible through the internet and through specific web services. But there is no standard approach. To make them useable it is necessary to ‘normalize’ the access to the different vocabularies through a standard interface: AUCTORITAS.

AUCTORITAS is developed at UCI. AUCTORITAS is a semantic based solution, a triple store of existing vocabularies that is queried through Sparql or through web services. It will contain different vocabularies. For the network CCS, Agrovoc, MESH are the most relevant. Local vocabularies can be added as well as simple authority lists (e.g. journal lists …). It is be possible to harvest the VIVO content as well as other non-structured lists. That way all authority information can be accessed through one service. (Tabares Martín, Fernández Peña, Leiva Mederos, Goovaerts, Calzadilla Reyes, & Ruano Álvarez, 2015, pp. 214-224). The National Library of Finland has developed SkosMos (Suominen, et al., 2015), a similar tool as Auctoritas focused on SKOS formatted vocabularies.

Use of Authority Systems in Moodle, Dspace and ABCD
To make the separate tools ‘network ready’ they have to integrate the authority system functionalities in many ways. The Moodle, DSpace and ABCD software needs further customization to get ‘network ready’.
It means that in Moodle metadata elements shall manage the authority files available through the authority tools AUCTORITAS, Skosmos or VIVO. The teachers is as well identified by his name as by a unique ID (ORCID or an alternative). The content of a course will be identified by terms out of controlled vocabularies and their unique identifiers.

The main weakness of repositories in general, and of DSpace specifically, is it weak metadata format. The document about the standardization of the metadata in the institutional repositories of the Network (Procedimiento para la recolección de metadatos en el Sistema de Repositorios Digitales Institucionales de las universidades de la Red TIC VLIR Cuba) will be adopted by the expert committee on September 1, 2016. Still this a first step to implement rich and granular metadata in DSpace. In the context of the transformation of DSpace 7 a new submission module that supports the authority tools AUCTORITAS, Skosmos or VIVO and uses a unique identifier system.
These developments will make it possible to link educational and research information of a person, to create bibliographic information for a course (semi-) automatically. It is a basis for educational and scientific information and communication services.

Development of the ELINF Ecosystem, a Reference Model for ICT supporting educational process and knowledge management
A reference model is a minimal set of unifying concepts, axioms and relationships within a particular problem domain, and is independent of specific standards, technologies, implementations, or other concrete details (MacKenzie, Laskey, McCabe, Brown, & Hamilton, 2006, p. 4).

The ELINF Ecosystem Reference Model is a minimal set of concepts, axioms and relationships to support educational processes and knowledge management in Cuban universities developed by the expert committee of ELINF on July 15 2015.

Figure 3. The ELINF Ecosystem

The three levels Authority Control, Information Management and Learning and Research are integrated through Policies and Cross-Cutting tools including standardized metadata and Semantic web solutions.
The implementation of the ELINF ecosystem is based on the basic level where organizational, personal and content information is defined. The cross-cutting standards that the project will use are based on OAI-PMH, OAI-ORE and related to Linked Open Data. It will use triple store solutions and Sparql end points.

**Other developments in the ELINF project**
While the project prepared the framework to develop an integrated approach of the research and educational network, partner universities worked on specific modules to enhance the functionalities of the different platforms. Two examples.

The Moodle research group at the University of Pinar del Rio has developed modules in Moodle that can access DSpace. Teachers can now directly access DSpace from Moodle to upload items using the Sword 2.0 protocol or create literature lists using the REST API of DSpace. A similar tool will be available for ABCD as soon as ABCD 3.0 will be finished.
Figure 5. Interoperability between Moodle and DSpace developed at UPR.

UCLV students developed the Darkaiv tool to create semi-automatically metadata for the local digital library. Important parameters are speed and quality. The ABCD group of Pinar del Rio worked with the metadata extractor Tika, storing the metadata in an ABCD collection. Their tests showed that it was possible to extract metadata on a large scale in a fast way. They got results of extracting metadata of 1000 documents in 6 minutes.

Figure 6. Workflow of semi-automatic metadata extraction proposed by D. Benitez (UCLV)
UCLV worked further on the metadata extraction tool and tried to combine speed with quality. A first workflow model included extra steps of metadata enhancement using services like Crossref and Google Scholar.

The first release of Darkaiv is a Desktop tool that has the following functionalities:

- Extraction of metadata using Grobid and Tika
- Metadata enrichment through external services (Crossref, Mendeley)
- Manual correction of metadata
- Automatic evaluation of the metadata completeness
- Submission to a repository and ABCD of the metadata and document file.

The tool is developed with a digital library in mind but the submission tool is relevant for repositories as well.

Figure 7. The Darkaiv Desktop Tool
Conclusions
The ELINF project of the VLIR-UOS programme ‘NETWORK University Cooperation: Strengthening of the ICT role in Cuban Universities for the development of the society’ is a work in progress. The result will support the further development of the Cuban university community. The five universities involved in the project are building up knowledge and capacity to create an integrated research, educational and information network.

The Open Source Software tools to manage the different aspects of the network are implemented: ABCD for library and information management, Moodle for the educational platform and DSpace and VIVO for the research information management. The ELINF Ecosystem is the framework where these tools can evolve from to an integrated research and education environment.

The ELINF project results in many interesting new software developments for the information community: ABCD 3.0, Auctoritas and Darkaiv are the most prominent but smaller contributions to Moodle, DSpace and VIVO will be relevant for the different communities. In the next two years the ELINF project will implement the developments on the level of interoperability. Finally it will share its knowledge about the development of an integrated research, educational and information network through specific outreach activities.

Bibliography


MOODLE-DSpace integration at the University of Pinar del Río. (2016).


