

## **Metadatos y sistemas de autoridad para VLIRED Cuba, una red virtual de investigación y educación para la información científico en Cuba.**

### *Metadata and authority systems for VLIRED Cuba, a virtual research and educational network for scientific information in Cuba*

Marc Goovaerts<sup>1</sup>, Febe Angel Ciudad Ricardo<sup>2</sup>, Luis Carlos Alvarez Fernández<sup>3</sup>, Leandro Tabares Martín<sup>4</sup>, Amed Abel Leiva Mederos<sup>5</sup>

#### **Resumen**

Este artículo se basa en las experiencias del programa VLIRED Cuba con la interoperabilidad entre los metadatos de diferentes plataformas y la reutilización de los datos en diferentes silos de información. VLIRED Cuba, un programa VLIR-UOS, comenzó con la implementación de plataformas específicas para la educación, la información y la producción de investigación, respectivamente Moodle, ABCD, DSpace y VIVO. El objetivo final es crear una red integrada de investigación, educación e información. Los autores definieron los conceptos metadatos básicos, sistemas de autoridad e interoperabilidad basados en un estudio de la literatura. La interoperabilidad es permitir que la información que se origina en un contexto se use en otro de

---

<sup>1</sup> Hasselt University Library, Bélgica, marc.goovaerts@uhasselt.be

<sup>2</sup> Universidad de las Ciencias Informáticas (UCI), Cuba, fciudad@uci.cu

<sup>3</sup> Universidad de las Ciencias Informáticas (UCI), Cuba, lcalvarez@uci.cu

<sup>4</sup> Universidad de las Ciencias Informáticas (UCI), Cuba, ltmartin@uci.cu

<sup>5</sup> Universidad de las Ciencias Informáticas (UCI), Cuba, amed@uclv.edu.cu

manera que sea lo más automatizada posible. Pero el éxito general de tales estructuras de datos aplicadas por una amplia variedad de productores de metadatos dependerá de la calidad del contenido de metadatos que proporcionen. VLIRED Cuba elige software de código abierto para desarrollar la red: ABCD para la biblioteca, DSpace y VIVO para información de investigación y Moodle para e-learning. Pero las plataformas usan diferentes formatos de metadatos, respetuosamente MARC, QualifiedDublin Core, un grupo de ontologías y LOM. Para hacer que los diferentes sistemas sean interoperables, el contenido debe describirse de manera estandarizada. Por lo tanto, el proyecto se centró en los elementos clave para estandarizar: personas, organización y palabras clave. Un principio básico es que estas entidades deberían definirse por un valor único. Esto tiene que estar integrado en los diferentes formatos de metadatos. Los sistemas de autoridad entregarán contenido estándar: para las personas ORCID, ID cubana o VIVO ID local; para vocabularios controlados por palabras clave como CCS, AGROVOC y MESH. Finalmente VLIRED Cuba está desarrollando herramientas como EsFácil y EsFácilAuthority para crear metadatos estandarizados en diferentes plataformas. EsFácilAuthority es un sistema de autoridad que puede administrar todo tipo de listas de autoridades y realizar búsquedas a través de una API REST. EsFácil recogerá y enriquecerá los metadatos de publicaciones y los pondrá a disposición para sistemas con diferentes formatos de metadatos.

**Palabras clave:** Metadatos, Interoperabilidad, Control de autoridad, Ambiente de educación, Ambiente de investigación, Sistema de biblioteca, Repositorio institucional, Identificación único.

### **Abstract:**

*This article is based on the experiences in the VLIRED Cuba programme with interoperability between the metadata of different platforms and the reusability of the data in different information silos. VLIRED Cuba, a VLIR-UOS network programme, started with implementing specific platforms for educational, library and information management and research output, resp. Moodle, ABCD, DSpace and VIVO. The final goal is to create an integrated research, education & information network. The authors defined the basic concepts metadata, authority*

*systems and interoperability based on a literature study. 'Interoperability is enabling information that originates in one context to be used in another in ways that are as highly automated as possible'. But the overall success of such data structures being applied by a wide variety of metadata producers will depend on the quality of the metadata content they provide. VLIRED Cuba chooses for open source software to develop the network: ABCD for the library, DSpace and VIVO for research information and Moodle for e-learning. But the platforms uses different metadata formats, respectively MARC, Qualified Dublin Core, a group of ontologies and LOM. To make the different systems interoperable the content should be described in standardized way. The project focused therefore on the key elements to standardize: people, organization and keywords. A basic principle is that these entities should be defined by a unique value. This has to be integrated in the different metadata formats. Authority systems will deliver standard content: for people ORCID, Cuban ID or local VIVO id's - for keywords controlled vocabularies like CCS, AGROVOC and MESH. Finally VLIRED Cuba is developing tools like EsFácil and EsFácil Authority to create standardized metadata in different platforms. EsFácil Authority is an authority system that can manage all type of authority lists and make then searchable through a REST API. EsFácil will collect and enrich publication metadata and make it available for systems with different metadata format.*

**Keywords:** *Metadata, Interoperability, Authority control, Learning environment, Research environment, Library and information systems, Institutional repositories, Unique identifiers.*

## **Introduction**

This article is focusing on the problems of interoperability between the metadata of different platforms and the reusability of the data in the different information silos. The use of standards in metadata and authority control systems is central in this approach.

The authors are involved in VLIRED Cuba, a VLIR-UOS network programme, which is creating a virtual research and educational environment in Cuba using open source software solutions. The development of the VLIRED Cuba is a cooperation of six Cuban universities (Camaguey,

Holguin, Pinar del Rio, UCI, UCLV + University of Oriente) and the Flemish universities coordinated by VLIR-UOS with the support of the Ministry of Higher Education of Cuba and other local stakeholders(Goovaerts, Ciudad Ricardo, & Benitez Erice, 2016).The article is based on the practical experience obtained during the project.

VLIRED Cuba startedwith implementing specific platforms for educational, library and information management and research output, resp. Moodle, ABCD, DSpace and VIVO.The final goal is to create an integrated research, education and information network.

It is not the goal of VLIRED Cuba to create one global platform. Services should be available in the platforms were they are relevant. Teachers and students will use primarily the e-learning environment, but when relevant will find there information from the other platforms. Researchers will use the research platform(s) and get the relevant information there as well.

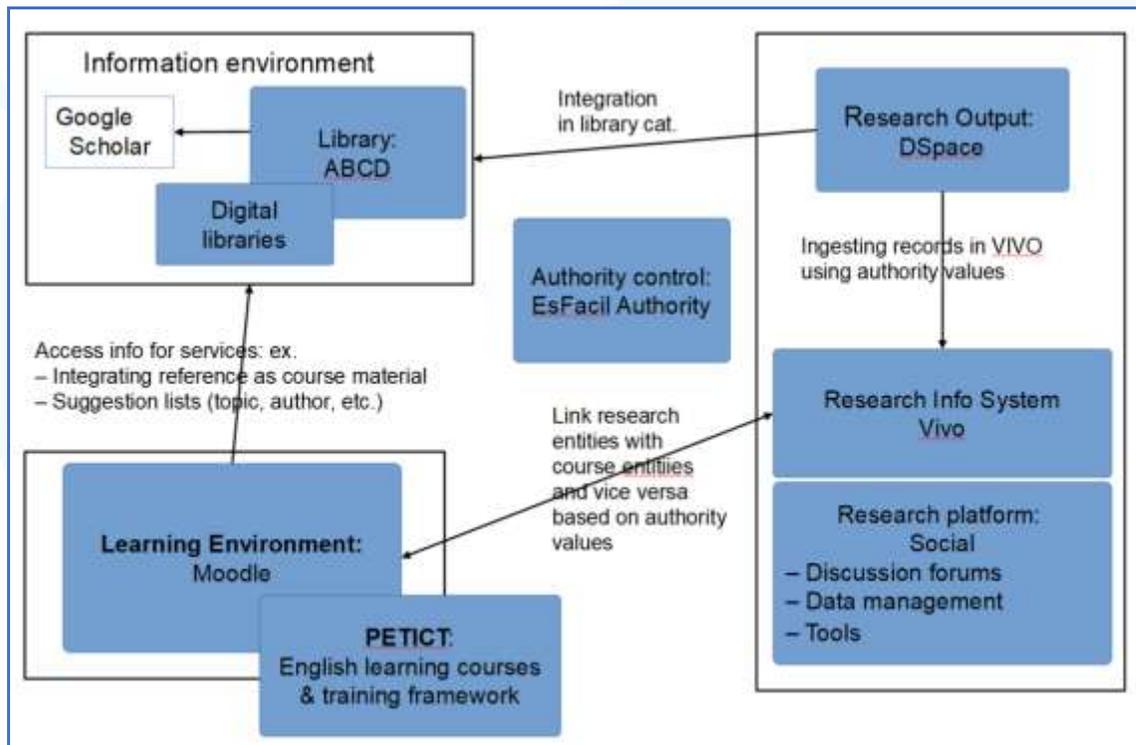


Figure 1. The VLIRED Cuba Virtual Research, Information and Education Network

It is therefore essential to standardize the approach to metadata. Metadata is a key part of the information infrastructure necessary to help create order in the chaos of the Web, infusing description, classification, and organization to help create more useful stores of information (Duval, Hodgins, Sutton, & Weibel, 2002, p. 6). The use of authority systems is an essential way to develop interoperability between different platforms.

## **Methodology**

This part focuses on the concepts of metadata, authority system and interoperability as defined in literature.

There are many definitions of metadata in the existing literature. The basic definition of metadata is data about data. Erik Duval added: Metadata are basically descriptive data. As such, metadata are at the heart of more general developments in the area of digital libraries. Basic metadata elements indicate the title, author, year of publication and similar simple bibliographic data. Richer metadata structures also cover technical features, copyright properties, annotations and so on. The purpose of metadata is 'to facilitate search, evaluation, acquisition, and use' of resources (Duval, 2001, pp. 591–592)

New descriptive metadata schemes such as Dublin Core, offer data structures for interoperability that are a promise for bibliographic control and access issues. But the overall success of such data structures being applied by a wide variety of metadata producers - with an equally wide variety of motivations, concerns and skills - will depend on the quality of the metadata content they provide (Vellucci, 1999). Authority control is central to create quality metadata.

Authority control is the process of selecting one form of a name and recording it, its alternatives, and the data sources used in the process. It is an important tool that boosts recall and precision in the retrieval of information resources. It provides consistency in the form of access points used to identify persons, families, corporate bodies, and subject headings (Sandberg & Jin, 2016, pp. 2–3). Barbara Tillett states from the point of view of a librarian that authority control is necessary

for meeting the catalogue's objectives of enabling users to find the works of an author and to collocate all works of a person or corporate body. (Tillett, 2004).

Interoperability is enabling information that originates in one context to be used in another in ways that are as highly automated as possible. (Duval, 2001, pp. 591–592) Semantics is about meaning; syntax is about form. Agreements about both are necessary for two communities to share metadata. Two communities may agree about the meaning of the term title or creator or identifier, but until they have a shared convention for identifying and encoding values, they cannot easily exchange their metadata (Duval et al., 2002, p. 4).

Finally the semantic web developments and linked data are a way to implement interoperability in a highly automated way. Linked Data is simply about using the Web to create typed links between data from different sources. These may be as diverse as databases maintained by two organisations in different geographical locations, or simply heterogeneous systems within one organisation that, historically, have not easily interoperated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets. (Bizer, Heath, & Berners-Lee, 2011)

The use of unique identifiers for metadata elements is essential in the approach of VLIREC Cuba in the development of interoperability between the different information silos and platforms. An analysis of the different platforms and their use of metadata is the following step .

## **Results**

*The network platforms and their metadata formats.*

VLIREC Cuba has implemented the following platforms: ABCD for library management, DSpace and VIVO for research output management and Moodle as e-learning platform.

1. ABCD 3.0

The higher education authorities in Cuba decided nearly 8 years ago to use open source software ABCD as the library system in every university. Since 2011 the UCI was called by the MES to support the implementation of ABCD in the Cuban universities. Since 2013, VLIREC Cuba, worked on the development of ABCD 3.0, a JAVA version based J-ISIS. A national development team was created, with members from all the network universities and coordinated by UCI and as special collaborators some specialists from INFOMED and IDICT. ABCD 3.0 was released in January 2017 and installed in the network and all MES universities during the same year.

As every traditional library system, the main metadata format of ABCD 3.0 is MARC. MARC was the first computer based standard for bibliographic data (1960s) developed Library of Congress. It is since the 70's an international standard in the library world. There are several versions of MARC in use around the world, the most predominant being MARC 21. (MARC standards, 2017). Still it is based on technology from the 60's and therefore outdated. MARC is, at its heart, a data format built to contain catalogue records; bibliographic items are described via the catalogue records rather than directly via the structured MARC data (Thomale, 2010).

The structure of MARC fields includes indicators and subfield codes:

<p>Indicators</p> <p>First - Type of personal name entry element</p> <p>0 - Forename</p> <p>1 - Surname</p> <p>3 - Family name</p> <p>Subfield Codes</p> <p>\$a - Personal name (NR)</p> <p>\$b - Numeration (NR)</p> <p>\$c - Titles and other words associated with a name (R)</p> <p>\$d - Dates associated with a name (NR)</p> <p>\$e - Relator term (R)</p> <p>\$f - Date of a work (NR)</p> <p>\$g - Miscellaneous information (R)</p> <p>\$j - Attribution qualifier (R)</p> <p>\$k - Form subheading (R)</p> <p>\$l - Language of a work (NR)</p> <p>\$n - Number of part/section of a work (R)</p> <p>\$p - Name of part/section of a work (R)</p> <p>\$q - Fuller form of name (NR)</p> <p>\$t - Title of a work (NR)</p> <p>\$u - Affiliation (NR)</p> <p>\$0 - Authority record control number or standard number (R)</p> <p>\$4 - Relationship (R)</p> <p>\$6 - Linkage (NR)</p> <p>\$8 - Field link and sequence number (R)</p>	<p>100 1# \$a Goovaerts, Marc G.J. \$q (Marc Gustaaf Jeanne), \$d 1956-</p>
--	---

Figure 2. example of MARC field: 100 - Main Entry--Personal Name

Other interesting fields with possibilities of authority control and implementing interoperability are : 110 - Main Entry--Corporate Name and 650 Subject Added Entry.

## 2. DSpace as repository software

All universities of the network are using version 6., but customized to a VLIREC Cuba format. Version. They are availability on Internet or through the national Cuban university network. (UHOLM, UC, UCLV and UPR).

DSpace basic metadata format is still Qualified Dublin Core (QDC), even when this format has been replaced byDCTerms. It is possible to add other metadata formats but it is only possible following the element-qualifier structure. Implementing a richer application format like BIBO or BIBFRAME is therefore not realistic.

<b>dc.contributor.author</b>	Ruano Alvarez, Wilber Alberto	
<b>dc.coverage.spatial</b>	1000003	en_US
<b>dc.date.accessioned</b>	2015-11-24T20:21:38Z	
<b>dc.date.available</b>	2015-11-24T20:21:38Z	
<b>dc.date.issued</b>	2015-09-11	
<b>dc.identifier.citation</b>	Tabares - Martín, L., Peña, F., Mederos, A., Goovaerts, M., Reyes, D., & Álvarez, W. (2015). Software Applications Ecosystem for Authority Control. In E. Garoufallou, R. Hartley & P. Gaitanou (Eds.), <i>Metadata and Semantics Research</i> (Vol. 544, pp. 214-224). Springer International Publishing 214-224	en_US
<b>dc.identifier.uri</b>	<a href="http://dspace.uclv.edu/cu/8089/handle/123456789/3626">http://dspace.uclv.edu/cu/8089/handle/123456789/3626</a>	
<b>dc.description.abstract</b>	Authority control is recognized as an expensive task in the cataloging process. This is actually an active research field in libraries and related research institutions even when several approaches have been proposed in this research area. In this paper, we propose AUCTORITAS, a tool for exposing high value services on the web for the authority control in a generic institution environment. This paper describes the application ecosystem behind AUCTORITAS and how the semantic web languages make possible the semantic integration of heterogeneous applications. Likewise we evaluate the applicability of the proposal for academic libraries.	en_US
<b>dc.description.sponsorship</b>	Proyecto VLIR	en_US
<b>dc.language.iso</b>	en_US	en_US
<b>dc.rights</b>	No se permite la reproducción del original sin autorización de la editorial	en_US
<b>dc.subject</b>	Authority Control	en_US
<b>dc.subject</b>	Linked Data	en_US
<b>dc.subject</b>	Open Data	en_US
<b>dc.subject</b>	Control de Autoridades	en_US

Figure 3. Record in QDC from DSpace@UCLV (<http://dspace.uclv.edu/cu/handle/123456789/3626?show=full>)

Metadata granularity is therefore limited. For example QDC has only identifier citation and DCTermsbibliographicCitation as element. Source, volume, pages and/or article id cannot be described separately, not to mention the description of conferences for proceedings papers. Different extensions have been realized to get a more granular metadata without a standardized approach (Lee, Rho, Lee, & Choi, 2017).

DSpace added from version 1.6 the interesting option of an identifier for an element. Already it is possible to link author names to ORCID. This functionality has been used by AgriOceanDSpace

to store the unique id of terms out of ASFA and AGROVOC thesauri (Goovaerts, Slipetsky, Subirats, & Dister, 2011).

With DSpace, VLIREC Cuba is confronted with two major challenges to resolve the problem of metadata quality and interoperability with other platforms: metadata granularity and implementation of authority systems. Therefore a group was set up to develop a specific submission module, EsFácil including metadata auto-extraction functionalities and the management of authority systems.

### 3. VIVO: a research information system

Research output is larger than publications. To manage the broader scope of research information VLIREC Cuba evaluated the possibilities of using a CRIS system, finally choosing to use VIVO, an open source software. Its semantic web approach was another reason for its selection. The universities of the project are experimenting with the software, specifically how to ingest and manage existing content.

For its (meta)data management VIVO uses a collection of ontologies to represent the research of an organization. The Integrated Semantic Framework ontology modules for VIVO (the VIVO-ISF ontology) provide a set of types (classes) and relationships (properties) to represent researchers and the full context in which they work. VIVO integrates existing ontologies in its Integrated Semantic Framework (Ontology Reference - VIVO 1.9.x Documentation - DuraSpace Wiki, n.d.). To describe publications it uses for example the following ontologies bibo, foaf, vcard, skos and vivo.

VIVO follows the principles of Linked Data. All concepts are defined by and in their relations.

All concepts (people, organisations, projects, publications) are uniquely identified by URIs.

(About VIVO | VIVO, n.d.)

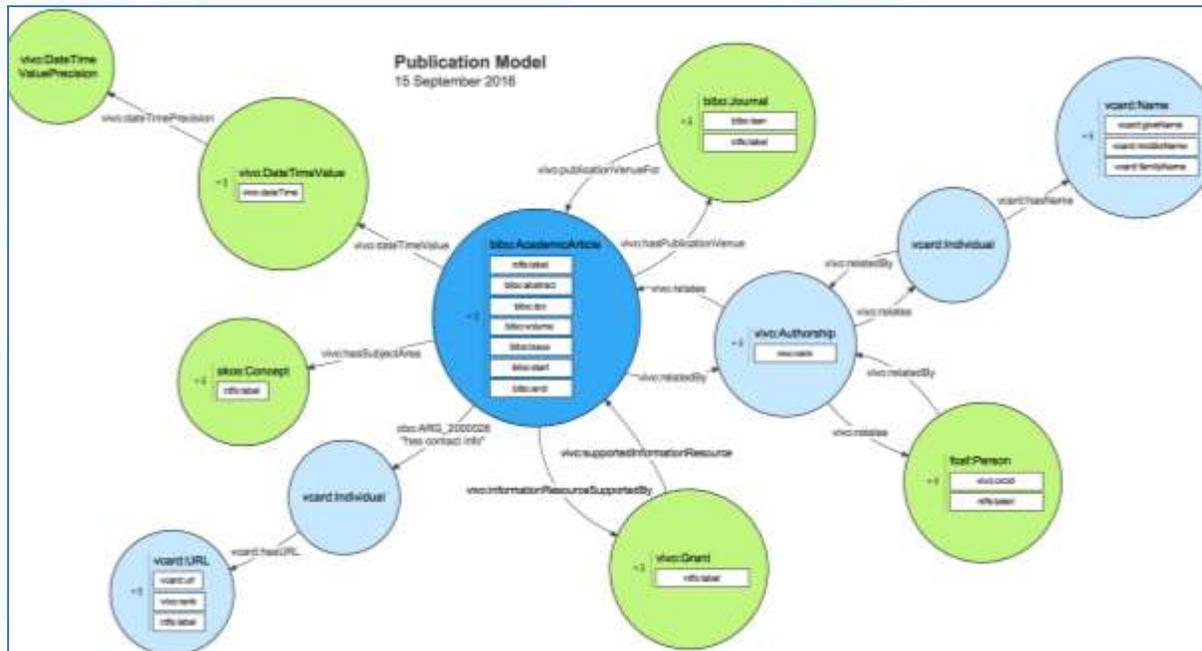


Figure 4. Entities of the publication model of VIVO  
 (<https://wiki.duraspace.org/display/VIVODOC19x/Publication+Model>)

#### 4. Moodle as a e-learning environment

The Cuban Higher Education Ministry (MES) has a national strategy for Virtual Learning Environments (VLE) to support education at every university and at distance, with MOODLE as the preferred platform. VLIREC Cuba has implemented the Moodle platform in the different universities of the network. On the level of the programme a central Moodle has been installed including the doctoral school courses and the English course (<http://redtic.uclv.cu/moodle/>).

Moodle is using LOM, the IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata, as metadata format. It is an internationally recognised open standard (published by the Institute of Electrical and Electronics Engineers Standards Association) for the description of “learning objects”. The LOM data model specifies which aspects of a learning object should be described and what vocabularies may be used for these descriptions; it also defines how this data model can be amended by additions or constraints. (Barker, 2005, pp. 1–5)

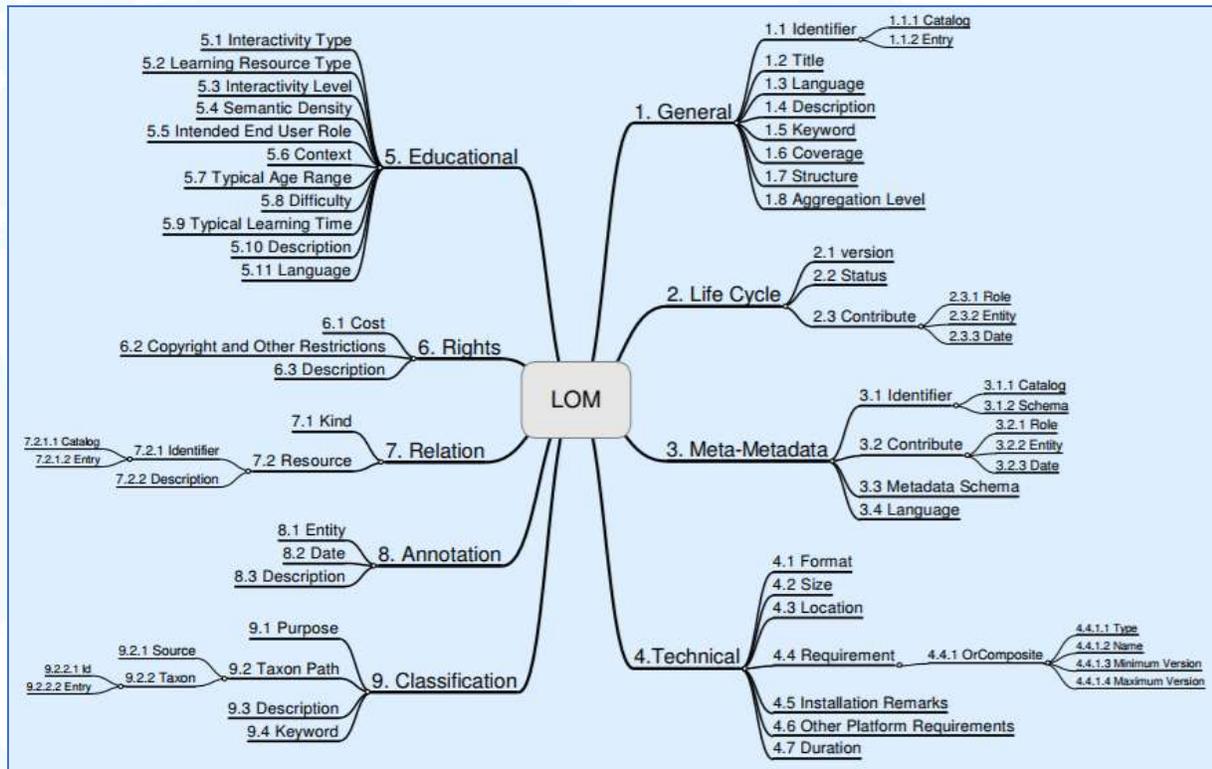


Figure 4. A schematic representation of the hierarchy of elements in the LOM data model (Barker, 2005, p. 2)

### *Interoperability between the different platforms of the information, education and research network VLIRED Cuba*

The different platforms of the network - Moodle for e-learning, ABCD as a library systems, DSpace as repository and VIVO as research information system - have specific functionalities which are valuable for their users. The project does not want to create an extra layer over the existing platforms but want to make exchange of metadata and data possible between them.

ABCD, DSpace, Moodle and VIVO are using different metadata formats and ontologies, because of the culture and necessities of the library, repository and educational communities respectively. But as Erik Duval stated 'Interoperability is enabling information that originates in one context to be used in another in ways that are as highly automated as possible. (E. Duval, 2001, pp. 591–592). A unique metadata format for all the platforms is not realistic. But there is clearly a need for shared convention for identifying and encoding values before metadata can be exchanged.

The VLIREDCuba workgroup on research information systems evaluated the elements and fields relevant for reuse in the different platforms. In the table below, you find the major elements relevant for reuse. In red are the ones that can be defined uniquely.

Moodle Course information	DSpace Publications	VIVO Publications	ABCD Books
<b>Teacher</b> Title Course level Finality Abstract <b>Organization</b> <b>Keywords - subject</b> Date (?)	<b>Authors</b> Title <b>Organization</b> <b>Keywords - subject</b> Source: <b>Journal title (issn)</b> <b>ISBN</b> <b>DOI</b> Date Abstract <b>Publisher</b>	<b>Authors</b> Title <b>Organization</b> <b>Keywords - subject</b> Source: <b>Journal title (issn)</b> <b>ISBN</b> <b>DOI</b> Date Abstract <b>Publisher</b>	<b>Authors</b> Title <b>Organization</b> <b>Keywords - subject</b> <b>ISBN</b> <b>DOI</b> Date Abstract <b>Publisher</b>

Over the different platforms the project focuses on the main elements that define persons, organizations and keywords. How can they be defined in the metadata format of the different platforms? Which authority files and controlled vocabularies can be used to standardize the content? How can unique identifiers be implemented?

Central identification systems for researchers and scientific authors are now implemented. The most important development is since its creation in 2010 ORCID. The core mission of ORCID is to provide a registry of persistent unique identifiers for researchers and scholars. ORCID is an open initiative where the ORCID identifier needs to be integrated into research workflows and linked to information on research activities such as publications, grants, patents, and datasets. (Haak, Fenner, Paglione, Pentz, & Ratner, 2012, pp. 259–264). ORCID has been rapidly accepted by the community. The main scientific publishers are partners in the organization. Publishers request the ORCID ID from authors when submitting articles. Funding organizations in Europe and the US request it too.

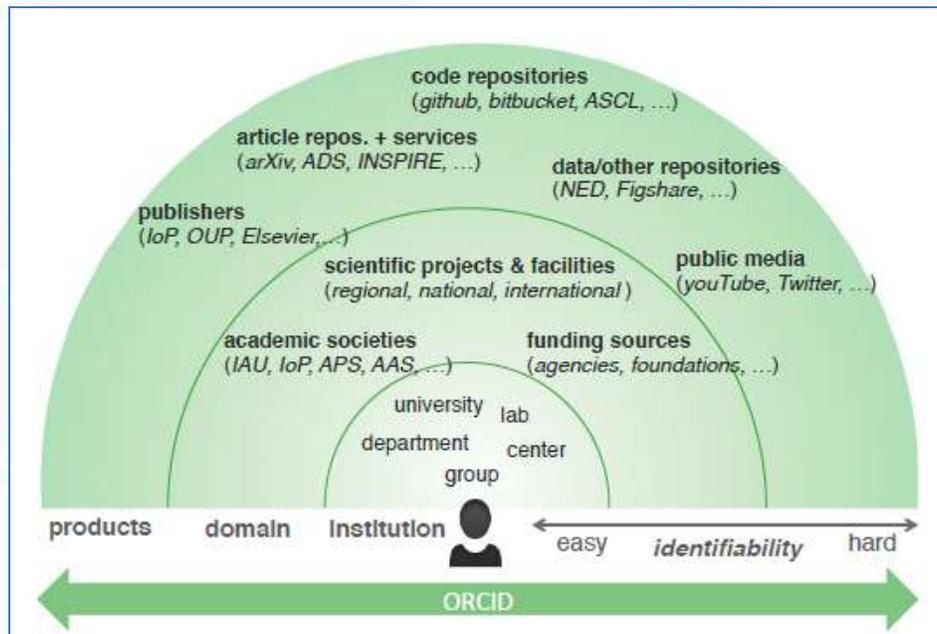


Figure 6.: Persistent, Global Identity for Scientists via ORCID (arXiv:1502.06274v1 [cs.DL])

Cuban researchers and academics can register in ORCID, simply to be part of that international community. The project is promoting ORCID in the different universities. Still, registration is a personal action.

To create an authority list of the authors of the network universities and in a broader context of Cuban authors, a more proactive approach has been developed: by collecting data from different resources: Moodle and DSpace of the network universities, ASSETS, IDICT databases, the Tocatoro project from UPR and international databases like Web of Science, Scopus, Google Scholar, PUBMED, Researchgate, Scielo, .... . The challenge is to collect this data and curate it. As a test in Web of Science the project group could download already 1508 publications from UCLV researchers. The idea is to upload this data in VIVO after curation. VIVO automatically generates an ID for every object, specifically for every person and organization. Other ID's can be added, in the first place the ORCID and if developed a Cuban ORCID.

The platforms have to use the same authority file for researchers and academics. A unique ID should be added to the metadata formats. In the following table a proposal of implementation is presented.

Moodle – LOM	DSpace - QDC	VIVO - Ontology	ABCD - MARC 21
2.3.1.Role: Adding an id attribute to the element	contributor.*: Adding an authority value	foaf: person with URI	e.g. Field 100 with subfield codes: \$4 - Relationship (R) \$6 - Linkage (NR) \$8 - Field link and sequence number ®

A way to define content unequivocally is by using controlled vocabularies. Different thesauri are translated to SKOS (“SKOS Simple Knowledge Organization System - home page,” n.d.). The project already uses CCS, AGROVOC, MESH. Other thesauri can be added. If they are not available in SKOS format, they can be transformed using tools like VocBench(VocBench | Agricultural Information Management Standards (AIMS),n.d.).

Metadata elements for subject should be using the same thesauri as authority files.

Moodle LOM	DSpace - QDC	VIVO - Ontology	ABCD - MARC 21
1.5 Keywords: adding an ID	subject with authority value (and vocabulary as a qualifier ?)	Skos:Concept with URI	e.g. Field 650 with subfield codes: \$4 - Relationship (R) \$6 - Linkage (NR) \$8 - Field link and sequence number (R)

EsFácil Authority is the tool developed by the project to manage the access to different vocabularies and authority systems locally. EsFácil Authority tool is fully implemented based on Free and Open Source Software (FOSS), which allows to share EsFácil Authority with the international community as FOSS through GitHub: <https://github.com/ELINF-Cuba-Network>.

EsFácil Authority aims to create an applications ecosystem enabling authority control capacities for external applications, by reusing semantically structured data shared by different institutions. It aspires to facilitate the processing of authority data in a standardized fashion, following the principles of Linked Data.

EsFácil Authority exposes vocabularies in SKOS format such as AGROVOC, covering all areas of interest of the Food and Agriculture Organization of the United Nations with over 32000 concepts, CCS vocabulary for Computer Sciences and MESH for Medicine and Life Sciences.

It also can integrate data from VIVO, which can be used as authority system for persons and organizations. EsFácil Authority queries to VIVO are done through VIVO's SPARQL endpoint.

The EsFácil Authority interface has four main functionalities exposed as REST web services:

- Search for personal authors information.
- Search for corporate authors information.
- Retrieve registered controlled vocabularies list.
- Search for an authorized term on a specified controlled vocabulary.

External applications, ABCD, DSpace, VIVO can send requests to EsFácil Authority, then EsFácil Authority queries its available information sources and retrieves the requested information structured as a XML or JSON. Figure 7 shows an answer to a query on the term “database” on the ACM Controlled Vocabulary.

```
<?xml version="1.0"?>
<vocabularyEntry>
<identifier>http://totem.semedica.com/taxonomy/The ACM Computing Classification System (CCS)#10002952</identifier>
<authorizedTerm>Data management systems</authorizedTerm>
</vocabularyEntry>
```

Fig. 7. EsFácil Authority answer to a query over ACM controlled vocabulary

Two main elements are sent as answer in this case, the identifier of the term in the requested vocabulary and the authorized term by itself. The identifier of the term is computer oriented for uniquely identify it by using an URI and the authorized term is what the person using the system sees. (Martín et al., 2015, pp. 214–224)

A logic step after the development of the authority system is the development of EsFácil as a submission module to create rich metadata. This is still work in progress but the main components, automated metadata extraction with DarkAIV and authority control (EsFácilAuthority) are realized. EsFácil is conceived as an independent module that will ingest records and files in repositories and information systems, in the first place in DSpace.

The idea is to store rich metadata and translate it to the format of the target applications. Internally, EsFácil uses a nosql approach using the JSONB type of Postgresql, with CSL JSON as metadata format.

The Citation Style Language (CSL) is an XML-Based stylesheet language for formatting of citations and bibliographies. The choice for CSLJSON, , also known as Citeproc JSON implementation (CSL-JSON — citeproc-js 1.1.73 documentation,n.d.), is based on the fact that it is a rather simple but rich format and that it is already used by services like Zotero, Mendeley, CiteProc, Crossref and Islandora(Citation Style Language,n.d.).

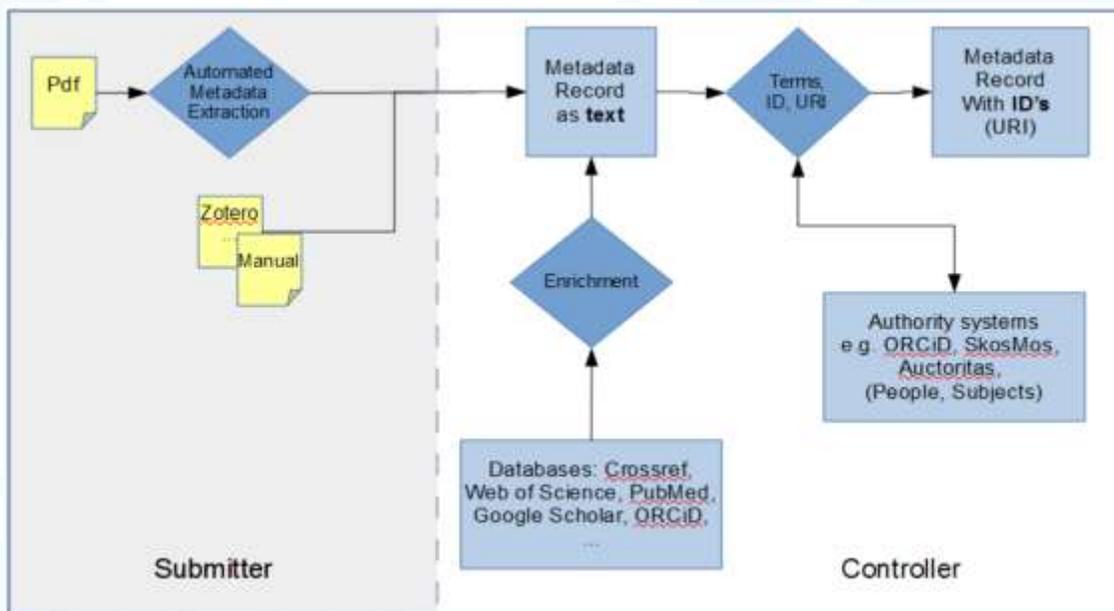


Figure 8: Extracting and enriching metadata from research documents

Metadata in CSL JSON for items have a format as described in figure 9. Crossref added an ORCID and an affiliation tag. EsFácil proposes to extend it with a URI tag. Another important element is subject. The tag subject can only contain the terms. To enrich it a new tag is created 'subject-id' which is enriched with a key as well as with the source.

CSL is very flexible. By extending the functionalities of CSL EsFácil still can import existing metadata from various resources like like Crossref, Zotero and Mendeley and at the same time using the EsFácil Authority tool to enrich it and present it through webservices to the information platforms of VLIREC Cuba: ABCD, DSpace, VIVO and Moodle.

<pre> CSL JSON  "author": [   {     "family": "de las Casas",     "given": "Bartolomé"   },   {     "family": "King",     "given": "Rev. Martin Luther Jr., Ph.D."   } ] </pre>	<pre> CSL JSON with EsFácil extension  "author": [   {     "family": "Machado Rivero",     "given": "Manuel Osvaldo",     "suffix": "Ph.D.",     "uri": "https://orcid.org/0000-0001-5095-8389"   },   {     "family": "Alvarez Fernández",     "given": "Luis Carlos",     "suffix": "Eng.",     "uri": "https://www.linkedin.com/in/lcav%C3%A1ndez"   } ] </pre>
<pre> "subject": ["Ocean"], </pre>	<pre> "subject-id": [   {     "value": "Ocean",     "key": "http://artemide.art.uniroma2.it:8081/agrovoc/c_39065",     "key-asserted-by": "AGROVOC"   } ] </pre>

Figure 9: Examples of CSL JSON and CSL JSON with EsFácil extension.

## Conclusion

The simplest solution to create interoperable metadata in different platforms is to use the same formats everywhere. In reality library, education and research have developed over time specific formats adapted to the methodologies of their domain. The platforms in VLIREC Cuba uses MARC, Qualified Dublin Core, LOM and different ontologies.

The approach of the project is to search for the main common elements used by the different platforms. The project workgroup identified persons, organizations and keywords as main common elements. The quality of the content depends on the same definition of the elements, preferably by using unique identifiers for every entity.

Modern authority systems deliver unique id's. For persons ORCID is an international standard, but national or local id's delivered by the research unit are relevant as well. There are initiatives to register organizations initiated by ORCID and GRID (Global Research Identifier Database - <https://www.grid.ac/>). For keywords there is growing availability of linked data based controlled vocabularies.

VLIRED Cuba is developing the tools EsFácil and EsFácilAuthority to create standardized metadata in different platforms, specifically ABCD, Moodle, DSpace and VIVO. But the tools can be integrated in other software packages too.

These efforts guarantees standardized metadata in the different platforms. The next step in the project is to create services that integrates the data from the different platforms. In the philosophy of the project this will not replace the existing platforms because each of them have their functionalities and users.

Different strategies are possible. VIVO for example, ingest all relevant information in its database. More in general the metadata will have to be exposed to other systems, preferably using a Linked Data approach. Finally in the project some universities are working to create relations between different platforms. The Moodle research group at the University of Pinar del Riohasfor example developed modules in Moodle that can access DSpace(Goovaerts et al., 2016, p. 14). More developments on services are expected to be realized in the following years.

## **Bibliography**

About VIVO | VIVO. (n.d.). Retrieved December 9, 2017, from

[ttp://www.vivoweb.org/info/about-vivo](http://www.vivoweb.org/info/about-vivo)

Barker, P. (2005). What is iee learning object metadata/ims learning resource metadata. *CETIS Standards Briefing Series, JISC (Joint Information Systems Committee of the Universities' Funding Councils)*.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked Data: The Story So Far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205–227.

<https://doi.org/10.4018/978-1-60960-593-3.ch008>

Citation Style Language. (n.d.). Retrieved December 10, 2017, from <http://citationstyles.org/>

CSL-JSON — citeproc-js 1.1.73 documentation. (n.d.). Retrieved December 10, 2017, from

<http://citeproc-js.readthedocs.io/en/latest/csl-json/markup.html>

Duval, E. (2001). Metadata standards: What, who & why. *Journal of Universal Computer Science*, 7(7), 591–601.

Duval, Erik, Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4). <https://doi.org/10.1045/april2002-weibel>

Goovaerts, M., Ciudad Ricardo, F. A., & Benitez Erice, D. (2016). Desarrollo de un red virtual de investigación y educación para la información científico en Cuba. Presented at the Info 2016, Habana, Cuba. Retrieved from <http://www.congreso-info.cu/index.php/info/2016/paper/view/139/48>

Goovaerts, M., Slipetsky, D., Subirats, I., & Dister, S. (2011). AGRI-OCEAN DSpace : FAO and UNESCO-IOC/IODE combine efforts in their support of open access. In *IAMSLIC Conference Proceedings 2011*. Zanzibar, Tanzania,. Retrieved from <http://darchive.mblwhoilibrary.org/handle/1912/5327>

Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*, 25(4), 259–264. <https://doi.org/10.1087/20120404>

Lee, M., Rho, J.-H., Lee, E.-J., & Choi, Y. K. (2017). The Development of Application Profile for the OAK Institutional Repository. *International Conference on Dublin Core and Metadata Applications*, 0, 100–103.

MARC standards. (2017, May 13). In *Wikipedia*. Retrieved from

[https://en.wikipedia.org/w/index.php?title=MARC\\_standards&oldid=780104200](https://en.wikipedia.org/w/index.php?title=MARC_standards&oldid=780104200)

Martín, L. T., Peña, F. O. F., Mederos, A. A. L., Goovaerts, M., Reyes, D. C., & Álvarez, W. A.

R. (2015). Software Applications Ecosystem for Authority Control. In *Metadata and Semantics Research* (pp. 214–224). Springer, Cham. [https://doi.org/10.1007/978-3-319-24129-6\\_19](https://doi.org/10.1007/978-3-319-24129-6_19)

Ontology Reference - VIVO 1.9.x Documentation - DuraSpace Wiki. (n.d.). Retrieved December 8, 2017, from <https://wiki.duraspace.org/display/VIVODOC19x/Ontology+Reference>

Sandberg, J., & Jin, Q. (2016). How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World. *Cataloging & Classification Quarterly*, 54(8), 537–552. <https://doi.org/10.1080/01639374.2016.1238429>

SKOS Simple Knowledge Organization System - home page. (n.d.). Retrieved December 10, 2017, from <https://www.w3.org/2004/02/skos/>

Thomale, J. (2010). Interpreting MARC: Where's the Bibliographic Data? *The Code4Lib Journal*, (11). Retrieved from <http://journal.code4lib.org/articles/3832>

Tillett, B. B. (2004). Authority Control: State of the Art and New Perspectives. *Cataloging & Classification Quarterly*, 38(3–4), 23–41. [https://doi.org/10.1300/J104v38n03\\_04](https://doi.org/10.1300/J104v38n03_04)

VocBench | Agricultural Information Management Standards (AIMS). (n.d.). Retrieved December 10, 2017, from <http://aims.fao.org/vest-registry/tools/vocbench>