

## Reusability of structured information: metadata and authority systems

Marc Goovaerts (Hasselt University Library) - [orcid.org/0000-0003-1809-4159](https://orcid.org/0000-0003-1809-4159)

CYTDES 2017, Camaguey, 6 June 2017

### Abstract

Information is now at reach for everybody everywhere on the internet. But originally the library was the centre of information. It has a tradition in the development of formats for description of publications and related materials. MARC is the bibliographic standard in the library even nowadays. Classification systems and thesauri helps to structure information. ISSN and ISBN are early systems of unique identification.

The internet saw an exponential growth of information. New metadata schema's and ontologies have been developed with Dublin Core as the most notorious format. On the Internet scientific information is expanding exponentially, with publications and other forms of information. A main feature of the internet is the capacity of linking. Therefore information has to be described as objects that are uniquely identified and can be interpreted by machines. While most of the internet metadata is still based on text, the philosophy of Linked Data is changing this approach.

The VLIR-UOS Network Cuba project is building an information network where information objects can be integrated and reused. It links library systems with learning environments, repositories and research information platforms. It is developing connections between different types of information: people, organizations, projects, courses and publications. Central in this approach are the authority tools, vocabularies, thesauri and ontologies. The development on eSFàcil, a submission module for DSpace, is an example of the implementation of interoperability in an information network. The article finishes with the problem of unique identification of people over the different platforms and possible solutions.

**Keywords:** metadata, thesaurus, authority systems, unique identifiers, Network Cuba, VLIR-UOS,

## Information: from text to related objects

Information is now at reach for everybody everywhere on the internet. But originally the library was the centre of information. Libraries have a century long tradition of describing publications and other library objects. This knowledge about metadata, authority systems and vocabularies is still relevant in the time of electronic information, even if it is going through a period of transition.

In the 1960s MARC (Library of Congress, n.d.) became the first computer based standard for bibliographic data supported by the Library of Congress. Originally it was developed to print out catalogue cards. The MARC format evolved and got new implementations like MARC 21, UNIMARC, MARCXML. Nowadays it is a standard in libraries, but it is clearly an outdated technology that does not take in account much of the evolutions in computer science like for example the relation model. (Thomale, 2010)

```
LDR          *****nam##22*****a#4500
001          <control number>
003          <control number identifier>
005          19920331092212.7
007          ta
008
820305s1991#### nyu#### ##### #001#0# eng##
020  ##      $a0845348116 :$c$29.95 (£19.50 U.K.)
020  ##      $a0845348205 (pbk.)
040  ##      $a<organization code>$c<organization code>
050  14      $aPN1992.8.S4$bT47 1991
082  04      $a791.45/75/0973$219
100  1#      $aTerrace, Vincent,$d1948-
245  10      $aFifty years of television :$ba guide to series and pilots, 1937-1988 /$cVincent Terrace.
246  1#      $a50 years of television
260  ##      $aNew York :$bCornwall Books,$cc1991.
300  ##      $a864 p. ;$c24 cm.
500  ##      $aIncludes index.
650  #0      $aTelevision pilot programs$zUnited States$vCatalogs.
650  #0      $aTelevision serials$zUnited States$vCatalogs.
```

Fig. 1. Example of MARC 21 record for a book

The library community is clearly aware of the problems of the MARC format. Less complex metadata format were developed like the XML-based MODS, made available in 2002. (Guenther, 2003). BIBFRAME is the latest evolution in metadata formats related to MARC following the linked data principles.

```

<titleInfo>
<title>Hiring and recruitment practices in academic libraries</title>
</titleInfo>
<name type="personal">
<namePart>Raschke, Gregory K.</namePart>
<displayForm>Gregory K. Raschke</displayForm>
</name>
<typeOfResource>text</typeOfResource>
<genre>journal article</genre>
<originInfo>
<place>
<placeTerm>Baltimore, Md.</placeTerm>
</place>
<publisher>Johns Hopkins University Press</publisher>
<dateIssued>2003</dateIssued>
<issuance>monographic</issuance>
</originInfo>
<language><languageTerm authority="iso639-2b">eng</languageTerm></language>
<physicalDescription>
<form authority="marcform">print</form>
<extent>15 p.</extent>
</physicalDescription>
<abstract>Academic libraries need to change their recruiting and hiring procedures to stay competitive in today's changing marketplace. By taking too long to find and to hire talented professionals in a tight labor market, academic libraries are losing out on top candidates and limiting their ability to become innovative and dynamic organizations.

```

Fig. 2. Example of MODS record for a book

Classification systems are used to classify nonfictional publications into subject categories. The subject is indicated by a three-digit numeral and further specification is given by numerals following a decimal point. Classification systems like LCC (Library of Congress, n.d.), UDC (UDC Consortium - About UDC, n.d.) and DDC (Dewey Services, n.d.) are also used as a way to place the books in the library.

One of the limitations of a classification is its hierarchical structure. Thesauri are more adapted to the linking world. As a reference work of terms, which can be global like the Library of Congress Subject Headings (Library of Congress, n.d.) or more specific like MESH (Medicine and life science) (NIH, n.d.), AGROVOC (Agriculture) (FAO of the United Nations, n.d.) and ASFA (Aquatic science) (FAO of the United Nations, n.d.), they show concepts and their relations.

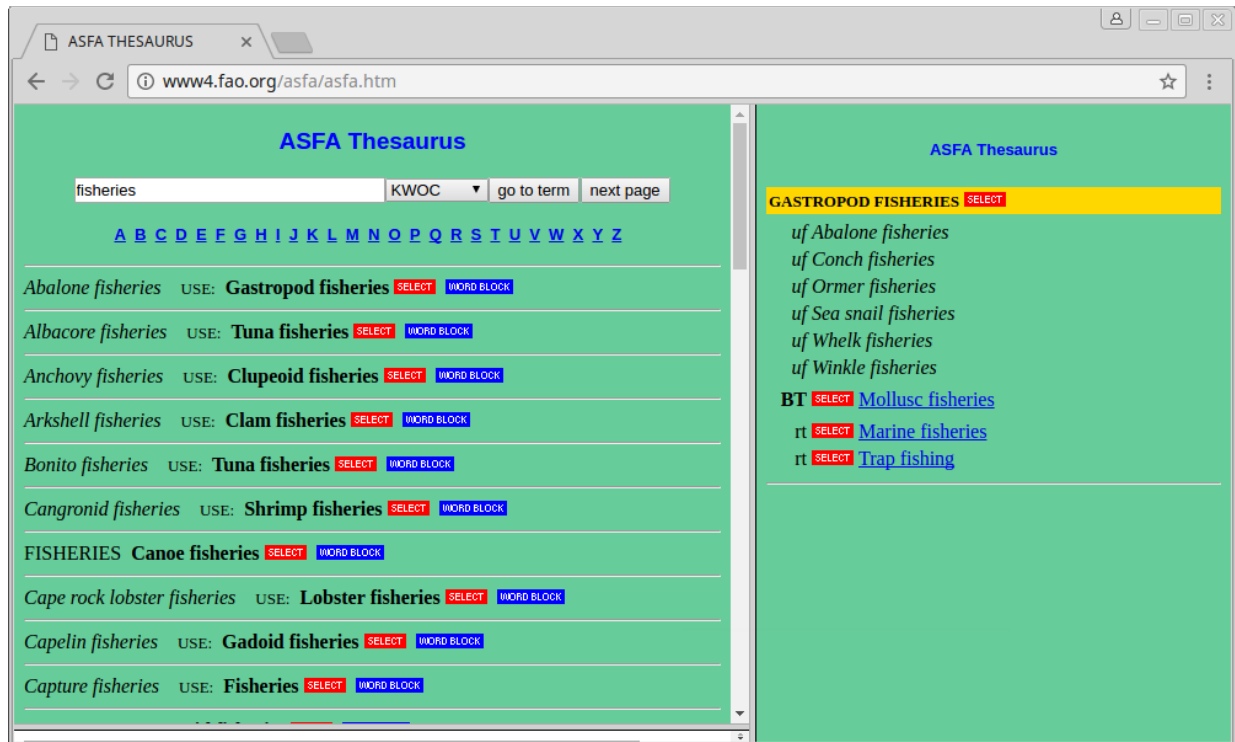


Fig. 3. ASFA Thesaurus. Terms with their relationships

The interest in thesauri and vocabularies is again growing because its similarity with the linking principles of the internet.

Another aspect of description is the problem of unique identification. ISBN and ISSN were developed in the past, respectively for books and journals, for commercial and distribution reasons. The Journal of Algebra has for example the ISSN number of 0021-8693.

In classic bibliographic formats metadata is mostly text based. It has the classical ambiguity problems. Names on publications can have different versions as well as misspellings. In the catalogue of the library of Antwerp University the name Karl Marx is written as follow: Marx Marx, K. Marx, Karl and even Marx, Carl. People with the same name cannot be identified. For concepts we have similar problem (homonyms and synonyms) besides translation issues as shown in the ASFA example.

Many indicators show an exponential growth of information on the internet. (Hobbes' Internet Timeline - the definitive ARPAnet & Internet history, n.d.). This is as well the case of scientific information.

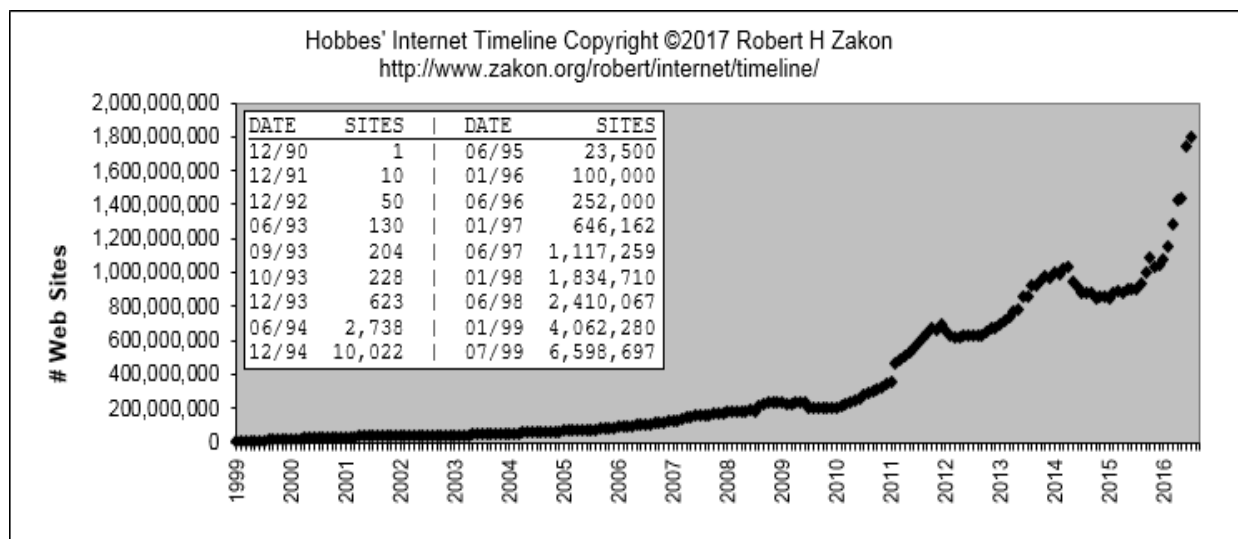


Fig. 4. The exponential growth of web sites between 1990 and 2016.

For example in the Open Access world, the collection harvested by the Base Search Engine has grown from 679.311 documents in September 2004 to 111.400.979 in May 2017. (BASE - Bielefeld Academic Search Engine | Statistics, 2017)

Still, the metadata behind the scientific databases and search engines is still mostly text based. Repository software like DSpace and Eprints manage mostly text as can be seen in the example of Apollo, the Cambridge University repository.

dc.contributor.author	Zhang, Yi
dc.contributor.author	Hatch, Kim A.
dc.contributor.author	Wernisch, Lorenz
dc.contributor.author	Bacon, Joanna
dc.date.accessioned	2011-06-16T16:29:55Z
dc.date.available	2011-06-16T16:29:55Z
dc.date.issued	2008-02-22
dc.identifier.citation	BMC Genomics 2008, 9:87
dc.identifier.uri	http://www.dspace.cam.ac.uk/handle/1810/237985
dc.identifier.uri	http://dx.doi.org/10.1186/1471-2164-9-87
dc.description	RIGHTS : This article is licensed under the BioMed Central licence at <a href="http://www.biomedcentral.com/about/license">http://www.biomedcentral.com/about/license</a> which is similar to the 'Creative Commons Attribution Licence'. In brief you may : copy, distribute, and display the work; make derivative works; or make commercial use of the work - under the following conditions: the original author must be given credit; for any reuse or distribution, it must be made clear to others what the license terms of this work are.
dc.description.abstract	<b>Abstract</b> Background Low oxygen availability has been shown previously to stimulate <i>M. tuberculosis</i> to establish non-replicative persistence in vitro. The two component sensor/regulator dosRS is a major mediator in the transcriptional response of <i>M. tuberculosis</i> to hypoxia and controls a regulon of approximately 50 genes that are induced under this condition. The aim of this study was to determine whether the induction of the entire DosR regulon is triggered as a synchronous event or if induction can unfold as a cascade of events as the differential expression of subsets of genes is stimulated by different oxygen availabilities. Results A novel aspect of our work is the use of chemostat cultures of <i>M. tuberculosis</i> which allowed us to control environmental conditions very tightly. We exposed <i>M. tuberculosis</i> to a sudden drop in oxygen availability in chemostat culture and studied the transcriptional response of the organism during the transition from a high oxygen level (10% dissolved oxygen tension or DOT) to a low oxygen level (0.2% DOT) using DNA microarrays. We developed a Bayesian change point analysis method that enabled us to detect subtle shifts in the timing of gene induction. It results in probabilities of a change in gene expression at certain time points. A computational analysis of potential binding sites upstream of the DosR-controlled genes shows how the transcriptional responses of these genes are influenced by the affinity of these binding sites to DosR. Our study also indicates that a subgroup of DosR-controlled genes is regulated indirectly. Conclusion The majority of the dosR-dependent genes were up-regulated at 0.2% DOT, which confirms previous findings that these genes are triggered by hypoxic environments. However, our change point analysis also highlights genes which were up-regulated earlier at levels of about 8% DOT indicating that they respond to small fluctuations in oxygen availability. Our analysis shows that there are pairs of divergent genes where one gene in the pair is up-regulated before the other, presumably for a flexible response to a constantly changing environment in the host.
dc.language.iso	en
dc.title	A Bayesian Change point model for differential gene expression patterns of the DosR regulon of <i>Mycobacterium tuberculosis</i>

Fig. 5. Example Record from Apollo, Cambridge University Repository.

The most common metadata format on the internet is Dublin Core (1995) and DCTerms (DCMI Metadata Basics) It has a core set of 15 elements for simple resource description: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights. Through application profiles DC can be enriched with other specialized vocabularies.

An example of such an application profiles is RIOXX. This application profile is developed to support research management standards in the UK. It includes extra attributes like an ORCID identifier for authors.

```
<rioxxterms:contributor id="http://orcid.org/0000-0002-1395-3092">
  Lawson, Gerald
</rioxxterms:contributor>
```

Fig. 6. RIOXX Metadata Application Profile (RIOXX, 2017): extension of the contributor element with an ORCID attribute.

Other very common services of scientific databases are the possibility to download metadata in reference manager formats like RIS, BibTex or Endnote.

But on the internet scientific information does not limit itself to publications. Current Research Information Systems collect information about organizations, people, projects. Learning systems are growing. There are specific scientific platforms like ResearchGate, Mendeley, Zotero. The availability of data sets is growing.

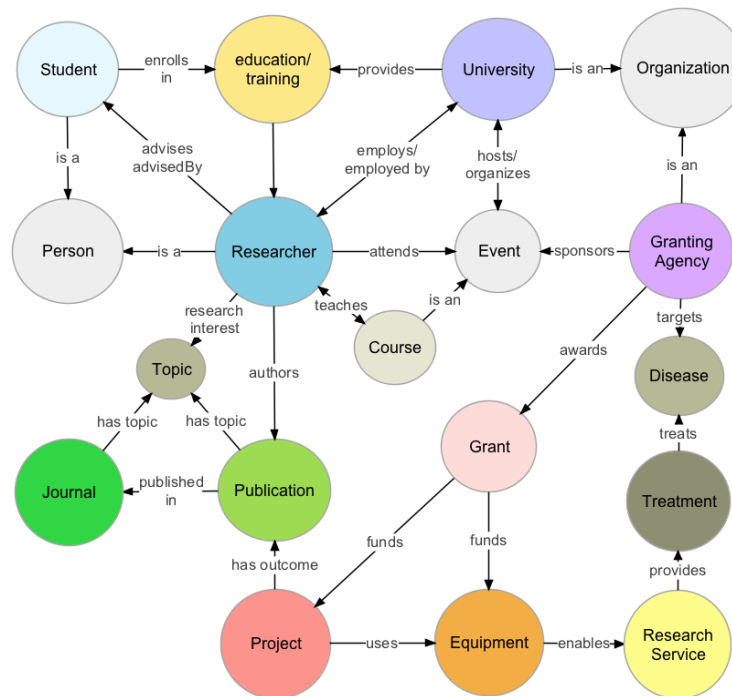


Fig. 7. The basic entities of VIVO and their relations

VIVO is an open source research Information software for description of the research of an organization: institutes and departments, people, projects, publications, ...

It follows the principles of linked data. All concepts (people, organizations, projects, publications) are uniquely identified by URIs. All concepts are defined by and in their relations using an RDF model. (Valeria Pesce, 2014)

Figshare is an example of a data repository in the cloud. By encouraging publishing of figures, charts, data, rather than being limited to the traditional entire 'paper', knowledge can be shared more quickly and effectively. (figshare - About, 2017)

The internet is moving from a network that links pages to a network that links data: Web of Data. The term denotes the evolution of the Web into an ecosystem of interconnected data and information contributed by individuals, governments, businesses and machines (e.g. sensors). The real value of the Web of Data lies in the relationships between the data. These relationships (referred to as links) put data in context and enrich their meaning and expressiveness. This means that putting the data online is not enough. Publishers need to ensure that data is made available in both human- and machine-understandable formats and is linked to other data. (Understanding Linked Data by example, n.d.)

To describe data on the internet new ontologies are created. For example schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond. Its core vocabulary currently consists of 589 types (e.g. creative work, book, movie, event, place, person), 860 Properties, and 114 enumeration values. (Home – schema.org, 2017).

```
<info:doi/10.1134/S0003683806040089> a bibo:Article ;
  dc:title "Effect of argillaceous minerals on the growth of phosphate-mobilizing bacteria
    Bacillus subtilis"@en ;
  dc:date "2006-01-01" ;
  dc:isPartOf <urn:issn:23346587> ;
  bibo:volume "42" ;
  bibo:issue "4" ;
  bibo:pageStart "388" ;
  bibo:pageEnd "391" ;
  dc:creator <http://examples.net/contributors/2> ;
  dc:creator <http://examples.net/contributors/1> ;
  bibo:authorList ( <http://examples.net/contributors/2>
    <http://examples.net/contributors/1> ) .

<urn:issn:23346587> a bibo:Journal ;
  dc:title "Applied Biochemistry and Microbiology"@en ;
  bibo:shortTitle "App Biochem and Biol"@en .
```

Fig. 8. bibo: Article description - example

Bibo, the Bibliographic Ontology Specification provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books,

articles, etc) on the Semantic Web. (Bibliographic Ontology Specification, 2017) In a semantic web environment information objects are identified by unique identifiers, URIs as we can see in the bibo example.

The first use of URIs for scientific information on the internet was with DOIs (digital object identifier) mostly used for publications and dataset. DOIs are a solution for the unstability of urls on the internet. (The DOI System, 2017)

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. (Open Access Academy, n.d.)

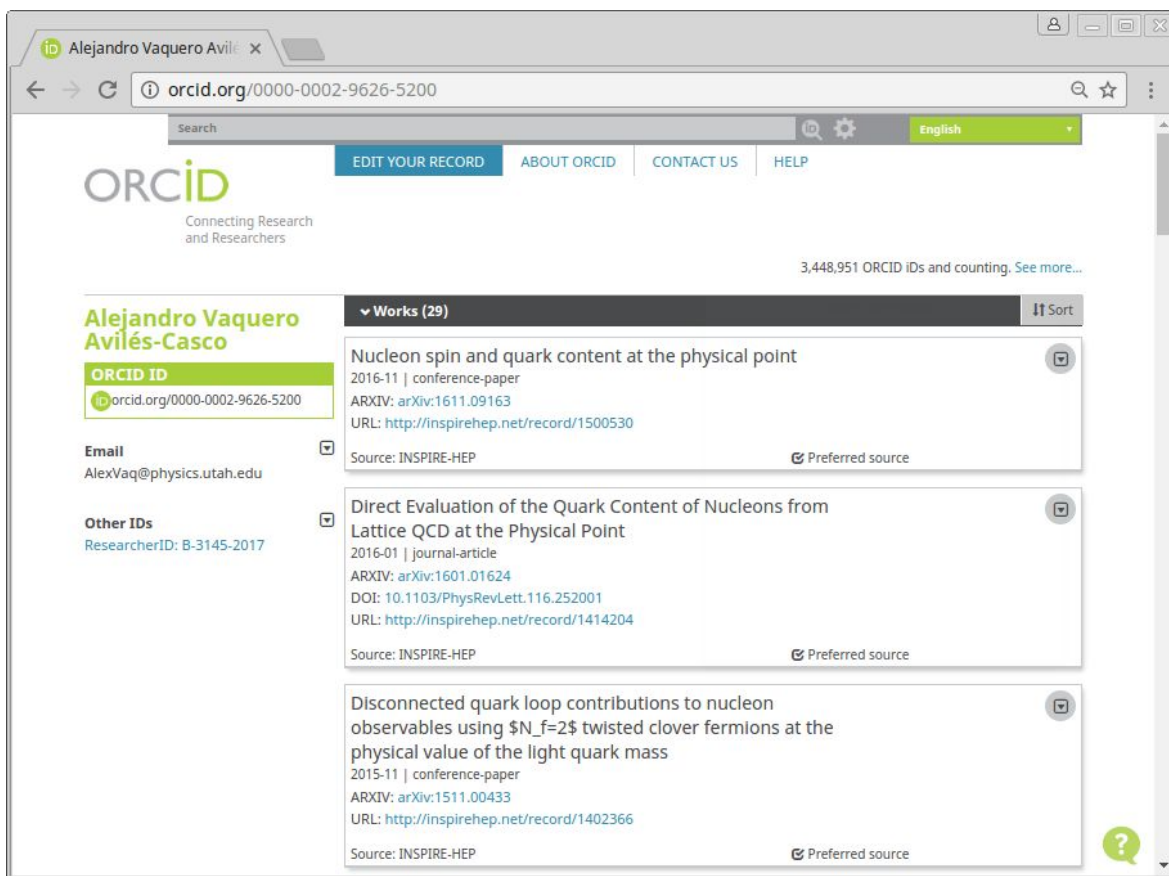


Fig. 9. The ORCID page of an author with his linked publications

The tools to identify subjects of information are for librarians thesauri and vocabularies. In the semantic world they get a new life. For example FAO has implemented AGROVOC as a linked data tool where every concept receives a URI and is described in RDF. (FAO of the United Nations, n.d.)

Property	Value
rdf:type	skos:Concept
skos:inScheme	http://aims.fao.org/aos/agrovoc http://voc.landportal.info/landterms
skos:topConceptOf	http://voc.landportal.info/landterms
skos:broader	http://aims.fao.org/aos/agrovoc/c_330888
skos:narrower	http://aims.fao.org/aos/agrovoc/c_331453 http://aims.fao.org/aos/agrovoc/c_3873 http://aims.fao.org/aos/agrovoc/c_2173 http://aims.fao.org/aos/agrovoc/c_331367 http://aims.fao.org/aos/agrovoc/c_9001025 http://aims.fao.org/aos/agrovoc/c_2674 http://aims.fao.org/aos/agrovoc/c_7035 http://aims.fao.org/aos/agrovoc/c_1696 http://aims.fao.org/aos/agrovoc/c_4987 http://aims.fao.org/aos/agrovoc/c_550 http://aims.fao.org/aos/agrovoc/c_4611 http://aims.fao.org/aos/agrovoc/c_37965 http://aims.fao.org/aos/agrovoc/c_644 http://aims.fao.org/aos/agrovoc/c_5659
skos:exactMatch	http://aims.fao.org/aos/asfa/c_5834 http://lod.nal.usda.gov/nalt/28131 http://d-nb.info/gnd/4017310-0 http://cat.aii.caas.cn/concept/c_55300
skos:broadMatch	http://cat.aii.caas.cn/concept/c_2890

prefLabel	altLabel	Lang
рыбное хозяйство	промышленное рыболовство профессиональное рыболовство товарное рыболовство	ru
渔业	捕捞业 商业渔业	zh
Pêches	Pêche professionnelle Pêche industrielle Industrie des pêches Pêche commerciale	fr
rybárstvo	komerčný rybolov rybný priemysel	sk
ជលផល		km
Pesca	Industria pesquera	es
การประมง	ธุรกิจการประมง	th
การประมง	อุตสาหกรรมประมง	th
మత్స్య శాస్త్రం	వేపలు పట్టి పరిశ్రమ	te

Fig. 10. An AGROVOC term as an URI.

The internet has the capacity to create relations between information (objects). Using unique identifiers makes information machine-readable. URI's are the most logical solution in the context of the internet. Authority systems are a way of standardizing the description of the content. It helps with deduplication of names, subjects, etc. But objects can be defined in different authority systems. Persons are identified with ORCID, LinkedInID, Institution ID and much more.

### Cuban Context: Interoperability in the VLIR-UOS Network Cuba

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.

The development of the VLIR-UOS 'Network Cuba' is conducted by five universities (Camaguey, Holguin, 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 ELINF project in the network is creating a virtual research and educational environment in Cuba.

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 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.

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 (Subirats et al., 2012, pp. 161-162) in the different tools and the implementation and development of extra tools.

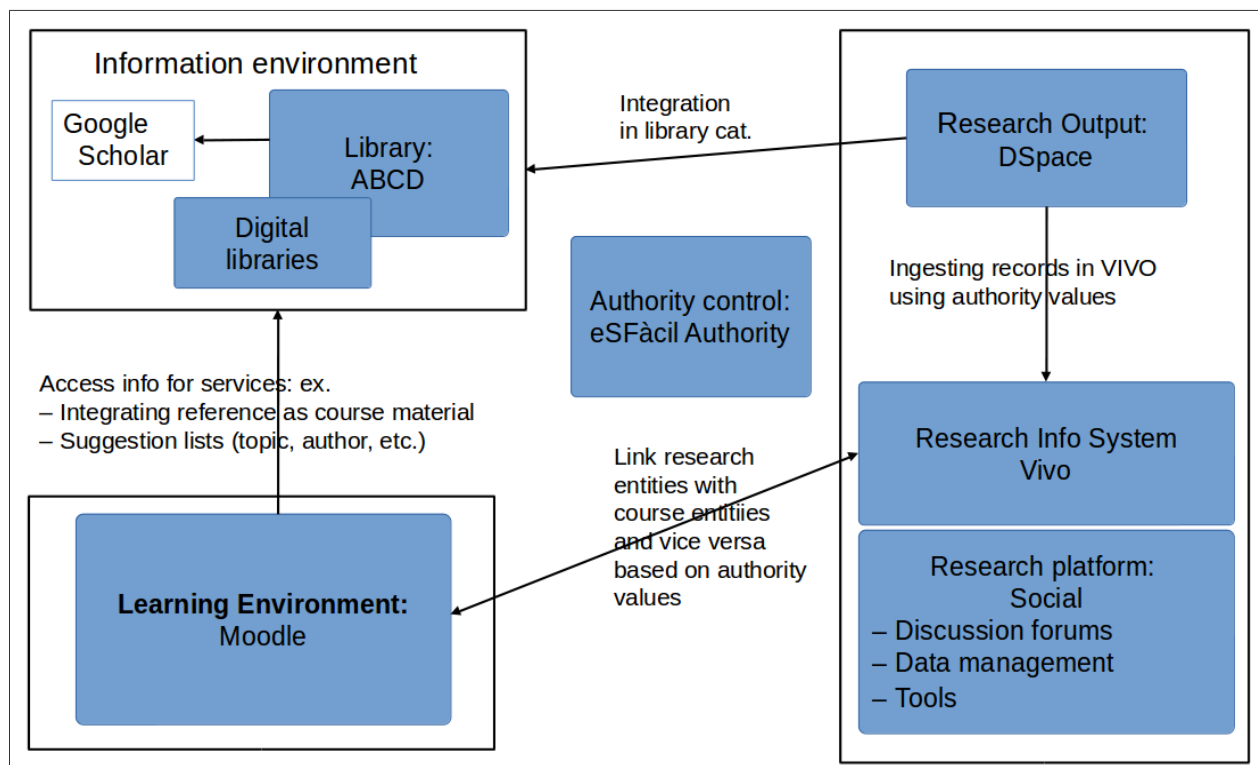


Fig. 11. Interoperability between the different platforms of the ELINF Network

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. But in the context of the

project it is not relevant to impose on the different platforms the use of a single metadata standard.

Therefore the project has to work out a minimal consensus of what can be standardized over the different metadata formats. 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. The issue of unique ID for persons will be discussed further on. The author focuses first on the issue of descriptors and keywords.

To describe content thesauri and vocabularies are the most standardized. 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. 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 usable it is necessary to 'normalize' the access to the different vocabularies through a standard interface: EsFàcil Authority.

EsFàcil Authority 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, CCS, Agrovoc and MESH . Local vocabularies can be added as well as simple authority (e.g. journal lists, ...) and other non-structured lists. That way all authority information can be accessed through one service. The National Library of Finland has developed SkosMos (Suominen, O. et al., 2015), a similar tool focused on SKOS formatted vocabularies.

### **Use of Authority Systems in DSpace: eSFàcil, a submission module with automatic metadata extraction and authority control**

To make the separate tools 'network ready' they have to integrate the authority system functionalities in many ways. The Moodle, ABCD and DSpace software needs further customization. The project is currently developing a specific submission module, eSFàcil, for DSpace that will support the interoperability goals of the network.

eSFàcil is conceived as an independent module that will ingest records and files in repositories and information systems, in the first place in Dspace 7. It will use postgres as a database. The JSONB format will be used to store the metadata of the records. Other data resources are DarkAIV, a tool for metadata extraction and enrichment and eSFàcil Authority for authority control. Both tools will be accessed as web services. Other tools could be attached as well. Angular 2 is chosen as the interface software.

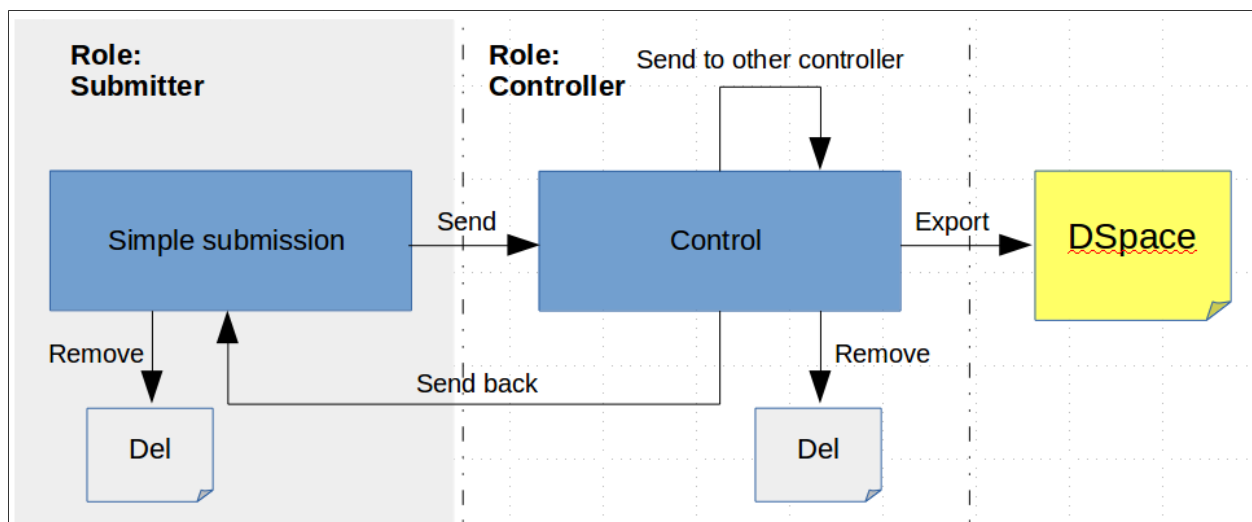


Fig. 12. Workflow of the eSFàcil submission module

The basic idea of eSFàcil is to create a workflow where authors can submit their publication. The module will extract metadata from the document. This is in the first place text information. The quality of the metadata will be controlled by information specialists and enriched using the tools of eSFàcil Authority. Specifically the translation of text to identifiers will prepare the metadata to be machine-readable.

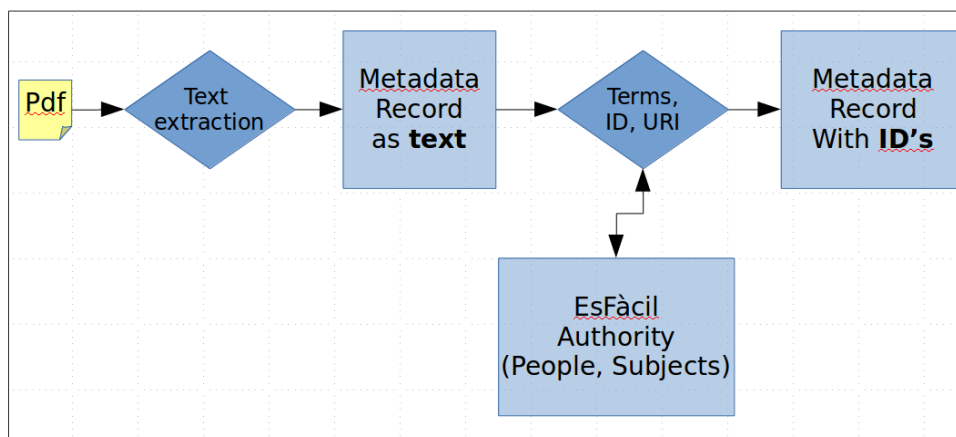


Fig. 13. eSFàcil: From text to information objects

### Development of a unique ID policy - the different options.

Organization, researcher and author databases defines people. Every individual has been identified in different systems starting with the national ID card. Every researcher has been registered by universities, funding agencies, societies, publishers, repositories, social media and other systems.

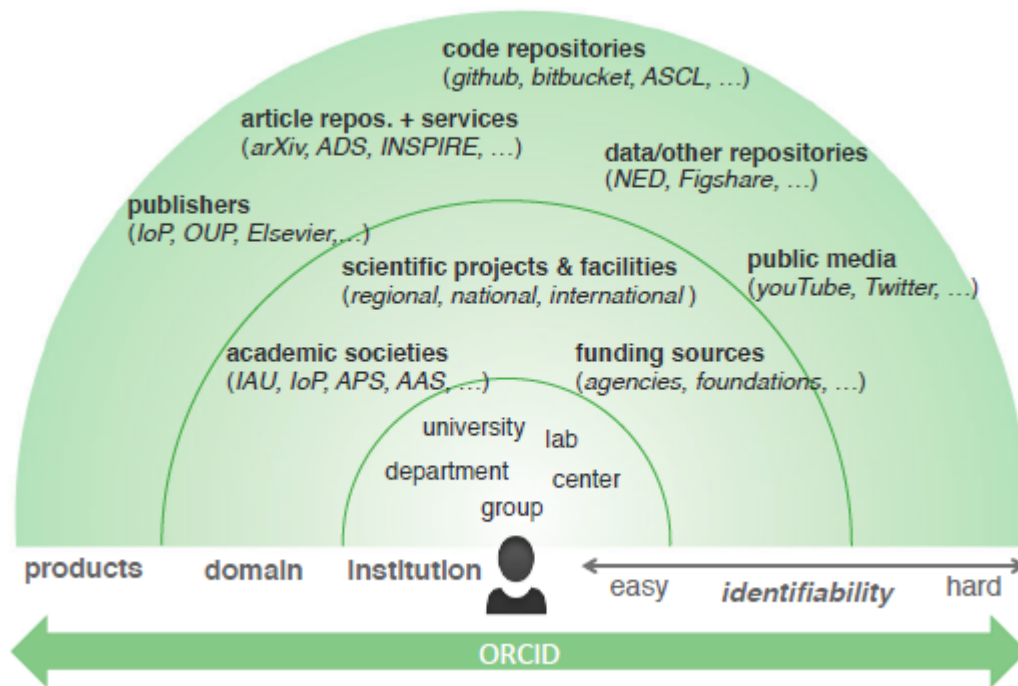


Fig. 14. Persistent, Global Identity for Scientists via ORCID (Evrard, Erdmann, Holmquist, Damon, & Dietrich, 2015)

The main problem is that all these identification systems are not related to each other. How can a global identification system be build up? Via a bottom-up decentralized or a top-down centralized approach?

A bottom-up decentralized approach

Universities and research institutes have or are developing research information systems. Mostly they use internal identifiers for every person. CRIS Systems like VIVO, which is a linked open data based, creates a URI for every object. This way every researcher in VIVO has a URI. If research information systems uses accepted standards while creating their ID, it would enhance the usability of these identifiers.

## Connecting researchers, resources, and clinical activities

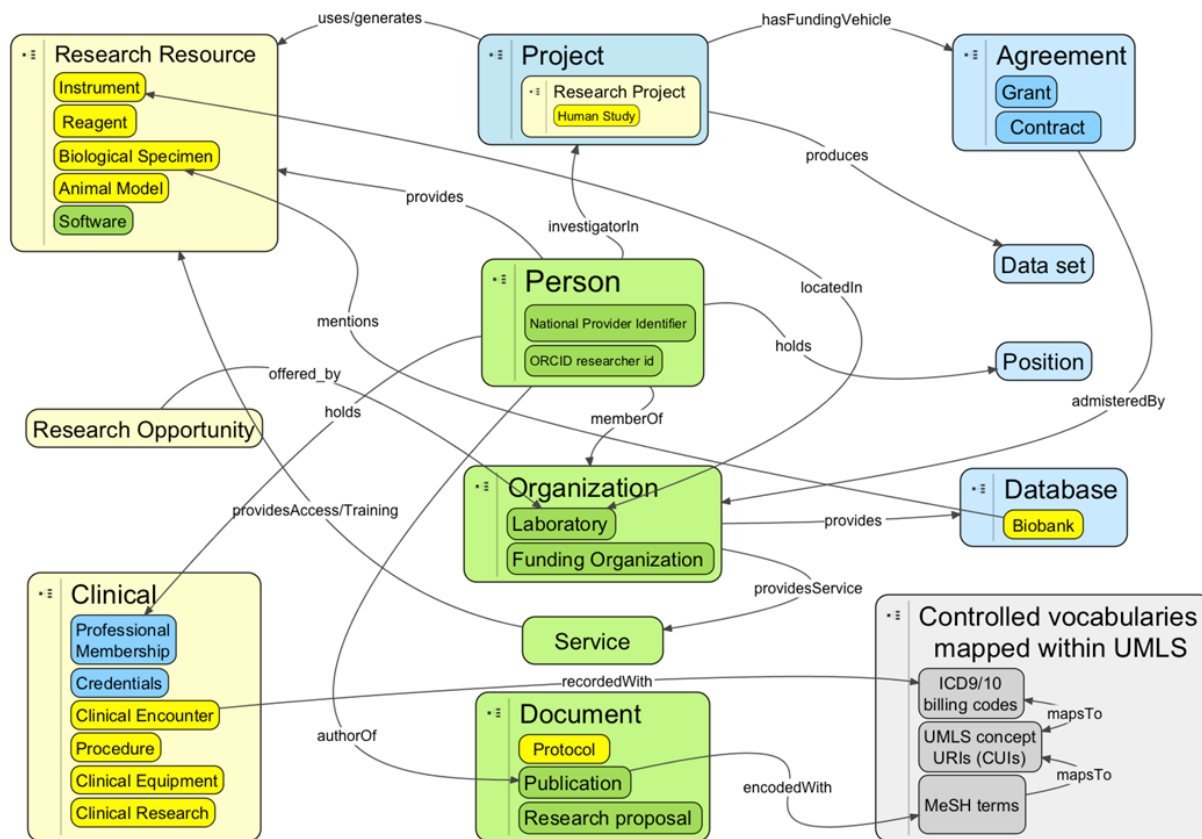


Fig. 15. Creating a Data Interchange Standard for Researchers, Research, and Research Resources: VIVO-ISF (Krafft & Lowe, 2013)

But researchers does not stay at one university. Sometimes they are active in different universities at the same time. This way researchers have more than one URI. The nice part of Linked Data is that it is based on relations and that relation between the different identifications of a person can be made explicit. The question is who will maintain these relations and who will be able to do that?

### A top-down centralized approach

Central identification systems for researchers and scientific authors are now implemented. Web of Science has its ResearcherID, Elsevier its ScopusID, but 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. From the start, the CrossRef and ORCID systems have been envisioned as complementary infrastructures for uniquely identifying researchers and enabling researchers to connect with their publications. (Haak, Fenner, Paglione, Pentz, & Ratner, 2012)

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. Universities and consortia of universities take licenses with ORCID to provide for every researcher a persistent identifier. For example, in 2016 an agreement was signed between the Flemish universities and ORCID.

But ORCID will need a management of duplications. Even when the records of ORCID will automatically populate itself with the papers published, grants received, codes authored, etc., it will need to be controlled and confirmed by the authors.

Most researchers already expose their professional profiles through personal web pages and/or existing commercial services offered by the likes of linkedin.com, researchgate.com, mendeley.com, and academia.edu. How does the orcid.org registry relate to these other profile solutions?

In that way, it is important to recognize that ORCID augments rather than replaces the balkanized set of researcher identities. Individual institutions such as universities and national laboratories will continue to maintain identity and access management services based on local identities. But, as Keith Hazelton points out, ORCID IDs could be used as an identifier to “cross-walk” among appropriately federated campus systems. (Evrard et al., 2015)

While registration by a single author is free, institutional licenses have a cost. Will the Cuban universities be able to fund such a license?

For any identifier system (for researchers but also in general) to be effective, it must be relevant on different levels of the research workflow and management: publication submission, dataset deposition, research grant and contract applications, faculty and staff profiles and evaluation, patent applications, etc. At university level it will be used to create relations between research information, management and evaluation systems, publication and data repositories, educational systems. In the VLIR-UOS network it will bring these tools on a higher level only if the network adopts standards in the development of the identification system that supports reusability of the local data in a broader context.

While ORCID has become in a short period the authority system for researcher ID on international level, alternative authority systems have their reasons to be developed. For example, simply because of the license cost of ORCID for (groups of) institutions.

The place to manage the researcher ID will be the Researcher Information System; VIVO, SIGENU or others. The project has now three main choices: a decentralized VIVO-like solution (still VIVO can include other IDs like ORCID), a Cuban authority system or ORCID as standard international system. It is possible of course to have mixed solutions too. Finally a unique identification system is not a goal by itself. It should only make the interoperability of data stocked in different silos more easily achieved.

## **Conclusions.**

The knowledge in libraries about information management is still relevant in the internet world. Management of bibliographic data, metadata formats, use of authority systems, thesauri and vocabularies will make access to information more easy and better structured. But these tools are evolving. Content is growing and also the type of information is changing. But mainly the internet is moving from a network that links pages to a network that links data: Web of Data. The real value of the Web of Data lies in the relationships between the data. These relationships (referred to as links) put data in context and enrich their meaning and expressiveness.

Unique identification of data is essential to make it machine-readable. Systems for identifying people, organizations and content are essential in the scientific internet world. DOIs, handles, and ORCIDs are used to identify publications and persons. Thesauri like AGROVOC are made linked data compatible.

The VLIR-UOS Network Cuba is developing an information network for research and education developing integrating different platforms for libraries, education and research. To integrate the different platforms standards have to be developed. Following the ideas of the web of data, the network will identify data elements in a standard way over the different platforms: people, organizations and content will get standardized unique identifiers. To support the standards a central authority system is developed: eSFàcil Authority. The use of this approach is now integrated in the different platforms used: Moodle, ABCD, DSpace, VIVO. For DSpace a specific submission module, eSFàcil, is in development that uses metadata autoextraction tools and the authority system of eSFàcil Authority to move the standard repository metadata from text to uniquely identified objects.

Finally, researchers and scientific authors have to be identified uniquely in a network. The network has different options, but finally has to choose between a decentralized or a centralized approach, having the option to use tools like VIVO or other local research system or the option of a centralized directory like ORCID or a specific Cuban identifier register;

## **Bibliography**

BASE - Bielefeld Academic Search Engine | Statistics. (n.d.). Retrieved June 2, 2017, from [https://www.base-search.net/about/en/about\\_statistics.php?menu=2](https://www.base-search.net/about/en/about_statistics.php?menu=2)

Bibliographic Ontology Specification | The Bibliographic Ontology. (n.d.). Retrieved June 3, 2017, from <http://www.biblontology.com/>

DCMI Metadata Basics. (n.d.). Retrieved June 2, 2017, from <http://dublincore.org/metadata-basics/>

Dewey Services. (n.d.). Retrieved June 2, 2017, from <https://www.oclc.org/en/dewey.html>

Evrard, A. E., Erdmann, C., Holmquist, J., Damon, J., & Dietrich, D. (2015). Persistent, Global Identity for Scientists via ORCID. *arXiv:1502.06274 [Astro-Ph, Physics:physics]*. Retrieved from <http://arxiv.org/abs/1502.06274>

FAO of the United Nations. (n.d.). AGROVOC - Multilingual Agricultural Thesaurus. Retrieved June 2, 2017, from <http://www4.fao.org/faobib/kwocinana.html>

FAO of the United Nations. (n.d.). AGROVOC Multilingual agricultural thesaurus | Agricultural Information Management Standards (AIMS). Retrieved June 3, 2017, from <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>

FAO of the United Nations. (n.d.). ASFA THESAURUS. Retrieved June 2, 2017, from <http://www4.fao.org/asfa/asfa.htm>

figshare - About. (n.d.). Retrieved June 2, 2017, from <https://figshare.com/about>

Guenther, R. S. (2003). MODS: The Metadata Object Description Schema. *Portal: Libraries and the Academy*, 3(1), 137-150. <https://doi.org/10.1353/pla.2003.0006>

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>

Hobbes' Internet Timeline - the definitive ARPAnet & Internet history. (n.d.). Retrieved June 2, 2017, from <https://www.zakon.org/robert/internet/timeline/#Growth>

Home - schema.org. (n.d.). Retrieved June 3, 2017, from <http://schema.org/>

Krafft, D. B., & Lowe, B. (2013, October 12). Creating a Data Interchange Standard for Researchers, Research, and Research Resources: VIVO-ISF Dean B. Krafft

Brian Lowe Coalition for Networked Information. - ppt download. Retrieved June 3, 2017, from <http://slideplayer.com/slide/4632327/>

Library of Congress. (n.d.). Library of Congress Classification Outline. Retrieved June 2, 2017, from <https://www.loc.gov/catdir/cpsol/lcco/>

Library of Congress. (n.d.). Library of Congress Subject Headings [webpage]. Retrieved June 2, 2017, from <http://id.loc.gov/authorities/subjects.html>

Library of Congress, N. D. and M. S. O. (n.d.). MARC STANDARDS [webpage]. Retrieved June 2, 2017, from <http://www.loc.gov/marc/>

NIH. (n.d.). Medical Subject Headings - Home Page [Product, Program, and Project Descriptions]. Retrieved June 2, 2017, from <https://www.nlm.nih.gov/mesh/>

Open Access Academy. (n.d.). What Is ORCID? Retrieved June 3, 2017, from [http://www.oaacademy.org/blog\\_what-is-orcid.html](http://www.oaacademy.org/blog_what-is-orcid.html)

RIOXX: RIOXX. (n.d.). Retrieved June 2, 2017, from <http://www.riox.net/>

Subirats, I., Malapela, T., Dister, S., Zeng, M., Goovaerts, M., Pesce, V., Keizer, J. (2012). Reorienting Open Repositories to the Challenges of the Semantic Web: Experiences from FAO's Contribution to the Resource Processing and Discovery Cycle in Repositories in the Agricultural Domain. In *Metadata and Semantics Research* (pp. 158-167). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-35233-1\\_17](https://doi.org/10.1007/978-3-642-35233-1_17)

Suominen, O., Ylikotila, H., Pessala, S., Lappalainen, M., Frosterus, M., Tuominen, J., Retterath, A. (2015). *Publishing SKOS vocabularies with Skosmos*. Retrieved from <http://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf>

The DOI System. (n.d.). Retrieved June 2, 2017, from <http://www.doi.org>

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

UDC Consortium - About UDC. (n.d.). Retrieved June 2, 2017, from <http://www.udcc.org/index.php/site/page?view=about>

Understanding Linked Data by example. (n.d.). Retrieved June 3, 2017, from <https://joinup.ec.europa.eu/community/semic/news/understanding-linked-data-example>

Pesce, V.. (2014). *VIVO as a CRIS platform*. Retrieved from <https://www.youtube.com/watch?v=jDPwbuqLfvE>