Towards Linked Agricultural MetaData: Directions of the agINFRA Project

Valeria Pesce[1], Guntram Geser[2], Vassilis Protonotarios[3], Caterina Caracciolo[1], Johannes Keizer[1]

[1] Food and Agriculture Organization of the United Nations, Rome, Italy
{Valeria.Pesce, Caterina.Caracciolo, Johannes.Keizer}@fao.org
[2] Salzburg Research, Salzburg, Austria
guntram.geser@salzburgresearch.at
[3] University of Alcala, Alcala de Henares, Spain
vprot@agroknow.gr

Abstract. The agINFRA project focuses on the production of interoperable data in agriculture, starting from the vocabularies and KOS used to classify and annotate them. In this paper we report on our first steps in the direction of contributing to a LOD of agricultural data. In particular we look at germplasm data and soil data, which are still widely missing from the LOD landscape, seemingly because information managers in this field are still not very familiar with LOD practices. This is why this paper also recaps the basics of LOD publishing, which will be applied in the agINFRA project.

Keywords: Agriculture, germplasm, soil, Knowledge Organization Systems, metadata sets, vocabularies, RDF, Linked Data, classification

1 Introduction

A discourse on agriculture requires notions and data coming from various perspectives. Therefore, in view of interlinking and integrating diverse data using coherent semantics, LOD is a very interesting technology to explore for those interested in working with agricultural data.

The agINFRA project (www.aginfra.eu) is a project co-funded by the European Commission (FP7 programme), aiming to provide tools and methodologies for creating large networks of agricultural data sources using grid- and cloud-based technology.

Any data set comes with some sort of metadata associated, in order to enable its description and retrieval. The first ingredient for having such metadata is a “vocabulary” to describe each individual piece of information that we may want to add to the data. For example, if we want to store information about the subject covered by the data, we will use a metadata element that may be rendered by using the property dct:subject of the Dublin Core vocabulary. The second key ingredient is some sort of “controlled vocabularies”, or “authority data”, from which values for those elements may be taken. For example, AGROVOC terms may be used as values for dct:subject. Usually, the former ingredients are referred to as “metadata sets,” while the latter are called Knowledge Organization Systems (KOSs). However, often both are referred to as “vocabularies” (as also noted in [1]). In the W3C Library Linked Data Incubator Group, a proposed terminology is: “metadata element sets” and “value vocabularies” [2].

Both types of vocabularies are crucial to “express” metadata. This is why vocabularies have been among the first things to be published as Linked Data: they are needed to understand the data.

For KOS, we adopt the definition given in [3], according to which Knowledge Organization Systems include “*classification and categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontologies.*”

In this paper, and in the agINFRA project, we do not consider name authority lists. However, although publishing name authority lists as Linked Data is a more complex business than publishing KOSs (mainly due to heavy disambiguation issues, as well as varying standards from country to country, transliterations etc.), one can assume that much of what is said in this paper about publishing KOS as Linked Data also applies to name authority lists.

2 Metadata Sets and KOSs Relevant to Agriculture

One of the goals of the agINFRA project is to enhance the interoperability of datasets related to agriculture, so as to allow for smooth harvesting and querying by services developed within the project. Those services will ultimately be put in the open domain for anybody to reuse. Table 1 below summarizes the data sets considered so far in agINFRA. The table focuses on the metadata schemas and the KOSs used by each dataset.

The analysis performed revealed that for educational resources the metadata sets adopted are rather homogeneous, and also the use of KOSs is rather consistent. The same applies to bibliographic resources. We found that most providers use either AGROVOC or thesauri like the Chinese Agricultural Thesaurus (CAT) or the ASFA thesaurus – both already mapped to AGROVOC. Also newly created ontologies like the Agroecology ontology are heavily based on AGROVOC. Besides, international standards exist and Linked Open Data (LOD) enabling recommendations for bibliographic resources have already been developed [5].

For germplasm data and soil data the formalization of metadata standards and the availability of common KOSs is not so advanced. This in a way makes them the most interesting use case for the publication of relevant vocabularies as LOD. In fact, since the collection and exchange of this type of data entails the adoption of scientific classifications and highly normative prescriptions, reference standards are not lacking.

However, they appear to have been rarely formalized as metadata sets or KOS and to have never been published as LOD.

Table 1. Datasets considered in agINFRA, with the metadata sets and KOS used in them.

Type of resource	Collection name	Metadata set used	KOS used
Educational	**Capacity Development Portal**	FAO AgLR AP[1]	AGROVOC[2]
	Organic.Edunet	Organic.Edunet IEEE LOM AP[3]	OA-AE ontology[4]
	LaFLOR	IEEE LOM AP [4]	ARIADNE subject classification system[5]
	OpenLearn	Own metadata schema	
Bibliographic	**AGRIS**	AGRIS AP[6] LODE-BD [5]	AGROVOC
	VOA3R	VOA3R AP[7]	AGROVOC
	CASDD (China)	Dublin Core[8] based metadata schema	Chinese Agricultural Thesaurus, Chinese Library Classification[9]
	FAO Open Archive	MODS[10]	AGROVOC
	Biodiversity Heritage Library (BHL)	Dublin Core, MODS, Darwin Core[11]	uBio Classification-Bank (species taxonomies)[12]

[1] Metadata application profile for FAO's agricultural learning resources, ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/learnap.pdf

[2] AGROVOC, http://aims.fao.org/standards/agrovoc/

[3] Organic.Edunet Metadata Application Profile, http://wiki.organic-edunet.eu/index.php/Organic.Edunet_Metadata_Application_Profile

[4] Organic Agriculture (OA) and Agroecology (AE) ontology, http://wiki.organic-edunet.eu/index.php/Organic.Edunet_Ontology

[5] ARIADNE, http://www.ariadne-eu.org

[6] AGRIS Application profile, http://www.fao.org/docrep/008/ae909e/ae909e00.htm

[7] VOA3R Metadata Application Profile, http://ieru.org/voa3r//wiki/index.php?title=VOA3R_Metadata_Application_Profile

[8] Dublin Core (DC), http://dublincore.org

[9] Chinese Library Classification, http://en.wikipedia.org/wiki/Chinese_Library_Classification

[10] Metadata Object Description Schema (MODS), http://www.loc.gov/standards/mods/

[11] Darwin Core (DwC), http://rs.tdwg.org/dwc/

[12] uBio ClassificationBank, http://www.ubio.org/browser/classifications.php

	Mendeley	Own metadata schema	Own classification schema
	INDUS (India)	Dublin Core	AGROVOC
	DOI Serbia	Dublin Core	
Germplasm	**CRA Germplasm (Italy)**	Multi-crop Passport Descriptors (MCPD) [6]	
	CGRIS (China)	Own set of germplasm descriptors	
Soil datasets and maps	**Italian Soil Information System (ISIS)**	ISO 19115/19139[13]	USDA Soil Taxonomy [7], World Reference Base for Soil Resources [8]

In the case of germplasm, the set of Multi-crop Passport Descriptors (MCPD V.1 2006, V.2 2012) is widely used for information exchange among crop conservation and research institutions worldwide. It is also used by the national germplasm inventories in Europe to provide their information to the EURISCO catalogue (with six additional descriptors for the specific purposes of EURISCO).[14] This includes the germplasm collections of the Italian Agricultural Research Council (CRA). The Crop Germplasm Research Information System (CGRIS) of the Chinese Academy of Agricultural Sciences (CAAS) uses an own set of passport descriptors which, however, represents the de facto standard in China and will be mapped to the MCPD.

Importantly, the MCPD does not include descriptors for Characterization and Evaluation (C&E) measurements of plant traits/scores which is the most important information for plant researchers and breeders. But initial sets of C&E descriptors for the utilization of 22 crops have been developed by Bioversity International together with CGIAR and other research centers [9]. C&E measurement data determine the value (e.g. resistance to specific pathotypes, grain yield, protein content, etc.) and, hence, selection of relevant germplasm. However, as assessed by the EPGRIS3 project, C&E data is difficult to standardize and integrate in central databases [10]. A major recent achievement therefore is the Darwin Core extension for genebanks (DwC-germplasm) which is represented in RDF/SKOS. The extension has been derived from the MCPD standard and includes basic descriptors for C&E measurements as suggested by EPGRIS3 [11].

[13] ISO 19115/19139: Geographic information – Metadata, and XML schema implementation, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=54904&published=on&includesc=true

[14] EURISCO, http://eurisco.ecpgr.org

With regard to authoritative plant names and taxonomies there is no shortage at all, and on the side of ontologies the Plant Ontology[15] (explicitly referenced in the DwC-germplasm), Trait Ontology[16] and Phenotypic Quality Ontology[17] provide important controlled vocabularies.

Concerning soil measurements, there exist (de facto) standards of data dictionaries of major databases which describe the dataset tables, i.e. provide the definitions of the data elements. Sometimes they are also called "metadata", for example by the U.S. National Soil Information System (NASIS) [12].

The most widely used metadata standard is ISO 19115 for geographic information and services, which is applied to catalog and fully describe datasets, including individual geographic features and feature properties. ISO 19139 provides the XML schema implementation, including the extensions for imagery and gridded data. Users of the Content Standard for Digital Geospatial Metadata (CSDGM) have been recommended by the U.S. Federal Geographic Data Committee (FGDC) transitioning to the ISO standards [13].

The main international KOSs are the Soil Taxonomy and the World Reference Base for Soil Resources. An important recent achievement is the multilingual soil thesaurus (SoilThes) that has been developed in the eContentplus project GS SOIL [14]. SoilThes was created as an extension of the General Multilingual Environmental Thesaurus (GEMET)[18] and contains the concepts of the World Reference Base, the soil vocabulary of ISO 11074[19] and additional soil-specific concepts.

3 Methodology and Tools for Conversion to LOD

Within the project, it was decided to publish the datasets considered as Linked Data, as a way to improve the interoperability of the data sources considered. Therefore the first step was to agree on a common set of RDF classes and properties for each type of resource, and on some reference KOSs to which the other local KOSs can be mapped. In both cases, preference was given to metadata sets and KOS already published as LOD. If those were not available, it was agreed that the project would take care of their publication as subsidiary agINFRA vocabularies.

The methodologies for publishing as Linked Data are slightly different for the two types of vocabularies, but in both cases they comply with the Linked Data rules [15]:

1. "Use URIs as names for things". In our case, "things" are both values in value vocabularies, and classes and properties in description vocabularies. Vocabularies themselves are "things" to be identified by URIs.

[15] Plant Ontology, http://www.plantontology.org

[16] Trait Ontology, http://www.gramene.org/plant_ontology/

[17] Phenotypic Quality Ontology, http://obofoundry.org/wiki/index.php/PATO:Main_Page

[18] GEMET, http://www.eionet.europa.eu/gemet/

[19] ISO 11074:2005 Soil quality - Vocabulary, http://www.iso.org/iso/catalogue_detail.htm?csnumber=38529

2. “Use HTTP URIs so that people can look up those names”. The URIs for concepts / values, classes and properties, as well as vocabularies, have to be resolved as HTTP URLs.
3. “When someone looks up a URI, provide useful information, using the standards”. This means that URI should resolve for both humans and machines. All URLs should then return an HTML page with useful information when requested by browsers, and RDF when requested by RDF software. Besides, vocabularies should be available for querying through a SPARQL endpoint.
4. “Include links to other URIs, so that more things can be discovered”. This is actually the essence of publishing Linked Data. In the case of metadata sets and KOSs, the URIs of concepts, classes and properties should whenever possible be linked to URIs in other vocabularies, for instance as a close match of another concept or sub-class of another class

3.1 Publishing Metadata Sets as LOD

For publishing metadata sets as Linked Data, the following steps have been planned in the agINFRA project for each data type covered by the project.

1. Identify common and standard RDF vocabularies

This step basically is about agreeing on a common set of RDF classes and properties for the types of resource to be described. Classes and properties from existing published vocabularies should be used whenever possible. When this is not possible, new classes and properties can be published in a new subsidiary vocabulary. Some classes and properties that are very specific to the agINFRA project, or for which no RDF vocabulary exists, will be published in an agINFRA RDF vocabulary.

This step is not needed if one only wants to publish a metadata set as RDF. But it is fundamental in order to achieve interoperability of vocabularies in a network (which is the case of the agINFRA vocabularies), and to provide a reference model to which similar metadata sets can link.

2. Express metadata sets in RDF

As indicated by the W3C Library Linked Data Incubator Group, metadata element sets are expressed as RDF Schemas or OWL Web Ontology Language ontologies.

Transforming an existing metadata set in RDF or OWL often requires some rethinking of the metadata architecture, especially in the case of transformation of a complex XML schema with many nested elements and many attributes for each element.

It is highly recommended that already at this stage, whenever possible, the elements are expressed using a class or property from a vocabulary already published as Linked Data (preferably, in the case of agINFRA, one in the common set of classes and properties agreed in step 1). Sometimes the adoption of the existing class or property is not possible, for example, because the name of the property has to remain the same as in the XML schema for legacy reasons. Then the new RDF class or property should be declared as a sub-class or sub-property of the existing class or property (this is part of both the vocabulary definition and its interlinking: see point 4). Only if this is not possible, for instance due to a difference in constraints compared to the existing RDF class or property, should a new class or property be created. These recommendations follow the guidelines provided by Tom Heath and Christian Bizer [16].

3. Publish namespaces for vocabularies
The resulting RDF definition needs to be published under a namespace, which will be the unique identifier for the vocabulary and will be appended to the class and property names to uniquely identify them. Following the Linked Data approach, this namespace should be an HTTP URI.

4. Interlink vocabularies
Whenever possible, classes and properties in the newly published vocabularies should be linked to classes or properties in other published vocabularies (preferably, in the case of agINFRA, the classes and properties agreed in step 1). This is done using standard RDF properties such as owl:equivalentClass and owl:equivalentProperty or rdfs:subClassOf and rdfs:subPropertyOf.

This process should be repeated over time, as new vocabularies are published.

Tools
There are tools that provide a graphic interface to create classes and properties, define them, associate a data type to them and in some cases even apply some constraints. Some of these tools allow users to export the resulting RDF vocabulary to be then published at a specific URL, which will constitute its namespace. Some other tools, deployed on line, automatically make the vocabulary available under a namespace.

A tool that has been tested in the agINFRA project for this purpose is the Neologism Drupal distribution[20], which is open source, easy to use, deployable online and dedicated to the building and online publication of simple RDF vocabularies. Neologism is listed in [16] together with TopBraid Composer[21] (a powerful commercial modeling environment), Protégé[22] (open-source ontology editor) and the NeOn Toolkit[23] (open-source ontology engineering environment for networked ontologies).

3.2 Publishing KOSs as LOD

Compared to the publishing of metadata sets, the process for publishing KOSs as Linked Data is more similar to the normal recommended process for publishing any data as Linked Data, as concepts are data. For this, the following steps have been planned in the agINFRA project.

[20] Neologism, http://neologism.deri.ie
[21] TopBraid Composer, http://www.topquadrant.com/products/TB_Composer.html
[22] Protégé, http://protege.stanford.edu
[23] NeOn Toolkit, http://neon-toolkit.org

1. Identify additional non formalized KOSs used in the data sources

Besides formalized KOSs, many of the data sources in agINFRA use internal controlled lists of values that are rarely based on commonly shared standards (or for which recognized standards do not exist): yet these controlled lists are essential for the interoperability of these sources. Examples are the controlled values used for the metadata element “document type” in bibliographic resources, the values used for “collecting/acquisition source” or “biological status of accession” in germplasm data and the values used for “learning resource type” in learning resources. For interoperability purposes, publishing these lists as Linked Data (and linking values between different homogeneous lists if there is more than one) can be very effective in the project.

2. Express a KOS in RDF, and establish the URI pattern

In RDF, KOSs are normally expressed using the SKOS vocabulary.

Transforming an existing KOS in RDF using the SKOS classes and properties is more straightforward than transforming metadata sets into RDF vocabularies. Because SKOS was designed to accommodate the structure of most KOSs, from flat mono-lingual classifications to multilingual thesauri with relations (except for complex ontologies with reasoning, for which OWL is usually necessary). Furthermore, SKOS supports languages and relations natively. Additional properties, like submitter or last update date, can be added seamlessly through the use of other RDF vocabularies (like Dublin Core).

Therefore, expressing a KOS in RDF probably will not require much re-thinking of the KOS structure.[24] While building the RDF representation of the KOS (for which a tool may be needed, especially for subsequent maintenance, see below), a decision should also be made on the URI pattern, both in relation to the hash vs. slash question[25] and in relation to the use of numeric identifiers or strings to identify entities within the namespace.

3. Publish a namespace for the KOS

The resulting SKOS needs to be published under a namespace, which will be the unique identifier for the KOS and will be attached to the concept IDs to uniquely identify them. Following the Linked Data approach, this namespace should be an HTTP URI. This can be done either by uploading a SKOS file (not recommended for huge KOSs) or by using a tool that allows the KOS’s owners to manage or import a SKOS and serve it as Linked Data on line.

4. Interlink KOSs

In order to achieve interoperability between KOS with similar or overlapping coverage, the terms / values in the KOS need to be linked to other related terms / values in

[24] The publication of AGROVOC as SKOS may be mentioned as an exception: It was previously built on an OWL model and many advanced OWL properties had to be converted to simpler SKOS properties.

[25] W3C HashVsSlash, http://www.w3.org/wiki/HashVsSlash

other KOSs. This is done using standard SKOS properties skos:exactMatch, skos:closeMatch, skos:broadMatch, skos:narrowMatch and skos:relatedMatch.

This process should be repeated over time, as new KOSs are published.

In order to improve the discoverability of the KOSs, they should then be registered in the CKAN Data Hub.[26]

Tools

Two tools that have been tested in the agINFRA project are the FAO VocBench[27] and the MediaWiki MoKi tool[28], both web-based.

VocBench is a multilingual editing and workflow tool developed by FAO for the management of various types of KOS, like thesauri, authority lists and glossaries using the RDF/SKOS model. VocBench provides tools and functionalities that facilitate both collaborative editing and multilingual terminology. It also includes administration and group management features that permit flexible roles for maintenance, validation and quality assurance. It allows users to build and maintain KOSs as well as export them in standard formats (SKOS) that are compatible with tools that publish Linked Data.

MoKi is based on MediaWiki and uses wiki pages to describe semantic terms: these wiki pages are then organized and the concepts and sub-concepts are visualized in a tree-based schema; concepts can be added, revised, translated and deleted, while the structure of the ontology can be modified by changing the hierarchy of the concepts.

Other existing tools include the open source SKOSJS[29], Protégé[30] and TemaTres Controlled Vocabulary server[31] along with commercial tools like PoolParty[32] or TopBraid Enterprise Vocabulary Net[33].

The current plan in agINFRA is hosting a VocBench v2.0 instance on the project Cloud over Grid system, where KOSs can be imported and managed directly by the owners.

4 Discussion and Future Work

Initial work done using the methodologies described in this paper has already shown interesting results. On the one hand, within the apparently linear process of expressing metadata sets and KOSs as RDF and interlinking them, many issues are arising, not all solvable, mostly due to the inherent differences between the conceptual models behind the vocabularies. Differences in granularity, in both types of vocabularies, are easier to reconcile, as appropriate relations exist (skos:broadMatch,

[26] CKAN Data Hub, http://datahub.io

[27] FAO VocBench, http://aims.fao.org/tools/vocbench

[28] MediaWiki MoKi , https://moki.fbk.eu

[29] SKOSJS, https://github.com/tkurz/skosjs

[30] Protégé, http://protege.stanford.edu

[31] TemaTres, http://www.vocabularyserver.com

[32] PoolParty, http://poolparty.punkt.at

[33] TopBraid, http://www.topquadrant.com/solutions/ent_vocab_net.html

skos:narrowMatch, rdfs:subProperty etc.), while differences in the ontological model are in some cases irreconcilable (this is why making vocabularies with a stronger "ontological commitment" interoperable is more difficult). On the other hand, the process of analyzing the existing metadata sets and KOSs and planning the steps for their publishing as LOD has made all data providers aware of the benefits of such an approach and willing to make an effort in this direction.

In particular, the study of current germplasm and soil data management practices revealed that experts in these two areas are actually looking forward to the adoption of LOD technologies to improve the interoperability of their data. The publication of germplasm and soil-related vocabularies will be a big step forward and will represent one of the really novel contributions that agINFRA makes to the agricultural data management community.

Another interesting aspect of the LOD work in agINFRA is the planned publication and mapping of smaller internal controlled lists of values used by different datasets in the project for similar concepts. One example is the varied use of values in different datasets for indicating the type of bibliographic resource: no comprehensive standard reference list has been published so far for such a common need. In agINFRA, all lists used in individual datasets will be published and one common combined list is under study.

Considering the option of mapping this list with a more specialized list like the types of learning resource, and the possibility of using these lists as allowed "schemes" for metadata properties, the potential of following such a path becomes evident. Another obvious case for publishing the internal lists of controlled values and mapping them is controlled values for specific germplasm metadata, for which data providers are lamenting the lack of authority control, like "collecting/acquisition source" or "biological status of accession", for instance.

Future work will consist first in the actual publication of the analyzed vocabularies as LOD and then in the exploitation of this framework in applications and information systems.

A few simple examples of what we foresee as results of this work in the project are:

1) Learning resources in a non-standardized repository, described with Dublin Core properties and some additional proprietary custom properties, will be easily harvested by learning platforms that manage resources described with IEEE-LOM properties by just looking up the vocabularies (i.e. without the need to write custom programming code).
2) Data management tools will be able to look up authority data published as LOD and allow for annotation with controlled values.
3) Germplasm data can be displayed alongside bibliographic resources when relevant to the search criteria.
4) Soil data can be displayed alongside other types of information resources for geographic searches or for AGROVOC terms-based searches, if the AGROVOC terms have an equivalent e.g. in the USDA Soil Taxonomy.

We foresee that publishing both types of vocabularies as Linked Data will amplify their power by making them machine-readable, easily re-usable and linked or potentially linkable to other vocabularies. This creates crosswalks that will subsequently increase the interoperability of data sources where the vocabularies are used, the discoverability of resources described and indexed with them, and both the precision and recall of searches in applications that leverage any of the vocabularies.

Acknowledgements: The research leading to these results has received funding from the European Union Seventh Framework Programme (*FP7/2007-2013*) under *grant agreement* n° *283770*. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The authors would like to thank all the consortium partners for their contribution in the analysis of the metadata used to describe agricultural data sources and collections.

5 References

1. Méndez, E., Greenberg, J.: Linked Data for Open Vocabularies and HIVE's Global Framework. El profesional de la información, vol. 21, pp. 236-244 (2012), http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_eng.pdf
2. Isaac, A. et al.: Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets. W3C Incubator Group Report (2011), http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/
3. Hodge, G.: Systems of Knowledge Organization for Digital libraries. Council on Library and Information Resources, Washington, DC, USA (2000), http://www.clir.org/pubs/reports/pub91/contents.html
4. IEEE Learning Object Metadata: Draft Standard for Learning Object Metadata (2002), http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
5. FAO: Meaningful Bibliographic Metadata - Recommendations of a set of metadata properties and encoding vocabularies (2012), http://aims.fao.org/metadata/m2b
6. FAO/Bioversity: Multi-Crop Passport Descriptors V.2 (MCPD V.2) (2012), http://eurisco.ecpgr.org/fileadmin/www.eurisco.org/documents/MCPD_V2_2012_Final_PDFversion.pdf
7. U.S. Department of Agriculture, Natural Resources Conservation Service: Soil Taxonomy - A Basic System of Soil Classification for Making and Interpreting Soil Surveys (1999), http://soils.usda.gov/technical/classification/taxonomy
8. IUSS Working Group: World reference base for soil resources, 2nd edition, 2006. World Soil Resources Reports No. 103. FAO, Rome (2006), http://www.fao.org/ag/agl/agll/wrb/doc/wrb2006final.pdf
9. Bioversity International: Key access and utilization descriptors for crop genetic resources (2011), http://www.bioversityinternational.org/index.php?id=3737

10. van Hintum, T.: Inclusion of C&E data in EURISCO - analysis and options. EPGRIS-3 Proposal, Wageningen (2009), http://edepot.wur.nl/186143
11. Endresen, D., Knüpffer, H.: The Darwin Core Extension for Genebanks opens up new opportunities for sharing germplsam data sets. Biodiversity Informatics, 8, 2012, pp. 12-29, https://journals.ku.edu/index.php/jbi/article/viewFile/4095/4064
12. U.S. Department of Agriculture, NRCS: NASIS-Related Metadata, http://soils.usda.gov/technical/nasis/documents/metadata
13. Federal Geographic Standard Committee (FGDC): Geospatial Metadata Standards (2012), http://www.fgdc.gov/metadata/geospatial-metadata-standards
14. GS SOIL: Establishment of a multilingual soil-specific thesaurus. Deliverable 3.5. Prepared by Herbert Schentz et al., 31 May 2012 (2012), http://www.gssoil-portal.eu/Best_Practice/GS_SOIL_D3.5_Soil_Specific_Thesaurus_final.pdf
15. Berners Lee, T.: Linked Data. In: Design Issues. World Wide Web Consortium (2006), http://www.w3.org/DesignIssues/LinkedData.html
16. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, pp. 1-136 (2011), http://linkeddatabook.com/editions/1.0/