Migrating Bibliographic Datasets to the Semantic Web: the AGRIS case

Editor(s): Jens Lehmann, University of Leipzig, Germany; Oscar Corcho, Universidad Politécnica de Madrid, Spain
Solicited review(s): Christophe Guéret, VU Amsterdam, The Netherlands; Daniel Vila-Suero, Universidad Politécnica de Madrid, Spain; Hugh Glaser, University of Southampton, UK

Stefano Anibaldi[a], Yves Jaques[a], Fabrizio Celli[a], Armando Stellato[ab] and Johannes Keizer[a*]
[a] *Food and Agriculture Organization of the United Nations (FAO of the UN), v.le Terme di Caracalla 1, 00154, Rome, Italy*
[b] *Department of Enterprise Engineering, University of Tor Vergata, Via del Politecnico 1, 00133, Italy*

Abstract. AGRIS is among the most comprehensive online collections of agricultural and related sciences information. It is a growing global catalog of 5 million high-quality structured bibliographic records indexed from a worldwide group of providers. AGRIS relies heavily on the AGROVOC thesaurus for its indexing. Following the conversion of that thesaurus into a SKOS concept-scheme and its publication as Linked Open Data (LOD), the entire set of AGRIS records was also triplified and released as LOD. As part of this exercise, OpenAGRIS, a semantic mashup application, was developed to dynamically combine AGRIS data with external data sources, using a mixture of SPARQL queries and web services. The re-engineering of AGRIS for the Semantic Web raised numerous issues regarding the relative lack of administrative metadata required to compellingly address the proof and trust layers of the Semantic Web stack, both within the AGRIS repository and in the external data pulled into OpenAGRIS. The AGRIS team began a process of disambiguation and enrichment to continue moving toward an entity-based view of its resources, beginning with the tens of thousands of journals attached to its records. The evolution of the system, the issues raised during the triplification process and the steps necessary for publishing the result as LOD content are hereby discussed and evaluated.

Keywords: Linked Datasets, Agriculture, Journal Archives, Digital Libraries

1. Introduction

The International Information System for the Agricultural Sciences and Technology (AGRIS) is an initiative set up by the Food and Agriculture Organization of the United Nations (FAO of the UN) in 1974, to make information on agricultural research globally available. One of the main objectives of the AGRIS initiative is to improve access and exchange of information serving "the information needs of developed and developing countries on a partnership basis" [1]. Since 1975, AGRIS has been collecting and disseminating bibliographic information on scientific and socio-economic publications, scholarly papers and grey[1] literature, issued on a wide variety of food and agricultural subjects from over 150 data providers located in more than 100 different countries.

With the growth of open access Institutional Repositories (IRs), AGRIS has considerably improved its methods for harvesting and indexing metadata from content providers. Initiatives such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) have greatly facilitated interoperability with our data providers.

Today AGRIS disseminates a collection of more than 5 million bibliographic references modeled using a qualified Dublin Core metadata format that facilitate sharing of information across information systems.

[1] Unpublished scientific and technical reports

*Corresponding author

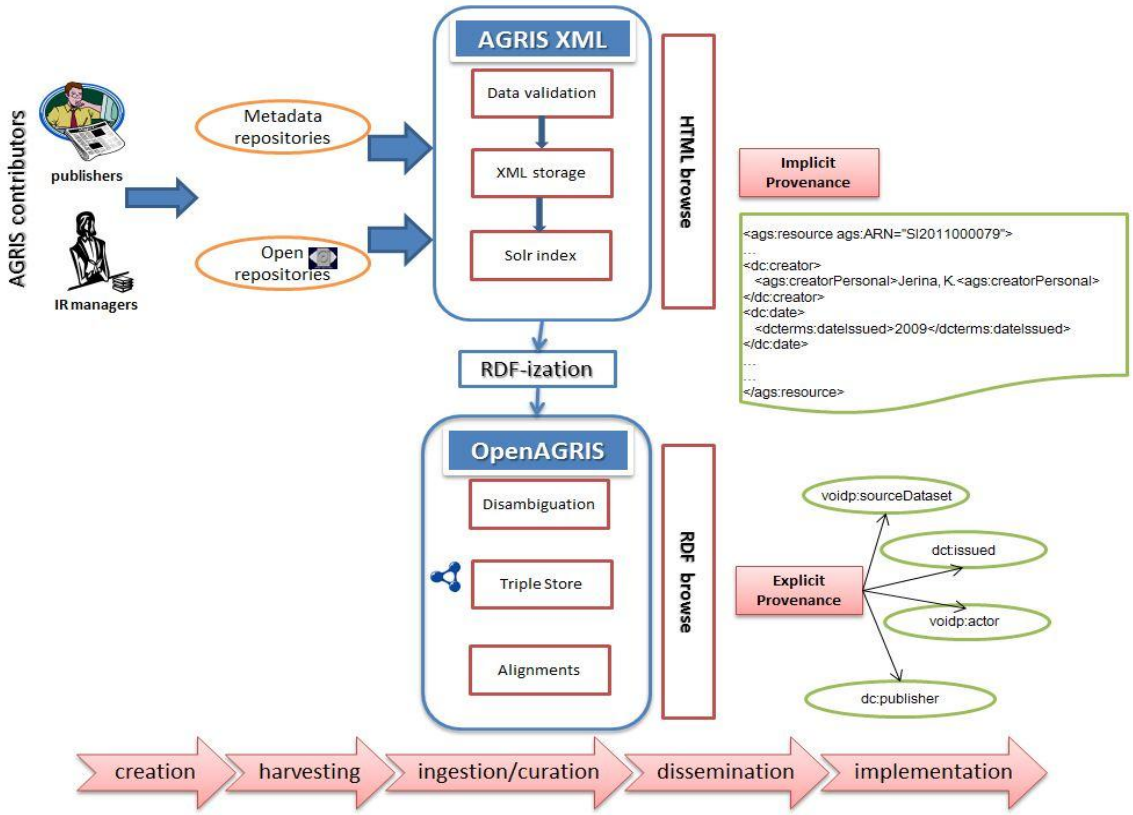

Fig. 1. Derivation history of an AGRIS artifact

Dissemination is handled by the AGRIS portal[2], one of the most trafficked FAO web sites, with over 200,000 unique visitors per month, augmented by the OpenAGRIS [2] semantic mashup that combines and merges information from different online data sources. The entire XML and RDF repository is freely downloadable[3] and can be reused and published, after a specific request to the AGRIS Secretariat, under the terms and conditions as stated in http://agris.fao.org/terms.

The AGRIS data is heavily indexed with the AGROVOC thesaurus. In order to take better advantage of its move to LOD, the entire set of AGRIS records was remodeled using Resource Description Framework (RDF) and deployed in an Allegrograph4 triple store. Well-known standard RDF vocabularies such as BIBO [3] FOAF [4] and Dublin Core Metadata terms [5] were reused to facilitate interlinking with other existing datasets. However, the triplification of the AGRIS content brought to light deficiencies in data production, in particular the handling of proof and trust in a world of machine-readable linked data. The paper covers a number of initial issues that were resolved and finishes with an overview of proposals aimed at partially or wholly remedying the remaining proof and trust deficiencies in the AGRIS dataset.

2. Why RDF

Bibliographic resources are static and often do not contain the information needed to answer a user's query, in particular when the record does not link to the full text of the described publication. A recent AGRIS site analysis shows that when end users reach an AGRIS result page (a reference) where no full text information is available, they typically exit the site and presumably search again – using other search engines or databases – for other online resources. The "...digital environment has increased the range of user needs and expectations beyond the scope of the collections of most consortia. End-user services are now required to operate at Web-scale and incorporate metadata from sources well beyond the traditional bibliographic record." [6]. As of August 2013, the

[2] http://agris.fao.org
[3] ftp://ext-ftp.fao.org/GI/Reserved/Agris/AgrisData/
[4] http://www.franz.com/agraph/allegrograph/

```
<bibo:Article rdf:about="http://agris.fao.org/aos/records/XS2010X00001">
  <dct:identifier>XS2010X00001</dct:identifier>
  <dct:title xml:lang="pt">Caracteristicas anatômicas ...</dct:title>
  <dct:title xml:lang="en">...</dct:title>
  <dct:creator>
    <foaf:Person><foaf:name>Mesquita, Alessandro Carlos</foaf:name></foaf:Person>
  </dct:creator>
  <dct:publisher>
    <foaf:Organization><foaf:name>Instituto Nacional de ...</foaf:name></foaf:Organization>
  </dct:publisher>
  <dct:issued>2010</dct:issued>
  <dct:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_6200"/>
  <bibo:abstract xml:lang="pt"><![CDATA[As estruturas envolvidas na produção ...]]></bibo:abstract>
  <bibo:abstract xml:lang="en"><![CDATA[The structures involved in latex production ...]]></bibo:abstract>
  <bibo:uri><![CDATA[http://www.scielo.br/scielo.php?pid=...]]></bibo:uri>
  <bibo:language>por</bibo:language>
  <dct:isPartOf rdf:resource="http://aims.fao.org/serials/c_e8d916a8"/>
</bibo:Article>
```

Fig. 2: The RDF/XML serialization of an AGRIS record

datahub[5] portal included 34 open bibliographic linked open datasets, such as Europeana[6] and the move to RDF has enabled AGRIS to align with such important hubs of information science. This is achieved by providing disambiguated, entity-based access to bibliographic and citation data and by mashing up this information with related data sources. "The ambition is not to collect comprehensively all bibliographic references in the subject area, but to use the latent knowledge in the AGRIS data to find, link and interpret relevant sources on the internet." [7].

3. The Road to Linked Data

With the growth of open access repositories and web resources, the AGRIS objectives required a thorough revision in the strategies and methods to use for such a large dataset.

Following the innovation begun by the porting of its indexing thesaurus AGROVOC to a Simple Knowledge Organization System (SKOS) concept-scheme published as LOD [8], the decision was made to connect the AGRIS content to the LOD cloud as well, and to fully exploit the potentialities of the Semantic Web.

When the AGRIS team decided to publish its records as linked data in RDF, it quickly became clear that crucial metadata necessary in addressing issues of proof and trust were missing. It goes without saying that "as the number of repositories and aggregators increase, so too does the number of potential formal or informal metadata sources" [9]. Figure 1 shows the long flow of an AGRIS artifact, from genesis to dissemination. Every phase generates administrative metadata, and for each record, AGRIS has always registered authors, titles, dates and the cataloguing institute.

3.1. AGRIS RDF datasets

Becoming part of the LOD cloud meant translating the repository of over 5 million XML bibliographic records to RDF, and publishing it on the Web. The translation process required a design step to define vocabularies and properties needed in order to model data as statements (triples), and the creation of a translation mapping between the AGRIS XML bibliographic repository and the AGRIS RDF. The AGRIS XML bibliographic repository is encoded in an XML qualified Dublin Core metadata format, so each XML element has a specific meaning. To move to RDF, it was important to map XML elements to RDF properties with the same meaning. Keeping in the spirit of reusability, the team avoided almost completely minting any new properties and to reuse as much as possible available standard vocabularies, such as BIBO, FOAF and Dublin Core. Reusing existing and widespread vocabularies helps applications to better know the meaning of a property, so to display, interlink, or use the content in the best way and with the correct meaning. We defined new properties only when there was nothing in the world that can express a desired relationship.

[5] http://datahub.io/
[6] www.europeana.eu/

Currently, apart from links to external linked datasets, AGRIS data comprises information from three different RDF data sources maintained in FAO:

- The *AGRIS Records* dataset: the direct translation of AGRIS XML records to RDF. This dataset consists of more than 130 million triples resulting from the conversion of the above mentioned XML repository. The relevancy of the dataset derives not only from the long process of disambiguation and data cleaning, but also on the multilingual assets obtained by crossing its data records with the multiple editions of records available in the diverse countries which provide it with data. Figure 2 shows an example of an RDF/XML AGRIS record.
- The *AGRIS Serials* dataset: since over 80% of AGRIS records are journal articles, it was thought to generate an appropriate authoritative dataset of agricultural journals, with comprehensive information about each entry such as ISSN, subjects, dates, frequency, publisher. The metadata information from the "AGRIS scientific journals", together with other important data sources such as the journals of the FAO catalogue, the Directory of Open Access Journals (DOAJ[7]), Centre for Agricultural Bioscience International[8] (CABI) and the National Agricultural Library Catalog[9] (AGRICOLA) was expanded by retrieving authoritative information from the ISSN Centre database[10]. Currently the team has disambiguated nearly 20,000 agricultural journals and rendered them unique entities, stored in a triple store, for a total of almost 320,000 triples.
- The *Agrovoc* dataset[11]: Agrovoc is the world's most comprehensive multilingual agricultural vocabulary. Agrovoc contains more than 32,000 concepts in over 20 languages, covering subject fields in agriculture, forestry and fisheries together with cross-cutting themes such as land use, rural livelihoods and food security. A great part of the AGRIS records are indexed with Agrovoc concepts, and interlinks to other thesauri, such as Eurovoc[12], NAL[13] and DBpedia[14].

[7] http://www.doaj.org/
[8] http://www.cabi.org/
[9] http://agricola.nal.usda.gov/
[10] http://www.issn.org/
[11] http://www.semantic-web-journal.net/system/files/swj274_1.pdf
[12] http://eurovoc.europa.eu/
[13] http://agclass.nal.usda.gov/
[14] http://dbpedia.org

3.2. The RDF-ization process

Both the AGRIS XML core dataset and the RDF dump are periodically updated and disseminated as Open Data at an FTP site. The extremely performing and flexible Apache Solr technology behind the index allows easy updates, specifically when there is a requirement to change parts or entire records. For 35 years or so, the AGRIS network always provided a stable environment, officialized by an agreement signed by all the United Nations Member Countries in 1974. In addition, in the last few years, journal publishers, Open Repositories and Service Providers adhered to AGRIS, and this increased enormously the AGRIS core data, which is continuously growing at a rate of 150,000 records each year.

The repository is accessed by many applications that pre-date the RDF porting of its content.

For this reason, the translation to RDF requires continuous automatic updates. Moreover, the process is based on operations of data transformation and enrichment, so it requires well defined and structured steps with a precise flow of information (again see Figure 1).

To begin with, a unique URI is generated for each AGRIS resource: each of the published URIs represents a unique means to identify a specific resource and provides researchers with a partial map of the global research community, linking formal outputs (papers, conference proceedings, etc.) with other information available in other datasets.

The URI is generated deterministically from available identifiers (see Figure 2, where the identifier associated to *dct:identifier* has been used to also generate the local name of the URI), thus there is no need of a component for managing the consistency of URIs along different translations. Then various steps cover the process of translation of the XML documents to RDF. Figure 1 shows these conversion steps. The process takes as input a set of AGRIS XML records and loads them into the database in an object-relational data model. This step is part of a data validation process necessary to ensure that no duplicate records will be added to the repository and to modify data using the common query language SQL. These data are consumed by a filter that transforms them (data type conversion, data cleaning, data validation) and enriches them with information taken from different controlled and authoritative sources, like the ISSN Centre database and the FAO Open Archive[15].

[15] http://www4.fao.org/faobib/

At the end, data are ready to be converted to RDF, loaded into a triple store and disseminated to the Web.

This process of RDF-ization is completely automatic: data come directly from structured AGRIS records, so mapping can be between the AGRIS XML and the AGRIS RDF, with a few filters to check the correctness of some information (e.g. date formats, ISSNs, etc.) and to add more data by using predefined data sources that have been previously mapped to AGRIS records. Thus, when new records are added to the AGRIS repository, they are automatically translated to RDF without any loss of reliability as no new uncertainty is added to the dataset. Where critical operations, such as mappings from AGROVOC to other datasets and interactions with data owners, such as the ISSN International Centre[16], require human intervention, the rest of the translation process is controlled by filters that check data consistency and correctness in the metadata ingested by AGRIS data providers.

Maintaining a wealth of curated knowledge - such as the OpenAGRIS Semantic application - entails the implementation of a mechanism for keeping track of provenance information of the data sources, as will be discussed in Section 4.

3.3. AGRIS Dataset Links

Information about AGRIS is available from the FAO web site[17], while more specific information about the AGRIS Dataset is available on AIMS[18].

A description file following the VoID (Vocabulary of Interlinked Datasets) specification [10] is available alongside the AGRIS Linked Open Dataset[19]. The VoID file contains statistical information about the linked dataset (partitioned into two subsets: AGRIS records and AGRIS journals), as well as coordinates (such as the http address of its SPARQL endpoint) for automatically accessing and properly querying it. More information on the VoID data about AGRIS is reported in section 4.

4. AGRIS Provenance and trust

The issues discussed in the previous section refer to the world of Linked Open Data (LOD) and are typically partially or wholly ignored in closed world systems. In open world systems, data is designed to be accessed by potentially any person or machine, and it is then that problems of proof and trust move to the forefront. OpenAGRIS, as a system designed to access such data, has to cope with these issues and has been an ideal vehicle to get to the top proof/trust layers of the LOD stack to experience the issues firsthand.

Provenance in particular is a broad term that may refer to various levels of granularity. Data provenance was not a historical concern as AGRIS always took for granted that the agreements in place with national governments (and the national libraries) provided AGRIS with a license allowing its secretariat to ingest and disseminate data without specific rights or provenance statements. With the shift to digital publishing and machine-readable records however, tracing the provenance chain gained new importance. In AGRIS, each record has an identifier known as ARN (AGRIS Record Number), which has a predefined structure and contains information on the data source together with the year of creation of the artifact. With this unique identifier, AGRIS provides precise and updated statements of the origins of the artifact itself. For instance, the ARN "ES2011001090" refers to a record created in 2011 from a specific AGRIS data provider located in Spain, whose progressive number is 1090. Especially for legacy data, ARNs are very important pieces of meta-information, since it is very difficult to retrieve provenance information for decades-old records with poor metadata [11]. Thus, the team triplified information about the AGRIS data providers, providing unique URIs for each and adding triples (using the property *dct:source*) to identify this aspect of provenance, which previously was only implicit in the ARN.

In such a way, using the ARN of a record, it is possible to assign provenance information to an entire record, which is a set of triples. But the AGRIS triplestore currently contains more than 130 million triples and, ideally, each triple should be linked to administrative metadata. In fact, information about a specific AGRIS record could come from different entities, each of whom would provide part of the metadata during the record lifecycle. It should be possible to know where the information was extracted from, who submitted it and the primary data source. Although this primary source – trustworthy or not – creates the information, it is often only possible to determine the last provider in the chain.

Questions remain on the appropriate set of properties with which to encode such metadata. The Vo-

[16] http://www.issn.org/

[17] http://agris.fao.org/knowledge-and-information-sharing-through-agris-network

[18] http://aims.fao.org/aos/agrovoc/

[19] http://agris.fao.org/void.ttl

cabulary of Interlinked Datasets (VoiD) has been a likely candidate. Intended as a bridge between publishers and consumers of RDF data, it is organized around four metadata areas: descriptive metadata, access metadata, structural metadata and *linking metadata* that is "helpful for understanding how multiple datasets are related and can be used together" [10]. For descriptive metadata, VoiD recommends Dublin Core (DC) and Friend of a Friend (FOAF) properties which together cover basic provenance issues. However to more fully cover provenance, extensions such as the EnAKTing Group's voidp Vocabulary for Data and Dataset Provenance [12] are desirable while earlier initiatives such as the Open Provenance Model [13] suffer from complex serializations and no reuse of existing properties.

Looking at metadata domains other than provenance, VoiD also covers license issues by using DC extensions and contains some recommendations regarding common license types. Where VoiD and its extensions are silent is in the trust layer area of quality which will become an important issue as more competing Linked Data comes online and machines need to make dynamic value judgments on which sources to prefer. Though beyond the scope of the proposed work in AGRIS, an interesting initiative which bears future examination is Olaf Hartig's tRDF, which "proposes a set of criteria to assess the quality of Linked Data sources" [14].

5. The OpenAGRIS Semantic Mashup

In the last two years the AGRIS project has focused its attention on the metadata which end users are mostly interested in and on the many ways to enrich these metadata. AGRIS references often suffer from a lack of complete information and in particular of full text links. Only 4% of the entire collection has a working full text link. Accordingly, if a user wants to get more information on a specific topic they must use Google or other search engines to retrieve the publication.

The AGRIS team thought then to treat AGRIS records abstractly as metadata sets that could be leveraged to automatically access and display related data. This approach led to the development of OpenAGRIS, a semantic mashup portal that aggregates information from different Web sources using AGRIS records exposed as sets of triples in a Linked Open Data environment. An AGRIS record represented in RDF (see Figures 2, 3) thus becomes the entry point for a mechanism that discovers related web resources primarily via AGROVOC keywords. AGROVOC contains many alignments to other vocabularies (e.g. DBpedia, FAO Geopolitical Ontology, etc.) that allow querying triple stores to retrieve external resources. Moreover, AGROVOC keywords can also be used to query traditional Web Services (e.g. World Bank, FAO fisheries dataset, etc.) to retrieve non-RDF data. The system currently displays production statistics of aquatic species, species occurrence maps, World Bank indicators and more, all dynamically queried through a constellation of related keywords and vocabulary alignments.

OpenAGRIS is an environment that allows the team to test and raise issues related to the implementation of end user systems based on the Semantic Web, and helps in spotting problems in the proof and trust layers of the AGRIS system as well as in most of the systems from which it retrieves data. As previously noted, the first major issue to arise was that of provenance. AGRIS records contained only partial provenance metadata, and most external sources that were accessed contained even less. DBpedia is a typical example in that by its very nature, wiki data is highly collaborative and almost immune to strong provenance tracking.

Licensing is another aspect which is often underestimated and/or kept off the explicit data level. Quality indicators are also typically non-existent. Finally, there are semantic issues related to vocabulary alignment that can impact the correctness of dynamically retrieved data. A common situation is that two vocabularies representing the same concept with the same name are attached to data that is contextually very different and even wrong from the user's point-of-view. For instance, one vocabulary may have a commercial view of a concept while another has a scientific view. SKOS' deliberately fuzzy definition of *exactMatch* and *closeMatch* properties, while avoiding ontological over-commitment, does little to assist in this regard.

5.1. Details of the OpenAGRIS platform

The core of the OpenAGRIS web application[20] is the AGRIS RDF dataset, stored in an Allegrograph[21] RDF store, which is the same source exposed on the LOD as a SPARQL endpoint.

By starting from the resource requested by the user and identified by the already mentioned ARN code

[20] http://agris.fao.org/openagris/
[21] http://www.franz.com/agraph/allegrograph/

OpenAgris queries the Agrovoc RDF repository to extract keywords for the specific resource and relationships to other datasets, such as DBpedia. If the resource is a journal article, the engine also queries the AGRIS Serials dataset, obtaining complete information about the journals and related articles from the same journal covering similar topics. This process can be extended to all areas of interest in the linked open data cloud, obtaining all possible information about the specific resource and its main topics.

From an architectural point of view, since filtered SPARQL queries on a triple store are not very efficient, OpenAGRIS uses an Apache Solr index to immediately display information about the record. These indexes, instead of merely providing indexed search over RDF literals, enable semantic shortcuts between elements bound by arbitrary RDF paths, in a way which is specialized for AGRIS constructs. Then, various threads are responsible for querying the other linked triple stores by means of their SPARQL endpoints, when available, or other APIs.

6. Conclusions

The process of moving the 5 million AGRIS data over to RDF and along the way disambiguating and enriching the records was one of the necessary steps in the process of rethinking bibliographic databases for the semantic web. Along the way the AGRIS team discovered a number of related benefits, in that the radical openness of the LOD paradigm put into discussion curation practices and shed light on historic weaknesses in the collection of administrative metadata.

The ways in which users search and discover information has changed enormously with the rise of the Internet. It is becoming less and less likely that users are browsing bibliographic data in order to find a print publication that they will then order from a publisher or research library. Users are more and more expecting to simply search subject categories in a generic search engine and click on the results to open full text journal articles, conference proceedings and book excerpts and whatever else will enable them to answer their queries. This implies that traditional bibliographic databases need to find new ways in which to offer value to end users. They must leverage their often rich set of authoritative, disambiguated and interrelated metadata to take advantage of the new sea of external data that is increasingly being exposed via web services and exploited using technologies such as Asynchronous JavaScript and XML (AJAX).

The OpenAGRIS pilot does just that. By combining disambiguated, authoritative records for journals together with a rich set of multi-lingual thesaurus concepts that are in turn broadly linked to other subject thesauri, OpenAGRIS is able to maximize a powerful set of related metadata by using it first against an ever increasing set of Semantic Web services and then pragmatically against other, useful web services whether semantic or not.

In its original conception OpenAGRIS was to be a purely RDF-based application. Experience showed that only a few domains intersecting with the agricultural domain offered RDF web services. Pragmatic considerations of basic data availability thus necessitated the use of non-RDF APIs.

Despite the relative lack of RDF-based APIs for agricultural use, the field is growing. As new, relevant RDF APIs become available, OpenAGRIS will continue to exploit them, while at the same time doing its part by producing AgriBase, a Freebase-like aggregation API for agricultural data. This challenging endeavor will necessitate wrapping and reinterpreting non-RDF APIs. It will be available via SPARQL endpoint as well as via RDF and non-RDF web services.

Along with a recent rise in the number of RDF-based services, data providers are increasingly providing web-service widgets returning complex functionality such as static and dynamic tables, maps, charts and graphs. OpenAGRIS will also begin to exploit such rich services, relying on its unique set of citation-based metadata relationships to deliver relevant information in graphical form to end users.

Finally, behind the scenes, the group responsible for the publication of AGRIS has been quietly publishing and linking an ever increasing set of authoritative RDF vocabularies, from thesauri, to journals, to corporate authors. This growing web of formally linked metadata will continue to improve the relevance of AGRIS records as they reach ever further into the linked data cloud.

References

[1] Food and Agriculture Organization of the United Nations, "AGRIS introduction," Food and Agriculture Organization of the United Nations, printed booklet, 1981

[2] F. Celli, S. Anibaldi, M. Folch, Y. Jaques e J. Keizer, «OpenAGRIS: using bibliographical data for linking into the agricultural knowledge web» [Online]. Available:

http://www.fao.org/docrep/015/an642e/an642e00.pdf, 2011

[3] B. D'Arcus and F. Giasson, "Bibliographic Ontology Specification,". [Online]. Available: http://bibliontology.com/, 2008

[4] "Friend Of A Friend Ontology (FOAF)," [Online]. Available: http://xmlns.com/foaf/spec/, 2010

[5] DCMI Usage Board, «DCMI Metadata Terms», [Online]. Available: http://dublincore.org/documents/dcmi-terms/ 14, June 2012

[6] G. Dunsire, «Bibliographic Data Cluster», [Online]. Available: http://www.w3.org/2005/Incubator/lld/wiki/Cluster_BibData 10 February 2011

[7] Food and Agriculture Organization of the United Nations (FAO of the UN), "AGRIS 2008-2010: A portal to access global knowledge in agricultural research and technology", 2008

[8] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jacques e J. Keizer, «Thesaurus Maintenance, Alignment and Publication as Linked Data - The AGROVOC Use Case», *International Journal of Metadata, Semantics and Ontologies (IJMSO)*, vol. 7, n. 1, pp. 65-75, 14 August 2012

[9] T. Emma e J. Allinson, «Signed metadata: method and application» in *International Conference on Dublin Core and Metadata Applications (DC2006)*, [Online]. Available: http://dcpapers.dublincore.org/pubs/article/view/861/857, Manzanillo, Colima, Mexico, 2006

[10] K. Alexander, R. Cyganiak, M. Hausenblas and J. Zhao, "Describing Linked Datasets with the VoID Vocabulary, W3C Interest Group Note 03 March 2011, " World Wide Web Consortium. [Online]. Available: http://www.w3.org/TR/void/, 2011

[11] Y. Jaques, S. Anibaldi, F. Celli, I. Subirats, A. Stellato and J. Keizer, "Proof and Trust in the OpenAGRIS Implementation," in *International Conference on Dublin Core and Metadata Applications (DC2012)*, The Kuching Proceedings [Online]. Available: http://dcpapers.dublincore.org/pubs/article/view/3665, 2012

[12] O. Tope e C. Gutteridge, «voidp: A Vocabulary for Data and Dataset Provenance,» EnAKTing. [Online]. Available: http://www.enakting.org/provenance/voidp/, 2011

[13] L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, B. Plale, Y. Simmhan, E. Stephan e J. Van den Bussche, «The Open Provenance Model core specification » (v1.1), Future Generation Computer Systems, Vol. 27, n. 6, pp. 743-756, [Online]. Available: http://dx.doi.org/10.1016/j.future.2010.07.005, 2011

[14] O. Hartig e A. Flemming, «Quality Criteria for Linked Data sources», [Online]. Available: http://sourceforge.net/apps/mediawiki/trdf/index.php?title=Quality_Criteria_for_Linked_Data_sources, 2010