IFLA 2014 LYON

Submitted on: 06/08/2014

The role of AGRIS in providing global agricultural information to boost productivity and food security

Thembani Malapela
Food and Agriculture Organization of the United Nations, Rome. Italy
thembani.malapela@fao.org

Fabrizio Celli
Food and Agriculture Organization of the United Nations, Rome. Italy
fabrizio.celli@fao.org

Imma Subirats
Food and Agriculture Organization of the United Nations, Rome. Italy
imma.subirats@fao.org

Johannes Keizer
Food and Agriculture Organization of the United Nations, Rome. Italy
johannes.keizer@fao.org

Abstract:

Access to agricultural information ensures that stakeholders in the farming system can make informed decisions towards increasing agricultural productivity. Agricultural information and data are stored in distributed systems and repositories (using different pieces of software and databases) that expose a variety of research outputs with various metadata formats. Often grey literature, journal articles and technical reports are lost as technologies are not employed to ensure that such resources are online, accessible and shared widely within the agricultural stakeholders. This paper reviews AGRIS (International System for Agricultural Science and Technology) as an international collaboration and partnership in collecting and using agricultural bibliographic data to enable researchers and policy makers retrieve related agricultural and scientific information and data. AGRIS database, therefore, uses bibliographic data as an aggregator of locating related content across information systems available on the Web through taking advantage of the Semantic Web and Linked Open Data technologies. Besides linking related content, AGRIS has a potential to gather like-minded researchers around AGRIS content to discuss and share ideas – this implies creating a social media layer by creating AGRIS social media.

Keywords: AGRIS, AIMS, AGROVOC, Agriculture, Linked Data, Agricultural Information, food productions, bibliographic datasets.

1.0 Introduction

Agricultural information is a key ingredient in boosting agricultural productivity; hence, institutions are investing in research to provide solutions to problems be-delving agriculture. Emerging technologies are set to improve agricultural research and examples have been adapted in geographic information systems, global positioning systems, remote sensing, agricultural statistical data, and early warning systems amongst many research based systems. Initiatives are on-going to embrace open data and the implementation of data infrastructures in the management of agricultural data and information both at technical and political level. These developments are ushering in a new era that will affect farmers' day-to-day operations and increase their ability to access agricultural knowledge and enable also researchers to access agricultural research information[1].

Currently, agricultural information is housed distributed in uncoordinated systems and repositories which expose a variety of research outputs using different pieces of software and databases. Today's researcher is lost in the following range of content types - photos, satellite images, statistical data, journal articles, published reports, books, graphs, maps and other geographic information, genomic information, and these coming in various formats and standards. It is from these sources that farmers, researchers, policy makers, academics, governments, and multinational institutions need to delve into during the research process (to both produce and consume these research outputs). The development in the semantic technologies provides a possibility to interlink different kinds of research outputs. In this sense, AGRIS (International System for Agricultural Science and Technology) is a mashup application that allows users to query the AGRIS content, interlinking all resources to external sources of information.

This paper seeks to review an international perspective of collaboration in collecting and exposing AGRIS bibliographic data to enable researchers and policy makers to retrieve agricultural and scientific information. A brief overview of the literature on initiatives in open data in agriculture and their possible impact on agricultural productivity and food security will be offered to contextualise the contributions of AGRIS. The paper will further explain the current status of the AGRIS repository. The AGRIS repository seeks to use bibliographic data as an aggregator of locating not only the full-text of the article, but also related content across information systems available on the Web through taking advantage of the possibilities offered by the Semantic Web. The paper proposes a plan to establish a social AGRIS layer over bibliographic records, where authors and scientists can discuss various issues emanating from the literature and of interest to agricultural scientists.

[1] Discussions are on-going on *Empowering Smallholder Farmers with Open Agricultural Data* by Agrilinks sponsored by USAID, see http://agrilinks.org/blog/empowering-smallholder-farmers-open-agricultural-data-recap-february-askag-twitter-chat [2014, April 23]

2.0 Brief Literature Review

Literature is awash [2] with authorities detailing the concept of Linked Open Data and numerous interest groups and communities in LOD have been established over the years. The AIMS team of the Food and Agriculture Organization have published articles[3] on the use of semantic Web standards to improve the open unrestricted availability and accessibility of agricultural information and data. Baker & Keizer (2010:178) reviewed the strategies of using descriptive metadata, thesauri, and ontologies for integrating access to information from a variety of sources for both scientific and technical audiences. Subirats et al., (2012) and Subirats (2014) provide a detailed explanation of how the AIMS team implemented Linked Open Data in the agricultural domain.

However, few studies are available about interlinking open data and food productivity or food security or the impact of thereof. Van Vark (2013) explained how open data in agriculture can achieve food security; he explained that the first step is to make agricultural data available, readable (interoperable) and thereafter develop user applications that facilitate data querying, and re-use. If this is achieved, as stated by the World Bank (2013) open data combined with agricultural knowledge can support farmers, and this was aptly phrased, 'Imagine creating the ability for farmers to use open data to understand what crops grow best where, or what prices can be expected after harvest, or how best to solve weather, blight or other challenges to yield. Open data combined with other tools such as cellular phones can do just that,' World Bank (2013, n.p).

A number of projects and organisations in the agricultural domain already exist to expose their datasets or either to implement open data projects. For instance, Kabore, Madalli and Keizer (2014) reviewed the latest developments of opening and linking wheat agricultural data within the framework Research Data Alliance[4] initiative. Other projects include CAB's Plant wise Knowledgebase, Cassava base, a number of open data projects within CGIAR centres, and the Global Open Data for Agriculture and Nutrition (GODAN)[5]. This paper presents how AGRIS seeks to provide to the community of agricultural information specialists and researchers access to agricultural data, information, and knowledge for them to tackle farm productivity and improve food security. The development path of AGRIS traditional bibliographic search engine to its present state has been well documented over the years by Jaques, et al., (2012), Anibaldi, et al., (2013), and Celli and Keizer (2014).

[2] For example google scholar search for linked open data yields 3,910,000 results http://scholar.google.co.za/scholar?hl=en&q=Linked+Open+Data&btnG=&as_sdt=1%2C5&as_sdtp= [2014, April 20]

[3] Search these by keywords FAO or FAO AIMS on E-LIS repository. http://eprints.rclis.org/ [2014, April 22]

[4] Research Data Alliance. https://rd-alliance.org/ [2014, April 26].

[5] Global Open Data for Agriculture. http://godan.info/ [2014, May 3].

3.0 Background to AGRIS

AGRIS can mean three different things – AGRIS as a network, AGRIS as web portal and AGRIS as a database. Firstly, AGRIS network refers to the collaborative network of more than 150 institutions from 65 countries. Secondly, AGRIS is a database with more than 7.7 million structured bibliographical records on agricultural science and technology. Finally AGRIS as a Web portal (http://agris.fao.org/) refers to the Web application that links the AGRIS knowledge to related Web resources using Linked Open Data methodologies. The purpose of AGRIS is to provide comprehensive scholarly research information in specific fields in the agricultural domain. The Food and Agriculture Organization of the United Nations and its technical partner Agro-Know[6] maintains the AGRIS database. AGRIS accepts content from publishers related to agriculture, forestry, animal husbandry, aquatic sciences, fisheries, and nutrition. Some of the records received have a link to the full text and AGRIS becomes an indexer to the respective repositories. AGRIS data is indexed in Google (since 2008), and this offers greater visibility of AGRIS content.

In the formative years, AGRIS grew from data received from AGRIS centers; however, AGRIS later began to ingest data from journal publishers. Celli, et al., (2013) explains how AGRIS can ingest metadata, either by pulling them through harvesting from clients (e.g. aggregators, institutional repositories) or by pushing data from clients (national libraries or journal publishers). Over the years technology has changed: however, AGRIS's stable and structured metadata, and the use of AGROVOC[7] concept scheme for indexing have provided bedrock for sustainability of content migration across technological changes. Currently, AGRIS tries to accept all metadata formats and the following formats/standards are still highly recommendable: PubMed NLM XML, METS XML, DOAJ XML, MODS XML, AGRIS AP, MARCXML, OAI-DC, ENDNOTE XML, and SIMPLE_DC. After the importing of AGROVOC to a Simple Knowledge Organisation System (SKOS) concept scheme published as Linked Open Data, also the AGRIS content was provided as LOD to fully exploit the potentialities of the Semantic Web. Thus, the AGRIS repository was born from a collaborative effort of AGRIS Centres, later journal publishers, research institutions, and FAO.

Before content linking (see Fig 1, below), AGRIS data comprise of three RDF sources which are the AGRIS records dataset, the AGRIS serials dataset and the AGROVOC dataset (Anibaldi et al., 2013). Over 80% of AGRIS records are journal articles with appropriate metadata description, for example showing the following properties - ISSN, date of publication, subjects, frequency, and publishers. Some of the records come with a URL linking to the full text: it was estimated that 25% of the records have links to full-text articles. Today AGRIS exists as an example of a collaborative and partnership arrangement, collecting bibliographic references and using semantic technologies to facilitate better retrieval of agricultural information. In December 2013, the new AGRIS website[8] (AGRIS

[6] Agro-Know. http://www.agroknow.gr/agroknow/ [2014, May 4].
[7] AGROVOC. http://aims.fao.org/standards/agrovoc/about [2014 May 2].
[8] AGRIS 2.0 http://agris.fao.org/agris-search [2014, March 28].

2.0) was launched with new search functionalities. AGRIS is heavily used with an average of 300,000 visits per month from more than 190 countries.

4.0 AGRIS path to RDF and Open AGRIS Mashup

AGRIS is an RDF-aware system, a mashup application that allows users to query the AGRIS content, interlink all resources to external sources of information. The linkages are made possible by the use of AGROVOC, a multilingual vocabulary containing more than 40,000 concepts available in more than 21 languages. AGROVOC itself is published as LOD and is part of the LOD cloud. As stated earlier, AGRIS data consumption is comprised of centralisation of data collected from data providers and then interlinking this data with other kinds of information related to the bibliographic records in AGRIS and the AGRIS domain. These kinds of information include maps, statistics, and country profile information.

In brief, the RDF-ization process includes the translation of AGRIS AP XML database into RDF and the selection of the external datasets to be interlinked to AGRIS. In the transition to RDF, a unique permanent URI is assigned to each ABGRIS record. Other URIs included are for AGROVOC keywords and AGRIS Journals. To assure provenance, each AGRIS records have a unique identifier (ARN) with a predefined structure and contains implicit information about the data source and year of creation. The information about AGRIS data providers is exposed as RDF, so that each AGRIS data provider has its own de-referentiable URI. To enable the interlinking to external databases, two options are available: first, the AGRIS application can rely on AGROVOC formal alignments to other thesauri (skos:exactMatch, skos:closeMatch) to access external sparql endpoints, where metadata have been indexed with thesauri different from AGROVOC; secondly, AGRIS can query external Web services using scientific names extracted from AGROVOC . This is necessary to ensure that the content automatically retrieved from databases available on the Web is related to an AGRIS record.

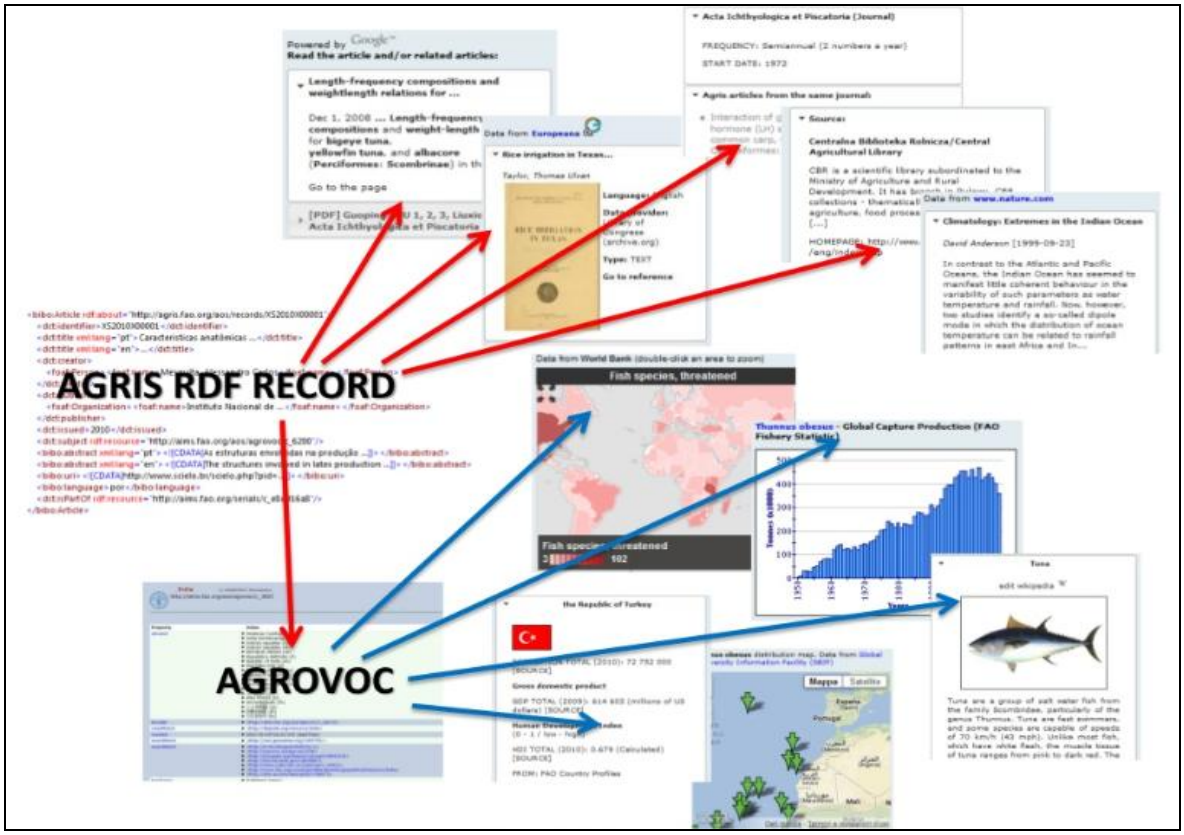

Fig 1: Content interlinking in AGRIS

To exemplify the benefits of interlinking AGRIS content to external datasets, one needs to play around with the platform and observe the system pulling content from different sources. The International Food Policy Research Institute (IFPRI) case exemplifies this process: when the user types a search on the AGRIS portal, the system shows some results and the user can select one of them; if the selected record was indexed with at least one AGROVOC keyword referring to a country name, the systems can query the IFPRI sparql endpoint to retrieve some information about the country and display it on the screen (See Fig 2).

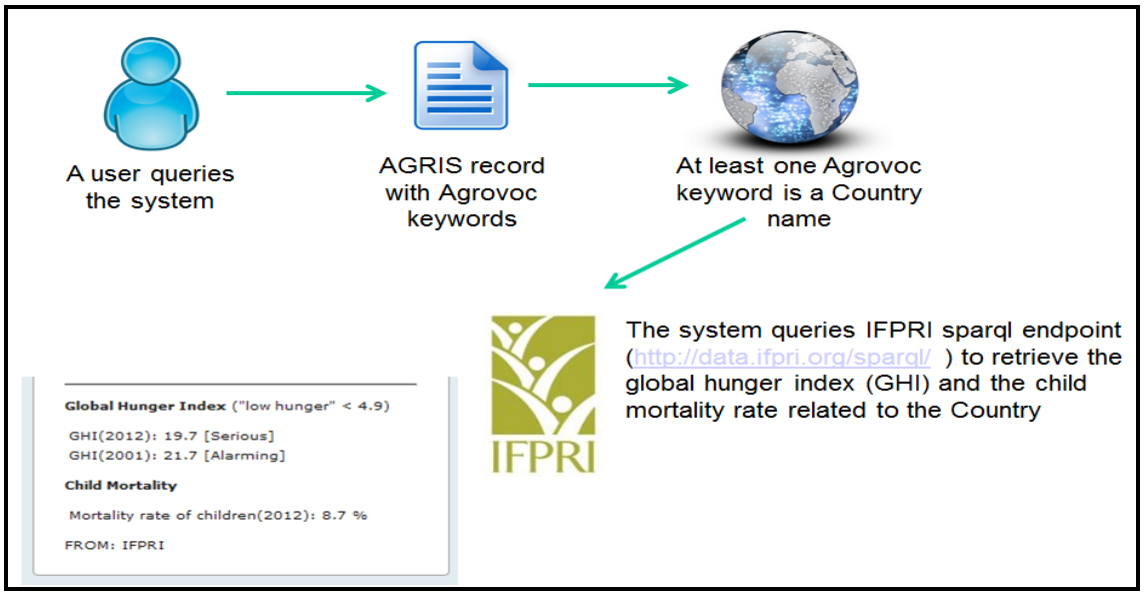

Fig 2: The IFRI case

5.0 AGRIS user needs survey

In order to ensure AGRIS' usability and to evaluate new features of AGRIS 2.0 a survey of users was carried out. One of the new feature of AGRIS 2.0 is the ability to filter records by collapsing and adding AGROVOC keywords. The purpose of the AGRIS survey was to get user's feedback on the platform so as to further inform and adapt existing information resources according to user requirements. The 30 question survey[9] was divided into the following parts, general appearance, search function, search results listing, search results display and general observations. The target audience was AGRIS users who include information management specialists, librarians and software developers, to researchers and students. The survey was distributed in English and Spanish. The overall result was a positive appreciation of AGRIS 2.0 features and the interface was considered to be clear and intuitive.

The recommendation included,

- improving the location of the advanced search function,
- reducing the number of boxes in the homepage,
- to add additional information like year and type of publication as well as information about the availability of full text was requested,
- to provide the original link of the source,
- to review the relevance of the information displayed from external resources

Work is underway to address these issues and improve the AGRIS 2.0 and assimilate the proposed and other new features. Many changes are foreseen, as the customization of the views. AGRIS users come from diverse backgrounds and disciplines, therefore AGRIS will provide users with an opportunity to register and choose their screen display. Some users may want to access the old fashion bibliographical view to make specific boolean queries to the

[9] Evaluation of the usability features of AGRIS 2.0.
http://aims.fao.org/sites/default/files/AGRIS_Survey_Results.pdf [2014 , April 30]

system, other users may want to select the external sources of information to be displayed on the screen, others may want to change colours. By the end of 2015, AGRIS 3.0 will be published, with many new features and the possibility of social navigating the information in agriculture.

6.0 Conclusion

The foregoing case has shown that using semantic technologies have enabled the aggregation of various research outputs into a single interface. AGRIS has shown the prowess of Linked Open Data and related semantics in the provision of agricultural information and data. The visits on AGRIS 2.0 and the collaboration in growing its content and technologies testify the AGRIS key role in providing essential information for agricultural productivity. However, even if agricultural information and data is available and used at research and farming system level, there remains a risk of falling back to a situation where there is no information available at all. There is a need to develop applications that can consume these data and make it available at farming level or user level; such applications could enable querying sub-disciplinary subjects as crops, weather, planting and pests data and information. For example, in Sub-Saharan Africa and Asia mobile agriculture is revolutionising and impacting greatly on farming information systems, which 'extentionists' and researchers could fully benefit from.

7.0 Glossary of Terms

AIMS stands for Agricultural Information Management Standards is a space for accessing and discussing agricultural information management standards, tools and methodologies, connecting information workers worldwide to build a global community of practice. AIMS is a community of more than 1,800 people. More information: http://aims.fao.org/

AGRIS stands for International System for Agricultural Science and Technology and currently contains more than 7 million bibliographic references on agriculture research and technology and links to related data on the Web. AGRIS can mean three different things – AGRIS as a network, AGRIS as web portal and AGRIS as a database. More information http://agris.fao.org/agris-search/index.do

AGROVOC is a controlled vocabulary covering all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations, including food, nutrition, agriculture, fisheries, forestry, environment etc. http://aims.fao.org/standards/agrovoc

8.0 References

Anibaldi, Stefano and Jaques, Yves and Celli, Fabrizio and Stellato, Armando and Keizer Johannes. 2013. Migrating bibliographic datasets to the Semantic Web: The AGRIS case. Semantic Web. *In Semantic Web Journal*. http://www.semantic-web-journal.net/system/files/swj463.pdf

Baker, Thomas and Keizer Johannes.2010. *Linked Data for Fighting Global Hunger: Experiences in setting standards for Agricultural Information Management.* http://www.fao.org/docrep/014/am324e/am324e.pdf [2014, May 6].

Celli, Fabrizio and Keizer Johannes. 2014. *AGRIS: an RDF-aware system in the agricultural domain*. A presentation in the LOD2014 Conference, Rome, Italy 20-21 February 2014.

Celli, Fabrizio and Jaques, Yves and Anibaldi, Stefano and Keizer, Johannes. 2013. Pushing, Pulling, Harvesting, Linking - Rethinking Bibliographic Workflows for the Semantic Web. In EFITA, Torino (Italy), 23-27 June 2013.

Jaques, Yves and Anibaldi, Stefano and Celli, Fabrizio and Subirats-Coll, Imma and Stellato, Armando and Keizer, Johannes. 2012. *Proof and Trust in the OpenAGRIS Implementation., 2012.* In International Conference on Dublin Core and Metadata Applications 2012, Kuching, Sarawak, Malaysia, 3-7 September 2012.

Kaboré, Esther Dzalé Yeumo, Madalli, Devika and Keizer, Johannes.2014.Opening and Linking Agricultural Research Data. *D-Lib Magazine*. Vol. 20 (1&2): http://www.dlib.org/dlib/january14/kabore/01kabore.html [2014, March 30].

Subirats-Coll, Imma and Malapela, Thembani and Dister, Sarah and Zeng, Marcia and Goovaerts, Marc and Pesce, Valeria and Jaques, Yves and Anibaldi, Stefano and Keizer, Johannes. 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*. 2012. In 6th Metadata and Semantics Research Conference (MTSR 2012), Cádiz, Spain, 28-30 November 2012.

WorldBank. 2013. *Open data+ agriculture can transform farmers response to crisis*. http://www.worldbank.org/en/news/feature/2013/04/26/open-data-can-transform-farmers-response-to-crisis [2014, March 13].