# AGRIS - From a bibliographical database to a Web data service on agricultural research information

Angela FOGAROLLI, Dan BRICKLEY, Stefano ANIBALDI, Johannes KEIZER

OKKAM Project, University of Trento (Italy), afogarol@disi.unitn.it
FAO of the UN, johannes.keizer@fao.org
FAO of the UN, stefano.anibaldi@fao.org
FAO of the UN, dan.brickley@fao.org

## Abstract

AGRIS has for many years provided a huge collection of bibliographic references, such as research papers, studies and thesis, each including metadata such as conferences, researchers, publishers, institutions, and keywords from different thesauri as AGROVOC. With the rise of full text search and online availability of more research material, the role for bibliographic metadata can appear redundant. When considered instead as a form of modelling that emphasizes relationships, connections and links, bibliographic metadata grows in value as the Web grows in connectivity, and can provide researchers with a map of the global research community, linking formal outputs (papers, data) with a wider grey literature (preprints, drafts) and with communication platforms (blogs, forums) that help researchers put formal findings into a wider context. This paper aims to describe the evolving role of the AGRIS bibliographic database that becomes a hub of agricultural research literature. The huge silo of 3 millions agricultural resources, collected by more than 150 institutions over the last 35 years, becomes the starting point to access the diverse knowledge in agricultural science and technology available globally on the Web.

## Introduction

Through exploring the evolving role of databases such as AGRIS, it has become clear that the connectivity patterns amongst the things described in the database (researchers, topics, institutes, places) can be better reflected online through a more explicit representation both in Web metadata and in user-facing Web sites. The distributed nature of the world described by AGRIS naturally fits a "linked data" deployment model, in which AGRIS becomes more than a document discovery portal - it becomes an entry point and map of the entire research landscape around some topic or theme.
The Linked-data[2] techniques foster the link between resources through the Web.
Such an approach requires an emphasis on sharing identifiers, names and descriptions of key real-world and abstract objects other than the bibliographic materials themselves: conferences, workshops, research centres, researchers, subject themes, homepages.
None of this is news to the bibliographic professional; such concerns have been at the heart of metadata work for years. What is new today is the presence of tools (standards, software) and community trends (open linked data, open archives, RSS/Atom syndication) that allow the full potential of such link-oriented metadata to be exploited.

## An entity centric approach to data aggregation

The idea of shifting the web from a huge graph of documents to a huge graph of data has become more and more popular since when Tim Berners-Lee proposed the idea of the Semantic Web. Scientists and practitioners have invested a lot of effort to realize this vision, often trying to adapt and reuse models and techniques originated from more traditional areas like databases and AI. However, there is a very important difference between traditional knowledge-based systems, and the current work aiming at reaching semantic computing at web scale: the notion of global interlinking of distributed pieces of knowledge.

At the base of such interlinking - and the resulting semantic interoperability of fragments of data - is the notion of identity of and reference to entities. Systems that manage information about entities (objects/individuals/instances...) commonly issue identifiers for these entities, just in the way relational databases issue primary keys for records. If these identifiers are generated by the information system itself, several issues arise that hinder interoperability and integration considerably: (i) a proliferation of identifiers is taking place, because the same object is potentially issued with a new identifier in several information systems; (ii) injectivity of identifiers cannot be achieved, i.e. one identifier can denote different entities in different information systems; (iii) reference to entities across information systems is very complicated or impossible, because there are no means to know how an entity is identified in another system.

To overcome this lack of data-level integration, OKKAM [1] proposes a global, public infrastructure, called Entity Name System (ENS), which fosters the systematic creation and reuse of identifiers for entities in the global space of information and knowledge. This “a priori” approach enables systems to reference the entities which they describe with a globally unique identifier, and thus create pieces of information that are semantically pre-aligned around those entities. Semantic search engines or integration systems are then able to aggregate information from distributed systems around entities in a precise and correct way. We call this the “entity-centric approach” to semantic interoperability, and the resulting information/knowledge space the Web of Entities.

## The OKKAMization Process

The OKKAMization is the process necessary to include entities in existing information sources in the web of entities. The process involves the identification of entities inside existing repositories and the creation of unique identifier (OKKAM ID) for entities which are not already present in the ENS system.

Creating an OKKAM ID for an object means get a unique identifier which is a non-ambiguous way to refer at that object without ambiguity. An OKKAM ID is a well formed URI which enables to semantically connect to other global resources.

To allow the correct creation of OKKAM identifier is necessary to collect a minimal set of information about an entity. This minimizes the risk of ambiguities. If, for example, we imagine to create an OKKAM ID about Mr. John Smith, just using his name, the result will be an OKKAM ID that refers to a person. However, OKKAM will not be able to identify uniquely this entity because there are many “John Smith” in the world. Building an OKKAM ID with more information as state, city, work, allows to better recognize the “right” “John Smith”.

The OKKAMization process of AGRIS repository is composed of four phases:

1. ***Corpus entity recognition.*** This activity focuses on entity recognition inside the AGRIS repository and related sources.
2. ***Associate OKKAM ids to extracted entities.*** This task is based on the matching of extracted entities against the OKKAM ENS. If a match can be found for an entity then the identifier is re-used otherwise a new entity profile is created and thus a new unique identifier for the extracted entity.
3. ***Enrichment of the AGRIS repository with OKKAM identifiers.*** The OKKAM identifier generated in phase two are included in the XML files of the repository as another type of metadata. This allows automatically identifying and aggregating entities inside the repository. The core point and objective of this phase is to enable entities based retrieval and to semantically connect entities in different contexts. Thus, from the user point of view this will translate in an efficient retrieval avoiding information overloading.
4. ***RDF enrichment of the AGRIS repository.*** This step has a big impact on the Web. It consists in describing the AGRIS repository using the RDF notation. Publishing the repository using RDF makes the content of the repository understandable by external semantic search engines (SIG.MA, Google Project...). Therefore hidden semantic connections among entities can be discovered and showed to the user. Entities form the AGRIS repository can be described in RDF or mircroformats in

other web resources such as the FAO website and this will increase the semantic information that can be aggregated for the same entity.

## The AGRIS linked-data model

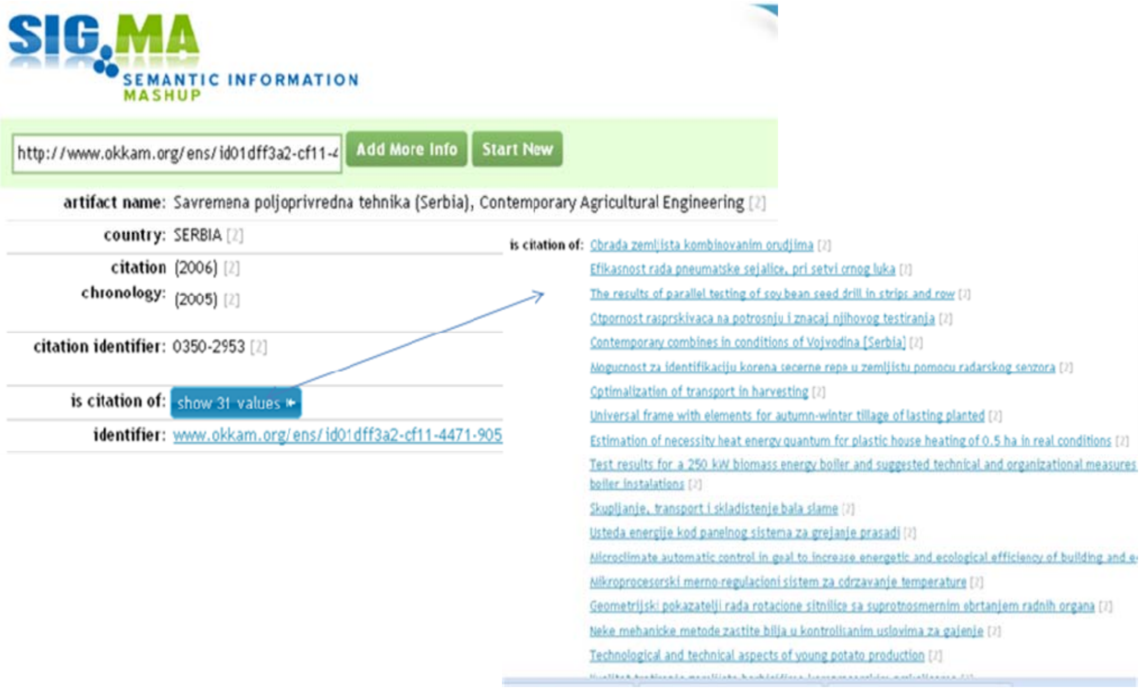
The AGRIS repository is a large and rich collection of bibliographic references encoded in a qualified DC XML format. Each XML document is structured in a metadata description for a resource which is sometimes available in PDF format.

In this section we report about a first experience for enabling linked data in AGRIS using the OKKAM ENS infrastructure.

In order to create a liked data model for AGRIS, we followed the OKKAMization steps described in the previous section. In more details we first (step 1) decided to focus the experience on the journal entity type in order to show an immediate advantage of applying a linked-data model to the AGRIS corpus for than extending the approach to other entity types such as author. Secondly, through the OKKAM ENS search API we obtained unique identifiers for each journal. Then we show how the unique identifiers are introduced in the original repository files (step 3) and lastly each file is translated in RDF format and submitted to a Semantic Web Search Engine (step 4).

The objective of assigning unique identifier to entities in the AGRIS repository leads to a light-weight data integration of entities and in this way enables inter-linkage among entities, which can come from different information sources, as shown in Fig.1 and 2. As a result, efficient information retrieval will be enabled within the AGRIS repository and in the global scale by interlinking with other information sources.

In Fig.1 and Fig.2 we show some examples of the result of the OKKAMization process exposed in a Semantic Web Search engine.



http://www.okkam.org/ens/id01dff3a2-cf11-4

artifact name: Savremena poljoprivredna tehnika (Serbia), Contemporary Agricultural Engineering [2]

country: SERBIA [2]

citation (2006) [2]

chronology: (2005) [2]

citation identifier: 0350-2953 [2]

is citation of: show 31 values

identifier: www.okkam.org/ens/id01dff3a2-cf11-4471-905

is citation of: Obrada zemljista kombinovanim orudjima [2]
Efikasnost rada pneumatske sejalice, pri setvi crnog luka [2]
The results of parallel testing of soy bean seed drill in strips and row [2]
Otpornost rasprskivaca na potrosnju i znacaj njihovog testiranja [2]
Contemporary combines in conditions of Vojvodina [Serbia] [2]
Mogucnost za identifikaciju korena secerne repe u zemljistu pomocu radarskog senzora [2]
Optimalization of transport in harvesting [2]
Universal frame with elements for autumn-winter tillage of lasting planted [2]
Estimation of necessity heat energy quantum for plastic house heating of 0.5 ha in real conditions [2]
Test results for a 250 kW biomass energy boiler and suggested technical and organizational measures boiler instalations [2]
Skupljanje, transport i skladistenje bala slame [2]
Usteda energije kod panelnog sistema za grejanje prasadi [2]
Microclimate automatic control in goal to increase energetic and ecological efficiency of building and e
Mikroprocesorski merno-regulacioni sistem za odrzavanje temperature [2]
Geometrijski pokazatelji rada rotacione sitnilice sa suprotnosmernim obrtanjem radnih organa [2]
Neke mehanicke metode zastite bilja u kontrolisanim uslovima za gajenje [2]
Technological and technical aspects of young potato production [2]

Fig.1 Semantic Search by OKKAM id for a journal

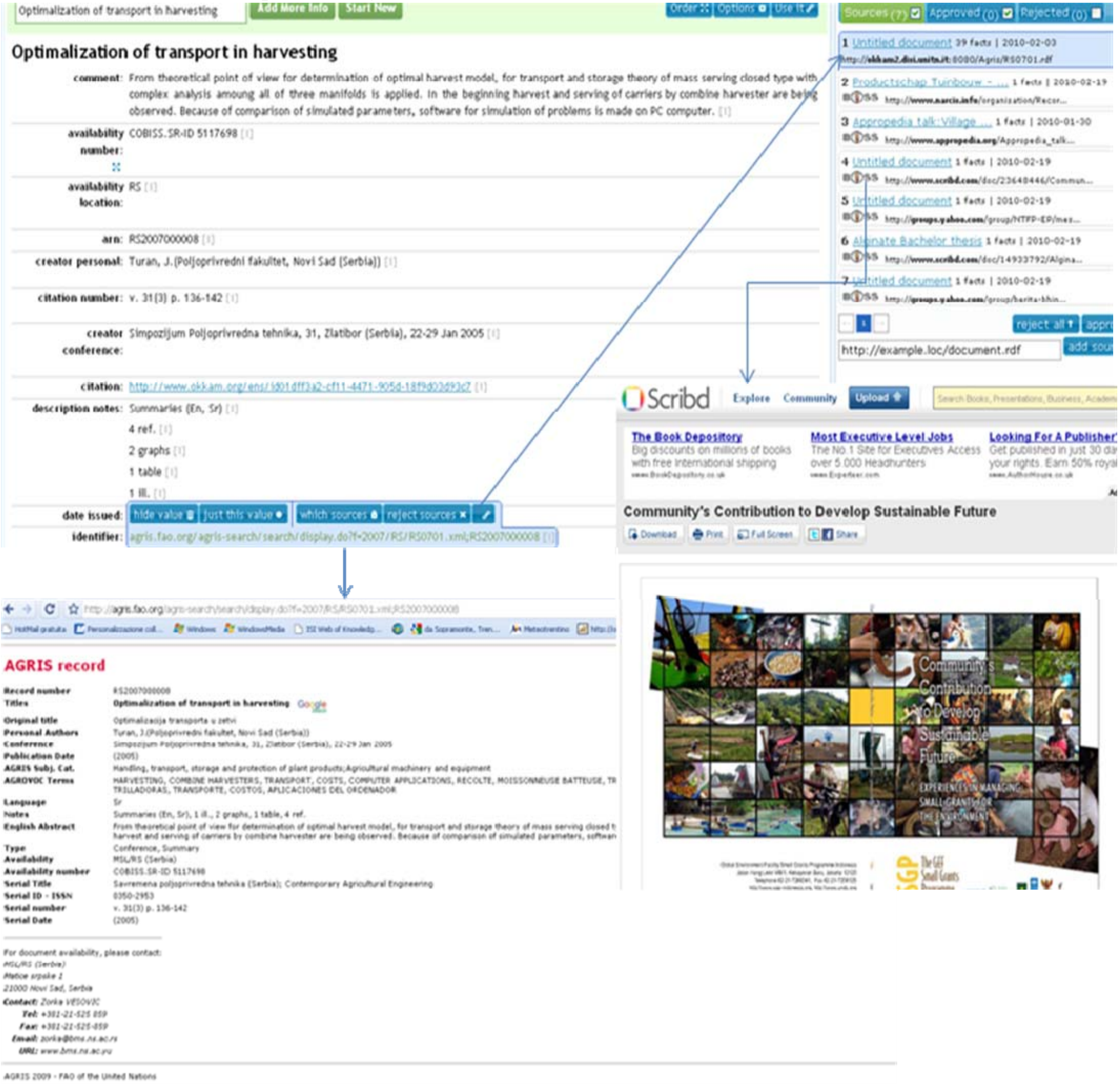
In Fig.1 it is shown a result of a search by unique identifier for a journal. The interface shows different statements about the journal resource and some of the attribute a clickable to enter in a deeper level of detail.

In this example, a click on the “is citation” attribute allows to display all the article titles for that journal. A further click on one of the article titles displays the details of the article (see. Fig.2).

Information about the OKKAMized resource can be aggregated from different sources.

The sources are displayed on the right side of Figure 2. And a click on an attribute can explore in more details the attribute itself. If the value of an attribute is an URL this can connect with external information sources. For example in Figure 2 from the article details it is presented the possibility to navigate to the AGRIS Website or to other external related sources.

The amplitude of the inter-linkage with external resources grows with the use of the same unique identifier whenever that particular journal is mentioned in the Web.



Optimalization of transport in harvesting

comment: From theoretical point of view for determination of optimal harvest model, for transport and storage theory of mass serving closed type with complex analysis amoung all of three manifolds is applied. In the beginning harvest and serving of carriers by combine harvester are being observed. Because of comparison of simulated parameters, software for simulation of problems is made on PC computer. [1]

availability number: COBISS.SR-ID 5117698 [1]

availability location: RS [1]

arn: RS2007000008 [1]

creator personal: Turan, J.(Poljoprivredni fakultet, Novi Sad (Serbia)) [1]

citation number: v. 31(3) p. 136-142 [1]

creator conference: Simpozijum Poljoprivredna tehnika, 31, Zlatibor (Serbia), 22-29 Jan 2005 [1]

citation: http://www.okkam.org/ens/id01dff3a2-cf11-4471-905d-18f9d03d93c7 [1]

description notes: Summaries (En, Sr) [1]
4 ref. [1]
2 graphs [1]
1 table [1]
1 ill. [1]

date issued:

identifier: agris.fao.org/agris-search/search/display.do?f=2007/RS/RS0701.xml;RS2007000008 [1]

1 Untitled document 39 facts | 2010-02-03
http://okkam2.disi.unitn.it:8080/Agris/RS0701.rdf

2 Productschap Tuinbouw - ... 1 facts | 2010-02-19
http://www.narcis.info/organization/Recor...

3 Appropedia talk:Village ... 1 facts | 2010-01-30
http://www.appropedia.org/Appropedia_talk...

4 Untitled document 1 facts | 2010-02-19
http://www.scribd.com/doc/23648446/Commun...

5 Untitled document 1 facts | 2010-02-19
http://groups.yahoo.com/group/NTFP-EP/mes...

6 Alginate Bachelor thesis 1 facts | 2010-02-19
http://www.scribd.com/doc/14933792/Algina...

7 Untitled document 1 facts | 2010-02-19
http://groups.yahoo.com/group/barita-bhin...

The Book Depository
Big discounts on millions of books with free international shipping
www.BookDepository.co.uk

Most Executive Level Jobs
The No.1 Site for Executives Access over 5.000 Headhunters
www.Experteer.com

Looking For A Publisher'
Get published in just 30 da
your rights. Earn 50% roya
www.AuthorHouse.co.uk

Community's Contribution to Develop Sustainable Future

Download
Print
Full Screen
Share
Community's Contribution to Develop Sustainable Future
EXPERIENCES IN MANAGING SMALL GRANTS FOR THE ENVIRONMENT
The GEF Small Grants Programme

http://agris.fao.org/agris-search/search/display.do?f=2007/RS/RS0701.xml;RS2007000008
HotMail gratuita
Personalizzazione coll...
Windows
WindowsMedia
ISI Web of Knowledg...
da Sopramonte, Tren...
Metaotrentino
http://

AGRIS record

| | |
|---|---|
| Record number | RS2007000008 |
| Titles | Optimalization of transport in harvesting |
| Original title | Optimalizacija transporta u zetvi |
| Personal Authors | Turan, J.(Poljoprivredni fakultet, Novi Sad (Serbia)) |
| Conference | Simpozijum Poljoprivredna tehnika, 31, Zlatibor (Serbia), 22-29 Jan 2005 |
| Publication Date | (2005) |
| AGRIS Subj. Cat. | Handling, transport, storage and protection of plant products;Agricultural machinery and equipment |
| AGROVOC Terms | HARVESTING, COMBINE HARVESTERS, TRANSPORT, COSTS, COMPUTER APPLICATIONS, RECOLTE, MOISSONNEUSE BATTEUSE, TR TRILLADORAS, TRANSPORTE, COSTOS, APLICACIONES DEL ORDENADOR |
| Language | Sr |
| Notes | Summaries (En, Sr), 1 ill., 2 graphs, 1 table, 4 ref. |
| English Abstract | From theoretical point of view for determination of optimal harvest model, for transport and storage theory of mass serving closed t harvest and serving of carriers by combine harvester are being observed. Because of comparison of simulated parameters, softwar |
| Type | Conference, Summary |
| Availability | MSL/RS (Serbia) |
| Availability number | COBISS.SR-ID 5117698 |
| Serial Title | Savremena poljoprivredna tehnika (Serbia); Contemporary Agricultural Engineering |
| Serial ID - ISSN | 0350-2953 |
| Serial number | v. 31(3) p. 136-142 |
| Serial Date | (2005) |

For document availability, please contact:
MSL/RS (Serbia)
Matice srpske 1
21000 Novi Sad, Serbia
Contact: Zorka VESOVIC
Tel: +381-21-525 859
Fax: +381-21-525-859
Email: zorka@bms.ns.ac.rs
URL: www.bms.ns.ac.yu

AGRIS 2009 - FAO of the United Nations

Fig. 2. AGRIS data linked to other web resources

As mentioned before, the references of the AGRIS repository are encoded in a XML format. This type of file can be enriched with unique identifiers and this will allow the future representation of the unique identifier on the AGRIS web page enabling record linkage to the Web of entities.

A snapshot of the XML of the AGRIS resource with a unique identifier for the journal in which the article appears is shown here:

```
<ags:citation>
        <ags:citationTitle>Savremena poljoprivredna tehnika Serbia)</ags:citationTitle>
        <ags:citationTitle>Contemporary agricultural engineering</ags:citationTitle>
        <ags:citationIdentifier scheme="ags:ISSN">0350-2953</ags:citationIdentifier>
        <ags:citationNumber>v. 31(1-2) p. 29-37</ags:citationNumber>
        1<ags:citationChronology>(2005)</ags:citationChronology>
        <OkkamID>http://www.okkam.org/ens/ida53b7142-5880-4684-aab3-f83c2a6d0ea8</OkkamID>
</ags:citation>
```

It follows an example of the automatic generated RDF file for each AGRIS resource article.

The unique identifier for the journal is enlightened in red colour. The journal attributes are described inside the *rdf:Description* tag for the resource with a specific unique identifier (.*rdf:about="*okkam_id value").

Whenever the journal is cited again inside the AGRIS record, the unique identifier will be used for its description. In green colour a reference to the original ARGIS website is underlined. This link will ensure reference to the AGRIS web search interface from any Semantic Web search engine.

```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:j.0="http://models.okkam.org/ENS-core-vocabulary.owl#"
   xmlns:j.1="http://purl.org/dc/terms/"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:j.2="http://purl.org/agmes/1.1/" >
  <rdf:Description rdf:about="http://www.okkam.org/ens/id01dff3a2-cf11-4471-905d-18f9d03d93c7">
   <j.2:citationChronology>(2006)</j.2:citationChronology>
   <j.0:artifact_name>Savremena poljoprivredna tehnika (Serbia), Contemporary Agricultural Engineering</j.0:artifact_name>
   <j.2:citationChronology>(2005)</j.2:citationChronology>
   <j.2:citationIdentifier>0350-2953</j.2:citationIdentifier>
   <j.0:country>SERBIA</j.0:country>
  </rdf:Description>
  <rdf:Description rdf:about="http://agris.fao.org/agris-search/search/display.do?f=2007/RS/RS0701.xml;RS2007000023">
   <j.2:subjectThesaurus>PROPIEDADES TECNICAS</j.2:subjectThesaurus>
   <j.2:subjectThesaurus>TRITICUM</j.2:subjectThesaurus>
   <j.2:subjectThesaurus>http://www.fao.org/aos/agrovoc#c_2572</j.2:subjectThesaurus>
   <j.2:subjectThesaurus>http://www.fao.org/aos/agrovoc#c_2386</j.2:subjectThesaurus>
   <j.2:creatorConference>Simpozijum Poljoprivredna tehnika, 32, Zlatibor (Serbia), 28 Jan - 4 Feb 2006</j.2:creatorConference>
   <j.1:abstract>The paper shows presentation of the exploitational examination results for the wheat drill sowing agregates. Some technical - technological drill solutions and the results of the working quality (norm, drilling depth) and the exploitational parameters (working speed, output) have been shown.</j.1:abstract>
   <j.2:creatorPersonal>Mehandzic, R.(Poljoprivredni fakultet, Novi Sad (Serbia). Departman za poljoprivrednu tehniku)</j.2:creatorPersonal>
   <j.2:creatorPersonal>Malinovic, N.(Poljoprivredni fakultet, Novi Sad (Serbia). Departman za poljoprivrednu tehniku)</j.2:creatorPersonal>
   <j.2:subjectThesaurus>SEMIS EN LIGNE</j.2:subjectThesaurus>
   <dc:type>K</dc:type>
   <j.2:subjectThesaurus>SEMOIR</j.2:subjectThesaurus>
   <j.2:ARN>RS2007000023</j.2:ARN>
   <j.2:descriptionNotes>3 tables</j.2:descriptionNotes>
   <j.2:subjectThesaurus>TECHNICAL PROPERTIES</j.2:subjectThesaurus>
   <j.2:descriptionNotes>2 ref</j.2:descriptionNotes>
   <j.2:citation rdf:resource="http://www.okkam.org/ens/id01dff3a2-cf11-4471-905d-18f9d03d93c7"/>
```

## Conclusions

In the AGRIS 2010 work [3], we have been prototyping a redesign for AGRIS that brings these concerns to the core of the system: both in our data modelling, and in the Web presence, AGRIS will emphasise the networked, linked nature of the things it describes. AGRIS has for many years provided a huge database of bibliographic references, such as research papers and thesis, each including metadata such as conferences, researchers, institutions, and keywords from different thesauri as AGROVOC.

For these reasons, the OKKAMization experiment explained above offers an effective and innovative solution for the global knowledge diffusion through semantic web technology.

The presented solutions allow information retrieval system to perform stronger automatic elaboration offering data identification and aggregation. OKKAM allows the AGRIS repository to acquire the added value of making its full content available to the global web and at the same time to combine and aggregate information between and outside the organizational boundaries.

## References

[1] Paolo Bouquet, Heiko Stoermer, Claudia Niederee, and Antonio Mana. Entity name system: The back-bone of an open and scalable web of data. *International Conference on Semantic Computing*, 0:554–561, 2008.

[2] Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), 1-22,2009

[3] Brickley, D., Anibaldi S., Picarella, A., Keizer, J. 2009. Designing AGRIS 2010- Information linking and Agricultural Research. URL: ftp.fao.org/docrep/fao/012/ak689e/ak689e00.pdf