



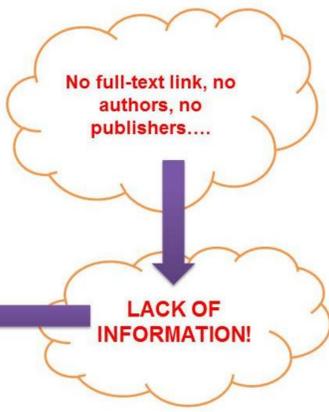
# The OpenAGRIS (s)mashup

## Using linked data and web services to augment AGRIS repository content

Fabrizio Celli, Stefano Anibaldi, Yves Jaques, Imma Subirats, Armando Stellato, Lim Ying-Sean and Johannes Keizer  
FAO of the United Nations, Rome (Italy)



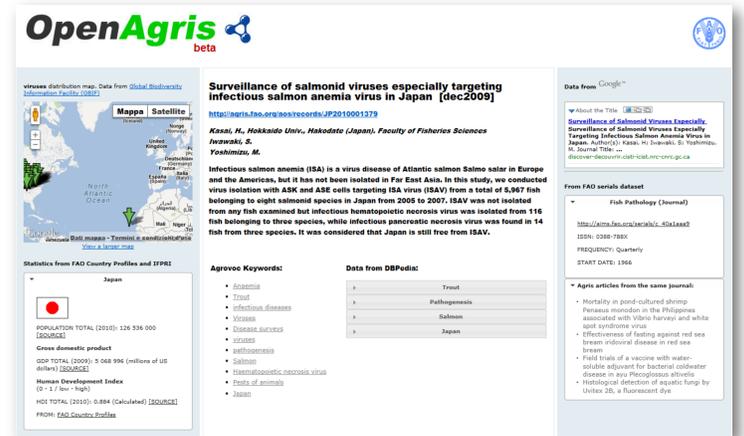
Dairy Herd Management (Jan 1977)	
How stable is your vacuum level? [Milking, dairy cows]	
Date of publication	Jan 1977
AGRIS Categories	Animal production
Language	English
Journal Title	Dairy Herd Management
Vol. No.	v.14(1) p. 10, 14-15



- Impossibility to interlink to other sources of data.
- The user is interested in the full-text of a document!!

AGRIS lack of information

Linked Open Data



The OpenAGRIS Web interface

### INTRODUCTION

“The AGRIS Network is an international initiative based on a collaborative network of institutions, whose aim is to promote free access to information on science and technology in agriculture and related subjects”. AGRIS (<http://agris.fao.org>), a collection of almost 3 million bibliographic references, is one of the most important world-wide information systems in the agricultural domain.

### PROBLEMS

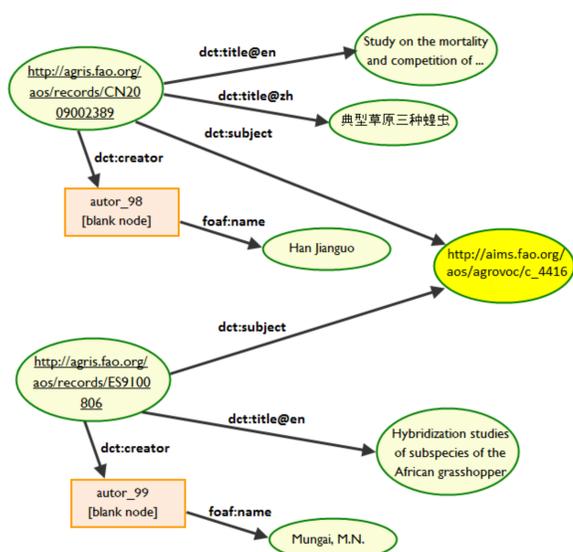
The quickly improving precision of commercial search engines coupled with the increasing availability of full-text documents was a mixed blessing for AGRIS. While on the one hand site traffic has never been higher, much of it is indexing activity from an army of web-bots and it remains difficult to increase user’s average time-on-site. Thus, there was the need to re-evaluate the possible uses to which bibliographic metadata may in the 21st century be put. Then, the analysis of the AGRIS content showed that sometimes AGRIS data lack of information, such as the full text link of the resource, so the resource is useless for the end user, who is interested in the entire document and not only in some incomplete metadata.

### DISAMBIGUATION & TRIPLIFICATION

- The team decided to transform AGRIS into a concept-based repository, requiring concept disambiguation and leading to the publication of records as Linked Data by carefully selecting RDF properties to allow reuse and interoperability
- The process of disambiguation was really challenging but within 2011 all keywords and journals were disambiguated and 60 million triples were published
- A set of challenges remain, such as disambiguating institutions and authors
- The, the team built a semantic mashup over the data, named OpenAGRIS
- OpenAGRIS uses AGRIS records to dynamically determine when it can retrieve related content from other providers such as the World Bank, Global Biodiversity Information Facility and DBPedia, which are then automatically displayed with the bibliographic record.

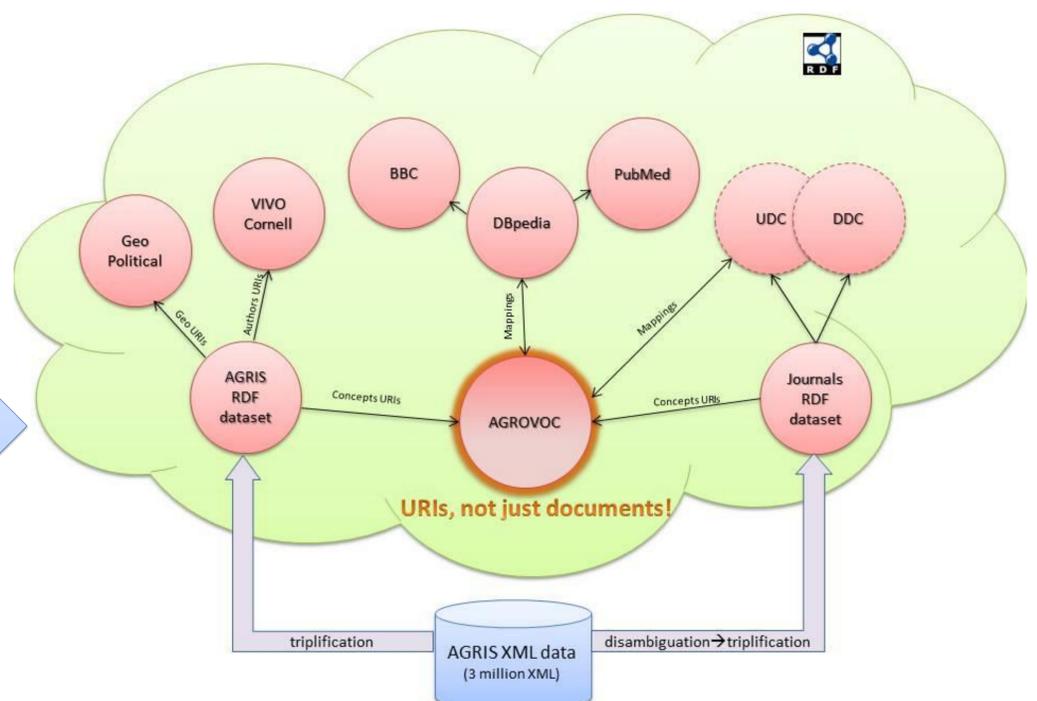
### CONCLUSIONS

- The conversion toward a concept-based repository and the services that exploit the additional capabilities such models provide show with a fresh eye the repository itself
- Revisiting the data helped the team to clean, disambiguate and improve the content and it also brought an updated selection of metadata properties that insured the content would be as widely interpretable as possible
- Moving towards a machine-readable language like RDF brought up interesting questions involving the proof and trust layers of the semantic web stack and how best to disseminate administrative metadata such as provenance, licensing and quality assurance
- The implementation of a mashup also made it clear how few repositories and other information systems are truly ready for the Semantic Web
- While OpenAGRIS was constrained to go beyond SPARQL and add internal ad-hoc metadata alignments and custom Web Services, it may encourage data providers to publish their data as Linked Data, providing SPARQL endpoints and services to access, reason and extract implicit knowledge



Fragment of the OpenAGRIS set of triples

AGROVOC is the backbone



The OpenAGRIS Data flow