

Work package number :	5	Start date or starting event:					
Work package title:	Business case 1: World-Wide Agriculture Monitoring						
Participant number:	2						
Participant short name	FAO	IXXO	USH	ONTO			
Person-months per participant: (i)							

Objectives (ii)

The objective of the work package is to create a dynamic linked data universe of agricultural scientific and technology knowledge. “Agricultural” in this context is understood in a broad sense covering all areas which fall under the mandate of the Food and Agriculture Organization of the United Nations, this includes fishery, forestry, nutrition and socioeconomic aspects of rural development.

This work package will transform the existing AGRIS (International Information Systems for Agricultural Science and Technology) database of FAO in a comprehensive gateway to relevant knowledge on the web.

Description of work (iii)

A cloud based linked data infrastructure will be created that should give comprehensive access to available agricultural scientific, research and technology knowledge on the web in various languages.

The starting point will be the existing 3 million records of the AGRIS database at FAO that will be the starting point to crawl the web for related content. These AGRIS records will be published as “linked data”. A crawler will be developed that finds related linked data sets, but also other data such as databases or unstructured text. Found “linked data” will be linked automatically to AGRIS linked data, whereas other material will be annotated and triples will be extracted to be added to the AGRIS linked data sets.

Initially approximately 70 million crawls will have to be conducted, to look for related information for all AGRIS information triples. (example: Title of a document, Author plus keywords, journal, Keyword combinations, Institutions plus keyword. Crawls with new information (about 20,000/month I guess have to be conducted with new AGRIS records that come in from human indexers in the AGRIS centers. The crawled material will be used twice to link to AGRIS records/triples to which it refers and will be triplified itself to enlarge the AGRIS knowledge base. The entire knowledge base will be used to periodically re-crawl the web to find newly added material.

Web services and an HTML interface will be developed to make this usable by machine agents and humans.

T5.1 Development of a “Agriadne”, a semantic annotator that triplifies unstructured text using custom vocabularies (especially the AGROVOC concept scheme) for identity extractions and creation of knowledge triples

T5.2 Development of AgriThresher a specific crawler that uses the AGRIS Index to crawl the web for related (linked data and unstructured) material and transfers this material to the Annotator for semantic analysis and extraction

T5.3 Development of an Infrastructure and Web Services which makes the AGRIS Linked Data Universe usable for search agents and persons.

T5.4 A user evaluation by agricultural scientists and extensionists will be conducted to measure the impact of the created infrastructure

Deliverables (iv)

D5.1.1 Data annotator v.0.5 M6-8

D5.2.1 Data crawler v. 0.5 M12-14

D5.3.1 Web Services and Interface v.0.5 M15

D5.4.1 User Evaluation 1 M 18

D5.1.2 Data annotator v.1 M 22

D5.2.2 Data crawler v. 1 M 26

D5.3.2 Web Services and Interface v.1 M 32
D5.4.2 User Evaluation 2 M 35*Notes*

- i. *For the coordinator include personnel effort needed for coordination and implementation of the network as well as any personnel effort for elaboration of deliverables; for all other beneficiaries, only the effort for elaboration of deliverables should be mentioned.*
- ii. *Provide a concise description of the objectives to be achieved within the work package and how these objectives will be pursued. Use quantifiable and verifiable elements. Refer to the tasks to be carried out.*
- iii. *Provide a short description of the work in this work package. State the role of the participants for the elaboration of each deliverable.*
- iv. *Provide a brief description of the deliverables including the month of delivery and the dissemination level. (note: it is expected that most of the deliverables will be publicly available):*