

Organizations at the workshop

"Improving Semantics in Agriculture"

Table of Contents

About the workshop	1
1. About the organizations	2
ASI, Aston, Bayer, CABI, Cornell, CGIAR, CRA-ABP, CIMMYT, Embrapa, INRA, IFPRI, ISI, IMC, KALRO, KTBL, Land Portal, Syngenta, USDA ARS, NAL, Wageningen UR Library, Wageningen Alterra	
2. Datasets	9
ASI, Bayer, CABI, CRA-ABP, CIMMYT, Embrapa, ISI, IMC, KTBL, Land Portal, Syngenta, USDA ARS, NAL, Wageningen UR Library, Wageningen Alterra	
3. Vocabularies	13
ASI, Aston, Bayer, CABI, Cornell, CGIAR, CRA-ABP, CIMMYT, Embrapa, Land Portal, ISI, IMC, KTBL, Syngenta, USDA ARS, NAL, Wageningen UR Library, Wageningen Alterra	
Appendix A: Links	18

Prepared by: Tom Baker, Caterina Caracciolo, Lori Finch, Phil Roberts
This report: http://aims.fao.org/sites/default/files/Agrisemantics_participants.pdf
Workshop report: http://aims.fao.org/sites/default/files/Report_workshop_Agrisemantics.pdf
Workshop site: <http://aims.fao.org/agrisemantics-workshop-2015>
Published: Rome: Food and Agriculture Organization of the UN
Date: 4 August 2015

About the workshop

With generous support from the Bill and Melinda Gates Foundation, a workshop was held in July 2015 at the headquarters of the Food and Agriculture Organization of the United Nations in Rome to assess ways of improving the role of "semantics" -- the vocabularies, ontologies, and schemas that give structure to repositories of research results and to agricultural datasets -- in order to strengthen our capacity to meet the challenges of global food security and climate change. For example, semantics can help us identify local varieties of a crop that are more resistant to climate change and to monitor their distribution across multiple language areas. Multilingual semantics can help users find and use data and research produced in different languages. Achieving this requires helping institutions, large and small, organize and publish their data in open and reusable formats.

By bringing together three major agricultural thesauri into a common global repository of conceptual and terminological information, the GACS ("Global Agricultural Concept Scheme") initiative provides a reference point around which we might construct a well-integrated clearinghouse of machine-readable semantic assets in agriculture and nutrition, such as vocabularies, code lists, ontologies, taxonomies, and statistical indicators. This work can benefit researchers, policymakers, extensionists and farmers by offering a backbone of aligned resources that simplify the common tasks of data standardization, comparison, integration, and interpretation. This can lead to accelerated innovation, facilitated academic research, more coherent and interoperable data, and systems and services that help farmers and smallholders become more efficient.

This report, which summarizes the answers to a questionnaire circulated to participant organizations before the workshop, briefly describes each organization and characterizes the datasets and vocabularies they maintain.

1. About the organizations

Agricultural Sustainability Institute (ASI)

The Agricultural Sustainability Institute at University of California, Davis, educates the next generation of leaders through hands-on experience at a 10-hectare Student Farm. Since 1993, ASI is advancing knowledge of water, energy, and soil management in a 100-year study (the Century Experiment) at the 120-hectare **Russell Ranch Sustainable Agriculture Facility**. ASI leads California-wide efforts to understand how agricultural research, education, and extension can improve the well-being of farm workers and food system workers. The **Sustainable Sourcing of Agricultural Raw Materials project** aims at promoting the long-term sustainability of the global food system by providing publicly available, scientifically validated information and tools to the broader food industry, using advances in information technology to measure and improve its sustainability. The project has prototyped a sophisticated **semantic web informatics structure** to manage the data and has developed a process for quickly winnowing the expansive **set of indicators** to a manageable set (approximately 10-20) that can be used to support a comprehensive understanding of sustainability in multiple dimensions for a given commodity in a given location.

In Rome, ASI is represented by two people. **Ruthie Musker** of the Sustainable Sourcing project. Ruthie helped with the initial use of controlled vocabularies for the platform and continues to update the list of terms used. She is also managing the use of controlled vocabulary of a dataset for the Century Experiment at the Russell Ranch. **Thomas Tomich**, founding director of ASI, has emphasized the importance of controlled vocabularies in ASI projects.

Aston Business School

Aston University (AST) is a long established research-led University known for its world-class teaching quality and strong links to industry, government and commerce. Aston Business School (ABS) is a triple-accredited business school (AMBA, EQUIS and AACSB), and has been ranked in the top 10 in the UK in the Eduniversal rankings for the third year running in 2013. ABS works in collaboration with industry, governments and the academic community to produce new research initiatives, with a particular focus on applied research that contributes to sustainable economic growth and development. The Business School has over 140 faculty members teaching and researching across all areas of business.

The **Operations and Information Management group** within the Business School has a world-class reputation in systems modelling and simulation, supply chain management, knowledge management, technology and operations management. The group is organised into four research teams concerning **Business Analytics, Operations Management, Global ICT Management, and Systems Modelling and Simulation**. It is a highly interdisciplinary group and has a growing body of research concerning agrifood, sustainability and the role of information and communication technology (ICT) in supply chains and operations.

In Rome, Aston Business School is represented by **Christopher Brewster** of Aston University of Bologna.

Bayer

With more than 400 employees, the **plant biotechnology Innovation center of Bayer in Ghent** is one of the most important R&D centers in the **Seeds division of Bayer CropScience**. Bayer CropScience is a global leader in the research, development, and marketing of agricultural and vegetable **seeds with added-value traits**. With headquarters in Monheim, Germany, Bayer CropScience has more than 23,100 employees in more than 120 countries. The center in Ghent focuses on agricultural crops such as oilseed rape, cotton, rice and wheat.

In Rome, Bayer is represented by **Erick Antezana** of the Computational Life Sciences (CLS) division, **Master Data for R&D department**, which maintains controlled vocabularies, taxonomies, and ontologies, as well as the platform which stores these artifacts.

CABI

CABI is an inter-governmental, not-for-profit organization, originally formed as the Commonwealth Agricultural Bureaux. CABI works in the **academic publishing, knowledge management, agricultural international development and scientific sectors**. CABI works to improve people's lives worldwide by providing information and applying scientific expertise to solve problems in agriculture and the environment.

In Rome, CABI is represented by **Anton Doroszenko, Thesaurus Manager** in charge of the management and upkeep of the CAB Thesaurus and associated authority files and participant in the GACS project during its first two phases; and **Phil Roberts** of the **Solutions and Architecture team**, who is in charge of data management, innovation, and the future strategy for knowledge creation.

Cornell University Library

John Ferreira works at Cornell University Albert R. Mann Library in the **Information Technology Services department** as a programmer/analyst/technology strategist with a focus on **Agriculture Information Systems and Semantic Web technologies**.

Consultative Group for International Agricultural Research (CGIAR)

CGIAR (the Consultative Group for International Agricultural Research) is a partnership addressing **agricultural research for development**. CGIAR contributes to the global effort to tackle poverty, hunger and major nutrition imbalances, and environmental degradation. Work is carried out by fifteen research centers, members of the CGIAR Consortium, with almost 10,000 scientists and staff in 96 countries, in collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations, and the private sector. CGIAR Research Programs are supported by a CGIAR Fund.

In Rome, CGIAR is represented by **Medha Devare** and **Melanie Bacou**.

Research Centre for Agrobiolgy and Pedology (CRA-ABP)

The Italian National Centre for Soil Mapping (CNCP) maintains the **Italian Soil Information System (SISI)** and collection of samples. CNCP was created in 1999 with a project that defined criteria and specifications for a soil map of Italy. CNCP belongs to the Agricultural Research Council (CRA), a National Research Organization operating under the Italian Ministry of Agriculture (MiPAAF) and located at the Research Centre for Agrobiolgy and Pedology of Florence (CRA-ABP). Currently, CNCP is funded by CRA and two international projects, the EU 7th Framework Programme (agINFRA) and the Azerbaijan Project (AZER), and belongs to the networks of other projects (INQUA Aeomed; CostAction Desertnub; Italy-Israel Ringo). CNCP maintains the soil database of Italy and a collection of several thousand of soil samples taken in Italy, Peloponnesus (Greece), Israel, and Azerbaijan.

In Rome, CNCP is represented by **Giovanni L'Abate**.

International Maize and Wheat Improvement Center (CIMMYT)

International Maize and Wheat Improvement Center (CIMMYT) belongs to a consortium of fifteen international agricultural research centers that is part of CGIAR. It is the lead center for **two CGIAR Research Programs, MAIZE and WHEAT**, and works as a partner in other programs. CIMMYT's **germplasm bank** holds large, unique collections of native maize and wheat varieties and wild relatives from key eco-regions. The seeds are conserved, studied, and shared under the terms of the International Treaty on Plant Genetic Resources for Food and Agriculture shipped to more than 600 partner organizations in 100 countries.

In Rome, CIMMYT is represented by **Rosemary Shrestha** of the **Data Management Unit (DMU)** under the Genetic Resources Program (GRP), which publishes data, with a focus on improving data management efficiency through integrated platforms, and implements CGIAR-compliant open-access policies. Rosemary introduces scientists and developers to new tools and standards and helps maintain the maize and wheat trait ontologies.

Brazilian Agricultural Research Corporation (Embrapa)

The Brazilian Agricultural Research Corporation (Embrapa) is a public corporation with private rights, which means it can reserve commercial rights on its products, services, or technologies. Embrapa works under the Brazilian Ministry of Agriculture, Livestock and Food Supply, **coordinates the Brazilian national system of agricultural research (SNPA)**, and so represents one of the major sources of agricultural information in Brazil. The SNPA consists of Brazilian public and private institutions and universities that collaboratively conduct agricultural research in centers all over Brazil and in some labs abroad, with a staff of 9,790.

In Rome, Embrapa is represented by **Ivo Pierozzi Júnior** of Embrapa Agricultural Informatics (EAI), a thematic research center in Campinas, São Paulo State, which develops computing and information technology applications for agriculture. Ivo implements KOS and terminologies to facilitate the collaborative production of databases and publications and to improve their global interoperability and visibility.

Food and Agriculture Organization of the UN (FAO)

The Food and Agriculture Organization (FAO) of the United Nations is an agency of the United Nations that leads international efforts to defeat hunger and malnutrition. FAO is a knowledge-based organization that helps its member states improve agriculture, forestry, and fisheries practices to work toward ensuring good nutrition and food security for all.

In Rome FAO is represented by **Johannes Keizer** of OPCCD, the open data team of the Office of Partnership and Capacity Development. OPCCD is responsible for promoting and supporting open data systems and common standards for sharing agricultural knowledge. Among the core products of OPCCD are **AGROVOC**, a multilingual agricultural concept scheme, **AGRIS**, a semantically-enabled agricultural bibliographic mashup, and **AIMS**, a community portal centered around good practices in agricultural information management.

French National Institute for Agricultural Research (INRA)

INRA is a public research organisation ranked second in the world and first in Europe for publications in **agricultural plant and animal sciences**. The missions of INRA are to serve the public interest by maintaining a balance between excellence of research and the demands of society; to produce scientific knowledge and innovation in the fields of agriculture, food and environment; and to contribute to the promotion of scientific and technical culture and debate. The INRA department of Scientific Information (DIST) provides innovative tools and services for managing data, information, and knowledge in support of research activities. DIST promotes open access to scientific and technical information and supports scientists in building and publishing domain vocabularies according to open standards.

In Rome, INRA is represented by **Esther Dzale Yeumo Kabore**, **Sophie Aubin** and **Pascal Aventurier**.

International Food Policy Research Institute (IFPRI)

IFPRI provides research-based policy solutions to sustainably **reduce poverty and end hunger and malnutrition in developing countries**. A research center of the CGIAR Consortium, IFPRI currently has more than 500 employees working in over 50 countries. IFPRI's Library and Knowledge Management unit (LKM) provides access to research results, supports IFPRI researchers in the use of tools, datasets, data visualizations, widgets, and internal blogs, and administers a library of books, journals, and datasets.

In Rome, IFPRI LKM is represented by **Melanie Bacou** and by **Cristina Perez**, Knowledge Management Systems Coordinator. Cristina manages the LKM website, creates LOD for IFPRI datasets and for VIVO, maintains the Agricultural Technology Ontology, and will coordinate an upcoming triple store for IFPRI publications.

Indian Statistical Institute (ISI)

Indian Statistical Institute (ISI) is devoted to research and higher education in statistics and to its application in natural and social sciences. Founded 1931, the institute gained the status of an Institution of National Importance by an act of the Indian Parliament in 1959. Its Documentation Research and Training Center (DRTC), established in 1962, aims at **developing expertise and excellence in different areas of information science**.

In Rome, ISI is represented by Devika Madalli.

Integrated Modelling Collaboratory (IMC)

IMC is an entity dedicated to modelling the semantic integration of many collaborators worldwide and publishes a semantic modelling platform. IMC maintains a set of ontologies and a core ontology for modeling and scientific observation. The main project under the integratedmodelling.org umbrella is the ARIES (Artificial Intelligence for Ecosystem Services), which provides a next-generation platform for enviro-social assessment and valuation, which is used by many groups worldwide.

In Rome, IMC is represented by **Fernando Villa**.

Kenya Agricultural and Livestock Research Organisation (KALRO)

KALRO is the premier institution in Kenya bringing together **research programmes in crops, livestock, biotechnology and socio-economics**. It envisions a vibrant commercially-oriented and competitive agricultural sector propelled by science, technology and innovation. KALRO promotes sound agricultural research, technology generation and dissemination to ensure **food security** through improved productivity and environmental conservation. KALRO goals include the dissemination and application of research findings in the field of agriculture and the establishment of a Science Park. KALRO coordinates **Kenya Agricultural Information Network (KAINet)**, a comprehensive national repository of research, extension, training and other materials related to research. It aims at making research open and accessible to various stakeholders.

In Rome, KALRO is represented by **Richard Kedemi**, head of KALRO's **Knowledge Management unit** and of KAINet.

Association for Technology and Structures in Agriculture (KTBL)

KTBL is a registered association in Darmstadt, Germany with a staff of approximately 70 people, supported by the German Federal Ministry of Food and Agriculture, with a mandate for **knowledge transfer of scientific findings into agricultural practice**. Its main objective is to promote environmentally friendly and society-embedded agriculture in accordance to the needs of consumers, particularly with regard to agricultural engineering techniques and methods in cultivation, species adapted livestock farming, landscape management, agricultural raw materials and energy production and recycling of organic wastes. With its activities, it supports policy makers and administrators in drafting legislation and helps to create regulations and legal instruments. Apart from printed publications containing planning and decision support data for farmers, consultants and administration, focus is increasingly leaning towards Internet-based information provision. With increasing value attributed to information, technologies of data exchange and information and knowledge management in agriculture have become a major field of activity for KTBL.

In Rome, KTBL is represented by **Daniel Martini**.

Land Portal Foundation

The Land Portal is a **global gateway for land related information** that is building a specialized community of land experts from civil society, government, intergovernmental agencies, and academia. The Land Portal Foundation aims to be *the* hub to centralize and exchange data and information about land. The Land Portal is a partnership facilitated by the International Land Coalition in collaboration with Actionaid, FAO, IFAD, AGTER, IALTA, and IEH. The Land Portal is an accessible, easy-to-use online platform to share land related information, monitor trends, and identify information gaps to promote effective and sustainable land governance.

At the workshop, the Land Portal Foundation is represented by **Laura Meggiolaro**.

Syngenta

The Swiss-based Syngenta AG is a **global agri-business committed to sustainable agriculture through innovative research and technology**. Its more than 28,000 employees in some 90 countries are dedicated to bringing plant potential to life.

In Rome, Syngenta is represented by **Derek Scuffell, R&D Data Strategist and Sensors Innovation Lead**.

USDA Agricultural Research Service (ARS)

The Agricultural Research Service (ARS) is the United States Department of Agriculture's (USDA) principal in-house research agency. ARS works to ensure that Americans have **reliable, adequate supplies of high-quality food** and other agricultural products. ARS accomplishes its goals through scientific discoveries that help solve problems in crop and livestock production and protection, human nutrition and the interaction of agriculture and the environment. ARS employs 2,000 scientists and postdocs and 6,000 other employees in 750 research projects at over ninety locations, including overseas laboratories.

National Agricultural Library (NAL)

Within ARS, the **National Agricultural Library (NAL)** is one of the four national libraries of the United States and houses one of the world's largest collections devoted to agriculture and its related sciences. The mission of NAL is to **collect, organize, preserve and provide access to global agricultural information**.

In Rome, ARS is represented by three people. **Dr Simon Liu, Associate Administrator for Operations and Management at ARS**, who provides leadership, manages, and is responsible for all aspects of ARS research operations. Simon recently served as Director of NAL, and was a principal collaborator in the Global Agricultural Concept Scheme (GACS). Previously, he was responsible for the **Unified Medical Language System**, which facilitates the development of computer systems that behave as if they "understand" the language of biomedicine and health. **Jeffrey White** of ARS conducts **research using simulation modeling** and related tools to understand potential impacts of climate change and adaptation to climate. **Lori Finch** is **Chief, Indexing and Informatics Branch at NAL**, and is responsible for the automated indexing of agricultural literature and the NAL Agricultural Thesaurus. Lori serves on the Working Group for GACS.

Wageningen UR Library

Wageningen University and Research Centre (Wageningen UR) is a Dutch public university in Wageningen, Netherlands, and includes agricultural research institutes formerly of the Dutch Ministry of Agriculture. Wageningen UR Library maintains the research information system for Wageningen UR, conducts bibliometric analyses, and maintains a **database of published datasets and of databases for Dutch vocational training**. The Library emphasizes research data management planning and provides courses on the subject for PhD researchers.

Wageningen Alterra research institute

Alterra, a Wageningen UR research institute, offers a combination of practical, innovative and interdisciplinary scientific research across many disciplines related to the green world around us and the **sustainable use of our living environment**. Aspects of our environment that Alterra focuses on include soil, water, the atmosphere, the landscape and biodiversity – on a global scale as well as regionally, from the Dutch polders to the Himalayas and from Amsterdam to the Arctic. Alterra promotes the Agriculture Model Intercomparison and Improvement Program

(AgMIP), which was formed in 2010 as a global network of agricultural modellers to provide better evidence of the impacts of climate change and food security. Circa 700 researchers from across the globe work on different aspects of agricultural systems modeling, which includes elements of regional and global integrated assessment of climate change impacts, modelling of sustainable farming systems and designing the next generation of agricultural models and data products.

In Rome, Wageningen UR Library is represented by **Hugo Besemer**, whose group is primarily involved with **bibliometric assessments and research data management**. Alterra and the AgMIP network are represented by **Sander Janssen**, leader of the **Earth Informatics team**.

2. Datasets available

Agricultural Sustainability Institute (ASI)

In the context of the **Sustainable Sourcing Project**, ASI has a database of over 2000 **indicators linked to controlled vocabulary terms**. ASI is connecting these indicators to **open-access spatial and tabular datasets** on a wiki. Ontology and data instances are made available as OWL and RDF files. ASI intends to make the datasets available as Linked Open Data but has not yet decided on a license for the data and needs to improve the quality of the formal representation of the data. Data from ASI's Russell Ranch project includes the **long-term core data from the Century Experiment**, including: crop yield, biomass and moisture data; cover crop biomass and moisture; crop elemental content; Normalized Difference Vegetation Index (NDVI) data, soil properties, weather, water elemental content, winter weed populations, and operational data including fertilizer and pesticide application amounts and dates, planting dates, planting quantity and crop variety, and harvest dates. A physical sample archive contains samples of crops, biomass, and soil.

Bayer

Bayer exploits publicly available resources, such as UniProt, in their research activities. They have developed proprietary datasets for specific R&D programs targeting specific **areas such as yield increase**. Some datasets get shared in specific situations, such as collaboration with academia. Bayer maintains an internal database of public and private datasets using Semantic Web standards and technologies, such as RDF and SPARQL, with RDF models customized to their needs.

CABI

On behalf of the UK government, CABI maintains **R4D (Research for Development)** -- an online portal about research funded by the UK Department for International Development (DFID) that is full open and available as RDF. CABI also maintains a more detailed version of **Geonames** and contributes its updates back into Geonames. CABI also has a number of products that are open access but have yet to be released as Linked Open Data. The plan is for these to be opened out further over the next six to twelve months. There are permission issues relating to how much can be Linked Open Data due to the nature of the data – especially those relating to pest distribution records around the world.

Cornell Library

The Cornell Library IT Services department manages a dataset called CUGIR. CUGIR is an active online repository in the **National Spatial Data Clearinghouse program**. CUGIR provides geospatial data and metadata for New York State, with special emphasis on those natural features relevant to agriculture, ecology, natural resources, and human-environment interactions. It is only available in English and is not published as Linked Data.

Research Centre for Agrobiolology and Pedology (CRA-ABP)

The infrastructure of the Italian Research Council (CRA) is curated by CRA Informative Systems Service, which maintains three datasets about soil: **Soilmaps**, which covers the national territory; the **Soil Information System of Italy (SISI)**, a WebGIS cloud service for **soil profiles**; and ARCAN, a database of **soil samples**. According to a

2013 policy document, collections are shared according to Italian Open Data License (IODL) v2.0 and the principle of "open by default", i.e., freely available to all, free of copyright or other forms of control which prevent the reproduction, and free of restrictions beside the obligation to cite the source. The data is published on the Web in an open format processable by generic tools.

International Maize and Wheat Improvement Center (CIMMYT)

CIMMYT provides access to a **repository for publications, images, and video**; Maize Finder, Maize Inbred Lines, the International Maize Trials Network, GRIN (Germplasm Resources Information Network) Global Maize; International Wheat Improvement Network and GRIN Global Wheat; CIMMYT evaluation of lines; and the recently launched CIMMYT Research Data Repository. All of the datasets are available for the public, with licensing arrangements possible, though are not published as Linked Data.

Brazilian Agricultural Research Corporation (Embrapa)

Embrapa's **databases of raw scientific data** are generally available only to their maintainers, even internally at Embrapa -- a crucial cultural issue that has a direct impact on Embrapa's adherence to the Open Data paradigm. A proposal on information governance, covering this issue, is currently under discussion. Key datasets include: Ideare, a Web-based projects database for use by Embrapa employees and project partners; Agricultural Research Database (BDP@) of **library collections**, which in turn includes Alice, a repository of Embrapa **research outputs**, Infoteca-e, a collection of materials on Embrapa **technologies for a wider audience of farmers and extension workers**, and the Open and Integrated Information System on Agriculture (Sabiia), which aggregates metadata from scientific data open access providers. All systems are in Portuguese but include English-language content. BDP@, Alice, and Infoteca-e publish open data though without specific policies. No semantic standards are used in the datasets.

Food and Agriculture Organization of the UN (FAO)

FAO datasets span many domains, from production and commodity time-series for **crops, livestock, forestry and fisheries to irrigation, and soil maps**. Indeed, with its vast and diverse holdings, data integration at FAO remains a challenge. Various integration initiatives exist, among the most long-standing is AGROVOC with its 30K+ concepts represented as a SKOS concept scheme. Part of the GACS concept-integration initiative, it is in wide-use across a number of organizations and is actively maintained in 22 languages by a distributed global team of experts.

Indian Statistical Institute (ISI)

The Documentation Research and Training Center (DRTC) hosts and maintains Librarians' Digital Library, Indus Asian Agricultural Resources gateway, and CALIS, a current **awareness service based on RSS feeds in library and information science**.

Integrated Modelling Collaboratory (IMC)

IMC does not authoritatively maintain its own datasets; rather, it **provides semantic annotation and distribution of global and local datasets for use** with their integratedmodelling.org platform. IMC maintains collaboration sites and have secure upload and semantic annotation facilities for users of the collaboratory.

Association for Technology and Structures in Agriculture (KTBL)

KTBL maintains a large collection of **planning data for agriculture**. It provides standard calculation values for investment, resource allocation and process planning for plant production as well as for livestock farming. Included are average purchase prices, useful lifespans of machinery and equipment, and average execution times of tasks such as harvesting and fertilization by crop or machinery used. Basic values for calculating nutrient balances and carbon footprint of production have recently been added. Descriptive labels and text fields are mostly in German, but parts have been translated into English. KTBL is preparing to answer requests for further languages and has started work on publishing the data as Linked Open Data. Data is currently published using the Creative Commons BYNC-SA 4.0 license, but KTBL is considering a switch to the more permissive BY-SA variant. The server provides data in RDF/XML, Turtle, JSON and XML format plus an HTML view for human consumption. It runs a Linked Data API and a SPARQL endpoint. The datasets use the RDF, RDFS and SKOS vocabularies along with QUDT for representation of units and dimensions.

Land Portal Foundation

Land library data: while there is technically a dataset available for all the resources, it is currently only accessible to people with strong technical expertise. For the **land book**, “original” (source) datasets are available. Both library and book datasets are made available using state of the art LOD technology. The Land Portal maintains a SPARQL endpoint allowing anyone to query all this data. A dataset on **Organisation and People** is not yet 100% settled. LP wants to publish a “public book” of organizations and people interested to share their profile with the Land community. The Land Portal section on “Policies and copyrights” really needs to be updated but information on Open data and copyrights related matters should be available from the About section on the website. In general the LP ought to (re)publish all information that goes through it as Open and Linked Data.

Syngenta

Syngenta datasets maintained in-house as part of the R&D innovation pipeline include: **small molecule bioactivity data for crop protection, genomic sequence and phenotype data**, regulatory study information for **crop protection** and **seed breeding** development, and environmental profile data such as **soil profiles, climate and weather**. In addition, Syngenta interacts with external datasets such as those maintained by the European Bioinformatics Institutes and similar public institutions. These datasets are serialized in RDBMS and RDF. The Syngenta Good Growth Plan data is published on the Web in both CSV and RDF using **various Creative Commons** licenses such as NonCommercial-NoDerivatives and Attribution-ShareAlike, an open data license. Each year, Syngenta reports progress on its “six commitments” and provides detailed data definitions in accordance with best practice.

USDA Agricultural Research Service (ARS)

USDA ARS makes available a wide range of datasets online, some of which in the context of three initiatives: the **GRACEnet Agricultural Research Service**, the **Agriculture Model Intercomparison and Improvement Program** (AgMIP), and **Decision Support System for Agrotechnology Transfer** (DSSAT). None of these are published as open data, and given the complexity of the datasets, open data is seen as a low priority compared to improving data organization and data acquisition. Data from ARS is US Federal data and has no protections except

for proprietary rights for a reasonable period prior to publication and in cases where security or specifically recognized commercial interests are a concern.

National Agricultural Library (NAL)

The National Agricultural Library publishes many datasets. AGRICOLA (AGRICultural OnLine Access) is a **bibliographic database** that serves as the catalog and index to NAL's modern and historical collections. PubAg is a portal to USDA-authored and other highly relevant agricultural research, with both **full-text articles and citations** to peer-reviewed journal articles. Both are **indexed using the NAL Thesaurus** and available under the CC0 license. In addition, NAL makes available Life qCycle Assessment Commons, which provides open access to **life cycle assessment (LCA) data sets and tools**. The Insect 5,000 Genome (i5K) dataset is provided by an international project to sequence the genomes of insects of importance to agriculture and human health. A Long-Term Agro-Ecosystem Research (LTAR) project will make data available through NAL in October 2015, starting with **near-real-time meteorological data**. The Ag Data Commons links existing repositories while providing a platform for datasets that require specialized metadata, visualizations or analytics, or do not otherwise have a home for long-term access, stewardship and preservation. The National Nutrient Database for Standard Reference holds **datasets on food, nutrients, weights and measures**, and sources of data useful to researchers, citizens, and agribusiness.

Wageningen UR Library

The Library maintains the **research information system for the Wageningen University** and Research Centre, conducts **bibliometric analyses**, and maintains a database of published datasets and of databases for Dutch **vocational training**. The Library emphasizes research data management planning and provides **courses on the subject for PhD** researchers.

Wageningen Alterra institute

The Alterra institute manages datasets and websites targeted at data visualization: **Land Use Netherlands**; **Object Height Netherlands**; a dataset of processed **satellite images** with some key indicators; **Soil Information** for the Netherlands); www.yieldgap.org, maintained together with partners on **yield gaps across Europe**; a geo-enabled **clearinghouse for forestry data**, with partners from the Trees4Future project with **some linked data components**; and the **Historic Weather Viewer**. The Agriculture Model Intercomparison and Improvement Program (AgMIP) has relevant datasets for which it is building a publication infrastructure, notably for **inputs and outputs of crop models**, with links to AgTrials (CGIAR) and USDA.

Alterra and Wageningen UR are involved in the establishment of the **Open Data Journal for Agricultural Research** as a new and peer reviewed publication avenue for data sets from agricultural research.

3. Vocabularies maintained

Agricultural Sustainability Institute (ASI)

- Uses vocabularies: AGROVOC, Library of Congress Subject Headings, MARC 21 Code List for Geographic Areas.
- Used our controlled vocabularies internally primarily to help analyze our database linking environmental issues with indicators, using Marxan, Matlab and R to do the analysis.
- Vocabulary maintenance: Google Spreadsheets, a Django-fronted web database, Protege, and a wiki developed using the Semantic MediaWiki platform.

Aston

- As a result of the FIspace project, Aston has formalised the GS1 EPCIS standard in the form of two ontologies:
 - The **EPCIS Event Model**
 - The **Core Business Vocabulary**
- A mapping is available between the the W3C PROV namespace and the EPCIS Event Model.
- The EPCIS Event Model and Core Business Vocabulary are intended to be used for the agri-food supply chain.
- There is a functioning Java Library for the generation of RDF triples from RFID/Barcode events.
- Currently there has been no commercial uptake.

Bayer

- **Internal vocabularies** only, all in English. Some linked to external ones (e.g. taxonomy of species), others are verbatim copies of external vocabularies (e.g. Gene Ontology)
- Contributes to expansion of public external vocabularies (e.g. trait ontology).
- Applications: range from pure Research tools (e.g. functional characterization of genes and proteins) to tools in the Development phase (e.g. plant breeding activities).
- Users: data scientists and wet scientists
- Vocabulary maintenance: custom pipelines, OBO edit, Protégé, Excel, among others.

CABI

- **CAB Thesaurus** contains more than 250,000 concepts, almost complete in four languages (English, Spanish, Portuguese and Dutch) and lesser content in seven other European languages.
- Part of **Global Agricultural Concept Scheme** (GACS), links to AGROVOC and NALT.
- LOD is in trial status but full scale LOD APIs anticipated by end of 2015.
- The URIs are live RDFa versions under a restrictive CC license, which will be opened up further by the end of 2015.
- R4D is used in numerous applications including DevTracker. CABI data is not currently consumed directly by other applications; rather they are fed to vendors and other organizations via FTP. This is changing in 2016. The majority of this is subscription-based and not open. These models will also change during the coming 12-18 months.
- Vocabulary maintenance: MultiTees is used for thesaurus maintenance, for web deployment, and for distributing files in multiple formats, including SKOS/RDF, to customers and online hosts for our bibliographic databases.

Cornell Library

- The Cornell Library does not maintain any specifically agricultural vocabularies. However, the **VIVO Semantic Web application includes the VIVO Ontology**, which is primarily used for representing **person, department, and publication information across multiple domains**. The application also

makes use of many other ontologies such as FOAF, SKOS, BIBO, the UN Geopolitical Ontology, and many more.

- Cornell is attending the work workshop more as a dataset consumer than a dataset provider. John Ferreira developed a portion of the External Concept Service in the VIVO application. This provides an integration layer for external datasets such as AGROVOC, Mesh, and Gemet and a lookup and assignment of concepts from these vocabularies for indicating research areas for a users profile.
- John also created a wrapper service that uses the same underlying code used by the **AgroTagger** for autotagging approximately 480 thousand research articles in the TEEAL project and have just started using it for mapping freetext expertise and area of interest terms to Agrovoc concepts.
- John is using VIVO to maintain and edit a version of the ONLD (**Organization Name Linked Data**) dataset for disambiguating organization names provided from multiple data sources providing user profile data.

CGIAR

- Contributes to maintenance of **crop ontology** and others in the domain of **genetic plant resources, breeding, agronomy**.
- Vocabularies mapped to Plant Ontology, Trait Ontology and Biodiversity plant descriptors. Will be mapped to ICASA variables of AgMIP project and to the Plant Environment Ontology
- **Agronomy ontology**, MultiCRopPassport Data (mapped to DARWIN CORE germplasm extension of GBIF), **Plant descriptors for genebanks**, collecting mission forms, household surveys on biodiversity.
- Users of crop descriptors: national genebanks, botanists.
- **Multi-Crop Passport Data** : used by all ex situ genebanks, databases, EURISCO, Genesys.
- Crop Ontology used by Breeding Management System, Next Generation Cassava Database, Global Agricultural Evaluation Trials Database, European Solanaceae breeding database, Australian Phenotyping Facility, Vitis Ontology (INRA), Soybean Ontology (USDA), POLAPGEN (barley), Triticieae-global (oats).
- Vocabulary maintenance at Crop Ontology based on google apps, uses Excel template called Trait Dictionary, conversion routines from Excel to OBO.
- Mappings done manually but looking to automatic mapping tools.

Research Centre for Agrobiolgy and Pedology (CRA-ABP)

- **AgINFRA Soil Vocabulary** English tool to represent data vocabulary.
- AGROVOC and VocBench.
- Tool to represent KOSs and Thesaurus – English-Italian.
- Repository AgINFRA – AllegroGraph WebView 4.11, English-Italian.
- External vocabularies used: USDA soil taxonomy and World Reference Base.
- Use of datasets and vocabularies: looking for implementation in AGRIS.
- Users: researchers and professional data users.
- Vocabulary maintenance: D2RQ, Neologism – data vocabularies, VocBench, AllegroGraph WebView 4.11.

International Maize and Wheat Improvement Center (CIMMYT)

- Contributes to expansion of crop ontology (in English, but with high demand for translation to Spanish, French and Chinese).
- Specifically, the **Maize trait ontology**, **wheat trait ontology**, and **crop research ontology** developed by CIMMYT.
- Vocabularies mapped to Plant Ontology and Plant Trait Ontology when possible.
- Breeding management system (BMS) uses the crop trait ontology and crop research ontology.
- Uses of vocabularies in near future: in FieldBooks and handhelds (KSU FieldBook and KDSmart, KDSmart v2 will be for use on Android phones).
- Users: Breeders, scientists and assistants, data managers and developers.

Embrapa

- Vocabularies used: **AGROVOC**, **NAL Thesaurus**, **Thesagro**, with links to each resource.
- Proposal for knowledge organization and engineering Embrapa, which includes use of automation to build vocabulary. Intend to make it available either to process actors or to users through web service. Using SKOS.
- Need for Brazilian Portuguese differing from European Portuguese, but also English, Spanish, French and Italian.
- Thesaurus maintenance: e-Termos (free access), and TheXML (proprietary).
- Applications for this new vocabulary not yet developed, but projects of interest:
 - re-engineering Thesagro with refined semantics,
 - OntoAgroHidro which is related to the interfaces of environmental and socioeconomic impacts of agriculture and climate changes on hydric resources,
 - Mapping between INRA And Embrapa vocabularies in agroecology to enable mutual visibility of open data in both institutions.
 - Creation of thematic arrangement of Embrapa RD&I projects based on terminological similarity indexes.

Food and Agriculture Organization of the UN (FAO)

- FAO OPCCD maintains **AGROVOC**, a multilingual agricultural concept scheme.
- Other vocabularies maintained at FAO include:
 - **GeoPolitical Ontology**, designed to be a master reference for geopolitical information that manages names in multiple languages (English, French, Spanish, Arabic, Chinese, Russian and Italian) including FAO official country names; maps standard coding systems (UN M49, ISO 3166, FAOSTAT, AGROVOC, etc); provides relations among territories (land borders, group membership, etc); and tracks historical changes.
 - **BioTechnology Glossary**, another SKOS-based concept scheme managed by the Department of Natural Resources.
 - The Fisheries and Aquaculture department maintains an ontology known as **FLOD**, connecting Marine Species, Water Areas, Land Areas and Exclusive Economic Zone classifications.
- FAO also maintains a number of other important and widely used classifications systems that are not currently semantically enabled, such as the FAO Commodity List and the Global Administrative Unit Layers both managed by the Statistics Division -- an excellent area for possible future work.
- Tools used to maintain and disseminate these vocabularies include VocBench, a web-based ontology maintenance platform now managed by the University of Tor Vergata as well as Allegrograph, Fuseki, and SKOSMOS.

Land Portal Foundation

- Vocabularies maintained include a **Land Portal taxonomy of topics**; and **regions and countries**.
- There may be other vocabularies related to documents and resources, such as type of material.
- The “new” Landportal has been launched 9 months ago with the landbook section.
- The public face of the Landportal (<http://landportal.info/book>) is build entirely using the LOD available from LP endpoint.
- The latest section, Land Library, has just been published.
- The goal is to have the LP regional hubs being the primary customers (as well as providers).
- As of today we do not have steady data customers nor can distinguish user groups.
- The vocabulary maintained essentially with a Drupal frontend, using standard modules (feeds import, mapping, filtering) to perform the import. They are then stored in a Linked data endpoint (Virtuoso).

- Some part (land book) use some more manual scripts and processes to gather, mashup and make re-available those information in a LOD form.

Indian Statistical Institute (ISI)

- Helps institutes and research organizations manage their vocabularies using a semantic framework built based on principles and postulates of LIS.
- Users: Researches of scholarly articles.
- Vocabulary maintenance software is internal, homegrown.

Integrated Modelling Collaboratory (IMC)

- Developed **k.IM language** which is the only descriptive framework available for **semantic annotation of data and models**. The language has facilities for integration of authorities, which bridge to existing vocabularies such as AGROVOC, GBIF encoders and IUPAC chemical identities so that they can be used within semantically explicit workflows.
- Uses of datasets and vocabularies: The ARIS platform is one of the major platforms for ecosystem services assessment and valuation.
- Vocabulary maintenance: An open source software stack (named k.LAB) is the main activity of the IMC and contains both a user-side platform (language, IDE and modeling engine) and a network node package that allows users to install semantic web nodes using the infrastructure and ontologies. Internally, our approach is based on OWL2 and DL reasoning.

Association for Technology and Structures in Agriculture (KTBL)

- agroRDF vocabulary, was previous available at iGreen project but is currently unavailable.
- There is a SPARQL endpoint for the vocabulary definition and **agricultural machine taxonomy**.
- Where possible, the vocabulary and taxonomy have been **mapped to AGROVOC concepts**.
- Users: two farm management system provides in Germany uses web services provided.
- Developing **smartphone application for the Android platform**. It is using the JSON serialization of the service and offers possibilities to do basic depreciation, fixed and variable cost and operating supply consumption calculations for agricultural machines.
- Vocabulary maintenance: text editors and standard source code versioning tools, d2rq for generating an RDF dump out of an Oracle database. This is then fed alongside the vocabularies into a Jena Fuseki instance to provide a SPARQL endpoint. When we reference an external vocabulary, that is imported either fully or as a subset into the Fuseki instance as well. Mappings are created manually.

Syngenta

- Maintains a vocabulary system, which hosts both **public and internal reference terms**.
- **Uses many public vocabularies**: FAO, obofoundry, bioportal, CGIAR, SWEET, QUDT, Semantic Sensor Network Ontology, SKOS, Organization Ontology, etc.
- Users: Crop Protection Research Scientists, data analysts, plant breeders, toxicologists, R&D researchers.
- Uses in-house and OTS tools for data capture and analysis.
- Environmental profile data is integrated with phenotype data.
- Vocabulary maintenance: uses in-house application. Soon to use Top Quadrant EVN, which will have links to Unichem and Identifiers.org.

USDA Agricultural Research Service (ARS)

- **ICASA dictionary is used by AgMIP and DSSAT**. Some preliminary mapping to Crop ontology.
- Datasets used in crop simulation modeling and general meta-analyses. AgMIP is a consumer of data for climate change scenario studies.
- Vocabulary maintenance: Google spreadsheets, SAS.

National Agricultural Library (NAL)

- **NAL Agricultural Thesaurus** contains more than 110,000 concepts, English-Spanish parallel versions, downloadable in SKOS, XML, PDF and MARC formats, CC0 license.
- Part of **Global Agriculture Concept Scheme (GACS)**, links to AGROVOC and CAB Thesaurus.
- Vocabulary maintenance: MultiTes for thesaurus maintenance and Web Development Kit for web delivery. In-house application enables the delivery of RDF for each skos:Concept URI.
- NALT used as controlled vocabulary for bibliographic databases (AGRICOLA, PubAg), and other NAL information centers and products. Global Change Master Directory used for LTAR.
- Users: Researcher, Policy Maker, Small Producer, Information Professional, Agribusiness Executive and Citizen (including educators, students).
- Applications: Fedora, Open LCA Framework, Drupal, Tripal, Chado, Web Apollo, DKAN.
- Apple and Android devices: Dietary Reference Intake calculator app.

Wageningen UR Library

- Previously maintained Dutch version of CAB thesaurus, but may not continue.
- Maintains a subject category tree.
- Users: research peers use archived datasets for verification and re-use.
- Using a locally developed thesaurus maintenance tool which will likely be replaced by Worldcat.

Wageningen UR Alterra

- Alterra contributes to **Seamless Ontology**, available from the Trees4Future project, some linked to ontologies and on the meta-data level. Managing **data structures on nature and biodiversity**, but not yet fully developed as vocabularies. AgMIP: ICASA dictionary.
- Using vocabularies to link data sets to models, or make data sets more discoverable online. Using some vocabularies to discover more data and information on a meta-data level for websites for the general public or practitioners in a relevant domain.
- Alterra datasets used for data visualization, data analysis, or collecting more data in crowdsourcing applications. Including smart phone apps as part of the products.
- Users: researchers and developers.
- Vocabulary maintenance: Protégé and other tools, no formal processes of updating.

Appendix A: Links

Aston

- EPCIS Event Model: <http://fispace.aston.ac.uk/ontologies/eem.html#>
- Core Business Vocabulary: <http://fispace.aston.ac.uk/ontologies/cbv.html#>
- Mapping of EPCIS Event Model to W3C Prov: http://fispace.aston.ac.uk/ontologies/eem_prov.html#
- W3C Prov namespace: <http://www.w3.org/ns/prov#>
- Paper about EPCIS:
<http://windermere.aston.ac.uk/~solankm2/papers/Enhancing-Visibility-in-EPCIS-Governing-Agri-Food-Supply-Chains-via-Linked-Pedigrees.pdf>

Bayer

- F.A.I.R. approach: <https://www.force11.org/node/6062>

CABI

- R4D portal: <http://r4d.dfid.gov.uk/>
- CAB Thesaurus: <http://www.cabi.org/cabthesaurus/>
- Bibliographic index: <http://www.cabdirect.org/>
- Linked Open Data trial: <http://id.cabi.org/cabt/page/54380>
- MultiTes: <http://www.multites.com/>

CGIAR

- Crop Ontology: <http://www.croponontology.org/>
- ISA-Tools: <http://www.isa-tools.org/>
- Ontomaton tools: <https://github.com/ISA-tools/OntoMaton>
- Ontology for Biomedical Investigations (OBI) project: <http://www.ob-ontology.org>
- Browsing OBI: <http://www.ontobee.org/browser/index.php?o=OBI>
- STATO - most commonly used statistical methods for experimental design: <http://www.stato-ontology.org>
- Biosharing: <https://www.biosharing.org/pages/about/>
- ContentMine project: <http://contentmine.org>

CRA-ABP

- CRA Informative Systems Service: <http://sito.entecra.it/portale/public/documenti/PSI/cra-directory.ods>
- Soilmaps: http://soilmaps.entecra.it/ita/ric_av.php
- Soil Profiles: <http://aginfra-sg.ct.infn.it/webgis/cncp/public/>
- Soil Samples: <http://93.63.35.107:8080/geoexplorer/composer/#maps/1>
- Tool to represent data vocabulary: <http://vocabularies.aginfra.eu/soil#>
- VocBench: <http://artemide.art.uniroma2.it/vocbench2/#Concepts>
- VocBench: <http://202.73.13.50:55481/vocbench/#Concepts>
- Tool to represent data in RDF formats: <http://rdf.entecra.it/soilmaps/>
- AllegroGraph WebView: <http://202.45.139.84:10035/catalogs/fao/repositories/agINFRA>
- USDA Soil Taxonomy, 10th Edition (2006): http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052172.pdf
- World Reference Base, 2nd edition (2006): <ftp://ftp.fao.org/agl/agll/docs/wsr103e.pdf>
- Smartphone application concept: <http://sito.entecra.it/portale/public/documenti/PSI/cra-directory.ods>

CIMMYT

- CIMMYT resources: <http://www.cimmyt.org/en/resources>
- CIMMYT Research Data Repository: <http://data.cimmyt.org/dvn/>
- Crop Ontology: <http://www.croponontology.org/>
- Planteome project: <http://planteome.org/node/9>
- Semantic GIS project: <http://www.semanticgis.net/>

Embrapa

- Embrapa: <http://www.embrapa.br>

- Ideare (projects database): <https://sistemas.sede.embrapa.br/ideare/>
- BDP@, Agricultural Research Database: <http://www.bdpa.cnptia.embrapa.br/consulta/>
- Alice, open access to Embrapa scientific information: <http://www.alice.cnptia.embrapa.br/>
- infoteca-e, Technological Information on Agriculture: <http://www.infoteca.cnptia.embrapa.br>
- Sabiia, Open and integrated information system on agriculture: <http://www.sabiia.cnptia.embrapa.br/sabiia/?initQuery=t>
- e-Termos: <https://www.etermos.cnptia.embrapa.br/index.php>
- TheXML: <http://www.viaapia.com.br/index.php/thexml>

FAO

- AGROVOC: <http://aims.fao.org/vesit-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>
- Biotechnology glossary: <http://www.fao.org/biotech/biotech-glossary/en/>
- Geopolitical Ontology: <http://www.fao.org/countryprofiles/geoinfo/en/>
- FLOD: <http://www.fao.org/fishery/topic/18046/en>

INRA

- GnpIS, plant and fungi data integration: <https://urgi.versailles.inra.fr/Tools/GnpIS>
- GenoToul: <http://bioinfo.genotoul.fr/>
- Oqali: http://www.oqali.fr/oqali_eng/
- Agroportal: <http://agroportal.lirmm.fr/>
- ANAEE Thesaurus: <http://agroportal.lirmm.fr/ontologies/ANAEEF>
- ATOL: <http://www.atol-ontology.com/index.php/fr/>
- BIOREFINERY: <http://ist.blogs.inra.fr/lovinra/2015/05/04/biorefinery/>
- BIOREFINERY: <http://agroportal.lirmm.fr/ontologies/BIOREFINERY>
- TRANSMAT: <http://ist.blogs.inra.fr/lovinra/2015/05/04/transmat/>

IFPRI

- Harvard Dataverse network: <https://dataverse.harvard.edu/dataverse/IFPRI>
- data.ifpri.org: <http://data.ifpri.org/>
- IFPRI copyright statement: <http://www.ifpri.org/copyright>
- AGRODEP: <http://www.agrodep.org/country/NGA>

ISI

- Librarians' Digital Library (LDL): <https://drtc.isibang.ac.in>
- Indus Asian Agricultural Resources gateway: <http://drtc.isibang.ac.in/indus>

IMC

- ARIES: <http://www.ariesonline.org>

KALRO

- KAINet and KALRO E-repository KALRO: <http://www.kalro.org:8080/repository>
- KAINet e-repository: <http://www.kainet.or.ke>
- OPENDOAR: <http://www.opendoar.org>
- OCSDNet: <http://ocsdnet.org/>

KTBL

- Subset of data about agricultural machinery: <http://srv.ktbl.de/data/MachineClass/>
- KTBL semantic search: <https://search.ktbl.de>
- Instance of Linked Data API: <https://github.com/UKGovLD/linked-data-api>
- SPARQL endpoint: <http://srv.ktbl.de/query>
- QUDT: <http://qudt.org>
- iGreen project: <http://data.igreen-services.com>
- KTBL vocabulary: <http://srv.ktbl.de/vocabulary#>
- KTBL machine taxonomy: <http://srv.ktbl.de/taxonomy/>
- d2rq: <http://d2rq.org>
- ELDA (Epimorphics Linked Data API): <http://github.com/epimorphics/elda>

Land Portal Foundation

- Land library: <http://landportal.info/library/>
- Land library data: <http://landportal.info/library/search>
- Land book catalog: <http://landportal.info/book/catalog>
- Land book dataset: <http://landportal.info/data/dataset>
- "Public book" (in development): <http://landportal.info/library/sources>
- "Public book" (in development): <http://landportal.info/debate/community>
- "Public book" (in development): <http://landportal.info/book/sources>
- Land Portal policies: <http://landportal.info/about/privacy-policy>
- "New" Landportal: <http://landportal.info/book>

Syngenta

- Progress on "six commitments": <http://www.syngenta.com/global/corporate/en/goodgrowthplan/progress/Pages/progress.aspx>
- The Good Growth Plan Progress Data:
<http://www.syngenta.com/global/corporate/en/GOODGROWTHPLANDATA/Pages/progress.aspx>
- Top Quadrant EVN: <http://www.topquadrant.com/products/topbraid-enterprise-vocabulary-net/>
- Unichem: <https://www.ebi.ac.uk/unichem/>
- identifiers.org: <http://identifiers.org/>

UCDavis ASI

- Checklist generator: <https://github.com/adhollander/checklist>
- Sustainable sourcing wiki: http://asi.ice.ucdavis.edu/sustsource/wiki/index.php/Main_Page
- Sustainable sourcing - OWL: <http://asi.ice.ucdavis.edu/sustsource/schemas/sustsource.owl>
- Sustainable sourcing - RDF: <http://asi.ice.ucdavis.edu/sustsource/schemas/sustsourceindiv.rdf>

USDA ARS

- GRACEnet: http://www.ars.usda.gov/research/programs/programs.htm?np_code=212&docid=21223
- AgMIP: <http://agmip.org>
- DSSAT: <http://dssat.net/data/exchange>
- ICASA dictionary:
<https://docs.google.com/spreadsheets/d/1MYx1ukUsCAM1pcixbVQSu49NU-LfXg-Dft-ncLbZGAM/pub?output=html>

USDA NAL

- AGRICOLA: <http://agricola.nal.usda.gov>
- PubAg: <http://pubag.nal.usda.gov>
- Life Cycle Assessment Commons: <http://www.lcacommons.gov>
- Insect 5,000 Genome "i5K": <https://i5k.nal.usda.gov/>
- Ag Data Commons: <http://data.nal.usda.gov>
- National Nutrient Database for Standard Reference: <http://ndb.nal.usda.gov>
- NAL Agricultural Thesaurus: <http://agclass.nal.usda.gov>
- GACS demo: <http://tester-os-kktest.lib.helsinki.fi/gacsdemo/en/>

Wageningen UR Library

- Research Information System: <http://library.wur.nl/WebQuery/wurpubs/show>
- Database of published datasets:
http://library.wur.nl/WebQuery/wurpubs?A240=OK&wq_inf_pre=pluspre&wq_rel=AND&wurpublikatie/publikatietype/hoofdtype==Dataset
- Databases for Dutch vocational training: <http://www.groenkennisnet.nl/nl/groenkennisnet.htm>
- Research data management planning: <http://www.Wageningenur.nl/en/Expertise-Services/Data-Management-Support-Hub.htm>
- Research data management planning - PhD course: <http://datamanagementplancourse.pbworks.com/>
- Subject categories:
<http://library.wur.nl/WebQuery/catbrowser/journal?recordtype=seriewerk%20OR%20monografie%20OR%20deel%20OR%20koepel>
- Data mining project - example:
<https://www.esciencecenter.nl/project/prediction-of-candidate-genes-for-traits-using-interoperable-genome-annotat>
- Data mining project - example:
<https://www.esciencecenter.nl/project/creation-of-food-specific-ontologies-for-food-focused-text-mining>

- ICASA variables: <http://research.agmip.org/display/dev/ICASA+Master+Variable+List>

Wageningen Alterra

- Processed satellite images of Netherlands: <http://www.groenmonitor.nl>
- Yield gaps in Europe: <http://www.yieldgap.org>