agINFRA – Building e-infrastructure and services for the agricultural research community

Guntram Geser[1], Johannes Keizer[2], Andrea Mulrenin[1]

[1]Innovation Lab
Salzburg Research
Jakob-Haringer-Str. 5/3
5020 Salzburg, Austria
guntram.geser@salzburgresearch.at

[2]Office of Knowledge Exchange
Food and Agriculture Organization of the United Nations
Viale delle Terme di Caracalla
00153Rome, Italy
Johannes.Keizer@fao.org

Abstract: The agINFRA project aims to provide the agricultural research community with e-infrastructure and services for open data access, sharing and re-use. This paper introduces the project's objectives and context, technical elements as well as the envisaged range of data resources. Furthermore opportunities for participation are highlighted.

1 Towards global sharing of agricultural research results

agINFRA[1] is an e-infrastructure project funded under the European Union's 7th Framework Programme. It started in October 2011 and will run for three years. agINFRA develops infrastructure and services for sharing agricultural research results that are managed by international, national, institutional as well as subject-based repositories.

The project involves technology and content partners from Europe, China, Ecuador and India. One of the lead partners is the Food and Agriculture Organization (FAO) of the United Nations, and the project goals are aligned with the Coherence in Information for Agricultural Research for Development (CIARD) initiative.[2] The initiative mobilizes and supports institutions in making agricultural research results more accessible globally.

Many agricultural research organizations already have content repositories and portals aimed to serve researchers, educators and extension workers. These institutional as well

---

[1] agINFRA, http://www.aginfra.eu
[2] CIARD, http://www.ciard.net

as subject-based repositories are increasingly integrated in information networks, mainly through deploying the Open Archives Initiative harvesting approach (OAI-PMH).[1]

The bulk of available content are publications while research data are often not archived and shared through Open Access repositories. Large surveys have shown that across all disciplines only about 6-8% of the researchers deposit datasets in an open archive. (Pa09; Sc11) Thus while the current emphasis of Open Access is on research publications the next frontier will be Open Data policies at least for publicly funded research. Yet it must be noted that also repositories of publications are often difficult to establish, promote and maintain. Integration in a larger pool of resources can allow for wider recognition, access and enhancement through added value services.

2 Goals of e-research infrastructures

Research is becoming increasingly distributed, collaborative, ICT and information-intensive. New terms such as „e-research", „e-science" or „cyberinfrastructure" emphasise the importance of digital tools and services for more efficient workflows and new ways of conducting research. As Hey and Hey (HH06) note, "e-Science is shorthand for the set of tools and technologies required to support collaborative, networked science. The entire e-Science infrastructure is intended to empower scientists to do their research in faster, better and different ways."

These goals have been taken up by major research policy initiatives such as the European Strategy Forum on Research Infrastructures (ESFRI) and the US National Science Foundation Cyberinfrastructure Panel. The European High-level Expert Group on Scientific Data understands that scientific data infrastructure "must be flexible but reliable, secure yet open, local and global, affordable yet high-performance" (Eu10).

Particularly important from the perspective of the providers and users are principles of collaboration and conditions for sharing of different resources in the networked research environment.

3 Elements of the agINFRA infrastructure and services

agINFRA aims to support the management, sharing and integration of content, services, software and computing resources in various ways. The following are some supportive elements agINFRA is developing.

An enhanced version of the CIARD - Routemap to Information Nodes and Gateways (RING) will act as central broker for data and service providers, datasets and vocabulaies. The CIARD-RING portal has been developed by the Global Forum on Agricultural Research (GFAR) and already comprises 614 services / sources registered by 384 pro-

---

[1] Open Archives Initiative, http://www.openarchives.org

viders.[1]

Linked Data methods will allow for leveraging the semantic interoperability of heterogeneous data. (BHB09) For example, FAO's Linked Data based OpenAgris (currently in beta stage) integrates many Web information resources in addition to the bibliographic data provided by the over 150 AGRIS information centres.

The e-infrastructure will comprise several software components that allow for enhancing repositories and information portals. This ranges from metadata extraction and indexing (e.g. through AgroTagger[2]) to semantic search and on to the visualization of research networks (e.g. based on VIVO[3]). For massive data aggregation and heavy infrastructure service tasks also dedicated Grid or Cloud services are in development.

4 Covering a broad range of content

At this stage, agINFRA is targeting the integration of five domains of information:

- *Bibliographic information on scientific and grey literature*, for example, FAO's AGRIS database (containing over 4 million bibliographic records) and other sources of research literature that are harvested by the Virtual Open Access Agriculture & Aquaculture Repository (VOA3R) project[4];
- *Geospatial information services* offering maps of land cover and soils, GIS datasets and other data with an agricultural or environmental theme, for example, the FAO GeoNetwork[5] and national resources such as the Italian Soil Information System (ISIS)[6];
- *Plant germplasm collections and genomic information*, for example, the Chinese Crop Germplasm Research Information System (CGRIS)[7] and other national and international collections (e.g. European National Inventories of germplasm as shared through the EURISCO data catalogue[8]) as well as DNA barcodes;
- *Agricultural statistics*, for example, FAOSTAT (over 3 million statistical entries, time-series data, etc.)[9], other United Nations databases, and the World Bank open data catalogue[10];
- *Learning and training resources*, for example, the Latin American Federation of Learning Object Repositories (LA FLOR)[11] and the Organic.Edunet[1] learning

---

[1] CIARD-RING portal, http://ring.ciard.net
[2] AgroTagger, http://aims.fao.org/agrotagger
[3] VIVO, http://vivoweb.org
[4] VOA3R, http://voa3r.cc.uah.es
[5] GeoNetwork, http://www.fao.org/geonetwork/
[6] ISIS, http://aginfra-sg.ct.infn.it/isis
[7] CGRIS, http://icgr.caas.net.cn/cgris_english.html
[8] EURISCO, http://eurisco.ecpgr.org
[9] FAOSTAT, http://faostat.fao.org
[10] World Bank open data, http://data.worldbank.org/data-catalog
[11] LA FLOR, http://laflor.laclo.org

ing resources for organic agriculture and agro-ecology.

Besides these (selected) resources of agINFRA partners and closely related organizations many others have been identified as potential candidates for integration in the data pool.

5 Opportunities for participation

agINFRA is designed as an open and collaborative initiative. Therefore it offers a number of ways for stakeholders to participate in the emerging infrastructure and service ecosystem. For example, agricultural research organizations can register information resources in the CIARD-RING to become part of a global pool of content and services; software developers can use agINFRA components, APIs or add-on plug-ins for enhancing existing tools and services.

Overall agINFRA aims at allowing agricultural research organizations open up their repositories and share the available content and data more effectively, in particular through applying Linked Data principles and methods. Moreover the project devotes particular attention to the semantics of shared data as well as criteria of reliability such as data provenance.

References

[BHB09] Bizer, C.; Heath, T.; Berners-Lee, T.: Linked data - the story so far. In: International Journal on Semantic Web and Information Systems 5.3, 2009, S. 1-22
[Eu10] European High-level Expert Group on Scientific Data Riding the wave. How Europe can gain from the rising tide of scientific data. A submission to the European Commission, October 2010
[HH06] Hey, T., Hey, J.: e-science and its implications for the library community. In: Library Hi Tech, vol. 24, no. 4, 2006, S. 515–528
[PA09] PARSE.Insight: Insight into digital preservation of research output in Europe. Project deliverable D3.4: Survey Report, December 2009
[Sc11] Science magazine: Special Issue "Dealing with Data". In: Science, vol. 331, no. 6018, 11 February 2011, S. 692-693

---

[1] Organic.Edunet, http://www.organic-edunet.eu