Setting up a Global Linked Data Catalog of Datasets for Agriculture

Valeria Pesce[1], Ajit Maru[1], Phil Archer[2], Thembani Malapela[3], Johannes Keizer[3]

[1] Global Forum on Agricultural Research, Rome, Italy
{valeria.pesce,ajit.maru}@fao.org
[2] ERCIM, Sophia-Antipolis, France
phila@w3.org
[3] Food and Agriculture Organization of the United Nations, Rome, Italy
{thembani.malapela,johannes.keizer}@fao.org

Abstract.

The movement to share data has been on the rise in the last decade and lately in the agricultural domain. Similarly platforms for publishing scientific and statistical datasets have sprouted and have improved visibility and availability of datasets. Yet there are still constraints in making datasets discoverable and re-usable. Commonly agreed semantics, authority lists to index datasets and standard formats and protocols to expose data are now essential. This paper explains how the CIARD RING provides a global linked data catalog of datasets for agriculture. The first part of this paper will describe the Linked Data layer of the CIARD RING focusing on the data model, semantics used and the CIARD RING LOD publication. The second part will provide examples of re-use of data from the RING. The paper concludes by describing the future steps in the development of the CIARD RING.

Keywords: Datasets · Data Catalogs · Directories · Linked Data · Interoperability · Semantic Web ·Vocabularies

1 Introduction

The need for better sharing and easier discovery of data has become more evident in the past few years with increasing calls and trends in open government data. In agriculture, the commitment was reinforced in 2013 by leaders at the G-8 International Conference on Open Data for Agriculture.[1] Repetition of research and difficulty in building upon other experts' findings in a timely manner hinders research uptake and innovation. This situation can be significantly improved if data and datasets used and produced in research are easily shared and found. "Sharing other products of research on the Web, including raw datasets and other re-usable results, is seen as essential for enabling innovation on important topics of agricultural research for development and

[1] https://sites.google.com/site/g8opendataconference/home

food security" [1]. This can only be achieved if the process of managing and sharing datasets is made easier and global registries of these datasets exist.

In addition, the need for integrated information systems in the agricultural domain is widely acknowledged (cf. [1, 2]). Integrated information systems should provide information gathered from as many relevant sources as possible and re-purposed for the specific needs of the prospected audiences. The main difficulty in building such integrated information systems is the little awareness of what information sources exist, how interoperable they are, how to tap into them and how to exploit their semantics. There is no comprehensive list or directory of agricultural information sources and technical details about these sources are often not documented and known only to the developers.

This is why the CIARD[2] movement set up the CIARD RING[3], managed by the Global Forum on Agricultural Research (GFAR).

The CIARD RING (henceforth shortened as the RING) is a global directory of web-based information services and datasets for agriculture such as search engines, databases, repositories, Open Archives, feeds, data sheets etc., associated with software tools that can process them. The services are described in details and categorized according to both content criteria such as thematic coverage, geographic coverage, content type, target audience; and technical criteria such as metadata sets adopted, vocabularies used, technologies used, protocols implemented. A new feature of the RING is the addition of a directory of dataset processing software tools and web services: datasets can be associated with software tools and APIs that can process them in different ways (convert, analyse, combine with other data etc.).

Our intent was that this information, besides being manually browsed by data and service managers, should be directly usable by the applications that needed it to build value integrated services on top of the data exposed by the datasets registered in the RING.

This paper will focus on how we used Linked Data technologies and semantics to make the RING a machine-readable hub / switchboard to datasets.

Our objectives in doing this were:

- Datasets registered in the catalog have to be found by applications
- Applications have to be able to read all the metadata about datasets and filter datasets according to their needs
- Applications have to find enough technical metadata in the catalog to:
  - Identify datasets with a specific coverage (type of data, thematic coverage, geographic coverage)
  - Identify datasets that comply with certain technical specifications (format, protocol etc.)
  - Access the dataset and get the data
  - Possibly identify APIs and software tools that can process the identified datasets

[2] CIARD is a global movement dedicated to open agricultural knowledge: http://www.ciard.info

[3] http://ring.ciard.info

To achieve this, we needed agreed semantics and authority lists to index datasets and standard formats and protocols to expose the data. This led us to the choice of creating an RDF store[4] using existing metadata vocabularies and Knowledge Organization Systems (KOS) and exposing all data using Linked Data[5] technologies.

This paper will initially give a brief overview of some related work that has already been carried out and explain why we think the RING fills a gap. The we will describe the Linked Data layer of the RING, focusing first on the data model and semantics used and then on the implementation of the Linked Data good practices.

1.1 Related Work

Recently, thanks to the open government and open data movements, dataset publishing platforms have become popular. Harvard University has made available the DataVerse[6] platform for publishing scientific and statistical datasets. Another popular publishing platform is CKAN[7], maintained by the Open Knowledge Foundation, which also provides a global dataset hub called the Datahub[8].

Some important agriculture-related datasets have been published using similar platforms. Government agricultural datasets are available on *data.gov* public platforms for some developed countries (US[9], UK[10], some statistics from European countries) and BRICS countries (India in particular has started a data.gov project that includes agricultural data; Brazil has an open data portal). Very little within the agricultural domain is available from developing countries (Kenya has started an open data portal including data on agriculture, while for Africa there is the Open Data for Africa portal[11]). Some agricultural research centers (IFPRI, Bioversity International, ICRAF) have started publishing their datasets on their own DataVerse instance and sharing them through the DataVerse Network.

At the regional and global level, OpenAIRE[12] and the European Union Open Data Portal[13] include agricultural datasets from Europe; the World Bank has been publish-

---

4 The Resource Description Framework (RDF, http://www.w3.org/standards/semanticweb/) is a family of specifications that has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources (http://en.wikipedia.org/wiki/Resource_Description_Framework). An RDF store is a way of storing data using a machine-readable "grammar" (RDF) and documented semantics (RDF vocabularies).

5 Linked Data is a “recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.” (Wikipedia). See http://www.w3.org/DesignIssues/LinkedData

6 http://dataverse.org/

7 http://ckan.org/

8 http://datahub.io/

9 http://catalog.data.gov/dataset?groups=agriculture8571

10 http://data.gov.uk/data/search?q=&publisher=department-for-environment-food-and-rural-affairs

11 http://opendataforafrica.org/

12 https://www.openaire.eu/

ing datasets for a while; the Food and Agriculture Organization of the United Nations (FAO) started working on *data.fao.org* a few years ago and some interesting general dataset catalogs and / or repositories that include agricultural data exist, like DataCite[14] (using re3data[15] to search repositories) and Dryad[16], a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable.

The current situation seems to be that datasets for agriculture are gradually being made available (especially from developed countries) but are not easily discovered and not easily accessible (remotely searchable, re-usable). The existing platforms and catalogs have of course improved the situation and help in finding relevant datasets for agriculture. However, there are still tough challenges in making datasets really discoverable and re-usable.

1.2 Challenges

An overview of the existing platforms showed that there were still gaps in the provided solutions in terms of general interoperability and our specific thematic interest.

- None of the existing catalogs and repositories has a coverage that is at once global and specific to agriculture; agricultural datasets can be identified in some catalogs using keywords, but with no further thematic specialization.
- Each platform uses different categorizations for datasets and metadata are usually not detailed enough to allow for federated searches or selective harvesting from these systems. Overall, the existing platforms do not seem to have very rich metadata or to follow common standards for describing dataset nor common authority reference data.
- No platform exposes machine-readable metadata about semantic and technical aspects of the datasets (dimensions / vocabularies, reference authority data, formats, protocols…), making it difficult for applications to automatically re-use the data.

Regarding the second point, many dataset publishing platforms have their own data model and their own metadata vocabulary (Dataverse [3], OpenAIRE (Datacite) [4], re3data [5], Dryad [6]), while very few[17] adopt for instance standard vocabularies like the W3C DCAT vocabulary[18] or the dataset properties recommended by CRIS standards like VIVO (Datastar[19]) [7] or CERIF[20] [8]. And very few adopt a Linked Data approach.

---

[13] Currently at http://publicdata.eu/ but expected to move to data.europa.eu in October 2015
[14] http://search.datacite.org/ui?q=subject%3Aagriculture
[15] http://www.re3data.org
[16] http://datadryad.org/
[17] http://www.w3.org/2011/gld/wiki/DCAT_Implementations
[18] http://www.w3.org/TR/vocab-dcat/
[19] http://sourceforge.net/projects/vivo/files/Datastar%20ontology/
[20] See https://cerif4datasets.wordpress.com/c4d-deliverables/

Therefore, our effort with the RING was towards filling these gaps: we wanted to create a global dataset hub for agriculture which is fully machine-readable, provides very rich metadata and uses standard vocabularies (integrating them when necessary) and concepts so that applications can automatically re-use the data.

2 Semantics for the RING Linked Data

We decided to use the Linked Data approach [9] and to aim for Tim Berners Lee's 5th star[21] because we wanted to achieve the maximum level of interoperability possible.

The first step was the definition of our semantics.

Semantics in Linked Data are defined by "vocabularies": this term is often used to indicate two types of vocabularies that are both needed for describing and indexing any resource: 1) the metadata elements used to describe a resource defining its characteristics: these are usually defined in what we call metadata vocabularies, metadata element sets, or simply vocabularies; 2) the controlled vocabularies allowed for any of the metadata elements: these are normally defined in "concept schemes" or "value vocabularies" and can be of different types: thesauri, authority lists, classifications, or more in general Knowledge Organization Systems (KOSs). We maintain this distinction [cf. 10] in this paper using the terms "metadata vocabulary" and "value vocabulary".

2.1 Data model and metadata vocabularies

We needed to identify a data model and related metadata vocabulary that was suitable for the catalog. The main type of resources that we wanted to cover in the RING is datasets and the definition of datasets that we adopted is the definition proposed by the W3C Government Linked Data Working Group: "A collection of data, published or curated by a single source, and available for access or download in one or more formats."[22]

Around this definition, the W3C Working Group created the Data Catalog Vocabulary[23]. There are several reasons why we chose this vocabulary as our core vocabulary:

- We limited our survey to RDF vocabularies. There are good vocabularies for datasets that have not been formalized as RDF (like the re3data metadata set), but we wanted to make our dataset "linked" and wanted to adopt vocabularies that are formalized as RDF and use URIs.
- We wanted to adopt a vocabulary that was widely endorsed and we thought having the W3C behind it made DCAT a good candidate. Besides, the EC has since made

[21] See http://www.w3.org/DesignIssues/LinkedData (bottom of page) and http://5stardata.info/. The 5th star is about "linking your data to other data to provide context"

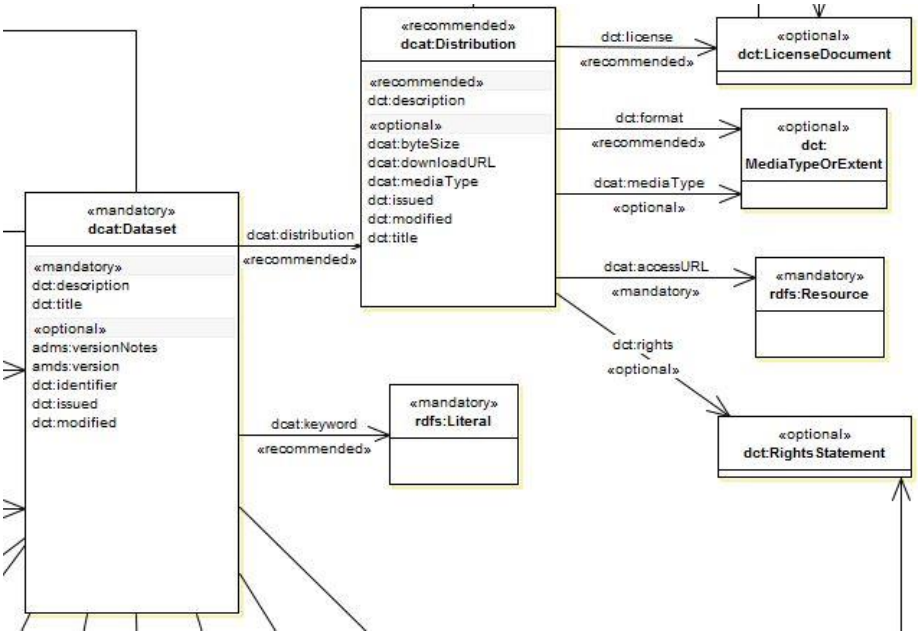
[22] http://www.w3.org/TR/vocab-dcat/#class--dataset

[23] http://www.w3.org/ns/dcat#

available the DCAT Application Profile – a set of recommendations for how to use DCAT in European data portals (see main text below).

- We needed a model that could represent the reality of the datasets we already had in our system, which in many cases had two or three "forms" of the same dataset. DCAT is designed around the relation between the dataset (the collection of data) and the "instances" of the dataset "available for access or download in one or more formats", called "distributions". This model suited our situation perfectly.
- We needed something sophisticated enough to distinguish between the dataset and its "distributions" but not so much specialized to be suitable only for very advanced cases (like VOID). We looked also at DataCite but the RDF version is still not official and the data model did not clearly distinguish between dataset and distributions.

In practice, since DCAT defines only new classes and properties for datasets while assuming the use of other existing vocabularies for the generic properties of any resource (like title, description etc.), we adopted an Application Profile that uses DCAT and formalizes also the re-use of other existing classes and properties from other vocabularies: the DCAT Application Profile for Data Portals in Europe (DCAT-AP)[24]. The figure below[25] shows the core entities of the DCAT-AP RDF model.



[24] https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final

[25] Full diagram: http://joinup.ec.europa.eu/site/dcat_application_profile/DCAT-AP_Final_v1.00.png

Fig. 1. Detail of the DCAT Application Profile RDF model

Besides the vocabularies already included in the DCAT-AP (Dublin Core[26], DCAT, FoaF[27], Vcard[28], SKOS[29]), in order to be interoperable with as many other systems as possible we also use other existing RDF vocabularies (VOID[30], DOAP[31], schema.org[32]) for additional (partial) descriptions of the datasets.

Furthermore, in order for the database to be fully interoperable by applications that needed more technical information on how to access the datasets, we needed a few additional properties that we published in a small extension to the DCAT vocabulary: the RING DCAT Extension.[33]

This small extension adds properties that support applications in accessing the datasets: for instance, the OAI-PMH[34] metadata prefix to specify the identifier of the metadata prefixes supported by the OAI-PMH target; or the subset ID to specify the name of the set or the URI of the graph that identifies the sub-set if a dataset is accessible through an API that supports the identification of a subset by limiting to a set (like OAI-PMH) or a graph (like SPARQL). This vocabulary also provides properties to link a dataset to a software tool or to an API method that can process it.

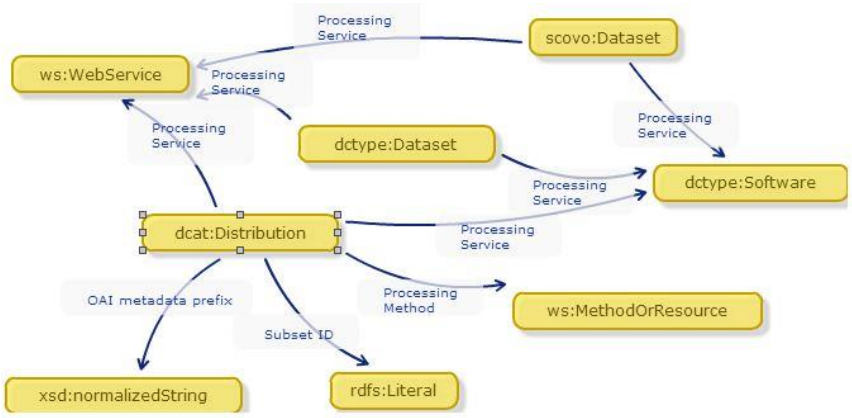


Fig. 2. A graph view of the small RING DCAT extension vocabulary

[26] http://purl.org/dc/terms/
[27] http://xmlns.com/foaf/spec/
[28] http://www.w3.org/TR/vcard-rdf/
[29] http://www.w3.org/TR/skos-reference/
[30] http://www.w3.org/TR/void/
[31] https://github.com/edumbill/doap/wiki
[32] http://schema.org
[33] http://vocabularies.aginfra.eu/dcatext#
[34] OAI—PMH is an exchange protocol for exposing metadata: https://www.openarchives.org/pmh/

2.2 Value vocabularies

This LOD layer, besides the RDF metadata vocabularies mentioned before, needs an infrastructure of LOD Knowledge Organization Systems (KOSs) or "value vocabularies" to univocally identify certain concepts that constitute the "values" for many of the dimensions that are essential to describe a dataset. Examples are: topics, geographic scope, data exchange protocols, metadata standards, file formats, data types etc.

Particular importance is given to the use of standards in the management of information: datasets are also linked to the vocabularies that they use. In order to provide comprehensive "authority" lists of existing information management standards that can be linked to the datasets, the RING harvests information from the registries available in the Agricultural Information Management Standards (AIMS)[35] website: the registry of metadata sets and the registry of Knowledge Organization Systems (KOS).

As regards other technical standards that are relevant to interoperability (protocol, notation), no comprehensive authority lists have been found, so the system provides either free-tagging lists that users can extend or local controlled lists. Such lists of controlled values are provided in the form of local SKOS Concept Schemes. These schemes have no pretense of becoming authority lists: they are used by the RING and by applications that use the RING until some the relevant authoritative bodies (e.g. IANA[36], W3C[37], Dublin Core) publish authority schemes using URIs. In the meantime, whenever possible the concepts in the RING local schemes have been mapped to the URIs of corresponding concepts in published schemes.

In order to have really "linked" data, whenever possible URIs in the RING are mapped to URIs in other authority data: for example, the RING local URIs of formats and notations are mapped, when possible, to the corresponding URIs (and in some cases URLs) from authoritative standardization bodies like IANA or W3C, as we said above; while local URIs for countries are mapped to the corresponding URIs in the FAO Geopolitical Ontology[38] and URIs of agriculture-related topics are mapped to the corresponding URIs in AGROVOC[39], the agricultural thesaurus published by FAO.

3 LOD publication approach for the RING Linked Data

Beyond the semantic aspects and the serialization of data as RDF, the actual publication of Linked Open Data (LOD) requires some additional steps and design work (cf. [12, 13]).

---

[35] http://aims.fao.org

[36] The Internet Assigned Numbers Authority (IANA) is responsible for the global coordination of the DNS Root, IP addressing, and other Internet protocol resources.

[37] The World Wide Web Consortium (W3C) is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards.

[38] http://www.fao.org/countryprofiles/geoinfo/en/

[39] http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

The first thing to consider for the actual publication of linked data is URI design and persistence [13, 14].

As for URI design, we initially decided to go for a simple URI pattern including the original RING domain name, a string identifier for the type of resource, and an ID for the resource. The URI for each resource in the RING is built as follows: {RING-domain}/node/{resource-ID}, e.g. http://ring.ciard.net/node/2417. The URI of each "concept" is built as follows: {RING-domain}/taxonomy_term/{concept-ID}, e.g. http://ring.ciard.net/taxonomy_term/108

This satisfied the requirements of having short unique and "opaque"[40] URIs for all entities. However, we have recently realized that URIs containing the domain name of an initiative or an institution are not ideal for persistence (see more on URI persistence in [14]): we are in the process of moving from the ring.ciard.net domain to the ring.ciard.info domain and we may lose control of the ciard.net domain in one year. So we decided to gradually move to PURL URIs: PURLs (Persistent Uniform Resource Locators) are Web addresses that act as permanent identifiers, allowing the underlying Web addresses of resources to change over time without negatively affecting systems that depend on them. RING URIs will become http://purl.org/net/ciardring/{resource type}/{resource-ID} and will resolve to the RDF and HTML versions of the resource at the URL where they are available at that moment.

The second thing is to provide machine access to the RDF data: the recommendations for Linked Data are to make them accessible through a) an RDF description at the resource URI; b) a SPARQL endpoint for querying the whole RDF store.

The RING was built with the Drupal[41] Content Management System, which provides modules that enable both the serialization of the metadata for each resource as RDF under a specific path and a SPARQL endpoint (for the RING: http://ring.ciard.info/sparql1). The RING also implements content negotiation [42] through Apache rewrite rules.[43]

In the end, the resulting LOD store publishes 74951 triples, 1186 concepts (around 500 of which mapped to external URIs), 1067 entities of type dcat:Dataset and 300 of type dcat:Distribution.

4 Examples of data re-use

Other applications can re-use data from the RING by sending SPARQL queries [15]. SPARQL queries are conceptually similar to SQL queries but rely on the published

[40] Not meaningful: humans or machines should not infer anything about the resource from the resource URI.

[41] http://drupal.org

[42] When an HTTP client attempts to dereference a URI, it can specify which type (or types) of content it would prefer to receive in response: if the client specifies HTML (like a normal browser), the system has to serve an HTML page; if the client specifies RDF, the system has to serve an RDF version of the resource.

[43] See http://www.w3.org/TR/swbp-vocab-pub/#recipe6

semantics of RDF vocabularies. These published semantics allow the application to write a query without the need to look at the internal structure of the database.

By just looking up the URIs of the entities and concepts in the RING[44], developers can send a query for instance to get all datasets available through the OAI-PMH protocol (the URI of the concept "OAI-PMH protocol" is http://ring.ciard.net/taxonomy_term/108): see an example of such a query at http://ring.ciard.info/get-all-datasets-available-through-oai-pmh.

The following examples illustrate how two types of applications (data aggregators and data processing tools) can leverage the RING to broaden the range of data sources they rely on. Using the RING instead of a local database of data sources a) allows data owners to update information on their datasets without the need of informing all the service / application providers that are using them (all applications using them will get the updated information in the query results from the RING); b) allows applications to dynamically find new suitable datasets without the need of constantly searching the web and updating their local lists, also exploiting work done by others; and therefore c) minimizes the duplication of effort and the creation of new silos.

4.1 Example 1: data aggregators using the RING as their collection database

Applications like data aggregators can register their data providers in the RING and then use it as a collection / dataset store to send queries and execute part of their workflows on them. An example of such usage of the RING data is AGRIS[45], a database of more than 7 million bibliographic references on agricultural research and technology and links to related data resources on the Web. AGRIS retrieves information on AGRIS data providers through a SPARQL query run against the RING looking for datasets that "belong to" (dc:partOf) the AGRIS network (http://ring.ciard.net/node/10687 is the URI of the AGRIS network in the RING):

```
... WHERE { ?dataset rdf:type dcat:Dataset . ?dataset
dc:partOf <http://ring.ciard.net/node/10687> ...
```

A similar use of the RING is made by AgriFeeds[46], an aggregator of news and events in agriculture that retrieves from the RING technical metadata about datasets available as RSS feeds. AgriFeeds makes a more dynamic use of the RING compared to AGRIS as it doesn't limit the query to datasets belonging to the AgriFeeds network but retrieves any dataset that is of type RSS and uses the "RSS metadata set", thus automatically increasing the number of feeds behind the service as new feeds are registered in the RING.

---

[44] http://ring.ciard.info/concept-uris and http://ring.ciard.info/entity-uris

[45] http://agris.fao.org

[46] http://www.agrifeeds.org

4.2 Example 2: data processing applications retrieving suitable datasets from the RING

Another example of data re-use is the iPython Notebook for estimating temperatures developed in the agINFRA project[47]. There are datasets, like the "European daily mean temperature series" maintained by the European Climate Assessment and Dataset project, that can be processed by this application. Since datasets in the RING are linked to software tools that can process them, the iPython Notebook can run regular queries to the RING to always retrieve the new datasets that might become available that are processable by the tool. The URI of the iPython Notebook in the RING is http://ring.ciard.net/node/19483, so the following fragment would filter all datasets that can be processed by the Notebook:

```
. ?distro dcat-ext:processingService
<http://ring.ciard.net/node/19483> .
```

5 Discussion and Further Work

Building the RING was partly a good generic exercise in creating a Linked Data dataset catalog and partly a practical community-specific implementation.

As concerns the exercise of building a Linked Data dataset catalog, what we think the Linked Data community may have to consider for the future is that existing dataset catalogs do not seem to be fully ready for data exchange, in either direction: a) aggregating data from them is in some cases possible but not to a high degree of granularity and not using shared semantics; b) most of these platforms, even the few that work as global directories (like CKAN), don't implement harvesting or aggregation: they require manual submission of datasets, thus implementing a centralized model and in the end building new silos.

As for the semantic aspects, what we have learnt is that there is a need on the one hand for extensions to the existing metadata vocabularies in order to better describe certain technical aspects of datasets (dimensions, syntax, reference standards…) and on the other hand for more authoritative reference lists exposed as Linked Data, possibly published by the relevant authorities, e.g. a comprehensive LOD reference list of serialization formats by IANA or an extension of the DCMI Type vocabulary.

Regarding the specific real case of the RING, our practical goal is to make it the reference dataset hub for agricultural information services: to get there, the RING has to reach a critical mass of registered datasets and a high level of metadata quality in order to become comprehensive and reliable enough for external services. To reach a critical mass of datasets registered, a move towards a federated approach is necessary. Past experiences show that centralizing the management of datasets is not a sustainable solution. Also forcing all providers to use the same platform will not work.

[47] agINFRA is an EC FP7 project completed in 2015 whose products are still accessible through the new website: http://aginfra.eu

Therefore, there is a need for a global directory of datasets in agriculture adopting a two-pronged approach: preferably, manual submission for higher quality of metadata and categorizations that are customized to agriculture and optimized for interoperability (this approach would also suit organizations that do not use any local platform and would provide them with a publishing platform); alternatively, exchange of metadata with existing platforms, in order not to force institutions to have a duplicate dataset publishing workflow. The implementation of a federation mechanism is the next step for the RING.

The objective remains that of making data produced by agricultural organizations more visible, better shared, easier to re-use and therefore actually consumable by integrated end-user services.

References

1. Chinese Academy of Agricultural Sciences, Global Forum on Agricultural Research, Food and Agriculture Organization of the United Nations: Interim Proceedings of International Expert Consultation on "Building the CIARD Framework for Data and Information Sharing", CIARD (2011)
2. Ballantyne, P.: ICTs transforming agricultural science, research and technology generation. Summary of the ICT Workshop at the Science Forum. GFAR, 2009
3. Crosas, M.: The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data. In: D-Lib Magazine, Volume 17, Number 1/2 (2011)
4. OpenAIRE Guidelines: For Data Archives, https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives
5. Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Fichtmüller, D., Pampel, H., Kindling, M., Kirchhoff, A., Göbelbecker, H., Klump, J., Bertelmann, R., Schirmbacher, P., Scholze, F.: Vocabulary for the Registration and Description of Research Data Repositories Version 2.0, re3data (2012)
6. Dryad Development Team: Dryad Metadata Application Profile, Version 3.0 (2010), http://wiki.datadryad.org/wg/dryad/images/8/8b/Dryad3.0.pdf
7. Ginty, K., Kerridge, S., Fairley, P., Henderson, R., Cranner, P., Bokma, A., Garfield, S.: CERIF for Datasets (C4D) – An Overview, C4D (2012)
8. Khan, H., Caruso, B., Corson-Rikert, J., Dietrich, D., Lowe, B., Steinhart, G.: DataStaR: Using the Semantic Web approach for Data Curation. In: The International Journal of Digital Curation, Issue 2, vol. 6 (2011)
9. Berners-Lee, T.: Linked Data. http://www.w3.org/DesignIssues/LinkedData.
10. Isaac, A., Waites, W., Young, J., Zeng, M.: Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets. W3C (2011)
11. DCAT Application Profile for Data Portals in Europe – Final, ISA Programme (2013)
12. Heath T., Bizer C.: Linked Data: Evolving the Web into a Global data Space (1st edition). In: Synthesis Lectures on the Semantic Web. Theory and Technology, 1:1, 1-136. Morgan & Claypool (2011)
13. Best Practices for Publishing Linked Data, W3C (2014). http://www.w3.org/TR/ld-bp/
14. Archer, P., Goedertier, S., Loutas, N.: Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. ISA Programme (2012) http://philarcher.org/diary/2013/uripersistence/
15. SPARQL Query Language for RDF, W3C (2007) http://www.w3.org/TR/rdf-sparql-query/