Europeana as a Resource for Social Scientists in Agriculture and Food: a Case Study

Maritina Stavrakaki, Nikos Manouselis, Vassilis Protonotarios, Thodoris Kontogiannis, Kostas Kastrantas, Panagis Katsivelis, Panagiotis Siokas[1]

A B S T R A C T

In this paper, we explore how we may give researchers from different disciplines new tools to enrich and use the data that are made discoverable and accessible through Europeana, the digital cultural aggregator of Europe. This paper presents a use case where selected content from Europeana is used as a resource for Social Scientists working in the agriculture and food domain. The process starts with the profiling and the identification of content requirements of the research community of the Greek Agricultural Economics Research Institute (AGRERI) to the enrichment of its library with quality content from Europeana and the development of a discovery microsite for AGRERI, providing access to the aforementioned selected content. The paper presents the connection of this content with AGINFRA, the data infrastructure for European agricultural research. This paper aims to showcase how researchers working in completely different disciplines may discover and exploit data sets of interest to them, from the vast amount of resources available through Europeana. By using these resources, agricultural (and not only) researchers can investigate various topics using different scientific methods and tools, thus making multi-disciplinary agriculture research more useful and meaningful.

1. Introduction

Europeana (http://www.europeana.eu) is the digital platform of the European Union (EU) that functions as a virtual European library showcasing the rich and diverse cultural heritage content of all participating countries. More specifically, working with a very large network of more than 3,000 museums, libraries and archives around Europe, more than 40 million digital cultural artefacts are currently made discoverable and accessible through the Europeana portal. Traditionally, the main audience of the Europeana content consisted of various types of museum stakeholders, historians, librarians, scholars and other stakeholders in history, art and social sciences, to name a few; on the other hand, despite the high quality of the content available through Europeana, there was no connection of Europeana with the research communities of other domains, like life sciences.

The Europeana Cloud Project (Europeana Cloud: Unlocking Europe's Research via The Cloud, http://pro.europeana.eu/structure/europeana-cloud) has been a project co-funded by the European Commission through the ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme (CIP). It aimed to create a digital infrastructure in order to support researchers from different disciplines discover valuable and rare resources from disperse sources and collections, namely Europeana (Benardou, Dallas & Dunning 2014).

Socioeconomics and policy research have become an integral part of agricultural research projects and programs. The concepts, framework and methodological tools and techniques for socioeconomic and policy research are derived from social sciences disciplines which help bring human dimension and issues in the research process and cover a wide range of research issues.

---

[1] Maritina Stavrakaki, Nikos Manouselis, Vassilis Protonotarios, Thodoris Kontogiannis, Kostas Kastrantas, Panagis Katsivelis, Panagiotis Siokas
Agroknow, 17 Grammou Str., 15235, Athens, Greece
maritina.stavrakaki@agroknow.com, nikosm@agroknow.com, vprot@agroknow.com, kontogiannis.thodoris@agroknow.com, kkastrad@agroknow.com, katsivelis.panagis@agroknow.com, panagiotis.siokas@agroknow.com

Social science can contribute critically in the analysis and conduction of research in crosscutting themes such as poverty, food security and natural resource management. They provide broad framework, research methods and tools for scarce resource allocations, identification of research agenda, research priority setting, adoption and impact assessment studies and can play important role by identifying policy gaps and constraints to technology development, uptake and scaling-up[2].

To this context, an interesting discipline that could evaluate the importance of such infrastructure is the research community of agricultural researchers at the Agriculture Economic and Policy Research Institute (AGRERI, http://www.agreri.gr). These researchers investigate the economic, cultural and societal factors affecting major aspects of the agricultural and food sector not only in Greece but also in other European countries and conduct high-level research. An infrastructure that will be able to integrate and present in a simple and visual way all this information from disperse sources will be of great importance not only to researchers but to anyone who would like to take advantage of all this knowledge. It will also change the way that researchers are currently storing and using the data produced as well as answering and making more sense of the data produced.

2. Methodology

The methodology followed and described aims to provide the means for extracting information from potential users of the service (Van den Berg, Parra, Jentzsch, Drakos & Duval 2014). The requirements extracted from these stakeholders will be used for shaping the expected outcomes of the service for the end users. The methodology comprises a series of steps involved in a process starting with the profiling and the identification of content requirements of the research community of AGRERI to the enrichment of AGRERI's library with quality content from Europeana and the development of a discovery microsite for AGRERI, providing access to the aforementioned selected content from Europeana and other sources.

More specifically, the first step in the proposed methodology is the identification of the different user profiles (personas) of interest to the service. It is very important to identify the content requirements (content analysis, information-related challenges and problems) and to have a very good description of the information services and systems and the usage scenarios related to them, in order to be able to design interface wireframes and describe the user-envisaged usage workflows. This will provide the resulting requirements for outcomes services. After validating these requirements with the users through interviews, workshops and surveys, final wireframes with the required features of the service are designed that will translate these features into requirements for the service. Reports from all workshops are used for the analysis of requirements that will drive the development of services by the technical team. A schematic representation of this process is shown in the figure below (Figure 1).

As mentioned earlier, the implementation of the methodology requires a profiling process and the identification of content requirements through the use of instruments and tools for the extraction, collection and organization of the user requirements as well as for their validation and finally, the evaluation of the suggested features of the service.

---

[2] https://ec.europa.eu/research/social-sciences/pdf/other_pubs/socio-economic-tools-for-sustainability-impact-assessment_en.pdf

Maritina Stavrakaki, Nikos Manouselis, Vassilis Protonotarios, Thodoris Kontogiannis, Kostas Kastrantas, Panagis Katsivelis, Panagiotis Siokas: Europeana as a Resource for Social Scientists in Agriculture and Food: a Case Study
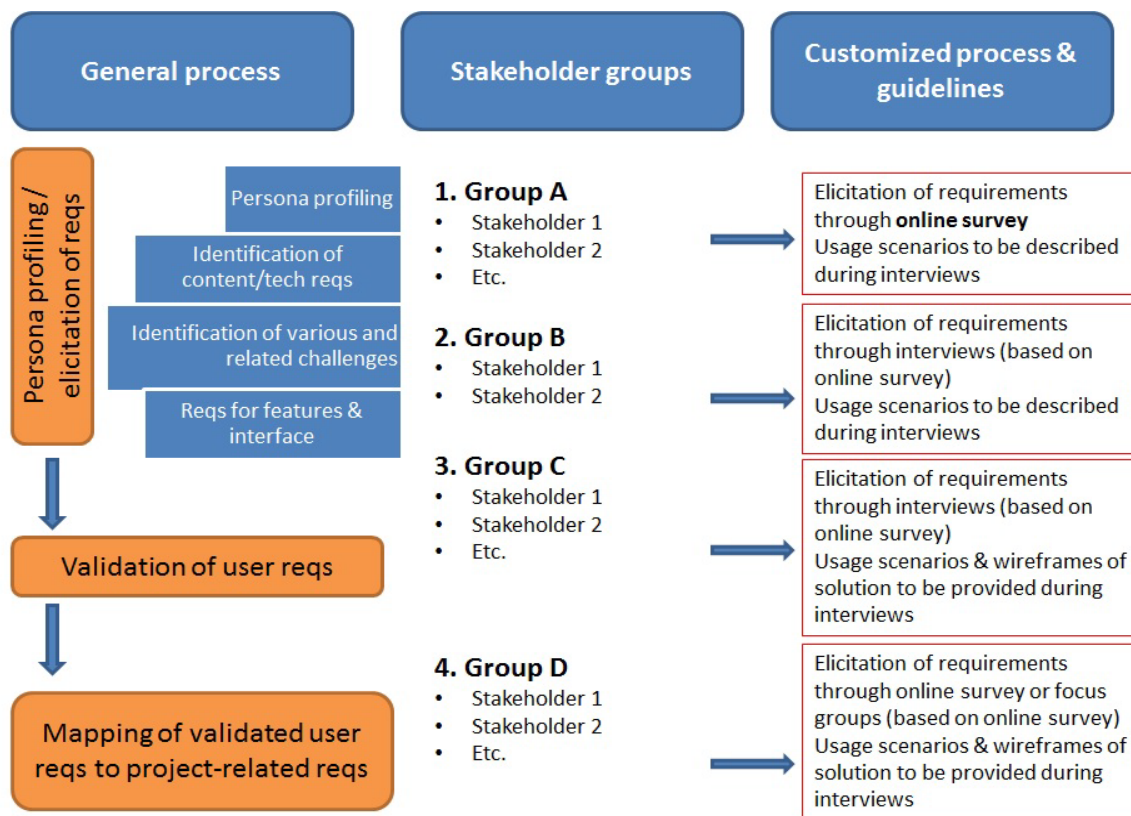
Figure 1. Workflow followed for the development of the services

The first tool used is an online questionnaire developed as a Google Form which consists of: a) questions related to the demographics of the personas, b) questions related to the data of relevance and interest to the specific persona, c) questions related to content-specific challenges faced by the persona and d) some suggestions that could be of use to the persona. In this case, fourteen (14) users answered the online questionnaire providing the first input for the user requirements elicitation.

In order to validate the challenges and problems collected via the online questionnaire, a set of face-to-face meetings and interviews takes place with potential stakeholders and representatives of the community in question, using the same set of questions in the online questionnaire but in a more in-depth way. In this case, nine (9) stakeholders participated in the interviews session providing sufficient feedback.

In order to evaluate the suggested features, a workshop was organized aiming to provide a hands-on evaluation of the service not only by the agricultural researchers of the AGRERI team but also with other users interested in the service. A total of fifteen (15) people participated in the hands-on workshop providing their feedback through discussion and an evaluation survey.

3. Results

The AGRERI Discovery Microsite (http://www.agreri.gr/en/external_library/browse) is a Microsite that can be used by researchers in the field of agricultural economics to search and discover relevant resources. The Microsite searches predefined collections of datasets based on user input and presents the results in a uniform way. It comes as a Microsite, built with lightweight web-technologies (HTML, CSS, HTTP, Javascript, AJAX) in order to be easily embedded in sites and web-applications, without the need to make changes for matching the existing technologies of the application. The first prototype of the Microsite is designed based on the needs of the AGRERI agricultural researchers. The Microsite is a tool in the following two ways: it is graphically designed

to be smoothly integrated with the web site of the AGRERI team and is built on top of collections that have been requested by stakeholders.

The main usage of the AGRERI Discovery Microsite is a faceted search interface that allows users to search and quickly filter the results (Figure 2) and then, get a more detailed view of the resource selected with the related resources of the same thematic. In addition, predefined categories that allow access over specific content (i.e. agricultural economics) are also available.

Figure 2. Screenshot from results listing after a search query

The prototype has been designed and developed with the constant feedback from the AGRERI team in order to better capture and cover their needs. During the development phase, a number of discussions were organized between the group and technical team in order to gather feedback concerning which collections to search, which categories to use, and which facets the stakeholders would like to use.

In order for the finder to allow faceted search and uniform representation of the metadata from resources coming from different collections, the Microsite uses the existing AGRERI infrastructure to store a repository with all these metadata. In the pilot, the resources and different collections are stored in the repository and in all cases an API is used to filter thematically resources for the dataset. At the same time, and in order to provide this uniform representation and make the resources available throughout the Microsite a transformation process took place, where all metadata records were transformed from their original scheme to an internal format. During this transformation procedure metadata records have also been enriched.

In the AGRERI Discovery Microsite, there are many different collections stored. The AGRERI Library contains resources (publications, reports and other material) of the AGRERI researchers. The External Libraries contain collections from external sources and this way these collections are made discoverable through the Microsite, as in the case of Europeana (Figure 3). Other content collections of interest to the users available through the Microsite are:

- Collections from AgEcon Search (http://ageconsearch.umn.edu)
- Collections from FAO AGRIS (http://agris.fao.org)
- Collections from the United States Department of Agriculture (http://www.usda.gov)
- Collections from the National Institute of Agricultural Research (http://www.inra.fr)
- Collections from the Food Policy Research Institute (http://www.ifpri.org)
- Collections from the Wageningen University and Research Center (http://www.wageningenur.nl)
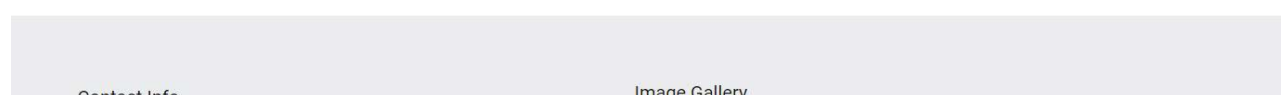

Figure 3. External libraries

All these collections consist of resources of interest to the AGRERI research community, combining traditional agricultural topics with ones on socio-economic ones. In addition, these data sources provide open access to their content therefore the aggregation of the corresponding metadata is possible.

3.1. Architecture

The AGRERI Discovery Microsite is implemented using modules at three layers, namely a) the data ingestion layer, b) the repository layer and c) the front end layer (Figure 4).
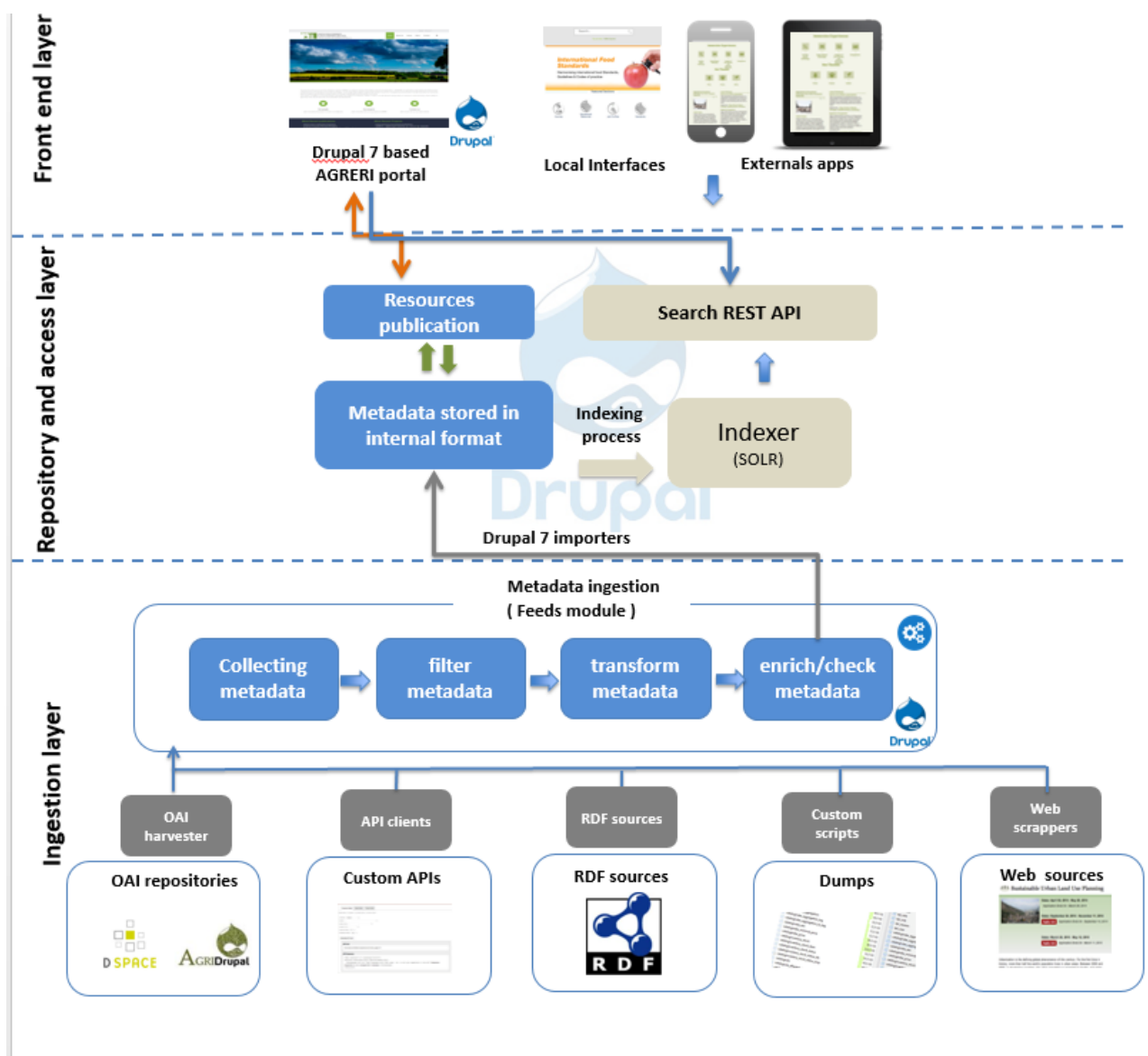
Figure 4. AGRERI Discovery Microsite architecture

More specifically, the following modules and components were deployed in the context of the Europeana Cloud project in order to implement the AGRERI Discovery Microsite:

- **REST API client** that collects relevant content from Europeana. A custom client has been developed to collect the relevant Europeana content in a JSON based format through Europeana's REST API.
- **Data ingestion module** that imports the metadata in AGRIS metadata Application Profileformat to the AGRERI site. This module can be used by the researchers to import new relevant content from Europeana. The data ingestion module includes the following components.
  - **Transformation component** that transforms the data from Europeana format to AGRIS AP that is suitable for agricultural research
  - **Enrichment component** that enriches the Europeana metadata with information about the thematic category of the content
- **Data indexing module** that indexes the metadata so they are available in the AGRERI Discovery Microsite.

● **Front end modules** (AGRERI Discovery Microsite) site that allows the discovery of the content from Europeana and connects this content to the existing content for agricultural economics that the AGRERI institution holds.

3.2. Ingestion (Data import)

The AGRERI Discovery Microsite offers a data ingestion mechanism. This data ingestion mechanism is responsible for importing both content and metadata records from multiple diverse sources that are publishing the metadata records through custom API (e.g. Europeana API), OAI-PMH protocol, RDF and dump files.

An importer can ingest content from raw files that are located to web apps directory or from a web API through Rest Services and to be imported into the AGRERI site. Also, before the data are stored in the database, the cleaning mechanism, if configured, removes unwanted content or curates it. For example, it can remove unwanted HTML characters. Another important functionality of the importer is the filtering mechanism that could be configured in order to filter out not relative content. For example, if a record is irrelevant to agriculture economics thematics, then this record will not be imported to the AGRERI Discovery Microsite

In order to provide to the user all the resources of the AGRERI Discovery Microsite, the powerful Apache Solr 1.4 engine is used. Furthermore, the ingested content is indexed with Apache Solr and offered via this engine to the user search interface under a specific schema

4. Connection with AGINFRA

AGINFRA (http://aginfra.eu) is the European research hub and thematic aggregator that catalogues and makes discoverable publications, datasets and software services (Drakos, Protonotarios & Manouselis 2015) developed by Horizon 2020 research projects on topics related to agriculture, food and the environment. The overall goal is to include them in the European research e-infrastructure "European Open Science Cloud" (https://www.openaire.eu/the-european-open-science-cloud-for-research-2), a synergy between various EU initiatives and networks, namely OpenAIRE (http://openaire.eu), EUDAT (http://eudat.eu), GEANT (http://www.geant.org), EGI (https://www.openaire.eu/egi.eu), LIBER (http://libereurope.eu). It is envisaged as evolving into an open data management and sharing infrastructure for research and extension on agriculture, food and the environment.

AGINFRA plays plays multiple roles in the global agri-food research ecosystem (Pesce, Geser, Protonotarios, Caracciolo & Keizer 2013). More specifically, it can function as:
● an atlas of agricultural research & extension (including institutions, people, publications, data sets, projects, courses, OERs);
● a semantic layer of processing, enriching & interlinking research information from distributed, heterogeneous sources & formats;
● a catalogue of software components (open source software stack & APIs) that anyone may use to process research information;
● a help desk service to support institutions & projects that wish to publish their research information openly;
● a set of data-rich service and application demonstrators for specific case studies (food safety, viticulture, crop composition, etc.)

4.1. Rationale

We wanted to investigate how Europeana collections may be linked to the domain-specific research infrastructures of the agriculture community, through the research data e-infrastructure for agriculture and food (AGINFRA), in order to make relevant collections discoverable also through the CIARD RING (http://ring.ciard.info) registry by both humans and software services, and see if and how the collections in Europeana can have value to agricultural social scientists. Also, whether there can be an automated dissemination mechanism by federating the Europeana and RING

registries and how we can use the APIs and support the Europeana portal in order to offer access to Europeana collections through the domain-specific data e-infrastructure of AGINFRA.

4.2. Implementation

CIARD RING is a core component of the AGINFRA e-infrastructure; in fact, it is a global directory of web-based information services and datasets for agricultural research for development (ARD). It is the principal tool created through the CIARD initiative to allow information providers to register their services and datasets in various categories and so facilitate the discovery of sources of agriculture-related information across the world (Pesce, Maru & Keizer 2011).

In the context of the AGINFRA, the RING provides a machine-readable Linked Data layer to meet the following requirements:
- Datasets registered in the RING 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 RING 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;

This machine-readable layer can for instance support the data aggregation workflows of external services. The CIARD RING is online accessible at http://ring.ciard.info/.

In order to register services and datasets in the RING, a user must first register to the service. By creating an account, a confirmation e-mail is sent to the user's email address. After the e-mail validation, the user can register a service or dataset. After registering to CIARD RING, the user can create an information source or dataset. In our case, we wanted to register a search query from Europeana using the keywords "agricultural cooperative" with the URL http://www.europeana.eu/portal/search?q=agricultural+cooperative. The user must complete all fields marked with an asterisk (*) in every tab and section and any other fields suggested. These include:
- Basic: name, description, URL, responsible body, language etc.
- Geo: location country, geographic coverage.
- Thematic: domain, other topics, specific topics (AGROVOC) etc.
- Content: type of service, languages covered, audience, open access mandate etc.
- Standards: metadata set(s) used, KOSs used etc.
- Access to data: distribution(s), special instructions etc.
- Aggregation: harvesting protocol, required format etc.
- Networks: belongs to networks etc.

After completing all the necessary fields, it becomes apparent how a simple initial search query in Europeana using specific keywords (i.e. agricultural cooperative) can become registered in the CIARD RING. This way, the content which exists in Europeana on agricultural cooperatives becomes available as a research dataset through AGINFRA and therefore, through AGINFRA-powered services like AGRIS where domain-specific and not only researchers can find it. Something which is extremely important taking into consideration the fact that AGRIS has more than 200.000 visitors per month, and thus enhancing the visibility and re-use of aggregators like the Europeana portal.

4.3. Recommendations on establishing a liaison between Europeana and AGINFRA

The information provided in the previous sections highlighted the potential of the linking between Europeana and domain-specific data registries like the CIARD RING of AGINFRA by providing a specific example with the steps that a user needs to follow in order to register a possible dataset. Through this process, a number of limitations occurred which led to some

recommendations in order to establish a more stable liaison and content exchange and discovery mechanism to help AGINFRA communities find collections of relevance and of interest in Europeana.

Indeed, Europeana provides access to content of interest to agricultural researchers so this content needs to be more easily accessible to them; however, Europeana does not include domain specific collections (e.g. ones on agricultural topics). In this context, new, alternative and innovative ways that will allow sharing of content of interest among researchers and other types of stakeholders will not only facilitate the sharing of high-quality content but at the same time it will significantly enhance the reuse of the Europeana content and its use in different domain specific applications. In the specific case described in the previous sections, a new collection is created based on the search results for a specific query.

This envisaged integration of services like CIARD RING in the specific use case presented in the document, as well as domain specific services from other domains, will require the consideration of a number of issues that were identified during the analysis presented earlier. Such issues are the following:

- Definition of the person responsible for the collection, so that it can be attributed and contacted when needed.
- Identification of the owner of the collection, so that he/she can be contacted for use and licensing issues.
- While the process can be completed by any user, a user with basic knowledge on metadata, licensing etc. may more easily complete the fields.
- Each URL is a dynamic list of results that are compiled after the query. Any revision in the content of the Europeana collections will result in the corresponding changes in the search results.
- A user has to register his/her organization if it doesn't already exist in CIARD RING. This might be an issue in the case of users that do not represent an organization but act as individual users.

In this context, addressing the aforementioned issues can provide a set of recommendations that will facilitate the linking of Europeana with domain specific services, like CIARD RING, as similar issues may have to be addressed in other contexts as well. This linking may be further enhanced by direct communication between the technical team of the Europeana and technical teams of other services that want to be linked to Europeana. This will ensure the optimal technical interoperability, in cases where automatic metadata and content sharing takes place. Last but not least, the definition of ownership over collections, like the one mentioned earlier in the case of CIARD, as well as the allowed uses through a well-defined licensing schema will further facilitate the use and reuse of Europeana content through third party services.

5. Conclusions and next steps

The specific piloted case study presented how Europeana may be used as a resource for Social Scientists and researchers that work in different scientific domains and disciplines, in this case in the domain of agricultural economics, how they can make use of the available resources and data sets and how they can be benefited from such digital infrastructures.

The methodology followed for the elicitation of requirements in this piloted case study has already been used, tested and validated in other case studies and it could also be used for the elicitation of requirements from targeted end-users in other case studies as well. The methodology consists of a number of well-defined steps that should be completed for the successful acquisition of requirements, steps that can be adapted and adopted accordingly to serve other cases.

In this specific piloted case study, the feedback received by the AGRERI researchers was positive since they were extremely satisfied both by the methodology followed and by the final outcome. They were impressed by the fact that Europeana proved to be quite useful even for researchers in their own discipline, since they had not anticipated to find useful and related

resources on agricultural economics and policy thematic areas, specifically when it comes to accessing rare and useful content, like the one that Europeana portal provides.

Through the connection that was established between Europeana and the domain-specific research data e-infrastructure AGINFRA, with the aim to examine how to make the relevant collections discoverable through AGINFRA's specific registries and services, it is obvious that there is potential for using Europeana content in other domains as well. This will allow to further facilitate the sharing of high-quality content and at the same time it will significantly enhance the reuse of the Europeana content and its use in different domain specific applications.

6. Acknowledgements

7. References

Benardou, A, Dallas, C, & Dunning, A 2014 'From Europeana Cloud to Europeana Research: The challenges of a community-driven platform exploiting Europeana content'. In Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Volume 8740 of the Series Lecture Notes in Computer Science pp. 802-810. Springer International Publishing. doi: 10.1007/978-3-319-13695-0_82

Drakos, A, Protonotarios, V, & Manouselis, N 2015 'agINFRA: a research data hub for agriculture, food and the environment'. *F1000Research*, 4:127.
doi: 10.12688/f1000research.6349.1

Pesce, V, Geser, G, Protonotarios, V, Caracciolo, C, & Keizer, J 2013 'Towards Linked Agricultural MetaData: Directions of the agINFRA Project'. In AgroSEM 2013, Thessaloniki (Greece), 19-22 November 2013. [Conference paper].

Pesce, V, Maru, A, & Keizer, J 2011 'The CIARD RING, an Infrastructure for Interoperability of Agricultural Research Information Services' [Article and Abstract]. *Agricultural Information Worldwide*, vol. 4, no. 1, pp. 48-53.

Van Den Berg, H, Parra, G, Jentzsch, A, Drakos, A, & Duval, E 2014 'Studying the history of philosophical ideas: supporting research discovery, navigation, and awareness'. In *Proc. of the 14th Inter. Conf. on Knowledge Technologies and Data-driven Business (i-KNOW '14)*, 16-19 September 2014, Graz, Austria. ACM, 12. doi: 10.1145/2637748.2638412