Maastricht University

Integrating Agriculture-related Data Provided by Thematic Networks into a High Impact Knowledge Reservoir

Citation for published version (APA):

Panoutsopoulos, H., Espejo Garcia, B., Verbist, P. E. G., Fountas, S., Spanoghe, P., & Brewster, C. (2021). Integrating Agriculture-related Data Provided by Thematic Networks into a High Impact Knowledge Reservoir. In G. Vlontzos, Y. Ampatzidis, B. Manos, & P. M. Pardalos (Eds.), *Modeling for Sustainable Management in Agriculture, Food and the Environment* (pp. 199-226). CRC Press. https://doi.org/10.1201/9780429197529

Document status and date:
Published: 01/01/2021

DOI:
10.1201/9780429197529

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 24 Apr. 2024

CHAPTER
7

Integrating Agriculture-related Data Provided by Thematic Networks into a High Impact Knowledge Reservoir

Hercules Panoutsopoulos[1*], Borja Espejo Garcia[1], Philip E.G. Verbist[2], Spyros Fountas[1], Pieter Spanoghe[2] and Christopher Brewster[3]

[1] Department of Natural Resources Management and Agricultural Engineering, Agricultural University of Athens, Greece
[2] Department of Plants and Crops, Faculty of Bioscience Engineering, University of Ghent, Belgium
[3] Institute of Data Science, Maastricht University, Netherlands

Introduction

Knowledge is conveyed through digital artifacts that people create as part of their everyday activities. Making these artifacts widely available may involve the use of database systems able to handle their storage and management. These systems deploy database models explicitly defining the internal database structure. Database models are specific types of data models, whose role is to formally describe entities, as well as their associations, and further enable data type specifications, relation definitions and constraints identification (Umanath and Scamell 2014). However, technology provides a wide spectrum of possibilities for the creation of digital artifacts and data generation. The abundance of the technologies and tools available allow for the production of data in a variety of formats. In many cases, knowledge is communicated through semi-structured and unstructured data rather than structured data. This has direct implications for the data store system to adopt the respective database models. The relational database management systems (RDBMSs)

*Corresponding author: hpanoutsopoulos@aua.gr

are not the optimal solution for the storage and management of semi-structured and unstructured data, especially when large volumes of such kinds of data need to handled. As a consequence, a new paradigm of database technologies and models has emerged (da Silva et al. 2015).

The aim of this chapter is to illustrate the initial attempts towards a formal definition of the concept of Thematic Networks, through a specially designed ontology, and a database model for storing Agriculture-related data into a centralized data store. The book chapter's focus is EURAKNOS[1], which is an EU-funded, H2020[2] research project collating Agriculture-related data, in various formats, in order to make it available to any interested parties and stakeholders for further analysis or development. The data considered in EURAKNOS is disseminated by Thematic Networks, which, in the project's context, are conceptualized as Knowledge Reservoirs.

EURAKNOS is a Thematic Network itself and as other Thematic Networks its overarching goal is to aggregate and further disseminate data conveying useful information and knowledge to a range of stakeholders targeted (namely, farmers, foresters, and advisors). In other words, its aim is to create a so called "High Impact Knowledge Reservoir". To this end, a data repository is going to be developed as the core component of the EURAKNOS digital platform. A similar approach has also been taken by the other Thematic Networks, which allow access to agricultural data stored into their own repositories. As an example we can mention: (i) the Inno4Grass[3] Thematic Network's "Encyclopedia pratensis" data repository[4]; (ii) The Smart-AKIS[5] Thematic Network's "Smart Farming Technologies Platform"[6]; and (iii) the 4D4F[7] Thematic Network's "Technology Warehouse"[8]. In addition, we may also refer to cases of Thematic Networks (e.g. AFINET[9]) that make their data available via general-purpose repositories such as Zenodo[10]. A thorough analysis of the repositories of the existing Thematic Networks, already undertaken in EURAKNOS, has revealed a number of design-related issues and inconsistencies with regard to the provision of the data to the user. More specifically, by taking account of the different scope of each Thematic Network, there is a great variance in the information and content needed

[1] https://www.euraknos.eu/
[2] https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020
[3] https://www.inno4grass.eu/en/
[4] https://www.encyclopediapratensis.eu/
[5] https://www.smart-akis.com/
[6] https://smart-akis.com/SFCPPortal/#/app-h/technologies
[7] https://www.4d4f.eu/
[8] https://www.4d4f.eu/content/technology-warehouse
[9] https://euraf.isa.utl.pt/el/afinet
[10] https://zenodo.org/

to conveyed, with a range of data formats being used. However, apart from the variance in the content/information and the data formats being used, the major deficiency, inherent to the design of all the existing Thematic Network repositories, is the inconsistency in the categorization and characterization of the data. In other words, there is no commonly accepted framework for the classification of the data disseminated by Thematic Networks, which, in turn, makes any attempt to proceed to a process of cataloguing them a difficult endeavor. This issue has direct implications for the annotation of data with metadata and the facilitation of the data search and retrieval process in compliance with contemporary recommendation and guideline frameworks such as the FAIR principles (Wilkinson et al. 2016). EURAKNOS aspires to help close this gap and contribute to a solution to the problem of the observed, design-related variances and inconsistencies by proposing a robust model that efficiently captures the entire spectrum of the data types and formats available by Thematic Networks.

Given the above, the efficient management of data heterogeneity and the need to structure data in an efficient manner, are core challenges for EURAKNOS and the various Thematic Networks, in the agricultural sector, which cover a range of activities, and topics, from dairy farms to agroforestry. This chapter begins with an outline of the concept of Thematic Networks and the rationale behind them. After that, a description of the EURAKNOS project takes place with an emphasis on the methodology adopted for identifying, collating and making available Thematic Network outputs (i.e. data) of interest. The ontology of Thematic Networks, aimed to be used for the needs of metadata definition, is presented next. From a technical perspective, the implementation of the High Impact Knowledge Reservoir will be based on an appropriately-designed digital platform. So, after an overview of the digital platform's architecture, a number of design-related considerations are discussed. These relate to: (i) the retrieval and integration of the Thematic Network outputs; (ii) the EURAKNOS data store and database model; (iii) the search technologies to be considered; and (iv) the user interaction. The chapter concludes with a summary of the key points of the work undertaken, its key contributions and limitations, as well as the steps to be taken next.

The Role of Thematic Networks as Multipliers of Agriculture-related Knowledge

EURAKNOS is a Thematic Network. Thematic Networks are a particular format of Multi-Actor projects, promoted by EIP-AGRI and funded by EU's Horizon 2020 programme, working on specific themes. They bring people from both the science and practice domains together to create practical and useful outputs. The concept of Thematic Networks is at the core of EURAKNOS. It has stemmed from the need to efficiently address Agriculture-/Forestry-related problems of ever-increasing complexity and identify novel solutions. The dynamic agricultural landscape calls for out-of-the-box thinking and challenges the top-down, linear model of knowledge transfer from the research community to practitioners. Scientists can no longer be viewed as the only innovation brokers in the Agri-food value chain. This means that other stakeholders also have an important role to play. In other words, solutions to problems need to occur as an outcome of coordinated, joint efforts of various actors bringing different perspectives and values to the process and, consequently, contributing to synergies between research and practice.

The above described approach is core to what the agricultural European Innovation Partnership (EIP-AGRI) calls an "*interactive innovation model*"[11] and constitutes the foundation of the Multi-Actor projects (EIP-AGRI 2017). The use of the term "*interactive*" is indicative of the way in which innovation is conceptualized as it is expected to occur from bottom-up processes involving seminal interactions among the various stakeholders (e.g. farmers, advisors, SME and NGO representatives, researchers) and aims to further the already existing knowledge. Thematic Networks are a particular format of Multi-Actor projects (EIP-AGRI 2016) and as such they:

- relate to specific themes (e.g. sustainable cropping systems, animal production systems, plant health, rural dynamics and policies, knowledge and innovation systems) and focus on real problems of practitioners;
- bring together partners with complementary backgrounds and expertise to collaborate, during the entire project lifecycle, and come up with innovative solutions to problems;
- collect, further develop, as well as disseminate ready-to-apply recommendations and practices and, thus, serve as multipliers of the Agriculture-related knowledge.

The key characteristics and properties of Thematic Networks are summed up in the following definition, which has been developed with the aim to facilitate the ontology creation:

Thematic networks are multi-actor projects collecting existing knowledge[12] and best practices on a given theme, which relates to a domain. Their aim is to produce outputs, having a specific purpose and addressing a particular topic, which are of different kinds (e.g. text, software, images,

---

[11] https://ec.europa.eu/eip/agriculture/en/eip-agri-concept

[12] In the context of EURAKNOS project, the term knowledge is used to denote the data that convey agricultural knowledge.

audio, video, and datasets) and formats. Outputs target end-users (e.g. farmers, foresters, advisors) and are produced by output creators, engaged in the Thematic Network, which can be distinguished into organizations (e.g. universities, SMEs, NGOs, research institutes), having different kinds of involvement in the Thematic Network (participatory or leading), and individual actors (e.g. farmers, livestock breeders, advisors).

EURAKNOS: A Meta-Thematic Network for Collecting, Compiling and Enabling Access to Existing Best Practices and Knowledge

The goal of the EURAKNOS project[13] is to collate, store and further disseminate already created, Agriculture-related data. It is a Thematic Network aiming to strengthen agricultural knowledge and promote innovation across Europe. By bringing together academic organizations, research institutes, advisory centers, government bodies, SMEs, NGOs and farmer organizations, EURAKNOS embraces the Multi-Actor approach and acts as an "umbrella" project aiming to connect all Thematic Networks. It can be considered as a Thematic Network about Thematic Networks, i.e. a meta-Thematic Network. Through the coordinated efforts of a consortium of 17 organizations from 11 European countries, with an extensive expertise in innovation in Agriculture and Forestry, EURAKNOS aims to attain a number of objectives relating to:

- the establishment of connections with existing Thematic Networks and the increase of their impact as creators and multipliers of Agriculture-related knowledge;
- the collection, analysis and evaluation, on the basis of well-established quantitative/qualitative criteria, of the knowledge and tools available by Thematic Networks;
- the development of guidelines and technical specifications for a centralized repository of agricultural knowledge, as well as recommendations for future Thematic Network-related initiatives;
- the development of a platform consisting of the EURAKNOS Thematic Network repository, search engine, and a web-based environment for querying and accessing Agriculture-related data;
- the promotion, sustainability and longevity of the project by making use of fit-for-purpose communication and dissemination channels and also pursuing links with other Thematic Networks, research projects (at the EU and national levels), as well as educational/ training programs.

---

[13] The EURAKNOS project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817863.

Methodology for Agriculture-related Data Collection, Storage and Dissemination

Methodology Overview and Involved Steps

As mentioned earlier, a term core to the project is that of "Knowledge Reservoir" (KR). EURAKNOS considers all Thematic Networks as KRs. A KR is a collection of digital artifacts conveying practical, ready-to-apply knowledge able to contribute to innovative solutions for a sustainable Agriculture and Forestry. The aim of EURAKNOS is to build a High Impact Knowledge Reservoir (HIKR), namely a KR that uses a content structuring model tailored to needs of the end-users targeted. The activities for the data collection, storage and dissemination are described in a four-phase methodology. The phases of the EURAKNOS methodology relate to: (i) Evaluating, (ii) Determining, (iii) Exploring, and (iv) Widening. Descriptions of the activities involved in each phase are provided in Table 1.

After having mentioned the phases of the methodology adopted in EURAKNOS and the activities involved in it, it is helpful to proceed to the presentation of the expected, from the Thematic Networks' evaluation, outcomes and the description of what the recommendations for an HIKR should be about.

State-of-the-Art Review and Recommendations for a High Impact Knowledge Reservoir

The first two methodology phases (namely, "Evaluating" and "Determining") set the ground for the development of the technical specifications and the creation of the EURAKNOS platform. As mentioned earlier, the "Evaluating" phase involves a rigorous analysis of how Thematic Networks have addressed the issues of: (i) data storage and management; (ii) KR design; (iii) communication, dissemination and exploitation; and (iv) Multi-Actor engagement. To attain this goal, a number of specially designed, fit-for-purpose tasks (namely, "data impact assessment", "design of knowledge reservoirs", "dissemination, communication and information strategies" and "Multi-Actor approach") have been executed. These tasks have involved a combination of desktop study and face-to-face interviews of Thematic Network representatives. Table 2 below lists all the tasks involved in the specific methodology phase together with references to the research tools used and the outcomes expected.

The "Evaluating" phase gives input to the "Determining" phase. This input is required for the creation of recommendations for developing the EURAKNOS HIKR. In order to better illustrate the synergy between

Table 1. Phases of EURAKNOS methodology, description of their scope and alignment with project work packages

#	*Phase*	*Involved activities*
1	Evaluating	• Definition of the qualitative and quantitative criteria (based on cultural, socio-economic and environmental aspects) for ranking Thematic Network outputs with regard to their applicability at the local, regional, national and EU level. • Evaluation of the ongoing and completed Thematic Networks. Operational Groups[1] and other relevant Multi-Actor projects are also taken into consideration. • Evaluation focused upon the knowledge and tools produced, as well as the communication and dissemination channels employed for reaching out to the end-users and maximizing the impact of Thematic Networks. • Evaluation of the Multi-Actor approaches (employed by Thematic Networks) with the aim to estimate the short-, medium-, and long-term involvement of the different types of actors.
2	Determining	• Development of recommendations and guidelines for: • The collection, storage and sharing of ready-to-be-used agricultural knowledge; • The set-up of an HIKR, tailored to the needs of the end-users targeted, allowing for easy access to high quality content; • The communication and dissemination of innovative knowledge to the end-users; • Interoperable KRs in future Thematic Network efforts.
3	Exploring	• Development of an ontology for describing the Thematic Networks domain. • Development of an architecture and technical specifications for the EURAKNOS platform. • Development of the EURAKNOS digital platform comprising: - A data repository, built on the basis of Open and FAIR data principles, for storing agricultural knowledge related digital artifacts; - A search engine for executing queries on the knowledge stored; - A user interface enabling the querying and access of the knowledge available.
4	Widening	• Design and execution of communication, dissemination, exploitation, and networking activities to maximize the impact of EURAKNOS.

(*Contd.*)

Table 1. (*Contd.*)

#	*Phase*	*Involved activities*
4	Widening	• Development of a strategy for reaching out to a variety of audiences with special attention being paid to the EURAKNOS platform's end-users (e.g. farmers, foresters, advisors). • Synergies and connections with other Thematic Networks and Multi-Actor projects so as to: - Contribute to the maximization of the impact of the existing Thematic Networks; and - Guide and structure future efforts in Knowledge Reservoirs' development. • Pursuit of the sustainability and longevity of EURAKNOS by drawing links with training initiatives and educational programs at various sectors (e.g. academic, advisory, farm sectors) and levels (local, regional, national, and EU level). • Connection of EURAKNOS with EIP-AGRI to help develop links with the existing and future Thematic Networks, Operational Groups and Focus Groups.

Table 2. Tasks in the "Evaluating" phase of the EURAKNOS methodology, the research tools employed and the outcomes expected

#	*Task name*	*Task description*	*Research tools employed and outcomes*	
1	Data impact assessment	• Investigation and analysis of the methods used for data generation and collection. • Investigation of the tools and technologies used for data storage. • Investigation of the data formats and types, as well as the content of the knowledge sources and tools available. • Estimation of the produced knowledge completeness and identification of potential knowledge gaps.	Desktop study	Research of Thematic Network websites to get insights into content provision and information architecture. Evaluation of the variations identified with regard to the different sectoral themes addressed by the Thematic Networks (e.g. livestock, crop production, forestry).
			Face-to-face interviews	Interviews of Thematic Network representatives (e.g. coordinators, WP/Task leaders, technical infrastructure administrators).
2	Design of knowledge reservoirs	• Investigation and analysis of: - The design of existing KRs; - The database models used for structuring knowledge; - The websites and data store systems of well-known international organizations (e.g. FAO, OECD, EFSA, etc.) contributing to the provision and dissemination of agricultural knowledge. • Analysis based on the principles for open access to knowledge provided by initiatives such as the OpenAIRE project.	Desktop study	Research existing KRs for data and knowledge formats, and the means for making it available to end-users. Research has focused upon: (i) open source software solutions; (ii) format of the data and knowledge available; (iii) search engines used; (iv) user interaction and experience.
			Face-to-face interviews	Interviews of Thematic Network representatives involved in the design of the KRs (e.g. software/database engineers, technical infrastructure administrators).

(*Contd.*)

Table 2. (*Contd.*)

#	*Task name*	*Task description*	*Research tools employed and outcomes*	
3	Dissemination, communication and information strategies	• Review of the state-of-the-art in communication and dissemination practices and tools. • Special attention on how specific types of end-users, of particular interest to EURAKNOS (e.g. farmers, foresters and advisors), have been reached. • Special consideration of social media.	Desktop study	Quantitative and qualitative evaluation of the dissemination and communication tools and material by taking account of the particular needs of the different end-user types, age, geographic location of end-users, frequency of use, etc.
			Face-to-face interviews	Interviews of the Thematic Network representatives involved in the development and implementation of the Thematic Networks' communication, dissemination, and exploitation strategies.
4	Multi-Actor approach	• Focus on the Multi-Actor approach employed by the existing Thematic Networks. • Investigation of how Multi-Actor approaches have been conceptualized, put into motion and executed. Emphasis upon the post-execution phase. • Specific issues taken into consideration: - The number and type of actors (e.g. in terms of the sector and organization represented, the geographic location, etc.); - The methods of approaching different actor types; - The delivery of the input by the actors involved; - The benefits in the short-, medium- and long-term.	Desktop study	Quantitative and qualitative analysis of the information retrieved from the Thematic Network websites.
			Face-to-face interviews	Interviews of Thematic Network partners and actors (in particular farmers, foresters and advisors) to further refine the outcomes of the desktop study.

these two methodology phases, Table 3 provides the description of the tasks involved in the "Evaluating" phase and their links to tasks of the "Determining" phase.

Towards an Ontology of Thematic Networks

Knowledge Representation is dedicated to domain-specific knowledge modeling. Two widely adopted Knowledge Representation models are database models and ontologies. Despite the fact that they have some differences (with many of them being just historical), these two constructs share a strong semantic heritage by using different formalisms, such as logic, to build conceptual models of some subject matter. Ontologies are explicit specifications of a conceptualization (Gruber 1993) with their aim being to represent meaning rather than data. They are based on Description Logic, a low-level Knowledge Representation technique able to be documented through various markup languages such as RDF, RDFS, OWL and OWL-DL. Ontologies can provide solutions to data heterogeneity and interoperability problems and, for that reason, they have played a significant role in the Semantic Web's development (Grimm et. al. 2011).

In order to develop the database model for the data representation in EURAKNOS, an ontology of Thematic Networks has been decided to be developed. More specifically, NOTICE (oNtology Of ThematIC nEtworks) is an ontology providing a formal description of the Thematic Networks' domain. The initial work undertaken on this ontology's development is presented in this chapter. NOTICE helps identify key entities related to the concept of Thematic Networks, the key attributes of the entities, as well as the relations between them. The NOTICE ontology, created on the basis of the methodology proposed by Noy and McGuiness (2001), aims to facilitate the definition of metadata for the annotation of data in the EURAKNOS data store. Figure 1 illustrates the major steps involved in the ontology creation process.

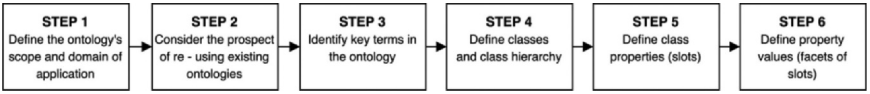

Figure 1. Methodology employed for building the NOTICE ontology (adapted from Noy and McGuiness 2001).

To the best of our knowledge, no systematic efforts regarding the creation of formal, explicit descriptions of the concept of Thematic Networks (based on some kind of Knowledge Representation structure) exist. Therefore, the NOTICE ontology has been built from scratch. The

Table 3. Tasks involved in the "Determining" phase of EURAKNOS methodology and their links to the "Evaluating" phase tasks

#	*Task name*	*Task description*	*Alignment with task of the "Evaluating" phase*
1	HIKR data best practices and methodologies	• Definition of the best methodologies and practices to generate, select, collect and store knowledge and data as the outputs of Thematic Networks. • Decisions about the data to be integrated in the EURAKNOS HIKR (decisions will be based on cultural, socio-economic, and relevant environmental aspects). • Recommendations for the development of a database model and data structuring specifications for the HIKR.	Data impact assessment
2	Design of HIKR	• Recommendations and guidelines for the EURAKNOS HIKR design by taking account of: - The needs of the end-users targeted; - Technology issues (e.g. ingestion of the data into the HIKR, data querying and access scenarios, data validation, user authentication and data security)	Design of knowledge reservoirs
3	HIKR dissemination, communication and exploitation strategies	• Identification and selection of the tools to communicate the project-related work, inform the end-users and disseminate the project's outcomes. • Exploitation of the tools identified with regard to best information, communication and dissemination channels to make the proposed innovation known to a wide range of potential end-users. • Proposal of a new format for practice abstracts.	Dissemination, communication and information strategies
4	HIKR Multi-Actor approach	• Identification and selection of the tools for Multi-Actor engagement and knowledge creation at different project phases, namely: conceptualization, initiation, execution and post-execution of the project. • Evaluation of the stakeholder needs with regard to the knowledge needed to be produced and stored in the HIKR.	Multi-Actor approach

identification of key terms, leading to the extraction of domain-related entities (which are, in turn, modeled as the ontology's classes and subclasses) and relations between them, has been based on the definition of the concept of Thematic Networks provided earlier. From that textual description, a number of statements containing references to domain-related entities, properties of those entities and relations between entities, have been extracted. In the statements presented below, the potential entities and entity properties are denoted with the use of bold characters and potential relations are highlighted with italicized characters. More specifically:

- A **Thematic Network** *is about* a **theme**.
- The **theme** (of a Thematic Network) *relates to* a **domain**.
- An **output** of a Thematic network *targets* **end – users**.
- An **output** *serves* a **purpose**.
- An **output** *addresses* a **topic**.
- An **output** *has* a **format**.
- A **topic** *relates to* a **theme** (namely, the theme of the Thematic Network).
- An **output** *is produced by* an **output creator**.
- An **output creator** *is distinguished into* **individual actor** and **organization**.
- An **individual actor** *belongs to* an **organization**.
- An **organization** *has* a specific **type of involvement** (namely, leading or participating organization) in the Thematic Network.

The above statements have helped to come up with a set of ontology classes, subclasses, and their properties all listed in Table 4 that follows.

The definition of the NOTICE classes, their subclasses, and a number of their properties has been based on the following schema.org types: *Thing*, *CreativeWork*, *DigitalDocument*, *MediaObject*, *AudioObject*, *ImageObject*, *VideoObject*, *PresentationDigitalDocument*, *SoftwareApplication*, *Person* and *Organization*. As mentioned in its official website[14], schema.org offers a vocabulary, able to be used with various encodings, for the definition of entities, relationships between entities and actions. The definition of the different kinds of digital artifacts, disseminated by Thematic Networks, has been based on the list of the media types (termed as Multipurpose Internet Mail Extensions or MIME types) provided by the Internet Assigned Numbers Authority[15]. The types included in this list are: *application*, *audio*, *example*, *font*, *image*, *message*, *model*, *text* and *video*. In order to propose an as inclusive as possible list of the Thematic Network—related output kinds, specifically tailored to the ontology's design needs, the {*application*, *audio*, *image*, *text*, *video*} MIME types sublist has been adopted and enriched

[14] https://schema.org/

[15] https://www.iana.org/

Table 4. The NOTICE ontology classes, subclasses and their properties

Class name	*Class definition*	*Class properties*	*Subclasses*	*Subclass properties*
Thematic Network	See definition provided in the section "The role of Thematic Networks as multipliers of Agriculture – related practical knowledge".	Acronym FullName Url	-	-
Theme	The subject a Thematic Network is about.	Title Domain	-	-
Output Creator	The actor or group of actors whose activity/-ies have led to the creation of a Thematic Network's output.	Name Description Country	IndividualActor A person who has been involved in the creation of a Thematic Network's output.	Email Type
			Organization An organization which has been involved in the creation of a Thematic Network's output.	AreaServed ContactPoint Involvement Type
Output	A digital artifact produced in the context of a Thematic Network.	Abstract Aggregate Rating dateCreatedinLanguage IsAccessibleForFree Keywords Purpose Size Title	AudioObject An audio file.	Duration Format PlayerType Transcript Type

Dataset A body of structured information describing some topic(s) of interest.[1]	Format Issn Measument Technique Type Variable Measured
Digital Document A document available in a digital format with its content being mostly text.	Format Type
Image Object An image file.	Height Format Type Width
Presentation Digital Document A file containing slides or used for a presentation.[2]	Format Type
Software Application Software designed to perform a group of coordinated functions, tasks, or activities for the benefit of the user.[3]	AvailableOnDevice Download Url Format MemoryRequirements OperatingSystem SoftwareRequirements SoftwareVersion StorageRequirements Type

(*Contd.*)

Table 4. The NOTICE ontology classes, subclasses and their properties

Class name	*Class definition*	*Class properties*	*Subclasses*	*Subclass properties*
			VideoObject A video file.	duration format playerType transcript type videoFrameSize videoQuality
Topic	The subject that an output of a Thematic Network is about.	Title	-	-
EndUser	The human individual that uses any computing-enabled device/appliance.[4]	Type	-	-

[1] Definition provided by schema.org: https://schema.org/Dataset
[2] Definition provided by schema.org: https://schema.org/PresentationDigitalDocument
[3] Definition provided by Wikipedia (https://en.wikipedia.org/wiki/Application_software)
[4] Definition provided by Technopedia: https://www.techopedia.com/definition/610/end-user

with the *presentation* and *dataset* types. Table 5 below lists all the kinds of potential Thematic Network outputs distinguished in NOTICE, which are modeled as subclasses of the "Output" class. It needs to be stressed that each output serves a specific purpose (namely, *access to data*, *best practice presentation*, *communication*, *decision support*, *dissemination*, *educational material*, *innovative practice presentation*, *training material*). This leads to another way of categorizing the different kinds of outputs, by taking account of the purpose-related types proposed. In other words, each subclass of the "Output" class has a property named "type" assigned to it, with its values relating to the purpose that the output serves.

Table 5. Subclasses of the "Output" class and respective values of the purpose-related "type" property

Subclasses of the "Output" class	*Values of "type" property*
AudioObject	{Advertising podcast, educational/training podcast, event capturing podcast, informational podcast, interview, on-demand seminar, tutorial}
Dataset	{Auditory data, crop-related data, geospatial data, graph-related data, imagery data, input-related data, network-related data, temporal data, textual data, video data, yield-related data}
DigitalDocument	{Article in conference proceedings, best practice guide, book, booklet, chapter in edited volume, deliverable report, factsheet, handbook/manual, journal article, milestone report, newsletter, practice abstract, press release, spreadsheet, review document, technical article, tutorial}
ImageObject	{Chart/graph, figure/image, info graphic}
Presentation-DigitalObject	{Demonstration, educational/training presentation, informational presentation, tutorial}
Software-application	{AI software, business software, data repository/database, decision support tool, educational/training software, Farm Management Information System (FMIS), game, scientific software, simulation}
VideoObject	{Advertising video, demonstration video, educational/training video, event capturing video, informational video, interview, testimonial, tutorial}

Distinguishing among different types of end-users is also important given that specific platform functionalities and solutions will be able to be provided given each user type's needs. On this basis, the following end-user types are defined: *adviser*, *bee keeper*, *consumer*, *distributor*, *entrepreneur*, *farmer*, *fisherman*, *forester*, *instructor*, *livestock breeder*, *policy maker*, *researcher*, *student*, *trader*, *trainer*, and *university professor*. As far as the relations between the ontology's classes are concerned, these are summarized in Table 6 that follows.

Table 6. The NOTICE ontology relations and the ontology classes participating in them

Referencing class	*Relation*	*Referenced class*
Thematic Network	*engages*	Output Creator
	is About	Theme
	is Coordinated By	Organization
Topic	*relates To*	Theme
Output	*addresses*	Topic
	is Produced By	Output Creator
	targets	End User
Individual Actor	*affiliated With*	Organization

All the ontology-related details are summarized in the diagrammatic representation of NOTICE illustrated in Fig. 2. However, it needs to be stressed again, that the work presented in NOTICE is preliminary. There are issues that need to be further taken care of such as the establishment of links between NOTICE and widely adopted, Agriculture-related ontologies and vocabularies for the needs of providing values to the domain and theme of Thematic Networks, as well as the topic of Thematic Network outputs.

Architecture of a Digital Platform for Agriculture-related Data

Architecture Overview

The EURAKNOS digital platform constitutes the technical implementation of the HIKR. The platform description is based on a three-layer architecture consisting of the data persistence layer, the application layer, and the presentation layer. The data persistence layer relates to the storage and management of the data. It is the layer where the data store is. The application layer is concerned with the actual logic of the platform. This is the layer where the EURAKNOS search engine resides. It handles data exchange between the presentation and data persistence layers and does not have to do with data persistence or how data is actually displayed to the end-user. However, it may well support multiple user interfaces (for instance, a web application and a mobile app). Finally, the presentation layer (in other words, the user interface) is what the end-users see. It is the part of the application that the end-users interact with. It displays the

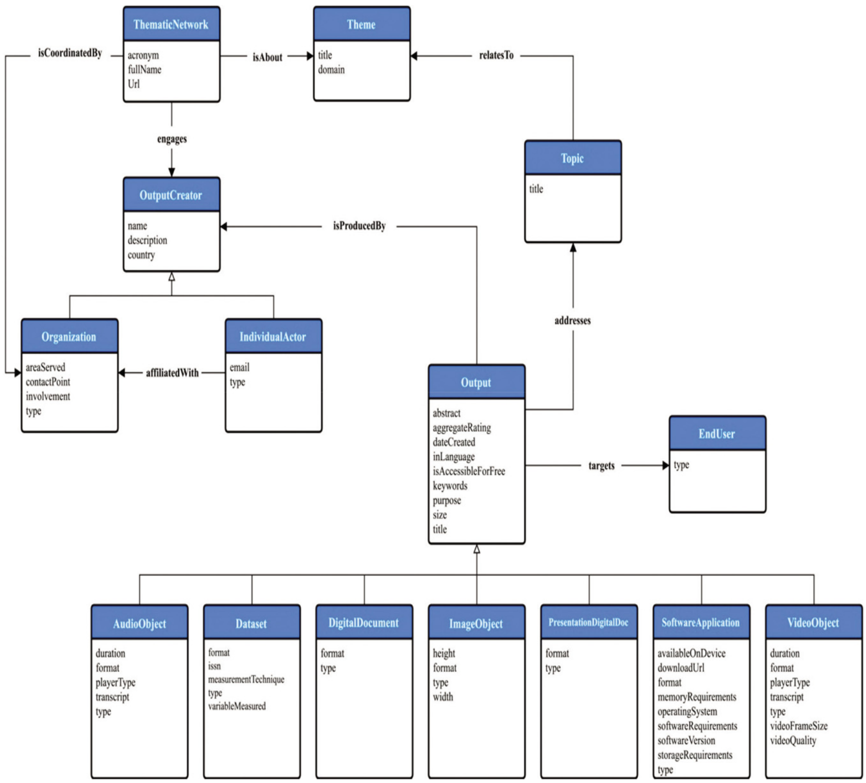

Figure 2. The NOTICE ontology.

data and receives requests from the end-users. This layer is not concerned with any logic and it just relies on the data. In order to make this stack of layers work, every layer needs to provide an Application Programming Interface (API) so that the other layers can communicate with it. So, for example, the EURAKNOS Search Engine may provide a REST Service and an OpenAPI (Swagger) specification upon which a user interface may be built. An overview of the EURAKNOS platform's architecture is provided in Fig. 3.

Issues Related to Data Collection and Ingestion into the EURAKNOS Data Store

EURAKNOS aims to collect and organize agricultural knowledge-related outputs available by existing Thematic Networks. Such outputs have so far been developed and managed in a decentralized and unstructured way. So, in order to make the data in the EURAKNOS data repository searchable and identifiable, indexes need to be built. Apart from that, a collection of large amounts of unstructured data will need to be based on

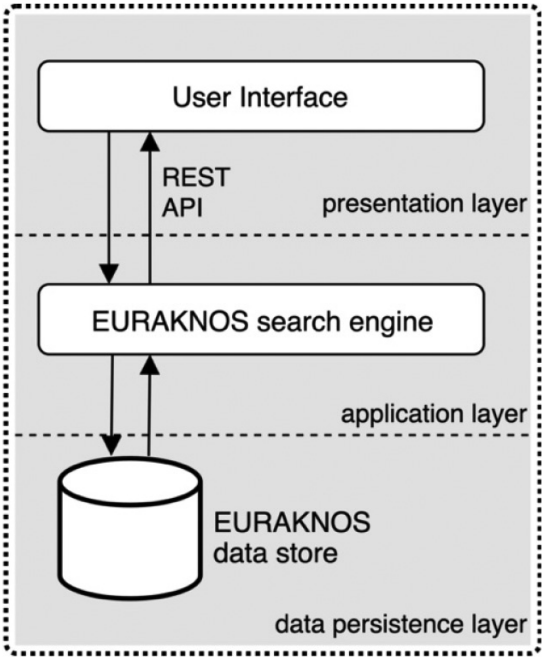

Figure 3. The architecture of the EURAKNOS digital platform.

some kind of automated process. Thematic Network outputs need to be aggregated into a single structured search index so that it can be searched. This is where web crawlers come into play. A web crawler is a program or automated script that browses the World Wide Web. It is usually employed to collect information, but it can also be used for automating maintenance tasks such as checking for dead links on a website or testing websites for errors.

A point to consider is that it is important not to "harm" the website that is being crawled. To this end, politeness policies are used to prevent the crawled website from being heavily affected. So, a polite web crawler should:

- respect the rules defined in the website's robots. txt file;
- avoid degrading the performance of the website it is crawling; and
- identify itself and its creator with contact information.

Humans are able to easily understand the content of a web page by just seeing it. However, it is not that easy for search engines to also interpret that same raw data. In order to help this interpretation process, a crawler collects metadata from the website. Search indexes can use metadata to interpret and link documents. The most obvious metadata provided by a web page are title-tags, the meta description, and keywords. It needs to be stressed that, quite often, web pages provide also descriptions for images. In the case of lack of metadata, web crawlers can resort to other ways of data interpretation.

The EURAKNOS Data Repository and Database Model

Thematic Networks compile information in many different ways. Therefore, it is hard to define a strict schema for data collection. One way to address this problem with an RDBMS is to employ the EAV/CR model (namely, Entity Attribute Value with Classes and Relationships). RDBMSs are highly consistent, but this comes at the cost of not being able to horizontally scale. NoSQL systems, on the other hand, allow for horizontal scalability, yet, at the cost of consistency. Horizontal scaling refers to the use of clusters of commodity hardware to store data, with each piece of hardware being responsible for the execution of processes such as look-ups and read/write operations on the data stored in it (Lake and Crowther 2013). This specific capacity of NoSQL data stores is pointed out in the definition of Cattell (2010) according to which NoSQL systems are "*designed to provide good horizontal scalability for simple read/write database operations distributed over many servers*" (p. 12). However, the advantages of NoSQL systems go beyond scalability-related issues. According to Vaish (2013), NoSQL data stores allow for schemaless data representation, which means that application developers can dynamically integrate changes in their design without needing to define a fixed data structure in advance. Schemaless data representation reduces the development time given that data access can be handled by application code rather than complex SQL queries. As a consequence, there is potential to develop applications able to efficiently respond to various workloads and deliver results very quickly.

The solution intended to be adopted for the deployment of the EURAKNOS data repository is MongoDB. MongoDB[16] belongs to the document-oriented category of NoSQL systems. By being a specific type of NoSQL data stores, document-oriented databases allow for the adoption of a dynamic and changeable schema, or no schema at all; a feature that makes them ideal for the storage of content changing over time (Vaish 2013). MongoDB is an open-source database management system providing high read and write throughput, as well as the ability for horizontal scaling and automatic failure recovery (Banker 2012). MongoDB has become popular mostly because of its capacity to efficiently represent and retrieve hierarchically structured information without a need to execute resource-intensive table joins (Banker 2012). Moreover,

[16] https://www.mongodb.com/

according to Banker (2012), MongoDB supports ad hoc queries[17] and data indexing. It also provides automatic data replication, which means distribution of data across nodes of a cluster with the aim to eradicate data loss due to hardware or network failure. Distribution of the data across nodes is internally handled by a mechanism called auto-sharding.

The model employed by MongoDB for data storage is called "document" and is based upon the JSON (i.e. JavaScript Object Notation) data exchange format. More specifically, MongoDB makes use of a binary representation of JSON called BSON (Lake and Crowther 2013). BSON documents can have a maximum size of 16 MBs; however, larger documents can also be stored with the help of the GridFS API. MongoDB's document can be better conceptualized if considered as the counterpart of a table row in an RDBMS. A document is made up of property names and values. Values can be of any BSON compatible data type (e.g. string, number, Boolean, date), as well as array, document (called embedded document) and array of documents. Thus, MongoDB can offer great flexibility for modeling complex data structures (Banker 2012). A "collection" is a group of documents and can be considered as the equivalent of a relational database's table. It needs to be highlighted that MongoDB allows for documents with different fields in the same collection.

Based on NOTICE, the database model proposed for the Agriculture-related data representation and storage is shown in Fig. 4.

Searching for Agriculture – Related Practical Knowledge

The primary aim of a search engine is to retrieve the most relevant documents to user queries, excluding other general content (Brin and Page 1998). To achieve this goal, a search engine usually instantiates crawlers that visit links, appearing in web pages, and download documents represented by those links. After that, an indexer parses the content of the documents (either textual or binary) and organizes it with the help of an index. Indexing is the method of storing data in index files and in a format helping the fast and efficient text searching. There are different index types each of which has strengths and weaknesses. The most popular index type is the inverted index, which is a data structure that stores mappings from terms to a set of documents. Apache Lucene[18] is the most popular open source library for document indexing. By also enabling query submissions, it allows the retrieval of query-matching results. The index can be stored in the file system or memory and can be

---

[17] According to Technopedia (https://www.techopedia.com/definition/30581/ad-hoc-query-sql-programming), an ad hoc query is "*a loosely typed command/query whose value depends upon some variable. Each time the command is executed, the result is different, depending on the value of the variable*."

[18] https://lucene.apache.org/

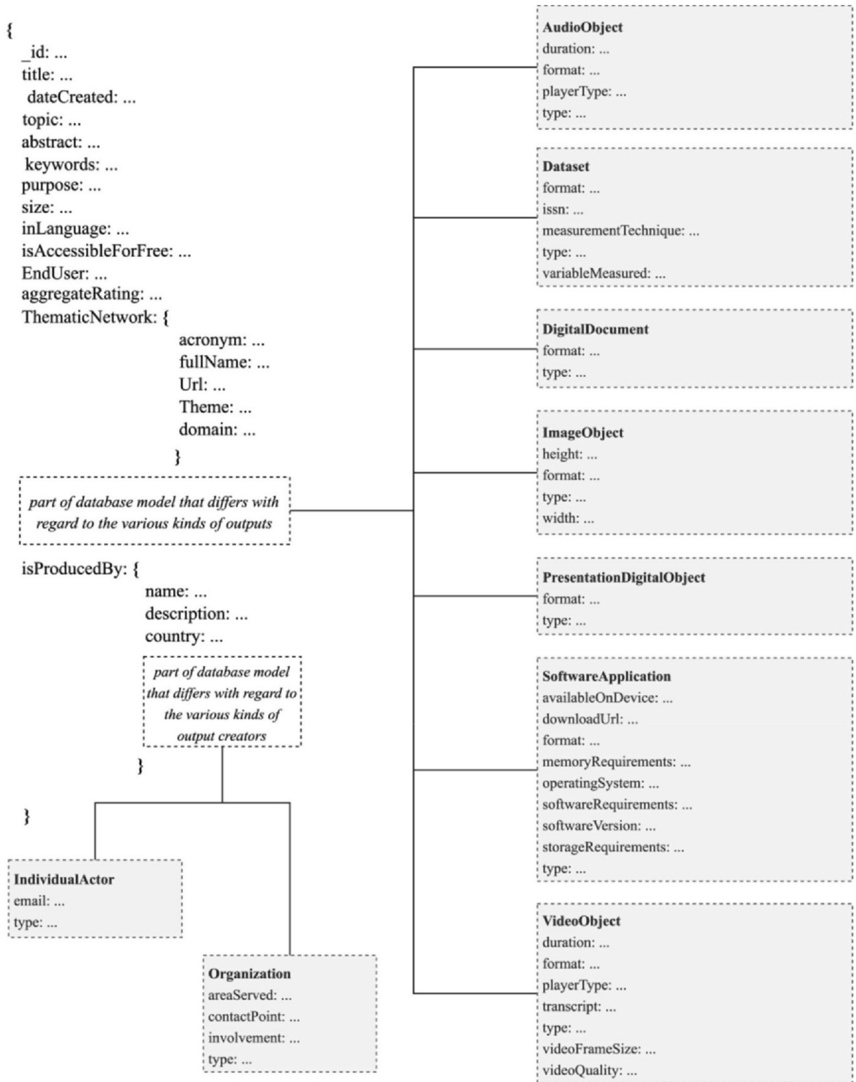

Figure 4. The EURAKNOS database model.

used to search the web, databases, and ontologies. Prior to storing data in index files, data is analyzed and parsed. Given Lucene's efficiency and precise search algorithms, it is widely adopted for being at the core of a search engine. The effectiveness of a search engine can be measured by two popular performance metrics, namely precision (i.e. the fraction of retrieved documents that are relevant) and recall (i.e. the fraction of relevant documents that have been retrieved).

There is no right or wrong choice with regard to search engine technologies. Choices depend on the application requirements and, based on them, the most efficient solution can be determined. Among a number of open source solutions available, the leading one, also qualifying for the deployment of the EURAKNOS search engine, is Elasticsearch[19]. Elasticsearch was first released in 2010 and is built upon the Apache Lucene library. It is stable and has a well-documented reference guide. Its core functionality has been enriched beyond simple text indexing and searching. Over the years, features such as faceting (currently known as aggregations), field collapsing, and language detection have been included. Elasticsearch exposes a REST API that allows, among others, to query and delete documents, as well as to create and manage indexes by using the HTTP GET, DELETE, POST and PUT methods. As far as the output format of the query results is concerned, Elasticsearch makes use of the JSON format. In terms of scalability, Elasticsearch uses a modern cluster configuration instead of the old-fashioned Master-Slave architecture.

Interacting with the EURAKNOS Digital Platform

By the time this chapter was authored, some preparatory work on the platform's back-end design had been done without, however, any final decisions having been made with regard to user interaction. Therefore, the aim of this paragraph is to highlight a number of interaction-related issues. Designing the interaction with the platform is heavily focused upon decisions for the facilitation of input provision from the end-user and the delivery of results. On one hand, the EURAKNOS repository will integrate digital artifacts of various formats and types waiting to be searched by the end-users targeted. On the other hand, the end-users come to interact with the system, with the expectation to get what is looked for, by bringing along presumptions and experiences from their interactions with other applications. These critical issues need to be carefully considered when designing input and output interfaces aiming to provide straightforward prompts and the right amount of clearly structured information.

As far as user input is concerned, the most commonly employed design pattern is that of a search box accompanied by a dedicated search button. Enabling users to submit free-text queries is the most commonly employed practice providing a great degree of freedom. In that case, the use of tools able to facilitate query syntax and submission needs to be investigated. According to Rosenfeld et al. (2015), such tools are spell checkers, which automatically correct misspelled query terms; stemming tools for retrieving results that contain terms sharing the same stem with terms in queries; Natural Language Processing tools having the capacity

[19] https://www.elastic.co/products/elasticsearch

to provide results after a syntactic analysis of the query; controlled vocabularies and thesauri enabling the search for results containing semantically similar terms; and autocompletion/autosuggestion tools able to significantly enhance the query development process.

Despite the intuitive nature of the free-text search and the advanced features that sophisticated add-ons can provide, the results yielded may not always be those expected. As a consequence, it is very important to investigate ways of further supporting the end users' searching endeavors. A solution to this direction is to provide the user with advanced search functionalities, able to lead to more specific queries, through the exploitation of logical operators. However, the use of this feature should be opted by the user (e.g. in cases where there are difficulties in retrieving the results required). So, in terms of design, access to this kind of functionality should be made available through separate interfaces. In this context, an option that is worth investigating is that of the interface segmentations called search zones (Rosenfeld et al. 2015), which offer search functionalities targeting at different user needs.

The design of search results display is also critical especially when the diversity of the Thematic Network outputs comes into play. A question requiring attention relates to the kind and amount of the results-related information needed to be displayed to the end-user. As Rosenfeld et al. (2015) point out, it is necessary to strike a balance between the display of representational information (for instance, the title, the creator and the date of creation of the result obtained) and descriptive information (in other words, information like the one mentioned above plus a content summary and maybe some keywords), aiming to help users not quite certain of the results they need. In addition to all the above, it should not be neglected that there also needs to be a focus on the order of the results display. Sorting and ranking are two popular options with the former appearing to be appropriate in cases where the user needs to make decisions or take actions, and the latter allowing for the establishment of a better understanding and learning purposes (Rosenfeld et al. 2015).

Conclusions and Further Steps

The aim of this chapter has been to present a systematic, ongoing effort for collecting Agriculture-related data and making them available. This process takes place within the context of EURAKNOS, a project that focuses on data disseminated by Thematic Networks. EURAKNOS is a Thematic Network itself aspiring to provide a single-point-of-access for Agriculture-related data available by the existing Thematic Networks. To this end, a fit-for-purpose methodology is implemented. Yet, there are many challenges that need to be addressed in order to conclude to a unified collection of Thematic Networks' outputs. For instance, there is a great diversity in the themes addressed by Thematic Networks, as well as in the types and formats of the outputs produced. In addition, there is no single technology used for the development of the existing Thematic Networks' Knowledge Reservoirs, and no consensus exists with regard to the data modeling approaches employed. As a consequence, the collation, storage, and provision of Thematic Network—related agricultural data, through a centralized repository, requires a robust and flexible data model. In this context, EURAKNOS makes a unique contribution by proposing the NOTICE ontology for the purpose of capturing, and formally describing, an extensive list of entities, entity properties, and relations tightly associated with any output (i.e. data object) disseminated by a Thematic Network. The ultimate goal of this Knowledge Representation structure is to help towards the identification and definition of a set of metadata for efficiently annotating the Thematic Network – related outputs in compliance with the FAIR data principles.

However, there are also some considerable limitations having mostly to do with EURAKNOS's scope. More specifically, EURAKNOS focuses solely on Thematic Networks and the data available by them. Thematic Networks are just a small fraction of Multi-Actor Projects and cannot be considered, in any case, as the sole sources of Agriculture-related knowledge. There are many more research-related efforts resulting in the creation of useful data that need to be considered. The NOTICE ontology is the first systematic attempt towards a formal, holistic description, and categorization, of agricultural data, but it cannot capture the full array of the characteristics and particularities of the data created in any Agriculture-related (research) effort. Therefore, by building upon the work conducted in EURAKNOS, further initiatives need to be taken towards this knowledge mapping direction[20].

In order to contribute to the realization of the main goal and objectives of EURAKNOS, there is a number of steps that need to be further taken. These steps relate to:

- The establishment of links between the NOTICE ontology and well-known Agriculture-related ontologies and vocabularies for the needs of providing values to the domain and theme of Thematic Networks, as well as the topic of the outputs of Thematic Networks.
- The facilitation of both human- and machine-enabled access to data through the achievement of compliance with the FAIR (i.e. "Findable",

[20] EUREKA (https://www.h2020eureka.eu/) is the follow-up project of EURAKNOS that extends upon its work through the adoption of a broader perspective (EUREKA focuses on the data objects produced by the entire set of Multi-Actor Projects. The EUREKA project receives funding from the EU's Horizon 2020 research and innovation programme under the grant agreement No 862790.

"Accessible", "Interoperable", and "Reusable") guiding principles and open data initiatives such as the OpenAIRE project[21].

- The development of the technical specifications necessary to assemble a robust system based on cutting-edge software solutions.
- The finalization of user interaction-related issues and the architecture of the information to be provided through the EURAKNOS digital platform, given the various end-user types targeted and their respective needs.

The work that has already been conducted together with the advancements expected, as the result of the steps to be taken next, will lead to the development of a body of knowledge of significant impact at both the small- and large-scale levels. At the small-scale level, EURAKNOS will frame the context for the future Thematic Networks by proposing and delivering a complete set of recommendations about their conceptual design and technology-related choices. By being exploited as it is or extended upon a fit-for-purpose basis, the NOTICE ontology is going to be a point-of-reference for the future Thematic Networks' conceptual design. In addition to that, the EURAKNOS platform's architecture and technical design will play a catalytic role to the creation and deployment of robust technological infrastructures capable of accurately serving the Thematic Network community members' needs. At the large-scale and long-term level, the achievements envisioned in EURAKNOS can also have implications for future policies. The benefits than can be reaped from the systematic exploitation of the unified EURAKNOS framework for Agriculture-related data collection and delivery may gradually lead to the establishment of standards and their embracement by policy-related authorities at the regional, national, and, probably, the international level.

References

Agricultural European Innovation Partnership. 2016. Thematic Networks under Horizon 2020: Compiling knowledge ready for practice. Retrieved from https://ec.europa.eu/eip/agriculture/sites/agri-eip/files/eip-agri_brochure_thematic_networks_2016_en_web.pdf

Agricultural European Innovation Partnership. 2017. Horizon 2020 multi-actor projects. Retrieved from https://ec.europa.eu/eip/agriculture/sites/agri-eip/files/eip-agri_brochure_multi-actor_ projects_2017_en_web.pdf

Banker, K. 2012. MongoDB in action. Manning Publications Co. New York. USA.

Brin, S. and L. Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *In*: Seventh International World-Wide Web Conference (WWW 1998), Brisbane, Australia.

[21] https://www.openaire.eu/

Cattell, R. 2010. Scalable SQL and NoSQL data stores. Acm Sigmod Record 39(4), 12–27.

da Silva, A.R., A. Silva and A. Ribeiro. 2015. Data modeling and data analytics: A survey from a big data perspective. Journal of Software Engineering and Applications 8(12): 720–726.

Grimm, S., A. Abecker, J. Völker and R. Studer. 2011. Ontologies and the Semantic Web. *In*: Domingue, J., D. Fensel and J.A. Hendler. [eds.]. Handbook of Semantic Web Technologies. Springer, Berlin, Germany.

Gruber, T.R. 1993. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93–04, Knowledge Systems Laboratory, Stanford University.

Lake, P. and P. Crowther. 2013. Concise Guide to Databases. Springer, London. UK.

Noy, N.F. and D.L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. Retrieved from https://protege.stanford.edu/publications/ontology_development/ ontology101.pdf

Rosenfeld, L., P. Morville, and J. Arango. 2015. Information Architecture: For the Web and Beyond. O'Reilly Media, Inc., Sebastopol, CA.

Umanath, N.S. and R. W. Scamell. 2014. Data Modeling and Database Design. Delmar Learning, New York, USA.

Vaish, G. 2013. Getting Started with NoSQL. Packt Publishing Ltd, Birmingham. UK.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne and J. Bouwman. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1), 1-9.