Deliverable 4.3
e-KRP documentation and guide

EURAKNOS

Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Task 4.2: Design and development of a prototype e-KRP

Summary

Task description: Based on the consultations and the work performed in WPs 2 and 3, a prototype e-KRP will be designed and developed to allow access to a centralised data repository, where data from the KRs of existing TNs will be stored. For the data collection, a data retrieval application was designed and developed (i.e., a web crawler) that is responsible for communicating with the KRs of existing TNs. Data from a sample of eight (8) appropriately selected TNs are integrated into the repository of the e-KRP by drawing upon the common HIKR format. The work presented in this document relates to the assembly of the input for the e-KRP, harvested from the KRs of the sample of the eight TNs, the identification and analysis of user needs, the development-related work that has taken place both for the back- and front-end of the platform, the testing and evaluation of the e-KRP, and the proposal of a sustainability plan. A user guide of the e-KRP is also provided, whereas the appropriate links with the work in progress in EUREKA (i.e., the follow-up project of EURAKNOS) are drawn.

Results: The key contributions that were feasible to be made as a result of the work in Task 4.2 are the following: (i) the e-KRP system's architecture; (ii) the set of the data objects integrated into the e-KRP repository; (iii) the EURAKNOS metadata set; (iv) the prototype e-KRP and; (v) a preliminary platform-related sustainability plan including a framework of guidelines and recommendations for future TNs about how to better protect their intellectual outputs.

Deliverable Number	Work Package
D4.3	WP4
Lead Beneficiary	**Deliverable Author(S)**
Agricultural University of Athens (AUA)	Hercules Panoutsopoulos (AUA)

Beneficiaries	Deliverable Co-Author (S)
Leap Forward Group (LFG) AKI Nonprofit KFT (AKI) Innovation for Agriculture (IfA)	Sofia Mouseti (AUA) Borja Espejo Garcia (AUA) Spyros Fountas (AUA) Laurens Van der Cruyssen (LFG) Szilveszter Palotay (AKI) Szabolcs Vago (AKI)
Planned Delivery Date	**Actual Delivery Date**
31/03/2021	29/03/2021

Type of Deliverable	R	Document, report (excluding periodic and final reports)	☑
	DEC	Websites, patents filing, press & media actions, videos	
	E	Ethics	

Dissemination Level	PU	Public	☑
	CO	Confidential, only for members of the consortium	

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Table of Contents

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

List of figures

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

List of tables

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

List of abbreviations and acronyms

AKIS – Agricultural Knowledge and Innovation Systems

EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs: towards a European Agricultural Knowledge Innovation Open Source System

EUREKA – European Knowledge Repository for Best Agricultural Practices

e-KRP – electronic Knowledge Reservoir Platform

EIP – AGRI – European Innovation Partnership for Agricultural productivity and Sustainability

EU – European Union

GA – Grant Agreement

GUI – Graphical User Interface

HIKR – High Impact Knowledge Reservoir

KR – Knowledge Reservoir

MAP – Multi-Actor Project

OG – Operational Group

TN – Thematic Network

UX – User Experience

WP – Work Package

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

1. Introduction

The current report documents the work implemented as part of Task 4.2 ("Design and development of a prototype e-KRP") in regard to the development of the EURAKNOS electronic Knowledge Reservoir Platform (e-KRP); i.e., a digital platform enabling open access to digital content (data objects) available from the Knowledge Reservoirs (KRs) of Thematic Networks (TNs). In order to showcase the technical feasibility of such an endeavor, all the activities carried out in EURAKNOS at the technical level have focussed on the development of a prototype digital platform (the prototype e-KRP) by taking account of specific, selected case studies (i.e., a subset of the existing pool of TNs). The aim has been to create a technological infrastructure having the capacity to deliver the content of high impact and importance to the end-users (i.e., farmers, foresters, and advisors), which has the potential to scale-up to an EU-wide, open-source, agricultural innovation and knowledge system linked with the EIP-AGRI Multi-Actor Projects (MAPs), Operational Groups (OGs), and other similar initiatives at the international and national levels (namely, a user-friendly database connecting all KRs in agricultural and forestry practice available at several levels in Europe).

Based on the context outlined above and to efficiently illustrate all the activities executed in Task 4.2, the report begins with a brief description of the broader context in which the task-related activities have been situated, the objectives that have been aimed, the involvement and contribution of project partners (Section 2), and an outline of the methodology (Section 3). Sections 5 ("User needs analysis and specification of requirements") to 8 ("Testing and evalustion of the EURAKNOS e-KRP") provide the details of the execution of the task-related methodology. Section 4 ("The EURAKNOS e-KRP system architecture") provides an outline of the architecture of the e-KRP system, Section 9 ("Sustainability of the EURAKNOS e-KRP") is a follow-up to the discussion about e-KRP-related sustainability issues, which has started in Deliverable 4.1[1], and Section 10 ("Recommendations for future Thematic Networks") serves the purpose of providing several recommendations to future TNs based on the lessons learned from the e-KRP development process. An e-KRP user guide is made available in Section 11 and the key points raised in this report, together with accounts of how the work in EURAKNOS should be followed up (and is being followed up) in EUREKA[2], are summarised in Section 12 ("Conclusions and Discussion").

1.1. Context

As mentioned above, the objective lying at the core of Task 4.2 has been the creation of the EURAKNOS prototype e-KRP. Before proceeding to the description of the broader context within which the task-related work and activities have taken place, it is important to start by shortly explaining some project- (and task-) related terms used in the present report. More specifically, "e-KRP" (electronic Knowledge Reservoir Platform) is the term used to refer to the system built in the project; that is, a digital platform enabling access to the digital content available from the system's Knowledge Reservoir. A "Knowledge Reservoir" (KR) is conceptualised as a collection of best materials, practices, instruments, methodologies and tools, able to contribute to the uptake and use of innovative solutions for sustainable agriculture and forestry. In a strictly technical sense, a KR is a digital repository where all data objects (conveying this content and information) are stored. The rationale behind the creation of the EURAKNOS prototype e-KRP system has been to investigate the (technical) feasibility and added value of a centralised platform enabling access to digital content of importance to end-users. In other words, to materialise a High Impact Knowledge Reservoir (HIKR); i.e., a KR aiming to maximise impact by making available the content, structures and methodologies appearing to best fit to urgent user needs, together with the channels and tools to reach the end-users (farmers, foresters, and advisors). Within the context of EURAKNOS, the HIKR is realised by the data repository residing at the core of our e-KRP system.

[1] Gernert M., et al. (2019). Report on added value and feasibility of e-KRP. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 4.1 – WP4.

[2] https://www.h2020eureka.eu/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

What is made evident from the definitions provided above is that the meaning of the terms "KR" and "HIKR" is not just technical. These terms have been initially coined with a broader conceptualisation in mind relating to content/information collections. Based on that, they are used throughout this report to denote any of these two meanings.

The context in which Task 4.2 has been situated is shown in Figure 1 below. As made evident from this diagrammatic representation, Task 4.2 is the final task in a pipeline of tasks implemented in WPs 2, 3 and 4. Figure 1 makes explicit all the dependencies between Task 4.2 and tasks from WP4 or other WPs by also showing the output generated from/input received by each one of those.

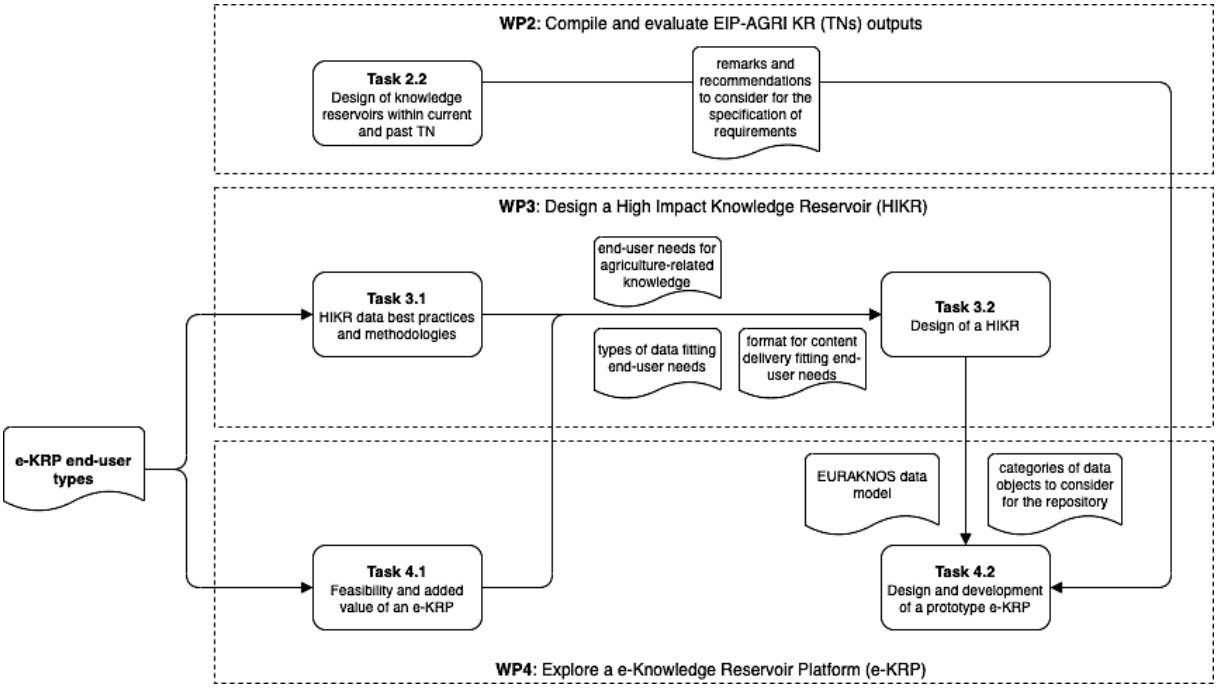


Figure 1. The context of Task 4.2

The methodology that has been followed in Task 4.2 is outlined in Section 3 ("Methodology") and then presented in detail in Sections 5 to 8.

1.2. Objectives

The main aim of Task 4.2 has been to create a digital platform (the e-KRP), collect data objects (namely, digital content and information in a variety of formats) from the KRs of a sample of TNs, and integrate them into the e-KRP repository as a proof-of-concept for the (technical) feasibility of this endeavour. In order to achieve this goal, a number of more specific objectives of a narrower scope have had to be set and achieved. These objectives related to:

- the identification of the needs of the end-users and specification of requirements;
- the design of the e-KRP system's architecture;
- the back- and front-end development of the e-KRP based on the identified requirements and the e-KRP system's architecture;
- the harvesting of digital content from the KRs of the sample TNs and the creation of a pool of data objects to consider as potential input for the repository of the e-KRP;
- the selection of digital content and information of importance to the end-users to integrate it into the repository;
- the annotation of the selected data objects by assigning the appropriate metadata values;

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- the integration of the selected data objects into the repository of the e-KRP; and
- the acquisition of insights into e-KRP-related usability and user-friendliness issues.

The objectives listed above relate directly to specific phases of the adopted methodology and concrete activities, calling for various kinds of expertise coming from the project partners, have been executed towards achieving them. The contribution of the project partners involved in the task and the expertise they have brought onboard are outlined in Section 2.

2. Task members

All the task-related activities have been coordinated by AUA (WP and task leader). The partners that have been involved in Task 4.2 are AKI, LFG, IfA, and the coordinators of the TNs whose KRs have been considered as sources of data objects for the EURAKNOS e-KRP repository. It needs to be stressed that the University of Maastricht (MU) has also participated in some of the activities implemented in Task 4.2, because of its WP leader role in the EUREKA project (i.e., it leads WP3: "Knowledge selection and developing a blueprint for the FarmBook") which takes on and extends the design- and development-related work done in EURAKNOS.

The involvement of each project partner in Task 4.2 and the expertise brought onboard are outlined in Table 1 below. Each partner's contribution is mentioned with regard to the phase of the methodology that its contribution relates to[3].

Table 1. Involvement of EURAKNOS partners in the activities related to Task 4.2 and the expertise brought onboard

Partner	Involvement in the task	Expertise brought to the task
AUA	• User needs analysis and specification of requirements. • Design of the e-KRP system architecture. • Set-up of the sample of TNs. • Assembly of the input for the e-KRP repository. • Integration of data objects into the e-KRP repository. • Back-end development. • e-KRP testing and evaluation.	• Software engineering/digital platform development/back-end development. • Data modelling. • Semantic web technologies. • Data (pre-)processing and analysis.
LFG	• User needs analysis and specification of requirements. • Design of the e-KRP system architecture. • Back-end development. • Front-end development. • e-KRP testing and evaluation.	• User Experience (UX) design and engineering. • Software engineering/front-end development/back-end development. • Software product testing.
AKI	• User needs analysis and specification of requirements. • Design of the e-KRP system architecture. • Harvesting of data objects from the KRs of TNs. • Assembly of the input for the e-KRP repository.	• Software engineering/back-end development/ front-end development. • Data (pre-)processing and analysis.
IfA	• User needs analysis and specification of requirements.	• Communication and dissemination expert. It has brought the end-users perspective and provided insights into the process of needs identification and requirement specification for the e-KRP.
TN coordinators	• Assembly of the input for the e-KRP repository.	• Knowledge of all the details related to the data objects created in their TNs.
MU	• Design of the e-KRP system architecture. • Integration of data objects into the e-KRP repository.	• Software engineering/back-end development. • Data modelling. • Semantic web technologies.

[3] An overview of the methodology adopted and implemented in Task 4.2 together with the phases it involved are described in Section 3 ("Methodology") and explicitly described from Section 5 ("Shaping up the repository of the EURAKNOS e-KRP") to Section 7 ("Testing the EURAKNOS e-KRP").

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

	• Back-end development.	

In addition to the above, IfA has also helped in shaping the EURAKNOS e-KRP's user manual, as well as the recommendations for future TNs, through the supply of insights and feedback.

3. Methodology

The methodology that has been followed in Task 4.2 is shown in Figure 2. What is made evident in the specific diagrammatic representation is the range and divergence of the activities that were needed to be executed to accomplish the task's main goal and the associated specific objectives. These activities ranged from the selection of the TNs that made up the sample of the data object sources for the e-KRP repository, to the analysis of user needs and the specification of requirements, the design of the e-KRP system's architecture and the back- and front-end development, as well as the assembly of the input for the repository and the integration of content into it. In the cases of the "Set-up of the sample of TNs", "Harvesting of data objects from the KRs of the sample TNs", and "Assembly of the input for the e-KRP repository" phases, there has been a breakdown of the involved activities in more detailed steps aiming to facilitate the fulfillment of the relevant task objectives. Furthermore, there has also been a parallel execution of activities. This has been indeed the case with the "Harvesting of data objects from the KRs of the sample TNs" and the "Set-up of the sample of TNs" phases, as well as with the "Back-end development" and "Front-end development" phases.

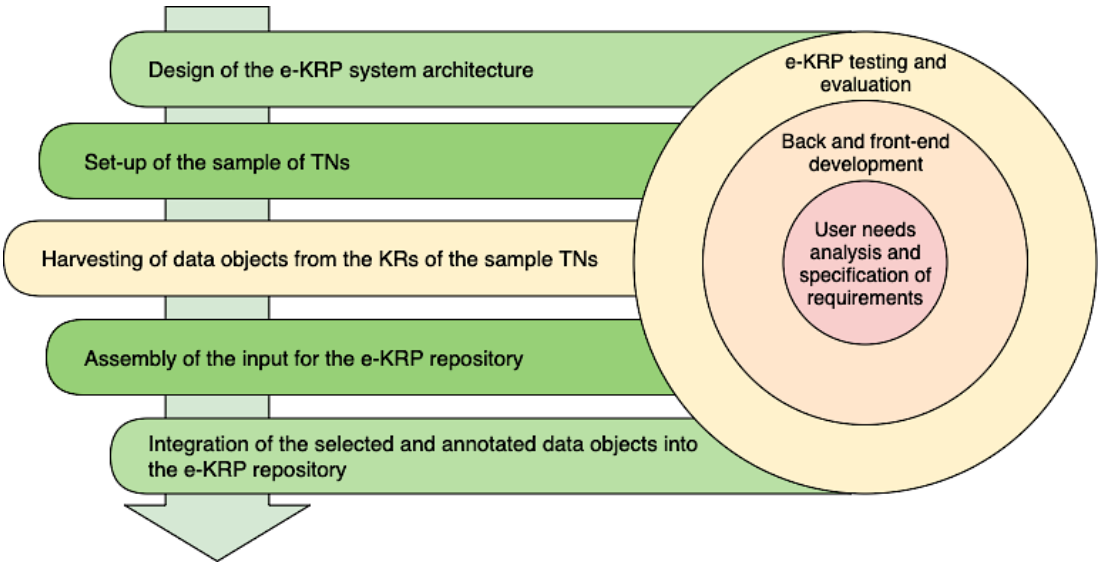


Figure 2. The methodology adopted in Task 4.2

The architecture of the e-KRP is presented in Section 4. The process and outcomes of the methodology phase related to the user needs analysis and the specification of requirements are described in Section 5. Section 6 ("Shaping up the repository of the EURAKNOS e-KRP") is about the activities implemented for: (i) setting-up the sample of TNs to consider as sources of data objects for the repository; (ii) the harvesting of data objects from the KRs of the sample TNs; (iii) the assembly of the input for the e-KRP repository (involving the cleaning, selection, and annotation of data objects); and (iv) the integration of the data objects into the repository. Details about the back- and front-end development are given in Section 7. Finally, the results and insights from the e-KRP's testing and evaluation are made available in Section 8 ("Testing and evaluatimg the EURAKNOS e-KRP") of the report.

Table 2 below presents the alignment of the task objectives with the phases of our methodology.

Table 2. Alignment of task-related objectives with phases of the methodology

#	Task-related objective	Phase of the methodology

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

1	Identification of the needs of the end-users and specification of requirements.	User needs analysis and specification of requirements.
2	Design of the e-KRP system's architecture.	Design of the e-KRP system architecture.
3	Back- and front-end development of the e-KRP based on the identified requiremens and the e-KRP system's architecture.	Back- and front-end development.
4	Harvesting of digital content from the KRs of the sample TNs and creation of a pool of data objects to consider as potential input for the repository of the e-KRP.	• Set-up of the sample of TNs. • Harvesting of data objects from the KRs of the sample TNs.
5	Selection of digital content and information of importance to the end-users to integrate it into the repository.	Assembly of the input for the e-KRP repository. • Data cleaning. • Selection of data objects for the repository.
6	Annotation of the selected data objects by assigning the appropriate metadata values.	Assembly of the input for the e-KRP repository. • Data object annotation.
7	Integration of the selected data objects into the repository of the e-KRP.	Integration of the selected and annotated data objects into the e-KRP repository
8	Acquisition of insights into e-KRP-related usability and user-friendliness issues.	e-KRP testing and evaluation.

4. The EURAKNOS e-KRP system architecture

Before delving into the details of our methodology deployment, it is important to set off by outlining the architecture of the EURAKNOS e-KRP system. This will help to better comprehend the activities and the steps of the work done in Task 4.2 both about the collection of the input for the repository, as well as for the needs of the technical (i.e., back-end and front-end) development. The architecture of the e-KRP is shown in Figure 3. It has been developed by drawing upon a stack of technologies (described in detail in the "Technology framework" section of the Deliverable 3.2 report[4]) and a philosophy that was decided and adopted to facilitate the follow-up of the work in the EURAKNOS project by its sibling project, EUREKA[5].

As reported in Deliverable 3.2 (Van Hal et al., 2020), the solution utilised as the data repository for the TN data objects is DSpace[6]. Based on the architecture presented below, data objects harvested from the KRs of a sample of TNs (Subsection 5.1 provides the details about the assembly of this TN sample) have been ingested into the system by using an ETL (Extract, Transform and Load) service described in Subsection 5.4 ("Integration of data objects into the e-KRP repository"). Storage for DSpace takes place in Solr[7] for the metadata of the data objects and a relational database (namely, PostgreSQL[8]) for all other information. Metadata is also stored in Fuseki, a triple store used for the storage of RDF triples encoding semantics with the data. The e-KRP backend is implemented as a microservice architecture. Each service sports a REST API[9] for communication.

Based on the technology framework presented and explained in Deliverable 3.2, the initial plan was to use OpenLink Virtuoso[10] as the triple store for metadata storage by circumventing Fuseki (i.e., the triple

[4] Van Hal, D. et al. (2020). Report on HIKR Design. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 3.2 – WP3.

[5] https://www.h2020eureka.eu/

[6] https://duraspace.org/dspace/

[7] Apache Solr and its features are described in detail in Deliverable 3.2 (Van Hal et al., 2020). For more details, the reader may also refer to Apache Solr's official website (https://solr.apache.org/).

[8] https://www.postgresql.org/

[9] A discussion about what an API is and the different types of existing APIs (including REST APIs) is provided in Deliverable 4.2 ("API documentation") (Mouseti et al., 2020).

[10] https://virtuoso.openlinksw.com/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

store built in DSpace). However, as mentioned in Section 4.3.4 of Deliverable 3.2 ("EURAKNOS e-KRP architecture design and challenges"), trying to utilise OpenLink Virtuoso instead of Fuseki could entail technical difficulties that had not been anticipated in the planning phase. The reason for initially opting for OpenLink Virtuoso was that it constitutes a solution used in many European Research & Innovation projects. While DSpace should support OpenLink Virtuoso according to its documentation, all efforts to utilise OpenLink Virtuoso for the RDFizer, which automatically converts metadata to semantically enhanced data, have proved fruitless. After communication with the DSpace developers, this technical issue was reported in the appropriate issue management to fix existing bugs in DSpace. Yet, no solution has become possible to be found. The reporting of this technical issue is available at https://jira.lyrasis.org/browse/DS-4520. Consequently, the triple store solution finally adopted was that of Fuseki (i.e., the triple store shipped with the DSpace framework). Details about Fuseki are available at https://jena.apache.org/documentation/fuseki2/.

The API layer is intended to function as a gateway API, serving as the only communication layer to the backend services. Any generic backend functionalities such as user management and security are implemented using standard (headless) Drupal and Contenta functionalities. The e-KRP may be leveraged either by directly accessing the EURAKNOS API[11] or by making use of the web interface (web GUI[12]).

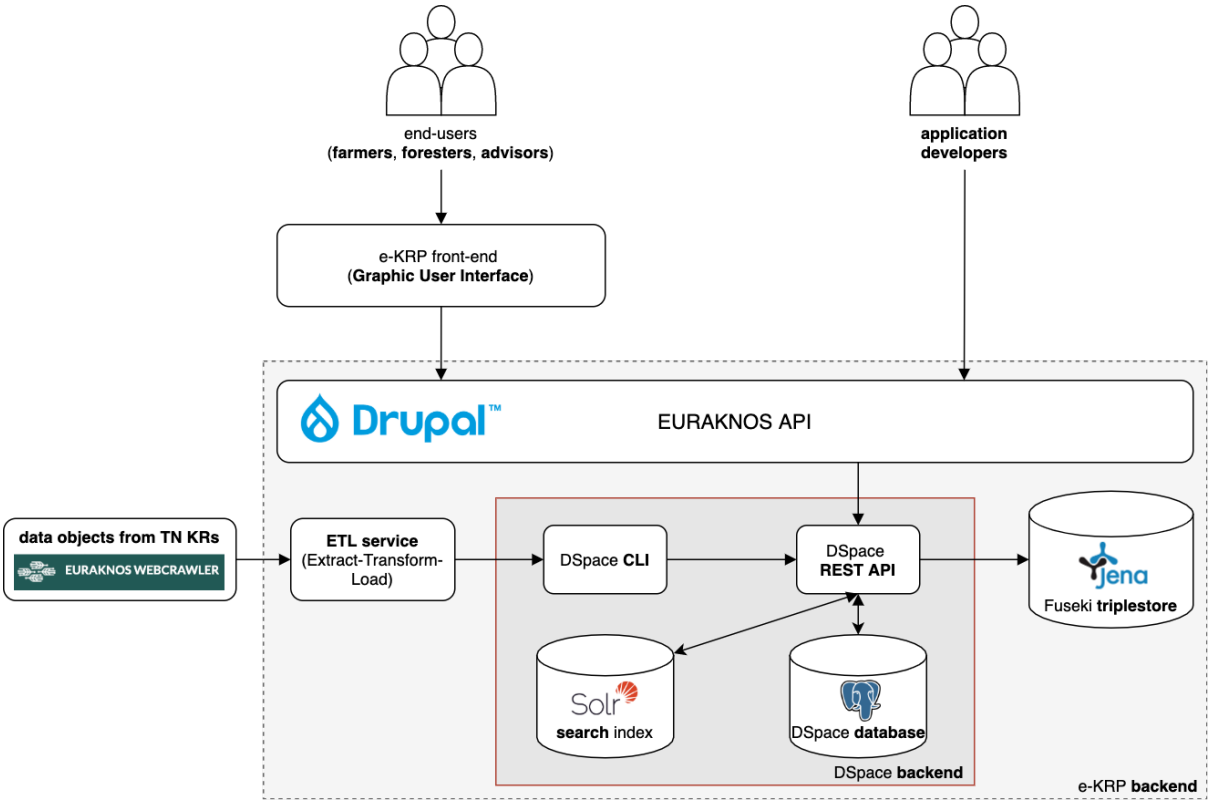


Figure 3. The architecture of the EURAKNOS e-KRP system

As a side note, it needs to be highlighted that the features of DSpace together with the flexibility of the ETL process, which has been implemented for the integration of the TN data objects into the e-KRP repository, have allowed for avoiding the use of Hadoop[13] (mentioned in the GA as the infrastructure

[11] The EURAKNOS API is reported in Deliverable 4.2 ("API documentation") (Mouseti et al., 2020).

[12] GUI is the abbreviation used for the term "Graphical User Interface".

[13] https://hadoop.apache.org/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

for the temporary storage of data objects before their actual integration into the e-KRP repository)[14]. This flexibility, together with the features of the EURAKNOS API (documented in Deliverable 4.2[15]), enable the e-KRP to be considered as a reference implementation for similar systems and data retrieval APIs developed in the context of efforts in Central Eastern Europe (CEE) countries.

5. User needs analysis and specification of requirements

The identification of the needs of end-users for the purpose of specifying requirements for the e-KRP, took place in the context of dedicated sessions and appropriately designed exercises, which were held: (i) in the EURAKNOS consortium meeting in Athens (17 and 18 June, 2019); and (ii) the EURAKNOS Paris workshop (11 to 13 December, 2019). The sessions implemented in those two events allowed to collect input from potential e-KRP users. The compilation and processing of the collected information assisted in acquiring useful insights into the actual needs and preferences of the end-users and thus, identify and define requirements for the e-KRP's UI and UX design and engineering (functional requirements of the e-KRP system). In addition to the above, the concluding remarks and recommendations reported in Deliverable 2.2 ("Report on format and design of knowledge reservoirs of Thematic Networks") (Marois et al., 2019) have also provided useful input.

5.1. Input collected from the wireframe-based exercise held in the Athens consortium meeting

A wireframe (a.k.a. a page schematic or screen blueprint) is a visual, diagrammatic representation of the structure of a webpage or a software application's interface (Brown, 2011). To initiate a discussion on the look and feel of the e-KRP and collect input about the needs of the end-users in regard to specific UI- and UX-related aspects, a wireframe-based exercise was held in the consortium meeting that took place in Athens, in June 2019.

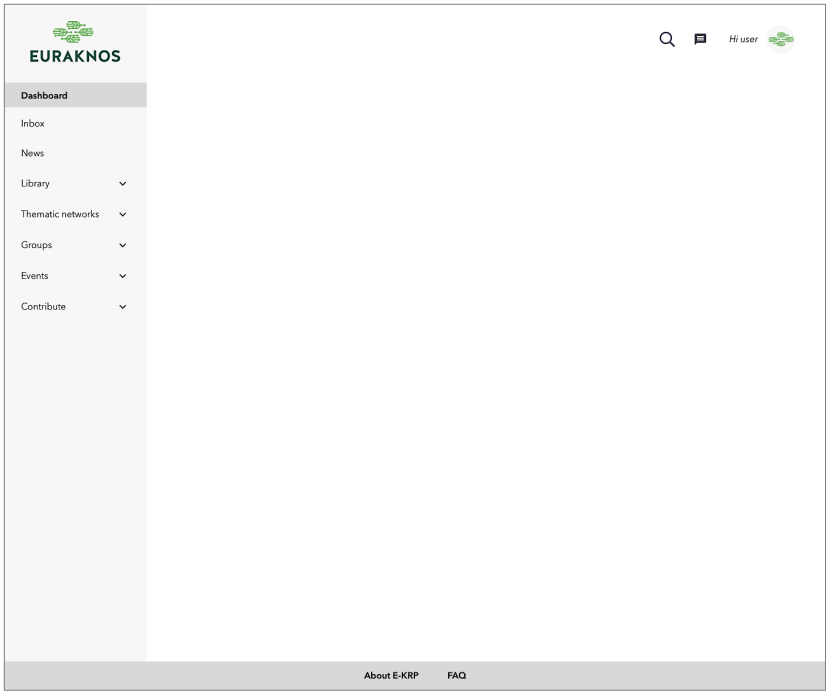


Figure 4. The navigational structures initially proposed for the e-KRP through the design of the respective wireframe

[14] The Hadoop framework could not serve the purpose of the e-KRP repository. It provides a distributed file system enabling the storage of large data blocks, but does not provide the advanced data storage- and curation-related functionalities sought for the e-KRP repository.

[15] Mouseti, M. et al. (2020). API documentation. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 4.2 – WP4.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

The exercise participants (28 consortium members) were assigned to three groups and asked to review the wireframes. Moreover, they were asked to provide comments and also suggest alternatives. Given the multi-actor synthesis of the project's consortium, an adequate representation of various end-user categories was achieved. The following aspects were investigated:

- **Navigation**
  The solution proposed for the e-KRP's basic navigational structure was that of a sidebar (see Figure 4 above), because of its flexibility and potential to easily expand and offer more navigation options. A menu on the top of the page was also suggested. Yet, the horizontal space required for the display of the menu items may impose significant layout changes in case a new option should be added.

- **Dashboard**
  In general, a dashboard is a type of GUI[16] that often provides at-a-glance views of key performance indicators (KPIs), progress reports, or data visualisations[17]. For the user needs analysis exercise, four different wireframe versions of the e-KRP's landing page dashboard were used to collect feedback in user preferences :
  - **Dashboard version #1** (see Figure 5 below): A classic dashboard of categories, where the user is able to view updates related to each one. In the case of a few category updates, this will become easily noticeable. However, many updates in a specific category will flood that category.
  - **Dashboard version #2** (see Figure 6 below): A dashboard version inspired by the widely-known Facebook feed. It is quite a dynamic dashboard, where updates can be viewed by scrolling down.
  - **Dashboard version #3** (see Figure 7 below): The third dashboard version was more dynamic in its layout. It resembles the layouts of online newspapers and gives the impression that there is always something new to discover.
  - **Dashboard version #4** (see Figure 8 below): A less dynamic dashboard focussing upon text-based content appropriately supported by visual aids.

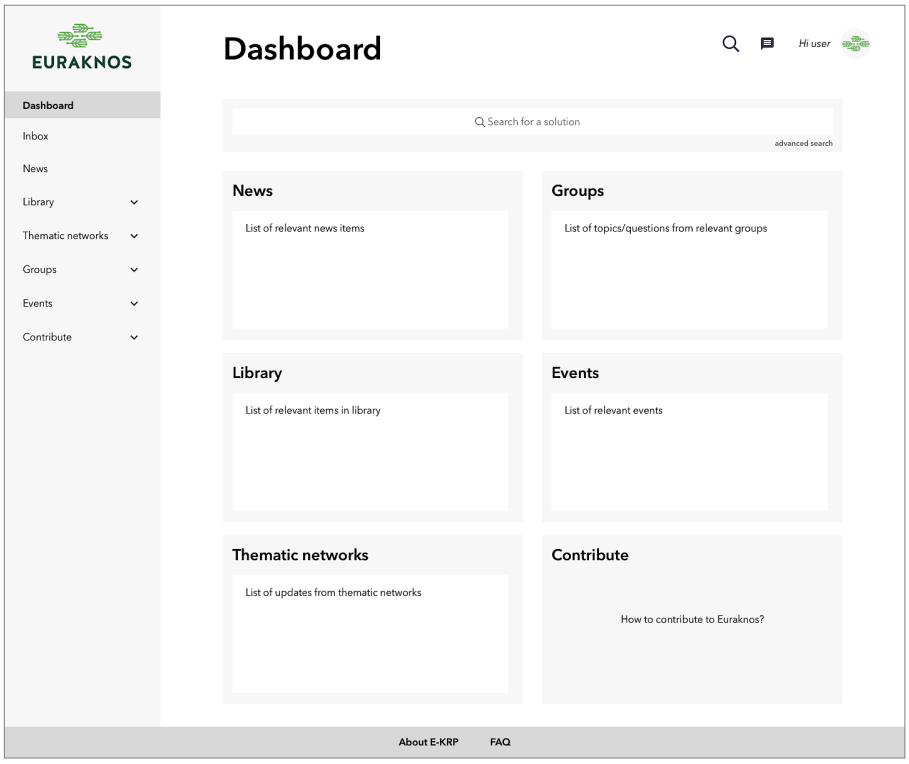


Figure 5. Version 1 of the e-KRP's landing page dashboard

[16] GUI is the abbreviation used for Graphical User Interface.

[17] https://en.wikipedia.org/wiki/Dashboard_(business)

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

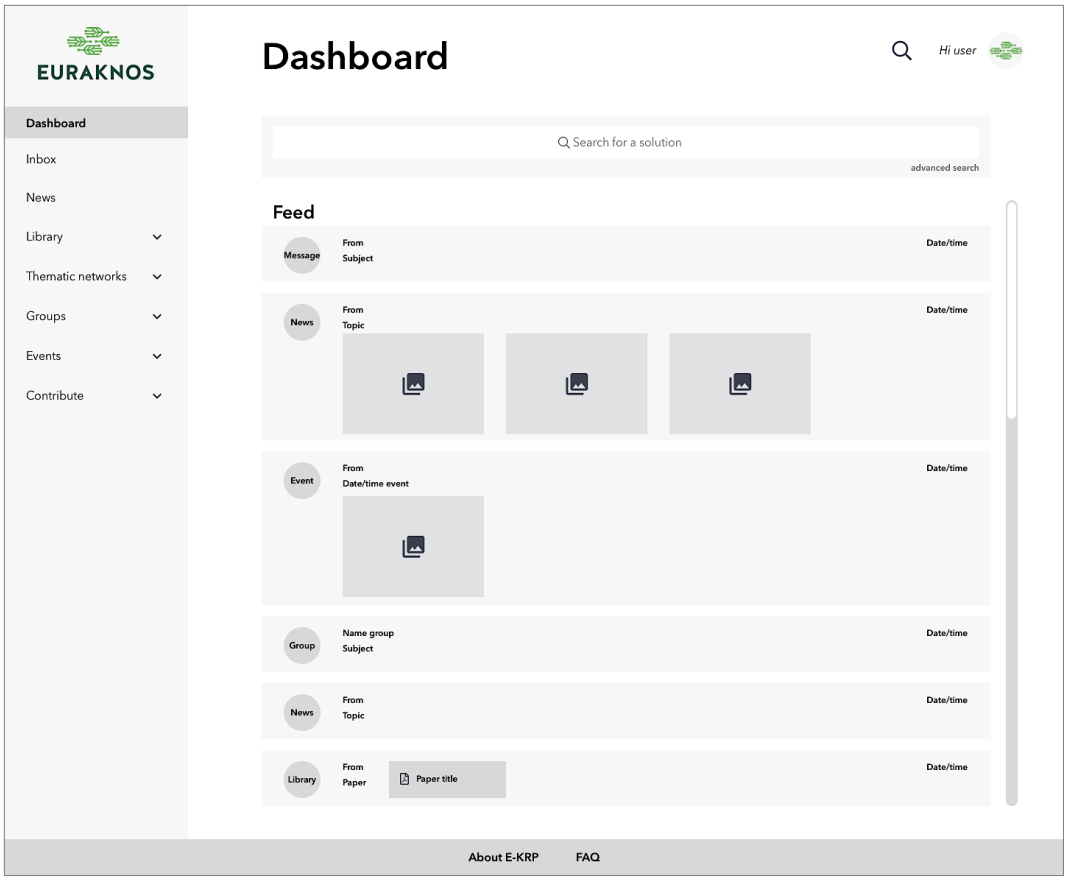


Figure 6. Version 2 of the e-KRP's landing page dashboard

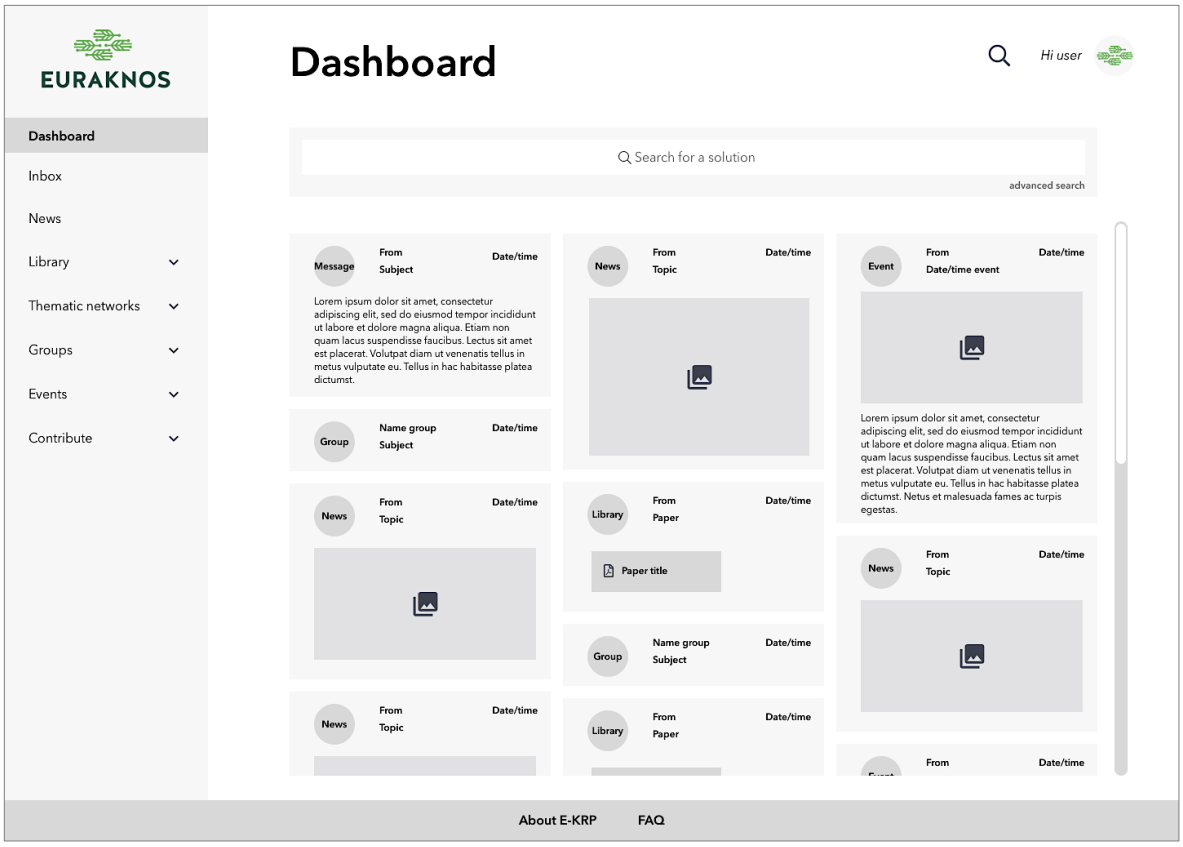


Figure 7. Version 3 of the e-KRP's landing page dashboard

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

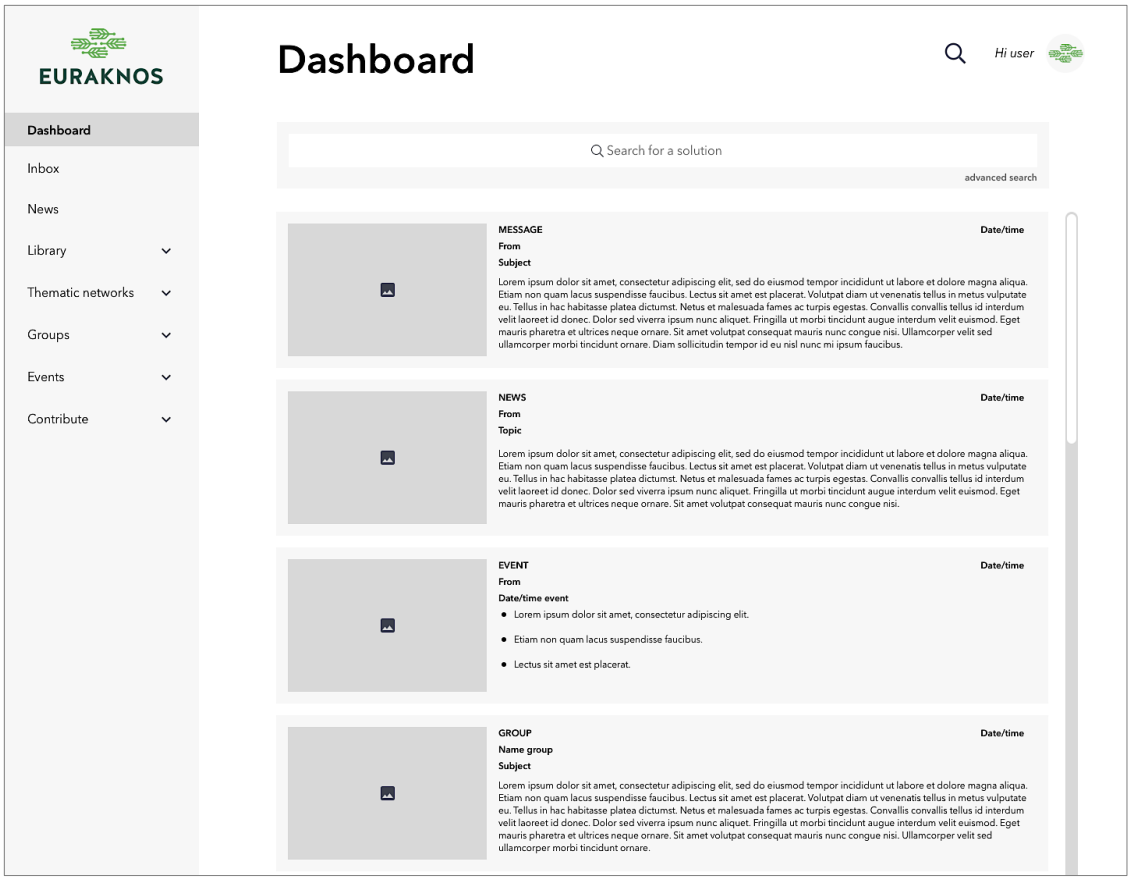


Figure 8. Version 4 of the e-KRP's landing page dashboard

- **Search bar**
  The solution proposed for searching for content and information in the repository of the e-KRP was that of a search function available in every e-KRP screen/page (see Figure 9 below).

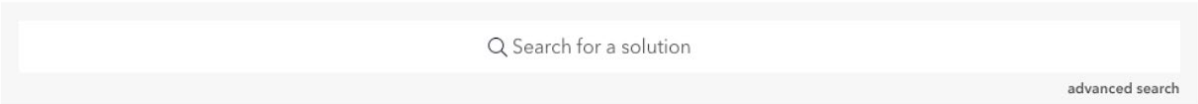


Figure 9. Search function provided in every e-KRP page with the aim to facilitate content and information search

- **News article**
  Two versions of news article-related views were created; namely, a view with a picture and a view without a picture illustrated in Figures 10 and 11 respectively.

- **Library overview**
  The library metaphor was proposed and shown to the consortium members in respect to the design of the e-KRP pages responsible for the display of content and information stored in the repository. The metaphor proposed is that of an online academic library, where the focus is on different ways to provide the overview of the requested content. Three different overview versions were proposed based on the wireframes shown in Figures 12, 13, and 14 below.
  - The first version allows a more graphical overview for navigating around the content (Figure 12).
  - The second version is oriented towards scrolling through textual content (Figure 13).
  - The third version constitutes a hybrid approach combining the advantages of the above versions (Figure 14).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Figure 10. Wireframe of news article view with a picture

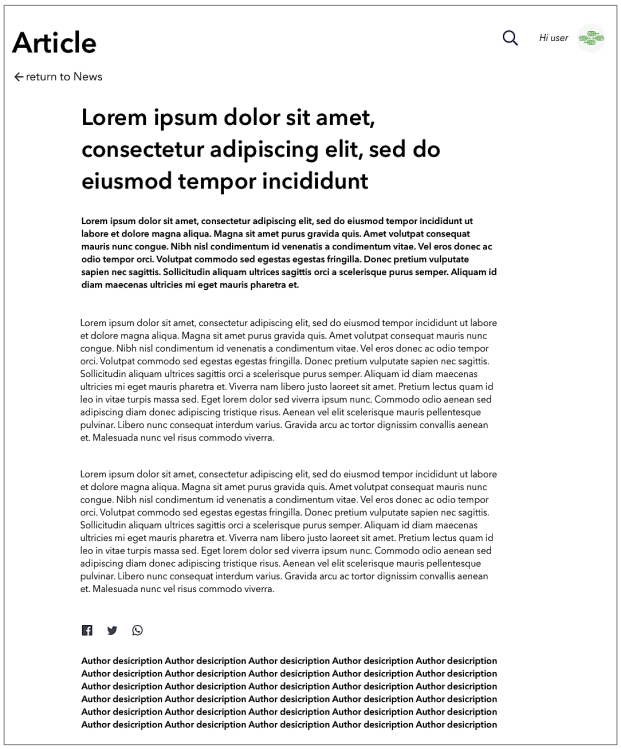


Figure 11. Wireframe of news article view without a picture

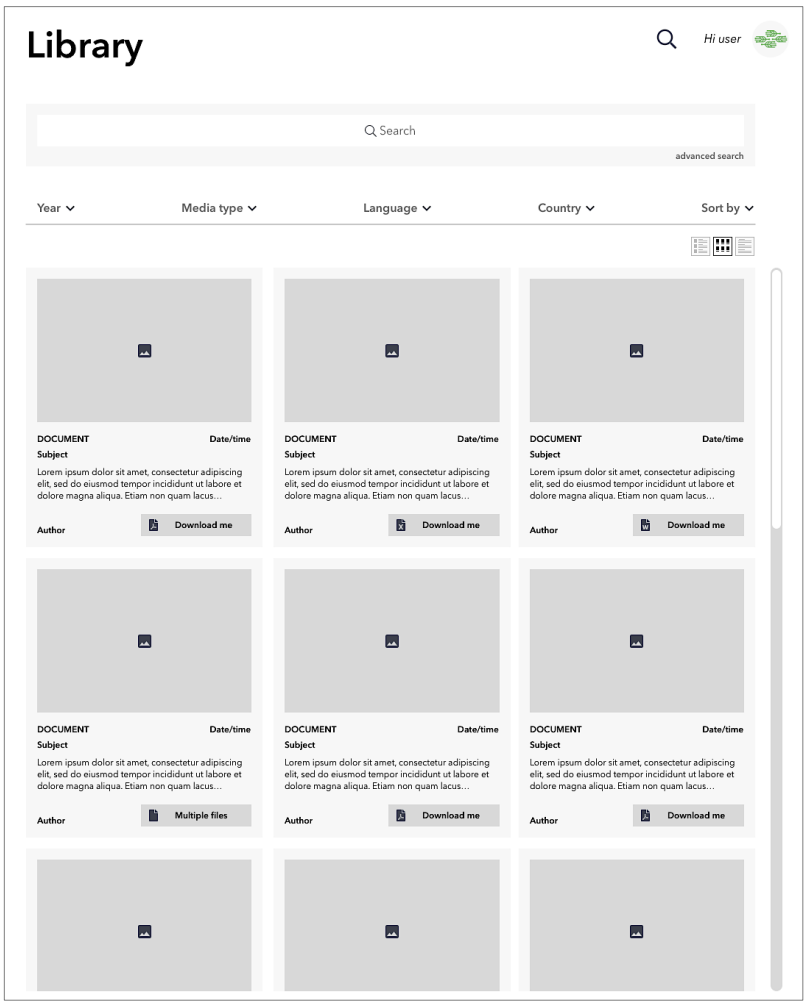


Figure 12. Version 1 of the library overview

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

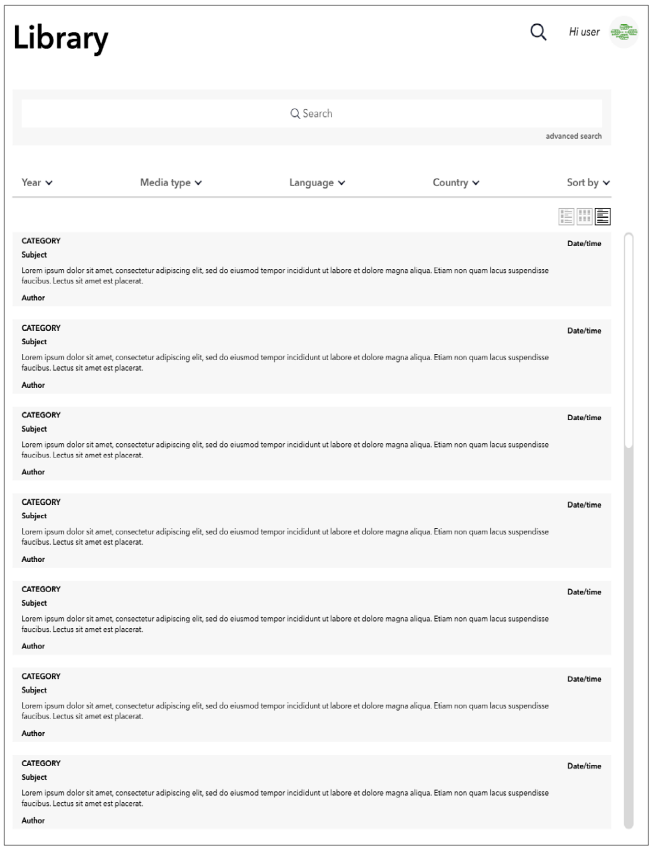


Figure 13. Version 2 of the library overview

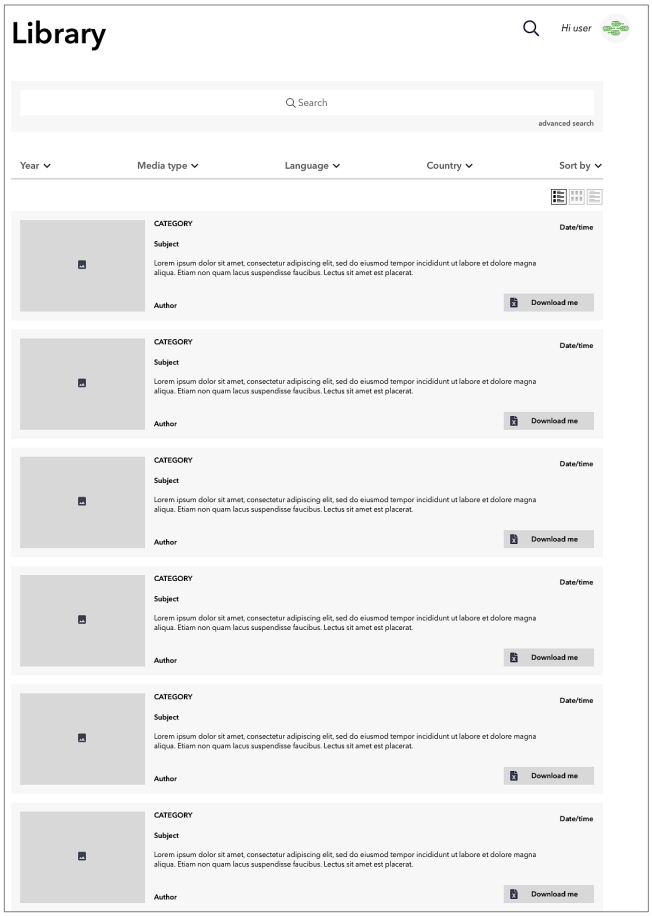


Figure 14. Version 3 of the library overview

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- **Library content**
  Regarding the delivery of specific types of content and information via the e-KRP, specially designed wireframes were used as indicative examples of how an article, a podcast, and a video could appear on the screen. In the case of an article/document, alternative display options were investigated. An example used in the exercise was that of the delivery of an academic article (see Figure 15 below). Example wireframes related the delivery of audio-visual content (namely, podcasts and videos) were also used for user needs identification and analysis exercise (see Figures 16 and 17 below).

Figure 15. Wireframe of academic article view

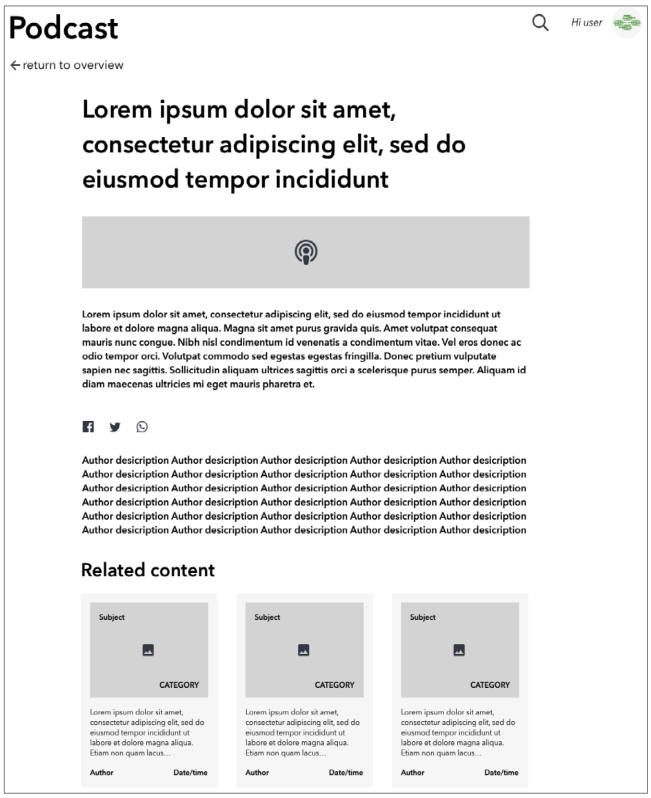


Figure 16. Wireframe of podcast view and access

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

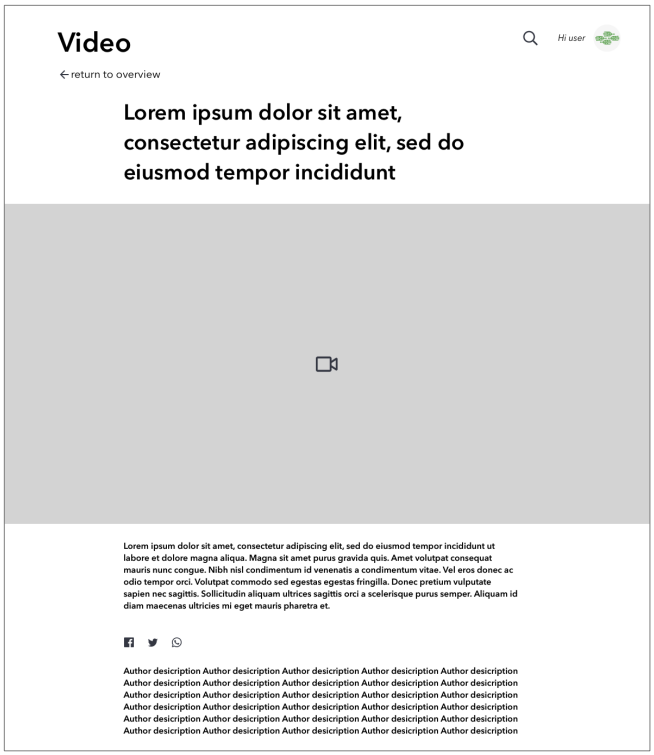


Figure 17. Wireframe of video view and access

The results obtained from the wireframe-based exercise have led to the following conclusions in regard to end-user preferences:

- **Dashboard**
  A clear preference in highly structured dashboards supported by fit-for-purpose visual aids has been reported. Thus, that the landing page dashboard solutions shown in Figures 5 and 8 were identified as the most preferred ones.
- **News articles**
  There has been a clear preference towards news content accompanied by a short text and graphical material. This means that the delivery of textual content with the support of graphical elements (for instance, a figure), as shown in Figure 10, is what the end-user mostly opt for.
- **Display of library content**
  The video display page received the greatest interest and the most positive feedback together with pages providing a brief textual description of the content. This means that solutions for the delivery of non-textual content (e.g., audio-visual content) like the one shown in Figure 17 are those that are mostly preferred by the end-users.

5.2. Input collected from the Marketplace session held in the Paris workshop

The wireframes provided also the basis for collecting insights and input on end-user needs as part of a Marketplace session that was held in the EURAKNOS Paris workshop (11 - 13 December, 2019). In that session, actors of the AKIS community/potential e-KRP users (namely, farmers/foresters, advisors, and researchers) provided feedback on design-related aspects (illustrated through the use of wireframes) and the basic concept underpinning the development of the e-KRP. The number of participants in the session was approximately 30. The feedback collected from this user needs analysis and identification session can be summarised to the following points:

- The platform needs to be responsive (i.e., work both on desktop and mobile devices).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- Some level of personalisation of the platform delivered to the end-user could be interesting through recommendations.
- A simple knowledge library would not attract much attention from farmers.
- It will be important that the platform provides better access to experts.
- It will be important that the platform is easily findable through the Google search engine.
- The way the content is structured and presented will be a great user-friendliness-related asset.
- Farmers should be able to look for information in their mother tongue.
- The community building element of the platform will be very important.
- Farmers should be able to provide feedback.
- The content on the platform should be monitored and curated.

5.3. Input from the work reported in other project deliverables

Apart from the insights and feedback that were obtained from the user needs analysis sessions in the Athens consortium meeting and the Paris workshop, useful input for the specification of e-KRP-related requirements has become available from the work documented in Deliverable 2.2 (Marois et al., 2019). The remarks and recommendations identified in that report as useful for specifying requirements are the following:

- Current TNs have different websites with different IT structures, formats and contents, depending on the actors involved in the project, the theme, the sector, the targeted audience and countries and the outputs of the project. To be of high impact the results which should not only be of high relevance to the end-user but also easily accessible and understandable, and this should be better facilitated.
- The sustainability of the website is a key issue in terms of long term impact. Therefore, a generic website or platform may be created with short descriptions and links to the individual project websites, which are built in a common format.
- It is essential to support end-users in their navigation in the KR and it is also important to think about finding a way to make the most appropriate content as easily accessible as possible.
- To be able to connect all KRs in a common open source infrastructure, a standardised framework should be developed. It should be a dynamic system with a self-improving feedback loop, with well overthought search options for specific farmers' needs, actions and different sectors.

5.4. Specifying requirements for the EURAKNOS e-KRP

The functional requirements for the e-KRP that have been specified based on the input from the user needs analysis sessions and the recommendations/concluding remarks reported in Deliverable 3.1 are listed in Table 3 below. Each requirement is mentioned in alignment with the specific piece(s) of input that have led to its definition.

Table 3. Functional requirements for the e-KRP presented in alignment with the input from the user needs analysis sessions and the project deliverable reports that has been used for specifying them

Requirement for the e-KRP	Input to the requirement's specification
Create views of the digital content available from the e-KRP based on the preferences of the end-users and the particularities of different data object categories (e.g., text-based material vs audio-visual content).	The way the content is structured and presented will be a great user-friendliness-related asset. • Preference in highly structured dashboards supported by fit-for-purpose visual aids. • Preference towards news content accompanied by a short text and graphical material. • Accompany the delivery of non-textual content (e.g., audio-visual material) with a brief textual description.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Provide functionalities that go beyond the delivery of content and information. Need to strengthen the interactions among community members via the e-KRP (e.g., functionalities for providing feedback to the content, discussion forum, access to expert-related information).	A simple knowledge library would not attract much attention from farmers.
	It will be important that the platform provides better access to experts.
	The community building element of the platform will be very important.
	Farmers should be able to provide feedback.
• Provide functionalities that will enable different ways of finding and retrieving relevant content and information (namely, search function along with an easy-to-use navigation structure). • Provide contextual information (i.e., metadata) for each data object to help the end-users identify content and information of relevenace and importance to them.	Current TNs have different websites with different IT structures, formats and contents, depending on the actors involved in the project, the theme, the sector, the targeted audience and countries and the outputs of the project. To be of high impact the results which should not only be of high relevance to the end-user but also easily accessible and understandable, and this should be better facilitated.
	It is essential to support end-users in their navigation in the KR and it is also important to think about finding a way to make the most appropriate content as easily accessible as possible.
Provide contextual information (i.e., metadata) for each data object to help the end-users identify content and information of relevenace and importance to them.	The sustainability of the website is a key issue in terms of long term impact. Therefore, a generic website or platform may be created with short descriptions and links to the individual project websites, which are built in a common format.
Provide functionalities that will enable different ways of finding and retrieving relevant content and information (namely, search function along with an easy-to-use navigation structure).	To be able to connect all KRs in a common open source infrastructure, a standardised framework should be developed. It should be a dynamic system with a self-improving feedback loop, with well overthought search options for specific farmers' needs, actions and different sectors.

Given the above, the functional requirements that have been identified for the e-KRP are the following:

- **Req#1**: Create views of the digital content available from the e-KRP based on the preferences of the end-users and the particularities of different data object categories (e.g., text-based material vs audio-visual content). Specifically:
  - Provide highly structured dashboards supported by fit-for-purpose visual aids.
  - Provide news content accompanied by a short text and graphical material.
  - Accompany the delivery of non-textual content (e.g., audio-visual material) with a brief textual description.
- **Req#2**: Provide functionalities that go beyond the delivery of content and information. Need to strengthen the interactions among community members via the e-KRP (e.g., functionalities for providing feedback to the content, discussion forum, access to expert-related information).
- **Req#3**: Provide functionalities that will enable different ways of finding and retrieving relevant content and information (namely, search function along with an easy-to-use navigation structure).
- **Req#4**: Provide contextual information (i.e., metadata) for each data object to help the end-users identify content and information of relevenace and importance to them.

In addition, the development of the e-KRP will be concerned with the aspect of enabling cross-platform access (i.e., access through a personal computer, tablet, or mobile phone) in response to the need for the e-KRP to be responsive (namely, "*The platform needs to be responsive (i.e., work both on desktop and mobile devices*"). Moreover, the e-KRP will deliver data objects conveying content and information in languages other than English given the need that has been reported in relation to the availability of

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

content/information in the end-users' mother tongue ("*Farmers should be able to look for information in their mother tongue*"). The need for curating and monitoring the content available from the e-KRP is of vital importance. However, it goes beyond the specification of requirements. It is an issue needed to be considered as part of a sustainability plan. A last point that was raised was that of personalisation and delivery of content-related recommendations (i.e., "*Some level of personalisation of the platform delivered to the end-user could be interesting through recommendations*"). This is a service that is made available in the digital platforms of several retail vendors. It requires to capture/track preference- and interaction-related data through the creation of user profiles and/or use of cookies. This aspect has not been fully considered for the Proof-of-Concept product built in EURAKNOS. It is an issue that will be further considered in the context of EUREKA (i.e., the follow-up project of EURAKNOS).

6. Shaping up the repository of the EURAKNOS e-KRP

This section of the report focusses on the presentation and explanation of the activities that have taken place as part of the phases of the methodology shown in Figure 18 below.

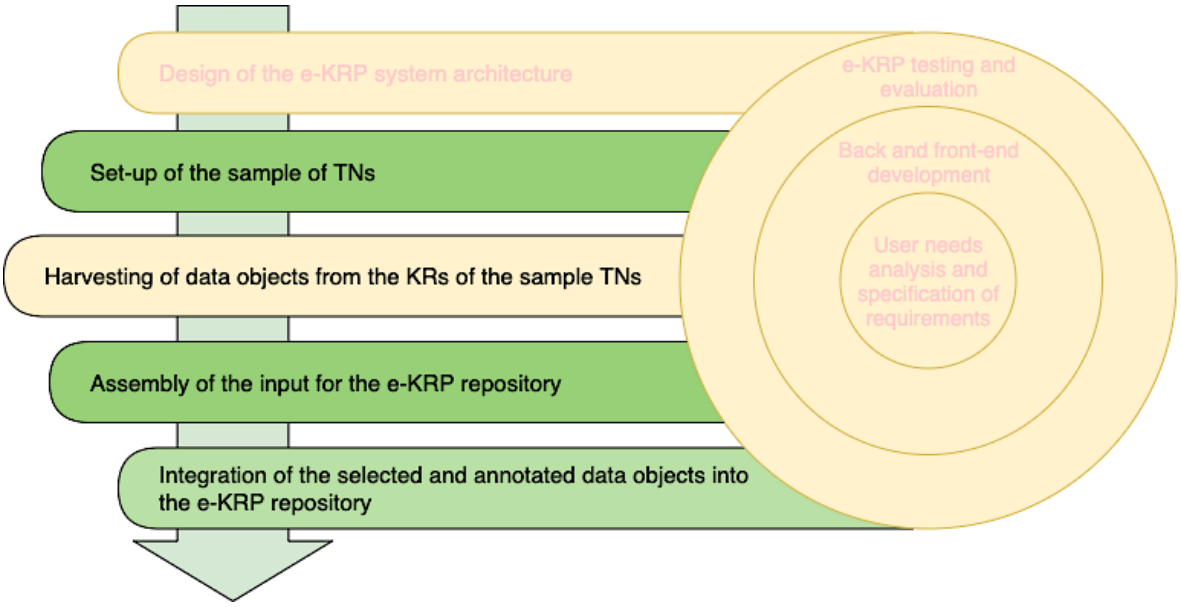


Figure 18. Phases of the Task 4.2 methodology relating to the shape up of the e-KRP repository

The methodology phases illustrated above and the activities executed in their context have served the purpose of populating the e-KRP repository with digital content, starting from the set-up of the sample of TN KRs to consider as sources of data objects and the harvesting of digital content from them. All four phases that appear in Figure 18 are described in detail below in Subsections 6.1 to 6.4.

6.1. Setting-up the sample of Thematic Networks to harvest content from

As explicitly mentioned in the EURAKNOS project's GA, "*data from selected TNs (8) will be put in HIKR format as a preparatory phase for the development of the prototype e-KRP as a pilot case study*". Based on that, one of the core activities that had to be executed for the needs of shaping up the EURAKNOS e-KRP repository has been to assemble the sample of the eight (8) TNs whose KRs would be the sources of data objects. The sample has been set up by being based on the selection criteria listed below:

- **Criterion #1**: Mapping of TNs with regard to the existence of their KRs

The existence of a TN's KR has been a requisite for considering the TN as a candidate for the TN sample. The application of this criterion has been based on the findings of the TN KR analysis reported in Marois et al. (2019). Therefore, the TNs that had not launched their KRs by the time at which the analysis of

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

the existing TNs' KRs took place[18] in Task 2.2 ("Design of knowledge reservoirs within current and past TN") would not be considered as sample candidates.

- **Criterion #2**: TNs rate of completion

Another criterion to consider has been the rate of each TN's completion. The information required to assess this criterion has been made available in the analysis of TNs reported in Deliverable 2.1 ("Report on content and structures of data provision and knowledge formats of TNs")[19]. However, given the time elapsed since the collection of information about the completion rates of TNs in Task 2.1 ("Data impact assessment of TNs")[20], a new process of identifying/estimating the TNs' completion rates took place in Task 4.2 to draw upon up-to-date information.

- **Criterion #3**: Domain the TNs belong to[21]

The term "domain" refers to the agricultural subsector that a TN's theme relates to[22]. These subsectors are "Crop farming", "Livestock", and "Forestry". The "Aquaculture & fisheries" domain has not been considered because there is no TN with a theme relating to it.

- **Criterion #4**: Involvement of EURAKNOS partners in the TN

The process of harvesting data from the TN KRs for the needs of assembling the input for the repository of the e-KRP, necessitated the establishment of robust communication channels with the coordinators of the TNs to efficiently address any issues that may arose during the input collection process. Thus, considering TNs in which EURAKNOS partners have had an active role (as coordinators or consortium partners) is a parameter that was identified as worth considering as well.

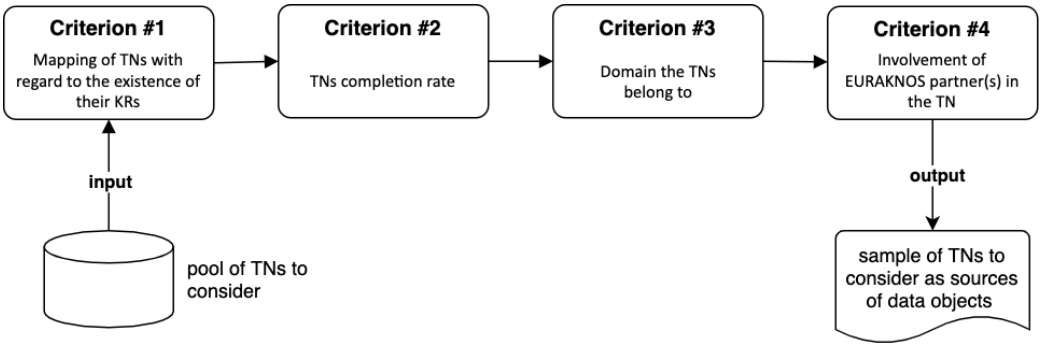


Figure 19. The sequence of criteria for setting-up the sample of TNs to consider as data object sources

The four (4) criteria described above have been applied sequentially with each one of them helping to filter out "irrelevant" TNs. Specifically, this sequential application of criteria has functioned as a funnel that has allowed to limit the initial number of TNs (28 in total[23]) to the final sample of the eight (8) TNs (output of the application of Criterion #4). The initial set of the 28 TNs provided as input to Criterion #1 are those that were considered for the analysis of the TN KRs undertaken in the context of Task 2.2 ("Design of knowledge reservoirs within current and past TN")[24] and reported in Deliverable 2.2

[18] Task 2.2 started in January 2019 (M01 of the project) and ended in August 2019 (M08).

[19] Mosquera Losada, M. R., et al. (2019). Report on content and structures of data provision and knowledge formats of TNs. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 2.1 – WP2.

[20] Task 2.1 started in January 2019 (M01 of the project) and ended in August 2019 (M08).

[21] "Crop farming", "Livestock", "Forestry", and "Aquaculture & fisheries" are according to FAO (http://www.fao. org/rural-employment/agricultural-sub-sectors/en/) the major subsectors into which the agricultural activity can be divided into.

[22] According to the EIP-AGRI's report "*Horizon 2020 thematic networks work on a specific theme*" (2016).

[23] This was the number of the TNs that were in progress or had been completed by the time at which EURAKNOS started (i.e., in January 2019).

[24] The TNs that have been considered for the creation of the sample are: 4D4F, AFINET, AGRI FOR VALOR, AGRISPIN, BEST4SOIL, CERERE, DISARM, ENABLING, EU-FRUIT, EUPIG, EURODAIRY, FERTINNOWA, HENNOVATION, HNV-LINK,

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

("Report on format and design of knowledge reservoirs of TNs")[25]. Figure 19 above shows the sequence of the four criteria application.

Table 4 summarises the results of the analysis executed in Task 2.2 in regard to whether the KR of a TN exists (or more specifically, existed at the time of the analysis) and the different types of existing KRs. Apart from listing the TNs having no KR implemented, it also allows insights into the different types of KR implementations such as the following (Marois et al., 2019):

- The TN uses its official website as its KR.
- The TN uses its official website as its KR together with a data repository dedicated to this purpose.
- The TN uses a general-purpose data repository as its KR (not its own, custom-made repository).
- The TN uses a custom-made data repository dedicated to the purpose of serving as its KR.

Table 4. Outcomes of the analysis of TNs in terms of KR existence and different implementations of their KRs (adapted from Marois et al., 2019)

Existence of the TN's KR	TN number	TN rate (%)	Name of the TN
The TN uses its official website as its KR	12	43%	• FERTINNOWA • 4D4F • CERERE • AGRI FOR VALOR • EUPIG • NEWBIE • SUWANU • AGRISPIN • SKIN • EURODAIRY • HNV-LINK • INCREDIBLE
The TN uses its official website as its KR together with a data repository dedicated to this purpose	1	4%	• HENNOVATION
The TN uses a general-purpose data repository as its KR (not its own, custom-made repository)	3	11%	• AFINET • SHEEPNET • NUTRIMAN
The TN uses a custom-made data repository dedicated to the purpose of serving as its KR	6	21%	• SMART-AKIS • OK-NET ARABLE • WINETWORK • OK-NET ECOFFED • INNO4GRASS • EU-FRUIT
No KR available	6	21%	• ENABLING • DISARM • LEGUMES TRANSLATED • PANACEA • INNOSETA • BEST4SOIL
TOTAL	28	100%	

Based on the results available in the table, there is a total number of six (6) TNs that had no KR available by the time of the analysis. These TNs are BEST4SOIL, DISARM, ENABLING, INNOSETA, PANACEA, and LEGUMES TRANSLATED. These TNs were excluded from the list of candidate TNs for the final sample. The outcome of the application of Criterion #1 has been the list of remaining TNs[26], which have served as the input necessary to apply Criterion #2.

By providing this list of 22 TNs as input to the implementation of Criterion #2, the intention has been to further narrow down the number of TNs to consider for the final sample. Criterion #2 relates to each TN's completion rate (if not ended). Such information was already available by Deliverable 2.1 ("Report on content and structures of data provision and knowledge formats of TNs")[27]. Nevertheless, an update was deemed to be necessary in order to draw upon up-to-date information. For this purpose,

INCREDIBLE, INNO4GRASS, INNOSETA, LEGUMES TRANSLATED, NEWBIE, NUTRIMAN, OK-NET ARABLE, OK-NET ECOFEED, PANACEA, SHEEPNET, SKIN, SMART-AKIS, SUWANU, and WINETWORK.

[25] Marois, M. et al. (2019). Report on format and design of knowledge reservoirs of Thematic Networks. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 2.2 – WP2.

[26] The TNs included in this list are: 4D4F, AFINET, AGRI FOR VALOR, AGRISPIN, CERERE, EU-FRUIT, EUPIG, EURODAIRY, FERTINNOWA, HENNOVATION, HNV-LINK, INCREDIBLE, INNO4GRASS, NEWBIE, NUTRIMAN, OK-NET ARABLE, OK-NET ECOFEED, SHEEPNET, SKIN, SMART-AKIS, SUWANU, and WINETWORK.

[27] Mosquera Losada, M. R., et al. (2019). Report on content and structures of data provision and knowledge formats of TNs. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 2.1 – WP2.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

a process of searching for TN-related information in the CORDIS database[28] took place. This information related to the start and end years, and months, of the 22 TNs (provided as input to the application of Criterion #2) to calculate their completion rates. The results are shown in Table 5 below.

Table 5. Completion rates of TNs (until June, 2020)

TN acronym	TN URL	Start year	End year	Duration	Completion rate
4D4F	https://4d4f.eu/	1 March, 2016	28 February, 2019	29 months	100%
AFINET	https://euraf.isa.utl.pt/afinet	1 January, 2017	31 December, 2019	36 months	100%
AGRI-SPIN	https://agrispin.eu/	1 March, 2015	31 August, 2017	30 months	100%
AGRIFORVALOR	http://www.agriforvalor.eu/pages/home	1 March, 2016	31 August, 2018	30 months	100%
CERERE	http://cerere2020.eu/	1 November, 2016	31 October, 2019	36 months	100%
EU Pig	https://www.eupig.eu/	1 November, 2016	31 October, 2020	48 months	89.60%
EU-FRUIT	http://www.eufrin.eu/index.php?id=55	1 March, 2016	28 February, 2019	36 months	100%
EURODAIRY	https://eurodairy.eu/	1 February, 2016	31 January, 2019	36 months	100%
FERTINNOWA	https://www.fertinnowa.com/	1 January, 2016	31 December, 2018	36 months	100%
HENNOVATION	http://www.hennovation.eu/	1 January 2015	31 August, 2017	32 months	100%
HNV-LINK	http://www.hnvlink.eu/	1 April, 2016	31 March, 2019	24 months	100%
INCREDIBLE	https://www.incredibleforest.net/	1 November, 2017	31 October, 2020	36 months	86.20%
INNO4GRASS	https://www.inno4grass.eu/en/	1 January, 2017	31 December, 2019	36 months	100%
NEWBIE	http://www.newbie-academy.eu/	1 January, 2018	31 December, 2021	48 months	60.50%
NUTRIMAN	https://nutriman.net/	1 October, 2018	31 March, 2021	30 months	66.70%
OK-NET ARABLE	http://www.ok-net-arable.eu/	1 March, 2015	28 February, 2018	36 months	100%
OK-NET ECOFEED	https://ok-net-ecofeed.eu/	1 January 2018	31 December, 2020	36 months	80.60%
SHEEPNET	http://www.sheepnet.network/	1 November, 2016	31 October, 2019	36 months	100%
SKIN	http://www.shortfoodchain.eu/	1 November, 2016	31 October, 2019	23 months	100%
SMART-AKIS	https://www.smart-akis.com/	1 March, 2016	31 August, 2018	30 months	100%
SUWANU	https://cordis.europa.eu/project/id/818088	1 January, 2019	30 June, 2021	30 months	43.40%
WINETWORK	http://www.winetwork.eu/	1 April, 2015	30 September, 2017	30 months	100%

In the case of non-completed TNs, the completion rate has been calculated by using the formula:

$$\boldsymbol{completion\ rate} = \frac{\boldsymbol{number\ of\ months}\ the\ TN\ is\ in\ progress}{\boldsymbol{total\ duration}\ of\ the\ TN\ (in\ months)} = \frac{(\boldsymbol{current}\ year\ \&\ month) - (\boldsymbol{start}\ year\ \&\ month)}{\boldsymbol{total\ duration}\ of\ the\ TN\ (in\ months)} \quad (1)$$

[28] https://cordis.europa.eu/en

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

To further filter the list of the TN sample candidates, only the TNs with a completion rate greater than or equal to 75% (≥75%) were shortlisted[29]. The application of Criterion #2 resulted in the preclusion of three more TNs, namely:

- NEWBIE (completion rate = 60.50% < 75%);
- NUTRIMAN (completion rate = 66.70% < 75%); and
- SUWANU (completion rate = 43.40% < 75%).

After the deployment of Criterion #2, the number of TNs that were provided as input to the application of Criterion #3 was limited to 19. The remaining TN candidates (for the sample of the 8 TNs) are listed in Table 6 concerning the domain (i.e., the agricultural subsector) that their themes relate to.

Table 6. List of TNs obtained as the outcome of the application of Criterion #2 and the domain their theme relates to

Domain	TN name	Number of TNs	%
Crop farming	AGRI FOR VALOR, AGRISPIN, CERERE, EU-FRUIT, FERTINNOWA, HNV-LINK, OK-NET ARABLE, SKIN, SMART-AKIS, WINETWORK	10	52.6%
Livestock	4D4F, EURODAIRY, EUPIG, HENNOVATION, INNO4GRASS, OK-NET ECOFEED, SHEEPNET	7	36.8%
Forestry	AFINET, INCREDIBLE	2	10.5%
	TOTAL	**19**	**100%**

The association of these 19 TNs with each of the "Crop farming", "Livestock", and "Forestry" domains has been based on lists of keywords used to describe them. These keywords have been extracted:

- from the EIP-AGRI database[30] that provides access to TN-related information; and
- by using the cortical.io web-based application[31] after having provided a short textual description of the TN[32] as input to it.

The last column of Table 6 provides a rate indicative of the distribution of the 19 candidate TNs across the domains considered. Specifically, the TNs related to the "Crop farming" domain account for 52.6% of the total number of the 19 candidate TNs. In the same line, the subset of the Livestock-related TNs includes 36.8% of the total number of TNs, and the Forestry-related TNs account for 10.5% of the pool of sample candidates. The creation of a representative sample of eight (8) TNs, to consider as sources of data objects, necessitated to keep the distribution shown in Table 6. Thus, in the final sample of the eight (8) TNs, 52.6% of the TNs should come from the "Crop farming" domain, 36.8% should relate to "Livestock", and 10.5% would need to be about "Forestry". Given this distribution, the number of the sample TNs considered per agricultural subsector (i.e., domain) was:

- Number of TNs related to "Crop farming" in the sample: $8 \times 0.526 = 4.208 \cong \mathbf{4}\ \boldsymbol{TNs}$
- Number of TNs related to "Livestock" in the sample: $8 \times 0.368 = 2.944 \cong \mathbf{3}\ \boldsymbol{TNs}$
- Number of TNs related to "Forestry" in the sample: $8 \times 0.105 = 0.84 \cong \mathbf{1}\ \boldsymbol{TN}$

The final step focussed on the identification of the final list of sample items. Doing so has involved the use of Criterion #4 (i.e., "Involvement of EURAKNOS partners in the TN"). Regarding the involvement of EURAKNOS partners in the 19 TNs (i.e., the output of Criterion #3), we distinguish the following:

- A partner in EURAKNOS is the coordinator of the TN.

[29] This threshold value has been selected based on the assumption that a TN completed at its ¾ has already delivered a quite considerable number of results.

[30] https://ec.europa.eu/eip/agriculture/en/find-connect/projects

[31] https://www.cortical.io/freetools/extract-keywords/

[32] The textual descriptions of TNs that have been provided as input to the cortical.io web application, for the extraction of TN-related keywords, were derived from the TN website pages giving background information on the TN (namely, the website pages able to be accessed by selecting the "Project" and/or "About" menu options).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- A partner in EURAKNOS is a member of the TN's consortium (i.e., a TN partner).
- A partner in EURAKNOS does not have a direct involvement in the TN but acts as an intermediary for coming in contact with the TN's coordinator or a consortium partner (i.e., a contact point).

Table 7[33] below presents the involvement of the EURAKNOS partners in the 19 TNs outputted from the application of Criterion #3.

Table 7. Type of involvement of EURAKNOS partners in the 19 TNs outputted from the application of Criterion #3.

Domain	Thematic Network	completion rate	EURAKNOS partner involved	Type of involvement
Crop Farming	AGRISPIN	100%	ACTA	contact point
	AGRI FOR VALOR	100%	NAK	contact point
	CERERE	100%	AU	contact point
	EU-FRUIT	100%	-	-
	FERTINNOWA	100%	PSKW	TN coordinator
	HNV-LINK	100%	USC	contact point
	OK-NET ARABLE	100%	IFOAM EU	TN coordinator
	SKIN	100%	USC	contact point
	SMART-AKIS	100%	AUA	TN coordinator
	WINETWORK	100%	ACTA	contact point
Livestock	4D4F	100%	IfA	TN coordinator
	EUPIG	89.60%	ACTA	contact point
	EURODAIRY	100%	AHDB	contact point
	HENNOVATION	100%	RAU	TN coordinator
	INNO4GRASS	100%	GLZ	TN coordinator
	OK-NET ECOFEED	80.60%	IFOAM EU	TN coordinator
	SHEEPNET	100%	IDELE	TN coordinator
Forestry	AFINET	100%	USC	TN coordinator
	INCREDIBLE	86.20%	USC	contact point

Showing a preference towards TNs where EURAKNOS partners were/are directly involved (either with the role of coordinator or as a consortium partner) is important, especially when it comes to issues of establishing a communication channel between EURAKNOS and the TN.

Based on this decision, OK-NET ARABLE, FERTINNOWA, and SMART-AKIS are three (out of the total number of four) TNs that were selected for the sample from the "Crop Farming" domain. Given that there is no TN where a EURAKNOS partner has been involved as a consortium partner, the fourth TN was been randomly chosen from the rest of the list. The TN finally selected was SKIN.

In the case of the TNs that relate to the "Livestock" domain, there are more TN cases where EURAKNOS partners have been involved with the role of coordinator. By taking account of these TNs' completion rates, the sample members were chosen among the Livestock-related TNs already finished (i.e., 4D4F, HENNOVATION, INNO4GRASS, and SHEEPNET) and coordinated by partners in the EURAKNOS project.

However, only three TNs were needed to be selected from the "Livestock" domain. Thus, the sample TNs from the specific domain were selected in a random way. After this random selection process, the Livestock-related TNs included in the sample were 4D4F, INNO4GRASS, and SHEEPNET.

The TN that has been selected from the domain of "Forestry" is AFINET (namely, an already completed TN coordinated by USC - a partner in EURAKNOS).

Figure 20 shows the TNs finally included in the sample and their domain.

[33] The information presented in Table 7 with regard to the type of involvement of EURAKNOS partners in TNs has been derived from the TN mapping process that has been undertaken in WP2.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Figure 20. The TNs selected for the sample to serve as data sources

6.2. Harvesting content from the Knowledge Reservoirs of the sample Thematic Networks

The categories/types of data objects to consider for storage into the EURAKNOS e-KRP repository that have been proposed in the report of Deliverable 3.2 ("Report on HIKR design")[34], together with the sample of TNs to serve as data object sources, have provided the input needed for the implementation of the "Harvesting of data objects from the KRs of the sample TNs" phase of the methodology. The data object harvesting process has been executed with the help of a software application particularly created for this purpose; i.e., the EURAKNOS web crawler. Before providing details on the data objects that have been harvested from the web crawler (i.e., the input to consider for the repository of the e-KRP), some background information on the web crawling information is presented below.

6.2.1. The EURAKNOS web crawler

A web crawler is a computer program used to automatically search for documents on the web. Web crawlers are programmed to execute repetitive actions so that the browsing process becomes automated. A web crawler looks for information on the web, categorises it, and then indexes and catalogues it so that the crawled information is retrievable and can be evaluated.

The operations performed by a web crawler need to be established before the crawling process starts and the computed instructions are automatically executed by the web crawler. An index is created with the results of the crawling process, which can be accessed with the help of output software. The content that a web crawler can harvest from the Web depends on the instructions provided in its development phase. The resources to be crawled are indicated by a list of URLs (also referred to as "seeds"). As shown in Figure 21 below, a crawling application is usually made up of the following components (Patil & Patil, 2016):

- the Crawler frontier that holds a list of URLs to be visited by the crawler;
- the Page Downloader that downloads pages from the Web; and
- a Web Repository that receives content for storage from the crawler.

Regarding the EURAKNOS web crawler, the points listed below offer a rough outline of the major steps taken as part of its development and testing. More specifically:

- The web crawler's design and development started in 2019. In this early stage, an initial version was created and tested with the aim to evaluate the diversity and quality of the available KRs of TNs.
  - The first version has been a static web crawler built in the Python programming language. The testing process of this initial version helped towards the identification of problems related to the metadata harvested from some TN websites and KRs.

[34] Van Hal, D. et al. (2020). Report on HIKR Design. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 3.2 – WP3.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- The results obtained from the testing of the crawler's initial version allowed for decisions about the programming language to use for its development. Therefore, there was a shift from Python to PHP.
- Improvements have been made to the first web crawler version in order to finally create a more flexible, responsive, and easy-to-use tool.
- An experimentation environment was set up in the AKI's private cloud, where the EURAKNOS web crawler and its database were hosted. AKI has granted permission to all WP4 partners to facilitate collaboration. As a result, an efficient testing process became feasible.
- Based on the considerations for the different categories/types of data objects to integrate into the e-KRP repository, a web-based Graphical User Interface (GUI) has been created for the web crawler and efforts have taken place towards allowing the tool to crawl/scrap content that is not only static.
- Existing, open-source technologies have been considered and used for the crawler's development; during the testing process, there have been modifications to the web crawler which allowed to also harvest data from websites developed with the use of dynamic JavaScript frameworks.

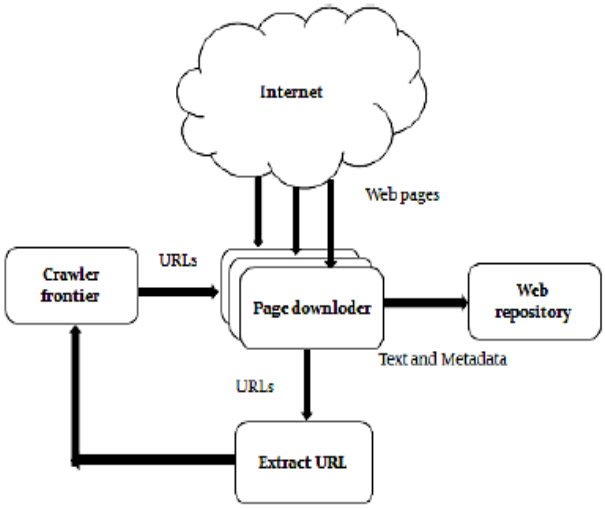


Figure 21. An indicative architecture of a web crawler (Patil & Patil, 2016)

The code developed for the EURAKNOS web crawling application has been published in Zenodo and is able to be accessed at https://zenodo.org/record/5386275#.YTnea54zbUI.

6.2.2. A pool of data objects to consider as input for the e-KRP repository

The process of crawling the KRs of TNs (i.e., the TNs in our sample) and harvesting of data objects to consider as input to the e-KRP repository took place in May and June 2020 after the consent of the TN coordinators. The data objects that were harvested from the KRs of the sample TNs by the web crawler were initially stored in the crawler's database (an instance of the MySQL database). The web crawler's landing page URL is http://onpremgw1.bioeast.eu/ and is shown in Figure 22 below.

It needs to be stressed that the EURAKNOS web crawling application is not a single crawler, but rather consists of a set of web crawlers with each of them focussed on harvesting data from a specific TN KR. Details about the web crawler's functionalities are provided in Annex I.

The web crawler's landing page allows access to a range of useful information. More specifically, there is an array of tiles at the top of the landing page that inform the user about the number of data items harvested per data object category (i.e., text/documents, images, presentations, videos, and audios). There is also a tile informing about the existence of harvested data objects that could not be classified in any of the categories instructed by the EURAKNOS data model (Van Hal et al., 2020). Such cases of data objects have been classified into the "Other" category. The number of data objects harvested per category allowed us to cross-validate the results of the research undertaken in Task 2.1 by Mosquera Losada et al. (2019). For instance, one of the conclusion that was drawn relates to the fact that specific

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

categories of data objects (e.g., videos), which are mostly preferred by the end-users, are not the most frequently occurring ones.

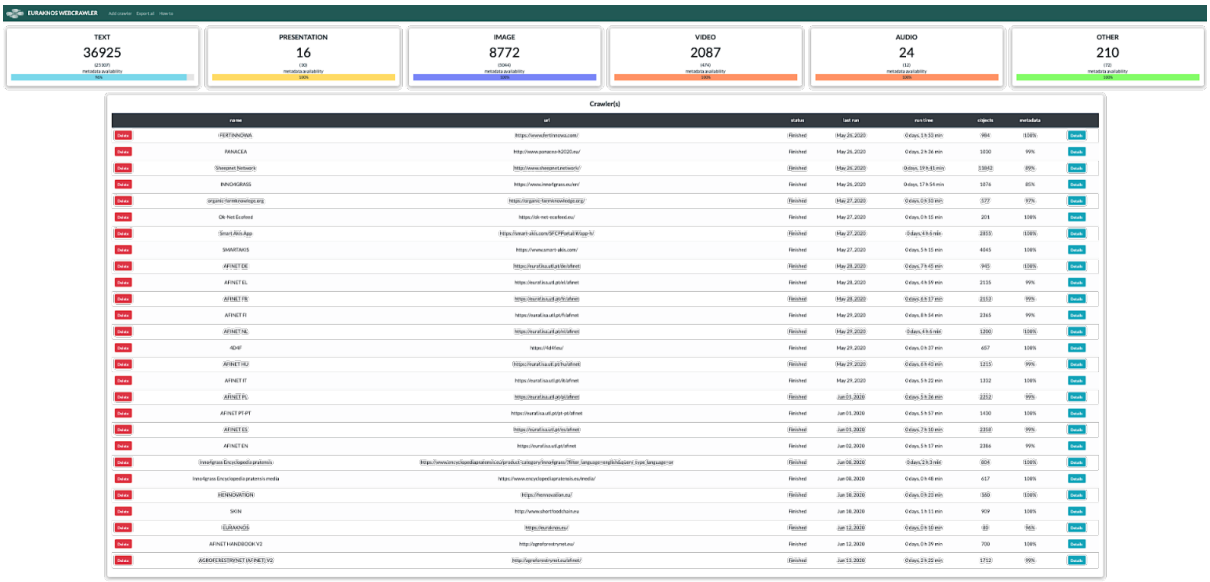


Figure 22. The landing page of the EURAKNOS web crawler

The list below the array of tiles provides details about each web crawler's status such as: (i) the name of the web crawler; (ii) the URL of the crawled TN; (iii) the web crawler's status (e.g., finished); (iv) the date of the last crawling execution[35]; (v) the time required for the completion of the crawling process; (vi) the number of the data objects harvested; and (vii) the availability of metadata. A "Details" button available on the right side of each line allows to access useful statistics about the kinds of data objects harvested from a TN's KR, as well as the existence of metadata per category of data objects. Figure 23 provides an overview of this statistics-related information for the data objects harvested from the KR of the FERTINNOWA TN.

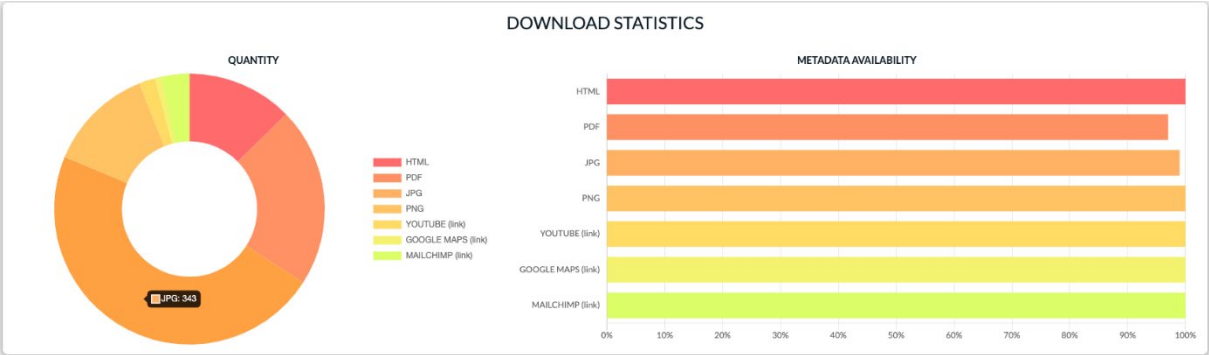


Figure 23. Statistics related to the data objects harvested from the KR of the FERTINNOWA TN

Table 8 below provides details about the categories/types and formats of the data objects harvested from each of the sample's TN KRs, together with information about the number of data items per type/ category and format. Also, information about the completeness of the metadata[36] is made available.

Table 8. Information about the number of data objects harvested from each TN's KR per data object category/type and format together with information about metadata availability

Domain	TN name	Number of data items harvested			Rate (%) of metadata availability		
		text/	html	92	text/	html	100%

[35] A web crawler can be executed multiple times.

[36] When referring to metadata, we actually mean the metadata available from the source KR of a data object.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Domain	TN name	Number of data items harvested			Rate (%) of metadata availability		
Crop farming	FERTINNOWA	documents	pdf	156	documents	pdf	97%
		videos	links to YouTube	14	videos	Links to YouTube	100%
		images	jpg	343	images	jpg	99%
			png	92		png	100%
		other	Google maps	6	other	Google maps	100%
			Mailchimp	24		Mailchimp	100%
	OK-NET ARABLE	text/ documents	html	224	text/ documents	html	100%
			pdf	138		pdf	88%
		videos	links to YouTube	9	videos	links to YouTube	100%
		images	jpg	77	images	jpg	100%
			png	24		png	100%
	SKIN	text/ documents	html	217	text/ documents	html	100%
			pdf	86		pdf	100%
		videos	links to YouTube	5	videos	links to YouTube	100%
		images	jpg	321	images	jpg	99%
			png	152		png	100%
		other	Google maps	-	other	Google maps	100%
	SMART-AKIS	text/ documents	html	2873	text/ documents	html	100%
			pdf	1654		pdf	98%
		presentations	pptx	-	presentations	pptx	100%
			prezi (link)	-		prezi (link)	100%
		videos	links to YouTube	27	videos	links to YouTube	100%
		images	jpg	118	images	jpg	99%
			png	78		png	100%
		other	Google maps	27	other	Google maps	100%
			Mailchimp	-		Mailchimp	100%
Livestock	4D4F	text/ documents	html	265	text/ documents	html	100%
			pdf	353		pdf	99%
			docx	-		docx	100%
			xlsx	-		xlsx	100%
		presentations	pptx	-	presentations	pptx	100%
		videos	links to YouTube	50	videos	links to YouTube	100%
		images	jpg	125	images	jpg	98%
			png	75		png	100%
	INNO4GRASS	text/ documents	html	1081	text/ documents	html	100%
			pdf	330		pdf	99%
		videos	links to YouTube	117	videos	links to YouTube	100%
		images	jpg	879	images	jpg	99%
			png	89		png	100%
			bmp	152		bmp	100%
	SHEEPNET	text/ documents	html	12286	text/ documents	html	100%
			pdf	571		pdf	94%
			docx	247		docx	97%
		videos	links to YouTube	93	videos	links to YouTube	100%
			links to Vimeo	-		links to Vimeo	100%
		images	jpg	582	images	jpg	97%
			png	274		png	100%
			svg	-		svg	0%
		other	Google maps	-	other	Google maps	100%
Forestry	AFINET	text/ document	html	311	text/ document	html	100%
			pdf	1147		pdf	97%
			docx	-		docx	100%
		presentations	pptx	-	presentations	pptx	100%

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Domain	TN name	Number of data items harvested			Rate (%) of metadata availability		
			prezi (link)	-		prezi (link)	100%
		videos	links to YouTube	51	videos	links to YouTube	100%
			links to Vimeo	11		links to Vimeo	100%
		images	jpg	476	images	jpg	100%
			png	54		png	100%
			bmp	-		bmp	100%
		other	Google maps	3	other	Google maps	100%
			Mailchimp	12		Mailchimp	100%

Figure 24 shows a screenshot of the input collected for the e-KRP (the pool of data objects to consider as input for the repository). The file type names (format of the data) illustrated at the top of the figure give an overview of the categories/types of the data objects that were harvested. The database of the web crawler contains also data objects harvested from TN KRs not included in the sample. These extra KRs have been considered for testing purposes.

url	domain	page	pdf	docx	xlsx	epub	pptx	jpg	png	swf	svg	youtube video	vimeo video	spotify	mp4	google_map	zip
https://www.fertinnowa.com/	FERTINNOWA	353	165	1	0	0	0	340	88	0	1	28	0	0	0	8	0
http://www.panacea-h2020.eu/	PANACEA	697	96	0	0	0	0	124	69	0	0	0	0	0	0	44	0
http://www.sheepnet.network/	Sheepnet Network	10081	533	205	0	0	0	553	243	0	1	188	8	0	0	30	0
https://www.inno4grass.eu/en/	INNO4GRASS	340	558	0	0	0	0	60	20	0	0	98	0	0	0	0	0
https://euraknos.eu/	EURAKNOS	58	3	0	0	0	0	0	3	0	3	6	0	12	0	0	0
https://organic-	organic-farmknowlege.org	319	134	0	0	0	0	83	20	0	1	20	0	0	0	0	0
https://ok-net-ecofeed.eu/	Ok-Net Ecofeed	130	14	0	0	0	0	28	21	0	0	8	0	0	0	0	0
https://smart-	Smart Akis App	2124	664	0	0	0	0	13	14	0	0	40	0	0	0	0	0
https://www.smart-akis.com/	SMARTAKIS	2338	1417	0	0	0	1	117	78	0	0	40	0	0	0	54	0
http://agroforestrynet.eu/	AFINET HANDBOOK	485	211	0	0	0	0	1	3	0	0	0	0	0	0	0	0
https://euraf.isa.utl.pt/de/afinet	AFINET DE	320	225	0	0	0	0	266	32	0	0	76	20	0	0	6	0
https://euraf.isa.utl.pt/el/afinet	AFINET EL	264	1245	2	0	0	1	439	61	0	0	95	22	0	0	6	0
https://euraf.isa.utl.pt/fr/afinet	AFINET FR	480	993	2	0	0	1	494	64	0	0	91	22	0	0	6	0
https://euraf.isa.utl.pt/fi/afinet	AFINET FI	495	1074	2	0	0	1	593	64	0	0	108	22	0	0	6	0
https://euraf.isa.utl.pt/nl/afinet	AFINET NL	353	339	0	0	0	0	350	45	0	0	85	22	0	0	6	0
https://4d4f.eu/	4D4F	129	348	2	2	0	1	93	62	0	0	20	0	0	0	0	0
https://euraf.isa.utl.pt/hu/afinet	AFINET HU	283	388	0	0	0	0	379	48	0	0	89	22	0	0	6	0
https://euraf.isa.utl.pt/it/afinet	AFINET IT	393	250	0	0	0	0	520	41	0	0	100	22	0	0	6	0
https://euraf.isa.utl.pt/pl/afinet	AFINET PL	375	1303	2	0	0	1	399	53	0	0	91	22	0	0	6	0
https://euraf.isa.utl.pt/pt-pt/afinet	AFINET PT-PT	451	212	0	0	0	0	594	40	0	0	104	22	0	0	7	0
https://euraf.isa.utl.pt/es/afinet	AFINET ES	411	1257	2	0	0	1	501	59	0	0	99	22	0	0	6	0
https://euraf.isa.utl.pt/afinet	AFINET EN	414	1264	2	0	0	1	517	62	0	0	97	22	0	0	7	0
https://www.encyclopediapratensi s.eu/product- category/inno4grass/?filter_langua	Inno4grass Encyclopedia pratensis	470	59	0	0	0	7	221	25	0	0	22	0	0	0	0	0
https://www.encyclopediapratensi s.eu/media/	Inno4grass Encyclopedia pratensis media	316	20	0	0	0	0	46	15	0	0	220	0	0	0	0	0
http://agroforestrynet.eu/afinet/	AGROFERESTRYNET (AFINET)	369	941	2	0	0	1	248	41	0	0	86	18	0	0	3	0
https://hennovation.eu/	HENNOVATION	120	0	0	0	0	0	35	5	0	0	0	0	0	0	0	0
http://www.shortfoodchain.eu	SKIN	342	83	0	0	0	0	321	150	0	0	10	0	0	0	3	0
https://www.henhub.eu/	HENHUB (Hennovation)	449	36	0	0	0	0	92	6	0	0	5	0	0	0	0	0

Figure 24. The pool of data objects to consider for the e-KRP repository

6.3. Assembling the input for the e-KRP repository

The assembly of the input for the repository (i.e., the data objects to be integrated into the repository and delivered via the e-KRP) has been a core methodology phase, which involved three main activities:

- The cleaning of the data objects harvested by the EURAKNOS web crawler.
- The selection of the data objects to integrate into the repository by taking account of the content-related preferences of the various end-user types (this information is provided by Deliverable 3.2[37]).
- The annotation of data objects (selected as input for the repository) with metadata values, based on the set of metadata instructed by the EURAKNOS data model.

The above activities have been executed for all the types/categories of the data objects identified and harvested by the crawler and therefore, considered for the e-KRP repository. These categories/types are: documents, presentations, videos, audio-based objects (podcasts), and images (infographics).

[37] Van Hal, D. et al. (2020). Report on HIKR Design. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 3.2 – WP3.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

6.3.1. Data cleaning

To better frame the context (and thus, help establish a deeper understanding) of this activity, it needs to be made clear what the outcome of the crawling process has been (i.e., what has been the output delivered by the EURAKNOS web crawler). As regards the general principles of the implementation and deployment of a web crawling application, what is being explicitly mentioned in Subsection 6.2.1 ("The EURAKNOS web crawler") is that "*the information/content that a web crawler can harvest from the Web depends on the instructions provided in its development phase*". In addition to the above, it needs to be noted that the EURAKNOS data model refers to specific data object categories/types[38] with each of them being associated with a range of file formats. However, the web crawler has not the capacity to "interpret" the data model's "abstract" types/categories as the types/categories of objects to search for, but rather "understand" rules relating to specific file formats. Thus, the "instructions" coded into the EURAKNOS web crawler have been to search for specific file formats. Table 9 lists the file formats of the data objects that were harvested by the EURAKNOS web crawler together with the categories/ types of data objects they are associated with.

Table 9. The file formats of the data objects harvested by the EURAKNOS crawler and the data object types/categories they are associated with

File formats of the data objects harvested by the EURAKNOS crawler	Type/category of data object
doc/docx, html, pdf, xls/xlsx	Document
ppt/pptx, pdf, prezi	Presentation
jpg, png, svg	Image
Video-related file formats supported by the YouTube[39] and Vimeo[40] platforms	Video

What is made evident from the above table is that some of the data object types/categories described in the EURAKNOS data model are missing. Specifically, there are no mentions of datasets and software applications. This is because no such data objects have been available in the KRs of the (sample) TNs. The same has been the case with audio-based content as well. Thus, the only instances of audio-based material delivered from the e-KRP are the EURAKNOS podcasts, mostly available in an mp3 file format, which have been manually retrieved.

Based on the descriptions provided above and given the long list of results available by the EURAKNOS web crawler, the data cleaning activity has involved the following steps:

- **STEP #1**: Association of the data objects harvested with the appropriate type[41] (classification task).
- **STEP #2**: Removal of duplicates (there were cases in which a data object existed more than once in the list of the crawling results).
- **STEP #3**: Identification of data objects available in file formats other than those that are primarily considered per data object type/category.

[38] The data object types/categories described in the EURAKNOS data model are documents, presentations, images, videos, audios, datasets, and software applications.

[39] https://www.youtube.com/

[40] https://vimeo.com/

[41] As an example, we can refer to the case of "factsheet", which is a data object type associated with the "document" category of data objects. For details about the range of the types/categories of data objects (named as "kinds" in the EURAKNOS data model) and their associated types, the reader may refer to Deliverable 3.2 ("Report on HIKR design").

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

In the case of data object file formats with regard to which a large number of results has been provided by the EURAKNOS crawler (documents available as pdf files[42] and videos[43]), the steps listed above were automatically implemented through the execution of specially designed code scripts[44]. In cases of file formats with not so many results retrieved (presentations available as ppt/pptx files and infographics), the above steps have been manually executed as part of the annotation process.

Concerning data objects in a pdf format, the basic idea behind STEP #1 has been to classify each object into the correct document-related type. According to the EURAKNOS data model, the types that a document may belong to are: "article in conference proceedings", "book", "booklet", "brochure/flyer", "chapter in edited volume", "deliverable report", "factsheet", "handbook", "guide", "journal article", "manual", "milestone report", "newsletter", "practice abstract", "press release", "review document", "spreadsheet", "report/paper", "technical article", "technical information/specifications card", "tutorial", and "other"[45].

Given the large number of the data objects retrieved by the EURAKNOS web crawler in a pdf file format, the classification task has been implemented in an automated way by running Python scripts especially developed for each TN in our sample. These scripts have a name of the form "<TN_name>_pdfs.ipynb" (e.g., the Python script that has been executed for the needs of classifying the pdf files available from the FERTINNOWA TN is named as "FERTINNOWA_pdfs.ipynb"[46]) and an indicative example is provided in Annex II.

The algorithm used for the classification of pdf documents into the appropriate document-related type is explained in Table 10 below. This automated classification process has also been complemented by a manual check of the accuracy of its results, undertaken by the annotators who were involved in the annotation activity[47].

After having associated the pdf files retrieved by the EURAKNOS web crawler with the appropriate (i.e., the "correct") document-related types, the next step was to remove duplicates. In the case of pdf files (i.e., documents), this duplicate removal process was implemented in an automated way through the execution of appropriately developed Python scripts. More specifically, a Python script was developed for each TN. Each of these scripts (used for the needs of the duplicate removal process) has a name of the form "Check4duplicates_in_<TN_name>_csv_files. ipynb"[48].

In the case of videos (i.e., a category of data objects with a large number of items/results harvested by the EURAKNOS web crawler as well), the cleaning process has involved the removal of duplicate entries from the available lists of results. This process has been implemented for videos created by the TNs in our sample and disseminated both from the YouTube and Vimeo platforms. The removal of duplicates has been made in an automated way through the execution of the "selecting_youtube_videos_for_ the_EURAKNOS_repo.ipynb" and the "selecting_vimeo_videos_for_the_EURAKNOS_repo.ipynb"[49] scripts, which have been particularly developed for this purpose. An indication of how important the

[42] The total number of pdf files harvested by the EURAKNOS web crawler is 16,935. This has been the majority of document-related data objects harvested by the EURAKNOS web crawler. Some documents available as doc or docx files have also been identified. However, given the fact that they were too few, they were not considered for the assembly of the input for the e-KRP repository.

[43] The initial list of videos retrieved by the EURAKNOS web crawler included more than 1,000 entries.

[44] The code scripts used for this purpose have been developed in Python.

[45] The "other" category has been used for the documents not qualifying for any of the document-related types defined in the EURAKNOS data model.

[46] The code development environment that has been used for the creation of the Python code scripts is Jupyter notebooks (https://jupyter.org/).

[47] The process of annotation is described in detail in Subsection 5.3.3.3 ("The process of the annotation of data objects with metadata values").

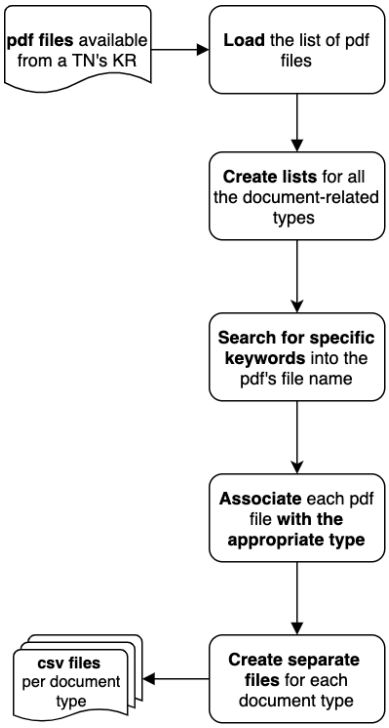
[48] An example case of such a Python script is provided in Annex III.

[49] The "selecting_youtube_videos_for_ the_EURAKNOS_repo.ipynb" Python Script is available in Annex IV.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

data cleaning activity has been is that only 404 items remained as videos to consider for the repository, from the initially available list of 949 links to YouTube videos harvested from the web crawler. Similarly, from the 80 Vimeo videos that were initially harvested, only 9 remained after removing the duplicates. The duplicate-free lists of the 404 YouTube and 9 Vimeo videos were the input to the activity relating to the selection of the video data objects to integrate into the e-KRP repository.

Table 10. The algorithm used for associating the pdf files harvested by the EURAKNOS web crawler with the correct document-related type

Algorithm overview	Explanation of algorithm steps
	The first step in the algorithm is to load and read all the pdf files that the EURAKNOS web crawler has harvested from each TN's KR. These lists have been made available as csv files and loaded into memory with the help of the Python Pandas Dataframe[50] data structure.
	For each of the types of documents described in the EURAKNOS data model, an empty list is created. By the time the implementation of the algorithm was finished, each list contained pdf files associated with their corresponding type.
	The algorithm searches for pdf files associated with each document type. The association has been intended to be made by seeking matches of the following keywords (indicative of the types described in the data model) in the pdf file names: "practice", "factsheet", "press", "newsletter", "tech", "guide", "article", "deliverable", "report", "handbook", "policy", "brief", "journal", "conference", "manual", "review", "book", "booklet", "paper", "tutorial", "proceedings", "information", "milestone", "broch", "flyer", and "research".
	Based on the keyword matches found (i.e., keywords of the above list found in the file name of a pdf file), a pdf file is classified into the appropriate document-related type. The classification is implemented by appending the specific pdf file in the respective list.
	For each type of document, a separate csv file is created. This file contains the records of the pdf files belonging to the specific category. For instance, the csv files that have been created for the FERTINNOWA TN are: "Fertinnowa**policy**_pdfs.csv", "Fertinnowa**newsletter**_pdfs.csv", etc. For instance, the Fertinnowa **newsletter**_pdfs.csv" file, contains the newsletters harvested from the KR of the FERTINNOWA TN.

In the cases of presentations, audio-based data objects, and images[51], the classification, and duplicate removal have been implemented in a manual way as part of the process of annotating the data objects with metadata.

The final step in the activity of data cleaning has been the identification of data objects available in file formats other than those primarily considered per data object category. For instance, the primary file formats of presentations are ppt and pptx. Yet, in the bulk of data objects harvested by the EURAKNOS crawler, there have also been presentations available in a pdf file format. The same has been the case with infographics as well. The "Presentations_in_pdf_format.ipynb"[52] and "Infographics_in_pdf_format.ipynb"[53] are Python scripts that were created and executed so as to identify presentations and infographics available in a pdf file format.

6.3.2. Selecting the data objects to integrate into the repository

The selection of the data objects that have constituted the input for the repository of the prototype e-KRP (i.e., the data objects integrated into the repository) has been based on considerations and insights

[50] For details, the reader may refer to https://pandas.pydata.org/pandas-docs/stable/user_guide/dsintro.html.

[51] The types of images we have been interested in in EURAKNOS are infographics.

[52] The Python script is available in Annex IV.

[53] The Python script is available in Annex IV.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

regarding the types and formats of the data objects helping to address the end-user preferences and needs. This information has been made available from the work reported by Van Hal et al., 2020 and can be summarised to the following points:

- Content and information delivered via video-based data objects is preferred the most by all types of end-users.
- There is a range of content and information provided in text-based formats (e.g., practice abstracts, factsheets, booklets, handbooks, brochures, guides, tutorials, education/training material) that can cover end-user needs; documents are a category of data objects that appears to be highly valued by both farmers and advisors.
- Evidence for preferences in slideshow-based material (i.e., presentations) has been made available only for advisors. Advisors are more in favour of slideshow-based material than content delivered in textual formats.
- Preferences in the image- and audio-based content/information have been reported only in the case advisors. Audio-based material (e.g., podcasts) conveying information related to innovative/ best practices, as well as guidelines, appears to be close to their preferences and needs.

On top of the content-related considerations highlighted above, some additional points have also been needed to be taken into account:

- Evidence for preferences in slideshow-based (i.e., presentations), as well as image- and audio-based content and information is only available for advisors. However, we may assume that these kinds/ formats of digital outputs could also be of relevance to farmers and foresters (by taking account of the tight interactions between advisros and farmers/foresters).
- The majority of digital outputs provided by TNs relates to types associated with text-based formats (e.g., practice abstracts, factsheets, press releases, research/technical articles, handbooks, reports, and guides). Video-based content and information is available at a lower rate (it counts for 41.7% of the digital outputs available). The same applies to image-based content and information, as well as podcasts (i.e., audio-based content and information) which count for 33.3% and 20.8% of the total content respectively (Mosquera Losada et al., 2019).

Given the above, the sample of data objects to be integrated into the repository could consist of:

1) **videos** of any type;
2) **documents** with an emphasis on document types such as practice abstracts, factsheets, tutorials, manuals, guides, handbooks, technical articles, booklets, brochures, newsletters, press releases, journal articles, chapters in edited volumes, articles in conference proceedings, deliverable reports, reports/papers, and books[54];
3) **presentations** with a focus on collecting (from the TN KRs) items relating to presentation types such as guides, tutorials, and educational/training presentations (if any and based on the extent to which presentations are available from the KRs of TNs).
4) **images** of any type (our core focus has been on infographics);
5) **audio-based outputs** with a focus on collecting (from the TN KRs) audio outputs like educational/ training podcasts, on-demand seminars, and guides (if any and based on the extent to which audio-based content is available from the KRs of TNs).

The above are also illustrated with the help of Figure 25 (the heights of the boxes are indicative of the importance that each type/category of data objects has for the types of the end-users targeted).

[54] By taking account of the fact that articles in conference proceedings, books, chapters in edited volumes, deliverable reports, journal articles, and reports/papers are not the most preferable types of data by the e-KRP end-user types targeted, only a few data objects from this subset will be collected and stored into the HIKR.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

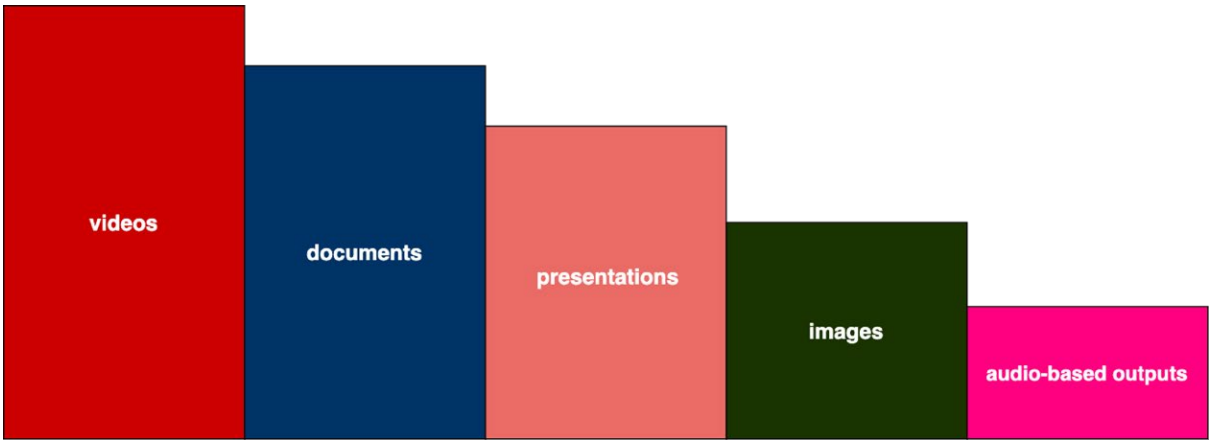


Figure 25. The data object categories integrated into the repository; the box heights are indicative of the relevance/importance of each category to the end-user preferences and needs

Before providing the details of the selection of data objects for the repository, it should be highlighted that the data selection process has been applied only for document- and video-related data objects, i.e. the data object categories with the largest numbers of items harvested by the EURAKNOS crawler. In the cases of presentations (i.e., slideshow-based material), as well as image- and audio-based data objects, all the items harvested by the crawler (or manually retrieved) have been considered as input for the e-KRP repository. The rationale behind this decision has been twofold: (i) the number of items relating to these categories is small; and (ii) there is not much evidence (from the research undertaken in EURAKNOS) for the preferences of end-users in these categories/types of data objects.

As regards text-based content coming from the KRs of the sample TNs (i.e., documents of various types available as pdf files[55]), the criteria employed for selecting data objects have been both qualitative and quantitative. These criteria were not intended to be used as strict rules, but rather as heuristics for the selection of content addressing the needs and preferences of the end-users in the best possible way. More specifically:

- There has been an intention to have a representation of as many document types as possible from those identified as important to the targeted end-user categories.
- Given the fact that EURAKNOS has aimed to the design and development of a prototype platform, a maximum number of 100 documents per TN[56] has been regarded as adequate for input as part of the proof-of-concept "exercise" undertaken.
- The number of documents classified into the "other" category, which should be considered as input for the repository, should not exceed those associated with the document types identified in the EURAKNOS data model (especially in cases in which the number of documents belonging to "other" category are more than those classified into the document types identified in the data model). For this purpose, the following rule of thumb has been applied:

  $$number\ of\ "other"\ documents\ = 0.6 \cdot (total\ number\ of\ items\ belonging\ to\ the\ document\ types)$$ [57]

[55] The EURAKNOS web crawler has also retrieved docx and xlsx files (also falling under the "document" category of TN-related data objects), which, however, have not been considered due to the very small number of items harvested in these formats.

[56] There are cases of TNs from our sample in which less that 100 documents have been available in total after the data cleaning process. In some other cases, we have been able to come up with more than 100 documents, from one TN's KR, to consider for the repository. Quantitative evidence is provided in tables 10 to 17.

[57] In cases in which the total number of documents belonging to the types defined in the EURAKNOS data model, which should be considered as input for the e-KRP, has been close to 100, the number of items falling into the "other" category has been decided to be the number of documents required to reach the threshold of 100 documents per TN. This has been the case of AFINET (see Table 17).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- In cases of document-related types with more than 100 items (i.e., $number\ of\ items > 100$), a 10% rate has been regarded as adequate (from the specific type) to be made available as input to the repository.
- In cases of document-related types with more than 10 and less than or equal to 100 items (i.e., $10 < number\ of\ items\ \leq 100$), a 40% rate has been regarded as adequate (from the specific type) to be made available as input to the repository.
- In the case of brochures/flyers (a specific type of document purely used for advertorial purposes), the rule applied for calculating an adequate number of items to consider as input for the repository was the following:
  - If $number\ of\ flyers\ \leq 5 \rightarrow$ all items should be considered as input for the e-KRP repository.
  - If $5 < number\ of\ flyers\ \leq 20 \rightarrow$ a 25% rate of them should be considered as input for the e-KRP repository.
  - If $20 < number\ of\ flyers\ \leq 40 \rightarrow$ a 15% rate of them should be considered as input for the e-KRP repository.
  - If $number\ of\ flyers > 40 \rightarrow$ a 10% rate of them should be considered as input for the e-KRP repository.

Tables 11 to 18 below show the number of documents per type and per TN available after the cleaning process together with the number of documents to be considered as input for the e-KRP repository. A point to be taken into account is that not all the document-related types described in the EURAKNOS data model are covered by the documents harvested from our sample's KRs. For instance, policy briefs have not become available from the KR of the OK-NET Arable TN (see Table 12 below).

Table 11. Types and numbers of documents harvested from the FERTINNOWA TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Press release	3	3
2	Policy brief	2	2
3	Newsletter	1	1
4	Booklet	1	1
5	Brochure/flyer	3	3
6	Book	2	2
7	Report (any report)	1	1
8	Technical/technology article	11	11
9	Article (any article)	3	3
10	Deliverable	9	9
11	Paper/research publication	1	1
12	Other	14	14
	TOTAL	51	51

Table 12. Types and numbers of documents harvested from the OK-NET Arable TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Guide	4	4
2	Technical/technology article	3	3
3	Handbook	1	1
4	Brochure/flyer	1	1
5	Manual	3	3
6	Paper/research publication	1	1

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

7	Other	6	6
	TOTAL	19	19

Table 13. Types and numbers of documents harvested from the Smart-AKIS TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Article (any article)	1	1
2	Press release	14	14
3	Policy brief	3	3
4	Guide	1	1
5	Paper/research publication	8	8
6	Manual	1	1
7	Factsheet	1	1
8	Review document	2	2
9	Brochure	3	3
10	Technical information/ specifications card	956	96 ($956 \cdot 0.1 = 95.6 \cong 96$)
11	Report (any report)	15	15
12	Flyer	3	3
13	Other	90	-
	TOTAL	1098	148

Table 14. Types and numbers of documents harvested from the SKIN TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Manual	1	1
2	Report (any report)	3	3
3	Press release	4	4
4	Technical/technology article	1	1
5	Paper/research publication	2	2
6	Newsletter	1	1
7	Deliverable	5	5
8	Guide	3	3
9	Handbook	1	1
10	Brochure/flyer	2	2
11	Other	54	14 ($23 \cdot 0.6 = 13.8 \cong 14$)
	TOTAL	77	37

Table 15. Types and numbers of documents harvested from the 4D4F TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Report (any report)	5	5
2	Paper/research publication	6	6
3	Guide	21	21
4	Technical/technology article	1	1
5	Other	151	20 ($33 \cdot 0.6 = 19.8 \cong 20$)
	TOTAL	184	53

Table 16. Types and numbers of documents harvested from the Inno4Grass TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Factsheet	2	2

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

2	Deliverable	1	1
3	Booklet	4	4
4	Press release	3	3
5	Brochure	1	1
6	Report (any report)	2	2
7	Paper/research publication	3	3
8	Guide	1	1
9	Technical/technology article	53	$21\ (53 \cdot 0.4 = 21.2 \cong 21)$
10	Flyer	8	$2\ (8 \cdot 0.25 = 2)$
11	Other	309	$24\ (40 \cdot 0.6 = 24)$[58]
	TOTAL	**387**	**64**

Table 17. Types and numbers of documents harvested from the SheepNet TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Paper/research publication	4	4
2	Press release	8	8
3	Policy brief	15	15
4	Guide	3	3
5	Practice abstract	16	16
6	Newsletter	6	6
7	Report (any report)	3	3
8	Technical/technology article	5	5
9	Factsheet	27	$11\ (27 \cdot 0.4 = 10.8 \cong 11)$
10	Flyer	1	1
11	Other	469	28
	TOTAL	**557**	**100**

Table 18. Types and numbers of documents harvested from the AFINET TN's KR and considered as input for the e-KRP repository

#	Type of document	Number of items harvested by the crawler	Number of items considered as input for the repository
1	Policy brief	7	7
2	Practice abstract	4	4
3	Newsletter	16	$6\ (16 \cdot 0.4 = 6.4 \cong 6)$
4	Press release	3	3
5	Deliverable	2	2
6	Book	3	3
7	Factsheet	152	$15\ (152 \cdot 0.1 = 15.2 \cong 15)$
8	Paper/research publication	9	9
9	Report (any report)	3	3
10	Brochure	1	1
11	Manual	1	1
12	Article (any report)	6	6
13	Technical/technology article	14	14
14	Guide	7	7
15	Review document	2	2
16	Flyer	2	2
17	Other	68	15[59]

[58] The number of documents to be considered from the "other" category is the equal to the rate of 60% of the documents coming from the document-related types defined in the EURAKNOS data model.
[59] Number of documents coming from the TN's KR (and belonging to the "other" category) in order to reach the threshold of the 100 items to consider (from the specific TN's KR) for the e-KRP repository.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

TOTAL	300	100

Before providing a summary of the total number of documents per type that were selected as the input for the e-KRP repository, the case of practice abstracts should be highlighted as a highly important type of document (i.e., text-based content) that has been considered for our repository. Practice abstracts have been intended to be automatically identified from within the bulks of the pdf documents available per TN (as a result of the crawling process) by seeking potential matches of the term "practice" in the file name of the pdf files. As made evident from the tables above, in some TN cases there has been an identification of documents having the word "practice" in their file name. This finding has served as an indication of the potential retrieval of practice abstracts. The manual check of the classification results, which took place during the annotation, aimed to identify misclassification cases including conclusions about whether documents with the term "practice" in their file names are indeed practice abstracts.

Given the importance of practice abstracts (a dissemination output of importance for TNs), the process of collecting practice abstracts[60] to integrate into the repository took place by: (i) coming into contact with representatives of the TNs (which have been entities also involved in EURAKNOS); or (ii) retrieving practice abstracts from the EIP-AGRI database[61]. Table 19 provides the numbers of practice abstracts per TN that have been considered as input for the e-KRP repository. The threshold applied in this case was 50 practice abstracts per TN at most.

Table 19. The number of practice abstracts per TN that have been considered as input for the e-KRP repository

#	Thematic Network	Number of practice abstracts
1	FERTINNOWA	43
2	OK-NET Arable	41
3	SKIN	50
4	4D4F	24
5	Inno4Grass	25
6	SheepNet	25
7	AFINET	26
TOTAL		234

The figures appearing in Table 19 about the practice abstracts considered from the SheepNet, AFINET, Inno4Grass, and 4D4F TNs are actually double the size because each of the practice abstracts retrieved from the KRs of those TNs is available in two languages.

The total numbers of documents per type (for all the sample TNs) are summarised in table 20.

Table 20. Number of documents per type available in total

#	Type of document	Number of items considered for the repository	#	Type of document	Number of items considered for the repository
1	Press release	35	10	Deliverable	17
2	Policy brief	27	11	Paper/research publication	34
3	Newsletter	14	12	Guide	40
4	Booklet	5	13	Handbook	2
5	Brochure/flyer	19	14	Manual	6
6	Book	5	15	Factsheet	29
7	Report (any report)	32	16	Review document	4

[60] The manual evaluation of the automated classification results that took place during the annotation process revealed that the documents containing the term "practice" in their file name were not practice abstracts and had been falsely classified in the specific category.

[61] https://ec.europa.eu/eip/agriculture/en/about/thematic-networks-%E2%80%93-closing-research-and

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

8	Technical/technology article	56	17	Technical information/ specifications card	96
9	Article (any article)	10	TOTAL		335

The number of documents that were classified into the "other" category as the result of the automated classification process was 121. As explained in Subsection 6.3.3 ("Providing metadata values to the data objects"), this list of 121 "other" documents served as a back-up pool of text-based data objects to be considered in the case that faulty documents were identified during the annotation process.

As regards the other types/categories of data objects (namely, videos, presentations, as well as image- and audio-based material), the numbers of items regarded as input to the e-KRP repository have been the following:

- videos: 209
  - videos available on the YouTube platform: 202
  - videos available on the Vimeo platform: 7
- presentations: 39 (both in ppt/pptx and pdf file formats).
- images (infographics): 24 (only available in a pdf file format)
- audio-based material (podcasts): 24 (only available from EURAKNOS; not available from the KRs of the sample TNs)

As mentioned above, apart from documents, videos are the second data object category for which a selection process has been applied to conclude to the set to finally integrate into the e-KRP repository. In this case, the applied criterion was purely quantitative; i.e., the selection of videos has been based on the number of views (namely, the number of video views at the time by which the crawling process took place)[62]. The number of views used as thresholds for the selection of YouTube and Vimeo videos has been 118.5 and 188.0 respectively. These values have been decided after the extraction of some descriptive statistics related to video views (Table 21).

Table 21. Descriptive statistics related to the number of YouTube and Vimeo videos harvested from the TNs' KRs

	YouTube videos		Vimeo videos	
Descriptive statistics	count (total number of videos)	404.0	count (total number of videos)	9.0
	mean (average number of views)	4311.7	mean (average number of views)	401.2
	std (standard deviation)	44778.2	std (standard deviation)	472.9
	min (minimum number of views)	5.0	min (minimum number of views)	59.0
	25% (first quartile)	34.0	25% (first quartile)	188.0
	50% (median)	118.5	50% (median)	225.0
	75% (third quartile)	418.3	75% (third quartile)	267.0
	max (maximum number of views)	872734.0	max (maximum number of views)	1526.0

As made evident from the table above, the number of views, which have been used as threshold values for the selection of the YouTube and Vimeo videos to make available through the e-KRP, is the median value, in the case of YouTube videos, and the first quartile, in the case of Vimeo videos. In other words, half of the YouTube videos have been considered as input to the repository. Regarding Vimeo videos, three-quarters of them (75% of the total number of Vimeo videos) were regarded as input. The reason for choosing a different threshold value for the Vimeo videos has been: (i) their small number (only 9 Vimeo videos were shortlisted after the data cleaning activity); and (ii) the more limited range of views (range in the number of Vimeo video views of = 1,467 vs. range in the number of YouTube video views = 872,729).

[62] May and June, 2020.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Regarding the presentations, all the outcomes of the crawling process have been considered as input. Infographics (available in a pdf file format[63]) are the only image types that have been accounted for. Finally, the podcasts considered (and made available via the e-KRP) were those created in EURAKNOS (no podcasts were harvested and retrieved from the KRs of the sample TNs as a result of the crawling process).

From the evidence provided above it is made clear that a total number of 631[64] data objects have been considered as input to the e-KRP repository. This set of data objects was given as input to the activity of annotation. However, not all the 631 data objects (outputted from the data selection activity) were finally annotated with metadata and integrated into the repository. As explained next, the reason for this was the identification of more duplicates (duplicate data objects appearing in different file names) or faulty data objects (e.g., dead links or broken documents).

6.3.3. Providing metadata values to the data objects

The annotation of the data objects delivered from the e-KRP with metadata was the last activity of the methodology phase related to the assembly of the input for the e-KRP repository. This activity involved three distinct steps: (i) the update of the EURAKNOS data model (and thus, the creation of version 2.0 of the EURAKNOS data model) to align some of the properties described in it with properties available from widely-adopted ontologies (in an attempt to better adhere to the FAIR Data principles[65]); (ii) the definition of the metadata to be used for the data object annotation; and (iii) the process of annotation of the data objects with metadata.

The annotation of data objects with the appropriate metadata values was the final activity before their integration into the repository. All the steps taken as part of this effort are described below.

6.3.3.1 Updating the EURAKNOS data model

The EURAKNOS data model and the steps taken to develop it are described in detail in Deliverable 3.2 (Van Hal et al., 2020). A diagrammatic representation of it is provided in Figure 26 below. As explicitly mentioned in Deliverable 3.2 (Van Hal et al., 2020), "*the data model needs to include the definitions of entities, relations between entities, and entity attributes helping to frame the context related to TN outputs*." The rationale for developing the EURAKNOS data model has been to identify and define the (bare minimum) set of metadata for effectively and efficiently characterising the data objects available from the e-KRP repository[66]. Delivering data objects that have been annotated with rich metadata is a requisite for facilitating successful search operations and an advanced user experience.

However, before the definition of the metadata to consider for the data object annotation, as well as the annotation process itself, it was necessary to update our data model by taking account of existing, widely-adopted ontologies. Such an ontology (the Schema.org ontology) had indeed been considered for the data model's creation. The update performed at this stage intended to align specific properties in the data model with relevant properties defined in external and widely-adopted ontologies. These ontologies are FOAF (http://xmlns.com/foaf/spec/) and the Dublin Core Metadata Initiative (https://dublincore.org/specifications/dublin-core/dcmi-terms/). This alignment was pursued in an attempt to strengthen the adherence of our conceptual design to the FAIR Data principles. Specifically, to enhance the our repository's FAIRness, it was needed to shift from custom properties to properties defined in well-known, widely-adopted ontologies.

[63] They have been retrieved from the bulk of data objects harvested by the EURAKNOS web crawler by executing the "Infographics_in_pdf_format.ipynb" Python scripts (see Subsection 6.3.1).

[64] The documents belonging to the "other" category are not included in this set of 631 data objects.

[65] Detailed definitions and examples of the FAIR Data principles are available at https://www.go-fair.org/fair-principles/.

[66] The digital outputs of Thematic Networks as they are called in the excerpt quoted from Deliverable 3.2.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

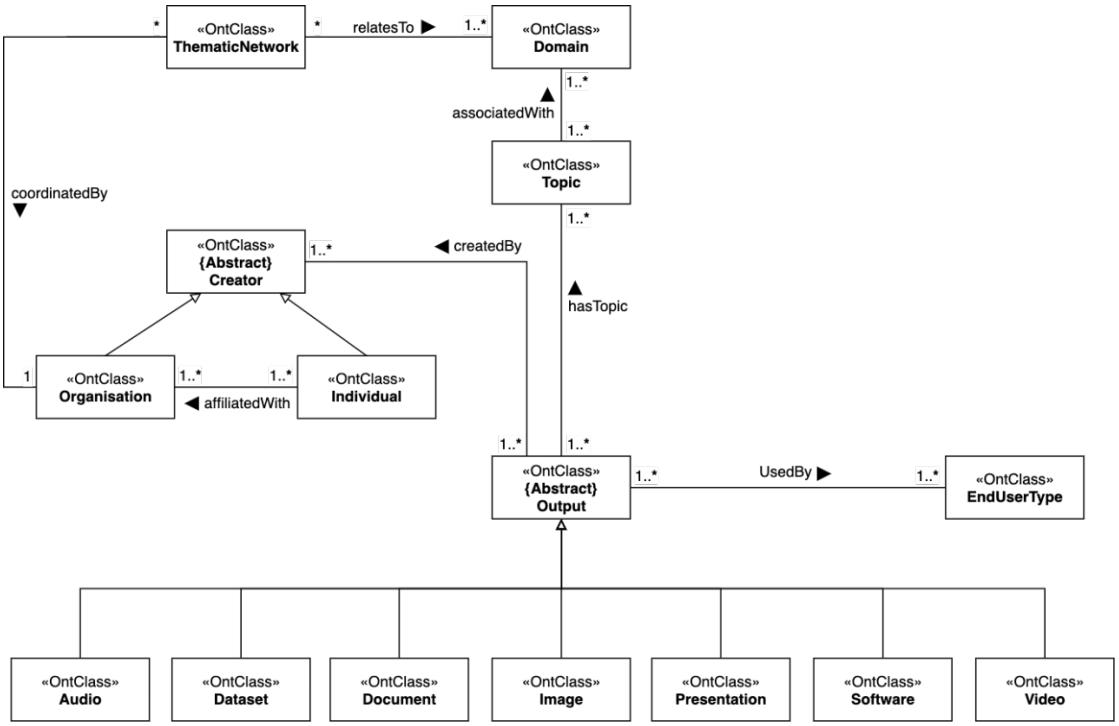


Figure 26. The EURAKNOS data model

It needs to be noted that the EURAKNOS data model identifies a large number of relations/properties, but not all of them were finally considered for the definition of the metadata to use for the annotation of data objects. As described below, the idea has been to conclude to a bare minimum set of metadata properties with the capacity to homogeneously characterise the various data objects outputted by the various Thematic Networks. The data model properties that have been finally considered for the needs of metadata definition are listed in Table 22. The table rows highlighted in light grey contain the model properties that were updated by defining them based on relevant properties from external ontologies. The table contains the property name, the property type (object or datatype property)[67], the domain and range of the property, and the URI[68] of the associated, from an external vocabulary, property.

Table 22. Alignment of relations/properties described in the EURAKNOS model with properties defined in external vocabularies

#	Property in the EURAKNOS data model	Type of property	Domain	Range	Matching property from external vocabulary
1	description	Datatype property	Output class	xsd:string[69]	http://schema.org/description
2	keywords	Datatype property	Output class	xsd:string	http://schema.org/keywords
3	createdBy	Object property	Output class	Creator class	http://schema.org/creator
4	geographicLocation	Object property	Output class	List of European countries provided in the data model	http://schema.org/contentLocation

[67] For a detailed discussion on the two different types of properties identified and used in the context of ontology engineering, the reader may refer to the document "A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools - Edition 1.1" available at http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_1.pdf.

[68] URI stands for Uniform Resource Identifier; for a detailed discussion on URIs and their utility in the context of Semantic Web technologies, the reader may refer to https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.

[69] "xsd:" is the prefix used for denoting the XML Schema namespace (i.e., http://www.w3.org/2001/XMLSchema#) in a URI. The XML Schema definition language enables to define the structure and data types for XML documents.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

5	url	Datatype property	Output class	xsd:anyURI	http://schema.org/url
6	hasTopic	Object property	Output class	Topic class	http://xmlns.com/foaf/0.1/topic
7	purpose	Object property	Output class	List of purposes defined in the data model	http://schema.org/potentialAction
8	format	Object property	Output class	List of file formats defined per category of data object	http://purl.org/dc/terms/format
9	fileSize	Datatype property	Output class	xsd:string	http://schema.org/fileSize
10	dateCreated	Datatype property	Output class	xsd:string	https://schema.org/dateCreated
11	license	Datatype property	Output class	xsd:string	http://schema.org/license
12	ThematicNetwork Acronym	Datatype property	Thematic Network class	xsd:string	http://schema.org/alternateName
13	ThematicNetwork FullName	Datatype property	Thematic Network class	xsd:string	http://schema.org/name
14	ThematicNetwork Url	Datatype property	Thematic Network class	xsd:anyURI	http://schema.org/url

Version 2.0 of the EURAKNOS data model and the documentation of it have been published in Zenodo and can be accessed at https://zenodo.org/record/5500341 and https://zenodo.org/record/5500348 respectively.

6.3.3.2 The EURAKNOS metadata set

The metadata that has been used to annotate the data objects available via the e-KRP has been defined by drawing on the updated EURAKNOS data model. Based on that, AUA drafted an initial metadata set to consider for the annotation process. This draft set of metadata was brought to discussion with AKI and MU[70]. Discussions were held in two virtual meetings that took place in July 2020. In these meetings, the participants that were involved had the chance to review the proposed metadata, comment on it, and suggest changes and adaptations, as well as metadata to further consider (i.e., metadata that were not included in the initial set).

The result of these interactions has been the list of metadata shown in Table 23. The list contains the metadata properties that have been used to annotate the data objects considered as the input for the e-KRP repository. Apart from each metadata property's name, there is also a short definition provided together with the property's URI (in case there is a match with a property from an external ontology). In the case that there is not a match of a metadata property with a property from an external ontology, the "no relevant property" indication is being provided.

The metadata in Table 23 constitutes the minimum set of properties that have been identified for the needs of a homogeneous description of all the data objects housed in the repository of the e-KRP. The development of this metadata set is a major accomplishment of the work in EURAKNOS, because it has allowed for the uniform annotation of all the data objects despite the heterogeneity in their types and formats.

[70] As mentioned in Section 2, the University of Maastricht (MU) is not a partner in EURAKNOS. MU leads the EUREKA project's WP that is related to the selection of the data objects to integrate into the FarmBook's repository, the data quality evaluation (including FAIRness), and the repository design (WP3). Given the fact that the work done in EURAKNOS is being taken up by EUREKA, the participation of MU in technical discussions and decisions would help the smooth transition from the one project to the other.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Table 23. The metadata properties used to annotate all the data objects available through the EURAKNOS e-KRP

#	Metadata property	Definition	Matching property from external ontology	Property URI
1	title	The title of the content available from the data object.	no relevant property from external vocabulary	
2	description	A short textual description of what the content of the data object is about.	**schema:**description	http://schema.org/description
3	keywords	Keywords or tags used to describe the content of the data object.	**schema:**keywords	http://schema.org/keywords
4	creator	The creator/author of the data object.	**schema:**creator	http://schema.org/creator
5	inLanguage	The language in which the content of the data object is available.	no relevant property from external vocabulary	
6	contentLocation	The location depicted or described in the content.	**schema:** contentLocation	http://schema.org/contentLocation
7	fileUrl	The URL allowing to download the data object.	**schema:**url	http://schema.org/url
8	topic	The topic addressed in the content of a data object.	**foaf:**topic	http://xmlns.com/foaf/0.1/topic
9	purpose	The purpose for which the data object has been created.	**schema:** potentialAction	http://schema.org/potentialAction
10	fileFormat	The file type of the data object.	**dct:**format	http://purl.org/dc/terms/format
11	type	The type property relates to what the content of the data object is about.[71]	no relevant property from external vocabulary	
12	fileSize	The size of the file/package. In the absence of a unit, KB is assumed.	**schema:**fileSize	http://schema.org/fileSize
13	fileChecksum	A sequence of characters used to check a data object for errors.	no relevant property from external vocabulary	
14	dateCreated	The date on which a data object was created.	**schema:** dateCreated	https://schema.org/dateCreated
15	license	A license document that applies to this content, typically indicated by URL.	**schema:**license	http://schema.org/license
16	ThematicNetwork Acronym	The acronym of the Thematic Network (e.g., AFINET).	**schema:** alternateName	http://schema.org/alternateName

[71] The definitions of all the different types of data objects are presented in detail in Deliverable 3.2 (Van Hal et al., 2020).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

#	Metadata property	Definition	Matching property from external ontology	Property URI
17	ThematicNetwork FullName	The full name of the Thematic Network (e.g., Agroforestry Innovation Networks).	**schema:**name	http://schema.org/name
18	ThematicNetwork Url	The URL of the Thematic Network or the URL directing to the TN-related record in the CORDIS database (in the case that the TN's URL is dead).	**schema:**url	http://schema.org/url
19	ThematicNetwork Domain	The agricultural sector (i.e., Crop farming, Livestock, or Forestry) to which the Thematic Network relates.	no relevant property from external vocabulary	

6.3.3.3 The process of the annotation of data objects with metadata values

The annotation of the data objects to be delivered via the e-KRP was a process of assigning values to the set of metadata properties (presented above) to each of the 631 data objects that have constituted the input set for the repository. The annotation process took place from the beginning of August 2020 till the end of October 2020, and was undertaken by a group of three annotators - members of AUA's team. The annotation process was implemented for each of the data object categories considered for the repository (i.e., documents, videos, presentations, images, and audios). Given the importance of annotating/characterising the data objects with the "correct" metadata values, a training session took place before the launch of the process of annotation of each data object category. Slides especially designed for this purpose were used to support training. These slides were provided to the annotators (after the completion of the training sessions) to use them as a reference during the process. Some example slides (developed for instructions on the annotation of videos) are shown in Figure 27 below.

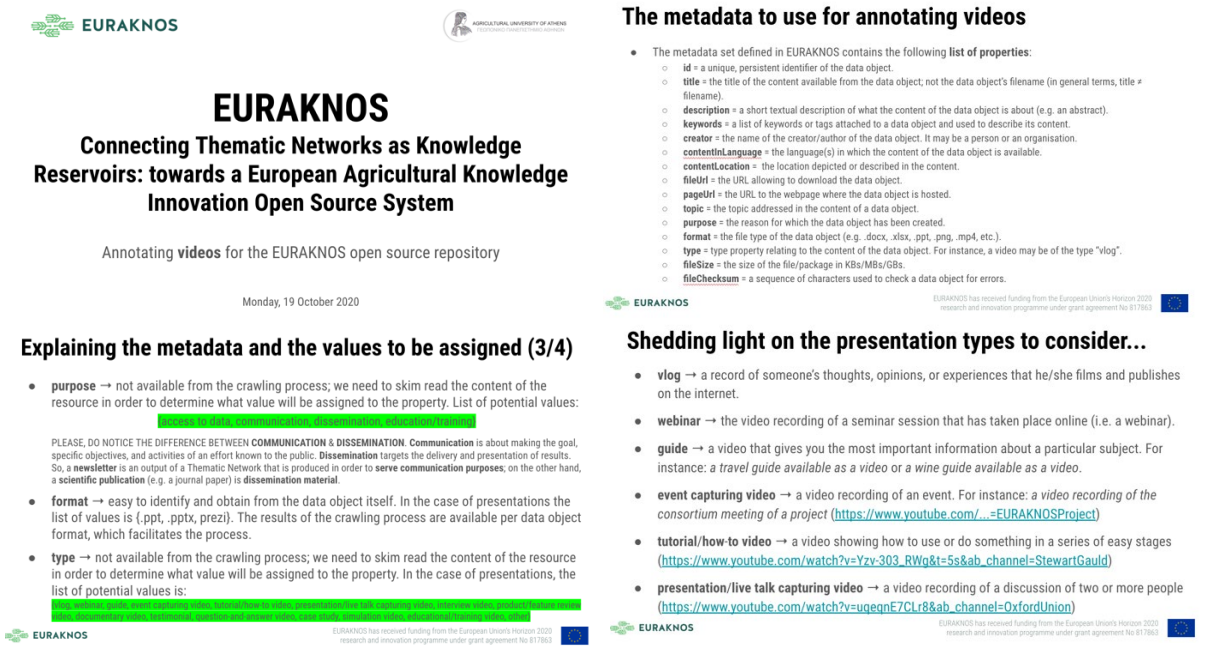


Figure 27. Example screenshots of the slides used to train the annotators on how to appropriately annotate the video-based content to be delivered through the e-KRP

Apart from instructions particularly focusing on the annotation of the specific categories/types of data objects, the delivered training sessions also aimed to inform the annotators about the importance of the annotation of data with metadata, why use metadata to characterise/describe data objects, as well as the different types of metadata and the purposes they serve.

Each training session was followed by the process of annotating the data objects of a specific category with the appropriate values. Support to the annotators was also provided during the annotation. So, apart from the training session implemented before the annotation process, one-to-one or group calls were held to discuss problems or the progress achieved. The results of the annotation were made available in spreadsheets. These spreadsheets were completed by each of the annotators. Each row in these spreadsheets contained the metadata values assigned to each concrete data object. By the end of the annotation process, we were able to collect spreadsheets containing the metadata values of all the data objects delivered from the e-KRP. Figure 28 below shows an example spreadsheet (namely, the spreadsheet with the metadata values of the podcasts available from the e-KRP).

As a side note, it needs to be mentioned that the number of the data objects finally annotated with metadata was less than 631 (the number of data objects selected from all data object categories/types as potential input). The reason is that during the annotation process, it has become feasible to identify faulty data objects. These faulty data objects related to dead links or damaged/corrupted files. It has also been feasible to further identify duplicates (not identified in the data cleaning process), as well as misclassification cases (i.e., cases of data objects falsely classified into a specific type)[72]. Finally, the bulk of the 121 objects classified into the "other" category has been utilised as a pool of back-up data objects to consider in cases where the numbers of data objects per type listed in Tables 11 to 18 could not be reached.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
#	title	description	keywords	creator	contentInLanguage	contentLocation	fileUrl	topic	purpose	format	type	fileSize		dateCreated
	http://www.semanticweb.org/hercules_pan/ontologies/2020/6/EURAKNOS_ontology#hasTitle	http://schema.org/description	http://schema.org/keywords	http://schema.org/creator	http://www.semanticweb.org/hercules_pan/ontologies/2020/6/EURAKNOS_ontology#hasContentInLanguage	http://schema.org/contentLocation	http://schema.org/fileUrl	http://xmlns.com/foaf/0.1/topic	http://schema.org/potentialAction	http://purl.org/dc/terms/format	http://www.semanticweb.org/hercules_pan/ontologies/2020/6/EURAKNOS_ontology#hasType	http://schema.org/fileSize		https://schema.org/dateCreated
1	A closer look at the Eu Pig Thematic Network with Ben Williams	A brief description of Eu Pig Thematic Network and its activities by Ben Williams.	pig; animal husbandry and welfare; best practice; farming practice; innovation; knowledge exchange	EURAKNOS	English	N.A.	https://open.spotify.com/episode/4BgQxzaWBpXBPyKJN2eMwr	animal husbandry and welfare practices; animal husbandry and welfare actors; animal husbandry and welfare systems; animal husbandry and welfare infrastructures; animal husbandry and welfare tools; animal husbandry and welfare digital technologies	communication	.mp3	interview	23.7 MB		2019-12-11
2	A quickfire podcast with the Euraknos consortium!	The biggest achievements and the next steps of EURAKNOS project are presented by the team members of EURAKNOS consortium.	N.A.	EURAKNOS	English	N.A.	https://open.spotify.com/episode/69FSJCRLFZgpcyY036Nx4	N.A.	communication; dissemination	.mp3	interview	14.2 MB		2020-06-19
3	AgriforValor: Exploiting side stream biomass resources	A brief description of Agriforvalor Thematic Network and its activities by Katalin Kurucz .	waste, by-products and residues management; forestry; plant production and horticulture; farming / forestry competitiveness and diversification; biomass; bioeconomy; knowledge exchange	EURAKNOS	English	N.A.	https://open.spotify.com/episode/3rVk6kcFzuvBaxBaGstY	crop farming practices; crop farming actors; crop farming systems; crop farming tools; crop farming digital technologies; socio-economic context; forestry practices; forestry actors; forestry systems; forestry infrastructures; forestry tools; forestry digital technologies; climatic zones; environment; socio-economic context	communication	.mp3	interview	28 MB		2020-05-19
4	AgriSpin: Space for Innovations in Agriculture	A brief description of AgriSpinThematic Network and its activities.	farming practice; best practice; innovation support service; AKIS	EURAKNOS	English	N.A.	https://open.spotify.com/episode/68RQ8JLNoI692vkaLTTa	crop farming practices; crop farming actors; crop farming systems; crop farming infrastructures; Crop farming tools; crop farming digital technologies;	communication	.mp3	interview	33.7 MB		2019-12-13
5	An insight into the EUfruit Thematic Network	A brief description of EUfruit Thematic Network and its activities by Michelle Williams.	fruit farming; agricultural production system; farming practice; knowledge exchange	EURAKNOS	English	N.A.	https://open.spotify.com/episode/6CcXuuuZ28fuPdu9GTmtXO	crop farming practices; crop farming actors; crop farming systems; types of crops; crop farming infrastructures; crop farming tools	communication	.mp3	interview	15.9 MB		2020-06-15
6	CEV's: Overcoming the challenges of Covid	Maria Gernert from IFOAM organisation describes her cross-exchange visit to EURAKNOS project and how it can be succesfully continuined during coronavirus period.	N.A.	EURAKNOS	English	N.A.	https://open.spotify.com/episode/19DJpIFxaNGfuruHA4Eb3c	N.A.	communication	.mp3	interview	10.3 MB		2020-09-17
7	Cross Exchange Visit: Identifying best practices for knowledge platforms	The most recent cross exchange visit of IFOAM, which it was focussed on the OK-NET Ecofeed Thematic Network knowledge platform, is presented by Maria Gernert.	N.A.	EURAKNOS	English	N.A.	https://open.spotify.com/episode/7vrutvbtWkyqPXwmTw84q4	N.A.	communication	.mp3	interview	11 MB		2020-09-17
8	DISARM: Innovative Solutions for Antibiotic Resistance Management	A brief description of DISARM Thematic Network and its activities by Laura Palczynski.	agricultural production system; farming practice; farming equipment and machinery; animal husbandry and welfare; pest / disease control;	EURAKNOS	English	N.A.	https://open.spotify.com/episode/7JZbvRm78vod758nVXAb	animal husbandry and welfare practices; animal husbandry and welfare actors; animal	communication	.mp3	interview	8.3 MB		2020-04-02

Figure 28. The metadata values assigned to the podcasts available through the e-KRP. Each row in the spreadsheet contains the metadata values of one data object (i.e., podcast)

The full list of the data objects housed in the repository of the e-KRP is contained in the "data_objects _in_e-KRP_repository.json" file, which is available in Zenodo (https://zenodo.org/record/5500336).

6.4. Integration of data objects into the e-KRP repository

The methodology phase that was implemented after the assembly of the input for the e-KRP repository was the one concerned with the integration of the annotated data objects into the repository. This has

[72] This has been the case with documents.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

been the last phase of the methodology related to shaping up the repository. In order to execute the integration, we had to consider the characteristics and particularities of DSpace[73] (i.e., the software solution adopted for the deployment of the e-KRP repository). More specifically, we had to take into consideration that, in the case of DSpace, the uploading of data objects is implemented using batch registration of items rather than a one-by-one registration. For this reason, an ETL (Extract-Transform-Load) service has been created to allow for the batch loading of data and metadata into DSpace.

The ETL service has been responsible for the ingestion of the data objects (harvested from the sample TNs' KRs) into the repository. The design and implementation of the ETL service were based on specific ways of making the TN data objects available; namely, by way of a csv format, where each row of a csv file corresponds to the metadata records associated with one data object. These csv files were created as the result of the annotation process. The corresponding data object was then supplied locally using an archive format such as ZIP, or by a download URL in the csv file.

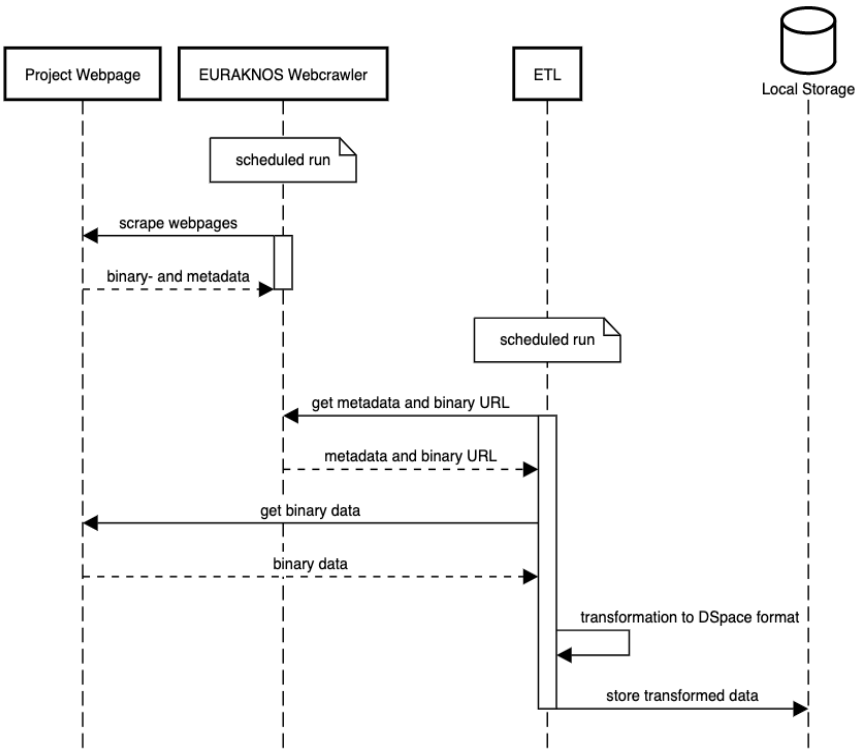


Figure 29. Sequence diagram displaying the ETL process for harvesting data objects from the KR of a TN and storing them to local storage

The process of harvesting data objects from the KRs of TNs and initially storing them into a local storage space is shown in Figure 29 above with the help of a sequence diagram. More specifically, by deploying the EURAKNOS web crawler it has become feasible to harvest data objects, and metadata associated with them, from the KRs of the TNs in our sample. The data objects have been retrieved from the KRs of the TNs and transformed to become DSpace compliant. The data objects have been initially stored in a local storage space reachable by DSpace.

As a preparatory step to loading all the data objects into DSpace, they have all first been converted to a proprietary Simple Archive Format[74] required for the registration in DSpace. This process has resulted in the storage of all data objects, and metadata contained in the supplied csv, on the local file system. For each entry in DSpace, a directory was created. This directory contained the data object and an XML

[73] Details about the reasons for deciding to make use of DSpace as part of the EURAKNOS technology stack are provided in Deliverable 3.2 (Van Hal et al., 2020).

[74] https://wiki.lyrasis.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

file containing descriptive metadata. After the loading of the data objects in DSpace, a checksum[75] was generated and added to the metadata. This has allowed for checking for identical files in the future, as well as providing a safeguard against corrupted data objects.

After having stored all data objects locally in the desired format, the ingestion into DSpace took place as shown in Figure 30. The registration of items into DSpace took place on a configurable time interval, during which the local file storage containing the prepared folders with the data object and metadata is read. Storage in DSpace was then made depending on the modality. Specifically, the data object was stored in a fully transaction-safe way on local storage. Relevant system information on the data object such as IDs was stored in the relational PostgreSQL database. Metadata was stored in the Solr document storage instance. Metadata was also converted to RDF triples. This process took place by a configurable mapping, which converted each metadata key-value pair to a corresponding triple in RDF. The resulting triples were stored in the Fuseki triplestore.

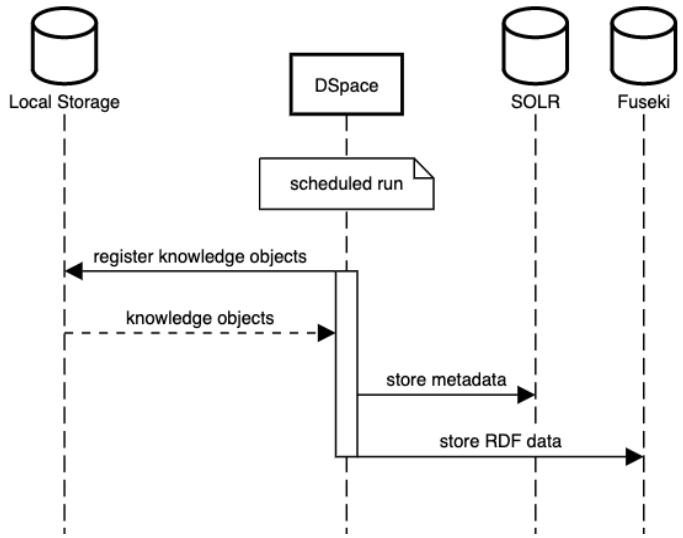


Figure 30. Sequence diagram illustrating the registration of data objects and their metadata into DSpace

As shown in Figure 31, the integration of data objects into the e-KRP repository was the final step in a pipeline of activities starting from the data cleaning process. This integration process has been the step required to make the data objects available and accessible to the e-KRP users.

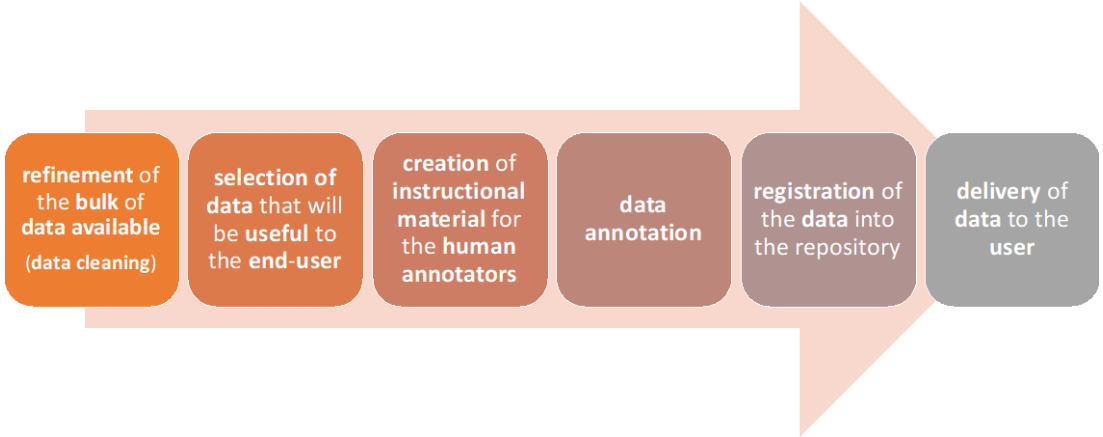


Figure 31. The integration and registration of the TN-related data objects (and their metadata) into the e-KRP repository as a final step before making them available to the end-users

[75] According to Wikipedia (https://en.wikipedia.org/wiki/Checksum), "checksum" is the term used to denote "*a small-sized block of data derived from another block of digital data for the purpose of detecting errors that may have been introduced during its transmission or storage*." Checksums are "*often used to verify data integrity but are not relied upon to verify data authenticity*."

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

6.5. Evaluating the FAIRness of the e-KRP

In the past few years, the FAIR Data principles (Wilkinson et al., 2016) have become prevalent as a set of recommendations for making datasets Findable (F), Accessible (A), Interoperable (I), and Reusable (R). The purpose of establishment of this set of recommendations has not been to enforce a mandatory use of them, but rather to provide guidelines to the entities generating data to make it useful to the community of interest. Given the added value that data has as raw material for the training of machine learning algorithms, and thus enabling breakthroughs in research and technology development, these recommendations have not been proposed only for the interest of people, who may make use of data, but also to make data "interpretable" and "consumable" by machine agents.

Based on the big interest that the compliance to the FAIR principles has attracted, both in the research and industry domains, several FAIRness assessment tools have been developed and published in an attempt to facilitate insights into the degree of adherence to the FAIR Data principles. SATIFYD (Self-Assessment Tool to Improve the FAIRness of Your Dataset)[76] is such an assessment tool, developed by the Data Archiving and Network Services (DANS)[77]. SATIFYD is made up of twelve (12) assessment items split into four sections (one section per each of the FAIR principles dimension with three items related to it). SATIFYD provides a final score of compliance with the FAIR principles in general, but also a degree of compliance with each dimension of the FAIR principles.

SATIFYD is the tool that has been used for evaluating the FAIRness of the e-KRP. Table 24 provides the items available in SATIFYD to assess the adherence to each dimension of the FAIR principles, the list of responses to select from, and the responses provided to evaluate the FAIRness of the e-KRP.

Table 24. Results of the EURAKNOS e-KRP's FAIRness evaluation

	Assessment item	Potential responses	Response provided	Score
FINDABLE	Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data?	• Rich metadata with as much information as possible. • Required metadata fields and some additional fields. • Minimal metadata but only required fields • No metadata.	Required metadata fields and some additional fields.	**55%**
	Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset?	• Controlled vocabularies. • Taxonomies (thesauri). • Ontologies. • There are no standards for my discipline.	Controlled vocabularies.	
	Did you provide rich and detailed additional documentation?	• Readme file. • Versioning. • Provenance.	None of the response options applied and thus, none of them was selected.	
ACCESSIBLE	Is the metadata publicly accessible even if the data is no longer available?	• Yes • No • I can't find this information in EASY.	Yes	**50%**
	Does your dataset contain personal data?	• Yes • No	Yes[78]	
	Which of the usage licenses provided by EASY did you choose to comply with the access rights attached to the data?	• Open access (CC0). • Restricted access. • Embargo. • Other access.	Other access.[79]	

[76] https://dans.knaw.nl/nl

[77] https://satifyd.dans.knaw.nl/

[78] We have provided the "Yes" response, because part of the data objects in the e-KRP repository are videos in which the faces of people are shown.

[79] The license decided to be attributed to the e-KRP's data objects is the Creative Commons CC BY-NC-ND 4.0 license. Details are provided in Subsection 8.2 ("Towards an e-KRP sustainability plan").

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

	Assessment item	Potential responses	Response provided	Score
		• I don't know which of these licenses apply to my data.		
INTEROPERABLE	Are the data in your dataset stored in preferred formats?	• All the data are in preferred formats. • Some of the data are in preferred formats. • None of the data are in preferred formats. • I don't know what preferred formats are.	All the data are in preferred formats.	**50%**
	Do you link to other (meta)data and is this (meta)data online resolvable?	• Yes, and these are online accessible. • Yes, but they are not online accessible. • No.	No	
	Did you provide contextual information about your dataset?	• Persistent identifiers. • Reference to other datasets. • Reference to publications. • No contextual metadata.	No	
REUSABLE	What kind of information did you provide about the provenance of your data?	• Origin of data. • Citations for reused data. • Workflow description for collecting data (machine readable). • Processing history of data. • Version history of data.	Origin of data.	**37%**
	Which usage license provided by EASY did choose for your dataset?	• Open access (CC0). • Restricted access. • Embargo. • Other access. • I don't know which of these licenses apply to my data.	Other access.	
	Does your (meta)data meet domain standards?	• Domain standards in metadata. • Minimal domain standards and minimal generic standards. • Generic standards. • No standards used.	Generic standards.	
			Total FAIRness score	**54%**

As made evident, the FAIRness score of the e-KRP has been 54%. This score is promising, especially if we consider the low degree of compliance with the FAIR principles achieved by other systems (Dunning et al., 2017.). This FAIRification effort has already been taken up in EUREKA in an attempt to advance the level of adherence to the FAIR Data principles. Figure 32 below shows the responses provided as part of the e-KRP's FAIRness evaluation with the help of the SATIFYD tool and the scores achieved.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

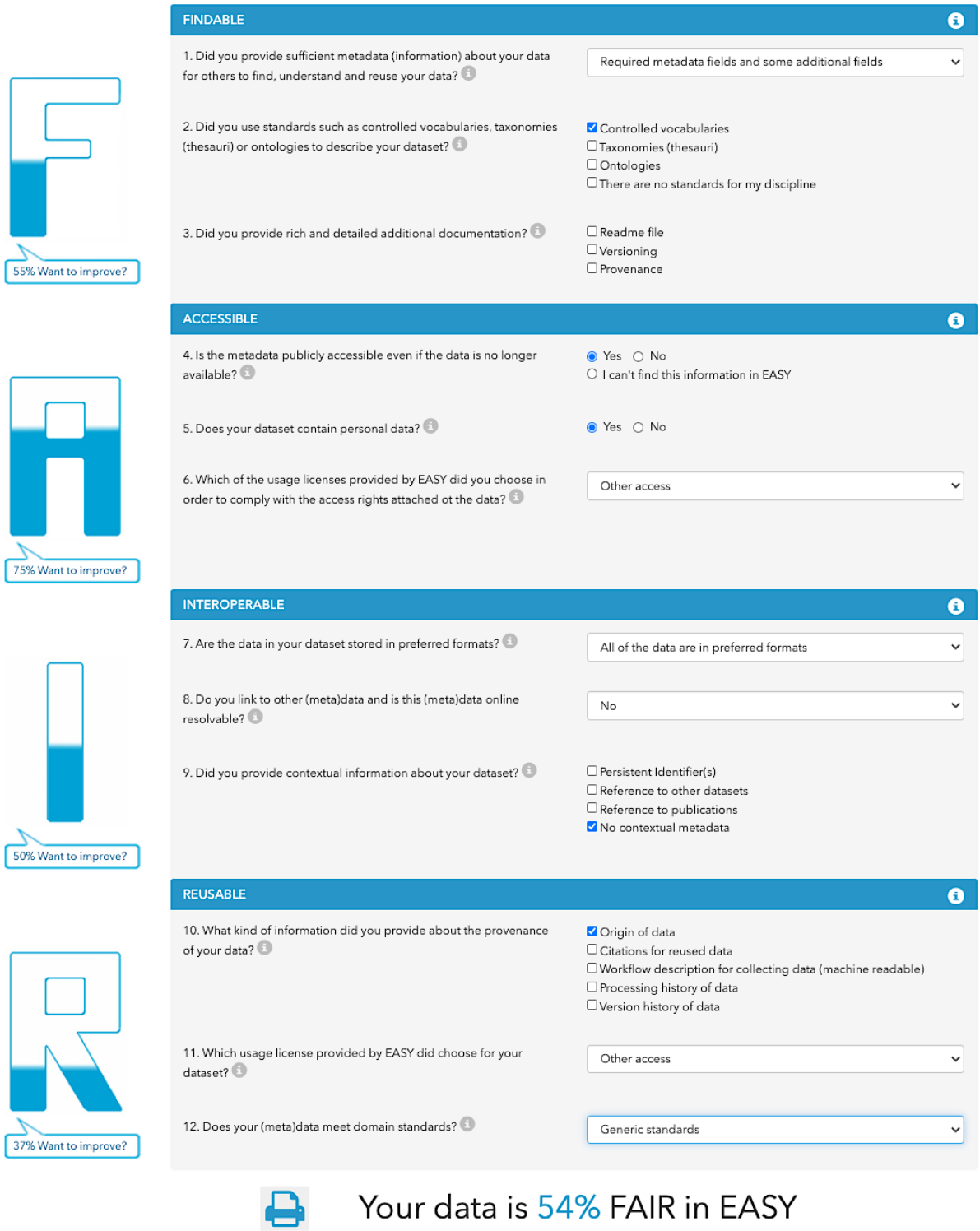


Figure 32. Evaluation of the EURAKNOS e-KRP's FAIRness with the help of the SATIFYD tool

7. The e-KRP's back- and front-end development

The back- and front-end development of the e-KRP has been a core part of the work and activities done in Task 4.2. It has helped bring all the components of the EURAKNOS e-KRP technology stack together and make them "talk" to each other in order to implement the architecture described in Section 4. The aim of the present section is not to dive deep into the technical details, but rather provide an overview of the work undertaken in both the foreground and background of the platform to turn it into a reality. We begin with a short discussion on what the back-end development is generally about and then with

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

accounts of what it has particularly involved in the EURAKNOS project (Subsection 7.1). Similarly, the work related to the front-end development of the e-KRP is presented in Subsection 7.2. The EURAKNOS e-KRP can be accessed at http://test.eufarmbook.eu/[80]. It needs to be stressed that the platform is in ongoing development continued in EUREKA, which has already taken up the work done in EURAKNOS. It is a live product the development of which is currently taking place in the EUREKA project by building upon the work in EURAKNOS. The code that has been developed for the EURAKNOS e-KRP is published in Zenodo (the link to the EURAKNOS project's virtual space in Zenodo, where the code can be found, is https://zenodo.org/communities/euraknos_project_data?page=1&size=20)[81].

7.1. Back-end development

The back-end development part of the work involved in a software application's creation is the "server-side" part of the work that cannot be seen but is necessary for the communication between the data repository and the browser and thus, the delivery of content and information to the end-user[82]. In the case of the e-KRP this communication establishment task has been "assigned" to Drupal[83]. This means that Drupal is responsible to make the front-end of the e-KRP "talk" to DSpace (i.e., the software used for the implementation of the e-KRP repository). Drupal is open-source software distributed under the terms of the GNU General Public License (GPL)[84]. It enables easy content authoring, has reliable performance, and provides excellent security features. Modularity is one of Drupal's core principles. More specifically, it allows the creation of versatile and structured content to facilitate dynamic web experiences. Besides, its functionality can be extended to enable the customisation of content delivery and presentation. Support in Drupal-based development can be sought and provided by the respective community, which is one of the largest open-source software communities in the world.

7.2. Front-end development

Front-end development is the term used in the IT community to refer to the "client-side"-related work that is done as part of a software application's creation[85]. It relates, among others, to the application's aesthetics and, for this reason, is inextricably linked to basic concepts and principles of User Experience design and engineering. To put it in simple terms, the work done as part of the front-end development has direct implications for the look and feel of the software application's user interface. However, the front-end development is not just limited to what the end-user sees. There are issues related to it that go beyond the aesthetics of a GUI. These issues have to do with decisions about how the content and information will be accessed. The use of menus and navigation structures are part of the solutions that can be provided (and thus, engineered in an application) to structure (and facilitate) access to content.

To develop a visually attractive and user-friendly platform, the front-end development in EURAKNOS has been based on the recommendations that have been identified and defined as the outcome of the user needs analysis (see Section 5: "User needs analysis and specification of requirements"). An initial, generic requirement that has been identified from the analysis of user needs relates to the potential to access the e-KRP from digital device (personal computer, tablet, or mobile phone); in other words, it relates to cross-platform access to the e-KRP. This requirement has been engineered into the e-KRP's front-end development and the result is shown in Figures 33 and 34 below, which illustrate the way in which the e-KRP's landing page can be viewed from a personal computer, a tablet, and a mobile phone.

[80] To access the e-KRP, use the following credentials: (i) username: euraknosplatform; and (ii) password: euraknos2020

[81] More specifically, the e-KRP code can be accessed at https://zenodo.org/record/5386227#.YTotSp4zb0o, https://zenodo.org/record/5386111#.YTotd54zb0o, https://zenodo.org/record/5386071#.YTot454zb0o, https://zenodo.org/record/5386041#.YTot_54zb0o, and https://zenodo.org/record/5385901#.YTouPp4zb0o.

[82] https://www.coursereport.com/blog/front-end-development-vs-back-end-development-where-to-start

[83] https://www.drupal.com/

[84] https://www.gnu.org/licenses/gpl-3.0.html

[85] https://www.coursereport.com/blog/front-end-development-vs-back-end-development-where-to-start

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

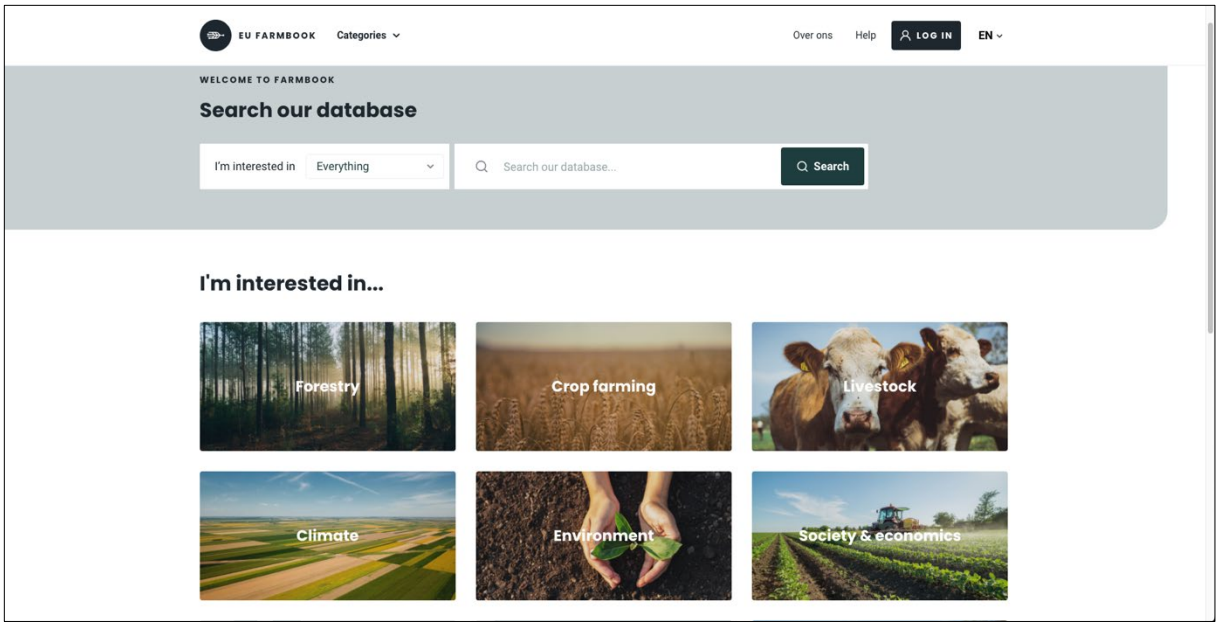


Figure 33. View of the e-KRP's landing page by using a personal computer

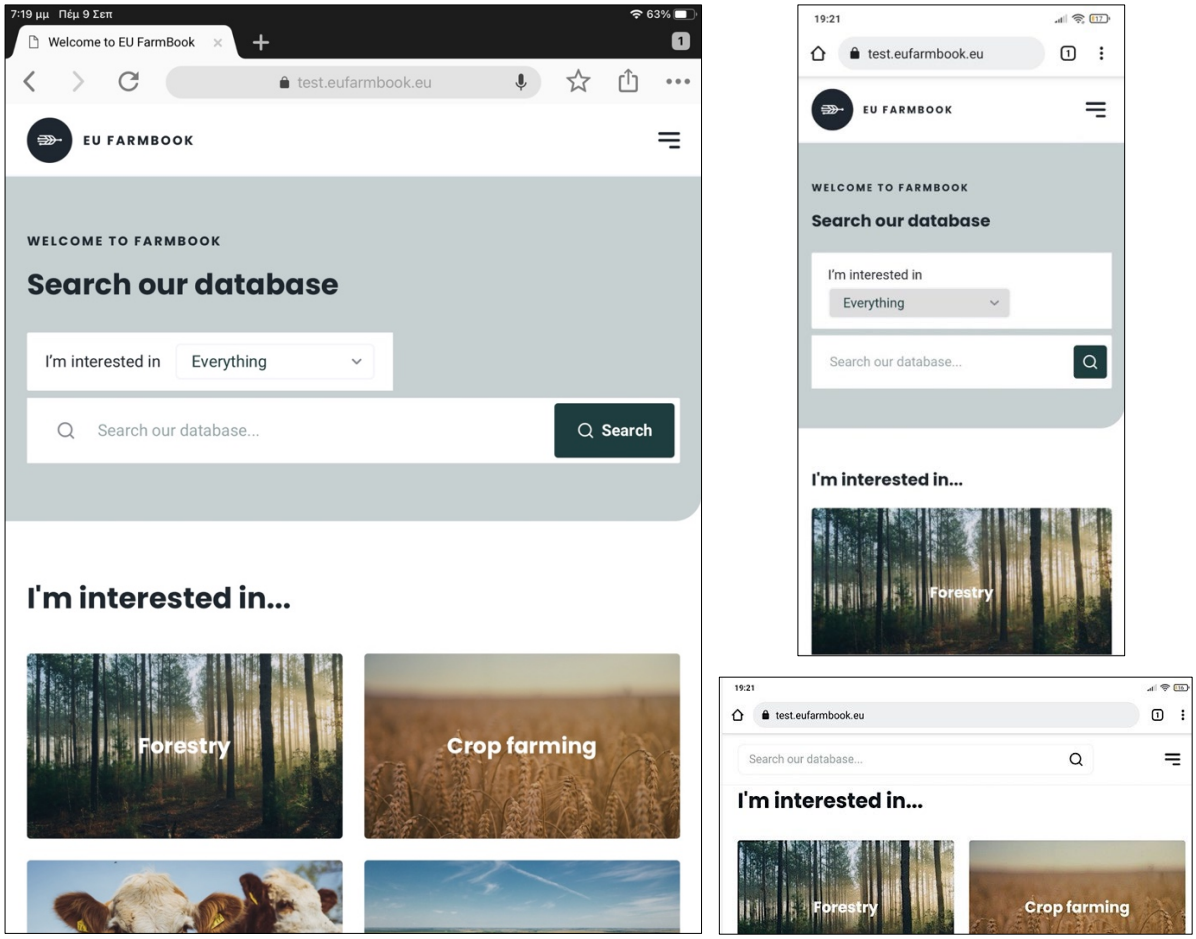


Figure 34. View of the e-KRP's landing page by using a tablet (left) and a mobile phone (right)

Regarding the guidance and support provided to the end-user in respect to identifying and finding the content and information of interest, Figure 35 shows the highly structured view that has been designed for the landing page of the e-KRP. What becomes evident from that screenshot are the clear and easily identified visual elements that help the user to easily select the content category/-ies of preference or

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

proceed to content search (**Req#1**: "*provide highly structured dashboards supported by fit-for-purpose visual aids*"). The use of visual aids is prevalent, whereas the text that is being used is very limited.

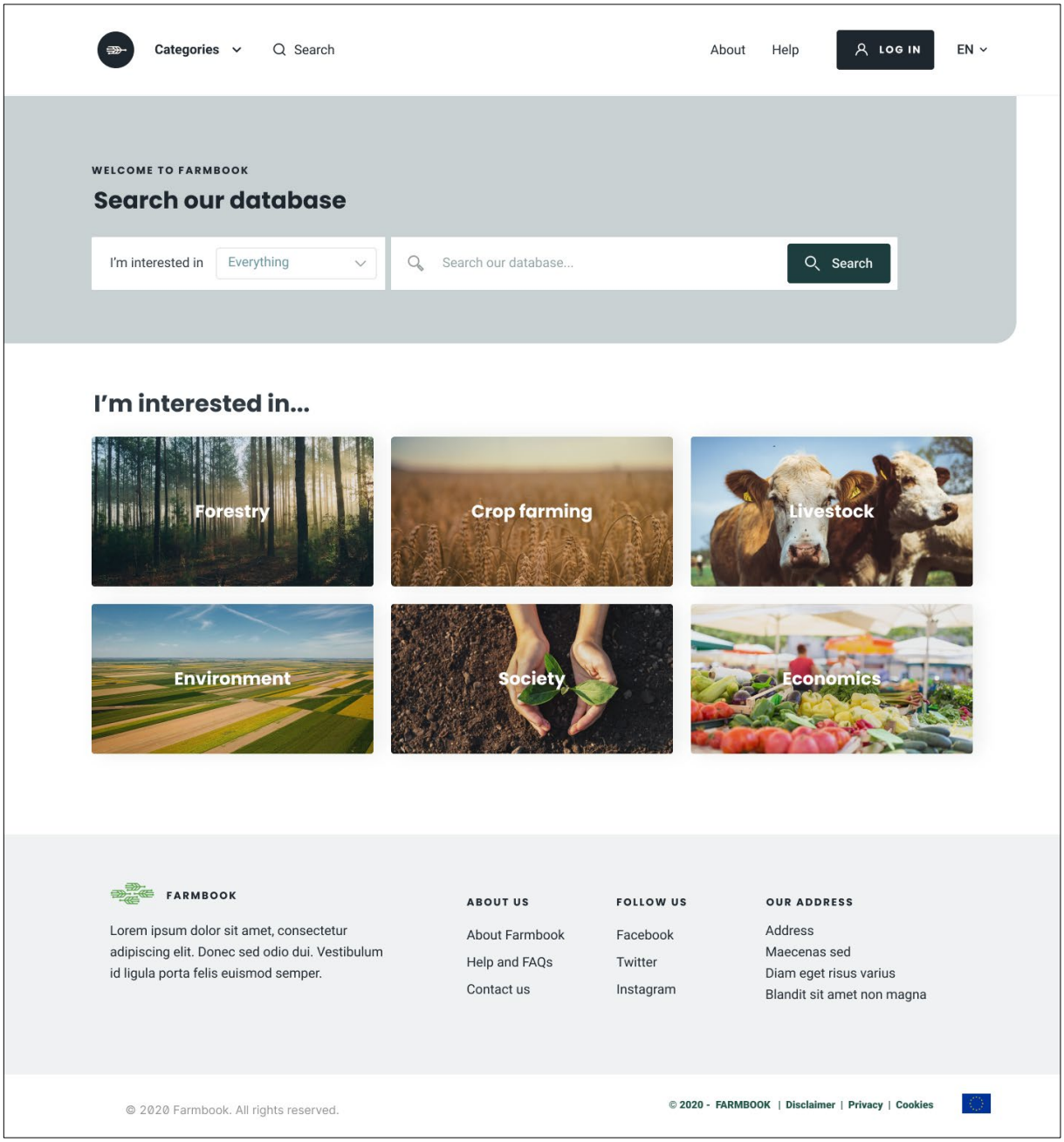


Figure 35. The highly structured landing page of the e-KRP allowing the user to easily select (or search for) the content of preference

In addition to the above, there has been a requirement to create different views of the digital content available from the e-KRP based on the particularities of the various categories of data objects (**Req#1**). Based on that requirement, different views have been designed. Specifically, the delivery of documents is based on page views in which visual elememts are made available (namely, a figure that is relevant to the content of the document) and text is provided at easy-to-consume chunks (see Figure 36 below). In the case of non-textual content (e.g., audio-visual material), the extent of text-based descriptions is minimal (**Req#1**: "*Accompany the delivery of non-textual content (e.g., audio-visual material) with a brief textual description*"). Figures 37 and 38 below illustrate the design decisions made in response to the relevant requirement, by showing the views related to a video and podcast delivery respectively.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

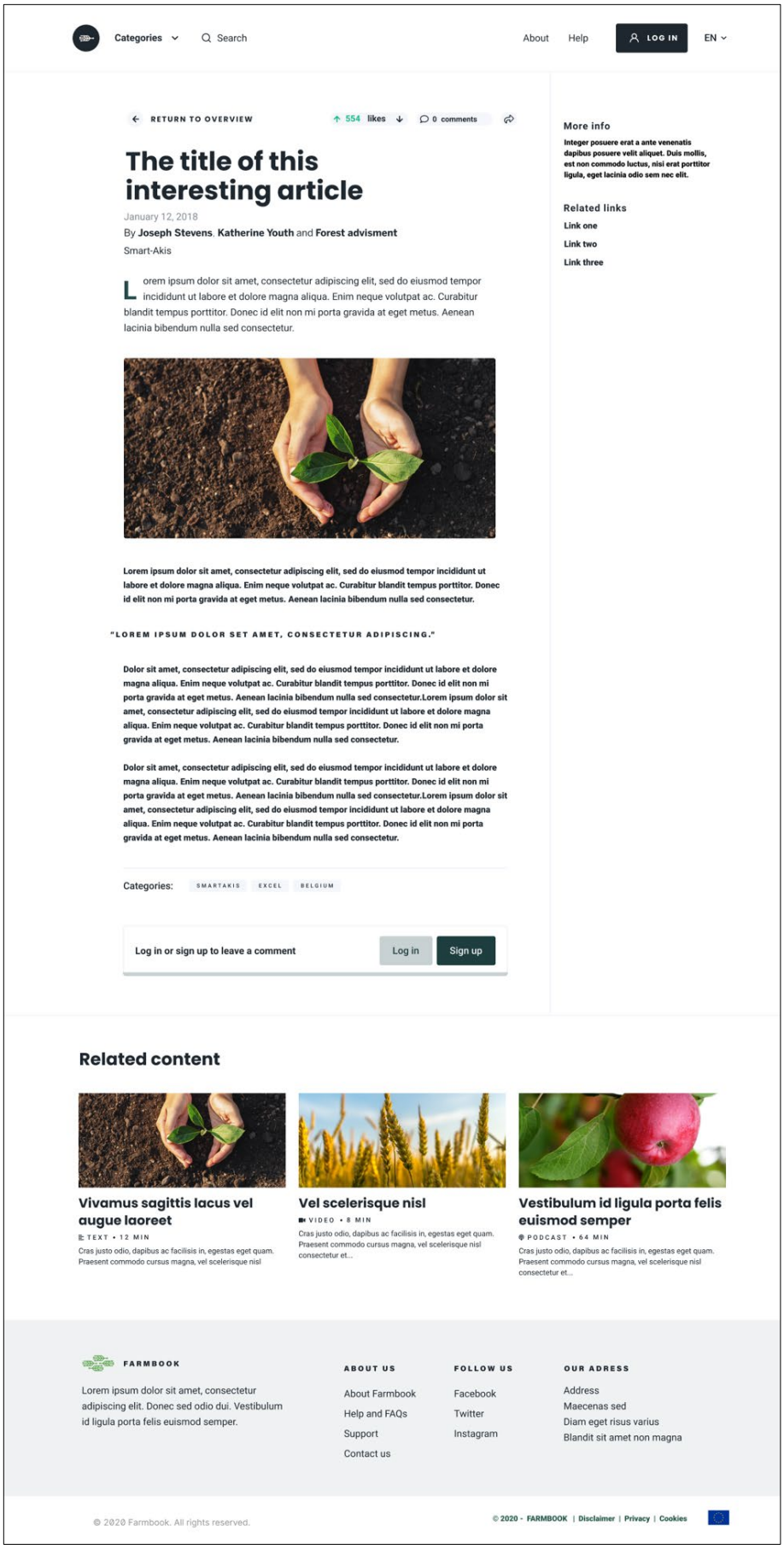


Figure 36. View of a document

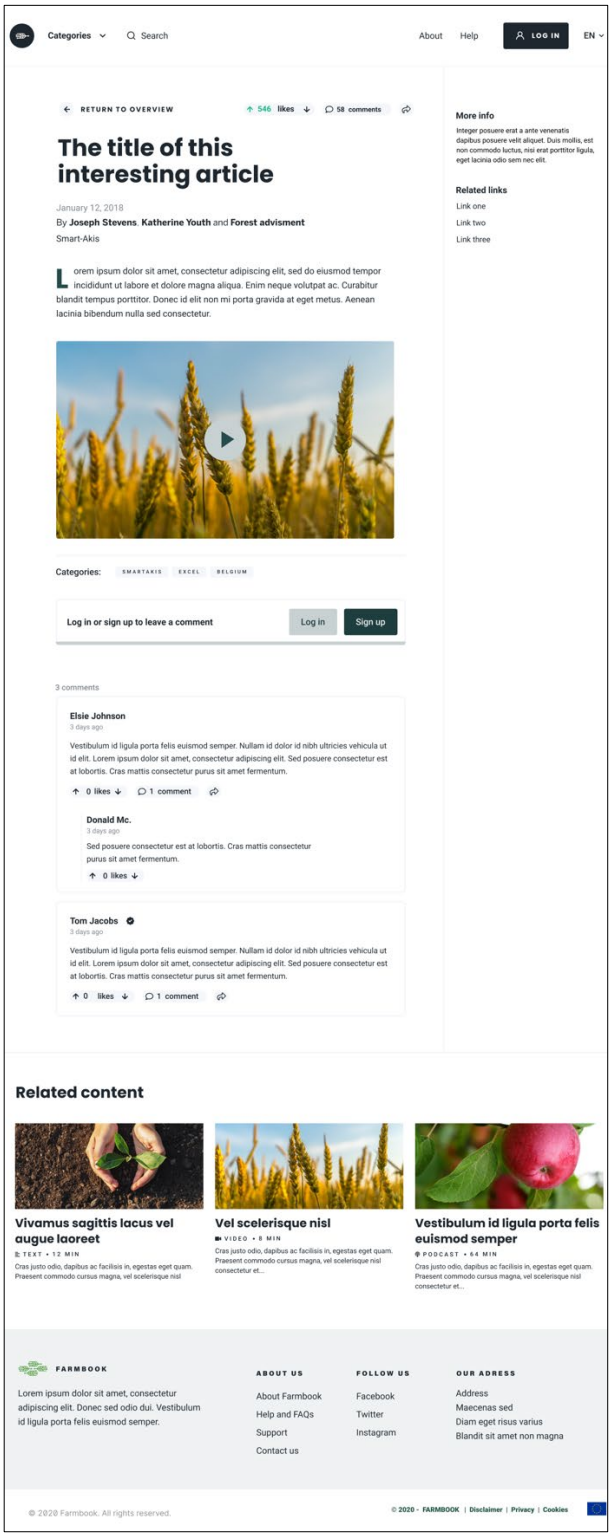


Figure 37. View of a video

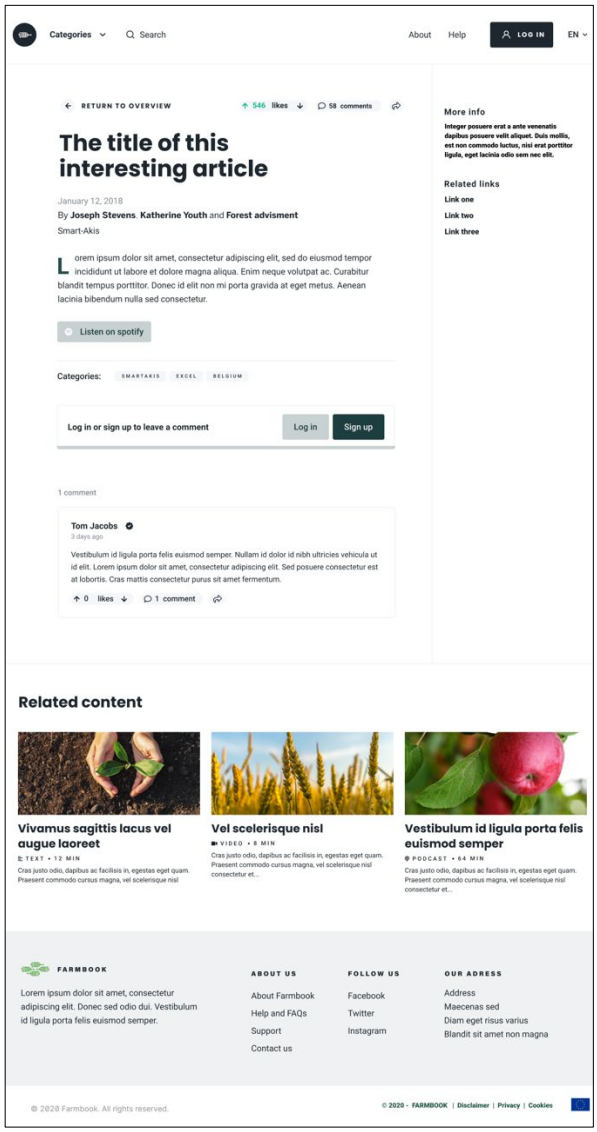


Figure 38. View of a podcast

The provision of a discussion forum where the users can create message threads by making their own posts or replying to existing ones is a functionality that has been considered and designed as a response to the requirement for functionalities that go beyond a mere delivery of content/information (**Req#2**). In the same line, another functionality that has been considered is that of providing access to details/ information about experts (the individuals involved in the creation of data objects). This functionality has been designed with the aim to infuse a sense of community building in the platform. The views related to the discussion forum and the page providing expert details are illustrated in Figures 39 and

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

40 below. In addition, the functionality of posting comments about data objects is another aspect that has been considered and designed towards the direction of bulding a platform not just limited content delivery (see Figures 37 and 38 above).

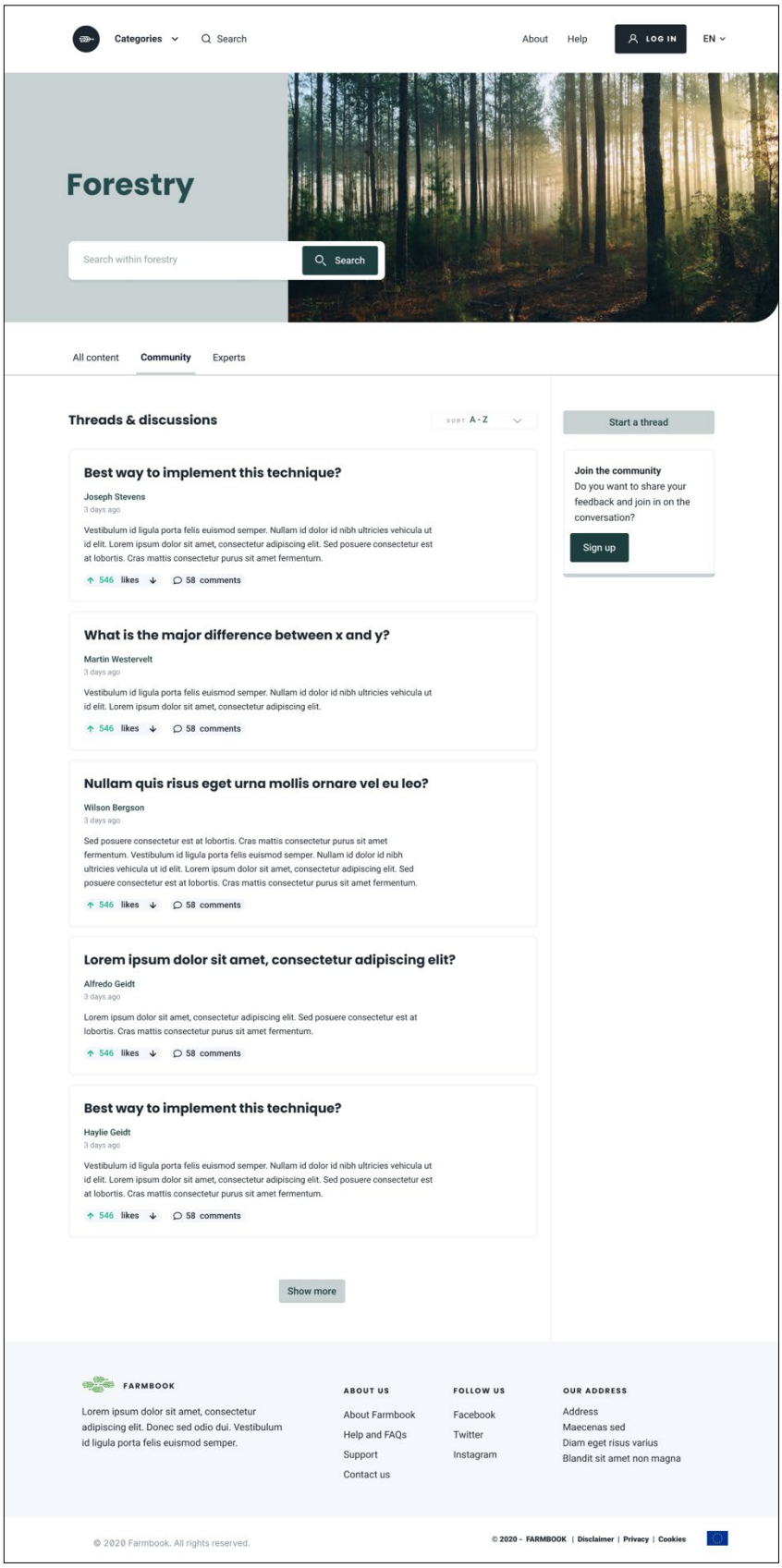


Figure 39. Discussion forum

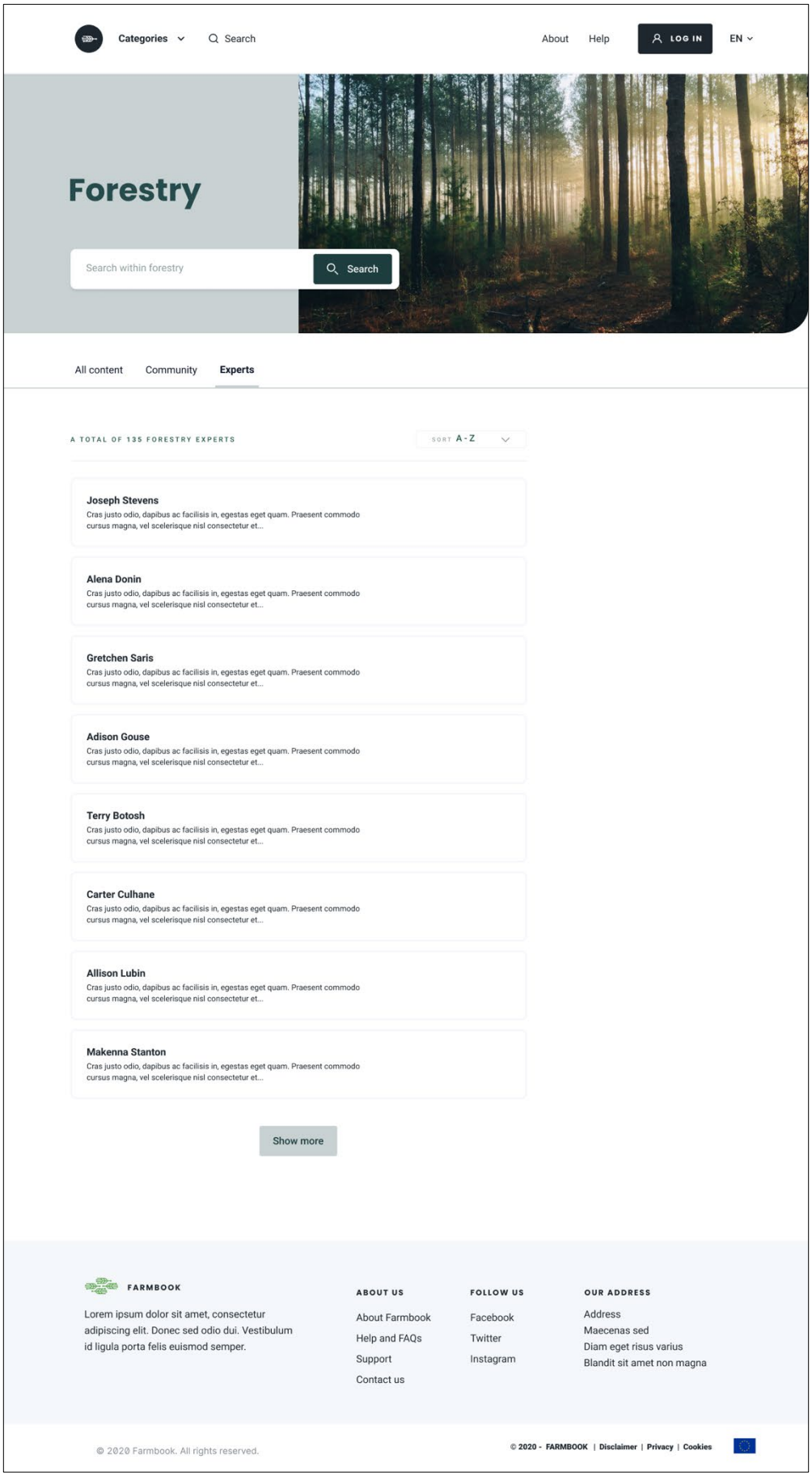


Figure 40. e-KRP providing expert details

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

The search for content and access to it in the e-KRP are enabled by means of: (i) a search function used with the help of a search bar that is placed in a prominent position on the e-KRP's landing page (see Figure 41); and (ii) a well-structured navigation structure. The design of these functionalities has been made as a response to **Req#3**: "*Provide functionalities that will enable different ways of finding and retrieving relevant content and information (namely, search function along with an easy-to-use navigation structure*".

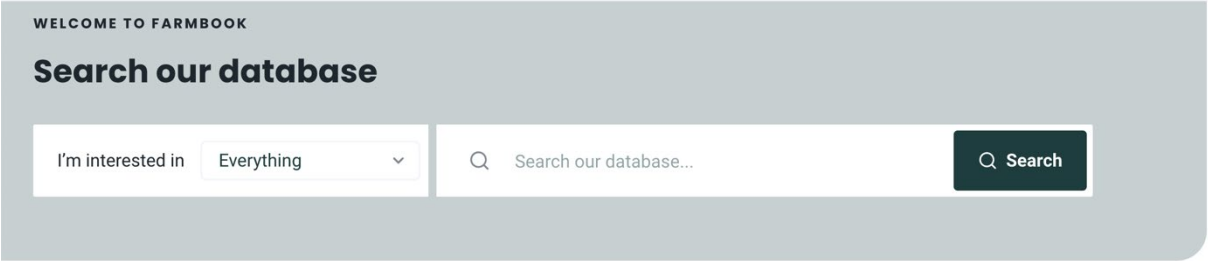


Figure 41. The search bar designed to implement the search function of the e-KRP

The structure illustrated in Figure 42 provides another way of facilitating the identification and retrieval of relevant data objects in the e-KRP by providing the blueprint for engineering the navigation of users through the available content and information.

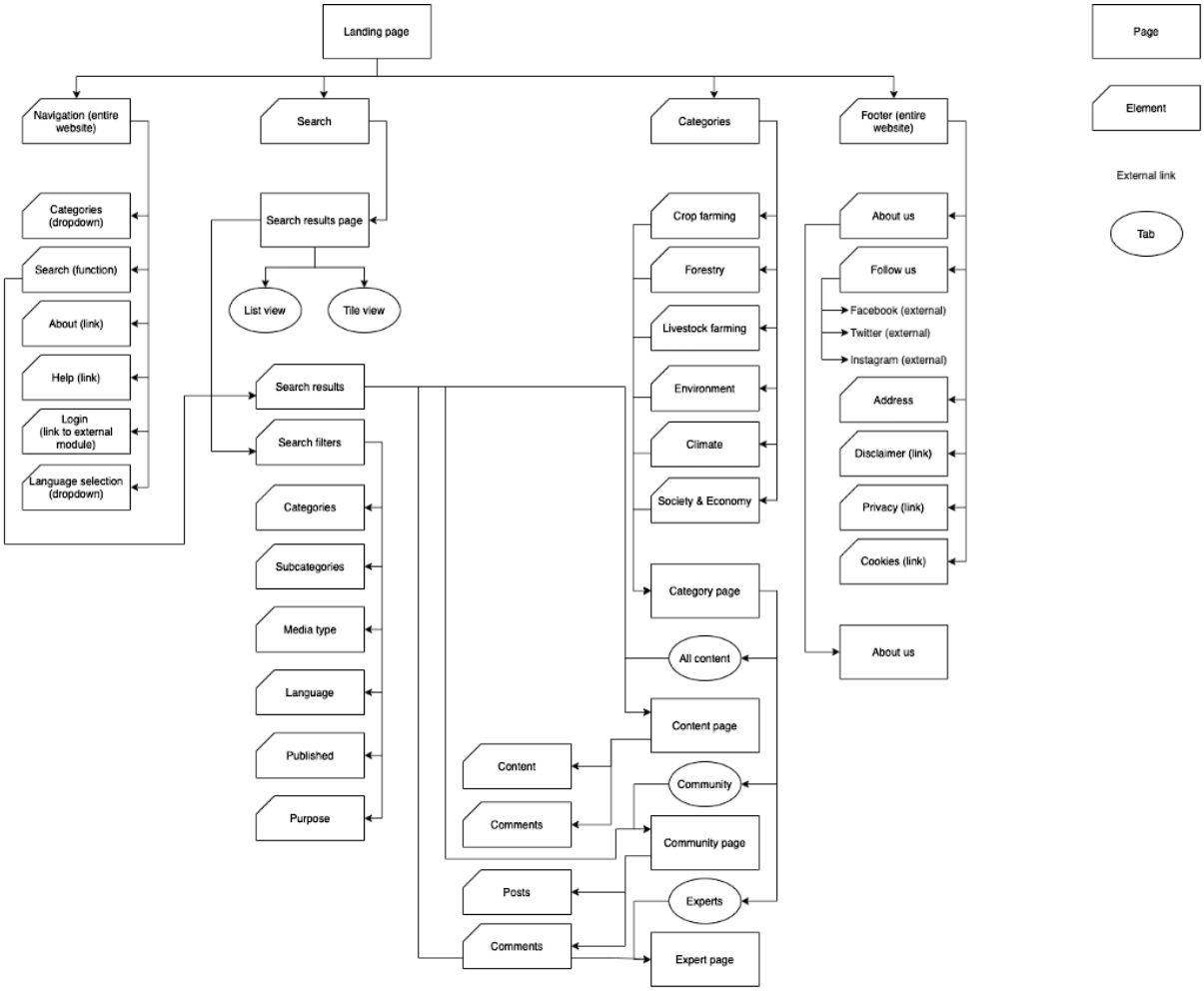


Figure 42. The EURAKNOS prototype e-KRP's Information Architecture

The elements illustrated in this design and their connections have been the input for the development of the e-KRP's navigation structures (e.g., menus) and functionalities such as the search result filtering

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

options. Figure 42 makes explicit the structure and navigation levels that have been adopted and used to facilitate access to content and information.

The "Navigation" element appearing on the left side of the diagram shows the menu options available at the top of each page of the e-KRP. The "Search" element, next to the "Navigation" element, indicates the available search result display options (the search results can be delivered in a list or in a tile-based mode) and the filtering options available for narrowing down the results of a search operation. The "Categories" element includes the categories of the topics that the content and information in the e-KRP relate to[86] (namely, "Crop farming", "Forestry", "Livestock", "Environment", "Society", and "Economics") and the pages that can be accessed after selecting one of these topic categories. These categories relate to agricultural sectors (namely, "Crop farming", "Forestry", and "Livestock") or cross-sectoral themes ("Environment", "Society", and "Economics"). Finally, the "Footer" element describes the information blocks related to basic, project- and platform-related details (e.g., access to the project's social media pages, contact form and details, privacy- and cookie-related policies).

The supply of appropriately selected contextual information (namely, metadata) for each of the items appearing in a search results list (see Figure 43) allows to further facilitate proceses of information and content identification and retrieval. This element has been decided to be incorporated into the front-end design of the e-KRP in respone to **Req#4**: "*Provide contextual information (i.e., metadata) for each data object to help the end-users identify content and information of relevenace and importance to them*".

Figure 43. Contextual information provided to the end-user in relation to a data object in order to facilitate the identification of relevant content and information

The front-end development followed a procedure that started with the design of the wireframes and the collection of feedback on them in the context of the user needs analysis sessions held in the Athens consortium meeting and the Paris workshop. Based on this feedback, it moved to the development of mock-up screens (i.e., non-functional versions of the platform's interfaces very close to the real ones) and the use of these mock-up screens to achieve the proof-of-concept for the prototype e-KRP's front-end[87]. To facilitate this process and the proper execution of all the steps, meetings among the project coordination team, the back- and front-end developers, as well as UX[88] researchers and designers were organised and held regularly.

From a technical perspective, the framework that was used for the front-end development is VueJS[89]; a well-known front-end JavaScript framework for building user interfaces and single-page applications. NuxtJS[90] has also been used to add extra features around VueJS. Both VueJS and NuxtJS are open-

[86] These are the categories identified in the EURAKNOS data model. For details, the reader may refer to Deliverable 3.2 (Van Hal et al., 2020).

[87] Details about this process are also provided in Deliverable 4.4 ("EURAKNOS visuals and specifications"). The Deliverable 4.4 report was drafted in parallel to the present report (both deliverable reports were due by the end of the project).

[88] UX is the abbreviation used to refer to User Experience.

[89] https://vuejs.org/

[90] https://nuxtjs.org/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

source frameworks providing the tools to keep the code clean and well-organised with a focus on performance and ease of use. In the end, as an end product, the code produced is compiled into three things, namely: HTML, CSS, and JavaScript (able to be interpreted by a browser and displayed on the end-user's screen).

8. Testing and evaluation of the EURAKNOS e-KRP

The e-KRP's testing and evaluation took place in a dedicated session in the consortium meeting of the EUREKA project held in January 2021. The feedback received from that testing and evaluation event helped to validate the design decisions that have been made and engineered into the e-KRP, as well as collect suggestions and recommendations for further improvements to be considered in the follow-up project (i.e., EUREKA). All the details of the user testing and evaluation of the e-KRP are described in Subsection 8.1 ("Results obtained from the user test session"). Apart from the feedback collected from the EUREKA consortium meeting's testing and evaluation session, testing has also been considered in regard to a comparison of the e-KRP with other systems addressing similar needs (i.e., supply of ready-to-apply knowledge to the agricultural community). To this end, a semi-structured comparison of the e-KRP with the EIP-AGRI database[91] has taken place. The outcomes of this comparison are presented in Subsection 8.2 ("Testing the e-KRP against other existing systems").

8.1. Results obtained from the user test session

User-led testing and evaluation of the e-KRP took place in an appropriately designed session that was held in the consortium meeting of the EUREKA project in January 2021. The number of participants in the testing and evaluation session was 28. Given the multi-actor nature of the consortium of EUREKA, an adequate representation of the end-user categories targeted in EURAKNOS (farmers and foresters, advisors, and researchers) was achieved. Additionally, there has been also an overlap in the synthesis of the end-user samples that have been considered for the user needs analysis and testing/evaluation phases (given the complementarity in the two project's scope and the fact that many of the EURAKNOS partners also participate in the EUREKA project). However, there were participants who contributed to the e-KRP's testing and evaluation and had not taken part in the sessions of analysis of user needs. This way, we have managed to limit the bias in the feedback collected from the user tests.

The session started with a brief presentation of the e-KRP and the demonstration of its functionalities. After that, the session participants were given time to interact with the platform themselves to be able to become familiar with the system. The session was virtual and the collection of feedback was based on the use of mentimeter[92] and the posting of comments in the virtual meeting software's chat tool.

The collection of feedback was based on the responses provided, by the session participants, to a list of evaluation-related questions, namely:

- **Q1**: What is your opinion about working with sector-specific and theme-related pages?[93]
- **Q2**: Is there anything missing from the sector- and theme-based segmentation of content that has been already proposed?
- **Q3**: What is your opinion on the subnavigation structure that has been adopted? Any suggestions/ comments?
- **Q4**: What is your opinion on the community-buiding aspects that have been infused into the design of the platform? Any suggestions/comments?

[91] https://ec.europa.eu/eip/agriculture/en/find-connect/projects?search_api_views_fulltext=&field_proj_funding_source_list=4

[92] https://www.mentimeter.com/

[93] The main categories that have been considered for the clustering/segmentation of content/information and its delivery to the end-users are: (i) "Crop farming", "Livestock", and "Foresrty" (sector-based content segmantation); and (ii) "Economics", "Society", and "Environment" (segmentation of content based on sector-independent, cross-sectoral themes).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- **Q5**: What is your opinion on the platform's visual design? Any suggestions/comments?

Regarding **Q1** ("*What is your opinion about working with sector-specific and theme-related pages?*"), the vast majority of participants (96.3%) reported a positive attitude towards the way of clustering the content (i.e., through the adoption of both sector-related and cross-sectoral categories) and delivering it to the users. The responses of the session participants are shown in Figure 44. Some comments that have been provided with the aim to be further considered in the context of the follow-up, development work already in progress in EUREKA, can be summarised to the following points:

- The categorisation being used is strongly aligned with that used by EIP-Agri for agricultural sectors. This is indeed a strong point. The sector-based and cross-sectoral themes that are currently in use, together with keywords associated with them in order to tag data objects and organise/structure the content in the platform, will need to be further (iteratively) developed.
- Some questions/concerns have been exerted in regard to how specific concepts could relate to the selected sector-specific and cross-sectoral content categories. These concepts are: food processing, rural development, organic, cooperation, and policy.

The above comments have come in response to **Q2** ("Is there anything missing from the sector- and theme-based segmentation of content that has been already proposed?").

Figure 44. Responses of the testing and evaluation session participants to the question: "What is your opinion about working with sector-specific and theme-related pages?"

Q3 ("*What is your opinion on the subnavigation structure that has been adopted? Any suggestions/ comments?*") has focussed on the collection on feedback on the appropriateness of the subnavigation structure that has been designed in the e-KRP. This structure is shown in Figure 45 below.

Figure 45. The e-KRP's subnavigation structure

The structure presented above has received some positive feedback. Apart from that, some comments suggesting further improvements have also been provided. The main feedback and comments received in relation to **Q3** are listed below:

- I like this navigation.
- Leave the subnavigation visible also, if the user goes through one specific article.
- Adapted to main theme.
- Perhaps the subnavigation cannot be the same for each sector.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- Too small letters. Maybe using images would be better.
- Maybe country could be added?
- Look at the keywords in the EIP-AGRI database, they are approximately 20.

Despite the fact that the feedback that has been received in regard to the subnavigation structure has been mostly positive, the decision that was made and engineered into the front-end design was to use the generic label "All content" for enabling access to content (instead of the more detailed "Practices", "Actors", "Systems", "Types", "Infrastructures", "Tools", & "Technologies" labels). As regards the infusion of community-building functionalities and aspects into the design (**Q4**), the feedback that has been received was largely positive (see Figure 46 below). Specifically, 88% of the participants provided a positive response (22 out of a total number of 25 responses).

Figure 46. Responses of the testing and evaluation session participants to the question: "What is your opinion on the community-buiding aspects that have been infused into the design of the platform? Any suggestions/comments?"

Some further comments and feedback on the community-building aspects and functionalities, which have been designed in the platform, have been also provided to be taken into consideration considered in the follow-up work in EUREKA.

Figure 47. Comments and further feedback on the infusion of community-building aspects and functionalities into the e-KRP

These comments are shown in Figure 47 above and can be grouped into the points listed below:

- Are users able to search for others from their local community?
- Local community.
- Some sectors have already their community (e.g., foresters). How do we link to these?

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- The comments should be for the public OR the friends only?
- Some moderation will be necessary.

Given the fact that several of the comments listed above relate to issues (and therefore, design-related decisions) that go beyond the scope of EURAKNOS, they are considered as points to be further touched upon in the context of work in progress in EUREKA.

Finally, the testing and evaluation session participants have been prompted to provide their feedback in regard to the aesthetic elements that have been incorporated into the platform and its overall visual identity (**Q5**). As it is clearly shown in Figure 48 below, the feedback that was collected is positive.

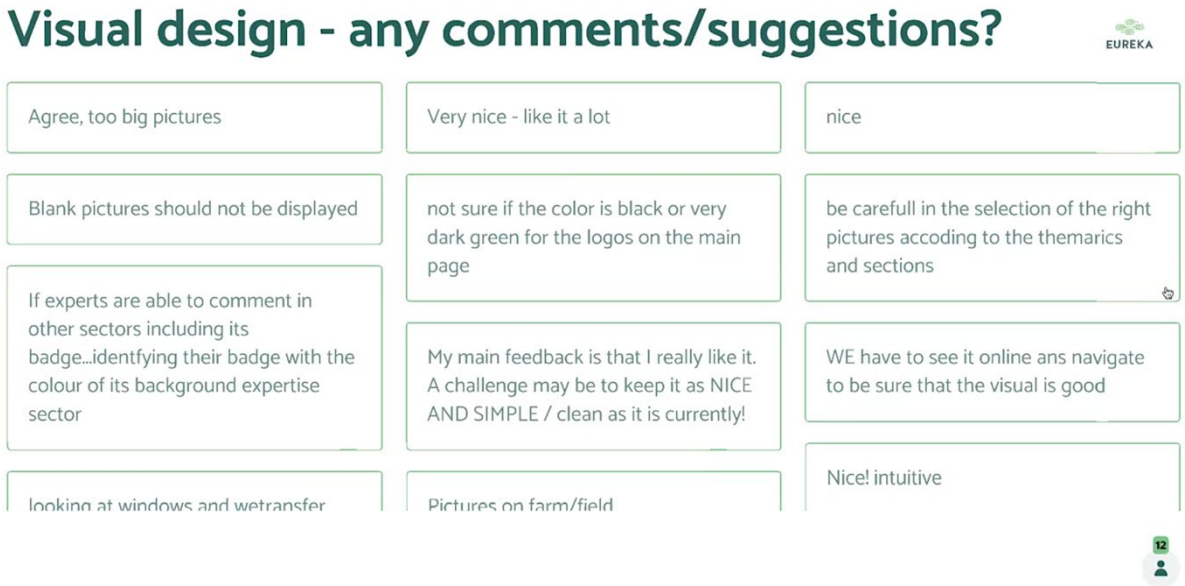


Figure 48. Responses of the testing and evaluation session participants to the question: "What is your opinion on the platform's visual design? Any suggestions/comments?"

8.2. Testing the e-KRP against other existing systems

In EURAKNOS, the testing of the e-KRP has not only been considered in terms of the supply of feedback, by end-users, but also by comparing it to other platforms aiming to serve similar needs. A system with a similar scope is the EIP-AGRI database able to be accessed via the EIP-AGRI service point's website[94]. Based on that, the current subsection aims to present the outcomes of a semi-structured comparison of the e-KRP with the EIP-AGRI's database. This comparison exercise focussed on the following aspects: (i) breadth of the content and information delivered; (ii) information architecture; and (iii) employment of visual elements and aesthetics.

8.2.1. The breadth of the content and information delivered

The initial motivation behind the conceptualisation of the idea of the EURAKNOS project and creation of the e-KRP has been the fact that there is a large amount of ready-to-apply knowledge developed by TNs, available in a variety of formats, which should be preserved and further disseminated. Given the fact that most of the TNs' KRs seize to exist after the end of the project, there is a need for a centralised repository system able to collect all this knowledge (packaged into data objects of various formats and types) and further distribute it to the agricultural community. A system aiming to serve this purpose is the EIP-AGRI database able to be accessed through the EIP-AGRI service point. The EIP-AGRI database

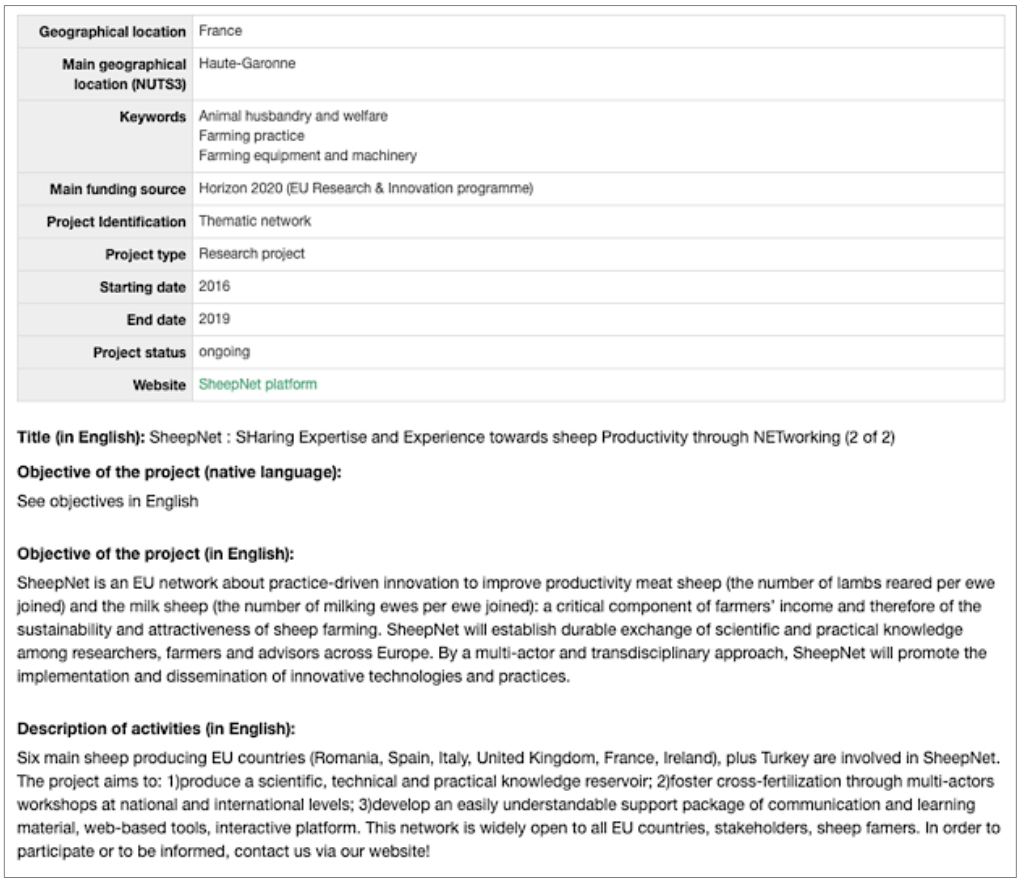
[94] https://ec.europa.eu/eip/agriculture/en/node

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

allows to access content/information mainly available in a textual format. More specifically, it provides direct access to the Practice Abstracts[95] created by the TNs (see Figure 49 below).

Figure 49. Access to the Practice Abstracts created by a TN from the EIP-AGRI database

As shown in Figures 50 and 51 above, pieces of information that become available from the EIP-AGRI's database are background information about the TN, its coordinating organisation, as well as the list of the partners involved in the TN. Access to other types of textual content/information (i.e., documents) is provided indirectly through links to the TN's KRs[96]. Access to audio-visual material is also available (see Figure 52 below).



Geographical location	France
Main geographical location (NUTS3)	Haute-Garonne
Keywords	Animal husbandry and welfare Farming practice Farming equipment and machinery
Main funding source	Horizon 2020 (EU Research & Innovation programme)
Project Identification	Thematic network
Project type	Research project
Starting date	2016
End date	2019
Project status	ongoing
Website	SheepNet platform

Title (in English): SheepNet : SHaring Expertise and Experience towards sheep Productivity through NETworking (2 of 2)

Objective of the project (native language):

See objectives in English

Objective of the project (in English):

SheepNet is an EU network about practice-driven innovation to improve productivity meat sheep (the number of lambs reared per ewe joined) and the milk sheep (the number of milking ewes per ewe joined): a critical component of farmers' income and therefore of the sustainability and attractiveness of sheep farming. SheepNet will establish durable exchange of scientific and practical knowledge among researchers, farmers and advisors across Europe. By a multi-actor and transdisciplinary approach, SheepNet will promote the implementation and dissemination of innovative technologies and practices.

Description of activities (in English):

Six main sheep producing EU countries (Romania, Spain, Italy, United Kingdom, France, Ireland), plus Turkey are involved in SheepNet. The project aims to: 1)produce a scientific, technical and practical knowledge reservoir; 2)foster cross-fertilization through multi-actors workshops at national and international levels; 3)develop an easily understandable support package of communication and learning material, web-based tools, interactive platform. This network is widely open to all EU countries, stakeholders, sheep famers. In order to participate or to be informed, contact us via our website!

Figure 50. TN details available from the EIP-AGRI database

[95] According to the EIP-AGRI's service point website (https://ec.europa.eu/eip/agriculture/en/eip-agri-common-format), Practice Abstracts (PAs) provide a common format for "*communicating about projects, activities and results*" and therefore, facilitate "*knowledge flows on innovative and practice-oriented projects from the start till the end of the project*". As a result, they constitute a means of communication and interaction among farmers, advisers, researchers and other actors in the EU.

[96] In the case that a TN's KR seizes to exist, then access to its content will also seize to be available. This is why a centralised repository system is needed.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Project partners

Contact person:	Jean-Marc Gautier
Address:	BP 42118 31321 CASTANET TOLOSAN CEDEX France
E-mail:	jean-marc.gautier@idele.fr
Phone:	+33 561754440
Partner category:	SME
Contact person:	INRA
Address:	24 Chemin de Borde Rouge, 31326 Castanet-Tolosan France
E-mail:	f.corbiere@envt.fr
Phone:	+33 (0)5.61.19.32.34
Partner category:	Researcher
Contact person:	AGRIS-Sardegna - Research Unit : Genetics and Biotechnology
Address:	S.S. Sassari-Fertilia Km 18,6, 07040 Olmedo ((SS)) ITALY
E-mail:	acarta@agrisricerca.it
Phone:	+39 (079) 2842 379
Partner category:	Researcher

Figure 51. Details of the partners involved in a TN available from the EIP-AGRI database

Further details

Links to other website(s):

Institut de l'Elevage platform
Inn'ovin
Pâtre

Audiovisual material: SheepNet explain in 2 minutes
SheepNet explain in 2 minutes
SheepNet explain in 2 minutes
SheepNet Solution - Body Condition Scoring Toolkit
SheepNet solution - Technical note on health control and management of abortion
SheepNet solution - Optimum lamb birth weight
SheepNet solution - Recording the causes of lamb losses
SheepNet solution - shed design and use
Solution SheepNet - Oscar un outil de diagnostic des avortements
SheepNet Solution - Planning des étapes clés
SheepNet solution - OSCAR

Description of the context of the project:

The EU is only 85% self-sufficient in sheep meat and is the largest importer of sheep meat worldwide. An increase in EU ewe productivity by 0.1 lambs reared per ewe joined would increase self-sufficiency in sheep meat to 92%. Ewe productivity is one of the main factors impacting on profitability of prime lamb production.

Additional information:

SheepNet is an open and easy access network, Stakeholders are fostered to participate, propose solutions and innovation via our interactive platform and via the different social medias.

Figure 52. Access to audio-visual material through the EIP-AGRI database

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

In EURAKNOS, the main focus has been on the creation of a platform having the capacity to address the needs of various end-user types by allowing to easily search and access a variety of content/information available in a wide range of types and formats. What has become evident from the research undertaken in Task 2.1 ("Data impact assessment of TNs") and reported in Deliverable 2.1 (Mosquera Losada et al., 2019) is that documents are the most prevalent category of data objects used for the delivery of content and information. However, the preferences of users in access to knowledge that have been reported in Deliverable 3.2 (Van Hal et al., 2020) offer a strong indication of the requirement to deliver content in other, more popular, formats and types as well (for instance, videos and presentations). Based on that, and as illustrated in Figure 53 below, the data object categories delivered from the e-KRP are:

- documents;
- videos;
- presentations;
- infographics; and
- podcasts.

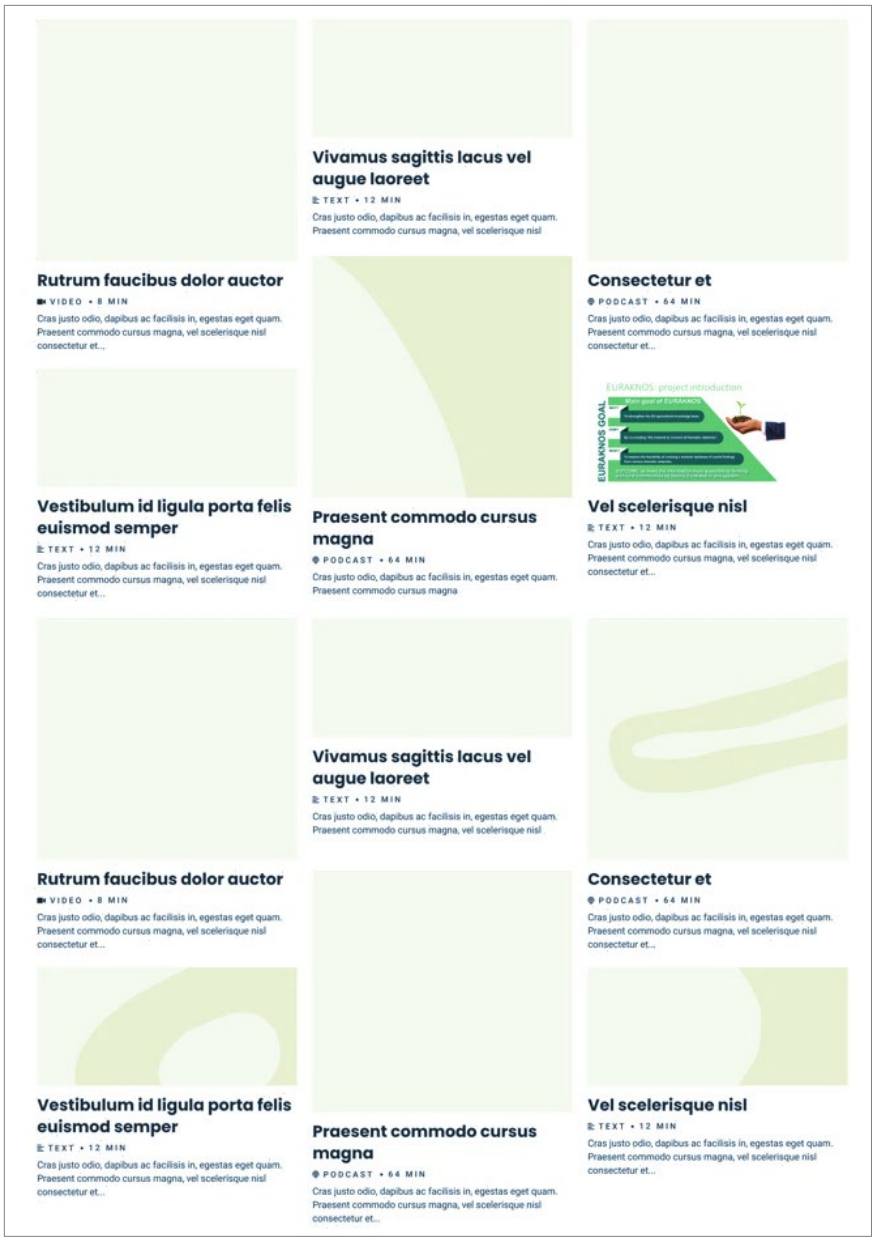


Figure 53. Access to a breadth of content and information, in a wide range of file formats, from the EURAKNOS prototype e-KRP

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

8.2.2. Information Architecture

Differences between the EIP-AGRI's database and the e-KRP exist in the Information Architecture of those systems as well. In the case of the EIP-AGRI database, a more complex navigational structure is used, as made evident from its menus (see Figure 54), mostly because of the fact that the database of EIP-AGRI is part of a broader system (the EIP-AGRI service point) aiming to deliver information/content that goes beyond the supply of information on TNs' results. As shown in Figure 54, the menu available at the top of every page of the website of the EIP-AGRI service point is both broad and deep. This may have implications for the system's ease of use and the way it actually facilitates the identification and retrieval of content and information in just a few clicks.

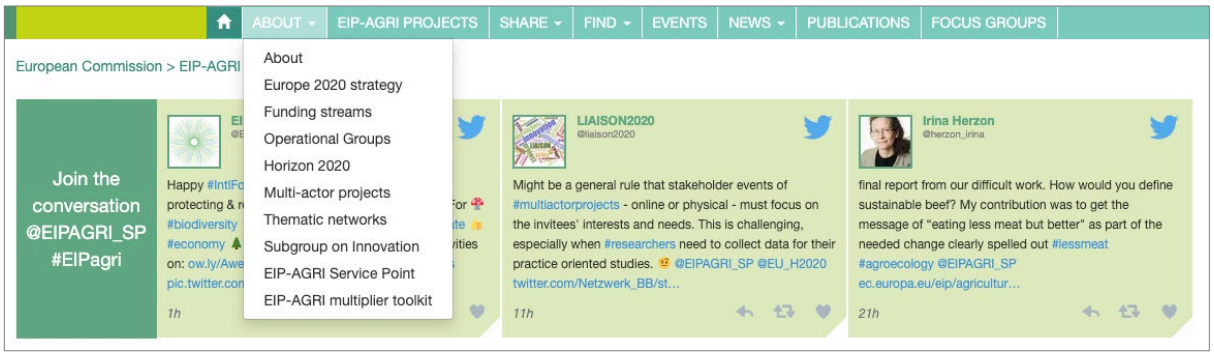


Figure 54. The EIP-AGRI service point website's main navigational structure

On the other hand, in the case of the EURAKNOS e-KRP, a clearer and more shallow structure has been employed for the platform's main menu to facilitate easy access to content and information (see Figure 55 below).

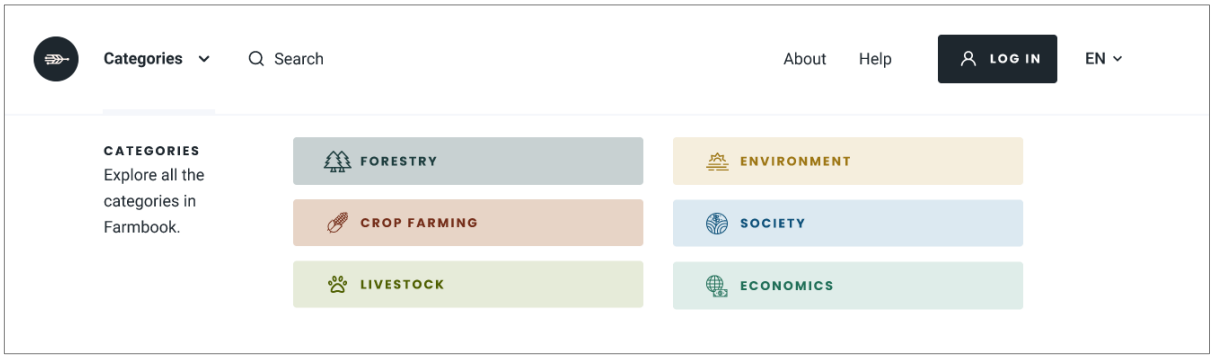


Figure 55. The EURAKNOS e-KRP's main navigational structure

Regarding the execution of search operations in the two systems, a dedicated search area is available in the case of the EIP-AGRI database. Even though no specific visual aids are used to demarcate this area in the context of the EIP-AGRI database, the user can proceed to a search operation by providing values to the text boxes being made available. The search operation is based on the submission of queries consisting of user-defined or predetermined terms. Additional search-related parameters can be used to get more elaborate results. These parameters relate to: (i) the use of logical operators (i.e., "AND", "OR", and "NOT") to make explicit how the terms of a search query should be used together; (ii) the selection of the funding scheme that the projects (to be retrieved as search results) should be associated with; and (iii) the geographical area to which the search results should relate to. Despite the fact that these options can indeed facilitate the identification and retrieval of fine-grained results, it is not always straightforward what the input to the text boxes should be or what the meaning of the pre-defined values is. The search function available available from the EIP-AGRI database is illustrated in Figure 56 below.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

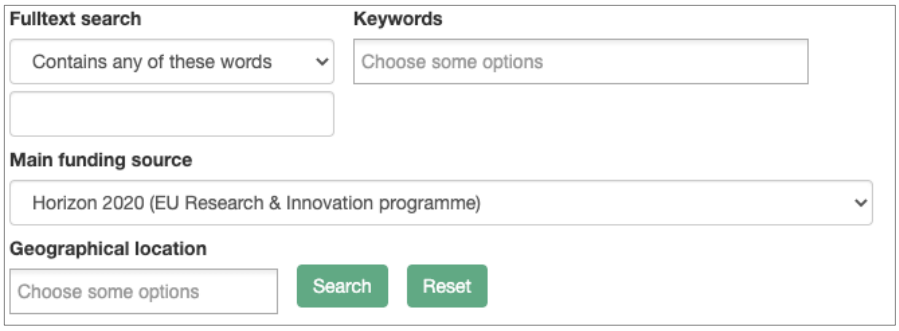


Figure 56. The search function available to search for content and information in the EIP-AGRI database

In the case of the e-KRP, there is use of more straightforward prompts and tools for the execution of search operations. More specifically, any search for content and information may be done through the submission of a search query consisting of any number of search terms of the preference of the user.

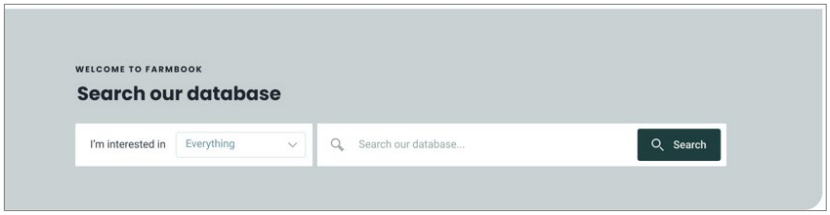


Figure 57. The search function of the EURAKNOS e-KRP

The search function of the e-KRP is very close to those available in state-of-the-art systems and therefore, very easy to use (see Figure 57 above). The refinement of search results can be made by using some filtering options available to the end-user (see Figure 58). As made evident from Figure 57, the search function of the EURAKNOS e-KRP is demarcated (and thus, easily spotted) and straightforward to be used without any additional prompts needed to be provided.

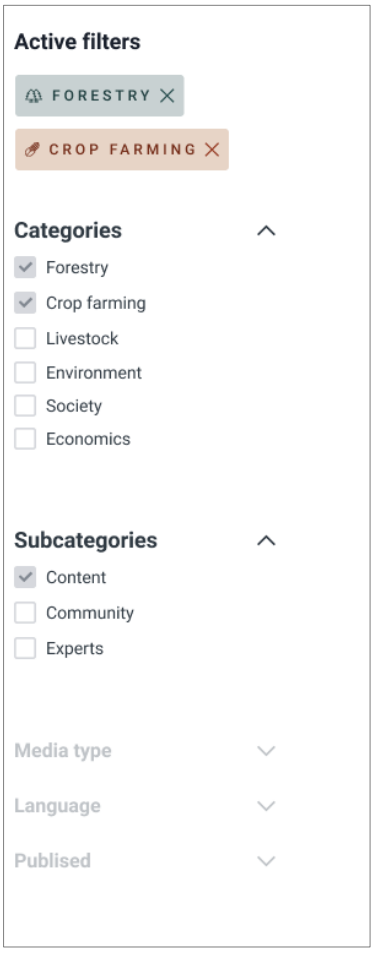


Figure 58. Filtering options provided in the e-KRP to elaborate the search results

As far as the delivery of search results is concerned, the mode used in the case of the EIP-AGRI database is that of a list, where the titles of the items constituting the search results are provided (see Figure 59 below). The EURAKNOS e-KRP employs two different modes for the delivery of search results: (i) a list view mode; and (ii) a tile-based mode. In both cases, apart from some basic result-related, text-based information, visual aids are also being provided (namely, a figure that is indicative of the content/scope of each result item). The screenshots illustrating these two different modes of search result display are provided in Figures 78 and 79 in Section 11 of the report ("e-KRP user guide").

A final remark regarding the delivery of content and information from the EIP-AGRI database and the EURAKNOS e-KRP has to do with how this information/content has been clustered. In the case of the EIP-AGRI database, content/information has been clustered around TNs. In EURAKNOS, there is another approach followed. More specifically, the end-user can access content and information clustered around agricultural sectors ("Crop farming", "Forestry", and "Livestock"), as well as broad, cross-sectoral themes (namely, "Environment", "Society", and "Economics"). This way of content and information organisation appears to be straightforward and closer to the needs of the end-users.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Published on	Title	Geographical location	Project type
08/02/2021	Soil Hydrology research platform underpinning innovation to manage water scarcity in European and Chinese cropping systems. Shui	Spain	Research project
08/02/2021	NIVA - A New IACS Vision in Action	Netherlands	Research project
20/01/2021	Eureka	Belgium	Research project
13/01/2021	LIAISON: Better Rural Innovation: Linking Actors, Instruments and Policies through Networks	Germany	Research project
24/11/2020	SUREFISH- FOSTERING MEDITERRANEAN FISH ASSURING TRACEABILITY AND AUTHENTICITY	Italy	Research project
12/11/2020	Optimising bio-based fertilisers in agriculture - Providing a knowledge basis for new policies (LEX4BIO), Optimising the use of bio-based fertilizers in the EU	Finland	Research project
09/11/2020	Confronting Obesity: Co-creating policy with youth	United Kingdom	Research project

Figure 59. Results of a search operation in the EIP-AGRI database. The display of the results is dominated by the use of textual information.

8.2.3. Employment of visual elements and aesthetics

From the aesthetics perspective, in EURAKNOS, there has been an effort to create a platform that looks and feels like many of the contemporary applications used in our everyday lives. A core consideration in the context of this effort has been to create easy-to-use and visually appealing interfaces, providing all the information to the user without overwhelming him/her with unnecessary textual details. As part of this effort, each of the agricultural sectors and themes (used for clustering content and information) have been associated with a colour. This colour coding allows the user to have an easy overview of the sector or theme that the content/information, or other features/functionalities, relate to. In the case of the EIP-AGRI database, less attention has been paid to the aesthetics part. The user interface is dominated by textual information, which makes the system informative but not that user-friendly. The difference between the aesthetics-related approach in the two systems is shown in Figures 60 and 61.

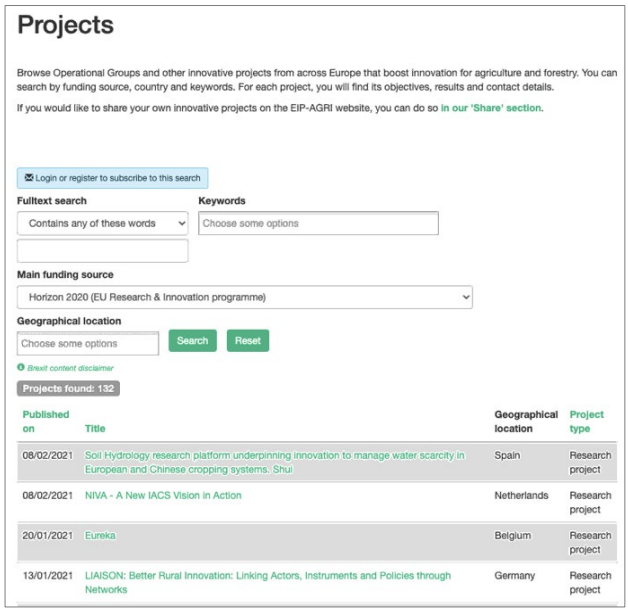


Figure 60. The landing page of the EIP-AGRI database

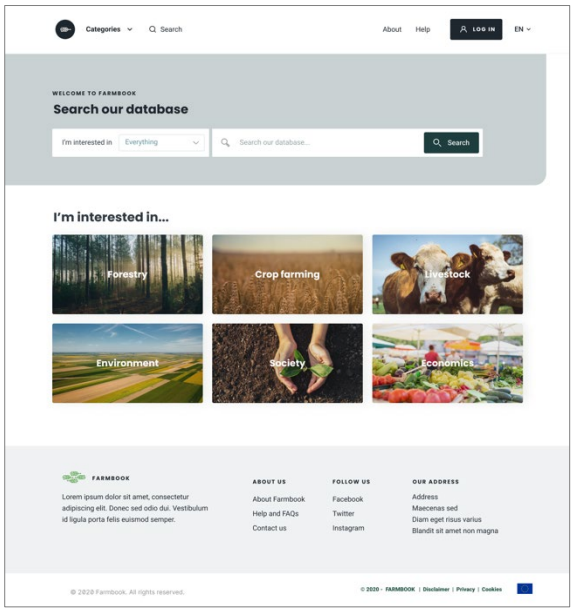


Figure 61: The landing page of the EURAKNOS e-KRP

Some key results of the comparison presented in this section are summarised in Table 25 below.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Table 25. Key results of the comparison between the EIP-AGRI database and the EURAKNOS e-KRP

	Content & information delivered					Information Architecture		Visual elements & aesthetics	
	docs	videos	presentations	infographics	podcasts	TN-oriented	sector- & theme-oriented	text-first	visual aspects - first
EIP-AGRI database	☑	☑				☑		☑	
EURAKNOS e-KRP	☑	☑	☑	☑	☑		☑		☑

9. Sustainability of the EURAKNOS e-KRP

What has been made evident from the outset of the EURAKNOS e-KRP design and development is that it constitutes a multi-faceted process with a range of issues (beyond the technical aspects of such an endeavour) requiring attention. One of these issues relates to the sustainability of the platform; i.e., the continuity in the infrastructure's existence and service provision beyond the lifetime of the project. To enable sustainability, there is a need for a plan explicitly identifying and framing all the dimensions that should be taken into consideration and the means of securing the necessary resources. Within the context of the work undertaken in EURAKNOS, we have been able to identify and perform an **in-depth analysis of all the dimensions relating to the sustainability** of the EURAKNOS e-KRP. The results of this analysis have been used to formulate and propose a sustainability framework (namely, the EURAKNOS sustainability framework) described in detail in Deliverable 4.1[97]. Some key aspects of it are presented in Subsection 9.1 below ("The EURAKNOS sustainability framework"). Specific solutions to consider as part of a concrete sustainability plan are provided in Subsection 9.2 ("Towards an e-KRP sustainability plan").

Given the fact that the issue of sustainability has not been considered and addressed to an adequate extent by the already implemented TNs (which has been the main motivation for the conceptualisation and launch of EURAKNOS), the suggestions made by EURAKNOS can be regarded as an initial attempt to touch upon this complex issue. It needs to be stressed that all the sustainability-related dimensions and solutions (identified and proposed in EURAKNOS) are not for the exclusive use of EURAKNOS itself, but rather constitute a useful resource to be considered by future TNs as well. The work that has been undertaken in EURAKNOS is being followed-up by EUREKA; therefore, all the suggestions and solutions available from EURAKNOS concerning sustainability will be taken up by EUREKA to further elaborate on this important matter.

9.1. The EURAKNOS sustainability framework

When discussing the sustainability of a technological infrastructure like the EURAKNOS platform, the issue that immediately comes to the forefront is that of ensuring the financial resources necessary for keeping the service provision going which entails, among others, hardware and technical equipment maintenance as well as updates in software and content. The financial dimension of sustainability is indeed the most critical one. However, what has been made evident from the review of literature and the analysis made in Task 4.1 ("Feasibility and added value of an e-KRP"), and reported in Deliverable 4.1, there are other dimensions to consider as well; yet, in direct relation to the financial aspect. The sustainability framework proposed in EURAKNOS, and used to make concrete suggestions about the e-KRP's existence beyond the project's end, is shown in Figure 62 below.

[97] Gernert, M. et al. (2019). Report on added value and feasibility of e-KRP. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 4.1 – WP4.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

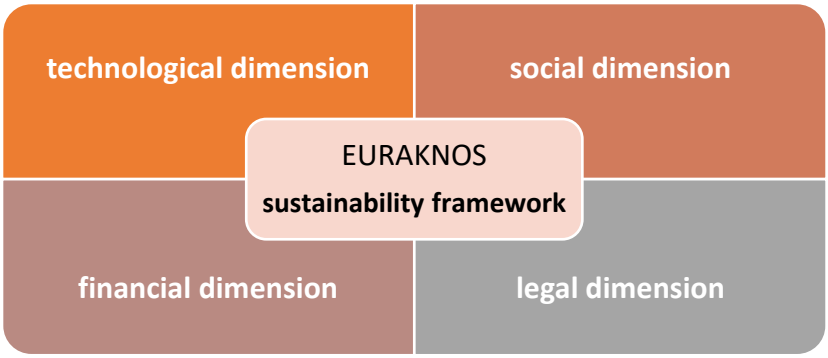


Figure 62. The EURAKNOS sustainability framework

The details of each of the dimensions that constitute the EURAKNOS sustainability framework (namely, the technological, financial, legal, and social dimensions) are provided in Deliverable 4.1. For reasons of completeness, Table 26 below contains brief descriptions of what each dimension is about.

Table 26. Brief description of each of the EURAKNOS sustainability framework's dimensions

#	Dimension of the sustainability framework	Brief description of the dimension
1	Technological dimension	The technological dimension encompasses all the technology solutions that need to be considered for the development of the platform and the provision of services.
2	Financial dimension	The financial resources required to continue maintaining and updating the hardware and software components of the e-KRP, as well as updating the platform's digital content, especially after the completion of the project. Costs for keeping the equipment running and ensuring the continuity in service provision, including the human labour required, need to be considered.
3	Legal dimension	The legal dimension is about: • providing guidelines and recommendations to future TNs about how to protect their intellectual outputs; • ensuring the provision of content under specific licensing terms; and • issues of copyrighting the EURAKNOS e-KRP solution to ensure the protection of the EURAKNOS partners' IPRs (especially, in the case that a third-party organisation assumes the responsibility of service provision after the project's end).
4	Social dimension	The social dimension touches upon issues such as: • the provision of high-quality and up-to-date content to ensure uptake by all stakeholders; • the assurance of easy access to any interested party; • the continuous delivery of content and services; and • the identification of issues of exploitation.

The sustainability-related issues raised in Subsection 9.2 focus on the financial and legal dimensions of our framework. However, there are strong interactions among all the dimensions of the sustainability framework. For instance, the description of the social dimension makes clear that there is a direct link to the technological dimension (especially in relation to the first three bullet points listed in the social dimension's description in Table 26). Moreover, the need to allocate financial resources for securing the e-KRP's viability makes evident an indirect relation between the social and financial dimensions.

Another point requiring attention is that the legal dimension of the sustainability framework is not just limited to protecting the EURAKNOS project's intellectual outputs, but also relates to the strengthening of the capacity of TNs to safeguard the outcomes generated in their context. Given the fact that the protection of the TNs' intellectual property is an issue that has attracted not much attention so far, it is indeed needed to provide future TNs with guidelines and recommendations on how to protect their

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

own intellectual property and disseminate it in a "safe" way. Incentivising TNs to pay attention to such issues is of vital importance to the e-KRP's viability and sustainability as well. It is the condition required to turn the e-KRP into a hub for disseminating and promoting creativity and innovation in agriculture. Therefore, a framework of recommendations and guidelines to future TNs for facilitating decisions on their intellectual property management and protection is provided in Subsection 9.2. Given this issue's importance, these recommendations are provided separately, in this section of the report, rather than in Section 10 ("Recommendations for future Thematic Networks"). Regarding the intellectual property management and protection issues that are relevant to the e-KRP, these are outlined in the EURAKNOS project's exploitation plan described in Deliverable 5.6 ("Dissemination activity report")[98].

9.2. Towards an e-KRP sustainability plan

9.2.1. Ensuring revenue streams for the e-KRP's financial sustainability

There are three (3) distinct scenarios that have been proposed and analysed in Deliverable 4.1 in regard to continue deploying the EURAKNOS e-KRP and providing services after the project's completion. Each scenario encompasses specific solutions for securing the required resources and covering costs. These scenarios are the following:

- **Scenario #1**: In-house hosting of the e-KRP by a consortium or third-party organisation.
- **Scenario #2**: Outsourced e-KRP hosting and service provision.
- **Scenario #3**: Hosting the e-KRP in data centers of the European Commission.

The fact that the work done in EURAKNOS is being followed-up by EUREKA, where the e-KRP continues to be developed from a prototype to an MVP[99], means that these three scenarios are of importance to EUREKA as well (or probably, of greatest importance to EUREKA).

Scenario #1 offers the potential of direct access to the technological infrastructure, thus allowing the technical personnel to execute configuration, maintenance, update, and upgrade tasks conveniently and directly. In this scenario case, there are two business models that may be adopted with the aim to ensure revenues, namely: (i) the "freemium" model; and (ii) the "crowdfunding" model. By taking into account the vision of EURAKNOS to create an EU-wide community for all the AKIS actors, by connecting the KRs of TNs, it appears that the "crowdfunding" model is a better fit for guaranteeing money flow and thus, the sustainability of the technological infrastructure and the service related to it. According to the SAGE Encyclopaedia of the Internet[100], crowdfunding is "*the sourcing of small contributions from a large number of individuals*". In the case of the e-KRP, these "individuals" are the TNs that can benefit from the supply of the e-KRP-related service. The benefits that future TNs can potentially obtain relate to:

- Avoiding all the work related to the design, development, and deployment of a dedicated repository for disseminating the content and information produced in their context, as well as the tasks related to the maintenance of this infrastructure and associated services.
- Enjoying greater visibility within the AKIS community by being able to disseminate their results to a larger audience via a centralised platform, of a well-known value, dedicated to the delivery of ready-to-apply, practical agricultural knowledge.
- Being able to design and implement more targeted dissemination strategies and achieving a higher impact by being able to disseminate their results to a larger audience.

[98] Deliverable 5.6 was also in the process of development by the time that the present report was drafted.

[99] MVP stands for "Minimum Viable Product"; for details about what a Minimum Viable Product is, the reader may refer to https://en.wikipedia.org/wiki/Minimum_viable_product.

[100] http://sk.sagepub.com/reference/the-sage-encyclopedia-of-the-internet-3v/i1840.xml

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Access to the service provided by the e-KRP (and thus, the potential to enjoy the benefits listed above) will be purchased by an interested TN by providing an appropriately estimated rate of its dissemination budget as a compensation for utilising the e-KRP infrastructure and the associated service. Adherence to this plan will become acknowledged from the TN's conceptualisation phase and the methodology proposed for its development.

Scenarios #2 and #3 present the advantage of removing the burden of maintenance from the shoulders of the project partner(s) or the organisation(s) in charge of keeping the service provided to the public, by outsourcing the task to the web service provider's or the European Commission's technical staff. In any case, opting for one of these scenarios may entail greater complexity and issues relating to security and quality of service assurance, protection of intellectual property, etc. More details on these options are available in the report of Deliverable 4.1 (Gernert et al., 2019). In the case of Scenario #2, the issue of ensuring revenue streams for outsourcing the maintenance of infrastructure and provision of service to a web service provider still exists. This may not be the case in Scenario #3 however.

The scenario case of having one or more project partners, or even a third-party organisation, in charge of the e-KRP maintenance and service provision, by using in-house technical equipment and hardware, is assumed to be capable of facilitating a greater sense of "belongingness" and "ownership" to the TNs and thus, it is regarded as the scenario solution that is worth considering.

9.2.2. Guiding Thematic Networks to protect their intellectual outputs

TNs are Research and Innovation projects oriented towards the identification and delivery of solutions to problems related to a specific theme. Solutions to problems are pursued in the context of the Multi-Actor approach, namely by bringing together stakeholders representing various actor categories from the AKIS community. This means that the process of pursuing solutions, and producing outputs as part of the solution-seeking process, is collective by nature. Hence, the acknowledgment of each partner's contribution and the protection of the produced intellectual outputs are of vital importance. However, what has been made apparent from the investigation in EURAKNOS is that the protection of intellectual property is an issue that has been poorly addressed so far. Table 27 below shows some results gathered from this investigation.

Table 27. Example cases of IPR management from existing TNs

	website of TN	data repository of TN
License available detailed enough and clear	**FERTINNOWA** […] *All rights in the design, text, graphics and other material on this site and the selection or arrangement thereof is copyright of FERTINNOWA or other third parties.* ***Users may read, download and/or take copies of*** *(by storing the content in an electronic form or by reducing to print)* ***the information in whole or in part for the users' own use in research and education, or other non-commercial purposes****. Ownership of the materials rests with the copyright owner. Any other use of the materials on our site (including reproduction for purposes other than noted above and alteration, modification, distribution or republication) without our prior written permission is strictly prohibited.* […] https://www.fertinnowa.com/wp-content/uploads/2017/11/FERTINNOWA-website-terms-and-conditions.pdf	**SMART-AKIS** […] *The* ***Smart-AKIS website and platform*** *are licensed under a* ***Creative Commons Attribution 4.0 International*** license (https://creativecommons.org/ licenses/by/4.0/). This license applies to all contents on the website except for cases, on the page or in the document, differently indicated and except for contents aggregated from other websites, in which case the original licensing applies. […] […] *The Smart-AKIS Platform is* ***open to farmers****,* ***advisory services****,* ***agriculture consultants****, and* ***farming equipment providers****, setting up an open community where these groups can interact.* […] https://www.smart-akis.com/index.php/legal-notice-privacy-cookie-policies/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

License available not many details available	**HNV-LINK**: Educational Material […] *All resources in the package are* ***Open-Source materials*** *under* ***CC BY-NC-SA****. You may* ***freely use, for non-commercial purposes only, any elements or as a whole, also modifying as fit, as long as you cite the project, its funding and the respective authors****. Observe copyrights for images: all images are by HNV-Link unless otherwise specified*. […] http://www.hnvlink.eu/outputs/educationalmaterials/	**AFINET** […] *Please note that users of content (“****Users****”) shall respect the* ***applicable license conditions****. Download and use of content from Zenodo does not transfer any intellectual property rights in the content to the User!* […]
No license available	**EURODAIRY** There is no clear information on how the content can be used; only a notice on pdf materials regarding the dissemination level.	**INNO4GRASS** There is no any kind of information in relation with IPR, nor on INNO4GRASS website neither on **encyclopaedia pratensis**.

From the information listed in the above table, it can be easily figured out that in nearly all the (sample) TN cases only some generic references are provided about whether their outputs are protected or not. This leaves by default a wide margin of interpretation. The only exception to this rule is the Smart-AKIS TN, which has disseminated data objects, conveying useful project-related information and outcomes, under specific terms and conditions of (re-)use specified in a Creative Commons license. However, the distribution of dissemination material under a license is only one of the issues needed to be considered in the context of an intellectual property management and protection strategy and plan.

In the light of these findings, the purpose of this section is to provide guidelines and recommendations to future TNs about issues requiring attention for the protection of their intellectual property. Making such a contribution is of big significance to the the e-KRP’s sustainability and longevity because it paves the way towards good result-sharing practices and avoidance of misuse. Before providing details, it is important to make clear that a discussion on the protection of intellectual property is very relevant to TNs, especially if we consider the multi-actor approach they adopt in their search for solutions and the creation of results. Intellectual property rights management is a fundamental aspect of the promotion of innovation in any domain. Intellectual property rights should be conceptualised as “*certain kinds of exclusive rights to intellectual capital and creations of the mind*” (European Commission, 2017).

To frame a context for the discussion about the issues that should be considered for the establishment of an intellectual property protection strategy, it needs to be stressed that there are three (3) different types of intellectual property rights: (i) industrial property rights; (ii) copyright and related rights; and (iii) trade secrets. Given the nature of the work undertaken in TNs and the solutions/results generated by them, copyright appears to be the means of intellectual property rights with the best fit to the need to protect the intellectual outputs of TNs. At this point, we should distinguish between the intellectual outputs that are created in the context of a TN and the dissemination material produced to make them known to the community (e.g., publications, presentations, promotional videos, leaflets, infographics, etc.). Based on this distinction, it is the intellectual outputs of a TN (e.g., a novel method that has been developed to provide a solution to a particular problem faced by cattle breeders) that (probably) need to be protected as the exclusive intellectual property of (some of) the TN’s consortium members. Taking steps for the protection of the TN’s intellectual capital is of vital importance for the exploitation of the TN’s results by the entities involved in it, as well as establishing awareness on how any external entity (not involved in the TN) can build upon these results. This does not mean that the dissemination material, produced to render the TN results known, should not be distributed under specific terms and conditions specifying how to (re-)use the information contained in it.

9.2.2.1 Guidelines and recommendations for the protection of the intellectual outputs of TNs

The guidelines and recommendations for the protection of the intellectual outputs of TNs come in the form of directions provided in the context of a product’s commercialization process, which also fits the

The EURAKNOS project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No: 817863

case of a TN-related solution development. According to the European Commission (2017), this process comprises the steps/phases listed below and shown in Figure 63 below:

- Idea;
- Concept;
- Prototype;
- Production; and
- Marketing.

Figure 63. The process of producing a solution in the context of a TN (adapted from European Commission, 2017)

As made evident from the diagrammatic representation illustrated in the above figure, the process of solution creation appears to be linear. However, given the fact that any feedback that may be collected from the application of the solution (i.e., the "marketing" phase in the chain of phases) could be utilised to further elaborate on the solution's idea (by feeding into the first step in the pipeline), we could also consider it as being circular in nature (see Figure 64 below).

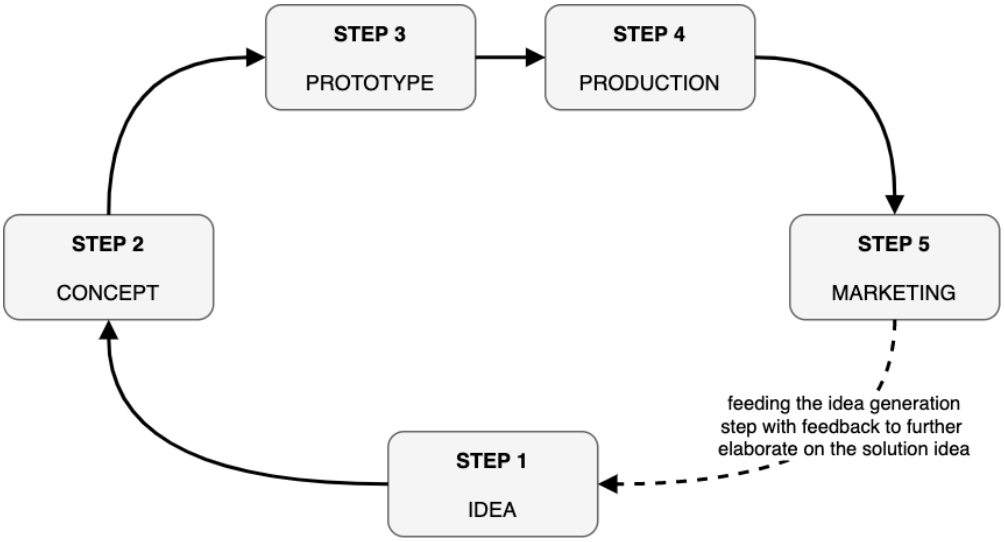
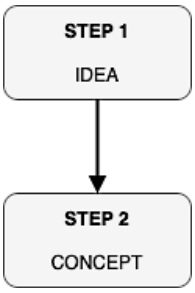
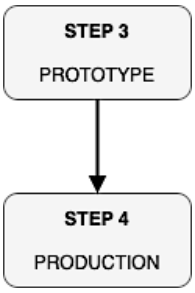


Figure 64. The process of producing a solution in the context of a TN conceptualised as being circular in nature (adapted from European Commission, 2017)

The guidelines and recommendations to future TNs in respect of helping them establish strategies and plans to protect their intellectual outputs are provided in Table 28 as concrete directions linked to each of the steps/phases of the solution creation process shown above.

#	Step/phase in the solution creation process	Guidelines and recommendations
1		• **Identify** the outputs of the project that should become exploitable. • **Identify** the form and format in which these outputs should become available. • **Decide upon** the project partners that should/can hold the right(s) of the output(s) exploitation (Are they the same with the partners that were involved in the solution creation process?). • **Identify** and **specify** the kind(s) of expertise and know-how required to produce the exploitable outputs. • **Decide upon** whether (and which of) the produced outputs need to be protected (e.g., via copyright).

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

		• **Identify** and **specify** whether there are any intermediate results that should be protected as well.
2		• **Identify** the partners that should be involved in the production of the exploitable outputs and **negotiate/co-decide** their involvement and contribution to the process. • **Make explicit** the contribution of each project partner to the process of creation of the exploitable outputs. – What does each partner bring to the process? – Is the expertise/know-how/expertise brought from the involved partners protected by some kind of intellectual property rights? • **Identify** the project partners whose work and contribution should be protected.
3		• **Specify** the project partners that should have the right to exploit the project's intellectual outputs and the outputs that are able to exploit (Will the project partners not involved in the output creation process have the right to expoit the project's outputs?). • **Specify** the terms and conditions under which any third-party entities (not involved in the project) should be able to use the project-related outputs and/or build upon them (e.g., via contractual aggrements). • **Specify** the project partners that should assume the responsibility of granting third-party entities the right to further distribute intellectual output-related dissemination material through their dissemination channels. • **Specify** the licensing schemes/types under which the dissemination material of the TN should become available, and further distributed, via third-party dissemination channels; **specify** terms and conditions of (re-)use.

9.2.2.2 Licensing of dissemination material in order to distribute it through the various dissemination channels

This section relates to the last point in the list of guidelines/recommendations for intellectual property protection provided in Table 27 ("***Specify** the licensing schemes/types under which the dissemination material of the TN should become available, and further distributed, via third-party dissemination channels; **specify** terms and conditions of (re-)use*."). More specifically, a license for the distribution of TN-related dissemination material via the e-KRP is proposed, given that fact that no such consideration has been made in (the sample of) TNs (except for the case of the Smart-AKIS TN). However, this issue needs to be explicitly addressed by TNs and for this reason, a recommendation particularly pointing to this has been outlined above. To this end, the licensing framework that has been considered is Creative Commons[101]. This choice has been made by taking account of the popularity of the framework and the range of the licensing options available from it to address various needs and data object (re-)use cases. Table 28 lists the main Creative Commons license types together with short descriptions of them.

Table 28. Main Creative Commons licenses together with a short description of them

#	License name	License description	URL
1	Attribution 4.0 International (**CC BY 4.0**)	This license allows re-users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. This license allows for commercial use. **CC BY** includes the following elements: • **BY** – Credit must be given to the creator	https://creativecommons.org/licenses/by/4.0/

[101] https://creativecommons.org/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

#	License name	License description	URL
2	Attribution-ShareAlike 4.0 International (**CC BY-SA 4.0**)	This license allows re-users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms. **CC BY-SA** includes the following elements: • **BY** – Credit must be given to the creator • **SA** – Adaptations must be shared under the same terms	https://creativecommons.org/licenses/by-sa/4.0/
3	Attribution-NonCommercial 4.0 International (**CC BY-NC 4.0**)	This license allows re-users to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. **CC BY-NC** includes the following elements: • **BY** – Credit must be given to the creator • **NC** – Only non-commercial uses of the work are permitted	https://creativecommons.org/licenses/by-nc/4.0/
4	Attribution-NonCommercial-ShareAlike 4.0 International (**CC BY-NC-SA 4.0**)	This license allows re-users to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms. **CC BY-NC-SA** includes the following elements: • **BY** – Credit must be given to the creator • **NC** – Only non-commercial uses of the work are permitted • **SA** – Adaptations must be shared under the same terms	https://creativecommons.org/licenses/by-nc-sa/4.0/
5	Attribution-NoDerivatives 4.0 International (**CC BY-ND 4.0**)	This license allows re-users to copy and distribute the material in any medium or format in non-adapted form only, and only so long as attribution is given to the creator. The license allows for commercial use. **CC BY-ND** includes the following elements: • **BY** – Credit must be given to the creator • **ND** – No derivatives or adaptations of the work are permitted	https://creativecommons.org/licenses/by-nd/4.0/
6	Attribution-NonCommercial-NoDerivatives 4.0 International (**CC BY-NC-ND 4.0**)	This license allows re-users to copy and distribute the material in any medium or format in non-adapted form only, for non-commercial purposes only, and only so long as attribution is given to the creator. **CC BY-NC-ND** includes the following elements: • **BY** – Credit must be given to the creator • **NC** – Only non-commercial uses of the work are permitted • **ND** – No derivatives or adaptations of the work are permitted	https://creativecommons.org/licenses/by-nc-nd/4.0/
7	CC0 1.0 Universal (**CC0 1.0**) Public Domain Dedication	CC0 (aka CC Zero) is a public dedication tool, which allows creators to give up their copyright and put their works into the worldwide public domain. It allows re-users to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions.	https://creativecommons.org/publicdomain/zero/1.0/

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

By taking into account the non-commercial nature of the EURAKNOS e-KRP[102] and the purpose that the platform aspires to serve, as well as the guide available in the Creative Commons website for enabling decisions on the type(s) of license(s) to choose[103], the license that has been decided to be attributed to the e-KRP's data objects, to further disseminate them, is Attribution-NonCommercial-NoDerivatives 4.0 International (**CC BY-NC-ND 4.0**)[104]. More details about this license type are provided in Table 29.

Table 29. Details of the CC BY-NC-ND 4.0 license selected to attribute to the e-KRP's data objects

Attribution-NonCommercial-NoDerivatives 4.0 International (**CC BY-NC-ND 4.0**)	**Permissions provided**
	The re-user can **share**; i.e., copy and re-distribute the material in any medium or format.
	License features
	• **Attribution:** The re-user must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses the re-user or the re-user's use. • **NonCommercial:** The re-user may not use the material for commercial purposes. • **NoDerivatives:** If the re-user remixes, transforms, or builds upon the material, he/she may not distribute the modified material. • **No additional restrictions:** The re-user may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

9.2.3. Concluding remarks

To sum up, the employment of the "crowdfunding" model in the context of the scenario encompassing the in-house hosting of the e-KRP by a consortium/third-party organisation, with the data objects being distributed under a CC BY-NC-ND 4.0 license, are the building blocks of the EURAKNOS sustainability plan. The establishment of a sustainability plan is, of course, an ongoing effort already taken up by the EUREKA project (the follow-up, sibling project of EURAKNOS). However, the sustainability of the e-KRP is inextricably linked to the source TN project's intellectual property management strategies and plans. Given the fact that this issue has not been adequately addressed so far, a framework of guidelines and recommendations to future TNs is outlines. The aim of this framework is to strengthen the capacity of the TNs to frame and protect their intellectual capital, a condition that comes as a requirement for the promotion of innovation in the domain of agriculture.

10. Recommendations for future Thematic Networks

The work that has been done during the two years of duration of EURAKNOS is multi-dimensional and is not limited to the technical implementation of the prototype e-KRP. However, given the focus of the activities implemented in Task 4.2 for the platform development, which are technical in nature, the recommendations provided in this section of the report relate to technical matters and serve the purpose of facilitating the establishment of a communication channel between EURAKNOS and future TNs. Specifically, these recommendations have originated from the EURAKNOS data model which has been created to capture, in a structured way, all the information necessary for coming up with the (minimum) set of metadata for the description/characterisation of the TNs' data objects.

To begin with, a significant effort has taken place to systematically categorise the types of data objects available per each data object category (i.e., the types of documents, videos, presentations, etc.). This categorisation is not considered to be definitive by any means, but rather a starting point for discussion

[102] A discussion on existing types of platforms, as well as the type the EURAKNOS e-KRP belongs to, is available in Deliverable 4.1 (Gernert et al., 2019).

[103] The guide is called "Creative Commons License Chooser" and is available at https://creativecommons.org/choose/#.

[104] This is the license under which the data objects from the Smart-AKIS KR are also made available.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

within the AKIS community. Based on the data object categories and associated types provided in our model, the first recommendation to future TNs is to make use of it as a reference point when creating their data objects and embarking upon a process of appropriately clustering them. In the case that the existing data object types (described in the EURAKNOS data model) are not sufficient to cover all categorisation needs (meaning that one or more new types should be considered), any suggestions for new data object types will be highly welcome. Such interaction cases, if any, will be the best way to prove that the EURAKNOS data model is open to changes and adaptations to better serve the need of the community for a robust representation of the produced knowledge.

Another issue directly related to the use of commonly agreed and accepted data object categories and types, for categorising the data objects to be created by future TNs, is that of being consistent when it comes to naming data objects. This can be achieved by making use of specific and commonly agreed naming conventions. Such an approach can be highly effective when needing to identify (and retrieve) specific types of content and information. For instance, we may refer to the automated process of data object classification that was executed as part of the assembly of the input for the e-KRP repository[105]. More specifically, to classify the data objects harvested by the EURAKNOS web crawler in one of the types identified in our data model, we searched for specific keywords in the data object filenames, indicative of the type that the data object could relate to. However, the conventions used by data object creators for naming data objects have not been consistent. For example, the use of the word "practice" in the filename of a data object could be a strong indication that the specific item could be a practice abstract. However, this was not always the case. Based on that, and by drawing upon the case of a text-based data object (i.e., a document), a recommendation to future TNs concerning this issue is to use the following naming convention:

- Start with the title of the document (or a short version of it).
- Include the name of the document type (e.g., practice abstract).
- Continue with the name of the Thematic Network in which the document has been produced.
- Provide the date on which the document was produced in the format YYYYMMDD.
- Use the underscore character ("_") to denote blank spaces.

To give an example of how this convention could be applied, the filename of a practice abstract about data object naming conventions, which was created on the 26th of November 2020, would be:

"naming_conventions_to_follow_practice_abstract_EURAKNOS_20201126.docx"

Similar approaches should be taken for naming other data object categories as well, by considering the types identified in the EURAKNOS data model per each data object category.

Apart from the above, we also need to consider that data objects become available as files of a specific file type. This file type is indicated by the extension in the name of the file. So, for instance, the ".docx" extension used in the above filename example is indicative of a document file able to be processed by a word processor. For the delivery of content and information, whatever the format for conveying it is, it is highly recommended to use file types able to be processed by non-proprietary software or file types able to be processed by software available in the majority of computer systems. In the case that a very specific software application would be needed to process a file of a specific type, this should be indicated by the person or the entity responsible for this data object's release.

As mentioned above, the basic motivation behind the development of the EURAKNOS data model has been to systematically identify all the entities related to the creation of data objects, any relations that may hold between them, as well as their properties. By combining all this with the help of a knowledge representation structure (data model/ontology) it has become feasible to identify and explicitly define

[105] See Subsection 6.3.1 ("Data cleaning") for details.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

the minimum set of metadata for the annotation/characterisation of the TN data objects[106]. To enable the dissemination of the data objects of future TNs through the EURAKNOS e-KRP (whether a TN opts to use the e-KRP as its KR from its outset or to make use of its own, dedicated repository), it is strongly recommended that the EURAKNOS metadata set is used as the minimum metadata set for annotating the data objects of TNs. This recommendation does not preclude the use of more metadata properties to capture specific features and complexities of the content and information intended to be conveyed via a TN's data objects. This recommendation serves the purpose of a minimum requirement for the adoption of a common "communication language", from all TNs, in an attempt to pursue a high impact for the knowledge produced by TNs.

A final recommendation, not related to the EURAKNOS data model, is linked to the need to deliver a nice user experience no matter what device/platform is used (i.e., personal computer, tablet, or mobile phone), especially when it is needed to access content/information in a textual format. What has been made evident from the work in Task 4.2 is that documents (namely, information available in a textual format) are mostly available as files in a pdf format (doc and docx files were also harvested by the web crawler, but they were too few). However, viewing content from a pdf file when using a device other than a personal computer is not user-friendly. Thus, when it comes to documents, content should also be delivered in a way able to facilitate cross-platform access with the text being adjusted to the screen size and/or orientation. To this end, apart from delivering documents as files of a text-related file type, we also recommend the delivery of documents as html files where tags are used to attribute specific properties to the text (indicative of how the text should be displayed to the user).

To sum up, the (technical) recommendations to future TNs can be summarised to the following points:

- **Recommendation #1**: Make use of the data object categories and types identified in the EURAKNOS data model for the categorisation of data objects.
- **Recommendation #2**: Make use of a commonly agreed convention for naming the TN data objects.
- **Recommendation #3**: Make use of the EURAKNOS metadata set as the minimum metadata set for the annotation of TN data objects.
- **Recommendation #4**: Deliver documents in an html format (in addition to providing them as pdf or other relevant file types) to facilitate cross-platform access to content.

11. e-KRP user guide

After having provided all the details of the activities related to the work in Task 4.2 for the development of the prototype e-KRP, and before making some concluding remarks and discussion about the steps to be further taken (as part of the work in progress in EUREKA), a detailed guide of the e-KRP is made available. The guide aims to present the basic functionalities and features of the platform by focusing upon the labelling and navigational structures and prompts used to facilitate access to the content, as well as the options provided to filter and narrow down search results.

All the information in this section is accompanied by screenshots added to facilitate the understanding of the platform functionalities and features. The design of the user experience has been driven by the need to deliver easy and straightforward access to content and information either in a user-led manner (by enabling the user to type and submit a search query indicative of what he/she is in search of) or in a more guided and structured manner by making available dropdown menus and content/information selection options. In any case, there has been an intention to provide content and enable access to it in just a few clicks and, on the other hand, not overwhelm the user with unnecessary information or distracting aesthetic elements.

[106] Details about the process that has led to the EURAKNOS metadata set creation and definitions of the respective metadata properties are available in Subsections 6.3.3.1 ("Updating the EURAKNOS data model") & 6.3.3.2 ("The EURAKNOS metadata set").

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

11.1. e-KRP landing page

The landing page of the e-KRP is shown in Figure 65.

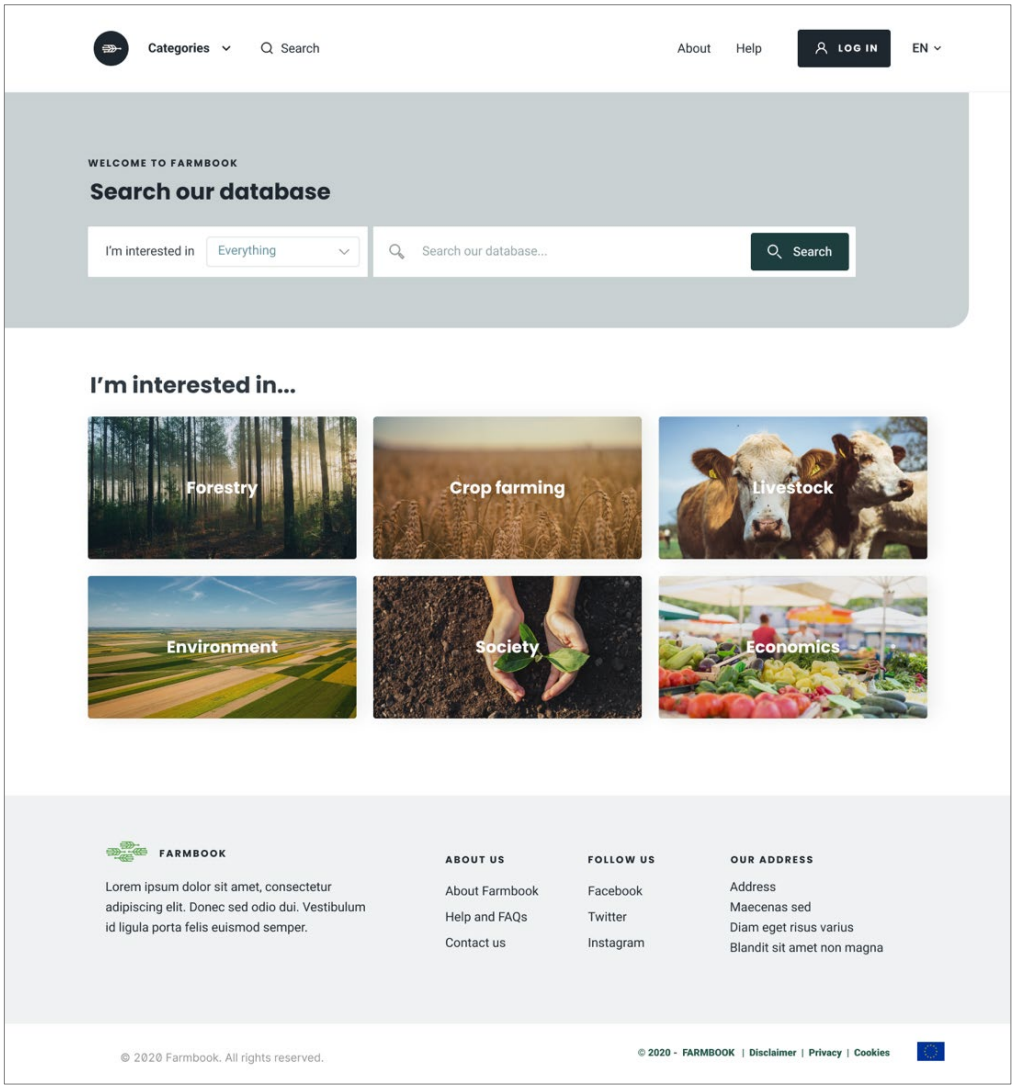


Figure 65. The landing page of the EURAKNOS e-KRP

On this page, as well as on all the interfaces/pages of the platform, there are visual elements indicative of the platform's identity. These elements can be seen both at the top and bottom of the page. The landing page provides two options for accessing content. More specifically, there is the option of typing a search query (using one or more search terms) in the search bar and then submitting it by clicking on the search button. The search bar has been placed at a prominent place on the page and is indicated with the help of the magnifying lens icon and the "Search our database…" text. The use of the gray box for framing the search area helps to make this free-form search option even more prominent.

Apart from the use of custom-made queries for searching for content in the e-KRP repository, there is also the option for more structured access to the content. This feature is made available with the help of the tiles beneath the "I'm interested in…" label. The tiles in the top row enable access to content relating to the major agricultural sectors (i.e., "Crop farming", "Forestry", and "Livestock") by clicking on them. The tiles in the second row do not relate to agricultural sectors, but rather enable access to content and information linked to broader, cross-sectoral themes. These themes are "Environment", "Society", and "Economics". In both cases, there is potential to access smaller portions of the content available in the repository; namely, the content associated with a sector (top-level tiles) or theme (bottom-level tiles). The same type of access to content is feasible by using the dropdown menu that

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

is made available when clicking on the "Categories" link at the top of the page (see Figure 66). This link is not provided just on the landing page but on every page of the e-KRP. Therefore, there is potential to access content about any sector or theme without needing to return to the platform's landing page.

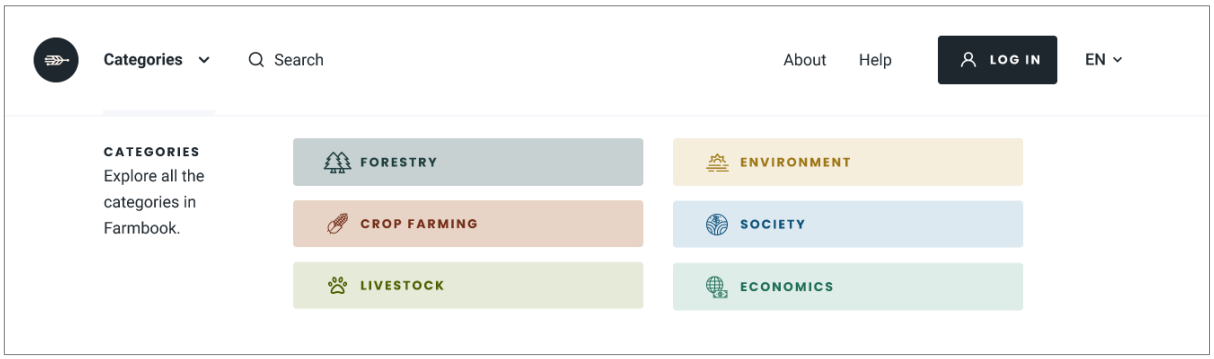


Figure 66. Dropdown menu of the sector- and theme-related categories of content able to be accessed by clicking on the link labelled "Categories" at the top of every platform page

The e-KRP allows also access to basic project information (namely, contact details, links to the project's social media pages, help, and FAQs), as well as its privacy- and cookie-related policies. These pieces of information can be accessed with the help of the links in the grey-coloured box at the bottom of every e-KRP page. The login button, as well as the option for language selection, relate to functionalities that will be added as part of the work in EUREKA.

The platform page with details on the privacy-related policy (accessed by clicking on the "Privacy" link) is shown in Figure 67 below. The form to be filled-in to come in contact with a platform representative is illustrated in Figure 68.

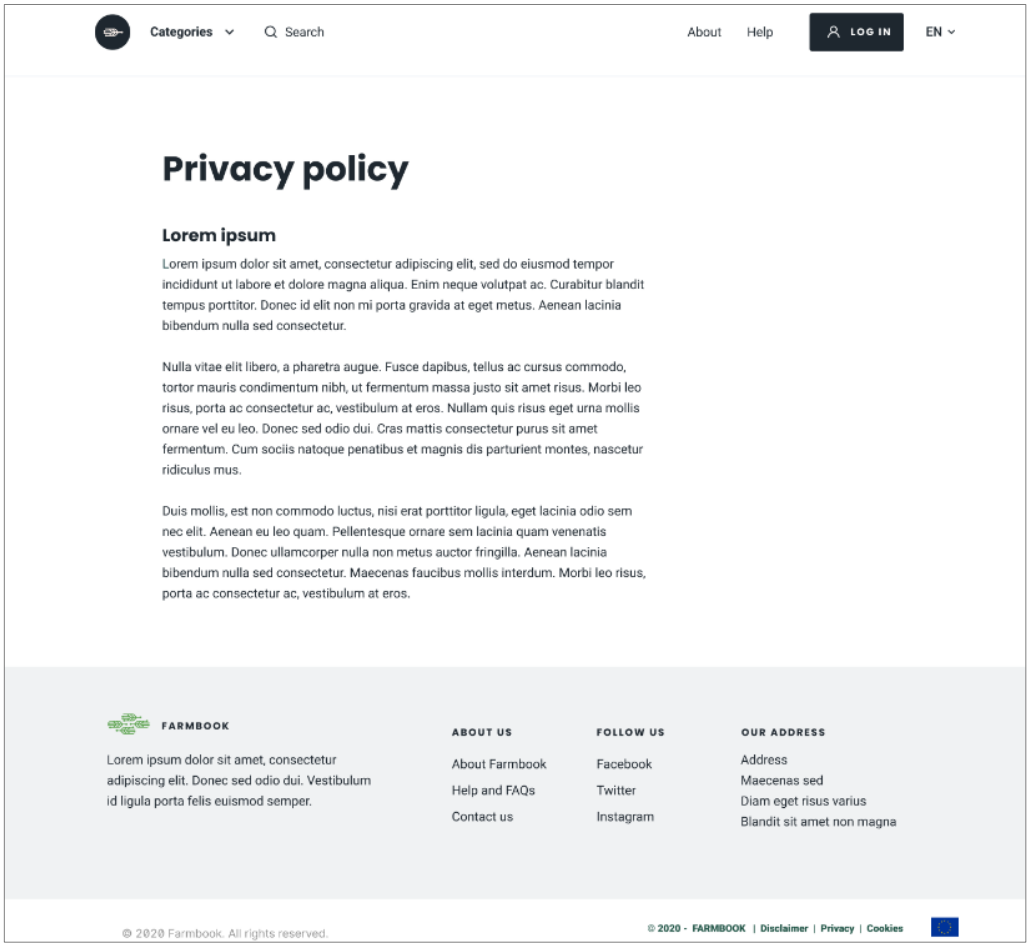


Figure 67. e-KRP page with details about the adopted privacy-relayed policy

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

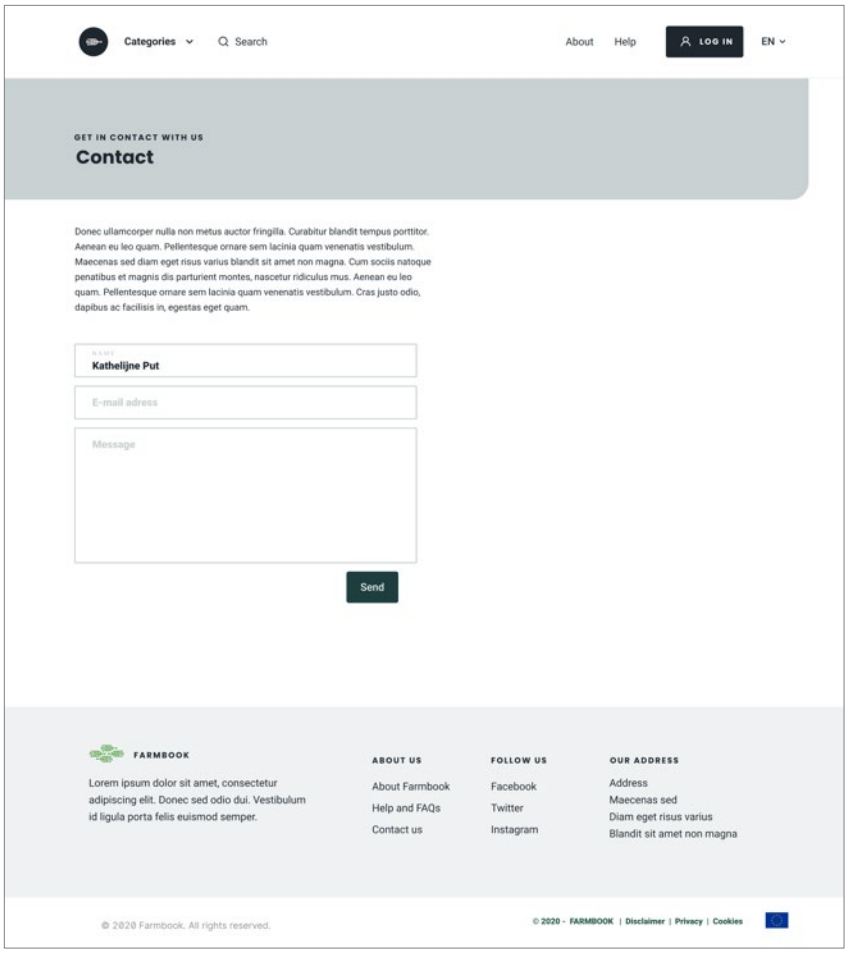


Figure 68. The "Contact us" page of the e-KRP

11.2. Sector- and theme-related pages

Figure 69 provides a bird's eye view of the e-KRP sector- and theme-related pages. These pages can be accessed by either clicking on one of the category links in the dropdown menu illustrated in Figure 66 above or by clicking on one of the tiles available in the e-KRP landing page. What is made evident from the screenshots shown in Figure 69 is that there is a set of common design elements in all these views. First of all, there is a specific colour associated with each sector/theme. This colour code is evident in every e-KRP page related to the sector or theme. In addition, at the top of each page, there is a search bar allowing to search for content and information by typing in and submitting a query.

Each sector- and theme-related page provides the same menu of options under the "All content", "Community", and "Experts" labels. By choosing the "All content" option, there is potential to access all the content and information associated with the specific sector or theme. This information/content is available via data objects in a variety of types and formats. The details related to each data object are the category they belong to (i.e., document/text, video, podcast, infographic, or presentation), an estimate of the time needed to go through their content, as well as a short description of their content. A figure, indicative of the content, is also available. The "Show more" button at the bottom of the page allows to load more sector- or theme-related data objects.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

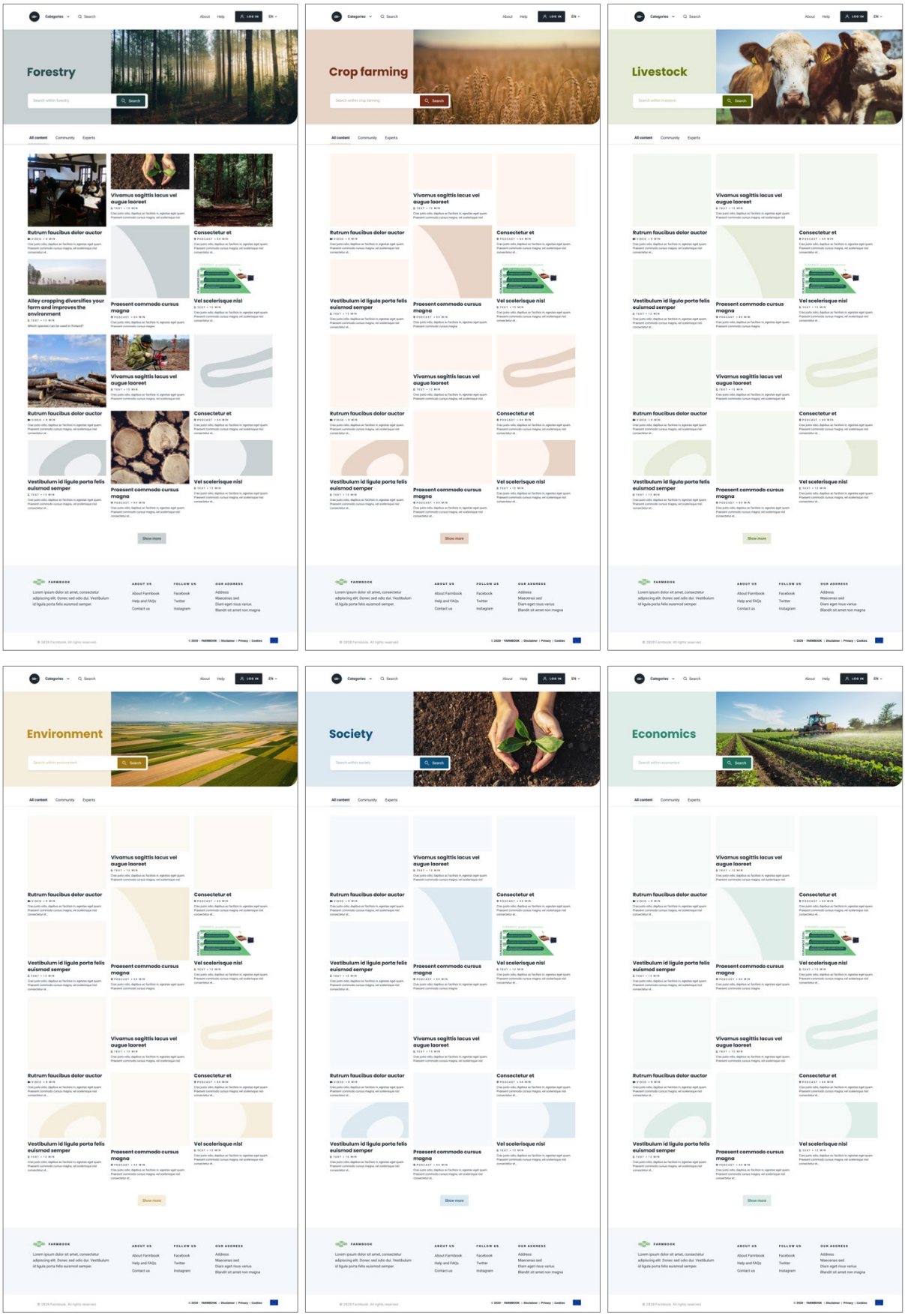


Figure 69. The sector- and theme-related pages able to visited by making a selection from the relevant menu or by clicking on one of the landing page's tiles

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Figure 70 below shows the e-KRP page related to the sector of Forestry.

The "Community" option directs to the page that can be used for communication and interaction with other community members/e-KRP users. On that page, there is potential to post a topic for discussion and start a thread of messages related to the specific topic[107]. The discussion page, as well as a thread of messages posted for a specific topic of discussion, are shown in Figures 71 and 72 respectively.

Apart from the option to interact and exchange opinions with other community members on a specific discussion topic, there is also potential to come into contact with experts in a sector or theme through the provision of those experts' contact details. As an example, a list of experts in the sector of Forestry is shown in Figure 73. The expert-related details are listed in alphabetical order in respect to the names of the experts. In the case that a long list exists, the "Show more" button allows to load more expert-related contact details.

Figures 74, 75, 76, and 77 show screenshots of the e-KRP pages associated with the views of an article, a factsheet, a podcast, and a video, regardless of the sector or theme they may relate to. What is made evident from these screenshots is that different content and information delivery options are adopted depending on the category/type of the data object (conveying the information/content) and its format. Users are able to post comments in regard to the content of a data object (see Figures 75, 76, and 77).

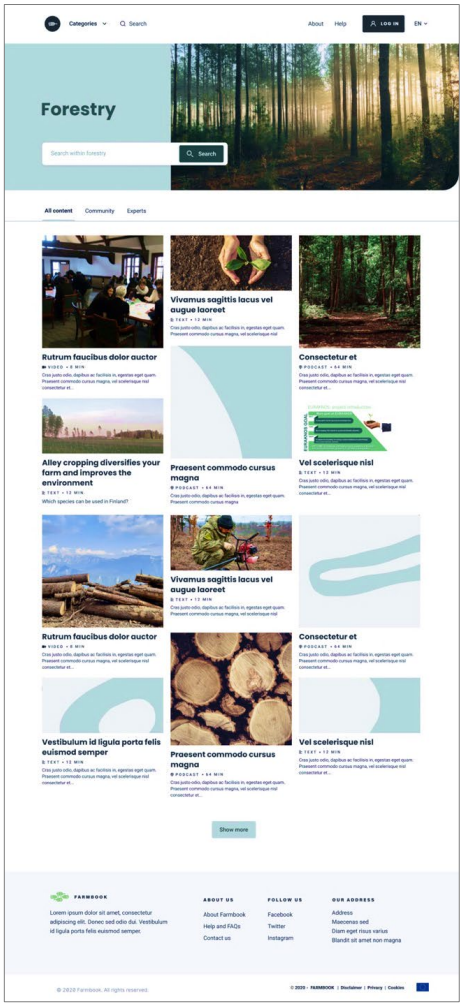


Figure 70. The e-KRP page related to the sector of Forestry

[107] This functionality has not been built in the e-KRP. It will be developed in EUREKA. However, the visual elements related to this functionality have been already designed in EURAKNOS so as to provide an overview of the look and feel of a functionality mentioned as an important feature both in the EURAKNOS Vision Paper and in the feedback collected from the marketplace session in the Paris workshop. The EURAKNOS Vision Paper is available at: http://www.acta.asso.fr/fileadmin/ressources/Europe/EURAKNOS_Vision_Paper.pdf.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

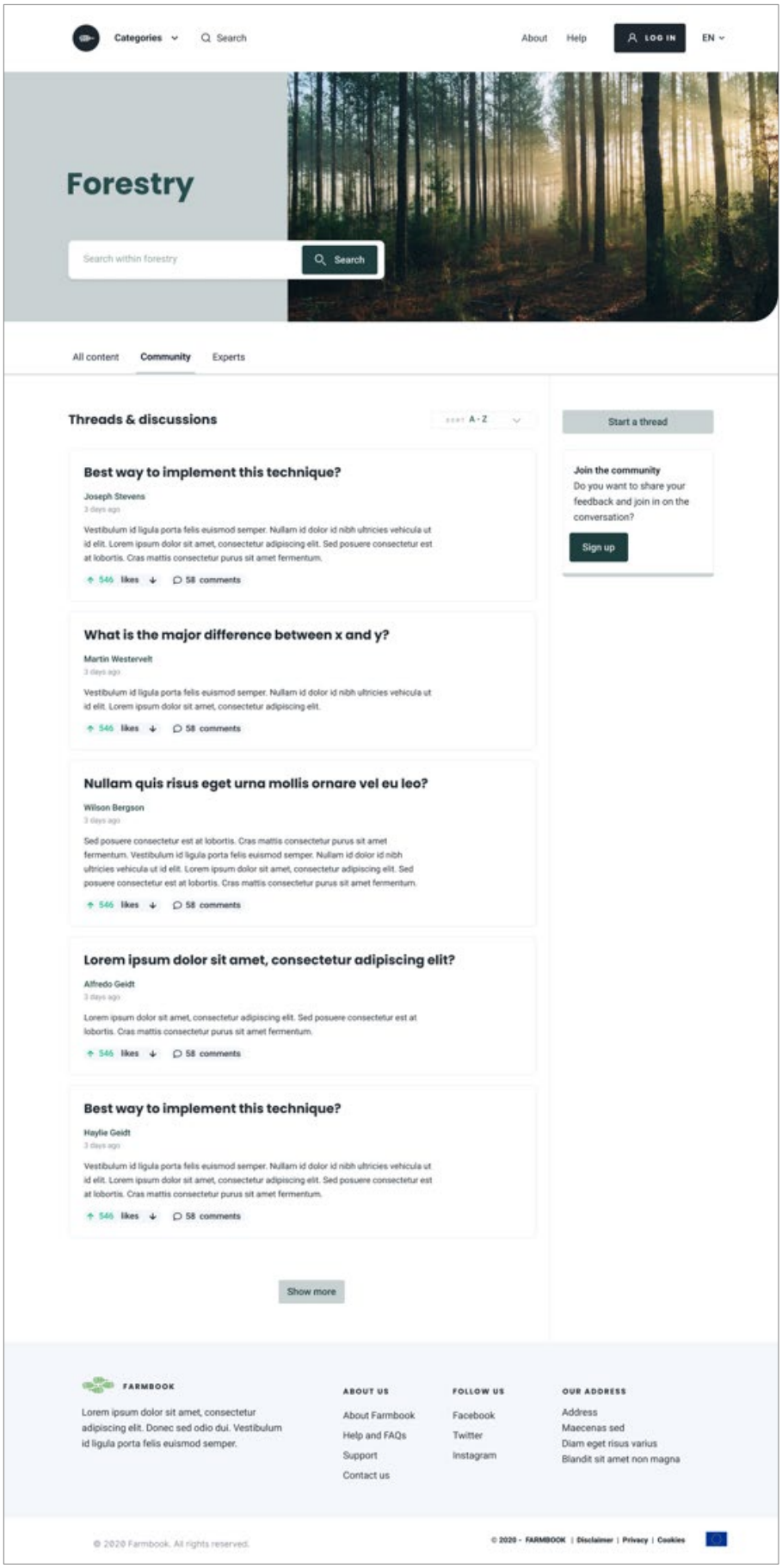


Figure 71. Topics of discussion in the "Community" page

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

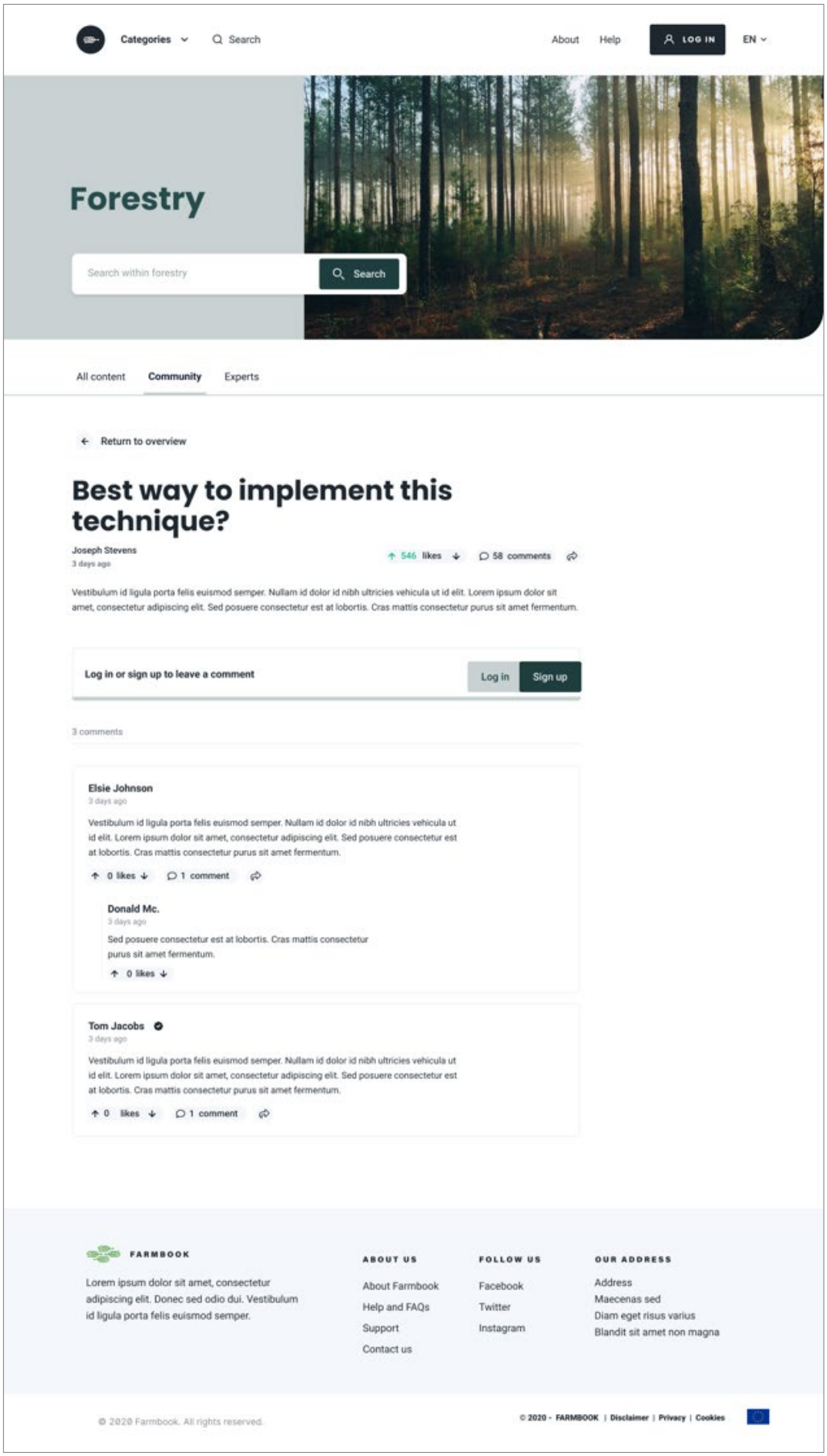


Figure 72. Message threads related to a topic of discussion

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

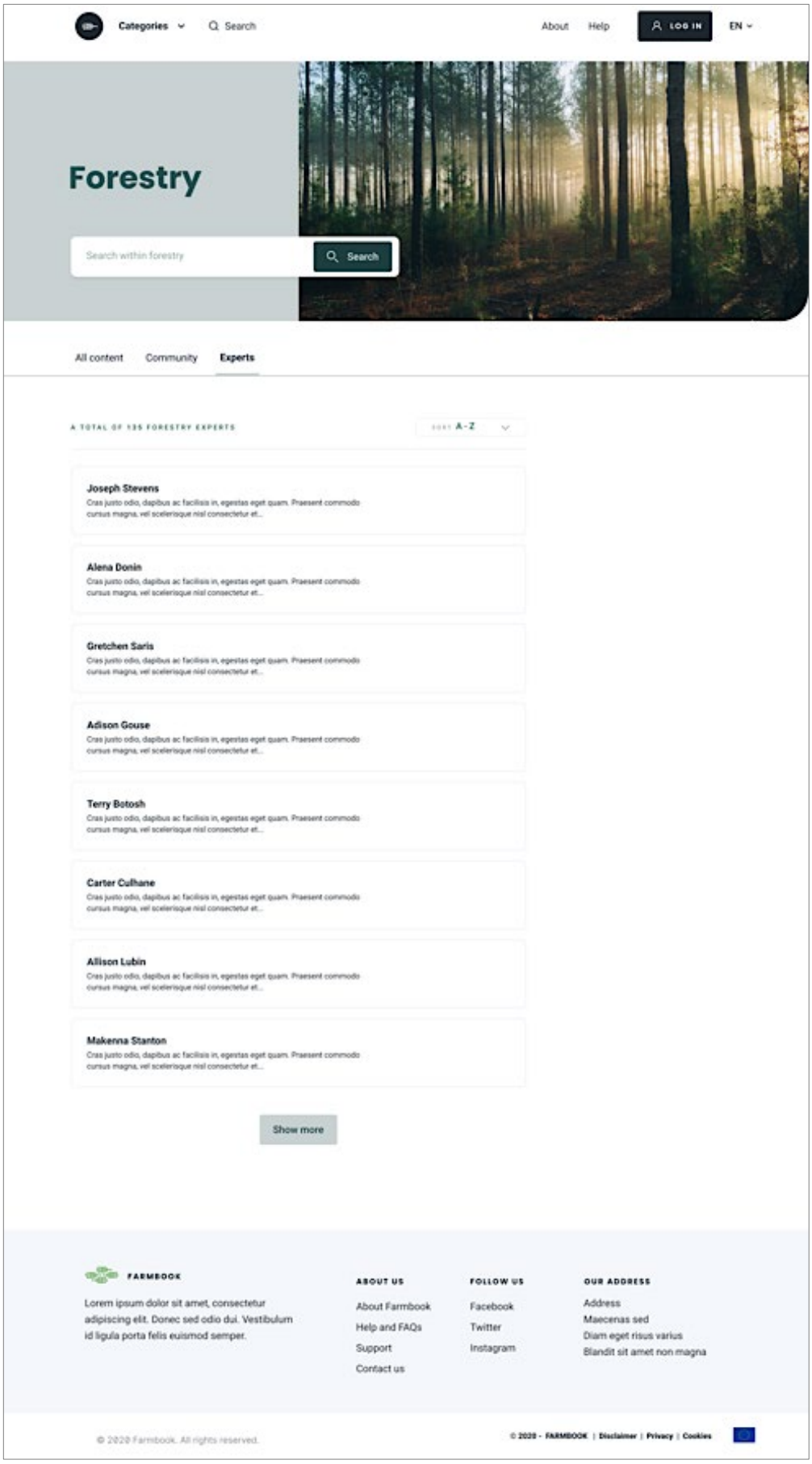


Figure 73. List of Forestry experts

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Figure 74. View of an article

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Categories Search About Help LOG IN EN

RETURN TO OVERVIEW 554 likes 0 comments

Alley cropping diversifies your farm and improves the environment

Which species can be used in Finland?

June 4, 2019

By **Mercedes Rois**, **Michael Den Herder** and **Iiris Mattila**

Afinet

THE WHAT AND WHY

Opportunities for agroforestry in Finland

Alley cropping, or planting woody perennials rows in arable or vegetable fields, is an innovative idea worthy of exploration by farmers seeking both an additional long term income, rather than income based solely on annual production, and to increase the environmental resilience of their system.It is advisable that the trees and shrubs planted should have some of the following characteristics: i) produce a product or multiple products (e.g., timber, nuts) with an acceptable local market, ii) have deep roots to reduce competition with the crops, iii) do not produce allelochemicals or acid foliage that would prevent some crops growing under them.Companion crops, planted in the alleys between the tree rows, may be: 1) cereal and forage crops; 2) fruits, berries, ornamental or aromatic/medicinal plants; or 3) biomass producing crops.In the initial stages, the growing environment in the alley will be favorable to row crops requiring full sun. As trees grow, they will increase shade, water and nutrient competition and humiditylevels, decreased temperatures and wind, so shade tolerant species will be more suitable then.The management of trees includes weed control, pruning and thinning.

Alley cropping system on contour lines with apple trees, cherries, pears and plums together with comfrey and aronia in southern Finland Iiris Mattila

Blossoming apple tree Iiris Mattila

More info

The Center for Agroforestry at the University of Missouri, established in 1998, is one the world's leading centers contributing to the science underlying agroforestry http://www.centerforagroforestry.org/pubs/training/chap3_2015.pdf

The USDA elaborates very practical and friendly factsheets on agroforestry https://www.fs.usda.gov/nac/documents/agroforestrynotes/an12ac01.pdf

Dupraz, C., Blitz-Frayret, C., Lecomte, I., Molto, Q., Reyes, F., Gosme, M. 2018. Influence of latitude on the light availability for intercrops in an agroforestry alley-cropping system. Agroforest Syst 1–15. https://doi.org/10.1007/s10457-018- 0214-x

Koivula, K. 2012. Peltometsäviljely mahdollisuutena tulevaisuuden Suomessa. Oulun seudun ammattikorkeakoulu. Available online: http://www.theseus.fi/handle/10024/53324

Related links

Document in French

Document in Dutch

Document in Romanian

Figure 75. View of a factsheet

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

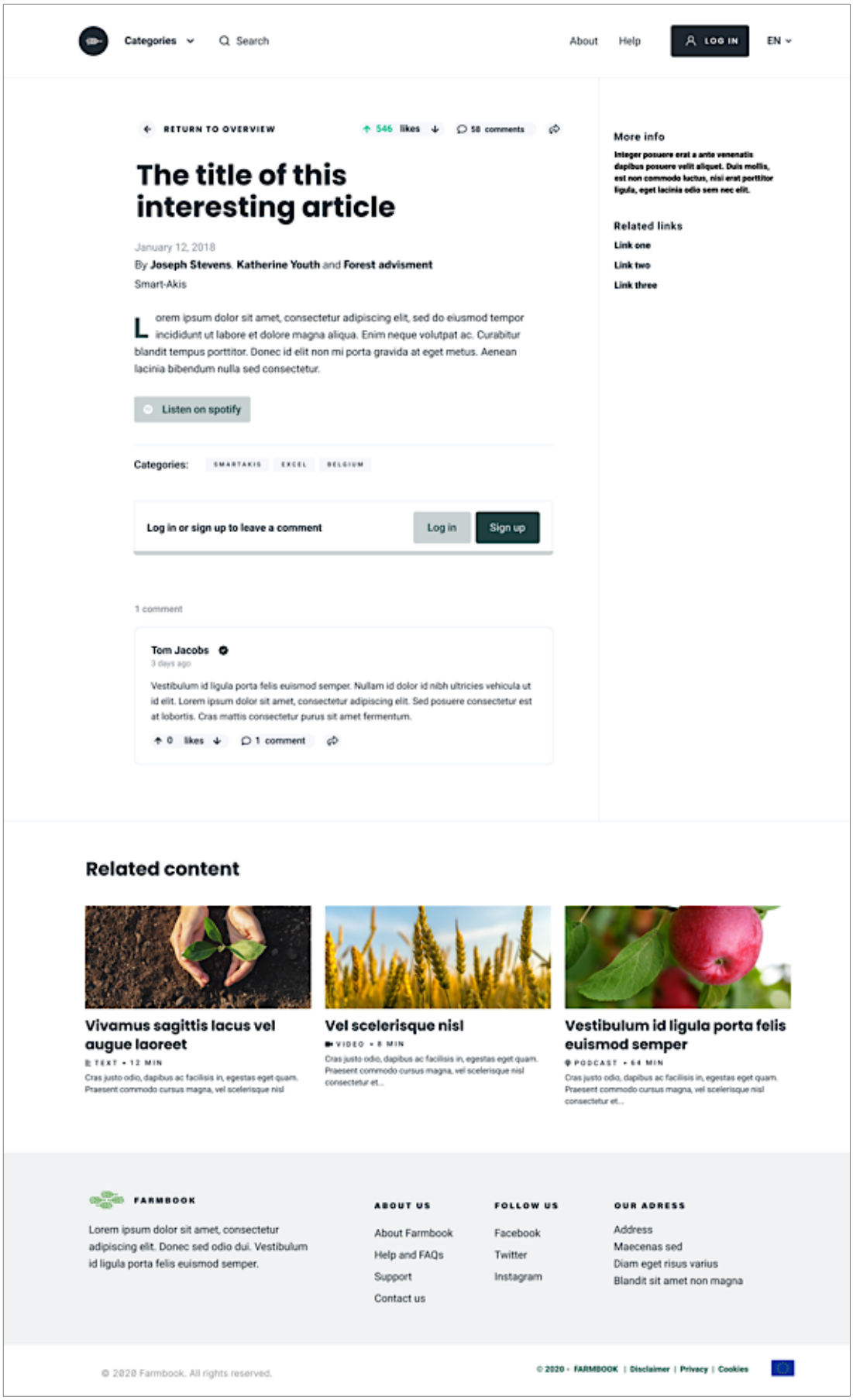


Figure 76. View of a podcast

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

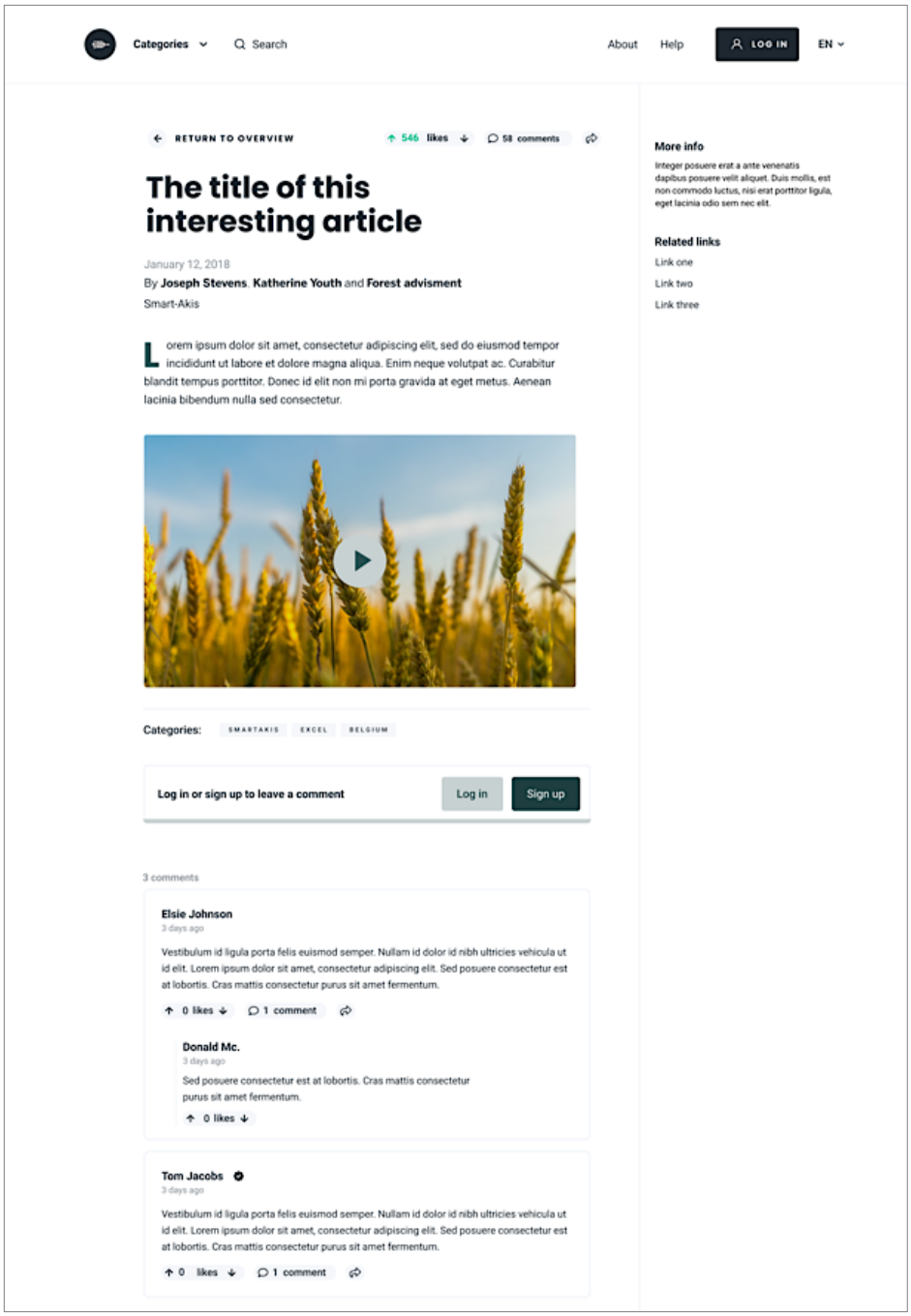


Figure 77. View of a video

11.3. Display of search results

When it comes to searching for information and content on the web, the toll that the users have been accustomed to using is that of a search bar where a search query (consisting of any number of search terms) can be typed in and submitted. This is one of the basic means to access content and information via the e-KRP as well. Figures 78 and 79 below illustrate the two different modes for the delivery of the results of a search operation. In Figure 78, the mode of results' display is that of a list, whereas in Figure 79 the results of the search are displayed in a tile mode. In both cases, some basic pieces of information are used to accompany each result. These pieces of information are: (i) the item's/result's title; (ii) the name of the agricultural sector or theme that the item/result relates to; (iii) the category/type of the data object (text/document, presentation, video, podcast, or infographic); (iv) an estimate of the time

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

needed to go through the data object's content; (v) a short description; and (vi) a figure indicative of the item's/data object's content.

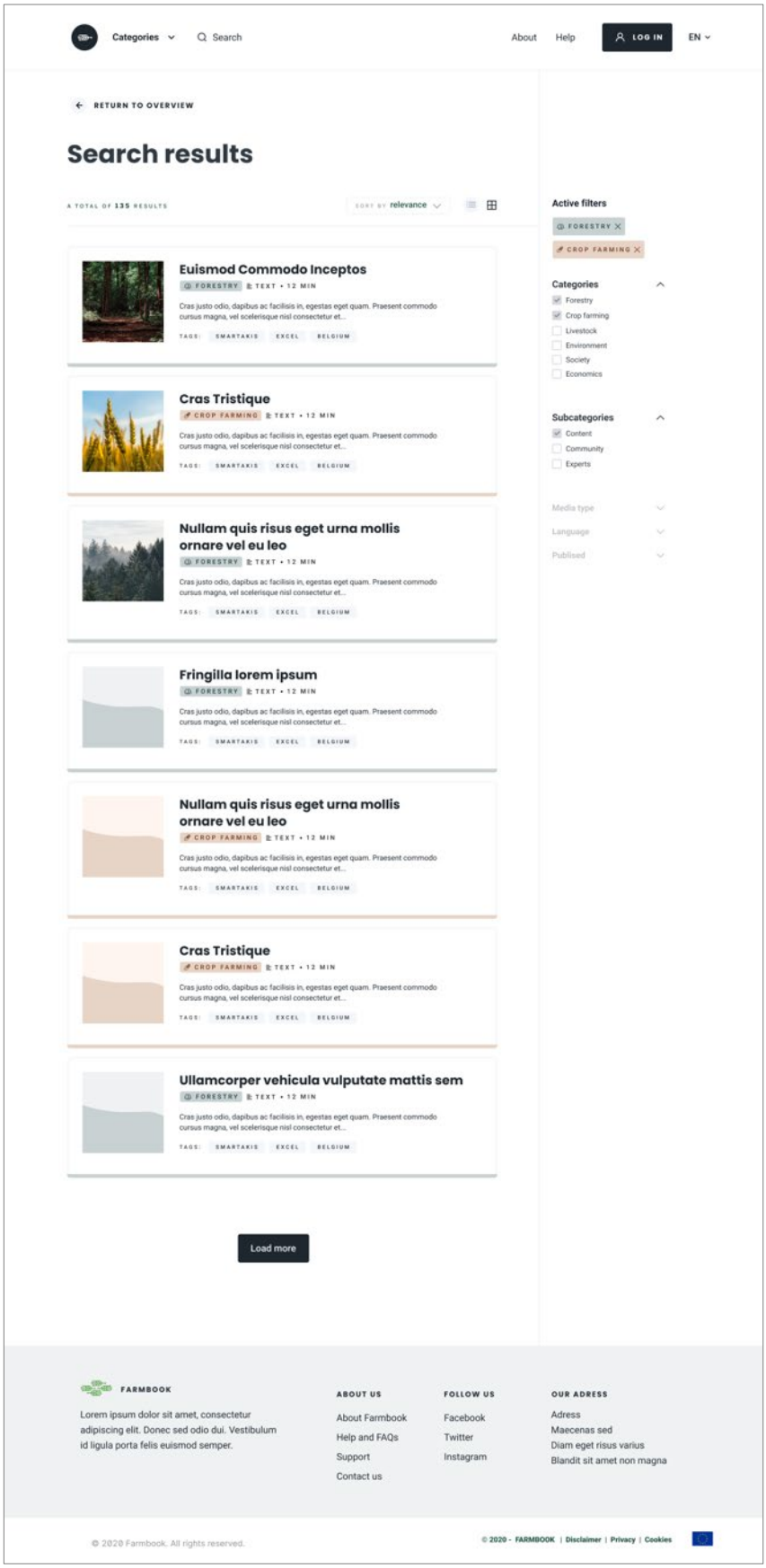


Figure 78. Display of search results as a list

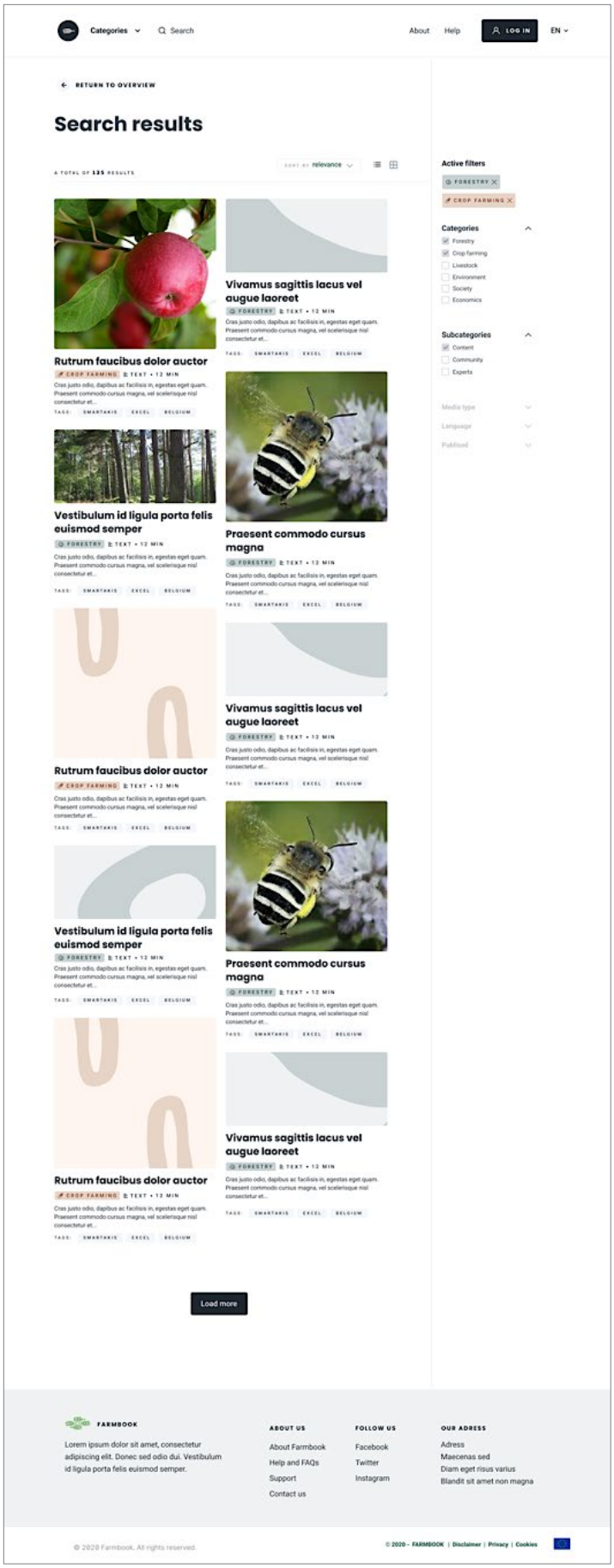


Figure 79. Display of search results by making use of tiles

Given the fact that in most search operation cases the number of the returned results is large, several filters are provided to help narrow down the range of the result items. These filters are shown at the top right corner in each of the two figures above (Figures 78 and 79). The filters used in the e-KRP allow to limit the search results by allowing options related to: (i) the sector and/or the theme that the search results should relate to; (ii) the category/type of the data object (text/document, video, presentation, podcast, and infographic); (iii) the language(s) in which content is delivered; (iv) the content delivery format; (v) the publication year; and (vi) the purpose of the content/information. In the case that TN-related information needs to be accessed, this can be provided through a page particularly dedicated to this purpose (see Figure 80 below).

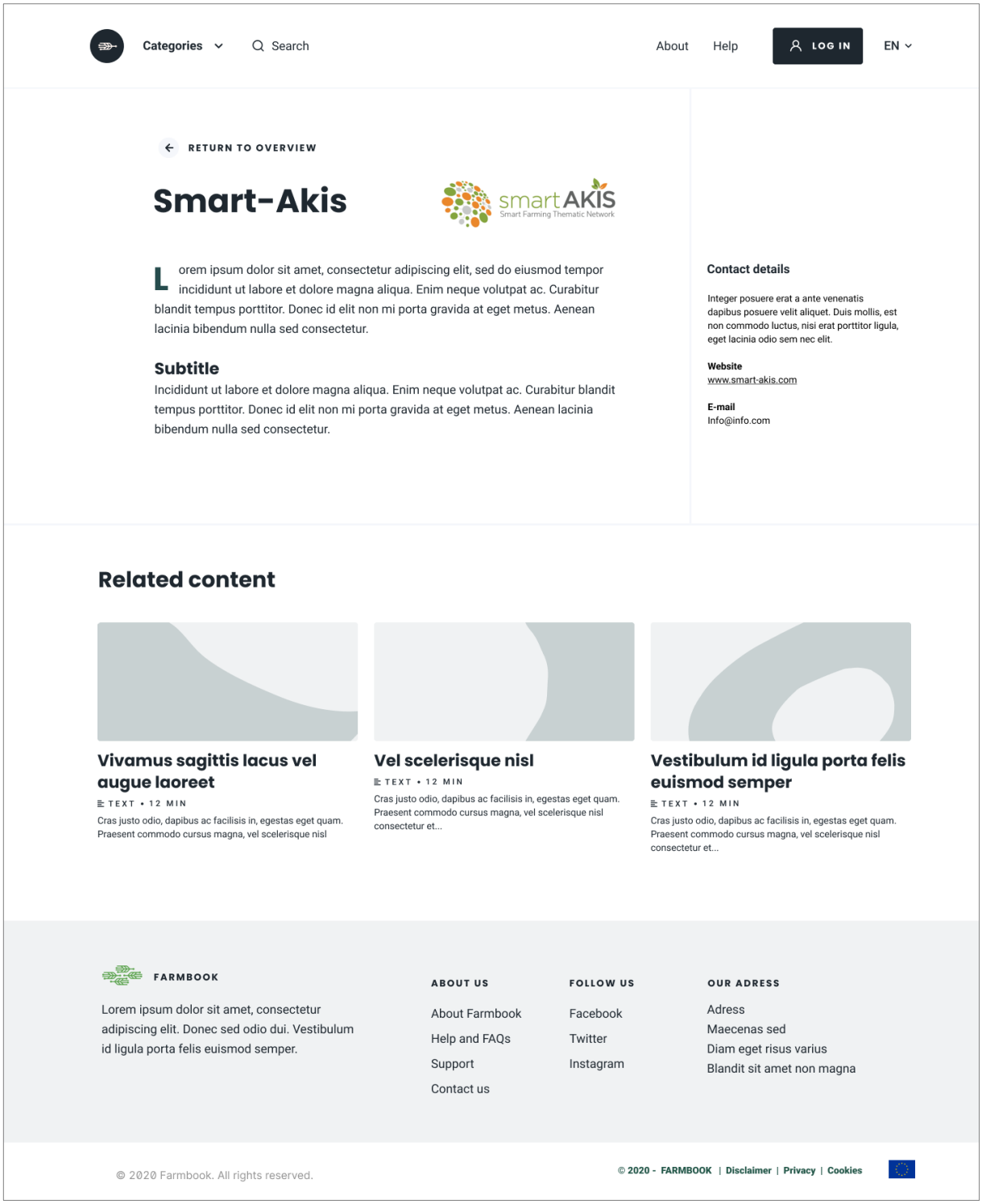


Figure 80. e-KRP page displaying information about a TN

12. Conclusions and next steps

The present report has provided all the details of the work done in Task 4.2 with regard to developing the EURAKNOS prototype e-KRP. The development process has involved diverse activities, requiring a broad range of expertise which has been provided by the partners involved in the task. The descriptions of the undertaken activities have revealed the tight connections between Task 4.2 and tasks in WP4 or other WPs. Specifically, it has become evident that the development-related work in Task 4.2 has built, among others, on the repository-/platform-related design documented in Deliverable 3.2. Within the context of this discussion, it needs to be also stressed that the (technical) work done in EURAKNOS has already been taken up by the EUREKA project (i.e., the follow-up, sibling project of EURAKNOS). This notion of continuity has been highlighted in many sections of the present report. This provides a strong indication of the interconnectedness of the two projects. It also makes clear that there are issues that remain open. These issues need to be further addressed and evolvements are still to be made. This is something that should be expected, given the fact that EURAKNOS has targeted the development of a prototype platform (as a proof-of-concept for the technical feasibility of this endeavour), whereas the EUREKA project aspires to evolve the prototype e-KRP to an MVP. Based on the above, the (technical) issues that remain open and need to be further addressed in EUREKA are the following:

- *Elaborate on the technical design of the e-KRP*
  EURAKNOS has followed a systematic approach for the identification and formal description of the entities (together with their relations and properties) relating to the data objects disseminated by TNs. This process has resulted in the development of the EURAKNOS data model, which has served as the basis for the definition of our metadata set. The model has undergone an update (leading to version 2.0 of the model) to better adhere to the FAIR Data principles. More specifically, this update has aimed to better align the properties described in our model with properties defined in external, widely-adopted ontologies. However, this effort of FAIRification can be viewed as a task in progress. The intention is to continue in EUREKA the investigation for ontologies and controlled vocabularies to consider for the definition of entities, concepts, and properties in order to come up with a robust (in terms of coverage of our needs for describing the data objects), flexible (able to be extended in the future), and elegant (as regards the trade-offs between size and descriptiveness) model. By the time this report was drafted, the technical report on the EUREKA FarmBook design (in the context of which the above issues have been addressed) had already been developed.

- *Updates in the EURAKNOS metadata set*
  As mentioned earlier, the set of metadata proposed in EURAKNOS (namely, the minimum metadata set to be used for an efficient annotation/description of the TNs' data objects) has been developed based on the EURAKNOS data model. This means that any changes/adaptations (to be) made in the context of EUREKA will be also be reflected on the metadata set. It needs to be stressed that both the EURAKNOS data model and the associated metadata set have not been created with a narrow focus in mind (by taking consideration only of the small set of our TN sample), but rather to capture (in the best possible way) all the features and particularities of TNs and MAPs in general (given that MAPs constitute a superset of TNs). Hence, we envisage that the metadata proposed in EURAKNOS will be sufficient (both in number and coverage aspects) for describing the digital outputs (i.e., the data/knowledge objects[108]) of MAPs as well. Any changes/updates to be made to the metadata set will (probably) have to do with the metadata property definitions given the ongoing process (in the context of EUREKA) of identifying and considering external ontologies and/or vocabularies for the needs of describing our model's elements (i.e., entities, concepts, relations, and entity properties).

- *Adaptations in the ETL process*
  As explicitly mentioned in Subsection 6.4 ("Integration of data objects into the e-KRP repository"), the ETL (i.e., Extract-Transform-Load) process is a service that has been developed and deployed to

[108] The term that is used in EUREKA to refer to the digital outputs available through the KRs of MAPs is "knowledge object".

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

ingest the data objects harvested by the EURAKNOS crawler (from the sample TNs' KRs) into the e-KRP repository (DSpace). However, the ETL process has been tested only in regard to the ingestion of data object belonging to the data object categories available from the TN KRs (i.e., documents, presentations, videos, podcasts, and infographics). Given the fact that MAPs have a broader scope than TNs and the greater breadth of the data/knowledge objects created by them, adaptations may need to be made in the ETL process to encompass further data/knowledge object categories that have not been considered in EURAKNOS (e.g., datasets and software applications).

- *Changes in the repository system's backend*
  In the section of the report describing the e-KRP system's architecture (Section 4), it is mentioned, among others, that a format in which metadata is stored is that of RDF triples. The database that is used for this purpose is Fuseki (the triple store included in the framework of DSpace). During the e-KRP development process, there have been discussions about circumventing Fuseki and making use of Virtuoso[109], which is adopted in many European projects. However, this process was not taken further in EURAKNOS because of the technical issues involved in it and mentioned in Section 4. This issue is intended to be further investigated in EUREKA[110].

- *Changes/adaptations in functionalities and features*
  Details about the feedback that has been collected from the testing and evaluation of the e-KRP are provided in Section 8 ("Testing and evaluation of the EURAKNOS e-KRP"). The session that has been held for the testing and evaluation of the e-KRP has led to the collection of feedback that constitutes valuable input for UI- and functionality-related improvements in EUREKA (given the fact that the e-KRP continues to be developed in EUREKA as part of advancing from the PoC to an MVP). Feedback relates mostly to community-related aspects and functionalities engineered into the platform, and the aesthetics part of the design as well[111]. In EUREKA, there has been a consideration for thorough user tests together with a WP particularly dedicated to the collection of useful information about the profiles or "archetypes" of specific end-user categories (WP2). The information intended to be collected from these tasks/activities will serve as valuable feedback to the functionalities already engineered, or considered at a conceptual level, in the e-KRP (the EUREKA's FarmBook platform is the follow-up of the e-KRP). This feedback will prove to be useful for the purpose of revisiting the platform functionalities and features already considered and provided, and probably come up with new ones. The EURAKNOS Vision Paper can also serve as an important source of input for drawing decisions about the functionalities to be further considered and developed in EUREKA. For instance, the availability of a SPARQL endpoint[112] (incorporated into the system's API layer), aimed to be used by researchers/scientists, is a feature that could potentially be considered. In the same line, issues of personalisation and supply of recommendations in regard to content and information tailored to users' needs, will be further investigated. Finally, in the context of building upon the functionalities and features already developed, it is also proposed to take stock of the work done in EURAKNOS in respext to the creation of a user guide for the e-KRP. Given that the development of the platform continues in EUREKA, an elaborate version of the user guide can become available as a standalone artefact. The user guide can be made available through the platform itself in various formats (e.g., as a booklet, presentation, or video). By providing instructions in a more user-friendly language, the idea is to offer an appealing manual able to be utilised as a reference point for helping users become familiar with the system.

[109] https://virtuoso.openlinksw.com/

[110] DSpace is a widely used, by many organisations and institutions worldwide, data repository solution. Nevertheless, it has not been considered and used in the context of European projects (at least, to the best of our knowledge). Therefore, in order to also embrace technical solutions favoured by other European projects as well, there needs to be some adjacency-related work which may have difficulties because of the technicalities behind making two systems "talk" to each other.

[111] Details are provided in Section 8.1 ("Results obtained from the user test session").

[112] For details about what a SPARQL endpoint is and its usefulness, the reader may refer to the "What is a SPARQL Endpoint, and why is it important?" Medium article available at https://medium.com/virtuoso-blog/what-is-a-sparql-endpoint-and-why-is-it-important-b3c9e6a20a8b.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- *Further investigation of sustainability issues*
  The issue of ensuring the sustainability of the e-KRP is of increased importance given the platform's intended role within the AKIS community; i.e., a hub of practical, agriculture-related knowledge. In EURAKNOS, a substantial effort has been undertaken for appropriately framing the sustainability of the e-KRP with several dimensions being associated with it. Each dimension has been appropriately analysed and explained and its impact on the overall platform sustainability has been estimated. In addition, some concrete solutions have been provided as part of an initial attempt to come up with an e-KRP sustainability plan. The pillars of this plan are: (i) the adoption of a "crowdfunding" model for securing the revenue streams needed to keep the technological infrastructure running and the associated service being provided; (ii) the continuity in service provision as part of a scenario related to the in-house hosting of the e-KRP by a consortium/third-party organisation; and (iii) the further dissemination of digital content/information under appropriate licensing schemes. In this context, the description of a framework of guidelines and recommendations related to the protection of the TNs' intellectual property is of high relevance. However, there is more to investigate. Consequently, the sustainability-related research and work has been already taken up by EUREKA. An issue to be further considered in EUREKA is the identification and use of a broader range of licensing schemes/ types given the breadth of the data/knowledge objects considered in EUREKA. For instance, some broadly-known licensing frameworks to take into account for the further dissemination of software-related work are Apache[113] and GNU[114].

13. 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.

Brown, **D.M.** (2010). *Communicating design: developing web site documentation for design and planning*. New Riders.

Dunning, **A**., **De Smaele**, **M.** & **Böhmer**, **J.** (2017). Are the fair data principles fair? *International Journal of digital curation*, *12*(2), pp. 177-195.

European Commission (2017). Intellectual Property Rights: Learn about IP Rights in Horizon 2020. Retrieved from https://cdn3.euraxess.org/sites/default/files/domains/tr/th2020_ipr_brochure_eng_apr17.pdf.

Gernert, **M.**, **et al.** (2019). Report on added value and feasibility of e-KRP. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 4.1 – WP4.

Marois, **M.**, **et al.** (2019). Report on format and design of knowledge reservoirs of Thematic Networks. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 2.2 – WP2.

Marois, **M.**, **et al.** (2020). Report on HIKR methodologies and practices for data generation, selection, collection and storage. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 3.1 – WP1.

Mosquera Losada, **M. R.**, **et al.** (2019). Report on content and structures of data provision and knowledge formats of TNs. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 2.1 – WP2.

[113] https://spdx.org/licenses/Apache-2.0.html

[114] https://www.gnu.org/licenses/gpl-3.0.html

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Mouseti, **S. et al.** (2020). API documentation. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 4.2 – WP4.

Patil, **Y.**, & **Patil**, **S.** (2016). Review of Web Crawlers with Specification and Working. *International Journal of Advanced Research Computer and Communication Engineering*.

Van Hal, **D. et al.** (2020). Report on HIKR Design. EURAKNOS – Connecting Thematic Networks as Knowledge Reservoirs towards a European Agricultural Knowledge Innovation Open Source System Deliverable 3.2 – WP3.

Wilkinson, **M. D.**, *et al*. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*.

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

14. Annex I: Functionalities of the EURAKNOS web crawler

Annex I aims to provide further details about the EURAKNOS web crawler's core functionalities. These functionalities are listed and briefly analysed below:

- **Functionality #1**: Create and start an individual web crawler
  - Using this function allows to create, configure and start one's own web crawler.
  - After selecting the "Add crawler" option available from the crawler's menu, one must set the name and the starting URL of the web crawler. Optionally, a number of parameter values can be configured, or extra parameters may be added to the web crawler. These parameters are presented in Table 30 below.

Table 30. Parameters and their values to be defined when a new, individual web crawler is being created

Parameters	Description	Mandatory	Default value
Starting URL	The starting URL of the web crawler.	Y	Empty
Base URL	The base URL defines the constant prefix of content's URLs. In some special cases, the starting and the base URL differ.	N	Empty
Crawler's name	The name of the web crawler.	Y	Empty
External content download	If set, the crawler will download content linked to the starting site but not hosted in it.	N	On
META Title	Harvesting meta titles from the website.	N	Off
META Keywords	Harvesting meta keywords from the website.	N	Off
META Description	Harvesting meta descriptions from the website.	N	Off
Heading 1	Harvesting level 1 heading (H1) tags from the website.	N	Off
Heading 2	Harvesting level 2 heading (H2) tags from the website.	N	Off
Heading 3	Harvesting level 3 heading (H3) tags from the website.	N	Off
Other searching elements	Add some extra searching options to refine the crawling process. The first field contains the DIV class name and the second one is its reference name in the database.	N	Empty, Empty

- **Functionality #2**: Details and Export details
  - By using this feature, it is possible to view the results obtained from the execution of an individual web crawler.
  - The displayed chart shows (Figure 81 below) the type, proportion, and number of downloaded files, as well as the percentage of the metadata extracted.

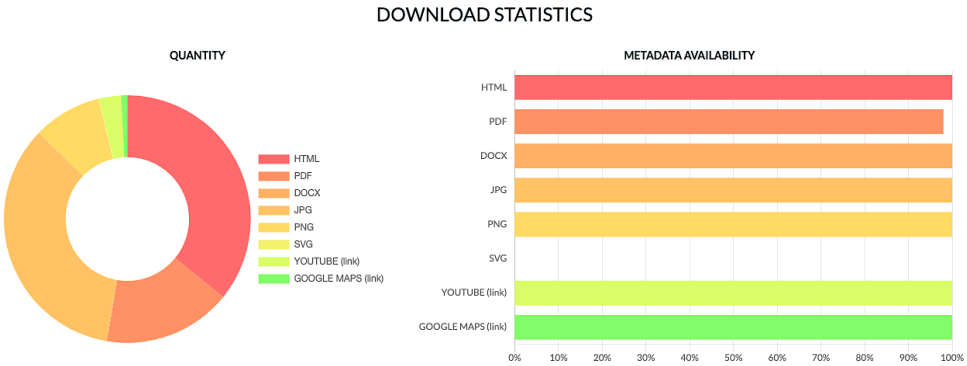


Figure 81. The information provided to the user by making use of the "Details" functionality

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

- By clicking on the “Export details” button, the statistics related to the current web crawler can be exported in the form of a spreadsheet.

- **Functionality #3**: Export all download statistics
  - By using this feature, one can export the download statistics of all the individual web crawlers, as a spreadsheet, either the crawling process is still running or has been completed.
  - The statistics export process will start by clicking on the “Export all” button.

- **Functionality #4**: Export all metadata
  - By using this feature, one can automatically export all the metadata extracted by all the web crawlers.
  - The extracted metadata will become available as a spreadsheet either the crawling process is still running or has been completed.
  - The export process starts by clicking on the “Export all meta” button.

15. Annex II: Example Python script for classifying documents into the appropriate document type

In this Annex section, the Python script for the classification of the FERTINNOWA TN’s documents is provided.

```
# coding: utf-8
# In[1]:
import pandas as pd
import re
from langdetect import detect
from gensim.summarization import keywords

# <p>We begin by loading the <code>csv</code> file containg all the <b>pdf documents</b>
harvested from the <b>Fertinnowa Thematic Network's repository</b> and <b>loading it</b> as a
<code>dataframe</code>.</p>

# In[2]:
FertinnowaPdfsDf = pd.read_csv("pdfs_per_TN/Fertinnowa_pdfs.csv", sep = ";", header = 0,
index_col = False)

# <p>The <code>dataframe</code> of the <b>Fertinnowa pdf outputs</b> is <b>loaded</b> and
<b>displayed</b>.</p>

# In[3]:
FertinnowaPdfsDf

# In[4]:
print("The number of records in the initial dataframe version is:\033[1m",
FertinnowaPdfsDf.shape[0], "\033[0m")

# <p><b>STEP 1</b>: We begin by <b>cleaning our pdf outputs dataset<b> from <b>all duplicate
records</b>.</p>

# In[5]:
FertinnowaPdfsDf.drop_duplicates(subset = "WCFILEPATH", keep = "first", inplace = True)

# In[6]:
FertinnowaPdfsDf

# In[7]:
```

The EURAKNOS project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No: 817863

```
print("The number of records in the clean dataframe version is:\033[1m",
FertinnowaPdfsDf.shape[0], "\033[0m")

# <p><b>STEP 2</b>: A <code>list</code> containg the names of the <b>most frequently appearing
document-related outputs</b> is being created. The intention is to <b>go through the entire set of
the pdf documents harvested</b> from the <b>Fertinnowa Thematic Network's repository</b> and
<b>find</b> any potential <b>matches</b>.</p>

# In[8]:
mostFrequentDocumentOutputs = ["practice", "factsheet", "press", "newsletter", "tech", "guide",
"article", "deliverable",
                    "report", "handbook", "policy", "brief", "journal", "conference", "manual", "review",
                    "book", "booklet", "paper", "tutorial", "proceedings", "information", "milestone",
                     "broch", "flyer", "research"]


# <p>We <b>iterate over the Fertinnowa pdf outputs dataframe</b> (and more specifically the
column named "<code><b>WCFILEPATH</code></b>") in order to <b>find those file urls</b> that
<b>contain the name of any of the most frequently appearing document-related outputs</b>
available in the <b>mostFrequentDocumentOutputs</b> <code>list</code>.</p>

# In[9]:
counter = 0
mostFrequentDocumentOutputfileUrlList = []

for fileUrl in FertinnowaPdfsDf["WCFILEPATH"]:

   for documentOutput in mostFrequentDocumentOutputs:

      if re.search(documentOutput, fileUrl, re.IGNORECASE):

         counter +=1

         print(fileUrl)

         mostFrequentDocumentOutputfileUrlList.append(fileUrl)

# <p>The <b>complete <code>dataframe</code> records</b> containing the above file URLs are
displayed.</p>

# In[10]:
filter = FertinnowaPdfsDf["WCFILEPATH"].isin(mostFrequentDocumentOutputfileUrlList)
FertinnowaPdfsDf[filter]

# <p>The <b>number</b> and <b>rate of matches</b> is being calculated. We define <b>match</b>
as the '<i>number of file URLs (and, thus, dataframe records) containing any of the terms in the
mostFrequentDocumentOutputs list</i>'. It needs to be kept in mind that <b>there may be duplicate
records</b>.</p>

# In[11]:
numOfMatches = FertinnowaPdfsDf[filter].shape[0]
print("Total number of matches:", numOfMatches)
print("The rate of matches is:", round((numOfMatches/FertinnowaPdfsDf.shape[0])*100,2), "%")

# <p><b>STEP 3</b>: We <b>split the pdf documents</b> that have been harvested by the crawler
(and match with the items of the <b>mostFrequentDocumentOutputs</b> <code>list</code>) into
lists by <b>taking into account their type</b> (the potential pdf output types are labelled by the
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

items of the mostFrequentDocumentOutputs <code>list</code>). In order to do so, we begin by creating a dictionary where:

 the key is the pdf output type; and

 the value is the file URL of the pdf output of the specific type.

```
# In[12]:
pdfOutputTypesDict = {}

for documentOutput in mostFrequentDocumentOutputs:

    pdfOutputTypesDict[documentOutput] = []

print(pdfOutputTypesDict)

# In[13]:
for key, value in pdfOutputTypesDict.items():

    for fileUrl in FertinnowaPdfsDf[filter]["WCFILEPATH"]:

        match = re.search(key, fileUrl, re.IGNORECASE)

        if match:

            pdfOutputTypesDict[key].append(fileUrl)

# In[14]:
total = 0

for key in pdfOutputTypesDict.keys():

    print(key,"=", pdfOutputTypesDict[key])

    print(key, "number =",len(pdfOutputTypesDict[key]))

    total += len(pdfOutputTypesDict[key])

    print()

print(total)
```

Given the fact that the words "book" and "booklet" have the same stem, we find instances of the "booklet" pdf type in the "book" list. Therefore, we need to remove them.

```
# In[15]:
bookletsToBeRemoved = []

for record in pdfOutputTypesDict["book"]:

    match = re.search("booklet", record, re.IGNORECASE)

    if match:

        bookletsToBeRemoved.append(record)

# In[16]:
bookletsToBeRemoved

# In[17]:
for item in bookletsToBeRemoved:

    pdfOutputTypesDict["book"].remove(item)
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# <p>Given the fact that both the words "<b>book</b>" and "<b>handbook</b>" contain the <b>term</b> "<b>book</b>", we find instances of the "<b>handbook</b>" <b>pdf type</b> in the "<b>book</b>" <b>list</b>. Therefore, we need to <b>remove</b> them.</p>

# In[18]:
handbooksToBeRemoved = []

for record in pdfOutputTypesDict["book"]:

    match = re.search("handbook", record, re.IGNORECASE)

    if match:

        handbooksToBeRemoved.append(record)

# In[19]:
handbooksToBeRemoved

# In[20]:
for item in handbooksToBeRemoved:

    pdfOutputTypesDict["book"].remove(item)

# In[21]:
pdfOutputTypesDict["book"]

# <p><b>STEP 4</b>: We <b>export each dataframe of pdf output type</b> as a <code>csv</code> file. For instance, all the records related to <b>Fertinnowa deliverables</b> harvested by the crawler are contained (and stored) in the <b>Fertinnowadeliverable_pdfs.csv</b> file.</p>

# In[22]:
def exportAsCsv (ThematicNetworkName, pdfOutputType, TNpdfsDf):

    filename = "Fertinnowa_pdf_outputs/" + ThematicNetworkName + pdfOutputType + "_pdfs.csv"

    TNpdfsDf.to_csv(filename, sep = ";", header = True, encoding = "utf-8")

# In[23]:
for key in pdfOutputTypesDict.keys():

    if len(pdfOutputTypesDict[key]) != 0:

        df =
FertinnowaPdfsDf[filter][FertinnowaPdfsDf[filter]["WCFILEPATH"].isin(pdfOutputTypesDict[key])]

        exportAsCsv("Fertinnowa", key, df)

# <p><b>STEP 5</b>: We <b>extract</b> all the <b>pdf documents</b>, harvested by the crawler, that <b>do not match</b> with the items of the <b>mostFrequentDocumentOutputs</b> <code>list</code> as a <b>separate <code>dataframe</code></b> (i.e. the <b>otherPdfTypesDf</b> <code>dataframe</code>).</p>

# In[24]:
otherPdfTypesDf =
FertinnowaPdfsDf[~FertinnowaPdfsDf["WCFILEPATH"].isin(mostFrequentDocumentOutputfileUrlList)
]

# In[25]:
otherPdfTypesDf

# <p><b>STEP 6</b>: We <b>retrieve a number of</b> "<b>other</b>" <b>pdf documents</b> based on the "<b>importance</b>" <b>of the words</b>/<b>terms</b> appearing in the respective
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

file URL. To do so, we cut the last part of the file URL (indicative of the pdf file's name) off the file URL. We consider this string as the pdf file's name.

```
# In[26]:
length = otherPdfTypesDf.shape[0]
counter = 0
text = ""

for rowNumber in range(length):

    languageShortName = detect(otherPdfTypesDf["WCFILEPATH"].iloc[rowNumber].split("/")[-1])

    if languageShortName == "en":

        print(otherPdfTypesDf["WCFILEPATH"].iloc[rowNumber].split("/")[-1])

        text = text + " " + otherPdfTypesDf["WCFILEPATH"].iloc[rowNumber].split("/")[-1].split(".")[0]

        counter +=1

print()

print("The total number of pdf files not associated with the most frequent document outputs is:",
counter)
```

We merge all the pdf file names (after having removed the ".<code>pdf</code>" file name extension) into a text excerpt. We then substitute all the "-" and "_" characters found in the text with a blank space.

```
# In[27]:
print(text)
```

```
# In[28]:
text = text.replace("-", " ")
text = text.replace("_", " ")
print(text)
```

We extract keywords (together with their scores) from the text produced by merging all the "other pdf documents" file names. The keyword extraction is executed by drawing upon the <code>gensim</code> library.

```
# In[29]:
for keywordInfo in keywords(text, words = 10, scores = True, lemmatize = True):

    print(keywordInfo)
```

We keep the keywords/terms with a score higher than, or equal to 19%. The pdf outputs whose file names contain these keywords will also be considered as input to the EURAKNOS repository.

```
# In[30]:
otherDocumentOutputs = []

for keywordInfo in keywords(text, words = 10, scores = True, lemmatize = True):

    if keywordInfo[1] >= 0.19:

        otherDocumentOutputs.append(keywordInfo[0])
```

```
# In[31]:
otherDocumentOutputs
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# <p><b>STEP 7</b>: Some <b>further pdf documents</b> will be considered by taking account of the <b>keywords</b>/<b>terms</b> with a <b>score higher than 19%</b> and their appearance in the pdf file names.</p>

# In[32]:

counter = 0
otherDocumentOutputfileUrlList = []

for fileUrl in FertinnowaPdfsDf["WCFILEPATH"]:

    for documentOutput in otherDocumentOutputs:

        if re.search(documentOutput, fileUrl, re.IGNORECASE):

            counter +=1

            otherDocumentOutputfileUrlList.append(fileUrl)

# In[33]:
filter = otherPdfTypesDf["WCFILEPATH"].isin(otherDocumentOutputfileUrlList)
otherPdfTypesDf[filter]

# <p>These <b>other pdf files</b> are <b>stored into a separate</b> <code>csv</code> <b>file</b> named "<b>Fertinnowa_other_pdfs.csv</b>".</p>

# In[34]:
otherPdfTypesDf[filter].to_csv("Fertinnowa_pdf_outputs/Fertinnowa_other_pdfs.csv", sep = ";", header = True, encoding = "utf-8")

# In[35]:
print("The total number of the Fertinnowa pdf outputs matching the most frequent pdf output types is:\033[1m", numOfMatches, "\033[0m")

# In[36]:
print("The total number of 'other' Fertinnowa pdf outputs is:\033[1m", otherPdfTypesDf[filter].shape[0], "\033[0m")

# In[37]:
print("Number of potential 'duplicates' among the most frequent pdf output types:\033[1m",  total - numOfMatches, "\033[0m")
```

16. Annex III: Python script for the removal of duplicate pdf files

In the present Annex section, the Python script for the removal of duplicates, after the FERTINNOWA TN documents' classification, is provided.

```
# coding: utf-8

# <h3>Check4duplicates_in_Fertinnowa_csv_files python script</h3>

# <p>The aim of this Python script is to <b>identify all the <code>csv</code> files in a directory</b>, i.e. the <code>csv</code> files containing <b>info about the different pdf output types</b> identified per Thematic Network.</p>

# In[1]:
from os import listdir
import glob
import csv
import pandas as pd
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# <p>The code script in the following block <b>reads all the <code>csv</code> files</b> contained in
a <b>specified directory</b>. The <b>directory</b> where the file search takes place is <b>specified
with</b> the help of the <code>.listdir()</code> (documentation related to the method can be found
[here](https://www.tutorialspoint.com/python/os_listdir.htm)).</p>

# <p>The <b>type of the files to be retrieved</b> is specified with the help of the repsective
<b>extension</b> (in our case ".csv"), which passed as a value to the <code>suffix</code> variable.
We then <b>iterate over the list of the files</b> contained in the directory and <b>print the
names</b> of those that are of the ".csv" type.<p>

# In[2]:
path = "Fertinnowa_pdf_outputs/"
suffix = ".csv"
directory = listdir(path)
csvFilesInDirectory = []
for filename in directory:

   if filename.endswith(suffix):

      csvFilesInDirectory.append(path + filename)

# In[3]:
for item in csvFilesInDirectory:
   print(item)

# <p>Calculate the <b>total number of pdf files</b> contained in the csv files available from the
directory. This operation is executed to <b>validate</b> that the <b><code>csv</code> files</b>
available in the directory <b>do contain</b> pdf output - related records.</p>

# In[4]:
csvFilesContentDataframe = []
for file in csvFilesInDirectory:

   df = pd.read_csv(file, delimiter = ";", header = 0, index_col = False)

   csvFilesContentDataframe.append(df)

result = pd.concat(csvFilesContentDataframe, ignore_index=True)

result

# <p>Displaying the <b>total number of the pdf files before removing</b> any duplicate records.</p>

# In[5]:
print("The total number the pdf outputs is:\033[1m", result.shape[0],"\033[0m")

# In[6]:
fileUrlsInCsvFile = []
with open(file) as csvfile:

   readCsv = csv.reader(csvfile, delimiter = ";")

   line_1 = next(readCsv)

   for value in line_1:

      if value == "WCFILEPATH":

         index = line_1.index(value)

   #fileUrlsInCsvFile = []

   for line in csvfile.readlines():
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
        fileUrlsInCsvFile.append(line.split(";")[index])
# In[7]:
def getStringFromDirectoryName(prefix, suffix, stringToConvert):
    convertedString = stringToConvert.replace(prefix, "").replace(suffix, "")
    if "_" in convertedString:
        convertedString = convertedString.replace("_", "")
    return convertedString
# In[8]:
csvFileUrlsDict = {}
pref = "Fertinnowa_pdf_outputs/Fertinnowa"
suff = "_pdfs.csv"
for item in csvFilesInDirectory:
    key = getStringFromDirectoryName(pref, suff, item)
    csvFileUrlsDict[key] = []
# In[9]:
print(csvFileUrlsDict)
# In[10]:
for file in csvFilesInDirectory:
    pref = "Fertinnowa_pdf_outputs/Fertinnowa"
    suff = "_pdfs.csv"
    fileName = getStringFromDirectoryName(pref, suff, file)
    for key in csvFileUrlsDict.keys():
        if key == fileName:
            with open(file) as csvfile:
                readCsv = csv.reader(csvfile, delimiter = ";")
                next(readCsv)
                for line in csvfile.readlines():
                    fileUrl = line.split(";")[index]
                    csvFileUrlsDict[key].append(fileUrl)
# In[11]:
for item in csvFileUrlsDict.items():
    print(item,"\n")
# <p><b>Total number of pdf files</b> in the dictionary <b>before removing</b> the duplicate
records.</p>
# In[12]:
counter = 0
for key in csvFileUrlsDict.keys():
    for value in csvFileUrlsDict[key]:
        counter +=1
# In[13]:
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
print(counter)
# In[14]:
for name in csvFileUrlsDict.keys():
    matchingUrls = []
    for fileUrlInCsvFile in csvFileUrlsDict[name]:
        print("\nChecking whether there is any match with\033[1m", fileUrlInCsvFile.split("/")[-
1],"\033[0m...")
        keys = list(csvFileUrlsDict.keys())
        keys.remove(name)
        for key in keys:
            print("\n",key)
            flag = False
            for url in csvFileUrlsDict[key]:
                print(url)
                if url == fileUrlInCsvFile:
                    flag = True
                    match = url
                    matchingUrls.append(match)
            if flag:
                print("\033[1mA match has been found! \033[0m")
                print("The matching URL is:", match)
            else:
                print("\033[1mNo match has been found! \033[0m")
    print()
    print(matchingUrls)
    for i in matchingUrls:
        for j in csvFileUrlsDict[name]:
            if i == j:
                csvFileUrlsDict[name].remove(j)
# In[15]:
for item in csvFileUrlsDict.items():
    print(item,"\n")
# <p><b>Total number of pdf files</b> in the dictionary <b>after having removed</b> the duplicate
records.</p>
# In[16]:
counter = 0
for key in csvFileUrlsDict.keys():
    for value in csvFileUrlsDict[key]:
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
        counter +=1
# In[17]:
print(counter)
# In[18]:
for key in csvFileUrlsDict.keys():
    for file in csvFilesInDirectory:
        pref = "Fertinnowa_pdf_outputs/Fertinnowa"
        suff = "_pdfs.csv"
        fileName = getStringFromDirectoryName(pref, suff, file)
        if key == fileName:
            df1 = pd.read_csv(file, sep = ";", header = 0, index_col = False)
            filter = df1["WCFILEPATH"].isin(csvFileUrlsDict[key])
            path = "duplicate_free_pdf_outputs/duplicate_free_Fertinnowa_pdf_outputs/"
            fileNamePrefix = "duplicate_free_"
            fName = file.replace("Fertinnowa_pdf_outputs/", "")
            finalName = path + fileNamePrefix + fName
            df1[filter].to_csv(finalName, sep = ";", header = True, encoding = "utf-8")
# <h4>Final count of the Thematic Network's pdf documents available per type.</h4>
# In[26]:
for key in csvFileUrlsDict.keys():
    if key != "other":
        print("The total number of\033[1m", key, "\033[0mpdf documents is:\033[1m",
len(csvFileUrlsDict[key]),"\033[0m")
print("\nThe total number of\033[1m other \033[0mpdf documents is:\033[1m",
len(csvFileUrlsDict["other"]),"\033[0m")
```

17. Annex IV: Python scripts for the selection of YouTube videos, presentations in a pdf file format, and infographics

Selection of YouTube videos

```
# coding: utf-8
# <h3>Cleaning the "youTubeVideosToSelectFrom" csv file from duplicate records.</h3>
# In[1]:
import pandas as pd
# <p><b>Loading the</b> "<code>youTubeVideosToSelectFrom</code>" <b>csv file</b> and
<b>reading it</b> as a <code>pandas dataframe</code>.</p>
# In[2]:
youTubeVideosDf = pd.read_csv("youTubeVideosToSelectFrom.csv", sep = ";", header = 0, index_col
= False, encoding = "ISO-8859-1")
# <p><b>Displaying</b> the <b>first five records</b> of the <code>dataframe</code>.</p>
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# In[3]:
youTubeVideosDf.head()

# <p><b>Total number of records</b> in the "<code>youTubeVideosDf</code>"
<code>dataframe</code>.</p>

# In[4]:
youTubeVideosDf.shape[0]

# <p><b>Total number of UNIQUE YouTube video URLs</b> in the
"<code>youTubeVideosDf</code>" <code>dataframe</code>.</p>

# In[5]:
len(youTubeVideosDf["WCFILEPATH"].unique())

# <p><b>Creating a list</b> of all the <b>unique YouTube video URLs</b>.</p>

# In[6]:
listOfIndexesOfUniqueYouTubeVideoUrls = []
for uniqueVideoUrl in youTubeVideosDf["WCFILEPATH"].unique():

   indexOfFirstVideoUrlOccurence = youTubeVideosDf.index[youTubeVideosDf["WCFILEPATH"] ==
uniqueVideoUrl][0]

   listOfIndexesOfUniqueYouTubeVideoUrls.append(indexOfFirstVideoUrlOccurence)

# <p><b>Displaying</b> the <b>total number</b> of the <b>UNIQUE video URLs</b> returned!</p>

# In[7]:
len(listOfIndexesOfUniqueYouTubeVideoUrls)

# <p><b>Discarding</b> the <b>rows</b> of the "<code>youTubeVideosDf</code>"
<code>dataframe</code> that <b>contain duplicate video URLs</b>.</p>

# In[8]:
rowNumCount = -1
for videoUrl in youTubeVideosDf["WCFILEPATH"]:

   rowNumCount += 1

   if rowNumCount not in listOfIndexesOfUniqueYouTubeVideoUrls:

      youTubeVideosDf.drop([rowNumCount], inplace = True)

# In[9]:
youTubeVideosDf.shape[0]

# <p>The <b>list of</b> the "<code>youTubeVideosDf</code>" <code>dataframe</code>
<b>column titles</b> that we are <b>interested in keeping</b>.</p>

# In[10]:
columnTitles = ["TITLE", "DESCRIPTION", "WCFILEPATH", "SITENAME", "PUBLISHEDAT",
"DEFAULTAUDIOLANGUAGE", "views"]

# In[11]:
duplicateFreeYouTubeVideosDf = youTubeVideosDf.copy()

# <p><b>Deleting</b> the <b>dataframe columns</b> that we <b>do not want to keep</b>!</p>

# In[12]:
for columnTitle in duplicateFreeYouTubeVideosDf:
   if columnTitle not in columnTitles:

      del duplicateFreeYouTubeVideosDf[columnTitle]
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# <p><b>Changing</b> the "<code>duplicateFreeYouTubeVideosDf</code>"
<code>dataframe's</code> <b>column titles</b> in accordance with the <b>mapping indicated</b>
with the help of the "<code>columnTitlesDict</code>" <code>dictionary</code>.</p>

# In[13]:
columnTitlesDict = {"SITENAME": "TNAcronym", "WCFILEPATH": "fileUrl", "TITLE": "title",
"DESCRIPTION": "description", "PUBLISHEDAT": "dateCreated", "DEFAULTAUDIOLANGUAGE":
"contentInLanguage"}

duplicateFreeYouTubeVideosDf = duplicateFreeYouTubeVideosDf.rename(columns =
columnTitlesDict)

# <p><b>Reseting the indexes</b> of the "<code>duplicateFreeYouTubeVideosDf</code>"
<code>dataframe</code>.</p>

# In[14]:
duplicateFreeYouTubeVideosDf.reset_index()

# <p><b>Displaying the first 30 rows</b> of the <b>newly created</b>
"<code>duplicateFreeYouTubeVideosDf</code>" <code>dataframe</code>.</p>

# In[15]:
duplicateFreeYouTubeVideosDf.head(30)

# <p><b>Extracting</b> some <b>descriptive statistics</b> from the "<b>views</b>"
<b>column</b> containg the number of views per YouTube video.</p>

# In[16]:
youTubeVideosDf["views"].describe()

# <p><b>Selecting those YouTube videos</b> (from the list of the duplicate free YouTube video
URLs) that <b>have a number of views higher that the median</b> (i.e. <b>118.5</b>)</p>

# In[17]:
selectedDuplicateFreeYouTubeVideosDf =
duplicateFreeYouTubeVideosDf[duplicateFreeYouTubeVideosDf["views"] > 118.5]

# <p><b>Total number</b> of <b>duplicate free YouTube videos</b> to integrate into the
EURAKNOS repository.</p>

# In[18]:
selectedDuplicateFreeYouTubeVideosDf.shape[0]

# <p><b>Reseting the indexes</b> of the "<code>selectedDuplicateFreeYouTubeVideosDf</code>"
<code>dataframe</code>.</p>

# In[19]:
selectedDuplicateFreeYouTubeVideosDf.reset_index()

# <p><b>Exporting</b> the "<code>selectedDuplicateFreeYouTubeVideosDf</code>"
<code>dataframe</code> as a <code>csv</code> <b>file</b>.</p>

# In[20]:
selectedDuplicateFreeYouTubeVideosDf.to_csv("selectedDuplicateFreeYouTubeVideosDf.csv", sep =
";", index = False, header = True)

# <p><b>Exporting</b> the "<code>selectedDuplicateFreeYouTubeVideosDf</code>"
<code>dataframe</code> as an <code>excel</code> <b>file</b>.</p>

# In[21]:
selectedDuplicateFreeYouTubeVideosDf.to_excel("selectedDuplicateFreeYouTubeVideosDf.xlsx",
index = False, header = True)
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

Identification and selection of presentations in a pdf file format

```
# coding: utf-8

# <h3>Searching for Thematic Network - related presentations in the list of the pdf files harvested</h3>

# In[1]:
import pandas as pd

# <p><b>STEP 1</b>: <b>Loading</b> and <b>reading</b> the <code>csv</code> file containing all the pdf-related records as a <code>Pandas dataframe</code>.</p>

# In[2]:
pdfsDf = pd.read_csv("pdfs_harvested.csv", sep = ";", header = 0, index_col = False, encoding = "utf-8")

# <p>What is the <b>total number of the pdf files</b> that have been harvested by the EURAKNOS crawler?</p>

# In[3]:
print("The total number of the pdf files harvested is:", pdfsDf.shape[0])

# <p>Displaying the <b>first five rows</b> of the <b>pdfsDf</b> <code>dataframe</code>.</p>

# In[4]:
pdfsDf.head()

# <p><b>STEP 2</b>: <b>Searching for</b> any pdf files (from the bulk of the pdfs harvested by the crawler) that may <b>contain the word</b> "<b>presentation</b>" in their <code>fileURL</code>.</p>

# In[5]:
counter = 0
for fileUrl in pdfsDf["WCFILEPATH"]:

    if "presentation" in fileUrl:

        counter +=1

        print(fileUrl)

print("\nThe total number of presentations in a pdf format that have been found is:", counter)

# <p><b>STEP 3</b>: <b>Creating a new <code>dataframe</code></b> containg all the <b>pdf records related to presentations</b>.</p>

# In[40]:
rowNum = -1
rowList = []
for fileUrl in pdfsDf["WCFILEPATH"]:

    rowNum +=1

    if "presentation" in fileUrl:

        #print(rowNum)

        rowList.append(pdfsDf.iloc[rowNum])

presentationsInPdf = pd.DataFrame(rowList)

# <p><b>Displaying</b> all the <b>rows of the <code>dataframe</code></b> holding <b>presentations in a pdf format</b>.</p>
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
# In[42]:
presentationsInPdf

# <p><b>STEP 4</b>: <b>Exporting the dataframe</b> containing all the <b>presentations in a pdf format</b> (i.e. the <code>presentationsInPdf dataframe</code>) as a <code>csv</code> file.</p>

# In[44]:
presentationsInPdf.to_csv("presentations_in_pdf/presentations_in_pdf.csv", sep = ";", header = True, encoding = "utf-8")
```

Identification and selection of infographics

```
# coding: utf-8

# <h3>Searching for Thematic Network - related infographics in the list of the pdf files harvested</h3>

# In[1]:
import pandas as pd


# <p><b>STEP 1</b>: <b>Loading</b> and <b>reading</b> the <code>csv</code> file containing all the pdf-related records as a <code>Pandas dataframe</code>.</p>

# In[2]:
pdfsDf = pd.read_csv("pdfs_harvested.csv", sep = ";", header = 0, index_col = False, encoding = "utf-8")

# <p>What is the <b>total number of the pdf files</b> that have been harvested by the EURAKNOS crawler?</p>

# In[3]:
print("The total number of the pdf files harvested is:", pdfsDf.shape[0])

# <p>Displaying the <b>first five rows</b> of the <b>pdfsDf</b> <code>dataframe</code>.</p>

# In[4]:
pdfsDf.head()

# <p><b>STEP 2</b>: <b>Searching for</b> any pdf files (from the bulk of the pdfs harvested by the crawler) that may <b>contain the word</b> "<b>infographic</b>" in their <code>fileURL</code>.</p>

# In[5]:
counter = 0
for fileUrl in pdfsDf["WCFILEPATH"]:

    if "infographic" in fileUrl:

        counter +=1

        print(fileUrl)

print("\nThe total number of infographics in a pdf format that have been found is:", counter)

# <p><b>STEP 3</b>: <b>Creating a new <code>dataframe</code></b> containg all the <b>pdf records related to presentations</b>.</p>

# In[7]:
rowNum = -1
rowList = []
for fileUrl in pdfsDf["WCFILEPATH"]:
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863

```
    rowNum +=1
    if "infographic" in fileUrl:
        #print(rowNum)
        rowList.append(pdfsDf.iloc[rowNum])
infographicsInPdf = pd.DataFrame(rowList)
# <p><b>Displaying</b> all the <b>rows of the <code>dataframe</code></b> holding
<b>infographics in a pdf format</b>.</p>
# In[9]:
infographicsInPdf
# <p><b>STEP 4</b>: <b>Exporting the dataframe</b> containing all the <b>infographics in a pdf
format</b> (i.e. the <code>infographicsInPdf dataframe</code>) as a <code>csv</code> file.</p>
# In[11]:
infographicsInPdf.to_csv("infographics/infographics_in_pdf.csv", sep = ";", header = True, encoding =
"utf-8")
```

The EURAKNOS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 817863