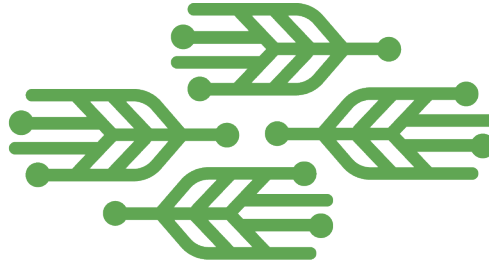


Deliverable 4.3
e-KRP documentation and guide



EURAKNOS

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



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 development-related work that has taken place both for the back- and front-end of the platform, the testing of the e-KRP, and the development of an initial 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.

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)	<input checked="" type="checkbox"/>
	DEC	Websites, patents filing, press & media actions, videos	
	E	Ethics	

Dissemination Level	PU	Public	<input checked="" type="checkbox"/>
	CO	Confidential, only for members of the consortium	



Table of Contents

List of figures	5
List of tables	6
List of abbreviations and acronyms.....	8
1. Introduction.....	9
1.1. Context	9
1.2. Objectives.....	10
2. Task members.....	11
3. Methodology	11
4. The EURAKNOS e-KRP system architecture.....	13
5. Shaping up the repository of the EURAKNOS e-KRP	14
5.1. Setting-up the sample of Thematic Networks to harvest content from	15
5.2. Harvesting content from the Knowledge Reservoirs of the sample Thematic Networks	20
5.2.1. The EURAKNOS web crawler	20
5.2.2. A pool of data objects to consider for the e-KRP repository	21
5.3. Assembling the input for the e-KRP repository	24
5.3.1. Data cleaning.....	25
5.3.2. Selecting the data objects to integrate into the repository	28
5.3.3. Providing metadata values to the data objects.....	35
5.4. Integration of data objects into the e-KRP repository	41
5.5. Evaluating the FAIRness of the e-KRP.....	43
6. The e-KRP's back- and front-end development.....	46
6.1. Back-end development	47
6.2. Front-end development	47
7. Testing the EURAKNOS e-KRP	49
7.1. Results obtained from preliminary user tests	50
7.1.1. Feedback from the wireframe-based exercise held in the Athens consortium meeting.....	50
7.1.2. Feedback from the Marketplace session held in the Paris workshop	57
7.2. Testing the e-KRP against other existing systems	58
7.2.1. The breadth of the content and information delivered	58
7.2.2. Information Architecture.....	61
7.2.3. Employment of visual elements and aesthetics	64
8. Sustainability of the EURAKNOS e-KRP.....	65
8.1. The EURAKNOS sustainability framework	65
8.2. Towards an e-KRP sustainability plan.....	67
9. Recommendations for future Thematic Networks.....	71
10. e-KRP user guide	73
10.1. e-KRP landing page.....	73
10.2. Sector- and theme-related pages.....	76
10.3. Display of search results.....	85



11. Conclusions and next steps	89
12. References	91
13. Annex I: Functionalities of the EURAKNOS web crawler	93
14. Annex II: Example Python script for classifying documents into the appropriate document type	94
15. Annex III: Python script for the removal of duplicate pdf files.....	99
16. Annex IV: Python scripts for the selection of YouTube videos, presentations in a pdf file format, and infographics	103



List of figures

Figure 1. The context of Task 4.2	10
Figure 2. The methodology adopted in Task 4.2	12
Figure 3. The architecture of the EURAKNOS e-KRP system	14
Figure 4. Phases of the Task 4.2 methodology relating to the shape up of the e-KRP repository.....	14
Figure 5. The sequence of criteria for setting-up the sample of TNs to consider as data object sources	16
Figure 6. The TNs selected for the sample to serve as data sources.	20
Figure 7. An indicative architecture of a web crawler (Patil & Patil, 2016)	21
Figure 8. The landing page of the EURAKNOS web crawler	22
Figure 9. Statistics related to the data objects harvested from the KR of the FERTINNOWA TN	22
Figure 10. The pool of data objects to consider for the e-KRP repository	24
Figure 11. 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	29
Figure 12. The EURAKNOS data model.....	36
Figure 13. 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.....	40
Figure 14. 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).....	41
Figure 15. Sequence diagram displaying the ETL process for harvesting data objects from the KR of a TN and storing them to local storage	42
Figure 16. Sequence diagram illustrating the registration of data objects and their metadata into DSpace	43
Figure 17. 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	43
Figure 18. Evaluation of the EURAKNOS e-KRP's FAIRness with the help of the SATIFYD tool.....	46
Figure 19. The EURAKNOS prototype e-KRP's Information Architecture	48
Figure 20. The navigational structures initially proposed for the e-KRP through the design of the respective wireframe	50
Figure 21. Version 1 of the e-KRP's landing page dashboard.....	51
Figure 22. Version 2 of the e-KRP's landing page dashboard.....	52
Figure 23. Version 3 of the e-KRP's landing page dashboard.....	52
Figure 24. Version 4 of the e-KRP's landing page dashboard.....	53
Figure 25. Search function provided in every e-KRP page with the aim to facilitate content and information search	53
Figure 26. Wireframe of news article with a picture	54
Figure 27. Wireframe of news article without a picture	54
Figure 28. Version 1 of the library overview	54
Figure 29. Version 2 of the library overview	55
Figure 30. Version 3 of the library overview	55
Figure 31. Wireframe of academic article view	56
Figure 32. Wireframe of podcast view and access.....	56
Figure 33. Wireframe of video view and access.....	57

Figure 34. Access to the Practice Abstracts created by a TN from the EIP-AGRI database	58
Figure 35. TN details available from the EIP-AGRI database.....	59
Figure 36. Details of the partners involved in a TN available from the EIP-AGRI database.....	59
Figure 37. Access to audio-visual material through the EIP-AGRI database	60
Figure 38. Access to a breadth of content and information, in a wide range of file formats, from the EURAKNOS prototype e-KRP	61
Figure 39. The EIP-AGRI service point website’s main navigational structure.....	61
Figure 40. The EURAKNOS e-KRP’s main navigational structure	62
Figure 41. The search function available to search for content and information in the EIP-AGRI database	62
Figure 42. The search function of the EURAKNOS e-KRP	62
Figure 43. Filtering options provided in the e-KRP to elaborate the search results	63
Figure 44. Results of a search operation in the EIP-AGRI database. The display of the results is dominated by the use of textual information.	63
Figure 45. The landing pages of the EIP-AGRI database	64
Figure 46. The landing pages of the EURAKNOS e-KRP.....	64
Figure 47. The EURAKNOS sustainability framework.....	66
Figure 48. The landing page of the EURAKNOS e-KRP	73
Figure 49. 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	74
Figure 50. e-KRP page with details about the adopted privacy-relayed policy	75
Figure 51. The “Contact us” page of the e-KRP	75
Figure 52. The sector- and theme-related pages able to be visited by making a selection from the relevant menu or by clicking on one of the landing page’s tiles	77
Figure 53. The e-KRP page related to the sector of Forestry	78
Figure 54. Topics of discussion in the “Community” page	79
Figure 55. Message threads related to a topic of discussion	80
Figure 56. List of Forestry experts	81
Figure 57. View of an article	82
Figure 58. View of a factsheet	83
Figure 59. View of a podcast	84
Figure 60. View of a video	85
Figure 61. Display of search results as a list	86
Figure 62. Display of search results by making use of tiles	87
Figure 63. e-KRP page displaying information about a TN.....	88
Figure 64. The information provided to the user by making use of the “Details” functionality.....	93

List of tables

Table 1. Involvement of EURAKNOS partners in the activities related to Task 4.2 and the expertise brought onboard	11
Table 2. Alignment of task-related objectives with phases of the methodology	12
Table 3. Outcomes of the analysis of TNs in terms of existence and the different implementations of their KRs (adapted from Marois et al., 2019)	16



Table 4. Completion rates of TNs (until June, 2020)	17
Table 5. List of TNs obtained as the outcome of the application of Criterion #2 and the domain their theme relates to	18
Table 6. Type of involvement of EURAKNOS partners in the 19 TNs outputted as the outcome of the application of Criterion #3.....	19
Table 7. 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	23
Table 8. The file formats of the data objects harvested by the EURAKNOS crawler and the data object types/categories they are associated with	25
Table 9. The algorithm used for associating the pdf files harvested by the EURAKNOS web crawler with the correct document-related type.....	27
Table 10. Types and numbers of documents harvested from the FERTINNOWA TN's KR and considered as input for the e-KRP repository	30
Table 11. Types and numbers of documents harvested from the OK-NET Arable TN's KR and considered as input for the e-KRP repository	31
Table 12. Types and numbers of documents harvested from the Smart-AKIS TN's KR and considered as input for the e-KRP repository	31
Table 13. Types and numbers of documents harvested from the SKIN TN's KR and considered as input for the e-KRP repository	31
Table 14. Types and numbers of documents harvested from the 4D4F TN's KR and considered as input for the e-KRP repository	32
Table 15. Types and numbers of documents harvested from the Inno4Grass TN's KR and considered as input for the e-KRP repository	32
Table 16. Types and numbers of documents harvested from the SheepNet TN's KR and considered as input for the e-KRP repository	32
Table 17. Types and numbers of documents harvested from the AFINET TN's KR and considered as input for the e-KRP repository	33
Table 18. The number of practice abstracts per TN that have been considered as input for the e-KRP repository	33
Table 19. Number of documents per type available in total	34
Table 20. Descriptive statistics related to the number of YouTube and Vimeo videos harvested from the TNs' KRs.....	35
Table 21. Alignment of relations/properties described in the EURAKNOS model with properties defined in external vocabularies	37
Table 22. The metadata properties used to annotate all the data objects available through the EURAKNOS e-KRP	38
Table 23. Results of the EURAKNOS e-KRP's FAIRness evaluation.....	44
Table 24. Key results of the comparison between the EIP-AGRI database and the EURAKNOS e-KRP	65
Table 25. Brief description of each of the EURAKNOS sustainability framework's dimensions	66
Table 26. Example cases of IPR management from existing TNs	68
Table 27. Main Creative Commons licenses together with a short description of them.....	69
Table 28. Details of the CC BY-NC-ND 4.0 license selected to attribute to the e-KRP's data objects...	70
Table 29. Parameters and their values to be defined when a new, individual web crawler is being created	93

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



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 (namely, 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 (“Shaping up the repository of the EURAKNOS e-KRP”) to 7 (“Testing the EURAKNOS e-KRP”) provide the details of the execution of each phase of the methodology. Section 4 (“The EURAKNOS e-KRP system architecture”) presents the architecture of the e-KRP system, Section 8 (“Sustainability of the EURAKNOS e-KRP”) is a follow-up to the discussion about e-KRP-related sustainability issues, which has started in Deliverable 4.1¹, and Section 9 (“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 10 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 EUREKA², are summarised in Section 11 (“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/>



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 the project tasks leading to Task 4.2 by also showing the output generated from/input received by each one of those.

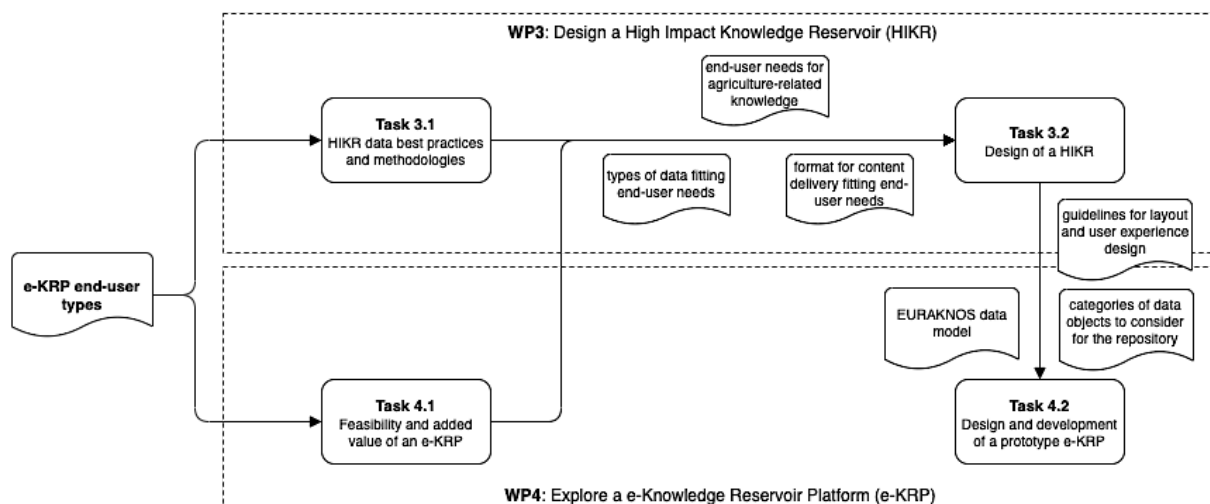


Figure 1. The context of Task 4.2

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 7.

1.2. Objectives

The main aim of Task 4.2 has been to create a digital platform (the e-KRP), collect data objects (digital content and information available in a range 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 design of the e-KRP system’s architecture;
- the back- and front-end development of the e-KRP based on 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 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 noted that the University of Maastricht (MU) has also participated in a part of the interactions undertaken in Task 4.2, because of its WP leader role in EUREKA (namely, in WP3: “Knowledge selection and developing a blueprint for the FarmBook”) which takes on and extends the design and development 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. The involvement of each partner is described with regard to the methodology phase in which it participated³.

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	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • Software engineering/digital platform development/back-end development. • Data modelling. • Semantic web technologies. • Data (pre-)processing and analysis.
LFG	<ul style="list-style-type: none"> • Design of the e-KRP system architecture. • Back-end development. • Front-end development. • e-KRP testing. 	<ul style="list-style-type: none"> • User Experience (UX) design and engineering. • Software engineering/front-end development/back-end development. • Software product testing.
AKI	<ul style="list-style-type: none"> • 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. 	<ul style="list-style-type: none"> • Software engineering/back-end development/front-end development. • Data (pre-)processing and analysis.
IfA	<ul style="list-style-type: none"> • e-KRP testing. 	<ul style="list-style-type: none"> • Communication and dissemination expert. It has brought the end-user’s perspective in the e-KRP testing process and provided insights.
TN coordinators	<ul style="list-style-type: none"> • Assembly of the input for the e-KRP repository. 	<ul style="list-style-type: none"> • Knowledge of all the details related to the data objects created in their TNs.
MU	<ul style="list-style-type: none"> • Design of the e-KRP system architecture. • Integration of data objects into the e-KRP repository. • Back-end development. 	<ul style="list-style-type: none"> • Software engineering/back-end development. • Data modelling. • Semantic web technologies.

3. Methodology

The methodology that has been adopted and followed in Task 4.2 is shown in Figure 2. What is made evident from that 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 have ranged from the selection of the TNs that made up the sample of the data object sources for the repository of the e-KRP, to the design of the e-KRP system’s architecture, the back- and front-end development, the assembly of the input for the repository and the integration of content into the repository. In the cases of the “Set-up of the sample of TNs”, “Harvesting of data objects from the KRs of the sample TNs”, as well as the “Assembly of the input for the e-KRP repository” phases, there has

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”).

also been a breakdown of the involved activities in a number of more detailed steps aiming to facilitate the fulfillment of the relevant task objective. In some cases, there has also been a parallel execution of activities. For instance, there has been a parallel execution of the activities that have been involved in the “Development and testing of the EURAKNOS crawler” and the “Set-up of the sample of TNs” phases, as well as in the “Back-end development” and “Front-end development” phases.

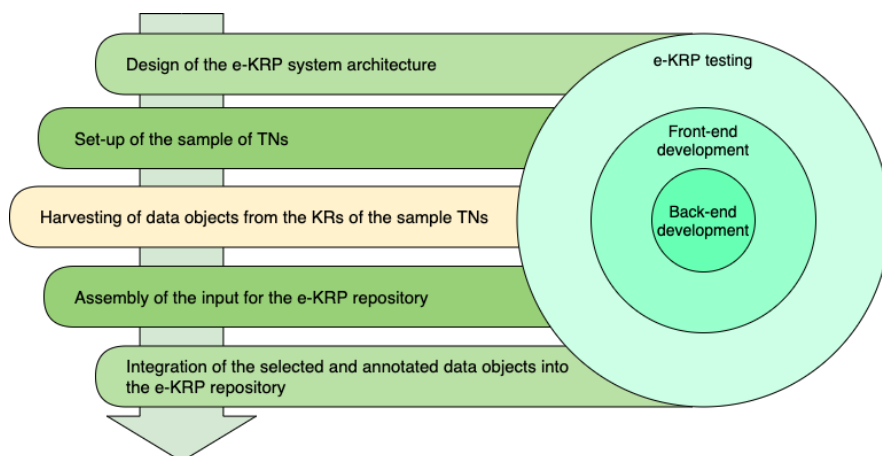


Figure 2. The methodology adopted in Task 4.2

The interactions that took place in Task 4.2 as part of the methodology phase relating to the design of the e-KRP system architecture and its outcome are presented in Section 4. In Section 5 (“Shaping up the repository of the EURAKNOS e-KRP”), there are detailed descriptions of the activities that have been implemented for: (i) setting-up the sample of TNs to consider as data object sources; (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, selecting and annotation of the data); and (iv) the integration of the data objects into the repository of the e-KRP. Details about the back- and front-end development of the e-KRP are provided in Section 6. Finally, insights from the e-KRP testing are available in Section 7.

Table 2 below presents the alignment of the objectives listed in Subsection 1.2 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
1	Design of the e-KRP system’s architecture.	Design of the e-KRP system architecture
2	Back- and front-end development of the e-KRP based on the e-KRP system’s architecture.	<ul style="list-style-type: none"> • Back-end development • Front-end development
3	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.	<ul style="list-style-type: none"> • Set-up of the sample of TNs • Harvesting of data objects from the KRs of the sample TNs
4	Selection of digital content and information of importance to the end-users to integrate it into the repository.	<ul style="list-style-type: none"> • Data cleaning • Selection of data objects for the repository
5	Annotation of the selected data objects by assigning the appropriate metadata values.	Data object annotation
6	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
7	Acquisition of insights into e-KRP-related usability and user-friendliness issues.	e-KRP testing

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 illustrated in Figure 3. It has been developed by drawing upon a stack of technologies (described in detail in Deliverable 3.2⁴) and a philosophy that was decided and selected to facilitate the follow-up of the work undertaken in EURAKNOS by its sibling project, EUREKA⁵.

As reported in Deliverable 3.2 (Van Hal et al., 2020), the solution utilised as the data repository for the TN data objects is DSpace⁶. Based on the architecture presented below, data objects harvested from the KRs of a sample of TNs (Subsection 5.1 provides all the details about the assembly of this sample of TNs) have been ingested into the system by using an appropriately created 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 Solr⁷ for the metadata of the data objects and a relational database (namely, PostgreSQL⁸) for all other information. Metadata is also stored in Fuseki⁹, 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 API¹⁰ for communication.

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 API¹¹ or by making use of the web interface (web GUI¹²) (details about the e-KRP front-end are provided in Section 7).

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 Hadoop¹³ (mentioned in the GA as the infrastructure for the temporary storage of data objects before their integration into the e-KRP repository)¹⁴. This

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 The initial intention has been to circumvent Fuseki, the triple store built in the DSpace framework, and use Virtuoso as the triple store for storing metadata in an RDF format (Virtuoso is a triple store currently used in many European projects). Such a bypass would require to make DSpace and Virtuoso “talk to each other”. After an initial investigation, it was found out that the establishment of a “communication channel” between DSpace and Virtuoso involved some technicalities requiring a substantial amount of time to be solved. Therefore, it was decided to explore this issue in more depth in EUREKA. For details about Fuseki, the reader may refer to <https://jena.apache.org/documentation/fuseki2/>.

¹⁰ 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).

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/>

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 any of the advanced data storage- and curation-related functionalities sought for the e-KRP repository. In any case, and for reasons clearly articulated in Deliverable 3.2 (where the choices have been made in regard to the EURAKNOS technology stack), the DSpace framework has proven to be the best solution for the implementation of the e-KRP repository.

flexibility, together with the features of the EURAKNOS API (documented in Deliverable 4.2¹⁵), 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.

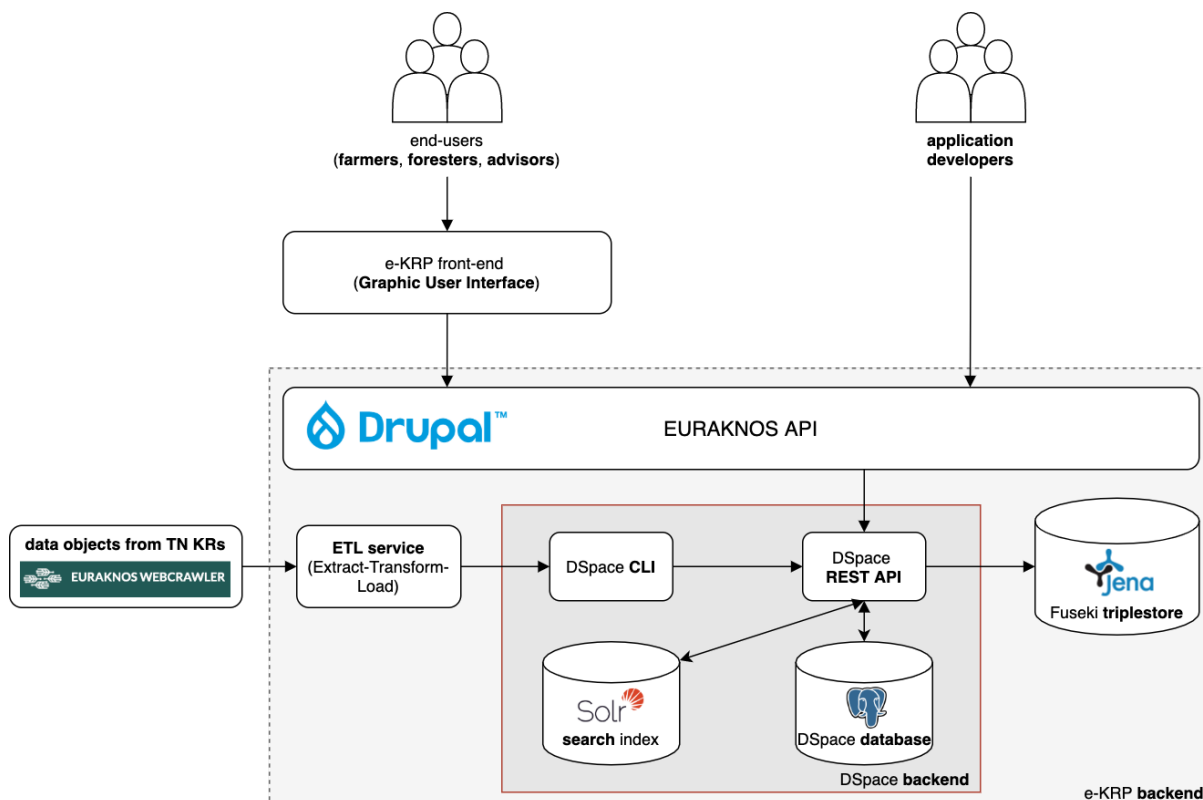


Figure 3. The architecture of the EURAKNOS e-KRP system

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

This section of the report focuses on the presentation and explanation of the activities that have taken place as part of the phases of the methodology shown in Figure 4.

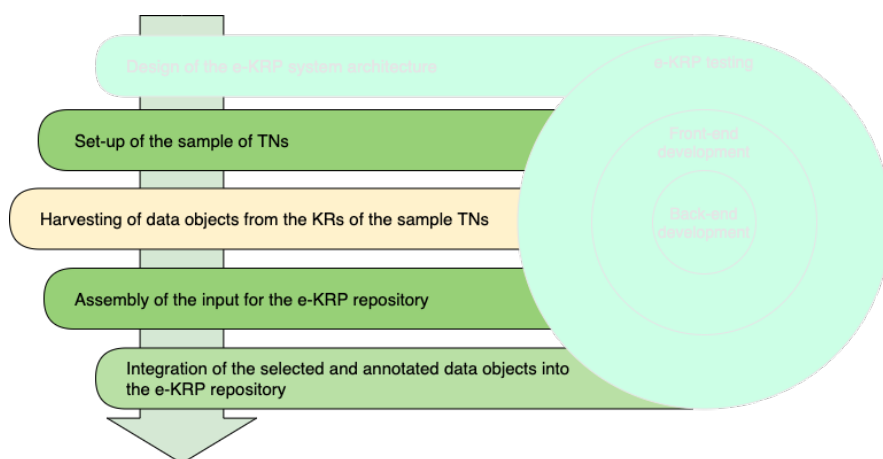


Figure 4. Phases of the Task 4.2 methodology relating to the shape up of the e-KRP repository

The phases of the methodology illustrated above and the activities implemented in their context have served the purpose of populating the e-KRP repository with digital content by starting from the set-up

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.

of the sample of TN KRs to consider as sources of data objects and the harvesting of the digital content from them. All four phases that appear in Figure 4 are described in detail in Subsections 5.1 to 5.4.

5.1. 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 existing TNs' KRs took place¹⁶ 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")¹⁷. 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")¹⁸, a new process of identifying/estimating the rates of TNs' completion took place in Task 4.2 to draw upon up-to-date information.

- **Criterion #3:** Domain the TNs belong to¹⁹

The term "domain" refers to the agricultural subsector that a TN's theme relates to²⁰. 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 arise 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.

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 total²¹) 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 considered for the analysis of the TNs' KRs undertaken in the context of Task 2.2 ("Design

¹⁶ Task 2.2 started in January 2019 (M01 of the project) and ended in August 2019 (M08).

¹⁷ 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.

¹⁸ Task 2.1 started in January 2019 (M01 of the project) and ended in August 2019 (M08).

¹⁹ "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.

²⁰ According to the EIP-AGRI's report *"Horizon 2020 thematic networks work on a specific theme"* (2016).

²¹ 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).



of knowledge reservoirs within current and past TN”)²² and reported in Deliverable 2.2 (“Report on format and design of knowledge reservoirs of TNs”)²³. Figure 5 shows the sequence in which the four (4) criteria have been applied.

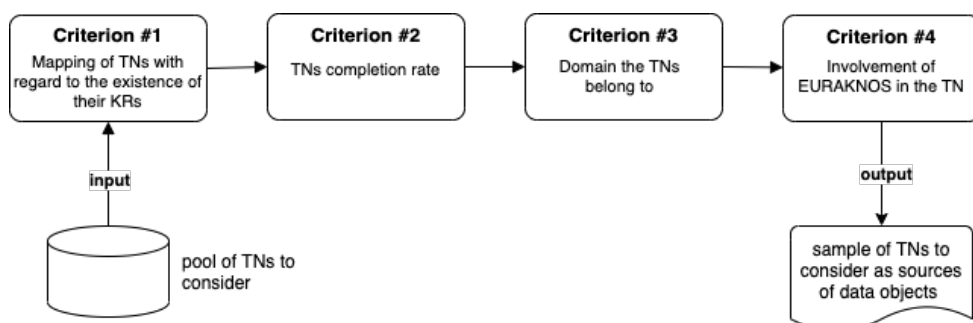


Figure 5. The sequence of criteria for setting-up the sample of TNs to consider as data object sources

Table 3 below summarises the results of the analysis executed in Task 2.2 regarding whether a TN’s KR exists (or more specifically, existed at the time of the analysis) and the different types of the existing KRs. Apart from listing the TNs having no KR implemented, it also allows insights into 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 3. Outcomes of the analysis of TNs in terms of existence and the 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%	<ul style="list-style-type: none"> • FERTINNOWA • AGRI FOR VALOR • SUWANU • EURODAIRY • 4D4F • EUPIG • AGRISPIN • HNV-LINK • CERERE • NEWBIE • SKIN • INCREDIBLE
The TN uses its official website as its KR together with a data repository dedicated to this purpose	1	4%	<ul style="list-style-type: none"> • HENNOVATION
The TN uses a general-purpose data repository as its KR (not its own, custom-made repository)	3	11%	<ul style="list-style-type: none"> • AFINET • SHEEPNET • NUTRIMAN
The TN uses a custom-made data repository dedicated to the purpose of serving as its KR	6	21%	<ul style="list-style-type: none"> • SMART-AKIS • WINETWORK • INNO4GRASS • OK-NET ARABLE • OK-NET ECOFFED • EU-FRUIT
No KR available	6	21%	<ul style="list-style-type: none"> • ENABLING • LEGUMES TRANSLATED • INNOSETA • DISARM • PANACEA • BEST4SOIL
TOTAL	28	100%	

²² 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, INCREDIBLE, INNO4GRASS, INNOSETA, LEGUMES TRANSLATED, NEWBIE, NUTRIMAN, OK-NET ARABLE, OK-NET ECOFEED, PANACEA, SHEEPNET, SKIN, SMART-AKIS, SUWANU, and WINETWORK.

²³ 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.

Based on the results available in the table above, there is a total number of six (6) TNs for which there was no KR available by the time at which the analysis was in progress. These TNs are BEST4SOIL, DISARM, ENABLING, LEGUMES TRANSLATED, INNOSETA and PANACEA. These TNs were excluded from the list of the candidate TNs for the final sample. Therefore, the outcome of the application of Criterion #1 has been the list of the remaining TNs²⁴, 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 or the degree to which a TN has been completed (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")²⁵. However, an update was necessary in order to draw upon up-to-date information. To this end, a process of searching for TN-related information in the CORDIS database²⁶ took place. This information related to the start and end years and months of the 22 TNs to identify/ calculate their completion rates. The results are shown in Table 4.

Table 4. 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%

²⁴ 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.

²⁵ 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.

²⁶ <https://cordis.europa.eu/en>



TN acronym	TN URL	Start year	End year	Duration	Completion rate
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:

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

To further filter the list of the TN sample candidates, only the TNs with a completion rate greater than or equal to 75% ($\geq 75\%$) were shortlisted²⁷. 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 5 concerning the domain (i.e., the agricultural subsector) their theme relates to.

Table 5. 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 database²⁸ that provides access to TN-related information; and
- by using the cortical.io web-based application²⁹ after having provided a TN-related description³⁰ as input to it.

²⁷ 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.

²⁸ <https://ec.europa.eu/eip/agriculture/en/find-connect/projects>

²⁹ <https://www.cortical.io/freetools/extract-keywords/>

³⁰ 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 last column of Table 5 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 necessitated keeping the distribution presented in Table 5. This means that 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: $8 \times 0.526 = 4.208 \cong 4 \text{ TNs}$
- Number of TNs related to “Livestock” in the sample: $8 \times 0.368 = 2.944 \cong 3 \text{ TNs}$
- Number of TNs related to “Forestry” in the sample: $8 \times 0.105 = 0.84 \cong 1 \text{ TN}$

The final step in the process of creating the sample of the eight (8) TNs, aiming to serve as the sources of data objects for the e-KRP, has focused on the identification of the final list of sample members from the pool of the 19 TNs produced as the output of the application of Criterion #3. Doing so has involved the use of Criterion #4 (i.e., “Involvement of EURAKNOS partners in the TN”). As far as the involvement of EURAKNOS partners in the 19 TNs is concerned, we distinguish the following:

- A partner in EURAKNOS is the coordinator of the TN.
- A partner in EURAKNOS is a member of the TN’s consortium (i.e., a TN partner).
- A partner in EURAKNOS does not have 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 6³¹ below presents the involvement of the EURAKNOS partners in the 19 TNs outputted from the application of Criterion #3.

Table 6. Type of involvement of EURAKNOS partners in the 19 TNs outputted as the outcome of 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

³¹ The information presented in Table 6 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.

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 related to the establishment of a communication channel between EURAKNOS and the TN.

Based on this assumption, OK-NET ARABLE, FERTINNOWA, and SMART-AKIS are three (out of the four) TNs selected for our sample from the “Crop Farming” domain. Given the fact that there is no TN where a EURAKNOS partner is involved under the role of consortium partner, the fourth TN, from the specific domain, has 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 also taking account of these TNs’ rates of completion, the sample members were chosen among the Livestock-related TNs already finished (namely, 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. To this end, the sample TNs, coming from the specific domain, were randomly selected. After this random selection process, the Livestock-related TNs included in the sample were 4D4F, INNO4GRASS, and SHEEPNET.

The TN selected from the “Forestry” domain is AFINET (an already completed TN coordinated by USC; i.e., a partner in EURAKNOS).

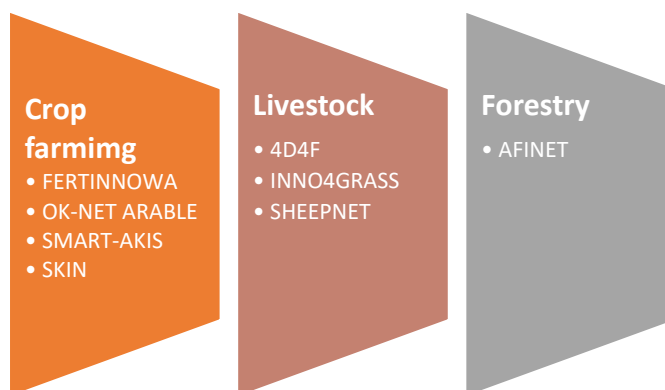


Figure 6. The TNs selected for the sample to serve as data sources.

Figure 6 shows the TNs finally included in the sample and their domain.

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

The categories/types of the data objects to store into the repository of the EURAKNOS e-KRP that have been proposed in Deliverable 3.2 (“Report on HIKR design”)³², together with the sample of TNs to serve as data sources, have provided the input required for the implementation of the “Harvesting of data objects from the KRs of the sample TNs” phase of the methodology. This data object harvesting process has been implemented with the support of a software application specifically created for this purpose; i.e., the EURAKNOS web crawler. Before providing a few details about the pool of the data objects that have been harvested from the web crawler (namely, the input to consider for the repository of the e-KRP), some background information on the web crawling information is presented below.

5.2.1. 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 gets initiated 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 information/content that a web crawler can harvest from the Web depends on the

³² 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.

instructions provided during its development phase. The resources to be crawled are indicated by a list of URLs (also referred to as “seeds”). As shown in Figure 7 below, in general terms, a crawling application is 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.

As far as the EURAKNOS crawler is concerned, the points listed below provide 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 TN KRs available.
- 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.
- 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 are hosted. AKI has granted permission to all WP4 partners to enable 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 several modifications to the web crawler, which have allowed for also harvesting data from websites developed with the use of dynamic JavaScript frameworks.

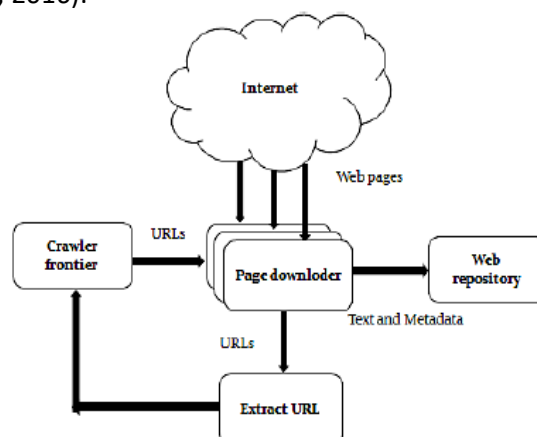


Figure 7. An indicative architecture of a web crawler (Patil & Patil, 2016)

5.2.2. A pool of data objects to consider for the e-KRP repository

The process of crawling the KRs of TNs (namely, the TNs of our sample) for harvesting the 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 have been harvested from the KRs of the sample TNs by the EURAKNOS 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 8 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 the harvesting of data from a specific TN’s 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 data objects that have been harvested, but not

being able to be classified in any of the data object 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 that have been harvested per category allowed to cross-validate the outcomes of the research undertaken in Task 2.1 by Mosquera Losada et al. (2019). Another conclusion relates to the fact that specific categories of data objects (e.g., videos), which are mostly preferred by the end-users, are not the most frequently occurring ones.

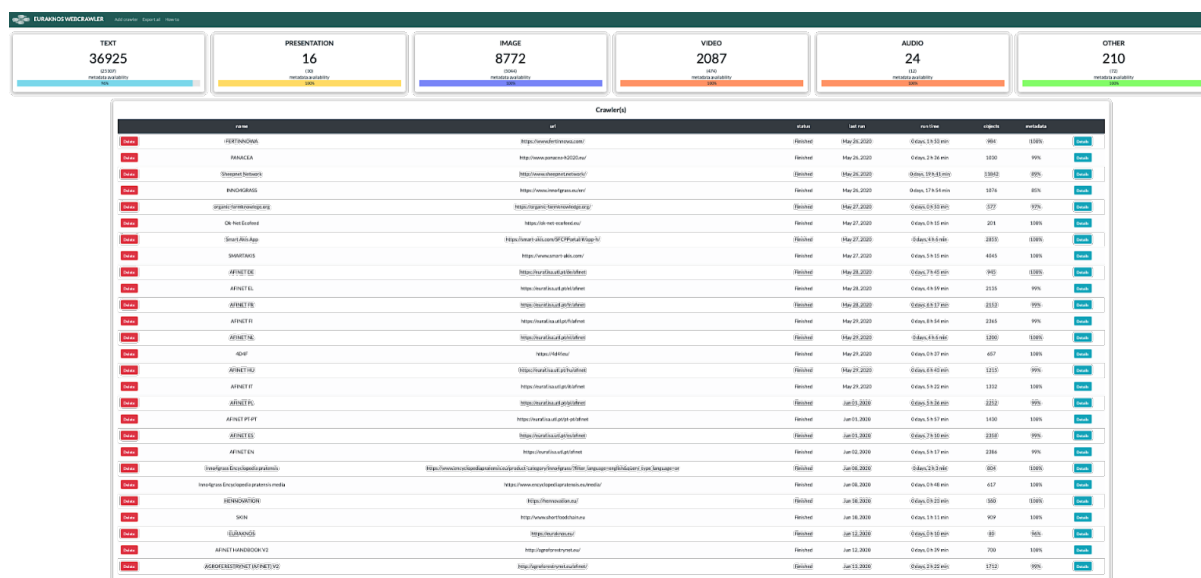


Figure 8. The landing page of the EURAKNOS web crawler

The list below the array of tiles gives details about each web crawler’s status. These details are about: (i) the name of the crawler; (ii) the URL of the crawled TN; (iii) the web crawler’s status (e.g., finished); (iv) the date of the last crawler execution³³; (v) the time required for the completion of the crawling process; (vi) the number of the data objects harvested; and (vii) the rate of metadata availability. A “Details” button available on the right side of each line allows to access some useful statistics related to the kinds of data objects harvested from a TN’s KR, as well as the existence of metadata per category of data objects. Figure 9 provides an overview of this statistics-related information for the data objects harvested from the KR of the FERTINNOWA TN.

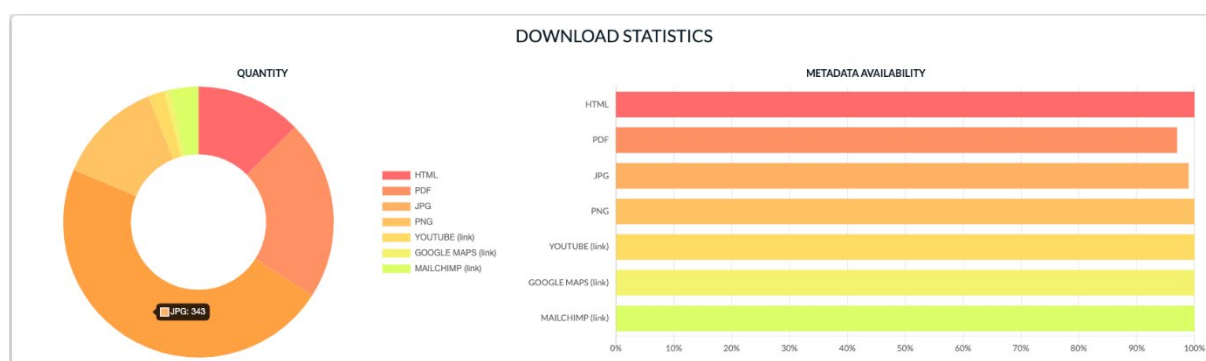


Figure 9. Statistics related to the data objects harvested from the KR of the FERTINNOWA TN

Table 7 below provides details about the categories/types and formats of the data objects harvested from each of the sample’s TNs KRs, along with information about the number of data items per type/category and format. Also, information about the completeness of the metadata³⁴ is made available.

³³ A web crawler can be executed multiple times.

³⁴ When referring to metadata, we actually mean the metadata available from the source KR of a data object.

Table 7. 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			
Crop farming	FERTINNOWA	text/ documents	html	92	text/ documents	html	100%	
			pdf	156		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
		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%	

Domain	TN name	Number of data items harvested			Rate (%) of metadata availability		
Forestry	AFINET	text/ document	html	311	text/ document	html	100%
			pdf	1147		pdf	97%
			docx	-		docx	100%
		presentations	pptx	-	presentations	pptx	100%
			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 10 shows a screenshot of the input collected for the e-KRP (the pool of data objects to consider as input for the e-KRP 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 harvested. The crawler's database contains also data objects harvested from TN KR not included in the sample. These extra TN KR have been considered for testing-related 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-farmknowledge.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://agroforestry.net.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/it/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.encyclopediapratisen-	Inno4grass Encyclopedia																
s.eu/product-	pratisen	470	59	0	0	0	7	221	25	0	0	22	0	0	0	0	0
category/inno4grass/?filter_langua	Inno4grass Encyclopedia																
https://www.encyclopediapratisen-	pratisen media	316	20	0	0	0	0	46	15	0	0	220	0	0	0	0	0
s.eu/media/	AGROFORESTRYNET																
http://agroforestry.net.eu/afinet/	(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 10. The pool of data objects to consider for the e-KRP repository

5.3. Assembling the input for the e-KRP repository

The assembly of the input for the e-KRP repository (i.e., the set of data objects to be integrated into the repository and delivered to the end-users via the e-KRP) has been a core phase of the methodology. This phase has involved three (3) 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 different end-user types targeted (this information has been provided by Deliverable 3.2³⁵ based on the work reported in Deliverables 3.1³⁶ and 4.1³⁷).
- 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 implemented for all the types/categories of the data objects identified and harvested by the web crawler and thus, considered for the repository. These categories/types are: documents, presentations, videos, audio-based objects (podcasts), and images (infographics).

5.3.1. 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 5.2.1 (“The EURAKNOS web crawler”) is that *“the information/content that a web crawler can harvest from the Web depends on the instructions provided during its development phase”*. In addition to the above, it needs to be noticed that the EURAKNOS data model refers to specific data object categories/types³⁸ with each of them being associated with a range of file formats. However, the web crawler cannot “interpret” our data model’s “abstract” data object 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 8 lists the file formats of the data objects harvested by the EURAKNOS web crawler together with the data object types/categories they are associated with.

Table 8. 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 made available by the (sample) TNs’ KRs. The same has been the case with audio-based content as well. Thus, the only instances of audio-based material delivered through the e-KRP are the EURAKNOS podcasts, mostly available as mp3 files, which have been manually retrieved.

³⁵ 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.

³⁶ 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.

³⁷ 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.

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/>

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 type⁴¹ (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 considered primary for the type/category the data objects belong to.

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 files⁴² and videos⁴³), the steps listed above were automatically implemented through the execution of specially designed code scripts⁴⁴. 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 file 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 9 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 activity⁴⁷.

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

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”).

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.

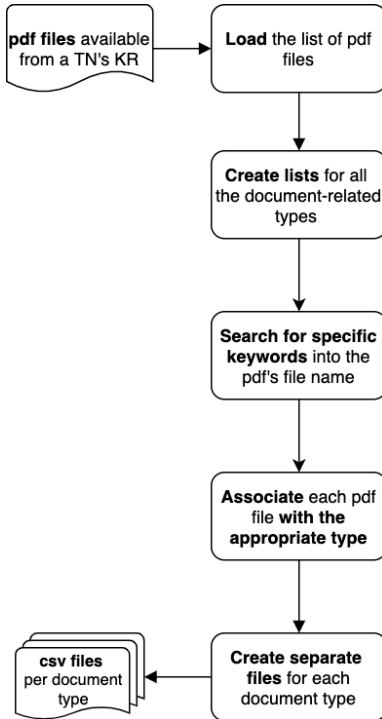
⁴⁶ The code development environment that has been used for the creation of the Python code scripts is Jupyter notebooks (<https://jupyter.org/>).

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”).

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 specially developed for this purpose. An indication of the importance of the data cleaning activity fact 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 EURAKNOS crawler. Similarly, from the 80 Vimeo videos initially harvested, only 9 remained after the duplicates’ removal. The duplicate-free lists of the 404 YouTube and 9 Vimeo videos constituted the input to the activity relating to the selection of the video-based data objects to integrate into the e-KRP repository.

Table 9. 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: “Fertinnowapolicy_pdfs.csv”, “Fertinnowanewsletter_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 images⁵¹, the classification, and duplicate removal process have been manually implemented as part of the process of the annotation of the data objects with metadata.

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.

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.

The final step in the data cleaning activity has been the identification of data objects available not only in their primary file formats. For instance, the primary file formats of presentations are ppt and pptx. However, in the bulk of data objects harvested by the EURAKNOS crawler, presentations in a pdf file format have also become available. The same has been the case with infographics. The “Presentations_in_pdf_format.ipynb”⁵² and “Infographics_in_pdf_format.ipynb”⁵³ are Python scripts developed and executed intending to identify presentations and infographics available as pdf documents.

5.3.2. 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 regarding the types and the formats of the data objects to be delivered through the platform to address the needs and preferences of the targeted end-user categories (mainly, farmers, foresters and advisors)⁵⁴. Such information has been made available from the work reported in Deliverable 3.2 (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:

- Some data object types/categories that appear to be of interest to the targeted end-user types may not be available from the (sample) TNs’ KRs.
- Evidence about preferences in slideshow-based, image-based, and audio-based information and content is available only for advisors. However, we may assume that these kinds of digital outputs could also be of relevance to farmers (and foresters).
- The majority of the data objects available by the TNs’ KRs are documents (i.e., the content available in a text-based format such as practice abstracts, factsheets, press releases, research and/or technical articles, handbooks, reports, and guides). Video-based content is available at a lower rate (it counts for 41.7% of the outputs of TNs). The same applies to image- and audio-based material which count for 33.3% and 20.8% of the total content of TNs respectively (Mosquera Losada et al., 2019).

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

- 1) **videos** of any type;

⁵² The Python script is available in Annex IV.

⁵³ The Python script is available in Annex IV.

⁵⁴ These are the main end-user categories targeted by the EURAKNOS e-KRP. However, some other user categories have been also taken into account (e.g., teachers/instructors and students).

- 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 books⁵⁵;
- 3) **presentations** with a focus on collecting (from the TNs' 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 TNs' 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 11 (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).

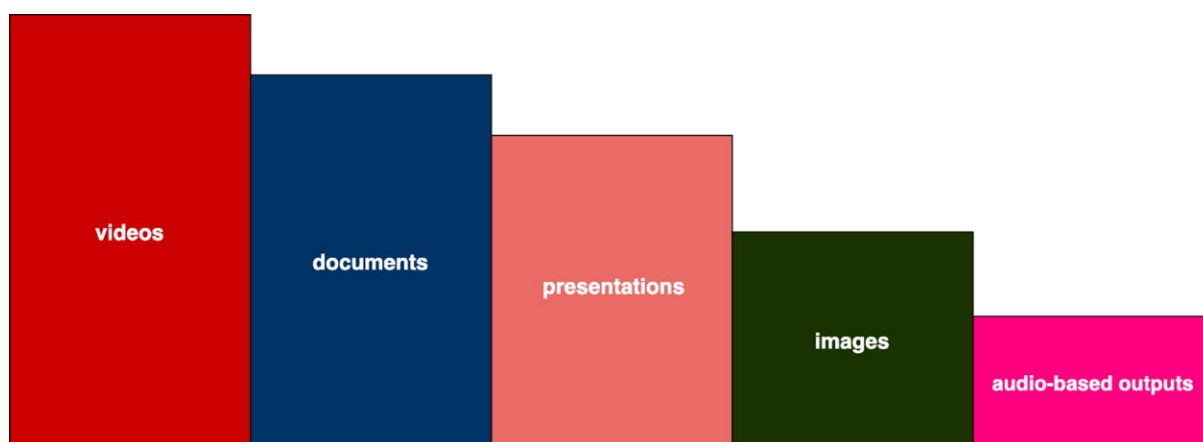


Figure 11. The data object categories integrated into the repository; the box heights are indicative of the relevance/importance of each category to the end-user preferences and needs

Before proceeding to the details of the selection of data objects for the e-KRP repository, it should be stressed that the data selection process has been applied only for the document and video data object categories, which have been those with the largest numbers of items harvested by the crawler. In the cases of presentations (slideshow-based material), as well as image- and audio-based objects, all the items harvested by the web 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 available (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 files⁵⁶), the criteria employed for selecting data objects have been both qualitative and quantitative. These criteria have not been intended to be used as strict rules, but rather as heuristics to enable the selection of content (as input to the EURAKNOS e-KRP repository) 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.

⁵⁵ 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.

⁵⁶ 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.

- Given the fact that EURAKNOS has aimed to the design and development of a prototype platform, a maximum number of 100 documents per TN⁵⁷ has been regarded as adequate for input as part of the proof-of-concept “exercise” intended to be solved.
- 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:

$$\text{number of "other" documents} = 0.6 \cdot (\text{total number of items belonging to the document types})^{58}$$
- In cases of document-related types with more than 100 items (i.e., $\text{number of items} > 100$), a 10% rate has been regarded as adequate (from the specific type) to be provided 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., $10 < \text{number of items} \leq 100$), a 40% rate has been regarded as adequate (from the specific type) to be provided 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 $\text{number of flyers} \leq 5 \rightarrow$ all items should be considered as input for the e-KRP repository.
 - If $5 < \text{number of flyers} \leq 20 \rightarrow$ a 25% rate of them should be considered as input for the e-KRP repository.
 - If $20 < \text{number of flyers} \leq 40 \rightarrow$ a 15% rate of them should be considered as input for the e-KRP repository.
 - If $\text{number of flyers} > 40 \rightarrow$ a 10% rate of them should be considered as input for the e-KRP repository.

Tables 10 to 17 show the number of documents per type and per TN available after the data cleaning process together with the number of documents that should be considered as input for the 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 TN KRs. For instance, policy briefs have not become available from the KR of the OK-NET Arable TN (see Table 11 below).

Table 10. 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

⁵⁷ There are cases of TNs from our sample in which less than 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.

⁵⁸ 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).

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 11. 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
7	Other	6	6
TOTAL		19	19

Table 12. 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 13. 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 14. 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 15. 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
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 ($40 \cdot 0.6 = 24$) ⁵⁹
TOTAL		387	64

Table 16. 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

⁵⁹ 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.

Table 17. 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 ⁶⁰
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 abstracts⁶¹ 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 database⁶². Table 18 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 18. 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

⁶⁰ 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.

⁶¹ 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.

⁶² <https://ec.europa.eu/eip/agriculture/en/about/thematic-networks-%E2%80%93-closing-research-and>

4	4D4F	24
5	Inno4Grass	25
6	SheepNet	25
7	AFINET	26
TOTAL		234

The figures appearing in Table 18 about the practice abstracts considered from the SheepNet, AFINET, Inno4Grass, and 4D4F TNs are double the size because each of the practice abstracts considered from these TNs is available in two languages.

The total numbers of documents per type (for all the sample TNs) are summarised in table 19.

Table 19. 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
8	Technical/technology article	56	17	Technical information/specifications card	96
9	Article (any article)	10	TOTAL		335

The number of documents classified into the “other” category as a result of the automated classification process was 121. As explained in Subsection 5.3.3 (“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 materials (podcasts): 24 (only available from EURAKNOS; not available from the KRs of the sample TNs)

As mentioned above, apart from documents, videos constitute the second data object type/category for which a selection process has been applied to conclude to the set of data objects to finally integrate into the repository of the e-KRP. In this case, the criterion applied has been purely quantitative; i.e., the selection of videos has been based on the number of views (the number of video views at the time by which the crawling process took place)⁶³. The number of views used as thresholds for the selection

⁶³ May and June, 2020.

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 20).

Table 20. Descriptive statistics related to the number of YouTube and Vimeo videos harvested from the TNs' KR

Descriptive statistics	YouTube videos		Vimeo videos	
	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 that 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. As far as Vimeo videos are concerned, 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) the small number of them (only 9 Vimeo videos were shortlisted after the data cleaning activity); and (ii) the more limited range of the numbers of views (range in the number of views of Vimeo videos = 1,467 vs. range in the number of views of YouTube videos = 872,729).

As far as the presentations are concerned, all the results of the crawling process have been considered as potential input. Infographics are the only image types that have been accounted for (only available in a pdf file format⁶⁴). Finally, the podcasts considered (and made available via the e-KRP) were those created in EURAKNOS (no podcasts were harvested and retrieved from the KR 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 631⁶⁵ 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).

5.3.3. Providing metadata values to the data objects

The annotation of the data objects (to be delivered through the EURAKNOS e-KRP) with metadata was the final activity of the phase of the methodology relating 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 to align the properties described in it with properties defined in widely adopted, controlled vocabularies (in an attempt to better adhere to the FAIR principles⁶⁶); (ii) the definition of the metadata to be used for the annotation of data objects; and (iii) the process of the data objects' annotation with metadata.

The annotation of the 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.

⁶⁴ 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 5.3.1).

⁶⁵ The documents belonging to the "other" category are not included in this set of 631 data objects.

⁶⁶ Detailed definitions and examples of the FAIR principles are available at <https://www.go-fair.org/fair-principles/>.

5.3.3.1 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 12 below.

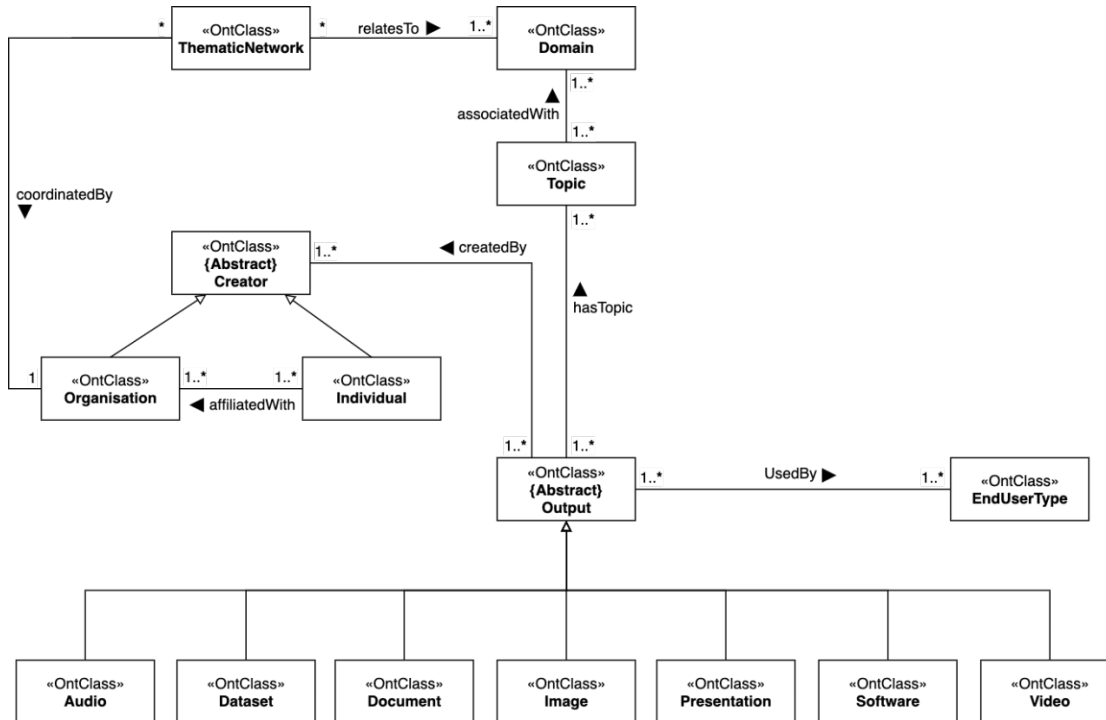


Figure 12. The EURAKNOS data model

As explicitly mentioned in Deliverable 3.2, “the EURAKNOS data model takes the form of an ontology that captures several entities of relevance to digital outputs of Thematic Networks, attributes of these entities, and how these entities link to each other.” The rationale for developing this ontology (i.e., data model) has been to identify and define the (bare minimum) set of metadata to use for adequately and efficiently describing/characterising the data objects available from the e-KRP repository⁶⁷. Delivering data objects that have been annotated with a rich metadata set is a requisite for facilitating successful search operations and an advanced user experience.

However, before the identification and definition of the minimum set of metadata to consider for the annotation of the data objects, as well as the annotation process itself, it was necessary to update our data model by taking account of existing controlled vocabularies. Such vocabularies had indeed been taken into consideration when developing our model. However, the update performed at this stage has intended to explicitly align specific relations (properties), described in the EURAKNOS data model, with relevant properties defined in external and widely adopted vocabularies. These vocabularies are schema.org (<https://schema.org/>), foaf (<http://xmlns.com/foaf/spec/>), and Dublin Core (<https://dublincore.org/specifications/dublin-core/dcmi-terms/#format>). This alignment has been pursued in an attempt to amplify the adherence of our design to the FAIR principles. More specifically, to further facilitate the FAIRness of our e-KRP repository, it was needed to shift from custom-defined properties to properties defined in the context of well-known and widely adopted vocabularies.

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 behind the identification of our metadata set has been to conclude to the bare minimum set of metadata properties with the capacity to characterise all these heterogeneous data objects outputted by the various Thematic Networks. The results of this alignment

⁶⁷ The digital outputs of Thematic Networks as they are called in the excerpt quoted from Deliverable 3.2.

process (of the properties/relations in our model with properties defined in external vocabularies) are shown in Table 21. The relations/properties in Table 21 are those considered for defining the set of our metadata. The table contains the name of each relation/property from our model, the type of the property (object or datatype property)⁶⁸, the property's domain and range (the model entities/classes whose instances are connected with this relation/property), and the URI⁶⁹ of the associated, from an external vocabulary, property.

Table 21. Alignment of relations/properties described in the EURAKNOS model with properties defined in external vocabularies

#	Relation/property of the EURAKNOS model	Type of property	Relation/ property domain	Relation/property range	Matching property from external vocabulary
1	description	Datatype property	Output class	xsd:string ⁷⁰	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 defined as instances of a class in the model	http://schema.org/contentLocation
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 as instances of a class in the 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	ThematicNetwork class	xsd:string	http://schema.org/alternateName
13	ThematicNetwork FullName	Datatype property	ThematicNetwork class	xsd:string	http://schema.org/name
14	ThematicNetwork Url	Datatype property	ThematicNetwork class	xsd:anyURI	http://schema.org/url

5.3.3.2 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 upon the updated data model described above. Based on the updated model, AUA drafted

⁶⁸ 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.

⁶⁹ 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.

⁷⁰ “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.

an initial set of metadata properties to consider for the annotation process. This draft set of metadata was brought to discussion with AKI and MU⁷¹. Discussions were held in two virtual meetings that took place in July 2020. In these meetings, the involved participants had the opportunity to go through the proposed list of metadata properties, comment upon it, and make suggestions about potential changes and adaptations, metadata properties to further consider (i.e., metadata not included in the initial set), as well as metadata properties, proposed in the initial/draft set, not too important to consider.

The result of these interactions has been the list of metadata presented in Table 22. This list contains the metadata properties that have been finally used to annotate the data objects considered as input for the e-KRP repository. Apart from the name of the metadata property, a short definition is provided together with an indication about whether there is a match with properties from external vocabularies and the respective URIs⁷². In cases of no such matches, the “no relevant property” indication is being provided.

The metadata in Table 22 constitutes the minimum set of properties for the characterisation of all the data objects delivered through the e-KRP. The development of this metadata set has been a major accomplishment of the repository-related design work in EURAKNOS, because it has allowed for the uniform annotation of all data objects despite the heterogeneity in their types and formats.

Table 22. The metadata properties used to annotate all the data objects available through the EURAKNOS e-KRP

#	Metadata property	Definition	Matching property from external vocabulary	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.	Thing > description	http://schema.org/description
3	keywords	Keywords or tags used to describe the content of the data object.	Thing > CreativeWork> keywords	http://schema.org/keywords
4	creator	The creator/author of the data object.	Thing > CreativeWork> 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.	Thing > CreativeWork> contentLocation	http://schema.org/contentLocation
7	fileUrl	The URL allowing to download the data object.	Thing > 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

⁷¹ 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 a smooth transition from the one project to the other.

⁷² See Subsection 5.3.3.1 (“Updating the EURAKNOS data model”) for a discussion on the alignment of the relations/properties of the EURAKNOS data model with properties of the schema.org vocabulary (<https://schema.org/>) as part of our data model's update.

#	Metadata property	Definition	Matching property from external vocabulary	Property URI
9	purpose	The purpose for which the data object has been created.	Thing > Property > potentialAction	http://schema.org/potentialAction
10	fileFormat	The file type of the data object.	terms > format	http://purl.org/dc/terms/format
11	type	The type property relates to what the content of the data object is about. ⁷³	no relevant property from external vocabulary	
12	fileSize	The size of the file/package. In the absence of a unit, KB is assumed.	Thing > CreativeWork > SoftwareApplication > 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.	Thing > CreativeWork > dateCreated	https://schema.org/dateCreated
15	license	A license document that applies to this content, typically indicated by URL.	Thing > CreativeWork > license	http://schema.org/license
16	ThematicNetwork Acronym	The acronym of the Thematic Network (e.g., AFINET).	Thing > Property > alternateName	http://schema.org/alternateName
17	ThematicNetwork FullName	The full name of the Thematic Network (e.g., Agroforestry Innovation Networks).	Thing > 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).	Thing > 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	

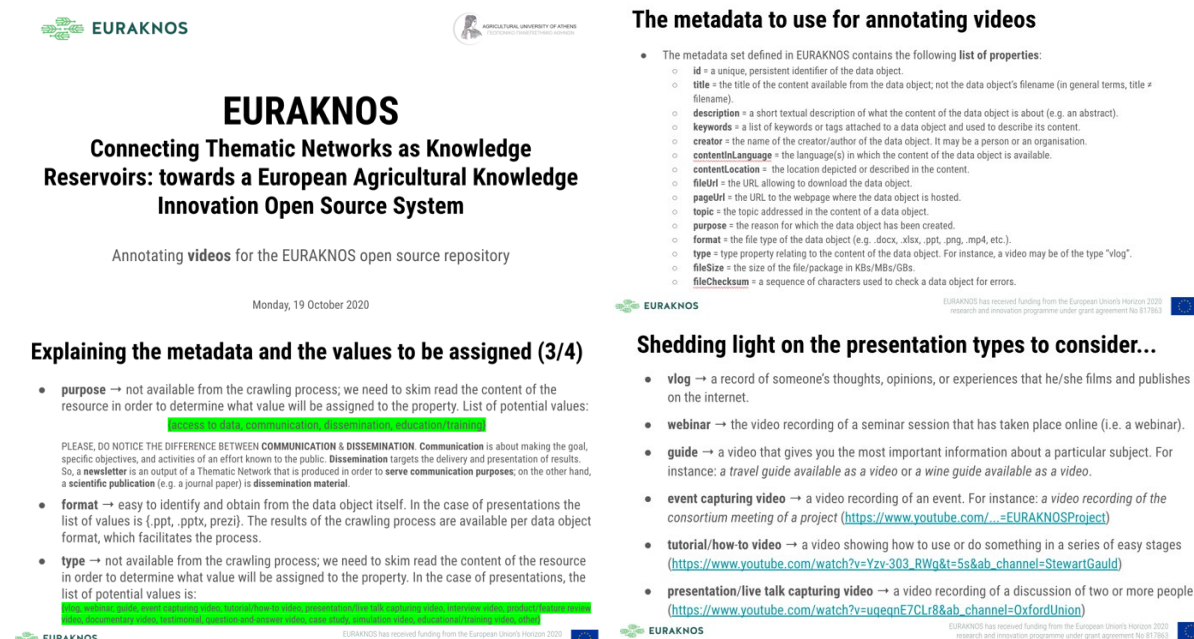
5.3.3.3 The process of the annotation of data objects with metadata values

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 e-KRP repository. This has been indeed a critical step in the work done in Task 4.2, because the characterisation of data objects with a rich metadata set is of increased importance when it comes to facilitating search operations and the overall user experience. The provision of a rich set of metadata is one of the recommendations for achieving data FAIRness⁷⁴.

⁷³ The definitions of all the different types of data objects are presented in detail in Deliverable 3.2 (Van Hal et al., 2020).

⁷⁴ See the F2 and R1 FAIR recommendations available at <https://www.go-fair.org/fair-principles/>.

The annotation process took place from the beginning of August till the end of October 2020 and was undertaken by three annotators - members of the team of the Agricultural University of Athens (AUA). The annotation process was implemented for each of the data object types/categories considered for the e-KRP repository (documents, videos, presentations, images, and podcasts). Given the importance of annotating/characterising the data objects with the “correct” (i.e., appropriate) metadata values, a training session took place before the launch of the process of each data object category’s annotation. A set of slides specially designed for this purpose was used to support training. These slides were used as a reference from the annotators during the annotation process. Some example slides (developed for instructions on the annotation of videos) are shown in Figure 13 below.



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

Annotating videos for the EURAKNOS open source repository

Monday, 19 October 2020

Explaining the metadata and the values to be assigned (3/4)

- purpose** → not available from the crawling process; we need to skim read the content of the resource in order to determine what value will be assigned to the property. List of potential values:

PLEASE, DO NOTICE THE DIFFERENCE BETWEEN COMMUNICATION & DISSEMINATION. Communication is about making the goal, specific objectives, and activities of an effort known to the public. Dissemination targets the delivery and presentation of results. So, a newsletter is an output of a Thematic Network that is produced in order to serve communication purposes; on the other hand, a scientific publication (e.g. a journal paper) is dissemination material.
- format** → easy to identify and obtain from the data object itself. In the case of presentations the list of values is {ppt, pptx, prez}. The results of the crawling process are available per data object format, which facilitates the process.
- type** → not available from the crawling process; we need to skim read the content of the resource in order to determine what value will be assigned to the property. In the case of presentations, the list of potential values is:

The metadata to use for annotating videos

- The metadata set defined in EURAKNOS contains the following list of properties:
 - id** = a unique, persistent identifier of the data object.
 - title** = the title of the content available from the data object; not the data object's filename (in general terms, title ≠ filename).
 - description** = a short textual description of what the content of the data object is about (e.g. an abstract).
 - keywords** = a list of keywords or tags attached to a data object and used to describe its content.
 - creator** = the name of the creator/author of the data object. It may be a person or an organisation.
 - contentLanguage** = the language(s) in which the content of the data object is available.
 - contentLocation** = the location depicted or described in the content.
 - fileUrl** = the URL allowing to download the data object.
 - pageUrl** = the URL to the webpage where the data object is hosted.
 - topic** = the topic addressed in the content of a data object.
 - purpose** = the reason for which the data object has been created.
 - format** = the file type of the data object (e.g. .docx, .xlsx, .ppt, .png, .mp4, etc.).
 - type** = type property relating to the content of the data object. For instance, a video may be of the type "vlog".
 - fileSize** = the size of the file/package in KBs/MBs/GBs.
 - fileChecksum** = a sequence of characters used to check a data object for errors.

Shedding light on the presentation types to consider...

- vlog** → a record of someone's thoughts, opinions, or experiences that he/she films and publishes on the internet.
- webinar** → the video recording of a seminar session that has taken place online (i.e. a webinar).
- guide** → a video that gives you the most important information about a particular subject. For instance: a travel guide available as a video or a wine guide available as a video.
- event capturing video** → a video recording of an event. For instance: a video recording of the consortium meeting of a project (<https://www.youtube.com/...=EURAKNOSProject>)
- tutorial/how-to video** → a video showing how to use or do something in a series of easy steps (https://www.youtube.com/watch?v=Yzy-303_RWg&t=5s&ab_channel=StewartGauld)
- presentation/live talk capturing video** → a video recording of a discussion of two or more people (https://www.youtube.com/watch?v=uqeqnF7Clr8&ab_channel=OxfordUnion)

Figure 13. 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 contains 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 through the e-KRP. Figure 14 below shows such a spreadsheet (namely, the spreadsheet with the metadata values of the podcasts available through 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)⁷⁵. Finally, the bulk of the 121 objects classified into the “other” category has been used as a pool of back-up data objects to consider in cases where the numbers of data objects per type listed in Tables 10 to 17 could not be reached.

#	title	description	keywords	creator	contentLanguage	contentLocation	fileURL	topic	purpose	format	type	fileSize	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/show/510...	animal husbandry and welfare practices; animal husbandry and welfare actions; 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/show/510...	N.A.	communication; dissemination	mp3	interview	14.2 MB	2020-06-19
3	AgriforValor: Exploring 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/show/510...	crop farming practices; crop farming actions; crop farming systems; crop farming tools; crop farming digital technologies; socio-economic context; forestry practices; forestry actions; forestry systems; forestry infrastructures; forestry tools; forestry digital technologies; climatic zones; environment; socio-economic context	communication	mp3	interview	28 MB	2020-05-19
4	Agrifor: Space for Innovations in Agriculture	A brief description of Agrifor Thematic Network and its activities.	farming practice; best practice; innovation support service; AKIS	EURAKNOS	English	N.A.	https://open.spotify.com/show/510...	crop farming practices; crop farming actions; crop farming systems; crop farming tools; crop farming digital technologies; socio-economic context	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/show/510...	crop farming practices; crop farming actions; crop farming systems; crop farming tools; crop farming digital technologies; socio-economic context	communication	mp3	interview	15.9 MB	2020-06-15
6	CEVs: Overcoming the challenges of Covid	Maria Genart from IFoAM organization describes her cross-exchange visit to EURAKNOS project and how it can be successfully continued during coronavirus period.	N.A.	EURAKNOS	English	N.A.	https://open.spotify.com/show/510...	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 focused on the ON-Net Ecoland Thematic Network knowledge platform, is presented by Maria Genart.	N.A.	EURAKNOS	English	N.A.	https://open.spotify.com/show/510...	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 Pilecynski.	agricultural production system; farming practice; farming equipment and machinery; animal husbandry and welfare; pest / disease control	EURAKNOS	English	N.A.	https://open.spotify.com/show/510...	animal husbandry and welfare practices; animal husbandry and welfare actions; animal	communication	mp3	interview	8.3 MB	2020-04-02

Figure 14. 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)

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

The phase in our methodology that has been the successor of the assembly of the input for the e-KRP repository is the one concerned with the integration of the annotated data objects into the repository. This has been the last phase of the methodology related to shaping up the e-KRP repository. To execute the integration, we had to consider the characteristics and particularities of DSpace⁷⁶ (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.

This 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 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.

The process of harvesting data objects from the KRs of TNs and initially storing them into a local storage space is shown in Figure 15 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 sample TNs. The data objects have been retrieved from the KRs of the TNs and transformed to become DSpace compliant. The TN data objects have been initially stored in a local storage space reachable by DSpace.

⁷⁵ This has been the case with documents.

⁷⁶ 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).

As a preparatory step to loading all the data objects into DSpace, they have all first been converted to a proprietary Simple Archive Format⁷⁷ 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 file containing descriptive metadata. After the loading of the data objects in DSpace, a checksum⁷⁸ 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.

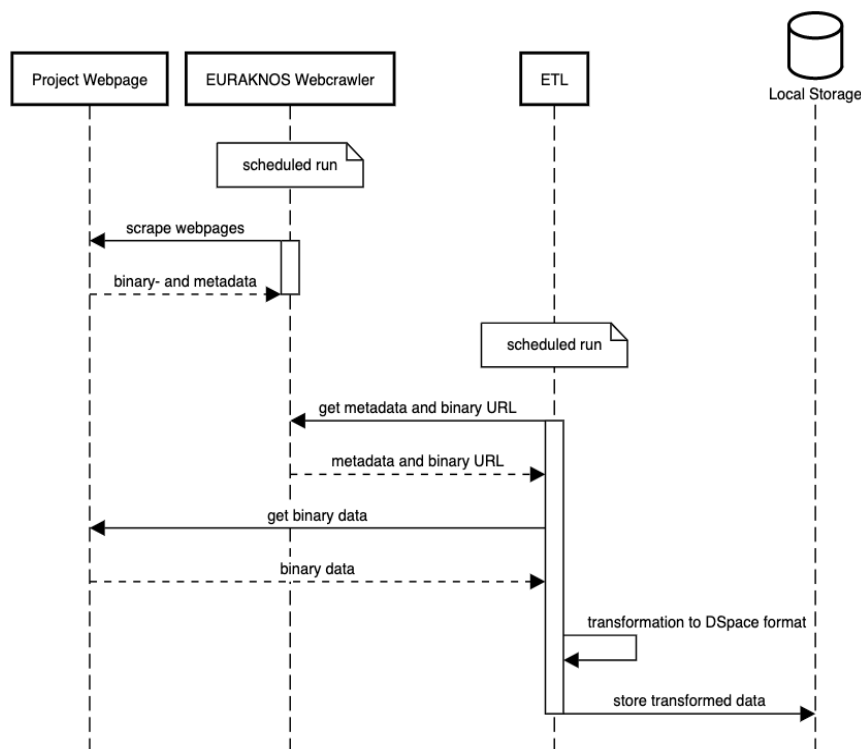


Figure 15. Sequence diagram displaying the ETL process for harvesting data objects from the KR of a TN and storing them to local storage

After having stored all data objects locally in the desired format, the ingestion into DSpace took place as shown in Figure 16. 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.

The total number of data objects finally integrated into DSpace (the e-KRP repository) was nearly 400. The difference between the number of data objects initially considered as input for the repository after the data selection process (631) and those finally integrated into DSpace (~400), is attributed to issues mentioned at the end of Subsection 5.3.3.3 (“The process of the annotation of data objects with metadata values”) and have related to the identification of faulty items. Some more faulty items were also detected during the implementation of the integration process.

⁷⁷ <https://wiki.lyrasis.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format>

⁷⁸ 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.”

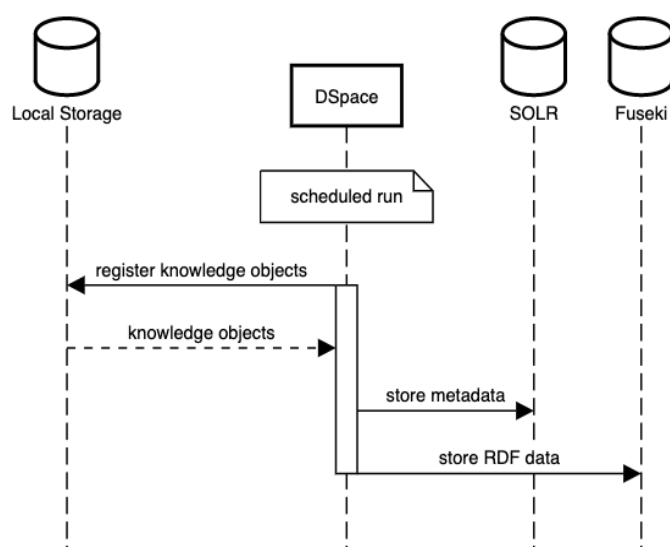


Figure 16. Sequence diagram illustrating the registration of data objects and their metadata into DSpace

As shown in Figure 17, 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.

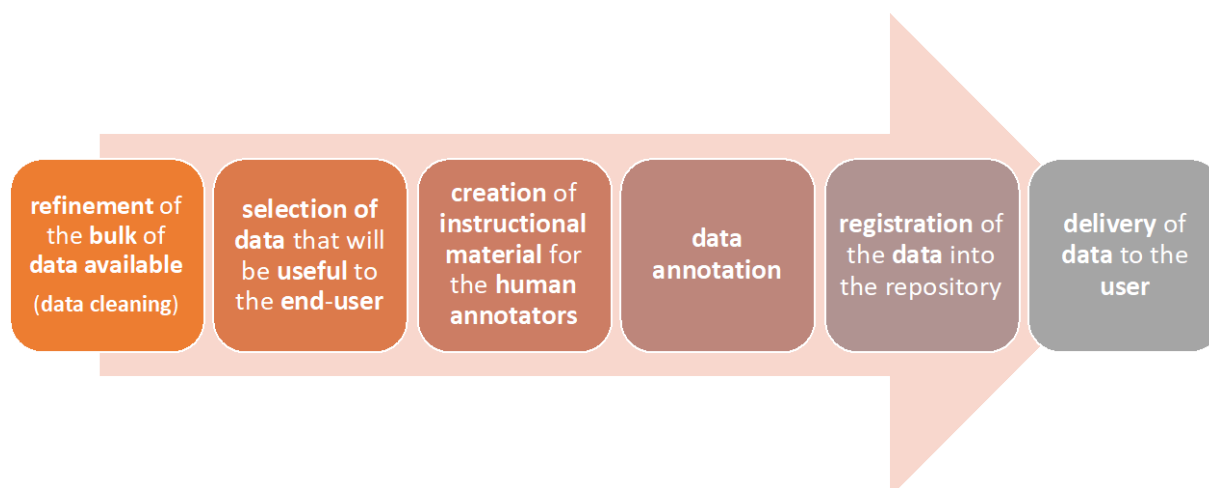


Figure 17. The integration and registration of the TN-related data objects (and their metadata) into the e-KRP repository as a final step before making them available to the end-users

5.5. Evaluating the FAIRness of the e-KRP

In the past few years, the FAIR data principles (Wilkinson et al., 2016) have become prevalent within the context of data published on the web by providing 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 the enforcement of a mandatory use of them, but rather to provide some guidelines to the entities generating data, in any domain, to make this data 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)⁷⁹ is such an assessment tool, developed by the Data Archiving and Network Services (DANS)⁸⁰. 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 23 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 23. 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?	<ul style="list-style-type: none"> 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?	<ul style="list-style-type: none"> Controlled vocabularies. Taxonomies (thesauri). Ontologies. There are no standards for my discipline. 	Controlled vocabularies.	
	Did you provide rich and detailed additional documentation?	<ul style="list-style-type: none"> 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?	<ul style="list-style-type: none"> Yes No I can't find this information in EASY. 	Yes	50%
	Does your dataset contain personal data?	<ul style="list-style-type: none"> Yes No 	Yes ⁸¹	
	Which of the usage licenses provided by EASY did you choose to comply with the access rights attached to the data?	<ul style="list-style-type: none"> Open access (CC0). Restricted access. Embargo. Other access. I don't know which of these licenses apply to my data. 	Other access. ⁸²	
INTEROPERABLE	Are the data in your dataset stored in preferred formats?	<ul style="list-style-type: none"> 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?	<ul style="list-style-type: none"> Yes, and these are online accessible. Yes, but they are not online accessible. No. 	No	

⁷⁹ <https://dans.knaw.nl/nl>

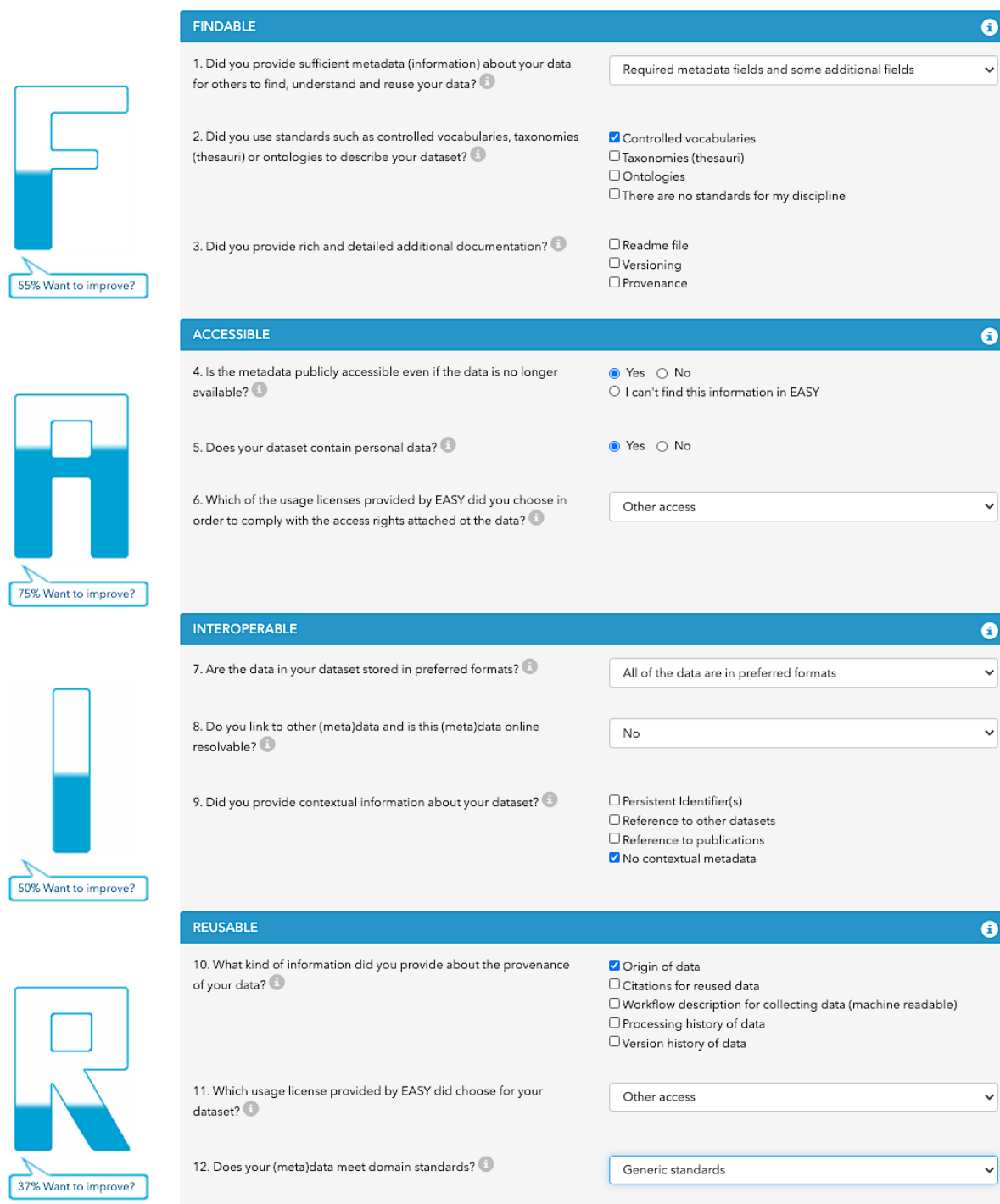
⁸⁰ <https://satifyd.dans.knaw.nl/>

⁸¹ 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.

⁸² 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").

	Assessment item	Potential responses	Response provided	Score
	Did you provide contextual information about your dataset?	<ul style="list-style-type: none"> • 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?	<ul style="list-style-type: none"> • 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?	<ul style="list-style-type: none"> • 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?	<ul style="list-style-type: none"> • 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 from Table 23 above, the total FAIRness score of the e-KRP has been evaluated to be equal to 54%. This score is promising, especially if we consider the low degree of compliance with the FAIR principles achieved by other data repository systems (Dunning et al., 2017.). This platform-related FAIRification effort has already been taken up in EUREKA in an attempt to further advance the level of adherence to the FAIR principles. Figure 18 below shows the responses provided as part of the e-KRP's FAIRness evaluation with the help of the SATIFYD tool and the evaluation scores that have been finally achieved.



FINDABLE

1. Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data? *i*

Required metadata fields and some additional fields

2. Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset? *i*

☒ Controlled vocabularies
☐ Taxonomies (thesauri)
☐ Ontologies
☐ There are no standards for my discipline

3. Did you provide rich and detailed additional documentation? *i*

☐ Readme file
☐ Versioning
☐ Provenance

ACCESSIBLE

4. Is the metadata publicly accessible even if the data is no longer available? *i*

☒ Yes ☐ No
☐ I can't find this information in EASY

5. Does your dataset contain personal data? *i*

☒ Yes ☐ No

6. Which of the usage licenses provided by EASY did you choose in order to comply with the access rights attached to the data? *i*

Other access

INTEROPERABLE

7. Are the data in your dataset stored in preferred formats? *i*

All of the data are in preferred formats

8. Do you link to other (meta)data and is this (meta)data online resolvable? *i*

No

9. Did you provide contextual information about your dataset? *i*

☐ Persistent Identifier(s)
☐ Reference to other datasets
☐ Reference to publications
☒ No contextual metadata

REUSABLE

10. What kind of information did you provide about the provenance of your data? *i*

☒ Origin of data
☐ Citations for reused data
☐ Workflow description for collecting data (machine readable)
☐ Processing history of data
☐ Version history of data

11. Which usage license provided by EASY did choose for your dataset? *i*

Other access

12. Does your (meta)data meet domain standards? *i*

Generic standards



Your data is **54% FAIR** in EASY

Figure 18. Evaluation of the EURAKNOS e-KRP's FAIRness with the help of the SATIFYD tool

6. 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 that have taken place in the context of Task 4.2. It has helped to bring all the components of the EURAKNOS e-KRP technology stack's together, and make them "talk" to each other, to implement the system architecture presented and explained in Section 4 ("The EURAKNOS e-KRP system architecture"). Based on that, the aim of the present section is not to dive deep into the technical details, but rather provide an overview of the work that has taken place both in the foreground and background of the



platform to make it a reality. To this end, we begin with a short explanation of what the back-end development is about, and what it has involved in EURAKNOS and the e-KRP development (Subsection 6.1). In the same way, the description of the front-end part of the development work is presented and explained in Subsection 6.2 (“Front-end development”).

6.1. 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-user⁸³. In the case of the e-KRP this communication establishment task has been “assigned” to Drupal⁸⁴. 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)⁸⁵. 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 the content delivery and presentation. Support in Drupal-based development can be sought and largely provided by the respective community, which is one of the largest open-source communities in the world.

6.2. Front-end development

Front-end development is the term used in the IT community to refer to the “client-side”-related work done in the context of a software application’s development⁸⁶. It relates to the software 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 navigational structures, in general, are part of the solutions that can be provided (and thus, engineered in an application) to structure (and facilitate) access to content. All these issues are dealt with in the context of Information Architecture, which is part of the User Experience design and front-end development.

To develop a visually attractive and user-friendly (from an Information Architecture point of view) platform, we have relied upon the recommendations reported in Deliverable 3.2⁸⁷, as well as the evidence related to user-needs that have been reported in Deliverable 4.1⁸⁸. The design of the e-KRP’s user interface has also been instructed by the employed Information Architecture illustrated in Figure 18 below. The elements appearing in this design and their connections have served as the input needed for the development of the navigational structures of the e-KRP (e.g., menus) and functionalities such as the search result filtering options provided to the user. Further details on how these elements have been incorporated into the e-KRP’s interfaces (and thus, delivered to the end-users), together with references to the respective screens, are available in Section 10 (“e-KRP user guide”). The Information

⁸³ <https://www.coursereport.com/blog/front-end-development-vs-back-end-development-where-to-start>

⁸⁴ <https://www.drupal.com/>

⁸⁵ <https://www.gnu.org/licenses/gpl-3.0.html>

⁸⁶ <https://www.coursereport.com/blog/front-end-development-vs-back-end-development-where-to-start>

⁸⁷ 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.

⁸⁸ 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.



Architecture diagram illustrated in Figure 18 helps make explicit the structure of the e-KRP and the navigation levels employed to enable access to all the 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 (namely, the search results can be displayed to the end-user with the help of 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 to⁸⁹ (namely, “Crop farming”, “Forestry”, “Livestock”, “Environment”, “Society”, and “Economics”) and the e-KRP 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).

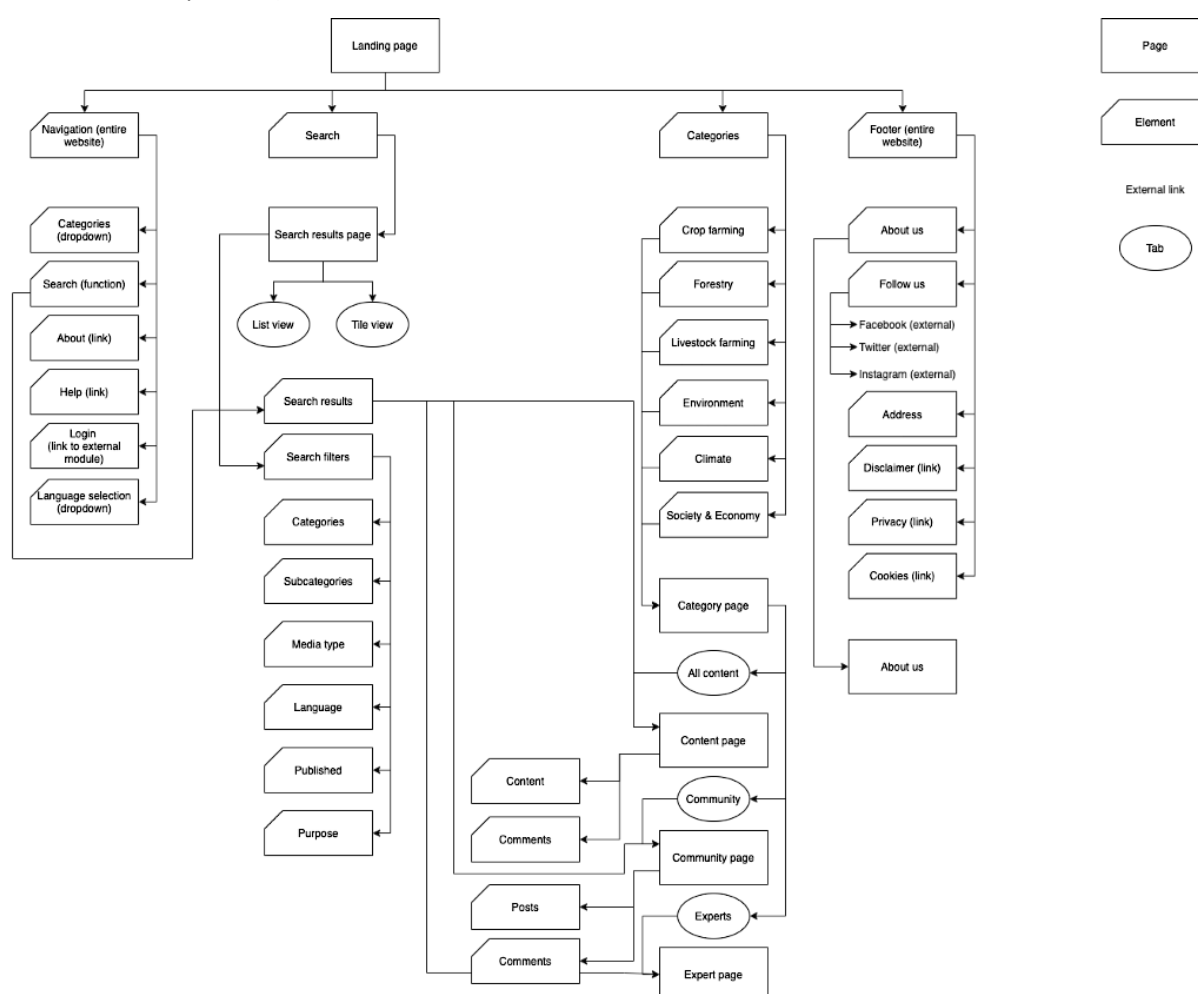


Figure 19. The EURAKNOS prototype e-KRP's Information Architecture

Although in EURAKNOS there was not a consideration for extensive user needs analysis and testing processes, input and feedback from potential e-KRP users was collected in the context of specially designed sessions in the EURAKNOS consortium meeting in Athens⁹⁰ and the Paris workshop⁹¹. Details

⁸⁹ 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).

⁹⁰ The Athens consortium meeting of EURAKNOS took place on the 17th and the 18th of June, 2019.

⁹¹ The Paris workshop of EURAKNOS took place from the 11th to the 13th of December, 2019.

about the feedback collected from those sessions are given in Subsection 7.1 (“Results obtained from preliminary user tests”). Overall, the design and development of the e-KRP’s front-end followed a procedure that started with the design of wireframes and the collection of feedback on them. 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-end⁹². To facilitate this process and the execution of all its steps, meetings among the project coordination team, the back- and front-end developers, as well as UX⁹³ researchers and designers were organised and held regularly.

From a technical perspective, the framework that has been used for the e-KRP front-end development needs is VueJS⁹⁴; a well-known front-end JavaScript framework for building user interfaces and single-page applications. Besides, NuxtJS⁹⁵ has also been used to add extra features around VueJS. Both are open-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 (3) things, namely: HTML, CSS, and JavaScript (able to be interpreted by a browser and displayed on the end-user’s screen).

7. Testing the EURAKNOS e-KRP

Processes of thorough user needs analysis and testing of both preliminary and close-to-final product versions form integral parts of any software system’s development, whether the traditional waterfall model is employed or a more intensive and iterative, agile-based approach is being favoured. As far as the collection of input concerning the needs of end-users in EURAKNOS, there has not been any such extensive process. However, insights into the actual needs of potential e-KRP end-users have been managed to be obtained from various sources. Such sources were the results of the EURAKNOS workshop in Paris (11 to 13 December, 2019), which are extensively reported in Deliverable 3.1⁹⁶ (“Report on HIKR methodologies and practices for data generation, selection, collection and storage”) (Marois et al., 2020), as well as the content-related recommendations provided in the Deliverable 3.2 report (“Report on HIKR design”) (Van Hal et al., 2020) also based on some of the findings documented in Deliverable 4.1 (“Report on added value and feasibility of e-KRP”) (Gernert et al., 2019). The insights obtained from these sources proved to be useful for the e-KRP’s design (and development), but rather more at the conceptual level and less to the UX engineering and design.

Concerning the collection of feedback from platform-related tests, such input has been feasible to be gathered from the wireframe-based exercise undertaken in the EURAKNOS consortium meeting in Athens (17-18 June, 2019) and the insights provided by representatives of the AKIS community in the Paris workshop’s Marketplace session. The compilation and processing of the pieces of the gathered information have helped to acquire useful insights into the actual needs and preferences of the e-KRP’s end-users and thus, move from the more abstractly designed wireframes to the closer-to-reality mock-up screens, which have finally served as the blueprints for the development of the front-end and the engineering of the User Experience. Details of this input/feedback collection are provided in Subsection 7.1 (“Results obtained from preliminary user tests”) below.

At this point, it needs to be, however, stressed that there is a twofold approach in the way that testing has been conceptualised in EURAKNOS. So, apart from the insights and feedback that has been feasible to be collected from potential users, testing is also considered in terms of a comparison of the e-KRP

⁹² 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).

⁹³ UX is the abbreviation used to refer to User Experience.

⁹⁴ <https://vuejs.org/>

⁹⁵ <https://nuxtjs.org/>

⁹⁶ For details, the reader may refer to Section 3 (“WG 3.1: participative workshops and the main results and inputs collected”) in the specific report.

with other systems aiming to address similar needs relating to the facilitation of access of ready-to-apply agricultural knowledge. To this end, the outcomes of a semi-structured comparison between the EURAKNOS e-KRP and the EIP-AGRI database⁹⁷ are described in Subsection 7.2 (“Testing the e-KRP against other existing systems”).

7.1. Results obtained from preliminary user tests

7.1.1. Feedback 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 set off the e-KRP’s front-end design and development process, a preliminary user test exercise was held in the EURAKNOS consortium meeting in Athens, in June 2019. The exercise was based on the use of wireframes designed and used to collect the consortium partners’ feedback on design aspects important to consider for the e-KRP’s front-end. These aspects have been the following:

- **Navigation**

The solution proposed about the e-KRP’s basic navigational structure was that of a sidebar (Figure 20 below), because of its flexibility and potential to easily expand to offer more navigation options. A menu on the top of the page was also suggested. However, the horizontal space required for the display of the menu’s labels may impose changes in the case that a new option needs to be added.

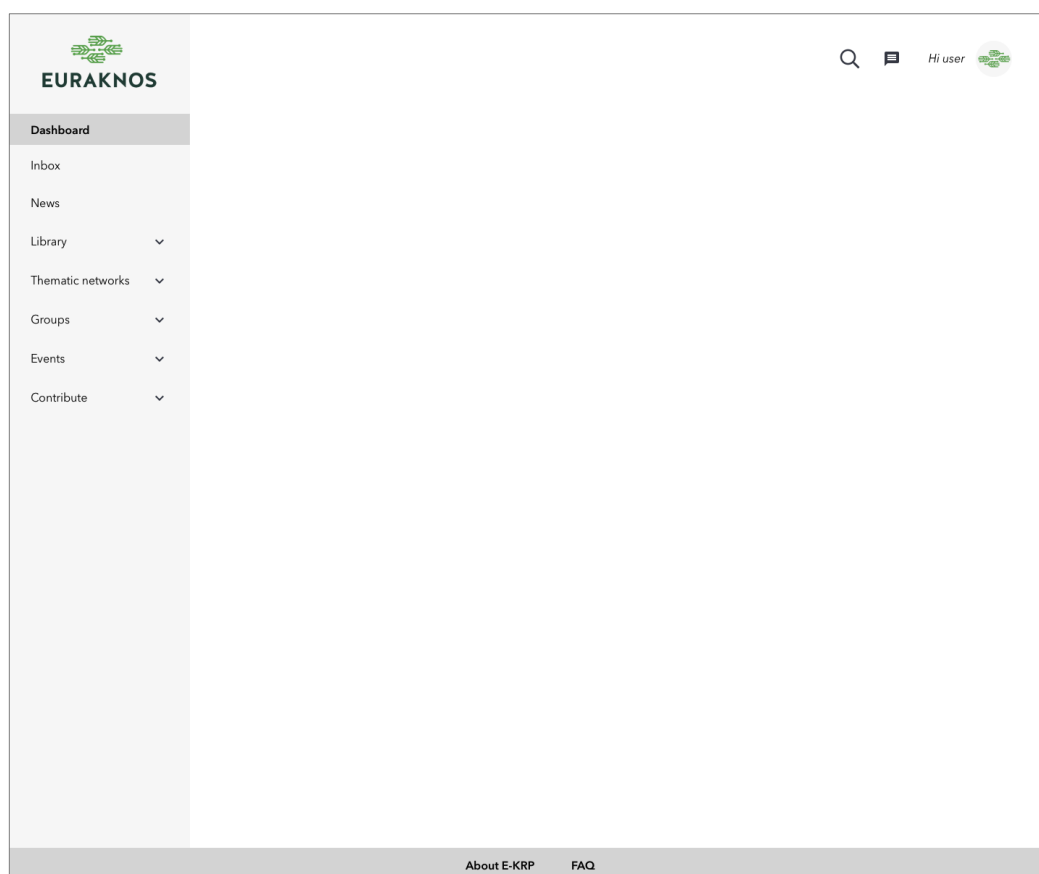


Figure 20. The navigational structures initially proposed for the e-KRP through the design of the respective wireframe

- **Dashboard**

⁹⁷ https://ec.europa.eu/eip/agriculture/en/find-connect/projects?search_api_views_fulltext=&field_proj_funding_source_list=4

In general, a dashboard is a type of GUI⁹⁸ that often provides at-a-glance views of key performance indicators (KPIs), progress reports, or data visualisations⁹⁹. For the needs of the user test exercise, four (4) different wireframe versions for the e-KRP's landing page dashboard were prepared:

- **Dashboard version #1** (Figure 21): A classic dashboard of categories where the user can view updates for each one. In the case of a few category updates, this will be very noticeable. On the other hand, many updates in a specific category will flood that category.
- **Dashboard version #2** (Figure 22): The second dashboard version has been inspired by the well-known Facebook feed. This is quite a dynamic dashboard, where the visitor can view updates by scrolling down.
- **Dashboard version #3** (Figure 23): The third version of the e-KRP landing page's dashboard was more dynamic in layout. It resembled the layouts of online newspapers and gave the impression that there is always something new to discover.
- **Dashboard version #4** (Figure 24): The fourth dashboard version is less dynamic and focuses on textual content supported by visual aids.

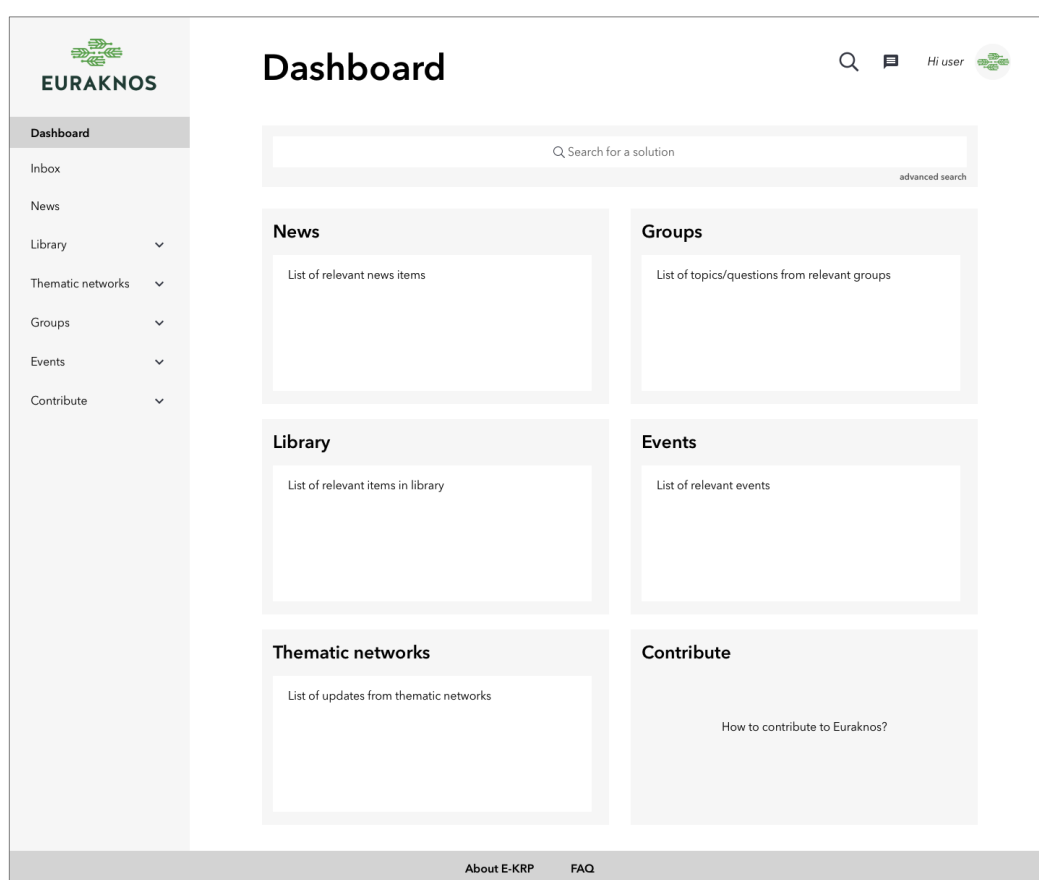


Figure 21. Version 1 of the e-KRP's landing page dashboard

⁹⁸ GUI is the abbreviation used for Graphical User Interface.

⁹⁹ [https://en.wikipedia.org/wiki/Dashboard_\(business\)](https://en.wikipedia.org/wiki/Dashboard_(business))

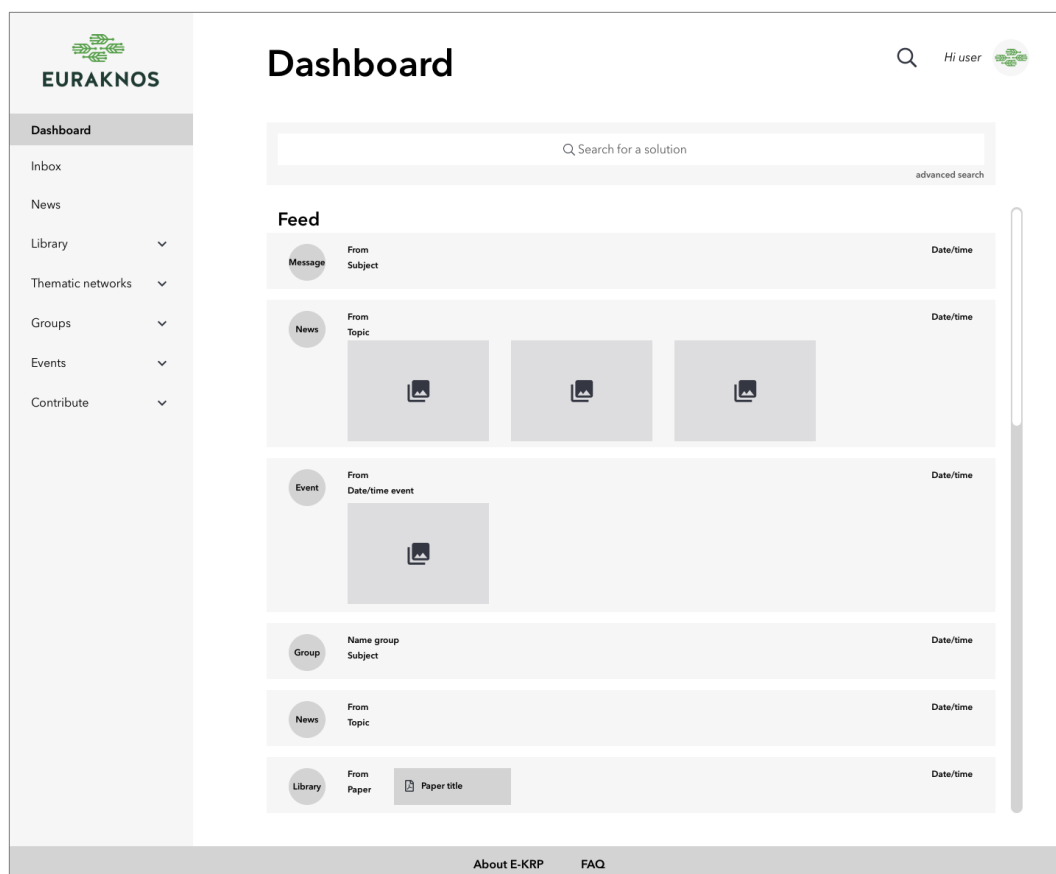


Figure 22. Version 2 of the e-KRP's landing page dashboard

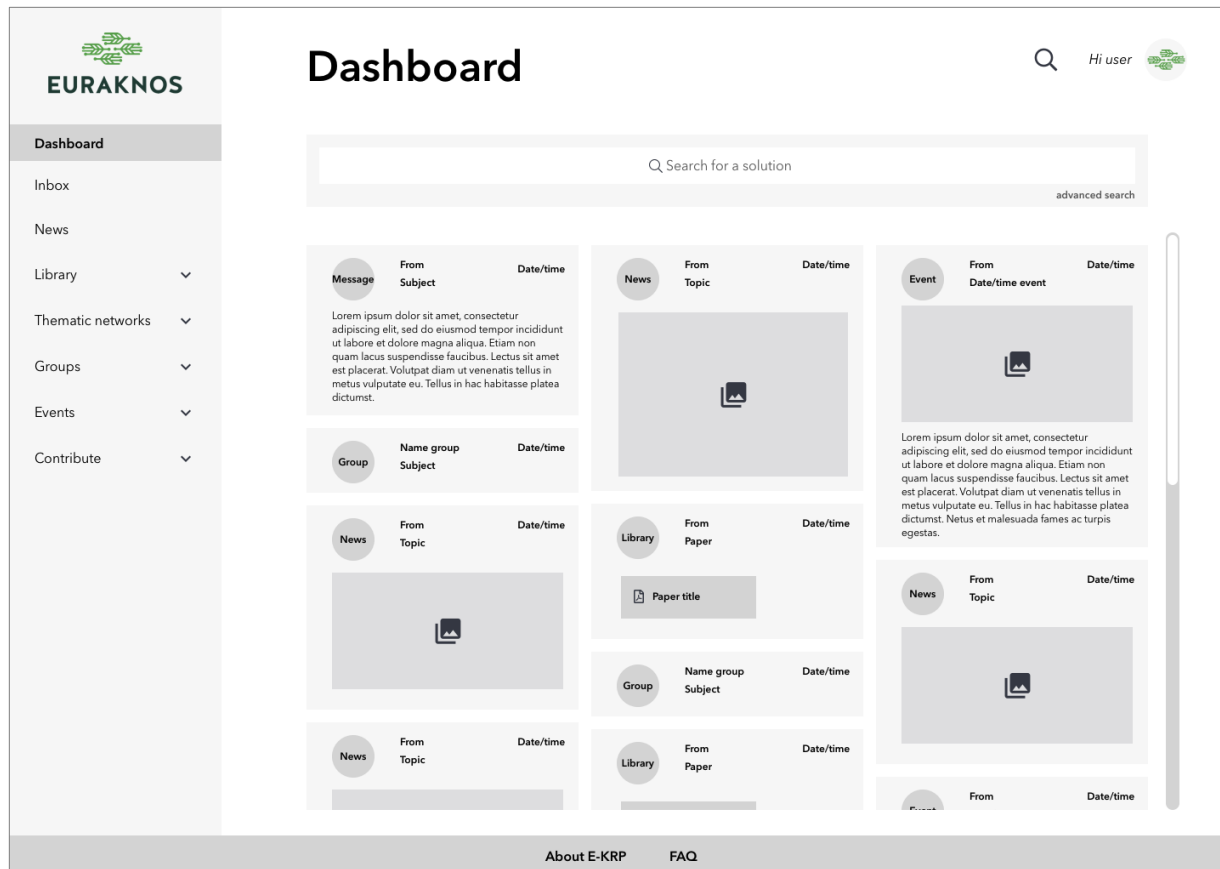


Figure 23. Version 3 of the e-KRP's landing page dashboard

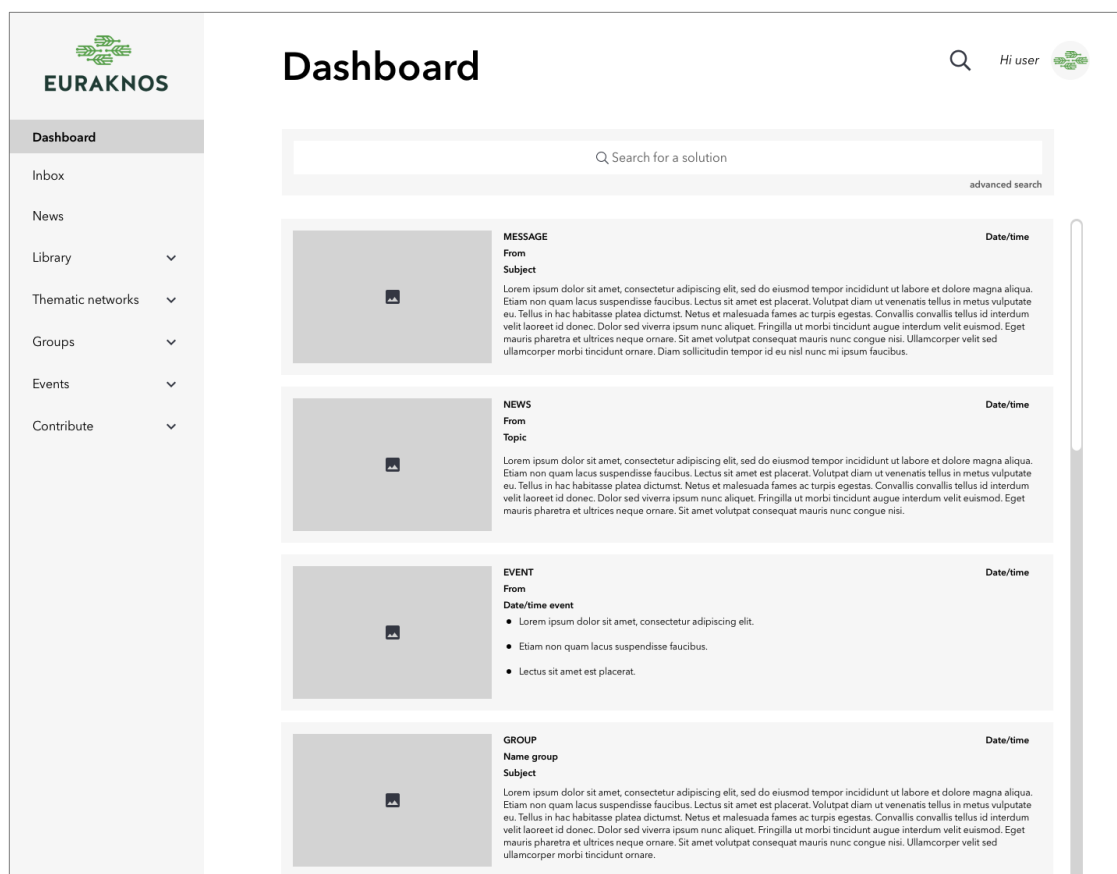


Figure 24. Version 4 of the e-KRP's landing page dashboard

- **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 25).

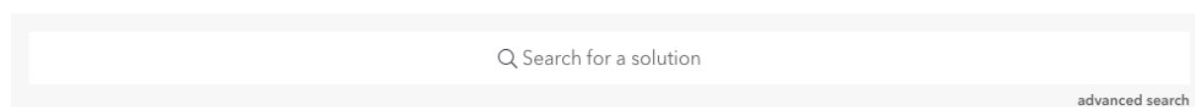


Figure 25. Search function provided in every e-KRP page with the aim to facilitate content and information search

- **News article**

Two versions of news article – related views were created; namely, a view with a picture and a view without a picture illustrated with the help of Figures 26 and 27 below.

- **Library overview**

The library metaphor has been adopted for the design of the e-KRP pages responsible for the display of the content and information stored in the repository. The library metaphor has been perceived in a way similar to that of an online academic library, where the focus is on different ways to present the overview of the content. Three (3) different library overview versions have been proposed with the help of the wireframes shown in Figures 28, 29, and 30.

- The first version allows a more graphical overview and fits best the needs of users for navigating around the content (Figure 28).
- The second version is very much oriented towards researchers scrolling through textual content, looking for something specific, who don't want to be distracted by graphical content (Figure 29).
- The third version is a combination of the previous versions and is best suited for users preferring to scroll vertically while still getting graphical content (Figure 30).

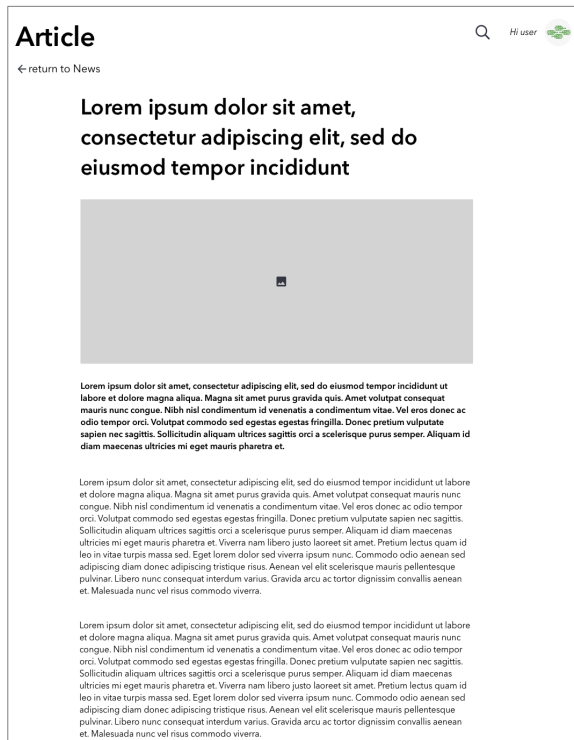


Figure 26. Wireframe of news article with a picture



Figure 27. Wireframe of news article without a picture

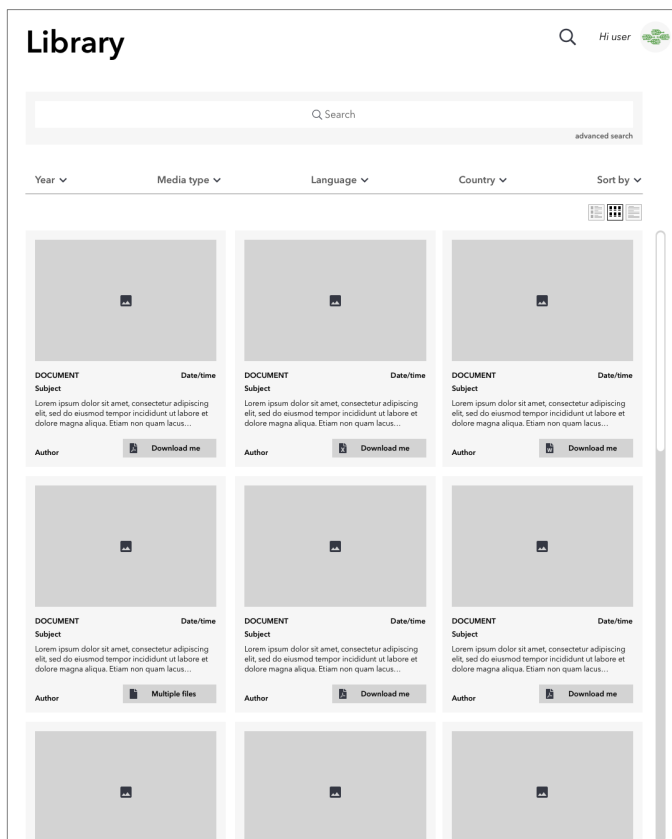


Figure 28. Version 1 of the library overview

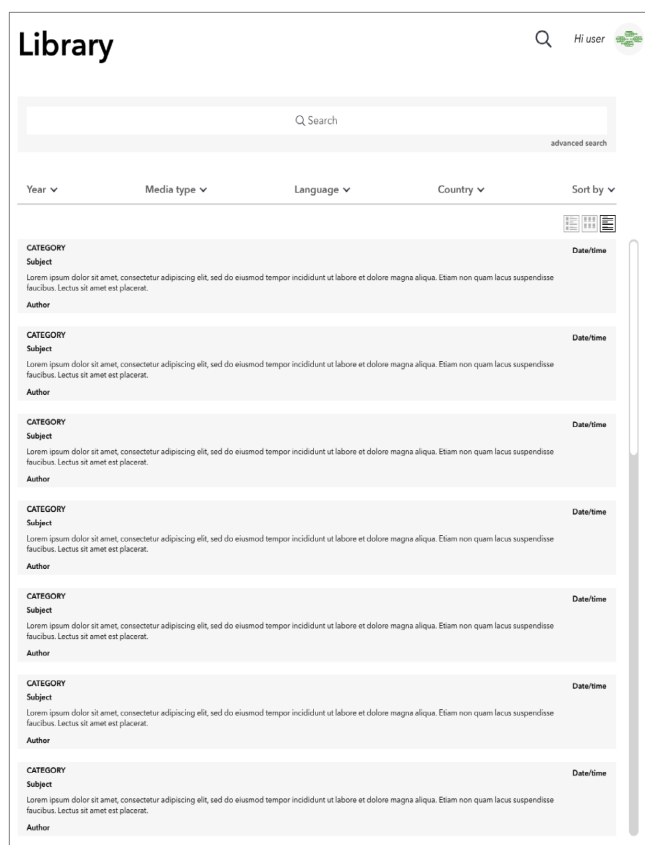


Figure 29. Version 2 of the library overview

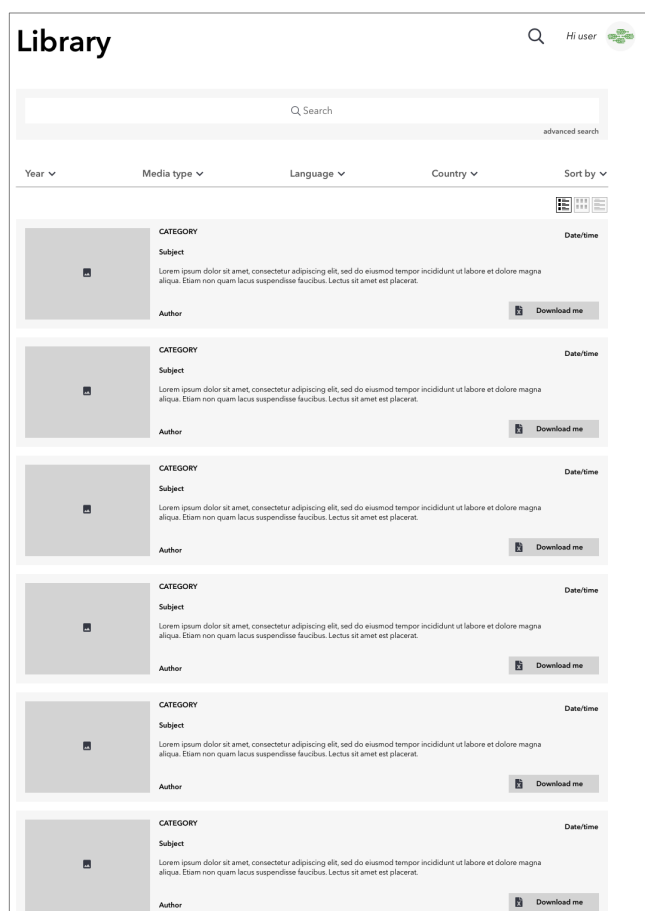


Figure 30. Version 3 of the library overview



- **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 option was the one related to the delivery of an academic article (Figure 31). Wireframes for audio-visual content (namely, podcast and video) were also developed and used for the needs of the user test exercise (see Figures 32 and 33).

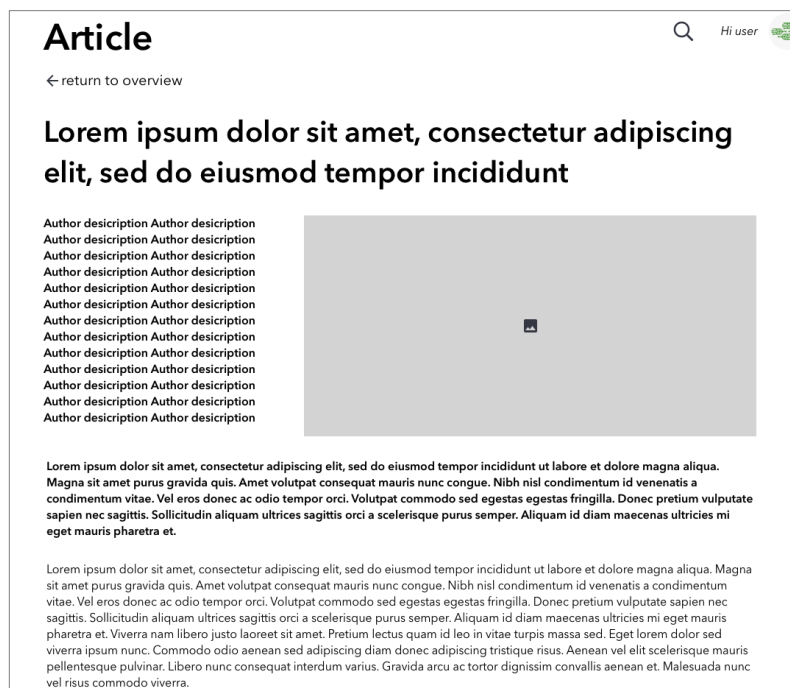


Figure 31. Wireframe of academic article view

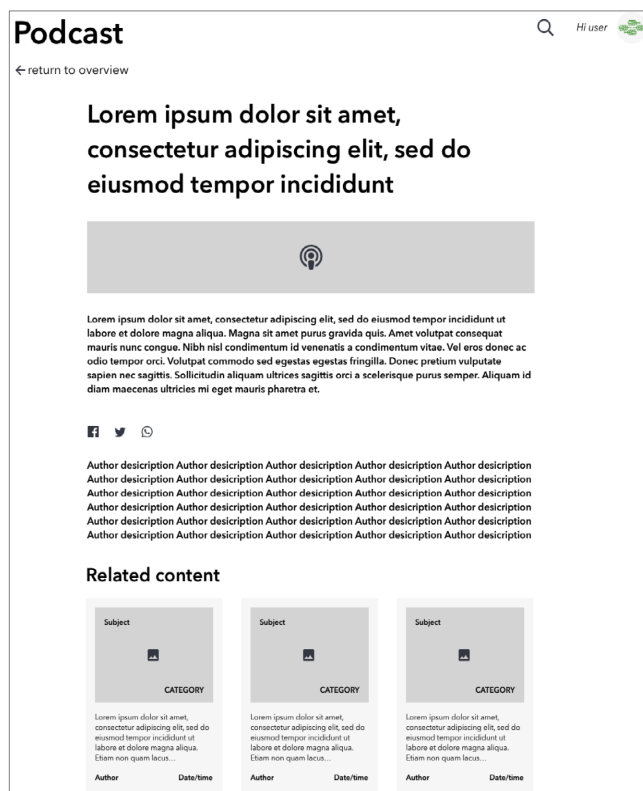


Figure 32. Wireframe of podcast view and access

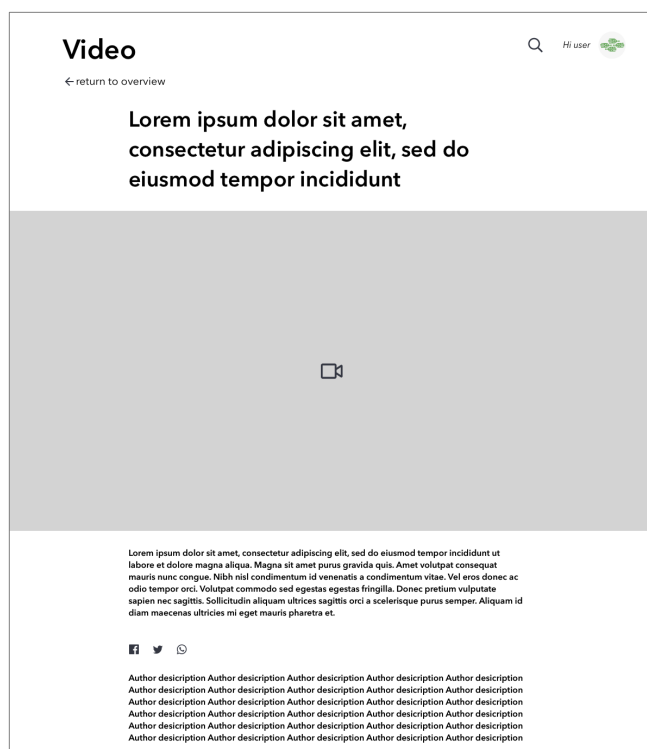


Figure 33. Wireframe of video view and access

The exercise participants (partners of the EURAKNOS consortium) were assigned to three (3) groups and asked to review the wireframes by providing comments and suggesting design-related alternatives depending on their comments. The results that were obtained from this exercise are listed below:

- **Dashboard**
A clear preference towards highly structured dashboards, supported by fit-for-purpose visual aids, was reported.
- **News articles**
There has been a clear preference towards news content accompanied by a short text and graphical material.
- **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.

7.1.2. Feedback from the Marketplace session held in the Paris workshop

The wireframes, which were used for the initial user-testing-related exercise in the Athens consortium meeting, served also as a basis for discussion in the context of a similar exercise held in the EURAKNOS Paris workshop. This user-test exercise took place as part of a Marketplace session, where potential e-KRP users (namely, farmers, advisors, and researchers) provided feedback concerning the work done for the e-KRP development, based on the wireframes and the basic platform concept. The feedback/ results obtained from this exercise, which have proven to be useful also through a cross-check with the results of the relevant Athens consortium meeting's exercise, are listed below:

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

The points listed above to summarise the input/feedback that has been able to be collected uncover a strong need for a system that: (i) can be easily searched and found through the use of general-purpose search engines; (ii) can be used in an easy/intuitive way for content and information access purposes; (iii) enables the search and retrieval of information in English and other languages; (iv) allows the user to provide feedback; and (v) provides personalisation and community building/strengthening features. In a nutshell, it not only allows for access to content but also facilitates community building processes.

7.2. Testing the e-KRP against other existing systems

In EURAKNOS, the testing of the e-KRP has not only been considered for the supply of feedback, by potential end-users, on various prototype versions, but also through having the e-KRP compared to other platform systems oriented towards serving similar needs of the agricultural community. Another system with a similar scope is the EIP-AGRI database accessed through the EIP-AGRI service point's website¹⁰⁰. Based on that, the current subsection aims to present the outcomes of a semi-structured comparison that has taken place between the EURAKNOS prototype e-KRP (being further developed in EUREKA) and the EIP-AGRI database. This comparison exercise has focused on the following aspects: (i) breadth of the content and information delivered; (ii) information architecture; and (iii) employment of visual elements and aesthetics.

7.2.1. 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, and made available in a range of formats, which should be preserved and further disseminated. Given the fact that most of the TNs' KRIs cease to exist after the end of the project,

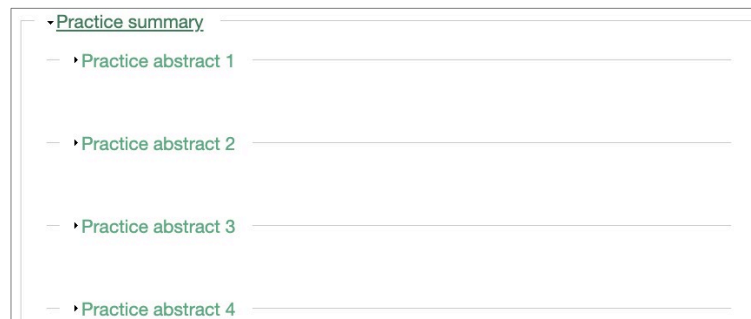


Figure 34. Access to the Practice Abstracts created by a TN from the EIP-AGRI database

there is a need for a centralised repository system able to collect all this available knowledge (packaged into data objects of various format categories and types) and further distribute it to the 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 allows to access content/information mainly available in a textual format. More specifically, it provides direct access to the Practice Abstracts¹⁰¹ created by the TNs (see Figure 34 above).

¹⁰⁰ <https://ec.europa.eu/eip/agriculture/en/node>

¹⁰¹ 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.

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 farmers. In order to participate or to be informed, contact us via our website!

Figure 35. TN details available from the EIP-AGRI database

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



As illustrated in Figures 35 and 36 above, pieces of information provided by the EIP-AGRI database are background information about the TN, the coordinating organisation, as well as the list of the partners involved. Access to other types of textual content/information (i.e., documents) is provided indirectly through links to the TN's KR¹⁰². Access to audio-visual material is also feasible (see Figure 37 below).

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

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 easy search and access to a breadth 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, presentations). Based on that, and as shown in Figure 38 below, the data object categories delivered from the e-KRP are:

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

¹⁰² In the case that a TN's KR ceases to exist, then access to its content will also cease to be available. This is why a centralised repository system is needed.



Figure 38. Access to a breadth of content and information, in a wide range of file formats, from the EURAKNOS prototype e-KRP

7.2.2. Information Architecture

Differences between the EIP-AGRI database and the EURAKNOS prototype e-KRP also exist concerning the Information Architecture employed in these systems. In the case of the EIP-AGRI database, a more complex navigational structure is used, as made evident from its menus, mostly because the EIP-AGRI database is part of a broader system (the EIP-AGRI service point) aiming to deliver information/content that goes beyond the a TN's results. As shown in Figure 39, the menu available at the top of every page of the EIP-AGRI service point's website is both broad and deep. This fact may have implications for the ease of use of the system and the potential to reach the content and information needed in just a few clicks.

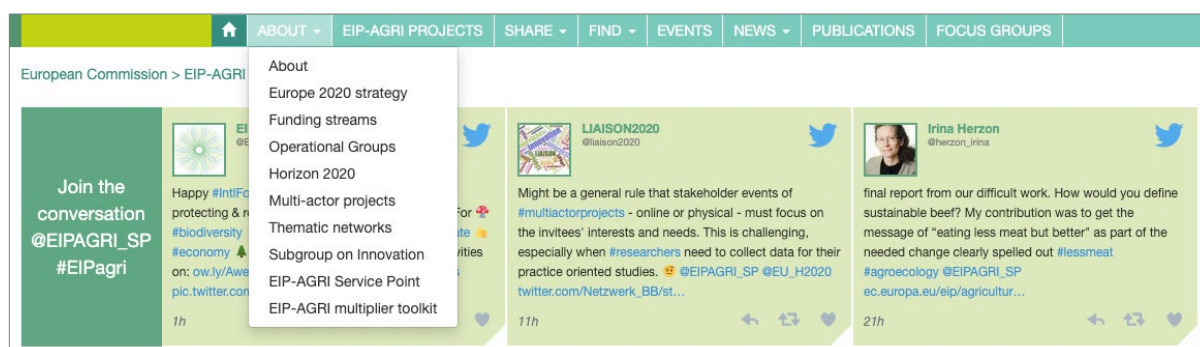


Figure 39. The EIP-AGRI service point website's main navigational structure

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 40 below).

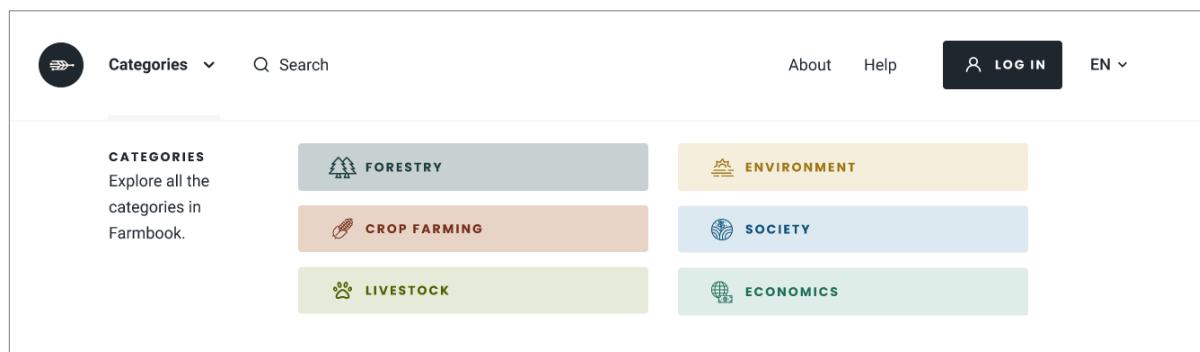


Figure 40. The EURAKNOS e-KRP's main navigational structure

As far as the execution of search operations in the two systems is concerned, a dedicated search area is made available to the user in the case of the EIP-AGRI database. Even though no visual aids are used to demarcate this area in the wider context of the respective 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: (i) to the use of logical operators ("AND", "OR", and "NOT") to make explicit how the terms of a search query should be used; (ii) the selection of the funding scheme that the projects (to be retrieved as search results) should be associated with; and (iii) the geographical region to which the search results need to relate to. Even though the options made available in the EIP-AGRI database can indeed facilitate the 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 from the EIP-AGRI database is illustrated in Figure 41 below.

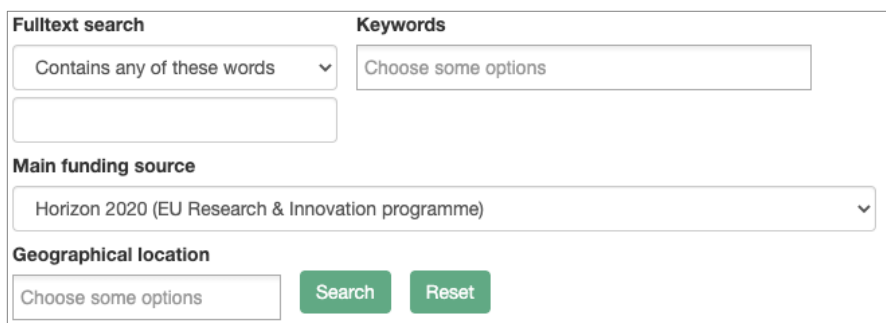


Figure 41. The search function available to search for content and information in the EIP-AGRI database

In the case of the EURAKNOS e-KRP, more straightforward prompts and tools are used for the needs of executing search operation. More specifically, any search for content and information may be done via the submission of a search query consisting of any number of search terms of the preference of the user.

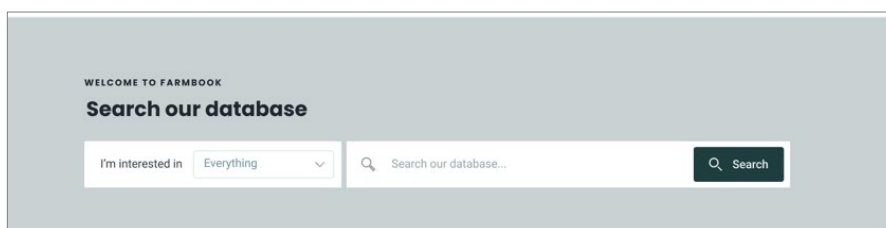
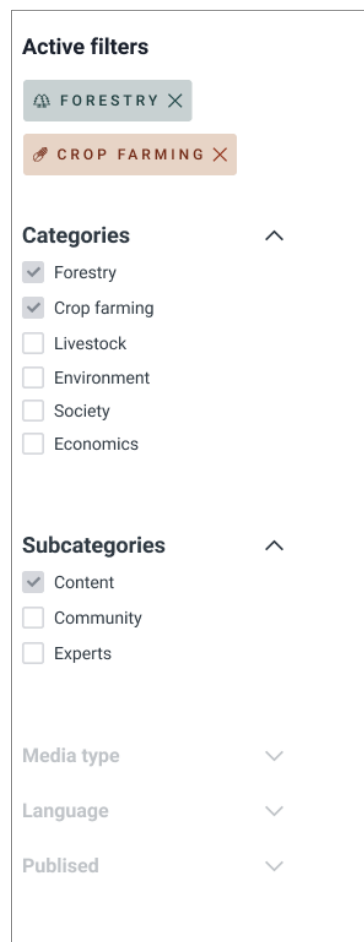


Figure 42. The search function of the EURAKNOS e-KRP

The search function of the e-KRP is very close to those available in state-of-the-art search engine tools and thus, very easy to be used (see Figure 42 above). The refinement of the search results can be made by using some filtering options available to the end-user (see Figure 43). As made evident from Figure 42, the search function of the e-KRP is demarcated (and therefore, can be easily spotted) and straightforward to be used without any additional prompts needed to be provided.

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 44 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 61 and 62 in Section 10 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. However, another approach is being followed in the EURAKNOS e-KRP. Specifically, the end-user can access content and information that has been clustered around agricultural sectors (i.e., “Crop farming”, “Forestry”, and “Livestock”), as well as broad, cross-sectoral themes (namely, “Environment”, “Society”, and “Economics”). This way of information and content organisation appears to be more straightforward and closer to the needs of the end-users.



Active filters

FORESTRY X

CROP FARMING X

Categories ^

☒ Forestry

☒ Crop farming

☐ Livestock

☐ Environment

☐ Society

☐ Economics

Subcategories ^

☒ Content

☐ Community

☐ Experts

Media type v

Language v

Published v

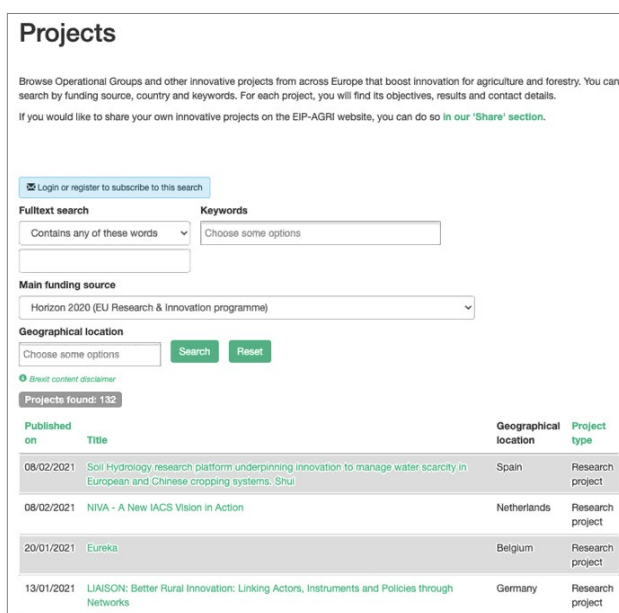
Figure 43. Filtering options provided in the e-KRP to elaborate the search results

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

7.2.3. 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 identity enables the user to have an overview of the sector or theme to which the content and information accessed, or other features and functionalities, relate to. In the case of the EIP-AGRI database, there has been less attention to the aesthetics part. The user interfaces are dominated by textual information, which makes the system informative but not that user-friendly. The difference between the two systems' aesthetics-related approach is evident in Figures 45 and 46 below.



Projects

Browse Operational Groups and other innovative projects from across Europe that boost innovation for agriculture and forestry. You can search by funding source, country and keywords. For each project, you will find its objectives, results and contact details.

If you would like to share your own innovative projects on the EIP-AGRI website, you can do so in our [Share](#) section.

[Login or register to subscribe to this search](#)

Fulltext search **Keywords**

Contains any of these words Choose some options

Main funding source

Horizon 2020 (EU Research & Innovation programme)

Geographical location

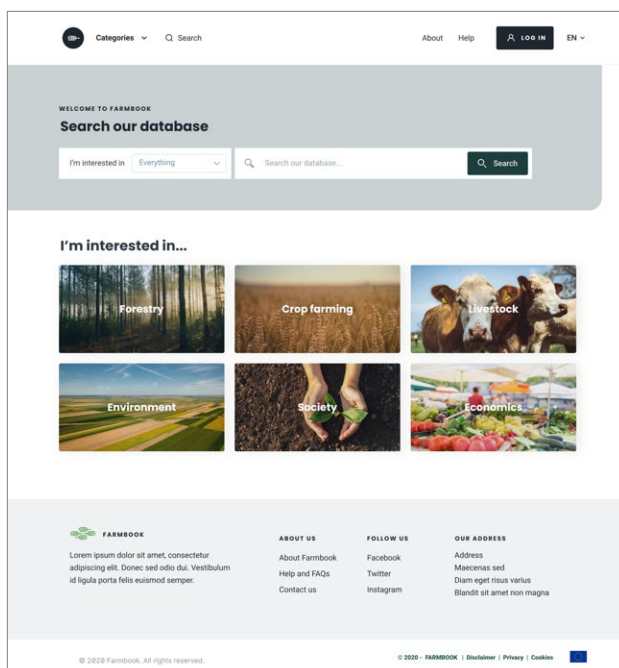
Choose some options [Search](#) [Reset](#)

[Reset content disclaimer](#)

Projects found: 132

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	LIASON: Better Rural Innovation: Linking Actors, Instruments and Policies through Networks	Germany	Research project

Figure 45. The landing pages of the EIP-AGRI database






Categories [About](#) [Help](#) [LOG IN](#) [EN](#)




WELCOME TO FARMBOOK

Search our database

I'm interested in [Search](#)

I'm interested in...

FARMBOOK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed odio du. Vestibulum id ligula porta felis euismod semper.

ABOUT US

About Farmbook
Help and FAQs
Contact us

FOLLOW US

Facebook
Twitter
Instagram

OUR ADDRESS

Address
Macenas sed
Diam eget risus varius
Blandit sit amet non magna

© 2020 Farmbook. All rights reserved. [Disclaimer](#) [Privacy](#) [Cookies](#)

Figure 46. The landing pages of the EURAKNOS e-KRP



Some key results of the comparison presented in this section are summarised in Table 24 below.

Table 24. 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	☑	☑	☑	☑	☑		☑		☑

8. 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 4.1¹⁰³. Some key aspects of it are presented in Subsection 8.1 below ("The EURAKNOS sustainability framework"). Specific solutions to consider as part of a concrete sustainability plan are provided in Subsection 8.2 ("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.

8.1. The EURAKNOS sustainability framework

When it comes to discussing the sustainability of a technological infrastructure like the platform built in EURAKNOS, the issue that immediately comes to the forefront is that of the financial resources necessary for enabling continuity in service provision which entails, among others, hardware/technical equipment maintenance, as well as regular 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 the literature and analysis made in Task 4.1 ("Feasibility and added value of an e-KRP"), and reported in Deliverable 4.1, there are also other dimensions to consider; yet, in direct relation to financial issues. 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 47 below.

¹⁰³ 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.

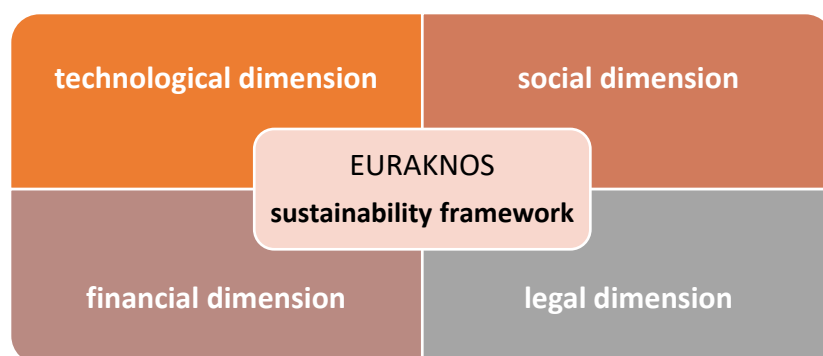


Figure 47. The EURAKNOS sustainability framework

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 25 below contains brief descriptions of what each dimension is about.

Table 25. 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 supply 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: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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 solutions described below (see Subsection 8.2) mostly focus on the financial and legal dimensions of our framework. However, there is an interplay among all four sustainability dimensions which makes them inextricably linked to each other and prompts us to consider them as a whole. For instance, the description of the social dimension of sustainability makes crystal clear that there is a direct link to the technological dimension of the framework (especially concerning the first three bullet points listed in the social dimension's description in Table 25 above). Moreover, the need to allocate financial resources for securing the viability of the e-KRP makes also apparent an indirect relation between the social and financial dimensions.

Another point requiring attention is the twofold approach taken towards the consideration of the legal dimension of the EURAKNOS sustainability framework. To provide recommendations paving the way towards an e-KRP sustainability plan, in Subsection 8.2 below, we have just considered the issue (and necessity) of delivering content and information to the community under a specific, and appropriately selected licensing scheme. Issues relating to copyrighting the EURAKNOS e-KRP solution to ensure the

protection of the EURAKNOS partners' IPRs are beyond the scope of this report. They constitute part of the EURAKNOS exploitation plan provided in Deliverable 5.6 ("Dissemination activity report")¹⁰⁴.

8.2. Towards an e-KRP sustainability plan

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 of these scenarios provides specific solutions for securing the required resources (to ensure continuity) 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 MVP¹⁰⁵, means that these three scenarios are of importance to EUREKA as well (or probably, of greatest importance to EUREKA).

The current implementation of the EURAKNOS e-KRP is hosted in the servers of the University of Gent. The experience obtained so far from the adoption of this specific solution is that there is potential for more direct access to the technological infrastructure allowing the technical partners of EURAKNOS to execute configuration-, maintenance-, update-, and upgrade-related tasks conveniently and directly. To be able to do so, they have been granted access from the Gent University's IT department the expertise of which has been proven to be a valuable ally in efficiently addressing deployment-related problems. Based on that, the solution/scenario appearing to fit best the purpose of long-term supply of the e-KRP-related services is Scenario #1 ("In-house hosting of the e-KRP by a consortium or third-party organisation").

In the case of Scenario #1, there are two business models that have been proposed to ensure revenues, namely: (i) the "freemium" model; and (ii) the "crowdfunding" model. By taking account of the vision of EURAKNOS, and the platform developed in its context, to create an EU-wide community for all AKIS actors by connecting the KRs of TNs, it can be said that the "crowdfunding" model appears to be best for securing money flow and thus, the sustainability of the technological infrastructure and the service related to it. According to the SAGE Encyclopaedia of the Internet¹⁰⁶, crowdfunding is "*the sourcing of small contributions from a large number of individuals*". In the case of EURAKNOS, these "individuals" are the TNs that can benefit from the supply of the e-KRP-related service. Specifically, 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.

Access to the service provided by the e-KRP (and thus, the potential to enjoy the benefits listed above) will be purchased by any interested TN by means of providing an appropriately estimated rate of its

¹⁰⁴ Deliverable 5.6 was also in the process of development by the time that the present report was drafted.

¹⁰⁵ 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.

¹⁰⁶ <http://sk.sagepub.com/reference/the-sage-encyclopedia-of-the-internet-3v/i1840.xml>

dissemination-related budget as compensation for using the e-KRP infrastructure and the associated service. This option will become available and reported from the TN's conceptualisation phase and its proposed development.

As far as the legal dimension of sustainability is concerned, there is a need to offer clear solutions concerning the terms of use of the data objects delivered via the e-KRP. A quick response to this is that all the data objects should be further disseminated under the license attributed to them upon their creation. However, what has been made evident from the investigation undertaken in the case of the TNs, is that the issue of IPRs, when it comes to the dissemination of data objects through the TNs' KRIs, has been poorly addressed. Table 26 below shows the results gathered from this investigation.

Table 26. 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/
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.hnmlink.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 provided in the above table, it can be easily understood that in nearly all (sample) TN cases only some generic references are provided in relation to how their data objects could be used (leaving a wide margin of interpretation). Even in cases of TNs in which detailed information about the data objects' terms of use are supposed to be available (e.g., in the case of FERTINNOWA), there is a complete absence of a specific licensing scheme explicitly framing these terms of use. The only case of TN where data objects are distributed under a specific license is that of Smart-AKIS.

Based on these findings and conclusions, it is imperative to contribute to good practices of data object use and further exploitation of knowledge by proposing, in the context of the EURAKNOS sustainability plan, a specific licensing scheme under which TN-related data objects should be further disseminated from the EURAKNOS e-KRP. The licensing framework that has been adopted for this purpose is Creative

Commons¹⁰⁷. This choice has been made by taking into consideration this framework's popularity and the range of licensing options available to address various needs and data use cases. Table 27 lists the main Creative Commons license types together with short descriptions of them.

Table 27. 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/
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-	This license allows re-users to copy and distribute the material in any medium or	https://creativecommons.org/licenses/by-nc-nd/4.0/

¹⁰⁷ <https://creativecommons.org/>

#	License name	License description	URL
	NoDerivatives 4.0 International (CC BY-NC-ND 4.0)	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: <ul style="list-style-type: none"> • 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 	
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/

By taking account of the non-commercial nature of the EURAKNOS e-KRP¹⁰⁸, based on the purpose the platform aims to serve, as well as the guide available in the Creative Commons website for facilitating decisions on which type of license to select¹⁰⁹, the license chosen 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)¹¹⁰. More details about this license type are provided in Table 28.

Table 28. 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
	<ul style="list-style-type: none"> • 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.

To sum up, the employment of the “crowdfunding” model in the context of a 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, however, an ongoing effort, which has been already taken up by EUREKA (the follow-up, sibling project of EURAKNOS). An issue needed to be investigated to a further extent is that related to the actors/stakeholders from whom consent needs to be provided to further disseminate data objects through the e-KRP. In the context of EURAKNOS, consent has been asked from the coordinators of our sample's TNs. However, we need to consider that EURAKNOS has aimed to develop a prototype platform aiming to serve as a proof-of-concept for the feasibility of this endeavour. Given the broader scope of EUREKA (it focuses upon 101 MAPs instead of 8 TNs, as well as

¹⁰⁸ 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).

¹⁰⁹ The guide is called “Creative Commons License Chooser” and is available at <https://creativecommons.org/choose/#>.

¹¹⁰ This is the license under which the data objects from the Smart-AKIS KR are also made available.

it progresses the development of the platform from a prototype to an MVP), the issues of consent, IPR management, and sustainability, in general, will be further explored.

9. 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 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 repository¹¹¹. 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. This was not always the case, however. 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”

¹¹¹ See Subsection 5.3.1 (“Data cleaning”) for details.

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 the minimum set of metadata for the annotation/characterisation of the TN data objects¹¹². 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 through a TN’s data objects. The specific 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.

¹¹² Details about the process that has led to the EURAKNOS metadata set creation and definitions of the respective metadata properties are available in Subsections 5.3.3.1 (“Updating the EURAKNOS data model”) & 5.3.3.2 (“The EURAKNOS metadata set”).

10. 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.

10.1. e-KRP landing page

The landing page of the e-KRP is shown in Figure 48.

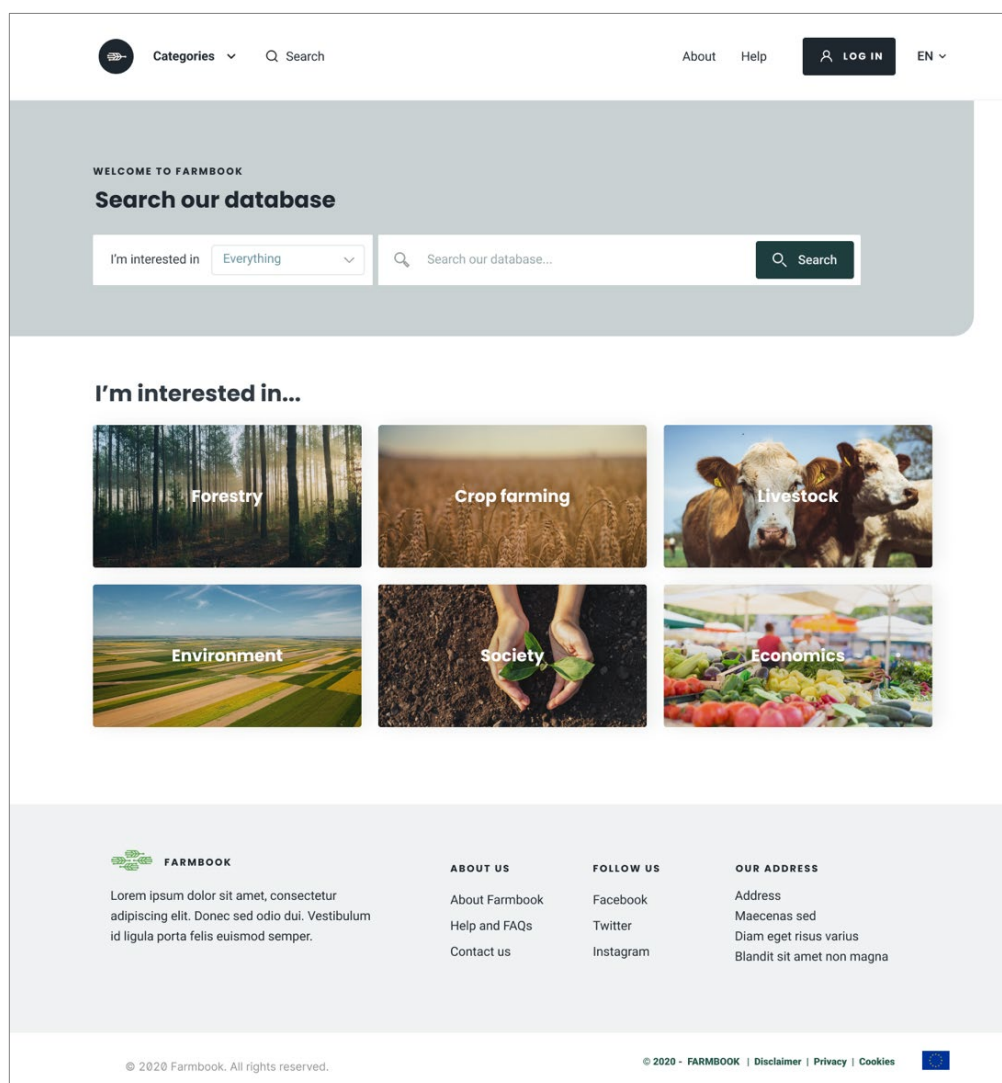


Figure 48. The landing page of the EURAKNOS e-KRP



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 the content. Specifically, there is the option of typing a search query (using one or more search terms) in the search bar and then submit it by clicking on the search button. The search bar has been placed at a prominent spot 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 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 about three 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 available when clicking on the "Categories" link at the top of the page (see Figure 49). This link is not provided only on the landing page but on every e-KRP page. Thus, there is potential to access content about any sector or theme without needing to return to the platform's landing page.

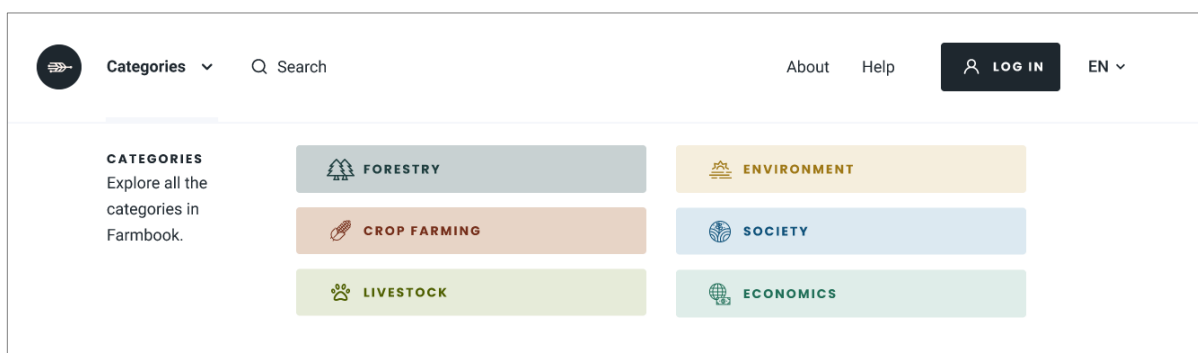


Figure 49. Dropdown menu of the sector- and theme-related categories of content able to be accessed by clicking on the link labelled "Categories" at the top of every platform page

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 50 below. The form to be filled-in to come in contact with a platform representative is illustrated in Figure 51.

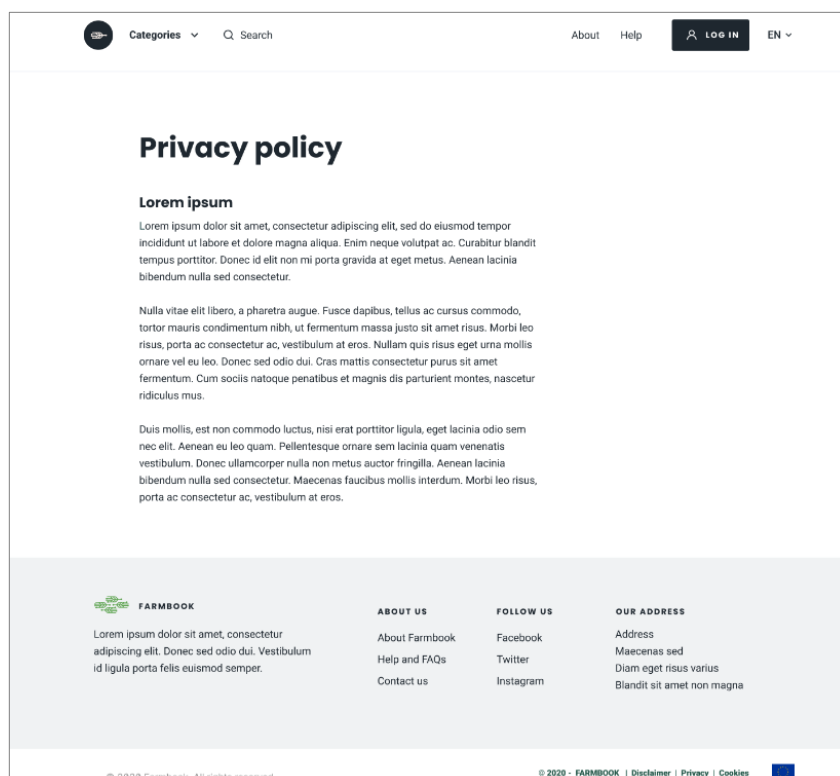


Figure 50. e-KRP page with details about the adopted privacy-related policy

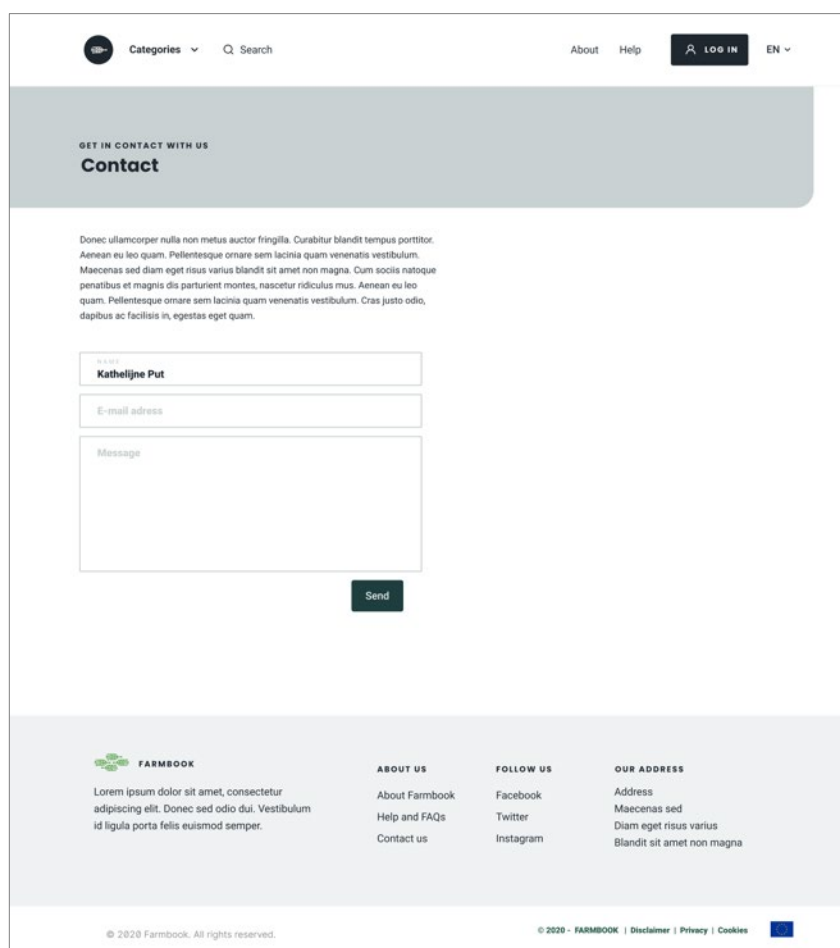


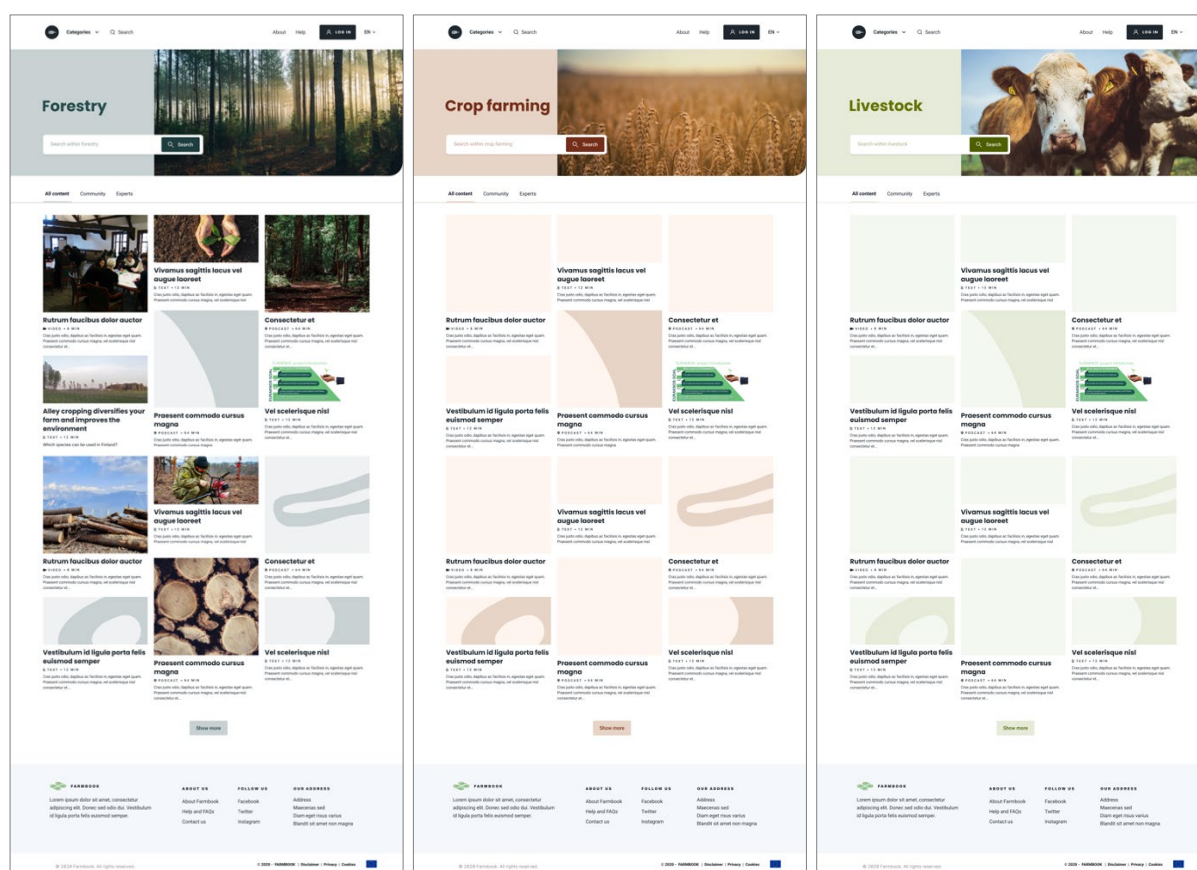
Figure 51. The “Contact us” page of the e-KRP



10.2. Sector- and theme-related pages

Figure 52 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 49 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 52 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 you to search for content and information by typing in and submitting a query, using the "Search" button.

Each sector- and theme-related page provides the same menu of options labelled as "All content", "Community", and "Experts". By choosing the "All content" menu option, there is potential to access all the content and information associated with the specific sector or theme. This information/content is available in a range of data object types and formats. The details available for each of the relevant data objects (conveying the sector- or theme-related content/information) is the category they belong to (i.e., document/text, video, podcast, infographic, or presentation), an estimation of the time needed to go through their content and a short description of the data object's 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.



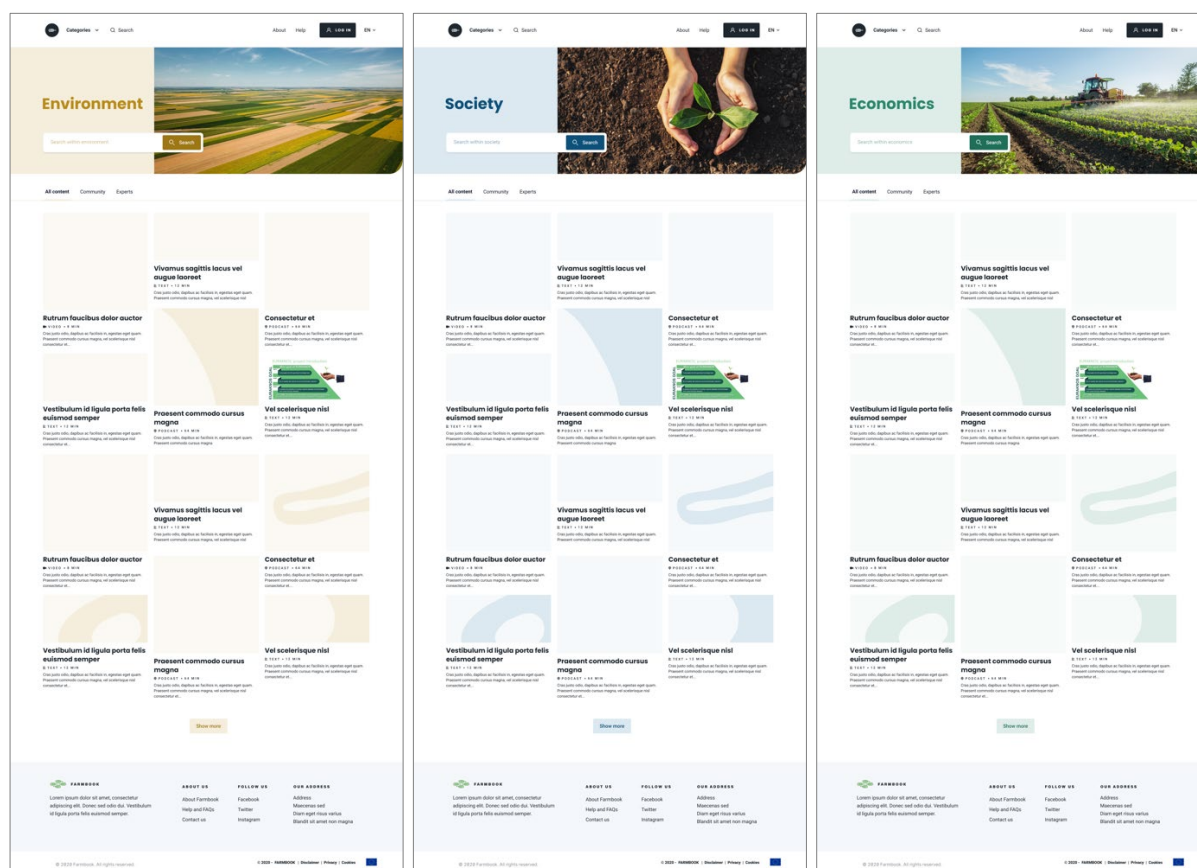


Figure 52. The sector- and theme-related pages able to be visited by making a selection from the relevant menu or by clicking on one of the landing page's tiles

Figure 53 below shows the e-KRP page related to the sector of Forestry and specifically the view of the page that is accessed when selecting the option of the menu having the label “All content”.

The “Community” option directs to the page of the e-KRP that can be used for communication and interaction with other actors. On that page, there is potential to post a topic for discussion and start a thread of messages related to the specific topic¹¹³. The discussion page, as well as a thread of messages posted for a specific topic of discussion, are shown in Figures 54 and 55 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 56. The expert-related details are listed in alphabetical order by taking account of their names. In the case that a large list of experts exists, the “Show more” button allows to load more expert-related contact details.

Figures 57, 58, 59, and 60 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.

¹¹³ 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.

In the screenshots showing the views of podcasts and videos (Figures 59 and 60 respectively), we can also see that the option to post a comment is provided.

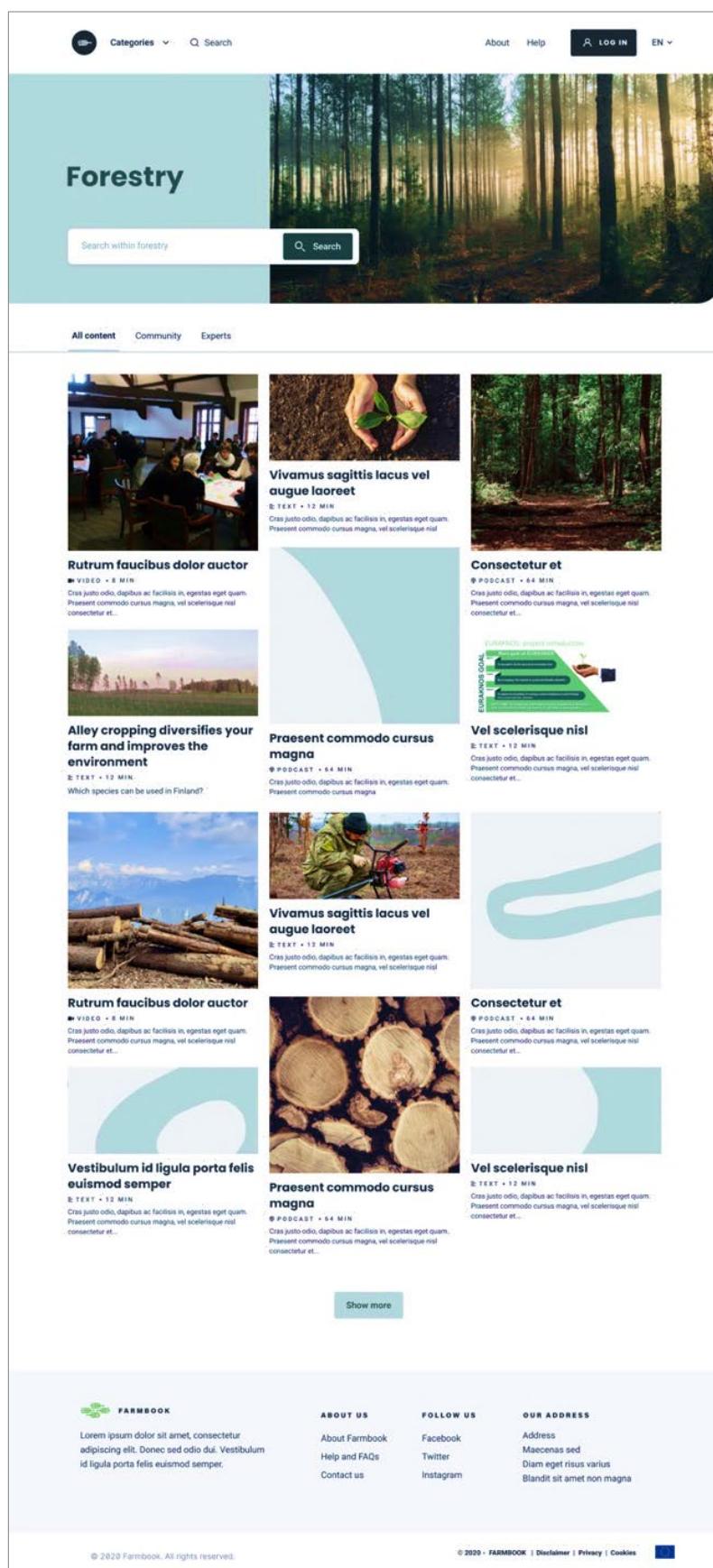
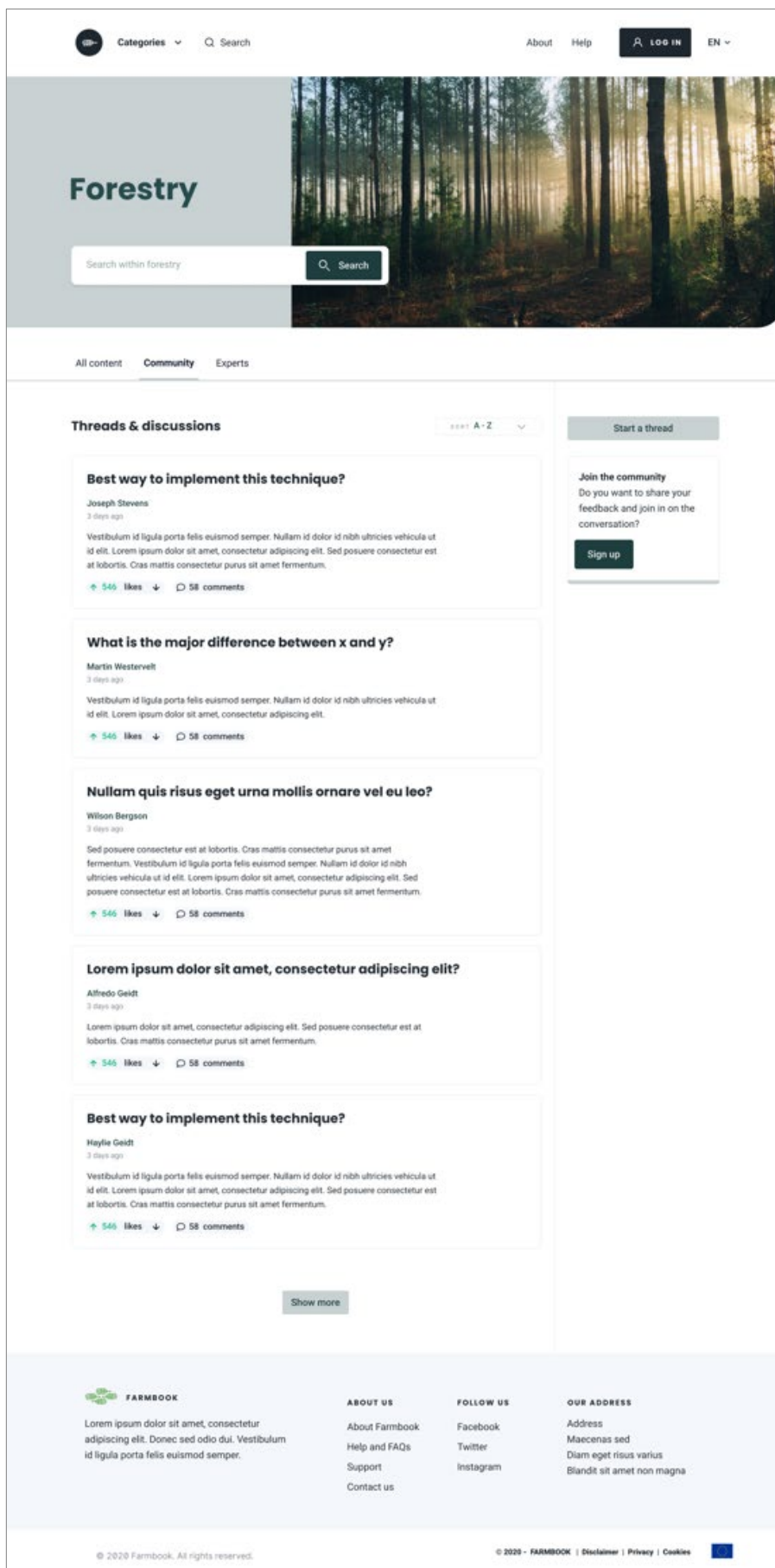


Figure 53. The e-KRP page related to the sector of Forestry





Forestry

Search within forestry

All content **Community** Experts

Threads & discussions

Sort A-Z

Best way to implement this technique?

Joseph Stevens
3 days ago

Vestibulum id ligula porta felis euismod semper. Nullam id dolor id nibh ultricies vehicula ut id elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum.

546 likes 58 comments

What is the major difference between x and y?

Martin Westervelt
3 days ago

Vestibulum id ligula porta felis euismod semper. Nullam id dolor id nibh ultricies vehicula ut id elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

546 likes 58 comments

Nullam quis risus eget urna mollis ornare vel eu leo?

Wilson Bergson
3 days ago

Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum. Vestibulum id ligula porta felis euismod semper. Nullam id dolor id nibh ultricies vehicula ut id elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum.

546 likes 58 comments

Lorem ipsum dolor sit amet, consectetur adipiscing elit?

Alfredo Geldt
3 days ago

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum.

546 likes 58 comments

Best way to implement this technique?

Haylie Geldt
3 days ago

Vestibulum id ligula porta felis euismod semper. Nullam id dolor id nibh ultricies vehicula ut id elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum.

546 likes 58 comments

Show more

FARMBOOK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper.

ABOUT US

About Farmbook
Help and FAQs
Support
Contact us

FOLLOW US

Facebook
Twitter
Instagram

OUR ADDRESS

Address
Maecenas sed
Diam eget risus varius
Blandit sit amet non magna

© 2020 Farmbook. All rights reserved.

© 2020 - FARMBOOK | Disclaimer | Privacy | Cookies

Figure 54. Topics of discussion in the “Community” page



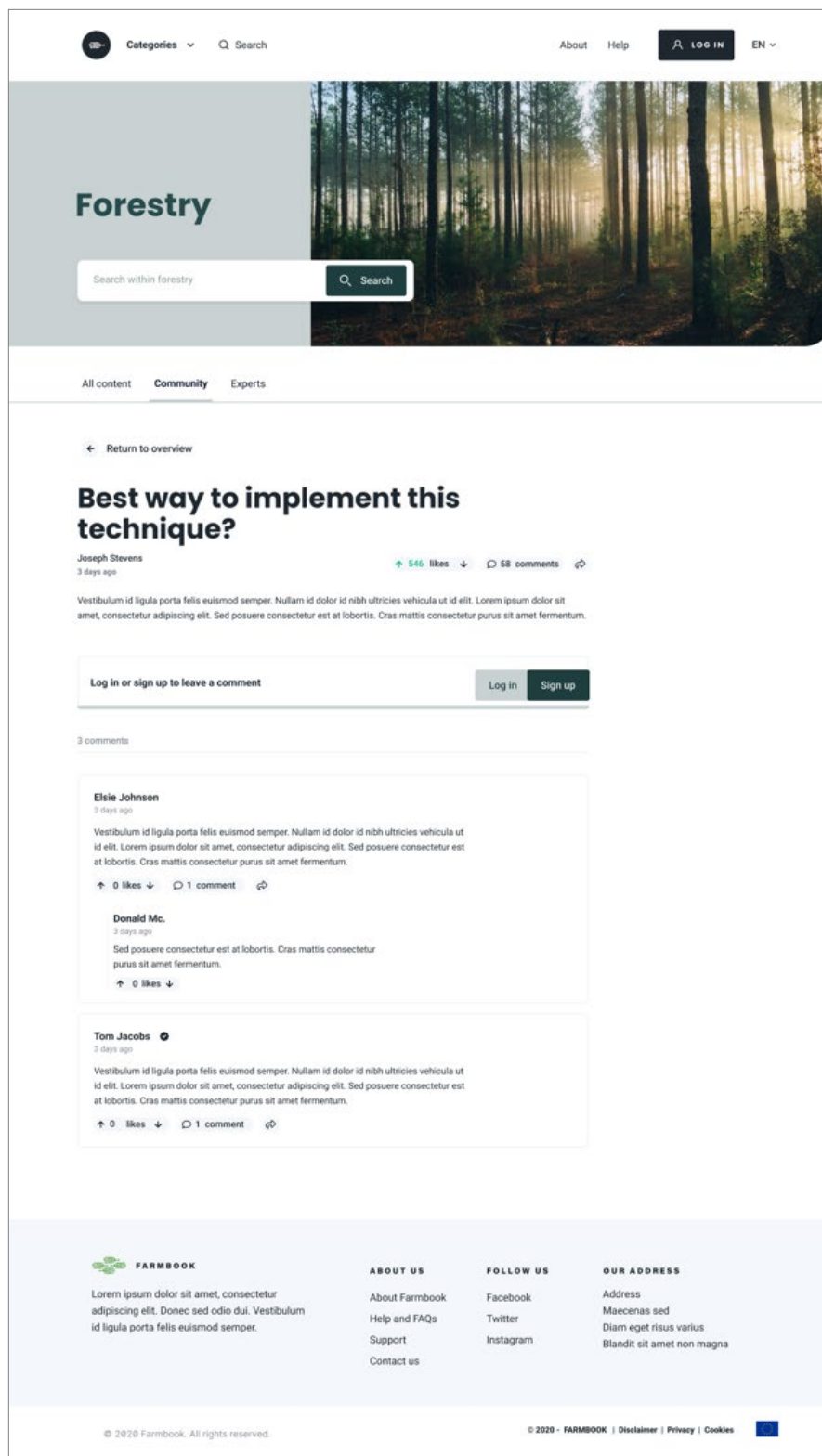
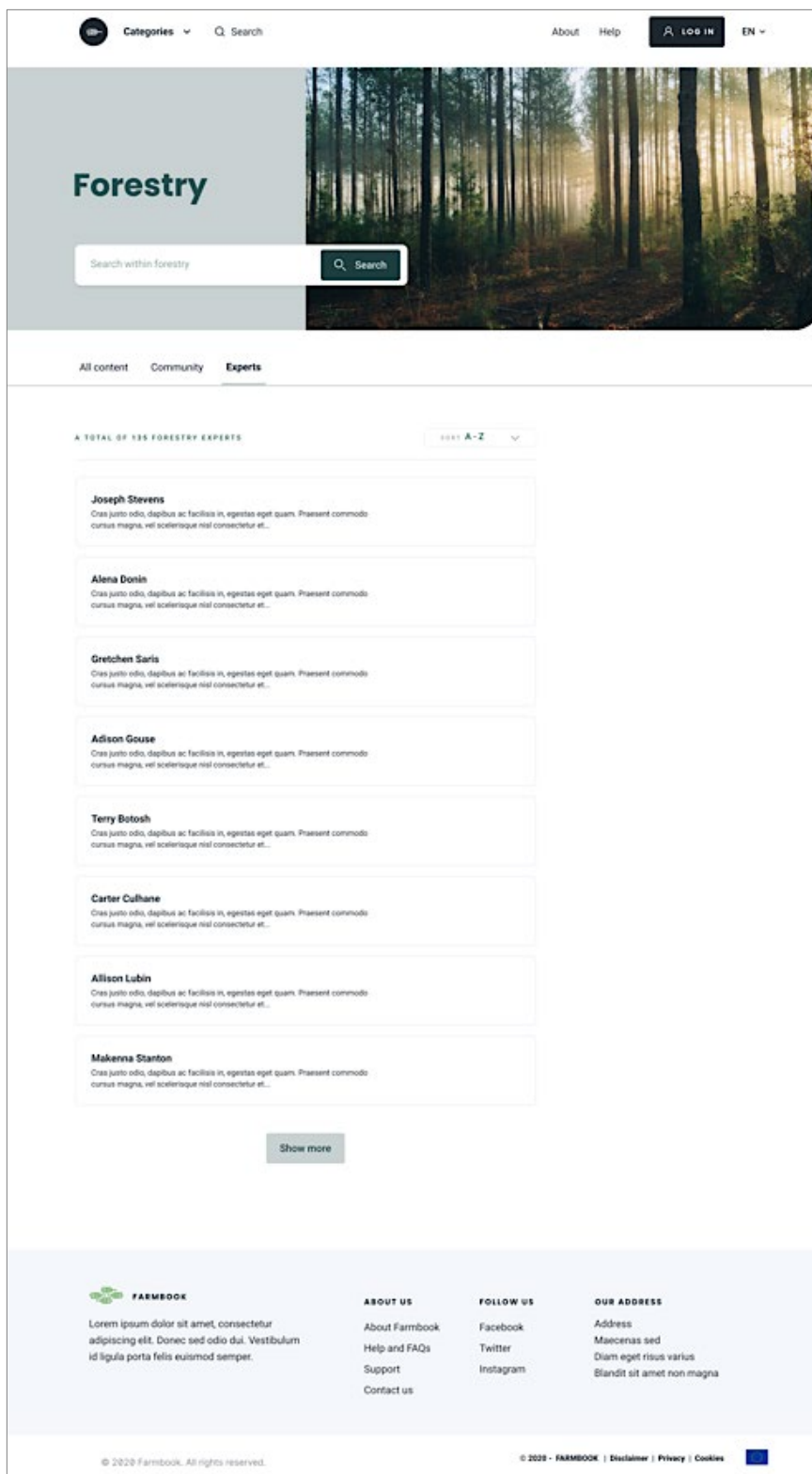


Figure 55. Message threads related to a topic of discussion



Forestry

Search within forestry

All content Community **Experts**

A TOTAL OF 135 FORESTRY EXPERTS sort A-Z

- Joseph Stevens**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Alena Donin**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Gretchen Saris**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Adison Gouse**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Terry Botosh**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Carter Culhane**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Allison Lubin**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...
- Makenna Stanton**
Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl consectetur et...

[Show more](#)

FARMBOOK
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper.

ABOUT US
About Farmbook
Help and FAQs
Support
Contact us

FOLLOW US
Facebook
Twitter
Instagram

OUR ADDRESS
Address
Maecenas sed
Diam eget risus varius
Blandit sit amet non magna

© 2020 Farmbook. All rights reserved. © 2020 - FARMBOOK | Disclaimer | Privacy | Cookies

Figure 56. List of Forestry experts

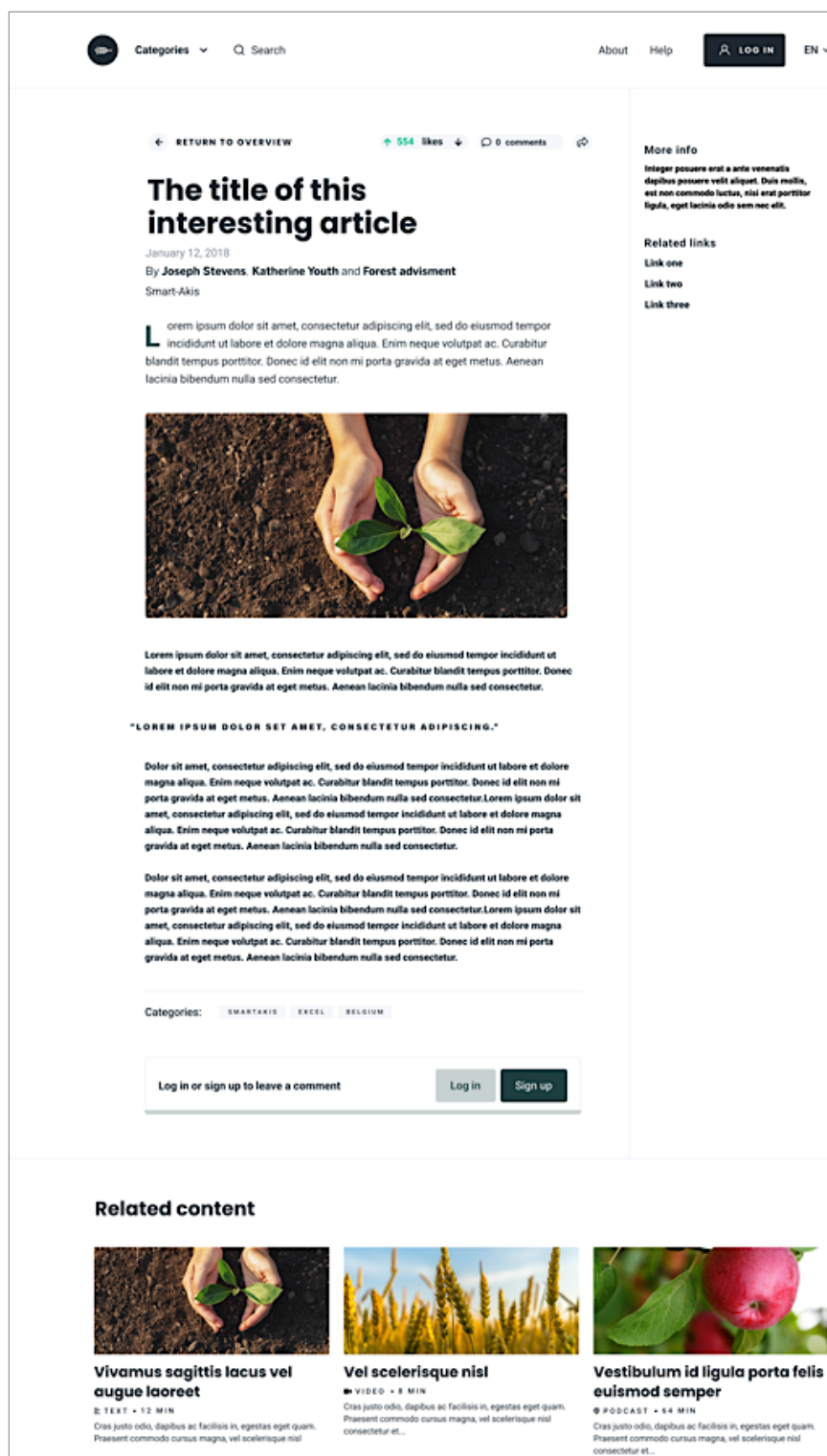


Figure 57. View of an article

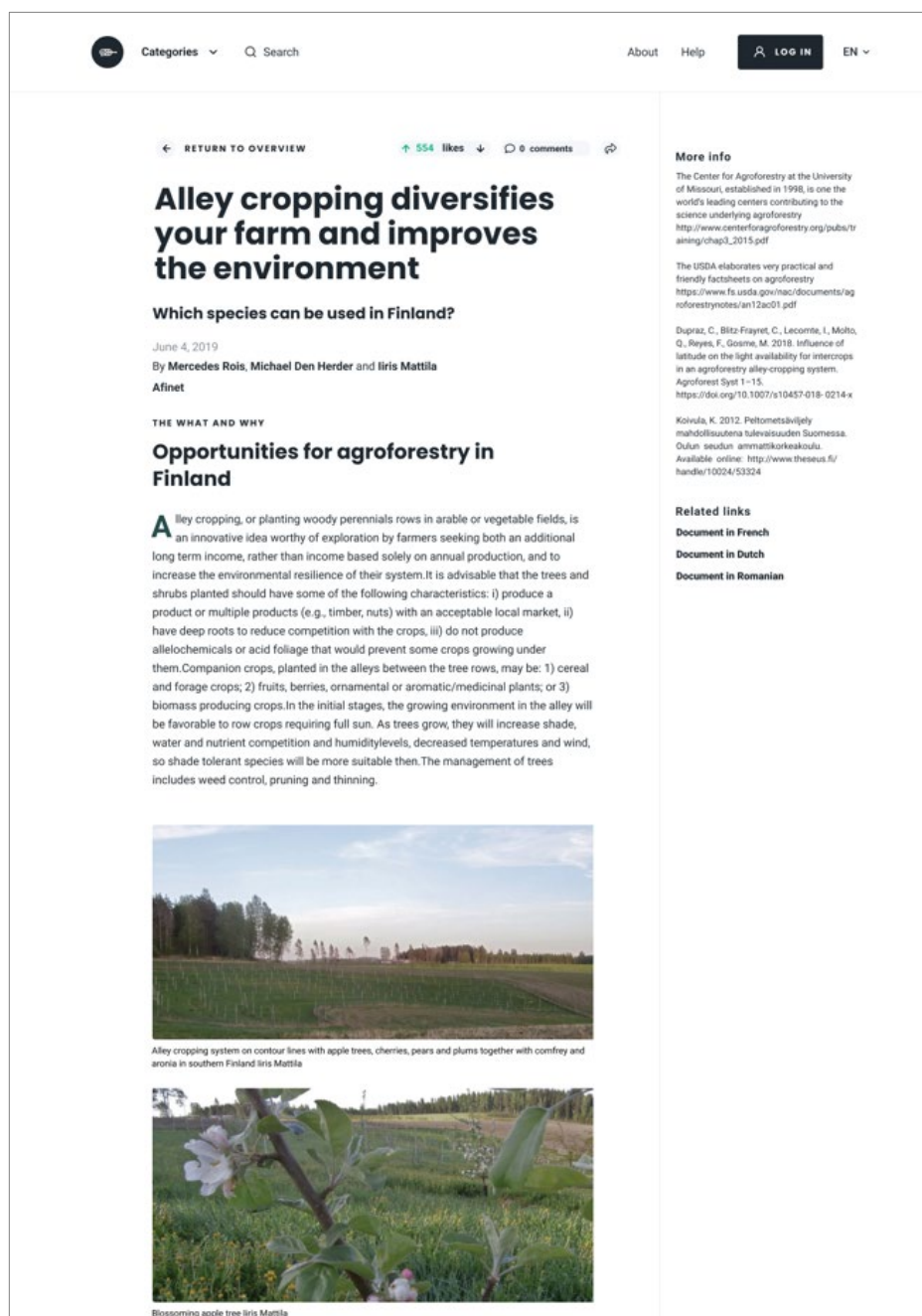
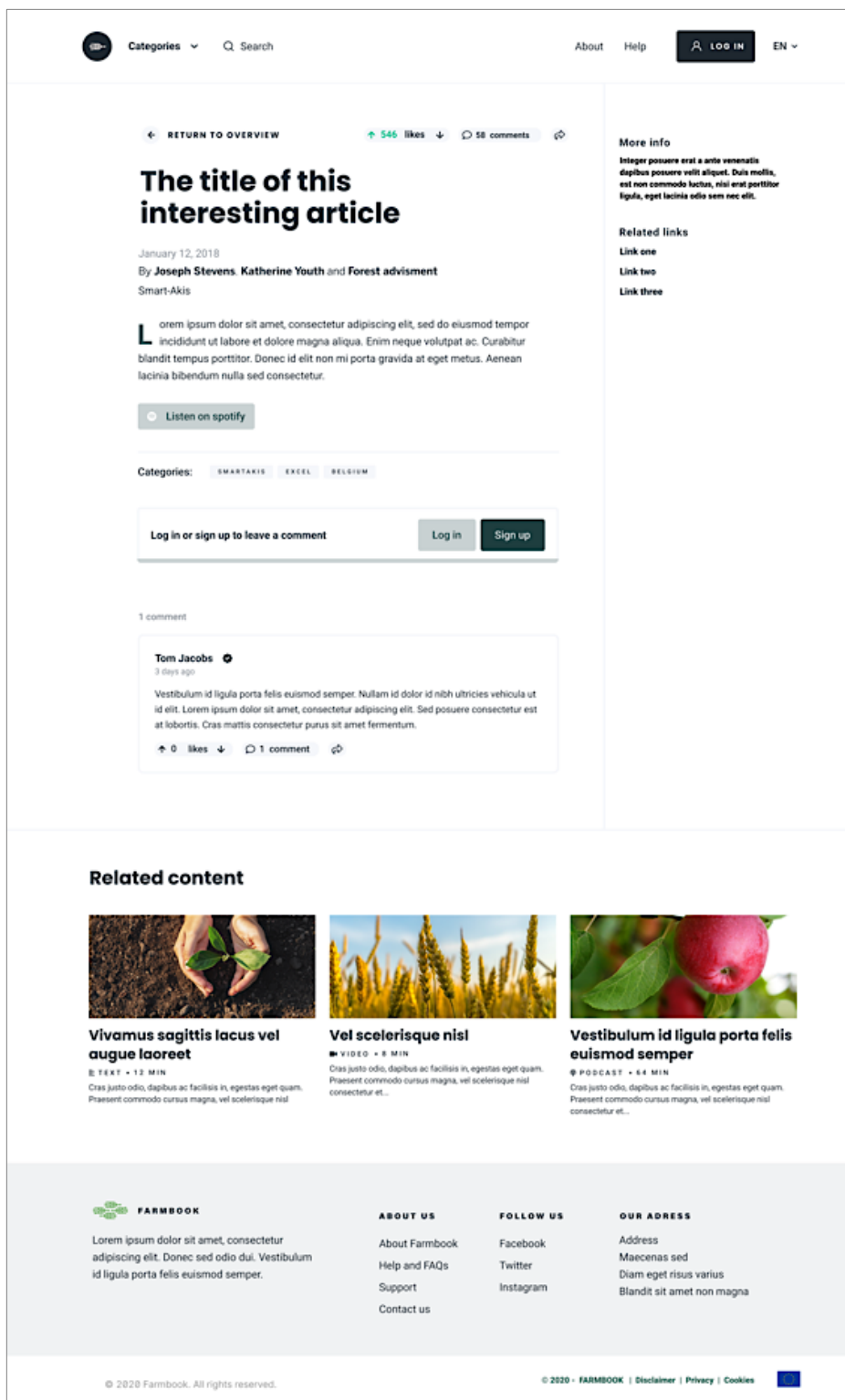


Figure 58. View of a factsheet



Categories ▾ Search

About Help LOG IN EN ▾

RETURN TO OVERVIEW 546 likes 58 comments

The title of this interesting article

January 12, 2018
By **Joseph Stevens, Katherine Youth and Forest advisement**
Smart-Akiss

>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Enim neque volutpat ac. Curabitur blandit tempus porttitor. Donec id elit non mi porta gravida at eget metus. Aenean lacinia bibendum nulla sed consectetur.

[Listen on spotify](#)

Categories: SMARTAKISS EXCEL BELGIUM

Log in or sign up to leave a comment Log in Sign up


1 comment

Tom Jacobs 3 days ago

Vestibulum id ligula porta felis euismod semper. Nullam id dolor id nibh ultricies vehicula ut id elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed posuere consectetur est at lobortis. Cras mattis consectetur purus sit amet fermentum.

0 likes 1 comment


Related content



Vivamus sagittis lacus vel augue laoreet

TEXT • 12 MIN


Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl.



Vel scelerisque nisi

VIDEO • 8 MIN

Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl.



Vestibulum id ligula porta felis euismod semper

PODCAST • 64 MIN

Cras justo odio, dapibus ac facilisis in, egestas eget quam. Praesent commodo cursus magna, vel scelerisque nisl.

FARMBOOK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper.

ABOUT US

About Farmbook
Help and FAQs
Support
Contact us

FOLLOW US

Facebook
Twitter
Instagram

OUR ADDRESS

Address
Maecenas sed
Diam eget risus varius
Blandit sit amet non magna

© 2020 Farmbook. All rights reserved. 2020 • FARMBOOK | Disclaimer | Privacy | Cookies

Figure 59. View of a podcast

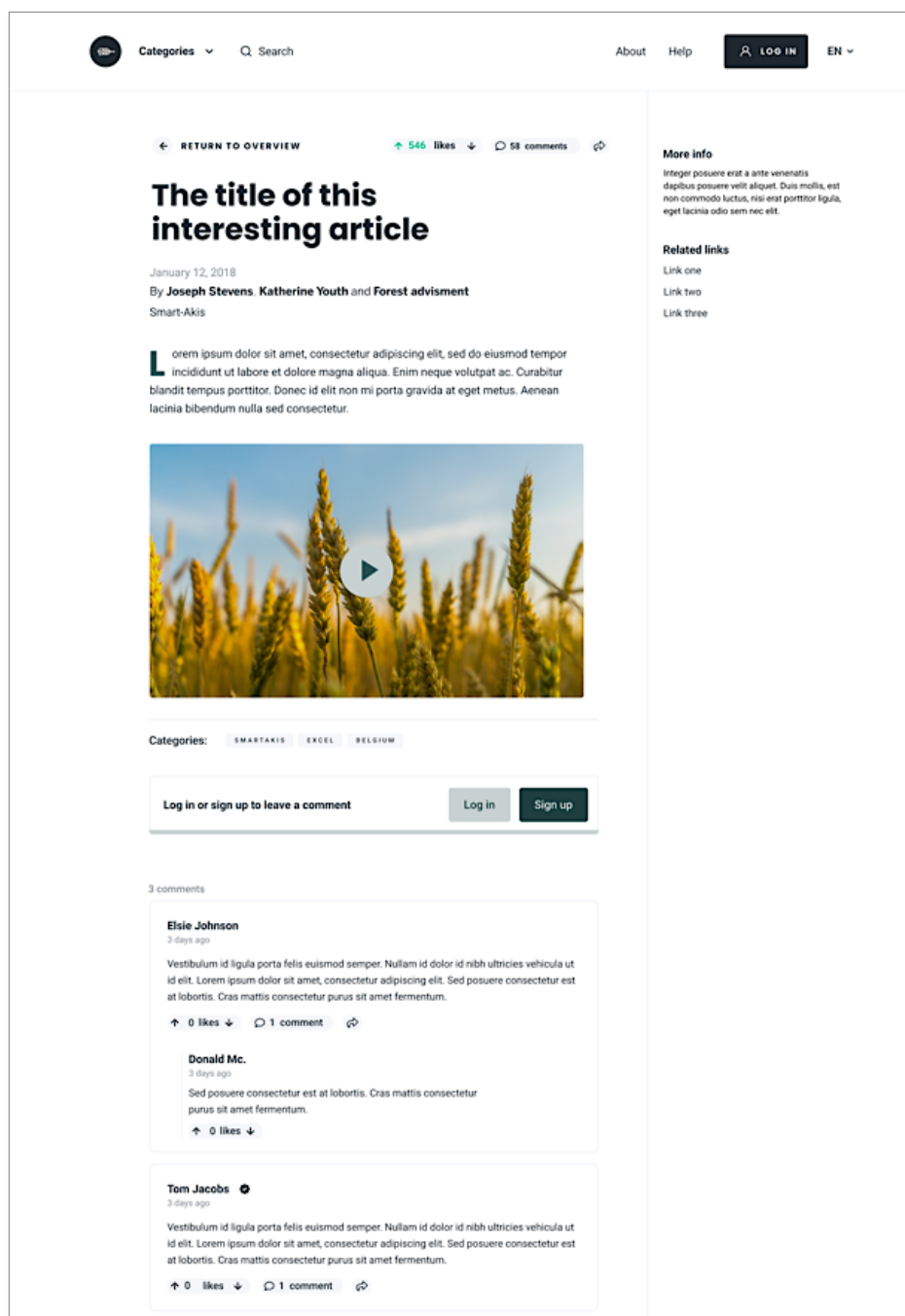
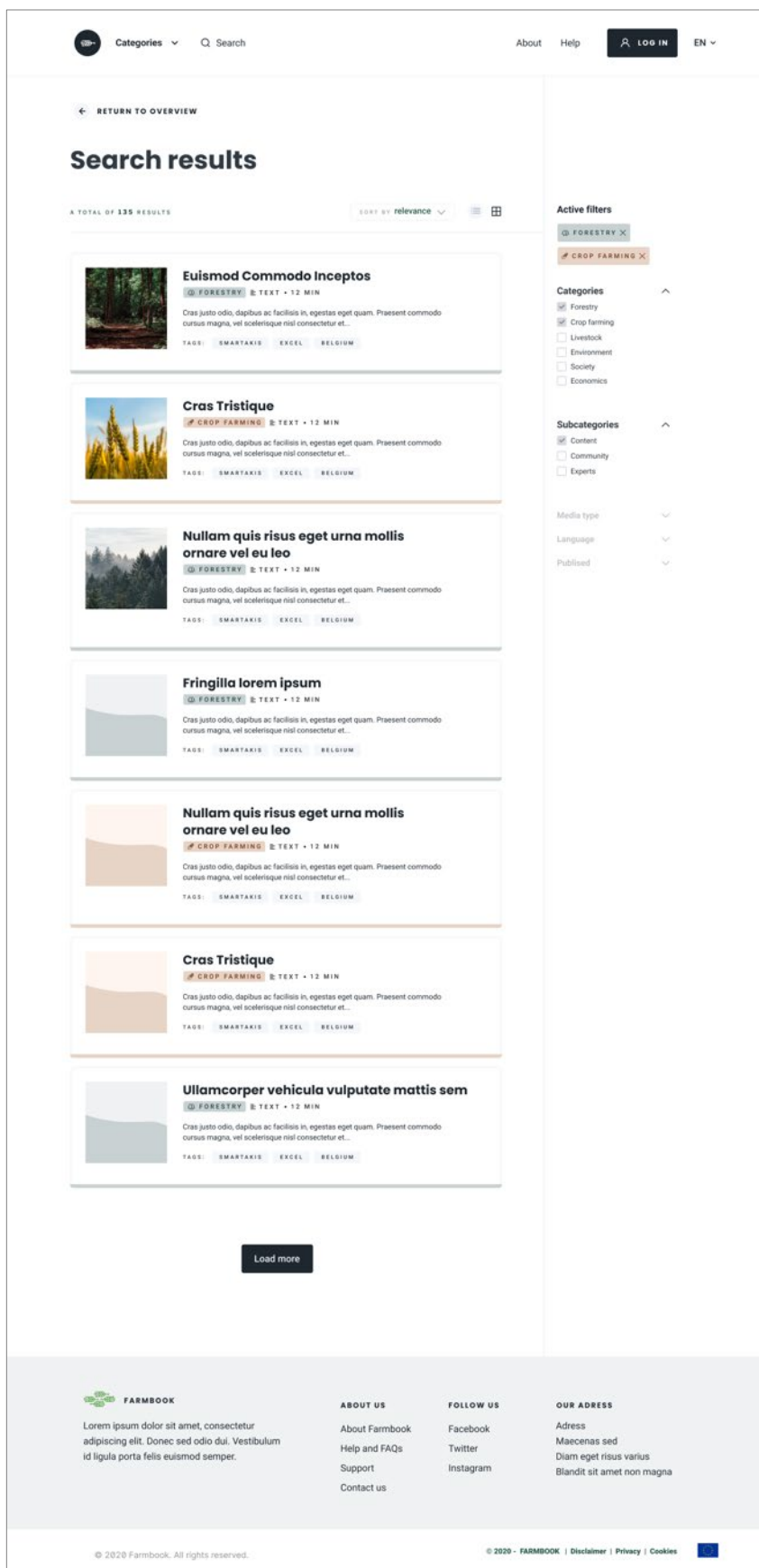


Figure 60. View of a video

10.3. Display of search results

When it comes to searching for information and content on the web, the functionality that users have been accustomed to using is that of a search bar where a search query can be typed (using any search terms) and submitted. This is one of the basic means to access content and information via the e-KRP as well. Figures 61 and 62 below show the two different modes for the delivery of results of a search operation. In Figure 61, the mode of results' display is that of a list, whereas in Figure 62 the results of the search are displayed in a tile mode. In both cases, there are some basic pieces of information 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 through which the content/information is conveyed (text/document, presentation, video, podcast, or

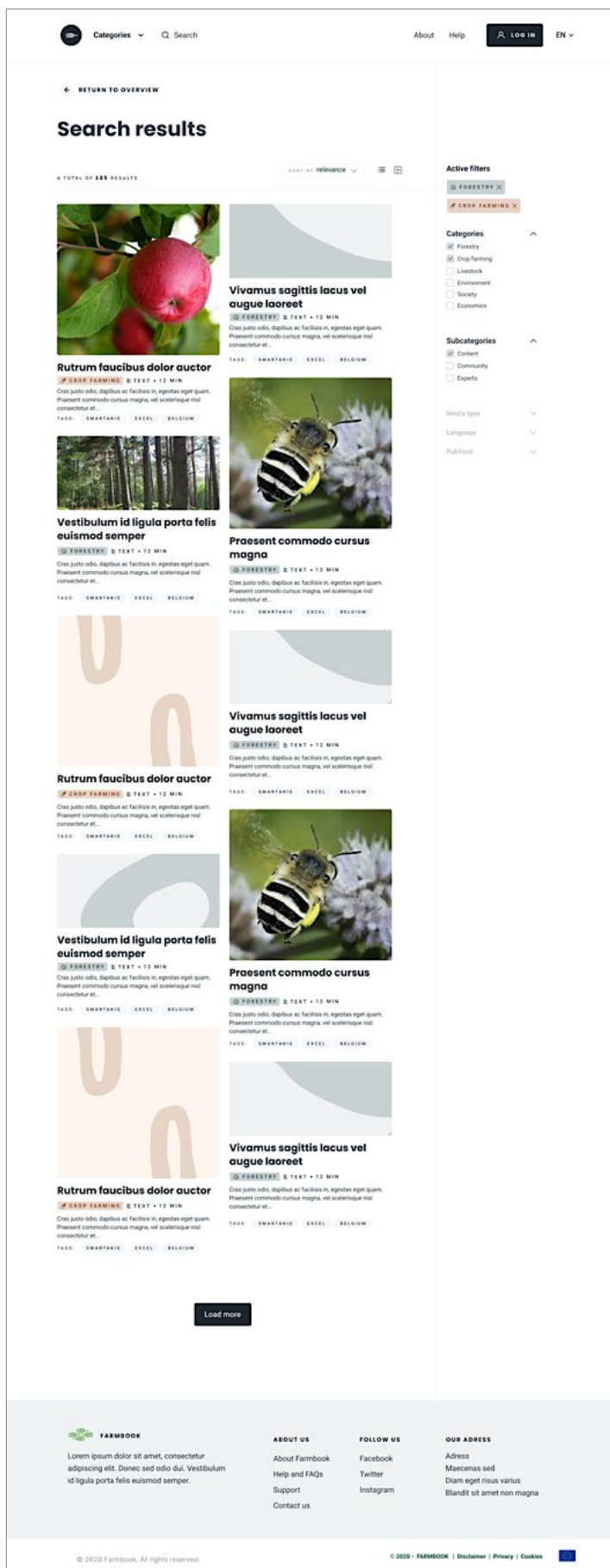
infographic); (iv) an estimate of the time needed to go through the data object's content; (v) a short description; and (vi) an image indicative of the content of each search result.



The screenshot displays the 'Search results' page of the Farmbook application. The page features a top navigation bar with 'Categories', 'Search', 'About', 'Help', 'LOG IN', and 'EN'. A 'RETURN TO OVERVIEW' link is located at the top left of the main content area. The search results are displayed as a list of cards, each containing a thumbnail image, a title, a category tag (e.g., 'FORESTRY', 'CROP FARMING'), a duration (e.g., '12 MIN'), a short description, and a list of tags (SMARTAXIS, EXCEL, BELGIUM). The results are sorted by 'relevance'. On the right side, there is a sidebar with 'Active filters' (FORESTRY, CROP FARMING), 'Categories' (Forestry, Crop farming, Livestock, Environment, Society, Economics), 'Subcategories' (Content, Community, Experts), 'Media type', 'Language', and 'Published'. A 'Load more' button is positioned at the bottom of the results list. The footer contains the Farmbook logo, contact information, social media links, and copyright details.

Figure 61. Display of search results as a list



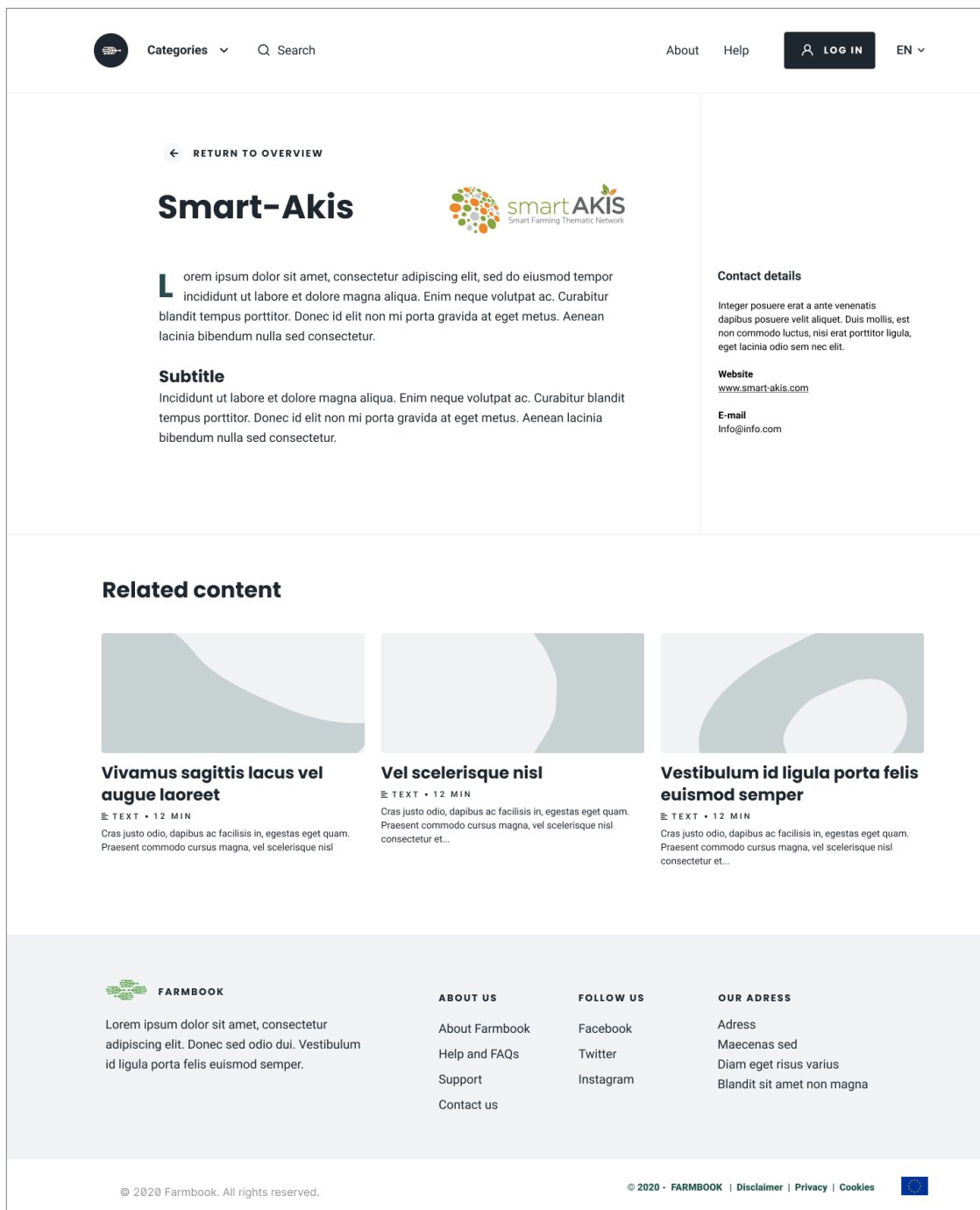


The screenshot displays the EURAKNOS search results interface. At the top, there's a navigation bar with 'Categories', 'Search', 'About', 'Help', and a 'LOG IN' button. Below the navigation bar, a 'RETURN TO OVERVIEW' link is visible. The main section is titled 'Search results' and shows 'A TOTAL OF 125 RESULTS'. The results are displayed in a grid of tiles. Each tile includes a thumbnail image, a title, a category tag (e.g., 'FORESTRY', 'CROP FARMING'), a duration (e.g., '12 MIN'), a brief description, and tags (e.g., 'SMARTAGIS', 'EXCEL', 'BELGIUM'). The tiles are arranged in two columns. On the right side, there are filters for 'Active filters' (Forestry, Crop Farming), 'Categories' (Forestry, Crop Farming, Livestock, Environment, Society, Economics), 'Subcategories' (Content, Community, Experts), 'Media type', 'Language', and 'Published'. At the bottom of the grid, there is a 'Load more' button. The footer contains the 'FARMBOOK' logo, contact information, social media links, and copyright information.

Figure 62. 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-related items. These filters are shown at the top right corner in each of the two figures above (Figures 61 and 62). The filters made available 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 conveying the content/information (namely, text/document, video, presentation, podcast, and infographic); (iii) the language(s) in which the content is delivered; (iv) the content delivery format; (v) the publication year; and (vi) the purpose for which the content/information has been released. In the case that information about a TN (i.e., the TN from which specific pieces of information and content have become available) needs to be accessed, this can be provided through an e-KRP page especially dedicated to this purpose (see Figure 63 below).



The screenshot shows a web page for 'Smart-Akis', a Smart Farming Thematic Network. The page layout includes a top navigation bar with 'Categories', 'Search', 'About', 'Help', 'LOGIN', and 'EN'. A 'RETURN TO OVERVIEW' link is at the top left. The main content area features the 'Smart-Akis' title and logo, followed by a paragraph of placeholder text. A 'Contact details' sidebar on the right lists 'Website' (www.smart-akis.com) and 'E-mail' (info@info.com). Below the main content is a 'Related content' section with three cards: 'Vivamus sagittis lacus vel augue laoreet', 'Vel scelerisque nisl', and 'Vestibulum id ligula porta felis euismod semper'. Each card includes a placeholder image, title, duration (12 MIN), and a short description. The footer contains 'FARMBOOK' information, 'ABOUT US' links (About Farmbook, Help and FAQs, Support, Contact us), 'FOLLOW US' social media links (Facebook, Twitter, Instagram), 'OUR ADDRESS' (Address, Maecenas sed, Diam eget risus varius, Blandit sit amet non magna), and copyright information (© 2020 Farmbook. All rights reserved.) along with a European Union flag.

Figure 63. e-KRP page displaying information about a TN



11. Conclusions and next steps

The present report has provided all the details of the work done in the context of Task 4.2 concerning the development of the EURAKNOS prototype e-KRP. The development process has involved diverse activities, requiring a broad range of expertise which has been brought by the project partners involved in the task. The descriptions of the undertaken activities have revealed the tight connections between Task 4.2 and tasks in the same 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 reported 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 report sections. This provides a strong indication of the interconnectedness of the two projects. It also makes clear that there are issues remaining open. These issues need to be further addressed and evolutions 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 the specific endeavour), whereas EUREKA 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 underwent a process of an update to better adhere to the FAIR data principles. This update has aimed to align the properties described in our model with properties defined in external, widely-adopted vocabularies. However, this effort of FAIRification of our conceptual design and model can be viewed as a task in progress. The intention is to continue, in the context of EUREKA, the investigation for controlled vocabularies that could be considered for defining entities, concepts, relations, and properties to come up with a robust (in terms of coverage of our needs for describing the data objects available via the platform), flexible (potential 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 (where 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 (i.e., 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). Therefore, we envisage that the metadata proposed in EURAKNOS will be sufficient (both in number and coverage aspects) for describing the digital outputs (i.e., data/knowledge objects¹¹⁴) of MAPs as well. Any potential changes/updates to the set of metadata will (probably) relate to the metadata property definitions given the ongoing (in the context of EUREKA) process of considering external vocabularies for the description of our model's elements (i.e., entities, concepts, relations, and entity properties).
- *Adaptations in the ETL process*
As explicitly mentioned in Subsection 5.4 ("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

¹¹⁴ The term that is used in EUREKA to refer to the digital outputs available through the KRs of MAPs is "knowledge object".

ingest the data objects harvested by the EURAKNOS crawler (from the sample TN KRs) into the e-KRP repository (DSpace). However, the ETL process has been tested only in regard to the ingestion of the data object categories available from the KRs of the TNs (namely, 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 not considered in EURAKNOS (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 Virtuoso¹¹⁵, which is adopted in many European projects. However, this process was not taken further in EURAKNOS because of the technical issues involved in it, which appeared to require a substantial amount of time to be solved. Consequently, it has been decided that this issue will be further considered in EUREKA. This is not a critical development-related issue but rather relates to an effort to better align with the technology choices made in other EU research projects¹¹⁶.

- *Changes/adaptations in functionalities and features*

Extensive and broad user tests have not been considered in the conceptualisation and methodology design of EURAKNOS. However, it has been feasible to collect feedback from potential users as part of project meetings and workshops that have taken place during the project (e.g., the feedback collected by the consortium partners in the wireframe-based¹¹⁷ exercise delivered by LFG in the Athens consortium meeting¹¹⁸, as well as the input of participants in the marketplace session in the Paris workshop¹¹⁹). Details about the feedback collected from the various testing exercises in the context of EURAKNOS are available in Section 7 of the report ("Testing the EURAKNOS e-KRP"). A thorough user-testing approach has been considered in EUREKA together with a WP particularly dedicated to the collection of useful information about the profiles or "archetypes" of specific end-user groups (WP2). The information intended to be collected from these tasks/activities will serve as valuable feedback to the functionalities already provided by the e-KRP (the EUREKA's FarmBook platform is the follow-up of the EURAKNOS's prototype e-KRP). This feedback will be exploited to revisit (if necessary) functionalities and features and probably consider new ones. The EURAKNOS Vision Paper will also serve as an important source of input regarding functionalities to be further considered and developed in EUREKA. Within this context, the provision of a SPARQL endpoint¹²⁰ (incorporated into the API layer of the system), aimed to be used by researchers/scientists, will also be under a feature under consideration.

- *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

¹¹⁵ <https://virtuoso.openlinksw.com/>

¹¹⁶ 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.

¹¹⁷ An interesting resource for getting to know what wireframes are and what they are used for is <https://balsamiq.com/learn/articles/what-are-wireframes/>.

¹¹⁸ The Athens consortium meeting was held on the 17th and 18th June, 2019.

¹¹⁹ The EURAKNOS Paris workshop was held from the 11th to the 13th of December, 2019.

¹²⁰ 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>.

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 financial resources required to keep the technological infrastructure running and the associated service being provided; (ii) the continuity in service provision as part of a scenario encompassing the in-house hosting of the e-KRP by a consortium/third-party organisation; and (iii) the distribution/further dissemination of digital content/information under an appropriate license type. However, there is more to investigate. To this end, the sustainability-related work is continued in EUREKA. Some of the issues to be further considered in EUREKA are:

- The investigation of what are the entities needed to be asked for granting permission to further disseminate the data/knowledge objects of TNs/MAPs through the e-KRP/FarmBook.
- The consideration of further licensing scheme/type alternatives given the greater breadth of the data/knowledge objects needed to be considered in EUREKA. For instance, some broadly-known licensing frameworks to consider for the needs of securing the IPRs of software-related work are Apache¹²¹ and GNU¹²².

12. 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.
- 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.
- 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*.

¹²¹ <https://spdx.org/licenses/Apache-2.0.html>

¹²² <https://www.gnu.org/licenses/gpl-3.0.html>



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.



13. 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 29 below.

Table 29. 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 62 below) the type, proportion, and number of downloaded files, as well as the percentage of the metadata extracted.

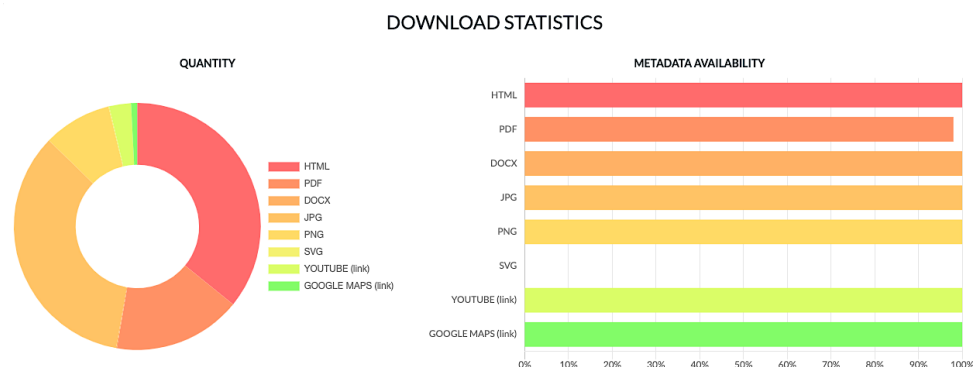


Figure 64. The information provided to the user by making use of the "Details" functionality

- 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.

14. 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]:
print("The number of records in the clean dataframe version is:\033[1m",
      FertinnowaPdfsDf.shape[0], "\033[0m")

# <p><b>STEP 2</b></p>: A <code>list</code> containing 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</b></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></p>: 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
```

items of the `mostFrequentDocumentOutputs` `list`). 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)
```

<p>Given the fact that both the words "book" and "handbook" contain the term "book", we find instances of the "handbook" pdf type in the "book" list. Therefore, we need to remove 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>STEP 4: We export each dataframe of pdf output type as a <code>csv</code> file. For instance, all the records related to Fertinnowa deliverables harvested by the crawler are contained (and stored) in the Fertinnowadeliverable_pdfs.csv 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>STEP 5: We extract all the pdf documents, harvested by the crawler, that do not match with the items of the mostFrequentDocumentOutputs <code>list</code> as a separate <code>dataframe</code> (i.e. the otherPdfTypesDf <code>dataframe</code>).</p>

In[24]:

```
otherPdfTypesDf =
```

```
FertinnowaPdfsDf[~FertinnowaPdfsDf["WCFILEPATH"].isin(mostFrequentDocumentOutputfileUrlList)]
```

In[25]:

```
otherPdfTypesDf
```

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

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.</p>

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

for rowNum in range(length):

    languageShortName = detect(otherPdfTypesDf["WCFILEPATH"].iloc[rowNum].split("/)[-1])
    if languageShortName == "en":

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

print()

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

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

# In[27]:
print(text)

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

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

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

    print(keywordInfo)

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

# 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
```

<p>STEP 7</p>: Some further pdf documents will be considered by taking account of the keywords/terms with a score higher than 19% 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 other pdf files are stored into a separate <code>csv</code> file named "Fertinnowa_other_pdfs.csv".</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")
```

15. 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 identify all the <code>csv</code> files in a directory, i.e. the <code>csv</code> files containing info about the different pdf output types identified per Thematic Network.</p>

In[1]:

```
from os import listdir
import glob
import csv
import pandas as pd
```

<p>The code script in the following block reads all the <code>csv</code> files contained in a specified directory. The directory where the file search takes place is specified with 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 type of the files to be retrieved is specified with the help of the respective extension (in our case ".csv"), which passed as a value to the <code>suffix</code> variable. We then iterate over the list of the files contained in the directory and print the names 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 total number of pdf files contained in the csv files available from the directory. This operation is executed to validate that the <code>csv</code> files available in the directory do contain 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 total number of the pdf files before removing 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():
```



```

        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]:

```

```

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]:

```



```

    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("\n\nThe total number of\033[1m other \033[0mpdf documents is:\033[1m",
            len(csvFileUrlsDict["other"]),"\033[0m")

```

16. 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> "youTubeVideosToSelectFrom" <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>

```



```
# 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]
```

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

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

```
duplicateFreeYouTubeVideosDf = duplicateFreeYouTubeVideosDf.rename(columns =
columnTitlesDict)
```

<p>Resetting the indexes of the "`duplicateFreeYouTubeVideosDf`" `dataframe`.</p>

```
# In[14]:
duplicateFreeYouTubeVideosDf.reset_index()
```

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

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

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

```
# In[16]:
youTubeVideosDf["views"].describe()
```

<p>Selecting those YouTube videos (from the list of the duplicate free YouTube video URLs) that have a number of views higher than the median (i.e. 118.5)</p>

```
# In[17]:
selectedDuplicateFreeYouTubeVideosDf =
duplicateFreeYouTubeVideosDf[duplicateFreeYouTubeVideosDf["views"] > 118.5]
```

<p>Total number of duplicate free YouTube videos to integrate into the EURAKNOS repository.</p>

```
# In[18]:
selectedDuplicateFreeYouTubeVideosDf.shape[0]
```

<p>Resetting the indexes of the "`selectedDuplicateFreeYouTubeVideosDf`" `dataframe`.</p>

```
# In[19]:
selectedDuplicateFreeYouTubeVideosDf.reset_index()
```

<p>Exporting the "`selectedDuplicateFreeYouTubeVideosDf`" `dataframe` as a `csv` file.</p>

```
# In[20]:
selectedDuplicateFreeYouTubeVideosDf.to_csv("selectedDuplicateFreeYouTubeVideosDf.csv", sep =
";", index = False, header = True)
```

<p>Exporting the "`selectedDuplicateFreeYouTubeVideosDf`" `dataframe` as an `excel` file.</p>

```
# In[21]:
selectedDuplicateFreeYouTubeVideosDf.to_excel("selectedDuplicateFreeYouTubeVideosDf.xlsx",
index = False, header = True)
```

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>STEP 1: Loading and reading 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 total number of the pdf files 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 first five rows of the pdfsDf <code>dataframe</code>.</p>

In[4]:

```
pdfsDf.head()
```

<p>STEP 2: Searching for any pdf files (from the bulk of the pdfs harvested by the crawler) that may contain the word "presentation" 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>STEP 3: Creating a new <code>dataframe</code> containing all the pdf records related to presentations.</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>Displaying all the rows of the <code>dataframe</code> holding presentations in a pdf format.</p>




```
# 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"]:
```

```
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")
```

