Check for updates

SOFTWARE TOOL ARTICLE

REVISED Knowledge sharing and discovery across heterogeneous research infrastructures

[version 3; peer review: 2 approved, 1 not approved]

Siamak Farshidi[1], Xiaofeng Liao[2], Na Li[2], Doron Goldfarb[3], Barbara Magagna[3], Markus Stocker[4], Keith Jeffery[5], Peter Thijsse[6], Christian Pichot[7], Andreas Petzold[8], Zhiming Zhao[2]

[1]Department of Information and Computer Science, Utrecht University, Utrecht, The Netherlands
[2]MultiScale Networked Systems (MNS), University of Amsterdam, Amsterdam, Netherlands, 1098 XK, The Netherlands
[3]Environment Agency Austria, Vienna, Austria
[4]TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
[5]British Geological Survey, London, UK
[6]MARiene Informatie Service, Nootdorp, The Netherlands
[7]French National Institute for Agriculture, Food, and Environment, Paris, France
[8]Forschungszentrum Juelich GmbH, Jülich, Germany

v3 **First published:** 14 Jun 2021, **1**:68
https://doi.org/10.12688/openreseurope.13677.1
Second version: 22 Nov 2021, **1**:68
https://doi.org/10.12688/openreseurope.13677.2
Latest published: 06 Jun 2023, **1**:68
https://doi.org/10.12688/openreseurope.13677.3

Abstract

Research infrastructures play an increasingly essential role in scientific research. They provide rich data sources for scientists, such as services and software packages, via catalog and virtual research environments. However, such research infrastructures are typically domain-specific and often not connected. Accordingly, researchers and practitioners face fundamental challenges introduced by fragmented knowledge from heterogeneous, autonomous sources with complicated and uncertain relations in particular research domains. Additionally, the exponential growth rate of knowledge in a specific domain surpasses human experts' ability to formalize and capture tacit and explicit knowledge efficiently. Thus, a knowledge management system is required to discover knowledge effectively, automate the knowledge acquisition based on artificial intelligence approaches, integrate the captured knowledge, and deliver consistent knowledge to agents, research communities, and end-users. In this study, we present the development process of a knowledge management system for ENVironmental Research Infrastructures, which are crucial pillars for environmental scientists in their quest for understanding and interpreting the complex Earth System. Furthermore, we report the challenges we have faced and discuss the lessons learned during the development process.

Open Peer Review

Approval Status ✓ ✗ ✓

	1	2	3
version 3 (revision) 06 Jun 2023			✓ view
version 2 (revision) 22 Nov 2021	✓ view		? view
version 1 14 Jun 2021	✗ view	✗ view	

1. **Robert Huber**, Swiss Federal Institutes of Technology Zurich ETHZ, Zürich, Switzerland
2. **Rebecca Koskela**, University of New Mexico, Albuquerque, USA
Ronin Institute, Montclair, USA
3. **Giacomo Marzi**, University of Trieste, Trieste, Italy

Keywords
Knowledge base, knowledge management, search engine, research infrastructure, software development lifecycle

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Horizon 2020 gateway.

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

This article is included in the Data Science gateway.

This article is included in the Research on Research gateway.

This article is included in the Cloud-based Technologies collection.

Corresponding authors: Siamak Farshidi (s.farshidi@uu.nl), Zhiming Zhao (z.zhao@uva.nl)

Author roles: Farshidi S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Liao X**: Data Curation, Visualization, Writing – Review & Editing; **Li N**: Investigation; **Goldfarb D**: Software, Writing – Review & Editing; **Magagna B**: Investigation; **Stocker M**: Resources, Writing – Review & Editing; **Jeffery K**: Resources, Writing – Review & Editing; **Thijsse P**: Resources, Writing – Review & Editing; **Pichot C**: Resources; **Petzold A**: Resources; **Zhao Z**: Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement Nos [824068] (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research [ENVRI-FAIR]), [862409] (Blue-Cloud: Piloting innovative services for Marine Research & the Blue Economy [BLUECLOUD]), [825134] (smART socIal media eCOsytstem in a blockchaiN Federated environment [ARTICONF]), and the LifeWatch ERIC project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860627.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Farshidi S *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Farshidi S, Liao X, Li N *et al.* **Knowledge sharing and discovery across heterogeneous research infrastructures [version 3; peer review: 2 approved, 1 not approved]** Open Research Europe 2023, **1**:68 https://doi.org/10.12688/openreseurope.13677.3

First published: 14 Jun 2021, **1**:68 https://doi.org/10.12688/openreseurope.13677.1

REVISED **Amendments from Version 2**

We made the following changes based on your feedback:

Sample Selection and Participant Details: We provided additional information to clarify the representativity of the sample and participant details. We conducted a webinar with experts from 26 research infrastructures involved in the ENVRI-FAIR project. The thirty-five domain experts who participated in the webinar were carefully selected based on their expertise and years of experience within their respective domains.

Online Survey Clarification: We acknowledged the lack of information regarding the online survey and apologized for the oversight. We clarified that the online survey was conducted using the Mentimeter platform after the webinar. The survey consisted of specific questions aimed at gathering information about the information needs, queries, and expectations of the ENVRI community. All responses were collected, analyzed, and prioritized based on recurring statements and common themes.

Table for Practical Implications: In response to the comment about practical implications, we agreed to provide a more concise and easily accessible summary. We created a table summarizing the main implications of the ENVRI-KMS project. The table highlights important aspects such as software engineering principles, software architecture decisions, database system selection, programming language ecosystems, cross-referencing of knowledge bases, the role of Ontowiki, versioning and backup strategies, continuous development efforts, and collaboration within the ENVRI community. This addition aims to enhance the usability of our paper for practitioners, providing a quick reference for applying our research in real-world scenarios.

Any further responses from the reviewers can be found at the end of the article

1 Introduction

Contemporary societies are faced with a new challenge for the 'globe' – the changing of the world's climate[1]. Climate change is unpredictable in its form and scope and is long-term rather than immediate in its impacts and remedies. Any practical solutions lie beyond any act of national will, requiring the international collaboration of unprecedented dimension and complexity. While an effective solution to address the challenge would play out over several decades, it is required to be shaped and put in place over the next few years[2].

Climate change has been identified as a major environmental problem for humanity by the United Nations and the European Union. Research is expected on potential scenarios on climate change that will drastically affect natural ecosystems, plants, habitat, and animals, contributing to speedup in biodiversity loss in some areas. The impacts would have knock-on effects for many communities and sectors that rely on natural resources, including agriculture, fisheries, fuels, tourism, and water. Additionally, the ocean plays a central role in regulating the Earth's climate[3].

Assessments of climate change and their association with the driving forces must be based on trustworthy and well-documented observations. This is a difficult task due to the many interactions that exist between the atmosphere, soil, and hydrosphere. The resulting impacts on ecosystems all need particular and focused, high-quality long-term observations. This forces us to have better observations and data on these essential preconditions to inform decision-makers better to take the measures necessary to maintain a thriving society[4].

Research Infrastructures (RIs) are vital for providing the required information to support science and fact-based policy development. Research infrastructures, including advanced computing and storage infrastructure, in environmental science, are essential requirements for scientists in this domain to understand and analyze the sophisticated earth system[5]. Interdisciplinary research communities and research infrastructures collaborate with the neighboring disciplines, namely atmosphere, biosphere, hydrosphere, and geosphere. Internal cooperation across different realms resulted in the formation of distinct research traditions, skills, and cultures. The interconnected essence of the earth system, on the other hand, requires the scientific community to transcend well-established divisions between disciplines and domains and work toward a common understanding of the world as a whole[6].

The data from the ICOS[1] Research Infrastructure, for example, aids climate science by informing scientists and the general public on natural and human-caused greenhouse gas emissions and uptake from the ocean, land ecosystems, and atmosphere. It gives access to high-quality data processed by the Thematic Centers as raw, near real-time, and final quality-controlled data and supplemented with elaborated (model) data and analyses, almost always licensed under a CC4BY[2] license. The IAGOS[3] research infrastructure provides atmospheric composition information, including greenhouse gas observations from commercial aircraft. IAGOS data are used by researchers worldwide for process studies, trend analysis, validation of climate and air quality models, and spaceborne data retrievals validation. Aerosols and their precursors are monitored by the ACTRIS[4] research infrastructure. Aerosols have a significant impact on the Earth's radiation balance, and consequently, the climate. Their levels are inextricably linked to human activity and emissions. Such RIs are part of a more significant worldwide effort to advance science-based, high-quality observations that will help people in making better decisions. As a result, the data and procedures are based on international, typically community-based standards.

Typically, RIs are domain-specific and are not connected, so that interoperability can be a critical issue for scientists involved in interdisciplinary research projects. Moreover, researchers/developers are not knowledgeable in all domains, so a knowledge management system is required to capture cross-domain environmental knowledge automatically and enable researchers to access data, software tools, and services from different

[1]Integrated Carbon Observation System (ICOS)[7]

[2]Creative Commons Attribution 4.0 international license (CC4BY)

[3]In-service Aircraft for Global Observing System (IAGOS)[8]

[4]Aerosol, Clouds, and Trace Gases Research Infrastructure System (ACTRIS)[9]

sources and integrate them into cohesive experimental investigations with well-defined, replicable workflows for processing data and tracking results' provenance. Accordingly, a knowledge management system is required for research communities that (1) discover cross-domain knowledge and capture them automatically, (2) answer any domain question without any limitation to its current search space, (3) deal with noisy sets of retrieved documents, likely consisting of many irrelevant documents and semantically and syntactically ill-formed documents, (4) have an advanced search engine to interpret and reformulate queries by information retrieval algorithms, (5) return a set of recommended solutions (answers) based on the retrieved documents, and (6) visualize its outcomes to facilitate the data analysis for research communities. This paper introduces a novel Knowledge management system, called ENVRI-KMS, to meet the ENVRI research community's requirements and make the research assets Findable, Accessible, Interoperable, and Reusable (FAIR[10]) for the community.

The rest of this study is structured as follows: Section 2 introduces knowledge discovery and sharing challenges, formulates the design research questions, and elaborates on the research methods that have been employed to capture knowledge regarding the ENVRI-KMS.

Section 3 outlines the development process of the ENVRI-KMS. Section 3.1 explains the online survey that we conducted to collect requirements of the ENVRI-KMS. Section 3.2 shows the use case scenarios that we identified based on the survey. Section 3.3 introduces the design decisions that we made to design the ENVRI-KMS architecture. Section 4 elaborates on the selected technologies that we employed to develop the ENVRI-KMS and demonstrates part of the current implementation. Section 5 analyzes the requirements and maps them to the survey questions and design research questions based on the participants' responses. Section 6 highlights the challenges and lessons learned during the development process of the ENVRI-KMS. Section 7 positions the proposed approach in this study among the other knowledge management approaches in the literature. Finally, Section 8 summarizes the proposed approach, defends its novelty, and offers directions for future studies.

2 Challenges regarding knowledge sharing and discovery

In this paper, we present a novel knowledge management system, called ENVRI-KMS, to meet the ENVRI research community requirements and make the research assets Findable, Accessible, Interoperable, and Reusable (FAIR[10]) for the community. The ENVRI-KMS is a Knowledge-as-a-Service (KaaS) for ENVRI-FAIR research communities to document the development and operation processes of RIs and support them with their engineering and design decisions. In general, the ENVRI-KMS should (1) ingest technical results from ENVRI-plus, FAIR assessment[5], the key sub-domains, and other tasks using a formal language for knowledge representation and proven semantic technologies; (2) provide services and tools to enable RI developers and data managers to browse, search, retrieve and compare RI technical statuses and technical solutions to development problems via available content; (3) provide content management tools for specialists in the ENVRI community to ingest new knowledge and control the quality of content; (4) also provide interfaces to other existing semantic resources, e.g., the service catalog of a future ENVRI-HUB[6], to enhance knowledge discovery and cross-RI search, between knowledge services and the online presence of ENVRI resources.

A significant number of advanced research infrastructures, such as ICOS[7] and IAGOS[8], are available to facilitate the access of researchers to research assets (e.g., data products, best practices, data service design decisions, software tools, and services). Such research assets are scattered among a wide range of heterogeneous knowledge resources[5]. Furthermore, operational policies of different domains typically restrict interoperability and accessibility of multidisciplinary research projects. Additionally, technical reports about architectural design, service interfaces, selections of metadata standards, controlled vocabularies, and ontologies are not shared effectively. Accordingly, the main design research question in this study is *"How to enable a domain-specific research community with their asset discovery challenges based on the FAIR principles?"*

As knowledge is scattered in a wide range of literature, forums, documentation, and tacit knowledge of domain experts, the following design research questions should be addressed to capture knowledge systematically: *RQ*1: Which sources of knowledge should be employed to build the search space of the the ENVRI-KMS? *RQ*2: How to capture knowledge regarding RI's research assets systematically? *RQ*3: How to keep the knowledge base of the ENVRI-KMS always up to date? *RQ*4: How to store and retrieve acquired knowledge when it is needed by the ENVRI-FAIR communities? *RQ*5: How to evaluate the recommended solutions of the ENVRI-KMS?

This study employs a mixed research method based on design science research, surveys, and documentation analysis to capture knowledge regarding knowledge management systems and answer the design research questions. The research approach for creating the proposed knowledge management system, called ENVRI-KMS, is Design Science, which addresses research by building and evaluating artifacts to meet identified business needs[11] in an iterative process[12]. Furthermore, we designed a survey form and asked several of our colleagues to critique it. We conducted an online survey in the context of 26 research infrastructures to collect their functional requirements and quality concerns. In total, 35 domain experts participated in the research to assist us with the ENVRI-KMS development life cycle and the requirement analysis phase. Moreover, to develop the ENVRI-KMS, we reviewed webpages, whitepapers, scientific articles, fact sheets, technical reports, product wikis, product forums, product videos, and webinars to collect data. A structured coding procedure

[5]FAIR data are data that meet principles of findability, accessibility, interoperability, and reusability.

[6]ENVRI-HUB is a one-stop-shop for access to environmental data and services provided by the contributing research infrastructures.

is employed to extract knowledge from the selected sources of knowledge.

Knowledge management systems employ problem-solving techniques, and knowledge discovery approaches to answer particular questions[13,14]. Knowledge discovery is the process of extracting useful and hidden information[15]. A variety of Knowledge management systems have been introduced in literature[16–18]. Most of the existing knowledge management systems in the literature bound to a limited search space and optimized to address questions in a particular context. Each question-answer-context tuple is well-formed, standardized, and generated rising from the context in which the question and answer were extracted.

3 ENVRI knowledge management system

The ENVRI-KMS[19] is a cluster-level knowledge base that allows different ENVRI users, such as RI developers and data managers, to effectively share their technical practices, identify common data and service requirements and design patterns, and facilitate the search and analysis of existing RI solutions for environmental RI interoperability challenges.

3.1 Requirement analysis

We organized a webinar to facilitate an online survey within the context of 26 research infrastructures actively engaged in the ENVRI-FAIR project[20]. The primary objective was to gather comprehensive functional requirements and quality concerns[21] from these infrastructures. In pursuit of this goal, a total of 35 domain experts were carefully selected to participate in the research, specifically to contribute to the ENVRI-KMS development life cycle and requirement analysis phase. The selection process took into account their extensive expertise and considerable years of experience in their respective domains. On average, the participants possessed over ten years of domain-specific experience, rendering them highly attuned to the potential challenges researchers within their communities and fields might encounter while performing their daily tasks.

To initiate the webinar, we introduced the potential functionalities of the ENVRI-KMS, drawing upon a literature review and internal meetings conducted with a select group of domain experts. We subsequently employed an online survey tool called Mentimeter[22], leveraging its capabilities to distribute a virtual questionnaire encompassing the following inquiries:

(Q1) What specific information do you typically seek from the ENVRI community?

(Q2) What typical queries would you pose to the ENVRI-KMS?

(Q3) How do you currently navigate the process of accessing information from the ENVRI community?

(Q4) From your perspective, which aspects of knowledge management system functionality prove most beneficial to you?

(Q5) What functionalities do you anticipate in the next version of the ENVRI-KMS?

Following the completion of the online survey, we meticulously collected and meticulously analyzed all responses. To prioritize the requirements, we conducted a thorough examination of the frequencies at which similar statements or themes emerged, allowing us to identify patterns and establish a hierarchy of importance.

It is noteworthy that the participants in our study were specifically selected from the Atmosphere, Ecosystem, Marine, and Solid Earth domains. We requested the research infrastructures involved in the ENVRI-FAIR project to nominate individuals who possessed expertise in their respective domains, ensuring a comprehensive understanding of their concerns and requirements. This deliberate selection process aimed to capture a representative sample of participants deeply knowledgeable about their domains, fostering the integrity and validity of our study.

Next, we have collected all responses and prioritized them based on analyzing the frequencies of similar statements[7].

Table 1 shows the requirements that we have extracted from the experts' responses. The color-coding indicates the importance of each requirement according to the number of responses that signified it.

The initial user stories for the ENVRI-KMS mainly focus on the data manager, RI service, or Virtual Research Environment (VRE)[23–25] developers, e.g., for enabling a developer to check the existence or details of data management solutions from different RIs. Accordingly, the following key technical requirements have been identified to design and implement the ENVRI-KMS:

Compatible with semantic web technologies. As the most common type for knowledge storage, representation, reasoning, the support of Resource Description Framework (RDF) is the core requirement in the design and development of the ENVRI-KMS. This requirement can include the following specific options: RDF import/export, RDF storage, owl import, SPARQL, and GeoSPARQL support. It is acknowledged that while providing many advantages, especially in the context of integrating and operating on heterogeneous knowledge sources and of linking to existing external resources, RDF, but also the overall concept of operating on a non-monolithic set of data collections, comes with specific limitations as well, such as lack of support for referential integrity. Nevertheless, it is assumed that the ENVRI-KMS content's nature is rather non-volatile, shifting this aspect into the background.

Semantic search and query functionality. An interface for searching and discovering ENVRI-KMS content should be provided; this could be the conventional keyword-based search or faceted search. A semantic search function is further expected to permit search based on 'similar' or 'related' terms across multiple ontologies/controlled vocabularies rather than strict adherence to a single controlled vocabulary or keyword set[26].

[7]We have published the responses of the domain experts who participated in the survey besides the data analysis phases on Mendeley Data[21].

Table 1. shows the requirements that we have extracted from the experts' responses. The color-coding indicates the importance of each requirement according to the number of responses that signified it.

ID	Requirements	doamin experts
R01	*The ENVRI-KMS should include all prospective RIs, datasets, repositories, best practices, service catalogs, and design decisions in its search space.*	54.29%
R02	*Contact lists of people in charge of specific tasks (authors, researchers, developers, etc.) must be provided by the ENVRI-KMS.*	54.29%
R03	*The ENVRI-KMS is required to employ a set of assessment criteria, such as FAIRness criteria, to evaluate search space entities.*	85.71%
R04	*Private, public, open-source, or premium search space entities should be indicated in the ENVRI-KMS.*	37.14%
R05	*Documentation, technical solutions, configurations, and compatible combinations should all be recommended by the ENVRI-KMS.*	34.29%
R06	*The ENVRI-KMS should provide technical discussions through Q&A forums and invite domain experts to participate.*	45.71%
R07	*Ontologies and semantic search should be supported by the ENVRI-KMS.*	60.00%
R08	*Multilingual inquiries should be supported by the ENVRI-KMS.*	8.57%
R09	*The ENVRI-KMS should be able to search for source code and provide suitable solutions to technical issues.*	14.29%
R10	*The contents of the RI websites have to be searchable through ENVRI-KMS (similar to what Google search engine does.)*	60.00%
R11	*The user interface (UX/UI) of the ENVRI-KMS should be similar to typical search engines.*	57.14%
R12	*The ENVRI-KMS should be able to connect to endpoints and support SPARQL queries.*	28.57%
R13	*High performance and availability have to be two essential quality attributes of the ENVRI-KMS.*	68.57%
R14	*APIs for connecting to virtual research environments (ENVRI-HUB) should be available through the ENVRI-KMS.*	34.29%
R15	*Automated knowledge ingestion should be possible with the ENVRI-KMS.*	28.57%
R16	*The outcomes and contents of the ENVRI-KMS should be visualized.*	48.57%
R17	*The ENVRI-KMS should be able to search numerous image categories (plots, etc.) and support image search.*	42.86%
R18	*Domain experts must be able to analyze the contents of the ENVRI-KMS using assessment techniques.*	48.57%
R19	*The ENVRI-KMS should support manual knowledge ingestion.*	11.43%
R20	*The knowledge base of the ENVRI-KMS should always be kept up-to-date.*	31.43%
R21	*One of the search criteria in the ENVRI-KMS should be the geolocation of datasets.*	48.57%
R22	*The metadata of the search space entities, such as datasets and APIs, should be available through the ENVRI-KMS.*	20.00%
R23	*Continuous integration and continuous delivery (CI/CD) should be supported by the ENVRI-KMS.*	28.57%
R24	*Different user groups, such as researchers, knowledge curators, developers, and high-level managers, should receive suggestions from the ENVRI-KMS.*	17.14%
R25	*The ENVRI-KMS should categorize and classifies its knowledge base contents.*	37.14%

Open and flexible knowledge ingestion. Due to the variance of source types in the ENVRI community, various methods should be supported for knowledge acquisition, like form-based manual RDF ingestion, Questionnaire-based RDF triple generation, existing RDF integration, structured and unstructured information transformation, etc. Specific measures should be considered to facilitate non-technical users straightforwardly adding knowledge.

Provenance and version control of the knowledge. Considering the typical case where multiple users contribute to the ENVRI-KMS, provenance is of fundamental importance for monitoring and tracking issues, for example, enabling the third party to reproduce the scientific workflow for an authority to audit the whole process. This primarily refers to tracking individual additions, deletions, and updates and their administration, i.e., approval, rejection, and reversion.

User-friendly and customizable user interface. A clear and straightforward user interface is needed to fulfill their objectives, like query, semantic search. Different user interfaces should be offered to meet the requirements of the general public and professional users.

Scaling and increasing performance. A choice between centralized or distributed storage should be considered to tackle the growing size of the ENVRI-KMS. Also should be considered includes the dynamic resource scheduling facing concurrent search/query requests. Other features like collaborative editing are required to enable comments on contributions by other users.

API interface. An application programming interface (API) abstraction layer can help make knowledge accessible through applications to facilitate knowledge via APIs.

Among such technical requirements, the ENVRI-KMS should play a key role in the ENVRI communities to develop FAIR data services and share their best practices.

3.2 Use case scenarios

Based on the survey we conducted (see section 3.1), we identified the following four types of users (see Figure 1) of the ENVRI-KMS:

(1) **End users** may use the ENVRI-KMS to find answers to their general questions about available sources of data, services, and tools, and to use the discovered information to perform further research activities using the other tools like Virtual Research Environments or services like the RI catalogs of data or services.

(2) **RI managers** or operators may use the ENVRI-KMS to check the status of the FAIRness of specific repositories or update the state of their RIs. The update process often needs the output of other third-party tools, e.g, FAIRness assessment tools.

(3) **RI developers** may use the ENVRI-KMS to check the existing technologies, e.g., those development results in the ENVRI portfolio or the demonstrators prepared for some known FAIRness gaps. They can also publish or update the technical descriptions using ENVRI-KMS components, such as an online description form.

(4) **Knowledge curator and knowledge base operators** may use the ENVRI-KMS to ingest content from new sources and respond to the possible errors that occurred during the ingestion or the operation.

3.3 Conceptual architecture

Based on the use case scenarios (see Figure 1), we design the key components of the ENVRI-KMS from the conceptual point of view. Note, the architecture is designed based on the Open Distributed Processing (ODP) framework[27–30]. Figure 2 shows the key components via three layers:

The **interface layer** atop contains components dealing with user-related activities. The ENVRI-KMS will be an open system for community users; the user management component is not for acquiring and processing users' personal information but more for providing customized user support based on their interaction or contexts. A user can log in to the system using an open identity provider. The User Interface (UI) components are the application parts that allow users to interact with it. It can be formatted and rendered into various presentations to address different users' requirements. Additionally, it validates and collects required data from users.

The **service layer** abstracts the functionality that the ENVRI-KMS offers; it can be roughly split into three sub-layers, namely:

(1) The Application sub-layer provides customized application logic (e.g., FAIRness Gap Analysis, Engineering support, or discovery knowledge from ENVRI community) based

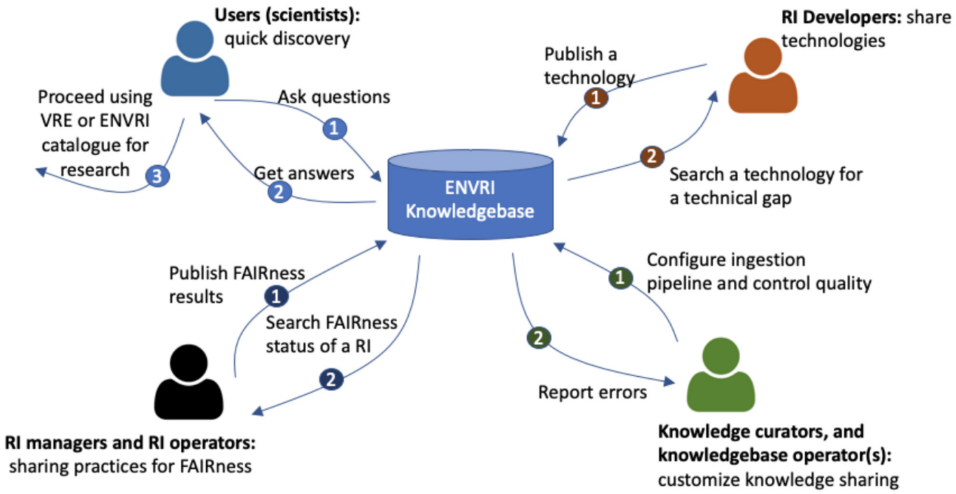


Figure 1. Shows an enterprise view of the ENVRI-KMS. The enterprise view highlights the key stakeholders (namely communities in the Open Distributed Processing (ODP) term) and their interaction scenarios with the ENVRI-KMS. The numbered circles indicate the possible orders of the interactions.

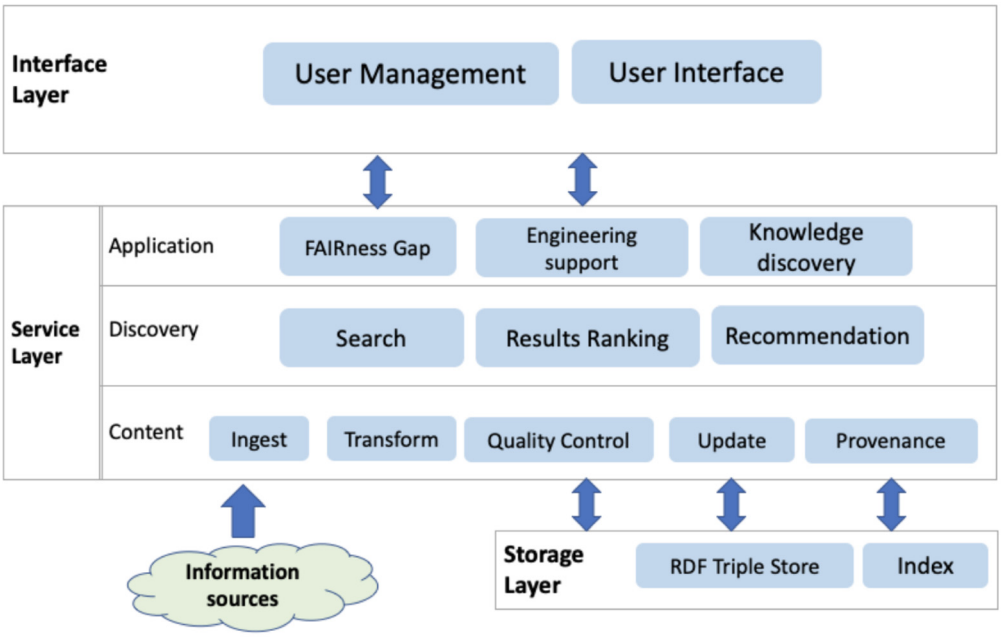


Figure 2. Shows the layered architecture of the ENVRI-KMS that is designed based on the Open Distributed Processing framework.

on the data passed from the underlying discovery sub-layer those results up to the User Interface Component.

(2) The Discovery sub-layer provides the functionality for searching the ENVRI-KMS, ranking the results, and recommending relevant content.

(3) The Content sub-layer provides functionality for managing the content in the ENVRI-KMS, typically in a pipeline covering: ingesting information, the transformation from information to knowledge, quality control of the knowledge generation, CRUD (Create, Read, Update, Delete) of the ENVRI-KMS content, and the provenance of these activities.

The **storage layer** at the bottom is responsible for data storing and access. The data storage options needed in this project include RDF Triple Store and Inverted Index. Currently, information collected in the ENVRI-KMS consists of two main parts, as illustrated in Figure 3. The structured data in the ENVRI-KMS is based on RDF and mainly includes: (1) OIL-e (ontology of the ENVRI Reference Model) based ENVRI RI description, (2) description of the service portfolio from the previous project, and the possibly new ones in ENVRI-FAIR, (3) FAIRness principles and the results of assessing the ENVRI research infrastructures, and (4) demonstrators for tackling the known gaps, e.g., those being identified during the FAIRness assessment.

The versions of the structure data currently can be managed via version control systems. Currently, GitHub is used. The dynamic data in the ENVRI-KMS will be ingested from different online sources of the ENVRI communities. Figure 4 depicts the necessary information flow of the knowledge ingestion.

(1) A significant amount of relevant information is represented in human-readable form, residing in Wikis, other content management systems, or even static web-pages, in the "offline" text found in various documents such as books, project deliverables, or scientific publications. In the ENVRI-FAIR context, the research infrastructure websites are an excellent resource of related information, including news/events, background knowledge, etc. Similar to ENVRI, ENVRI-FAIR, the community websites also contain lots of related information, like news/events, community introduction, community landscape, project information, progress, etc. These information sources have different formats, such as a webpage, word document, and pdf file.

(2) Another approach to populate the ENVRI-KMS would be to process such free-text information to extract structured, machine-readable information. Named entity recognition would represent the first step in this regard, while the application of more complicated Natural Language Processing operations could be a valuable field of research in its own right.

(3) Information from the available catalogs of data and services. It should be clear that the indexes generated from those sources will not aim to replicate the entire catalogs but provide a quick searching capability for community users. For some RI, such information will be already managed in RDF format and accessible from triplestores.

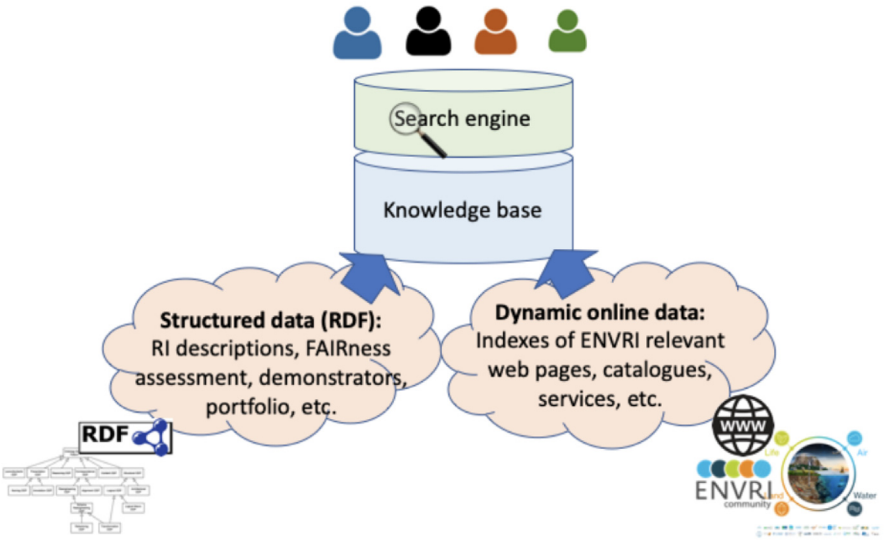


Figure 3. Illustrates the ENVRI-KMS content components. The ENVRI-KMS can be used by end users to search different contents.

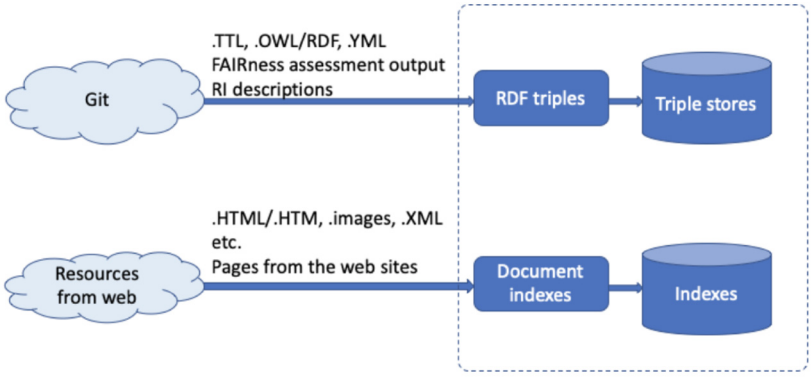


Figure 4. Shows the basic information flow of the knowledge ingestion.

4 Prototype

The ENVRI-KMS development follows an interactive approach, in which the requirements based on the experts' responses (see Section 3.1) have been analyzed, and technical choices have been selected according to the state-of-the-art review published in 31. We use Ontowiki to manage the RDF triples and Open Semantic Search to develop the ENVRI-KMS's search engine in the current prototype. Several tools were developed for ingesting specific knowledge, e.g., a technology description form for describing the service portfolio, interactive graph visualizer for the search results, and dynamic online data ingestion pipeline. These tools will be described in the following sections.

4.1 Knowledge storage

The comparison of existing RDF content management platforms is summarized in 31. Note, we have selected OntoWiki for managing RDF content. The main reasons for this decision were as follows:

(1) Direct operation on RDF triples: Ontowiki can directly operate on a triplestore as the underlying storage layer and provides an API to populate it with RDF.

(2) Integrated User management and statement-level provenance: Ontowiki supports user management with varying permissions and offers a detailed create/update/delete history on the RDF statement level.

(3) Named-graph-based separation of RDF content and administrative data: RDF data ingested via Ontowiki is directly written as-is into the underlying triplestore, while all the administrative statements such as provenance etc., are stored separately.

(4) Plugin-based extensions: Ontowiki offers a framework for developing plugin extensions.

The choice of Ontowiki directly affected the selection of the underlying Triplestore since Ontowiki provides a pre-configured

connector to the Openlink Virtuoso data management system, which members of the ENVRI-KMS team already had experience with from previous projects. The open-source edition of Openlink Virtuoso[32] (Version 7.2.5.1) was therefore deployed for that purpose and configured for Ontowiki (and vice-versa).

4.2 Tools for ingesting knowledge

The population of knowledge bases can take different routes. On the one hand, existing collections of information can sometimes be transformed so that they can be "bulk" imported into the ENVRI-KMS, which includes rearrangements and mappings of existing collections of structured information but potentially also the extraction of structured content from unstructured sources such as free text, which is by no means an easy task considering the complexity in the natural language processing/understanding. On the other hand, it is usually possible to manually add ENVRI-KMS's contents, "fact by fact". However, manual input can be slow, tedious, and error-prone if not supported by dedicated tools. In the context of the ENVRI-KMS, it should be possible to provide content in both ways.

As far as manual data entry is concerned, the system supports the creation of valid RDF data via custom HTML Web forms. They are dynamically created using the RDForms[33] Javascript library based on formal JSON descriptions of the underlying data model. This also includes the specification of constrained SPARQL queries for the dynamic retrieval of menu options to maintain consistent RDF relationships between the described entity and related terminology and other entities already stored in the ENVRI-KMS.

4.3 FAIRness status sharing and gap analysis

To improve the findability, accessibility, interoperability, and reusability of digital research objects for both researchers and machines, the ENVRI-KMS offers a FAIR assessment dashboard[8]. It supports RIs by discovering gaps in FAIR principle implementation at the granularity of their repositories and the discovery of possible technology solutions to address such gaps. For instance, the FAIRness assessment of a particular RI can be modified to indicate whether the repository contains machine-readable provenance information. By selecting an RI, the user interface gives an overview regarding its FAIRness status and gap analysis.

4.4 Ontowiki as a knowledge management platform

OntoWiki is a free and open-source semantic wiki web application that serves as an ontology editor and a knowledge acquisition system. Additionally, Ontowiki is a suitable RDF data management platform. A test instance is configured[34] and slightly customized to use the ENVRI logo and display the ENVRI RSS news feed on the front page. It currently serves as a data gateway for the facts added via forms based on the FAIR assessment dashboard. Ontowiki was found to perform well as RDF "middleware" used to ingest data from the RDF forms.

4.5 Search Engine

In this section, we present a running example of the ENVRI-KMS Search Engine. To facilitate the general users to explore the ENVRI-KMS easily, we build the ENVRI-KMS Search Engine based on the Open Semantic Search's fundamental concepts and components[35]. Figure 5 illustrates the search interface[9].

A searcher can go to the landing page of the ENVRI-KMS (See Figure 5 (a)) directly and enter her search query in the search box and see the results immediately (See Figure 5 (b)). The results and their relevance to RIs and be visualized based on the graph visualization of the ENVRI-KMS (See 5 (c)). Note, the searcher can limit the search space of the ENVRI-KMS to a particular category, such as Webpages and RIs, as well. For instance, Figure 5 (d) shows the dataset search of the ENVRI-KMS.

The ENVRI-KMS can automatically capture, extract, and index knowledge regarding research assets based on the URL of the RIs (See Figure 5 (e)). Additionally, knowledge curators can ingest research assets manually to the knowledge base of the ENVRI-KMS (See Figure 5 (f)). Note, the ENVRI-KMS checks the indexed documents periodically to keep its knowledge base always up-to-date.

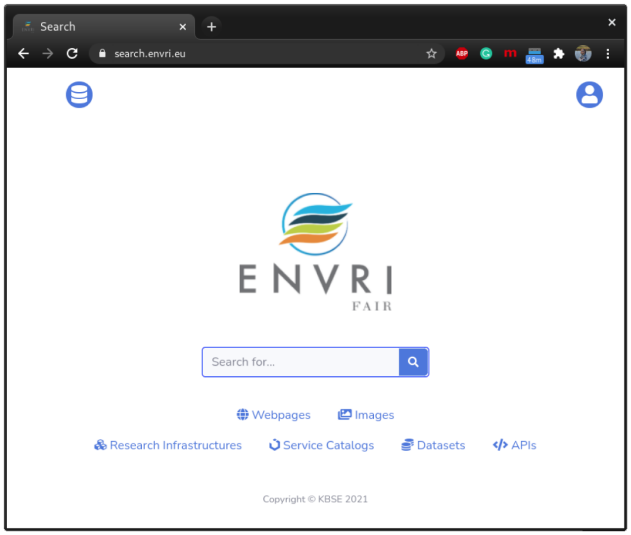
4.6 Operational workflow

This section elaborates on the operational workflow of the ENVRI-KMS[10] and presents its constituent components (See Figure 6). **Research Infrastructures**, such as ACTRIS and ANAEE, are the primary sources of knowledge that contain knowledge assets, including webpages, datasets, APIs, service catalogs, publications, design decisions, best practices, devices, and data provenance. The **Sitemap Extractor** explores and extracts the site structure (the list of URLs) of the RIs. Then, the **Web Crawler** browses the extracted URLs and by employing **NLP & ETL** (Natural Language Processing and Extract/Transform/Load) techniques, such as Named-Entity Recognition (NER) and Relation Extraction (RE), tries to index documents and classify the extracted knowledge. Typically, a web crawler is a bot or software agent. It starts with a

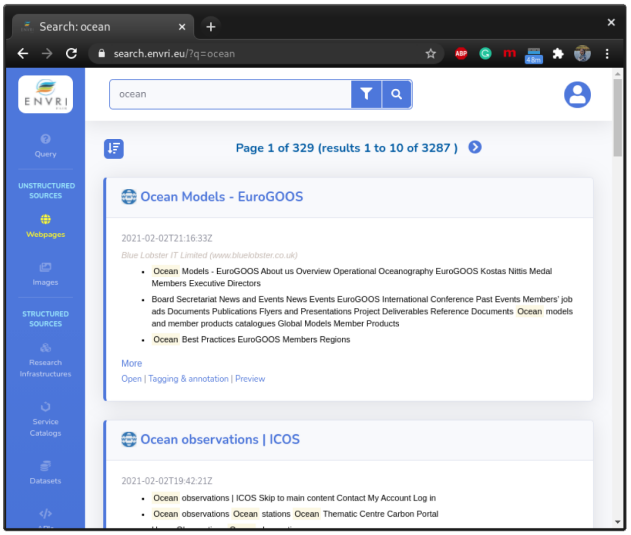
[8]The FAIR dashboard can be accessed through the following link: https://envri-fair.github.io/knowledge-base-ui/

[9]The ENVRI-KMS Search Engine is available through the following link: https://search.envri.eu

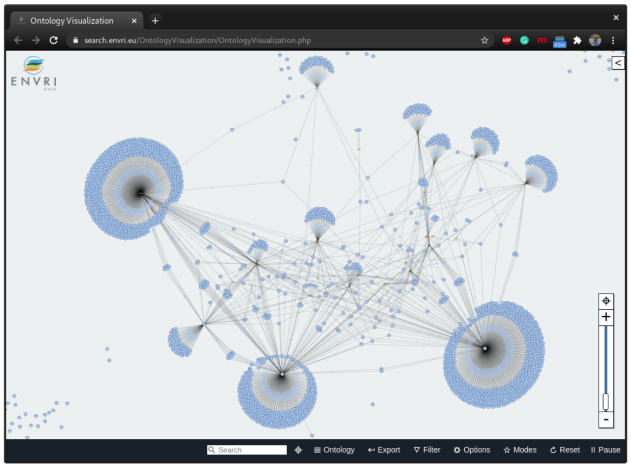
[10]In order to implement the ENVRI-KMS, we have employed the proposed design decisions and solutions that Open Semantic Search[35] and WebVOWL[36] as two open-source projects have been offered.



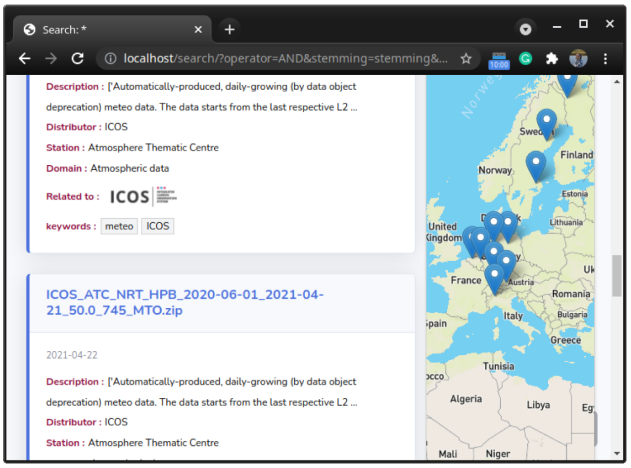
(a) The ENVRI-KMS landing page: `https://search.envri.eu/`



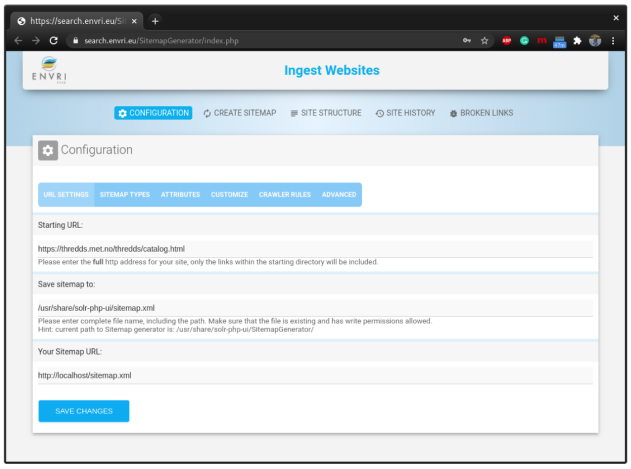
(b) The ENVRI-KMS webpage search results



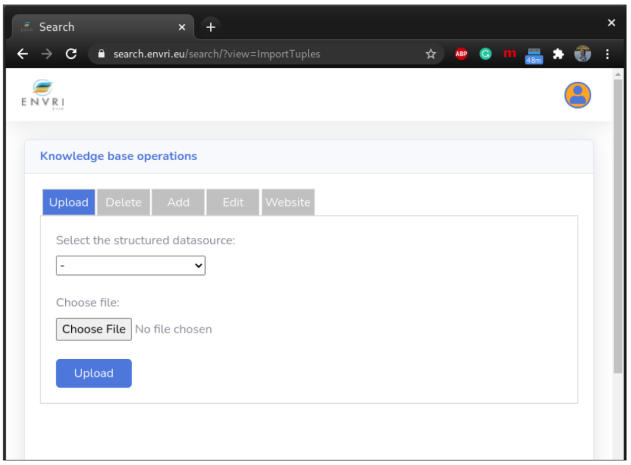
(c) Graph visualization of the ENVRI-KMS



(d) The ENVRI-KMS dataset search results



(e) Automatic knowledge ingestion



(f) Manual web page ingestion (KB operations)

Figure 5. shows a set of functionality of the ENVRI-KMS.

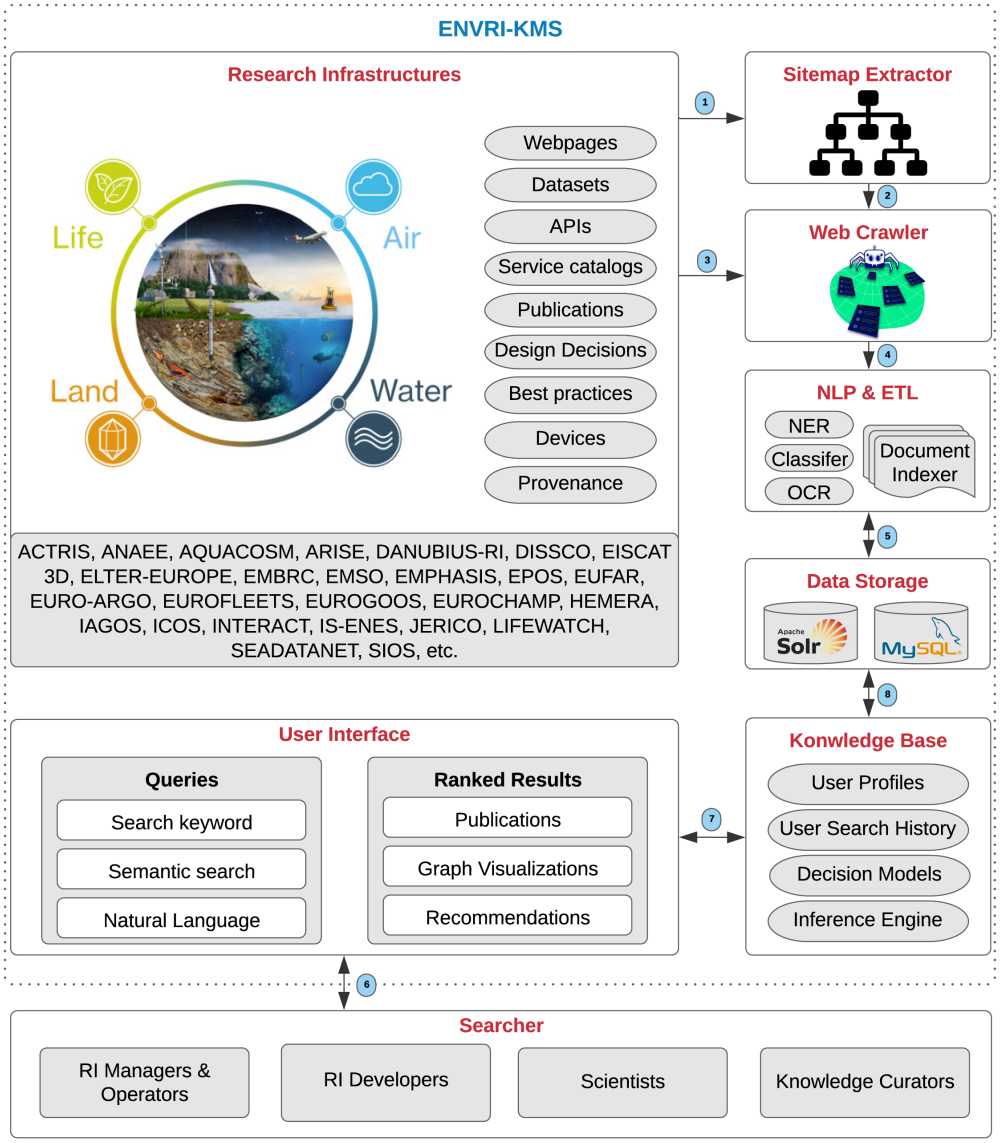


Figure 6. Represents the operational workflow of the ENVRI-KMS.

list of URLs to visit, called the seeds. As the crawler visits such URLs, it identifies all the hyperlinks on webpages by the aim of a sitemap extractor, and adds them to the list of its URLs to visit, called the crawl frontier. For instance, in the knowledge extraction process, the NER and RE approaches identify the entities represented in documents and their relations as fundamental knowledge extraction processes. The extracted knowledge is used to build the knowledge graph in the knowledge base of the ENVRI-KMS. **Data Storage** technologies, including Apache Solr and MySQL, are used to store the acquired knowledge systematically. The **Knowledge Base** of the ENVRI-KMS integrates user profiles, user search histories, decision models (e.g., meta-models), and infers solutions (results) based on searchers' queries. the **User Interface** receives user queries, such as keywords and user stories, and demonstrated the results (e.g., publications, graph visualizations, and recommendations) to the **Searchers the process of extracting useful and hidden information 11** (See section 3.2).

5 Analysis

In this subsection, we reflect on each of the proposed design research questions based on our observations during the development process, the online survey, and documentation analysis.

5.1 Design decisions

We revisit the requirements and analyze the gap for the tools or platforms we investigated in terms of the requirements identified in Section 3.1.

Compatible with Semantic Web technologies. The two storage solutions (Apache Jena and Virtuoso) are triplestores dedicated to storing RDF data, thus fully meeting semantic web technology compatibility requirements. Regarding the knowledge management solutions, as the comparison in 31 indicates, both Semantic Mediawiki and Ontowiki are RDF compatible.

Semantic search and query functionality. Though the several Knowledge management systems investigated (like Ontowiki, Semantic Mediawiki) allow users to explore, search and edit the ENVRI-KMS's content via GUI tools, they still lack easy user experience in terms of the technology required. The original purpose of both Semantic Mediawiki and Ontowiki is a semantic annotation of wiki pages and as a knowledge base editor, respectively.

Open and flexible knowledge ingestion. As shown in 31, knowledge management systems, such as Semantic Mediawiki and Ontowiki, support RDF import, facilitating the ingestion of knowledge. However, to prepare RDF triples or transform the information needed into knowledge, some customized tools needed to be designed and implemented considering the diversity of our project's information sources.

Provenance and version control of the knowledge. As far as the considered knowledge management platforms are concerned, Ontowiki meets the requirements by providing detailed user management and statement-level provenance for RDF data, allowing tracking and potentially editing individual user contributions to the ENVRI-KMS.

User-friendly and customizable user interface. As already analyzed, although the Knowledge management systems provide a GUI for search and query, their targeted users are knowledge base administrators considering the technology barriers. For general users without much technical knowledge of the SPARQL or triplestores, a straightforward user interface for searching and exploration is expected to increase the user experience.

Scaling and increasing performance. Apache Jena Fuseki does not currently support horizontal scale-up, but there are workaround solutions like coordinating the updates from a staging server and publishing (read-only) to external clients. Based on the comparison, it is clear that no one single solution satisfies all the requirements. The optimal solution should be combining existing options, and other software such as Blazegraph could be a candidate.

5.2 Design research questions

To answer the first two design research questions (*RQ*1 and *RQ*2), we have conducted an extensive literature review besides a set of expert interviews with domain experts at the RIs to build the search space (including webpages, datasets, etc.) of the ENVRI-KMS and capture knowledge systematically. The current search space[11] of the ENVRI-KMS includes all research infrastructures which are mentioned on the ENVRI community knowledge base[12]. It is essential to highlight that the search space is not limited to the initial sets and grows automatically. Accordingly, the third design research question (*RQ*3) can be addressed based on the natural language processing approach and Open Semantic Search that we have employed in the implementation of the ENVRI-KMS[31]. To answer the fourth design research question (*RQ*4), we have evaluated a set of technologies that can be employed to store and retrieve data. The last design research question (*RQ*5) is one of the key challenges in this research. We plan to build a community around the ENVRI-KMS and ask the stakeholders, including domain experts, practitioners, and researchers, actively assess the search results and recommendations.

The FAIRness of the ENVRI-KMS should be elaborated in order to answer the study's main design research question. As a result, research assets become **Findable** when adequate metadata characterizes them and a searchable resource efficiently indexes them, allowing them to become recognized and available to potential users. A unique and persistent identifier should also be established so that the data may be referred and mentioned in research communications without ambiguity. The identifier facilitates data discovery and reuse by allowing persistent mapping between data, metadata, and other associated resources. The code or models required to utilise the data, research literature that provides additional insights into the data's development and interpretation, and other related information are examples of related resources. The ENVRI-KMS indexes research assets and assigns them a unique identifier, allowing them to be shared among RIs.

Accessibility means that a human or a machine is given the exact conditions under which research assets can be accessed via metadata. Researchers in research communities can use the ENVRI-KMS to access research assets in accordance with RI policies and regulations.

The ENVRI-KMS search entities are characterized using normative and community-accepted specifications, vocabularies, and standards that define the precise meaning of concepts and qualities represented by the data. **Interoperability** is a crucial aspect of research assets' value and usefulness. It is not only semantic interoperability that is important, but also technological and legal interoperability. Technical interoperability refers to the research assets being encoded using a standard that can be read by all systems involved.

The FAIR principles highlight the necessity for extensive metadata and documentation that match relevant community

[11] https://search.envri.eu/

[12] https://envri.eu/research-infrastructures/

standards and give information about provenance in order for research materials to be reusable. The ability of humans and machines to evaluate and select research assets based on provenance information criteria is critical to their reuse. **Reusability** also necessitates the publication of research assets with a "clear and accessible usage license," which means that the terms under which the assets can be utilized should be transparent to both humans and machines.

Table 2 represents the mapping among the extracted requirements (R01 to R25) based on the responses of the participants to the survey questions (Q1 to Q5) and the design research

Table 2. The mapping among the extracted requirements (R01 to R25) based on the responses of the participants to the survey questions (Q1 to Q5) and the design research question (RQ_1 to RQ_5). Additionally, the last column shows how far we have addressed the requirements up to now.

Requirements		Survey Questions					Research Questions					Addressed?
		Q1	*Q2*	*Q3*	*Q4*	*Q5*	*RQ1*	*RQ2*	*RQ3*	*RQ4*	*RQ5*	
R01	**Completeness of the ENVRI-KMS search space**	X				X	X					Partially
R02	**List of the contact persons**	X		X			X					Not yet
R03	**FAIRness criteria**	X	X		X						X	Yes
R04	**Entitiy types (private, open-source, etc.)**		X				X					Not yet
R05	**Recommendations**	X	X		X	X				X		Partially
R06	**Q&A forums for technical discussions**		X		X						X	Not yet
R07	**Ontologies and semantic search**		X			X				X		Partially
R08	**Multilingual queries**		X						X	X		Not yet
R09	**Source code and API search**		X							X		Not yet
R10	**Search RI website's contents**		X	X						X		Yes
R11	**Standard user interface**			X		X				X		Yes
R12	**SPARQL queries**				X					X		Not yet
R13	**High performance and availability**				X					X		Partially
R14	**APIs to be connected to VREs**				X					X		Partially
R15	**Automatic knowledge ingestion**				X	X		X				Yes
R16	**Visualization**				X					X		Partially
R17	**Image search**				X	X				X		Partially
R18	**Assessment tools**		X		X	X					X	Not yet
R19	**Manual knowledge ingestion**				X			X				Yes
R20	**Updatable knowledge base**					X			X			Yes
R21	**Dataset geolocations**		X			X	X					Yes
R22	**Metadata of the search space entities**			X		X	X					Not yet
R23	**Continuous integration and continuous delivery**					X			X			Yes
R24	**Different user categories**					X					X	Partially
R25	**Categories & classifications**			X	X		X					Yes

question (*RQ*1 to *RQ*5). Additionally, the table shows that more than half of the identified requirements (62%) are at least partially addressed so that the main components of the ENVRI-KMS are functional.

6 Discussion

This section summarizes our observations and highlights several lessons learned during the development process of the ENVRI-KMS.

Software engineers have a broad knowledge of software development technologies, and they apply software engineering principles to develop software products. By employing such engineering principles in the software development lifecycle, from requirement analysis to software implementation and then deployment, they can build customized software products for individual stakeholders. The demand for highly skilled and qualified software engineers seems to have no end. This demand is growing in a changing economic landscape and fueled by the necessity of software development technologies. On the one hand, billions of dollars are spent annually on software products[37] that are produced and maintained by software engineers. On the other hand, business processes are introduced and managed by stakeholders and top-level managers who principally understand businesses[38].

Software architecture deals with the base structure, subsystems, and interactions among these subsystems, so it is critical to the success or failure of any software system[39]. Software architecting can be thought of as a decision-making process in which software architects consider a collection of possible solutions for solving a system design problem and choose the one that is evaluated as the optimal[40]. Software architecture decisions are design decisions that meet both functional and quality requirements in a system. Design decisions are concerned with the system's application domain, architectural patterns employed in the system, Commercial off-the-shelf components, other infrastructure selections, and other aspects needed to satisfy all requirements[41]. According to Avgeriou *et al.*[42], failing to make architectural design decisions during software development has well-known implications, such as costly system evolution, weak stakeholder communication, limited reusability of architectural assets, and poor traceability between specifications and implementation.

In order to make the design decisions to design the architecture of the ENVRI-KMS, we analyzed several alternative tools that could be used to build the fundamental components of the knowledge base. Selecting the right database system(s) (DBMS) was one of those design decisions. The DBMS selection problem is a subclass of the Commercial off-the-shelf (COTS) selection problem, and both problems are a subclass of Multi-Criteria Decision-Making (MCDM) problems[43]. Accordingly, we used a decision support system that has been introduced by Farshidi *et al.*[44] to evaluate potential alternative solutions that we can employ to store and retrieve data. After performing an extensive evaluation, we decided to use *Apache Solr* to indexing the search entities and *MySQL* to manage user profiles and user search histories.

Judging the suitability of a set of technologies, such as programming languages, for developing a knowledge base system is a non-trivial task. For instance, a purely functional language like Haskell is the best fit for writing parallel programs that can, in principle, efficiently exploit huge parallel machines working on large data sets[45]. However, while developing a dynamic website, a software engineer might consider ASP.net as the best alternative, and others might prefer using PHP or a similar scripting language. It is interesting to highlight that successful projects have been built with both: StackOverflow is built in ASP.net, whereas Wikipedia is built in PHP. Furthermore, a software engineer might prefer particular criteria, such as scalability in enterprise applications, whereas other criteria, such as technology maturity level, might have lower priorities.

We realized that we needed to select the right programming language ecosystems for developing the ENVRI-KMS. We used the decision model in the knowledge base of the decision support system[46] to evaluate potential programming languages that we can use to develop the ENVRI-KMS. Note, as mentioned earlier, we use an open-source tool, called open semantic search, in which its backend was implemented in *PHP* and *Python*, as the initial phase of the development process of the ENVRI-KMS. So that the first two solutions for us were these programming languages. However, the decision support system suggested *C#* , *Java*, and *Ruby* as three more alternative solutions. Finally, we decided to continue using Python, as we had more experience with it and found so many open-source projects on Github, which were implemented in Python, that could boost the development process.

Some issues were discovered regarding the cross-referencing of statements between knowledge bases (named graphs). A workaround published in a newsgroup provided a potential fix for static data but would have to be extended for a continuously growing data collection. A possible solution would be to store information that is expected to change/grow, e.g., the entity descriptions and the user terminology collected from the RDF forms, in a typical named graph and to configure Ontowiki filters for its efficient navigation while storing more static content, such as external ontologies, in separate graphs. While Ontowiki supports flexible navigation and data editing at the RDF statement level, the interface is arguably not

appropriate for the vast majority of RI managers or developers. We conducted some experiments with the atmospheric domain, but RIs did not engage with the user interface. This is to be expected since Ontowiki relies on a good understanding of the RDF data model. Moreover, presenting information at the RDF statement's granularity is typically inadequate for high-level information needs, e.g., discovering FAIR gaps in the data centers of an RI. We thus suggest that Ontowiki can act as an RDF-based middleware that powers high-level user applications and services. A critical aspect of using Ontowiki to manage the generated RDF data will be the question of versioning. While built-in features such as the statement-level provenance in principle allow detailed tracking of changes/revisions of the provided data, a backup strategy using external means should be considered as well. One straightforward step would be to export complete RDF dumps of the provided content in regular intervals and to track their versions in source code repositories such as Github.

Table 3. The bullet points in the table are a concise summary of the practical implications based on the lessons learned during the development process of the ENVRI-KMS.

Practical Implications
Software engineers with broad knowledge and software engineering principles are essential for successful software development projects.
Software architecture, encompassing base structure and subsystems, significantly impacts the success or failure of a software system.
Design decisions in software architecture should meet both functional and quality requirements, considering the application domain and other aspects.
Selecting the appropriate database system(s) (DBMS) is a crucial design decision that affects data storage and retrieval in the ENVRI-KMS.
Evaluating and selecting the right programming language ecosystem is crucial for developing a knowledge base system like the ENVRI-KMS.
Cross-referencing statements between knowledge bases (named graphs) requires careful consideration, with separate graphs for static and dynamic data.
Ontowiki can serve as an RDF-based middleware, supporting high-level user applications and services in the ENVRI-KMS.
Versioning and backup strategies are crucial for managing RDF data in the ENVRI-KMS, including exporting RDF dumps and using external means.
Continuous development and growth of the ENVRI-KMS depend on the efforts and contributions of the ENVRI subdomains and research infrastructures.
Interaction and collaboration with other subdomain developers and semantic search workgroups provide valuable input for the ENVRI-KMS.
Future development efforts will focus on continuous content ingestion and curation, improvement based on community feedback, DevOps practices, and community involvement in content maintenance.

7 Related work

In this research, Snowballing was the primary method to investigate the existing literature regarding tools and techniques that address the knowledge management challenges. A subset of selected studies is presented in Table 4.

Since 1990, business publications have started to publish an extensive list of research articles on knowledge management and decision support systems[47]. Wielinga *et al.*[48] explained knowledge-based systems' development as a modeling activity. Sapuan[49] reported a set of knowledge management systems'

Table 4. The results of the systematic literature review based on Snowballing (citation tracking) are presented here. The table shows the comparison of the selected studies and this study against a set of key factors, including research methods, publication types, research types, emphasized lifecycle phases, and contexts.

Study	Year	Research Method	Research Type	Lifecycle Phase	Context
This study	2021	Literature Study Document Analysis Survey Design Science	Research Paper	Planning Requirement Analysis Architecture Design Implementation	Knowledge Engineering Knowledge Management Knowledge Discovery Knowledge Acquisition Knowledge Representation
48	1992	Literature Study	Research Paper	Architecture Design	Knowledge Engineering Knowledge Acquisition
49	2001	Literature Study	Research Paper	Architecture Design	Decision-Making Process
50	2019	Literature Study	Research Paper	Planning	Knowledge Management
51	2018	Survey	Research Paper	Planning Requirement Analysis	Knowledge Management
52	2002	Literature Study	Research Paper	Planning	Knowledge Management
53	2005	Literature Study	Research Paper	Maintenance	Knowledge Management
54	2020	Literature Study Experiment	Research Paper	Architecture Design Implementation	Knowledge Discovery Knowledge Representation
55	2017	Literature Study	Research Paper	Planning	Knowledge Management
56	2019	Case Study	Research Paper	Planning	Knowledge Management
57	2019	Case Study	Research Paper	Planning	Knowledge Engineering Knowledge Management Knowledge Discovery
58	2018	Literature Study	Research Paper	Planning	Knowledge Management Decision-Making Process
59	2019	N/A	Tool Paper	Implementation	Knowledge Discovery
60	2017	Literature Study	Tool Paper	Architecture Design Implementation	Knowledge Discovery Knowledge Acquisition
17	2020	Literature Study Document Analysis Design Science	Tool Paper	Architecture Design Implementation	Decision-Making Process Knowledge Management
61	2018	Literature Study	Tool Paper	Architecture Design Implementation	Knowledge Discovery Knowledge Acquisition
62	2002	N/A	Tool Paper	Architecture Design Implementation	Knowledge Management
63	2006	Literature Study	Tool Paper	Architecture Design Implementation	Knowledge Management

architectures, concepts, and development processes. Additionally, the author highlighted the importance of knowledge-based systems in the context of concurrent engineering. Lee and Hong[52] define knowledge management concepts and distinguish them from business process reengineering and learning organization in terms of information technology application. Chau and Chuntian[62] proposed a knowledge management system on flow and water quality to simulate human expertise and heuristics in problem-solving and decision-making in the coastal hydraulic and transport processes. Akhavan *et al.*[53] explained and analyzed the main failure factors of implementing a knowledge management system in a pharmacist company. Park and Kim[63] proposed a framework for designing and implementing a knowledge management system for the fourth generation of Research and Development (R&D). Wachsmuth *et al.*[60] introduced a search engine framework for acquiring, mining, assessing, indexing, querying, retrieving, ranking, and presenting arguments while relying on standard infrastructure and interfaces. GIGGLE[64] is a genomics search engine that identifies and ranks the significance of genomic loci shared between query features and thousands of genome interval files. Santoro *et al.*[51] investigated the relationship among knowledge management systems, open innovation, knowledge management capacity, and innovation capacity. Farshidi *et al.* introduced a framework and knowledge management system to build decision models for database management systems[65], cloud service providers[66], software architecture patterns[17,67], model-driven development platforms[68], programming languages[46] blockchain platforms[69], and decentralized autonomous organization platforms[70]. Based on the literature study phase of the research, we realized that publications could be categorized into *knowledge management approaches* and *knowledge management systems*. There is a vast amount of literature about knowledge management approaches, such as[48–58], that are research papers that mainly introduce methodologies that can be employed to discover, gather, manage, and apply a particular type of tacit or implicit knowledge. Alternatively, a significant number of tool papers can be found in literature[17,60–63,66], that mainly present knowledge management systems that can be employed in a particular domain, such as software engineering or environmental science.

Additionally, we identified the phases of the software development life cycle (such as Planning, Requirement Analysis, and Architecture Design, Implementation)[71,72] that selected studies have reported. During the literature study, we observed that the selected publications mainly researched within the following contexts: (1) **Knowledge Engineering** which emphasizes how to represent human knowledge (tacit knowledge) in a system and to extract interpretable information that can be turned into knowledge (explicit knowledge). (2) **Knowledge Management** that explains the process of creating, sharing, using, and managing the knowledge and information of an organization. (3) **Knowledge Discovery** that refers to the process of finding explicit knowledge in data and emphasizes the "high-level" application of particular data mining methods. The main goal is to extract such knowledge from data in the context of large databases. (4) **Knowledge Acquisition** which is the process used to define the rules and ontologies required for a knowledge-based system and is the process of extracting, structuring, and organizing knowledge from one source, usually human experts. (5) **Knowledge Representation** that translates information from the real world into a machine-understandable form and then utilizes acquired knowledge to solve complex decision-making problems. (6) **Decision-Making Process** which is a reasoning process based on assumptions of values, preferences, and beliefs of decision-makers. It leads to suggesting a set of solutions among several possible alternative options.

Table 3 shows the key factors of the selected studies and compares them against our study. The table shows that employed a combined research method that is based on literature study, document analysis, survey, and design science. Moreover, we reported the planning, requirement analysis, architecture design, and implementation phases of the ENVRI-KMS. The paper's primary contexts are Knowledge Engineering, Knowledge Management, Knowledge Discovery, Knowledge Acquisition, and Knowledge Representation.

8 Conclusion and future Work

The development and operation of the ENVRI-KMS will be continuous. It will grow during the project while the development results and knowledge accumulate. The ENVRI-KMS development and operation depend on the development effort from the ENVRI subdomains and research infrastructures. The ENVRI-KMS should play a role in supporting developers from RIs to share best practices and find existing solutions, but the ENVRI community provides valuable input to the ENVRI-KMS and keeps it alive.

Currently, the ENVRI-KMS team closely interacts with the other subdomain developers (via workshops, meetings, and workgroups organized by subdomains). Through members, there is valuable input of the catalog of services, authentication and authorization, persistent identifier, triple store, license and usage tracking, and ENVRI-HUB.

The ENVRI-KMS team also closely interacts with semantic search workgroups in subdomains, e.g., a semantic search use case in ACTRIS[13] reported in the Semantic Search Working Group Final Report []FinalReport.

The ENVRI-KMS will continue in the rest of the ENVRI-FAIR project. In the next phase, the development effort will mainly focus on the following aspects: (1) Continuous content ingestion and curation. The ENVRI-KMS team will improve the knowledge ingestion tool and continuously ingest the description (metadata) of high-quality results from the ENVRI community (e.g., sub-domain or RI developers), including development results (e.g., best practices, software technologies, recommendations, updated FAIRness assessment possibly generated by new tools) in the ENVRI-KMS, and make those descriptions FAIR for the community.

(2) Continuous improvement of the ENVRI-KMS based on the feedback is received from the community. Extra features, e.g., for ENVRI-KMS discovery and recommendation, will be further explored.

[13]https://github.com/xiaofengleo/actris

(3) The development and operation of the ENVRI-KMS will also follow the software engineering DevOps practices. The continuous testing, integration, and deployment pipeline will be established.

(4) We will also extend the content maintenance to community specialists. In this way, we hope the community will play a key role in the ENVRI-KMS.

Data availability

Underlying data

Mendeley Data: ENVRI-KMS. https://doi.org/10.17632/ntxypfsvds.1[21]

This project contains the following underlying data:

- Knowledgebase-discussion.xlsx (Raw survey outcome data)

Extended data

- Analysis of the Survey: https://doi.org/10.17632/ntxypfsvds.1[21]
- Technology review, system design, documentation of the implementation, a demo of the ENVRI-KMS: https://doi.org/10.17632/n3khm4pnsd.1[31]

This project contains the following extended data:

- ENVRI-KMS (1).pdf (Summary of data analysis)
- Knowledgebase-discussion.pdf (Visualized survey outcomes)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability

Software available from: SciCrunch: ENVRI-KMS,RRID:SCR_021235

Source code available from: https://github.com/SiamakFarshidi/solr-php-ui.git

Archived source code at time of publication: http://doi.org/10.5281/zenodo.4882766[19]

License: https://opensource.org/licenses/Apache-2.0

References

1. Urry J: **Climate change and society.** In: *Why the social sciences matter.* Springer, 2015; 45–59.
Publisher Full Text
2. Michie J, Cooper C: **Why the social sciences matter.** Springer, 2015.
Reference Source
3. Tanhua T, Pouliquen S, Hausman J, *et al.*: **Ocean fair data services.** *Front Mar Sci.* 2019; **6**: 440.
Publisher Full Text
4. Vermeulen A, Glaves H, Pouliquen S, *et al.*: **Supporting cross-domain system-level environmental and earth science**. In: *Towards Interoperable Research Infrastructures for Environmental and Earth Sciences.* Springer, 2020; 3–16.
Publisher Full Text
5. Zhao Z, Liao X, Martin P, *et al.*: **Knowledge-as-a-service: A community knowledge base for research infrastructures in environmental and earth sciences**. In: *2019 IEEE World Congress on Services (SERVICES).* IEEE, 2019; **2642**: 127–132.
Publisher Full Text
6. ENVRIplus: **Research infrastructures.** 2021.
Reference Source
7. ICOS: **Integrated carbon observation system.** 2021.
Reference Source
8. IAGOS: **In-service aircraft for a global observing system.** 2021.
Reference Source
9. ACTRIS: **European research infrastructure for the observation of aerosol, clouds and trace gases.** 2021.
Reference Source
10. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **Addendum: The fair guiding principles for scientific data management and stewardship.** *Sci Data.* 2019; **6**(1): 6.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Hevner AR, March ST, Park J, *et al.*: **Design science in information systems research.** *MIS Q.* 2004; **28**(1): 75–105.
Publisher Full Text
12. Simon HA: **The Sciences of the Artificial (3rd Ed.)**. MIT Press, Cambridge, MA, USA, 1996.
Reference Source
13. Baumeister J, Striffler A: **Knowledge-driven systems for episodic decision support.** *Knowl Based Syst.* 2015; **88**: 45–56.
Publisher Full Text
14. Power DJ, Sharda R: **Model-driven decision support systems: Concepts and research directions.** *Decis Support Syst.* 2007; **43**(3): 1044–1061.
Publisher Full Text
15. Velampalli S, Jonnalagedda MV: **Graph based knowledge discovery using mapreduce and subdue algorithm.** *Data Knowl Eng.* 2017; **111**: 103–113.
Publisher Full Text
16. Becker C, Kraxner M, Plangg M, *et al.*: **Improving decision support for software component selection through systematic cross-referencing and analysis of multiple decision criteria**. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on.* IEEE, 2013; 1193–1202.
Publisher Full Text
17. Farshidi S, Jansen S: **A decision support system for pattern-driven software architecture**. In: *Proceedings of the 14th European Conference on Software Architecture, ECSA 2020.* ACM, 2020; **1**: 1–12.
Reference Source
18. Castellano G, Vessio G: **Towards a tool for visual link retrieval and knowledge discovery in painting datasets**. In: *Italian research conference on digital libraries.* Springer, 2020; 105–110.
Publisher Full Text
19. Farshidi S: **SiamakFarshidi/solr-php-ui: ENVRI-KMS (Version 1.0).** *Zenodo.* 2021.
http://www.doi.org/10.5281/zenodo.4882766
20. gra.fo.: **Envri-fair research infrastructures**. 2021.
Reference Source
21. Farshidi S, Zhao Z: **Envri-kms**. 2021.
Reference Source
22. Mentimeter: **An application for creating interactive presentations & meetings**. 2021.
Reference Source
23. Martin P, Remy L, Theodoridou M, *et al.*: **Mapping heterogeneous research infrastructure metadata into a unified catalogue for use in a generic virtual research environment.** *Future Gener Comput Syst.* 2019; **101**: 1–13.
Publisher Full Text
24. Calyam P, Wilkins-Diehr N, Miller M, *et al.*: **Measuring success for a future vision: Defining impact in science gateways/virtual research environments.** *Concurr Comput.* 2021; **33**(19): e6099.
Publisher Full Text
25. Zhao Z, Koulouzis S, Bianchi R, *et al.*: **Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment.** *Software: Practice and Experience.* 2022; **52**(9): 1947–1966.
Publisher Full Text

26. Liao X, Bottelier J, Zhao Z: **A column styled composable schema matcher for semantic data-types.** *Data Sci J.* 2019; **18**(1): 25.
Publisher Full Text

27. ISO: **Iec 10746-1 information technology–open distributed processing–reference model: Overview.** 1998.
Reference Source

28. ISO: **Iec 10746-2 information technology–open distributed processing–reference model: Foundations.** 2009.
Reference Source

29. ISO: **Iec 10746-3 information technology–open distributed processing–reference model: Architecture.** 2009.
Reference Source

30. ISO: **Iec 10746-4 information technology–open distributed processing–reference model: Architecture semantics.** 1998.
Reference Source

31. Mentimeter: **Technology review, system design, documentation of the implementation, a demo of the kms-envi**. 2021.
http://www.doi.org/10.17632/n3khm4pnsd.1

32. Openlink Virtuoso: **Open-source edition**. 2021.
Reference Source

33. RDForms: **Rdf in html-forms**. 2021.
Reference Source

34. ontowiki: **A knowledge management platform**. 2021.
Reference Source

35. Open Semantic Search: **Integrated research tools for searching and text mining**. 2021.
Reference Source

36. WebVOWL: **Visual notation for owl ontologies**. 2021.
Reference Source

37. Bhattacharya P, Neamtiu I: **Assessing programming language impact on development and maintenance: A study on c and c++.** *Proceedings of the 33rd Int Conference on Software Engineering.* ACM, 2011; 171–180.
Publisher Full Text

38. Olariu C, Gogan M, Rennung F: **Switching the center of software development from it to business experts using intelligent business process management suites.** *Soft Computing Applications.* Springer, 2016; 993–1001.
Publisher Full Text

39. Clements P, Kazman R, Klein M, *et al.*: **Evaluating software architectures.** Tsinghua University Press Beijing, 2003.

40. Lago P, Avgeriou P: **First workshop on sharing and reusing architectural knowledge.** *ACM SIGSOFT Software Engineering Notes.* 2006; **31**(5): 32–36.
Publisher Full Text

41. Bosch J: **Software architecture: The next step.** In *European Workshop on Software Architecture.* Springer, 2004; 194–199.
Publisher Full Text

42. Avgeriou P, Kruchten P, Lago P, *et al.*: **Sharing and reusing architectural knowledge--architecture, rationale, and design intent.** In *29th International Conference on Software Engineering (ICSE' 07 Companion).* IEEE, 2007; 109–110.
Publisher Full Text

43. Farshidi S: **Multi-Criteria Decision-Making in Software Production.** PhD thesis, Utrecht University, 2020; **2020-35**: 1–306.
Publisher Full Text

44. Farshidi S, Jansen S, de Jong R, *et al.*: **Multiple criteria decision support in requirements negotiation.** *23rd International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018).* 2018; **2075**: 100–107.
Reference Source

45. Peyton Jones S, Leshchinskiy R, Keller G, *et al.*: **Harnessing the multicores: Nested data parallelism in haskell.** In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science.* Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008.
Publisher Full Text

46. Farshidi S, Jansen S, Deldar M: **A decision model for programming language ecosystem selection: Seven industry case studies.** *Inform Software Tech.* 2021; **139**: 106640.
Publisher Full Text

47. Sprague RH Jr, Watson HJ: **Bit by bit: toward decision support systems.** *Calif Manage Rev.* 1979; **22**(1): 60–68.
Publisher Full Text

48. Wielinga BJ, Schreiber AT, Breuker JA: **Kads: A modelling approach to knowledge engineering.** *Knowl Acquis.* 1992; **4**(1): 5–53.
Publisher Full Text

49. Sapuan SM: **A knowledge-based system for materials selection in mechanical engineering design.** *Mater Des.* 2001; **22**(8): 687–695.
Publisher Full Text

50. Martins VWB, Rampasso IS, Anholon R, *et al.*: **Knowledge management in the context of sustainability: Literature review and opportunities for future research.** *J Clean Prod.* 2019; **229**: 489–500.
Publisher Full Text

51. Santoro G, Vrontis D, Thrassou A, *et al.*: **The internet of things: Building a knowledge management system for open innovation and knowledge management capacity.** *Technol Forecast Soc Change.* 2018; **136**: 347–354.
Publisher Full Text

52. Lee SM, Hong S: **An enterprise-wide knowledge management system infrastructure.** *Ind Manag Data Syst.* 2002; **102**(1): 17–25.
Publisher Full Text

53. Akhavan P, Jafari M, Fathian M: **Exploring the failure factors of implementing knowledge management system in the organizations.** *J Knowl Manag Pract.* 2005; **6**.
Reference Source

54. Castellano G, Lella E, Vessio G: **Visual link retrieval and knowledge discovery in painting datasets.** *Multimed Tools Appl.* 2021; **80**: 6599–6616.
Publisher Full Text

55. Iskandar K, Jambak KI, Kosala R, *et al.*: **Current issue on knowledge management system for future research: a systematic literature review.** *Procedia Comput Sci.* 2017; **116**: 68–80.
Publisher Full Text

56. Albassam BA: **Building an effective knowledge management system in saudi arabia using the principles of good governance.** *Resour Policy.* 2019; **64**: 101531.
Publisher Full Text

57. Orenga-Roglá S, Chalmeta R: **Methodology for the implementation of knowledge management systems 2.0.** *Bus Inf Syst Eng.* 2019; **61**: 195–213.
Publisher Full Text

58. Hellebrandt T, Heine I, Schmitt RH: **Knowledge management framework for complaint knowledge transfer to product development.** *Procedia Manuf.* 2018; **21**: 173–180.
Publisher Full Text

59. Kopanos C, Tsiolkas V, Kouris A, *et al.*: **Varsome: the human genomic variant search engine.** *Bioinformatics.* 2019; **35**(11): 1978–1980.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

60. Wachsmuth H, Potthast M, Al-Khatib K, *et al.*: **Building an argument search engine for the web**. In *Proceedings of the 4th Workshop on Argument Mining*. Copenhagen, Denmark, Association for Computational Linguistics. 2017; 49–59.
Publisher Full Text

61. Chantamunee S, Fung CC, Wong KW, *et al.*: **Knowledge discovery from thai research articles by solr-based faceted search.** In *International Conference on Computing and Information Technology.* 2018; 337–346.
Publisher Full Text

62. Chau KW, Chuntian C, Li CW: **Knowledge management system on flow and water quality modeling.** *Expert Systems with Applications.* 2002; **22**(4): 321–330.
Publisher Full Text

63. Park Y, Kim S: **Knowledge management system for fourth generation r&d: Knowvation.** *Technovation.* 2006; **26**(5–6): 595–602.
Publisher Full Text

64. Layer RM, Pedersen BS, DiSera T, *et al.*: **Giggle: a search engine for large-scale integrated genome analysis.** *Nat Methods.* 2018; **15**(2): 123–126.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

65. Farshidi S, Jansen S, de Jong R, *et al.*: **A decision support system for software technology selection.** *J Decis Syst.* 2018; 98–110.
Publisher Full Text

66. Farshidi S, Jansen S, de Jong R, *et al.*: **A decision support system for cloud service provider selection problems in software producing organizations.** In *IEEE 20th Conference on Business Informatics (CBI).* 2018; **1**: 139–148.
Publisher Full Text

67. Farshidi S, Jansen S, Martijn J, *et al.*: **Capturing software architecture knowledge for pattern-driven design.** *J Syst Softw.* 2020; **169**: 110714.
Publisher Full Text

68. Farshidi S, Jansen S, Fortuin S: **Model-driven development platform selection: four industry case studies.** *Softw Syst Model.* 2021; **20**; 1525–1551.
Publisher Full Text

69. Farshidi S, Jansen S, España S, *et al.*: **Decision support for blockchain platform selection: Three industry case studies.** *IEEE Transactions on Engineering Management.* 2020; **67**(4): 1109–1128.
Publisher Full Text

70. Baninemeh E, Farshidi S, Jansen S: **A decision model for decentralized autonomous organization platform selection: Three industry case studies.** *Blockchain: Research and Applications.* 2023; 100127.
Publisher Full Text

71. Pressman RS: **Software engineering: a practitioner's approach.** Palgrave macmillan, 2015.
Reference Source

72. Ruparelia NB: **Software development lifecycle models.** *ACM SIGSOFT Software Engineering Notes.* 2010; **35**(3): 8–13.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✘ ✔

Version 3

Reviewer Report 12 June 2023

https://doi.org/10.21956/openreseurope.17438.r32534

© **2023 Marzi G.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Giacomo Marzi**
Department of Management, Mathematics and Statistics (DEAMS), University of Trieste, Trieste, Italy

Dear authors,

Thank you for addressing all of my comments. From my perspective, the paper is now ready for indexing.

I would also like to express my appreciation for your submission to Open Research Europe.

Best wishes,
GM

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: innovation management, knowledge management

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 09 September 2022

https://doi.org/10.21956/openreseurope.15455.r30062

© **2022 Marzi G.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Giacomo Marzi
[1] Department of Management, Mathematics and Statistics (DEAMS), University of Trieste, Trieste, Italy
[2] Department of Management, Mathematics and Statistics (DEAMS), University of Trieste, Trieste, Italy

Dear authors,
Thanks for your submission to Open Research Europe. Since I have stepped in during the second round of revision of the present paper, I will ground my comments even on the previous reviews and the previous reviewers' comments.

As a first impression, I have noticed a strong improvement on the paper, especially about setting the boundary of your study on ENVironmental Research Infrastructures (ENVRI), while I see additional room for improvement in some aspects of the paper.

In particular, there are two points that require an additional revision, also considering the previous reviewers' comments:

- It is still not clear the representativity of the sample and the details of the participants. You stated you conducted an online webinar with experts while not providing details about the questions, the topic of discussion, and the approach to the interviews.

  At the same time, you did not disclose any information about the participant in the subsequent online survey that is the core of your study.
  I would suggest a table (or some text) detailing the sample and how the sample has been selected.

- The second focal point to be addressed is about the practical implications. In Section 6, you aim to summarize the lesson learned during the ENVRI-KMS project. However, a practitioner could find it difficult to apply such findings or could be in difficult to find them inside the paper. As a result, I suggest adding a table summarizing in few bullet points the main implications of your study.

Thanks for your time and good luck with the review.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: innovation management, knowledge management

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 May 2023

Siamak Farshidi

Comment 1: We conducted a webinar with experts from 26 research infrastructures involved in the ENVRI-FAIR project to gather their functional requirements and quality concerns for the ENVRI-KMS development. However, we understand that the representativity of the sample and participant details need further clarification.

Thirty-five domain experts participated in the webinar, and their selection was based on their expertise and years of experience. On average, the participants had over ten years of experience within their respective domains. They possessed a deep understanding of the challenges faced by researchers in their fields during their daily tasks.

To address the lack of information regarding the online survey, we apologize for the oversight. The online survey was conducted after the webinar using the Mentimeter platform. The virtual questionnaire included the following questions:

What information will you typically search for from the ENVRI community?
What will be the typical queries you would ask the ENVRI-KMS?
How do you currently search for information from the ENVRI community?
Which functionality of the knowledge management system do you find most beneficial?
What functions do you expect from the next version of the ENVRI-KMS?

All the responses to the survey questions were collected and analyzed, with priority given to recurring statements. This helped us identify common themes and prioritize the requirements based on the frequency of similar responses.

To address the comment's suggestion, we will provide additional text that details the sample selection process and how the participants were chosen for the webinar and subsequent survey. This will ensure transparency and better demonstrate the representativity of the sample in our study.

Comment 2: Thank you for your valuable feedback regarding the practical implications discussed in Section 6 of our ENVRI-KMS project. We appreciate your insight and the suggestion to provide a more concise and easily accessible summary of the lessons learned. After carefully considering your comment, we agree that presenting the practical implications in a more organized and condensed format would greatly benefit practitioners

and readers seeking to apply the findings of our study. To address this, we have incorporated your suggestion and created a table summarizing the main implications of the ENVRI-KMS project. This table presents the key takeaways in a clear and concise manner, allowing for quick reference and application.

We believe that this addition will enhance the usability of our paper for practitioners, providing them with an at-a-glance overview of the practical implications derived from our research. The table highlights the importance of software engineering principles, software architecture decisions, database system selection, programming language ecosystems, cross-referencing of knowledge bases, and the role of Ontowiki as an RDF-based middleware. Additionally, it emphasizes the significance of versioning and backup strategies, continuous development efforts, and collaboration within the ENVRI community.

We genuinely appreciate your input, as it has helped us improve the accessibility and practicality of our findings. We hope that the inclusion of this table will facilitate the application of our research in real-world scenarios and foster further discussion and collaboration among practitioners in the field.

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 December 2021

https://doi.org/10.21956/openreseurope.15455.r28067

© 2021 Huber R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robert Huber
[1] Agricultural Economics and Policy AECP, Swiss Federal Institutes of Technology Zurich ETHZ, Zürich, Switzerland
[2] Agricultural Economics and Policy AECP, Swiss Federal Institutes of Technology Zurich ETHZ, Zürich, Switzerland

The authors have significantly improved the manuscript by restructuring and condensing the previous draft.

Is the rationale for developing the new software tool clearly explained?
Partly

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: geosciences, environmental sciences, data management

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 12 July 2021

https://doi.org/10.21956/openreseurope.14751.r27110

© 2021 Koskela R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rebecca Koskela
[1] University of New Mexico, Albuquerque, NM, USA
[2] Ronin Institute, Montclair, NJ, USA
[3] University of New Mexico, Albuquerque, NM, USA
[4] Ronin Institute, Montclair, NJ, USA

The paper summarizes the process for creating the Knowledge Management System for the Environmental Research Infrastructures (ENVRI) community.

Introduction:
- Assertions in this section require citations and they are missing. For example, "Due to population growth and economic development, human impacts on natural resources are continuing to grow." It also does not introduce the reader to ENVRI.

Section 2:
- This section is supposed to include the formulation of research questions. It doesn't accomplish this - the questions are generic and the connection to ENVRI is not obvious. Table 1 only adds confusion. Why would a KMS for mechanical engineering or product complaints be relevant? More discussion is needed for the choices in Table 1.

Section 3:
- Describes the survey that was used to gather requirements. It would be helpful to understand what breakout of the 35 participants, especially with respect to Figure 1. Was there sufficient participation by all the identified stakeholders? If there are 26 RIs, is 35

sufficient to discover the requirements?

- In addition, the acronym "KB" is used in Section 4 but not defined. It is most likely a typo. Another issue is the organization of the paper. The survey is discussed in Section 4 and then again in Section 6.

It would be interesting to read more about what is different about a knowledge management system for environmental science.

Is the rationale for developing the new software tool clearly explained?
Partly

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
No

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Competing Interests: I am co-chair of the Research Data Alliance Metadata Interest Group with one of the authors, Keith Jeffery. The work of that group is not related to this paper. This conflict of interest did not affect the objectivity of my review.

Reviewer Expertise: High-performance computing; data management, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 24 June 2021

https://doi.org/10.21956/openreseurope.14751.r27111

© 2021 Huber R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robert Huber
[1] Agricultural Economics and Policy AECP, Swiss Federal Institutes of Technology Zurich ETHZ, Zürich, Switzerland
[2] Agricultural Economics and Policy AECP, Swiss Federal Institutes of Technology Zurich ETHZ, Zürich, Switzerland

The authors present a new knowledge management system to be used for environmental research infrastructures. Overall, this is certainly an innovative approach to make subject-specific knowledge and technical expertise easily findable and available also outside the own community. Unfortunately, the paper in its current form still has significant shortcomings, so that I would suggest indexing only after a fundamental revision and renewed review.

In detail, the paper leaves too much room for technology reviews and generalities. On the other hand, many questions remain open concerning the implementation and use of the tool. Many descriptions of the individual components are unfortunately still fragmentary and incoherent. In addition, a profound discussion is missing, which explains, for example, what distinguishes the proposed tool in comparison with commercial solutions and how the individual components and their interaction (given the identified user requirements and use cases) represent added value for the community here.

Abstract

- typo: ENVironmental

Introduction

Para 1:

- Would recommend not put population growth at the first place. Developped countries are main drivers.

Para 2:

- IODE comment on data echange does not really fit here. (Instead a sentence on the necessity of observation systems would be required.

Para 5:

- Avoid acronym CC4BY
- Expand IAGOS and ACTRIS acronyms
- Focus on atmosphere should be explained

Para 6:

- Please explain type of missing connects , assume you mean more fields than just data ?
- Please explain shortly what kind of information shall be captured?

Section 2:

Para 1:

- Please differentiate between research environment and RI
- First sentence seems to be broken 'such as ICOS are IAGOS'

Para 2:

- These are not research questions but design questions,

Para 3:

- Para could imho be moved down, maybe also Para 1. Introduce the ENVRI-KMS then explain the methodology.

Para 5:

- I would say KMS is not covering every aspect of FAIR.

- Sure you mean development not operating and consuming communities in 'Knowledge-as-a-Service for the RI development communities.'

- Are these listed requirements a result of the methods described above? Maybe you should differentiate between community specific and generic requirements

Section 3

- Table 1: explain terms used in the table to classify

- Section lacks an overview on existing commercial knowledge management systems.

Section 3.1 and 3.2

- Should be condensed and shortened and maybe moved to Introduction.

- The purpose of this sections is unclear to me. You describe literature covering 3 decades of research. Maybe you should extract a short, historically organised overview on past achievements instead of list every paper individually (which is already done in the Table).

- What are the consequences, conclusions of the lists described?

Section 4.1

- Please condense and shorten this section.

Para 1:

- Online event?

- Here you mention 26 RI but you only introduced three above.

- Could you really select experts out of a larger pool?

- 'Firstly, we introduced the potential functionality' -> How did you get there? Is this the result of the literature analysis?

Para 2:

- At least the first few requirements sound as if they have been translated into some jargon. Are these the original reqs defined by RI experts?

- A bullet point list would be easier to read the whole section probably could be replaced by a table or appendix.

- Sentence broken: '(R01) have all potential RIs,'

- Define 'search space entity.'

- R03: please rewrite.

Para 3:

- So all asked RI experts were from the data domain? Please explain the background of experts above.

- User stories are not explained at all above this is however of key importance to understand the scope of the paper.

Section 4.2
Move section up and explain use cases in more detail e.g. explain the focus on data (see comment above)

(2)
- Hard to understand, please rewrite.
- Update the state of their RI? Rather : get updated on ...
- Assessment wizard tool is not explained.
- Explain acronym KB.

(3)
- Check which technologies? Implemented in RI or existing outside?
- Please rewrite this section what e.g. means 'prepared for FAIRness gaps'?

Figure 2: figure caption missing,

Section 4.4
- Please strongly shorten and focus on the technologies chosen. You need not justify this by such a detailed technology overview.
- Consider to skip this section, the selected technologies are mentioned below anyway.

Section 5

Instead of 'prioritized user stories' you should mention the use case you considered (data management etc.)

Section 5.1

'It was suggested to consider OntoWiki' -> Did you choose this tool or another one?

Section 5.3
- This section is hard to understand and needs to be rewritten. It is not clear if this functionality is already in place or still planned.
- Links given as footnotes lead to binder linked notebooks which are not very useful to demonstrate how FAIR assessments are performed.

Section 5.4
- 'via forms based on the FAIRness analysis' -> This is not explained before.
- 'Some issues were discovered' which ones, is this important to describe the tool in this paper? Sections dealing with bug fixing are not necessary.
- Would suggest to move the section dealing with user engagement to the discussion.

Section 5.5

- This very short section suddenly jumps into document management issues.
- Please explain a bit how this was done, I assume here Elasticsearch was used?
- It is hard to understand the connection with Ontowiki is there any?

Section 5.6

- Web Crawler and Sitemap extractor are mentioned but not explained. In comparison to the lengthy description of OntoWiki this is not sufficient.
- Also sections on NLP etc. this would be more interesting in comparison to the tool selection parts above.
- Please expand and explain how this relates to Ontowiki and Search engine.

Section 6.1

Many of the arguments listed here are already mentioned in the previous sections. Please merge and shorten (a lot) and ideally move this above/prior the technology selection section.

Section 6.2

- Literature: is this what has been selected to feed the system? Content is not very clear to me.
- Overall this section is much too long and a bit hard to follow, maybe this can be merged with some other section. The FAIR principles are not helpful in this context to describe how research questions have been addressed.

Section 7

Please do not discuss 'Software architecture' and Implementation in general, instead discuss the architecture and implementation of the ENVRI-KMS. The whole section needs to be rewritten in way that it focuses on the ENVRI KMS.

General comment

Usually KM systems also care a lot about 'tacit knowledge' but I assume you rather deal with documents and data?

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

No

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: geosciences, environmental sciences, data management

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Comments on this article

Version 1

Reader Comment 05 Jul 2021

Daniel Spichtinger, self-employed, Vienna, Austria

This is a very topical article, given that sharing and discovery across different areas promises new insight, but is, in practice, difficult to do.

From my side, I would be interested in:

a) update and use of the described tool

b) implications for European policies, in particular in regards infrastructures (ESFRI & co) but also broader data management and sharing policies in Horizon Europe and the ERA.

Competing Interests: No competing interests were disclosed.