Supporting eosc

Skills 4 eosc

Milestone: Pilot learning path for Data Stewards

Lead Partner:	TU Delft, T4.1
Version:	**1.0**
Status:	**Version open for feedback (co-creation) within WP4**
Dissemination Level:	
Document Link:	https://doi.org/10.5281/zenodo.14506356

Deliverable Abstract

This short report invites comment on the pilot training on Ontologies which is part of the curriculum for Data Stewards designed by the Skills4EOSC project, WP4, T4.1. the EOSC.

Co-funded by the European Union

co-funded by UK Research and Innovation

Skills4EOSC has received funding from the European Union's Horizon Europe research and innovation Programme under Grant Agreement No. 101058527 and from UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10040140]

In person participants at KIT for the pilot training on Ontologies

29 October 2024

COPYRIGHT NOTICE

DELIVERY SLIP

Date	***Name***	***Partner/Activity***	***Date***
From:	Nida van Leersum Saba Sharma	TU Delft, 4.1	15.11.2024
Contributors to pilot	Fernando Aguilar Achim Winandi Fabian Schubö Nicky Daniels Lorna Wildgaard Dominique Green Kasper Drazewski Sonja Filiposka Vojdan Kjorveziroski	KIT/CSIC, 4.1 KIT, 4.1 KIT, 4.1 TU Delft, 4.1 DKB, 4.1 UDEIN, 4.1 KU Leuven, 4.1 UKIM, 2.3 UKIM, 2.3	
Reviewed by:			
Approved by:			

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
v.1	***4 November 2024***	*Initial draft based on internal outputs of task T4. and WP4,*	*Nida van Leersum and Saba Sharma (4.1)*
…			
…			
v.n			

TERMINOLOGY

https://eosc-portal.eu/glossary

Terminology/Acronym	*Definition*

Contents

1 Executive summary

Overview of the Pilot

On the 29th of October, a hybrid (in person at KIT, Germany, and online) pilot for one of the modules from the Data Steward Curriculum "Introduction to Ontologies" was conducted. Online participants took part using Big Blue Button and materials were published on the Skills4EOSC learning platform. The training was facilitated by Dr. Fernando Aguilar Gómez. Fernando works at the Spanish National Research Council (CSIC) in Santander (IFCA), but as one of the fellows from the Skills4EOSC fellowship programme, he was working at the library at KIT (Karlsruhe Institute of Technology). He was assisted with the preparation by Achim Winandi (KIT) and Fabian Schubö (KIT).

As this was a hybrid training, it consisted of participants attending the training in person at KIT and online. Online attendance was moderated by Saba Sharma (TU Delft) and Nicky Daniels. Saba Sharma is co-leading Task 4.1 as part of 4TU.ResearchData at TU Delft and Nicky Daniels currently works at Hasselt University but through the Skills4EOSC fellowship programme was part of 4TU.ResearchData at TU Delft as well.

The focus of the half day pilot was to learn what an ontology is and how it is used to model knowledge. The training established some links to its use within a data management context and explored the possibilities of making queries within the semantic web using linked data. The workshop included hands-on exercises using SPARQL. No prior knowledge of ontologies was required.

Participant Evaluation Results

A total of 39 learners participated in the pilot training. 19 completed the feedback form (49% response rate). The results of the feedback from the 19 particpants are summarised below.

Nine respondents were satisfied or very satisfied with the pilot, and five were neutral. Most participants had novice to intermediate knowledge of ontologies, metadata, XML, RDF, SPARQL and persistent identifiers/URIs. Post-training evaluations indicated an improvement in knowledge about ontologies, RDF, SPARQL and persistent identifiers/URIs.

Positive Aspects and Areas for Improvement

Participants praised the well-prepared material and hands-on aspect of the pilot (a good mixture of theory and exercise). The most valueable aspects of the training was the introduction on ontologies, the practical examples, the hand-on SPARQL exercises.

Some feedback for improvement was that there was too much information, not enough preparation time for particpants to install and test tools and that the pacing was too fast.

Additional training materials should be developed to address how to structure queries, how to make complex SPARQL queries, and real example ontologies.

Challenges of hybrid training

This training was offered as a hybrid session. The instructor was at KIT with the face to face participants and more than half of the total training particpants joined online. We had anticipated this challenge and had secured two online facilitators to ensure similar online engagement. However this still remained challenging. Some comments for improvement include:

- Sound quality in the room so that online particpants can better hear face to face participants
- The instructors pace and speed was sometimes fast and difficult to understand for the online particpants
- The interface on the shared screen was in Spanish which made it difficult to follow for the online particpants

Alignment with MVS for Data Stewards

The ontology module is well aligned with the MVS and provides practical approaches to making data FAIR, including linked and semantic data. The learners gained an understanding of the concepts and logical structure of an ontology. SPARQL is used to query information from databases/datasources to gain new knowledge or understanding of the ontology. The activities in the ontologies module are supported with deeper discussions of FAIR, data curation, good data stewardship, data sharing and publication, and metadata elsewhere in the curriculum.

As the MVS deals with beginner Data Stewards, this module is well aligned to the requirements for an embedded and coordinator Data Steward. Topics have been identified for further development of intermediate and advanced data stewards but it is beyond the scope of the current MVS.

Sources for Pilot Training

The pilot was hosted on the Skills4EOSC Learning Platform and was adapted from content availible in an editable library of training materials on GitHub developed by Task 4.1. Together, these sources provide a foundation for creating a comprehensive and effective pilot training on ontologies.

2. Detailed Training Feedback

2.1 Description of the training

The pilot on 'Introduction to Ontologies' was held on 29th October 2024. The goal was to test one module of the comprehensive Data Steward curriculum being developed by Task 4.1 as part of the Skills4EOSC project. This pilot was is part of the milestone deliverable for the task.

The pilot was held in hybrid mode, with online participants using the the Big Blue Button as the online meeting room, and participants were asked to access materials published on the Skills4EOSC learning platform during the pilot. The pilot was facilitated by an instructor and covered three main areas: ontologies, linked data and hands-on exercises. The focus of each area was as follows:

- Ontologies: Introduction, key concepts, benefits and "types", technologies and applications (examples), ontology modelling.
- Linked data: Introduction, principles, basic rules, challenges and trends
- Hands-on: SPARQL

The slides used in the pilot can be found on the Skills4EOSC learning platform and are published on GitHub (https://github.com/Skills4EOSC-DSCurriculum/Pilot-Training-Ontologies/tree/main).

The pilot was developed, organised and delivered and facilitated by an entire team as follows:

- Instructor: Fernando Aguilar (KIT/CSIC)
- Organisers: Achim Winandi (KIT), Fabian Schubö (KIT), Saba Sharma (TU Delft), Nida van Leersum (TU Delft)
- Online facilitator: Nicky Daniels (TU Delft)
- Fair by Design Methodology: Achim Winandi (KIT), Fernando Aguilar (KIT/CSIC)
- Observers: Dominique Green (UDEIN) and Kasper Drazewski (KU Leuven)
- Analysis of training with MVS Data Stewards: Lorna Wildgaard (DKB)

- Advice on badges, training platform and Fair by Design Methodology: Sonja Filiposka (UKIM), Vojdan Kjorveziroski (UKIM), Andrea Corleto (GARR)

2.2 Feedback Results

2.2.1 Participant Background

The pilot had a total of 39 learners. Out of that, 19 completed the evaluation survey (49% response rate). 12 respondents were satisfied or very satisfied with the pilot, and 5 were neutral.

The professional profiles of the learners is presented in Table 1. The "other" category was defined as project worker, RDM coordinator and project manager.

Table 1: Professional profiles of the learners who responded to the evaluation survey

Which of the following best describes your current role:		
Response	**Average**	**Total**
Researcher	43%	9
Data Scientist	19%	4
Data Manager	24%	5
Data Steward	5%	1
Other	19%	4

2.2.2 Knowledge of the topic

The majority of the learners described themselves as having novice to intermediate knowledge of ontologies, metadata, XML, RDF, SPARQL, persistent identifiers/URIs (Table 2).

Table 2: Participants knowledge of each area prior to participating in pilot

Responses	Novice	Intermediate	Expert	Total
Ontologies	**10** (53%)	9 (47%)	0	**19**
Metadata	3 (16%)	**13** (68%)	3 (16%)	**19**
XML	**10** (53%)	9 (47%)	0	**19**
RDF	**12** (63%)	6 (32%)	1 (5%)	**19**
SPARQL	**15** (79%)	3 (16%)	1 (5%)	**19**
Persistent Identifiers/URIs	6 (32%)	**11** (58%)	2 (11%)	**19**

After the training, the participants were asked to rate if their knowledge of the above topics had improved (Table 3).

The most improvement was on their understanding of ontologies with minor improvement in knowledge of metadata, RDF, SPARQL and persistent identifiers/URIs.

Table 3: Participants knowledge of each area after participating in pilot

Responses	Novice	Intermediate	Expert	Total
Ontologies	1 (5%)	**16** (84%)	2 (11%)	**19**
Metadata	2 (11%)	**13** (68%)	4 (21%)	**19**
XML	**10** (53%)	9 (47%)	0	**19**
RDF	**10** (53%)	8 (42%)	1 (5%)	**19**
SPARQL	**12** (63%)	7 (37%)	0	**19**
Persistent Identifiers/URIs	4 (21%)	**11** (58%)	4 (21%)	**19**

2.2.3 What went well and areas of improvement

The instructor was complimented by the learners for his good explanations, insights, an understandable introduction, well-prepared material and hand-on exercises.

We asked the pilot participants in the evaluations if there are any related topics not covered that should be added? In their responses, the following suggestions came up: How to structure queries, highly complex SPARQL queries, transfer from OWL to RDF, real example ontology.

Below, is a summary of what went well and where there is room for improvement from both the respondents point of view and from the observers who were present in the training to evaluate it.

What went well (participant feedback)

- Theoretical background
- Hands-on
- Overview table
- Introduction
- Examples; basic introduction
- The focus on RDM through ontologies
- Example (hamster) made it easy to understand the essence of ontologies
- The practical example
- SPARQL exercises
- Good examples
- Notebook
- Mention & show new tools
- Query examples in SPARQL
- SPARQL hands-on examples
- Refresh RDF
- First half on ontologies
- Heard buzz words

What went well (Observers feedback)

Overall the training was a really interesting topic which generated interest and interaction amongst the participants. Having different ways of explaining 'classes' was particularly good. Showing what a class is in the transportation discipline and in medicine is useful for the different types of people within the audience and is useful for the reusability of the training.

Step by step explanations by the instructor of each component of ontologies/knowledge was also very good. The mentimeter polls scattered throughout the lecture were also a nice addition because it broke up the lecture and allowed participants think about what's coming next.

The structure of the presentation was also nicely done. The introduction to ontologies in the first part of the session was easy to follow and well-laid out. Providing contact details to answer questions at the end of the lecture was an added bonus, as the training was a bit behind time.

Areas for improvement (Participant feedback)

- The exercises/practical was too quick
- Instructor based explanations on different examples instead of sticking to one example and explaining in a consistent way
- Better to install/try out the tool beforehand
- Too much information
- Interaction part could be better

Areas for improvement (Observers feedback)

- Given that it was a hybrid session, it was hard to hear the questions in the room. We should either have the instructor repeat the question for those online or consider a room mic.
- It would be useful to have the steps to accessing the activity pre-loaded in the learning platform. We didn't know that we had to save the file, for instance, to access it.
- It seems a couple of people online didn't have their questions answered fully. This could be because the question was outside of the scope of the workshop, but saying that clearly would be helpful.

- In the practical exercise, the learning curve felt a little steep at times and it was easy to get lost in the expressions of the query language – perhaps it could be easier if a handout were provided listing the key items.
- It was sometimes difficult to follow the spoken explanations due to the presenter's fast manner of speaking. The presenter's microphone was capturing a lot of noise from the room / through the open window. A condenser microphone typically resolves such issues.
- It would be best to follow up the menti responses/activities with some discussion about what's been uncovered. It was difficult to know if what people selected were right or wrong because there was little follow up to it.

2.2.4 Feedback on learning experience

In their evaluations, participants commented on their experience of learning. Table 4 provides an overview.

The majority of partipants who provided feedback found the content relevant, the information was well presented, the instructor was effective and the learning environment and the online facilitation was good. Areas that could have scored better were about pacing and having activities that were helpful.

Overall feedback includes: wanting follow-up/advanced course, difficult to follow as a non-programmer. There was also feedback on the feedback form and having more space under the questions to provide more detailed feedback.

Table 4: Participants experience of the training

Responses	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Total	N/A
The content was relevant	**9** (47%)	**9** (47%)	1 (5%)	0	0	**19**	0
The information was presented clearly	**6** (32%)	**6** (32%)	5 (26%)	2 (11%)	0	**19**	0
The pacing was right (not too slow, not too fast)	3 (16%)	**7** (37%)	3 (16%)	5 (26%)	1 (5%)	**19**	0
The activities were helpful	4 (21%)	**7** (37%)	6 (32%)	2 (11%)	0	**19**	0
The instructor was effective	6 (32%)	**8** (42%)	5 (26%)	0	0	**19**	0
The learning environment was good	**7** (37%)	**7** (37%)	4 (21%)	1 (5%)	0	**19**	0
The online facilitation was good	4 (33%)	**6** (50%)	2 (17%)	0	0	**12**	7

2.2.5 Feedback from the trainer

Below is a summary of the feedback from the trainer from his perspective of the pilot training.

Timing: The timing for the first learning unit was underestimated. With a bit of readjustment, the content can still be covered effectively within a four-hour schedule. Re-adjustments would be needed to allow for a more comfortable pace, ensuring participants have enough time to grasp the concepts fully.

Audience Engagement: Using Mentimeter was beneficial for keeping the audience engaged. However, while it facilitated connection, it could be further optimized to stimulate interaction more actively, perhaps by

incorporating more frequent, interactive questions or small quizzes to deepen engagement.

Hybrid Mode: The hybrid setup was manageable, and overall it worked well. Participants seemed to adapt to this format without major issues, allowing for a balanced experience for both in-person and remote attendees.

Interest and Participation: There appeared to be genuine interest in the workshop, as evidenced by the questions and comments from several participants. Many of their questions were practical, tied to their daily work, which added relevance to the session and enriched the overall experience. This level of engagement always makes the workshop more dynamic and attractive.

Hands-On Session: During the hands-on portion, it was noticeable that some participants became disengaged, likely due to the programming requirements. While those who were comfortable with coding found the exercises useful, programming acted as a barrier for others. I would recommend that, in future sessions, participants have some preparatory training on Jupyter Notebook beforehand. This could help alleviate anxiety around programming and make the hands-on activities more accessible to all.

3. Alignment with the MVS data steward

In this section we explore further the alignment between the activities described in the MVS for Data Steward (coordinator and embedded) and the training materials used for the pilot. Particular attention is paid to which activities are covered and which content, if any, is missing in the MVS.

3.1 Skills and competencies from the training materials

The ontology module, exemplified in the pilot training provides practical approaches to making data FAIR, including linked and semantic data. The learners gained an understanding of the concepts and logical structure of an ontology. SPARQL is used to query information from databases/datasources to gain new knowledge or understanding of the ontology. The activities in the ontologies module are supported with deeper discussions of FAIR, data curation, good data stewardship, data sharing and publication, and metadata elsewhere in the curriculum.

3.2 Comparison: MVS Data Steward and Ontologies Pilot

Table 5 below lists how the module on ontologies supported elements of the MVS.

Table 5: How the module supports the MVS Data Steward

This module contributes to the following open science outcome for a Data Steward: Application of ontologies in Research Data Management. This module contributes to the DS' skills in making research data more discoverable and interoperable, thus enabling sharing, finding and preservation. In summary. the materials in this unit support practical skills in how to implement the FAIR Principles: • Digital data are as FAIR and open as possible and as closed as necessary. • Research data and other digital objects are effectively managed to ensure their suitability for archiving and sharing, and advancement of research methods appropriate to the discipline(s). • Familiarity with how ontologies support the creation of structured metadata that enhances data searchability, interoperability and retrieval, meeting FAIR

metadata standards.

The module provides material that would help the Coordinator and Embedded Data steward in the following main activities:

Coordinator Data Steward:

- Develops institutional guidance on adoption and awareness of linked data and semantic web, e.g. **advice on resources and technical capabilities needed in organisations to support interoperability**, data quality (assurance) and understanding need for **competence development to align with institutional policies surrounding linked and FAIR data**.
- Understanding data and informing decision makers – data described with ontologies and linked represents domain knowledge in formal and explicit ways thus data can be **queried efficiently and the value of data in informing organisational decisions increases**.

Embedded Data Steward:

Develops Data Management Plans, tailoring the metadata description of the data for the project to relevant ontologies and implements the technical management of the data by applying the ontologies. The Embedded data steward can also query databases/data resources by applying their knowledge of the logical structure of ontologies to advice on the content of a data source and possible relevance for the research project. The Embedded data steward:

- **Advises on the use of disciplinary standards and ontologies,** and relevant community practices that are applied in producing FAIR research outputs.
- Implements **good practice on data description, like the use of controlled vocabularies and its relationships.**
- **Advises and supports researchers on data-infrastructure and tools,** and adoption of innovative techniques or tools, including those provided by relevant (inter)national data-infrastructure and tools.

Essential skills and competencies covered are:

- **Service provision to support cross-domain/domain specific Open Science practices** including: use of FAIR and CARE principles, Open Access, **data optimization**, data preservation, archiving and responsible re-use.

- Knowledge about Research Data Management, (personal) data governance and ethics, **Open Science data publication and exchange(sharing) services**, information security and risk management.

- **Advise/provide support on the use of infrastructure and tools at the appropriate levels of the organisation**, e.g. for data storage, data versioning and documentation, FAIR software and databases.
- **Training design and delivery to support Open Science practices**, policies and practices.
- **Knowledge/awareness of programming,** FAIR code and FAIR software and **use of standards and ontologies**

As indicated above, the proposed module offers the possibility of acquiring intermediate, technical skills in specific tools. The module takes the concept of ontologies from the MVS and exemplifies how to query data using ontologies with Python and SPARQL. Other workflows and tools can be used to apply ontologies and the queries can be more advanced than those used in the module.

Below, Table 6, is a list of skills from MVS's Essential skills and competences that are aligned with the skills acquired as a result of the learning path.

Table 6: The alignment between the MVS skills and competencies and the training materials

MVS's Essential skills and competencies	Covered in training material skills
Cross-domain/ domain-specific knowledge on Open Science practices, policies and regulation and translating these (when necessary) to the appropriate levels of the organisation.	No – but the participant becomes aware of disciplinary ontologies that can translate into practices and policies
Service provision to support cross-domain/domain specific Open Science practices including: use of FAIR and CARE principles, Open Access, data optimization, data preservation, archiving and responsible re-use	Technical skills to put FAIR principle "Interoperability" into practice which in turn contributes to data optimization and findability/reuse.

Knowledge about Research Data Management, (personal) data governance and ethics, Open Science data publication and exchange(sharing) services, information security and risk management.	Yes, the instruction points to the 5 Stars of Open Data, that links best practices with the technicalities of sharing data. To score the maximum five stars, data must (1) be available on the Web under an open licence, (2) be in the form of structured data, (3) be in a non-proprietary file format, (4) use URIs as its identifiers and, (5) include links to other data sources.
Raising awareness of the value of good data management among data creators and users, researchers, organisational colleagues, and decision-makers.	Knowledge of the value of linked/semantic data
Advise/provide support on the use of infrastructure and tools at the appropriate levels of the organisation, e.g. for data storage, data versioning and documentation, FAIR software and databases.	Can support and advise the use of ontologies to create the metadata, particularly link different relationships among versions, software or any other research component.
Training design and delivery to support Open Science practices, policies and practices.	With practice, the learner can begin to train others in how to apply an ontology to a dataset and for which purpose
Monitor the research and funding ecosystem, including possible conflicting motivations, drivers and incentives among different stakeholders.	No
Knowledge/awareness of programming, FAIR code and FAIR software and use of standards and ontologies	This is the essential skill supported in the module. The learner develops skills in applying ontologies and also arguing for why this is important.
Advocacy, analysis and assessment on FAIR data criteria, FAIR code and software preservation.	Awareness about advocacy for linked data (Findability and Interoperability)

Rather than adding more ontology-related competencies to the MVS, it is expected that any curriculum would be tailored to the level of the learner – novice, intermediate or advanced/expert. More advanced curriculum in ontologies may cover:

1. Ontology Development and Customisation
Understanding the process for creating or extending ontologies to accommodate the specific data needs of research projects or domains. Data stewards should support researchers in selecting or adapting ontologies that align with disciplinary standards and research needs, ensuring data interoperability across projects. More details about specific tools for achieving these tasks should be included.

2. Ontology Selection and Implementation
Practice in finding and assessing existing ontologies for suitability, evaluating alignment with FAIR principles, data standardisation, and interoperability requirements. Data stewards should advise researchers on ontology choice by considering both disciplinary standards and interdisciplinary compatibility, supporting FAIR-compliant data organisation and metadata enrichment.

3. Semantic Data Integration and Interoperability
Developing proficiency cards in semantic integration techniques thus building expertise in how to facilitate data integration across domains/projects, enabling broader data reuse, discoverability, and interoperability.

4. Training and Advocacy in Ontologies and Semantic Technologies
Competence in training researchers on the value of ontologies for enhancing data quality, interoperability, and long-term reuse. To be a trainer requires awareness of common challenges and solutions related to ontology use in a research community/organisation.

5. Engagement with Ontology Standards Bodies and Community Practices
Engagement in and awareness of ontology and semantic web communities, staying updated on best practices and standards. (WP6 Professional Networks for lifelong learning) thus the ability to advise on the adoption of community-recognised ontologies and standards.

4. Lessons learned and follow up actions

In this section, we summarise lessons learned from the pilot with the aim to optimise the curriculum for Data Stewards that is being produced by Task 4.1. The aim of the pilot was to assess if the curriculum gives a trainer enough information to prepare a training (for their local context).

We do not explore the areas for improvement for a training. This report has covered those areas in extensive detail.

4.1 Curriculum for Data Stewards

The curriculum is quite extensive providing guidelines to instuctors on what are the most importannt aspects of each module. It links to resources, makes suggestions on what sort of learning activities to incorporate into the training and the key messages to be covered by the trainer. The curriculum does not provide slides or content. This is because each trainer in their local context should be able to use the information and tailor their training according to their local context.

The trainer for this pilot confirmed that the curricum provides enough guidance from which to develop the training. He confirmed that even if there were slide decks provided, he would still develop his own material based on his own experience and content.

4.2 Quality Assurance

The Fair by Design Methodology gives ample advice on how to ensure that materials are FAIR and enable reuse. It also guides trainers how to quality assure their training before delivery.

The trainer and organising team found the materials very useful. Their only suggestion for improvement is to have all the resources in one place for easier use.

4.3 The Minimum Viable Skillset

The MVS provides a good reference point for curriculum design. The large turnout for this pilot training indicates the wide interest.

During the evaluation, there was feedback for further training which is confirmed by the review of the training with the MVS that basic knowledge on this topic was sufficiently covered and there can be many possiblities to develp the knowledge further.

4.4 Adaption of Learning Path

Further adaptions of the training materials are recommended to reduce the amount of content and use more time on discussions. Most of the content addressed licensing data, develop more in-depth material about software and code.

As learners come to the training with different levels of knowledge of the topics, use effort facilitating the discussion groups to ensure assignments are understood and conversations are flowing.

4.5 Revision of the MVS Data Steward

The ToT (Train of Trainers) pilot addressed only a small sub-set of skills described in the MVS Data Steward. The MVS DS deals with skills and competencies in a fairly general way. The training material on the other hand is more operational, showing use cases where the arguments for convincing people to use the licenses are clearly identified (slides 16 and 20). We recommend including additional text in the MVS regarding licensing as a part of the essential skills for data stewards, for example "Working knowledge of copyright and licensing as it relates Open Science and other research products". Hence, the MVS would gain in precision by proposing a section dedicated to the legal aspects of the RDM, in particular concerning the dissemination and sharing of FAIR data in compliance with legal recommendations (obligations).

If anything, the training exemplifies how much knowledge is needed for each single concept listed in the MVS as this learning path provides more intermediary material that goes beyond the MVS and is also relevant for MVS for Legal Experts.

We consider the training materials also to be in alignment with the MVS for Legal Experts as a supplement to their technical skills and competencies in

copyright licensing rules and existing frameworks. Likewise the materials, especially unit 1 concerning why researchers should license their research outputs and common myths why not to do so, are highly relevant for ELSI professionals to further their understanding of the psychosocial-disciplinary context and concerns of the researcher.