Developing a Reference OntoUML Conceptual Model for Data Management Plans: Enhancing Consistency and Interoperability

Jana Martínková[a], Marek Suchánek[b] and Robert Pergl[c]
Faculty of Information Technology, CTU in Prague, Prague, Czech Republic
{jana.martinkova, marek.suchanek, robert.pergl}@cvut.cz

Keywords: Data Management Plan, FAIR Principles, Ontology, Conceptual Model, OntoUML.

Abstract: The growing significance of Data Management Plans (DMPs) has highlighted the need for standardized and accurate data management practices. Current DMPs often suffer from inconsistent terminology, leading to misunderstandings and reducing their effectiveness. This study proposes the development of a DMP OntoUML conceptual model to address these issues. The model aims to clearly define all relevant concepts and their relationships, ensuring consistency and interoperability, particularly by connecting with the FAIR principles OntoUML model. The research follows a structured approach: specifying necessary concepts using existing templates and ontologies, defining terms and their relationships within the OntoUML model, and verifying the model's syntax. The resulting conceptual model will standardize terminology, promote interoperability, and support future DMP development and education.

1 INTRODUCTION

In recent times, the importance of developing a data management plan (DMP) has grown significantly. Effective data management practices ensure more accurate data collection, secure storage, proper handling, and utilization beyond the primary project scope. However, existing DMPs and their templates often employ varying terminology to describe the same concept or use identical terms for different concepts. This inconsistency can cause misunderstandings at human and machine levels, lowering the value of DMPs due to incomplete or incorrect information.

These misunderstandings can cause errors in DMPs as data stewards may misinterpret terms, leading to incorrect data management. This fragmentation hinders collaboration, reducing research quality and impact. Inconsistent terminology also complicates training for new researchers and data managers, making it harder to adopt best practices.

In recognition of this challenge, our proposal focuses on developing a DMP OntoUML conceptual model. This model will accurately describe all the concepts used within DMPs and establish clear relationships between them. Additionally, it will be connected to existing OntoUML model of Findable, Accessible, Interoperable, Reusable (FAIR) principles showing the connection with concepts of FAIR principles. Moreover, the conceptual model will aid in standardizing terminology, promoting interoperability between systems working with DMPs, and ensuring scalability for future DMPs development. Furthermore, it will serve as a valuable resource for training and education.

To accomplish this, we set the following partial steps:

- **G1.** Specify the concepts that needs to be covered using existing DMP templates, ontologies, and knowledge models related to DMPs;
- **G2.** Define the terms and their relationships in the OntoUML conceptual model using existing ontologies and vocabularies related to DMPs;
- **G3.** Verify the syntax of the OntoUML model and validate its content by using an example that ensures it comprehensively covers an existing DMP.

2 CONCEPTUAL MODELLING

Conceptual modelling is an activity aimed at developing a formal description of relevant aspects of reality, involving the domain, its concepts, and activities within it. The resulting output of this pro-

[a] https://orcid.org/0000-0001-8575-6533
[b] https://orcid.org/0000-0001-7525-9218
[c] https://orcid.org/0000-0003-2980-4400

Martínková, J., Suchánek, M. and Pergl, R.
Developing a Reference OntoUML Conceptual Model for Data Management Plans: Enhancing Consistency and Interoperability.
DOI: 10.5220/0012940000003838

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 2: KEOD*, pages 159-166
ISBN: 978-989-758-716-0; ISSN: 2184-3228

cess, conceptual model, is used in various contexts and domains, offering several advantages. As described by (Gonzalez-Perez, 2018), conceptual models serve as an excellent method for documenting knowledge with low ambiguity and high simplicity, making them easily understandable without additional explanations. As pointed out in (Robinson et al., 2015), they also bridge the gap between different mindsets and areas of knowledge.

The benefit also lies in the precise definition of the scope and purpose of tools and techniques used for work, and parts of the conceptual model can be reused in entirely different contexts, as described in (Robinson et al., 2015). Having a conceptual model also opens possibilities for comparing and connecting information from various sources with a higher certainty of understanding the authors' intentions. As described in (Gonzalez-Perez, 2018), conceptual modelling enables the exploration of complex fragments of the world that initially seem very tricky. Using a modelling language helps overcome obstacles by reducing complexity, allowing for a problem-solving approach that addresses one issue at a time.

During the development phase of conceptual models, a language is imperative to accurately, unambiguously, and clearly represent knowledge. Domain-specific languages such as DEMO (Dietz and Hoogervorst, 2015) or domain-agnostic ones like OntoUML can be employed (Pergl, 2019).

OntoUML (Guizzardi, 2005a) is an ontologically well-founded language for ontology-driven conceptual modelling based on Unified Foundational Ontology (UFO) and as a Unified Modeling Language (UML) Profile constructed using UML Class diagram notation. The aim behind its creation was to establish a unified language for developing ontologically correct conceptual models (Guizzardi, 2005b).

UFO (Guizzardi et al., 2015) is a resulting ontology of a research analysis of conceptual modelling languages with the aim of developing an ontological foundation for these languages. The research was motivated by the notion that explicit definition of fundamentals and adherence to a certain ontological commitment are crucial for conceptual modelling. Any attempt to develop foundations for conceptual modelling should take into account human knowledge and linguistic capabilities. In order to provide explicit definitions of entities and their relationships, crucial for obtaining a valid domain description, the UFO is enriched with theories from philosophical formal ontologies, cognitive sciences, linguistic logic, and philosophical logic (Guizzardi et al., 2015).

3 DATA MANAGEMENT PLANS

A DMP is essential for ensuring effective data management throughout the lifespan of a project. It describes the lifecycle of the data generated or collected, detailing how it will be managed and ensuring its future usability and accessibility. Having such a plan is essential for (research) data management (RDM), offering important insights into the origins, usage, and availability of data. Recently, the importance of data management planning has increased, especially among funding bodies and scientific institutions, to ensure that data remains useful beyond the original project. Good data management practices improve data collection accuracy, secure storage, and proper handling, enhancing data's value and relevance across various research fields (Smith, 1998).

Several resources are critical in defining the terminology used in a DMP, whether for human or machine readability. These resources include existing DMP templates, knowledge models from various DMP tools, and data management-related ontologies. To achieve the **G1** goal, it was essential to select and thoroughly analyse these resources.

3.1 DMP Templates

A DMP is commonly structured using a template to ensure all essential components are covered. However, certain sections may be adapted based on the specific project, funding source, or organization. For this work, the *Horizon Europe Template* (European Commission, 2020), *National Institutes of Health Data Management and Sharing Plan Template*(National Institutes of Health, 2023), and most importantly *Science Europe Template* (Science Europe, 2021) were selected due to their widespread adoption on a global scale. The concepts in these templates are recurring, although in different contexts, making additional templates unnecessary. The *Science Europe Template* (Science Europe, 2021) was the main source of knowledge because it aligns European DMP templates from various domains. This template covers core requirements that can be adjusted to specific needs, ensuring all essential components are addressed.

3.2 DSW Knowledge Models

One of the existing tools designed to help create and manage DMP is the Data Stewardship Wizard (DSW) (Pergl et al., 2019), recommended as an interoperability resource by ELIXIR and several other institutions like UB-BOTT (Universities of Norway).

This tool uses questionnaires, structured by so-called knowledge models defining specific questions and their interconnections; thus, offering a comprehensive perspective on the terms used for our work.

For the use of the DSW in the context of DMPs, several predefined knowledge models are available. These models are based on a mind map developed by Rob Hooft (Hooft, 2019). Although primarily focused on the natural sciences, the insights from this mind map are applicable to other domains as well. Among these models, the fundamental one derived from Rob Hooft's mind map is called the Common DSW Knowledge Model (DSW Team, 2018).

3.3 DMP-Related Ontologies

Numerous ontologies related to DMPs were analysed in (Martínková and Suchánek, 2023b), where nine ontologies were examined to identify overlaps and interconnections. In this work, some of these ontologies were used as dictionaries to understand the usage and context of various terms, leveraging the provided overview model (Martínková and Suchánek, 2023a).

As highlighted in (Martínková and Suchánek, 2023b), different terms are often used to define the same concept even within DMP-related ontologies. However, in the case of ontologies, the terminology is a secondary concern as meaning is established through relationships with other classes.

4 FAIR PRINCIPLES

FAIR (Wilkinson et al., 2016) was created in connection to conference *Jointly designing a Data FAIRPORT*. Result of this event was an agreement of creation of principles ensuring Findability, Accessibility, Reusability and Interoperability of data.

These principles do not prescribe specific implementation methods or technologies. Instead, they serve as a set of guidelines or approaches to achieve data reusability and accessibility. This openness in implementation can lead to inconsistencies in interpreting these principles. As noted in (Jacobsen et al., 2020) this can result in potentially incompatible implementations, which contradicts the original intent of the FAIR principles. To address this, the FAIR authors provided further explanations (Jacobsen et al., 2020) for the intended interpretations and implementation considerations for each principle.

In response to the ongoing controversy, an OntoUML model (Bernasconi et al., 2023) was developed to address the issues surrounding the interpretation of the FAIR principles. This model aims to clarify any ambiguities and uncertainties within the FAIR principles and provide guidelines for designing a dataset's FAIR classification based on a detailed analysis of these principles.

The model is divided into three parts, covering Findability and Interoperability, Accessibility, and Reusability. It addresses the overall concept of the FAIR principles as well as each sub-principle. The core of the model consists of *Data*, with their content described as *Data Items*, and their *Metadata*. This is supplemented with additional concepts to clarify and describe each sub-principle. The model employs an undefined relation called *externalDependence*, which does not align with OntoUML definitions. However, as we understand it, this relation makes sense within the context of the model. For our purposes, we retained this relation in the original model, but we did not use it in our extension due to our uncertainty about its proper definition and usage. Additionally, the model incorporates navigable relations, which are not supported by standard OntoUML; therefore, we have omitted them.

5 OUR APPROACH

To achieve the study's objectives, we initiated the development of the DMP OntoUML model by identifying the parts and concepts within the scope of the DMPs that the model must encompass. As is described in Section 3, various resources play a crucial role in defining the terminology used in the DMP. These resources were analysed to compile a comprehensive list of the necessary parts and concepts, detailed in Section 5.1. After establishing what needs to be included, each component was thoroughly analysed and incorporated into the OntoUML model. This effort aimed to meet the **G2** goal and link the DMP components of the model to the existing FAIR model, as described in Section 5.2. Finally, the model was validated to ensure syntax correctness, as stated in the **G3** goal, confirming its accuracy and completeness in representing the DMP, as described in Section 5.3.

5.1 Resource Analysis and Concept Identification

In order to determine all of the concepts needed for the model, we analysed various resources, including ontologies related to DMP, knowledge models serving as a knowledge base in DSW and DMP templates, as described in Section 3.

Using DMP templates and knowledge models from DSW, we determined the primary components of the domain. Below is a list of the main parts that need to be addressed according to our analysis. Our list includes all the essential details required by the *Science Europe Template* (Science Europe, 2021), which was created to align and cover requirements from various European DMP templates used by different funders and institutions. The model captures each item in detail, including concepts that might not be explicitly listed here.

- *Administrative Information* includes the project's name, identifier, start and end date.
- *Funding Informations* covers the involved funders, their identification, and the resources they offer.
- *Research Process* section covers data creation, reuse (including relevant considerations), and preservation. The Common DSW Knowledge Model (DSW Team, 2018) also includes data processing and interpretation; however, these aspects were excluded from our model as they detail automated steps, their compute environment, and visualizations, which are beyond the scope of this model. Nevertheless, these elements could be documented within the general resource documentation.
- *Data Preservation* deals with long-term strategies for data preservation and ongoing accessibility beyond the project lifecycle. Data preservation is, according to (DSW Team, 2018) a part of research process.
- *Data and Their Roles in the FAIR Context* includes descriptions of the data itself, datasets (including their format, value, purpose, identification, etc.), and their metadata.
- *Personal Data* covers related concepts such as informed consent and its potential reuse, accessibility to personal data, and the committee overseeing personal data.
- *Distribution* includes details about the data repository and the distribution itself.
- *Access and Reuse Requirements* includes authorization processes and tools requirements for accessing and reusing data.
- *Cost* includes all necessary resources during the project, especially concerning their availability, reusability, and preservation.
- *Compliance* focuses adherence to legal and ethical standards, including GDPR and other relevant regulations.
- *Members Engagement* encompasses involvement of various members in data management processes.
- *Data Quality Assurance* includes procedures and criteria for ensuring the accuracy, reliability, and validity of the data.

5.2 Implementation of the Model

As a foundational model for our development, we used the aforementioned FAIR model (Bernasconi et al., 2023) and extended it with concepts from the DMP domain. We aimed to align with the FAIR model by retaining all the core classes unchanged and distinguished them in grey in the developed model (Martínková et al., 2024). The main part of the FAIR model incorporates the DATA, METADATA and GROUND DATA (see Figure 1).

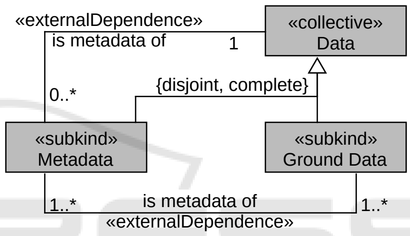

Figure 1: Data Representation within the FAIR model (Bernasconi et al., 2023).

In the FAIR model the DATA represents the dataset, while GROUND DATA represents data that cannot serve as metadata and therefore can be collected into a dataset. In DMP, the terms *datasets* and *data* are often used ambiguously. For instance, the *Horizon Europe Template* (European Commission, 2020) includes the following question (in Section 3.1 of the template): *Will all **data** be made openly available? If certain **datasets** cannot be shared...*

As seen, these two terms are used interchangeably within a single question, referring to the same concept with different terminology. According to (U.S. Geological Survey, nd), data and datasets are distinguished hierarchically: data, such as measurements or observations, can be organized into a structured collection, forming a dataset.

In the context of DMP, the term *dataset* is usually preferred because DMP typically refers to a structured collection of data currently being created or reused from a previous creation proces. While it is perfectly acceptable to use the term *data*, it should be used accurately, as this term is often overused. The absence of *dataset* class in the FAIR model may seem confusing; however, the authors of the FAIR model have effectively addressed mentioned nuances by distinguishing DATA and GROUND DATA.

The following sections explore the most contentious areas within the domain, focusing on the use and interpretation of terminology. These areas present challenges that require careful consideration to achieve a consistent and accurate understanding.

5.2.1 Data Access Evaluation

Data accessibility in DMP templates typically includes a detailed overview of the requirements necessary to access the data. This encompasses authorization protocols, authentication mechanisms, and the tools or instruments needed for data retrieval. Additionally, it outlines any potential restrictions that may be applied to data accessibility. Furthermore, these templates often describe the procedures for validating access requests through an access committee that ensures compliance with policies and regulations.

As accessibility is one of the FAIR principles, it is a core aspect covered by the FAIR model (Bernasconi et al., 2023). As shown in Figure 2, data accessibility requirements are supplemented by the necessary instruments and, most importantly, by the access request itself and its evaluation process. Typically, DMP templates inquire about the presence of a data access committee responsible for validating access requests based on the established data accessibility requirements. For instance, the *Horizon Europe Template* (European Commission, 2020) includes the question: *Is there a need for a data access committee?* Further details are more thoroughly elaborated in the DMP model (Martínková et al., 2024).

5.2.2 Data Availability

One term that is often incorrectly interchanged with *accessibility* is *availability*, and the domain of DMPs is no exception. In the FAIR model, accessibility is accurately captured as a role of DATA, which is fitting given that accessibility is one of the core principle.

In (Cambridge University Press, nda), *accessibility* is defined as the ability to be reached or obtained easily, whereas *availability* is defined in (Cambridge University Press, ndb) as the fact that something can be reached. In the context of data, *available* means that the data can be reached, without specifying by whom or how, or even if anyone is currently having the ability to access it—they are simply somewhere reachable. On the other hand, *accessible* data means we know how to reach them and who can access them, even if they are not fully-open. In other words, if data is available it does not ensure they are accessible to certain type of users.

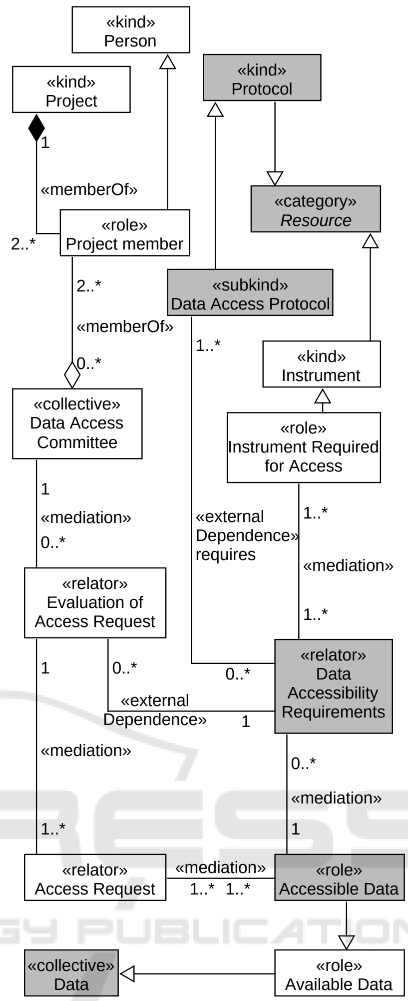

Figure 2: Part of the DMP model (Martínková et al., 2024) describing data access.

To accurately capture *availability* in the model as a role of DATA, we need to determine the criteria that establish this availability. What serves as the evidence or proof of the data's availability? According to the DCAT ontology (Albertoni et al., 2020), the class for distribution *dcat:Distribution* is defined as an available dataset. Therefore, the presence of an existing distribution acts as a strong indicator, or a "truthmaker", that the data is available. This means that if a dataset has been distributed, it can be considered available, as the distribution itself serves as a witness to the dataset's availability. The connection between available data and their distribution is included in the DMP model, see Figure 3.

To establish connections between roles or phases of data, it is essential to identify any connections or hierarchical relationships. From the definitions of accessibility and availability, we have already deduced that there is a hierarchical relationship: data must be available before it can be considered accessible.

The relationship between reusable and available data is similarly important. As noted in (Yoon et al., 2017), data availability is a prerequisite for reuse. The part of the model that captures these relationships is illustrated in Figure 3.

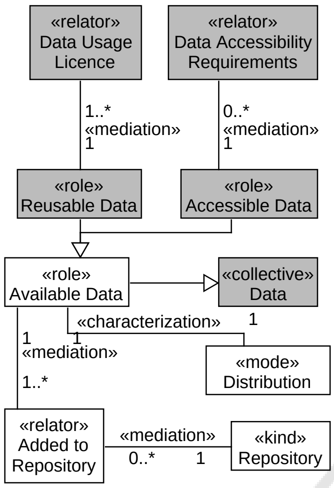

Figure 3: Part of the DMP model (Martínková et al., 2024) describing availability of data.

5.2.3 Projects Research

Another crucial aspect that needed detailed description is the research phase of the project, which involves the activities where data are collected. These research activities are directly tied to the project's objectives, as illustrated in Figure 4. An important part of research is the reuse, or more broadly, the consideration of reusing data. Questions about data consideration for reuse and actual reuse are typically part of DMP templates, with reuse consideration hierarchically above the actual reuse, as depicted in Figure 4.

Data creation is included, but reused data, which requires harmonization, is a special case. In case some data are reused and needs to be harmonized, this new data results from harmonization can be considered as created. Therefore, CREATED DATA has a subrole HARMONIZED REUSED DATA.

According to (DSW Team, 2018) a part of research process is also data preservation and ongoing accessibility beyond the project lifecycle. Therefore one of the activities of research is also DATA PRESERVATION, which is distinguished to preservation during and after the project.

5.3 Verification and Validation

To validate the model's syntax, we used the OpenPonk platform, an open-source tool for conceptual modeling, diagram development, and simulations (Uhnák and Pergl, 2016). The platform includes numerous extensions and modules for standardized notations, such as OntoUML, which includes a comprehensive framework for verifying OntoUML models (Bělohoubek, 2019). This extension ensures that all defined entities and relationships adhere to the specified rules of the OntoUML language. Additionally, it includes automatic detection of OntoUML anti-patterns (Bělohoubek, 2021), which identifies suspicious structures within the model that typically indicate errors as well. Using OpenPonk's OntoUML extension, we conducted a thorough validation of the model, thereby achieving the verification part of our goal **G3**. This involved systematically checking each entity and relationship to ensure they were correctly defined and aligned with OntoUML standards.

To ensure the model captures all required elements of the DMP domain, we used the *Science Europe Template* (Science Europe, 2021) as a base for our proposal. This template consolidates requirements from various European DMP templates, providing a comprehensive standard.

To further verify the completeness and accuracy of the proposed model, we created an instantiation model (Martínková et al., 2024) using an existing DMP that adheres to the *Science Europe Template*. We began by identifying relevant concepts in the DMP text, ensuring they aligned with the concepts in our proposed model and that our model contained all the required concepts. We then constructed an instantiation model based on these identified concepts. This approach allowed us to validate our model against a real-world example.

6 CONCLUSION

In this study, we addressed the critical issue of inconsistent terminology and incomplete information in DMP by developing an OntoUML conceptual model. We analysed existing DMP templates, knowledge models used in DSW tool and ontologies related to DMP in order to specify concepts that needs to be covered, in accordance with **G1**. For accomplishing **G2**, we selected concepts with their relations and developed an OntoUML conceptual DMP model extending the existing FAIR model.

To ensure the model's accuracy and completeness, we conducted a thorough validation using the Open-

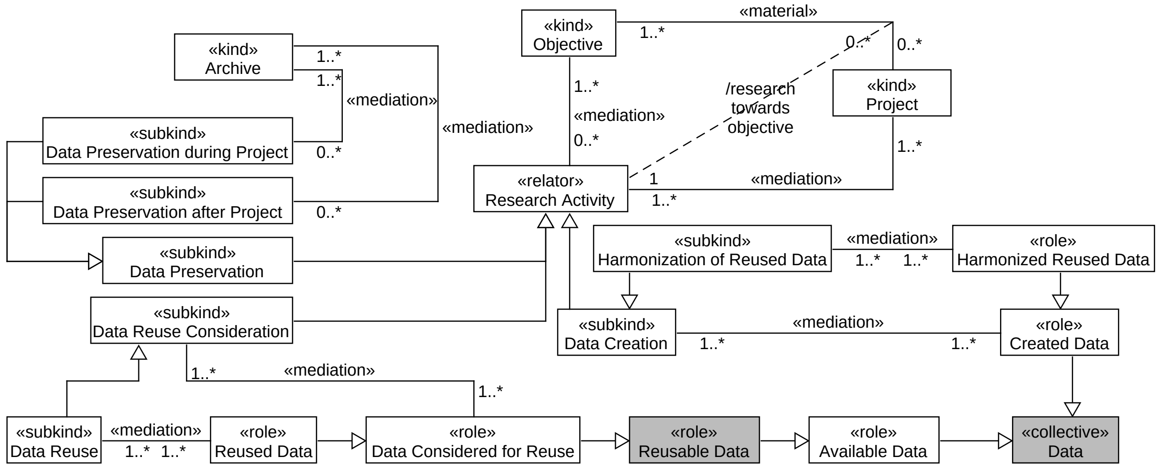

Figure 4: Part of the DMP model (Martínková et al., 2024) describing research.

Ponk platform and its OntoUML extension, achieving our goal **G3**. This validation process involved systematically checking each entity and relationship for correct definition and alignment with OntoUML standards. The platform's tools enabled us to identify and rectify any issues, ensuring the model's structural integrity and adherence to best practices in conceptual modelling. Additionally, we validated the proposed model using a real-world example by analyzing an existing DMP, identifying relevant concepts from our proposed model in its text to test for any missing concepts, and constructing an instantiation model based on those concepts.

This result DMP model captures essential concepts around DMPs to promote consistency and unambiguity, making the domain easily understandable without additional explanations or vocabularies. It opens floor for more connections and extensions that can go deeper in the details, while keeping the main core of the DMP domain. Finally, the model can also serve as a foundation for ontology development suitable for the area of DMP.

ACKNOWLEDGEMENTS

This work was supported by by the Czech Technical University in Prague grant: Advance Research In Software Engineering, No. SGS23/206/OHK3/3T/18 and ELIXIR LM (LM2023055).

REFERENCES

Albertoni, R., Browning, D., Cox, S., Beltran, A. G., Perego, A., Winstanley, P., et al. (2020). Data catalog vocabulary (dcat)-version 2. [Accessed 21-Jan-2023].

Bernasconi, A., Simon, A. G., Guizzardi, G., Santos, L. O. B. d. S., and Storey, V. C. (2023). Ontological representation of fair principles: A blueprint for fairer data sources. In Indulska, M., Reinhartz-Berger, I., Cetina, C., and Pastor, O., editors, *Advanced Information Systems Engineering*, pages 261–277, Cham. Springer Nature Switzerland.

Bělohoubek, M. (2019). OntoUML Models Verification for the OpenPonk platform. Bachelor's thesis, Czech Technical University in Prague, Faculty of Information Technology.

Bělohoubek, M. (2021). Extending OntoUML Modelling Capabilities on the OpenPonk Platform. Master's thesis, Czech Technical University in Prague, Faculty of Information Technology.

Cambridge University Press (n.d.a). Accessibility. [Accessed: 2024-06-12].

Cambridge University Press (n.d.b). Availability. [Accessed: 2024-06-12].

Dietz, J. L. and Hoogervorst, J. (2015). Enterprise engineering theories–introduction and overview. Technical report, CTU in Prague.

DSW Team (2018). Common DSW Knowledge Model. [Accessed 2023-03-19].

European Commission (2020). Horizon 2020 dmp. [Accessed 2023-12-15].

Gonzalez-Perez, C. (2018). *Benefits and Applications of Conceptual Modelling*, pages 17–21. Springer International Publishing, Cham.

Guizzardi, G. (2005a). *Ontological Foundations for Structural Conceptual Models*. CTIT PhD thesis series. Centre for Telematics and Information Technology, Telematica Instituut.

Guizzardi, G. (2005b). *Ontological foundations for structural conceptual models*. Phd thesis - research ut, graduation ut, University of Twente.

Guizzardi, G., Wagner, G., Almeida, J., and Guizzardi, R. (2015). Towards ontological foundations for conceptual modeling: The unified foundational ontology (ufo) story. *Applied ontology*, 10.

Hooft, R. W. W. (2019). Data stewardship mindmap.

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W., Imming, M., Jeffery, K. G., Kaliyaperumal, R., Kersloot, M. G., Kirkpatrick, C. R., Kuhn, T., Labastida, I., Magagna, B., McQuilton, P., Meyers, N., Montesanti, A., van Reisen, M., Rocca-Serra, P., Pergl, R., Sansone, S.-A., da Silva Santos, L. O. B., Schneider, J., Strawn, G., Thompson, M., Waagmeester, A., Weigel, T., Wilkinson, M. D., Willighagen, E. L., Wittenburg, P., Roos, M., Mons, B., and Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, 2(1-2):10–29.

Martínková, J. and Suchánek, M. (2023a). Conceptual Models for Data Stewardship Ontologies.

Martínková, J. and Suchánek, M. (2023b). Laying foundations for connecting data stewardship domain ontologies. In *New Trends in Intelligent Software Methodologies, Tools and Techniques*, pages 125–136. IOS Press.

Martínková, J., Suchánek, M., and Pergl, R. (2024). Conceptual model for data management domain. [Accessed 2024-06-23].

National Institutes of Health (2023). Data management & sharing plan. [Accessed 2023-08-13].

Pergl, R. (2019). Conceptualisation: Chapters from harmonising enterprise and software engineering.

Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., and Slifka, J. (2019). "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. *Data Science Journal*, 18:59.

Robinson, S., Arbez, G., Birta, L. G., Tolk, A., and Wagner, G. (2015). Conceptual modeling: Definition, purpose and benefits. In *2015 Winter Simulation Conference (WSC)*, pages 2812–2826.

Science Europe (2021). Practical guide to the international alignment of research data management-extended edition.

Smith, J. (1998). *The Book*. The publishing company, London, 2nd edition.

Uhnák, P. and Pergl, R. (2016). The openponk modeling platform. In *Proceedings of the 11th edition of the International Workshop on Smalltalk Technologies*, pages 1–11.

U.S. Geological Survey (n.d.). What are the differences between data, a dataset, and a database? [Accessed: 2024-06-12].

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

Yoon, A., Jeng, W., Curty, R., and Murillo, A. (2017). In between data sharing and reuse: Shareability, availability and reusability in diverse contexts. *Proceedings of the Association for Information Science and Technology*, 54(1):606–609.