Anna Bellotto, Cristiana Bettella, Linda Cappellato,
Yuri Carrer, Giulio Turetta

Modelling (Meta)Data in a Digital Repository

Methodological Tips in Practice

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 135–161
https://doi.org/10.25364/97839033742329

Anna Bellotto | ORCID iD: 0000-0003-1148-5456
Cristiana Bettella, University of Padua, Library Centre | ORCID iD: 0000-0001-5268-9522
Linda Cappellato, University of Padua, Library Centre | ORCID iD: 0000-0002-8649-4691
Yuri Carrer, University of Padua, Library Centre | ORCID iD: 0000-0002-1823-1646
Giulio Turetta, University of Padua, Library Centre | ORCID iD: 0000-0002-5430-6852

Abstract

Metadata play an essential role in making a digital object discoverable, accessible, usable, and interpretable. By adopting the conflation of data and metadata in the expression (meta)data modelling, the key aim of the contribution is to properly highlight the multi-layer dimension of the descriptive representation levels informed by a digital object, showing how this constitutes the foundational basis on which a (meta)data model is grounded. The contribution offers insights into the essential principles of shaping a (meta)data model, their applicability and interoperability challenges, with the aim to serve as an entry point for anyone interested in the field looking for good practices.

Keywords: (Meta)data modelling; digital object; FAIRness; metadata standards; interoperability; metadata application profile

Zusammenfassung

Modellierung von (Meta)Daten in einem digitalen Repository. Methodische Tipps für die Praxis

Metadaten spielen eine wesentliche Rolle dabei, ein digitales Objekt auffindbar, zugänglich, nutzbar und interpretierbar zu machen. Durch die Verschmelzung von Daten und Metadaten im Begriff der (Meta-)Datenmodellierung besteht das Hauptziel des Beitrags darin, die mehrschichtige Dimension der beschreibenden Repräsentationsebenen eines digitalen Objekts angemessen hervorzuheben und zu zeigen, wie diese die grundlegende Basis für ein (Meta-)Datenmodell bildet. Der Beitrag bietet Einblicke in die wesentlichen Prinzipien der Gestaltung eines (Meta-)Datenmodells, ihre Anwendbarkeit und die Herausforderungen der Interoperabilität, mit dem Ziel, als Einstiegspunkt für jeden zu dienen, der sich für dieses Gebiet interessiert und nach bewährten Verfahren sucht.

Schlagwörter: Metadaten; Datenmodellierung; Digitales Objekt; FAIRness; Metadatenstandards; Interoperabilität; Metadaten-Anwendungsprofil

1. Introduction

The theoretical and methodological decisions in modelling (meta)data lay the foundation of a digital repository while involving a substantial, and often challenging, combination of critical thinking, domain expertise, computational knowledge, user requirements, and therefore an evaluation of several heterogeneous aspects. The present contribution is a step towards this vast landscape by offering a holistic but practice-driven overview of what is meant by (meta)data modelling, delineating complementary benefits and complexities that may arise.

The reader will be accompanied through a hands-on agenda that reflects the key areas of the process of creating, implementing, managing, evaluating and disseminating (meta)data in digital repositories. The contribution opens with a preliminary orientation on basic concepts and definitions aiming to frame the discourse on common ground. Following this contextual introduction, the text then leads to an analytical exploration of the first topics of crucial priority: which theoretical decisions need to be made when defining a (meta)data model; to what extent the descriptive representation of data determines the richness of metadata; and which practical steps its implementation entails. The paper will raise awareness of existing methodological approaches, the primary role of community standards and the identification of the designated community. Simultaneously, it will seek to anchor a deep understanding of the model as a binomial of syntax and semantics, meant, on the one hand, as the language and structure of the model, or the design model tout court; on the other hand, as the capacity of formal representation provided by the model itself. Another section of the paper covers the central topics of interoperability and FAIRness of both human- and machine-actionable (meta)data. It will be shown how these principles have direct and profound relevance to the accessibility, quality, and discoverability of a digital repository while offering the concurrent benefit of enhancing its trustworthiness. Finally, the significance of tools and practices will be discussed by presenting specific examples of their use in order to efficiently and systematically analyse and evaluate the quality of the outcomes.

Placed within the layered but interrelated framework of digital repository management, this contribution will offer insights into the essential principles of shaping a (meta)data model, their applicability and technical challenges, with the aim to serve as an entry point for anyone interested in the field looking for good practices.

2. "Setting the stage"[1]

Metadata plays an essential role in making a digital object discoverable, accessible, usable, and interpretable. According to Berg-Cross, Ritz, and Wittenburg, if a digital object is defined as a complex entity that is "represented by a bitstream, is referenced and identified by a persistent identifier and has properties that are described by metadata", metadata "contains descriptive, contextual and provenance assertions about the properties of a digital object"[2]. By adopting the conflation of data and metadata in the expression (meta)data modelling, the multi-layer dimension of the descriptive representation levels informed by a digital object is properly highlighted, showing how this constitutes the foundational basis of a (meta)data model.

So what do we mean by (meta)data modelling and what is its context? Borrowing the definition offered by Flanders and Jannidis, we could state that "data modeling refers to the activity of designing a model of some real (or fictional) world segment to fulfill a specific set of user requirements using one or more of the metamodels available in order to make some aspects of the data computable and to enable consistency constraints"[3]. The next few paragraphs will attempt to dissect the cornerstones of the modelling process wholly encompassed in this definition, providing a foundational overview of what is considered one of the research practices at the heart of the Digital Humanities, and in a broader sense of the Digital Scholarship tout court[4].

In an integrated view that combines the "two cultures"[5] of humanities – here, philosophy and formal logic – and science – here, database design and software engineering –, data modelling emerges as an intellectual and computational activity, pushing towards what Willard McCarty claims to be a "Philosophy of modelling":

> the manipulatory essence of modelling, with its connotation of embodied action, physically or metaphorically; the mediating role and ternary relationship modelling establishes between knower and known; the directed, vector-like engagement of the inquirer's attention, through the model he or she has made to the object of study; and the model's consequent function as an artificial agent of perception and instrument of thought[6].

1 Gilliland, A. J. (2016)
2 Berg-Cross,G.; Ritz, R.; Wittenburg, P. (2015). For a further discussion on the concept of Digital Object, see below, p. 6f.
3 Flanders, J.; Jannidis, F. (2015), p. 15.
4 Ciula, A.; Eide, Ø.; Marras, C.; Sahle, P. (2018), pp. 7–29.
5 Snow, C. P. (1959)
6 McCarty, W. (2005), p. 55.

At its first step, as stated by Eide and Ore[7], the development of a data model cannot preclude an ontological analysis of the field or objects we are intended to describe in the digital realm. Within this context, the usage of the word "modelling" refers to the identification and description of the entities that form the part of the world a modeller is modelling, along with their relevant properties as well as their relationships[8].

On the one hand, this operation necessarily involves a substantial level of subjectivity. A model does not have the purpose of being a copy of the object it represents, capturing all the features it may depict[9]. Rather, it should just select the ones that are considered relevant, allowing "questions about the one [the object] to be answered by examining the other [the model]"[10]. The act of selection and implicit reduction inherently holds both the modeller's point of view and assumptions about that universe of discourse, as well as the intended usage of the objects being represented. Indeed, when modelling has a utilitarian as opposed to a pedagogical or self-reflexive goal[11], the most significant aspect that impacts the complexity and richness of the data model is its function[12]. The discussion around this factor will be unfolded below by illustrating the two main types of approaches – curation-driven as opposed to research-driven – arising from the different digitization communities.

On the other hand, by considering the dimensions of "adequacy" and "robustness", the task of modelling may still embed a form of objectivity. "Modeling does not simply mirror an external reality" – Jannidis and Flanders[13] comment – "but is an active process that depends on the social construction of a segment of the world". Having "a body of pre-existence knowledge" in common with their specific community of reference[14], modellers have the possibility to operate a negotiation of meaning between their own interpretations and the expectations of the community of peers. Aligning project-specific models with the community's understanding through the use of standard reference models, models can overcome the mere private context and reach mutual agreement and social consensus, while potentially

7 Eide, Ø.; Smith Ore, C.-E. (2019), p. 184.
8 Flanders, J.; Jannidis, F. (2019b), p. 28, p. 82.
9 Flanders, J.; Jannidis, F. (2019b) p. 28.
10 Sperberg-McQueen, C. M. (2019), note 8, p. 286.
11 Sperberg-McQueen, C. M. (2019), note 10, p. 286.
12 Flanders, J.; Jannidis, F. (2019b), p. 84.
13 Flanders, J.; Jannidis, F. (2019b), p. 90.
14 Pierazzo, E. (2019), p. 129.

successfully performing also across diverse settings and applications[15]. The relevance of standardisation in the realm of data modelling is another main topic that the following section will examine.

A conceptual model per se, however, cannot be treated and processed by a computer without first further operations. The features of the observed reality and its assumptions need to be formally and explicitly specified in a language that could be understood not only by humans but also by machines. At this subsequent stage of the developmental workflow, the word modelling enters the sphere of the technical implementation.

The codification of the abstract model in a machine-readable and actionable form distinguishes two levels of analysis to which two components of formal data modelling correspond: the metamodel and the data model. The metamodel refers to the formally defined syntax selected as an encoding format for the model representation. This syntax works as an organisational construct for the data structure, informing about the relationships and the information properties entities can hold, such as "hierarchy, inheritance, one-to-one vs. many-to-one relationships, cyclicality, nesting, sequencing, and so forth"[16]. The most widely used metamodels are the relational models used in database systems, the eXtensible Markup Language (XML) and the Resource Description Framework (RDF). All three model information in a different way: a relational database structures it as a table, XML as a tree, and RDF as a graph[17]. At this layer of analysis, formats and encodings serve as technical agreements that influence the potential exchange and reuse of data represented by the model. Their differences across multiple systems bear pivotal responsibility towards interoperability issues: as stressed by Zeng, "[w]ithout syntactic interoperability, data and information cannot be handled properly [...]"[18].

At the next level, independent of any encoding syntax selected at the layer below[19], the data model – also called schema – is the machine-processable translation of the ontological representation of the universe of discourse[20]. In these terms, a data model is a set of rules and constraints that express information about which data elements the modelled object is allowed or required to include, which attributes each element can have, how they must be ordered and how many times they can

15 Flanders, J.; Jannidis, F. (2019b), note 8, p. 90.
16 Flanders, J.; Jannidis, F. (2019a), p. 322.
17 Riley, J. (2017)
18 Zeng, M. L. (2019a), pp. 122–146.
19 Zeng, M. L. (2019b), note 18.
20 Tomasi, F. (2018), pp. 170-179.

appear, while concurrently providing data typing information[21]. The reference to community standards at this tier greatly benefits its corresponding layer of interoperability – i.e. the "structural layer" – illustrated by Zeng[22].

Within the paradigm of digital repositories, along with the key principles of data modelling, a clear understanding of the meaning of the concept of Digital Object (DO) should be offered to the reader. Digital objects, at a level of abstraction, can be considered as artefacts, encapsulating and virtualizing atomic elements that afford the online distribution of digital assets in terms of storage, dissemination, management, exchange and reuse. Starting from the seminal conceptualization given by Robert Kahn and Robert Wilensky in 1995's A framework for distributed digital object services, where the authors formally define a digital object as

> an instance of an abstract data type that has two components, data and key-metadata. The data is typed [...]. The key-metadata includes a handle, i.e., an identifier globally unique to the digital object; it may also include other metadata, to be specified[23],

we have been witnessing the evolution of the concept towards what the FAIR Digital Object Forum calls now "[a] technical essence of a 'thing' in cyberspace" binding "all critical information about an entity in one place and creat[ing] a new kind of actionable, meaningful and technology independent object that pervades every aspect of life today"[24]. A FAIR digital object (FDO), according to the technical definition[25], is therefore

> a unit composed of data that is a sequence of bits, or a set of sequences of bits, each of the sequences being structured (typed) in a way that is interpretable by one or more computer systems, and having as essential elements an assigned globally unique and persistent identifier (PID), a type definition for the object as a whole and a metadata description (which itself can be another FAIR digital object) of the properties of the object, making the whole findable, accessible, interoperable and reusable both by humans and computers for the reliable interpretation and processing of the data represented by the object[26].

21 Flanders, J.; Jannidis, F. (2019a), note 16, p. 328; note 17, p. 16.
22 Zeng, M. L. (2019b), note 18.
23 Kahn, R.; Wilensky, R. (2006), pp. 115–123.
24 https://fairdo.org/, where FAIR stands for Findable, Accessible, Interoperable, and Reusable.
25 https://fairdo.org/library/
26 Within the context of the Reference Model for an Open Archival Information System (OAIS), it might also be worth mentioning, it is the information object that yields the information represented by the data object – either a physical and a digital object as well –, and it is properly the

3. Transitioning from one generation to another

Digital innovations that occurred in the past few decades are exerting pressure on institutions to pursue their transition "to the next generation of metadata"[27]. Changes in standards, infrastructures and tools are having an impact on the way metadata are modelled and created, pushing forward a semantic evolution of the concept of metadata to Linked Open Data[28]. The outline of this current framework and its evolving modelling practices will be the topic of the next few paragraphs.

According to the Organization for the Advancement of Structured Information Standards (OASIS), a reference model is "an abstract framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment"[29]. In other words, a reference model represents a conceptual formalisation of a certain domain of knowledge, providing a common semantics that can be used unambiguously across and between different implementations[30]. As stressed already, a data model informs the design, as well as the conceptual structure of the data, from the double point of view of the syntax and the semantics assumed with respect to the reference information context. The degree of openness of a data model is expressed through the inherent capacity of being adaptable to different information contexts, and for information purposes and needs that might be diverse and unpredictable in principle. "By nature", Willard McCarty explains, "modelling defines a ternary relationship in which it mediates epistemologically, between modeller and modelled, between researcher and data or between theory and

knowledge representation of data content which the information system must guarantee to preserve, hence "data" provide "[a] reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing" (CCSDS 650.0-M-2, 2012, p. 10 and ISO 14721:2012: Space data and information transfer systems – Open archival information system (OAIS). Geneva, ISO 2012). Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen. OAIS identifies four parts to the digital object, i.e. an object composed of a set of bit sequence, described by them as the information package. These are the content information and the preservation description information, which are packaged together with packaging information, and which is discoverable by virtue of the descriptive information.

27 Smith-Yoshimura, K. (2020); Bahnemann, G.; Carroll, M.; Clough, P. et al (2021). Both reports served as background reading and inspiration for the eight virtual round table discussions promoted by OCLC Research, and held in six different European languages, throughout the month of March 2021. "How do we make the transition to the next generation of metadata happen at the right scale and in a sustainable manner, building an interconnected ecosystem, not a garden of silos?" was the primary question to lead the whole discussion among participants. An overview of the discussion series, accompanied by summary reports and recordings, is available from https://www.oclc.org/go/en/events/next-generation-of-metadata.html

28 Bellotto, A.; Bettella, C. (2019), pp. 167-184.

29 ISO 14721:2012, note 26.

30 Bekiari, C.; Bruseker, G.; Doerr, M. et al. (2021). Riva, P.; Le Bœuf, P.; Žumer, M. (2017).

the world"[31]. Hence, following Jannidis and Flanders[32], modelling data might be characterised by two main approaches: the curation-driven approach "which emphasizes the open-ended usefulness of the data", and the research-driven approach "where data is being created to support the creator's own research needs", both affecting and affected by the semantic extent from which the data model seeks to be elicited through the manipulative, iterative, and interactive process of modelling.

Such adaptive compliance – between modeller and model, and between data modeller and data model – defines the edge of the so-called data profile[33], by establishing a set of rules and constraints that should be declared into a documented schema, and certified by the adoption of community standards[34]. In these terms, a data profile formally translates the reference model into a (meta)data specification, becoming as such potentially applicable to other information contexts and for other information purposes, enabling the meaningful information integration and exchange. It acts as a Metadata Application Profile (abridged MAP) – the notion of which has been coined by the Dublin Core Community in 2000[35]: "a metadata design specification that uses a selection of terms from multiple metadata vocabularies, with added constraints, to meet application-specific requirements"[36]. Sliding over

31 McCarty, W. (2005), note 6, p. 24.

32 Flanders, J.; Jannidis, F. (2019b), note 8, p. 86.

33 "A profile identifies a set of base standards, together with appropriate options and parameters necessary to accomplish identified functions for purposes including: (a) interoperability, and (b) methodology for referencing the various uses of the base standards, meaningful both to users and suppliers" (ISO/IEC TR 10000-1:1998, quoted at https://www.loc.gov/z3950/agency/profiles/about.html by the Library of Congress, designated as Maintenance Agency and Registration Authority for ANSI/NISO Standard Z39.50 and ISO 23950:1998).

34 By way of example, and partially drawn from the typology of metadata standards outlined by Anne J. Gilliland (Gilliland: Setting the Stage (Note 1), based on the classification by Karim Boughida, 2005), we can distinguish: standards related to data structure, e.g.: MARC/BIBFRAME, Dublin Core Metadata Elements Set, MODS, VRA Core, EAD, TEI; to data content, e.g.: AACR2, RDA; to data value, e.g.: semantic artefacts such as subjects, classifications, thesauri, controlled vocabularies, ontologies; or standards for data exchange, e.g.: ISO 2709-2008, MARCXML, RDF, JSON-LD. A Metadata Standards Catalog applicable to scientific data, and maintained by the RDA Metadata Standards Catalog (MSC) Working Group, is available from https://rdamsc.bath.ac.uk/. For a graphic representation of the metadata landscape, it is well worth citing Riley, Jenn (2010): Seeing Standards. A Visualization of the Metadata Universe. Graphic design funded by the Indiana University Libraries White Professional Development Award: http://www.jennriley.com/metadatamap.

35 The coinage of MAP is by Rachel Heery and Manjula Patel, firstly introduced at the 8th Dublin Core™ workshop of October 2000 (http://www.ariadne.ac.uk/issue/25/app-profiles/). See also: Baker, Thomas (2011): Dublin Core™ Application Profiles at eleven years (2011). DCMI Blog posts: https://www.dublincore.org/blog/2011/application_profile/, and Coyle, Karen; Baker, Thomas (2013): Application Profiles as an alternative to OWL ontologies. In: DC-2013, Lisbon, Portugal.

36 https://www.dublincore.org/resources/glossary/application_profile/, ideally based on, or compatible with, vocabularies defined in RDF, and Coyle, Karen; Baker, Thomas (2009): Guidelines for Dublin Core™ Application Profiles. Dublin Core Metadata Initiative (DCMI): https://www.dublin-

the past two decades of history of data, the Program for Cooperative Cataloging (PCC) Task Group on Metadata Application Profiles has recently framed a comprehensive definition of MAP:

> A metadata application profile (MAP) is a set of recorded decisions about a shared data target for a given community. MAPs declare what models are employed (what types of entities will be described and how they relate to each other), what controlled vocabularies are used, the cardinality of fields/properties (what fields are required and which fields have a cap on the number of times they can be used), data types for string values, and guiding text/scope notes for consistent use of fields/properties. A MAP may be a multipart specification, with human-readable and machine-readable aspects, sometimes in a single file, sometimes in multiple files (e.g., a human-readable file that may include input rules, a machine-readable vocabulary, and a validation schema)[37].

From this perspective, we can argue that MAPs function as a crucial construct that performs and enhances semantic interoperability in its broadest sense[38]. Let us describe closely how this takes place in the contemporary landscape of modelling data.

Assuming that any data model is created with the goal of meeting local functional requirements, a schema that fulfils solely this necessity is usually of little use outside the specific implementation. When integration and interoperability across heterogeneous systems and applications are listed as additional targets, machine-readable meaning and context of structured data play a pivotal role. The disclosure of the intended meaning of data is what is needed by computers to decode data into knowledge and from the very beginning, computational ontologies and Linked Data technologies were conceived with this goal in mind.

Adapting it from the philosophical field to the digital environment, the computer science community started to use the term ontology to describe an engineering ar-

core.org/specifications/dublin-core/profile-guidelines/. It is worth mentioning at least the Europeana Data Model (EDM): https://pro.europeana.eu/page/edm-documentation (retrieved 07.01.2022), on which it is based the Metadata Application Profile of the Digital Public Library of America: https://pro.dp.la/hubs/metadata-application-profile (retrieved 07.01.2022); the DCAT Application Profile for Data Portals in Europe, based on the specification of the Data Catalog Vocabulary (DCAT) of 16 January 2014 and the Data Catalog Vocabulary (DCAT) – Version 2, W3C Recommendation, 04 February 2020: https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe. Linked Art based on CIDOC CRM: https://linked.art/

37 https://www.loc.gov/aba/pcc/taskgroup/Metadata-Application-Profiles.html

38 Zeng, M. L. (2019b), note 18.

tefact, which can be defined as a "formal, explicit specification of a shared conceptualisation"[39]. This specification commonly takes the form of a set of classes and properties representing concepts and relationships of that piece of the world, expressed through logical axioms usually based on the so-called Description Logics, i.e. a formal language for the knowledge representation that gives the capability of deducing new information from an explicated group of data[40]. As stressed by Eide and Ore, although there is sometimes some fuzziness when discussing models of the real world versus implementations of these models, ontologies should denote "a special kind of data model dealing with formalized conceptualizations rather than implementation issues"[41].

Thanks to their nature, ontologies have played a fundamental role in the development of the Semantic Web, an extension of the World Wide Web able to automatically read and process data and information without human intervention[42]. Providing context, i.e. a clearly expressed description of concepts, terms, and relationships within a knowledge domain, and serving as sources of shared meaning, ontologies affirmed their main significance towards this goal. Alongside, the new extension of the Web needed a set of established standards and technologies – what the term Linked Data collectively refers to – to create relationships among different datasets and formally explain to computers how to access and associate information[43]. Key standards for encoding and connecting data are the Resource Description Framework (RDF)[44], Web Ontology Language (OWL)[45] and Simple Knowledge Organization System (SKOS)[46], while Simple Protocol And RDF Query Language (SPARQL)[47] is the most used language to query and retrieve data.

39 Studer, R.; Benjamins, R. V.; Fensel, D. (1998), pp. 161-198. p. 184. For a detailed account of the notions of "conceptualization" and explicit "specification" according to Gruber's definition of ontology (Gruber, Thomas R. (1993): A translation approach to portable ontology specifications. In: Knowledge Acquisition 5 (2), pp. 199–220), while discussing the importance of *shared* explicit specifications introduced by Borst (Borst, Willen Nico (1997): Construction of Engineering Ontologies, PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands), see: Guarino, Nicola; Oberle, Daniel; Staab, Steffen (2009): What Is an Ontology? In: Staab, S.; Studer, R. (2009).

40 Biagetti, M. T. (2016), pp. 49–50.

41 Eide, Ø.; Ore, S. (2019), note 7, p. 182.

42 Berners-Lee, T.; Hendler, J.; Lassila, O. (2001)

43 https://www.w3.org/standards/semanticweb/data. Yoose, B.; Perkins, J. (2013), pp. 197-211.

44 https://www.w3.org/TR/rdf11-concepts/

45 https://www.w3.org/TR/owl2-overview/

46 https://www.w3.org/TR/skos-reference/

47 https://www.w3.org/TR/sparql11-query/

Given the possible outcomes of the use of Linked Data technologies, such as powerful data sharing and integration, efficient communication and meaningful information discovery, many repositories are currently investing time and efforts in the process of semantic enrichment, i.e. a procedure which consists in providing metadata of "more contextualized meanings" by expressing various types of relationship[48].

On the one hand, RDF gives the possibility to uncover the semantics of the data model by providing shareable and machine-processable definitions of metadata elements. Expressing information through the use of assertions called statements[49], the RDF format would require the components of a triple to be unambiguously identified through unique identifiers (URIs). In this context, using URIs that refer to external formal ontologies created by discipline-specific communities – to cite some examples: BIBFRAME[50] (for library resources), CIDOC CRM[51] (for cultural heritage data) or FRAPO[52] (for research project administrative information) – or developed with more general purposes – such as DCMI Metadata Terms[53] or Schema.org[54] –, is the key to enable computers to identify what the URI is and to understand the concept the metadata field refers to. The vast range of options that modellers deal with when selecting ontologies for the local data model requires a careful examination of the assumptions and agreements left implicit in the preliminary ontological analysis.

On the other hand, the link to thesauri, controlled vocabularies and classification schemes that adhere to the principles of Linked Open Data and are expressed in a standardised formal language (such as SKOS) allow metadata values populating the local datasets to gain substantial benefits. Namely, these are a more efficient interlinking with heterogeneous external resources described with the same concepts, an increased visibility and accessibility, and a potential supplementation of (multilingual) information, such as translated labels or broader labels. The use of standardised URI-values coming from authoritative and well-established LOD reference resources[55] – such as the Art & Architecture Thesaurus (AAT) by The Getty Research

48 Zeng, M. L. (2019b), p. 7.
49 In RDF each statement takes the form of a triple made of three elements: subject, i.e. the thing that is described; predicate, i.e. the property, attribute or relationship that is attributed to that thing in order to describe it; object, i.e. the value of that property.
50 https://www.loc.gov/bibframe/
51 https://www.cidoc-crm.org/Version/version-7.1.1
52 Shotton, D. (2017)
53 https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
54 https://schema.org/
55 https://id.loc.gov/

Institute[56] or the Resource Type vocabulary by COAR[57] – facilitate semantic interoperability, assuring that the meaning of the language and terminology used is correctly understood. As for the choice of ontologies illustrated before, the selection of the targets for the semantic enrichment at the value metadata level is far from trivial. It requires "a good knowledge of the source data in terms of topic coverage, gaps, quality issues" as well as a rigorous analysis to assure a matching granularity and coverage between sources and targets[58].

4. Metadata as a FAIR enabler

Metadata is key when dealing with the interoperability of digital objects managed, curated and archived in a repository. The data model of the digital repository should consider from the outset what characteristics of the metadata elements are necessary for effective interoperability. The semantics of the data model plays a fundamental role in ensuring the interoperability of digital assets so that they can be shared and reused by the user community.

Semantic interoperability "ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood'"[59] and can be defined as "the ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems"[60].

Semantic networks[61] – a network of semantic relations – are the lowest common level of strong semantic interoperability. Well-known thesauri can be used to achieve basic semantic interoperability[62]. Machine-actionability further strengthens interoperability allowing to query and aggregate relations from existing semantic networks, thus improving or even creating novel networks. According to this framework, adherence to community-driven standards enables a robust representation of domain-relevant data and metadata.

56 https://www.getty.edu/research/tools/vocabularies/aat/
57 https://vocabularies.coar-repositories.org/resource_types/
58 https://pro.europeana.eu/page/europeana-semantic-enrichment
59 https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/semantic-interoperability
60 Corcho, O.; Eriksson, M.; Kurowski, K. et al. (2021).
61 Pirnay-Dummer, P.; Ifenthaler, D.; Seel, N. M. (2012).
62 Hugo, W.; Le Franc, Y.; Coen, G. (2020), p. 14.

An advantageous and operative strategy to enable semantic interoperability in the context of digital repositories is the uptake of FAIR principles for data and metadata management, first introduced in the article The FAIR Guiding Principles for scientific data management and stewardship[63] in 2016. The FAIR principles consist of 15 high-level principles[64], providing guidelines to improve the findability, accessibility, interoperability and reuse of digital assets. Data and metadata can hardly be separated from each other when dealing with FAIR guidelines, although some principles state specifically what characteristics metadata should hold to be considered FAIR. As an example, Principle F3, which concerns the findability of data, states that metadata must explicitly mention the global and persistent identifier assigned to the described data. Therefore, if the data is assigned a DOI, Handle, or ARK identifier[65], the identifier must be encoded in the metadata. Letter I of the FAIR principles targets the interoperability of digital assets, and hence metadata plays a substantial role in it. The three principles involved are:

1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
2. (Meta)data use vocabularies that follow the FAIR principles
3. (Meta)data include qualified references to other (meta)data

Principle I1 states that the knowledge representation language of the metadata must be readable by humans and machines. To ensure machine readability, the RDF knowledge representation model and at least a subset of RDF serialisation formats, namely Turtle, RDF/XML, and Javascript Object Notation for Linked Data (JSON-LD) should be utilised. Principle I2 concerns the findability and documentation of Knowledge Organization Systems (KOS) used by the data model of the digital repository. The BARTOC.org registry[66], a database of KOS resources, is a well-established and reliable resource for finding FAIR vocabularies and ontologies. Finally, Principle I3 prescribes the use of qualified relationships in the digital repository. Relationships must be captured in the metadata so that meaningful links between digital objects (or parts of the objects) can be established. The more specific the meaning of the relationships, the greater the interoperability of the related digital assets. Suitably qualified relationships may come from general-purpose metadata schemas – such as the DCMI Metadata Terms isPartOf relationship – or discipline-specific ontologies – such as the Europeana Data Model has Type relationship.

63 Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J. et al. (2016).
64 https://www.go-fair.org/fair-principles/
65 The Library of Congress maintains the Standard Identifier Schemes list at https://id.loc.gov/vocabulary/identifiers.html
66 https://bartoc.org/

When considering the adoption of FAIR principles, it should be clear from the start who the designated community of the digital repository is. According to the CoreTrustSeal glossary[67], the designated community is

> an identified group of potential consumers who should be able to understand a particular set of information. The designated community may be composed of multiple user communities. A designated community is defined by the repository and this definition may change over time.

Namely, the designated community for a disciplinary research data repository consists of researchers, students, and scholars who interact with the repository (e.g. by depositing a dataset or making use of the data) or the end user of the data, or someone who should be able to use the data applying disciplinary standards.

Adhering to the FAIR principles also means understanding the level of FAIRness the digital repository aims to achieve. The FAIRness of a digital object can be defined as the measure of the extent to which the object is FAIR. A possible way to achieve the desired FAIRness level is by adopting the FAIR principles incrementally: the most useful and straightforward principles (concerning the repository needs) should be taken into account first. Identifying digital objects using globally unique and persistent identifiers (Principle F1) could be a starting point. Furthermore, recording repository details on re3data.org[68] and FAIRsharing.org[69] registries of data repositories and indexing the metadata of the objects in discovery tools such as OpenAIRE Explore[70] (Principle F4) are low-effort tasks that can improve the findability of digital objects and help pinpoint the strengths and weaknesses of the repository. Principle F4 could also include the use of custom search engines for FAIR datasets[71] and tools for FAIR metadata to be "presented on the Web"[72]. The I (Interoperability) subset of the principles can be addressed by exposing metadata elements through metadata crosswalks. Schema.org metadata schema for structured data could be a possible choice to ensure improvements in the interoperability of digital assets without modifying the underlying data model of the repository. To this end, a subset of entities and relationships, serialised as JSON-LD, from the Schema.org vocabulary can be easily embedded in the HTML source code of the landing page of the digital object. The subsequent and more challenging step would

67 CoreTrustSeal Standards and Certification Board (2022)
68 https://doi.org/10.17616/R3D
69 Sansone, S. A.; McQuilton, P.; Rocca-Serra, P. et al. (2019), pp. 358-367.
70 https://explore.openaire.eu/
71 https://www.dtls.nl/fair-data/find-fair-data-tools/
72 https://www.fairdatapoint.org/

be to design the repository data model using a comprehensive RDF representation of the metadata of the digital assets.

Several initiatives and projects have been initiated among the stakeholder community since the publication of the FAIR guiding principles. GO FAIR[73] is one of them. GO FAIR is an initiative that aims to promote the implementation of FAIR principles and the coherent development of a network of FAIR services. GO FAIR also proposes technical solutions for metadata that are not natively FAIR: metadata can undergo a FAIRification process[74] by means of FAIRification tools[75], or they can be exposed on FAIR Data Points[76]. A digital repository data model that needs to maintain backward compatibility with legacy metadata can consequently become FAIR without losing any previous feature.

To evaluate progress in adopting FAIR principles, the use of metrics and assessment tools should be considered. For this purpose, the FAIRsFAIR project[77] has developed the FAIRsFAIR Data Object Assessment Metrics and two practical tools, FAIR-Aware and F-UJI[78]. It is worth noting that the FAIRsFAIR object metrics and tools assume the assessment of research data objects as a subset of the possible digital objects that a repository can manage; therefore, the results of the assessment should be tailored to the specific use case of digital objects.

FAIR-Aware[79] is a disciplinary-agnostic self-assessment tool intended to raise awareness about FAIR principles. It consists of a questionnaire comprising ten questions provided with practical tips. Comprehensive understanding and adoption of the FAIR principles are key for the designated community; FAIR-Aware can help create a liaison between the community and the repository managers. The self-assessment tool could be included as part of the repository deposit workflow to improve data and metadata quality and user community awareness.

The FAIRsFAIR metrics are mainly drawn from the set of indicators provided by the RDA FAIR Data Maturity Model Working Group[80]. The FAIRsFAIR metrics focus on what and how these indicators can be evaluated in practice[81].

73 https://www.go-fair.org/
74 https://www.go-fair.org/fair-principles/fairification-process/
75 https://github.com/FAIRDataTeam/OpenRefine-metadata-extension/
76 https://www.go-fair.org/how-to-go-fair/fair-data-point/
77 https://fairsfair.eu/
78 Devaraju, A.; Mokrane, M.; Cepinskas, L. et al. (2021), pp. 1-14.
79 https://fairsfair.eu/fair-aware
80 RDA FAIR Data Maturity Model Working Group (2020)
81 Devaraju, A.; Mokrane, M.; Cepinskas, L. et al. (2021), note 78.

F-UJI[82] is an automated assessment tool for programmatically assessing published digital objects for their level of FAIRness. Provided an identifier of the object, the tool performs practical tests on the FAIRsFAIR Data Object Assessment Metrics. As an example, to check compliance with Principle I3 involving the FsF-I3-01M metric, F-UJI performs the following tests:

- Related resources are explicitly mentioned in the metadata
- Related resources are indicated by machine-readable links or identifiers

The first test succeeds if at least one relationship is found in the metadata. For instance, the test checks whether the is Part Of relationship, which might state that an object belongs to a collection of objects, is listed in the object metadata. The second test checks if the identifier of the related resource points to the resource through a resolvable URI. F-UJI allows one to set up an iterative workflow to incrementally adopt the FAIR principles by using the tool to test the improvements made to the repository data model after every implementation step.

In the process of adopting the FAIR principles, the foundational elements of knowledge representation in the Semantic Web and Linked Data landscape come into play, namely RDF and triples, especially when interoperability, FAIRness, and long-term preservation are the main objectives. RDF is a way to represent the world using triples that comprise a subject, a predicate, and an object. Asserting that 'Beethoven is the author of the Ninth Symphony', a suitable triple could be:

- Subject: the Symphony No. 9
- Predicate: has author
- Object: Beethoven

This approach is formal enough for a consistent representation of knowledge[83]. Once a set of triples is created, an RDF graph is generated. The graph consists of subjects and objects of the triples as vertices and predicates as edges. This kind of graph is known as the graph of knowledge. Metadata and relationships become triples as well. Furthermore, controlled vocabularies, ontologies, and classification schemes can be represented as triples.

A knowledge graph does not offer by itself a way to query it or retrieve data from it. To query a knowledge graph, the SPARQL query language is needed. SPARQL is a semantic query language that can be used by humans and automated agents to query a SPARQL endpoint. For example, if a 'title' property exists in the metadata of a digital object – i.e. the 'has title' predicate of an RDF triple – the endpoint can

82 Devaraju, A.; Huber, R. (2021)
83 Oldman, D.; Doerr, M.; Gradmann, S. (2016)

be queried for that metadata element using an SQL-like syntax. More complex queries can be built, for instance, to list the titles of objects belonging to a collection or to find the depositors of a set of objects alongside the number of objects they have deposited. The repository should provide a public SPARQL endpoint for querying the metadata elements that describe the digital assets. A well-known example of a public SPARQL endpoint is the Wikidata Query Service[84].

RDF is suitable for exposing metadata of digital objects as well as for their internal representation by the Digital Asset Management System (DAMS)[85] of the repository. Metadata are internally represented as documents, usually stored in RDF/XML or JSON-LD serialisation formats. The DAMS transforms the metadata elements into RDF triples, usually stored in a database. These triples can then be indexed by search engine software[86] for internal repository functionality or external uses that could comprise end-user searches and queries by agents on machine-readable endpoints.

As the internal representation of metadata may not be unique, repositories should support bulk uploads of digital objects, data manipulation tasks, and external data services queries on common ground. Therefore, repositories should develop an Application Programming Interface (API). An HTTP REST API[87] uses the HTTP protocol to send data to or retrieve data from the repository, enabling a simple interface to exchange and operate the data in the repository. It is best practice to detail the support level (timespan, rate limits) for each published version of the API specifications and the standards being implemented. The URL of the API endpoint should also be recorded on the aforementioned public repository registries. API responses return metadata depending on the metadata serialisation formats available in the repository. XML and JSON are commonly used formats that ensure a high degree of interoperability. A repository that models metadata following RDF specifications could benefit from exposing metadata in the JSON-LD format, an RDF serialisation format that encodes linked data by means of the JSON format. A reliable repository API can enhance the confidence of the designated community in the repository and increase the use and reuse of digital assets by external data services.

84 https://query.wikidata.org/. Several example queries can be found at https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples

85 Fedora is a DAMS platform commonly utilised by digital repositories. Fedora documentation can be found at https://duraspace.org/fedora/

86 Digital repositories often use Solr https://solr.apache.org/ or ElasticSearch https://www.elastic.co/elasticsearch/ as search engine software

87 https://csrc.nist.gov/glossary/term/rest

Generally, digital assets stored in a repository must remain unchanged for a prolonged period of time to ensure long-term preservation. Metadata again comes into play in the form of an integrity check indicator. These metadata values are not usually disclosed to end users, but doing so can improve the trustworthiness of the repository. A nearly effortless approach is to compute a fingerprint of the data and metadata of the digital object in the form of short sequences of alphanumeric characters using Message-Digest 5 (MD5) checksums. The result of the computation is 32-character sequences that uniquely represent the content of the data and metadata. The computed fingerprint is stored and recomputed when users modify the metadata[88]. Whenever something unexpected occurs to the data or metadata, the recomputed checksums will differ from the stored ones. Secure Hash Algorithm (SHA) is a more advanced hash algorithm and a common alternative to the MD5 algorithm, but can serve the same purpose. An automated procedure should be set up to periodically check if data and metadata undergo any unexpected change by recalculating the checksums and comparing them to the stored ones. As periodic recalculation might be computationally expensive, especially when a large number of objects are involved, the procedure can be scheduled weekly or monthly or even performed offline on data backups. A digital repository should choose the best approach for its use case, but managing integrity check metadata is a cornerstone to ensure the long-term preservation and reliability of archived data and metadata.

5. Quality of metadata

Metadata quality is a fundamental aspect of a digital repository, but it is not so simple to deal with it since no general consensus has been reached on 'what metadata quality is': it is a multidimensional and context-specific concept[89], so it is not possible to provide universally valid indications on how to manage metadata quality. An efficient metadata quality control is essential to realise a good repository. It helps to ensure the ease of finding the desired objects, user satisfaction, data interoperability, and increase the possibility of data sharing and reuse. The presence of quality metadata, in fact, allows the user to find digital objects that respond correctly to his request, to interpret them, to understand their context and consequently to reuse the data.

Metadata errors are more common than one might think. For this reason, it is necessary to provide control and verification systems and try to provide the necessary

88 The same MD5 checksum could potentially have the same value for different binary contents, but it is very rare (from a statistical point of view) that if an unexpected error occurs in the data the MD5 checksum remains the same.

89 Tani, A.; Candela, L.; Castelli, D. (2013), pp. 1194-1205.

tools to metadata creators to reduce such errors as much as possible. The errors generally concern typing or spelling mistakes, inconsistency in the formatting of dates, fields in the metadata editor that are incomplete or left blank, incorrect use of punctuation and separators, inaccuracies in the attribution of keywords and entering values in the wrong fields.

Thinking about the quality of metadata means taking into consideration different aspects, in particular: criteria for assessing the quality of metadata, quality control procedures and mechanisms to ensure quality.

The parameters that can be used to evaluate the quality of metadata are many, and each author who has dealt with this topic has identified different characteristics to be evaluated and different metrics to be used. The main ones are as follows[90]:

- Completeness: each digital object should be accompanied by all the metadata necessary to be found and correctly interpreted. There is no degree of completeness that is generally valid because it may depend on the type of resources and some choices of the repository managers.
- Accuracy: the metadata must be accurate both in terms of content and in terms of form.
- Logical consistency and coherence: this is the most difficult element to ensure, especially when several people create metadata in the same repository. Having consistent metadata within the repository means that the same value is always used to express the same concept and that each value is used with the same meaning for different objects; it also means consistently using the different fields of the metadata set, always inserting the same type of value in a given field.
- Conformance to expectations: this parameter is very strictly related to context and depends on the characteristics of the repository and of the user community.
- Timeliness
- Accessibility
- Provenance: refers to the origin of the metadata (who created them or, in case of machine-derived metadata, how they were generated)
- Shareability: it depends on the interoperability of the repository (e.g. use of standard communication protocol). For a single digital object to be shareable, it should also have consistent and complete metadata, thus providing context.

90 Park, J.-R.; Tosaka, Y. (2010), pp. 696-715. Hillmann, D. I.; Dushay, N.; Phipps, J. (2004). Tani, A.; Candela, L.; Castelli, D. (2013). McCarthy, K. (2015).

Creating good quality metadata means ensuring that they remain meaningful even outside the local context, for example, when (meta)data are shared with other repositories. Interoperability plays a fundamental role: when metadata exit from the original context, the difficulty of maintaining high metadata quality grows exponentially (the purpose can change, some information can get lost, etc.).

Quality control can take place before or after the ingest phase of the data (to also check the display of the data) and can be[91]:

Primary (by those who create the metadata, also through batch mechanisms):

- Fill in all required fields
- Filling the fields with the correct values
- Absence of typing and spelling errors
- Correct formatting, for example in the use of separators or date formats
- Secondary (by those who create the metadata or other experts):
- The ability of entered values to accurately describe digital objects
- Consistency of metadata
- Completeness

Some measures can help improve the quality of metadata. Repository staff should, for example, ensure accurate training of the staff in charge of metadata creation, provide tools for requesting support, regularly update the staff, and also carry out recurring checks and report any inaccuracies found to the creator of the metadata to avoid repetition.

Also, some tools can be useful to guarantee metadata quality, in particular:

- Clear guidelines and examples
- An intuitive metadata editor, with drop-down menus and links to thesauri, which can help improve data consistency
- Mandatory fields to ensure at least a basic completeness of metadata
- Use of widespread and well-established vocabularies, ontologies, standards, both to reach data consistency and to improve interoperability of data into the Semantic Web
- Templates and other tools that make compilation easier, for instance, tools for automatic creation of metadata and conversion scripts that transform metadata already created and checked (for example, for bibliographic catalogue) into the format required by the repository

91 UCLA Library (2015)

- Data quality checklists, to facilitate self-analysis of data quality by the creator of the metadata

Writing guidelines is especially important when there are multiple people depositing digital objects in the repository and compiling metadata. Two different people, in fact, can interpret the meaning of the fields in a slightly different way or decide to enter the same information in different fields. In addition, some formal aspects, such as the formatting of the dates or the use of separators, can differ according to the preferences and habits of those who compile (for example, a date can be expressed as 06 September 2021 or 09/06/21 or even 6th September 21). To maintain uniformity within the repository, it is therefore necessary to have a certain rigidity in the metadata editor or to provide clear indications through guidelines or contextual help. Furthermore, the existence of guidelines is an excellent reminder for those who upload digital objects only occasionally. It is much easier to create high-quality metadata from the beginning than to intervene in existing metadata to improve its quality, so it is important that tools to ensure quality and control mechanisms are planned from the beginning. To give a practical example, you can think of a repository with digital objects uploaded by several subjects, where the field to enter the keywords is a free text field. In the absence of clear indications (e.g. guidelines), we will easily have differences in the use of separators and inconsistency/incoherence in the use/meaning of terms. This situation may confuse users and represents a problem with interoperability.

Measures to clean up and harmonise pre-existing content are long and complex, they involve many manual interventions, object-by-object, and require getting in touch with formal metadata creators one by one. Additionally, some inaccuracies may get out of control and remain excluded from the corrections made.

It is much easier (and much more efficient) to design a repository with clear guidelines and practical examples, links to widespread thesauri and ontologies and 'stricter' fields to enter values (in this example, if each keyword were to be entered in a different field, there would be no problem with separators). However, providing quality tools and control mechanisms is not enough: the process of FAIRification of (meta)data, together with continuous technical improvements of the repository and of other databases related to it, implies recurring updates of the guidelines, tools and existing data. It is also necessary that the repository manager knows as much as possible about the content of the repository to ensure a rapid update of tools and workflows and to understand which objects need to be intervened with from time to time and what types of improvements are necessary. A reasoned work

of metadata reconciliation should be considered to enrich metadata and improve their quality[92].

6. Final remarks

The reader has been accompanied through a hands-on agenda that reflects the key areas of the process of creating, implementing, managing, evaluating, and disseminating (meta)data in digital repositories. In this analytical exploration, the viewpoint of the model as a binomial of syntax and semantics, meant as the design model tout court and the capacity of formal representation provided by the model itself, has been stressed. Additionally, technical implementations towards the improvement of semantic interoperability and FAIRness have been described, highlighting how they can have a direct and profound relevance to the accessibility, quality and discoverability of a digital repository.

Grounded on the direct experience in managing repositories dealing with a heterogeneity of digital assets, the strategies and approaches illustrated throughout this contribution have been drawn from the acknowledgement to sharpen a needed balance between machine and human actionability in the process of modelling data. By way of practical expertise, it has also been demonstrated that the awareness of semantic interoperability, the increment of FAIRness, and the trustworthiness of data and metadata are fundamental to pave the way for meeting the requirements for repository certification. Among the various existing certifications, achieving the CoreTrustSeal[93] certification can build stakeholder trustfulness in the repository while ensuring a core-level standardisation of processes and good practices. The repository will then become a trustworthy digital repository capable of fulfilling data management mandates from national and international funding bodies.

Bibliography

Bahnemann, Greta; Carroll, Michael; Clough, Paul et al. (2021): Transforming Metadata into Linked Data to Improve Digital Collection Discoverability: A CONTENTdm Pilot Project. Dublin, OH: OCLC Research. https://doi.org/10.25333/fzcv-0851

Bekiari, Chryssoula; Bruseker, George; Doerr, Martin et al. (eds.) (2021): ISO 21127. Information and Documentation – A Reference Ontology for the Interchange of Cultural Heritage Information, Geneva: ISO 2014; ISO/IEC 21838-1:2021: Information Technology – Top-Level Ontologies (TLO) – Part 1: Requirements. Geneva: ISO 2021; ICOM/CIDOC: Definition of the CIDOC Conceptual Reference Model. Produced by the ICOM/CIDOC

92 Khalid, H.; Zimanyi, E.; Wrembel, R. (2018). Tillman, R. K. (2016). See also https://freeyourmetadata.org/reconciliation/

93 https://www.coretrustseal.org/

Documentation Standards Group, Continued by the CIDOC CRM Special Interest Group. Version 7.1.1. https://www.cidoc-crm.org/version/version-7.1.1 (retrieved 07.01.2022)

Bellotto, Anna; Bettella, Cristiana (2019): Metadata as Semantic Palimpsests. The Case of PHAIDRA@unipd. In: Manghi, Paolo; Candela, Leonardo; Silvello, Gianmaria (eds.) (2019): Digital Libraries: Supporting Open Science. IRCDL 2019. (Communications in Computer and Information Science 988). Cham: Springer, pp. 167–184. https://doi.org/10.1007/978-3-030-11226-4_14

Berg-Cross, Gary; Ritz, Raphael; Wittenburg, Peter (2015): Data Foundation and Terminology Work Group Products. https://doi.org/10.15497/06825049-8CA4-40BD-BCAF-DE9F0EA2FADF

Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001): The Semantic Web. A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities. In: Scientific American 284 (5), pp. 1–5.

Biagetti, Maria Teresa (2016): An Ontological Model for the Integration of Cultural Heritage Information: CIDOC-CRM. In: JLIS.it 7 (3), pp. 49–50. http://dx.doi.org/10.4403/jlis.it-11930

Ciula, Arianna; Eide, Øyvind; Marras, Cristina et al. (2018): Modelling. Thinking in Practice. An Introduction. In: Historical Social Research, Supplement 31, pp. 7–29. https://doi.org/10.12759/hsr.suppl.31.2018.7-29. The authors have collected the results of their combined efforts to clarify the use and role of models in humanities research supported by computational methods in: Ciula, Arianna; Eide, Øyvind; Marras, Cristina et al. (2023): Modelling Between Digital and Humanities. Thinking in Practice. Cambridge: Open Book Publishers. https://doi.org/10.11647/OBP.0369

Corcho, Oscar; Eriksson, Magnus; Kurowski, Krzysztof et al. (2021): EOSC Interoperability Framework. Report from the EOSC Executive Board Working Groups FAIR and Architecture. Luxembourg: Publications of the European Union. https://data.europa.eu/doi/10.2777/620649

CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2023-2025. https://doi.org/10.5281/zenodo.7051125

Devaraju, Anusuriya; Huber, Robert (2021): An Automated Solution for Measuring the Progress Toward FAIR Research Data. In: Patterns 2 (11) 100370. https://doi.org/10.1016/j.patter.2021.100370

Devaraju, Anusuriya; Mokrane, Mustapha; Cepinskas, Linas et al. (2021): From Conceptualization to Implementation. FAIR Assessment of Research Data Objects. In: Data Science Journal 20 (4), pp. 1–14. https://doi.org/10.5334/dsj-2021-004

Eide, Øyvind; Ore, Christian-Emil Smith (2019): Ontologies and Data Modeling. In: Flanders, Julia; Jannidis, Fotis (eds.): The Shape of Data in the Digital Humanities. Modeling Texts and Text-based Resources. New York: Routledge, pp. 178–196.

Flanders, Julia; Jannidis, Fotis (2019a): Glossary. In: Flanders, Julia; Jannidis, Fotis (eds.): The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources. New York: Routledge, pp. 331–351.

Flanders, Julia; Jannidis, Fotis (2019b): A Gentle Introduction to Data Modeling. In: Flanders, Julia; Jannidis, Fotis (eds.): The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources. New York: Routledge, pp. 26–98.

Flanders, Julia; Jannidis, Fotis (2015): Knowledge Organization and Data Modeling in the Humanities. White paper. urn:nbn:de:bvb:20-opus-111270

Gilliland, Anne J. (2016): Setting the Stage. In: Baca, Murtha (ed.): Introduction to Metadata. 3rd ed. Los Angeles: Getty Publications. https://www.getty.edu/publications/intro-metadata/ (retrieved 04.04.2023)

RDA FAIR Data Maturity Model Working Group (2020): FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. https://doi.org/10.15497/RDA00050

Hillmann, Diane I.; Dushay, Naomi; Phipps, Jon (2004): Improving Metadata Quality. Augmentation and Recombination. In: DC-2004--Shanghai Proceedings. https://dcpapers.dublincore.org/pubs/article/view/770 (retrieved 17.08.2023)

https://www.library.ucla.edu/help/services-resources/digital-projects-for-special-collections/ (retrieved 25.01.2024)

Hugo, Wim; Le Franc, Yann; Coen, Gerard et al. (2020): D2.5 FAIR Semantics Recommendations Second Iteration (1.0). https://doi.org/10.5281/zenodo.5362010

Kahn, Robert; Wilensky, Robert (2006): A Framework for Distributed Digital Object Services. In: International Journal on Digital Libraries 6 (2), pp. 115–123. https://doi.org/10.1007/s00799-005-0128-x

Khalid, Hiba; Zimanyi, Esteban; Wrembel, Robert (2018): Metadata Reconciliation for Improved Data Binding and Integration. In: Kozielski, Stanisław; Mrozek, Dariusz; Kasprowski, Pawel et al. (eds.): Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety. 14th International Conference, BDAS 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings. (Communications in Computer and Information Science 928). Cham: Springer, pp. 271–282. https://doi.org/10.1007/978-3-319-99987-6_21

McCarthy, Kate (2015): Guide to Metadata Quality Control (2015–2019). Digital Repository of Ireland. https://doi.org/10.7486/DRI.sj13pg68d-1

McCarty, Willard (2005): Humanities Computing. New York: Palgrave Macmillan.

Oldman, Dominic; Doerr, Martin; Gradmann, Stefan (2016): Zen and the Art of Linked Data. In: Schreibman, Susan; Siemens, Ray; Unsworth, John (eds.): A New Companion to Digital Humanities. 2nd ed. Chichester: Wiley-Blackwell, pp. 251–273. https://doi.org/10.1002/9781118680605.ch18

Park, Jung-Ran; Tosaka, Yuji (2010): Metadata Quality Control in Digital Repositories and Collections. Criteria, Semantics, and Mechanisms. In: Cataloging & Classification Quarterly 48 (8), pp. 696–715. https://doi.org/10.1080/01639374.2010.508711

Pierazzo, Elena (2019): How Subjective Is Your Model? In: Flanders, Julia; Jannidis, Fotis (eds.): The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources. New York: Routledge.

Pirnay-Dummer, Pablo; Ifenthaler, Dirk; Seel, Norbert M. (2012): Semantic Networks. In: Seel, Norbert M. (ed.): Encyclopedia of the Sciences of Learning. Boston: Springer. https://doi.org/10.1007/978-1-4419-1428-6_1933

Riley, Jenn (2017): Understanding Metadata. What is Metadata, and What Is It For? Baltimore: National Information Standards Organization. https://www.niso.org/publications/understanding-metadata-2017 (retrieved 04.04.2023)

Riva, Pat; Le Bœuf, Patrick; Žumer, Maja (2017): IFLA Library Reference Model. Den Haag: IFLA 2017. https://repository.ifla.org/handle/123456789/40

Sansone, Susanna Assunta; McQuilton, Peter; Rocca-Serra, Philippe et al. (2019): FAIRsharing as a Community Approach to Standards, Repositories and Policies. In: Nature Biotechnology 37, pp. 358–367. https://doi.org/10.1038/s41587-019-0080-8

Shotton, David (2017): FRAPO, the Funding, Research Administration and Projects Ontology. https://sparontologies.github.io/frapo/current/frapo.html (retrieved 04.04.2023)

Smith-Yoshimura, Karen (2020): Transitioning to the Next Generation of Metadata. Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343

Snow, Charles Percy (1959): The Two Cultures. Cambridge: Cambridge University Press (= The Rede Lecture).

Sperberg-McQueen, C. Michael (2019): Playing for Keeps: The Role of Modeling in the Humanities. In: Flanders, Julia; Jannidis, Fotis (eds.): The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources. New York: Routledge.

Staab, Steffen; Studer, Rudi (eds.) (2009): Handbook on Ontologies. International Handbooks on Information Systems. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92673-3

Studer, Rudi; Benjamins, Richard V.; Fensel, Dieter (1998): Knowledge Engineering. Principles and Methods. In: Data & Knowledge Engineering 25 (1–2), pp. 161–198. https://doi.org/10.1016/S0169-023X(97)00056-6

Tani, Alice; Candela, Leonardo; Castelli, Donatella (2013): Dealing With Metadata Quality. The Legacy of Digital Library Efforts. In: Information Processing & Management 49 (6), pp. 1194–1205. https://doi.org/10.1016/j.ipm.2013.05.003

Tillman, Ruth Kitchin (2016): Extracting, Augmenting, and Updating Metadata in Fedora 3 and 4 Using a Local OpenRefine Reconciliation Service. In: The Code4lib Journal 31. https://journal.code4lib.org/articles/11179 (retrieved 04.04.2023)

Tomasi, Francesca (2018): Modelling in the Digital Humanities. Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain. In: Historical Social Research. Supplement 31, pp. 170–179. https://doi.org/10.12759/hsr.suppl.31.2018.170-179

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan et al. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. In: Scientific Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Yoose, Becky; Perkins, Jody (2013): The Linked Open Data Landscape in Libraries and Beyond. In: Journal of Library Metadata 13 (2-3), pp. 197–211.

Zeng, Marcia Lei (2019a): Interoperability. In: Knowledge Organization 42 (2), pp. 122-146. Also available in: Hjørland, B.; Claudio Gnoli, C. (eds.): Encyclopedia of Knowledge Organization. https://www.isko.org/cyclo/interoperability (retrieved 04.04.2023)

Zeng, Marcia Lei (2019b): Semantic Enrichment for Enhancing LAM Data and Supporting Digital Humanities. In: El profesional de la información 28 (1) e280103, p. 7. https://doi.org/10.3145/epi.2019.ene.03 (retrieved 25.01.2024)

Anna Bellotto graduated in Italian Philology and Digital Humanities, has been focusing on topics of data curation, metadata modelling and controlled vocabularies. She previously worked with the team of Europeana Foundation and the digital repositories of the University of Padova and University of Vienna.

Cristiana Bettella graduated in Romance Philology and Digital Humanities, works at the Digital Library Office of the University of Padua Library Centre as Metadata and Electronic Resources Services coordinator. She is engaged in digital scholarship, data modelling and curation of the digital repository Phaidra.

Linda Cappellato graduated in Archival and Library Science at the University Ca' Foscari of Venice, works at the Digital Library Office of the University of Padua Library Centre. She is involved in services related to digital collections, institutional archives, open science, virtual exhibitions and information literacy.

Yuri Carrer graduated in Computer Engineering at the University of Padua, in his thesis dealt with digital objects, an activity he prosecuted since 2003 at the Library System of the University of Padua. He has developed the Padua@ institutional research and research data archives and manages Phaidra as technical manager, dealing with the FAIRification of the repository.

Giulio Turetta graduated in Telecommunications Engineering, works as Digital Services Librarian at the Digital Library Office of the University of Padua Library Centre. He is a manager of the library discovery service and the Phaidra digital repository.