Marcel Riedel, Fabian Riebschläger, Giulia Russo

Terms and concepts of publishing and citing information resources in archaeology and beyond. A perspective from the CiVers project and the iDAI.world.

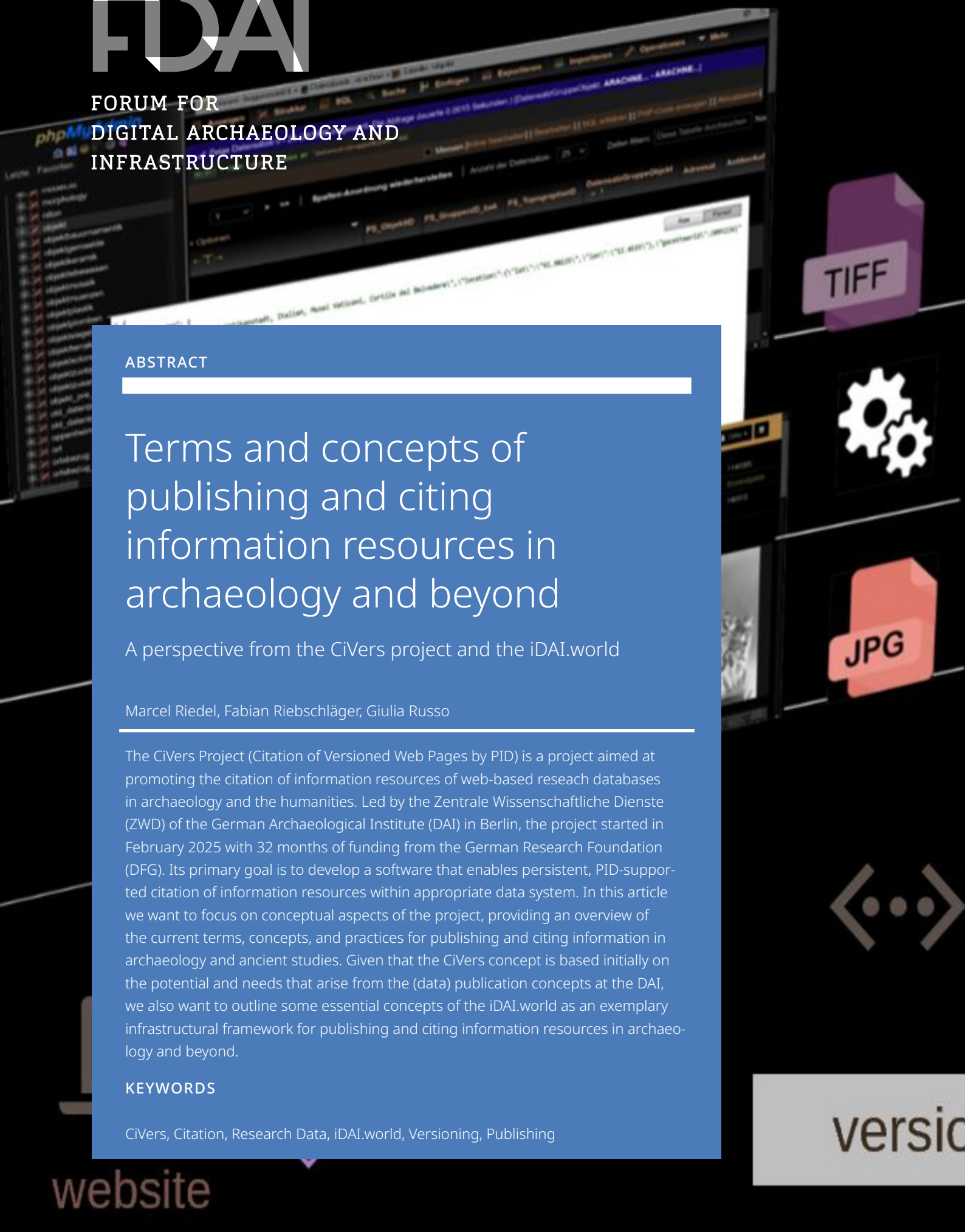
Forum for Digital Archaeology and Infrastructure Faszikel 2025 1–34 (§)

https://doi.org/10.34780/6k764r03

Herausgebende Institution / Publisher:
Deutsches Archäologisches Institut

AN ARTICLE FROM THE

FORUM FOR
DIGITAL ARCHAEOLOGY AND
INFRASTRUCTURE



ABSTRACT

Terms and concepts of publishing and citing information resources in archaeology and beyond

A perspective from the CiVers project and the iDAI.world

Marcel Riedel, Fabian Riebschläger, Giulia Russo

The CiVers Project (Citation of Versioned Web Pages by PID) is a project aimed at promoting the citation of information resources of web-based reseach databases in archaeology and the humanities. Led by the Zentrale Wissenschaftliche Dienste (ZWD) of the German Archaeological Institute (DAI) in Berlin, the project started in February 2025 with 32 months of funding from the German Research Foundation (DFG). Its primary goal is to develop a software that enables persistent, PID-supported citation of information resources within appropriate data system. In this article we want to focus on conceptual aspects of the project, providing an overview of the current terms, concepts, and practices for publishing and citing information in archaeology and ancient studies. Given that the CiVers concept is based initially on the potential and needs that arise from the (data) publication concepts at the DAI, we also want to outline some essential concepts of the iDAI.world as an exemplary infrastructural framework for publishing and citing information resources in archaeology and beyond.

KEYWORDS

CiVers, Citation, Research Data, iDAI.world, Versioning, Publishing

MARCEL RIEDEL, FABIAN RIEBSCHLÄGER, GIULIA RUSSO

Terms and concepts of publishing and citing information resources in archaeology and beyond

A perspective from the CiVers project and the iDAI.world

Introduction

1 The CiVers project (Citation of Versioned Web Pages by PID)[1] is an initiative aimed at promoting the citation of information resources of web-based research databases in archaeology and the humanities. Led by the Central Scientific Services[2] of the German Archaeological Institute (DAI) in Berlin, the project started in February 2025 with 32 months of funding from the German Research Foundation (DFG).[3] Its goal is to develop a software (and methodologies) that enable persistent citation of information resources within appropriate data systems and without requiring extensive modifications to database software or existing data models.

2 Key components of the CiVers project include the development of reusable information infrastructure software[4] and its implementation as a service for information systems such as iDAI.objects (object catalogs) and iDAI.field (excavation documentation) as case studies.[5] Essential aspects of the project

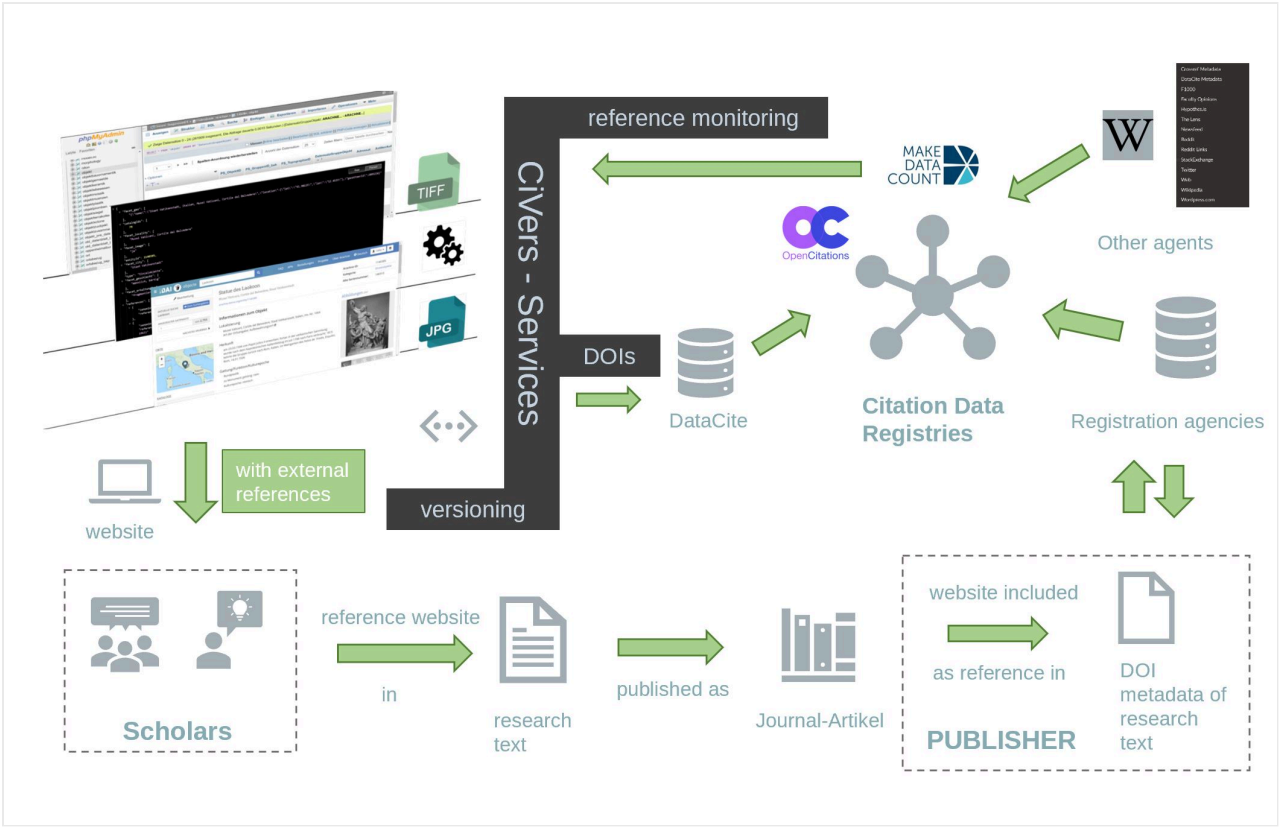
1 CiVers project page at dainst.org: https://www.dainst.org/en/research/projects/citation-of-versioned-web-pages-by-pid-civers/5927.

2 Zentrale Wissenschaftliche Dienste (ZWD) at dainst.org: https://www.dainst.org/dai-standorte/zentrale/wer-wir-sind/organisation/zwd. We thank Henriette Senst, Reinhard Förtsch, Sebastian Cuy, Wolfgang Schmidle, Simon Hohl, Daniel de Oliveira, Benjamin Ducke, Peter Baumeister, Benedikt Boxyen, Lisa Steinmann and Ammar Saeed for their support and input throughout the course of this ongoing project. We also thank Federico Buccellati for his suggestions and for proofreading this article in English.

3 CiVers project at GEPRIS (DFG): https://gepris.dfg.de/gepris/projekt/537548255.

4 DFG-Term: “Informationsinfrastruktur-Software”.

5 It is also planned to integrate research databases of our project partners, Propylaeum and GBV (VZG), as further test cases for CiVers in an early phase. For example, see Propylaeum Vitae: https://sempub.ub.uni-heidelberg.de/propylaeum_vitae/.



1

Fig. 1: Schema: CiVers in a nutshell

involve leveraging web archiving techniques, integrating open interfaces for DOI registration (e.g. via DataCite) along with their metadata and persistent identifiers (PIDs), and continuously extracting external references from citation event data registries (Fig. 1).

3 The effectiveness of resource (or knowledge) graphing[6] within global information infrastructures depends largely on the ability of publication service providers to ensure that relationships between self-distributed and external resources are represented through machine-readable metadata. To enhance the impact of the project's technical components, CiVers is complemented by community-driven initiatives aimed at evaluating principles, practices, and techniques of (data) citation and publishing.

4 In this article, addressing researchers and infrastructure stakeholders, we focus on conceptual aspects of the project, providing an overview of the current terms[7], concepts, and practices for publishing and citing information in archaeology and ancient studies - without aiming for a systematic analysis. Alongside the continuous development of CiVers as information infrastructure

6 CiVers is addressing contributions to the (DataCite) PID Graph and other schorlary network graphs: https://support.datacite.org/docs/datacite-graphql-api-guide#the-pid-graph. For the potential of the PID-Graph, see Cousijn 2021. Stocker 2021. To learn more about knowledge graphs (e.g. Wikidata, DataCite PID Graph) in general - often built using Linked Open Data (LOD) principles - see Schmidt 2022. They provide a very clear and understable introduction to Linked Open Data within the archaeological domain. Task Area 5 of NFDI4Objects provides the NFDI4Objects Knowledge Graph as a Discovery Services (DiS), developed and maintained by the GBV (VZG): https://www.nfdi4objects.net/portal/services/nfdi4objects-graph/

7 The concepts defined here in separate text boxes reflect our definitions of various terms that are central to CiVers.

software, more extensive surveys and workshops about data publishing and citation practices are planned during the project to consider the various aspects and concepts – described below - in depth. Given that the CiVers concept is based initially on the potential and needs that arise from the (data) publication concepts at the DAI, we also want to outline some essential concepts of the iDAI.world as an exemplary infrastructural framework for publishing and citing information resources in archaeology and beyond.

Publishing and citing information resources in archaeology and ancient studies

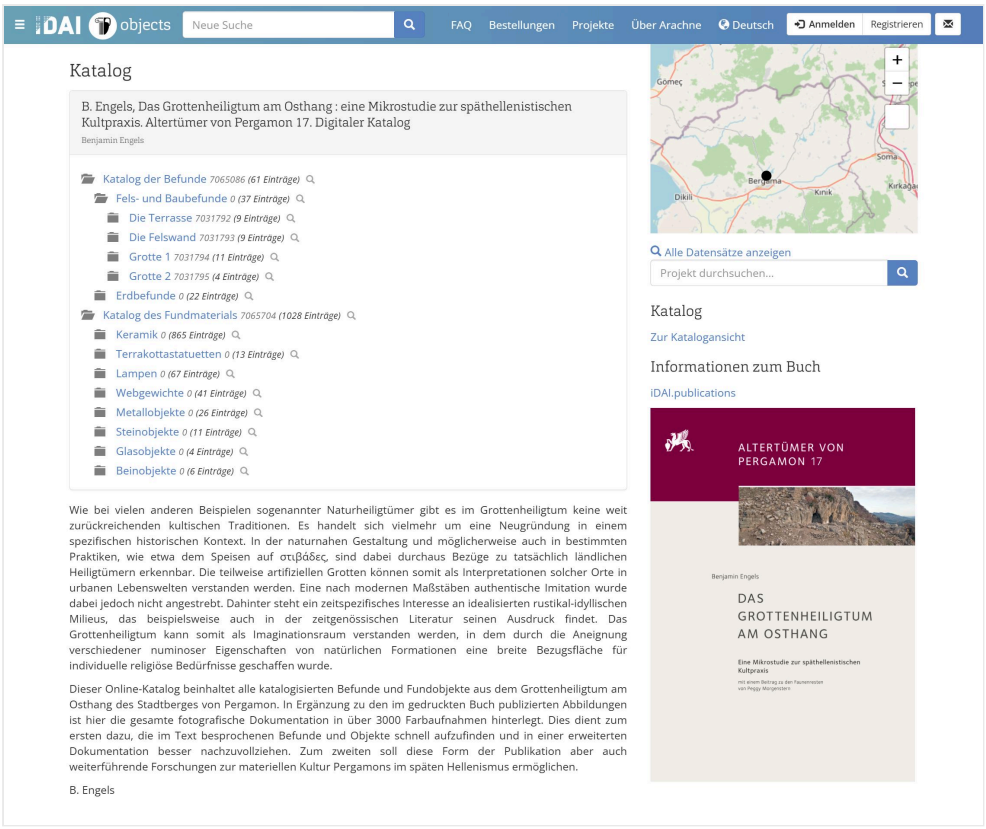
5 Scientific publishing is currently undergoing a technological transformation (the digital turn) with far-reaching implications from accessibility (open science, open access, open data) and responsibility roles to financial, legal, and social factors such as reputation and participation in the scientific processes (citizen science).[8] This shift is driven by the availability of new digital formats for publishing (or sharing) research texts and research data. Today, the primary communication space for science is no longer the (physical) library but the Internet.

6 The term publication format is commonly – and often indifferently – used for organizational, structural, representational and technical forms related to the organization of the distribution of research outputs, such as research texts and research data.

> **Research texts** are written documents that describe, analyze, and present the findings, methodologies, and conclusions of a research project. Research texts provide context and a synthetic view to research data. **Research data** are raw or processed (factual) information collected, generated, or analyzed during a research project. Research Data (e.g. catalogs, images, databases, data tables, GIS data, protocols, software code) form the empirical basis for research texts.

7 Taking into account journals and book series that are institutionally organized by editors and publishers, publication formats can be related to organizational forms providing credibility (peer review, editorial oversight) and an established channel (platform) for research dissemination of research texts and data. The Journals and Books Series as a publication venue (platform) use standardized publication types (structural forms), including traditional formats such as journal articles, conference papers and monographs, but also increasingly formats such as data papers and data publications. On a representational level, these formats can be separated further from forms that are defined by distribution media (print, digital/online) and by design and interactivity features

8 Comprehensively and diversely presented by Taubert 2016. See also Wissenschaftsrat 2022 (recommendations for transforming academic publishing into open access) and Breuer 2021 on expanding the concept of "publication". On the ongoing transformation of scientific publishing see also Kuhlen 2023. On „Open Science" see Tochtermann 2023.



2

Fig. 2: Project page representing a digital catalog in iDAI.objects. The catalog was published as supplementary content to: B. Engels, Das Grottenheiligtum am Osthang: Eine Mikrostudie zur späthellenistischen Kultpraxis. Mit einem Beitrag zu den Faunenresten von Peggy Morgenstern, AvP 17, 2022, doi: 10.34780/21b5-01te.

(paged-media or „paper-views“, enhanced web views, file format related data views, e.g. tables). Last but not least, all formats are available in specific file formats that are often derived from other file formats (e.g. .docx, .csv, .tiff).

A **data paper** – or data article – is a scholarly publication that describes datasets or data collections in detail, especially its creation, structure and potential reuse – without presenting traditional research conclusions or the entire data itself. It focuses on making data discoverable, citable, and reusable by providing metadata (e.g. collection methods, formats, spatial/temporal coverage), data access instructions (e.g., repository links, licenses) and contextual information (e.g. research purpose, quality controls). Data Papers are often published in peer review journals, for example the Journal of Open Archaeology Data (JOAD) that features peer reviewed data papers describing archaeology datasets with high reuse potential.

8 In the humanities, scientific texts continue to be published in both analog and digital formats. However, digital research data is primarily disseminated in the file formats and database systems in which it was originally created (born-digital). Unlike traditional publication formats such as printed books,

research data cannot be fully displayed in terms of quantity and quality through conventional publishing methods. In both practice and academic discourse, as well as in science policy strategy papers, research data is increasingly recognized as a reusable and citable research output, deserving equal consideration and recognition as an integral component of the overall research process.[9] Within the representational and technical scope there are often multiple formats for one information object like a single research text. Certain information objects (e.g. databases, 3D-Models) are often not fully transferable into conventional paper-formats, other information objects (e.g image data) are transferable as derivatives (illustrations), but tend to lose information without a parallel distribution in appropriate forms (e.g image databases or repositories). Digital catalogs (**Fig. 2**) and excavation documentations (**Fig. 3**), published via iDAI.objects or iDAI.field are examples of transferring complex information objects into (web) representations that were formerly (and still are) arranged in traditional publication formats.

Web representations are the various (layout) forms in which any kind of content is collected, rendered or delivered on a web page or website. These forms can include visual, textual, or interactive formats. Representations are often intentionally compiled (by system providers) as a collection of information with different levels of granularity and organized according to semantic data models. The way to represent information objects on a web page (like a dataset of an archaeological object) often differs significantly from the representations on the backend side (e.g. data are stored in various tables and images in a file system). On the frontend or client side it is a document and shares design concepts of research texts or data arranged on paper pages.

A **web page** is a single document that is (mostly) delivered by web servers on the World Wide Web and accessible through a web browser. It is typically written in HTML (HyperText Markup Language), styled by CSS (Cascading Style Sheets) and integrates JavaScript for interactivity. It may include additional multimedia content like images, videos, or audio. Web pages are often part of a larger collection of related pages, known as a website (e.g. a research database). They are often dynamically generated on the fly and linked together through hyperlinks. Web pages can be archived and are comparable to small documents (or pages that are part of a book).

9 Data Citation Principles 2014: “Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications”. See Pfeiffenberger 2017, 333-335, for a summary of the positions and developments regarding the question of “Zitationswürdigkeit” of research data.

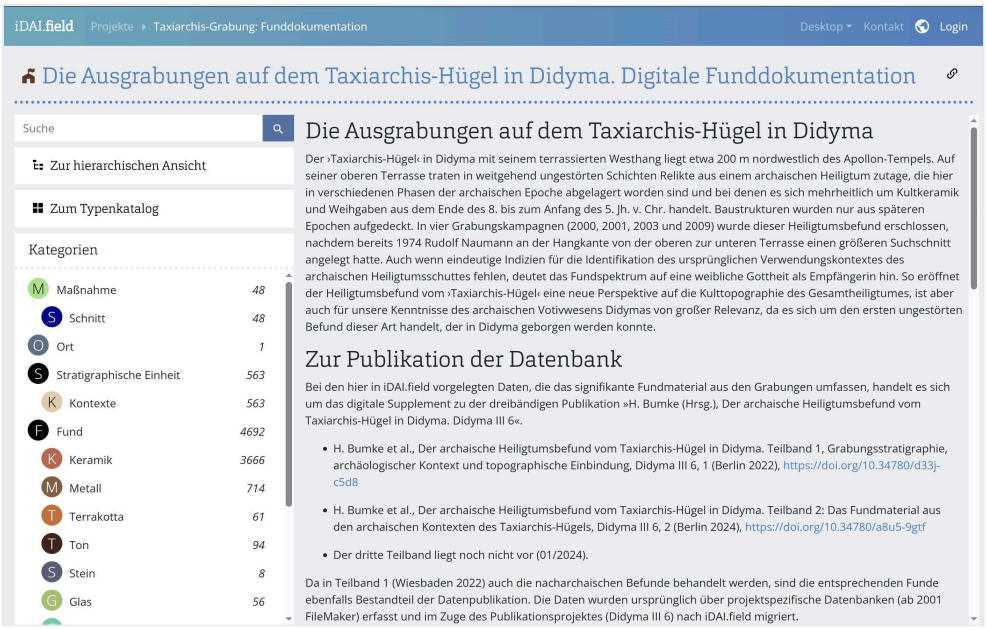


Fig. 3: Example of a digital find documentation published via iDAI.field, see: https://field.idai.world/project/didyma-project

3

9 Contrary to the prevailing discourse, we additionally – when describing it here – need to distinguish between two domains: publishing of articles and books – or research texts – on the one hand and sharing or publishing research data on the other. These two domains are still often loosely connected and differ in various aspects (e.g. publication formats, responsible service providers or work units).[10]

Publishing articles and books

10 In the research culture of archaeology and ancient studies[11], which is grounded in the examination and documentation of material and written legacies, these legacies are central to scholarly publications. Whether in the form of excavation reports, exhibition catalogues, or historical-epigraphic source editions, the publication of extensive material and source catalogs – including numerous illustrations, drawings, and plans – is a key focus. These empirical components of a publication are often the most frequently used, enduring for long periods and being referenced across diverse study contexts. In contrast, the interpretative results and narrative classifications can become outdated within a decade. A distinctive feature of the scholarly culture in this field is the accessibility of the material legacies themselves. Their development and conservation are regulated by national and international antiquities authorities, museums, and archives. Various requirements and control mechanisms (such as excavation permits, publication rights, and obligations) are in place to ensure that

10 The increasing prevalence of data papers can be seen as an intermediating element between these domains. The German Archaeological Institute has started to support data paper submissions for regular DAI-Journals (e.g. Archäologischer Anzeiger).
11 Among others: Classics, Egyptology, Assyriology, Philology, Epigraphy, Numismatics, History.

documentary research respects both scientific quality and preservation qualities related to archaeological records itself.[12] These projects often lead to the production of more extensive doctoral and postdoctoral theses.

11 In terms of publication formats, scholarly publications in ancient studies typically fall within conventional categories such as journals, monographs, and conference proceedings. The monograph – a book of considerable length – remains a popular format, as it allows for the presentation of extensive material catalogs. Newly discovered or previously known objects are also featured, though to a lesser extent, in journal articles. However, the primary focus of these journals is often on brief reports about ongoing excavation and research projects, overviews, short individual studies, and treatises on theory and methodology. The same is true for specialist articles published in collected works ("Sammelwerke"), typically conference proceedings. These contributions – written in various languages – are often short discussions in which new material is rarely introduced. Preprint formats, by contrast, have been somewhat uncommon in the ancient studies. Finally, we should also highlight museum and exhibition catalogues, as well as travel and archaeological site guides[13], which are particularly relevant to ancient studies. These "trade books" primarily target a broad public, including tourists, but also serve as valuable resources for the academic exploration of material monuments. These publications are almost exclusively distributed in print and available through museum shops.

12 In contrast to the life and natural sciences, where digital publishing predominates, publication practices in the humanities and cultural studies still rely on a print culture.[14] In these fields, digital publication formats and techniques (including persistent identifiers and FAIR principles) play a minor or secondary role so far. As a result, scholarly disciplines and many publication service providers[15] remain primarily focused on the production, distribution, and management of traditional print publications, especially monographs.

Persistent identifiers (PIDs) are long-lasting, unique references to digital objects, resources, or entities (such as documents, datasets, researchers, or organizations). They are defined as unique combinations of numbers and letters (e.g. https://doi.org/10.5438/axvs-my78) that enable reliable and consistent referencing of knowledge resources. Open, non-commercial PID systems (e.g. DOI, ORCID, ROR) have been developed to meet the needs of the scientific community, contributing to making research data FAIR. They ensure accurate attribution and disambiguation, enhance citation tracking and linking, facilitate open and FAIR data principles, and support building knowledge graphs and performing scientometric analysis.

12 Archaeological excavations are inherently destructive processes; as such, they cannot be replicated. This underscores the need to provide thorough, precise, and accessible documentation of the research conducted in the field.
13 "Site guides" in the form of printed travel guides to individual cultural heritage sites are, particularly for archaeological sites in non-European countries, sometimes the only up-to-date literature available on these sites and their material remains.
14 See also WR 2022, 43-44.
15 Söllner 2017, 248.

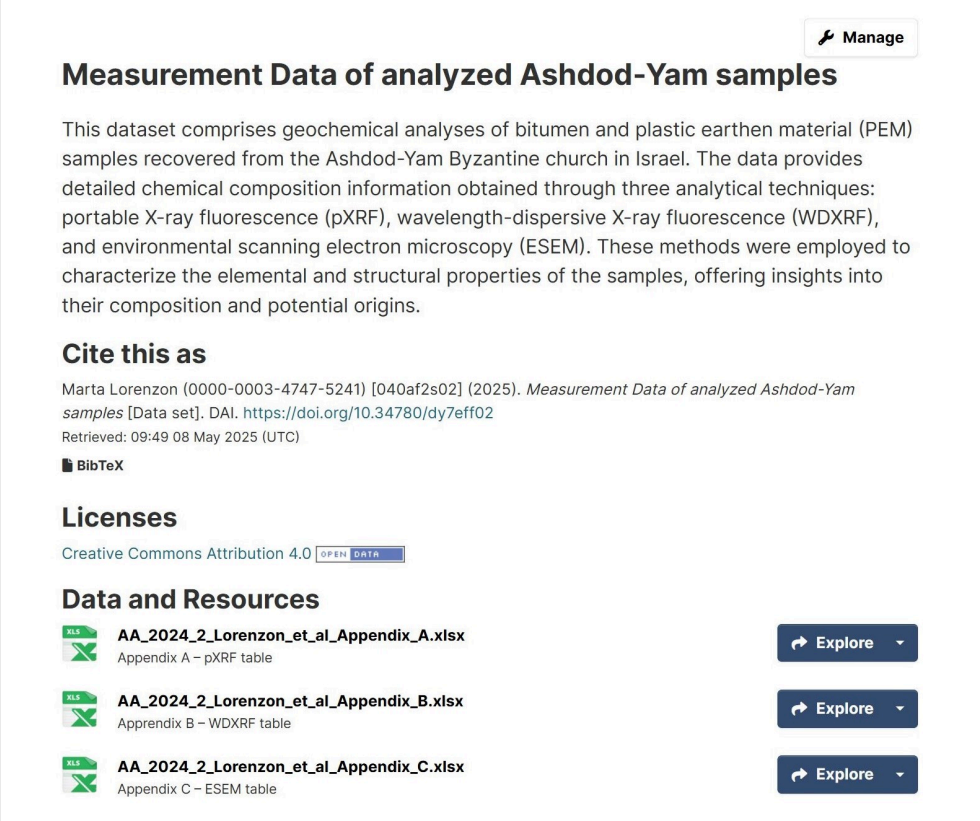


Fig. 4: Detail of a dataset in iDAI.repo: M. Lorenzon, Measurement Data of analyzed Ashdod-Yam samples, iDAI.repo (2025), 10.34780/dy7eff02.

4

The **FAIR (Data) Principles** are a set of guidelines designed to make digital research data (and research texts as well) FAIR by both humans and machines, that means: Findable (easily discoverable with metadata and persistent identifiers like DOIs), accessible (openly retrievable with clear access protocols), interoperable (formatted for integration across tools and disciplines), reusable (well-documented with licenses and provenance for replication). The FAIR Principles were formally defined in 2016 by a consortium of researchers, funders, and organizations in a seminal paper published in Scientific Data (Nature). See Wilkinson et al. 2016.

Publishing research data

13 Research data is published through both[16] research data repositories and subject-specific research database systems based on semantic data models, both often flanked by authority-data systems:

14 In research data repositories, research data can be stored, archived and made available (via viewer or as a download) in a low-threshold manner (commonly in files with a wide range of different formats) and under standardized

16 Reiter 2024, 2-3 for the distinction between data repository and (relational) database: ”It is important to distinguish between an archaeological data repository and an archaeological relational database, as the two are often confounded (see also Appendix 1 for a list of archaeological databases and repositories)”.

licenses, such as CC licenses (Fig. 4). [17] In addition, they can be secured for the long term in accordance with defined standards such as the OAIS standard (Open Archive Initiative System Reference Model[18]) and using suitable media, such as magnetic tapes (bitstream preservation). Permanent referencing of the entire research data provided is often ensured by using persistent identifiers, such as DOIs[19]. The level of granularity with which the data can be referenced varies depending on the repository and the requirements of the respective discipline. Data repositories therefore meet important requirements for implementing the FAIR principles for the reusability of research data[20].

15 According to Pampel and Elger (2021)[21] four types of repositories can be distinguished: institutional, discipline-specific (e.g. Pangea for Earth and Environmental Science)[22], multidisciplinary (e.g. Zenodo[23] and Figshare[24]) and project-specific repositories. Multidisciplinary repositories have the clear disadvantage that, due to the generic profile of the services, discipline-specific standards, metadata and context information can only be supported to a limited extent. Project-specific repositories can map the respective research data in a very fine-grained data model. However, this may limit the comparability and extensibility of the data (models), which must meet very specific requirements. In the field of archaeology and ancient studies, there are also discipline-specific data repositories, such as the Archaeology Data Service[25], which is operated by the University of York, the DANS Data Station Archaeology[26], the Journal of Open Archaeology Data Dataverse instance[27], tDAR - The Digital Archaeological Record[28] and the research data center IANUS[29], which was developed within a DFG project under the coordination of the DAI and is available to the community within the framework of NFDI4Objects.[30]

16 The usually less standardized subject specific database systems are distinguished from data repositories by the fact that the data modeling and the associated functionalities place greater emphasis on the semantics and their relations. Unlike research data repositories, where datasets are typically stored as files, these specialized database systems represent individual semantic entities (e.g. objects, contexts, places) in the first place.[31] Depending on the system, these entities can be structured in a fine-grained manner and presented in semantically enhanced, media-rich formats. The data can also be standardized and contextualized within a semantic network, using (domain) specific authority data

17 https://creativecommons.org/.
18 http://www.oais.info/: ISO Standard 14721:2012
19 https://en.wikipedia.org/wiki/Digital_object_identifier
20 The global registry for data repositories re3data offers a comprehensive overview of existing research data repositories: https://www.re3data.org.
21 Pampel 2021, 524-526.
22 https://pangaea.de/.
23 https://zenodo.org/.
24 https://figshare.com/.
25 https://www.archaeologydataservice.ac.uk/.
26 https://dans.knaw.nl/en/data-stations/archaeology/.
27 https://dataverse.harvard.edu/dataverse/JOAD.
28 https://core.tdar.org/.
29 https://ianus-fdz.de/.
30 https://www.nfdi4objects.net/.
31 For example: Pergamon as topographical entity at iDAI.objects: https://arachne.dainst.org/entity/1269; Pottery Shard 1003-5 at field.idai.world: https://field.idai.world/document/meninx-project/73aa71da-62ff-5a8d-1121-709d0cf96516; Bone CGM22418 (Animal Bone) at opencontext.org: https://opencontext.org/subjects/09B6F893-0E8D-4443-F1E6-499B355C88DA.

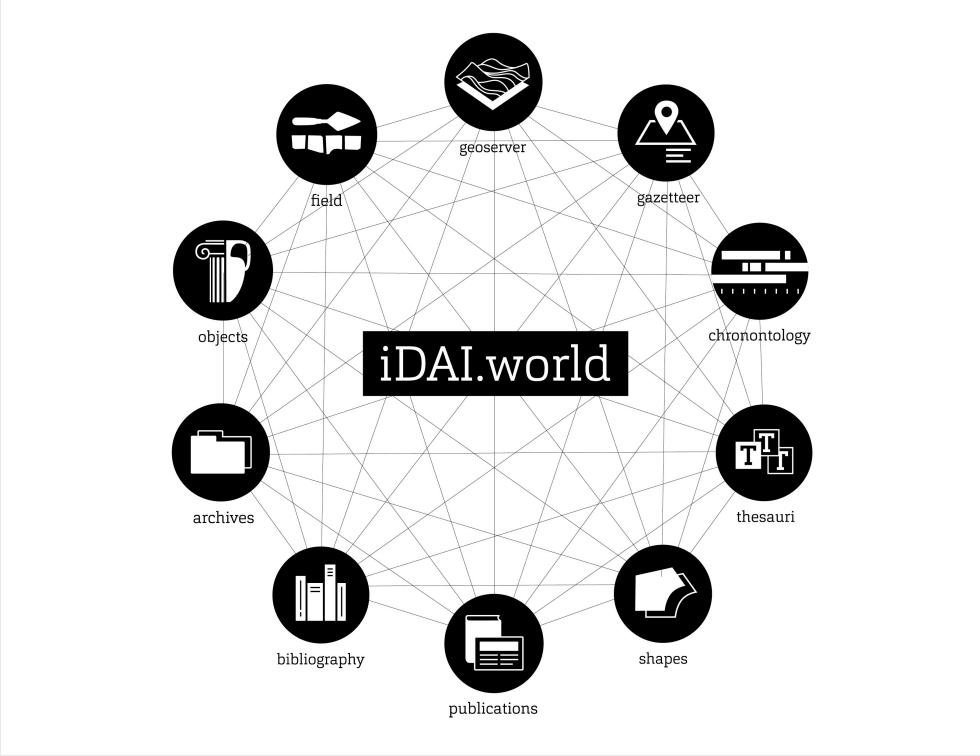


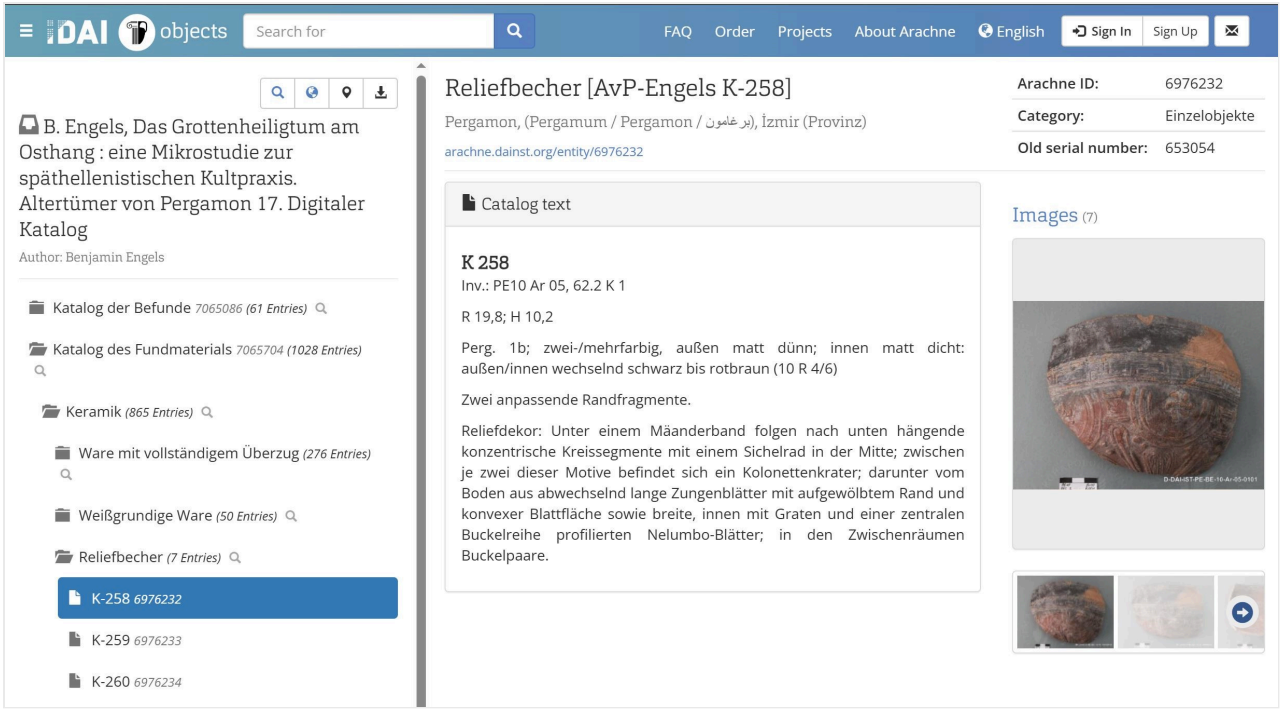
Fig. 5: Schema of the iDAI.world

5

and linking it to other data sets. Open Context[32] and iDAI.world[33] are examples of this kind of networked database infrastructure focussing on relations between semantic entities.

A **data model** is an abstract specification of entities, their attributes, and the relationships between them, designed to systematically represent and govern data within a defined domain. Entities (objects) are discrete, identifiable units (e.g. a place, a stratigraphic unit, an image, an artifact collection, a resource). Attributes are characteristics describing an entity (e.g. name, materials, measurements). Relationships are logical connections between entities (e.g. isPartOf, wasFoundAt). Semantic data models focus meaning, context and relationship rather than just structure and technical connections. They use real-world vocabularies (thesauri) that are often domain-specific (e.g. Getty AAT) and create ontology-based relationships (e.g. CIDOC-CRM) between entities enabling humans and machines to reason about data. For example: "Excavation → yielded → Artifact → createdBy → Culture“.

32 https://opencontext.org/.
33 https://idai.world/.



6

Fig. 6: View of a catalog entry in iDAI.objects (cf. Fig. 2).

Granularity refers to the degree of detail or scale at which information objects are represented. It can describe how much a larger entity is subdivided into smaller, more detailed parts (fine granularity), or how smaller, less distinct entities are grouped into larger, more distinguishable units (coarse granularity). For example: A single dataset, a single file or an article are more fine grained compared to an entire data collection, a file collection or a journal volume that typically contain multiple distinguishable parts of smaller scale.

The iDAI.world: Bridging research texts and research data

17 The DAI operates the iDAI.world (**Fig. 5**), a networked research data infrastructure comprising multiple repositories, databases, and services that provide extensive information resources online (as web resources).[34] In addition to a repository (iDAI.repo: a CKAN instance managed by the DAI)[35] and a publicly accessible long-term archive (IANUS)[36], many resources are accessible through specialized databases. These databases semantically contextualize information and interlink it using authority data systems for chronological information (iDAI.chronontology)[37], locations (iDAI.gazetteer)[38], and concepts (iDAI.world Thesaurus).[39] The research data includes information on individual objects,

34 https://idai.world/how/idai-world-architecture. For iDAI.world see also Senst 2022, 1-10.
35 https://repo.dainst.org/.
36 https://ianus-fdz.de/.
37 https://chronontology.dainst.org/.
38 https://gazetteer.dainst.org/.
39 http://thesauri.dainst.org/.

Fig. 7: Detail of an article displayed by the DAI-Journal-Viewer. See: F. Pirson, Pergamon – Die Arbeiten in der Kampagne 2021, Archäologischer Anzeiger, 2022, 2, 1–197 (§), doi:10.34780/gb88-g9bh.

digital object catalogs, excavation and survey documentation, geospatial data, archival material and natural science data which are published either as stand-alone data projects or in supplementary conjunction with scientific articles and books.[40]

18 The CiVers project primarily focuses on research databases like iDAI.objects[41] and iDAI.field[42], providing among other things (digital) object catalogs and excavation and survey documentations:

19 Object catalogs in iDAI.objects consist of structured object descriptions and image media (2 and 3-dimensional), accompanied by author-specific catalogue texts (**Fig. 6**). These catalogues can be organized into individualized works using a structured framework. While conceptually rooted in traditional object catalogues, they extend into a modern publication format that enables comprehensive, technologically advanced, and contextually enriched representations of

40 https://idai.world/what/publications. See „Data Enhanced Print Publications“.
41 The entire inventory of iDAI.objects/arachne includes almost 3 million image media and over 350,000 contextualized object data sets, see https://arachne.dainst.org/info/about.
42 https://field.idai.world/.

various objects. They can be accessed through dedicated entry pages, structured to align with related publications or the specific research questions of associated projects (via project pages).[43]

20 iDAI.field is designed for documenting archaeological field projects, including excavations, building and landscape surveys. The collected data can be published as digital excavation documentation via iDAI.field-web.[44] The content is structured as resources, encompassing measures (e.g., sections), archaeological categories (e.g., stratigraphic units, finds), subcategories, and image media. These resources can be interconnected through relational links, arranged in hierarchical structures, provided with geometries and visualized. The data model is flexible and can be customized for specific projects based on global default configurations.

21 These object catalogs and field documentations highlight two key aspects of scientific publishing: the conception of scholarly works and the role of the publication medium (as representational forms). Since such datasets can no longer be fully represented in print or PDF formats[45] they should be recognized as independent media forms that surpass traditional text-based publications in structural and technological quality—offering advantages such as semantic networking, contextualization, accessibility, and reusability. This applies not only to data collected within these systems (e.g., during excavations) but also to datasets migrated from text-based formats and local databases into specialized data models. As a result, iDAI.world has evolved from a supporting resource platform of digitized information resources into a central publication platform for object catalogues, excavation documentation, and other domain-specific content. Moving forward, the DAI will continue along this path, further integrating iDAI.world as a publication medium for both internal and external research projects.

22 As another noteworthy example for bridging the gap between research texts and research data we like to close this section with the DAI-Journal-Viewer[46]: The DAI-Journal-Viewer – a customized version of the eLife Lens Viewer – allows displaying excerpts of referenced data entities parallel to the article text.[47] Supported datasets include object and image records from iDAI.objects, resources from iDAI.field, and location data from iDAI.gazetteer (Fig. 7). The article content, including the entity-references, is structured according to the JATS-XML standard, which allows machine-readability and continuous re-use, allowing for future software which can also to represent the article

43 Examples of recently published digital object catalogs (among others): Simitthus 6: https://arachne.dainst.org/project/beck_simitthus_6; AVP 17: https://arachne.dainst.org/project/avp17_engels; MKT 20: https://arachne.dainst.org/project/mkt20_boecker and FAAK 18: https://arachne.dainst.org/project/faak18.

44 Examples of recently published projects via iDAI.field.web: Taxiarchis: https://field.idai.world/project/didyma-project and Henchir Bourgou: https://field.idai.world/project/bourgou-online.

45 For example, limitations to a small selection of images or the depiction of color images as black-and-white photos due to growing printing costs.

46 See Baumeister 2022

47 This pioneering approach of the Archäologischer Anzeiger 2020 was awarded with the Digital Publishing Award in the category “Process/Technology” at the Frankfurter Buchmesse (Frankfurt Book Fair).

content in feature-rich web formats.[48] But what does this kind of entity referencing mean in terms of citing and referencing of external resources? Is it a special type of referencing?

Machine-readability refers to the ability of data or text to be automatically processed by a computer without direct human intervention. Machine-readable formats follow structured, standardized rules and formats (such as XML, JSON, or CSV), allowing software to efficiently access, process and interpret the information. Adding semantic depth by means of structured metadata, controlled vocabularies, ontologies, semantic markups, identifiers and linked data is key to enable technical services (= machines) to integrate, recompose and connect information resources.

Web resource is an abstract term referring to any type of content, tool, or service available on the internet that can be identified, named, addressed or handled. It can be a single item (e. g. a web page, an image, a record of a person) or collection of various items (in a networked information system) often encompassing other digital assets like datasets, images, software tools and services. In the context of the Semantic Web the term is commonly used in the sense of identifying, locating and describing resources via web addresses (URLs or URI) using structured data (RDF, OWL). In that sense it is comparable to standard library services, aiming to organize, describe, and provide access to information in reliable, structured and meaningful ways.

Citing (and referencing) information resources

23 Citing and referencing are essential components of scientific writing and publishing. They fulfill fundamental purposes and ethical standards in scientific research, serving the needs of data producers and users. Citation is "the process of indicating what external sources have been used to create content".[49] By acknowledging the sources of research and the creators of the data used, citing and referencing ensure accessibility, transparency, reproducibility, and proper attribution of research outputs. Typically, citations appear in the form of footnotes or in-text citations. For every citation, a corresponding reference is

48 To name just a few advantages over paged-views or PDF formats: better usability trough more flexibility in layouting content (responsive designs), barrier-free accessibility, automatic translations based on browser services etc.
49 EU Publications Office 2022, 7.

```
<relatedIdentifiers>
        <relatedIdentifier relatedIdentifierType="ISSN" relationType="IsPartOf">2748-8861</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="IsPartOf">10.34780/chzkvr64</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.4103/1008-682X.171582</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.34780/CF6E-PRCF</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.48440/allianzoa.041</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1016/j.patter.2020.100180</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.5438/0001</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.25490/A97F-EGYK</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.14454/MZV1-5B55</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110769043-023</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.5438/K81T-ZQ43</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.2830/59387</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.3346/jkms.2018.33.e139</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110769043-052</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110657807-028</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110494068-038</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/opar-2024-0015</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1108/JD-10-2022-0234</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.3390/digital2030019</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.5282/o-bib/5844</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110494068-028</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.5438/jwvf-8a66</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110448115-001</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.1515/9783110769043-060</relatedIdentifier>
        <relatedIdentifier relatedIdentifierType="DOI" relationType="References">10.57674/fyrc-vb61</relatedIdentifier>
</relatedIdentifiers>
```

8

Fig. 8: Modelling of <relatedIdentifiers> according to the DataCite Metadata Schema.

provided with all relevant information, such as author, year, title of the cited source, place of publication, and publisher. References are usually – but not necessarily – detailed in a reference list.[50]

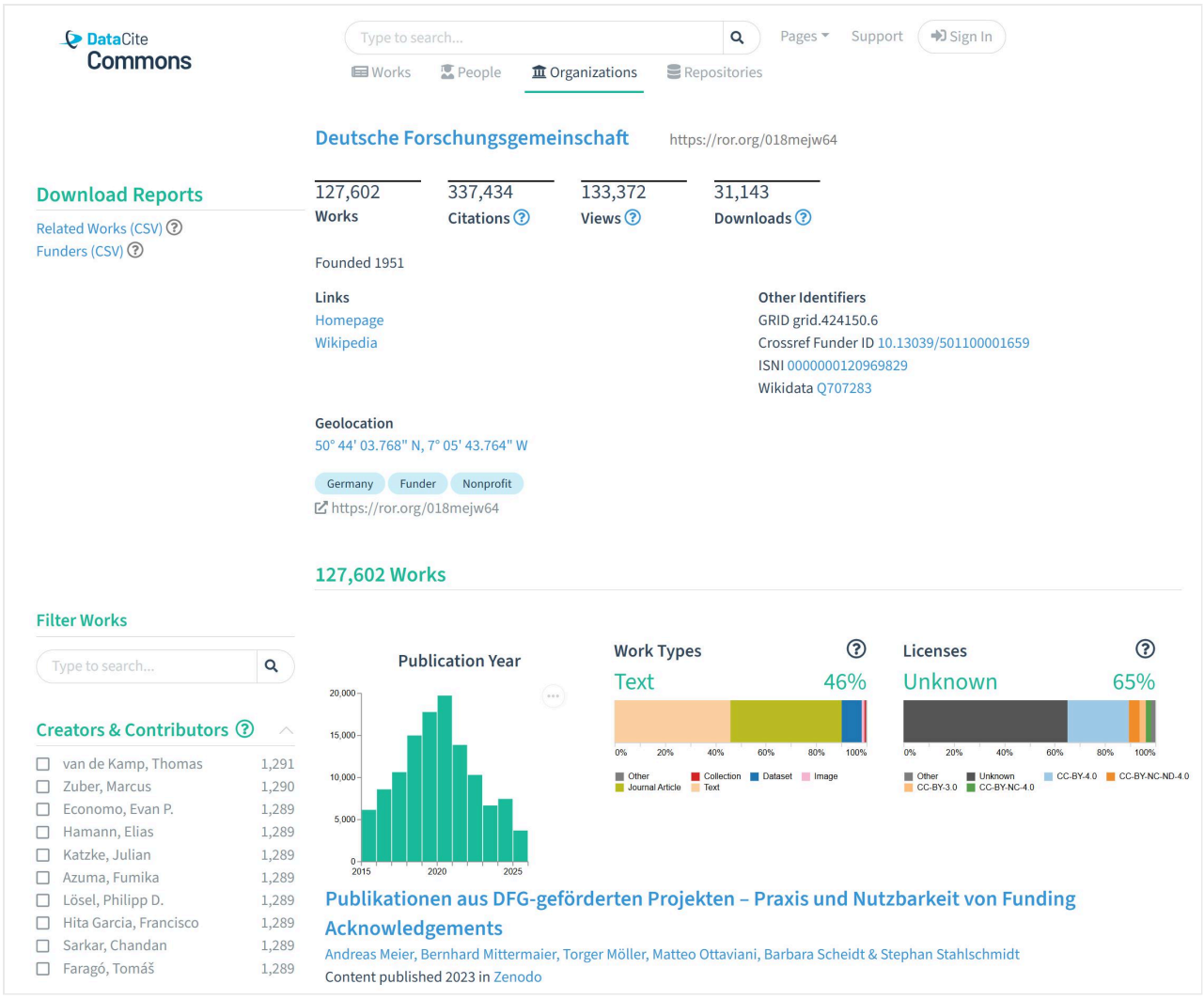
Machine-Readability and Knowledge Graphs

24 While references in research papers and research (meta)data can appear in various places (e.g. in-text citations or reference lists) other types of references are used to define terms, provide additional information, or simply link to related resources. For example, the XML articles displayed by the DAI Journal Viewer are enhanced with references to external objects in order to enrich terms used within the research text, e.g. naming archaeological objects or places.

25 In any case, there is often reference information behind the scenes: machine-readable metadata that are often standardized and extendable allowing to describe semantic relations between knowledge resources of various kinds. Common international standards for describing (electronic) resources are the Dublin Core Metadata Terms (DCMT) and the DataCite Metadata Schema, which is DublinCore compliant.[51] The DataCite Metadata Schema is used to describe research (data) resources, registered with DataCite. These standards offer a wide set of terms (properties) and recommendations to describe meta information related to a single resource (e.g. identifier, creator, contributor, title, language, date, format) but also its relations to other resources (e.g. relatedIdentiers, isVersionOf, isPartOf) or entities (e.g. nameIdentifier, funderIdentifier,

50 Not to be confused with a bibliography, which offers references for further reading or background information. A bibliography can and does include sources not cited in the text, hence not directly used for the research presented in a given paper. For a technical implementation of certain terms and concepts see Data Cite: https://support.datacite.org/docs/contributing-citations-and-references.

51 See DataCite Metadata Working Group 2011. For the current schema see DataCite Metadata Working Group 2024.



9

Fig. 9: DataCite Commons

rightsIdentifer).[52] To take this article as an example (Fig. 8): Following the DataCite metadata schema each reference listed in the reference-list below are described as relatedIdentifier with relationType = "references". As part of a journal (issue), it „isPartOf" this certain issue, and further versions of the article text would be described as „isVersionOf".

26 When it comes to citing and referencing information, these relation types are key elements (among other metadata terms) for improving discoverability and constant enhancement of (domain-specific) knowledge graphs. To name a few advantages: Researchers can enhance the discoverability of their own works (including associated dataset), libraries and information service providers can make their cataloging and resource management more efficient, research institutions can accurately track dynamically and evaluate research output and its dependencies from other resources based on (citation) event data.

27 Citation (event) data influence the circulation of research outputs and data, citation and referencing affect the visibility of researchers in terms of their contribution to their fields. This can be a deciding (social) factor for early career

52 El-Gebali 2024.

researchers when seeking to secure jobs and grants. Various metrics (publication count, citation count, H-index, article-level metrics, altmetrics) are currently in use to assess the impact of scholarly publications.[53] However, these metrics differ greatly in how they are calculated, and there is currently no single comprehensive metric for assessing the value of researchers´ contributions.[54] For instance, while some metrics consider the number of articles published in peer-reviewed, high-impact journals (productivity), others emphasize instead the frequency of citations of a work (impact), and some take into account both productivity and impact factors. Moreover, the impact and ranking of the journals in which a work is published can also influence these metrics, which, in turn, might misrepresent the contributions of individual researchers in collaborative works. Beside these widely criticized metrics, maintained by proprietary service providers, non-profit oriented services (e.g. DataCite, Crossref, Open Alex[55]) and initiatives (like I4OC[56], Make Data Count[57]) are pushing toward open, responsible metrics (Fig. 9). [58]

28 Assuming a continuous growth of research resources, especially data and entity objects (in multiple versions and in a fine-grained manner), there is a growing need to extend machine-readability of scientific resources. The re-use potential, especially for semantic purposes, increases with fine-grained meta information that need fine-grained entities in the first place: When resource-type specific research data infrastructures, such as image databases, offer citation modalities based on persistent identifiers, allowing researchers to reference each image as an individual resource in their texts, researchers take the first step toward building semantic relations between the research texts and each cited image resource. If publishing providers transform these images references into a machine-readable form (Research Text A references Image A) into an open metadata repository, all service providers can extract and display these relations, e.g. for enhancing the reference list of the image resource (referencedBy: Research Text A).

Knowledge Graphs (general or domain-specific) are structured semantic representation of entities such as sites, artifacts, research texts, and researchers, along with their relationships, enabling integrated analysis, linking of heterogeneous data sources, and enhanced discovery across historical and archaeological research resources. Knowledge graphing is the process of modeling, querying, or visualizing resources and their relationships in a graph structure. It consists of dots (called nodes or vertices) that represent things (like people, places, or objects), and lines (called edges) that show the relationships or connections between them.

53 Agarwal 2016.
54 Gasparyan 2018.
55 https://openalex.org/.
56 https://i4oc.org/.
57 https://datacite.org/make-data-count/.
58 Dorsch 2022.

(Citation) **event data** refers to detailed records of when, how, and where one scholarly work cites (or references) another. Each "event" captures a specific act of citation - essentially a timestamped connection between two pieces of academic content. Several major services manage citation event data, especially for scholarly publications, datasets, and other research outputs. The most widely used, reputable and non-profit ones are: Crossref, DataCite and OpenCitations. Beside these there are proprietary services like Scopus (restricted access), Web of Science (restricted access) or Dimensions.ai (partly restricted).

Challenges for infrastructure providers and researchers

29 All this comes along with technical and content related requirements on both sides (publishing and referencing), possibly causing additional workloads that are not easy to handle by research database providers, by researchers or by publishing service providers. The ongoing integration of external information objects (like images and other data resources) into research products made for publishing increases the publishing-specific requirements these data objects have to fulfill better than before: for example, credit attribution in terms of contributorship are often not fully implemented, technically or with regard to the metadata descriptions of individual resources (datasets). Many research databases are originally made for digitizing or for sharing, not for publishing, and the amount of items that – thankfully – have been shared by digitizing projects are immense. However, referencing resources in a standardized form in terms of metadata standards is possible only when the resource itself, which may exist in multiple versions, as well as its creators and contributors (of various kinds) are correctly identified and given credit, ideally by also acknowledging the specific roles of their contributions.[59]

30 Citing and referencing starts with the creation of research texts and research data. In the case of small scale data resources (such as images) it is necessary to collect and manage all these citations for all these resources, often coming from various data providers offering different concepts, policies or citation formats (e.g. APA, Harvard). Yet, as each discipline has its own citation style, citing and referencing across disciplines can be challenging, especially in a multidisciplinary field like archaeology.[60] Additionally, when it comes to individual online resources like web pages or websites there is still often no formal advice on how to cite individual resources at all (or at least in the manner of regular citation guidelines, expecting author, title and date information).

59 To standardize the acknowledgment of different authors´ contributions, the NISO Credit Guidelines produced a Credit Taxonomy for research papers in 2015: https://credit.niso.org/. This taxonomy employs a simple, controlled vocabulary to ensure that contributors are recognized according to the type (e.g. Conceptualization, Writing, Data Curation) and extent of their contribution(s).

60 For example, the DAI provides guidelines for referencing and using abbreviations in archaeological publications based on the author-date-system, see: https://www.dainst.org/en/research/publications/publishing-at-the-dai/dai-citation-style-and-abbreviations. The Romano-Germanic Commission, the Commission for Ancient History and Epigraphy, and the Eurasia Department use slightly modified versions for their publication, each following international - but specific - standards within (!) the wide field of archaeology.

31 Fortunately, common citation guidelines and policies have already been addressing the peculiarities of citing research data or web-resources.[61] Persistent identifiers (like DOIs) play an important role here, forcing infrastructure providers to manage and to describe their resources according to the manner of regular citation guidelines. But many web-resources (including subject specific web-databases, blogs, news-articles) do not offer persistent identifiers based or version-controlled content at all, allowing a persistent referencing in terms of requesting resources and reading the content that is likely to changed (especially its semantics).

32 There are many technical solutions to handle the growing amount of standards and technical requirements of scholarly referencing: For example, the Citation Styles[62] repository provides an open-source, XML-based language to standardize the formatting of citations and bibliographies in scholarly publications, aiming to automate citation formatting. Reference management and writing tools (like Mendeley, Papers, Zotero, etc.)[63] have adopted the Citation Styles to provide consistent and accurate formatting of citations in order to streamline the writing and publishing process. The repository stems from the close collaboration between authors, publishers, and developers in finding, implementing, and maintaining citation styles, and integrating them into relevant software.[64] Nonetheless, errors in citing and referencing persist, thus stressing the need for standardization, cooperation with publishers, and better training in formatting citations.[65]

33 Finally, publishing service providers are confronted with all these requirements in the end when proof-reading and layouting all citations and reference lists. The editorial curation of references of any kind is a time-consuming process, with or without taking machine-readability into account. Considering the growth of multiple small-scale data resources (like images), the ongoing change in publication designs requires editorial strategies and more intensive usage of technical tools and services to facilitate the processing of these requirements in the interest of the field of archaeology and ancient studies.

34 Within the CiVers project it is planned to evaluate standards, tools and services to facilitate the curation and transformation of references, with special focus on modelling reference-lists in human- and machine-readable forms (e.g. according to Dublin Core or DataCite Metadata Schemas). Some requirements described can be easily fullfilled by adapting existing (standard) solutions provided by various initiatives and non-profit services: Open-source publishing systems like Open Journal System (OJS)[66] or Open Monograph Press (OMP)[67] – used by many small or medium-size publishing providers to manage and publish

61 For example, APA: https://apastyle.apa.org/style-grammar-guidelines/references/examples/webpage-website-references; MLA: https://owl.purdue.edu/owl/research_and_citation/mla_style/mla_formatting_and_style_guide/mla_works_cited_electronic_sources.html; Vancouver: https://guides.library.uq.edu.au/referencing/vancouver/webpages; DAI: https://www.dainst.org/en/research/publications/publishing-at-the-dai/dai-citation-style-and-abbreviations. See also EU Publications Office 2022, 33-48 (for citing data resources).
62 https://citationstyles.org/.
63 https://www.zotero.org/, https://www.citavi.com/, https://www.mendeley.com/.
64 As of now, the crowdsourced repository includes more than 10.558 (!) free citation styles (CSL) with a Creative Commons Attribution-ShareAlike (BY-SA) license available on GitHub.
65 Santos 2023.
66 https://pkp.sfu.ca/software/ojs/.
67 https://pkp.sfu.ca/omp/.

articles and books – are by default able to store reference lists and provide plugins for registering DOIs (including machine readable-metadata) at DOI registations agencies such as DataCite. These plugins can be extended and improved by contributions, e.g. to map references as related identifiers or to integrate references of all kind into external knowledge resources. For this specific requirement we want to highlight the KOMET project (started as OPTIMETA) that has been developing OJS-plugins for integrating open citations and spatio-temporal metadata into openly accessible data sources such as Wikidata.[68] Combining all this together can help to decrease workloads and to increase the discoverbility of scientific information resources at all.

68 https://projects.tib.eu/komet/en/.

References

Agarwal 2016 A. Agarwal et al., Bibliometrics: tracking research impact by selecting the appropriate metrics, Asian Journal of Andrology 18, 2, 2016, 296, https://doi.org/10.4103/1008-682X.171582

Baumeister 2022 P. Baumeister, Das neue maschinenlesbare Zeitschriftenmodell des Deutschen Archäologischen Instituts. Ein Werkstattbericht, Forum for Digital Archaeology and Infrastructure, 2022, §1-67, https://doi.org/10.34780/CF6E-PRCF

Breuer 2021 C. Breuer – P. Trilcke, Die Ausweitung der Wissenschaftspraxis des Publizierens unter den Bedingungen des digitalen Wandels, 2021, https://doi.org/10.48440/allianzoa.041

Cousijn 2021 H. Cousijn – R. Braukmann – M. Fenner – C. Ferguson – R. van Horik – R. Lammey – A. Meadows – S. Lambert, Connected Research: The Potential of the PID Graph, Patterns 2, 1, 2021, 100180, https://doi.org/10.1016/j.patter.2020.100180

DataCite Metadata Working Group 2011 DataCite Metadata Working Group, DataCite Metadata Schema for the Publication and Citation of Research Data v2.0: Documentation, 2011, https://doi.org/10.5438/0001

Data Citation Synthesis Group 2014 Data Citation Synthesis Group, Joint Declaration of Data Citation Principles, https://doi.org/10.25490/A97F-EGYK

DataCite Metadata Working Group 2024 DataCite Metadata Working Group, DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.6, 2024, https://doi.org/10.14454/MZV1-5B55

Dorsch 2022 I. Dorsch – S. Haustein, B 11 Bibliometrie, in: R. Kuhlen – D. Lewandowski – W. Semar – C. Womser-Hacker (Hrsg.), Grundlagen der Informationswissenschaft (2022) 271–280, https://doi.org/10.1515/9783110769043-023

El-Gebali 2024 S. El-Gebali, Connecting the Dots with DataCite DOI Metadata, 2024, https://doi.org/10.5438/K81T-ZQ43

EU Publications Office 2022 Publications Office of the European Union and P.Jessop, Data citation – A guide to best practice, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2830/59387

Gasparyan 2018 A. Y. Gasparyan et al., Researcher and Author Impact Metrics: Variety, Value, and Context, Journal of Korean Medical Science, 2018, Apr;33(18):e139, https://doi.org/10.3346/jkms.2018.33.e139

Kuhlen 2023 R. Kuhlen, "E 1 Informationsmarkt", in: R. Kuhlen et al. (Hrsg.), Grundlagen der Informationswissenschaft (Berlin 2023), 603-620, https://doi.org/10.1515/9783110769043-052

Pampel 2021 H. Pampel – K. Elger, 5.6 Publikation und Zitierung von digitalen Forschungsdaten, in: M. Putnings – H. Neuroth – J. Neumann (Hrsg.), Praxishandbuch Forschungsdatenmanagement (2021) 521–536, https://doi.org/10.1515/9783110657807-028

Pfeiffenberger 2017 H. Pfeiffenberger, "8. Data Publishing und Open Access", in: K. Söllner und B. Mittermeier (Hrsg.), Praxishandbuch Open Access, 2017, 333-340, https://doi.org/10.1515/9783110494068-038

Reiter 2024 S. Reiter et al., The BIAD Standards: Recommendations for Archaeological Data Publication and Insights From the Big Interdisciplinary Archaeological Database, Open Archaeology 2024, 10(1), https://doi.org/10.1515/opar-2024-0015

Santos 2023 E. A. D. Santos – S. Peroni – M. L. Mucheroni, An analysis of citing and referencing habits across all scholarly disciplines: approaches and trends in bibliographic referencing and citing practices, JD 79, 7, 2023, 196–224, https://doi.org/10.1108/JD-10-2022-0234

Schmidt 2022 S.C. Schmidt, F. Thiery, M. Trognitz, Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata. Digital 2022, 2, 333-364. https://doi.org/10.3390/digital2030019

Senst 2022 H. Senst – F. Riebschläger – J. Watson, „iDAI.world“: Die vernetzte Forschungsdatenplattform des Deutschen Archäologischen Instituts, o-bib. Das offene Bibliotheksjournal / Herausgeber VDB 9, 4, 2022, 1–10, https://doi.org/10.5282/o-bib/5844

Söllner 2017 K. Söllner, 5a. Geisteswissenschaften in: K. Söllner und B. Mittermeier (Hrsg.), Praxishandbuch Open Access (Berlin 2017), 247-253, https://doi.org/10.1515/9783110494068-028

Stocker 2021 M. Stocker – M. Fenner, Introducing the PID Graph, 2021, https://doi.org/10.5438/jwvf-8a66

Taubert 2016 N. Taubert and P. Weingart, "Wandel des wissenschaftlichen Publizierens – eine Heuristik zur Analyse rezenter Wandlungsprozesse". Wissenschaftliches Publizieren: Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung, edited by Peter Weingart and Niels Taubert, Berlin, 2016, pp. 3-38. https://doi.org/10.1515/9783110448115-001

Tochtermann 2023 K. Tochtermann - A. M. Höfler, "E 9 Open Science", in: R. Kuhlen et al. (Hrsg.), Grundlagen der Informationswissenschaft (Berlin 2023), 703-714, https://doi.org/10.1515/9783110769043-060

Wissenschaftsrat 2022 Wissenschaftsrat, Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access, 2002, Köln. https://doi.org/10.57674/fyrc-vb61

SOURCE OF ILLUSTRATIONS

Fig. 1: Design: Marcel Riedel
Fig. 2: Screenshot: Marcel Riedel
Fig. 3: Screenshot: Marcel Riedel
Fig. 4: Screenshot: Marcel Riedel
Fig. 5: Design: Tanja Lemke-Mahdavi
Fig. 6: Screenshot: Marcel Riedel
Fig. 7: Screenshot: Marcel Riedel
Fig. 8: Screenshot: Marcel Riedel
Fig. 9: Screenshot: Marcel Riedel

METADATA

https://doi.org/10.34780/6k764r03

AUTHORS

Marcel Riedel
German Archaeological Institute, Head Office

Fabian Riebschläger
German Archaeological Institute, Head Office

Giulia Russo
German Archaeological Institute, Head Office

IMPRINT

Forum for Digital Archaeology and Infrastructure
https://doi.org/10.34780/8FE9-PX79

Deutsches Archäologisches Institut
Podbielskiallee 69–71
14195 Berlin
Deutschland
www.dainst.org

Editing and Typesetting

Publishing editor: FdAI, DAI Zentrale, Arbeitsstab Kommunikation

Contact for article submissions: fdai-journal@dainst.de

Editing: Elianor Sket

Corporate Design: LMK Büro für Kommunikationsdesign, Berlin