

The International Journal of Digital Curation

Volume 8, Issue 1 | 2013

Linking Data and Publications: Towards a Cross-Disciplinary Approach

Maarten Hoogerwerf,
Data Archiving and Networked Services, NL

Mathias Lösch and Jochen Schirrwagen,
Bielefeld University, DE

Sarah Callaghan,
STFC Rutherford Appleton Laboratory, UK

Paolo Manghi,
National Research Council, IT

Katerina Iatropoulou and Dimitra Keramida,
National and Kapodistrian University of Athens, GR

Najla Rettberg,
University of Göttingen, DE

Abstract

In this paper, we tackle the challenge of linking scholarly information in multi-disciplinary research infrastructures. There is a trend towards linking publications with research data and other information, but, as it is still emerging, this is handled differently by various initiatives and disciplines. For OpenAIRE, a European cross-disciplinary publication infrastructure, this poses the challenge of supporting these heterogeneous practices. Hence, OpenAIRE wants to contribute to the development of a common approach for discipline-independent linking practices between publications, data, project information and researchers. To this end, we constructed two demonstrators to identify commonalities and differences. The results show the importance of stable and unique identifiers, and support a “by reference” approach of interlinking research results. This approach allows discipline-specific research information to be managed independently in distributed systems and avoids redundant maintenance. Furthermore, it allows these disciplinary systems to manage the specialized structures of their contents themselves.





Introduction

The rise of data-driven science has led to a change in the process of scholarly communication. To date, traditional publishing has mainly focused on textual publications. However, the publication and citation of research data is gaining a similar importance in the scientific dissemination process. In particular, the publishing of research data beside publications, and linking these in an open and meaningful way, is becoming a key requirement to increase the quality and transparency of scholarly communication (Kunze et al., [2011](#)).

OpenAIRE¹ is a European cross-disciplinary publication infrastructure aiming to support researchers, data providers and funders by providing transparent access to research output and information on funding, licenses and usage statistics. Its most visible result is the OpenAIRE portal, where these contents can be browsed by users. The successor project OpenAIREplus aims to expand the OpenAIRE infrastructure, including its portal, by providing support for the linking of publications, research data and other contextual information. This ambition presents OpenAIREplus with a challenge, as the linking of research contents is still an emerging topic that is approached differently by different communities.

On the one hand, the DRIVER II² project suggested the concept of “enhanced publications”, providing an overview of a theoretical model and formulating a broad definition:

“An Enhanced Publication is a publication that is enhanced with research data as evidence of the research, extra materials to illustrate or to clarify, or post-publication data like commentaries and ranking” (Woutersen-Windhower et al., [2009](#)).

The project results have been continued by various projects, leading to different instances of enhanced publications such as the European Value Studies (EVS), the Journal of Archaeology in the Low Countries (JALC) and an oral history book on interviewed veterans. EVS allows their publications to be linked to the underlying concepts and variables from questionnaires in DDI3 format³ (DDI3 is an XML specification for social science metadata that is designed to support the full data life cycle⁴). JALC allows online publications to embed sortable/filterable data tables, geographic data on maps and image collections (Smits et al., [2009](#)). The book on oral history allowed its authors to define fragments in a collection of recorded veteran interviews and cite these as streaming fragments within the different publications within the book (Berg et al., [2010](#)). More generic instances were created using the Escape tool, which allows researchers to define random aggregations of resources such as authors, publications, datasets and events.⁵ The implementations of this model


¹ OpenAIRE: <http://www.openaire.eu>

² DRIVER II: <http://www.driver-community.eu/>

³ DataPlus Enhanced Publication Editor: http://www.centerdata.nl/en/TopMenu/Wat_doen_wer/ICT-toepassingen/dataplusepe.html

⁴ For an introduction to DDI3, see: <http://www.ddialliance.org/DDI/ddi3/index.html>

⁵ For example, see Zijdemans (2009).



delivered OAI-ORE resource maps (Lagoze et al., 2008), which are aggregations of references and descriptions of all constitutive entities. We call this approach the “package” model of linking research results.

On the other hand, there is the concept of connecting research results contained in different infrastructures by using hyperlinks. For instance, in the area of the Life Sciences, the European Bioinformatics Institute (EBI) provides freely available data and services, such as nucleotide sequences, gene expression, protein information, chemicals and biological pathways. More recently, the EBI has led the development of Europe PubMed Central (Europe PMC)⁶, a literature database containing abstracts and full text articles from the Life Sciences. The core content is enriched through the addition of citation information (i.e. who is citing who); text mining; allowing the user to highlight and browse keywords, such as gene names, organisms, and diseases; and links to respective records in biological databases (McEntyre et al., 2011).

On a broader level, the issue of data citation has recently been discussed extensively (cf. Lawrence et al., 2011, Ball & Duke, 2012). Data citation differs from merely linking to datasets in that it provides a formalised method for the user to locate and discover information about the data. For instance, this includes the use of stable identification systems like Digital Object Identifiers (DOIs) for datasets. Moreover, data citation credits the data producers for their efforts in creating the dataset, and the data publisher for managing and archiving the data. Several discipline-specific and cross-disciplinary initiatives exist to encourage and standardise the process of data citation. A discipline-specific example is PANGAEA⁷, a data deposition and citation service for the Earth & Environmental Sciences. On the cross-disciplinary side, DataCite (cf. Brase, 2009) aims to establish easier access to research data on the Internet, increase acceptance of research data as legitimate, citable contributions to the scholarly record and support data archiving that will permit results to be verified and re-purposed for future study. We call the data linking/citation method the “by reference” model of connecting data and publications.

These different approaches of connecting research results make it difficult for third parties (like the OpenAIRE infrastructure) to provide sustainable, automated services that interpret, manage and exploit the added value of such related research output. Moreover, OpenAIRE itself aims to further enrich the network of related assets by adding information on projects, funding and usage statistics. This paper presents our contribution towards the creation of a common approach for linking related research results. We explore how different types of existing, interlinked outputs can be managed by cross-disciplinary infrastructures. To this end, we built two demonstrators showing examples of interlinked research results from the Life Sciences, the Social Sciences and the Humanities. In the process of building these demonstrators we explored commonalities, differences and other issues that can contribute to a general model for linking publications and datasets. The results shall support further discussion on a general model.

In general, the publication is put *into context*. This context helps researchers assess a publication and discover related resources. The examples create context on different levels: some embed (e.g. by embedding research data into the actual publication),

⁶ Europe PMC: <http://europepmc.org>

⁷ PANGAEA: <http://www.pangaea.de/>

while others link to it via the metadata. Since the OpenAIRE portal is primarily based on metadata, the work described in this paper focuses on the linking of publications and data on the metadata level.

Exploring Data to Publication Linking in Different Disciplines

This section describes two demonstrators that were built by OpenAIREplus. The demonstrators show examples of interlinking data and publications within the Life Sciences, the Social Sciences and the Humanities. The goal of constructing these demonstrators is to forward the development of a common, discipline-independent model for linking publications, data and other contextual information. It does so by identifying issues, commonalities and differences during the iterative construction of the demonstrators, and by providing concrete examples to support further discussion.

First Demonstrator: Life Sciences

In the case of the Life Sciences demonstrator⁸, the focus is on the problem of how an aggregation infrastructure (such as OpenAIRE) can re-use ‘added value’ elements produced by Europe PMC, which actively links publications and biological research data. More precisely, a publication that has been previously enriched by Europe PMC services should remain connected to its related objects and be further enhanced within the demonstrator, e.g. get attributed to relative project information. While this requirement seems straightforward at first, the challenges lie in the lack of standardized exchange formats that encode bibliographic metadata, whilst at the same time endeavouring to preserve a network of related objects as created by Europe PMC.

The demonstrator is a simple web application for displaying, browsing and searching publications. Its core entities are publications, authors, datasets and projects from the Seventh Framework Programme (FP7)⁹, all of which are represented as HTML splash pages identified by stable URIs in the application’s front end.

To identify publications, we use PubMed IDs – the well-established universal identifier in the Life Sciences. For the sake of simplicity, we ignore most bibliographic details except for the publication title, the authors and potential external identifiers (such as DOIs). Thus, a publication is displayed under its title, together with an author list and a tabbed view of context information from Europe PMC (see Figure 1).

⁸Life Sciences Demonstrator: <http://ub.unibi.de/oademo>

⁹ Running from 2007 to 2013, the Seventh Framework Programme (FP7) currently is the European Union's main instrument for funding research in Europe: http://ec.europa.eu/research/fp7/pdf/fp7-factsheets_en.pdf

OpenAIRE Enhanced Publications Demonstrator – Life Sciences Edition

Home Publications Authors Datasets Projects Search publications jane

Publications / Diversity of a complex centromeric satellite and molecular characterization of dispersed sequence families in sugar beet (Beta vulgaris)

Diversity of a complex centromeric satellite and molecular characterization of dispersed sequence families in sugar beet (Beta vulgaris)

Gerhard Menzel, Daryna Dechyeva, Torsten Wenke, et al.

Enhance this Publication!

Abstract References (67) Cited by (7) Database Links Text-Mined Terms Metrics Versions Projects MeSH

Database links are curated by experts at the [European Bioinformatics Institute \(EBI\)](#)

- EMBL
 - Beta vulgaris dispersed repeat pDvu11-2
 - Beta vulgaris dispersed repeat pDvu12-1
 - Beta vulgaris dispersed repeat, clone pDvu11-1
 - Beta vulgaris dispersed repeat, clone pDvu11-3
 - Beta vulgaris dispersed repeat, clone pDvu11-4
 - Beta vulgaris dispersed repeat, clone pDvu11-5
 - Beta vulgaris dispersed repeat, clone pDvu12-2
 - Beta vulgaris subsp. vulgaris tandem repeat pRv1-1, clone 46C21
 - Beta vulgaris subsp. vulgaris tandem repeat pRv1-2, clone 65O23
 - Beta vulgaris subsp. vulgaris tandem repeat pRv1-3, clone 59E3

Figure 1. A screenshot of a publication splash page in the Life Science Demonstrator.


The application is populated with roughly 120 sample publications. These were imported from Bielefeld University’s repository “PUB”¹⁰ which uses PubMed IDs in its metadata model. More precisely, we selected the publications by querying PUB’s OAI-PMH interface for metadata from the last two years and by filtering for publications with assigned PubMed IDs. Next, we connected all FP7-funded publications in the demonstrator to the respective EC-funded project. Links to projects are provided by the index of the OpenAIRE infrastructure. If a publication could be attributed to a project, we also imported some details of the project, such as the acronym, duration and subject area. In the demonstrator, projects are identified by their FP7 grant numbers.

While all of the afore-mentioned steps were carried out during the development and deployment of the application, the mechanism for importing linked information from Europe PMC was designed to be performed by actual users of the system. To be more precise, a user can trigger queries to the CiteXplore¹¹ web service – an interface to Europe PMC’s data corpus that can be queried for linked information for a publication. For a given publication, the following types of information can be imported from CiteXplore (if available in Europe PMC):

1. The publication’s reference list;
2. A list of publications that cite this publication (from Europe PMC’s citation database);
3. References to biological databases that have been manually attributed to the publication by EBI’s data curators;

¹⁰ PUB: <http://pub.uni-bielefeld.de/>

¹¹ CiteXplore: <http://europepmc.org/WebServices>

- 
4. Text-mined biological terms that link to different resources at EBI/Europe PMC;
 5. Links to different versions of the full text with information on accessibility;
 6. Medical subject headings (MeSH)¹² for this publication.

If any of this information is available, it is represented in the demonstrator, connected to the respective publication, and displayed on the public splash page visible to all users. Most importantly, all of the imported objects are represented by external hyperlinks where the target depends on the information type. References and citations are again identified by PubMed IDs and link to the respective abstracts in Europe PMC. Database links point to entries in biological databases, identified by unique accession numbers. The nature of the underlying datasets depends on the database. For instance, the UniProt¹³ database contains information on genes and proteins, while the ChEBI¹⁴ database contains chemical entities.¹⁵ Both PubMed IDs and accession numbers can easily be converted to resolvable URIs by concatenating them with a stable URL root of the hosting infrastructure.

The link targets of text-mined terms are more heterogeneous, as they depend on the domain of the detected term. For instance, if a term has been identified as a chemical, the link target is an identifiable entry in the ChEBI database. Conversely, if the term has been categorized as gene or protein, the link target is a search query in the UniProt database. Full text links can point to various journal websites, and MeSH terms trigger a literature search in Europe PMC.

It becomes evident that these relations can be categorized along different dimensions. Firstly, we can distinguish them according to cardinality, as known from the domain of relational databases, i.e. one-to-one, one-to-many/many-to-one, or many-to-many. Secondly, the relations carry different levels of trustworthiness, dependant on whether they have been created by a human expert or a machine. Thirdly, there is the dimension of scope, denoting whether the link points to an internal or external representation of the entity. Fourthly and finally, the target type of link varies between a uniquely identifiable object and a (potentially empty) set of objects, as returned by a search query.

Second Demonstrator: Social Sciences and Humanities

The second demonstrator provides examples from the Social Sciences and one example from the Humanities. It builds upon NARCIS, a national portal that has a comparable role to the OpenAIRE portal. The demonstrator investigates how the OpenAIRE portal can best be extended to support different practices for linking publications with data from the Social Sciences and the Humanities, as well as generic interlinking of publications, data, projects and researchers. The examples present different kinds of links and support the discussion how such relations can be captured, displayed and navigated.

¹² Medical Subject Headings: <http://www.nlm.nih.gov/mesh/>

¹³ UniProt database: <http://uniprot.org>

¹⁴ ChEBI database: <http://www.ebi.ac.uk/chebi/>

¹⁵For a complete list of EBI databases, refer to: <http://www.ebi.ac.uk/Databases/>

The examples contain bibliographic descriptions of publications, data sets, research information, people and organisations. The publications and data sets have been harvested from the Dutch institutional repositories via their common standards.¹⁶ The publications and datasets are all identified by OAI-identifiers, and most of them also have persistent identifiers for the copy in the repository as well as a persistent identifier for the publisher-copy. The research projects, people and organisations come from the Dutch Research Databank – a database that is maintained by editors from DANS. Its entries are managed using local identifiers. Around half of the researchers in the database also have an external Digital Author Identifier (DAI).¹⁷

In the demonstrator, each bibliographic description can relate to any other bibliographic description using a semantically labelled link. The example for the oral history book (Zijdeman, 2009) illustrates this. The book is related to many researchers, of whom some are *authors* and others are *editors*. This also holds for related publications: one is a *related* paper which discusses how the interview fragments were processed, one is *cited* by this publication and another one is the *same* publication within another repository.

One of the relations already available within the existing NARCIS infrastructure is the relation between publications and researchers with a DAI. To experiment with other relations, the demonstrator added a manually composed list of relations that specify the subject identifier, the object identifier and a predicate to label the relation (see Table 1).

Subject	Predicate	Object
publication:oai:aup.nl:72413	doap:editor	dai:136423701
publication:oai:aup.nl:72413	doap:author	dai:102167834
publication:oai:aup.nl:72413	doap:fromProject	project:12345
publication:oai:aup.nl:72413

Table 1. Example triples between a publication, researchers and a project.

The relations of a bibliographic description to any other are displayed as contextual information in the sidebars: persons, projects and organizations on the left side, data sets and publications in the right side. These links allow the users to navigate to the corresponding items within the portal. The relations are bi-directional. When a user navigates from a publication to a *cited* dataset, this dataset will be displayed in the center. From here, the user can navigate back to the publication that the dataset is *cited by* (see Figure 2).

The demonstrator uses two existing examples of subject-specific references between publications and data. The Social Sciences examples use the publications from EVS that reference the concepts and variables behind a questionnaire. These references are described in a DDI3-formatted XML file. The demonstrator displays the references with the publication using an XSL stylesheet transformation.

¹⁶ See SURFshare Application Profiles: <http://wiki.surf.nl/display/standards/Application+Profiles>

¹⁷ Digital Author Identifier (DAI): <http://www.surf.nl/en/themas/openonderzoek/infrastructuur/pages/digitalauthoridentfierdai.aspx>

PUBLICATION

WELFARE STATE EFFECTS ON SOCIAL CAPITAL AND INFORMAL... (2005) [Open access](#)

<p>Research</p> <hr/> <p><i>Input</i></p> <p>European Value Studies ></p>	<p>Persons</p> <hr/> <p><i>Author</i></p> <p>Halman, Dr. L.C.J.M. ></p> <p>Halman, L.C.J.M. ></p> <p>Oorschot, W.J.H. van ></p> <p>Arts, W.A. ></p>	<p>Organisations</p> <hr/> <p><i>Repository</i></p> <p>Tilburg University ></p>	<p>External Database Links</p> <p>Concept: ethnocentrism Variable: V59: dont like as neighbours: muslims (Q7H) Variable: V60: dont like as neighbours: immigrants/foreign workers (Q7I) Variable: V64: dont like as neighbours: jews (Q7M)</p>	<p>Data</p> <hr/> <p><i>Cites</i></p> <p>European Values Study 199... ></p> <p>EVS'99/2000 : Release I... ></p>	<p>Publications</p> <hr/> <p><i>Cites</i></p> <p>Welfare States, Solidarit... ></p> <p>Three worlds of welfare c... ></p> <p>Who should get what, and... ></p> <p>Individual motives for co... ></p> <p>Wij en zij in Europa: De... ></p>	<p><i>Cited by</i></p> <p>Making the difference in... ></p> <p>Culture and social policy... ></p> <p>Multi-level determinants... ></p>
---	--	---	---	--	---	---

TILBURG UNIVERSITY

Figure 2. Screenshot of a Social Sciences publication *in context*.

In addition, the identifiers of the variables are transformed into hyperlinks to their corresponding descriptions in the data repository at GESIS.¹⁸ The example for oral history uses a similar approach. The references to streaming fragments listed in the book are also described in an RDF/XML file. Like the EVS demonstrator, an XSL stylesheet transformation displayed these as short transcriptions with links to the streaming fragments so the user can listen to them. Note that the complete interviews are also referenced as data citations, but these are not accessible due to privacy issues. However, the fragments were evaluated and approved for public access.


Discussion

In this section the demonstrators from the previous section are compared and discussed. It describes the commonalities, the differences and other issues that were encountered and will propose solutions. Though such findings are supposed to be generic, they are the result of demonstrators that represent just selected examples. These findings are therefore only meant for discussion and for comparison with other examples.

Generic Entities and Relations

Both demonstrators are based on bibliographic descriptions for publications, datasets, research projects and researchers, interlinking these and allowing users to navigate between them. The bibliographic descriptions could all be fetched from their original sources in their original metadata. Many of them could be retrieved using OAI-PMH, but other sources were also used (normal web resources, dedicated APIs). The available metadata could be reused for the purpose of these demonstrators. It is beyond the scope of this paper to discuss the need for more standardized or specialized metadata schemas.

¹⁸ GESIS: <http://zocat.gesis.org/>



The linking depended on the identification of these resources. Commonly used identifiers, such as PubMed IDs, DAIs and DDI3-identifiers, are good examples of well-defined identifiers that support automatic linking, though even these have their limitations: PubMed IDs are only globally unique and persistent within their own domain and DAIs are only assigned systematically within The Netherlands. We recommend more standardized application of globally unique persistent identifiers for resources, as well as standardized, globally unique naming schemes for project (funding) and researchers. The existing metadata schemes could need slight adjustments to reference these identifiers. It is important that the assignment is taken care of by the appropriate stakeholder at the appropriate moment in the workflows, allowing others to define trusted references to these entities.

The “by reference” model appeared useful for these demonstrators: the identification of entities using well-established schemes such as PubMed IDs for publications or DAIs for Dutch authors allowed other entities to contain references to them. The demonstrators do not yet show a need for the “package” model. Moreover, we foresee many challenges when supporting such packages, as they introduce a new entity with no clear stakeholder in the distributed environment of scholarly communication.

The relations between the different resources need to be semantically typed, as they can have different meanings. The relation can also be defined in different ways: by the creator of the resource, by an expert curator, by an automated inference algorithm or by crowdsourcing. It is therefore important to register this origin, as it implies different levels of reliability.

Discipline Specific Entities and Relations

An important difference between the demonstrators can be observed in the concept of what a “dataset” is. Different databases in the Life Sciences, as well as DDI3-encoded data in the Social Sciences, are very well structured which allows detailed identification and relations with their contents. However, their internal structures differ from each other. The Humanities’ data is the most heterogeneous, so the most common structure is defined by files and folders.

The diversity of the discipline-specific data, its complex structures and the advanced visualization that is required to interpret it, suggests that its management should be delegated to the original data sources. This fits the “by reference” model, where all relations to objects are encoded as hyperlinks to their representations in the original infrastructures.

Early feedback from both researchers and repository managers indicates that such access to detailed data entities is not of primary concern. It is more important that a researcher can discover other publications and data sources related to, for example, a concept in a questionnaire, rather than being able to analyse it from the portal. This raises an important challenge: how can a cross-disciplinary portal provide different subject-specific indexing for all its resources?

Other feedback primarily concerned the example from the Humanities that showed interview fragments with the publications. The feedback taught us that such references

were of no value without their context, which is described in the publication itself. However, it is useful to be able to discover the availability of the complete interviews. Another valuable feature that researchers and repository managers mentioned was discovery via categorizations in terms of a detailed academic discipline, time, space and persons.

Conclusion

The examples in the demonstrator share general concepts, such as publications, datasets, and projects. The “by reference” model is very well suited to connect these concepts, but relies on globally unique and stable identification schemes. OpenAIREplus is therefore to endorse the use of persistent identifiers for publications and datasets, but also schemes for identifying researchers, such as ORCID¹⁹ or ISNI²⁰, and a (to be developed) scheme for project funding. Furthermore, vocabularies are needed to describe the different types of relations. With regards to linking publications to datasets, we therefore recommend considering the activities of initiatives like DataCite.

The main characteristics of the different disciplines are the way they structure and describe the objects within their own specific databases. It is infeasible for infrastructures like OpenAIRE to manage such objects from each and every discipline. Stable and globally unique identifiers allow OpenAIRE to index these relations between these objects so users can search for publications that reference these objects or browse them in their original sources. We recommend that OpenAIREplus should collaborate with the different communities to specify the relevant schemes.

The relations among the different objects can be captured in different ways. We recommend that they are captured early in the workflows by the most knowledgeable stakeholder, usually the author or creator of a resource. This implies the availability of the identifiers for the items to be referenced. The availability of vocabularies and facilities for smart, auto-complete forms can support the establishment of those relations. To overcome the absence of relations among existing materials, methods for automatic association of digital objects that have been explored recently (Boland et al., 2012) need to be investigated further in the future.

The process of constructing the demonstrators was only the first step towards a common approach of discipline-independent interlinking of research information. Further discussion and comparison with other examples is the next step. The demonstrators support this discussion by providing concrete examples.

Acknowledgements

The work presented in this paper has been funded by the European Commission as part of the project OpenAIREplus (FP7-INFRA-2011-2, Grant Agreement no. 283595). The authors would like to thank Alessia Bardi, Natalia Manola, Mateusz Kobos, Jo McEntyre, Oliver Kilian for their support and contributions.

¹⁹ ORCID: <https://orcid.org/>

²⁰ ISNI: <http://www.isni.org/>



References

- Ball, A., Duke, M. (2012). *Data Citation and Linking*. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/briefing-papers/>
- Berg, H., Scagliola, S.I. & Wester, F.P.J. (Eds.). (2010). *Wat veteranen vertellen : Verschillende perspectieven op verhalen over ervaringen tijdens militaire operaties*. Amsterdam University Press. [doi:10.5117/9789085550341](https://doi.org/10.5117/9789085550341)
- Boland, K., Ritze, D., Eckert, K. & Mathiak, B. (2012). Identifying references to datasets in publications. In P. Zaphiris, G. Buchanan, E. Rasmussen, and F. Loizides (Eds.), *Theory and Practice of Digital Libraries* (pp. 150–161). Lecture Notes in Computer Science: Vol. 7489. Springer. [doi:10.1007/978-3-642-33290-6_17](https://doi.org/10.1007/978-3-642-33290-6_17)
- Brase, J. (2009). Datacite: A global registration agency for research data. In *Cooperation and Promotion of Information Resources in Science and Technology, 2009 (COINFO'09)*. IEEE.
- Kunze, J., Cruse, P., Hu, R., Abrams, S. et al (2011). *Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines*. CDL Staff Publications. Retrieved from <http://n2t.net/ark:/13030/c7jw86m55>
- Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S., Sanderson, R. & Johnston, P. (2008). *Object re-use & exchange: A resource-centric approach*. Retrieved from <http://arxiv.org/abs/0804.2273v1>
- Lawrence, B., Jones, C., Matthews, B., Pepler, S. & Callaghan, S. (2011). Citation and Peer Review of Data: Moving Towards Formal Data Publication. *International Journal of Digital Curation* 6(2), 4-37. [doi:10.2218/ijdc.v6i2.205](https://doi.org/10.2218/ijdc.v6i2.205)
- McEntyre, J., Ananiadou, S., Andrews, S., Black, W., Boulderstone, R., Buttery, P., Chaplin, D., Chevuru, S., Cobley, N., Coleman, L. et al. (2011). UKPMC: A full text article resource for the life sciences. *Nucleic acids research* 39(suppl 1), D58–D65. [doi:10.1093/nar/gkq1063](https://doi.org/10.1093/nar/gkq1063)
- Smits, E. & van der Plicht, J. (2009). Mesolithic and Neolithic human remains in the Netherlands: physical anthropological and stable isotope investigations. *Journal of Archaeology in the Low Countries*. Retrieved from <http://dpc.uba.uva.nl/jalc/01/nr01/a04>
- Woutersen-Windhout, S. et al. (2009). *Enhanced Publications: Linking Publications and Research Data in Digital Repositories*. Amsterdam: Amsterdam University Press. [doi:10.5117/9789089641885](https://doi.org/10.5117/9789089641885)
- Zijdeman, R.L. (2009). Like my father before me: Intergenerational occupational status transfer during industrialization (Zeeland, 1811-1915). Retrieved from <http://www.narcis.nl/vpub/RecordID/escape-demo:rem:2679/id/1/Language/EN>