# Semantically Enhanced Interactions between Heterogeneous Data Life-Cycles

## Analyzing Educational Lexica in a Virtual Research Environment

Basil Ell[1], Christoph Schindler[2], and Marc Rittberger[2]

[1] Karlsruhe Institute of Technology (KIT),
Karlsruhe, Germany
basil.ell@kit.edu,

[2] German Institute for International Educational Research (DIPF),
Frankfurt am Main, Germany
{schindler, rittberger}@dipf.de

**Abstract.** This paper highlights how Semantic Web technologies facilitate new socio-technical interactions between researchers and libraries focussing research data in a Virtual Research Environment. Concerning data practices in the fields of social sciences and humanities, the worlds of researchers and librarians have so far been separate. The increased digitization of research data and the ubiquitous use of Web technologies change this situation and offer new capacities for interaction. This is realized as a semantically enhanced Virtual Research Environment, which offers the possibility to align the previously disparate data life-cycles in research and in libraries covering a variety of inter-activities from importing research data via enriching research data and cleansing to exporting and sharing to allow for reuse. Currently, collaborative qualitative and quantitative analyses of a large digital corpus of educational lexica are carried out using this semantic and wiki-based research environment.

**Keywords:** Virtual Research Environments, Research Infrastructures, Digital Libraries, Semantic Web technologies

## 1 Introduction

As more and more digital libraries such as the Europeana[3] or services of the German National Library[4] come to exist and make their content available using Semantic Web standards, new opportunities arise for research projects to exploit these resources. These resources offer new capacities for research, which can be addressed by Virtual Research Environments offering a variety of methods and tools [7, 18].

[3] http://www.europeana.eu

[4] http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html (2013-08-19)

In this paper, we describe in detail how Semantic Web technologies can be applied to enable new socio-technical interactions between researchers and libraries based on research data in a Virtual Research Environment. While typically the data life-cycles of digital libraries and of researchers are separated, we exemplify how semantically enhanced Virtual Research Environments offer concrete capacities for interaction between these worlds. Semantic Web technologies offer to these kinds of research environments a possibility for heterogeneous data production, publication, and sharing, supporting a main affordance of data practices in these fields: multiple layers for overlapping and interacting data life-cycles.

An example is given of a concrete research practice, i.e. the qualitative and quantitative analysis of a large digital corpus of educational lexica. The semantically enhanced research environment is built upon Semantic MediaWiki [10]. Additionally, new SMW extensions addressing requirements of research practices are developed. The VRE is created within the project "Semantic MediaWiki for Collaborative Corpora Analysis" (SMW-CorA) in view of concrete needs, taking a participatory and agile design approach based on ongoing requirement elicitation with researchers carrying out their long-term research project within this environment.

The paper is organized as follows: Section 2 gives an overview of the state of the art in the fields of Virtual Research Environments, libraries and the Semantic Web. Afterwards, the design approaches and the support of researchers are described in Section 3, followed by a description of the realized heterogeneous research data life-cycle and its interactions within the semantically enhanced Virtual Research Environment in Section 4. These activities are highlighted in respect of the role of Semantic Technologies in Section 5. The paper concludes with a discussion and an outlook in Section 6.

## 2 Virtual Research Environments, Digital Libraries, and the Semantic Web

Today, Virtual Research Environments (VREs) aim to enhance research by using the capabilities of networked technologies, distributed resources and computational power.[5] While previously the facilitation of research in the fields of science and technology has been focussed, the World Wide Web is beginning to impact on the fields of humanities and social sciences, addressing and enforcing collaboration of researchers in projects and beyond. Some VREs have started to use Semantic Web technologies in order to enhance research practices. *MyExperiment* is one example, which allows users to create, share and publish workflows of scientists [5]. Accordingly, resources are semantically described and published as RDF data. Furthermore, the VRE *ourSpaces* [6] uses an ontological framework for provenance [12], semantic policy reasoning for access management of resources [17], and a user interface to create metadata. There is still a lack of VREs missing that offer semantically-enhanced interactions on the research data

[5] Some VREs are listed at `http://misc.jisc.ac.uk/vre/projects`. (2013-08-19)

level. Our VRE SMW-CorA allows users to semantically interact in various ways with research data for which the import process and the formal annotation is described in [16, 14, 15].[6] In comparison to other VREs, SMW-CorA integrates the research data itself in a semantic environment and enables researchers to carry out research directly on a semantic level: underlying research data as well as data created via research interactions are represented semantically whereas in previous approaches an additional translation process is necessary.

In the domain of Semantic Web, Auer et al. [1] recently described a life-cycle of Linked Data. Beyond the scope of this work by Auer, we would like to emphasize the possible intersections and overlappings of various life-cycles, which we address in a research environment and its interactions with a digital library in a concrete field of humanities and social sciences. The background is that data practices in research and the prospects of data sharing are identified as a conundrum, a problem which has to be addressed for particular research communities and their concrete interactions in research practice [3]. In this paper, the interaction between the library and research data practices is considered in detail where the data and its different life-cycles will be focused.

## 3 Design and Method

A recent large international survey of social media indicates that Semantic Web technology offers great capacities to enhance research and data practices. The study shows that social media are already used "at all points of the research life-cycle, from identifying research opportunities to disseminating findings at the end" [13]. Nevertheless, the challenge is raised especially in the fields of humanities that requirements should be articulated by the scholars themselves [2] and the solutions should be aligned to specific research communities [3].

Therefore, a specific approach is used for designing concrete capacities of interaction in research practice: a participatory design approach with agile development has been carried out, as required for Virtual Research Environments [4, 18]. Apart from researchers, staff members working with a digital library (Research Library for the History of Education, BBF) were involved as active participants in the requirement elicitation and realization process. For establishing ongoing feedback loops we organized several on-site meetings and continued to hold online team meetings in the course of the project. Further, the iterative development, the requirement elicitation as well as the testing of the realized functionalities relied on two researchers in the field of history of education who are carrying out their research in the VRE. After introducing the VRE, in particular the features, the syntaxes of MediaWiki and Semantic MediaWiki to the researchers, they learned to carry out their research within this environment and to explore the data by writing queries while being assisted when necessary by the developers. For two years the two PhD projects manually integrated more than 60 lexica and performed more than 17.700 edits. For the future it is planned to

[6] Further realizations of SMW as a VRE are Docupedia (`http://www.docupedia.de`), [9] as an archaeological infrastructure, and [11] as an archaeological corpus.

evaluate i) the applicability of the VRE to further research projects within the domain of social sciences and humanities, and ii) to assess the usability via cognitive walkthroughs and expert interviews. Recently, we started to realize a VRE for analyzing historical school magazines as a first step towards the assessment of its applicability.

## 4 The Research Data Life-cycle and its Interactions

Data providers may adhere to a data life-cycle model such as the DCC Curation Life-cycle Model [8]. While this model covers actions such as the access to the data by designated users and reusers and the action of receiving data in accordance with documented collecting policies, this model focuses the curator perspective to research data and needed interactions. Given a landscape with multiple data collections maintained by diverse initiatives where overlap exists regarding the digital objects and databases they are centered around, interactions not foreseen by the DCC model can take place. Data consumers e.g. researchers who perform research on data retrieved from a digital library may add new levels of data, enrich it with further information, add missing pieces, create abstractions and aggregations, complement it with new data, add new perspectives or identify and correct errors in the data. The main interactions – described in more detail in subsequent sections – in our realization of a heterogeneous life-cycle model are as follows:

**Importing research data:** Research data such as historical lexica that are hosted by a source such as a digital library and that are relevant for a particular research project carried out within the VRE are imported.

**Enriching research data:** Research data imported into the VRE or created within the VRE are enriched with further information, missing pieces are added, abstractions and aggregations are created, they are complemented with new data, and perspectives are added.

**Data cleansing:** Errors in the data are identified and corrected.

**Exploring and analyzing:** Unstructured and structured content can be explored; structured content can be qualitatively and quantitatively analyzed.

**Export and sharing:** Content can be selected and exported to allow for reuse by third parties.

Note that these interactions need not be executed in any specific sequence. They can be carried out at any time, in parallel, in any order and an arbitrary number of times. Moreover, they are dependent on and enabled by semantic metadata. In the following sections each of these interactions are discussed. Beyond performing interactions with the digital library, the research projects carried out within the VRE may interact due to overlap in lexica relevant to their projects, as well as with life-cycles of data within the Semantic Web in general.

### 4.1 Importing Research Data

As mentioned above, the research data of the exemplarily realization of the VRE are historical lexica. Lexica that are of interest here are mainly available at the digital library Scripta Paedagogica Online (SPO),[7] hosted by the Research Library for the History of Education (BBF)[8] which has indexed the lexica and rendered them accessible online as image files as part of their library life-cycle. The corpus contains a total amount of nearly 22,000 articles and more than 25 lexica. Each lexicon is bibliographically described as a collected edition in the library database allegro-C and the digital library environment Goobi.[9]

The data consists of images of scanned pages and pertinent metadata. Therein, four levels of entities, their properties and relations are formalized (lexicon, volume, lemma, and image). This collection is accessible via an OAI interface.[10] A custom-made application of the VRE communicates with this interface, creates representations reflecting the levels of entities within the VRE, and imports the scanned images [16]. How this tool creates representations of the data within the VRE can be specified using XSLT (Extensible Stylesheet Language Transformations) documents. The development of a custom import tool is necessary, if the data to be imported are not available via an OAI interface.

Several Semantic Web vocabularies were imported into the VRE to represent the available metadata. These are: FOAF, PRISM, BIBO, SKOS, and DC.[11] Using these well-known vocabularies has benefits since this offers the reusability of exported data by third parties as discussed in Section 4.5.

Additionally to the life-cycle of the digital library the researcher themselves have their own life-cycles of creating and using research data. Thereby they define the scope of relevant lexica in respect to their research question which is iteratively adjusted while getting new insights in the research process. To offer these capacities of integrating lexica which are not digitized and available at the digital library SPO, the *OfflineImport* feature was developed. This feature allows creating pages for lexica, volumes, lemmata, and images given a minimal set of metadata with minimal effort for the researcher. Researchers can annotate the pages of these lexica and perform their analysis in the same way as the automatically integrated lexica. Thus, the data life-cycles of the digital library and the research carried out within the VRE are interacting thus offering feedback in an additional direction: Digital libraries are thus informed about potential consumers of relevant research data, they can set priorities in their digitization activities, and inform the researchers once the requested content has been digitized.[12]

---

[7] The lexica are available at `http://bbf.dipf.de/digitale-bbf/scripta-paedagogica-online/ digitalisierte-nachschlagewerke`.

[8] `http://bbf.dipf.de/en`

[9] See `http://www.allegro-c.de/` and `http://www.goobi.org`

[10] An OAI interface implements the Protocol for Metadata Harvesting (OAI-PMH) defined by the Open Archives Initiative (`http://www.openarchives.org/`).

[11] FOAF (Friend of a Friend ontology), PRISM (Publishing Requirements for Industry Standard Metadata), BIBO (The Bibliographic Ontology), SKOS (Simple Knowledge Organization System), and DC (DCMI Metadata Term).

[12] While these interactions are in principle possible and are technologically already realized within our VRE, the processes are yet to be established with the BBF.

### 4.2 Enriching Research Data

Research data such as the imported lemmata need to be annotated to facilitate analysis. In the social sciences this process of annotating segments of text is referred to as *coding* where annotation facilitates qualitative and quantitative analysis of the content. Since the research data, the historical lexica, are integrated as images from scanned pages, parts of images are annotated instead of parts of texts. Therefore, the *SemanticImageAnnotator* was developed and published as an extension of MediaWiki that allows to select and annotate (either with free annotations or existing thesauri or classification systems) rectangular areas on images.[13] [14]

As an example of an annotated part of an image, a section of a lemma of a lexicon can be annotated with i) the type of argument used by the author, such as a moral or ideological argument, ii) the topic that is subject of the argument, and iii) the position taken by the author towards this topic. Furthermore, the affiliation of the author, such as the affiliation to a religious institution, can be stored on the author's page. Thus, by coding and linking the imported and created research data, a semantic network of relevant entities is created by the researchers. Each annotation is stored as an object that links to the image and has semantic properties such as the coordinates on the image, categories, tags, and any other property the researcher wishes to assign. This network can then be subject to qualitative and quantitative analysis (as discussed in Section 4.4) where this network serves the purpose of a surrogate for the underlying research data which is not computer-processable.

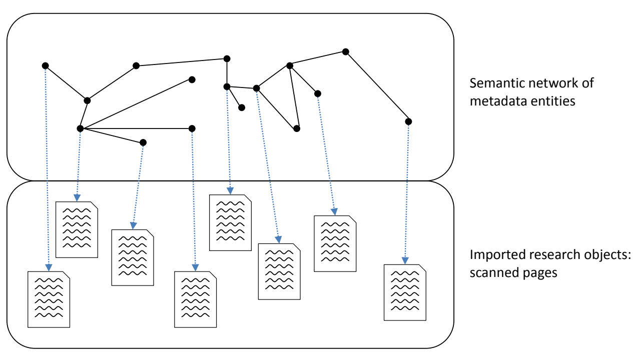


**Fig. 1.** The layers of research data and the semantic network.

[13] This technique can be applied in further use cases of qualitative social sciences and humanities projects where images have to be coded (annotated). Further uses for tool are the annotation of technical diagrams (construction plans, floor plans), pieces of visual art, photos where people and objects are annotated, and the like.

[14] Available at: `http://www.mediawiki.org/wiki/Extension:Semantic_Image_Annotator`.

This semantic network (depicted in Figure 1) consists of two layers. The bottom layer consists of the imported images which are not computer-processable. The layer above consists of the semantic network where research data are represented as nodes that are linked with further nodes depicting the imported metadata (lexicon, volume, lemma, etc.) as well as entities created within the VRE. Data in both levels can be maintained by multiple life-cycles, thereby establishing interactions between digital libraries, research projects and further initiatives and data sources. When referring to entities during coding, these entities do not need to exist and can be created automatically.

To give an example for an analysis that exploits this network, imagine the following situation: The concept of *Affenliebe* (infatuation) is mentioned in three lemmata L1, L2 and L3 which are authored by three persons P1, P2, and P3 who are affiliated to the institutions I1, I2, and I3, respectively. The term is negatively connotated in L1 and L2 but positively connotated in L3 thus the set of institutions is clustered into two groups regarding the attitude towards this concept: {I1, I2} and {I3}. Furthermore for I3 it is known that, contrary to the institutions I1 and I2, lemmata authored by members of I3 are usually characterized by a religious perspective. This distinction may serve as a basis for explaning the clustering.

Besides linking research data by coding and linking created entities with other created entities, entities can also be linked to entities outside of the VRE. For this purpose we co-developed the *SemanticWebBrowser*[15] extension. On an entity's page, further URIs referring to this entity can be stored. For example, on the page representing an author of a lemma, the author's unique identifier[16] as used by the authority file of the German National Library (DNB) can be stored. At the bottom of each page a fact box displays all property and value pairs, such as `Profession: Tutor`, that are stored within the local page. The *SemanticWebBrowser* extends this list by external facts – property and value pairs retrieved by retrieving RDF data available at this URI. For example, the property and value pair `Date of birth: 1900` could be retrieved from data provided by the DNB. Therefore, researchers can become aware of externally available data and decide for manually importing this data.[17]

### 4.3 Data Cleansing

Data imported from external sources such as a digital library could contain errors. Further possible sources of error are the import process or the work of other researchers within the VRE. Since these errors can distort results of analyses, researchers need to be able to correct an error or to mark an error and ask for help. Therefore, the researchers can specify in the VRE that they

[15] Available at `http://www.mediawiki.org/wiki/Extension:Semantic_Web_Browser`.
[16] For example for Josef Spieler: `http://d-nb.info/gnd/117483885/about/rdf`
[17] Remaining issues of the beta version of this SMW extension are: 1) to automatically identify additional URIs for an entity, and 2) to enable the users to select which externally available facts to import.

identified and corrected an error, whereby they can describe in a *DataCorrection* element the value they replaced, the reason why it had to be changed and the source this decision is based on. Here, the corrective action can be the removal of a triple (such as *Josef Spieler*, *date of birth*, *1900*) or addition of a triple or both. These corrections are stored as objects carrying semantic properties which offers the possibility for tracing back the changes of structured data within the VRE and between the life-cycles. For example, a list of modifications can automatically be compiled with background information for a source such as a digital library, to inform the original provider of this data about the encountered errors and to publish these corrections so that other consumers can adapt their data as well. This list of modifications and justifications can be exported using the semantic data format RDF. Once confirmed by a data provider, a correction performed within a VRE can be turned into e.g. a SPARQL UPDATE or SQL INSERT command and executed by the provider of the remote data source.

The data correction workflow is depicted in Figure 2. In order to correct the data (which is either imported or created within the VRE), the researcher edits the respective wiki page where the imported data is stored, adds the code `{{DataCorrection}}` and saves the page. This leads to the display of a small box on the page which contains a link to a form where data about the correction can be entered. After editing the correction object, this box displays data entered via the form.

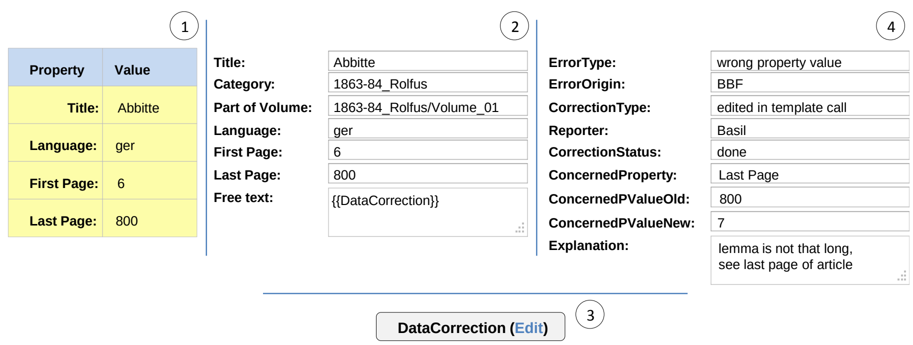


**Fig. 2.** The data correction feature. The value 800 in (1) is wrong. The page can be edited as shown in (2) and the value can be corrected. When saving the page the DataCorrection box is displayed on the page as shown in (3). The edit button within this box leads to an edit form as shown in (4) where details about the correction can be provided.

At the time of the submission of this paper to the best of our knowledge no ontology exists that allows to represent patches to ontologies. While a Graph Update Ontology[18] exists, this ontology is intended to describe which changes

[18] http://webr3.org/specs/guo/

to apply to an ontology automatically. It does not allow specifying the error and to provide evidence or arguments in natural language targeted at the data maintainer, which needs to decide whether to agree and apply each patch. This is a main need of the library to start their quality maintenance activities and requirement for interaction capacities on this level between library and research life-cycles.

### 4.4 Exploring and Analyzing

For the purpose of exploring the VRE's content, researchers can create and embed queries written in ASK – the query language of SMW – and thus create dynamic views of the content. Examples for these queries are a list of relevant lexica for a certain project that contains a lemma that is annotated with certain terms from a taxonomy or a depiction of dates of birth of lemma authors for a certain lexicon on a timeline. Figure 3 shows an example of a qualitative analysis: the visualization of reference types in annotated lemmata. A lemma can refer to another lemma within the same lexicon (internal reference), to an author (reference to author), and to another publication (reference to literature). For each type of reference each annotation is depicted with a square which is either grey-colored if it does not represent a reference of this type or colored depending on the type of reference it represents.


**Fig. 3.** Qualitative analysis: visualization of reference types in annotated lemmata

### 4.5 Export and Sharing

SMW provides facilities to export query results in non-semantic formats such as CSV and JSON, but also to export content as RDF data. Besides sharing information and patches about incorrect data available in external sources as discussed in Section 4.3, results of the researchers' efforts (such as the efforts to enrich and interlink the research data and to create new entities) can be exported as well. Here, the identifiers from imported vocabularies, as described in Section 4.1 are used. For example, if the property *knows* is imported from the FOAF vocabulary, while exporting RDF data for an entity that uses this property then instead of the wiki's own identifier for this property, *foaf:knows* is used which is the identifier of the imported property. This increases the prospects for the exported data to be readily reusable and integrable in other contexts due to the use of well-known vocabularies.

## 5 Potentials of Semantic Web Technologies

The exemplarily realization of the VRE for the analysis of educational lexica offers several capacities for interactions between the life-cycles of digital libraries and research. On this basis it is possible to summarize the following potentials:

- Import of research data from a digital library is preceded by importing existing vocabularies into the VRE. Research data can then be stored and represented using terms from these standardized vocabularies. The benefits of using these vocabularies are the increased prospects for the data to be readily reusable and integrable in other contexts by third parties. Storing information about equivalent resources residing outside the VRE enables referencing of these entities and display of externally available data.
- The result of the enrichment activities, the semantic network, can be queried and serves as input to qualitative and quantitative analyses. Nevertheless, the layers – where data has been imported from – remain separable but individually semantically addressable. Results of these analyses are dynamic since they depend on the query results as input thus reflecting the current state of the semantic network.
- Annotations created with the *SemanticImageAnnotator* allow to specify which research project an annotation belongs to. This additional information allows separating the annotations of the distinct projects.
- The schema used to represent and link entities can be updated at any time by introducing and using new categories and properties. By using a so far unknown property or category the VRE's ontology is extended thus easily offering a semantic continuum for providing a free degree of formalization in articulation.
- Each object created using the *DataCorrection* feature stores information that enable users to compile lists of correction proposals for each source. Informing the data provider about these errors can benefitial for the provider as well as for other consumers of this data. Patches can be shared with other consumers as well.
- A missing value can be made explicit by using so-called gardening properties. Regarding a missing value for a property $P$, for the property *missing value for property* the value $P$ can be stored. A gardening page lists all pages where a certain property value is missing. This can guide the enrichment process.

## 6 Discussion and Outlook

In this paper we describe a semantically enhanced and wiki-based Virtual Research Environment which addresses socio-technical interactions between researchers and digital libraries offering new ways of collaboration throughout their different life-cycles. In detail, we show how these interactions are supported and enhanced through Semantic Web technologies balancing thoroughly and fine-grained the intersections between library, research and the Semantic Web in general. While the benefits of the utilized Semantic Web technologies are manifold and enable addressing and capturing of heterogeneous data practices, the realization exemplifies the need to adjust the environment to these concrete practices. Different tasks and quality aspects of the life-cycles have thus been implemented. While the realization targets a specific community rooted in the history of education, the supported tasks, the technology and the method developed within the project can be transferred to and be reused in multiple research endeavours since the functionality they provide is not specific to the needs of this particular research. It remains to be evaluated whether in future expansion of our focus these tools are applicable. Concrete examples for transfer and reuse of technical developments are the *SemanticImageAnnotator*, the *SemanticWebBrowser* and the established workflows for data corrections.

Furthermore, designing for concrete research practices and offering a flexibility of the environment needs to take into account that an ongoing support is required until the end of a research project in order to stabilize the endeavor and ensure the scientific output. Currently we are preparing to extend the interaction capacities of the VRE with tools for further types of research data such as videos and interview transcripts. Designing these interaction-aware Virtual Research Environments is one step towards future ecologies of small to large research projects and data providers where data flows between the participants lead to enriched and improved data thus providing benefits in a multitude of collaborations.

## Acknowledgements

This project is funded by the German Research Foundation (DFG) entitled: Entwicklung einer Virtuellen Forschungsumgebung für die Historische Bildungsforschung mit Semantischer Wiki-Technologie - Semantic MediaWiki for Collaborative Corpora Analysis (RI 803/10-2, STU 170/21-2) in the domain of Scientific Library Services and Information Systems (LIS). It is realized in a cooperation between the German Institute for International Educational Research (DIPF), the Karlsruhe Institute of Technology (KIT), the Library for Research on Educational History (BBF), and historical educational researchers mainly of the Georg-August-University Göttingen. We are grateful that Rudi Studer, Denny Vrandečić, Elena Simperl, Cornelia Veja, Klaus-Peter Horn, Anne Hild, Anna Stisser, Benedikt Kämpgen, Martin Wünsch, Sabine Liebmann, Stefan Cramme, and Gwen Schulte have actively supported our endeavor.

## References

1. S. Auer, S. Tramp, B. van Nuffelen, R. Isele, J. Lehmann, L. Bühmann, C. Dirschl, P. N. Mendes, H. Williams, O. Erling, et al. Managing the Life-Cycle of Linked

Data with the LOD2 Stack. In *The Semantic Web–ISWC 2012*. Springer Berlin Heidelberg, 2012.

2. C. L. Borgman. The Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quarterly*, 3(4), 2010.
3. C. L. Borgman. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6):1059–1078, 5. April 2012.
4. A. Carusi and T. Reimer. Virtual Research Environment collaborative landscape study. *JISC, Bristol*, 2010.
5. D. De Roure, C. Goble, and R. Stevens. The design and realisation of the Virtual Research Environment for social sharing of workflows. *Future Generation Computer Systems*, 25(5):561 – 567, 2009.
6. P. Edwards, E. Pignotti, A. Eckhardt, K. Ponnamperuma, C. Mellish, and T. Bouttaz. ourSpaces–Design and Deployment of a Semantic Virtual Research Environment. In *The Semantic Web–ISWC 2012*, pages 50–65. 2012.
7. M. Fraser. Virtual research environments: overview and activity. *Ariadne*, 44:31–40, 2005.
8. S. Higgins. The dcc curation lifecycle model. In R. L. Larsen, A. Paepcke, J. L. Borbinha, and M. Naaman, editors, *JCDL*, page 453. ACM, 2008.
9. I. Huvila. Being Formal and Flexible: Semantic Wiki as an Archaeological e-Science Infrastructure. In *Revive the Past: Proceeding of the 39th Conference on Computer Applications and Quantitative Methods in Archaeology, Beijing, 12-16 April 2011*, pages 186–197, 2012.
10. M. Krötzsch, D. Vrandečić, and M. Völkel. Semantic MediaWiki. In *The Semantic Web – ISWC 2006*, pages 935–942. Springer Berlin Heidelberg, 2006.
11. E. Leclercq and M. Savonnet. Access and annotation of archaelogical corpus via a semantic wiki. In *Fifth Workshop on Semantic Wikis-Linking Data and People (Semwiki)*, 2010.
12. L. Moreau, J. Freire, J. Futrelle, R. McGrath, J. Myers, and P. Paulson. The open provenance model: An overview. *Provenance and Annotation of Data and Processes*, pages 323–326, 2008.
13. I. Rowlands, D. Nicholas, B. Russell, N. Canty, and A. Watkinson. Social media use in the research workflow. *Learned Publishing*, 24(3):183–195, 2011.
14. C. Schindler and B. Ell. Semantic MediaWiki for Collaborative Corpora Analysis: Analyzing Educational Reference Books in a Virtual Research Environment. Peer-reviewed abstract, September 2012. ECER 2012.
15. C. Schindler and B. Ell. Kollaborative Analyse von historischen Netzwerken: Virtuelle Forschungsumgebung für die Historische Bildungsforschung. In A. H. u. P. M. Hans-Ulrich Grunder, editor, *Netzwerke in bildungshistorischer Perspektive*. Verlag Julius Klinkhardt, Bad Heilbrunn, Oktober 2013.
16. C. Schindler, C. Veja, M. Rittberger, and D. Vrandečić. How to teach digital library data to swim into research. In *Proceedings of the 7th International Conference on Semantic Systems*, pages 142–149, New York, NY, USA, 2011. ACM.
17. M. Şensoy, T. Norman, W. Vasconcelos, and K. Sycara. OWL-POLAR: Semantic policies for agent reasoning. In *The Semantic Web – ISWC 2010*, pages 679–695. Springer Berlin Heidelberg, 2010.
18. A. Voss and R. Procter. Virtual research environments in scholarly work and communications. *Library Hi Tech*, 27(2):174–190, Juni 2009.