

Fusing Text and Image Data with the Help of the OWLnotator

Giuseppe Abrami¹, Alexander Mehler¹, and Dietmar Pravidà²

¹ Johann Wolfgang Goethe-Universität Frankfurt am Main, Frankfurt, Germany
abrami@em.uni-frankfurt.de

² Frankfurter Goethe-Museum/Freies Deutsches Hochstift, Frankfurt, Germany

Abstract. A central challenge for any approach to mining multimedia data concerns the availability of a *unified* semantics that allows for the *fusion* of multicodeal information objects. To meet this challenge, a format is needed that enables the representation of multimedia data even across the border of different (e.g. iconic and symbolic) codes using the *same* ontology. In this paper, we introduce the *OWLnotator* as a first step to meeting this dual challenge by example of text-image relations. The *OWLnotator* is presented as part of the *eHumanities Desktop*, a browser-based, platform-independent environment for the support of collaborative research in the digital humanities. It focuses on modeling and analyzing multicodeal, multimedia information objects as studied in the humanities. The eHumanities Desktop contains a wide range of tools for managing, analyzing and sharing resources based on a scalable concept of access permissions. Within this framework, we introduce the *OWLnotator* as a tool for annotating intra- and intermedia relations of artworks. The OWLnotator allows for modeling relations of symbolic and iconic signs of various levels of resolution: ranging from the level of elementary constituents to the one of complete texts and images. To this end, the *OWLnotator* integrates *TEILex* (a system for interrelating corpus and lexicon data as part of the eHumanities Desktop) with the expressiveness of *OWL*-based ontologies in order to meet the first part of our twofold challenge. As an evaluation, we illustrate the *OWLnotator* by means of “*Illustrations of Goethes Faust*”.

1 Introduction

At present, the field of digital humanities – and especially of digital art history – is still in its inception phase. Its main goals are to provide researchers access to large collections of images and to enable them to handle those materials. The first task is well understood and its importance is sufficiently appreciated. The digitizing of extensive bodies of images housed in archives and museums is under way all over the world, and it will continue to be a priority for most institutions for some time to come. There is an acute awareness of the necessity of standards for digitizing collections, some best practices have been established and are widely accepted. On the other side, the tools for exploring those research materials in-depth are still in need of being developed. For the descriptive metadata of images, museums have

generally adopted the CIDOC *Conceptual Reference Model* [5], which provides an extensible ontology for concepts and information in cultural heritage and museum documentation. However, the development of adequate content research tools for large volumes of digitized images is, owing to the inherent complexity of the task, not so far advanced. And yet, there is a need for devices that allow fine-grained searches in order to find similarities, non-obvious affiliations, patterns and connections between images and their relations to other media such as texts. At the present stage, one possible next step is to focus on single small- or mid-size corpora and develop viable solutions for some well-defined tasks of possibly general interest that allow further extension. A relatively coherent and thematically connected corpus of illustrations to a single literary work seems to be a good starting point for further explorations in digital art history and literary studies. Literary illustrations, by definition, stand in a relationship to the literary work they are supposed to illustrate. A vast part of the corpus will inevitably be of a conventional quality and thereby prevent that proposed solutions do not satisfy the requirements of extensibility and non-ad-hocness.

A written text presents a story world, whose inconspicuous conventional elements do not need to be explicitly described or can even dispense with any explicit mention, because the cognitive abilities of the reader will readily supply them. An illustrator cannot help but making such elements explicit and thereby supplementing and commenting on the text. The initiative is left to the artist, but at the same time there is a steady tradition of iconography that supports his endeavors and ensures that he will be understood. Thus, the relation between text and image and the devices of rendering complex narrative situations by a combination of typical pictorial elements is necessarily a complex one, and yet, the complexity should stay within the limits of a manageable one.

It will be an arduous task to develop the tools that help scholars accustomed to the highly developed and sophisticated methods and standards of the humanities in their daily work. Until that point will be reached, a large amount of elementary work needs to be done. But even on a much lower level, the digital exploration of images (such as literary illustrations) of a comparatively large corpus will offer possibilities that could not otherwise exist and allow scholars to ask and answer new questions. At the same time, the digital capturing of the contents of images by description, analysis and classification should allow a more intense and diverse use of image archives and collections in museums and for many educational purposes.

The present paper is about a tool in support of these goals. It concerns the mining of multimedia data based on a *unified* semantics that allows for the *fusion* of multicode information objects. This is exemplified by text-image relations that are easily established by human beings to a degree still unequalled by any approach to automatic text or image understanding. To meet this and related challenges, a format is needed that enables the representation of multimedia data even across the border of different (e.g. iconic and symbolic) codes using the *same* ontology. In the line with previous work of the project “Illustrations of Goethes Faust”¹,

¹ <https://www.goethehaus-frankfurt.de/sammlungen/digitaler-katalog>.

images are the initial point for our paper. In the context of this project², we created a corpus of 2500 illustrations. The illustrations are segmented [1] in order to relate their subimages to segments of the “Faust” text corpus. In the present paper, we investigate the information content of images. To answer questions about this content by means of large corpora, a computer based solution is necessary. It has to face that based on ever changing research interests, the focus of information to be explored will vary. Therefore, flexibility is an indispensable requirement for the system to be developed. In this paper, we describe the so-called *OWLnotator*, a highly flexible system for the annotation of multimedia corpora of texts and images using OWL-based ontologies. The *OWLnotator* allows for modeling relations of symbolic and iconic signs of various levels of resolution: ranging from the level of elementary constituents to the one of complete texts and images. The *OWLnotator* integrates *TEILex* (a system for interrelating corpus and lexicon data as part of the eHumanities Desktop and based on the *Text Encoding Initiative*³) with the expressiveness of OWL⁴-based ontologies in order to meet the first part of our twofold challenge.

The paper is organized as follows: Sect. 1.1 briefly describes related work. Section 2 deals with the scope of the OWLnotator in terms of multimodal, multimedial and multimedia data. In Sect. 3, we describe the image and text corpus underlying the present paper. Section 4 is devoted to the technical description of the OWLnotator, while Sect. 5 contains a brief evaluation of this system. Finally, Sect. 6 gives a conclusion and a prospect on future work.

1.1 Related Work

There is a lot of previous work in the area of virtual research platforms. First of all, the digital image archive *Prometheus* [6], started in 2001, connects a lot of distributed image-databases for research. For the relation-based management and image segmentation, Prometheus uses the tool *Meta-Image* [7], which is based on one of the most used tools, that is, *HyperImage* [16]. *HyperImage* and its follower named *Yenda* [17] are used in the project *Hachiman Digital Handscrolls* [18]. The goal of the project is to present “monumental or moved imageformats” [18] to the research community. The project deals with seven illuminated Japanese cross-roles from the 14th to the 17th century. With the help of *Yenda*, all the functionality of *Hyper-Image* is included and extended with the possibility of semantic annotations for the content of the cross-roles. The tool is very promising and will be in the focus of further investigation as soon as being published in Summer 2015. Other research projects like *CLAROS*⁵ and *TheWissKI Project*⁶ developed effective techniques for information integration and retrieval.

² Founded by LOEWE – *The State Offensive for the Development of Scientific and Economic Excellence* (<https://www.proloewe.de/en>).

³ www.tei-c.org.

⁴ *Web Ontology Language*.

⁵ <https://www.clarosnet.org/XDB/ASP/clarosHome>.

⁶ <https://wiss-ki.eu/node/23>.

2 Aspects of Fusing Multimodal and Multicodal Information Objects

In [10], we briefly described intra- and intermodal relations of textual and pictorial units. In this section, we extend this outline by additionally distinguishing multicodal relations and interpretation relations (see Table 1: from the point of view of the sender or receiver of a sign aggregate we speak of *multimodal* relations if producing/processing this aggregate involves different sensory channels (cf. Weidenmann [23]).

Table 1. Intra- and inter- as well as mono- and multimodal sign relations as object of interpretation processes using the OWLnotator.

| → | text (Segment) | image (Segment) | code unit | interpretation unit |
|------------------------|---|--|---|------------------------------------|
| text (segment) | intratextual / intertextual relation | intermedial / multimodal relation | intracodal / intercodal abstraction | instantiation / fusion relation |
| image (segment) | intermedial / multimodal relation | intrapictorial / interpictorial relation | intracodal / intercodal abstraction | instantiation / fusion relation |
| code unit | intracodal / intercodal manifestation | intracodal / intercodal manifestation | intracodal / intercodal relation | instantiation / fusion relation |
| interpretation unit | abstraction / fusion relation | abstraction / fusion relation | abstraction / fusion relation | ontological relation |

From the point of view of the underlying sign systems we speak of *multicodal* relations if producing/processing the aggregate involves different (e.g., linguistic or pictorial) codes [23]. A linguistic example of multicodality is given by multilingual descriptions of the same image – making use of the code of different languages (e.g., terminological ontologies). From the point of view of the sign vehicles and media involved, we, finally, speak of *multimedia* relations if transferring the aggregate involves different (e.g., linguistic or pictorial) media or multicodal signs [23]. Generally speaking, we speak of multimedia, -modal or -codal signs when referring to one of these views. Whatever the complexity of a sign is along these distinctions, scholars need to present them in a simplified and unified manner that makes operations on the resulting representations manageable. An analog to this requirement comes from cognition in terms of *fusion*.

That is, we extend our triadic distinction in order to account for situated, cognitive processes of fusion that result in multicode, multimodal sign representations.⁷ The notion of information fusion is applied here to interpretation processes in the humanities where the same aggregate is object of ongoing interpretation processes contextualized by different goals, traditions, schools etc. Any such process involves a mapping of multimodal/multicode signs onto the same (monocode) interpretation language. The representational underpinning of this interpretation-related “interlingua” is the object of the OWLnotator: it maps from multiple codes and modes using the same unifying format. Its cognitive correlate guarantees the ease by which human beings can switch, for example, between images, diagrams, texts in multimodal documents [2, 3, 15] to manifest the same concepts. The distinction we make here is between long-term codes (distributed over a corresponding language community) and short-term interpretations based on these codes by means of fusing their manifestations.

What is at stake here is the possibility to account for a wide range of relations on the level of metadata-related descriptions as well as on the level of the form and meaning of signs [11]. Think, for example, of rhetorical relations interrelating images and texts on the level of their (pragmatic) meanings [21]. Whatever the signature of such a relation is (dyadic or polyadic, types of its arguments etc.), a tool like the OWLnotator has to face the openness of their inventory and of the structures being representable by means of them. That is, a kind of expressiveness is required that makes it prohibitive to pre-establish ontologies of text-image relations. Rather, this establishment has to be delegated to the interpreter (humanities scholar) in a way that the OWLnotator guarantees applicability of the resulting ontology for annotating any sign aggregate within the eHumanities Desktop by analogy to the openness and flexibility of processes of cognitive fusion. Moreover, by analogy to the hermeneutic circle [20] the OWLnotator has to additionally account for situations in which scholars repeatedly change their interpretation models in the light of ongoing interpretation processes. Meeting these two requirements (expressiveness and extensibility) is exactly the task of the OWLnotator to be described in the following sections.

3 Description of the Corpus

In this section, we briefly describe the corpus by means of which the OWLnotator is evaluated.

As argued in Sect. 1, literary illustrations are a worthwhile subject of research for digital art history on grounds of both their inherent art-historical interest and their suitability for digital exploration. The illustrations of Goethe’s “Faust” drama are a case in point. The publication of the first part of Goethe’s “Faust” in 1808 was a major event in the literary history of the 19th century. To be sure, his youthful “Sorrows of Young Werther” won him international fame and had a lasting impact on the sentimentalist and early romanticist tendencies at the end of the 18th and the beginning of the 19th centuries all over the continent (Fig. 1).

⁷ Analogously, *fusion* denotes the process of producing/manifesting such signs (cf. [19]).



Fig. 1. Eugène Delacroix, Gretchen in the Cathedral, scene from “Faust”, 1827, Freies Deutsches Hochstift - Frankfurter Goethe Museum, Inv-Nr. III-13280/001c.

But only with “Faust” did Goethe emerge as a leading figure and universally recognized authority in the world of letters. With Goethe’s drama, the “Faust” legend of the 16th century became one of the principal narratives of modernity, its possibilities and dangers. In particular, its Europe wide reception was an inspiration for artists to dedicate their work to the illustration of the “Faust” drama which thereby became the most frequently illustrated subject of Goethe’s oeuvre and of world literature itself. The spectrum of artistic approaches that have been tried is very broad, ranging from a large amount of conventional illustrations of figures and objects mentioned in the drama, to many highly original interpretations that try to convey the somber and uncanny atmosphere of its scenes in another medium. The existing body of illustrations shows the different attitudes artists have had towards their subject: from renderings that try to keep as close to the text as possible to imaginative explorations of the possibilities of the pictorial medium which are barely hinted at in the corresponding text they are supposed to illustrate.

As a dramatic text necessarily lacks the wealth of descriptive detail other types of literary writing – especially narrative forms – have, there is ample latitude for individual solutions which might, in turn, be the starting point of pictorial traditions of their own.

Because of its wide variety of possibilities within a thematically defined corpus, the tradition of “Faust” illustrations is a rich source for the research into external and internal relations of images (relations between two or more images and relations between details within one image) and text-image-relations. The study of the diverse forms of artistic reception, appropriation and interpretation of Goethe’s “Faust” has been the subject of numerous studies and research

activities, mostly focused on individual artists or particular traditions; a new, comprehensive major treatment that follows earlier works such as Boehn 1924 [4], Wegner 1962 [22] or Giesen 1998 [8] is an important desideratum. The largest collection of illustrations to Goethe's "Faust" (of all type, especially drawings and prints) is held by the Goethe Museum in Frankfurt. The collection currently consists of about 2500 drawings and prints of "Faust" from the early 19th century to the present. The collection is completely digitized and fully accessible online. The images are provided with the necessary descriptive metadata which are represented on a first level in a metadata schema for museum objects (LIDO [12]) and on the second level in an ontology based on the CIDOC CRM.

Any form of in-depth research into the content of the corpus of "Faust" images will need a basis of systematically stored semantic information about the depicted objects and figures and their place within the dramatic action. Since there is no way of exhausting the material by some all-encompassing description, there has to be a decision of where to begin. First endeavors should start from elementary questions known to be of relevance to art historical and literary studies. One starting point might be the gestures in pictures, as they may relate as well to the stage directions in the dramatic text as to iconographic traditions of representing human conduct. At the same time, the descriptions of gestures are likely to be of interest to many other projects. For the semantic representation of gestures an ontology has to be devised that allows to capture the relevant information and complies both with the demands of precision, generalizability and of further extensibility to different corpora. This is a typical task to be performed with the help of the OWLnotator.

4 The OWLnotator

The OWLnotator is a highly flexible annotation system for annotating inter- and intramedial relations in multimedia corpora. As an annotation module it is part of the eHumanities Desktop [9], a browser-based, platform-independent research environment for the support of collaborative research in the digital humanities. The eHumanities Desktop contains a wide range of tools for managing, analyzing and sharing resources based on a scalable concept of access permissions. Based on ontologies, written in OWL⁸, the *OWLnotator* is able to typecast elements with *OWL-Classes* and can annotate every resource of the eHumanities Desktop (words, texts and their segments as well as images and their segments). With the typecast, the elements become *OWL-Objects* with all the properties related to the base class. New properties can be added by means of drag and drop operations. While adding new properties, the *OWLnotator* assists researchers with an ontology-based pre-selection of available *OWL-properties* and the existing objects for selection (if there are no literals). Figure 2 shows the interface of the *OWLnotator*. In the center, the resource to be annotated is shown. Based on the media-type, the center-area is displayed appropriately. Pictures are presented as

⁸ <http://www.w3.org/TR/owl-features/>.

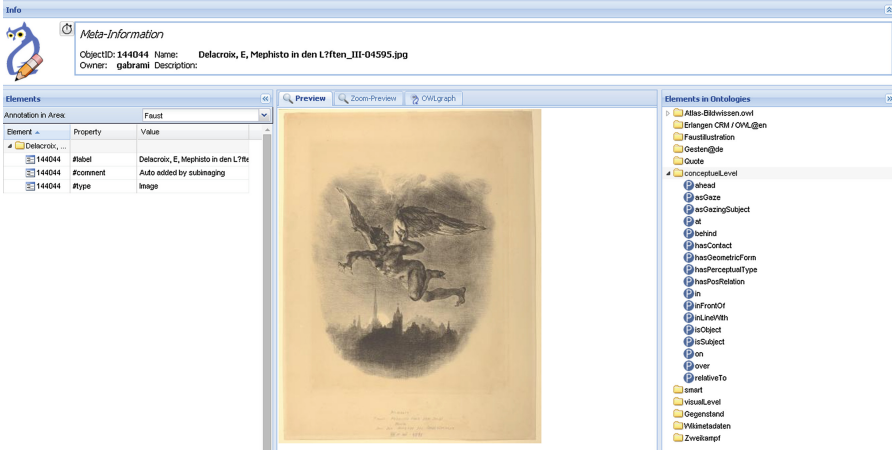


Fig. 2. The interface of the OWLnotator.

they are; text encoded in TEI P5 is displayed by means of its logical document structure and also in plain-text. The left side displays the current annotation of the element and on the right side the available ontologies are shown associated with their properties available for them.

As part of the eHumanities Desktop, which provides an environment for collaborative research, the *OWLnotator* can be used with a flexible system of access permissions. The *OWLnotator* is based on the free and open source Java framework of the Apache project for building Semantic Web and Linked Data applications, called *Jena*.⁹ We use Jena because of its built-in reasoning-mechanisms and its flexible import and export features. Jena operates on a Quad-Store containing triples of subjects, predicates and objects embedded in the fourth dimension: *the Model*. Based on this, we create new resources within the eHumanities Desktop, *Annotation Locations* and *Annotation Areas*. *Annotation Locations* represent a Jena-TDB-Store¹⁰ and contain the *Annotation Areas*. *Annotation Areas* are the representations of the *Model* mentioned above. Researchers can get access permissions on *Annotation Locations* as well as on *Annotation Areas*. The distinction of *Annotation Locations* and *Annotation Areas* allows for a detailed access management where the user can decide which other users or user groups¹¹ have access on *their Annotation Areas* within the eHumanities Desktop.

At the same time we can handle a clear division between *Annotation Areas* used for definitions and *Annotation Areas* used for annotations. Depending on the access permissions (see Fig. 3) they can add, modify or remove annotations in *Annotation Areas* or they can manage the access permissions directly. Every user, registered in the eHumanities Desktop, owns his own area, called

⁹ <https://www.jena.apache.org>.

¹⁰ TDB is a RDF-Store.

¹¹ To which users can belong.

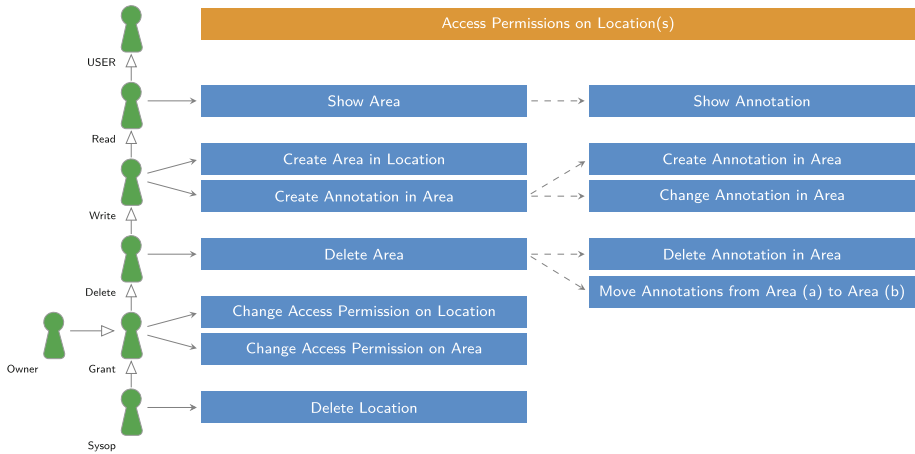


Fig. 3. Access Permissions on *Annotation Locations*.

Home Area. It is very easy to add new ontologies into the *OWLnotator*: Researchers only have to create a new *Annotation Area* and upload, from remote or local resources, the ontology into it. The ontology has to be valid. Its validity may be provided with the help of some open source tools like Protégé¹². In the same way, it is possible to change views on annotations by uploading a new ontology which interprets the current annotations by means of other features, classes etc.

5 A Snapshot of Data Managed by the OWLnotator

The OWLnotator is used by several annotation projects in the digital humanities. This includes the *Illustrations of Goethe's Faust* and the Project on *Historical Knowledge of Pictures* [13, 14]. Further, the OWLnotator contains several image and text-related ontologies as well as more than numprint758464¹³ (See Fig. 4, left) annotations in 174 *Annotation Areas* of 7 *Annotation Locations* (see Fig. 4, right) – this amount of data increases every day. The time for creating one annotation is below one minute

Researchers can work independent of each other if their topics do not share any information objects. All the annotations are done based on the same software solution. Users can share their annotation results. With the help of a time measurement tool, which allows user-related measurements of annotation times, it is also possible to create annotations.

¹² <https://www.protege.stanford.edu>.

¹³ Counted at 02/18/2015.

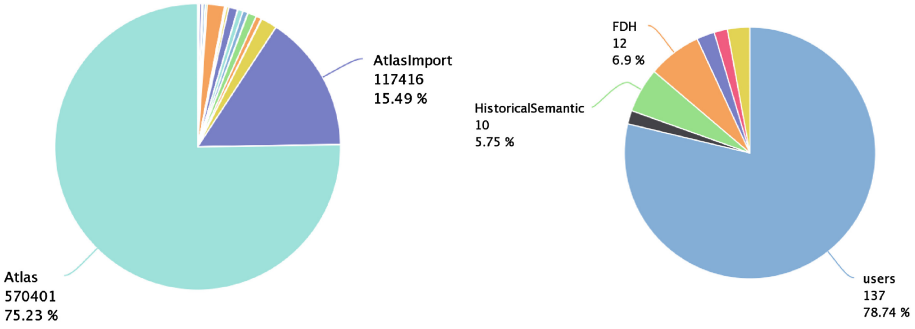


Fig. 4. Left: annotations in *Annotation Areas*. Right: *Annotation Areas* managed by the OWLnotator.

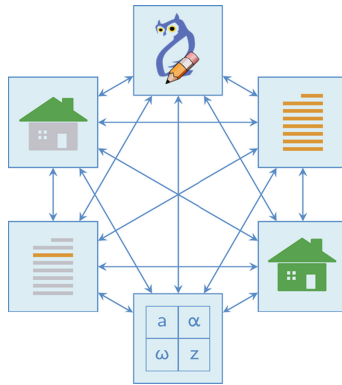


Fig. 5. The OWLnotator connects different types of media.

6 Conclusion and Future Work

With the *OWLnotator* we developed a tool for annotating multimedia resources as shown in Fig. 5. As part of the eHumanities Desktop, the OWLnotator allows for using all services of this platform for collaborative research, especially its tool for managing access permissions. After creating and uploading an ontology with the OWLnotator, the researchers can directly start with annotating art work.

The next development step is to make the OWLnotator a stand alone *web-service* to allow for adding it into other software projects in the digital humanities.

References

1. Abrami, G., Freiberg, M., Warner, P.: Managing and annotating historical multimodal corpora with the eHumanities desktop - an outline of the current state of the LOEWE project illustrations of Goethe's Faust. In: Proceedings of the Historical Corpora Conference, 6–9 December 2012, Frankfurt (2012)

2. Bateman, J.A.: Text and Image: A Critical Introduction to the Visual/Verbal Divide. Taylor & Francis, NewYork (2014)
3. Bateman, J.A., Kamps, T., Kleinz, J., Reichenberger, K.: Towards constructive text, diagram, and layout generation for information presentation. *Comput. Linguist.* **27**(3), 409–449 (2001)
4. von Boehn, M.: Faust und die Kunst. In: Goethe, Johann Wolfgang: Faust. Hundertjahrs-Ausgabe, pp. 3–221. Askanischer Verlag, Berlin (1924)
5. CIDOC CRM Special Interest Group (SIG): Definition of the CIDOC Conceptual Reference Model, 5.0.4 edn. (2011). <http://www.cidoc-crm.org/docs/cidoc-crm-version-5.0.4.pdf>
6. Dieckmann, L.: Prometheus: the distributed digital image archive for research and education. In: *L’Art et la Mesure: Histoire de l’art et méthodes quantitatives*, sous la direction de Béatrice Joyeux-Prunel, avec la collaboration de Luc Sigalo Santos, pp. 141–151. Paris (2010)
7. Dieckmann, L., Kliemann, A., Warnke, M.: Meta-Image: forschungsumgebung für den bilddiskurs in der kunstgeschichte. *cms-j. Comput.- und Medienservice* **35**, 11–17 (2012)
8. Giesen, S.: Den Faust, dächt’ ich, gäben wir ohne Holzschnitte und Bildwerke: Goethes “Faust” in der europäischen Kunst des 19. Jahrhunderts. Ph.D. thesis, Technische Hochschule Aachen, Aachen (1998)
9. Gleim, R., Mehler, A., Ernst, A.: SOA implementation of the eHumanities Desktop. In: *Proceedings of the Workshop on Service-oriented Architectures (SOAs) for the Humanities: Solutions and Impacts, Digital Humanities 2012*, Hamburg, Germany (2012)
10. Gleim, R., Warner, P., Mehler, A.: eHumanities Desktop - an architecture for flexible annotation in iconographic research. In: *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST 2010)*, Valencia, 7–10 April 2010
11. Hollink, L., Schreiber, A.T., Wielinga, B.J., Worring, M.: Classification of user image descriptions. *Int. J. Hum.-Comput. Stud.* **61**(5), 601–626 (2004)
12. ICOM-CIDOC Working Group Data Harvesting and Interchange: LIDO - Lightweight Information Describing Objects, 1.0 edn. (2010). <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>
13. Jussen, B. (ed.): Atlas des Historischen Bildwissens 1: Liebig. Digitale Bibliothek, Berlin (2009)
14. Jussen, B. (ed.): Atlas des Historischen Bildwissens 2: Reklamesammelbilder. Digitale Bibliothek, Berlin (2009)
15. Kress, G., Leeuwen, T.: *Multimodal Discourse*. Arnold, London (2001)
16. Kuper, H.G., Loebel, J.M.: Hyperimage: of layers, labels and links. In: *Proceedings of RENEW the 5th edition of the International Conference on the Histories of Media Art, Science and Technology*. Riga (2014)
17. Kuper, H.G., Loebel, J.M.: Yenda - Picture Knowledge, Open-Source semantische virtuelle Forschungsumgebung (2015). <http://yenda.tools>
18. Loebel, J.M., Kuper, H.G., Arnold, M., Decker, E.: Hachiman digital handscrolls semantische anreicherung mit hyperimage und yenda. In: Bienert, A., Hemsley, J., Santos, P. (eds.) *Elektronische Medien & Kunst, Kultur und Historie*, pp. 262–267. Konferenzband, Berlin (2014)
19. Lücking, A., Pfeiffer, T.: Framing multimodal technical communication. with focal points in speech-gesture-integration and gaze recognition. In: Mehler, A., Romary, L. (eds.) *Handbook of Technical Communication, Handbooks of Applied Linguistics*, vol. 8, chap. 18, pp. 591–644. De Gruyter Mouton, Berlin and Boston (2012)

20. Poser, H.: *Wissenschaftstheorie. Eine philosophische Einführung*, 2nd edn. Reclam, Stuttgart (2012)
21. Taboada, M., Habel, C.: Rhetorical relations in multimodal documents. *Discourse Stud.* **15**(1), 59–85 (2013)
22. Wegner, W.: *Die Faustdarstellung vom 16. Jahrhundert bis zur Gegenwart*. Erasmus Buchhandlung, Amsterdam (1962)
23. Weidenmann, B.: Multicodierung und Multimodalität im Lernprozess. In: Issing, L.J., Klimsa, P. (eds.) *Information und Lernen mit Multimedia*, pp. 45–62. Beltz, Weinheim (1997)