Data-rich Web Annotations: Embedding Datasets to Link Complex Metaphor Analyses With Their Textual Basis

Philipp Tögel[1,*,†], Henning Gebhard[2,*,†], Frederik Elwert[2,†], Stefanie Dipper[2], Makar Fedorov[2], Vandana Jha[1], Volkhard Krech[2] and Danah Tonne[1]

[1]Karlsruhe Institue of Technology, Germany
[2]Ruhr University Bochum, Germany

Abstract

Annotating is a central activity with a long history in the humanities. For the purpose of digital annotations, the Web Annotation Data Model (WADM) is an established W3C standard that enables data sharing and is supported by a wide array of applications. Storing simple annotations is quite easy, but storing complex data is difficult. We propose a generic extension mechanism for the WADM that allows storing structured data inside the body of a Web Annotation. In contrast to previous research, our proposal uses the base WADM without custom extensions (except for the embedded data themselves), and thus facilitates data sharing. As a use case, we show how a domain ontology is used to model structured information about metaphors in religious texts, and how we apply our approach to store the information in data-rich annotations that can be used for queries that support comparative research across languages.

Keywords

Web Annotation, Metaphor Studies, Linked Open Data

1. Introduction

Annotating texts or images has a long history in the humanities and, unsurprisingly, it is a central activity in the digital humanities (DH) as well. The Web Annotation Data Model (WADM) [1] is an established W3C standard for digital annotations, building upon widely supported semantic web technologies like the Resource Description Framework (RDF) and the JavaScript Object Notation for Linked Data (JSON-LD). Web Annotations have become a standard for representing stand-off annotations within RDF in a FAIR (Findable, Accessible, Interoperable, and Reusable) and Linked Open Data (LOD)-compliant manner across digital humanities and other fields [2, 3, 4]. A Web Annotation is essentially a RDF resource, which binds together two other resources, asserting that one of them (the body) is in some way or form about the other (the target). As such, it provides a very generic, flexible, and extensible building block for modeling a wide array of use cases, from simple classifications to complex analyses. This makes it an attractive approach for the Collaborative Research Center (CRC) 1475 "Metaphors of Religion"[1].

CRC 1475 addresses annotating the occurrence of metaphors in texts from a variety of cultures and languages throughout time and history. While some parts of the research results (like the usage of metaphorical language in a given text segment) are easily expressed as annotations by assigning simple labels to a span of text, others are not. The results of the analysis of a metaphor require a rich structure beyond text labels and a more sophisticated data model, which facilitates machine readability and elaborate queries on the research data (for a more detailed explanation of our data model, see section 3).

SemDH 2025: Second International Workshop of Semantic Digital Humanities. Co-located with ESWC 2025, June 02, 2025, Portoroz, Slovenia.

*Corresponding author.
†These authors contributed equally.
philipp.toegel@kit.edu (P. Tögel); henning.gebhard@rub.de (H. Gebhard)
0000-0002-1791-0268 (P. Tögel); 0000-0001-7853-1950 (H. Gebhard); 0000-0001-9149-9377 (F. Elwert); 0000-0003-4357-9078 (S. Dipper); 0000-0002-6629-0042 (V. Jha); 0000-0002-9013-4461 (V. Krech); 0000-0001-6296-7282 (D. Tonne)

[1]CRC 1475 "Metaphors of Religion" consists of researchers from the Ruhr University Bochum and the Karlsruhe Institute of Technology in 14 scientific subprojects.

Therefore, we need to integrate our domain-specific data model with the WADM in a way that retains its strengths, i.e., its interoperability and generality. This is a common challenge in research projects which utilize Web Annotations, and as a matter of fact a number of projects from different fields have proposed their specific solutions in recent years (see section 2). Although the details of their approaches differ, they all extend the WADM by closely intertwining it with their domain-specific ontologies. While integration is necessary as it creates the basis for relevant queries and research, such tight coupling carries the risk that the resulting models are only applicable to the particularities of their domain, effectively forcing each project to come up with their own solution. Starting from comparative metaphor research in religious texts as a use case, this paper pursues two aims: first, we propose a replicable way to interface Web Annotations with complex domain models. This integrates the data for easy querying without tight coupling, thus strengthening the reusability and adaptability of both the data and the approach itself. Second, we show how a domain ontology is used to model structured information about metaphors in religious texts, allowing for data-rich annotations which in turn enable comparative research across languages.

2. Related work

Extending Web Annotations has sparked the interest of scholars across disciplinary boundaries. As Web Annotations are suitable for annotating any resource (`Target`) identified by an Internationalized Resource Identifier (IRI), they have been used to annotate a variety of media types from a wide range of time periods: from digitized historical data — in the form of historical trace data [5], scanned historical manuscripts [6] and 3D images of clay tablets [7, 8] — to born digital contemporary material — like the results of multimedia analysis on digital content [9], interactive system design and development [4], multimedia content for architects, designers and video game creators [10], and fake news [11]. Furthermore, they also have been used to store the results of object detection and layout analysis of images [12, 6], as well as for working with musicological objects [13, 14, 15].

The diversity of research is reflected in the amount of different approaches taken to extend the WADM. A common strategy among the listed projects is to replace the `Annotation` as a whole or some of its components with a custom subclass. This can be motivated by the wish to restrict what can or cannot be used as the value of its properties, be it its `Body`, its `Target`, or some metadata. Alternatively, it can also be used to include additional relations or properties. Some projects use a combination of both, such as the V4Ann model [9]. This model defines an `Annotation` subclass which closely resembles a regular Web Annotation, but which replaces the standard relations to its `Body` and `Target` with custom ones, to enforce that only certain bespoke classes, `v4d:View` and `v4d:MediaType` respectively, are valid value ranges of these relations. These in turn have additional relations on top of those which are normally present in annotation bodies and targets. In some cases, an extension can be as simple as adding metadata fields to the `Annotation` or its `Body`. This encompasses various labels, provenance information of non-digital objects, or key-value pairs specifying their physical dimensions.

Another practice which has frequently been employed by previous projects can be summarized as establishing new semantics inside the existing formal structure of Web Annotations. An `Annotation`'s `motivation` and a `Body`'s `purpose` properties have recurrently been assigned values which carry special meaning only in the context of the particular project. While seemingly an unintrusive adjustment, as it does not change the data model per se, this can also make it hard to understand for potential users. This can be mitigated by extending the `motivation` values which are available in the WADM with another controlled vocabulary that better expresses the intended meaning. This is where Web Annotations, as being built on top of RDF principles, really shine: it is very easy to include other vocabularies and ontologies as additional `contexts` to integrate external knowledge graphs, allowing the reuse of domain-specific data models which are in itself not concerned with annotation practices.

As these examples show, the WADM is indeed very extensible. However, the multitude of options and the complexity it entails make it non-trivial to come up with solutions which are both suitable for the individual project's use case and understandable for others. This is especially true for projects which are not primarily concerned with data modeling but with the querying capabilities that come with it. Some only want to have something which is sufficient for their use case, as is made apparent by the fact that some of the extensions, like adding otherwise unspecified properties, are not formally valid. Nonetheless, while coupling the WADM with domain-specific models is powerful and seems convenient, it hinders shared modeling approaches and shared software solutions. This applies not only when it comes to potential external users. It also affects one's own work, especially when there are different kinds of annotations involved, or when the data requirements evolve over time. It is often hard enough already to model one's actual research domain without the intricacies of Web Annotations. That is why we are proposing an approach where the Web Annotation Data Model is left largely untouched and different domain-specific ontologies can be plugged in as needed, while still retaining a tight enough integration to make it easy to harness the rich querying capabilities of linked data.

3. Including structured data in Web Annotations

3.1. Use case: Metaphor analysis

CRC 1475 annotates metaphor in texts from a variety of cultures and languages throughout time and history. Since "metaphor" means different things in different fields, a clear operationalization of the concept was fundamental for the project. We annotate metaphors on two levels: First, we identify individual Metaphor Related Words (in short MRW) according to the "Metaphor Identification Procedure Vrije Universiteit" (MIPVU) guidelines [16, 17]. Then, we apply Steen's "Five Steps" [18] to transparently go from individual, figuratively used words to conceptual cross-domain mappings that inform the concrete linguistic expression. As a standardization measure, we append an additional step, where each part of an extracted conceptual mapping is linked to a conceptual thesaurus.[2] While some parts of the research results, i.e., whether a given word is a Metaphor Related Word according to MIPVU or not, are easily expressed as annotations by assigning simple labels to a span of text, others are not. The conceptual mappings which constitute the metaphor in a text form a rich structure beyond text labels. In addition, parts of the mappings are frequently not given explicitly in the text and therefore have no clear annotation target. All of this requires a more sophisticated data model which facilitates machine readability and elaborate queries on the research data. To achieve that, we need to extend the WADM in a way which retains its strengths, i.e., its interoperability and generality, while giving room for (potentially evolving) data modeling requirements that come with complex research data.

3.2. General approach

Our decision to use the WADM, which models stand-off annotations, is rooted in the need for flexibility and scalability in handling annotations. Stand-off annotations are a long-established practice [3] to decouple annotations from the annotated resources, in our case XML documents compliant to the Text Encoding Initiatives-guidelines [20]. This separation allows the resource to evolve independently of its annotations and simplifies the use of annotation targets overlapping each other. Furthermore, it fosters collaboration as users can annotate the same passage of text without interfering with the annotations of others; instead, each creates a separate annotation resource. However, this approach requires robust synchronization to ensure annotations align with the resource as updates occur.

By representing annotations in JSON-LD format, the system achieves both human readability and machine interoperability, fostering a more accessible and efficient workflow for researchers inside the CRC. Using a well documented W3C Recommendation (the WADM) eases the reuse and allows for better interoperability of our data, which is central for FAIR data. Furthermore, the WADM allows us to contribute to the semantic web and enables the inclusion of our data in knowledge graphs.

Both our own research as well as the existing extensions of the WADM illustrate the need to use complex, highly structured and well-defined data as the body of a Web Annotation. Many projects tackle this by establishing their own conventions inside of the WADM or by extending the WADM itself. We

[2]For a detailed elaboration of the CRC's understanding of "metaphor" and the used methodology see [19].

propose instead to use existing idioms of the WADM, namely a combination of `SpecificResources` and `type: "Dataset"`, to incorporate RDF data into Web Annotations while sticking to fully generic Web Annotations. By doing this, any modifications are limited to the place that is inherently concerned with user-provided data, namely the body. The annotation logic which constitutes the core functionality of the WADM remains unaffected.

Each annotation is treated as a micro-publication, attributing authorship to the creators. As we decouple our metaphor analysis data model from the WADM, we can use the Web Annotation Protocol Server [21] as our storage solution, which was developed as a generic Web Annotation backend and offers advanced querying capabilities.[3] Through SPARQL queries, researchers can extract, aggregate, and analyze data efficiently, making the annotation system not just a repository, but a tool for discovery and analysis. As the WADM covers how the analyses refer back to specific parts of a text, we can query not only analysis data itself but also its textual source (see 3.5). Furthermore, this also allows us to easily interface with other established vocabularies in the LOD ecosystem, like Simple Knowledge Organisation System (SKOS), which we use to link cross-domain mappings to entries in a conceptual thesaurus for normalization and retrieval purposes. Finally, by specifying our annotation bodies as `Dataset` and including our ontology as additional `context`, we can provide additional metadata and explanations on how to make use of it directly inside of the Web Annotation, similar to how one would specify, e.g., a file type, media type, and language when using an audio file on the web as annotation body (see [1]).

3.3. Body model

As mentioned in section 3.2, we use `SpecificResources` for our annotation bodies, as is common in the WADM.[4] To indicate their special role, the WADM provides us with two mechanisms. Firstly, the use of `oa:hasPurpose`. For its values we reuse the TaDiRAH taxonomy [25] (e.g. `tadirah:analyzing` in case of the metaphor analysis Web Annotations). Furthermore we specify the type of the `SpecificResource`'s `source` as `dctypes:Dataset`. The source's value is then either a `MetaphorAnalysis` or a `MRW`, as defined by our own ontology[5] (see figure 1a "Web Annotation containing a metaphor analysis"). Our domain-specific modeling is entirely encapsulated inside these entities. In the case of `MRW`s this includes their type[6], with its possible values determined by yet another bespoke controlled vocabulary. For `MetaphorAnalyses`, this encompasses their cross-domain mappings, including references to corresponding `skos:Concepts` from our conceptual thesaurus, and also links to the `MRW`s which are integral for each analysis (see figure 1b "Excerpt of ontology modeling a metaphor analysis" and listing 1 "Simplified embedded `MetaphorAnalysis` resource"). As each `MRW` and `MetaphorAnalysis` resource has its own `IRI`, we can directly refer to the actual resource as opposed to the Web Annotation in which it is embedded. All of our own vocabularies and ontology are going to be published under persistent URLs and can be easily integrated into the Web Annotations via additional `contexts`, both for machine-readability and as documentation for humans readers.

3.4. Target model

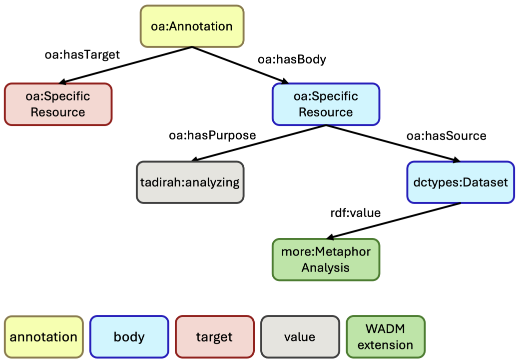
CRC 1475 annotates the occurrence of metaphors in texts. These texts have been converted to XML files compliant to the TEI-guidelines [20] and are tokenized. Each token is assigned an `xml:id`, which we use

[3]We use the base-repo [22], wap-server [21] and SKOSMOS [23], but other projects can use any database for storing XML files, and any server implementing the Web Annotation Protocol and the WADM to store annotations, respectively. The tool for creating annotations (amongst other tools) is still in development and will be published at the end of the project. For an overview on all infrastructural components, see [24].

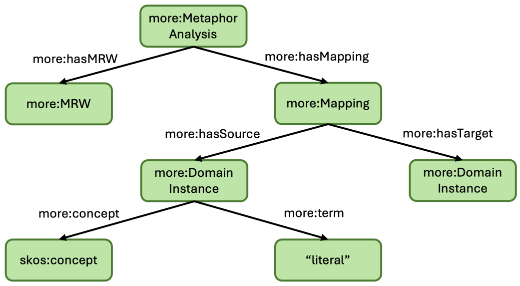
[4]This paper only contains reduced examples; full examples can be found at: https://doi.org/10.5281/zenodo.15235098. The complete dataset is not yet publicly available. All infrastructure components used by the CRC and the stored data (texts, annotations, `skos:Concepts` etc.) will be published at the end of the project.

[5]An article focusing on the ontology is currently in preparation.

[6]MIPVU differentiates between different kinds of Metaphor Related Words, depending on their concrete linguistic manifestation. For details, see [16].



(a) Web Annotation embedding a metaphor analysis



(b) Excerpt of ontology modeling a metaphor analysis and its relationships to Metaphor Related Words and cross domain mappings

Figure 1: Generic WADM and domain-specific metaphor analysis data model

```
"body": [
  {
    "type": "SpecificResource",
    "purpose": "tadirah:analyzing",
    "source": {
      "id": "moredata:6fabb347123",
      "type": "Dataset",
      "value": {
        "id": "moredata:6fabb347123MetaphorAnalysis",
        "type": "more:MetaphorAnalysis",
        "more:hasMapping": [
          {
            "more:hasSource": {
              "type": "more:DomainInstance",
              "id": "moredata:dbebd22e27dc1",
              "more:hasTerm": "shepherd",
              "more:hasConcept": ["https://example.org/concepts/345"]
            },
            "more:hasTarget": {
              "type": "more:DomainInstance",
              "id": "moredata:dbebd22e27dc2",
              "more:hasTerm": "guide",
              "more:hasConcept": ["https://example.org/concepts/567"]
            }
          }
        ],
        "more:hasMRW": ["moredata:6fabb347123MRW"]
      }
    }
  }
]
```

Listing 1: Simplified embedded `MetaphorAnalysis` resource

in an `XPathSelector`. These `xml:ids` serve as a stable point of reference, which will not be affected by unrelated changes to the document. However, we do provide a `TextQuoteSelector` as a fallback option, if a consuming application can not deal with XPath. Furthermore, it makes it easy to include the text selected by a user in the results of queries, as the exact value of the `TextQuoteSelector` is stored in the triple store as well (see listing 2 "Text selection"). We are aware of the danger that changes

```
"target": [
  {
    "selector": {
      "type": "XPathSelector",
      "value": "id(\"w.5\")"
    },
    "source": "https://example.org/texts/Psalms.xml"
  },
  {
    "selector": {
      "type": "TextQuoteSelector",
      "exact": "shepherd",
      "prefix": "The LORD is my ",
      "suffix": "; I shall not"
    },
    "source": "https://example.org/texts/Psalms.xml"
  }
]
```

Listing 2: Text selection

in the document can lead to inconsistency with the `TextQuoteSelector`, and therefore recommend resolving the `XPathSelector` to get the annotated text.

Thanks to the cardinality of bodies and targets of the WADM, where "each Body is considered to be equally related to each Target individually" [1], marking the "same" metaphor in multiple text "manifestations" is quite easy. When different so-called "manifestations" (transcription, transliteration etc.) of a text are available and stored in individual files, the span of text containing the same passage can be selected and included in the annotation by just adding another `target`.

3.5. Extensive query possibilities

Despite our results being cleanly encapsulated in the body as a `SpecificResource`, they are still fully included in their corresponding Web Annotations, so we can leverage the rich capabilities of SPARQL to make complex queries. Not only can we retrieve the "content" of the metaphorical mappings, which are the most important part of the analysis, but also their textual basis and domain-specific metadata. This enables comparative research across languages and religious traditions. To give one example, one can now retrieve all annotations that refer to a given concept as the target domain of a metaphorical mapping, like the concept "Guide" (https://example.org/thesaurus/567) in listing 1. Further information stored inside the annotation like the source domain of the mapping and the text in which the metaphor is located (the target of the annotation) can be retrieved and show that the concept "Guide" is mapped to different concepts in different texts written in various languages (see listing 3 "Example query" and figure 2 "Example query result"). While SPARQL queries are a great tool in itself for users with technical expertise, this also facilitates the creation of sophisticated graphical tools for scholars in the humanities to explore the data in a comparative way and to assist in their research.

```
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX more: <https://example.org/MoRe-SFB1475/ontology#>
PREFIX tadirah: <https://vocabs.dariah.eu/tadirah/>

SELECT DISTINCT ?metaphorAnnotation ?sourceConcept ?text {GRAPH ?g
    {
        ?metaphorResource a more:MetaphorAnalysis ;
            more:hasMapping ?mapping .
        ?mapping more:hasTarget/more:hasConcept "https://example.org/
           concepts/567" .
        ?mapping more:hasSource/more:hasConcept ?sourceConcept .
        ?metaphorAnnotation a oa:Annotation ;
            oa:hasBody ?body ;
            oa:hasTarget/oa:hasSelector ?selector .
        ?body oa:hasPurpose tadirah:analyzing ;
            oa:hasSource/rdf:value ?metaphorResource .
        ?selector rdf:type oa:TextQuoteSelector ;
            oa:exact ?text .
    }
}
```

Listing 3: Example query covering metaphor analysis Web Annotation, linked concept, and original text

metaphorAnnotation, sourceConcept, text
https://example.org/annos/52e, https://example.org/concepts/345, Der HERR ist mein Hirte
https://example.org/annos/d10, https://example.org/concepts/345, The LORD is my shepherd
https://example.org/annos/483, https://example.org/concepts/345, מִזְמוֹר לְדָוִד יְהוָה רֹעִי לֹא אֶחְסָר
https://example.org/annos/5e2, https://example.org/concepts/345, Ḫa-am-mu-ra-pí re-iu-um ni-bi-it dEn-líl a-na-ku
https://example.org/annos/09a, https://example.org/concepts/789, Ḫa-am-mu-ra-pí re-iu-um ni-bi-it dEn-líl a-na-ku
https://example.org/annos/01d, https://example.org/concepts/921, 故從事而道者，道德之

Figure 2: Example query result

4. Discussion

While far from being the standard use case, the WADM supports such complex annotations via the verbatim inclusion of `SpecificResourcess` as annotation bodies. The specification does not explain the inclusion of resources verbatim, but we believe using `rdf:value` to include a `Dataset` to be reasonable.[7] The idea of embedding a graph was part of the Open Annotation draft,[8] which is the predecessor of the WADM. Even the final Open Annotations talk about embedding resources. They propose to assign a `IRI` to the resource, so it can be referenced instead of just being a blank node. We have taken this into account and we refer to the `IRI` of the `MRW` resource to link it to the `MetaphorAnalysis` resource. Furthermore, we now add metadata to the resources like the Open Annotation model suggests [26].[9] Grossner 2019 discussed the inclusion of a `Dataset` in Web Annotations for the Linked Traces format [5] as well.

[7]Also the validator for Web Annotations provided by the Apache Annotator project (see Apache Annotator GitHub page: https://annotator.apache.org/) deems our annotations as valid.
[8]See Open Annotation Data Model Module: Publishing. Community Draft, 08 February 2013: https://web.archive.org/web/20221226053344/http://www.openannotation.org/spec/core/publishing.html#Graphs.
[9]Van de Sompel, a co-author of the Open Annotation data model, presented the idea of using inline and structured bodies in a talk given in 2013 (see his slideset: https://swib.org/swib11/vortraege/swib11-herbert-van-de-sompel-open-annotation.pdf).

4.1. Referring to resources by `IRI` or as External Web Resources

The easiest way to use any kind of resource as the body of a Web Annotation is to just refer to it by its `IRI`. If the resource is available online, we can also specify it as an `External Web Resource` (see [1]), which allows us to include some metadata, like the media format or the language, in addition to the `IRI` of the resource. This is also the general recommendation given by the WADM specification, because it can often be useful for clients to know about the general type of a resource and whether they have a way to appropriately deal with it. Our approach is very similar, and in fact all our `MetaphorAnalysis` and `MRW` resources stand fully on their own and could be stored separately and be referenced by their `IRI`. This would make no difference at all for the data model of our resources, which in itself could be seen as a convenient feature of our approach, as it gives flexibility for adjustments. However, embedding the resources in a `SpecificResource` has proven for us to be the more practical solution, as we can more freely provide even more metadata without changing the WADM itself. It also slightly simplifies storing and querying, as it eliminates the need for federated queries.

4.2. Annotating annotations

Another, more fundamentally different solution would be to model (some of) the semantic interrelations by using annotations of annotations. One could, for example, use a third annotation with a bespoke `purpose` to link a metaphor analysis Web Annotation with one or more MRW Web Annotation which it addresses. We argue, however, that this would wrench the WADM too much into a use case which it was not designed to handle. For instance, it is not even clear which resource would be "in some way about" the other one, which is the intuitive meaning of an annotation given in the WADM specification. What is more critical, it would force us to express the different semantics involved in existing WADM fields like `purpose` or `motivation`. While it would be possible to have a controlled vocabulary of values to express the *intentions* of such linkings, it would not enforce the appropriate constraints on the modeling side, e.g., that a certain kind of relationship can only be between exactly one metaphor analysis Web Annotation and one or more MRW Web Annotations. These kind of constraints are outside of the scope of the WADM and are better handled by a custom ontology. We have found that an abundance of meta annotations and the indirection and overhead it creates for clients makes working with it unnecessarily unwieldy without any gains.

5. Conclusion & Outlook

Incorporating RDF data in Web Annotations in a `SpecificResource` as a `Dataset` allows for inclusion of complex, domain-specific data in an encapsulated way, minimizing its repercussions on the standardized generic data model of Web Annotations. It offers a way to include data modeled according to a domain ontology in the annotation and to store everything in a single triple store. By solely extending the body section, our data model remains compatible with generic Web Annotation software. This approach can be easily adjusted for different domains and adopted by other projects. These projects can then store their data as RDF triples and leverage the facilities of SPARQL to explore their data and answer their questions. If consuming applications are unable to handle the verbatim inclusion of a `Dataset`, they can still extract information from the remaining parts of the Web Annotation. By proposing this approach we hope to improve the quality of "semantic web data" by enhancing its comprehensibility and reusability.

Regarding the use in our project, our model provides a functional way to deal with our requirements. However, a few open questions remain. Until now, we have not come to a conclusion on how to deal with comments on either parts or the entirety of an annotation. Should a comment about the `SpecificResource` reside inside it or this better served by creating a simple additional Web Annotation with a `motivation` of `"commenting"`? While we have generally tried to avoid creating intricate chains of Annotations, this decision is not absolute and should always consider the specific context. If some kind of comment is deemed to be an integral part of an analysis, in the fashion of a critical apparatus, it seems reasonable to include it in the resource itself. If it is an addendum, potentially even by different authors, this evaluation might change significantly. Another point of investigation is the question of additional metadata. Currently, information about the creation date or the language of a text is stored in the TEI documents. Inquiries into the usage of a certain domain in different cultures, or at different points in time, thus require us to extract these information from the document instead of having them readily available via SPARQL queries. One could duplicate these information in the annotation containing the `MetaphorAnalysis`, the `MetaphorAnalysis` itself or another Web Annotation as well, or make use of an intermediate index or query service. As the WADM is not only generic and flexible about the bodies, but also the target, an extension of our ontology is feasible. If scholars wish to annotate the occurrence of metaphors in pictures, the class `MetaphorRelatedWord` might not be sufficient to describe the complexity of images and therefore a class `MetaphorRelatedObject` could be added to the ontology. Conveniently, our chosen approach ensures that these kind of extensions only affect our domain-specific modeling, but not how it interacts with Web Annotations, nor the WADM itself. This enables backward compatibility with existing data as well as with tools.

While we developed our approach in the context of the CRC to enable cross-cultural metaphor research, the above mentioned extensibility makes it applicable to other domains and research contexts as well. By leveraging both the flexibility of this approach and the generic design of the WADM, arbitrarily complex data can be modeled and utilized for queries and tools to assist research. This opens up new ways of exploring and analyzing research data, which is stored in a comprehensible and standard compliant way, remaining compatible to the shared foundations of Web Annotations.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1475 – Project ID 441126958.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] R. Sanderson, P. Ciccarese, B. Young, Web Annotation Data Model. W3C Recommendation 23 February 2017, 2017. URL: https://www.w3.org/TR/annotation-model/.

[2] A. F. Khan, C. Chiarcos, T. Declerck, D. Gifu, E. G.-B. García, J. Gracia, M. Ionov, P. Labropoulou, F. Mambrini, J. P. McCrae, et al., When linguistics meets web technologies. Recent advances in modelling linguistic linked data, Semantic Web 13 (2022) 987–1050. doi:10.3233/SW-222859.

[3] P. Cimiano, C. Chiarcos, J. P. McCrae, J. Gracia, Linguistic Linked Data. Representation, Generation and Applications, Springer, 2020. doi:10.1007/978-3-030-30225-2.

[4] M. Winckler, P. Palanque, J. L. Hak, E. Barboni, O. Nicolas, L. Goncalves, Engineering Annotations: A Generic Framework for Gluing Design Artefacts of Interactive Systems, Proceedings of the ACM on Human-Computer Interaction 6 (2022) 1–36. doi:10.1145/3535063.

[5] K. Grossner, R. Simon, R. Light, R. Scholtz, Linked Traces annotations. v0.2 Draft for comment, 19 October 2019, Technical Report, 2019. URL: https://github.com/LinkedPasts/linked-traces-format.

[6] G. Götzelmann, D. Tonne, Aristoteles annotieren – vom handschriftendigitalisat zur qualitativ-quantitativen annotation, in: C. Hastik, P. Hegel (Eds.), Bilddaten in den digitalen Geisteswissenschaften, Harrassowitz, 2020, pp. 53–66. doi:10.13173/9783447114608.053.

[7] T. Homburg, R. Zwick, H. Mara, K.-C. Bruhn, Annotated 3D-Models of Cuneiform Tablets, Journal of Open Archaeology Data (JOAD) 10 (2022) 4. doi:10.5334/joad.92.

[8] T. Homburg, PaleOrdia: Semantically Describing (Cuneiform) Paleography using Paleographic Linked Open Data, in: Proceedings of the 1st International Workshop of Semantic Digital Human-

ities (SemDH 2024) co-located with the European Semantic Web Conference 2024 (ESWC 2024), 2024. URL: https://ceur-ws.org/Vol-3724/short1.pdf.

[9] G. Meditskos, S. Vrochidis, I. Kompatsiaris, V4Ann: Representation and Interlinking of Atom-Based Annotations of Digital Content, in: Semantic Systems. The Power of AI and Knowledge Graphs: 15th International Conference, SEMANTiCS 2019, September 9–12, 2019, Proceedings 15, Springer, 2019, pp. 124–139. doi:10.1007/978-3-030-33220-4_10.

[10] M. Rousi, G. Meditskos, S. Vrochidis, I. Kompatsiaris, Supporting the discovery and reuse of digital content in creative industries using linked data, in: 2021 IEEE 15th International Conference on Semantic Computing (ICSC), IEEE, 2021, pp. 100–103. doi:10.1109/ICSC50631.2021.00025.

[11] G. Rehm, J. M. Schneider, P. Bourgonje, Automatic and Manual Web Annotations in an Infrastructure to handle Fake News and other Online Media Phenomena, in: Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 2416–2422. URL: https://aclanthology.org/L18-1384.

[12] T. Arnold, L. Tilton, Enriching Historic Photography with Structured Data using Image Region Segmentation, in: Proceedings of the 1st International Workshop on Artificial Intelligence for Historical Image Enrichment and Access, 2020, pp. 1–10. URL: https://aclanthology.org/2020.ai4hi-1.1.pdf.

[13] D. M. Weigl, W. Goebl, A. Hofmann, T. Crawford, F. Zubani, C. C. Liem, A. Porter, Read/Write Digital Libraries for Musicology, in: Proceedings of the 7th International Conference on Digital Libraries for Musicology, 2020, pp. 48–52. doi:10.1145/3424911.3425519.

[14] D. M. Weigl, C. VanderHart, D. Rammler, M. Pescoller, W. Goebl, Listen Here! A Web-native digital musicology environment for machine-assisted close listening, in: Proceedings of the 10th International Conference on Digital Libraries for Musicology, 2023, pp. 109–118. doi:10.1145/3625135.3625144.

[15] D. Lewis, E. Shibata, A. Hankinson, J. Kepper, K. R. Page, L. Rosendahl, M. Saccomano, C. Siegert, Supporting Musicological Investigations With Information Retrieval Tools: An Iterative Approach to Data Collection, in: Proceedings of the 24th International Society for Music Information Retrieval Conference, 2023, pp. 795—-801. doi:10.5281/zenodo.10265407.

[16] G. J. Steen, A. G. Dorst, J. B. Herrmann, A. A. Kaal, T. Krennmayr, T. Pasma, A Method for Linguistic Metaphor Identification: From MIP to MIPVU, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2010. doi:10.1075/celcr.14.

[17] S. Nacey, A. G. Dorst, T. Krennmayr, W. G. Reijnierse (Eds.), Metaphor Identification in Multiple Languages: MIPVU around the world, John Benjamins Publishing Company, Amsterdam, 2019. doi:10.1075/celcr.22.

[18] G. Steen, From Three Dimensions to Five Steps: The Value of Deliberate Metaphor, metaphorik.de (2011) 83–110. URL: https://www.metaphorik.de/de/journal/21/three-dimensions-five-steps-value-deliberate-metaphor.html.

[19] S. Dipper, F. Elwert, Annotating Metaphorical Mappings: An Implementation of Steen's Five Step Method, Metaphor Papers (2024). doi:10.46586/mp.258.

[20] L. Burnard, S. Bauman, TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.7.0, 16. November 2023, TEI Consortium, 2007. URL: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html.

[21] D. Tonne, G. Götzelmann, P. Hegel, M. Krewet, J. Hübner, S. Söring, A. Löffler, M. Hitzker, M. Höfler, T. Schmidt, Ein Web Annotation Protocol Server zur Untersuchung vormoderner Wissensbestände, in: DHd 2019 Multimedial & Multimodal. 6. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2019), 2019. doi:https://doi.org/10.5281/zenodo.4622128.

[22] T. Jejkal, A. Vondrous, A. Kopmann, R. Stotzka, V. Hartmann, KIT Data Manager: The Repository Architecture Enabling Cross-Disciplinary Research, in: C. Jung, A. Streit (Eds.), Large-Scale Data Management and Analysis (LSDMA) - Big Data in Science., Karlsruher Institut für Technologie (KIT), 2014, pp. 9–11. URL: https://publikationen.bibliothek.kit.edu/1000043392/3212685.

[23] O. Suominen, H. Ylikotila, S. Pessala, M. Lappalainen, M. Frosterus, J. Tuominen, T. Baker,

C. Caracciolo, A. Retterath, Publishing SKOS vocabularies with Skosmos, 2015. URL: https://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf, manuscript submitted for review, June 2015.
[24] H. Gebhard, V. Jha, P. Tögel, S. Dipper, F. Elwert, D. Tonne, Metaphors of Religion. Towards a shared infrastructure for metaphor analysis, in: DHd 2023 Open Humanities Open Culture. 9. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2023), 2023. doi:10.5281/zenodo.7715325.
[25] L. Borek, C. Hastik, V. Khramova, K. Illmayer, J. D. Geiger, Information Organization and Access in Digital Humanities: TaDiRAH Revised, Formalized and FAIR, in: Information between Data and Knowledge, volume 74 of *Schriften zur Informationswissenschaft*, Werner Hülsbusch, Glückstadt, 2021, pp. 321–332. URL: https://epub.uni-regensburg.de/44951/.
[26] R. Sanderson, P. Ciccarese, H. Van de Sompel, Designing the W3C open annotation data model, in: Proceedings of the 5th Annual ACM Web Science Conference, 2013, pp. 366–375. doi:10.1145/2464464.2464474.

A. Online Resources

This paper only contains reduced examples; full examples can be found at: https://doi.org/10.5281/zenodo.15235098.