



## King's Research Portal

DOI:

[10.1109/DEST.2010.5610635](https://doi.org/10.1109/DEST.2010.5610635)

*Document Version*

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Blanke, T., Hedges, M., & Palmer, R. (2010). Digital Surrogates for Humanities Research. In *4th IEEE International Conference on Digital Ecosystems and Technologies (DEST), 2010* (pp. 275-280). IEEE. <https://doi.org/10.1109/DEST.2010.5610635>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Digital Surrogates for Humanities research

Tobias Blanke  
King's College London  
tobias.blanke@kcl.ac.uk

Mark Hedges  
King's College London  
mark.hedges@kcl.ac.uk

Richard Palmer  
King's College London  
richard.d.palmer@kcl.ac.uk

## Abstract

*In this paper we present various advanced means of embedding high performance computing in a digital ecosystem to serve Humanities research. We present how high performance computing can be used to create automatic chains of readings and how digital surrogates can help embed these chains in the digital ecosystem. Three different types of surrogates are presented, each catering for a different research need.*

## 1 Introduction

As suggested in [4], High Performance Computing (HPC) is proving to be a useful tool for researchers in the e-Humanities, particularly in document analysis. Under e-Humanities, we understand the advanced use of computing technologies and methodologies for Humanities research [3]. In this paper, we continue our work from [4], where we present a self-organizing text mining agent based on web services that acts on behalf of e-Humanities users to create an 'automatic chain of reading' [5], which we call 'paths'. This path will help the e-Humanities discovery and analysis process of resources.

This paper demonstrates more effective means of integrating such an agent into a digital ecosystem and discusses advanced means to embed high performance computing possibilities into the digital ecosystem for e-Humanities. Digital ecosystems are defined as 'agents-based, loosely coupled, domain-specific [...] communities which offer cost-effective digital services and value-creating activities' [6]. We investigate how they can be built to support e-Research activities, in particular in the domain of e-Humanities. To this end, effective digital surrogates are presented that provide multiple views (in our case stemming from different research usages) on the same object and represent it in such a way, that agents in the ecosystem can process and understand them. We present three alternative surrogates in Section 3, each catering for different needs within the e-Humanities ecosystem.

The paper is organised as follows: In Section 2, we introduce the background and the experiments of the Forging Restful Services for Humanities (FReSH) project<sup>1</sup> at King's College London, which is the basis of our current investigation. In the project, it could be shown how the computational needs for document analysis of an institution in e-Humanities can be served by implementing a Campus Grid, which is connected to the UK National Grid infrastructure. We develop light-weight mechanisms to connect to this Campus Grid, which we discuss in Section 2. In Section 3, we introduce the next steps that generalise the results from FReSH and show different ways of representing digital surrogates within the e-Humanities ecosystem.

## 2 Background

FReSH supports the creation and use of scholarly resources in research. We have developed a prototype for a light-weight e-Humanities ecosystem service that could demonstrate how high performance computing (HPC) for text mining could be integrated in web-resource-oriented research. As discussed in [4], both with FReSH and its predecessor projects, we follow a strict resource-oriented approach, as we consider the results of our textual analysis algorithms as resources, where both processing and data can be found in a resource. All services and all processing results are exposed as ReSTful services that return results in machine-readable formats, so that they can be integrated in web development environments by users with knowledge of web applications but no specialist expertise in HPC.

'Running' new textual analysis algorithms, therefore means to invoke GET on a preprocessed resource. `http://freshpaths.org/sims/coll/TheLeader/item/ldr-1851-08-16-ad02212.xml` would then be the call of a web service to produce similarity relationships between collection elements. At the moment FReSH only implements two types of clustering algorithms: latent semantic indexing and vector spaces [4]. New algorithms can, however, be added at any moment in time as well as new parameters

<sup>1</sup><http://www.fresh.cerch.kcl.ac.uk>

which generally improve such text mining algorithms like, e.g., document length normalisation, etc.

The first aim of the FReSH project is to deliver text mining services for e-Humanities as ReSTful services. The second aim is to expose the results of the text mining processing in a machine-readable format so that they can be integrated in web development environments. We offer a set of alternative ways to create simple machine-to-machine interfaces using digital surrogates, which we describe in detail in Section 3. We use languages that are understood by most web development environments and belong to the standard core of the emerging semantic web.

The experiments we ran for FReSH were mainly to show the research value of our text mining, that it can indeed enhance Humanities research. To this end, we mainly processed the Nineteenth Century Serials Editions (NCSE).<sup>2</sup> The NCSE is a free, online scholarly edition of nineteenth-century periodicals and newspapers. It has been created as a collaboration between Birkbeck, University of London, King's College London, the British Library and Olive Software, and was funded from January 2005 to December 2007 by the Arts and Humanities Research Council in the UK. The NCSE corpus contains circa 430,000 articles that originally appeared in roughly 3,500 issues of six 19th Century periodicals. Published over a span of 84 years, materials within the corpus exist in numbered editions, and include supplements, wrapper materials and visual elements. We processed as an example 10 years of 'The Leader' from 1815-1860 and have made results available through various feeds and read-only ReSTful web services. We could show how easy it would be to integrate references to related material into existing collections. We use feed2js<sup>3</sup> to create automatically HTML pages from ATOM feeds generated by the text mining agent.<sup>4</sup> feed2js automatically produces the JavaScript to transform feeds into HTML, which in turn can be easily integrated in any Humanities website.

In a second experiment, we tried to link authority content from the well-maintained Serving Soldier<sup>5</sup> collections with the community collections of the BBC's People's War<sup>6</sup>. Serving Soldier is maintained by King's College Archives and covers over 20,000 historical archives held in the Liddell Hart Centre for Military Archives. The project's theme is the multi-faced role of the soldier. For the People's War collections, the BBC collected memories from World War Two between June 2003 and January 2006. This archive contains 47,000 stories by veterans and 15,000 images. In the experiment we could show that it is possible to generate links between the authority archive of Serving Soldier

and People's War and therefore enhance the content in the authority archive.

Having shown in the experiments that FReSH services can enhance Humanities research, in the current paper we explain the more advanced digital surrogates we used in FReSH to achieve the kind of links we generated in our experiments. The paper presents the current state of our work that would effectively embed our text mining agent in a larger digital ecosystem beyond individual experiments.

### 3. Digital Surrogates

In [4], we have shown how effective different combinations of standard web technologies such as ReSTful services and ATOM feeds can be used to deliver text mining services to end users in the Humanities. The main focus of our original work has been the delivery of such services to human creators of Humanities websites and to allow these human creators to access the results of text mining as easily as possible. However, machine agents in the worldwide digital ecosystem are often better served by other representations. In this paper, we would like to concentrate on those and how to make the information coming from the FReSH text mining agent better machine-readable. To this end, we develop in this section so-called digital surrogates that will enable the processing of three dominant forces on the world wide web: Google, general text mining agents and finally the emerging agents in the scholarly Linked Data cloud.

We begin with the analysis of Google, as much of the work we have done in [4] can be re-used here to produce data pages from XHTML representations. But, before that we briefly introduce digital surrogates and their traditional importance for Humanities research.

#### 3.1. Traditional View on Digital Surrogates

Digital surrogates [8] are the result of a digitisation process of a physical object. The project *E-Curator: 3D colour scans for remote object identification and assessment*<sup>7</sup>, for instance, uses University College London's collections and state of the art 3D-colour scanner to produce exact 3D surrogates of cultural heritage artifacts.

The quality of the digitisation is then measured in the information loss that the digitisation involves for the consuming agents. Traditionally, digital surrogates are to be consumed by humans and their quality is decided by the fact whether humans find the advantages of having the digital surrogate outweigh the information loss associated with it. The aim is to represent the physical object as accurately as possible, with no noise. This is particularly important in research, which bases its conclusion on the quality of the information.

<sup>2</sup><http://www.ncse.ac.uk>

<sup>3</sup><http://feed2js.org/>

<sup>4</sup><http://www.freshpaths.org>

<sup>5</sup><http://www.kcl.ac.uk/iss/archives/servingsoldier/>

<sup>6</sup><http://www.bbc.co.uk/ww2peopleswar/>

<sup>7</sup><http://www.museums.ucl.ac.uk/research/ecurator/>

Digital surrogates play an important role in e-Humanities and are often used to help with access to a fragile or expensive object. Medieval history, for instance, has been revolutionized in recent years by the existence of high resolution digital surrogates for manuscripts that were otherwise inaccessible and hidden in highly securitized archives [3]. For the information professional supporting e-Humanities, a digital surrogate is attractive as it allows virtually infinite reproduction of the object and eliminates the necessity for physical delivery. Though physical originals are still preferred in many disciplines, without doubt digital surrogates allow for less time and money spent on research travel and on complicated bureaucratic procedures to gain access. Supplementing real access, digital surrogates have led to better research performance without doubt.

A digital surrogate is useful where remote and concurrent access is required or analysis tools exist that help discover different features from a pure physical examination [8]. Traditionally, the risk associated with a digital surrogate is linked to the fact whether it is a near perfect substitute for physical examination. We would like to take a different perspective and look at surrogates from the point of view of how machine agents in a digital ecosystems are able to identify them. The first agent will be Google as the most universally used online research tool.

### 3.2. Data Pages

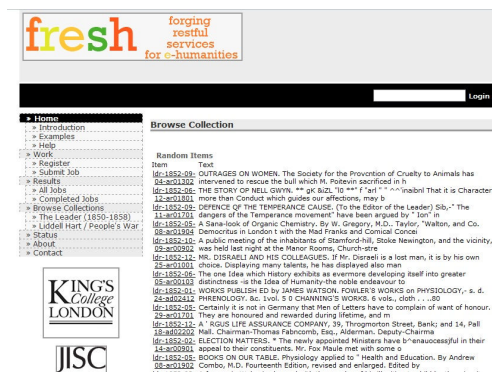


Figure 1. FReSH path data page

Traditionally, so-called landing pages are those pages that appear when a visitor enters a website. The page displays content that is an extended version of the information visible to a search engine. Landing pages play an important role in search engine optimization and in the general visibility of information on the web. In FReSH, we intended to expand the concept of a landing page to the concept of a data page that would be the first point of access for researchers to discover the chains of readings we create.

Most research users name Google as the one inevitably used digital tool in their research processes [9]. It has been the experience of most data providers, that Google is used even though there might be dedicated search services to retrieve information from remote research repositories. The concept of data pages recognizes this and tries to use Google instead to aid the discovery of online collections and the data in them, optimized to feature specific keywords or phrases for indexing by search engines. Figure 1 shows our example of such a data page for our chains of reading produced by FReSH.

The concept of data pages has been innovated by the Australian National Data Service (ANDS) if owners of data sets can register descriptions of data collections with ANDS.<sup>8</sup> ANDS only publishes these descriptions so that the data itself remains under the control of the owners. Data pages are therefore also an example of digital surrogates that help with discovery on the web while maintaining ownership.

Data pages are a very good representation to allow for search engine consumption of collections (chains of reading) but sometimes we would like to go deeper and present the individual document or article. The next two representations allow this, with two different machine agents in mind. The first representation allows other text mining agents to process articles in our collections, while the second one exposes these articles to the Linked Data cloud [2].

### 4. Surrogates for text mining

FReSH delivers a text mining agent. However, FReSH services only deliver a particular perspective. They cluster documents using standard vector space and latent semantic indexing algorithms. These clusters lead to document paths that humans can follow to find related information. It is important in a more generic digital ecosystem that we do not prescribe how these paths could be built. We want to enable other text mining agents to deliver different perspectives on the data. To this end, we introduce a digital surrogate of the source data that suits their need, while at the same time being abstract enough not to impede with restrictions on the access of the actual content.

In short, a text mining system [1] is based on two fundamental processing steps. In the first so-called indexing step, it aims to find a representation of the available information that as good as possible models the available information content in the considered documents, while at the same time delivers a computationally viable representation. In the second step, in the actual mining of documents, the system analyses the large sets of documents presented to it to extract and discover information. These might be new

<sup>8</sup><http://ands.org.au/>

relationships, as FReSH discovers new previously unknown links between documents, or new facts. The first indexing step is often based on extracting term frequency (TF) and inverted document frequency (IDF) information to evaluate how much information a word contains for identifying facts in a document collection. The assumption, that a word is more important the more it occurs in a collection, offset by the number of documents it occurs in, has proven to be a powerful representation. This TF-IDF information is finally collected in indexes and used to do the actual analysis step.

For FReSH, we decided that it often will be enough to give the TF-IDF index representation of the document for other text mining agents to do their analysis. In Listing 1, we present an abstract of such a representation of one document in the NCSE collection. We tested this by creating links to secondary literature collections in JSTOR using their research interface.<sup>9</sup> The processing worked though the actual results were limited, given that the NCSE is a collection of newspaper articles, which were only recently published and therefore cannot be found in the JSTOR collection.

```
<?xml version="1.0" encoding="utf-8"?>
<terms total="32">
  <term>
    <name>stohy</name>
    <frequency>3</frequency>
    <idf>8.466</idf>
  </term>
  <term>
    <name>tbavem</name>
    <frequency>2</frequency>
    <idf>8.451</idf>
  </term>
</terms>
```

**Listing 1. TF-IDF representation**

Please also note that, as discussed in [4], the main problem with text mining for historical collections is the quality of the input from the OCR'ing. Listing 1 shows a good example for this issue, as none of the ocr'ed terms is error-free.

The information in 1 helps other text mining agents consume information from FReSH. Still missing, however, are the link structures that are exposed through FReSH, i.e. the relationships between documents. In the next section, we shall investigate how Linked Data standards can be used to achieve this.

## 5. Representation of the link structure using Linked Data principles

Linked Data refers to a set of best practices for publishing and connecting structured data on the web using URI's

and RDF [2]. It directly builds upon the general architecture of the web, using its principles to publish data and not just web documents. To this end, URI's are used to name things, while various RDF standards provide useful information about the things described by the URI's. Finally, RDF links [2] take the form of RDF triples, that link two data sets and their name spaces. As of May 2009, the resulting web of data (the Linked Data cloud) consists of 4.7 billion RDF triples, which are interlinked by around 142 million RDF links [2].

Within the Linked Data community, it is considered to be good practice to reuse existing vocabularies such as FOAF, SKOS, SIOC, Dublin Core or OAI-ORE whenever possible. This ensures a wide-ranging interoperability with other existing web applications. Only if absolutely necessary, new dedicated terminology should be introduced. So, if we would like to publish our data sets as Linked Data, we first need to have URI's for each of them and associate them with some standard RDF description, which shall allow us to represent all the information we would like to publish.

In order to represent the links between the articles in a machine-readable manner, we use the emerging OAI-ORE standard.<sup>10</sup> OAI-ORE is a standard widely used in the online scholarly communications communities to describe aggregations on the web. So-called resource maps make it possible in OAI-ORE to map out web aggregations and describe their structure and semantics. Resources in an OAI-ORE aggregation are called aggregated resources.

In our case, the aggregations are the collections of documents, such as for instance all documents in 'The Leader' from NCSE. These collections are the original context of each document. Using text mining, we create a new context, its paths as a relationship of semantically related documents. A document can be in any number of different paths, while each path only exists within the context of a particular collection. In order to represent a resource that stands for an aggregated resource within the context of a specific aggregation, we use OAI-ORE proxies. A proxy is a resource that represents an aggregated resource in the context of a particular aggregation [10]. This way, we can avoid reification problems. Now, it is possible to represent both facts that document X is part of path Y and at the same time part of path Z by introducing the reified statement that in the context of collection C1 document X is in path Y, while in the context of collection C2 document X is in path Z.

Listings 2, 3 and 4 show the ATOM serialisation of our resource map. ATOM is an attractive protocol to serialise aggregations, as Web 2.0 applications can then reuse them [10]. We have seen how this might work in the NCSE experiments, described in Section 2. This way an ecosystem of related text mining and analysis agents is enabled using simple and proven web technologies. However, there are

<sup>9</sup><http://dfr.jstor.org/>

<sup>10</sup><http://www.openarchives.org/ore/>

also inherent difficulties to represent any kind of RDF in the XML-based ATOM format. According to [10], aggregations need to be represented as entries in ATOM feeds and the ATOM link element with an OAI-ORE value for its *rel* attribute is used to represent aggregations. Finally, awaiting a solution from the ATOM community, ore:triples element wrap further RDF descriptions, which are in our case the proxies. Their paths are represented as OWL enumerations.

Listing 2 shows how to represent the aggregation.

```
<?xml version="1.0" encoding="UTF-8"?>
<atom:entry
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:atom="http://www.w3.org/2005/Atom"
  xmlns:ore="http://www.openarchives.org/ore/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <!-- This Atom entry describes a specific ORE Aggregation -->
  <atom:link href="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/"
    rel="http://www.openarchives.org/ore/terms/describes"/>
  <atom:category term="http://www.openarchives.org/ore/terms/Aggregation"
    scheme="http://www.openarchives.org/ore/terms/" label="Aggregation"/>
  <!-- The Aggregation aggregates ... -->
  <atom:link href="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01210/"
    rel="http://www.openarchives.org/ore/terms/aggregates" title="... " type="text/html" hreflang="en"/>
  <atom:link href="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01211/"
    rel="http://www.openarchives.org/ore/terms/aggregates" ... />
  ...
```

### Listing 2. OAI-ORE Aggregation

Listing 3 delivers further metadata about the collection (in particular access rights), in this case 'The Leader'.

```
<!-- Metadata about the Aggregation: title and authors -->
<atom:title>The Leader Journal</atom:title>
<atom:author>
  <atom:name>...</atom:name>
</atom:author>
<!-- The HTTP URI of this Resource Map -->
<atom:link href="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/"
  rel="self" type="application/atom+xml"/>
<!-- Metadata about the Resource Map: datetimes, rights, and author -->
<atom:updated>2009-12-03T14:30:34Z</atom:updated>
```

```
...
<atom:rights>This Resource Map is available under the Creative Commons Attribution-Noncommercial Generic license</atom:rights>
<atom:link href="http://creativecommons.org/licenses/by-nc/2.5/rdf" rel="license" type="application/rdf+xml"/>
...
```

### Listing 3. OAI-ORE Metadata

Finally, Listing 4 demonstrates the serialisation of proxies.

```
<oreatom:triples>
<!-- More additional properties ... -->
<!-- Proxy for Aggregated Resource http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01210/ -->
<rdf:Description
  rdf:about="http://oreproxy.org/r?what=http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01210/&where=http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/">
  <ore:proxyFor rdf:resource="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01210/" />
  <ore:proxyIn rdf:resource="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/" />
</rdf:Description>
<!-- Proxy for Aggregated Resource http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01211/ -->
<rdf:Description
  rdf:about="http://oreproxy.org/r?what=http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01211/&where=http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/">
  <ore:proxyFor rdf:resource="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/item/ldr-1854-04-15-ar01211/" />
  <ore:proxyIn rdf:resource="http://www.freshpaths.org/browse/collection/The%20Leader%20(1854)/" />
</rdf:Description>
...
</oreatom:triples>
```

#### Listing 4. OAI-ORE Proxies

To extract the triples from the ore:triples extension, we can reuse an existing GRDDL transform service.<sup>11</sup>

To embed our paths as a relationship of the proxies we use the XML serialisation of OWL's enumerations.<sup>12</sup> This allows for the reuse of another standard vocabulary. We can simply embed the XML in the content element of ATOM. However, in many ways it would have been better to embed this information in the ATOM feed using another ore:triples like extension. But, we lacked the necessary extractor to deal with such triples. Implementing one should not, however, be a problem and can be done in future work.

```
...
  <owlx:Class owl:name="2046" owl:complete="
    true">
    <owlx:OneOf>
      <owlx:Individual owl:name=".../ldr
        -1857-08-22-ad02308/" />
      <owlx:Individual owl:name=".../ldr
        -1857-01-03-ad02214/" />
      <owlx:Individual owl:name=".../ldr
        -1857-01-03-ad02214/" />
      ...
    </owlx:OneOf>
  </owlx:Class>
</atom:content>
...
</atom:entry>
```

#### Listing 5. OWL serialisation

As we have a highly visible data page for of the created paths, we also experimented with using RDFa<sup>13</sup> to embed the OAI-ORE directly in these data pages.

Our text mining agent is another example of how to create data links in the Linked Data cloud automatically. Automation of links is subject to intense research [2] but also common practice as data resources often contain large number of entities. For instance, in [7] the authors develop algorithms to interlink different music-related data resources. As seen in Section 2, by employing simple clustering algorithms we are able to create chains of readings or in the language of Linked Data interlinked data resources. In this section, we have seen how to represent them using various Web 2.0 standards.

## 6. Conclusion

In this paper, we presented effective means of representing a text-mining agent in a digital ecosystem for e-Humanities. We demonstrated the results of the FReSH

project, which supports the creation and use of scholarly resources in research. With FReSH and its predecessor projects, we follow a strict resource-oriented approach. In [4], we have shown how effective different combinations of standard web technologies such as ReSTful services and ATOM feeds can be to deliver text mining services to end users in the Humanities. Here, we concentrated on more effective means to deliver text mining results to machine agents in the digital ecosystem. To this end, we offered three alternative versions of digital surrogates, with which we experimented in FReSH. The first delivered results to the agent in the digital e-Humanities ecosystem that is most commonly used by researchers: Google. The second one delivered FReSH processing outcomes to other text mining agents, while the final one makes FReSH processing visible in the emerging scholarly communications web.

## References

- [1] M. W. Berry and M. Castellanos. *Survey of Text Mining II: Clustering, Classification, and Retrieval*. 2007.
- [2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 2009.
- [3] T. Blanke, A. Aschenbrenner, M. Küster, and C. Ludwig. No claims for universal solutions. In *E-SCIENCE '08: Proceedings of the IEEE e-Humanities Workshop*, Washington, DC, USA, 2008. IEEE Computer Society.
- [4] T. Blanke, M. Hedges, and R. Palmer. Restful services for e-humanities. In *3rd IEEE International Conference on Digital Ecosystems*, Washington, DC, USA, 2009. IEEE Computer Society.
- [5] W. Brockman, L. Newmann, C. Palmer, and T. Tidline. Scholarly work in the humanities and the evolving information environment. 2001.
- [6] E. Chang and E. Damiani. Call for papers IEEE-DEST 2007. <http://www.ieee-dest.curtin.edu.au/2007/>, 2006.
- [7] Y. Raimond, C. Sutton, and M. Sandler. Automatic interlinking of music datasets on the semantic web, 2008.
- [8] J. Unsworth. The value of digitization for libraries and humanities scholarship, 2004.
- [9] J. Unsworth. The draft report of the american council of learned societies' commission on cyberinfrastructure for humanities and social sciences, 2006.
- [10] H. Van de Sompel, C. Lagoze, M. L. Nelson, S. Warner, R. Sanderson, and P. Johnston. Adding escience assets to the data web. Technical Report arXiv:0906.2135, Jun 2009. Comments: 10 pages, 7 figures. Proceedings of Linked Data on the Web (LDOW2009) Workshop.

<sup>11</sup><http://www.openarchives.org/ore/atom-grddl>

<sup>12</sup><http://www.w3.org/TR/owl-xmlsyntax/>

<sup>13</sup><http://www.w3.org/TR/rdfa-syntax/>