# Chapter Nine

# Scholarly communication: a System in Transition*

***Herbert Van de Sompel***
Los Alamos National Laboratory
Email: hvdsomp@gmail.com

***Jenny Walker***
Independent Consultant
Email: jenny-walker@msn.com

*Review status: Peer reviewed*

## Introduction

This paper, written by Jenny Walker, is based on the keynote presentation delivered by Professor Herbert Van de Sompel of Los Alamos National Laboratory. Professor Herbert Van de Sompel, Information Scientist, from the Los Alamos National Laboratory, New Mexico, United States of America, has been instrumental in addressing many issues on the path of transition from a paper-based scholarly communication system to a web-native digital environment. More than fifteen years ago, Van de Sompel worked on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Open Archives Initiative, 2015), a protocol for the recurrent exchange between systems of XML-structured metadata including Dublin Core.

More recent initiatives in which he has been instrumental include the Open Annotation Project, which ran from 2009 to 2013 (Open Annotation, 2015), Memento, which also ran from 2009 to 2013 (Memento, 2015), ResourceSync, running from 2102 to 2014 (NISO, 2014) and Hiberlink, running from 2013 to 2015 (Hiberlink, 2015).

All the above projects share some characteristics in that they concern the use of the web, the web architectures and the tools of the web trade in order to use the web as a platform for scholarly communication. Van de Sompel and his team are extremely interested in matters of interoperability across scholarly systems. For example, the protocol for metadata harvesting comes from the perspective of repositories and interoperability between repositories. Van de Sompel's work in the last decade has involved issues not only of interoperability between repositories, but also interoperability with the broader web. Further, Van de Sompel's work is targeted at making information systems more robust, which leads to better scholarly communication.

This paper covers some of the challenges involved in moving from a print-based scholarly communication system towards one that is based solely on the web, thus creating a web-native digital environment for scholarly communication in which there is no equivalent paper form. This transition, whilst offering advantages, also engenders difficulties, which this paper explores.

The current scholarly communication system is largely based on the paper-based systems with which we have long been familiar. Even those examples of web-based systems are scarcely more than a scanned version of the paper-based system and do not represent a novel approach: the current system was not reinvented from scratch, but simply copied into the web. What might be expected from a radical transition to a web-native

*The editors would like to acknowledge the contributions of Peter G. Underwood for this chapter.

digital environment where scholarly communication truly embraces the web in all its aspects?

**Background**

The Open Annotation project looked at means of sharing annotation of resources on the web in a form that is independent of the platform used. The intention is to allow users to add, modify or remove information from a web resource without affecting the integrity of the resource itself, thus allowing social commentary, evaluation and other guidance to be added by users without affecting how others wish to use the web resource. The initiative is a direct reflection of the needs of scholars in many disciplines, derived from a study of how annotation has traditionally been used in a paper-based environment allied with consideration of solutions that could be used in a web-based environment. The scope has also been widened to consider how annotation could be used in areas other than scholarly communication.

Memento (2009-2013) (Memento, 2015) was a project funded by the United States National Digital Information Infrastructure and Preservation Program (NDIIPP) aimed at making content on the web that has been archived more easily located and used. By supplying the original web address and a chosen target date, the user was able to request an earlier version of a web resource, such as a web page. This would enable scholars to view the revision history and other versions of a web resource, though the intended scope of use is wider than scholarly communication. The Internet Archive Wayback Machine (https://archive.org/web/) and Time Travel (http://timetravel.mementoweb.org) are practical examples of the application of this idea.

ResourceSync (2012-2014) (NISO, 2014) was a joint project of the US National Standards Organization and the Open Access Initiative which revisited the design of the protocol for metadata harvesting, earlier explored in the OAI-PMH project. ResourceSync extends the notion of synchronising metadata across systems to synchronising the web resources themselves across systems, thus facilitating the orderly updating of the contents of repositories. By automating this process, the new standard will economise on the time, effort, and resources required to manage repositories, also increasing the general availability of content available and reduce problems created by content that has become outdated, inaccurate, or has been superseded.

The background to this paper is the ecology of scholarly communication, but it will focus on one aspect – long term access to the scholarly record. In the future, will it be possible to revisit the materials that we publish today by recreating the context of the publication as it existed at a certain point in time? Is this possible for the paper-based system, for the current system and for a future system?

Users of the web often find that web addresses no longer function after a period of time: the "File 404 error" or "Page not found" often occurs, especially for older material. In many cases, the resource is still available but has been moved to a different address. This "linkrot" or "reference rot" is a substantial nuisance and threatens the integrity of the scholarly web. Furthermore, even working links may lead to material that has changed substantially as newer versions have superseded the original files. Hiberlink, a project which ran from 2013 to 2015, (Hiberlink, 2015a), is at the core of this paper: it is about combatting "reference rot" in the scholarly web.

Fundamental to an understanding of the issues to be discussed, is to understand the terms "HyperText Transport Protocol" (HTTP) and "Uniform Resource Identifier" (URI). Each item stored in association with a computer system has a name that enables the system to find it. The most familiar manifestation of a URI is a Uniform Resource Locator (URL), often termed a "web address", where the location is combined with a designation of the scheme (for example, http) that is used to view the item.

## Long term access to the scholarly record

### *Paper-based systems*

The familiar paper-based system represents one that tends to be stable and where its contents persists and can be located after long periods of time. The assumption built into the scholarly communication system is that if reference is made to an item in the system by providing its location and identity, then that item will be capable of being found. For example, in the paper-based journal system, an article that is published in a journal would contain references to other publications, mostly other articles, all printed on paper and held in library collections and archives. Thus, sometime after original publication it was possible to access a journal article and also the surrounding context such as the article references and the other articles in the same journal. The journals containing the referenced articles were likely to be held on the shelf of the same or other libraries; in the worst case scenario the reader would have needed to take the train or the bus to visit a couple of libraries or the inter-lending system between libraries would have been used. In a paper-based environment, it is eventually possible to reconstruct the entire context for an article.

### *Web-based systems*

In the current version of the scholarly communication system articles are published in e-journals that are accessible via the web. An article is published in an e-journal and that article references other publications. In contrast with the paper-based paradigm, the references are "live" and link the user via web links to those referenced articles and other web-based items. It is no longer necessary to travel to other libraries or request an inter-library loan to get the article from another library. However, libraries are no longer archiving the electronic journals to which they subscribe; archiving is now in the hands of specialist organizations such as Portico, Lots of Copies Keep Stuff Safe (LOCKKS) and others (Portico Digital Preservation Service, 2015; LOCKKS, 2014).

In order to revisit an article on the web, the links in the article and the article references must remain valid, even after a significant period of time. Further, access to the archival material must be possible. However, links on the web are sometimes brittle, being vulnerable to "reference rot", and those associated with scholarly publications on the web share this weakness, despite the controlled environment of scholarly publishing.

This problem of "reference rot" was identified early on in the transition from print to e-journals on the web. In order to address the issue of publisher mergers and the move of a publication from one location to another, the use of persistent identifiers for journal articles – the Digital Object Identifier (DOI) – was introduced about twenty years ago (DOI, 2015a). In addition, the DOI resolver system was developed to ensure that links pointing at these articles by means of a DOI continue to be valid and work, even when the articles change web location (DOI, 2015b). Each article is assigned a unique DOI on publication and this is a persistent identifier for that article throughout its life, regardless of whether responsibility for the publication moves from one publisher to another or the location of the web hosting is moved. The operation of the DOI effectively is to remove the actual web location of an article: rather, it transparently leads the user to its current location. More than fifteen years ago, a consortium of publishers created Crossref to assign DOIs and manage citation links for scholarly publications (Crossref, 2015). Crossref now has over seventy-five million DOIs registered for scholarly articles and other content items (books, chapters, data, theses, technical reports) and provides a very effective infrastructure for interoperability across the scholarly web.

Archiving in the journal system is not very effective. David Rosenthal, the inventor of the LOCKSS system, in his blog posting, "Patio Perspectives at ANADP II: Preserving the Other Half", states that, on average, less than 40% of the typical journal collection of a North American mem-

ber of the Association of Research Libraries is archived (Rosenthal, 2013: para 3). That is very low and what is archived is unlikely to be endangered because it consists typically of the materials owned by large publishers, such as Elsevier and Springer, that are unlikely to vanish in the foreseeable future. Also, typically, what is archived is what can easily be found. Some of the journals that exist on the web are a both harder to find and less likely to be archived. At the University of Edinburgh, the UK Joint Information Systems Committee (JISC) has been instrumental in the development of a centre for digital expertise and online service delivery, known as EDINA (EDINA, 2015): as part of this work, the Keepers Registry has been introduced to keep track of the extent to which the digital journal literature is archived and by which organizations. These include the Archaeology Data Service, the British Library and the Library of Congress, Portico, the Global LOCKSS Network and the Scholars Portal (Keepers Registry, 2015).

Interestingly, the Keeper's Registry is based on the use of the International Standard Serials Number (ISSN), volume and issue number, rather than DOIs or URIs. This reflects a print-based view of journals rather than a growing reality in which consumers of scholarly information think in terms of disaggregated journal articles that exist on the web, each with its unique identifier. What is needed is a catalogue that works with DOIs to uniquely identify the articles and to easily determine whether an article is available in a particular archive.

In this web-based environment, therefore, it is not known whether the context surrounding an article can be recreated because it is not known what is being archived. But, understanding that less than 40% of the journal literature is archived, it is clear that attempts at re-creating the original context of a publication, its chain of references and other links, is unlikely to be successful because of the extent of the material that is missing from the archive.

Looking further at the current environment there is another problem to consider. When a new article is published, reference links are given to publications in the journal system, but there are also links to the web-at-large; links for items that are available on the web but are not considered to be at the core of the scholarly communications system. For example, project websites, software, ontologies, online debates, slides, blogs, videos: all are used or created as part of scholarly endeavour, but the community does not necessarily consider such items as a publication. Web resources are dynamic, changing over time; they are ephemeral, vanishing from the web. This is the real, evanescent, nature of resources on the web. So, when a journal article links to a web-at-large resource, there is a high probability that, sometime after publication of the article, the resource has disappeared or its content has changed. This is not hypothetical: it is happening today. The Hiberlink Project studies the extent of this problem and considers what hinders recreation of the original context surrounding an article.

**Hiberlink**

The Hiberlink Project is studying, on a very large scale, this phenomenon of references to vanishing or changing web-at-large resources (Hiberlink, 2015a). The Hiberlink Project is a collaboration between the Los Alamos National Laboratory (LANL), New Mexico, and two groups at the University of Edinburgh: EDINA, which provides a number of information services to higher education in the United Kingdom, and the School of Informatics (Hiberlink, 2015b).

The problem domain being considered in Hiberlink is the basic scholarly communication domain, with a focus on the journal system in which journal articles may contain a mixture of references to resources on the web, consisting of some formal citations to scholarly resources and others to general web resources – "web-at-large" resources – that are not considered to be at the core of the scholarly record. Critical to the Hiberlink Project is an assessment of the extent of the

problem and, to this end, an assessment of the extent of "reference rot" has been undertaken, by assessing several large collections – corpora – of scholarly publications available online to determine what links still work and what material has been archived.

One of the corpora used in the Hiberlink research is that based on a subset of PubMed Central (PMC), which consists of about 480,000 articles published between 1997 and 2012. All these articles were checked for the presence of URIs. The number of links to web-at-large resources was shown to be growing steadily with each publication year. For example, of the articles in the PubMed Central corpus that were published in 2012, there were approximately180,000 links to articles and 120,000 links to other web-at-large resources. For articles published in 2009 there were 100,000 links to articles and 40,000 links to web-at-large resources (Klein, Shankar, Van de Sompel & Wincewicz 2014).

**Reference Rot**

The Hiberlink Project introduced the term "reference rot" to describe the combination of two problems involved in using URI references. Both these problems relate to the dynamic and ephemeral nature of the web and were both considered in the Hiberlink project.

1. Link rot
   "Link rot", as referred to earlier, whereby the link stops working as a result of the referenced resource at the end of the link having disappeared or having been moved to another location.
2. Content Drift
   The resource identified by a URI may change over time and the content at the end of the URI may evolve to such an extent that it is no longer representative of the content that was originally referenced.

In certain cases, such as with popular culture, "content drift" is not such a problem; however, in the context of scholarly communication, it is very important for the integrity of the scholarly record and the integrity of referenced items: at the time an article was written and a web resource was referenced, that web resource contained a certain item of interest which further illuminated the meaning of the article. However, when that web resource is revisited, some years later, the content may well have significantly changed and may no longer be so strongly linked to the content of the citing article.

For example, today, the URI of the Digital Library Conference in the year 2000 (DLC) no longer resembles anything like a web page for the Conference. Those responsible for the domain name did not keep up their registration payments and the domain name has been taken over by others. From a closer look at the DLC sites for 2004, 2005, and 2008 it is evident that there have been three different owners over the last eight years. This is a good example of "content drift". Although this example may illustrate an extreme case, in that the domain owner did not maintain the domain, content drift also frequently occurs at the level of individual web pages that remain under ownership of the same custodian. Project web sites, for example, change as projects evolve and new content becomes available with the result that what was referenced at one time is no longer available at a later date.

The Hiberlink Project explored the extent of the "reference rot" problem. To assess this, Hiberlink has used articles published between 1997 and 2012 from three corpora: PubMed Central, arXiv, and a random sample of Elsevier articles. The PubMed Central corpus contained about 480,000 articles of which about 400,000 had URI references to other web-based resources. Of those, about 240,000 pointed to other articles and about 156,000 (or one third of all articles) pointed to web-at-large resources. Also, 1,600,000 URIs were found when parsing all three corpora: of these about 750,000 were references to articles and 480,000 to web-at-large resources, about a third of all references (Klein, Van de Sompel, Sanderson, et al., 2014).

The first part of the study looked at "link rot",

determining whether or not the links still worked after a period of time. It was found that the older the publication and, therefore, the older the link, the higher the chances that the link no longer worked. For example, 80% of links that were created in 1997 no longer worked. For 2012 publications in the corpus, by early 2014, around 20% of links no longer worked. Within five years of publication, the rate of "dead" references was around 40%. Extrapolating from these numbers, in 2020 it is estimated that about 50,000 links in articles published in 2012 will no longer work. These statistics are indicative of a severe problem.

The situation is serious because it is not known whether the content at the end of a link put in place in 2005, for example, would be the same in 2014. Hiberlink looked at web archives to see if the content to which links that were created in 2005 and successive years could still be found. Web archives contain snapshots of resources as they evolve over time, so using the Memento infrastructure, web archives around the world were automatically searched to find the relevant snapshot of the linked web resource around the time it was referenced (Memento, 2015).

However, knowing whether or not a snapshot exists is not sufficient; through "content drift" the content of that web resource may have changed. If, for example, a link is put in place in 2005 and an archive snapshot is found that dates back to 1999, and another archive snapshot dated 2011, it is unlikely that either snapshot will be representative of what was referenced in 2005.

**The window of representativeness**

Rather than looking for any archived copy of a particular resource, the Hiberlink Project looked for an archived copy that was likely to be representative of what the author creating the reference originally saw: typically, this was an archived copy dated to within a one-month period around the period of publication — 14 days prior to publication and 14 days after publication. Archived copies that fitted into this interval were found for only 20% of the referenced resources; for 80% of the referenced resources, no suitable archived copy was found. This is not particularly surprising as what is presently available in web archives is a result of incidental archiving by systems that go about the regular business of crawling the web and placing items into the archives. This, then, is not necessarily the result of purposive action taken after determining what needs to be archived in relation to the needs of the scholarly communications system and the integrity of scholarly literature.

A closer examination of the referenced links in the three corpora reveals that the six most–used top-level domains of the URIs are **.org**, **.edu**, **.com**, **.gov**, **.uk** and **.de**. PubMed Central articles have a lot of links to the .org domain; Elsevier has a lot of links to the **.com** domain and arXiv to the **.edu** domain.

**Pockets of persistence**

The Hiberlink Project not only quantified the problem, it also considered solutions to address the problem. Out of that work emerged the notion of creating "pockets of persistence", relating to the need to be able to revisit the scholarly web of the past and the web of the present at some point in the future in a persistent, precise and seamless way. The challenge exists for the entire web, but some communities care more about addressing the issue than others, including the scholarly community, the legal fraternity, online journalism, and socially-constructed sources of information, such as Wikipedia.

The results of the Hiberlink Project suggest that there are three main components to address the "pockets of persistence" challenge:

1. Capturing referenced resources pro-actively
2. Appropriately referencing the resources once they are captured
3. Accessing the resources and/or their captures.

Rather than the haphazard approach to archiving currently adopted by web archives, Hiberlink proposes a pro-active archiving of web re-

sources that integrates seamlessly into the life cycle of an article and requires little, if any, explicit intervention by the author. Hiberlink involves the concept of a "seed collection", for example of scholarly literature, legal documents, online journalism or Wikipedia articles. The seed collection for a particular community becomes the starting point for capture; it becomes the corpus of interest for which pockets of persistence can be created. All the items in a seed collection, or in the collection of interest, point at other items on the web which are vulnerable and may eventually disappear.

Hiberlink proposed that at important moments in the life cycle of items in the seed collection, snapshots be taken of the referenced items and that these are placed into a web archive, together with the referencing item. The point of intervention in the life cycle of items in the collection very much depends on the kind of collection. For example, the Hiberlink Project, identified that, for a corpus of scholarly literature, there are a number of really significant moments in the life cycle of journal publications. These include the research stage when the author is finding relevant items to cite and reference and decides to record them in a reference manager tool, the automated submission process, and during the review process just prior to publishing, or maybe even later on when the document is being interacted with. Such life cycle moments also exist for information sources such as Wikipedia. For example, when a substantially new version of an article is created and that article has external references, these could all be archived. The Wikipedia article and all external references could be archived.

There is already a lot of activity around this idea, such as the newly-created web archive, Perma.cc (Perma.cc, 2015). This is an effort led by the University of Harvard Law Library as part of a consortium of law libraries around the world that maintains a web archive of web resources referenced in legal documents. Perma.cc was created to address the substantial reference rot in legal literature, including US Supreme Court decisions (USA, Supreme Court, 2015). This archive allows someone who is altering a legal document to enter the URI that they are planning to reference in their document and a snapshot of the resource will be taken. The URI of the archived snapshot will be provided and this can then be used for referencing rather than the URI of the document on the "live" web.

The Hiberlink Project is looking at ways in which to make this process more seamless. In the case of Perma.cc the author specifically has to go to the archives, enter a URI and retrieve a new URI that can be used for referencing. Hiberlink is experimenting with using the Zotero reference manager tool so that while an author is taking notes for a new publication, each time a page is bookmarked using Zotero, the system will automatically 'push' that resource into a web archive. The author is not necessarily aware that this action is taking place. It is being done behind the scenes just in case the author wants to reference that resource later on, in which case there will be a snapshot in a web archive. An experimental Zotero extension archives web resources as an author bookmarks them during the research process and ensures that the author has access to the original URI, the URI of the snapshot that was taken and archived and the date and time when the snapshot was taken (GitHub, 2015). These three components are all extremely important as will be described later.

Another experimental service, Hiberactive, integrates with a manuscripts submission system or a repository: when an author submits an article, Hiberactive issues requests to web archives to archive all web-at-large resources referenced in the article (Klein, Shankar, Van de Sompel & Wincewicz, 2014). At the moment when an author submits a manuscript, the repository to which the manuscript was submitted sends a notification to the Hiberactive service, indicating that there is a new document in the repository. Hiberactive fetches that document, and identifies all the URIs that are referenced, just as was done

with the PubMed Central corpus. For each retrieved URI that points to a web-at-large resource, Hiberactive places a snapshot in one of the web archives around the world. Again, this process provides the original URI, the URI of the capture, and the date and time of capture.

Once we have these captures, how are we going to actually reference them? Having the URI of the capture is not sufficient. The web archive has the date and time on which the snapshot was taken but this is not recorded in the document itself. This is problematic as having disposed of the original URI, it is not possible to revisit the original resource should you wish to do so. Software development is a prime example of why you might want to do this. If a piece of software was referenced two years ago you would want to be able to revisit and see exactly how that software was at that moment and not how it is today. Further, for some purposes, you might also want to see how it has evolved and see what the current version of that software is. This would not be possible without the original URI. More importantly, by linking to the capture of the URI, you are now completely dependent on the long term existence of the archive in which the capture was placed. You have thereby replaced one link rot problem with another as web archives are also not guaranteed to last forever (Van de Sompel et al., 2013).

WebCite, which commenced operation ten or more years ago, was the very first web archive to address the issue of link robustness for scholarly communication (WebCite, 2015). An author could submit a URI, get a snapshot of a resource as well as a URI for that resource. In 2014 the service was suffering financial problems and was at a point where it could no longer accept new captures. There was considerable uncertainty about whether the system could be kept active. Despite active fund-raising, the service is still in jeopardy.

Mummify.it was one of the very first commercial web archiving services. The service existed for about six months before closing down; the only remnant of this web archive is a snapshot in the Internet Archive (Mummify.it, 2013).

While web archives can disappear, they can also become (temporarily) inaccessible. Some time ago, a Russian blocking order prevented Russian citizens from accessing the parent site of the Internet Archive's Wayback Machine, a non-profit library of over 485 billion snapshots of web pages taken on various dates (Internet Archive, 2015). The blocking order was originally made to ban a single web page, entitled "Solitary Jihad in Russia", a brief text detailing the "theory and practice of partisan resistance" (Carey, 2015). The implication here is that, had one included the URI of the capture in a reference and that capture reference was to the Internet Archive, it would not be possible to visit it from Russia. There would simply be no access.

The examples above illustrate that it is not a good idea to dispose of the original URI because the original URI is a key to finding snapshots in all the web archives around the world. Without the original URI it is impossible to find sources in web archives other than the one in which the specific source was deposited. This means that the success of the approach is fully dependent on the long term existence of that one archive.

**The proposed solution**

Hiberlink proposes that the original URI be used for referencing but that, to ensure a robust link, the reference be augmented – or "decorated" – with other archival information, namely, the URI of the specific archived capture of the referenced resource and/or the date and time of referencing. Further, it is proposed that the reference information be made available in a form that a computer can process. This will allow consumption of the information in various ways. Using the original URI the original referenced resource can be revisited as it evolves over time. Using the URI of the capture, the user can revisit that capture for as long as the web archive in which it was created remains operational. And, thanks to the protocol developed in the earlier-mentioned

Memento Project and its associated infrastructure, a combination of the original URI and the date and time of referencing can link to other captures which can be found in web archives taken close to the time the item was originally referenced.

The Memento extension for the Chrome browser is available for download (Shankar, 2015). With this extension, and by right-clicking on a link, you can revisit the linked resource as it was at some point in the past. You can specify the date for which you would like to see the resource and in this way you can "time-travel" the web as it used to exist. This extension also supports the aforementioned link decorations, which are formally specified at http://robustlinks.mementoweb.org/spec/. The link decorations can also be made operational using a JavaScript application, robustlinks.js, that can be linked or embedded in a web page that contains decorated links (Robust Links, 2015). In November 2015, *D-Lib Magazine* published a paper by Herbert Van de Sompel and Michael Nelson that illustrates decorated links in action (Van de Sompel & Nelson, 2015).

**Conclusion**

The archiving of online scholarly resources is not at all comprehensive. Today, the archival activity has granular perspective in which each resource is treated as an autonomous item without dependencies on other resources. This perspective is no longer realistic and it will become less so in the future.

The examples given in this paper primarily concern journal literature. However, the environment is changing and, in the future, journal literature may no longer be central to the scholarly record. There is a wide variety of other means of communication emerging which will be web-native and will have a greater degree of interconnectedness with other resources and will be more dynamic, changing more rapidly over time. It will be increasingly important, if not essential, to be able to wind back the clock to review what the scholarly record looked like at a particular moment in time, together with its temporal context. Ultimately, this is about the long-term integrity of the scholarly record.

**References**


Carey, G. 2015. Russia won't go Wayback, blocks the Internet Archive. *Digital Trends*, (26 June). Available: http://www.digitaltrends.com/web/russia-censors-wayback-machines-485-billion-archived-web-pages/ [27 November 2015].

Crossref. 2015. *About Crossref*. Available: http://www.crossref.org/ [27 November 2015].

DOI (Digital Object Identifier). 2015a. Home page. Available: http://www.doi.org [27 November 2015].

DOI (Digital Object Identifier). 2015b. *Resolve a DOI Name*. Available: http://dx.doi.org [27 November 2015].

EDINA. 2015. *Who we work with*. Available: http://edina.ed.ac.uk/about/whoweworkwith.html [27 November 2015].

GiTHub. 2015. *Hiberlink Zotero plugin*. Available: https://github.com/hiberlink/zotero_hiberlink [27 November 2015].

Hiberlink. 2015a. *About the Project*. Available: http://hiberlink.org/about.html [27 November 2015].

Hiberlink. 2015b. *Partners*. Available: http://hiberlink.org/partners.html [27 November 2015].

Internet Archive. 2015. *Wayback Machine*. Available: https://archive.org/web/ [27 November 2015].

Keepers Registry. 2015. *Information about archiving agencies*. Available: http://thekeepers.org/registry.asp?action=agencies [27 November 2015].

Klein, M., Shankar, H., Van de Sompel, H. & Wincewicz, R. 2014. HiberActive: Pro-active archiving of web references from scholarly articles. Presentation at *Open Repositories 2014*, Helsinki, Finland, 9-13 June. Available: http://www.doria.fi/handle/10024/97609 [27 November 2015].

Klein, M., Van de Sompel, H. , Sanderson, R., Shankar, H., Balakireva, L., Khe Zhou & Tobin, R. 2014. Scholarly context not found: one in five articles suffers from reference rot. *PLOS One*, 9(12): e115253. doi: 10.1371/journal.pone.0115253. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253 [27 November 2015].

LOCKSS (Lot of Copies Keep Stuff Safe). 2014. Home page. Available: http://lockss.org/ [27 November 2015].

Memento. 2015. *About the Memento Project*. Available: http://mementoweb.org/about/ [27 November 2015].

Mummify.it. 2013. Home page of discontinued web archive service Mummify.it [http://mummify.it], captured on 14 February 2014. Available: https://web.archive.org/web/20140214233752/https://www.mummify.it/ [27 November 2015].

NISO (National Information Standards Organization). 2014. *Open Archives Initiative ResourceSync Framework Specification*. Available: http://www.openarchives.org/rs/toc [27 November 2015].

Open Annotation. 2015. *Welcome*. Available: http://www.openannotation.org/spec/core/ [27 November 2015].

Open Archives Initiative. 2015. Home page. Available: https://www.openarchives.org/ [27 November 2015].

Perma.cc. 2015. *About Perma*. Available: https://perma.cc/about [27 November 2015].

Portico Digital Preservation Service. 2015. Home page. Available: http://www.portico.org/ [27 November 2015].

Robust Links. 2015. *Robust links: links that refuse to die*. Available: http://robustlinks.mementoweb.org/ [27 November 2015].

Rosenthal, D. 2013. Patio perspectives at ANADP II: preserving the other half. *DSHR's blog* [20 November]. Available: http://blog.dshr.org/2013/11/patio-perspectives-at-anadp-ii.html [27 November 2015].

Shankar, H. 2015. *Memento time travel* [Chrome browser extension]. Chrome web store. Available: https://chrome.google.com/webstore/detail/memento-time-travel/jgbfpjledahoajcppakbgilmojkaghgm?hl=en&gl=US [27 November 2015].

USA (United States of America), Supreme Court. 2015. *What's new* (section 2). Available: http://www.supremecourt.gov [27 November 2015].

Van de Sompel, H. & Nelson, M. L. 2015. Reminiscing about 15 years of interoperability efforts. *D-Lib Magazine*, 21(11/12). doi: 10.145/november2015-vandesompel. Available at: http://dx.doi.org/10.1045/november2015-vandesompel [27 November 2015].

Van de Sompel, H., Klein, M. Sanderson, R. & Nelson, M. L. 2013. *Thoughts on referencing, linking, reference rot*. Available: http://mementoweb.org/missing-link/ [27 November 2015].

Web Cite. 2015. *WebCite Consortium FAQ*: What is the history of this service? Available: http://webcitation.org/faq [27 November 2015].