

Remembering the Past in the *Dynarchive*: The State of Knowledge in Digital Archives
Paper submitted for the conference Media in Transition 7, *Unstable Platforms: The Promise and Peril of Transition*. Boston, MIT, May 13-15, 2011.

Julia Noordegraaf, University of Amsterdam

Introduction

The proliferation of digital technologies has changed the way we perceive of and use audiovisual archives and their holdings. As Rick Prelinger, founder of the online collection archive.org recently pointed out, YouTube has become the standard of what people expect audiovisual archives to be – unlimited online access and active user participation have become crucial for an archive’s visibility and public existence. (Prelinger 2009: 269-270) Although the institutions still function as the principal gatekeepers – if only because of copyright restrictions – the emergence of virtual archives and online portals is changing the relation between the keepers and users of audiovisual heritage, challenging the role of the archivist as principal expert on the knowledge the collection represents. In this paper I investigate the implications of these developments for the status of the (audiovisual) archive as a gatekeeper of knowledge. I analyze a recent experiment with social tagging (the Netherlands Institute for Sound and Vision’s video labeling game *Waisda?*) and ask to what extent these kinds of experiments destabilize the existing archival platforms for validating and describing audiovisual heritage.

The Archive as Repository of Knowledge

Archives, like libraries and museums, are custodians of cultural heritage. Archives founded as institutions, often related to the state and funded with public money, have the legal task of collecting the documentary traces of activities of organizations, individuals, families, companies or other social groups. The aim of archival work is to safeguard this documentary heritage and make it accessible and meaningful for contemporary and future users. (Cook 1997: 18) From the publication of the first archiving handbook, known as the Dutch *Manual* (Muller, Feith and Fruin 2003 [1898]), archivists have organized archival records according to the principle of *provenance* which stipulates that records produced by a single administrative body should be kept together. (Cook 1997: 21) This ensures that the records are not only preserved but also kept *contextually meaningful* to the original purpose and function of the document. It is this context that ultimately gives records their evidentiary value. (Horsman 1994; Winget 2005)

The context of archival records is documented in the so-called *metadata*, the sets of information accompanying each individual object or *record*. These metadata are crucial for retrieval and for determining the exact nature and meaning of the objects in question. Without a catalogue that classifies and describes the content of the archive it would be impossible to find what one is looking for. Besides, these metadata also inform users on the what, when, where and by whom of the records in question and thus are crucial in safeguarding the evidentiary status of the material.

Traditionally, archivists have understood the task of creating a reliable record of the past as requiring systematic, transparent and objective methods of describing archival records. The International Council of Archives’ guidelines for description clearly define these requirements:

- I.2 The purpose of archival description is to identify and explain the context and content of archival material in order to promote its accessibility. This is achieved by

creating *accurate and appropriate representations* and by organizing them in accordance with *predetermined models*. [...] These processes make it possible to institute the intellectual controls necessary for *reliable, authentic, meaningful and accessible* descriptive records to be carried forward through time. (ICA-ISAD(G) 2000: 7, emphasis added)

Seen from this perspective safeguarding reliability, authenticity, meaningfulness and accessibility requires that archival material is described in accordance with predetermined models that guarantee that the meaning of the original documents is preserved and made accessible for future users. ‘Meaning’ here is understood primarily as a reflection of the author’s or maker’s intention. As information resources specialist Eileen Peterson points out, methods like Aristotle’s influential classification system or the Dewey Decimal Classification (DDC) widely used in libraries assign objects a fixed place in pre-defined thematic categories based on the author’s intent. (Peterson 2006)

In the past two decades this view on archival description has been challenged. On the one hand, archival scholars have indicated that the archival processing of records, including their classification and description, never was an entirely neutral and objective process to begin with. John Ridener’s discussion of the history of archival theory (2009) shows how each ‘paradigm’ in archival science has struggled with the tension between objective methods and subjective decision-making moments. In fact, he argues, almost all steps in the archival process require value judgments that can never be fully objective. For example, at the Netherlands Institute for Sound and Vision, the Dutch broadcasting archive, there is a clear hierarchy in the level of detail of the descriptions: news, actualities and sports are described in great detail, while lesser valued programs like quizzes and reality TV are only described at item-level. (De Jong 2009: 10-11) Archival scholar Terry Cook concludes that, rather than viewing the archivist ‘as a kind of honest broker between the creator of records and the records’ later use by researchers,’ one should recognize that in fact s/he is ‘an active agent shaping the archive, a mediator and interpreter of meaning.’ (Cook 2009: xv-xvii)

On the other hand, the view on archival description has been challenged by the emergence of electronic records. New ways of communicating information, such as exchanging information via e-mail and linking digital files via Intranet and Internet obscured the connection between the record at hand and the context in which it performed its function. Although in the end the archival task of comprehending and elucidating contextual linkages remains the same, the validation of these digital records involves a more active participation of the archivist. (Cook 1997: 41-42) As David Bearman, a leading expert on the impact of digitization on archives and other heritage institutes, asserts, ‘the important point of these challenges to the traditional document is that the boundaries of the document have given way to a *creative authoring event in which user and system participate*. Only the context in which these virtual documents are created can give us an understanding of their content.’ (quoted in Cook 1997: 42, emphasis added)

An example of this development is the impact of digitization on the workflow of broadcasting archives. With the digitization of television *production*, the place of the archivist has moved from the end of the production chain – taping full programs right after they were broadcast on video tapes – to the centre of the digital production environment, making decisions about archivable content from a virtual and dynamic collection of media objects from which editors ‘publish’ on different platforms, such as television, radio, websites, or mobile phones. As a consequence, the reality of the audiovisual archive has become extremely *dynamic*: as an integral part of the digital workflow the process of archiving no longer has one clear location and takes place throughout all phases of the production process. And the description of content is fed by input of both human (cataloguers)

and non-human actors (such as the data on the time and place of a recording accompanying footage shot by digital cameras). (De Jong 2009: 13)

From Archive to *Archiving*: The Archive as Process

The dynamism of contemporary archival practice is also reflected in the theoretical conception of archiving as a *process*. This conception is part of what literary scholars Astrid Erll and Ann Rigney describe as ‘a larger shift of attention within cultural studies from products to processes, from a focus on discrete cultural artefacts to an interest in the way those artefacts circulate and interact with their environment.’ (2009: 3) In the case of archives this processual approach is based on the idea that the meaning of archival records is located in their *use*, including their use at the various stages of the archival process itself. As archival scholar Eric Ketelaar explains, archival records have performative power, in that they incite actions: ‘Records are not only evidence, they communicate, and through communication they can have performative power, they can accomplish something, make a difference in status before and after.’ These activations or performances of archival records add meaning to them: ‘Each activation adds a branch to [...] the semantic genealogy of the record and the archive.’ (2009: 10 and 7) As such, these activations influence the evidentiary status of archival records.

Media scholar Wolfgang Ernst argues the dynamism of digital archives from a materialist perspective. In the digitized audiovisual archive, the objects are defined in binary code, with the bit as the smallest information unit whose duality allows the archival encoding of words, images, sounds and times. As a consequence, the object in question loses its exclusivity towards other forms of data object, such as the metadata describing the objects. Besides, as Ernst explains, the object is latent and dynamic in that every time it is retrieved it is re-created: ‘Algorithmic objects are objects that always come into being anew and processually; they do not exist as fixed data blocks. It is a question of archiving the source codes with which (...) a new whole can be regenerated – a latent archive.’ (2010: 83-84)

These implications of digitization for the archival object also change the structure of the archive itself. According to Ernst the essence of the digital archive is less the archived material per se than a dynamic conception of the idea of the archive: the classical, file-oriented archival practices yield ‘to the use-oriented (“to be completed”) “dynarchive”.’ (81) A key change in the archival process is that it no longer focuses on the content of the files, but rather on creating meaningful links between them, their ‘logistical interlinking.’ (84-85) This has a great impact on the knowledge created by and stored in the archive:

The archival infrastructure in the case of the Internet is only ever temporary in response to its permanent, dynamic rewriting. Ultimate knowledge (the old encyclopedia model) gives way to the principle of permanent rewriting or addition (Wikipedia). The memory spaces geared to eternity are replaced by series of temporally limited entries with internal expiry dates that are as reconfigurable as the rhetorical mechanisms of the *ars memoriae* once were. (86)

Digitization challenges traditional ideas about the role of the archive as a repository of knowledge in two important ways. First, the idea that archival records are relatively stable carriers of meaning has yielded to a view of records as having a lifecycle of use that influences what they come to mean. Second, this usage also changes because of the wider accessibility of digitized archival collections online. Many archives currently experiment with *tagging*: allowing general users to assign and evaluate index terms to digital still or moving images and sound collections at an online platform, often in the form of a game. In this way general users help archives to describe parts of the collection and thus increase the

collection's searchability. The implementation of these user-generated descriptions in the archive's database also influences the status of the metadata as sources for the evidentiary and informational value of the records. In the next sections I investigate the implications of these experiments for the status of archives as gatekeepers of knowledge.

Tagging Audiovisual Content

Tagging is a form of annotating audiovisual content. It's an example of *crowdsourcing* in that it uses 'the crowd' to 'outsource' the creation of metadata on parts of the archive's holdings. The main difference between professional annotation and tagging is that the professionals use a controlled vocabulary, such as a thesaurus, whereas the general users can assign any term they like, not being bound to an existing vocabulary or taxonomy. The latter principle is described as 'folksonomy' (a conflation of 'folk' and 'taxonomy'), an ordered set of categories that emerges from how people tag items. (Weinberger 2008: 165)

Tagging is considered particularly useful for large collections of audiovisual content. Creating detailed descriptions for audiovisual content is highly labor intensive and time consuming: depending on the level of detail a professional annotator needs between one and four hours to describe one hour of content. As a consequence, professional cataloguers at audiovisual archives usually make detailed descriptions of select, highly valued categories of material (news, actualities and sports) and create only general, item-level descriptions of other categories (such as talk shows, reality TV programs, quizzes and other entertainment programs). Users, however, would like to have as many detailed descriptions of content as possible. For media producers, the main user group of broadcasting archives, detailed, shot-level descriptions of all content would make their search for relevant footage much more effective and efficient.

Heritage institutions have different motives for engaging in tagging. One motive is to stimulate general user interaction with heritage that previously was harder to access. Besides, tags can help to close what is known as the *semantic gap*: the difference between the keywords assigned to objects by a professional annotator (usually from a controlled vocabulary) and the search terms the general public uses for referring to or finding the same document.¹ The assumption is that general users will assign tags that they would also use for retrieving the same content through a search query. Finally, tagging can be a means to enrich the database with factual and contextualized information. (Oomen et al. 2010)

The involvement of various other actors in the production of metadata on archival content has significant consequences for the status of the archives as gatekeepers of knowledge. The implementation of user-generated metadata potentially entails a shift from *authoritative knowledge* (metadata created and managed by professionally trained cataloguers) to *knowledge by consensus* (most votes count). This democratic conception of knowledge is a key feature of all Web 2.0 technologies, with Wikipedia as the best known example.

Although this democratization of knowledge is welcomed by some, there is also a lot of resistance among traditional information professionals. As information resources specialist Elaine Peterson argues the underlying philosophy behind folksonomies is philosophical relativism, in that objects are no longer assigned a fixed place in pre-defined categories based on the author's intent but are instead classified according to their meaning as perceived by individual users. Because user-generated tags can be inaccurate, irrelevant and inconsistent, Peterson fears a 'breakdown' of the retrieval system. (Peterson 2006) Adverts of social tagging by contrast celebrate this breakdown of the traditionally authoritative archival

¹ The term originates from the domain of automatic image retrieval in computer science, where it indicates the 'lack of coincidence between the information that one can [automatically, JN] extract from the visual data and the interpretation that the same data have for a user in a given situation.' (Smeulders et al. 2000; see also Jørgensen 2007)

structure and advocate the advent of truly ‘participatory archives’ that are ‘based on an understanding that together the participants are more knowledgeable about the archival materials than an archivist alone can be.’ (Huvila 2008: 26)

In order to investigate to what extent these experiments actually change the type of knowledge associated with the audiovisual archive I now analyze one specific case of social tagging: the Video Labeling Game *Waisda?*

Video Labeling Game *Waisda?*

Waisda? (slang for the Dutch equivalent of ‘What’s that?’) is a multi-player Video Labeling Game, developed by the Netherlands Institute for Sound and Vision, the Dutch national broadcasting archive, as part of the large-scale digitization project Images for the Future.² It is an example of a ‘game with a purpose’ (GWAP): a computer game in which people, as a side effect of playing, perform tasks computers are unable to perform. (Gligorov et al. 2010: 2; von Ahn 2006) The model for *Waisda?* was the first online tagging game: the *ESP* (Extra Sensory Perception) *game*, where randomly provided images have to be tagged in a short time span by two randomly coupled players who have to guess what words the other player will assign (hence the name): points are assigned on the basis of matching tags.³ *Waisda?* uses this principle for the tagging of videos. Earlier examples of annotating video are the *Yahoo! Video Tag Game* and *Popvideo*.⁴

Waisda? allows various players to simultaneously assign keywords to pre-selected episodes of television programs – in this pilot phase full episodes of the reality TV show *Farmer Seeks a Wife* (provided by the public broadcaster KRO) and of a popular Dutch talk show, *Barend and Van Dorp* (which from the early 1990s until 2005 was screened by the commercial broadcaster RTL and in 2009 was donated to the broadcasting archive). Players can choose among four videos. They generate points by assigning tags that were also assigned by other players and there are bonus points for inventing tags that are picked up by others soon after by other players – the shorter the time span, the more points awarded. Players always play against at least one other player; if no other player is present, the single player is coupled to a *bot*, a computer player that simulates a real one by ‘submitting’ the tags entered by players in previous games for the same video (the bot is indicated with #, see van Ees 2010). Registered players can save their score which is ranked in one of four score lists on the main page.

The aim of the game is in line with the motivations of archives for using tagging outlined above: to generate more detailed descriptions of full television programs and to get better matches between search queries and results. The pilot of *Waisda?* has been very successful. From May to November 2009, 2,296 players added 340,551 tags to 604 items. In total forty percent of the tags was matched within ten seconds by two or more players. (Baltussen et al. 2010: 4-5) The first evaluations of the quality of the tags seem to indicate that they indeed complement the professional metadata for audiovisual content and thus can help to bridge the semantic gap. (Gligorov et al. 2010: 6) Besides, the evaluation of the *Waisda?* tags by a professional senior cataloguer showed that the type of content influences the usefulness of the tags: programs with a great variety of topics, such as actuality programs, are tagged with more specific terms than programs that focus on one topic, such as the reality series *Farmer Seeks a Wife*. This research also showed that in order to generate sufficient tags the game needs to attract a large group of players, inviting small adjustments to game design.

² www.waisda.nl; www.imagesforthefuture.com. After a first run of the pilot in 2009-2010 *Waisda?* is currently being redesigned and therefore cannot be played – the website directs visitors to the blog with information on the game (in Dutch): <http://blog.waisda.nl/>.

³ <http://www.gwap.com/gwap/gamesPreview/espgame/>

⁴ <http://videotaggame.sandbox.yahoo.com/> and www.gwap.com.

(Baltussen et.al 2010; van Ees 2010) In the new version of the game these issues will be taken into account. Eventually, the user-generated tags will be implemented in the Institute's catalogue.

Besides *Waisda?* the Netherlands Institute for Sound and Vision has also developed a wiki on their collection: the 'Beeld en Geluidwiki'.⁵ This is a similar experiment with *crowdsourcing* metadata on parts of their collection – in this case contextual data on people and programs from Dutch media history.

Consequences: From Authoritative Knowledge to 'Knowledge by Consensus'?

The evaluations of the first experiments with tagging cultural heritage show that it can be a very successful strategy for generating annotations that complement the professional descriptions of the heritage institutions. Users tend to create more subjective, associative tags that allow for searches on unorthodox elements, such as the colors, background, and specific details of paintings. (Trant 2009) In the case of audiovisual heritage, taggers focus mostly on what they directly see and here, whereas professional cataloguers assign more general, abstract terms borrowed from a thesaurus. (Gligorov et al. 2010: 5) When certain conditions are met, user-generated metadata can thus help to generate more information and contextualization and improve the retrievability of cultural heritage. But what are the consequences of this introduction of 'participatory knowledge' for the status of the archive as a repository of knowledge?

The first examples of including user-generated tags in the catalogues of heritage institutions show that archives will not obliterate the differences between professional and non-professional data. For example, at the online collection database of the Powerhouse Museum Sidney the 'user tags' are visually separated from the descriptions made by the professional cataloguers.⁶ The left-hand column presents objects descriptions made by the museum's own cataloguers (who, by the way, are not identified). The user-tags are listed in the right-hand column, indicated in a different color, and with an option – indicated by a red cross – to flag inappropriate terms for deletion. This right-hand column further contains categories such as 'related subjects', 'theme containing this object', 'parent object', 'this object belongs to [this and that collection, JN]', 'similar objects' (based on the formal classification categories of the object in question) and a list of 'auto-generated tags' (produced by passing the object record through the OpenCalais service from Reuters, a service that makes a 'best guess' at the names, places, companies and specialist terminology that is in the text of the object records – accompanied with the same option of flagging for deletion). This hierarchy in the presentation of different types of metadata demonstrates that the infrastructures of authoritative knowledge production are still firmly in place. (See Beaulieu and de Rijcke 2010)

Also, the first evaluations of the tagging projects at cultural heritage institutions indicates that the second generation of tagging projects will implement a higher level of steering by the institution, such as providing faceted thesauri that can help users to disambiguate tags. An example is the T3 project of *steve.museum: Text, Terms, Trust*, which 'will focus on ways in which users can disambiguate tags (for instance by using a faceted thesaurus) and can rank the usefulness of tags, in order to create more structure in the vast amount of collected tags.' (Baltussen 2010: 28)⁷

At the same time, these initiatives might affect the way we think about the status of knowledge. For example, the usefulness and validity of the tags generated through the Video Labeling Game *Waisda?* is judged on the basis of *consensus*: if a tag is assigned more than

⁵ <http://beeldengeluidwiki.nl>

⁶ <http://www.powerhousemuseum.com/collection/database/>

⁷ See also http://steve.museum/index.php?option=com_content&task=blogsection&id=1&Itemid=2

once within 10 seconds, it is considered ‘matched’. So tags that are mutually agreed on are considered as ‘verified’. (Gligorov et al, 2010: 1) This clearly is a form of *knowledge by consensus*: a term is valid if the majority agrees. This might lead to a new conception of knowledge, that is more open and responsive to the knowledge production outside the recognized knowledge institutions. (Nowotny et al. 2003; Fuller 2002)

Yet one may wonder how this will really change expert knowledge production. The principle behind the multiplayer games like *ESP game* and *Waisda?*, where points are scored on the matching of tags, is based on the idea that people will enter *reasonable* keywords in order to have any chance on agreeing on one. (Gligorov et al. 2010: 2) Also, the first evaluation of the tags generated through *Waisda?* reveals that the user-generated tags greatly differ from the keywords from the thesaurus used by the archive, but that this, besides from mistakes or sloppiness in the tags can also be explained from the fact that these tags are simply *different*, and thus may *complement* the keywords made by professionals using a controlled vocabulary. (Gligorov et al. 2010: 4, see also Baltussen et al. 2010) Users tend to focus on other, more subjective elements of pictures or videos than professional cataloguers do. (Trant 2009) In the case of *Waisda?* taggers focused on what they saw and heard at specific moments within the program, rather than on the more general subject categories used for longer sections of a program by the experts. (Gligorov et al. 2010: 5) So user-generated tags are located on a different, lower semantic level, closer to what actually is seen and heard in the videos.

Conclusion

To what extent, then, do experiments with participatory culture, like the ones with crowdsourcing that I discussed, really transform the archive?

In theory, the digitization of the archive, both of its holdings and of its workflow, implies the dissolution or even ‘liquidation’ of the classical concept of archive. As Wolfgang Ernst argues, the real ‘archive’ in the Internet is a system of technological protocols, while the collection of objects is latent, making its meaning dynamic and subject to permanent rewriting and addition. (2009: 87) Seen from this perspective the digitization of archival holdings has great consequences for their evidentiary value: ‘The testimonial function of archival records was once firmly rooted in their material authenticity. (...) With the digital, physical signals become *information*. The intrinsic value of the documents yields to their media-technological nature, consisting of alphanumeric and hardware.’ (90-91)

In the practice of institutional archives, though, it appears the infrastructures of authoritative knowledge production are much more hardwired than could theoretically be expected. In my view, even though these new forms of access may involve the introduction of a new type of participatory knowledge in the archive, this does not signal a radical break with the previous practice and leaves the epistemology of the archive in tact. In fact, the effects of digitization on the archival workflow only show that the archival process has always been characterized by multiple authorship and subjective moments of appraisal. In the end, digitization only exposes the archives’ inherently dynamic, performative nature.

Bibliography

Ahn, Louis Von. 'Games With a Purpose.' *IEEE Computer* 39 (2006) 6: 92-94.

Baltussen, Lotte Belice, Maarten Brinkerink and Johan Oomen. *Evaluatieverslag Waisda? Taggen van Audiovisueel Materiaal in Spelvorm*. Hilversum: Nederlands Instituut voor Beeld en Geluid, 2010.

http://maartenbrinkerink.net/temp/waisda/Rapport_Waisda_DEF.pdf (accessed 5 May 2011).

Baltussen, Lotte Belice. *Barbarians versus Gatekeepers? Tagging as a Way of Defining the Emergent Living Archive Paradigm*. Master's thesis, University of Amsterdam, 2010. <http://dare.uva.nl/scriptie/339629> (accessed 5 May 2011).

Beaulieu, Anne and Sarah de Rijcke. 'The Enactment of Museum Collections in Digital Infrastructures: Authorizing Actors, Configuring Expertise.' Paper presented at the conference *Reimagining the Archive: Remapping and Remixing in the Digital Age*, UCLA, November 11-14, 2010. Available from http://polaris.gseis.ucla.edu/reimagining/Papers/Panel33_DeRijckeBeaulieu.pdf (accessed 6 May 2011).

Cook, Terry. 'What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift'. *Archivaria* 43 (1997): 17-63.

Cook, Terry. 'Preface.' In John Ridener, *From Polders to Postmodernism: A Concise History of Archival Theory*. Duluth, MN: Litwin Books, 2009: xi-xix.

Ees, Annelies van. *How to Tag 700.000 videos? The Power of Games*. Master's thesis, VU University Amsterdam, 2010.

Erll, Astrid and Ann Rigney, eds. *Mediation, Remediation and the Dynamics of Cultural Memory*. Berlin: De Gruyter, 2009.

Ernst, Wolfgang. 'Underway to the Dual System: Classical Archives and/or Digital Memory.' In Dieter Daniels and Gunther Reisinger, eds. *Netpioneers 1.0: Contextualizing Early Net-based Art*. Berlin: Sternberg Press, 2010: 81-99.

Fuller, Steve. *Knowledge Management Foundations*. Boston etc.: Butterworth-Heinemann, 2002.

Gligorov, Riste, Lotte Belice Baltussen, Jacco van Ossenbruggen, Lora Aroyo, Maarten Brinkerink, Johan Oomen and Annelies van Ees. 'Towards Integration of End-User Tags with Professional Annotations.' *Proceedings of the Web Science Conference 2010: Extending the Frontiers of Society On-Line*, April 26-27th, 2010, Raleigh, NC, USA. http://journal.webscience.org/363/2/websci10_submission_65.pdf (accessed 5 May 2011).

Horsman, Peter. 'Taming the Elephant: An Orthodox Approach to the Principle of Provenance.' *The Principle of Provenance: Report from the First Stockholm Conference on Archival Theory and the Principle of Provenance 2-3 September 1993*. Stockholm: Svenska Riksarkivet, 1994: 51-60.

ICA- ISAD(G), *General International Standard Archival Description, Second Edition*. Ottawa: ICA, 2000. Available from [http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf) (accessed 5 May 2011).

Jørgensen, Corinne. 'Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift.' In Joan Lussky, ed. *Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*. The University of Arizona Digital Library of Information Science & Technology (DLIST), 2007. <http://hdl.handle.net/10150/105339> (accessed 6 May 2011)

Ketelaar, Eric. 'The Performative Power of Records.' Unpublished paper, presented in Japan, 2009.

Muller, S., J.A. Feith and R. Fruin. *Manual for the Arrangement and Description of Archives*. Translated by Arthur H. Leavitt. Chicago: The Society of American Archivists, 2003 [1898].

Nowotny, Helga, Peter Scott and Michael Gibbons. 'Introduction: 'Mode 2' Revisited: The New Production of Knowledge.' *Minerva* 41 (2003): 179-194.

Oomen, Johan and Lora Aroyo. 'Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges.' Paper presented at the 5th *International Conference on Communities and Technologies, Queensland University of Technology*, Brisbane, Australia, 29 June-2 July 2011.
<http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf> (accessed 5 May 2011).

Oomen, Johan, Lotte Belice Baltussen, Sander Limonard, Annelies van Ees, Maarten Brinkerink, Lora Aroyo, Just Vervaart, Kamil Asaf and Riste Gligorov. 'Emerging Practices in the Cultural Heritage Domain: Social Tagging of Audiovisual Heritage.' *Proceedings of the Web Science Conference 2010: Extending the Frontiers of Society On-Line*, April 26-27th, 2010, Raleigh, NC, USA.
http://journal.webscience.org/337/2/websci10_submission_23.pdf (accessed 5 May 2011).

Peterson, Elaine. 'Beneath the Metadata: Some Philosophical Problems with Folksonomy.' *D-Lib Magazine* 12 (2006) 11: <http://www.dlib.org/dlib/november06/peterson/11peterson.html> (accessed 11 March 2010).

Prelinger, Rick. 'The Appearance of Archives.' In Pelle Snickars and Patrick Vonderau, eds. *The YouTube Reader*. Stockholm: National Library of Sweden, 2009: 268-274.

Ridener, John. *From Polders to Postmodernism: A Concise History of Archival Theory*. Duluth, MN: Litwin Books, 2009.

Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta and R. Jain. 'Content-based Image Retrieval at the End of the Early Years.' *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000) 12: 1349 – 1380.

Trant, Jennifer. *Tagging, Folksonomy and Art Museums: Results of steve.museum's Research – 31 January 2009*, steve.museum/Archives & Museums Informatics, 2009:
<http://conference.archimuse.com/files/trantSteveResearchReport2008.pdf> (accessed 30 March 2010).

Weinberger, David. *Everything is Miscellaneous: The Power of the New Digital Disorder*. New York: Henry Holt and Company, 2008.

Winget, Megan. "'It was just in the air, at any place...'" The Principle of Provenance & Modern Archival Systems.' Unpublished paper, 7 March 2005: 1. Available from
<http://www.unc.edu/~winget/research/DigitalProvenance.pdf>, accessed 26 May 2010.