Julia Noordegraaf*

Crowdsourcing Television's Past

The State of Knowledge in Digital Archives

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.[1] 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 article I investigate the implications of these developments for the status of the (audiovisual) archive as a gatekeeper of knowledge. I do so by analyzing a recent experiment with the online annotation or *social tagging* of audiovisual content: the Netherlands Institute for Sound and Vision's video labeling game WAISDA?. This example of *crowdsourcing* – 'outsourcing' tasks traditionally performed by professionals to the general public or 'the crowd' – is a particularly apt example to study how digitization affects the epistemology of the archive, since it introduces untrained professionals in the core of the archive's knowledge-base: the catalogue descriptions containing contextual data about the production, distribution and use of the collection items. These so-called *metadata* are crucial for retrieving archival holdings and for interpreting them: without knowledge about who created a document, when and for what purpose, the document becomes meaningless. Based on a discussion of the first evaluations I argue that social tagging projects in the heritage

---

*   Julia Noordegraaf is universitair hoofddocent bij het departement Mediastudies van de Universiteit van Amsterdam, waar ze ook het internationale masterprogramma Preservation and Presentation of the Moving Image coördineert. Ze is de auteur van *Strategies of Display: Museum Presentation in Nineteenth- and Twentieth-Century Visual Culture* (NAi Publishers/Museum Boijmans Van Beuningen, 2004) en eerste redacteur van de bundel *Preserving and Exhibiting Media Art: Challenges and Perspectives* (Amsterdam University Press, verschijnt in 2012). In haar huidige boekproject (Performing the Archive: Tracing Audiovisual Heritage in the Digital Age) onderzoekt ze de impact van virtuele archieven, online portals en andere vormen van hergebruik van audiovisueel erfgoed op de status van het audiovisueel archief als kennisbewaarder. Daarnaast ontwikkelt ze onderzoeksprojecten op het terrein van media en onrecht (The Audiovisual Memory of (In)Justice) en de conservering van computergebaseerde kunst.

sector entail a new form of *participatory knowledge* that allows for alternative searches and connections between archival holdings.

In order to understand how participatory knowledge affects the status of the archive as a knowledge keeper, I start with a discussion of the principles underlying the authoritative knowledge traditionally associated with archives. I describe how archival description is based on principles of transparency and objectivity, using hierarchically structured, standardized datasets for the description of archival records. I then argue that in the past decade this view on description has been challenged, on the one hand by changing ideas on the supposed neutrality of archival practice and on the other by the specific nature of electronic records, which for their description require a more active role of the archivist as an interpreter of the meaning of the object in question.

A subsequent discussion of Media Scholar Wolfgang Ernst's ideas on the impact of digitization on the archive serves to show how, at a fundamental level, digitization changes not only the nature of the archival record – transforming an originally material object in a latent source code – but therewith also the way these records are contextualized and described. Ernst argues that digital objects can be interrelated in diverse and constantly changing ways, therewith exchanging the relatively stable model of encyclopedic knowledge for the dynamic and interactive one of Wikipedia. The latent and dynamic nature of digital objects and metadata is the condition for the social tagging experiments by heritage institutions, which are only possible in an online environment and which aim to make cultural heritage accessible and interpretable in new and different ways, involving the users-participants in the process.

Ernst's argument focuses mostly on the ontology of the archive and its holdings. In my view, the knowledge associated with archives, besides relating to the records themselves and the metadata describing them, is also the product of concrete institutional practices. All phases of the archival process, from selection, description and preservation to access and reuse, add meaning to the object, and thus influence its evidentiary status. In the final section of the paper, then, I analyze a social tagging project as realized in institutional practice. I outline the motivations of heritage institutions to engage with this type of crowdsourcing and describe and analyze the results of WAISDA?. In order to judge to what extent these experiments affect the authoritative knowledge associated with heritage institutions I also study the first implementation of these 'user-generated metadata' by the Power House Museum in Sidney in their online catalogue. Based on the analysis of this case study I finally argue that, even though these experiments introduce a new type of knowledge in the archive, at an institutional level the infrastructure for authoritative knowledge production is left intact.

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 and 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.[2] From the publication of the first archiving handbook, known as the Dutch *Manual*, 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.[3] 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.[4]

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.[5]

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.[6]

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 shows how each 'paradigm' in archival science has struggled with the tension between objective methods and subjective decision-making moments.[7] 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.[8] 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'.[9]

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 email and linking digital files via Intranet and Internet, obscure 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.[10] 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.'[11]

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 on video tapes right after they were broadcast – 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).[12]

Towards a Latent, Dynamic Digital Archive

According to media scholar Wolfgang Ernst digitization profoundly affects both the ontology and epistemology of the (audiovisual) archive. 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 objects, 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.'[13]

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"'.[14] 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'.[15] This, thus Ernst, 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'.[16]

So in Ernst's view, the inherent instability of the digital archive, with its latent collection of source codes, also destabilizes its function as a repository of knowledge. He describes the digital archive as a collection of data that can be retrieved in numerous combinations, in the process changing the knowledge associated with them. Notwithstanding the rigor in thinking through the implications of the digital object at an ontological level, this materialist perspective seems to ignore the fact that an archive, besides being a collection of objects and information, also is an institution, with a long tradition of developing standards and procedures for the validation and description of its holdings. To what extent does the involvement of untrained users in archival description in tagging experiments in real institutions actually change these practices of knowledge production?

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.[17]

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 five 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.[18] 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.[19]

The involvement of various other actors in the production of metadata on archival content potentially entails a change in the type of knowledge associated with the archive, whereby the *authoritative knowledge* traditionally associated with the archive (metadata created and managed by professionally trained cataloguers) is being replaced by *participatory knowledge* (metadata created and managed by different types of actors, human and technological, trained and untrained). 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 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

Screen of the labeling game WAISDA?

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.[20] Adverts of social tagging by contrast celebrate this breakdown of the traditionally authoritative archival 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.'[21]

In order to investigate to what extent these tagging 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.[22] 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.[23] 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.[24] WAISDA? uses this principle for the tagging of videos. Earlier examples of annotating video are the YAHOO! VIDEO TAG GAME and POPVIDEO.[25]

WAISDA? allows various players to simultaneously assign keywords to pre-selected episodes of television programs – in the first pilot phase full episodes of the reality TV show BOER ZOEKT VROUW (Farmer Seeks a Wife, provided by the public broadcaster KRO) and of a popular Dutch talk show, BAREND EN 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), and in the current, redesigned version of the game episodes from the *faits divers* program MAN BIJT HOND. Players can choose among five 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. Registered players can save their score which is subsequently ranked in the top score list 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.[26] 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.[27] 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 BOER ZOEKT VROUW. 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.[28] In the new version of the game these issues have be taken into account, in particular by providing a greater variety of shorter clips. 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 its collection: the 'Beeld en Geluidwiki'.[29] 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.

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.[30] 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.[31] 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.[32] 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]', '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.[33]

Also, the first evaluations of the tagging projects at cultural heritage institutions indicate 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.'[34]

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 once within 10 seconds, it is considered 'matched'. So tags that are mutually agreed on are considered as 'verified'.[35] 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.[36]

Yet one may wonder how this will really change expert knowledge production. The principle behind 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.[37] 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.[38] Users tend to focus on other, more subjective elements of pictures or videos than professional cataloguers do.[39] 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.[40] 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 as a repository of knowledge? It is clear that the tagging experiments generate a different type of description that remains close to what can be directly seen and heard in the videos. This makes user-generated tags particularly appropriate for retrieving *stock shots*: shots with a general subject matter – sunset, camping site, market scene – that can be re-used as illustration of items set in similar scenes, or for identifying specific people and places at the specific moments they occur in the clip. The fact that they lack the more abstract terms used in controlled vocabularies, such as the indication of a program's genre or topic chosen from a hierarchically ordered list, implies that they are less useful for interpreting the content of the item in question in a the broader social and cultural historical framework in which it was produced, distributed and seen. The low semantic level on which user tags are located allows for establishing connections between objects in more subjective areas, such as colours, atmospheres or emotions – areas that general users also borrow their search terms from. And the interactive nature of the tagging projects, such as the option to delete and replace

tags in the Power House Museum catalogue, does evoke the principle of permanent rewriting and edition that Wolfgang Ernst identifies as characteristic of digital archives. In this sense social tagging does introduce a new form of participatory, democratic knowledge into the archive.

Yet the analysis of the first implementations of the results of these tagging experiments has shown that this new form of knowledge does not replace the infrastructure of authoritative knowledge production and distribution traditionally associated with the archive. The Power House Museum presents the user tags in a separate column, clearly separated from the professional catalogue description which, contrary to the user tags, cannot be revised. In addition, the new generation of tagging projects in heritage institutions will demonstrate a higher level of steering, providing users with closed vocabularies in order to stimulate them to create less ambiguous tags. Here, the archive firmly confirms its custodial role, training the users in working with professional standards. So where from a materialist point of view digitization leads to a highly unstable collection of archival holdings, a 'latent archive' in Ernst's terms, in practice the infrastructure of archival institutions proves more hardwired than can theoretically be expected. 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 nature.

Noten

1   R. Prelinger, 'The Appearance of Archives', in: P. Snickars and P. Vonderau (eds.), *The YouTube Reader*, National Library of Sweden, Stockholm 2009, p. 269-270.
2   T. Cook, 'What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift', in: *Archivaria* vol. 43, 1997, p. 18.
3   Cook, 'What is Past is Prologue', in: S. Muller, J.A. Feith and R. Fruin, *Manual for the Arrangement and Description of Archives*, translated by Arthur H. Leavitt, The Society of American Archivists, Chicago 2003 [1898], p. 21.
4   P. Horsman, 'Taming the Elephant: An Orthodox Approach to the Principle of Provenance', in: The Principle of Provenance: Report from the First Stockholm Conference on Archival Theory and the Principle of Provenance, 2-3 September 1993, Svenska Riksarkivet, Stockholm 1994, p. 51-60 and M. Winget, '"It was just in the air, at any place…": The Principle of Provenance & Modern Archival Systems', unpublished paper, 7 March 2005, p. 1. Available from http://www.unc.edu/~winget/research/Digital Provenance.pdf, accessed 26 May 2010.
5   ICA- ISAD(G), *General International Standard Archival Description. Second Edition*, ICA, Ottawa 2000, p. 7, emphasis added. Available from http://www.icacds.org.uk/eng/ISAD(G).pdf, accessed 5 May 2011.
6   E. Peterson, 'Beneath the Metadata: Some Philosophical Problems with Folksonomy', in: *D-Lib Magazine*, 11, vol. 12, 2006, http://www.dlib.org/dlib/november06/peterson/11peterson.html, accessed 11 March 2010.
7   J. Ridener, *From Polders to Postmodernism. A Concise History of Archival Theory*, Litwin Books, Duluth, MN 2009.
8   A. de Jong, *De nieuwe toegang tot audiovisuele content. Ontwikkelingen in de netwerkcultuur*, Nederlands Instituut voor Beeld en Geluid, Hilversum 2009, p. 10-11.
9   T. Cook, 'Preface', in: Ridener, *From Polders to Postmodernism*, p. xv-xvii.

10 Cook, 'What is Past is Prologue', p. 41-42.

11 Quoted in Cook, 'What is Past is Prologue', p. 42, emphasis added.

12 De Jong, *De nieuwe toegang*, p. 13.

13 W. Ernst, 'Underway to the Dual System: Classical Archives and/or Digital Memory', in: D. Daniels and G. Reisinger (eds.), *Netpioneers 1.0. Contextualizing Early Net-based Art*, Sternberg Press, Berlin 2010, p. 83-84.

14 Ernst, 'Underway to the Dual System', p. 81.

15 Ibidem, p. 84-85.

16 Ibidem, p. 86.

17 D. Weinberger, *Everything is Miscellaneous. The Power of the New Digital Disorder*, Henry Holt and Company, New York 2008, p. 165.

18 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] extract from the visual data and the interpretation that the same data have for a user in a given situation'. A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, 'Content-based Image Retrieval at the End of the Early Years', in: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, vol. 22, 2000, p. 1349–1380. See also C. Jörgensen, 'Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift', in: J. 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 & Technolgy (DLIST), 2007, http://hdl.handle.net/10150/105339, accessed 6 May 2011.

19 J. Oomen, L.B. Baltussen, S. Limonard, A. van Ees, M. Brinkerink, L. Aroyo, J. Vervaart, K. Asaf and R. Gligorov, 'Emerging Practices in the Cultural Heritage Domain: Social Tagging of Audiovisual Heritage', in: Proceedings of the Web Science Conference 2010. Extending the Frontiers of Society On-Line, 26-27 April 2010, Raleigh, NC, USA, http://journal.webscience.org/337/2/websci10_submission_23.pdf, accessed 5 May 2011.

20 Peterson, 'Beneath the Metadata'.

21 I. Huvila, 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management', in: *Archival Science*, 1, vol. 8, 2008, p. 26.

22 www.waisda.nl; www.imagesforthefuture.com. After a first run of the pilot in 2009-2010 WAISDA? has been redesigned and was re-launched in October 2011 as MAN BIJT HOND WOORDENTIKKERTJE: http://woordentikkertje.manbijthond.nl. The new version retains much of the functionalities of the original pilot. More information on both versions can be found at the blog (in Dutch): http://blog.waisda.nl.

23 R. Gligorov, L.B. Baltussen, J. van Ossenbruggen, L. Aroyo, M. Brinkerink, J. Oomen and A. van Ees, 'Towards Integration of End-User Tags with Professional Annotations', in: *Proceedings of the Web Science Conference 2010. Extending the Frontiers of Society On-Line*, 26-27 April 2010, Raleigh, NC, USA, http://journal.webscience.org/363/2/websci10_submission_65.pdf, accessed 5 May 2011, p. 2 and L. Von Ahn, 'Games With a Purpose,' in: IEEE *Computer*, 6, vol. 39, 2006, p. 92-94.

24 www.gwap.com/gwap/gamesPreview/espgame.

25 http://videotaggame.sandbox.yahoo.com/ and www.gwap.com.

26 L.B. Baltussen, M. Brinkerink and J. Oomen, *Evaluatieverslag Waisda? Taggen van Audiovisueel Materiaal in Spelvorm*, Nederlands Instituut voor Beeld en Geluid, Hilversum 2010, p. 4-5. Available from http://maartenbrinkerink.net/temp/waisda/Rapport_Waisda_DEF.pdf, accessed 5 May 2011. In the new version, MAN BIJT HOND WOORDENTIKKERTJE, it goes even faster: during the test phase over 300.000 thousand tags have been assigned in three months, of which over 170.000 were matched.

27 Gligorov et al., 'Towards Integration of End-User Tags', p. 6.

28 Baltussen, Brinkerink and Oomen, *Evaluatieverslag Waisda?* and A. van Ees, *How to Tag 700.000 videos? The Power of Games*, master's thesis, VU University Amsterdam, 2010.

29 http://beeldengeluidwiki.nl.

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

31 Gligorov et al., 'Towards Integration of End-User Tags', p. 5.

32 http://www.powerhousemuseum.com/collection/database.

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

34 L.B. Baltussen, *Barbarians versus Gatekeepers? Tagging as a Way of Defining the Emergent Living Archive Paradigm*, master's thesis, University of Amsterdam, 2010, p. 28. http://dare.uva.nl/scriptie/339629, accessed 5 May 2011. See also http://steve.museum/index.php?option=com_content&task=blogsection& id=1&Itemid=2.

35 Gligorov et al., 'Towards Integration of End-User Tags', p. 1.

36 H. Nowotny, P. Scott and M. Gibbons, 'Introduction: 'Mode 2' Revisited: The New Production of Knowledge', in: *Minerva*, vol. 41, 2003, p. 179-194 and S. Fuller, *Knowledge Management Foundations*, Butterworth-Heinemann, Boston etc., 2002.

37 Gligorov et al., 'Towards Integration of End-User Tags', p. 2.

38 Ibidem, p. 4. See also Baltussen, Brinkerink and Oomen, *Evaluatieverslag Waisda?*.

39 Trant, *Folksonomy and Art Museums*.

40 Gligorov et al., 'Towards Integration of End-User Tags', p. 5.