

Authenticity in open data: contributions from archival science

Erik Borglund

Mid Sweden University
851 70 Sundsvall, SWEDEN
erik.borglund@miun.se

Tove Engvall

Mid Sweden University
851 70 Sundsvall, SWEDEN
tove.engvall@miun.se

Keywords: Archival science, Authenticity, Data quality, Open data, Public records, Trustworthiness, User satisfaction

Introduction

Open data in e-government has its foundation in the Public Sector Information Directive (PSI-Directive) from 2003, in which the EU set out regulations about re-use of public information. The PSI-directive is driven by both economic aspects and citizens' rights of access to public information (European Parliament 2003). Since 1766 the Swedish citizens have, grounded in the constitution, free access of public records. In Sweden the work on open data has its background in the E-delegation, a governmental initiative (E-delegationen 2009; E-delegationen 2010a; E-delegationen 2010b). Free access to public records in Sweden is governed by public archives. One can say that open data and reuse of open data is something citizens in Sweden have long experienced, i.e. through access to and use of public records. However the open data initiatives in Sweden are currently more focused on stimulating innovation and building new "apps" which differ from archives tasks.

In the DeLone & McLean IS success model (DeLone and McLean 1992; DeLone and McLean 2002; DeLone and McLean 2003) information quality is seen as a variable dependent on the success of an information system, which requires that the "open data" which is meant to be usable must also be of high quality, but also the information systems in which the data are created and also the metadata about the open data, because high quality is dependent on these two factors. Historically the archive is the source of trustworthy records which can and will be used by others due to its quality (McKemmish et al. 2005; Thomassen 2001).

In Sweden the PSI directive aims to extend the use of public records (Finansdepartementet 2009), and is regulated by an Act (Riksdagen 2010) that should make it easier for actors to re-use public information for innovations. Currently there are some initiatives where archival institutions take an active role in ensuring access to open data. For example, the Stockholm City Archive is one archival institution that argues for the opportunity archives have to support the establishment of new open data initiatives. The benefits of open data are presented by the Stockholm City Archive as¹:

- It should support economic growth, by enabling improved public data in new information services
- Make it easier for companies to use public data
- Increase the benefits for citizens and make public business more efficient

In 2016 the National Archives in Sweden was given a commission to coordinate the available public open data, which can be seen as an officially defined relationship between open data and archival institutions. The Stockholm City Archives for example is building their e-archive as the information/data hub for the whole Stockholm City administration. The archive will by that be the information storage of all public data in Stockholm.

¹ <http://dataportalen.stockholm.se/dataportalen/>

In this article we discuss and argue that records quality criteria are also relevant to open data initiatives. The objective of this article is to present the basis for a study of the success and usefulness of open data in relation to quality criteria for records, which would implicitly motivate an even stronger relation between archival institutions and open data initiatives.

Theoretical basis

This article is based on a perspective that public information can be seen as public records. Records are:

“Information created, received, and maintained as evidence and as assets by any organization or person, in pursuit of legal obligations or in the transaction of business” (International Standards Organization 2016, p. 2)

Records and archives are the two fundamental concepts that form the discipline of archival science (Thomassen 2001), in which the evidential value of a record is central and records are preserved for the evidence they represent (e.g. Saarinen and Sääksjärvi 1990; Schellenberg 1956/1998; Sprehe 2000; Thomassen 2001). In order to have evidential value, records must achieve four other important quality criteria: authenticity, reliability, integrity, and usability. A record can never serve as evidence unless the four quality criteria are attained (e.g. Duranti 2001; International Standards Organization 2016; Reed 2005). However many argue that authenticity is the most important of the four quality criteria for ensuring the evidential value of records.

ISO 15489-1 defines authenticity and reliability as:

“An authentic record is one that can be proven

- a) to be what it purports to be,*
- b) to have been created or sent by the person purported to have created or sent it, and*
- c) to have been created or sent at the time purported”*

(International Standards Organization 2016)

Why authenticity in open data?

Authenticity is important because it makes people trust records (Duranti and Rogers 2012; Lemieux 2016). One of the challenges for modern archives, i.e. those that deal with electronic records, is how to maintain authenticity of the records over time. In the research project InterPARES, and in the Digital Records Forensics Project (Duranti 2009; Duranti and Rogers 2011; Duranti and Thibodeau 2001; Heather MacNeil et al. 2001; Jansen 2010) this was one of the core issues: strategies for ensuring the authenticity over time of records that are born digital is very problematic. For example, in the Digital Records Forensics Project, one of the problems that the research focused on was how to ensure that the authenticity could still be maintained when electronic records are moved out of the original digital context in which they were created (Duranti 2009; Duranti and Rogers 2011; Jansen 2010).

Thurston (2012) argue that well-managed records result in reliable records which are a cornerstone for success in open government. In this article we argue that trustworthy records can increase the potential and usability of open data. It is debatable whether open data that not are records actually exist, and in this article the majority of open data are seen as records. Trustworthiness of records springs from the authenticity and evidential value of records. The trustworthiness of records is often guaranteed by trusted digital repositories in which records are kept (Thurston 2012). The archivist is often seen as the custodian of trustworthy records, who should guarantee that the authenticity of the records is maintained over time and that no one can tamper with the stored records. Open data is about re-use of public data. When the open data is within a public data repository it can be managed and kept so that authenticity is maintained. But when the data is moved outside of the repository, the data will still be treated and interpreted as trustworthy public data but its' authenticity is no longer guaranteed.

The 8 principles of open data: it's about quality

Within the Open Government Working Group's 8 principles of open government data, several quality-related principles are found, which support our argument that archival theory can contribute to work with open data. The 8 principles are:

1. Data Must Be Complete
2. Data Must Be Primary
3. Data Must Be Timely
4. Data Must Be Accessible
5. Data Must Be Machine processable
6. Access Must Be Non-Discriminatory
7. Data Formats Must Be Non-Proprietary
8. Data Must Be License-free

(Open Government Working Group 2007)

From an archival perspective these principals are applicable for public records as well, as we here show: In Sweden the first principle (1) is found in the Swedish Constitution related to the free access to public records. Archives hold primary records and due to adherence to the principle of provenance, an aggregated record can always be broken down into smaller units without breaking links to their creator (2). Modern e-archives capture records as they are born and automatically make them available (3). Records are made to be used and e-archives deal with accessibility (4). The Swedish Constitution makes it possible to access public records while remaining fully anonymous. (6). The long term preservation of records requires open formats to ensure their readability over time (7).

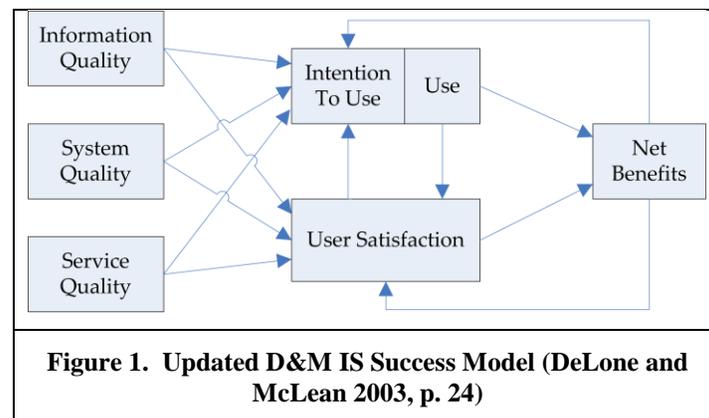
Setting the agenda for a research project.

It is not enough to argue that authenticity is important for the usefulness of open data. In this paper we set out the basis for a study planned to be carried out together with the Stockholm City Archives.

The research study will be inspired by the relationship between quality and user satisfaction found in the information systems success model, by DeLone & McLean (DeLone and McLean 1992; DeLone and McLean 2002; DeLone and McLean 2003). The model is based on the ambition to find and identify “dependent variables” (DeLone and McLean 1992, p. 61) in management information systems (MIS) research. Variables which could measure the outcomes of MIS research i.e. efficiency and success in information systems. By an extensive literature study of more than 100 empirical studies they identified a set of categories indicating information system success:

- System quality
- Information Quality
- Use
- User Satisfaction
- Individual Impact
- Organizational Impact

DeLone & McLean found that the success of an information system cannot be seen from only one or a few of the above categories, and they presented an updated version (DeLone and McLean 2003) in which service quality has been added. The updated model is presented in Figure 1.



In the proposed study we aim to further understand the relationship between the authenticity of records and *the intention to use records, the use of records and user satisfaction*. This will be carried out as a study at the Stockholm City Archives during Spring 2017, using a mixed methods approach (Johnson and Onwuegbuzie 2004) where questionnaires, interviews and data logs will be used as primary data sources. A natural follow up study would be to extend the scope of the research to further investigate whether the authenticity of records (open data) also has impact on the user satisfaction of apps and software developed upon the open data. A relationship based upon the DeLone & McLean IS success model should exist. The latter is not yet planned and designed, since it much depends upon the findings from the Stockholm City Archives study.

Summary

From an archival science perspective public information is records, and basically this is what makes the discussion of authenticity relevant. A record should be authentic even if it has left the control of a government-controlled repository. This is necessary if citizens are to be able to trust re-used information presented by non-government providers in a true open data initiative. By applying an archival science perspective in which authenticity as an important quality criteria we believe that authenticity will also affect user satisfaction with the systems or “apps” developed from open government data. Implicitly we therefore argue that archival science as a scientific domain can make a significant contribution to research on open data, given its history of preserving authentic records over time. Also significant is that much of the debate around open data is about providing access to the data. But around the world many e-archives are set up and an e-archive is nothing more than a data center that can provide secure and trustworthy access to public data, i.e. open data.

We argue therefore that further research on authenticity of open government data can contribute to the already existing archival research in the field of e-government and recordkeeping (e.g. Anderson et al. 2010; Asproth et al. 2010; Borglund et al. 2009; Samuelsson and Öberg 2008; Samuelsson et al. 2007).

References

- Anderson, K., Borglund, E. A. M., and Samuelsson, G. 2010. "Strategies for High Quality Recordkeeping in Public E-Services," *iRMA - Information & Records Management Annual 2010*, pp. 73-82.
- Asproth, V., Borglund, E. A. M., Samuelsson, G., and Öberg, L.-M. 2010. "E-Tjänstens Framtida Historia: Informationsbevarande, Ett Bortglömt Ansvarsområde?," in *Förvaltning Och Medborgarskap I Förändring* K. Lindblad Gidlund, A. Ekelin, S. Eriksén and A. Ranerup (eds.). Lund: Studentlitteratur, pp. 167-184.
- Borglund, E. A. M., Anderson, K., and Samuelsson, G. 2009. "How Requirements of Record Managers Change after Implementing New Electronic Records Management Systems," in: *3rd European Conference on Information Management and Evaluation*. Gothenburg, Sweden.
- DeLone, W. H., and McLean, E. R. 1992. "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research* (3:1), pp. 60-95.
- DeLone, W. H., and McLean, E. R. 2002. "Information Systems Success Revisted," in: *Hawaii International Conference on System Sciences*. Waikoloa, Big Island, Hawaii: IEEE Computer Society.
- DeLone, W. H., and McLean, E. R. 2003. "The DeLone and Mclean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems* (19:4), pp. 9-30.
- Duranti, L. 2001. "Concepts, Principles, and Methods for the Management of Electronic Records," *The Information Society* (17), pp. 271-279.
- Duranti, L. 2009. "From Digital Diplomatics to Digital Records Forensics," *Archivaria* (68), pp. 39-66.
- Duranti, L., and Rogers, C. 2011. "Educating for Trust," *Archival Science* (11:3-4), pp. 373-390.
- Duranti, L., and Rogers, C. 2012. "Trust in Digital Records: An Increasingly Cloudy Legal Area," *Computer Law & Security Review* (28:5), pp. 522-531.
- Duranti, L., and Thibodeau, K. 2001. "The Interpares International Research Project," *The Information Management Journal* (Vol. 35:1), pp. 44-50.
- E-delegationen. 2009. "Statens Offentliga Utredningar. 2009:., Strategi För Myndigheternas Arbeta Med E-Förvaltning : Betänkande." Stockholm: Fritze, pp. 286, [284] s.
- E-delegationen. 2010a. *Så Enkelt Som Möjligt För Så Många Som Möjligt : Från Strategi Till Handling För E-Förvaltning : Betänkande*. Stockholm: Fritze.

- E-delegationen. 2010b. *Så Enkelt Som Möjligt För Så Många Som Möjligt : Under Konstruktion - Framtidens E-Förvaltning : Betänkande*. Stockholm: Fritze.
- European Parliament, C. 2003. "Directive 2003/98/Ec of the European Parliament and of the Council of 17 November 2003 on the Re-Use of Public Sector Information," in: *2003/98/EC T.E.P.a.t.C.o.t.E. Union* (ed.).
- Finansdepartementet. 2009. *Vidareutnyttjande Av Handlingar [Elektronisk Resurs] : Genomförande Av Psi-Direktivet*. Stockholm: Finansdepartementet Fritze [distributör].
- Heather MacNeil, Chen We, Luciana Duranti, Anne Gilliland-Swetland, Maria Guercio, Yvette Hackett, Babak Hamidzadeh, Livia Iacovino, Brent Lee, Sue McKemmish, John Roeder, Seamus Ross, Wai-kwok Wan, and Zhao Zhon Xiu. 2001. "Inter pares Authenticity Task Force Report."
- International Standards Organization. 2016. *Information and Documentation – Records Management – Part 1: Concepts and Principles (Iso 15489-1:2016, Idt)*. Geneva: International Standards Organization.
- Jansen, A. 2010. "Digital Records Forensics: Ensuring Authenticity and Trustworthiness of Evidence over Time," in: *2010 Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*. pp. 84-88.
- Johnson, R. B., and Onwuegbuzie, A. J. 2004. "Mixed Methods Research: A Research Paradigm Whose Time Has Come," *Educational Researcher* (33:7), pp. 14-26.
- Lemieux, V. L. 2016. "Trusting Records: Is Blockchain Technology the Answer?," *Records Management Journal* (26:2), pp. 110-139.
- McKemmish, S., Piggott, M., Reed, B., and Upward, F. (eds.). 2005. *Archives: Recordkeeping in Society* Wagga Wagga: Charles Sturt University, Centre for Information Studies
- Open Government Working Group. 2007. "8 Principles of Open Government Data." Retrieved 2 October, 2012, from <http://www.opengovdata.org/home/8principles>
- Reed, B. 2005. "Records," in *Archives: Recordkeeping in Society* S. McKemmish, M. Piggott, B. Reed and F. Upward (eds.). Wagga Wagga: Charles Sturt University, Centre for Information Studies, pp. 101-130.
- Riksdagen. 2010. "Lag (2010:566) Om Vidareutnyttjande Av Handlingar Från Den Offentliga Förvaltningen."
- Saarinen, T., and Sääksjärvi, M. 1990. "The Missing Concepts of User Participation - an Empirical Assessment of User Participation and Information System Success," *Scandinavian Journal of Information Systems* (2), pp. 25-42.
- Samuelsson, G., and Öberg, L.-M. 2008. "The Future History of E-Services. Long-Term Preservation of Complex and Integrated E-Services," in *Collaboration and the Knowledge Economy: Issues, Applications, Case Studies. Part 1* :. pp. 286-293.
- Samuelsson, G., Öberg, L.-M., and Borglund, E. 2007. "E-Services and Long-Term Preservation," in: *E-gov 07 conference, Regensburg (Germany), September 3-7*. Regensburg (Germany), September 3-7.
- Schellenberg, T. R. 1956/1998. *Modern Archives : Principles and Techniques*, (Repr. ed.). Chicago: SAA. (Original work published 1956).
- Sprehe, T. J. 2000. "Integrating Records Management into Information Resources Management in U.S. Government Agencies," *Government Information Quarterly* (17:1), pp. 13-26.
- Thomassen, T. 2001. "A First Introduction to Archival Science," *Archival Science* (1:4), pp. 373-385.
- Thurston, A. C. 2012. "Trustworthy Records and Open Data," *The Journal of Community Informatics; Special Issue: Community Informatics and Open Government Data* (8:2).