

Making the strange familiar: Bridging boundaries on database preservation projects

Peter Francis
Public Record Office Victoria
99 Shiel St
North Melbourne VIC 3051
+61 3 9348 5645
peter.franzis@prov.vic.gov.au

Alan Kong
Public Record Office Victoria
99 Shiel St
North Melbourne VIC 3051
+61 3 9348 5720
alan.kong@prov.vic.gov.au

ABSTRACT

Archive authorities develop information resources to enable public offices to meet their obligations under their jurisdiction's public records laws. Particular care is taken to ensure that these materials equip their audience with the necessary context and knowledge. Our current work with the evaluation of tools and processes for the preservation of relational databases causes us to question whether good documentation will be enough.

In this paper we describe our experiences at the Public Record Office Victoria (PROV), Australia, in developing processes and guidance for the preservation of relational databases. We find that these projects are different to 'traditional' transfers, and that their novelty and technical challenges may be made more difficult by organizational and conceptual complexities. We posit that the nature of such projects may require more than the knowledge of what must be done and how it should be done. We reason that these projects may be hindered by the lack of a shared language to communicate across organisational or functional boundaries.

Using database preservation projects as an example, we discuss the potential contribution that theoretical perspectives such as boundary objects (Star), transmission theory (Shannon) and externalization (Norman) may make to our development of guidance and how this may assist the support of cross-functional dialogue. While focused on database preservation projects, this approach may be generalisable to other cross-disciplinary and cross-functional work.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – user issues.

General Terms

Management, Documentation, Design, Human Factors, Theory.

Keywords

boundary objects, public records, database preservation, SIARD

iPres 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

Disclaimer: This paper is part of an exploratory research project and as such should not be regarded as endorsed policy by the Public Record Office Victoria.

1. INTRODUCTION

Public records form part of the Victorian jurisdiction's critical information infrastructure. They embody much of our community's civic and personal memory. Further, they play an essential role in the legislative and judicial systems, being relied upon as a true account in forensic legal investigations such as Royal Commissions and citizen's requests under Freedom of Information laws.

1.1 The responsibilities of archiving authorities

The *Public Records Act 1973* (Act) requires that the Keeper of Public Records establish recordkeeping standards for the efficient management of public records.

Underneath these standards is a comprehensive suite of recordkeeping documents including specifications, guidelines and fact sheets, each tailored for a specific audience including records managers, public officers, commercial entities and researchers.

The Act also specifies that the officer in charge of a public office¹ is responsible for carrying out a program of records management in accordance with the standards.

Our focus in this paper is on our role in the production of this guidance.

1.2 'Traditional' records management

The records management function in many public offices will be seen as largely concerned with management of physical records and dedicated electronic document and records management systems (eDRMS). Typically, the records management function is led by the records team within the agency.

The exponential increase of both physical and digital records, combined with the emergence of a number of disruptive technologies, has caused us to reassess the way we develop guidance.

Further, the manner in which information is stored, managed and used has changed dramatically over the years. This has reached a point where no one single unit within an agency could operate in isolation without the expertise and cooperation from other units.

¹ For the precise definition, see: http://www.austlii.edu.au/au/legis/vic/consol_act/pr1973153/s2.html#public_office

1.3 The SIARD Research project

Archive authorities² in Australasia have been developing their capacity to archive public records that are stored in non-records management systems, such as business systems. Earlier studies by PROV have resulted in a suite of projects to address this new landscape. One current project, SIARD Research, was commenced to develop our capacity to preserve relational databases from business systems³. There are not currently in place the tools or processes to ensure the continuum [14] management of public records in business systems. Trigger events may be when the business system is being decommissioned or otherwise deemed to be at risk.

The SIARD Research project is evaluating the database archiving tool, SIARD⁴, for its use in the transfer of public records from business systems to the state archive. In addition to the technical evaluation, we are exploring the end-to-end management processes, the design of our archive infrastructure, and the resource implications of a full-scale program.

This project has led us to consider the similarities and differences presented. For the purposes of this paper, we will discuss those of particular relevance to our topic – those relating to communication and shared understanding.

1.4 Boundary Objects

In their article, *Institutional Ecology, 'translations', and boundary objects: Amateurs, and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39* [12], Susan Leigh Star and her co-author James Griesemer examined the heterogeneity of scientific work within the Berkeley Museum.

Expanding the intersement model developed by Latour [5] and Callon [1], Star and Griesemer proposed the use of boundary objects as a mediator to engage the diverse actors to obtain and cooperation across multidisciplinary operations, ultimately to achieve a common goal. This model has been widely cited and the concept of boundary objects has been adopted in disciplines including computer science and public policy.

In their 1989 paper, Star and Griesemer [12] identified four types of boundary objects from their case study, although at that time and subsequently [11] made it clear that there were likely to be more. The initial four types were:

1. A repository that is standardized in a manner that allows access by different actors (i.e. a library catalogue)
2. A representation or abstraction that plays the role of an ideal type, serving as a platform to promote cooperation among different actors (i.e. circuit diagram)
3. An object that could be framed in a manner shared by different actors although the content within that object could vary

² For the purposes of this paper, archive authorities are bodies charged with responsibilities for the archiving of the public records for a jurisdiction.

³ We define business systems as information systems that are not specifically designed to support records management. Databases in business systems may contain public records.

⁴ Developed by the Swiss Federal Archives, SIARD stands for Software Independent Archiving of Relational Databases. See <http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en>

4. A form that is standardized in a manner that could be used by different actors

In another words, boundary objects could be viewed as a language which is translated and agreed upon, understood and used by two separate yet related actors across disciplines, facilitating them to achieve a common goal [3]. While clearly facilitating the co-ordination of work, however, boundary objects themselves should not be viewed as possessing co-ordinating features [9].

This paper describes some of the communication issues that may be presented by database preservation projects, and our application of a boundary objects perspective to them.

2. COMMUNICATION AND DATABASE PRESERVATION PROJECTS

Database preservation projects indicate a need for considerable use of cross-disciplinary and cross-organisational communication. This may be problematic as mis-communication between parties may introduced inefficiencies or rework into projects. In some cases, it may even contribute to viable projects being deemed unfeasible.

Cross-disciplinary and cross-functional communication problems are not unique to database preservation projects. Many ICT initiatives, for example, must deal with them. ICT projects, however, will generate considerable design documentation – 'as is' and 'to be' models that can be used in discussions with stakeholders. In contrast, our 'project manager' may be the records manager, who may not be widely recognized across the agency. Further, the preservation of databases for transfer to the state archive is unlikely to attract the resources or authority accorded a transformational ICT project, so the budget will not sustain elaborate documentation and the project will not enjoy high visibility. Our task then, is to support these projects within such constraints.

2.1 The draft process

We will first consider a simple process (Figure 1), where we embed the technical processes for database preservation into one that is similar to that used for the transfer of physical records or those from electronic records management systems. In short, PROV provides the standards and guidance for public offices to localise and execute.

The agency (public office) in the model contacts PROV (or accesses our online resources) for guidance on performing the preservation of a database. Armed with these materials, the agency works through the initial preparation (feasibility, planning), the technical preparation, determining the sentencing actions required (what to transfer to archive, what to leave in place, what to delete), the application of the sentencing, conversion to archival format, transfer to PROV, and ingest into our archive.

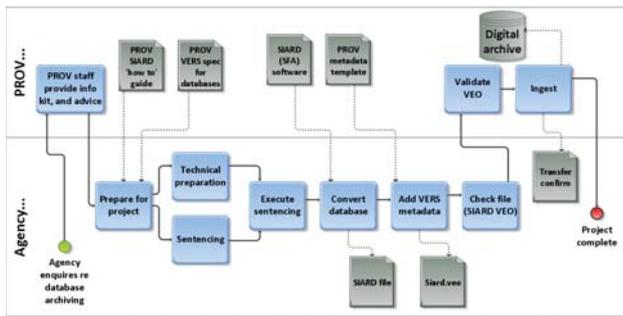


Figure 1: A simple view of the possible process for preserving databases using SIARD.

This is, however, a simplistic view of the process, and one that presumes a homogeneity that is rare in reality. The reality is often more complex, and the progress of such projects made problematic, due to technical, organizational and conceptual complexities not encountered in ‘traditional’ records management.

2.2 A more realistic view

In reality, accessing, preserving and managing a database within a public office to meet both legislative and organizational requirements will require considerable consultation and collaboration across functional, discipline and organizational boundaries.

2.2.1 Architecturally and technically complex

The tools and techniques for the long term preservation of relational databases continue to improve, however, it remains technically complex when applied to real business systems. The data models may be large and complex, and documentation sparse or non-existent. In the case of older systems, the staff who possess an intimate working knowledge of the system may have moved on.

As depicted in Figure 2, the business system may be accessing data from multiple databases, or the database may be supporting multiple business systems.

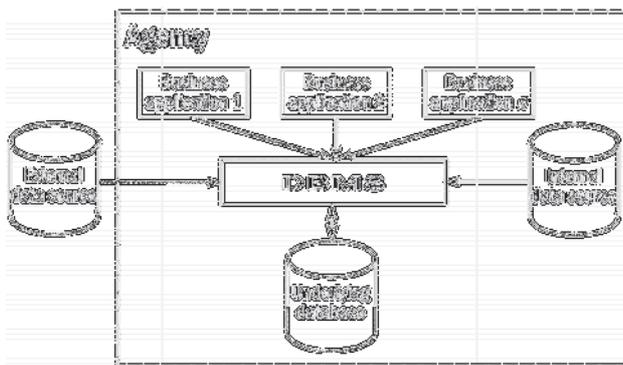


Figure 2: The data forming a ‘record’ may be aggregated from a number of sources.

2.2.2 Organisationally complex

Modern business systems are rarely managed by a wholly in-house IT function, for example:

- The business system and the database may be managed or hosted by a third party service provider.
- The business system’s vendor may maintain control over access to the application layer and the database.

- The business system and the database may be under the management of two different parties.
- The business system or the database may be hosted outside of the jurisdiction.

Gaining access to the database in order to perform the analysis, preparation, preservation and transfer tasks necessary may involve early and ongoing communication among various representatives of the public office (such as records, management and information systems staff), the application vendor, and the IT service provider.

2.2.3 Conceptually complex

The parties who will be involved in a database preservation project will likely bring their own conceptual models and perceptions of the project. As way of illustration, below is a non exhaustive list of the different actors that may have an operational, legislative or contractual interest and responsibility to the same database.

- The **records manager** has an invested interest with the data in the database and will see the database from a records perspective. To a records manager, the primary focus is to ensure that records in the database are preserved, managed, controlled appropriately.
- The **third party service provider** will be contracted to deliver IT services detailed in a suite of service level agreements. This may limit their ability to provide staff or resources to projects, particularly if they are not clearly defined or in terms that could be related to an SLA..
- The **vendor**. The responsibilities of the vendor are usually spelled out in the service agreement with the organization. To a vendor, intellectual property, privacy matters and financial considerations are a priority. They may also perceive general approaches regarding data transfer as an indication that the product is under review.
- The **database administrator**. Someone with database administration duties, and specific knowledge of the source data for the project will need to work on preservation planning and the execution of sentencing and export of the data. They will likely see the database in terms of its data model and stored procedures.
- The **application analyst**. Someone familiar with the business process supported by the system.
- The **data custodian** is someone with responsibility for the data involved to ensure governance obligations are met. In some instances, such as eDRMS may not have data custodians and, if not, this is another group that may have limited exposure to records management.

The participation of many of these people cannot be easily isolated to discrete tasks within the project. In many cases, they may need to work together productively to develop the project from the feasibility stage onwards.

Not only does each actor have their respective discipline based perception on a particular issue, he/she will also have their own psychological biases and work history which will vary even within the same discipline.

Many of these are people who have had little or no prior exposure to the records management environment, which indicates that records management concepts and terms may not be a natural option for a common language.

2.3 Example scenario: sentencing

The sentencing of the records may be done by a records manager, while the execution of the sentencing done by a database administrator or similar.

Records managers need to apply disposal plans to the records in the system (sentencing). To do so, they will need to see the data from a records perspective.

Once sentenced, they will likely need a DBA or similar role to execute that sentencing. The DBA will need the sentencing actions to be in a form that can unambiguously applied to the data model.

When executed, it is prudent that the action be validated - there is a risk here that miscommunications at any point may introduce errors - the wrong data may be preserved/destroyed. That is, neither the records manager, nor the DBA is able to judge that the other's work has been done correctly.

The database and/or the business system may be managed by a third party service provider. This has a number of implications: any work may come at a cost, and that cost may depend upon how 'actionable' the execution plan is (it is better to be in terms that they are familiar with and which do not need reinterpretation). The provider's representative will need to be involved at the feasibility stage - any miscommunication may result in advice that may prove prohibitively expensive making the project unviable, or may result in advice that the project is not technically feasible, or may be interpreted as impinging upon the vendor's IP (which we believe is in fact unlikely in most cases).

Addressing the technical complexities may be made more difficult due to the organizational and conceptual complexities that are likely on database preservation projects.

2.3.1 Addressing the performance gap

We can see that to implement database archiving projects will go beyond existing practices and perceived roles.

- If we expect that the work on SIARD projects will necessarily be across disciplines, as records managers will need to make sentencing decisions but data managers will need to execute them,
- and if records managers and data managers use different concepts and terms, and view the data in different ways,

then we should consider measures to reduce these barriers when developing our guidance materials.

3. BOUNDARY OBJECTS AND DATABASE PRESERVATION PROJECTS

3.1 Relevant qualities of a boundary object

3.1.1 Translation

Further, we believe that the language used and the form of the 'object' must not disenfranchise or subordinate collaborators – there will likely be a leader, but the object should not determine who that will be.

When we think about translations, we do it with Shannon's [10] model in mind (Figure 3), which, although developed for telecommunications, has been found more widely applicable to human communication.

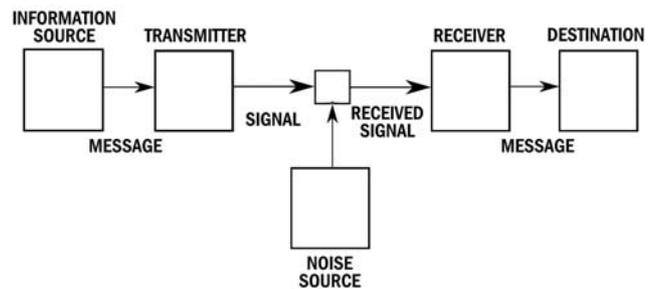


Figure 3: Shannon's schematic diagram of a general communication system [10].

In the non-technologically mediated case of two people speaking to each other, the Transmitter could be regarded as the language and concepts used by the speaker (what they say and how they say it). The Receiver may be the interpretive filter (of their role and experience) that may influence what the listener hears. Although originally a technical model, we find the concept of messages undergoing encoding and decoding helpful. The role of a boundary object may minimize the need for both parties to 'translate' for the other.

In the earlier sentencing example, communication is depending upon the forming of the request by the records manager and the interpretation of the request into database operations by the DBA. Where the need for interpretation, or re-analysis, is high, so too is the risk of error or unnecessary rework.

3.1.2 Externalisation to aid cognition

Although not a strict quality of boundary objects, we anticipate most will have a material quality that will support individual and shared thinking. Externalisations have long been considered to enable memory and computational offloading, freeing the mind of some of the burden during problem solving (see, for example [4, 7]).

We see that a boundary object in database preservation projects that enables a database administrator and records manager to relate the 'record' and the data model to the business system would reduce cognitive load on both parties.

3.1.3 Non-directive and unbiased

A boundary object is non-directive, it does not embody any responsibilities or agreements, and implies no obligation on the parties. Where such mechanisms are necessary, they can be managed outside of, not through, the object.

The planning model, as demonstrated by Suchman [13], is flawed. We should take care not to build our logic into the object and introduce further barriers to use.

3.2 A boundary object for database preservation

We look for possible common concepts, ones that directly relate to the system, but in which each party can derive meaning for their own work. For example, one candidate that is neither a record nor a data model is the business object.

3.2.1 A business objects perspective of the data

If public records are to be identified and appraised in business systems, it will be necessary to look at the business system's data (a relational database model) from a records perspective. Once records management decisions have been made, they must be

translated into requirements that a database administrator can execute.

From a database perspective, Olson [8] describes business objects as either ‘entities’ or ‘transactions’. Entities persist for long periods of time, and are subject to change over time. Transactions are records of events that are created and completed in a relatively short period of time.

3.3 The sentencing scenario revisited

If we consider the case of a fictitious government agency, the Dept of Science. The records manager has identified the Service Delivery System (SDS) as likely holding public records. The SDS supports the department’s role in providing advice to research organizations. The Advisory Services function is covered by a Retention & Disposal Authority (RDA), developed by the department to identify their public records and detail their management.

The RDA has been used to manage Advisory Services records stored in the department’s electronic records management system, however, the records manager believes that the SDS system contains data that would also be required to be preserved permanently and transferred to PROV.

Figure 4 depicts a simple business object model of the fictional SDS. This view may map well onto the records management concept of a record.

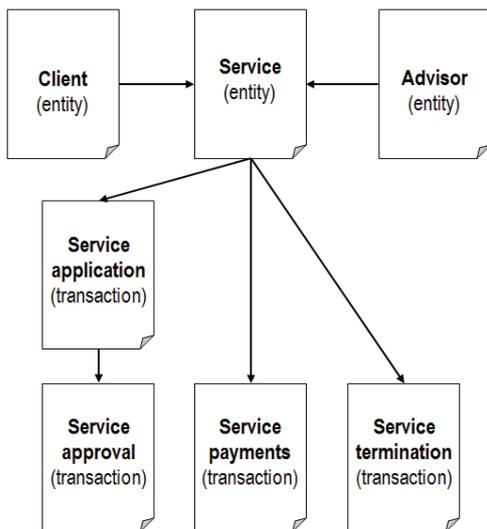


Figure 4: An example of a simple business objects perspective as a boundary object (using a fictitious Dept of Science service delivery system).

It may be that by jointly analyzing the business system and expressing it terms of business objects the records manager and database administrator will establish a shared understanding of the system.

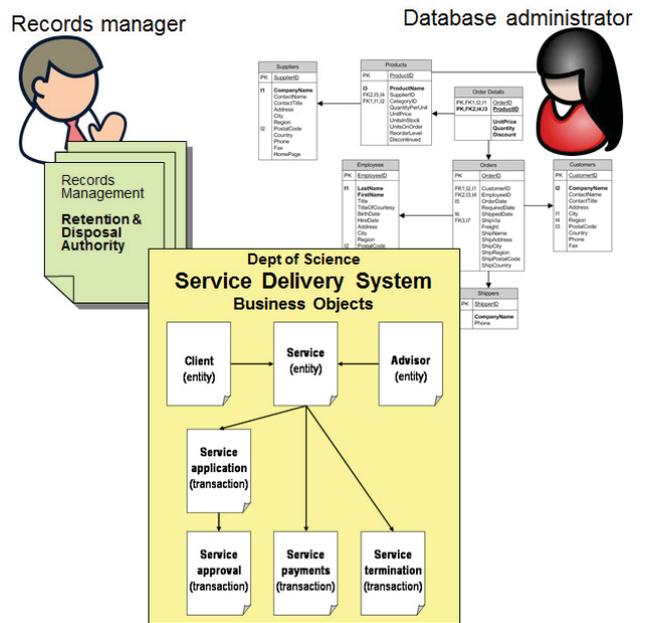


Figure 5: Example of the use of a business objects concept to facilitate communication between a records manager and database administrator.

3.3.1 As a translation support

In the example of use depicted in Figure 5, the business object model may serve as a useful bridge for the records manager to describe the data requiring action, and the criteria for determining action (such as retain, transfer to PROV, destroy, etc.). The RM may find it easier to express the functional descriptions of the RDA into relevant business objects, than on a database schema. For their part, the DBA may be more confident in tracing the database tables and fields supporting a business object, than from the descriptions commonly found in an RDA.

3.3.2 As a form of externalization

By providing a physical model that is able to be expressed as a diagram (as above) a table or list, both the RM and the DBA can reduce the need to retain both the conceptual model and the past determinations as they deal with a problem at hand.

3.3.3 Non-directive and unbiased

The business objects model may be useful to both the RM and the DBA but does not clearly belong to either world. In this way, it does not confer ownership to either.

This exchange highlights another potential benefit in that it may simplify the identification of the data required, in instances where the data is distributed by providing a logical rather than physical perspective.

There are a number of potential barriers that may hinder the adoption of boundary objects. One particular assumption is that each actor, given he/she is fully aware of the type of boundary object that is at play, is willing to adopt the object to achieve an outcome. However, this level of willingness is dependent on a number of factors including the actor’s trust of the approach, past history, relationship with the other actor and other behavioral biases.

In addition, the boundary object itself is silent on whether the achieved outcome reflects work policy or the organisation’s

overall strategic direction. Without addressing these fundamental concerns, it is likely that despite the boundary object being effectively used, there will be no support from the executive or stakeholders.

Boundary objects are unique in that they are designed to address one particular given circumstance which may become ineffective when applied elsewhere.

3.4 Evaluation

We will be using data generated during the SIARD Research project to map records management definitions and concepts onto data models and vice versa. In the process, we will look for opportunities for the development of general principles that can be used as the basis for the development of a transformation tool.

Our initial evaluation of this approach and of any potential boundary objects will be through iterative co-design and collaboration with our project partners. We believe that this field development will give our work a form of member validation [6] and we leave the judgment as to our success to those who it is intended to support.

4. CONCLUSION

The motivation for the work described in this paper is founded on a number of questions: We ask, as we always do, are our guidance materials fit for purpose? Are they accurate? Do they reflect policy? Are they within our scope, not straying into areas beyond our brief? Are they generalisable, do they work for all our public offices?

Our work to date on the SIARD Research project causes us now to ask, will our usual approaches be successful? Is there more than knowing what to do, and how to do it? We must anticipate that database preservation projects will rarely enjoy the resources, design documentation, or profile that would accompany an ICT project. Our proposed approach, outlined in this paper, is shaped by two constraints: the almost infinite variety of installations in public offices, and the clearly finite resources that archiving authorities are able to allocate to any problems.

Business systems and the underlying databases are implemented in a variety of ways and under a variety of management arrangements. Even at the data level, there will be the possible need for operator intervention, and the use of a variety of export and conversion tools. "The processing of the finding aids has taught us many useful lessons relevant to preservation of databases and other structured data. It revealed that there is no such thing as a standard way to import data. Most of the 3.1 million records needed some kind of human intervention during the import process. The data of the DTNA project was imported using a variety of different methods such as direct database connections and exporting data as CSV from the source." [2] p.9

Archiving authorities cannot always 'be there' for the agency, to assist or facilitate – it is not sustainable for them to do so. They can, however, continue to reflect upon the guidance materials they provide.

We have identified that some new approaches to the preservation of public records may be impeded by organizational and conceptual complexities not generally encountered during more traditional public records transfers. The archiving authority may

not necessarily be able to address them simply by providing better advice on what should be done, however, including a boundary objects perspective into our thinking as we develop resources to support public offices may assist in better communication and collaboration on cross-disciplinary public records preservation projects.

5. ACKNOWLEDGMENTS

Our thanks to our colleagues at Public Records Office Victoria who worked on this project with us or acted as reviewers for this paper.

6. REFERENCES

- [1] Callon, M. 1986. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *Power, Action and Belief: A New Sociology of Knowledge?* J. Law, ed. Routledge, London. 196–223.
- [2] Essen, M. van, Rooij, M. de, Roberts, B. and Dobbelsteen, M. van den 2011. *Database preservation case study: Review*. National Archives of the Netherlands.
- [3] Fox, N.J. 2011. Boundary objects, social meanings and the success of new technologies. *Sociology*. 45, 1 (2011), 70–85.
- [4] Hutchins, E. 1995. *Cognition in the wild*. MIT Press, Cambridge, Mass.
- [5] Latour, B. 1987. *Science in action: How to follow scientists and engineers through society*. Harvard University Press, Cambridge, Mass.
- [6] Neuman, W.L. 2003. *Social research methods: Qualitative and quantitative approaches*. Allyn and Bacon, Boston.
- [7] Norman, D.A. 1993. *Things that make us smart: Defending human attributes in the age of the machine*. Addison-Wesley Publishing Co., Reading, Mass.
- [8] Olson, J.E. 2010. *Database archiving: how to keep lots of data for a very long time*. Morgan Kaufmann.
- [9] Schmidt, K. and Bannon, L. 1992. Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)*. 1, 1-2 (1992), 7–40.
- [10] Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*. 27, (1948), 379–423, 623–656.
- [11] Star, S.L. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science, Technology & Human Values*. 35, 5 (2010), 601–617.
- [12] Star, S.L. and Griesemer, J.R. 1989. Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*. 19, 3 (1989), 387–420.
- [13] Suchman, L.A. 1987. *Plans and Situated Actions: The Problem of Human-machine Communication*. Cambridge University Press.
- [14] Upward, F. 1996. Structuring the records continuum - part one: Postcustodial principles and properties. *Archives and Manuscripts*. 24, 2 (1996).