

Multidisciplinary Research Collaboration: A Case Study of Data Management for Reuse of Qualitative Data Over Time

Molin-Juustila, Tonja

EveLine Group and Department of Information Processing Science
University of Oulu, Finland
tonja.molin-juustila@oulu.fi

Abstract. This position paper presents some initial grassroots experiences and challenges while designing a shared data repository and related practices in order to support multidisciplinary collaboration. From the point of view of infrastructuring, this case is an effort to build small scale data management technology and related work practices to support long-term research collaboration across disciplinary and organizational boundaries. The main problem area, the privacy challenge of qualitative data, is explained in more detail.

1 Introduction

In 2006 a group of researchers (called MOTTI at that time) started a project to increase cross-disciplinary collaboration in research training and supervision (see <http://www.oulu.fi/motti/in-english.html>). The main idea was to develop students of higher education in their skills concerning interdisciplinary work and research. At that time, however, the concrete efforts of the project were put on interdisciplinary and collaborative practices of teaching and training of basic

research skills. The group had researchers from many different disciplines (English and Finnish language, sociology, marketing, information processing science, etc.) within the University of Oulu, Finland. After two years of teaching collaboration, this group of researchers – currently called the EveLine group – realized that in order to really improve the multidisciplinary research practices the group should change the perspective from merely supervising individual, and typically disconnected students to a more research focused approach; meaning research collaboration around issues with shared, multidisciplinary interests.

The core members of the EveLine group are faculty members (professors, lecturers, post docs, etc.) from the different disciplines sharing a common interest: currently the everyday life of technology-rich neo-communities. The main goal¹ of this group is to conduct long-term research around the main subject area with more focused, project based research efforts around some special issues and topics. In order to do this, the group applies for funding from different sources for different, focused purposes and therefore the research work of this multidisciplinary collaboration is typically managed via projects. Also the members actively involved may dynamically vary according to the projects.

Very soon we realized that a shared repository of data was needed in order to support and manage such a long-term multidisciplinary research efforts across disciplinary and organizational boundaries. In this position paper, the initial experiences and major challenges of this multidisciplinary collaboration case will be discussed in order to provide our empirical experience of collaboratively designing a data repository and related management practices. This case could be considered as an end user initiated case of e-Research (see Jirotko, Procter, Rodden & Bowker, 2006).

2 Setting the Stage for Shared Data Repository

From September 2008 until January 2009 the EveLine group conducted its first collaborative research project. This multidisciplinary research project called “MOTTI” was considered as an initial starting project with a common multidisciplinary research interest (on children and technology) and an aim to learn more about the necessary practices of this type of multidisciplinary research collaboration. The research was implemented through a student project work course for students of information processing science. The course is part of the necessary advanced studies for 4th year students during which the students work 4-

¹ With the concept neo-community we refer to the modern social life as a complex mixture of real and virtual due to the wide spread of modern information and communication technology.

6 months (300 hours each) on any subject typical for their future profession. Their assignment was to study everyday technologies of school children, aged 9-11 and to practice participatory design with these children concerning a portfolio of their school work. The project was managed by a multidisciplinary steering group with members from the EveLine group and a teacher representative from the school.

Since the very beginning, this first project had an aim to collect important research data for multidisciplinary analysis. This data was supposed to form the basis for long-term research collaboration on the subject area and therefore there was a need for a data repository shared by the EveLine group. However, there was no such a repository available yet. Part of the project work was also to establish such a repository – together with rules and practices to elaborate with the repository. The idea was to collect data for future research efforts to be used by both the members of the EveLine group as well as by the current or future students (of any discipline) interested in doing their thesis work based on the collected data. As part of their project work assignment, the students' were supposed to examine different issues (like study agreements, metadata, legislation, etc.) related to this type of long-term data management work and their current practices.

Although this MOTTI project had a data repository and related practices as one of its main goals, the students did not understand this goal. All their efforts were put on the research on children and technology and therefore, until the mid steering group meeting of the project, they considered this long-term data management issue more like a practice related only to their own project, i.e. informed consent forms and metadata descriptions for their own data. This is quite interesting and partly shows how new and timely issue this long-term data management is for our students as well as for us researchers. Students did not understand the issue and we as the researchers could not provide the necessary basis and guidelines to trigger such efforts and state of mind. However, as guided in the steering group's mid meeting, the MOTTI project group started to pay attention also to the time after the project and what will happen then with their data. At this point the challenge with Finnish legislation was realized. As guided by the steering group, the group started to look for regulations related to the area and very soon found problems with the data already collected. Because of these problems, the details of which will be discussed in the next section, the project finally ended with a "disagreement" concerning the rights to use their data in the long-term research efforts of the EveLine group.

3 The Privacy Challenge of Quality Data

In their report on this long-term data management issue (Asamäki et al., 2009), the MOTTI project group describes the challenge quite well. The main issue of

concern turn out to be the personal data collected and the privacy of the research subjects. It is the matter of balance between privacy and the long-term archiving of authentic qualitative data collected with necessary details for research purposes (see also Carusi & Jirotko, 2007). In this case the conflict focused around diary types of videos recorded by the children aged 10-11.

According to the MOTTI project's informed consent form, participation in the research was made voluntary and by assigning the form the parents of the school children gave permission for their child to participate the study. In the informed consent form it was explained that digital material in the form of text, images, audios and videos might be collected and possibly archived as such in a repository supporting long-term research and teaching purposes. It was also made very clear that this material might be used by the researchers (and their partners in cooperation) as well as by the students of the University of Oulu only in purposes of research (including thesis work) and teaching. In the publications based on the material issues concerning privacy will be taken care of and individual participants will not be identifiable.

It might be harmful for people if this type of research data is further delivered into, or otherwise become available to, the "wrong" hands meaning that it will also be used disrespectfully. People have rights to prevent this of course and the Finnish laws (like the Constitution of Finland, the Personal Data Act, the Act on the Protection of Privacy in Electronic Communications and the Copyright Act) has protected these rights. Especially *Personal Data Act* concentrated on the privacy issue. One of the core issues around the video recordings was the potential person data file (register) created. According to the *Personal Data Act* (Finlex, 1999) "*personal data* means any information on a private individual and any information on his/her personal characteristics or personal circumstances, where these are identifiable as concerning him/her or the members of his/her family or household." This means that the video recordings produced by the children, with an identifiable person face and voice, it was considered as personal data. When this digital material is stored in a shared repository, the data forms a personal data file. "*Personal data file* means a set of personal data, connected by a common use and processed fully or partially automatically or sorted into a card index, directory or other manually accessible form so that the data pertaining to a given person can be retrieved easily and at reasonable cost" (Finlex, 1999). Based on their search for these regulations, the MOTTI project group considered the archiving of the diary videos as unethical. The group felt that the diaries included "sensitive" data (young children at their homes wearing nightdresses etc.) and yet we, the EveLine group, had only quite a vague intention for long-term archiving of the videos without any practices and guidelines existing concerning the rights to access and permission to use the material, not to mention the hosting of the repository.

Although these practices and guidelines were considered as one of the project task, The MOTTI group did not have enough resources left to create these practices or any other proper ways to address these conflicting issues. It was also discussed that anonymisation of those videos was not an option because necessary information would have been lost and therefore the value of the material would have lessened substantially.

Interestingly, the Finnish *Copyright Act* protects rights concerning images, audios and videos. Therefore the MOTTI project group considered that the EveLine group does not own the rights for the videos and therefore it is the MOTTI group's right, or better yet an ethical responsibility, to deny us from archiving the videos. We discussed the consent form providing the necessary authority for archiving, but the MOTTI project group felt that the form did not include necessary statements concerning the fact that archiving the material in a digital repository will form a register with personal data. Because of this, the consent form should have included also the necessary controller's information concerning the personal data register (e.g. information concerning the controller, the purpose for collecting personal data, description of the data collected, security information etc.). They came to a conclusion that the original videos should be destroyed unless, within a year, the EveLine group is able to solve these controversial issues either by reconciling opacity related to these various legislations² or creating reasonable practices concerning the long-term data management. Only transcriptions made by the MOTTI project group or the EveLine group (during this year) may be archived.

Therefore, in order to solve this disagreement, in January 2009 (ending at the end of June) the EveLine group started a new project, called "Datalog", again a multidisciplinary project to design the shared data repository for the EveLine group. It was seen, that the data repository together with reasonable practices concerning its use might solve this conflict. Also the MOTTI project group agreed that this would solve our disagreement.

4 Building the Shared Data Repository

Again, the implementation of the data repository was started through a study project, this time with a group of students from two departments. In the project assignment it was further clarified that the EveLine group also needs reasonable tools (e.g. improving the consent form) and practices for long-term data acquisition and archiving as well as for metadata and data management including

² Interestingly it turned out that currently even the central research and innovation services of the University of Oulu has no instructions concerning these types of legal issues.

repository controlling. Special attention was to be given to the privacy issues and practices concerning the archiving and use of sensitive personal data.

While defining the requirements for the repository, the Datalog project decided to structure the repository data into groups of certain types of data the use of which was allowed for certain types of user groups. This way it was made possible for the system to restrict unauthorized use of the data. In addition the necessary roles of administrative as well as main users were identified. The data itself was considered to be of various elements (e.g. texts, images, audios and videos) and with possible links between the different data elements. Since the very beginning of the project also diversified metadata was considered important for easy and efficient use of the data.

During the Datalog project it turned out that creating the use practices for the repository were more demanding than expected. This was also closely related to the members of the EveLine group and their varying roles and responsibilities related to the data being archived. Who (the administrator or the main user, or both) should have the rights to create new users or user groups for the repository? Who should have the rights to give permissions for the use of certain data elements? Considering the very sensitive data, is it acceptable to have administrator or main user (possibly several of them) to have access to the data? On the other hand, how would it be possible to manage the repository if individual researcher (i.e. conventional users) themselves would be given the right to authorize the use of the data elements added to the repository. Also it was not clear if all the conventional users would be given the right to search the metadata of data elements with restricted access rights.

Currently the Datalog project has ended, the resources of which unfortunately were not enough to meet the goals. As for the foundation of the repository it was decided that suitable open source solutions should be searched. Based on this, the DSpace open source software (see <http://www.dspace.org/>) was considered the most promising and therefore selected. It was considered as reasonable and sufficient application to be tailored for the project's purpose, but unfortunately it was realized only later that also changes at the code level would be needed causing demands for more technical expertise currently not included in the multidisciplinary EveLine group. Especially one key area in need for better solution is the possibilities to manage metadata. At the very beginning of the project the importance of the metadata was highlighted. It was even considered that a simple repository of only the metadata might be sufficient enough for the EveLine group providing only references to people responsible for the different data elements. One important characteristic concerning the metadata was

automatically forcing to fill in certain fields of metadata for each data element. Unfortunately the current DSpace solution does not support this.

This position paper has provided a description of a challenging case of our micro (or grassroots) level attempts for data management of long-term research data. The multidisciplinary research group, EveLine, has an acute data management problem while establishing the basis for its emerging multidisciplinary research collaboration. A shared data repository of mainly qualitative type of data is considered as a necessity for the project-based evolution of the multidisciplinary shared research area. This paper has offered our initial experiences and challenges when collaboratively designing such a repository and related micro level management practices. We hope you will find it as an interesting position for the “collaborative infrastructuring” workshop.

5 Acknowledgements

The author gives warm thanks to all the members involved in the MOTTI and Datalog projects.

6 References

- Asamäki, M., Halonen, S., Juusola, E., Kangas, A., Kinnunen, R., Muhonen, S. & Vallinmäki, T. (2009): Johdatus empiirisen aineiston pitkäaikaiseen hallintaan tulevaisuuden opetus- ja tutkimuskäyttöä varten, Dep. of Information Processing Science, University of Oulu, Finland. (in Finnish).
- Carusi, A. & Jirotko, M. (2007): From data archive to ethical labyrinth. In Proceedings of the Third International Conference on e-Social Science. (<http://ess.si.umich.edu/papers/paper171.pdf>)
- Finlex. (1999). Personal Data Act (unofficial translation). Finland's Legislation. 1999/523. (http://www.finlex.fi/en/laki/kaannokset/1999/?_offset=20&_max=44)
- Jirotko, M., Procter, R., Rodden, T. & Bowker, G.C. (2006): Special Issue: Collaboration in e-Research. *Computer Supported Cooperative Work*, 15, 251-255).