Publisher: IISI - International Institute for Socio-Informatics
ISSN 1861-4280

international reports on socio-informatics

volume 6 issue 4
2009

Proceedings of the ECSCW 2009 Workshop on Collaborative Infrastructuring – Conceptualizing Emergence and Design of Information Infrastructures

Guest Editors:

Karen Baker
Pelle Ehn
Sara Eriksén
Helena Karasti
Michael Twidale
Volkmar Pipek

Editors:

Volkmar Pipek
Markus Rohde

Table of Content

The 'international reports on socio-informatics' are an online report series of the International Institute for Socio-Informatics, Bonn, Germany. They aim to contribute to current research discourses in the fields of 'Human-Computer-Interaction' and 'Computers and Society'. The 'international reports on socio-informatics' appear at least two times per year and are exclusively published on the website of the IISI.

Impressum

IISI - International Institute for Socio-Informatics	fon:	+49 228 6910-43
Stiftsgasse 25	fax:	+49 228 6910-53
53111 Bonn	mail:	iisi@iisi.de
Germany	web:	http://www.iisi.de

Workshop Summary: Collaborative Infrastructuring – Conceptualizing Emergence and Design of Information Infrastructures

Karen Baker[1], Pelle Ehn[2], Sara Eriksén[3], Helena Karasti[4], Volkmar Pipek[5], Michael Twidale[6]

[1] Scripps Institution of Oceanography, University of California San Diego, USA; [2] School of Arts and Communication, Malmö University, Sweden; [3] School of Computing, Blekinge Institute of Technology, Sweden; [4] Department of Information Processing Science, University of Oulu, Finland; [5] Institute for Information Systems, University of Siegen, Germany; [6] Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, USA

[1] kbaker@ucsd.edu, [2] pelle.ehn@mah.se, [3] sara.eriksen@bth.se, [4] helena.karasti@oulu.fi, [5] volkmar.pipek@uni-siegen.de, [6] twidale@illinois.edu

Abstract. The workshop examined issues around the collaborative design and use of information infrastructures through a collective sharing and analysis of case studies. We welcomed as position papers analyses on empirical studies or descriptions of cases that the authors are familiar with. The workshop approach was a collaborative activity involving a 'live metareview' over participants' case studies. That was, the group will consider in turn a number of issues emerging from the cases. For each issue we discussed whether and how it manifests in the particular infrastructure settings that each participant is familiar with or has studied. This enabled the participants to gain a richer understanding of the research space around infrastructure design and use. Goals of the one-day workshop were: case studies explored, key issues and special problematics identified, a poster prepared for the conference poster session, a journal special issue planned – and networking.

Introduction

There is a growing interest in information infrastructures and their multiple facets. With a focus on large-scale technological systems, they have been studied as increasingly ubiquitous (e.g. e-society, e-government, e-science/cyberinfrastructures, e-research, e-health) conglomerates of technologies and practices (Ciborra et al. 2000, Atkins et al. 2003, Jirotka et al. 2006, Edwards et al. 2007, Olson et al. 2008). With a focus on their taken-for-grantedness independent of their scale, they have been studied as ‘work infrastructures’ (Hanseth & Lundberg 2001, Pipek & Syrjänen 2006, Pipek & Wulf 2009) even on smaller scales. Although there is an impressive history of research on infrastructures (Star 1999, Star & Ruhleder 1996, Edwards et al. 2007, Ribes & Finholt 2007), and although much of that research remains highly relevant even as technologies have changed and become more pervasive, nevertheless there remains much to be learned.

We particularly struggle to design good, robust, flexible and adaptable infrastructures that can scale well and remain useful over time. Our success is mixed, and to a certain extent many approaches assume a passive perspective on infrastructure processes as developing phenomena. For a design turn, we need to assume an active perspective on infrastructure processes as phenomenon development. Can we do better than a Darwinian model of constantly building new ones and hoping that some will prove fit enough to survive and be replicated? The need emerges to complement existing research with a closer look at the stakeholders and collaborations that produce and manifest infrastructure (Lee & Dourish 2006), and to consider approaches that see infrastructure-making as a process.

Information infrastructures are multi-scale in terms of spatial extent, temporal orientation, and conceptual breadth. In recent times there has been a growing emphasis on how to create infrastructures that are large scale and can operate for the long term (Karasti & Baker 2004). This provides impetus to comparative studies across the continuum of network activities from micro to macro. As large-scale initiatives, they involve top-down elements such as coordination and provision of access to scarce resources. However, at the same time they may need to be integrated with local practices and existing systems sometimes referred to as the ‘installed base’. As ‘taken-for-granted’ systems they develop a certain dynamic of innovation around breakdowns and the resolution of reverse salients. Infrastructures come about through wide-ranging chains of innovation that eventually bring their usages into effect. These innovations can involve the end users themselves, either by intention of the infrastructure designers (Dittrich et al. 2009) or unbeknownst to them (Karasti & Baker 2008, Twidale & Floyd 2008).

People in multiple roles – not only professional designers – establish and shape an infrastructure through various kinds of encounters over extended timeframes. People can innvovate by programming, but also by tailoring, or appropriating an application for a purpose its designers did not intend. They can combine existing familiar computer applications into a complex workflow, and exploit novel applications and web services as they become available. The activities involved are varied, relating to a) selection, adoption, combination, design, development, deployment, enactment, and implementation of systems and environments b) mediation, interpretation, elicitation, and articulation of issues by professionals in emergent roles, and c) adaptation, appropriation, tailoring, redesign, and maintenance by diverse individuals over time. Strategies for before-use, during-use, and future-use are intertwined. The suite of actions are intricately interconnected, interdependence made explicit by the infrastructures themselves. Long-term infrastructures can be thought of as a network of processes, i.e. multiple simultaneous, interleaved processes that require constant tending to shift in response to unexpected contingencies, to develop in response to local insights, and to counteract tendencies to drift from alignment. They are also often networks of application and services that can change as new technologies emerge and use patterns and needs evolve. It is this multifaceted, complex phenomenon that was explored in our workshop.

The workshop aimed to help develop a richer understanding of issues related to the analysis and design of infrastructures:

1) the concepts, issues and theories that can inform our analysis both of the infrastructures themselves, and of the processes of collaborative infrastructure design
2) the concepts, issues, theories and methods that can improve the processes of doing collaborative infrastructure design

In order to achieve these aims, the workshop involved a collective sharing and analysis of case studies.

Position papers were invited that include one or more case studies, empirical research or at least some description of an infrastructure setting that the workshop participant was familiar with and could discuss at the workshop.

References

Atkins, D. E., K. K. Droegemeier, et al. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure, Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.

Ciborra, C. U., Braa, K., Cordella, A., Dahlbom, B., Failla, A., Hanseth, O.,

Hespø, V., Ljungberg, J., Monteiro, E. and Simon, K. A. (2000) From control to drift - the dynamics of corporate information infrastructures. Oxford University Press, Oxford.
Dittrich, Y., S. Eriksén, et al. (2009). From knowledge transfer to situated innovation: Cultivating spaces for co-operation in innovation and design between academics, user-groups and ICT provides. Research report. Blekinge, Sweden, Blekinge Institute of Technology.
Edwards, P.N., Jackson, S.J., Bowker, G.C., & Knobel, C. (2007). Understanding infrastructure: dynamics, tensions, and design, NSF Report of a Workshop: History and theory of infrastructure: lessons for new scientific cyberinfrastructures.
Hanseth, O. and N. Lundberg (2001). Designing Work Oriented Infrastructures. Computer Supported Cooperative Work 10(3-4): 347-372.
Jirotka, M., R. Procter, et al. (2006). Special Issue: Collaboration in e-Research. Computer Supported Cooperative Work 15(4): 251-255.
Karasti, H. and K. Baker (2004). Infrastructuring for the Long-Term: Ecological Information Management. Hawaii International Conference on System Sciences 2004 (HICSS'37), January 5-8 2004, Hawaii, USA.
Karasti, H. and K. S. Baker (2008). Community Design: Growing One's Own Information Infrastructure. Participatory Design Conference (PDC'08), Oct 1- 4 2008; Bloomington IN, USA, CPSR, ACM.
Lee, C. P., P. Dourish, et al. (2006). The Human Infrastructure of Cyberinfrastructure. CSCW'06, Banff, Alberta, Canada; November 4-8, ACM.
Olson, G., Zimmerman, A. and Bos, N. (2008) Scientific Collaboration on the Internet. MIT Press.
Pipek, V. and Syrjänen, A.-L. (2006) Infrastructuring as capturing in-situ design. In 7th Mediterranean Conference on Information Systems, Association of Information Systems, Venice, Italy.
Pipek, V. and Wulf, V. (2009) Infrastructuring: Towards an integrated perspective on the design and use of information technology. Journal of the Association for Information Systems (JAIS), Special Issue on e-Infrastructures 10 (4), to be published.
Ribes, D. and T. Finholt (2007). Tensions across the Scales: Planning Infrastructure for the Long-Term. GROUP'07, Sanibel Island, Florida, USA; Nov 4-7 2007, ACM.
Star, S. L. (1999). The Ethnography of Infrastructure. American Behavioral Scientist 43(3): 377-391.
Star, S.L. & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research: 7(1), 111-134.
Twidale, M. B. and I. Floyd (2008). Infrastructures from the bottom-up and the top-down: Can they meet in the middle? Participatory Design Conference (PDC'08), Oct 1-4 2008; Bloomington IN, USA, CPSR, ACM.

The AUGMENT Project: Co-constructive accessibility mapping for supporting people in their realities

Annelie Ekelin, Peter Anderberg
Blekinge Institute of Technology, Blekinge County Board of Health
annelie.ekelin@bth.se, peter.anderberg@ltblekinge.se

Abstract. This position paper describes an ongoing pilot study which is part of the start-up of a three-year local research and development project called AUGMENT. The aim of the project is to explore and develop participatory methods and cultivate shared transformational spaces concerning user-driven up-date of accessibility information via mobile applications. There are a multitude of heterogeneous user-groups, as well as a number of other local stakeholders, whom we would like to involve in this case. Some of them are more difficult to reach and communicate with than others, whether due to bureaucracy, lack of time/disinterest, physical and/or psychological disabilities and/or lack of representation giving them an official voice and communication channel in the local society. In this position paper we outline and discuss some dilemmas and issues which have been high-lighted through the pilot study so far and which are of both technical and social character.

Introduction

Recent research on rehabilitation engineering (Anderberg, 2006) has shown that disability is dependent on the situation, not primarily on the individuals, which means that problems are possible to jointly minimize or solve to a greater extent than the mainstream understanding of disability problems has previously assumed. Accessibility, thus, should be seen as an act of co-construction, not something which someone has to provide for someone else. Accessibility information needs to be re-formulated and customized, depending on the individual's circumstances and current location in space and time, rather than simply and statically presented as one-size-fits-all and relying on individuals learning generalized strategies of how to use "off-the shelf information". New solutions providing access for disabled groups are frequently developed - this is not an issue any longer- the issue is rather: how are these new solutions communicated to those who need the information and in which way are they contributing to the understanding of accessibility? How can shared spaces be opened and cultivated for co-construction of accessibility? The AUGMENT project aims to work with groups and individuals who have experience of accessibility problem and who are not satisfied with current accessibility solutions, which have primarily concentrated on regarding accessibility as a reificated artifact (Ekelin, 2007) rather than a situation which is dependent on re-interpretation.

The organization of physical places affects disabled people's possibilities of participation. The physical environment in Sweden to some extent lacks relevant customization and there are also gaps in accessibility for groups of people with various disability problems. But the picture is not one-sided. In relation to rebuilding of physical environment, a number of accessibility problems are solved over time. The issue is rather how these changes and reinforcements are communicated to the affected groups and individuals who are dependent of such information. A repeatedly formulated wish from representatives of these groups is the possibility to describe environments with the help of images and other examples of ”rich pictures”, where the user her/himself can decide about and evaluate the offered accessibility.

In a recent charting of different EU-initiatives, HANDISAM, the Swedish Agency for Disability Policy Coordination, point out that the aim of steering development and research towards more inclusive projects and solutions is based on the i2010 strategy which is the guiding framework for accessibility issues. There are ongoing discussions about legislation of eAccessibility within EU and the European Commission highlights the importance of prioritizing a coherent, mutual and effective strategy for eAccessibility, or web accessibility, in order to boost the development of the eSociety in line with a new social agenda. (KOM2008: 412, cited in Axelson, 2009, p.12)

This is not, however, the main priority today. There is also great demand for flexibility and mobility, and a new generation of mobile web tools has been developed, contributing to supporting and enhancing this mobility and flexibility. Interactive features make it possible for individuals to contribute on various levels by posting experience based information on the web site. The figurative expression of "*lowering the tresholds*" has thus taken on a new meaning, beyond, yet based on, reflecting back to and further enriching, as it were, the original metaphor. Providing accessibility is not simply about providing information, but also about providing means of co-construction of the expressions of accessibility as well as form and content – providing space for exploring a multitude of experiences of variations of disability in relation to accessibility issues.

Some interesting research issues in the AUGMENT project, as we see them evolving now, through our experiences within the on-going pilot project, are:

- In what ways will development towards more specific demands of flexibility and mobility and in design of place-based accessibility information reconfigure the notion of accessibility along with the experience of accessibility hindrances?
- What are the consequences for the design and management of a specific accessibility application?
- How might a WIKI[1] solution based on cooperative reworking of material support participation by users? Could it support increased participation? Might it become a hindrance for some while supporting others? How could such problems be addressed in design and management of the application, if so?
- How can future navigation tools deliberately be designed for and make direct use of user generated content? What are the main issues concerning accountability and transparency which need to be addressed in design and management of such tools?

Research approach and methods

The researchers who are involved in the AUGMENT project come from various academic backgrounds, combining perspectives from Human Work Science, Informatics, Interaction Design and Rehabilitation Engineering. This has motivated a multi- and interdisciplinary approach in the project. The researchers have all, however, in one way or another, been working within the field of participatory design, eInclusion and eDemocracy for more than 10 years, which gives them, in some sense, a common ground and shared vision within the project.

Recent R&D involvement which has inspired the AUGMENT project was our participation in *The Planning Portal*, a 3-year (2006-2009) national project coordinated by The Swedish National board of Housing and Planning. This was a project which had the aim of developing integrated map-based planning services

[1] A WIKI (from the Hawaiian word *wikiwiki*, which means quick, fast) is a shared web site for cooperative reworking of versions and cooperative responsibility for content information.

for supporting national, regional and local planning authorities, but also for promoting enhanced public map-based e-services to citizens. Another source of inspiration was the charting of on-going practices of eParticipation on a national and international level which Ekelin was commissioned to carry out for the Swedish Government (2008). Ekelin was also appointed “teamleader” in a national consultation about future e-democracy policy by the Swedish Government during 2009. Anderberg has a Ph D in Rehabilitation Engineering and has developed a model called FACE which discusses disability as a situation dependent attribute rather than depending on the individual.

The research approach we have chosen concentrates on case studies, small-scale action-oriented R&D projects with a base in using qualitative ethnographic studies coupled with engineering development work. The basis for this approach is the Scandinavian tradition of workplace democracy with a deliberate use of multiple perspectives through iterative negotiation processes in ICT development. The aim is also to achieve conceptualization based on the interplay of practice and theory with a focus on participatory design processes.

Project description

The aim of this project is to develop public digital spaces, processes and infrastructure for user-driven co-construction of accessibility information, making use of existing handheld mobile phones which offer the possibility to upload pictures and comments via an application with a map-based interface such as that provided by Google Earth. The local development consists of constructing a suitable interface, a customer-generated database, and a wiki-solution for handling and maintaining data. The task is mainly charting of “unaccessible” places and the aim here is working with various groups of users and cases, for instance hospitals, public places, and common recreation places. The main issue here is to offer possibilities for direct participation by those affected. There is also a sister project running in Tamil Nadu, India called *The Walk-on-Water project* (Eriksén & Ekelin, 2008; Eriksén *et al*, in press), which has a different focus, but which we are using for trans-cultural comparison of evolving practices of user-driven and participatory design of public e-services based on co-construction among multiple local stakeholders of databases containing current, meaningful local information.

In order to be able to concentrate on design of an easy-to-use solution with the user as co-constructer, the development process in the pilot project is focused on a specific modification (or module) of an existing and established application. The maintenance issue of the shared database is both a key to success and a serious challenge. An example of a basic solution is Google maps Street View (http://maps.google.com/) where you can walk around on the streets virtually and examine pictures and surroundings by assessing information on the map. In real life, use of GPS-based technology with positioning makes it possible to contribute to the map content with personal photos and comments.

The locally developed interface contains a set-up of a user-driven accessibility database combined with a wiki solution in order to handle different versions of information (the information could be exemplified by scaling, individual evaluations, location of for instance toilets and so on). The aim of the project is to find new methods for continuous up-dating and ways to secure accurate, up-dated and high quality of status of accessibility in the local area.

Project benefits

The project benefits for the involved group of stakeholders are primarily practical: to jointly develop new ways of working around provision of accessibility information. It is also a way to gain goodwill for local authorities by the introduction of a customer-driven accessibility database which makes use of the implementation of a Wiki-solution in order to handle information. This is in line with recent development of new methods for accessible update of information and visions of creating good governance as well as shared responsibility for the quality of accessibility information. For the region, the suggested project is a way to offer improved accessibility for citizens at the same time as the affected groups are given a possibility of greater influence on the content of accessibility data as well as the presentation form and management of the data. On the political level, the issue of inclusion of all citizens is crucial, and the establishment of more well informed and democratic decision-making concerning accessibility issues is in line with visions of good governance. For the involved researchers, the project is expected to contribute to the development of more inclusive methods for participation, and a re-conceptualization of notions of accessibility and disability, as well as providing material for development of a new agenda for the Scandinavian approach to systems design with an even broader scope of direct participation than previously. We are also exploring differences between the related research traditions of end-user innovation and participatory design, and what we can learn from these differences, concerning how to provide useful feedback efficiently and effectively to software providers, software engineers and interaction designers, and thus support the development of sustainable infrastructures for inclusive design-in-use (Dittrich *et al*, 2002; Dittrich *et al*, 2009).

Issues to discuss further

Identified critical issues which are suitable for further discussion are listed below:

The possibilities/problems of shared database maintenance

The issue of maintaining a culture of engagement/participation beyond the use of rational solutions based on the rationale of CRM

The issue of accurate, valid and usable information seen as a basis for local continuous negotiations rather than the basic outcome

References

Anderberg, P. (2006*) FACE- Disabled people, Technology and Internet*. Doctoral dissertation. Available online at: http://www.certec.lth.se/doc/face/ .

Anderberg P, Olander E, Jönsson B. and Sperling L, (2007). Enabling design. Invited paper in Kumar, S (Ed.) *Rehabilitation Ergonomics.* CRC Press, Taylor and Francis Group

Anderberg, P., Flodin, E., Jönsson, B., Malmborg, L., Nordgren, C. and Svensk, A. (2005) Ethics in the making. *Design Philosophy Papers*, no 4, 2005

Axelson, von H. Kartläggning över initiativ för eInkludering i EU och Sverige. Delrapport: Regeringsuppdrag N2008/5985/ITP

Bjerknes, G and Bratteteig, T . 1995. 'User participation and democracy: A discussion of Scandinavian research on system development", *Scandinavian Journal of Information Systems* Vol. 7, No. 1: 73-98.

Blomberg, J., Giacomi, J., Mosher, A. and Swenton-Wall, P. 1993. 'Ethnographic Field Methods and Their Relation toDesign'. In *Participatory Design: Perspectives on Systems Design*. D. Schuler & A. Namioka (eds.) Hillsdale NJ: Lawrence Erlbaum

Blomberg, J, Burrell M and Guest, 2002. 'An Ethnographic Approach to Design,' in *Handbook of Human-Computer Interaction in Interactive Systems*. Jacko, J. and A. Sears (eds.), Lawrence Erlbaum Associates,Inc., New Jersey

Dittrich, Y., Eriksén, S. and Hansson, C. (2002), PD in the Wild; Evolving Practices of Design in Use. In Binder, T., Gregory, J. and Wagner, I. (eds.), *PDC 2002 Proceedings of the Participatory Design Conference*, Malmö, Sweden, June 2002, pp.124-134.

Dittrich, Y, Ekelin, A, Elovaara, P, Eriksén, S and Hansson, C., 2003. 'Making *e-government happen: Everyday co-development of services, citizenship and technology*', in Sprague, R H Jr. (ed.), *Proceedings of the 36th annual Hawaii International Conference on System Sciences [*HICSS'36], 6-9 January 2003, Big Island, Hawaii. Los Alamitos, California, & Washington, Brussels, Tokyo: IEEE Computer Society

Dittrich, Y., Eriksén, S. and Wessels, B. (2009), *From Knowledge Transfer to Situated Innovation: Cultivating spaces for co-operation in innovation and design between academics, user-groups and ICT providers*. Blekinge Institute of Technology Research Report 2009:01.

Ehn, P. 1993. 'Scandinavian design: On participation and skill'. In Schuler & Namioka, (eds.) *Participatory design, Principles and Practices. Participatory Design: Perspectives on Systems Design.* Hillsdale NJ, Lawrence Erlbaum

Ekelin, A. (2007): *The Work to Make eParticipation Work*, Dissertation No. 2007:11, Blekinge Institute of Technology

Ekelin, A. (2007b) Situating eParticipation. In *Understanding eParticipation. Contemporary PhD eParticipation Research in Europe.* [eds. Avdic, A ; Hedström, K ; Rose, J ; Grönlund, Å. *Örebro:* Örebro University, Department of Business, Economics, Statistics and Informatics

Eriksén, S. (2002), Designing for Accountability. In Berthelsen, O., Bødker, S. and Kuutti, K. (eds.), *NordiCHI 2002 Proceedings of the Second Nordic Conference on Human-Computer Interaction*, Tradition and Transcendence, Oct. 19-23, 2002, Århus, Denmark (pp.177-86).

Eriksén, S., Kawlra, A., Jhunjhunwala, A. and Ekelin, A. (in press), Planning beyond the Urban-Rural Divide: Participatory design of ICT for sustainable development in India and Sweden. Chapter in Bengs, C. [ed.], INSTEC book to be published autumn 2009.

Eriksén, S. and Ekelin, A. (2008), Beyond the Buzz: Participatory, sustainable, convergent and high quality public e-services - developing methods and practices in India and Sweden. In *Proceedings of the 31st Information Systems Research Seminar in Scandinavia IRIS 2008*

Macintosh, A. 2006. eParticipation in Policy-Making: The research and the challenges, in Cunningham, P &Cunningham, M. (eds.) *Exploiting the Knowledge Economy: Issues, Applications and Case Studies*. IOS press, 364-369.

Suchman, L**,** 1994. 'Working relations of Technology Production and Use,' *Computer Supported Cooperative Work (CSCW*). Vol. 2, pp. 21-39, Kluwer Academic Publishers, Netherlands.

Designing Cyberinfrastructure for Future Users

Matthew J. Bietz, Charlotte P. Lee

University of Washington, USA
{mbietz/cplee}@u.washington.edu

Abstract. Scientific information infrastructures are expected to operate over long time scales, but this creates challenges for the design of those infrastructures. This paper uses the example of cyberinfrastructures for metagenomics research to illustrate some of the issues that can arise when scientists attempt to use legacy cyberinfrastructures to answer new research questions. New science brings new forms of data, new analysis tools, and the need to recontextualize existing data. Cyberinfrastructure design is complicated by the difficulty of predicting future user requirements. We discuss three strategies for addressing these issues that are emerging in the metagenomics domain.

1 Introduction

Information infrastructures operate over relatively long time scales, but this creates certain challenges for the designers of these infrastructures. Often the infrastructure is expected to persist through funding cycles, changes in technologies, the coming and going of people involved in the project, and larger social and policy changes (Edwards, et al., 2007; Ribes & Finholt, 2007). One particularly difficult challenge is that as the infrastructure evolves, the user base may change. As users change their focus or new users arrive, they present a new set of requirements and infrastructure needs. Here we use the emergent field of metagenomics research to illustrate some of the challenges that arise when scientists begin to use existing information infrastructures to answer new research questions.

Metagenomics, sometimes called population genomics or environmental genomics, is a “new science” that allows scientists to study the genetic composition of populations of microorganisms to understand biological diversity, microbes’ functional roles, and microbial impacts on and adaptations to their environments. Metagenomics is an interdisciplinary approach, using the analysis of genetic sequence data to answer questions in fields as diverse as environmental remediation, cancer research, drug discovery, marine microbiology, and power generation (National Research Council (U.S.). Committee on Metagenomics: Challenges and Functional Applications, 2007).

Metagenomics is enabled by new laboratory methods, advances in sequencing technologies, and cutting edge information infrastructures. In the past, geneticists and genomicists had to isolate individual organisms and grow them in the laboratory in order to study their DNA. However, it has been estimated that less than 0.1% of the world’s microorganisms are amenable to culturing in the laboratory. New techniques and technologies have been developed that make it possible to bypass this culturing step while significantly lowering the cost of DNA sequencing. These changes give scientists access to a wealth of genetic information from organisms that previously could not be studied. Metagenomics also makes it possible to ask new questions about the relationships among organisms and their relationship to their environment.

The field of metagenomics provides an interesting case study in part because of its rapid growth. Indeed, the term was only coined in 1998 (Handelsman, et al., 1998), and by mid-2005, nine major metagenomic sequencing projects had been completed (Chen & Pachter, 2005). Interest in these techniques is growing: for example, the Metagenomics 2008 conference attracted more than 250 participants, and the NIH is embarking on a major project to study the human microbiome [http://nihroadmap.nih.gov/hmp/]. Here we use the emergence of metagenomics to demonstrate some of the ways that the introduction of a new community of scientists with new research questions and new information needs can challenge existing information infrastructures.

2 Our Study

This research reports on an ongoing ethnographic study of the development of cyberinfrastructure to support metagenomics research. This study includes both an in-depth examination of one particular cyberinfrastructure development project, and a broad survey of information infrastructures serving metagenomics researchers. We have conducted thirty-three interviews with metagenomics researchers, computer scientists, bioinformaticists, and others involved in the development of metagenomics cyberinfrastructures. We have also conducted over 100 hours of formal and informal observation, including attending development meetings, laboratory meetings, workshops and conferences. Interview transcripts and field notes were analyzed using a grounded theory approach (Glaser &

3 Cyberinfrastructures for DNA Sequenz Data

Scientific cyberinfrastructures are distributed enterprises supported by advanced technological infrastructures such as supercomputers and high-speed networks. In the genetic sciences, scientists have long recognized the importance of sharing sequence data, and have developed significant infrastructures for doing so. GenBank, for example, has been collecting and distributing DNA sequence data since 1982 (National Center for Biotechnology Information). GenBank is only one of many infrastructures that provide storage of DNA sequence data and facilities for analyzing and visualizing that data.

Data in these databases is submitted by the scientists who conduct the DNA sequencing and analysis. The field has strong norms around data sharing, backed up by a commitment by journal publishers not to publish analyses of genetic data unless the data is made public and submitted to GenBank or other databases (Marshall, 2001). While the databases may have their own underlying architectures, data sharing among the scientists and databases is supported by a strong standard called FASTA, which specifies a uniform file format for representing sequences using individual letters to stand for amino acids (National Center for Biotechnology Information). Many of these systems also provide tools like the Basic Local Alignment Search Tool (BLAST) which allow scientists to

4 New Questions for Old Infrastructures

Metagenomic analyses use the same sequence data that is used in other geneticsbased fields, and tools like BLAST are still useful to compare new genetic sequences to sequences generated by other scientists. Metagenomicists need some of the same basic functionality provided by infrastructures like GenBank. But at the same time, these infrastructures become more valuable when the design of the tools and databases have a good fit to the scientific questions being asked (Bietz & Lee, 2009). In this section, we discuss the ways that new metagenomics questions challenge the design of cyberinfrastructure.

4.1 New Data and Tools

Metagenomics and its associated laboratory techniques bring a new set of data storage and analyses requirements to existing cyberinfrastructure. One of the consequences of new DNA sampling and sequencing technologies is that DNA sequencing has become relatively inexpensive. While sequencing costs were around $10 per base pair in 1990 (Powledge, 2003), today researchers pay a few cents per thousand base pairs. The amount of DNA sequence data being produced is overwhelming, to the extent that data storage and computation requirements are outpacing Moore's law (Dooling, 2009).

In addition to simply having more data, metagenomics also assumes a different unit of analysis. Rather than focusing on the gene or even whole genome of an organism, metagenomicists work at the level of a community or population of microorganisms. Many existing sequence databases cannot easily represent this level of relationships among data.

One of the key focus areas in metagenomics is the relationship between microbes and the environment, but studying these relationships requires scientists to also collect contextual "metadata" that describe where the samples were found, including location, temperature, pH, etc. Most genetic and genomic databases were not designed to handle this level of data complexity.

Along with this new data, scientists need new tools to analyze and visualize the data. For example, a common question in marine metagenomics involves understanding how ocean temperature affects the diversity of the local microbiome. Not only would this require temperature data, but also the ability to query it, include it in analyses, and create visualizations around it. This kind of question would be almost impossible to answer with the data structures and tools provided in cyberinfrastructures created for traditional genetics and genomics researchers.

4.2 Recontextualizing Existing Data

Metagenomicists bring new data to existing infrastructures, but they also want to ask their new questions about old data. Often to ask a new question requires putting the old data into the new metagenomic context. For example, even if metadata were not stored in the database originally, there may be sources (like the publication record) that could be used to populate new fields in the database. However, reformatting data or retrospectively adding metadata are expensive tasks, especially when the work may need to be done again for the next group of scientists who pose a new question.

Another issue arises in that new metagenomic data may change the interpretation of legacy data. As metagenomic data is added at a phenomenal rate to theses database, the computational problems are becoming immense. One database developer told us:

> So you do need to go back from time to time and do all [the analyses] from scratch…. So the problem there is that we need to do periodic updates and periodic updates are every three months…. Now if new data is coming at an increasing pace, we are already at the point where even really big infrastructures and big computer clusters cannot really support all that.

Beyond these issues of computational power, scientists are also refining and expanding theory. In genetics and genomics, for example, scientists are finding that some prior assumptions about how genes operate, the role of "non-coding" regions of DNA, and evolutionary processes can necessitate a reconsideration of old data and interpretations.

5 Difficulty of Requirements Prediction

One question that arises is why these systems were not designed originally to support these new questions. If we accept the history told by many metagenomics researchers, metagenomics is a "logical progression" from genetics and genomics, and these future needs could have been predicted.

> The concept was simple: Take seawater and capture all the microorganisms swimming in it on filters with microscopic pores, isolate the DNA from all the captured organisms simultaneously.... Rather than focusing on the hunt for one particular type of life, we would obtain a snapshot of the microbial diversity in a single drop of seawater-a genome of the ocean itself. This was, to me, a straightforward extension of work that had started with the EST method and led to the whole-genome shotgun approach, then the first genome of an organism in history, and then of course to the human genome. (Venter, 2007, p. 345)

While this version of the origin of metagenomics creates a compelling narrative, it does not recognize two important features of these scientific changes. First, as science has "progressed" through these phases, it has not left old questions behind. There are still scientists who are studying the functions of individual genes, and there are still scientists who are studying the genomes of individual organisms. Metagenomics has not supplanted these fields. In fact, it is essential for metagenomicists that research continues in genetics and genomics:

> It would help us tremendously in doing metagenomics if we had a wide range of reference genomes.... The NIH is funding 400 complete genomes of microbes that live in humans. And again, these are to give us standards and to allow us to interpret metagenomic data more rigorously. So first of all, as far as I'm concerned, we've only begun to sequence. We need to sequence - whole genome studies need to go on to expand the opportunities in studying evolution and getting many specific genes and models for human disease and for understanding biology.

Not only are genetic and genomic studies important for metagenomics, new metagenomic techniques are also changing the way geneticists and genomicists do their work. For example, shotgun sequencing not only allows for the sequencing of populations of microorganisms, it also makes it possible to sequence genomes from organisms that could not be cultured in a laboratory.

Secondly, the progression from genetics through genomics to metagenomics is logical only in retrospect. The development of metagenomics was by no means a foregone conclusion, and scientists found that they had to work hard to convince their peers that these techniques were valid. One scientist explained the difficulties she experienced finding a venue in which to publish her work this way:

> Not only has there been this distrust between the two fields, the genomics and the traditional fields—I think it's becoming more acceptable—but now metagenomics has come in too. So we're not just talking about sequencing entire genomes, we're talking about populations of genomes and defining what's there based solely on sequence similarities to those genomes.
>
> So what I've - I'm taking a huge leap here. I'm saying I have these 50,000 sequences. They're very distantly related to these sequences from [other] genomes. I know nothing about their physiology. I don't know what they infect. I don't know their reproductive lifecycle. I don't know anything about them. I'm just giving them a name based on the history of those sequences. So I think I'm taking an even farther leap.... And I think we try not to tread too heavily upon people's toes. We don't want people to think we're trying to take over their fields and that these approaches are the end all to the field.

Traditional approaches to identifying microbes rely on direct examination of microbes' physiology, pathogenesis, and reproduction, this scientist found that using only metagenomic techniques was not readily accepted by peer reviewers. Even though some scientists see metagenomics as a "straightforward extension" from earlier techniques, this new way of looking at microbial populations was not predicted by early geneticists and genomicists, the science is not without its detractors, and it is not entirely clear how these techniques will unfold into the future.

These observations highlight significant challenges for the development and maintenance of cyberinfrastructure. As science changes over time, scientists will need different things from cyberinfrastructures. While some research questions will persist, others will change and new research questions will be asked. A new science like metagenomics brings new questions and new communities of scholars with different ways of understanding the world. The requirements for information infrastructures develop and change as the science and communities change. Just as it is impossible to predict with any certainty how a scientific field will develop, it is equally impossible to predict all future information infrastructure requirements.

6 Infrastructure Adaptation

So far we have focused on the challenges that a new science can pose for an existing scientific information infrastructure. Determining the best methods to address these challenges remains an open question, but the metagenomics field provides examples of three different approaches.

One strategy that has been adopted has been to create work-arounds for existing infrastructures to adapt them to new uses and questions. For example, GenBank does not provide much support for the contextual metadata that is key to metagenomics approaches. Metagenomics researchers have begun to add metadata to free-text comment fields, sometimes using a “structured comment” that mimics a table of fields and data. This provides the benefit that it can be used immediately and without much disruption to the existing infrastructure, and allows the new metagenomics researchers to store contextual data without affecting how other geneticists or genomicists use the system. On the other hand, work-arounds like these are often difficult to use, lack standardization, and often do not provide full integration of the new science.

A second approach involves modifying or extending existing infrastructures to support metagenomic data. This seems to be happening in systems like IMG (Markowitz, et al., 2008) and The SEED (Overbeek, et al., 2005), which have extended their systems to include new metagenomics tools and support for metagenomics data. While this provides greater integration than the workarounds, there can still be a disconnect between legacy data and new tools.

A third approach, taken by projects like CAMERA (Seshadri, et al., 2007), creates new infrastructure from scratch specifically to support the new science. While this may provide the best fit to the scientific questions, it can also be a very expensive option, and can make it more difficult to use legacy data and tools. Splitting off from existing infrastructures may also reinforce the separation between communities that may benefit from greater interaction.

7 Conclusion

Scientific information infrastructures that persist over long time scales must respond to the emergence of new science. New science brings with it a new set of research questions, data and tools, scientific communities, and ways of understanding legacy data. There is a need for a deeper understanding of how developers of cyberinfrastructure can manage the evolution of user needs and requirements, and to understand when it is better to extend an existing information infrastructure and when it is necessary to create new infrastructure. The introduction of metagenomic approaches in molecular biology highlights the dynamic nature of both the human and technological aspects of cyberinfrastructure.

8 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990): 'Basic local alignment search tool', Journal of Molecular Biology, vol. 215, no. 3, Oct 5, pp. 403-410.

Bietz, M. J., & Lee, C. P. (2009): 'Collaboration in metagenomics: Sequence databases and the organization of scientific work' ECSCW 2009: Proceedings of the Eleventh European Conference on Computer Supported Cooperative Work, London: Springer-Verlag.

Chen, K., & Pachter, L. (2005): 'Bioinformatics for whole-genome shotgun sequencing of microbial communities', PLoS Comput Biol, vol. 1, no. 2, Jul, pp. 106-112.

Dooling, D. (2009): Maximizing utility of genome sequence data. Paper presented at the The Biology of Genomes. from http://www.politigenomics.com/2009/04/cshl-biology-ofgenomes-2009.html

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007): 'Understanding infrastructure: Dynamics, tensions, and design', from http://www.si.umich.edu/~pne/PDF/ui.pdf

Glaser, B. G., & Strauss, A. L. (1967): The Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Aldine de Gruyter.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998): 'Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products', Chem Biol, vol. 5, no. 10, Oct, pp. R245-249.

Markowitz, V. M., Ivanova, N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., et al. (2008): 'IMG/M: a data management and analysis system for metagenomes', Nucleic Acids Research, vol. 36, pp. D534-D538.

Marshall, E. (2001): 'Bermuda Rules: Community Spirit, With Teeth', Science, vol. 291, no. 5507, 16 February 2001, pp. 1192.

National Center for Biotechnology Information: 'FASTA Format Description', Retrieved 14 April, 2009, from http://www.ncbi.nlm.nih.gov/blast/fasta.shtml

National Center for Biotechnology Information (April 2, 2008): 'GenBank Overview', Retrieved February 23, 2009, from http://www.ncbi.nlm.nih.gov/Genbank/index.html

National Research Council (U.S.). Committee on Metagenomics: Challenges and Functional Applications (2007): New science of metagenomics: Revealing the secrets of our microbial planet. Washington, D. C.: National Academies Press.

Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Chuang, H., Cohoon, M., et al. (2005): 'The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.', Nucleic Acids Research, vol. 33, no. 17, pp. 5691-5702.

Powledge, T. M. (2003): 'How many genomes are enough?', The Scientist, vol. 4, no. 1, pp.

Ribes, D., & Finholt, T. A. (2007): 'Tensions across the scales: Planning infrastructure for the long-term' Proceedings of the 2007 International ACM Conference on Supporting Group Work, New York: ACM, pp. 229-238.

Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007): 'CAMERA: A community resource for metagenomics', PLoS Biology, vol. 5, no. 3, pp. e75.

Venter, J. C. (2007): A Life Decoded: My Genome: My Life. New York: Viking.

Condor and Philosophy of Flexibility: A Case Study of Two Scientific Collaborations Using High-Throughput Computing

Sara Kraemer and Chris Thorn

Wisconsin Center for Education Research
University of Wisconsin-Madison
sbkraeme@wisc.edu, cathorn@wisc.edu

Abstract. In this paper, we summarize two scientific collaborations using Condor at the University of Wisconsin-Madison, a distributed High-Throughput Computing (HTC) resource that supports a range of scientific collaborations, including cross-disciplinary, virtual and co-located teams, among others. The Condor project embodies a 5-component philosophy of flexibility that addresses the barriers to unpredictable distributed operating environments. We interviewed 2 scientific collaborations using Condor as the primary source of HTC resources to describe how the philosophy of flexibility works within each of the collaborations. We have adopted a sociotechnical approach to investigating these teams and outlined some preliminary characteristics of HTC teams using Condor.

1 Introduction

An increasing range of scientific research is addressing problems that can best be addressed using computationally-intensive analytical tools. High-throughput computing (HTC) addresses this need by supplying large amounts of computing power over distributed networks. We define HTC as an environment that can deliver large amounts of processing capacity over long periods of time. In addition to computational cycles delivered, there is a second, critical measure of system quality: HTC systems are designed to be extremely fault-tolerant and require minimum human intervention (Thain et al., 2005). By design, these technologies enable and support distributed teams. These and other characteristics channel the interactions and forms of collaboration that emerge when users from various scientific domains use HTC resources to work on computational problems.

Scientists in many disciplines have begun revolutionizing their fields by using HTC resources in technology-mediated, distributed-work environments. Some of these trends have included the extension of complex simulation and modeling from classic approaches to scientific research (i.e., theoretical/analytical, experimental/observational) (Atkins et al., 2003). In addition to these new trends in computational science, one of the central HTC considerations is sociotechnical implications of shared computational resources, ownership, and cooperation (Thain et al., 2003).

Some HTC systems, such as the Condor© project at University of Wisconsin-Madison, have unique characteristics to address these implications, such as high degrees of resource flexibility, end-user control, open-ended planning, and distributed resource management (Thain et al., 2005). The Condor project embodies a *philosophy of flexibility*, which has allowed HTC systems to flourish in highly unpredictable distributed operating environments (Thain et al., 2005).

In this paper, we will explore the sociotechnical characteristics of Condor's philosophy of flexibility with 2 case studies of scientific collaborations at the Grid Lab of Wisconsin (GLOW). GLOW is a cluster of cooperative computing resources using Condor at the University of Wisconsin-Madison (GLOW, 2009). We will first describe the overall GLOW project and Condor's philosophy of flexibility followed by a case study analysis of two GLOW teams.

2 GLOW, Condor, and Philosophy of Flexibility

GLOW is a distributed scientific computing resource at the University of Wisconsin-Madison that combines and enhances autonomous sites of computing resources. GLOW is an interdisciplinary effort that spans 10 scientific domains: Biostatistics and Medical Informatics, Chemical and Biological Engineering, Chemistry, Computer Sciences, Engineering Physics, Genomics, Genetics, Materials Science and Engineering, Medical Physics, Physics, and Astrophysics. The laboratory consists of 8 physical sites and individually provides the necessary hardware, software, and support infrastructure for the development and experimental evaluation of HTC applications. Each of the sites focuses on addressing local computational needs and maintains full control over local resources while sharing unused computing power and storage space across site boundaries according to a group defined policy. The goal of the *laboratory* is to bring together domain and computer scientists to make HTC computing an effective tool for scientific research by harnessing and sharing the power of commodity resources. GLOW members collaborate in the development, implementation, testing, and deployment of grid-enabled capabilities while cultivating interdisciplinary science.

The underlying computational cyberinfrastructure and resource management for GLOW is supplied by the Condor project. Condor has characteristics that differentiate it from other HTC resources. For example, because Condor runs on many computing platforms and operating systems and can execute any software that does not require user interaction. Condor also offers a wide range of tools that are readily available for users—from commercial research software to scripting engines and compilers. In addition, the available scientific tools allow individual scientists or teams to engage with the Condor HTC environment using tools familiar to them. Further, enabling of existing tools in an HTC setting provides critical social and technological gateways for new adopters of HTC. Access to the HTC environment also exposes new adopters to tools and methods used by others to address similar computational and/or analytical problems. In this way, scientists' skills and knowledge are affected by the capabilities and characteristics of HTC technologies and tools.

The Condor project embodies a philosophy of flexibility; this philosophy that has served the allowed the design to flourish in a highly unpredictable distributed operating environment (Thain et al., 2003). International distributed systems are heterogeneous in numerous ways: they are composed of many types and brands of hardware; they run various operating systems and applications; they are connected by unreliable networks; they change configuration constantly as old components become obsolete and new components are become online, and they have many owners with local policies and requirements that control their participation in the community. Condor has adopted a 5-component flexibility philosophy to address these barriers:

(1) *Let communities grow naturally.* Given tools of sufficient power, people will organize the computing structures they need. However, human relationships are complex, and people invest their time and resources to varying degrees and relationships and requirements change over time. Therefore, Condor design permits but does not require cooperation.

(2) *Leave the owner in control, whatever the cost*. To attract the maximum number of participants in a community, the barriers to participation must be low. Users will not donate their property to the "common good" unless they maintain some control over how it is used. Therefore, owners of computing resources are given the tools to set policies and retract resources for private use.

(3) *Plan without being picky*. Plan for slack resources as well as resources that are slow, misconfigured, disconnected, or broken. The designers of Condor

spend more time and resources contemplating the consequences of failure than the potential benefits of success.

(4) *Lend and borrow*. The Condor project has developed a large body of expertise in distributed resource management and aims to give the research community the benefits of their expertise while accepting and integrating knowledge and software from other sources. They have also instituted a mechanism for collective problem-sharing and solving among its users.

(5) *Understand previous research*. The Condor project continually updates its organizational knowledge with previous research to apply well-known fundamentals as well as cutting-edge techniques to emergent problems. The inclusion of current user innovations keeps the work focused on the edge of discovery rather than wasting effort remapping known territory.

As outlined in the philosophy of flexibility, the Condor approach is more than a complex set of computational resources. The Condor team maintains a close intellectual partnership with GLOW teams and works together on the challenges of HTC in the context of break-through science. Condor has advanced HTC technology via improvements in their software coupled with innovations in the computational approaches with the domain scientists. These interactions have made Condor privy to numerous sets of interdisciplinary virtual/co-located teams as well as numerous types of sociological and technological factors encountered in research settings. In this next section, we will highlight some of these factors in 2 GLOW teams: IceCube and the Laboratory for Molecular and Computational Genomics.

3 GLOW Teams: IceCube and The Laboratory for Molecular

We have conducted group interviews with 2 GLOW teams to assess how Condor's philosophy of flexibility is implemented in live research settings (see Appendix for interview guide). The first team is IceCube, a collaboration supporting a neutrino detector at the South Pole; the second team (lab) is the Laboratory for Molecular and Computational Genomics (LMCG) at the University of Wisconsin. The IceCube group interview consisted of the P.I. and computational resource scientist who manages the HTC for IceCube group at UW-Madison. The LMCG group interview consisted of the P.I., 2 research scientists, and 1 post-doc.

UW-Madison is the lead institution for the construction and operation of IceCube, as well as the largest group of faculty, scientists, post-docs, and students in the international IceCube collaboration of over 250 people in 35 institutions. IceCube, a telescope under construction at the South Pole, will search for neutrinos from the most violent astrophysical sources: events like exploding stars, gamma ray bursts, and cataclysmic phenomena involving black holes and neutron stars. The IceCube telescope is a powerful tool to search for dark matter, and could reveal the new physical processes associated with the enigmatic origin of the highest energy particles in nature. IceCube encompasses a cubic kilometer of ice and uses a novel astronomical messenger called a neutrino to probe the universe. (IceCube Neutrino Observatory, 2009).

GLOW's role has touched all aspects of commissioning and operation of the detector, especially in the challenging production of real-time detector simulations. The scientific analyses and simulations of for UW-Madison scientists rely completely on GLOW resources; these analyses and simulations will also completed for other collaborators on the IceCube team once the operational phase of the detector is fully employed.

GLOW has assisted in the timeliness of delivering the results of simulations and experiments to IceCube team members, which would yield a competitive advantage. Further, since the IceCube collaboration is about ¼ the size of typical particle physics collaboration they are able to more easily communicate, share, and work on findings which fosters an overall culture of cohesiveness (compared to typical particle physics collaborations). Further, the on-time delivery of data results has also increased the level of the group's productivity; because many tasks and data results are interdependent to ongoing work in other scientists.

The use of Condor has also emerged independently in several IceCube collaborating institutions. Some of their European collaborators share access to distributed researchers with groups using grid technologies.

The LMCG investigates single molecule phenomena for the creation of new systems in the biological sciences (LMCG, 2009). The size of the LMCG is smaller, 12-13 people in this lab, and the entire lab is located at UW-Madison. Within this group, many disciplines are represented; including but not limited to chemistry, statistics, bioinformatics, engineering, and genetics. The LMCG team has also had an in-house computational resource "expert" facilitate the work between Condor and LMCG (although this position is currently vacant). They felt that this facilitator role was an immense resource for identifying LMCG's scientific needs and effectively translating them into computational requirements for Condor. This role also proactively anticipated any problems or potential concerns, and mitigated them on behalf of the LMCG to Condor. LMCG attributed the accomplishment of many large research goals to close collaboration between the LMCG facilitator and the Condor team of experts.

3.1 Philosophy of Flexibility in GLOW

Let communities grow naturally, leave the owner in control, plan without being picky

GLOW was formed as a collective effort of domain scientists to share computational resources at UW-Madison in collaboration with computer scientists at Condor. They have a shared interest in consuming a large amount of computational resources for their research problems and therefore found that sharing slack resources would benefit their projects in key ways, such as computational speed, efficiency, and minimizing complexity. Because of the open-ended nature of the Condor philosophy, GLOW participants each have individual input for decision-making of the collective.

For example, a critical organizational design piece of GLOW is the monthly meeting of GLOW teams. At this monthly meeting, decisions about resource allocation and usage are made among the group; attendees are typically those who organize computational resources for each physical site. This meeting was described as a time for "allocating opportunistic time". In addition to these monthly meetings, individual research sites contact and request resources and expertise directly to the Condor team. The IceCube and LMCG projects both reported that their computational resource needs have always been met, although they may experience differences among GLOW team members about how resources should be divided (this is very rare). IceCube noted because of this intra-group structure and organization of available slack resources, there is very little to no competition among the collective for slack resources.

Lend and borrow, understand previous research

Both GLOW teams relied heavily on Condor's expertise in distributed resource management to both deliver the computational power needed to execute their scientific algorithms, models, and simulations, but also contributed to creative approaches to complete their work in ways that allowed the scientists to more efficiently and collaboratively. In addition, the Condor team both supports the use of Condor generally as well as engages very deeply with local clients' and their projects, supporting them in the defining and redefining of their research problems, assisting with interdisciplinary collaboration, and providing technical assistance, which was echoed at both sites. The LMCG noted that over the years-long collaboration with Condor, they have found working with the Condor system has progressively become easier to use and they attributed this improvement to continuous feedback Condor elicits from user groups as well as current research from in the field of study.

Some examples of Condor's help have included assistance re-tooling complex algorithms to run on Condor. At the LMCG, the computational resource staff person would interact with Condor as well as the domain scientists (a mathematician in this example) to figure out a way to re-tool the algorithm.

Further, because of the availability of computational resources, the revised algorithm did not have to be written in a way that conserved computational resources. As a result, the time needed to complete the new algorithm was lessened considerably. This scientist noted that Condor provided "the gift of time" to work on other problems.

4 Future Research

This paper summarizes some of our preliminary work in investigating how Condor's philosophy of flexibility manifests in interdisciplinary collaborations. We are currently studying other GLOW teams and aim to further explore how these tenets contribute to or hinder the effectiveness of teams' collaboration and scientific output. We also are very interested in developing our understanding of virtual team performance using Condor and the similarities and differences of those various teams, from a sociotechnical perspective. An analysis of the social impacts of the technical configurations of Condor, may lead to deeper understanding of how HTC is used as an effective enabler of new scientific problem sets, solutions, and collaboration configurations. In addition, this investigation can also influence the design of Condor, and HTC technology in general, to meet emerging scientific problems and configurations.

Appendix: Interview Guide

1. Do you think that using Condor and interacting with Condor team members has contributed to breakthroughs in your science?
    a. How has it specifically contributed to breakthroughs? Examples needed.
2. Does Condor allow you to organize your computational resources, and by extension, team and work, in a way that best reflects your research needs and project tasks?
    a. If not, why? If so, how? Examples.
    b. Has using Condor changed the way you work with other scientists, both within and outside of the team? How?
3. How do you set Condor policies and settings within your project?
    a. Do you feel any aspect of your work is affected by sharing computational resources with other Condor users?
    b. Does sharing Condor use facilitate or hinder any aspect of your science? Does it vary at all across different scientific disciplines?
4. Have you encountered any problems with using Condor as a computation resource? If so, how?

a. What are the best aspects of using Condor for virtual teamwork and interdisciplinary scientific collaboration?
5. Has interacting with the Condor team affected the quality of scientific output and/or the collaboration among team members?
b. If so, how? If not, why not? Examples?

Acknowledgments

We would like our collaborators at the University of Wisconsin: Miron Livny at the Condor Project, Francis Halzen at the IceCube project, and David C. Schwartz at the Laboratory for Molecular and Computational Genomics.

5 References

Thain, D., Tannenbaum, T. and Livny, M. (2005): Distributed computing in practice: The Condor experience. Concurrency and Computation: Practice and Experience, 17 (2-4). 323-356.

Atkins, D.E., Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P. and Wright, M.H. (2003): Revolutionizing science and infrastructure through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on cyberinfrastructure: National Science Foundation.

Thain, D., Tannenbaum, T., Livny, M., Berman, F., Fox, G. and Hey, T. (2003): Condor and the Grid. in D. Hutchinson (Ed.), Grid Computing: Making the Global Infrastructure a Reality (pp. 299-335): Johh Wiley & Sons.

GLOW. (2009): http://www.cs.wisc.edu/condor/glow/.

IceCube Neutrino Observatory. (2009): http://www.icecube.wisc.edu/info/.

LMCG. (2009): http://www.lmcg.wisc.edu/.

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 an 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 goal[1] 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 beconsidered as an end user initiated case of e-Research (see Jirotka, 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 & Jirotka, 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 legislations[2] 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 if 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. & Jirotka, 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). Finlands Legislation. 1999/523. (http://www.finlex.fi/en/laki/kaannokset/1999/?_offset=20&_max=44)

Jirotka, M., Procter, R., Rodden, T. & Bowker, G.C. (2006): Special Issue: Collaboration in e-Research. Computer Supported Cooperative Work, 15, 251-255).

Shared care and the use of collaborative information systems: A new health reform and old pitfalls

Vigdis Heimly

Department of Computer and Information Science, Design and use of Information System Group, Norwegian University of Science and Technology, Trondheim, Norway and Norwegian Centre for Informatics in Health and Social Care, Trondheim, Norway
vigdis.heimly@idi.ntnu.no

Abstract. The paper addresses the increasing need for collaboration in the Norwegian health sector, and how information infrastructures[3] can facilitate exchange and sharing of health information. An upcoming national health reform in 2009 will have focus on how the patient can get health services in, or closer to, their homes. The change in the cooperation processes between primary and specialized care will trigger the need for better collaboration platforms. ICT-systems that support collaboration have been available for more than 10 years, but they are still in limited use. Some indications of why this process has been so difficult are given as a basis for development of new systems that can support the health reform. The paper is based on a survey to the Norwegian hospitals in 2008, semi-structured interviews at hospitals and with GPs, participation in meetings with end users and available documentation.

1 Introduction

Information infrastructures offer a shared resource for delivering and using services in places where users interact. In a CSCW-context (Harrison 1996), place

[3] Information Infrastructures (Hanseth 2004) are defined as a shared, evolving, heterogeneous base of IT Inbased on open and standardized interfaces.

is defined as a cultural and communally-held understanding of the appropriateness of styles and behaviour and interaction, which can be organised around spatial features. ICT-systems that support shared care can be used in places where health workers from different organizations and patients interact. Shared care is cooperative healthcare across organizational- and often also geographical borders. Shared care will typically involve a diversity of health workers as General Practitioners (GPs), medical specialist, nurses, midwifes or physiotherapists. Design of collaborative systems for shared care is challenging, because it requires an understanding of the nature of the collaborative work processes and an ability to forsee how new collaborative tools can support existing or future work-processes.

ICT has been used as a tool to support the clinician's work-processes in Norway for more than two decades. The first Norwegian Electronic Health Record (EHR)-systems for GPs were in use as early as in 1984. 98% of the GPs have had these systems in daily use since 2001 and EHR-systems are also present at all Norwegian hospitals. These systems started as administrative tools, but have over time emerged to be systems that support daily clinical work-processes. The focus has also changed towards shared care that involves several caretakers in primary and specialized care. The electronic collaboration between the caretakers in different organizations has so far mainly been based on electronic messaging, but web-based solutions and access to shared core medical information are also coming up as new options. Deployment of electronic messaging has been much slower than initially expected. This has proven to be more related to organizational challenges than technical barriers (Heimly 2007).

2 Challenges and changes in the Norwegian health sector

The Norwegian health system has changed a lot since the first EHR-systems were developed. A major health reform in 2001 led to the organization of the 81 Norwegian hospitals under 5 health enterprises that are owned and operated by the government. All patients are assigned to **one** GP's patient list. All primary contacts with the health care system, except acute care, should be channeled through the GP. Most patients who are admitted to the hospital have been referred by their GP. When the patient has finished the treatment at the hospital, the normal procedure will be to return the patient to community care under the GP's responsibility.

The Norwegian health system has obvious challenges that also are visible in other European countries. The hospital administration wants to keep the patient stay as short as possible in order to reduce hospital costs, but patients who have finished the specialized care they need at the hospital but are waiting for transfer to nursing homes or are not well enough yet to move to their own homes, are filling up hospital beds. As people live longer and longer due to better health care, more and more patients will need care on their elderly days. Many people are also rescued from a sudden death as early newborns or in traffic accidents, but may need specialist care for a long period.

A new Norwegian health reform is expected in the autumn of 2009. The reform will have focus on how the patient can be provided with more health care in community care, closer to their homes, and reducing the need for expensive specialized care. This health reform is also likely to be followed by economic incentives, and resources will be transferred from the hospitals to the municipalities. The municipalities will have to pay the hospitals according to the number of patients they refer to specialized care, and there will also be a high cost to pay for patients who have finished their hospital stay, but have to wait for community care to be organized.

Orlikowski has had a focus on the need for better understanding for how technological systems interact with political actions and human choices (Orlikowski 1992, Orlikowski 2001). The implementation of the new health reform in Norway will need to be followed by both organization changes in the health sector and the development of improved ICT-solutions for shared care. These change processes should be coordinated.

3 Computer Supported Collaboration between caretakers in different health organizations

The coming health reform is likely to put an even higher pressure on the need for collaboration. Higher speed in the treatment line, and the possibility of rising costs for the municipalities due to delays and prolonged hospital stays will make the need for availability to the right information at the right time essential. It is likely that ICT-solutions for sharing of essential health information in core databases will become more common. It is also a trend towards web-based solution that are owned and operated by hospital or private actors where there is a strict control both on which input should be registered in the systems and which information should be shared.

3.1 The technical infrastructure is available

The motorway for information sharing and exchange in the Norwegian health sector in available to many actors. The Norwegian Health Net (NHN) is a closed secure high speed network that connects almost all hospitals and GPs. An increasing number of municipalities with nursing homes and home care offices are also connected to the net. One of the main uses of the health net is broadband communication between the hospitals, but more and more information is also exchanged between hospitals and primary care. The main challenge so far has been that a very limited number of services are available. The Norwegian health net is a technical infrastructure, but only to a limited degree an information infrastructure. Development of end user services has so far mainly been the communicating parties' responsibility. The new health reform will suggest that NHN shall be owned by the government and not the 4 Regional Health Authorities that operate the hospital as today. This intension is to emphasize that the health net is available for all actors in the health sector. The new NHN will also get an extended responsibility for adding new services to the net. This will probably also include collaborative systems as a national core EHR.

Existing services that are available in NHN are message exchange (discharge summaries, referrals, lab requisitions and results..), web-based solution for requisition of laboratory tests and different telemedicine solutions.

3.2 Some tools to support collaboration are present

In order to make the treatment chain between primary care and the hospital as efficient as possible, there is a need to register, communicate, and interpret the information that is exchanged by all the involved parties. The information can either be sent as a message, the receiver can actively get access to information that is stored by the other party, or the sender can actively register information in a system held by the cooperation partner. It might also be possible to share information in a system held by a third party. The selected technical solution can depend on national legislation, and agreements between the communicating actors.

In Norway the most commonly used alternative is messaging between GPs and hospitals (referrals and discharge letters..). A few hospitals use a web-based referral system where the GP registers the referral in the hospitals system. Core EHR-systems that includes the most essential information about medication and contact are at a pilot stage.

4 Lessons learnt

Obviously changing the cooperation process from paper to electronic should involve much more that just replacing the paper form with a corresponding electronic form that is sent electronically. Different alternatives will have implications on the involved health-worker's work-processes. How can I make sure that I get access to the right information when I need it? How can new possibilities for collaboration be used as a means to improve the quality of the information that is shared? How can I be aware that new information is present, at how can I make other parties that I have added new content that might be of interest? If the work-processes are changed, and the workload is shared between the health workers in new ways, how can we assure that the actors trust each other and support the new changes? A series of semi-structured interviews with users of existing systems used in shared care haves, a survey to the hospitals, participation in meetings with project managers and reading of reports and other documentation has provided some clues to factors that should be paid special attention when new systems should be designed and developed.

4.1 Awareness in collaborative health systems

4.1.1 To whome should I display my actions, and whose actions should I monitor?

Souza (Souza 2007) focuses the problem of "To whome should I display my actions, and whose actions should I monitor"? These questions are highly relevant in shared care because health workers need access to health information that is updated by many parties. Awareness of when new content is added is important, but should on the other hand not be to disturbing in the daily work-process. GPs that have been involved in a Norwegian core medical chart project (Heimly 2009) were very concerned that they should be disturbed in their daily work by flags of alarms that are popping up on their screen. They did not want be informed immediately when medication is prescribed for their patients by other doctors, but wanted to check this on a list at a daily basis.

4.1.2 Enough, but not too much information

Why would GPs want to send information as messages when information could have been shared? First of all: Legislation in many countries does not permit doctors at different levels in the treatment chain to share medical information. Information sharing requires the patient consent, and consent-based systems are not always practical in daily use. The legislation in Norway is changing very slowly, and is still quite restrictive. The introduction of a proposal for a law-change that will permit sharing of core-EHR information based on consent has led to heated debates in the media. Patient seem to be very reluctant when it comes to how much information should be shared, and patient organizations seem to be more concerned with the possibility for sensitive information in the wrong hands than the possibility for better treatment if the clinicians have access to the right information at the right time.

But the most important factor is probably that the doctors only want to have access to the information they need, and not all the information that could possibly be available about the patient. A better structure of the medical record and better possibilities for filtering of information could have helped on this problem, but we not there yet. Most of the EHR-information is just a big lump of free text. Important information can be hidden in the hospitals EHR-information, and the GP does not want to have the responsibility for searching through all this information in search for something he or she does not even know is present. Instead of sharing all information, doctors seem to be more happy with getting the information they need transferred as an abstract, og getting access to some core information about the patient as current medication, diagnoses, allergies and updated demographic information.

4.1.3 The purpose of the information, documentation for you, me or the patient?

Documentation that can be present in shared care can be produced for use in one context, but can be used by other actors in a different context. When a specialist writes information into an EHR, the recorded documentation might be used in several contexts:

- Documentation as a part of the internal work-process that covers the treatment of the patient at the hospital. The hospital-stay should be as short as possible, but on the other hand, the patient should also be well enough to not be readmitted within a short period of time The patient will normally be treated by many doctors and nurses at different shifts, and accurate information about the patient’s medical condition, medication and treatment plans needs to be available at a “need to know” basis.
- Another goal is to document the work he or she has done in order to satisfy the legislation Complaints from the patients about procedure failures and maltreatment is getting more and more common, and thus documentation of the actual treatment and procedures followed is getting more and more important.
- Documentation for the patient. The patient is getting closer and closer to a customer, and requests access to their own EHR. Many patients even have bedside access to their own EHR. This also means that the EHR-documentation must be written in a language that is understandable for non-experts.

- Documentation for the next level in the treatment chain. The GP would request EHR-documentation that is important for further treatment when the patient returns to primary care. The GP would typically not be interested in details regarding surgery or a cure that was given during the hospital stay. Information about current medication when the patient leaves the hospital is on the other hand important, and information about the outcome of the hospital stay, scheduled appointments with the specialist and expectations for further treatment in primary care.
- Documentation for reporting to national registers, eg a “patient register” with administrative information about hospital stays or quality assurance registers as the Norwegian Cancer Registry.
- Documentation for reimbursement. In Norway hospital get paid from the government according to have many patients and which diagnoses they treat on an annual basis.
- Documentation for research purposes.

The GPs are very concerned with the amount of time that is spent on documentation and the registration process has to be as efficient as possible.

His or hers income is likely to depend on the number of patients treated, and the time for each consultation is very limited. Documentation of the outcome of the consultation, suggested treatment plan, scheduled appointments and medication are examples of information that should be present in the GP’s EHR. If the GP decides to refer the patient to a specialist, sufficient information for making the appointment should be provided.

4.2 Trust

Trust is important in collaborative work, but it is a challenge for health workers, as for most other people, to trust others recommendations. This can particularly seem difficult when you interact with people that you do not know very well. As an example, the waiting-list coordinator commented during an interview that a project where GPs could refer patients directly for hernia surgery at the hospital ward without passing through the outpatient clinic was terminated because there had been several cases where the hernia could not be found when the patient was admitted to the hospital. The specialists at the hospital meant that the GPs were not qualified for choosing patients for surgery. In interview with a representative from the hospital management later, it was on the other hand claimed that “missing hernia” would also often be the case even if the patient was admitted via the hospitals outpatient clinic, and that the problem was not necessarily related to the GPS competence.

The health-workers in different organizations seem to need to get a better understanding of the cooperating actor's work-processes. Norway has so far had positive experiences with practice consultants who are GPs that work in part-time positions at the hospital. This could typically be 2 days a month. Their mandate is to work with improvement of procedures that are related to collaboration between primary and specialized care. Some examples of activities are: revisions of procedures for referrals, templates for documents that are communicated, eg discharge summaries, referral og lab. reports. The practice consultant will also often be used as resource persons in projects where new ICT-solutions to support shared care are introduced. The practice consultants practice would often be used as a pilot site. According to the survey to the hospitals in 2008, 75% of them have practice consultants, and the hospital reported that they have good experiences with their effect on improvement on collaboration.

5 Results and recommendations

Information that is supposed to be shared need to be suited for the context in question. A common understanding of the needs of actors who are going to share the health information should be developed over time, and should also imply changes in both specifications of data, user interfaces and technical solutions over time.

"Me deciding requirements for You" is seldom a good solution. Collaborative systems need to develop over time and changes in user interfaces need to be easy to implement. The tension between doctors in primary and is likely to remain, and it is not evident that new technical solutions will be more used than the existing ones if they do not support the health-workers work-processes to a sufficient degree at all levels. Extended use of practice-consultants can be beneficial for a better understanding of other actors' needs.

New technical solutions will facilitate new possibilities for collaboration, but many of the existing organizational barriers will still remain, and should be carefully considered when designing new technical solutions. Use of qualitative research methods can be used to get a better understanding of how future collaborative support for shared care can be designed and used. Further use of semi-structured interviews (Kvale 2007, Holstein 1995) with future users and data analysis based on grounded theory (Clarke 2005) can be beneficial.

6 References

Clarke, A.: Situational analysis, grounded theory after the postmodern turn, Sage 2005

Hanseth O., Lyytinen K., Theorizing about the Design of Information Infrastructures: Design, Theorizing about the Design of Information Infrastructures: Design Kernel Theories and Principles, Sprouts, Vol 4 2004, Article 12

Harrison, S., and Dourish, P. "Re-place-ing space: the roles of place and space in collaborative systems," ACM conference on Computer supported cooperative work, ACM Press, Boston, Massachusetts, United States, 1996, pp. 67-76

Heimly V. , Berntzen K. E., Concent-based Access to core EHR Collaborative Approaches in Norway, Methods of Information in Medicine, 2/2009

Heimly V., Standardization, innovation and deployment of electronic referral software in Norway, J Telemed Telecare 2008;14

Holstein, J. A. and Gubrium J.F: The active interview, Sage 1995

Kvale, S: Doing interviews, Sage 2007

Orlikowski, W.J, Technology and Institutions: What can Research and Information Technology and Research Organizations learn from each other?, MIS Quarterly Vol 25 No.2, June 2001, pp 145-165

Orlikowski, W.J, The Duality of Technology, Rethinking the Concept of Technology in Organizations, Organization Science, August 1992, pp. 398-427

Souza, C.R.B.d., and Redmiles, D. "The Awareness Network: To Whom Should I Display My Actions? And, Whose Actions Should I Monitor?," ECSCW'07: Proceedings of the Tenth European Conference on Computer Supported Cooperative Work, Springer, Limerick, Ireland, 2007, pp. 99-117.

Immediacy Lost: Managing Risks in Oil and Gas Production

Thomas Østerlie
Norwegian University of Science and Technology
thomas.osterlie@idi.ntnu.no

Vidar Hepsø
BI Norwegian School of Management
vidar.hepso@bi.no

Abstract. This paper presents material in support of the position that risk management in oil and gas (O&G) production is characterized by a loss of immediacy to the production process. We illustrate how loss of immediacy is an outcome of the O&G information infrastructure's mediation of the production process, and how engineers cope with the loss of immediacy when managing risk situations.

1 Introduction

This paper reports on research in progress. It presents material in support of the position that risk management in oil and gas (O&G) production is characterized by a loss of immediacy to the production process.

O&G production has been characterized as infrastructure work (Hepsø et al. 2009). Risk management is inextricably intertwined with the daily activities of this infrastructure work. Offshore installations recover O&G from subsea wells deep beneath the seabed. An extensive information infrastructure (II) collects real-time data from sensors along the production line and distributes it throughout a large network of specialized computer systems offshore as well as onshore.

Engineers in onshore production centres plan and optimize offshore O&G production. Physically removed from the offshore installations, the onshore engineers reach across this boundary with ICTs. These ICTs range from advanced expert applications visualizing real-time data from offshore O&G installations, to simple spreadsheets with historic production data. These ICTs are part of the larger O&G II, and are the same ICTs that engineers use to manage risks during operations.

As a society, we are concerned with technological accidents and the risks we shoulder through increased technological dependence (Beck 1992). Within computing research, this concern has spawned the subfield of software safety engineering. Software safety engineering is concerned with the construction of software for safety-critical systems. Safety-critical systems are systems whose failure may lead to human injury or environmental harm (Leveson 1995). Risk management in software safety engineering emphasises the need to identify risks to system safety, and to design software to detect, prevent, and recover from system failures.

Software safety engineering is based on the risk management strategy of anticipation. Anticipation is understood as "sinking resources into specific defenses against particular anticipated risks" (Wildavsky 1988, p. 220). Anticipation makes sense in stable and predictable environments where risks and their remedies can be anticipated. Yet, anticipation has clear limitations as a risk management strategy for IIs. IIs support a wide range of activities, with multiple actors continuously adapting the installed base with limited centralized control over the evolution. Unexpected events are likely to arise in such an environment and quickly propagate through the II (Hanseth et al. 2006). For safe and reliable operations in IIs anticipation needs to be complemented with other risk management strategies.

We therefore ask: How does an O&G company manage risks during operations?

Drawing upon materials collected during an ongoing study of ICT use for managing risks during O&G production, this paper emphasises how organizations mobilizes its information infrastructure to cope with unexpected events as they arise during operations. This is a shift of focus from anticipating risks during software design and planning. The object of study is therefore risk situations rather than risks themselves. Risk situations are understood as situations that hold the potential of human injury or environmental harm.

Managing risk situations requires the engineer to balance between safety and other organizational concerns such as the survival of wells and production goals – all of this in an ICT-mediated environment with a lack of immediacy to the production process. To argue the position that risk management in O&G production is characterized by a loss of immediacy to the production process, this paper is organized as follows. Section 2 presents the research setting and methods.

Section 3 offers a preliminary analysis of the collected data in support of the position argued here. Section 4 concludes the paper with a brief discussion the presented materials.

2 Research setting and methods

This paper reports from an ongoing study of ICT use in risk management during O&G production in Alpha Petroleum Company (APC). APC is a global energy company with much of its O&G operations on the Norwegian Continental Shelf (NCS). O&G production started on the NCS in the early 1970s. Many of APC's original fields have therefore entered tail-end production. Tail-end production is the last phase of an O&G field's lifecycle. Yet, new production technology, increased knowledge about the reservoirs, and better recovery methods have extended the lifetime of the fields. Tail-end production, however, is more expensive than regular O&G production. Not only does tail-end production require closer monitoring. The technical condition of the original, now aging, systems is also degrading. Increased production costs coincide with an increasingly competitive market situation.

APC has invested heavily in technological and organizational changes to meet this challenge. In addition to making better use of personnel, the transition to centralized onshore production centres is also regarded as a key to saving production costs. While there have always been onshore staff, these are now co-located with personnel that have traditionally be offshore. Co-located onshore, engineers from multiple disciplines work to together in cross-disciplinary teams. The transition to onshore production centres therefore requires closer integration and collaboration across geographical, organizational, and professional boundaries.

The purpose of our study is to develop a substantive theory on ICT use in managing risk during O&G operations. We therefore draw upon the grounded theory approach developed by Glaser and Strauss (1967) and elaborated by Strauss and Corbin (1998). Grounded theory is a *constant comparative method* for generating theory from data. With basis in coded interviews and fieldnotes, we have identified conceptual categories and their properties. The material presented in the next section is taken from our ongoing conceptualization of ICT use for managing risks during O&G production.

3 Loss of immediacy

This section presents material in support of the argument that risk management in O&G production is characterized by a loss of immediacy to the production process. This is inherent in the infrastructural work of O&G production. Since our conceptualization is still in early phases, we draw upon a concrete incidence as springboard into the conceptualization.

As reservoirs on the NCS are entering tail-end production, sand in the well flow becomes an increasing concern. O&G is recovered from reservoirs deep within geological formations beneath the seabed. Through this process the formations loose their integrity, and sand particles start mixing with the O&G flowing out of the well, the *well flow*. Sand production emerges in situations when the sand content in the well flow is high.

Sand production is a significant safety problem. The well flow moves through the piping towards the surface at high speeds. Sand increases the erosion on pipes and valves. Unchecked sand may erode through the equipment, with the potential of significant human injury and environmental harm. Sand production may also interrupt production, and is a threat to operational reliability. Sand weighs down the well flow. Too much sand makes the well flow too heavy to lift towards the surface. In these situations compact sand may fill hundreds of meters of the subsea pipeline. This stops the production, and may even mean the end of the well.

The incidence we draw upon here starts when an onshore engineer on call duty receives a call from the offshore control room. Being on call duty means that the engineer has to be available to the offshore personnel in case of unexpected events in production.

> The phone calls around two a.m. It's the offshore control room. They have an erosion alarm. Their onboard sensors register the highest erosion rate ever measured on the platform. The engineer gets out of bed and goes downstairs to the kitchen.
>
> Interviewing the engineer two days after the incident, I ask: "Are you at home when this happens?"
>
> "Yes", the engineer answers, "I am sitting at home in my kitchen. It is important that we have proper VPN access to our software applications in these situations. If I had to get in the car and drive to the operations centre in the middle of the night, I would have been inclined to tell the offshore control room to await the course of events somewhat. You know, there is a certain resistance to getting up in the middle of the night to start the working day."
>
> The engineer turns to his computer and points at an application window on the desktop: "On all flow lines, in all wells, we have a probe. Or a set of probes. Erosion probes that detect sand. When we have erosion on the probes, resistance changes."

3.1 Mediating the production process

In the above example, the O&G II takes us from wells deep beneath the seabed, along the flow line leading the O&G topside onto the offshore platform, all the way to the laptop on the engineer's kitchen table. Situated between the production process and the engineers, the O&G II forms a connection between the two. The II mediates between the non-quantified well flow and the quantified world of engineering. Mediation is not merely a matter of data circulation, but also of data transformation (Latour 1999). *Mediating is therefore the socio-technical process where a mediator transforms a source to a representation*.

A. Mediation chain

The erosion probe in the above example is only one of many sensors placed along the production line. Sensors are place along the production line from the wells deep in geological formations beneath the seabed along the pipeline onto the platform. Other sensors measure the well flow pressure, temperature, and chemical mix of the fluid pumped up from the well. The sensors transform aspects of the well flow into electrical signals.

Engineers, however, never relate to the electrical signals generated by the sensors. Rather, between the well flow and the engineers there is a chain of mediators taking the representation of the previous mediator as the source of its own transformation. This chain of mediators consists of production control systems, corporate production reporting databases, expert applications to support production planning and operations, as well as electronic document systems containing reports. Data undergoes multiple transformations along the mediation chain of the O&G II.

B. Transformations

The ICTs that engineers use are found at the end of the mediation chain, where each mediator along the chain performs different transformations. We will mainly focus on two of these transformations here: deconstruction and visualization.

By placing sensors along the production line, the well flow is deconstructed. It is deconstructed in two ways. First, sensors are localized at specific points along the production line. This reduces the continuous stream of the well flow to data points. Second, the sensors decompose the well flow by measuring a single aspect of it like temperature, pressure, sand content, as well as mix of oil, gas, and water. The well flow is the source of this transformation; sensors are the mediators that represent the well flow as electrical signals. The electrical signals represent changes in the well flow over time.

As the data passes along the O&G II's mediation chain it undergoes a series of transformations. The raw data of the sensors are timestamped and stored in production databases. To save storage space and bandwidth during onshore transmition, data is chunked by calculating average values or simply by picking a single value at regular intervals.

At the end of the mediation chain are the ICTs that engineers use in their daily work. These ICTs can be production control systems, expert applications, or even Excel spreadsheets. These systems visualize the data. Expert applications usually visualize time series data as graphs. These graphs enable engineers to see changes in a single sensor reading over time. It is also common that time series from multiple sensors may be displayed at the same time, enabling engineers to compare multiple measurements for similarities and differences. While time series is the most common form of visualization, visualizations can also be represented in form of standard reports ready to be printed on paper or gauges in the expert application.

C. Loss of immediacy

Each mediator along the mediation chain removes the data from the production process that it is to represent. Through the process of mediation, there is therefore a loss of immediacy. Immediacy is used in an epistemological sense: that of knowing the represented objects directly (Bolter and Grusin 1999). Engineers often talk about this loss of immediacy as an issue of data quality. However, rather than being a property of the data, loss of immediacy is a product of the mediation process. Loss of immediacy has multiple dimensions. We will limit our presentation to two dimensions: loss of necessary causal relations and fragmentation.

Returning to the sand production incident, we see how data may lose necessary causal relations to the represented object:

This effect (points to a spike in the graph in the application) may be caused by erosion, but it can also be caused by a change in the velocity of the well flow.

With no necessary causal relation between the data visualized in the expert application and the represented object, it is difficult for the engineer to fully determine what is happening offshore. Fragmentation is another dimension:

The focus may be on the erosion probe where you actually measure erosion, but you need to relate the data from the erosion probe to the whole system. [The probe is in] the ten inch flow line topside. Subsea you have six inch piping to the ten inch manifold. The erosion potential subsea is therefore thirty times higher than topside.

Deconstructing is necessary in order to quantify the well flow. Yet, it provides engineers with fragmented data of the well flow. In reaching across the physical boundaries separating engineers from the objects they are working with, the O&G II offers data on the well flow. Yet, for the engineers to have knowledge and understanding of the mediated phenomenon, they need to reach across the boundary between representation and the represented.

3.2 Determining situation trajectory

Let's return to the concrete incident at hand. When the onshore engineer receives the call, it is still unsure whether or not they have a risk situation at hand. The offshore control room has received an erosion alarm, but the onshore engineer is not convinced this means there is sand production. There is no necessary causal relation between the measure erosion and sand in the well flow:

> The effect may be caused by erosion, but it can also be caused by a change in the velocity of the well flow. Yet, something is usually amiss when we the data spikes like this.

When unexpected events occur, like a spike in erosion data, it is the onshore engineer's task to clarify the situation. Risk situations develop over time. When unexpected events occur, responsible engineer has to *define the temporal dimension* of the situation:

> It is important that we have quick access to data about the situation. We will then look at the data to see if this may be a false alarm, and await further developments. On the other hand, if the situation is critical, we have to act immediately.

Defining the temporal dimension of the situation is linked to *possible outcomes of the situation*: "An option could be to say 'let us await further developments until tomorrow, and keep producing oil as normal', but if there is sand in the well flow it may flow back and block the pipe and a well costing like 40 to 50 million dollars would be lost". Yet, to determine possible outcomes of the situation, the engineers need to determine what the situation at hand really is.

Through mediation the available data is fragmented and lacks necessary clear causal relations. The data is available through multiple ICTs that offer only a partial view of the situation. Downhole data from the subsea well is available through the onshore engineer's expert applications, but the offshore control room's control systems may also offer relevant data. While the data offered by multiple ICT's available to the O&G engineers lacks immediacy, the *act skilfully together* to assemble the data into a sufficiently working picture of the situation:

> We (the onshore and offshore engineers) sit down together to figure out what we need to do. You can say, they (the offshore engineers) do not analyse pressure data like we do. They look at erosion data, and measure sand in the well flow.

3.3 Risking

At some point, even though the situation is not entirely grasped, a decision has to be made:

> There is a bit of qualitative data evaluation, but then there is a matter of guesswork, too. It is simply a matter of daring to act. Taking a change. Not saying 'I have to wait until tomorrow to check with my supervisor'. Making a mistake here is expensive. Drilling a new well could cost about 40 to 60 million dollars. On the other hand: not acting is not an option either.

Such *risking* is accompanied by an *organizational non-accountability*. There is an explicit understanding within the organization that loss of immediacy often makes it difficult to fully grasp risks situations. Risking is therefore necessary.

Indeed, it is even encouraged. In a risk situation an onshore engineer reacted to what he found to be undue caution on behalf of another engineer: "What have you got to loose? Go for it, man!".

It is important for the organization to review after the fact, but individual engineers are never held directly accountable for their risking. Instead, the organization has focus on *developing routines* for handling similar situations in the future. There is focus on developing new expert applications that inscribe standard ways of combining information to deal with the loss of immediacy in known risk situations.

4 Concluding remarks

We have presented material in support of the position that risk management in O&G production is characterized by a loss of immediacy to the production process. This material illustrated how loss of immediacy is an outcome of the O&G II's mediation of the production process (Section 3.1). We then illustrated how this loss of immediacy creates uncertainty during risk situations, and how O&G engineers skilfully use ICTs to handle these situations (Section 3.2). Finally, we illustrated how the organizational practices of risking and situational non-accountability are necessary to manage risks in situations of uncertainty (Section 3.3).

Software safety engineering view safety and reliability as a systems property (Leveson 1995). Yet, the above conceptualization illustrates how safety and reliability is achieved in the interaction between engineers and ICTs. Safety and reliability is therefore better understood as the organization's ability to respond to and recover from risks that can lead to failure in performing the required activities. Safety and reliability is not a residual factor of well-designed production systems (Hepsø 2006). Rather, the above conceptualization suggests that *safety and reliability is an organizational capability*.

Amit and Schoemaker (1993, p. 35, italics in original) define capability as "a firm's capacity to deploy *Resources*, usually in combination, using organizational processes, to effect a desired end". Capability is therefore more than people acting knowledgably to achieve organizational reliability. It is also matter of mobilizing appropriate resources to achieve safety and reliability in risk situations. Data is among these resources. Safety and reliability is therefore not a system property, but rather a continuous achievement.

5 Acknowledgements

The reported research is part of the project 'Dependable ICT for the energy sector', funded by the Research Council of Norway (GOICT, RCN proj.no. 183235), and lead by Torbjørn Skramstad of NTNU/DNV. The authors thank the engineers at APC and members of the Integrated Operations in the High North consortium who have given generously of their time to advance this study. We also wish to thank Eric Monteiro for his comments.

6 References

Amit, R., and Schoemaker, P.J. "Strategic Assets and Organizational Rent," *Strategic Management Journal* (14:1), January 1993, pp. 33-46.

Beck, U. *Risk Society: Towards a New Modernity*, SAGE Publications, London, UK, 1992, p. 260.

Bolter, J.D., and Grusin, R. *Remediation: Understanding New Media*, The MIT Press, Cambridge, MA, 1999, p. 295.

Glaser, B.G., and Strauss, A. *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aidline, Chicago, Ill., 1967, p. 271.

Hanseth, O., Jacucci, E., Grisot, M., and Aanestad, M. "Reflexive Standardization: Side Effects and Complexity in Standard Making," *MIS Quarterly* (30), August 2006, pp. 563-581.

Hepsø, V. "When Are We Going to Address Organizational Robustness and Collaboration as Something Else Than a Residual Factor?", in *Proceedings of the 2006 SPE Intelligent Energy Conference*, Amsterdam, Netherlands, April 11-13, 2006.

Hepsø, V., Monteiro, E., and Rolland, K.H.R. "Ecologies of E-Infrastructures," *Journal of the AIS* (10:5), May 2009, pp. 430-446.

Hollnagel, E., Woods, D.D., and Leveson, N.G. (Eds.) *Resilience Engineering: Concepts and Precepts*. Ashgate Publishing Limited, Aldershot, UK, 2006, p. 397.

Latour, B. *Pandora's Hope*, Harvard University Press, Cambridge, MA, 1999, p. 324.

Leveson, N.G. *Safeware: System Safety and Computers*, Addison-Wesley Publishing Company, Reading, MA, 1995, p. 680.

Strauss, A., and Corbin, J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, SAGE Publications, Thousand Oaks, CA, 1998, p. 312.

Weick, K.E., Sutcliffe, K.M., and Obstfeld, D. "Organizing for High Reliability: Processes of Collective Mindfulness," *Research in Organizational Behavior* (21), 1999, pp. 81-123.

Wildavsky, A. *Searching for Safety*, Transaction Books, New Brunswick, NJ, 1988, p. 253.

Rationalising Shared Care: The Case of the Referral

Tariq Andersen[1] and Troels Mønsted[2]
[1]Dept. of Computer Science, University of Copenhagen
[2]Dept. Management Engineering, Technical University of Denmark
tariq@diku.dk, *trmo@man.dtu.dk*

Abstract. As part of shared care the referral proves to be crucial for establishing inter-organizational communication around patients. In this workshop position paper we preliminarily describe the referral by foregrounding activities and artefacts that constitutes it. When engaging in infrastructural inversion (Bowker and Star 2000) we find an inherent set of multiple interdependent actions and artefacts mobilized. These we present with an ambition to provoke a discussion on analytic issues but moreover to engage in a dialogue on how to approach a collaborative redesign of the infrastructure that constitutes and surrounds referrals. In particular, we are concerned with consequences of rationalisation i.e. standardization and automation in relation to design of such complex, interdependent, and extremely contingent collaboration.

1 Introduction

The communication and coordination involved in the referral of patient from one health care provider to another, is an important part of modern shared care. In the following we report on a case study of referrals in the treatment of ICD[4] patients in the national region of Denmark. This treatment is distributed among a large number of health care providers, such as general practitioners (GPs), home care and hospitals. In this case we draw specific attention to the communication and

---

[4] Implantable Cardioverter Defibrillator

coordination between cardiologists at satellite hospitals, who perform the preliminary diagnosis of the patients, and ICD specialists at Copenhagen University Hospitals Heart Centre, who assess referrals and implant ICD devices.

In the case study we specifically focus on the process whereby a referral of non-urgent patients are transmitted from one hospital to another, and how the referral is transformed in a sequence of activities and through use of a number of artefacts. We delimit ourselves from looking at the referral of urgent patients and patients diagnosed internally at Copenhagen University Hospital, as these cases follow different patterns.

The case study is based on a combination of observations and interviews. In total 6 observations were conducted at two referring satellite hospitals and at the Heart Centre at Copenhagen University Hospital. There were performed a total number of 8 interviews of nurses, doctors and secretaries from these organizations. The goal of the case study is to present a preliminary analysis of this empirical body and point out the direction for the future analysis and design process.

2 Framework for analysis and design

At first hand, a cardiologist' work with assessing referrals prior to admission of the patient Copenhagen University Hospital seems like a simple task. The same is apparent when the patient has arrived at the ward. What is hidden is the large amount of work that has been carried out to ease the tasks of assessing referrals and admitting patients. It is such seemingly simple activities that are afforded by the referral that we wish to examine through our case study. We wish to foreground all the background work of rendering the referral part of the infrastructure supporting the distributed communication.

Inspired by Bowker and Star's (2000) methodological trick of 'infrastructural inversion' we approach the phenomena of the referral as a compound artefact but also as a set of activities that together become infrastructure for working doctors and nurses. When employing infrastructural inversion the complexity of the infrastructure becomes visible. That is, large amounts of collaborative work activities, various artefacts, technologies, software, paper documents, terminologies, and standards etc. become centre stage for our case study. Bowker and Star formulate this powerful methodological trick as *"[..] to question every apparently unnatural easiness in the world around us and look for the work involved in making it easy."* (ibid. pp. 39).

With an extended analysis framed by infrastructural inversion, we are moreover interested in redesigning the "referral". As part of the CITH-project[5] we

[5] CITH is short for "Co-Constructing IT and Healthcare". More information about the research project can be found on www.cith.dk

ambitiously wish to combine analytical insights with attempts to intervene and do actual design work on the referral. At present it is our plan to carry out collaborative design activities that will change the current infrastructure of the referral. With this interest we enter a discussion on how to take on design activities i.e. how should we continue our analysis and how can it inform the design process?

In relation to design we particularly wish to investigate the well known challenge of rationalising medical work (Berg 1997, Berg and Timmermans 2003). By rationalisation we mean formalising and standardising but also, and quite importantly to us, automating. The design process will therefore become engine for exploring both methods and techniques but also enter the growing interest in how rationalisation should be dealt with when designing for integrated care. Relevant for the CSCW community is that we wish to add to e.g. Winthereik and Vikkelsø's (2005), and thereby enter a discussion on the dilemmas of standardising inter-organisational healthcare communication. In the following sections we present the preliminary analysis of the case study.

3 The infrastructure of referrals

In the following, we focus our description of the tasks performed at four locations: 1) The satellite hospital. 2) The Visitation at the Heart Centre[6]. 3) The Bed Ward at Copenhagen University Hospital. 4) The operating room.

3.1 Activities at the satellite hospital

There are several satellite hospitals and specialised general GP who refer patients for ICD implantation. Patients get referred to ICD implantation because of multiple reasons (that we do not go into here). In the following, we simplify this process by describing the creation of a referral at hospital X on the basis of a somehow general trajectory. This work involves the following sub activities.

A) Medical consultation and dictation of referral (doctor)

As a consequence of a heart patient's health examinations and medical treatment the doctor can decide to refer the patient for ICD implantation. The decision is based on the doctor's training and knowledge about heart diseases and ICDs and is carried out by him/her dictating onto a cassette tape (part of the hospital paper record), reasons and indicators for referring. Part of this activity involves informing the patient about ICD implantation.

B) The referral is prepared and sent

---

[6] We refer to the office where assessment of referrals are performed as the "Visitation"

When the secretary receives the doctors' oral dictation, she transcribes it and starts creating a referral. This involves making the front page in a blank word document, finding the address for the Heart Centre and copy-pasting transcribed text onto it. The secretary also retrieves the prescribed documentation, such as copies of health examinations and electrocardiograms. The secretary can use three means of communication to send the referral to Copenhagen University Hospital: letter, fax machine, a computer system or a combination[7]. By performing these actions, the secretary transforms a selection of documents into what is now a referral.

3.2 Activities at the Visitation at the Heart Centre

The overall task of the Visitation is to assess the referrals from the satellite hospitals prior to the admission of the patient on the Heart Centre. This work involves three roles; 1) Nurse X. 2) Nurse Y. 3) Cardiologist. The work can be divided into the following sub activities.

A) The referral is prepared for assessment (nurse X and nurse Y)
At the visitation all incoming referrals are handled as paper documents. This implies that referrals received through the computer system is printed, and thereby converted to paper format. When a new referral arrives, nurse X puts a stamp on the front page. This stamp contains a number of fields that will later be filled out. Nurse X writes the current date in one of the fields.[8] She then open a record in the computer system "GS!ÅBEN" and enters the master data of the patient.

The referral is then handed over to nurse Y, who runs through the attached documentation, e.g. description of the diagnosis and electrocardiograms. By experience, nurse Y knows what the assessing cardiologist requires from this documentation. Nurse Y acquires any missing documentation by contacting the secretary on the referring hospital, most often by telephone.

The referral is put on hold until the new documentation has arrived. The referral is then handed over to the cardiologist by placing it in one of four letter trays. Referrals concerning ICD patients are categorised as "Electrical". The main task for the nurses at this stage is to prepare referrals to be effectively assessed by the cardiologist. In an infrastructural perspective, the nurses and the artefacts in use, therefore becomes a part of the underlying infrastructure that ease the work of the cardiologist.

---

[7] As the computer system cannot handle attachment such as electrocardiograms, referrals send by computer must always be supplemented by either a letter or a fax.

[8] Later, this stamp will be referred to as "the assessment stamp"

B) The referral is assessed (cardiologist)
A cardiologist visits the visitation office every day to assess new referrals. The cardiologist takes the pile of referrals from the letter tray, and carefully reads the documentation for each referral. She either approve the referral, reject the referral or in cases of doubt, requests for further medical examinations of the patient. The cardiologist writes the verdict in a field in the assessment stamp at the front page of the referral, at places it in the same letter tray as it was taken from. Due to the amount of documentation, this work can be time consuming, although the task is made significantly easier by the nurses' preparation.

C) The time for operation is booked
When the referrals have been assessed, nurse X reads what is annotated in the assessment stamp, and takes action as prescribed. If the referral is approved, she uses the computer system "ORBIT" to book a time for the implantation, writes the dates on the front page of the referral, and places it in the corresponding letter tray.

D) Nurse Y writes a letter of notification to the patient
When the referred patient has been appointed a time for implantation and admission, nurse Y picks up the referral from the letter tray, and uses the computer system GS!ÅBEN to write a notification to the patient. The letters are printed and send to the patient by mail. She then writes information on the appointment in a paper calendar placed next to the computer. This calendar acts as a backup system to the computer. Finally, nurse Y sorts all new referrals by the date of the admission, and places them in a ring binder.

E) The referral is brought to the bed ward
One week before the operation, the nurse X removes the referrals from the ring binder and carries them the bed ward where she places them in a letter tray.

3.3 Activities at the ward at the Heart Centre

The majority of activities surrounding the referral at the bed ward are carried out by a secretary and her assistant. This is mainly work on preparing or shaping the referral for becoming front page of the hospital paper record, i.e. ready for doctors' and nurses' work with the admission patients. Referrals are also input to the secretary's management of the ward's 30 beds. Below, we sequentially describe the activities and the artefacts involved.

A) Managing bed occupancy (secretary)
The arrival of referrals from the visitation triggers several activities at the bed ward. Most importantly the secretary uses the referrals in the management of the ward's 30 beds. She uses the information to allocate beds to admitted patients, and the referrals inform her on e.g. how many patients will arrive the following week,

which day, and indicators of length of stay. Going through the pile of referrals she adds each patient (barcode labels and annotation) to the overview of bed occupancy of each specific weekday. She incrementally builds up A4 paper sheets for every weekday for creating an overview of patient admission and bed occupancy. She refers to them as “the brain”.

In this case, each referral feeds into creating overviews of patient arrivals that in the course of a day are vital parts of the infrastructure, both for the secretary, nurses and doctors at the ward.

B) The referral is reshaped and becomes front page of the hospital paper record (secretary and secretary assistant)

While creating the sheets of bed occupancy, the secretary also examines each referral for the assessment stamp and checks in GS!ÅBEN whether the patient have an existing hospital paper record. If not she needs to create one. If the patient is readmitted and a record already exists, the secretary acquires the hospital paper record by creating a collection of GS!ÅBEN "print-screen"-printouts. This makes it possible for the assisting secretary to dispatch the journals from different locations at the hospital. If the patient is new to the hospital she creates a new hospital paper record using data from the referral.

For all referrals she fills out a local bed ward referral form. The annotations are results from e.g. bookings of various examinations including blood testing. She makes a copy and places it in the nurses’ chart in the hospital paper record and the original becomes front page of the hospital paper record. Finishing the preparation of a referral is its placement in the day pigeonhole according to the weekday that the patient arrives.

Jointly these activities result in reshaped referrals augmented by annotated bed ward referral forms. They are now front pages of patients’ hospital paper records placed according to the weekday of the patient admission. It is this new shape and the placement in the pigeonhole that renders the reshaped referral an indispensable part of the infrastructure when nurses and doctors prepare for patient admission.

C) Receiving patients using the referral (doctors and nurses)

At this point in time referrals are ready to enter nurses’ and doctors’ work of admitting patients. However, the responsibility of each arriving patient needs to be delegated to nurses. This is carried out by three nurses working the night shift. What they do is picking up the referrals (that are now front pages of the hospital paper record) stacked for the coming day (in the weekday shelve system) and delegate by writing the patient name etc. together with the responsible nurse’s name on a whiteboard in the nurses’ office.

The following morning, nurses who are responsible of receiving patients start by checking the whiteboard and then pick up the corresponding hospital paper record (the nurses chart herein) in the secretary’s office weekday shelves. They browse the referral and the record to get acquainted with the patient beforehand, understanding reasons for admission, length of stay and other important indicators from the patient's trajectory. The same is the case for doctors when they visit the patient the first time. They use it together with a dialogue with the patient to write an admission note.

Again, the referral has shown to be important piece of the infrastructure supporting nurses' and doctors' work. Moreover, information from the referral gets transformed further onto a whiteboard and embedded into doctors' admission notes. After admitting the patient the referral ends up as an enclosed document in the hospital paper record.

3.4 Activities at the operating room at the Heart Centre

The referral ends its trajectory when it arrives at the operation room as enclosed in the hospital paper record. By now it has moved into being a part of the hospital paper record alongside other enclosed documents.

4 Approaching Design

The process of referring patients from one hospital to another is seemingly an easy task for the doctors involved, although this case study reveals a number of severe problems. Although not emphasised in our description above, the most important finding is that the work of assessing referrals at the visitation is permeated by exceptions; more often than not, nurse Y have to acquire supplementary documentation from the referring hospital, which increases her work load and delays the treatment of patients.

The main reason for these exceptions is found to be poor quality of the referrals from the satellite hospitals; often, relevant documents are missing, and often, medical examinations are of such poor quality, that they must be repeated. In relation to design an obvious solution is to rationalise this process, for instance by designing a referral system in a way that insures that and adequate amount of documentation is attached to the referral. Also it is obvious to automate a number of the many sequential actions and thereby reduce the average time it takes a referral to pass assessment.

Before entering the design phase with the goal of rationalising and automating the referral process, it is necessary with thorough consideration on the possible downsides of this approach in the context of shared care.

5 References

Berg, Marc (1997): Rationalizing medical work: decision-support techniques and medical practices Cambridge, Mass.: MIT Press.

Bowker, Geoffrey C. and Susan Leigh Star (1999): Sorting things out: classification and its consequences. Cambridge, Mass.: MIT Press.

Timmermans, Stefan and Marc Berg (2003): The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care. Philadelphia, Temple University Press.

Winthereik, Brit Ross and Signe Vikkelsø (2005): ICT and Integrated Care: Some Dilemmas of Standardising Inter-Organisational Communication. Computer Supported Cooperative Work 14: 43–67, Springer.