Journal of Documentation

# Using the domain analytical approach in the study of information practices in biomedicine

Annikki Roos Turid Hedlund

JDOC
72,5



Received 14 November 2015
Revised 13 March 2016
Accepted 3 April 2016

# Using the domain analytical approach in the study of information practices in biomedicine

Annikki Roos
*Hanken School of Economics, University of Helsinki, Helsinki, Finland, and*
Turid Hedlund
*Hanken School of Economics, Helsinki, Finland*

**Abstract**

**Purpose** – The purpose of this paper is to analyze the information practices of the researchers in biomedicine using the domain analytical approach.

**Design/methodology/approach** – The domain analytical research approach used in the study of the scientific domain of biomedicine leads to studies into the organization of sciences. By using Whitley's dimensions of "mutual dependence" and "task uncertainty" in scientific work as a starting point the authors were able to reanalyze previously collected data. By opening up these concepts in the biomedical research work context, the authors analyzed the distinguishing features of the biomedical domain and the way these features affected researchers' information practices.

**Findings** – Several indicators representing "task uncertainty" and "mutual dependence" in the scientific domain of biomedicine were identified. This study supports the view that in biomedicine the task uncertainty is low and researchers are mutually highly dependent on each other. Hard competition seems to be one feature, which is behind the explosion of the data and publications in this domain. This fact, on its part is directly related to the ways information is searched, followed, used and produced. The need for new easy to use services or tools for searching and following information in so called "hot" topics came apparent.

**Originality/value** – The study highlights new information about information practices in the biomedical domain. Whitley's theory enabled a thorough analysis of the cultural and social nature of the biomedical domain and it proved to be useful in the examination of researchers' information practices.

**Keywords** Information services, Information science, Research work, Information practices, Biomedicine, Domain analytical approach

**Paper type** Research paper

## 1. Introduction

Within the field of information science (IS) research into domains of scientific and scholarly disciplines has a long tradition. Hjørland and Albrechtsen (1995) introduced a new terminology "domain analysis" and argue that the best way to understand information is to study the domains where knowledge is presented and that knowledge domains should be studied as thought and discourse communities. The domain analytical approach, according to Hjørland and Albrechtsen (1995), suggests a social perspective that individuals should be seen as a member of a discourse community, a research or working group or a discipline.

Palmer and Cragin (2009) point out that in recent decades research on disciplinary practice has been growing in the social sciences in general. In IS, the number of

The authors would like to acknowledge the suggestions and comments that the authors received from the anonymous reviewers.



Journal of Documentation
Vol. 72 No. 5, 2016
pp. 896-915

0022-0418
DOI 10.1108/JD-11-2015-0139

disciplinary-oriented studies has also increased. However, numerous existing studies (see e.g. Nicholas *et al.*, 2006), which aim to look at access and use of information resources in scientific domains, consider the use of a resource as an isolated entity. According to Talja (2005) it is well known that there are field differences in researchers' information practices but a comprehensive understanding about factors that determine the differences is missing. Domain analysis as an approach to understand scientific or scholarly research and the information practices in a discipline or research field or even as small a unit as a research group is not widely used in IS research. Advocates of the domain analytical approach are Jenny Fry (2006) and Sanna Talja (2005). Both Fry and Talja see the domain analytical approach as a mean to understand and explain the field differences in scientific and scholarly work and the information practices and use.

According to Whitley (2000) scientific disciplines may be understood in similar ways as work organizations. However, compared with other work organizations, differences exist. The outcomes of the work tasks are more uncertain because a key feature in research work is the commitment to innovative and new knowledge. This uncertainty leads to a particular structure for organizing and controlling research. Mainly the novelty of research results is rewarded. The scientific community and researchers in the same field in co-operation control and evaluate the novelty and contribution of the research work. Whitley (2000) explains that there exists variation between scientific fields in how the work is controlled and organized. He presents the differences between scientific fields by two factors: the "degree of mutual dependence" between researchers and the "degree of task uncertainty." These factors will be explained in more detail in the fourth chapter.

Fry and Talja (2007) join forces in an article contributing to the theoretical understanding of differences in scientific disciplines and how the differences affect the shaping of for example net-based resources. The findings of their study support Whitley's theory. In environmental biology and high-energy physics, where the task uncertainty is low and mutual dependence high, scientists "know where their research is located in relation to the rest of the field and how to make a contribution to it. This means that they have no need to use unrestricted topic-oriented lists as an aid to scanning the field and assessing where their research fits." (Fry and Talja, 2007, p. 123) In the fields, where mutual dependence is lower and task uncertainty higher, like in social/cultural geography or literature and cultural studies, researchers need to use various informal channels, for example mailing lists, in order to be able to locate themselves within the existing discussion in their field.

In this study, we will apply the domain analytical perspective when studying information practices in the field of biomedicine. According to the dimensions used by Whitley (2000) biomedicine resembles environmental biology as a field with high mutual dependence in research communication and controlled research and publishing methods (Hedlund and Roos, 2007).

We understand information practices as a set of socially and culturally established activities to seek, use and share information and data available in different sources. As noticed by Savolainen (2007, p. 110), the concept of information practices is not neutral but is based on the presumption that information-related activities are constituted socially and dialogically rather than based on individual motives, needs or ideas as is the case in the cognitiv research approach in IS. We consider that among one of the most important factors to consider in understanding information practices, is the social dimension of scholars' domain or discipline (Palmer and Cragin, 2009). The historical, collective and material aspects of the context of the practice, in our study, the scientific domain, have effects and modify the information practice itself (Rivera and Cox, 2014).

Practices are not valuable as such, but become significant as tools that help in accomplishing the main objectives of activities (Savolainen, 2007, 2008). In the case of research work in biomedicine, the main objectives of the research work could be to find answers to research questions and to create new knowledge. The final outcome of the research work is to prevent or cure diseases (Roos, 2012).

Hjørland (2005) finds that the domain analytic approach emphasizes domains instead of individuals as a unit of analysis. However, it also recognizes the individual cognitive information-related processes and how they interrelate with others in the same domain. The adaption and learning process of the individual in relation to others in the same domain is described as what is relevant information in a certain research domain. This affects the information practices and how they are shaped.

In this study, the level of analysis is the research group in biomedicine. However, we recognize that groups consist of individuals, and that the information practices of groups are shaped and developed by individual researchers resembling the socio-cognitive approach as described by Bates (2005).

The domain analytical perspective allows us to concentrate on the specific features of the practices applied in the field of biomedicine. Our aim is to initially open up the concepts of mutual dependence and task uncertainty and study how they might be interpreted in the field of biomedicine. Our research approach is abductive in the sense that we start from the general theory described by Whitley in analyzing our data and in the results we are able to categorize the research practices in the field of biomedicine.

With this research we aim to acquire a better understanding of the domain specific features of the information practices in biomedicine. This, we believe, is the only way to be able to produce information services that support biomedical researchers at optimal manner in their work.

The research questions we ask are:

*RQ1.* How can we describe and interpret the concepts "mutual dependence and task uncertainty" in the domain of biomedicine?

*RQ2.* In which ways does the domain affect information practices in biomedicine?

Next we describe the biomedical domain and the research settings, as well as the analyses of the characteristics of the biomedical domain using Whitley's theoretical frame. We will continue with the analyses of information practices in biomedicine and end with discussion and conclusions.

## 2. The domain of biomedicine

Biomedicine is comprehended as medicine or a branch of medical science, which is based on the application of the principles of natural sciences, especially biology and biochemistry to clinical practice (www.merriam-webster.com/dictionary/biomedicine; www.memidex.com/biomedicine). Biomedicine has developed profoundly on the tradition of empiricism and positivism. As a result of this, the experimental method, basing on hypothesis testing is valued and used in biomedical research. This does not mean that, in particularly in the studies of medical practice, the critic of the nomothethic methodology would be excluded (see e.g. Hjørland, 2011; Wilson, 2000). Research topics in this domain are broad and varied, covering areas from basic cellular biology to the treatment and medication of complex diseases and health services. Research in biomedicine can be divided roughly into three broad and partly overlapping categories: basic science research, clinical research and translational research (comp. Henderson and Bunton, 2013). Basic research constitute research, for example, from scientific domains like biochemistry, microbiology and pharmacology. Laboratory research is included in it almost by rule. The focus in biomedicine is mainly on the underlying mechanisms behind diseases and health. Clinical research in biomedicine is more applied and patient and disease oriented, aiming more directly at the prevention and treatment of diseases and health promotion. Clinical research can be divided into sub-fields according to numerous clinical specialties, for example, pediatrics, geriatrics and oncology. Translational, or bench-to-bedside research is a rather new research category in biomedicine. It aims to disseminate the outcomes from the basic research to clinical research. The purpose is to assist in the practical application of the results of the basic research in the clinical research and in this way to enable the development of new therapies and medical procedures (e.g. Henderson and Bunton, 2013). An example of the outcome of translational medicine is the launching of the Cancer Genome Atlas (http://tcga-data.nci.nih.gov/tcga/) where clinical data and research data have been integrated. The rough and partly overlapping categorization of the biomedical domain is presented in Figure 1.

Molecular medicine (MM) is an interdisciplinary, the practice oriented and applied subfield of biomedicine (Roos *et al.*, 2008) mostly belonging to the basic science research area. It means "the field of medicine concerned with understanding the biochemical basis of health and disease involved in developing diagnostic and therapeutic methods that utilize molecular biology techniques (www.ncbi.nlm.nih.gov/mesh/?term=molecular+medicine)." MM consists of biomedical and molecular biological research. Molecular biology for its part is a combination of biochemistry, genetics, cell biology and virology (Roos *et al.*, 2008).

Trent (2012) uses the term MM extensively to describe the effect that knowledge of DNA and RNA are having on medical practice. In MM, the biological processes and the mechanisms of the human diseases are studied using molecular and genetic techniques. Research focusses partly in single gene Mendelian disorders and for the most part in complex genetic disorders. The former are quite rare in population but the complex diseases, like cancer, heart diseases and diabetes comprise a real challenge to public health. Understanding the molecular mechanism of a complex disease is also a

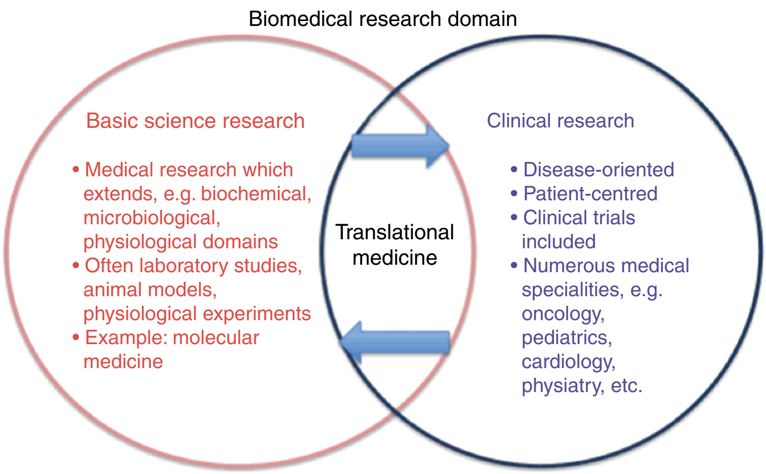


**Figure 1.**
The biomedical research domain

complicated research question causing challenges to the research work. "Omics," like genomics, developed in the twentieth century after Human Genome Project, is the answer to the questions rising from the study of complex disorders. In Omics, all or many molecules within an organism, a tissue or a cell are characterized (Trent, 2012).

The research results in biomedicine are published dominantly in scientific journals. The amount of published articles has been growing at a double-exponential phase during the recent decades (Hunter and Cohen, 2006; Lu, 2011). In 2015, the number of references in PubMed, the most important searching tool of published literature in biomedicine, is over 24 million (www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.FAQs). Researchers in biomedicine and particularly in MM produce and use huge amounts of research data. The amount of data is growing exponentially. An example of this is the growth rate of the GenBank, a comprehensive public database of nucleotide sequences. The size of the database has doubled since 1982 about every 18th month (www.ncbi.nlm.nih.gov/genbank/statistics/). The amount of tools and services, which help in the processing and use of data is considerable (Roos *et al.*, 2008).

## 3. Methods and research settings

In this study, we have reanalyzed data, which was collected during the years 2007 and 2012 (earlier results reported by Roos *et al.*, 2008; Roos, 2012, 2015). In the 2008 article, the analysis focussed on the quantitative part of the data, while in the 2012 and 2015 articles the activity theoretical research frame was used in the analyses and based on data collected from interviews and observations. In this study, the entire data set is approached using Whitley's theoretical frame to explore the domain specific characteristics in the field which in turn form the information practices.

In the 2008 article, the data were collected from two MM research units, located in Helsinki (Unit A) and Tampere (Unit B), in Finland. From these units, we gathered basic information by two separate surveys, both conducted in 2007. We got totally 116 answers, 63 from Unit A, and 53 from Unit B. The majority (66, 56.9 percent) of the answers come from doctoral students. We received 50 answers (43.1 percent) from senior and post-doctoral researchers. After the survey six semi-structured interviews were made to complement the data. The survey form is included as an attachment (Attachment 1) in this paper. Detailed information about the surveys is presented in the Table I.

Later, in the article, which was published in 2012, the survey data from unit A, was supplemented with additional semi-structured interviews. For practical reasons, the members of the unit B were not applicable for the interviews and were not included in this part of the research. The director of the unit, six group leaders and two doctoral students were interviewed. In addition, one of the authors occasionally shared the office with one of the research groups during the autumn 2007, and took part in the unit and

| Survey data | Unit A | Unit B | Total |
|---|---|---|---|
| No. of juniors | 63 | 50 | 113 |
| No. of seniors | 28 | 53 | 81 |
| Target audience totally | 91 | 103 | 194 |
| *Answers received* | | | |
| From juniors | 41 | 25 | 66 |
| From seniors | 22 | 28 | 50 |
| Answers received totally | 63 | 53 | 116 |

**Table I.**
Information about the surveys

group meetings as well as some other activities in the unit. As noticed in the Attachment A, we collected data about the publishing plans of the researchers in 2015, PubMed was searched and the actual execution of the plans of researchers was checked. Those articles that the group leaders or PIs in the Unit A had published between 2007 and 2010, were searched from the database.

In the paper, published in 2015, data were collected by interviewing researchers and research groups from Helsinki University Central Hospital and the University of Helsinki. The majority of the researchers were medical scientists, which means that they were doing biomedical research in clinical settings (i.e. clinical research). We gathered data by interviewing totally 12 researchers. They represented five different biomedical subfields: obstetrics and gynecology, otorhinolaryngology, pediatrics and public health. Here, we include only data collected from those ten researchers who were doing clinical research. Three of these medical scientists were seniors and seven were doctoral students. An overview of the data sets is presented in the following table (Table II).

Information about the interviews conducted at various stages of this study is collected in Table III.

## 4. Characteristics of the biomedical scientific domain according to the framework presented by Whitley

In the beginning of this chapter, we will present more closely two dimensions, "mutual dependence" and "task uncertainty," which according to Whitley (2000) explain differences between scientific fields. After that follows a more thorough analyzes of these dimensions in the biomedical domain.

The dimension of mutual dependence according to Whitley (2000) is related to the extent that a field is dependent on knowledge produced in order to make a contribution to science. There are two analytically distinct aspects to mutual dependence: the "degree of functional dependence" and the "degree of strategic dependence" (Whitley, 2000),

| Date | Type | Field/domain Research subject |
|---|---|---|
| Year 2007 | Survey data total 116 answers | Molecular medicine Unit A and B |
| Year 2007 | Interview data total 6 interviews | Molecular medicine Unit A and B |
| Year 2007 | Interview data 9 interviews | Molecular medicine Unit A |
| Year 2012 | Interview data 12 interviews | Biomedical research in clinical settings |

**Table II.** Overview of the data sets

| Year | No. of interviewed researchers totally | Of which in Unit A | Of which in Unit B | Of which in Helsinki University and HU Hospital |
|---|---|---|---|---|
| 2007 | 6 | 4 | 2 | |
| 2007 | 9 | 9 | | |
| 2012 | 12 | | | 12 |
| Totally | 27 | 13 | 2 | 12 |

**Table III.** Information about the interviews where the degree of the functional dependence refers to the extent to which researchers need to use specific results, procedures and ideas to be able to make competent contributions to the field. The degree of strategic dependence is related to the amount that researchers have to convince the scientific field of the significance of the chosen research problem and research approach (Whitley, 2000). These dimensions might, for example, influence the extent that the researchers need to use specific research procedures and earlier results in order to convince their scientific field of the legitimacy of research results.

The dimension of task uncertainty (Whitley, 2000) refers to the degree of predictability and visibility of the outcomes and research processes. This dimension is associated with the patterns of work organization and control. Whitley argues that there are two aspects in task uncertainty: technical and strategic. Technical task uncertainty is related to the visibility, stability and uniformity of task outcomes. Strategic task uncertainty refers to the similarity, stability and integration of research goals and strategies (Whitley, 2000). When task uncertainty is low, a consensus about the technical procedures and goals and priorities of research exists. In most cases, Whitley emphasis, both aspects of the high degree of mutual dependence and the task uncertainty are unlikely to occur with a very low aspect of that dimension's other aspect. We concur with Fry and Talja (2007) that mutual dependence and task uncertainty are best conceived as an increasing and decreasing continuum.

Our initial starting point to the domain analysis was the fact that biomedicine appears to clearly possess the characteristics of a scientific field with a low level of task uncertainty and a high level of mutual dependence as in the framework by Whitley (2000), Hedlund and Roos (2007). The aim of our exercise is to explore practical exemplifications of how the framework can be utilized and represented when studying how and why information practices have developed in the field of biomedicine.

In Table IV the two dimensions task uncertainty and mutual dependence are presented in the left column according to the characterization of Whitley. In the right column are listed characteristics found in our data set on biomedicine associated with these two dimensions. In biomedicine, low task uncertainty is represented by a common goal for research, strong emphasis on group work, a clear division of labor and a clear leadership and supervision. While again mutual dependence is represented by a preference for sharing resources, a strong dependence on technology, a collaborative approach to research within one's own unit but also a favorable attitude to national

**Table IV.**
The dimensions of task uncertainty and mutual dependence in biomedicine

| | |
|---|---|
| Task uncertainty (Whitley)<br>Characterized by: the degree of predictability and visibility of the outcomes and research processes. This dimension is associated with patterns of work organization and control | Low task uncertainty (biomedicine)<br>Represented by:<br>a common goal for research<br>strong emphasis on research groups<br>clear division of labor<br>clear leadership and supervision |
| Mutual dependence (Whitley)<br>Characterized by the extent that a field is dependent on knowledge produced in order to make a contribution to science. This dimension might for example influence the extent that the researchers need to use specific research procedures and earlier results in order to convince their scientific field of the legitimacy of research results | High mutual dependence (biomedicine)<br>Represented by:<br>sharing of resources<br>dependence on new technology to produce competitive research<br>collaboration within the research unit as well as national and international collaboration<br>strong competition in the research field |

and international collaboration. Above all biomedicine is characterized by strong competition in the research field.

In the following, we present a closer analysis of the domain specific features in biomedicine, which indicate a low task uncertainty and high mutual dependence.

### *4.1 Features indicating low task uncertainty*

*4.1.1 Common goal.* Despite the broad variation of research topics, the final objective of biomedical research is clearly quite common: to prevent and make interventions to human diseases by combining the efforts of the physical, chemical, biological and medical sciences. In MM, there are usually two basic research questions: which gene/genes cause a disease or disorder and what is the molecular mechanism behind the disease. Developments in information technology and molecular biology are the main factors, which have generated the basis for the expansion of the MM. Clinical research is concentrating more to the application of the results of the basic biomedical research to the prevention and treatment of diseases.

*4.1.2 Research groups*. The research environment in these cases is characterized by group formation. Research groups form the smallest entity in the Molecular Medicine Research Institute. All researchers work in groups, which are usually a part of a bigger research program or a unit. The group is lead by the principal investigator (PI), other members usually including post-doctoral researcher/s, doctoral student/s, often also some graduate students and laboratory personnel. All of the medical scientists belonged also to a research group. Compositions of the groups were quite similar to MM. Groups could have members in other universities or institutes. However, the position of the researchers differed. In Molecular Medicine Research Institute research work was the main duty of the scientists. A part of those medical scientists who were PhD students were participating in the clinical work and were doing part-time research. Only 50 percent of the medical scientists had research work as their main responsibility. The PIs in the clinical research groups were sharing their time mainly between clinical work, education, research and administrative work. In MM, the PIs were able to concentrate more to the research and planning of research work along with administrative duties.

*4.1.3 Division of labor*. Research students are extremely important to the research work in both of our cases. They carry out the laboratory experiments, collect the results, do the analysis and report the results. In MM, the supervision of the work of the research students is mainly the duty of the group leaders but many senior researchers also have the responsibility. In clinical research, the role of the senior medical scientists seemed to be quite important. The reason for this perhaps was that the PIs in our cases appeared to be extremely busy. The director of the research unit of MM supervises directly few students' work but is actually involved partly in all of the youngest researchers' work. Among the most important and time consuming duties to the PI in the MM research unit, was to seek funding for the group.

*4.1.4 Leadership and supervision*. The role of the PI is to direct the research work. In MM, she/he is the source of the new research topics and ideas as well as the decisions about the laboratory and research methods. Senior researchers in the group or colleagues from other groups or organizations are important partners to the PI in this stage of the research. The role of the senior medical scientists in clinical research was perhaps more important in comparison with their colleagues in the MM research unit. Previous research is examined before the hypothesis is ready to be tested. PIs usually design the work and doctoral students do it in the laboratory.

Junior researchers usually do the first analyses of the results and will discuss them with a senior or the PI. A new test on the laboratory might be needed before the results are ready to be reported.

*4.2 Features indicating high mutual dependence*

*4.2.1 Sharing of resources*. Commonly shared, globally used and produced information and data resources are typical to the biomedical domain. There exist websites or portals, which integrate various databases, tools and literature. Examples of these are Entrez, a federated search engine or web portal produced by National Centre for Biotechnology Information (NCBI), and collections of services, which is produced by European Molecular Biology Laboratory/The European Bioinformatics Institute. Gene Ontology Project (http://geneontology.org/) is an example of a common initiative, which has been organized by the Gene Ontology Consortium (more about Gene Ontology see e.g. Mayor and Robinson, 2014). The consortium consists of collaborative groups all over the world. NCBI is behind PubMed, the dominating search tool for biomedical literature. The US federal government has a long history of supporting the development of PubMed and the database (MEDLINE) behind it (Weiner, 2009). NCBI is a part of the United States National Library of Medicine (NLM), which on its' part belong to the National Institute of Health (NIH).

*4.2.2 Competitive environment*. International competition in biomedicine is hard. Research groups act in a global environment where the research topics are rather commonly shared. In MM, the competition is extremely hard when, for example, DNA sequencing laboratories all over the world are in the running to the similar score. In addition they compete on funding from the same sources. The funders evaluate the performance of individual researchers, research groups and organizations by using various metrics. The most common tool for evaluating achievements is to use citation analyses. Medical faculties use bibliometrics when comparing and selecting individual researchers to open positions in the faculty. The competition between researchers in our cases seemed to be quite open. This was demonstrated by the use of a tool developed by the library. Scholar Chart makes an open, constantly updating ranking list of all biomedical researchers in Finland (www.terkko.helsinki.fi/scholarchart/).

*4.2.3 Dependence on technology*. Biomedicine is dependent on expensive equipment and technologies. (Massoud and Gambhir, 2007; Roos, 2012; Trent, 2012). As noticed earlier, it is data intensive, producing and processing a huge amount of data. The development of DNA-sequencing technology, for example, has improved dramatically during last decade and caused the sequencing costs to decrease at an exponential rate (Stein, 2010). However, data storage appears as a growing problem (Pennisi, 2011).

*4.2.4 Collaboration*. Co-authorship is the most typical form of collaboration among researchers. The network of research collaboration is broad ranging from local and national to international cooperation. As Newman (2001) notices, it is typical to biomedical research that it is organized under laboratories where the PI is supervising many post-doctoral researchers, doctoral and other students who work in different projects. This was the situation also in our cases. In addition, our research groups joint forces with researchers in various local organizations, Scandinavian and European scientists, and they had partners in the USA and Canada. Genome-wide analyses are an example of the studies where a broad international collaboration is a necessity. Many of the international relationships were based on the personal contacts of the senior researchers.

## 5. Information practices in biomedicine

In the following analysis of the information practices in biomedicine we will use as a framework the domain specific features described above. We will analyze information practices in relation to the indicators that are signaling about the low task uncertainty and high mutual dependence.

### *5.1 Indicators of low task uncertainty*

*5.1.1 Common goal for research.* As noticed above, biomedical research on the highest level aims to achieve a common goal: prevent and cure diseases. One indication of the common goal could be the nature of the publishing platforms. The results of the research work in this domain are reported as a rule in scientific journals. Articles are usually short, between 2,000 and 4,000 words. Language of the articles is precise and controlled compared, for example, with social sciences and humanities. Because of this, it is quick and easy to identify the significance and the novelty of the new research outcomes, which is an indication of the low task uncertainty. (comp. Fry and Talja, 2007). There are also articles, which publish the results of studies conducted by a big International Consortium. In this way, the co-ordination within the field becomes more certain, which in turn enables the division of the field into separate still interdependent specialisms (Whitley, 2000, pp. 130-131). In biomedicine, the consequence has been the emergence of various journals for every biomedical specialty.

*5.1.2 The research group as the standard form of research work.* Biomedical researchers work in groups. The members of the research groups have different roles in the research work and information practices. Our cases showed that the PI or the senior researcher was an important source of all kinds of information for the PhD students. They advised younger researchers with literature and methods as well as followed the progress of the work. Almost all (92 percent) researchers in MM and all medical scientists used PubMed as their main literature search engine. When information was searched from PubMed, researchers did it by themselves. Most of them had a profile for current awareness service in PubMed. We noticed that in MM it was typical that expertise was searched and used outside the group but inside the unit, from other groups of the institute. This was the case especially bioinformatics or statistics. Researchers in general seemed to share information about the information and data related tools. Particularly group leaders in MM gave guidance and help with the use of data resources. Colleagues from other groups acted as an information source especially for the selection of the research/laboratory methods. The searching of literature was done quite alone as well as reading and writing, even though the connection between the members of the group was quite active during the process. Researchers, who work physically close to each other, even though in a different group, seem to have a significant role.

*5.1.3 Division of labor.* As already mentioned, labor in biomedical research work is quite clearly divided. The doctoral students usually have the main responsibility of the writing process. The PI supervises the work of the doctoral and graduate students and is usually a co-writer in the articles. The amount that PIs participate may vary depending on the stage of the studies of the doctoral student: in the beginning perhaps more, later less. It was typical in our cases that the PI provided the most important literature for the doctoral student in the beginning. Later, perhaps after the first article, doctoral students became more mature to be able to follow the latest updates and search literature from PubMed by themselves. Usually, the PI will choose the journal where researchers aim to publish results. The choice is made according to the quality and scope, which is measured mainly by the impact factor (IF) in the field.

*5.1.4 Leadership and supervision.* PIs are leading and supervising the research work. The most significant implication of this to the information practices is perhaps the way they guide younger researchers to search information or data and, for example, which tools they prefer. They have an important role also in transferring the most important elements of the biomedical research culture to junior researchers. An example of this is publishing practices.

*5.2 The indicators of strong mutual dependence*

Scientific publications and research data are shared widely and organized into a relatively coherent system. The portal of NCBI (Entrez) is an example of a centralized service, which has been able to integrate different data sources and literature as an advantageous service to the global biomedical community. In 2015, Entrez integrates literature and data from 39 different databases (www.ncbi.nlm.nih.gov/books/NBK3837/). However, the amount of data and its growth rate are so extensive that the integration and interoperability of different data sources is a growing challenge (Buetow, 2005; Gadaleta *et al.*, 2011; Hoehndorf *et al.*, 2015). It is typical to many biomedical data resources that these have been based on a common effort and open access compared with many other domains. It is a shared rule that, for example, sequence data are deposited for the use of the whole research community in the common database. The amount and variety of data has caused the development of diverse tools (Roos *et al.*, 2008).

The development and use of the biomedical ontologies has been one solution for the data management and integration problem. The use of ontologies started about 1998 with the development of the Gene Ontology. Since then the amount of ontologies has increased and because of this, the repositories of ontologies have been established. This has created a need to co-ordinate them internationally as is done in efforts like Open Biomedical Ontologies Foundry and the National Centre for Biomedical Ontologies (BioPortal).

There is a long tradition in biomedicine to use a shared indexing tool of literature, called Medical Subject Headings. This controlled vocabulary thesaurus is updated centrally but in global collaboration. The data of the thesaurus are also openly available (www.nlm.nih.gov/pubs/factsheets/mesh.html).

The vast majority (91 percent) of MM researchers in our case study were searching data and using data resources during their current project. Among medical scientists the situation was different. Most of the researchers did not have enough knowledge to use the databases even though the databases were found important for their research.

PubMed, produced by NCBI, is the most important searching tool for scientific literature. In total, 95 percent of MM researchers and all medical scientists in our studies were using it. Information is searched from PubMed in a very simple way, often like laymen use the general search engines (Roos *et al.*, 2008; Roos, 2015).

Even though commercial publishers dominate in the biomedical journal market, there has been a strong demand of open access. NIH has been among the first research funders who demand that research results that are financed by NIH, have to be openly available to all. In 2000, NIH/NLM launched an open repository for biomedical literature. NIH demands that even though the outcomes of the research that they are financing are published in traditional commercial platforms, the paper has to be submitted also to PubMedCentral.

*5.2.1 Competitive environment*. Hard competition in biomedicine has effects on publishing speed and could be one reason behind the exponential growth of the number of published scientific articles. The contents of PubMed have been studied by using text mining techniques (e.g. Jensen *et al.*, 2006; Wang and Zhu, 2013; Zhou, 2012). Jensen *et al.* (2006) have an example of how the popularity of some topics as a research object may vary during time. The increase of certain "hot" topics may be so extreme that it is impossible to a human read all new papers. Because of this a great number of text mining tools have been developed (see e.g. Nadkarni and Parikh, 2012).

Because of the publishing speed, current awareness is very important for biomedical researchers. We observed that all medial scientists in our case study were actively following developments in their research area during every stage of the research process. They used varied methods and tools for this purpose. The most common among researchers was to make a constant search to PubMed with an e-mail alert. This was also an ordinary practice among MM researchers.

Another publishing-related indication of high mutual dependence could be the selection of the publishing platform. Biomedical journals seem to form quite an obvious hierarchy where the status of commercial and commercial society publishers is strong. The MM researchers in our study indicated quite clearly that they intend to publish in those journals that have the highest IF. As 51 percent of MM researchers named IF as the most important selection criteria, 36.5 percent named the focus of the journal as the most significant reason to choose the publishing platform. In our MM case, the researchers named totally 66 different journals when asked to name the most important journal titles, which they read. The most common being *Science* (38.1 percent), followed by the *American Journal of Human Genetics* (*AJHG*) (23.8 percent), *Human Molecular Genetics* (20.6 percent) and *Cell* (12.7 percent). When asked about publishing plans and suitable publishing forums they named totally 44 journals, most often *Nature Genetics* (34.1 percent), *Human Molecular Genetics* (25 percent), *Molecular Psychiatry* (22.7 percent) and *AJHG* (20.5 percent). When the actual publishing practices were checked from PubMed, it appeared that these researchers really published in 60 different journals between 2007 and 2010. The most popular journal was *Nature Genetics* (15.8 percent, IF 25.556 in 2007), *PLoS Genetics* (7.6 percent, IF 8.721 in 2007), *AJHG* (7 percent, IF 11.092 in 2007) and *Human Molecular Genetics* (5.8 percent IF 7,806 in 2007), these four journals covered 36.3 percent of the whole publishing volume in 2007-2010.

*5.2.2 Dependence on technology*. The dependence on technology is more characteristic of the researchers in MM than among medical scientists. The reason for this is the reliance on computational tools and methods that are needed in MM researchers' work. These result in the production of huge amounts of data, which need to be analyzed and affects directly information practices. Our observation was that MM researchers named a broad category of different databases, tools and services, which they used interlaced. Another finding was that the number of actively used databases was quite limited (Roos *et al.*, 2008).

*5.2.3 Collaboration*. Co-authorship is in general the most typical form of collaboration in scientific research work. In biomedicine researches write articles together following a quite clear division of labor. PhD researchers are usually the responsible writers and PIs are included in the process. There are often several researchers from other local, national or international research groups included in the writing process, but their role seem to be more supportive. Genome-wide research needs global co-operative work and those articles that report the results may include hundreds of contributors (see e.g. www.nature.com/ng/journal/v46/n11/pdf/ng.3097.pdf).

We were a little surprised that none of the groups had a collective reference database or any group working tool which could help in the writing process. Sharing information about the published literature was based mainly on what Talja (2002) calls directive sharing. Directive sharing happened between the PI or senior and the doctoral student. Medical scientists had so called "journal clubs," regular meetings in the clinic. The purpose of these gatherings is to inform other clinicians about new research results. Predominantly young researchers have the duty to review a certain topic.

The development of certain common, collaboratively developed and centrally coordinated tools have been important to the development of the whole biomedical domain. Examples of this are the work for the biomedical ontologies, the development of various open databases and tools and integrated services that NCBI and NLM have developed.

## 6. Discussion and conclusions

We identified in this study several indicators, which represented task uncertainty and mutual dependence in biomedicine. Low task uncertainty is represented by: a common goal for research, strong emphasis on research groups, clear division of labor and clear leadership and supervision. Following indicators represent high mutual dependence: sharing of resources, strong competition in the research field dependence on new technology to produce competitive research and collaboration within the research unit as well as national and international collaboration. Some of the indicators seem to be interrelated. Dependence on an expensive technology may, for example, be a reason for research groups to find collaborators. A clear tension appears in the field because of the need for collaboration when there is prevailing a strong competition between research groups.

Analyzing information practices against the found indicators helped us to understand how the specialty of the domain may affect information practices. Hard competition in the field has many effects. It is one factor behind the explosion of the data and publications, which again is directly related to the ways information is searched, followed, used and produced. The need for developments in information services in this area is obvious. Easy to use literature and data searching tools, text and data mining tools and current awareness services are outstandingly important in the scientific domain of biomedicine. The constant competition is also related to the direction that the publishing sector has developed in biomedicine. The status of commercial publishers is strong. The novelty and the value of the contribution of new research are controlled by the established system of commercial scientific journals. Journals have a clear ranking system of IFs. Researchers take into consideration that when aiming to publish their research outcomes in the most valued journals.

Our study supports the view that biomedical domain has a quite low task uncertainty and the mutual dependence of the researchers is considerably high. It seems, as noticed in previous studies (Hedlund and Roos, 2007) that when the degree of mutual dependence between researchers is increasing, the methods for scientific communication become more controlled. In biomedicine, where researchers are mutually dependent, competition increases and citation patterns become more important. This has affected publishing patterns, which have become more standardized and restricted, emphasizing the journals of the highest rank and IFs.

Because biomedical research culture is based on working in groups, where the work is quite clearly divided, it could be beneficial to exploit these features when information services are developed and organized. As the work is organized according to laboratories or clinics where PI is supervising various projects simultaneously, this should be taken into consideration. From the point of view of the library it would be beneficial if that key person would be well informed about the usefulness and availability of various tools and services that the library offers. It would help the other members of the group to adopt and take the advantage of these services. Because of these facts, research in IS should focus more to the needs and specialties that working in groups are causing in the biomedical and other domains. With the help of this information it would be possible to create services that would support research work in the best possible way.

Hard competition and rapid publishing speed cause a real challenge to biomedical researchers. Information about new research results needs to be available and accessible as soon and as easy as possible. All researchers, during the whole research process need this information. To be able to develop tools and organize services for searching information about new research results is among the most important services of the library to the biomedical researchers. More attention in libraries should be paid to the text mining technologies. These become more and more important as the amount of biomedical literature is rising exponentially. It is possible that in the not so far future, traditional library services in certain biomedical fields will be replaced totally by text mining and related technologies.

The hard competition is a stress factor for biomedical researchers and makes them exceptionally aware of time. In addition, medical scientists do clinical work, which causes strict, time bound requirements to the research work. Because of this, information practices have to be easy, accessible and as integrated as possible. The status of the globally shared and centrally coordinated tools and services is stable. These facts need to be respected when library services and tools are developed. The integration of all relevant resources is a necessity.

Information practices need to be studied in context, not understood as independent, separate doings. Domain approach in the study of information practices seems to be useful for this purpose. With the help of Whitley's theory, we were able to make a thorough analysis of the cultural and social nature of the biomedical domain. This aided in the examination of the information practices of researchers. In the future, we think, it would be beneficial if the research in IS would focus more on the special requirements in a particular scientific field or domain. Because of this, it would be advantageous to utilize Whitley's theoretical frame on the study of information practices in various other fields or domains. We believe that this understanding would lead to tools and services that would be the most beneficial to the scientific community in their work of creation of new knowledge and innovations in that particular field.

## 7. Limitations of the study

There are some limitations to this study. Originally, Whitley's theoretical frame was developed for the comparison of different domains or fields. In spite of that we found it useful in the analyses of the specialties within one domain. Another limitation is that the data to some parts are collected as early as in 2007. This might have affected the results in a rapidly developing field like biomedicine. It is also important to realize that the majority of researchers in this study were junior researchers, often PhD students. If only senior or post-doctoral researchers would have been included, the results might have differed at least in some points.

**References**

Bates, M.J. (2005), "An introduction to metatheories, theories, and models", in Erdelez, F.S. and Lynne McKechnie, K.E. (Eds), *Theories of Information Behavior*, Information Today, Medford, OR, pp. 1-24.

Buetow, K.H. (2005), "Cyberinfrastructure: empowering a 'third way' in biomedical research", *Science*, Vol. 308 No. 5723, pp. 821-824.

Fry, J. (2006), "Scholarly research and information practices: a domain analytic approach", *Information Processing & Management*, Vol. 42 No. 1, pp. 299-316.

Fry, J. and Talja, S. (2007), "The intellectual and social organization of academic fields and the shaping of digital resources", *Journal of Information Science*, Vol. 33 No. 2, pp. 115-133.

Gadaleta, E., Lemoine, N.R. and Chelala, C. (2011), "Online resources of cancer data: barriers, benefits and lessons", *Briefings in Bioinformatics*, Vol. 12 No. 1, pp. 52-63.

Hedlund, T. and Roos, A. (2007), "Open access publishing as a discipline specific way of scientific communication: the case of biomedical research in Finland", in Garten, E.D. *et al.* (Eds), *Advances in Library Administration and Organization*, Vol. 25, Elsevier, Amsterdam, pp. 113-131.

Henderson, M.K. and Bunton, S.A. (2013), *Handbook of Academic Medicine: How Medical Schools and Teaching Hospitals Work*, Association of American Medical Colleges, Washington, DC.

Hjørland, B. (2005), "Empiricism, rationalism and positivism in library and information science", *Journal of Documentation, Emerald Group Publishing Limited*, Vol. 61 No. 1, pp. 130-155.

Hjørland, B. (2011), "Evidence-based practice: an analysis based on the philosophy of science", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 7, pp. 1301-1310.

Hjørland, B. and Albrechtsen, H. (1995), "Toward a new horizon in information science: domain-analysis", *Journal of the American Society for Information Science*, Vol. 46 No. 6, pp. 400-425.

Hoehndorf, R., Schofield, P.N. and Gkoutos, G.V. (2015), "The role of ontologies in biological and biomedical research: a functional perspective", *Briefings in Bioinformatics*, Vol. 16 No. 6, pp. 1069-1080.

Hunter, L. and Cohen, K.B. (2006), "Biomedical language processing: what's beyond PubMed?", *Molecular Cell*, Vol. 21 No. 5, pp. 589-594.

Jensen, L.J., Saric, J. and Bork, P. (2006), "Literature mining for the biologist: from information retrieval to biological discovery", *Nature Review Genetics*, Vol. 7 No. 2, pp. 119-129.

Lu, Z. (2011), "PubMed and beyond: a survey of web tools for searching biomedical literature", *Database: The Journal of Biological Databases and Curation*, Vol. 2011, baq036, available at: http://doi.org/10.1093/database/baq036

Massoud, T.F. and Gambhir, S.S. (2007), "Integrating noninvasive molecular imaging into molecular medicine: an evolving paradigm", *Trends in Molecular Medicine*, Vol. 13 No. 5, pp. 183-191.

Mayor, C. and Robinson, L. (2014), "Ontological realism, concepts and classification in molecular biology: development and application of the gene ontology", *Journal of Documentation*, Vol. 70 No. 1, pp. 173-193.

Nadkarni, P. and Parikh, C. (2012), "An eUtils toolset and its use for creating a pipeline to link genomics and proteomics analyses to domain-specific biomedical literature", *Journal of Clinical Bioinformatics, BioMed Central*, Vol. 2 No. 1, pp. 1-7.

Newman, M.E.J. (2001), "Scientific collaboration networks. I. Network construction and fundamental results", *Physical Review E*, Vol. 64 No. 1, pp. 16131.

Nicholas, D., Huntington, P. and Jamali, H.R. (2006), *Determining the Impact of Open Access Publishing on Use and Users: A Deep Log Analysis of Nucleic Acids Research*, CIBERUCL Centre for Publishing, London.

Palmer, C.L. and Cragin, M.H. (2009), "Scholarship and disciplinary practices", *Annual Review of Information Science and Technology*, Vol. 42 No. 1, pp. 163-212.

Pennisi, E. (2011), "Will computers crash genomics?", *Science, American Association for the Advancement of Science*, Vol. 331 No. 6018, pp. 666-668.

Rivera, G. and Cox, A. (2014), "An evaluation of the practice based approach to understanding the adoption and use of information systems", *Journal of Documentation*, Vol. 70 No. 5, pp. 878-901.

Roos, A. (2012), "Activity theory as a theoretical framework in the study of the information practices in molecular medicine", *Information Research*, Vol. 17 No. 3, available at: http://informationr.net/ir/17-3/paper526.html#.UKue0OPZ8w1

Roos, A. (2015), "Medical scientists' information practices in the research work context", *Health Information & Libraries Journal*, Vol. 32 No. 1, pp. 23-36.

Roos, A., Kumpulainen, S., Järvelin, K. and Hedlund, T. (2008), "The information environment of researchers in molecular medicine", *Information Research*, Vol. 13 No. 3, pp. 7.

Savolainen, R. (2007), "Information behavior and information practice: reviewing the 'umbrella concepts' of information-seeking studies", *The Library Quarterly*, Vol. 77 No. 2, pp. 109-132.

Savolainen, R. (2008), *Everyday Information Practices: A Social Phenomenological Perspective*, Scarecrow Press, Lanham, MD, pp. 1-7.

Stein, L.D. (2010), "The case for cloud computing in genome informatics", *Genome Biology*, Vol. 11 No. 5, p. 207.

Talja, S. (2002), "Information sharing in academic communities: types and levels of collaboration in information seeking and use", *New Review of Information Behavior Research*, Vol. 3 No. 3, pp. 143-159.

Talja, S. (2005), "The domain analytic approach to scholars' information practices", in Erdelez, F.S. and Lynne McKechnie, K.E. (Eds), *Theories of Information Behavior*, Information Today, Medford, OR, pp. 123-127.

Trent, R.J. (2012), *Molecular Medicine, Molecular Medicine*, 4th ed., Elsevier, Amsterdam, available at: http://doi.org/10.1016/B978-0-12-381451-7.00001-3

Wang, Y. and Zhu, F. (2013), "Mining out hot points in Biomedicine", *Global Journal on Technology*, Vol. 3, pp. 719-724.

Weiner, S.A. (2009), "Tale of two databases: the history of federally funded information systems for education and medicine", *Government Information Quarterly*, Vol. 26 No. 3, pp. 450-458.

Whitley, R. (2000), *The Intellectual and Social Organization of the Sciences*, 2nd ed., Oxford University Press, Oxford.

Wilson, H.J. (2000), "The myth of objectivity: is medicine moving towards a social constructivist medical paradigm?", *Family Practice*, Vol. 17 No. 2, pp. 203-209.

Zhou, Y. (2012), "Extracting top hot hits in biomedicine from PubMed", *IEEE International Symposium on Information Science and Engineering (ISISE)*, pp. 266-269.

## Appendix

### The Information Environment of Researchers at the Department of Molecular Medicine, KTL

The aim of this inquiry is to find out the types and usage of information resources at MLO. Please, kindly fill in the form as precisely as possible.

**Your name** ______________________

**Your main discipline (eg. biochemistry)** ______________________

**Academic degree**

- ☐ PhD
- ☐ MD
- ☐ Bsc
- ☐ MSc
- ☐ Undergraduate
- ☐ Other

**Status**

- ☐ Group leader
- ☐ Senior researcher
- ☐ Post doc.
- ☐ Doctoral student
- ☐ Graduate student

**How long have you been doing reseach work?**

- ☐ more than 15 years
- ☐ 10-15 years
- ☐ 5-10 years
- ☐ 2-5 years
- ☐ 1-2 years
- ☐ less than one year
- ☐ I'm a real beginner

**Students, please name your supervisor(s)**

______________________________________________________

______________________________________________________

______________________________________________________

______________________________________________________

**Name the leader of your research group ( if not mentioned above)** ______________________

**Your current research subject(s)**

______________________________________________________

______________________________________________________

______________________________________________________

______________________________________________________

**With which work tasks are you most occupied at the moment (you can choose several alternatives)**

- ☐ working in the laboratory
- ☐ searching information from literature databases or other sources
- ☐ reading literature
- ☐ writing an article or a report
- ☐ searching data from data collections/databases
- ☐ scientific computing
- ☐ developing of programs or other tools
- ☐ planning (e.g. a new project)
- ☐ reseach administration
- ☐ Other, please specify #1 ______________________
- ☐ Other, please specify #2 ______________________
- ☐ Other, please specify #3 ______________________

**Do you have close partners or collaborators, people outside your research group you contact frequently concerning your current project(s)**

| | |
|---|---|
| MLO | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |
| KTL | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |
| elsewhere in Finland | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |
| in Europe | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |
| in the U.S.A. | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |
| or in the other parts of the world | ☐ no<br>☐ 1-5<br>☐ 5-10<br>☐ 11-20<br>☐ more than 20 |

The following section contains questions, about information resources that you have used during your current research project(s)

**When you have had a need for information, have you preferred to**

- ☐ contact a colleague or a supervisor
- ☐ use the internet/intranet (all kinds of databases and resources on the net)
- ☐ use information sources located in the laboratory or office
- ☐ contact library or information specialist
- ☐ Other, please specify ____________

**Which type of published material have you used in current project(s), e.g. books, scientific articles?**

______________________________
______________________________
______________________________
______________________________

**Please, list journal titles you follow regularly**

______________________________
______________________________
______________________________
______________________________

**Which types of literature search engines and other finding tools have you used to find literature, e.g. PubMed, Web of Science, library catalogs, GoogleScholar, Nelli?**

______________________________
______________________________
______________________________
______________________________

**If you have used internet to find information (other than information about literature), which tools did you use, e.g. Google, Wikipedia, blogs, collaborative news sites, other www sites/pages?**

______________________________
______________________________
______________________________
______________________________

**Which collections of data have you used, e.g. Nucleotide Sequences, UCSC Genome Browser, Entrez databases, EMBL etc?**

______________________________
______________________________
______________________________
______________________________

**What other information resources, like personal data collections have you used during your current project(s)?**

______________________________
______________________________
______________________________
______________________________

**What tools, like Primer3, Sequence Manipulation Suite, BCM Search Launcher have you used?**

______________________________
______________________________
______________________________
______________________________

**Please, choose from the above mentioned at least 3 most useful resources for your current project(s) at the moment**

________________________________________
________________________________________
________________________________________
________________________________________

**Do there exist any particularly useful www-sites, which you would like to mention?**

________________________________________
________________________________________
________________________________________
________________________________________

**How many articles or other material are you planning to publish this year?**

| | |
|---|---|
| as the 1st, 2nd or last author | ☐ None<br>☐ 1<br>☐ 2<br>☐ more than 2 |
| as a co-author | ☐ None<br>☐ 1<br>☐ 2<br>☐ more than 2 |
| conference papers or posters | ☐ None<br>☐ 1<br>☐ 2<br>☐ more than 2 |
| theses | ☐ None<br>☐ 1<br>☐ 2<br>☐ more than 2 |
| other material | ☐ None<br>☐ 1<br>☐ 2<br>☐ more than 2 |

**Please, name three journal titles which have a scope that you find suitable for publishing**

________________________________________
________________________________________
________________________________________
________________________________________

**What is the most important factor in choosing the journal for publishing?**

☐ impact
☐ publishing speed
☐ open access
☐ scope
☐ Other, please specify ____________________

**Corresponding author**
Annikki Roos can be contacted at: annikki.roos@nic.fi