Challenges of transformation of research data into open data: the perspective of social sciences and humanities[1]

Aleksei Kelli, Tõnis Mets, Kadri Vider and Age Värv, University of Tartu
Lars Jonsson, Uppsala University Holding Company
Krister Lindén, University of Helsinki
Ramūnas Birštonas, Vilnius University

Abstract
The authors address the transformation of research data into open data. The article draws on the experience in four countries: Sweden, Finland, Estonia and Lithuania. The transformation process presents several challenges where legal, organizational and individual aspects influence the process. Research data often contain personal data. Research data could also covered with intellectual property (IP) rights. This means that personal data and IP regulations should be integrated into the dissemination model. While there is a potential conflict between the policies for open data that aim to make data freely available and those of an entrepreneurial university that emphasize commercialization of research results, these policies need to be made compatible. Researchers producing data are vital for reconciling the two, but they currently lack the motivation to contribute towards the implementation of the open data policy due to missing career incentives.

Keywords
open data
research data
entrepreneurial university
personal data protection
intellectual property
academic career
incentives

Introduction
The article addresses the transformation of research data into open data. Digital technologies have created opportunities for unlocking the potential of research data. The process has several rationales. First, since taxpayers' money is used for research, the results should be freely available to everyone. Second, free access to data should result in valuable new knowledge. In many research areas, we experience the evolution of disruptive knowledge when large data volumes are collected and analysed efficiently. An excellent example from the field of medicine is the increased survival rate for some types of cancer because of a more individualized treatment based on data analysis. Another example is the changed view of gender equality in the medieval Swedish society based on the data analysis of court protocols stemming from research in the field of history. Data sharing and big data analysis will soon be essential research tools in many more fields of social sciences and humanities. It creates new demands on both the university's administration to set up the necessary infrastructure for gathering, assuring quality and storing the data, and on the research community to make sure that there is a fair incentive structure in place for those researchers who make valuable data available for others to analyse and publish on. Individual researchers often have a vested interest in keeping their research data to themselves until they have fully utilized the data, which might conflict with general policies promoting the open sharing of research data. The goal to share research data interacts with legal restrictions such as personal data and

intellectual property, and with political and economic considerations favouring wide dissemination and utilization of publicly funded research results and general ambitions to have entrepreneurial universities contribute towards a knowledge society.

The article conceptualizes issues related to open data at policy (strategy), regulatory, organizational (research institutions) and individual (researcher) levels. The aim is to provisionally map the current situation and make recommendations on how to enhance the open data policy further. The article has a practical focus on identifying and trying to overcome potential obstacles. To some extent, open data issues are discipline specific, which means that challenges relating, for example, to chemistry could be different from those of social sciences and humanities. This article focuses mainly on the transformation of research data into open data in the realm of social sciences and humanities by studying language research as an example.

Since research data and open data are the central concepts framing the article, it is necessary to grasp them from the outset. The European Commission (EC) (2017a: 4) defines research data as 'information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation'. Statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images are provided as examples (EC 2017a: 4). The OECD (2015a: 8) similarly defines research data as 'factual records used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings'. Despite these policy-level definitions (EC, OECD), it is not always easy to define data in specific situations. For instance, books, articles, pictures and motion pictures could be treated as research data if they are studied for some purpose (e.g. text and data mining). The restrictive definition of research data fails to address these issues. Therefore, we loosely define research data as any data used for research.

The concept of open data is developed within the context of open science. The OECD (2015: 7) defines open data as 'data that can be used by anyone without technical or legal restrictions. The use encompasses both access and reuse'. According to Open Knowledge International 'open data and content can be freely used, modified, and shared by anyone for any purpose'. Finnish policy documents define open data as 'data reusable by anyone, free of charge, with permission and in a machine-readable format' (Ministry of Finance of Finland 2015: 7). Estonian policy documents have adopted a similar approach (Estonian Research Council 2016: 14) (for further discussion on openness see Kelli et al. 2018).

In the first part of the article, the authors explore the general issues of open data such as personal data and intellectual property protection, dissemination models and conditions and known barriers. This general outline is used as a theoretical framework for case studies. In the second part, the article expands on the open data case studies originating from four jurisdictions (Sweden, Finland, Estonia and Lithuania). Since some of the authors are involved in national and international language research, several examples and cases are related to this domain. The authors argue, first and foremost, that there are several individual, organizational and state-level barriers that influence the implementation of open data policy. Open data policy is not the thing-in-itself. Its ultimate objective is to enhance social prosperity. Therefore, there is a need to strike a fair balance with other policies such as protection of individuals' privacy (implemented through personal data protection) and entrepreneurial university policy supporting the transformation into a knowledge-based economy.

The article is structured as follows: in Section 2, we introduce the background from regulatory and policy perspectives on legal rights covering data, requirements and limitations to data collection and time and quality issues affecting the dissemination of open data. Also, individual and organizational barriers are addressed. In Section 3, we explain our research methodology, followed by a presentation of our findings in Section 4. The issue of a possible contradiction between the policies of open data dissemination and the entrepreneurial universities is examined in Section 5, followed by a general discussion in Section 6 and conclusions in Section 7.

Background

Legal rights covering research data

Depending on the nature of research data (factual data, statistics, interviews and so forth) these can be covered with personal data and intellectual property rights (copyright, related rights to copyright and industrial property). Since the scope and limitations of legal rights covering data are not the primary focus of the article, they are addressed in brief to identify issues requiring attention when implementing the open data policy. First, we shall briefly outline the essence of these rights. At the end of the section, an example of language data is provided to demonstrate that research data are sometimes simultaneously covered with several rights.

Personal data protection

Personal data protection could pose challenges not only to the implementation of open data policy but also to research in general (for further discussion see Kelli et al. 2018; Klavan et al. 2018) if implemented too restrictively.

First, research data often contain personal data, the use of which is restricted. The General Data Protection Regulation (GDPR) defines personal data as 'any information relating to an identified or identifiable natural person' (Art. 4 [1]). According to the Article 29 Working Party (advisory body on personal data protection, WP29), personal data cover

> information available in whatever form, be it alphabetical, numerical, graphical, photographic or acoustic, for example. It includes information kept on paper, as well as information stored in a computer memory by means of a binary code, or on a videotape. (WP29 2007: 7)

The name of a person (C-101/01) and even incorrect data are personal data (WP29 2007: 6). Publicly available personal data are also protected (C-73/07). The critical issue here is how far we should go when it comes to identifiability. Research literature distinguishes between absolute and relative approaches depending on the effort used to identify the person (Spindler and Schmechel 2016: 165–66). WP29 expresses the view that 'a mere hypothetical possibility to single out the individual is not enough to consider the person as "identifiable"' (2007: 15). It is pointed out that data can become identifiable through combination with other datasets and identifiability is context dependent (Oostveen 2016: 306). Even if we adopt the relative approach, the dissemination of a large amount of research data is still restricted due to data protection reasons.

Second, while the open data policy aims to make large amounts of data freely available and usable, data protection restricts data usage. Purpose limitation (for further discussion on

purpose limitation see WP29 2013) and data minimization are central GDPR principles setting a framework for data processing (GDPR Art. 5). Therefore, it is indicated in the research literature that conflicts between personal data protection and big data are fraught with difficulties in finding a legitimate ground for processing data and the application of the principles of purpose limitation and data minimization (Oostveen 2016: 309).

Open data policy and personal data protection entail different policy considerations. Open data aim to make data available for the promotion of social prosperity, but personal data protection preserves the privacy of a natural person. Therefore, there is a need to strike a fair balance between the two policies. Neither openness despite the costs and suffering of individuals nor locking up data are the desired outcomes.

Copyright and database protection

If data include works (e.g. books, articles, web posts), copyright protection applies. This means that the author of such works enjoys an exclusive right to use and to authorize or prohibit the use of the work. The use of this type of data can rely on the copyright exceptions and limitations (e.g. research use), which are not analysed here.

It is a widely recognized approach that copyright does not protect data as such (restrictively defined as excluding copyright-protected works). For instance, the Estonian Copyright Act (CA) does not apply to facts and data (§ 5 [7]). Systematically arranged data, however, can be protected as a database (§ 4 [3] 22); § 75^2). The basic framework of database protection has been established at the international level. According to the Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPS Agreement), compilations of data that due to the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection shall not extend to the data itself (Art. 10 [2]). The same legal norm is found in the World Intellectual Property Organization (WIPO) Copyright Treaty (Art. 5). Database protection is harmonized at the EU level as well. The Database Directive defines a database as 'a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means' (Art. 1 [2]). Besides copyright, the Database Directive provides additional *sui generis* protection of databases. Under the Database Directive, the main criterion of the *sui generis* protection is 'a substantial investment in either the obtaining, verification or presentation of the contents' (Art. 7 [1]). The Court of Justice of the European Union (CJEU) has explained that investment as a protective criterion refers to 'the resources used to seek out existing independent materials and collect them in the database. It does not cover the resources used for the creation of materials which make up the contents of a database' (C-203/02, point 1).

There is an ongoing discussion about whether to introduce a data producer's right for raw machine-generated data at the EU level. According to the European Commission (2017b), 'A right to use and authorise the use of non-personal data could be granted to the 'data producer'', i.e. the owner or long-term user (i.e. the lessee) of the device. This approach would aim at clarifying the legal situation and providing more choice to the data producer by opening up the possibility for users to utilize their data and thereby contribute to unlocking machine-generated data. Some intellectual property scholars are critical of this proposal (see Hugenholtz 2017; Drexl et al. 2016). We share their concerns.

Trade secret protection

Data can also potentially be protected as industrial property (e.g. patent, industrial design, trade secret). Research data often qualify for trade secret protection. In this section, we focus

on trade secret protection (undisclosed information) since it is most likely to conflict with an open data policy (see Section 5 on the interaction between open data and entrepreneurial university policies). At the international level, the obligation to protect undisclosed information is stated in Art. 39(2) of the TRIPS Agreement. At the EU level, trade secret protection was recently harmonized by the Trade Secrets Directive, which obliges the EU member states to take measures against the unlawful acquisition, use or disclosure of trade secrets by other parties.

According to the Directive, a trade secret is information which is secret, has commercial value because it is secret, and the person who controls the information has taken steps to keep it secret (Art. 2).

The Trade Secrets Directive acknowledges the importance of trade secret protection both for businesses and for non-commercial research institutions that invest in acquiring, developing and applying know-how and information that is the currency of the knowledge economy and provides a competitive advantage (Recital 1). It is a widespread practice that universities sell their knowledge in the form of trade secrets.

The example of language data

Language data well exemplify the legal difficulties in the transformation of research data into open data. The following figure outlines legal rights covering research data (language resources):
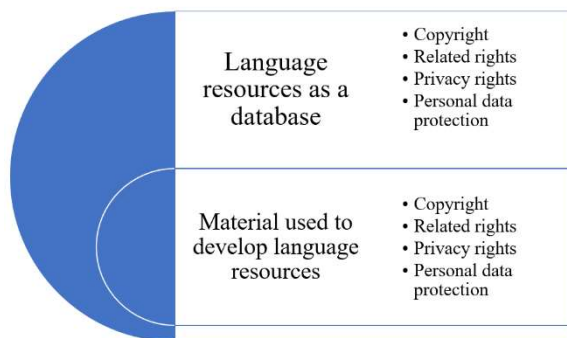

Figure 1: Two tiers of rights covering language resources (Kelli et al. 2015).

Different layers of rights cover language data: first, input materials used for language research (e.g. books, oral speech, articles, blogs). They are often protected by copyright, related rights and personal data rights. Second, research results in the form of language data are protected as a database (Kelli et al. 2015).

Personal data and intellectual property protection regulations do not render open data policies useless. However, they should be combined to create a holistic approach to research data management. Intellectual property and personal data rights can be addressed through contractual instruments such as open licences and consents (for further discussion see Kelli et al. 2018) and statutory limitations (e.g. research exemption). It should also be mentioned that the GDPR does not apply to anonymized data (Recital 26).

Dissemination of open data

Requirements and limitations of data collection

A considerable amount of geographical and financial information, population, migration, welfare and health statistics and other economic and demographic data are produced using standardized procedures accepted by international and national institutions, and statistical offices of countries (Salgé 1999; Hermoso et al. 2015). Hence, standardization can be enforced *de jure* according to international agreements and applied de facto. An example of a coordinating function is the International Organization for Standardization (ISO) – a data standardization organization whose recommendations are followed by institutions worldwide. The World Bank, the OECD, the European Commission and a few others are institutions providing open, standardized data. The data in their repositories are freely available for multiple purposes (including research). Although the interpretation of standardized data might be different, context-based, variances are minimal.

Similar tendencies are observed at the national level. For instance, according to the Finnish National Interoperability Framework Observatory (2016)

> […] diverse information resources are provided as open data, from geodata to weather, climate, sea, soil related, transport, financial, statistical and cultural data. Also, a growing number of municipalities are opening up their data. Such data were not interoperable regarding their content or technical aspects, and need to be standardised.

Therefore, the standardization of data is a crucial issue in the implementation of the open data policy.

The following stages frequently describe the traditional data lifecycle: planning, collecting, quality assurance and quality control and analysis. Making data public needs data documentation (metadata) and archiving (in a public repository), which enables the next stages: discovery, integration and analysis (Rüegg et al. 2014). Many other researchers can use the data collected by the original investigator for their studies. It allows for the additional validation of data and research conclusions, and for new analyses and outcomes.

The process of data creation and collection should be analysed to assess the nature of research data in open repositories. The description of the data life cycle is oblivious to the aim, for what purposes (goal and tasks) and how (methodology) the datasets were created. The problem may arise in many original social studies, for example, where the context of the research and conditions of data collection exert a profound influence on the survey questionnaire or interview answers. The impact may be stronger in qualitative studies using specific methods. This means that the data life cycle overlaps with the research life cycle. In the research life cycle, research gaps (problems), hypotheses and research models, and methodology, become more critical for tracking the phenomena than the final data collected from a particular sample. Thus, opening a dataset in some cultural context enables duplication of the analysis for this particular dataset only and does not say much about the validity of the research model and phenomena in the cultural diversity context. For example, the study on learning organizations in Estonia (Mets and Torokoff 2007), published in an open-access journal, was repeated in Turkey (Yaşlıoğlu et al. 2014), validating the research concept, model and methodology designed by the original authors. Using only the open data and repeating the data analysis would not have verified the general applicability of the

methodology of the original study equally well. This indicates the need for a broader understanding of the interaction between open data and methods.

Not all data can meet the requirements of ensuring complete freedom and openness of data for research purposes (cf. personal data). However, the data may be available in a more limited way, and yet still be useful for research. The FAIR principles set forth minimum requirements for making data available for research in a repository, i.e. data should be (FORCE11 2017):

- Findable
- Accessible
- Interoperable
- Re-usable.

The FAIR principles are also advocated by the Horizon 2020 programme in the Guidelines on FAIR Data Management in Horizon 2020 (EC 2016a).

From the example above where the social study was repeated in Turkey on another sample in another cultural environment, we see that data compatible with the FAIR requirements do not guarantee the right context, analysis or interpretation of results. There are no specific formal rules to manage know-how limitations of data usage. This reveals the need for a more extensive application of open science principles (open methodology, open source), which is not only confined to data. This highlights the importance of open access to the documentation of the data to enable scientifically comparable and relevant reuse.

Time of dissemination, responsibility for the quality and dissemination model

There are numerous practical issues such as timing, the extent of liability and the dissemination model that need to be addressed when implementing the open data policy.

It is emphasized that big data analytics rely on real-time analyses (Drexl 2016: 14). The OECD (2015b: 323) points out that researchers do not have incentives to disclose datasets at the pre-publication stage. Researchers and research institutions delay the dissemination of data if they plan to patent research results (publication of data results in a loss of novelty) or use them for commercial purposes relying on trade secret protection. In this context, academic entrepreneurship takes precedence over the open data policy. However, the vast bulk of research data are not likely to be used for commercial purposes, and it can theoretically be disseminated (provided that other obstacles/disincentives have been removed).

Potential liability is also a relevant issue that should not be ignored. It is possible that the open data contain errors. The real question is who should assume the potential risks. If we say that researchers and research institutions are legally liable for the accuracy of the data that they openly share for free, then why would they do it? It is wiser not to share, and this way avoid liability. Therefore, the central principle should be that data are made available on an as-is and as-available basis. For instance, this approach is followed by Common Language Resources and Technology Infrastructure (CLARIN), which makes digital language resources available. It uses the following procedure to disseminate language data (see Kelli et al. 2015):

> o Data are deposited with repositories, and a CLARIN Deposition Licence Agreement is concluded;
> o The user accepts the CLARIN Terms of Service and End-user Licence Agreement (EULA).

The CLARIN contractual framework stipulates that repositories are not to be held responsible for research data, i.e. CLARIN disclaims warranty and liability. Standard licences have a similar approach. For instance, Creative Commons (CC) provides that 'the Licensor offers the Licensed Material as-is and as-available and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other' (Sec 5). The European Union Public Licence (EUPL) also disclaims warranty and liability (Sec 7 and 8). To sum up, it is recommendable to disclaim liability and warranty for research data.

The successful implementation of an open data strategy requires that research data be made available together with legal metadata. The FAIR Data Principles also set forth that re-usability requires a clear and accessible data usage licence. The idea behind this is that the user understands the terms of service. It should be clear whether research data can be used for research purposes only or also for commercial purposes. To address the requirement, CLARIN has created a laundry tag system to facilitate the use of language data (also called language resources). CLARIN has three main categories of research data (CLARIN licensing framework):

1) Public (PUB)
2) Academic (ACA)
3) Restricted (RES).

According to the experts Oksanen et al. (2010), who contributed to the establishment of the system, the category 'publicly available' allows for the use of research data without limitations. The category 'academic use' includes a stipulation that data are used for research only by a member of the research community. The 'restricted use' category refers to a situation where the use is restricted to an individual (e.g. a copy is available only for a particular research purpose because of personal data protection). One of the advantages of the CLARIN tripartite division is its simplicity and user-friendliness. Legal metadata in the form of PUB, ACA, RES categories provide necessary information on what is allowed.

For open data, Finland recommends Creative Commons Attribution 4.0. (JHS Recommendations 2014) for public sector information (PSI) or other similarly open data. However, there are legacy data and personal data that cannot be made fully open but can still be made accessible along the FAIR principles. For such data, FIN-CLARIN (a consortium partner of CLARIN) uses the CLARIN licences (ACA) for academic community-wide use and (RES) for individual research use. Similarly, personal data collected specifically for research purposes most often fall in the RES category because access to RES data needs a motivation for why access is necessary, which in turn can be evaluated against the research purpose specified for the personal data when these were collected.

The ACA category serves a similar purpose by having the data available in a physical archive or museum open to research, i.e. it is available with some effort to researchers. Such data may be personal interviews with the deceased people (technically making it non-personal data), possibly still mentioning some existing names or places that may identify a third person. It should be noted that any personal data such as YouTube videos may be provided with a CC licence by an individual publishing his or her data even if personal data can typically only be released after anonymization if publishing someone else's data. Note that the GDPR also allows archiving and using personal data unmodified, provided that protective measures guarantee that it is used only for scientific or historical research purposes (Art. 5 [1] [e]), which in CLARIN can be controlled by the RES category.

It could also be argued that research data are similar to other PSI since these are also created using taxpayers' money. The question is whether these two categories of data should be treated differently. The Directive on the Re-use of Public Sector Information sets forth the following four conditions for dissemination of PSI: (1) Re-usable for commercial or non-commercial purposes (Art. 3); (2) Charges shall be limited to the marginal costs. Any applicable conditions and the actual amount of those charges shall be pre-established and published (Art. 6, 7); (3) Re-use conditions shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition. Standard licences for the re-use of public sector documents are available in a digital format (Art. 8); (4) Prohibition of exclusive arrangements and non-discrimination (Art. 10, 11). The authors find the described PSI dissemination model suitable for open data as well.

Individual and organizational-level barriers to the implementation of the open data policy
Open data policy must be implemented in a coordinated manner at state, organizational and individual levels (Figure 2). Open data policy is not an established practice as yet. It is more often discussed than followed. The OECD (2015b: 46) correctly explains that several barriers to data sharing still remain. Some are of a technical nature, such as issues related to storage, the technical infrastructure to allow data sharing, interoperability and standards. Other types of barriers are related to the lack of an open data culture or the disincentives that researchers and scientists face with respect to the disclosure and sharing of datasets, especially related to research at the pre-publication stage.
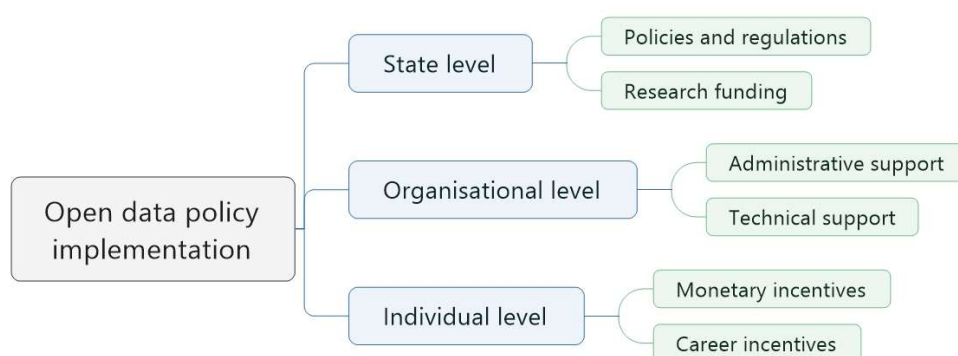

Figure 2: Open data policy implementation.

It is explained in policy documents that researchers are not motivated to share their data. For instance, the report commissioned by the Research Information Network (RIN) describes the following five researcher-level barriers (RIN 2008: 8):
1. Researchers wish to retain exclusive use of the data to use the data for as many publications as possible
2. Lack of career rewards for data creation and sharing
3. Lack of time and resources
4. Lack of experience and expertise in data management (e.g. metadata, legal issues); and
5. Fear of exploitation or inappropriate use of the data.

The practical measure should be the creation of incentives for researchers. Researchers should obtain academic credit and research funds (Reilly et al. 2011: 5; RIN 2008: 9; OECD 2015:

13) for the creation and dissemination of data. There is a need for a standardized citation system such as DataCite (see DataCite) and technical support.

Methodology

The empirical research for this article is drawn from four countries around the Baltic Sea: two traditional democratic market economies Sweden and Finland, and two post-transition countries Lithuania and Estonia. Historically, Finland belonged to the Swedish Kingdom from the twefth century to the beginning of the nineteenth century. In Estonia, the Swedish period was shorter, lasting from the sixteenth century until the beginning of the eighteenth century. Both Estonia and Finland have long-term academic traditions of Swedish origin – the University of Tartu (1632) and the University of Helsinki (1640) were founded in the Swedish Kingdom. These universities were founded based on the academic traditions of Uppsala University in Sweden, founded in 1477, and thus being the first university in Scandinavia.

Lithuania also has long-standing academic traditions, but of local origin – the Vilnius University was established in 1579. In the nineteenth and twentieth centuries Finland, Estonia and Lithuania belonged to the Russian Empire until First World War. Lithuania and Estonia were occupied by the Soviet Union for approximately 50 years and regained independence only at the beginning of the 1990s by embracing democracy and market economy. Therefore, the countries in the sample have different development trajectories, different academic traditions and different traditions of openness in society and data management at the state level. All these factors have contributed towards the culture of open data in research.

The case study approach has been selected to map and compare the challenges of transforming research data into open data in the sample countries. The cases were outlined by the representatives of the relevant countries in the authors' team and edited during the comparative study. The compilation of case studies is based on the documentary sources and regulations of the countries. To some extent, the authors as researchers also rely on their own experience and on four semi-structured interviews conducted with researchers and university managers at Uppsala University, Sweden. The research questions pursued in the case descriptions were in part formulated based on the theoretical overview above, while also reliant on the authors' previous research (Kelli et al. 2017).

The main topics of case studies are discussed in the following sub-sections and are further elaborated on by the authors' in the section on findings. First, the themes were mapped and reported country by country by the representatives of every country in question based on the information sources detailed above. In effect, the case studies in Finland and Estonia are mainly based upon the authors' own experiences from their participation in the CLARIN research infrastructure (CLARIN). The methodology applied can thus be said to be a version of Participatory Action Research (PAR) (Blake 2007), where the participating researchers outside and in parallel to the original research protocol have made separate observations about the pros and cons of the applied open data access policy. Those parallel observations have been used as research data to be analysed in this study. In the Swedish case, four semi-structured interviews were conducted, three face-to-face and one by telephone, each lasting 45 minutes to one hour. One interview was carried out with the principal researcher in the bigdata research programme Gender and Work at the department of history at Uppsala University (Gender and Work 2018), and one with the head of the research programme U-CAN (2018), which is one of the governmental strategic research programmes at the national level headed by Uppsala University. One interview (by telephone) was conducted with the

Vice-Rector of the faculty of Medicine and Pharmacy and one with the Pro-rector (Deputy Rector), both at Uppsala University, to explore the more general guiding principles for open data policies.

Second, the leading author compiled extracted material from reports and returned to co-authors for a critical check of context and feedback. The data and its presentation in the article have been jointly agreed upon.

The four case studies resulted in a number of common and general findings, but also some dissimilarities as presented and discussed in more detail below.

Findings

Open data as an integral part of open science and the open society framework

Open data policy should be conceptualized in a broader framework of the open science policy. The OECD (2015: 7) describes open science as 'efforts to make publications and the research data accessible in digital format with no or minimal restriction'. According to the European Commission (2016b: 33), open science moves 'towards sharing and using all available knowledge at an earlier stage in the research process'.

The open data policy has several stakeholders. First, we have researchers who are the key players. In case they do not support the open data policy, its implementation becomes very complicated. Second, we have research institutions and universities. Third, we have institutions allocating funds for research.

All four countries studied are currently taking steps to enhance the open data policy. In the next sections, we shall introduce the situation in those four countries, respectively. The situations in the four countries analysed can be described as follows:

Sweden

In Sweden, open access to public documents has a long tradition dating back to 1766 stipulating that almost all the documents of a state agency, except for working papers, are public and can be shared with all citizens. The Public Access to Information and Secrecy Act excludes health journals and state and business secrets. Universities in Sweden are considered state agencies. Chalmers Technical University in Gothenburg, Stockholm School of Economics and Jönköping University College are run by private foundations but have, under the contract with the State, agreed to follow the same publishing policy as the state universities. This means that almost all university documents, other than working papers or those subject to the Secrecy Act but including databases, are public.

Finland

Finland has similar legal traditions to Sweden, which means that many documents and datasets produced by state agencies are openly available. A notable difference compared with Sweden is that all Finnish universities became private foundations in 2009 and the traditional openness of public agencies no longer applies. However, several programmes have recently been introduced in Finland to promote freely available open data and open research data (see e.g. the Ministry of Finance of Finland on open data; the Ministry of Culture and Education of Finland), with a recent conclusion presented in March 2017 by the Prime Minister's Office (Government's Analysis, Assessment and Research Activities Report 2017). Moreover, requirements laid down by research funding agencies stating that research data should be made freely and openly available have effectively contributed towards the willingness of

researchers to make their research data openly available. Recently, the stringent data publishing requirements set by research agencies to obtain funding awards have become a significant incentive for publishing research data. However, under a thin veil of concern for the protection of personal data, many researchers may now develop a habit of restricting access to research data devoid of any personal data as well.

Estonia
The preliminary framework for the Estonian open data strategy has been prepared (see Kelli et al. 2017). Previously, the Estonian Research Council expert committee on open science set forth principles and recommendations for the development of the national open science policy in Estonia (see Estonian Research Council 2016). No official and general open data strategy has been adopted. The open data strategy will most likely be implemented using research funding conditions that require open dissemination of research data. A government-run open data portal exists in Estonia (Open Data Portal of Estonia).

Lithuania
The underlying national regulation for open data in Lithuania is the Law on Higher Education and Research (LHER). The law expressly provides for the obligation to make state-funded research results available:

> In order to ensure the quality of research conducted with funds of the state budget, the transparency of the use of funds of the state budget, to enhance scientific progress, the results of all research works carried out in higher education and research institutions with funds of the state budget must be announced publicly (through the Internet or some other way), to the extent this is in compliance with the legal acts regulating the protection of intellectual property, commercial or state and official secrets. (Art. 51)

Another relevant legal document (based on Art. 51 of LHER) is the Resolution Regarding the Approval of the Guidelines on Open Access to Scientific Publications and Data (the Guidelines), which was enacted by the Research Council of Lithuania on 29 February 2016. The Research Council of Lithuania is responsible for the coordination of the Open Access Policy at the national level. Chapter V of the Guidelines focuses on the open access to research data and limitation thereof. Many universities have enacted their local access policy regulations. Vilnius University, which is the most prominent university in Lithuania, has had such regulation in place since 2009.

Apart from the above-mentioned documents, there is no general national strategy on open data and research data. However, it has been suggested by the Research Council of Lithuania that such a document is inevitable in the future. The Lithuanian practice is also influenced and shaped by the European open access regulations and other instruments.

Individual and organizational barriers to dissemination of data
The individual country studies revealed the following barriers at individual and organizational levels and the possible means to address them.

Sweden
For Sweden, the main barrier hindering the implementation of open data policies is the lack of incentives for the following stakeholders:

- Individual researchers and registrar officers retrieving the data and feeding it into the database
- Moderators responsible for the quality assurance of the database
- University administrations allocating resources for this often time-consuming and labour-intensive work.

The reward system for the data producers, mainly in the humanities, must be improved. The data are not usually shared until the data producers have been able to conduct an initial analysis of the data and write a scientific paper based on this analysis.

Publication of data in open databases is commonly postponed until the data-producing Ph.D. students have had time to analyse and publish their results. A research project on gender and work carried out at Uppsala University exemplifies that senior researchers must also have career protection when working with big databases. One of the database moderators responsible for quality, reliability and timeliness of the data was not able to bolster his academic standing in a traditional way due to the heavy workload. He was in a worse position compared to his peers when later competing for new grants and university positions.

Finland
The Finnish study also found it problematic that there is no separate scientific reward for publishing research data. Data as such are not counted in the official publication record of a researcher, which means that in practice the data are made publicly available as part of a regular publication or a piece of software. Like in Sweden, the Finnish universities are encouraged by the government to provide open access to research data, especially the programme on Open Science in Finland (the Ministry of Culture and Education of Finland). Money has been spent on opening databases collected by public administration agencies. An open access data management plan in compliance with the FAIR principles for research data is required for receiving public research funding. It has had a noticeable effect on the opening of research data and the attitude towards making datasets available as part of the research process. There is still a need to include datasets as one of the publication types acknowledged in the official research output database. Publication of data increases the reliability and reproducibility of research, which, in the long run, also promotes citation numbers affecting university rankings.

Estonia
According to the Estonian experience, there are no tangible incentives for researchers to implement the open data policy. Researchers do not have much to gain if they invest time and resources in making their data open. First, it takes a considerable amount of time and technical assistance (making data compatible with the required standards, adding metadata). Second, making research data available and reusable does not result in any significant reward. Data citation is not a widespread practice, and the implementation of the open data policy is not taken into account in the tenure decision process (at least it is not emphasized in the regulations of Estonian universities).

Lithuania
Currently, no real administrative or legal barriers could be identified. The mentality and habits of the Lithuanian research community constitute the principal barrier. Open access mostly depends on the individual decisions of researchers, and they are not used to sharing their research data. Sometimes there is a fear that archiving of research data would bring about additional costs and this, in turn, can adversely affect the overall budget of the research.

Perceived in this way, storing and opening up the research data is associated with a financial barrier that is best avoided.

There are no clear rewards for researchers to make their research data open. No data citation system is fully operational in Lithuania. As a rule, data are cited in connection with the publication.

The interaction between open data and entrepreneurial university policies

Open data policies require research institutions to share research data without restrictions. Innovation policies, however, introduce the concept of an entrepreneurial university, which means that the university is involved in commercial knowledge transfer (see Kelli et al. 2014, Kelli et al. 2013; Mets 2010). The central legal mechanisms for knowledge transfer are licensing, assignment of intellectual property (including know-how protected as a trade secret), spin-off creation, consultancy and so forth (for further discussion on knowledge transfer see EC 2009). This means that research results need to be controlled, which could easily conflict with open data policies.

It should also be emphasized that open data are not the ultimate and only goal of academia. While open data policy facilitates further research, academia needs to support economic development and transformation into a knowledge-based economy as well. These policies can and should be amalgamated.

The concept of the entrepreneurial university contributing towards a knowledge economy can be interpreted differently. One approach emphasizes that universities should merely act as commercial entities, aiming to maximize profit when possible from new inventions and new venture creations. An entrepreneurial university can also be based on the view that universities should actively interact with its surrounding society, promoting open innovation and knowledge dissemination when suitable, but also exhibit awareness and readiness to be able to encourage commercialization through licensing deals and new venture creation. As to the first interpretation, profit maximization for the university becomes central while according to the second interpretation, the dissemination of knowledge should be maximized to benefit the society at large while the goal of the universities' commercialization efforts should not be profit maximization for the university alone. The level of tension between the policy of open data sharing and that of an entrepreneurial university depends on which interpretation of the concept the country and university favour.

Commercially useful datasets are often of an accruing nature, whereas scientific research results most often are independent of last-minute updates to a dataset. Limited access to recent updates is usually enough for research, provided that the data become fully available within a limited time frame. It is also well known that technology without knowing how to apply it is of limited use. Therefore, selling consultancy and bleeding edge data are still possible while furthering open science.

Most likely only commercially valuable knowledge is subject to commercialization efforts by the university or its researchers. Research data that do not have a direct commercial application could be shared. These data, however, can be used as a basis for the development of commercial applications. To sum up, when there is commercial potential, and no enforceable obligations exist to disseminate the data, these are used for academic entrepreneurship. The respective country studies revealed the following aspects.

Sweden

According to the Swedish approach, entrepreneurial universities do not solely focus on their profit maximization. Instead, they focus on offering the society their maximum research utilization and dissemination of knowledge. Therefore, no apparent contradiction between the interests of disseminating open data and the policy to commercialize research findings has been identified as yet in the discussion.

Finland

According to the Finnish case study, open data and entrepreneurial university policies can coexist in the open science policy, i.e. companies and researchers alike can contribute datasets for non-commercial or pre-competitive purposes. Commercially useful datasets are often of an accruing nature, providing access to fresh data from which to mine new information, whereas scientific research results are by their nature independent of last-minute updates to a dataset as the results are intended to be reproducible and verifiable. Limited access to recent updates is usually sufficient for research, provided that the data become fully available within a limited time frame.

Estonia

The interaction among open data and entrepreneurial university policies has not been fully acknowledged in Estonia and dissemination of research data is not an established practice as yet. Both policies are relevant, but they are not integrated into a holistic approach. This can sometimes yield undesired results. Research results with commercial potential should be transferred to the economy. Research data that are commercially inapplicable should be made available.

Lithuania

The interaction between open data and entrepreneurial university policies has not been fully acknowledged in Lithuania either. The Lithuanian case study revealed that until now universities have been more concerned with open access and seem indifferent to entrepreneurial issues. Therefore, no real clash between open data policies and the entrepreneurial university has arisen so far.

Discussion

Lack of incentives and citation rewards

Our study has produced similar key findings for all four countries. There is a need for a substantial reward for data producers (researchers), who are the first ones to analyse data and publish results. It seems to be a common practice that the data in open databases are not shared until the data researchers have prepared initial publications based on the data. In general, data citation is part of the ethical code of research no different from any other references listed in a publication. However, data references do not have the same impact as traditional citations. This is especially true for the registrars and moderators of the database ('the everyday workers').

There is a need to acknowledge data citations as academic merit to make the publishing of datasets more appealing for researchers. Currently, there is a category for registering the software in a national research publication record but nothing for datasets. The current procedure allows for publication of a paper describing the creation of a dataset along with particular cases of use while asking everyone using the dataset to refer to the article. It becomes slightly problematic if the reuse of the data is for a different purpose, in which case the real reference to a dataset should be made through an independent description of the

dataset. Such a data manual is usually too detailed for a scientific publication and is often best published as a technical report, which is not counted as a scientific publication. Sometimes this leads to a situation where there is insufficient funding left for publishing the dataset at the end of a project. Since at least some recognition is received for creating software, projects often create a database interface, but typically there is no maintenance of the database and its interface after the completion of the project. If the data can be made publicly available, a cheap way to comply with the publication requirements is to publish the data on GitHub, where, due to the current data deluge, the data in most cases will stay "hidden in plain sight" until the researcher has published a scientific paper advertising the data.

The sharing of data can increase the scientific community's awareness of the scientist and his or her citation rankings. The data-sharing requirement could be instituted as a precondition for project funding. Some financial measures should also be introduced. An additional part of the budget could be earmarked specifically for the preparation and storage of research data. This would be independent of the central budget and would not affect the researchers' salaries.

Potential harm
Disseminating data could be detrimental to researchers for several reasons. First, possible errors might become evident. Second, researchers do not have a monopoly on the data and cannot publish several articles based on the same data. Therefore, it is crucial to motivate researchers by providing actual incentives. Formal requirements (e.g. researchers are obliged to make data available if they qualify for funding) do not contribute meaningfully to the publication of valuable data. Attaching value to data citations could be a way forward. In case researchers are not incentivized to share data, the open data policy cannot be implemented successfully. Formal requirements are just not enough. They would ensure formal compliance, but the disseminated data would not necessarily be of high quality. Data sharing and data citations should be considered in tenure decisions. Data citations are not equal to citations in publications.

Technical compatibility
Technical compatibility can also pose problems. Indeed, institutional databases should comply with specific (uniform) technical standards. Otherwise, it can be rather challenging to integrate them with other national/regional open access databases and for users to access and use them.

Databases as part of research infrastructure in need of allocated resources
Universities must prioritize the implementation of the open data policy and create incentives for researchers. Currently, universities are not actively encouraged to implement the open data policy. One significant difference between scientific disciplines is that at least the humanities have no history of being dependent on this kind of infrastructure, and so there are no research assistants and laboratory assistants as in natural sciences, medicine and pharmacy. It is therefore essential to develop budgets that will enable employment of senior staff in charge of the database quality assurance work and less qualified assistants specializing in registering the data and feeding the database with raw material.

The future of research, both in the humanities and in social sciences, will probably become more contingent on this kind of new infrastructure, which must be integral to planning and financial management work carried out at faculty and university administration levels. It is time for the university's administration to realize that great databases are a new kind of infrastructure that will become as crucial for high-level research in humanities and social

sciences as laboratories are today for medicine, pharmacy and natural sciences. They must get this message through to the governments and research grant foundations to ensure that financial support for the building and maintenance of these databases is available. Databases are expensive, especially sizeable ones with high maintenance costs. To be able to defend the expenses in applications to secure the support needed, it must be shown that as many qualified researchers as possible can use the infrastructure. It must be openly available.

The Swedish Research Council stipulates that active support is needed from at least three different Swedish universities to be granted financial aid for this kind of infrastructure. This also goes for laboratories and equipment within life sciences and materials science and in other research areas. Moreover, a university hosting national or international infrastructure enjoys greater prestige and various other advantages. These facilities function as hubs for the research community. They probably also affect the rankings of universities, in some cases directly, in other cases more indirectly, since first-class university facilities attract collaboration, high-profiled peers and so forth, resulting in more high-scoring publications from its researchers.

Our study revealed that open data and an entrepreneurial university are compatible, although it ought to be carefully considered when and what knowledge and findings are shared. The bulk of research data can be made available with relative ease since it has no direct commercial application. Commercial interests can be protected by delaying the dissemination of data or keeping some specific know-how secret.

Conclusions

The successful implementation of the open data policy calls for the introduction of measures at state, organizational and individual levels.

The state-level measures concern the integration of different policies and the development and implementation of relevant regulations. Open data policies are not the only policies. Academic entrepreneurship as part of an innovation policy should be integrated with the open data policy. These policies can be compatible with each other. Most research data do not generate commercial value, or they clearly remain beyond the scope of commercial exploitation and consequently can be made available without access and use limitations. Commercially valuable data can be transferred to the economy. Thus, it can be concluded that there is no real contradiction between the implementation of entrepreneurial university and open data policies.

Personal data protection complicates the process of transformation of research data into open data. However, its objective is not just to interfere with openness, but to protect individuals and their privacy along the way. These concepts are not mutually exclusive. There are plenty of research data that do not include personal data (personal data can also be anonymized) and can thus be disseminated freely.

Intellectual property law also has mechanisms to support open data policies. For instance, copyright does not protect data and facts. This means that they can be freely disseminated.

It is debatable whether data standardization and dissemination models are state – or organizational-level measures. It often takes place in international settings involving states and organizations. It is clear that to reap the benefits of open data, these must be standardized. However, even if data are standardized, it is not enough to avoid misinterpretation since data in the field of social sciences and humanities are context specific.

There is a need to establish dissemination models, which subsequently requires a suitable contractual framework. The model should limit the liability for the accuracy of data and ensure the reusability of data.

Organizational-level measures are related to administrative and technical support. It is not realistic that researchers themselves develop and maintain repositories containing open data and so forth. Databases, especially big databases, must be considered research infrastructure much the same way as laboratories. University administrations, research foundations and national governments must allocate enough resources both for physically housing the databases and engaging sufficient technical support, but also ensure proper funding so that the database managers can employ data collectors and implement an adequate quality assurance programme. It needs both the competence of well-qualified research staff and personnel with lower qualifications just as we see in biomedical and technical laboratories.

Last but not the least, there are the individual-level measures. If researchers do not support open data policies, the latter cannot possibly thrive. Compliance with formal requirements is not sufficient. Researchers should be able to benefit from clear career and monetary incentives. Data creation and dissemination should be considered research and have a direct bearing on career decisions. Data citation systems should be integrated into research evaluation systems. There is also a compelling need for technical and administrative support. The quality of research data warrants further consideration; in this respect, perhaps data citations could serve as indicators of the quality of data.

References

Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPS Agreement) (1995), 15 April, https://www.wto.org/english/docs_e/legal_e/31bis_trips_01_e.htm. Accessed 22 July 2018.

Article 29 Working Party (WP29) (2013), 'Opinion 03/2013 on purpose limitation', 2 April, http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf. Accessed 7 November 2018.

_____ (2007) 'Opinion 4/2007 on the concept of personal data', 20 June, http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf. Accessed 23 July 2018.

Autoriõiguse seadus (Copyright Act) (1992), 'Entry into force', 12 December, https://www.riigiteataja.ee/en/eli/519062017005/consolide. Accessed 22 July 2018.

Ball, A. (2014), 'How to license research data', DCC How-to Guides, Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides/license-research-data. Accessed 22 July 2018.

Blake, M. (2007), 'Formality and friendship: research ethics review and participation action research', *ACME: An International E-Journal for Critical Geographics*, 6:3, pp. 411–21.

The British Horseracing Board Ltd and Others vs. William Hill Organization Ltd. (2009), Case C-203/02. 9 November, http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1512926505978&uri=CELEX:62002CJ0203. Accessed 12 November 2018.

CLARIN (Common Language Resources and Technology Infrastructure), 'CLARIN – European research infrastructure for language resources and technology', https://www.clarin.eu/. Accessed 22 July 2018.

CLARIN licensing framework, https://www.clarin.eu/content/clarin-licensing-framework. Accessed 7 November 2018.

Creative Commons. Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/legalcode. Accessed 7 November 2018.

Criminal Proceedings against Bodil Lindqvist (2003), Case C-101/01, 6 November, http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1521039149443&uri=CELEX:62001CJ0101. Accessed 7 November 2018.

DataCite, 'Our mission', https://www.datacite.org/mission.html. Accessed 24 July 2018.

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database directive) (1996), OJ L 77, 27 March, pp. 20–28, http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31996L0009&rid=1. Accessed 22 July 2018.

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-use of Public Sector Information. Amended by Directive 2013/37/EU of the European Parliament and of the Council (Text with EEA relevance of 26 June 2013) (Directive on the Re-use of Public Sector Information), http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003L0098-20130717. Accessed 22 July 2018.

Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) against their Unlawful Acquisition, Use and Disclosure (Text with EEA Relevance) (Trade Secrets Directive) (2016), OJ L 157, 15 June, pp. 1–18, http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1513091051135&uri=CELEX:32016L0943. Accessed 22 July 2018.

Drexl, J. (2016), 'Designing competitive markets for industrial data – between propertisation and access', Max Planck Institute for Innovation & Competition Research Paper, no. 16-13, pp. 1–70, https://ssrn.com/abstract=2862975. Accessed 22 July 2018.

Drexl, J., Hilty, R., Desaunettes, L., Greiner, F., Kim, D. Richter, H., Surblyte, G. and Wiedemann, K. (2016), 'Data ownership and access to data – position statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European

Debate', Max Planck Institute for Innovation & Competition Research Paper, no. 16-10, pp. 1–12, https://ssrn.com/abstract=2833165. Accessed 22 July 2018.

Estonian Research Council (2016), 'Avatud teadus Eestis. Eesti Teadusagentuuri avatud teaduse ekspertkomisjoni põhimõtted ja soovitused riikliku poliitika kujundamiseks' ('Open Science in Estonia. The principles and recommendations on the development of national open science policy by the expert committee of the Estonian Research Council'), http://www.etag.ee/wp-content/uploads/2016/03/AT_soovitused_20160229.pdf. Accessed 22 July 2018.

European Commission (EC) (2009), Metrics for Knowledge Transfer from Public Research Organisations in Europe Report from the European Commission's Expert Group on Knowledge Transfer Metrics, https://era.gv.at/object/document/332. Accessed 22 July 2018.

_____ (2016a), 'H2020 programme. Guidelines on Fair Data Management in Horizon 2020', https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accessed 22 July 2018.

_____ (2016b), 'Open innovation, open science, open to the world – a vision for Europe', https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe. Accessed 24 July 2018.

_____ (2016c), 'Proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market', Brussels, 14 September, COM(2016) 593 final 2016/0280 (COD), http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016PC0593&from=EN. Accessed 22 July 2018.

_____ (2017a), 'H2020 programme. Guidelines to the rules on open access to scientific publications and open access to research data in Horizon 2020', Version 3.2, 21 March, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Accessed 23 July 2018.

_____ (2017b), 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Building a European Data Economy"', Brussels, 10 January COM(2017) 9 final. {SWD(2017) 2 final}, http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1505218931648&uri=CELEX:52017DC0009. Accessed 22 July 2018. FIN-CLARIN – Integrating the language resources in Finland, https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/KielipankkiFrontpage _ Accessed 22 July 2018.

Finnish National Interoperability Framework Observatory (2016), 'Finnish open data programme eliminates obstacles to data reuse (API, JulkICTLab, ODIP, CKAN, JUHTA)', https://joinup.ec.europa.eu/document/finnish-open-data-programme-eliminates-obstacles-data-reuse-api-julkictlab-odip-ckan-juhta. Accessed 23 July 2018.

FORCE11 (2017), 'Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0', https://www.force11.org/fairprinciples. Accessed 22 July 2018.

Gender and Work (2018), http://gaw.hist.uu.se. Accessed 30 October 2018.

Government's Analysis, Assessment and Research Activities Report (2017), *Data utilisation Boosts Innovation and Growth*, http://vnk.fi/en/article/-/asset_publisher/raportti-datan-hyodyntaminen-lisaa-innovaatioita-ja-kasvua. Accessed 22 July 2018.

Hermoso, R., Cardoso, H. L., Fasli, M. (2015), 'From roles to standards: A dynamic maintenance approach using incentives', *Information Systems Frontiers*, 17:4, pp. 763–78.

Hugenholtz, P. B. (2017), 'Data property: Unwelcome guest in the House of IP', https://pure.uva.nl/ws/files/16856245/Data_property_Muenster.pdf. Accessed 22 July 2018.

JHS Recommendations (2014), 'JHS 189 The license for using open data', http://jhs-suositukset.netum.fi/web/guest/jhs/recommendations/189/full . Accessed 12 November 2018.

Kelli, A., Mets, T. and Jonsson, L. (2014), 'Management of intellectual property rights at academia: The Estonian and Swedish perspectives', *International Journal of Technology Management & Sustainable Development*, 13:3, pp. 219−36.

Kelli, A., Vider, K. and Lindén, K. (2015), 'The regulatory and contractual framework as an integral part of the CLARIN infrastructure', in Koenraad De Smedt (ed.), *Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland*, Linköping University Electronic Press, Linköping, pp. 13–24, http://www.ep.liu.se/ecp/article.asp?issue=123&article=002. Accessed 22 July 2018.

Kelli, A., Mets, T., Vider, K. and Kull, I. (2017), *Avatud teadus Eestis ja Euroopas: Õiguslik ja majanduslik lähenemine* (*Open Science in Estonia and Europe: Legal and Economic Perspectives*), Report commissioned by the Estonian Research Council and supported by the European Regional Development Fund through the Programme for Addressing Socio-economic Challenges of Sectoral R&D, Tartu, http://www.etag.ee/wp-content/uploads/2018/02/Avatud_teadus_Eestis_ja_Euroopas_%C3%B5iguslik_ja_majandusl ik_l%C3%A4henemine_T%C3%9C.pdf. Accessed 22 July 2018.

Kelli, A., Mets, T., Jonsson, L., Pisuke, H. and Adamsoo, R. (2013), 'The changing approach in academia-industry collaboration: From profit orientation to innovation support'*, Trames: Journal of the Humanities and Social Sciences*, 17:3, pp. 215−41.

Kelli, A., Lindén, K., Vider, K., Labropoulou, P., Ketzan, E., Kamocki, P. and Straňák, P. (2018), 'Implementation of an open science policy in the context of management of CLARIN language resources: A need for changes?', in Maciej Piasecki (ed.), *Selected Papers from the Clarin Annual Conference 2017*, Linköping University Electronic Press, Linköping, pp. 102−11, http://www.ep.liu.se/ecp/147/009/ecp17147009.pdf. Accessed 23 July 2018.

Kelli, A., Lindén, K., Vider, K., Kamocki, P., Birštonas, R., Calamai, S., Kolletzek, C., Labropoulou, P. and Gavriilidou, M. (2018), 'Processing personal data without the consent of the data subject for the development and use of language resources', in Inguna Skadin and Maria Eskevich (eds), *CLARIN Annual Conference 2018 Proceedings: CLARIN Annual Conference 2018, 8–10 October 2018 Pisa, Italy*, CLARIN, pp. 43−48, https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf. Accessed 8 November 2018.

Klavan, J., Tavast, A. and Kelli, A. (2018), 'The legal aspects of using data from linguistic experiments for creating language resources', *Frontiers in Artificial Intelligence and Applications*, 307, pp. 71–78, http://ebooks.iospress.nl/volumearticle/50306, Accessed 8 November 2018.

Lietuvos mokslo tarybos 2016 m. vasario 29 d. nutarimas dėl Atvirosios prieigos prie mokslo publikacijų ir duomenų gairių patvirtinimo (Resolution Regarding the Approval of the Guidelines on Open Access to Scientific Publications and Data of Research Council of Lithuania (2016), 29 February, http://www.lmt.lt/data/public/uploads/2016/09/eng_-atvira-prieiga-_-galutinis.pdf. Accessed 22 July 2018.

Lietuvos Respublikos mokslo ir studijų įstatymas (Law on Higher Education and Research) (2009), 'Entry into force', 12 May, https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/136176908a8711e4be90fee65c21bf37?jfwid=ck9gyalf7. Accessed 22 July 2018.

Mets, T. (2010), 'Entrepreneurial business model for classical research university', *Inzinerine Ekonomika/Engineering Economics*, 21:1, pp. 80–89.

Mets, T. and Torokoff, M. (2007), 'Patterns of learning organisation in Estonian companies', *TRAMES. A Journal of the Humanities and Social Sciences*, 11:2, pp. 139–54.

Ministry of Culture and Education of Finland, 'Open science in Finland', https://openscience.fi/. Accessed 22 July 2018.

Ministry of Finance of Finland, 'Open data', http://vm.fi/en/opendata. Accessed 22 July 2018.

_____ (2015), 'Open data policy in the Finnish Government', http://vm.fi/documents/10623/360816/Avoimen+tiedon+tavoitteet+2015-2020/c7e9c09c-c492-4f04-ac52-2449b0c4973b. Accessed 23 July 2018.

Oksanen, V., Linden, K. and Westerlund, H. (2010), 'Laundry symbols and license management: Practical considerations for the distribution of LRs based on experiences from CLARIN', *Proceedings of LREC 2010: Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management*, https://helda.helsinki.fi/handle/10138/29359. Accessed 22 July 2018.

Oostveen, M. (2016), 'Identifiability and the applicability of data protection to big data', *International Data Privacy Law*, 6:4, pp. 299–309.

Open Data Portal of Estonia, 'Frequently asked questions', https://opendata.riik.ee/en/faq/. Accessed 24 July 2018.

Open Knowledge International, 'The open definition', http://opendefinition.org/. Accessed 8 November 2018.

Organisation for Economic Co-operation and Development (OECD) (2015a), 'Making open science a reality', *OECD Science, Technology and Industry Policy Papers*, vol. 25, Paris: OECD Publishing, http://dx.doi.org/10.1787/5jrs2f963zs1-en. Accessed 23 July 2018.

_____ (2015b), *Data-Driven Innovation: Big Data for Growth and Well-Being*, Paris: OECD Publishing http://dx.doi.org/10.1787/9789264229358-en. Accessed 22 July 2018.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN. Accessed 22 July 2018.

Reilly, S., Schallier, W., Schrimpf, S., Smit, E. and Wilkinson, M. (2011), 'Report on integration of data and publications', http://libereurope.eu/wp-content/uploads/ODE-ReportOnIntegrationOfDataAndPublication.pdf. Accessed 22 July 2018.

Research Information Network (RIN) (2008), *To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs*, http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf. Accessed 22 July 2018.

Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G. J., Felzer, B. S., McIntyre, N. E., Soranno, P. A., Vanderbilt K. L. and Weathers, K. C. (2014), 'Completing the data life cycle: using information management in macrosystems ecology research', *Frontiers in Ecology and the Environment*, 12:1, pp. 24–30.

Salgé, F. (1999), 'National and international data standards', in P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (eds), *Geographical Information Systems: Principles, Techniques, Management and Applications*, New York: Wiley, pp. 693–706.

Spindler, G. and Schmechel, P. (2016), 'Personal data and encryption in the European general data protection regulation', *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 7 (2), pp. 163–77.

Tietosuojavaltuutettu vs. Satakunnan Markkinapörssi Oy and Satamedia Oy (2008), Case C-73/07, 16 December, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62007CA0073&qid=1536154290371&from=EN. Accessed 7 November 2018.

U-CAN (2018), http://u-can.uu.se. Accessed 30 October 2018.

WIPO Copyright Treaty (1996), 20 December, Geneva, http://www.wipo.int/treaties/en/text.jsp?file_id=295166#P59_6206. Accessed 22 July 2018.

Yaşlıoğlu, M. M.; Şap, Ö. and Toplu, D. (2014), 'An investigation of the characteristics of learning organizations in Turkish companies: Scale validation', *Procedia-Social and Behavioral Sciences*, 150, pp. 726–34.

Contributor details
Contact:
E-mail: aleksei.kelli@ut.ee
https://orcid.org/0000-0002-0856-1000

E-mail: tonis.mets@ut.ee
https://orcid.org/0000-0003-3972-5204
E-mail: lars.jonsson@holding.uu.se
https://orcid.org/0000-0003-2047-656X
E-mail: krister.linden@helsinki.fi
https://orcid.org/0000-0003-2337-303X
E-mail: kadri.vider@ut.ee
https://orcid.org/
E-mail: ramunas.birstonas@tf.vu.lt
https://orcid.org/
E-mail: age.varv@ut.ee
https://orcid.org/

Note

---

[1] The article partly draws on and extends the previously completed research (see Kelli et al. 2017).