Summer 6-15-2016

A COMMONS PERSPECTIVE ON GENETIC DATA GOVERNANCE: THE CASE OF BRCA DATA

Polyxeni Vassilakopoulou
University of Oslo, xvasil@ifi.uio.no

Espen Skorve
University of Oslo, espesko@ifi.uio.no

Margunn Aanestad
University of Oslo, margunn@ifi.uio.no

A COMMONS PERSPECTIVE ON GENETIC DATA GOVERNANCE: THE CASE OF BRCA DATA

Research

Vassilakopoulou, Polyxeni, University of Oslo, Oslo, Norway, xvasil@ifi.uio.no

Skorve, Espen, University of Oslo, Oslo, Norway, espesko@ifi.uio.no

Aanestad, Margunn, University of Oslo, Oslo, Norway, margunn@ifi.uio.no

Abstract

The advances that led to cost-efficient genome sequencing cannot bring significant benefits in medical care and public health unless genetic data interpretation is facilitated. Essentially, the mainstreaming of genetics within healthcare and the realisation of their transformative potential is increasingly becoming an information systems' problem of providing adequate decision support and enabling data sharing. However, providing this support is far from straightforward as data governance in the field has been the subject of contestations between actors committed to finding ways for pooling all available data and actors advocating restrictive approaches. In this paper we analyse the tensions around the governance of genetic data related to two genes which influence humans' susceptibility to breast and ovarian cancer (BRCA1 and BRCA2). Our research adopts a process approach that follows the evolution over time and draws theoretically from the Commons literature. Our findings show how "exclusion" and "subtractability" are not external data properties but sociotechnical achievements that are still in the making. Hence, aiming to handle genetic data as indisputably common goods or simply setting-up infrastructures with the expectation that they will be used, is futile. Instead, an effort that entails combining new technological potentialities with regulatory provisions is required.

Keywords: Genetic data, Commons, Governance, Institutional Work

1 Introduction

Advances in sequencing technologies, medical genetics, pharmacogenetics and bioinformatics are already inducing changes in healthcare affecting diagnostics, treatment planning and prevention (Battista et al. 2012; Bennett et al. 2010; Haga et al. 2012; Skirton et al. 2005; Snyderman 2012). More importantly, they are raising expectations for radical transformations in the near future. For instance, the Council of Europe's Committee of Ministers issued a recommendation where it was stated that : "the development of genetics in health care services has a major impact on the organisation of health care, leading to shifting from curative to preventive services, from in-patient to out-patient treatment, from specialised genetic services to genetics as an integral part of general health services" and that "governments should consider investing in the development of common platforms of genetic knowledge, where collaboration between genetic services, research organisations, patient support groups and the public can take place" (Council of Europe 2010). We are already seeing genetics embedded in some aspects of healthcare but there are significant challenges that need to be addressed before we experience the radical reconfigurations foreseen.

The greatest challenge for extending the use of genetic technologies in healthcare relates to the interpretation of genetic data (Quintáns et al. 2014; Sboner et al. 2011; Slade et al. 2015; Vrijenhoek et al. 2015). The rapid decrease in run times and cost of genetic data generation has not been matched by advances to speed up genetic data interpretation and conversion to actionable information, hence, interpretation is becoming a bottleneck. The pressing need to facilitate interpretation is instigating the interest on "common platforms for genetic knowledge". Essentially, the mainstreaming of genetics within healthcare and the realisation of their transformative potential is increasingly becoming an information systems problem. The progress in genetic technologies needs to be complemented by digital health initiatives for the development of information sharing platforms and decision support tools to accelerate interpretation. However, this is far from straightforward as data governance in the field has been the subject of controversies and contestations. The struggles around data governance can be summarised as tensions between those committed to finding ways for pooling together and collectively governing processed genetic data and those advocating restrictive approaches. These struggles are impeding advancement and need to be better understood. Providing information systems' support for genetic data interpretation is not solely a technical problem but rather a convoluted sociotechnical problem and has to be addressed as such in order to move forward.

In this paper we investigate the ongoing data governance developments in the genetics field by focusing on one specific type: genetic data related to two genes which influence humans' susceptibility to breast and ovarian cancer (*BRCA1* and *BRCA2*). This is a significant domain within genetics as "a

large proportion of the work in genetic services is the management of familial breast and ovarian cancer, and this clinical area exemplifies both the opportunities and challenges to increasing access to gene testing" (Slade et al. 2015). The analysis we present spans the two decades that have elapsed since the identification of the two specific genes. During this period, scientists around the world have analysed the genetic material of thousands of individuals and contributed to the accumulation of rich datasets, valuable for better understanding cancer biology and for diagnosing cancer susceptibility. The governance of these rich data sets has been challenged by tensions between those that support and those that oppose their objectification as common goods. Our paper represents an attempt to tease out a fruitful approach to the analysis of this kind of large-scale, technology intensive information arenas where distributed actors play out. To do so we are leveraging concepts from new-institutionalism and specifically from theory on commons governance (Ostrom 1990, 2011) which can offer a way to unpack experiences and gain insight valuable for policy makers, technology providers, health professionals and the academia when pursuing the much needed common platforms of genetic knowledge.

The remainder of the paper is structured as follows. First, we lay out the theoretical background and we describe the method used to collect empirical material. Subsequently, we provide a brief overview of BRCA genetic testing. Then, we present our case analysis. Finally, we conclude by discussing insights from our analysis, pointing also to the limitations of our work and to possible directions for further research.

2 Theoretical Background and Method

2.1 Theoretical Background

Our interest on the ongoing tensions between actors committed to finding ways for pooling together and collectively governing genetic data and those advocating restrictive approaches has oriented our attention to institutionalisation phenomena, i.e. "the processes by which structures, including schemas, rules, norms, and routines, become established as authoritative guidelines for social behaviour" (Scott 2005). Lawrence and Suddaby developed the concept of "institutional work" to describe "the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions" (Lawrence and Suddaby 2006; Lawrence et al. 2011) and to explore the distributed nature of institutionalisation. The concept helps us to follow a process approach, examining how "things change over time" (Pettigrew 1997) by looking at action types, rather than action accomplishment (Lawrence et al. 2009).

To support the conceptualisation of complex institutional situations Ostrom developed a comprehensive framework for Institutional Analysis and Development (IAD) that directs attention to the particu-

larities of the action situation that is to be analysed (Ostrom 2011). This framework builds upon a typology of goods that was introduced during the late 70s by Ostrom and Ostrom to disambiguate between property regimes and the nature of goods. For some time, economists struggled with classifying goods as either private or public. In their typology, the Ostroms clearly identified that there are more than two types of goods (Ostrom and Ostrom 1977). They used two attributes from the political economy literature that help identify four broad classes of goods (Table 1). The first attribute is that the benefits consumed by someone may subtract to a greater or lesser extend from the benefits available to others. The second attribute relates to how difficult it is to exclude individuals from using the flow of benefits (Hess and Ostrom 2003).

		Subtractability	
		Low	High
Exclusion	Easy	Club Goods	Private Goods
	Difficult or Infeasible	Public Goods	Common Pool Resources

Table 1. Typology of goods based on exclusion potential and subtractability. Adapted from Ostrom & Ostrom 1977 and Hess & Ostrom, 2003.

The critical factor in this approach is to begin with the nature of the goods involved (Ostrom and Ostrom 1977). The category of goods for which it is difficult or infeasible to implement exclusion measures is characterised as "commons". The "commons" can be governed as public goods or as common pool resources. The classification of goods in one of the four categories does not automatically lead to an optimal governance regime but it contours the action arena. For example, in their paper on information as common pool resource, Hess and Ostrom clearly state that there is not a one-to-one association between common-pool resources and common-property regimes: "common-pool resources may be owned by national, regional, or local governments, by communal groups, by private individuals or corporations, or used as open-access resources by whomever can gain access. Each regime has different sets of advantages and disadvantages" (Hess and Ostrom 2003).

Essentially, the Ostroms coined two characteristics that are pivotal for explaining a specific regime for goods' governance. What we have found intriguing in their conceptualisation is that the nature of goods is considered in the IAD framework to be part of the "external variables" (i.e. exogenous factors not to be shaped within the arena of action). While this might hold for biophysical goods (like forests, bread, sunshine, electricity) it is not always the case with non-material, manmade goods where the exclusion and subtractability characteristics are not necessarily to be taken as given as they can be fabricated and technologically contingent. We have explored the making of exclusion and subtractability characteristics for BRCA datasets as institutional work.

2.2 Method

The impetus for our study comes from our involvement in a research and development project within genetics. This is a collaborative project between the Department of Medical Genetics in the X University Hospital and the University of X. The aim of the project is to develop a secure IT platform that facilitates distributed collaboration and access to a high-performance analysis and storage facility. As one of the research activities in this project, we have conducted interviews and observations of how molecular biologists and other specialists perform their work. During our observations we were struck with the role of external databases and tools. We started further investigations into the international 'ecology' of information resources via secondary data collection (documents, reports, research papers etc.). Our data collection focused on the resources that support BRCA data interpretation, and in particular the dynamics that have precipitated around the Breast Information Core, a globally shared database established in 1995. We have downloaded the information in the database, analysed the patterns of submissions and examined the content of the records. Starting from BIC we expanded our review to other related initiatives and collated data to produce an account of the information landscape evolution. In summary, our paper is based on data collected using a combination of fieldwork, documents' analysis and hands-on inspection of medical genetics' data repositories (Table 2).

Source	Description	Topics Covered
Interviews	12 semi-structured interviews with scientists engaged in medical genetics (biologists, bio-informaticians, medical doctors).	Current work practices, information exchanges with other scientists that are remotely located, usage of tools, insights on ongoing developments in the field related to data sharing.
Data repositories inspection	Inspected the BIC and ClinVar repositories	For BIC: analysed the patterns of submissions and examined the content of the records. For ClinVar: retrieved BRCA variant data and examined the information consolidation mechanisms employed.
Document analysis	Journal Papers, Media Publications, Scientific Guidelines, Nomenclature Documents, Legal Proceedings, Project Reports, Presentations prepared for various audiences.	Key events that mark the evolution of the overall information infrastructure, argumentation and public debates for controversial issues.

Table 2. Data Sources.

We have analysed the data collected by focusing on the evolution of events related to BRCA data governance following an interpretive approach (Klein and Myers 1999; Walsham 1993). Being theoretically informed by the streams of literature on institutional work and the governance of the commons, we have developed the following research question which directed our data analysis:

"How do actors within the field of medical genetics for hereditary breast cancer shape the exclusion and subtractability characteristics of BRCA datasets?".

3 Empirical Background: BRCA testing

In 1994, a gene (*BRCA1*) which influences susceptibility to breast and ovarian cancer was identified (Miki et al. 1994) and soon after that, a second gene related to the same diseases (*BRCA2*) was also identified (Wooster et al. 1995). The identification of those two genes made genetic testing relevant for a much wider share of population than in the past: it was now possible to perform genetic testing not only for diseases that are inherited from generation to generation (like thalassemia or cystic fibrosis) but also for diseases were genes are one of many causative factors (e.g. both inheritance and environment play a role). Genetic BRCA testing can be performed for diagnostic purposes but also, presymptomatic tests can be performed in order to check for disease predisposition.

BRCA genetic testing entails mapping the gene sequence for a specific individual and comparing it with what is most commonly encountered in the general population. Differentiations (variants) from the common sequence are assessed by genetic experts and classified as: variants that indicate pathogenicity; variants that do not indicate pathogenicity; or variants of uncertain/unclassified clinical significance (VUS). Based on genetic test results, doctors can plan clinical actions; some of these actions can be as radical as mastectomy or ovary removal. The process of BRCA genetic testing can be summarized in seven major parts: a) sampling, b) lab preparation of the sample, c) sample processing by a specialized machine (sequencer), d) variant identification using informatics tools, e) variant lookup in genetic databases, f) functional analysis, g) clinical and family data evaluation (Figure 1).

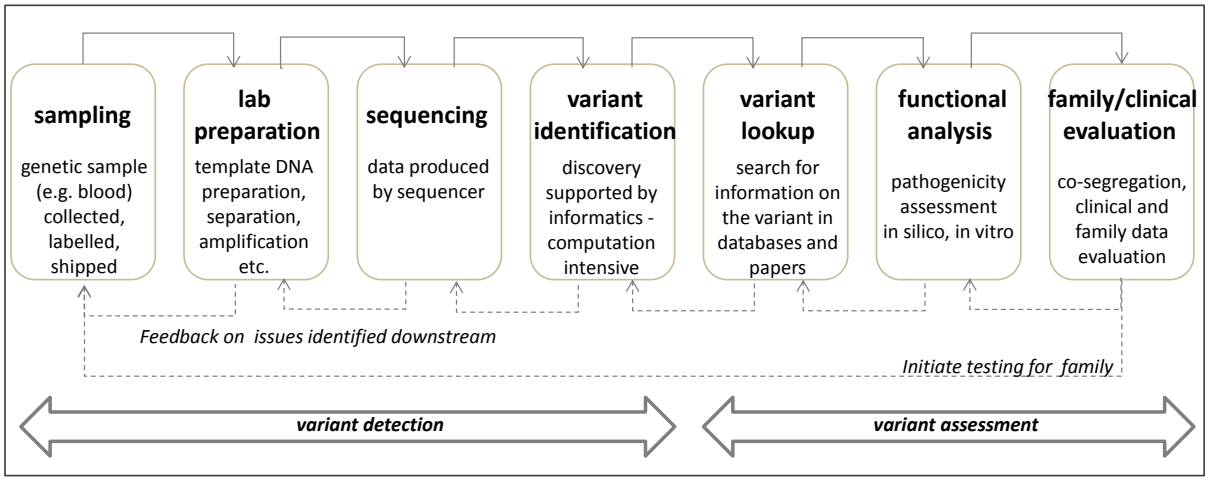


Figure 1. Overview of BRCA genetic testing.

Sampling and lab preparations are important for obtaining good quality sequence data (Morey et al. 2013). Similarly, having a good understanding of possible problems with the sequencer that might impact raw data production is important. Nevertheless, variant detection is less challenging than variant assessment, hence, the availability of good quality information for assessment purposes is critical.

4 Findings

4.1 Tensions around the "exclusion" property for BRCA datasets

In 1995, ten scientists from universities, hospitals, a research institute and a private company in different European countries and the USA, created a web-accessible data repository named Breast Cancer Information Core (BIC) (Friend et al. 1995). BIC's aim is explained in a 1996 paper: "One of the serious impediments to achieving clinical benefits from the isolation of the *BRCAl* gene is finding and assessing the significance of mutations in this new cancer susceptibility gene. This will be greatly facilitated by coordinated detection and interpretation of mutations and the dissemination of this information to as many qualified investigators as possible. To this end, the BIC has created and maintains a central repository for information regarding mutations and polymorphism" (Couch & Weber, 1996).

Currently, BIC contains more than 30 000 entries on *BRCA1* and *BRCA2* variants that include characterisations of variants' clinical importance, complementary data about samples' origin, detection methods, and depositor's contact details. The data are deposited by individual investigators, research, hospital-based and commercial labs and are published after being examined and edited by members of BIC's steering committee. Registration is open to all and access to registered users is unrestricted. Nevertheless, several labs around the world are not contributing information. Most notably, Myriad, which is the world's largest molecular diagnostic clinical laboratory (Myriad Genetics 2014) and has been the largest BIC contributor in the past, stopped contributing in 2004 (Figure 2). Myriad explained this decision by claiming that the common database lacked operational and clinical standards (Tucker 2014) and also questioned the data quality by pointing to the high proportion of variants classified as of unknown clinical importance (Angrist and Cook-Deegan 2014). It should be noted that in the 2002-2013 period, Myriad achieved to reduce by 84% the rate of variants of uncertain/ unclassified clinical significance (VUS) in their test results, bringing it down to 2.1% (Eggington et al. 2013).

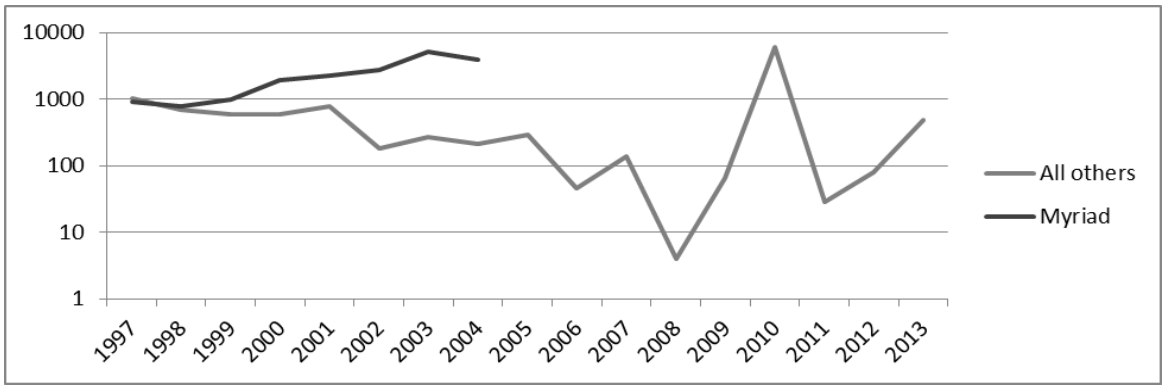


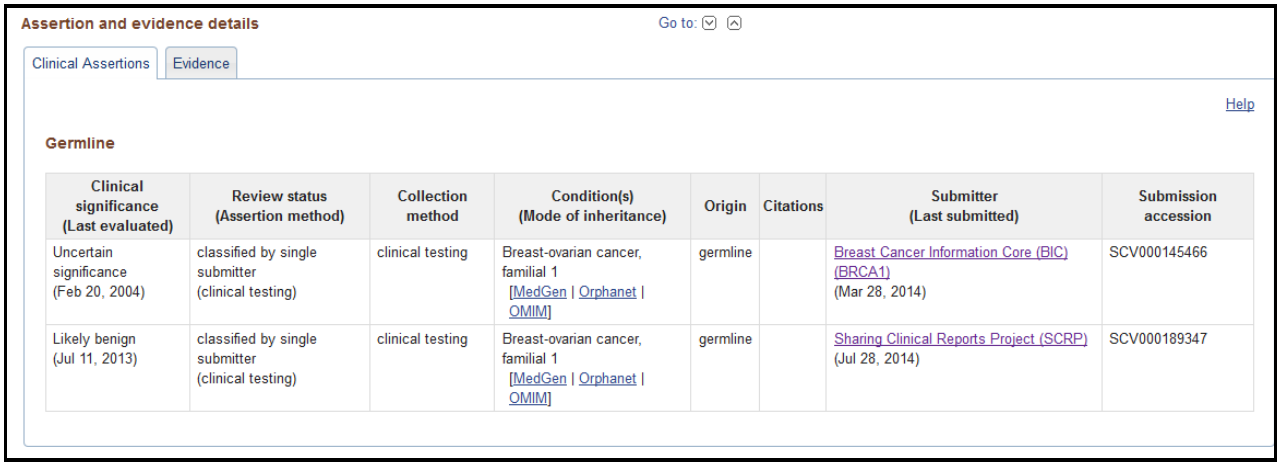
Figure 2. Breakdown of annual submissions to the Breast Cancer Information Core 1997-2013 (analysis of database instance downloaded on September 17th 2014).

Myriad was founded in 1991 as a spin-off from the University of Utah's Centre for Genetic Epidemiology. During the 1997-2000 period the company was granted with nine patents in the United States on the *BRCA1* and *BRCA2* genes. The patents gave them control over the use of diagnostic tests on those genes in the USA which explains its large number of BIC submissions. In 2013, the US Supreme Court invalidated Myriad's patents on *BRCA1* and *BRCA2* genes after the Association for Molecular Pathology along with a number of research centres and patients' associations challenged the patents' validity. Still, Myriad retained a privilege in BRCA testing by having accumulated data in a proprietary database: "by becoming the world's largest testing service, Myriad also discovers new variants and incorporates those into its database… for BRCA testing, Myriad has a distinct advantage, even over the best academic centres, because of its unique data set" (Baldwin and Cook-Deegan 2013).

As a reaction to Myriad's decision to stop depositing information, two senior medical geneticists initiated a project named "Sharing Clinical Reports Project" (SCRP) reaching out for physicians that receive test reports from Myriad (Nguyen and Terry 2013). The idea for this initiative was to collect Myriad's variant assessments via the distributed network of physicians that keep records of individual reports for their patients. SCRP was initiated in 2012 (Kolata 2013) and up to August 2014 the project received 5416 submissions related to 2048 unique BRCA variants (Sharing Clinical Reports Project 2014). Additionally to SCRP there is as "sister project" addressed to patients titled "Free the Data" (Free the Data 2014). This additional initiative exploits a decision taken in February 2014 by the US Department of Health and Human Services which provides for patient access to their test reports directly from the laboratory (before that, physicians were information gatekeepers) (Angrist and Cook-Deegan 2014). The "Free the Data" project not only provides a platform for patients to upload their BRCA reports but goes one step further by giving them the option to contribute additional personal health related information that can be invaluable for interpreting their genetic data.

In 2012, the US National Center for Biotechnology Information (NCBI) released a new open repository aimed to host data on variant interpretations (Landrum et al. 2013) from any type of genetic test. This new repository (named ClinVar) invited scientific groups that hold data from clinical testing, research and literature to contribute their information. ClinVar is a variant-centric repository that covers all human gene variations – it is not dedicated only to BRCA genes. Both BIC and SCRP registered as submitters and have contributed interpretations for all their variants. A number of hospitals and labs that were not submitting actively to BIC started submitting data to ClinVar. This is maybe related to the implementation of several mechanisms to strengthen data release. Today the big journals in the field demand disclosure of both software code, algorithms and sequence data upon publication (Krol 2014). Nevertheless, estimates by experts suggest that all the various public databases combined have amassed about 20–25% of Myriad's data up to 2014 (Conley et al. 2014).

Available interpretations for a particular variant are presented in ClinVar through comprehensive views where multiple submissions are associated (Figure 3). Conflict resolutions have to be provided from an expert panel or a professional society (Landrum et al. 2013).



Assertion and evidence details Go to:

Clinical Assertions | Evidence

Help

Germline

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter (Last submitted)	Submission accession
Uncertain significance (Feb 20, 2004)	classified by single submitter (clinical testing)	clinical testing	Breast-ovarian cancer, familial 1 [MedGen | Orphanet | OMIM]	germline		Breast Cancer Information Core (BIC) (BRCA1) (Mar 28, 2014)	SCV000145466
Likely benign (Jul 11, 2013)	classified by single submitter (clinical testing)	clinical testing	Breast-ovarian cancer, familial 1 [MedGen | Orphanet | OMIM]	germline		Sharing Clinical Reports Project (SCRP) (Jul 28, 2014)	SCV000189347

Figure 3. Screenshot from ClinVar: a consolidated view of interpretations for a specific BRCA variant.

Unlike BIC where the content of submissions is reviewed by members of its steering committee and the variant interpretation for each entry is periodically updated to reflect the latest findings, in ClinVar the original content of the submissions is not curated or modified. By introducing a "submitter-driven" logic ClinVar faced its own share of challenges. During one week in October 2013, a single lab retracted 80 percent of its submitted BRCA variant data: "The clinical lab had gathered variant information by curating the literature, before launching a commercial test. The lab then voluntarily shared these interpretations from the literature into ClinVar but a lot of those classifications were wrong. When these mistakes came to light, the unnamed lab retracted the initial submission from ClinVar and then re-entered the variant information without the pathogenicity claims drawn from the literature" (Ray 2014). The proponents of ClinVar's open approach point to the value of the community reaction that led to submissions' retraction and use it as an "example of the benefits of greater transparency within clinical labs' interpretation work" (Ray 2014).

In 2015 two more initiatives to put in place platforms for genetic data accumulation were announced. The commercial laboratory Quest Diagnostics and Inserm (the French National Institute of Health and Medical Research institution) announced the launch of BRCA Share. This is a public-private data sharing initiative, with Quest licensing the data and forming sub-license agreements with participants. To participate in BRCA Share, laboratories have to commit on sharing past, present and future data. Commercial labs need to pay a subscription fee according to their size, while research entities get access at no cost. This financial arrangement allows BRCA Share invest on data curation arrangements to attend to data quality and to conduct functional studies on the effects of mutations, without depend-

ing on research or public funding. Essentially, BRCA Share's alternative model resembles a "club arrangement" and was described by a Nature editorial as a "walled garden" (Nature Editorial 2015). Following a different, open approach, in June 2015 a new web portal named BRCA Exchange was launched under the auspices of the Global Alliance for Genomics and Health. The Global Alliance is an impressive network of major healthcare providers, research funders, research institutions, disease advocacy organizations, and life science and information technology companies. The network's aim is to create a common framework of harmonized approaches to enable the responsible, voluntary, and secure sharing of genomic and clinical data (Global Alliance for Genomics and Health 2015). The new portal aims to facilitate open sharing of expert-reviewed BRCA variant classifications and is already launched as a "beta". A number of working groups with members located around the world are working within this initiative to identify sources of classified variants to be aggregated and also, unclassified variants with evidence, which can then be interpreted, classified, and aggregated. Essentially, BRCA does not rely to the depositors' initiative for data collection but rather, proactively accumulates data through a "pull model". At this moment, ClinVar is used as a major source of information. The classified variants are then displayed at the BRCA Exchange portal for "use by clinicians, patients, and labs to enable accurate understanding of any individual BRCA1/2 variant, for superior clinical care and decision-making" (BRCA Exchange 2015). In Table 3 we summarise the key events in the trajectory we described in the previous paragraphs.

Event	When	Description
BIC repository creation	1995	A web-accessible repository that facilitates the accumulation and sharing of BRCA variant assessments is launched. Data are accessible by all and can be deposited after being reviewed by volunteering experts.
Myriad stops contributing to BIC	2004	The world's largest molecular diagnostic clinical laboratory stops contributing data to BIC.
SCRP	2012	A mechanism for collecting BRCA variant assessments via doctors that receive test results is put in place in USA.
ClinVar	2012	An aggregator of variant interpretations (in the form of an open, web-accessible repository) is launched. The repository covers all genes (not only BRCA variants). The data are deposited by multiple contributors and are not curated.
Free the Data	2014	A mechanism for collecting variant assessments directly from patients is launched in USA.
BRCA Share	2015	A public-private data sharing initiative for BRCA variant data is launched. Participation is possible for all labs on a fee basis (free for research) as long as they share all their data.
BRCA Exchange	2015	A new web portal to share expert-reviewed BRCA variant classifications is launched. A new pull model is implemented for the proactive identification and aggregation of data.

Table 3. Key events related to tensions around the "exclusion" property of BRCA data.

4.2 Tensions around the "subtractability" property for BRCA datasets

Intuitively, the information that relates to BRCA variant assessments comes to mind as a non-subtractable good in the sense that normally, someone that gets hold of this information can not impair its value for others. In general, information goods are easily conceptualised as non-subtractable: benefits consumed by someone are not subtracting from the benefits available to others. Of course, this is an oversimplistic understanding that does not take into account the rivalries and races among researchers to establish 'priority' and the maze of intellectual property rights regimes related to commercially valuable information (David 2003). Although the classification or reclassification of a variant "*is not a discovery big enough to warranty the publication of a paper"* (interview with genetics' specialist) there are strong incentives for not disclosing variant information as this could level-out the competitive advantage of labs that have a significant information base to draw upon. When the information on a particular variant becomes part of the common knowledge base it can be valuable in places where labs lack the full range of resources and skills for variant assessment. For instance, to come up with variant assessments, Myriad utilises multiple methods that include: in vitro assessments, segregation analysis, co-evaluation with mutations in other genes associated with the same syndrome, and the application of a scoring algorithm developed using the large numbers of data available to the company (Eggington et al. 2012; Eggington et al. 2013; Pruss et al. 2014). A New York Times reader posted a comment in the newspaper's webpage describing her experience with the process: "*I was recently informed that I have one of those "uncertain mutations."* (…) *Myriad will offer free testing to my family members if I complete an extensive family history--including cousins, aunts, uncles, nieces and nephews, in addition to immediate family*" (Anonymous 2013). Practically this means that Myriad incurs a high cost and leverages an advanced infrastructure in order to ensure good quality assessments, to minimise uncertainty and to complete the testing procedure as quickly as possible. Although apparently Myriad does not charge a premium price it certainly uses the shorter turnaround time than other providers and the quality of the results as competitive advantages in order to retain a high market share (Cook-Deegan et al. 2010; Karow 2013).

To counteract on the tendency of labs to keep valuable information for their own use guidelines state that labs should report when *"a variant of uncertain clinical significance becomes clearly pathogenic, or a variant is not pathogenic anymore"* (Wallis et al. 2013). Indeed, Myriad has been contributing information on variants identified through scientific publications. In an effort to reduce the value of nondisclosure in February 2014, the Cancer Genetic Counselling Program of Yale School of Medicine issued a position statement where it is declared that laboratory choices will be made on the basis of four criteria: quality, time, cost and open access: "whenever possible we will choose laboratories that have pledged to make all of their past, present, and future gene data publicly available in order to al-

low this important information to be freely accessible to all clinicians and researchers, to further the advancement of medical knowledge and to best serve patient care. We will not support laboratories that hoard data." (Yale Cancer Genetic Counceling 2014). If more healthcare providers were to introduce similar policies, the competitive advantage of not sharing data would be diminished. Actually, the relationships among providers, users, funders, regulators and custodians of data impact the processes through which data acquire value, the distribution of this value, and the value accrual dynamics. The value accrual dynamics are in turn configuring the "subtractability" property for BRCA datasets.

An opposite trend is generated by web 2.0 technologies for patient involvement that are creating the conditions for BRCA datasets to become anti-rival goods. This means that value can increase for all each time someone uses the dataset as additional insights might be generated. Since March 2014, the "American BRCA Outcomes and Utilization of Testing Patient-Powered Research Network" has been active in enrolling patients and soliciting health-related information directly from them (Daugherty et al. 2014; Fleurence et al. 2014). The initiative is funded by the Patient-Centered Outcomes Research Institute (PCORI), a US government-supported agency. Their aim is to support patient-reported data collection, real-time data sharing across the network, and automated data transfer with organisations including genetic testing labs. Data collection is addressed to multiple subgroups: BRCA variant carriers, breast and ovarian cancer patients, women with metastatic disease, and family members (Patient-Centered Outcomes Research Institute 2014). The initiative "Free the Data" (see section 4.1) has been embraced by ABOUT which provides a functionality to upload BRCA test results (Friedman 2014). Additionally, network participants can contribute personal health information, experiences and outcomes, extracts from medical records and medical claims. The long term follow-up of BRCA variant carriers creates significant new possibilities for variant interpretations by leveraging longitudinal data.

5 Discussion and Conclusion

The narrative of the developments in the domain of BRCA genetic data shows that exclusion and subtractability in our case are not pre-existing data properties but sociotechnical achievements that are still in the making. Technically, it is not possible to exclude individuals from reusing existing BRCA data as the SCRP and "Free the data" initiatives prove. Technology offers the possibility for information to be easily digitised and shared, and it can start flowing towards public databases. Still, this kind of flow can be painstakingly slow. In the meantime, one powerful actor created the conditions for building a sheltered stock of valuable information. Other actors in the field were also active early: they created common data repositories already 20 years ago and performed "advocacy work" (Lawrence and Suddaby 2006) aiming to mobilise political and regulatory support. Still, the outcome of their efforts did not lead to regulatory measures for open sharing of all existing BRCA datasets.

The classification of BRCA datasets as public, or common pool, or private or club goods is still to be settled. Hence, aiming to handle them as indisputably common goods or simply setting-up infrastructures with the expectation that they will be used, is futile. Instead, a continual effort to put in place appropriate conditions for establishing a "commons logic" is required. This effort will entail introducing measures with sociotechnical nature combining new technological potentialities with regulatory provisions. Even more importantly, the effort will entail negotiations and settlements among multiple stakeholders (Borgman 2015; Hess and Ostrom 2006) with the aim to establish good working relationships between funders, scientists, clinicians, commercial actors and patients.

Striking a balance among the different interests is challenging and new governance arrangements will require thinking of actors' relations by taking into account also technology design. Regulators can impose digitally facilitated transmissions of test results in semi-structured formats via secure networks through trusted third parties (e.g. national health authorities) to ensure data accumulation (interestingly, even today, test results are routinely transmitted by labs via fax). The connectivity afforded by web technologies can be leveraged for enrolling multiple qualified scientists from all over the world in massive curation and quality assurance activities (as in the BRCA Exchange initiative). Furthermore, new arenas for value generation could be created by catering for the growing needs of clinicians for decision support tools. The development of specialised information services (with powerful graphic visualizations, inference models built upon multiple lines of evidence, targeted data feeds containing variant reclassifications and alerts) can be new sources of competitive advantage for companies.

Michael Heller introduced the term "tragedy of the anticommons" to describe the underutilisation of resources when multiple actors with partial but exclusionary rights are involved and fail to coordinate effectively (Heller 1998). Similarly, in the genetics' data case, the advances that made possible the continuous cost decrease for genome sequencing cannot bring significant benefits in medical care and public health unless the coordination of the multiple right-holding actors is facilitated and accelerated.

A significant limitation of the case that we have studied is that it does not pose significant privacy issues. Isolated anonymised data on single gene's variants (as for *BRCA1* and *BRCA2*) cannot be used for the re-identification of patients. Furthermore, the volumes of data discussed are relatively low. The advent of next generation sequencing (Metzker 2010; Pabinger et al. 2014) will bring-in additional challenges related both to privacy and to the sheer volume of information. Future research may proceed in two general directions. The first direction is towards extending the analysis and performing further research in the domain of next generation sequencing. Another equally important direction is towards the operationalisation of insights from the current study in order to have an impact in the governance of the currently dispersed datasets on individual genes' data.

References

Angrist, M. and Cook-Deegan, R. (2014), Distributing the future: The weak justifications for keeping human genomic databases secret and the challenges and opportunities in reverse engineering them, Applied & Translational Genomics, (in press, electronic pre-publication).

Anonymous (2013), Comment under the article "DNA project aims to make companys data public". <http://www.nytimes.com/2013/04/13/health/dna-project-aims-to-make-companys-data-public.html?pagewanted=all&_r=1&>.

Baldwin, A L. and Cook-Deegan, R. (2013), Constructing narratives of heroism and villainy: Case study of Myriad's BRACAnalysis® compared to Genentech's Herceptin®, Genome medicine, 5 (1), 8.

Battista, RN, et al. (2012), Genetics in health care: an overview of current and emerging models, Public health genomics, 15 (1), 34-45.

Bennett, C., et al. (2010), A toolkit for incorporating genetics into mainstream medical services: Learning from service development pilots in England, BMC health services research, 10 (1), 125.

Borgman, C. (2015), Big data, little data, no data: scholarship in the networked world: MIT Press

BRCA Exchange (2015), History of the BRCA Exchange. <http://brcaexchange.org/about/history>, accessed 15 October 2015.

Conley, J. M., Cook-Deegan, R., and Lázaro-Muñoz, G. (2014), Myriad after myriad: the proprietary data dilemma, North Carolina journal of law & technology, 15 (4), 597.

Cook-Deegan, R., et al. (2010), Impact of gene patents and licensing practices on access to genetic testing for inherited susceptibility to cancer: comparing breast and ovarian cancers with colon cancers, Genetics in Medicine, 12, S15-S38.

Council of Europe (2010), Recommendation CM/Rec(2010)11 of the Committee of Ministers to member states on the impact of genetics on the organisation of health care services and training of health professionals. .<https://wcd.coe.int/ViewDoc.jsp?id=1678099&Site=COE&BackColorInternet=DBDCF2&BackColorIntranet=FDC864&BackColorLogged=FDC864a>, accessed 25 November 2015.

Daugherty, S. E., et al. (2014), Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research, Journal of the American Medical Informatics Association, amiajnl-2014-002758.

David, P.A. (2003), The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information: a primer, in J. Esanu and P. Uhlir (eds.), The role of the public domain in scientific and technical data and information (Washington, DC.: The National Academies Press), 19-34.

Eggington, JM., et al. (2012), Current variant of uncertain significance rates in BRCA1/2 and Lynch Syndrome testing (MLH1, MSH2, MSH6, PMS2, EPCAM), ACMG Annual Clinical Genetics Meeting. Charlotte, NC.

Eggington, JM., et al. (2013), A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes, Clinical genetics.

Fleurence, R. L., et al. (2014), Launching PCORnet, a national patient-centered clinical research network, Journal of the American Medical Informatics Association, amiajnl-2014-002747.

Free the Data (2014), Genetic information is more valuable when shared. <www. free-the-data. org>, accessed 15 September 2014.

Friedman, S. (2014), ABOUT Patient-Powered Research Network, FORCE newsletter, spring 2014. <http://www.facingourrisk.org/understanding-brca-and-hboc/publications/newsletter/archives/2014spring/about-network.php>.

Friend, S., et al. (1995), Breast cancer information on the web, Nature genetics, 11 (3), 238-39.

Global Alliance for Genomics and Health (2015), About the Global Alliance. <https://genomicsandhealth.org/about-the-global-alliance/frequently-asked-questions>, accessed 26 October 2015.

Haga, S. B., et al. (2012), Survey of genetic counselors and clinical geneticists' use and attitudes toward pharmacogenetic testing, Clinical Genetics, 82 (2), 115-20.

Heller, M. (1998), The tragedy of the anticommons: property in the transition from Marx to markets, Harvard Law Review, 621-88.

Hess, C. and Ostrom, E. (2003), Ideas, artifacts, and facilities: information as a common-pool resource, Law and contemporary problems, 66 (1/2), 111-45.

Hess, C. and Ostrom, E. (2006), 'A framework for analyzing the knowledge commons', in C. Hess and E. Ostrom (eds.), Understanding Knowledge as a Commons: from Theory to Practice: MIT Press.

Karow, J (2013), Competition for Myriad heats up as US testing labs launch BRCA tests, hereditary cancer panels, Clin. Sequencing News (26 June, http://www. genomeweb. com/sequencing/competition-myriad-heats-us-testing-labs-launchbrca-tests-hereditary-cancer-pan).

Klein, H. K. and Myers, M. D. (1999), A set of principles for conducting and evaluating interpretive field studies in information systems, MIS Quarterly, 23 (1), 67-93.

Kolata, G. (2013), DNA Project Aims to Make Public a Company's Data on Cancer Genes, The New York Times, (April 12 2013). <http://www.nytimes.com/2013/04/13/health/dna-project-aims-to-make-companys-data-public.html?pagewanted=all&_r=0>, accessed 15 September 2014.

Krol, A. (2014), As Genetics Moves to the Clinic, Pathogenic Variants Still Subject to Doubt and Debate, Bio-IT World, (April 17, 2014). <http://www.bio-itworld.com/2014/4/17/genetics-moves-clinic-pathogenic-variants-still-subject-doubt-debate.html>, accessed 11 November 2014.

Landrum, M. J., et al. (2013), ClinVar: public archive of relationships among sequence variation and human phenotype, Nucleic Acids Research, gkt1113.

Lawrence, T. B. and Suddaby, R. (2006), 'Institutions and institutional work. ', in S. R. Clegg, et al. (eds.), Handbook of organization studies.; London: Sage, 215-54.

Lawrence, T. B., Suddaby, R., and Leca, B. (2009), 'Introduction:Theorizing and studying institutional work', in T. B. Lawrence, R. Suddaby, and B. Leca (eds.), Institutional work : actors and agency in institutional studies of organizations; Cambridge, UK: Cambridge University Press, 1-27.

Lawrence, T. B., Suddaby, R., and Leca, B. (2011), Institutional Work: Refocusing Institutional Studies of Organization, Journal of Management Inquiry, 20 (1), 52-58.

Metzker, M. L. (2010), Sequencing technologies—the next generation, Nature reviews genetics, 11 (1), 31-46.

Miki, Y., et al. (1994), A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, Science, 266 (5182), 66-71.

Morey, M., et al. (2013), A glimpse into past, present, and future DNA sequencing, Molecular genetics and metabolism, 110 (1), 3-24.

Myriad Genetics (2014) Myriad Genetics Corporate Presentation, <https://myriad-web.s3.amazonaws.com/myriad.com/pdf/Myriad-Corporate-Presentation.pdf>

Nature Editorial (2015), Thank you for sharing, Nature 520 (7549), 585.

Nguyen, S. and Terry, SF. (2013), Free the data: the end of genetic data as trade secrets, Genetic testing and molecular biomarkers, 17 (8), 579-80.

Ostrom, E. (1990), Governing the commons: The evolution of institutions for collective action: Cambridge University Press

Ostrom, E. (2011), Background on the institutional analysis and development framework, Policy Studies Journal, 39 (1), 7-27.

Ostrom, V. and Ostrom, E. (1977), 'Public Goods and Public Choices', in E. S. Savas (ed.), Alternatives for delivering public services: toward improved performance.

Pabinger, S., et al. (2014), A survey of tools for variant analysis of next-generation genome sequencing data, Briefings in bioinformatics, 15 (2), 256-78.

Patient-Centered Outcomes Research Institute (2014), American BRCA Outcomes and Utilization of Testing Patient-Powered Research Network (ABOUT Network). <http://www.pcori.org/research-results/2013/american-brca-outcomes-and-utilization-testing-patient-powered-research>, accessed 14 November 2014.

Pettigrew, A. (1997), What is processual analysis?, Scandinavian Journal of Management, 13 (4), 337-48.

Pruss, D., et al. (2014), Development and validation of a new algorithm for the reclassification of genetic variants identified in the BRCA1 and BRCA2 genes, Breast cancer research and treatment, 147 (1), 119-32.

Quintáns, B, et al. (2014), Medical genomics: The intricate path from genetic variant identification to clinical interpretation, Applied & Translational Genomics, 3 (3), 60-67.

Ray, T. (2014), In Tackling the VUS Challenge, Are Public Databases the Solution or a Liability for Labs?, Pharmacogenomics Reporter, (February 12, 2014). <http://www.genomeweb.com/clinical-genomics/tackling-vus-challenge-are-public-databases-solution-or-liability-labs>, accessed 15 September 2014.

Sboner, A., et al. (2011), The real cost of sequencing: higher than you think!, Genome biology, 12 (8), 125.

Scott, W. R. (2005), 'Institutional theory: contributing to a theoretical research program', in K. G. Smith and M. A. Hitt (eds.), Great minds in management : the process of theory development; New York: Oxford University Press, xiii, 600 p.

Sharing Clinical Reports Project (2014), Sharing Clinical Reports Project. <http://www.iccg.org/about-the-iccg/collaborations/sharing-clinical-reports-project/>, accessed 15 September 2014.

Skirton, H., Patch, C., and Williams, J. (2005), Applied genetics in healthcare: A handbook for specialist practitioners: Garland Science

Slade, I., et al. (2015), Development of cancer genetic services in the UK: A national consultation, Genome medicine, 7 (1), 18.

Snyderman, R. (2012), Personalized health care: From theory to practice, Biotechnology Journal, 7 (8), 973-79.

Tucker, K.I. (2014), 'Genetics Lab Refuses To Share Data That Could Save Lives', Jewish Daily Forward, 12 August 2014.

Vrijenhoek, T., et al. (2015), Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects, Eur J Hum Genet, 23 (9), 1142-50.

Wallis, Y., et al. (2013), Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics, Association for Clinical Genetic Science of the British Society for Genetic Medicine. <http://www.acgs.uk.com/ media/

774853/evaluation_and_reporting_of_sequence_variants_bpgs_june_2013_-_finalpdf.pdf>, accessed 8 November 2014.

Walsham, G. (1993), Interpreting information systems in organizations, John Wiley series in information systems; Chichester, West Sussex, England ; New York: Wiley

Wooster, R., et al. (1995), Identification of the breast cancer susceptibility gene BRCA2, Nature, 378 (6559), 789-92.

Yale Cancer Genetic Counceling Titl', <http://yalecancergeneticcounseling.blogspot.gr/2014/02/genetic-testing-lab-position-statement_2573.html>, accessed 7 July 2015.