

E-Infrastructures and the divergent assetisation of public health data: expectations, asymmetries and uncertainties

Paraskevas Vezyridis, Stephen Timmons

Nottingham University Business School, UK

Abstract

Normative, scientific and economic pledges to Electronic Health Record (EHR) data-driven research (for health and wealth) attempt to reconfigure public health data as an asset for realising multiple values across healthcare, research and finance. In this paper, we examine some of the expectations, frictions and uncertainties involved with the assetisation of de-identified NHS patient data by (primary care) research services in UK. We introduce the concept of ‘asymmetric assetisation divergence’ to study the various practices of configuring and using this data, both as a continuously generated resource to be extracted and as an asset to be circulated in the knowledge economy. As data assetisation and exploitations grow bigger and more diverse, the capitalisation of these datasets may constitute EHR data-driven research in healthcare as an attractive technoscientific activity, but one limited to those actors with specific sociotechnical resources in place to fully exploit them at the required scale.

Keywords: NHS patient data; infrastructure; expectations; assetisation; uncertainty

Introduction

I believe, however, that as we seek to unlock a new age of enterprise, we might need to go further in exploring ways of unlocking new growth without increasing public spending. As with a business in a cash crisis, we need to shore up the profit and loss account by reducing waste, as the Government have so quickly done. Equally, as with real business growth, we need to look creatively at our balance sheet and think about our assets and our competitive advantage. Everyone in government, in every Department at every level, should be asking themselves, "What can we sell to the rest of the world, in order to repair our damaged public finances?"...The third is the National Health Service. I know from my own experience that we are sitting on billions of pounds-worth of patient data. Let us think about how we can unlock the value of those data around the world. (Freeman, 2010)

It was November 2010, in the aftermath of the last global financial crisis, and the UK House of Commons was debating the country's growth policy. George William Freeman MP (Con), who four years later was appointed Minister for Life Sciences at the Department of Health *and* the Department for Business, Innovation and Skills, unwrapped his own vision for the country's economic growth. Problematizing the state as a 'business in a cash crisis' that needs to enter 'a new age of enterprise', the MP asked public services to start looking for public assets they could sell to potential investors so as to restore public finances.

While such national policy expectations and narratives of wealth creation out of NHS patient data are not new, they have intensified after the 2008/9 financial crisis as (corporate) biomedical innovation is being outsourced to the academic sector (Robinson, 2018) and (personal) data has emerged in the economic literature as the world economy's new asset class (cf. BIGT, 2003; Department of Health, 2011; HM Government, 2018; WEF, 2011). At the

same time, national health systems are ‘morally obliged’ to become data-driven so as to stop ‘flying blind’ and, start ‘saving lives’ and taxpayers’ money (du Preez, 2015). They are now expected to become not only more (cost-)effective in personalised healthcare but also expand their wealth-creating role as investors in biomedical Research and Development (Department of Health, 2011; Kelsey and Cavendish, 2014).

The UK has been consistently advertised as the home to one of the best and biggest healthcare datasets in the world (HM Government, 2018). Its universal, (largely) computerised and centrally managed public healthcare system (NHS) has created volumes of linked datasets from longitudinal Electronic Health Records (EHRs) across all levels of care. In fact, the demand for, and availability of, NHS patient data, for research into anything from pharmacovigilance, drug prescribing and safety, standards of care and trial recruitment, continues to grow year after year (NHS Digital, 2018). Commercial models are currently being debated in the pursuit of lawful and publicly acceptable contractual arrangements between the NHS and private sectors for creating and realising the values of this public asset (HM Treasury, 2018; Harwich and Lasko-Skinner, 2018). Recently, the economic value of curated NHS patient data was estimated at £9bn per year for both the NHS and the patients (Wayman and Hunerlach, 2019). However, these valuations neglect the investments in technoscientific and market infrastructures required to reap such financial and epistemic benefits.

Our aim with this paper is to explore how narratives of innovation and normative pledges to the interrelatedness of health and wealth attempt to mobilise and configure public health data as an asset for realising multiple values across healthcare, research and finance (Welsh and Wynne, 2013; Hogarth, 2017). For this, we take as our case study the work of research services in England that synthesise and release NHS patient data to actors in healthcare and beyond, such as the Clinical Practice Research Datalink (CPRD), QResearch,

The Health Improvement Network (THIN), CALIBER and ResearchOne. Understanding ‘valuation as a social practice’ (Helgesson and Lee, 2017: 533), we focus on the intertwined societal, economic, and sociomaterial expectations and logics such research services translate and enact. This enables them to realise and extract the *values* of this public asset for their organisational sustainability and for other actors across the economy (cf. Star, 1985; Brown and Michael, 2003; Birch, 2017). Responding to Birch’s (2019) call for more empirical STS investigations of *how* ‘things are turned into assets’ (Muniesa et al., 2017) and controlled for the extraction of rents (such as licences or fees), we show that the assetisation of NHS patient data by research services is a complex and laborious process. It involves the configuration of competing and complementary frictions, as well as uncertainties around (*inter alia*) coding practices, regulation, acceptability, supplies of datasets, technoscientific capital and user demand.

We demonstrate that NHS patient data assetisation by these research services serves four main purposes: maintaining organisational sustainability, developing their capability to continue assetising data, developing the epistemic and human capital of this field (e.g. development of new scientific methods and training of new data scientists) and, consequently, strengthening their role in the valorisation, performativity and financialisation of EHR data-driven research. For this, we have coined the concept of *asymmetric assetisation divergence* to explore their expectations and valuations as well as their normative, scientific and economic discourses, and practices. (Birch, 2017a; Muniesa et al., 2017). In this way, we elucidate the asymmetries around the sociotechnical and financial infrastructures that are configured for this assetisation, including the control of data flows for research, the public’s participation in decision-making and the various knowledge assets (e.g. phenotypes, biomarkers, quality improvement reports, recruitment pools of research participants, clinical

risk prediction algorithms, and scientific publications) that are made (im)possible for healthcare, academic and biomedical networks of actors.

This paper draws on ethnographic material mainly from 27 interviews conducted in 2016 with 7 GPs who were involved in clinical commissioning, information governance, medical ethics teaching or academic EHR data-driven research, 9 citizens who reported having opted out from and/or had campaigned against programmes of NHS patient data capitalisation (i.e. care.data) and 11 health data researchers. The last group of participants comprised statisticians, epidemiologists, data architects and research facilitators who have worked with data from and/or for one of the more established research services in England that have been collecting and curating de-identified NHS patient data from a network of contributing GP practices across the country to support EHR data-driven research (Vezyridis and Timmons, 2016). These included QResearch – a partnership between the IT supplier EMIS Health and (since 2019) the University of Oxford (previously the University of Nottingham) – that holds data from 1500 (out of approximately 9800) GP practices in the UK; the CPRD – the oldest research service with a history dating back to 1987 – that extracts and curates data from GP practices using IMS Health’s computer system and (since 2018) also from EMIS Health; THIN by Cegedim SA that extracts data from over 550 GP practices that use the Vision primary care software (approximately 6% of the UK population); CALIBER at the University College London that has a licence with CPRD since 2012 and ResearchOne by TPP and the University of Leeds that (since 2013) houses data from both primary and secondary care providers using the company’s SystemOne software. All interviews focused mainly on the opportunities and the (technical, social and ethical) challenges of realising the benefits of NHS patient data-driven research, particularly in English primary care. Health data researchers and GPs were also asked more specific questions around the challenges of developing and maintaining such research services,

including issues of sociotechnical infrastructure and information governance, as well as of conducting observational studies with NHS patient data. These interviews were supplemented with documents, reports and online material from these research services' website. Our analysis is also informed by the first author's non-participant observation of team meetings at some of these research services, national health data analytics workshops and public consultations on the ethics of NHS patient data exploitations as well as by completing university training courses for data researchers related to the opportunities and challenges of conducting observational studies with EHR primary care data in the UK.

In the next section, we theorise data assetisation drawing from the sociology of expectations and science and technology studies (STS) (Borup et al., 2006; Birch, 2017a, 2019). This forms our analytical approach for studying the performative promises, normativities and sociotechnical practices involved in these situated processes of NHS patient data capitalisation (Muniesa et al., 2017). Following this, we examine some of the expectations and uncertainties as well as frictions and risks involved in the process of transforming NHS patient datasets into research assets. We focus not only on the scientific but also on the economic, political, social and ethical valuations these research services have to navigate through. We conclude speculating on the future of EHR data-driven research and argue that asymmetric assetisation divergence may indeed foster innovation in EHR data-driven research and development (cf. Kleinman and Vallas, 2001). However, it is also creating unequal configurations of access to knowledge production and public scrutiny as these research services compete for data and funding. More centrally controlled flows of data and assets for the extraction of rents may, in the end, benefit only those networked actors that have the sociomaterial resources (including financial capital) and knowledge expertise to capitalise on these (public) datasets for narrowly specified purposes (cf. Birch, 2019; Muniesa et al., 2017).

Theorising research expectations, service valuations and data assetisations

From a STS perspective, expectations have been theorised as both products and producers of innovation, mobilising, and institutionalising people and resources (Brown and Michael, 2003; Brown, 2003; Borup et al., 2006; Tutton, 2011). Scholars in this field have shown how expectations produce, in their own material-discursive way, the necessary ‘dynamism and momentum’ (Brown and Michael, 2003) as well as the ‘incentives and obligations’ (Brown, 2003) for human and non-human actors to come together and ‘wishfully enact’ (Tutton, 2011) particular versions of the future.

Research that examined the role of biomedical initiatives in the creation of new forms of scientific, social, political and economic expectations has demonstrated, for example, how human tissue, medical data and genomics are assetised for scientific collaborations and circulations in knowledge economies (cf. Cooper and Waldby, 2014; Dagiral and Peerbaye, 2016; Tarkkala et al., 2018). Geiger and Gross (2019) have shown how consumer genomics firms mobilise a ‘platform business model’ in order to assetise genomics information via specific processes of ‘accumulation’, ‘augmentation’ and ‘obscuration’ of related uncertainties. These platforms simultaneously maintain constant flows of data and values between sellers and customers across different markets. Barrett et al. (2016) demonstrated how specific sociomaterial configurations enacted by digital platforms create multiple epistemic, ethical, and financial expectations from online health communities. Likewise, Timmons and Vezyridis (2017) and Vezyridis and Timmons (2017) demonstrated how scientific, ethical, social and economic valuations (re)configure a range of new socio-material processes and relationships between patients, hospitals, universities and biomedical industries in the process of assetising medical waste and patient data.

A main purpose of this paper, therefore, is to understand how these research services enact NHS patient data assetisations in order to frame, sustain and expand the ‘liminal space’ between hopes and promises and, concrete products and assets (cf. Hogarth, 2017; Brown, 2003). For this, we conceptualise assetisation as: ‘a process in which value is constituted by the management of value and valuation, especially as they relate to organisational entities and their capacities’ (Birch, 2017a: 470). Consequently, we treat any scientific, ethical and economic values around NHS patient data not as something ‘stable and predefined, but rather as something grappled with, articulated, and made in concrete practices’ (Dussauge et al. 2015: 2). Values, as Birch (2017a: 462, 466) asserts, are ‘immanent or latent in material things (e.g., commodity) and/or discursive claims (e.g., hope)’ and they require ‘active, ongoing, and performative management’. Both conceptually and empirically, as Brown (2003: 5) argues, expectations and values eventually become both ‘inseparable’ and ‘tradable’, forming ‘the basis of exchange relationships within “communities of promise”’.

Expectations of EHR data-driven research for speculative valuations of health and wealth have, therefore, their own important role in how GPs, NHS patients and the data they co-produce are imagined and valued to perform for EHR data-driven research within a high stakes biomedical knowledge economy (Dussauge et al., 2015; Birch, 2017a; Wienroth et al., 2019). They are fundamental for the mobilisation of various state, commercial, and academic actors as well as capital for the assetisation of NHS patient data and the performativity of specific types of health data access markets (cf. Brown, 2003; Birch, 2017a; Vezyridis and Timmons, 2017).

During our fieldwork, we noted, for example, how health data scientists valued the availability of EHR data to answer different types of research questions in healthcare. Their anticipatory discourses highlighted the unique opportunities now available to researchers from

the mining of data – anything from clinical to demographic and geographical – to identify causes of diseases, complication and management rates and, prescribing practices. They were often excited about the potential for life sciences in the UK to lead the development of new, EHR data-driven, practices for the prevention, and treatment of diseases or for conducting more ‘pragmatic’ randomised controlled trials to assess the effectiveness of drugs (Powell et al., 2017).

Research services, such as those examined here, appear to have assumed this technoscientific and economic role of materialising and capitalising on the multiple scientific, normative and economic promises and values of EHR data-driven biomedicine out of this public good (i.e. NHS patient data) (cf. Brown and Michael, 2003; Martin, 2015; Birch, 2019). They do so by enacting the (negotiated and contested) infrastructural work, i.e. the situated narratives, heterogeneous relationships and sociomaterial practices, necessary to bring this imagined future forward (cf. Brown, 2003; Dagiral and Peerbaye, 2016).

Thus, for the present study we understand the capitalisation of these public datasets as a sociotechnical performance that translates imaginaries and transforms resources into assets in order to realise the expectations of specific actors for value and capital, i.e. future earnings whether that is ‘money, or something comparable’ (Muniesa et al., 2017: 12; Birch, 2017a; Birch, 2019; Dussauge et al., 2015). Following Muniesa et al. (2017), we treat assetisation as a process that extends organisational boundaries to include social, technological, and economic infrastructuring of institutions, practices and individuals across society for the production of those resources deemed appropriate for capitalisation. Research services studied here mediate and service ‘across boundaries between different scales, levels, times and communities’ (Borup et al., 2006: 293; Dagiral and Peerbaye, 2016), i.e. between patients, GP practices, IT suppliers and researchers, for new, contemporary and future, ‘secondary data uses’.

As we show below, GPs (or other healthcare professionals) and the data they co-produce during consultations and treatments are disentangled from the local sociomaterial networks and transformed, via EHRs, from objects of clinical labour to objects and means of scientific inquiry and economics (cf. Denis and Goëta, 2017). It is through specific relational sociomaterial infrastructures that they are turned into intangible knowledge assets for a range of scientific and other purposes (cf. Dagiral and Peerbaye, 2016; Birch, 2017a; Muniesa et al., 2017; Birch, 2019). Their uninterrupted participation into the valuation practices of these research services facilitate their ceaseless assetisation (Wienroth et al., 2019) for the continuous capitalisation of NHS patient data and the financialisation of EHR data-driven research (cf. Muniesa et al., 2017; Birch, 2017a).

However, there is not a uniform approach to such performances. As Lilley and Papadopoulos (2014: 979) also remind us: ‘valuation is the outcome of a complex set of intra-actions inside the worlds in which the clashes of valuation unfold’. We see these research services as navigating through various political, economic and epistemological promises and valuations of NHS patient data-driven research. We echo Kleinman and Vallas’ (2001) observations that ultimately such technoscientific-intensive services ‘asymmetrically converge’ their market, academic and public healthcare expectations and logics in the process of transforming NHS patient data into ‘promissory assets’ for monetary circulations in knowledge economies (cf. Martin, 2015; Cooper and Waldby, 2014).

We assert that research services, such as the ones studied here, have already been hybridised as they pursue their organisational sustainability (and/or profitability) for the capitalisation of NHS patient data (cf. Kleinman and Vallas, 2001), albeit to various creative configurations of norms and practices. As they practice different *modi operandi* to financialise their operations (Birch, 2017a), which increasingly resemble those of other digital platforms

(cf. Geiger and Gross, 2019), different (non-)economic outcomes are made (im-)possible (cf. Waldby and Mitchell, 2006; Dussauge et al., 2015; Helgesson and Lee, 2017).

Central to our study, therefore, is how, during the process of assetisation, such research services embed specific value assumptions, structure practices, and (re-)configure relations of co-production (with GPs and patients). They (re-)order data flows, legitimise data uses and validate data users (Martin, 2015; Dagiral E and Peerbaye, 2016) for multiple purposes: creating their asset-based incomes via the licensing of curated datasets (cf. Birch, 2017); supporting the (further) development of healthcare knowledge products and services (e.g. scientific publications and quality improvement reports) (Dagiral and Peerbaye, 2016); pursuing their own research and other innovation projects (e.g. analytic scripts and phenotypes for GP practices and the research community). To explore and describe the ‘diverging registers of value’ (Dussauge et al., 2015) and the (in)commensurate valuations these research services are making and articulating, the asymmetries in knowledge practices (Tsoukas, 1997) and configurations of sociotechnical networks involved in the assetisation of NHS patient data, we introduce the situated concept of *asymmetric assetisation divergence*.

We use the above concept to consider, firstly, asymmetries between the operational (including financial) and, the scientific and public health logics and valuations that affect the assetisation (cf. Kleinman and Vallas, 2001; Dussauge et al., 2015). Secondly, we use it to consider whether the co-producers and providers of this *public* data, i.e. patients and healthcare professionals (GPs) have enough information and power to participate in the shaping and direction of this assetisation (Brown, 2003). Thirdly, to consider whether data assetisation can diverge (asymmetrically) because some of the processes involved can take place at different scales between these research services based on resources that are available to them, e.g. funding, access to data, human and technical resources, as well as networks of data providers

and customers. Following that, we reflect on the kind of new or existing asymmetries and inequalities of access to resources, information and knowledge research services may (re-)produce as they assetise NHS patient data (Brown, 2003).

Expectations of networked materialities and alternative epistemologies for divergent assetisations

This section explores which and how social and economic expectations, motivations and uncertainties are valued by stakeholders for the capitalisation of NHS patient data (cf. Tutton, 2011; Tarkkala et al., 2018). We focus on how the current state of affairs (Callon, 1984) in healthcare provision and research is problematised and we describe some of the political, economic and technoscientific arrangements in place for these research services to extract, prepare and circulate more and more data (Vezyridis and Timmons, 2017; Geiger and Gross, 2019). Lastly, we discuss how such research services, in their attempt to respond to the competition for data, resources, customers and epistemic impact, may differentiate their operations and diverge in their assetisation of NHS patient data (Barman, 2002).

Among researchers and *some* GPs, understandings of EHR data-driven research focused on expectations that by collecting and analysing data from all aspects of healthcare at a large scale *potentially* huge opportunities for the improvement of human health could be derived. The data tracking of individual events (e.g. treatments) related to patients, as they navigate the healthcare system, was portrayed as a new and unique approach to reconstruct medical histories and follow patient journeys ‘from cradle to the grave’. They asserted their determination to study ‘pretty much every disease that there is out there’ and make sense of the complexity of human health and illness based on the unprecedented availability of (decontextualised) information stored in EHRs (Tsoukas, 1997). By mobilising economies of scale and converging diversified datasets – from the many isolated and small data repositories

across (primary and secondary) care into fewer centralised ones – they anticipated the reduction of the time, cost, and the uncertainties involved with research and the provision of healthcare. By narrowing temporal, spatial and epistemological asymmetries in the collection and analysis of data, the speed of knowledge production was expected to increase as multiple reuses of these datasets from greater pools of participants and for a variety of purposes were being made possible:

...everything in a unified database, everything in one place, so it's accessible, it is structured, is very convenient...because you don't know what might be important later on down the line and to collect everything from time zero is a more efficient way of collecting data than collecting your data, then finding out that you need like seven new different measurements, and to go back and collect that would be very difficult, very costly so collecting everything with the hope that one day it might be useful to someone is a very bad idea (laughs) but it's efficient. (Researcher 2)

While for more 'traditional' epidemiological studies of rare or long-term conditions and treatments, researchers have to collect data from hundreds or thousands of individual participants, with routinely collected EHR data the number of research participants can be in the range of millions. It is the size, breadth and representativeness of the populations covered that constituted, in the eyes of these stakeholders, these datasets as unique resources for conducting observational studies in a much faster, cheaper and more 'pragmatic' way than ever before (Powell et al., 2017; Harwich and Lasko-Skinner, 2018). NHS patient data was, thus, valued and valorised by ascertaining the effectiveness of EHR data-driven research in the modelling of healthcare by other means: a very *promising alternative* approach to research and innovation of *similar* effectiveness.

While the assertion that ‘size is an advantage’, that science and society cannot just ignore, constituted a strong motivation for the EHR data-driven research communities, this is not to say that the significant issues of data quality and the susceptibility of these datasets to many biases were not acknowledged by data researchers (we will return to these issues in the next section). The growing concerns around the epistemological misconception that bigger is necessarily better, and critiques of the reproducibility of EHR data-driven research were accepted (Lipworth et al., 2017). However, the risks and uncertainties that this *asymmetrical* quality of EHRs introduce in EHR data-driven research (e.g. research bias) were considered as epistemological challenges (rather than actual barriers to EHR data use) that do not have to be addressed in full *before* pursuing the establishment of a research service for the assetisation of NHS patient data. There is not a uniform approach to be followed. For instance, some of these research services had proceeded more cautiously in this area and have been trying to recruit GP practices with ‘good’ coding practices, in order to then carefully curate the extracted data. Others worried less about data quality and more on finding a way to gather everything in one place *first* and deal with issues of data quality *later*. In any case, researchers in the field overwhelmingly anticipated the overcoming of such challenges soon enough and as datafication of healthcare and coding standardisation across the NHS continue to grow (cf. Brown, 2003).

...we've got this wealth of data that we can get hold of; we don't know quite how to use it. We must use it because there must be stuff in there that is going to be really valuable basically, and from [the IT supplier's] side, they knew that and they want it to be used.
(Researcher 9)

The above motivation that there *must* be a way forward with this data because now we *can*, which Fourcade and Healy (2016: 16) so aptly asserted as ‘the ceremonial aspect of the data

imperative’, was based not only on arguments of size, availability and effectiveness but also on an anticipatory uncertainty, for example, around (already) problematic numbers participating in research. Welsh and Wynne (2013: 543) have argued that ‘scientific authorities...declare the public meaning of technoscientific innovations and controversies to be matters of risk or science’. Here, we noted declarations around data assetisation, and by extension around the support of the work of such research services, as a matter of concern in overcoming (existential) risks *within* the life sciences, e.g. low trial recruitment rates (see Powell et al., 2017). The ‘reworking [of] epistemological asymmetries’ (Brown, 2003: 18), was, therefore, framed as a necessity for both society *and* the rest of the technoscientific communities ‘because 10 years from now we won’t have any other cohorts remaining’ and the future of population studies will be compromised. Consequently, for data researchers the development of new methodologies for EHR data-driven research are materialising their expectations of ‘getting all that we can out of this data’ and, continue to advance biomedical knowledge and improve healthcare against an uncertain epistemological future.

At the same time, there are also other stakeholders that come with their own expectations, valuations and role for the assetisation of this public asset. For example, there is the role of and expectations by state actors to drive and facilitate, via investment and regulatory frameworks, the assetisation and capitalisation of NHS patient data for the benefit of the national economy (cf. Timmons and Vezyridis, 2017). As one data researcher noted:

Government Ministers have been quite open about it, wanting to make the UK and the UK data sources world leaders in this kind of research, which usually means they want it to be a money spinner in some way or other. (Researcher 5)

For suppliers of primary care computer systems maintaining or contributing to a research service can materialise unique, albeit diversified, expectations. While our participants talked

asymmetric assetisation divergence

about certain research services that just want to capitalise EHR data and ‘licence just about everything it’s possible to licence’ to expand their presence in the biomedical research data market, they also identified several opportunities for the other stakeholders. For example, they discussed how the establishment of ‘knowledge transfer partnerships’ with academic institutions that can analyse these datasets has allowed them to advance their portfolio of services they provide to GP practices that purchase their system. GP practices have now been given a unique opportunity to contribute to epidemiological research and also to the development of new digital clinical decision support tools to improve healthcare. In exchange, they offer to such research services some new (and asymmetrical) advantages for the assetisation of NHS patient data, particularly when compared to smaller research services and institutions that rely on bespoke data extractions for research. Through their piecemeal arrangements with their contributing GP practices and other healthcare organisations, whole networks of local data providers and infrastructures for data extractions can now be maintained, making the continuous updating of research databases with new data a relatively seamless and near-real time process. These databases can then be managed by either the IT supplier exclusively, the IT supplier in conjunction with the academic or governmental partner, or exclusively by the academic partner.

For instance, CPRD and THIN provide quality improvement reports to collaborating GP practices. CALIBER has developed and curates a common pool of more than 50 computable phenotypes for researchers to identify and analyse the EHR of patients with a particular condition. QResearch has developed a number of risk calculators (e.g. for fractures, cancer, stroke, diabetes) (Hippisley-Cox et al., 2017). These and other research assets can then be provided to the research and GP communities (and some of them to the public) either free of charge and/or as licensed knowledge products via private limited companies (see <https://clinrisk.co.uk>). In the process of realising the values of EHR data-driven research,

research studies and publications that have used research services' data are also getting assetised and valorised under specific *targets* of 'knowledge transfer', increasing their reputational (and financial) value in this technoscientific market (cf. Barman, 2002) (Vezyridis and Timmons, 2016). For instance, QResearch has supported (as of March 2019) more than 165 peer-reviewed scientific publications, THIN more than 600 and, CPRD more than 2200 with targets of number of supported studies set annually (see MHRA, 2016).

Thus, there is not a single but multiple problematisations (Callon, 1984) for the assetisation of NHS patient data and the multiple, divergent, assets that can be made possible to fulfil common as well as diverse expectations and valuations of this public asset. On the other hand, expectations, no matter how ambitious or effective they might first appear in mobilising stakeholders and various forms of capital (e.g. human, technical, financial), they do not necessarily secure stable sociomaterial entanglements across science and/or the economy. While they become indispensable parts of data valuations and assetisations within particular communities, the gap between hype and reality has to be maintained by such research services in order to keep capitalisations going (cf. Brown, 2003).

Sociomaterial frictions and organisational uncertainties in asymmetrical asset productions

In this section, we look into how research services and researchers attempt to internalise and render manageable the various uncertainties and asymmetries involved in the production, assetisation and use of this data (cf. Star, 1985; Dagiral and Peerbaye, 2016; Denis and Goëta, 2017; Powell et al., 2017). Specifically, we explore how processes of assetisation reveal and mitigate diverse epistemological, economic, social and material 'frictions' (Edwards, 2010) in data production and analysis. They also shape how medical phenomena are expected to be coded so as to increase the quality and value of this asset for research (Petersen et al., 2019).

During our interviews, we were repeatedly exposed to data researchers' frustration with and, in some cases, limited understanding of, the way GPs code medical phenomena in EHRs. They often expressed their surprise, for example, at the fact that the codes GPs use to record even the same single disease could vary substantially from one GP practice to another. They also emphasised the impact such inadequate coding practices have on the overall quality and integrity of the datasets with which they work as well as the time and effort they have to put to mitigate it in preparation for analysis. As one researcher highlighted:

..a woman with prostate cancer, okay, fine, I have to remove this observation, it's kind of most obvious, but ... when you're thinking whether this person has osteoporosis or not, whether they have prescriptions, maybe these prescriptions are not right, maybe they're for cancer, not for osteoporosis ... When it's definitely bad data, you remove the observation. For example, if it's 50 people altogether when you use a sample of 50,000, that's okay, so it doesn't affect the conclusion. (Researcher 11)

This problem with data inconsistencies and redundancies was attributed by researchers and GPs mainly to the fact that the data is not collected primarily for research purposes. As they explained, data is collected to do the GP's clinical *and* 'business' work (cf. Petersen et al., 2019). Coding is, first, a practice highly contingent on the resources and effort (e.g. staff, time, information technology and skills) each GP practice can allocate for this purpose (Verheij et al., 2018). This is why research services often aim at recruiting mostly 'big' GP practices with 'good' management that can contribute patient data of good quality. At the same time, GP coding practices are also sensitive to the particularities and (unintentional) influence of the particular GP computer system in use (Verheij et al., 2018). As one experienced researcher described to us:

I'm sure you've seen those patterns of prescribing trends and how they differ nationally when actually it's just picked up that different clinical systems order their drugs differently, so obviously clinicians pick the drugs at the top of the screen and it's just picked up a difference in the clinical systems, rather than GPs' underlying behaviour, which was the same. (Researcher 10)

Moreover, GP computer systems allow the recording of structured ('coded') and unstructured patient (health and administrative) data ('free text') to diverse levels of granularity but without the use of a universally-adopted health coding scheme, like those used in hospitals (i.e. ICD-10). In effect, between regions within the UK different classification systems and versions of the same system (e.g. Read codes – version 2 and 3) have been adopted by the IT suppliers and, thus, GPs. To overcome this geographical clustering of systems and datasets (Kontopantelis et al., 2018), 'simply a side effect of where IT suppliers have been able to sell their software', a research service has to go to great lengths to map these coding schemes used by (primary and secondary) healthcare providers across the country into more 'formalised and universal structures' (cf. Denis and Goëta, 2017).

Thus, contrary to common misconceptions about the ready-made value of big health data, these extracted datasets are not just clinical artefacts ready for research after some technical translations of variables. And the availability of more EHR data does not necessarily make the NHS more 'transparent' to outsiders for all sorts of research purposes. Our interviews underline the situatedness of coding practices and how entangled they are with other diverse local sociotechnical practices, normativities and valuations beyond the clinical aspects of healthcare provision (cf. Denis and Goëta, 2017). In order to address and mitigate the uncertainty such 'knowledge asymmetries' (Tsoukas, 1997), between the observer (i.e. data researcher) and the observed (i.e. healthcare professional), introduce to the generalisability of

asymmetric assetisation divergence

any research outcome, research services also work on the introduction and mapping of more advanced classification systems (i.e. SNOMED CT). More standardisation of coding (of even more data) is expected to allow, in the future, the assetisation of more symmetrical and, therefore, more researchable data (Dagiral and Peerbaye, 2016).

Until then and in contrast to GPs and their familiarity with individual patient cases and coding practices, data researchers continue to be uncertain (and often less transparent in their scientific publications) as to whether any patterns in disease prevalence and trends they observe in the data should be attributed to the actual disease, demographic shifts, the patient management software, changes of clinical coding guidelines, preferences for information disclosure and coding during patient consultations or the particular pay-for-performance scheme, such as the Quality and Outcomes Framework (QOF), used to incentivise specific coding practices (Verheij et al., 2018). QOF has, in fact, been instrumental in improving data coding and collection across GP practices contracted by the NHS. Introduced in 2004, this point-based reward and incentive national programme for GPs, which currently constitutes up to 10% of their income, aims at standardising and improving the care they provide by having them meet specific quality targets in certain clinical areas (e.g. coronary heart disease, heart failure, stroke, diabetes mellitus, chronic obstructive pulmonary disease, cancer, dementia, depression and obesity). However, as Forbes et al. (2016) observed, QOF has not actually achieved its primary purpose as it conflated high performance with high quality of the provided care:

A few years ago they brought in QOF around depression, so that if you coded somebody as having depression, you then had to set on a pathway and had to do certain things by certain times, and if you didn't, you lost money. So there was a huge drop in anybody being diagnosed, coded as depression, even though that's what they may have been

treated for. So you can have knock-on effects. People also might ask not to be coded, because of the stigma, you don't want something in your medical record that says something that you think might have an adverse effect on, I don't know, insurance. So I think things aren't ever black and white, are they? (Researcher 9)

It is because of such intricacies of the entangled sociomaterial and economic practices of data coding for divergent purposes of use and often conflicting *local* assetisations and capitalisations (e.g. QOF remuneration) that researchers and GPs interviewed emphasised the considerable scientific and technical expertise required to assetise and use NHS patient data for research. Consequently, the assetisation of this data is not only about investments and maintenance of technologies of data storage, curatorial practices and analytical methods. It is also about investments in interdisciplinary teams of 'knowledge workers' (Kleinman and Vallas, 2001) who are not only experts in computer and data science but also in healthcare in order to make sense of the knowledge and practice that shape the coding of medical phenomena and, thus, the local production and use of datasets.

Among data researchers, however, there was also a wide acknowledgment of 'the huge shortage of skills across the board' to fully realise the values of these datasets. While (university) training was thought to be improving, issues around its quality and cost as well as 'letting the younger generations know that [health data science] is a career path that's available to them' remain a challenge. They were worried that considerable pitfalls may occur in this field in the *future* if training and recruitment are not addressed satisfactorily, particularly as more and more data providers are joining these – or developing their own – research services and more researchers 'with less and less skill' from other data-driven (e.g. financial) industries are getting access to these datasets. Other experienced researchers underlined the *asymmetries of capability* some (smaller) research services have been facing as they compete for expertise

and resources within this field. As EHR data-driven research is expanding and datasets collected ‘grow dramatically’ in size and variety (e.g. genomics), the acquisition and maintenance, for example, of ‘high performance computer systems’ to analyse these datasets becomes a significant challenge.

The skills required to design and model the systems are the same skills that are required to model systems in finance. And the salaries of science, they’re much lower, so it’s really difficult to find the right people to do the job. Plus the resources, a lot of the software, hardware...they’re under licence, for instance...and when you want to use some good software packages, in order to be able to do your work, in the academic and public sector, they’re not readily available, and that puts a barrier up of the things that you can do. (Researcher 2)

The process of assetising NHS patient data, therefore, reveals and involves several noteworthy asymmetries of information and practice. There are asymmetries regarding the maintenance of local sociomaterial infrastructures and processes, including GP practices, NHS patients and IT suppliers, capable enough to facilitate the production of datasets of acceptable levels of quality for research. The mitigation of such asymmetries (and uncertainties) often involves finding ways to overcome other processes of assetisation and capitalisation of data and coding (e.g. QOF, computer systems and their market distribution, consultation and coding preferences). Lastly, it requires the maintenance of appropriate research infrastructures (resources and skills) of enough capability to clean, prepare and guarantee quality datasets for use.

Politics, economics and societal valuations for asymmetrical asset flows

As we started identifying research services’ own normative understandings of how and by whom these datasets should be capitalised, we became interested in exploring the way they are gradually maintaining an essential, mediating, role (Callon, 1984) in the assetisation and

circulation of data. Here, our focus is on the particular political economic, social and professional environment within which they operate but also (re)enact to normatively increase the responsibility of other stakeholders (e.g. GPs, patients, regulators) for the disentanglement of NHS patient data from its local sites of production. We also explore some of the scientific, organisational, social and ethical asymmetries of this assetisation, produced as research services are re-entangled with data producers and other infrastructures required for the continuous extraction of data and the management of expectations, flows and values (Brown, 2003; Birch, 2017a).

As we showed previously, there has been considerable interest from all stakeholders to share and analyse NHS patient data for diverse purposes. However, notions of the (perceived) usefulness of EHR data-driven research were not shared uniformly across stakeholders, particularly when compared to more established types of scientific inquiry. For instance, data researchers were frustrated by the fact that few people outside their (academic) communities fully ‘understood’, and prioritised in their work the production and sharing of good quality data. This was often due to the fact that, as some GPs explained, the data GP practices provide to research services was not expected to come back to them in some ‘useful’ (to their practice) form, e.g. clinical guidelines for treating patients. While GPs generally acknowledged *some* of the capabilities of EHR data-driven research for improved healthcare, such as for healthcare planning and identification of public health trends, some of them were also more cautious about placing this type of research among the so called ‘cutting edge medicine’:

Certainly the capabilities of [EHR data-driven research] and the scope to look at information is great. I mean I think if you're looking at genuine treatments for diseases, the bottom line comes down to the hard graft with looking at molecules and cancer and

such. It comes down to laboratory technicians testing on cells and in petri dishes and microscopes...to try and work out and find drugs that are effective. (GP 3)

Moreover, the competitive environment of expectations and anticipations in which the capitalisation of NHS patient data takes place does not appear to be maintained only *between* the various groups of stakeholders that may benefit from it. The process of assetising intangible and reusable resources, within an increasingly marketised environment, reveals specific kinds of competition *within* particular groups of stakeholders. For instance, even within the data research communities there can be a combination of factors, that makes the sharing of this data difficult task across their networks. Issues that were often attributed to the marketisation of higher education and public healthcare, such as personal agendas and career aspirations, fears of litigation, intellectual property rights and competition for research funds, were thought to impede the establishment of a culture of openness in data sharing for the wider benefit of science.

One particular problematic situation that stood out during interviews was around the extent to which a research service could support an increasing number of applications for data access. The capabilities of research services to support other (especially non-affiliated) data researchers with data and expertise are not identical amongst them. Whether it is fees or research outputs some form of (mutually beneficial) collaboration and/or exchange has to take place to compensate for the work involved in preparing datasets for analyses.

It's true that people that have access to the databases do tend to set themselves up as gatekeepers and be a bit choosy about who they work with. Sometimes that's in furtherance of their own careers, other times it's because the data are complicated and they can't support every person to do a project, so inevitably you support the ones that you're interested in. (Researcher 5)

Contrary to some common societal expectations that any research team (with a project approved by the research services' *own* Independent Scientific Advisory Committee for its scientific benefits and information governance compliance) could access data, some of these research services have their own additional restrictions on eligibility. Various *asymmetries of collaboration* between research services and the wider academic and biomedical research communities shape divergent forms of NHS patient data assetisation. For example, CPRD can provide data to any research team from across the world as long as some of its members have experience working in UK primary care. QResearch, on the other hand, provides access to data mainly to teams from UK universities, with at least one member registered with the General Medical Council, while pharmaceutical companies are excluded unless the research project is about drug safety.

At the same time, the capitalisation of NHS patient data under an industrialised environment means that data researchers and research services are expected to internalise and enact the associated financial logics of EHR data-driven research in the UK. Research services operate usually (but not exclusively) on a not-for-profit basis. While they do not usually charge directly for the datasets, they do charge for the support they provide to researchers to access data (e.g. infrastructure, application and project set-up, training, data preparation and release). They may charge (fixed or variable) fees for individual projects or for their annual (commercial and non-commercial) licences of data access and support. For instance, CPRD's new pricing model charges £75,000 (excluding VAT) for its non-commercial (i.e. academic, government, charity) multi-study annual licence and more than four times that for the commercial one. CALIBER charges a fixed fee of £25,000 per approved project.

Since public information on how these cost-recovery charges are calculated by research services is limited, we can only assume that this valorisation of access is a flexibly interpretive

asymmetric assetisation divergence

activity for research services to sustain (financially) their operations and mitigate any *asymmetries of resources and capital* they experience for competitive and quality data assetisation. These include; direct financial support from research charities and commercial or governmental partners, research grant success rates, whether personnel and infrastructure is hosted by another organisation (e.g. University), physical location of the research service (e.g. corporate, governmental or academic buildings in London or the Midlands) and whether any other special financial arrangements with GP practices are in place (e.g. for the number of active patients they include in the database and/or for additional data beyond the EHR). In cases where researchers want to link GP records with hospital records and mortality data, these research services have (for an additional fee) to go through, the statutory trusted third party responsible for the handling and linking of NHS patient records (NHS Digital).

As demand for up-to-date datasets is constant, data assetisation requires not only new circulations of data to more users but also continuous data production facilitated by an appropriate regulatory framework for data researchers and research services to lawfully handle sensitive (NHS) patient data. This framework in England evolves around a rather outdated legislature piece of executive power that allows the Secretary of State for Health to override the common law duty of confidentiality and the need for patient consent for specified medical purposes (e.g. medical research) (see Section 251 of the NHS Act 2006 and Health Service Regulations 2002). While it has facilitated an unprecedented expansion of the so called ‘secondary use’ of NHS patient data, during the interviews, it was still problematised as an unavoidable but also necessary *burden within* organisational and scientific routines: a bureaucratic activity separate from the wider societal and historical context of unconsented medical research and part of the costs involved with the assetisation of these datasets.

Some researchers understood information governance in this area to be a ‘hygienic’ practice in ethics (Vezyridis and Timmons, 2019) of critical importance for minimising moral frictions and, thus, maintaining ‘social sustainability’ (cf. Brown and Michael, 2003; Tupasela, 2017) with the public and the healthcare professionals that co-produce the data. Others, acknowledging their limited understanding of the complicated legal framework surrounding the use of these datasets, suggested the introduction of new disciplines within research teams (e.g. information governance experts) for advising researchers on the legal and ethical legitimacy of research designs. Others directed their frustration towards regulators who could do more to ‘tighten up on things’ for faster (and cheaper) access to curated datasets and, therefore, to knowledge production, especially in an industry still operating under a fragmented framework of multiple (public and private) providers acting as ‘owners’ of these datasets:

..whenever we go to for a question to one of these [data providers], they say that they have to do the linkage for us ... so if you go to NHS they say they have to do the linkage, if you go to the GP data, they say they have to do it, and other sources like clinical registries for cancer they are not easy to access and even if you [have] a very good high quality protocol then the problem is how should we link the data, because they will not hand it to you. (Researcher 4)

Intertwined *financial and regulatory asymmetries of access* become more evident as data researchers often struggle to conduct their work within a competitive technoscientific environment that requires them to be at the forefront of *fast* and always *relevant* scientific knowledge production (Brown and Michael, 2003). While they ‘understand’ that research services ‘sell their product’, the ‘whacking costs’ for buying bespoke extractions and linked datasets, particularly by those not affiliated with these research services, have created considerable and asymmetrical (financial) barriers to research (Gilbert et al. 2015). In the

asymmetric assetisation divergence

dynamic field of healthcare, where practices and objects of study are constantly in flux (e.g. prescriptions of medications), data is not only ‘expensive to work with because of all the processes that are involved to generate it, keep it clean and keep it secure’ but it becomes ‘old’ quickly. Speed of knowledge production becomes of paramount importance for EHR data-driven research:

It all costs money and if you have money, then it's easier...I've been doing lots of studies without funding, because it's all for hot topics, and then by the time you get funding, it's gone, the public is not interested. But then if you have funding you can buy linkages to other sources, like a cancer registry or other registers. (Researcher 11)

Lastly, and notwithstanding some criticism on issues of medical confidentiality (Brown et al., 2010), these research services have not attracted the public outcry that other exploitations of these public datasets did in the past, namely NHS England’s defunct care.data (see Vezyridis and Timmons, 2017; Freeman, 2016), and Alphabet’s DeepMind unlawful contractual data arrangements with certain London NHS Trusts (see Powles and Hodson, 2017). Citizens interviewed were generally supportive of research that benefits the common good but were increasingly sceptical of NHS patient data exploitations that often take place away from the public eye (cf. Skovgaard et al., 2019; Vezyridis and Timmons, 2019). They also highlighted the NHS’s characteristic information ‘paternalism’ around NHS patient data capitalisations and expressed their frustration with the fact that information about such research services are ‘basically non-existent’. All this *asymmetry of information*, while it was still a great challenge for them, as NHS patients, to get access to their EHR and try to understand what is in there that such research services continuously assetise. For those few interviewees that were more aware of these research services, it was because of their own personal research and privacy concerns, especially after the aforementioned debacles. This is not to say that such research services do

not advocate openness and transparency about their operations. However, there is limited information to be found, for example, on their websites about which GP practices contribute to them with NHS patient data and/or which research projects they have approved (or rejected) for data access. As such, it is only through scientific publications that an interested citizen could conclude how NHS patient data is used by these research services, while having to rely solely on their GP practice to let them ‘know whether [they are] on these databases’ and how they could opt-out from a research service, if they so wish.

Thus, research services’ simultaneous assetisation of highly complex data and management of data assets flows becomes a relatively competitive and contradictory exercise in organisational stability (cf. Tupasela, 2017; Denis and Goëta, 2017; Geiger and Gross, 2019), introducing new or incorporating existing asymmetries in the capitalisation of NHS patient data and healthcare research: prioritisation of EHR data-driven research (over other disciplines), competition for funding and scientific career development, cumbersome information governance frameworks, gated access to data and linkages, competition for human and technological capital (within and beyond the sector), variable distribution of benefits, limited public awareness and engagement.

Conclusion

In this paper, we examined some of the expectations, frictions and uncertainties involved with the assetisation and capitalisation of NHS patient data (Birch, 2017a) by UK research services. Drawing on the sociology of expectations and (economic) STS literature, we brought into focus the performativity of these research services for EHR data-driven research to demonstrate the way they reconfigure practices, responsibilities and accountabilities for materialising specific promises of innovative healthcare research (cf. Brown, 2003; Petty and Heimer, 2011). For this, we treated these research services as heterogeneous assemblages (Callon, 1984) enacted in

order to assetise NHS patient data. We showed that this assetisation is both the outcome and the driver of various competing epistemic, and economic expectations and valuations (Brown, 2003; Birch, 2019). Divergent assets for healthcare and biomedical research are then produced via complex and situated practices that are enacted within a specific competitive and regulated environment increasingly entangled into (financial) industry expectations, logics and practices (Martin, 2015; Robinson, 2018). In this way, research services maintain a balance between establishing and expanding the health data science field, securing their organisational (including financial) sustainability, supporting biomedical and healthcare innovation and, contributing to the (knowledge) bio-economy, while also maintaining their public acceptability.

NHS patient data was found to be configured and used both as a continuously generated resource to be extracted from the NHS as well as an invaluable asset to be circulated in an asset-based biomedical knowledge economy for the enactment of multiple and divergent ontologies in healthcare and beyond (cf. Dagiral and Peerbaye, 2016; Denis and Goëta, 2017). These include observational studies, patient recruitment for research, pragmatic trials, clinical risk predictions, income and so on. As producers, users and brokers of such assets, at the intersection of the public healthcare, academic and biomedical sectors (Timmons and Vezyridis, 2017), these organisational entities are gradually becoming valued and valorised ‘obligatory passage points’ (Callon, 1984) in a social, political and economic shift of the ‘performative management of value’ (Birch, 2017a) of these public healthcare datasets and EHR data-driven research.

These research services are constantly attempting to realise and capitalise on their multiple epistemic, ethical, reputational, research service, platform and financial values out of (asymmetric and divergent) data assetisations (cf. Barrett et al., 2016). They not only converge

and translate expectations and promises but also construct new notions of productivity, adequacy, quality and transparency in EHR data-driven research (cf. Birch, 2019). They reconfigure EHR data-driven research as a (cost-)effective and more pragmatic alternative type of population health studies for a more promising and less risky future of improved healthcare (Brown, 2003; Borup et al., 2006). They also participate in; the reconfiguration of healthcare as a datafied technoscientific and ‘transparent’ practice (Tsoukas, 1997); the NHS as both the means and the subject of data labour; and the patient-citizen as a scientific, corporate, and moral resource for continuous data co-production (with the NHS) (cf. Brown, 2003). The numbers of contributing GP practices, sizes of populations covered, data fields, time periods included and speeds of updating their databases with new NHS patient data are valued and valorised. Information governance, including de-identification of EHRs, consent and public engagement are mobilised as a political economic apparatus (Birch, 2017b) of institutionalised altruism and volunteerism for minimising external moral frictions around the direction of NHS patient data assetisation.

In effect, they have assumed the overall role of ‘de-risking’ the costly assetisation of these datasets for actors in the academic and biomedical industries (cf. Brown, 2003; Cooper and Waldby, 2014; Robinson, 2018). They internalise the uncertainties and risks involved – and externalised by the NHS (cf. Robinson, 2018) – in valuating, creating, qualifying and mobilising the data assets as well as the capital of disjointed professional communities for shaping renewed promises of health and wealth (Tarkkala et al., 2018). They are pragmatic in the frictions and challenges they face around datasets’ completeness and integrity, access delays and costs (Powell et al., 2017), clustered geographical coverage of populations, varied cultures of collaboration and data sharing or inadequate public engagement. At the same time, however, they participate in the externalisation of other risks and uncertainties related to data assetisation back to the NHS and society at large. These include: costly datasets and

infrastructural incapacity, inadequate standardisation of coding practices and extractions, burdensome information governance for data extraction and linkage, limited public or professional acceptability, restricted funding and investments and, inadequate supplies of data scientists to produce data assets (cf. Denis and Goëta, 2017; Kleinman and Vallas, 2001). In this way, they ‘extensify’ valuations and, thus, they attempt to commit stakeholders in ever more data assetisations (cf. Lilley and Papadopoulos, 2014; Borup et al., 2006; Robinson, 2018).

Lastly, we found that this assetisation of NHS patient data is asymmetric in several noteworthy ways. First, although research services attempt to maintain their organisational (financial) sustainability and/or market dominance, the rents for these assets are seen as prohibitive for researchers conducting these observational studies (Gilbert et al. 2015). They now have to compete for funding in order to acquire increasingly diverse, and complex datasets, produced out of various asymmetries of coding and other knowledge practices (cf. Tsoukas, 1997; Barman, 2002). Asymmetries were also found with regards to the limited information and role GP practices and the public receive in order to participate in the shaping of the normative direction of NHS patient data capitalisation for the benefit of public healthcare. Finally, as these research services compete for data and customers, while embarking on their own scientific endeavours, they asymmetrically diverge in their assetisation of NHS patient data. Each service operates under different agreements with IT suppliers and other data providers across the country. In an attempt to differentiate their service for an intertwined epistemic and market dominance (Barman, 2002), they not only provide access to carefully curated datasets but also to various other (digital) assets of objectified information and knowledge about healthcare (cf. Tsoukas, 1997; Birch, 2019), such as: disease-specific code lists, phenotypes, biomarkers, quality improvement reports, recruitment pools of research

participants, clinical risk prediction algorithms, analytic scripts, data dictionaries, scientific publications and so on.

We conclude by arguing that the capitalisation of (NHS) datasets by such hybridised organisational entities (Kleinman and Vallas, 2001) may replace restrictions of access to data from the many individual organisational silos (e.g. GP practices) that hold disparate datasets to fewer and bigger silos where data is converged into licensed databases out of closed-source, proprietary clinical systems. Data assetisation continues to grow bigger and more diverse in size, complexity and scope (e.g. genomics, wearables and medical imaging) under a political-economic environment of industrialised scientific competition and production of expectations, assets and capital (Brown, 2003). Financial and other epistemic restrictions of access to NHS patient data-driven research may, in the end, stabilise an oligopolistic ‘data access market’, attractive and affordable only to those actors that have the capital and the sociomaterial infrastructure in place to ‘buy’ access to data and undertake this kind of research at the required scale and depth (Powles and Hodson, 2017).

We suggest that checks and balances regarding research services’ ontological, epistemological and ethical role as biomedical knowledge makers and brokers for the benefit of society at large should be considered. Accountable requirements for the fulfilment of their ‘long-term promises’, rather than of their short-term return-on-investment opportunities, should be elaborated for transparent, inclusive and equitable research agendas and knowledge-making practices. This is especially urgent at a time when the role of EHR data-driven research is gradually moving beyond the concept of ‘secondary use’ and onto the longitudinal management of population health and the contemporary planning of (national) health systems.

Acknowledgments

We are grateful to the study participants for their time and patience. We would also like to thank the SSS reviewers and, the editors Adam Hedgecoe and Sergio Sismondo.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was funded by the European Commission (H2020-MSCA-IF-EF-2014-659478).

References

- Barman EA (2002) Asserting Difference: The Strategic Response of Non-profit Organizations to Competition. *Social Forces* 80(4): 1191–1222.
- Barrett M, Oborn E, Orlikowski W (2016) Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. *Information Systems Research* 27(4): 704–723.
- BIGT (2003) *Bioscience 2015: Improving national health, increasing national wealth. Report of the Bioscience Innovation and Growth Team*. London: Department of Trade and Industry. Available at: <https://web.archive.org/web/20120107081132/http://www.bioindustry.org/document-library/bioscience-2015/1bia-1103-bioscience-2015.pdf> (accessed 15 October 2015).
- Birch K (2017a) Rethinking *Value* in the Bio-economy: Finance, Assetization, and the Management of Value. *Science, Technology, & Human Values* 42(3): 460–490.

- Birch K (2017b) The problem of bio-concepts: biopolitics, bio-economy and the political economy of nothing. *Cultural Studies of Science Education* 12(4): 915–927.
- Birch K (2019) Technoscience Rent: Toward a Theory of *Rentiership* for Technoscientific Capitalism. *Science, Technology, & Human Values*: 016224391982956.
- Borup M, Brown N, Konrad K, et al. (2006) The sociology of expectations in science and technology. *Technology Analysis & Strategic Management* 18(3–4): 285–298.
- Brown I, Brown L and Korff D (2010) Using NHS Patient Data for Research Without Consent. *Law, Innovation and Technology* 2(2): 219–258.
- Brown N (2003) Hope against hype-accountability in biopasts, presents and futures. *Science & Technology Studies* 16(2): 3–21.
- Brown N and Michael M (2003) A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis & Strategic Management* 15(1): 3–18.
- Callon M (1984) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review* 32(1_suppl): 196–233.
- Cooper M and Waldby C (2014) *Clinical Labor: Tissue Donors and Research Subjects in the Global Bioeconomy*. Experimental futures. Durham: Duke University Press.
- Dagiral E and Peerbaye A (2016) Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. *Science & Technology Studies* 29(2): 44–61.
- Denis J and Goëta S (2017) Rawification and the careful generation of open government data. *Social Studies of Science* 47(5): 604–629.

Department of Health (2011) *Innovation Health and Wealth, Accelerating Adoption and Diffusion in the NHS*. December. Leeds.

du Preez D (2015) NHS director argues big data in healthcare is a ‘moral obligation’ – care.data relaunched. *diginomica*, 20 January. Available at:

<https://web.archive.org/web/20190306022832/https://government.diginomica.com/2015/01/20/nhs-director-argues-big-data-healthcare-moral-obligation-care-data-relaunched/> (accessed 28 March 2019).

Dussauge I, Helgesson C-F, Lee F, et al. (2015) On the omnipresence, diversity, and elusiveness of values in the life sciences and medicine. In: Dussauge I, Helgesson C-F, and Lee F (eds) *Value Practices in the Life Sciences and Medicine*. Oxford University Press, pp. 1–28.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, Mass: MIT Press.

Forbes L, Marchand C and Peckham S (2016) *Review of the Quality and Outcomes Framework in England: Final Report 2016*. London: Policy Research Unit in Commissioning and the Healthcare System (PRUComm). Available at: <https://web.archive.org/web/20190901185357/http://blogs.lshtm.ac.uk/prucomm/files/2017/02/Review-of-QOF-21st-December-2016.pdf> (accessed 30 December 2018).

Fourcade M and Healy K (2016) Seeing like a market. *Socio-Economic Review* 15(1): 9-29.

Freeman G (2010) House of Commons debates: Policy for growth. Available at: <https://web.archive.org/web/20190901190645/https://publications.parliament.uk/pa/cm201011/cmhansrd/cm101111/debtext/101111-0003.htm> (accessed 5 March 2019).

Freeman G (2016) Review of health and care data security and consent. Available at:

<https://web.archive.org/web/20190725165037/https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent> (accessed 7 July 2016).

Geiger S and Gross N (2019) A Tidal Wave of Inevitable Data? Assetization in the Consumer Genomics Testing Industry. *Business & Society* DOI: 10.1177/0007650319826307.

Gilbert R, Goldstein H and Hemingway H (2015) The market in healthcare data. *BMJ*: h5897. DOI: 10.1136/bmj.h5897.

Harwich E and Lasko-Skinner R (2018) *Making NHS data work for everyone*. December.

London: Reform. Available at:

<https://web.archive.org/web/20190729101454/https://reform.uk/sites/default/files/2018-12/NHSdata%20-%20no%20gif%20SP.pdf> (accessed 7 March 2019).

Helgesson C-F and Lee F (2017) Valuations as Mediators Between Science and the Market:

How Economic Assumptions Shape Pharmaceutical Trial Designs. *Science as Culture* 26(4): 529–554.

Hippisley-Cox J, Coupland C and Brindle P (2017) Development and validation of QRISK3

risk prediction algorithms to estimate future risk of cardiovascular disease:

prospective cohort study. *BMJ*: j2099.

HM Government (2018) *Industrial Strategy: Life Sciences Sector Deal 2*. Crown. Available

at:

https://web.archive.org/web/20190803200430/https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/768570/life-sciences-sector-deal-2-web-ready-update.pdf (accessed 10 May 2019).

HM Treasury (2018) *Getting smart about intellectual property and other intangibles in the public sector: Budget 2018*. October. London: HM Treasury. Available at: https://web.archive.org/web/20190724073411/https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/752003/Getting_smart_about_intellectual_property_and_other_intangibles_in_the_public_sector_-_Budget_2018.pdf (accessed 8 March 2019).

Hogarth S (2017) Valley of the unicorns: consumer genomics, venture capital and digital disruption. *New Genetics and Society* 36(3): 250–272.

Kelsey T and Cavendish W (2014) *Personalised Health and Care 2020: Using Data and Technology to Transform Outcomes for Patients and Citizens. A Framework for Action*. National Information Board. Available at: https://web.archive.org/web/20190802124912/https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf (accessed 28 March 2019).

Kleinman DL and Vallas SP (2001) Science, capitalism, and the rise of the “Knowledge Worker”: The changing structure of knowledge production in the United States. *Theory and Society* 30(4): 451–492.

Kontopantelis E, Stevens RJ, Helms PJ, et al. (2018) Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open* 8(2): e020738.

Lilley S and Papadopoulos D (2014) Material Returns: Cultures of Valuation, Biofinancialisation and the Autonomy of Politics. *Sociology* 48(5): 972–988.

- Lipworth W, Mason PH, Kerridge I, et al. (2017) Ethics and Epistemology in Big Data Research. *Journal of Bioethical Inquiry* 14(4): 489–500.
- Martin P (2015) Commercialising neurofutures: Promissory economies, value creation and the making of a new industry. *BioSocieties* 10(4): 422–443.
- MHRA (2016) *Medicines and Healthcare products Regulatory Agency Annual Report and Accounts 2015/16*. London, UK: Medicines and Healthcare Products Regulatory Agency. Available at:
https://web.archive.org/web/20190725154058/https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/539679/MHRA_annual_report_and_accounts_2015_to_2016.pdf (accessed 25 July 2019).
- Muniesa F, Doganova L, Ortiz H, et al. (2017) *Capitalization: A Cultural Guide*. Sciences sociales. Paris: Mines ParisTech.
- NHS Digital (2018) Data Access Request Service (DARS) performance reporting. Available at:
<https://web.archive.org/web/20190901190130/https://digital.nhs.uk/services/research-advisory-group/data-access-request-service-dars-performance-reporting> (accessed 8 March 2019).
- Petersen I, Welch CA, Nazareth I, et al. (2019) Health indicator recording in UK primary care electronic health records: key implications for handling missing data. *Clinical Epidemiology* 11: 157–167.
- Petty J and Heimer CA (2011) Extending the rails: How research reshapes clinics. *Social Studies of Science* 41(3): 337–360.

- Powell GA, Bonnett LJ, Tudur-Smith C, et al. (2017) Using routinely recorded data in the UK to assess outcomes in a randomised controlled trial: The Trials of Access. *Trials* 18(1): 389.
- Powles J and Hodson H (2017) Google DeepMind and healthcare in an age of algorithms. *Health and Technology* 7(4): 351–367.
- Robinson MD (2018) Financializing epistemic norms in contemporary biomedical innovation. *Synthese*. DOI: 10.1007/s11229-018-1704-0.
- Skovgaard LL, Wadmann S and Hoeyer K (2019) A review of attitudes towards the reuse of health data among people in the European Union: the primacy of purpose and the common good. *Health Policy*. 123(6): 564-571.
- Star SL (1985) Scientific work and uncertainty. *Social Studies of Science* 15(3):391–427.
- Tarkkala H, Helén I and Snell K (2018) From health to wealth: The future of personalized medicine in the making. *Futures* 109: 142-152.
- Timmons S and Vezyridis P (2017) Market-driven production of biospecimens and the role of NHS hospital-led biobanks. *Sociology of Health & Illness* 39(7): 1242–1257.
- Tsoukas H (1997) The tyranny of light: The temptations and the paradoxes of the information society. *Futures* 29(9): 827-843.
- Tupasela A (2017) Data-Sharing Politics and the Logics of Competition in Biobanking. In: Pavone V and Goven J (eds) *Bioeconomies*. Cham: Springer International Publishing, pp. 187–206.

- Tutton R (2011) Promising pessimism: Reading the futures to be avoided in biotech. *Social Studies of Science* 41(3): 411–429.
- Verheij RA, Curcin V, Delaney BC, McGilchrist MM (2018) Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *Journal of Medical Internet Research* 20: e185.
- Vezyridis P and Timmons S (2016) Evolution of primary care databases in UK: a scientometric analysis of research output. *BMJ Open* 6(10): e012785.
- Vezyridis P and Timmons S (2017) Understanding the care.data conundrum: New information flows for economic growth. *Big Data & Society* 4(1): 205395171668849.
- Vezyridis P and Timmons S (2019) Resisting big data exploitations in public healthcare: free riding or distributive justice?. *Sociology of Health & Illness* 41(8): 1585-1599.
- Wayman C and Hunerlach N (2019) *Realising the value of health care data: a framework for the future*. London, UK: EY. Available at: https://web.archive.org/web/20190828180130/https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/life-sciences/life-sciences-pdfs/ey-value-of-health-care-data-v20-final.pdf (accessed 28 August 2019).
- Welsh I and Wynne B (2013) Science, Scientism and Imaginaries of Publics in the UK: Passive Objects, Incipient Threats. *Science as Culture* 22(4): 540–566.
- Wienroth M, Pearce C and McKeivitt C (2019) Research campaigns in the UK National Health Service: patient recruitment and questions of valuation. *Sociology of Health & Illness* DOI: 10.1111/1467-9566.12957.

asymmetric assetisation divergence

World Economic Forum (2011) *Personal Data: The Emergence of a New Asset Class*. World

Economic Forum (WEF). Available at:

https://web.archive.org/web/20190803001919/http://www3.weforum.org/docs/WEF_I

[TTC_PersonalDataNewAsset_Report_2011.pdf](#) (accessed 15 March 2019).