Transferring Health Big Data within the European Legal Framework: What Role for National Healthcare Services?

Davide Golinelli, Fabrizio Toscano, Andrea Bucci and Gherardo Carullo*

The main objective of this article is to describe the legal principles governing the selection by European public authorities, such as National Health Services (NHS) of third parties, when entering into agreements for the transfer of health data. According to Directive 2003/98/EC, and in light of the provisions of the Treaties of the European Union, the choice as to how a public authority makes its data available to third parties needs to be transparent, non-discriminatory and may not in any case benefit a specific company at the expense of others. For this reason, we maintain that a hypothetical agreement by which a public authority grants exclusive access to a large amount of health data to a private company selected with non-transparent criteria appears highly questionable. We advocate that the NHS should adopt more appropriate data policies aimed at promoting the sustainability of the NHS, following the legal framework analysed in this article.

INTRODUCTION

The term "Big Data" (BD) in health care refers to volumes of large, complex, linkable information[1], that can be defined as routinely or automatically collected datasets, which are electronically captured and stored[2]. It is re-usable in the sense of multipurpose data[3] and comprises the fusion and connection of existing databases for the purpose of improving health and health system performance.[4]

---

* Davide Golinelli: Department of Biomedical and Neuromotor Sciences, Alma Mater Studiorum, University of Bologna, Bologna, Italy. Fabrizio Toscano: Post-Graduate School of Public Health, Department of Molecular and Developmental Medicine, University of Siena, Siena, Italy; currently Visiting Fellow at the Department of Healthcare Policy and Research, Weill Cornell Medical College, New York, USA. Andrea Bucci: Department of Economics and Social Sciences, Marche Polytechnic University, Ancona, Italy. Gherardo Carullo: Department of Italian and Supranational Public Law, University of Milan, Italy.

[1] I de la Torre Díez et al, "Big Data in Health: A Literature Review from the Year 2005" (2016) 40 *Journal of Medical Systems* 209.

[2] A Laschkolnig et al, European Commission, Directorate-General for Health and Food Safety, *Study on Big Data in Public Health, Telemedicine and Healthcare: Final Report* (Publications Office, Luxembourg, 2016) <http://dx.publications.europa.eu/10.2875/734795>.

[3] R Margolis et al, "The National Institutes of Health's Big Data to Knowledge (BD2K) Initiative: Capitalizing on Biomedical Big Data" (2014) 21 *Journal of the American Medical Informatics Association* 957.

[4] N Brennan et al, "Leveraging the Big-data Revolution: CMS Is Expanding Capabilities to Spur Health System Transformation" (2014) 33 *Health Affairs (Millwood)* 1195.

[5] Weldon D, *Big Data Investments to Top $57B in 2017* (25 October 2017) <https://www.information-management.com/news/big-data-investments-to-top-57b-in-2017>.

From an economic point of view, global spending on BD technologies surpassed $US57 billion in 2017 and will increase by at least 10% per year up to 2020.[5] Collecting data is easier than ever and implementing tools to work with BD has also become much more accessible.

Through BD-processing software, it is possible to group huge and complex information, which can be analysed by few operators. This represents an opportunity and an important tool for health care. As a matter of fact, in this sector BD applications can target many different fields (individual and personalised care, clinical research, epidemiology, health services research, etc) through prospective data monitoring or retrospective data analysis.[6] These innovative tools have the potential to help to improve health care by providing insights into the causes and outcomes of diseases, better drug targets for precision medicine and enhanced disease prediction and prevention.[7]

When considering the use of BD in health care it is crucial to understand what the sources of data and the processes behind their collection are.[8] There are many potential sources of data related to health care[9]: routinely recorded health data, pharmaceutical data, genomic data, wearable devices, social media and internet cookies, and the internet of things, just to name a few.[10] This vast number of data sources reflects the numerous players involved in the BD challenge in health care (Table 1) and may entail a potential conflict between them. Two recent events suggest the need for a deeper analysis of the connection between different BD players and national health care services (NHS).

TABLE 1. Big Data players in health care: National health care service; IBM: International Business Machines Corporation.

Category	Players
Traditional players	Pharmaceutical firms, Hospitals and Medical-technology companies
Forced players	Health insurers, Pharmacy-benefit managers, and Single-payer national health care services such as Britain NHS or Italian Servizio Sanitario Nazionale.
New players	Amazon, IBM, Google, Apple, etc.

In the United Kingdom, Alphabet Inc. (the parent company of Google Inc.), through its British-based artificial intelligence subsidiary (DeepMind Technologies Limited, DeepMind), reached an agreement with the Royal Free London NHS Foundation Trust (Royal Free) between July 2015 and October 2016. Royal Free agreed to transfer to DeepMind identifiable patient records, without obtaining the prior explicit consent of patients, so that the latter may use such data for developing a clinical alert app for kidney injury.

Julia Powles and Hal Hodson have highlighted several critical aspects of the agreement.[11] They have concluded that existing institutional and regulatory responses are insufficiently robust and agile to respond properly to the challenges presented by data politics and the rise of algorithmic tools in health

---

[6] de la Torre Díez, n 1; CS Kruse et al, "Challenges and Opportunities of Big Data in Health Care: A Systematic Review" (2016) 4 *JMIR Medical Informatics* e38.

[7] MJ Khoury and JPA Ioannidis, "Big Data Meets Public Health" (2014) 346 *Science* 1054.

[8] de la Torre Díez, n 1; Kruse, n 6; ID Dinov, "Methodological Challenges and Analytic Opportunities for Modeling and Interpreting Big Healthcare Data" (2016) 5 *GigaScience* [Internet] <https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-016-0117-6>.

[9] H Chang and M Choi, "Big Data and Healthcare: Building an Augmented World" (2016) 22 *Healthcare Informatics Research* 153.

[10] DV Dimitrov, "Medical Internet of Things and Big Data in Healthcare" (2016) 22 *Healthcare Informatics Research* 156.

[11] J Powles and H Hodson, "Google DeepMind and Healthcare in an Age of Algorithms" (2017) 7 *Health and Technology* 351.

[12] Powles and Hodson, n 11; Information Commissioner's Office, *Royal Free – Google DeepMind Trial Failed to Comply with Data Protection Law* (July 2017) <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/>.

care. Indeed, in July 2017 the United Kingdom data protection authority ruled that DeepMind was unlawfully given the health records of 1.6 million patients.[12]

In Italy, in March 2016 IBM announced plans to launch its first Watson Health European Center of Excellence in Milan, near the Human Technopole Italy 2040 research campus. In doing so, IBM would financially support the Italian government initiative to establish an international hub for the advancement of genomics, BD, ageing and nutrition.[13] At the time of writing, the details of the agreement remain unknown.

According to the available information, it appears that the economic investment by IBM would be subordinated to the acquisition of Lombardia region residents' health data (10,019,166 population in 2014), the so-called "Protected Health Information", including "health care-related data", "personal clinical health records" and "nominal or anonymous fiscal information".[14]

The outcome of the agreement could represent both an opportunity and a risk. Hypothetically, if the tools developed by IBM are integrated into the Italian NHS, in a way that is economically viable and favourable for the latter, there could indeed be mutual benefits. However, if the divergent interests of IBM and of the Italian NHS prevent such a positive result, considering the different mission of the two players, critical health data may be exposed without any actual benefit for Lombardia region residents.

PERSPECTIVE AND AIM OF THE ARTICLE

The examples of IBM/Italy and DeepMind/United Kingdom bring to light the need for the NHS – and for all the subjects involved – to take into due consideration the legal framework within which agreements on the transfer and re-use of health data can be entered upon.

For this reason, the main objective of this article is to describe the legal principles governing the selection by European public authorities (such as NHS) of third parties when entering into agreements for the transfer of health data.
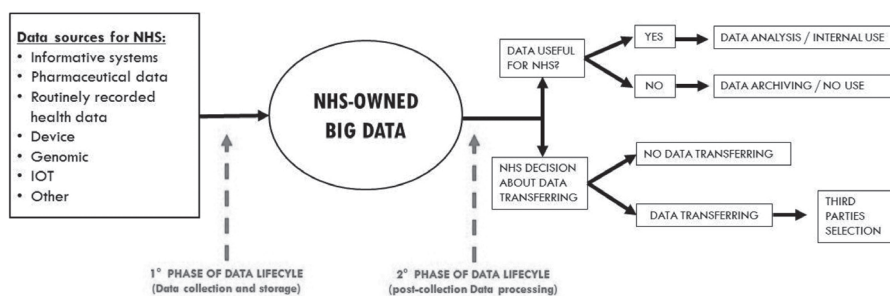
Having defined such legal principles, the article discusses the role of the NHS in the challenge of BD exploitation. The need for updated, adequate and innovative data policies is a very current topic, and the debate on data-related processes and their lifecycle (Figure 1), particularly in health care, is intense. This phenomenon is well evidenced by the relevance of health data in Regulation 2016/679/EU on the protection of personal data (GDPR). Several provisions of the GDPR are aimed at informing and regulating policies on the use of health data.[15] Given the limits established by the GDPR for processing health data, this article takes for granted that data can be transferred to third parties only insofar as permitted by privacy regulations. For this reason, aspects regulated by the GDPR (eg data subjects' consent, anonymisation, data safety, etc) will not be considered. The contribution of the article is to highlight the further – and less obvious – legal provisions that regulate the selection of third parties to whom health data is to be transferred. In other words, the article focuses on the second phase of the data lifecycle (Figure 1), the one in which public authorities (ie data owners, such as the NHS) are in a position to decide whether or not to transfer their data to third parties.

FIGURE 1. Structure of the data lifecycle in health care, from data collection to use and/or transfer. NHS: National health care service; IOT: Internet of things.

---

[13] IBM News, *IBM Plans First Watson Health European Center of Excellence in Italy* (31 March 2016) <https://www-03.ibm.com/press/us/en/pressrelease/49436.wss>.

[14] C Poggi, *IBM Looking to Help P.A. Enter New Digital Age, GM Cereda Says* (9 September 2016) <http://www.italy24.ilsole24ore.com/art/markets/2016-09-08/interview-cereda-ibm-113713.php?uuid=AD3cz2GB>.

[15] See, eg, GDPR Art 9, para 2.

Data sources for NHS:
- Informative systems
- Pharmaceutical data
- Routinely recorded health data
- Device
- Genomic
- IOT
- Other

NHS-OWNED BIG DATA

DATA USEFUL FOR NHS?

YES → DATA ANALYSIS / INTERNAL USE

NO → DATA ARCHIVING / NO USE

NHS DECISION ABOUT DATA TRANSFERRING

NO DATA TRANSFERRING

DATA TRANSFERRING → THIRD PARTIES SELECTION

1° PHASE OF DATA LIFECYLE
(Data collection and storage)

2° PHASE OF DATA LIFECYLE
(post-collection Data processing)

TRANSFER AND RE-USE OF DATA IN THE EUROPEAN LEGAL FRAMEWORK

Data held by public authorities are increasingly treated by European law as a raw material. Directive 2003/98/EC (the Directive), provides a comprehensive regulation of the re-use of public sector information (ie data), and in doing so its fifth recital expressly states that "public sector information is an important primary material for digital content products and services".[16] As provided by Art 1, para 1, the aim is to establish "a minimum set of rules governing the reuse and the practical means of facilitating re-use of existing documents held by public sector bodies of the Member States (MS)".

According to Art 11, para 1, of the Directive, rules for accessing documents shall be based on "fair, proportionate and non-discriminatory conditions". Additionally, "contracts or other arrangements between the public sector bodies holding the documents and third parties shall not grant exclusive rights". The second paragraph of Art 11 clarifies that exclusive agreements may be entered upon only if "necessary for the provision of a service in the public interest". Recital 20 of the Directive explains that "this may be the case if no commercial publisher would publish the information without such an exclusive right". In such a case, Art 11, para 2, provides that "the validity of the reason for granting such an exclusive right shall be subject to regular review, and shall, in any event, be reviewed every three years". Moreover, the agreement "shall be transparent and made public", to disclose the reasons which prompted the MS to adopt an exclusive deal.

However, the Directive applies only to data that MS had made accessible. As clarified by recital 9, the "Directive does not contain an obligation to allow re-use of documents", and it is left to MS the choice of whether making data accessible. In other words, the Directive is applicable to the information held by public authorities only if it has been made accessible.

One could therefore doubt that the Directive applies to the agreement between the Italian Government and IBM, since the data being transferred to the latter have never been made accessible and have been granted solely to one company. In light of the spirit of the Directive, however, it must be maintained that the provisions regulating the limits within which exclusive agreements can be entered into must apply to exclusive agreements on the transfer of non-accessible data, like the one between the Italian Government and IBM.

This latter conclusion is indeed confirmed by the recent proposal for a revision of the Directive.[17] According to the explanatory memorandum by the European Commission, Art 12 of the proposal – corresponding to Art 11 in the current numbering[18] – has been amended precisely "to specify that the

---

[16] Official Journal of the European Union, Directive 2003/98/EC of the European Parliament and of the Council (17 November 2003).

[17] Proposal for a Directive of the European Parliament and of the Council on the re-use of public sector information, COM (2018) 234 final (25 April 2018).

[18] A new Art 10 of the proposal will shift subsequent numberings by one.

[19] COM (2018) 234 final [11].

[20] *AG2R Prévoyance v Beaudout Père et Fils SARL* (Court of Justice of the European Union, First Chamber, C-437/09, 3 March 2011) [24]–[25].

prohibition of exclusive arrangements shall also extend to such arrangements that do not expressly grant an exclusive right in the re-use of documents, but may lead to a situation where access is limited to one or very few re-users".[19]

Regardless, even if the Directive were to be deemed not applicable, European law could still preclude an agreement like the one at issue. According to the case law of the European Court of Justice, "Article 101 TFEU *[Treaty on the Functioning of the European Union]*, read in conjunction with Article 4(3) EU, Member States are required not to introduce or maintain in force measures, whether legislative or regulatory, which may render ineffective the competition rules applicable to undertakings" and that "under Article 106(1) TFEU […], in the case of public undertakings and undertakings to which Member States grant special or exclusive rights, Member States may neither enact nor maintain in force any measure contrary to the rules contained in the Treaties, in particular to those rules provided for in Article 18 TFEU and in Articles 101 TFEU to 109 TFEU, subject to Article 106(2) TFEU".[20]

Under such case law, exclusive agreements between public authorities and undertakings can be entered into only within the limits of Art 106, para 2, TFEU, that is only if the company is "entrusted with the operation of services of general economic interest", and only "insofar as it is necessary for performing the particular tasks assigned to them".

Thus, it can be maintained that the choice on how a public authority, such as a NHS, makes available to third parties, for commercial purposes, its data – even if Directive 2003/98/EC were to be deemed not applicable – still needs to be transparent and non-discriminatory and may not in any case benefit a specific company at the expense of others.

In conclusion, an agreement by which a public authority grants exclusive access to a large number of health and personal data to a company selected with non-transparent criteria, appears highly questionable, whether under Directive 2003/98/EC, or in light of the fundamental principles and provisions of the EU Treaties.

WHAT ROLE FOR NATIONAL HEALTH CARE SERVICES?

In light of the legal framework highlighted in the previous paragraph, which in essence requires public authorities to act in a transparent and non-discriminatory manner when transferring health data, some questions may arise: What should be the role played by NHS in the BD challenge? What could be the risks related to BD exploitation and use?

A first consideration relates to the definition of NHS' primary interest. In fact, both United Kingdom and Italian NHS, as virtually any NHS, aim at preserving their communities' health. As a consequence, agreements such as those cited above can be deemed admissible only insofar as they are proved to be functional to an NHS' interests.

Additionally, given that the economic sustainability of health care systems represents a concrete and current challenge for governments, every opportunity of self-financing should be taken into consideration. This includes properly assessing the possibility to benefit economically from a transparent and lawful transfer of data to third parties.

Single-payer NHSs should avoid acting passively – as in the cases of Italy and United Kingdom – and should act in a structured and consistent manner, by adopting stronger policies regarding health data creation, analysis and trade, contributing to their own financial sustainability. This can be done through the creation and/or implementation of NHS-owned data-platforms or data-warehouses, which can be made accessible to third parties for data transfer, re-use or analysis. Such an approach has been adopted, for example, in the United States by the Centers for Medicare and Medicaid Services, which has made available the claims data to researchers and analysts. The availability of claims data, moreover, has

---

[21] H Kharrazi et al, "Comparing Population-based Risk-stratification Model Performance Using Demographic, Diagnosis and Medication Data Extracted From Outpatient Electronic Health Records Versus Administrative Claims" (2017) 55 *Medical Care* 789.

[22] AB Frakt and SD Pizer, "The Promise and Perils of Big Data in Health Care" (2016) 22 *American Journal of Managed Care* 98.

generated a significant interest in the composition and implementation of the aforementioned claims data with the increasing adoption of electronic health records across the country. The combination of the two, in fact, constitutes an interesting set of BD not only to answer the increasingly value-based payment policies, but also to provide risk stratification and predictive modelling tools.[21]

Another consideration calls into question the quality of re-usable data and of its analysis. In fact, an ambiguous interpretation of BD can lead to potential risks for both individuals and populations and may consequently affect the sustainability of NHS. It could in particular lead to over- or under-use of health services and goods.

While gathering data is extremely important, BD players should evaluate how to use these new assets correctly and effectively.[22] In fact, although new tools to create and analyze data are in the process of being implemented and perfected, letting data speak for themselves could lead to unrealistic relationships between outcomes and predictors, due to the so-called "spurious" correlation.

A real-world example of this misapprehension is provided by Google Flu Trends, which gives estimates of influenza spread by aggregating Google search queries. Google Flu Trends drastically overestimated 2013 peak flu levels using flu-related Internet searches[23].

All of the above suggests that one mandatory point in question in BD analysis is the need to interpret and/or map the proper dimensions of data correctly to extrapolate their meaning in specific settings and contexts, to protect the health of individuals and populations and to ensure the economic sustainability of NHS.[24]

CONCLUSIONS

This article illustrated the European legal framework within which BD players in the field of health care should work when they intend to co-operate with national public authorities.

According to the rules of national law, where public authorities decide to make their health or clinical data accessible, they will have to do so through transparent and predetermined criteria, to ensure that all economic operators are guaranteed equal access to data, without unjustified preferential or exclusive treatment. Moreover, NHS must favour the interests of patients, both as individuals and as part of a community. It follows that the legal criteria mentioned above should ensure that data are released in a manner consistent with the health protection mission pursued by NHS.

New thinking and a deeper debate about the role of NHS in the challenge of BD is needed to preserve the sustainability of NHS. This also raises the economic elements of the operation. National and local public authorities should take into due consideration the monetisation opportunities arising from the exploitation of BD, by adopting long-term structured data policies aimed at promoting the sustainability of national healthcare services, following the legal framework analyzed in this article.

---

[23] D Butler, "When Google Got Flu Wrong" (2013) 494 *Nature* 155.

[24] ID Dinov, "Volume and Value of Big Healthcare Data" (2016) 4 *Journal of Medical Statistics and Informatics* 3.