The Privacy and Security Implications of Open Data in Healthcare

A Contribution from the IMIA Open Source Working Group

Shinji Kobayashi[1], Thomas B. Kane[2], Chris Paton[3]

[1] Chair of IMIA OSWG, EHR Department of Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2] Centre for Algorithms, Visualisation and Evolving Systems, School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

[3] Co-Chair of IMIA OSWG, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Summary

Objective: The International Medical Informatics Association (IMIA) Open Source Working Group (OSWG) initiated a group discussion to discuss current privacy and security issues in the open data movement in the healthcare domain from the perspective of the OSWG membership.

Methods: Working group members independently reviewed the recent academic and grey literature and sampled a number of current large-scale open data projects to inform the working group discussion.

Results: This paper presents an overview of open data repositories and a series of short case reports to highlight relevant issues present in the recent literature concerning the adoption of open approaches to sharing healthcare datasets. Important themes that emerged included data standardisation, the inter-connected nature of the open source and open data movements, and how publishing open data can impact on the ethics, security, and privacy of informatics projects.

Conclusions: The open data and open source movements in healthcare share many common philosophies and approaches including developing international collaborations across multiple organisations and domains of expertise. Both movements aim to reduce the costs of advancing scientific research and improving healthcare provision for people around the world by adopting open intellectual property licence agreements and codes of practice. Implications of the increased adoption of open data in healthcare include the need to balance the security and privacy challenges of opening data sources with the potential benefits of open data for improving research and healthcare delivery.

Keywords

Keywords: Open data, FLOSS, Open source, healthcare, privacy, security

Yearb Med Inform 2018:
http://dx.doi.org/10.15265/
Published online

1 Introduction

The International Medical Informatics Association (IMIA) Open Source Working Group (OSWG) is a voluntary group supported by IMIA that brings together researchers and practitioners from multiple countries with a diverse range of informatics experience but common interest in the adoption of open approaches to advancing the use of informatics to improve healthcare. Although not engaged on a common research project, working group members foster international discussions and debates on current issues related to the adoption of open source approaches and have supported the development of an open access database of Free, Libre, and Open Source Software (FLOSS), called MedFLOSS (www.medfloss.org), to be used in the medical domain.

In response to the theme of the 2018 issue of the IMIA yearbook, "Sharing data: balancing access and privacy to advance healthcare", the IMIA OSWG members have collaborated on a working group discussion to discuss the interplay and commonalities between the open source and open data movements in the healthcare domain.

The openness of software has accelerated innovation in recent decades [1–3]. FLOSS has been widely used as an infrastructure for powering in the Internet and has also been widely adopted in the medical domain [4–6]. This trend towards openness in software development is now being applied to the publication of datasets. In recent years, datasets have been opened to the public under FLOSS-like licences on the Internet as "open data" [7]. There is still some debate about the definition of "open data" but, in general, "open data" is the phrase that has come to mean available "for anyone to use, for any purpose, at no cost" [8].

The Open Knowledge Foundation (OKF) has defined "open data" as datasets released with these conditions:

1. **Availability and access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the Internet. The data must also be available in a convenient and modifiable form;

2. **Reuse and redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable;

3. **Universal participation:** everyone must be able to use, reuse, and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

In this article, the term "open data" follows the OKF definition.

The open data movement has rapidly spread across the world since Barack Obama issued his memorandum on transparency and open government in 2009, which emphasized how openness would strengthen democracy and promote government efficiency and effectiveness [9]. Following this memorandum, the US government implemented a Web-based platform to open their data called Data.Gov [10]. Other countries followed this activity and there are now many open data websites

operated by governments around the world. The social coding website, GitHub, now hosts a wide range of open source tools to manipulate open data and the site is frequently used by the open data community to discuss implementation guides, data requirements, and other processes needed to use open data effectively [11]. There are various FLOSS tools to utilize "open data", such as geographical mapping tools [12] and tools for genetic analysis [13], and there are now a wide range of internet companies using open data to create online services such as Citymappers and Transport for London who use these open data sources to provide visualised information [14, 15]. The beneficial economic effect of open data has been estimated to be between 3 to 5 trillion USD per year, in the global economy, and the use of open data in health care has been claimed to save 450-550 billion USD per year in US [16]. The European commission estimated the total market size of "open data" to between 193 and 209 billion EUR for 2016, and expected this figure to grow by 36.9% by 2020 in the EU 28+ [17].

In the scientific domain, genomics data have been made open through internet repositories and these have accelerated the completion of the human genome project [18] and cultivated a wide range of bioinformatics and other -omics research through the availability of data [19] computer science, mathematics, and statistics. Data intensive, large-scale biological problems are addressed from a computational point of view. The most common problems are modeling biological processes at the molecular level and making inferences from collected data. A bioinformatics solution usually involves the following steps: Collect statistics from biological data. Build a computational model. Solve a computational modeling problem. Test and evaluate a computational algorithm. This chapter gives a brief introduction to bioinformatics by first providing an introduction to biological terminology and then discussing some classical bioinformatics problems organized by the types of data sources. Sequence analysis is the analysis of DNA and protein sequences for clues regarding function and includes subproblems such as identification of homologs, multiple sequence alignment, searching sequence patterns, and evolutionary analyses. Protein structures are three-dimensional data and the associated problems are structure prediction (secondary and tertiary. Free access to scientific databases has supported a wide range of large-scale scientific projects globally and is now largely regarded as part of the information technology (IT) infrastructure for many projects. Organisation for Economic Co-operation and Development (OECD) guidelines have described these beneficial effects of open data and now promote the disclosure of research data from public-funded projects to maximise the effects of government investment [20].

Healthcare data usually contains sensitive information that needs to be protected for privacy reasons, which, at first thought, may be considered as a barrier to openness. However, many non-healthcare open data repositories already contain human subject data and open data availability is increasingly being adopted in healthcare despite this concern. There are now rising pressures to adopt open data for research and healthcare operational transparency but the use of open data will still require a careful balancing of free access and privacy in order to protect healthcare clients' confidentiality [8, 21]. This is because although we would like the patient to have access to his or her own data, we also want to preserve the right for any patient to keep his or her health data private if (s)he so wishes. This is highlighted in the "second pillar" of the Estonian e-Society [22] which states that citizens can have access to their data records but that other government agencies have access to the data only if the citizen authorises access via his or her personalised digital signature.

Citizens' right to personal privacy has been extended to patients by physicians from the time of Hippocrates of Cos (400 B.C.), who established a school of medicine and expected medical practitioners to subscribe to a medical oath, now known as the Hippocratic oath [23], essential variants of which are still in use today. New clinicians promise that "whatsoever I shall see or hear in the course of my profession, as well as outside my profession in my intercourse with men, if it be what should not be published abroad, I will never divulge, holding such things to be holy secrets".

The oath's respect for privacy and human dignity also has a medical dimension: "I will use treatment to help the sick according to my ability and judgment, but never with a view to injury and wrong-doing." This can be relevant in situations where a patient's health may be adversely affected by having little or no control over his or her own medical privacy [24], or where societal forces allow employers to pursue an expectation of having access to personal medical data by presuming that employee medical data will be made available to them because they can buy it, or because they can expect all medical data of their employees to be made open to them. The new General Data Protection Regulation coming into force to the state members of European Union in 2018 [25] will threaten worldwide companies with large fines if they misuse personal data from Europe. Alongside data protection issues, there arise questions about the use of medical health data that categorise individuals [26] by commercial organisations. And yet, at the same time, perhaps there are medical details patients will want to share, for purposes of medical research [27]. Perhaps, too, patients would like to have access to all of their own medical data.

With the topic of this year's Yearbook of Medical Informatics issue being how to balance the various requirements of the various stakeholders in medical health data, an important medico-socio-technical issue can be seen as one of seeking to allow persons comprehensive access to their own medical data, where their own interests are served, and using healthcare data to improve medical research and healthcare delivery while protecting their medical data privacy, where it is medically justifiable to do so.

2 Methods

Through a combination of Open Source Working Group discussions and desk research, we selected a number of existing open healthcare data repositories to serve as examples of the types of standards and approaches to open data in healthcare. Working group members also independently reviewed recent issues and controversies surrounding the open data movement to produce a series of mini-case studies.

We evaluated whether health data repositories meet the OKF open definition [28] in the following categories:

1. Open licence;
2. Free accessibility;
3. Machine readability;
4. Open format.

Licences must be public domain or compatible with open licences and should be downloadable via the Internet at reasonable cost. The data should be in a machine-readable form that can be easily processed by a computer using an open data format [28].

Through a snowball sampling technique, we aimed to discover open data repositories in use in healthcare by examining grey and white literature retrieved from internet searching by keywords, such as "open data health", journal database searches, and the Open Health Data Journal recommendations [29]. While the "data portal" site [30] shows 524 open data repositories, we sampled 13 data repositories that include healthcare data shown below (Table 1) as examples.

3 Results

Table 1 shows an overview of each of the example healthcare data repositories and how they align with the openness principles defined by the OKF which we summarise here:

Organisation

The maintenance of the majority of the data repositories identified was performed by local or national governments, or both working together. Repositories #3, #8, #9, #10, #11, and #13 were maintained by academic groups. The remaining repositories #1, #2, and #6 are "mash-up" sites that collect relevant data repositories of their nations to provide each dataset.

Repository #11 is a data repository site for researchers that enables them to freely share data within a size limit in a private repository and provides unlimited storage with open conditions. If users need more storage over the limit, a premium service is also available.

Data Categorisation

Every data repository reviewed has various categories of data, such as public health, epidemiological data, health services, disease distribution (by age groups, areas, or other risk factors), profiles of healthcare providers, and geolocation of hospitals.

Data Accessibility

All data sets were accessible via the Internet without any charge. Most of the data sets could be opened and freely downloaded, but the repositories #3, #9, and #12 requested registration or permission in order to download data.

Repositories #1, #2, #9, #10, and #11 also provide data via web application programming interface (API) for developers to query data programmatically. One example of open data API use is the rHealthDataGov project on GitHub which enables API access to HealthData.gov for users of the R statistical analysis package [31].

Table 1 Health Data Repositories and their profiles.

No.	Data repository name	Issuer	URL	Accessibility	License	Machine readability	Available formats
1	HealthData.gov	U.S. Department of Health & Human Services	https://www.healthdata.gov/	open	open	3531/3542	JSON, CSV, XML, RDF
2	DATA.GOV.UK	United Kingdom, government digital service	https://data.gov.uk/	open	open	2121/2132	CSV, XLS, HTML
3	The human mortality database	The Human Mortality Database Project	http://www.mortality.org/	registration required	not open	12103	CSV
4	Global health observatory data	World Health Organisation	http://www.who.int/gho/database/en/	open	not open	>1000	JSON, CSV, XML, XLS
5	Big cities health coalition	A forum for the leaders of America's largest metropolitan health departments	https://bchi.bigcitieshealth.org/	open	not explicit	53/53	CSV
6	DATA GO JP	Cabinet secretariat of Japan	http://www.data.go.jp	open	open	112/624	CLS, HTML, PDF
7	Dryad	Non-profit organisation	http://datadryad.org	open	open	974/974	XLS, MATLAB, RMD, SOLR
8	UKDA	Academic group	http://www.data-archive.ac.uk	open	open	>1502	PDF, XLSX, CSV
9	Physionet	NIH granted project	http://www.physionet.org	open but partial registration required	open	>111	R, MATLAB, Database, API
10	Open Health Data dataverse	Harvard university	https://dataverse.harvard.edu/dataverse/openhealthdata	open	not open	About 62900	Text, CSV, XSLC, R, SPSS
11	Figshare	Academic group	https://figshare.com/	open	open	(more than 40)	CSV, XSLX
12	SND	Swedish national data service	https://snd.gu.se/sv	permission required	not explicit	0	-
13	eResearch South Australia	Joint venture of universities	https://data.sa.edu.au/	open	open	1	Text

License

Repositories #1, #2, #6, #8, #9, #10, #11, and #13 have open licenses that satisfy the open data definition, such as Creative Commons CC-BY [32], Open Data Commons Open Database Licence [33], or other open licences. Other sites did not show explicit information on the licences used for data distribution or modification.

Even though most of the datasets contained in the repositories described were released under open licences on "open data" repositories, some of the datasets were not published under open licences, because of legal restrictions related to the governance of health data.

Machine Readability, Available Formats

In most of the sites, common data formats such as CSV (Comma Separated Values), or exchange data formats such as JSON (JavaScript Object Notation), and XML (eXtensible Markup Language) are available. The number of formats were categorized in #1, #2, and #6, but not shown in other sites. However, some of the datasets provide only PDF (Portable Document Format). For example, repository #6 provides 624 datasets, but 512 of them are in PDF format. Therefore, only 112 of 624 datasets were available as machine-readable formats. Even though PDF is an ISO 32000-2 standardized format, it is not a preferred option for processing data for analysis by most scientists. There are also many datasets published in spreadsheets using the XLS format that is widely used in Microsoft Excel 2003 or earlier. Since the specification of XLS format is openly published by Microsoft Corporation [34], it might be categorized as an open format. The Office Open XML format, that was adopted in Microsoft Excel 2007 or later, is one of the ISO standards [35] but it was less frequently adopted than XLS in published datasets.

Data Protection

There are more than 80,000 open datasets in all the repositories we identified but personally identifiable information was not found in the example database we selected although more detailed analysis is needed to confirm this. Repositories #1, #2, #4, #6, #7, #10, #11, and #12 addressed their privacy policy to their datasets, and we could not identify any fault against their policies in their datasets.

4 Case Studies

The Working Group discussion and individual working group member desk reviews identified a number of Case Studies that serve to highlight common themes related to the adoption of open data in healthcare, and ethical issues about how medical data sets are shared, used, and reused by various agencies.

4.1 Case 1 - Care Providers and Global Businesses

Accenture PLC [36] is one of the largest consulting companies in the world. When analysing the medical domain, they found various forms of disruption entering the industry over the next decade [37]. For example, Accenture found examples of device companies (such as hearing-aid suppliers) and pharmaceutical companies turning towards service support. They also found that new partnerships developed between digital companies, healthcare providers, and service support industries. There seems to be a particular interest for advanced cognitive systems such as IBM Watson [38], in areas such as genomics, oncology, care management, and drug discovery. Google's DeepMind Health systems are being employed to serve patients, nurses, and doctors [39]. The DeepMind "Streams" app, for example, allows clinicians to be informed when patient vital signs deteriorate using data from patient-monitoring technology that deliver in real-time significant patient life-sign indicators to the clinician's mobile device. The potential for such support, particularly in developing countries, has been highlighted as a major medical benefit. Many of these big data and cognitive solutions are proprietary technologies which are engaged under contract. Ethical questions regarding policy, data ownership, and data usage by national governments and private solution providers arise. Deployment of free and open source eHealth solutions may take a greater national importance when considered as an option to national implementations in this environment.

4.2 Case 2 - Medical Data Aggregated with Data from Other Sources

In 2003, NHS England launched care.data [40] in order to combine all healthcare records stored by general practitioners with all information stored by social services and hospitals, the data being loaded into national Health and Social Care Information Centre (HSCIC) databases. Another aspect of HSCIC databases is the Hospital Episode Statistics dataset which collects and collates data from 125 million individual inpatients, outpatients, and Accident and Emergency records yearly in England. Questions naturally arise regarding how this data will be combined, how it will be used, under what circumstances it needs to be anonymised, and whether it can be anonymised. National implementers face numerous eHealth medical, ethical, and governance challenges [41, 42], and a range of tools for framing questions is emerging [43, 44]. These difficult issues aside, national healthcare providers who have made significant movements in healthcare include Estonia, Sweden, Norway, Denmark, Japan, Canada, and England. More research is needed to compare and contrast the merits and demerits of various approaches taken by national governments.

4.3 Case 3 – Selling Medical Data to Commercial Organisations

Care.data has been a valuable research resource for such tasks as resource allocation and monitoring of treatment effectiveness. Controversially, the combined data was made available to pharmaceutical companies, insurance companies, health charities, hospital trusts, think tanks, and other private companies. In 2014, as part of an audit of sales, it was disclosed that anonymous, pseudonymous, and identifiable data was sold to 160 organisations [40]. In response to a Freedom of Information request [45], HSCIC stated: "We recognise that there will however remain a latent risk that when combined with other sources of data, the identity of the individual may be ascertained". Questions arise regarding the importance of anonymisation, and the suitability of pseudonymisation in cases where pseudonymisation does not protect the identity of patients from commercial organisations.

Regarding the responsibility of organisations such as HSCIC in their provision of anonymous and pseudonymous data, the Article 29 working party [46] has specified that "to identify if a person is identifiable, ac-

count should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person." Questions regarding commitment to patient data security arise.

4.4 Case 4 - Patient, Treatment, Research, and Citizen's Rights

In a case from 2008, a patient who had suffered a life-threatening cancer as a youth [24] later became a clinician. Details of the patient's condition were passed on to researchers, and researchers subjected the adult clinician to intrusive and upsetting telephone calls in the name of research. The clinician complained to the hospital where the researchers were based, but no action was taken to remove the patient from their register. Eventually, through recourse to the law, the situation was resolved. The hospital apologised for its treatment of the clinician and removed the clinician's name from the register. The hospital, in its apology, noted that the concept of medical confidentiality is paramount. Questions arise regarding the steps patients need to take to ensure their wishes and data security are respected so that their previous illnesses don't cause them further distress after treatment has ended.

4.5 Case 5 - Responding to Concerns, Learning Lessons

Because of numerous concerns, opt-outs (over 1 million people, including 40% of GPs) and governmental criticisms, and following a government review of care.data [26], care.data was closed down in July 2016. However, bringing together social records, hospital records, and general practitioner records is ongoing. One of the outcomes from the DeepMind work, the Streams app, which allows clinicians to remain informed of a patient's life signs, has been so successful that Taunton and Somerset NHS Foundation Trust and DeepMind Healthcare signed a 5-year contract in June 2017 to develop and evaluate a system able to detect early signs of kidney-failure [47]. Taunton and Somerset have been pioneers in the use of open source software, with their contract with IMS Maxim. That said, there are many ethical questions still outstanding. Over 1.6 million live NHS data records were given to Google, via DeepMind, inappropriately, and it is neither clear what other such mistakes will be made, nor how the business relationship will develop over the years between DeepMind, Google, and NHS England. In this regard, it is encouraging to note the emergence of an ethics unit within DeepMind Health Systems [47]. Important questions arise over how to learn lessons from past failures as technical rollouts proceed and whether sophisticated business information system solutions, as they become the norm, will be at odds with ethical considerations and, in particular, the Hippocratic oath.

These questions are pertinent to a new age of digitised medical health provision. Where service providers need to find a balance between data security, data openness, and citizen data rights, such cases can help us to focus attention on important ethical questions. Strong opinions need to be traded one against the other in a dialectical process in which all data stakeholders are sufficiently represented.

5 Discussion

We outlined some of the key "open data" repositories in the healthcare domain and reviewed recent cases that highlight the privacy and security issues associated with open data. The main concerns around "open data" are how to manage the benefits and transparency effects that we discuss below.

The open data movement has been productive for a positive cycle of creation for both datasets and tools [7]. Even though the spectrum of datasets for open data is different from person identifiable data sets (Figure 1), there is frequent scepticism about open health data initiatives. Questions that arise here are not only those about mixing open data with private medical information, there are also questions about providing commercial organisations with live patient data that could be used in a business space, for business advantage.

To protect individual privacy, open data repositories set their own privacy rules and usually oblige data issuers to perform de-identification, pseudonymisation, and anonymisation. From the open government memorandum of Barack Obama, legislation was advanced for US, EU, and other countries as safeguards [48]. For example, opt-out rules from private data collections were obligated in many countries. However, there are differences between the privacy regulation of the EU [25] and the US [49, 50]. Since data sources are distributed worldwide, privacy legislation needs to be harmonized over countries or regions.

Even though health data sets are freely available, five out of 13 of the repositories did not provide datasets under open licences. Some of them addressed publication guides for research but there were no explicit licences for secondary use of their datasets. While these data repositories are primarily designed to use data for research which is published in the academic arena, open data have also been used to develop software to visualize the datasets with or without analysis. In 2011, open data were used to facilitate emergency responses to disasters. After the east Japan earthquake, OpenStreetMap [12] was used to share the disaster information for logistics with victims [51]. In the same year, Hurricane Irene approached New York city and nursing home capability data were used to build a plan to optimize the evacuates [52]

Because health data contains sensitive private information, the level of openness of health datasets needs to be restricted. Most of the data for clinical research cannot be opened because the data needs to be person identifiable to be clinically useful. Another issue that has been problematic recently is data fraud in clinical research [53]. The International Committee of Medical Journal Editors (ICMJE) proposed a data sharing framework for clinical trials to improve the transparency for research [54]. An open data approach might help to build up such frameworks for transparency of clinical trials.

One of the most widely discussed controversies is about who owns the data in healthcare data repositories. Consensus is building that personal data should belong to each person, but population data could be considered as a resource of communities or societies. Since the open data movement is such a social movement, sharing de-identified or anonymised health data as open data could be an approach to advocate eHealth for new generations.

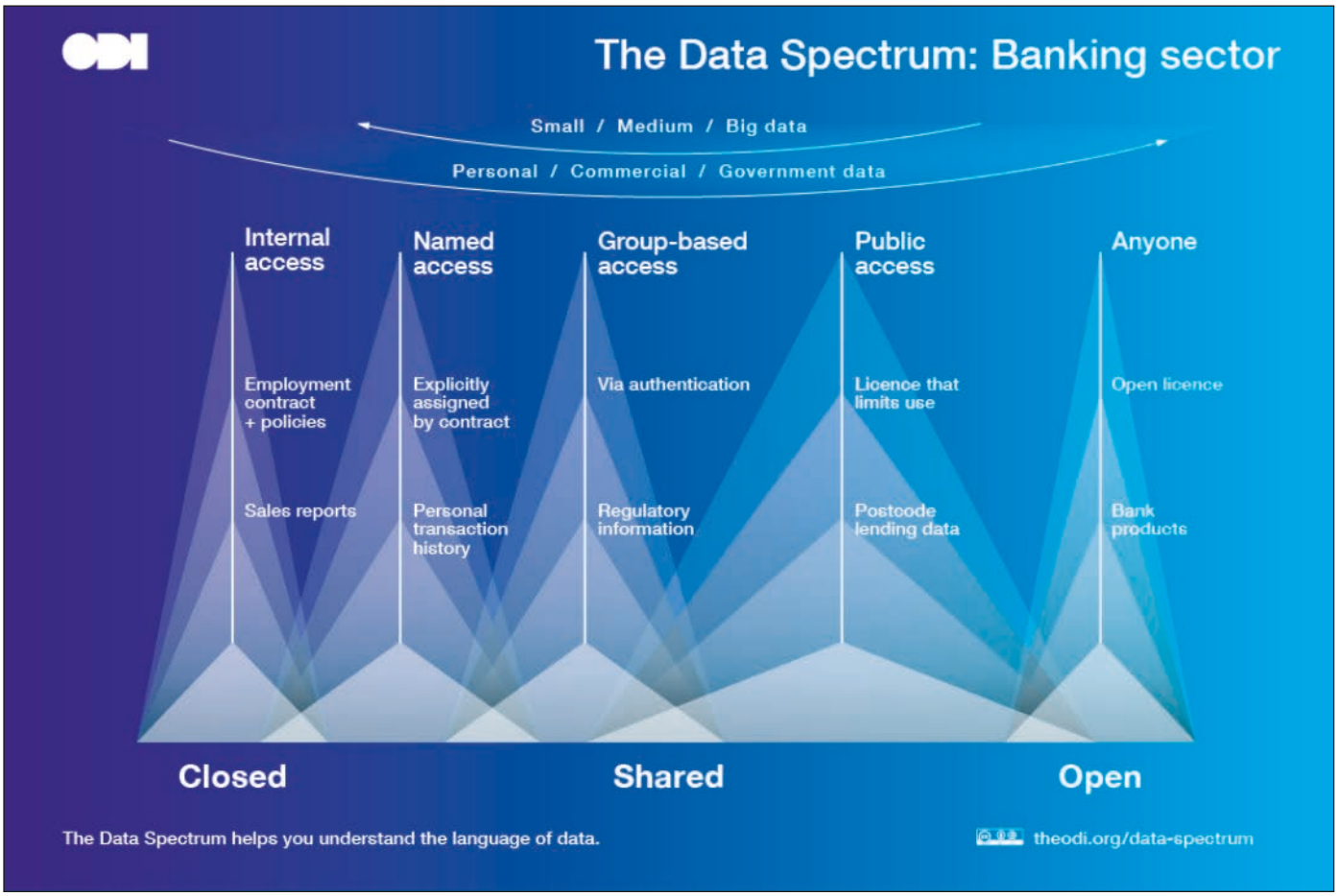

Fig. 1 Data spectum of data sets. Reproduced from the Open Data Institute website

6 Conclusion

Many aspects of the open data movement originated from the open source software movement with the aims of improving transparency and fostering innovation. The benefits of open data have been established in many cases and, as health data are increasingly included in open data repositories, issues of anonymisation and de-identification need to be addressed and appropriately managed. The open data and open source movements share many common philosophies and approaches including using developing international collaborations across multiple organisations and domains of expertise. Both movements aim to reduce the costs of advancing scientific research and improving healthcare provision for people around the world by developing intellectual property licence agreements and codes of practice that are increasingly being adopted in the software development and scientific communities.

Acknowledgements

We would like to thank Thomas Karopka (past Chair of the OSWG) for his advice and direction for this paper. The paper was conceived jointly by the authors who participated in an OSWG discussion in response to the call for working group contributions to the IMIA Yearbook. SK contributed to the selection and analysis of the Open Data repositories, TK contributed to the Case Studies discussing issues relating to the adoption of open data. All the authors revised the manuscript and contributed to the overall structure of the paper.

Funding Statement

CP was funded by the Health Systems Research Initiative grant (MR/N005600/1) (jointly supported by the Department for International Development (DFID), the Economic and Social Research Council (ESRC), the Medical Research Council (MRC), and the Wellcome Trust (WT)) to investigate the adoption of open source software in low and middle income countries which informed this WG discussion paper.

References

1. Open Source: A Platform for Innovation | WIRED [Internet]. [cited 2017 Nov 29]. Available from: https://www.wired.com/insights/2013/11/open-source-a-platform-for-innovation/
2. Von Krogh G, Spaeth S, Lakhani KR. Community, joining, and specialization in open source software innovation: a case study. Res Policy 2003 Jul;32(7):1217–41.
3. How Open Source Is Changing Software Innovation [Internet]. [cited 2017 Nov 29]. Available from: http://www.digitalistmag.com/cio-knowledge/2017/05/17/open-source-changing-software-innovation-05081271
4. Paton C, Karopka T. The role of free/libre and open source software in learning health systems. Yearb Med Inform 2017 Aug;26(1):53–8.
5. Erickson BJ, Langer S, Nagy P. The role of open-source software in innovation and standardization in radiology. J Am Coll Radiol 2005 Nov;2(11):927–931.
6. Kobayashi S, Yahata K, Goudge M, Okada M, Nakahara T, Ishihara K. Open source software in medicine and its implementation in Japan. Journal on Information Technology in Healthcare 2009;7(2):95–101.
7. Chignard S. A brief history of Open Data [Internet] 2013. Available from: http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data
8. Public views on open data. 2013 [cited 2018 Mar 14];(June). Available from: https://forum.kodujdlapolski.pl/uploads/default/original/2X/d/df977225a37ba192982cf408c08926572af775f2.pdf
9. Obama B. Transparency and Open Government [Internet] 2009 [cited 2017 Mar 30]. Available from: https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government
10. Data.gov [Internet]. [cited 2017 Nov 29]. Available from: https://www.data.gov/
11. Rouder JN. The what, why, and how of born-open data. Behav Res Methods 2016 Sep;48(3):1062–1069.
12. OpenStreetMap. Planet dump retrieved from https://planet.osm.org [Internet]. 2017 [cited 2017 Dec 1]. Available from: https://www.openstreetmap.org
13. GenomeTools [Internet]. [cited 2018 Mar 11]. Available from: http://genometools.org/
14. Citymapper executive to governments: "Open more data so we can improve your cities" | News | Open Data Institute [Internet]. [cited 2017 Nov 30]. Available from: https://theodi.org/news/citymapper-government-open-data-improve-cities
15. Open data users - Transport for London [Internet]. [cited 2017 Nov 30]. Available from: https://tfl.gov.uk/info-for/open-data-users/
16. Open data: Unlocking innovation and performance with liquid information | McKinsey & Company [Internet]. [cited 2017 Nov 30]. Available from: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information
17. Creating Value through Open Data. [cited 2018 Mar 14]; Available from: https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf
18. Marturano A. Human genome and open source: balancing ethics and business. Rev Derecho Genoma Hum 2011 Dec;(35):225–238.
19. Can T. Introduction to bioinformatics. Methods Mol Biol 2014;1107:51–71.
20. OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD Publishing; 2007.
21. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B. Hart G, et al. Who owns the data? open data for healthcare. Front Public Health 2016 Feb 17;4:7.
22. Priisalu J, Ottis R. Personal control of privacy and data: Estonian experience. Health Technol (Berl) 2017 Jun 15;7(4):441–51.
23. Henderson J. Hippocrates of Cos, The Oath. Loeb Classical Library; 1923.
24. MacDonald V. New leaked data fiasco - Channel 4 News. Channel 4 News;2008.
25. EUGDPR. EU General Data Protection Regulation [Internet]. [cited 2017 Nov 21]. Available from: http://eugdpr.org/eugdpr.org.html
26. Presser L, Hruskova M, Rowbottom H, Kancir J. Care.data and access to UK health records: patient privacy and public trust. Technology Science 2015;
27. Senior M. The case for open-source electronic patient records. Computing; London. 2014;11–2.
28. The Open Definition [Internet]. Open Knowledge International. Available from: http://opendefinition.org/od/2.1/en/
29. Open Health Data Journal [Internet]. [cited 2017 Feb 11]. Available from: https://openhealthdata.metajnl.com/
30. Search - Data Portals [Internet]. [cited 2018 Mar 7]. Available from: http://dataportals.org/search
31. rOpenHealth/rHealthDataGov: This package provides an R interface to the HealthData.gov Data API. [Internet]. [cited 2017 Dec 2]. Available from: https://github.com/rOpenHealth/rHealthDataGov
32. Attribution 4.0 International (CC BY 4.0) [Internet]. Creative Commons. Available from: https://creativecommons.org/licenses/by/4.0/
33. Open Data Commons Open Database License(ODbL) [Internet]. Open Knowledge Foundation. Available from: http://www.opendatacommons.org/licenses/odbl/1.0/
34. [MS-XLS]: Excel Binary File Format (.xls) Structure [Internet]. [MS-XLS]: Excel Binary File Format (.xls) Structure. 2017 [cited 2018 Mar 16]. Available from: https://msdn.microsoft.com/en-us/library/office/cc313154(v=office.12).aspx
35. ISO/IEC 29500: Office Open XML File Formats. ISO. 2016.
36. Perrine CL, Tar JJ, Asper AE. Accenture. interactions 2002 Mar 1;9(2).
37. Elton J, O'Riordan A. Healthcare disrupted: Next generation business models and strategies. John Wiley & Sons; 2016.
38. Chen Y, Elenee Argentinis JD, Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Clin Ther.2016 Apr 21;38(4):688–701.
39. Iacobucci G. Patient data were shared with Google on an "inappropriate legal basis," says NHS data guardian. BMJ 2017 May 18;357:j2439.
40. Bhatia N. care.data - in detail. care.data - in detail. 2016;
41. He DD, Yang J, Compton M, Taylor K. Authorization in cross-border eHealth systems. Information Systems Frontiers; New York. 2012 Mar;14(1):43–55.
42. George C, Whitehouse D, Duquenoy P, editors. eHealth: legal, ethical and governance challenges. 1st ed. Heidelberg ; New York: Springer; 2013.
43. Hopia H, Punna M, Laitinen T, Latvala E. A patient as a self-manager of their personal data on health and disease with new technology--challenges for nursing education. Nurse Educ Today 2015 Dec;35(12):e1–3.
44. Menvielle L, Audrain A-F, Menvielle W. The digitization of healthcare : new challenges and opportunities; 2017.
45. Bhatia N. Register of approved data releases - a Freedom of Information request to NHS Digital. WhatDoTheyKnow; 2014 Apr;
46. WP A 29. Article 29 Working Party Opinions and recommendations - European Commission; 2016;
47. Temperton J. DeepMind's new AI ethics unit is the company's next big move [Internet]. [cited 2017 Nov 21]. Available from: http://www.wired.co.uk/article/deepmind-ethics-and-society-artificial-intelligence
48. What is open data? [Internet]. Open data institute. [cited 2017 Feb 20]. Available from: https://theodi.org/what-is-open-data
49. Department of State Privacy [Internet]. [cited 2017 Dec 1]. Available from: https://www.state.gov/privacy/
50. Summary of the HIPAA Security Rule | HHS.gov [Internet]. [cited 2017 Dec 1]. Available from: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html
51. Koshimura, S., Hayashi, S., Gokon, H. The impact of the 2011 Tohoku earthquake tsunami disaster and implications to the reconstruction. SOILS AND FOUNDATIONS. 2014 Aug;54(4):560–572.
52. Martin EG, Helbig N, Shah NR. Liberating data to transform health care: New York's open data experience. JAMA 2014 Jun 25;311(24):2481–2.
53. George SL, Buyse M. Data fraud in clinical trials. Clin Investig (Lond) 2015;5(2):161–73.
54. Devereaux PJ, Guyatt G, Gerstein H, Connolly S, Yusuf S; International Consortium of Investigators for Fairness in Trial Data Sharing. Toward fairness in data sharing. N Engl J Med 2016 Aug 4;375(5):405–7.
55. The Data Spectrum | Open Data Institute [Internet]. [cited 2017 Nov 29]. Available from: https://theodi.org/data-spectrum

Correspondence to: