

The Role of Free/Libre and Open Source Software in Learning Health Systems

C. Paton¹, T. Karopka²

¹ Group Head for Global Health Informatics, Centre for Tropical Medicine and Global Health, University of Oxford, UK

² Chair of IMIA OS WG, Chair of EFMI LIFOSS WG, Project Manager, BioCon Valley GmbH, Greifswald, Germany

Summary

Objective: To give an overview of the role of Free/Libre and Open Source Software (FLOSS) in the context of secondary use of patient data to enable Learning Health Systems (LHSs).

Methods: We conducted an environmental scan of the academic and grey literature utilising the MedFLOSS database of open source systems in healthcare to inform a discussion of the role of open source in developing LHSs that reuse patient data for research and quality improvement.

Results: A wide range of FLOSS is identified that contributes to the information technology (IT) infrastructure of LHSs including operating systems, databases, frameworks, interoperability software, and mobile and web apps. The recent literature around the development and use of key clinical data management tools is also reviewed.

Conclusions: FLOSS already plays a critical role in modern health IT infrastructure for the collection, storage, and analysis of patient data. The nature of FLOSS systems to be collaborative, modular, and modifiable may make open source approaches appropriate for building the digital infrastructure for a LHS.

Keywords

FLOSS; Open Source; Learning Health Systems; Secondary Use

Yearb Med Inform 2017;53-8

<http://dx.doi.org/10.15265/IY-2017-006>

Published online May 8, 2017

Introduction

As more patient data are collected electronically through Electronic Health Records (EHRs) and other information systems used in healthcare organisations, the opportunity to reuse the data these systems collect for research and quality improvement becomes more apparent. There is now a wide range of large-scale research projects [1, 2] that are reusing routinely collected data for analysis and the concept of a Learning Health System (LHS) has been coined to summarise this cycle of data collection, analysis, and health service improvement [3–5].

There is a plethora of different information systems in use across healthcare systems around the world. Although many user-facing systems are proprietary in nature, meaning that source code for the software is protected and cannot be modified, many systems use infrastructure components that are open source (meaning that the source code is licenced in a way that allows for modification and reuse). This paper aims to discuss the role of Free/Libre and Open Source Software (FLOSS) in modern healthcare information technology (IT) infrastructure and will highlight some of the major open source projects that are contributing to the development of LHSs that reuse patient data for research and health service improvement.

Methodology

We have undertaken a broad environmental scan utilising a snowballing approach to the discovery of FLOSS that contributes to

LHSs, examining the academic literature, the MedFLOSS database that catalogues FLOSS in use in healthcare [6], and the grey literature from websites, reports, and personal communications with experts in the area of FLOSS adoption in healthcare.

Although some EHRs and research systems are fully open source, such as the Veterans Information Systems and Technology Architecture (VistA) [7], modern healthcare infrastructure is often a combination of open source and proprietary systems. In our analysis, we will therefore examine how the open source (OS) ‘stack’ from infrastructure to user interfaces is used in healthcare applications. We will then examine the range of OS data analysis tools and look at how new modular EHR “apps” that utilise open source software and open standards can close the loop to create a LHS that reuses patient data to enable better clinical decision making.

Results

Our environmental scan has resulted in the identification of a wide range of FLOSS that contribute to health IT projects critical for LHSs. We have therefore divided our findings into general purpose FLOSS building blocks such as operating systems, frameworks, and databases (Core Infrastructure FLOSS); FLOSS projects used to analyse patient data but not specifically designed for clinical purposes such as general purpose statistical software; FLOSS projects that are used in healthcare but may or may not contribute to LHS development depending on how they are implemented, such as open

source EHR systems (Clinical FLOSS); and Clinical Data Management FLOSS that, when combined with EHR systems and core technologies, can form the digital infrastructure for LHSs. Table 1 displays the set of the general categories identified along with a list of examples in each category.

Core Infrastructure FLOSS

Open source operating systems are now used in a wide variety of clinical environments. Servers running Linux often power in-house data-centres and cloud-based EHR systems. Databases such as MySQL, PostgreSQL and newer NoSQL languages are now storing clinical data for a wide range of EHR systems including the NHS ‘spine’ that uses the OS Riak NoSQL system at its core [8]. Most modern EHR systems allow clinicians to input and retrieve data through mobile devices or web interfaces (or web interfaces on mobile devices). The two major mobile operating systems are based on OS projects: Android is a fully open source system and iOS is based on the BSD version of Unix. Clinical apps use a variety of OS frameworks including AngularJS, the Spring framework, and Django to deliver the user interface. Table 2 shows a few examples of core infrastructure FLOSS (although the number of projects in use in various healthcare systems around the world is too numerous to list in full here). As an example, the development of a nursing intelligence system in a large university hospital [9] illustrates how these building blocks can be put together to build a LHS. The authors use different open source tools for data warehousing and business intelligence. They used the Talend Open Studio for data integration (www.talend.com), a PostgreSQL (www.postgresql.org) database for data management, and for data analysis, business intelligence, and result presentation, they used the BIRT (Business Intelligence and Reporting Toolkit, www.eclipse.org/birt) toolkit.

Data Analysis FLOSS

In order to reuse data from clinical systems, they must first be analysed to create actionable insights such as mortality rates

Table 1 General categories of FLOSS

Categories	Examples
Core Infrastructure FLOSS	Operating Systems, Frameworks, Databases, Core Applications (See Table 2)
Data Analysis FLOSS	Statistical Software, Big Data Software, AI/Machine Learning (Table 4)
Clinical FLOSS	EHRs, HIS, LIMS, PACS, Interoperability Engines (See Table 3)
Clinical Data Management FLOSS	OpenClinica, TransSMART, i2b2, CohortExplorer, SEMCARE, OMOP

Table 2 Example of FLOSS used in Core Infrastructure

Operating Systems	Frameworks	Databases	Applications
Linux	AngularJS	MySQL	Apache
Android	Bootstrap	PostgreSQL	Chromium
	Apache Spark	Riak	Firefox
	Java Spring		Eclipse
	Django		

Table 3 Examples of FLOSS used in Data Analysis

Statistical Software	Big Data/AI
R	HDFS
DataMelt	MapReduce
GNU Octave	Weka
Mondrian	Torch
	OpenNN

or levels of adherence to clinical guidelines. A simple analysis can often be performed by exporting information from an EHR database into a CSV file for analysis in a spread sheet. As more data are collected and analyses become more automated as part of a formal LHS, this approach becomes more difficult and organisations and researchers use more sophisticated systems for managing and analysing the clinical data. The R open source statistical analysis package is often used as part of the infrastructure of a LHS as there is now a large library of open source tools that allow data to be imported, analysed, and exported to dashboards or other visualisation systems often in an automated or semi-automated fashion. As datasets grow larger and analyses become more complex, the limits of statistical tools such as R mean that new “Big Data” tools are required. These tools enable researchers

to run complex analyses over large datasets by breaking the dataset down into chunks which are analysed in parallel using open source systems such as Apache Hadoop. Table 3 details some of the core FLOSS projects designed for data analysis that are used as building blocks for LHSs.

Clinical FLOSS

There are now a number of large scale FLOSS projects that enable the collection of patient data in clinical settings. Most of these are not specifically designed for data reuse as part of a LHS but, as with the core infrastructure tools and statistical software described above, can form part of a LHS when combined with other software. Some high-profile examples from the MedFLOSS database are detailed in Table 4.

Clinical Data Management FLOSS

At the heart of a LHS is the ability to transform clinical data into new knowledge. This knowledge may be generalizable and may contribute to new scientific understanding or it may be specific to a particular healthcare organisation and contribute to local quality improvement. Over recent years, a number

Table 4 Examples of Clinical FLOSS

Application Types	Examples
Electronic Health Record (EHR)	WorldVista, OpenMRS, GNU Health, OpenMAXIMS, LibreHealth EHR,
Picture Archive and Communication System (PACS)	MRIdb, Orthanc, DICoogle, Xebra, OSPACS, OpenSourcePACS, ClearCanvas, Conquest DICOM software, CDMEDIC PACS WEB, DCMTK - DICOM Toolkit, dcm4che
Health Information System (HIS)	DHIS2
Laboratory Information Management System (LIMS)	OMERO, FreeLIMS, OpenELIS
Interoperability	Mirth, OpenXDS, IHE open source, OpenARNA, IHE Gazelle Tools, Open eHealth Integration Platform (IPF)

of FLOSS projects have been developed to assist in the analysis of clinical data. Many of these have been developed for the purpose of large scale observational trials but can also be used for local quality improvement. In this section, we highlight some of the recent literature around FLOSS clinical data tools to give a broad overview of some of the current major projects and how they can work together and support data and terminology standards:

i2b2: i2b2 (Informatics for Integrating Biology & the Bedside) [10] is an NIH-funded suite of open source tools for supporting clinical and basic biological science research projects and is now used in several major academic medical centres and research institutes in the USA. The system has also been used for secondary use of claims data from the Austrian Health Insurance system [11] and has undergone successful viability tests for implementation in networked medical research in Germany although some challenges were identified when importing from heterogeneous data sources [12].

The open source nature of the i2b2 platform has resulted in multiple projects extending and adding new capabilities to the existing tools. For example, Bauer *et al.* reported on work where they have developed a suite of tools for loading clinical data from various formats such as CSV, SQL, CDISC ODM or biobanks [13]. Gabetta *et al.* have developed a flexible extension of i2b2 able to exploit different statistical engines [14]. The same group has developed BigQ, an extension of the i2b2 framework, which integrates patient clinical phenotypes with genomic variant

profiles generated by Next Generation Sequencing [15]. Westra *et al.* were able to demonstrate the feasibility of creating a hierarchical flowsheet ontology in i2b2 using data-derived information models and determined the underlying informatics and technical issues [16]. Rance *et al.* have built a platform for translational research in cancer research using i2b2 and tranSMART [17, 18] (described further below).

OpenEHR: OpenEHR is an open data specification supported by an international community that have also developed a number of open source tools for modelling clinical data. Haarbrandt *et al.* have implemented a solution for the automated population of an i2b2 clinical data warehouse from an openEHR-based data repository [19]. With this approach openEHR-based data can be made available to researchers for secondary use.

OpenClinica: OpenClinica is an open source clinical research data management platform that is now in use in over 100 countries [20]. The system has a number of features for conducting clinical trials including randomization and support for patient-reported outcome measures (PROMs) and it can connect with other clinical systems such as Electronic Health Records and Picture Archive and Communications Systems (PACS).

REDCap: REDCap is another free platform for data capture (although not licensed as open source). This platform has been adopted for many clinical studies in low-resource settings such as the Clinical Information Network in Kenya where it is integrated with a suite of open source data capture, processing, and analysis tools [21].

TranSMART: TranSMART is another open source clinical and basic science research platform, originally developed by Johnson and Johnson and then established as an open source platform supported by the tranSMART Foundation from 2013 [22].

Both the tranSMART and OpenClinica platforms have been extended, connected, and enhanced by the international open source community. For example, Camacho Rodriguez *et al.* implemented Mirth Connect as a Communication Server (CS) to convert HL7 messages either to Operational Data Model (ODM) data for the automatic import in OpenClinica or tabular-delimited text format files, whose data is uploaded in tranSMART using the tMDataLoader tool [23]. Firnkorn *et al.* have addressed the problem of extracting data from data warehouses like i2b2 and tranSMART. They have implemented a Generic Case Extractor for clinical items out of these two databases with the aim of extracting data for more complex statistical analysis [24]. Satagopam *et al.* have combined tranSMART with two other server-based platforms addressing different aspects of data processing in translational medicine research: data integration and exploration, bioinformatics workflow construction, and interpretation of analysis results in the disease context [25]. In particular, they have integrated MINERVA (Molecular Interaction NETwork Visualization) [26], with tranSMART and Galaxy, a web-based platform for computational biomedical research [27].

Weasis: A remaining challenge to using electronic data capture systems (EDCSs) in the context of secondary use of patient data is in respect to the use of image data. Although this is usually provided in a standardized form (DICOM), there are still challenges in connecting PACS systems with EDCS. Haak *et al.* have implemented a web-based solution integrating EDCS, PACS, and a DICOM viewer using the open source projects OpenClinica, dcm4che, and Weasis [28,29]. The system is used to perform decentralized computed tomography (CT) screening.

CohortExplorer: Another challenge is the underlying database model of EDCSs which is usually based on generic entity-at-

tribute-value (EAV) schemas. This structure is cumbersome to query with SQL and even more challenging when combining distinct sources with different underlying schemas. The application CohortExplorer [30] provides a solution to this issue. It is available under a GPL 3+ and available with several APIs for connecting to other systems (e.g. for Opal and REDCap). Future versions will also provide an API for OpenClinica.

SMART: The Strategic Healthcare IT Advanced Research Projects (SHARP) program is an initiative from the Office of the National Coordinator for Health Information Technology (ONC) in the USA that supports breakthrough health IT research. One area they are supporting is secondary use of EHR data. In one of the SHARP projects, Klann *et al.* are using Substitutable Medical Apps Reusable Technologies (SMART) to build a patient summarization app which displays a problem-oriented view of medications and presents a line-graph display of laboratory results [31, 32]. The summarization app can run in any EHR environment that either supports SMART or runs SMART-enabled i2b2. In this way data can be extracted from an i2b2 repository reusing clinical data extracted from EHRs. Area four of SHARP (SHARPn) is aimed at developing open-source tools that could be used for the normalization of EHR data for secondary use using high throughput phenotyping [33]. One of the outcomes of this project is that clinical element models proved capable of capturing data originating from a variety of sources within the normalization pipeline and are capable of serving as suitable normalization targets [34].

FHIR: Fast Healthcare Interoperability Resources (FHIR) is an emerging open HL7 standard [35]. Mandel *et al.* have adopted the emerging FHIR standard for use on the SMART framework and called the platform SMART-on-FHIR [36]. Waghlikar *et al.* have implemented an interface for an i2b2 installation using SMART-on-FHIR and have transformed i2b2 into an apps platform [37]. Authors from the same group have implemented an open source toolchain to connect any app developed with Apple's ResearchKit to

i2b2 and called the solution C3-PRO [38]. C3-PRO stands for Consent, Contact, and Community framework for Patient Reported Outcomes and aims to collect personal health data and patient reported outcomes (PRO) from distributed populations.

OMOP: One of the core challenges to implementing LHSs in practice is the standardisation and adoption of common data models to ensure consistency between data used in day-to-day clinical practice and data analysed to generate recommendations for quality improvement or research. The Observational Medical Outcomes Partnership (OMOP) [39] developed a common data model and suite of open source tools to facilitate the adoption of the model and analysis of patient data gathered from clinical IT systems.

The open source nature of these tools has fostered a range of collaborations and extensions enabling them to work with each other, adopt data standards, and support the integration of clinical systems such as EHRs with data management tools to build digital infrastructures that can support LHSs.

Discussion

In order to promote the adoption of LHSs and the reuse of patient data, a number of new clinical solutions will be required and existing systems will need to be adapted and extended to allow for LHS functionality. The role of FLOSS in these projects will range from building on top of core FLOSS infrastructure (such as operating systems and databases) to fully open source clinical and research systems.

For software developers, the adoption or not of open source software in clinical solutions may depend on the types of business models necessary to support business or organisational sustainability. It may be cost-effective for a clinical software supplier to leverage open source software for core infrastructure (such as the operating system and database) and then use an open source programming language to create copyrighted code for the user interface layer which is released as a commercial product with the proceeds supporting

business sustainability and growth. This method of building on an 'open source stack' may encourage more entrepreneurial activity and be a cost effective way for new start-ups to enter the market.

For healthcare providers, utilizing FLOSS may also offer a more sustainable approach. Due to the complex and interconnected nature of LHSs, it may be beneficial for healthcare providers to adopt open source systems rather than purchasing proprietary systems that may be difficult to integrate with other systems (for example, if an EHR does not allow for integration with automated data analysis systems). Various commercial business models, implemented for reasons of business sustainability and growth, may work against the development of a LHS. For example, encouraging customers to purchase a single system across an organisation may be problematic if the system doesn't include data analysis functionality and does not offer a way to easily integrate research and data management systems.

The use of open standards (such as HL7 FHIR) may mitigate some of these issues if they are widely adopted by commercial vendors capable of meeting the needs of both interoperability and data analysis. However, if the proprietary systems on offer focus only on more basic messaging standards (such as HL7 v2), it may be preferable to adopt an open source infrastructure that allows for modification and collaboration between institutions to develop APIs, metadata, and modular extensions (such as seen with the OpenMRS system [40]).

Although the use of patient data for research has been underway now for many years in academic medical centres, large healthcare providers are now starting to gather data from across multiple healthcare facilities (sometimes at a national level) to build larger scale LHSs. The use of this data for research and commercial exploitation is generating interesting discussions on ownership and privacy rights for the individuals that data reference. In the UK, a large scale clinical data aggregation project called Care.Data made headlines about whether citizens should have to opt-in and perhaps explicitly consent to such data collection, to what degree the data should be made

anonymous, and whether or not such data should be sold to researchers and commercial organisations [41]. A recent report by Dame Caldicott offers some useful guidance on how we should communicate with patients about data reuse and when and where we need to obtain explicit consent [42]. Collaboratively developed FLOSS for helping manage consent and data security may help to mitigate some of these issues.

Conclusions

FLOSS infrastructure including operating systems, databases, and software frameworks have enabled much of the technology we use every day as we access information on our smartphone apps or through the internet. Many of these tools also power the core infrastructure behind current health IT systems and will likely form the digital infrastructure for the learning health systems of the future. FLOSS can help new clinical software companies rapidly ‘spin up’ new solutions and, on a larger scale, healthcare providers, research organisations, and governments can take advantage of the collaborative and extensible nature of FLOSS to implement clinical and data management systems that can speed up quality improvement cycles and enable large scale clinical research projects that utilize learning healthcare systems.

Funding Statement

CP is 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).

References

- About PCORnet - PCORnet [Internet]. PCORnet. [cited 2017 Mar 24]. Available from: <http://www.pcor.net/about-pcornet/>
- Delaney BC, Curcin V, Andreasson A, Arvanitis TN, Bastiaens H, Corrigan D, et al. Translational Medicine and Patient Safety in Europe: TRANSFoRm—Architecture for the Learning Health System in Europe. *Biomed Res Int* [Internet]. 2015 Oct 11 [cited 2017 Mar 24];2015. Available from: <https://www.hindawi.com/journals/bmri/2015/961526/>
- Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* 2015 Jan;22(1):43–50.
- Roundtable on Value & Science-Driven Health Care, The Learning Health, Institute of Medicine. *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. Grossmann C, Powers B, McGinnis JM, editors. Washington (DC): National Academies Press; 2011. 320 p.
- English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al. Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical Care in Low- and Middle-Income Countries. *PLoS Med* 2016 Apr;13(4):e1001991.
- Medical Free/Libre and Open Source Software | Structured project database, service providers, publications, events, [Internet]. [cited 2017 Mar 24]. Available from: <http://medfloss.org>
- Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA—U.S. Department of Veterans Affairs national-scale HIS. *Int J Med Inform* 2003;3;69(2–3):135–56.
- Todd R. Spine2 built on Riak database | Digital Health [Internet]. Digital Health. 2013 [cited 2017 Mar 24]. Available from: <https://www.digitalhealth.net/2013/10/spine2-built-on-riak-database/>
- Hackl WO, Rauchegger F, Ammenwerth E. A Nursing Intelligence System to Support Secondary Use of Nursing Routine Data. *Appl Clin Inform* 2015 Jun 24;6(2):418–28.
- i2b2: Informatics for Integrating Biology & the Bedside [Internet]. [cited 2017 Mar 24]. Available from: <https://www.i2b2.org/>
- Endel F, Duftschmid G. Secondary Use of Claims Data from the Austrian Health Insurance System with i2b2: A Pilot Study. *Stud Health Technol Inform* 2016;223:245–52.
- Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU. Unlocking Data for Clinical Research – The German i2b2 Experience. *Appl Clin Inform* 2011 Mar 30;2(1):116–27.
- Bauer CRKD, Ganslandt T, Baum B, Christoph J, Engel I, Löbe M, et al. Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data. *Methods Inf Med* 2016;55(2):125–35.
- Gabetta M, Malovini A, Bucalo M, Zini E, Tibollo V, Priori SG, et al. Beyond Cohort Selection: An Analytics-Enabled i2b2. *Stud Health Technol Inform* 2016;228:572–6.
- Gabetta M, Limongelli I, Rizzo E, Riva A, Segagni D, Bellazzi R. BigQ: a NoSQL based framework to handle genomic variants in i2b2. *BMC Bioinformatics* 2015 Dec 29;16:415.
- Westra BL, Christie B, Johnson SG, Pruinelli L, LaFlamme A, Park JI, et al. Expanding Interprofessional EHR Data in i2b2. *AMIA Jt Summits Transl Sci Proc* 2016 Jul 20;2016:260–8.
- Rance B, Canuel V, Countouris H, Laurent-Puig P, Burgun A. Integrating Heterogeneous Biomedical Data for Cancer Research: the CARPEM infrastructure. *Appl Clin Inform* 2016 May 4;7(2):260–74.
- Dunn W Jr, Burgun A, Krebs M-O, Rance B. Exploring and visualizing multidimensional data in translational research platforms. *Brief Bioinform* [Internet]. 2016 Sep 1; Available from: <http://dx.doi.org/10.1093/bib/bbw080>
- Haarbrandt B, Tute E, Marscholke M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *J Biomed Inform* 2016 Oct;63:277–94.
- Cavelaars M, Rousseau J, Parlayan C, de Ridder S, Verburg A, Ross R, et al. OpenClinica. *J Clin Bioinform* 2015;5(Suppl 1):S2.
- Tuti T, Bitok M, Paton C, Makone B, Malla L, Muinga N, et al. Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient pediatric care and research in Kenya. *J Am Med Inform Assoc* 2016 Jan;23(1):184–92.
- Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Jt Summits Transl Sci Proc* 2013 Mar 18;2013:6–8.
- Camacho Rodriguez JC, Stäubert S, Löbe M. Automated Import of Clinical Data from HL7 Messages into OpenClinica and tranSMART Using Mirth Connect. *Stud Health Technol Inform*. 2016;228:317–21.
- Firnkorner D, Merker S, Ganzinger M, Muley T, Knaup P. Unlocking Data for Statistical Analyses and Data Mining: Generic Case Extraction of Clinical Items from i2b2 and tranSMART. *Stud Health Technol Inform* 2016;228:567–71.
- Satagopam V, Gu W, Eifes S, Gawron P, Ostaszewski M, Gebel S, et al. Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases. *Big Data* 2016 Jun;4(2):97–108.
- Gawron P, Ostaszewski M, Satagopam V, Gebel S, Mazein A, Kuzma M, et al. MINERVA—a platform for visualization and curation of molecular interaction networks. *NPJ Syst Biol Appl* 2016 Sep 22;2:16020.
- Goecks J, Nekrutenko A, Taylor J, Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010 Aug 25;11(8):R86.
- Home - Weasis - dcm4che.org Wiki [Internet]. [cited 2017 Mar 24]. Available from: <https://dcm4che.atlassian.net/wiki/display/WEA/Home>
- Haak D, Page C-E, Reinartz S, Krüger T, Deserno TM. DICOM for Clinical Research: PACS-Integrated Electronic Data Capture in Multi-Center Trials. *J Digit Imaging* 2015 Oct;28(5):558–66.
- Dixit A, Dobson RJB. CohortExplorer: A Generic Application Programming Interface for Entity Attribute Value Database Schemas. *JMIR Med Inform* 2014 Dec 1;2(2):e32.
- Klann JG, McCoy AB, Wright A, Wattanasin N, Sittig DF, Murphy SN. Health care transformation through collaboration on open-source informat-

- ics projects: integrating a medical applications platform, research data repository, and patient summarization. *Interact J Med Res* 2013 May 30;2(1):e11.
32. Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J Am Med Inform Assoc* 2016 Sep;23(5):909–15.
 33. Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPN project. *J Biomed Inform* 2012 Aug;45(4):763–71.
 34. Oniki TA, Zhuo N, Beebe CE, Liu H, Coyle JF, Parker CG, et al. Clinical element models in the SHARPN consortium. *J Am Med Inform Assoc* 2016 Mar;23(2):248–56.
 35. Index - FHIR v3.0.0 [Internet]. [cited 2017 Mar 24]. Available from: <http://hl7.org/fhir/index.html>
 36. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc* 2016 Sep;23(5):899–908.
 37. Waghlikar KB, Mandel JC, Klann JG, Wattanasin N, Mendis M, Chute CG, et al. SMART-on-FHIR implemented over i2b2. *J Am Med Inform Assoc* [Internet] 2016 Jun 6; Available from: <http://dx.doi.org/10.1093/jamia/ocw079>
 38. Pfiffner PB, Pinyol I, Natter MD, Mandl KD. C3-PRO: Connecting ResearchKit to the Health System Using i2b2 and FHIR. *PLoS One* 2016 Mar 31;11(3):e0152722.
 39. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med* 2010 Nov 2;153(9):600–6.
 40. Wolfe BA, Mamlin BW, Biondich PG, Fraser HSF, Jazayeri D, Allen C, et al. The OpenMRS system: collaborating toward an open source EMR for developing countries. *AMIA Annu Symp Proc* 2006;1146.
 41. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics* 2015 May;41(5):404–9.
 42. Review of data security, consent and opt-outs - GOV.UK [Internet]. [cited 2017 Mar 24]. Available from: <https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs>

Correspondence to :

Dr Chris Paton
 Centre for Tropical Medicine and Global Health
 Peter Medawar Building for Pathogen Research
 South Parks Road
 Oxford OX1 3SY, UK
 E-mail: chris.paton@ndm.ox.ac.uk