Genome Medicine

Commentary

# Genomic electronic health records: opportunities and challenges

Mohammad Al-Ubaydli* and Rob Navarro†

Addresses: *UCL Centre for Health Informatics and Multiprofessional Education, Archway Campus, Highgate Hill, London N19 5LW, UK. †Sapior, 16 Byron Avenue, London E18 2HQ, UK.

Correspondence: Mohammad Al-Ubaydli. Email: me@mo.md

**Abstract**

There is value to patients, clinicians and researchers from having a single electronic health record data standard that allows an integrated view, including genotype and phenotype data. However, it is important that this integrated view of the data is not created through a single database because privacy breaches increase with the number of users, and such breaches are more likely with a single data warehouse. Furthermore, a single user interface should be avoided because each end user requires a different user interface. Finally, data sharing must be controlled by the patient, not the other end users of the data. A preferable alternative is a federated architecture, which allows data to be stored in multiple institutions and shared on a need-to-know basis. The data sharing raises questions of ownership and stewardship that require social and political answers, as well as consideration of the clinical and scientific benefits.

In the May issue of *Genome Medicine*, Belmont and McGuire [1] make the case for a 'uniform electronic health record' (EHR) that includes both genotype and phenotype information. By uniform they mean a single data standard across different EHR databases and user interfaces, rather than a single database or a single user interface (this has been confirmed by personal communication with the authors).

It is certainly true that a clearer picture of a patient's health is possible when their genotype data are combined with phenotype data. The quantity and quality of these data are improving, along with the analytical tools that allow us to interpret them. Patients, clinicians and researchers can all benefit from a better understanding of these data, and Belmont and McGuire's article [1] describes efforts in Europe and the USA to unify the datasets.

However, other parties that would benefit from better understanding include public health officials, government bureaucrats, insurance companies and employers. And in some cases, there are conflicts of interest; for example, an insurance company could use genetic information to raise premiums or deny cover, whereas a patient might use the same information to seek increased cover when they learn of the risk for future diseases.

There are ways to solve the conflicts of interest that can arise from the use and availability of patient data. First, as Belmont and McGuire [1] describe, efforts such as the Personal Genome Project [2] allow patients to opt in to fully disclose their genetic information for the benefit of researchers. PatientsLikeMe.com [3] has an openness policy alongside their privacy policy so that participants can agree to share all their data, and tens of thousands of people from around the world have already agreed to do so. The value to researchers is currently limited because the data are self-submitted rather than independently verified, but the proof that patients are willing to share their personal information is there.

The principle must still stand, however, that data sharing begins with and is controlled by the patient. This favors a single personal health record (PHR) as a database rather than a single electronic health record. PHRs are records owned and controlled by the patient [4], as opposed to EHRs, which are owned and controlled by health care practitioners.

Useful data standards for PHR and EHR communication should be expanded to fit the genomic vision that Belmont and McGuire [1] outline. In particular, the Continuity of Care Record (CCR) data format is the digital equivalent of a referral letter from one clinician to another about a patient [5]. It is supported by PHR providers such as Google Health and Microsoft HealthVault; pharmacies such as Walgreens and CVS; and providers such as MinuteClinic [6]. The Department of Health and Human Services at the National Cancer Institute unveiled a standard earlier this year for family history [7]. However, a single genomic data standard is not yet available or widely adopted.

Second, de-identification algorithms that work for genotype data are needed. De-identification is a better term than anonymization because the latter implies a binary process, which is misleading, while the former accurately conveys a spectrum. We know that de-identification algorithms are already in use when the public interest

EHR, electronic health record; PHR, personal health record.

demands phenotype sharing but patient consent is not possible or practicable. Examples include notifiable disease surveillance, public health planning and large-scale research. In these cases, looking after the patient's privacy requires measures that ensure they cannot be identified through illicit use of those data. But de-identification algorithms for genotype data are not mature enough.

Re-identification becomes more likely as the number of users increases. Illicit patient re-identification has three sources of risk: the research team, all other people who have access to these data and finally the inherent readability of the data itself [8]. Building a single system to be accessed by hundreds or thousands of researchers across tens or hundreds of projects is simply inconsistent with minimizing these three sources of risk. Such systems can therefore never be adequately private.

What might work, when public interest demands but consent is not possible, are schemes that separately copy just the minimum of phenotype and genotype data from various health management systems for a specific group of vetted researchers working within a highly protective legal context. Any change in project purpose would necessitate a re-assessment of the prevailing risks. A system in which highly vetted organizations were permitted to collect and link minimal data from all its various sources would be ideal.

In addition, the architecture for a single EHR or PHR is not a simple one. It is desirable and correct to view all the relevant data at the time of making a clinical decision or coming to a research conclusion. However, that does not mean all the data should be viewable.

For the person viewing the data, their storage in a single place does mean faster access and allows data normalization. But for the people whose data are viewed, such a data warehouse is ripe for abuse. Citizens have expressed their distrust of such systems on many occasions [9], and security experts have repeatedly pointed out the risks of data warehouses [10]. Federated architectures, where data are spread across multiple sites and queried as needed, have been deployed [11] and are made easier by new approaches, such as service-oriented architecture. And knowing how much protection to put in place is made easier by couching privacy concerns in terms of the risk of illicit patient re-identification.

## Conclusions

All of the above discussion is not to say that a single EHR is a bad idea. Belmont and McGuire [1] make a good case for the need to unify data in the service of laudable aims, including providing good patient care and advancing medical research. However, just because something can be done does not mean that it should be done, and in health care it is patients who should decide what should be done. They will be the most affected by privacy breaches, so they must be the ones who decide which of the benefits to take advantage of. The danger is when professionals confuse their convenience with the benefit of patients. The good news is that mature technologies exist that do put patients in control. As professionals we need to earn their trust by using these technologies when we ask for data sharing that makes our jobs easier.

## Competing interests

MA is the CEO of Patients Know Best, a company that makes and sells personal health record software. RN is the CEO of Sapior, a company that makes and sells de-identification software for the private sharing of health data.

## Authors' contributions

MA wrote the sections on personal health records and RN wrote those on de-identification.

## References


1. Belmont J, McGuire A: **The futility of genomic counseling: essential role of electronic health records.** *Genome Med* 2009, **1:**48.
2. **Personal Genome Project** [http://www.personalgenomes.org/]
3. **PatientsLikeMe** [http://www.patientslikeme.com/]
4. **Markle Foundation: Connecting for Health** [http://www.connectingforhealth.org/resources/final_phwg_report1.pdf]
5. **Continuity of Care Record Standard** [www.ccrstandard.com]
6. **Medpedia: Continuity of Care Record (CCR) Standard** [http://wiki.medpedia.com/Continuity_of_Care_Record_(CCR)_Standard]
7. **Cancer Biomedical Informatics Grid** [https://gforge.nci.nih.gov/projects/fhh]
8. Navarro R: **An ethical framework for sharing patient data without consent.** *Inform Prim Care* 2008, **16:**257-262.
9. McKie Robin: **Icelandic DNA project hit by privacy storm.** *The Observer* 16 April 2004 [http://observer.guardian.co.uk/international/story/0,6903,1217842,00.html]
10. Anderson R, Brown I, Dowty T, Inglesant P, Heath W, Sasse A: *Database State*. York: Joseph Rowntree Reform Trust; 2009 [http://www.cl.cam.ac.uk/~rja14/Papers/database-state.pdf]
11. Gruman G: **Massachusetts takes a spoonful of SOA.** *InfoWorld* 2 May 2005 [http://www.infoworld.com/d/architecture/massachusetts-takes-spoonful-soa-904]


Published: 22 July 2009
doi:10.1186/gm73