This is the published version of a paper published in *Journal of Internal Medicine*.

Electronic health records: new opportunities for clinical research

■ P. Coorevits[1,2], M. Sundgren[3], G. O. Klein[4], A. Bahr[5], B. Claerhout[6], C. Daniel[7], M. Dugas[8], D. Dupont[9], A. Schmidt[10], P. Singleton[11], G. De Moor[1,2] & D. Kalra[12]

From the[1]Department of Medical Informatics and Statistics, Ghent University, Ghent;[2]The European Institute for Health Records (EuroRec), Sint-Martens-Latem, Belgium;[3]AstraZeneca R&D, Mölndal, Sweden;[4]University of Science and Technology, Trondheim, Norway;[5]Sanofi R&D, Chilly-Mazarin, France;[6]Custodix NV, Sint-Martens-Latem, Belgium;[7]Paris Descartes University, INSERM, Paris, France;[8]Institute of Medical Informatics, University of Münster, Münster, Germany;[9]Data Mining International SA, Geneva;[10]Pharma Product Development, F Hoffmann-La Roche Ltd, Basel, Switzerland;[11]Cambridge Health Informatics, Cambridge; and[12]University College London, London, UK

Abstract. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, Dugas M, Dupont D, Schmidt A, Singleton P, De Moor G, Kalra D (Ghent University, Ghent; The European Institute for Health Records (EuroRec), Sint-Martens-Latem, Belgium; AstraZeneca R&D, Mölndal, Sweden; University of Science and Technology, Trondheim, Norway; Sanofi R&D, Chilly-Mazarin, France; Custodix NV, Sint-Martens-Latem, Belgium; Paris Descartes University, INSERM, Paris, France; Institute of Medical Informatics, University of Münster, Münster, Germany; Data Mining International SA, Geneva; F Hoffmann-La Roche Ltd, Basel, Switzerland; Cambridge Health Informatics, Cambridge; University College London, London, UK). Electronic health records: new opportunities for clinical research. (Review). *J Intern Med* 2013; doi: 10.1111/joim.12119.

Clinical research is on the threshold of a new era in which electronic health records (EHRs) are gaining an important novel supporting role. Whilst EHRs used for routine clinical care have some limitations at present, as discussed in this review, new improved systems and emerging research infrastructures are being developed to ensure that EHRs can be used for secondary purposes such as clinical research, including the design and execution of clinical trials for new medicines. EHR systems should be able to exchange information through the use of recently published international standards for their interoperability and clinically validated information structures (such as archetypes and international health terminologies), to ensure consistent and more complete recording and sharing of data for various patient groups. Such systems will counteract the obstacles of differing clinical languages and styles of documentation as well as the recognized incompleteness of routine records. Here, we discuss some of the legal and ethical concerns of clinical research data reuse and technical security measures that can enable such research while protecting privacy. In the emerging research landscape, cooperation infrastructures are being built where research projects can utilize the availability of patient data from federated EHR systems from many different sites, as well as in international multilingual settings. Amongst several initiatives described, the EHR4CR project offers a promising method for clinical research. One of the first achievements of this project was the development of a protocol feasibility prototype which is used for finding patients eligible for clinical trials from multiple sources.

Keywords: clinical research, electronic health records, research ethics, research techniques.

Introduction

We are currently on the edge of a golden era of medical understanding, with the amount of available information to support healthcare increasing at an enormous rate. Computer and information science concepts and tools are now part of the framework of biomedical science. Scientific computing platforms and infrastructures allow new types of experiments that were impossible to conduct only 10 years ago, changing the way scientists 'do science' [1]. The past decades of progress in health information technology (HIT) have undoubtedly reshaped the way health care is carried out and how health data are being documented. At present, healthcare practice generates data exchanges and stores huge amounts of patient-specific information [2] in electronic health records (EHRs) and ancillary databases, including in some cases emerging genome sequence data and vast

amounts of information from digital imaging examinations. This generation of electronic health data holds great promise not only to significantly contribute to healthcare provision but also to transform biomedical research.

At the same time, the knowledge explosion and an ageing society create an escalation in healthcare expenditures placing unprecedented organizational and economic pressures on healthcare systems as well as expectation on the pharmaceutical industry for the rapid development of innovative medicines [3]. The development of new medicines is critical to deliver improvements in healthcare. Most new medicines are developed by the pharmaceutical industry in collaboration with academic and healthcare organizations which, for example, conduct clinical trials and observational research. In parallel, healthcare authorities and provider organizations and academic biomedical researchers are increasingly looking at secondary uses of clinically recorded data towards optimizing the reach, success and efficiency of disease prevention, disease management and public health strategies and programmes [3].

Researchers use various methods to investigate, for example, disease comorbidities, patient stratification, drug interactions and clinical outcome from various clinical databases and registries. A critical factor for successful utilization of available health data for research is the access, management and analysis of integrated patient data, within and across different functional domains. For example, most clinical and basic research data are currently stored in disparate and separate systems, and it is often difficult for clinicians and researchers to access and share these data. Furthermore, inefficient workflow management in clinics and research laboratories has created many obstacles to medical/clinical research, decision-making and assessment of outcomes. The vitally needed change in contributing to biomedical research and other important areas such as drug discovery cannot be achieved without the availability of trustworthy and scalable reuse of EHRs [4]. Various innovative methods are being used to find meaning in these large sets of information [5].

Here, we first provide an overview of the different methods for obtaining data for clinical research processes, and then describe the fascinating possibilities provided by the new types of federated EHRs. The challenges and obstacles to increasing the scale of EHR use will be considered next, along with ways to overcome these problems, including semantic interoperability, privacy and legal concerns. Finally, the structural and political challenges to a sustainable system for clinical research in cooperation with EHR systems and important initiatives for federated EHR systems for clinical research will be described, with particular emphasis on the Electronic Health Records for Clinical Research (EHR4CR) project [6].

Obtaining data for clinical research processes

What is clinical research?

There are many different types of research questions and methodologies covered by the term 'clinical research'. The pharmaceutical industry focuses in particular on controlled clinical trials. This type of research remains very important, and there is a need to improve the efficiency and lower the cost of conducting trials whilst responding to increasing demands from regulatory bodies for more and better quality evidence of effectiveness and outcomes. Although academic clinical scientists often participate in such studies, they are also concerned with many other types of studies including comparative effectiveness research with older drugs and unselected patients with multiple diseases and various characteristics that were exclusion criteria at the time of the market approval study.

Many clinical research projects are not primarily concerned with therapy at all but investigate, for example, the natural course of diseases, criteria for diagnosis, the role of patient education and continued surveillance. Clinical research now often includes studies on the role of genes and metabolic pathways in relation to health and disease development. Some clinical research is also concerned with the function of the health system at large, with the function and effectiveness of various organizational structures and collaborations including the care and above all the costs of health care. Such studies require clinical records but also data that may be stored in various administrative databases for patient care or provider reimbursement.

Not a one-size-fits-all approach

These various types of clinical research inevitably use structured and narrative health records – increasingly from EHRs – as well as special

Table 1 *Characteristics of some sources of clinical information for research*

Data sources	Advantages	Disadvantages
Electronic health record (EHR) at a single institution	Easy management of rights and consents. Full clinical content, structured and unstructured data. Possibly same semantics for all	Too few cases for many important studies. No general purpose research tools
Special disease registers at a regional or national level (often termed quality registers)	Collect data from several institutions. Allow comparisons of results and larger samples. Well-defined data variables	Limited and relatively fixed data set. Changed rarely at the most yearly. Does not allow analyses of types of variables other than those collected. More complicated rights and consent management. Extra work to record the data. In some cases, though, it is possible to transfer data from an EHR. Often double registration in EHR and quality register
Special research database system for a specific project (e.g. a regulated clinical trial)	Very well-controlled variables including functions to ensure project process support and reasonable compliance	Expensive to set up for one project. Extra work because data cannot be retrieved from EHRs and extra work for clinical staff to transfer data from screen or paper to the research system
Federated system of electronic health records and special research project tools	May allow very large case populations, especially if federation across national borders	Semantic interoperability and consent are difficult to manage

databases for images and laboratory data including sequence data from genetic analyses that, in most cases, are stored in separate systems.

Table 1 shows some of the principal sources of health information that may be used for research. We believe that the new paradigm of federated EHRs will become an essential tool; however, different methods will continue to be explored for some aspects of clinical research for many years.

Possibilities with new types of EHRs

The era of the EHR

What is now most commonly referred to as the EHR started to enter clinical care as early as the 1960s. It is interesting to note that many of the pioneers were already at that time seeing the improved possibilities for follow-up and research as one of the most valuable reasons for the transfer from paper-based to electronic recording systems. Whilst these objectives where maintained to some

degree, the further development of clinical information systems has largely focused on improving administrative processes (including reimbursement) and, more recently, the direct provision of clinical care. Early attempts to structure data input were unfortunately replaced by large free-text narrative (letters, reports and progress notes), in most locations dictated by a physician, sometimes with speech-to-text assistance. The move to EHRs has been far from uniform in different parts of the world and has not mirrored general IT developments. In some regions, including Scandinavia and the UK, electronic systems were first adopted by primary care, whereas in others, the development was led by university clinics in large hospitals.

However, whilst the world as a whole is still far from seeing the end to paper records, there has been a very rapid expansion in the last 5–10 years to the point where now in some countries, nearly 90% of all healthcare records are digital. Indeed, a very

dramatic recent increase in the USA has been largely due to government financial incentives for EHRs with 'meaningful use' criteria [7]. Despite a few relatively new EHR system products that provide important support for some institutional research needs, most EHR systems today do not provide a good basis for clinical research.

Improving the quality of EHR data

To use EHR systems efficiently for clinical research, a number of features are required that unfortunately have often been lacking. In addition to structured data capture, functions are required to ensure the correctness, completeness and accuracy of the data within the EHR systems [8, 9]. Equally important is the assurance within EHR systems of security, with confidentiality, integrity and general trustworthiness to meet the requirements for high-quality research data [10–12] including regulated clinical trials where good clinical practice is mandated [13].

Quality assurance mechanisms may be needed to ensure that the EHR systems themselves adhere to certain quality characteristics. Third-party certification is essential in the EHR quality assurance process. The Healthcare Interoperability Testing and Conformance Harmonization (HITCH) project has provided a roadmap of how eHealth interoperability quality labelling and certification should be organized in Europe. As part of the EHR-Q Thematic Network, quality labelling and certification of EHRs have been promoted in Europe by organizing more than 70 workshops in 27 European member states, and 'data quality' has been identified as one of the key issues. The European Institute for Health Records (EuroRec) has developed and currently maintains a repository of more than 1700 EHR quality criteria (functional descriptive statements), and tools to facilitate the process of EHR quality labelling and certification.

Data quality has many dimensions such as completeness, correctness, concordance, plausibility and currency [9, 14]. A more direct involvement of the patient and next-of-kin in EHR data collection can also contribute to EHR data quality. For instance, Porter *et al.* [15] demonstrated that parental data entry is more complete than recording by physicians. New mobile computing devices enable patient questionnaires to be directly connected to EHRs [16].

On the other hand, evidence for the benefits of EHRs, in particular related to data quality, has been challenged [17]. In addition to regulatory obstacles to the reuse of EHRs, inaccurate diagnostic codes and problem lists can cause errors [18]. Botsis *et al.* [19] analysed 10 years of EHR data regarding pancreatic cancer from a major clinical data warehouse and reported between 6% and 46% incompleteness for some study variables. Similar findings regarding completeness of EHR data for recruitment of clinical trials were reported by Kopcke *et al.* [20].

Given the importance of EHR data quality, a process for quality assessment – such as monitoring of EHR data quality – should be implemented. Kahn *et al.* proposed determining the priority of variables, iterative cycles of assessment and 'detailed documentation of the rationale and outcomes of data quality assessments to inform data users' [21].

Given the poor quality of many legacy EHR systems, it is not surprising that their use for clinical research has been limited. In many cases, registries have been created with special reporting outside the normal clinical record, to serve research purposes. Some countries have invested substantially in such registries; for example, Sweden's 'quality registers', which include more than 70 conditions on a national scale and collect high-quality data with coverage that may be near 100% of all cases for some of these conditions. This has created much valuable data, many international publications and a significant impact on the practice of medicine [22]. However, the registry structures are inflexible and create significant work, even if EHR extracts using modern standards can partially automate registry population, as has been demonstrated for the Swedish Heart Failure Register.

Semantic challenges regarding the integration of EHRs

The analysis of EHRs for research, on a European scale, shares many challenges with the communication of EHRs between systems for patient care. Not only do EHR systems have markedly different repositories, the way clinical information organized within them by different teams and care settings is radically different. Some aspects are uniform in one country or institution, but other aspects of clinical recording vary between individual clinicians without any evidence-based reason.

When using EHRs for clinical research studies, different types of information need to be integrated – protocol eligibility criteria, clinical research data items and EHR data – to enable the distributed queries across multiple patient-centred sources in support of cohort identification. Health informatics research over the past two decades has focused on developing approaches to bridge heterogeneous EHRs to facilitate their consistent interpretation (known as semantic interoperability) [23].

Layered semantic models in clinical care and clinical research

In the domain of patient care, the collective international efforts of multiple standards development organizations have resulted in standards for both the structure and the semantics of clinical information that enables computable semantic interoperability between diverse systems. Three major contributions currently dominate internationally.

First, ISO EN 13606 is a generic and comprehensive representation for the exchange of EHR information between heterogeneous systems, deliberately kept as simple as possible to minimize the vendor burden of mapping to and from this intermediate representation [24]. It is ideally suited to the extraction, communication and/or mapping of longitudinal EHR data including fine-grained parts of an EHR.

Secondly, the openEHR Foundation maintains a more detailed model, catering for the widest set of use cases for patient level data, ideally suited to the implementation of a comprehensive EHR system as its persistence model [25]. This model can be seen as an extension of the formal ISO standard 13606.

Thirdly, HL7 Reference Information Model (RIM) and HL7 Clinical Document Architecture (HL7 CDA) [26] are designed to communicate a single clinical document as a message and are therefore ideally suited to a messaging environment in which HL7 version 3 is already in use for other purposes, and where the communication needed is for a single document at a time (e.g. a discharge summary).

These standards all take a 'semantic-layered' approach to representing the meaning of the clinical information they contain [27, 28]: (i) generic reference information models that can represent the common characteristics of any clinical information, such as authorships and responsibilities, dates and times of observations and healthcare activities, version management, access policies and digital signatures – it is important to note that these models require an associated, robust data type model such as that defined by ISO 21090; (ii) more detailed clinical information structures (13606/openEHR archetypes and HL7 CDA templates) that reflect the needs for documenting particular details within EHRs, such as how breathing difficulties, heart sounds, an echocardiogram, a differential diagnosis or a drug prescription should be structured [29]; and (iii) clinical terminology systems such as the International Classification of Diseases or SNOMED-CT that provide the domain of possible values for each element within an information structure.

In the domain of clinical research, the Clinical Data Interchange Standards Consortium (CDISC) has developed a number of platform-independent standards that support the electronic acquisition, exchange, regulatory submission and subsequent archiving of clinical research data. In particular, the recently released Protocol Representation Model (PRM) and Study Design Model (SDM) allow organizations to provide rigorous, machine-readable, interchangeable descriptions of the designs of their clinical studies [30, 31]. In addition, the Operational Data Model (ODM) defines the organization, structure and syntax of data captured for analysis and reporting over the course of a clinical trial [32]. Recently, the Clinical Data Acquisition Standards Harmonization (CDASH) initiative has specified the unambiguous semantics of a number of common data elements that are deemed 'common' to all trials [33]. Lastly, the Biomedical Research Integrated Domain Group (BRIDG) model, resulting from a joint effort between CDISC, HL7, the National Cancer Institute (NCI) and the US Food and Drug Administration, provides representations of the semantics of clinical research data consistent with the semantic layers described above for clinical care [34].

Achieving broad-based, scalable and computable semantic interoperability across multiple domains requires the integration of multiple standards, which therefore must be mutually consistent, coherent and cross-compatible [35–37]. Unfortunately, standards in this field have often been developed in parallel and are therefore somewhat incompatible with each other.

Towards standard-based use cases and cross-domain semantic models

Integrating the Healthcare Enterprise (IHE) has sought to address this compatibility challenge through 'integration profiles' that specify how one or more standards might be tailored and applied together to serve the interoperability needs of particular focused use cases [38]. The IHE domain Quality, Research and Public Health (QRPH) defines the information exchange profile for sharing information for quality improvement in patient care and clinical research [39]. This set of integration and content profiles addresses the issue of multivendor, scalable interoperability required for EHR-enabled research. Initially focusing on syntactic interoperability for the reuse of EHR data, a recently developed profile – data element exchange – provides a solution for sharing cross-domain semantic models. Major research efforts currently focus on defining shared sets of semantically unambiguous and context-neutral (to enable reuse) common data element definitions. The US National Cancer Institute has developed the Cancer Data Standards Repository (caDSR) initiative to standardize common data elements used in cancer research [40, 41]. Similarly, CDISC Shared Health and Research Electronic Library (CSHARE) aims to build a global, accessible electronic library, to enable data element definitions [42]. CSHARE, which is similar to NCI caDSR, utilizes the ISO/IEC 11179 standard as the semantic basis for the metadata repository of Common Data Elements [43]. In the EHR4CR project [6], in collaboration with the European project SALUS [44], we explore the advantage of using a variety of semantic web tools and technologies in support of the representation and sharing of cross-domain semantics [45, 46].

Privacy: ethical and legal challenges to federated research

Legal and ethical aspects of using EHRs for research

It is essential to use patients' medical information for secondary purposes, beyond care of the individual concerned, for the high quality of healthcare delivery and the effectiveness of scientific research [47]. The use of EHRs for clinical research is inevitably challenged both by legal and ethical considerations [48]. A balance must be found to enable scientific research progress within a framework in which the privacy of patients is not compromised.

The ethical issues are generally similar across different cultures and healthcare systems [8], although priorities and practical solutions may vary considerably from one environment to another.

Additionally, laws and regulations differ substantially for processing personal data in different countries. Even where some harmonization exists in the general data protection legislation, in the EU achieved by the Data Protection Directive (presently undergoing revision that may lead to a uniform EU-wide regulation), many additional laws regarding medical research vary between jurisdictions. This fact and possible misinterpretations of the spirit of the law can create difficulties and prevent multicountry collaborative research projects involving several jurisdictions.

These differences in laws and ethical approaches and their interpretations create a number of pragmatic issues (see Table 2) surrounding the reuse of EHR data for clinical research.

The 'consent model' and the 'trust model' are two possible approaches to address some of these challenges for a research network based on federated EHRs.

The consent model
It is debatable whether explicit consent is required for reuse of key-coded (pseudonymized) EHR data for research and statistical purposes [51]. In legal terms, it is possible as it may be considered a 'compatible use' consistent with the original collection of the data (for healthcare) and it may fall outside the scope of the principles of personal data protection regulations [52]. In some countries, special legislation may require primary EHR data to be submitted for public health purposes to national or regional registries without the need for consent of the data subject.

Many difficulties arise if explicit consent is required for a clinical research project, as outlined in Table 2. Alternatively, or more often in addition to consent of data subjects or their proxy, a collective decision or 'social consent' by a research ethics committee or similar body might be possible or necessary.

The trust model
The second approach is to reduce the information content of the data so that individuals can no longer be identified. In this case, there would be no privacy risks and consent would no longer be required; this could be termed 'effectively anonymized'

Table 2 *The most common issues encountered in collaborative projects where different laws and/ or institutional ethical frameworks apply*

Issue	Identified problems
Gaining retrospective consent	Too difficult, too costly or requires disproportionate effort (e.g. patients may have moved or changed their names)
Gaining broad prospective consent	Difficult to ensure that the data subject is 'fully informed' [49]. Also, research methods and detailed research questions may change over time. Is the broad consent still valid?
Gaining dynamic consent	This model in which the data subjects are continuously informed about the project progress and asked to reaffirm their consent with new directions may seem to be the solution in the Internet age, but there are also good arguments against close inclusion of patients in research project steering [50]
Gaining early consent (as part of treatment)	May be deemed 'coercive'
Legal position of 'nearly anonymized' data	It would help scientists to understand what is really expected from them to ensure compliancy when reusing EHRs for research
Use of the 'precautionary principle' by data 'gatekeepers'	Practical interpretation will be more restrictive than legislators intended
Lack of consistency in interpretation of the legal position between regulators or approval bodies, such as research ethics committees	This is especially important where the consent process may be affected

data, although there is no clear definition and, with the levels of information currently available online, it can be hard to ensure that any data set is fully anonymized [53].

The uncertainties of the legal position of 'nearly anonymized' data make it difficult for researchers to know when they are being compliant with the law whilst reusing EHRs for research. There are similar uncertainties for the representatives of the 'data controller' at a healthcare institution to know what levels of data they can safely release. It is often easier for such 'data gatekeepers' to use the 'precautionary principle' [54] and not release the data. This is further compounded by different interpretations and approval processes at each institution [55]; what is acceptable at one institution may not be acceptable or practical at another. Thus, finding a common approach can be nearly impossible.

Privacy protection and security measures

De-identification
One of the important questions for privacy protection is whether microdata (data pertaining to

discernible individuals) are required for research or whether aggregated results are sufficient. Numerous approaches and techniques have been proposed and studied with respect to the de-identification (anonymization) of microdata. Their main objective is to maximize the information content level whilst minimizing the re-identification risk with respect to the individuals involved (with mathematically provable guarantees). These approaches usually encompass a combination of techniques such as generalization [56, 57], suppression [58], global recoding [59], Post RAndomisation Method (PRAM) [60], microaggregation [61], top and bottom coding [59] and slicing [62, 63].

At the same time, various grouping-based transformation strategies have been defined for determining whether a data set is safe for disclosure, the most well known of which is 'k-anonymity' [64–69].

The above techniques do not, however, solve de-identification problems as unfortunately they tend to excessively reduce the amount of information. The concept of 'contextual anonymity' [70] was introduced in the Advancing Clinico-Genomic

Trials on cancer: Open Grid Services for improving Medical Knowledge Discovery (ACGT) EU research project (i.e. an operational environment in which data can be considered *de facto* anonymous). The proposed Data Protection Framework combines de-identification with a contractual framework (managed by the nonprofit organization Center for Data Protection [71]) and a wide range of technical security measures. This framework and its tools (e.g. the Custodix Anonymisation Services [70]) have been successfully used in several EU projects for reusing medical data.

In addition to relying solely on de-identification, application information flows can also be designed in such a way that no microdata are required beyond the original hospital environment, for instance, by introducing distributed privacy-preserving data mining algorithms [72]. Some data reuse applications inherently only require aggregate results from the EHR (e.g. trial protocol feasibility studies only need patient counts). Nevertheless, even in these cases, it remains necessary to perform a proper risk assessment. For example, applications that query an EHR only to retrieve aggregate results might still need specific disclosure control protection when the query results return too small aggregated groups.

Security
'Basic' security (authentication, authorization and audit) is a fundamental requirement of each IT system. However, some topics are of particular interest when dealing with data reuse, especially when relatively large distributed networks are involved (e.g. trial protocol feasibility studies, patient recruitment and data export to registries).

1 Access control management and enforcement
  Crossorganization EHR data reuse (sharing) translates into complex security policies that need to be uniformly managed and enforced. New complex requirements include for example the capability of dealing with data-binding concepts such as 'purpose of use' and 'conditions on use' (cf. privacy metadata, 'sticky' policies [73]).

2 Consent management
  Consent is closely related to authorization (it can be seen as a kind of access policy determined by the data subject). When consent is electronically managed, it can be included into the overall governance and be 'enforced' automatically [74].

In this context, there are two interesting projects working at the forefront of this area, EHR4CR [6] and EURECA [75]. The former focuses on the practical side of public–private cooperation in this newly developing area, the latter on defining a unified security framework (alongside a legal framework) with the aim of offering regulatory compliance 'by design' [76].

Structural and political challenges

Given a growing healthcare demand and limited resources, health technologies must provide meaningful benefits to different stakeholders, such as improved health outcomes to patients and cost optimization to payers [77–79]. Considering that patients will soon navigate between healthcare points along with their EHR and other data, health systems must evolve to take advantage of all the data available in this new landscape driven by information technologies. Consequently, there is a need to develop scalable integrated healthcare platforms, as well as potent aggregators for managing health data across different systems and data sources [3].

In particular, for patients and their families and care givers, EHR-integrated research platforms will provide a secure environment to share health data, for advancing clinical research towards achieving faster access to safe and effective innovative medicines. For the research community, EHR-enabled research will optimize research and development platforms, processes and timelines. For the pharmaceutical industry, the reuse of EHR data will maximize the R&D value chain by generating high-quality clinical evidence faster through better protocol feasibility assessment, improved patient identification and recruitment, and more efficient clinical study conduct, including for reporting serious adverse events. For contract research organizations, EHR-enabled clinical research will maximize the value to customers and diversify revenue streams. For clinical investigators and primary and secondary care physicians, having access to the most modern, trustworthy and efficient EHR-integrated research environments will enable their participation in a larger number of clinical trials. For regulatory agencies, the reuse of EHR health data for research will generate comprehensive clinical evidence more rapidly for assisting regulatory decision-making. For public and private payers, EHR health data mining will enable further cost-effectiveness research to assist

optimal reimbursement decisions. For hospitals and healthcare organizations, participating in EHR-integrated research will enhance EHR data quality, as well as management reporting, performance benchmarking, optimization of care pathways and research revenue. For academic centres, mining EHRs will generate more research opportunities and funding, including in emerging domains. For the industry of HIT, technical vendors, trusted third parties and service providers, EHR research platforms will open new business opportunities facilitated by sustainable business models.

Overall, the reuse of EHR data for clinical research will optimize clinical development towards achieving faster access to innovative medicines. Considering R&D costs of €1.1 billion for each new chemical or biological entity [80, 81], and the large number of clinical trials that the pharmaceutical industry must conduct to achieve regulatory approval and reimbursement, the efficiency gains from EHR-integrated research platforms will provide key competitive assets. The deployment of value-based innovation across the R&D framework also involves integration of patient-oriented programmes, evidence-based approaches and multi-stakeholder strategies, from early clinical research phases to lifecycle management, and beyond [78, 79, 82, 83]. These opportunities will be maximized with the adoption of EHRs by patients, health providers and researchers, and by achieving interoperability [79, 84, 85]. Such integrated approaches will enrich health data and will improve clinical research and patient care [5, 79, 82].

For healthcare systems, the opportunity to optimize health outcomes of target populations through the timely delivery of healthcare interventions, including innovative medicines, and to monitor their effectiveness in real-life settings using EHR-integrated research platforms, will provide an important strategic tool for addressing public health priorities.

Important initiatives for federated clinical research

There are currently several ongoing projects dealing with the (re)use of EHR data for the purpose of clinical research. In the USA, initiatives such as i2b2 [86], the eMERGE network [87], the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH) [88] and the Million Veteran Program [89] are focusing on integrating EHRs and genomic data [5]. The Stanford Translational Research Integrated Database Environment (STRIDE) is an example of a US project that aims to create an informatics platform supporting clinical and translational research [90].

In Europe, several research projects and initiatives such as the i4health network [91], EMIF (European Medical Information Framework) [92], eTRIKS (Delivering European translational information & knowledge management services) [93], EURECA (Enabling information re-use by linking clinical research and care) [75], INTEGRATE (Integrative cancer research through innovative biomedical infrastructures) [94], Linked2Safety [95], SALUS (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies) [44], TRANSFoRm (Translational Research and Patient Safety in Europe) [96] and EHR4CR (Electronic Health Records for Clinical Research) [6] are all concerned with re(using) EHRs for facilitating clinical research, thereby focusing on different disease domains and addressing different use cases and scenarios. The EHR4CR project is addressing many of the challenges discussed in this review and will therefore be described in detail below.

The EHR4CR project

Overview and objectives

The EHR4CR project is part of the European Innovative Medicines Initiative (IMI) programme. The 4-year project is ongoing (2011–2014), has a budget of more than 16 million Euros and involves 35 academic and private partners (including 10 pharmaceutical companies. The consortium includes also 11 hospital sites in France, Germany, Poland, Switzerland and the United Kingdom. The authors of this publication are all members of this consortium. An aim of the EHR4CR project is to demonstrate how data held in EHRs can be reused to enhance clinical research processes, in a multi-national context, whilst protecting privacy. The project will provide a robust platform accompanied by a portfolio of relevant services (protocol feasibility, patient identification and recruitment, clinical trial conduct and serious adverse event reporting services) to demonstrate sustainable, scalable and cost-effective solutions. The EHR4CR platform will also be supported by an innovative business model (e.g. governance model, accreditation and financial mechanisms) and a customized value proposition [81].

Technical approach

The EHR4CR platform will be developed and implemented as a common set of components and services that will allow the integration of the lifecycle of clinical studies with heterogeneous clinical systems, thereby facilitating data extraction and aggregation, workflow interactions, privacy protection, information security, and compliance with ethical, legal and regulatory requirements. This will help to speed up the protocol feasibility refinement process with rapid feedback on population numbers and their geographical distribution, to assist in identifying suitable patients via their nominated care providers, and to accelerate and improve the accuracy of patient recruitment and trial execution, and to enable more complete and real-time safety monitoring. The organizational model, with inclusion of an independent trusted third party, will also allow for additional kinds of data transactions between different stakeholders and environments [e.g. platform-level audit trial (re)construction and specific (de-identified) data exchanges outside the scope of the standard scenarios].

Pilot sites will use de-identified EHR data from the EHR4CR hospital partner sites to validate the platform and the proof-of-concept services and to provide input to the EHR4CR business model. The EHR4CR consortium and the hospital sites involved have been chosen intentionally in such a way as to ensure the necessary success factors for obtaining future solutions for the reuse of EHR data across different legal frameworks. The project will primarily address the following disease areas included in the pilot sites: oncology, inflammation, neuroscience, diabetes and cardiovascular and respiratory diseases. These areas are relevant to current pharmaceutical industry research interests, and align with clinical research and data resources at the pilot sites.

Business model approach

The EHR4CR business model will provide a systematic, structured and scalable approach to the use of EHR data for clinical research. It will define how the platform and its complementary services will be funded and sustained in the long term. The project uses a formal approach and business model innovation best practices [97, 98] for guiding the design of a sustainable and operational business model framework. This process includes the design and development of EHR4CR sustainability strategies, governance model and business model

core capabilities, namely: (i) EHR4CR service offering and value propositions; (ii) customer segmentation and management; (iii) organizational infrastructure (resources, activities and processes, including accreditation and certification); and (iv) financial schemes (cost structure and revenue streams).

The business model involves the development of comprehensive and customized value propositions describing the expected benefits that an organization offering the service promises to deliver to its stakeholders in relation to their needs [97–99].

Results after 2 years of progress

During the first 2 years, the project has produced a number of deliverables. A first version of the EHR4CR information model (a platform-independent conceptual model) has been developed, based on generic reference models for representing clinical data (e.g. ISO/HL7 RIM and CDISC/HL7 BRIDG) and data elements of standard data types [46].

Software requirement specification for the protocol feasibility service (PFS) and patient identification and recruitment service (PRS) has been completed. The first version of the EHR4CR platform, including the PFS, has been developed based on a service-oriented architecture (SOA) in which service providers and consumers can dynamically connect. As such, the primary goal of the EHR4R architecture is the specification of clearly defined interfaces and responsibilities supporting potentially any physical location of service consumers and providers. Data end-points (e.g. the connections between the platform and each hospital) are key service elements in the EHR4CR platform from which the different scenarios can be built.

The viability and performance of the EHR4CR platform and the PFS have been tested with good results by connecting 11 hospitals to the platform using a list of the 82 most important EHR data elements. Feasibility queries from 10 different (recently performed) clinical studies were evaluated in real time using a graphical user interface allowing specification of Boolean and temporal constraints between individual eligibility criteria (Fig. 1).

In assessing the PFS, all 10 European Federation of Pharmaceutical Industries and Associations (EFPIA) partners participated in user acceptance
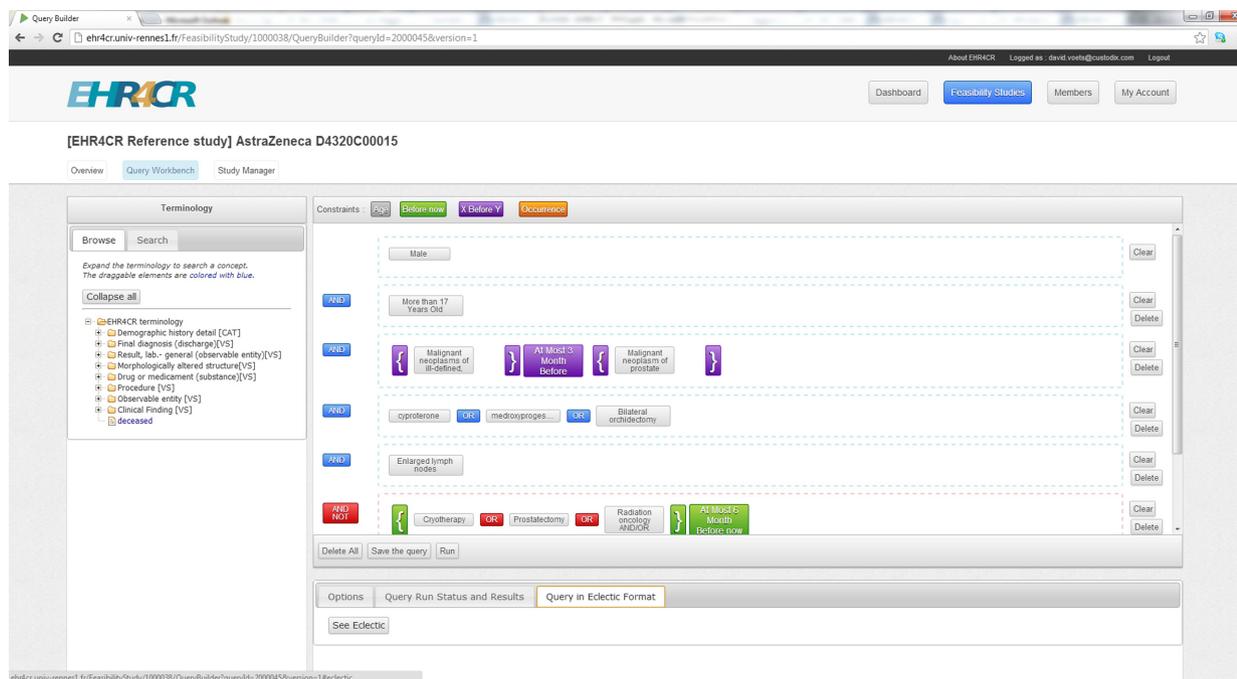
Fig. 1 *Screenshot of the EHR4CR platform user interface for the protocol feasibility service.*

testing. Overall, 373 free-text eligibility criteria were reviewed by clinical trial experts; 175 feasibility criteria were transformed into a computable representation. In addition, pilot sites mapped approximately 300 codes from their local terminologies. After running an eligibility query, the results can be visualized by showing the overall results and with the possibility to analyse separately on the basis of patient demographics (age categories and gender) and individual eligibility as well as for individual sites.

The EHR4CR business model framework has been developed, and preliminary simulations suggest that the model would be profitable (for different parties including the pharmaceutical industry, system vendors and hospitals) and sustainable over a 5-year time period, contingent upon swift adoption of EHR4CR services at project completion and steady market uptake thereafter. Further simulations using consolidated market assumptions are currently in progress.

Conclusion

EHRs have a great potential to support clinical research, including but certainly not limited to

clinical trials for new medicines. However, there are a number of challenges to achieving this on a European scale and it may be some time before the analysis of routinely collected EHR data can replace traditional clinical trial workflows. Nevertheless, we believe that modern quality-controlled EHRs, combined with a platform that supports semantic interoperability, protects privacy and provides various clinical research tools, can offer very important opportunities for new clinical research, beyond the single institution and in some cases beyond national borders. This research will be faster, of higher quality and use fewer resources, towards a goal where each patient case can be used to improve knowledge, that is, basic biomedical understanding as well as new insights into the currently most effective and efficient diagnostic and therapeutic processes. The European research initiative EHR4CR has an important part in developing a number of innovative services to support federated clinical research based on the semantic integration of different EHR system products, across organizations and across countries. Attention is being paid to the ethical considerations and to ensuring appropriate security measures for de-identification, paired with security measures for confidentiality, integrity, availability and auditability,

using cryptographic techniques and public key infrastructures.

Hence, advanced EHR-integrated platforms will provide truly innovative solutions which promise to revolutionize clinical research, to advance clinical care, and to bring significant benefits to many stakeholders, including patients, health systems, researchers, industry and society.

Conflict of interest statement

No conflict of interests were declared.

Acknowledgements

References

1 'Towards 2020 Science' report. Available from: http://research.microsoft.com/towards2020science. 2006.

2 Beck T, Gollapudi S, Brunak S *et al.* Knowledge engineering for health: a new discipline required to bridge the "ICT gap" between research and healthcare. *Hum Mutat* 2012; **33:** 797–802.

3 Kairos Future. The Data Explosion and the Future of Health: What every decision-maker in the health and healthcare industries need to know about the coming revolution. Global Strategic Analysis Report. 2011.

4 Geissbuhler A, Safran C, Buchan I *et al.* Trustworthy reuse of health data: a transnational perspective. *Int J Med Informatics* 2013; **82:** 1–9.

5 Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012; **13:** 395–405.

6 Electronic Health Record Systems for Clinical Research (EHR4CR). Available from: http://www.ehr4cr.eu.

7 Desroches CM, Audet AM, Painter M, Donelan K. Meeting meaningful use criteria and managing patient populations: a national survey of practicing physicians. *Ann Intern Med* 2013; **158:** 791–9.

8 Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Informatics* 2008; **77:** 291–304.

9 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *JAMIA* 2013; **20:** 144–51.

10 De Moor G, O'Brien J, Fridsma D *et al.* Policy brief on the current status of certification of electronic Health Records in the US and Europe. *Stud Health Technol Inform* 2011; **170:** 83–106.

11 Hoerbst A, Ammenwerth E. Electronic health records. A systematic review on quality requirements. *Methods Inf Med* 2010; **49:** 320–36.

12 Hoerbst A, Ammenwerth E. Quality and Certification of Electronic Health Records: an overview of current approaches from the US and Europe. *Appl Clin Inform* 2010; **1:** 149–64.

13 International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Guideline for Good Clinical Practice (GCP), 2002.

14 McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. *AMIA Annu Symp Proc* 2012; **2012:** 1302–9.

15 Porter SC, Mandl KD. Data quality and the electronic medical record: a role for direct parental data entry. *Proc AMIA Symp* **1999:** 354–8.

16 Fritz F, Balhorn S, Riek M, Breil B, Dugas M. Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. *Int J Med Informatics* 2012; **81:** 303–13.

17 Black AD, Car J, Pagliari C *et al.* The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011; **8:** e1000387.

18 Weng C, Appelbaum P, Hripcsak G *et al.* Using EHRs to integrate research with patient care: promises and challenges. *JAMIA* 2012; **19:** 684–7.

19 Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *AMIA Summits Transl Sci Proc* 2010; **2010:** 1–5.

20 Kopcke F, Trinczek B, Majeed RW *et al.* Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak* 2013; **13:** 37.

21 Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care* 2012; **50(Suppl):** S21–9.

22 Ovretveit J, Keller C, Hvitfeldt Forsberg H, Essen A, Lindblad S, Brommels M. Continuous innovation: developing and using a clinical database with new technology for patient-centred care–the case of the Swedish quality register for arthritis. *Int J Qual Health Care* 2013; **25:** 118–24.

23 Stroetman VKD, Rector A, Rodrigues J-M, Stroetman K, Surjan G *et al.* Semantic Interoperability for Better Health and Safer Healthcare. 2009.

24 EN ISO 13606-1:2012, Health informatics – Electronic health record communication – Part 1 – Reference model, International Organization for Standardization, Geneva 2008 and CEN – The European Committee for Standardization, Brussels, 2012.

25 Kalra D, Beale T, Heard S. The openEHR Foundation. *Stud Health Technol Inform* 2005; **115:** 153–73.

26 Dolin RH, Alschuler L, Boyer S *et al.* HL7 clinical document architecture, release 2. *JAMIA* 2006; **13:** 30–9.

27 Bointner K, Duftschmid G. HL7 template model and EN/ISO 13606 archetype object model - a comparison. *Stud Health Technol Inform* 2009; **150:** 249.

28 Kalra D, Musen M, Smith B, Ceusters W, De Moor G. ARGOS policy brief on semantic interoperability. *Stud Health Technol Inform* 2011; **170:** 1–15.

29 Goossen W. Representing knowledge, data and concepts for EHRS using DCM. *Stud Health Technol Inform* 2011; **169:** 774–8.

30 Clinical Data Interchange Standards Consortium (CDISC) Study Design Model in XML (SDM-XML). Available from: http://www.cdisc.org/stuff/contentmgr/files/0/8c85b168e 80d6834ded59339b55fdbc7/misc/cdisc_sdm_xml_1.0.pdf.

31 Clinical Data Interchange Standards Consortium (CDISC) Protocol Representation Model (PRM). Available from: http://www.cdisc.org/protocol.

32 Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). Available from: http://www.cdisc.org/odm.

33 Clinical Data Interchange Standards Consortium (CDISC) Clinical Data Acquisition Standards Harmonization (CDASH). Available from: http://www.cdisc.org/CDASH.

34 Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. *JAMIA* 2008; **15:** 130–7.

35 Hammond WE, Jaffe C, Kush RD. Healthcare standards development. The value of nurturing collaboration. *J AHIMA* 2009; **80:** 44–50; quiz 1–2.

36 Mead CN. Data interchange standards in healthcare IT–computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? *J Healthc Inf Manag* 2006; **20:** 71–8.

37 Semantic Interoperability for Better Health and Safer Healthcare. Deployment and Research Roadmap for Europe. 2009.

38 Integrating the Healthcare Enterprise (IHE). Available from: http://www.ihe.net.

39 IHE Quality, Research and Public Health (QRPH) domain. Available from: http://www.ihe.net/Quality_Research_and_ Public_Health.

40 Komatsoulis GA, Warzel DB, Hartel FW *et al.* caCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. *J Biomed Inform* 2008; **41:** 106–23.

41 Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P. Common data element (CDE) management and deployment in clinical trials. *AMIA Annu Symp Proc*: 2003;1048.

42 Jiang G, Solbrig HR, Iberson-Hurst D, Kush RD, Chute CG. A collaborative framework for representation and harmonization of clinical study data elements using semantic MediaWiki. *AMIA Summits Transl Sci Proc* 2010; **2010:** 11–5.

43 Ngouongo SM, Lobe M, Stausberg J. The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research? *J Biomed Inform* 2013; **46:** 318–27.

44 Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies (SALUS). Available from: http://www.salusproject.eu.

45 Laleci G, Yuksel M, Dogac A. Providing semantic interoperability between clinical care and clinical research domains. *IEEE J Biomed Health Inform* 2012; doi: 10.1109/ TITB.2012.2219552

46 Ouagne D, Hussain S, Sadou E, Jaulent MC, Daniel C. The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. *Stud Health Technol Inform* 2012; **180:** 534–8.

47 Jones JB, Stewart WF, Darer JD, Sittig DF. Beyond the threshold: real-time use of evidence in practice. *BMC Med Inform Decis Mak* 2013; **13:** 47.

48 Opinion 03/2013 of Article 29 data protection working party on "purpose limitation" adopted on 2 April 2013.

49 Hempel C, Lomax G, Peckman S. Broad consent in biobanking. *Nat Biotechnol* 2012; **30:** 826.

50 Steinsbekk KS, Myskja BK, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? *Eur J Hum Genet* 2013; **21:** 897–902.

51 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Article 13 para 2; article 32 para 3. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

52 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Recitals 28 & 29; Article 6 para 1 (b).Available from: http://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

53 Ohm P. Broken premises of privacy: responding to the surprising failure of anonymization. *UCLA Law Rev* 2010; **57:** 1701.

54 Communication from the European Commission on the precautionary principle, COM (2000) 1. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do? uri=CELEX:52000DC0001:EN:NOT

55 Eurobarometer. Data Protection in the European Union. Data controllers' perceptions. Analytical Report. 2008.

56 LeFevre K, DeWitt DJ, Ramakrishnan R. Incognito: efficient full domain k-anonymity. *Proc ACM SIGMOD* 2005; 49–60.

57 Sun X, Li M, Wang H, Plank A. An efficient hash-based algorithm for minimal k-anonymity. *Proc Thirty-First Aust Conf Comp Sci* 2008; **74:** 101–7.

58 Ciriani V, De Capitani di Vimercati S, Foresti S, Samarati P. k-Anonymity. Yu T, Jajodia S, eds. *Secure Data Management in Decentralized Systems*. Philadelphia, PA: Springer US, 2007; 323–53.

59 μ-ARGUS. EESnet-project. Statistics Netherlands, 2008.

60 Gross B, Guiblin P, Merrett K. Implementing the Post Randomisation method to the Individual Sample of Anonymised Records (SAR) from the 2001 Census. Statistical Disclosure Control Centre Methodology Group, Office for National Statistics, 2004.

61 Hansen SL, Mukherjee S. A polynomial algorithm for optimal univariate microaggregation. *Knowl Data Eng* 2003; **15:** 1043–4.

62 Arnab Bhattacharya RG. t-closeness: Privacy Preserving Data Mining. 2009.

63 Li TC, Li NH, Zhang J, Molloy I. Slicing: a new approach for privacy preserving data publishing. *Knowl Data Eng* 2012; **24:** 561–74.

64 Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. l-diversity: privacy beyond k anonymity. *Proceedings of the 22nd Conference on Data Engineering* 2007, doi:10.1145/ 1217299.1217302

65 El Emam K, Dankar FK, Issa R *et al.* A globally optimal k-anonymity method for the de-identification of health data. *JAMIA* 2009; **16:** 670–82.

66 Domingo-Ferrer J, Sebe F, Solanas A. An Anonymity Model Achievable Via Microaggregation. In: Jonker W, Petkovic M, eds. *Secure Data Management.* Heidelberg, Berlin: Springer-Verlag, 2008; **5159:** 209–18.

67 Ninghui L, Tiancheng L, Suresh V., t-closeness: privacy beyond k-anonymity l-diversity. *Proceedings of the IEEE 23rd International Conference on Data Engineering* 2007: 106–15.

68 Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. . *Proceedings of the IEEE Symposium on Research in Security and Privacy* 1998: 384–93.

69 Truta TM, Campan A, Meyer P. Generating microdata with p-sensitive k-anonymity property. *Sec Data Manag Proc* 2007; **4721:** 124–41.

70 Claerhout B, Forgo N, Krugel T, Arning M, De Moor G. A data protection framework for trans-European genetic research projects. *Stud Health Technol Inform* 2008; **141:** 67–72.

71 Center for Data Protection (CDP). Available from: http://www.privacypeople.org.

72 Aggarwal C, Philips Y. Privacy-Preserving Data Mining. 2008.

73 Chadwick D, Lievens S. Enforcing "Sticky" Security Policies throughout a Distributed Application.

74 Consent in a Trial and Care Environment (CONTRACT). Available from: http://www.contract-fp7.eu.

75 Enabling information re-use by linking clinical research and care (EURECA). Available from: http://eurecaproject.eu

76 Vdovjak R, Claerhout B, Bucur A. BRIDGING THE GAP BETWEEN CLINICAL RESEARCH AND CARE - Approaches to Semantic Interoperability, Security & Privacy. *Proceedings of the International Conference on Health Informatics* 2012: 281–6

77 Epstein RS, Sidorov J, Lehner JP, Salimi T. Integrating scientific and real-world evidence within and beyond the drug development process. *J Comp Eff Res* 2012; **1(Suppl. 1):** 9–13.

78 Russo MJ, Balekdjian D. Weighing the outcomes: the emergent healthcare value era, its consequences, and what drug firms need to do about it. *Nat Biotechnol* 2008; **26:** 173–82.

79 Salimi T, Lehner JP, Epstein RS, Tunis SR. A framework for pharmaceutical value-based innovations. *J Comp Eff Res* 2012; **1(Suppl. 1):** 3–7.

80 EFPIA. The Pharmaceutical Industry in Figures. Key Data 2009. Update. 2009; 1–27.

81 Kalra D, Schmidt A, Potts HWW, Dupont D, Sundgren M, De Moor G. Case report from the EHR4CR Project – A European survey on electronic health record systems for clinical research. *Health Conn* 2011; **1:** 108–13.

82 Lehner JP, Epstein RS, Salimi T. Integrating new approaches for clinical development: translational research and relative effectiveness. *J Comp Eff Res* 2012; **1(Suppl. 1):** 15–21.

83 Epstein RS. R&D transformation and value-based innovation. *J Comp Eff Res* 2012; **1(Suppl. 1):** 1–2.

84 Hillestad R, Bigelow J, Bower A *et al.* Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff* 2005; **24:** 1103–17.

85 Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. *Health Aff* 2005; Suppl Web Exclusives: W5–10-W5-8.

86 i2b2 initiative. Available from: https://www.i2b2.org/

87 eMERGE network. Available from: http://emerge.mc.vanderbilt.edu.

88 Kaiser RPGEH Program. Available from: http://www.dor.kaiser.org/external/dorexternal/rpgeh.

89 Million Veteran Program. Available from: http://www.research.va.gov/mvp/default.cfm.

90 The Stanford Translational Research Integrated Database Environment (STRIDE). Available from: https://clinicalinformatics.stanford.edu/research/stride.html.

91 i4health network. Available from: http://www.i4health.eu.

92 European Medical Information Framework (EMIF). Available from: http://www.imi.europa.eu/content/emif.

93 European Translational Information and Knowledge Management Services (eTRIKS). Available from: http://www.etriks.org.

94 Integrative Cancer Research Through Innovative Biomedical Infrastructures (INTEGRATE). Available from: http://www.fp7-integrate.eu.

95 A Next-Generation, Secure Linked Data Medical Information Space For Semantically-Interconnecting Electronic Health Records and Clinical Trials Systems Advancing Patients Safety In Clinical Research (Linked2Safety). Available from: http://www.linked2safety-project.eu.

96 Translational Research and Patient Safety in Europe (TRANS-FoRm). Available from: http://www.transformproject.eu.

97 Osterwalder A, Pigneur Y. *Business Model Generation.* Amsterdam,The Netherlands, Self Published, 2009. ISBN 978-2-8399-0580-0, 2009.

98 Zott C, Amit R, Massa L. The Business Model: Theoretical Roots, Recent Developments and Future Research. Working Paper WP-862, IESE Business School, University of Navarra 2010: 1–45.

99 Barnes C, Blake H, Pinder D. *Creating and delivering your value proposition: managing customer experience for profit.* Kogan Page. 2009.

Correspondence: Pascal Coorevits, Department of Medical Informatics and Statistics, Ghent University, c/o Business Complex Groeninghe – Building F, Zwijnaardsesteenweg 314, 9000 Ghent, Belgium.
(fax +32 9 3313350, email: pascal.coorevits@telenet.be).■