Strategic Report on the Re-use of Clinical Information for Research in the European Union

January 2012
Possible inaccuracies of information are under the responsibility of the project team. This report reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein.
Executive Summary

Purpose of this document and the endeavour

This strategic report has been prepared in order to assess the opportunities and obstacles that confront us as Europe plans to facilitate the re-use of clinical data in biomedical research. It represents the outcome of state-of-the-art literature reviews, expert consultations and five hours of intensive discussions within a Think Tank convened at UCL on 24 June 2011 under the auspices of the INBIOMEDvision Consortium (ICT-270107; http://www.inbiomedvision.eu).

It represents a consensus among the 27 invited experts who participated. These were drawn from a wide range of backgrounds, including clinicians, physicians, engineers, medical scientists, IT directors within hospitals, industry representatives, and scientists active on translational research projects. The Think Tank was co-chaired by Prof Peter Coveney (University College London, UK), Prof. Dr. Norbert Graf (University of Saarland, Germany), and Dr. Nour Shublaq (University College London, UK). The names and affiliations of the experts are listed in the Appendix.

The report is divided into two sections. Since all discussions around the re-use of clinical data assume that such data will be principally available electronically, we look first at the necessary IT considerations for the design and development of the proposed systems. We then move on to consider ways of exploiting medical data in research, looking at the expected benefits of such research, the therapeutic areas where it should be of most value, and further informatics concerns. The voice of the patient was kept central in all the discussions, and that emphasis is reflected in the report.

I would like to acknowledge the input from the experts, and in particular Dr. Clare Sansom for her assistance in preparing this report. Special thanks to Prof. Peter Coveney for his valuable comments and support throughout.

Dr. Nour Shublaq
University College London
January 2012
n.shublaq@ucl.ac.uk
# Table of Contents

**EXECUTIVE SUMMARY** ........................................................................................................ 1

**PURPOSE OF THIS DOCUMENT AND THE ENDEAVOUR** ...................................................... 1

1. **INTRODUCTION** .................................................................................................................. 1
   1.1. **SETTING THE SCENE** ..................................................................................................... 1
   1.2. **THE IMPORTANCE AND IMPACT OF THIS STRATEGIC REPORT** .............................. 2

2. **IT CONSIDERATIONS FOR MEDICAL INFORMATICS** .................................................... 3
   2.1. **DATA INTEGRATION AND MANAGEMENT OF CLINICAL SYSTEMS** ......................... 3
   2.2. **INFORMATION ASSURANCE** ....................................................................................... 6
   2.3. **SUMMARY** ................................................................................................................... 8

3. **EXPLOITING MEDICAL DATA IN RESEARCH** ................................................................... 9
   3.1. **RESEARCH BENEFITS FROM CLINICAL DATA AVAILABILITY / HEALTHCARE INFORMATION INCORPORATION INTO "OMICs" AND MOLECULAR LEVEL RESEARCH** ................................................................ 9
   3.2. **CLINICAL AREAS TO BE TARGETED BY RESEARCH FUNDING** ................................ 9
   3.3. **PERSONALISING ELECTRONIC HEALTH RECORDS / VALUE IN INDIVIDUAL ACCESS TO PERSONAL HEALTHCARE DATA** ................................................................................................................................. 10
   3.4. **DESIRABLE EVOLUTION OF BIOMEDICAL INFORMATICS IN RELATION TO RE-USE OF CLINICAL INFORMATION** ................................................................. 11
   3.5. **SUMMARY** ................................................................................................................... 13

4. **CONCLUSIONS AND RECOMMENDATIONS** ................................................................. 14

**REFERENCES** ....................................................................................................................... 15

**APPENDIX** ............................................................................................................................ 17
1. Introduction

1.1. Setting the Scene

Basic information technology considerations are of primary importance in the design of health informatics systems. The first section of this report examines data integration and the management of digitised clinical systems, and information assurance issues including security and information governance. If clinical data are to be re-used for research, data quality standards need to be set and imposed. Clinicians are busy people, often working in stressful situations, and most are not IT experts, so data capture systems must be reliable and easy-to-use. The interoperability of IT systems should be promoted to allow data integration and to build increasingly complex environments in a scalable and re-usable fashion; where possible, systems should be interoperable between hospitals, across regional, national and international boundaries. Standards should therefore be adopted on at least a Europe-wide scale. Security and privacy are key concerns, particularly with individual patient records; there is value in exploring a range of solutions, including but certainly not restricted to Cloud storage systems. Patients are already becoming more knowledgeable about and involved with their own healthcare. Patient empowerment is a key component of the approach we recommend, and we must ensure that their voice is always listened to.

Information technology systems in hospitals will need to be adapted to ensure that data needed for research are captured and included. These systems should be flexible enough to allow the frequent and simple incorporation of new data types. Today, much clinical information is still recorded on paper only, and one of the most important challenges is to facilitate the full transition to electronic health records. This can only be achieved by making future electronic patient data capture systems more user-friendly, with substantially more added value compared to existing paper bound approaches. Moreover, re-use of clinical information requires proper data and text mining together with semantic tools to aid the analysis of retrospective data. This further emphasises the need for complete standardisation wherever possible. Although almost every medical discipline is likely to benefit from the integration of clinical data with more basic research, the emphasis should be on those therapeutic areas where there is greatest un-met medical need, and where early results may be of practical clinical use.
1.2. The importance and impact of this Strategic Report

The European Union has set out a vision and strategy for enabling “personalised, predictive and preventative” healthcare (3P’s) by 2020 (ICT Results, 2010). Exploiting information technology in both clinical medicine and biomedical research is a key component of this strategy. The use of properly validated clinical data in biomedical research is expected to have a dramatic impact on future healthcare outcomes. In addition, a number of interesting papers have been recently published: (Fernald et al., 2011) on challenges for personalized medicine, (Lo and Parham, 2010) on the impact of Web 2.0 and social networks, and (Zheng et al., 2011) on collaborative search in Electronic Health Records. A very important White Paper concerning the re-use of clinical information for research purposes has also been published by the American Medical Informatics Association (2007) in JAMIA, where key issues and recommendations in Biomedical Informatics have been highlighted.

This report aims to raise awareness in the European Union and beyond, about the current problems regarding data, IT security, information governance, and legal issues with regard to clinical data, as well as to suggest solutions. A "digital vision" and agenda are needed within Europe, to cover the next five years and beyond; EU member states are urged to commit significant resources into this effort, and to adopt clear and fully aligned legal positions to allow the optimum re-use of data in research and to support clinical decision-making. This would substantially facilitate the digital revolution in healthcare provision that is urgently required.
2. IT considerations for Medical Informatics

2.1. Data integration and management of clinical systems

It is important to distinguish between data that are used for administrative purposes in hospitals, and the medical and clinical data used for patient care. It is the latter that is of most interest to the biomedical research community, but it is difficult for researchers to access this data for many reasons, including ethics, issues of data ownership, security and confidentiality (discussed at more length in Section 2.2), the heterogeneity of data systems, as well as the currently hotly-debated issue of the choice of cloud-based versus distributed local and/or federated data storage. These matters were discussed in turn.

Different European countries, regions, and even hospitals use different systems for storing patients’ medical records. In many countries, clinical and administrative data are captured in different ways. A single hospital may use a number of different systems, all too many of which are often still paper-based. An EU task force for ICT in health and social care carried out a survey of 900 hospitals in the 27 EU member countries (EC Information Society, 2011), and found that there are many hospitals still without common electronic health record systems, and that many different systems are in place in each hospital. Electronic exchange of clinical data between countries is even rarer, and was only possible in 5% of the hospitals surveyed. Linking data systems between hospitals and primary care is also a challenge; for example, in the UK, primary care records are more often computerised than hospital ones.

Methods of data gathering for electronic patient records must become quick and easy-to-use in a busy clinical setting: using mobile devices for patient data entry on-the-fly (for instance hosted by iPads or iPhones) would be helpful, but the digital input would need to be “as quick as pen and paper” for clinicians to adopt the system readily. Even now, it is still often faster to collect paper-based records than online ones. One reason why take-up of electronic systems is proving to be slow is a perceived lack of demand: the IT industry is not being encouraged enough to search for solutions to outstanding problems such as these. The EHR4CR project (Electronic Health Records for Clinical Research; http://www.ehr4cr.eu/) is a notable exception, coordinated by AstraZeneca, and has been funded through the EU’s Innovative Medicines Initiative to develop a platform and a common set of tools to optimise data sharing with clinical researchers.

Some participants cited examples of systems that were being used to share data within or between hospitals, and forms of good – and not so good – practice that could be learnt from them.

- The Neonatal Network in the UK (http://www.neonatal.org.uk/) is a system for storing data on babies admitted to neonatal units on the Cloud for easier sharing, as seriously ill babies are expected to move between hospitals. Data are obtained in a standardised form. From this experience, the participants concluded that one additional incentive is for the doctors involved to know that the data they are collecting is used for audit and exchange/research purposes.
• The “OpenEyes” system at Moorfields Hospital (UK) ensures that a mobile patient record system is used, similar to that used by the Neonatal Network. This has proved to be successful; it is based on integrating systems already in use rather than obliging clinicians to use new ones.

• On the other hand, the US National Cancer Institute’s “caBIG” (cancer Biomedical Informatics Grid) project is thought to have made a fundamental mistake. After a highly critical review by the National Cancer Institute (NCI) Board of Scientific Advisors (NCI, 2011), caBIG was made to refocus its efforts from the development of software tools to the devising of standards for interoperability and data exchange, and ensuring these standards are adopted into clinical and basic science research (The Cancer Letter, 2011). Moreover, the 12-15 year programme “Connecting for Health”, which started in 2003 and incurred a cost of £12 billion, is considered to have been a very costly failure (The Cancer Letter, 2011), after a damning UK National Audit Office Report on it in May 2011 (Department of Health, 2011; BBC News Health, 2011). This proves that a major IT replacement in hospitals is not always going to be effective and worth going for.

• An international effort of great relevance is then mentioned as being the NIH-funded i2b2 research centre (Informatics for integrating Biology and the Bedside) (Murphy, 2009; Murphy, 2010). The i2b2 project aims at building an informatics infrastructure to support biomedical research by storing and integrating data from clinical practice, and making them accessible in anonymised form to biomedical researchers. To this end, i2b2 has developed a software architecture structured as a “hive” with different software “cells” devoted to data extraction, data manipulation or data analysis tasks. The software has been developed in Java and exploits web services for the communication between the cells. The i2b2 software has been exploited in a variety of clinical and research context, and is currently used in the U.S, Europe and Asia.

There was a consensus that the sharing of all types of data, within hospitals, regionally, nationally and internationally, is already essential, and is becoming more so. Clinical data to be shared will certainly involve images, if it does not already do so. The town of Basel, located between Switzerland, Germany and France, is a pioneer in this area, already using a standard for the exchange of clinical documents.

One barrier to data sharing over and above technical bottlenecks that applies particularly over national boundaries is the perception of clinicians and researchers that “data is power” and that sharing data may dilute its value and make it harder for research to be published. Creating genuine collaborations between research groups that are built on mutual trust can overcome this barrier. The EuResist Network (www.euresist.org) has in the last few years demonstrated the strength of such an approach: hundreds of centres from Italy, Germany, Sweden, Spain, Portugal, Belgium, Luxembourg contribute their data (anonymised clinical and genomic data) to create the largest existing physical database of antiretroviral resistance data: the EIDB (EuResist Integrated DataBase) has now data from more than 50,000 patients and attracts researchers from all over the world to make publishable studies on it. The Swedish system “InfCare” (www.infcare.se) developed by Karolinska Institutet for holding and analysing HIV data, recently extended to hepatitis data, has now been adopted throughout all Scandinavian and Northern European countries. In this context, the Shrine initiative (Shared Health Research Information Network;
The panel had a lively discussion about the nature and importance of standards, without achieving full agreement. All participants agreed that standards for the type and quality of clinical data to be stored were essential, but they disagreed about how rigorous they should be and how and when they should be developed. Some attempts to create formal standards, such as a proposal from IBM to HL7 for standards for HIV data exchange, have not been adopted. On the other hand, the HICDEP standard for HIV data handling (developed by CHIP, the Copenhagen HIV Programme (http://www.chip.dk/HICDEP/tabid/60/Default.aspx) has emerged and is imposing itself, showing how sometimes standards happen to emerge from usage, in a bottom-up process. Sophisticated data standardisation is very time-consuming and much of this work can be wasted. Many delegates expressed their preference to an incremental approach, starting with the most basic standards and ground rules, and building up from there. Different scientific communities have different priorities and needs; in particular, researchers and clinicians will generally have different needs and expectations. There was some support for the idea of building a basic data system for clinicians and customising it later for research use. The importance of quality assurance was also highlighted.

Several publications providing guidance and supporting standards in creating and sharing data records electronically were highlighted. The UK Data Archive has published a booklet, Managing and Sharing Data: Best Practice for Researchers (2009); this, however, is designed for general purposes rather than specifically for use with clinical and biomedical data. Within the EU, an independent, not-for-profit organisation, the EUROREC Institute (EuroRec), has been set up to promote the use of high quality electronic health record systems (EHRs) throughout Europe. EuroRec has published online an inventory of standards that are relevant to EHRs (see http://www.eurorec.org/services/standards/index.cfm); this can be used to ensure that EHRs can be shared across and, perhaps, outside Europe.

The issue of the choice of data storage in “the Cloud” (where the data are held externally on servers that are usually owned by third party, usually commercial, providers and that may be located anywhere, including outside the country where the data are generated) versus institutionally controlled, “federated” distributed and in some cases purely local data storage was raised. This brought up a number of questions, technical, legal and ethical; the issue is articulated at greater length in the next session, “Information assurance”.

http://catalyst.harvard.edu/services/shrine/) is a crucial effort undergoing in the United States in order to find suitable models for sharing data without jeopardising control and ownership rights of the data (Weber, 2009).
2.2. Information assurance

Questions of data ownership, security and privacy underlie the question of information assurance. There was no consensus among the participants about the “ownership” of clinical data, or even about precisely what that term means. Legally, in many countries, a patient might be thought to own all his/her clinical data; but patient data in general can also be said to belong to his/her government, for example, its Ministry of Health. Whoever owns data, patients are often (and should always be) able to control how their data are used. Data ownership can be seen as less important than the decisions that are made about how data are used.

Using the Cloud to store data can be a useful solution, but one that raises problems with data security. With a Cloud-based solution, data are held a long way away from the patients who “own” that data and the clinicians and researchers who will be using it, perhaps in a different country, and neither the owners nor the users of the data may know where it is stored. Many of the technical and security considerations involved in Cloud-based data storage are not specific to the health domain, but all citizens, not only those who are already patients, can be understandably nervous about their personal data being held in this way. In many countries, including the UK, IT security is not perfect, and it is possible for data that are traceable to individuals to end up in the wrong hands.

Two of the current projects funded by the 7th Framework Programme (FP7) under the VPH Initiative are looking explicitly at security issues in the Cloud and at comparing the costs and benefits of Cloud-based and local (federated) data storage solutions. The p-medicine project, coordinated by the University of Saarland, Germany (http://www.p-medicine.eu/) is an Integrated Project (IP) that is developing IT infrastructure, data and models for combining clinical and multi-level data (from whole human and organ level right down to the genetic level) so as to move towards more personalised medicine in the cancer domain. VPH-SHARE, coordinated by the University of Sheffield, UK (http://projects.kmi.open.ac.uk/vph-share/) is another IP that is developing the infrastructure needed to facilitate collaboration and data sharing between several pre-existing VPH projects. Enabling secure data storage is an important consideration in both projects, despite the divergent approaches they are taking (the former leaning towards a federated model, the latter predicated on cloud solutions).

There are clear data protection considerations pertaining to the storage and use of data related to the health of particular individuals. There are a number of technical solutions to this: data can be anonymised or pseudonymised (pseudo-anonymised), while data held on doctors’ laptops can be encrypted. However, some data, particularly genetic data, cannot be fully anonymised (Heenay et al., 2011). Thus, for example, fully anonymised data cannot be used to select patients with particular characteristics for clinical trials.

Patient consent is also a key issue here. Currently, patients need to give consent for each specific research use of their data. This makes retrospective use of data very difficult. It should be possible to move to a system where patients give general consent for their data to be used when they enter the healthcare system. This, however, raises different problems: the more generic a consent form is, the less informative it will be. A cohesive
European approach to this issue is required. There are EU directives in place but it is likely that each country may interpret them differently through their own rules for information governance.

Patients are more likely to give their consent if the system is seen as secure, and if they are aware that their data may be used to improve treatments as well as for “blue skies” research. Several examples of good practice in different countries were presented. Scotland uses an “opt-out” system that in practice has a very low opt-out rate. Denmark has a system where records are available to any medical professional. All data accesses are logged, and patients can find out who looked at their data and when. A “social contract” model was proposed, to be discussed at European level and undoubtedly requiring a long lead time, whereby all citizens (not just patients) would have an obligation to allow access to their data for research, while clinicians and researchers acquire reciprocal obligations to behave ethically and in the best interests of their subjects.

It was noted that, in the age of social networking, people are becoming less concerned about making personal information, including health information, available online. Many people search for disease-specific information online and are not concerned about revealing information about their symptoms that way. The related concept of “Patient Empowerment” was raised. Patients, as the real “data owners”, may be willing to give out very personal information, even non-anonymised, if they think that they may be more likely to benefit from innovative treatments that way.

Patients Like Me (http://www.patientslikeme.com/) is a US-based social networking and data sharing platform for people with a range of mainly chronic and serious conditions. Patients are encouraged to share information about their disease and its treatment, questions and concerns with a network of people suffering from the same condition, and this data are used, transparently, in clinical trials and other research. Initially, all information was openly available to all, but more recently a range of privacy settings have been included so patients may choose how much data to make publicly available. The pharmaceutical company GSK has been using this data and data stored by the personal genomics company 23 and Me (https://www.23andme.com/) which stores and analyses the genotypes of thousands of individuals at over 500,000 different positions in genetic research. They have noticed that most participants in these studies are not particularly concerned about whether their data are shared.

However, although many people are willing to be open about the use of their data, it must be borne in mind that there is probably a significant number who are not. Some will undoubtedly change their minds over time. Any system that is introduced will need to offer the facility for patients to opt in and out, and to withdraw their data from research studies in progress. Such systems are already in use in Spain, namely the “Personal Health Folder” (http://www.gencat.cat/especial/esalut/eng/carpeta.htm) and Canada (http://www.bmj.com/content/332/7556/suppl/DC1), and in the Sintero project in Cardiff, Wales (http://www.healthinformatics.cardiff.ac.uk/research/current-projects/180-sintero-class-server-technology) where medical data are standardised but continually updated.
2.3. Summary

- Europe must put forward a clear vision for the digitisation and standardisation of clinical data and medical records, with an agenda to cover at least the next five years. Both the member states and the EU as a whole will have roles in this.
- There must be clarification about what is happening nationally and differences across the EU, particularly concerning interpretations of the regulations.
- Interoperability of IT systems should be promoted to allow data access and integration, including across national borders.
- The quality of, and standards for data capture and its re-use for research must be optimised, especially quality standards for data entry. It should, however, be borne in mind that not everything can be standardised and reasonable compromises should be made.
- Some elements in the system should be made mandatory to allow better quality of data entry and/or capture.
- Clinicians are very busy, and often interrupted during routine work, which can lead to the generation of incomplete data sets. It is important to consider incentives to improve the data entry (e.g. automatic generation of discharge reports, laboratory findings etc...).
- At the same time, clinicians will need to be assured of quality assurance and control to use such systems of data entry, particularly as data entry will often need to be done during difficult and stressful situations, and the moment of data entry can never be fully controlled. The systems must also be easy to learn and use.
- All clinicians, not only those with specific interest and expertise in information technology and/or clinical research, should be involved in the development of systems for electronic data entry.
- The role of the patient is also vital. We need to ask their opinions and listen to them, and to encourage them to consent to data sharing. Websites such as “Patients like Me” can help to create patients who are empowered and confident in their knowledge of their disease and its treatment. Ultimately, patients should benefit from involvement in research as findings could have an effect on their recovery.
- Questions of consent need to be explored more fully. “Blanket” consent may be desirable but may not be possible, and opt-outs should always be available.
- Projects such as EHR4CR led by AstraZeneca should lead to important advances in capturing clinical data that take these factors into account, but this will necessarily be a long-term enterprise. This is particularly important if, as is desirable, the proposed solutions are to be world-wide and not just implemented within the EU.
3. Exploiting Medical Data in Research

3.1. Research benefits from clinical data availability / Healthcare information incorporation into “omics” and molecular level research

There was general agreement on the basic value of combining clinical data with biomedical research data, and of incorporating clinical data into medical research; there were more questions over the issue of how this can best be done. Building connections between databases of clinical trials and medical information databases is likely to be useful. Currently, it is easier to obtain data on individual patients in advance, for prospective clinical trials, than it is to access similar clinical data retrospectively. Data should be collected systematically, with data from different formats and sources integrated, and it should be of consistently high quality. Several recent publications have discussed uses of electronic health records in genetic research and drug discovery, including genotype-phenotype studies and investigations of drug re-purposing and off-label use (Yao et al., 2011; Kohane, 2011).

Data collected retrospectively can be useful, but it has important limitations. It may be necessary to go back and test old data against newer models, technical improvements, and/or improvements in diagnosis and treatment since the data were collected. This ought to be taken into account.

The pharmaceutical industry would like clinical data to be used not just to test medical hypotheses, but to make use of them as key input into the design of clinical trials.

3.2. Clinical areas to be targeted by research funding

There was a general consensus that clinical data will benefit biomedical research very widely; indeed, the question posed to kick off this part of the discussion, “Which research areas would benefit most?” was changed to “Which areas of research will not benefit from the use of clinical data?”, with the answer that such areas would be only very limited in number.

There was, however, also a consensus that it is necessary to rank disease areas by priority. There are a number of criteria that may possibly be used for this, including disease prevalence, costs of the disease to the individual patient and to society, and the likelihood that the research will lead to clear clinical benefit. There is currently perhaps too great a focus on cancer research as compared to, for example, research in hypertension and diabetes (both extremely common diseases). It was suggested that researchers should focus on diseases with high cost and/or high levels of morbidity, and those where the gaps in current knowledge are the most obvious. Chronic heart failure was cited as an example of a disease where relatively small technical advances in its detection, diagnosis and treatment might lead to significant improvements in outcomes, and allergy management as an area where more research funding is needed. Another cited example was Hepatitis C, among the most diffused infectious diseases, for which the first selective treatments are becoming available during 2011 and 2012, thus
turning it into a field where research on clinical (and related genetic) data will have a huge impact in terms of clinical effects.

There was a substantial discussion of the related area of education for disease prevention. This may reduce, but can never remove, the need for treatment of acute and chronic diseases. An optimistic vision for the long term is for better knowledge of the causes of all diseases, whether genetic, environmental and/or lifestyle, to feed into setting the priorities for, and designing, clinical and biomedical research studies.

As it is far more costly for patients to be cared for in a hospital than at home, there is a great deal of interest in extending the time for which patients, particularly those with chronic diseases, can be treated at home and in the community. The use of clinical data and IT systems is already important in this and will become more so, for example in computerised, online decision-support systems for general healthcare purposes. Research data and information from research studies clearly have a key role in improving diagnosis and selecting optimum treatment regimens. As all humans age, and ageing is the main risk factor for many chronic conditions, discussion of the collection and use of biomedical data is relevant to all citizens, not only those with specific known conditions and needs.

It was noted that in the UK, the UK Clinical Research Collaboration (UKCRC; http://www.ukcrc.org/) offer useful guidelines for clinical research studies. Similarly the National Institute for Health and Clinical Excellence (NICE; http://www.nice.org.uk) helps guide clinical practice and service provision following critical appraisal of the clinical effectiveness and health economics of therapeutics and/or interventions, often very important in UK commissioning of health services. Although NICE is more engaged in disseminating knowledge of best practice and/or acceptable treatments by disease category (see http://www.nice.org.uk/CG36 for Atrial Fibrillation, and (NICE, 2007, pp.154) for Atopic eczema in children), rather than promoting research or providing research guidelines per se, it does definitely promote patient access to clinical trials, particularly where there is a need for a better evidence base. Similar guidelines are likely to be available in many other countries.

3.3. Personalising electronic health records / Value in individual access to personal healthcare data

Issues raised around personal healthcare records and individual access to health data are directly concerned with both the development of personalised medicine and patient empowerment. Patients are already using simple devices to monitor their own conditions and, to some extent, take control of the management of their diseases. Monitors for blood pressure and blood glucose levels, for example, are available at relatively low cost over the counter from, for example, stores such as Boots in the UK. These are likely to be used by the “worried well” and those who know they are at risk of developing relevant conditions as well as by current patients.

Citizens as well as patients are also becoming more knowledgeable about their own and their families’ medical conditions, with the phenomenon of the Internet-savvy “expert patient”. This is particularly noticeable in the care of patients suffering from rare
conditions. This trend is likely to become even more prevalent in the future, and it may be possible to involve patients more in delivering complex treatments such as chemotherapy for cancer. Consumer-led healthcare would itself generate more information and data for biomedical research.

With healthcare moving from the hospital to the community, and with the patients playing an increasingly important role in managing their own healthcare, the role of the doctor is diminishing relative to that of other health and social care professionals (e.g. social workers). These are more numerous than doctors and generally less well qualified. IT-based tools can help move healthcare decision-making from the doctor to the patient in the community (with the support of these less “elite” professionals), but only if adequate training is given to these health and social care professionals. Doctors would also need to be more IT “savvy” to continue to be part of this transition, and health education need to be more available to the general member of the public.

Manifestly, it is also essential that computer-based tools developed for use in the community are reliable and trustworthy. Research results that support the adoption of new tools and new treatments must be evaluated thoroughly before they can be adopted, and it should always be recognised that experimental treatments tested in clinical trials often give worse outcomes than the established ones. The risks involved in developing IT systems for healthcare are different from and can be greater than those involved in developing such systems for other applications.

Given these caveats, however, there is great potential for using computer-based tools to support clinical decision-making in the community, which will allow already knowledgeable patients to play a greater part in designing their own healthcare. This can only increase as the technology becomes even more mobile and miniaturised, with hand-held devices offering great potential. IBM is developing its “Watson” supercomputer, currently best known for its prowess in the word game Jeopardy, as a clinical decision-making tool. Although this is still some years away from having any real world health impact, the company has already had some success with an application to HIV/AIDS (IBM, 2009). In this environment, and with the concept of patient empowerment in the foreground, it is at least desirable and may even be necessary for patients to have access to their complete electronic health records: what may be thought of, in shorthand, as a “Digital Me”; or the “Digital Patient”, a newly funded EU VPH FP7 project.

### 3.4. Desirable evolution of Biomedical Informatics in relation to re-use of clinical information

Biomedical informatics resides at the interface between the academic discipline of bioinformatics and informatics for health. These two disciplines are continuously moving closer together, particularly as we are moving into an era where personal genome sequencing is becoming low-cost and therefore widespread. The challenges of collecting, storing and analysing this torrent of individualised genetic data, and putting it to practical use in the clinic are being discussed in the bioinformatics literature (e.g. Fernald et al., 2011), and one topic under discussion is how electronic health records can best assist with this (Busis, 2010).
There was general agreement that if data are to be useful for both the clinical and the biomedical research communities, it should be collected and collated in a fully structured manner with its associated meta-data. It would be useful if the data and meta-data could be made accessible by semantic web technologies without the need for further technological developments. Clinicians will only take up tools if they are straightforward and easy-to-use with accessible, most likely web based, interfaces.

There was a short discussion of one specific example: the use of research data to personalise drug dosage, rather than using standard doses for all patients with a given condition. The optimum dose of a drug for a particular patient depends on the drug’s pharmacokinetics and the patient’s genetic make-up, as that affects how strong the drug will interact with its biological target(s), as well as how it will be metabolised. Data on the genotypes of individual patients at relevant positions will clearly be useful here. Examples were provided by clinicians from their own fields to illustrate this.
3.5. Summary

- The issues around the re-use of clinical data for research purposes are complex; finding optimum solutions will require a multi-disciplinary approach, including IT and data security specialists, clinicians, biomedical researchers and bioethics experts among others.
- The roles of patients and their carers are crucial. These people must be listened to and empowered. It should be possible for patients to become more actively involved in clinical trials related to their conditions.
- Industry – not only the pharmaceutical and biotech industries, but also the IT industry – is an important player. IT system suppliers for hospitals must be informed about research needs, and IT system designers should be told what types of information will need to be captured by the systems developed.
- It is important to raise awareness in the EU about the current problems regarding IT security, information governance, and legal issues with particular regard to clinical data.
- Developing and maintaining the highest possible standards for data collection is essential.
- Existing IT systems in hospitals should be developed to include data needed in research, and these should be flexible enough to allow the frequent and simple incorporation of new data types.
- The analysis of retrospective clinical data presents particular challenges: not least concerning how the quality of data collected even years ago can be assured. This may require the development of further text mining / semantics tools.
- Probabilistic models can usefully be incorporated into IT systems and these should be upgraded continuously.
- Although almost all therapeutic areas may benefit from the use of clinical data in research, this research should be focused, wherever possible, on those areas where results are likely to prove of practical clinical use.
- Preventive medicine is an important priority and is becoming increasingly more so. Clinical research can certainly be useful in this field, although it is less immediately obvious where the data will come from although certainly whole exome/genome analyses will be very valuable here.
4. Conclusions and Recommendations

- The EU should invest significant time and resources in establishing a clear, Europe-wide vision for the digitisation and standardisation of clinical data and medical records over the next five years.
- Provided that standardisation is seen as a priority, it may be best for standards to be allowed to evolve pragmatically rather than trying to impose too many in advance.
- These are complex issues and finding optimum solutions will require a multi-disciplinary approach, including IT and data security specialists, clinicians, industrial and public sector biomedical researchers and bioethics experts etc.
- The systems proposed should operate as far as possible across regional and country boundaries, and between hospital and community care settings.
- The role of the clinician is crucial and data entry systems, in particular, will need to be validated, to be seen as reliable and to have user-friendly interfaces to ensure optimum take-up by the clinical community.
- Health data raises issues of security, privacy and ethics beyond those of other types of data. These issues should be of the first importance in systems design; however, we cannot yet (and may never) propose either purely Cloud based or purely federated solutions.
- Patients, and their carers, should be at the forefront of all discussions and proposed solutions, and their views should always be taken into account. The possibilities for “empowered” patients to play a greater role in planning their treatment, and to take part in research and clinical trials, are already significant and will increase as the Internet becomes ubiquitous.
- Patients can be encouraged to share their data, but this must never be made compulsory, and it should always be possible for them to opt out of and/or withdraw their data from studies.
- Clinicians and healthcare centres should be on their hand encouraged to share their completely anonymised or permitted to share data.
- Systems, including data entry systems, should be flexible enough to allow the ready incorporation of new types of data.
- Analysis of retrospective clinical data is likely to be of great value, but this poses particular challenges, especially in the area of quality assurance.
- Almost all disciplines are likely to benefit from the use of clinical data in research, but the emphasis should be on those where there is greatest unmet medical need and those where results are likely to prove of practical benefit in the clinic. This includes preventative medicine, which may require a rather different approach to data collection.
- All systems should eventually be able to be used worldwide, rather than just within the EU.
REFERENCES

- BBC News Health (2011) 7bn pounds NHS electronic records ‘achieving little’ for patients, by Nick Triggle
  
  [http://www.bbc.co.uk/news/health-13430375 (17/07/2011)]


  


  

  


- The Cancer Letter (2011) 37, 1-6
  
  http://ontology.buffalo.edu/smith/articles/Cancer_Letter_Editorial.pdf
  (17/07/2011)


  (17/07/2011)

- UK Data Archive, (2009). Managing and Sharing Data: Best Practice for Researchers


APPENDIX

**Venue:** Executive Suite Lecture Theatre, Roberts Building, University College London, London, UK

<table>
<thead>
<tr>
<th>Participants</th>
<th>Institutional Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof. Peter Coveney – co-chair</td>
<td>Centre for Computational Science, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Dr. Norbert Graf – co-chair</td>
<td>University of Saarland, Germany</td>
</tr>
<tr>
<td>Dr. Nour Shublaq – co-chair</td>
<td>Centre for Computational Science, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Dipak Kalra</td>
<td>Centre for Health Informatics &amp; Multi-Professional Education, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Kathy Pritchard Jones</td>
<td>UCL Partners, Programme Director Cancer Institute of Child Health, University College London, U.K.</td>
</tr>
<tr>
<td>Dr. Elizabeth Murray</td>
<td>Academic Primary Care, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Patrick Maxwell</td>
<td>Division of Medicine, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Richard Begent</td>
<td>Cancer Institute, University College London, U.K.</td>
</tr>
<tr>
<td>Mr. Tony Lezard</td>
<td>UCL Partners, U.K.</td>
</tr>
<tr>
<td>Mr. Mike Andersson</td>
<td>Moorfields Eye Hospital, U.K.</td>
</tr>
<tr>
<td>Mr. Joel Bacquet</td>
<td>European Commission – ICT for Health Unit DG Information Society, Belgium</td>
</tr>
<tr>
<td>Dr. Robert Davey</td>
<td>Toshiba, Medical Visualisation Systems Europe</td>
</tr>
<tr>
<td>Dr. Stephen Mayhew</td>
<td>GlaxoSmithKline, R&amp;D Strategy</td>
</tr>
<tr>
<td>Mr. Tom Wilton</td>
<td>GlaxoSmithKline, U.S. Pharmaceuticals</td>
</tr>
<tr>
<td>Mr. John Crawford</td>
<td>IBM Europe, Healthcare Industry</td>
</tr>
<tr>
<td>Prof. Riccardo Bellazzi</td>
<td>University of Pavia, Italy</td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dr. Francesca Incardona</td>
<td>EuResist Network GEIE</td>
</tr>
<tr>
<td>Dr. Miguel Ángel Mayer</td>
<td>Universitat Pompeu Fabra, Spain</td>
</tr>
<tr>
<td>Mrs. Sandra Pla</td>
<td>Fundació Institut Mar d'Investigacions Mèdiques, Spain</td>
</tr>
<tr>
<td>Ms. Maria Saarela</td>
<td>Fundació Institut Mar d'Investigacions Mèdiques, Spain</td>
</tr>
<tr>
<td>Dr. Juan Miguel Gargia-Gomez</td>
<td>Universidad Politécnica de Valencia, Spain</td>
</tr>
<tr>
<td>Dr. Ing. Andrea Remuzzi</td>
<td>Mario Negri Institute, Italy</td>
</tr>
<tr>
<td>Prof. David Hawkes</td>
<td>Centre for Medical Image Computing, University College London, U.K.</td>
</tr>
<tr>
<td>Dr. John Hipwell</td>
<td>Centre for Medical Image Computing, University College London, U.K.</td>
</tr>
<tr>
<td>Prof. Luigi Martinelli</td>
<td>Niguarda Hospital, Italy</td>
</tr>
<tr>
<td>Dr. John Apps</td>
<td>Institute of Child Health, University College London, U.K.</td>
</tr>
<tr>
<td>Dr. Lianghao Han</td>
<td>Centre for Medical Image Computing, University College London, U.K.</td>
</tr>
</tbody>
</table>