DEBUGIT
Detecting and Eliminating Bacteria UsinG Information Technology
Large-scale integrating project (FP7, 2008 - 2011)
Seventh Framework Programme

First yearly report "M12"

DebugIT WP11 Ethical & legal issues report

FP7-ICT-2007-1
DebugIT
Grant agreement no 217139

D11.2 Ethical & legal issues report (M12)

WP11-lead	José Verguts, jose.verguts@agfa.com
Contributing WP-partners	UCL: Dipak Kalra, Peter Singleton

Document History

Version	*Date*	*Description*	*Author*
0.1	2009-02-01	First version including an overview of general issues and a template to be filled-in for site/WP-specific details	Dipak Kalra Peter Singleton
0.1.1	2008-02-01	Transformed into DebugIT template for deliverables	José Verguts

[Distribution List]

Version	*Date*	*To*
0.1	2009-02-01	Agfa, EMP
0.1.1	2009-02-01	DebugIT Community

[Optional Approval List between partners within a WP]

Version	*Date*	*By*	*Comment*

Table of contents

Guideline

There is a form on the last page which needs to be completed by each site. Best is to copy the table to a new document and complete it there. This is a first *brief* survey of what each site is doing with personal (or anonymised) data, so we need to find a balance between brevity and exhaustive detail.

Apart from the contact details and ethics approvals sections, there are four main sections:
1) about data you will hold (possibly sourced from elsewhere);
2) data you will receive (as a partner in DEBUGIT, possibly from your home institution)
3) data you will share (which you may provide onto other partners or after processing with the outside world, including your home institution)
4) data you will use or analyse (so received fleetingly as part of queries)

These do overlap to some degree. Feel free to repeat some of these sections, say, if you receive different data from more than one source - especially if that makes it easier for you to complete (as it will also make it easier for the rest of us to understand).

The idea is that the completed form would only run to 2 or 3 pages once completed - by all means, attach some other documents that explain particular points in more detail.

Please copy in Peter Singleton (peter.singleton@chi-group.com) , Dipak Kalra (d.kalra@chime.ucl.ac.uk) and José Verguts, jose.verguts@agfa.com to your responses. If you have any queries then email Peter for more guidance

1. Introduction

This product is part of the Work Package 11 – Project Management, produced by the task, T11.4 Ethics, data protection and other legal issues.
The description for Task T11.4 is[1]:

> *The work to be performed in the project will be faced with a variety of ethical and legal issues, in particular in respect of consent procedures, personal data security, confidentiality and privacy. 'Personal data' (relating to individuals) will only be held at the clinical institutions involved in the treatment of the patients concerned; only anonymised data will be held at the consortium-level and shared between consortium members. There will be a case reference generated by the 'treating institution' specifically for this purpose so that processed information can be passed back to the 'treating institution' and, where pertinent, re-combined with the patient record to inform treatment, otherwise the consortium will have no knowledge of the individuals, only about episodes of care.*
> *This work task is led by an appropriately qualified expert able to provide highly competent and independent counsel on these issues to the whole of the project team. Where partners will provide access to patient data to be used for our research, submissions are to be prepared for the respective ethics committee responsible at each of the sites. The ethics and data protection advisor will be responsible for coordinating ethics committee submissions.*

This report, representing deliverable, D11.2, is a review of the ethical and legal issues surrounding this area of research, this project in particular, and the various activities within the project by partner site.

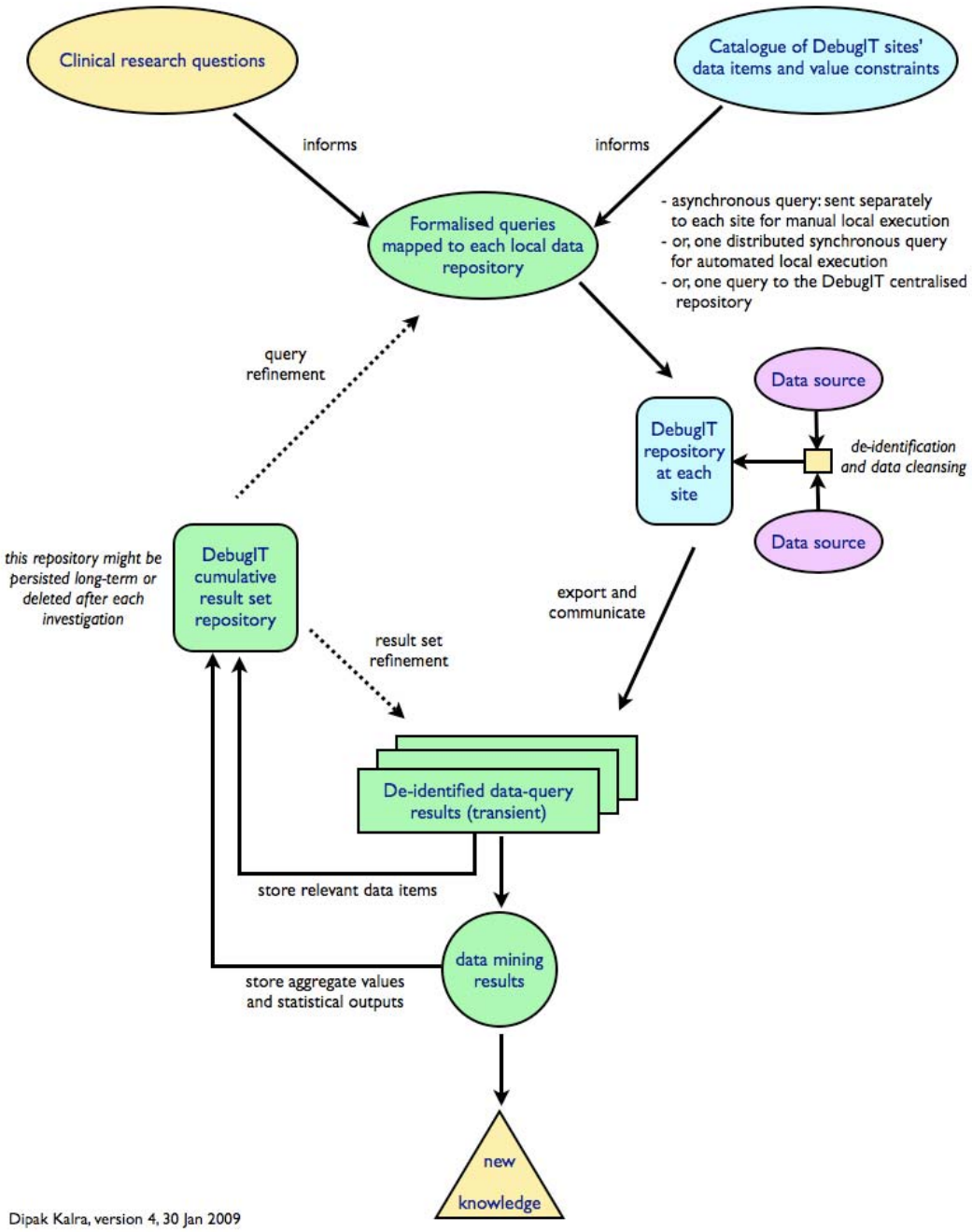
2. Summary of Project

The DebugIT project will use clinical and operational information from Clinical Information Systems (CIS) at several pilot sites across the European Union through the view of a virtualized, fully integrated Clinical Data Repository. It will provide transparent access to de-identified data from the original CIS together with data aggregations that are held in local stores . Highly advanced new text, image and structured data mining on individual patients, as well as on populations, will render valuable information and temporal patterns of patient harm from bacterial sources.
The knowledge gleaned will be fed into a Medical Knowledge Repository and combined with domain information coming from external sources (guidelines and scientific evidence). After validation, this knowledge will be used by a decision-support and monitoring tools in the clinical environment to prevent harmful patient safety issues and report on them.
Outcomes and benefits of clinical and socio-economic terms will be measured. The results will be integrated into CIS of participating European hospitals, industry and their clients and become available globally. This will happen through a European or world-wide Disease Control Centre/Public Authority. It will be an Open Source solution.
Advanced ICT applications and innovations address the virtualization of the Clinical Data Repository. This will happen through ontology and terminology binding and mediation, advanced data mining techniques, the use of machine reasoning related to real, point-of-care patient data, as well as consolidation of all these techniques in a comprehensive but open framework. Output will be applicable to other clinical fields.
For more detail on the project refer to: www.debugit.eu/news/documents/DebugIT_Brochure_Web.pdf

[1] From FP7-ICT-2007-1 ECGA Annex I: WP11_Description-1.pdf

2.1 Data use within the project

The following diagram illustrates the intended use of data within the project at a high-level:



Note: it is not the intention of DebugIT to store centrally any identified health records, or to store complete de-identified health records. Queries for specific clinical data items will be executed at each clinical site, and only de-identified results sets will be pooled in order to perform the data mining. Any data items that have not contributed to the knowledge pool will be purged.

There are three main data flows: clinical data-source to DebugIT partner, partner to partner, and project to external organisation (possibly including original data source).

Clinical data-source to partner: These arrangements will vary across the partner sites depending on local requirements, facilities, and resources. Ideally data-sets provided should be de-identified at the source before transmission to the partner, but in some cases it may be necessary to transfer identifiable data to the partner under a 'safe harbour' arrangement where the data will then be de-identified for use within the project. Even where data is transferred in de-identified form the partner site may apply further data-cleansing or de-identification routines to meet internal project standards.

Each such data-sharing is likely to require individual ethical approval from the data-source ethics research committee, though there may be local arrangements for multi-centre approval, which will limit the amount of paperwork and effort.

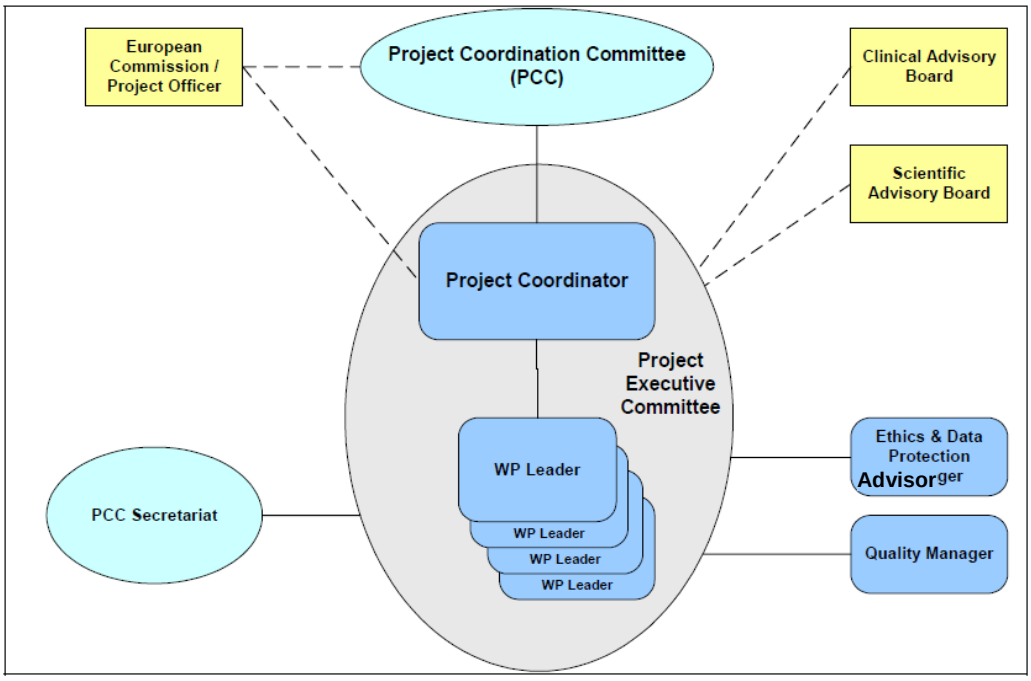
Partner-to-partner: Once the data is standardised and stored in the local DebugIT repository, then it will be available to be queried by other partners. The output data-sets will be either aggregate data results or individual-level data-set extractions in anonymised form.

Conceptual results of queries and further analysis will be recorded within the knowledge-base – this will consist of clinical facts rather than data about individuals or individual clinical events

Project to external organisation: These will consist solely of clinical facts from the knowledge base, though possibly with the rider that these facts are immediately relevant to the institution, viz. that there are cases under treatment to whom these may apply (e.g. where some correlation shows that the current treatment plan for one or more patients should be revised to either avoid possible adverse results or to simply improve likely outcomes).

There will be no indication or reference to specific cases and no attempt to re-identify the cases – that will remain the responsibility of the originating organisation. Indeed, the clinical facts may apply more widely, including cases where results were not shared with the DEBUGIT project.

2.2 Project Control structures



2.2.1 Project Coordination Committee

This committee is responsible for the overall coordination of the project, including seeking and gaining ethics approvals as necessary and meeting legal and ethical requirements. It is supported in day-to-day matters by the Project Coordinator and Work Package Leaders (forming the Project Executive Committee), and in its direction by the two Advisory Boards, who may also provide opinion on some ethical and legal aspects, based on experience at other projects and wider research norms.

2.2.2 Role of Ethics & Data Protection Advisor

The role of Ethics & Data Protection Advisor is to investigate the ethical and legal environment in all participating Member States as well as at EU level, follow the project activities, raise potential ethical and data protection issues related to the use of real patient data, and suggest appropriate measures for addressing such issues. The Ethics & Data Protection Advisor can take part in Project Executive Committee meetings on demand, and works in close cooperation with Advisory Board members.
The Ethics & Data Protection Advisor reports directly to the Project Executive Committee and is also available to advise the Work Package Leader on issues within their own Work Package(s).

3. Issues around this area of research

The use of patient data for 'medical research' raises a number of legal and ethical issues. Some of this arises through the lack of clarity around whether medical research (or if drawn more widely, healthcare research) is a 'secondary' and distinct purpose from the primary purpose of gathering patient information in order to fulfil the immediate clinical needs of a patient with a given condition or conditions[2]. This issue is important to a project like DebugIT which cannot afford to obtain explicit consent from each patient included in the data mining analyses being performed.

3.1.1 Purposes and uses

Clearly the main reason for recording a patient's health information is as a record of fact of what has been done in medical terms when treating a patient, both to help direct future care for the patient and also for potential legal defence by practitioners. Healthcare organisations and regulators variously require clinicians to 'keep good records'.

As an example, the UK General Medical Council (GMC-UK) has stated[3] the need for doctors to keep 'clear, accurate and contemporaneous patient records which report the relevant clinical findings, the decisions made, the information given to patients and any drugs or other treatment prescribed'. It also requires keeping colleagues well informed when sharing the care of patients, and taking part in regular and systematic medical and clinical audit.

In terms of distinctions between primary and secondary uses, the Royal College of Physicians in London state[4]:

> Medical records serve many functions in the modern healthcare environment. These can be broadly divided into primary and secondary functions (table 1).

Primary functions	Supporting direct patient care • Aide memoire • Support clinical decision making • Communication
Secondary functions	Medico-legal record Source of information for: • Clinical audit and research • Resource allocation • Epidemiology • Service planning

Table 1 – Primary and secondary functions of medical records

For many, clinical audit (although classified above as a secondary function) is definitely an intrinsic part of healthcare as it should be performed as a matter of course by the same clinicians and clinical teams that provide treatment to patients, so would not consider that as 'secondary'. For others, research is without doubt wholly separate from the primary function of healthcare, no matter who performs it.

In legal terms, the matter is really what the public has been told as part of the fair processing information they are given. Unfortunately, this is rarely clear – hence the suggestion of the Article 29 Data Protection Working Party that use of EHRs should require statutory support to define what should be expected. However, this has little practical meaning for researchers until member states choose to create such statutory support.

3.1.2 Legal questions

The main European-wide legal aspects are:

[2] See *The added value of electronic health records (EHRs): Linking patient care, clinical research and public health* available at: www.eurorec.org/files/filesPublic/High_Level_Statement_EHR%20in%20EU__FINAL%20_July%2014.pdf

[3] in 'Good Medical Practice': www.gmc-uk.org/guidance/good_medical_practice/GMC_GMP.pdf

[4] DRAFT RCP record keeping standards – inpatients - http://hiu.rcplondon.ac.uk/clinicalstandards/recordsstandards/draft_std_5-0_standards.pdf

- EU Convention for the Protection of Human Rights and Fundamental Freedoms[5]

  Article 8 of the European Convention of Human Rights:
  1 Everyone has the right to respect for his private and family life, his home and his correspondence.
  2 There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

- EU Data Protection Directive (95/46/EC)[6]

  Specifically concerning healthcare: "Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and **the processing of data concerning health or sex life**." Article 8 (1)
  Exceptions:
  - Consent of data subject
  - Processing of (medical) data by health professionals
  - Substantial public interest exemptions

- Directive 2002/58/EC of the European Parliament and the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector ("e-Privacy Directive").

  The e-Privacy Directive elaborates on Directive 95/46/EC, in areas such as confidentiality, billing and traffic data, rules on spam, etc. The objective of this Directive is to allow the free circulation of legally processed personal data within the Member States of the EU. It also sets the foundations of confidentiality as fundamental principles applicable to all forms of electronic communications. Consequently, any shape of interception or storage of the communications deprived without the preliminary approval of the users (e.g. opt-in system) or not based on other specific grounds in the law has to be forbidden.
  However, Directive provides that Member States may adopt legislative measures to restrict the protection of personal data, e.g. on the sensitive issue of data retention, in order to allow criminal investigations or to safeguard national security, defence, and public security. Such action may be taken only where it constitutes a "necessary, appropriate and proportionate measure within a democratic society".

- EU Charter of Fundamental Rights

  While this Charter has been agreed and adopted by the EU Parliament, the Council and the Commission – it is not actually legally binding as yet, though may yet become so. In one sense, it merely encapsulates the rights already laid out in existing legislation. However, the wording is at some points noticeably different:
  - Article 7 (Respect for private and family life);

Everyone has the right to respect for his or her private and family life, home and communications.

[5] www.echr.coe.int/NR/rdonlyres/D5CC24A7-DC13-4318-B457-5C9014916D7A/0/EnglishAnglais.pdf
[6] http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

- Article 8 (Protection of personal data)

> Protection of personal data
> 1. Everyone has the right to the protection of personal data concerning him or her.
> 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis lay down by law. Everyone has the right of access to data, which has been collected concerning him or her, and the right to have it rectified.
> 3. Compliance with these rules shall be subject to control by an independent authority.

3.1.2.1 Specific supporting guidance

- Article 29 papers on Electronic Health Records
  - WP131[7]: Recommends specific legislation in member states to regulate other purposes of EHR in member states, and
  - WP136[8]: definition of 'Personal Data'

  Both documents are "working papers" which means possible revision in the future and are seen as guidelines for EU member states in developing specific legal responses to the use of EHRs

3.1.2.2 Vital interests

In some cases, such as the UK Data Protection Act 1998[9], the term 'vital interests' has been interpreted quite broadly as 'medical purposes'. The recent Article 29 paper, WP131, clarified the Working Party's interpretation of this phrase to be very strict: a matter of life and death where there is no opportunity to seek consent.
This shows one of the core conflicts at issue here between personal autonomy and the wider public interests (safe healthcare and efficient delivery of healthcare, especially where healthcare is funded through tax or state insurance).
It is clearly in the 'vital interests' of a patient that treatment is provided as safely as possible, but if clinical audit and research are not considered part of normal healthcare, then seeking consent for such activities will necessarily increase the organisational burden of delivering either healthcare or the audit or research. This would mean that either healthcare costs increase, healthcare is restricted, or that less audit and research are done. Either of the last two affects the 'vital interests' of the patient.
Increasing healthcare costs in a state-funded healthcare system will mean an increase in taxes which will have other implications for the citizen. In a non-state-funded system, such as the USA, this usually means that either poor citizens cannot afford healthcare or insurance premiums go up, which again usually means that more poor citizens cannot afford healthcare insurance.
This may also show some of the differences between a strict 'legal' interpretation of the Directive and a more nuanced 'ethical' interpretation, where the spirit and impact of the Directive is considered as well as the terminology used.

3.1.2.3 Differing legal codes

There are a number of variations across member states that affect the actual impact of these legal requirements on citizens:

- Criminal and anti-terrorism laws to permit or require data-sharing for specific purposes
- Variable interpretation and enactment of the DP Directive
- Local exemptions from DP Directive

[7] http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp131_en.pdf
[8] http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf
[9] www.opsi.gov.uk/Acts/Acts1998/ukpga_19980029_en_1

- Other statutes or legal precedence, such UK common law of confidentiality, based on case law rather than explicit statute

3.1.3 Ethical Issues

Dr William Lowrance has covered many of the issues around secondary uses of medical data in his Nuffield Trust paper[10], where he notes in particular the very different scenarios between clinical trials where recruitment and consent is the norm and epidemiology where the *post hoc* analysis of actual events and experience makes formal consent difficult for practical and theoretical reasons (as the purpose of the research is not usually known when the data is collected).
Dame Onora O'Neill in 'Rethinking Informed Consent'[11] covers many issues around patient consent and re-use of medical records, and argues in 'Autonomy and Trust in Bioethics' that there may be some wider social obligation to permit re-use of medical records for the wider public benefits and the benefit of future generations. With the ongoing battle against new strains of disease, one might argue that this is particularly apposite.
There is a basic conflict between the 'right to privacy' and the wider social need for effective and safe healthcare, though there may be mechanisms through which the conflict of interests can be mitigated.

3.1.3.1 Security and approvals

One of the key ways that the risk to individuals can be reduced is simply by ensuring that their data is only used and shared safely and securely. People are more likely to be comfortable with research use of data if they feel their data is safe. Fortunately, there is some evidence[12] that researchers are generally trusted by the public and that the public is positive about medical research[13], though this may be through ignorance of what they do or failure to understand their motivations[14].
Having approvals processes around research access to data ensures that there are likely to be gains from the research to counterbalance the (hopefully, remote) risks to the individuals involved. The difficulty is, of course, in assessing the benefits in advance of the research, where it may be the unanticipated insights which give the greatest benefits. Too strong an emphasis on assured outcomes may actually limit the possible outcomes by eliminating more adventurous research.
One key point which is unresolved is the retention of research data at the end of a research project. This needs to cover any original identifiable data, and derived data at the individual level (which may have been linked with other data-sets, so is richer in content and so possibly riskier too), as well as summary data-sets, which should have a very low risk of privacy breach. If data is automatically destroyed at the end of a project, then there may be no opportunity to build on the investment in time and effort in previous research projects. Archiving of the data may be a sensible approach, but only if it can be re-used. This presents a number of practical and organisational problems, which are outside the scope of the paper, though the core question of what to do with the data at the end of a project remains.

3.1.3.2 Opt-in vs. Opt-out

There is still great debate about opt-in versus opt-out approaches to consent: is opt-in the only valid basis for consent? Certainly, it is accepted that where treatment is involved (for ordinary care or clinical trials) that

[10] Lowrance W (2002) Learning from Experience: Privacy and the Secondary Use of Data in Health Research, Nuffield Trust, www.nuffieldtrust.org.uk/members/download.aspx?f=%2fecomm%2ffiles%2f161202learning.pdf&a=skip
[11] Manson NS and O'Neill O (2007) Rethinking Informed Consent in Bioethics, Cambridge University Press
[12] Various Ipsos Mori polls: www.mori.com/polls/2003/scientificalliance-top.shtml and www.ipsos-mori.com/polls/2006/rcp.shtml
[13] Ipsos Mori (2007) The Use of Personal Health Information in Medical Research, Medical Research Council, www.mrc.ac.uk/consumption/idcplg?IdcService=GET_FILE&dID=10983&dDocName=MRC003810&allowInterrupt=1
[14] Armstrong et al (2007) Public Perspectives on the Governance of Biomedical Research, Wellcome Trust, www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtx038443.pdf

patients must express their consent to treatment, though this need not be formal written consent, except where the physical risk is high or irreversible (e.g. surgery).
Where the risk is low or the treatment minimally invasive, then the consent process may be more passive: 'Let me take a blood sample from you" and if there is no objection or there is an accepting action (e.g. rolling up a sleeve) then consent is inferred. The consent process is tuned to the clinical need and the risks involved. This clinical practice has built up over the years and allows treatment to be provided efficiently for both clinicians and patients.
In the past, this form of consent had been taken to include research use of medical records as well, though then most of the research would have been done by the clinicians themselves, whereas now clinical teams may be completely different from research teams, especially outside clinical trials.

3.1.4 Consent/Anonymisation

As noted above, the gaining of consent for epidemiological research or statistical analysis is difficult and possibly impracticable[15], so either a more efficient approach to 'gaining consent' (perhaps by pre-consenting all patients) or the risk to individuals must be removed so no consent is required.
This is recognised in the EU Data Protection Directive:

> *(26)Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the' principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible;*

Unfortunately, while this is fine in principle, it is fraught with difficulty in practical interpretation, especially with the explosion of data available through the Internet (not such an issue when the Directive was drafted). The recent Article 29 Working Paper (WP136) on 'Personal Data' has helped in some ways, if only by highlighting the difficulties of interpretation.

> *(27)Whereas the protection of individuals must apply as much to automatic processing of data as to manual processing; whereas the scope of this protection must not in effect depend on the techniques used, otherwise this would create a serious risk of circumvention; whereas, nonetheless, as regards manual processing, this Directive covers* ***only filing systems, not unstructured files****; whereas, in particular, the content of a filing system must be structured according to specific criteria relating to individuals allowing easy access to the personal data; whereas, in line with the definition in Article 2 (c), the different criteria for determining the constituents of a structured set of personal data, and the different criteria governing access to such a set, may be laid down by each Member State; whereas files or sets of files as well as their cover pages, which are not structured according to specific criteria, shall under no circumstances fall within the scope of this Directive;*

In the UK, The Office of the Information Commissioner had published guidance[16] around the 'temp' test for relevant filing systems, as a practical gauge of whether data related to an individual or not, recognising that any useful data about individuals had some theoretical risk of re-identification, but that a pragmatic answer was needed to allow useful work to proceed. Unfortunately, this type of guidance has not been reflected across Europe where a more rigid legalistic interpretation has often applied, simply stating that there must be 'express consent'.

3.1.5 Data Protection: for patients and professionals

It is important to remember that there is often information, not only on patients, but also on the clinicians involved in treating them. For some purposes, say clinical audit, it will be necessary to be able to identify healthcare professionals, either as persons or by pseudonymised identifiers.
It is important that when de-identifying healthcare records that the de-identification processes are not just applied to the patient alone, but also to fields that may reference healthcare staff.

[15] Singleton, P and Wadsworth, M: 'Practical aspects of obtaining consent for the use of personal medical data in research', British Medical Journal, Jul 2006; 333: 255 - 258
[16] Data Protection Technical Guidance Note: Frequently Asked Questions and Answers about relevant filing systemswww.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/technical_guidance_note_faqs_relevant_filing_systems.pdf

3.1.6 Third-party information

Similarly, there may be references to third-parties: family members or informal carers (as opposed to care professionals). These may be particularly difficult to identify as the references may not appear as structured data which can be recognised as such, but only appear within free-text fields.
For certain areas of study, family relationships are important, but such links represent additional risks as an individual' personal details might be available through another's record. Certainly any use involving genetic details may create such linkages accidentally, especially where rare genetic markers are concerned.
In strict legal terms, these present no different requirements, except in terms of data subject access requests, where such third-party details can be withheld from the data subject. As noted earlier there can be some question of when is a person referenced in a record the 'data subject' or not, and so whether subject access rights apply.

3.1.7 Statistical Disclosure Control

It is recognised in the academic literature that even aggregate information can be subject to statistical analysis to discover information about individuals, either through the identification of individual or rare occurrences or the discovery of 'data rules' (e.g. there are no red-headed psychologists earning under €100,000 as that cell in the report is blank, so any red-haired psychologist that you know must earn over €100,000). This has led to certain strategies for cell suppression, cell perturbation, or Barnardisation[17].
While there is no statute laying down any specific requirements in this regard, there have been a few instances of legal judgements, most recently, a Freedom of Information appeal by NHS Scotland[18], though this ultimately did not resolve whether the barnardised data was 'personal data' or not – sending it back to the Scottish Information Commissioner to decide.

[17] A random modification adding 0, +/-1 to values 2, 3, & 4, and 0/1 to value 1 wherever they occur in a cell
[18] Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland) [2008] UKHL 47 - 9 July 2008: www.publications.parliament.uk/pa/ld200708/ldjudgmt/jd080709/comm-1.htm

4. Issues for this project

The DEBUGIT project seeks to meet legal requirements and protect patient interests by exchanging only anonymised or aggregate data between institutions. This does mean that there is some loss of capability through failure to be able to recognise that cases in different institutions may relate to the same individual, leading to some error in statistical results.
However, in relation to investigations of high-risk infections, it is unlikely that more than one institution would actually be treating a patient, so the significance of this error should not practically affect results.
The decision to communicate only anonymised or aggregate data between the sites and centrally removes some of the obligations on DebugIT that would otherwise be required if identifiable data was being shared and used. However, this still raises three areas that still need careful attention.
1. The processes used to de-identify the data at each site, and in particular how patient identifiers are to be managed in order to support the longitudinal linkage within each site of periodic exports.
2. Security policies that enforce good practice in the handling of data within and between sites, for example including access controls and policies on network access, mobile devices etc.
3. The Ethics and other approvals that are still required to use and communicate patient data even if it is anonymised or aggregated.
4. Audit and other ways in which DebugIT will demonstrate that it is operating in a diligent and transparent manner with respect to protecting the confidentiality of patients (and healthcare staff) and complying with the law.
Each of these four areas is discussed briefly below. Some will be the subject of further work in year 2 of DebugIT.

4.1.1 Anonymisation and pseudonymisation

Anonymisation is a one way transformation in which all personally-identifiable information is removed from a data set, so that it is irreversibly disconnected from the source data. This includes removal of demographic traits, person identifiers and system/resource identifiers (i.e. including laboratory numbers, test id numbers etc.) and also any fine grained details that might allow tracing back to whom the individual record subject is, including dates (which might be blurred to years), locations (which might be blurred to city or country, for example) and organisations (such as hospital names).
It should be recognised that anonymisation is almost always relative rather than absolute. A balance has to be struck between the removal of so many fields as to make re-identification virtually impossible but leaving the data nearly useless for research, and the retention of so many fields that re-identification is readily possible. Kalra and Singleton et al have published a more detailed discussion of this challenge [19].
Pseudonymisation is similar to anonymisation except that one unique link is created between the source data and the released data. This is commonly done by issuing a “study-id” to each data subject, which is added to the anonymised data set. The knowledge of which study-id corresponds to which patient in the source data needs to be carefully guarded, and is usually held (as linkage tables or as a transformation algorithm applied to a real identifier) by the data controller of the source data or a nominated trusted party. If a pseudonymisation approach is pursued it is vital that policies are declared and adhered to for when the linkage tables may be accessed. Commonly this is at the time of re-export, to ensure that each data subject is always mapped to the same study-id to permit longitudinal research and data mining. However, there are sometimes grounds declared for re-linkage if a patient is identified as being suitable to be approached for recruitment into a new study or if a serious risk is identified through data mining and medical staff or patients need to be alerted of this. The linkage tables and algorithms need to be covered by stringent security policies which must include data destruction or retention policies for the end of the research.
Since re-linkage does to some extent break the absolute protection of confidentiality, the policies for re-linkage need to be specified in advance when ethical approval is sought, including the grounds on which re-linkage will be performed and on whose authority. The approvals might need to include permission for some

[19] Kalra,D., Gertz,R., Singleton,P., Inskip,H. (2006). Confidentiality of personal health information used for research. BMJ 333, 196-198. ISSN: 0959-8138

research staff to view fully identifiable data at each data source (i.e. in each CIS) in order to determine which fields to export and to work out the best measures to be applied to anonymise or pseudonymise the data.
At this stage of the project there are no plans to permit the linking of patient details across institutions (and with the slight loss of accuracy of results as noted above) or to allow the re-identification of individuals (mainly to support improved care).
Where results establish possible improvements to care processes or treatments, these will be shared back to project institutions at a protocol level with selection criteria, so that the institutions may search their own identifiable records for cases which may benefit from the revised treatment protocols. This will allow much of the knowledge benefits to be gained without risking patient privacy.
The same de-identification and use of criteria-only results would apply to any outcomes that might apply to care professionals, perhaps as carriers or showing poor infection control.
A more detailed specification of good practice in anonymisation and pseudonymisation has been published as ISO TS25237: 2008 Pseudonymisation.

4.1.2 Security Policies

Because reasonably-anonymised or pseudonymised data does still carry some (albeit small) risk of re-identification, it is considered wise to still secure the data as if it were identifiable, and to follow good practice in managing access controls, applying technical protection measures and disclosure controls.
Each DebugIT site will need to define policies that:

- list the assets to be protected
- list users and their various roles ad corresponding privileges
- specify measures to protect the assets
- define rules for the re-linkage of data sets
- define data retention and data destruction arrangements
- define how audit logs will be maintained and regularly reviewed, and how internal and external audit reviews will be performed or enabled
- specify an escalation pathway in the event of problems being identified, or complaints being received.

Further work in this area will be undertaken in Year 2 of DebugIT as part of Workpackages 2 and 11.

4.1.3 Approvals

4.1.3.1 Ethics issues addressed in EU Submission

As this is a data-only project, there were none of the ethical issues surrounding clinical trials or other activities which directly involved patients.
The fact that only anonymised information would be shared also greatly simplified the ethical background for the project, but given that patient data is involved within each institution, it was important that a strong element of privacy and security oversight was involved to ensure that personal data would not be inadvertently shared.
To this end an Ethics & Data Protection Advisor has been included within the project brief to advise project staff and managers on the appropriate controls and assist with the submission of the various other approvals that would be required. The Ethics & Data Protection Advisor reports into the Project Executive Committee to ensure that any concerns or possible issues are properly addressed and resourced.
The project received EU approval on [DN: complete with appropriate detail]

4.1.3.2 Country/partner-specific approvals needs

Each partner will require appropriate ethics approval for the sharing of derived patient data with other partners in the DEBUGIT project. While, these will be the same in principle, there are variations within each

member state and local institution as to the paperwork required and the level of detail that may need to be submitted.
Each Research Ethics Committee is likely to have different concerns and may require specific questions to be addressed. The Ethics & Data Protection Advisor will be able to help partners present the appropriate responses to such questions in relevant terminology and quoting relevant precedents or support.

4.1.3.3 Internal project approvals

There will need to be appropriate internal controls concerning access to patient data by members of the project team in their capacity as researchers in the DEBUGIT project. Before such access begins, the project managers will need to ensure that appropriate controls are in place to support this access and to prevent or limit possible abuse.
Identification of these controls is one of the tasks for each institution, followed by their implementation.
Each institution will also need to be confident that any other institution due to receive data has appropriate controls around the subsequent use of that data before the data would be sent. This will require each institution to establish the relevant controls and to publish to the other project participants that these controls have been established.

4.1.4 Audit

Members of the UCL team will support the development of coherent and consistent policies across all of the sites.
The DebugIT Ethics & Data Protection Advisor will liaise with each site to ensure that satisfactory arrangements are in place, including policies and approvals.
As part of diligence and transparency each site will need to conduct a periodic audit review of its policies and of their implementation. The DebugIT Ethics & Data Protection Advisor will oversee this process across the consortium, and collate audit results, but is is clearly not scalable for the Advisor personally to undertake all such reviews. In general a process of internal audit review will be encouraged. The approach to audit reviews will be elaborated, in partnership with each site, in year 2 of the DebugIT project.

5. Issues for specific Work Packages or Tasks

The involvement of the various DEBUGIT partners is detailed below:

Organisation	Data Set Size	Data Source/Data type	Data Structure	Codes	Start
GAMA	98,000 patient records	HIS data, including patient demographics, appointment scheduling, examinations, procedures and tests, admission data, hospital stay data, discharge data, death and autopsy data	Structured, semi-structured and free text	ICD 9/10	2006
HUG	1,000,000 patient records	Lab results, clinical results, discharge letters etc.	Structured, semi-structured, free text	ICD 10	2000
INSERM/ HEGP, Pharmacovigilance	6,000 case reports	Patient demographics, patient history, medication history	Structured and free text	ICD 9, WHO – ART, MedDRA	1983
INSERM / HEGP, EHR	350,000 discharge summaries	Patient demographics, diagnostics, patient history, medication ordering, lab results, imaging reports	Structured, Semi-structured, free text	ICD 10, CCAM	2000
IZIP	1,000,000 patient records	Patient demographics, basic patient information (blood group, allergies, chronic diseases, patient history, vaccination, ambulatory examinations, emergency information, lab results, x-ray, ultrasound, CT, MRI – scans (description, PACS), hospitalisation (including hospital reports etc.), medication history, patient notes	Structured, free text	ICD 10	2002
UKLFR	50,000 discharge summaries	Lab data, nursing documentation, admission and discharge information, selected scores	Structured, semi-structured, free text	ICD 10, TNM	2000
LIU	80% of admission data to adult general intensive care, 100% of paediatric, 20% of neurology and 40% of cardiothoracic ICU admission (2007)	Data on patient demographics, care data, risk score, diagnosis, death in ICU, procedures, complications etc.		ICD 10, NOMESO 1.9, APACHE III	2001
TEILAM	approx. 10 patient record sets / week	Patient demographics, admission and discharge information, temporal sequences of ICU data including radiographic images, prescriptions and antibiograms.	Structured, semi-structured, free text		2007

This clearly shows that ‘sensitive personal data’ will be at the source of the DEBUGIT data, but protocols for extracting relevant data from source sites have not yet been established in full enough detail to comment on any remaining risks or issues at this level.

5.1.1 Detail on the specific DebugIT partner sites

The following templates, provided by each site, are the first version of a data custodianship summary which outlines the kinds of data held and the purposes of use and/or communication. This template will inform the DebugIT Ethics & Data Protection Advisor of the kinds of approvals, policies and measures that need to be verified as being in place, and the areas for periodic audit review.
A comparison across templates will also enable sites to share useful security resources, and enable UCL to help contribute to generic project-wide policies.

Site Name:	[e.g. Agfa]										
Contact:	[main contact for ethics & security]										
Address:											
Telephone:											
Email:											
WP involvement:	1	2	3	4	5	6	7	8	9	10	11
[enter Y or N as appropriate]											Y
Role(s) within project:	Holding data repository								Y / N		
	Gathering data								Y / N		
	Sharing data/servicing queries								Y / N		
	Using/analysis of data								Y / N		
Approvals being sought (& status):	[Description of any ethics applications done or underway]										
Approvals needed (& when)	[Description of any further ethics applications needed]										
Data being Held											
Type of data	[Content, quantity, types of individual]										
Individual or Aggregate	[whether identifiable or anonymised or aggregate]										
Key identifiers	[May include locally generated IDs, or geographical information, even occupation]										
De-identification used											
Purpose/use of data	[Why you hold it as part of DEBUGIT]										
Data Being Received	From:										
Type of data											
Individual or Aggregate											
Key identifiers											
De-identification used											
Purpose/use of data											
Data being Shared	To:										
Type of data											
Individual or Aggregate											
Key identifiers											
De-identification used											
Purpose/use of data											
Data being Used	Purpose:										
Type of data											
Individual or Aggregate											
Key identifiers											
De-identification used											
Purpose/use of data											

END OF DOCUMENT