Deliverable 7.1.1

Project ID	654241
Project Title	A comprehensive and standardised e-infrastructure for analysing medical metabolic phenotype data.
Project Acronym	PhenoMeNal
Start Date of the Project	1st September 2015
Duration of the Project	36 Months
Work Package Number	7
Work Package Title	Privacy and Ethics
Deliverable Title	D7.1.1 Workshop on best practices in handling sensitive human data, taking into account National and Institutional legal policies
Delivery Date	M3
Work Package leader	ICL
Contributing Partners	ICL, EMBL-EBI, IPB, UB, UL, UOXF, SIB, UU, BBMRI

Authors: Robert Glen, Tim Ebbels, Daniel Schober, Steffen Neumann, Marta Cascante, Kenneth Haug and Namrata Kale

Contents

1. Executive Summary

For a health care infrastructure project on human metabolomic data like PhenoMeNal it is key to follow best practices in information privacy and ethics. To this end, two workshops to disseminate best practice were planned. This deliverable relates to the first workshop held on 20th November 2015 at Imperial College London.
The workshop was attended by 22 people from the PhenoMeNal partners and included 5 invited speakers around the themes of ethics, data security, clinical phenotyping, legal frameworks and European data sharing efforts. Clear ways forward were identified to aid the implementation of state of the art privacy and ethics standards within the PhenoMeNal infrastructure. The main conclusions of the workshop were:

1. Privacy and ethics issues are mainly governed by the consent given by the participants, and the level of anonymisation of the data.
2. There is no consistent set of laws or policy around privacy and ethics in the EU and even less outside it. The EU is currently revising its data protection legislation. Projects such as BiomedBridges have developed tools, which could be of use for PhenoMeNal here.
3. The European Genome-Phenome Archive (EGA) is a good model on which to base PhenoMeNal activities. We may consider the EGA data access forms and agreements as a starting point for the PhenoMeNal data submission form.
4. The AIRWAVE project has potential as a good use-case for PhenoMeNal. This will most likely need an application to the AIRWAVE access committee.

2. Project Objectives

No.	Objective	Yes	No
1	D7.1.1 Workshop on best practices in handling sensitive human data, taking into account National and Institutional legal policies	X	

3. Detailed report on the deliverable

a. Background

Patient and research subject data is very sensitive, and it is paramount to establish a robust governance framework for overall information management including sensitive data. The PhenoMeNal e-infrastructure will be able to cope with data generated from comprehensive clinical, genotypic, 'omics and analytic sources including medical records, electronic health records, clinical measurements, genotypic data, phenotypic and metabolomics data from tissue and biofluid analysis, image and pathology data. Primarily, all data collected and held within the project will

comply with all local laws, regulations and ethics. All personal information will be processed in accordance with accepted Data Protection Principles. Responsibility for data will be with the host institution/data provider.
In accordance with the above Ethical, Legal and Social Implications (ELSI) requirements, two workshops on best practice in handling sensitive human data are to be organised. The first of these took place on 20th November 2015 at Imperial College in London, UK, organised by Prof. Robert Glen and Dr. Tim Ebbels.

b. Workshop Overview

The workshop was conducted by Imperial College London on 20th of November at Imperial College, South Kensington Campus. The details of the agenda are attached as **Annex 1.**

c. Introduction to the workshop

The workshop began with five talks from invited speakers who were experts in the field of data protection, privacy and ethics. Below, the agenda is summarised, with notes on each individual talk and discussion.

Prof. Holm (University of Manchester) discussed the problems of data sharing across national boundaries. He highlighted the reasons why data protection was important, and approaches to protecting privacy. Key points are that the consent signed by the patient governs what can be done, and that several data providers (e.g. UK Biobank) rely on users signing a data access agreement which is legally binding.
Prof Nicholson (Imperial College London) gave an overview of the collection and use of omics data at Imperial College. He illustrated the pipeline joining research and clinical practice using the Institute of Translational Medicine and Therapeutics (ITMAT). A key technology in the vanguard of translational research is the 'intelligent knife', a surgical tool which allows surgeons to obtain real time metabolic information about the tissue being cut, assisting in the surgeon's real time decisions.
Dr Ansari (Addenbrook's Hospital) described the operation of the research governance office at Addenbrook's Hospital in Cambridge, UK. At any time, over a thousand studies are being conducted within the trust, with several hundred going through the approvals process. She outlined the different types of anonymised data and the requirements of the UK Data Protection Act (similar legislation exists in other EU member states). A key point is that the latter only applies to living individuals, although other restrictions (e.g. medical confidentiality) would apply to deceased individuals too.
Dr Saunders (EMBL-EBI) gave a technical overview of the EBI approach to controlled data access, exemplified by the European Genome-phenome Archive (EGA)(Nature Genetics 47, 692–695, (2015)| doi:10.1038/ng.3312). The key points were that all access is controlled by Data Access Committees (DACs) who set terms

of access and approve applications for access. Each project (e.g. Wellcome Trust Case Control Consortium) has a separate DAC. Each DAC can administer several different data sets and control access according to several different Data Access Agreements (DAAs). The DAA specifies the terms under which a user may access a given data set and is prepared by the DAC. Technically, a series of public/private encryption technologies are used to deposit or retrieve data to/from the archive, ensuring only those with permission can access. It was discussed that the EGA forms a very good prototype on which to base the PhenoMeNal architecture.
Dr Schlünder (Scientific Consultant for Legal Affairs & Data Protection at TMF-ev.de) described the legal aspects of data protection and anonymisation at the European level. It was pointed out the a new EC Directive on privacy is currently being negotiated and is hoped to be agreed in the next few months. (For more information and details on the negotiation agenda see http://www.eppgroup.eu/news/Data-protection-reform-timetable). It is clear that the situation in the USA is different, with defined lists of variables (the Health Insurance Portability and Accountability Act (HIPAA) list) which, if present, jeopardize 'anonymity' (see https://privacyruleandresearch.nih.gov/pr_08.asp for more details on "How can Covered Entities Use and Disclose Protected Health Information for Research and Comply with the Privacy Rule"). In Europe we will likely have to rely on 'broad consent' allowing reuse of data for future ethically approved studies. The importance of considering risk as a criteria for data protection was a theme discussed.
Dr Ibrahim Karaman (Imperial College London) presented a possible exemplar data set to be used by PhenoMeNal. The AIRWAVE project (http://www.police-health.org.uk/) is a large epidemiology study involving 3000 individuals who have had NMR and UPLC-MS metabolic profiles of plasma and urine recorded at the UK National Phenome Centre. The data set comprises around 15000 NMR spectra and 24000 UPLC-MS profiles. Usage of the data will probably require a successful application to the data access committee. It was felt that this would be an excellent data set / use case to develop / demonstrate PhenoMeNal capabilities, and that it would be desirable to have access to the real data, rather than a 'scrambled' version of it (which might be easier to distribute).

d. Workshop participants

A total of 22 people attended the workshop including representatives from the partner organisations and 4 experts with previous handling of sensitive data and ethics. The details of the participants are attached as **Annex 2.**

e. Discussion and Outcomes

The following is a short summary of the discussion.

- Identifiable in a legal/ethical context means if at least one of the participants can be identified from the data. So for example, even if 1 out of 1000 study

participants could be identified, this would count as identifiable data. So while the average person might not be identifiable from their metabolic profile, some individuals with extreme phenotypes might be, and this would make the data 'identifiable'.

- The EGA seems to be an excellent 'prototype' for PhenoMeNal. The EGA DAA would be a good place to start for the PhenoMeNal data provider form.
- Re-analysis (e.g. meta-analysis) of data: would a user have to apply to every DAC to search for data which might be relevant for reanalysis?
    - Some DACs already make some data available for searching purposes. Generally at EGA, gender, sample alias & 'high level phenotype' are public, and in the future more will be so, to enable effective searching.

The main outcomes of the workshop were as follows.

1. We will explore following the example of EGA as a model for addressing privacy issues.
2. Specifically, we will look at EGA's DAA and associated documents to develop the PhenoMeNal data provider form.
3. We will make an application to use the AIRWAVE data as a use case for PhenoMeNal. We will aim to use the full data (not scrambled) so that it is a real test of the infrastructure to handle patient sensitive data.

Early feedback from the participants suggest that this was a useful and constructive introduction to ELSI requirements and will form a firm basis for software development plans.

4. Delivery and Schedule

The delivery is delayed: No

5. Background information

Patient and research subject data is very sensitive, and it is paramount to establish a robust governance framework for overall information management including sensitive data. The PhenoMeNal e-infrastructure will be able to cope with data generated from comprehensive clinical, genotypic, 'omics and analytic sources including medical records, electronic health records, clinical measurements, genotypic data, phenotypic data from tissue and biofluid analysis, image and pathology data. Primarily, all data collected and held within the project will comply with all local laws, regulations and ethics. All personal information will be processed in accordance with accepted Data Protection Principles outlined above. Responsibility for data will be with the host institution/data provider.

Work package number	WP7	**Start date or starting event:**	M1
Work package title	Privacy & Ethics		
Participants	EMBL-EBI, ICL, IPB, UB, UL, UOXF, SIB, UU, BBMRI		

Objectives

Objective 7.1 Develop appropriate policies, procedures and management accountability and structures to provide a robust governance framework for information management.

Objective 7.2 Raise awareness of information governance within the consortium and assure ongoing compliance

Objective 7.3 provide a forum for information exchange on best practice in clinical data sharing and disclosure.

Task 7.1 Organise workshops involving Scientific and Technological Advisory board (STAB) and the Management Committee (MC) and other necessary players to devise best practices in handling sensitive human data, taking into account National and Institutional legal policies. (ICL, EMBL-EBI)

6. Annexes

1. Agenda
2. List of participants

Annex 1: Agenda

Time	Description
09.30 - 10.00	Refreshments
10.00	Brief overview and introductions by participants (Glen, Ebbels)
10.15	Prof Søren Holm (Chair in Bioethics at the Centre for Social Ethics and Policy, part of the School of Law at the University of Manchester). ELSI, consent and omics data.
11.00	Prof Jeremy Nicholson (HOD Surgery and Cancer, Imperial). Clinical and omics data within an ELSI environment
11.30	Dr Tracy Ansari (Research Governance Coordinator, Addenbrooke's Hospital). Ethics and approval for clinical data studies.
12.00	Dr Gary Saunders (European Genome-Phenome Archive at EBI). Technical aspects of storing and accessing genomic data.
13.30	Dr Irene Schluender (Lessons learned from BioMedBridges)
14.00	Discussion
17.00	Close of meeting

Annex 2: List of attendees

Organisers:

- Robert Glen
- Tim Ebbels

PhenoMeNal Project:

- Ken Haug
- Namrata Kale
- Reza Salek
- Merlijn van Rijswijk
- Theo Reijmers
- Etienne Thevenot
- Daniel Schober
- Philippe Rocca-Serra
- David Johnson
- Pablo Moreno
- Christoph Steinbeck
- Ibrahim Karaman
- Kim Kultima
- Payam Emami

Invited Speakers:

- Søren Holm
- Jeremy Nicholson
- Tracy Ansari
- Gary Saunders
- Irene Schluender
- Stephanie Suhr - *was not able to attend due to illness*.