Deliverable 7.4

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	D7.4 Process to extract maximum information from sensitive datasets with minimum compromise, in collaboration with BBMRI and BioMedBridges
Version	Revised
Delivery Date	M12
Work Package leader	ICL

Contributing Partners	All
Authors	Robert Glen, Timothy Ebbels, Namrata Kale

Abstract: In this deliverable we provide workflows and a data provider form to guide users through the process of considering ELSI constraints for use of their (or other) data in PhenoMeNal. In addition we specify a registration process for PhenoMeNal and provide additional guidance on accessing relevant ELSI information.

History of Changes:
Incorporated the change: 95/46/EC is Data Protection Directive, not the General Data Protection Regulation

Table of Contents

1. EXECUTIVE SUMMARY

PhenoMeNal is creating a compute infrastructure to process data, principally from metabolomics, but also including related 'omics technologies. Depending on the data intended to be processed, Ethical Legal and Social Implications (ELSI) may have a strong bearing on the use and re-use of the data. Researchers should consider these implications carefully before using the infrastructure. To assist in this process of assessment, we are developing a step-wise approach to allow researchers to properly evaluate ELSI constraints in the context of their own datasets (**Task 7.5**: Develop workflows in collaboration with BBMRI and BioMedBridges to extract maximum information from sensitive datasets with minimum compromise within legal, ethical and privacy constraints).

This is based on two previous workshops, the deliverables (**D7.1** Workshop on best practices in handling sensitive human data, taking into account National and Institutional legal policies, and **D7.2** Report on policies and procedures for sensitive human data management) included guidance on handling sensitive human data in PhenoMeNal, as well as Terms of Use for accessing the PhenoMeNal Infrastructure.

The general use-case for PhenoMeNal is to provide a secure environment to process input data and to provide a series of data outputs for further analysis - **it is not envisaged that data will be archived and stored in PhenoMeNal** (e.g. as in the European Genome-phenome Archive). However, PhenoMeNal will allow access to Open Databases of metabolic and other related 'omics data such as MetaboLights[1] (at EBI), and the deposition of data after processing (if appropriate in MetaboLights). Access to data within open databases is often still subject to terms of use[2].

There are many variations on how PhenoMeNal can be used including:

- Data processing engine e.g. PhenoMeNal can be used to process confidential data,
- Input and output data are controlled by the researcher e.g PhenoMeNal is installed locally in the researcher's home institution, taking the compute to the data
- Researcher gives access to their data for processing by another researcher. The onus therefore in making PhenoMeNal available for use is to ensure that before

[1] http://www.ebi.ac.uk/metabolights/
[2] http://www.ebi.ac.uk/about/terms-of-use

data is uploaded into PhenoMeNal, the researcher takes into account any ELSI constraints.

To assess the ELSI constraints of data before submission to PhenoMeNal, the use of a **Data Provider Form** was discussed in detail at the previous workshops, and in deliverable **D7.3** (Evaluation report for the introduction of a data provider form, M8)**.** It was concluded that the approach taken in the EU project BioMedBridges[3] (BMB) was the most appropriate and advanced, and with suitable modification would serve as the template for PhenoMeNal. In BMB, the guidance on the use of data and the relationships between parties was explored and this was formulated in a number of Templates for data sharing. These are based on the document: **BioMedBridges Ethical Governance Framework**[4].

The Data Provider Form is to ensure consideration by a data supplier that all data collected and held within the project will comply with all local laws, regulations and ethics and all personal information will be processed in accordance with accepted **Data Protection Principles**. In the EU, the law on the use of private information in the clinical setting is under review (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data)[5].

In particular, data used for medical research is specifically considered. In paragraph (33) it is proposed "*It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose*" and in paragraph (157*)* "*By coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression. On the basis of registries, research results can be enhanced, as they draw on a larger population….In order to facilitate scientific research, personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards set out in Union or Member State law*".

[3] http://www.biomedbridges.eu

[4] http://www.biomedbridges.eu/sites/biomedbridges.eu/files/documents/deliverables/BioMedBridges%20Ethical%20Governance%20Framework_V1-1.pdf

[5] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

The informed consent of the participating patients in data sets that have been collected, and which are intended to be used for testing of the PhenoMeNal infrastructure, was considered in deliverable **D7.5** (Report to the EC/REA with ethical approvals, informed consent forms and patient information material of datasets to be used within PhenoMeNal e-infrastructure, M8). The re-use of data in open repositories (or otherwise) and the transfer of data between researchers using the PhenoMeNal infrastructure will be facilitated by the use of the PhenoMeNal Data Provider Form.

As reiterated in previous reports, the final responsibility for any data submitted to PhenoMeNal will be with the host institution/data provider.

To assist in this process, we are providing:

- Workflows to take the user through a number of key steps to allow proper consideration of ELSI
- A registration form to collect user information
- The provision of PhenoMeNal Terms of Use (Appendix 1a)
- A Data Provider Form

2. CONTRIBUTION TOWARDS PROJECT OBJECTIVES

This deliverable contributes towards the following objectives:

- Develop appropriate policies, procedures and management accountability and structures to provide a robust governance framework for information management.
- Raise awareness of information governance within the consortium and assure on going compliance.
- Provide a forum for information exchange on best practice in clinical data sharing and disclosure.

The deliverable thus aims to:

- To generate understandable and easy to follow workflows describing steps to be taken prior to data transmission, data storage, data usage and publication.

- Provide guidance on ethical and legal issues and requirements towards sensitive data.
- Provide an example of a data provider form to guide users through the process of considering ELSI constraints for use of their (or other) data in PhenoMeNal.
- Define a registration process for PhenoMeNal.

3. DETAILED REPORT ON THE DELIVERABLE

3.1. Background

Two PhenoMeNal workshops were held on ELSI (20th November 2015, 27th January 2016) The outputs from these workshops formed the basis of the deliverables D7.2 and D7.3 as mentioned above and based on these discussions, we now provide **flowcharts** giving step-by-step guidance for consideration of ELSI before submission of data to PhenoMeNal.

These workshops also formed a consensus on the appropriate and cost effective way of adopting/developing data provider criteria, which resulted in a Data provider Form and advice on completing the form. Additionally to confirm that the terms of use and ELSI have been considered, a registration page has also been created to collect user ID, institutional location, confirmation that the terms of use have been read and that a data transfer/use form (if required) has been completed.

3.2. Registration for PhenoMeNal

The registration page will capture the following information (Figure 1):

- Account Details
  - Email
  - Password
- Personal Details
  - Full Name
  - Display Name
  - Organisation
- Terms and Conditions must be checked:
  - I am fully aware of the ELSI requirements.

ELSI requirements are linked to documents outlining ELSI guidance (section 3c).

- I have Agreement to use PhenoMeNal with my data.
  Agreement is linked to the PhenoMeNal Data Provider Form (3d).
- I accept the PhenoMeNal Terms and Conditions.
  Terms and Conditions is linked to the PhenoMeNal Terms of use (Annex 1a).

- Helpful Information:
  Additional links to PhenoMeNal Tutorials, Manuals, Publications.

Additionally, the user will agree to cite PhenoMeNal in any publications:

The Recipient agrees to acknowledge PhenoMeNal Resources and applicable publications, the source of the data disclosed/used in any publications or other public disclosures reporting use of PhenoMeNal. The following form of words should be used: "We acknowledge [PhenoMeNal References], funded by [EU (Horizon 2020, 654241)] for the provision of the PhenoMeNal Infrastructure".

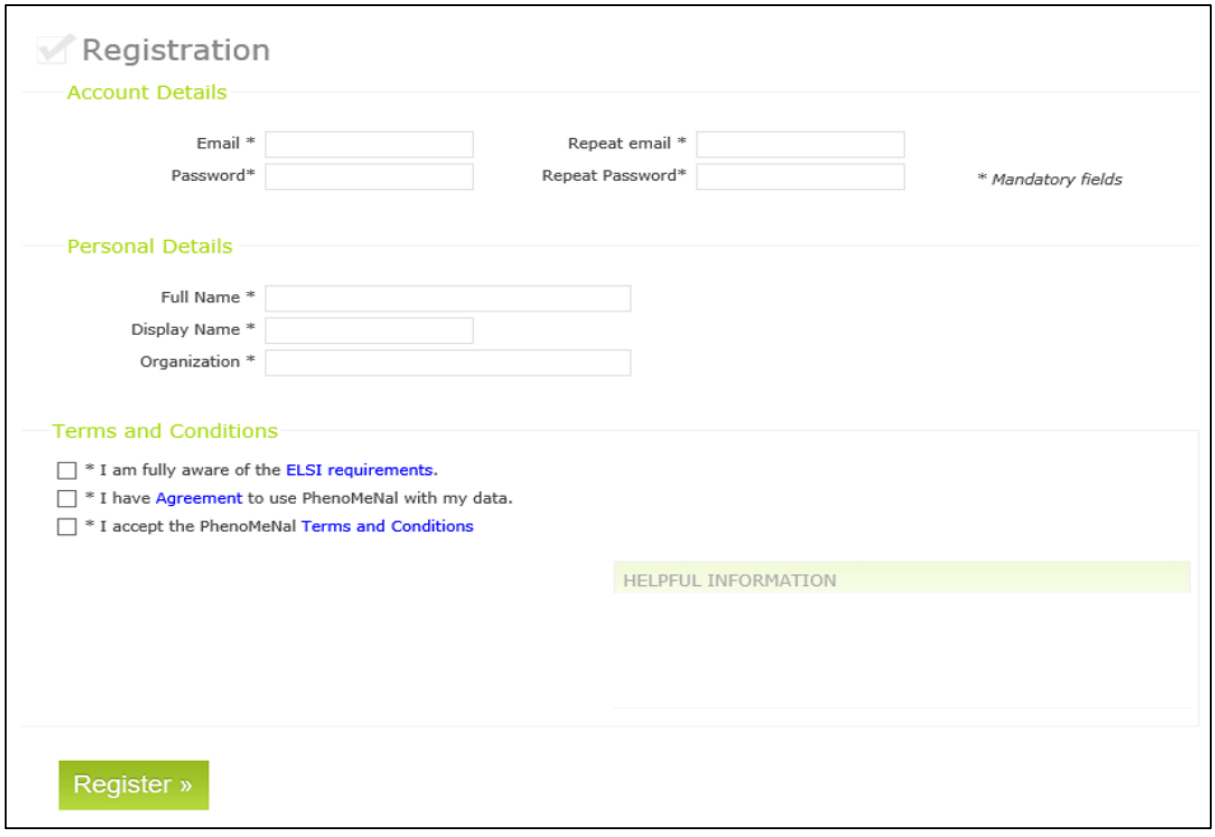

Figure 1. The image of a registration page currently under development.

3.3. Flowcharts for ELSI consideration in PhenoMeNal

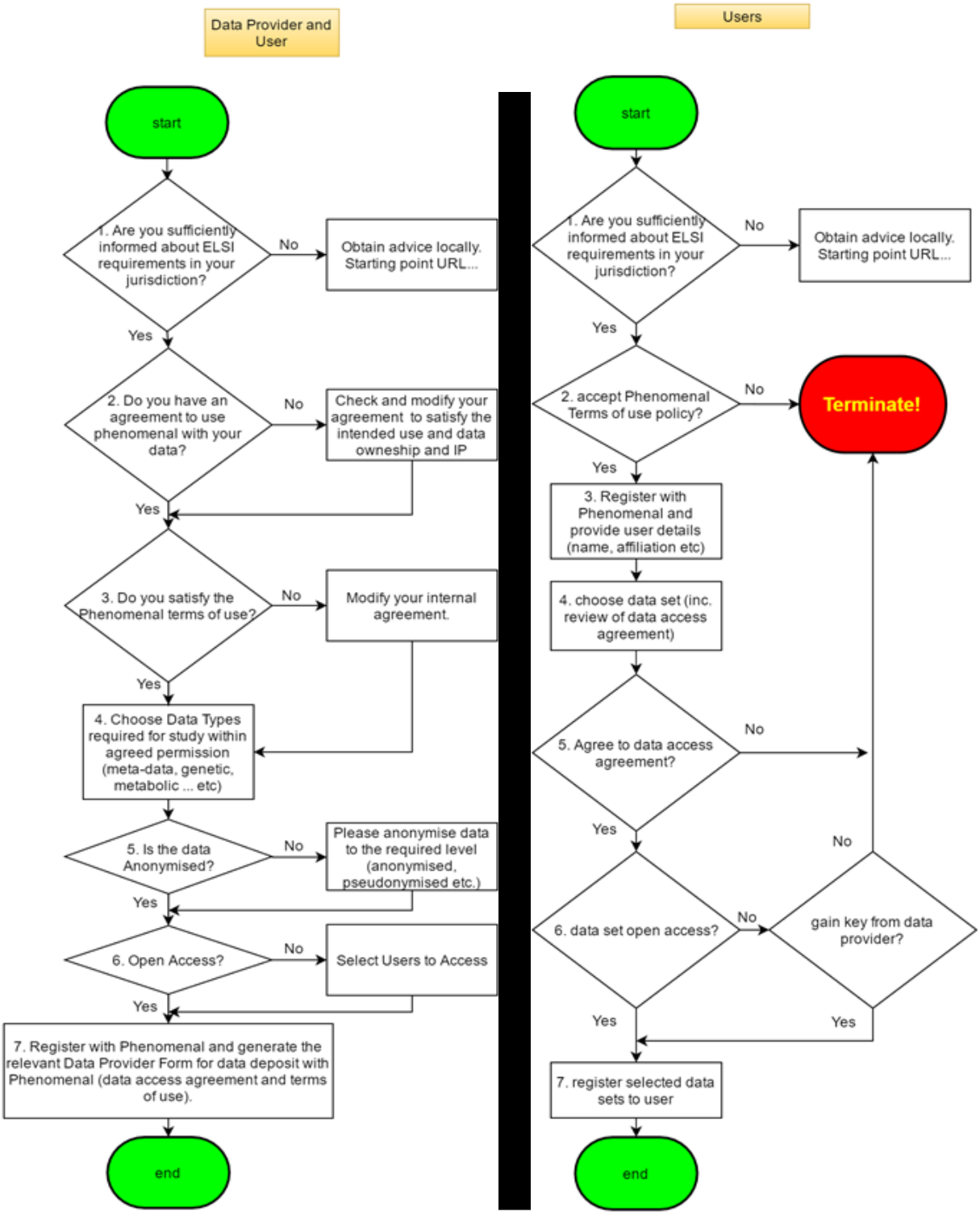

Figure 2. Flowcharts for ELSI consideration

Two flowcharts were produced (Figure 2). In the first case (left), the user (**provider**) will make data available to be analysed by another user. In the second case (right), data is already available (e.g. from the MetaboLights repository) and the user wishes to process or analyse the data in PhenoMeNal (in this case the user is an **analyst**). Typically, a provider will also be a user, in that they will bring their own data and process/analyse it within the PhenoMeNal environment. In this case both flow charts will be relevant.

As shown in the chart above, in both cases (*provider/analyst*) the process starts by ensuring that the user has considered whether they are sufficiently informed about ELSI requirements in the jurisdiction where they will run the workflows. As a simple example, a user may have data generated in their own lab in the UK and wish to use the PhenoMeNal cloud (based at EBI in the UK) to process this data. In this case, the user would need to be aware of ELSI requirements of UK law, but not those of other EU states. A more complex scenario might consist of users accessing data provided from one state, and analysing it in another state, in which case the ELSI requirements of more than one state are pertinent. We point users to detailed information at a number of websites (below), which can help in this regard. If further information of guidance is required then the user should seek assistance at their institution.

- **Global Alliance for Genomics and Health**. The Global Alliance is seeking to make it easier for researchers and clinicians to access data and combine data sets, whilst also considering the ethical, legal and societal implications of the work they are doing and ensuring that research participants' privacy is protected [6].
- **BioMedBridges Ethical Governance Framework** The BioMedBridges Ethical Governance Framework sets out policies that specifically relate to ethical and regulatory issues with regard to the access of sensitive data [7].
- **Regulation (EU) 2016/679 of the European Parliament** and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (Data Protection Directive).

For the *provider*, the next step is to check that they have permission within their collaboration/organisation to use PhenoMeNal with the particular data set at hand. For example, this could consist of applying to an access committee for permission to analyse data from an external project, or simply informally checking with collaborators that the PhenoMeNal cloud application is appropriate. Any restrictions found at this stage might

[6] https://www.ebi.ac.uk/training/online/course/quick-tour-global-alliance-genomics-and-health-els/elsi-standards

[7] http://www.biomedbridges.eu/sites/biomedbridges.eu/files/documents/deliverables/biomedbridges_ethical_governance_framework_v1-2.pdf

be addressed by deploying a local instance of the PhenoMeNal VMIs behind a firewall, as would normally be done in the case of sensitive clinical data.

Before using the system, both types of users will need to ensure that they accept the PhenoMeNal terms of use (see Annex 1a). This ensures that users understand their obligations and responsibilities, particularly with respect to ELSI. After this point the two workflows diverge.

For the *provider*, the next step is to decide which items of data they are able to share/upload. For example, they may be able to upload anonymised metabolomics data, and some basic clinical metadata, but not genomic or more specific clinical data. This is up to the *provider* to decide on a case by case basis, depending on their understanding of ELSI requirements and agreements within their organisation. Next, they must ensure that the data is suitably anonymised. This means it must be either fully anonymised (impossible to identify participants) or pseudo-anonymised (coded, with codes identifying patients only available to a trusted third party). We recommend the following resources for information on state of the art practices in anonymization:

- Managing and Sharing Data. UK Data Archive, van den Eynden V. et al., University of Essex[8] ISBN: 1-904059-78-3.
- Hrynaszkiewicz et al. Trials 2010, 11:9[9].
- Guidance on the anonymisation of clinical reports for the purpose of publication, European Medicines Agency[10].

The *provider* will then need to decide which other PhenoMeNal users will be able to access the data. It is recommended that only the data cleared for analysis is uploaded into PhenoMeNal. The access to data may be controlled by an access control system and access should be sought from the data provider. Such a system is present in the EGA.

Finally, the *provider* must generate and complete the relevant data provider form using the templates provided or the appropriate local consent forms.

For the *analyst,* after accepting the PhenoMeNal terms of use, the user will need to register their account and select which data set(s) they wish to access. This will include review and acceptance of the relevant Data Access Agreement (set up by the provider). Note that even public data sets will be subject to some access agreement. If the data is not publically available, a key must be obtained from the system (ultimately at the decision of the data host) for the *analyst* to access each data set. The (external data

[8] http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
[9] https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-11-9
[10] http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2015/09/WC500194087.pdf

provider) system should then register the selected data sets to the *analyst* and allow them to proceed with the computational workflow in PhenoMeNal.

3.4. PhenoMeNal Data Provider Form

For users considering giving access to anonymised clinical data for use in PhenoMeNal this Data Provider Form is provided as a guide to ensure that ELSI considerations are considered and also as a record of the transaction. Local laws and regulations may dictate that a different form is used, users should obtain guidance before proceeding.

PhenoMeNal Data Provider Form

This form is intended to be used as a template for parties providing data to another party or where there are underlying restrictions on the use of the data imposed by consent requirements, ethics committee approvals, national regulations or any other reasons for use within PhenoMeNal.

NOTE: Custody of all data made available within the project remains with the provider (and, where applicable, the participants).

The data provider has full responsibility for the data when making it available within the project under the terms agreed within the PhenoMeNal Terms of Use.

Please provide the following information:

I. Name of data provider

II. Name and address of data providers' research institute/university

III. Name of dataset

IV. Please list the restrictions on use imposed by consent requirements, ethics committee approvals, national regulations or any agreements, such as, research collaboration agreements, material transfer agreements, and data access agreements, including those made with other parties who may have originally supplied the data or other material facts.

V. Is the data linked or unlinked anonymised*?

VI. If linked anonymised, name the person(s) holding the linkage key

VII. If linked anonymised, please give the name and address of the linkage key holder's research institute/university

VIII. If applicable, please state if there is a date by when this dataset must be removed/deleted/returned.

IX. Please state any decisions made regarding the management and communication of findings of individual clinical significance, including any obligations data requestors may have to communicate findings, and any pre-set time limits for the feeding back of results

Please sign to indicate that you have read the PhenoMeNal Ethical Governance Framework document and you agree to abide by the conditions contained therein and that the donor consent provisions and/or ethical approval, and national laws and regulations, allow the use of the data in PhenoMeNal. If you are unsure whether the current consent provisions or ethical approval adequately allow the use of the data in PhenoMeNal, we recommend you seek advice from an appropriate ethics committee

If required:

Signature of PhenoMeNal Coordinator representative

Date of approval

*Linked anonymised (or pseudoanonymised) means that the data is coded and can be linked back to the participant by the holder of the linkage 'key', but not by the third party researcher accessing the data. Unlinked anonymised means that no one is able to identify which participant the data originated from.

4. WORK PLAN

4.1. Structure and management of WP7 Tasks

WP7 tasks are coordinated by **ICL** with specific contributions from EMBL-EBI, IPB, UB, UL, UOXF, ISB, UU and in collaboration with the other members of the consortium.

The work plan follows the identification and workflow of the tasks identified. These are:

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)

Task 7.2 Establish and generate a robust workflow for data dissemination and disclosure taking into account guidance for task 7.1.

Task 7.3 Identify governance managers in each collaborating research group who are responsible for overseeing day to day Information Governance issues; developing and maintaining policies, standards, procedures and guidance; coordinating Information Governance in the consortium; and ensuring on-going compliance from all the project participants. (ICL, UL, EMBL-EBI)

Task 7.4 Evaluate the introduction of a data provider form to be completed by each data provider to the project, the intention of which is to ensure that all ethical aspects of making data available within PhenoMeNal have been addressed, similarly to what implemented in the BioMedBridges project. (ICL, EMBL-EBI)

Task 7.5 Develop processes in collaboration with BBMRI and BioMedBridges to extract maximum information from sensitive datasets with minimum compromise within legal, ethical and privacy constraints. (ICL, EMBL-EBI)

Task 7.6 Raise awareness of information management within the consortium and the user community during regular PhenoMeNal outreach. (ICL, EMBL-EBI, UL, UB).

4.2. Performance metrics and management of tasks

The achievement for the work package (WP7) has been as measured by the following performance metrics:

- Reports on privacy management successfully completed and workshops outcomes (D7.2, submitted M6, PhenoMeNal terms and conditions)
- Data disclosure process and procedure established (D7.3)
- Agreed disclosure form (work in progress for the PhenoMeNal data provider form, PhenoMeNal registration progress)

Project management was integrated using Pivotal Tracker with other PhenoMeMal tasks.

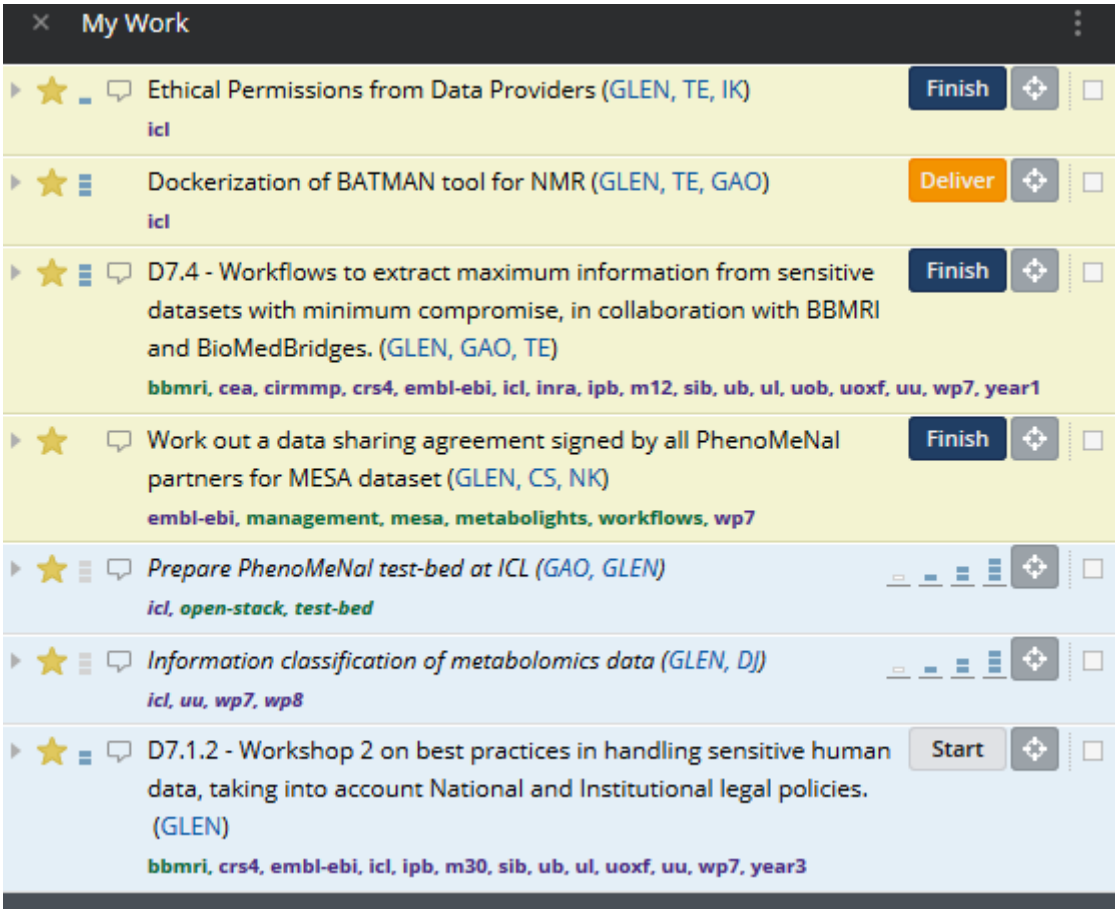

Figure 3. Screen shot from pivotal tracker showing distribution and assignment of tasks

Utilization of Resources:

The total PMs (person months) utilised until M12 (inclusive)

Partner	EMBL-EBI	ICL	IPB	UB	UOXF	SIB	UU
PMs	3.5	11	2.4	1.5	1.6	0.25	0.8

4.3. Project Risk Assessment

WP7 was established to mitigate the ELSI risks in Phenomenal. The Risks and the mitigation steps taken are outlined below.

Principal Risks identified:

- Metabolomics data could be classified as “data which relates to a living individual who can be identified from those data”
- Data access could be obtained without permission
- Data could be copied and re-used without permission
- Inadequate permission could be obtained from the access committee
- Use of Phenomenal breaks one or more laws or ethical constraints
- Consideration of ELSI is not applied adequately

Mitigation

- Registration process forces users to consider ELSI
- Adequate information is provided to educate users
- Users are warned that they have personal responsibility for their use of data in Phenomenal
- Phenomenal is available as a locally installed process within the institutions firewall (taking the software to the data)
- A data sharing form is provided to allow consideration of ELSI in data sharing before use of Phenomenal.

5. DELIVERY AND SCHEDULE

The delivery is delayed: No.

6. CONCLUSION

In this deliverable we provide workflows and a data provider form to guide users through the process of considering ELSI constraints for use of their (or other) data in PhenoMeNal. In addition we specify a registration process for PhenoMeNal and provide additional guidance on accessing relevant ELSI information.

7. REFERENCES

A. www.p3g.org/resources/ipac
B. http://www.bbmri-wp4.eu/wiki/index.php/Main_Page
C. http://www.hsern.eu/index.php/
D. http://www.biomedbridges.eu/deliverables/52-0
E. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
F. https://wellcome.ac.uk/what-we-do/our-work
G. http://www.mrc.ac.uk/research/policies-and-resources-for-mrc-researchers/data-sharing/data-sharing-population-and-patient-studies/

8. ANNEXES

8.1. PhenoMeNal Terms of Use version 1.0

The following is the first version of the PhenoMeNal Terms of Use which will be available on the PhenoMeNal website and associated infrastructure pages. Users will be required to accept these terms before accessing the PhenoMeNal infrastructure.

PhenoMeNal Terms of Use version 1.0

PhenoMeNal is an integrated, secure, on-demand service-driven, privacy-compliant and sustainable European e-infrastructure for processing, analysis and information-mining of metabolomics data. The project has been designed to enable maximum benefit from research by making data as accessible as possible to the research community, while protecting the interests of participants from whom the data originate with regard to Ethical, Legal and Social Implications (ELSI) and within the scope of their consent. These Terms of Use reflects PhenoMeNal's commitment to provide this service and impose no additional constraints on the use and transfer of the contributed data than those provided by the data owner.

- All users have an obligation of confidentiality and must conform to data protection principles to ensure that data is processed in compliance with the legal and ethical requirements.

- The data owners must ensure that they have sought and obtained, where necessary, all appropriate approvals, ethical and legal, for the data collected.
- For animal data, the data owner must ensure that national guidelines for their welfare and care during the collection of data has been followed.
- PhenoMeNal does not guarantee the accuracy of any provided data.
- PhenoMeNal has implemented appropriate technical and organisational measures to ensure a level of security which we deem appropriate, taking into account the sensitivity of data we handle. However, the data provider holds sole responsibility for the usage and distribution of data.
- Computing of personal and sensitive data on PhenoMeNal infrastructure should be run internally by the users on their secure cloud infrastructures under appropriate firewalls. PhenoMeNal will not hold any liability for any loss or damage to data.
- While we will retain our commitment to privacy of sensitive data, we reserve the right to update these Terms of Use at any time. When alterations are inevitable, we will attempt to give reasonable notice of any changes by placing a notice on our website, but you may wish to check each time you use the website. The date of the most recent revision will appear on this, the 'PhenoMeNal's Terms of Use' page. If you do not agree to these changes, please do not continue to use our services. We will also make available an archived copy of the previous Terms of Use for comparison.
- Any questions or comments concerning these Terms of Use can be addressed to: PhenoMeNal-help@ebi.ac.uk