

Deliverable 3.1.2

Project ID	654241
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	3
Work Package Title	Dissemination and Outreach
Deliverable Title	D3.1.2 Report of the PhenoMeNal Stakeholder Meeting
Delivery Date	M24
Work Package leader	UoB
Contributing Partners	EMBL-EBI, UoB
Authors	Ralf Weber, Mark Viant, Namrata Kale, Merlijn van Rijswijk, Daniel Schober, Philippe Rocca-Serra, Pedro de Atauri, Marta Cascante, Pablo Moreno, Etienne Thevenot, Christoph Ruttkies, Michael van Vliet, Kenneth Haug.
Abstract: The deliverable is a report on the outcomes of the second annual PhenoMeNal Joint Stakeholder & SAB Meeting held on 15 th June 2017 at EMBL-EBI.	



PhenoMeNal

Contents

1	Executive Summary	3
2	Contribution towards Project Objectives	3
3	Detailed Report on Deliverable	4
3.1	Introduction	4
3.2	Thematic sessions - Feedback and Discussions	5
3.3	Recommendations Scientific Advisory Board, and associated action points and activities.	6
4	Annex 1 – Agenda from the stakeholder meeting	10

1 Executive Summary

PhenoMeNal provides well-tested and portable metabolomics workflows in the cloud. As noted previously, an effective outreach strategy to inform stakeholders about the project progress and new opportunities is an important element to assess and foster project impact. This deliverable is a report on the stakeholder engagement by the consortium and is built on the previous activities under this work package (WP3).

The PhenoMeNal consortium hosted its 2nd annual stakeholder meeting on 15th June 2017 at EMBL-EBI, UK. The meeting, a joint event with the annual scientific advisory board (SAB) meeting, gathered 36 participants from across the stakeholder community including members of the consortium. The primary goals were to inform the stakeholders on the recent developments in the project and gain an understanding of how to further work with the user community to foster impactful research.

The stakeholders, representing clinical, academic and industry users, including the advisory panel were positively impressed by the overall e-infrastructure, the services provided including available tools and integration of the data standards. The meeting fully met its objectives with the outputs of a constructive feedback and recommendations from the advisory panel. The consortium used these results to define its future dissemination and outreach strategies in achieving its goals

2 Contribution towards Project Objectives

The activities described in the present report contribute to the achievement of the following PhenoMeNal project objectives:

Objective 3.1: Establish extra-consortium links with other e-infrastructures, standards communities, publishers and funders;

Objective 3.2: Raise awareness for standards, services and tools provided by the PhenoMeNal GRID;

Objective 3.4: Arrange PhenoMeNal stakeholder meetings.

3 Detailed Report on Deliverable

3.1 Introduction

The stakeholder meeting, a one-day event organised by the PhenoMeNal coordinator EMBL-EBI, gathered 11 stakeholders from across diverse communities including:

- SME's
- Clinicians
- Metabolomics experts
- Other e-infrastructures
- Health data experts

The specific objectives laid for the meeting were:

- Demonstrating, using & testing PhenoMeNal VRE for data processing and analysis across cloud services;
- Discussing data sharing, privacy and ethics issues for the case of human data.

To fulfil its objectives, the meeting was structured into *thematic sessions* that included:

- Lightning talks on the reuse of health data, data privacy and clinical metabolomics presented by the expert panel
- Demonstrations on PhenoMeNal VRE deployment and tools
- Demonstrations on data analysis using PhenoMeNal workflows

The consortium took the opportunity to use these talks to initiate discussion and receive feedback and advice from the participants. For a detailed agenda see Annex 1. A summary of the issues identified, recommendations and outcomes of the meeting is outlined below:

3.2 Thematic sessions - Feedback and Discussions

- Lightning talks on the reuse of health data, data privacy and clinical metabolomics were presented by the expert panel. **Dipak Kalra** (President of EuroRec) - "Electronic" healthcare is becoming more and more important. The European Institute for Innovation through Health Data (i~HD) has been formed as one of the key sustainable entities arising from the Electronic Health Records. i~HD promotes best practices for information governance, which is an important link to PhenoMeNal. He highlighted the importance of standards and good practice (e.g. European Network of Excellence for Hospitals and i~HD Quality Seal for Research Platforms). **Andres Floto** (Professor of Respiratory Biology, University of Cambridge) and **Kim Kultima** (Caramba-eng - Department of Medical Sciences - Uppsala University) both presented (metabolomics) studies that are potential use cases for PhenoMeNal (i.e. data-processing and analysis workflows).
- A short technical overview of the project components and their interplay was given by **Pablo Moreno** (15 minutes). Stakeholders and SAB members were impressed by the ability of PhenoMeNal to encapsulate that much complexity and expose it through the portal. Stakeholders and SAB members from Cambridge University showed special interest in trying the infrastructure on the university's OpenStack installation, which was followed on later during coffee breaks. Main questions were regarding the integration of own data sources into the analysis pipelines, and what did data packing in this context meant exactly.
- The Virtual Research Environment was demonstrated by **Sijin He** (EMBL-EBI). Stakeholders and SAB members were impressed by the ease of use when it came to making cloud deployments of PhenoMeNal, which was demonstrated. There were some comments regarding the type of input given to users in terms of the progress of the deployment.
- The fluxomics pipeline was shown through a video (available in the documentation) and through a demonstration of a newer step of visualisation of metabolic networks with fluxes values derived from the analysis.
- We should explore different approaches and methods to transfer large amounts of data (>1TB) into the phenomenal e-infrastructure¹.

¹ Data Streaming for Metabolomics: Accelerating Data Processing and Analysis from Days to Minutes, DOI: 10.1021/acs.analchem.6b03890



PhenoMeNal

- Our duty is to take reasonable precautions; make people aware that they are responsible for their deployments and the data that they upload. We cannot ensure that people will use, for instance, non-anonymised data. This must be made clear to users, very prominently (e.g. on deployment pages).
- We should define cost based models on certain use cases (e.g. cluster size, amount of disk space used, cpu time, etc). This should be shown in the PhenoMeNal Portal when deploying. How cost-effective is PhenoMeNal's cloud model.
- Need for UX Workshops (Oliver Kohlbacher expressing doubts on 'Bring your own Data')
- Applicability of WFs for different data sets: C. Ruttkies (IPB) tested different MetaboLights (MTBLS) datasets by running the univariate and multivariate statistical workflow that could be successfully applied to 24 MTBLS studies of which 19 showed statistically meaningful results (D9.3)

3.3 Recommendations Scientific Advisory Board, and associated action points and activities.

1. Focus on a specific problem or challenge that the e-infrastructure addresses rather than providing a general infrastructure for any metabolomics task.

Action points:

- Develop and present up to three showcased applications that would not have been possible otherwise. Following discussions, we are proposing that these will be:
 - The univariate and multivariate statistical workflow was applied to all available MTBLS studies (276, MetaboLights²) with an API (D9.3) accessing the PhenoMeNal e-infrastructure. At the moment around 10% of these studies meets the criteria for automated analysis by the current version of the workflow, and were processed successfully. Ongoing developments to increase the number of analysed datasets include the modification of the workflow (e.g. for quantitative factors) as well as the curation of the MTBLS repository (e.g. some features are reported in datasets although their intensity is 0 in all samples).
 - The BATMAN tool (for estimating metabolite concentrations from Nuclear Magnetic Resonance spectral data using a specialised MCMC

² <http://www.ebi.ac.uk/metabolights/>



PhenoMeNal

algorithm) has been dockerised as part of PhenoMeNal. The team at Imperial College London (ICL) have worked on running several instances of this tool in parallel to process the MESA³ data set. This parallelisation is based on the PhenoMeNal infrastructure using the KubeNow Stack. The experiments have been successfully run on a powerful local server and they are currently running on a large cluster (~1000 CPUs) on the EMBASSY cloud.

- o The MTBLS233 workflow previously implemented as R nodes in the KNIME platform. The objective of this showcase was to demonstrate the scalability and reproducibility of an existing workflow (applied on the MetaboLights study MTBLS233⁴), re-implemented in Docker containers and run by the Luigi workflow system.
 - o Start-to-end mass spectrometry data processing workflow. A Galaxy workflow was designed to process raw mass spectrometry data and perform downstream analysis. The workflow consists of 27 nodes and is capable of performing quantification and identification as well as statistical analysis. The pipeline was run on mass spectrometry data from multiple sclerosis patients and resulted to a number of metabolites altered between different multiple sclerosis subtype. All the resulting metabolites were part of pathways previously shown to be related to multiple sclerosis.
 - o The workflow for fluxomics. The current version allows for analysis of flux distribution from ¹³C-propagation data assuming steady state. In order to extend the applicability of the workflow a new tool for dynamical analysis is ready. Also, the different tools and the formats are being discussed and upgraded looking for the most general applicability.
- Execute the Outreach plan, as described in D2.3 (M22) on how to attract and service key users (champions or ambassadors) to use the PhenoMeNal infrastructure for their research. Following discussions, we are proposing that these will be:
 - o Up to three webinars will be presented to introduce and demonstrate PhenoMeNal. A recording of the webinar will be made available online to increase coverage. A number of face-to-face workshops to introduce and demonstrate the PhenoMeNal e-infrastructure. Standardised training material will be developed and distributed across

³ <https://www.mesa-nhlbi.org/>

⁴ <http://www.ebi.ac.uk/metabolights/MTBLS233>



PhenoMeNal

trainers/demonstrators to assure a high standard and consistency across demonstrations/training.

- o Introduce and demonstrate PhenoMeNal to potential users, champions and ambassadors, and contributors that have shown an interest. A dedicated trainer or demonstrator will visit the "customer's" site to provide a 1/2 day of face-to-face (or online) training to (further) introduce and demonstrate PhenoMeNal.
 - o Customers that show further interest will be contacted again to plan face-to-face (or online) a more extensive training session, a 1-day follow-up training will be provided at the "customer's" site to let the customer process and analyse their own datasets using Phenomenal.
 - o Organise UX session(s) where potential users, who are not yet familiar with PhenoMeNal, can:
 - be introduced to our services;
 - do hands-on exercises using example data;
 - and can be observed by UX specialist(s) to collect insights about possible usability issues.
- Focus on the biologists and clinicians, not bioinformaticians.
 - Develop a roadmap for a follow-up project.
2. Improve and increase the support and guidance in how to use the tools and workflows within the infrastructure.
- Action points:
- Organise a UX meeting where users are tracked and their feedback is captured in a formal way to deliver intelligible hints for usability improvements and software ergonomics (see point 1 - Outreach plan).
 - Develop tutorials and clear guidance on how to use the different services.
 - Organise the tools in clear categories and make them easy to find.

3. Implement Best Practices and Standardisation throughout the e-infrastructure where possible.

Action points:

- Liaison with FAIR Data Group and FairSharing.org to ensure FAIR compliance via good standards coverage in PhenoMeNal workflow area
- **Improve quality of metabolomics study metadata descriptions.** Overall, several lines of evidence indicate shortcomings in the way deposited studies are annotated. It ranges from lack of essential annotations, something which is context dependent. For instance, basic patient information about gender, body mass index or age are all too often missing, thus significantly limiting reuse potential and dropping on the scale of FAIRness. UOXF is



PhenoMeNal

demonstrating the need to regularise the reporting of study independent variables using ISA syntax. UOXF has also issued recommendations for representing QC runs as well as ELSI related information in ISA format.

- **Build standard compliant tools for supporting QC runs:** On the issue of QC and study design, the organisation of a meeting specifically dedicated to taking into consideration the tools needed to support definition and tracking of the conditions defining how experimental samples and controls are run has been welcomed. This meeting will be organised by WP8 to feed in functional specifications for the ISA API. The aim is to make it easier for wet lab scientists to plan their experimental runs and format study information using the PhenoMeNal vetted metadata standard, ISA.
- **Enhance NMR data standard envelop to support 2D applications.** This step is a natural evolution of the nmrML standard, currently limited to 1D applications. With the increase of 2D NMR techniques and refinements to signal processing methods, it will become necessary to develop the necessary support, in terms of data structure but also in documentation, reference implementation guidelines and training.
- **Deliver a data standard compliant VRE.** It is necessary to demonstrate the PhenoMeNal app library consumes and produces standard compliant datasets. To this end, it is essential that PhenoMeNal workflows clearly markup inputs and output data files with resources such as EDAM and liaise with FAIRSharing/BioSharing.
- **Normalised descriptions of workflows:** To increase the discoverability and awareness of PhenoMeNal-produced workflows (mostly Galaxy workflows), the necessity of regularising their description was highlighted. A naming convention is therefore being considered⁵ but a tension exists between remaining meaningful while being compact.
- **Improve documentation and guidelines:** Provide users and demonstrators with “standardised” datasets showcasing information representation best practice and implementation patterns. These aspects also include regularisation of names in ISA documents (work with Birmingham and Imperial College London Phenome Centres is underway).
- Add functionalities to export workflows and associated parameters, using for instance the Common Workflow Language (CWL)⁶ currently being considered by efforts such as GA4GH⁷.

⁵ <https://github.com/phnmnl/container-galaxy-k8s-runtime/issues/51#issuecomment-318822471>

⁶ "A Gentle Introduction to the Common Workflow Language." <http://www.commonwl.org/draft-3/UserGuide.html>. Accessed 30 Aug. 2017.

⁷ "Global Alliance for Genomics and Health: Home." <http://genomicsandhealth.org/>. Accessed 30 Aug. 2017.

4 Annex 1 – Agenda from the stakeholder meeting

Agenda *PhenoMeNal Joint Stakeholder & SAB Meeting*

Date: Thu 15th June 2017

Venue: Rosalind Franklin Pavilion, Genome Campus Conference Centre

Time: 9:00 - 17:30

Objectives:

The meeting will discuss the following issues:

- Using/testing PhenoMeNal VRE for data analysis and platforms
- Data sharing, privacy and ethics issues in case of human data

9:00 - 9:10

Introduction

Official Kick-off and Welcome Address

Speaker - *Chris Steinbeck* (Co-ordinator, PhenoMeNal)

9:10 - 10:40

Session 1: Presentations

Chair: Christoph Steinbeck (University of Jena/EMBL-EBI)

Progress towards greater reuse of health data for research

Dipak Kalra (President of EuroRec)

Exploring regulation of lung inflammation through gut microbiome metabolites

Andres Floto (Professor of Respiratory Biology, University of Cambridge)

Clinical metabolomics

Kim Kultima (Uppsala University)



PhenoMeNal

10:30 - 10:45

Tea/Coffee break

10:45 - 12:15

Session 2:

Demo on PhenoMeNal deployment and tools

VRE portal tour/Deployment of VRE on various cloud providers

Pablo and Sijin (EMBL-EBI)

Data import/re-utilisation from public metabolomic data repositories

Philippe (UOXF)

12:15 - 13:30

Lunch

13:30 - 15:00

Session 3:

Demo on PhenoMeNal workflows

NMR workflow

Reza (EMBL-EBI)

Fluxomics workflow demo

Marta/Pedro And Pablo (UB & EMBL-EBI)

Open MS/XCMS workflow

Kim Kultima

Statistics workflow

Etienne Thevenot (CEA)



PhenoMeNal

15:00 - 15:15

Coffee/tea break

15:15 - 16:00

Discussion and Feedback

16:00 - 17:30

Preparation of SAB report

SAB panel only

17:30

Closing Remarks

Adjourn

19:00 - 21:00

Dinner



PhenoMeNal

List of Stakeholders/SAB attending the meeting:

	Name	Affiliation
1	Nicolas Schauer	Metabolomic Discoveries, Germany
2	Heidi Howard (TBC/online)	Centre for Research Ethics & Bioethics, Uppsala University
3	Andres Floto	University of Cambridge
4	Oliver Kohlbacher	U of Tübingen
5	Theodore Alexandrov	EMBL-Heidelberg
6	Dipak Kalra	The European Institute for Innovation through Health Data
7	Jules Griffin	University of Cambridge
8	Sabine Bahn	University of Cambridge
9	Rima (online participation)	Duke Univer. sity School of Medicine
10	David Wishart (online participation)	University of Alberta
11	Sameer Velankar	EMBL-EBI (West-Life)
12	Smarason Sigurdur	EURAC
13	Hutan Ashrafian	ICL