Deliverable 3.4.1

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	D3.4.1 Two training workshops on omics data deposition, grid processing, dissemination and access
Delivery Date	M24
Work Package leader	UoB
Contributing Partners	All
Authors	Ralf Weber, Mark Viant, Reza Salek, Daniel Schober
Abstract: This deliverable reports the workshops developed and provided for dissemination and outreach to the user community.	

Table of Contents

1 EXECUTIVE SUMMARY

Now that early releases of the PhenoMeNal e-Infrastructure have occurred, the consortium has increased its efforts to establish links with the metabolomics and wider scientific communities to form a strong user base for the services offered. By providing face-to-face training activities, it enables us to introduce the PhenoMeNal e-infrastructure to potential users as well as demonstrate its capabilities. Here we provide a summary of the face-to-face training and workshop activities that we have delivered to date as well as the consortium's extensive new plans to deliver face-to-face training in the final 12 months of the PhenoMeNal project.

2 CONTRIBUTION TOWARDS PROJECT OBJECTIVES

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

Objective 3.2: Raise awareness for standards, services and tools provided by the PhenoMeNal GRID;
Objective 3.3: Provide training for users interested in the PhenoMeNal grid

3 DETAILED REPORT ON DELIVERABLE

3.1 Face-to-face workshops delivered:

***1. Data Sharing, Standardisation and Workflow for reproducible analysis in Metabolomics*, 13th Annual International Conference of the international Metabolomics Society, June 26-29, 2017, Australia [http://metabolomics2017.org]**

Workshop description: Workflow management environments applied in metabolomics and cross-omics analyses are an essential requirement to allow standardisation of bioinformatics analysis, provide access to the metabolomics community, and produce high-quality, reproducible results in a time-effective manner: on the one hand, experimenters should be able to easily select the tools via a graphical interface, choose the parameters, run the workflow and save/share the results; on the other hand, developers should be able to integrate new tools seamlessly into the environment. There is a growing need for the international metabolomics community to understand the current availability and capability of these workflow environments, and to provide input into their ongoing development and interoperability.

- In this workshop, we discussed and demonstrated computational workflows for metabolomics data analysis within the PhenoMeNal e-infrastructure.
- We also demonstrated how emerging metabolomics data sharing platforms, such as MetaboLights and Metabolomics Workbench, aim to make use of data standards (e.g. mzML, nmrML, ISA-tab, mzTab) to promote data-sharing.
- Discussions on ontology based metadatabases of metabolomics software, databases and tools

Workshop description; Reproducing results in any science is quite challenging. In the field of metabolomics, for results to become reproducible, descriptions of an investigation in a manuscript are not sufficient. To surpass this, and increase the chance of result reproducibility, standard frameworks for data sharing and sharing of experimental data are invaluable. We now have several data sharing platforms such as MetaboLights and Metabolomics Workbench that aim to make use of such standards to promote data-sharing. In this workshop, we will discuss data sharing as well as metabolomics data formats, much of which are adopted from the efforts in the proteomics HUPO-PSI initiative. Examples and application of data formats such as mzMl and nmrML would be given. We will also present and discuss mzTab developments for metabolite identification. We will also demonstrate how emerging metabolomics data sharing platforms can promote open, accessible data sharing standards. Finally we will discuss computational workflows for metabolomics data analysis within the PhenoMeNal e-infrastructure. We hope to make the metabolomics community aware of such efforts and ideally how to get involved.

Workshop Organiser; Metabolomics Society data standard Task group and PhenoMeNal H2020

Number of participants: about 40 people attended, comprising of PhD students, postdoctoral researchers, group leaders, as well as scientists from industry.

Presenters; Reza Salek (Chair), Pablo Moreno, Saravanan Dayalan, Oliver Jones, Andrew Jones, Masanori Arita, Adam Caroll

Workshop format; traditional workshop with panel discussion

Duration; 90 minutes
20 min Overview: Data Standards and data sharing in metabolomics, current and future! - an overview
Reza Salek, EMBL-EBI, Cambridge, UK
An overview of past and current initiatives on data standards carried out by the Metabolomics Society data standards task group and the future plans on data standards in metabolomics. Why do we need it and is it important? Discussion on: Community standards, how to make it easier to disseminate metabolomics data through life science. How to support workflows for a broad range of metabolomics applications. Are metabolomics data FAIR? that is Findable, Identifiable, Accessible and Reusable

15 min - Complying with international data standards through data management software - an automated approach

Saravanan Dayalan, Metabolomics Australia, The University of Melbourne
The quantity of data generated by high throughput technologies is on constant rise. To tackle this ever-growing mountain of data, the scientific community (users and publishers) has started

to emphasis the importance of standardizing data formats and exchange methods. The two leading standards in Metabolomics are the ISA format adopted by the MetaboLights public repository and the mwTab format adopted by the Metabolomics Workbench public repository. Whilst the repositories provide tools for researchers to manually describe their projects, experiments and sample data in the standardized data formats, it would be more efficient if data management softwares were able to do so in an automated way. In this talk, we explore the different software solutions that attempt to achieve this.

20 min - “Complying with standards in a multi-user environment: the case of MassBank”

Masanori Arita (National Institute of Genetics, Japan)
We present the ongoing effort at MassBank, to which over 30 research groups contributed their spectra under different licenses and different format in past 10 years. The publication style of MassBank is unique in that we have provided an interface to view contributions (raw spectra) and even allowed retraction or edits. This policy, however, needs some update to enable smooth data exchange and remix. We emphasize the importance of separation between searchable library and spectral repository and introduce our curation strategy to provide a useful reference.

15 min -”PhenoMeNal technical overview and how to bring your tools to the framework”

Pablo Moreno, EMBL-EBI, Cambridge, UK
Computational analysis of high-dimension and high-volume metabolomics data is a complex, time-consuming process including many steps, some of which still being the focus of intense research. Workflow management environments applied in metabolomics and cross-omics analyses are therefore an essential requirement to allow standardisation of bioinformatics analysis, provide access to the metabolomics community, and produce high-quality, reproducible results in a time-effective manner: on the one hand, experimenters should be able to easily select the tools via a graphical interface, choose the parameters, run the workflow and save/share the results; on the other hand, developers should be able to integrate new tools seamlessly into the environment. A few open-source workflows have recently been applied in different environments including Galaxy-M, Workflow4Metabolomics and MetaboAnalyst. There is a growing need for the international metabolomics community to understand the current availability and capability of these workflow environments, and to provide input into their ongoing development and interoperability.

15- Ontology based metadatabase of metabolomics software, database and tools

Adam Carroll, Australian National University Canberra
Formalised ontology for data processing steps, data types and formats and the relationships between them to build a more systematic and consistent annotation style for software tools and algorithms is needed more than ever. Key relevant projects to address the need for ontology building around metabolomics data processing to complement and extend the existing EDAM and Software Ontology projects will be discussed.

10 min Discussion and questions
Final discussion and question on the topics presented.

3.2 Workshops planned for the next 12 months, including those just outside of the time window of D3.4.1:

***1. CloudMET 2017: Cloud-based Metabolomics Data Analysis and Collaboration,* 11th - 15th September 2017, Italy [http://cloudmet2017.crs4.it]**
Workshop description: In this School (ca. 25) students will have the opportunity to learn about current topics in metabolomics, with a slant on the integration of cloud computing technologies where they are beneficial to the effectiveness and efficiency of research and analysis work. Top-level lecturers in the field will provide their insight and will be available for the entire duration of the school, with ample opportunity for interaction with the students. Importantly, the School will also include practical sessions where students can put their new knowledge into practice *under the guidance of tutors (from up to six different PhenoMeNal work packages) and run analyses using the new PhenoMeNal cloud-based metabolomics platform*.

2. *Metabolomics workflows*, 30th October - 2nd November, European Bioinformatics Institute (EMBL-EBI), United Kingdom
[https://www.ebi.ac.uk/training/events/2017/metabolomics-workflows]
Workshop description: This course will provide an introduction to metabolomics data analysis using publicly available software and tools. Participants will become familiar with the current state of data sharing and data standards in metabolomics, particularly through using the EMBL-EBI's MetaboLights repository. In addition, participants will have a hands-on session using the PhenoMeNal Compute infrastructure. There will be a large practical component, where participants will participate in hands-on tutorials on data submission and using a workflow-based approach and compute infrastructure for data analysis. These tutorials will take place under the guidance of the lecturers and teaching assistants

3. Satellite training course to MetaboMeeting2017, December 2017, Birmingham, UK [http://metabomeeting2017.thempf.org]
Workshop description: Course content will be developed in November 2017, building on the CloudMET training material.

4. *Cloud-based Metabolomics Data Analysis using PhenoMeNal Compute infrastructure,* The 13th International Conference of the Metabolomics Society, June 25th – 29th, Seattle, Washington, United States.
[http://metabolomics2018.org]
Workshop description: Workshop has been proposed to the International Organising Committee, we are awaiting their feedback..

4 Delivery and Schedule

The deliverable is submitted on time. Due to slight delays in publishing the first stable release of the Phenomenal e-infrastructure (August 2017) we have shifted the temporal distribution of the workshops slightly, such that within this deliverable period we have reduced these activities from two to one; however, for next deliverable period we are now planning to deliver at least three more major workshops in the next 12 months (meeting the overall EU project deliverable) and ideally delivering four workshops in the final year of the project, as listed above.