



PhenoMeNal

Deliverable 1.4.1

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	1
Work Package Title	Management
Deliverable Title	D1.4.1 Biannual progress Report
Delivery Date	M6
Work Package leader	EMBL-EBI
Contributing Partners	EMBL-EBI, ICL, IPB, UB, UoB, CIRMMMP, UL, UOXF, SIB, UU, BBMRI, CEA, INRA
Authors	Christoph Steinbeck, Namrata Kale, Kenneth Haug, Pablo Moreno, David Johnson, Tim Ebbels, Ola Spjuth, Daniel Schober, Merlijn van Rijswijk, Kristian Peters, Steffen Neumann, Ulrich Guenther, Michael van Vliet, Robert Glen
Abstract: This deliverable is a comprehensive report of the PhenoMeNal consortium's activities and performance towards meeting the objectives and goals of the project from M1-M6.	



Version	Revised
History of Changes: <ol style="list-style-type: none">1. Tracking of progress based on objectives2. Detailed description of current and future work plans3. Reporting of deviations or delay4. Inclusion of performance metrics	
Comments: Details on the resource consumption and costs will be provided during the periodic reporting (M18)	



Contents

1. Executive Summary	4
2. Objectives, work progress and achievements, project management	5
2.1. Project objectives for the period	5
2.2. Work Progress and Achievements of PhenoMeNal until M6 (September 2015 - February 2016)	6
2.2.1. WP2- Sustainability	6
2.2.2. WP3 – Dissemination and Outreach.....	9
2.2.3. WP4- Interfacing with Biomedical European Infrastructures	15
2.2.4. WP5 – Operations and Maintenance of PhenoMeNal grid/cloud e-infrastructure	18
2.2.5. WP6 - PhenoMeNal Virtual Research Community Gateway	21
2.2.6. WP7 – Privacy and Ethics.....	23
2.2.7. WP8 - Data Provenance, Compliance and Integrity.....	25
2.2.8. WP9 - Tools, workflows, audit and data management.....	28
2.2.9. WP1- management activities.....	30
3. Conclusions.....	34



1. Executive Summary

Description of project context and Objectives

The PhenoMeNal project aims to establish a comprehensive and standardised e-infrastructure for data processing, analysing and information mining of extremely large medical metabolic phenotype data thereby being one of the key enabling infrastructures addressing the H2020 Societal Challenges in Health, Demographic change and Wellbeing.

It will focus on integrating existing open source tools and methods for the management, dissemination and computational analysis of the very large datasets of human metabolic phenotyping and genomic data into a secure and sustainable e-infrastructure. In addition, PhenoMeNal will work to improvise and scale-up the tools to cope up with such large datasets while complying with the privacy-protection methods that allow working with highly sensitive data. The project will also work towards establishing a watertight audit trail for the processing of human metabolic phenotyping data from the raw data acquisition all the way to the generation of high-level biomedical insights.

PhenoMeNal aims to operate and consolidate the e-infrastructure based on the existing internal and external high performance computing (HPC) and the grid resources such as European Grid Initiative (EGI, <http://www.egi.eu>), and to extend it to worldwide grid infrastructures. In order to maximise the impact, it will foster worldwide adoption via dissemination, networking and training activities. A model to ensure sustainability of PhenoMeNal beyond its initial funding period would also be developed and implemented.

Description of overall work performed

The PhenoMeNal work plan is organised in 9 Work Packages (WP) designed to meet the general project activities as well as the detailed objectives outlined under the individual WPs. During the first six months, the consortium work was focussed on activities to meet these objectives in line with the action points that were the outcome of the initial kick off meeting held at EMBL-EBI in September (see D1.3 Minutes of meeting for details).

As part of the effective management of the project under WP1, activities including online meetings between groups, task distribution and coordination between the partners using Pivotal Tracker (www.pivotaltracker.com) and a comprehensive and systematic documentation system to record all the activities of the consortium via Google docs stored in a shared Google drive were achieved.

Regarding the sustainability and interfacing with other e-infrastructures, a list of majority of relevant e-infrastructures was prepared and workshops/meetings were conducted to interact with them. These workshops have concentrated a significant part of the consortium's effort and resources to maximise the impact of the project by communication and interfacing with existing European Strategy Forum on Research Infrastructures (ESFRI) and Innovative Medicines Initiative (IMI) infrastructures. As



for the dissemination and outreach of the project to the broader community, the project website (<http://phenomenal-h2020.eu/home/>) was developed, dissemination materials in form of newsletters and press releases were produced, and the project was presented in several workshops, training events and conferences throughout Europe and UK.

As part of the service activities, several technological solutions and frameworks were evaluated for applicability in PhenoMeNal. To this end, a Jenkins (<https://jenkins.io>) server was established for continuous integration. An effort towards the PhenoMeNal Virtual Research Community (VRC) was initiated by identification of the potential users and requirements analysis by conduction of an initial User Experience (UX) workshop.

Resulting from the project's major consideration for using sensitive data in compliance with the Ethical, Legal and Social Implications (ELSI) regulations, workshops were held to understand and devise best practices in handling sensitive human data taking in compliance with the National policies.

As for the data provenance, compliance and integrity, a survey on 'Standards, Compliance and Data Integrity' was commissioned in collaboration with the PhenoMeNal partners. In addition, real world use cases that have distinct practice-driven demands on data storage/federation and privacy were also identified.

As such all the tasks this period were carried out as planned without any deviations or delays in terms of the deliverables submitted and milestones achieved.

2. Objectives, work progress and achievements, project management

2.1. Project objectives for the period

During the first six months, the consortium has successfully worked towards the following the general project objectives:

- To integrate existing open source tools and methods for the management, dissemination and computational analysis of very large datasets of human metabolic phenotyping and genomic data into a secure and sustainable e-infrastructure.
- To operate and consolidate the PhenoMeNal e-infrastructure based on existing internal and external HPC and grid resources, including the EGI, and to extend it to worldwide grid infrastructures.
- To improve and scale up tools within the infrastructure to cope with very large datasets
- To establish privacy-protection methods that allow working with highly sensitive molecular phenotype data
- To foster the worldwide adoption of PhenoMeNal through a wide range of outreach, dissemination, networking and training activities.
- To develop a model to ensure sustainability of PhenoMeNal.

Details on how these objectives were pursued are discussed in detail under the description of progress work for each WP.



The main deliverables and milestones achieved during this period were:

- D1.1 Project plan
- D1.2 Consortium agreement
- D1.3 Minutes of kick off meeting
- D1.4.1 Biannual progress report
- D1.5.1 Data management plan
- D2.1 Report on mapping of e-infrastructures, users, investments for supporting policy developments in the field of metabolomics, biomarkers and biobanks
- D6.1 User Experience document on the VRC design guide
- D7.1.1 Workshop on best practices in handling sensitive human data taking into account National and Institutional legal policies
- D7.2 Report on the policies and procedures for sensitive data management
- D9.1 Report on the existing software tools, workflows and analytical pipelines initially supported in the PhenoMeNal grid.
- MS1.1 PhenoMeNal Project Plan
- MS1.2 Initial release of Data management plan
- MS6.1 Release of VRC design guide
- MD7.1 Agreement on policies and procedures for sensitive human data management.

2.2. Work Progress and Achievements of PhenoMeNal until M6 (September 2015 - February 2016)

2.2.1. WP2- Sustainability

The main achievements were:

- Identification and mapping of relevant e-infrastructures, users, investments in the field of metabolomics, biomarkers and biobanks.
- Establishment of relations with relevant ESFRIs.
- Submission of deliverable D2.1 (Report on mapping of e-infrastructures, users, investments for supporting policy developments in the field of metabolomics, biomarkers and biobanks (M6) on time.

Work plan

Following the kick off meeting, dedicated hangouts regarding the strategies and work distribution under WP2 were conducted. The overall concept of sustainability and measures to sustain what is developed under PhenoMeNal were discussed in great detail. A three-tier approach towards the sustainability was proposed:



I. Building strong relationships with ESFRIs, and international & national e-infrastructures:

In general infrastructure initiatives have a longer and more sustainable (funding) horizon, likely in general to continue beyond the initial 3-year funding period of PhenoMeNal. Therefore we want to build strong relationships not only with ESFRIs, but also with international & national e-infrastructures and organisations like the European Bioinformatics Institute (EMBL-EBI) and the European Genome-Phenome Archive (EGA), to ensure that they will support and if possible continue to sustain the development of PhenoMeNal, also after the initial 3-year funding period. As part of this:

- An ESFRI workshop, together with WP4 (Interfacing with Biomedical European Infrastructures) and relevant ESFRI infrastructures was organised on February 5th 2016. In this workshop the needs of the different ESFRIs were discussed. A detailed report of this workshop is added as Appendix 1 of the D2.1 report.
- Efforts have been made to involve organisations like EGA and representatives of BIOMEDBRIDGES/CORBEL to participate in PhenoMeNal workshops. They participated for instance in the Ethical Legal and Social Implications (ELSI) workshops/meetings hosted by Imperial College of Science, Technology and Medicine (ICL, WP7 lead), which were held on November 20th 2015 and January 27th 2016 respectively.
- We have also mapped the relevant international and national (e-infrastructures). This report is added as Appendix 2 of the D2.1 report. The overview indicates that there are many e-infrastructures in the different countries and that individual partners of PhenoMeNal have strong ties with the national e-infrastructures. The design choice of PhenoMeNal to establish a virtual research community ensures that PhenoMeNal can be easily integrated into these e-infrastructures and that no duplications occur.

II. Building strong relationships with users:

We will take initiatives to reach out to academic and clinical users and companies at a very early stage of the PhenoMeNal project, to ensure that PhenoMeNal is widely accepted and used. We envisage that once PhenoMeNal is thoroughly used, it is far more likely that users will continue to sustain the development of PhenoMeNal also after the initial 3-year funding period. As part of this:

- On 19th November 2015, hosted by EMBL-EBI (WP6 lead), a workshop was organised to ensure a **user centric design and usability** of the Phenomenal Virtual Research Community, which will be the central access point for all users. Twenty future VRC users from the consortia and potential users from the metabolomics community participated in a hands-on workshop to make sure the



future VRC is designed to suite the users. This is described in more detail in the M3 report of. D6.1 User Experience document on VRC Design Guide.

III. Establishing a PhenoMeNal Foundation and/or evaluate pro's and cons of collaboration with eTRIKS and the Phenotype foundation and tranSMART foundation

For the long-term sustainability of large open source software initiatives, sometimes foundations are being set up. These foundations unite users, and drive the further development of the software. In the period M6-12 we will have meetings with parties having set up the tranSMART (<http://transmartfoundation.org>), Phenotype foundation (<http://phenotypefoundation.org>) and eTRIKS (<https://www.etriks.org>) to discuss pro's and con's of setting up a Phenomenal Foundation. Furthermore we will discuss with more general research foundations in the field of metabolomics about the pros and cons of establishing a dedicated software foundation versus integrating Phenomenal into a more general research foundation. Based on this we will develop a more detailed sustainability strategy, which we will report in the M12 progress report.

Future plans:

- We will organise **Industry workshops**. On March 14th-15th we will discuss PhenoMeNal at the quarterly meeting of the EMBL-EBI Industry Consortium meeting (<https://www.ebi.ac.uk/industry>). In the EMBL-EBI Industry group many European Pharma and Food companies are represented. In addition, we will organise a workshop aimed at instrument vendors and software companies, during the Metabolomics Society meeting in June in Dublin. They will be consulted for specifications and challenges relating to grid-computing establishment from industry point of view. A Memorandum of Understanding (MoU) will be signed in year 3 with the willing industry partners in the later stages of the project, who wish to sustain PhenoMeNal in future, recognising its scientific utility.
- We have scheduled a **Clinical users workshop** to define the state-of-the-art in Metabolic Phenotyping in the clinic on May 27th 2016 in Barcelona.
- We will organise a **Stakeholder meeting** during the first annual Consortium meeting, which will take place on June 14th as part of the annual PhenoMeNal consortium meeting.
- We will organise a **Research users workshop** during the annual (global) Metabolomics Society conference, which will be held from June 27-30 in Dublin (Ireland).



2.2.2. WP3 – Dissemination and Outreach

During this period, the main achievements were:

- Dissemination through presentations, posters and workshops
- PhenoMeNal website
- Participation in the production of the eInfra booklet for ICRI2016.

These activities correspond to the consortium's initiative to meet the project objective of fostering worldwide adoption of PhenoMeNal through dissemination.

Work Plan

In November 2015, the project partners selected a logo for the project:



Figure 1. PhenoMeNal logo.

The consortium used channels such as scientific publications, workshops and presentations at metabolomics conferences for dissemination and outreach.

Dissemination of PhenoMeNal in terms of workshops, meetings, conference and presentations for PhenoMeNal partners

European Molecular Biology Laboratory (EMBL-EBI):

- *Conference:* e-Infrastructures & RDA for data intensive science - Pre-RDA plenary workshops 22 September 2015, CNAM Paris, France (<https://rd-alliance.org/plenary-meetings/sixth-plenary/programme/e-infrastructures-rda-data-intensive-science/service>)
 - Presentation: Phenomenal, an e-Infrastructure for Clinical Metabolomics Data (Kenneth Haug)



- **Poster presentations:** Metabomeeting 2015 (<https://selectbiosciences.com/conferences/index.aspx?conf=Metabo2015>, 7th - 9th December):
 - Title: PhenoMeNal Large Scale Computing for Medical Metabolic Phenotyping Data (Pablo Moreno)
- **Workshops:** BioMedBridges Data sharing for Advancing Health (<http://www.biomedbridges.eu/workshop-data-sharing-advancing-health>, 2015)
 - Presentation: Sharing data from clinical and medical research (Christoph Steinbeck)

Imperial College of Science, Technology and Medicine (ICL):

- **Short courses:**
 - “Metabolic Phenotyping in Disease Diagnosis and Personalised Health Care”, 26th-30th October 2015, Imperial College London (Tim Ebbels, lectures)
 - “Hands-on data analysis for metabolic profiling”, 3rd-6th November 2015, Imperial College London (Tim Ebbels, course lead).
 - “Hands-on NMR for metabolic profiling”, January 2016, Imperial College London (Tim Ebbels - lectures)
 - “Hands-on NMR for metabolic profiling” 23rd-26th February 2016, Imperial College London (Tim Ebbels, course lead)

Leibniz Institute of Plant Biochemistry (IPB):

- **Meetings:**

Workflows and infrastructure for computational mass spectrometry also in environmental research (Steffen Neumann, Eawag, CH, 4th-6th November 2015).
- **Workshops:**
 - RSC Workshop “Mass spectrometry imaging – challenges and opportunities for next-generation capabilities”, Kavli Royal Society Centre, <https://royalsociety.org/events/2015/11/mass-spectrometry/> (Invited presentation, Steffen Neumann, 23rd-24th November 2015)
- **Presentations:**
 - Dagstuhl Seminar “Computational Metabolomics”, Dagstuhl, Germany (Steffen Neumann, 30th November-4th November 2015)
 - deNBI center Bochum (Steffen Neumann, 25th January 2016. deNBI is a german network on bioinformatics with strong workflow and service aspects)
 - Standards support, Bruker Daltonics, Bremen, Germany (Steffen Neumann, 26th January 2016)



Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP):

- **Meetings:**
 - The Metabolomics Innovation Center (TMIC, Edmonton, Alberta) and Ann Arbor, Michigan. (Claudio Luchinat, 4th-11th November, 2015. to establish links between PhenoMeNal and EXCEMET (<http://www.excemet.org>) on one side and the Michigan Infrastructure on the other.

University of Leiden (UL):

- **Meetings:**
- DTL programmers meeting
(https://wiki.dtls.nl/index.php/Agenda_DTL_Programmers_Meeting_2016/01/22
Michael van Vliet, 22nd January 2016)

Uppsala University (UU):

- **Meetings:**
 - PhenoMeNal discussions, IPB, WP9 (Steffen Neumann in Uppsala, 28th September, 2015)

Commissariat à l'énergie atomique et aux énergies alternatives (CEA):

- **Conference:** Metabomeeting 2015 (7th - 9th December, Cambridge, UK)
 - Presentation: The Workflow4Metabolomics computational infrastructure: meeting the workflow challenge (Etienne Thevenot)
- **Meetings:** European Bioconductor Developers' Meeting (<https://sites.google.com/site/eurobioc2015/timetable>, 8th December, Cambridge, UK)
 - 'The statisticians' and experimenters' needs for reproducible workflows with Bioconductor and Galaxy: the example of the *ropls* and *biosigner* package integration into the Workflow4metabolomics computational infrastructure (Etienne Thévenot)

Institut National de la Recherche Agronomique (INRA):

- **Conference:** Metabomeeting 2015 (7th - 9th December)
 - Presentation: MetExplore and Genome-Scale metabolic networks. (Fabien Jourdan)
- **Training:**
 - Hosting Karl Burgess (head of University of Glasgow metabolomics facility) for a week (February 8th to 12th 2016) to work on ab initio network reconstruction based on high resolution LC-MS data



University of Barcelona (UB):

- *Conference*: 4th Conference on Constraint-Based Reconstruction and Analysis (COBRA 2015) (<http://www.aiche.org/sbe/conferences/conference-on-constraint-based-reconstruction-and-analysis-cobra/2015>, 16th-18th September 2015)
 - *Poster Presentation*: A Dynamic Model of Metabolism that Integrates ^{13}C Isotopomer Data (Carles Foguet, Silvia Marin, Vitaly Selivanov, Eric Fanchon, Pedro de Atauri and Marta Cascante)
- *Training course*: EMBO Practical Course on Metabolomics Bioinformatics for Life Scientists, Hinxton, Cambridgeshire, UK.
- Metabolomics flux introduction and Metabolomics flux and genome scale sessions (Dr. Marta Cascante and Dr. Igor Marin de Mas 19th February 2016)

Publications, News and articles:

- Rocca-Serra, Philippe et al. "Data standards can boost metabolomics research, and if there is a will, there is a way." *Metabolomics* 12.1 (2016): 1-13.
- Ruttkies, Christoph et al. "MetFrag relaunched: incorporating strategies beyond in silico fragmentation." *Journal of Cheminformatics* 8.1 (2016): 1-16.
- "Data standards can boost metabolomics research, and if there is a will, there is a way", P. Rocca-Serra et al, *Metabolomics*. 2016;12(1):14. Epub 2015 <http://www.ncbi.nlm.nih.gov/pubmed/26612985>
- Benjamin Merlet, Nils Paulhe, Florence Vinson, Clément Frainay, Maxime Chazalviel, Nathalie Poupin, Yoann Gloaguen, Franck Giacomoni and Fabien Jourdan. "Computational solution to automatically map metabolite libraries in the context of genome scale metabolic networks". *Frontiers in Molecular Biosciences, section Metabolomics*. (2016).
- EMBL-EBI Press Release: PhenoMeNal: a gateway to personalised medicine (<http://www.ebi.ac.uk/about/news/press-releases/phenomenal-clinical-metabolomics>)
- Toxalim (INRA) Press Release: The european project PhenoMeNal (Horizon 2020) launched on September 1st, 2015 (https://www6.toulouse.inra.fr/toxalim_eng/Home/News/Phenomenal)
- Newsletters:
 - Life Science Sweden (Swedish newspaper only): Tar fram infrastruktur för metabolomik (<http://www.lifesciencesweden.se/forskning/tar-fram-infrastruktur-for-metabolomik/>, September 2015)
 - EGI newsletter: PhenoMeNal: towards an e-Infrastructure for the pheno- and genotyping data (http://www.egi.eu/news-and-media/newsletters/Inspired_Issue_21/PhenoMeNal.html, Issue 21, October 2015)



- MetaboNews: PhenoMeNal—An e-infrastructure for analysis of metabolic phenotype data (http://www.metabonews.ca/Jan2016/MetaboNews_Jan2016.htm#spotlight, January 2016) Spanish Society of Biochemistry and Molecular Biology (SEBBM): Marta Cascante - Febrero 2016: La flujómica en el estudio del cáncer (<http://www.sebbm.es/web/es/divulgacion/acercate-nuestros-cientificos/1333-marta-cascante-febrero-2016-la-flujomica-en-el-estudio-del-cancer>)

Social media:

- Twitter - #PhnmIH2020
- Facebook- <https://www.facebook.com/PhenoMeNal-1650363338550686/?ref=hl>
- Broadcast of Workflows4Metabolomics Demonstration on YouTube at <https://www.youtube.com/watch?v=f2WoAZ24Gjo>
- Post on CompMS @ Dublin: <https://plus.google.com/+SteffenNeumann/posts/EgF1BoMuKEV>

User groups:

- IPB: Several postings in www.metabolomics-forum.com
- UL: <http://www.dtls.nl/phenomenal-project-to-build-an-e-infrastructure-for-clinical-metabolomics-data/>
- EMBL-EBI:
 - PhenoMeNal-Compute: <https://groups.google.com/forum/#!forum/phenomenal-compute>

Websites:

- PhenoMeNal website with content management system is now available: <http://phenomenal-h2020.eu/home/>

PhenoMeNa (Phenome and Metabolome aNalysis)
noun /fiːnɒmɪn(ə)l/
 It is a comprehensive and standardised **e-infrastructure** that will support the data processing and analysis pipelines for molecular phenotype data generated by **metabolomics** applications. PhenoMeNa will provide services enabling computation and analysis to improve the understanding of the causes and mechanisms underlying health, healthy ageing and diseases.

LATEST NEWS

Convert your MS raw files into Open Data formats: In the cloud and anywhere

By Namrata • No Comments • What's new • All, Technical, Tools

One of the main routes to mzML-formatted data is using Open Source converter msconvert. It can convert to mzML from most of the vendor formats. In PhenoMeNa, we have now succeeded to package msconvert in a Docker container and execute it in the WINE

Tweets by @PhnmIH2020

PhenoMeNa-H2020 Retweeted

Ola Spjuth
@ola_spjuth

Our @PhnmIH2020 partners at CARAMBA platform for clinical #metabolomics have a new website: caramba.clinic

- CARAMBA - Clinical Analsi...

Figure 2. Screen shot of the latest version (23.06.2016) of the website

- GitHub: <https://github.com/phnmnl>
- Uppsala University: Institutionen för farmaceutisk biovetenskap (<http://www.farmbio.uu.se/forskning/researchgroups/pb/PhenoMeNa/>)

Digital4Science platform:

PhenoMeNa is part of the European Commission's (EC) e-Infrastructure project groups for H2020 Work Programme 2014-15 and has contributed towards the production of the eInfra booklet for ICRI2016.

WP3 performance metrics

- Active participation to e-science conferences (3)
- Active participation in metabolomics-related meetings (5)
- Active participation to general biomedical/life science conferences (6)
- Number of articles (scientific or general public) published (6)



2.2.3. WP4- Interfacing with Biomedical European Infrastructures

During this period, the main achievements were:

- Initiating communications and interactions with other European infrastructures with similar interests.

These activities correspond to the consortium's initiative towards the project's objective towards developing a model to ensure sustainability of PhenoMeNal.

Work plan

In the WP4 we aim at maximising communications with European infrastructures with an interest in biomedical data generation and analysis. The PhenoMeNal project through this work package, will actively participate in the concertation activities, consultations and other meetings and events related to e-infrastructure. The objective is to optimise synergies between projects by providing input and receiving feedback from other stakeholders. To this end, the PhenoMeNal project is being implemented within the framework of numerous European initiatives in the field of biological medicine that are already on going or starting up in parallel with PhenoMeNal itself. Such initiatives will thus provide natural reference points to ensure that the development of PhenoMeNal services is aligned with current/foreseeable needs of biomedical research in Europe. Furthermore, this network of contacts is useful to provide input to the activities of WP2 by identifying interested parties that can impact on the sustainability of PhenoMeNal. The activities of WP4 will leverage the involvement of various PhenoMeNal partners in many of the European initiatives mentioned above. For example:

- a) Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP, WP4 lead) is a member of the Instruct European Strategy Forum for Research Infrastructures (ESFRI) and of Coordinated Research Infrastructures Building Enduring Life-science Services (CORBEL).
- b) The Chancellor, Masters and Scholars of the University of Oxford (UOXF, WP8 lead) is a partner in the IMI eTRIKS (European Translational Information & Knowledge Management Services) project.
- c) European Molecular Biology Laboratory (EMBL-EBI) is a key partner in ELIXIR (<https://www.elixir-europe.org>, a distributed infrastructure for life-science information).

First workshop

A joint workshop with the CORBEL partners (see also the subsection on "*Leveraging existing involvement in European Infrastructures*" below) focused on the molecular aspects of biomedicine was held at CIRMMP's premises (Florence, Italy) from February 4th to 5th, 2016. The workshop involved participants from various European



infrastructure projects including: ELIXIR, PhenoMeNal, Euro-BioImaging (<http://www.eurobioimaging.eu>), the Infrastructure for systems biology Europe (ISBE, <http://project.isbe.eu>), INFRAFRONTIER (<https://www.infrafrontier.eu>), INSTRUCT (<https://www.structuralbiology.eu>), EU-OPENSREEN (<http://www.eu-openscreen.eu>), and EMBRC (European Marine Biological Research Centre, <http://www.embrc.eu>).

The workshop included discussions on the usefulness of metabolomics for the activities of the various ESFRIs and the possible relevant services. For example, metabolomics is explicitly taken into account into WP4 of ISBE (<http://project.isbe.eu/preparatory-phase/wps/wp4/>). It was agreed that the points raised during this discussion would be summarised in a “white paper”, whose first draft is to be assembled by the PhenoMeNal partners involved in WP4. The draft will then be circulated to the participants in the Florence workshop.

More in detail, the topics below were addressed:

- What is PhenoMeNal, how it is organised, what are its expected outputs.
- Relationship with BBMRI (not present at the meeting), including EXCEMET.
- Usage of metabolomics in biomedicine: from the definition of individual phenotypes (metabotypes) to the quantitative assessment of the impact of disease and/or treatments on the metabotype.
- Usage of metabolomics for the prediction of features such as survival rates.
- Recognition of current/perspective usage of metabolomics at the RIs and possible specific scientific topics.

Scouting

Efforts have also been made to establish links between PhenoMeNal and [Excemet](#) on one side and the Michigan Infrastructure on the other and also to liaise with the [Metabolomics Innovation Center](#) (TMIC) in Edmonton (Alberta) and with one of the six metabolomics centers recently financed by the NIH in the US (Ann Arbor, Michigan) with a total of about 150 M\$.

Leveraging existing involvement in European Infrastructures

The IMI-funded eTRIKS (European Translational Information & Knowledge Management Services) project is a public-private partnership between the European Union IMI programme and EFPIA (European Federation of Pharmaceutical Industries and Associations), consortia of big pharmaceutical companies. eTRIKS focuses on translational research in pre-competitive studies, developing an e-infrastructure platform based on the open-source tranSMART data warehouse project, infrastructure provision and data curation services to disease or drug research area-specific IMI projects, and



looking towards long-term sustainability of the platform development and associated services by building up a rich user community. **UOXF** (WP8 lead) is currently a partner on the eTRIKS project and there is opportunity for PhenoMeNal to build a relationship with eTRIKS and its partner organisations via this bridge. In the previous Coordination of Standards in Metabolomics (COSMOS, <http://www.cosmos-fp7.eu>) project, the interaction with BioMedBridges (a joint effort of twelve biomedical sciences research infrastructures on the ESFRI roadmap to develop a shared e-infrastructure, <http://www.biomedbridges.eu>) was particularly effective in identifying ESFRIs with a distinct interest in metabolomics. The following were explicitly mentioned in a joint statement by COSMOS and BioMedBridges and was approved by the Steering Committee of BioMedBridges (More details in Appendix to D2.1, WP2):

- **BBMRI-ERIC** - Biobanking and Biomolecular Research Infrastructure (<http://bbmri-eric.eu>)
- **ELIXIR**
- **Euro-BioImaging** - a large-scale pan-European research infrastructure project on the ESFRI Roadmap
- **EU-OPENSREEN** - a large-scale Research Infrastructure (RI) with an 'open' pre-competitive character that makes all generated tools and data publicly accessible
- **INSTRUCT**
- **EATRIS** - European infrastructure for translational medicine (<http://www.eatris.eu>)
- **ECRIN** - European clinical research infrastructure network (<http://www.ecrin.org>)

After BioMedBridges came to its conclusion, 11 biomedical ESFRIs joined in the CORBEL consortium. The CORBEL project will develop the tools, services and data management required by cutting-edge European research projects involving European RIs in the field of Biological and Medical sciences. In particular, CORBEL aims to leverage the results of BioMedBridges to develop innovation pipelines supporting research in health and molecular bioscience. **CIRMMP**, as a node of INSTRUCT, was a partner in BioMedBridges and is now a partner in CORBEL. In this context CIRMMP will work to maintain the communication channel implemented by COSMOS and BioMedBridges, in order to provide a direct link between PhenoMeNal and the RIs that are partners of CORBEL.

Future Plan

At the 2016 Metabolomics conference (Dublin, Ireland - 27-30 June 2016) we will organise a meeting of EXCEMET partners to obtain further input on the needs of other infrastructures as well as of initiatives not funded by the EC. Furthermore, EXCEMET



has close interactions with the BBMRI ESFRI, which was not involved in the first workshop.

In parallel, we will maintain awareness of CORBEL activities, to identify other possible occasions of interaction with its partner ESFRIs. As an outcome of these workshops, we plan to install the first working group and select a specific aspect of the biomedical and/or eScience fields that are most likely to impact on the metabolomics field and PhenoMeNal direction.

2.2.4. WP5 – Operations and Maintenance of PhenoMeNal grid/cloud e-infrastructure

During this period, the main achievements were:

- A continuous integration established and available at <http://phenomenal-h2020.eu/jenkins/>.
- Interaction with cloud providers and EGI-ENGAGE.
These activities correspond to the consortium's initiative towards the project's objective towards operating and consolidating the PhenoMeNal e-infrastructures based on existing internal and external HPC and grid resources, including EGI.

During the first 6 months, WP5 has carried out several technology evaluations and made thorough investigations on the type of e-infrastructure needed to accomplish this.

Work Plan

Virtual meetings

Joint virtual meetings (Google hangouts) jointly with WP9 were held on 2015-10-16, 2015-11-02, 2015-11-13, and 2015-12-10 with 10-15 participants in each call.

Technology evaluations

Many technological solutions and frameworks exist for e-infrastructure and platform development, and WP5 has evaluated a range of these for their applicability in PhenoMeNal:

- *Cluster management/scheduling*: Mesosphere, Kubernetes
- *Cloud provisioning*: Vagrant, Ansible, Terraform, Cloud-config
- *Microservices*: Virtual Machines and Docker containers
- *Microservices/container orchestration*: Stackstorm and MANTL were evaluated. Contact has been established and support obtained by MANTL developers.



Continuous integration

A Jenkins server has been established for continuous integration, available at <http://phenomenal-h2020.eu/jenkins/>. Given its need to build Docker images for other applications, and that Docker-inside-Docker is not a viable solution; the Jenkins server has attached a variable number of slaves (Ubuntu VMs), which can build both Docker images and Vagrant images. This ability requires root access on those slave machines, for which reason a two-way authentication system was installed on the Jenkins master server (slaves are not accessible from outside the private network) and the slave machines have no access to private keys that would allow access to other machines in the network. The Jenkins instance is hooked to the Github repos of all the current tools that have been Dockerized so far within PhenoMeNal, so new builds are triggered on new code being pushed. Finally, each Docker image built is pushed by Jenkins to the PhenoMeNal internal Docker registry (CoreOS VM, runs the registry on Docker), which in turn makes the images available to the deployed Kubernetes Cluster. All the machines and the cluster mentioned run inside the EMBL-EBI EMBASSY Cloud PhenoMeNal Tenancy (OpenStack).

From Galaxy Workflows to Scalable Computation Clusters:

We have completed a first prototype tool that can be invoked within the Galaxy Workflow environment and can offload the computation directly onto a Kubernetes (k8s) Cluster, using a dockerized tool (that would otherwise run locally on the same machine as the galaxy instance), which the k8s Cluster consumes from our local Docker registry. This first prototype relies on a shared filesystem (GlusterFS based). This is running at the EMBL-EBI EMBASSY Cloud. The Galaxy instance (accessible at <http://phenomenal-h2020.eu/galaxy/root>) is based on the Workflow4Metabolomics (W4M) available image. This instance deployed is just for internal development purposes, and is closed to external users. The Figure below shows how all the elements deployed interact.

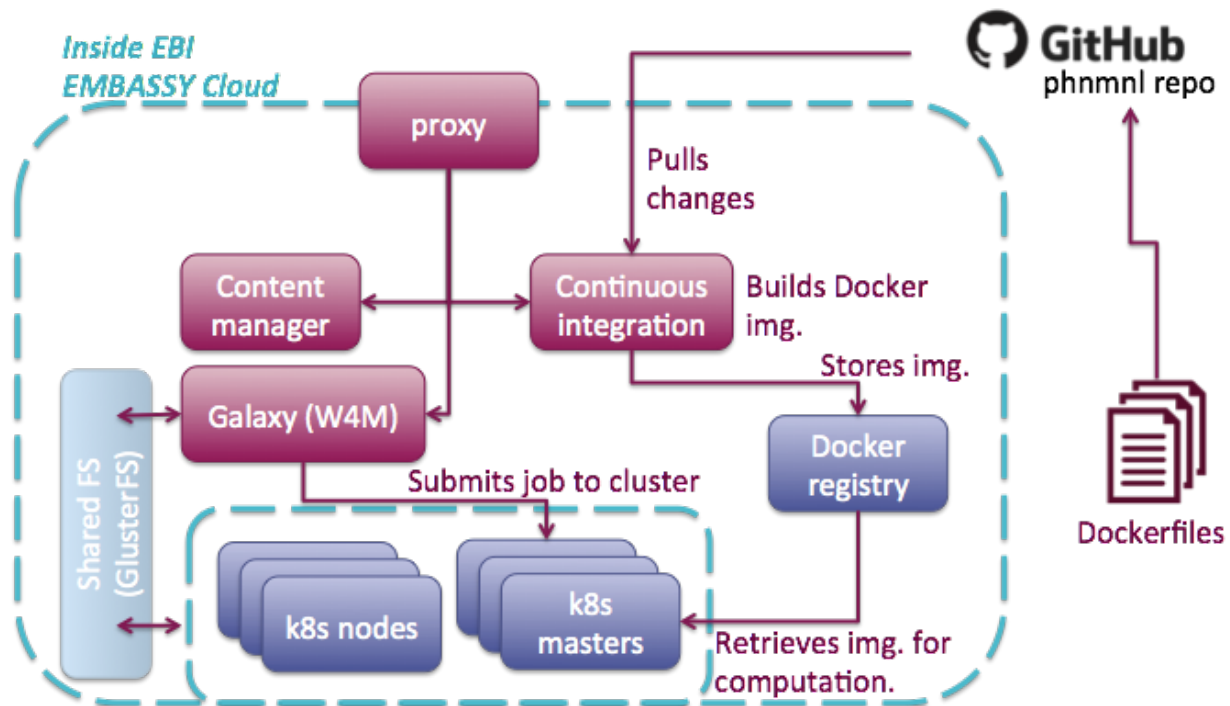


Figure 3: Architecture of PhenoMeNal



Cloud providers

A decision was made to initially operate with local Openstack installations and Google Cloud Platform (GCP). Contact was established with Google and a shared PhenoMeNal project has been set up with an initial credit of 2,500 USD kindly provided by Google. Local OpenStack installations at Uppsala University and at the EMBL-EBI EMBASSY Cloud have been used for prototyping.

Interaction with EGI-ENGAGE

PhenoMeNal has been selected as one of the scientific use cases of the ELIXIR Competence Centre as part of the EGI-Engage project. Discussions are on going how EGI-Engage can best support PhenoMeNal via the planned ELIXIR Compute Platform (ECP).

Data federation

A decision was made to focus during M1-6 on computing frameworks with a mounted shared file system. We will from month 6 start evaluating iRODS and object storage systems for their feasibility in PhenoMeNal e-infrastructure.

Training/workshops

A training session/workshop on e-Infrastructure is planned for 2016-02-29 -- 2016-03-02 in Uppsala, Sweden. The workshop participants will be exposed to and given the opportunity for hands-on activities using a selected set of the technology frameworks listed above.

Future Plans

Continue the work on continuous integration in terms of security, scalability and sustainability for deliverable D5.1 – Build system with continuous integration, providing development snapshots of PhenoMeNal Virtual Machine images (M9).

2.2.5. WP6 - PhenoMeNal Virtual Research Community Gateway

During this period, the main achievements were:

- Identify potential users from the community
- Identify tools for the computational analysis of metabolomics data
- Submission of deliverable D6.1 User experience document on VRC design guide (M3) on time.



- Release of VRC design guide (M3) on time.

These activities correspond to the consortium's initiative towards the project's objective towards integrating existing open source tools for analysis of metabolomics data.

Work plan

The PhenoMeNal Virtual Research Community (VRC) will be the central access point for all users. This is the online portal enabling user interactions, virtual machine distribution, access to compute resources and information exchange. Users can search for existing public datasets and perform online analysis. The VRC will interact with the PhenoMeNal middleware developed in WP5 to federate on-demand job requests between provisional data, compute, and analysis resources. Expert help and training will be available through the VRC.

User Experience Meeting for the VRC Portal

The Virtual Research Community (VRC) User Experience (UX) workshop was held on November at the EMBL-EBI, with 23 participants that spanned different levels of seniority and diverse domain fields within metabolomics. The workshop was held at the EMBL-EBI, took a complete day, and included a number of User Experience (UX) activities:

- Identification of user groups
- Personas design
- Experience map
- Identify and prioritise tools/services

These activities were aimed at understanding:

- User Roles: expected users of the VRC; their aims and what they would expect from the gateway. Activities 1 and 2.
- Workflows: relevant workflows for the users defined. Activity 3.
- Tools: a prioritised set of software tools in metabolomics that the project should include among those to be offered in containers. Activity 3 and 4.

Detailed results can be obtained from the UX deliverable "D6.1 User Experience document on VRC Design Guide" at <https://docs.google.com/document/d/1ns9nXkid5qDORtCSprLZXluPTXbUCtSY14Y8mHEF-xg/edit>.

Future plans

Plans for the provision expected at the VRC:



Discussions with partners have led to initial decisions on what the PhenoMeNal VRC will offer in terms of services, being:

- a) A helpdesk to provide expert help and training. Knowledge that will be shared back to the community as FAQ/How to. Information on how to access compute resources.
- b) Virtual Machines: downloadable pre-configured easy to use images of all (or part) of the services for testing and/or teaching purposes. Limited technical skills required!
- c) Containers: portable versions of the individual services. Containers allow developers to package a tool/service with all of its dependencies into a standardised unit for software development and deployment.
- d) Infrastructure-as-code: The complete (or parts of) PhenoMeNal e-infrastructure codified in proper provisioning standards (Terraform, Ansible, etc.) to be automatically deployed on cloud instances. This will be the easiest and preferred way to fire up the developed e-infrastructure in both private and public clouds.
- e) Installation guides and provisioning scripts with detailed step-by-step instructions on how to deploy the services. By sharing information on how to install, configure and run tools and/or pipelines, we ensure technical sustainability of the PhenoMeNal WP 5, 6 and 9 deliverables. In addition to, supporting a sustainable future, it will also encourage data processing standardisation (WP8) within the metabolomics community by providing easy access to stable and proven services.
- f) All services developed (or improved) within the PhenoMeNal project will be shared under a suitable¹ open source license. Changes to services are submitted and documented using a software version control system (<https://github.com/phnmnl>). This allows developers to contribute improvements or fixes and review changes made over time.
- g) The VRC should serve users with different needs in terms of granularity and diverse levels of computational expertise. It will contain as well a proper documentation to aid users in the exploitation of these different options. We are observing as well other examples of scientific “App stores” such as the EGI’s Application Catalogue, bioboxes and biodocker; and non-scientific ones, such as Bitnami.

2.2.6. WP7 – Privacy and Ethics

During this period, the main achievements were:

¹ Existing licenses may influence the choice of license under which we share software changes or additions.



- Understanding the ELSI requirements
- Interactions with projects such as BioMedBridges
- Submission of deliverables: D7.1.1 Workshop on best practices in handling sensitive human data, taking into account National and Institutional legal policies (M3) and D7.2 Report on the policies and procedures for sensitive human data management (M6) on time.
- MS7.1 Agreement on policies and procedures for sensitive human data management (M6) on time.

These activities correspond to the consortium's initiative towards the project's objective to establish privacy protection methods that allow working with highly sensitive molecular phenotype data.

Work Plan

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 the data generated from the 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, regulation 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.

To this end, two workshops to assimilate and disseminate best practice were planned. The WP7 is now finalising its second set of deliverables - a report on policies and procedures for handling sensitive data, and a data provider form.

First ELSI Workshop:

The first workshop was held on 20th November 2015 at Imperial College London and attended by 22 people from PhenoMeNal partners plus 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:

- a) Privacy and ethics issues are mainly governed by the consent given by the participants, and the level of anonymisation of the data.
- b) There is no consistent set of laws or policy around privacy and ethics in the EU and even less outside it. The EU has recently revised its data protection



legislation. Projects such as BiomedBridges have developed tools, which we will develop and adapt to the PhenoMeNal case.

- c) The European Genome-Phenome Archive (EGA) is a good model on which to base PhenoMeNal security models and data deposition. We may consider the EGA data access forms and agreements as a starting point for the PhenoMeNal data submission form.
- d) The AIRWAVE project has potential as a good use-case for PhenoMeNal. This will most likely need an application to the AIRWAVE access committee.

Second ELSI Workshop:

The second workshop was held on 27th January at Imperial College London and attended by 10 people, including two representatives of the BiomedBridges project plus four accessing over Skype. The workshop focussed on deliverables D7.2 (policies and procedures report) and D7.3 (data provider form). The main discussion points were as follows:

- a) Various public data sets were proposed as test data for PhenoMeNal including at least one set of data on each for NMR, GC-MS and LC-MS.
- b) The AIRWAVE study was considered to be a good use case for “bringing the compute to the data”, since in discussions with the Ethics committee, it was decided that local access was required due to ethics. The MESA study was considered as a use case for large-scale human phenotyping. We expect that this data will be made public in March so can be distributed to the rest of the consortium. Both projects are similar in the size (n~4000), type (NMR/LCMS) and complexity of the data.
- c) The BiomedBridges data provider form provides a good starting point for the PhenoMeNal data provider form.

Future Plans

- a) Preparation for D7.4 - workflows to extract information from sensitive data with minimum compromise.
- b) A review on Ontologies and Terminologies assisting in Privacy and Data anonymisation schemes has been collated by WP 8 at https://docs.google.com/document/d/1myK0AQbJwqOSuZ_Ao5twsCCz59Wmk8HKA98GS-Xansk/edit

2.2.7. WP8 - Data Provenance, Compliance and Integrity

The main achievement during this period were:



- Commissioning of a survey on ‘Standards, Compliance and Data Integrity’ to gain a better understanding of the requirements for data standards, compliance and integrity, in the context of metabolomics research.

This survey corresponds to the consortium’s initiative towards the project’s objective for data exchange and storage standards while improving and scaling up tools.

Work plan

Building a federated infrastructure, by essence, requires the stakeholders to agree on reporting and communication standards for messages and information to be reliably brokered and exchanged between systems. Contingent to a central demand to ensure data evaluation, audit and review imposes that provenance, traceability and integrity of data is delivered.

WP8 will therefore develop a set of modular format specifications to achieve maximum efficiency during data exchange with each module focusing on specific component of data management, while retaining compatibility with existing community standards and addressing distinct and complementary scenarios of usage. Data safety and patient de-identification schemes developed in WP7 will be considered.

Virtual Meetings

Google Hangouts were held with WP8 contributing partners on 13th November 2015 and February 3rd 2016, with 10 and 18 participants in attendance respectively. A Google group forum was also established to keep a record and focus for WP8 email/on-going discussions.

Preliminary work towards D8.1

D8.1 is to produce a “Report on community standards for reporting, access and integrity supported in the PhenoMeNal grid” that is to be disseminated in a dedicated BioSharing page and via the project website, **by project month 12**. To achieve this, we have commissioned a survey on ‘Standards, Compliance and Data Integrity’ to gain a better understanding of the requirements for data standards, compliance and integrity, in the context of metabolomics research. The purpose of the survey is to gather information about metadata management needs in the metabolomics community.

In collaboration with PhenoMeNal partners, via Hangout discussions and through email exchanges, a survey was drafted up to solicit this information from various target groups. Types of studies, metabolomics applications used, awareness of patient consent, legal frameworks, how data is shared (databases, other places), standards, LIMS system use, are all queried.

At the time of writing, the preview survey can be viewed at this link:



https://www.surveymonkey.com/create/survey/preview?sm=yQtC4E3CYbMfAOoAGXbH7yo6X2NPW6fndjaYpuzXjT4_3D

The survey is to be initially disseminated to the following target audiences:

- a) All PhenoMeNal partners are expected to provide responses; including contacting any clinical sites based within each partner institution.
- b) Metabolomics Society Data Standards task group.
- c) French Metabolomics and Fluxomics Network, RFMF, via their newsletter.
- d) MetaboNews email newsletter, published in partnership between TMIC and the Metabolomics Society.
- e) Linked from relevant partner websites, e.g. PhenoMeNal project, MetaboLights database website.
- f) EMBL-EBI mailing lists (including training lists and the MetaboLights users lists).

The survey will go out imminently and results to be gathered over the next 2-3 months.

Cross-Work Package activities

There are WP8-related activities working with WP5 and WP9 on integrating data reporting standards into the wider PhenoMeNal infrastructure. To this end WP8 is represented in the aforementioned Work Package's Hangout meetings as well as in attendance to:

- a) Training session/workshop on e-Infrastructure, Uppsala, Sweden, February 29th-March 3rd 2016
- b) Galaxy/W4M hackathon, CEA, France, March 14th-16th 2016.
- c) WP8 representatives were also present at the two WP7 ELSI workshops to be informed about relevant standard needs to ethics/privacy issues that may feed back to standards development. A review on standardisation efforts for ELSI issues using ontologies has been created at https://docs.google.com/document/d/1myK0AQbJwqOSuZ_Ao5twsCCz59Wmk8HKa98GS-Xansk/edit#heading=h.wgltex7iack6

Future plans

Work is already underway towards the remaining deliverables that are standards and supporting software deliverables. Good progress is being made to ISA metadata reporting development (www.github.com/ISA-tools/isa-api) and to mzML and nmrML (www.github.com/nmrML) work. Tools are already being worked on in collaboration with WP5 and WP9 towards prototyping some data management microservices.



Recently WP8 participants have started looking into an ontology-based metabolite identification and evidence scheme² and the Evidence Code Ontology (<http://www.evidenceontology.org/>), which will allow to encode data provenance and reliability of identification assertions in the future, i.e. allowing to set confidence thresholds for search and retrieval tasks.

Note that both the aforementioned survey results on 'Standards, Compliance and Data Integrity' and the cross-work package interactions will inform and guide the development on the standards and tools developed out of WP8.

2.2.8. WP9 - Tools, workflows, audit and data management

During this period, the main achievements were:

- Identification of use cases and tools to be developed within PhenoMeNal.
- Submission of deliverable D9.1 Report on existing software tools, workflows and analytical pipelines initially supported in the PhenoMeNal grid (M6) on time.

These activities correspond to the consortium's initiative towards the project's objective to improve and scale up tools.

Work Plan

Our goal is to develop and maintain the primary scientific and technological tools as well as corresponding interfaces. We specify and integrate software pipelines and tools utilised in the PhenoMeNal e-Infrastructure into Virtual Machine Images (VMI), adhering to data standards developed in WP8 and supporting the interoperability and federation middleware developed in WP5.

We collected a list of distributed tools for phenomics, metabolomics and bioinformatics processing pipelines and workflows suitable for packaging into VMIs (see D9.1 report). We have started to use public repositories and continuous integration to always provide development snapshots of the infrastructure VMIs. In order to avoid data lock-in and ensure continuous availability of the infrastructure we are closely working together with WP5 and WP8. A primary goal is to hide the complexity of the underlying infrastructure to the actual user (e.g. biologists, clinicians), while giving easy to understand technical instructions to bioinformaticians for installing the supplied PhenoMeNal VMIs in a short time while preserving data privacy and security. We have already begun to produce development-snapshots of the VMIs, which are available in the PhenoMeNal Continuous Integration Service Jenkins (<http://phenomenal-h2020.eu/jenkins/>).

² Sumner, Lloyd W et al. "Proposed quantitative and alphanumeric metabolite identification metrics." *Metabolomics* 10.6 (2014): 1047.



We arranged for 6 hangouts (with usually 10-15 attendees each) so far:

- a) 2015-10-16: initial planning, evaluation on technical infrastructure and suitable technologies
- b) 2015-11-02: discussing possible architecture of the PhenoMeNal cloud infrastructure, technical evaluation, upcoming workshops
- c) 2015-11-13: monthly status meetings, integrating software standards, integration of new tools such as IPO and mzml2isa, integration of GALAXY and W4M
- d) 2015-12-10: reports on status meetings, privacy and VRE workshop, finding real use cases and integrating them into the proposed workflow with GALAXY and W4M, microservice orchestration, organising workshops
- e) 2016-01-13: discussion on how data federation and virtualisation interface with WP9 and how to integrate them in a transparent way
- f) 2016-02-10: reports on status meetings and workshops, discussing how work packages better integrate with each other, preparations for upcoming workshops in Paris and Uppsala, collection of use cases and tools

One Hangout contained a demonstration of a data analysis using the Workflows4Metabolomics Galaxy platform, which was recorded and is available on YouTube <https://www.youtube.com/watch?v=f2WoAZ24Gjo> .

Report on collected use cases and standardised tools to be used within PhenoMeNal

We collected four real world use cases that have distinct practice-driven demands on data storage/federation and privacy. The use cases are discussed in detail in the D9.1 Report on existing software tools, workflows and analytical pipelines:

- a) Data processing for MESA metabolomics data Use Case
- b) Data processing for CoLaus Use Case
- c) Data processing for the Uppsala Use Case
- d) Data processing, statistical analysis, and annotation of the “Physiological Variations of the Urine Metabolome” Use Case
- e) Data processing for fluxomic analyses

We collected a minimum list of tools to be used within PhenoMeNal. This first list of required software-tools was gathered at the WP6 Virtual User Community workshop and will be extended in the future, also based on the WP8 Survey results and the four primary use cases in D9.1. We started to collate this list in a table with the following scheme, the full list being available in the D9.1 report.



Technology	Functionality	Tool Name	Open Source	Virtualisation status	As required by
NMR	nmrML&FID2processedData conversion	nmRIO	yes	scheduled	Own tool involvement
NMR	assignment & quantification of metabolites	rNMR	yes	Add constructors for nmrML, Galaxy based docker	VRC Meeting
NMR	automated metabolite identification for mGWAS	MetaboMatching	Yes?	MatLab dependencies?	CoLaus Use Case
MS	Vendor to mzML conversion	msconvert	yes	Proteowizard VM / Docker image with msconvert pre-installed	Use cases
...

Figure 4. Tools to be integrated into the PhenoMeNal compute infrastructure were collected from different sources.

Future plans

Work towards development of PhenoMeNal Virtual Machine Images (VMIs).

2.2.9. WP1- management activities

The main achievements during this period were:

- Kick-off meeting
- Establishment of PhenoMeNal SAB
- Several workshops and staff-exchanges for knowledge exchange
- Submission of deliverables D1.1 Project plan (M2), D1.2 Consortium agreement (M1), D1.3 Minutes of kick off meeting (M3) and D1.4.1 Bi-annual progress report (M6) and D1.5.1 Data Management Plan (M6) on time.
- MS1.1 Project Plan released (M2) on time
- MS1.2 Initial release of Data Management Plan (M6) on time.

Work Plan

All the tasks related to legal, financial and administrative management of the project were performed as indicated in the Grant agreement. The pre-financing amount received from the European Commission was distributed between the partners.



Distribution and progress of tasks according to the project plan according to the WP were monitored via an agile project management tool Pivotal Tracker:

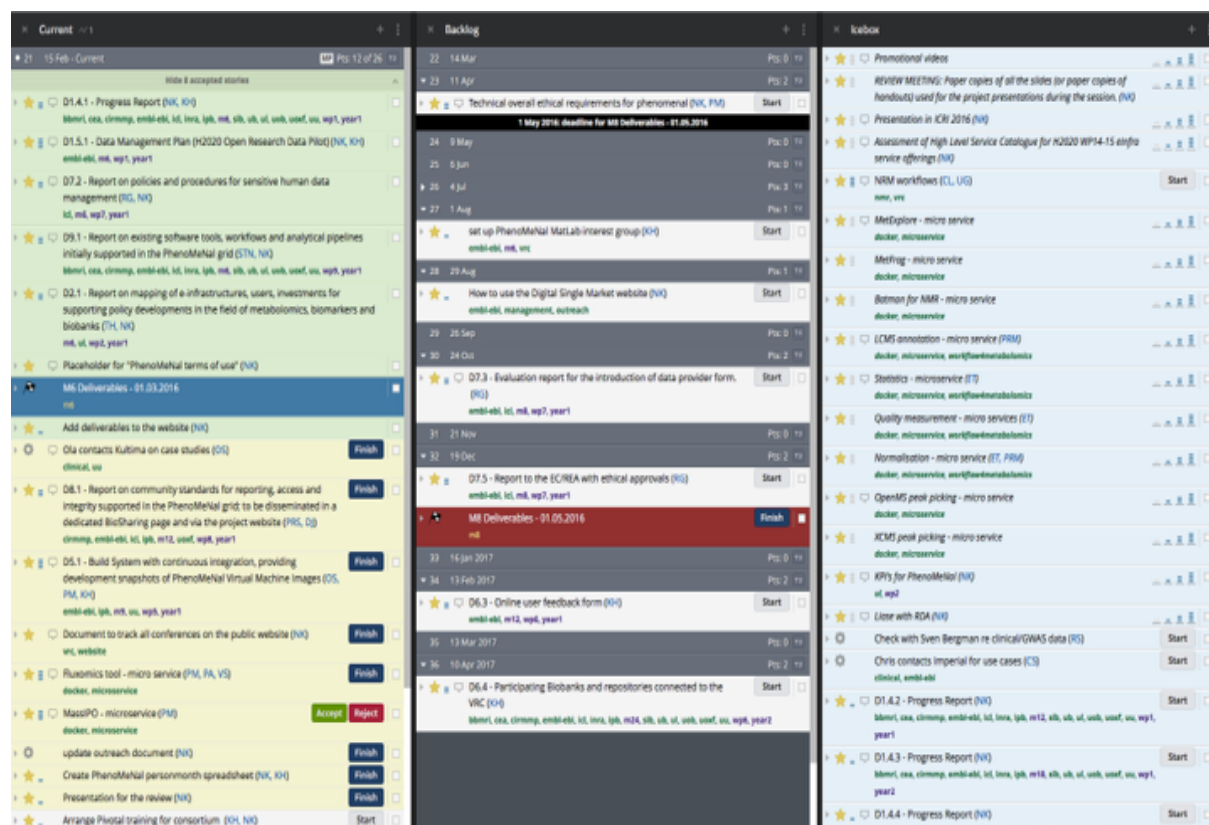


Figure 5 Screen shot of the Pivotal Tracker.

The project guidance and monitoring of progress was based on monthly online meetings and WP specific hangouts:

PhenoMeNal online meetings:

One of the action points from the kick-off meeting was to organise regular teleconferences to track the progress and discuss issues related to PhenoMeNal. As a part of this WP, we have organised:

- a) PhenoMeNal monthly-status update meeting (repeating every four weeks on a Friday) via Google hangouts:

Led by the coordinator and attended by all the partners, the focus of these meetings is the overall progress of the project, progress of the individual WPs and their upcoming deliverables as well as any other issues or important announcements that need attention of the consortium. Discussions and decisions are minuted using a Google document that is shared and stored in the Google drive.



b) Google hangouts for individual WPs:

These meetings are targeted specifically towards the progress, breaking down of the work package tasks and its distribution between the contributing partners and other issues related to individual WPs. The meetings are scheduled bi-weekly/monthly and are led by the respective WP leader. The discussions and decision made are extensively documented. Currently these include for:

- Sustainability of PhenoMeNal (WP2) led by University of Leiden (UL, WP2 lead)
- Data provenance, compliance and integrity (WP8) led by The Chancellor, Masters and Scholars of the University of Oxford (UOXF, WP8 lead)
- Joint meetings for Operations and maintenance of PhenoMeNal GRID/Cloud (WP5) and Tools, workflows, audit and data management (WP9) under the flagship of Uppsala University (UU, WP5 lead) and Leibniz-Institute for plant biochemistry (IPB, WP9 lead).

Additional hangouts for other WPs are also being planned.

c) Google hangouts for general outreach activities:

These include occasional dedicated hangouts for planning and organising general outreach activities in the form of workshops and hackathons.

Project Meetings:

PhenoMeNal officially started in September 2015 with a kick-off meeting in Cambridge, UK organised by EMBL-EBI (coordinator) from 8th to 10th September 2015. The meeting was structured as a three-day event with a plenary session on the overview of the whole project structure, the consortium agreement and legal implications and introductions to different partners. The meeting also included presentations on the PhenoMeNal work packages (WPs) and role of each partner in contributing towards the different deliverables. A comprehensive session on the overall project management and outreach of the project was also accomplished. Lightning workshops followed by brief discussions on the privacy and ethics with respect to the sensitive data, tools, workflows and grid/cloud resources that will form the basis of the PhenoMeNal infrastructure were also carried out. The consortium agreed on a tentative plan for a stakeholder meeting, workshops and staff-exchange meetings as foreseen for year 1. Other activities included recommendations for the scientific advisory board (SAB).



PhenoMeNal Scientific Advisory Board meeting:

As an outcome of the suggestions made during the kick-off meeting, a PhenoMeNal SAB representing experts in the field of metabolomics, bioinformatics, clinical medicine, and bioethics, was established. The board comprises of the following members:

- a) *Prof. Hannelore Daniel* - Professor, Molecular Nutrition Unit, Technical University of München
- b) *Prof. Murielle Bochud* - Professor, Institut universitaire de médecine sociale et préventive, Lausanne
- c) *Dr. Nicolas Schauer* - Managing Director, Metabolomic Discoveries, Germany
- d) *Prof. Barend Mons* - Professor, Biosemantics Group, Leiden University Medical Centre, the Netherlands, Chair of the high level expert group for the European open science cloud at European Commission
- e) *Prof. Rima Kaddurah-Daouk* - Professor, Department of Psychiatry / Behavioral Medicine, Duke Psychiatry and Behavioral Sciences, Duke University School of Medicine
- f) *Dr. Theodore Alexandrov* - Team leader, European Molecular Biology Laboratory, Heidelberg.
- g) *Prof. Oliver Kohlbacher* - Professor, Applied Bioinformatics Group, University of Tübingen
- h) *Dr. Martin Ingelsson* - Department of Public Health and Caring Sciences, Geriatrics; Molecular Geriatrics / Rudbeck laboratory, Uppsala University
- i) *Dr. Heidi Howard* - Centre for Research Ethics & Bioethics, Uppsala University
- j) *Prof. Whei-Mei Teresa Fan* - Professor, Toxicology and Cancer Biology, University of Kentucky College of Medicine
- k) *Prof. Sabine Bahn* - Professor, Department of Chemical Engineering and Biotechnology, University of Cambridge
- l) *Dr. Jules Griffin* - Group leader, MRC Human Nutrition Research, Cambridge
- m) *Prof. David Wishart* - Professor, Departments of Biological Sciences and Computing Science, University of Alberta

The first SAB meeting was held on 17th December 2015 via WebEx. The meeting was kicked off by a presentation on the general overview and aims of the PhenoMeNal project. The key points that were discussed included:

- Methods for standardising phenotyping data: PhenoMeNal should take efforts in standardising phenotyping data.
- Data privacy issues and sensitivity of metabolomics data.
- Types of data that can be deposited into PhenoMeNal that is if data such as cell



culture, animal models etc. can be deposited into PhenoMeNal.

- User access to clinical data and data governance and policies of PhenoMeNal.

Views on the willingness of the community to deposit their clinical data and approach of PhenoMeNal in acquiring such sensitive patient data.

The staff-exchanges so far:

- a) PhenoMeNal Hackathon on Software Standards, EMBL-EBI, November 11, 2015. The 1 day hackathon was focussed on software development and “containerisation standards” for PhenoMeNal with the aim to come up with an initial draft of “The PhenoMeNal cloud”.
- b) PhenoMeNal staff exchange for interactions between W4M-Galaxy and scalable clusters infrastructure, EMBL-EBI, January 25 - 29, 2016. The main purpose of the exchange is that developers from W4M (Pierrick Roger Mele) and EMBL-EBI (Pablo Moreno) work together towards enabling W4M instance at the EMBASSY Cloud to deploy jobs at the previously deployed Kubernetes (K8S³) Cluster at the same infrastructure.
- c) PhenoMeNal staff exchange for interactions with MetXplore, EMBL-EBI, February 15-19, 2016 focussed on integration of Network visualisation currently developed by INRA.
- d) PhenoMeNal staff exchange for e-Infrastructure training at Uppsala University, February 29- 2nd March 2016 focussed to gain hands-on experience with cloud computing environments and microservices.

Future plans

- Organising Annual Consortium and SAB meeting before M12

3. Conclusions

During the first six months of the project, the PhenoMeNal consortium has adhered to its original project objectives. The tasks were completed on time and are under progress according to the project plan defined during the initial stages of the project. The deliverables and milestones were achieved within the assigned framework of time. No minor or major deviations, critical risks technical or administrative, were identified during this period and the consortium plans to continue its activities towards accomplishing its aim of an integrated, secure and sustainable e-infrastructure.

³ The abbreviation K8S stems from the fact that in Kubernetes there are 8 letters between K and S, hence K8S.