# Deliverable 1.4.4

<table>
<thead>
<tr>
<th><strong>Project ID</strong></th>
<th>654241</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Title</strong></td>
<td>A comprehensive and standardised e-infrastructure for analysing medical metabolic phenotype data</td>
</tr>
<tr>
<td><strong>Project Acronym</strong></td>
<td>PhenoMeNal</td>
</tr>
<tr>
<td><strong>Start Date of the Project</strong></td>
<td>1st September 2015</td>
</tr>
<tr>
<td><strong>Duration of the Project</strong></td>
<td>36 Months</td>
</tr>
<tr>
<td><strong>Work Package Number</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Work Package Title</strong></td>
<td>Management</td>
</tr>
<tr>
<td><strong>Deliverable Title</strong></td>
<td>D1.4.4 Biannual progress Report</td>
</tr>
<tr>
<td><strong>Delivery Date</strong></td>
<td>M24</td>
</tr>
<tr>
<td><strong>Work Package leader</strong></td>
<td>EMBL-EBI</td>
</tr>
</tbody>
</table>
Contributing Partners  |   EMBL-EBI, ICL, IPB, UB, UoB, CIRMMP, UL, UOXF, SIB, UU, BBMRI, CEA, INRA, CRS4, UL
---|---
Authors | Namrata Kale, Pablo Moreno, Ken Haug, Daniel Schober, Steffen Neumann, Sijin He, Ola Spjuth, Merlijn van Rijswijk, Michael van Vliet, Reza Salek, Vitaly Selivanov, Rico Rueedi, Claire O'Donovan

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 March 2017 (M19) - August 2017 (M24) inclusive.

Contents

1 EXECUTIVE SUMMARY......................................................................................................................................................... 3

2 OBJECTIVES, WORK PROGRESS, ACHIEVEMENTS AND PROJECT MANAGEMENT.............. 3

2.1 Work progress and achievements during this period ........................................................................................................ 7
   2.1.1 WP2 - Sustainability of PhenoMeNal .......................................................................................................................... 7
   2.1.2 WP3 - Dissemination and Outreach .......................................................................................................................... 13
   2.1.3 WP4 - Interfacing with Biomedical European Infrastructures .................................................................................. 19
   2.1.4 WP5 - Operations and maintenance of PhenoMeNal GRID/CLOUD ......................................................................... 21
   2.1.5 WP6 - PhenoMeNal Virtual Research Community Gateway .................................................................................. 26
   2.1.6 WP7 - Privacy and Ethics ............................................................................................................................................ 29
   2.1.7 WP8 - Data provenance, compliance and Integrity .................................................................................................... 30
   2.1.8 WP9 - Tools, workflows, audit and data management .............................................................................................. 33
   2.1.9 All WP’s (joint effort) .................................................................................................................................................. 39

2.2 Project management during this period .......................................................................................................................... 40
   2.2.1 WP1 - Management .................................................................................................................................................. 40

3 CONCLUSION:........................................................................................................................................................................ 53
1 EXECUTIVE SUMMARY

In this 4th reporting period, from March to August 2017, the consortium continued to focus on the further development of the infrastructure to facilitate scalable and secure data analysis. The services provided by PhenoMeNal, including the infrastructure, virtual research environment (VRE) portal, workflows and tools, underwent rigorous system testing for functional and system requirements. To this end, in August 2017, a 2nd major release with the name **Bucetin**\(^1\) helped to improve the overall security and stability of PhenoMeNal. The number of tools and workflows available were also increased with detailed documentation on their usage and operation. As part of the release procedure, an indepth heuristic analysis of the PhenoMeNal portal is scheduled for completion in September 2017 to detect and resolve any usability issues with the interface prior to the main UX testing with the user community, scheduled for the later part of this year. The PhenoMeNal App Library/service catalogue was updated and the number of applications available via Galaxy workflows and Jupyter libraries through the PhenoMeNal VRE were increased to 45, from 32 reported in D1.4.3 (end of February 2017). Annual consortium and advisory board meetings, stakeholder engagement, user workshops, and completion of the periodic reporting were the other major activities accomplished in this reporting period.

2 OBJECTIVES, WORK PROGRESS, ACHIEVEMENTS AND PROJECT MANAGEMENT

During the period, the consortium has continued to work towards the general objectives of the project\(^2\) which were pursued through the combination of its networking, service and joint research activities integrated across its 9 work packages (WP).

As part of the consortium’s networking activities, a joint annual stakeholder and scientific advisory board meeting (SAB) was organised in June 2017 which included live demonstrations and presentations on the deployment of VRE and workflows. Dissemination and outreach included workshops, blogs on website, video tutorials and presentations and posters at international and national metabolomics and life sciences conferences. Ethical and legal implications (ELSI) continued to be considered and integrated across all work packages (WPs):

The main activities in the reporting period were:

---


- Continued establishment of collaborations with other e-infrastructures including working closely with the clinical users to understand their needs and incorporate them.
- Updated version of the PhenoMeNal App library with 45 applications³.
- The Beta version of PhenoMeNal VRE Portal after the first release (Alanine, March 2017), includes Single Sign-On (SSO) and automated VRE deployment to OpenStack, AWS and GCP.
- Detailed documentation, tutorials and developer guidelines were made available via PhenoMeNal wiki and “help pages”.
- Integration with the GO FAIR initiative⁴. Analysing areas where standards are missing and surveying and updating existing ones (e.g. nmrML).
- 4 new or expanded workflows available
  - **Fluxomics** - 13C traced MS fluxomics data analysis⁵,⁶
  - **MS-MetFrag-XCMS** - annotates molecules from compound (metabolite) databases to MS/MS (tandem mass spectrometry) spectra⁷,⁸
  - **Univariate and Multivariate Statistics** - exploratory data analysis, prediction, and feature selection⁹,¹⁰
  - **NMR1d** - 1D NMR Workflow¹¹
- PhenoMeNal production release¹² “Bucetin” in August 2017 that includes the following features:
  - Improved stability and performance of the deployments and running of the underlying infrastructure for a better user experience.
  - New and updated workflows, with detailed written and video tutorials, describing how to combine several tools to process and analyse data.
  - Active support by helpdesk.
  - Improved user experiences for the deployment of the PhenoMeNal VRE on OpenStack.
  - Introduction of metadata storage to persist user metadata for better user experience.

---

⁶ [https://www.youtube.com/watch?v=SaXjdFDFlJU&feature=youtu.be](https://www.youtube.com/watch?v=SaXjdFDFlJU&feature=youtu.be)
⁸ [https://www.youtube.com/watch?v=V5M-dcE3qz8&feature=youtu.be](https://www.youtube.com/watch?v=V5M-dcE3qz8&feature=youtu.be)
¹⁰ [https://www.youtube.com/watch?v=ABW99825gY&feature=youtu.be](https://www.youtube.com/watch?v=ABW99825gY&feature=youtu.be)
○ User experience testing of the existing prototypes to inform the development of the next release which is due Q1 2018.

The main deliverables achieved during this period were:

<table>
<thead>
<tr>
<th>No</th>
<th>Deliverable name</th>
<th>WP</th>
<th>Lead participant</th>
<th>Delivery date</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.5.2</td>
<td>Updated Data management plan</td>
<td>1</td>
<td>EMBL-EBI</td>
<td>Resubmission post review</td>
<td>Submitted</td>
</tr>
<tr>
<td>D1.4.4</td>
<td>Bi-annual Progress report</td>
<td>1</td>
<td>EMBL-EBI</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D2.3</td>
<td>Updated report on mapping of e-infrastructures, users investments for supporting policy developments in the field of metabolomics, biomarkers and biobanks</td>
<td>1</td>
<td>UL</td>
<td>M22</td>
<td>Submitted</td>
</tr>
<tr>
<td>D3.1.1</td>
<td>Report of the annual stakeholder meeting</td>
<td>3</td>
<td>UoB</td>
<td>Resubmission post review</td>
<td>Submitted</td>
</tr>
<tr>
<td>D3.2</td>
<td>Report on establishing and maintaining relations with publishers for supporting data deposition</td>
<td>3</td>
<td>UoB</td>
<td>Resubmission post review</td>
<td>Submitted</td>
</tr>
<tr>
<td>D3.4.1</td>
<td>Two training workshops on omics data deposition, grid processing, dissemination and</td>
<td>3</td>
<td>UoB</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D3.1.2</td>
<td>Report of the annual stakeholder meeting</td>
<td>3</td>
<td>UoB</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>D5.3</td>
<td>Operational gris/cloud allowing for combining data, tools, and compute VMIs. Most services available.</td>
<td>5</td>
<td>UU</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D6.4</td>
<td>Participating Biobanks and repositories connected to the VRC</td>
<td>6</td>
<td>EMBL-EBI</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D8.2</td>
<td>Modularised ISA model format: biospecimen centric schema, corresponding XML schemas, reference implementation guidelines and validation rules</td>
<td>8</td>
<td>UOXF</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D8.4</td>
<td>Signal processing and analysis data exchange format</td>
<td>8</td>
<td>UOXF</td>
<td>M20</td>
<td>Submitted</td>
</tr>
<tr>
<td>D8.4.1</td>
<td>Specifications for derived data matrices specifications and terminology for description of analysis and statistical results</td>
<td>8</td>
<td>UOXF</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D9.2.5</td>
<td>Portal Virtual Machine image 5</td>
<td>9</td>
<td>IPB</td>
<td>M24</td>
<td>Submitted</td>
</tr>
<tr>
<td>D9.3</td>
<td>Report API access to</td>
<td>9</td>
<td>IPB</td>
<td>M24</td>
<td>Submitted</td>
</tr>
</tbody>
</table>
Table 1. List of deliverables submitted until M19-M24

Milestones achieved in this period:

<table>
<thead>
<tr>
<th>No</th>
<th>Milestone name</th>
<th>WP</th>
<th>Lead participant</th>
<th>Delivery date</th>
<th>Means of verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS4.2</td>
<td>Establishment of second working group(^{13})</td>
<td>4</td>
<td>CIRMMP</td>
<td>M24</td>
<td>On project website</td>
</tr>
<tr>
<td>MS5.1</td>
<td>Operational GRID/Cloud</td>
<td>5</td>
<td>UU</td>
<td>M24</td>
<td>On project website(^{14}) (blog post)</td>
</tr>
<tr>
<td>MS9.2</td>
<td>Re-analysis of several high-profile data analysis with the PhenoMeNal infrastructure</td>
<td>9</td>
<td>IPB</td>
<td>M24</td>
<td>On project website (link to the deliverable)</td>
</tr>
</tbody>
</table>

Table 2. List of Milestones achieved until M19-M24

2.1 Work progress and achievements during this period

2.1.1 WP2 - Sustainability of PhenoMeNal

WP leader: University of Leiden (UL)

The main achievements during this period are:

\(^{13}\) http://phenomenal-h2020.eu/home/about/management-structure/work-groups/

\(^{14}\) http://phenomenal-h2020.eu/home/
- Organised a workshop “The Future of Metabolomics in ELIXIR” held April 25, 2017 in Frankfurt, Germany;
- Presented proposal for establishing a Metabolomics Use Case in the ELIXIR work programme 2019-2023 during the ELIXIR Heads of Nodes meeting on June 13, 2017;
- Proposal for establishing a PhenoMeNal Foundation during the PhenoMeNal annual consortium meeting in Hinxton on June 16, 2017;

A summary of the progress in this period:

**Task 2.1:** Report on other e-infrastructures, users, investments in the field of metabolomics, biomarkers and biobanks for supporting policy developments.

The original report on mapping of e-infrastructures and users submitted in M6 was updated to include more countries, policy developments and an overview of the potential champions or ambassadors who could enable reaching out to the broader community for accepting and implementing PhenoMeNal. The activity was an outcome of the recommendations from the review panel during the 1st periodic review held in March 2017. The report was a collaborative effort with input from people outside the PhenoMeNal consortium (see D2.3 Updated report on mapping of e-infrastructures, users, investments for supporting policy developments in the field of metabolomics, biomarkers and biobanks)

**Task 2.3:** Establishing and maintaining relations with publishers for supporting data deposition services.

In conjunction with the dissemination and outreach activities under WP3, we focused our activity on high profile publishers that are seen as leaders in their respective field. Considerable progress has been made with the Public Library Of Science publishers, specifically as part of a collaboration with an industry partner, as well as with Springer Nature’s Scientific Data journal and Gigascience publishers. For a detailed report, see revised D3.2 (Report on establishing and maintaining relations with publishers for supporting data deposition).

**Task 2.4:** PhenoMeNal business plan.
A series of activities have been undertaken in this reporting period, towards the preparations for a sustainability plan also submitted as report D2.4 (Business plan beyond the end of the project, M24).

A. Workshop on "Establishing a node for FAIRifying metabolomics and phenomics data in the European Open Science Cloud"

The European Open Science Cloud\(^{15}\) (EOSC) was launched by the European Commission (EC) in 2016 with the aim of ensuring that science, business, and public services reap benefits from the big data revolution. Following the recommendations from our SAB about collaborating closely with the EOSC, a 2-day strategic workshop on “Establishing a node for FAIRifying metabolomics and phenomics data in the European Open Science Cloud” was hosted at the LU from 9th – 10th March 2017.\(^{16}\) The workshop was jointly hosted with the GO FAIR\(^{17}\) initiative. As an outcome of the workshop, on 10 March 2017, the ‘Metabolomics Implementation Network’ of the European Open Science Cloud was launched. The network aims to implement the FAIR data principles\(^{18}\) in the field of metabolomics in Europe. This will help scientists to exchange research data and workflows and to combine metabolomics data with other phenotypic data, enhancing the efficiency and impact of European life sciences research. To quote Professor Karel Luyben, chairman of the GO FAIR taskforce, “The GO FAIR initiative welcomes the Metabolomics Implementation Network as one of the first GO FAIR Implementation Networks. We actively support this network as a key element in the European Open Science Cloud.”

The Metabolomics Implementation Network consists of key players in the European metabolomics community, including EMBL-EBI, Imperial College London, Leiden University, Oxford University, CIRMMP, CNRS, University of Jena, Janssen Pharmaceuticals, DSM, the Hyve, and linked Data experts. The group signed the ‘Leiden Declaration on FAIR metabolomics’ and has set itself the following objectives:

- To form and develop the open Metabolomics Implementation Network on metabolomics data and services; we invite others to join;
- To collectively implement standards compliant with FAIR principles in the wider research community and to actively communicate these;

\(^{17}\) [https://www.dtls.nl/go-fair-european-open-science-cloud/](https://www.dtls.nl/go-fair-european-open-science-cloud/)
\(^{18}\) [https://www.dtls.nl/fair-data/](https://www.dtls.nl/fair-data/)
• To work closely together with other communities on better capturing and understanding phenotypes enabling integrated approaches;
• To support the GO FAIR initiative with the widespread implementation of the FAIR principles in the EOSC, contributing to a global open internet of FAIR data and services.

Figure 1: Demonstrating the GO FAIR initiative during the workshop


The main goal of this project is to create a better harmonised virtual Metabolomics research infrastructure in Europe, by uniting the leading metabolomics facilities in Europe and at the same time educate the next generation of young researchers. MetaStar would build on the outcomes of PhenoMeNal with a strong focus and standardisation of wet analytical procedures. The project had a budget request of over €
5 million, for a duration of 48 months and included over 25 partners, including all major instrument vendors and academic outreach partners in Estonia, the Czech Republic, Greece and Portugal. Although the proposal was not funded in the end, it helped us plan for the future of PhenoMeNal and brought in new partners for potential future collaborations.

C. Workshop “The Future of Metabolomics in ELIXIR”

The workshop was organised in collaboration with ELIXIR-NL\(^\text{19}\) and ELIXIR-DE\(^\text{20}\) on 25th April in Frankfurt, Germany\(^\text{21}\). The one-day workshop showcased presentations on the metabolomics resources and services available through the ELIXIR nodes, ELIXIR use cases\(^\text{22}\) and their alignment with the ELIXIR platforms\(^\text{23}\). This was followed by discussion sessions on identifying the needs and challenges of the metabolomics community, formulating use cases and their alignment with the existing ELIXIR platforms. The highly interactive workshop resulted in a position paper for establishing a Metabolomics Use Case in ELIXIR. The latter was presented at the ELIXIR Heads of Nodes (HoN) meeting on June 13th 2017\(^\text{24}\) and has recently been accepted for publication in F1000. The Use case was well received and people recognised the bottom up effort from the Metabolomics community. As a follow up, an application for the formal proposal for metabolomics as a Use Case for ELIXIR is currently under preparation. This will be decided at the next HoN meeting in September 2017\(^\text{25}\).

\(^\text{19}\) https://www.elixir-europe.org/about-us/who-we-are/nodes/netherlands  
\(^\text{20}\) https://www.elixir-europe.org/about-us/who-we-are/nodes/germany  
\(^\text{21}\) http://phenomenal-h2020.eu/home/about/events/the-future-of-metabolomics-in-elixir/  
\(^\text{22}\) https://www.elixir-europe.org/use-cases  
\(^\text{23}\) https://www.elixir-europe.org/platforms  
\(^\text{24}\) https://www.elixir-europe.org/events/elixir-hons-retreat-meeting  
\(^\text{25}\) https://www.elixir-europe.org/events/elixir-hons-meeting
D. Establishing a PhenoMeNal Foundation

For the sustainability of PhenoMeNal, it is key to have a “organising entity” which can be at the core of PhenoMeNal community activities beyond the end of the project. During the annual consortium meeting in June 2017, it was decided that setting up a PhenoMeNal foundation would be a good way forward. Leiden University is in the process of preparing draft Articles of Association, the governance and a (partnership fee based) business model. This foundation is also going to be the driving force for the Metabolomics Implementation Network in the European Open Science Cloud (see under heading A).

E. Updated Report on mapping of e-infrastructures, users, investments for supporting policy developments in the field of metabolomics, biomarkers and biobanks (D2.3)
The report was submitted as a deliverable in M22 and included details about:

- Overview of potential users of the PhenoMeNal.
- Potential “champions/ambassadors”, biological and medical researchers influential in their field, who can enable us to reach out to a broader community for the acceptance and implementation of PhenoMeNal.
- Overview of potential contributors (instrument vendors, pharma companies, open source software companies)
- Overview of the (national) policy measures which can help sustain the PhenoMeNal infrastructure and how PhenoMeNal has positioned itself in these.
- Update on newly formed (e-)infrastructures
- Overview of the interactions with the (e-)infrastructures identified in the D2.1 report
- New country reports on Greece, Estonia, Czech Republic and Portugal, identifying potential users in these countries and relevant e-infrastructures

As a logical consequence of these mapping activities, an action plan has also been drafted for the strategic and systematic outreach activities in the period M24-M36.

Future Plans

- Establish a European Metabolomics Foundation before M30 as a “driving force” to steer follow-up activities of the PhenoMeNal project and the Metabolomics Implementation Network of the EOSC.
- Organise a workshop in Autumn 2017 to discuss potential follow-up project(s) and/or coordination actions in calls described in the Horizon2020 work programme 2018/2019.
- Organise a follow-up workshop of the Metabolomics Implementation Network of the EOSC in Winter 2017/2018 to discuss the progress.
- Execute the Dissemination and Outreach activities to potential users and champions as detailed in D2.3
- Execute the activities according to the business plan, D2.4

2.1.2 WP3 - Dissemination and Outreach

WP leader: University of Birmingham (UoB)

The main achievements during this period are:
- Establishing contacts with publishers for supporting data deposition.
- Continued work towards establishing a broader user community through workshops and champions/ambassadors.

A summary of the progress of the work in this reporting period:

**Task 3.1** We will initially employ the usual channels for the dissemination of PhenoMeNal tools and services, including scientific publications, workshops and presentations at the metabolomics conferences to reach the wider metabolomics community.

As part of its dissemination and outreach, the consortium presented its ongoing work in the form of presentations and posters\(^{26}\) in various national and international conferences.

<table>
<thead>
<tr>
<th>No</th>
<th>Date and Location</th>
<th>Event</th>
<th>Contribution</th>
<th>Attendee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>March 2017, Florence, CERM</td>
<td>Giornata di studio l'evoluzione digitale applicata al mondo dell'imaging, della medicina personalizzata e delle terapie</td>
<td>Presentation</td>
<td>CIRMMP</td>
</tr>
<tr>
<td>2</td>
<td>May 2017, Bordeaux, France</td>
<td>Inserm - Management and reuse of health data: methodological issues(^{27})</td>
<td>Presentation</td>
<td>EMBL-EBI</td>
</tr>
<tr>
<td>3</td>
<td>May 2017, Paris, France</td>
<td>Workflow4Experimenters 2017(^{28})</td>
<td>Presentation</td>
<td>CEA and INRA (organisers), ICL, IPB, UoB</td>
</tr>
<tr>
<td>4</td>
<td>May 2017, Leipzig, Germany</td>
<td>Central German Conference on Bioinformatics</td>
<td>Presentation</td>
<td>IPB</td>
</tr>
<tr>
<td>5</td>
<td>May 2017, EMBL-EBI, EMBL-EBI</td>
<td>EMBL-EBI Industry Meeting</td>
<td>Presentation</td>
<td>EMBL-EBI</td>
</tr>
</tbody>
</table>


\(^{27}\) [https://eudat.eu/events/inserm-workshop-246](https://eudat.eu/events/inserm-workshop-246)

<table>
<thead>
<tr>
<th></th>
<th>Date</th>
<th>Location</th>
<th>Event Description</th>
<th>Type</th>
<th>Organiser</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMBL-EBI,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>June 2017,</td>
<td>Sansepolcro, Italy</td>
<td>2nd NatMedDay</td>
<td>Presentation</td>
<td>IPB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>June 2017,</td>
<td>UK</td>
<td>ResOps: Delivering Science Across Clouds (EMBL-EBI+Sanger Institute)</td>
<td>Presentation</td>
<td>EMBL-EBI</td>
</tr>
<tr>
<td></td>
<td>EMBL-EBI,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>June 2017,</td>
<td>Giessen, Germany</td>
<td>DeNBI summer school on cloud computing</td>
<td>Presentation</td>
<td>IPB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>June 2017,</td>
<td>Brisbane, Australia</td>
<td>Metabolomics 2017</td>
<td>Presentation, Workshop</td>
<td>EMBL-EBI, IPB, INRA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>June 2017,</td>
<td>Brussels</td>
<td>OLEA R&amp;I Project Networking Event</td>
<td>Presentation</td>
<td>CIRMMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EMBL-EBI,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>July 2017,</td>
<td>Cagliari, Italy</td>
<td>BITS 2017: 14th Annual Meeting of the Bioinformatics Italian Society</td>
<td>Presentation</td>
<td>CRS4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>July 2017,</td>
<td>ITT, Prato</td>
<td>Focus on breast cancer</td>
<td>Presentation</td>
<td>CIRMMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>May 2017,</td>
<td>EMBL-EBI</td>
<td>EMBL-EBI day</td>
<td>Poster</td>
<td>EMBL-EBI</td>
</tr>
<tr>
<td></td>
<td>EMBL-EBI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>July 2017,</td>
<td>Cagliari, Italy</td>
<td>BITS 2017</td>
<td>Poster</td>
<td>CIRMMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3:** List of dissemination activities by the consortium

---

29 [http://natmed.aboca.com/program/](http://natmed.aboca.com/program/)
30 [https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/Education/courses/content/denbilss2017](https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/Education/courses/content/denbilss2017)
The project was also disseminated through meetings with other relevant projects, infrastructures and the user community:

- iNext and INSTRUCT meeting in Brno
- Knowledge sharing session at The Hyve\(^{31}\).
- BBMRI-ERIC, meeting of the working group on metabolomics.
- BBMRI-ERIC-QM-meeting.
- Cloud Computing & Big Data module on Software Engineering Programme (MSc).

Some of the publications produced during the period are:

- Opinion article (accepted) “The Future of Metabolomics in ELIXIR” in F1000 research (article number: 12342) in August 2017;

**Task 3.2** Using new and already established links we will build a network of e-infrastructure users and resource providers

**Development of User community**

A number of strategic workshops were held to create awareness about the PhenoMeNal services amongst the user community:

- FAIRifying Metabolomics and Phenomics data in the EOSC, Leiden
- PhenoMeNal Workflow Workshop, IPB Halle\(^{34}\)
- The Future of Metabolomics in ELIXIR, Frankfurt

---

\(^{31}\) [http://thehyve.nl](http://thehyve.nl)

\(^{32}\) doi: 10.1016/j.jbiotec.2017.05.018

\(^{33}\) doi: 10.1007/s11306-017-1213-z

These workshops gathered participants from across the user community with the outcome of generating recognition and creating a strong user base in Europe and beyond.

The Brisbane workshop was much appreciated by the participants, especially by vendors such as Sciex36, who were quite impressed and showed interest in reusing PhenoMeNal. Talks are currently in progress and they have promised to generate a “container” with their proprietary libraries to allow PhenoMeNal to convert their raw formats into open formats that can be digested by most tools in the infrastructure.

The awareness about PhenoMeNal e-infrastructure has also been raised through existing networks and communities that the consortium members are part of. The University of Luxembourg37 and the Luxembourg Elixir node38 have expressed particular interests in PhenoMeNal solutions specially due to their involvement in a large study on Parkinson’s disease39 (Dr Venkata Satagopam40, personal communication).

PhenoMeNal stakeholder, the Hyve41, a dutch company providing software support and solutions in the area of translational research involved in IMI eTRIKS project42 and other several dutch initiatives is currently in the process of evaluating and testing the PhenoMeNal e-infrastructure. The feedback so far has been extremely positive.

Efforts have been made by PhenoMeNal members under WP8 activities to reach out to Metabolomics users beyond the clinical world. Through long standing involvement with Plant Science research, we have contacted high-throughput phenotyping groups involved in Plant Breeding efforts. Taking the opportunity of an Elixir Belgium43 supported workshop held in Ghent in June 2017, the PhenoMeNal e-infrastructure was
informally presented. Following this preliminary contact, further discussions are planned.

**Task 3.4** PhenoMeNal will also build an intensive dialog between mass spectrometry and NMR instrument vendors, search engine providers, experimentalists, data resources, and publishers.

Following the feedback from reviewers during the periodic review, the consortium redefined its communication strategy with the publishers. The publishers were categorised into a) publishers that already promote the use of metabolomics data repositories and an open access philosophy and, b) those that have published metabolomics data but have not yet adopted a policy for making said data open access. The letter to group 1 will be intended to not only make the publisher aware of the services provided by the PhenoMeNal e-infrastructure but additionally requests the publisher considers *requiring* the submission of metabolomics data to an open-access data repository as a prerequisite of publishing (see Annex 7.2). For group 2, we would request that the publisher considers *recommending* the submission of metabolomics data to an open-access data repository, and our letter also increases their awareness that PhenoMeNal can standardise the data generation process by pipelining processing tools into citable re-runable workflows.

As part of its networking activities with the vendors, Bruker has continued to show considerable interest in the services provided by PhenoMeNal. A letter of intent to express their interest in working with the consortium has been signed. The project (through IPB) has applied for free AWS research credits and are now in the final stages to get these approved and applied to a common project Amazon account. This will enable testing of advanced, hence fairly costly, AWS features by consortium members. In addition, our partners at the ICL have been awarded a grant from Microsoft to conduct an evaluation of PhenoMeNal on their Azure cloud. The aim is to use the Azure cloud to test the scalability of PhenoMeNal resources with large data sets.

The consortium has also continued to concentrate its efforts in outreach to SMEs and the industry. The services provided by PhenoMeNal including live demonstrations on data analysis using its tools and workflows were showcased during the quarterly EMBL-EBI Industry meeting that was attended by experts from the Pharmaceutical and Agri-food. The meeting had a positive outcome with the possibility of some potential collaborations.

**Task 3.5** We will develop and disseminate a web tutorial and the software workflows established by the PhenoMeNal e-infrastructure.
The consortium continued to develop video tutorials and user documentation on PhenoMeNal tools and workflows. These were made available on the PhenoMeNal Wiki\textsuperscript{44}, “Help” page on the portal\textsuperscript{45} and the project website\textsuperscript{46}. A web tutorial has also been made available on the EMBL-EBI training page\textsuperscript{47} and the ELIXIR training portal\textsuperscript{48}.

**Task 3.6** We will provide training workshops on metabolomics on the e-infrastructure services and tools, dissemination and access at the conferences through partners outreach departments.

The following two hands on training workshops are planned for the next period:

- Scientific school on Cloud-based metabolomics data analysis and collaboration, 11-15th September, Sardinia
- Metabolomics workflows at EMBL-EBI\textsuperscript{49}, 30th October- 2nd November 2017

**Future plans**

- PhenoMeNal booth at the upcoming MetaboMeeting\textsuperscript{50} 2017 in Birmingham, UK
- PhenoMeNal satellite training course to this MetaboMeeting.
- Webinars and training courses.

2.1.3 WP4 - Interfacing with Biomedical European Infrastructures

**WP leader:** Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP)

The main achievements during this period are:

- PhenoMeNal participated as a panelist in the Round Table “Bringing together the bio-medical scientific communities: the role of research infrastructures”, co-organised by the H2020 project iNEXT and by the Virtual Research Community project West-Life\textsuperscript{51}.
- The document by the working group on Metabolomics in Systems Biology has been finalised and is being submitted to Metabolomics.

\textsuperscript{44}https://github.com/phnml/phenomenal-h2020/wiki
\textsuperscript{45}https://portal.phenomenal-h2020.eu/help
\textsuperscript{46}http://phenomenal-h2020.eu/home/training-online/
\textsuperscript{47}https://www.ebi.ac.uk/training/online/course/phenomenal-accessing-metabolomics-workflows-galaxy
\textsuperscript{48}https://tess.elixir-europe.org/materials/phenomenal-accessing-metabolomics-workflows-in-galaxy
\textsuperscript{49}https://www.ebi.ac.uk/training/events/2017/metabolomics-workflows
\textsuperscript{50}http://metabomeeting2017.thempf.org
\textsuperscript{51}https://twitter.com/WestLifeSB/status/867316716124155904/
Successful establishment of a 2nd working group engaging clinicians (MS 4.2 achieved). A draft document on Metabolomics in clinical applications is currently in progress.

A summary of the progress of the work in this reporting period

Task 4.1. Collecting and reporting on the requirements of research and/or industry centers that produce or make use of metabolomics

A detailed list of requirements was put together in the previous reporting period. In this reporting period we focused on fostering the interaction of PhenoMeNal with large infrastructures, especially those on the roadmap produced by the European Strategy Forum on Research Infrastructures (ESFRI), operating in the field of biological and medical sciences (BMS). To this end, we involved a representative of PhenoMeNal as a panelist in a Round Table event co-organised by iNEXT\(^{52}\) and West-Life\(^{53}\), which took place in Brno, Czech Republic, on May 24th, 2017. The INSTRUCT ESFRI was also involved in its organisation.

The round table discussed on the exploitation of BMS Research Infrastructures, both physical and electronic, by European researchers working in biological and biomedical sciences. The panelists included representatives of eleven BMS RIs and H2020 health related projects. The assessment of the services offered, the corresponding awareness by the target communities and the scientific community at large, existing/viable options for integrated offers, how to increase appreciation of RI services, and training of users were among the topics addressed. These are all themes relevant to the long-term impact of the achievements of PhenoMeNal and thus contributed to shaping up the development of PhenoMeNal services.

Task 4.2: Establish and convene working groups involving the PhenoMeNal consortium as well as participants in other biomedical infrastructure and research projects

Two working groups are currently operational in the frame of the PhenoMeNal initiative. The first working group, which has been active for nearly one year, has focused on the different aspects of the relationship between the discipline of systems biology (SB) and metabolomics. The second working group, which was established during the previous reporting period, has the aim of identifying the needs of the clinicians from a

\(^{52}\) http://inext.ceitec.eu
\(^{53}\) http://about.west-life.eu/network/west-life/west-life
metabolomics point of view and to advise on functions required in PhenoMeNal tools and workflows by clinical users.

Both groups have been working through a combination of teleconferences and off-line preparation of the drafts of their output documents, which will be released in the public domain as open access documents. The working group on systems biology has finalised its document in the month of August 2017, and is planning to submit it as a publication in a peer-reviewed journal. This would have the benefit of enhancing the visibility of the initiative, and therefore of the PhenoMeNal project as well, and of increasing its impact. In the next reporting period we will seek possible synergies between the outputs of the two groups.

Future Plans

- Submission of position paper currently being prepared in collaboration with the clinical working group.
- Explore synergies between the two working groups.
- Continue to present the achievements of PhenoMeNal at events involving large infrastructures, especially ESFRIs, and national/regional research and clinical labs with a potential interest in metabolomics.

2.1.4 WP5 - Operations and maintenance of PhenoMeNal GRID/CLOUD

**WP Leader:** Uppsala University (UU)

During this period, much of the work was focused on developments for the PhenoMeNal VRE M24 release (“Bucetin”).

The main achievements during this period are as follows:

**T5.1: Operation and Maintenance of the GRID/cloud infrastructure**

- VRE now has HTTPS support via CloudFlare.
- Storage systems: Heketi managed GlusterFS, improves reliability and community support.
- Galaxy: Update to version 17.05-pheno.
- Hardening deployment: Refactoring of KubeNow code. Improve maintainability.
- Testing: CI-testing is implemented for KubeNow immutable images and the complete PhenoMeNal VRE deployment.
A new KubeNow/PhenoMeNal client was developed: A turnkey-solution with all deployment dependencies now included in a single container image.

- KVM support for local deployment, more reliable than previously used Virtualbox.
- Single node local deployment due to advances in KubeAdm and KubeNow.
- nip.io domain names support for deployed VRE installations.
- Support for Kubernetes 1.6.x including Kubernetes role based access control (RBAC).

T5.2: Operation and Maintenance of the PhenoMeNal VRC

Reference installation at EMBL-EBI Embassy Cloud

This installation has been described in detail in earlier deliverables. The following advancements have been recently added:

- As of mid-August 2017, the EMBL-EBI Galaxy Public instance VRC, https://public.phenomenal-h2020.eu/ has been redeployed to a larger Kubernetes cluster running Kubernetes version 1.6. This instance was tested as development for a month before releasing it as the newer public instance, and stress tested with some large datasets. This instance reflects the Bucetin release.
- A secondary instance54, holds the cutting edge version of the PhenoMeNal Galaxy deployment.
- Both instances now have only secure access through the HTTPS.
- Accounts can now only be created by registered PhenoMeNal users (after going through the PhenoMeNal Portal’s SSO, which allows anyone with a Google, LinkedIn, ORCID or recognised academic institutional login to access). This is both the case for the public and the publicdev instances.
- The deployment has been tested with native NFS provided by EMBL-EBI on an hybrid cloud capacity, for future use.
- All the instances are monitored for uptime through the free service StatusCake (see Figure 3).

---

54 https://publicdev.phenomenal-h2020.eu/
**Figure 3:** StatusCake uptime monitoring for EMBASSY instances of Galaxy and PhenomeNal Portal running, both for production and dev. Downtimes for the Galaxy public instance in the older setup have been sorted, so we expect a better uptime since the new deployment in mid-August.

**EMBL-EBI TSI Portal mediated automatic VRE deployment**
- TSI API provides a solution that allow automatic VRE deployment
- Extensive testing has been carried out for the automatic VRE deployment

**T5.3: Provisioning of the PhenomeNal Services**
- Inclusion of Kubernetes Dashboard in PhenomeNal VREs, enabling a GUI for on-demand service management
- APIs of fluxomics tools have been adjusted to facilitate working with docker images and to implement them as a workflow in Galaxy.

**T5.4: Operation and Maintenance of a reference site for analysis on private data**
ICL has implemented a local instance of PhenomeNal to test the feasibility of processing sensitive NMR data in a local compute environment with BATMAN. The medbio server has 40 cores, 2 threads per core which means it has 80 CPUs. In terms
of RAM, it has 1TB. For the tests we have access to several internal datasets\textsuperscript{55}, including MESA (Multi-Ethnic Study of Atherosclerosis), to evaluate workflows such as the NPC pre-processing pipeline and BATMAN\textsuperscript{56}. Virtual clusters of varying sizes and capacities were used to test BATMAN. Initial tests show that a data set which takes 3 days to process on a desktop could be processed in just 6 hours on the medium scale cluster, suggesting that similar improvements can be expected by further increasing the number of cores. An abstract has been submitted to showcase this work at the upcoming MetaboMeeting 2017.

A second site has been established by UU within the CARAMBA group, which is a mass spectrometry research lab as well as an accredited lab for clinical diagnostics, PhenoMeNal VREs were deployed on local servers, Google Cloud, and on SNIC Science Cloud and used for analysis of data with varying degree of sensitivity. The most sensitive analysis were done on a local server (ncores: 72, RAM: 256 GB) and non-sensitive runs were executed on Google Cloud and on SNIC Science Cloud. Data sets analysed included cerebrospinal fluid (CSF) samples from thirteen relapsing-remitting multiple sclerosis (RRMS) and fourteen secondary progressive multiple sclerosis (SPMS) patients as well as ten healthy controls, all of which were processed on the local server. In addition, the experiments contained multiple quality control and identification samples which were pre-processed on Google Cloud, and on the SNIC Science Cloud. The downstream analysis was performed using the PhenoMeNal Jupyter notebook deployed on the local server.

**T5.5: Operation and Maintenance of Continuous Integration system**

The PhenoMeNal Continuous Integration (CI) system Jenkins, previously reported in D5.1, is operational and available at [https://phenomenal-h2020.eu/jenkins/](http://phenomenal-h2020.eu/jenkins/). Access to the CI is now available as well through secure connections. The CI now pushes all release tagged containers to the biocontainers registry, besides our usual PhenoMeNal container registry at EMBL-EBI.

**MS5.1 Operational Grid/Cloud**

Figure 4. Blog post on the project website

Future Plans

- Implement Microsoft Azure support for KubeNow and PhenoMeNal.
- Implement provider specific shared block-storage cloud integrations for Amazon, Google and Azure clouds, to improve performance.
- Implement Windows support for PhenoMeNal/KubeNow command line client.
- Develop the high-availability features of KubeNow when the dependency kubeadm supports it, or move to a deployment scheme that provides high-availability features.
- Move Luigi deployment in Kubernetes to the Helm package manager. Ultimately to have the whole PhenoMeNal software stack as a single Helm-chart.
- Work on understanding resource usage (CPU, IO, memory) for tools that are bottlenecks in workflows, to set sensible minimal requirements and limits per container in the Galaxy deployment. This should avoid cluster choke scenarios for workflows that parallelise too many jobs, as it can be the case for the main NMR and MS workflows.
- Add support for containers that are behind authenticated docker registries.
- Reduce deployment time for Galaxy, mostly linked to the database schema creation.

Augment our continuous integration server with continuous delivery, to have long term running instances of the PhenoMeNal Portal and PhenoMeNal Galaxy public instance roll-update when newer versions of the containers appear.

2.1.5 WP6 - PhenoMeNal Virtual Research Community Gateway

**WP leader:** European Molecular Biology Laboratory (EMBL-EBI)

The main achievements during this period are:

- The PhenoMeNal App Library/Service Catalogue was updated to include 45 applications, available via Galaxy workflows and Jupyter libraries through the PhenoMeNal VRE.
- Contribution to the EMBL-EBI TSI Portal with the functionality to deploy to OpenStack private cloud installations.
- Ability for the PhenoMeNal Portal to deploy to OpenStack, in addition to the previously available ability to deploy to Google GCE and Amazon AWS.
- Introduction of metadata storage to store user metadata and other settings for improved user experience.
- Migration of the different components of the PhenoMeNal Portal (portal frontend, metadata backend, wiki and app library) to individual containers, all open source and freely available. Containers follow the release process that is used for all our tools.
- Creation of a Kubernetes Helm chart for simplified and parameterised deployment of the Portal and all of its components on a Kubernetes cluster.
- Parallel production\(^58\) and development\(^59\) instances for the deployment of release and development versions of PhenoMeNal respectively. These instances have been deployed through the mentioned Helm charts and containers.
- Secure access to portal instances.
- Migration of the PhenoMeNal Portal to use a newer version of the EMBL-EBI TSI Portal for cloud deployment on the “Bucetin” release. The development version of the PhenoMeNal Portal references the development version of EMBL-EBI TSI Portal.
- Planning started with an UX expert for a third usability round.

A summary of the progress of the work in this reporting period

---


\(^{59}\) [https://portaldev.phenomenal-h2020.eu/](https://portaldev.phenomenal-h2020.eu/)
Task 6.1 Deployment of the PhenoMeNal VRC

OpenStack API Integration into PhenoMeNal Portal
We implemented the UI and backend for fetching the necessary OpenStack environment information through OpenStack APIs, based on user supplied OpenStack credentials. These settings are needed because each OpenStack installation has different machines, network names, floating ip pools and other differently named and configured elements, as opposed to consistent offerings from commercial cloud providers. Such OpenStack environment information can be used within the PhenoMeNal Portal for creation and deployment of a Cloud Research Environment on OpenStack installations. This feature improves the cloud deployment experience of the users and widens the set of supported cloud providers from two (Amazon AWS and Google GCE) to three (and OpenStack). The backend functionality for OpenStack support was contributed by PhenoMeNal developers to the EMBL-EBI TSI Portal API code base, and is the first external code contribution done to the EMBL-EBI TSI Portal.

Development of User Audit Metadata Storage
We implemented a backend for metadata storage in the PhenoMeNal portal. This feature not only persists user generated data when they use the PhenoMeNal portal, but can also simplifies adding new features in the future. The main purpose of the metadata storage backend is to keep track of whether the user has accepted all the required terms and conditions and the ELSI conditions necessary to use the PhenoMeNal infrastructure in regards to clinical data. The backend stores this data through a MySQL database. User credentials for cloud providers or institutional logins are not stored, and pairing with the user identity is solely based on the Single Sign-On (SSO) provided user ID, which makes the user identifiable through the Elixir security framework.

Containerisation of PhenoMeNal Portal Components
We created containers for all the separate components of the PhenoMeNal Portal (portal-frontend, wiki, app-library, backend). Portal containers follow the same specifications and guidelines as all other containers created in PhenoMeNal (ie. for tools), and we produce both development and release versions (tags) of each container through our continuous integration server.
For a simplified deployment on Kubernetes clusters, or on simple single machine Minikube\textsuperscript{60} installations, a PhenoMeNal Portal Helm chart\textsuperscript{61} was created. This deployment method makes use of all the portal containers to setup a fully working portal. We used this method to provision the current instances for production\textsuperscript{62} and development\textsuperscript{63} of the PhenoMeNal Portal. Further details of these developments can be found in our earlier deliverable D9.2.2\textsuperscript{64}.

**Task 6.7 Host VRC tutorials, online help and provide user support.**
All user documentation and tutorials can be accessed through the PhenoMeNal portal via the “help” page\textsuperscript{65}. User support and help was also provided through “contact us” form available on the project website\textsuperscript{66}.

![Tickets dashboard](image)

**Figure 4.** Help requests/tickets on the project website.

**Future Plans**

\textsuperscript{60}https://kubernetes.io/docs/getting-started-guides/minikube/
\textsuperscript{61}https://github.com/phnmnl/helm-charts/
\textsuperscript{62}https://portal.phenomenal-h2020.eu/
\textsuperscript{63}https://portaldev.phenomenal-h2020.eu/
\textsuperscript{64}https://zenodo.org/record/832590#.WaVAMNOGPnM
\textsuperscript{65}https://portal.phenomenal-h2020.eu/help
\textsuperscript{66}http://phenomenal-h2020.eu/home/help/
- Improvement of system logging information to end users during deployment, including detailed information of technical deployment aspects to aid debugging where this would be required.
- Complete OpenStack deployment ability in the user interface (UI).
- UI improvements originating from the schedule UX analysis to be completed in Q4 2017 / Q1 2018.
- Integrate a containerised version of the EMBL-EBI TSI Portal, to enable direct local deployments from the Portal UI behind firewalls.
- Support federated deployments, i.e. link geographical deployments.
- Support scaling of deployments, i.e. change the deployed servers.

2.1.6 WP7 - Privacy and Ethics

WP leader: Imperial College London (ICL)

All WP7 deliverables for the project have been successfully completed at month 12. We continue to monitor and promote privacy and ethics (ELSI) within the project.

The main achievements during this period are:
- Successful installation and testing of the PhenoMeNal infrastructure on local servers to facilitate handling of sensitive data.
- Incorporation of ISA-Tab within the NPC preprocessing pipeline, allowing standards-compliant handling of metadata.

A summary of the progress of the work in this reporting period

Task 7.1 Organise workshops involving Scientific and Technological Advisory board (STAB), the Management Committee (MC), projects/infrastructures funded explicitly in the ELSI area, as well as with existing expertise to devise best practices in handling sensitive human data, taking into account National and Institutional legal policies. (ICL, EMBL-EBI)

It was decided to arrange the final ELSI workshop in spring 2018, rather than summer 2017, in order to obtain feedback relating to the full release of PhenoMeNal in summer/autumn 2017. Preparations are under way.

Task 7.2 Establish and generate a robust workflow for data dissemination and disclosure taking into account guidance for task 7.1.
We have successfully installed and tested the PhenoMeNal infrastructure on a local server, in order to examine the feasibility of handling sensitive patient data behind an institutional firewall. We created a Kubernetes virtual cluster and deployed Jupyter, Galaxy and Luigi on it using KubeNow. This cluster was deployed on a powerful server located behind ICL’s institutional firewall and only accessible from the ICL network. We have been successful in running instances of the NMR processing software BATMAN in parallel on this cluster, with a large test data set (1980 NMR spectra) from the MESA consortium (see D9.1). See also D5.3. This is an example of taking the compute to the data to improve usability of PhenoMeNal within a secure ELSI compliant environment.

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

The NPC Preprocessing Pipeline has been augmented to read multiple metadata formats, including the international standard ISA-Tab. We have finalised a fully working version of the module that reads ISA-Tab and it is now being used internally within ICL. We currently only support ISA-Tab files that contain one MS Assay and/or one NMR Assay descriptions. Our next step is to support multiples of these.

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

ICL has prepared a chapter for the forthcoming book “The Handbook of Metabolic Phenotyping” (Elsevier, eds. Lindon, Holmes & Nicholson). The chapter focuses on large scale computing for metabolomics and explicitly addresses the issues of privacy and ethics with particular emphasis on the PhenoMeNal architecture.

**Future Plans**

ELSI is a dynamic and changing landscape and we will continue to monitor ELSI across the wider community (bioinformatics, clinical, societal) and implement best practice.

**2.1.7 WP8 - Data provenance, compliance and Integrity**

**WP leader:** The Chancellor, Masters and Scholars of the University of Oxford (UOXF)
The main achievements during this period are:

- Formalisation of Data Matrix for Univariate analysis and delivery of JSON data package
- Implementation of MAF format for metabolite identification and annotation by PhenoMeNal services
- Reference implementation guidelines for ISA archives.
- Development of new ISA configurations implementing reference guidelines (inclusion of ELSI annotation requirements, enable reporting of quality control runs).
- First Draft and usage of nmrML and tool environments within PhenoMeNal Workflows
- Submitted paper on nmrML to Analytical Chemistry journal.
- Paper draft: Missing data standards for workflow management systems
- Development of Naming Convention guidelines as a pragmatic interim solution for the time being where covering CV standards are lacking. Examples are WF module, Output file naming and the overall Workflow naming\(^67\).

Task 8.4.1: Harmonisation of data matrices and analytical results

The focus of the work over the period has been on driving standard adoption in the various components and workflows produced by PhenoMeNal. In particular, this meant liaising with developers to ensure consumption and production of standard compliant datasets and files.

- The first step towards regularisation of data matrix output from PhenoMeNal component tackled the reporting of metabolite identification and annotation from spectral processing tools from either mass spectrometry (MetFrag) or 1D NMR (BATMAN). In particular, WP8 developers’ participation in the IPB Workflow meeting in April 2017, pointed out that these key services (MetFrag for mass spectrometry or BATMAN R for NMR) produced an output that was not compatible with data deposition to EMBL-EBI’s MetaboLights\(^68\) repository, which mandates the use of the Metabolite Assignment File (MAF) format or possibly mzTab whose specifications for metabolomics applications are still in progress. Since then, both MetFrag and BATMAN R microservices have been upgraded to produce MAF data matrices. Furthermore, the work led to the refinement of the MAF file specifications to accommodate annotation requirements specific to the methodology deployed in the BATMAN component.

\(^67\) [https://github.com/phnmnl/container-galaxy-k8s-runtime/issues/51#issuecomment-318822471](https://github.com/phnmnl/container-galaxy-k8s-runtime/issues/51#issuecomment-318822471)

\(^68\) [http://www.ebi.ac.uk/metabolights/](http://www.ebi.ac.uk/metabolights/)
• A second line of efforts were aimed at dealing with the structuring of data matrices produced by univariate analysis modules developed by CEA. Following regular meetings and discussions between CEA and UOXF teams, specifications have been drawn up. Several JSON datapackages profiles have been defined. They clarify the amount of information and the semantic by providing a complete semantic markup of each column header, which up until now were simply shorthands. The notion of contrast, p-value, corrected p-value, confidence bounds (upper and lower) as well as the effect size estimator and false discovery rate are now fully accessible following their anchoring to STATO terms. STATO is a controlled terminology developed by UOXF which provided coverage for all that is related to statistical methods. The resource largely addresses the gap in knowledge representation which was affecting the domain up until now. Several JSON data packages had to be developed to accommodate the variations necessary to represent the specifics of different univariate statistical tests. For instance, effect size estimates are computed differently and these estimates can be a difference between means or a correlation coefficient. The JSON data packages make this explicit. The ability to provide unambiguous human and machine readable outputs from PhenoMeNal services is a major step towards delivering a FAIR infrastructure and will add to the appeal of the work delivered.

**D8.4.2: reference implementation guidelines and validation rules**

• Finally, a significant effort has been made to further drive delivery of essential experimental metadata descriptors. As the PhenoMenal infrastructure matured, it was possible to execute a batch analysis of the entire content of public repositories. This was done in the form of robustness testing for the tools but it quickly identified quality issues with the data available from public repositories. The failure of workflow execution was not down to faults with the tools but with the limited information available from the input metadata files. Collaboration between CEA, UOXF, IPB and EMBL-EBI pointed to ways to improve on how ISA archives are structured. WP8 is currently incorporating these observations into a set of functions as part of the ISA-API to help regularise ISA documents so the metadata payload can be used to integrate ideally with PhenoMeNal services. The ISA-API create mode in particular, is being developed and will be part of the coming release. It is supported by a series of ipython notebooks, easing training and use. Also, major advances towards D8.4.2, a guideline for reference implementation were achieved.

**Future Plans**
- Organise an implementation meeting during the fall of 2017 with the objective of making sure PhenoMeNal workflows consume and produce standards compliant data, including the implementation of JSON data packages defined under D8.4.1.
- Produce online tutorials showcasing the use of ISA-API for managing metabolomics data.
- Develop a PhenoMeNal workflow for data deposition to EMBL-EBI Metabolights repository.
- Organise a meeting in early 2018 with the Phenome Centres (Birmingham, London, Uppsala) about study design and run randomization.
- Coordinate with EMBL-EBI MetaboLights curation efforts on ISA document following the batch analysis performed by UOXF.
- Submit manuscripts detailing the work done.
- Follow-up on the FAIRification process of metabolomics datasets.

2.1.8  WP9 - Tools, workflows, audit and data management

**WP leader:** Leibniz Institute of Plant Biochemistry (IPB)

The main achievements during this period are:
- New LC-MS/MS Metabolite Annotation workflow developed and added to release.
- New 1D NMR Workflow developed and added to release.
- Improved development guidelines for testing.
- Performed batch processing of all applicable MetaboLights studies with the Univariate and Multivariate Analysis workflow.
- Add and improve documentation on infrastructure, testing, workflow and tools.

A two-day workshop was organised at the IPB on 04./05. April 2017. The 24 attendees were consortium members and external participants. Objectives of the workshop were to 1) Showcase workflows and progress since January/February demos, 2) Improve existing workflows to achieve broader applicability to more diverse studies, to add more functionality and to expand workflows, 3) Improve interoperability 4) Add FAIR information 5) Add and improve documentation, and 6) Improve quality and stability of our workflows through testing. Together, the workshop contributed to all tasks in WP9.
A summary of the progress of the work in this reporting period. The partners in WP9 developed new workflows, and have continued to work on the improvement of existing ones. They are part of the 2017.08 Bucetin release, and their availability on the public Galaxy instance and in newly deployed PhenoMeNal CRE instances is also reported in D5.3:

**Task 9.1: NMR Data processing pipelines**

A focus of the work in this task group was on assembling the early steps of NMR data processing, constructing an NMR1D workflow in Galaxy.

The **1D NMR workflow** performs the processing of 1D NMR experiments from raw data to a data matrix required for visualisation and statistics. The first tool imports NMR data from a relevant MetaboLights study of choice, which is then converted to nmrML, a standardised open source data format. The rnmr1d tool will run the fourier transformation (fft), zero-filling, line broadening, phase-correction and baseline correction. These operations are considered the pre-processing steps. After these steps have been performed, the alignment and bucketing of spectra takes place. The data matrix file is produced as output and then used as input in the next step that will
produce a stacked plot as an overview of all spectra. The Galaxy modules have been containerised by Kristian Peters (IPB), Reza Salek (EMBL-EBI) provided feedback and testing, and created the video tutorial\textsuperscript{69}. The Rnmr1D NMR data processing tool in "R" was written by Daniel Jacob (INRA) and integrated into the PhenoMeNal instance with the help of Kristian Peters (IPB) and Pablo Moreno (EMBL-EBI).

**Task 9.2: Data analysis pipelines**

Work has been continued to improve existing and add new workflows and corresponding tools.

**LC-MS/MS Metabolite Annotation**

This new workflow concerns MS/MS compound spectra extraction, annotation and identification using XCMS, CAMERA and MetFrag.

Metabolite identification in clinical studies is a crucial step when trying to understand e.g. a disease on the metabolomic level. The MetFrag workflow provides a first step into this direction as it annotates molecules from compound (metabolite) databases to MS/MS (tandem mass spectrometry) spectra. This annotation is based on the mapping of in silico generated fragments to the experimental spectra and scoring of these mappings based on several criteria.

Continuous integration results are made available on Jenkins\textsuperscript{70}

**Fluxomics workflow**

Development of the fluxomics workflow has continued, and the new version of the workflow and contained tools has been improved:

- The R-program CDF2MID developed at UB that reads netCDF files of new experiment extracts baseline corrected mass spectra of metabolites of interest and produces output in the format of the file exchangeable with the MetaboLights database.
- The R-program RaMID supports reading mass spectrometer-generated raw data from netCDF files, corrects baseline, and extracting mass spectra at the time point of maximum intensity. New functionality added to RaMID: if a netCDF file contains mass spectra for many metabolites, RaMID finds the spectrum of

\textsuperscript{69}https://www.youtube.com/watch?v=pHB9pN2jXMA

interest based on the known retention time and m/z value of the lightest isotopomer.

- MIDcor corrects the mass spectra extracted by RaMID for natural isotope abundance and overlapping. New functionality added to MIDcor: if the set of netCDF files processed by RaMID contain data from several experiments that should be analysed separately, MIDcor separates them between corresponding output files to facilitate the next step of workflow.
- Isodyn simulates the mass isotopomer distributions provided by MIDcor. New functionality added to Isodyn: It automatically adjusts the calculation of deviations from the experimental values to the list of metabolites for which mass spectra were registered in a given experiment, improving the fitting algorithm.
- Examples of particular fluxomic analysis based on stable isotope tracer data were added.

**Task 9.3: Data integration and Management**

Metabomatching\(^{71}\), a tool for automated metabolite identification in metabolome-wide genome-wide association studies (mGWAS) that use untargeted NMR metabolome data as phenotypes, has been implemented as a container. We also created the required Galaxy tool description, and metabomatching is now available in Galaxy on the public PhenoMeNal instance and in VRE deployments.

Continuous integration build history for metabomatching was made available on Jenkins\(^{72}\).

**Task 9.4: Data Integrity, Audit Trail and Quality Control**

One of the aims of data standards is that data can be analysed and checked in an automatic manner. We have applied the statistics workflow maintained by CEA. The workflow processes a MetaboLights study and generates univariate and multivariate statistics reports\(^{73}\). The script uses wft4galaxy\(^{74}\) which is accessing the API of a Galaxy instance within a deployed PhenoMeNal e-infrastructure. The analysis revealed several issues, which resulted in corrections on different levels. The statistics workflow itself was consequently improved to handle studies with multiple assays, and several studies with

---


\(^{74}\) [http://wft4galaxy.readthedocs.io](http://wft4galaxy.readthedocs.io)
inconsistent metadata annotations were reported to the MetaboLights curation team. The batch analysis code is also available\textsuperscript{75}.

Currently, 24 MetaboLights studies are applicable to this batch processing on a technical level. However, the statistical results must not be used without ensuring that the statistical analysis itself is appropriate for the study design. This has spawned efforts within WP8 to improve the representation of the study design, which will then help users to decide which data analysis is appropriate for a study.

The workflows developed in the framework of PhenoMeNal, in fact, complement the initial data, designed for the analysis, with additional data, thus ensuring the integrity of the whole assembly of data. For instance, the workflow for fluxomics starts the analysis from some fragmentary 13C tracing of some metabolites contained in netCDF files along with some unnecessary information. At the first step, RaMID extracts only useful information (mass spectra of metabolites of interest in the time points of their maximum) and then MIDcor complements this information with data corrected for natural isotope abundance. Next Isodyn complements this data with the corresponding set of metabolic fluxes and some other characteristics (parameters of the model). As a result, this analysis ensures integrity of the whole assembly: initial data, spectra, corrected spectra, and set of corresponding fluxes. The functions providing statistic analysis of the obtained data and validation perform quality control for the whole assembly.

**Task 9.5:** Provide continuous development snapshots with life-cycle management for building, testing, and deploying code, tools, and workflows

The continuous build infrastructure has already been in place for the first release (“Alanine”) of the PhenoMeNal infrastructure. Work has continued to streamline the release process as part of the life-cycle management.

**Task 9.6:** Migrate existing prototype software from proprietary environments to open, distributable software of high performance and scalability.

**BMFL Quality control pipeline, NL**

The BioMedical Metabolomics Facility Leiden (BMFL) quality control pipeline for targeted MS data has been re-written from MatLab to Python. The first node of the pipeline, the vendor data converter\textsuperscript{76}, has been added to PhenoMeNal Galaxy service and is accessible in the latest release. It supports conversion for Agilent, Sciex, Shimadzu, and Waters exported data. The remaining steps: blank-effect, inter- and

\textsuperscript{75} https://github.com/c-ruttkies/phmnl_statistical_workfow_api
\textsuperscript{76} https://github.com/phnmnl/container-ms-vfetc
intra-batch QC correction, and retention time shift reporting are planned to be added to the upcoming release (2018-02). Current development is focussed on adding interactive reporting for easy interpretation and providing support for other vendor exports.

The MRC-NIHR National Phenome Centre (NPC), UK
The NPC Preprocessing Pipeline is designed to read multiple metadata formats, one of these is ISA-Tab. We have finalised a fully working version of the module that reads ISA-Tab and it is now being used internally within ICL on sensitive data. We currently support ISA-Tab files that contain one MS Assay and/or one NMR Assay descriptions. Our next step is to support multiple assays in one study. The corresponding NPC Pipeline publication is in preparation, covering the whole NPC Preprocessing Toolbox. The toolbox has various modules and functionality which will make this publication reach out beyond just QC processing.

NPC Toolbox Release: the entire toolbox has been developed using Python 3 and will be open sourced once the publication is out.

MetaboLab, UK
The MetaboLab software (developed by partner UoB) has been packaged as a container image. This software is an integrated software package for NMR data processing and analysis, closely linked to the previously developed NMRLab software. It includes tools for batch processing and gives access to a wealth of algorithms. Algorithms within MetaboLab help to optimise the flow of metabolomics data preparation for statistical analysis. This software has been adapted to function without a graphical user interface in a more automated mode to be used here in a docker container. The MetaboLab container is designed to process 1D data from Bruker and Agilent Spectrometers, and provides automated phase correction, alignment of spectra, automated baseline correction, exclusion of unwanted spectral regions, segmental alignment, TSA/PQN scaling, bucketing, noise filtering and generalised logarithmic transformation.

Milestone MS 9.2
“Re-analysis of several high-profile data analysis with the PhenoMeNal infrastructure”. Here we aimed to analyse all applicable studies in the MetaboLights repository with our workflows.

As described in D9.3 in detail, we applied the univariate and multivariate statistical workflow to all available MetaboLights studies. The workflow processes a MetaboLights study and generates univariate and multivariate statistics reports. The analysis revealed issues which resulted in corrections on different levels. Currently, 24 MetaboLights
studies can be processed with the workflow on a technical level. However, the statistical results must not be used without ensuring that the statistical analysis itself is appropriate for the study design. This has spawned efforts within WP8 to improve the representation of the study design, which will then help users to decide which data analysis is appropriate for a study.

But the above successes also revealed several shortcomings. The UU and CARAMBA team tried to identify public studies that could be reproduced, and the initial conclusion is that there quite often something missing, such as necessary parameters, metadata, or sufficient information on the analysis steps. The lesson we learned from that exercise was that there are very few publicly available datasets that can be completely reproduced without additional information from the original authors, which also proved very hard to get.

Despite not being able to process all studies in MetaboLights, the approach showed the applicability of PhenoMeNal workflows to around 24 studies. We will continue to perform batch-processing of workflows on available data sets, to improve the applicability, stability and robustness of the workflows. Metadata curation is an ongoing activity by the MetaboLights curation team, and aligning metadata to enable the above mentioned processing is being retrospectively addressed.

Future Plans

- Integration of tools from external developers, such as Workflow4Metabolomics and OpenMS.
- Improving the interoperability of different tools, to allow e.g. the exchangeability of the XCMS feature detection with those from OpenMS.
- Export of tools developed in PhenoMeNal for use by other infrastructures, such as Workflow4Metabolomics.
- Generalise scalability investigations (see D5.3, work at ICL) to other workflows.
- Improve and increase the number of modules of the workflows in order to increase their functionality and to make their use more agile.

2.1.9 All WP's (joint effort)

PhenoMeNal Release

The 2017-08 release of PhenoMeNal, also known as “Bucetin”, has been released August 16th 2017. In comparison to the previous release, this was an upgrade of the 2017-02 beta release, with a richer set of tools. For a complete list of tools including information about their container versions and partner affiliations, visit the PhenoMeNal
GitHub page\textsuperscript{77}. We have improved the automated testing of tools, containers and workflows, and have introduced usability/user testing. Test results have been conducted and collected using a spreadsheet\textsuperscript{78} with linked documents describing test protocols (clinician/ technician\textsuperscript{79} and bioinformatician\textsuperscript{80}) and guides to coordinate test. The spreadsheet gives a complete overview and is used to find issues in our UI and/or functionality.

2.2 Project management during this period

2.2.1 WP1 - Management

WP leader: EMBL-EBI

The main achievements of the reporting period are:

- Annual Joint SAB-Stakeholder meeting
- Annual Consortium meeting
- Changes to the project website
- Submission of deliverables on time
- Hackathon for dissemination and knowledge management

A summary of the progress of tasks is reported below:

Task 1.1 Coordination at the consortium level of the “technical activities” of the project

Project Meetings

The management continued to manage the progress of the project through monthly consortium, dedicated WP and cross-WP and release update meetings. Minutes were documented and shared on Google drive.

Annual Consortium Meeting

A second annual meeting was organised by the coordinator at EMBL-EBI on 16th June 2017 with the following objectives:

- Review the current status of the project
- Develop plans for PhenoMeNal sustainability and broader Outreach

\textsuperscript{77} https://github.com/phnmnl/container-galaxy-k8s-runtime/blob/v17.05-pheno-bucetin/README.md

\textsuperscript{78} https://docs.google.com/spreadsheets/d/1o5fATAx4EXDq12LXDkm79EMFL8qwUb-JWq3orseqol/edit#gid=0

\textsuperscript{79} https://docs.google.com/document/d/1840zT1McxOy5qQUaU-z1MGsvlTGnn5hRyVTjAevNt64/edit#

\textsuperscript{80} https://docs.google.com/document/d/1jMguE9uH9OB4Fp5B2-yKoiK15_IUOrM_7E94LkFtJpU/edit#heading=h.izxhl6fxmab
The meeting was attended by 32 representatives from the consortium and included dedicated sessions to discuss financial reporting following feedback during the periodic review, project progress in terms of the tools and infrastructure, outreach and sustainability, and ELSI compliance. A summary of the key issues and outcomes from the discussions are listed below:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Issues</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Reporting</td>
<td>1. Underspends / overspends by partners</td>
<td>1. Partners to provide quarterly forecasts listing their proposed tasks and corresponding utilisation of resources. These will be supported by the actual resources consumptions</td>
</tr>
<tr>
<td></td>
<td>2. Resource deployments and cost claims</td>
<td></td>
</tr>
<tr>
<td>Tools and Infrastructure</td>
<td>1. Annotation of workflows</td>
<td>1. Need for reasonable naming conventions for the tools and workflows[^81]</td>
</tr>
<tr>
<td></td>
<td>2. Versioning of tools and workflows</td>
<td>2. Improved documentation/tutorials</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Need for individual UX sessions</td>
</tr>
<tr>
<td>Outreach and dissemination</td>
<td>1. Establishing contacts with publishers</td>
<td>1. Requesting publisher to either require or recommend submission of metabolomics data to an open-access data repository as a prerequisite of publishing[^82]</td>
</tr>
<tr>
<td></td>
<td>2. Engaging the user community</td>
<td>2. Planning of user workshops and training courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Prepare webinars</td>
</tr>
</tbody>
</table>

[^81]: [GitHub Issue](https://github.com/phnmnl/container-galaxy-k8s-runtime/issues/51#issuecomment-318822471)
2. Identify potential “champions/ambassadors”, biological and medical researchers influential in their field, who can enable to reach out to a broader community for accepting and implementing PhenoMeNal

<table>
<thead>
<tr>
<th>Privacy and ethics</th>
<th>1. GDPR implications</th>
<th>1. Use of anonymised data</th>
</tr>
</thead>
</table>

**Table 4**: Outcomes of the Annual Consortium meeting

**Joint Scientific Advisory Board and Stakeholder Meeting**

The joint SAB-stakeholder, held on 15th June 2017, gathered experts from the PhenoMeNal scientific advisory panel\(^3\) including experts from academia, clinical metabolomics and health care. The meeting had the objectives to discuss issues with data sharing, privacy and ethics while handling human data and to review the PhenoMeNal progress and receive recommendations from the participants on the way forward. The aspects of the project progress was reviewed in the light of live demonstrations of PhenoMeNal deployments and workflows for data analysis.

Recommendations were received by the participants in the following areas:

- **Tools and workflows**
  - PhenoMeNal needs to show use and importance of the tools offered in the form of user stories
  - Detailed tutorials/documentation of tools and workflows
  - Develop/support best practices

- **Communications and visibility**
  - Develop showcases of project impacts for key users to better communicate its results
  - PhenoMeNal should strategise its communication activities with the users

The details about the outcomes of the meeting have been submitted as D3.1.2 Report on the stakeholder meeting.

**Workshops and hackathons**

---

The consortium continued to host workshops and hackathons as part of PhenoMeNal’s infrastructure innovation.

**Figure 5:** Blog post from the PhenoMeNal Workflow workshop

**Task 1.2 The overall legal, contractual, ethical, financial and administrative management of the consortium**

All tasks related to the legal, contractual, ethical, financial and administrative management of the project are being performed as indicated in the grant agreement. The interim payments received from the European Commission post submission of the first periodic review were distributed amongst the partners.

---

84 http://phenomenal-h2020.eu/home/about/events/
Task 1.3. Coordination of knowledge management, IPS and other innovation-related activities.

Project Website:

The Project website was improved to showcase the services offered and links to the user documentation. The consortium continued to use the website to disseminate information about the latest developments and upcoming workshops/events.

Figure 6: New and improved Project website
As part of its contribution to open science, all project deliverables including presentations and posters were made available on the project website\textsuperscript{86,87} and Zenodo\textsuperscript{88} All software developed is fully open source and hosted on GitHub.\textsuperscript{89}

**Task 1.5 Maintaining communications with the commission**

The coordinator continued to be the prime link between the consortium and commission regarding the schedule of deliverables, submission of the first periodic report and other important communications.

Utilisation of resources by the partners:

![Consortium Effort Summary](http://phenomenal-h2020.eu/home/about/objectives/)

![Phenomenal Actual Efforts](http://phenomenal-h2020.eu/home/about/presentations-posters/)

![Phenomenal Actual Efforts](https://zenodo.org/communities/phenomenalh2020/?page=1&size=20)

![Phenomenal Actual Efforts](https://github.com/phnmnl)

---

\textsuperscript{86} http://phenomenal-h2020.eu/home/about/objectives/
\textsuperscript{87} http://phenomenal-h2020.eu/home/about/presentations-posters/
\textsuperscript{88} https://zenodo.org/communities/phenomenalh2020/?page=1&size=20
\textsuperscript{89} https://github.com/phnmnl
The actuals reported for M24 (August) is an estimate as the financial reporting cycle is not complete. Some partners financial reporting are estimates due to their internal policies. An updated version of this report will be made available in the next bi-annual progress report (D1.4.5)

KPI report
<table>
<thead>
<tr>
<th><strong>Project Objective</strong></th>
<th>To develop a model to ensure sustainability of PhenoMeNal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work Package</strong></td>
<td>WP2</td>
</tr>
</tbody>
</table>
| **Measures**          | # MoU or SLA from biomedical infrastructures, national or European metabolomics/phenome centers and/or biobanks  
                        | # LoI from an Industrial partner                          
                        | # Inclusion of PhenoMeNal into national funding schemes and EU or related projects |
| **Rationale**         | ● Building strong relationship with stakeholders to support policy development  
                        | ● Support development of the PhenoMeNal business plan |
| **Achievements**      | ![Achievements Bar Chart](image)                          |
| **Comments**          | ● MoU between EuroBioImaging Multimodal Molecular Imaging node and Instruct Core Center (signed) and National Phenome Centre (in progress)  
                        | ● Letter of support from ELIXIR                           
                        | ● LOI from Bruker (Appendix XXX)                          
<pre><code>                    | ● Inclusion in MetaStar, Go FAIR initiative               |
</code></pre>
<p>| <strong>Project Objective</strong> | To foster the worldwide adoption of PhenoMeNal through a wide range of outreach, dissemination, networking and training activities |</p>
<table>
<thead>
<tr>
<th>Work Package</th>
<th>WP2, WP3, WP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td># Agreement with publishers</td>
<td></td>
</tr>
<tr>
<td># Participation in e-science, biomedical, metabolomics meetings/conferences</td>
<td></td>
</tr>
<tr>
<td># Increase in members joining the RDA metabolomics IG</td>
<td></td>
</tr>
<tr>
<td># Positive feedback from users at workshops</td>
<td></td>
</tr>
<tr>
<td># Articles published</td>
<td></td>
</tr>
<tr>
<td># Working groups implemented</td>
<td></td>
</tr>
<tr>
<td># European infrastructures/projects in working groups</td>
<td></td>
</tr>
<tr>
<td># Documents produced by working groups</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td></td>
</tr>
<tr>
<td>• Raise awareness for services and tools provided</td>
<td></td>
</tr>
<tr>
<td>• Establishing relations with publishers for data deposition</td>
<td></td>
</tr>
<tr>
<td>Achievements</td>
<td></td>
</tr>
<tr>
<td>Conferences/meetings</td>
<td></td>
</tr>
<tr>
<td>Agreement with publishers</td>
<td></td>
</tr>
<tr>
<td>Publications</td>
<td></td>
</tr>
<tr>
<td>Feedback at workshops</td>
<td></td>
</tr>
<tr>
<td>Working groups</td>
<td></td>
</tr>
<tr>
<td>Representation of other projects</td>
<td></td>
</tr>
<tr>
<td>Documents in progress with</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>• Nine publishers have recommended MetaboLights as a repository for metabolomic data deposition</td>
<td></td>
</tr>
<tr>
<td>• NIH recommends MetaboLights for metabolomic data deposition</td>
<td></td>
</tr>
<tr>
<td>• Systems biology and clinical working group</td>
<td></td>
</tr>
<tr>
<td>• Representation of ELIXIR, ISBE, Hospital of Prato and Istituto Toscano Tumori, Hospital Clinic Barcelona(^\text{90}), Hospital of Uppsala(^\text{91}), UCL and EuroRec in the working</td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{90}\) [http://www.hospitalclinic.org/en](http://www.hospitalclinic.org/en)

\(^\text{91}\) [http://www.akademiska.se/en/](http://www.akademiska.se/en/)
<table>
<thead>
<tr>
<th>Project Objective</th>
<th>To establish technology for a water tight audit trail for the processing of human metabolic phenotyping data from the raw data acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package</td>
<td>WP5</td>
</tr>
</tbody>
</table>
| Measures          | # % Data Analysis Stack elements with high level of logging
# Interfaces available to view logs |
| Rationale         | ● The PhenoMeNal data analysis stack is comprised of multiple software components (Galaxy, Kubernetes cluster, analysis tool) in interaction to enable users to run data analysis. The higher this percentage, the higher coverage we have in terms of logging.
● The more interfaces available increases the ability of users/admins to obtain the logged information. |
| Achievements      | ● % Data Analysis Stack elements with high level of logging: 66%
● Interfaces: 3 (k8s dashboard, cli (command line interface), galaxy log view) |
| Comments          | ● Accessible on each VRE deployed. |

<table>
<thead>
<tr>
<th>Project Objective</th>
<th>To establish privacy-protection methods that allow working with highly sensitive molecular phenotype data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package</td>
<td>WP7</td>
</tr>
</tbody>
</table>
| Measures          | # Agreed disclosure form
# Data disclosure process and procedure established |
| Rationale         | ● To ensure ELSI compliance |
| Achievements      | ● User consent form for accepting a) ELSI requirements b) use |

2 manuscripts in preparation with systems biology and clinical working groups
of fully anonymised data  
c) PhenoMeNal terms and conditions  
- Moved to secure “https” site

<table>
<thead>
<tr>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 7. Screenshot of the portal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project Objective</th>
<th>To improve and scale up tools used within the infrastructure to cope with very large datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package</td>
<td>WP5, WP8, WP9</td>
</tr>
</tbody>
</table>
| Measures          | - # Services on project cloud: number of tools available in VRE deployments  
|                   | - # Release of production grade xml schemas  
|                   | - # Publication of functional specifications: The ISA configuration acts as such  
|                   | - # Adoption and implementation by academic and commercial entities                      |
| Rationale         | Community                                                                                |
| Achievements      | - 158 tools explicitly available through Galaxy (97 of them based on OpenMS).  
|                   | - Release of ELSI compliant PhenoMeNal ISA metadata configuration for patient based studies, with terminology constraints supported by the ‘Data Use Ontology’ adopted by GA4GH and EMBL-EBI’s European Genotyping Archive.  
|                   | - Ongoing work to validate ISA export from 2 Phenome centres (London and Birmingham)    |
Swedish SME Consid ([www.consid.se](http://www.consid.se)) evaluated KubeNow in 2017 for instantiating Kubernetes clusters on-demand.

<table>
<thead>
<tr>
<th>Project Objective</th>
<th>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 world-wide grid infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package</td>
<td>WP5</td>
</tr>
</tbody>
</table>
| Measures          | # Sites supporting project cloud  
# CPU’s available  |
| Rationale         | ● To ensure sufficient resources for adequate testing of PhenoMeNal VREs  |
| Achievements      | ● 6 sites are supporting project cloud  
● 168 CPUs available  |
| Comments          | Sites supporting the project cloud: de.NBI, ICL, IPB, EMBL-EBI, UU, Da Vinci European Data bank |

Project Objective
To use existing open source community standards, integrate tools, resources 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

<table>
<thead>
<tr>
<th>Work Package</th>
<th>WP6, WP5, WP9</th>
</tr>
</thead>
</table>
| Measures     | # Number of visits to PhenoMeNal VRC gateway/portal  
               # Number of available and documented standard protocols  
               # Number of online tutorials on the VRC web portal  
               # Support requests from the users and other projects  
               # Number of Users  
               # VMs available  
               # VM downloads  
               # High profile data analysis that can be fully reproduced using workflows |
| Rationale    | ● To track user growth over time  
               ● Provision of expert help and training |
| Achievements | ![Bar Chart](https://github.com/phnmnl/phenomenal-h2020/wiki) |
| Comments     | ● The metabolomics containers at Bioconductor maintained by IPB have been pulled a total of 287 times.  
               ● The standard protocols are guidelines for developers or advanced users[^52] |

3 CONCLUSION:

In this reporting period from March - August 2017, the consortium has continued to adhere to its original objectives and deliver on time. The feedback during the 1st periodic review was taken on board and implemented by the consortium.