



Deliverable 2.2

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	2
Work Package Title	Sustainability of PhenoMeNal
Deliverable Title	D2.2 A first version of the sustainability plan, which details business models options as well as the design choices and risk mitigation if any of the components fails to be sustainable
Delivery Date	M12
Work Package leader	UL
Contributing Partners	EMBL-EBI, ICL, CRIMMP, UL, UOXF, IPB, SIB
Authors	Merlijn van Rijswijk, Ola Spjuth, Steffen Neumann, Daniel Schober, Roger Mallol Parera, Reza Salek, Michael van Vliet, Thomas Hankemeier, Ken Haug, Namrata Kale
Abstract This document presents our current insights in the sustainability for the PhenoMeNal infrastructure, including a) business model options for economic sustainability as well as b) the design choices and risk mitigation if any of the components	



fail to be sustainable, the technical sustainability. From the economic sustainability point of view, the most relevant element is a broad acceptance and involvement of users of the e-infrastructure, to create a large user base as well as a good link with other e-infrastructure initiatives and research infrastructures in biomedicine. Different scenarios for the sustainability of PhenoMeNal are presented. From the technical sustainability point of view, PhenoMeNal is implementing a Virtual Research Environment (VRE) using an architecture based on workflow systems, micro services and containers. Collaborations exist with EGI, Indigo Datacloud and Elixir, to ensure that PhenoMeNal use technologies that are well supported, thus ensuring technical sustainability. PhenoMeNal is part of the long-term commitment of the leading European groups to work towards standardisation in a global perspective. PhenoMeNal, fuelling the Metabolomics Standards Initiative, under the umbrella of the International Metabolomics Society.



Table of Contents

1. EXECUTIVE SUMMARY.....	4
2. CONTRIBUTION TOWARDS PROJECT OBJECTIVES.....	5
3. DETAILED REPORT ON THE DELIVERABLE.....	6
3.1. ECONOMIC SUSTAINABILITY.....	6
3.2. TECHNICAL SUSTAINABILITY.....	15
3.3. RISK MANAGEMENT	16
4. WORK PLAN	18
5. DELIVERY AND SCHEDULE.....	20
6. CONCLUSIONS.....	20
7. APPENDIX.....	21



1. EXECUTIVE SUMMARY

This document presents our current insights in the sustainability for the PhenoMeNal infrastructure, including a) business model options for economic sustainability as well as b) the design choices and risk mitigation if any of the components fail to be sustainable, the technical sustainability.

From the economic sustainability point of view, the most relevant element for the sustainability of PhenoMeNal is a broad acceptance and involvement of users of the e-infrastructure, to create a large user base as well as a good link with other e-infrastructure initiatives (like ELIXIR) and research infrastructures in biomedicine (like BBMRI and [ISBE](#)¹). We want to position PhenoMeNal as *the* biomedical metabolomics e-infrastructure of the European Open Science Cloud (EOSC), and obtain long-term funding from the EOSC for the upkeep of our services. A paid "metabolomics only" approach is not likely to attract a large number of users, and we believe that a wider biomedical if not clinical scope is what we should aim for. The PhenoMeNal infrastructure will be endorsed and supported (by having seamless access/interfaces) by EBI and Instrument Vendors. Together with WP8 we assure that data parsers for all major vendor formats are available for use case workflows.

From the technical sustainability point of view, PhenoMeNal is implementing a Virtual Research Environment (VRE) using an architecture based on workflow systems, microservices and containers. Collaborations exist with [EGI](#)², Indigo Datacloud and Elixir, to ensure that PhenoMeNal use technologies that are well supported, thus ensuring technical sustainability. We avoid the development of new middleware as much as possible, since this would imply considerable maintenance efforts in the future, and could result in incompatibility with other infrastructures. At the same time, we avoid both dependencies on specific technologies and frameworks, as well as dependencies on specific data-centers. Both ensure that the VRE can be used without a particular vendor lock-in afterwards. Similarly, we adopt an "upstream first" principle, which means that our infrastructure developments are integrated into the used workflow tools. PhenoMeNal is part of the long-term commitment of the leading European groups to work towards standardisation in a global perspective. PhenoMeNal, but also the FP7

¹ <http://project.isbe.eu/>

² "Federated cloud - EGI." 2012. 30 Aug. 2016 <<https://www.egi.eu/infrastructure/cloud/>>



project [COSMOS](http://cosmos-fp7.eu/)³ are all initiatives fuelling the [Metabolomics Standards Initiative](http://www.metabolomics-msi.org/)⁴, under the umbrella of the [International Metabolomics Society](http://metabolomicssociety.org)⁵.

This report will be updated towards the end of the project for inclusion in the business plan. Input for this report has been provided by Industrial users on June 28th during an Industry workshop held at the Annual Metabolomics Society meeting in Dublin and by other Industrial and academic stakeholders on June 7th during the PhenoMeNal Stakeholder meeting in Rhodes.

2. CONTRIBUTION TOWARDS PROJECT OBJECTIVES

Continuity of PhenoMeNal after the project funding period is of paramount importance to assure its widespread usage and further development for the processing, analysis and information-mining of the massive amount of medical molecular phenotyping and genotyping data that will be generated by metabolomics applications now entering research and clinic

Key activities identified so far for achieving PhenoMeNal sustainability include, but are not limited to, the interaction with similar efforts, the involvement of the industry and the development of a business plan. Therefore, the activities described in the present report contribute to the achievement of the following objectives:

- **Objective 2.2** Establishing and maintaining relations with publishers for supporting data deposition services
- **Objective 2.3** Coordination with national and international related activities and support to the development of global and sustainable approaches in the field
- **Objective 2.4** Promotion of long-term sustainability, including the involvement of funders
- **Objective 2.5** Preparation of a business plan beyond the end of the project

Deliverables and Milestones:

- **Deliverable 2.2** - A first version of the sustainability plan, which details business models options as well as the design choices and risk mitigation if any of the components fails to be sustainable
- **Milestone 2.1** - Sustainability Plan released

³ <http://cosmos-fp7.eu/>

⁴ <http://www.metabolomics-msi.org/>

⁵ <http://metabolomicssociety.org>



3. DETAILED REPORT ON THE DELIVERABLE

3.1. Economic Sustainability

Overall goal

The overall goal is to assure the widespread usage, continuity and further development of PhenoMeNal, also after the end of the H2020 funding period.

Users

PhenoMeNal is primarily designed as a workflow for bioinformaticians analysing metabolomics and phenotyping data in the biomedical domain. In addition we want to facilitate end users, e.g. clinicians and epidemiologists, in applying metabolomics in the context of other phenotypic data in their daily practice.

User benefits, our niche and unique selling proposition

PhenoMeNal will provide data analysis services capable to work on exabyte-scale amounts of biomedical phenotyping data, for which no viable solution exists at the moment, by:

- i) facilitating the integration and interpretation of human metabolomic data in combination with most types of omics data PhenoMeNal, providing new R&D opportunities to EU researchers and SMEs and thereby increasing the attractiveness of the EU as a leading centre for advanced bioinformatics research;
- ii) providing novel and user-friendly statistical/computational solutions to gain new knowledge from human patient and cohort information, resolving bottlenecks encountered by end users (e.g. clinicians, epidemiologists) in timely fashion and turning this into everyday clinical practice, and;
- iii) creating an international research network able to raise, resolve, and disseminate results from end-user defined methodological questions, for the benefit of the European research effort and European competitiveness in this fast-moving field.

PhenoMeNal is designed to be able to interact with all instrument vendors' file formats and proprietary software.



Ecosystem

Before planning for sustainability, we have developed a clear understanding of the situation of PhenoMeNal within the e-infrastructure ecosystem (Figure 1). We have performed an extensive mapping of national and international e-infrastructures (D2.1, M6). PhenoMeNal operates in a complex bioinformatics ecosystem and we have and will continue to engage with these initiatives.

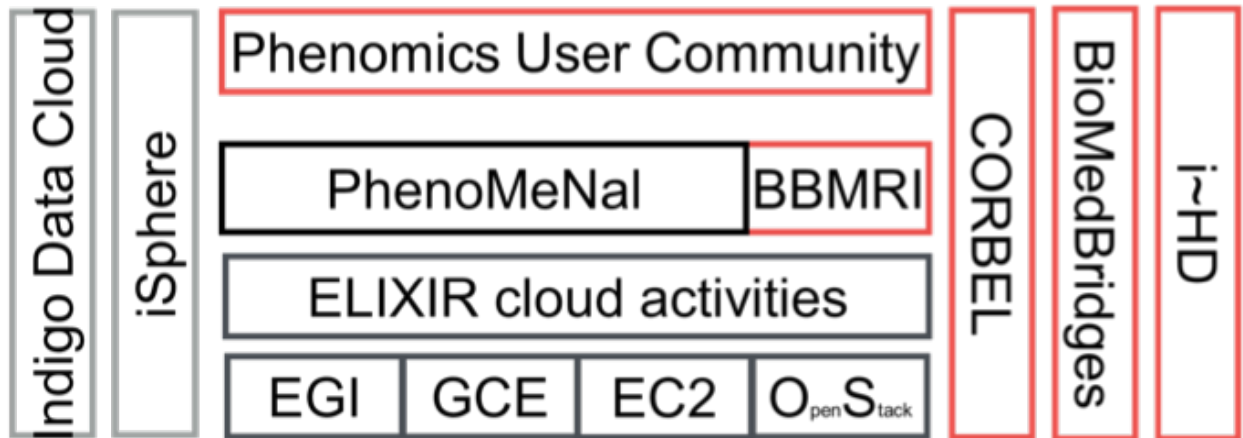


Figure 1. PhenoMeNal ecosystem comprised of research infrastructures in biomedicine and systems biology, as well as other e-infrastructures, which together provide cloud computing, data services, biological samples and medical technologies, to name a few. These existing initiatives will be involved through concertation activities and meetings in order to providing input and receiving feedback from working groups addressing activities of common interest.

Business model options

There are different business models for software in general, all of which we could consider.

- (one-off) Selling a software licenses for an individual user / institution or annual user fees
- Pay per use
- “Freemium” model, offering a free (limited) version of PhenoMeNal and a premium (extended, professional) version of PhenoMeNal against a license fee
- Development of services based on the open source PhenoMeNal

We think it is highly unlikely that users in research (especially from an omics perspective) are willing to pay for a “metabolomics-only” (software) suite. Therefore working with a license fee model, annual user fee or a pay per use, is likely not a wise route for broad adaptation of PhenoMeNal. On the other hand, although Horizon2020 regulations on e-infrastructures prescribe that virtual access to PhenoMeNal must be free of charge, there are no free



infrastructures. All costs associated with use, upkeep and development must be taken up by someone.

Funding sources

In general, we can divide funding sources for the upkeep and development of PhenoMeNal into the following categories:

- Direct public funds
- Funding by national institutions
- Direct private funds
- Indirect funding by research grants
- New funding sources

Direct public funds and funding by national institutions: Creating synergies between national and European public funding has been recognised as a priority⁶. Moreover, providing aggregate funding to national infrastructures in areas related to PhenoMeNal will be of great “EU added value”. Therefore we will continue to engage with national institutions and European Strategy Forum for Research Infrastructures (ESFRIs) (WP4).

Direct private funds: direct private funds can be generated in different ways:

- a) (consultancy) services (also to private companies) paying new developments partly from the surplus generated by services. Moreover, having an open development environment, where (open source based) software companies can offer (paid) continuous software development services. During our stakeholder meeting in Rhodes, one such company [The Hyve](http://thehyve.nl/)⁷ also involved in tranSMART, showed their interest in this. We will continue to engage with The Hyve and other (open source) software companies for that reason.
- b) Having an organised community, for instance organised in a foundation like the [tranSMART foundation](http://transmartfoundation.org/)⁸ and the [Phenotype foundation](http://phenotypefoundation.org/)⁹, could be highly beneficial, as this ensures a well-structured continuous development. "Partnership fees" of both public and private organisations could be one way to pay for this continuous development.

⁶ Report on the Consultation on Long Term Sustainability of Research Infrastructures, May 2016

⁷ <http://thehyve.nl/>

⁸ <http://transmartfoundation.org/>

⁹ <http://phenotypefoundation.org/>



Indirect funding by research grants: It is likely that after the project ends, we will need to do more development, especially in the integration of PhenoMeNal with other "OMICS"-infrastructures. Also here lies a considerable "EU added value", and therefore we will consider applying for a follow-up H2020 grant, in a broader biomedical "omics" perspective. In addition we will build on the long term commitment of EMBL-EBI to advance and support the field.

New funding sources: We will explore a new source of funding that will be "Closed pays for open": the use of the PhenoMeNal workflow is 'free of charge' when used in 'Open Science Mode (green button)', whereas embargo or complete privacy (for whatever reason) modes are 'paid'.

European Open Science Cloud

In a recent communication of the European Commission the [European Open Science Cloud](https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud)¹⁰ was announced as a major initiative to ensure science, business and public services reap the benefits of big data revolution, by bringing together current and future data infrastructures. In figure 2 an overview of the EOSC initiative is given.

¹⁰ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

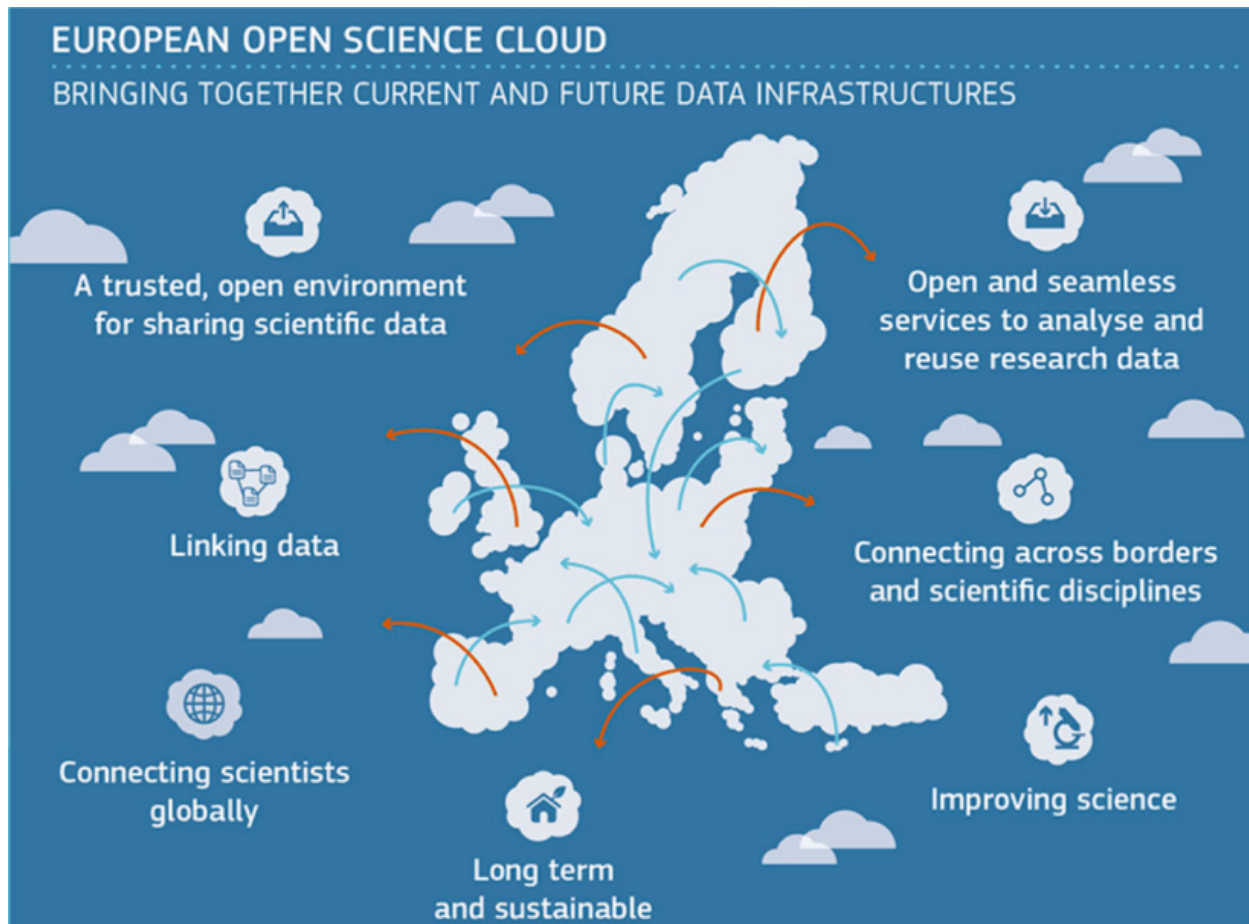


Figure 2. A schematic representation of the European Open Science Cloud initiative

Open and seamless services, like PhenoMeNal, to analyse and reuse research data are a key component of the EOSC. It is envisaged that the EC and member states will invest billions in the next year in realising the EOSC, and already plans are being made for Implementation Studies. PhenoMeNal intends to apply for funding, to become the (permanent) metabolomics service to analyse and research data in the biomedical field, in collaboration with other initiatives like BBMRI and other OMICS domains.

Sustainability strategies

There are four different sustainability strategies for PhenoMeNal, depending on the number of users PhenoMeNal is able to attract and the degree of involvement of the users. With the degree of involvement we mean, how important PhenoMeNal is for its users. This is depicted in the figure 3 below.

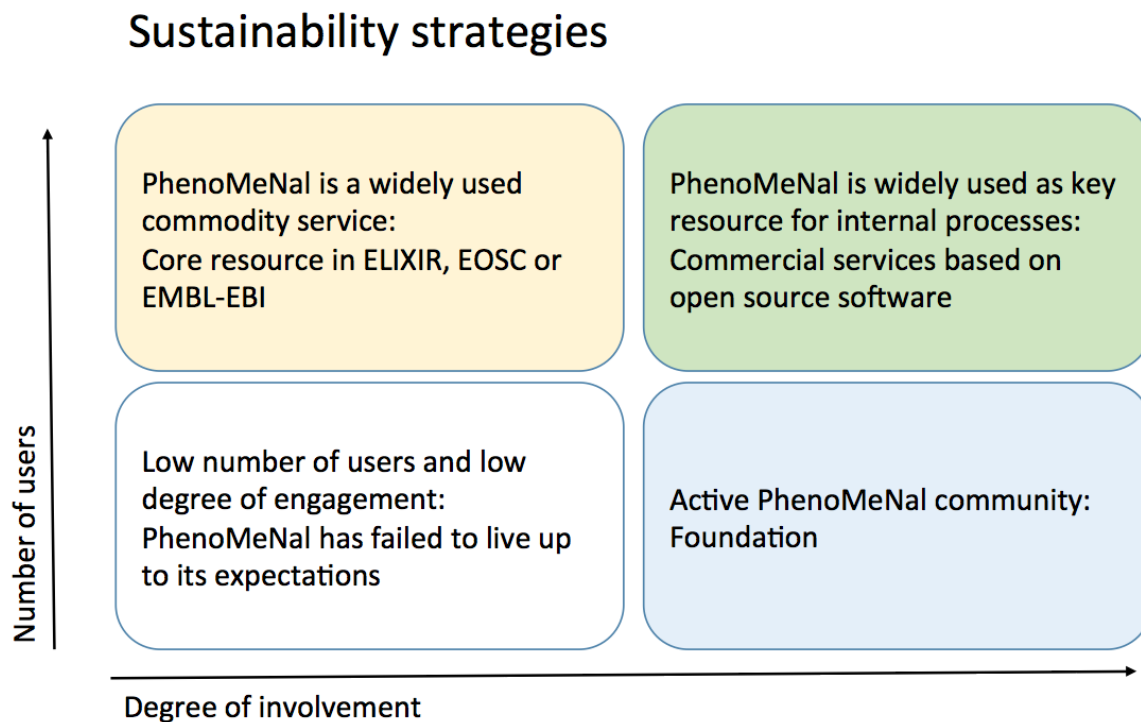


Figure 3. Sustainability strategies for PhenoMeNal

The term “degree of involvement” needs some explanation: if PhenoMeNal is a commodity for its users, for instance a software tool in the cloud that is sometime been used for specific data-analytical tasks, then the degree of involvement is said to be low. If the PhenoMeNal is a (business) critical tool for internal data analytical processes, the degree of involvement is said to be high. We distinguish four different strategies, based on the success PhenoMeNal is able to reach in the next years:

- **Commodity service** (high number of users, low involvement): PhenoMeNal can be positioned as a core resource in ELIXIR, the European Open Science Cloud or as a key service supported by EMBL-EBI. Examples of such core resources are the [Human Metabolome Database](http://www.hmdb.ca/)¹¹ or [Metabolights](https://www.ebi.ac.uk/metabolights/)¹², the metabolomics data repository hosted by EMBL-EBI.
- **Active community** (low number of users with high involvement): PhenoMeNal is a very important tool for a limited number of organisations (can be both academic, clinic and for instance pharma, biotech and nutritional companies applying metabolomics in their

¹¹ <http://www.hmdb.ca/>

¹² <https://www.ebi.ac.uk/metabolights/>



research). A foundation model, like the tranSMART or Phenotype foundation, could be a good way of organizing the further development of PhenoMeNal.

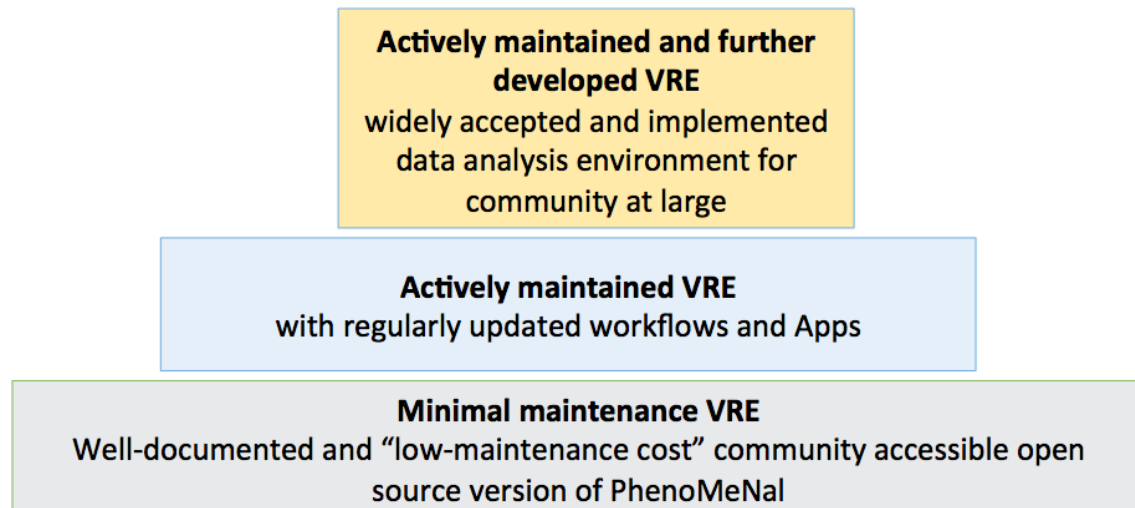
- **Key resource** (high number of users with high involvement). If PhenoMeNal is a big success, likely software development companies will start offering commercial services and instrument vendors will integrate (access to) PhenoMeNal in their data analytical software suites. Public and private parties, will invest from their own investments means in the further development of PhenoMeNal.
- **Failure** (low number of users, low involvement). If PhenoMeNal is not able to attract a user base and is not of (critical) importance of its users, it has failed and is likely not to be further developed.

PhenoMeNal is still being developed, but this matrix shows that the most relevant element for the sustainability of PhenoMeNal is a broad acceptance and involvement of users of the e-infrastructure, to create a large user base as well as a good link with other e-infrastructure initiatives (like ELIXIR) and research infrastructures in biomedicine (like BBMRI and [ISBE](http://project.isbe.eu/)¹³). Therefore we put a lot of efforts in testing the different components of the PhenoMeNal infrastructure, and getting inputs from users, both “expert” users like bio-informaticians as well as “knowledgeable” end users p.e. clinicians and epidemiologists.

¹³ <http://project.isbe.eu/>



Sustainability scenario's



We conceive different options/scenarios (VRE maintenance modes) depending on the degree of funding we are able to mobilize in the next years:

- 1) **Minimal maintenance of PhenoMeNal VRE.** In this scenario, active development of tools and workflows are not carried out, but the focus is on ensuring that the existing e-infrastructure components are available and can operate. An important task is the maintenance of the continuous integration server, and supervision of all existing projects ensuring that they conform to PhenoMeNal VRE standards and interoperability best-practices over time as individual tools are further developed by external partners. Should there be major updates that are not possible to adapt for PhenoMeNal VRE, such tools would have to be removed from PhenoMeNal VRE. There would only be very limited public compute resources contributed by the PhenoMeNal project to the community, and users would have to rely on public cloud providers, internal datacenters with IaaS capabilities, or international consortia such as ELIXIR/EGI.
 - Maintenance of e-infra to sustain PhenoMeNal over time: indicative effort 2 FTE (200 k€/year)
- 2) **Actively maintained VRE.** This constitutes a scenario where the PhenoMeNal VRE is supported by an active community, that apart from maintaining the continuous



integration server in scenario 1 above, also would provide regularly updated workflows and Apps and with active interaction and participation in tool development. The documentation is continuously improved, and outreach activities assist the community with further adopting the e-infrastructure.

- Maintenance of e-infra to sustain PhenoMeNal over time: 2 FTE
- Tool maintenance: 2 FTE
- Workflow maintenance: 1 FTE
- Documentation and outreach: 1 FTE
- Indicative effort (total): 6 FTE (600 k€/year)

3) **Actively maintained and further developed VRE.** This scenario includes, apart from all in scenarios 1 and 2 above, also future development of tools, new workflows in other scenarios, continuous work on data formats, refinement of virtual infrastructure components such as contextualization, more demonstrations on projects within sensitive installations, and interaction and harmonization with related projects (ELIXIR, EOSC, BBMRI and emerging projects on microservices in other domains e.g. genomics).

- Maintenance and further development of e-infrastructure: 5 FTE
- Tool maintenance and further development: 4 FTE
- Workflow maintenance and further development: 2 FTE
- Data format development: 2 FTE
- Additional demonstrators: 4 FTE
- Management, documentation, interaction and outreach: 3 FTE
- Indicative effort (total): 20 FTE (2 M€/year)

Raising awareness with policy makers

During the past Industry Workshop launched by this consortium within the 12th Annual Conference of the Metabolomics Society held in Dublin, we agreed to organize a future meeting explaining the benefits (in particular for clinical users) of this infrastructure to the European Commission (e.g. the Personalised Medicine unit), for instance with a presentation of Prof. Jeremy Nicholson from Imperial College London and initiator of the Phenome Centre network.



3.2. Technical sustainability

Upstream-first principle: For the establishment of the comprehensive and standardised e-infrastructure for analysing medical metabolic data we have to improve several existing data analysis software tools, and develop new ones only if they are missing links required in the workflows. In addition, we need to establish and adapt the middleware, and integrate the aforementioned data analysis methods. In all cases, we aim to contribute to the upstream software project, such that our developments become part of the normal code development, testing and maintenance. This way the software basis of our e-infrastructure will be available beyond the runtime of PhenoMeNal.

PhenoMeNal has made a strategic decision to implement VREs using an architecture based on workflow systems, microservices and containers, as we believe that this is the next-generation e-infrastructure that will also develop/emerge/mature in other fields. We base this architectural decision on interviews with scientific organisations such as EGI, Indigo Datacloud and Elixir, as well as by studying state-of-the-art virtual organisations in industry. In order to reduce risks within PhenoMeNal and establish a sustainable e-infrastructure, we have decided to follow these general rules:

1. **Avoid development of new middleware as much as possible.** We will primarily rely on open source tools from industry with a large user- and development community.
2. **Avoid dependencies on specific technologies and frameworks.** We will focus on supporting emerging standards, not specific implementations. We will always make sure that there are alternatives to critical components.
 - a. Example 1: We embrace Software Containers where Docker is now the dominating implementation, but we acknowledge and will keep the door open for other emerging container solutions.
 - b. Example 2: We demonstrate the use in the workflow system Galaxy, but also ensure that the microservices are also usable with e.g. Jupyter notebooks, again keeping the door open for other emerging solutions.
3. **Avoid dependencies on specific data-centers.** We will support several types of Infrastructure-as-a-Service solutions, and not be tied into e.g. only Amazon Web Services or Google Cloud Platform. This allows to also operate as part of the European Science Cloud.
4. **Implement continuous integration and delivery, validated by extensive testing and with clear maintenance responsibilities.** Of uttermost importance for a sustainable e-infrastructure is that all components need to be operational and work together. We are placing high focus on extensive test suites that ensure that all PhenoMeNal components -- such as virtual infrastructures contextualization, software tools and containers, and workflows -- are fully



covered with tests that signals to a responsible person if anything changes in e.g. dependencies, source, APIs that break interoperability with PhenoMeNal so that we can react on the raised issues.

5. **Work closely with end-users to iteratively define requirements.** During the design and development of the VRE and e-infrastructure we closely work with end-users, in particular data analysts who today uses many of the tools for metabolomics data analysis. These users are given early access to beta-versions long before public release, and their feedback and input ensures that the systems and interfaces we design will become usable and accepted by scientists in the targeted community.
6. **Engage and interact with other communities on e-infrastructure level.** We will interact closely with european initiatives such as ECP/EGI and Indigo Datacloud to drive these to support the features required in PhenoMeNal. So far, we have:
 - a. Engaged with ECP so that they develop the tools needed for PhenoMeNal to host a VRE using EGI portal.
 - b. Engaged with Indigo Datacloud (<https://www.indigo-datacloud.eu/>) by promoting the use cases in PhenoMeNal of multi-datacenter federation/integration on data and container level.
7. **Interact with other scientific communities to promote microservices as a basis of interoperability.** We will reach out to e-infrastructure projects in other life science domains, such as genomics, transcriptomics, toxicogenomics etc. So far we have:
 - a. Interacted with the OpenTox/ToxicoGenomics consortium (www.opentox.org and www.toxbank.org), which lead to an approved H2020-proposal (OpenRiskNet, currently in negotiations phase) that will establish an e-infrastructure based on microservices for toxicological predictions on the same underlying architecture as PhenoMeNal. There will be great synergy between the PhenoMeNal and OpenRiskNet consortia.

We here list the major software components required to implement the PhenoMeNal infrastructure, highlight potential risks and provide mitigation strategies for risk management:

Component	Risks	Risk management
Docker https://www.docker.com/	Security inadequate. New technologies emerge	Assess and document security. Make implementations container-technology agnostic.



MANTL http://mantl.io/	Missing critical features. Other projects become market leader.	Work with developers to ensure required functionality in roadmap. Evaluate and test alternative contextualization frameworks.
Kubernetes http://kubernetes.io/	Missing necessary features.	Keep up-to-date with state-of-the-art frameworks for container orchestration.
Mesos http://mesos.apache.org/	Missing necessary features.	Keep up-to-date with state-of-the-art frameworks for container orchestration.
Galaxy https://galaxyproject.org	Users prefer to work with scripts	Have a textual workflow system as alternative
Jupyter http://jupyter.org/	Users prefer to work with GUI	Have a GUI workflow system as alternative
Data Centers	National IaaS systems do not deliver services with sufficient quality	Ensure VRE also operates on public cloud providers supported by industry

Long-term commitment

PhenoMeNal is part of the long-term commitment of the leading European groups to work towards standardisation in a global perspective. PhenoMeNal, but also the FP7 project [COSMOS](#) are all initiatives fuelling the [Metabolomics Standards Initiative](#)¹⁴, under the umbrella of the International Metabolomics Society¹⁵. We closely work and coordinate our efforts with the International Metabolomics Society [Data standards task group](#), which aims to foster and harmonise efforts globally in enabling efficient storage, compression, exchange and verification of information within metabolomics datasets¹⁶. The data standards task group mandate is to engage with those who are producers or users of data formats such as; database providers, software engineers and instrument vendors working towards standardisation and agreements

¹⁴ <http://www.metabolomics-msi.org/>

¹⁵ <http://metabolomicssociety.org>

¹⁶ Salek, R et al. "Embedding standards in metabolomics: the Metabolomics Society data standards task group." *Metabolomics* 11.4 (2015): 782-783.



set by [MSI](#)^{17&18}, [HUPO-PSI](#)¹⁹, [COSMOS](#)²⁰ and other similar community wide accepted initiatives. This close collaboration and relationship would ensure long term applicability, usage, and sustainability of the resource built, coordinated with the larger metabolomics community, same time officially endorsed by [The International Metabolomics Society](#). Other Task group that our effort would also coordinate with are the [Data Quality Task Group](#)²¹. This task group aims to promote robust Quality Assurance (QA) and Quality Control (QC) in the metabolomics community through increased awareness via communication, outreach and education and through the promotion of best working practices. Improvements in data quality will facilitate data exchange, improve between-lab repeatability, enhance the usefulness of publications and improve submissions to metabolomics data processing, analysis and submission to repositories. [The Computational Mass Spectrometry task group](#) represents the Metabolomics interests in the [www.CompMS.org](#) initiative, which promotes the efficient, high quality analysis of mass spectrometry data with state-of-the art computational tools and algorithms through dissemination and training in existing, and coordination of new, innovative approaches. This TG also synergies between different application domains, in particular proteomics and metabolomics. Many members of the PhenoMeNal consortium are also members or leading various Task Groups and committees within the International Metabolomics Society and other international initiatives (HUPO-PSI). This close and long term relationship, as the community build up progress and engagement with the standards effort would ensure long-term technical suitability within the PhenoMeNal project.

4. Work plan

Deliverable D2.2 describes the initial sustainability plan for the PhenoMeNal infrastructure which is based on the first 12 months into the project. By now we have:

- Adapted a data-center agnostic approach where the VRE can be run on any IaaS provider such as public clouds and local servers

¹⁷ Salek, Reza M et al. "The role of reporting standards for metabolite annotation and identification in metabolomic studies." *GigaScience* 2.1 (2013): 1.

¹⁸ Fiehn, Oliver et al. "The metabolomics standards initiative (MSI)." *Metabolomics* 3.3 (2007): 175-178.

¹⁹ Hermjakob, Henning et al. "The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data." *Nature biotechnology* 22.2 (2004): 177-183.

²⁰ "COSMOS - COordination of Standards in MetabOlomicS | [http ...](http://www.cosmos-fp7.eu/)" 2012. 30 Aug. 2016

²¹ Bearden, Daniel W et al. "The New Data Quality Task Group (DQTG): ensuring high quality data today and in the future." *Metabolomics* 10.4 (2014): 539.



- Identified technologies and frameworks developed by industry to build upon in order to avoid developing our own middleware
- Established the Continuous Integration system as a focal point for development and testing for PhenoMeNal (<https://phenomenal-h2020.eu/jenkins/>), ensuring that the project has a single point of control for the integrated components (Submitted as Deliverable 5.1).
- a first static version of the PhenoMeNal VRE Portal (<http://phenomenal-h2020.eu/portal/>) which allows us to collect valuable feedback from two of our primary end-users: clinicians and bioinformaticians (deliverable D6.2),
- the infrastructure has reached a stage where proof-of-concept integration between Virtual Machine Images (VMIs) and containers is operational (deliverable D5.2),
- A first set of tools have been wrapped in containers and are available in publicly available repository for the research community.

In the next 12 months we will:

- Take part in upcoming European Open Science Cloud meetings and discussions on Implementation studies
- Together with WP5 and WP9 organize a testing workshop at EBI, developing best-practices for unit testing and integration testing of tools and containers. This will pave the way towards a sustainable software ecosystem in PhenoMeNal.
- Together with WP8 we organise a workshop with experts of Instrument vendors, as follow up of the Industry workshop in Dublin (Appendix 1), to assure that data parsers for all major vendor formats are available for use case workflows.
- Visit and interact with at least 10 national nodes of major infrastructure initiatives, ensuring that PhenoMeNal is well visible and that co-development and early adaptation takes place.
- Reach out to other microservice based architectures, such as the new OpenRiskNet initiative, to align design, architectures, and exchange experiences.
- Organise a two-day workshop for ESFRIs together with WP4.
- Interact with the tranSMART and Phenotype foundation on their business model and “lessons learned” from their operations.
- Interact with users in different phases of the development of PhenoMeNal to implementation measures to simplify PhenoMeNal usage for new and non-expert users.
- Organize a meeting explaining the benefits (in particular for clinical users) of this infrastructure to the European Commission



Utilization of resources:

The total PMs (person months) utilised in WP2 until M12 (inclusive):

Partners	EMBL-EBI	ICL	CRIMMP	UL	UOXF	SIB	UU
PMs	0.25	0.2	0.75	4	0.5	1.5	1

5. DELIVERY AND SCHEDULE

The delivery is delayed: No

6. CONCLUSIONS

From the economic sustainability point of view, the most relevant element is a broad acceptance and involvement of users of the e-infrastructure, to create a large user base as well as a good link with other e-infrastructure initiatives (like ELIXIR) and research infrastructures in biomedicine (like BBMRI and ISBE). Therefore we put a lot of efforts in testing the different components of the PhenoMeNal infrastructure, and getting inputs from users, both “expert” users like bioinformaticians as well as “knowledgeable” end users p.e. clinicians and epidemiologists.

From the technical sustainability point of view, PhenoMeNal is implementing a Virtual Research Environment (VRE) using an architecture based on workflow systems, microservices and containers. Collaborations exist with EGI, Indigo Datacloud and Elixir, to ensure that PhenoMeNal use technologies that are well supported, thus ensuring technical sustainability. PhenoMeNal is part of the long-term commitment of the leading European groups to work towards standardisation in a global perspective. PhenoMeNal, fuelling the Metabolomics Standards Initiative, under the umbrella of the International Metabolomics Society.



7. APPENDIX

Appendix 1: Report of PhenoMeNal industry workshop

Tuesday June 28th, Dublin Conference Centre

Conclusions

- There is a strong support of industry (in particular instrument vendors) for the PhenoMeNal infrastructure. If we can get Europe moving towards more standardized data solutions, we build a lot of value.
- The **presentation on the PhenoMeNal e-infrastructure** of Thomas Hankemeier **will be shared** with all participants
- PhenoMeNal would like to know which **technical requirements** for PhenoMeNal to be optimally interfacing your products, we will send you the details in the follow-up email.
- All industrial partners are invited, if interested, to **nominate one person per company for the PhenoMeNal Industry Panel**. This Industry panel will be a sounding board for the consortium, and be consulted at least once a year on the development of the e-infrastructure.
- A suggestion is to organize a meeting explaining the benefits (clinical) metabolomics to the EC (eg Personalized Medicine unit), for instance with a presentation of Jeremy Nicholson of Imperial College London and initiator of the Phenome Centre network.

Meeting report

- The aim of the meeting is to create a lasting interaction with industry, in order to raise awareness of the PhenoMeNal initiative and to ensure optimal interoperability of PhenoMeNal infrastructure and instrument vendor's data formats and tools.
- Sciex: is the privacy handling of PhenoMeNal in alignment with EMEA and FDA regulations? Chris Steinbeck (EBI) answers that we take a pragmatic approach. All encapsulation will work through standardized workflows, which can be fired up on Amazon web services and Google cloud or on your own local stack. We bring compute to the data. It is the responsibility of the one using the data to comply with the regulations. We build the tools and follow the requirements posed by others. ISO and FDA certification is a whole other world. We don't expect to get certifications done. Maybe this can be done in collaboration with industry. Thomas Hankemeier adds that the European Science cloud will also work towards this, and we don't need to solve this within the project, but need to be compliant. Sciex comments that they work together with Illumina for our ONE OMICS products with cloud computing, and they are well aware of the sensitive data issues in cloud computing. Every lab has its own IT infrastructure and workflow. If we can get Europe moving towards more standardized solutions, we build a lot of value. This approach PhenoMeNal is great.
- Thomas Hankemeier raises the issue that we as PhenoMeNal have taken the design perspective to use open source tools. He explains that PhenoMeNal will not use so-called "viral" open source licenses. What does this design choice look like from a vendor perspective? It will



create chances for new business models, for instance service companies like The Hyve (www.thehyve.nl) offering commercial (development) services around open source software, but also on the compute site new business models appear, eg. hybrid business models (local compute + industry/company capacity) and new compute rent business models. Vendors could offer their cloud capabilities in parallel to licensed tools.

- The general consensus from industry is that PhenoMeNal needs to be interoperable, meaning that PhenoMeNal can work with instrument vendor's tools and file format formats, so that they can seamlessly be combined with each other. Proprietary tools can thus be linked to the PhenoMeNal e-infrastructure. The only constraint is that we need to make sure, as a common challenge for academia and industry, that "interoperable formats" do not increase file sizes 2-10 fold, as now sometimes is the case. Sciex asks how to make sure that request/bugs are handled in open formats? Thomas Hankemeier answers by working together if issues arise, and PhenoMeNal will have to plan and react accordingly. Steffen Neumann "PhenoMeNal scales beyond metabolomics, but also for instance for proteomics developments. We link as well to these developments, we don't want to reinvent the wheel. Sciex: [Skyline](#) is a good example, where open source software has been developed, that is now a standard for industry as well. Chris Steinbeck suggests to organize a hackathon, with informal participation of (data) experts from industry, for instance to share their knowledge on interoperability, pe in the conversion of proprietary data formats to open source data formats.

- Chris Steinbeck offers all companies to nominate one person to participate in the PhenoMeNal industry advisory board. This will be an informal group, with a low workload, which will be used to get more advice from industry on the future directions of the PhenoMeNal initiative. All companies respond positive to the initiative.

- Sciex: Who are your users? Thomas Hankemeier answers, there are different types of users: (1) researchers, but we also want to reach out to (2) clinicians, which in the future will be users of metabolomics data. It is our ambition that omics and clinical data can be combined, this requires integration on the different levels. That is the reason why PhenoMeNal is linked ELIXIR and European Open Science cloud, and therefore we also need High Performance Computing. Chris Steinbeck adds "We need much more human data, so that we can inform data on their decision making in the clinic. Now: metabolomics data are primarily used for early statistical analysis, with the Phenome Centre network this will be much broader. Sciex: visualization poses lots of challenges how to reflect data to physicians. Bruker: do you engage with doctors? Chris Steinbeck "Yes we already did organize a meeting for clinicians and this is in fact a continuous process. We had Manfred Spraul at the last consortium meeting. We have the same vision, we need to resolve the roadblocks together. We will work in a transparent and open manner. We have the clinical research as our customs for this project."

- SCIEX: do you engage with EC and politicians on the Life Sciences OMICS, and clinical metabolomics in particular? The HoU Irene Norsted of personalized medicine, is engaging mostly with genomics. Marta Cascante explains that she is involved in an advisory committee on personalized medicine. The suggestion is to organize a meeting explaining the benefits (clinical)



metabolomics and the Phenome Centre network to the EC, for instance with a presentation of Jeremy Nicholson.

· Bruker: Do you have a plan to export this project to Asia (and North America)? Chris Steinbeck explains that we have strong partners in Asia (Masanori Arita) and US (NIH Metabolomics cores). The concept is global. Sciex: are you aware of similar initiatives in the world? Chris we have strong links with NIH networks, they develop also some (local) solutions, but there are no large-scale projects going on in the world. It is bad to sell “paper-ware”, but we need to outreach based on a product we have. We will have this by the end of the year. MetabolomeXchange is a global network, where we build upon.

· Shared SOPs are also important (Bruker), to ensure data quality, by controlling sampling and the data acquisition process. Thomas: in repositories we now ask for the analytical methods being used in the data acquisition. Ideally we would require an error model per metabolite. Quality of data is key. That is not the focus of PhenoMeNal, we will rely on global agreements on this. Phenome centre network is also on developing these SOPs. Sample preparation is key, and needs global training Partners in PhenoMeNal are linked to other projects and initiatives. We make sure that SOPs can be uploaded.

List of participants

1. Daniel Stoessel	Metabolomics Discoveries
2. Ken Haug	European Bioinformatics Institute
3. Chris Steinbeck	European Bioinformatics Institute
4. Baljit Ubhi	Sciex
5. Jörg Dojahn	Sciex
6. Jean-Baptiste Vincendet	Sciex
7. Noriyuko Ojima	Shimadzu
8. Aiko Barsch	Bruker
9. Kim Colson	Bruker
10. Claire Cannet	Bruker
11. Ian Reah	Non-linear Dynamics
12. Steffen Neumann	IPB Halle
13. Merlijn van Rijswijk	Leiden University
14. Thomas Hankemeier	Leiden University
15. Marta Cascante	University of Barcelona
16. Lee Gethings	Waters
17. Tim Ebbels	Imperial College London

In addition ThermoFisher and Agilent are informed and aware of the project