

Guest posting: Building a PhenoMeNal metabolomics e-infrastructure

Scott Edmunds 

Published July 19, 2016

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/51w44-grp73>

Keywords

Medicine, Technology, Big Data, Guest Post, Medical Data
Feature Image

Copyright

Copyright © Scott Edmunds 2016. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Building an open, community-supported, e-infrastructure for medical metabolomics data.

The European Union's Horizon 2020 research and innovation programme is funding the [PhenoMeNal](#) (Phenome and Metabolome aNalysis) project that aims to support data processing and analysis pipelines for molecular phenotype data generated from metabolomics applications. We aim to build an open, community-led and community-supported e-infrastructure by leveraging existing cloud infrastructures, tooling, and data repositories, under one umbrella of services dedicated to the European biomedical community to begin with, and eventually worldwide. However, the experience and tooling can be applied to non clinical settings.

The project was formulated to address the complexity and high volumes of data being generated, collected, and analysed that are quickly going beyond current data management and computational capabilities. For example, it is estimated that a single National Phenome Centre managing only around 100,000 human samples per year might generate a velocity of data amounting to more than 2PB annually. And this is a conservative estimate.

When thinking of the near to middle-term future where phenotyping efforts on a population-wide scale are on the horizon, such e-infrastructures will need to deal with Exabyte-scales of biological phenotyping data. It is essential to build such data infrastructures, in particular for metabolomics phenotyping, as there is an urgent need to improve the understanding of the causes and mechanisms underlying health, healthy ageing and diseases beyond what can be done solely with genomic approaches.

Strategy for knowledge generation enabled by PhenoMeNal e-infrastructure. Source: PhenoMeNal consortium

Strategy for knowledge generation enabled by PhenoMeNal e-infrastructure. Source: PhenoMeNal consortium

There are also additional challenges beyond simply addressing the volumes, velocity and variety of data. Ethics and privacy are of central concerns in the project since we aim to handle clinical metabolomics data, and we are engaged with experts on ethics, legal, and social implications of biological data. The project also aims to build sustainable infrastructure that will live-on beyond the three years of EU funding awarded to PhenoMeNal. This involves ensuring that all of the infrastructure we develop is open source, well documented, and community supported, and that we engage with existing research e-infrastructures across Europe and worldwide.

****Building part public-part private clouds**

****So how are we building the PhenoMeNal e-infrastructure to address these challenges?**

In our first year, we have completed plenty of groundwork towards setting up exemplar pipelines on public and private cloud computing infrastructure. We are working on setting up the computing infrastructure across European boundaries, with part hosting on the European Bioinformatics Institute's EMBASSY Cloud in the UK, on an installation at UPPMAX in Uppsala in Sweden (infrastructure previously [covered by GigaScience](#)), and within CRS4's datacenter in Sardinia, Italy. Alongside these, we are also working with Google on creating a hybrid public-

private cloud infrastructure that we can then demonstrate the migration of computing tools to data that needs to be ring-fenced within institutions (e.g. within a clinic's own infrastructure where data cannot be moved) while at the same time still leveraging the power of the public cloud.

We have also been working on packaging metabolomics software to become 'cloud-ready'. We are utilizing the [Docker system](#) for containerizing software, where we have already packaged tools such as [BATMAN](#) (Bayesian AuTomedated Metabolite Analyser for NMR spectra), [OpenMS](#), [XCMS](#), as well as data management tooling such as the [ISA metadata framework](#) API.

All of this work is in collaboration with clinical and research sites, to support some example use cases. We work closely with clinical metabolomics centers such as [CARAMBA](#) (Clinical Analysis and Research Applying Mass spectrometry and Bioinformatics at Akademiska) at Uppsala University (Sweden), the [MRC-NIHR National Phenome Centre](#) (UK), the [Phenome Centre Birmingham](#) (UK), [Hospital Clínic of Barcelona](#) (Spain), and the [Netherlands Metabolomics Centre](#) (Netherlands).

****Survey on Metabolomics Data: Call for participation**

****To gain a better understanding of the requirements for data infrastructures for metabolomics research, PhenoMeNal has commissioned a survey to gain insight into current practices, future needs, and opinions, on how metabolomics research data is managed, published, and disseminated. The survey welcomes contributions from the community of researchers, including clinicians, and also from those involved with data management, engineering and software development involved with research and application of metabolomics.**

If you work in any role relating to metabolomics, be it in clinical application, basic research, or in technical support, please help us build PhenoMeNal as an open, community-led, metabolomics infrastructure by letting us know your thoughts.

You can contribute to the survey (closes end August 2016) here: <https://www.surveymonkey.co.uk/r/metabolomics-data>

Acknowledgements

The PhenoMeNal project is led by a consortium of 14 partners at: the European Molecular Biology Laboratory-European Bioinformatics Institute (UK), Imperial College of Science, Technology and Medicine (UK), Leibniz-Institute of Plant Biochemistry (Germany), Universitat de Barcelona (Spain), The University of Birmingham (UK), Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (Italy), Universiteit Leiden (Netherlands), The University of Oxford (UK), Swiss Institute of Bioinformatics (Switzerland), Uppsala Universitet (Sweden), Biobanking and BioMolecular Resources Research Infrastructure-ERIC (Austria), Commissariat à l'énergie atomique et aux énergies alternatives (France), Institut national de la recherche agronomique (France), and Centro Di Ricerca, Sviluppo E Studi Superiori in Sardegna (Italy).

The post [Guest posting: Building a PhenoMeNal metabolomics e-infrastructure](#) appeared first on [GigaBlog](#).