

Data management pilot project report

Report date	2016 12 20	Report #	1
Project title	ActionableBiomarkersDK		
Grant holder	Prof. Søren Brunak, KU (UCPH)		
Partner institutions	SDU + DTU + Zealand University Hospital		
Project start	01.08.2016	Project end	30.06.2018

Details about the reporting

On site visit (PI's office) December 20th 2016, 11.00 – 12.00

Participants:

Project side: PI Søren Brunak + Peter Løngreen + Stig Ejdrup Andersen + Jose Maria Gonzalez-Izarzugaza + Nanna Birch Andersen + Ole Nørregaard Jensen

Evaluator side: Henrik Pedersen + Rene Belsø + Diba Markus

Overall assessment at this point in time

The project is on track and momentum is good.

Action points for grant holder, DM Secretariat or others

	Assignee	Deadlines	Who's in the loop
To get proteomics included, it should be investigated if the support for Computerome could be offered to scientist at SDU by including part of Abacus in the Computerome cloud system	Henrik Pedersen	ASAP	SDU, Peter Løngreen

Project progress so far

- WP1: Data capture, data harmonization, conversion of unstructured data into structured biomarker formats**
 Milestone: An updatable workflow for a comprehensive aggregated database of human biomarkers (M24).
 In progress.
- WP2: Data management effort addressing primary data types: genome and proteome sequences**
 Milestone: A cloud compatible, implemented workflow for genomics and proteomics data preparing for biomarker extraction (M12).
 Not begun yet. A challenge is lack of people who understand the particular life science data analysis workflows, plus Abacus support lacks.
- WP3: DTU text mining effort addressing full length papers for novel biomarker detection**
 Milestone: An updatable workflow for controlled vocabularies relevant for biomarker detection in scientific literature (M12).
 In progress – and close to accomplished. A serious challenge is conversion from .pdf to .txt (.pdf structure corrupts original structure).
- WP4: Data management workflows implementing the condensation of genomic and proteomic data into actionable biomarkers**
 Milestone: An updatable workflow for a comprehensive biomarker annotation (M18).
 Milestone: An improved reference for better biomarker identification in the Danish population (M9).
 In progress – and close to accomplished.
- WP5: Secure private cloud effort for biomarker workflows on Computerome (DTU/KU) and ABACUS 2.0**
 Milestone: Virtual integration of sensitive data through cloud bursting (M24).
 In progress for genomics - proteomics is not in the loop from the start, but is expected to join shortly. VM containerization down to network level (=> HW agnostic) and APIs for automating the ordering of resources have already been completed. Establishing tunnel and call back between Computerome and Abacus for storing and updating the biomarkers is in progress. Likewise, with resource pooling.
 Discussion: Establish a "joint cloud" of Abacus and Computerome for the life sciences, i.e. application of seamless access to both (mutually exchangeable seen from researcher perspective). Henrik Pedersen has been in contact with Martin Zachariassen, Claudio Pica and Ole Nørregaard Jensen to promote this idea.

Budget and timeline

The budget is geared through other project funding and the project is on track to reach all milestones in time.