DeiC
DANISH e-INFRASTRUCTURE COOPERATION

Deff
Danmarks Elektroniske Fag- og Forskningsbibliotek

Data management pilot project report

Report date	2017 06 15	Report #	2

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

Video meeting June 15th 2016, 8.30 – 9.00
Participants: Project side: PI Søren Brunak + Jose Maria Gonzalez-Izarzugaza + Nanna Birch Andersen + David Westergaard
Evaluator side: Henrik Pedersen + Josva Kleist + 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
Henrik will get in touch witb Ole Nørregaard Jensen	Henrik Pedersen	ASAP	
Henrik will talk to Per Løngreen regarding Block Chain	Henrik Pedersen	ASAP	
Next meeting, we look forward to get enlighted on the subject of how protomics and genomics together enhances biomarkers for e.g. diabetes	Søren Brunak		

Project progress so far

- **WP1: Data capture, data harmonization, conversion of unstructured data into structured biomarker formats**

  M1: An updatable workflow for a comprehensive biomarker annotation. Status: Completed.

  M2: An improved reference for better biomarker identification in the Danish population. Status: Two papers coming out of the project – one of them is to be published in Nature.

- **WP2: Data management effort addressing primary data types: genome and proteome sequences**

  M1: A cloud compatible, implemented workflow for genomics and proteomics data preparing for biomarker extraction. Status: Implemented. Registry data is utilized. The workflow is in play in Tryggve, which in turn in some respects seems be a role model to Elixir to a certain extend.

- **WP3: DTU text mining effort addressing full length papers for novel biomarker detection**

  M1: An updatable workflow for controlled vocabularies relevant for biomarker detection in scientific literature. Status: Completed. One paper coming out of this.

- **WP4: Data management workflows implementing the condensation of genomic and proteomic data into actionable biomarkers**

  M1: Consolidate the existing pipeline for whole genome sequencing data. Status: Huge progress. The former performance challenges with the protomics worksflows are overcomed with the setup of a large VM on Computerome.

  M2: Improved identification of variants by mapping to the Genome Denmark reference. Status: In progress – and close to accomplished. Also, MedDRA - international medical terminology dictionary (and thesaurus) used by regulatory authorities in the pharmaceutical industry during the regulatory process - is now being translated to Danish (due to the fact that Danish patient records are written in Danish).

- **WP5: Secure private cloud effort for biomarker workflows on Computerome (DTU/KU) and ABACUS 2.0**

  M1: Virtual integration of sensitive data through cloud bursting. Status: See WP4 M1.

Budget and timeline

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