Data, Platforms, and Collaborative Agreements
Agenda - Data, Platform, Agreements

11.00-12.00: data handling
Jesper Grarup, Magnus Fontes, Amanda Mocroft
• Data sources for data warehouse - current and future
• Standardisation of data formats
• Data warehouse structure and platform
• Solutions for data sharing/export – LabKey, FTP, direct peer-to-peer links
• Hardware resource for bioinformatics analyses (access/sharing)

12.00.13.00: bioinformatics analyses
Finn Cilius, Magnus Fontes

13.00-13.30: Structures and formality for establishment of collaborative with external partners
Magnus Fontes, Amanda Mocroft, Lars Fugger, Jens Lundgren
11.00-12.00: Data Handling
Jesper Grarup, Magnus Fontes, Amanda Mocroft

- Data sources for data warehouse - current and future
- Standardisation of data formats
- Data warehouse structure and platform
- Solutions for data sharing/export – LabKey, FTP, direct peer-to-peer links
- Hardware resource for bioinformatics analyses (access/sharing)
Current Data Capture

Data currently captured via real time import from
• Hospital databases: LABKA, MADS, PatoBank and manual updates from EPM
• National data via MedCom: national biochemistry and microbiological data
Future Data Capture

Core database
*Collects data from a variety of sources*

- Microbiology lab
- Biochemistry lab
- Medication, FMK/EPM
- Pathology PatoBank
- Biobank inventory
- Clinical data - KISO
- Host and microbial genomics database
- Immunological characterisation database

- National data registries
- Pharma data
- Imaging
- Ad hoc data collection
- Results of data analysis
- Not yet identified source of data
Structure of data warehouse

External databases

- EPM (RH)
- LABKA (RH)
- MADS (RH)
- Biochemistry (External A)
- Virology (External)
- PatoBank (RH)
- DANBIO (National Quality Assurance)

MATCH database

- EPM (RH) Import
- LABKA (RH) Import
- MADS (RH) Import
- Biochemistry (External A) Import
- Virology (External) Import
- PatoBank (RH) Import
- DANBIO (National Quality Assurance) Import

Biochemistry MATCH
Virology MATCH
Clinical MATCH

User interface MATCH
Interface to database for clinicians
Data capture challenges

• Patients in prospective cohort from 2\textsuperscript{nd} of March when PM biobank opens
  – Script underway searching for the PM LABKA package number and comparing with list of already enrolled patients

• Capture of data from retrospective patient
  – Clinics to supply the PM Core function with lists of patients to enrol retrospective data from
  – Process to be discussed with clinics

• Linking up with the Danish national registries (e.g. cancer, death, etc) requires a formal and time consuming process.
  – Plans for how this is done?
Input.....

Output.....
Personalised interventions to reduce infectious complications
Some thoughts on data management

- Secure, central repository for data
- Documented QA program
- Standard operating procedures across all data sources
- Secure transfer of data between partners
- Standardised data formats across different data sources with central documentation
- Centralised record of all data files, latest version of data-files, where data has been sent
- Central repository for data analysis programs and results
From the statistician’s view

• Consistency between updates of data
• Consistency in data formats in different datasets
• Point of contact to help with data queries
• Agreed protocol of responsibility and actions for data cleaning
• Input into data collection, formats, analyses
From database to researcher

Core database
*Distribute data*

Anonymize

Data selection criteria

- Data set
- Data set
- Data set
- Data set
Solutions for data sharing/export (1)
Solutions for data sharing/export (2)

- FTPS
  - dedicated users/projects folders (CHIPXNET)
- API
  - eg. LabKey or services
- Private cloud service
  - eg https://owncloud.org/features/
Resource for DM, bioinformatics and stats

• Capabilities and number of persons at HQ/RH/partner institutions
  – DM
  – Bioinformatics
  – Stats

• How to ensure communication and collaboration?
Project database

- Portfolio of projects currently in progress, and the resources identified to execute them.
- Principle there are three different types of projects that should be regarded as PM projects:
  - a) those financed directly by funding from DNRF incl all projects that uses the PM biobank;
  - b) projects within the PM mission statement that are funded by PM co-funding from other sources;
  - c) other projects than those mentioned under a) and b) but which uses data collected via the PM centre.
- Discussion of possible projects and how they should be classified under this general classification would be welcome – in particular under b) and c).
Hardware & software platforms & recourses

CHIP’s servers and programmes on CIMT network
• Microsoft Windows Server 2012+
• Microsoft SQL Server 2012+, MySQL
• ASP.NET, C#, IIS, PHP, JAVA, Tomcat
• R, SAS, REDCap

• Server requirements to run LabKey?

Hardware resource for bioinformatics analyses
• Availability Rigshospitalet, Lund University, Institute Pasteur
• access/sharing/collaboration
Standardisation of data formats

Data source formats:
• Diagnosis: ICD-10
• Biochemistry: NPU
• Microbiology: Local classification (NPU)
• Medication: ATC (http://www.whocc.no/atc_ddd_index/)
• Pathology: SNOMED
• Health system admin: SKS, SHAK, SOR, CPR-No. (Danish)
  http://www.ssi.dk/Sundhedsdataogit/National%20Sundheds-it/TerminologiOgKlassifikationer/

Other areas needing standardisation?
• Assay definitions
• Genomics, proteomics
• Flow cytometry

• In general: translate to common international standard eg. CDISC?
12.00.13.00: bioinformatics analyses
Finn Cilius, Magnus Fontes
• Finn and Magnus to add slides
13.00-13.30: Structures and formality - collaboration with external partners
Magnus Fontes, Amanda Mocroft, Lars Fugger, Jens Lundgren

- The DNRF contract requires a collaborative agreement to be entered with external partners within 3 months after contract signature (deadline 23 April 2015)
  - Details to be added...
University College London

• Long standing (> 20 years) extremely successful collaboration between CHIP and UCL
• Open, collegiate and respectful from leaders of individual groups and all researchers
• Based on mutual respect, good communication encouraged by all
• Shared responsibilities drawing on strengths of individuals
Current CHIP-UCL collaborations

Copenhagen (Clinical)  
Jens Lundgren

> 20 past or present PhD/MDs

UCL (Statistical)  
Amanda Mocroft

Clinical trials  
INSIGHT network, including START / FLU)  
Local trials (MaxCMin, PASS)

Observational studies  
MATCH  
HIDES  
EuroSIDA  
DAD  
PARTNER  
Many ad-hoc studies

Funding  
European commission  
Pharma  
NIH  
NIHR  
Local funding  
Ad-hoc funding

UCL (Hans Stauss)  
Immunity and transplantation

Danmarks Grundforskningsfond  
Danish National Research Foundation
Joint publications between UCL and CHIP (October 2014)

Published items in each year

Citations in each year

Results found: 384

Sum of the Times Cited [:] : 17028
Sum of Times Cited without self-citations [:] : 15919
Average Citations per Item [:] : 44.34
h-index [:] : 64

*search for Lundgren and (Mocroft or Phillips or Sabin or Cozzi Lepri or Fox)
Going forward.....

• Formality of collaboration
• Project by project basis
• Agreed tasks and deliverables by all partners
• How to manage funding for external partners
• Recognition for external institutions
• Maintain excellence across all disciplines