Phenotype database: To store your study design and integrate data

Lars Verschuren/ Jildau Bouwman

European Nutritional Phenotype Assessment and Data Sharing Initiative
Intervention studies (www.dbnp.org)
Schematical view of database

- Study Meta-data
  - Raw data
  - Clean data
  - Statistical results
  - Processed results (e.g. pathways)

- Clinical chemistry data (can be stored in the phenotype database)
- Transcriptomics (locally or in ArrayExpress)
- Metabolomics (locally or in Metabolights)

- Multi-study analysis and results

- Study capture (based on ISA-TAB)
  - Data processing workflows
  - Search, visualize
New study

Define the basic properties of your study

Enter all the basic information of your study. Keep in mind that the more specific the information that is filled out, the more valuable the system will be.
Templates

**Diet intervention (switch)**

Currently, this template contains the following fields. Drag fields to reorder. Drag fields to the list of available fields to remove the field from the template.

- **name (Short text)**
- **Event name (STRING) (Short text)**
- **Intervention/Challenge (Dropdown selection of terms)**
- **Event-type (Dropdown selection of terms)**
- **Route (Dropdown selection of terms)**
- **Description (Long text)**
- **Diet description (File)**
- **Diet carbohydrate-level (Energy%, Decimal number)**
- **Diet fat-level (Energy%, Decimal number)**
- **Diet protein-level (Energy%, Decimal number)**
- **Migration (Short text)**

**Available fields**

These fields are available for adding to the template. Drag a field to the template to add it.

- **2nd Event-type (Dropdown selection of terms)**
- **Compound (DB:id name) (Dropdown selection of terms)**
- **Compound dose (Decimal number)**
- **Compound dose unit (Dropdown selection of terms)**
- **Compound frequency (Short text)**
- **Compound full name (Dropdown selection of terms)**
- **Vehicle (Dropdown selection of terms)**

Create new field
Any ontology from bioportal can be used

Creation of new vocabularies
Define or import your study design

The study design consists of treatment types and sample types, grouped together in sample & treatment groups. Sample & treatment groups can be assigned to groups of subjects.
Output possibilities

- For study comparison computer access to software that can analyse your data, e.g. R (via API)
- Search options through database
- Excel export
- Visualisation (Pathvisio, simple graphs)
Licensing

Data
• Studies.dbnp.org: can be used on request
• Server: at TNO behind firewalls, daily back-up to other servers, access only by administrators (TNO personnel)
• Data: data access only for the data owner and data manager

Code
• Code: open (apache license)
• Safety: authentication/authorization system of grails framework sustained by VMware
Take home message

• Phenotype database is a system that can capture study design (meta-data)

• Via API link with databases e.g. ArrayExpress and PRIDE, etc.
Objective

- To deliver:
  - Open access research infrastructure
  - Containing data from a wide variety of nutritional studies
  - Including mechanistic/interventions and epidemiological studies
  - With a multitude of phenotypic outcomes
  - To facilitate combined analyses in the future
ENPADASI Project

- 9 countries
- 51 partners (in 15 national consortia)
- Initial national selection of partners
- Consortium formation in Rome May 2014
- Started December 2014
- 2-year project
- Nationally funded
Why?

- Connect similar studies to resolve chronic diseases with lifestyle related solutions
- Combining studies will increase the power thereby limiting the needs for new and larger intervention studies
- Validation of study results in a different cohort/study will enhance the biological applicability of the conclusions
- Data comparison is instrumental to improve the interpretation and validation of results
- Increase the knowledge and understanding of how food and nutrition can improve human health
Data sharing in Nutrition (DASH-IN)

- Intervention studies (design only)
- Intervention studies (incl clean data)
- Observational studies (incl clean data)
- Observational studies (design only)

- Is there a study with similar design? \(\rightarrow\) improve power
- Is there a study with similar outcome? \(\rightarrow\) validate outcome
- Is there a study with similar design? \(\rightarrow\) Extend measures
- Are there studies with specific interventions and outcome
Goal

Federated database

Intervention study

Phenotype database

Observational study

FAIR

DASH-IN

Test case
Ultimate goal

Biomaterial and data collections

All resources to measure
e.g. genomics, transcriptomics, proteomics, metabolomics, bioimaging, microscopy, quantified self, lifestyle, nutritional studies

Design of Multi-disciplinary Experiments

All resources for data stewardship and analytics:
e.g. bioinformatics, informatics, biostatistics, computational (systems) biology, e-science, ICT, ...

Medical

Consumer

Improve public health

Model systems

Information & Insight

e-health & quantified self data

International reference data
Context & sustainability

WP1: Sustainability

WP2: Collect studies

WP3: Federated database
- Interventional
- Observational

WP4: - Ontologies
- Integration

WP5: Legal and ethics

WP6: Training

elinxir

BBMRI-ERIC

DEDIPAC

ECRIN

The Food Biomarker Alliance
ENPADASI in the context of other initiatives

• Data from the medical field are important for ENPADASI and DISH-RI
• Therefore, general data standards in life science are needed (ELIXIR and Corbel)
• DEDIPAC delivers important data for determinants of intake
• Foodball requires the integrated infrastructure for their biomarker development
Acknowledgements

- All ENPADASI partners
- WP2 – Study data (Lars Dragsted)
- WP3 – Database (Corrado Priami)
- WP4 – Integration (Duccio Cavalieri)
- WP5 – Regulation (Martine Laville)
- WP6 – Training (Lorraine Brennan)

See also: test.dbnp.org (user=user password=useR123! )
MEDICAL DATABASE OF USEFUL INFORMATION, 5k.
MEDICAL DATABASE OF USELESS INFORMATION, 500,000,000 MB.