

Data from the Biobanking Perspective

concepts - issues - challenges

*Joerg Geiger
Head of Fluids Biobank @ idbw*

RDA 2017 - IG healthdata

Interdisciplinary Bank of Biomaterial und Data Würzburg

ibdw



agenda

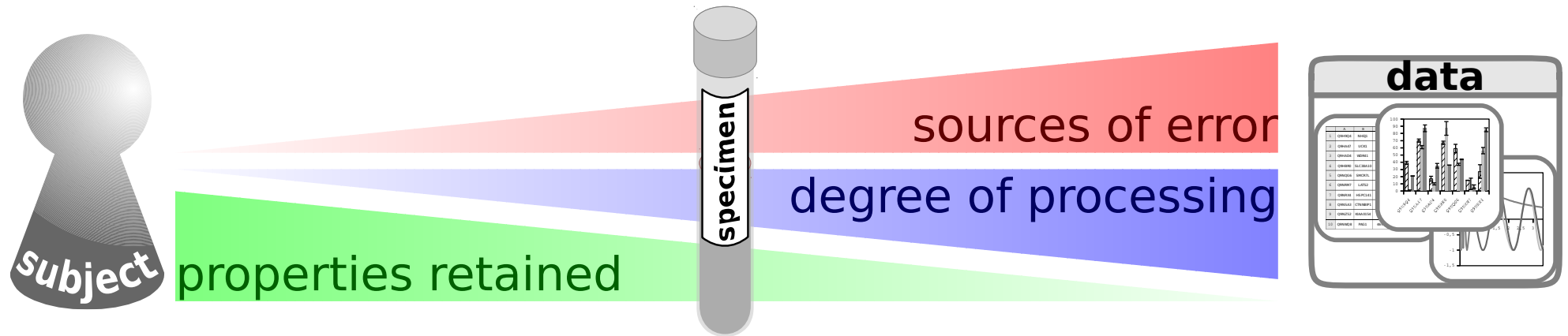
- the biobank concept
- the sample concept
- data sources
- quality issues
- ELSIssues
- interoperability
- standardization initiatives

biobank definition

- Hewitt, R., and Watson, P. (2013). Defining biobank. Biopreservation Biobanking 11, 309–315.:
„A biobank is a facility for the collection, preservation, storage and supply of biological samples and associated data, which follows standardized operating procedures and provides material for scientific and clinical use.“

samples and data

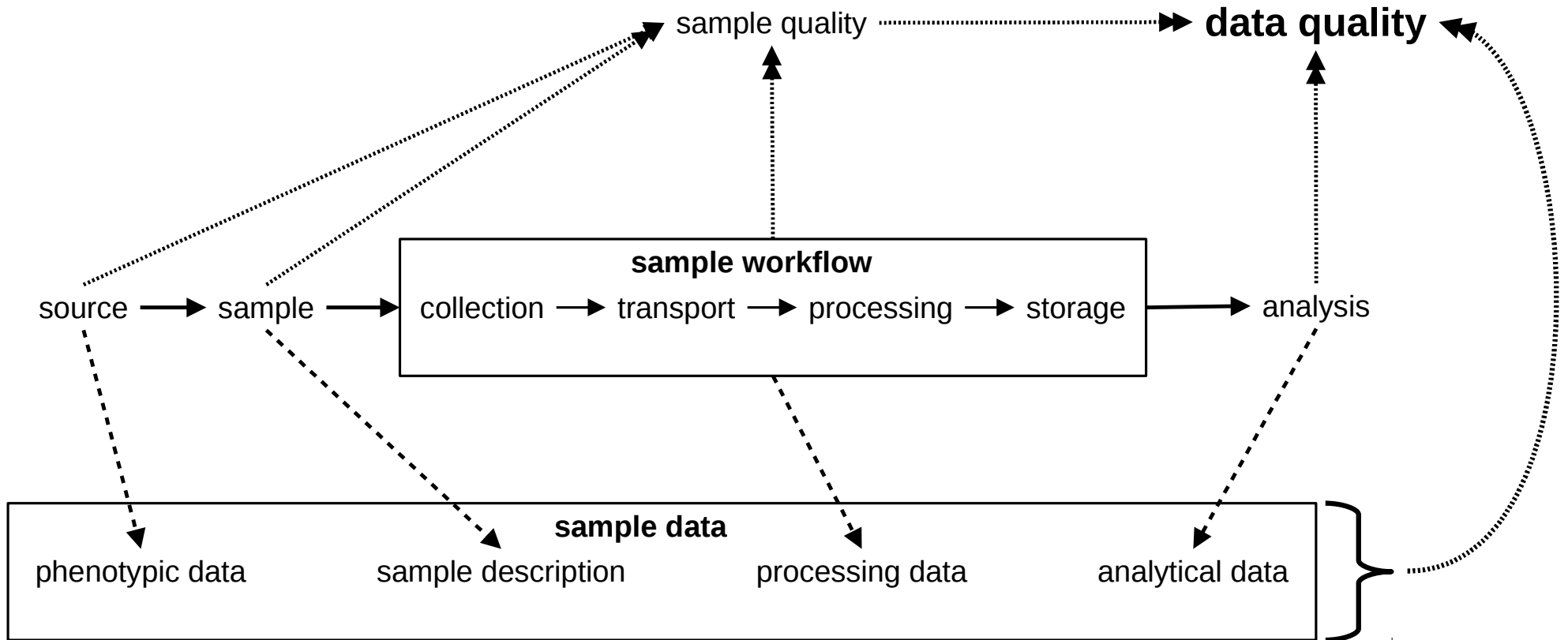
- samples are the source of data
 - sample quality \sim data quality
 - \Rightarrow data quality must be assessed on basis of sample quality
- sample data reflect sample quality
- sample data add further dimensions to analytical data



sample data

- sample description data
(type of sample, material characteristics)
- sample processing data
(sample history, preanalytics)
- source phenotypic data
(biometric, health)
- analytical data
(data generated from the sample)

data quality



data quality issues

- sample description and processing data
 - *documentation guidelines (specifying items, no quality indicators!)*
- phenotypic data and analytical data
 - *different purposes (accounting, therapy)*
 - *different sources (non/curated)*
 - *comparability (longitudinal data collection)*
- in any case:
 - *provenance (no tracking, w/o quality controls, qualification)*
 - *traceability (validity)*

ethical and privacy issues

- right to access personal information
- non-disclosed information
- user anonymity (data granularity and dimensionality)
- re-identification risk
- identity matching
- patient tracking

■ interoperability

- *data format*
- *data integration*
- *data semantics*
- *comparability*
- *validity*

■ ethics and privacy

- *effective anonymization*
- *tracking/managing consent and restrictions*
- *donor feedback*

■ ISO/TC 276 Biotechnology

- *Microbial resources data— specification on data management and publication in microbial resource centres*
- *Minimal requirements for downstream data processing and integration workflows for interfacing and linking heterogeneous data, models and corresponding metadata*
- *Data sharing and publication*
- *Methods to evaluate the quality of the massive sequencing data*
- *Cell metadata specification*
- *Genome compression*
- *Documentation requirements in biobanking*

Summary

- highly diverse data
- many unresolved problems
- several quality issues
- ethical and legal questions
- standardization efforts



ORCID