Data sharing and accessibility of data (data mining) open access data warehouse

Ana Sofia Figueiredo

Medical Faculty Mannheim

University of Heidelberg

Germany

Brussels, 8th of December 2016

UMM

UNIVERSITÄTSMEDIZIN MANNHEIM

Medizinische Fakultät Mannheim der Universität Heidelberg

Universitätsklinikum Mannheim

More than 11 million animals are used each year for research purposes in the European Union

(data from 2011)

http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm

How can we reduce animal testing in research?

- Actively implement the "3Rs" principle of replacement, reduction and refinement
- Improve study design
- Use alternative models of disease
  - Computer Models and simulations
    - **Systems Biology** and **Bioinformatics** approaches
  - Cell lines / Blood from donors / Animal waste
  - Artificial Animal Models
- ***Share your research data***

Share your research data to minimize laboratory animal use

Challenges and opportunities of data sharing

Data accessibility

Open access data warehousing (DW)

Share your research data in five steps

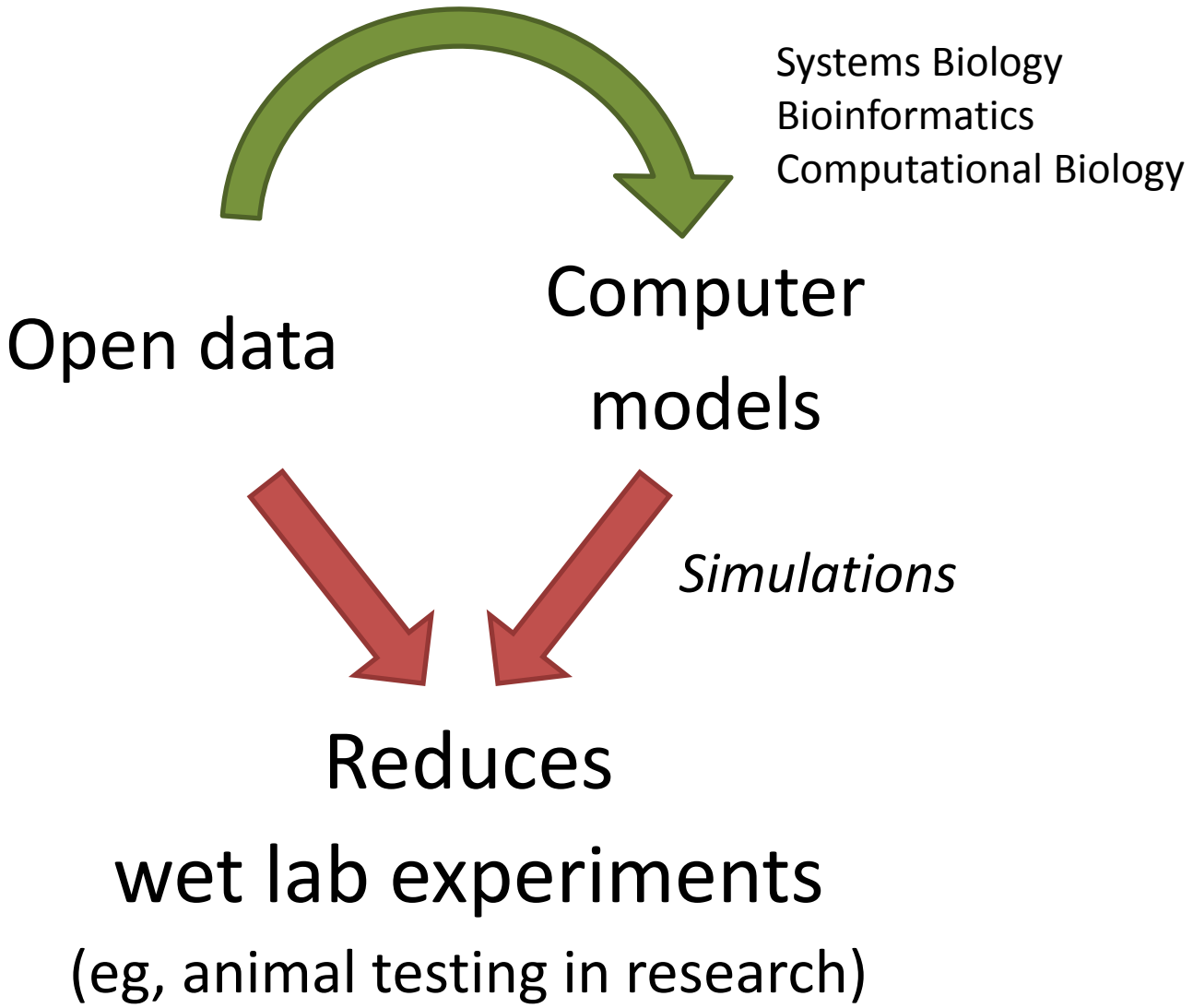
Challenges and opportunities of data sharing

- Avoids duplication of experiments
- Parallel re-evaluation of the same data set
- Reduces animal experiments
- Returns the investment back to society

- Competitors with full access to the data can publish faster
- Data misinterpretation by re-users
- The investment of sharing the data is high and has no immediate return

More knowledge, fewer experiments

Open Data and Computer Models synergize to reduce animal testing



Share or protect Data?

- Clinical trials
- Patient data
- Results from questionnaires

Data protection in the EU is currently governed by the Regulation (Eu) 2016/6 250 79 of the European Parliament and of the Council of 27 April 2016

Protect *personal data.*

Personal data:

"(…) any data relating to a natural person who can be identified either directly or indirectly by it (…)"

source: General Data Protection Regulation (GDPR)

http://www.appliedclinicaltrialsonline.com/data-protection-regime-change-eu-impact-pharmaceutical-industry
http://eurlex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Share de-personalized clinical data

Taichman DB, et al, „Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors". JAMA. 2016

Data Accessibility

Improve access to research data:

- (Meta)data have a unique identifier
- A standardized protocol handles these identifiers
- This protocol is open, free, and universal
- It allows for an authentication procedure
- Metadata are accessible, even when the data are no longer available

Wilkinson et al, „*The FAIR Guiding Principles for scientific data management and stewardship*“ Scientific Data 3, 2016 http://www.nature.com/articles/sdata201618

The importance of being FAIR

Findable – Accessible – Interoperable – Reusable

- Guides data *producers* and *controllers*
- Maximizes the added value of data
- Ensures

transparency, **reproducibility**, **ability to search** and **reusability**

Live update:
http://datafairport.org/fair-principles-living-document-menu

Wilkinson et al, „*The FAIR Guiding Principles for scientific data management and stewardship*“
Scientific Data 3, 2016 http://www.nature.com/articles/sdata201618

Open access data warehousing

Data Warehouses are **central repositories of integrated data.** They store and curate the deposited data and create analytical reports for (re-)users.

Repositories for research data

A data repository is:

- Persistent, Robust
- Curated
- Recognized by your community

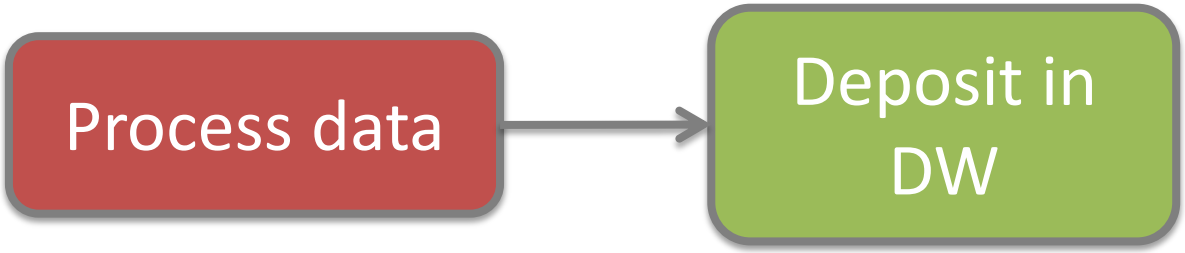
re3data.org registry

A data repository provides:

- Good technical assistance
- Restricted access to the data
- Track system of your (meta)data usage and citations

http://research.omicsgroup.org/index.php/Data_warehouse

Bottleneck



- Sharing data implies well planned experiments, well structured and identified data
- Experimentalists are not trained to curate their data
- Data sharing means extra work with no immediate return
- Solution: hire or train a data scientist

Share your research data in five single steps

Step 1: Improve experimental design

- Check if your idea is feasible in terms of time and resources
- Design your experiment
- Create a data management plan
  - Data generation
  - Data processing and quality control
  - Data preservation
  - Sharing plans

A well designed experiment will be easier to understand and re-use than a not so well-designed one.

Step 2: Think Metadata and Use standards

Metadata

- Metadata makes data searchable and findable
- Structured representation of the data
- Identifies the data
- Locates a resource

Metadata *is not* data

Its quality *determines* data reusability

Use Standards

- Science is more complex and interdisciplinary
  - Team work
  - *lingua franca*
- Standard file formats are preferential than proprietary ones
- Ideally, researchers use existing standards adopted by the data repositories.
- When standards do not exist, use generalized data serialization formats such as YAML or JSON

http://biosharing.org

can help you select the most appropriate standart for your metadata

Step 3: License your data

With whom can you share your data?
Only with Scientists or with Society?
Non profit organizations?

Controlled Access - Licensing
Open Data Commons (ODC)
Creative Commons (CC)

permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Step 4: Deposit / Publish your data

Repositories

- Dryad
- Figshare
- GitHUB (code version control)

Peer reviewed periodicals

- Scientific data, NPG
- Data in Brief
- Gigascience Journal
- Codata – Data Science Journal

Step 5: Budget your costs!!

- Data sharing is costly in terms of time, money and resources
- Funders and regulators must be aware of the challenges behind data sharing
- Be aware of the costs of:
  - producing FAIR data
  - storing / publishing your data
  - gaining the expertise to perform these tasks

Outlook

- Research data support scientific discovery
- Data sharing can speed up scientific progress and knowledge discovery
- ***Data sharing is one step-stone to avoid the unnecessary duplication of animal studies***

…and it paves the way for a more open, ethical and sustainable science