Zukunft Reproducible Science
1. Introduction
Reproducibility crisis is partly caused by:

- Misinterpretation,
- Misunderstanding,
- Incompleteness, and/or
- Unreadability.

Part of the Reproducibility Problem

■ “An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship.”

■ “The actual scholarship is the complete ... set of instructions which generated the figures.”

■■ David Donoho, 1998
3 Data Science Applications of Bio Research

- 3 Topics Today
- Molecular Biology
- Systems Biology
- Collective Behaviour
2. Molecular Biology
Membrane Generation and Simulation

in CELLmicrocosmos 2.2 MembraneEditor (CmME)

- solving the Lipid Packing Problem
- movement along 2-dimensional area
- shape-based collision detection operates in 3 dimensions
- all packing settings are stored in a protocol

2. Molecular Biology
Defining a Membrane Composition in CmME

B. Sommer et al. JCM 5(51):1165–82, 2011.


Improved methods to isolate and subfractionate rat liver mitochondria.

Rud Hovius, Hilde Lammers, * Klaas Nolting and Ben de Kruijf

Cardiolipin

Phosphatidylethanolamine

Phosphatidylserine

Phosphatidylglycerol

Cardiolipin

Phosphatidylethanolamine

Phosphatidylserine

Phosphatidylglycerol

nmol phospholipid/mg protein

1110 ± 150 (4)

OM

48 ± 3 (7)

10 ± 0.3 (3)

9.9 ± 1.6 (6)

2 (4)

Hovius et al. 1990
Applied Databases by CmME
- PDB for proteins
- OPM for protein placement
- PDB_TM for transmembrane protein placement
- lipid import: various lipid databases
Membranes by Seed

- example from CmME
- using the same random seed with the same algorithm and configuration, the same membrane will be generated.
- using different seeds results in various models of the same configuration
Membrane Simulation

- heterogeneous membrane based on CmME
- simulated with Gromacs
Multiple Membrane Simulations

- vary random seeds creates multiple MD simulations with one membrane configuration
- multiple runs can be used to validate the model
- example:
  - a single 100x100 Ångstrom$^2$ membrane
  - 1,000 ns simulation
  - 64 cores on the SCC Konstanz cluster
  - requires ca. 2 weeks
  - creates 260 GB of data to be analyzed
- then: analysis and comparison to wet lab evaluations
3. Systems Biology
Examples from Systems Biology

- A network with 10^3 nodes
- A network with 10^3 nodes
- Metabolic network, source: KEGG, 2012
- A network with 10^4 nodes
- Protein interaction network, source: DIP, 2013
- A network with a custom layout
- Cyclins and Cell Cycle Regulation, source: Abcam 2009

3. Systems Biology

Björn Sommer, University of Konstanz
Most English speaking country

Quebec

Iran

China

Israel

Singapore

Norway

Poland

USA and Canada

- Standardized symbols as inspiration

3. Systems Biology
Three languages in one:

- Systems Biology Graphical Notation
- Le Novère et al. Nature Biotech., 2009

### Process Description maps
- Temporal courses of biochemical interactions

### Entity Relationships maps
- Relationships in which a given entity participates, regardless of time

### Activity Flow maps
- Information flow between biochemical entities in network

3. Systems Biology
Three languages in one:
- Systems Biology
- Graphical Notation
- Le Novère et al. Nature Biotech., 2009

Process Description maps
- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships maps
- Unambiguous
- Mechanistic
- Non-Sequential

Activity Flow maps
- Ambiguous
- Conceptual
- Sequential
Three languages in one:

- Unambiguously describe biochemical and cellular events in graphs
- Limited amount of symbols (~30) → smooth learning curve
- Can graphically represent quantitative models, biochemical pathways, at different granularity levels
- Developed since 2006 by a growing community, part of COMBINE

3. Systems Biology
3. Systems Biology
Repositories affiliated to COMBINE standards

- https://pathwaycommons.org
- https://biomodels.net
- https://synbiohub.org
- https://models.physiomeproject.org

3. Systems Biology
5. Collective Behaviour
MoveBank in 2017
- Over 600 million animal locations
- Over 1.1 billion measurements from non-location sensors

- 4,200 studies
- 756 taxa
- 5,000 contributors
- 800 active data feeds

5. Collective Behaviour
Research Questions Bird Behaviour

- How do birds move with the wind?
- Who leads a flock?
- Where and how do birds die?
- Can we predict disasters based on animal behavior?

Next: Accumulated Bird Data over several years projected to one year
5. Collective Behaviour
Life of a Storch

Name
Karamell / DER AL582 (eobs3031)

Species
Ciconia ciconia

Description
White Stork female, born 2013 in Radolfzell-Bingingen (Southern Germany). Siblings Zuzu (AL581), Swag (AL 583) and Eamy (AL585). Karamell was the third chick to hatch.
Data is collected via GPS sensors

- Differ in timely resolution
- 2 times 24h, or
- every second
- Usually only Longitude/Latitude are recorded – Altitude is missing
- here: storks
ICARUS

- International Cooperation for Animal Research Using Space
- Driven by Martin Wikelski et al.
- ICARUS <5-g tags including GPS, acceleration and other sensors, ground-based base stations and software supporting two-way communication
- Data will be distributed to owners via automated feeds in Movebank

5. Collective Behaviour

Björn Sommer, University of Konstanz
Data is collected via GPS sensors
- here: storchs

- next:
  - storch & thermals

- After next:
  - Fly with the storchs in VR!
5. Collective Behaviour
Immersive Analytics of Collective Bird Behaviour

Stand at BLS 6 2018 in Konstanz

5. Collective Behaviour
6. Acknowledgements
6. Acknowledgements

Thanks for your attention!!!

---

**CELLmicrocosmos project**
- Funding
- DFG GK635
- BMBF, DAAD, NNSF China
- And our many talented students over the years!
- [http://team.CELLmicrocosmos.org](http://team.CELLmicrocosmos.org)

---

**MPI for Ornithology**
- Martin Wikelski
- Kamran Safi
- Mate Nagy
- Michael Quetting
- Andrea Flack
- Wolfgang Fiedler

**Computational Life Sciences Konstanz**
- Falk Schreiber
- Karsten Klein
- Hieu Nim (Monash Uni.)
- Our bird expert students:
  - Alexej Gluschkow
  - Artur Baltabayev
  - Dimitar Garkov
  - Kim Rehberg
  - Stefan Feyer
  - Thorsten Breitkreuz
  - Ying Wang

---

**Universität Konstanz**

---

**Björn Sommer, University of Konstanz**
4th International Symposium on
Big Data Visual & Immersive Analytics 2018

OCTOBER 17-19, KONSTANZ, GERMANY

General Chair
Falk Schreiber

Program Chairs
Eduard Gröller, Kim Marriott

Organising Chairs
Alexandra Diehl, Björn Sommer

University of Konstanz

Follow @BDVASymposium