The Role of *Elephants* in Complex Workflows in Electrophysiology

Analysis of Spikes and Population Signals

HBP Code Jam 2016
Shrigley Hall, Manchester, UK
with Andrew Davison, Detlef Holstein, Vahid Rostami, Alper Yegenoglu

Jan. 13, 2016 Michael Denker

The illusion of experiments as a neuronal Rosetta Stone

Variability

Complexity

Consequences of disillusionment for reproducible research in electrophysiology

JÜLICH FORSCHUNGSZENTRUM

Reproducibility…

…is undervalued

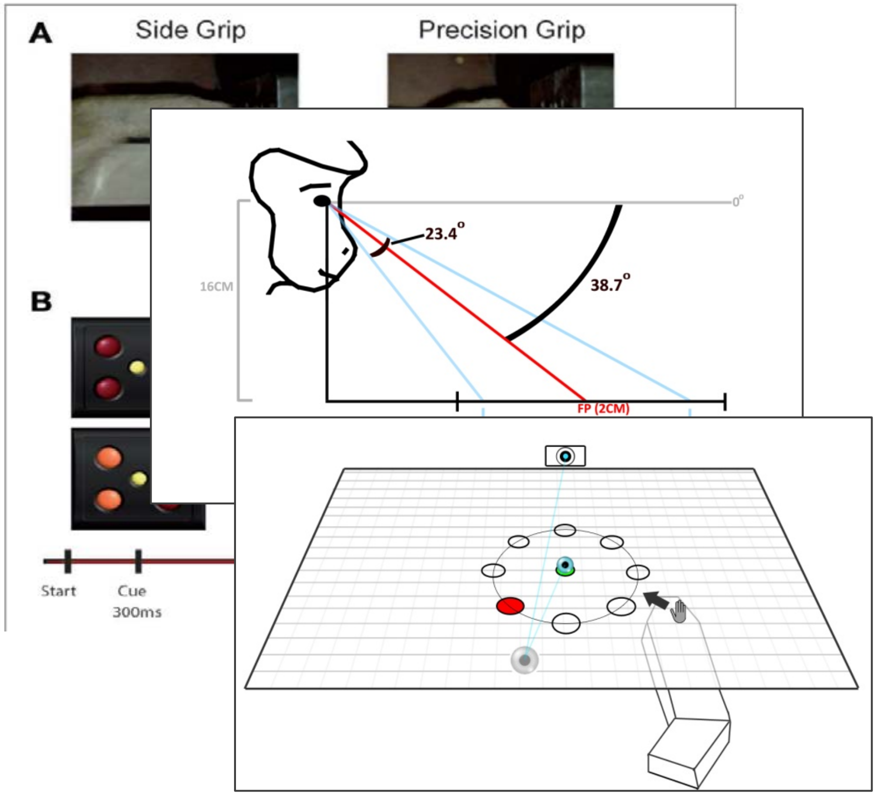
(variability)

…is a difficult task

(complexity)

Why?

Complex metadata | behavioral experiment



- Complex, natural behavior (e.g. Reach-to-grasp task)
- Involves long training
- Complicated cue presentations
- Registration of events (e.g. reaction time)
- Control of behavior
- Measurement of behavior
- Many parallel recording channels

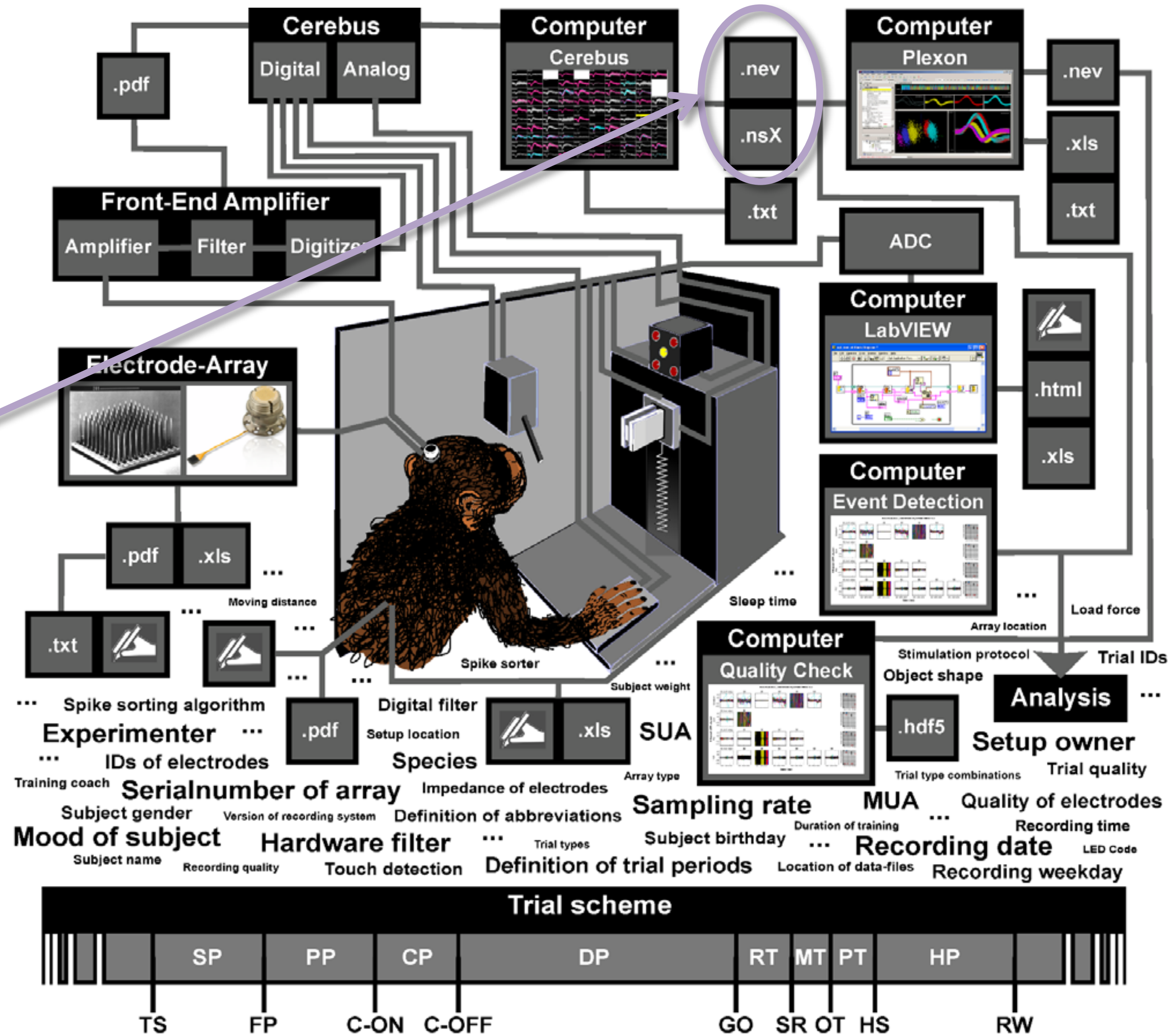
Riehle et al (2013) Front Neural Circuits

Complex metadata | behavioral experiment

Reach-to-grasp study:

- 120 trials / recording
- ~ 5 recordings / day
- ~ 70 days / monkey
- 3 monkeys

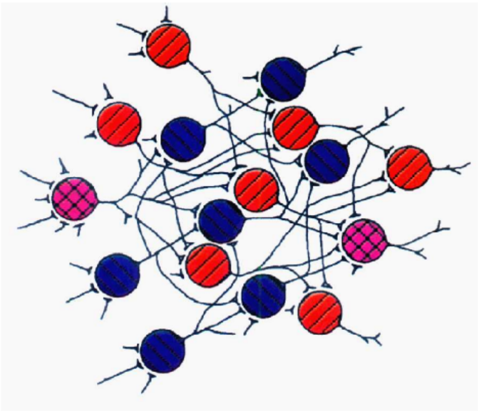
Actual neural data in only two files!

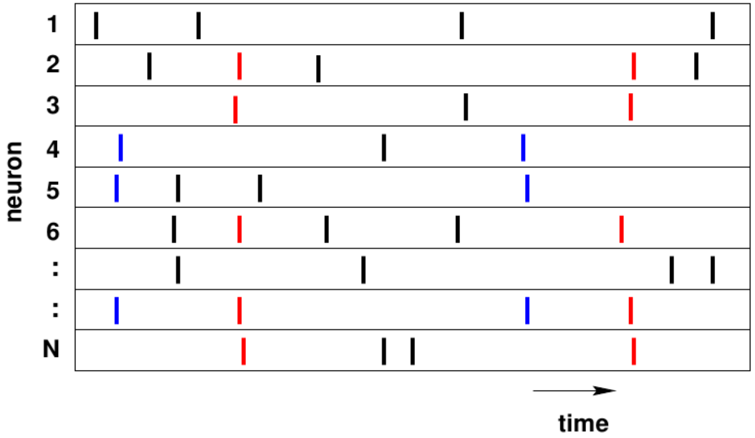


Zehl, Jaillet, Stoewer, Grewe, Sobolev Wachtler, Brochier, Riehle, **Denker, Grün** (submitted)

Complex analysis | finding cell assemblies

- Cell **assembly** hypothesis: Representation of information (percepts, actions, timing…) by the *transient*, *precise co-activation* of a specific neuronal assembly
  e.g., Kilavik, Ponce-Alvarez, Confais, **Grün**, Riehle (2009) *J Neuroscience*

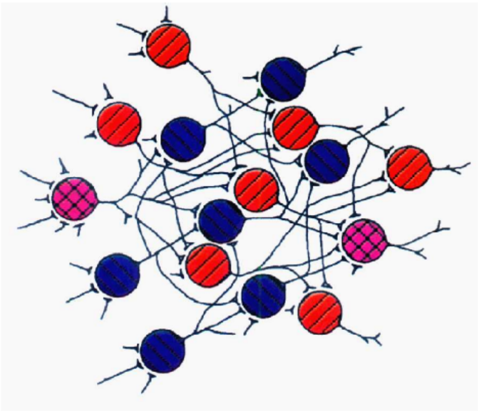


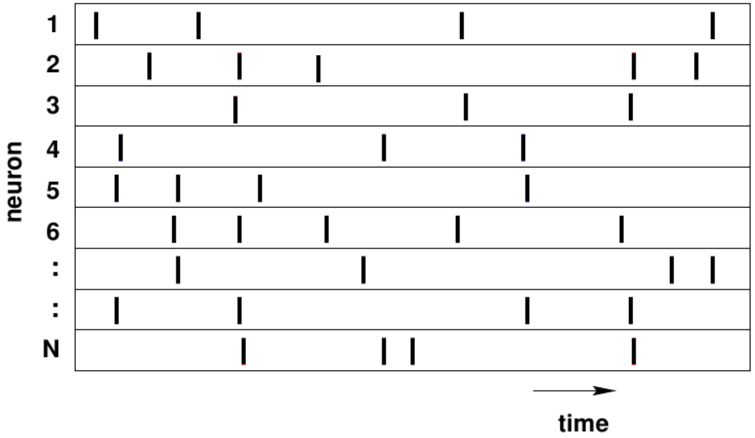


courtesy: M. Abeles

Complex analysis | finding cell assemblies

- Cell **assembly** hypothesis: Representation of information (percepts, actions, timing…) by the *transient*, *precise co-activation* of a specific neuronal assembly
  e.g., Kilavik, Ponce-Alvarez, Confais, **Grün**, Riehle (2009) *J Neuroscience*
- **Challenge:** Detection of assembly activations by pair-wise and higher-order analysis of spiking activity

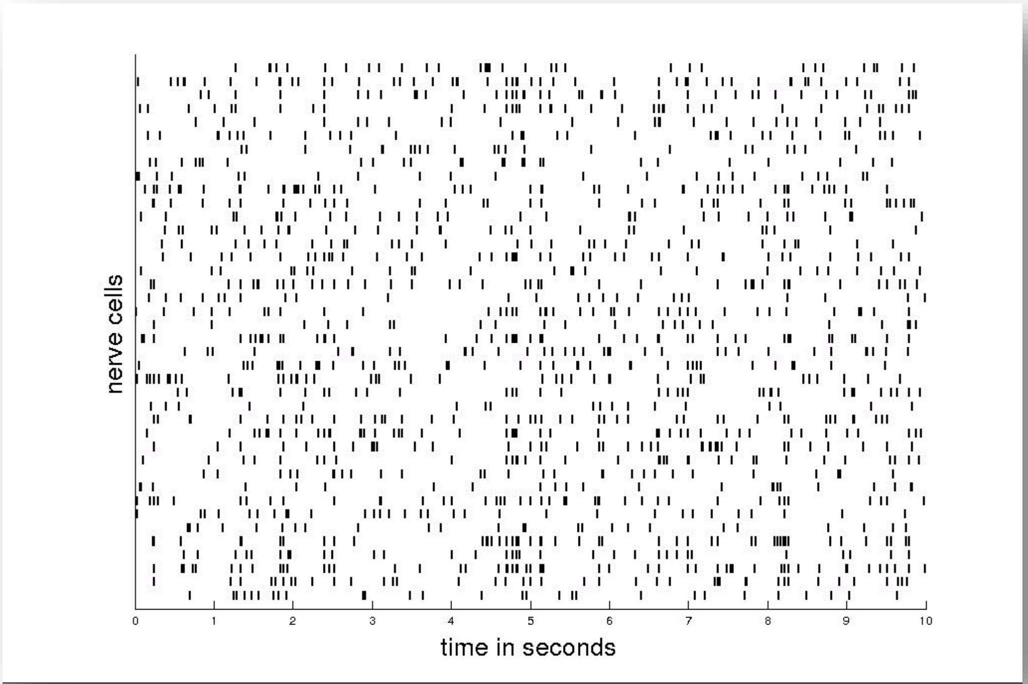




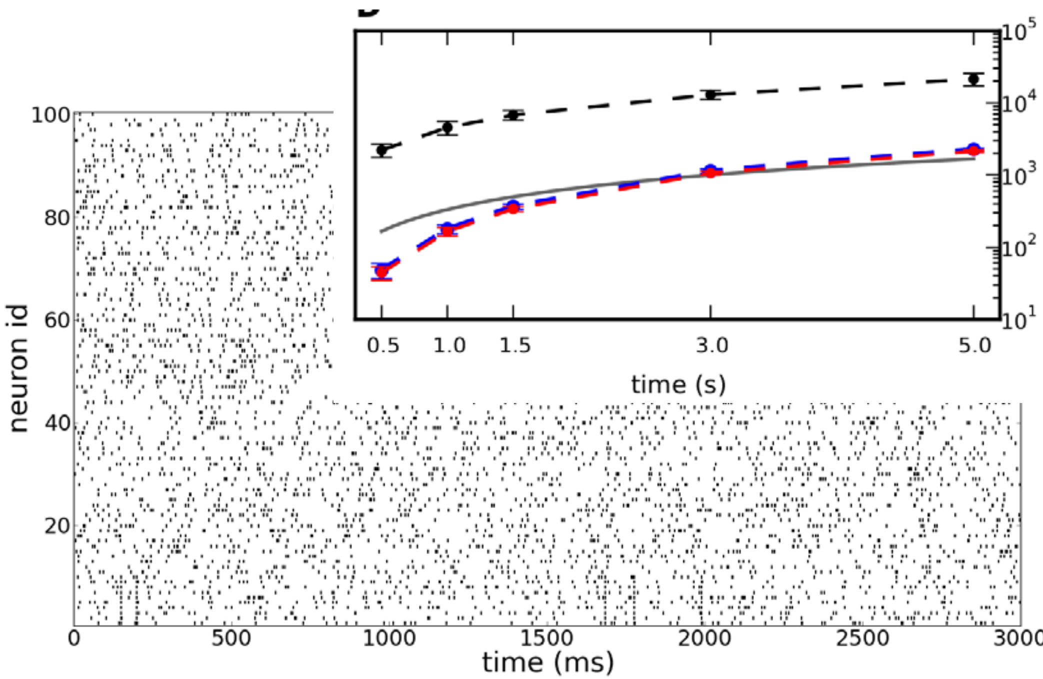
courtesy: M. Abeles

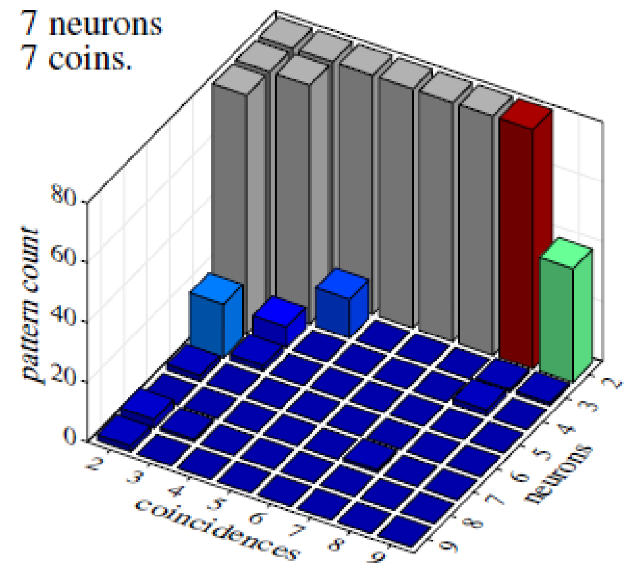
Complex analysis | finding cell assemblies

- Cell **assembly** hypothesis: Representation of information (percepts, actions, timing…) by the *transient*, *precise co-activation* of a specific neuronal assembly
  e.g., Kilavik, Ponce-Alvarez, Confais, **Grün**, Riehle (2009) *J Neuroscience*
- **Challenge:** Detection of assembly activations by pair-wise and higher-order analysis of spiking activity



Complex **analysis** | detection of spike synchrony patterns

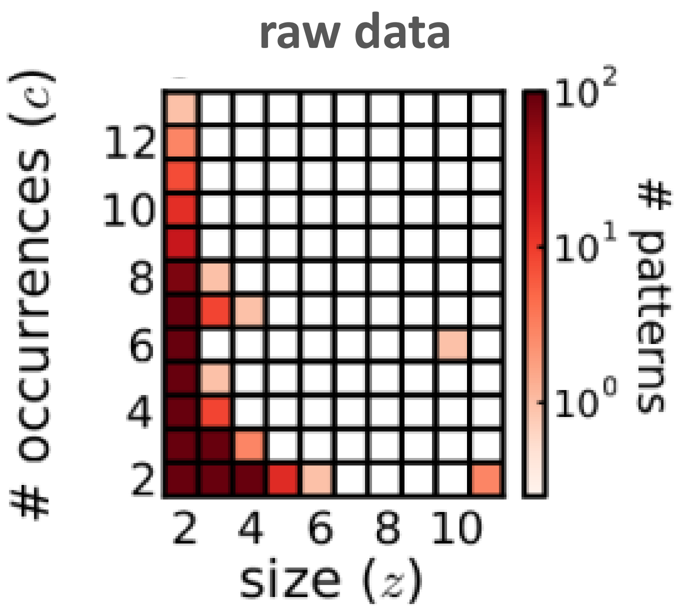




- Employ frequent itemset mining to count patterns efficiently
- Avoid massive multiple testing by pooling patterns of identical size and occurrence count in pattern spectrum

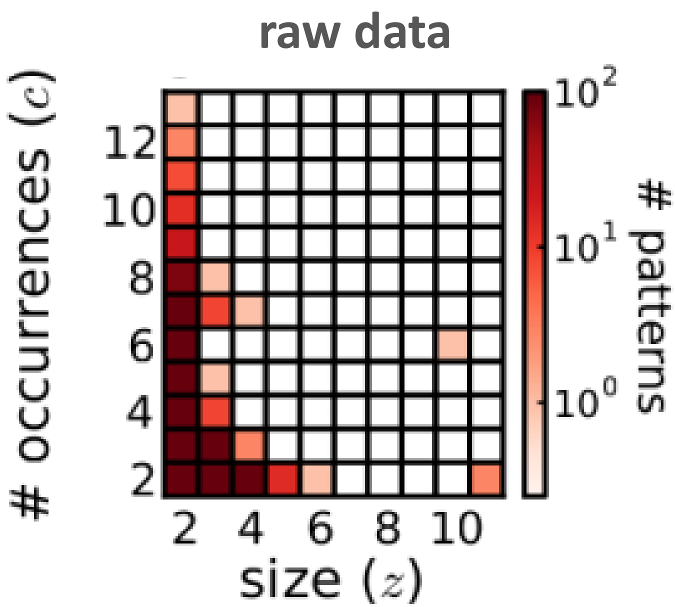
Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci
Picado-Muiño, Borgelt, **Berger**, Gerstein, **Grün** (2013) Front Neuroinform

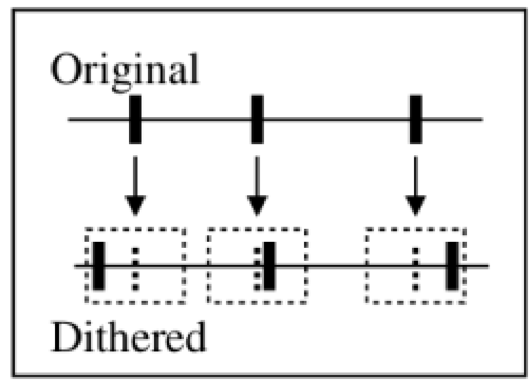
Complex analysis | estimation of significance of synchronous spike patterns



Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci

Complex analysis | estimation of significance of synchronous spike patterns

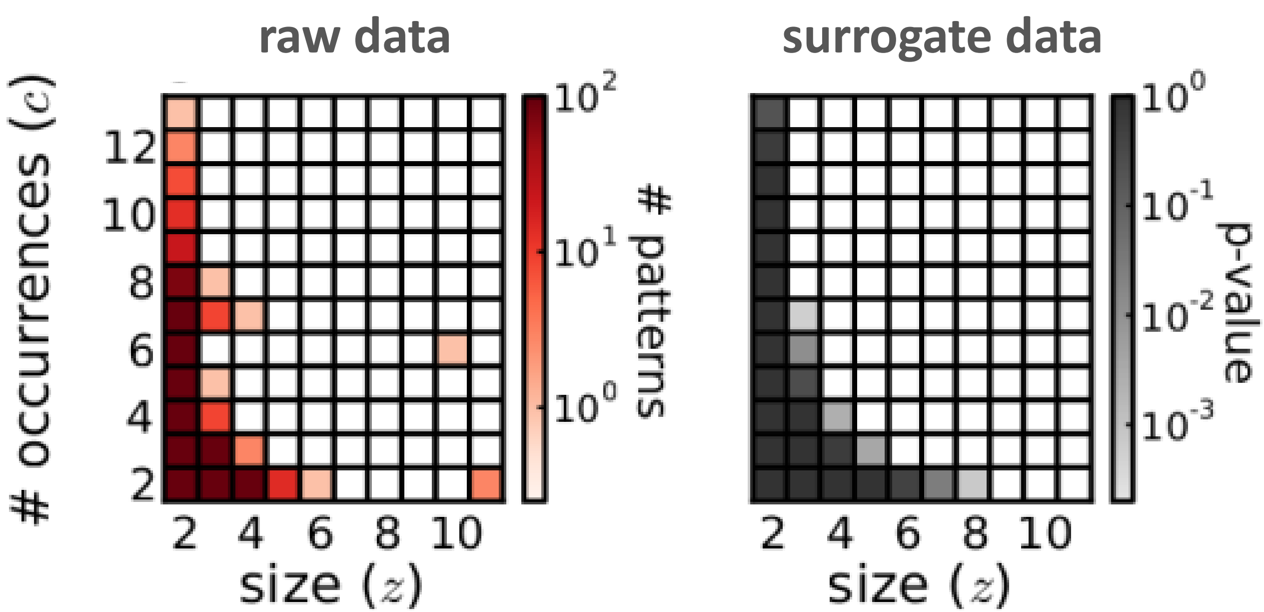




- Extraction of significant spike synchrony patterns by comparison to (independent) surrogate data

Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci

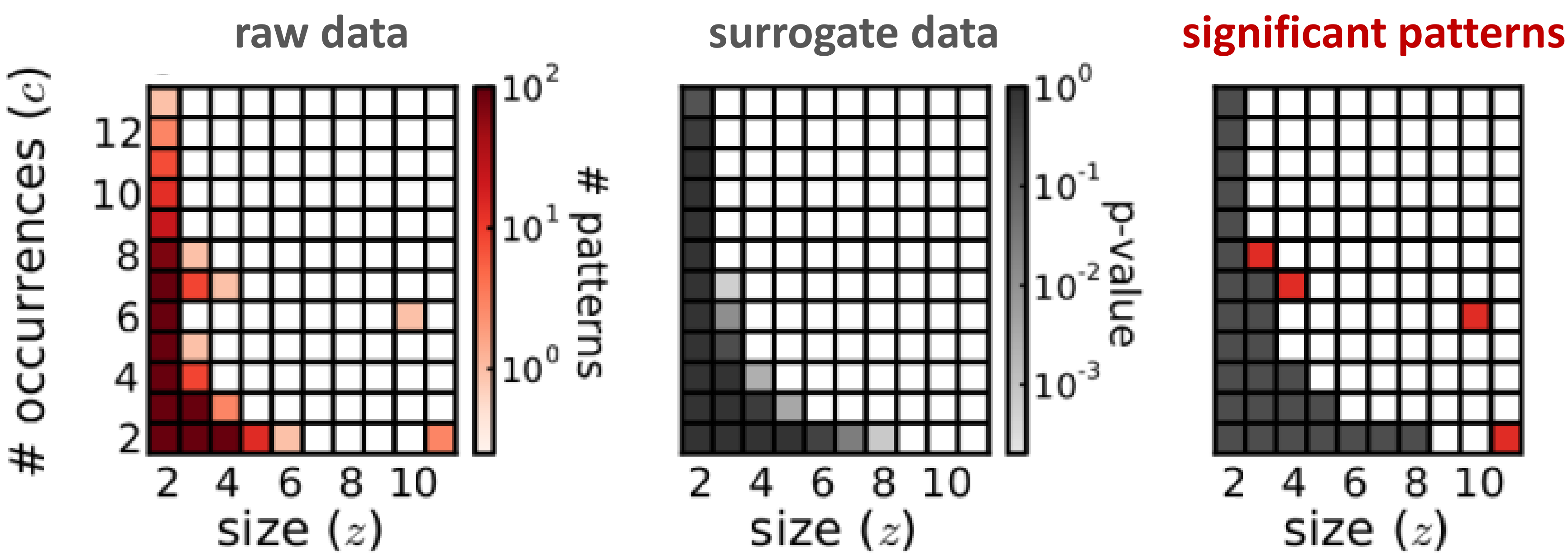
Complex analysis | estimation of significance of synchronous spike patterns



- Extraction of significant spike synchrony patterns by comparison to (independent) surrogate data

Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci

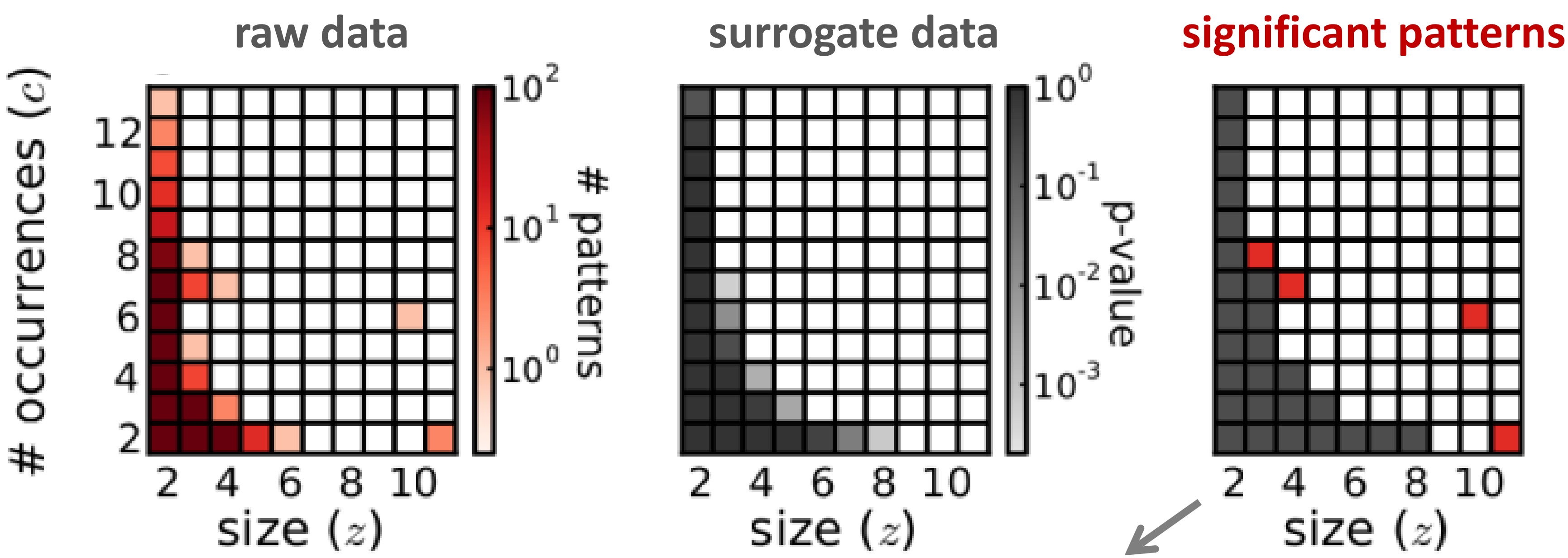
Complex analysis | estimation of significance of synchronous spike patterns



- Extraction of significant spike synchrony patterns by comparison to (independent) surrogate data

Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci

Complex analysis | estimation of significance of synchronous spike patterns



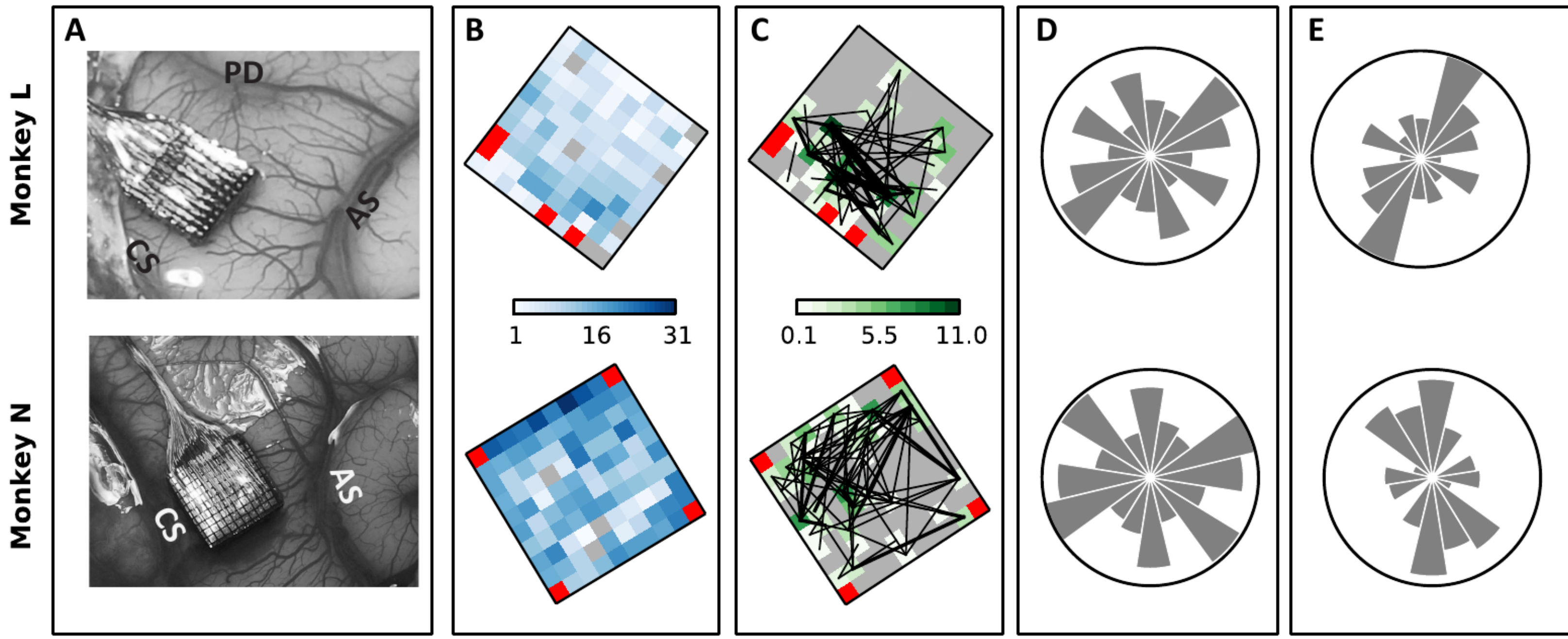
{1,2,3,4,5,6,7,8,9,10,80}: 2
{1,2,3,4,5,6,7,8,9,10,24}: 2
{4,6,9,10}: 7

{1,2,3,4,5,6,7,8,9,10,28}: 2
{1,2,3,4,5,6,7,8,9,10}: 6
{4,6,10}: 8

→ *Conditional tests on pre-screened patterns*

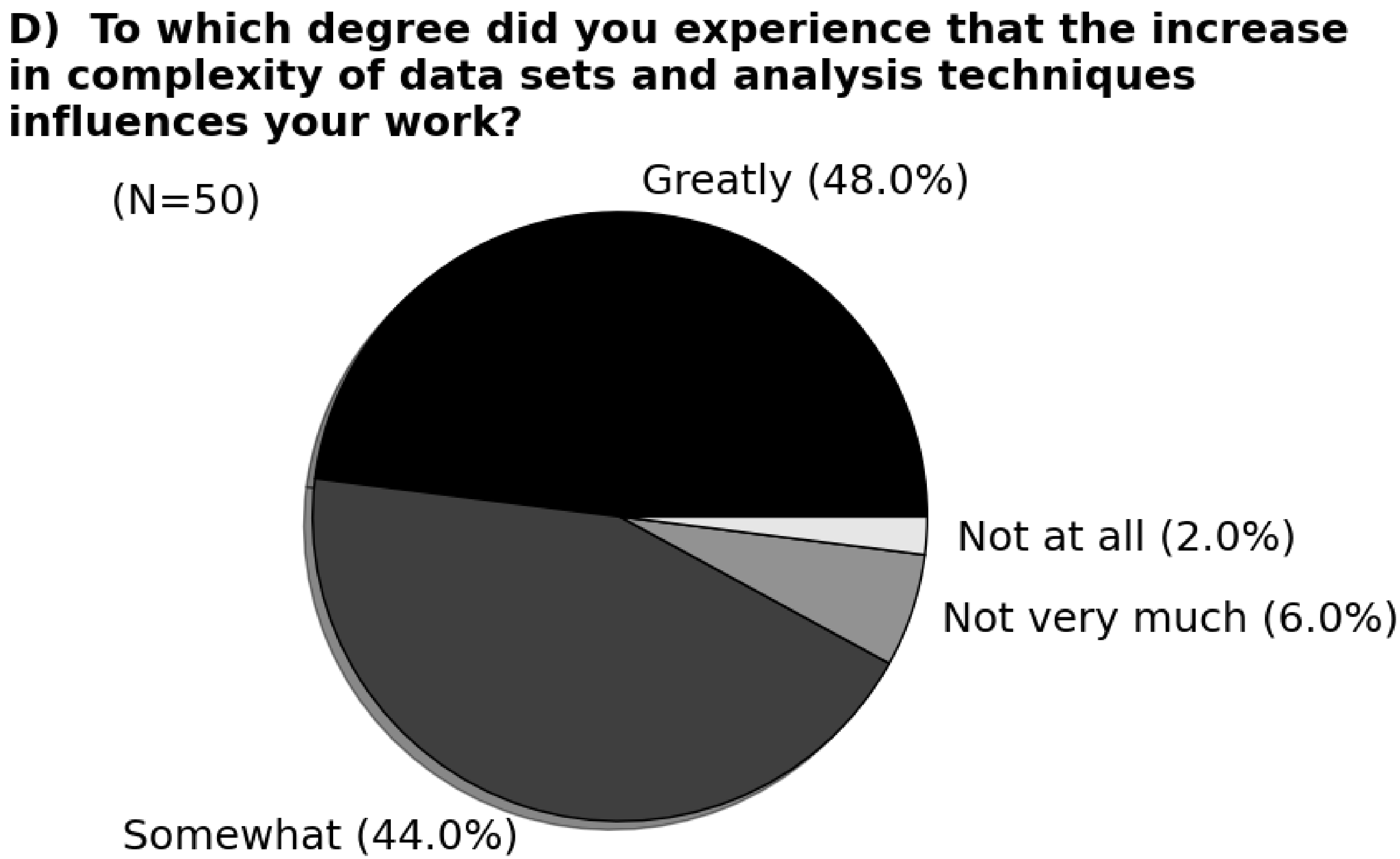
Torre, Picado-Muiño, **Denker**, Borgelt, **Grün** (2013) Front Comput Neurosci

Complex analysis | finding patterns in actual experimental data

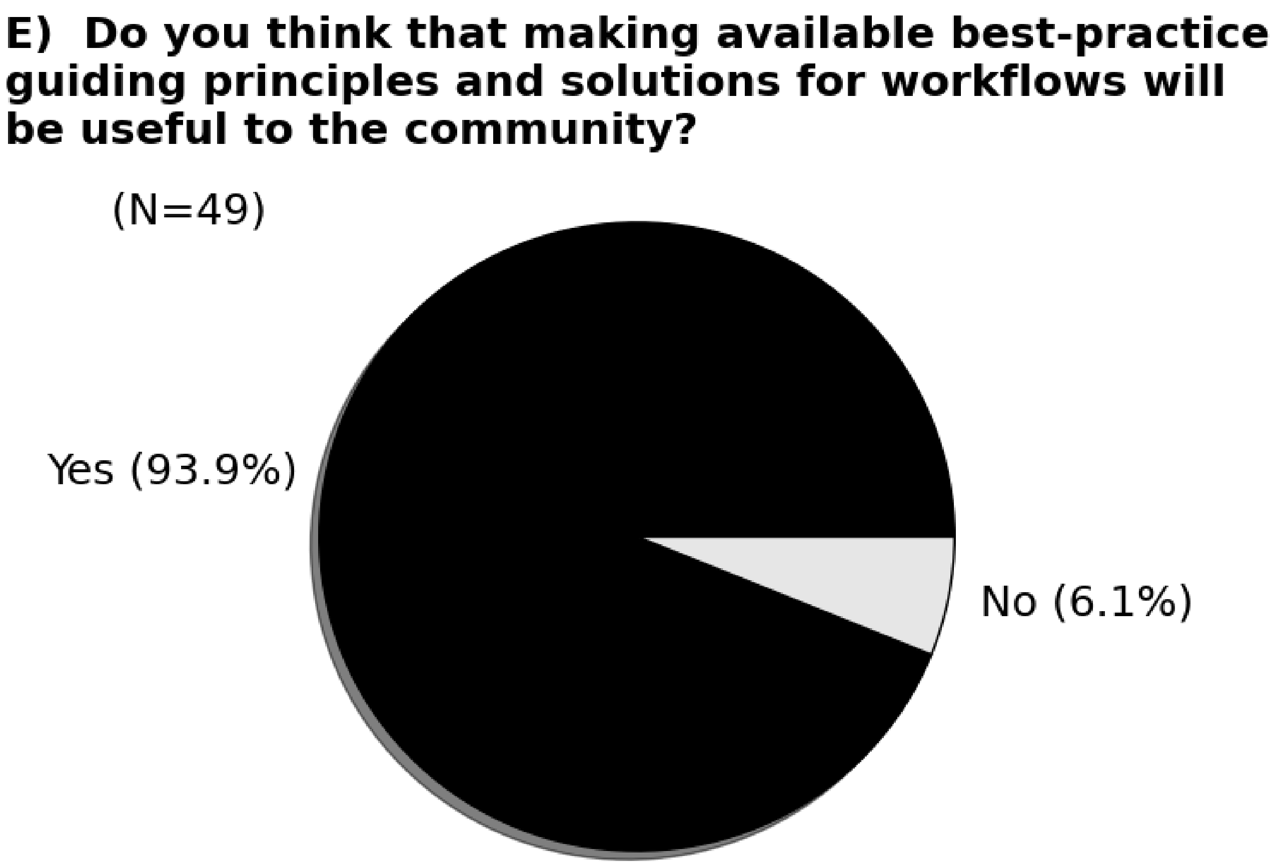


Torre et al. (in review)

Survey | community in need for better tools

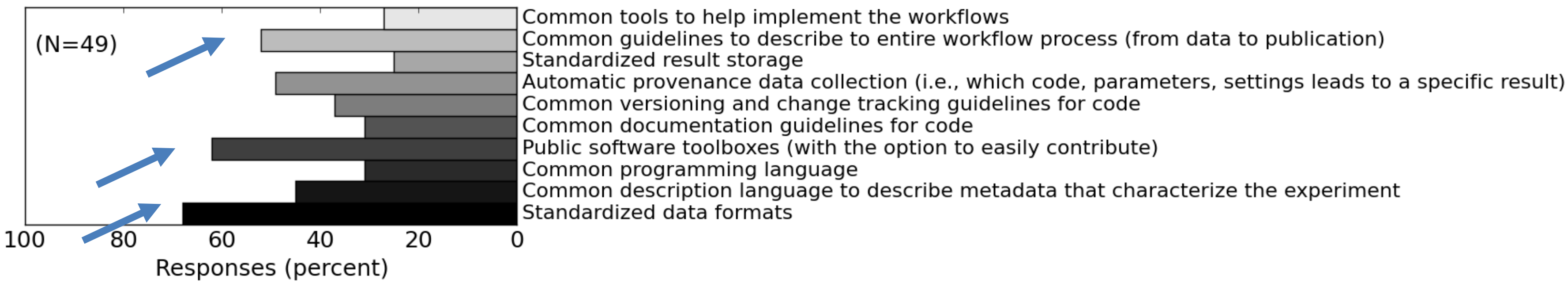


Survey | community in need for better tools



Survey | community in need for better tools

C) Which of the following are most important to implement in the community to ensure that results can be reliably reproduced, verified and extended by other researchers? (Check all that you consider very important)



Outcomes | INCF Workshop "New perspectives on Workflows and Data Management for the Analysis of Electrophysiological Data"

JÜLICH FORSCHUNGSZENTRUM

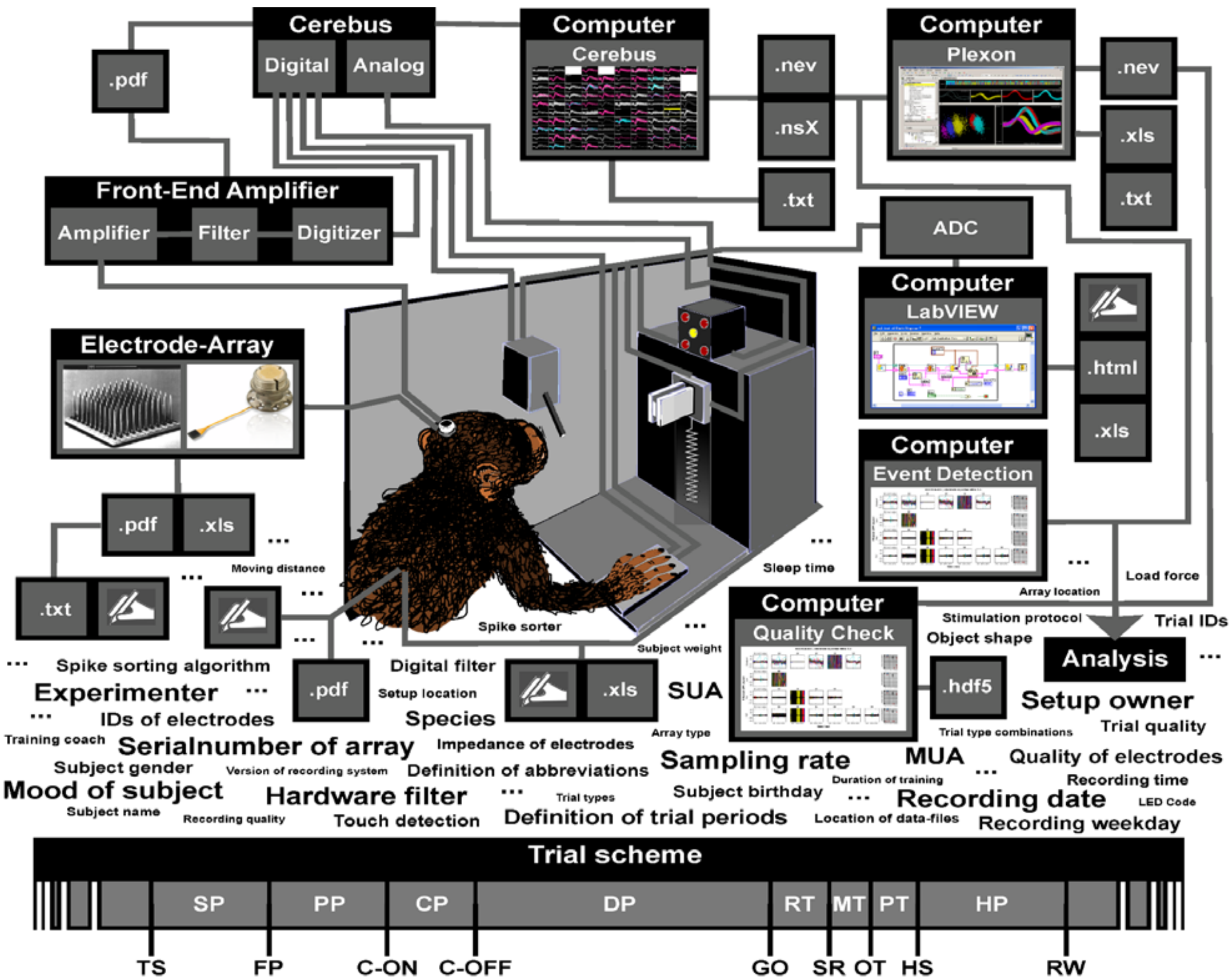
Recommendations to the INCF

- **Establish best practice guidelines for developers working on provenance and interoperability.**
- **Develop standards for metadata.** INCF's work on metadata should focus on capturing metadata, integrating it with data, and easy entry/management, not on determining which metadata are important since this will vary with user needs. This work should be coordinated with efforts of existing INCF task forces working on metadata, provenance, and workflows.
- **Create a website for community exchange.** The participants believed the current webpage with resources for data sharing in electrophysiology should be expanded to a website for community exchange that contains practical information about experiments, the type of metadata required for certain experiments, tool tutorials, and a forum feedback on experimental design.
- **Engage with vendors to coordinate standardization.** INCF should encourage vendors to adopt common formats for data/metadata, include annotation tools that include standard terms, have automatic metadata capture at all steps, and automated save function. Also, the participants thought that INCF should also encourage vendors to develop easy, user-friendly tools.
- **Support training activities on versioning, software, and relevant concepts.** This recommendation was motivated by the fact that many of the tools used today target people with some programming experience, not people who are unable to write code. INCF should hold courses to teach users basic coding, as well as offer courses to help developers develop easy to learn, easy to use, and well-documented tools. In addition, the INCF should encourage organizers of data analysis courses to base their teaching on open standards, formats and tools to promote their use to the young generation of scientists

http://incf.org/activities/workshops/scientific-workshops

Workflows | assembling software tools to sustain a reproducible data analysis

Integrated Solutions



metadata

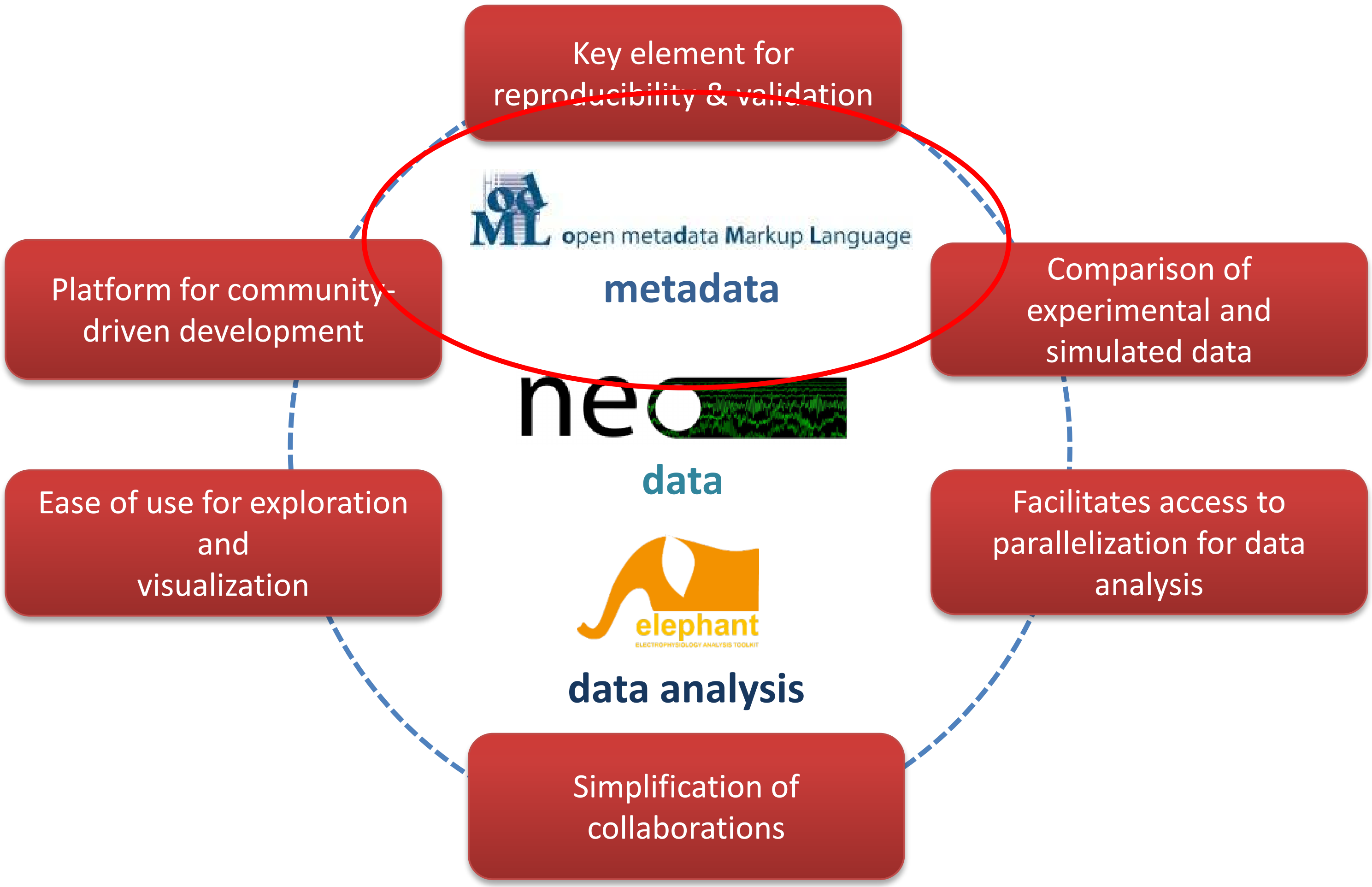
(Zehl et al., submitted)

data representation

data analysis

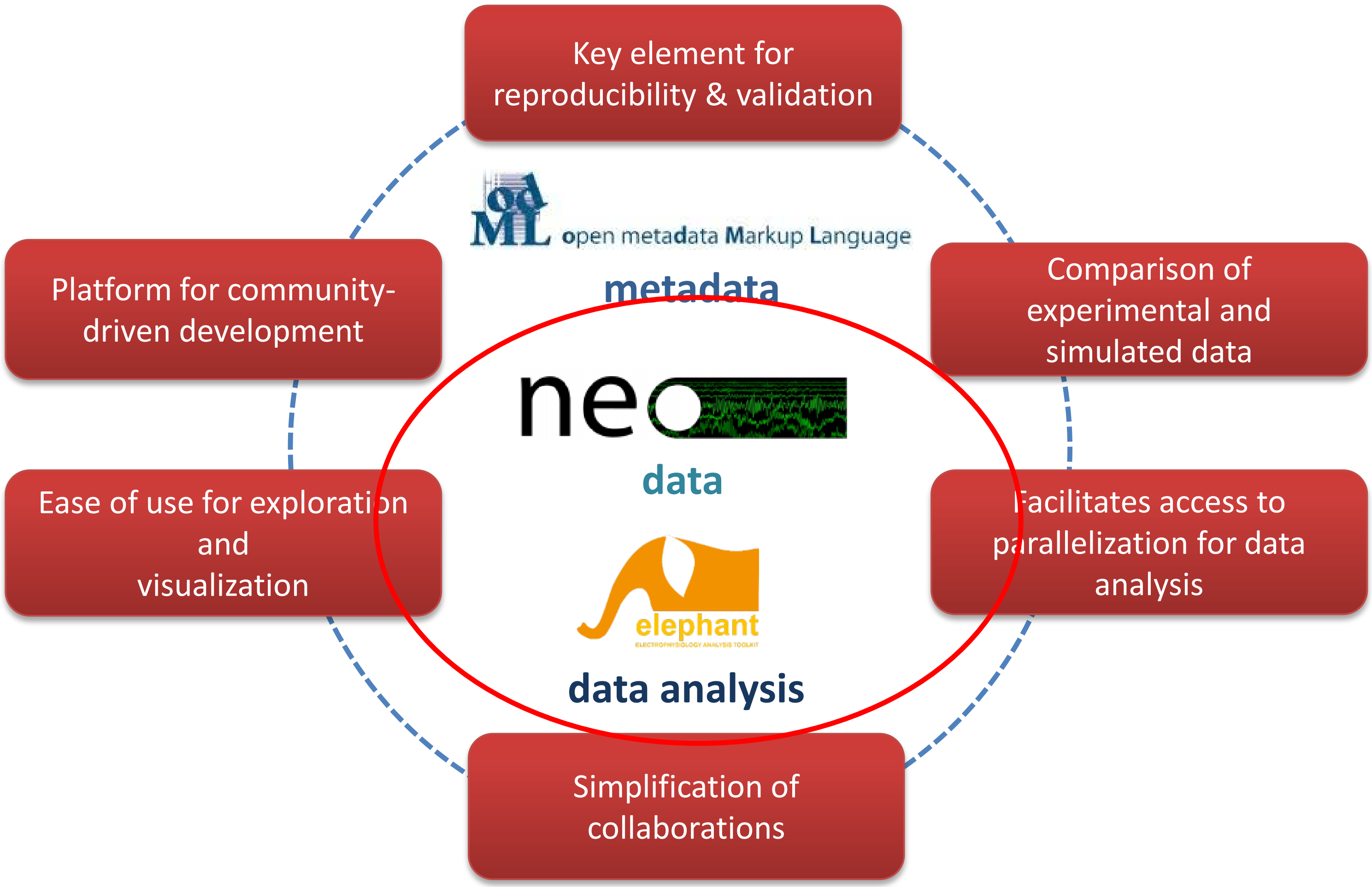
(Denker et al., SfN 2015)

Reproducible workflows | open software tools



Reproducible workflows | open software tools

JÜLICH FORSCHUNGSZENTRUM



Overview | reproducible data analysis using *Elephant*

FORSCHUNGSZENTRUM JÜLICH

Aims:

- provide generic tools to analyse
  - **brain dynamics** from **experiments** and **simulations**
  - **large neuron populations (**massively parallel spike trains, local field potentials)
  - relationship of such **multi-scale** data
- create **toolbox** for hosting **a broad range of methods**
- based on the data models provided by the **Neo library**
- **modular design** of analysis functions

elephant
ELECTROPHYSIOLOGY ANALYSIS TOOLKIT

github.com/NeuralEnsemble/elephant
elephant.readthedocs.org/en/latest

Overview | reproducible data analysis using *Elephant*

JÜLICH FORSCHUNGSZENTRUM

Development history and strategy:

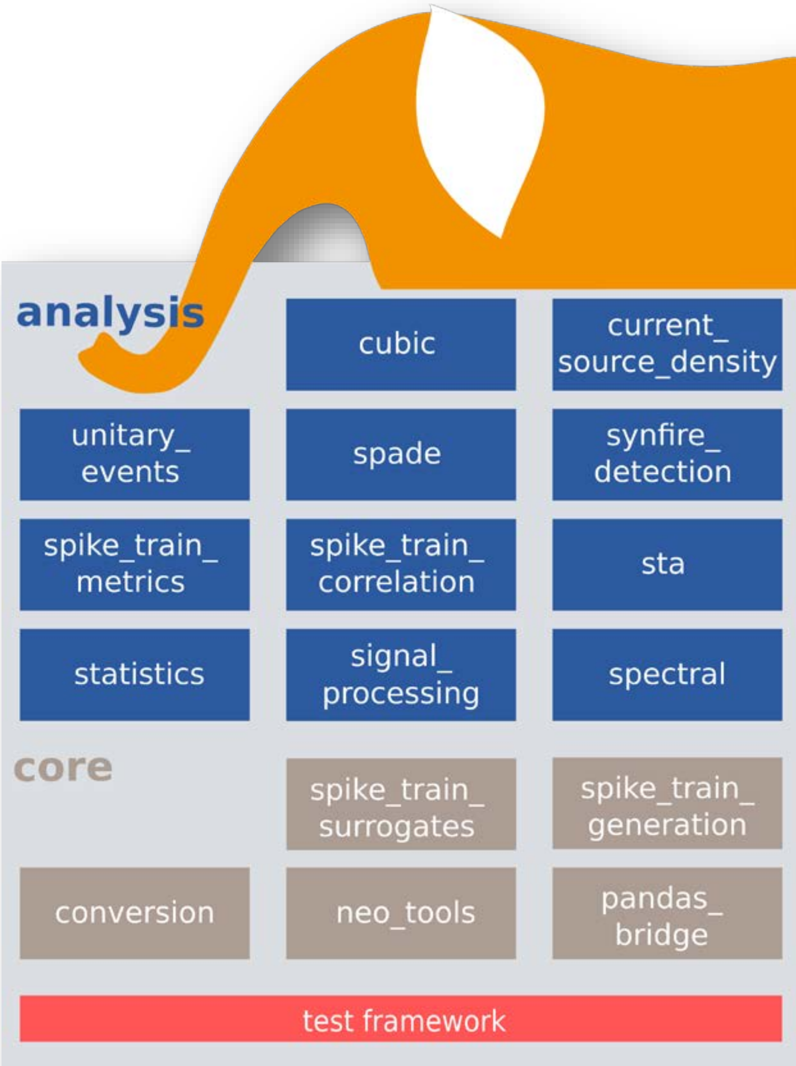
- **community-centered**, open-source, curated
- successor of NeuroTools developed in EU projects **Facets and BrainScaleS**
- Recognized as one of four **key innovations** of the EU BrainScales project
- Model of success: **co-design** (software+application developed in parallel by scientists and software engineers)

elephant
ELECTROPHYSIOLOGY ANALYSIS TOOLKIT

`github.com/NeuralEnsemble/elephant`

`elephant.readthedocs.org/en/latest`

Scope | structure of *upcoming* contributions

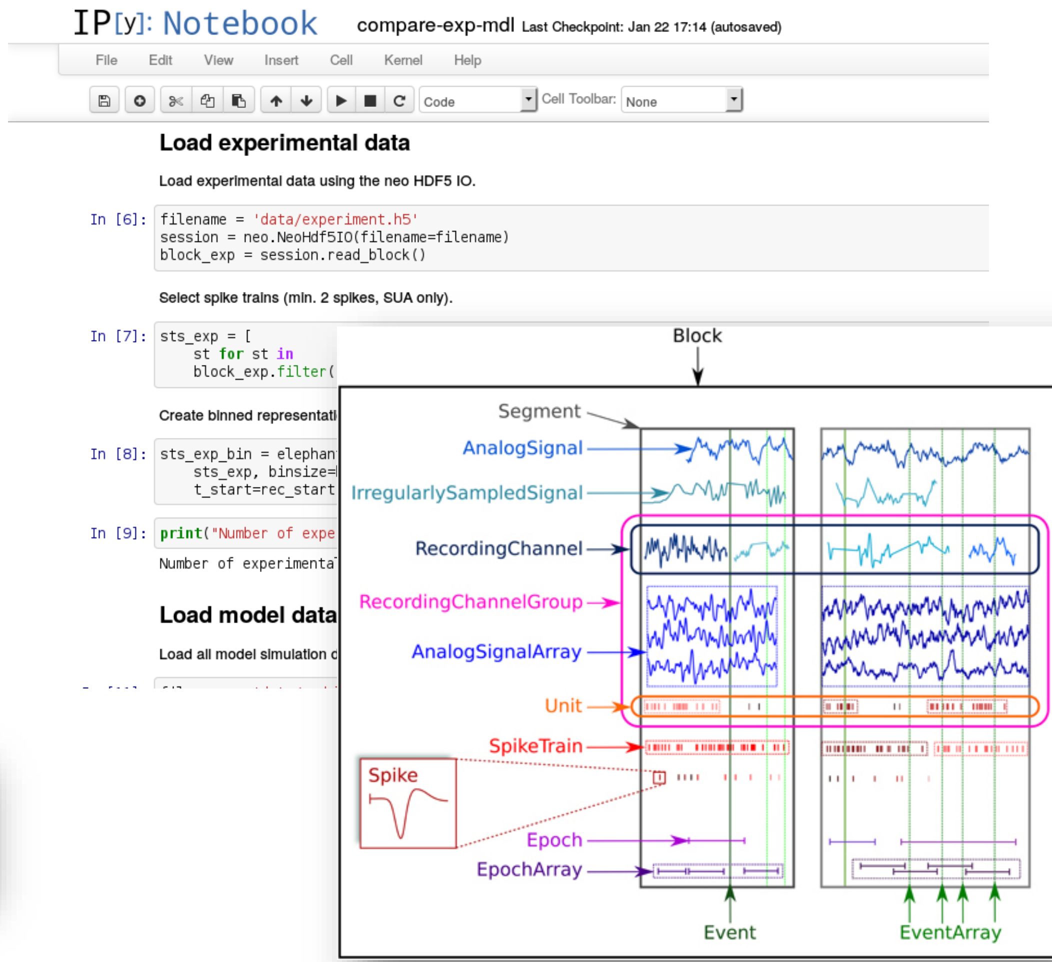


in the making…

- Spike detection from intracellular data
- Spike train metrics
- Correlation methods
- Higher order correlation
  - (Population) Unitary Events
  - SPADE
  - Synfire detection
  - CuBIC
  - State Space Analysis (Shimazaki)
  - Gravity clustering
- Current source density
  - Inverse CSD (Hagen)
  - Kernel CSD (Wojcik)
- Rate change point detection (Schneider)
- LFP / LFP-spike (phase) analysis
  - Spiketrain-field measures
  - Ridge detection
- Instantaneous rate estimation
- Kernels

Neo | common, vendor-independent representation of data

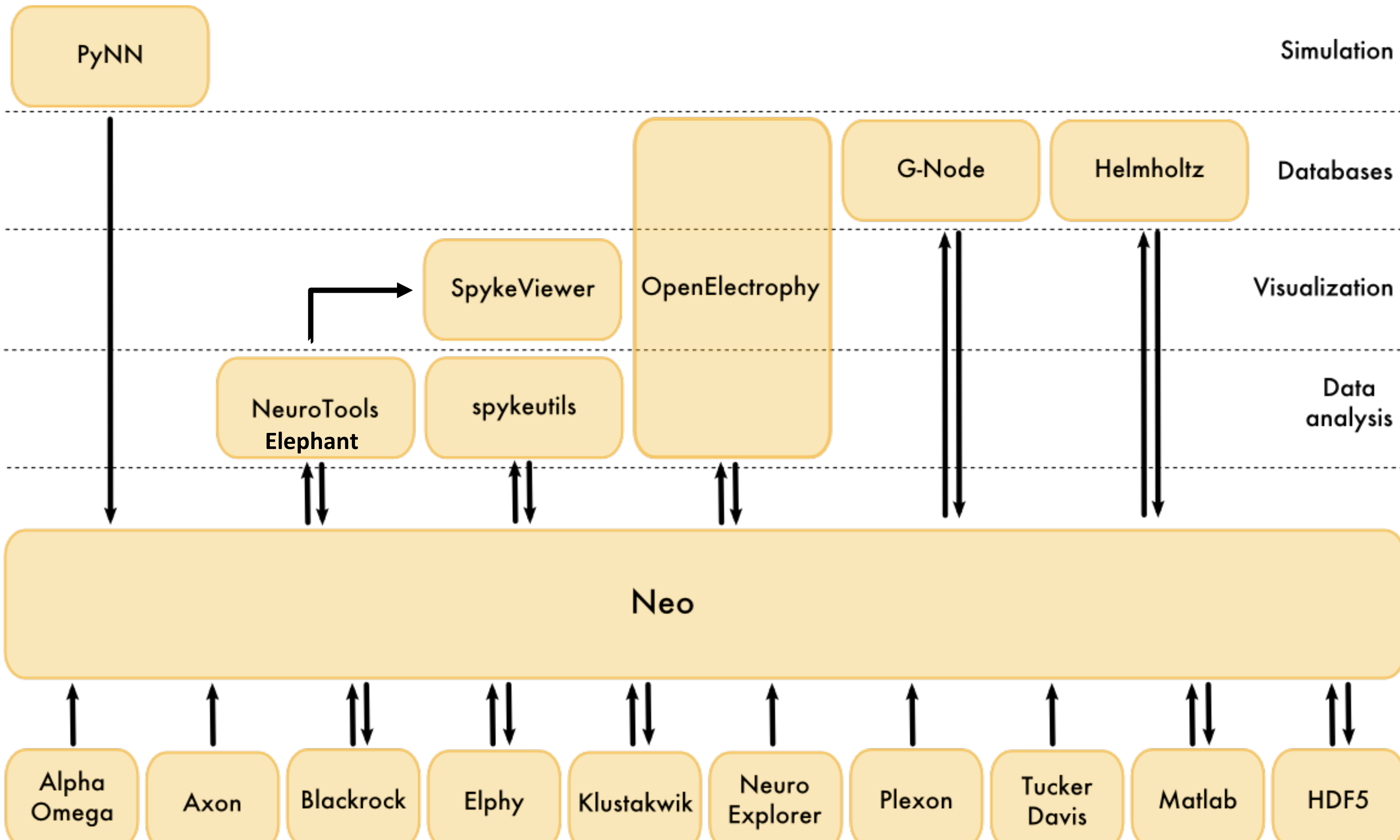
- *Elephant*: employs **Neo** for common internal data representation
- **load data from different (proprietary) formats into Neo** data object model
- Key concept: not a common file format, but **I/O bridge** to common object model
- Semantics delivered by annotations



Garcia, … , Davison (2014) Front Neuroinform

Neo | connecting diverse tools

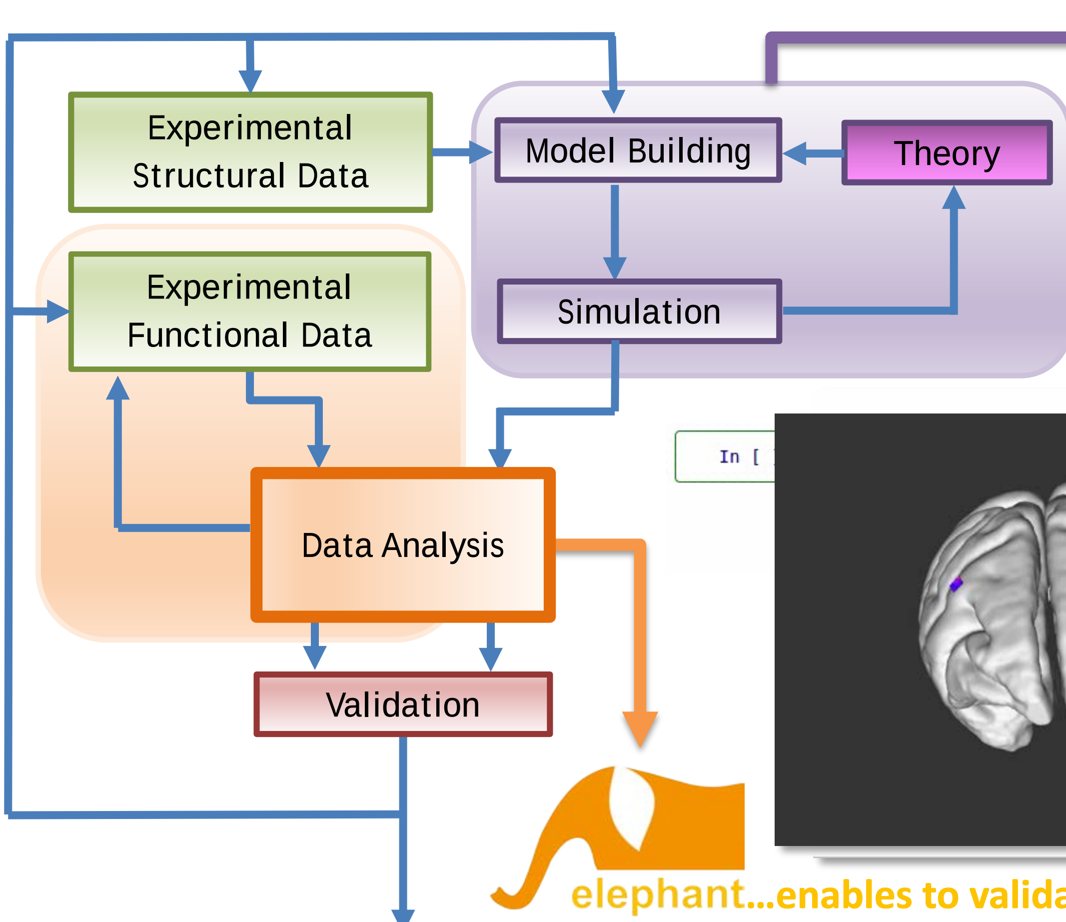
Use of Neo as a common data model and API to connect diverse software tools



Interactive Loops | validate neuronal simulations using experimental data

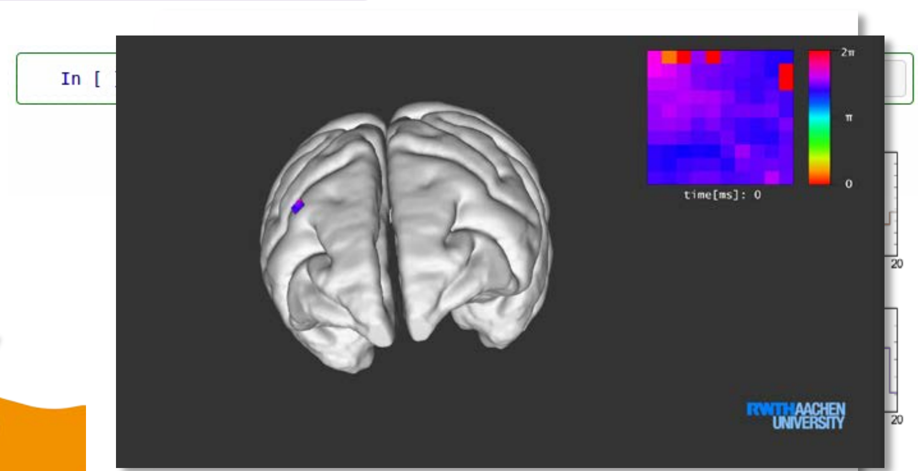
JÜLICH FORSCHUNGSZENTRUM

Work in *Interactive Loops* unlocks the potential of reproducible workflows built on a common software infrastructure



nest:: initiative

etc…



elephant ELECTROPHYSIOLOGY ANALYSIS TOOLKIT

neo

…enables to validate simulations on the level of measurements

…enables to validate simulations across scales of network dynamics