Slide 1

Data Sharing: Monkey multi-electrode array recordings and Neo

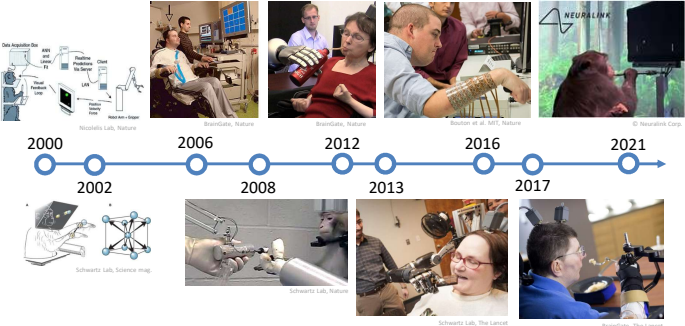Andres Agudelo-Toro and Hans Scherberger

NFDI Neuro 5th Community Workshop:
Integrating the Neuroscience Community
July 2021

Slide 2

Impressive progress of Brain Computer Interfaces

Slide 3
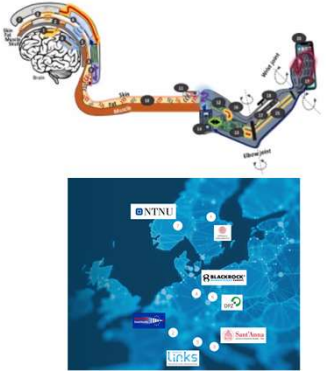
Project B-CRATOS (2021-2025)

- Wireless **B**rain-**C**onnect inte**Rf**A**ce **TO** machine**S**

- Large multi-site EU project (€ 4.6 m.)

- Aim: wireless **bidirectional** communication between the brain and a robotic device
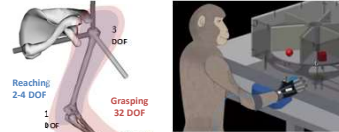
- More info: b-cratos.eu

Slide 4

Multi-lab Big-Data sharing challenge

- 4 brain areas
- 8 x 32 recording sites @ 24 KHz
- 32 joint angles @ 100 Hz
- 50 different objects
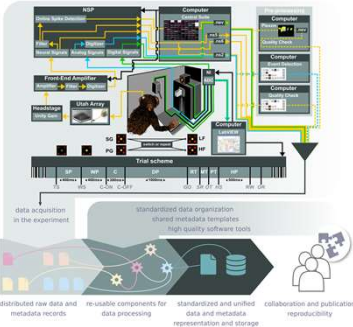- Hundreds of sessions
- Typical raw session: ~ 100 GB

- Preparation for real-time experimental phase requires simulation with existing data

- EU working groups focused on different levels:

  – Neural decoding

  – Robotic grasping

  – Neural stimulation, etc.

Preventing a data sharing "perfect storm"

DPZ

- Complex data set where details matter

- Sharing to non-electrophysiologists

- Until now customized formats (Lab's own protocols)

- Ideal case: automated tools before the experiment (not always possible)

5

Requirements

DPZ

- Open data formats

- Ideally open and freely accessible tools

- Compatible with collaborator's workflows (De facto Python)

  – But consider existing Matlab code

- Reusable

  – Training of Master's and PhD students

  – Future collaborations

- Good performance/low storage footprint

6

Considered formats

DPZ

- Neuroshare (defunct)

- **NWB: Neurodata Without Borders**

  – HDF5 Based. Python, Matlab support
  – Complex specs, more rigid
  – Relatively new, toolset still limited

- **NIX/Neo ecosystem:**

  – HDF5 Based. Python, Matlab, Java support
  – Simpler specs, more flexible
  – Actively maintained, more available tools

7

Experience so far with Neo

DPZ

- Fitted the project well:

  – Version 1.0: Use the Python tools but save as Matlab files

  – Version 2.0: Neo structure can be also be produced in Matlab using struct(...) and cell(..)

- Related tools (e.g. Elephant) can be easily used for common routines (binning, PETHs)

- Well documented (still lacks richer code examples)

neo

Used by 219

Contributors 69

GitHub

8

Some metrics

DPZ

- **Half a day** to write the code for basic translation of existing behavioral data
- **Two days** to add neural information
- Rough code length:
  - Version 1.0: .m **80 loc**, .py **40 loc**
  - Version 2.0: .m **160 loc**
- Spikes and behavioral data stored in .mat files in very manageable sizes (**1 GB per session**)
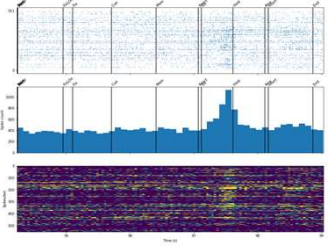
9

Code: a lot in very few lines

DPZ

10

Interaction with collaborators

DPZ

- So far neural data shared with the Swedish and Italian team
- Required Python modules shared to collaborators as a **conda** recipe
- Setting up the Python environment was possible in **10's of minutes**
- Collaborators could plot basic binned data from **day zero** using elephant
- Data shared with 1 Master's student, successful machine learning project completed in **8 weeks**

11

Summary

DPZ

- Ecosystem keeps evolving with standard formats and tools
- Python and Conda greatly facilitate sharing work environments
- More than the storage format (e.g. nix vs nwb): it is important to agree upon a **data structure**
- That is what you will deal with 90% of the time!
- If there is a common structure, community tools will develop around it

12