



**HAL**  
open science

# Open science: A journey from sharing research artefacts to collaborative research

Camille Maumet

## ► To cite this version:

Camille Maumet. Open science: A journey from sharing research artefacts to collaborative research. Rigor and Reproducibility Seminar Series - UF Interdisciplinary T32 in Movement Disorders and Neurorestoration, Jul 2022, Gainesville (Online), United States. hal-03716748

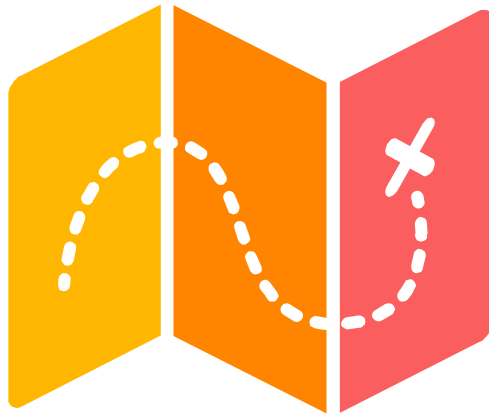
**HAL Id: hal-03716748**

**<https://inria.hal.science/hal-03716748v1>**

Submitted on 7 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

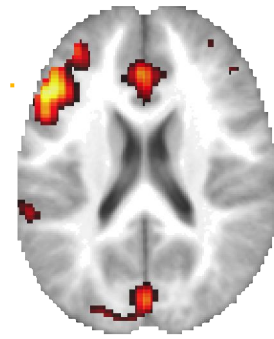


# Open science A journey from sharing research artefacts to collaborative research

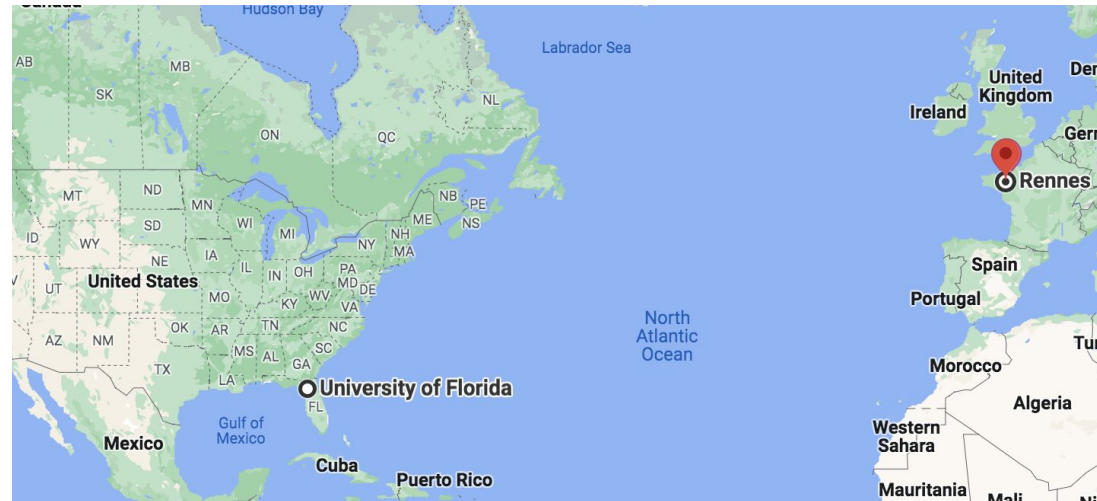
Camille Maumet

Inria, Univ Rennes, Inserm, CNRS, IRISA

# Hi! I'm Camille



Research scientist in neuroinformatics  
Rennes, France

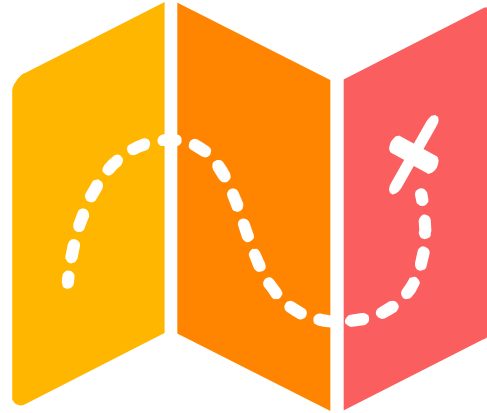


PhD in brain imaging methods

PostDoc in neuroimaging data sharing → ✨ open science ✨



# A journey into Open Science



## Step 1

More  
**reproducible**  
research



SCIENCE

## A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



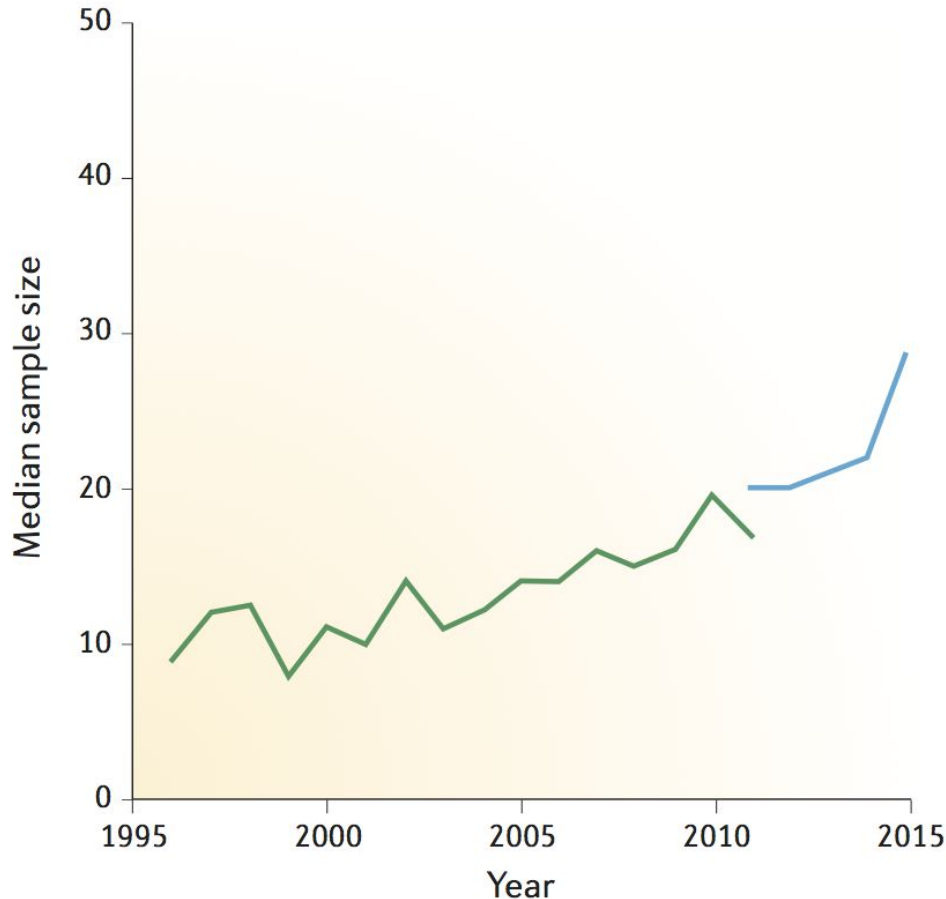
SEAN NEL / SHUTTERSTOCK

In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

**(Ir)-Reproducibility**

# Sample sizes in neuroimaging



2015 : 30 participants per study

[Poldrack et. al, Nature Neuroscience 2017]



# Open data

Unique study  
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T Q R V  
S M Y C O N N E C T O M E Q  
G S P K K W A R R G K E H R

NITRC

OSF

- + Images
- + Homogenous
- Datasets



# Open data

Unique study  
30 participants



OpenNEURO

studyforrest.org

Consortium  
1000 participants



ABIDE  
Autism Brain Imaging  
Data Exchange

1000 Functional  
Connectomes Project

NEUROVAULT

L I E L N L F T Q K T Q R V  
S M Y C O N N E C T O M E C  
G S P K K W A R R G K E H R



NITRC

OSF

CORR  
CONSORTIUM FOR  
RELIABILITY AND  
REPRODUCIBILITY



- + Images
- + Homogenous
- Datasets





# Open data

Unique study  
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T O R V  
S M Y C O N N E C T O M E C  
G S P K K W A R R G K E H R

Consortium  
1000 participants



ABIDE  
Autism Brain Imaging  
Data Exchange

1000 Functional  
Connectomes Project



NITRC

OSF

CORR  
CONSORTIUM FOR  
RELIABILITY AND  
REPRODUCIBILITY



Cohort  
100 000 participants



HUMAN  
Connectome  
PROJECT

biobank<sup>uk</sup>  
Improving the health of future generations

- + Images
- + Homogenous
- Datasets

# Towards FAIR datasets



# Towards FAIR datasets



*Donders Repository*



# Towards FAIR datasets



OpenNEURO



LORIS

*Donders Repository*



**Findable** we can search for open datasets in each repository

# Towards FAIR datasets



*Donders Repository*



**Findable** we can search for open datasets in each repository

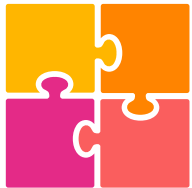


**Accessible** Direct download or with Data User Agreement

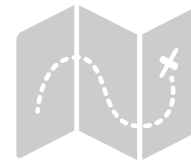
# Towards FAIR datasets



*Donders Repository*



**Interoperable?**



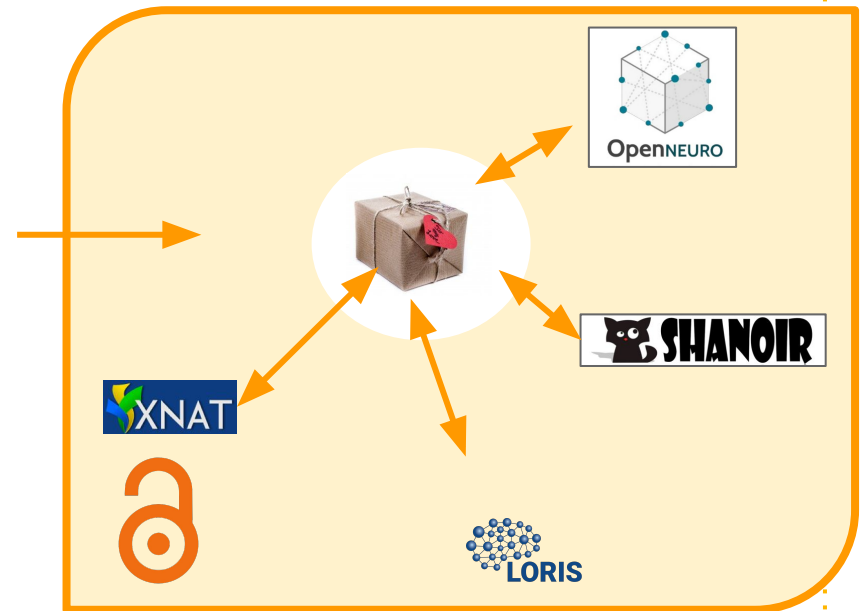
**Findable** we can search for open datasets in each repository



**Accessible** Direct download or with Data User Agreement

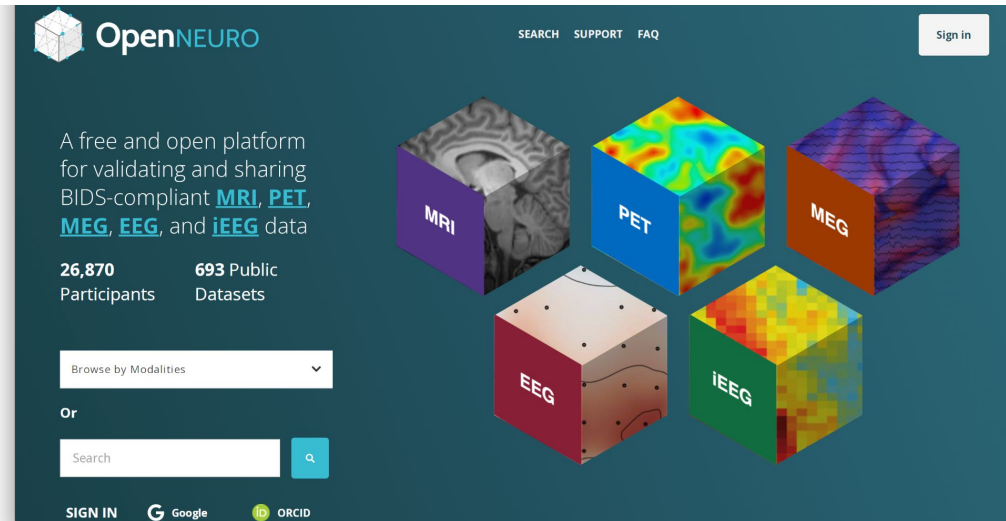
# Standards for data sharing

- Data and **meta-data**
- A common “language”



# Standards for data sharing - BIDS

```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   ├── func/  
│   │   ├── sub-01_task-rest_bold.nii.gz  
│   │   └── sub-01_task-rest_bold.json  
│   └── dwi/  
│       ├── sub-01_dwi.nii.gz  
│       ├── sub-01_dwi.json  
│       ├── sub-01_dwi.bval  
│       └── sub-01_dwi.bvec  
├── sub-02/  
├── sub-03/  
└── sub-04/
```



<https://openneuro.org/>

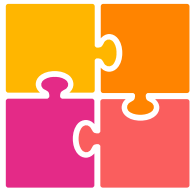
[Gorgolewski et. al, Scientific Data 2016]



# Towards FAIR datasets



*Donders Repository*



**Interoperable?**



**Findable** we can search for open datasets in each repository

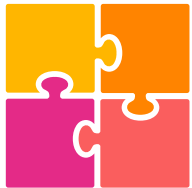


**Accessible** Direct download or with Data User Agreement

# Towards FAIR datasets



*Donders Repository*



**Interoperable** A common languages across tools



**Findable** we can search for open datasets in each repository

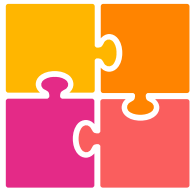


**Accessible** Direct download or with Data User Agreement

# Towards FAIR datasets



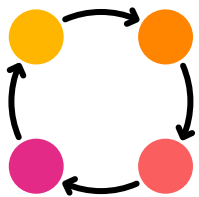
*Donders Repository*



**Interoperable** A common languages across tools



**Findable** we can search for open datasets in each repository



**Reusable** Metadata selected to enable reuse



**Accessible** Direct download or with Data User Agreement

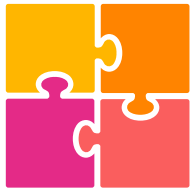
# Towards FAIR datasets



*Donders Repository*



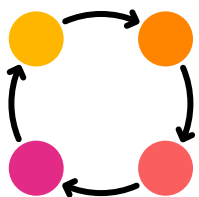
**F**indable **A**ccessible **I**nteroperable **R**eusable



**Interoperable** A common languages across tools



**Findable** we can search for open datasets in each repository

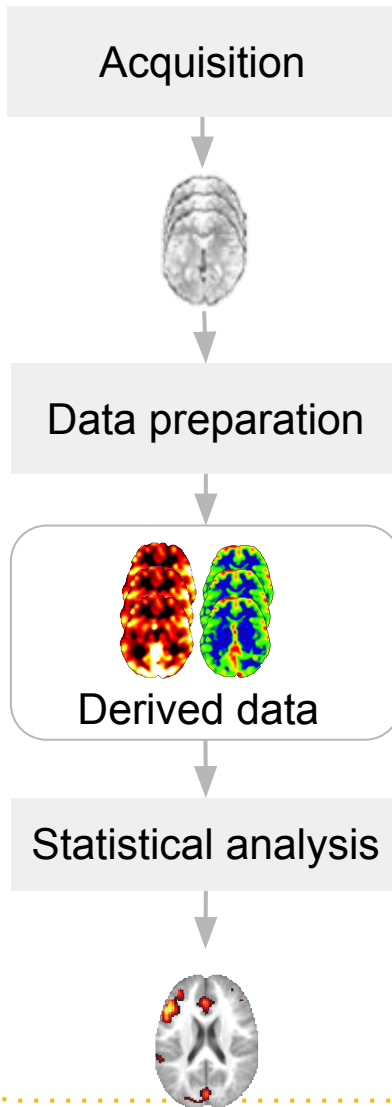


**Reusable** Metadata selected to enable reuse

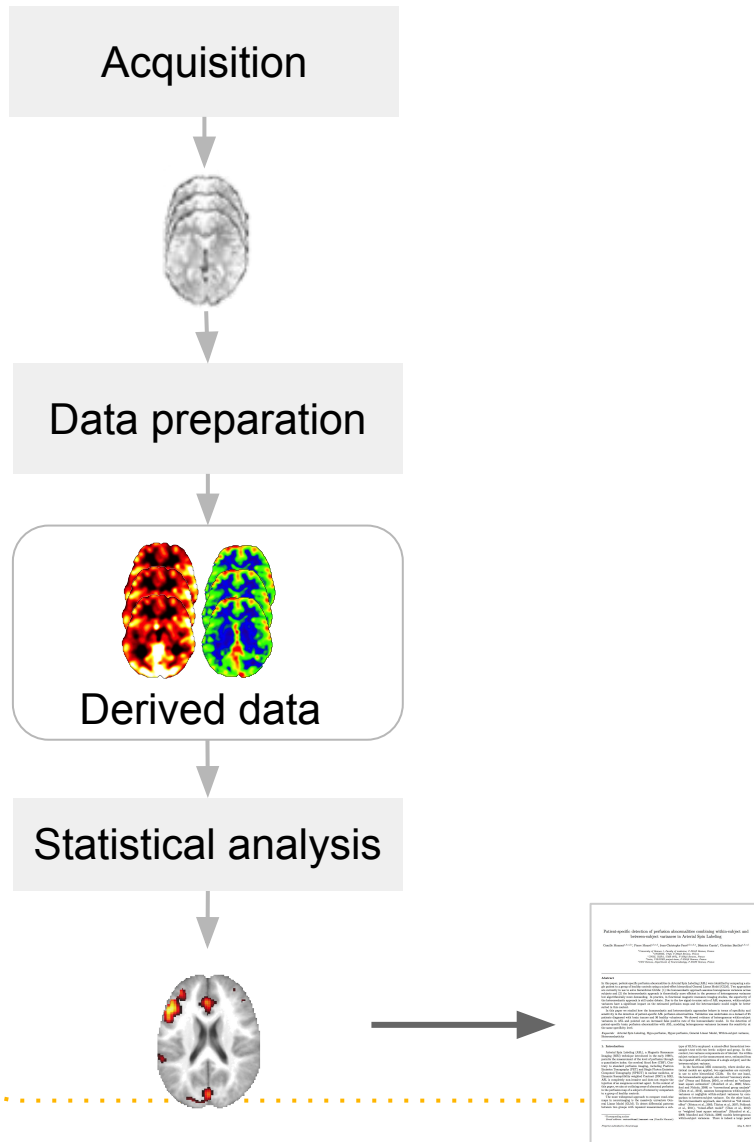


**Accessible** Direct download or with Data User Agreement

# Sharing more research products

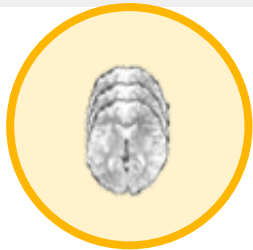


# Sharing more research products



# Sharing more research products

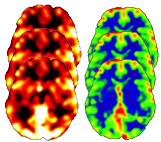
Acquisition



*Donders Repository*



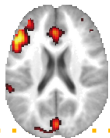
Data preparation



Derived data



Statistical analysis



# Sharing more research products

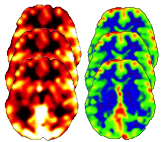
Acquisition



*Donders Repository*



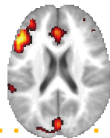
Data preparation



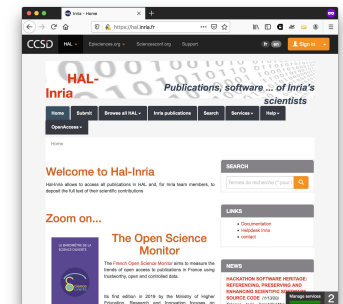
Derived data



Statistical analysis



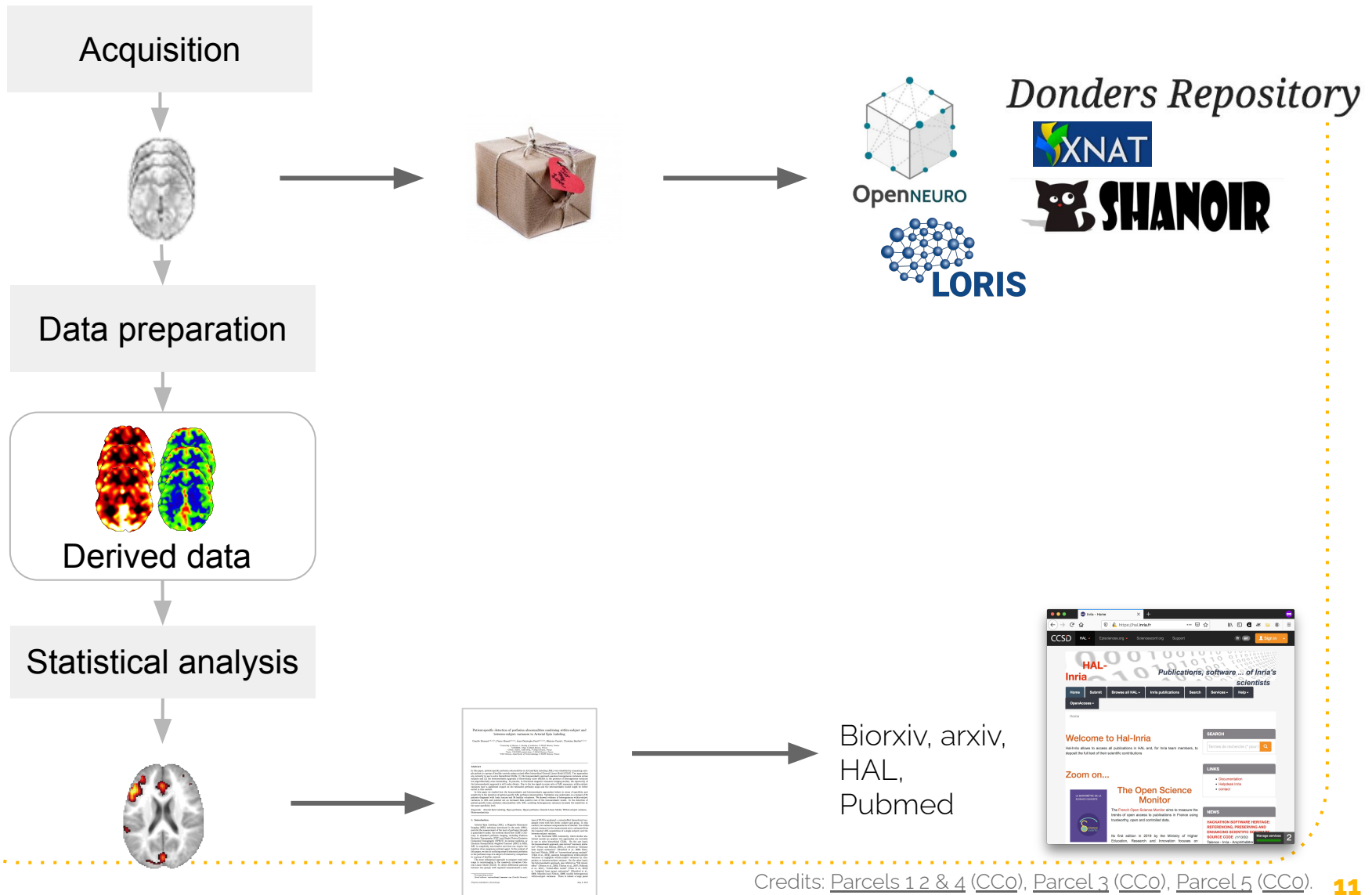
Biorxiv, arxiv,  
HAL,  
Pubmed



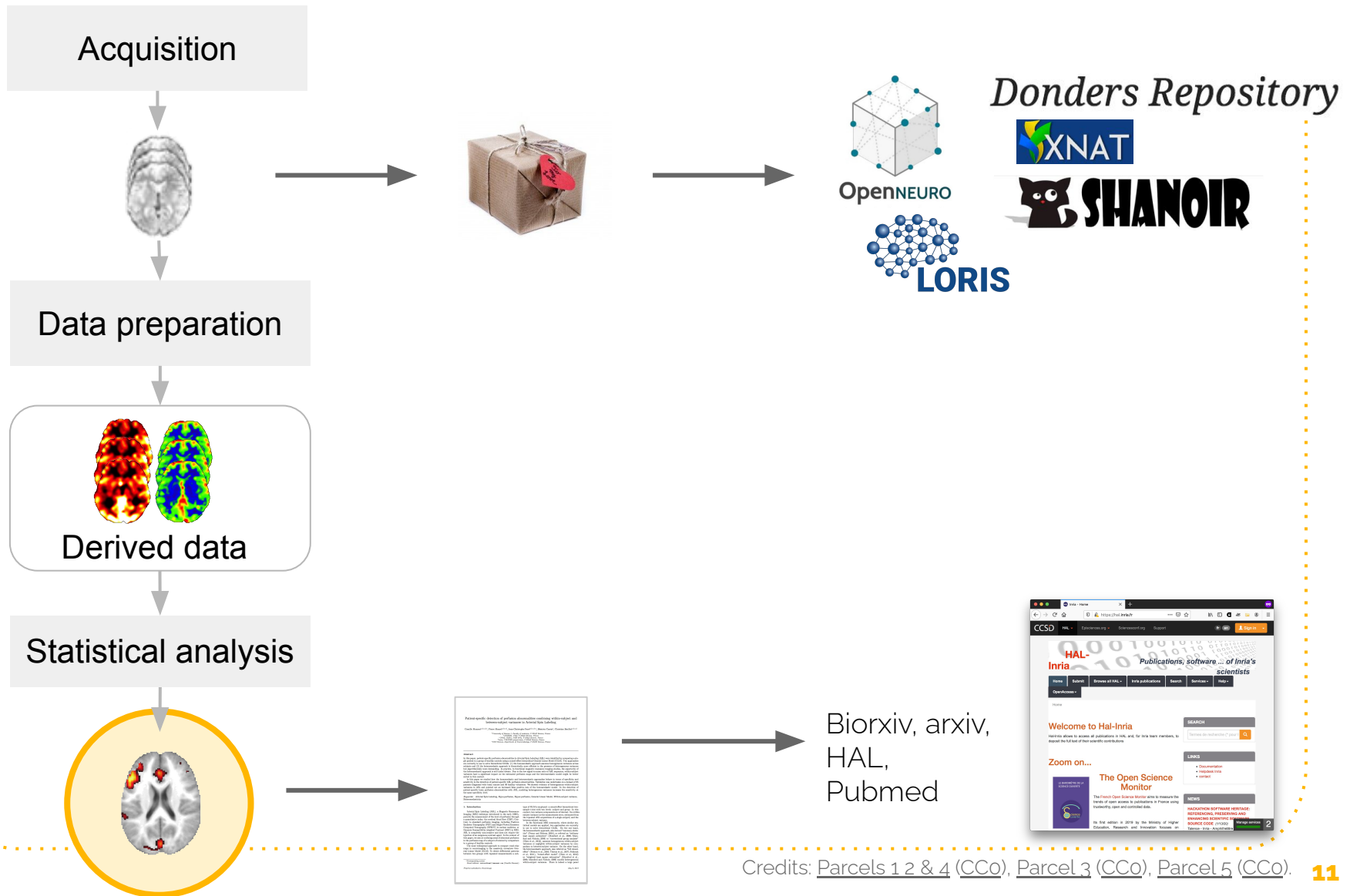
Credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).



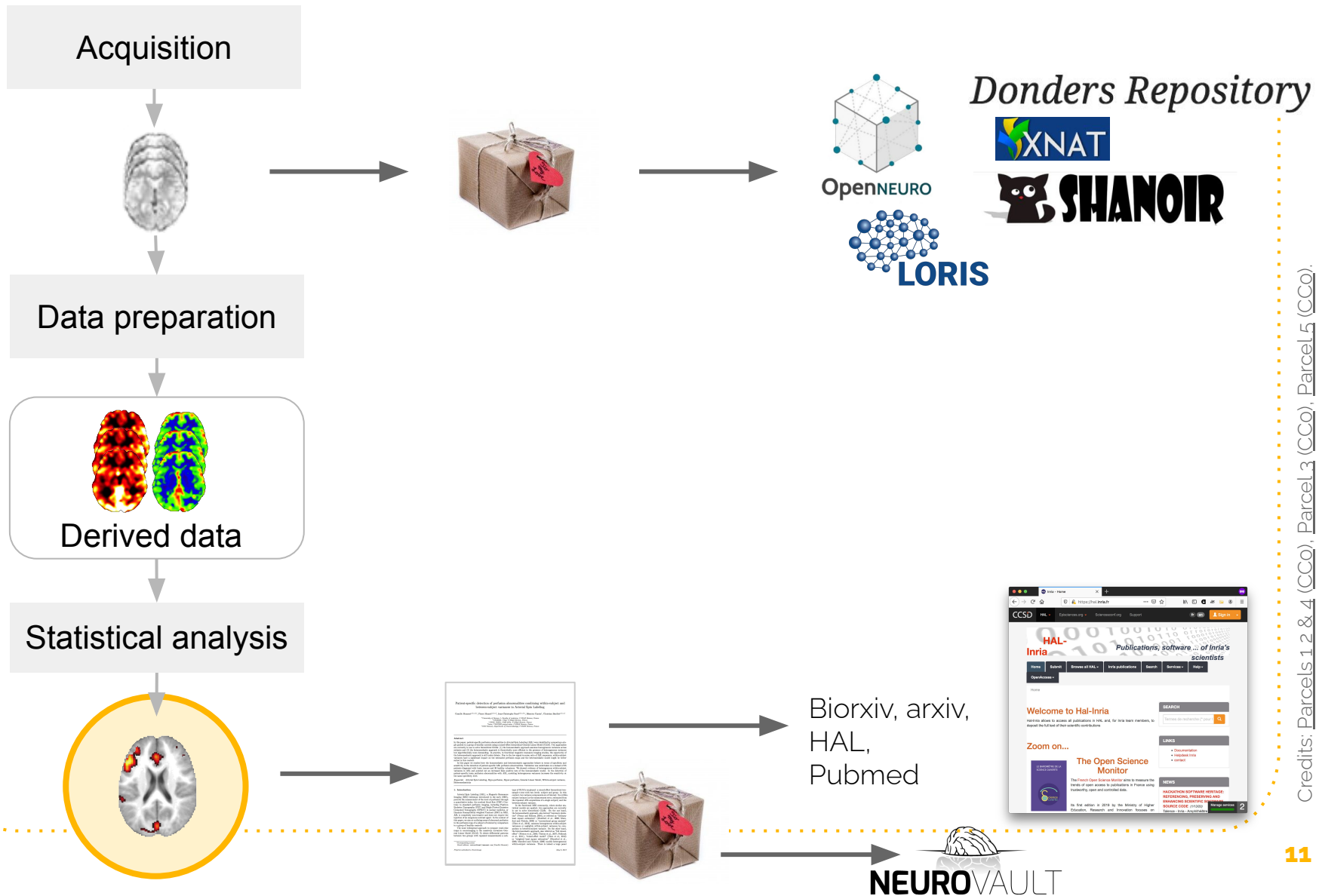
# Sharing more research products



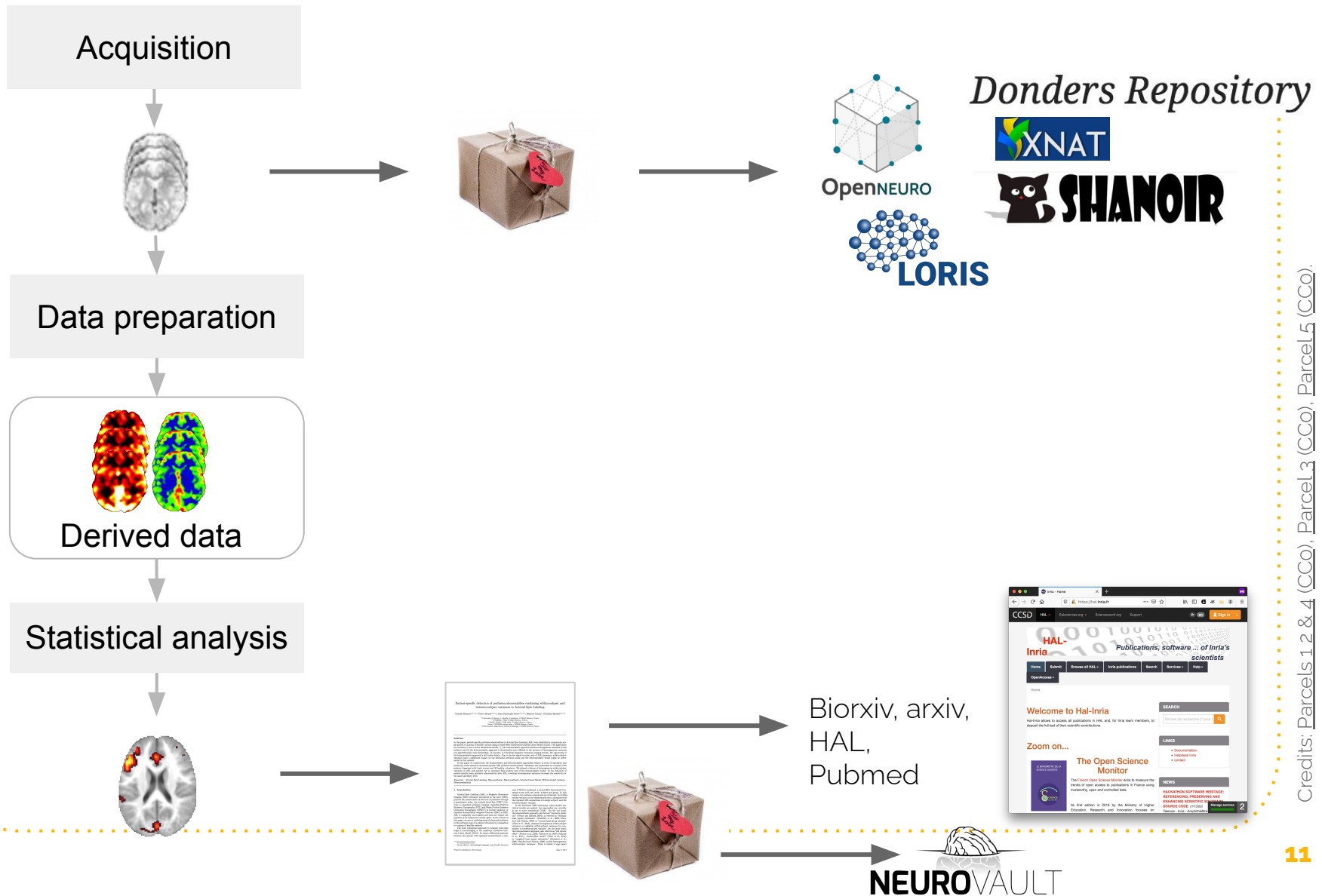
# Sharing more research products



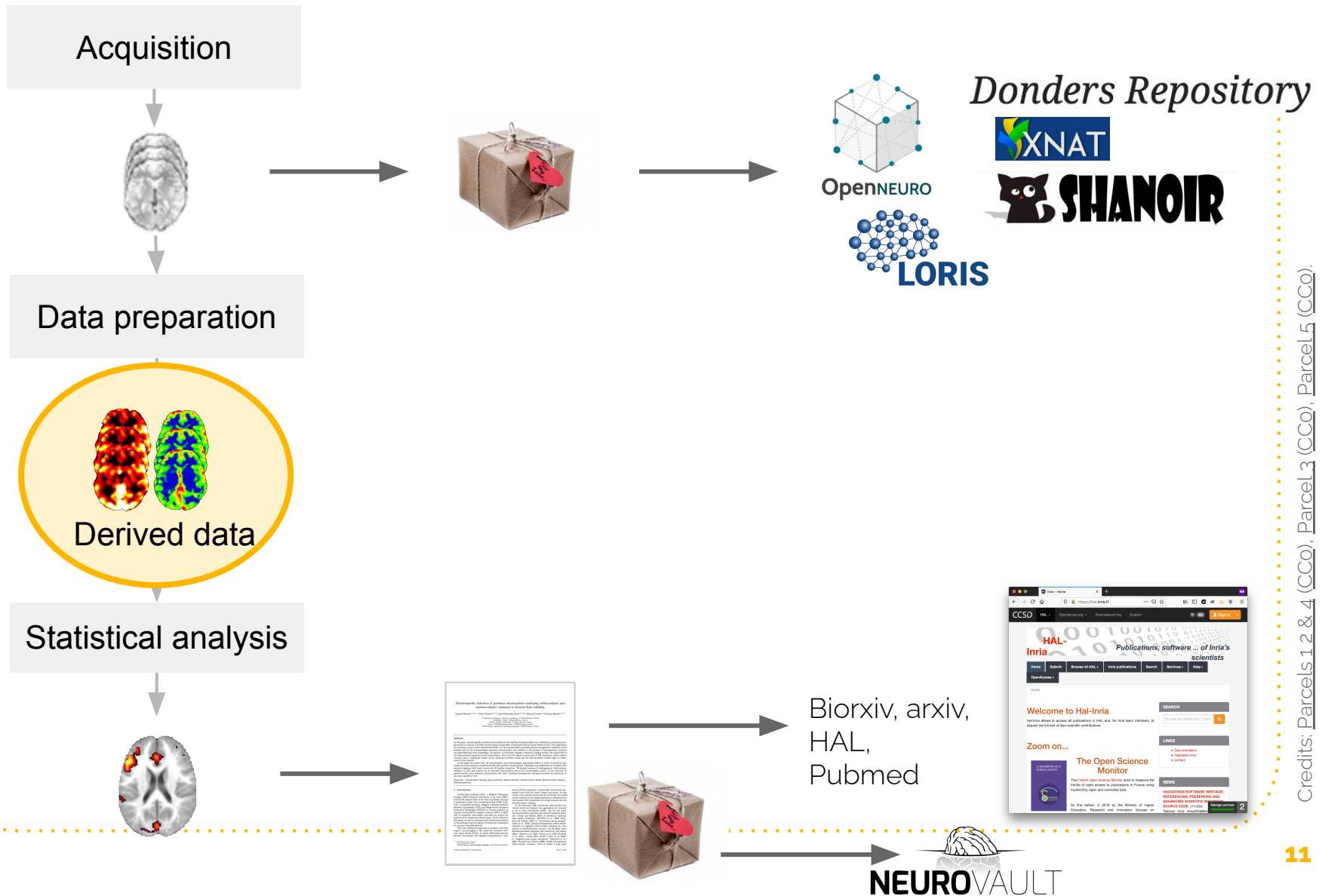
# Sharing more research products



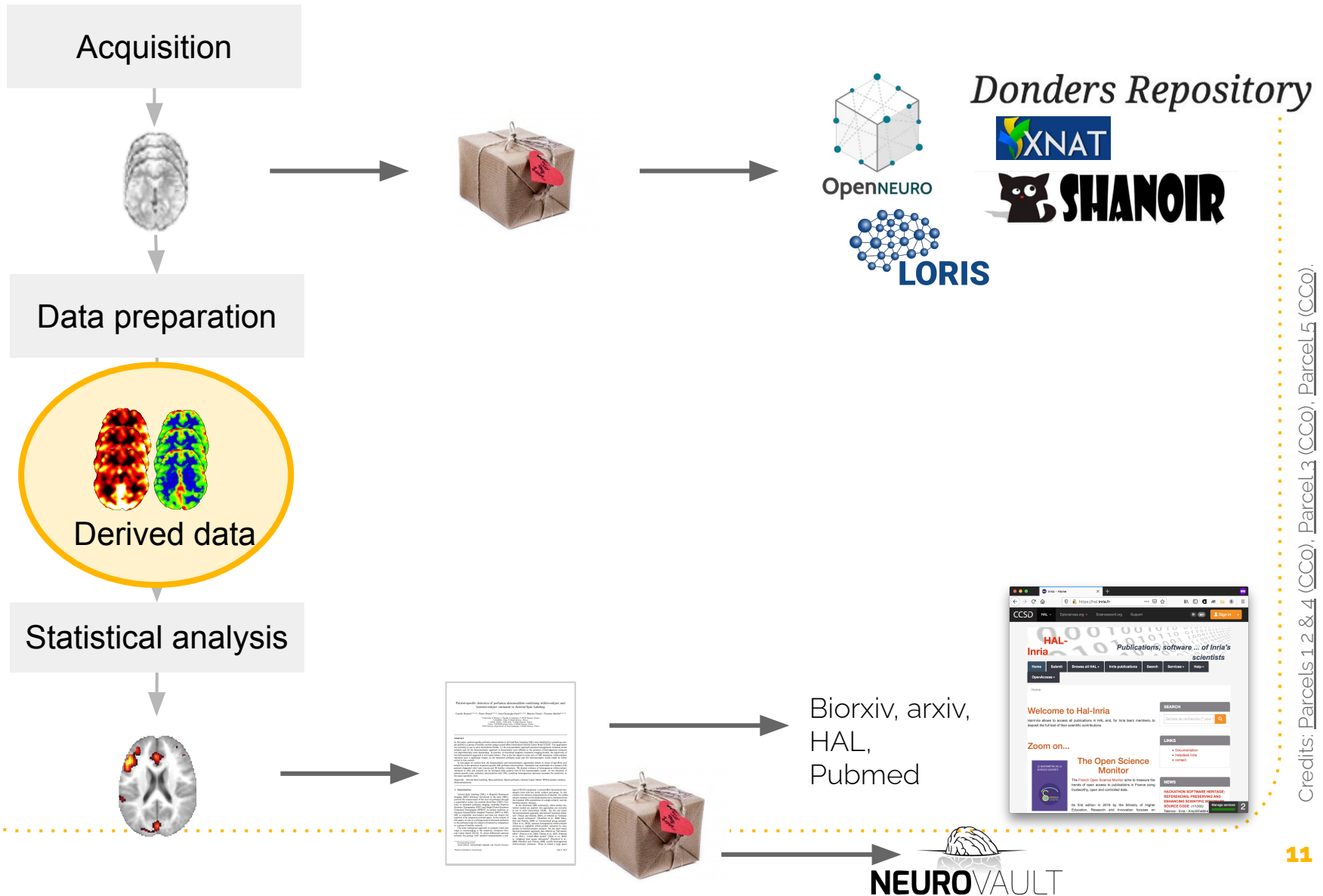
# Sharing more research products



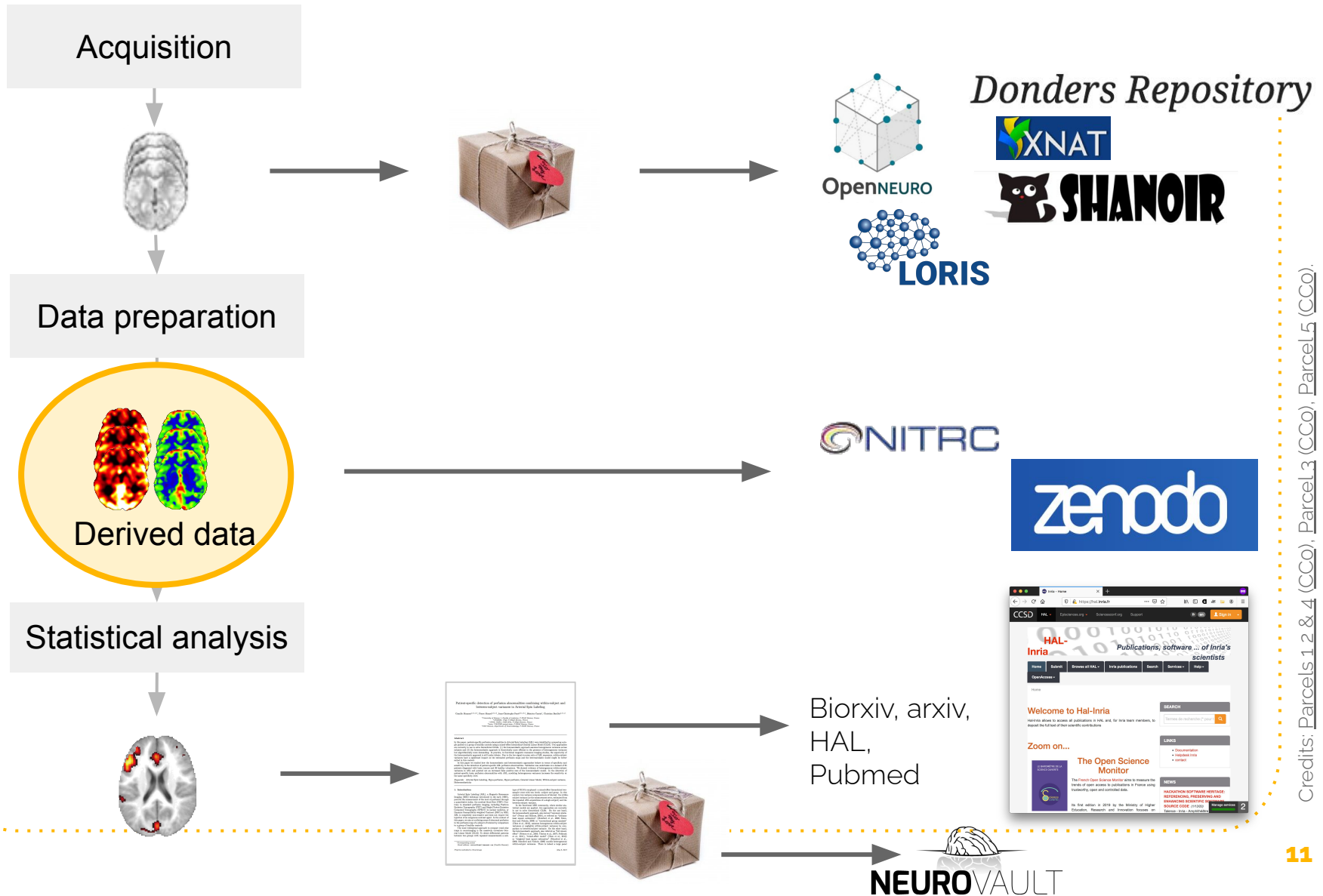
# Sharing more research products



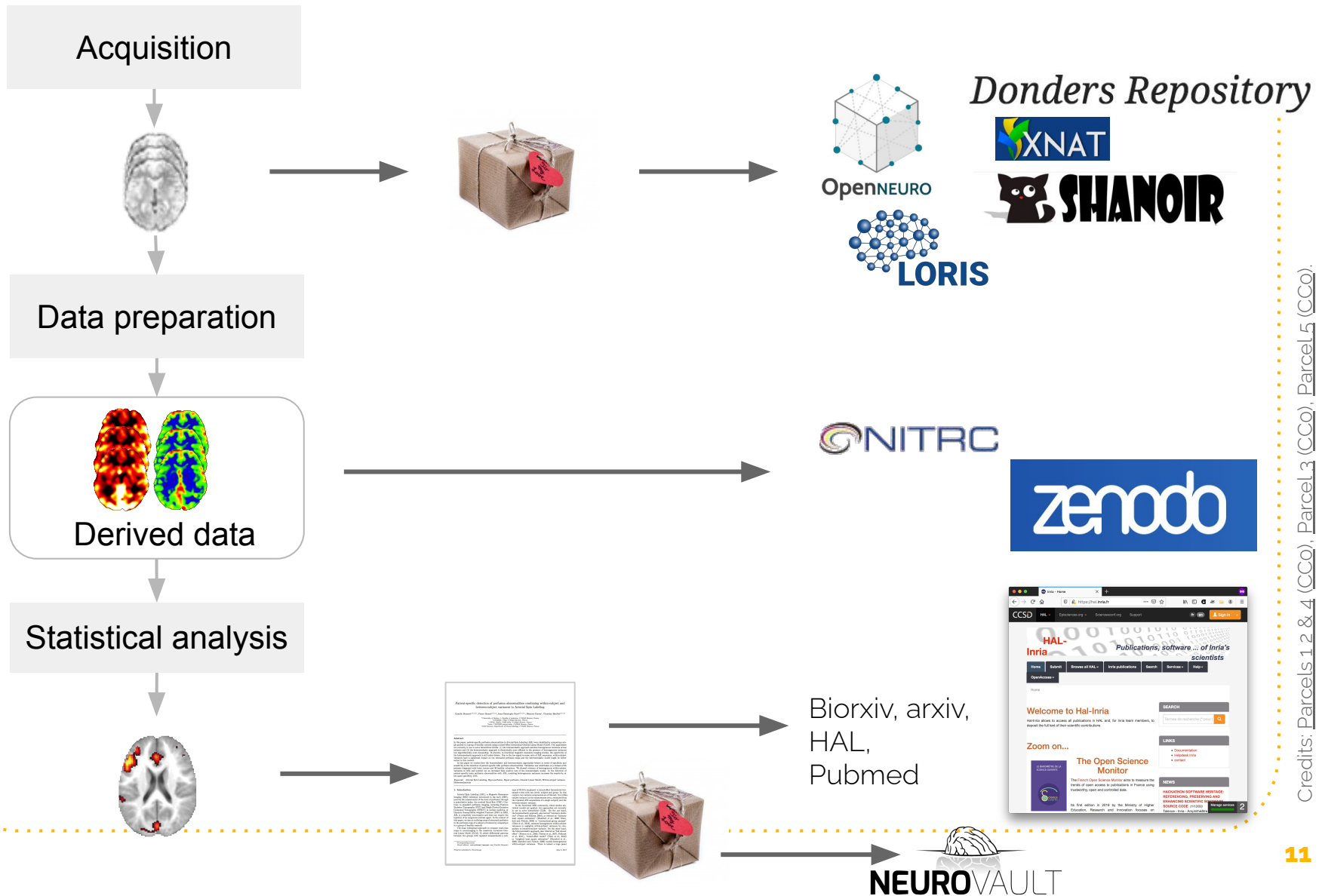
# Sharing more research products



# Sharing more research products

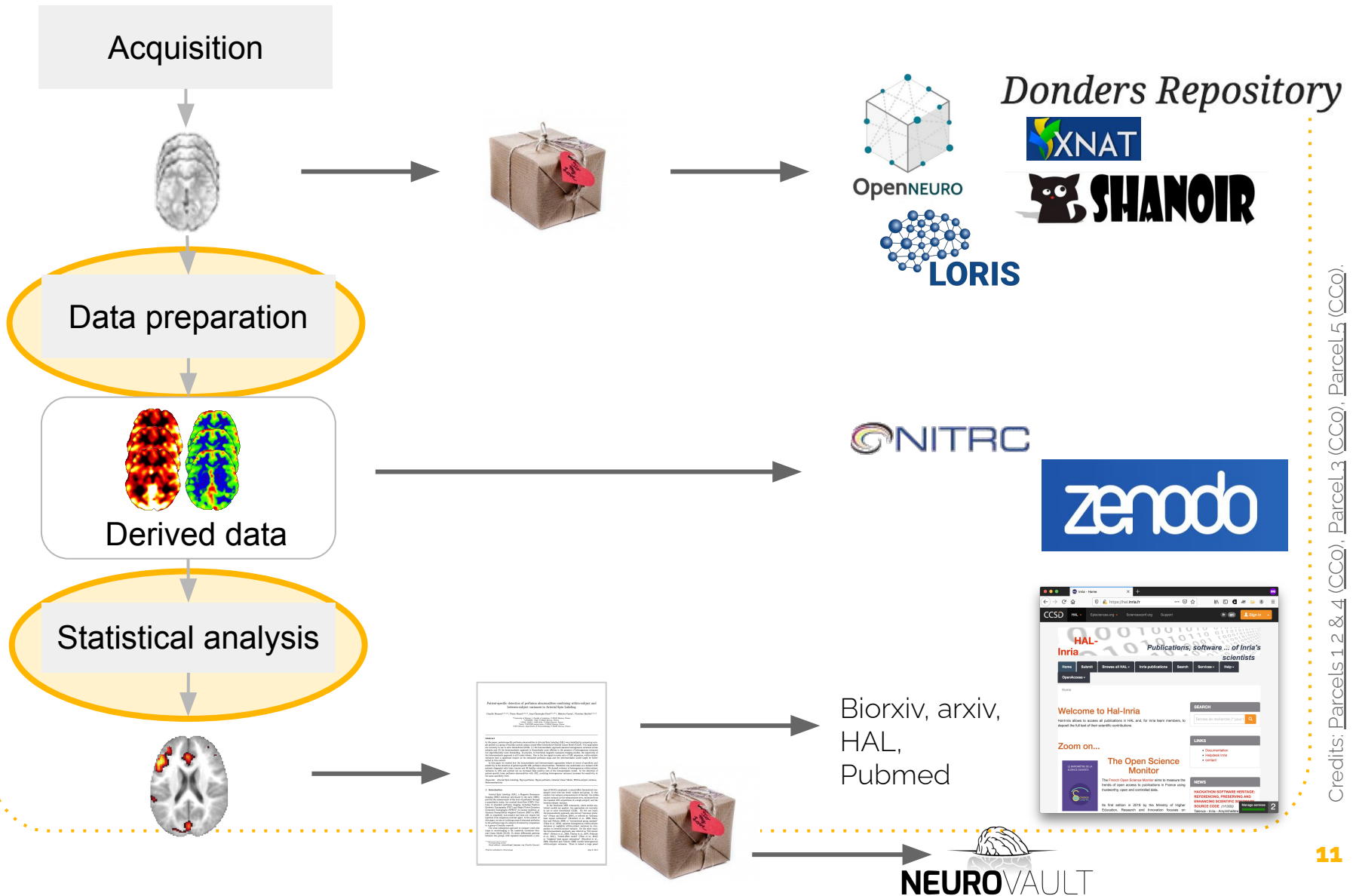


# Sharing more research products





# Sharing more research products



# Sharing analysis code

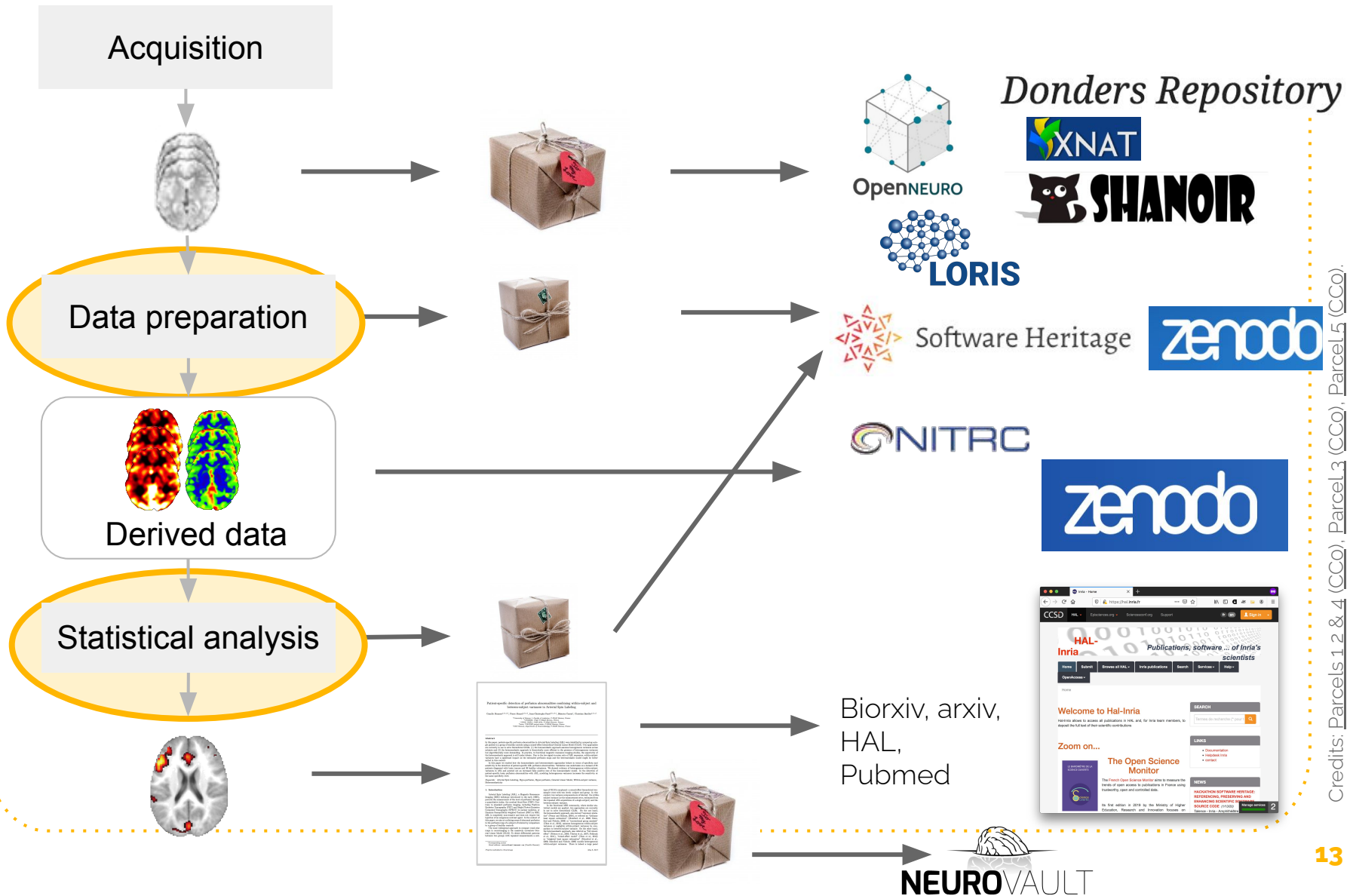


**Different requirements** for sharing code compared to datasets

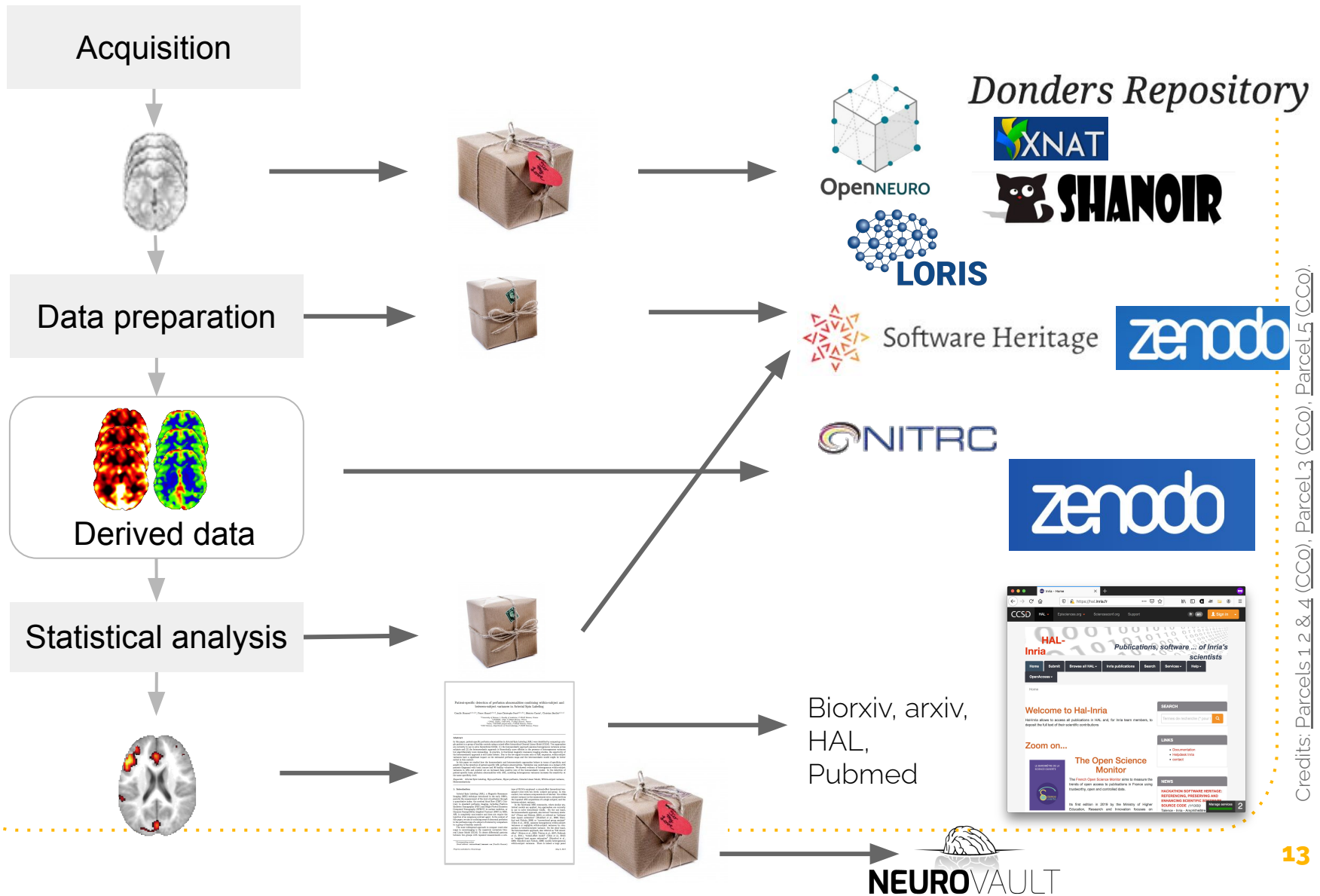
- Often stored in **online repositories** (e.g. Github, Gitlab)
- **Many** versions (commits)
- Long term preservation
- **Code structure**, README

```
.
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|   -- birds_count_table.csv
|-- doc
|   -- notebook.md
|   -- manuscript.md
|   -- changelog.txt
|-- results
|   -- summarized_results.csv
|-- src
|   -- sightings_analysis.py
|   -- runall.py
```

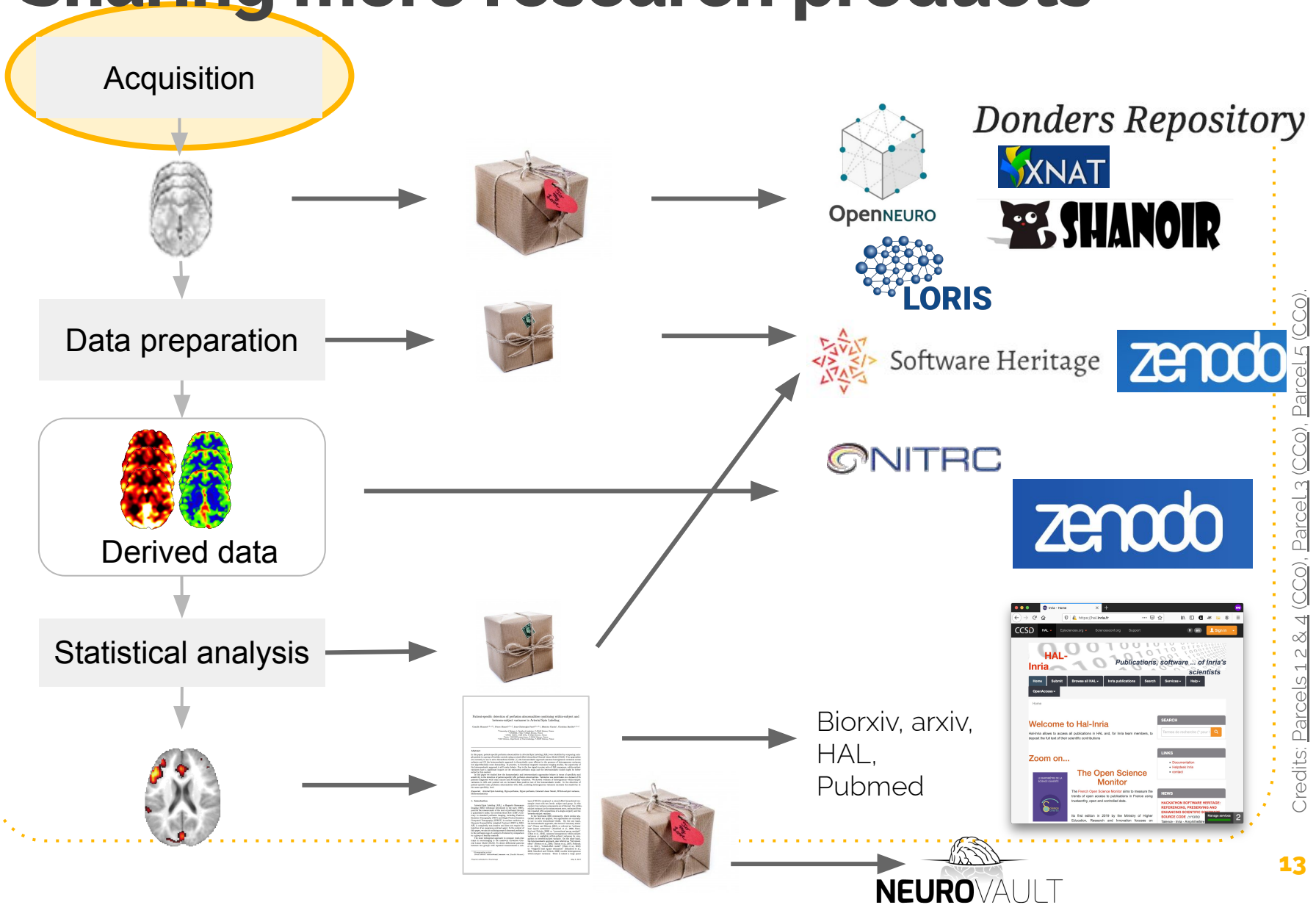
# Sharing more research products



# Sharing more research products



# Sharing more research products



# Let's get started...



- **Overwhelmed?** Lack of training in open science practices
- **Lack of incentives** How can we as individuals acts towards more reproducible science?

# Reproducible Science



Beyond technical solutions, a culture change

- Making resources **available to the scientific community**
- **Reusing** existing resources (datasets, tools) that were created by others

**towards more collaborative research**



**Chris Holdgraf**

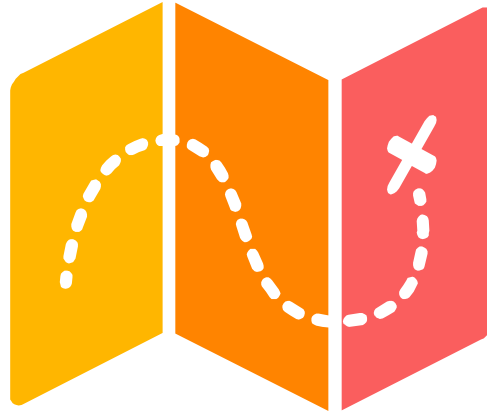
@choldgraf

Something I'm thinking about today: an important part of Open Science and Open Infrastructure is the idea that the end-product is not the only thing that matters. The *\*process to get there\** is equally important. 1/

[Traduire le Tweet](#)

11:31 AM · 30 juin 2022 · Twitter Web App

# A journey into Open Science



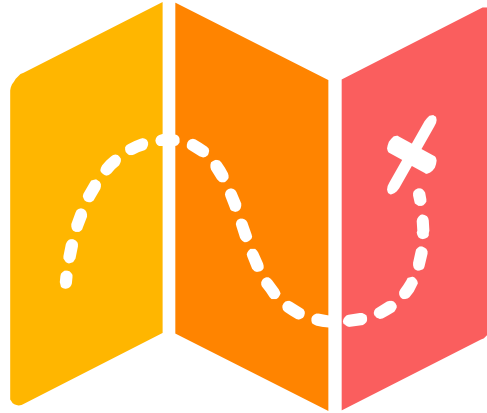
## Step 1

More  
**reproducible**  
research





# A journey into Open Science

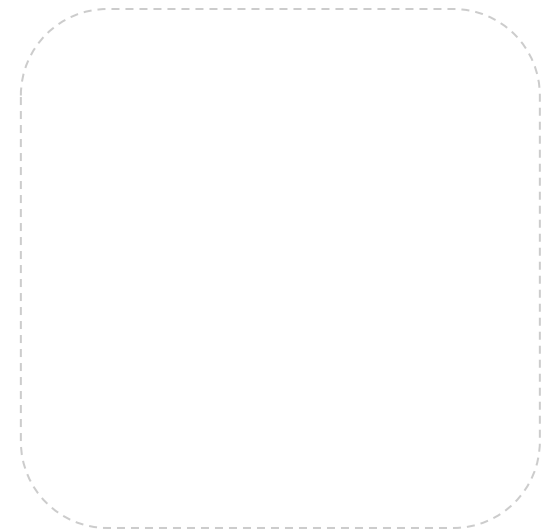


## Step 1

More  
**reproducible**  
research

## Step 2

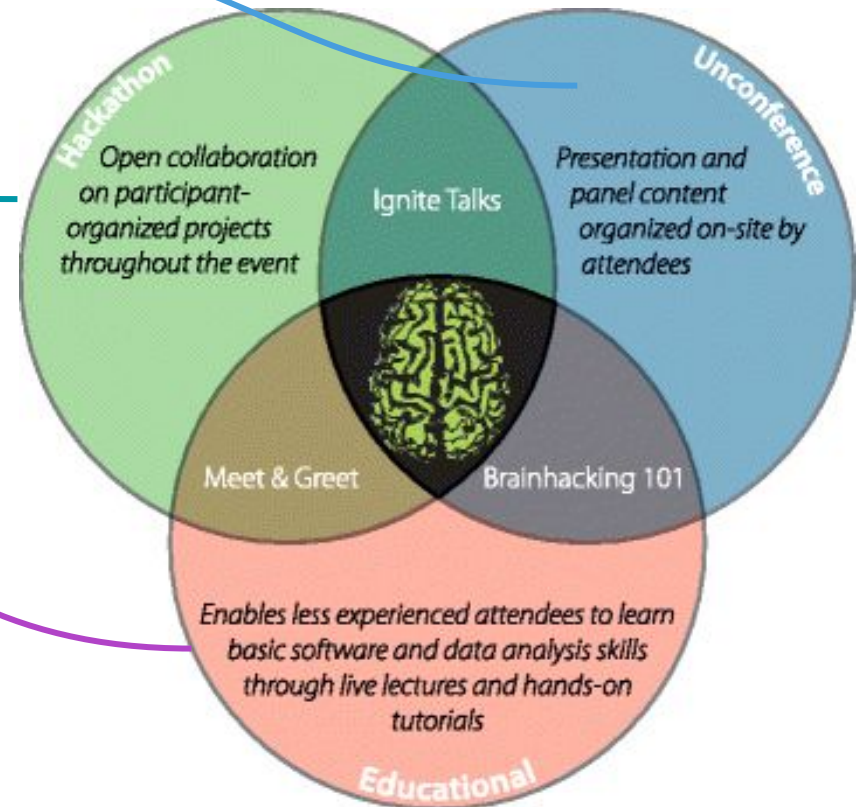
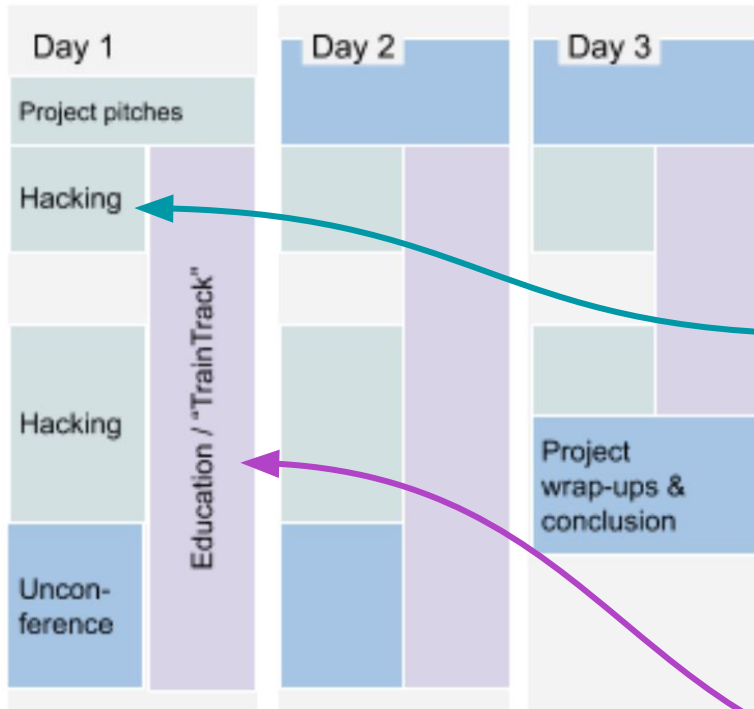
More  
**collaborative**  
research



# Hackathons

*"Hackathons based on collaboration, not competition"*

## B. Timeline of a Single Event



# Hackathon: a space for learning

## **Traintrack**

Courses covering many aspects of open science practices (e.g. version control, containers, etc.)



## **Unconferences**

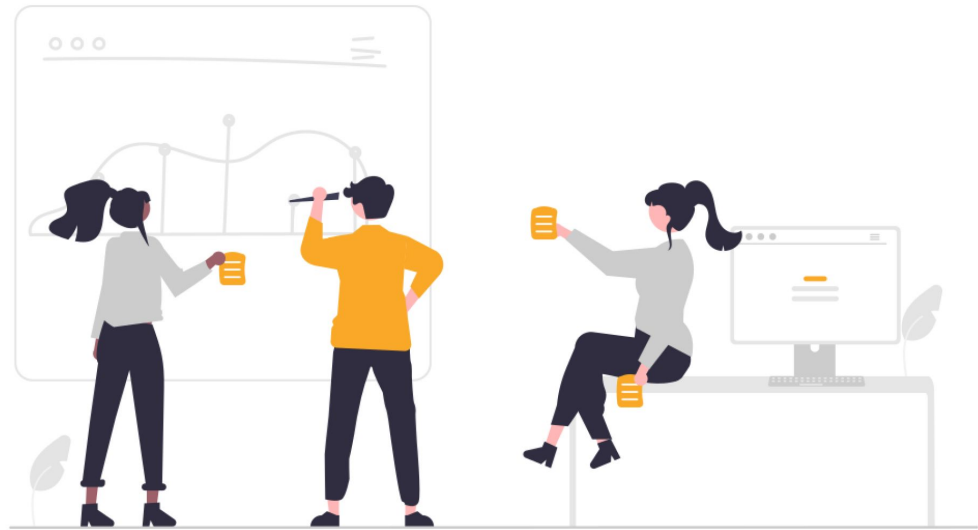
Spontaneous short sessions of a variety of topics

## **Community learning**

Hearing about new projects, discussing with peers, etc.

# Hackathon: a space for building

**Hacktrack** Finding critical mass and diverse opinions to work on projects that you care about



# Data privacy & sharing



How to inform participants, collect consent for data sharing in compliance with European General Data Protection Regulation (GDPR)?

## Open Brain Consent

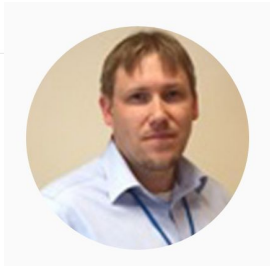
[Open Brain Consent WG, HBM 2020]



Stephan Heunis



Yarik Halchenko



Cyril Pernet

Open Brain Consent  
stable

Search docs

- Sample consent forms
- Recommendations
- Ultimate consent form
- GDPR edition
- Anonymization tools
- Contribute
- Contact information
- Discussions

Docs » Make open data sharing a no-brainer for ethics committees. [Edit on GitHub](#)

### Make open data sharing a no-brainer for ethics committees.

[Zenodo badge](#)

#### Statement of the problem

The ideology of open and reproducible science makes its ways into various fields of science. Neuroimaging is a driving force today behind many fields of brain sciences. Despite possibly terabytes of neuroimaging data collected for research daily, just a small fraction becomes publicly available. Partially it is because management of neuroimaging data requires to confirm to established legal norms, i.e. addressing the aspect of research participants privacy. Those norms are usually established by institutional review boards (IRB, or otherwise called ethics committees), which are in turn "governed" by national, federal and supra-national regulations.

Flexibility in interpretation of original regulations established in the past century, decentralization of those committees, and lack of a "community" influence over them created the problem: for neuroimaging studies there was no commonly accepted version of a Consent form template which would allow for collected imaging data to be shared as openly as possible while providing adequate guarantees for research participants' privacy. In majority of the cases, used Consent forms simply did not include any provision for public sharing of the data to get a "speedy" IRB approval for a study. Situation is particularly tricky because major granting agencies (e.g. NIH, NSF, RCUK) nowadays require public data sharing, but do not provide explicit instructions on how.

DigitalOcean Save time & money w/ the cloud platform loved by devs. Try for Free  
Sponsored - Ads served ethically

# Data privacy & sharing



## Consent form

[Direct link](#)

To be approved and signed by participants

“[...] While the collection, use and storage of your data are done for the purpose of conducting the study to which you are currently participating, **these data might also be used for other future research projects in the field of medical and cognitive neuroscience.** [...]”

## Data user agreement

[Direct link](#)

To be approved and signed by researchers reusing the data

“[...] 2. I **will not attempt to establish or retrieve the identity of the study participants.** I will not link these data to any other database in a way that could provide identifying information. [...]”

3. I **will not redistribute these data** or share access to these data with others, unless they have independently applied and been granted access to these data, i.e., signed this Data Use Agreement. This includes individuals in my institution.  
[...]”

## Translations

- Bosnian
- Czech
- German
- Greek
- Spanish
- Finnish
- French
- Italian
- Dutch
- Norwegian
- Turkish

## French version



Anne Hespel



Elise Bannier

# Hackathon: a space to meet with the community

Finding **people who care about the process** of how we do science



Finding **collaborators** for your next research project

Greater sense of belonging into the **Community**

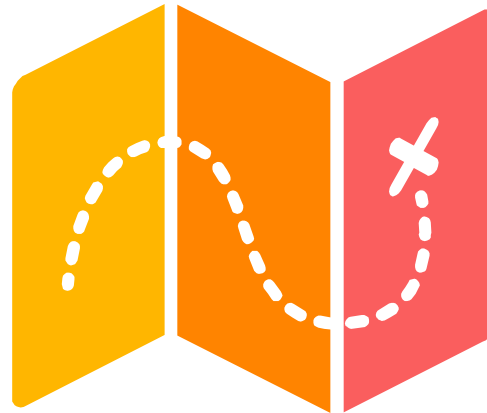
# Let's continue the journey...



- We found a **community**
- **But did everyone get this opportunity?** Who is missing?



# A journey into Open Science



## Step 1

More  
**reproducible**  
research

## Step 2

More  
**collaborative**  
research

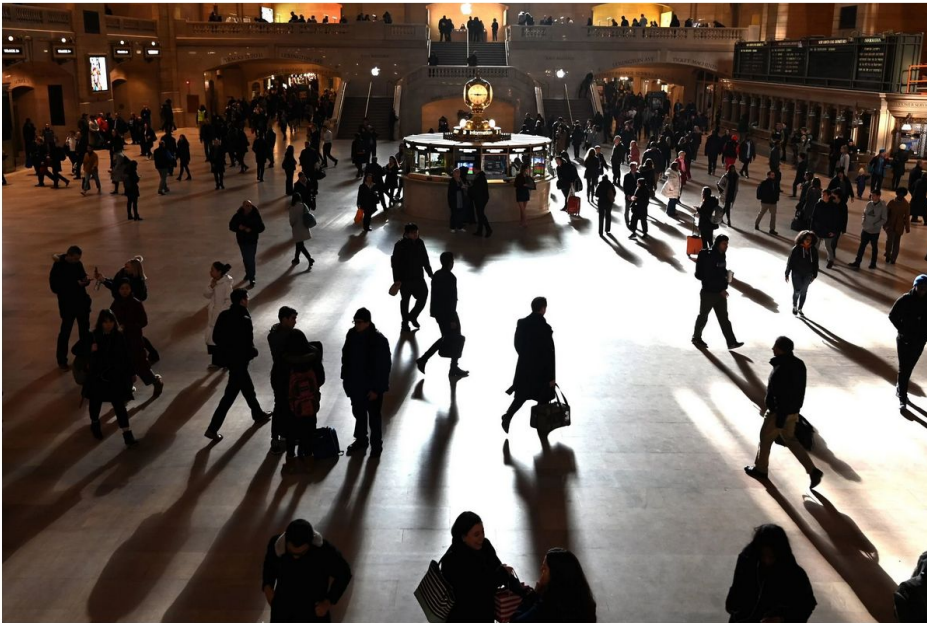
## Step 3

More  
**inclusive**  
research

The New York Times

## *Many Facial-Recognition Systems Are Biased, Says U.S. Study*

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images



Source: TED talk

[Joy Buolamwini, MIT 2017]

**Algorithmic bias**

# Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis

Agostina J. Larrazabal, Nicolás Nieto, Victoria Peterson   +1, and Enzo Ferrante   [Authors Info & Affiliations](#)

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved April 30, 2020 (received for review October 30, 2019)

May 26, 2020 | 117 (23) 12592-12594 | <https://doi.org/10.1073/pnas.1919012117>

 1,936  80



## Abstract

Artificial intelligence (AI) systems for computer-aided diagnosis and image-based screening are being adopted worldwide by medical institutions. In such a context,

[Larrazabal et al, PNAS 2020]

[Responsible AI Seminar](#)  
(DTU Denmark, online)

Talk by Enzo Ferrante

X-ray: Lung opacity detection



# The Diversity–Innovation Paradox in Science

Bas Hofstra , Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, , and Daniel A. McFarland  [Authors Info & Affiliations](#)

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved March 16, 2020 (received for review September 5, 2019)

April 14, 2020 | 117 (17) 9284–9291 | <https://doi.org/10.1073/pnas.1915378117>

 5,276  199



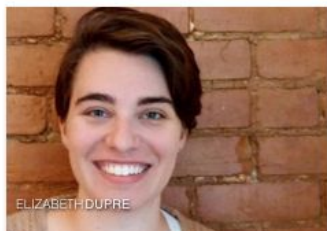
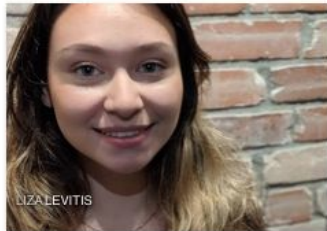
## Significance

By analyzing data from nearly all US PhD recipients and their dissertations across three decades, this paper finds demographically underrepresented students innovate at higher rates than majority students, but their novel contributions are discounted and less likely to earn them academic positions. The discounting of minorities' innovations may partly explain their underrepresentation in influential positions of academia.

[Hofstra et al, PNAS 2020]

# Designing online conferences for inclusivity

# OHBM Open Science SIG 2020



# OHBM Open Science Room



Cass  
Gould van Praag



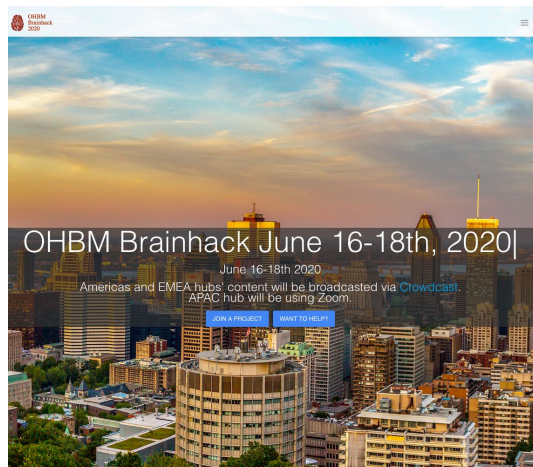
Stephan  
Heunis



Open  
Science  
Room



# OHBM Brainhack 2020



Rémi Gau



Liza Levitis

# Designing online conferences for inclusivity



## Unlocking access



### Schedule purposefully

3 time zone hubs and flexible personal schedule



### Build an accessible space

Geographical restrictions and accessibility



### Use appropriate fee system

Support hidden costs



### Adapt content locally

Translations, language groups



# Available for all at a reasonable time

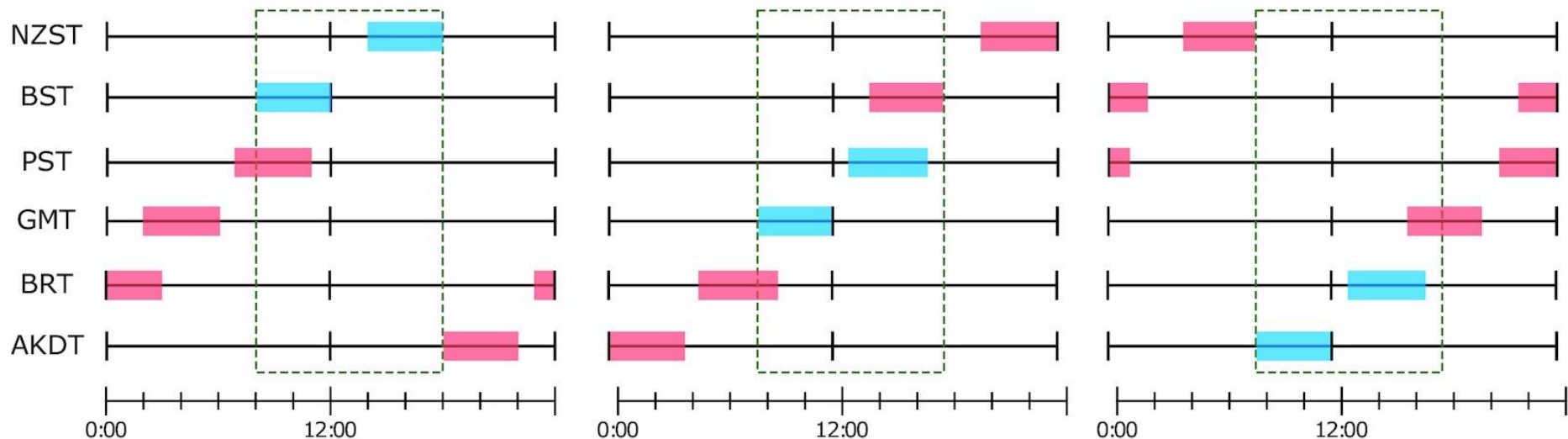
Hub 1: Asia, Pacific



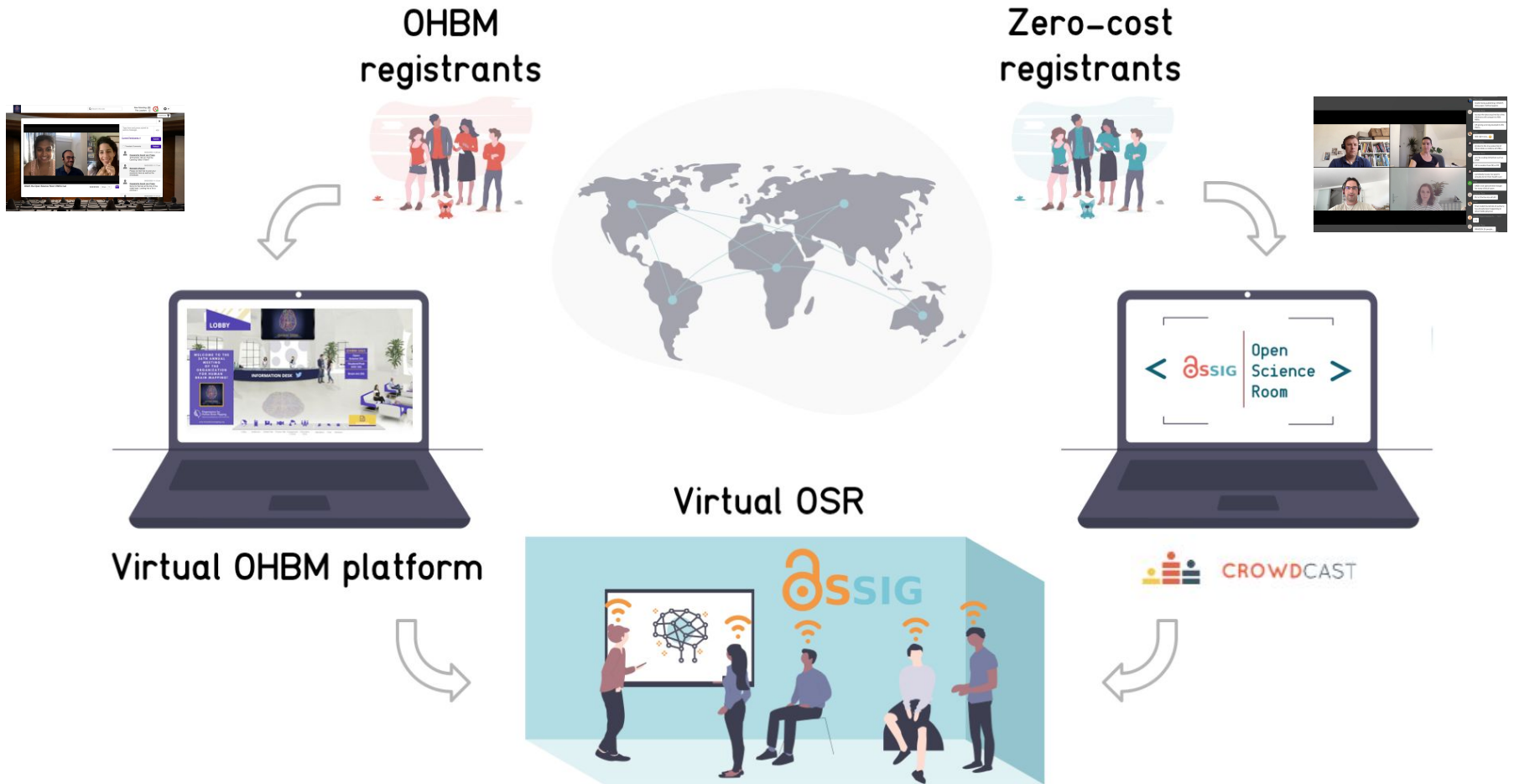
Hub 2: Africa, Europe, Middle East



Hub 3: Americas



# Multi-road Open Science Room



# Hackathon & OSR from China



**R-fMRI Journal Club** 举报 科技文化 · 科普 · 新知科普

1023 关注

RfMRILab 3262 0友邻 1.1kg

鱼吧 | 视频

福星挑战 0 > 超级粉丝团 >

登录领 x10 分享 福利签到

斗鱼 房间号: 993594 | 2020.06.23

任务大厅 鱼塘 礼物红包 盲充礼包 666 背包

快来抢! 今天免费领2w鱼丸+30办卡!

鱼丸 0 鱼翅 0 充值 背包



## OHBM骇客松活动公告

Original: OHBM OS-SIG OHBM国际人脑图谱学会 5/16

这里是OHBM开放科学特别兴趣小组 (OS-SIG), 很高兴能与您分享以下信息:

2020年OHBM 大脑骇客活动 (OHBM Brainhack, 由OS-SIG组织的骇客松/编程马拉松活动) 现已开始注册, 活动将于6月16日至18日在线上举行。快来注册吧! (<http://www.humanbrainmapping.org/HackathonReg>)

我们将提供50个价值100美元的微型基金, 以帮助骇客松的参加者们设置在线会议。通过上文链接提交申请。

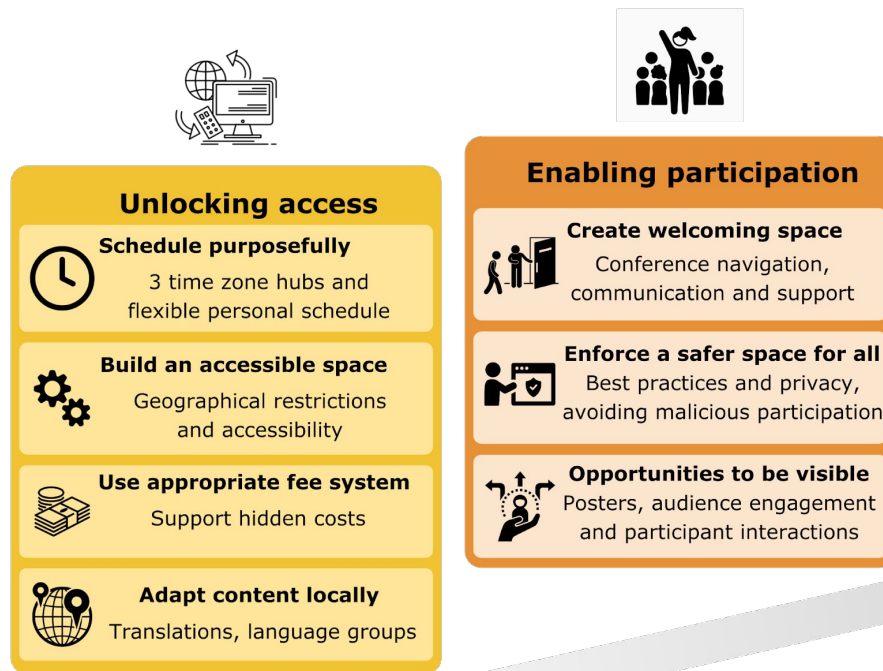


Scan to Follow

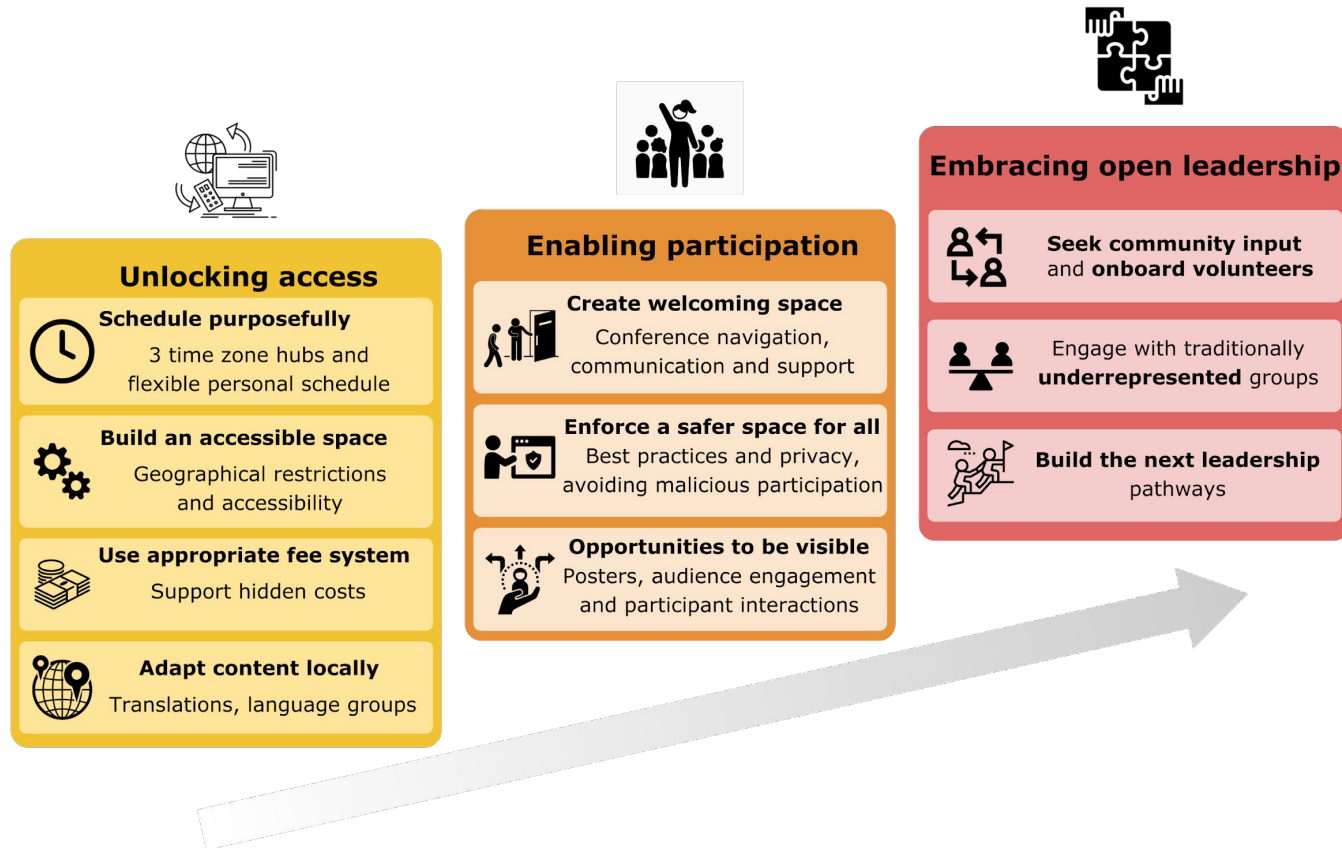
何为OHBM 大脑骇客活动?

观看短片了解更多信息!

# Designing online conferences for inclusivity



# Designing online conferences for inclusivity



# 100 volunteers! Thank you



**Alexandra Lautarescu**  
Centre for the Developing Brain, King's College London



**Amanda Robinson**  
The University of Sydney, Australia



**Frantisek Vasa**  
King's College London



**Irene Dupong**  
University of Paris



**Manoj Kumar**  
Princeton University



**Marta Topor**  
University of Surrey



**Sin Kim**  
Korea Advanced Institute of Science and Technology, Daejeon, Korea



**Soroosh Afyouni**  
University of Oxford, Oxfordshire, United Kingdom



**Angela Tam**  
National University of Singapore



**Ayan Sengupta**  
University of Cambridge, University of London



**Jakub Vohryzek**  
University of Oxford, Oxford, UK



**Jess Stevenson**  
CUBRC, Cardiff University, UK



**Matteo Frigo**  
Iarla Sophia Antipolis - Mediterranée, Nice, France



**Morgan McIntyre**  
Queensland Brain Institute



**Tibor Auer**  
School of Psychology, University of Surrey, Guildford, Surrey, England



**Tjii Grootswagers**  
Western Sydney University, Sydney, Australia



**Ben Fulcher**  
The University of Sydney, Australia



**Cooper Smout**  
Institute for Globally Distributed Open Research and Education



**Johanna Bayer**  
The University of Melbourne



**Jon Haltz Legarreta Gorroño**  
Universidad de Shobrenke, Sherbrooke QC Canada



**Rohit Goswami**  
Faculty of Physical Sciences, University of Iceland



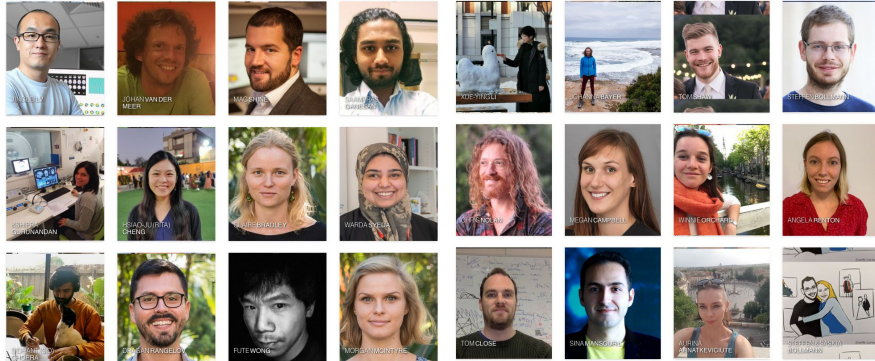
**Rotem Botvink-Nezer**  
Dartmouth College, Hanover, NH, USA



**Tom Johnstone**  
Swinscoe University of Technology, Melbourne, Australia



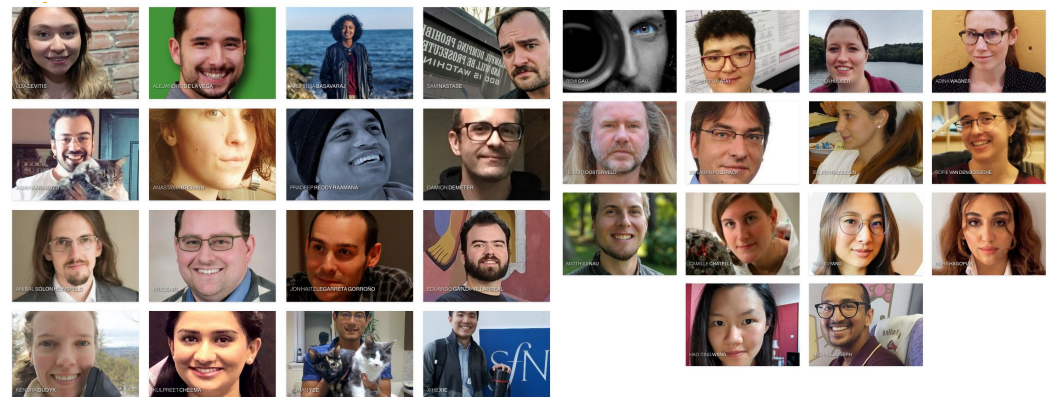
**Tom Kirk**  
Institute of Biomedical Engineering, University of Oxford



**Tulika Nandi**  
University of Oxford, Oxford, UK

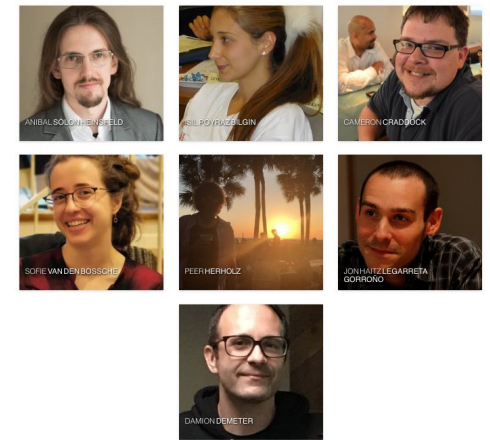


**Vittorio Iacovella**  
CIMAC - Center for Mind-Brain Sciences, The University of Trieste (Italy)



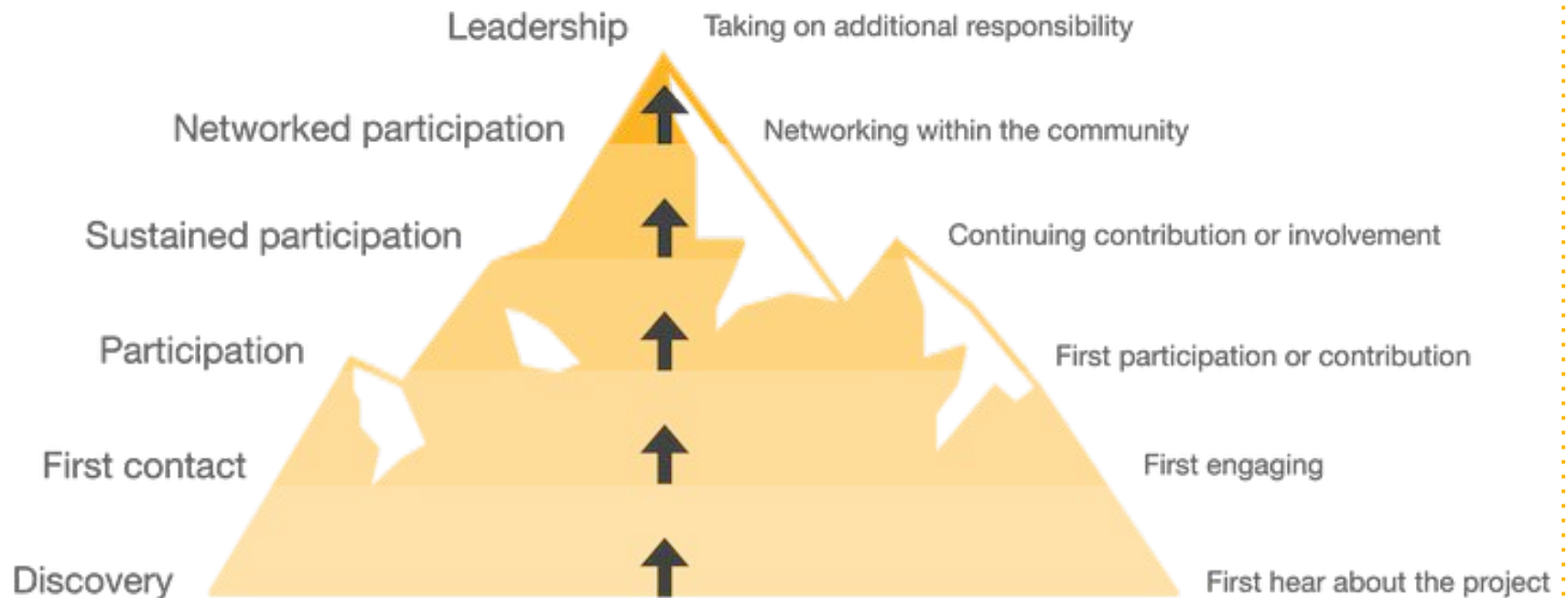
## Brainhack

The purpose of Brainhack is to bridge the data science and neuroscience research communities to advance the progress of brain science research. Brainhack is a unique conference that convenes researchers from across the globe and myriad disciplines to work together on innovative projects related to neuroscience. Year after year, global Brainhack events have brought together researchers to participate in open collaboration and regional Brainhack events keep the momentum going throughout the year.



# Going further!

## Mountain of engagement



[By WIN Open Research](#)

# Continuing on our journey...

## Online training & communities

### Open Life Science

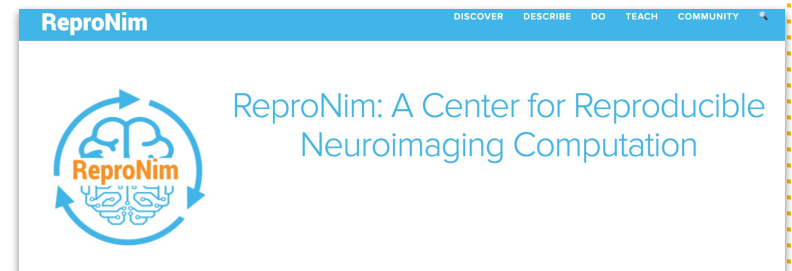


#### A mentoring & training program for Open Science ambassadors

The **Open Life Science (OLS)** program is for people interested in **applying open principles** in their work and **becoming Open Science ambassadors** in their communities.

<https://openlifesci.org/>

### Repronim



<https://www.repronim.org/>

### The Turing Way



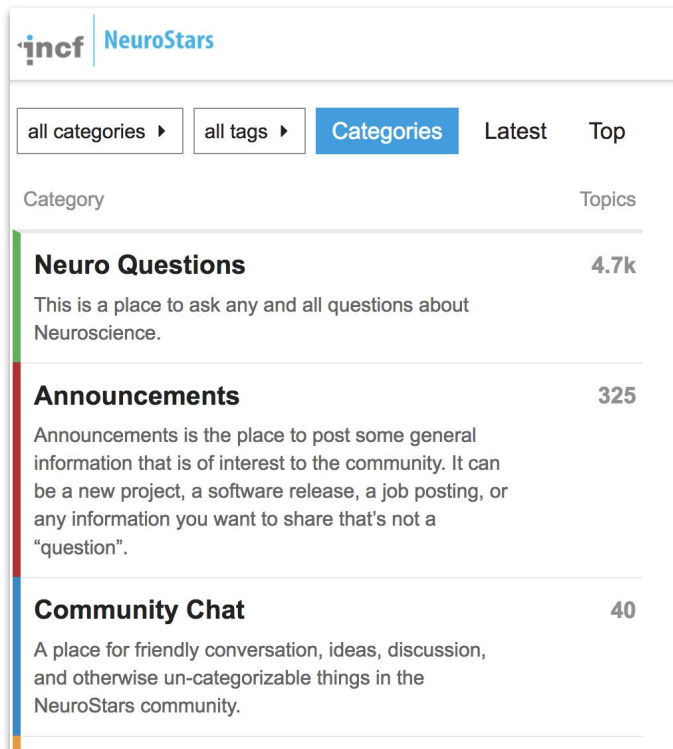
<https://the-turing-way.netlify.app>



# Continuing on our journey...

## Online discussions

### INCF Neurostars



The screenshot shows the INCF NeuroStars website interface. At the top, there is a navigation bar with the INCF logo and the text "NeuroStars". Below this, there are three dropdown menus: "all categories", "all tags", and "Categories" (which is currently selected). To the right of these menus are two buttons: "Latest" and "Top". Below the navigation bar, there is a table with two columns: "Category" and "Topics". The table lists three categories: "Neuro Questions" with 4.7k topics, "Announcements" with 325 topics, and "Community Chat" with 40 topics. Each category has a brief description below it.

Category	Topics
<b>Neuro Questions</b> This is a place to ask any and all questions about Neuroscience.	4.7k
<b>Announcements</b> Announcements is the place to post some general information that is of interest to the community. It can be a new project, a software release, a job posting, or any information you want to share that's not a "question".	325
<b>Community Chat</b> A place for friendly conversation, ideas, discussion, and otherwise un-categorizable things in the NeuroStars community.	40

<https://neurostars.org/>

### Brainhack mattermost

Text-based discussions

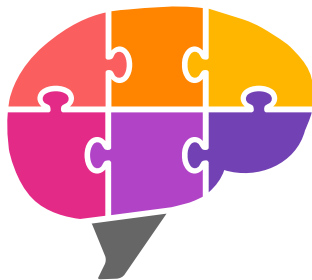


<https://brainhack.org/>

# **Open science** A journey from sharing research artefacts to collaborative research

Camille Maumet

# Thank you!



**We are hiring**

<https://team.inria.fr/empenn/job-offers/>