Open Access: what, how and why?
Paola Masuzzo

@pcmasuzzo       paola.masuzzo@ugent.be
Computational omics and systems biology group, VIB
Ghent University, Ghent, Belgium

Data Sharing in Neuroimaging
Daniele Marinazzo

@dan_marinazzo       daniele.marinazzo@ugent.be
Department of data analysis, faculty of psychology and pedagogical sciences
Ghent University, Ghent, Belgium

OpenAIRE Workshop, 31 May 2016
Open Access explained!
https://www.youtube.com/watch?v=L5rVH1KGBCY
What?

How?

Why?
What?

How?

Why?
What is Open Access?

• Open Access (OA) refers to the removal of major obstacles to accessing, sharing and re-using the outputs of scholarly research.

• The research process is then facilitated by ensuring rapid and widespread access to research findings such that all communities have the opportunity to build upon them.

• Read full definition @Budapest OA Initiative: http://www.budapestopenaccessinitiative.org/read
There are currently over 700 OA policies and mandates recorded worldwide

Policies Adopted by Quarter

- Research organisation
- Funder
- Sub-unit of research organisation
- Funder and research organisation
- Multiple research organisations

http://roarmap.eprints.org
What?

How?

Why?
There are basically two routes to OA

- **The Gold route**: freely accessible research articles at the point of publication (sometimes accompanied by APCs).

- **The Green route**: author self-archiving; a version of the peer-reviewed article is posted online to a repository or website.
What?

How?

Why?
There are several reasons to go for OA:

- Researchers in developing countries can see your work
- Taxpayers get value for money
- More exposure for your work
- Compliant with grant rules
- The public can access your findings
- Practitioners can apply your findings
- Higher citation rates
- Your research can influence policy
There are several reasons to go for OA!

- More exposure for your work
- Researchers in developing countries can see your work
- Taxpayers get value for money
- Compliant with grant rules
- The public can access your findings
- Practitioners can apply your findings
- Higher citation rates
- Your research can influence policy
The academic case for OA

• OA enables higher documented impact of scholarly articles through availability and re-use

The academic case for OA

- OA enables higher documented impact of scholarly articles through availability and re-use

- OA non-restrictively allows researchers to use automated tools to mine the scholarly literature (TDM)
And even more reasons...!
If you want to know more about OA (and you should!)

• The academic, economic and societal impacts of Open Access: an evidence-based review (http://f1000research.com/articles/5-632/v1)
• Budapest OA Initiative: http://www.budapestopenaccessinitiative.org
• FOSTER EU OA Resources: https://www.fosteropenscience.eu/foster-taxonomy/open-access
• SPARC OA: http://sparcopen.org/open-access
• OpenCon 2016: http://www.opencon2016.org
Open Access in the Open Science taxonomy

https://www.fosteropenscience.eu/taxonomy/term/5
The pillars of open science

Gorgolewski and Poldrack, BiorXiv 2016
Chris Gorgolewski

Stanford Center For Reproducible Neuroscience

Helping neuroscientists move from data to discovery

A practical guide for improving transparency and reproducibility in neuroimaging research

Krzysztof J. Gorgolewski, Russell Poltrack

do: http://dx.doi.org/10.1101/039354

biRxiv

http://www.slideshare.net/chrisfilo1/share-and-reuse-how-data-sharing-can-take-your-research-to-the-next-level
An overview of the main public fMRI datasets
Current Sample Characteristics (As of July 2015)

The NKI-RS is designed to be a representative, heterogeneous community sample. Exclusion criteria are minimal, and subjects vary widely in many respects. The following figures provide basic demographic information for the NKI-RS as of the 689th subject. Of these 689 participants, 657 have imaging data available.

Age

[Histogram showing age distribution]

[Chord diagram showing gender distribution]

Table Of Contents
- Studies
- Recruitment
- Participant Schedule
- MRI Protocol
- Assessments
  - Sample Characteristics
    - Age
    - Sex
    - DSM-IV Diagnoses
    - Data Sharing
    - NKI-RS Life Releases
    - NKI-RS Team

Quick search

An open neuroscience project brought to you by.
Human Connectome Project

Project Goals I

- Study a large population:
  - 1,200 healthy adults
  - 300 twin pairs and their non-twin siblings
- Cutting-edge neuroimaging methods
  - 3T Skyra MRI, customized gradient (UMinn -> Wash U)
  - 7T MRI (UMinn, 200 subjects); perhaps also 10.5T
  - dMRI/tractography; R-fMRI; Task-fMRI
  - MEG/EEG (100 subjects)
- Extensive behavioral testing
- Blood samples for genotyping
<table>
<thead>
<tr>
<th>Query terms</th>
<th>Number of images &amp; distinct subjects returned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>T1 SCZ + CON</td>
<td>3178</td>
</tr>
<tr>
<td></td>
<td>737</td>
</tr>
<tr>
<td>3T T1 20–60 Yr</td>
<td>1348</td>
</tr>
<tr>
<td></td>
<td>345</td>
</tr>
<tr>
<td>T1 healthy controls</td>
<td>1846</td>
</tr>
<tr>
<td></td>
<td>439</td>
</tr>
<tr>
<td>Sternberg SCZ 3 T &lt;= 30 Yr</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>18</td>
</tr>
<tr>
<td>Resting BP + CON</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>101</td>
</tr>
<tr>
<td>Sensory motor + T1, SCZ + CON, 30–50 Yr</td>
<td>1310</td>
</tr>
<tr>
<td>(multimodal)</td>
<td>119</td>
</tr>
<tr>
<td>All MRI</td>
<td>21,309</td>
</tr>
<tr>
<td></td>
<td>1029</td>
</tr>
</tbody>
</table>
ElectroEncephalography (EEG) data

IEEG.org is a collaborative initiative funded by the National Institutes of Neurological Disorders and Stroke. This initiative seeks to advance research towards the understanding of epilepsy by providing a platform for sharing data, tools and expertise between researchers. The portal includes a large database of scientific data and tools to analyze these datasets.

You can also:
- Sign up
- Download Toolkit
- Read Documentation

IEEG.ORG is a collaborative initiative funded by the National Institutes of Neurological Disorders and Stroke. This initiative seeks to advance research towards the understanding of epilepsy by providing a platform for sharing data, tools and expertise between researchers. The portal includes a large database of scientific data and tools to analyze these datasets.

Temple University EEG Corpus

Downloads: To request access to the TUH EEG Corpus, please fill out this form. You will receive an automatically-generated username and password via email.

- TUH EEG Epilepsy Corpus (v0.0.0): This is a subset of the TUH EEG Corpus that contains 100 subjects with and without epilepsy. We plan to expand this to 1,000 patients of each type. To request access to this data, please fill out the attached form.
- TUH EEG Corpus (v1.0.0): This will be the final official release of all data collected from 2002-2014. Contact us for further details.
- TUH EEG Corpus (v0.0.0): A beta release used to collect feedback from the community.
- TUH EEG Corpus (v0.2): This is our first public release of the TUH EEG Corpus. This is a beta release intended to allow users to give us feedback on the data. There are 247 sessions, 615 EDF files, and over 150 hours of EEG data. The uncompressed data occupies about 8.3G of disk space.

The TUH EEG Corpus is freely available. The only reason we require registration is that we need to track who downloads the data. We also want to be able to inform you of any updates to the releases.

Once you have obtained the username and password, you can selectively download portions of the corpus using your browser. Due to the size of the data, the best way to transmit this data is via a hard disk. Email us for details.

Useful Links

- Recent news
  - PhysioNet/GinC Challenge 2016

Looking for data?
- Physionet Google Search: Use the keyword search at the top right of every page.
- Physionet Database Index: All of our databases sorted by category (EEG, ECG, etc.)
- Physionet Record Search: Record search with key input information. Advanced user guide.
- MINIC-III: A massive healthcare dataset collected from over 40000 critical care patients.

Looking for software?
- PhysioToolkit Software Index: All the useful software contributed to our website.
- WFDB Software Package: Our own large collection of software for signal reading, writing, processing, and automated analysis.
- WFDB Matlab Toolbox: A Matlab implementation of the WFDB
NeuroImage Special Issue, January 2016
Data Sharing Saves Money

Gorgolewski et al. OHBM 2015
Data Sharing Concerns

- Fear of being scooped
- Fear that someone finds a mistake in your analysis
- Misconceptions about data ownership
Data Sharing and Quality of Results

http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0026828
Even more credits for data sharing: data papers

- NeuroInformatics
- GigaScience
- Scientific Data
- F1000 Research
- Data in Brief
- Journal of Open Psychology Data

Gorgolewski, Milham and Margulies 2013
Repositories and formats

- OpenfMRI.org
- FCP/INDI
- COINS
- OSF.io
- figshare
- DataVerse
- DataDryad

Quality assessment

mriqc: image quality metrics for quality assessment of MRI

This pipeline is developed by the Pollock Lab at Stanford University for use at the Center for Reproducible Neuroscience (CRN), as well as for open-source software distribution.

Preprocessed Connectomes Project

Data Format, Guidelines, Metadata

http://bids.neuroimaging.io/
BrainBox, OpenScience Prize Finalist

Collaboratively visualize and segment any brain image available online
Conclusions: why share data

• It’s an ethical thing to do
• The journal (PLOS) or the funder (NIH and soon EU) requires it
• Increases exposure and impact of your research (and so your chances of getting the next grant)
• It’s related to higher citation rate (Piwowar et al. 2007, 2013)
• Shows that you care
Thank you!