Workshop : Open and Big Data for Life Imaging

Christian Barillot

Michel Dojat

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- Scientific

« Power failure: why small sample size undermines the reliability of neuroscience »
Button et al. Nat. Neurosc 2013
« Why most published research findings are false" **Ioannidis Plos Med 2005**
"Data from preclinical animal studies appear to be associated with even greater bias »
Ioannidis ScienceTranslationalMed 2012,
« Why small low-powered studies are worse than large high-powered studies and how to protect against trivial findings in research **» Ingre Neuroimage 2012**

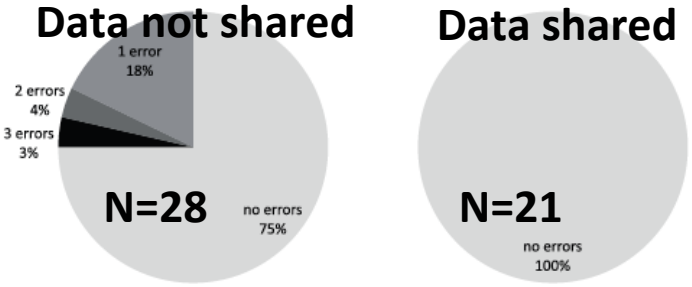
- **Underpowered is endemic** (false positive, false negative)

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- Scientific
  - **Data analysis and reporting are often selective and biased**

Errors in reporting statistical results



Wicherts et al. PLOS One 2011

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- Scientific
  - **Rarely replicated so false solutions persist**
  - **New questions, increasing the speed of discovery**

Large cohort, Re-analysis, Replication, MetaAnalysis

« collective mind » Fox et al. Ann Rev Neurosci 2014

« crowd science » Franzoni & Suerman Res Policy 2014

Three Good Reasons for refraining Sharing Data and Tools ….

- Economic

Data Acquisition: e.g. neuroimaging 300€/H 30 subjects x 2 groups 18 K€ …
Many efforts for image processing tools development and validation
dont replicate existing data and tools ….

Reducing cost doing science, maximize investment

- Ethic

Healthy subject and patients give their consent for contributing to knowledge advancement.
For preclinical studies, ethical obligation to ensure that the animals are not wasted.

Maximize their participation

Three Good Reasons for refraining Sharing Data and Tools ….

- Scientific

How can I have a return on investment ? What direct benefit I will get ?
Who benefits from, when, and how?
Degree of control, fear of adverse use or flawed interpretation or, exposition of some errors
« Too much effort », appropriateness of others' data

Wallis et al.: » goods are bartered between trusted colleagues rather than treated as commodities » PLoS ONE 2013

- Economic
- Ethic & Regulation

Three Good Reasons for refraining Sharing Data and Tools ….

- Scientific

- Economic

The cost of sharing. Data should be kept and curated.
Who will make the investments?
Which Bussiness model?
Van Horn & Gazzaniga « Have a plan for what happens AFTER your grant funding runs out » Neuroimage 2013

- Ethic

Three Good Reasons for refraining Sharing Data and Tools ….

- Scientific
- Economic

- Ethic & Regulation

Who is the data owner ? use data for new non initially expected applications
Have images some specificities ? (see UKBioBank)
Research subjects vs patients
Securisation: potential failure in the authentification system =>
Balance risk & security level &…. Bypassing the rules

Researchers, journals and funding agencies must work together to identify and implement technical solutions that allow the most effective data sharing without greatly increasing the burden on researchers.

Poldrack and Gorgolewski Nat Neuro 2014

Researchers, journals and funding agencies must work together to identify and implement technical solutions that allow the most effective data sharing without greatly increasing the burden on researchers.

Poldrack and Gorgolewski Nat Neuro 2014

The slow adoption of tools and services such as data repositories are indications that **technology alone cannot change scientists' practices; social and cultural factors must also encourage datasharing…. Much sharing is private rather than public**.

Wallis et al. PLOS One 2013

We conclude that research data cannot be regarded as knowledge commons, but **research policies that better incentivise data sharing** are needed to improve the quality of research results and foster scientific progress.

Fecher et al. PLOS One 2015

We conclude that research data cannot be regarded as knowledge commons, but **research policies that better incentivise data sharing** are needed to improve the quality of research results and foster scientific progress.

Impediments regarding sharing supplementary data could include clear and elaborate reasons to opt out. In order to remove risk aversion and ambiguity, an understandable and clear legal basis regarding **the rights of use is needed to inform researchers on what they can and cannot do with data they collected.**

Fecher et al. PLOS One 2015

An Open Discussion

- Context of the workshop

Introduction to FLI-IAM project

Christian Barillot, Michel Dojat, Michael Kain (FLI-IAM, FR)

- Legal and ethics questions

What can be shared? To whom data own? Status of data, metadata and derived data?
What are the requirements for sharing?

Paul Oliver Gibert (Digital Ethics, FR)

- Institutional aspects

What are the institutional actions to promote data sharing?
Which Economical models?
What credit for data and tools sharing ?

Christine Balagué (CCNum, FR)

An Open Discussion

- Technical aspects

What are the existing solutions? What are the main technical challenges?

P. Mouillard (Vigisys, FR)

D. Kennedy (INCF, US)

- Pros and cons of working architectures

positive and negative aspects of sharing to large scale in practice? user acceptance Efforts for sharing.

Wiro Niessen (Rotterdam cohorte, NL)

Gabriel Robert + Dimitri Papadopoulos (Imagen Cohort, FR, UK)

ROUND TABLE

Round Table

- Prior Issues
  - Standards for sharing data
    - Do we need standard for data model (e.g. ontologies) ?
    - Standard for raw/derived data format ?
    - Do we need standard for interoperability (on concepts, on communication) ?
    - wait for standards or go forward ?
  - Quality control in Data Sharing
    - Ratio cost / added value
    - Automatic vs Human Quality Control

Round Table

- Infrastructure Issues:
  - Data bases:
    - Big Data Centers vs Distributed Storage vs Peer to Peer storage (bioTorrent) ?
  - High Performance Computing (HPC) and Big Data for *in vivo* imaging
    - Is HPC infrastructure generic or specific to usages?
    - Cloud computing: Clusters vs Grids vs Crowd Computing ?
  - Image Processing Code Sharing ?
    - Is it different than Data Sharing?
    - Is Crowd Science meaningful for in vivo imaging?

Round Table

- Socio-economic issues:
  - Data Sharing credit ?
    - Co-authorship *vs* citation *vs* payment ?
    - Data rights
  - Economic model
    - Cost for using shared data ?
    - Who support the cost ?
    - Can data sharing become a business?
  - Is regulation / ethics a real limit?
  - How funding agencies/institutions can help to promote data sharing?

Round Table

- Finally, Bottlenecks are:
  - Socio-economical?

  Or
  - Technological?