Edinburgh Imaging
www.ed.ac.uk/edinburgh-imaging

Sharing brain MRI data under GDPR

Cyril Pernet, PhD
Centre for Clinical Brain Sciences, University of Edinburgh

GliMR COST action Working Group2:
Open Brain Consent: 27th November 2019

@CyrilRPernet
cyril.pernet@ed.ac.uk

Overview action points

- GDPR core

-> basic info in any PIS and consents

- Brains are personal data

-> what type of sharing (control via DUA)

- Sharing on the internet

-> users vs patients rights

GDRP core

GDPR core

- General Data Protection Regulation against the misuse of a people's data
- **Personal data** must be processed lawfully, transparently, and for a specific purpose. *Pseudonymised personal data may also be subject to GDPR rules, depending on how easy or hard it is to identify whose data it is.*
- A **data controller** is responsible for setting out how and why data is collected – typically the PI; the **data processor** is the one collecting the data (PI, student, post-doc, research nurse, etc).
- **Consent** must be an active, affirmative action : already the case in research – except minors and incapacitated participants

Legal basis

- As researchers in universities and hospitals, ‘public task’ is the typical legal basis
- show that the processing is necessary to achieve it (you must collect MRI brain data for your research)
- balance it against the individual’s interests, rights and freedoms (can sharing do harm?)

https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/?template=pdf&patch=248#link19

GDPR not for us?

- Articles 5-30 and 35-50 of the GDPR state that this does not apply to journalistic, academic, or artistic works. True for non-personal and fully anonymized data.

- **Issues**:

(1) MRI brains are personal data.

(2) Share publicly: within vs outside EU? how to remove public data?

Brains are personal data

Brain MRI data are personal

- Always remove dicom ID related tags with/without nifti conversion = de-identification
- Often keep a key to reverse the ID as there are expectations that MRI could reveal future problem
- Often apply defacing algorithms
- Pseudo-anonymization is a process, it doesn't change the status of the data – after making your defaced .nii(.gz) files with a new sub-XX ID, if we can get back to the subjects this is personal data (=always the case if you keep a key/record).

https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/

COMPUTER LAW & SECURITY REVIEW 34 (2018) 222–233

Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK

Miranda Mourby [a,*], *Elaine Mackey* [b], *Mark Elliot* [b], *Heather Gowans* [a], *Susan E. Wallace* [a,c], *Jessica Bell* [a], *Hannah Smith* [a], *Stergios Aidinlis* [a], *Jane Kaye* [a]

Under GDRP there are only two options: data are either personal or anonymous

Pseudo-anonymization (a process) reduces risk of direct identification

→ GDPR definition makes it that pseudo-anonymized data can be personal data

→ De-identification must be 'irreversible' to be considered 'anonymisation',

→Rare clinical cases can be singled out = personal data

Brains are personal data

"the data controllers should focus on the concrete means that would be necessary to reverse the anonymisation technique, notably regarding the cost and the know-how needed to implement those means and the assessment of their likelihood and severity"

- After removing ID and defacing:

(i) is it still possible to single out an individual?

(ii) is it still possible to link records relating to an individual?

(iii) can information be inferred concerning an individual?

https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Singling-out and Re-identification

nature COMMUNICATIONS

https://doi.org/10.1038/s41467-019-10933-3 OPEN

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher [1,2,3], Julien M. Hendrickx[1] & Yves-Alexandre de Montjoye[2,3]

DE-ANONYMIZATION ATTACKS ON NEUROIMAGING DATASETS

arXiv:1908.03260v1 [cs.CR] 8 Aug 2019

Vikram Ravindra
Department of Computer Science
Purdue University
West Lafayette, IN 47906
ravindrv@purdue.edu

Ananth Grama
Department of Computer Science
Purdue University
West Lafayette, IN 47906
ayg@cs.purdue.edu

Singling-out and Re-identification

	concrete means	cost and the know-how	likelihood
Rocher et al. (2019)	Access to freely available datasets like post-codes, census, etc	Gaussian-copula modelling from (incomplete) metadata - not attempt to re-identify here, estimate likelihood of success given the data Julia and Python (free) Good stats, data science knowledge	> 95%
Ravindra & Gramma (2019)	Access to multiple datasets and possibly original data	Imaging and connectomic knowledge, HPC - infer subject across sets, task performance, possibly ID if access to some original data	~ 90%

Refacing to linkage

REFACING: RECONSTRUCTING ANONYMIZED FACIAL FEATURES USING GANS

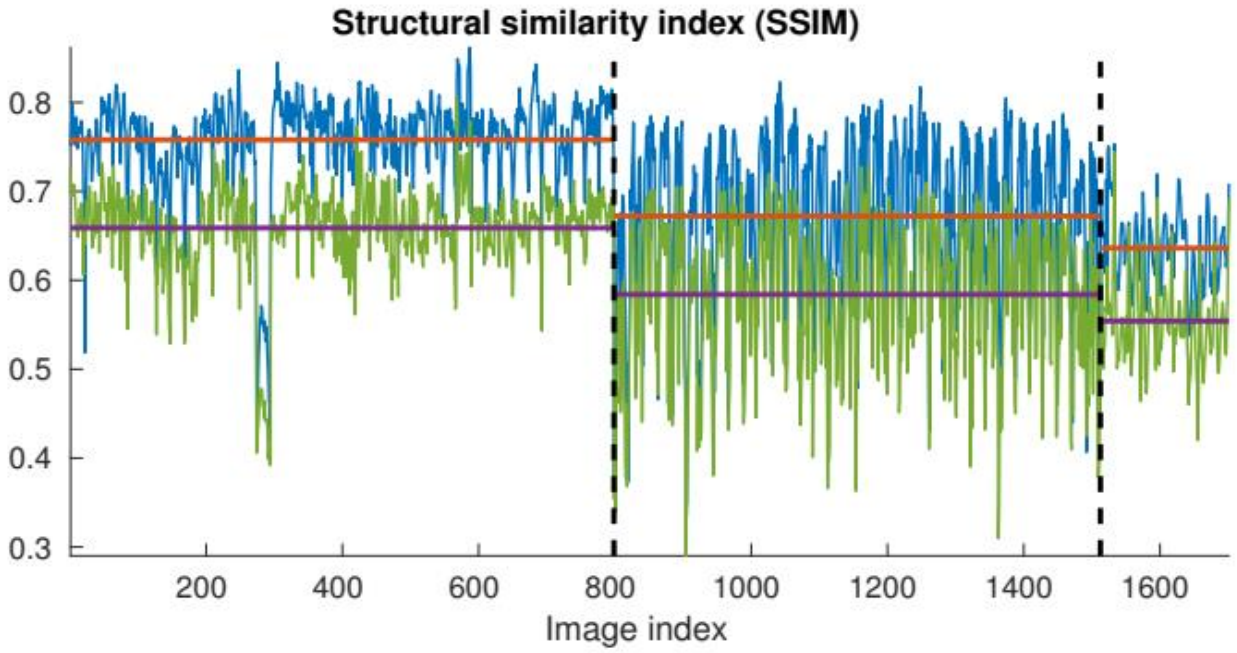
David Abramian[a,b] *Anders Eklund*[a,b,c]

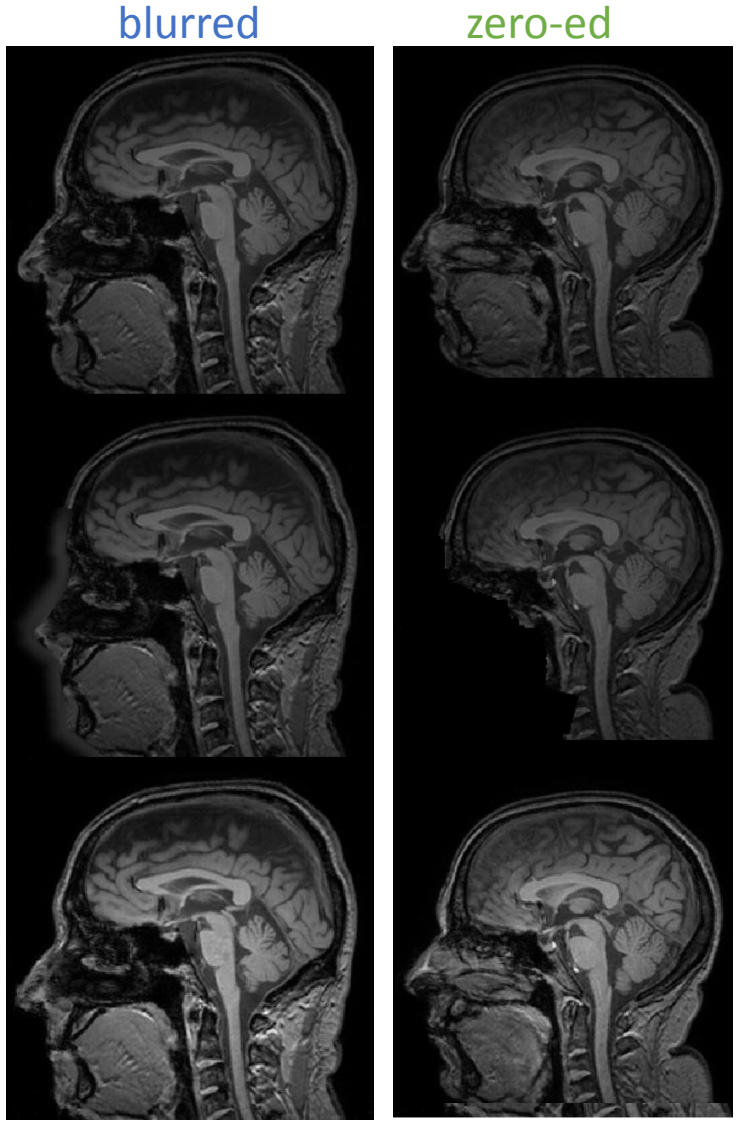
Refacing using Generative Adversarial Networks (GAN) based on Convolution Neural Networks: means, cost, know-how much higher but for how long?

Refacing to linkage

→IXI dataset from multiple sites

→only 2D (i.e. from sagittal slice) but it's a matter of computation force to have a 3D version





Sharing openly

A funder requirement !

Sharing openly

“a transfer shall take place only where it is necessary or legally required on important public interest grounds”

→GDPR restricts transfers of personal data outside the EEA

→Data Sharing Agreement seems the solution – better than open licencing which allows redistribution

https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-5-identifiability-anonymisation-and-pseudonymisation/

https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_2_2018_derogations_en.pdf

Sharing openly

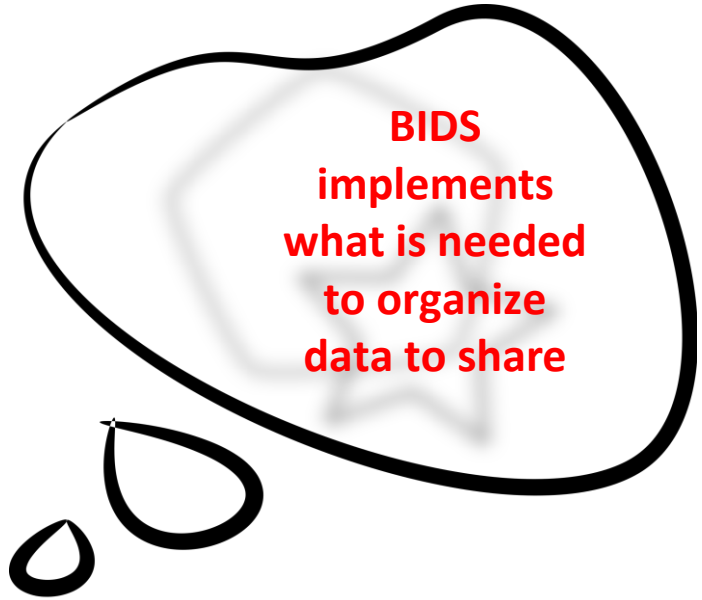
OXFORD

Briefings in Bioinformatics, 00(00), 2019, 1–10

doi: 10.1093/bib/bbz044
Advance Access Publication Date: 29 June 2019
Review article

Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives

Charles Vesteghem, Rasmus Froberg Brøndum, Mads Sønderkær, Mia Sommer, Alexander Schmitz, Julie Støve Bødker, Karen Dybkær, Tarec Christoffer El-Galaly and Martin Bøgsted



Idea

Sharing openly

Central or Local

- Central platform (e.g. ENBIT) gives more visibility, allows easier meta-search, and meta-analyses
- Many local repositories easier to manage but needs to be findable (DOI and Google datasearch can help)

Fully open or gate keeper

- Under GDRP the gate keeping can be a simple data user agreement, necessary as users outside EU have limited rights for personal data.
- Gate keeping for clinical data is necessary, to keep track of users and often restrict on commercial usage

GDPR compliant open brain consent

https://open-brain-consent.readthedocs.io/en/stable/

What needs to be in information sheet?

- Who is data controller and the data processor
- The legal basis is that you rely on to make the research lawful
- The participants rights to access and withdraw data
- Complaints contact and the contact details of this organisation's DPO

GDPR requires you collect and process personal data for a purpose – that's your study
If you share under that study, in principle, any reprocessing should be on the the same goal (makes your study reproducible).

To extend re-usage, a separate PIS and consent related to sharing seems warranted – sharing for research only, commercial or not, for anything, it's up to you but make sure you have the means to achieve what you state.

What needs to be in information sheet?

- How data will be pseudo-anonymized and what to expect from that
- What infrastructure to store securely personal data
- Where will you share the data (what implication for removing data)
- How access will be granted

- must have secured servers for raw data
- share only relevant metadata (yes limits usage)
- zeroing faces from structural images
- minimal gate keeping access