The hidden cost of open science: what's our footprint?

Dr Charlotte Rae
c.rae@sussex.ac.uk
@NeuroRae

UNIVERSITY OF SUSSEX

Open science & sustainability

We all agree open science is a good thing…

…but there are hidden costs

- Time & effort
- Disproportionately falls on ECRs
- **Sustainability** (environmental impact)

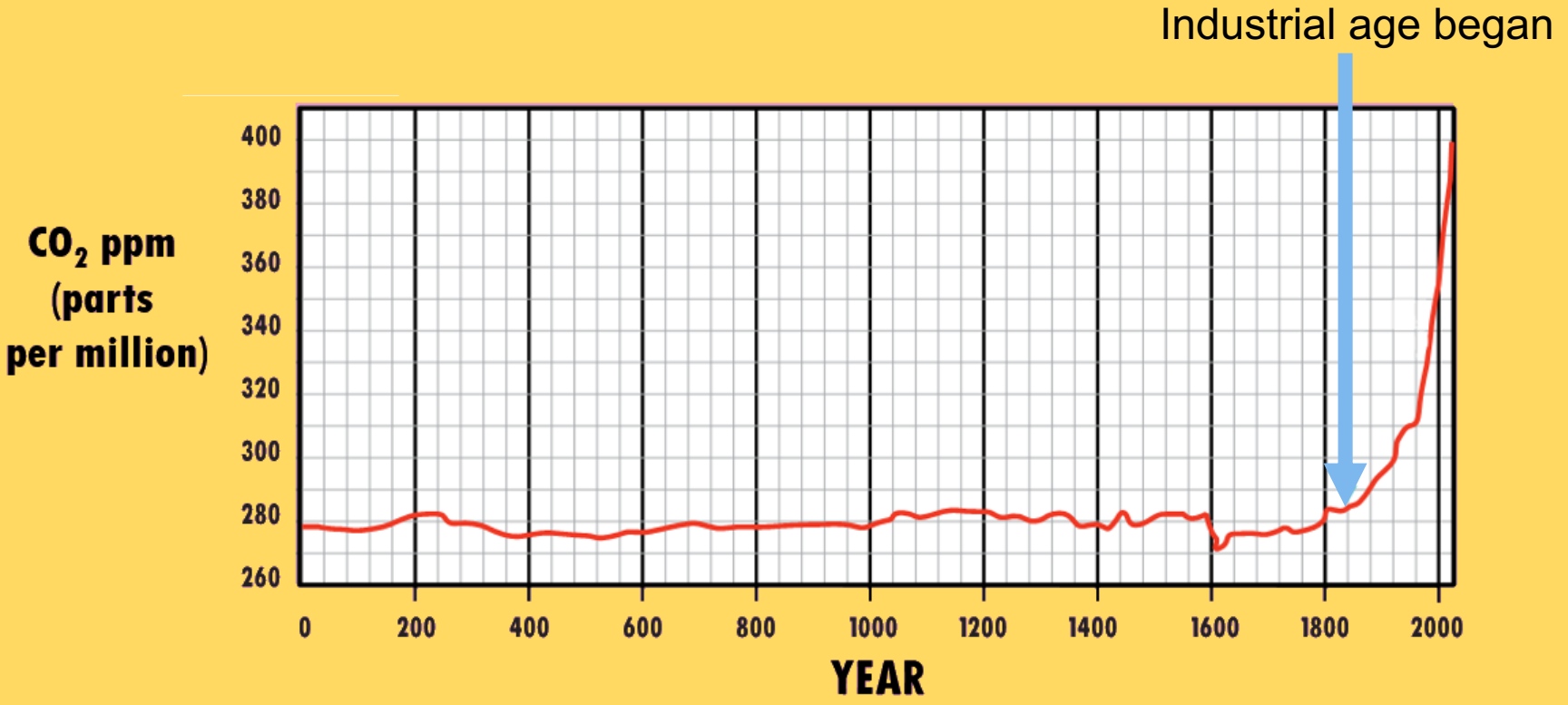
Focus today is on MRI & M/EEG in particular

- Large file sizes; liquid helium

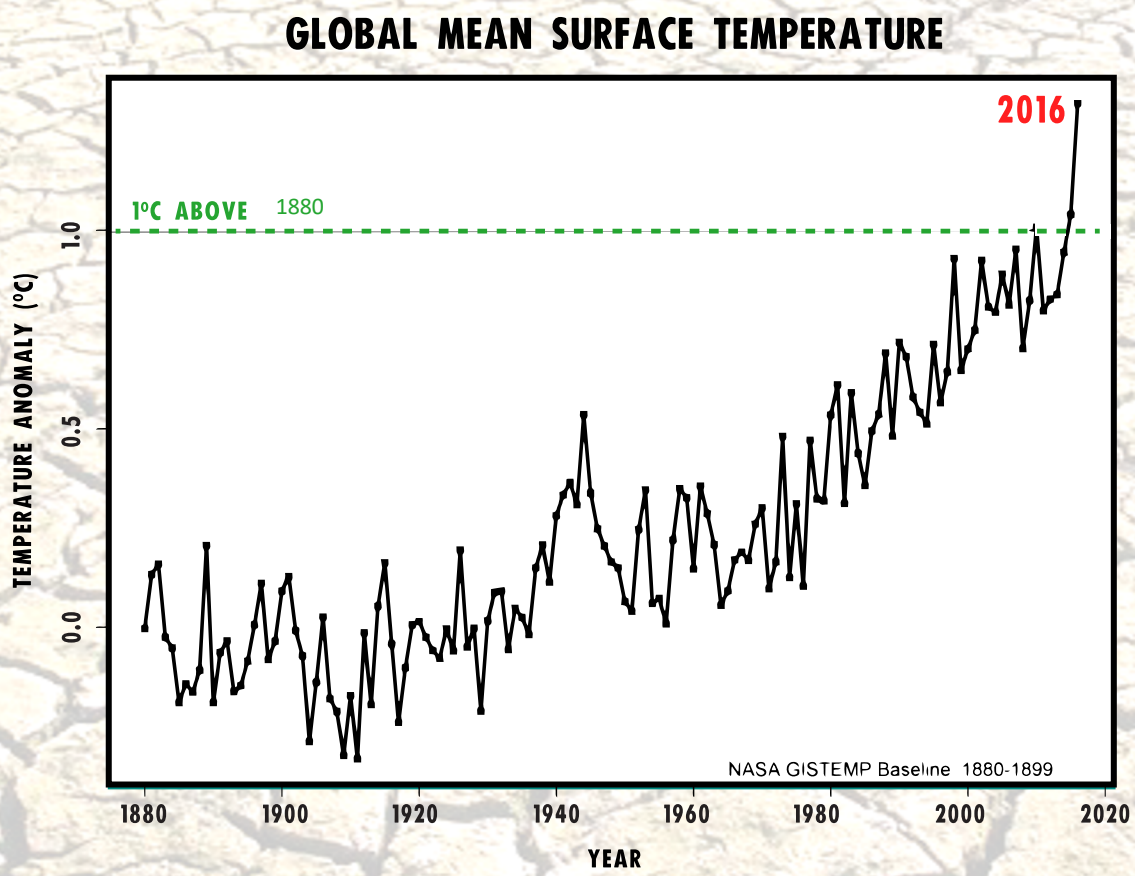
Repositories – who gets green flag, who gets lump of coal

Reduce, reuse, recycle (‘slow science’)

GREENHOUSE GAS EMISSIONS ARE SKYROCKETING



GLOBAL TEMPERATURES ARE SKYROCKETING



THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC)

30th Anniversary
ipcc
INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE
1988 – 2018
WMO UNEP

NEWS · 08 OCTOBER 2018

IPCC says limiting global warming to 1.5 °C will require drastic action

Humanity has a limited window in which it can hope to avoid the worst effects of climate change, according to climate report.

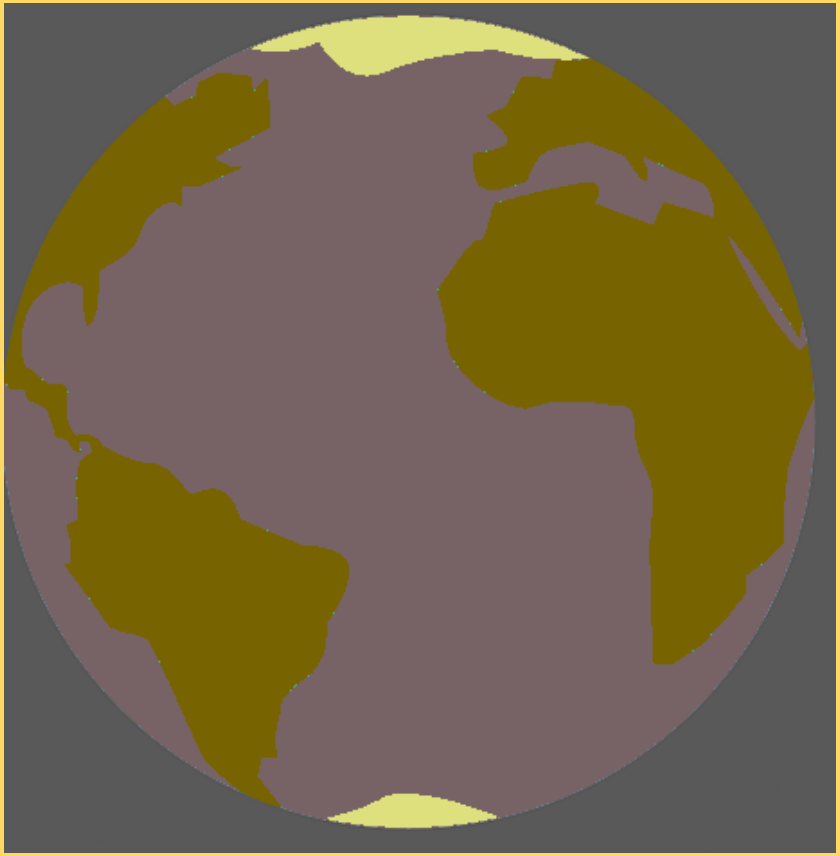
HEATING IMPACTS ALREADY BEING FELT GLOBALLY

CONSEQUENCES FOR "BUSINESS AS USUAL"

- Accelerating sea level rise due to melting ice and thermal expansion; coastal flooding & storm surges
- Stronger atmospheric weather systems with strong storms, droughts, wildfires, desertification & crop failure
- Deaths due to heat stress
- Famine, water shortages, displacement of people from their homes
- War

WE ARE ALSO DESTROYING THE PLANET'S

ECOLOGY



HIGH-INTENSITY AGRICULTURE HAS REPLACED WILDERNESS

DE-FORESTATION IS REDUCING CARBON DRAWDOWN, BIODIVERSITY & MOISTURE RETENTION

AND BY DESTROYING OUR FOOD WEB, WE ARE SAWING OFF THE BRANCH WE ARE STANDING ON

The climate catastrophe & ecological emergency

- Happening now
- We must act urgently
- The causes extend across all domains of human activity:
    - scientific research is one of these

Sustainability impacts of open science

Servers

- Building: construction
- Manufacture: mining of finite resources
- Usage: energy (including cooling)

MRI and MEG

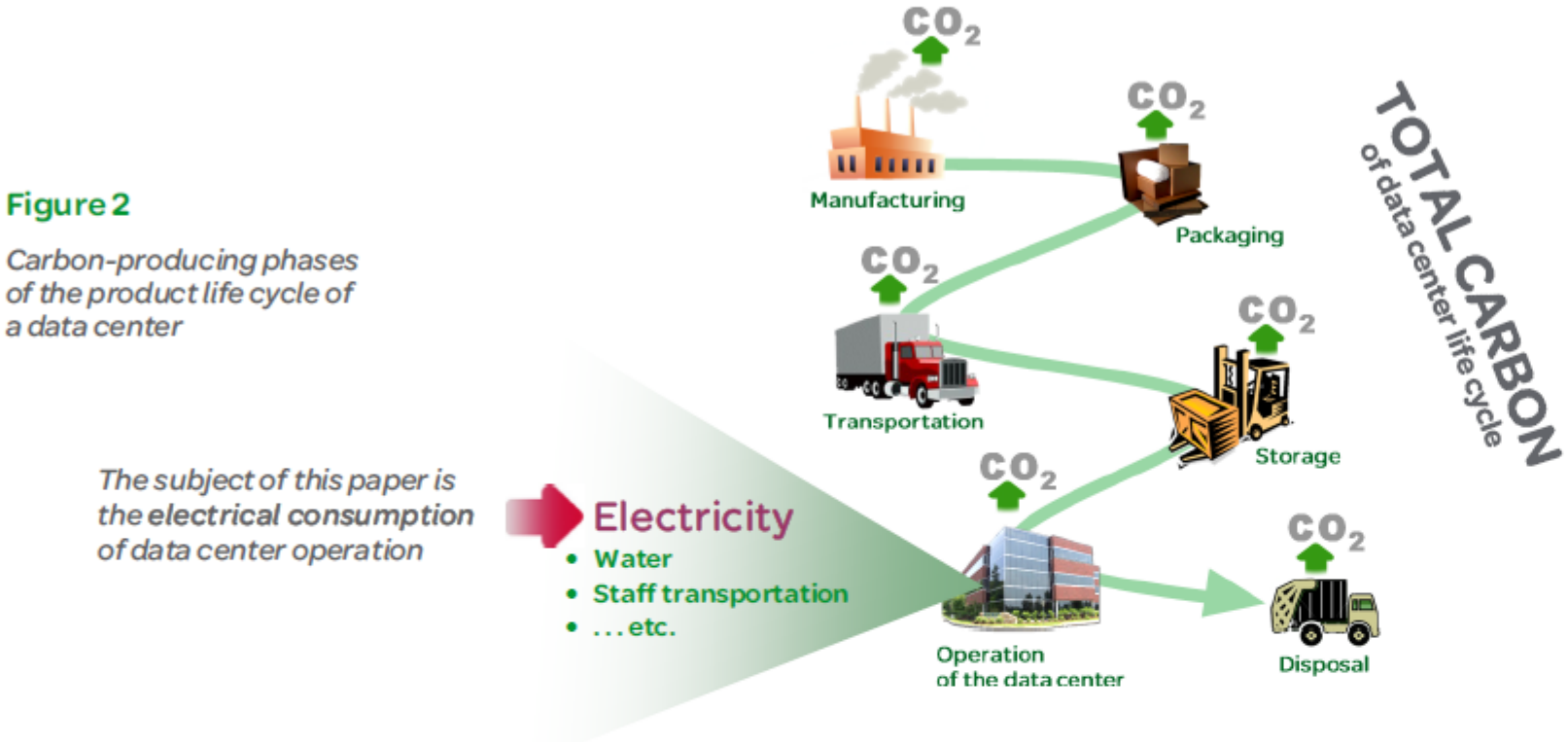
- Large file sizes
- Liquid helium: finite resource obtained via fossil fuel extraction

Data centres: multiple issues

Estimating a Data Center's Electrical Carbon Footprint

White Paper 66

by Dennis Bouley



https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: construction

> The building "shell"

Quantified for a 1MW data center

Concrete	
Masonry	Brick, stone, grout
Metals	Steel beams, lead pipes, copper wires, aluminum sheet metal, stairs, railings, floor plates, grates, nails, screws, bolts, aluminum flashing, sheet metal, aluminum ventilation, louver systems
Wood, plastic, composite	Room framing, wire coatings, doors, windows
Thermal/moisture protection	Insulation, vapor barriers
Water	Cleaning, cooling, fire suppression
Chemicals	Glue, glycol, cleaners, water repellants sealants, fire suppression
Glass	
Tar	Roofs, roads, sidewalks, parking lots
Shingles, tiles	

Materials - building shell 5,700 ft^2 (530 m^2) office facility	Tonnes of CO_2	Percentage of total
Foundation (concrete)	4.7	4%
Flooring (concrete slab, insulation)	39.9	31%
Ceilings (plaster board)	2.3	2%
Structure (steel beams)	15.4	12%
External walls (brick, insulation)	32.1	25%
Internal walls (wood frame and plasterboard)	8.7	7%
Stairs (concrete)	1.1	1%
Windows (glass and frame)	0.59	0.4%
Internal doors (particle board)*	-0.4	-0.3%
External doors (plastic)	0.6	0.5%
Roof (wood, concrete, insulation)	23.4	18%
TOTAL	128.3	100%

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: server manufacture

> The data center

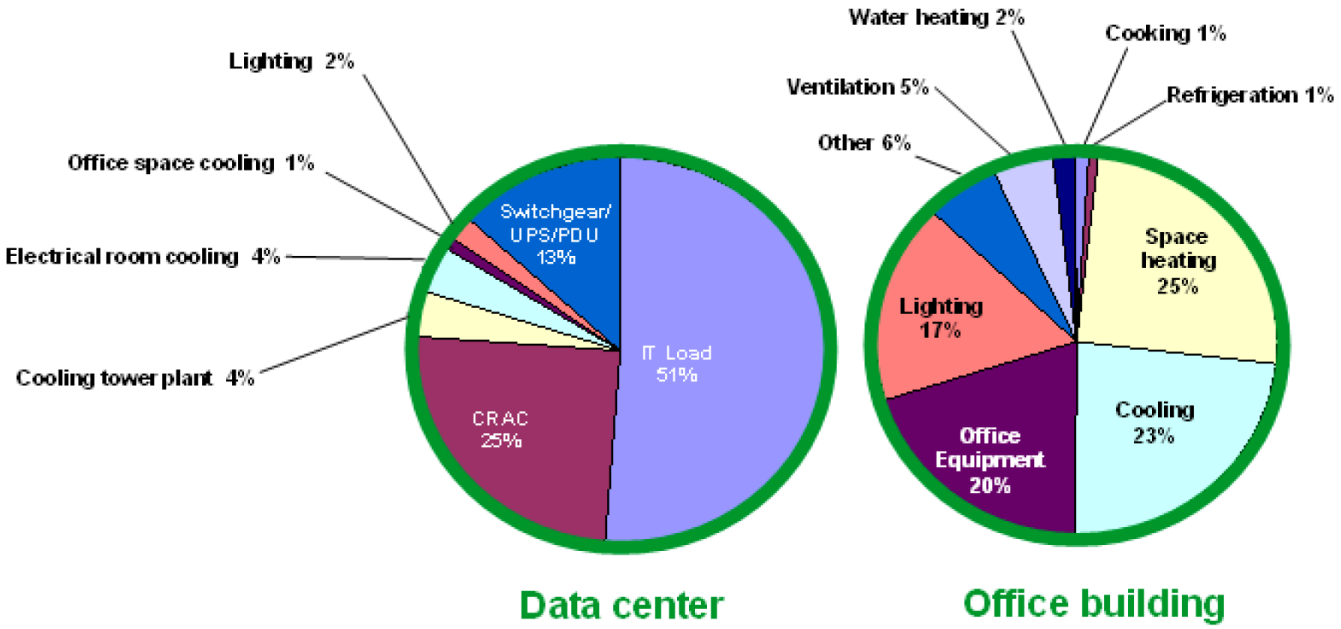
Quantified for a 1MW data center

Electricity	177,000,000 kW-hr
Water	60,000,000 gal (227,000,000 l)
Copper	145,000 lbs (65,771 kg)
Lead	21,000 lbs (9,525 kg)
Plastic	33,000 lbs (14,968 kg)
Aluminum	73,000 lbs (33,112 kg)
Solder	12,000 lbs (5,443 kg)
Steel	377,000 lbs (171,004 kg)

Assumptions: 10-year lifetime, high redundancy, two IT refreshes, includes power/cooling/racks/IT, does not include the building

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: energy



"Data centres can be **40 times** as energy intensive as an office building"

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: energy

3 key factors

- IT load

  - more activity, more energy required

- Location

  - extremes of temperature require more cooling

- Efficiency

  - system design; local vs cloud

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: energy

Where is the energy coming from?

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Data centres: server disposal

> The data center

Quantified for a 1MW data center

Electricity	177,000,000 kW-hr
Water	60,000,000 gal (227,000,000 l)
Copper	145,000 lbs (65,771 kg)
Lead	21,000 lbs (9,525 kg)
Plastic	33,000 lbs (14,968 kg)
Aluminum	73,000 lbs (33,112 kg)
Solder	12,000 lbs (5,443 kg)
Steel	377,000 lbs (171,004 kg)

Assumptions: 10-year lifetime, high redundancy, two IT refreshes, includes power/cooling/racks/IT, does not include the building

"10 year lifetime"

"Two IT refreshes"

https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-estimating-data-centers-carbon-footprint.pdf

Neuroimaging dataset size

This applies to any online activity: streaming, shopping, Googling*…

…but neuroimaging datasets are particularly large

- My local multimodal MRI dataset (n=100) GB
- HCP Young-Adult (n=1200) TB
- UK Biobank (n=100,000) PB
  - Select sequences TB
  - ‘Imaging Derived Phenotypes’ (IDPs) GB (MB)

*use www.ecosia.org instead to plant trees

Conclusion #1

The more you store & share,

the bigger your footprint

(even if data centre runs on renewable energy)

Liquid helium

(NB. OPM-MEG no helium required)

https://www.resilience.org/stories/2019-05-19/helium-is-a-finite-resource-who-knew/

https://www.quora.com/What-is-the-carbon-footprint-of-a-3T-MRI-scanner

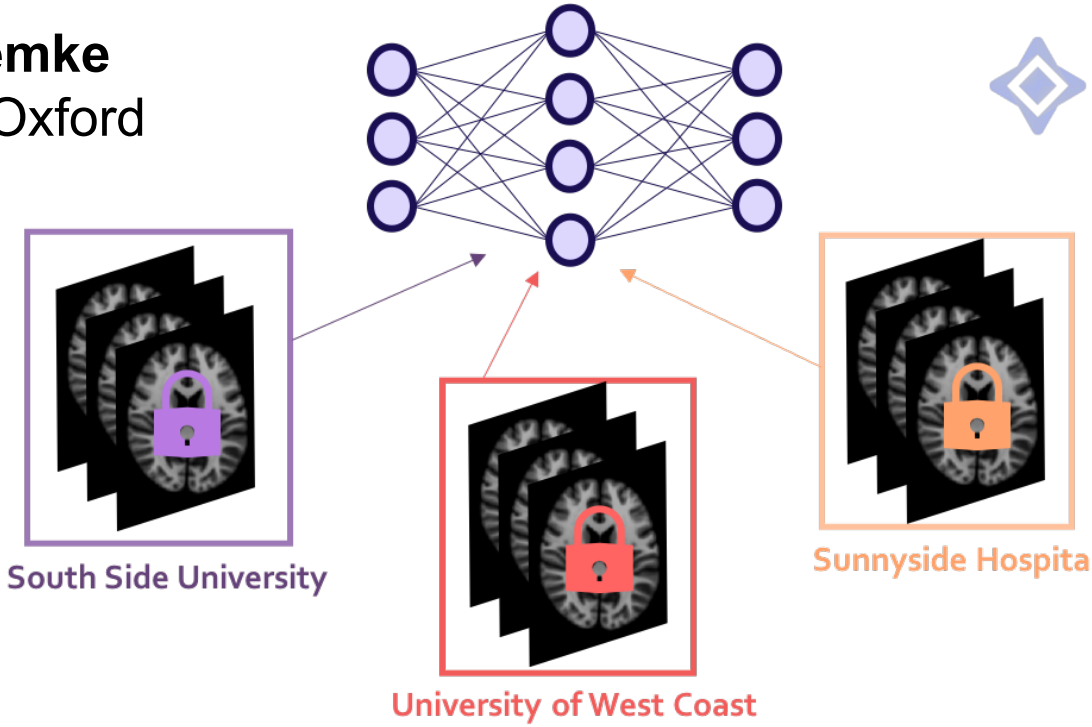
Valuing reusable datasets

Let's say we have 3 open neuroimaging datasets

- My local multimodal MRI dataset (n=100) — GB ?
  Decent quality, lowish n
- HCP Young-Adult (n=1200) — TB ✓
  Top quality, large n
- UK Biobank (n=100,000) — (PB)
  Solid quality, huge n, IDPs available — **TB/GB** ✓

Efficient open workflows

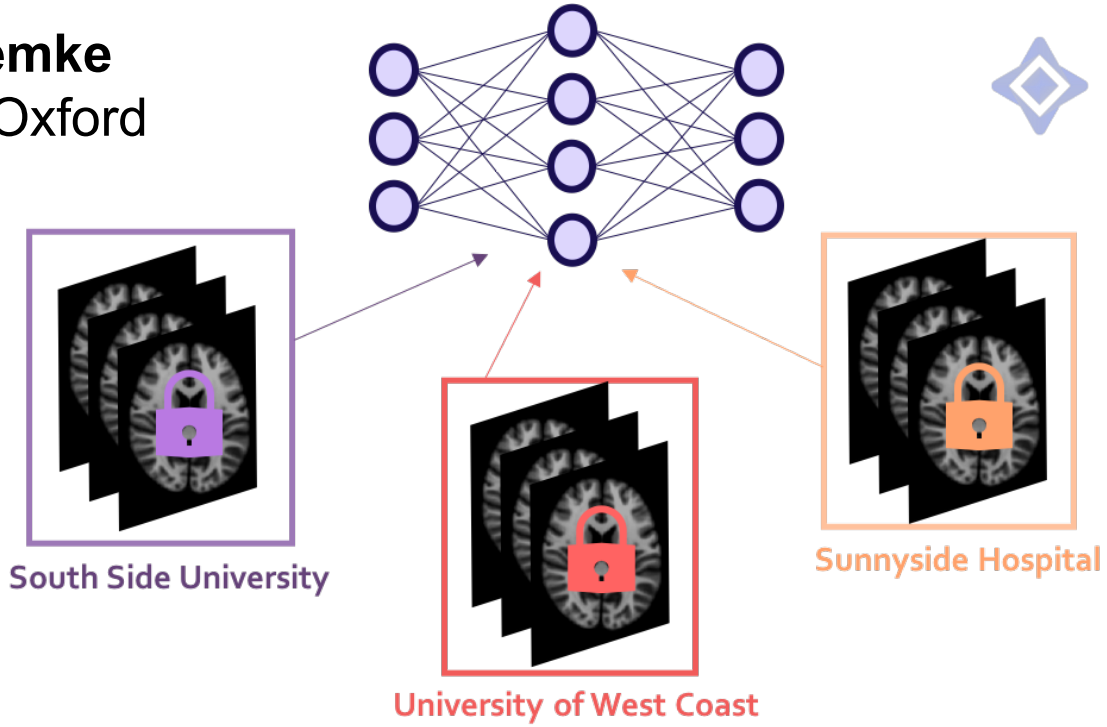
Emma Bluemke
University of Oxford



"these privacy-preserving developments allow us to train our model on data from multiple institutions, *without sharing the patient data*…in other words, I no longer have to request a copy of a dataset in order to use it in my statistical study"

Efficient open workflows

Emma Bluemke
University of Oxford



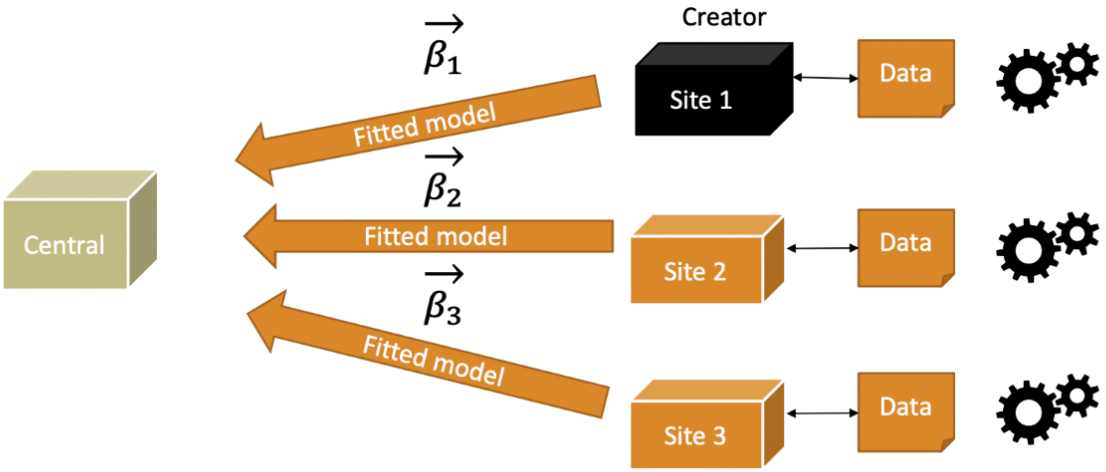
HOWEVER: “certain privacy preserving techniques increase the energy needed for compute, and some have a huge overhead for communication between the parties performing the computations, so I’m not actually sure if it would be net environmentally friendly”

Efficient open workflows

Eric Verner
Georgia State University

COINSTAC

TReNDS



"The sites only access their own data, not a remote, shared dataset…this may be more environmentally friendly than a remote, shared dataset because the costs of transfer of large amounts of data are avoided."

Conclusion #2

Reusing existing, popular datasets,

over installing new scanner & acquiring your own,

may lower your footprint

(while sharing average quality datasets, with limited interest,

may not be worth the carbon price)

Repository green credentials

OSF

- Uses Google: 100% renewable
- “matches 100% of energy consumed by our operations with renewable energy”

NEUROVAULT

- Uses Amazon Web Services
- AWS currently ~50% renewable; Oregon storage offset
- NeuroVault currently looking at switching to guaranteed renewable

https://aws.amazon.com/about-aws/sustainability/#progress

Repository green credentials

OSF

- Uses Google: 100% renewable
- “matches 100% of energy consumed by our operations with renewable energy”

NEUROVAULT

- Uses Amazon Web Services
- AWS currently ~50% renewable; Oregon storage offset
- NeuroVault currently looking at switching to guaranteed renewable

OpenNEURO

- Uses AWS (storage location?)

Conclusion #3

OSF currently the ‘green standard’

(and if your data have low reuse potential, sharing just summary on NeuroVault, rather than full data on OpenNeuro, will lower your footprint)

Reducing before reusing

"**Reduce**, reuse, recycle"

- We need to fundamentally reduce our consumption first and foremost, across all domains of life

  - this includes scientific research

- "Slow science": (Frith, 2020, *TICS*)

  - less, but better

  - restrict output

Serious summary

- Significant environmental impacts of data acquisition, and data sharing (although we need formal analysis to draw thresholds)
- “What do I really need to share, and how can I best do that cf environmental impact?”
- **Reduce**, reuse, recycle: we need to have honest conversation about reducing consumption in all areas of life
    - reductions need to be substantial
    - they need to happen very fast
- Tech solutions for climate crisis do not exist & won’t arrive in time; offsetting inappropriate & not possible on scale or pace we need
- If we do not urgently address the biggest challenge humanity has ever faced, we will not be able to do science full stop – within our lifetimes

Community action

- We need quantitative assessment of relative sustainability merits
  - different data acquisition,
  - storage,
  - sharing,
  - workflow practices,

  specialised for neuroimaging
- We need to collaborate internationally as a community, to work out how to minimise the impact of neuroimaging research

…OHBM Environment SIG

https://tinyurl.com/ybgxar3t

Organization for Human Brain Mapping

Advancing Understanding of the Human Brain

Further reading

- Aron et al (2020) *Neuron*

  "How can neuroscientists respond to the climate emergency?"

- Rae lab website

  https://www.sussex.ac.uk/psychology/abc-lab/climate-change

- Absolute Zero report

  https://www.repository.cam.ac.uk/handle/1810/299414

- IPCC

  https://www.ipcc.ch/

The hidden cost of open science: what's our footprint?

Dr Charlotte Rae
c.rae@sussex.ac.uk
@NeuroRae

UNIVERSITY OF SUSSEX