



HAL
open science

From NIDM to BIDS-Prov: A personal perspective on 10 years of metadata standardization towards neuroimaging data sharing

Camille Maumet

► To cite this version:

Camille Maumet. From NIDM to BIDS-Prov: A personal perspective on 10 years of metadata standardization towards neuroimaging data sharing. 2024 - The future of FAIR data management and engineering for open and reproducible neuroimaging, Jul 2024, Oldenburg, Germany. pp.1-24, 2024. hal-04651624

HAL Id: hal-04651624

<https://inria.hal.science/hal-04651624>

Submitted on 17 Jul 2024

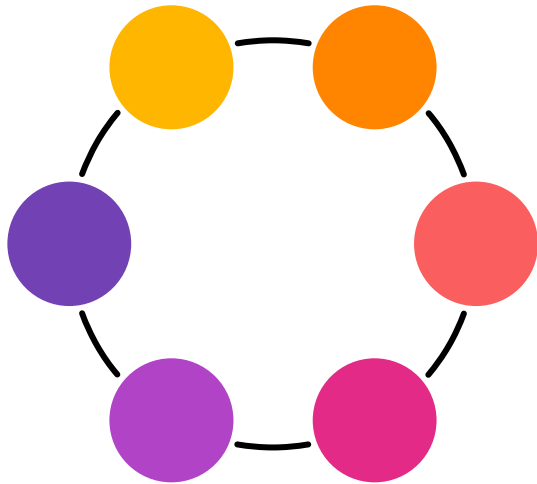
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.



Distributed under a Creative Commons Attribution 4.0 International License

Session 2: Metadata standardization



From NIDM to BIDS-Prov

A personal perspective on 10 years of
metadata standardization towards
neuroimaging data sharing

Camille Maumet

Inria, Univ Rennes, Inserm, CNRS, IRISA

10 years of standards for data sharing...

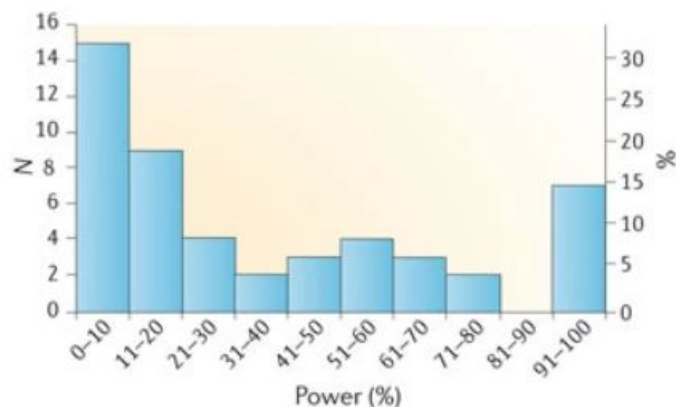
2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024



Why sharing neuroimaging data?

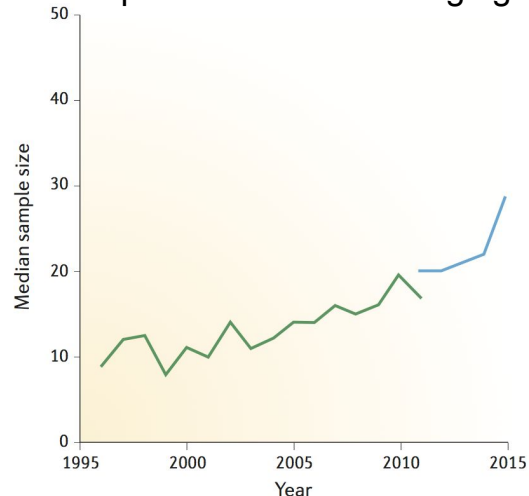
Small samples and low statistical power

Power of neuroscience studies



(Button et. al, 2013)

Sample sizes in neuroimaging



(Poldrack et. al, 2013)

High rate of **false positives** in the literature

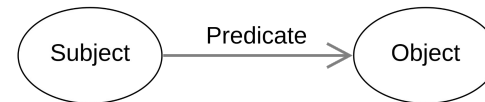


Solution: Data sharing to get to larger samples

10 years of standards for data sharing...



incf Neuroimaging Data Sharing (NIDASH) Task Force



Members of NIDASH (2014)

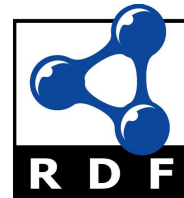
David Kennedy (USA), **Jean-Baptiste Poline** (France), Cameron Craddock (USA), Guillaume Flandin (UK), Stephan Gerhard (Switzerland) Chris Gorgolewski (UK) Satrajit Ghosh (USA) Yaroslav Halchenko (USA) Michael Hanke (Germany) Christian Haselgrove (USA) **David Keator** (USA) Arno Klein (USA) Daniel Marcus (USA) Camille Maumet (UK) Franck Michel (France) Nolan Nichols (USA) Thomas Nichols (UK) Simon Milton (Australia) Russ Poldrack (USA) Rich Stoner (USA) Jessica Turner (USA).

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024



incf Neuroimaging Data Sharing (NIDASH) Task Force



Members of NIDASH (2014)

David Kennedy (USA), **Jean-Baptiste Poline** (France), Cameron Craddock (USA), Guillaume Flandin (UK), Stephan Gerhard (Switzerland) Chris Gorgolewski (UK) Satrajit Ghosh (USA) Yaroslav Halchenko (USA) Michael Hanke (Germany) Christian Haselgrove (USA) **David Keator** (USA) Arno Klein (USA) Daniel Marcus (USA) Camille Maumet (UK) Franck Michel (France) Nolan Nichols (USA) Thomas Nichols (UK) Simon Milton (Australia) Russ Poldrack (USA) Rich Stoner (USA) Jessica Turner (USA).

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

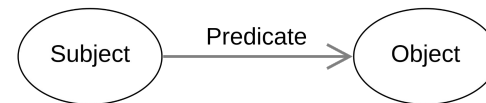
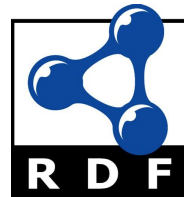


incf Neuroimaging Data Sharing (NIDASH) Task Force

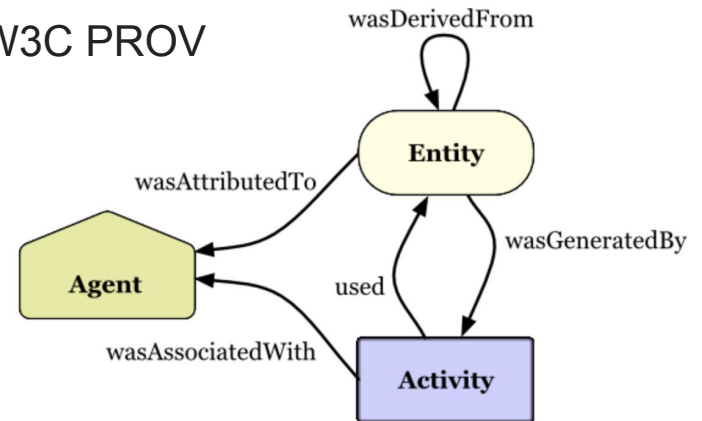


Members of NIDASH (2014)

David Kennedy (USA), **Jean-Baptiste Poline** (France), Cameron Craddock (USA), Guillaume Flandin (UK), Stephan Gerhard (Switzerland) Chris Gorgolewski (UK) Satrajit Ghosh (USA) Yaroslav Halchenko (USA) Michael Hanke (Germany) Christian Haselgrove (USA) **David Keator** (USA) Arno Klein (USA) Daniel Marcus (USA) Camille Maumet (UK) Franck Michel (France) Nolan Nichols (USA) Thomas Nichols (UK) Simon Milton (Australia) Russ Poldrack (USA) Rich Stoner (USA) Jessica Turner (USA).



W3C PROV



10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

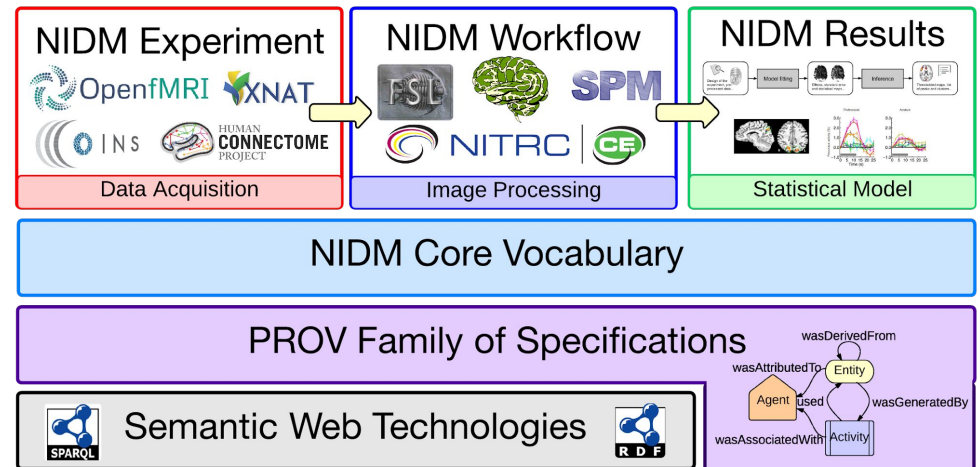


incf NeuroImaging Data Sharing (NIDASH) Task Force



Members of NIDASH (2014)

David Kennedy (USA), **Jean-Baptiste Poline** (France), **Cameron Craddock** (USA), **Guillaume Flandin** (UK), **Stephan Gerhard** (Switzerland) **Chris Gorgolewski** (UK) **Satrajit Ghosh** (USA) **Yaroslav Halchenko** (USA) **Michael Hanke** (Germany) **Christian Haselgrove** (USA) **David Keator** (USA) **Arno Klein** (USA) **Daniel Marcus** (USA) **Camille Maumet** (UK) **Franck Michel** (France) **Nolan Nichols** (USA) **Thomas Nichols** (UK) **Simon Milton** (Australia) **Russ Poldrack** (USA) **Rich Stoner** (USA) **Jessica Turner** (USA).



[Keator et al, NeuroImage 2013]

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

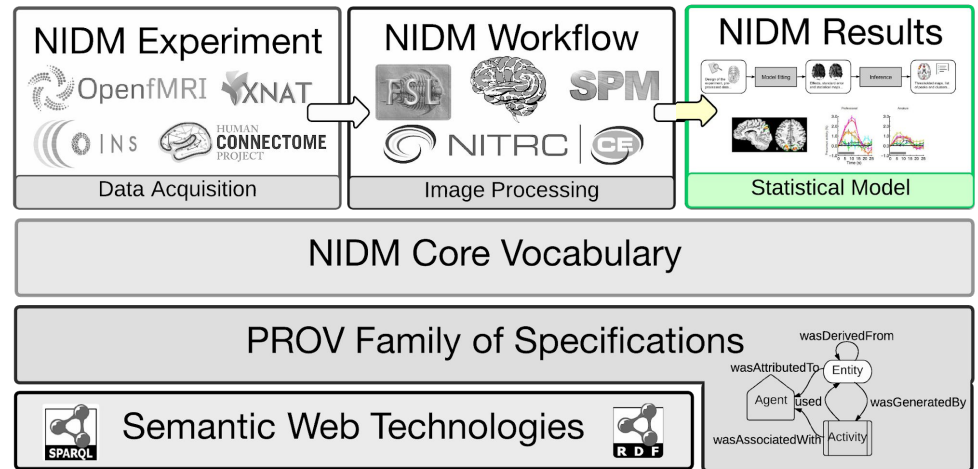


incf NeuroImaging Data Sharing (NIDASH) Task Force



Members of NIDASH (2014)

David Kennedy (USA), **Jean-Baptiste Poline** (France), **Cameron Craddock** (USA), **Guillaume Flandin** (UK), **Stephan Gerhard** (Switzerland) **Chris Gorgolewski** (UK) **Satrajit Ghosh** (USA) **Yaroslav Halchenko** (USA) **Michael Hanke** (Germany) **Christian Haselgrove** (USA) **David Keator** (USA) **Arno Klein** (USA) **Daniel Marcus** (USA) **Camille Maumet** (UK) **Franck Michel** (France) **Nolan Nichols** (USA) **Thomas Nichols** (UK) **Simon Milton** (Australia) **Russ Poldrack** (USA) **Rich Stoner** (USA) **Jessica Turner** (USA).



[Keator et al, NeuroImage 2013]

[Maumet et al, Sci Data 2016]

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

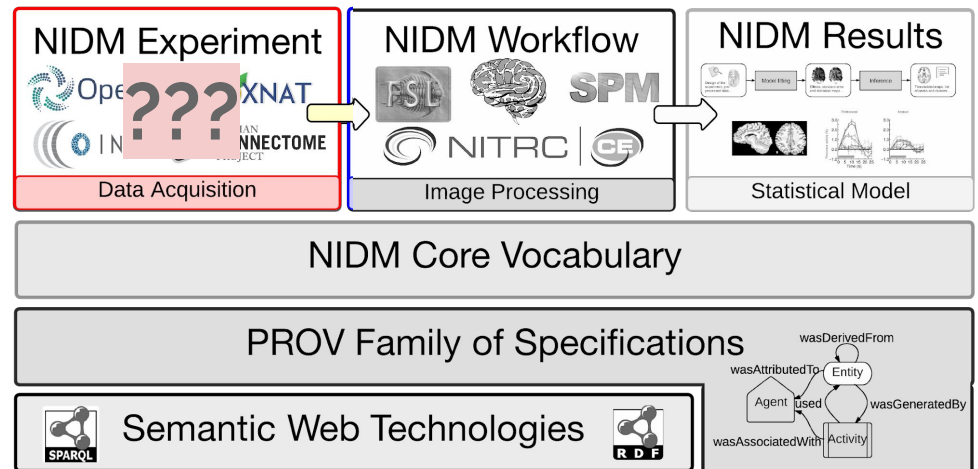


incf NeuroImaging Data Sharing (NIDASH) Task Force



Members of NIDASH (2014)

David Kennedy (USA), **Jean-Baptiste Poline** (France), **Cameron Craddock** (USA), **Guillaume Flandin** (UK), **Stephan Gerhard** (Switzerland) **Chris Gorgolewski** (UK) **Satrajit Ghosh** (USA) **Yaroslav Halchenko** (USA) **Michael Hanke** (Germany) **Christian Haselgrove** (USA) **David Keator** (USA) **Arno Klein** (USA) **Daniel Marcus** (USA) **Camille Maumet** (UK) **Franck Michel** (France) **Nolan Nichols** (USA) **Thomas Nichols** (UK) **Simon Milton** (Australia) **Russ Poldrack** (USA) **Rich Stoner** (USA) **Jessica Turner** (USA).



[Keator et al, NeuroImage 2013]

[Maumet et al, Sci Data 2016]

10 years of standards for data sharing...



10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024



Brain Imaging Data Structure



[Gorgolewski et al, Sci Data 2016]

```
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/
```

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024



Brain Imaging Data Structure

First
Governance
For BIDS



[Gorgolewski et al, Sci Data 2016]



(Holdgraf et al., 2019)



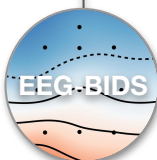
(Karakuzu et al., 2022)



(Bourget et al., 2022)



(Niso et al., 2018)



(Pernet et al., 2019)



(Moreau et al., 2020)



(Clement et al., 2022)



(Norgaard et al., 2022)

```
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/
```

>1 million BIDS-compliant images
in 2024

Source: [MRIQC Api](#)

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

Why sharing neuroimaging data?



10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

Why sharing neuroimaging data?



Generalizability and fairness

X-ray: Lung opacity detection

Model trained on male images, tested on female images



(Larrazabal et. al, 2020)

Biased outcomes



Solution: Decentralized data sharing
to get **more diverse samples**

France →

shancir

10 years of standards for data sharing...

2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

Generalizability and fairness

X-ray: Lung opacity detection

Model trained on male images, tested on female images



(Larrazabal et. al, 2020)

Biased outcomes



Solution: Decentralized data sharing to get more diverse samples

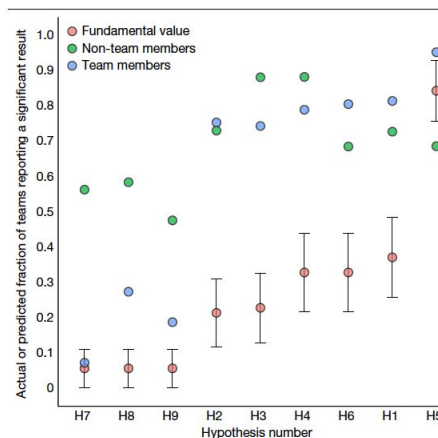
France →

shaneir

Why sharing neuroimaging data?



Different pipelines, different results



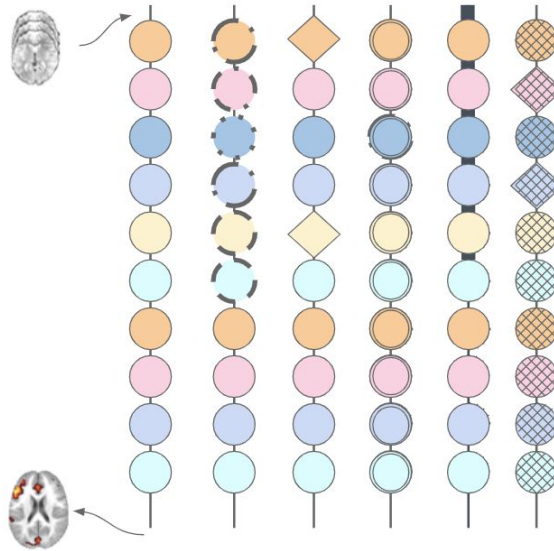
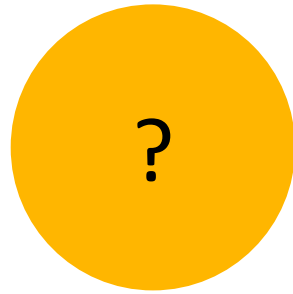
(Botvinik-Nezer et. al, 2020)

Analytical variability

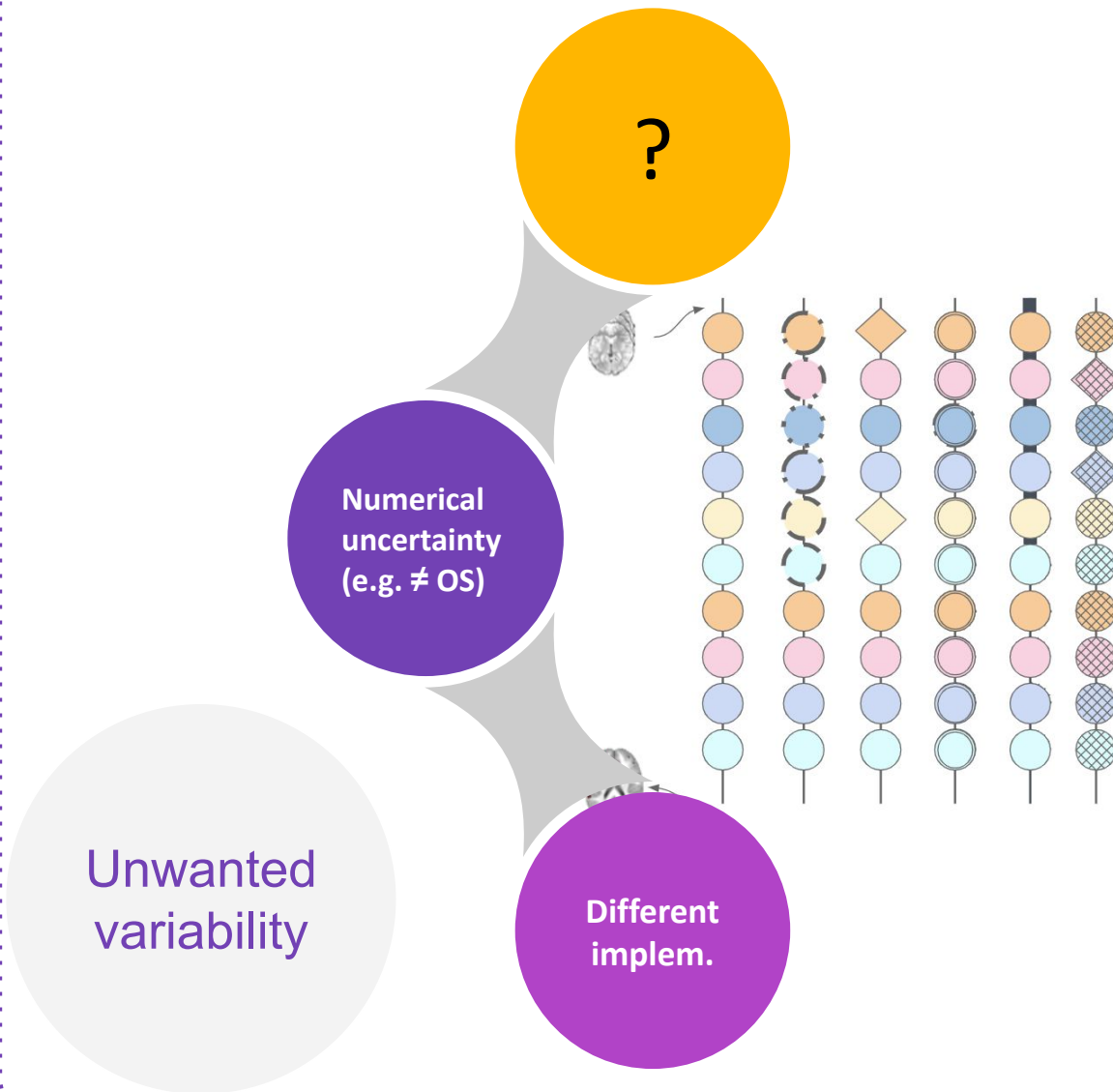


(Partial) Solution: Transparency on the pipeline used

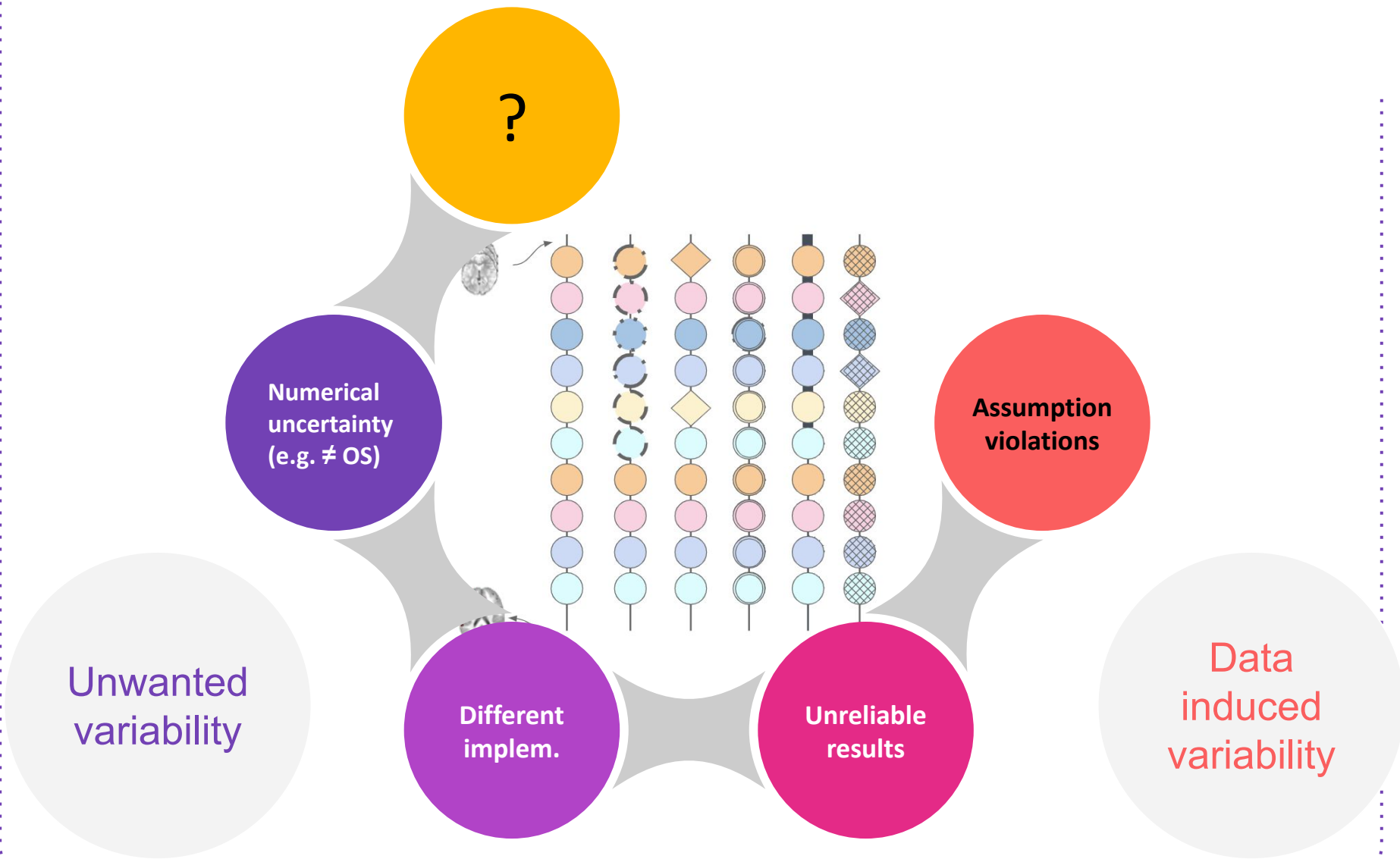
Understanding analytical variability



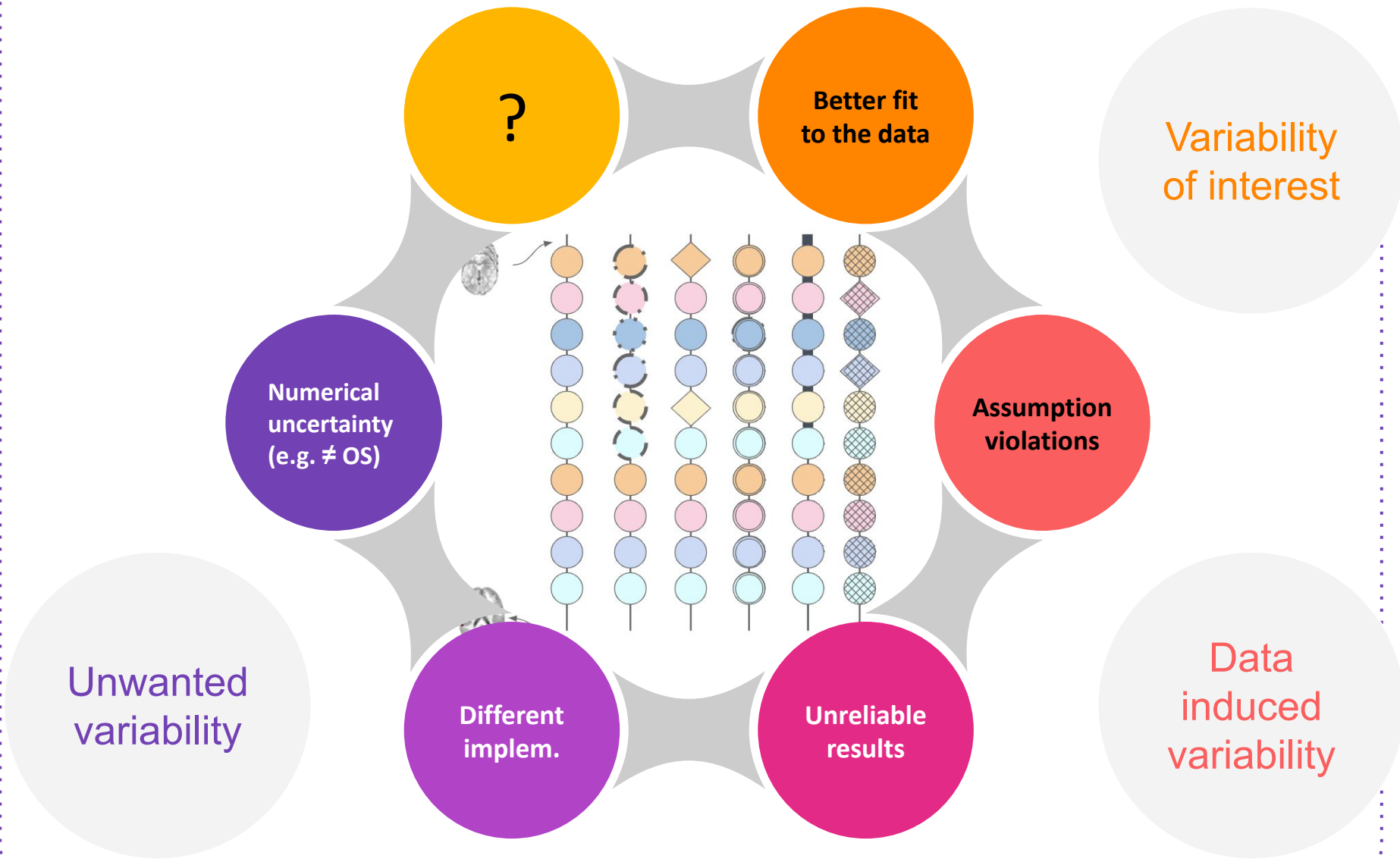
Understanding analytical variability



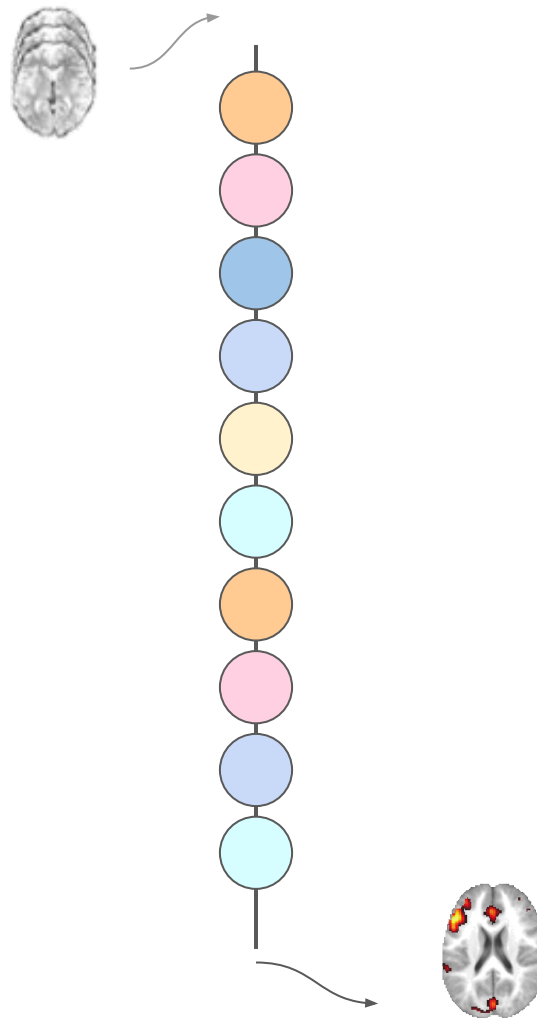
Understanding analytical variability



Understanding analytical variability



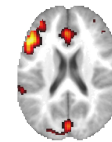
What is BIDS-Prov



What is BIDS-Prov

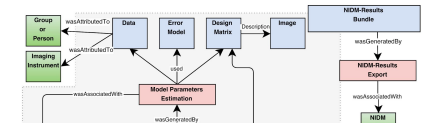
```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/
```

BIDS
“raw”



BIDS-derivatives
(NIDM-Results)

```
my_dataset/  
├── participants.tsv  
└── sub-01/
```

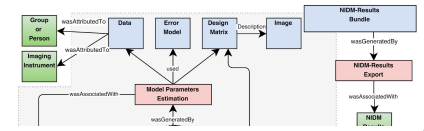
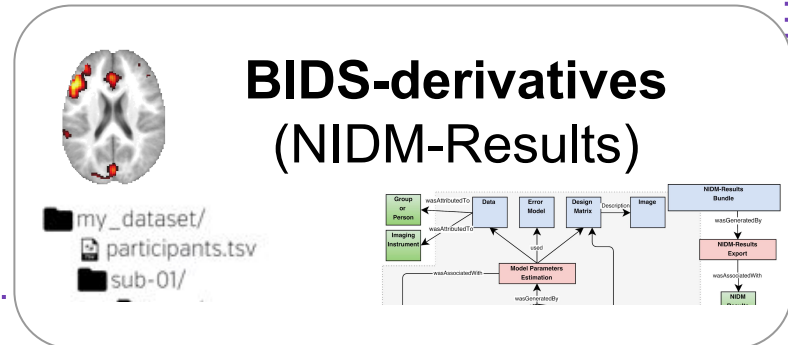


What is BIDS-Prov



BIDS-Prov

- BIDS focus on metadata that **maximize (re)usability**
- Relies on expert knowledge
- For **derived/intermediary data**, we sometimes lack expert knowledge to identify the metadata that “matter” in neuroimaging pipelines and that will impact reusability.



What is BIDS-Prov

- BIDS focus on metadata that **maximize (re)usability**
- Relies on expert knowledge
- For **derived/intermediary data**, we sometimes lack expert knowledge to identify the metadata that “matter” in neuroimaging pipelines and that will impact reusability.

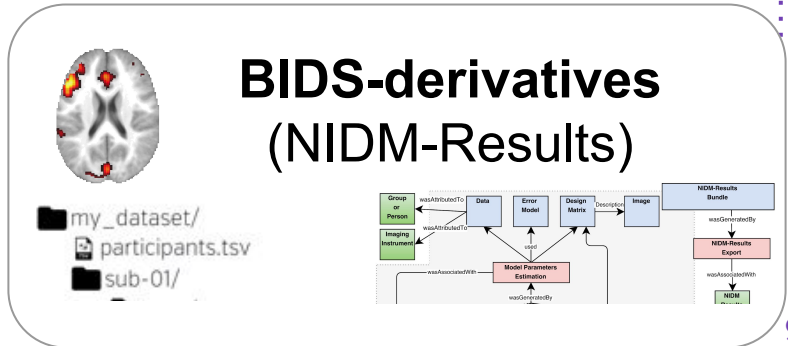
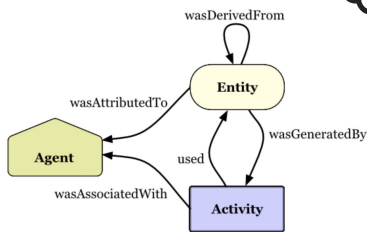


JSON-LD

RDF +



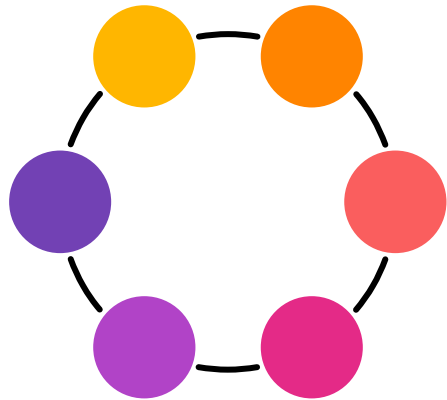
BIDS-Prov



Open discussion

- Transparency of data processing
 - How to we keep the **ease of use of BIDS** while having **large expressivity** (many parameters)?
- Different levels of **granularity**...
- Or **should we better understand analytical variability first** to know which parameters are most impactful?
- Standards and **large-scale data sharing** efforts (ABCD, HCP, UK Biobank) not always aligned
- **Decoupling metadata and data sharing** is particularly important for the EU (GDPR)

Session 2: Metadata standardization



From NIDM to BIDS-Prov

A personal perspective on 10 years of
metadata standardization towards
neuroimaging data sharing

Camille Maumet

Supported by: ANR JCJC VICUNA



 @cmaumet.bsky.social

 @cmaumet@fediscience.org

 @cmaumet