



**HAL**  
open science

## BIDS-Prov: Recording neuroimaging provenance

Camille Maumet

► **To cite this version:**

Camille Maumet. BIDS-Prov: Recording neuroimaging provenance. Brain Imaging Data Structure (BIDS) derivatives meeting, Jun 2023, Copenhagen, Denmark. hal-04137329

**HAL Id: hal-04137329**

**<https://inria.hal.science/hal-04137329>**

Submitted on 22 Jun 2023

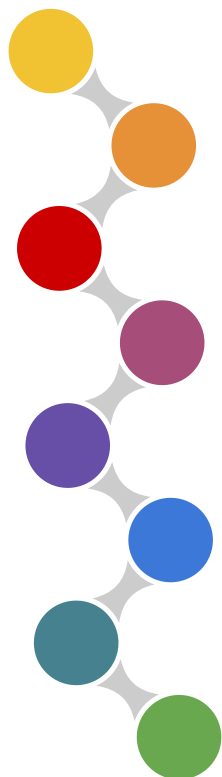
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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

Workshop BIDS-derivatives  
June 21-23, 2023



# BIDS-Prov

Recording neuroimaging provenance

Camille Maumet

Inria, Univ Rennes, Inserm, CNRS, IRISA



# The BIDS-Prov Team

## BIDS Extension Proposal 28

**Moderators:** Satra Ghosh and Camille Maumet

**Contributors:** Rémi Adon, Tibor Auer, Stefan Appelhoff, Thomas Betton, Hermann Courteille, Michael Dayan, Eric Earl, Dorota Jarecka, Yaroslav Halchenko, Matthieu Joulot, Chris Markiewicz, Cyril R. Pernet, Jean-Baptiste Poline, Cyril Regan, Omar El Rifai, Sarah Saneei, Ghislain Vaillant.

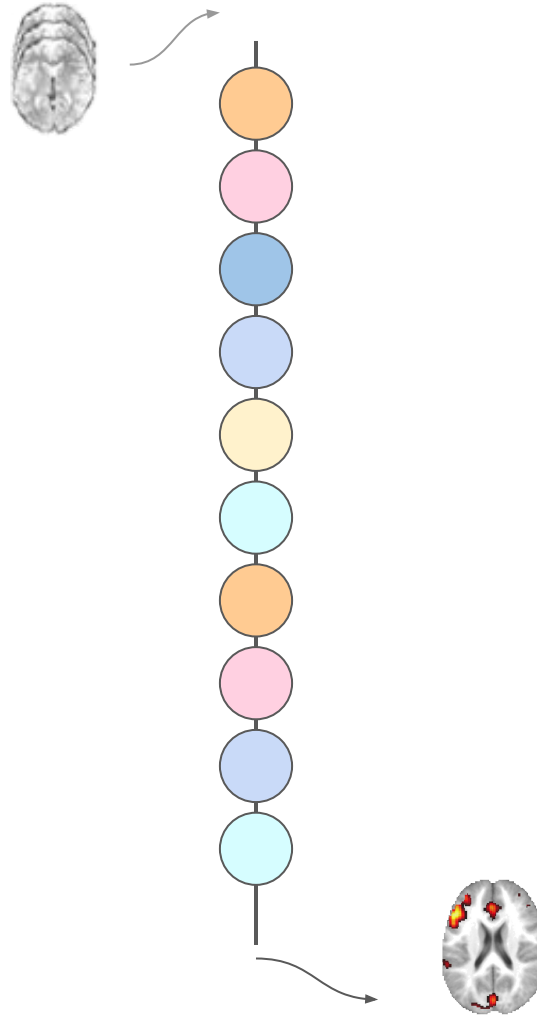
Thank you!



**BIDS steering committee:** Guiomar Niso, Melanie Ganz, Robert Oostenveld, Russell Poldrack, Ariel Rokem.

**BIDS maintainers:** Stefan Appelhoff, Chris Markiewicz, Franklin Feingold, Taylor Salo, Rémi Gau, Ross Blair, Anthony Galassi, Eric Earl, Christine Rogers.

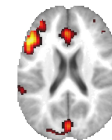
# What is **BIDS-Prov**?



# What is **BIDS-Prov**?

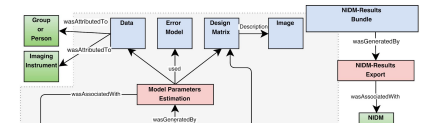
```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/
```

**BIDS**  
“raw”

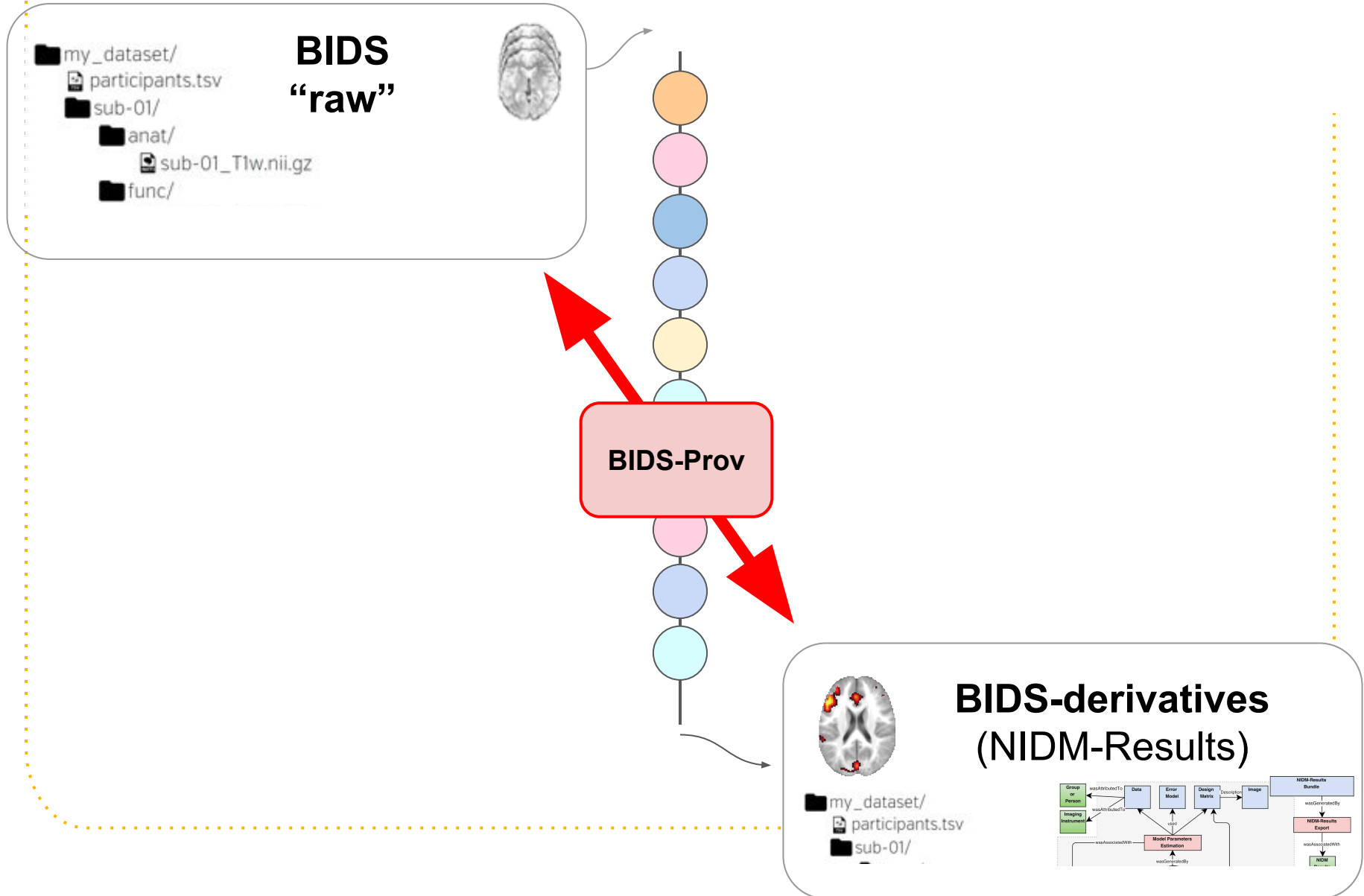


**BIDS-derivatives**  
(NIDM-Results)

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



# What is **BIDS-Prov**?



# Goal of BIDS-Prov

- Describing **the provenance** of a BIDS dataset.
- Description is **retrospective**,  
i.e. set of steps that were executed in order to obtain the dataset.

# Why BIDS-Prov?

- BIDS focus on metadata that will maximize the **(re)usability** of datasets
  - E.g. onsets for task-fMRI
- Relies on expert knowledge to identify metadata that “matter”
- But for derived datasets this can get complicated...

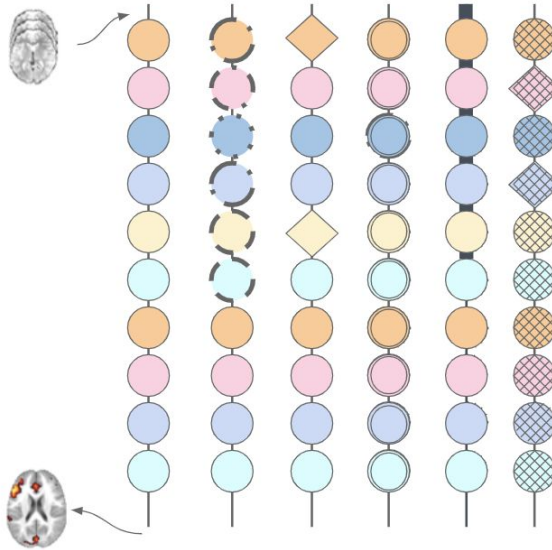


# What is a reusable derived dataset?

- **Can depend on the software package** (different stopping points) → Paul
- We sometimes **lack expert knowledge** to identify metadata that “matter” in neuroimaging pipelines

# Understanding analytical variability...

?



# Understanding analytical variability...

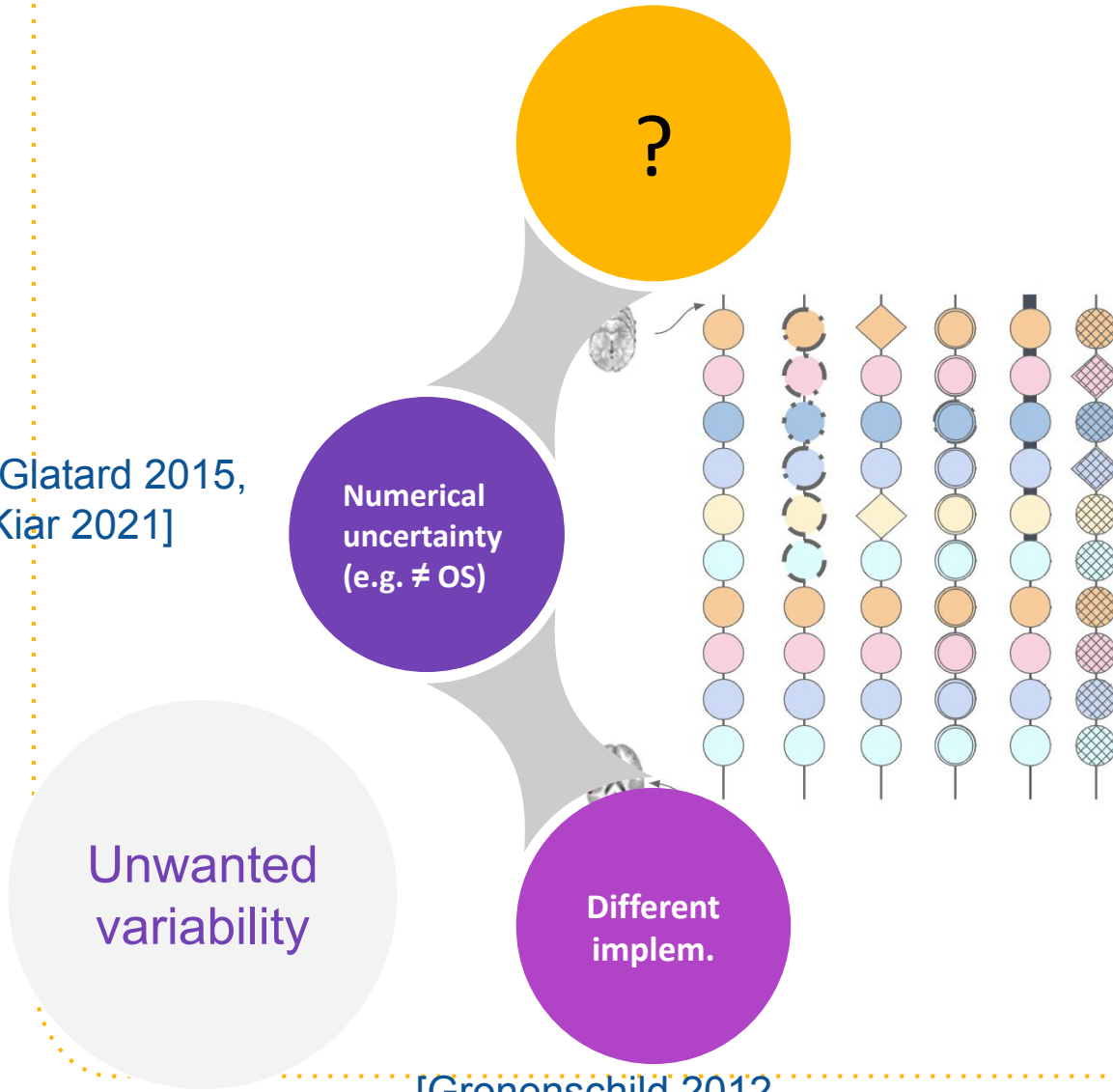
[Glatard 2015,  
Kiar 2021]

Numerical  
uncertainty  
(e.g.  $\neq$  OS)

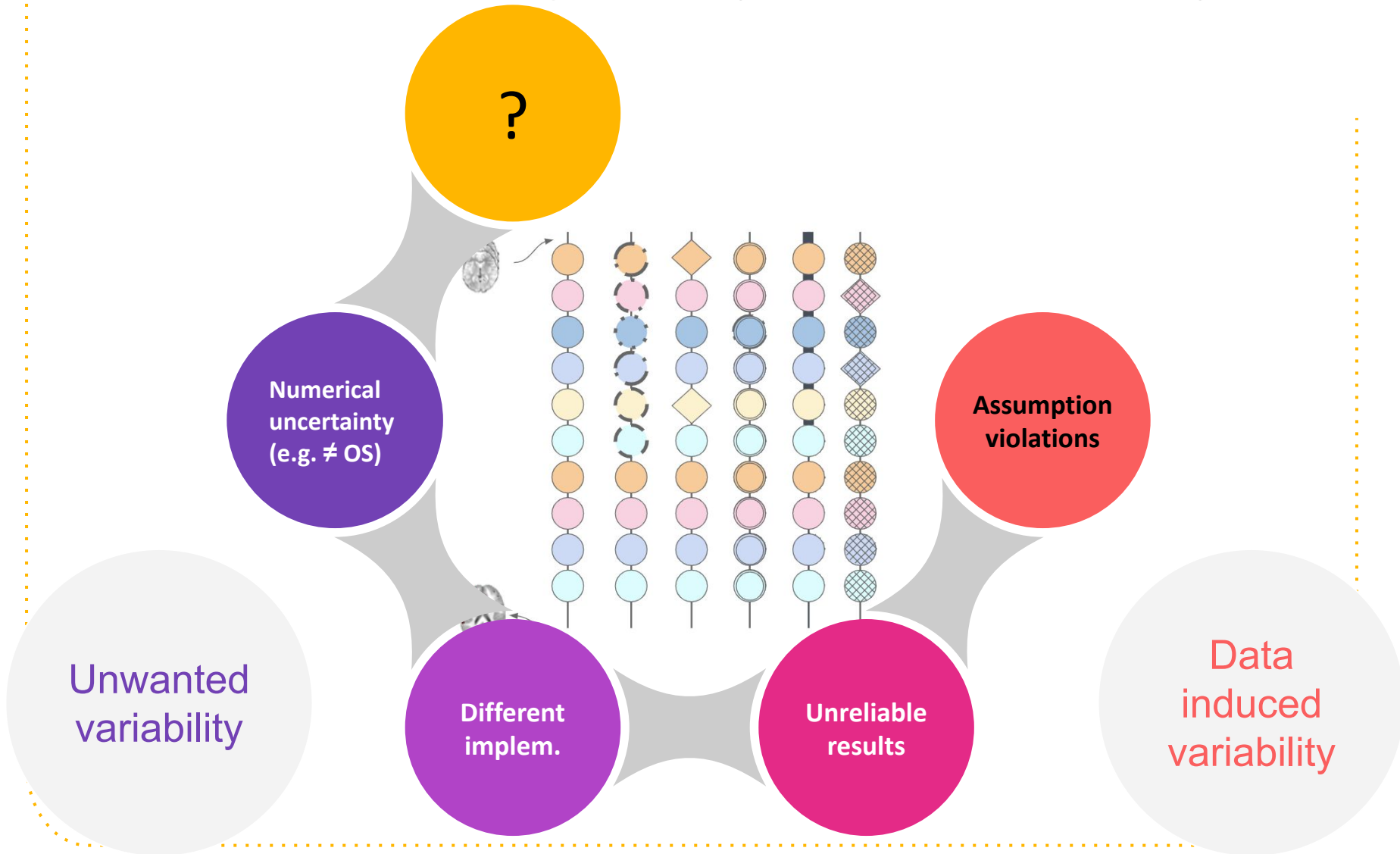
Unwanted  
variability

Different  
implem.

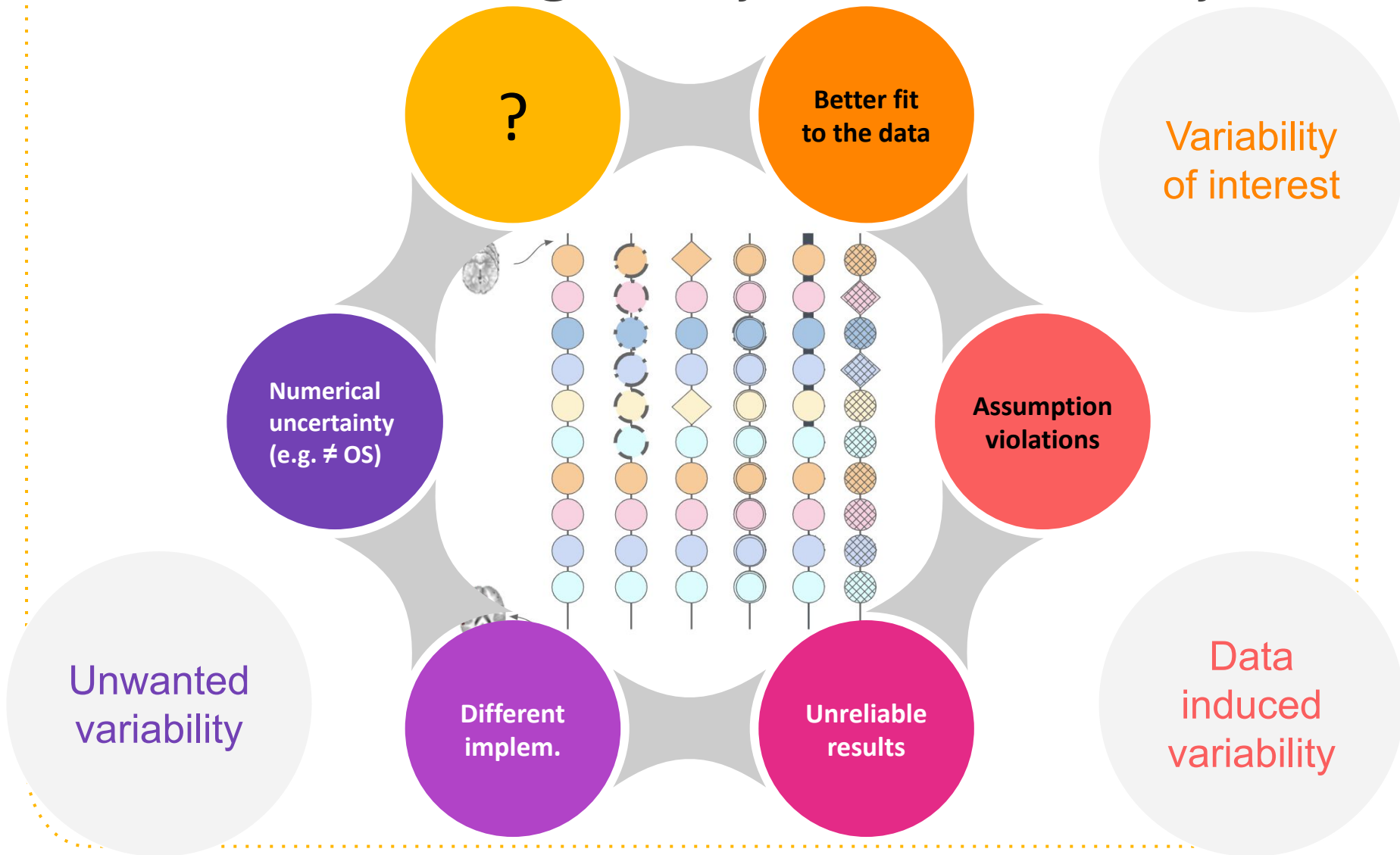
[Gronenschild 2012,  
Bowring 2019]



# Understanding analytical variability...



# Understanding analytical variability...



# What is a reusable derived dataset?

- **Can depend on the software package** (different stopping points) → Paul
- We sometimes **lack expert knowledge** to identify metadata that “matter” in neuroimaging pipelines
- **Expert knowledge can evolve** and in that case it is really useful to be able and go back to previous studies.

# What is a reusable derived dataset?

- **Can depend on the software package** (different stopping points) → Paul
- We sometimes **lack expert knowledge** to identify metadata that “matter” in neuroimaging pipelines
- **Expert knowledge can evolve** and in that case it is really useful to be able and go back to previous studies.

We need detailed description of "what happened" to a dataset (i.e. provenance)

# Provenance in neuroimaging software

- **AFNI history** is stored in the BRIK files
- **MINC** also has a mechanism to store provenance
- SPM stores some provenance in the **SPM.mat**, saving the **matlabbatch** as a batch script is also provenance.
- Similarly in FSL with the **fsf** files.

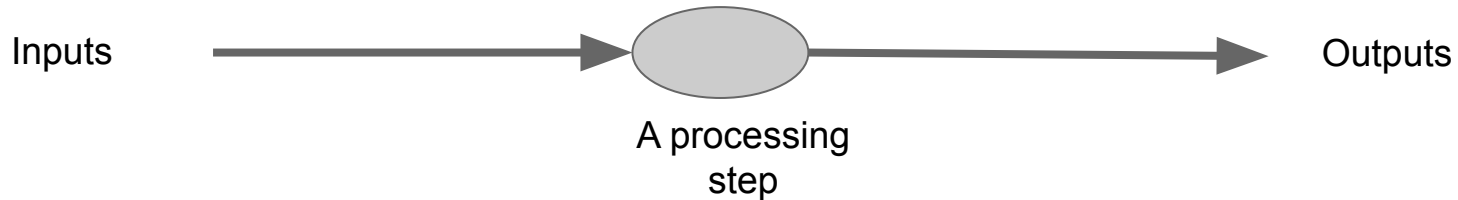
BIDS-Prov = an attempt at a software agnostic representation



# BIDS-Prov extension proposal

Key idea:

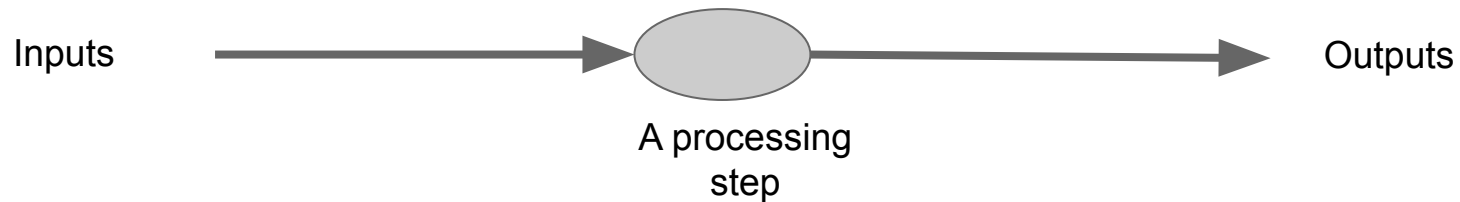
A common model to represent **chaining of processing** (for now no description of the 'meaning' of those processing)



# BIDS-Prov extension proposal

Key idea:

A common model to represent **chaining of processing** (for now no description of the 'meaning' of those processing)



BIDS-Prov files can be stored at two-levels:

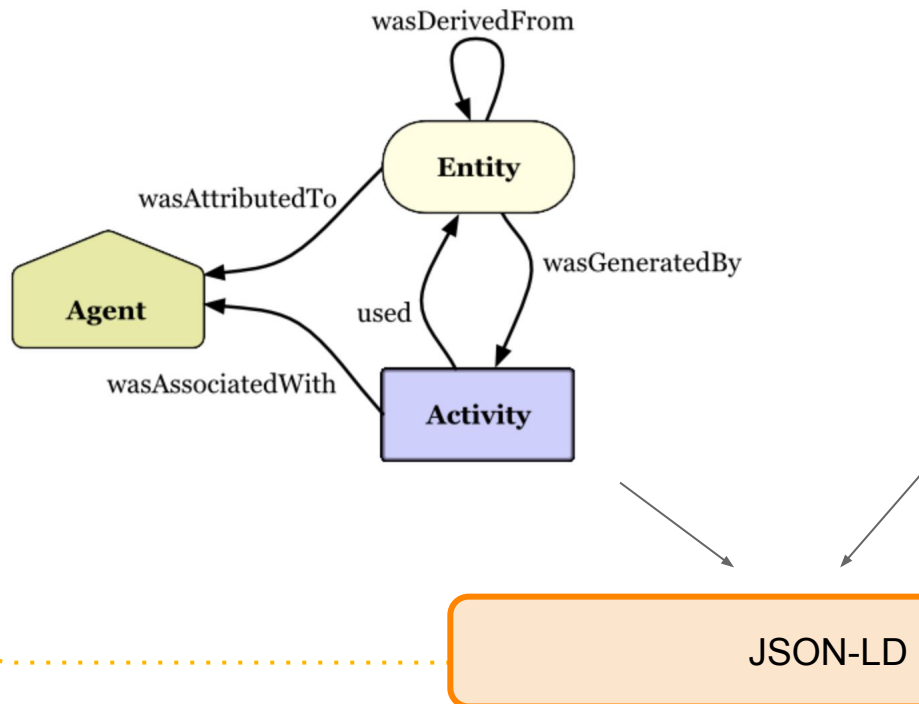
- File-level provenance `<path-to-file><file-name-with-ext>.prov.jsonld`
- Dataset-level provenance `prov/[<subfolders>*/]<label>.prov.jsonld`

# BIDS-Prov extension proposal

## RDF

BIDS-Prov uses semantic web technologies and is based on the **W3C PROV**

## BIDS



# Example

```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```

# Example

```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```

# Example

```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```

Software used

# Example

```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```

Software used

SPM

# Example

```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```

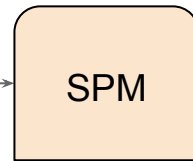
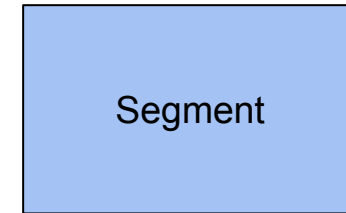
Processing  
applied ("steps")

SPM



# Example

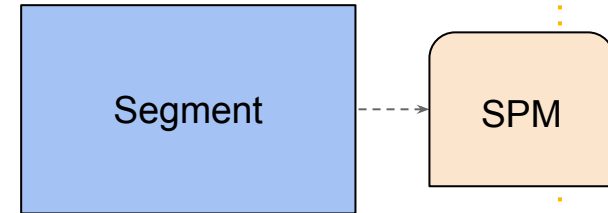
```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      {"@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224"}
    ],
    "prov:Activity": [
      {"@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }"
    ],
    "prov:Entity": [
      {"@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0"}
    ]
  }
}
```



Processing  
applied ("steps")

# Example

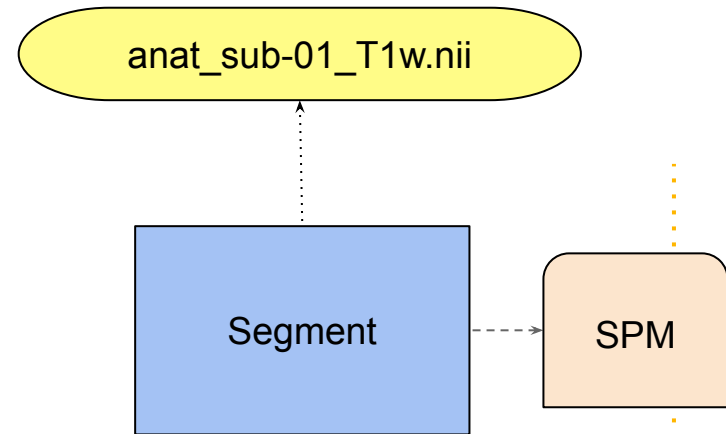
```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```



Inputs and  
outputs

# Example

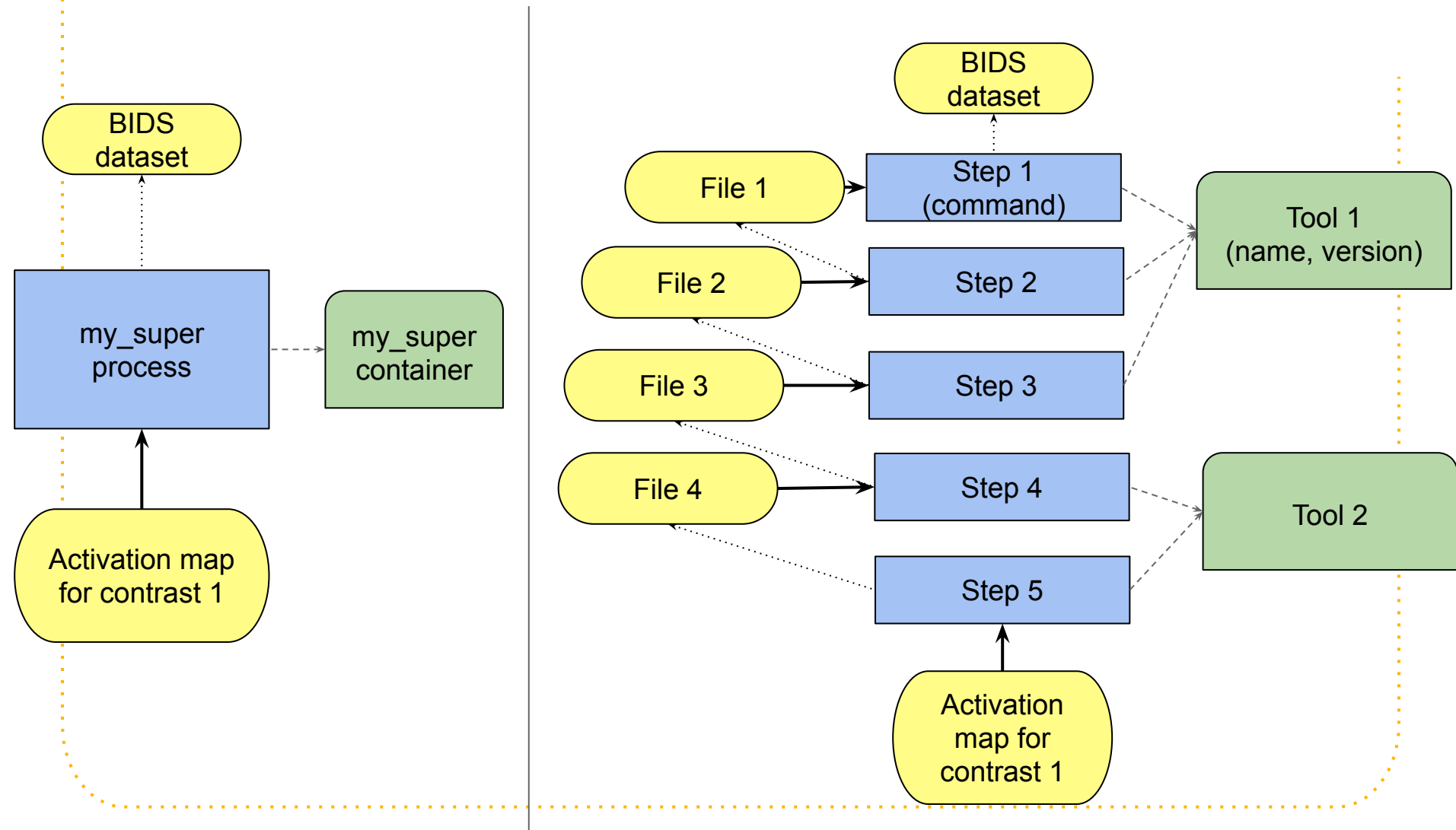
```
{ "@context": "https://raw.githubusercontent.com/bids-standard/BEP028_BIDSprov/master/context.json",
  "BIDSProvVersion": "dev",
  "records": {
    "prov:Agent": [
      { "@id": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "RRID": "RRID:SCR_007037",
        "@type": "prov:SoftwareAgent",
        "label": "SPM",
        "version": "SPM12r7224" }
    ],
    "prov:Activity": [
      { "@id": "urn:d124a1ec-ede9-4c55-9f3b-5534ba2d4bf1",
        "label": "Segment",
        "used": [
          "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6"
        ],
        "associatedWith": "urn:030f431e-c8bd-4ad2-a374-1f115df2930e",
        "command": "matlabbatch{7}.spm.spatial.preproc.channel.vols(1) = [...] }
    ],
    "prov:Entity": [
      { "@id": "urn:7d6b4ad3-03f6-4196-a307-35cc3bd624d6",
        "label": "anat_sub-01_T1w.nii",
        "generatedBy": "urn:771015e0-2dcd-4c90-b82f-1910bff41ca0" }
    ]
  }
}
```



Inputs and  
outputs

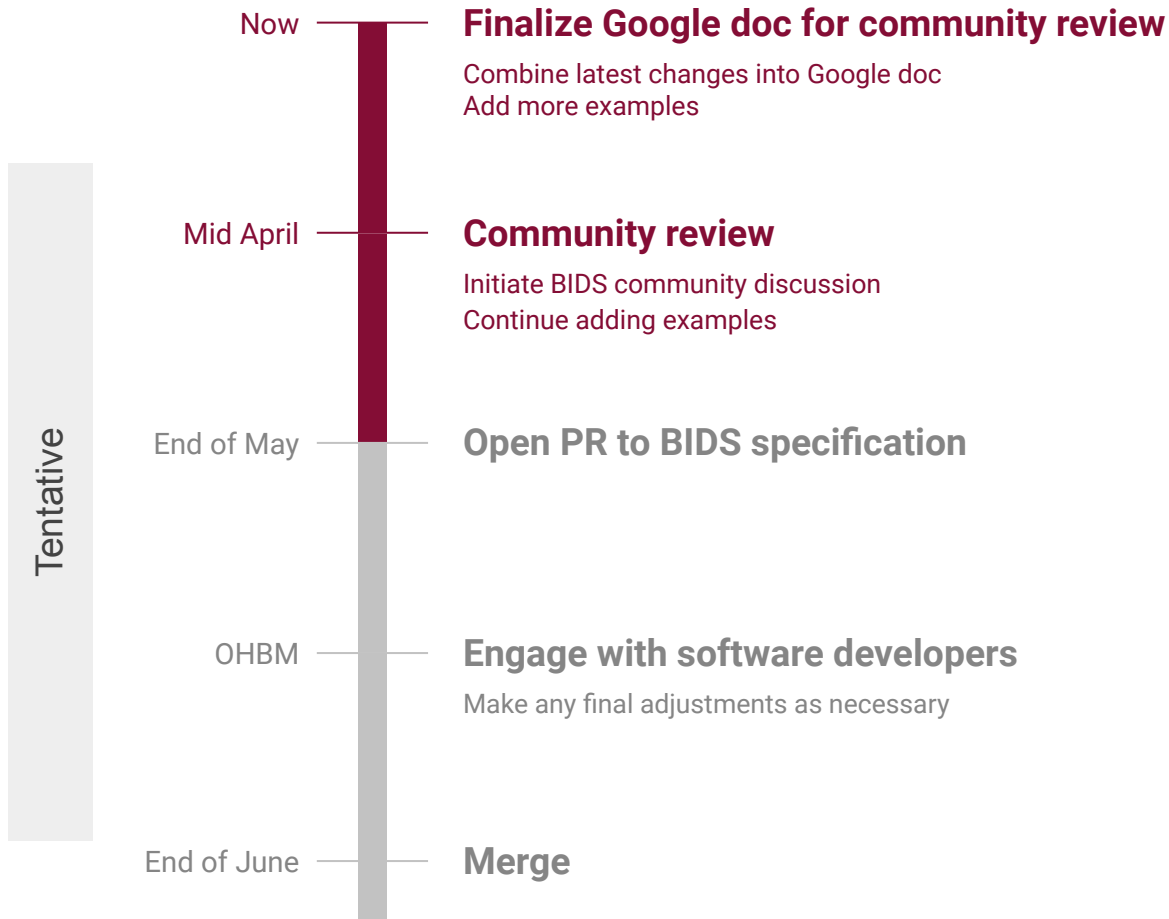


# Different "levels" of provenance



# Timeline and next steps

Presented to  
BIDS steering in  
March 2022



# BIDS-Prov extension proposal

**BIDS-Prov adds provenance information to the BIDS structure** in the form of sidecar JSON-LD files, which constitute a graph of provenance information.

Open questions / thoughts:

- Some provenance is already in BIDS
- Different levels of provenance will enable different use cases
- Ability to keep track of provenance might depend on the application and the context (which metadata is available)
- Satra : Possibility to implement via message passing

Workshop BIDS-derivatives  
June 21-23, 2023



# BIDS-Prov

Recording neuroimaging provenance

Camille Maumet



# Thank you!



[@cmaumet@fediscience.org](mailto:cmaumet@fediscience.org)

Credit: Presentation template and icons by SlidesCarnival, adapted



[@cmaumet](https://twitter.com/cmaumet)