

# Extending NI-DM to share the results and provenance of a neuroimaging study: an example with SPM

Camille Maumet<sup>1</sup>, Thomas Nichols<sup>2</sup>, B. Nolan Nichols<sup>3</sup>, Guillaume Flandin<sup>4</sup>, Jessica Turner<sup>5</sup>, Karl Helmer<sup>6</sup>, Jason Steffener<sup>7</sup>, Jean-Baptiste Poline<sup>8</sup>, Satrajit Ghosh<sup>9</sup>, David Keator<sup>10</sup>

1. Warwick Manufacturing Group, University of Warwick, Coventry, UK; 2. Dept. of Statistics and Warwick Manufacturing Group, University of Warwick, Coventry, UK; 3. Integrated Brain Imaging Center, University of Washington, Seattle, WA, USA; 4. Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK; 5. Department of Neurology, Columbia University, Atlanta, USA; 6. Martinos Center for Biomedical Imaging, Massachusetts General Hospital; Dept. of Radiology, Boston, MA, USA; 7. Department of Neurology, Columbia University, New York, USA; 8. Helen Wills Neuroscience Institute, BIC, University of California, Berkeley, CA, USA; 9. McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, 10Department of Psychiatry and Human Behavior, Dept. of Computer Science, University of California, Irvine, CA, USA

**incf** International Neuroinformatics Coordinating Facility

**Acknowledgments:** We would like to acknowledge the work of all the INCF task force members as well as of many other colleagues who have helped the task force. We are particularly indebted to Mathew Abrams, Linda Lanyon, Roman Valls Guimera and Sean Hill for their support at the INCF. Further we acknowledge the long-standing support of Derived Data Working Group activities by the BIRN coordinating center (NIH 1 U24 RR025736-01), and the Wellcome Trust for support of CM & TEN.

## Introduction

While data sharing is prevalent, if not mandatory, in other disciplines, **data sharing in neuroimaging** is largely still confined to a table of coordinates in a published paper.

Routine sharing of imaging data (e.g., 3D "activation" volumes, raw/pre-processed data and associated metadata) would facilitate **reproducibility/replication studies, meta-analyses** and eventually support **new discoveries** in neuroscience or medicine [2].

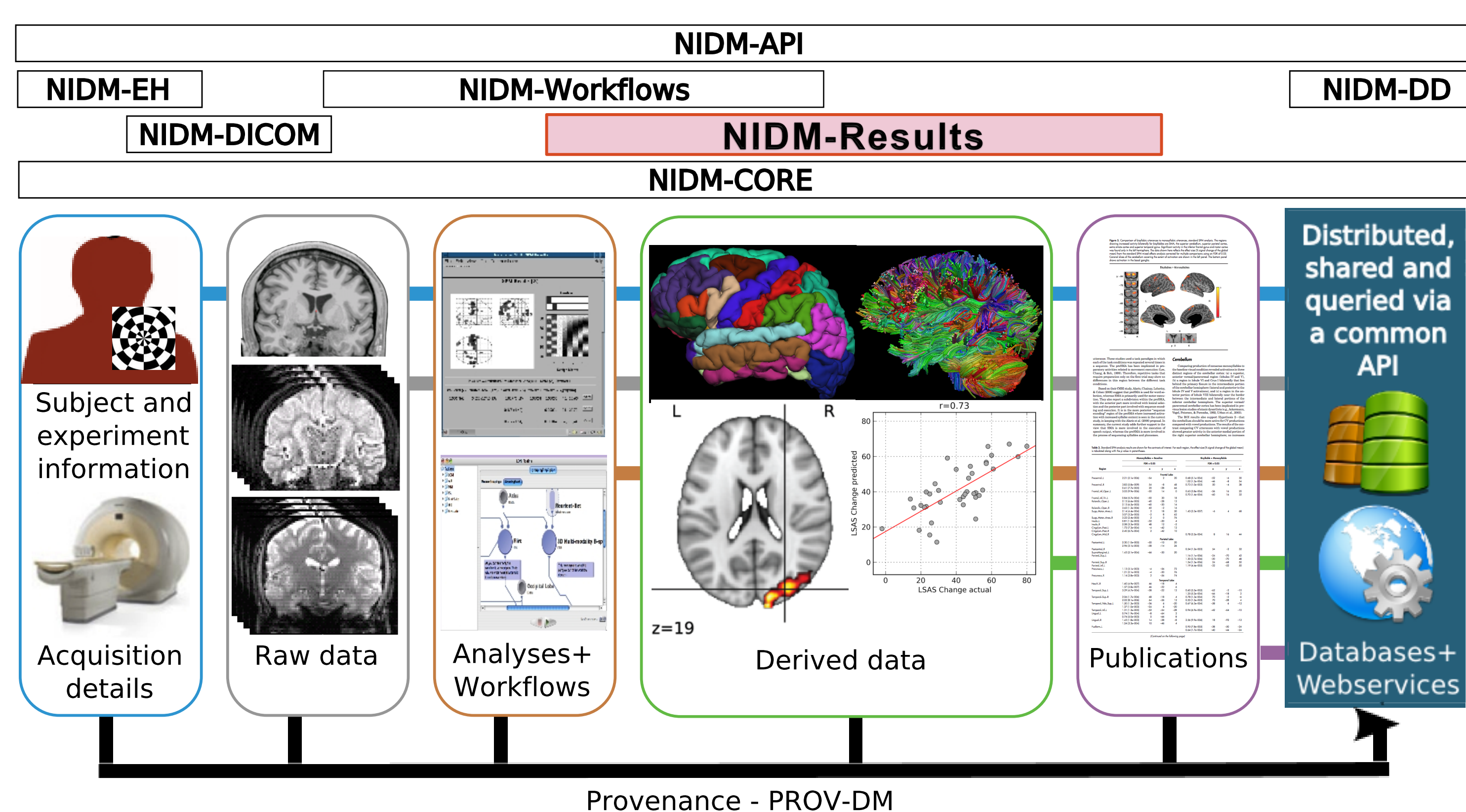


Fig. 1: Overview of a neuroimaging experiment workflow and NI-DM components.

In [3], we introduced the **Neuroimaging Data Model (NI-DM)**, a domain-specific extension of the recently-approved W3C recommendation, **PROV-DM** [6].

Here, we **extend NI-DM** with **NIDM-Results** to **model the results of statistical parametric mapping studies**, such as fMRI brain mapping results, and their provenance (from model fitting to inference). We also provide a publicly-available specification of a domain-specific object model for the SPM software [7].

## Methods

Through a weekly video conferences and focused workshops with experts in neuroimaging data analysis, we defined a recommended **minimal set of neuroimaging metadata** to be reported for **SPM analyses**. Each metadata term was assigned a **Uniform Resource Identifier (URI)** and **definition** to unambiguously represent a given concept.

Working with the SPM developers, we created an **object model** that balances complexity with expressivity and developed a **formal specification** document following the guidance of the W3C.

## Results

A **lexicon of terms** was defined to capture general brain mapping analysis terms and, when appropriate terms were not found in existing terminologies, SPM-specific terms. Using the defined terms, the SPM results were modeled as PROV **entities**, **activities**, and **agents** with corresponding properties and attributes.

Figure 1 shows a simplified graph of the proposed data model and Figure 2 presents an example entity describing a cluster-statistic.

An automatic RDF serialisation was developed within the SPM software to implement **NI-DM as part of the official SPM12 release**.

## SPM Data Model: Overview

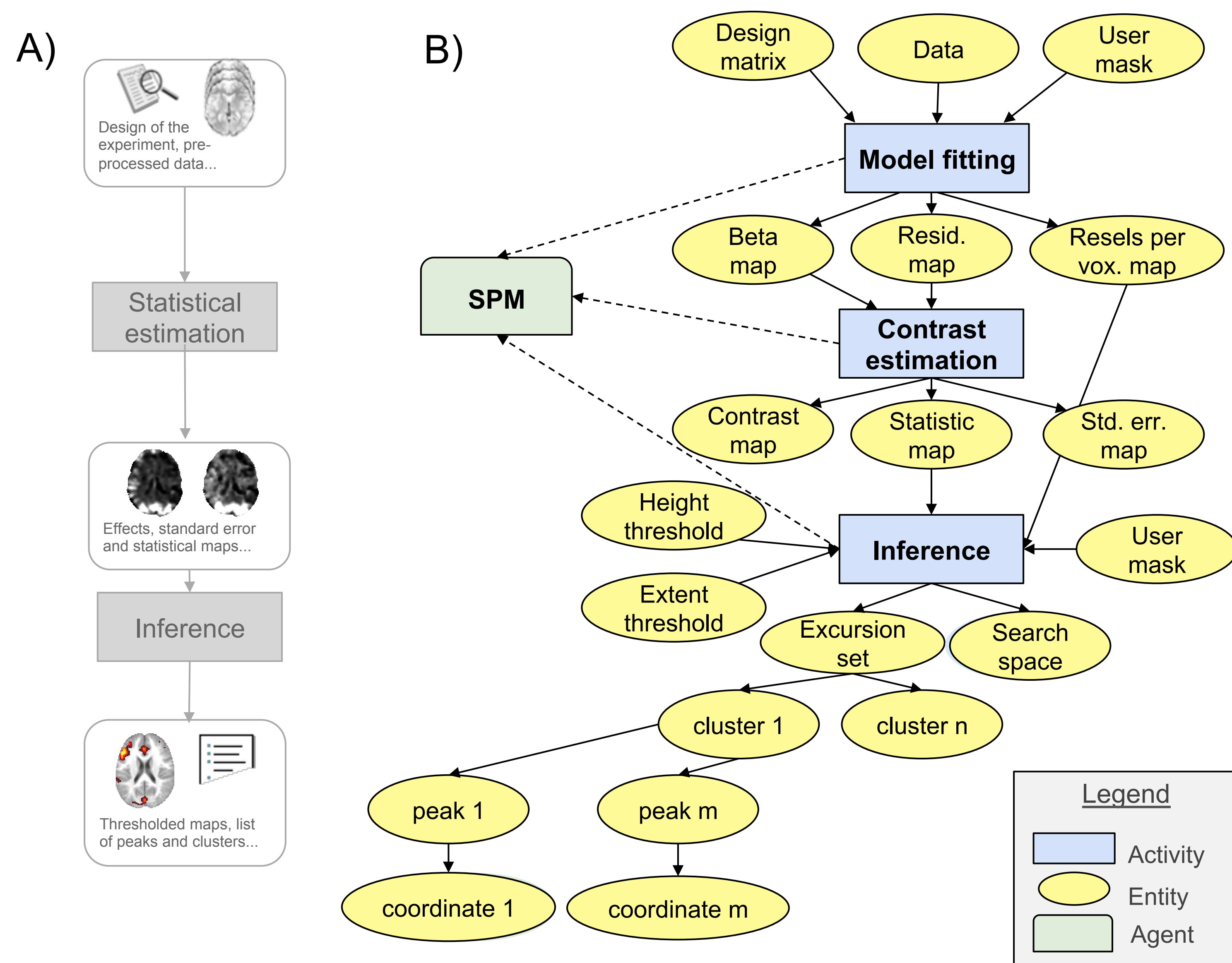
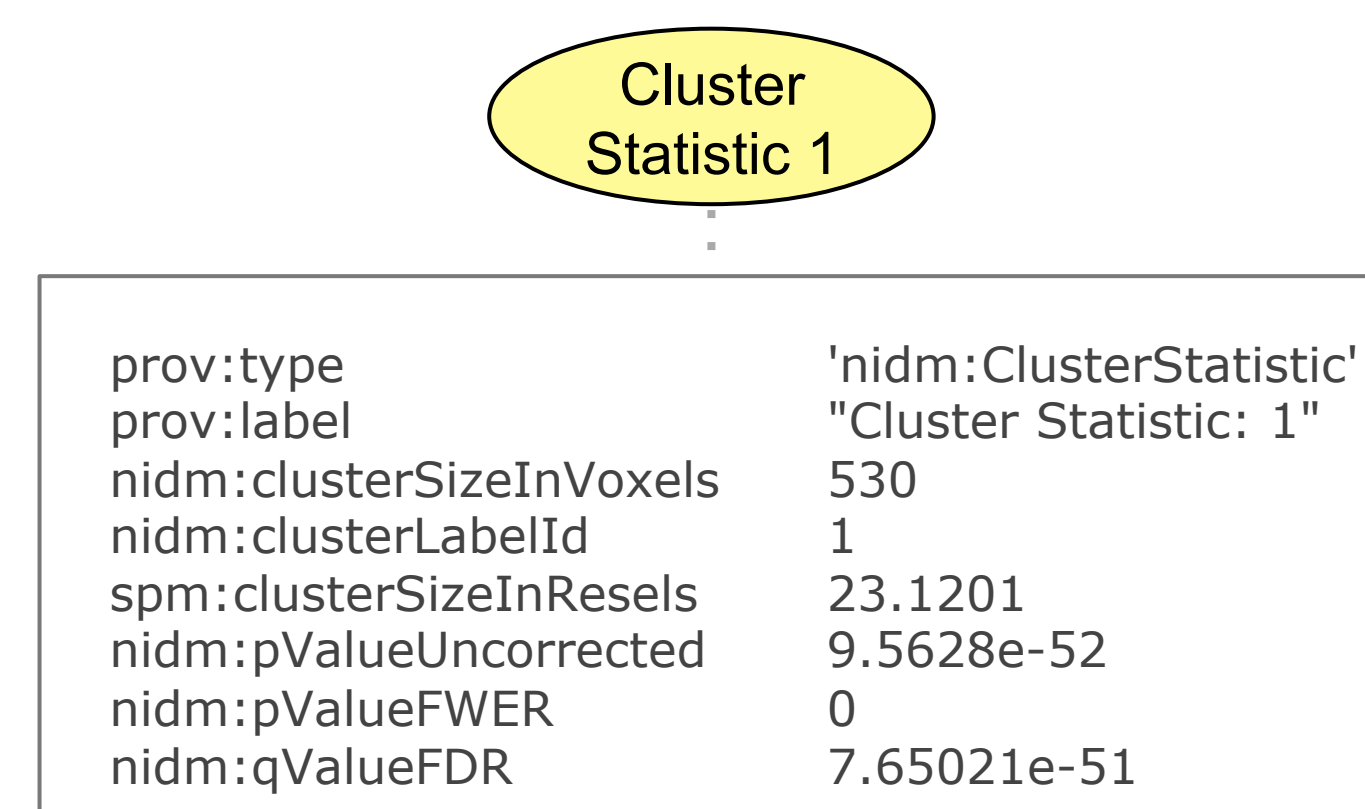


Fig. 2: A) Conceptual overview of the SPM statistical estimation and inference workflow. B) Overview of the proposed NI-DM extension which models the results of a neuroimaging study in SPM including entities (yellow), activities (blue) and agent (green).

Full specification available at: <http://nidm.nidash.org/specs>

## Example of ClusterStatistic entity

Graphical representation:



PROV-N notation:

```
entity(niiri:cluster_0001,
[
  prov:type = 'nidm:ClusterStatistic',
  prov:label = "Cluster Statistic: 1",
  nidm:clusterSizeInVoxels = "530",
  nidm:clusterLabelId = "1",
  spm:clusterSizeInResels = "23.1201",
  nidm:pValueUncorrected = "9.5628e-52",
  nidm:pValueFWE = "0",
  nidm:qValueFDR = "7.65021e-51"
])
```

SPARQL query: "Find cluster entities with p<0.05 FWE corrected"

```
SELECT ?cluster_id ?pvalue
WHERE {
  ?cluster_id a nidm:ClusterStatistic ;
  nidm:pValueFWE ?pvalue .
  FILTER (?pvalue < 0.05)
}
```

More queries at: <http://nidm.nidash.org/specs>

## Conclusion

We have used NI-DM to represent brain mapping statistical results in SPM. Future work consists of publishing the neuroimaging terms in a public index and repeating the process for FSL [8] and AFNI [9]. We envision this work to ease the process of submitting neuroimaging results to journals during publication and facilitate meta-analyses.

## References

- [1] Button et al. (2013). Nature reviews. Neuroscience, 14(5), 365–76.
- [2] Poline et al. (2012). Frontiers in Neuroinformatics. 6:9.
- [3] Keator et al. (2013). Neuroimage. 15;82:647-61
- [4] Nichols et al. (2013) OHBM.
- [5] Nichols et al. (2013). Neuroinformatics 2013.
- [6] [www.w3.org/TR/prov-dm/](http://www.w3.org/TR/prov-dm/)
- [7] [www.fil.ion.ucl.ac.uk/spm/](http://www.fil.ion.ucl.ac.uk/spm/)
- [8] <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>
- [9] <http://afni.nimh.nih.gov>