Bridging the gap between brain imaging standards: towards federated queries in life sciences

Stage Ingénieur / Master - 2019

Supervisors

Camille Maumet, Research scientist at Inria, VisAGeS team, Rennes, France
camille.maumet@inria.fr, http://camillemaumet.com

Alban Gaignard, Research engineer at CNRS, Institut du Thorax, Nantes, France
alban.gaignard@univ-nantes.fr, http://albangaignard.github.io

Yao Chi, Research engineer at Inria, VisAGeS team, Rennes, France
yao.yao@inria.fr

Location:
Inria / IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex, France

Duration: 2-6 months

Keywords
Brain Imaging, Standardisation, Semantic Web, Interoperability, Data integration.

Context
Precision medicine aims at tailoring medical treatments to the individual characteristics of each patient. It typically consists in selecting appropriate and optimal therapies based on a combination of the patient's genetic, molecular, cellular and phenotypic profile. However, reconciling the diseases complexity with patient-specific data requires an integrated approach and remains a major open challenge for data science.

There are currently more than 1500 complementary life science reference databases. But, most of these databases have been designed independently, have heterogeneous schemas and rely on technologies that do not support their interoperability (Gomez-Cabrero et al. 2014).

Over the last decade, Semantic Web technologies have established a relevant framework for addressing both interoperability and scalability issues (Shadbolt, Berners-Lee, and Hall 2006). It supports the creation of persistent identifiers for referencing data and has been widely adopted in the Life science community. It also evolved into the FAIR principles for ensuring that the available data are Findable, Accessible, Interoperable and Reusable (Wilkinson et al. 2016).

This internship is part of the "RDF Datahub for Precision Medicine" project which aims at studying the best strategies for querying multiple datasets efficiently, in imaging and whole genome sequencing. The project involves three computer science teams (Dyliss Rennes, VisAGeS Rennes, LS2N Nantes) and three medical teams (Institut du Thorax Nantes, MicMac Rennes and CHU Rennes bioinformatics unit).

More specifically, the goal of this internship is to build a semantic web representation for brain imaging datasets by leveraging existing standards and taking an active role in ongoing international initiatives.

Detailed description
The successful applicant will interact with two international standardisation efforts from the neuroimaging community:

- NIDM: a standard based on Semantic Web technologies (Keator et al. 2013; Maumet et al. 2016). While those technologies are particularly well suited for aggregation across complex datasets, using them can be costly in terms of initial development time.

- BIDS: a standard based on file naming convention and JSON files (Gorgolewski et al. 2016). BIDS has been particularly well-accepted by the community and adoption is growing steadily with more than 20 000 open BIDS-compliant datasets available. While those data constitute a very rich resource, due to the very nature of the BIDS standard, using them in conjunction with other resources requires specific developments and manual metadata alignments.

Here, we propose to get the best of both worlds and extend BIDS with Linked Data principles. To bridge the two standards, the successful candidate will use JSON-LD ("JSON-LD 1.1" n.d.) (a JSON representation for semantic graphs) and extend ongoing work (Maumet et al. 2019).

This work will be applied to ICAN (Bourcier et al. 2017), a nation-wide collaborative project aimed at better understanding the development and the evolution of intracranial aneurysms. More than 30 French university hospitals have collected around 30 000 medical imaging datasets for a total of approximately 3 000 participants. The enriched BIDS model will be used to filter medical image candidates to be processed (i.e. select available brain images matching specific criteria) but also to export derived information on the quantification of the brain bifurcation network.

The selected candidate will:

1. Bridge the gap between BIDS and NIDM to provide a semantically-enriched version of BIDS.
2. Interact with the international BIDS community to get feedback on the suggested updates.
3. Attend weekly conference calls with the NIDM working group to get feedback on suggested updates.
4. Apply this work to the ICAN project.

Required skills

- Interest (even better knowledge) in standardisation in general and in interoperability.
- Interest in medical imaging
- Good communication skills
- Good programming skills in Python
- Very good understanding of English

Preferred skills

- Knowledge of JSON (or even better JSON-LD)
- Knowledge of semantic web technologies

References

Bourcier, Romain, Stéphanie Chatel, Emmanuelle Bourcereau, Solène Jouan, Hervé Le Marec, Benjamin Daumas-Duport, Mathieu Sevin-Allouet, et al. 2017. "Understanding the Pathophysiology of Intracranial Aneurysm: The ICAN Project." Neurosurgery 80 (4): 621–26.

Gomez-Cabrero, David, Imad Abugessaisa, Dieter Maier, Andrew Teschendorff, Matthias Merkenschlager, Andreas Gisel, Esteban Ballestar, Erik Bongcam-Rudloff, Ana Conesa, and Jesper Tegnér. 2014. "Data Integration in the Era of Omics: Current and Future Challenges." BMC Systems Biology 8 Suppl 2 (March): I1.

Gorgolewski, Krzysztof J., Tibor Auer, Vince D. Calhoun, R. Cameron Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin, et al. 2016. "The Brain Imaging Data Structure, a Format for Organizing and Describing Outputs of Neuroimaging Experiments." Scientific Data 3 (June): 160044.

"JSON-LD 1.1." n.d. Accessed January 7, 2019. https://json-ld.org/spec/latest/json-ld/.

Keator, D. B., K. Helmer, J. Steffener, J. A. Turner, T. G. M. Van Erp, S. Gadde, N. Ashish, G. A. Burns, and B. N. Nichols. 2013. "Towards Structured Sharing of Raw and Derived Neuroimaging Data across Existing Resources." NeuroImage 82 (November): 647–61.

Maumet, Camille, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, et al. 2016. "Sharing Brain Mapping Statistical Results with the Neuroimaging Data Model." Scientific Data 3 (160102). https://doi.org/10.1038/sdata.2016.102.

Maumet, Camille, Satrajit Ghosh, Yaroslav Halchenko, Jarecka Dorota, Nolan Nichols, Jean-Baptiste Poline, and Michael Hanke. 2019. "The Best of Both Worlds: Using Semantic Web with JSON-LD. An Example with NIDM-Results & Datalad". Preprint, available at: https://www.hal.inserm.fr/inserm-01972649.

Shadbolt, Nigel, Tim Berners-Lee, and Wendy Hall. 2006. "The Semantic Web Revisited." IEEE Intelligent Systems 21 (3): 96–101.

Wilkinson, Mark D., Michel Dumontier, I. Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3 (March): 160018.