

# INEX-MED: INtegration and EXploration of heterogeneous bio-MEDical data

Kirsley CHENNEN<sup>1,2,3,\*</sup>, Maxime FOLSCHETTE<sup>1,4,\*</sup>, Alban GAIGNARD<sup>5</sup>, Richard REDON<sup>5</sup>, Hala SKAF-MOLLI<sup>4</sup>, Olivier POCH<sup>2</sup>, Jocelyn LAPORTE<sup>3</sup>, Julie THOMPSON<sup>2</sup> and the INEX-MED CONSORTIUM

<sup>1</sup> Institut Français de Bioinformatique, CNRS UMS 3601, France

<sup>2</sup> Equipe CSTB - Laboratoire iCUBE, CNRS UMR 7357, Université de Strasbourg, France

<sup>3</sup> Dpt Médecine translationnelle et neurogénétique, IGBMC, INSERM U1258, CNRS UMR7104, Illkirch, France

<sup>4</sup> LS2N, Laboratoire des Sciences du Numérique de Nantes, CNRS UMR 6004, Université de Nantes, France

<sup>5</sup> Institut du Thorax, Inserm UMR 1087, CNRS UMR 6291, Université de Nantes, France

Corresponding author: [kchennen@unistra.fr](mailto:kchennen@unistra.fr), [maxime.folschette@ls2n.fr](mailto:maxime.folschette@ls2n.fr), [alban.gaignard@univ-nantes.fr](mailto:alban.gaignard@univ-nantes.fr)

## Abstract

The new era of modern biology and medicine calls for the development of novel integrated approaches leveraging massive multi-disciplinary, multi-scale, and multi-modal heterogeneous biomedical data. These biomedical datasets (ex: clinical, imaging, or “omic”) are currently stored in data silos [1], which makes their cross-exploitation challenging.

Here, we present INEX-MED, a unified Knowledge-Graph based framework that aims to enable systematic integration of diverse data modalities and clinical observations to accelerate both the statistical/semantic data exploitation of biomedical big data with diverse clinical purposes. Following the “FAIR” data principles (Findability, Accessibility, Interoperability, Reusability) [2], we propose a prototype system that (i) integrates clinical, imaging and genomics data from cohorts into a dedicated knowledge-graph, (ii) allows secure access and integrated query processing on heterogeneous biomedical data, and (iii) performs statistical analysis and machine learning to improve the diagnosis/prognosis of studied diseases.

The INEX-MED prototype is currently being developed for two use-case cohorts:

- (i) The ICAN cohort [3] covering 3000 individuals affected by intracranial aneurysm for which clinical records, MRI imaging and exome sequencing data are acquired. The aim is to identify biomarkers and risk factors characterising the development of this disorder.
- (ii) The MYO-lico cohort with 1200 congenital myopathy patients having clinical records, histopathological imaging data, and exome sequencing data. The aim is to identify novel genes causing congenital myopathies and classify them [4].

INEX-MED already benefits from the Cloud infrastructure provided by the French Bioinformatics Institute (IFB). As a result of this project, the developed prototype will be made available to the biomedical community as an IFB resource, providing methods and technological guidelines to address other biomedical use cases.

## References

- [1] Akram Alyass, Michelle Turcotte, and David Meyre. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Medical Genomics*, 8(1):33, Jun 2015.
- [2] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 2016.
- [3] Romain Bourcier, Stéphanie Chatel, Emmanuelle Bourcereau, Solène Jouan, Hervé Le Marec, Benjamin Daumas-Duport, Mathieu Sevin-Allouet, Benoit Guillon, Vincent Roualdes, Tanguy Riem, et al. Understanding the pathophysiology of intracranial aneurysm: The ICAN Project. *Neurosurgery*, 80(4):621–626, 2017.
- [4] J Böhm, R Schneider, E Malfatti, V Schartner, X Lornage, I Nelson, G Bonne, B Eymard, J Nectoux, F Leturcq, et al. Integrated analysis of the large-scale sequencing project “Myocapture” to identify novel genes for myopathies. *Neuromuscular Disorders*, 27:S195, 2017.

---

\*. These authors contributed equally to this work