



Integration and provenance control of proteomics data using SWOMed, a Product Lifecycle Management framework for biomedical research

Amel Raboudi, Marianne Allanic, Pierre-Yves Hervé, Daniel Balvay, Joevin Sourdon, Philippe Boutinaud, Bertrand Tavitian

► To cite this version:

Amel Raboudi, Marianne Allanic, Pierre-Yves Hervé, Daniel Balvay, Joevin Sourdon, et al.. Integration and provenance control of proteomics data using SWOMed, a Product Lifecycle Management framework for biomedical research. SMMAP2017, Oct 2017, Marne la vallée,, France. hal-01654383

HAL Id: hal-01654383

<https://hal.science/hal-01654383>

Submitted on 3 Dec 2017

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.

Integration and provenance control of proteomics data using SWOMed, a Product Lifecycle Management framework for biomedical research

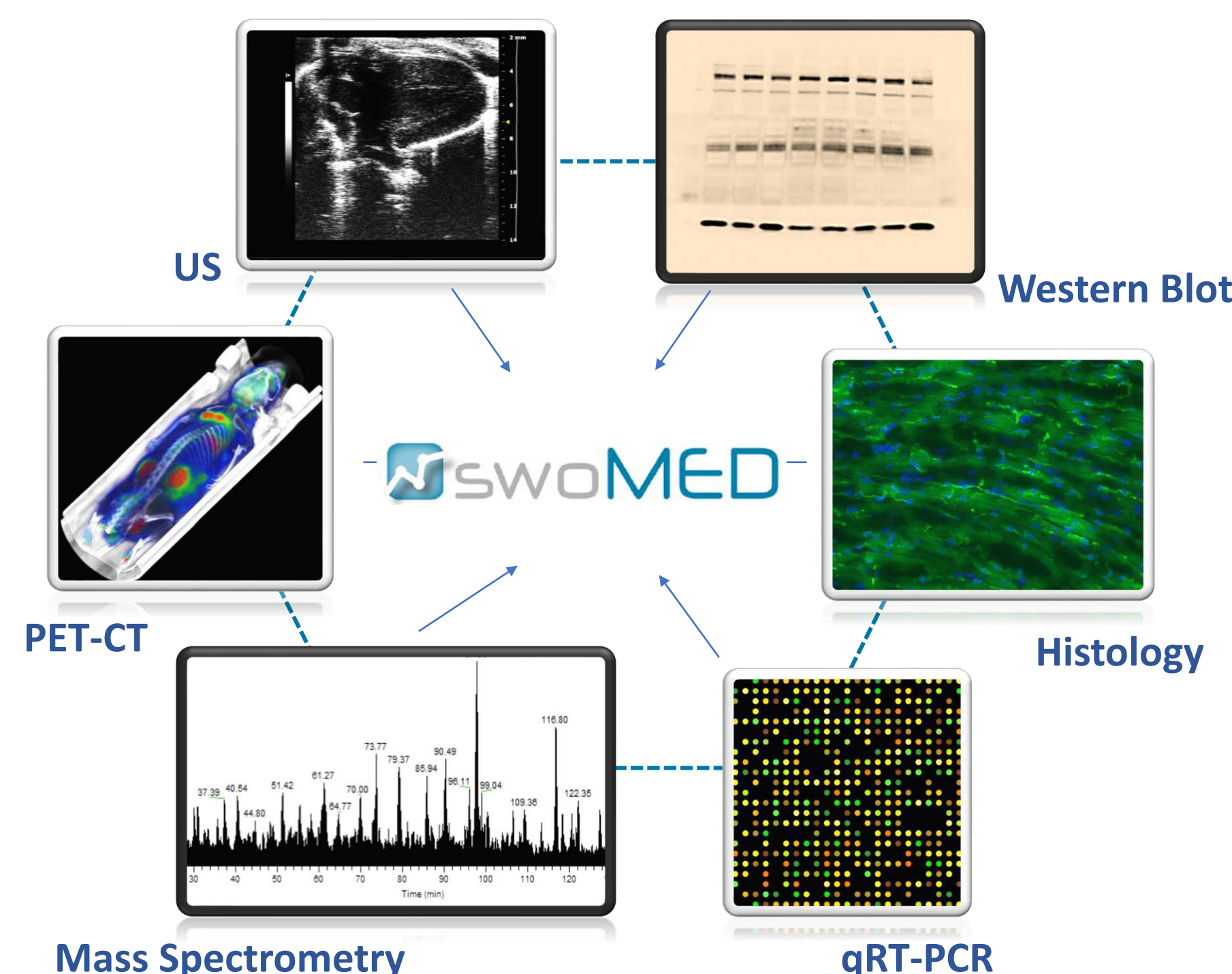
AMEL RABOUDI^{1,2,3}, **MARIANNE ALLANIC**¹, **PIERRE-YVES HERVÉ**¹, **DANIEL BALVAY**^{2,4}, **JOEVIN SOURDON**^{2,4}, **PHILIPPE BOUTINAUD**¹, **BERTRAND TAVITIAN**^{2,4,5}

1. FEALINX, 37 rue Adam Ledoux 92400 Courbevoie, France 2. INSERM, UMR970, Paris-Cardiovascular Research Center at HEGP, Paris, France
3. Université de Technologie de Compiègne (UTC), Roberval Laboratory, Compiègne, France 4. Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, F-75006 Paris, France
5. Department of Radiology, Georges Pompidou European Hospital, Paris, France

Context

Because of the complexity of living organisms, biomedical research makes use of multiple data sources from multiple instruments, techniques and protocols, e.g. various in vivo and in vitro imaging techniques, various omics methods, physiology, pharmacology, etc. Presently, there is a **lack of tools to integrate efficiently multiple heterogeneous biomedical data** and exploit their significance for addressing specific research issues.

Product Lifecycle Management (PLM) was developed by the industry to provide collaborative, secure, and reliable tools for industrial manufacturing. It provides **traceability, versioning, strict access rights and data integrity to complex data from multiple sources in multiple formats.**



SWOMed is a biomedical PLM system, recently developed during the interdisciplinary research project BIOMIST (ANR-13-CORD-0007). It provides a collaborative framework for biomedical data lifecycle management, with a focus on cohort imaging and human cognitive neuroscience studies (Allanic et al. [1]), but was not tested in the context of an experimental preclinical study incorporating proteomics data.

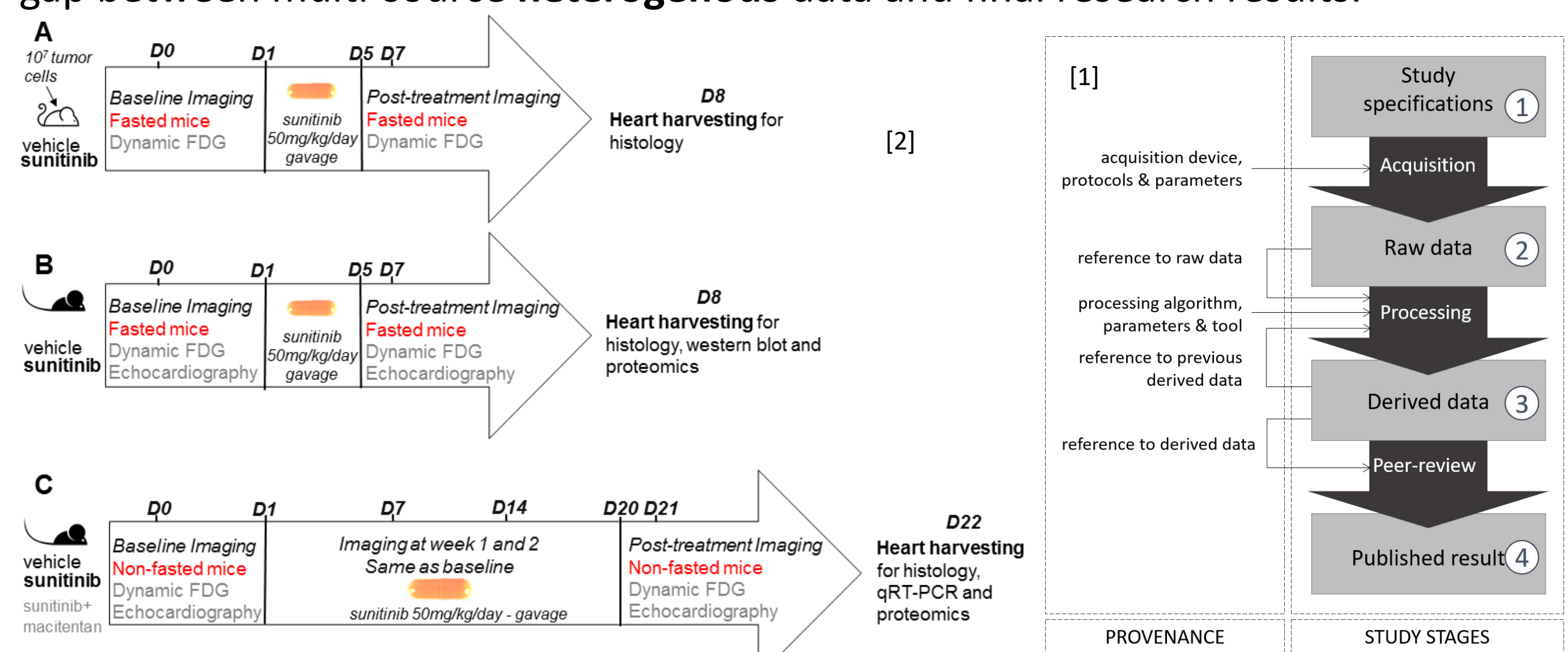
[1] M. Allanic *et al.*, « BIOMIST: A Platform for Biomedical Data Lifecycle Management of Neuroimaging Cohorts », *Front. ICT*, vol. 3, janv. 2017.

Objective

Our main objective is to integrate **proteomics** and experimental multimodal preclinical studies using SWOMed. Specific objectives are to guarantee research data quality, improve data sharing and collaboration, ensure reproducibility and reuse of **heterogenous** study data.

Case study

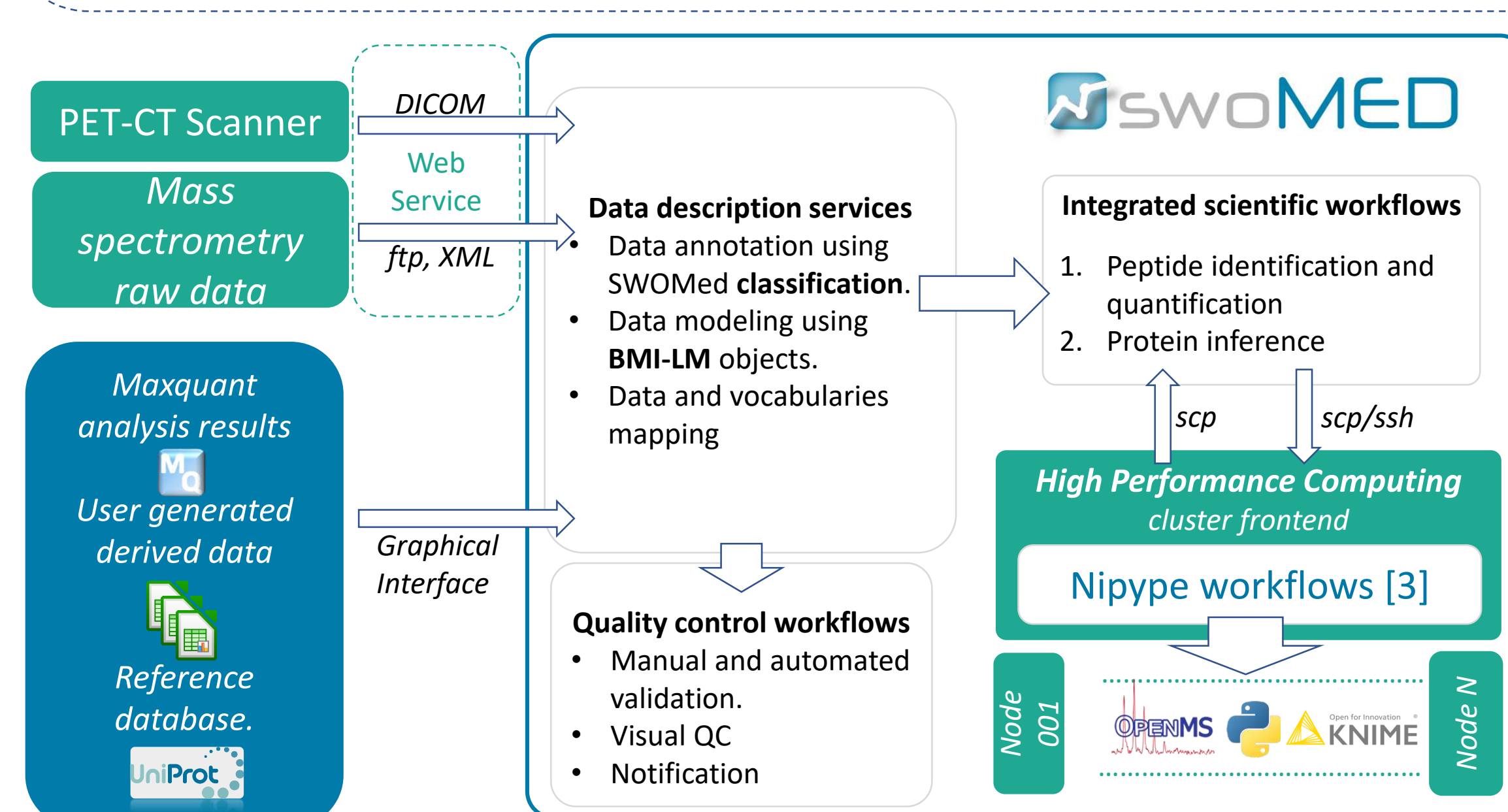
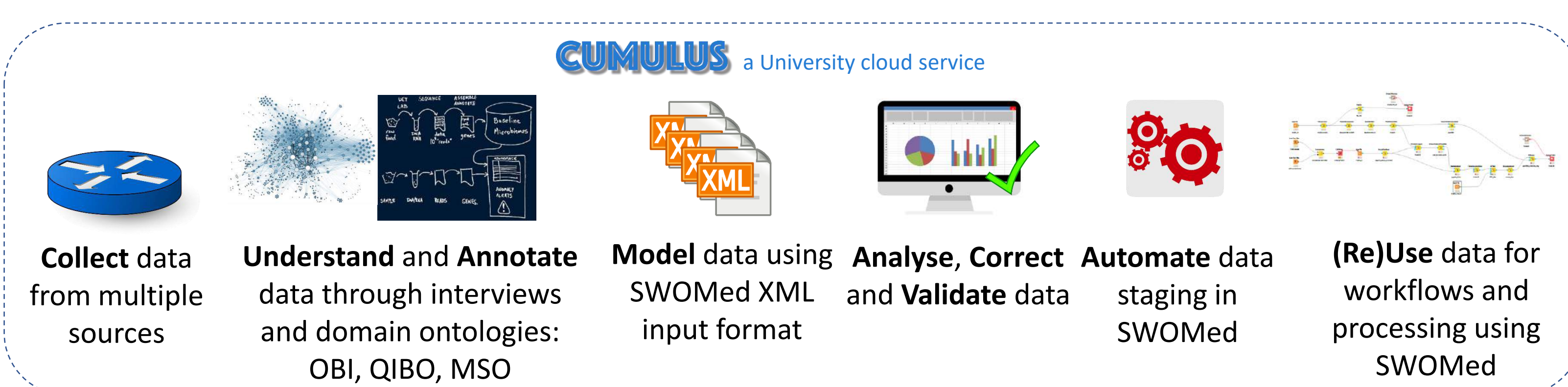
We adapted the generic data model (BMI-LM) of SWOMed to the needs of **DRIVE-SPC** (Déploiement du Réseau d'Images du Vivant de Sorbonne Paris Cité), a joint project of PARCC-Inserm laboratory and Fealinx company aiming at bridging the gap between multi-source **heterogenous** data and final research results.



Our first use case is an experimental **cardiotoxicity** study combining proteomics, histology and two imaging modalities (Positron emission tomography and cardiac ultrasound) results with the aim to understand the mechanisms underlying the cardiotoxicity of an anti-angiogenic anticancer treatment in mice [2].

[2] J. Sourdon *et al.*, « Cardiac Metabolic Deregulation Induced by the Tyrosine Kinase Receptor Inhibitor Sunitinib is rescued by Endothelin Receptor Antagonism », *Theranostics*, vol. 7, n° 11, p. 2757-2774, 2017.

Materials and Methods

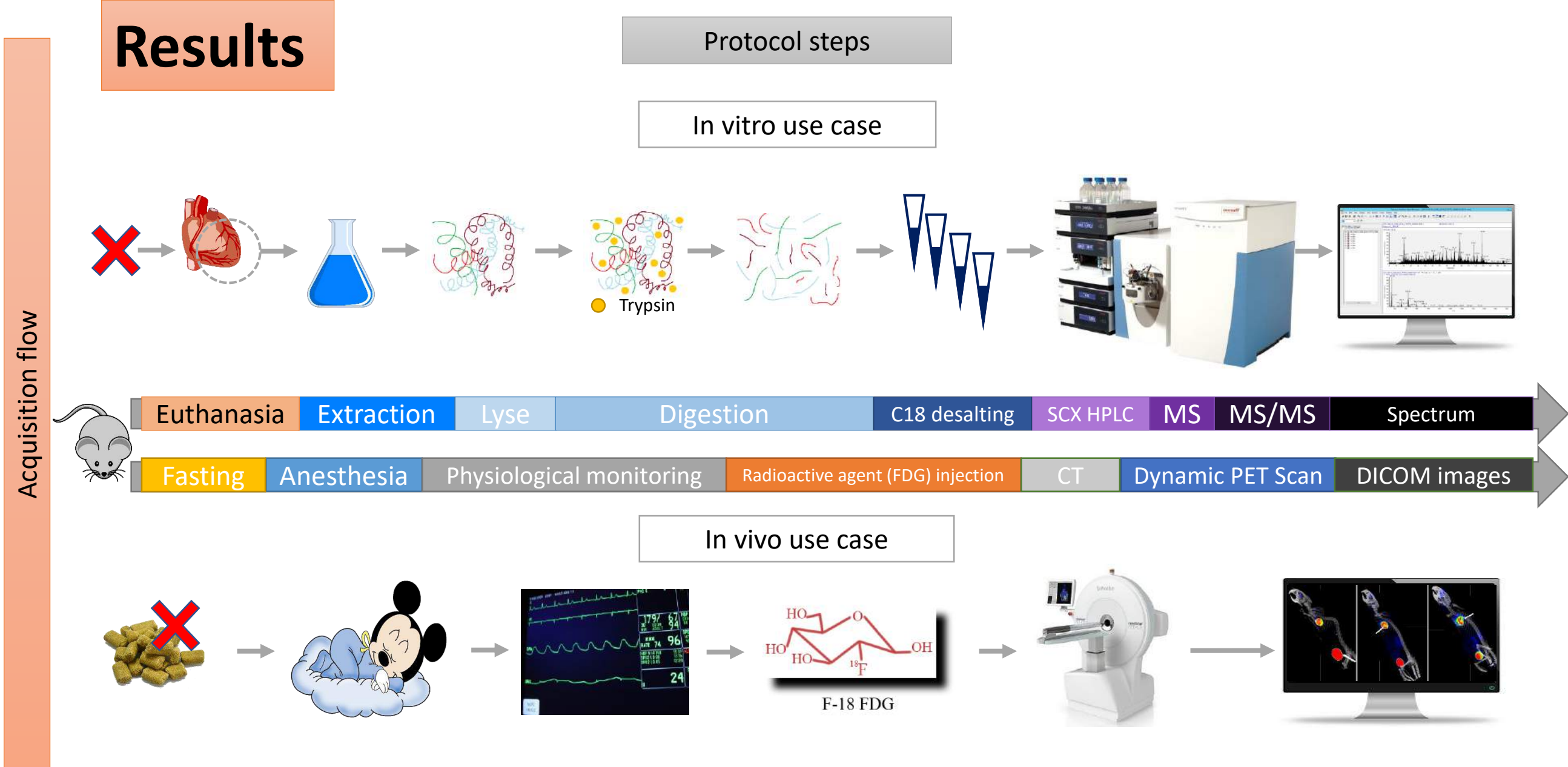


[3] K. Gorgolewski *et al.*, "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python," *Front. Neuroinformatics*, vol. 5, 2011.

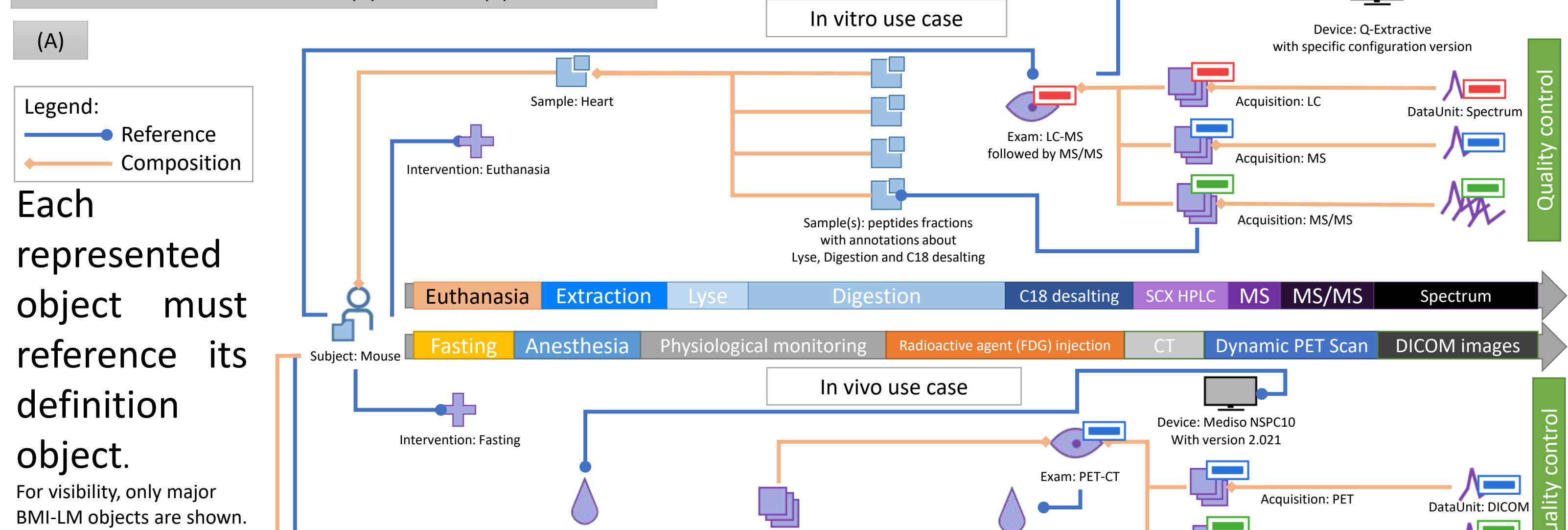
Features

- Traceability
- Provenance
- Versioning
- Multisite studies
- Strict access rights
- Access to previous research data
- Integrated workflows
- FAIR guidelines
- Comprehensive metadata
- Use of ontologies

Results



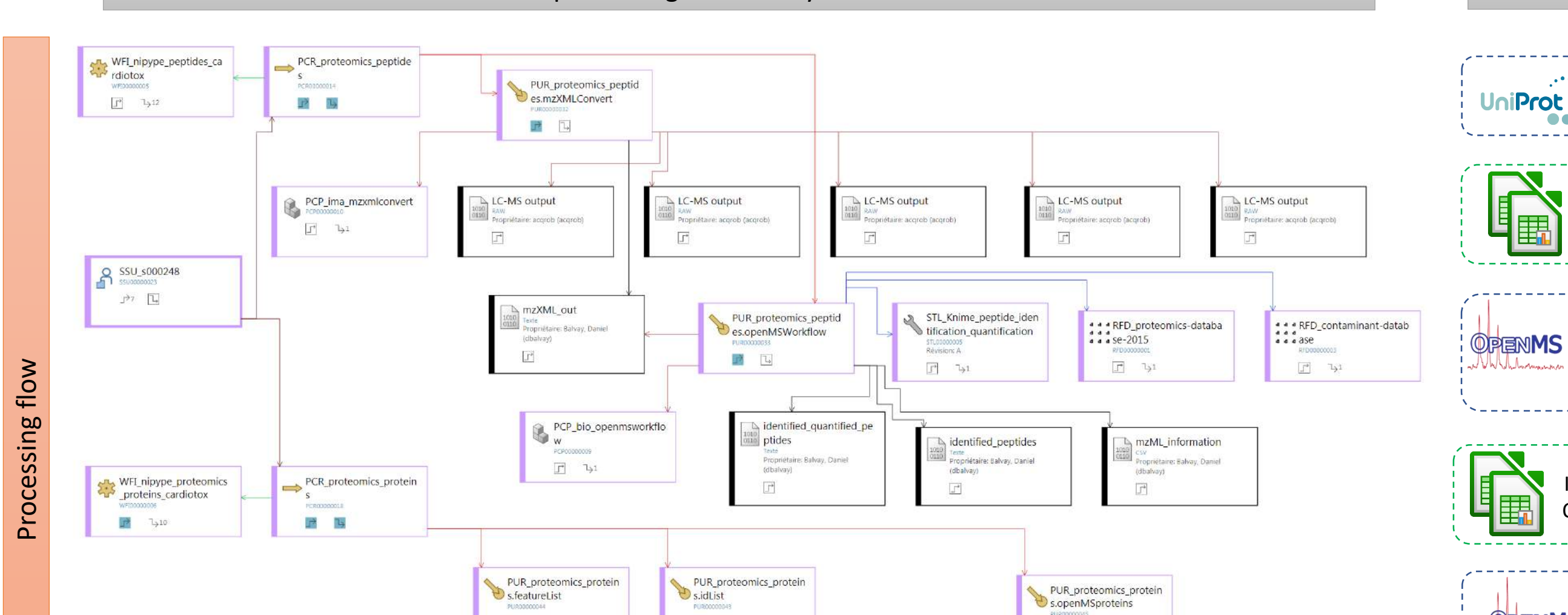
Protocol data model in SWOMed. (A) Raw data. (B) Derived data.



Each represented object must reference its definition object.

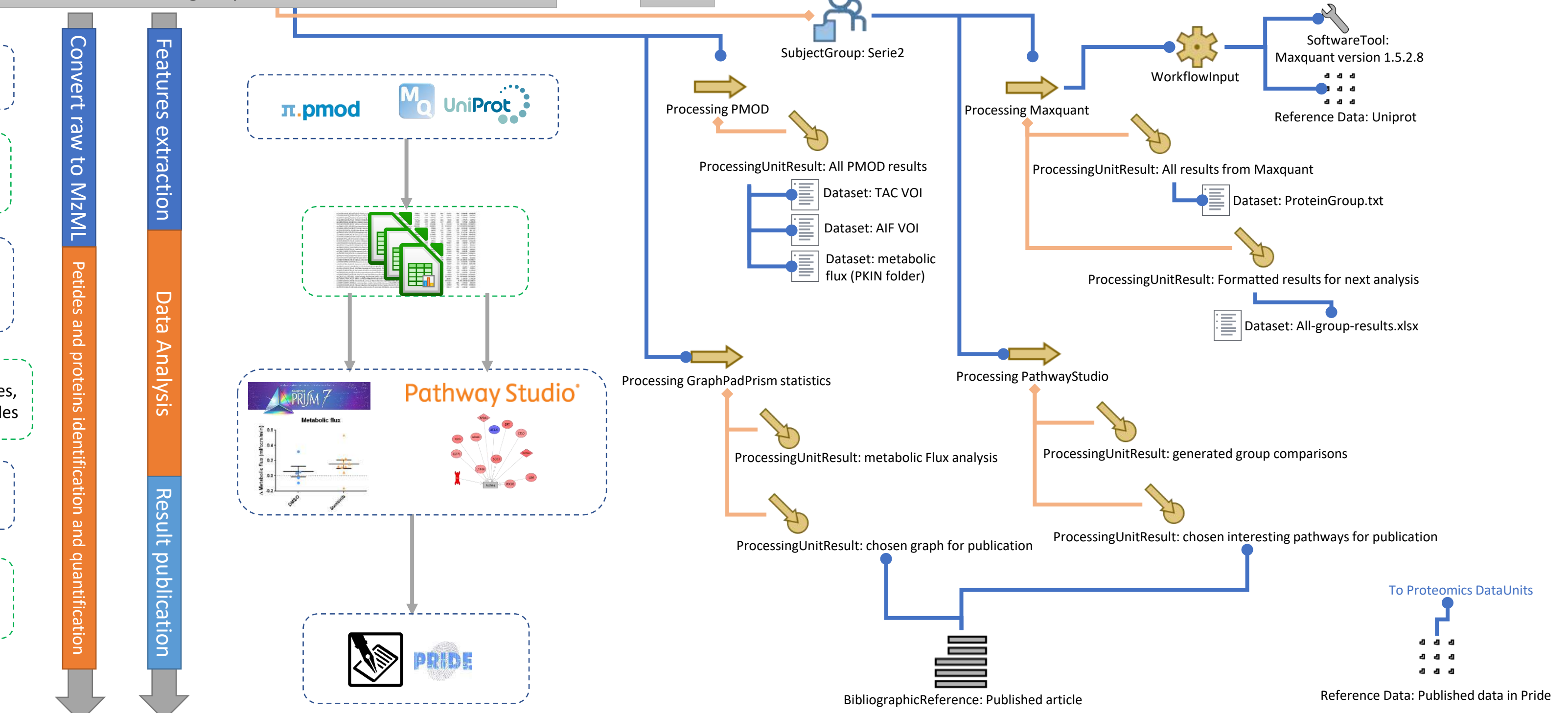
For visibility, only major BMI-LM objects are shown.

Screenshot of processing traceability in SWOMed web client



Results from integrated proteomics workflows. Above, is shown the workflow for raw to MzXML files conversion, and the results from workflows for peptide (PCR_proteomics_peptides) and protein (PCR_proteomics_proteins) identification and quantification.

Processing steps



Conclusion

We have built a centralized management framework for **heterogenous** research data, including imaging and proteomics data lifecycle. It uses standard based methodology that guarantees research data quality and ensures **comprehensive metadata**. We now wish to extend this centralized data management solution to **complex workflows** integrating more and more diverse data sources. Moreover, during the course of this study we encountered an unexpectedly high rate of protocol changes and system evolutions. Therefore, we will develop new tools and approaches taking into account the **evolutions and mutations of biomedical research ecosystems** in order to adapt PLM methods to high protocol mutation rates and improve the **stability** and **resilience** of our management framework for heterogenous research data.