



**HAL**  
open science

## Implementation of a Product Lifecycle Management System for Biomedical Research

Amel Raboudi, Pierre-Yves Hervé, Marianne Allanic, Daniel Balvay, Roberto Duarte, Ismael Bakayako, Thomas Viel, Anais Certain, Thulaciga Yoganathan, Philippe Boutinaud, et al.

► **To cite this version:**

Amel Raboudi, Pierre-Yves Hervé, Marianne Allanic, Daniel Balvay, Roberto Duarte, et al.. Implementation of a Product Lifecycle Management System for Biomedical Research. 18th IFIP International Conference on Product Lifecycle Management (PLM), Jul 2021, Curitiba, Brazil. pp.185-199, 10.1007/978-3-030-94399-8\_14 . hal-04195233

**HAL Id: hal-04195233**

<https://inria.hal.science/hal-04195233v1>

Submitted on 4 Sep 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



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Implementation of a Product Lifecycle Management System for Biomedical Research

Amel Raboudi<sup>1,2,3</sup>[0000-0003-3870-046X], Pierre-Yves Hervé<sup>3</sup>[0000-0002-5525-1750], Marianne Allanic<sup>3</sup>[0000-0002-8816-5744], Daniel Balvay<sup>2</sup>[0000-0003-3802-2224], Roberto Duarte<sup>2</sup>, Ismael Bakayako<sup>2</sup>, Thomas Viel<sup>2</sup>[0000-0002-0756-9166], Anais Certain<sup>2</sup>, Thulaciga Yoganathan<sup>2</sup>[0000-0003-0525-0407], Philippe Boutinaud<sup>3</sup>[0000-0003-3662-1119], Alexandre Durupt<sup>1</sup>[0000-0001-7668-9703], Bertrand Tavitian<sup>2</sup>[0000-0002-5349-8194], Benoît Eynard<sup>1</sup>[0000-0002-5393-4363]

<sup>1</sup> Université de technologie de Compiègne, France

amel.raboudi@utc.fr

benoit.eynard@utc.fr

<sup>2</sup> Inserm U970, Paris Cardiovascular Research Center - PARCC, Paris, France

bertrand.tavitian@inserm.fr

<sup>3</sup> Fealinx Company, Courbevoie, France

pboutinaud@fealinx.fr

**Abstract.** Today's biomedical research shows similar characteristics with manufacturing industry 20 years ago. Biomedical data are more and more complex and heterogeneous due to recent scientific discoveries and technological innovations. In a previous work, the Product Lifecycle Management (PLM) paradigm was applied to neuroimaging research to manage heterogeneous data and their provenance. A BMI-LM data model and an associated platform were proposed to enable data reuse and sharing in neuroimaging. Here, this application is extended to biomedical data related to Histology, Proteomics, and PET-CT modalities. Data originate from the IVIR laboratory in preclinical research of the Paris Cardiovascular Research Center (PARCC). We proposed a data and workflow integration method and applied them to the collected preclinical data. An extended version of the data model, that we called BMS-LM for BioMedical Study – Lifecycle Management, was developed. Three concepts were added to the data model: Agent, Sample and Intervention. As a proof of concept, samples of imported data from the three identified modalities are given. Results support the use of the BMS-LM data model and system to ameliorate trust of biomedical data and prospectively their reuse.

**Keywords:** PLM, information integration, workflow, traceability, provenance, heterogeneous data.

## 1 Introduction

Due to technological innovations and scientific discoveries, a variety of data acquisition modalities and analysis opportunities are available for biomedical researchers.

Thus, research activities are more and more collaborative, use multitude of fine biomedical expertise and generate heterogeneous data i.e., multi-format, multi-disciplinary, multi-sites, and multi-sources. Heterogeneity, combined to poor data annotations and missing data provenance information, hinders data reuse and sharing [1]. To solve this problem for the biomedical research field, mature Research Data Management (RDM) tools and methods are needed. RDM is related to open science, a worldwide movement for a better management, sharing and reuse of scientific data that should be FAIR (Findable, Accessible, Interoperable, Reusable) [2].

Several existent systems are available for biomedical data management. The well-known ones are the Electronic Health Records (EHR) for clinical data, the Hospital Information Systems (HIS) for administrative, financial and insurance data, the Clinical Data Warehouses (CDW) for data archiving and reuse, the Laboratory Information Management Systems (LIMS) for biological data, the Picture Archiving and Communication Systems (PACS) for imaging data, and the electronic laboratory notebooks used by researchers to report on daily activities. None of the available systems is specialized in heterogeneous research data management, neither prepares the data for it sharing and reuse.

Product lifecycle management (PLM) is an integrated approach for data management throughout the lifecycle of a product: starting from specification, design, manufacturing, distribution, maintenance, to recycling [3]. It was mainly applied in aeronautics and automobile domain [4]. Similarities between biomedical research data and industrial data can be identified: multi-site data, multi-disciplinary context, strong collaboration, data heterogeneity, procedures complexity, etc. In a previous work [5] [6], the Product Lifecycle Management (PLM) paradigm was applied to neuroimaging research for a better RDM and better data provenance reporting. Provenance is about retracing the history of data from its creation to its sharing [7]. This resulted in a PLM system that manages heterogeneous neuro-imaging research data: brain images, neuropsychological research questionnaires, and connection to calculation grid with a maximum of traceability. Considering results of [5] [6] conducted on a subdomain of biomedical research, we hypothesize that applying the PLM paradigm for heterogeneous biomedical data may empower data reuse and sharing by improving trust in data with provenance reporting and lifecycle management.

Here, we describe our methods for PLM application to heterogeneous scientific data, and some preliminaries results. The application field of our proposition is the preclinical research domain. It is defined here as research conducted on animals with a therapeutic or diagnostic objective of output to human beings. The neuroimaging domain and the preclinical domain are both subdomains of biomedical research.

First, the PLM applications in biomedical fields are reviewed and the previous work about the PLM application to neuroimaging research is summarized. Then, our methods to implement a PLM solution in the preclinical field are explained. The proposed BMS-LM data model for BioMedical Study Lifecycle Management is described in the results section together with examples from the preclinical domain. A discussion and a conclusion ends the paper.

## 2 State of the Art

### 2.1 PLM application to Other Fields

The application of PLM to other fields is a known research practice. The authors of [8] provide a list of PLM applications in mechatronics, Computer Assisted Engineering (CAE), building design and architecture, services, education, etc. In the biomedical domain, the PLM approach reduced the TTM (Time To Market) of drugs and medications [9]. Drug development value chain starts with laboratory research and ends with the validation of clinical trials. After that, starts the pharmaceutical supply chain. The medical and economics responsibilities behind the pharma-industry make it a critical process-centric domain where every step is quantified and studied in advance, which has motivated the use of PLM to manage pharma-industry data. Main reused PLM functions are: product specification, role definition, project planning, knowledge management, predisposition to change. In addition, the emergence of Industry 4.0 may offer more technological functions to pharma-industry using PLM systems [10].

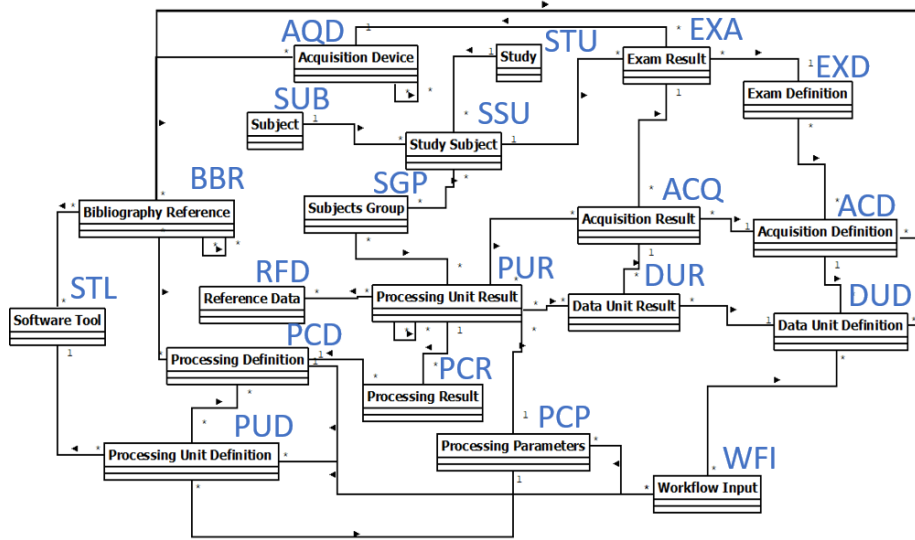
PLM was also applied in biomedical engineering for the construction of medical devices [11]. It was applied to personalized prosthesis due to similarities between 3D design of a car component and 3D reconstruction of an imaging scan of a human body. In both cases, the need for data and information management in a collaborative CAD context is present. Several applications were identified: prosthesis manufacturing [11], prosthesis implantation [12], PLM for personalized face implant [13], lower limb prosthesis development [14], etc.

### 2.2 PLM Application to Neuroimaging

During the BIOMIST (ANR-13-CORD-0007) project, a PLM system was adapted to neuroimaging data where biomedical study lifecycle was defined. It starts with (1) Study Specification, then (2) Raw Data Acquisition and (3) Derived Data Production and ends with (4) Results Publication that opens the door to another (1') Study Specification. Mainly, the reused PLM functionalities were: complex data visualization, standardization using a data model and a Classification tree for neuroimaging, and basic PDM functionalities like database administration and document management. This resulted in a data model for BioMedical Imaging Lifecycle Management (BMI-LM) and a platform using it, called BIOMIST [15], both proposed to ameliorate neuroimaging data managing through traceability and provenance reporting. In what follows, each one is briefly presented.

#### The BMI-LM Data Model

The BMI-LM (BioMedical Imaging – Lifecycle Management) data model is composed of 20 generic classes that structure biomedical data. The UML diagram **Fig.1** presents them, their links and their short names. They belong to two main categories:



**Fig. 1.** UML diagram of the BMI-LM data model [16]

- Definition class (D): A plan specification that describes how results are obtained (acquisition protocol, processing steps, etc.). It can be reused from one study to another. This category is important for provenance reporting. Example: « Data Unit Definition– DUD » for the description of an expected dataset.
- Result class (R): An entity containing real data: acquisition configuration used, processing parameters, raw or derived study data, document or image file, etc. Example: « Data Unit Result – DUR » for the description of a resulting dataset.

Referring to **Fig.1.**, a « Subject (SUB) » participates in a « Study (STU) » as a « Study Subject (SSU) ». The latter undergoes « Exam Result (EXA) » defined by an « Exam Definition (EXD) » and composed of an « Acquisition Result (ACQ) » using an « Acquisition Device (AQD) » that generated « Data Unit Result (DUR) ». The DUR generated from the EXA is used in a « Processing Result (PCR) » composed of « Processing Unit Result (PUR) » and using « Software Tool (STL) ». All have their associated definitions that not always end with the letter D. For instance, « Processing Parameters (PCP) » and « workflow input (WFI) » are parts of the definition classes. At the end, the study results are published as a « Bibliography Reference (BBR) ». The BMI-LM data model did not explicit all relations that exists between its classes. Only two types are used:

- Traceability relation: a link of data provenance traceability between two result classes (R) or two definition classes (D).
- Identification relation: a link between a result class and its corresponding definition class.

In addition to the 20 generic classes of the BMI-LM data model, classes that are more specific are proposed as a tree of concepts called «Classification» in order to specifically describe biomedical data. Each generic class is associated to a tree branch of specific classes. For example, the Study Subject class is associated to a tree of potentially studied organisms: « Homo Sapiens », « Rodent », « Mouse », etc.

### The BIOMIST Platform

The BIOMIST platform was constructed on the top of the PLM software Teamcenter edited by Siemens Industries Software<sup>1</sup>. BIOMIST has a 4-Tiers architecture (Database layer, Enterprise application layer, Web application layer, Client layer) and reused the standard PLM functionalities of the Teamcenter PLM system. Several software modules were added in order to adapt to a neuroimaging laboratory context:

- The BMI-LM data model [15]
- Data export using a web service.
- Simplified data exploration using ODBC (Open DataBase Connectivity) connection between the BIOMIST platform and compatible spreadsheet or statistical software. ODBC is a software interface for standard connection between different database types. It is compatible with several software and database management system (DBMS), and ODBC client libraries are available in many programming languages.
- Medical imaging devices integration using the DICOM communication protocol. DICOM is the reference standard in medical imaging<sup>2</sup>. DICOM is both a file format and a communication standard for image sharing between scanners, visualization stations and *Picture Archival and Communication Systems* (PACS).
- Data processing and communication with computing clusters, achieved through integration of a Python scientific computing workflow management software: Nipype [17]. Nipype command line interface specifications and processing pipelines are imported in the PLM. This way, pipelines can be parameterized and executed over distant computing clusters, managed by schedulers such as SLURM<sup>3</sup>. Results are automatically added to the PLM data graph, using the graph structure of the processing pipeline as template.
- The BIOMIST platform was tested with structural and functional MRI data and associated psychology data. The obtained results were encouraging. The work of [5] [6] [15] [16], is being extended in this paper to enable preclinical data management using the BIOMIST platform and the BMI-LM data model.

---

<sup>1</sup> <https://www.plm.automation.siemens.com/global/fr/products/teamcenter/>

<sup>2</sup> <http://dicom.nema.org/>

<sup>3</sup> <https://fr.wikipedia.org/wiki/SLURM>

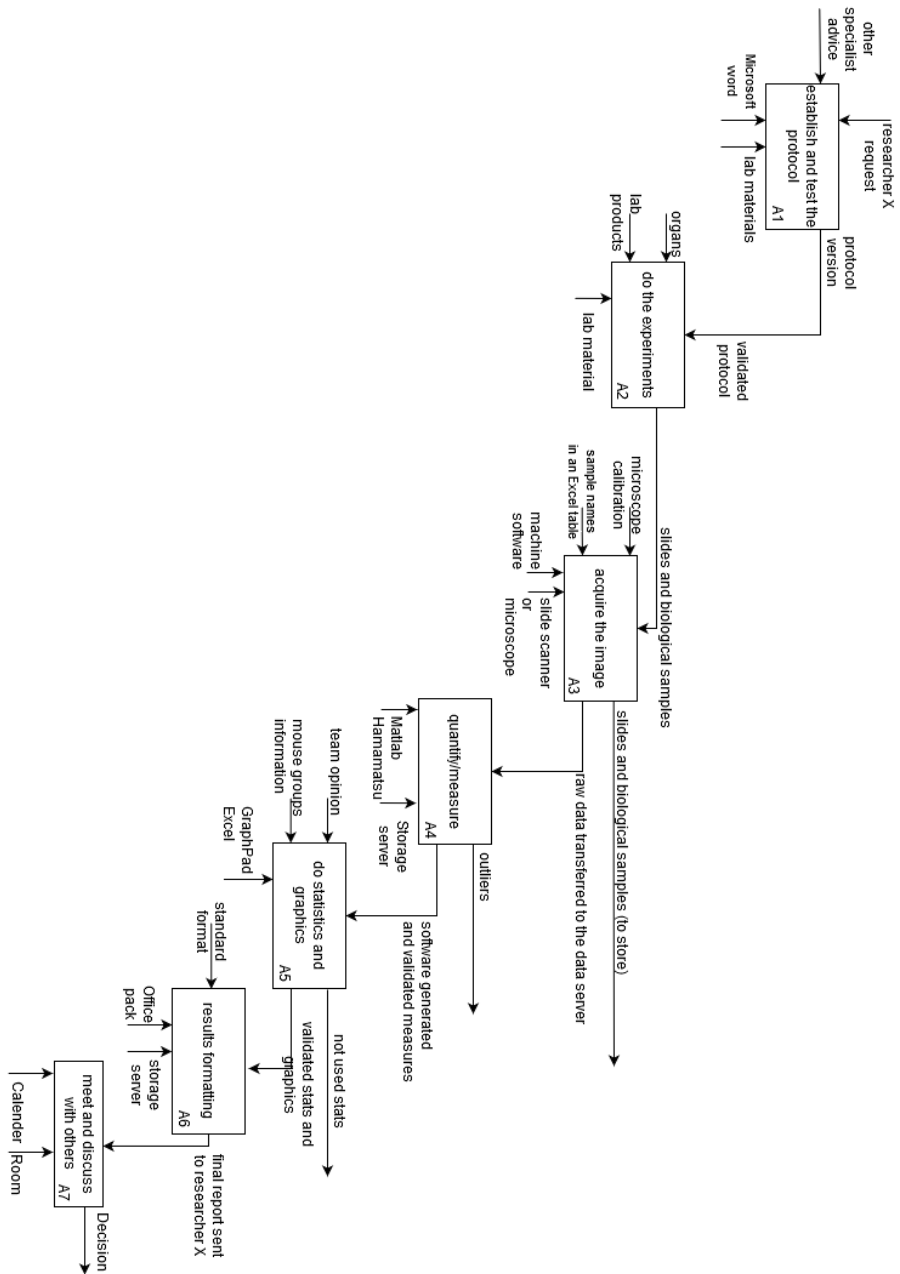
### **3 Methods: Specification and Implementation of a PLM System for Biomedical Data Management**

A data model and a PLM system were proposed for the neuroimaging data management. To extend them to other research domains, the following proposed methods were proposed and applied to preclinical data of the In-Vivo Imaging Research (IVIR) Laboratory in the Paris Cardiovascular Research Center (PARCC).

#### **3.1 Data and Workflow Identification**

As a starting point for the PLM application to preclinical research, typical preclinical data and processing were identified and studied. They are the input for our PLM extension. Data from Histology, Proteomics, and PET-CT imaging were collected and experts in those fields were interviewed. Histology is a method that study biological tissues and samples using microscopes, slide scanner, observations, etc. Proteomics is the study of all the proteins present in a biological sample. PET-CT (Positron Emission Tomography – Computed Tomography) is a medical imaging technique to detect distribution of radiolabeled molecules injected in a living organism. It is used to detect tumors for example after injection of radioactive glucose accumulated in tumors). As an example, the following IDEF0 diagram (**Fig.2.**) explain the research lifecycle for Histology data acquisition and processing. It results from the histology expert interview in the IVIR lab. The IDEF0 **Fig.2.** covers the lifecycle of a research study: the first step relates to the specification step, the second and third describes the raw data acquisition, the next two produces derived data and the final two are related to publication and valorization of results. In the same spirit, other IDEF0 have been produced for other data acquisition modalities (PET-CT and Proteomics) in the lab and could be consulted in [18].



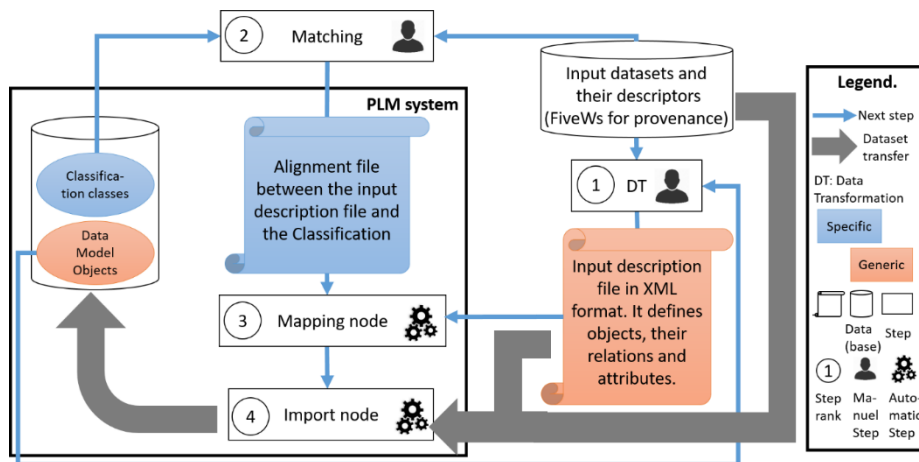


**Fig. 2.** IDEF0 for histology research lifecycle steps: sample preparation, data acquisition and analysis

### 3.2 Data Integration Method

To integrate heterogeneous data in the PLM system, we propose a data integration method that can be applied to different data types. Its modules are described in **Fig.3.** First, we collect provenance metadata about the dataset (to be imported) using the FiveWs questions: what, who, when, where, and why, preconized by [19]. Then, we prepare a table describing the dataset that uses its original metadata and the provenance ones. The description table is then converted via the «(1) Data Transformation» step to an XML file that contains, in addition to provenance and original metadata, the data model generic objects to be created and relations between them.

Each generic data model object is associated to a specific «Classification» instance that aggregates a set of metadata. During the integration of a dataset in the PLM system, data model objects and relations between them are created and linked to the dataset files. Whereas, the dataset annotations are stored in the attributes of the «Classification» tree classes. To make this work, a list of correspondence (Alignment) is established between data model object and «Classification» classes for each new type of data. It is the output of the «(2) Matching» step **Fig.3.** The Alignment file is used by the «(3) Mapping node». The latter establishes automatic routes for the conversion of the XML file contents to a set of: data model objects and relation between them, together with «Classification» instances and their attributes. The XML description file and the initial dataset are sent later to an «(4) import node», that imports the dataset to the PLM system.



**Fig. 3.** Data integration modules used to import heterogeneous biomedical data in a PLM system

### 3.3 Scientific Workflow Integration Method

After integrating heterogeneous data in the PLM system, we need to integrate scientific workflows to process them. The DIMP (Integrated Data Management and Processing,) method was proposed [5]. It aims at reinforcing traceability and provenance

of processing executions as a workflow. The «Workflow Input (WFI)» object in the BMI-LM data model serves the purpose of configuring the workflow execution: its inputs data, its script version, its input parameters and eventual other resources. The “Processing Definition (PCD)” defines a processing chain, composed of «Processing Unit Definition (PUD)» that harbors the source code of the workflow to be run. Each “Processing Parameters (PCP)” stores the parameters to be applied to its «processing unit». These parameters are passed to the command line formatted by Nipype [17] at runtime.

The different steps of this workflow integration method are described in [5]. In the first step, the WFI and the «Subject Group (SGP)» providing the data are chosen by the user. Then, the data are staged on the distant server over ssh, and the processing chain is submitted to the high-performance computing server scheduler. In the last step, the resulting data, including the log files, are uploaded in the PLM system and the user is notified of the end of the process.

In this work, we reused the DIMP method for automatic and mature workflows and we have proposed a second method to exploratory, graphical, and user-dependent workflows in order to guarantee a minimum of traceability even when automation through Nipype [17] is impossible. This method consists on the use of an API (Application Programming Language) to connect between user-dependent software application and the PLM system. It enables communication between both endpoints. We leverage this communication to (1) retrieve and send data from and to the PLM system and (2) send provenance and traceability information during the local workflow execution.

## 4 Results and Application: The BMS-LM Data Model

The first integration tests of Histology, Proteomics and PET-CT data have resulted in an evolution of the BMI-LM data model. First, it was renamed to BioMedical Study - Lifecycle Management (BMS-LM) to adapt more to its extension, and second, three concepts have been added to the data model: Agent, Sample, and Intervention. Each of them was divided in two classes: a Definition class and a Result Class. The list of added classes and their definitions are present in the table 1. Relations between them are detailed in the UML diagram **Fig.4**.

**Table 1.** Added classes list and their definitions

| Added class             | Short name | Definition   |
|-------------------------|------------|--|
| Sample Result           | SAR        | biological samples collection from a study subject                   |
| Sample Definition       | SAD        | biological sample collection protocol and steps                      |
| Intervention Result     | ITR        | action conducted on a study subject that transforms it               |
| Intervention Definition | ITD        | protocol of an action conducted on a study subject                   |
| Agent Result            | AGR        | action of pharmacological product administration to a study subject. |
| Agent Definition        | AGD        | protocol of the administration of a pharmacological product          |

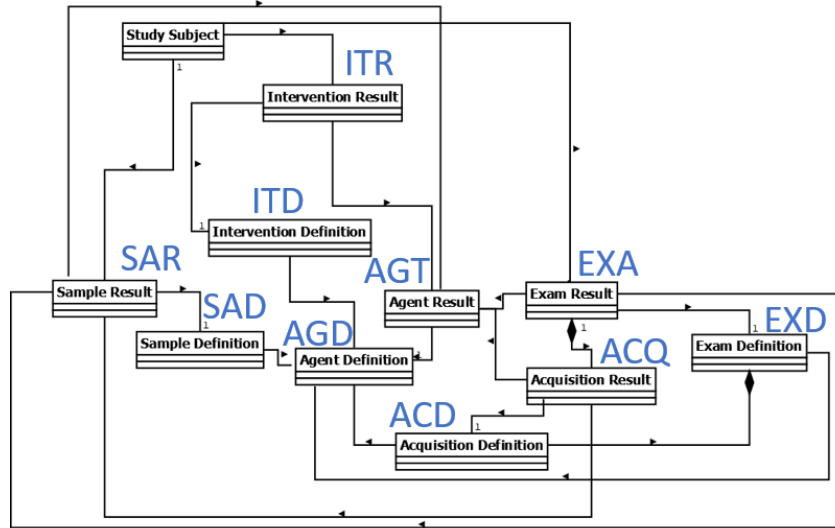


Fig. 4. UML diagram of added classes and their relations in the BMS-LM data model

#### 4.1 Preclinical Research Data Integration

The BMS-LM data model and the proposed integration methods were used to import the IVIR laboratory data in the BMS-LM system (name attributed to the PLM system used for biomedical study lifecycle management). In what follows, examples are presented from each identified modality (Histology, Proteomics, PET-CT) after its import in the BMS-LM system.

research study (STU)

study subject (SSU)

realized exam (EXA) referencing its realized data acquisitions (ACQ). ACQs reference produced data (DUR). DURs references histology images (IF-Composite in the Preview).

STU\_Renotox

- AcquisitionDevice
  - AQD\_Axiomager\_Aptome
- Classification
- ExamDef
- Group
- InterventionDef
- SampleDef
- StudySubject
  - SSU\_000386
  - ExamResBO
    - EXAbio\_JF-NEPHRINE-CD31
      - Acquisition
        - ACQbio\_JF-NEPHRINE-CD31\_1
          - ACD\_JF-Nephrine-CD31
            - Agent
              - DataUnit
                - DURbio\_JF-NEPHRINE-CD31\_1\_CD31
                  - DUD\_JF-CD31
                    - IF-CD31
                      - DURbio\_JF-NEPHRINE-CD31\_1\_Composite
                        - DUD\_JF-Composite
                          - DURbio\_JF-NEPHRINE-CD31\_1\_DAPI
                            - DUD\_JF-DAPI
                              - DURbio\_JF-NEPHRINE-CD31\_1\_Nephrine
                                - DUD\_JF-Nephrine
                                  - IF-Nephrine
                                    - DURbio\_JF-NEPHRINE-CD31\_1\_XML-metadata

IF-Composite

Owner: acqrob.jacqrobh | Date Modified: 17-Apr-2018 19:06 | Release Status: TIF | Type: TIF

Overview | Audit Logs

Properties

      - Object: IF-Composite
      - Name: IF-Composite
      - Description:
      - Owner: acqrob.jacqrobh
      - Group ID: dba
      - Last Modifying User: acqrob.jacqrobh
      - Checked-Out: No Value
      - Checked-Out By:
      - Checked-Out Date:
      - Checked-Out Change Id:

More Properties...

Classification Properties

Preview

Actions: Copy, Open, New Workflow Process...

Fig. 5. Histology exam and its produced images imported in the BMS-LM system

A preview of an histology image of a mouse kidney is presented in Fig.5.. This image is encapsulated in a series of interconnected data model objects in order to describe its provenance, acquisition parameters and context. The examination of the displayed data model objects in the left of Fig.5.. delivers information about the study (STU\_Renotox), the acquisition device (AQD\_Axioimager\_Apotome), the Study Subject involved (SSU\_s000387), the exam name (Exabio\_IF-NEPHRINE-CD31), the acquisition (ACQ), the dataset produced (DUR) and the final image displayed (IF-Composite).

**Fig.6.** presents a PET-CT exam (EXA\_T002167\_201804...) from the study (STU\_AGA-PETRUS...) and the subject (SSU\_s000957). It is composed of two acquisitions: the Computed Tomography (ACQ\_CT) and the Positron Emission Tomography (ACQ\_PT). Status of validity and completeness (incomplete-valid) has been associated automatically to the exam for better quality control.

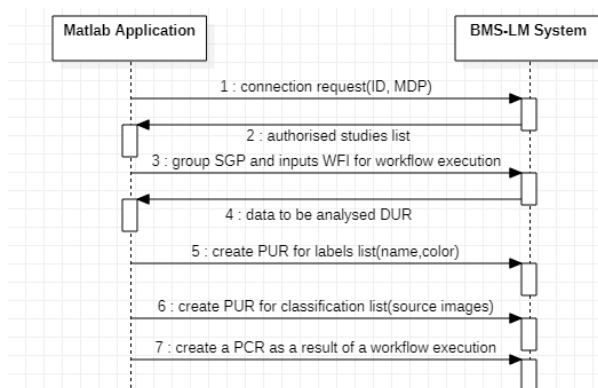
**Fig. 6.** PET-CT exam and its produced images imported in the BMS-LM system

The same data integration method was used for raw and derived proteomics data retrospectively. Traceability links have been established starting from the study specification to the results publication, to demonstrate their role for data provenance reporting and trust. Proteomics data of a finished study in the IVIR lab (STU\_Cardiotox) were reused for this data integration. In the Cardiotox study, Study Subject Groups (SGP) have been defined for data acquisition (SGP\_Control, SGP\_Sunitinib, SGP\_Macitantan) and data processing (SGP\_Proteomics\_analysis). Each Study Subject (SSU) from the proteomics analysis group (SGP), was linked to objects describing its samples (SAR\_peptides\_fraction), and exams (EXA\_proteomics\_MS/MS). Each EXA is related to its acquisitions (ACQ), data units (DUR) and final raw files (ThermoFisher raw spectra). Raw data were analyzed using Maxquant software and the resulting protein list was also integrated to the system as a processing unit result (PUR). The Maxquant analysis processing chain (PCR) is a central node of the chain that links raw data to derived data. It references its input raw data and its output derived data. After a series of connected Processing Results (PCRs) via inputs and outputs as derived data, some interesting results for publication emerged and are integrated to the final

publication (object: BBR\_JSO\_Article<sup>4</sup>). Thus, data from the beginning of the study (STU\_Cardiotox) to its publication in a research paper (BBR\_JSO\_Article) are chained step by step using the BMS-LM data model and system. An illustrated figure of the proteomics integration is accessible via this link<sup>5</sup>.

## 4.2 Preclinical Workflow Integration

As for processing, a graphical Matlab application in the IVIR lab was reused to test our proposed workflow integration method, which is based on the BMS-LM API. It is used to quantify histology images through two key data: a «list of labels» for categories to be quantified in the image (normal tissue, pathologic regions, vessels, etc.) and a «list of classification» that links pixels in the image to predefined labels. The input (images) and the output data (labels list, classification list) in each Unit of the Processing (PCD/PUD, PCR, PUR) were identified. We have also predefined possible Workflow Inputs (WFIs) that references the input data types to be used.



**Fig. 7.** Sequence Diagram of the integration through API of a Graphical Matlab application from the IVIR lab in the BMS-LM system

The sequence diagram in **Fig.7.** presents the implemented communications between both software. Calls to the BMS-LM API functions were added throughout the execution of the Matlab application to retrieve (and send) data and traceability information from (and to) the BMS-LM system. When the Matlab application is configured to use the BMS-LM API, the user must connect to the BMS-LM system otherwise he/she cannot be authorized to list studies and retrieve data from the server. The user selects or creates the Study Subject Group (SGP) and the (WFI) to be used for the current execution. The BMS-LM system sends the images (under DUR), corresponding to the SGP and WFI selection, to the local machine. The user quantifies the images and sends the generated parameters and data (PURs corresponding to «label list» and «classifica-

<sup>4</sup> <https://www.thno.org/v07p2757>

<sup>5</sup> [https://gitlab.com/AmelieSpark/plm21\\_bmslm/-/blob/main/proteomics-sample\\_fullsize.jpg](https://gitlab.com/AmelieSpark/plm21_bmslm/-/blob/main/proteomics-sample_fullsize.jpg)

tion list») to the server throughout the execution. Once the execution finished a Processing Result PCR is sent to the BMS-LM system with all the traceability information of the current execution. It is linked to the input data used for the quantification (DUR encapsulating the input images).

## 5 Discussion and Conclusion

In this paper, PLM systems were applied to a new field: the preclinical research (Histology, Proteomics, and PET-CT.data). This original work ensure a better trust in biomedical data as traceability is performed from the beginning of a research study to its end. Prospectively, this enables a better data reuse and sharing. Heterogeneous data management using a PLM system in the biomedical field was enabled using the proposed integration methods for data and workflow. In addition, the original BMI-LM data model was extended to broader biomedical fields with the adding of 6 classes. It was renamed to BMS-LM for BioMedical Study – Lifecycle Management.

The «Classification» corresponding to each BMS-LM object was enriched with specific concepts from well-known standards and ontologies of the related biomedical fields (QIBO [20], MSO [21], OBI [22]). In this extension, flexibility and heterogeneity of data were taken into consideration. However, some limitations of the «Classification» tree were identified: the difficulty to easily refer to the original ontology concept of a class, the impossibility to express advanced relations other than the inheritance relation, the difficulty to easily find classes for annotation in the «Classification» tree for a non-initiated user. Thus, a loss of interoperability between the used ontologies and the BMS-LM data model and «Classification» could emerge.

Regarding relations in the BMS-LM data model, they convey more specific meaning than identification and traceability. For instance, «Aggregation» is expressed between an SGP and an SSU; «Composition» is found between an EXA and its ACQ; «Contribution» is when an SSU participates in an STU; «Production» and its inverse «Ingestion» references respectively inputs and outputs relations to a processing (PCR); «Transformation» is when a SAR is retrieved from an SSU, etc. Therefore, they must be made explicit.

For all previously presented facts, it has been proposed to convert the data model BMS-LM and its corresponding «Classification» to an ontology in the continuity of our work on the integration of biomedical heterogeneous data in the BMS-LM system. Ontologies are defined as «formal, explicit specification of a shared conceptualization» [23] and are believed to be more explicit and better understandable than data models. This transformation aims at improving the interoperability of heterogeneous data through the BMS-LM system.

## Acknowledgements

The work presented in this article was funded by a 2015REMP9HE085 (DRIVE) grant from Paris University. It was also partially supported by the PACIFIC project (BPI n°2018- PSPC- 07). Authors want to thank all biomedical researchers in the IVIR

lab at PARCC, 3P5 platform for proteomics data, the PARCC Histology platform and engineers in the Fealinx Company: Arthur Grioche, Thierry Brial, and Olivier Menuel.

## References

1. Borgman, C.L.: The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*. 63, 1059–1078 (2012). <https://doi.org/10.1002/asi.22634>.
2. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.
3. Terzi, S., Bouras, A., Dutta, D., Garetti, M., Dimitris Kiritsis: Product lifecycle management – from its history to its new role. *International Journal of Product Lifecycle Management*. 4, 360–389 (2010). <https://doi.org/10.1504/IJPLM.2010.036489>.
4. Eynard, B., Gallet, T., Nowak, P., Roucoules, L.: UML based specifications of PDM product structure and workflow. *Computers in Industry*. 55, 301–316 (2004). <https://doi.org/10.1016/j.compind.2004.08.006>.
5. Allanic, M., Hervé, P.-Y., Durupt, A., Joliot, M., Boutinaud, P., Eynard, B.: Processing and Visual Analyze of Heterogeneous and Multidimensional Data in Biomedical PLM Context. In: Harik, R., Rivest, L., Bernard, A., Eynard, B., and Bouras, A. (eds.) *Product Lifecycle Management for Digital Transformation of Industries*. pp. 385–398. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-54660-5\\_35](https://doi.org/10.1007/978-3-319-54660-5_35).
6. Allanic, M., Hervé, P.-Y., Durupt, A., Joliot, M., Boutinaud, P., Eynard, B.: PLM as a strategy for the management of heterogeneous information in bio-medical imaging field. *International Journal of Information Technology and Management*. 16, 5–30 (2017).
7. Simmhan, Y.L., Plale, B., Gannon, D.: A survey of data provenance in e-science. *Association for Computing (ACM) Machinery Special Interest Group On Management of Data (SIGMOD) Record*. 34, 31 (2005). <https://doi.org/10.1145/1084805.1084812>.
8. Fielding, E.A., McCardle, J.R., Eynard, B., Hartman, N., Fraser, A.: Product lifecycle management in design and engineering education: International perspectives. *Concurrent Engineering*. 22, 123–134 (2014). <https://doi.org/10.1177/1063293X13520316>.
9. Prajapati, V., Dureja, H.: Product lifecycle management in pharmaceuticals. *Journal of Medical Marketing*. 12, 150–158 (2012). <https://doi.org/10.1177/1745790412445292>.
10. Branke, J., Farid, S.S., Shah, N.: Industry 4.0: a vision for personalized medicine supply chains? *Cell Gene Therapy Insights*. 2, 263–270 (2016). <https://doi.org/10.18609/cgti.2016.027>.
11. Lantada, A.D. ed: *Handbook on Advanced Design and Manufacturing Technologies for Biomedical Devices*. Springer US (2013). <https://doi.org/10.1007/978-1-4614-6789-2>.
12. Ngo, T.N.: A PLM based approach for supporting collaboration and knowledge management in the medical domain: Application to the treatment process requiring prosthesis



- implantation, PhD Thesis, Ecole centrale de Nantes, France (2018). <http://www.theses.fr/2018ECDN0013>.
13. Ardila-Mejia, C.C., López-Gualdrón, C.I., Martínez-Gómez, J.M.: Product Lifecycle Management Strategy for the Definition and Design Process of Face Implants Oriented to Specific Patients. In: Chiabert, P., Bouras, A., Noël, F., and Ríos, J. (eds.) *Product Lifecycle Management to Support Industry 4.0*. pp. 181–190. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01614-2\\_17](https://doi.org/10.1007/978-3-030-01614-2_17).
  14. Martínez Gómez, J.M., López Gualdrón, C.I., Murillo Bohórquez, A.P., Garnica Bohórquez, I.: PLM Strategy for Developing Specific Medical Devices and Lower Limb Prosthesis at Healthcare Sector: Case Reports from the Academia. In: Stark, J. (ed.) *Product Lifecycle Management (Volume 4): The Case Studies*. pp. 201–221. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-16134-7\\_16](https://doi.org/10.1007/978-3-030-16134-7_16).
  15. Allanic, M., Hervé, P.-Y., Pham, C.-C., Lekkal, M., Durupt, A., Brial, T., Grioche, A., Matta, N., Boutinaud, P., Eynard, B., Joliot, M.: BIOMIST: A Platform for Biomedical Data Lifecycle Management of Neuroimaging Cohorts. *Frontiers in ICT*. 3, (2017). <https://doi.org/10.3389/fict.2016.00035>.
  16. Allanic, M.: Gestion et visualisation de données hétérogènes multidimensionnelles: application PLM à la neuroimagerie, PhD Thesis, Université de technologie de Compiègne, France (2015). <https://www.theses.fr/199231729>.
  17. Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S.: Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinform.* 5, (2011). <https://doi.org/10.3389/fninf.2011.00013>.
  18. Raboudi, A.: Gestion du cycle de vie des études multimodales de recherche biomédicale préclinique pour le partage et la réutilisation des données hétérogènes, PhD Thesis, Université de technologie de Compiègne, France (2021).
  19. Sahoo, S.S., Sheth, A., Henson, C.: Semantic provenance for eScience - Managing the deluge of scientific data. *IEEE Internet Computing*. 12, 46–54 (2008). <https://doi.org/10.1109/MIC.2008.86>.
  20. Buckler, A.J., Ouellette, M., Danagoulian, J., Wernsing, G., Liu, T.T., Savig, E., Suzek, B.E., Rubin, D.L., Paik, D.: Quantitative Imaging Biomarker Ontology (QIBO) for Knowledge Representation of Biomedical Imaging Biomarkers. *Journal of Digital Imaging*. 26, 630–641 (2013). <https://doi.org/10.1007/s10278-013-9599-2>.
  21. Mayer, G., Montecchi-Palazzi, L., Ovelheiro, D., Jones, A.R., Binz, P.-A., Deutsch, E.W., Chambers, M., Kallhardt, M., Levander, F., Shofstahl, J., Orchard, S., Antonio Vizcaíno, J., Hermjakob, H., Stephan, C., Meyer, H.E., Eisenacher, M.: The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary. *Database (Oxford)*. 2013, (2013). <https://doi.org/10.1093/database/bat009>.
  22. Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Frago, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M.A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A.L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J.A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Schober, D., Smith, B., Soldatova, L.N., Stoeckert, C.J., Taylor, C.F., Torniai, C., Turner, J.A., Vita, R., Whetzel, P.L., Zheng, J.: The Ontology for Biomedical Investigations. *PLoS One*. 11, (2016). <https://doi.org/10.1371/journal.pone.0154556>.
  23. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*. 25, 161–197 (1998). [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6).