



Processing and Visual Analyze of Heterogeneous and Multidimensional Data in Biomedical PLM Context

Marianne Allanic, Pierre-Yves Hervé, Alexandre Durupt, Marc Joliot, Philippe Boutinaud, Benoit Eynard

► To cite this version:

Marianne Allanic, Pierre-Yves Hervé, Alexandre Durupt, Marc Joliot, Philippe Boutinaud, et al.. Processing and Visual Analyze of Heterogeneous and Multidimensional Data in Biomedical PLM Context. 13th IFIP International Conference on Product Lifecycle Management (PLM), Jul 2016, Columbia, SC, United States. pp.385-398, 10.1007/978-3-319-54660-5_35 . hal-01699675

HAL Id: hal-01699675

<https://inria.hal.science/hal-01699675>

Submitted on 2 Feb 2018

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

Processing and visual analyze of heterogeneous and multidimensional data in biomedical PLM context

Marianne Allanic¹, Pierre-Yves Hervé², Alexandre Durupt³, Marc Joliot², Philippe Boutinaud¹, Benoit Eynard³

¹CADESIS company, Courbevoie, France
mallanic@cadesis.com (corresponding author); pboutinaud@cadesis.com

²Université de Bordeaux, GIN, IMN, UMR 5293 CNRS CEA, Bordeaux, France
pierre-yves.herve@u-bordeaux.fr, marc.joliot@u-bordeaux.fr

³Sorbonne Université, Université de Technologie de Compiègne (UTC), UMR 7337, Compiègne, France
alexandre.durupt@utc.fr; benoit.eynard@utc.fr

Abstract. The emergence of PLM for biomedical imaging lifecycle management highlights the needs for management and analysis of heterogeneous, complex and multidimensional data in PLM systems. Data provenance in biomedical imaging domain is complex, notably provenance of processing data, and to ensure full traceability in a purpose of reuse, processing operations must be integrated to PLM systems and processing provenance must be easily analyzable by users. The DIMP (Data Integrated Management and Processing) method was designed for this objective: it allows user to launch easily processing chains from PLM systems and ensures a full management of provenance. The MDG (Multidimensional Dynamic Graph) representation is introduced to formalize complex provenance and data relationships. JGEX (Json Graph EXchange) file format and NeuroGraphViewer web graph visualization client have been developed to facilitate the analysis of MDG. An application of the DIMP method to the study of functional brain connectivity through MDG analysis encourages further work on analysis of complex relationships in PLM systems.

Keywords: heterogeneous data, multidimensional data, graph, processing integration, biomedical, PLM (Product Lifecycle Management)

1 Introduction

The application of Product Lifecycle Management (PLM) concepts to the management of biomedical imaging study lifecycle [3] raises new challenges for the PLM community. Biomedical imaging data are heterogeneous (nature, type and source), due to the natural interdisciplinary of the domain. The lifecycle of a biomedical imaging study is composed of four stages: (1) study specifications, (2) raw data, (3) derived data and (4) published data. The high cost of data (both acquisition and processing) and the need for reproducibility make data reuse and sharing a necessity [15]. Therefore,

keeping data provenance throughout stages of study is a strong concern for the community.

Data processing in biomedical imaging domain is complex: many steps of registration (temporal, spatial) and reconciliation are required to get readable images, and even more steps to analyze them. Obviously, biomedical PLM users cannot set up processing provenance by hand, neither analyze it at a glance in the system in order to reuse data (raw, derived and processing steps). The paper addresses these last issues, by proposing a method for integrated processing management in PLM systems and a way of representing complex relationships so they can be easily analyzed.

Section 2 introduces existing approaches for the processing and analyze of complex relationships and heterogeneous data, from biomedical imaging domain and PLM points of views. Section 3 presents the Data Integrated Management and Processing (DIMP) method, which ensures full provenance of processing data. Section 4 presents the Multidimensional Dynamic Graph (MDG) representation in order to analyze provenance and complex data in biomedical imaging domain. In section 5, an application of the DIMP method to the study of functional brain connectivity through MDG analysis is developed. To end with, the results are discussed in section 6.

2 Approaches for processing and analyze of complex relationships and heterogeneous data

First, data processing is introduced both for biomedical imaging and PLM domains. Second, representation of complex relationships with graphs is discussed. To end with, a synthesis of existing approaches towards our concern is proposed.

2.1 Data processing

Data processing is the operation of transforming data through algorithms, which results in simplified, combined or formatted data. In order to be able to understand how data were obtained and how to reproduce exactly a processing chain, keeping data provenance is unavoidable.

2.1.1 Processing of biomedical imaging data

Provenance is crucial in biomedical imaging domain [13]: cohorts get bigger and bigger, data acquisitions are expensive and results must be reproducible – both for scientific cross validation and longitudinal studies. As images processing is very complex (dependencies, loops, multi-inputs...), the community of biomedical imaging developed pipeline tools that handle workflows and calls to required libraries, notably LONI pipeline [8], Nipype [9] and PSOM [6]. Figure 1 shows as a graph a Nipype workflow to compute raw images as a graph in Tulip visualization software. Processing

steps are complex and the chain is difficult to understand at a glance, which implies using interactive visualization tools.

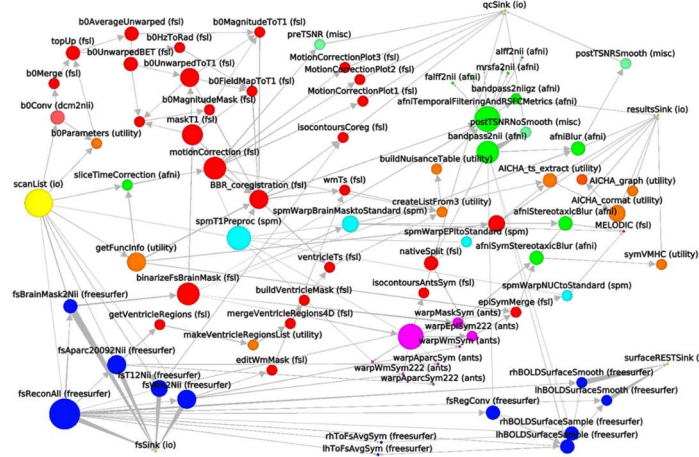


Figure 1. Graph representation of a Nipype workflow, visualized in Tulipe software. Processing units are linked by input data flows. Colors represent software applications.

More and more data repositories in biomedical imaging propose integrated features to launch workflows, however no one offers provenance management of derived data once the computation is done. Processing provenance have been described by two major models: process-oriented model [17], that suits to frameworks, and data-oriented model [16], that suits better for sharing data between laboratories.

2.1.2 Data processing in PLM systems

Data processing integration was not an initial concern in PLM systems. However, simulation lifecycle management has become very valuable to manufacturing companies, as simulation predicts the behavior of a system without performing a physical experiment, which saves time and costs. A simulation process is composed of three steps, as defined by [2]: (1) modelling, (2) solving and (3) post-processing. At each of these steps, choices are made: input Computer-Aided Design (CAD) model, parameters, hypotheses and Computer-Aided Engineering (CAE) tool to use. As these steps are, most of the time, embedded in CAE systems, the traceability of simulation processing is not fully covered in PLM systems.

2.2 Graph analyze of complex relationships

Data dependencies can become complex when it deals with the modelling of systems and processes, whatever their nature – physical, biological or software. One way to analyze these complex relationships is to visualize them with graphs. A graph G is commonly defined as a set of vertices V and a set of edges E , such as (where u and v are any vertices of the graph):

$$G = (V, E) \text{ with } E \subseteq \{(u, v) | u, v \in V, u \neq v\} . \quad (1)$$

Table 1 present the definitions of the main types of graphs used to represent heterogeneous data with complex relationships, depending on data characteristics. Multivariate graphs allow to take data attributes into account during analysis. The types of graphs can be combined, by example a multivariate dynamic graph.

Table 1. Definitions of the main types of complex graphs.

<i>Graph type</i>	<i>Definition</i>
Multivariate M	Graph with n attributes A are added to vertices and edges, such as (for vertices, similar for edges): $A = \{A_1, \dots, A_n\} = (a_{ij}) (j = 1 \dots V ; i = 1 \dots n)$ with A_i a column of the table of attributes and $a_u = (a_{u1}, \dots, a_{un})$ describes all attribute values of node u [11].
Compound C	$C = (G, T)$ is composed of $G = (V, E_G)$ and a tree $T = (V, E_T, r)$, where r is the root vertex, sharing a set of edges, such as $\forall e = (v_1, v_2) \in E_G, v_1 \notin \text{path}_T(r, v_2)$ and $v_2 \notin \text{path}_T(r, v_1)$. Vertices sharing a parent in T belong to the same group and vertices linked to the same parent in G share a generic relationships [1].
Dynamic Γ	Sequence of graphs $\Gamma = (G_1, G_2, \dots, G_n)$ where $G_i = (V_i, E_i)$ are static graphs whose subscript refers to a time point $\tau = (t_1, t_2, \dots, t_n)$ [5].

2.2.2 Graph analysis in biomedical domain

Graphs are used a lot in biomedical domain, to analyze proteins, genes, brain organization, etc. and also processing workflows (see section 2.1.1). Data to be represented by graphs are multivariate (characteristics of brain regions, proteins, algorithms...), dynamic (evolution of brain organization through ageing, evolution of genes combinations, longitudinal studies...), multidimensional (comparison of subjects and groups of subjects in a cohort depending on their characteristics, comparison of families of proteins, comparison of processing chains...). However, graph analysis (both algorithmic and visual) in these domains is at its beginning, as it is for instance in neuroimaging [10].

2.2.2 Relationships analysis in PLM systems

PLM systems manage heterogeneous data (concepts, file formats, metadata...), complex relationships (dependencies) and multidimensional data (BOMs, versioning...). Limits of current PLM interfaces to analyze complex relationships have already been highlighted [4]: no features allow users to check relationships consistence, to browse efficiently relationships, to analyze dependencies or to detect patterns.

2.3 Synthesis of presented approaches

Data processing workflows in biomedical imaging domain are very complex, so processing should be integrated to be able to manage full provenance in PLM systems. Therefore, a first concern of the paper is to address the integration of processing chains in PLM systems, from workflow launch to the management of resulting data.

Graphs are a good representation for complex relationships. Provenance data and biomedical data in general are multivariate, dynamic and multidimensional. However, current graph types cannot represent all these types of data combined. A second concern in the paper is to propose a suitable graph representation to enable multidimensional dynamic graph analysis.

3 Data Integrated Management and Processing (DIMP)

In a context of biomedical PLM (Biomedical imaging Lifecycle Management – BiLM), data is traced at every step of study: from study specifications to published results. In imaging domain, processing chains to obtain useful derived data are complex (multi-inputs, dependencies, loops). To ensure full traceability, the *Data Integrated Management and Processing* (DIMP) method propose to integrate processing tasks to the PLM system: users launch workflows from PLM interface and resulting data are automatically uploaded and linked to inputs, definitions and parameters data. First an extension of the BMI-LM data model for biomedical PLM is presented: it allows to reuse easily processing chains on new data. Second, the stages of the DIMP method are described.

3.1 Workflow Input (WFI): definition of the integrated processing

The BioMedical Imaging – Lifecycle Management (BMI-LM, see details in [3]) data model is composed of generic objects representing concepts associated with specific classes based on domain ontologies. The nineteen generic concepts (see table 2 below) are divided in three categories:

1. *Definition objects*: they described how *result objects* were obtained and they can be reused from one study to another. They are part of the provenance strategy.
2. *Result objects*: they store data of the study, raw and derived, in shape of datasets (files) and metadata. They belong to a specified study.
3. *Ambivalent objects*: depending on the context, these objects can be used as a *definition object* or a *result object*. They are part of the provenance strategy.

Table 2. Generic objects of the BMI-LM data model according to study stages and categories.

Study stages	Definition objects	Result objects	Ambivalent objects
1.Specifications		Study	
2.Raw data	Subject	StudySubject	
	Exam Definition	Exam Result	
	Acquisition Definition	Acquisition Result	
	Data Unit Definition	Data Unit Result	
	Acquisition Device		
3.Derived data	Processing Definition	Processing Result	Reference Data
	Processing Unit Definition	Processing Unit Result	Subject Group
	Processing Parameter		
	Software Tool		
4.Published results			Bibliography Reference

Basically, to launch a processing, users must define: (1) data to compute, (2) algorithms to apply, (3) values of algorithmic parameters. In biomedical imaging domain, a major concern is the reproducibility of results, both on same data and on new data: in longitudinal imaging studies, subjects are having imaging exams regularly on a long period of time (two to ten years), and exactly the same processing chains must be applied to data can be compared.

To meet this objective, a generic object is added to the BMI-LM data model: *WorkFlow Input* (WFI). Its role is to gather all the definition objects needed to launch a processing chain: the processing chain itself (object: *Processing Definition*), processing parameters (object: *Processing Parameters*) and the definition of input data (objects: *Data Unit Definition*, *Processing Unit Definition*). These last data are crucial: they allow the PLM system to query the right data, at any moment, for the subjects selected by the user. Figure 2 shows how using WFI is particularly valuable to reproduce same processing chain several time on new data (acquisitions on the fly, longitudinal studies, new studies).

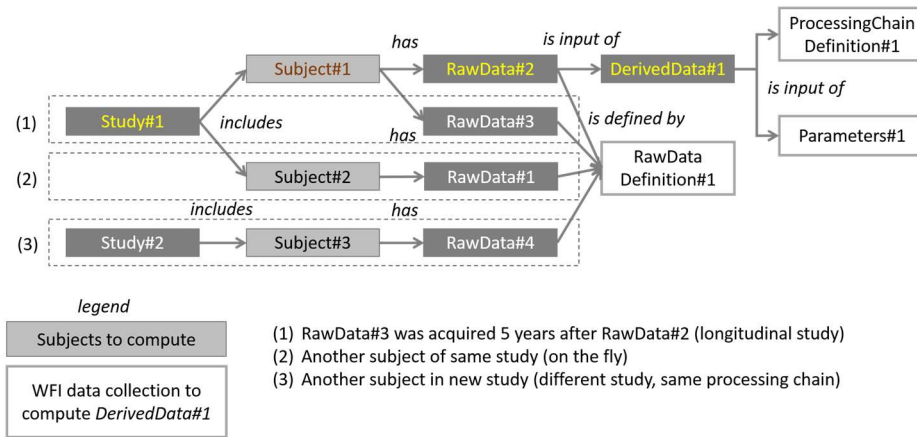


Figure 2. Diagram showing interest of WFI for three use cases in a simplified representation of data management in PLM systems with BMI-LM concepts. Definition of input data (raw data in the figure, but it could be derived data), definition of processing chain and parameters are collected in WFI by users. When a processing chain has to be computed again on new data, WFI is reused and the targeted subjects are given to the system to query corresponding input data. For use case (1), appropriate raw data is found by excluding data that has already been computed with the processing chain and parameters of the WFI.

3.2 Stages of integrated processing in PLM

The main objective of Data Integrated Management and Processing (DIMP) method is to ensure quality provenance of derived data by reducing manual operations from users: data resulting from processing chains are automatically linked to input data, definition of processing chain and parameters. The DIMP method is defined by the following stages:

Initialization

1. (user) build or identify WFI
2. (user) launch integrated processing workflow
 - select WFI
 - select subjects

Workflow execution

3. (PLM system) query input data
4. (PLM system) export in working folder
 - input data
 - definition of the processing chain
 - parameters of the processing chain
5. (pipeline script stored in the definition of the processing chain) execute processing operations locally or on computing grid

Traceability operations

6. (PLM system) upload resulting data
 - create corresponding *result objects*

-
- link *result objects* to input data, and *definition objects* (definition of the processing chain and parameters)
7. (user) receive an email: data are ready

4 Multidimensional Dynamic Graph (MDG) analysis

Section 2.2 highlighted that existing types of graphs are not suitable to represent multivariate data evolving through several dimensions. This section presents a new way of representing and analyzing complex relationships between heterogeneous data. First, a new type of graphs, *Multidimensional Dynamic Graphs* (MDG), is introduced. Second, a file format, *JGEX*, and a web graph visualization client, *NeuroGraphViewer*, are introduced: they were developed to fit the characteristics of MDGs and allows their storage and analysis.

4.1 Multidimensional Dynamic Graphs (MDG) to analyze complex provenance

A *Multidimensional Dynamic Graph* (MDG) Γ is defined by a sequence of graphs:

$$\Gamma = (G_1, G_2, \dots, G_n) \quad . \quad (2)$$

where $G_i = (V_i, E_i)$ are static graphs, called *configurations*, whose subscript refers to a *dimensional* moment $M = (m_1, m_2, \dots, m_n)$: every dimensional moment represents a snapshot of the MDG. Attributes may be associated to every element of the MDG (vertices and edges) and the graph itself. These attributes may evolve according to the dimensions. An element of a dimension is called a *condition*. An illustration of a MDG is given in the figure 3 below to help reader's understanding.

The MDG allows multivariate data, as well as complex and compound relationships that evolve according to many dimensions.

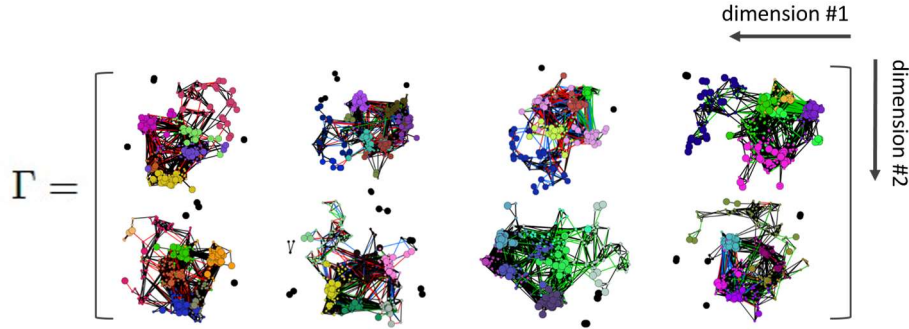


Fig. 3. The Multidimensional Dynamic Graph Γ is composed in height configurations, according to conditions of dimensions: four conditions for dimension #1 and two conditions for dimension #2.

4.2 Json Graph Exchange (JGEX) format to store MDG data

Many graph formats are available; in order to choose a suitable one for MDGs, eleven of them (most commonly used and referred) were tested according to the following characteristics – all required to store MDGs: weighted graphs, attributes on elements of the graph, visualization attributes, default value on an attribute, hierarchical graphs, dynamic graphs, multidimensional dynamic graphs, many graphs, attributes on graphs, groups of nodes and references across graphs.

The result of the comparison is presented in table 3. No format is currently able to store MDG characteristics. GEXF (Graph Exchange Format) can be extended and is only missing multidimensional dynamic graphs and references across graphs, however XML language is quite wordy which implies heavier files. Therefore, there is a strong need to create a new file format to store MDGs.

Table 3. Comparison of the characteristics of existing graph format.

	Weighted graphs	Attributes	Visualisation attributes	Default value of an attribute	Hierarchical graphs	Dynamic graphs	Multidimensional dynamic	Many graphs	Attributes on graphs	Groups of nodes	References across graphs
CSV											
DL Ucinet											
DOT Graphviz											
GDF											
GEXF											
GML											
GraphML											
NET Pajek											
TLP Tulip											
VNA Netdraw											
DGS											

Json Graph Exchange (JGEX) format has been designed to support the exchange of dynamic multidimensional data between programs and applications. JGEX format is an extension of JSON format, and its schema can be found online¹. The main structure of a JGEX file is composed of some metadata, a list of graphs and a list of definitions of attributes. Any defined attribute can be a dimension, and any attribute can vary along a dimension, which allows infinite possibilities.

4.2 NeuroGraphViewer web client

Besides topological analysis, interactive visual analysis is useful to understand complexity: it plays the role of an external cognitive support [12]. However, existing graph viewers do not allow browsing of multidimensional dynamic data.

¹ <http://www.swocloud.net/redmine/projects/biomist-public/wiki/En-jgex>

NeuroGraphViewer is web client developed for BIOMIST project (see section 5.1) that responds to MDG requirements, its last stable version is available online².

Distinctive features of NeuroGraphViewer are (1) the management of filters and display parameters as specific graphs, which allows them to be exported and imported easily, and (2) the possibility to connect to Teamcenter PLM system (edited by Siemens Industry Software), in order to query the database and visually browse and analyze data relationships. Other main features are (3) browsing through dimensions of MDGs and through multi-views display, (4) import and export of JGEX files (including filters and display parameters) as well as other graph formats (TLP, GEXF) and CSV so that any user can build a graph without learning specific formats, and (5) the connection to a selection of graph analysis libraries to perform topology and layout algorithms, in particular some algorithms were developed to fit MDG analysis.

5 Application: management and analyze of neuroimaging data in Teamcenter PLM center

This section presents an application of the DIMP method to the study of functional brain connectivity through MDG analysis, in the context of BIOMIST project.

5.1 Use case: study of functional brain connectivity with biomedical PLM

Functional brain connectivity studies aim at improving understanding of brain organization and how brain regions are working together, with Magnetic Resonance Imaging (MRI) techniques. Subjects' brains are segmented in regions, and the connectivity of each pair of regions is measured, which can be represented by graphs (vertices are regions and connectivity values are edges). Subjects' characteristics (sex, handedness, psychology, genetics...) affect brain organization, therefore MDG are used to represent brain connectivity according to many dimensions.

The proposed use case covers all phases of a study: from study specification and raw data to derived data analysis and publication of results. The stages of the uses case are presented in figure 4.

² <http://212.83.190.113/>

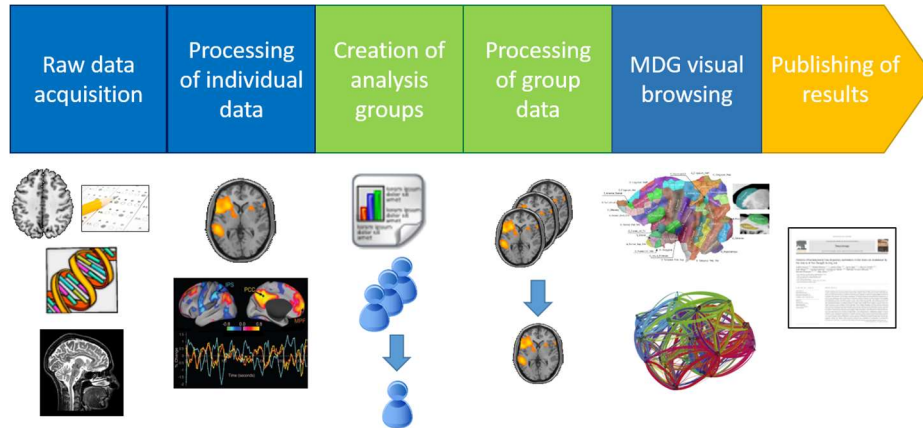


Fig. 4. Stages of the use case

5.2 Application to the BIL&GIN dataset

The BIL&GIN dataset (Mazoyer et al., 2015) has been created by the GIN research group to study hemispheric specialization from cognitive, behavioral, genetic and functional points of view. 45% of the 453 subjects in the dataset are left-handed, which is a bigger proportion than average stated in population.

The dataset is managed for the BIOMIST project [7] in Teamcenter 10 PLM system (edited by Siemens Industry Software) with the BMI-LM data model. A domain classification has been designed based on existing neuroimaging ontologies, and it was implemented in Teamcenter module of classification. Confidentiality and Protected Health Information (PHI) are managed thanks to regular access management and to the project structure (one project is one study) available in the software.

The use case is applied, processing data are computed with DIMP method and final MDG data is stored in JGEX format and analyzed with NeuroGraphViewer web client. DIMP method is implemented with Nipype pipeline tool that drives processing workflows on local computing grids.

Figure 5 presents results obtained with the DIMP method on the BIL&GIN dataset: the complete processing chain is traced in the PLM system, including data inputs, processing definition and processing parameters. MDG of brain connectivity are processed in JGEX format and can be analyzed in NeuroGraphViewer web client.

NeuroGraphViewer also allows to query objects from Teamcenter. In the interface of the web client, users may analyze queried relationships with available libraries of algorithms, apply filters and adjust the layout in an interactive way. The graph created from Teamcenter objects and relationships can be saved in JGEX format for further analyses.

A video showing the whole steps of the use case, from individual raw data to dynamic graphs of subjects' groups, in Teamcenter and in NeuroGraphViewer is available online at the web site of BIOMIST project (<http://biomist.fr>).

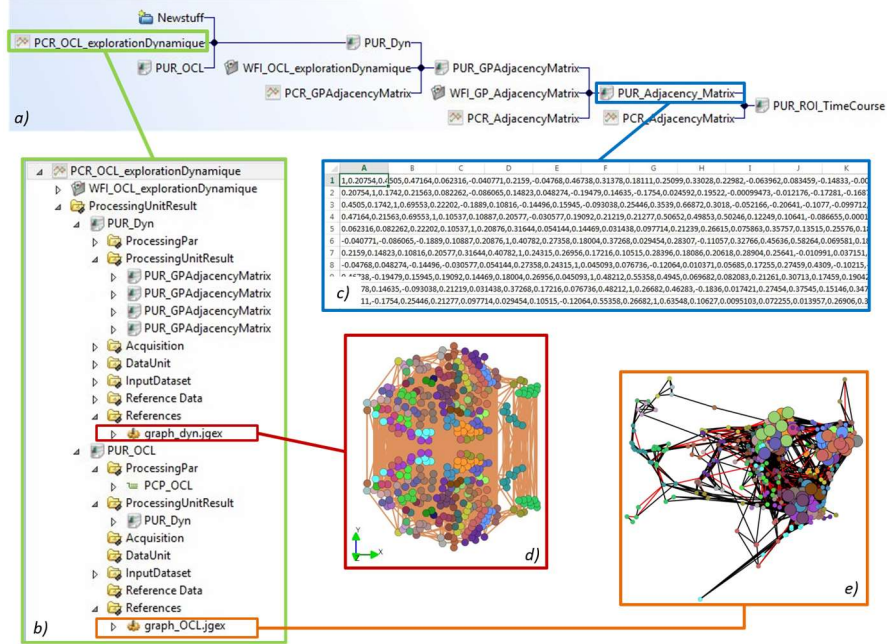


Fig. 5. View of some data resulting from the application of DIMP method to the BIL&GIN dataset in Teamcenter PLM system. a) Processing chain from an individual adjacency matrix to the final MDG to study functional brain connectivity. b) Data and relationships of the final layout processing of the MDG: data inputs (adjacency matrices of the 4 groups), processing definition, processing parameters, final JGEX dataset. c) Adjacency matrix of an individual. d) and e) Visualization of the MDG in NeuroGraphViewer web client, respectively with a 2D-anatomical layout and a OCL-force layout.

6 Conclusion and discussion

The work presented addresses integrated processing management and analysis of complex relationships in biomedical PLM systems. The DIMP method ensures a complete and integrated provenance of processing data in PLM systems. MDG representation allows to formalize heterogeneous data and complex (dynamic and multidimensional) relationships. A new file format, JGEX, allows to store and to exchange MDG data in PLM systems and with NeuroGraphViewer web client.

The combination of these methods and tools constitutes an efficient way to manage quality derived data with full processing provenance: provenance is not set up by users, which prevents mistakes and omissions, and it can be easily analyzed, both quantitatively (graph topology analysis) and qualitatively (graph visual analysis).

The application of this work to the study of functional brain connectivity shows that biomedical imaging domain would benefit from using PLM systems. The work done on the BIL&GIN dataset in Teamcenter PLM system is understandable by users external to the study – thanks to a complete provenance –, which means that derived

data can be reused, and that the processing chain that was computed can be applied to another dataset. In biomedical imaging research – and in particular in neuroimaging –, correlations between subjects' characteristics and behavior are looked for, which implies that data are analyzed through many dimensions. By enabling multidimensional representation of data, MDGs open promising perspectives for finding patterns and correlations.

However, graph drawing and information visualization domains have started to focus on dynamic graphs very recently (the last ten years) and there are still some aspects that have not been addressed, no speaking of MDGs. So future work should focus on proposing topology and layout algorithms for MDG analysis, which would be useful both for provenance analysis in PLM systems and biological networks analysis.

Even if this work was developed for biomedical imaging study management, there is nothing preventing manufacturing industry from benefiting of it. First, the DIMP method could be used to enhance simulation lifecycle management. Second, relationships between the different BOMs of a product (requirements, eBOM, mBOM...) are complex and MDG representation could be of great interest for understanding impacts of a change in the requirements, to check for consistency through BOMs or to analyze the evolutions of configurations of a product.

Acknowledgments. The work presented in the paper is conducted within the ANR (Agence Nationale de la Recherche)-funded project BIOMIST (no. ANR-13-CORD-0007) for thematic axis no. 2 of the Contint 2013 Call for Proposal: from content to knowledge and big data. This study also benefitted from ABACI, a project that is supported by a public grant from the French Agence Nationale de la Recherche within the context of the Investments for the Future Program, referenced ANR-10-LABX-57 and named TRAIL. The authors would like to thank Thierry Brial, Nicolas Boulic and Arthur Grioche from Cadensis, and Pierre-Yves Hervé from the GIN, who supported technically their work.

References

1. Abello, J., Kobourov, S. G., & Yusuf, R. 2005. Visualizing large graphs with compound-fisheye views and treemaps Graph Drawing. Springer, 2005, pp. 431–441.
2. Adams, V. A Designer's Guide to Simulation with Finite Element Analysis, NAFEMS (2008)
3. Allanic, M., Durupt, A., Joliot, M., Eynard, B., & Boutinaud, P. (2014). Towards a data model for PLM application in Bio-Medical Imaging. Proceedings of TMCE 2014, May 19–23, 2014, Budapest, Hungary, edited by I. Horvath and Z. Rusak - Organizing Committee of TMCE 2014, pp. 365–376.
4. Allanic, M., Durupt, A., Eynard, B., Joliot, M., Brial, T., & Boutinaud, P. (2014). Towards an enhancement of relationships browsing in mature PLM systems. Product Lifecycle Management for a Global Market.
5. Beck, F., Burch, M., Diehl, S., & Weiskopf, D. 2014. The State of the Art in Visualizing Dynamic Graphs. EuroVis STAR, 2014 9th-13th InfoVis Conference.
6. Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J., Zijdenbos, A., & Evans, A. C. (2012). The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Frontiers in neuroinformatics, 6, 7.

-
7. BIOMIST, <http://biomist.fr> (visited on 04/01/2016)
 8. Dinov, I. D., Van Horn, J. D., Lozev, K. M., Magsipoc, R., Petrosyan, P., Liu, Z., ... & Toga, A. W. (2009). Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. *Frontiers in neuroinformatics*, 3(JUL).
 9. Gorgolewski, Krzysztof, Burns, Christopher D, Madison, Cindee, Clark, Dav, Halchenko, Yaroslav O, Waskom, Michael L, & Ghosh, Satrajit S. 2011. Nipype : A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, 5, 13.
 10. Hutchison, R Matthew, Womelsdorf, Thilo, Allen, Elena A, Bandettini, Peter A, Calhoun, Vince D, Corbetta, Maurizio, Della Penna, Stefania, Duyn, Jeff H, Glover, Gary H, Gonzalez-Castillo, Javier, et al. 2013. Dynamic functional connectivity : promise, issues, and interpretations. *Neuroimage*, 80, 360–378
 11. Jusufi, I. *Multivariate Networks: Visualization and Interaction Techniques*. 2013, PhD thesis.
 12. Keller, T., & Tergan, S.O. Visualizing knowledge and information: An introduction. *Knowledge and information visualization*, pp.1–23. (2005)
 13. MacKenzie-Graham, A. J., Van Horn, J. D., Woods, R. P., Crawford, K. L., & Toga, A. W. (2008). Provenance in neuroimaging. *NeuroImage*, 42(1), 178-195.
 14. Mazoyer, B, Mellet, E, Percey, G, Zago, L, Crivello, F, Jobard, G, Delcroix, N, Vigneau, M, Leroux, G, Petit, L, et al. BIL&GIN : a neuroimaging, cognitive, behavioral, and genetic database for the study of human brain lateralization. *NeuroImage* (2015)
 15. Poline, J-B., Breeze, J. L, Ghosh, S., Gorgolewski, K., Halchenko, Y. O, Hanke, M., Haselgrove, C., Helmer, K. G, Keator, D. B, Marcus, D. S, Poldrack, R. A, Schwartz, Y.,
 16. Simmhan YL, Plale B, Gannon D. A survey of data provenance in e-science. *Sigmod Record*. 2005. pp.31–36.
 17. Zhao J, Goble C, Stevens R, Bechhofer S. Semantically linking and browsing provenance logs for e-science. *Semantics of a Networked World: Semantics for Grid Databases*. 2004. pp.158–176.