A study of Information System mutations (changes) using Knowledge Organization System (KOS) applied to biomedical research study

Amel Raboudi[1,2,3]

[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), UMR 7337 Roberval, Compiègne, France
amel.raboudi@utc.fr

Abstract. A biomedical research study is a collaboration process where several experts, institutions, disciplines and data sources are involved. Traceability of data provenance and efficient data management are essential in order to guarantee results integrity. Researchers, however, use different Information Systems (IS) in order to collect, process and analyze their increasingly complex datasets. Accordingly, Research Data Management (RDM) is seen as a tedious, time consuming and error prone task. In a previous work, an IS based on Product Lifecycle Management (PLM) technology was proposed to manage data heterogeneity and provenance throughout the biomedical research lifecycle. However, the fast-changing context of biomedical research causes data, information and knowledge changes, that we hereby call mutations. Mutations can affect IS components and impact IS consistency. In a collaboration context, it is important to have the same shared knowledge for all actors. Therefore, the use of a Knowledge Organization Systems (KOS) is proposed in order to model shared knowledge and enable dealing with knowledge level mutations.

Keywords: Information System, Knowledge Organization System, Ontology evolution, Research Data Management, Change Management, Mutation.

1 Introduction

Biomedical studies are multisource, multidisciplinary, multimodal, multi-partners and include longitudinal series. Biomedical researchers are constantly moving back and forth from one data source to another, in order to collect, curate, process and analyze heterogenous data. Therefore, biomedical Research Data Management (RDM) is particularly a complex, time consuming and error prone task. Issues about research reproducibility, data provenance, data sharing, data interoperability and reuse are major in biomedical research field. Solutions such as Information Systems (IS), experts, accurate methods and processes are essential to data management for biomedical RDM. Biomedical research labs rarely use intensive processes and methods for RDM; in most cases lab investment in that field is negligible: In 2016, European Union decided for the first time to allocate 5% of project budget to RDM [6].

Several ISs are used for biomedical RDM, each one is dedicated to a type of data: Laboratory Information Management System (LIMS) [8] for biological samples information; PACS (Picture Archiving and Communication System) and RIS (Radiology Information System) for imaging data, etc. They are either domain specific or data format centric or both. Besides, they don’t cover all aspects and steps of a biomedical research study. In a previous work, an IS based on Product Lifecycle Management (PLM) technology: Biomedical PLM [1] was proposed to manage heterogeneous data provenance. It aims at managing all types of biomedical data throughout the biomedical research lifecycle: (1) specification, (2) acquisition of raw data, (3) processing of derived data, and (4) scientific publication. Biomedical PLM is a study centric, lifecycle-oriented data management system, that enables sharing among actors, processes, organizations and distant sites.

The context of our work is the DRIVE-SPC[1] project. Its goal is to develop an integrated solution to manage biomedical research data of the Imaging Research Laboratory (LRI) - team 2 - at Paris Cardiovascular Research Center (PARCC). Data used by LRI are mainly preclinical imaging exam results for oncology and cardiology research: PET-CT, MRI, ultrasound, and histology. A biomedical PLM instance is currently deployed for the lab members in order to manage data of several studies: cardiotoxicity of a cancer treatment, tumors metabolism, etc.

The fast-changing context of biomedical research is an additional issue for data sharing among the laboratory members. For instance, a researcher changes the way he organizes his data without informing data users, a change in the measurement unit of a crucial parameter for analysis after software update, internal changes in a partner laboratory that affect project schedule and eventually data, discovery in science that must be taken into consideration, etc. Observations done on the LRI lab reveal that previously listed unexpected changes can have serious consequences on the IS and the research project. For instance, limited access for a period of time, errors in automatic toolkits/routines, loss of data, inconsistency of information, loss of time, issues about results integrity, impossibility of data sharing and loss of data provenance etc. All these events have in common data and information changes (lost or gain) that were not anticipated by the IS original design. Our hypothesis is that unexpected changes are mainly related to knowledge sharing issues, because motivation behind unexpected changes exists in people minds and organizations memory. Therefore, changes cannot be studied without an accurate modeling of the shared knowledge among all IS actors (users/developers).

The research problem addressed in this article is how to manage unexpected changes in an IS (applied to biomedical PLM) in order to guarantee its consistency and continuous usability, taking into consideration the fast-changing context of biomedical research and the variable lifespans of different IS components and partners. First, a literature review is presented. Then, we focus on our proposed preliminary approach to manage IS changes, and finally our research methodology and discussion are drawn.

1 a collaboration between Fealinx company and LRI, financed by USPC university.

2 State of the art

Data-Information-Knowledge-Wisdom (DIKW) framework for IS research highlights the importance of knowledge: Knowledge is data and/or information that have been organized and processed to convey understanding, experience, accumulated learning, and expertise as they apply to a current problem or activity [20]. IS definition at [4] describes two parts of an IS: the known part since the IS design phase, and the unexpected part, that depends on all other systems (technical/social) composing the IS environment. In Information Systems, data and information are explicitly managed, but, knowledge is implicit which prevents unexpected changes to be properly managed. To make knowledge explicit among actors, and therefore manage the impact of knowledge changes on IS, we must look for a way to manage knowledge.

2.1 Knowledge Organization System (KOS)

Knowledge Organization Systems (KOS) are designed to manage knowledge: KOS are defined as all types of schemes for organizing information and promoting knowledge managements [12]. KOS includes classification schema, standardized terminology, structured vocabulary, glossaries, semantic networks, ontologies, etc.

Knowledge Organization System and IS. Managing the knowledge level with KOS is an important step in the process of ensuring IS consistency: The work on [2] proposed and validated CoMIS-KMS, a process to guide the conversion of an existing Information System to a knowledge management system with the use of a knowledge base (KOS). Haase PhD [10] proved that ontology (KOS) evolution allows to manage changes in distributed ISs in a consistent manner.

Knowledge Organization System evolution. Dos Reis et al. [5] studied KOS mapping maintenance issues after KOS evolution in the biomedical field. Studied KOSs are: NCIT (Thesaurus), ICD-9-CM (Classification), SCT (Ontology) and MedDRA (Dictionary). General KOS evolution is rarely addressed in the literature, instead, the evolution study of certain types of KOS exists and especially ontology [10] [16] [19] [21] [22] [24].

2.2 Ontology evolution

A KOS classification proposed by [3] underlines that ontologies are the most semantically clear type of KOS. Ontology is an explicit specification of a shared conceptualization of a domain of interest [9]. Ontology change is a largely addressed field in the literature. An exhaustive analysis is presented in [7] where ten subareas for ontology change research fields are defined together with their mutual boundaries. Some of them deals with heterogeneity resolution between ontologies as a proposed method to change an ontology, through ontology mapping and matching [24]. Another concern is about ontology fusion (integration and merging) issues [21] [10]. The same article also presents ontology evolution and separates it from ontology debugging. Both research fields focus on incorporating changes in an existing ontology while avoiding inconsistency.

A state of the art at [23] presents ontology evolution as a five step cycle as follow: (1) Detecting the need for evolution, (2) Suggesting changes, (3) Validating changes, (4) Assessing evolution impact, (5) Managing changes: (5.a) Recording changes and (5.b) Versioning. For each step, a review of related literature is presented. Briefly, (1) is the starting point for ontology evolution process by detecting a need for change that can be initiated from user behavior [19] or data sources [22]. (2) is the phase of change representation with the help of structured or unstructured resources, for example online ontologies were used by [22] as a background knowledge for integrating newly discovered concepts. (3) assesses the relevance of the suggested change to the domain and its impact on ontology consistency. (4) treats impact on dependent applications and external artifacts that uses the ontology under change. (5) applies and traces the change throughout ontology versions. An example is in [14], it applies provenance W3C standard [15] in order to trace ontology changes.

3 Proposed approach

Our approach consists of two main propositions: first, we develop an analogy between IS unexpected changes and mutations in genetics, and second, we propose a model for shared knowledge among IS actors in order to manage unexpected changes. Both are in their preliminary steps.

Expected and unexpected change. Based on IS definition at [4], we propose to consider two types of changes: expected and unexpected changes. Expected changes are managed according to a defined and established process of change management and IS evolution: identifying a need for change, scheduling the change operation and executing it. Unexpected changes cannot be managed through a regular process. They can be: an IS dependency change, a wrong use of an IS functionality, external systems software update, etc. We focus on unexpected changes as they have serious effects on systems consistency and usability such as data sharing and data interoperability issues.

Unexpected changes and mutation. We propose to consider unexpected changes as mutations. A mutation in genetics is a sudden change in DNA code that continuously occurs in bio-cells and that is fundamental for species evolution with regard to natural selection [13]. In IS research or in genetics, an evolution process for an object (thing) is a migration from a consistent state A to a consistent state B. The complete analogy with genetics is not treated in this article but it will be addressed in the course of this PhD. IS mutations can affect any IS component and can lead to an IS evolution (best case scenario) or an IS inconsistency otherwise.

DIKW and Knowledge engineering. We choose the DIKW framework to consider IS mutation concept. Mutations in DIKW framework can occur at each level: data mutation, information mutation, and knowledge mutation. Observations done at the LRI lab revealed that available data and information in the IS do not give any indication about knowledge involved. Besides, knowledge mutations have their origin in external knowledge sources and Information System technical and social environment. Thus, we propose to study IS mutations with the use of knowledge engineering in order to focus on knowledge mutations. Our aim is to ensure knowledge sharing among all actors. Then, we propose to study knowledge evolution and link it to IS evolution in order to manage mutations phenomenon in all DIKW levels. To this end, we must deal with two challenges. First, lab data integration in Biomedical PLM for IS usability enhancement. Therefore, we proposed a generic data integration method that allows different research data types to be imported in Biomedical PLM [17] [18]. And second, modeling Biomedical PLM knowledge taking into consideration change management and mutations issues, which is a work in progress.

Knowledge Organization System (KOS) for Biomedical PLM. IS manages explicitly data and information levels, whereas Knowledge Organization Systems (KOS) manages knowledge level. We propose to manage knowledge mutations with the help of a Biomedical PLM KOS. As presented in the state-of-the-art, ontologies are the most semantically clear type of KOS. The proposed biomedical PLM KOS is based on an ontology that models the knowledge of (1) the Biomedical PLM Information System and (2) the shared knowledge of biomedical research studies: general biomedical ontologies, domain ontologies, laboratory vocabulary, etc.

4 Methodology

The aim of this research is to manage unexpected changes, that we call mutations, in Information Systems in order to ensure IS consistency and continuous usability.

Key steps of our work methodology are (1) exploration of the DRIVE-SPC project context and identification of unexpected changes, (2) state-of-the-art of IS changes related literature, (3) development of biomedical PLM KOS for biomedical RDM with the use of ontologies, (4) proposition and modeling of the whole process of managing mutations in IS with the use of KOS in biomedical research study and (5) test and validation of the proposed approach according to DRIVE-SPC project context.

Our first PhD year focused on the deployment of the Biomedical PLM IS in LRI lab in order to identify mutations. It aims to increase usability of Biomedical PLM in the lab in order to track every type of mutations. Therefore, data of a pilot research study were integrated retrospectively in the IS in order to give a real-life use case for researchers [20]. And recently, some speed-interviews sessions with five key users were organized in order to identify more relevant use cases scenarios.

Presently, ontology representing Biomedical PLM knowledge is under construction. Exploration of available methodologies in literature for ontology construction and integration, together with ontology development framework choice (Protégé, NEON

Toolkit)[2] and language selection (OWL Lite, OWL DL, SKOS, RDFS) are some of the ongoing work.

Next, matching and mapping between (1) Biomedical PLM ontology, and (2) Biomedical PLM Information System, and traceability of mutations in the whole system will be developed.

5 Discussion

In this article, we introduced Information System mutation concept (unexpected change) based on a preliminary analogy with genetic mutations, and we proposed an approach to analyze Information System mutations with the help of Knowledge Organization System (KOS). It aims to provide a shared reference for IS actors and then facilitate mutations management. In order to succeed the shared knowledge design, a further literature review must be done concerning ontology integration, mapping and matching, together with ontology versioning, evolution and debugging.

Actors involvement. IS needs to maintain up-to-date functionalities despite the fast-changing context of biomedical research studies and the multiple partners involved. Our methodology involves partners from an early stage. This is done through interviews, data management behaviors tracking, data preparation, etc. Our proposition depends on IS actors reactivity and understanding of knowledge acquisition importance. This is a risk factor to consider in our research.

System modeling. With the design of Biomedical PLM KOS, a cover of DIKW levels mutations is assured and a larger vision of system changes is provided, which offers a comprehensive framework for IS mutations management. However, this choice adds complexity upon the Biomedical PLM system. Thus, it is interesting to consider a model for the whole system (IS+KOS) to better clarify the role, boundaries and impact of each one.

IS customization. When a mutation occurs, IS providers are supposed to act rapidly in order to manage mutations and ensure continuous IS relevance to customer's need. Haller [11] describes the dilemma of IS providers "in-between" the management of all IS changes and the reduction of IS customization related costs. This is an interesting requirement to consider while proposing mutation management strategy.

2 https://www.w3.org/wiki/Ontology_editors

Acknowledgments

The author wants to thank her direct supervisor at Fealinx Company Dr. Marianne Allanic. Her PhD directors Pr. Bertrand Tavitian and Pr. Benoît Eynard. Her supervisors Dr. Alexandre Durupt (UTC), Dr. Philippe Boutinaud (Fealinx), and DRIVE-SPC project colleagues: Dr. Pierre-Yves Hervé (Fealinx) and Dr. Daniel Balvay (Inserm). Special thanks to LRI key users: Dr. Thomas Viel, Anais Certain, Thulaciga Yoganathan, and Caterina Facchin for their time and help.

References

1. Allanic, M. et al.: PLM as a strategy for the management of heterogeneous information in bio-medical imaging field. International Journal of Information Technology and Management. 16, 1, 5–30 (2017).
2. Anderson, R., Mansingh, G.: Migrating MIS to KMS: A Case of Social Welfare Systems. In: Osei-Bryson, K.-M. et al. (eds.) Knowledge Management for Development: Domains, Strategies and Technologies for Developing Countries. pp. 93–109 Springer US, Boston, MA (2014).
3. Bergman, M.: An Intrepid Guide to Ontologies. 24 (2007).
4. Beynon-Davies, P.: The 'language' of informatics: The nature of information systems. International Journal of Information Management. 29, 2, 92–103 (2009).
5. Dos Reis, J.C. et al.: Analyzing and supporting the mapping maintenance problem in biomedical knowledge organization systems. In: Proc. SIMI Workshop at ESWC. pp. 25–36 (2012).
6. European Commission, Directorate-General for Research and Innovation: Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud. Publications Office, Luxembourg (2016).
7. Flouris, G. et al. : Ontology change : classification and survey. The Knowledge Engineering Review. 23, 2, 117–152 (2008).
8. Gibbon, G.A.: A brief history of LIMS. Laboratory Automation & Information Management. 32, 1, 1–5 (1996).
9. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition. 5, 2, 199–220 (1993).
10. Haase, P.: Semantic Technologies for Distributed Information Systems. (2006).
11. Haller, K.: Information System Maintenance Costs: The "In-between "Challenge. Presented at the Workshop Software-Reengineering, Germany (2010).
12. Hodge, G.: Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Digital Library Federation, Council on Library and Information Resources, 1755 Massachusetts Ave (2000).
13. Jacob, F.: Evolution and tinkering. Science. 196, 4295, 1161–1166 (1977).
14. Kondylakis, H., Papadakis, N.: EvoRDF: evolving the exploration of ontology evolution. The Knowledge Engineering Review. 33, (2018).
15. Lebo, T. et al.: Prov-o: The prov ontology. W3C recommendation. 30, (2013).
16. Noy, N.F., Klein, M.: Ontology Evolution: Not the Same as Schema Evolution. Know. Inf. Sys. 6, 4, 428–440 (2004).

17. Raboudi, A. et al.: Integration and provenance control of proteomics data using SWOMed, a Product Lifecycle Management framework for biomedical research. Presented at the SMMAP Congress October (2017).
18. Raboudi, A. et al. : Traçabilité de l'intégration de données biomédicales hétérogènes dans le système SWOMed de gestion du cycle de vie des études biomédicales. In: actes du symposium SIIM 2017. , Toulouse (2017).
19. Stojanovic, L. et al. : User-Driven Ontology Evolution Management. In: Gómez-Pérez, A. and Benjamins, V.R. (eds.) Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. pp. 285–300 Springer Berlin Heidelberg (2002).
20. Turban, E. et al.: Introduction to Information Technology. Wiley, New York (2004).
21. Xuan, D.N. et al.: Ontology Evolution and Source Autonomy in Ontology-based Data Warehouses. 21 (2006).
22. Zablith, F.: Evolva: Towards Automatic Ontology Evolution. Technical report. Knowledge Media Institute (KMi) (2008).
23. Zablith, F. et al.: Ontology evolution: a process-centric survey. The knowledge engineering review. 30, 1, 45–75 (2015).
24. Zurawski, M.: Distributed multi-contextual ontology evolution–a step towards semantic autonomy. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 198–213 Springer (2006).