DILS 2017 Data Integration in the Life Sciences

Detailed Schedule

14th November

Time	Section
9:30	Welcome Coffee
10:00	Opening ceremony, Prof. Eric Dubois, Head of ITIS department of LIST Presentation of funding opportunities for research in Luxembourg by Michèle Weber (FNR)
10:20	A Simple Tool to Enrich Clinical Trial Data with Multiontology-Based Tags **Holger Stenzhorn** In this demonstration submission the re-engineering works on ObTiMA are presented which focus on 1) ease-of-use allowing clinicians to employ ontologies and concepts in their established work approaches with little training via a plain yet responsive user interface, 2) allowing the use of established ontologies in conjunction with semantic-web standard technologies to foster interoperability.
10:40	DyKOSMap : from a research prototype to a Web application Marwa Chaabane, Silvio Domingos Cardoso, **Cédric Pruski** and Marcos Da Silveira DyKOSMap is a research prototype developed in the framework of the DynaMO project. It aims to maintain existing mappings established between knowledge organization systems (KOSs) by taking into account information about the dynamics of these KOSs. The aim of this work is to implement a Web application, in order to make a more robust and easier-to-manipulate tool starting from this prototype.
11:00	Keynote: Robert Stevens University of Manchester Title: Ontologies and data integration: A state of nation
12:00	Lunch
14:00	Semantic Interoperability of Clinical Data **Jacqueline Santo** and Claudia Medeiros The interoperability of clinical information systems is particularly complicated due to the use of outdated technologies and the absence of consensus about standards. The literature applies standard-based approaches to achieve clinical data interoperability, but many systems do not adopt any standard, requiring a full redesigning process. Instead, we propose a generic computational approach that combines a hierarchical organization of mediator schemas to support the interoperability across distinct data sources. Second, our work takes advantage of knowledge bases to be linked to clinical data, and exploit these semantic linkages via queries. The paper shows case studies to validate our proposal.

14:30	Enabling Data Integration Using MIPMap **Giorgos Stoilos**, Despoina Trivela, Vasilis Vassalos, Tassos Venetis and Yannis Xarchakos In previous work we have analysed the infrastructure of the Human Brain Project Medical Informatics Platform focusing on the challenges related to data integration based on a visual data exchange tool, called MIPMap. In this paper we present new MIPMap features that enhance the integration process and data access.
15:00	Break
15:30	An Integrated Ontology-based Approach for Patent Classification in Medical Engineering **Sandra Geisler,** Christoph Quix, Rihan Hai and Sanchit Alekh Medical engineering is an interdisciplinary domain with short innovation cycles in which usually researchers from several fields cooperate in research projects. To support the identification of suitable partners for project, we present an integrated approach for patent classification that combines ideas from topic modeling, ontology modeling & matching, bibliometric analysis, and data integration. First evaluation results show that the use of semantic technologies in patent classification can indeed increase the quality of the results.
16:00	Evaluating and improving annotation tools for medical forms Ying-Chi Lin, **Victor Christen**, Anika Gross, Silvio Domingos Cardoso, Cédric Pruski, Marcos Da Silveira and Erhard Rahm The annotation of entities with concepts from standardized terminologies and ontologies is of high importance in the life sciences to enhance semantic interoperability, information retrieval and meta-analysis. Unfortunately, medical documents such as clinical forms or electronic health records are still rarely annotated despite the availability of some tools to automatically determine possible annotations. In this study, we comparatively evaluate the quality of two such tools, cTAKES and MetaMap, as well as of a recently proposed annotation approach from our group for annotating medical forms. We also investigate how to improve the match quality of the tools by post-filtering computed annotations as well as by combining several annotation approaches.
16:30	End
16:40	Departure from the hotel to the guided tour
17:00	Guided tour
19:30	Gala Diner (Hotel Royal)

15th November

Time	Section
9:30	Interactive Map Visualization System Based on Integrated Semi-Structured and Structured Healthcare Data **Milena Frtunić Gligorijević**, Darko Puflović, Evgenija Stevanoska, Tatjana Jevtović Stoimenov, Goran Velinov and Leonid Stoimenov Big data in the healthcare industry is overwhelming, not only because of its volume but also because of its variety. In order to use such data, it needs to be pre-processed and integrated first. An additional problem is the visualization of such big data and making it valuable, readable and easier to come to the conclusions. This paper presents a system that uses interactive maps for presenting data and services for integrating healthcare data and combining it with other external sources. The purpose of this system is to show a presence of some disease in the country, how many patients with that diagnosis had to travel to some other location in order to get the medical examination and how far they had to go. Such information can be valuable in process of organizing and optimizing healthcare resources and creating models for cheaper and more optimal healthcare both from system's and patient's perspective.

10:00	Layout aware Semi-automatic Information Extraction for Pharmaceutical Documents **Simon Harmata**, Katharina Hofer, Phuong-Ha Nguyen, Christoph Quix and Bujar Bakiu Pharmaceutical companies and regulatory authorities are also affected by the current digitalization process and transform their paper-based, document-oriented communication to a structured, digital information exchange. The documents exchanged so far contain a huge amount of information that needs to be transformed into a structured format to enable a more efficient communication in the future. In such a setting, it is important that the information extracted from documents is very accurate as the information is used in a legal, regulatory process and also for the identification of unknown adverse effects of medicinal products that might be a threat to patients' health. In this paper, we present our layout-aware semi-automatic information extraction system LASIE that combines techniques from rule-based information extraction, flexible data management, and semantic information management in a user-centered design. We applied the system in a case study with an industrial partner and achieved very satisfying results.
10:30	Break
11:00	A Conceptual Approach for Modelling Social Care Services: the INSPIRE project **Elaheh Pourabbas**, Antonio D'Uffizi and Fabrizio L. Ricci This paper proposes a conceptual model of social care services that takes into account the different configurations of service setting and activities. This model encompasses a multidimensional conceptualization of social and health care service integration and considers those characteristics that are relevant to answer a fragile user's need. On the basis of the proposed model, this paper outlines the methodological considerations with regard to information management, modelling and design of data related to social care services. To this end, a taxonomy of services categorized according to different targets, i.e., elders and people with disabilities, is presented. Our model that is based on the UML standard is going to be used within the INSPIRE project to design and develop its required goals.
11:30	Variant-DB: a tool for efficiently exploring millions of human genetic variants and their annotations **Joachim Kutzera** and Patrick May Next Generation Sequencing (NGS) allows sequencing (determining the precise order of nucleotides) of a human genome within hours, enabling large scale applications such as sequencing the genome of each patient in a clinical survey. Each individual human genome has about 3.5 Million genetic differences to the so called reference genome, the consensus average genome of a healthy human. These differences, called variants, determine individual phenotypes, and certain variants are known to indicate disease predispositions. Finding associations from variant patterns and affected genes to these diseases requires combined analysis of variants from multiple individuals and hence, efficient solutions for accessing and filtering the variant data. We present Variant-DB, our in-house database solution that allows such efficient access to millions of variants from hundreds to thousands of individuals. Variant-DB stores individual variant genotypes and annotations. It features a REST-API and a web-based front-end for filtering variants based on annotations, individuals, families and studies. We explain Variant-DB and its front-end and demonstrate how the Variant-DB API can be included in data integration workflows.
12:00	Lunch

Social Event

Guided Tour: On Tuesday 14th/Nov, from 18:00

Enjoy walking around the city during 1,5 hours and discover the most beautiful corners and attractions of the city centre in company of an experienced guide: Place d'Armes, Constitution Square, Government District, Corniche, Monument of the Millenium, Old Town, Palace of the Grand Dukes (outside), Place Guillaume II.

- Departure: 17:00 at Constitution Square
- *Or, we will leave at 16:40 from the Parc Belle-Vue Hotel*
- End: 19:00 at Hotel Le Royal

Conference Dinner: On Tuesday 14th/Nov evening, at Hotel Le Royal

This restaurant proposes gourmet cuisine with many traditional French plates and typical Luxembourgian specialties.

- Aperitif 19:00
- Start: 19:30
- Address: 12, boulevard Royal | L-2449 LUXEMBOURG