HEALTH DATA INTEREST GROUP

(HD-IG)

Here follows a **Proposal for an Interest Group** on **"Health Data" (HD-IG)**, as a long-term initiative in the framework of **RDA**, Research Data Alliance (https://rd-alliance.org). It follows a rather successful BoF Session during the 6th RDA Plenary Meeting in Paris, which was attended by over 35 researchers and professionals from diverse backgrounds, who discussed several relevant issues, expressed significant interest in forming the proposed Interest Group, and helped shape its focus as presented here. The Interest Group will fill a gap in the RDA subject map formed by its current WGs and IGs, as at the moment, there is no RDA group focusing on the intricacies of Health Data, especially as it relates to privacy and security issues in Healthcare. Establishment of this IG will also enrich the set of communities involved in and contributing to RDA, as there are several professions as well as research disciplines that revolve around Health Data.

Timeline: we're looking forward to having established the IG before the 7th RDA Plenary Meeting, 1-3 March 2016 in Tokyo, so that we may have the first meeting of the group then.

BACKGROUND

This proposal to form the HD-IG is rooted in a long series of European, international, and national projects in the area of biomedical informatics, in which the proposers have been involved in the past decade. These include projects Health-e-Child (www.health-e-child.org), Sim-e-Child (www.sim-e-child.org), MD-Paedigree (www.md-paedigree.eu), p-medicine (www.p-medicine.eu), and others. Different techniques of de-identification were adopted (pseudonymisation and anonymisation) and ad-hoc privacy guidelines were developed during these projects, not only to meet the requirements of the in-force legislation but also to face future challenges in the possible exploitation of the projects. The scientific and practitioner community developed during these projects is quite extensive and several members of them are expected to join HD-IG.

CHARTER

Data based Healthcare characterizes a fundamental shift in the way biomedical data are collected and processed, as well as how biomedical research is performed. The application of data techniques in Healthcare will allow us to capitalise on growing patient and health system data availability and generate healthcare innovation. However, to bring about this revolution in healthcare there are legal, technical and cultural/societal barriers that must be overcome. The proposed "Health Data" Interest Group (HD-IG) seeks to bring together

stakeholders from all relevant sides and provide a forum for discussion on the specific issues that arise when using advanced data management and analytics techniques in a Healthcare setting, particularly (although not exclusively) focusing on the impact of privacy and security concerns.

Bottom-up (evidence-oriented) analysis, seeking to extract useful knowledge by mining the daily routine's streaming data, is of fundamental interest in model-guided personalized medicine. In this context, advanced techniques are applied aiming to identify latent factors (disease signatures) that can explain and predict variability in drug therapies and disease evolution, reveal similarities among patients stratifying patient groups and build patient specific simulation and prediction models. Such an approach goes beyond classical flat file data analysis, batch learning procedures, and simple data analysis techniques commonly focused only on a few variables of interest and a well specified dataset from a specific clinical trial. On the contrary, Knowledge Discovery and Data Mining (KDD) platforms in this area should be able to handle massive volumes of uncertain, streaming heterogeneous biomedical data, to curate, validate and analyze them in an incremental/on-line fashion from multiple points of view and under different assumptions, as well as to include or exclude dimensions, combine different modalities and incorporate existing knowledge and previous beliefs, all while preserving the privacy of the patients whose data is being analysed..

The still growing potential of modern data management and analysis is today fully acknowledged, but it may remain partly undeveloped or lead to undesirable outcomes or misuses of data if not carried in parallel with a deeper understanding of the regulatory and legal challenges it poses to patients' privacy and data protection. At the same time, harming innovation and putting restrictions on research should be avoided. Indeed, as debates and proposals held in different countries show (such as on a "Magna Charta for Big Data")[1], there's a societal need for a more adequate legislative framework for ethical leveraging of data applications, balancing the needs and rights of data providers and owners.

It must be remarked that privacy and regulatory issues related to the process of "data-intensive scientific discovery" are a greater matter of attention for the EU, in particular in view of the new General Data Protection Regulation, which is expected to determine a more comprehensive legal framework to refer to, with the aim to strengthening individuals' trust and confidence in the digital environment and enhancing legal certainty. The EU debate on data policies turns around three core themes: the need to ensure that citizens' data are adequately protected; the need for Open Access to data for research purposes; the need to develop a vibrant Data Value industry in health, enabling EU to remain a major competitor

[1] A "Magna Carta for Data" was proposed and discussed during a seminar hosted by the Insight Centre for Data Analytics titled "Insight: Frontier Data Analytics: Towards a Magna Carta for Data", held in Brussells on the 4th February 2015. The discussion document is available on the following link: https://www.insight-centre.org/sites/default/files/basic_pages_file/magna_carta_for_data_pdf.pdf.

in this field. As a consequence, it is necessary to strike the appropriate balance between individual privacy concerns in the healthcare setting and research purposes and innovation, which can greatly benefit patients.

Given the lack of legal harmonisation at the EU level and the different national implementations of data protection, unless the Regulation enters into force as much as more comprehensive global standards of protection, different approaches and protocols will be adopted (many of them in accordance to HIPAA, which still is, as yet, the strongest worldwide de-identifying constraint expression). Comparing and discussing these approaches is a fundamental need for the improvement of data technology in Healthcare.

The HD-IG will provide its members with a forum to discuss and highlight the legal, technological, ethical and societal challenges to the adoption of advanced data management and analysis techniques in Healthcare, to exchange opinions and compare experiences, and form Working Groups to address these challenges.

TOPICS OF INTEREST

The initial focus of the HD-IG includes the following topics:

- Data access and protection
    - sharing best practice on pseudonymisation and anonymisation
    - developing models for consent that protect patients while enabling research
    - providing a forum for discussing, explaining and responding to data protection regulation
    - secure opening up of data to facilitate research
- Data-based Healthcare for Personalised Medicine
    - disease signatures identification
    - stratification of patient groups
    - patient-specific simulation and prediction
- Data literacy in Healthcare
    - providing materials for education of healthcare professionals on use and misuse of data
- Patient data repositories
- *In silico* drug development and clinical trials
- Policy making
    - representing interests of the data-based healthcare community to policy makers
    - identifying and discussing related challenges, interdisciplinary research needs and potential roadmaps

PARTICIPATION

HD-IG is open to all RDA members to participate. Particularly, but not exclusively, HD-IG welcomes individuals with the following expertise to actively participate in its activities:

- Clinicians wanting to use data technology to improve practice
- Biomedical researchers using data heavy analytical techniques
- Healthcare Data Analytics with data mining, machine learning, physiological modelling and image processing expertise
- HPC and distributed computing experts
- Policy-makers for Healthcare
- Health bioinformatics legal experts
- Healthcare administrators and Health Maintenance Organisations
- Pharmaceutical industry researchers and manufacturers
- Medical equipment researchers and manufacturers
- *In silico* modelling, testing and clinical trial experts

A quick survey during the P6 BoF Session identified participants with all but the last expertise in the above list, an indication of both the diversity of relevant stakeholders and the strength of current interest in the focus areas of HD-IG.

COLLABORATION

The discussions during the P6 BoF Session shifted its original focus from "Big Health Data" to simply "Health Data" and intensified the already high importance of privacy and security aspects in the field. This proposal reflects the conclusions of those discussions and concentrates on a rich set of issues that are critical for Health Data and are not covered by any currently active RDA IGs. On the other hand, HD-IG will seek to collaborate with those IGs that have affinity to aspects it will address, as well as with external organisations. Below is a list of RDA IGs that are of particular relevance to HD-IG. The leaders of some of them were among the participants of the P6 BoF Session:

- Active Data Management Plans
- Big Data
- ELIXIR Bridging Force
- Ethics and Social Aspects of Data
- Long tail of research data
- RDA/CODATA Legal Interoperability
- Structural Biology

Interest Group Chairs

Anthony Chang, Children's Hospital of Orange County (USA)
(http://www.choc.org/findadoc/index.cfm?id=P00348&pid=1142)

Yannis Ioannidis, Athena Research & Innovation Center and University of Athens (GR)
(http://www.di.uoa.gr/~yannis)

Edwin Morley-Fletcher, Lynkeus Srl, Rome (IT)
(http://www.lynkeus.eu)

Timos Sellis, RMIT University, Melbourne (AUS)
(http://www1.rmit.edu.au/staff/timos-sellis)

Supporters

Bruno Dallapiccola, Ospedale Pediatrico Bambino Gesù (IT)
(http://www.palazzochigi.it/bioetica/eng/curriculum/Dallapiccola.pdf)

Laura Palazzani, Italian National Bioethics Committee (NBC)
http://www.governo.it/bioetica/eng/curriculum/Palazzani_14-10-13.pdf

Norbert Graf, University of Saarland (DE)
(http://www.uniklinikum-saarland.de/einrichtungen/kliniken_institute/kinder_und_jugendmedizin/klinik_fuer_paediatrische_onkologie_und_haematologie/)

David Manset, Gnùbila (FR)
(https://gnubila.fr/)

Patrick Ruch, University of Applied Sciences Western Switzerland (CH)
(http://bitem.hesge.ch/content/patrick-ruch)

Hans Westerhoff, Universiteit van Amsterdam and University of Manchester (NL & UK)
(http://www.uva.nl/en/contact/staff/item/h.v.westerhoff.html?f=westerhoff)

Ludovica Durst, Lynkeus Srl (IT)
(http://www.lynkeus.eu/company-profile/)

… and approximately 35 other professionals who attended the P6 BoF Session.