

## **LERU Summer school 2016**

*Data Stewardship for Scientific Discovery and Innovation*

Leiden, 10-15 July 2016

### **Organizers**

Prof. Barend Mons - Department of Human Genetics, LUMC, BioSB Research School

Prof. Ton Raap - LUMC Graduate School and Department of Molecular Cell Biology, LUMC

### **Theme**

Existing methods of scholarship, scientific discovery and communication (publishing research articles) have been in place for hundreds of years. But as the amount of data being generated, either collectively or in any single study increases, methods that worked so well before, are now impeding the discovery potential latent within large and heterogeneous data. Out-dated methodologies for and incentives behind data generation, capture, processing, interlinking and analysis make data hard to find, hard to integrate and impossible to analyze. Without effective data management plans and data stewardship protocols, valuable data sets are increasingly at risk of being lost.

Robust data stewardship and data publication in machine-readable format is swiftly becoming an essential skill for participation in professional science, and are becoming more and more the responsibility of the data-generator whether they come from academic research or private companies and governmental organizations. Beyond technical challenges, there are also interdependent ethical, legal and societal issues at stake.

The LERU Summer School *Data Stewardship for Scientific Discovery and Innovation* will prepare its participants for the future of complex, data-intensive, multi-stakeholder and multi-disciplinary scientific methods and collaborations.

### **Participant requirements**

The LERU Summer school *Data Stewardship for Scientific Discovery and Innovation* is intended for PhD students of LERU partners (2 per partner) Candidates for the summer school will be nominated by their institution. They should be well advanced in their doctoral project and keenly motivated to work on the topic of the summer school (data stewardship). Candidates must be available for the entire duration of the summer school, and must be prepared to participate in all activities. All sessions will be conducted in English, and therefore fluency in both spoken and written English is essential.

### **Objective**

Introduce next generation scientists to emerging methods of data-intensive research and discovery.

### *Learning Objectives*

- Gain a broad understanding of the many complex issues surrounding robust data stewardship practices and how this is necessary for optimal discovery in data-intensive sciences (covered in 5 Keynotes).
- Gain insight in the (emerging) responsibilities of academic researchers as ‘data generators’ and how the re-usability and reproducibility of data is impacting scientific methods. (also mainly in the Keynotes)
- Gain experience in machine-assisted knowledge discovery in large and complex datasets. Examples will be drawn from biomedical, omics, and clinical research, law, and literature as well as current trends in social networks, the quantified self and the internet.
- Exposure to emerging principles, protocols, standards and formats related to both technical and social (including ELSI) mechanisms of data stewardship, including practical sessions using standards, technologies and protocols.
- Practical exposure to real-world examples of data capture, processing, interlinking and the subsequent pattern recognition in interlinked massive data with existing tools.
- Exposure to current trends in academic funding that increasingly require investigators to formulate data stewardship plans as an intrinsic part of modern research proposals.
- Skill building for writing grounded and comprehensive data stewardship plans for data intensive project proposals or investment plans, that make data Findable, Accessible, Interoperable and Re-usable (FAIR) for both human and machine.
- Gain critical understanding about the ‘value of data’ in the distinct, but interrelated processes of discovery, innovation and valorisation.

### **Participation costs**

The costs for the summer school are € 500 per participant. This includes course materials, accommodation for 5 nights from 10 -15 July (Sunday night – Friday morning) and meals (Monday Breakfast –Friday Lunch).

### **Additional costs**

Travel to and from Leiden is not included in the summer school cost. Participants will also be expected to cover the cost of any additional social activities undertaken.

## **Program overview**

Day 1. Historical Overview & Data Capture

Take-away: “We are the new generation of information scientists.”

Evening Event: Rector and Deans reception

Day 2. Data Linking

Take-away: Making data ‘sharable’: Pre-processing and quality control

Evening Event: Welcome reception in Town Hall offered by the Mayor of Leiden

Day 3. Publication and Curation of Data

Take-away: New modes of digital publication, communication, exchange

Evening Event: Tour of the LUMC Anatomical Museum

Day 4. Analysis Across Data Sets

Take-away: What becomes possible when data is interoperable

Evening Event: Official dinner (Boerhaave Museum)

Day 5. Data Ecosystems

Take-away: Putting it all together: Data ecosystems

Farewell

## Program outline

### Day 0. Arrival, Evening Informal get together at Hotel lobby

Prof. Ton Raap (Leiden University, LUMC)

### Day 1. Historical Overview & Data Capture

Take-away: "We are the new generation of information scientists."

*Opening remarks* Prof. Ton Raap and Prof. Barend Mons (LUMC)

*Morning Plenaries & Keynote*

- Growth of knowledge [historical context] - Dr. Erik Schultes (LIACS) -
- Data stewardship & FAIR Data [current trends in data science] - Prof. Barend Mons
- **Keynote:** -Big Science is Open Science [scientific questions beyond the scope of individual labs and experts] - Dr. Myles Axton (NPG Nature Genetics) (confirmed)

Lunch

*Afternoon Workshop*

- Ontology engineering [what is an ontology, vocabulary, mark-up languages] - Dr. Luiz Bonino (VU Amsterdam, DTL)
- Demos tools for automatic semantic mark-up [Rightfield Google Refine] - Dr. Katy Wolstencroft (LIACS)

*Summary and preview day 2* Prof. Barend Mons

*Evening Event* Rector & Deans reception

*Home Work* Engineer an ontology for your data / write a nanopublication

### Day 2. Linking Data

Take-away: Making data 'sharable'

*Morning Plenaries & Keynote*

- Linking data & BYODs - Dr. Marco Roos (LUMC)
- Recent advances at Elsevier Labs - Dr. Paul Groth (Elsevier) (to be confirmed)
- **Keynote:** Rise of the Social Machines— Prof. Frank van Harmelen (VU) (to be confirmed)

Lunch

*Afternoon Workshop*

- Data integration & licensing: The legal issues - Timothy Vollmer, Manager, Public Policy (Creative Commons) (to be confirmed)
- Big Data and Judicial Decisions - Prof. Jaap van den Herik (LCDS) (confirmed) and Rob van Eijk, E Law@Leiden Center For Law And Information Society

*Presentations by LERU students*

*Summary and preview day 3* Prof. Barend Mons

*Evening Event* Welcome reception in Town Hall by mayor of Leiden

*Home Work* License your data

### **Day 3. Publication and Curation of Data**

Take-away: New modes of digital communication and exchange

*Morning Plenaries & Keynote*

- Nanopublications - Dr. Tobias Kuhn (ETH) (to be confirmed)
- Huntington Disease: a case study in data publication - Eleni Mina (LUMC)
- **Keynote:** Prof. Carole Goble (Univ Manchester, ELIXIR, FAIRdom)

Lunch

*Afternoon Workshop*

- Behind the scenes of the FAIR Skunk Team - Prof. Mark Wilkinson (Spain) (to be confirmed)
- Manual, machine & community annotation; annotator tools demo - Dr. Mark Thompson (LUMC)

*Presentations by LERU students*

*Summary and preview day 4* Prof. Barend Mons

*Evening Event* Tour of the LUMC Anatomical Museum

*Home Work* Link your data sets

### **Day 4. Analysis Across Data Sets**

Take-away: What becomes possible when data is interoperable

*Morning Plenaries & Keynote*

- The Monarch Initiative: Semantic Phenotyping for Disease Diagnosis and Discovery - Prof. Melissa Haendel (OHSU Library) (to be confirmed)
- The Euretos Knowledge Platform - Arie Baak (Euretos) (to be confirmed)
- **Keynote:** Dr. Niklas Blomberg (director ELIXIR) (to be confirmed)

Lunch

*Afternoon Workshop*

- Gene-Disease Associations: Concept profiles & the implicitome - Dr. Kristina Hettne (LUMC)
- IBM Watson Team (to be confirmed)

*Presentations by LERU students*

*Summary and preview day 5* Prof. Barend Mons

*Evening Event* Official dinner (Boerhaave Museum)

### **Day 5. Data Ecosystems**

Take-away: Putting it all together: Data ecosystems

*Morning Plenaries & Keynote*

- The Astrophysics Data ecosystem: an example of how to share data and do world class science - Dr. Lourdes Verdes-Montenegro (Instituto De Astrofisica Granada, Andalucia, Spain) (to be confirmed)
- Life Science Data Stewardship can promote Scientific Discovery and Innovation in Developing Nations - Prof. Barend Mons

- **Keynote Closure:** The future infrastructure for Open Science - Prof. Paul Ayris (LERU)