



Universiteit
Leiden
The Netherlands



Data Stewardship for Scientific Discovery and Innovation

LERU Summer School 2016

Leiden, 10-15 July 2016

Organizers

Prof. Barend Mons, PhD, Department of Human Genetics, LUMC and BioSB Research School
Prof. Ton Raap, PhD, LUMC Graduate School and Department of Molecular Cell Biology, LUMC

Context

Sharp formulation of scientific questions, well-designed experimental set-ups, sound methodologies for data generation, -capture, -processing, -interlinking and -analysis as well as critical interpretation of results are key to the process of scientific discovery and innovation.

Given the current and ever increasing ability of the experimental sciences to rapidly generate *and* virtually instantaneously share and link vast amounts of data from a variety of sources, proper data stewardship is swiftly becoming an essential responsibility of experimental and data scientist, if not all researchers (see also the [LERU Roadmap for Research Data](#)).

In taking this responsibility they face intellectual challenges that are too complex to be mastered by their individual expertise. Consequently, the scientists of the future should solve big data related challenges as respectfully collaborating specialists.

Dealing with the high complexity of the data sources and the opportunity to grant (open) access under a responsible data stewardship policy is a prerequisite for academic research, but equally so for private companies and governmental and non-governmental organizations. Evidently for all these sectors, ethical, legal and societal issues of data stewardship are at stake.

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

Target Groups

The LERU Summer School *Data Stewardship for Scientific Discovery and Innovation* is intended for PhD students of LERU partners (2 per partner with selection by partner institutes). Candidates from any discipline are invited to apply.

Overall Objective

To gain thorough theoretical understanding of and extensive practical experience (Bring Your Own Data, BYOD!) with the data stewardship and discovery cycle so as to be able -in a joint effort of experimental, data scientists, IP- and research integrity specialists- to define for a given research project within a particular sector, a technically sound and ethically responsible data stewardship format. In addition the use of data after completion of the project and the use of those data for actual innovation will be part of the skills set acquired.

Learning Objectives

1. A thorough understanding of the complexity of adequate data stewardship and optimal discovery in 'science 2.0' (covered in 5 Keynotes)
2. Clear insight in the (changed) responsibilities of academic researchers as 'data generators' and how the re-usability and reproducibility of experiments is impacting the running of scientific experiments (also mainly in the Keynotes)
3. A deep understanding of the real issues pertaining to the shift to machine-assisted knowledge discovery in large and complex datasets as present in digital archives of history, law and literature or generated in e.g. socio-demographic-, (bio)medical 'omics' studies and clinical trials
4. An insight how current (publishing) practices impair the above process: non-findability, non-citability, lack of incentives for data sharing, non-interoperability, licensing issues of datasets etc.)
5. Proficiency in the needs and possibilities of standards, protocols, formats and approaches (including ELSI aspects) related to the full data stewardship and discovery cycle (practical sessions using standards, technologies and protocols on exemplar data)
6. A deep understanding of the new ways in which data are published as a 'first class citizen' object in contemporary science, the complexity and the costs involved to make data Findable, Accessible, Interoperable and Re-usable (FAIR) for both human and machine users and how funders increasingly demand this to be an intrinsic part of modern research and research funding
7. Hands-on experience in (exemplar) data capture, processing, interlinking and the subsequent pattern recognition in interlinked massive data with existing tools (practical sessions, including BYOD type of sessions)
8. The skill to write a grounded and comprehensive data stewardship plan for future grants, in supported writing sessions (creating a data stewardship plan for a fictive or real scientific experiment/project)
9. A clear understanding about the 'value of data' both in the discovery and in the subsequent innovation and valorisation processes and understanding of the significant differences and confusions about discovery, innovation and valorisation as distinct, but interrelated processes.