

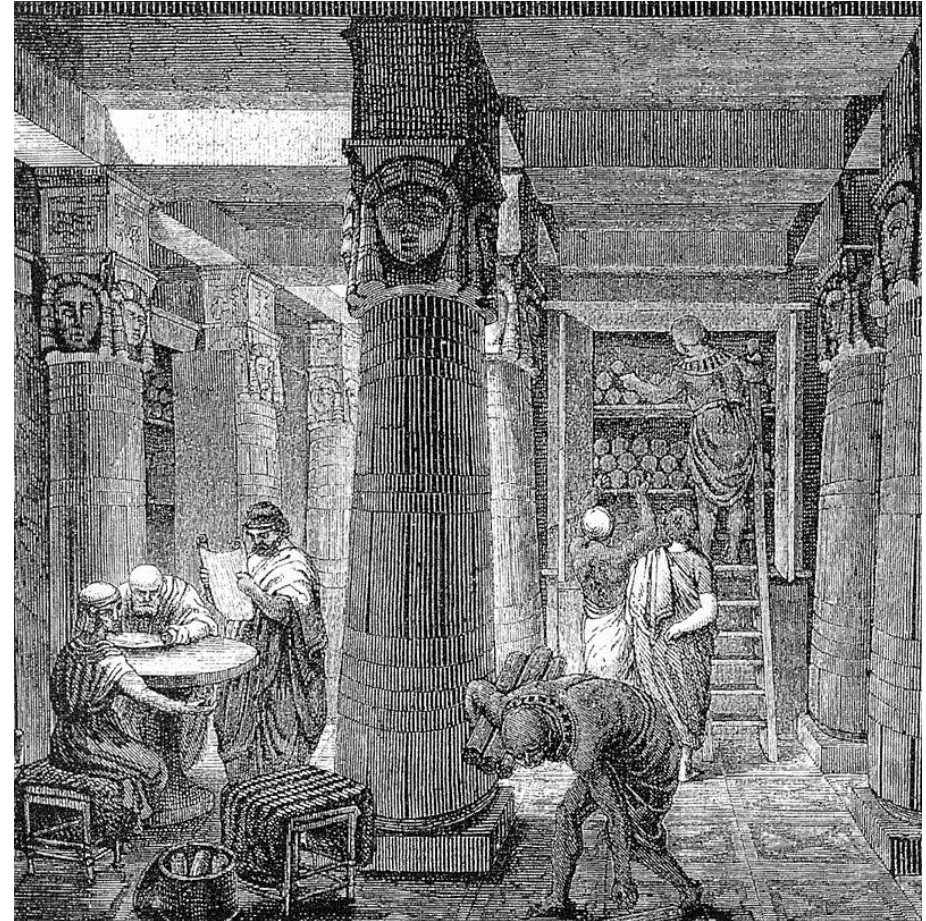
Facing the challenges of research data management

Lessons learnt at KU Leuven libraries

Tom Willaert, Jacob Cottyn, Simon D'haenens, Ulrike Kenens, Thomas Vandendriessche

Introduction

- Organizing, transferring and preserving knowledge @ Library of Alexandria (3rd century BC)
 - Philosophers and scholars
 - Scribes
- Bridging gaps between
 - Knowledge management strategies (taxonomies, ontologies)
 - Knowledge management practices (scholarly primitives such as writing and transcribing)



Introduction

- Contemporary knowledge culture
 - Debates concerning openness and transparency
 - open science, open data
- Open scholarship

[...] requires a culture change in the way stakeholders in the research, education and knowledge exchange communities create, store, share and deliver the outputs of their activity.



Introduction

- EC's pillars of open science
 - Culture built around 'data'
- Research data
 - Not as ontological fact
 - But as indicator of (new) strategies and practices for dealing with resources (Niels-Oliver Walkowski)
 - eScience, big data, machine learning, advanced analytics

Introduction

- Well-implemented strategy for research data management (RDM) as practical counterpart
 - Digital curation center: ‘maintaining, preserving and adding value to digital research data throughout its lifecycle
 - KU Leuven: also physical data
- Data life cycle
 - From data management strategy to implemented, sustainable business plan for dealing with active data, as well as storing, archiving and cataloguing data at the end of a research project

Introduction

- Economic, ethical and scientific benefits of RDM
 - Avoid duplicate data
 - Open data to scrutiny and criticism
 - Make researchers accountable for findings
 - Increase findability, accessibility, interoperability, reusability of data (FAIR)

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtis.nl).
#A full list of authors and their affiliations appears at the end of the paper.

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

1

Introduction

- RDM requires infrastructure ('infrastructuring')
 - Interpreting guidelines
 - Developing best practices
 - Writing DMPs
 - Providing training
 - Studying requirements for IT-infrastructure
 - Etc.
- Techno-cultural setup with many stakeholders and partners, including libraries
 - Experience in data and information management
 - Structuring data
 - Assigning metadata
 - Developing and engineering ontologies
 - Information retrieval
 - Scholarly communication and open access

Introduction

- KU Leuven
 - Research Coordination Office
 - KU Leuven Libraries
 - ICT services
 - Legal department
 - Research and development
 - LIBIS
- Involvement since 2015, matured into concrete plan of action in strategic plan
 - Library experts + network of data stewards



Introduction

- Experimental case: 2018 call for proposals of Research Foundation Flanders (FWO)
 - How can facilitating RDM-support leverage research libraries' position of leadership in the context of open scholarship? What value can be created by evaluating DMPs and surveying researchers from various career stages?
 - How can RDM-support and a data-oriented perspective cement the research library's position as an educational centre, notably with regards to data and information literacy and interdisciplinary programmes such as 'digital humanities'?
 - To what extent can experiences in the field of RDM be used as a measure for the validity and applicability of such theoretical principles and guidelines as 'FAIR' data', or frameworks such as RISE?

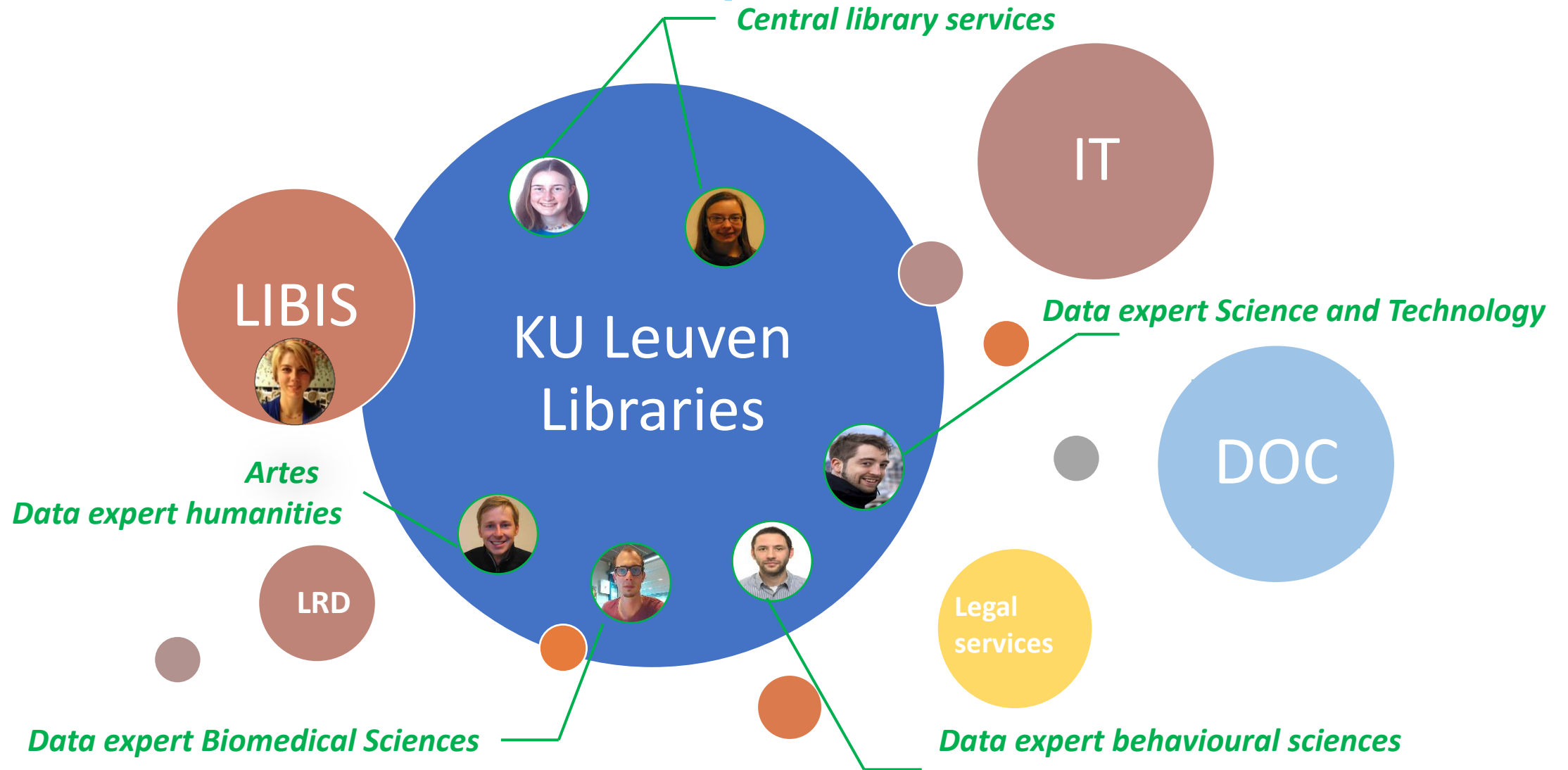
Open science, leadership and innovation

- Adapt services and organizational structure to changing circumstances
- KU Leuven
 - Research data policy established December 2014
 - Data should be made available upon request by KU Leuven in the course of research and should be retained for period of at least five years after publication or after the end data of the research grant agreement
 - DMP for projects with substantial internal funding
 - Ties in with Horizon2020 and Research foundation Flanders

Open science, leadership and innovation

- Division of tasks
 - Research Coordination Office: funder compliance + liaises with Ethical Committees
 - Central and Local IT-services provide infrastructure and guidance for gathering, processing and storing active research data
 - Legal department offers tailored advice on processing personal data under GDPR and handles questions on ownership
 - Technology Transfer Office (Leuven Research & Development) takes care of agreements of third parties
- KU Leuven Libraries
 - Generic and discipline-specific advice on data management planning, data classification and metadata collection
 - Together with Research Coordination Office: highly engaged in training researchers via online resources and workshops
 - Library systems developer LIBIS specializes in tailored solution for preservation, access, publishing and discovery of research data

Open science, leadership and innovation



Open science, leadership and innovation

‘Layer cake’

- Central support desk
- Network of Data Experts
- Network of Data Stewards



Open science, leadership and innovation

- Take-aways

- No longer exclusively passive supplier of research literature, but active involvement in all phases of research life cycle
- Data experts and stewards gain valuable oversight of the research projects being set up in their departments
- Provide valuable feedback on strategies and infrastructures developed at central level
- While performing traditional support role, libraries put themselves at the forefront of innovation, thus flipping the model on its head

Education and training

- Importance of interpersonal networks and collaboration for cultural shift
 - Heavy investment in networks of knowledge transfer and training
 - Inspired by data stewards at other universities
- Philosophy of 'train the trainer'
 - Data stewards come from heterogeneous backgrounds
 - No new staff was hired

Education and training

- Internal training
 - Introduction to scholarly landscape, funder requirements, benefits of RDM
 - Specific information about FWO questions
 - No templates, but 'on the job' learning (together with data experts)
 - Effective method for training heterogenous group of data stewards
 - Different backgrounds
 - Already took up tasks and responsibilities
 - Varying degrees of familiarity with research methods and practices

Education and training

- External training
 - Optional workshop for PhD students in which the basic principles of RDM are explained and illustrated by the library's data experts.
 - Optional workshop on peer reviewing DMPs that is taught by data experts and members of the Research Coordination Office
 - Optional session on publishing data and research funding provided by the Research Coordination Office
 - Mandatory session on research integrity that also features basic guidelines on research data management

Education and training

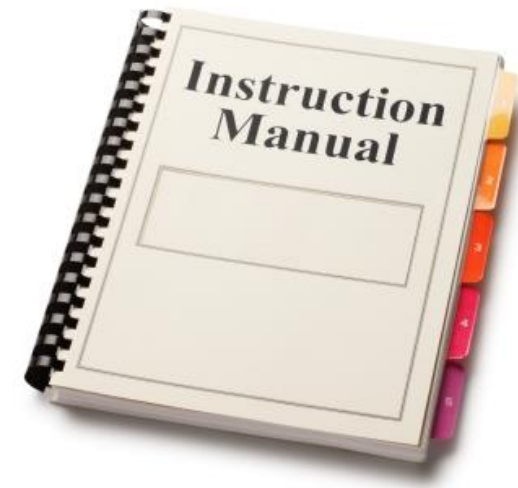
- Curriculum
 - part of the Doctoral School of Biomedical Sciences,
 - Doctoral training programme in digital humanities at the Faculty of Arts, undergraduate course on data processing, course on linked data scholarship in advanced masters in DH
- Data life cycle and RDM as instruments to structure curricula in data- and information-intensive programmes

Education and training

- Take-away
 - Offering trainings, information sessions and generally teaching others about how to deal with data and research data in particular provides information professionals with an important opportunity to take up a position of leadership, as it enables them to help educate the researchers of tomorrow.

Data stewardship principles and guidelines 'in the field'

- Changes on many fronts and at different speeds
 - Definitive guide on how to implement RDM support still has to be written
 - Useful guidelines, principles and frameworks to bridge gaps between research data policy and practice



Data stewardship principles and guidelines 'in the field'

- RISE (Research Infrastructure Self-Evaluation Framework)
 - Grade maturity of infrastructure
 - Range of topics
 - RDM Policy & strategy, Business plans and sustainability, Advisory Services, Training, Data management planning, Active data management, Appraisal and risk assessment, Preservation, Access, publishing and discovery
- Best practices at other universities
 - Conference visits and other networking events
 - Institutional websites

Data stewardship principles and guidelines 'in the field'

- Philosophical inspirations
 - Open Science (LERU roadmap)
 - FAIR data principles
 - Anticipate challenges of doing research alongside machines, finding a middle ground between diversity of data and standardization
- Benefits
 - Acknowledge that cultural shift is imminent and required
 - Confirm that data stewardship is a profession in its own right
 - Can convince policy makers and research funders to support RDM initiatives

Conclusions

- Insight into the process of innovation that was required to implement research data support at KU Leuven Libraries
- Try and do justice to multi-faceted nature of building infrastructure at the core of fundamentally new way of doing scholarship and science
- Cultural shifts take time to occur and to a large extent need to come from bottom-up
- Diverse stakeholders requires trade-off between flexibility and standardization
- Not just future oriented, also to justice to legacy ('analog') data

Q&A