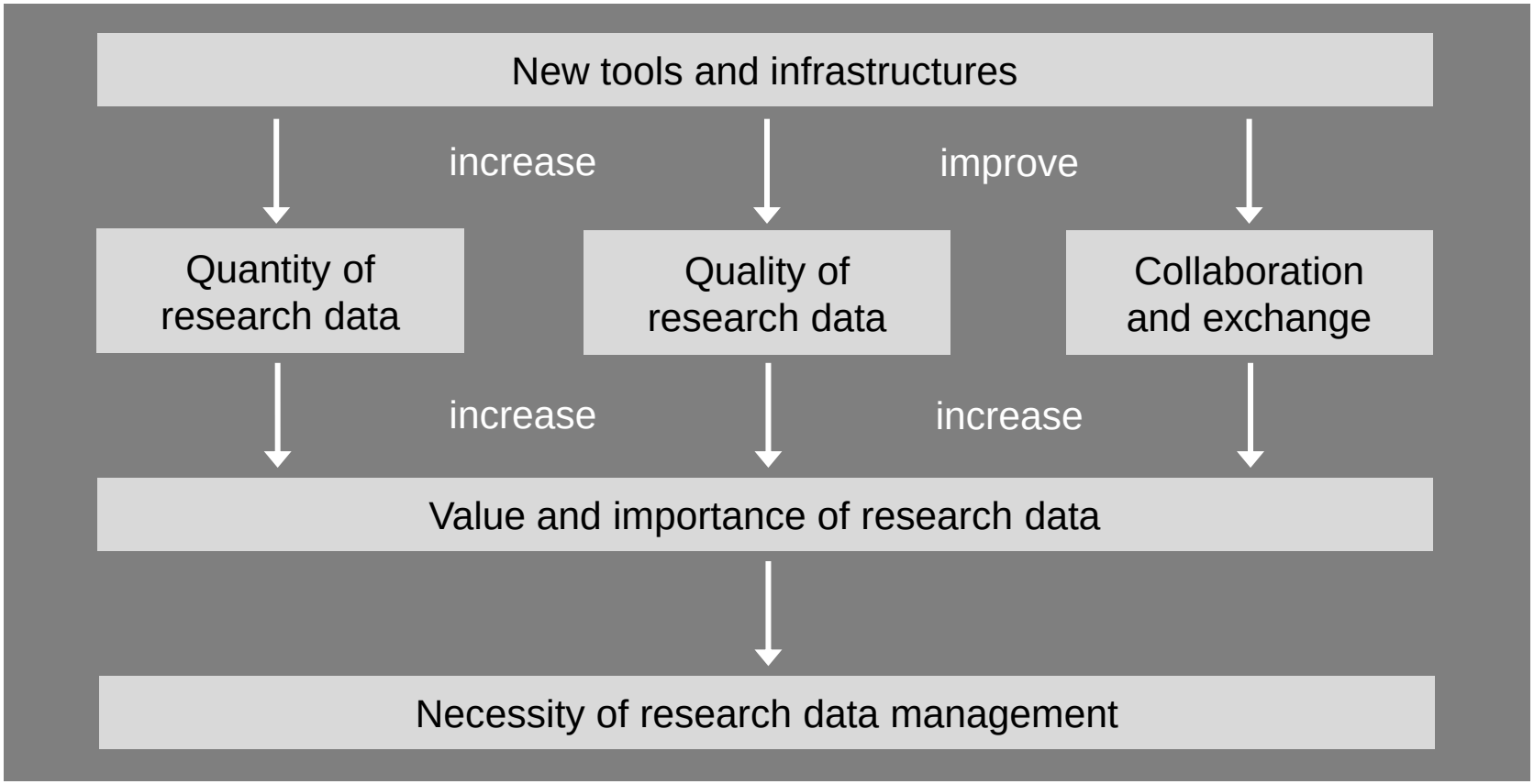
How Do Researchers Manage Their Data?

Anne Thoring, Dominik Rudolph and Raimund Vogl
WWU Münster

Research Data Management

Why is it necessary to manage research data professionally?



EUNIS THESSALONIKI 2016

Westfälische Wilhelms-Universität Münster

Research Data Management

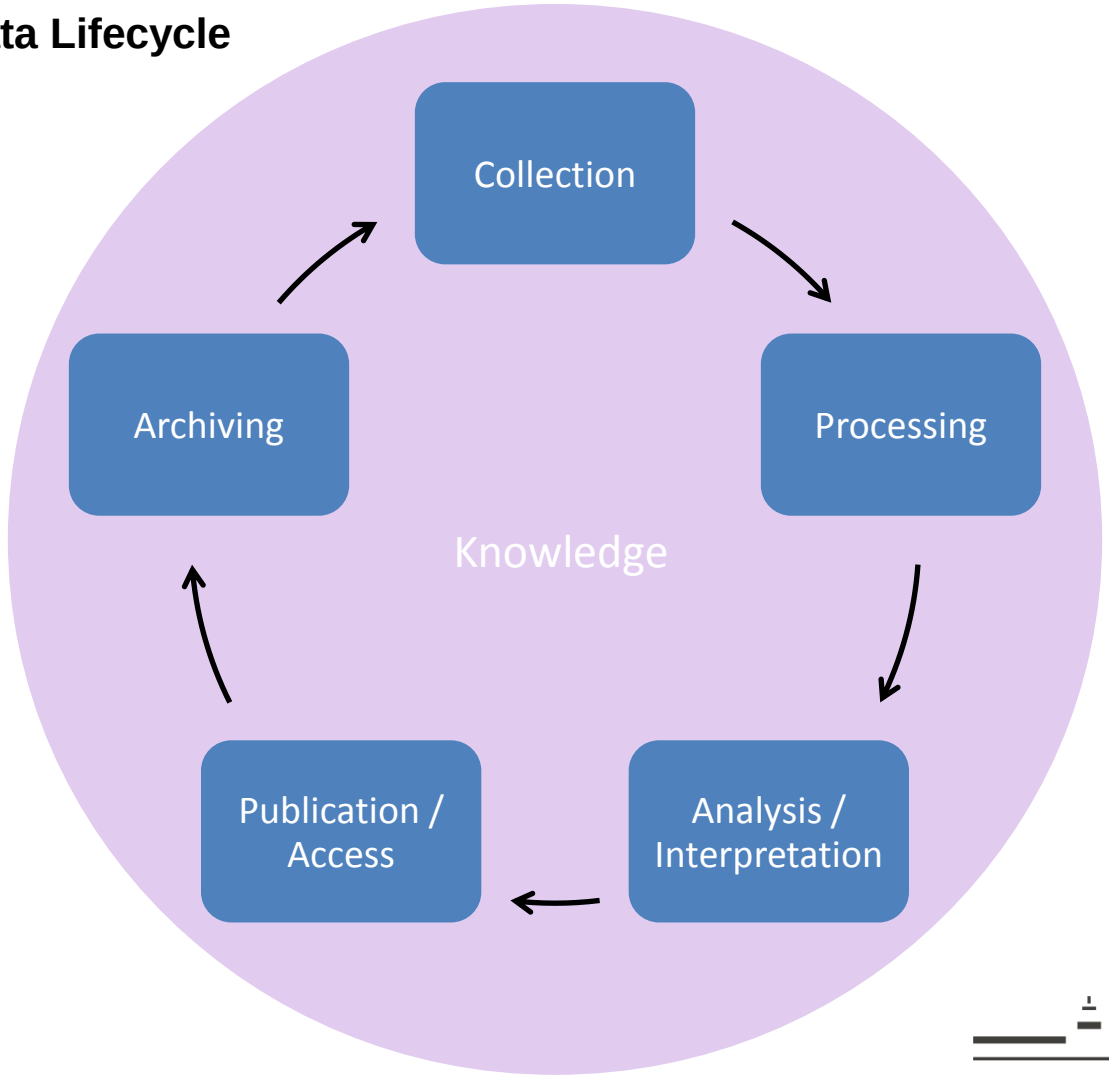
What is RDM?

Research data is all data generated in the course of scientific work.

RDM is the management of this data throughout the whole research data lifecycle, aiming for long-term storage, accessibility and reusability of research data.

Research Focus

Research Data Lifecycle



Research Questions

RQ1 – Open Science

What relevance does the idea of open access have for making research data available in scientific practice?

RQ2 – Archiving

How far have researchers progressed in terms of professional archiving?

RQ3 – Knowledge

How do researchers assess their knowledge of dealing with research data?

Research Method

Method

- online questionnaire (21.07.-08.08.2014)

Population

- scientific staff of Münster University (6,000 individuals)

Sample

- 1,042 participants (17%)
- 667 completed questionnaires
    - 19% professors
    - 79% non-professorial academic staff
    - 2% other

Open Science

RQ1

What relevance does the idea of open access have for making research data available in scientific practice?

Open Science Criteria

- OC1: Making available of research data
- OC2: Regulations of disclosure by binding guidelines

Open Science

Making Available

	Ø
Making data available	27.0
Thereof: via a subject-specific repository	4.2
Thereof: in the context of a publication by a publishing house	17.4

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Open Science

Conclusion

× OC1: Only a minority makes research data available to other scientist

× OC2: Guidelines for disclosure are mostly unknown

→ ***The idea of open access is of minor relevance in the scientific practice***

Archiving

RQ2

How far have researchers progressed in terms of professional archiving?

Archiving Criteria

- AC1: Non-local storage
- AC2: Long-term storage
- AC3: Regular backups
- AC4: Binding regulations (safe storage & systematic recording in databases)
- AC5: Involvement of professional data specialists
- AC6: Targeted archiving with a clear purpose of use

Archiving

Storage Locations

	Ø
Internal storage locations	
Office computer	69.9
Server of the department	48.3
Server of the computing center	34.5
External storage providers	
Subject-specific repository	7.5
External cloud provider	17.5
Other locations	
Private computer	35.7
External data storage media	62.8

(Results in %, N=667)

Archiving

Archiving Routines

	Ø
Storage duration: at least 5 years	52.5
Backup routine: regular, at least quarterly, backups	43.5

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Archiving

Knowledge of Guidelines for Storage and Recording

	Ø
University directives	
Data backup for a certain duration	19.9
Systematic recording in internal reference databases	5.8

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Archiving

Persons in Charge for Data Archiving

	Ø
Professors	52.6
Non-professorial academic staff	91.0
Student assistants	54.1
IT staff	29.2
Library staff	1.5
External service providers	6.9

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Archiving

Storage Purposes

	ø
Proof of replicability	84.9
Researchers‘ own re-analyses	84.9
Others‘ re-analyses	42.4
Scientific education	27.3
Exclusion of legal risks	42.1
Preservation as historically relevant information	14.2
Without cause	15.3

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Archiving

Conclusion

× AC1: Primarily internal storage locations

– AC2: Tendency towards long-term storage

× AC3: Backups are common, but often not on a regular basis

× AC4: Regulations for storage and recording are mostly unknown

× AC5: Professional data specialists are rarely involved

– AC6: Major purpose are proof of replicability and own further research

→ ***Archiving has not reached a professional level as demanded by RDM***

Knowledge

RQ3

How do researchers assess their knowledge of dealing with research data?

Knowledge Criteria

- KC1: Knowledge of research data management
- KC2: Need for advice

Knowledge

State of Knowledge and Need for Advice

	Ø
Good to very good knowledge	20.0
Need for advice	83.7
General questions	38.7
Publishing and quotation	33.1
Technical questions	48.4
Legal questions	52.9
Data management plans	28.5
Third-party funded projects	29.8

C1: Humanities and social sciences, C2: Life sciences, C3: Mathematics, C4: Natural sciences, C5: Economics and law

(Results in %, N=667)

Knowledge

Conclusion

- × KC1: Lack of knowledge about research data management
- × KC2: Considerable need for advice (mainly legal and technical aspects)

→ *The majority of scientists has only little knowledge about RDM*

Conclusion

Scientists are highly interested in RDM, but it has not affected their work in a vital way.

- **Open Science** (Open Access) is of minor relevance

  → Binding guidelines + incentives for researchers

- **Archiving** has not reached the necessary degree of professionalism

  → Binding guidelines + data specialists

- **Knowledge** of RDM is highly deficient

  → Integration of RDM into education + further training