FOSTER

Course for Doctoral Students

RESEARCH DATA MANAGEMENT AND OPEN DATA

25th July 2015, Social Science Data Arhives,
Faculty of Social Sciences, University of Ljubljana

ECPR Summer School 2015

RESEARCH DATA LIFECYCLE: ROLE OF DATA SERVICES

Sonja Bezjak, Social Science Data Archives

Social Science Data Archives

Content

- About Social Science Data Archives (ADP)

- Complexity and diversity
- Open Data project in Slovenia
- Research Lifecycle
- Research Data Lifecycle
- Roles and Responsibilities in Research Data Lifecycle

FOSTER

Social Science Data Archives, UL

- 1997

- national data repository for social sciences

- 600 social science surveys

- depositors from all 4 (3 public) universities, private research centres, Statistical Office of Slovenia (8-10 research centres per year)

- cca. 700 users yearly (90 % education, 10 % scientific/research purpose)

FOSTER

Single researcher
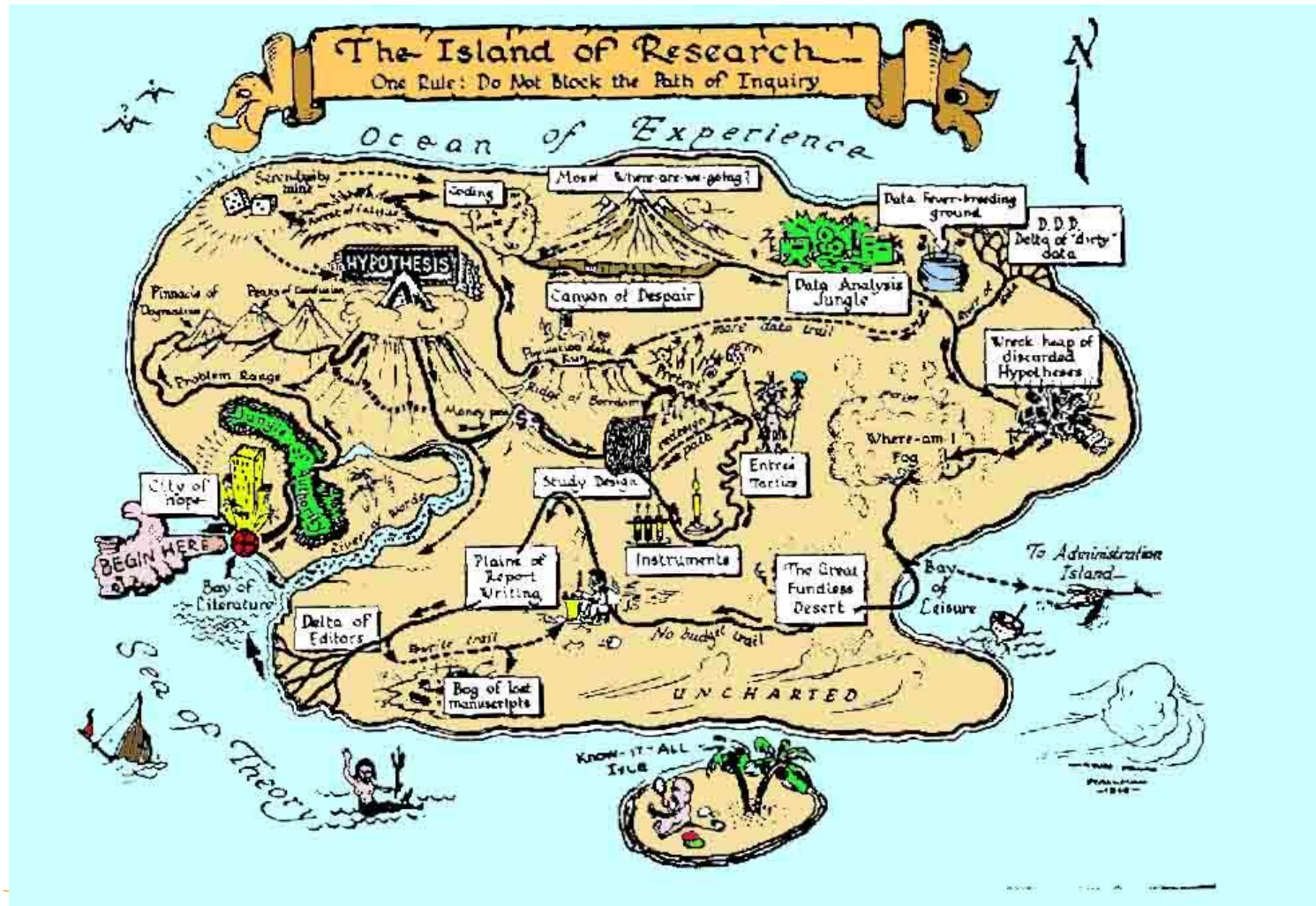
"Open Data Excuse" Bingo

#openDataExcuses

Lawyers want a custom License	Data Protection	Poor Quality	People may misinterpret the data
It's too big	There's no API	It's too complicated	Thieves will use it
It's not very interesting	There's already a project to...	We will get too many enquiries	What if we want to sell it later
We'll get spam	We might want to use it in a paper	I don't mind, but someone else might	Terrorists will use it

For open data teams; print out a copy and put it on your office wall. Cross out each excuse people give you. There are no prizes, but you can tweet "bingo! #openDataExcuses" if you think it might make you feel better*.
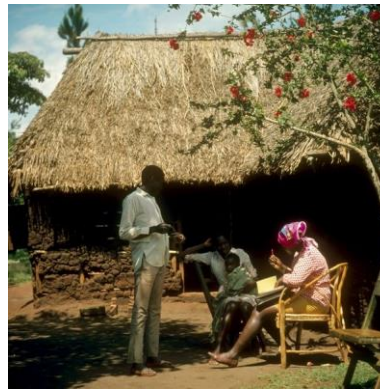
* it won't

Generate your own bingo grids at http://data.dev8d.org/devbingo/

FOSTER

Diversity of Methodologies I

- Interview
- Medical survey
- Geographical location
- ...

Source: Students working with an artificial patient (Faculty of Biomedical Engineering, CTU in Prague)

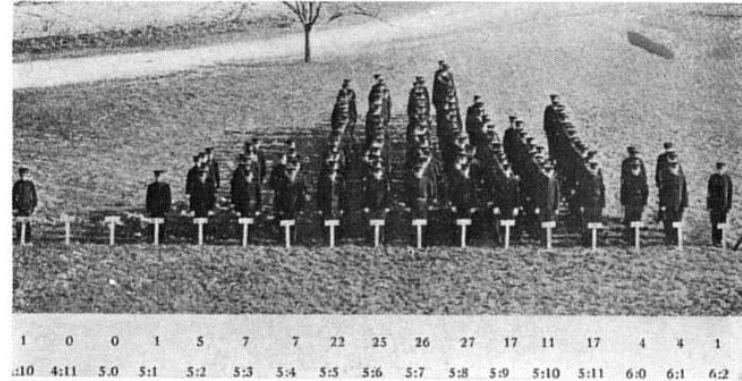Source: Tropenmuseum, part of the National Museum of World Cultures

Source: Stone hand axes, from Acheulean, by: Didier Descouens

FOSTER

Diversity of Methodologies II

Source: The archive of the Thesaurus Linguae Latinae with some of the 10 million slips used in creating the dictionary, by N p holmes

Source: Students arranged according to size. (After Blakeslee.), by: Project Gutenberg

Source: EthnoMuse, by: Matija Marolt

Source: John Paul Thomas' analysis of Vermeer painting "The Love Letter" grid overlay #2 showing module and primary axes, by: John Paul Thomas estate

FOSTER

Diversity of Methodologies III

Source: Project SCREEN, Center for Climate Change, Universitat Rovira i Virgili, Spain.

Source: The CERN datacenter with World Wide Web and Mail servers, by Hugovanmeijeren

Source: Mars Science Laboratory parachutes, by: NASA

Source: Forest Climate Observatory near Magdeburg, Germany; data retrieval, by André Künzelmann

FOSTER

Diversity of...

- Types
- Formats
- Size
- Sensitive data (Human, State secret)
- Long term / Short term value
- ...

One solution for all disciplines?

FOSTER

Open Data Project (2010-2013)

- **Goal:** to prepare drafts of national **policy and strategy**, needed for establishing a system of open access to research data in Slovenia

- **Principle of Flexibility**: ‚*Specific national, social, economic and regulatory implications should be considered when organisations develop research data access arrangements, and when governments develop policies to promote data access and review the implementation of these Principles and Guidelines.'* (OECD, 2007, 15)

FOSTER

Methodology/Approach: ‚bottom to top'

1)22 semi-structured interviews:

- researchers, librarians, heads of research institutes
- 10 research institutes, 6 faculties
- 17 research disciplines: physics, biology, medicine, civil engineering, archeology, social work, economy, musicology, anthropology, languages…

2) 3 workshops:

- Workshop 1: Problems and Solutions in the Field of Data Services in Slovenia
- Workshop 2: Policy of Research Data Management in Slovenia
- Workshop 3: Advanced Technology for Establishing Data Infrastructure in Slovenia

3) Individual working visits/consultations:

- Information Commissioner
- Intellectual Property Institute
- Research Centre of the Slovenian Academy of Sciences and Arts
- DARIAH, CLARIN representatives in Slovenia …

FOSTER

What we found out?

- Regarding documentation, preservation and access to research data <u>in Slovenia</u>:

  ➢ Researchers have different habits and views

  ➢ Research institutes follow disunited, unwritten rules and practices

- But mainly they face identical problems:

  ➢ lack of knowledge, time and finances for dealing with research data,

  ➢ extremely competitive scientific environment

FOSTER

What we found out?

Major barriers in achieving open data:

1) big differences in the development of data infrastructure and services

2) dilemmas and fears emerging from questions related to law (intellectual property, personal data protection)

3) absence of culture of data sharing

4) absence of framework policy, which would help in research data management
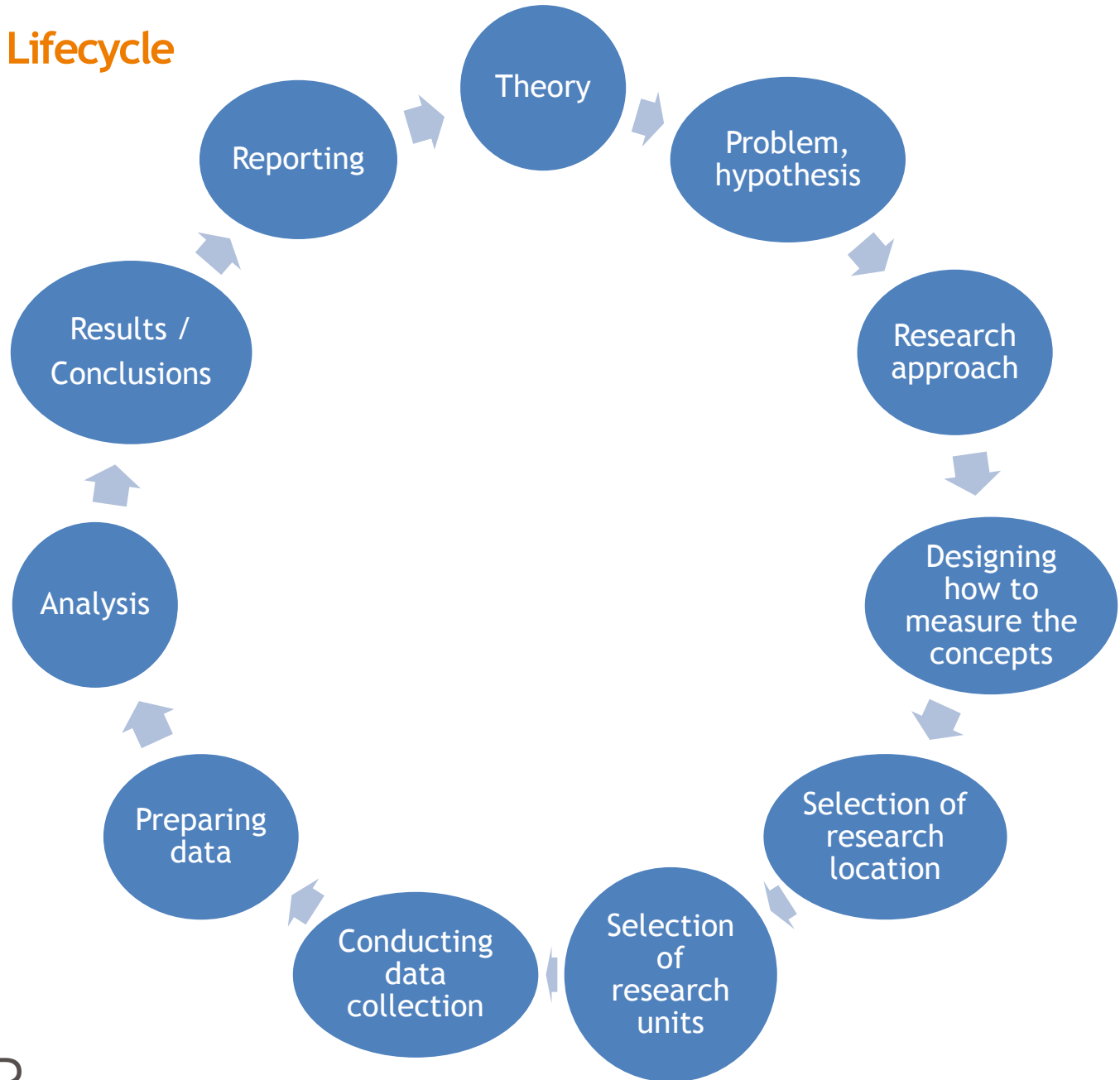
FOSTER

One solution for all disciplines?

To overcome the differences in development of infrastructure and services, expactations of researchers, capabilitites of funders and diversity of disciplines (formats, methods etc.):

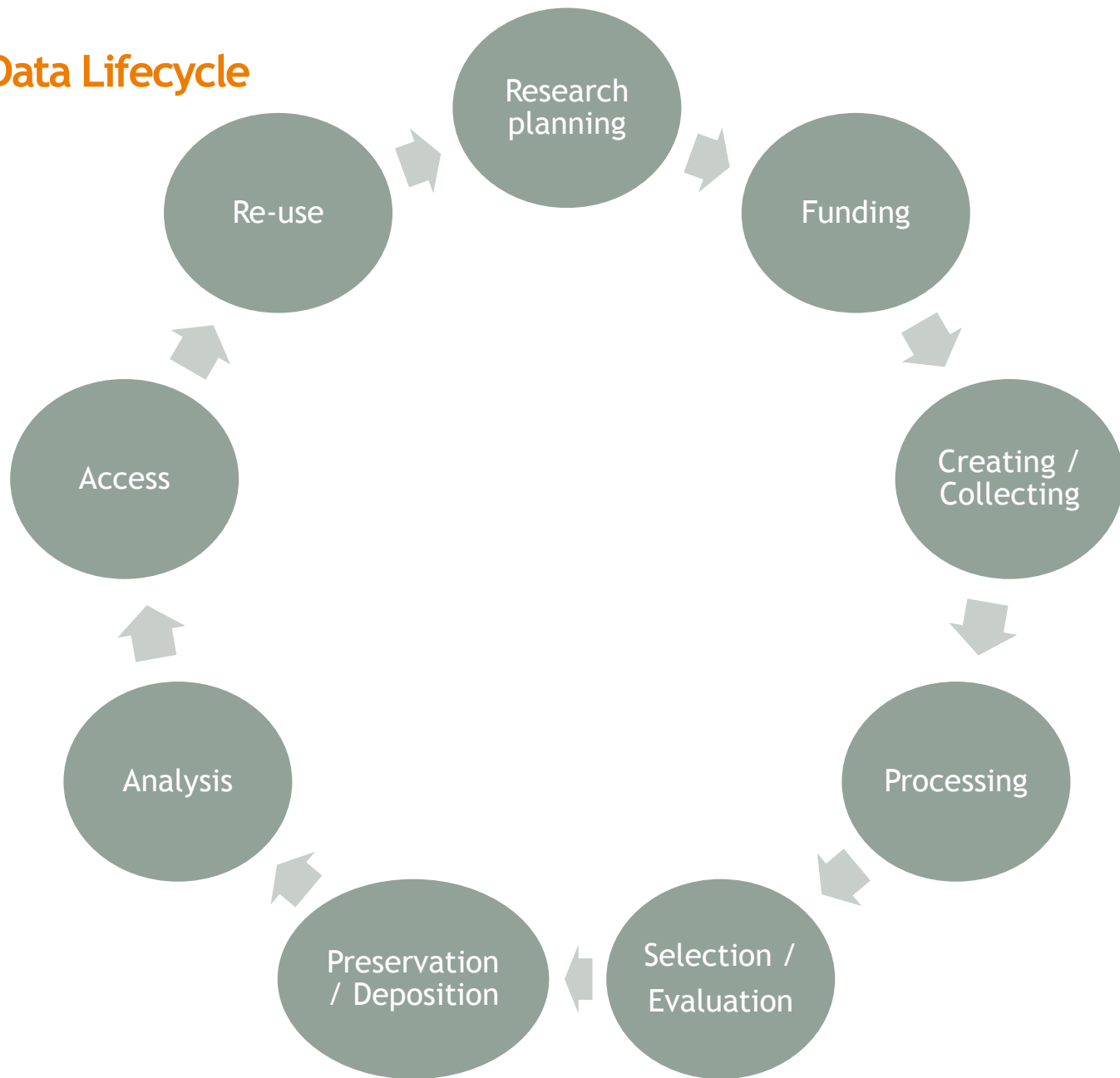- Data management knowledge, data services and data infrastructure

FOSTER

- **Research Lifecycle**
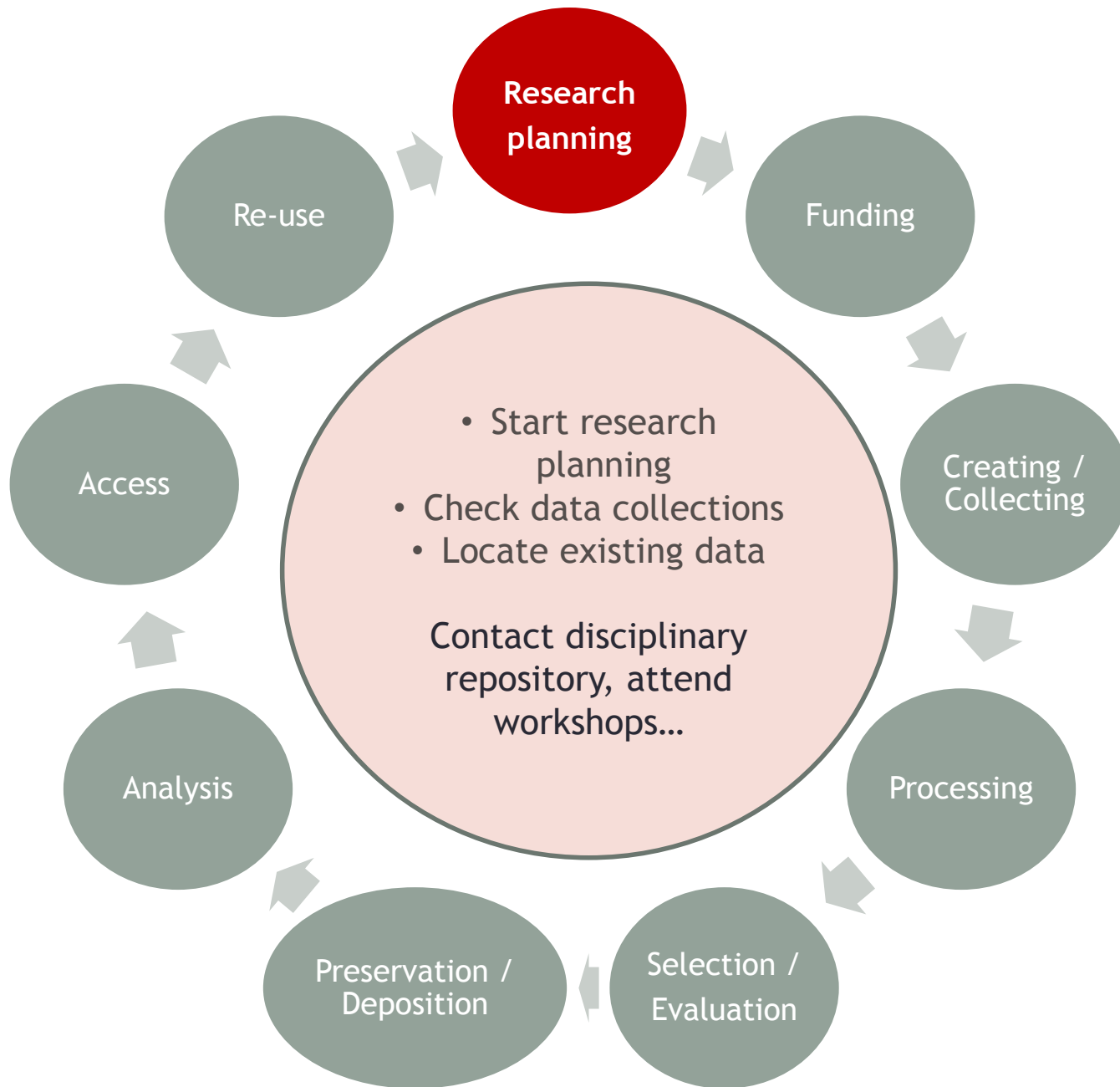
- **Research Data Lifecycle**

Research Lifecycle

Research Data Lifecycle

- Research planning
- Funding
- Creating / Collecting
- Processing
- Selection / Evaluation
- Preservation / Deposition
- Analysis
- Access
- Re-use

FOSTER

Research data lifecycle diagram (FOSTER)

- Research planning
- Funding
- Creating / Collecting
- Processing
- Selection / Evaluation
- Preservation / Deposition
- Analysis
- Access
- Re-use

Center:
- Design research
- Develop / adjust RDMP
- Develop consent form
- Collect data
- Capture metadata

Get involved with repository, ask for assistance, when needed

Research planning

Re-use

Funding

Access

Creating / Collecting

- Quality of data
- Type of access
- Interesting data
  - Rare data
- Important data

Get involved with trusted repository or journal

Analysis

Processing

Preservation / Deposition

Selection / Evaluation

FOSTER

Research data lifecycle diagram:

Research planning → Funding → Creating / Collecting → Processing → Selection / Evaluation → Preservation / Deposition → Analysis → Access → Re-use → (back to Research planning)

Preservation / Deposition (center content):

- Appropriate format
- Suitable medium
- Back-up and store
- Assure metadata and documentation
- Deposit

Get in communication with repository: protocols, standards for depositing

FOSTER

Research planning

Funding

Re-use

Access

- Sharing data
- Regulation access
- Copy-right
- Promotions

Help repository in promotional activities: lectures, hands-on, report about your publications…

Creating / Collecting
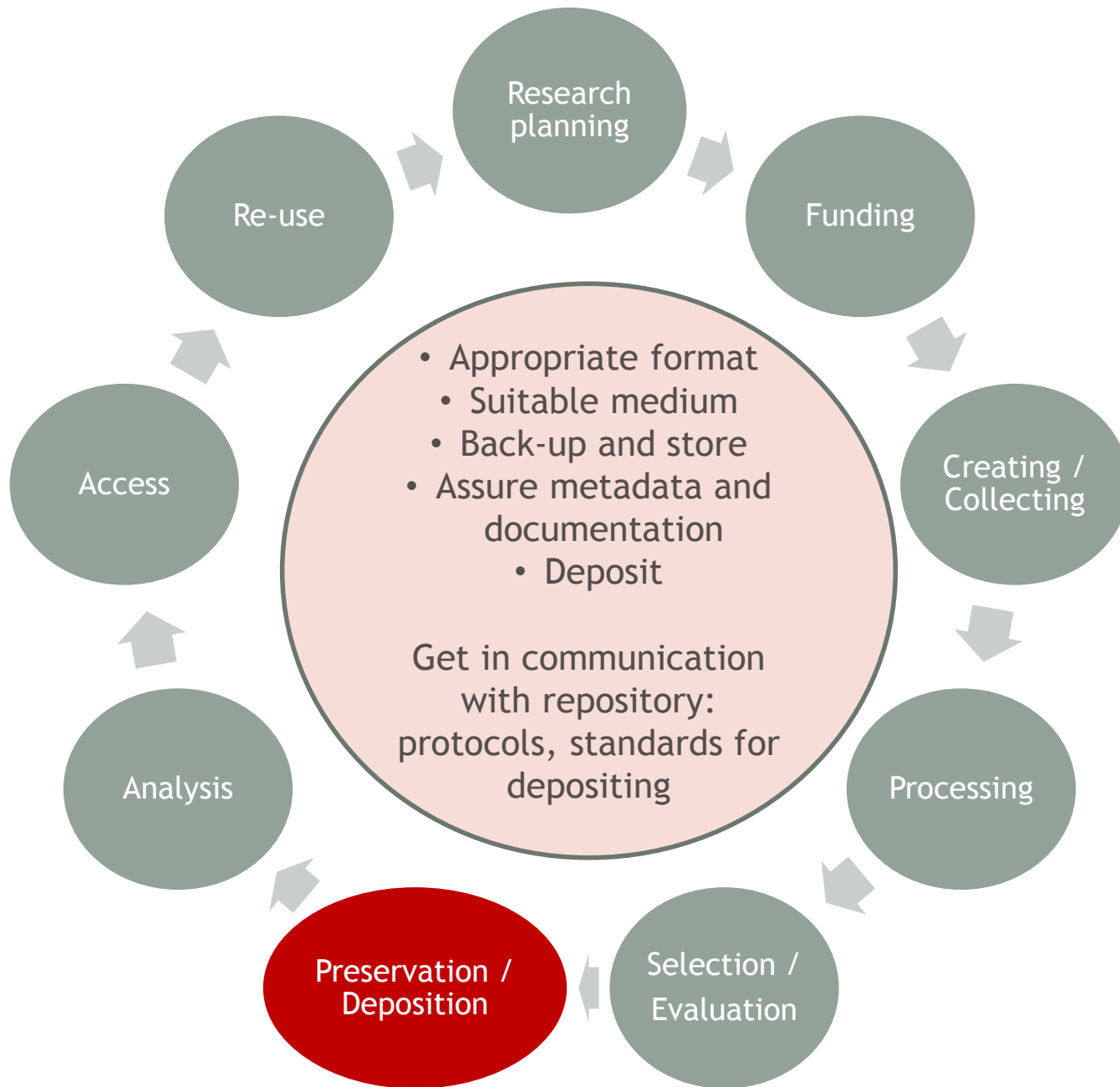
Processing

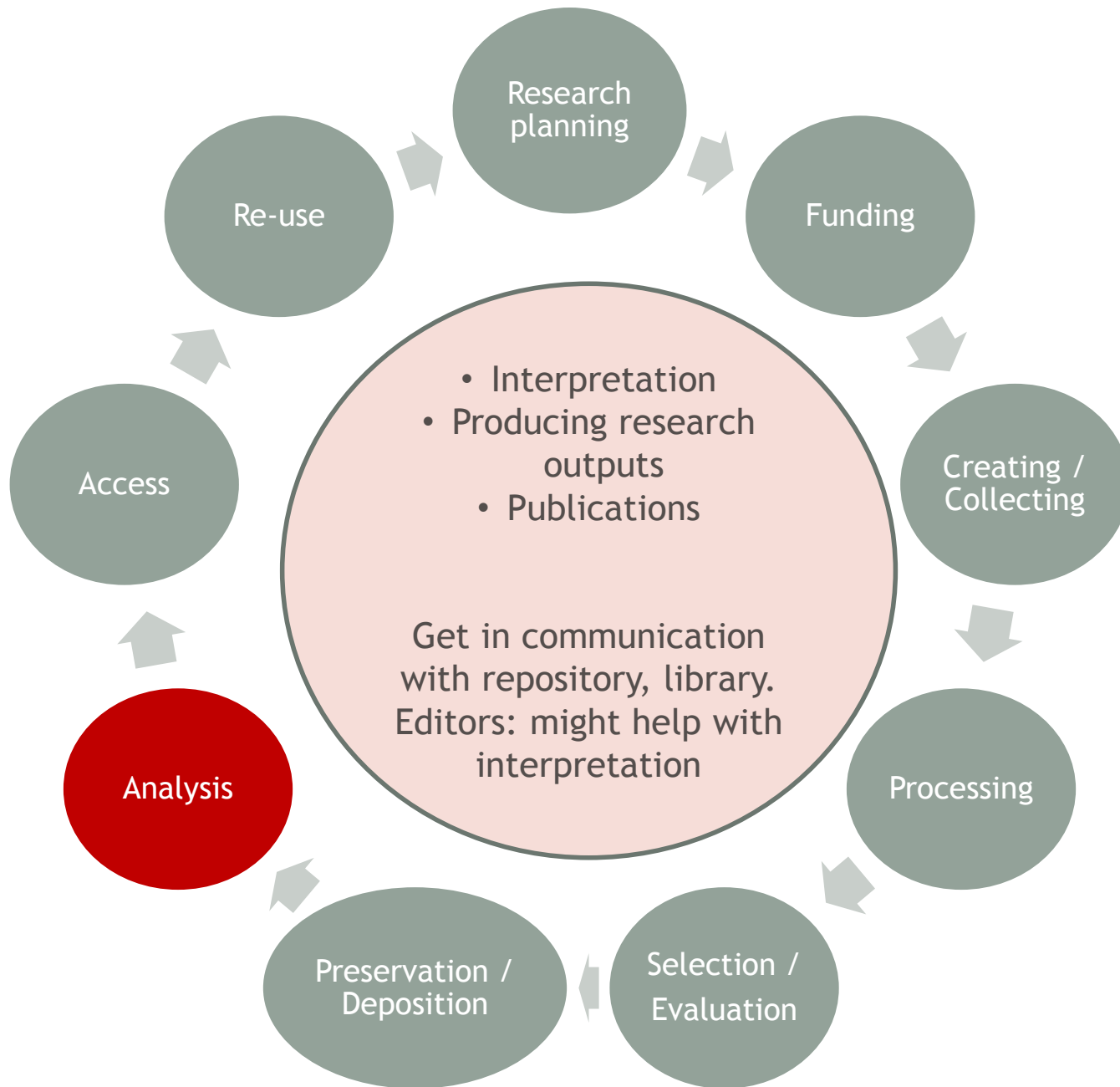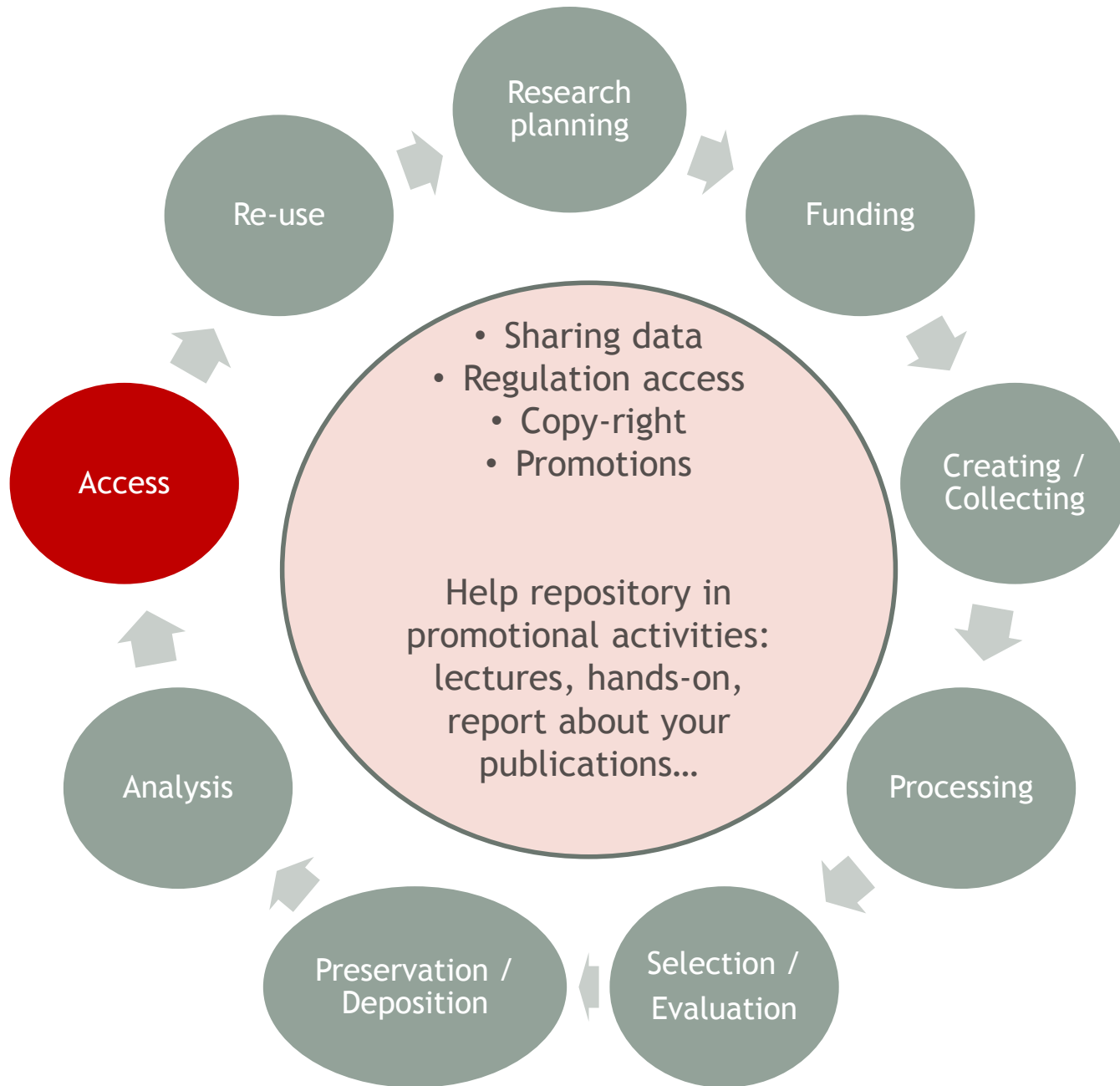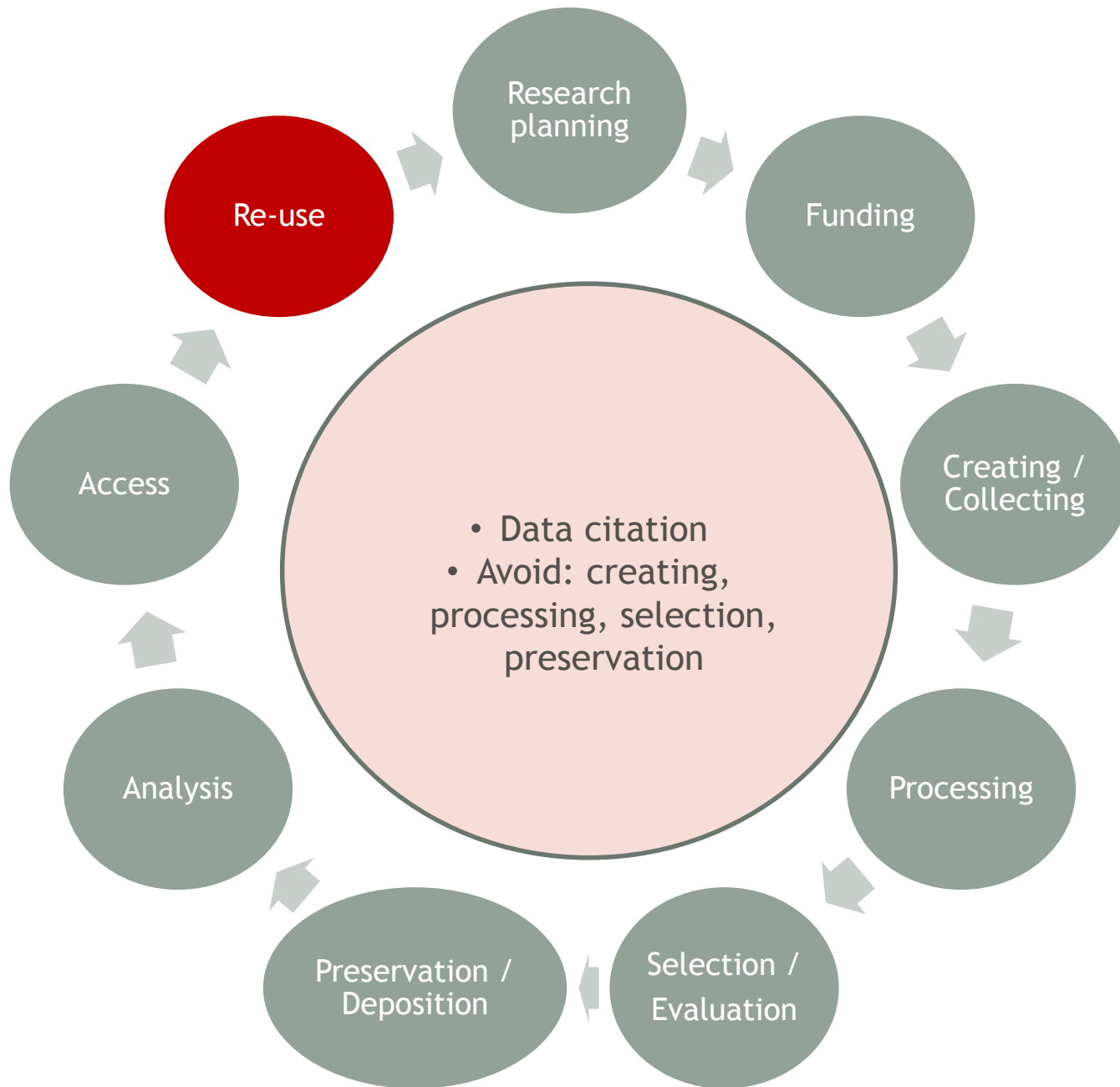Selection / Evaluation

Preservation / Deposition

Analysis

FOSTER

Roles and responsibilities in RDLC

RESEARCHER:

- Scientific principles: quality, transparency, data as public good
- RDMP
- Trainings
- OA

FOSTER

Roles and responsibilities in RDLC

RESEARCH INSTITUTION:

- OA policies
- trainings
- infrastructure
- common services and tools

FOSTER

Roles and responsibilities in RDLC

LIBRARY:

- information about data sources
- information depositing data
- Helps select data centre or data archive
- Information about OA
- Preparation of DMPs
- Support with preparation of basic study metadata and documentation, author's rights, and explains other deposition requirements

FOSTER

Roles and responsibilities in RDLC

FUNDER:

- national / disciplinary policies (RDM, OA)
- funds to cover OA costs
- OA obligations

FOSTER

Data services

Data + metadata + accompanying material

DATA DEPOSITORS:

- Formats
- Standards
- Consent
- Licenses
- Bibliography
- Cobiss
- ...

ADP

- Selection
- Added value
- Curation
- Access
- ...

DATA USERS:

- Search
- Use
- Tools
- Citation
- ...

FOSTER

Questions?

FOSTER