

Data stewardship: from security to documentation and sharing

Brian Kleiner, FORS - 11.12.2018
VariaForMea

Some background: a changing research landscape in Europe

- Funders' requirements
- Journals' requirements
- Technical environments and tools
- Open data and research transparency
- International standards
- Research team composition
- Focus on **data management**

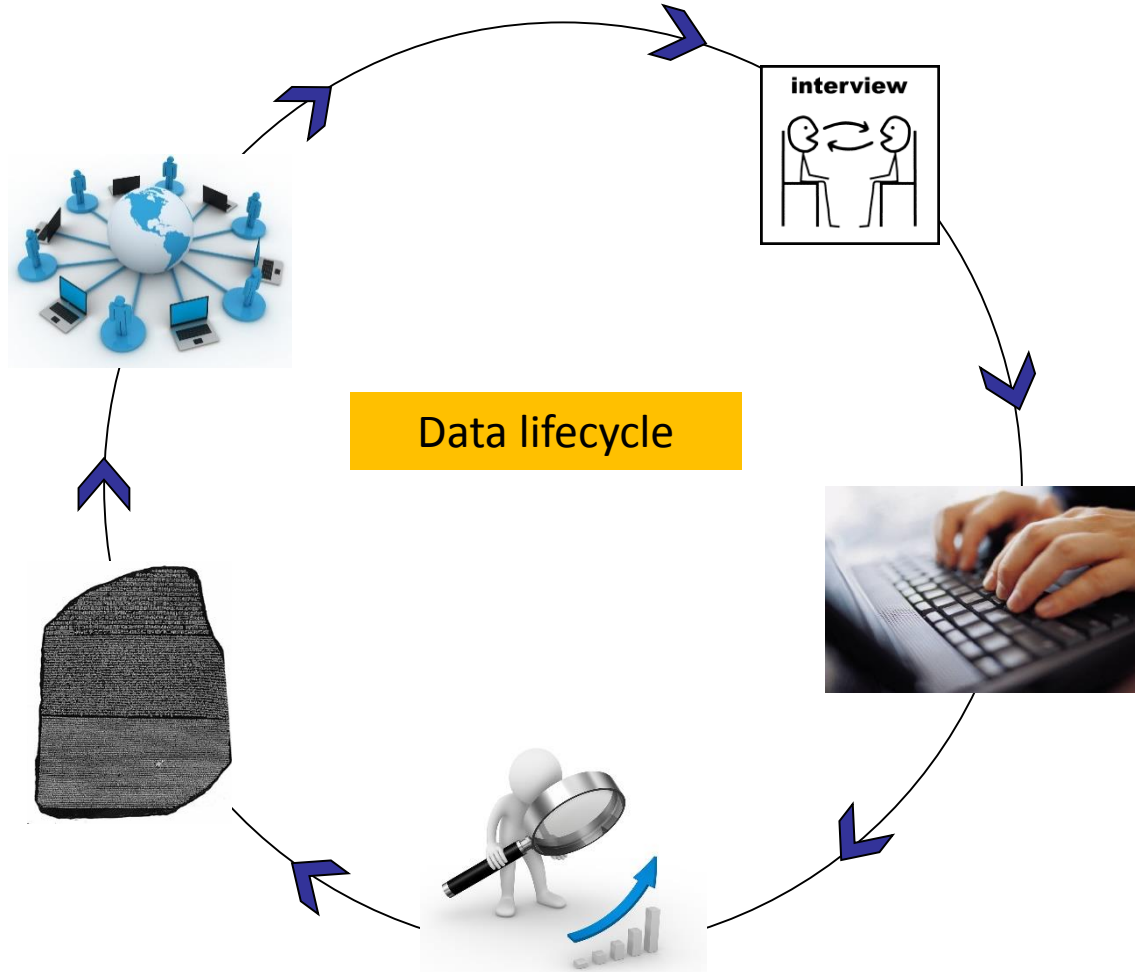
Data management throughout the lifecycle

Global:

- Project organisation
- Security and access
- Documentation
- File organisation
- Research ethics

Local:

- Informed consent
- Data preparation
- Anonymisation
- Copyright
- Long-term preservation
- Data sharing and reuse



Data management throughout the lifecycle

Global:

Project organisation

Security and access

Documentation

File organisation

Research ethics

Local:

Informed consent

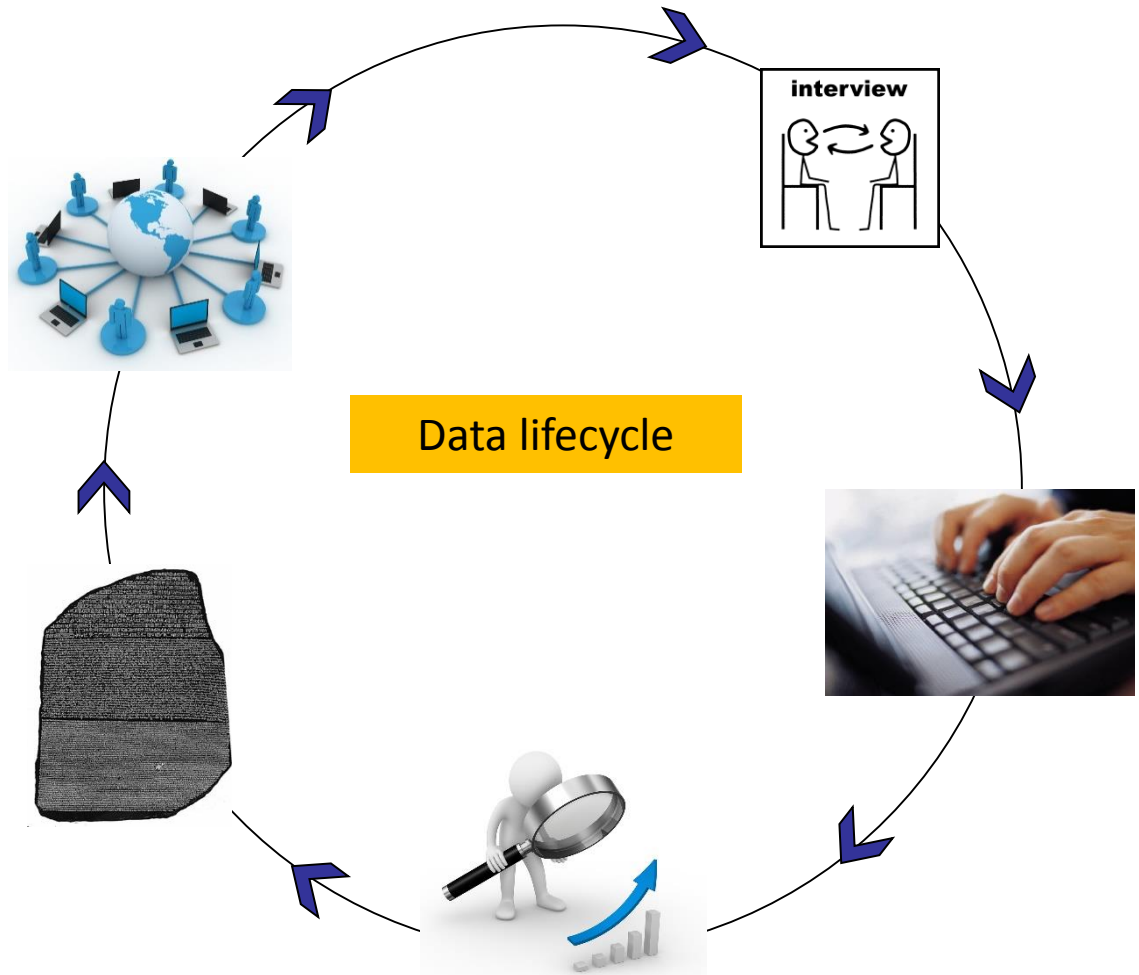
Data preparation

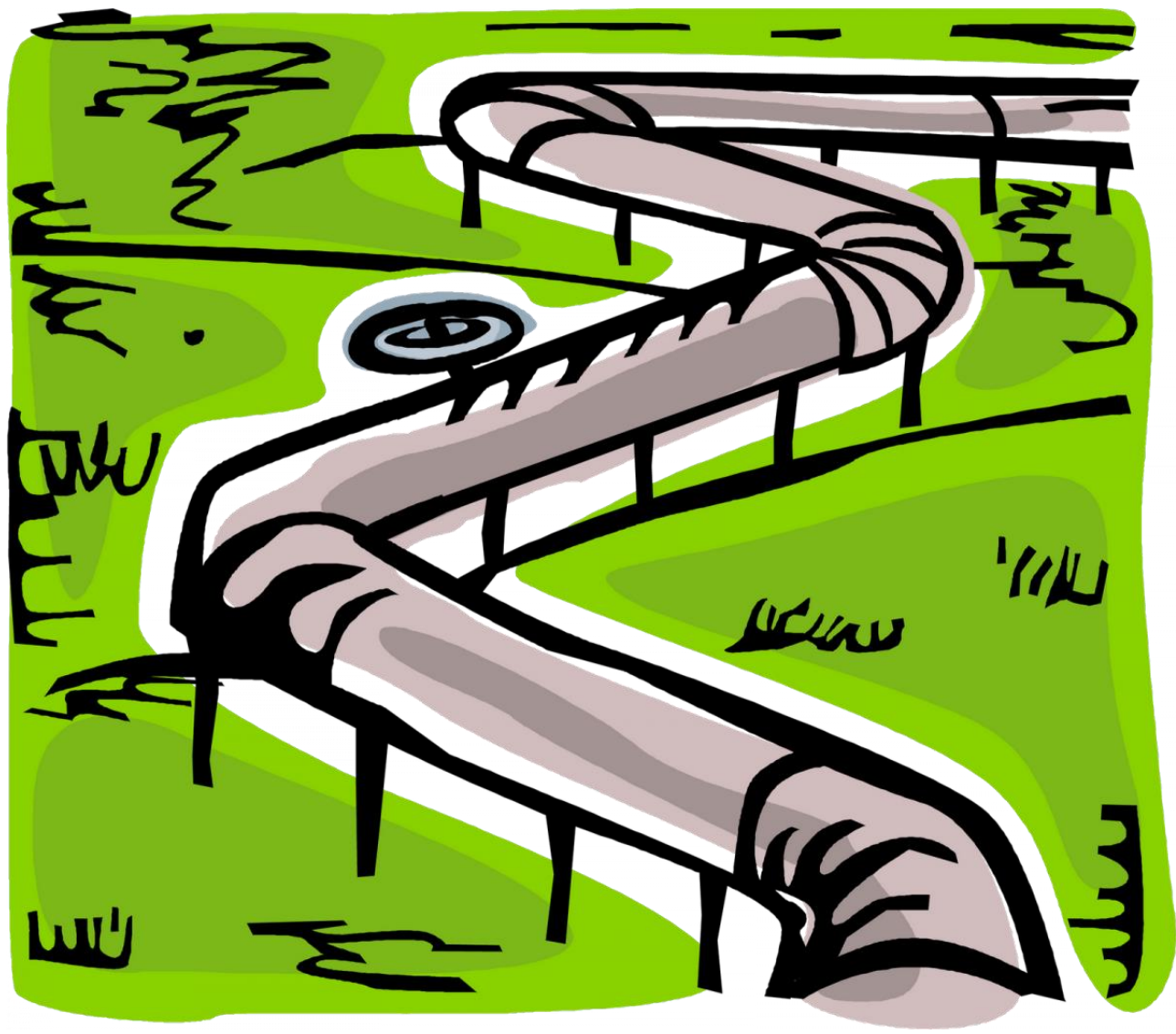
Anonymisation

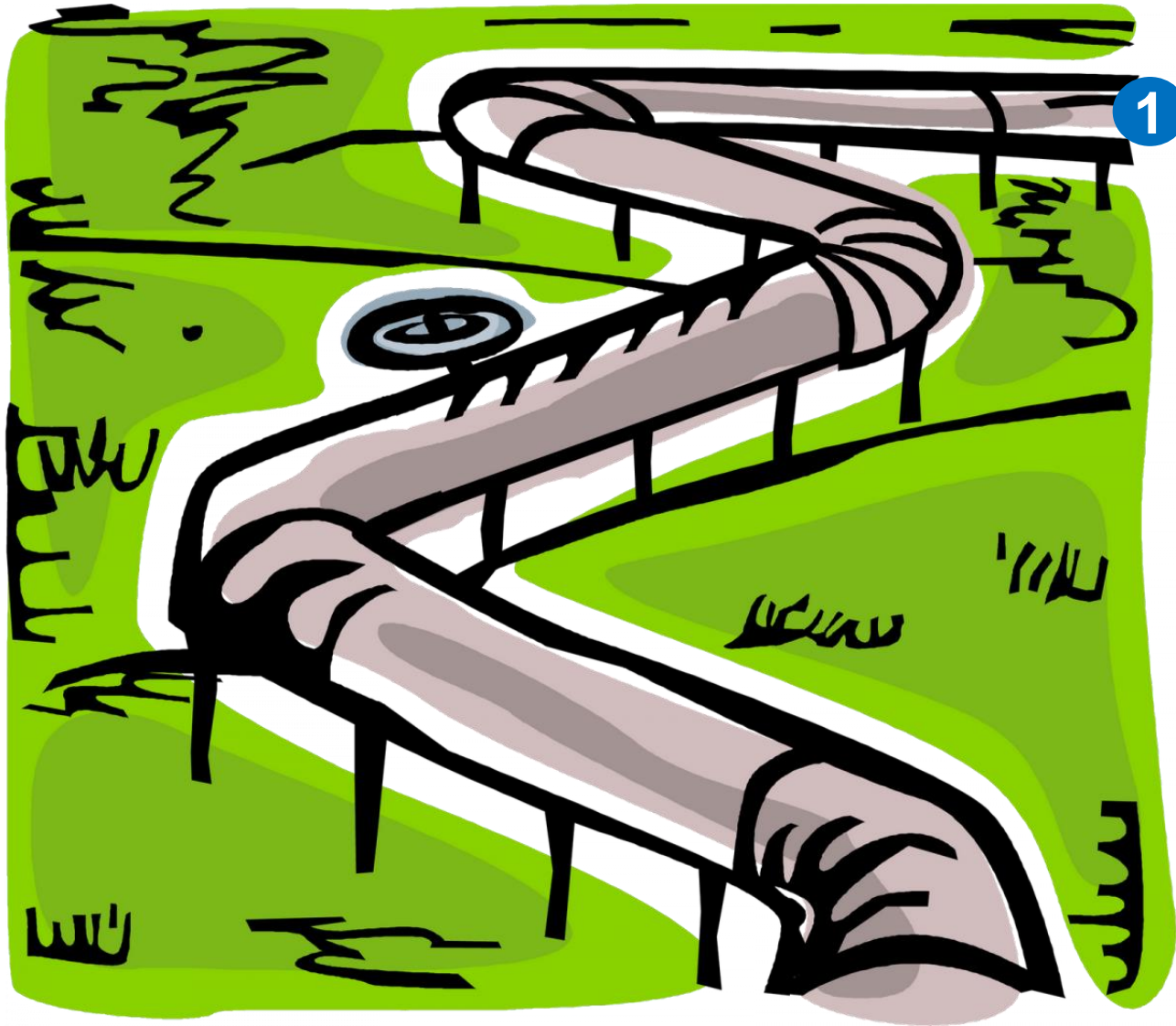
Copyright

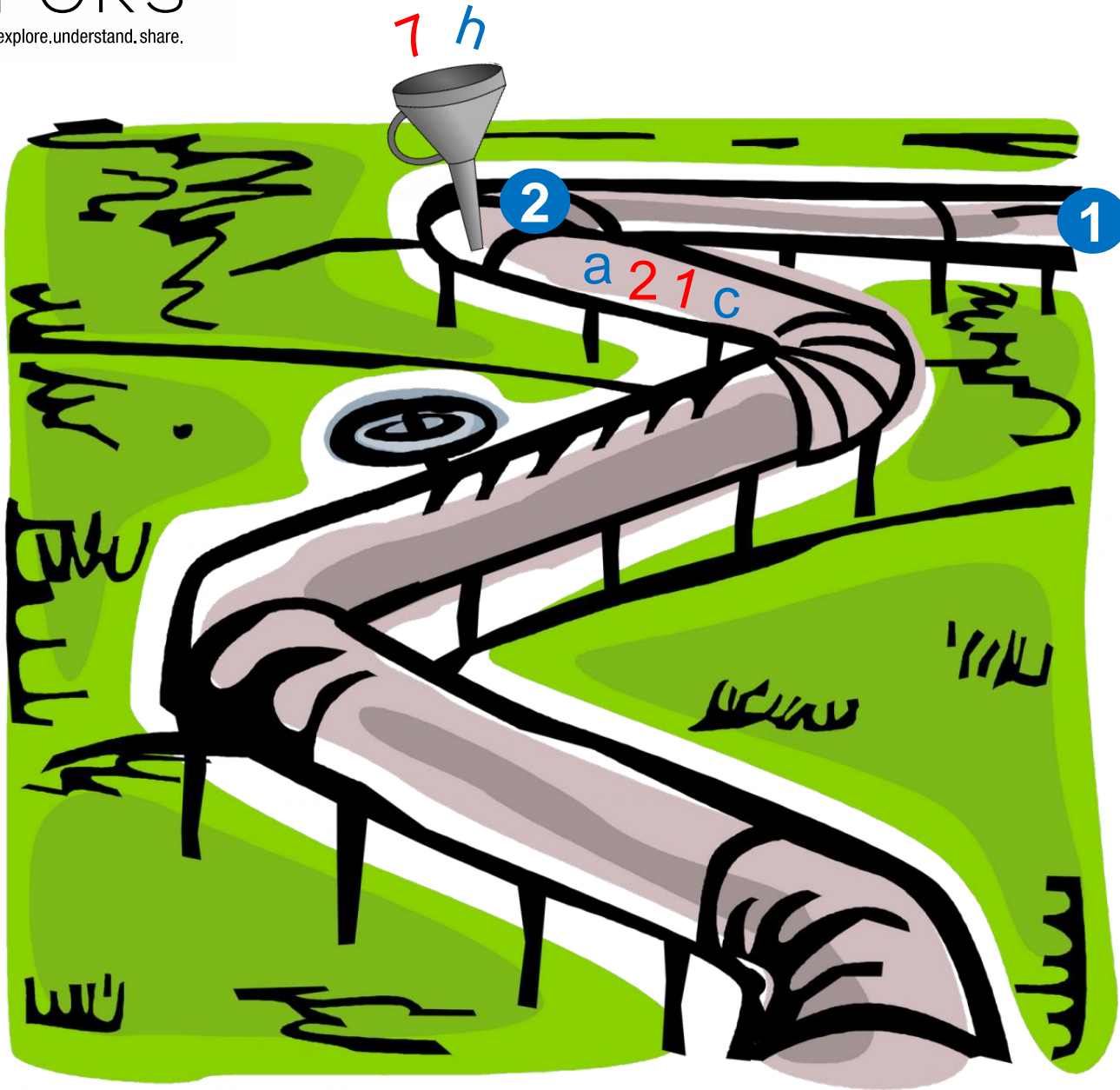
Long-term preservation

Data sharing and reuse



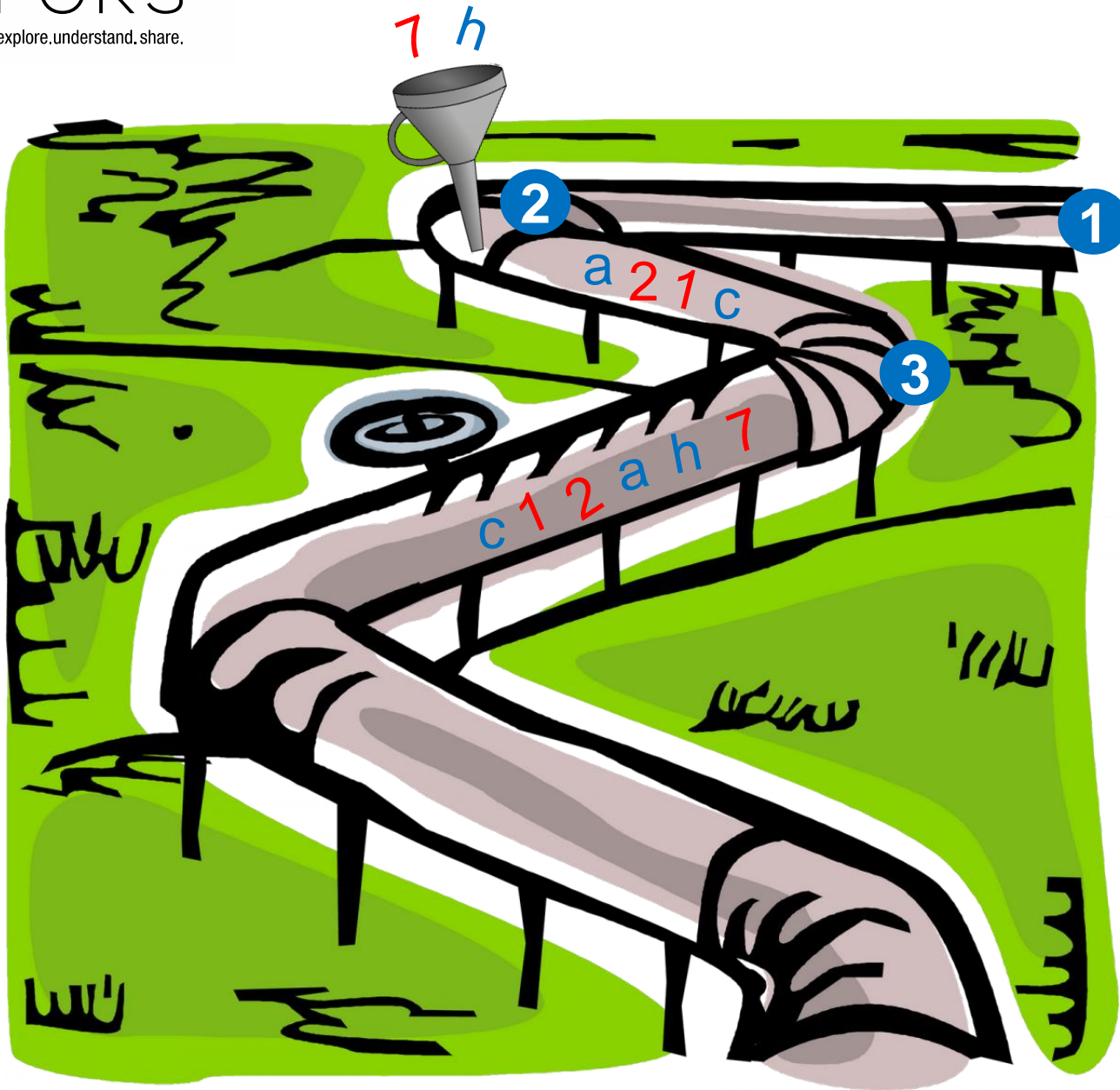


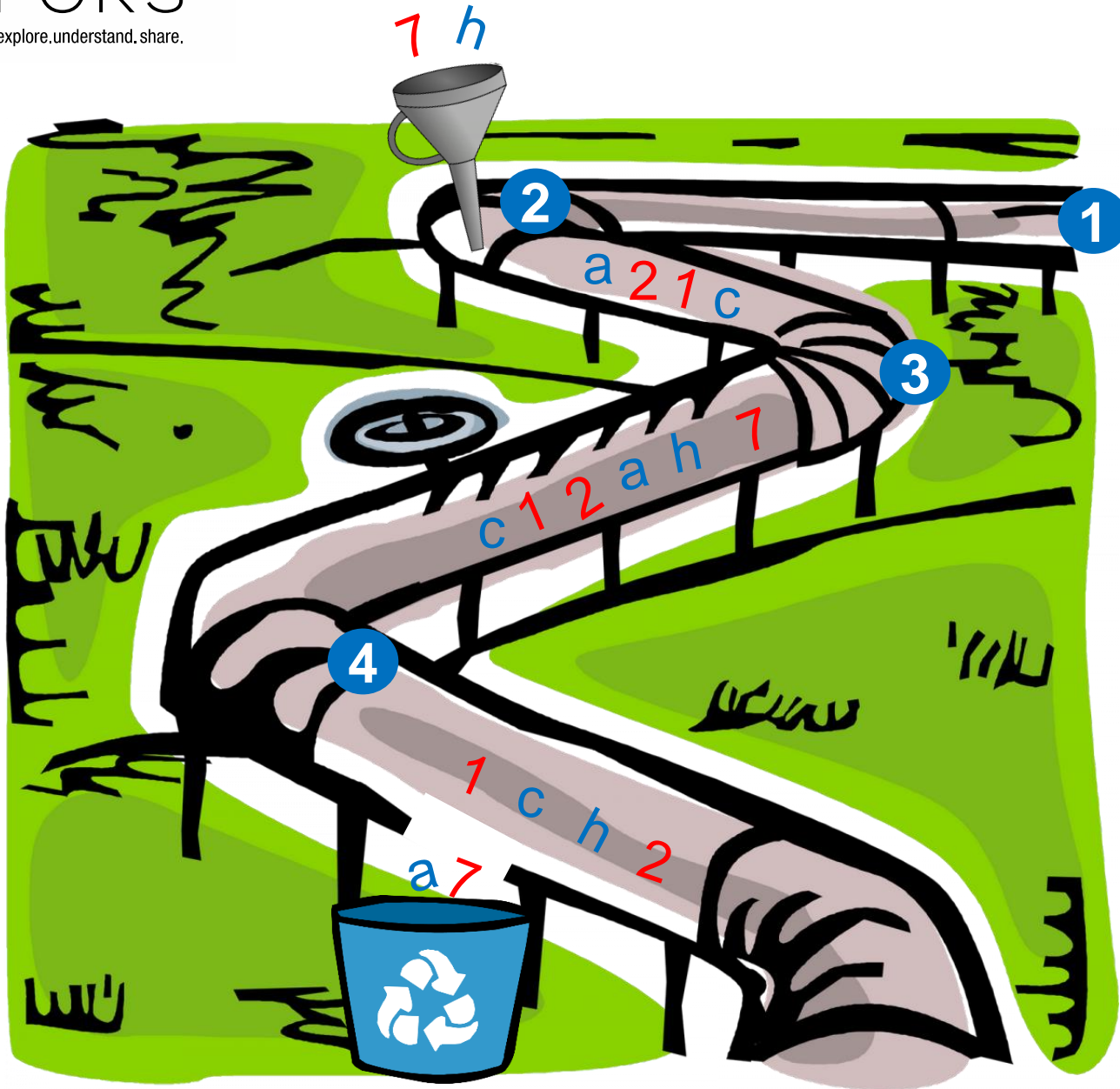


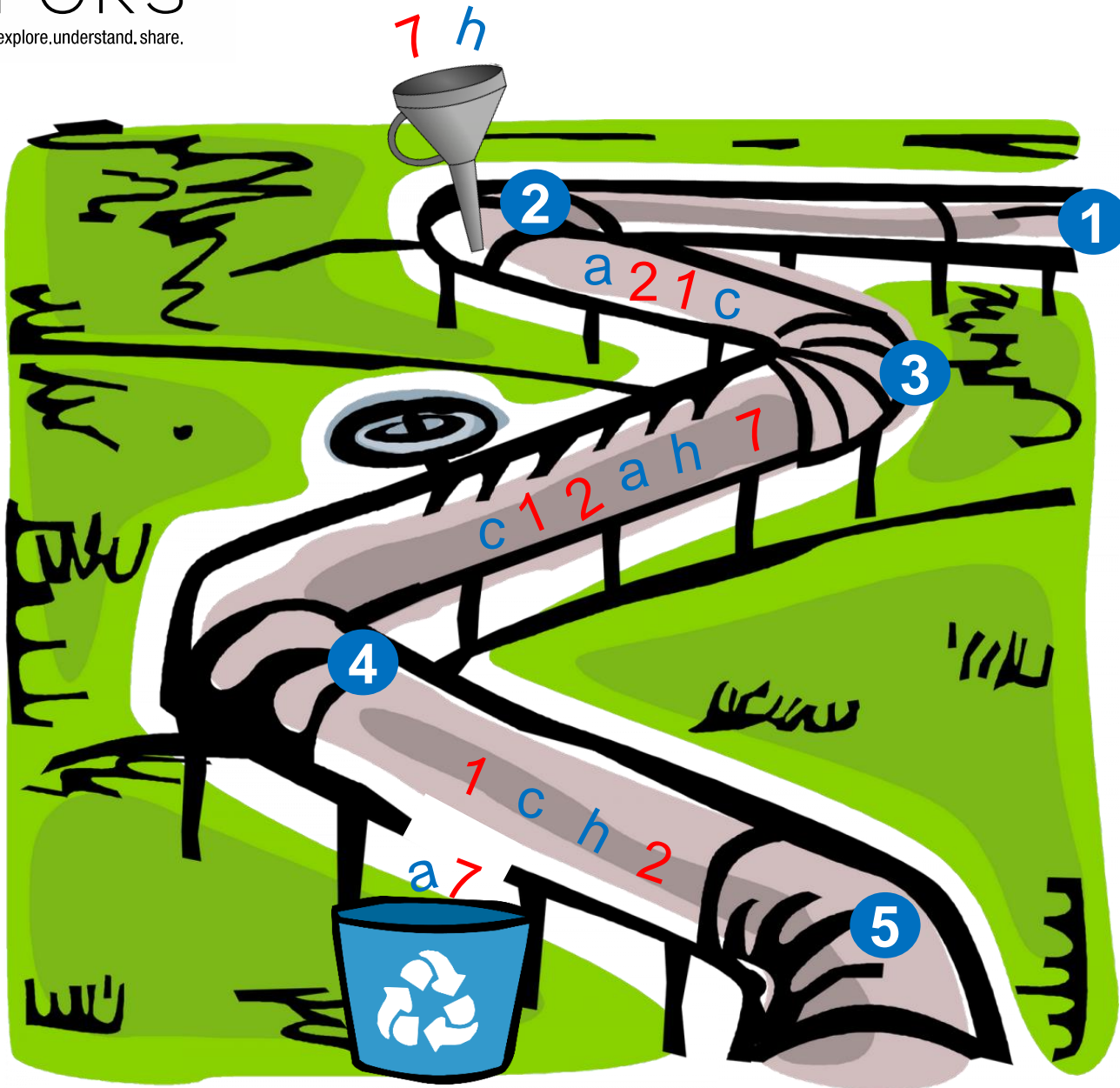


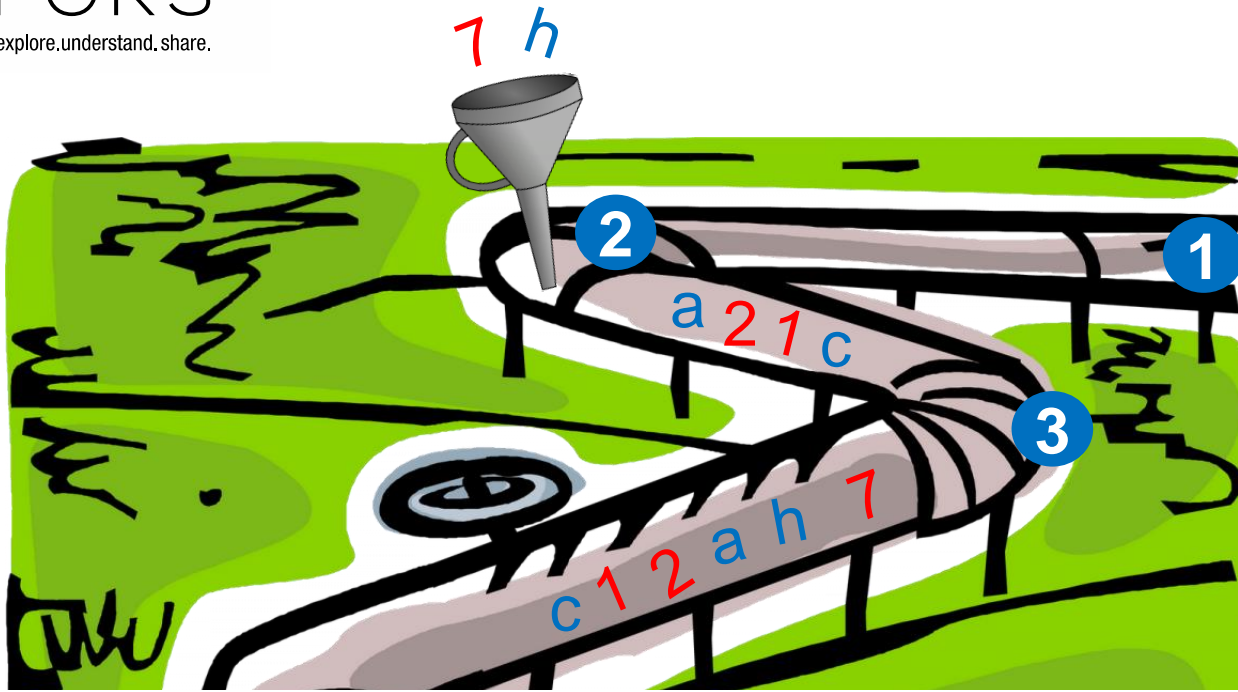
FORS ⁺

explore.understand.share.



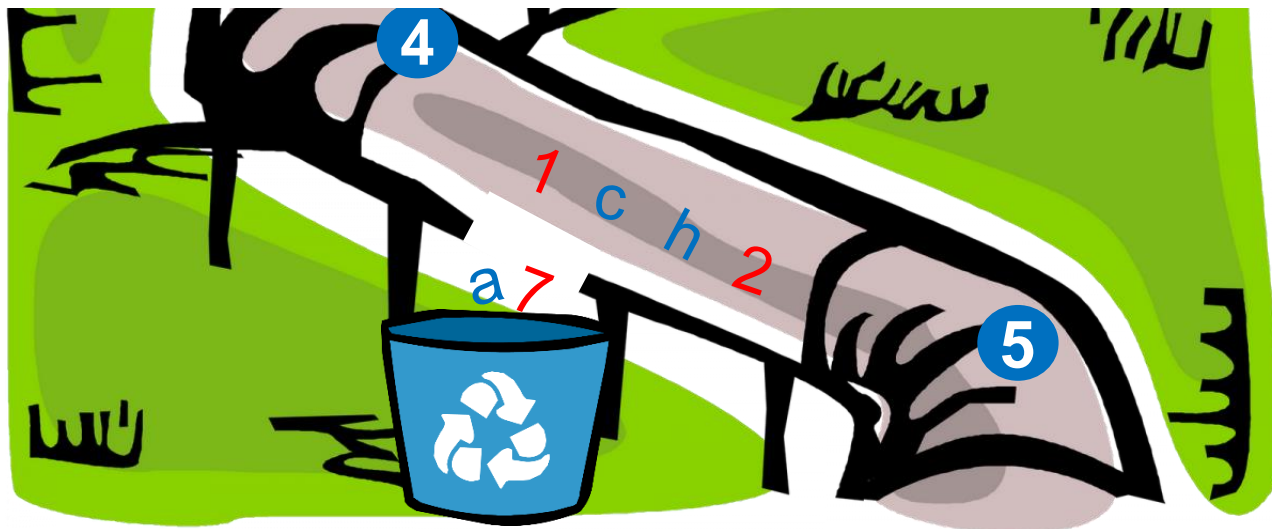






Project

Storage
Security
Access
Data protection
Documentation



Post-project

Security
Data protection
Documentation
Preservation
Sharing and reuse

Setting goals and parameters for data management

- Individual or team-based?
- Quantitative or qualitative data?
- Sensitive data or not?
- Later use planned or not?
- Sharing or not?
- Long or short project?
- Resources

DM should be planned as a **coherent whole**, and not piecemeal.

Data security

Data security is the means of ensuring that research data is kept safe from corruption and that access is suitably controlled.



Some data security risks

- Accidental or malicious damage/modification to data
- Theft of data
- Inappropriate access

In sum, your data can be damaged, altered, lost, or stolen, or accessed by the wrong people – your research and/or respondents can be harmed!

Securing the data environment: basic considerations

- What is the technical environment for file storage? Where are the servers?
- During the project, who can access the data and under what conditions?
- How are your data backed up?

Backing up data

- Will all data or only changed data be backed up?
- How often will full and incremental backups be made?
- How long will backups be stored?
- How much hard drive space or number of Digital Video Discs (DVDs) will be required to maintain the backup schedule?
- If the data are sensitive, how will they be secured and (possibly) destroyed?
- What backup services are available that meet these needs and, if none, what will be done about it?
- Who will be responsible for doing backups?

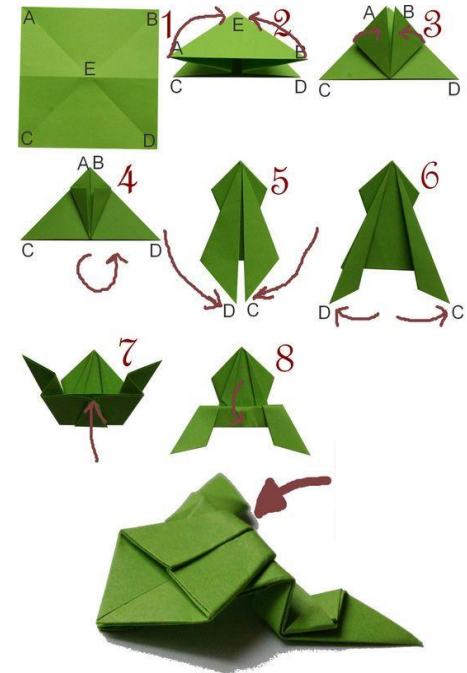
Documentation

Documentation is any information that serves as a record of a research project and that renders data usable and meaningful.

It:

- explains the study context;
- explains the research process and how data were created, including instruments used;
- describes the content and structure of data;
- describes manipulations that have taken place

Metadata are defined as data providing information about one or more aspects of specific data



Documentation during a project

Documentation is also important in order to have a written record of internal project processes. This is especially important for the ongoing functioning of large-scale complex projects.

Project rules should be documented on responsibilities, roles, and procedures for different DM activities and tasks, such as security and backup, file organisation, and data preparation. These can be revised over time as needed.

Documentation for others – how far to go?

Documentation should justify the needed effort. In deciding what to document and how much to document, always try to answer:

- Who might need the information?
- How important is it that the information is available to target users?
- For what purposes is the information needed?
- What might happen if the information was not available in the future – to you, to your team, to others?

Data preservation – after the project

Long-term preservation of data and documentation allows you to:

- Come back to your data in the future
- Offer your data and results for replication purposes
- Offer your data to others for reuse - new analyses and discovery

Professional archives provide long-term preservation services

Data sharing

Some benefits of sharing your data through an archive:

- Data preserved and safe from loss
- Controlled access to eligible users
- Recognition and citation
- Expansion of professional network

Some obstacles:

- Confidentiality
- Ownership
- Resources

Reuse of data

Secondary data can be useful for:

- seeing how others have treated a topic,
- getting inspiration for methodological approaches
- addressing new research questions
- looking at changes over time, comparing existing data, or possibly compare existing data with data that you collect yourself
- material for teaching

Exercise

- What are the *biggest risks* to your data in leading them along “the pipeline”? How can these risks be addressed?
- For your project, what are the pros and cons of data sharing? What might prevent you from sharing your data? In what ways might your data be re-used by others?