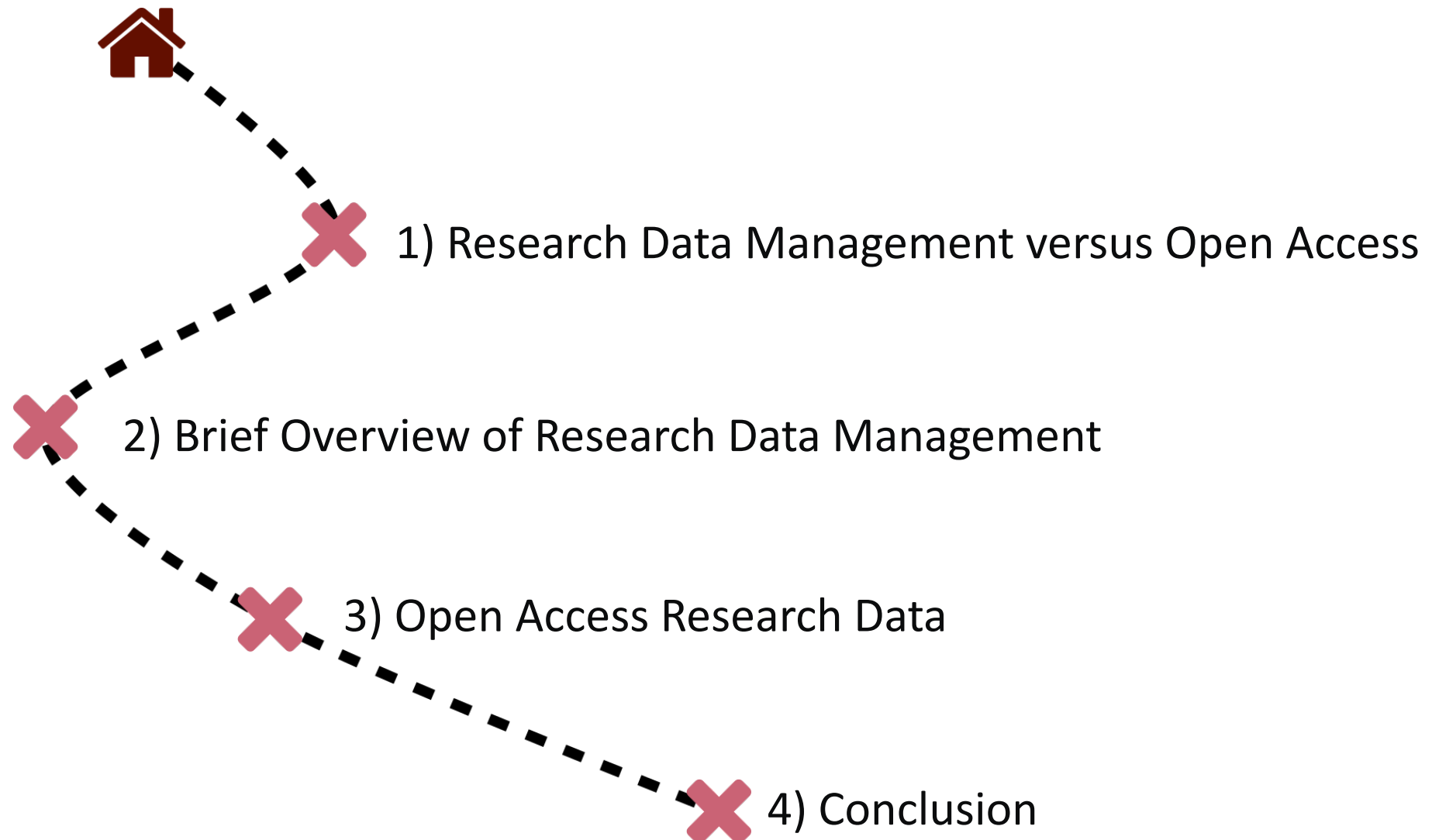


From Research Data Management to Open Access Data



Samuel Simango
Manager: Research Data Services
ssimango@sun.ac.za
28th February 2023

Presentation Overview

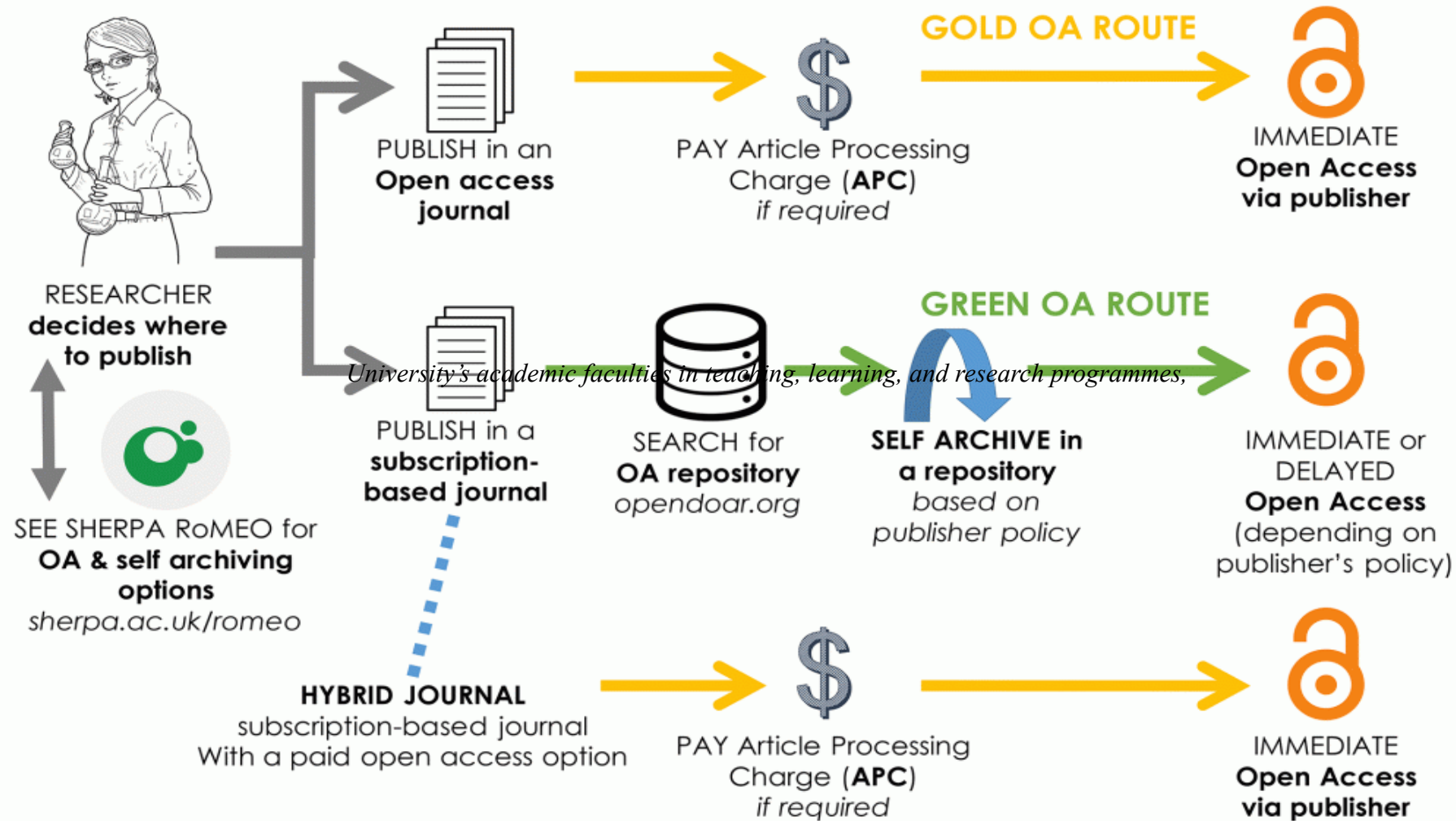


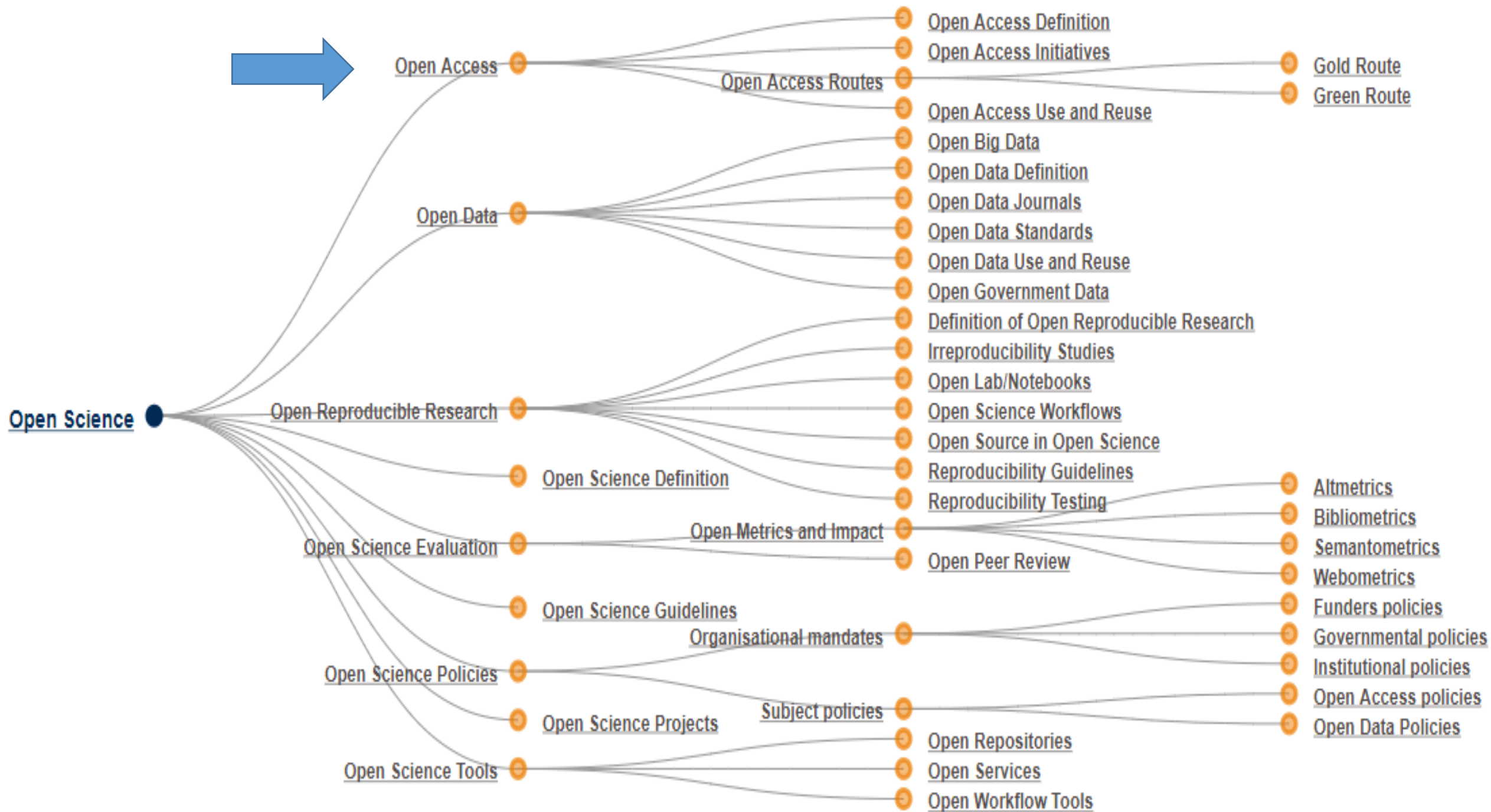
(1) Open Access vs Research Data Management

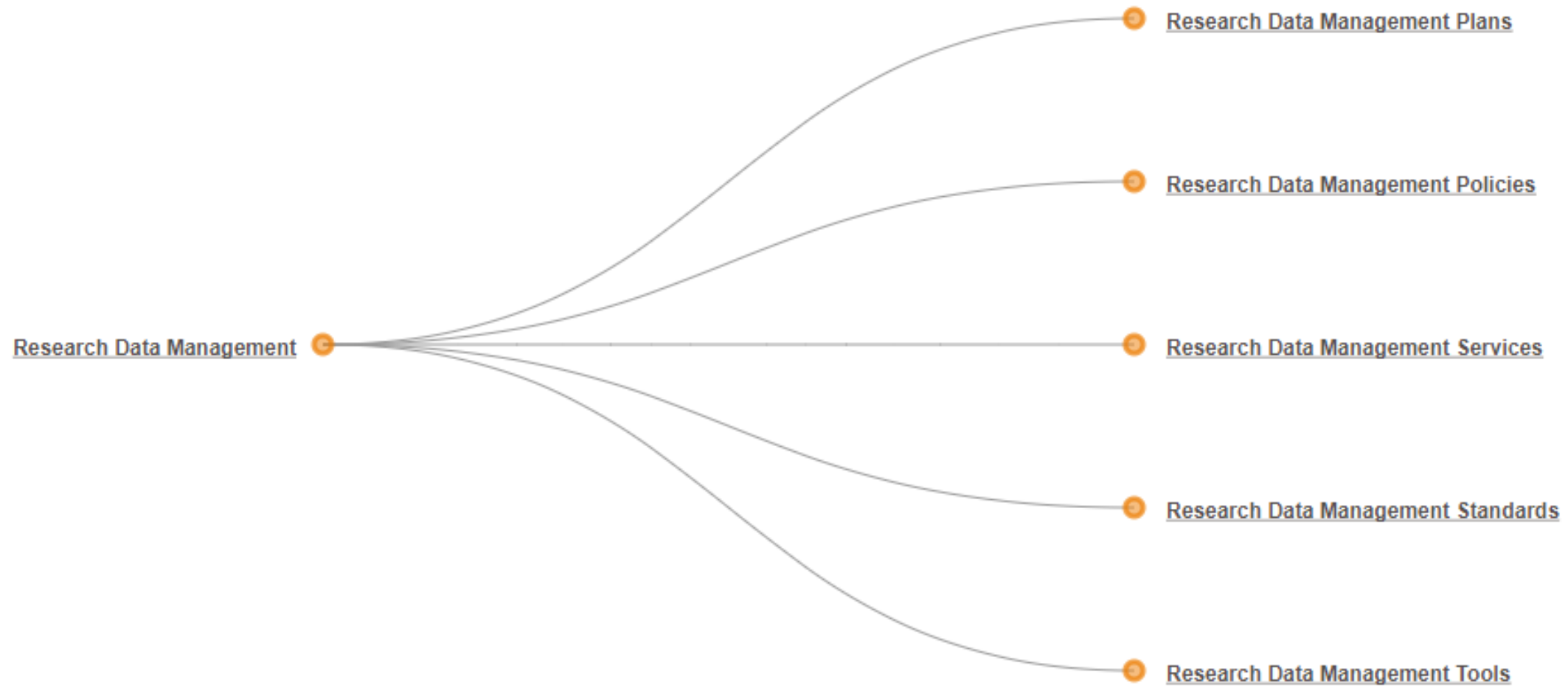
Berlin Declaration Definition of an Open Access Contribution

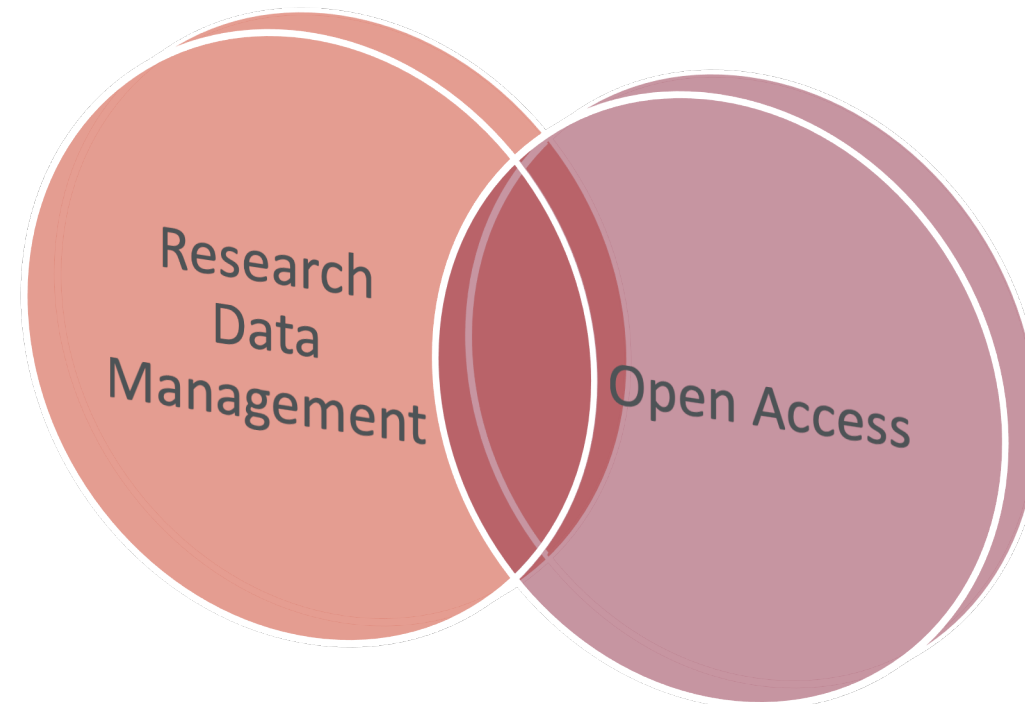
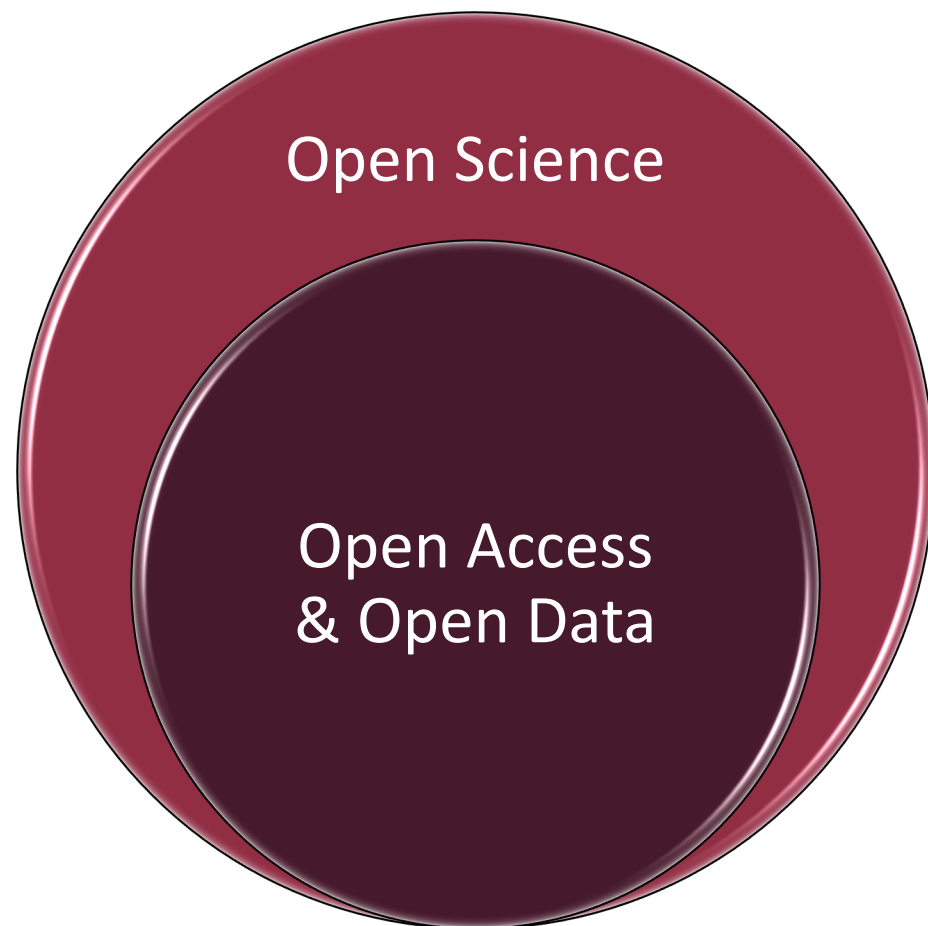
- Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.
 1. Open access contributions must satisfy two conditions: The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.
 2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.

Open Access Publishing









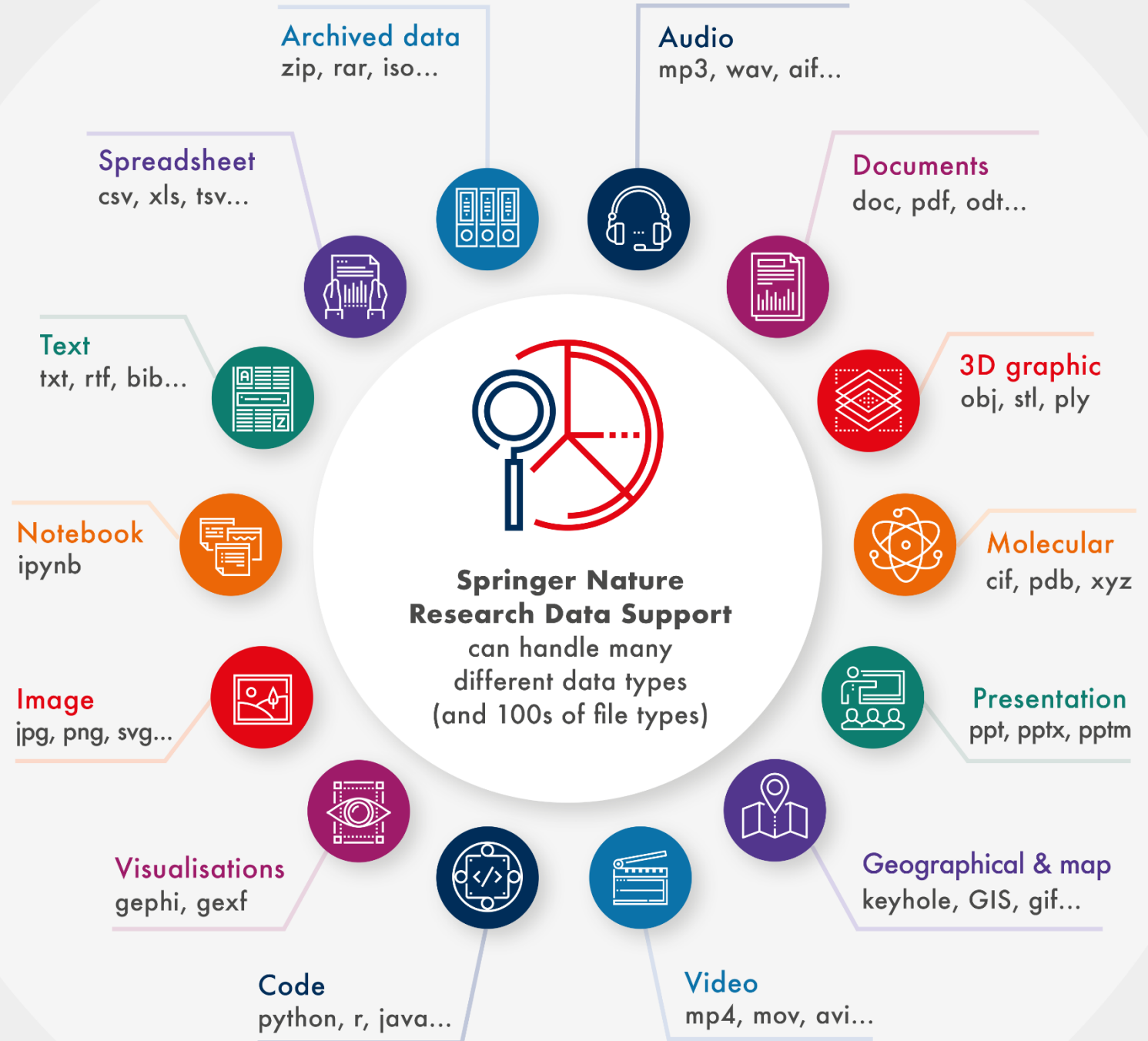
- Research Data Management is not the same thing as Open Access.



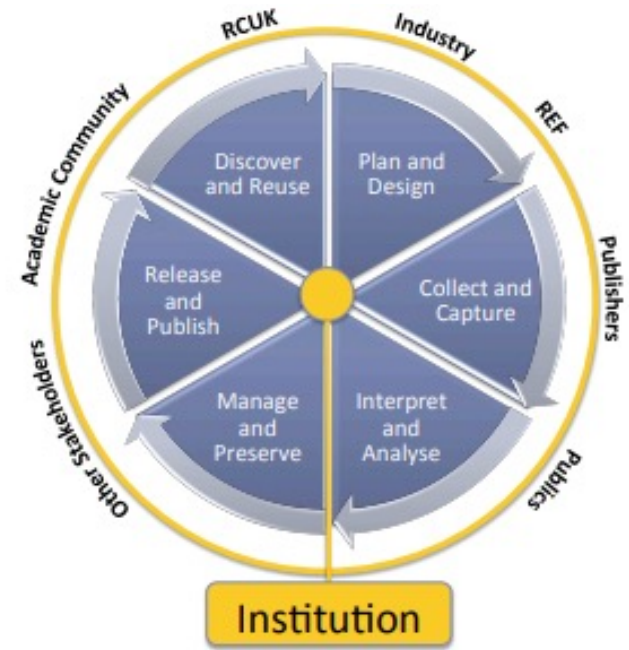
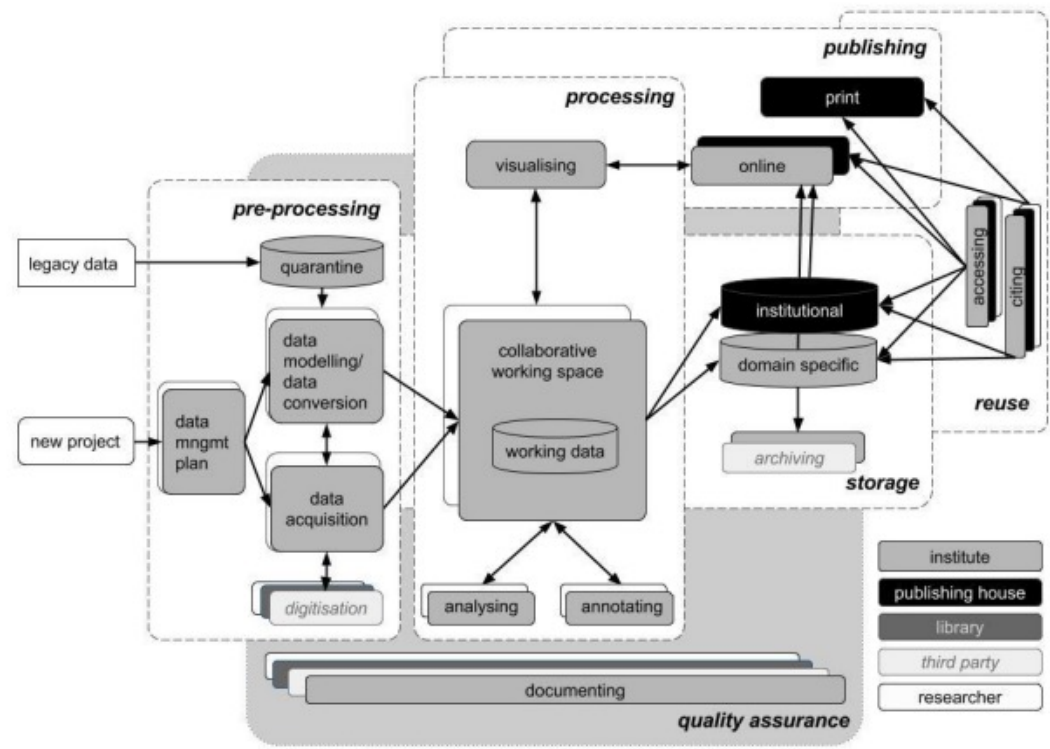
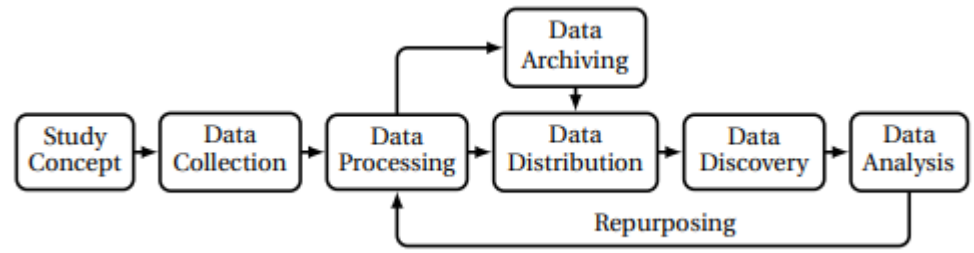
(2) Brief Overview of Research Data Management

What are research data?

Any data collected or generated for the purpose of analysis, in order to generate or validate scientific claims.

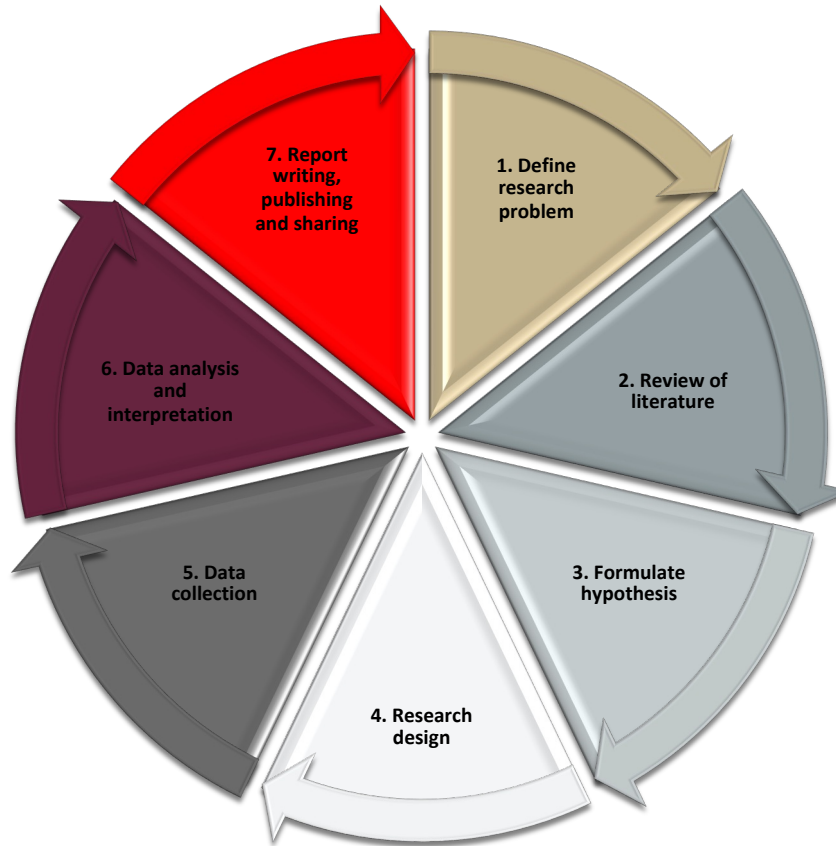


Research Data Lifecycles

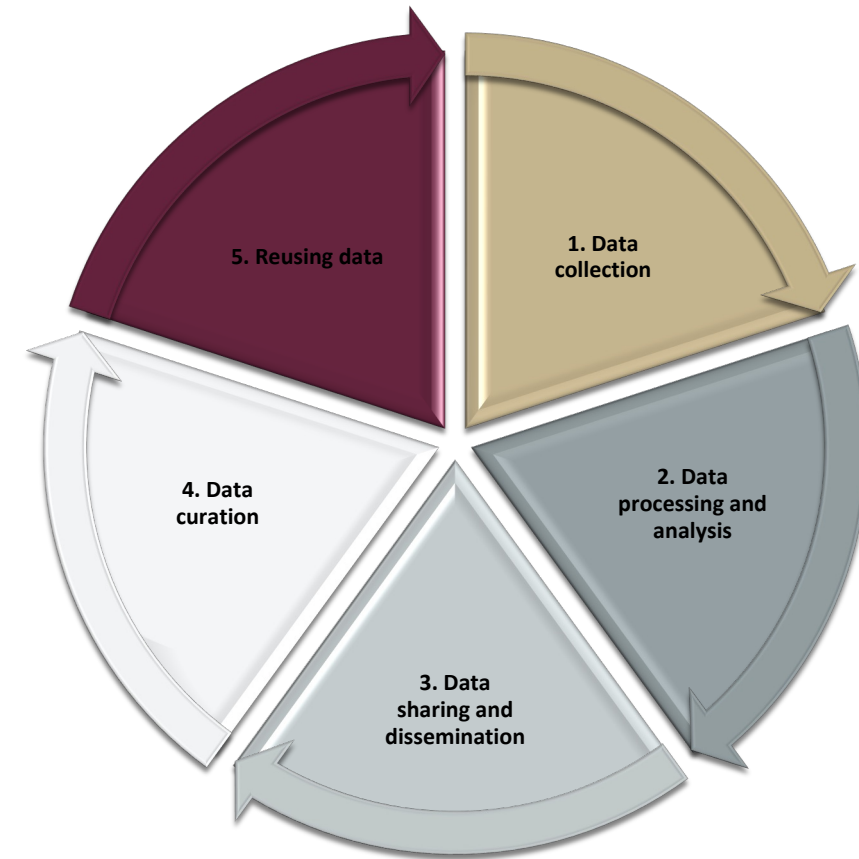


Conceptualising Research Data Management

Research lifecycle vs data lifecycle

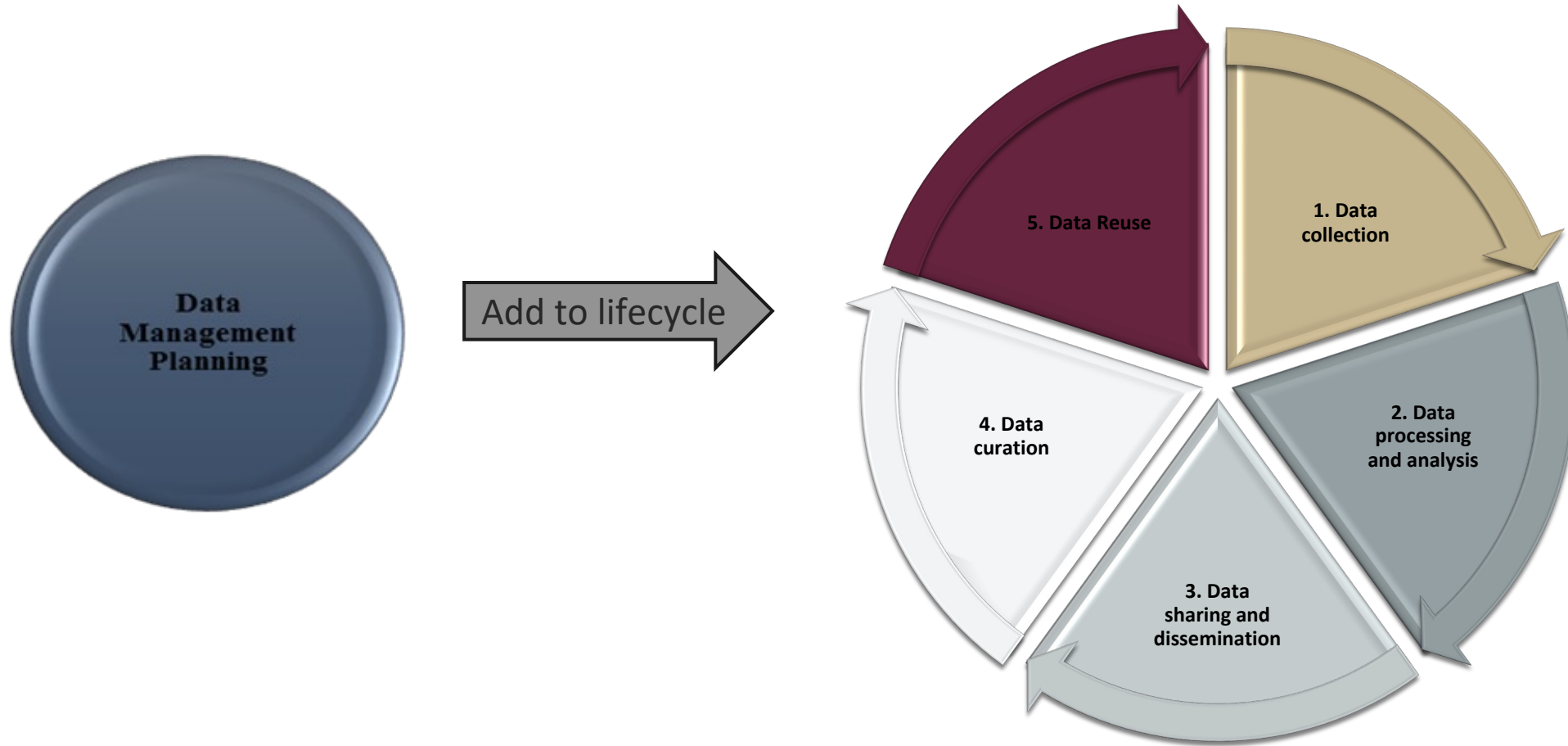


Research lifecycle



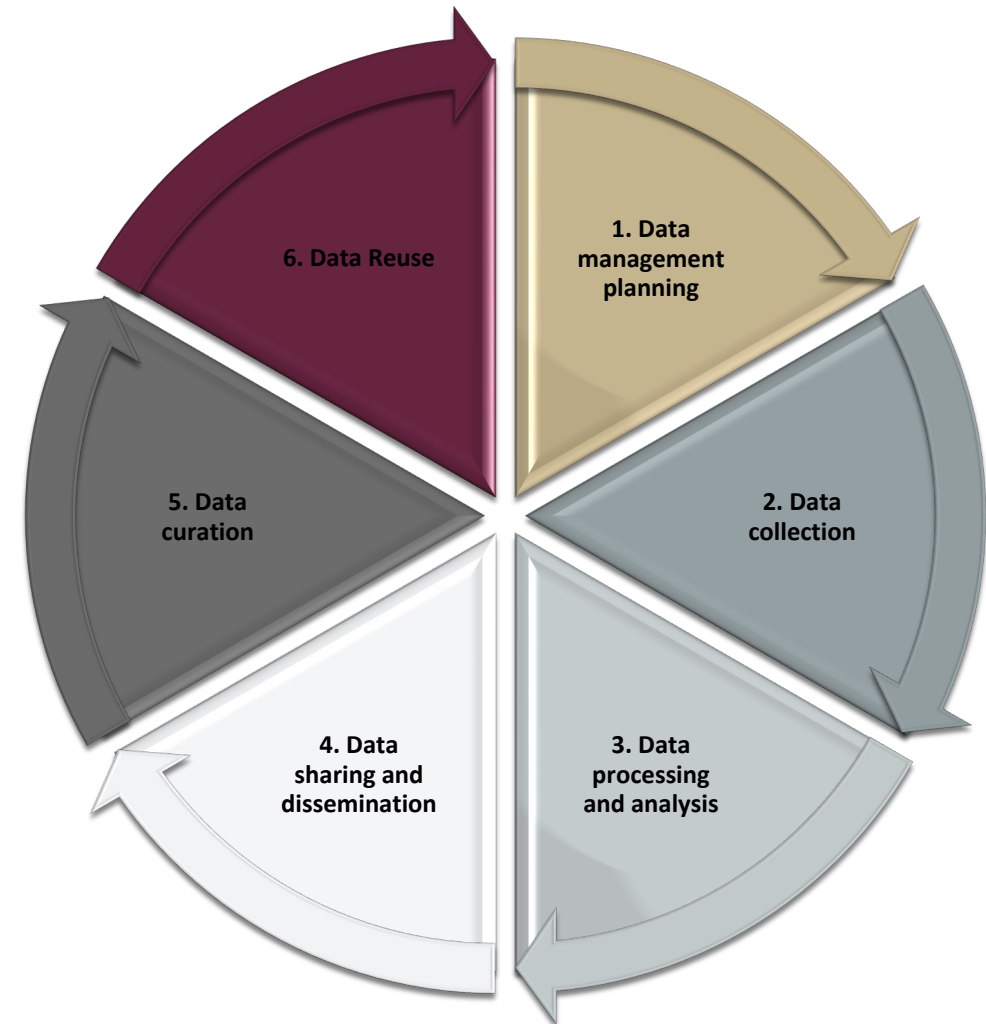
Research data lifecycle

Emergence of Research Data Management Lifecycle



Research Data Management Lifecycle

1. Data management planning
2. Data collection
3. Data processing and analysis
4. Data sharing and dissemination
5. Data curation
6. Reusing data



Definition of Research Data Management

What is Research Data Management?

- **Research Data Management is a process which entails:**
 - 1) **planning for the manner in which research data will be managed during and after a research project; and**
 - 2) **controlling the collection, processing, analysis, sharing, dissemination, curation and reuse of research data.**

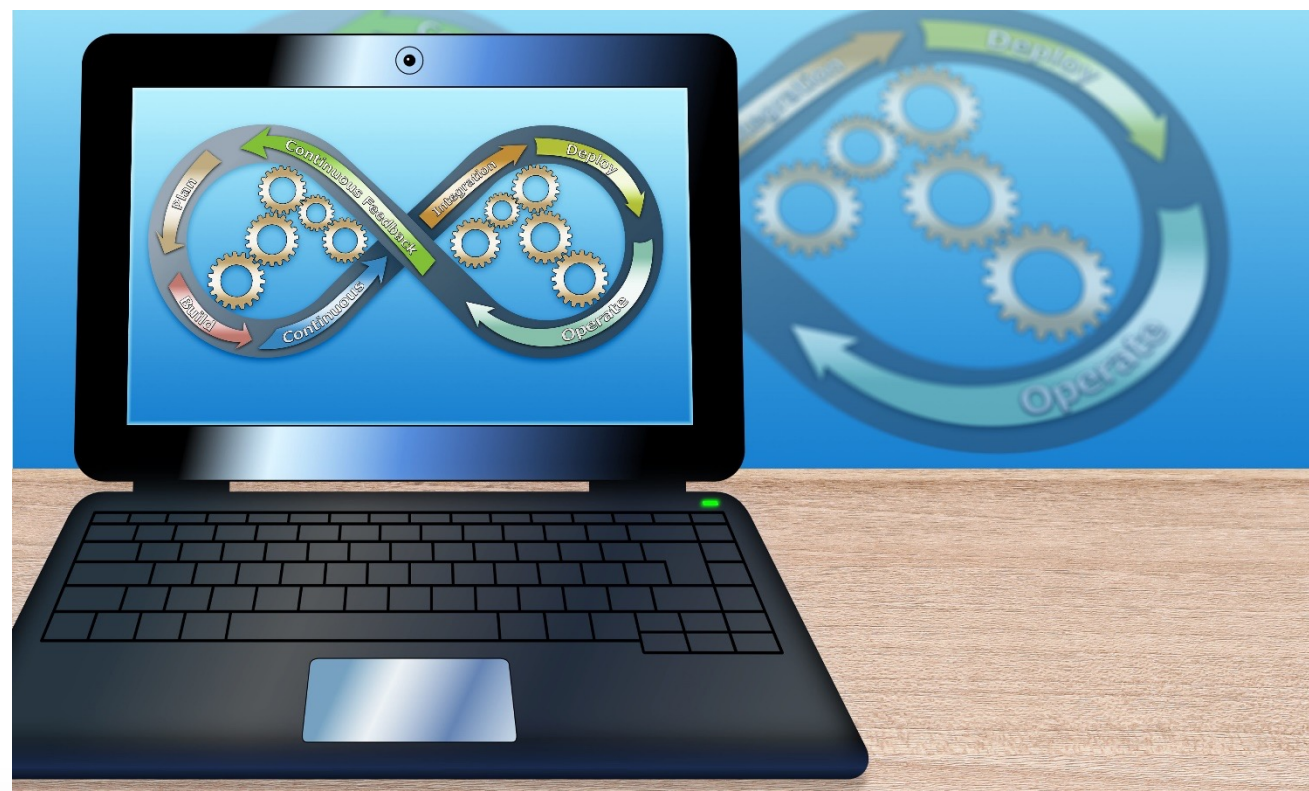
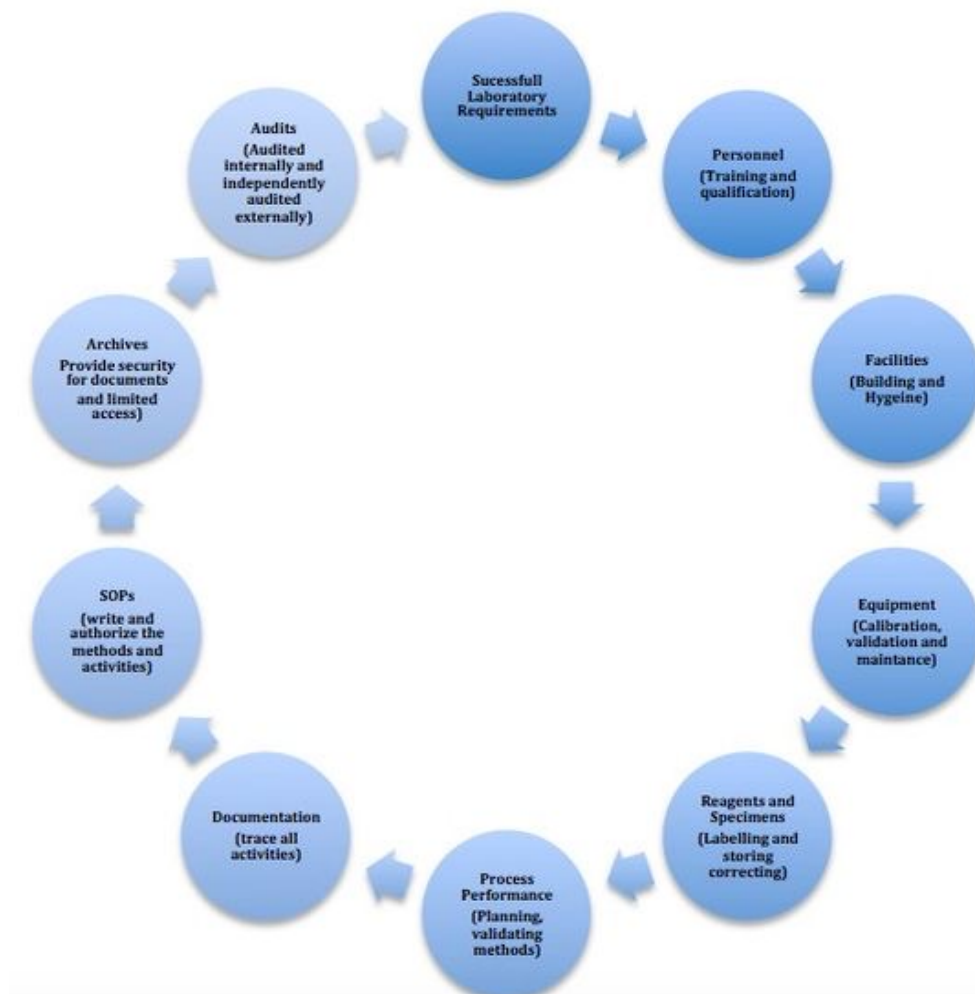
Goals of Research Data Management

- Researchers should take cognisance of the three goals of research data management. These three goals are:
 1. to assure that data are capable of supporting the analysis,
 2. to assure that all operations performed on data are traceable and
 3. to facilitate use of data by people other than those involved in their original collection, management, and analysis.

Goals of Research Data Management

- Researchers should take cognisance of the three goals of research data management. These three goals are:
 1. to **assure** that data are capable of supporting the analysis,
 2. to **assure** that all operations performed on data are traceable and
 3. to facilitate use of data by people other than those involved in their original collection, management, and analysis.

- Where quality assurance needs to be addressed



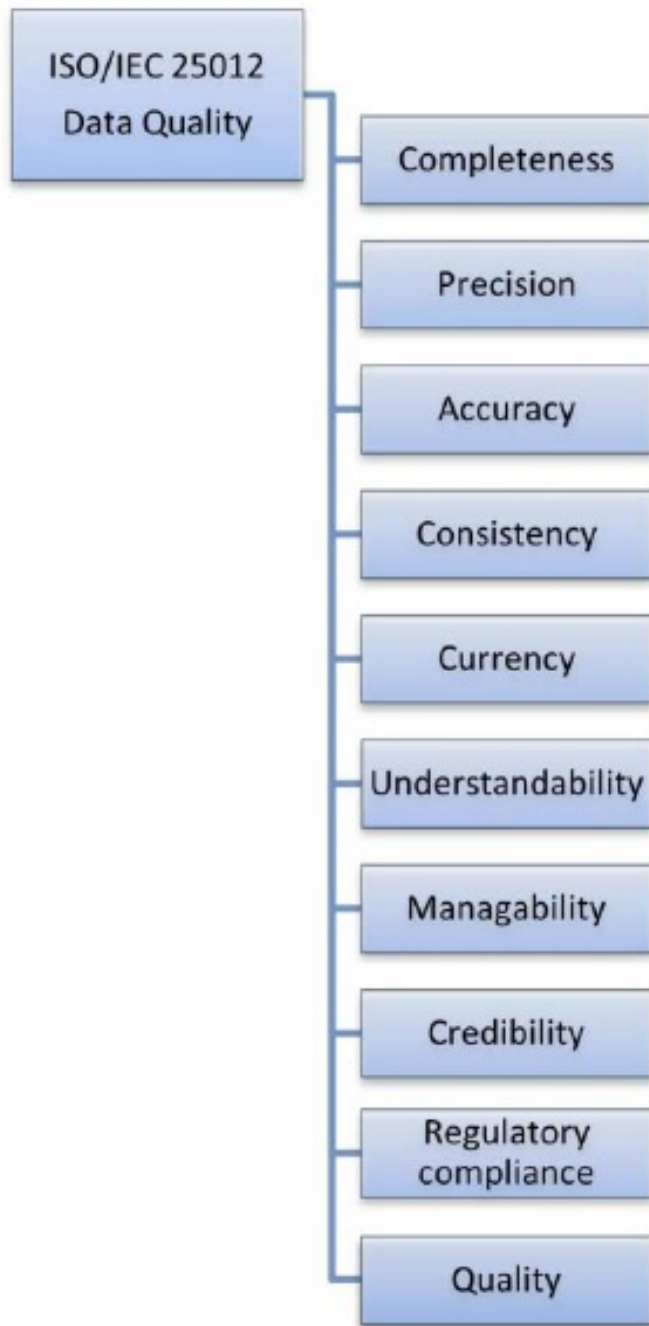
- Where quality assurance needs to be addressed
- Quality control(s) must also be in place



- What is quality?

An **attribute** or a **property** characteristic of an object

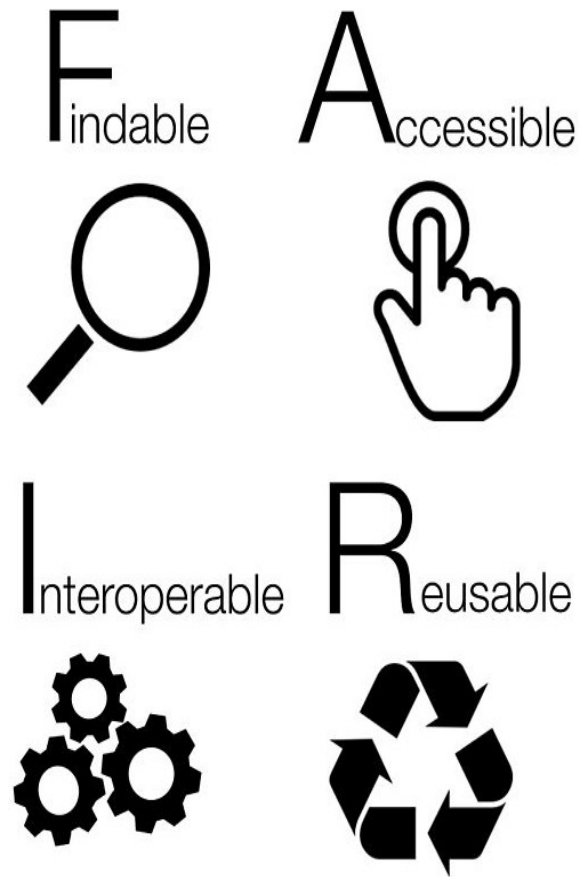
Quality



Data quality attributes

Attribute	What it means	Example of good practice	Example of bad practice	Metrics
Consistency	No matter where you look in the database, you won't find any contradictions in your data.	Your payment system shows that Jane Brown has made 5 purchases this month, and CRM system contains the same information.	Your payment system shows that Jane Brown has made 5 purchases this month, while CRM system shows she has made only 4.	The number of inconsistencies.
Accuracy	The information your data contains corresponds to reality.	Your customer's name is Jane Brown. And this is exactly how it's reflected in your CRM.	In your CRM, the customer's name is spelled Jane Brawn, though her actual name is Jane Brown.	The ratio of data to errors.
Completeness	All available elements of the data have found their way to the database.	You know that Jane Brown is born on 11/04/1975.	You have no idea how old Jane Brown is, as the date of birth cell is empty.	The number of missing values.
Auditability	Data is accessible and it's possible to trace introduced changes.	You can track down the changes made in Jane's data record. For example, on 12/5/2018, her phone number was changed.	It's impossible to trace down the changes in Jane's record.	% of cells where the metadata about introduced changes is not accessible.
Orderliness	The data entered has the required format and structure.	The entry for December 11, 2018 is in the format 12/11/2018.	The entry for December 11, 2018 is in the format 12/11/18, 12/11/2018 and even 11/12/18 (in your European stores).	The ratio of data of inappropriate format.
Uniqueness	A data record with specific details appears only once in the database.	You have only one record for Jane Brown, born on 11/04/1975, who lives in Seattle.	You have multiple duplicate records for Jane Brown.	The number of duplicates revealed.
Timeliness	Data represents reality within a reasonable period of time or in accordance with corporate standards.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. The customer's name was corrected the next day.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. Her name was corrected only in a month.	Number of records with delayed changes.

FAIR Data Principles



FAIR Principles

Compliance



Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.

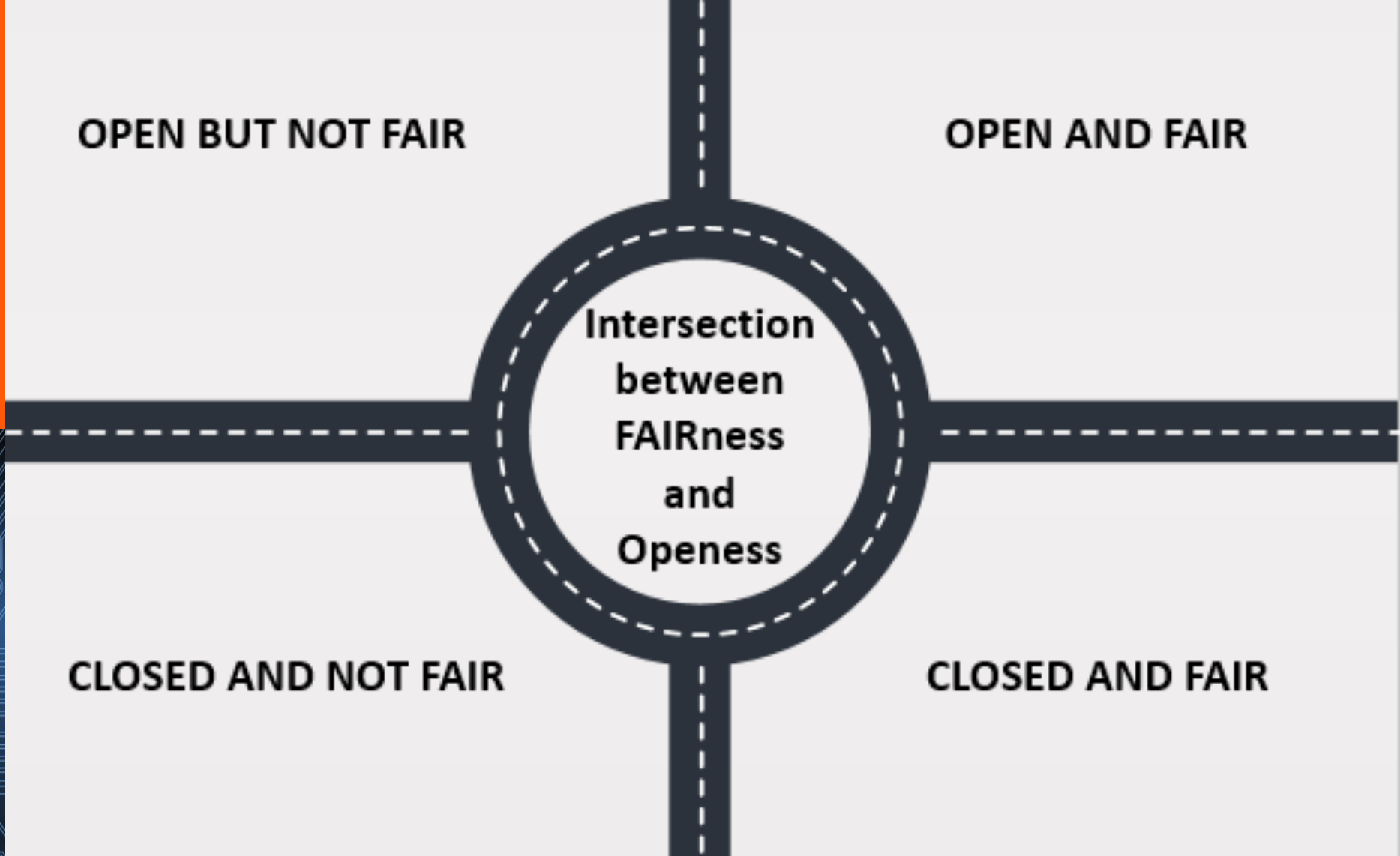


Reusability

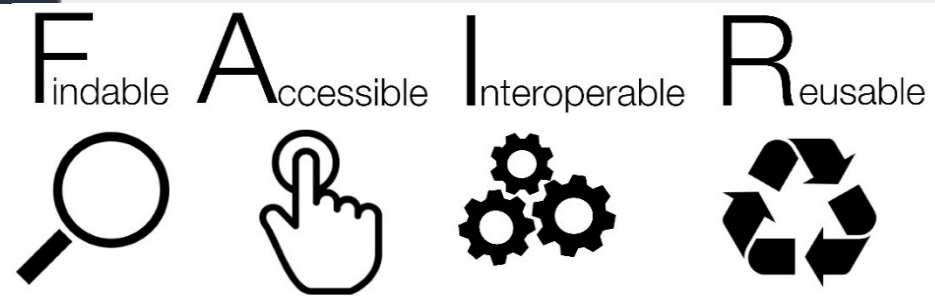
Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines

- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.





Not FAIR



(3) Open Access Research Data

Goals of Research Data Management

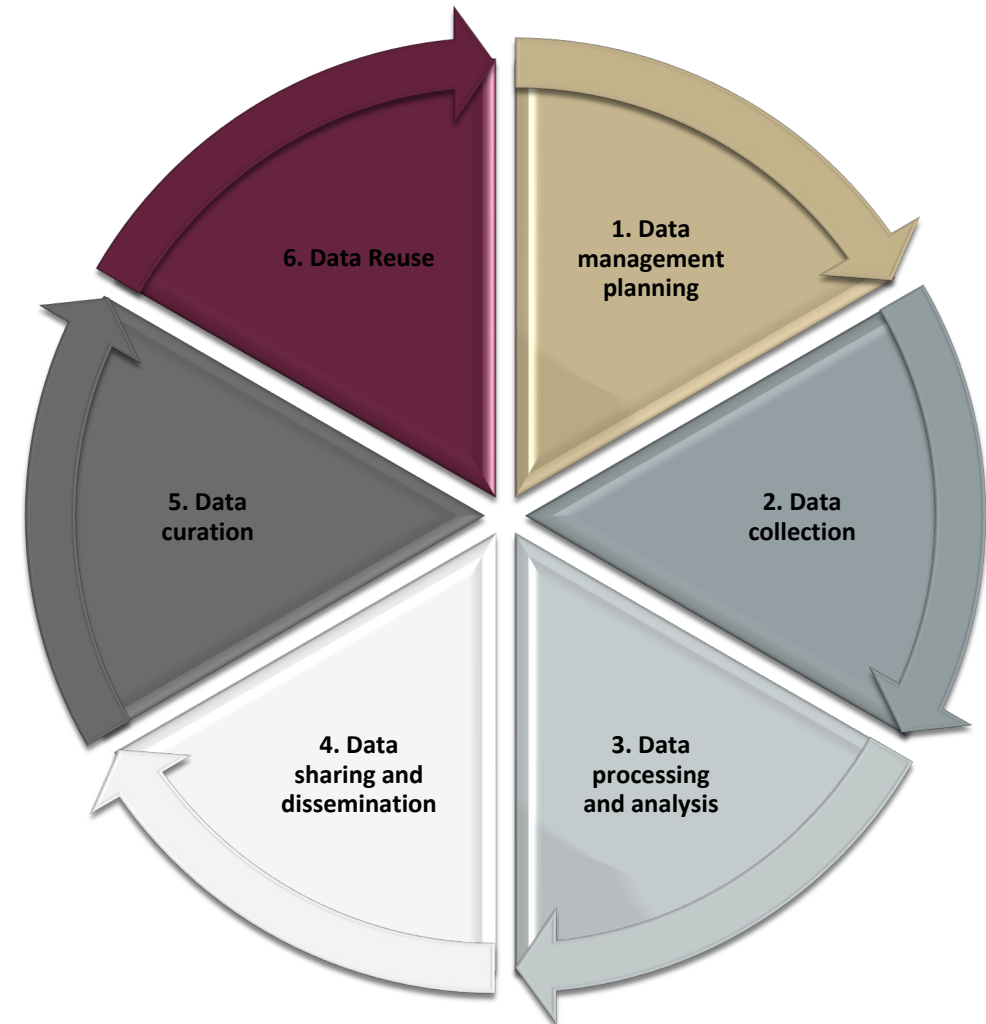
- Researchers should take cognisance of the three goals of research data management. These three goals are:
 1. to assure that data are capable of supporting the analysis,
 2. to assure that all operations performed on data are traceable and
 3. to facilitate use of data by people other than those involved in their original collection, management, and analysis.

Goals of Research Data Management

- Researchers should take cognisance of the three goals of research data management. These three goals are:
 1. to assure that data are capable of supporting the analysis,
 2. to assure that all operations performed on data are traceable and
 3. to facilitate use of data by people other than those involved in their original collection, management, and analysis.

Research Data Management Lifecycle

1. Data management planning
2. Data collection
3. Data processing and analysis
4. Data sharing and dissemination
5. Data curation
6. Reusing data



Modern Academic Research Publication Model

Researcher conducts research

Generates

Research data

Published

Manner of publication

Generates

Scholarly paper

Published

Manner of publication

Supplementary data

Journal article

Data deposit

Research data repository

Data paper

domain-specific

Generic/multi-disciplinary

Data journal

Third-party

Third-party

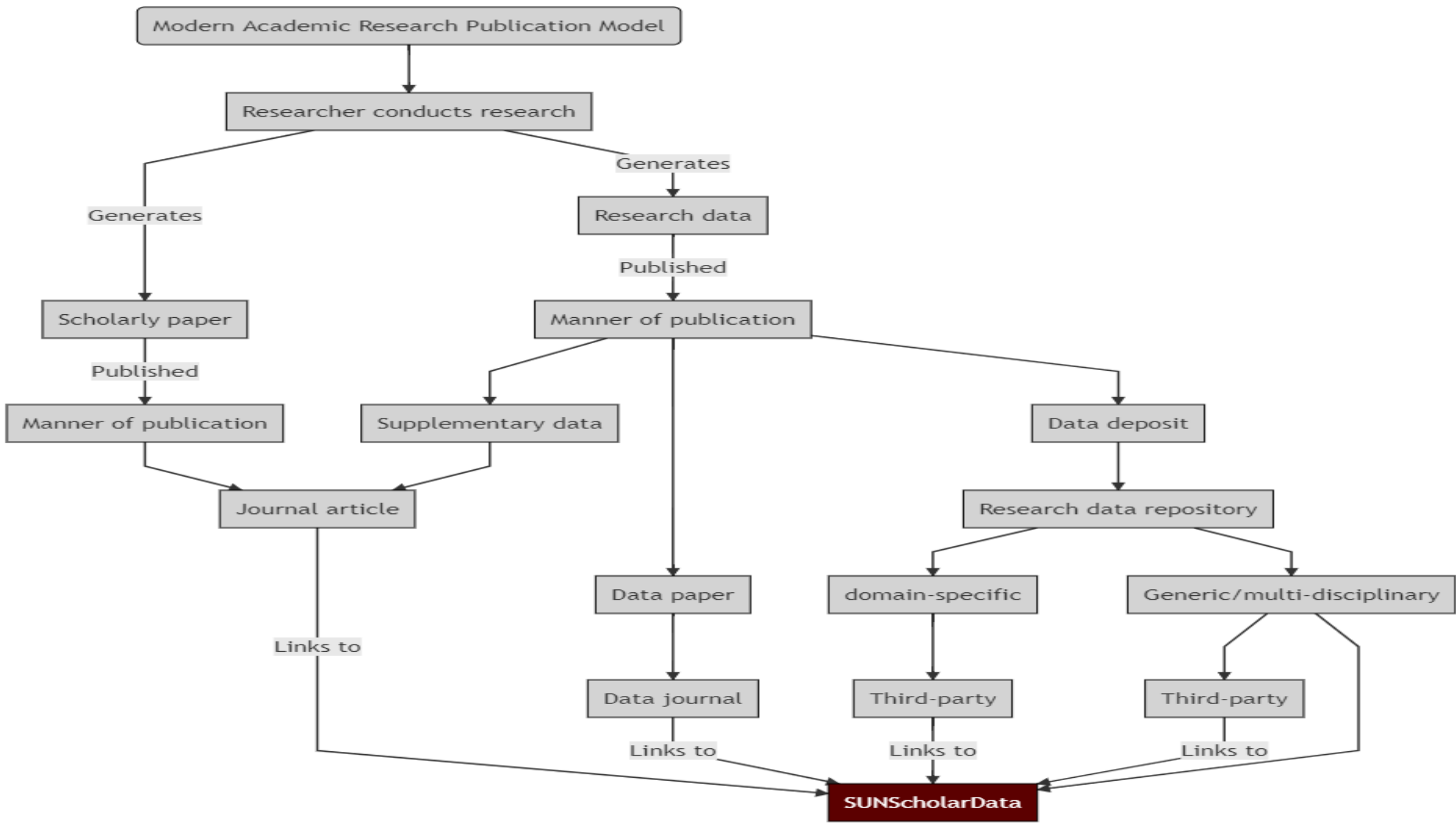
Links to

Links to

Links to

Links to

SUNScholarData



(4) Conclusion

Takeaways

1. Research Data Management is not the same thing as Open Access.
2. Research Data Management is a process which entails planning and controlling for the manner in which research data are treated during and after a research project.
3. Research Data Management has a quality management aspect.
4. Research Data Management actually serves a quality management function whilst also contributing to the openness of research data.
5. Research data can be made openly accessible in a variety of ways.

Thank You
Dankie
Enkosi

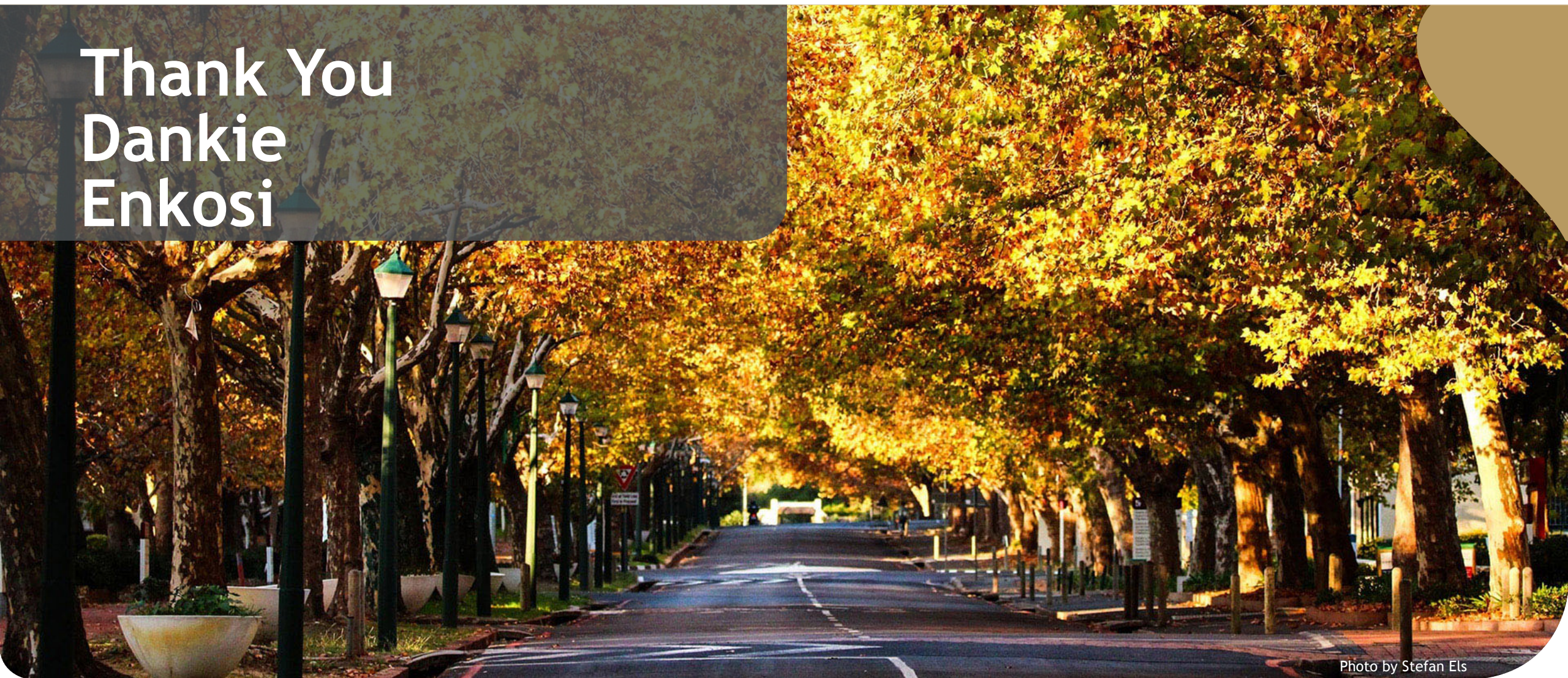


Photo by Stefan Els