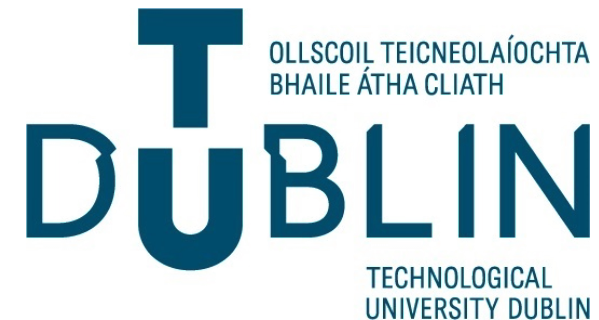


Data Management

Data Lifecycle

Dr Emma Murphy



Data Lifecycle

- Like other assets, data has a lifecycle. To effectively manage data assets, organizations need to understand and plan for the data lifecycle.
- Well-managed data is managed strategically, with a vision of how the organization will use its data.
- A strategic organization will define not only its data content requirements, but also its data management requirements.

Data Lifecycle

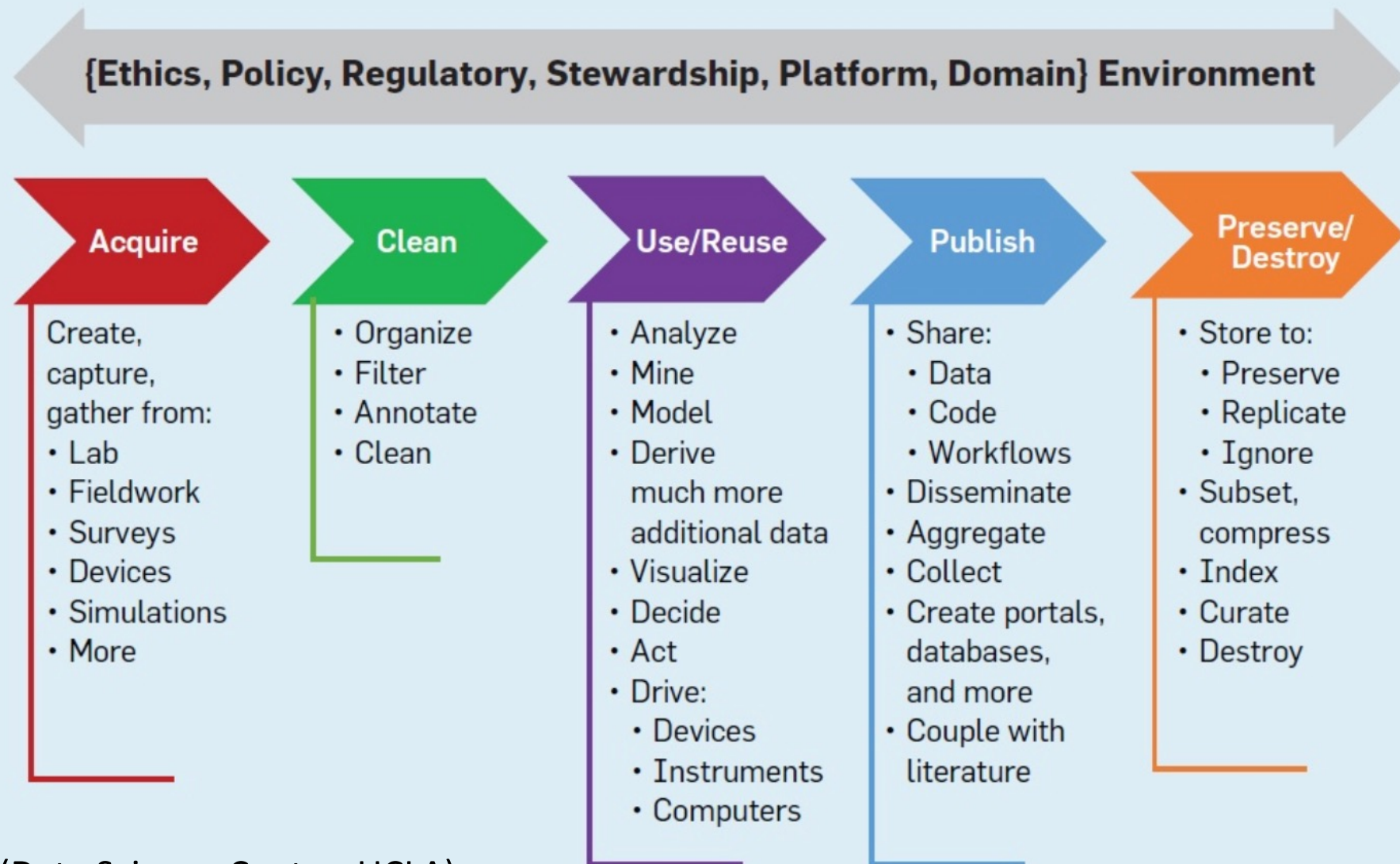
- The data lifecycle is based on the product lifecycle. It should not be confused with the systems development lifecycle.
- It includes processes that create or obtain data, those that move, transform, and store it and enable it to be maintained and shared, and those that use or or apply it as well as dispose of it.

Data Lifecycle

- As data is used or enhanced, new data is often created, so the lifecycle has internal iterations that are not shown on the diagram.
- Data is rarely static.
- Managing data involves a set of interconnected processes aligned with the data lifecycle.

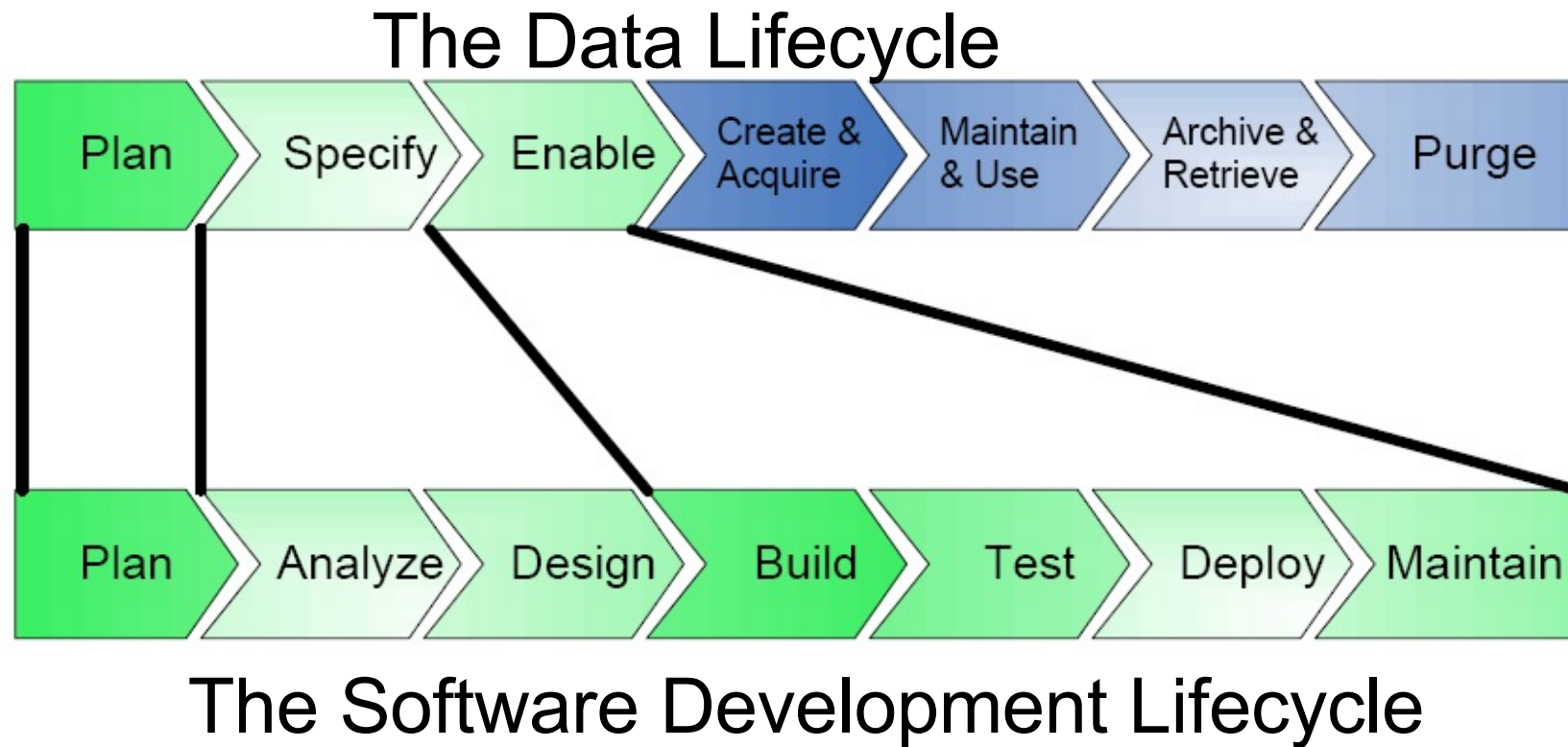
(DAMA, 2017)

Data Lifecycle



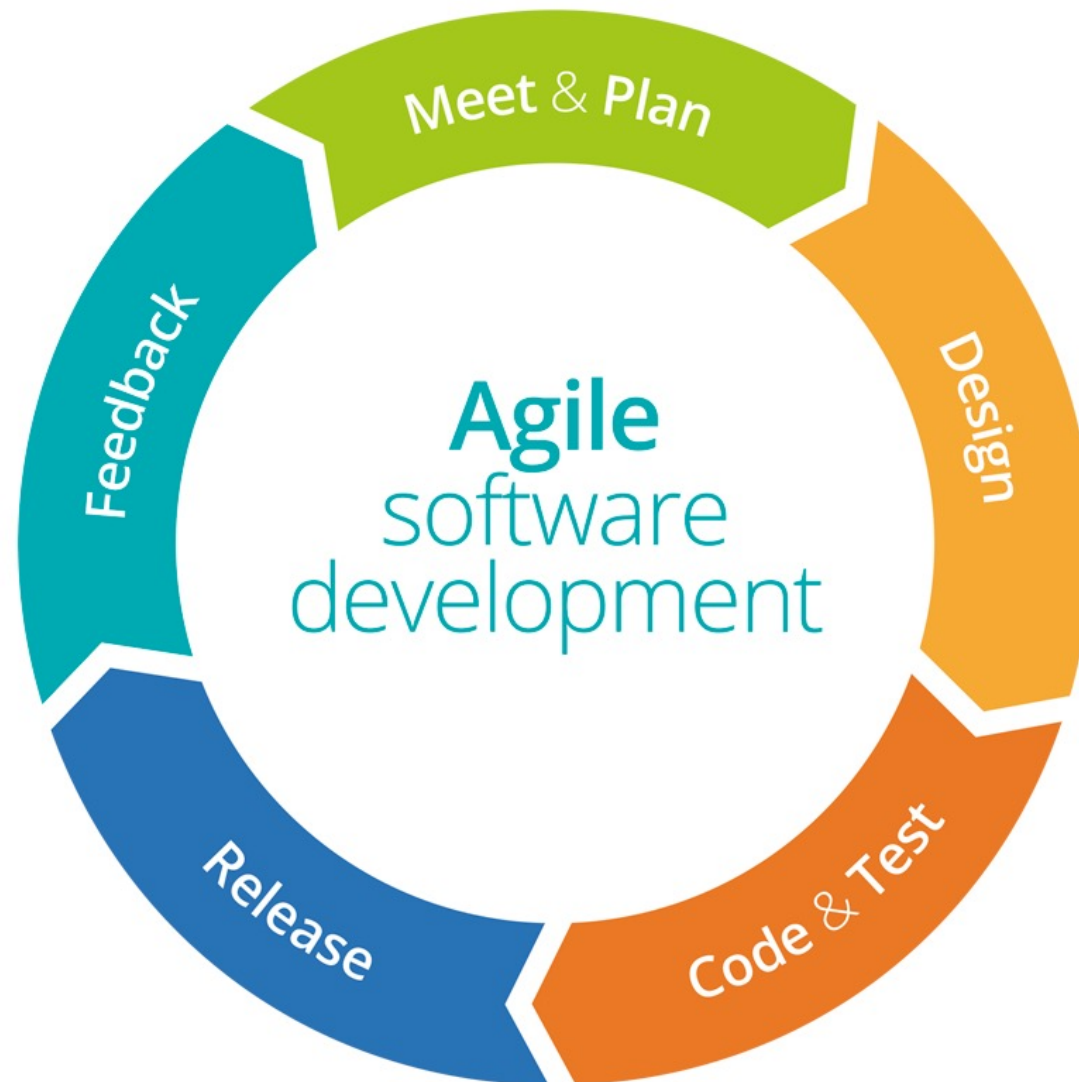
(Data Science Centre, UCLA)

Data vs Software Lifecycle

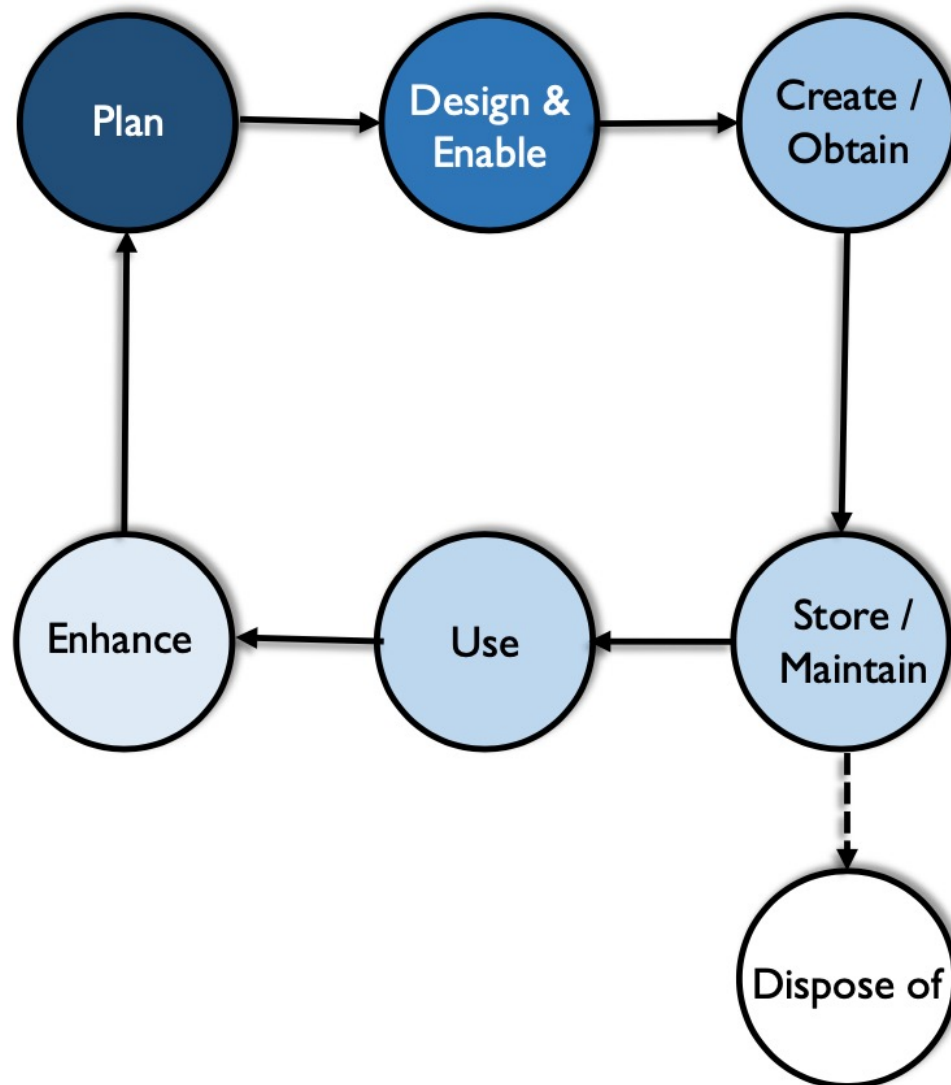


(DAMA, 2009)

Software Lifecycles



Data Lifecycle



Data Lifecycle Key Activities
(DAMA, 2017)

Digital Curation

- For long term use, preserving the data and protecting it from loss or obsolescence, particularly if it is important, irreplaceable information, is essential in order to sustain the data assets. “Digital curation is an ongoing process not a one-off action. It is a chain of activities only as strong as its weakest link”. Abbot (2010)
- Part of digital curation is discovering and re using data and ensuring the information is in a readable format for future use and to ensure the data is a trust worthy reliable source.

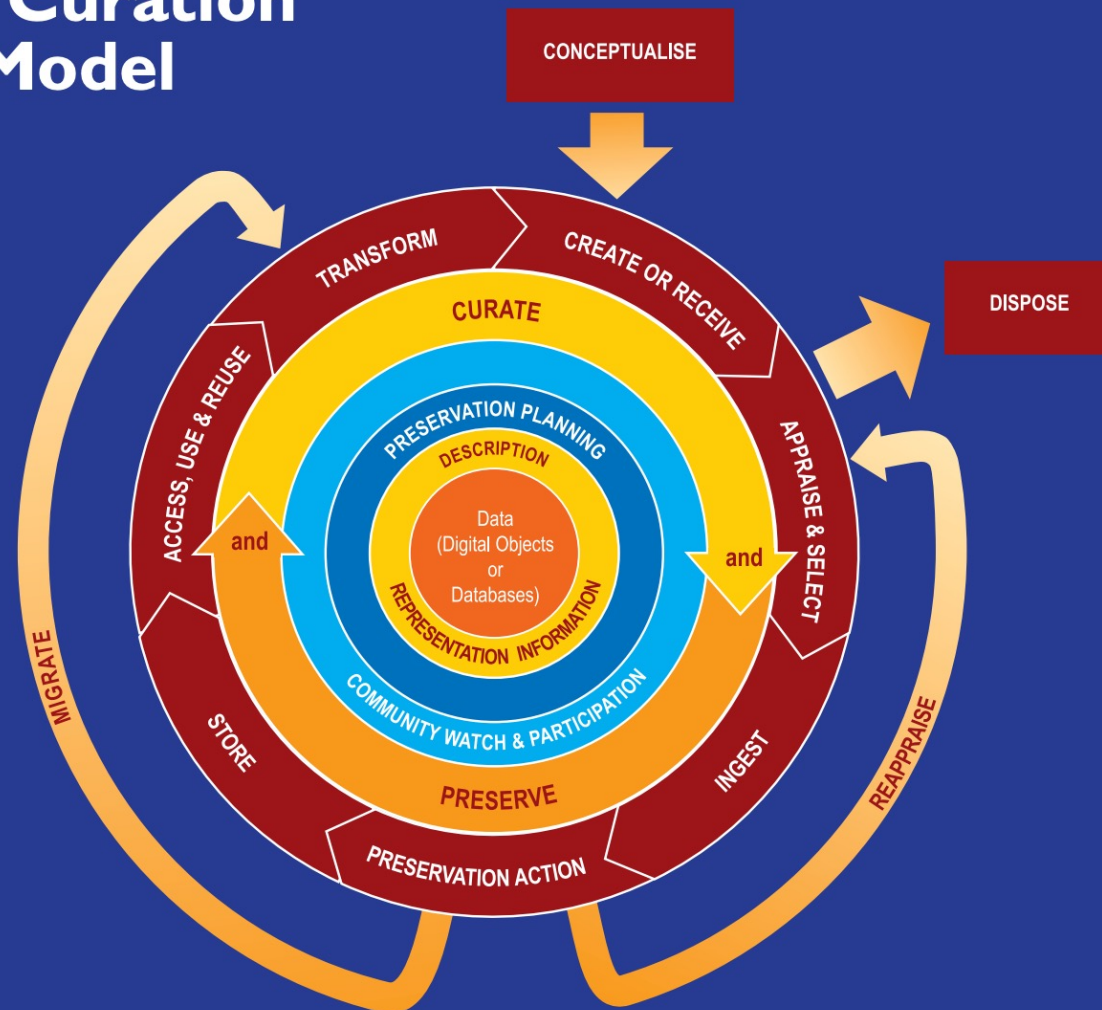
<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Digital Curation

- The Digital Curation Centre (DCC) - centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management.
- University of Edinburgh
- https://www.youtube.com/watch?v=6cuOdgvYRGM&feature=emb_logo

Research DM Lifecycle

The DCC Curation Lifecycle Model



Data Curation – Lifecycle actions

- **Description and Representation Information**

Assign administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, to ensure adequate description and control over the long-term. Collect and assign representation information required to understand and render both the digital material and the associated metadata.

- **Preservation Planning**

Plan for preservation throughout the curation lifecycle of digital material. This would include plans for management and administration of all curation lifecycle actions.

- **Community Watch and Participation**

Maintain a watch on appropriate community activities, and participate in the development of shared standards, tools and suitable software.

- **Curate and Preserve**

Be aware of, and undertake management and administrative actions planned to promote curation and preservation throughout the curation lifecycle.

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Data Curation - Sequential Actions

- Conceptualise

- Create or Receive

- Create data including administrative, descriptive, structural and technical metadata. Preservation metadata may also be added at the time of creation.

- Appraise and Select

- Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements.

- Ingest

- Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance, policies or legal requirements.

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Data Curation - Sequential Actions

- **Preservation Action** - Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats.
- **Store** - Store the data in a secure manner adhering to relevant standards.
- **Access, Use and Reuse** - Ensure that data is accessible to both designated users and reusers, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable.
- **Transform** - Create new data from the original, for example: by migration into a different format, or by creating a subset, by selection or query, to create newly derived results, perhaps for publication

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Data Curation - Occasional Actions

- **Dispose** - Dispose of data, which has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Typically data may be transferred to another archive, repository, data centre or other custodian. In some instances data is destroyed. The data's nature may, for legal reasons, necessitate secure destruction.
- **Reappraise** - Return data which fails validation procedures for further appraisal and re-selection.
- **Migrate** - Migrate data to a different format. This may be done to accord with the storage environment or to ensure the data's immunity from hardware or software obsolescence.

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Data lifecycle

- The focus of data management on the data lifecycle has several important implications:
- Creation and usage are the most critical points of the data lifecycle
- Data quality must be managed throughout the data lifecycle
- Metadata quality must be managed throughout the data lifecycle
- Data security must be managed through the data lifecycle
- Data management efforts must focus on the most critical data

Data Disposal

- Effective data management requires clear policies and procedures, regarding retention and disposal of records.
- A retention and disposition policy will define the timeframes during which documents for operational, legal, fiscal or historical value must be maintained.
- This defines when inactive documents can be transferred to a secondary storage facility, such as off-site storage.

Data disposal

- Non-value-added information should be removed from the organization's holdings and disposed of to avoid wasting physical and electronic space, as well as the cost associated with its maintenance.
- There is also risk associated with retaining records past their legally required timeframes. This information remains discoverable for litigation.