Research Data Toolkit

This toolkit is designed to support researchers with Research Data Management throughout a project. Research Data Management (RDM) is split into different stages, commonly known as the Research Data Lifecycle.

This toolkit includes key misconceptions and answers to Research Data myths, and how good RDM practices are beneficial to research projects. It also contains sections of each stage of the research data lifecycle, with each one containing relevant resources and helpful tips.

Toolkit sections:
- General Terms & Resources
- Data Myths
- Research Data Policies

- Data Planning
- Data Collection
- Data Processing & Analysis
- Data Storing & Preserving
- Data Sharing & Publishing
- Data Reuse & Citing

Research Data Lifecycle

- Data Planning
- Data Collection
- Data Analysis
- Data Storing and Preserving
- Data Sharing and Publishing
- Data Reuse

Research Data

Research Data is any data found or generated through a research project. Research data can take a variety of forms, including spreadsheets, lab work and field notes, interview transcripts or surveys and questionnaires, or multimedia and audiovisual material.

Research Data Management

Research Data Management (RDM) are the processes of managing and maintaining research data within any research project. This includes planning and documenting data in understandable and accessible ways as well as managing and storing research data securely and in line with university policies and data protection legislation.

General RDM resources

- <u>DDC Curation Life Cycle Model:</u> This resource focuses on RDM workflows for different roles and responsibilities within a research project or organisation.

- <u>JISC RDM Toolkit:</u> This toolkit is designed for the general research community and contains a helpful RDM checklist.

- <u>RDM Toolkit:</u> This toolkit focuses on the data life cycle for scientific studies and projects and provides guides on how to make data as FAIR as possible.

This section highlights 4 key common misconceptions of Research Data and Open Data. These misconceptions are addressed by key benefits of RDM and highlight some useful tools of the University's data repository, RaYDaR.

RDM isn't applicable to some projects as not all research has data, therefore there is nothing to share.

All research contains data. Data can consist in different forms depending on research type and academic discipline. The way to consider data in relation to RDM is how it links to the key research questions and how it validate, support and demonstrate findings. In the case of Art and Humanities based disciplines this could include sound recordings, transcripts, sketches or annotations generated or built on throughout a project.

Data Myths

Sharing data increases chances of others taking and manipulating data, and then taking the credit. This is particularly problematic when sharing a dataset before the research output is published.

Sharing research data is part of the Open Data landscape. It is actively encouraged that open research data is built on by others.*

- Creators are protected by Open Licences which are chosen by the creator and attached to a dataset on RayDaR. Creative Commons Licences also ensure creator attribution.

- RaYDaR has an embargo option so datasets can be hidden from public view until a research output has been published. The use of private links can then be used with publishers to view and discuss data.

 *Published, open datasets cannot be adapted in the original file by others in a repository – only the creator of a dataset has the power to change and reupload the original file with the assistance of repository admins.

Data Myths

Some research data has identifiable information so it can't be shared openly. By sharing this type of data researchers break the terms of ethical policies and contracts.

All data containing identifiable information should be anonymised in all research and research data outputs. Information about the storing and sharing of personal information and data generated from human participants or other sources should be outlined in the Data Management Plan and discussed with research supervisors or ethical leads. Human participants should give informed consent to their data being shared openly.

In some cases data where highly sensitive, such as NHS or government based projects. Processes of data handling and storage should be outlined in additional policies.

YORK
ST JOHN
UNIVERSITY

Library &
Learning Services

Est. 1841

Publishing a research output can be time consuming and may have financial implications– publishing research data will be similar.

Publishing research data is less likely to involve rigorous publisher workflows such as peer review, and it may not have any publisher involvement at all. Publishing research data on repositories like RaYDaR can be done quickly and at no extra cost to researchers. It is normal for it to take longer when using RaYDaR for the first time and it depends on how many datasets, and their size, are being uploaded. For more information contact ray@yorksj.ac.uk

Est. 1841

YORK
ST JOHN
UNIVERSITY

Library &
Learning Services

Before starting any research project or research data plan you need to be aware of RDM policies and frameworks.

Key YSJU Policies

- **YSJU Research Data Management Policy**
- **YSJU Open Research Policy**
- **YSJU Data Protection Policy**
- **YSJU Ethics and Integrity**
- **YSJU Intellectual Property Policy**

Additional policies

- **UKRI framework on research data** - this provides seven core principles surrounding data sharing. The YSJU research data policy aligns with the UKRI framework.
- **Concordat on Open Research Data** – We ask all YSJU researchers to be aware of the Concordat Principles

YORK ST JOHN UNIVERSITY
Est. 1841
Library & Learning Services

FAIR Data Principles

The <u>FAIR data principles</u> are a **key standard for high quality research data management.** It is something to be aware of at all stages of a research project and the research data lifecycle. There are four FAIR principles.

Findable - Data should be easy to find. It should be described well and be machine readable.

Accessible – Data should be accessible in different formats and should be clear to understand.

Interoperable – Data should have standardised vocabularies so it can work across different systems and processes.

Reusable – Data should be well described so others can replicate or combine it.

FAIR Data Resources:

- **Go FAIR – The FAIR Data Principles**
- **How to make the most out of your metadata**

Est. 1841
YORK ST JOHN UNIVERSITY
Library & Learning Services

The planning stage touches on all areas of the research data cycle. Researchers decide on the types of data that will be collected, how it will be documented and analysed and how it may be shared in the future.

Data Management Plans

A data management plan (DMP) is a formal document that guides a research project. It allows researchers to consider and address risks or issues related to working or managing data.

Writing a DMP is good research practice and is a growing requirement of most research funders.

DMPs are expected to evolve throughout a project and all changes should be recorded within it.

Key DMP resources

- **DMPOnline**: A resource provided by the Digital Curation Centre that provides guidance and examples of DMPs.

- **DMPTool:** This website provides DMP templates relating to funder criteria.

What to include in DMPs

Every research project is unique so what to include in DMPs differs. However as a general guide you should consider:

- **The type of data that will be created or collected.** This includes different data formats or sensitive data.
- **Data Management responsibilities.** This includes ethical or legal responsibilities or complying with funder requirements.
- **Policies which relate to managing the data.** This includes YSJU data policies and research funder policies.
- **Information about ownership of data or access rights.** This is important when working with third party data.
- **Information about data archiving or sharing arrangements.**
- **Explanation of how the data will be organised and stored securely.**

Research Data Toolkit

Data Collection

The data collection stage involves considering collection methods and tools which should be outlined in DMPs. Different tools may be used for different types of data and projects may have a large range of data collection methods.

Documenting collected data

It's important that all data is documented to ensure the research project is beneficial in the long term.

Clear and consistent documentation of data is valuable to additional researchers who may re-use data or replicate the data collection methods of the project.

Tips for successful long-term documentation of data:

- Keep naming conventions of files and folders consistent and meaningful such as *Transcript 1, Transcript 2, Transcript 3.*

- Keep a record of where you sourced the data, and make sure research data is cited correcting in research reports.

- Make sure that any documented data is placed and stored in an accessible format. This allows other researchers to understand the data and convert data formats efficiently.

YSJU software for data collection:

- **Microsoft Forms:** This software is part of Office 365 and can provide personalised survey designs. It returns information directly into an Excel spreadsheet.

- **Qualtrics Surveys:** This software is a comprehensive online tool. Everyone at York St John can access this with university credentials at https://yorksj.eu.qualtrics.com/login

Further information about software can be found through YSJU's Digital Training and Support webpages.

YORK ST JOHN UNIVERSITY
Est. 1841
Library & Learning Services

This section focuses on different processes and methods of analysis of collected and documented data. It also provides advice for processing and analysing data when part of a collaborative research group.

Data Processing is different for every project. Workflows can include cleaning data, combining different pieces of collected data or converting the format and files of data.

Processing and analysis workflows both centre on the quality and transparency of data, making sure that it meets the **FAIR principle criteria.**

Data Analysis is the interpretation and interrogation of data which create findings that underpins the research output.

Findable - Data should be easy to find. It should be described well and be machine readable.

Accessible – Data should be accessible in different formats and should be clear to understand.

Interoperable – Data should have standardised vocabularies so it can work across different systems and processes.

Reusable – Data should be well described so others can replicate or combine it.

Est. 1841 YORK ST JOHN UNIVERSITY Library & Learning Services

For collaborative research projects there are additional considerations when it comes to the data collection, data processing and data analysis stages.

- **Ensure the research roles and responsibilities of the group are clearly defined.**

- **Ensure that everyone has access to secure, data storage spaces, and the same data tools for processing and analysis.**

- **Different people within a group may interrogate data differently.** Make sure everyone uses the same file management system and naming conventions for data.

YSJU Software for Data Processing and Analysis (available through AppsAnywhere):

- NVIVO: This software assists in organising, analysing and sharing qualitative data.

- SPSS: This software assists in the editing and analysis on quantitative data.

Further information about software can be found through YSJU's Digital Training and Support webpages.

Data Storing and Preservation

This section provides information about data preservation and storage process to ensure that data is suitable in the long term. Data Preservation is a key part of Research Integrity, as research data can be continually used for future analysis, data validation and support for research or research data outputs.

Data Preservation are the processes that ensure that research data and its metadata are suitable for future use and is not affected by any technological changes.

Not all data from a project is preserved – it depends on each research project. However, it is encouraged you keep all data until the project's completion unless stated otherwise in the DMP. Collected data could also be used for future research projects so it's important data is stored and preserved effectively.

Digital Preservation is the activity in accessing data in the long term. It is futureproof over physical storage tools including software and hardware.

Resources for digital preservation:

- **DigitalPreservationCoalition** – an organisation that advises best practice in creating and preserving digital objects.

- **DATACC** - This platform supports RDM for the physics and chemistry disciplines and has a section about digital preservation.

YORK ST JOHN UNIVERSITY
Est. 1841
Library & Learning Services

Tips for successful Data Preservation

- **Preservation planning**. Consider and investigate ways to preserve your data while making sure it is constantly findable and accessible. This should be outlined in the DMP.

- **Keep in mind technological and legal considerations of storage spaces and platforms**. Is the space well known and secure? Is there anything to be aware of in the terms and conditions if you are storing data externally? The University recommends using OneDrive for storing collected data and the institutional repository, **RaYDaR** for completed, published datasets.

- **Make sure published research data receives a digital object identifier (DOI)**. This allows the resource to always be findable if its digital location changes. RaYDaR assigns every published dataset a DOI automatically.

Est. 1841 YORK ST JOHN UNIVERSITY Library & Learning Services

This section provides key definitions, guides and tools in how to share and showcase research data.

Key terms in Research Data Sharing:

- **Open Data:** Research Data that can be accessed, distributed and used freely (subject to open licence). Open Data must also align with the **FAIR data principles** (data should be *findable, accessible, interoperable and reusable*)

- **Creative Commons Licences:** A type of Open Licence that can be used for open research data. This allows the creator of a data set to decide on the terms their data set can be used.

- **Data Repositories:** A digital space to deposit datasets along with its metadata. YSJU's data repository, **RaYDaR**, is an open access repository which means published data outputs can be accessed publicly and freely.

- **ORCID iD:** An author's persistent digital identifier which is attached to published datasets. This links to an author's academic profile where all their research outputs are listed. ORCID iDs are free but require registration.
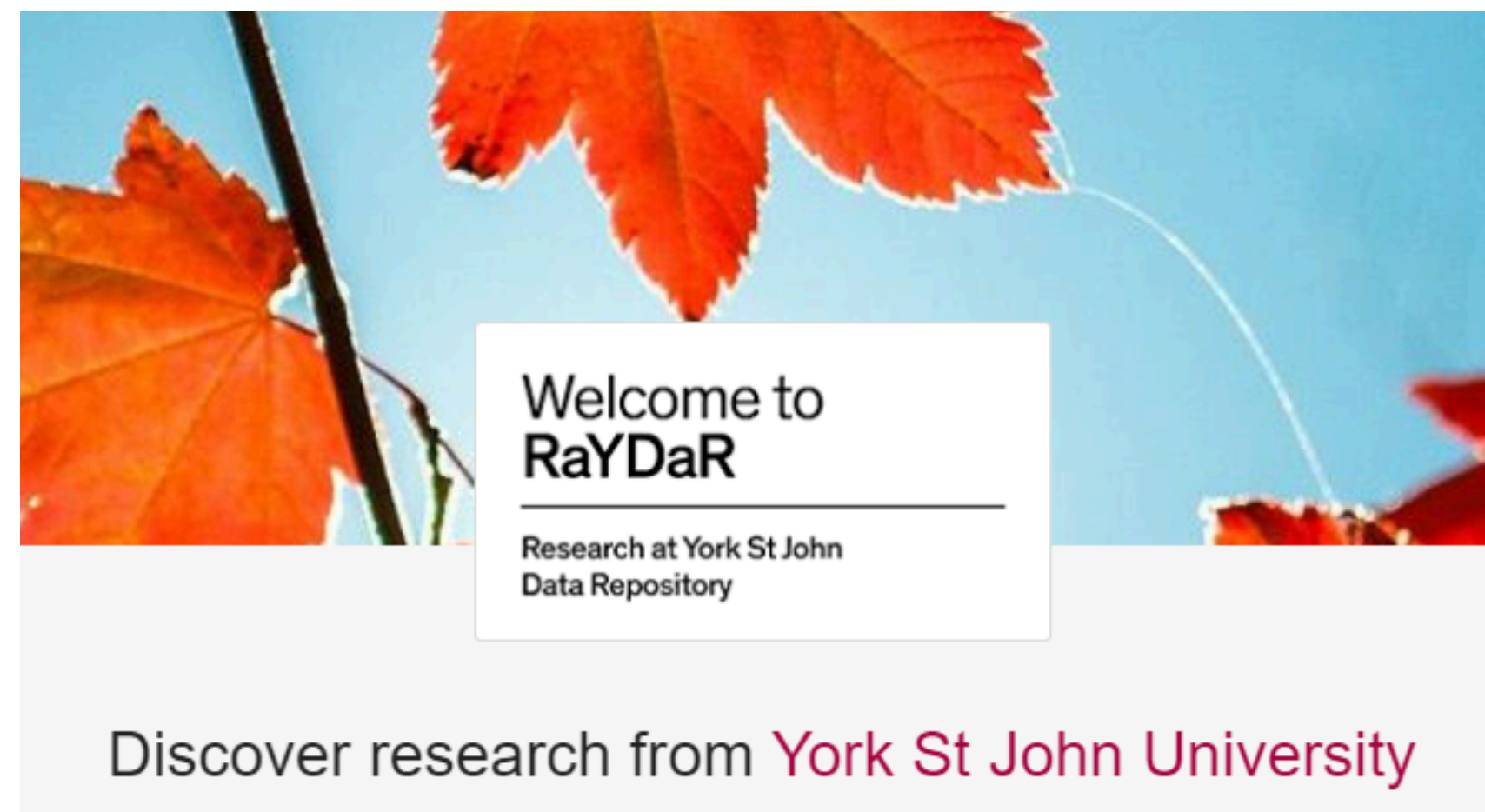
Research Data Toolkit

Data Sharing and Publishing RaYDaR

Research data can be shared during a project or exclusively at the end. While research data is normally attached to a published research output, research datasets can also be treated as a standalone output. This allows that dataset to be used and built on in future research outputs.

Welcome to
RaYDaR
Research at York St John
Data Repository

Discover research from York St John University

YSJU Data Management Policy

As part of YSJU's data management policy, researchers are required to deposit research data that supports research outputs into the University repository, **RaYDaR** unless specified otherwise in the data management plan. This applies to students whose research data is included in published research outputs, however YSJU expect all staff and students to uphold principles of data management.

This section provides advice in how to ensure your research data is accessible and reusable, and, and best practice in how to use the datasets of others.

When undergoing research, you may find data in repositories or within other research outputs. When data is used in other research outputs it's important you check the citation and source details to ensure the data can be used, and to what extent.

We recommend using <u>RaYDaR</u> for searching datasets or to see how a dataset has been formatted and licenced. As RaYDaR is a public repository you don't need to log in to search.

As best practice when using research datasets from others it is always encouraged to:

- Always cite data sources
- If you generate new data from an original work, use the same open licence as the original as well as attributing the original author.

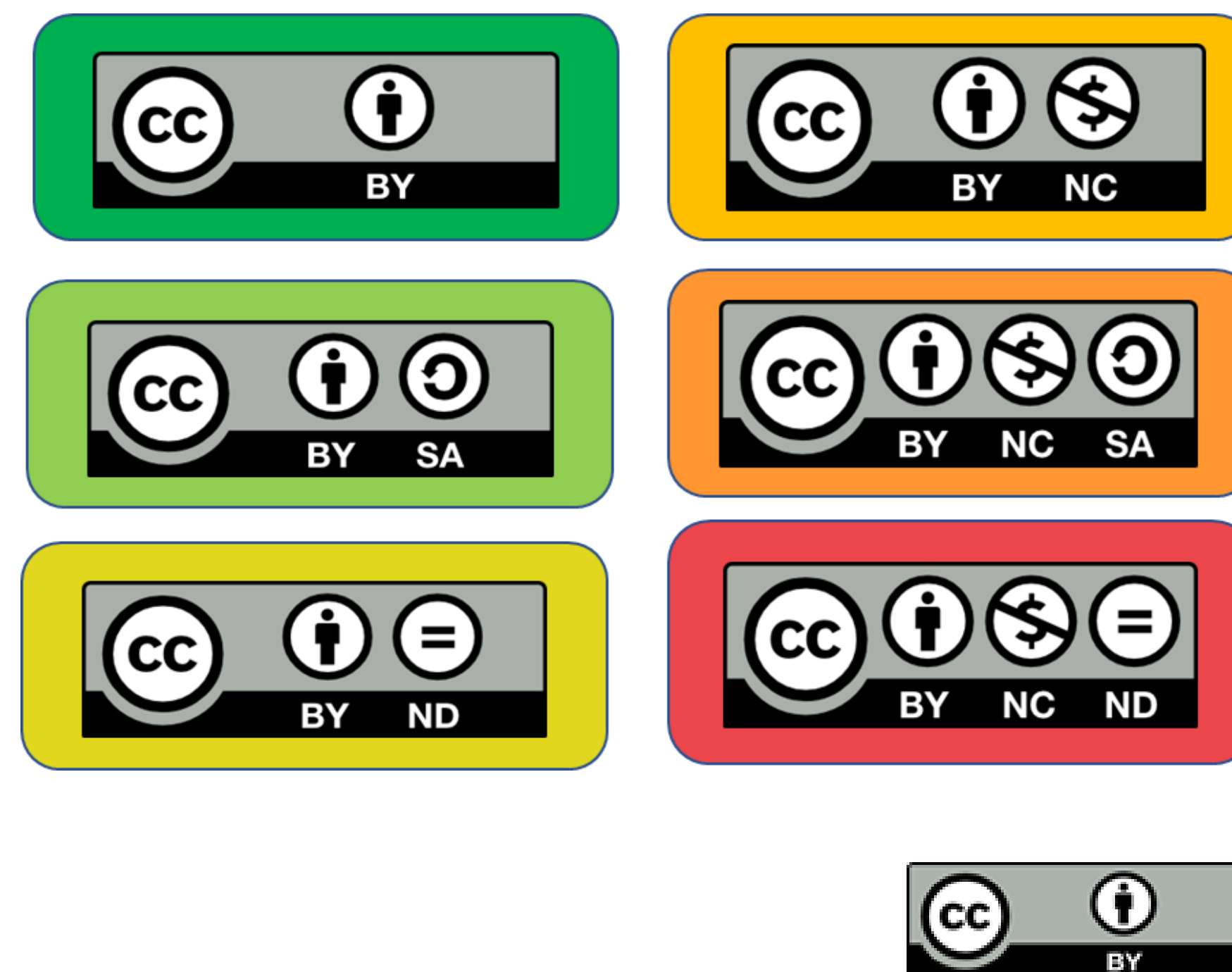Data in repositories will likely have an **Open Licence** attached to it – such as a **Creative Commons** licence. This will outline how the data can be used – and ensure that the original author of the data is credited.

If licence details are not clear on datasets, or you want to use the data in a different way to the licence terms you will have to contact the author of the dataset to seek additional permissions.

Find out more about Creative Commons and the six licence types here:

https://creativecommons.org/

YORK ST JOHN UNIVERSITY Est. 1841
Library & Learning Services