Access Controls and Licensing of Data

Dr Scott Summers
UK Data Service
University of Essex

Managing and Sharing Research Data: Best Practice for Data Protection – Lift, London
28th and 29th November 2018

UK Data Service

Overview

Licensing and access controls can help share sensitive and confidential data in a safeguarded way

Areas to be covered

- Licensing models
- Licensing considerations
- Access controls models
- Access controls in practice
- Access and licence control strategies

UK Data Service

Licensing

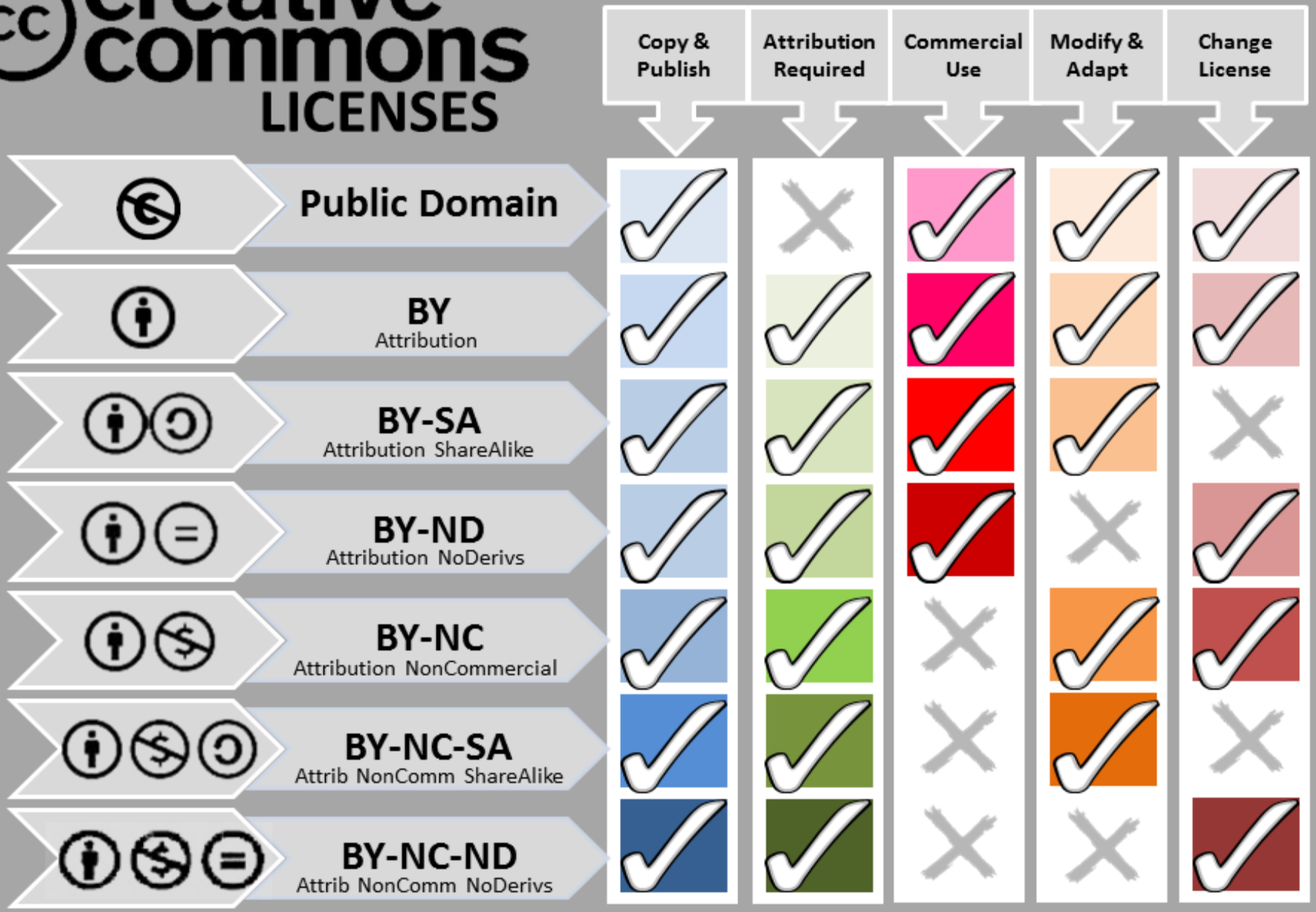
Publishing data – licensing

- If you publish data with a data repository of your choice, a licence agreement will be applied to your data
- A licence agreement is a legal arrangement between the creator/depositor of the dataset and the data repository, signifying what a **user** is **allowed to do** with the **data**
- To make reuse as likely as possible one should use a licence that:
  - makes data available to the widest audience possible; and
  - makes the widest range of uses possible

Creative Commons Licenses

- Creative Commons forge a balance inside the traditional “all rights reserved” setting that copyright law creates
- There are 6 licenses
- Specifying different requirements
- All require attribution

creative commons LICENSES



Creative Commons Licenses – Notes

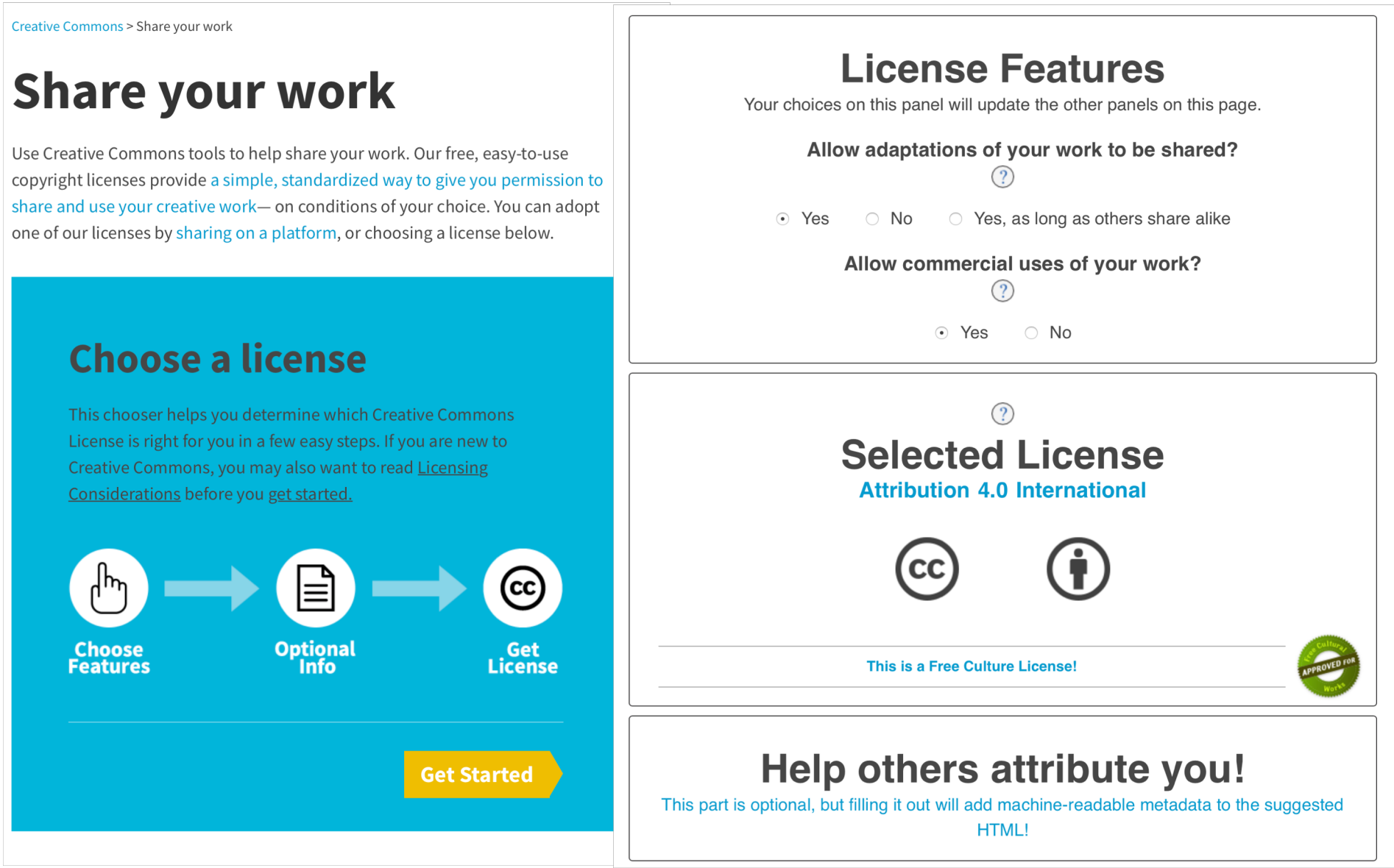
CC0

- CC0 is a completely open CC licence
- The copyright owner waives all its rights, including the database right and the right to be identified as the creator

Notes

- A CC licence cannot be revoked once it has been issued
- The licence choice may be limited or determined by the data repository of your choice

In Practice: CC tool to help choose a licence



Licensing considerations

- Rights and ownership, establish who owns what
  - Questions to ask:
    - Were secondary sources used? Are the necessary permissions in place to republish?
    - Were any of the data purchased? What agreement was in place for the future archiving of the data?
    - Any additional copyright considerations?
    - Can you sign a licence on behalf of rights owners?

- Sharing more disclosive data may require a data sharing agreement and access procedures

- Depositor licence agreement
  - Responsibilities and liabilities
  - Copyright

Access Controls

Publishing data – access controls

- Publishing data in a data repository does not automatically make them openly accessible
- Personal data can still be protected by limiting access to the data
- Access controls can permit control down to an individual file level, meaning that mixed levels of access control can be applied to a data collection

Access conditions

Most data repositories operate a three-tiered approach to data access:

1. Open access
2. Access for registered users (safeguarded)
3. Restricted access

* Embargo

In practice: managing access to data at the UK Data Archive

Open

- available for download / online access under open license without any registration

Safeguarded

- available for download / online access to logged-in users who have registered and agreed to an End User License (*e.g. not identify any potentially identifiable individuals*)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled

- available for remote or safe room access to authorised and authenticated users whose research proposal has been vetted and who have received training

In practice: data with access conditions

Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

- Interviews (audio and transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users (Safeguarded)
- 3 interviews and 5 diaries were embargoed until 2015 (Safeguarded – Embargoed)
- Audio files archived and only available by permission from researchers (Safeguarded – Special Agreement)

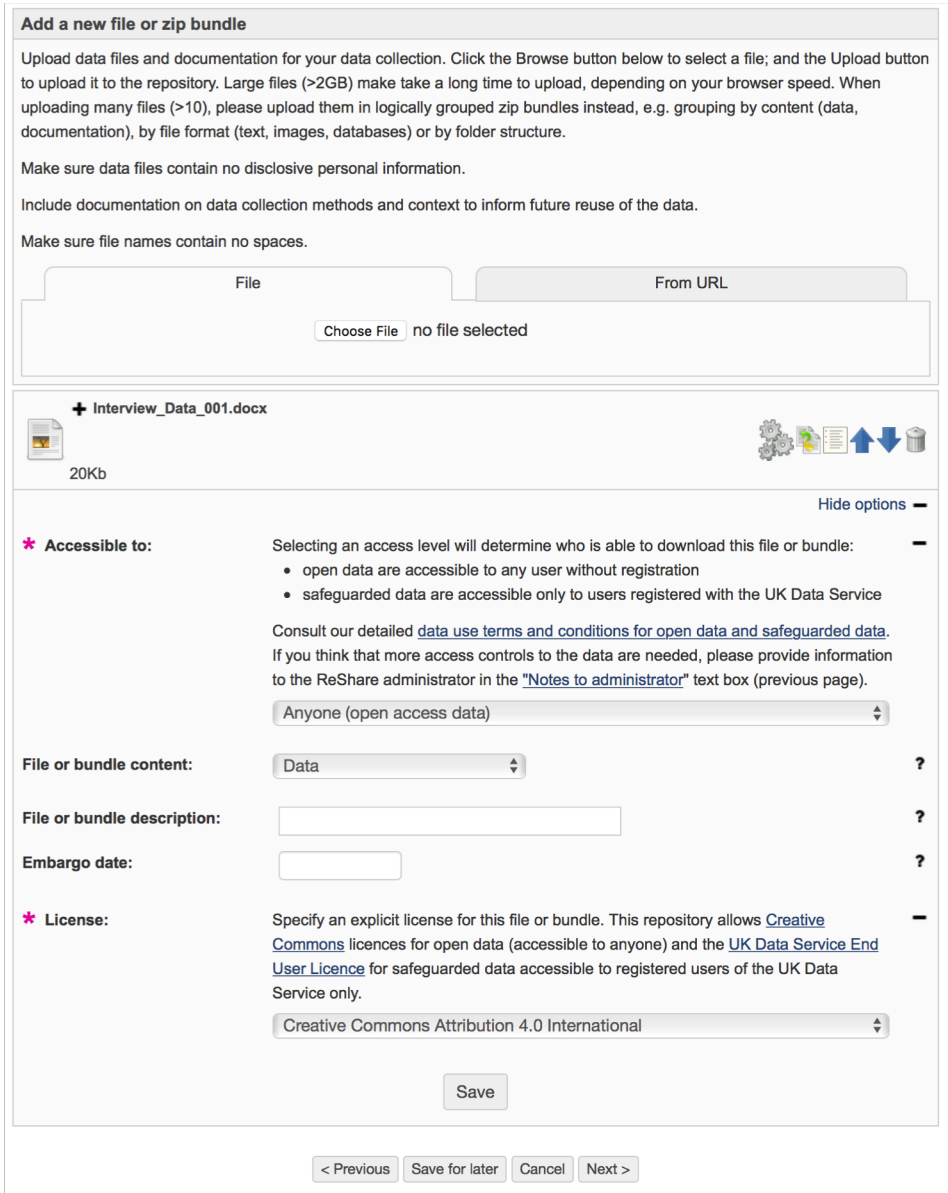
discover.ukdataservice.ac.uk/catalogue/?sn=5407

doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf

In practice: access & licensing ReShare

Use terms and conditions for open access data

- Data files deposited as Open Data are licensed under the Depositors choice of one of two Creative Commons Attribution 4.0 licenses:
- **CC-BY-NC-SA** Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International
- **CC-BY-SA** Creative Commons Attribution-ShareAlike 4.0 International (this licence allows for commercial use)
- ReShare emphasises that Creative Commons licenses can only be agreed to by the copyright holder(s) and should not be used if there are third party rights holders.



Strategy for enabling safe access

Fives safes enables safe access to data that meet the needs of data protection. Yet fulfils the demands for open science and transparency

- Safe data - treat data to protect confidentiality
- Safe people - educate researchers to use data safely
- Safe projects - research projects for ‘public good’
- Safe settings - Secure Lab system for sensitive data
- Safe outputs - Secure Lab projects outputs screened

5 Safes Animation

UK Data Service

In practice: controlled access

- Controlled access requires:
  - Approved/Accredited Researcher application = SAFE PROJECT
  - User Agreement signed by researcher and their institution = SAFE PEOPLE
  - Face-to-face training session = SAFE PEOPLE
  - Secure access via: remote desktop or safe room = SAFE SETTING
  - Outputs checked = SAFE OUTPUTS

Access control strategy

When choosing an access category, consider the following:

- Does the data contain identifiable information?
- Can the information in this data collection be linked with anything in another data collection which might lead to participant's identities being disclosed?
- What did participants consent to?
- If 'restricted access' is to be chosen who will manage the access to this request?

Open metadata for sensitive data

- Even if data cannot be published in open access, it is always possible to publish the metadata which belongs to the dataset
- Openly publishing metadata is the only way to make such datasets discoverable
- Metadata is always freely accessible meaning that:
  - No registration is needed for searching in the metadata;
  - No registration is needed for harvesting the metadata (e.g. by search engines)
- Metadata of sensitive datasets should never contain confidential or identifying elements or characteristics, like names
- When someone finds a dataset under restricted access, they can submit an access request to the dataset holder. If this is granted, the dataset will be available to download by the user

Concluding remarks

- Sensitive and confidential data can be safeguarded by regulating or restricting access to – and use of – the data
- Access controls should always be proportionate to the kind of data and level of confidentiality involved
- When regulating access and licensing data, consider who would be able to access the data, what they are able to do with it, whether any specific use restrictions are required, and for how long the data are to be available for

Questions

Contact Details:

Scott Summers
UK Data Service
University of Essex
ukdataservice.ac.uk/help/get-in-touch