



#### Managing Data Access in CESSDA I: The Data Catalogue and Beyond

#### IASSIST Conference, 9 June 2022

Sharon Bolton, Taina Jääskeläinen, Darren Bell



cessda.eu



@CESSDA Data





Licence: CC-BY 4.0



#### **CESSDA**

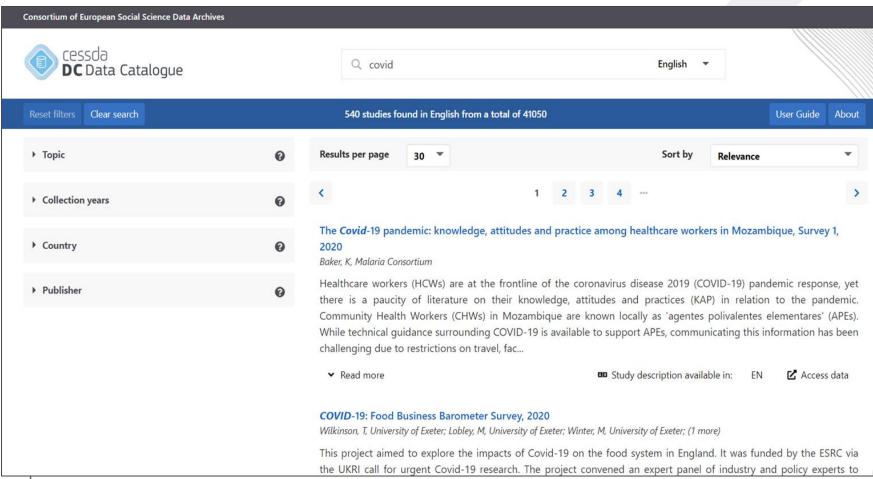
- Consortium of European Social Science Data Archives: <a href="https://www.cessda.eu">https://www.cessda.eu</a>
- Started as informal umbrella organisation for European national social science data archives in 1976
- Became an ERIC (European Research Infrastructure Consortium) in 2017
- 22 member countries across Europe (EU and non-EU) and one observer
- Aims to "provide a full scale sustainable research infrastructure enabling the research community to conduct high-quality research ... and to facilitate teaching and learning in the social sciences"
- Funds services/products and projects



# CESSDA Data Catalogue (CDC)

- Multilingual 'one-stop shop' for searching and finding data: <a href="https://datacatalogue.cessda.eu/">https://datacatalogue.cessda.eu/</a>
- Access to study-level metadata records of European social science data
- Catalogue records of CESSDA's national Service Provider (SP) organisations are harvested for the multilingual CESSDA Data Catalogue.
- Over 40,000 records
- 11 languages, covering 20 European countries
- Covers many kinds of data: quantitative, qualitative, mixed-methods, crosssectional, longitudinal, census, register, historical
- 'Access data' button takes user to study information on SP's website







# Finding data

- Basic search, advanced search, apply filters comprehensive online user guide – lots of options
- "If we build it, will they come?" What do users really want from this resource?
- User feedback: would like to narrow down their searches to find open access data they can access easily and quickly
- Open access filters have significantly increased data use in other catalogues.
- How to enable this?



# Enabling search by data access

- Include a search filter in CDC so that users can easily select access type
- Each SP has their own data access classification
- Classifications differ due to differences in national legislation and agreements with data owners
- SPs may have multiple types of data in their collection, with varying access conditions across the spectrum (completely open > highlyrestricted)
- Easiest way to manage this in CDC?



### Harmonising access types

- Scoping work undertaken by two CESSDA projects:
  - MDO: two large SPs with varied data collections and types of access (UKDS, Gesis), developed set of access types based on their access conditions.
  - Could other SPs' access conditions harmonise to these types? Multiple diverse organisations, various complications
- Later project (led by Sikt): gathered feedback, concentrated on simple 2-type classification for filter (more granularity later)
  - 'Open': no restrictions on use or user type
  - 'Restricted': any kind of condition attached to use of the data, across access spectrum from simple registration/agreement to a user licence to strict secure remote or on-site access



#### CDC: DDI Profiles

- Set of DDI profiles created to clean and harmonise CDC metadata (Metadata Office)
- Profiles help validate DDI documents that CDC receives from participating SPs (available for DDI1.2.2, DDI2.5 and DDI3.2):
  - Formal, machine-actionable XML (eXtensible Markup Language) documents that specify additional constraints on the content of a DDI XML document, over and above those specified by the document's associated XSD schema
  - Contain rules that metadata documents need to follow; human-readable information, e.g. whether use of an element is mandatory, recommended, or optional and whether uses CV (controlled vocabularies)
- CDC uses DDI CVs, multilingual versions in CESSDA Vocabulary Service (CVS): <a href="https://vocabularies.cessda.eu/#!discover">https://vocabularies.cessda.eu/#!discover</a>



#### CDC: DDI Profiles

- Profiles support machine-actionable validation by the CESSDA Metadata Validator (CMV) <a href="https://cmv.cessda.eu">https://cmv.cessda.eu</a>
- Validation with the CDC can be done by providing individual records via URL or direct upload
- Profiles ensure consistency in CDC metadata records and provide cleaned, categorised information for search filters
- Lots of work done with SPs to resolve metadata issues



#### DDI elements and further work

- No vocabulary element exists for access in DDI 2.5, suggested for DDI 2.6 (awaiting publication)
- Lack of DDI element meant delay but provided extra time to consider implications and work with SPs
- Developed a common access vocabulary for profile SPs can map their own categories (currently Open/Restricted, more granularity later)
- Plan to add filter as soon as possible; will enhance user experience and make CDC an essential research tool for finding open data (EU Open Science Policy) and meeting FAIR recommendations (R1.1 and F2)



### Further plans for CDC

- In European Open Science Cloud (EOSC) Marketplace https://marketplace.eosc-portal.eu/services/cessda-data-catalogue
- Constant improvement planned service management:
  - regular technical releases
  - regular review of profiles
  - more filters
  - support for SPs to improve their metadata; CDC creates impact for them by widening potential user community
  - include catalogue metadata from outside CESSDA
- Promotion to increase use articles, conference presentations, encourage SPs to feature link to CDC on own websites and catalogues, e.g. <a href="Progedo">Progedo</a>



# Any questions?

Thank you for listening!

Further information: contact <a href="mailto:metadata-office@cessda.eu">metadata-office@cessda.eu</a>

