UK • DATA ARCHIVE

METADATA MARVELS
AT THE UK DATA ARCHIVE

CITATION

Datacite
- Digital Object Identifiers (DOI) ensure a persistent web address for digital collections
- British Library is the registration agency that provides our DOIs
- not-for-profit organisation managed by the German National Library of Science and Technology with other international members

Our DOIs
- Datacite DOIs are assigned to each data collection
- our DOIs support changes, so that if the location of the data changes, the DOI will always link to the data cited

APA
- standardised referencing format for various materials
- innovation: we have developed a paragraph-level reference list citation format:

e.g. Marsden, D., 'Interview with Clough' in Mothers Alone : Poverty and the Fatherless Family, 1955-1966, UK Data Service [distributor], February 2005. SN:5072, Paragraphs 2-4, http://dx.doi.org/10.5255/UKDA-SN-5072-1 from http://discover.ukdataservice.ac.uk/UKQualiBank/Document/?cid=q-1ed6fef1-3f9d-4284-b23d-d9f7a4266745

Data Citation Index (DCI)
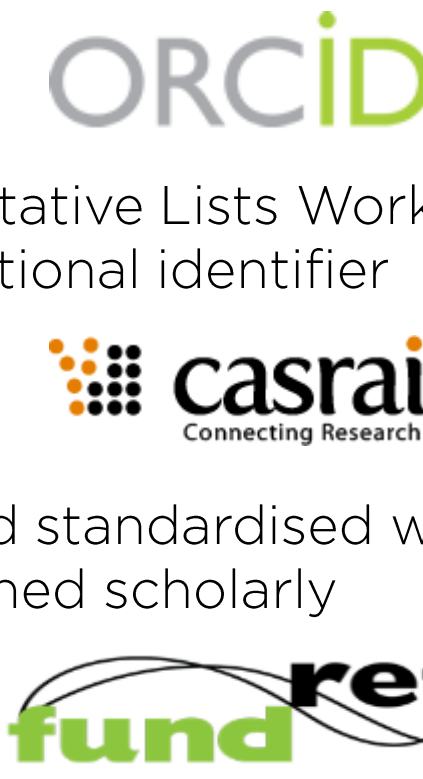- all metadata available via the DCI

Digital Object Identifiers (DOIs)

CONTROLLED VOCABULARIES

The UK Data Archive is guardian of the Humanities and Social Science Electronic Thesaurus (HASSET) and of its multi-lingual sister, the European Language Social Science Thesaurus (ELSST)

- ELSST has been translated into nine languages, with three more on the way
- HASSET has been in use and developed by the UK Data Archive over more than 30 years:
  - used for indexing data, at study level and variable level
  - allows retrieval of data and related documentation using hierarchies of keywords
  - recently converted to SKOS, using the Pubby tool
- between them HASSET and ELSST represent 7,695 concepts and 4,032 synonyms
- in SKOS, that's 101,808 relationships as triples

HASSET thesaurus

ELSST thesaurus

ENTITY IDENTITY

- We support ORCID in our self-deposit repository

- We are part of Jisc-CASRAI Authoritative Lists Working Group for a recommended organisational identifier standard

- We will support FundRef, the agreed standardised way to report funding sources for published scholarly research

ReShare repository

QUDEX

Schema that enables discovery, locating, retrieving and citing complex qualitative data collections in context:
- highly structured and consistently marked-up data
- rich descriptive metadata for files e.g. interview characteristics, interview setting, type of object
- logical links between data objects: text to related audio, images, and other research outputs
- preserves references to annotations performed on data
- common metadata elements that enable federated catalogues across providers and borders
- maintained by the UK Data Archive, University of Essex
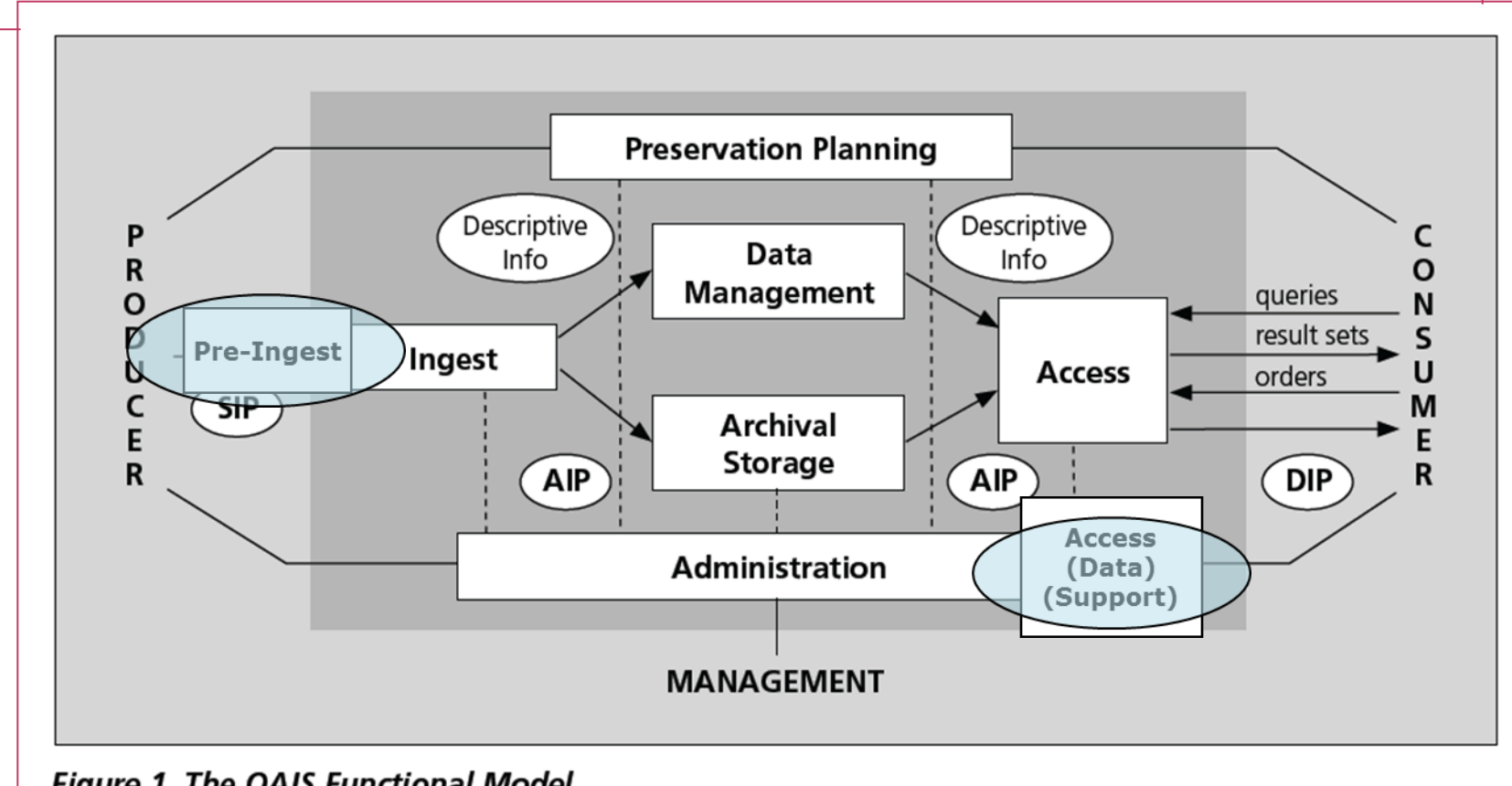
UK Quali Bank

UK DATA ARCHIVE

With over 40 years of use and promotion of metadata standards for collection description and data description in the social science domain, the UK Data Archive:

- is an active promoter and guardian of standards and is engaged in:
  - Discovery open metadata principles
  - Jisc-CASRAI
  - DDI and DDI Controlled Vocabularies Group
  - DDI Qualitative metadata working group

- undertakes R&D metadata projects:
  - ESRC Digital Futures, UK Quali Bank
  - Jisc Research Data Registry
  - Wellcome Data Discoverability
  - Jisc SKOS-HASSET
  - Jisc U.Geo

PRESERVATION

- We support the Open Archival Information System (or OAIS) for organisational structure and systems for archives with responsibility for preserving information and make it available to a Designated Community.

- We will support Premis to provide granular structured machine-actionable preservation event statements.

Figure 1. The OAIS Functional Model.

DDI CODEBOOK 2.5

The Data Documentation Initiative (DDI) is a specification for capturing metadata about social science data:

- originally created to capture the information found in survey codebooks, which remains the focus of earlier versions
- used for:
  - basic study-level catalogue metadata
  - rich variable description for survey files
- maintained by the Data Documentation Initiative Alliance, a membership-driven consortium including universities, data archives, and international organisations
- made available via OAI-PMH and as DDI XML

Study level: Discover portal

Data level: Nesstar online browsing

TEI

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form.

Many profiles but UK Data Service uses it for:
- structural mark-up of textual qualitative data
- three mandatory TEI header elements
- body elements: turn takers, paragraphs, headers
- inline tags: corrections, errors

UK Quali Bank

SDMX

The Statistical Data and Metadata Exchange (SDMX) technical specifications comes out of the world of official statistics and:
- aims to foster harmonisation and standards for the exchange of statistical information
- has cooperating international organisations include the IMF, Eurostat, World Bank and OECD

UKDS.stat

www.data-archive.ac.uk