

The Application of ISAD(G) to the Description of Archival Datasets

Elizabeth Shepherd

University College London



Research project objectives

- evaluate ISAD(G) multi-level rule for the description of datasets
- evaluate ISAD(G) elements in the description of datasets
- identify omissions in ISAD(G) elements
- compare the use of ISAD(G) for listing datasets with other approaches

Project results

- commentary and practical guidelines on the application of ISAD(G) for the description of datasets
- new ISAD(G) element set for description of datasets

Original research published:

- Elizabeth Shepherd and Charlotte Smith
‘The application of ISAD(G) to the
description of archival datasets’ *Journal of
the Society of Archivists* 21: 1 (April 2000):
55-86

Multi-level description rule as applied to datasets

- general to the specific
- from the *fonds* to the *item*
- allows for macro level descriptions (creator's name and biographical or administrative history) in linked authority files
- multi-provenance records problem

Macro and micro level descriptions for datasets

- ISAD(G)'s higher levels of description (fonds, sub-fonds) derive from administrative function of the database
- lower levels (file, item) derive from internal structure (relationships that exist between the tables and fields of the dataset)
- middle level (series) may not exist organically

Example higher level descriptions

- Fonds - Department of Health
- Sub-fonds - Anatomy Office
- [NDAD - Department of Health: Anatomy Office: Anatomy Dataset]
- Herriot, Peter - Birkbeck College, Department of Occupational Psychology - Principal Investigator
- Rothwell, Charles - Birkbeck College. Department of Occupational Psychology - Principal Investigator and Data Collector
- Economic and Social Research Council - Sponsor
- [Essex Data Archive - SN2058 Effects of Role Expectations on the Graduate Selection Interview, 1980-1982]

Series definition

- *Datasets* arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation, or the same activity;..... A series is also known as a record series *and when applied to datasets is taken as being related annual accruals of a dataset, or regular snapshots of an accruing system etc.*

Static and active datasets

- ‘static’ datasets (ie contain data from a finite activity which will not be amended)
- identify organic series
- ‘active’ database files (ie continuously evolving and data amended or added)
- create ‘artificial’ series

Four types of dataset (1)

- One-off project related datasets
- data collected over specific time period, then closed
- treat as a series of a single dataset
- **Example:**
 - Series Level - Taxation Database, 1291-1292
 - File Level - 1 Dataset: Taxation Database, 1291-1292
 - [History Data Service - SN3897: Taxation Database, 1291-1992]

Four types of dataset (2)

- Active or static database
- files relate to specific data collection period
- no further data added to closed files
- **Example:**
 - Series Level - Attitudes of Students at the London School of Economics
 - File Level - 3 datasets: 1987-1989 survey, 1989-1990 survey, 1991-1992 survey etc.
 - [Essex Data Archive - SN 33131 Attitudes of Students at the London School of Economics]

Four types of dataset (3)

- active database, not closed regularly
- old data updated and overwritten
- data amendments regular or continuous
- create ‘artificial’ archive files by snapshots and/or log files
- **Example:**
- banking system for personal accounts where data overwritten with new transactions

Four types of dataset (4)

- Active database but new data sits alongside old data, does not overwrite
- no pre-determined period of data collection
- create ‘artificial’ archive files by snapshots of entire database at intervals

Combination datasets

- ‘Snapshots of DOMUS are expected to be transferred to NDAD annually. Each snapshot will consist of data held in the DOMUS system at the time of transfer except for data in the ANNUAL table relating to the most recent DOMUS survey, which will be transferred in the following year. Data relating to one-off special surveys will also be transferred after one year. Collectively the snapshots will allow the comparison of data from those files of DOMUS where data is not preserved from year to year, i.e. where data is overwritten as new data is received.’
- [NDAD - Museums and Galleries Commission: DOMUS, Series Catalogue]

Item level descriptions of datasets

- Fonds level - Department of Health
- Sub-fonds level - Anatomy Office
- Series level - Anatomy Office Dataset
- File level - Snapshot of a dataset, October 1998
- Item level - Table 1: Bodies
- Item level - **Table 2: Schools**

Item level description

- | ■ <i>Field Name</i> | <i>Description</i> |
|---|--|
| ■ School | Medical School [PRIMARY KEY] |
| ■ Address | Address of school |
| ■ Licencee0 | Person Authorised to sign for disposal of body |
| ■ Licencee1 | 2nd Person Authorised to sign for disposal of body |
| ■ [NDAD - Health Department: Anatomy Database: Snapshot, October 1998: Table 2 Schools] | |

3.1 Identity statement area

- 3.1.1 Reference code(s)
- 3.1.2 Title
- 3.1.3 Date(s)
 - *3.1.3.1 Date(s) of creation of the structure of u/d*
 - *3.1.3.2 Date(s) of contents of the unit of description*
 - *3.1.3.3 Date of last input*
 - *3.1.3.4 Date of last access*
- 3.1.4 Level of description
- 3.1.5 Extent of the unit of description (quantity, bulk, or size)

3.2 Context Area

- 3.2.1 Administrative *Context*/ biographical history
- 3.2.2 *Aim and purpose*
- 3.2.3 *Statement of responsibility*
- 3.2.4 Archival history
- 3.2.5 Immediate source of acquisition or transfer

3.3 *Scope And Content Area*

- 3.3.1 Scope and content
- 3.3.2 Appraisal, destruction and scheduling information
- 3.3.3 Accruals

3.4 Structure Area

- *3.4.1 Logical structure and schema*
- *3.4.2 The active/working nature of the database*
- *3.4.3 Data Capture and Validation before Transfer to the Archive*
- *3.4.4 Constraints on the Data Reliability*

3.5 Conditions Of Access And Use Area

- 3.5.1 Conditions governing access
- 3.5.2 Conditions governing reproduction
- 3.5.3 *Data Protection requirements*
- 3.5.4 Language/scripts of material
- 3.4.5 Finding aids

3.6 Allied Materials Area

- 3.6.1 Related units of description
- 3.6.2 Publication note
- 3.6.3 *Publications produced by originating body*

3.7 Original System Attributes

- *3.7.1 Hardware*
- *3.7.2 Operating system*
- *3.7.3 Application software*
- *3.7.4 User interface*

3.8 Archive Management Area

- *3.8.1 Digital processing and arrangement*
- *3.8.2 Content validation*
- *3.8.3 Transformation validation*

Note areas

- 3.9 Note Area
 - 3.9.1 Note
- 3.10 Description Control Area
 - 3.10.1 Archivist's Note
 - 3.10.2 Rules or Conventions
 - 3.10.3 Date(s) of Descriptions

The Application of ISAD(G) to the Description of Archival Datasets

Conclusions

e.shepherd@ucl.ac.uk

