Contribution ID: **246**   Type: **Poster**

Baseline protocols for archiving

Monday 13 October 2025 18:00 (1h 30m)

Many current solutions for data management are expensive or require considerable technical underpinnings (or both). The global data community needs to consider simpler approaches in order to include more participants and to improve equity, but this requires guidance about minimal requirements. The Protocols for the Implementation of Archival Repository Services are an attempt to start the process of establishing a baseline for reliable archiving which can guide the development of tools and services which will then be available to a broad range of organisations.

Working with researchers and community groups who have data of interest to our project, the Language Data Commons of Australia (LDaCA), we encounter a variety of scenarios. Some data has made its way into Archival Repositories, places of safe keeping, such as the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), which takes a principled approach to managing resources for the long term, including having simple access controls in place to make sure that access to some materials can be restricted.

However there are two categories of data which are very common where there is no effective governance or technology in place to ensure data is well cared for:

Firstly, a large amount of data is sitting in researchers'offices, garages, or community centres on analogue media and/or at-risk digital media such as shelved hard drives or tapes. A large amount of contemporary research currently being conducted has ad-hoc data storage on individual computers, institutional or individually provided cloud services (Dropbox, One Drive etc). If metadata is present at all it is ad hoc and not linked to data assets.

Secondly, a great deal of data is 'locked in'to applications such as websites or proprietary content management systems, many of which use the term 'archive'in their sales materials but which do not follow archival practices.

These current 'solutions'for data management are sub-optimal in terms of sustainability as they can be expensive financially and require complex technical underpinnings (or both).

The LDaCA team collaborated with PARADISEC staff and our network of partners and stakeholders to produce a set of protocols which are aimed to ensure that data can be managed for the long term, Protocols for Implementing Long-term Archival Repositories Services (PILARS).

The high-level aims of these PILARS protocols are to:

• Maximise autonomy for data custodians/stewards

• Maximise return on investment in data and data infrastructure

• Maximise long-term sustainability for data and for data systems and management

(Source: https://w3id.org/ldac/pilars)

We will present the protocols with examples of how they have been implemented and show the extensive Open Source toolkit that has been created to implement the protocols, not just for language data but for any data that needs to be stored for the long term.

Primary author: SACAL BONEQUI, Moises (Language Data Commons of Australia)

Co-author: Dr SEFTON, Peter (Language Data Commons of Australia)

Presenter: SACAL BONEQUI, Moises (Language Data Commons of Australia)

Session Classification: Poster Session

Track Classification: SciDataCon Persistent Themes: Data, Society, Ethics, and Politics