

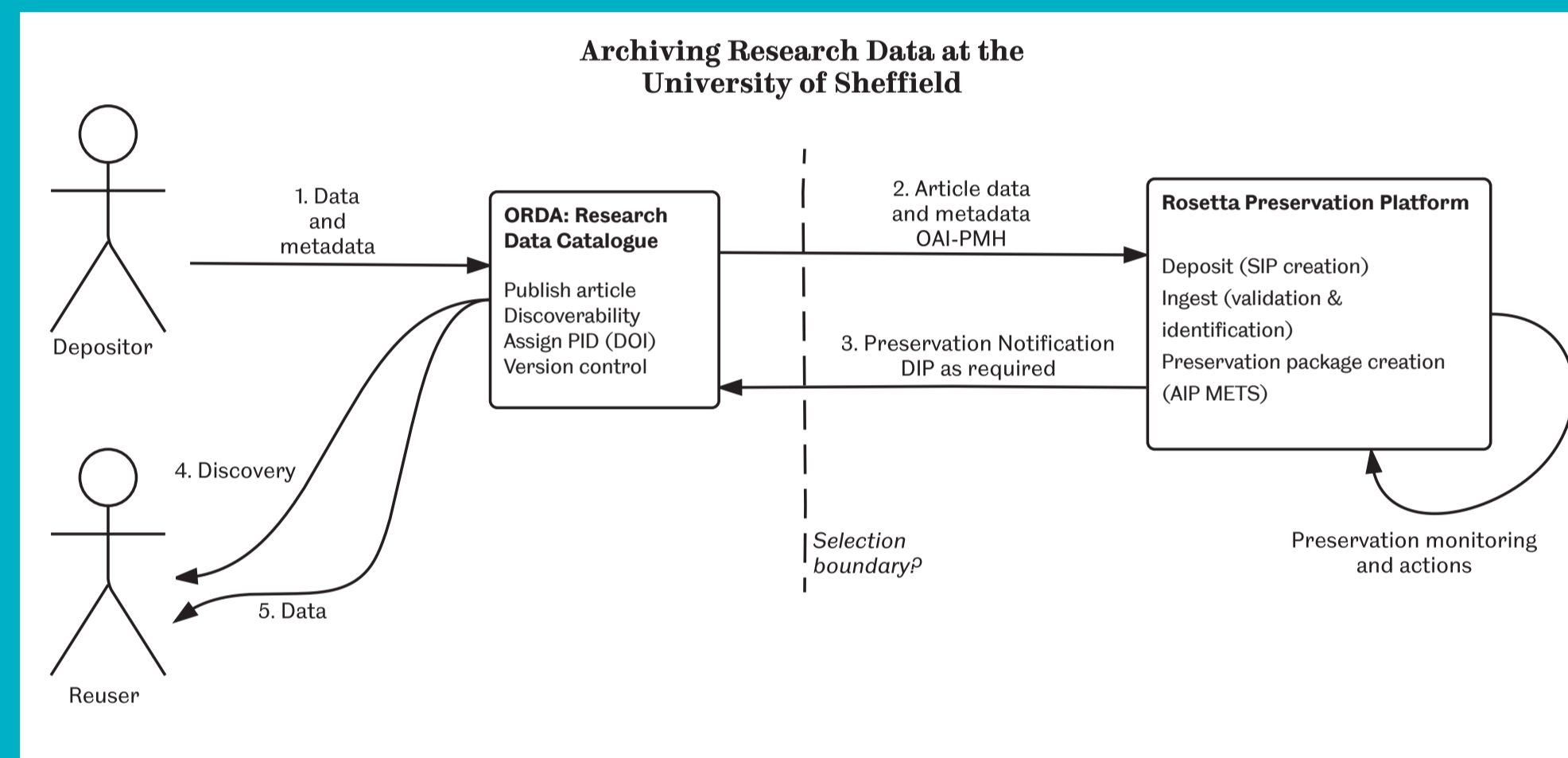
The University Library at the University of Sheffield is taking the leading role in supporting the active management and curation of research data within the institution. We have recently implemented a research data catalogue and repository, ORDA (Online Research Data, powered by figshare for institutions). We have also begun safeguarding library collections and key

administrative assets of the University using Rosetta, a digital preservation platform from Ex Libris. We are now working with figshare and Ex Libris to integrate both services to provide seamless preservation of published research data across the research lifecycle.

In the longer term, this work will enable us to provide complete lifecycle data management service for the

university's research community; identify, understand and act on risks associated with preserving data sets; better inform advice and guidance around use of data formats for sharing and preservation purposes; and encourage researchers to share their data more openly with others by guaranteeing the long term sustainability of that data.

Integration work

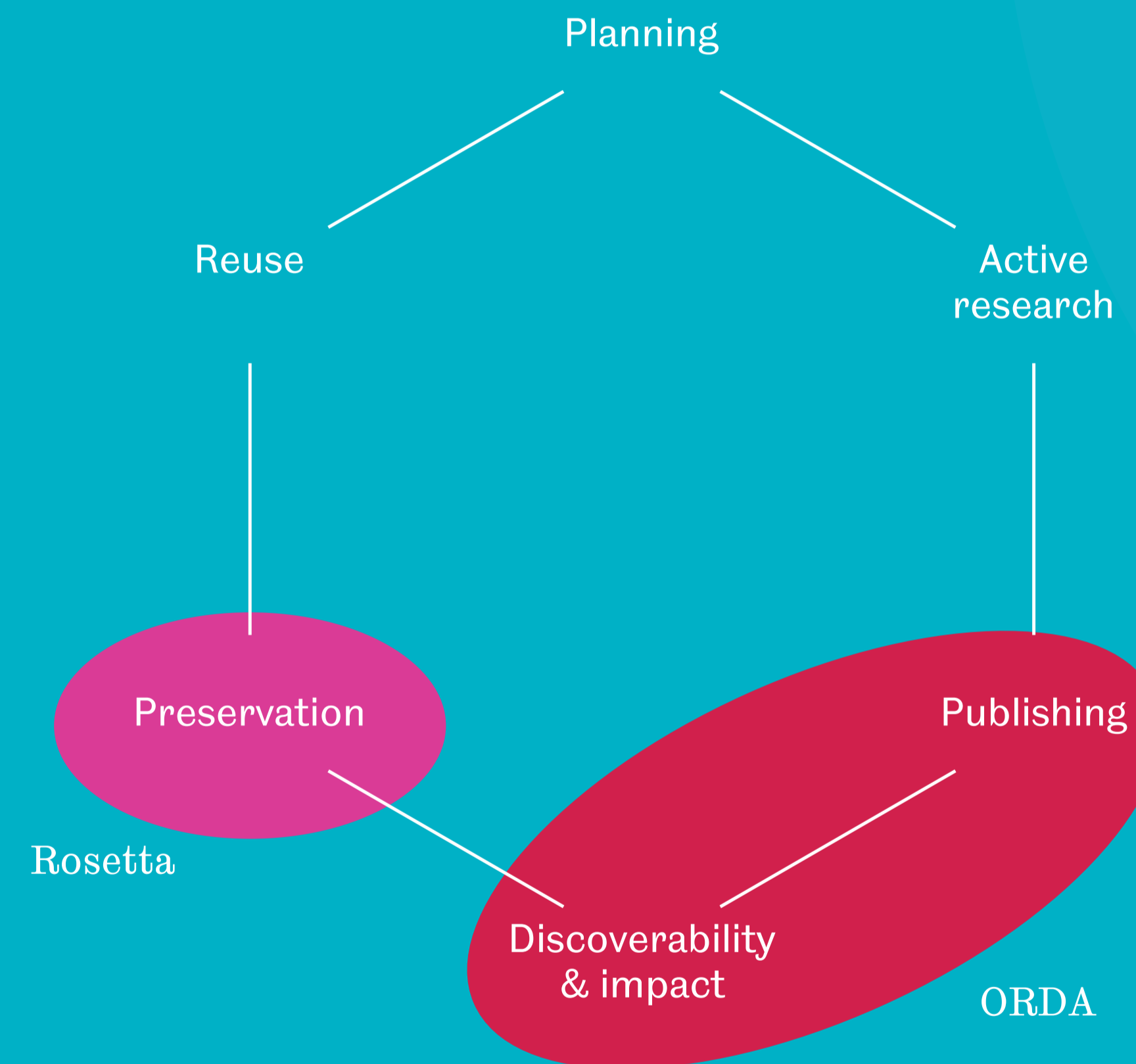


Our initial focus is on using the OAI-PMH protocol with METS metadata to allow efficient transfer of information. While figshare remains the interface for researchers and external users, Rosetta will act as a dark archive, giving us a secure preservation copy of all published data. ORDA will display a badge for each item which has been preserved in Rosetta.

Preparation included an analysis of the data held in ORDA using DROID to identify and articulate key format / software-related risks that may be associated with long term retention of data in our research collections. Initial profiling showed that DROID was only able to successfully identify 21% of the content deposited by researchers. This leaves a large proportion of files in our sample (79%) unidentified.

Further analysis will be required to identify why this is the case and how to manage it. Is it simply that the formats are not included in PRONOM? Is it due to proprietary methods of creating files (incorrect file extensions for example). We know, for example, that DROID struggles to identify software code.

Lifecycle



File Format Assessment

