



# Omega Catalogue - DORIS

## Integration Analysis and Feedback

[TNA-OCDA22/1]

Prepared for: Jone Garmendia, The National Archives

Prepared by: Adam Retter, Evolved Binary  
[adam@evolvedbinary.com](mailto:adam@evolvedbinary.com)

4 November 2022

Document version: 0.2.0



© The National Archives, 2022

This information is licensed under the Open Government Licence v3.0.

To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/>.



EVOLVED BINARY

ABSTRACT

As part of The National Archives’ Project Omega to create a modern catalogue system, the eventual decommissioning and replacement of the ILDB and miniILDB database systems underlying the current catalogue is a complex task, as there are a number of other legacy core operational systems within The National Archives that rely on access to these databases.

One such system is DORIS (Document Ordering Record Information System). Previously it was envisaged that Project Omega and the IT department would need to agree and accommodate some form of integration between Omega and the legacy DORIS system. More recently The National Archives has decided to also investigate replacing the DORIS system with a more modern option. In light of this welcome news, we examine the current state of affairs at a high-level, our plans for a Linked Data Pan-Archival Catalogue Service, and make several suggestions towards development of a modernised future DORIS system.

This document first outlines the integration challenges facing TNA when updating its legacy systems, discusses in brief the current approach of Project Omega, highlights the key points of integration between the current Catalogue Editorial System and DORIS, and then suggests opportunities for collaboration and integration between the two.

Cover photo by [C.M](#) on [Unsplash](#).

TABLE OF CONTENTS

1. Background .....3

2. Omega Catalogue Replacement.....4

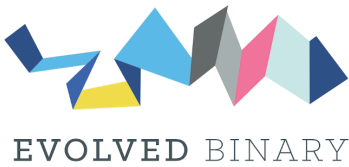
3. Current DORIS / Catalogue Integration .....5

4. DORIS and Project Omega Opportunities.....6

    4.1 High-Level Outlook.....7

        4.1.1 Outlook Rationale .....8

Glossary .....10



## 1. BACKGROUND

The fundamental consideration with replacing any of TNA's legacy core operational systems is that there is a myriad of delicate dependencies between these systems where points of integration have been setup in the past.

Almost all of these legacy systems make use of the same legacy technology stack (e.g. Microsoft Access, SQL Server, ASP 2.0/.net, SQL Server Reporting, Internet Explorer/Edge). The difficulty with replacing one or more of these legacy systems is two fold:

1. The procurement or development of a modern replacement is unlikely to yield the same technology stack. Many of the technologies used in these legacy systems are either obsolete or no longer considered good choices for modern systems. Integration between legacy and modern technologies is not without challenges, especially where legacy applications may use non-standardised or proprietary interfaces.
2. The current points of integration between the legacy systems are predominantly in the form of SQL Server Jobs that replicate data tables from one database to another. Unfortunately due to that approach, each legacy application is tightly coupled to the existing database technology, structures, and format of TNA's data; thus making incremental replacement a complex proposition.

Just focusing on the data used across the systems, replacing any one of TNA's legacy core operational systems presents two main challenges:

1. Extracting and converting the existing data (for which that system is the canonical source) into a modern format that both, (a) meets TNA's business requirements, and (b) aligns with the chosen replacement technologies.
2. Maintaining/Replacing the integration points between the replaced legacy application, and the other existing legacy applications still in service.

Whilst it is relatively common in technology refresh projects to extract and convert existing data into a more modern format, subsequent difficulties arise from this data still needing to be used by the remaining legacy systems. There are typically three approaches to resolving this:

1. Replacing further existing legacy systems, so that all systems with points of integration are modernised and can work together. Typically standardised data formats and APIs must be agreed across the organisation to make this a reality. If the extent of integration between the existing legacy systems is not well understood, this can sometimes have the effect of "Pulling on a thread" whereby for organisations that have existing complex and tightly-coupled integrations, it quickly becomes apparent that all legacy applications may need to be replaced.

2. An ETL in the opposite direction ("reverse ETL"), or legacy connectors may be developed. In this approach the modern data format is converted back into the legacy format for access by the legacy applications. This system may be part of the new application, added to each legacy application, or form a central bridge. This may work well when data synchronisation is read-only or uni-directional, but can become complex, or even impossible, for read-write or bi-directional workloads.
3. A hybrid approach utilising each of the above. Through identifying areas of overlap in data requirements between groups of legacy applications, the choice is then made as to which legacy applications should be replaced, and which will stay in service and be catered to by a reverse ETL or supported by legacy connectors.

The DORIS system as one of TNA's legacy core operational systems is no exception, and provides several challenges for modernisation.

## 2. OMEGA CATALOGUE REPLACEMENT

Project Omega is working on a replacement system for the current Catalogue. The replacement system will in the first instance replace the following systems:

- PROCat Editorial
- PROCat Viewer
- ILDB
- miniILDB
- Taxonomy
- *(possibly aspects of)* MYC Authority Files

Project Omega eventually aims to be the central canonical store for all metadata of TNA's assets. As such, the project has identified additional "catalogue" like systems across the organisation that it eventually intends to replace. The first such system identified for replacement is the Digital Catalogue from within the DRI (Digital Repository Infrastructure) project of the Digital Archiving department. In the mid-term, the ePro (Electronic Public Records Office) and SAR (System for Access Regulation) systems have also been identified for replacement by Project Omega. Replacement of other identified systems are still pending further discussion.

In short, due to both: (a) data management issues with previously hierarchical and tabular approaches to records information management at TNA, and (b) international advances in archival data management practice (e.g. Matterhorn, ICA RiC, etc.), Project Omega has adopted, and is promoting, a Graph data model for storing



---

## PROJECT OMEGA

records and asset metadata. This is a radical departure from the existing tabular approach used with SQL Server by most existing legacy core operational systems at TNA. Such a change presents many challenges for integration with existing systems, but offers opportunities for: reducing effort and duplication of data storage, reduce operational costs, and unlocking the potential of a vast array of existing rich data.

Project Omega is being developed as a layered architecture, whereby the required business services are to be developed atop a Graph data store and made available to any department within the organisation (assuming suitable authorisation). These business services will be developed in a technology agnostic manner, utilising agreed international standards so that they may be consumed by any other system of technologies. Such services will be accessed asynchronously via a messaging fabric to decouple individual systems. Individual TNA departments may then use, or augment, these business services to develop their own applications that can reuse and enhance the metadata stored at the heart of the organisation. This approach architecturally is not dissimilar to that of an ESB (Enterprise Service Bus) for the purposes of EAI (Enterprise Application Integration). We believe that such an approach can provide The National Archives with an Enterprise Architecture that could, if desirable, work towards eliminating many of the issues it experiences today due to its proliferation of disjoint systems and data silos.

In addition, by utilising these business services, Project Omega will build and deliver a dedicated Catalogue Editorial application to enable the existing Catalogue users to perform the day-to-day business of Catalogue Management.

Until this time, integration with other existing legacy systems, such as DORIS, has only been considered at a high-level.

## 3. CURRENT DORIS / CATALOGUE INTEGRATION

DORIS requires access to a number of databases, some of these are owned by the DORIS system, whilst others are solely for the purpose of integration with other legacy core operational systems at TNA.

The Catalogue (i.e. the systems maintained and operated by/for The Cataloguing, Taxonomy, and Data Team), which include PROCat Editorial, PROCat Viewer, ILDB, miniILDB, and Taxonomy, sit at the heart of TNA's records management strategy.

DORIS has existing integration points that require access to the following SQL Server databases owned by systems within The Catalogue:



1. ILDB (and ILDB Audit?)
  1. There is a daily SQL Server job that replicates data from the ILDB SQL Sever database into the DORIS SQL Server database.
2. miniILDB
  1. There is a daily SQL Server job that replicates subset of the data in ILDB itself into a miniILDB database.
  2. This miniILDB database is used by the larger DORIS system.
  3. There is a component of the DORIS system called the "*KBS App*" that also requires access to the miniILDB database.

In addition it is worth noting that stored within ILDB (and presumably miniILDB by extension?), is an "*Orderability*" flag, which we understand to be the canonical flag for controlling whether a Record is orderable by users interactive with the DORIS system. The data ownership of this flag is the responsibility of The Cataloguing, Taxonomy, and Data Team.

## 4. DORIS AND PROJECT OMEGA OPPORTUNITIES

There is a great deal of cross-over in both the data and functionality requested by users of both DORIS and The Catalogue. As TNA is seeking to replace both of these legacy operational systems within an overlapping time-frame, it would be financially prudent to identify commonalities and overlap between the two systems and their requirements. Where possible, common components need only be built or procured once, and canonical data need only be stored once without duplication and/or transformation. This has the advantage of reducing both costs of development and procurement, and subsequent operational costs of infrastructure after deployment.

Regardless, we must in the least find an agreeable strategy for integration between the two new systems. Project Omega will be decommissioning ILDB and miniILDB, which are two systems that the existing DORIS system is dependent on in favour of a new graph database; presumably the replacement DORIS system will still require access to Catalogue data to operate. The degree of integration is as yet unclear, however, as good practice for design of modern Enterprise Systems demands decoupling data and services, it would seem sensible to work up from the data level initially, and then secondly, identify business services (operations) that need to be provided atop this data.



## 4.1 High-Level Outlook

There is an obvious need for DORIS to be able to work with data from The Catalogue. Conversely, there is metadata which is created and managed by DORIS which, from an Enterprise Architecture perspective, would be highly desirable to have within a centralised datastore at TNA.

One example of such metadata, is that of the physical location of physical records. Project Omega already has the concept of Realisations/Manifestations of records. These Realisations are updated as the tangible asset changes over time. The provenance of these changes is recorded to form a strong auditable history of change. The location (URL) of Digital Records will be stored as a property of their Realisations. Therefore it would seem an odd asymmetry to not also store the location of physical records as a property of their Realisations, and as such we would welcome such metadata in the replacement Catalogue System.

We suggest that there would be value in undertaking an exercise to analyse and group the data used by DORIS into two categories:

1. Pure Records/Assets Metadata
2. Application State and Process information.

From the results of that analysis, it could be advantageous for the pure Records/Assets metadata to be integrated into the Project Omega Data Model of the replacement catalogue system. This data would then be considered canonical and reside in the centralised datastore of the new Project Omega Catalogue System. As this data may then subsequently be accessed by the replacement DORIS system, or any other TNA system, via the planned Business Services APIs, the organisation as a whole can benefit through utilising and enhancing this data. Governance and Ownership of such data would still need to be decided, but would not necessarily have to reside with the existing Cataloguing, Taxonomy, and Data Team.

Like DORIS, Project Omega also has a need to store Application State and Process information for its own applications (e.g. The Catalogue Editorial application). It is currently planned that this data will be held in a number of separated graphs, within the same graph datastore. There could be cost and resource savings that can be identified, should DORIS also consider storing its own Application State and Process information into the same graph datastore as Project Omega (albeit this would likely be in separated graphs designated solely for DORIS applications).

In short, Project Omega has already built several software components and infrastructure that could be re-used, and/or extended for a replacement DORIS system at a lower cost than building, or possibly purchasing, and integrating a new system. The option of a centralised and linked data store, providing technology agnostic business services to all departments of the organisation offers many tantalising options for gaining new value from past data investments, and developing new applications in future.

Finally, we recognise that Project Omega has limited resources currently available to it. At this time those resources available to Project Omega are focused on replacing PROCat and ILDB. However, due to the software architecture approach taken in Project Omega, where data, business services, and consumers are highly decoupled, individual services can potentially be rapidly developed not only by Project Omega, but by any department within TNA. Therefore, we believe that a collaborative approach to development could be embarked upon by The Catalogue and DORIS teams working together, each contributing resources and building parts of a larger system. Such an approach in the general sense is not new, but is widely and successfully used in many modern large Open Source software projects, whereby many different parties come together to collaborate on building a single system together.

### 4.1.1 Outlook Rationale

We believe this outlook aligns well with many of the requirements identified and outlined for TNA by Chaucer in the document "DORIS Next Generation Feasibility Study" (Draft Version 1.0 - September 2022). Specifically we would like to highlight the requirements from the pages of the Chaucer document that align well, or overlap, with Project Omega:

- Page 7 - the possibility for Project Omega to reduce, or even remove, the need for data duplication between DORIS and The Catalogue (i.e. ILDB and miniILDB databases), by instead unifying aspects of the two data models, and facilitating services upon this data for use by DORIS, ETNA, and the rest of the organisation at large.
- Page 10 - identifies the limitations of their existing use of Microsoft SQL Server Reporting Server. Likewise the current Catalogue has this limitation, and Project Omega aims to replace this with a reporting workbench for users which empowers them to construct their own reports utilising a "low-code" like approach.
- Page 13 - identifies another integration between DORIS and a Digital Surrogates database. This Digital Surrogates database has been identified as one such ancillary catalogue that should be subsumed by the new Project Omega Catalogue.
- Page 14 - states that there is no current integration between DORIS and the Surveyor system. The technology agnostic business services planned for Project Omega, could easily form the basis of such an integration. This would, without effort, naturally include an integration with DORIS data if some, or all, of DORIS's data were to form part of the Project Omega Catalogue. This would remove today's need of duplicated effort whereby DORIS and Surveyor each have to be updated manually by staff to keep them in synchronisation.
- Page 14 - states that there is no current integration between Conservation and DORIS. Conservation make use of two systems currently: ResearchSpace, and Handwell. ResearchSpace is constructed atop many of the same technologies (e.g. JVM, Graph Data Storage, and Open Standards) as used in Project Omega.



---

## PROJECT OMEGA

Project Omega has been in discussion about developing integrations with Conservation's ResearchSpace system. If DORIS and ResearchSpace were to utilise the technology agnostic business services planned for Project Omega, both could benefit from a reduced path to integration.

- Page 15 - states that the existing KBS App needs real-time access to document, i.e. catalogue, references for validation purposes. Project Omega is already scheduled to provide services for retrieval and validation of catalogue references as part of its technology agnostic business services; as such any replacement to the KBS App could benefit from these.
- Page 39 - states that there is no advanced functionality for extracting data or creating charts using DORIS. The technology agnostic business services planned for Project Omega include facilities for importing and extracting data. DORIS could also benefit from utilising either components of these facilities or perhaps even the facilities in total.
- Pages 39, 40, and 41 - state that no record is kept in DORIS as to which member of staff has handled the documents, which makes auditing very challenging. Should DORIS wish to store such information into the Project Omega Catalogue Model, then such a facility is available intrinsically. This comes about as in the Project Omega data model, all instances of objects are immutable, and an entire chain of provenance is recorded using W3C Prov for every update made to the system.
- Page 40 - states that it would be useful for the Surveyor system to link with the a potential new DORIS system to allow users to input and save the space requirements of different records (maps, folios, specialty items such as charters, etc.). Such physical information of storage requirements could be linked to the Realisations of the Records held in Project Omega, for use by DORIS, Surveyor, and other systems via Project Omega's planned technology agnostic business services.
- Page 41 - discusses records copying, the ability to determine if a record has been copied before, and where its copy is held. This Records Copying database has a relationship with the Catalogue References maintained in PROCat and ILDB (which will be replaced by Project Omega), as such it was identified as one possible ancillary catalogue that could be subsumed by the new Project Omega Catalogue.
- Page 42 - discusses User Interface Accessibility requirements. On this point we would just highlight that Project Omega's Catalogue Editorial application that is under development shares many of the same concerns and requirements, e.g.: WCAG, and that we have spent quite some time comparing available UI frameworks against such requirements. We are happy to share our knowledge in this area.

## GLOSSARY

*Glossary Table*

Term	Description
DORIS	Document Ordering Record Information System
DRI	Digital Repository Infrastructure
EAI	Enterprise Application Integration
ePRO	Electronic Public Records Office
ESB	Enterprise Service Bus
Project Omega	Project to develop a Pan-Archival Catalogue to subsume and replace existing legacy Catalogue Systems within The National Archives.
SAR	System for Access Regulation
TNA	The National Archives; successor of PRO