

THE
NATIONAL
ARCHIVES

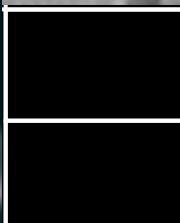


The Architecture

:: Project Omega



Adam Retter
Project Omega Architect
(Consultant)
23 October 2023



Phase 1

:: Develop the Vision and Recommendations

November 2019 to June 2020

Phase 2

:: Develop a Pilot Implementing the Recommendations

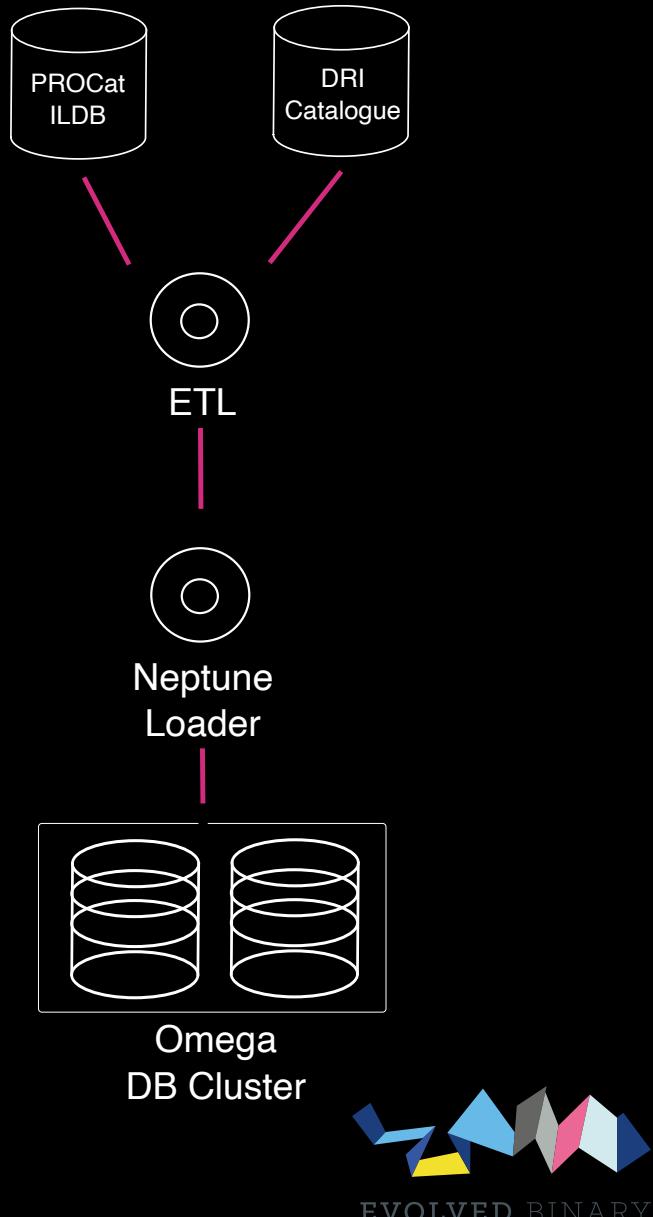
January 2021 to November 2023



EVOLVED BINARY

Data Architecture

- **Existing Paper Catalogue Data**
 - Relational Model
 - with embedded EAD XML document fragments
 - with embedded key/value property model
 - Microsoft SQL Server
 - 12+ million records
- **Other Catalogues:**
 - Variety of Models - Relational, Document, Graph
 - Variety of DBMS - SQL Server, Jena, Mongo, etc.
- **New Omega Catalogue**
 - Graph Model
 - ETL (Extract, Transform, and Load) existing data to new model
 - Extensive cleanup of existing data is required
 - Hitachi Pentaho (migrating to Apache HOP)



The New Data Model

- It's a graph! RDF (Resource Description Framework)
- Based on Matterhorn Model
 - Reuse well known open vocabularies
 - Feels like a bit of a Chimera
- Kind of an expression of RiC-CM
 - In my opinion, RiC is not (yet) fit for purpose
- Incorporates ideas from FRBR (Functional Requirements for Bibliographic Records)
- Provenance is mandatory (via W3C Pro.)

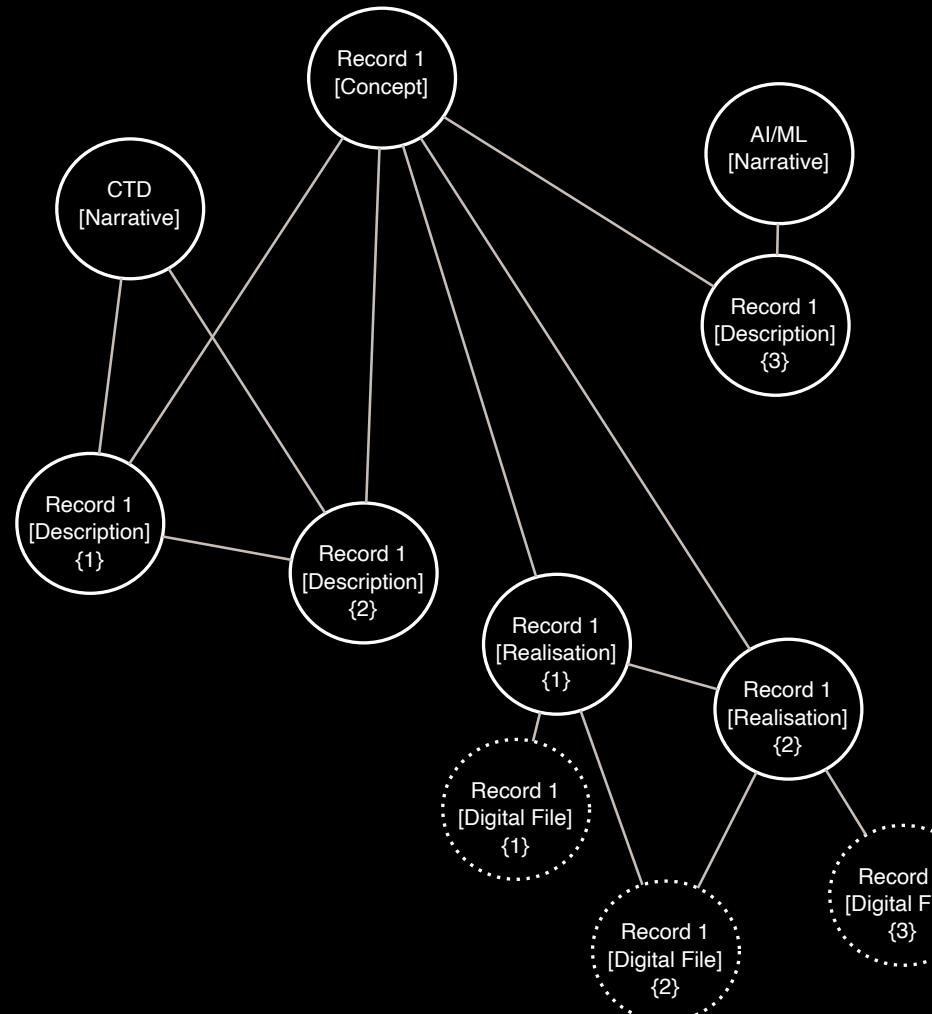


The New Data Model

- **Assets**
 - Record, Record Set, Person, Corporate Body, FOI Exemption, and Taxonomy
- **Supports:**
 - Physical, Digital, and Hybrid assets
 - Redactions, and Surrogates
 - Non-Hierarchical Systems of Arrangement
 - File Format Migration, and Digital File deduplication
 - Taxonomic classification
 - Multiple Narratives
- **Facilitates Transparent Government**
 - All data is immutable
 - Create new “versions”
 - Preserve every change
 - Track and preserve provenance

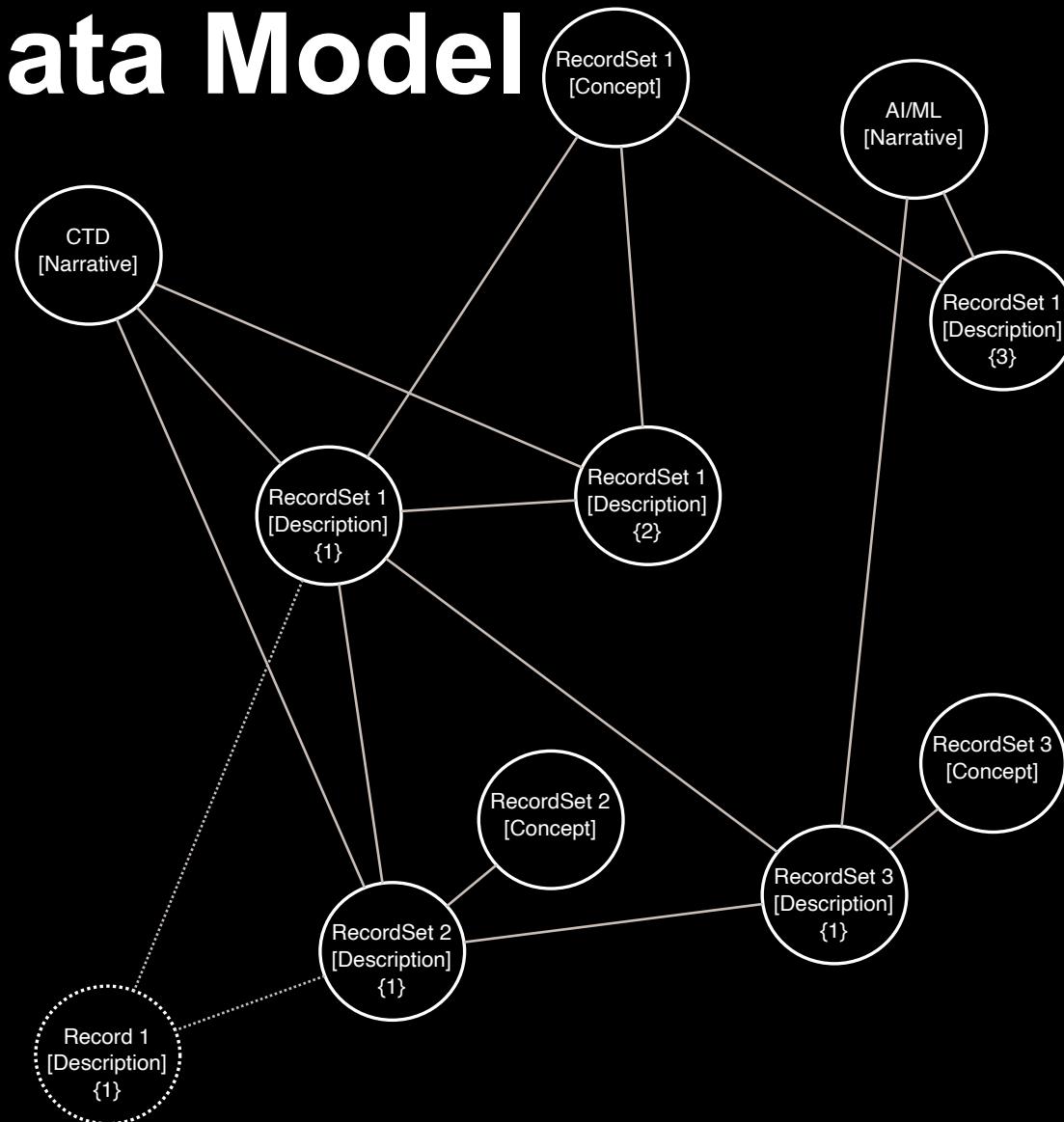


The New Data Model



EVOLVED BINARY

The New Data Model



EVOLVED BINARY

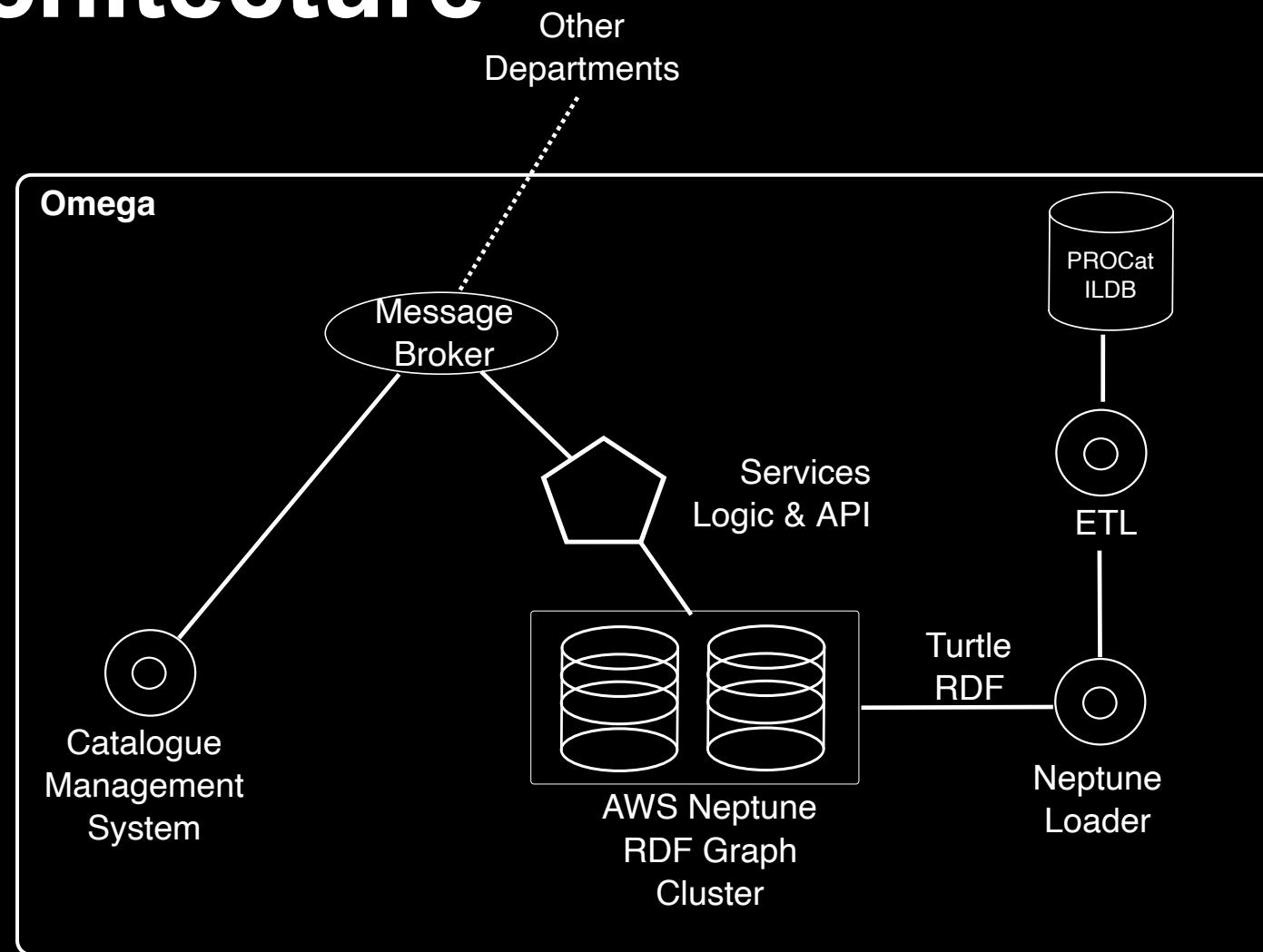
System Architecture

- Separate the Data from the Services
- Separate the Catalogue Management System (Editorial)
- Provide the Services to the entire organisation
- Decouple Systems and Departments
 - Messaging
 - Asynchronous Communication
 - Synchronous like Request/Reply over Messaging
 - You/I can continue even if I/You are unavailable



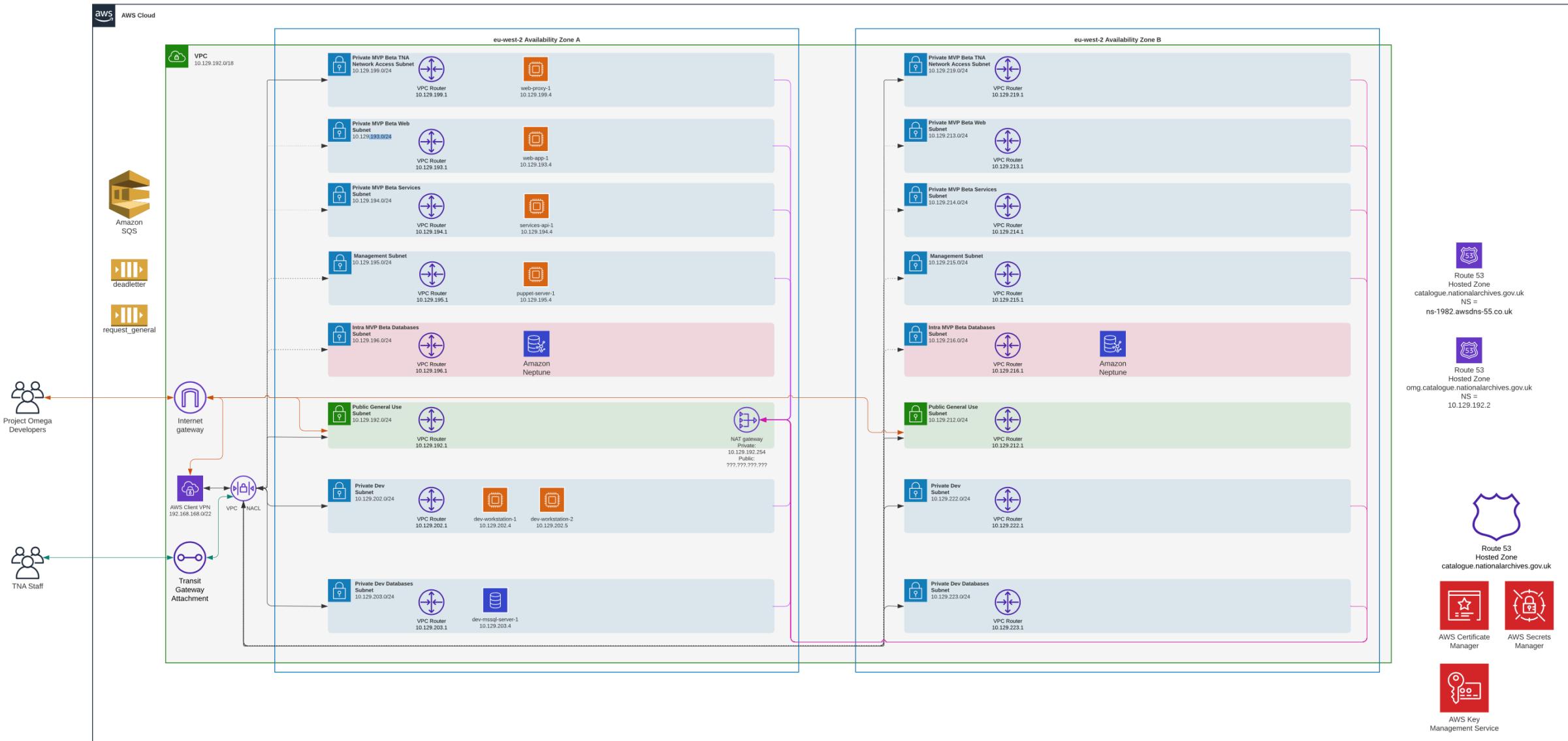
EVOLVED BINARY

System Architecture



EVOLVED BINARY

Cloud Architecture



For more information:

- Internal Website - www.omg.catalogue.nationalarchives.gov.uk
 - Online from 2023-10-27
- Project Omega Catalogue Model Proposal (TNA-CMP20/1)
 - <https://cdn.nationalarchives.gov.uk/documents/omega-catalogue-model-proposal.pdf>
- Project Omega - Catalogue Data Model (TNA-OCDM20/1)
 - <https://cdn.nationalarchives.gov.uk/documents/omega-catalogue-data-model.pdf>
- Project Omega - DORIS Analysis (TNA-OCDA22/1)
- Various Articles and Presentations
 - 2021-09-13 - Developing a Pan-Archival Linked Data Catalogue - <https://www.youtube.com/watch?v=iLOQo7412r4>
 - 2021-07-23 - Omega API Initial Talking Points - <https://slides.com/adamretter/omega-api-initial-talking-points>
 - 2021-04-07 - What is a Record? - <https://slides.com/adamretter/omega-what-is-a-record>
 - 2020-05-29 - Omega Catalogue Data Model - <https://slides.com/adamretter/omega-catalogue-data-model>
 - 2020-01-28 - Pan-Archival Cataloguing - <https://slides.com/adamretter/tna-pan-archival-cataloguing>



Omega Today

- Running in TNA's Amazon Web Services
- Catalogue Services are available to the organisation
 - TNA Developers - If you want to try them... Ask us!
- Catalogue Management System (Editorial) is available to the organisation
 - Communicating with the Catalogue Services
 - TNA Staff - If you want to access it... Ask us!
- Private SPARQL Query environments???
 - TNA Data Scientists - Ask us!





Thank you

Adam Retter
Director
Evolved Binary
London, UK

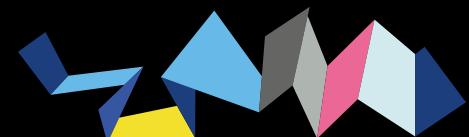
www.evolvedbinary.com



adam@evolvedbinary.com



@adamretter



EVOLVED BINARY