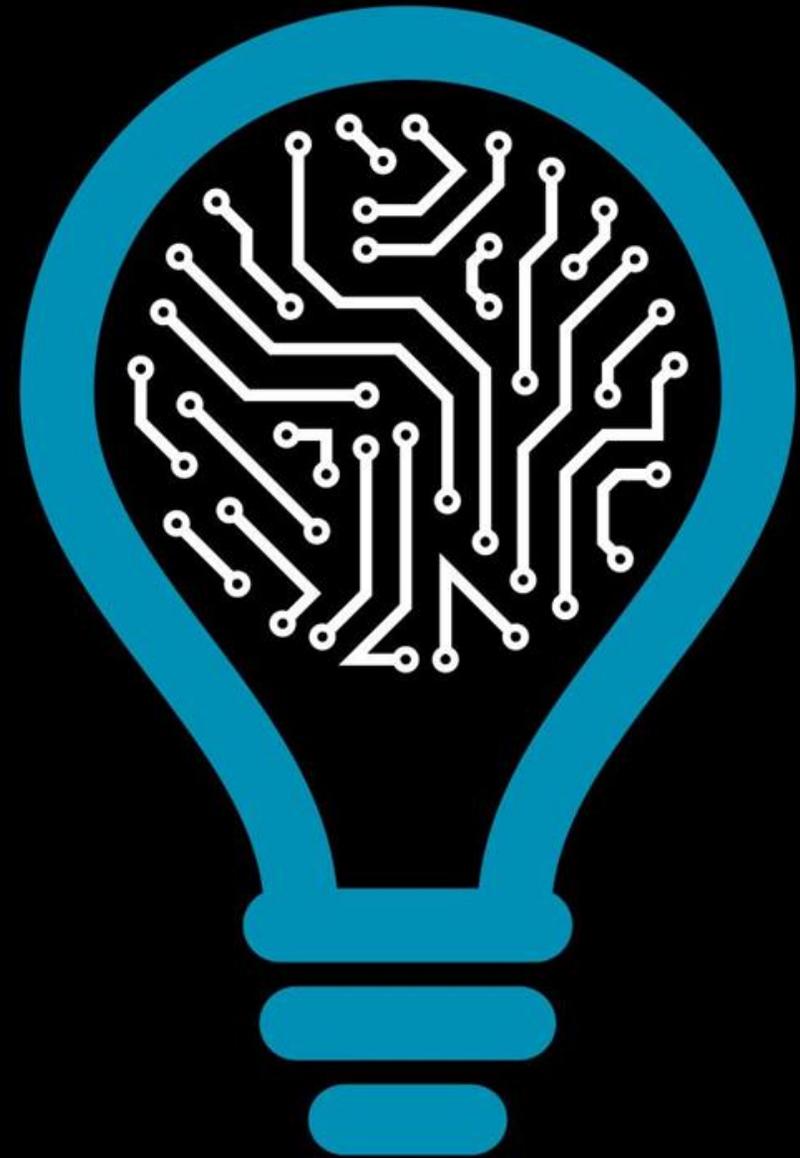


EPISA - ENTITY AND PROPERTY INFERENCE FOR SEMANTIC ARCHIVES

55th CIDOC CRM and 48th FRBR CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
Sérgio Nunes

This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project DSAIPA/DS/0023/2018. Inês Koch is also financed by FCT within the research grant 2020.08755.BD.



Inês Koch



Ph.D. candidate - Doctoral Programme in Digital Media
EPISA Project - Entity and Property Inference for Semantic Archives

Master in Information Science

Porto, Portugal

Outline

- EPISA Project
- ArchOnto
- EPISA Platform

EPISA

Entity and Property Inference for
Semantic Archives

EPISA Project

- Part of the ongoing renewal of the The Portuguese National Archives' existing data infrastructure
- Goals:
 - Develop a prototype for an open-source knowledge graph platform representing archival information on a linked data model.
 - Find ways to guarantee the migration of contents stored according to ICA (International Council on Archives) standards to an ontology-based model, the CIDOC CRM (Conceptual Reference Model).

EPISA team

INESC TEC: Cristina Ribeiro, Carla Teixeira Lopes, Gabriel David, Sérgio Nunes, Bruno Giesteira, Inês Koch, Mariana Dias, Cláudia Martins, Tiago Silva, Tiago Gomes, Catarina Pires, Camilla Oliveira, Pedro da Silva

U.Évora: Irene Rodrigues, Dora Melo, Davide Varagnolo

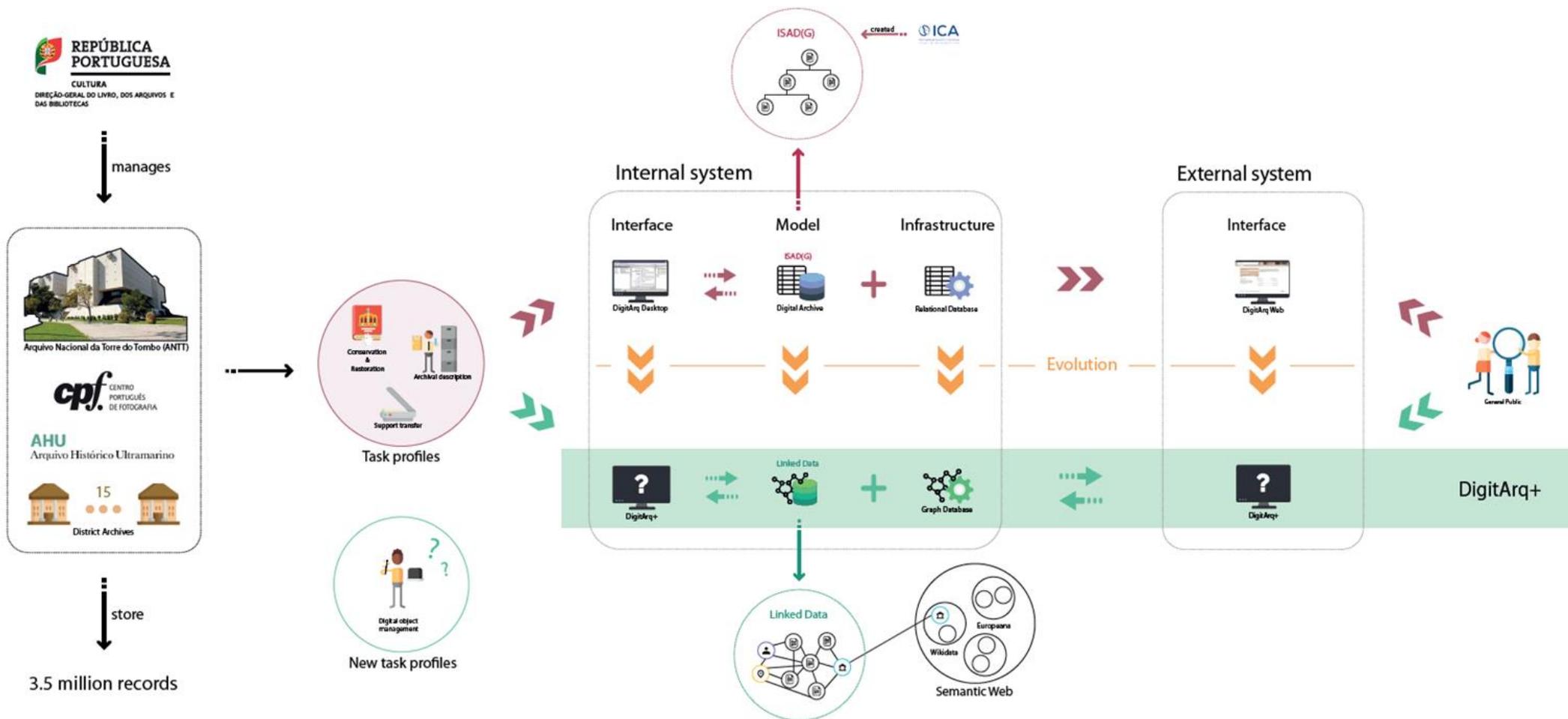
DGLAB: Ana Maria Rodrigues, Lucília Runa, Maria José Almeida, Cristiane Casaca, Silvestre Lacerda



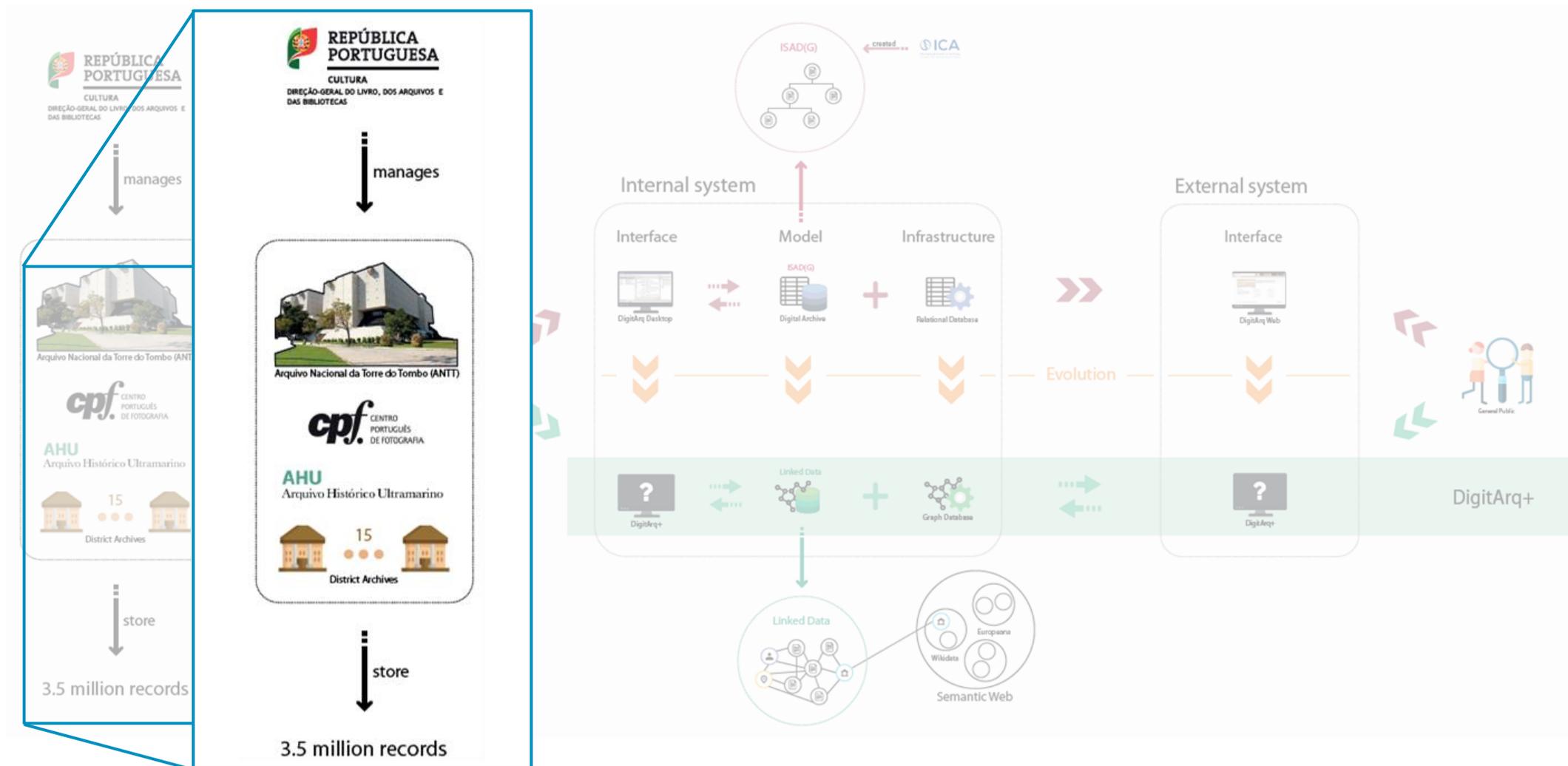
EPISA tasks

- T1: Project management and dissemination of results
- T2: Linked data model for cultural content
- T3: Semantic migration with relations inference
- T4: Knowledge propagation
- T5: Document mining for automatic metadata records
- T6: Knowledge graph management platform KGMP
- T7: Exploration and querying interface
- T8: User evaluation

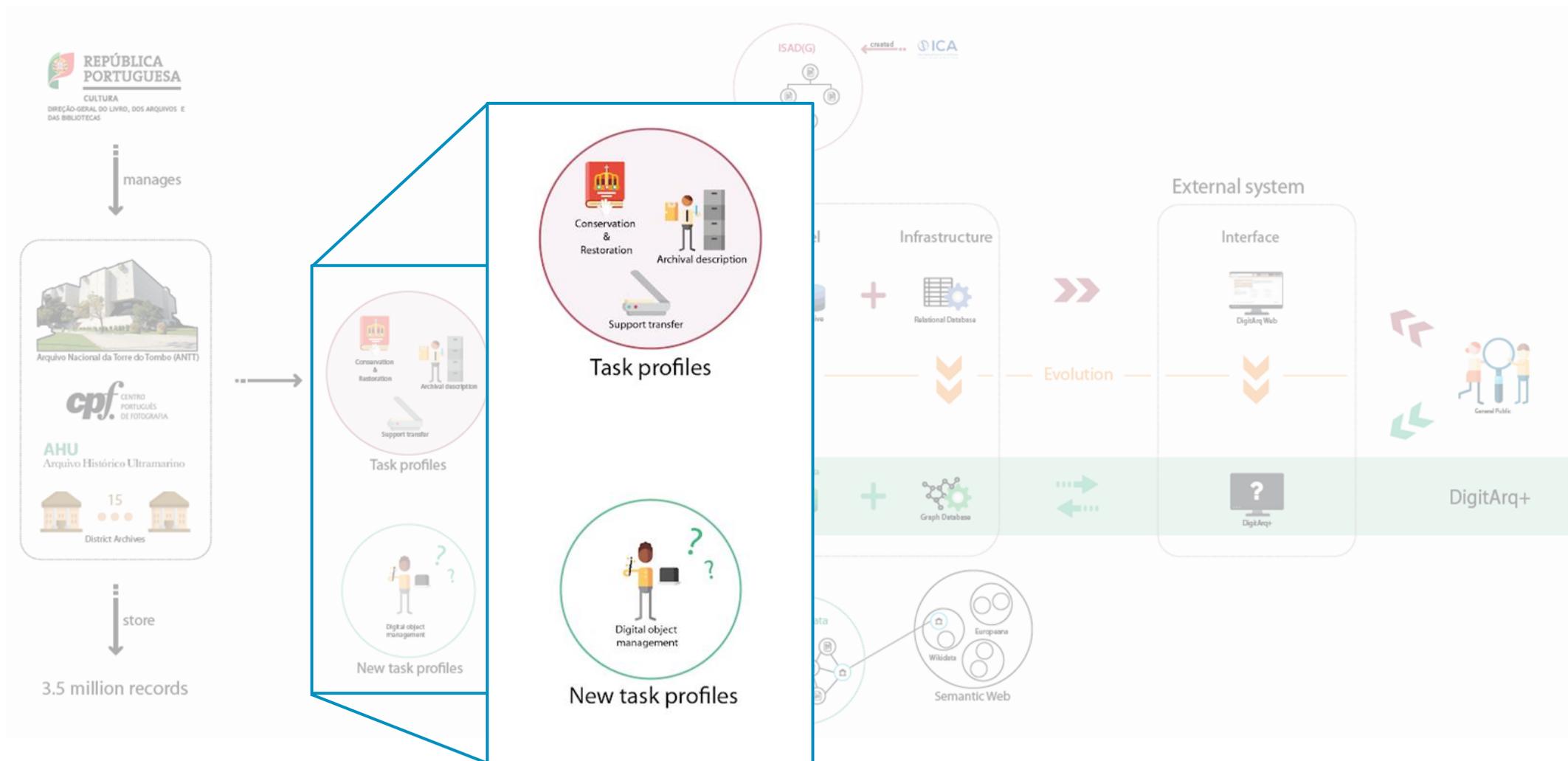
From Digitalarq to Digitalarq+



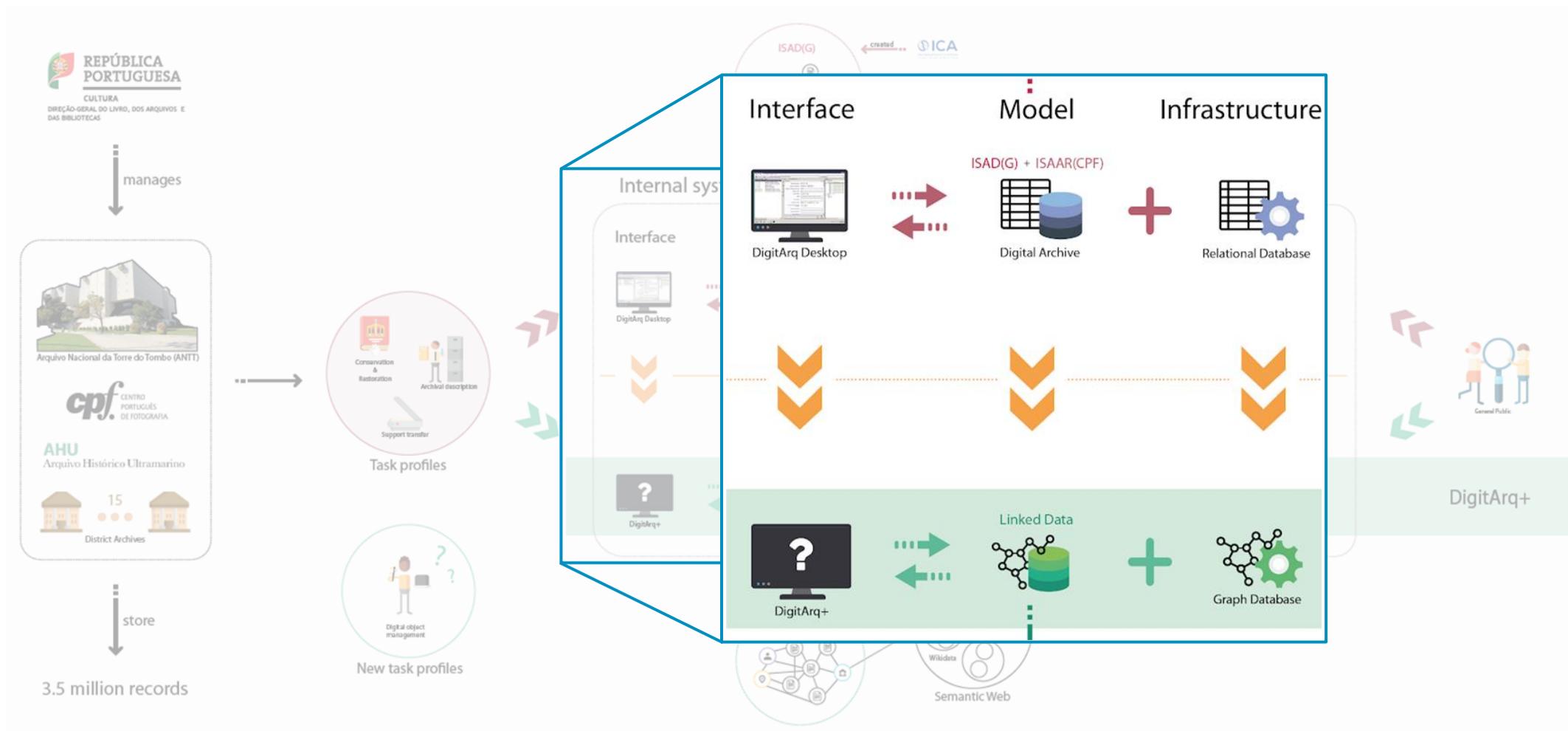
From Digitalarq to Digitalarq+



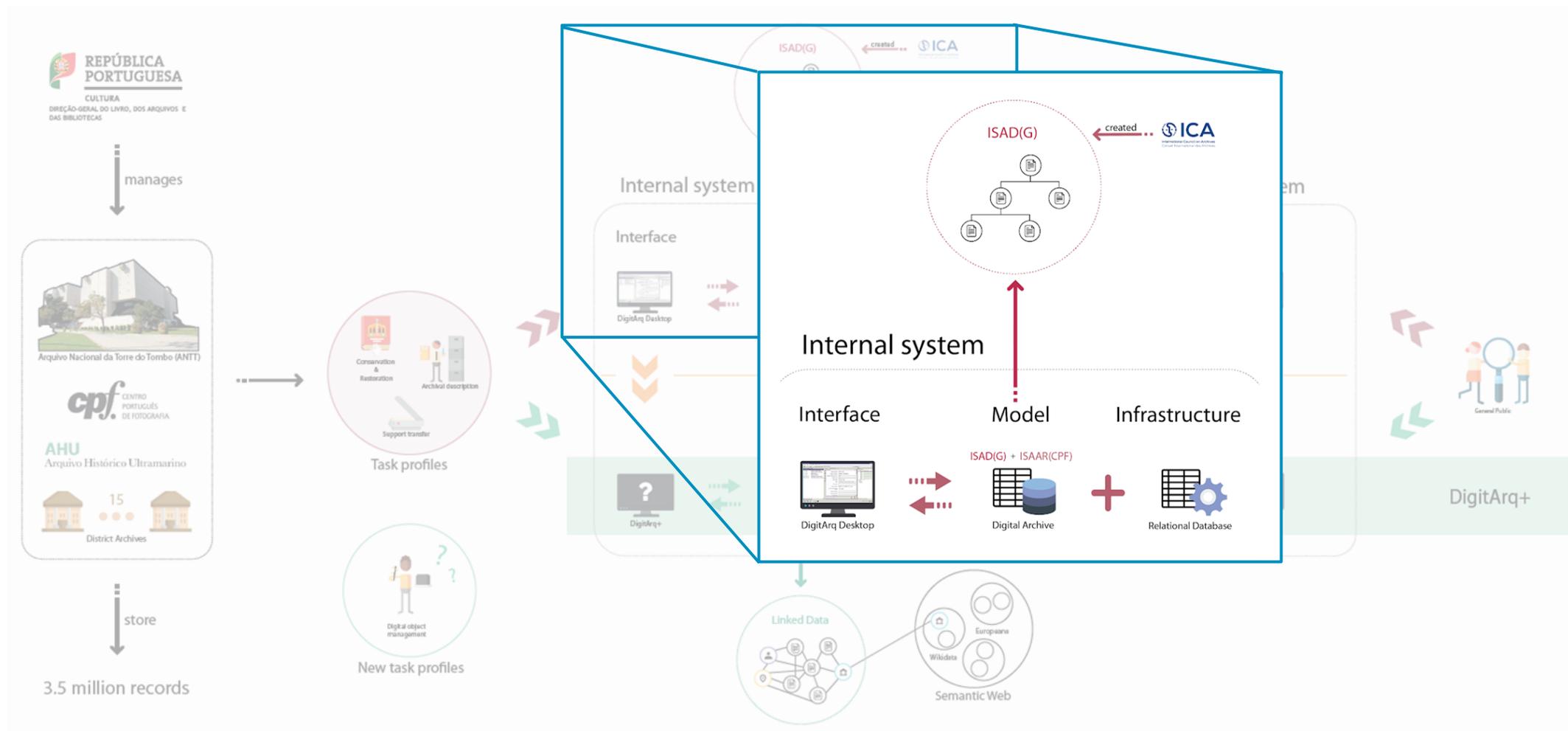
From Digitalarq to Digitalarq+



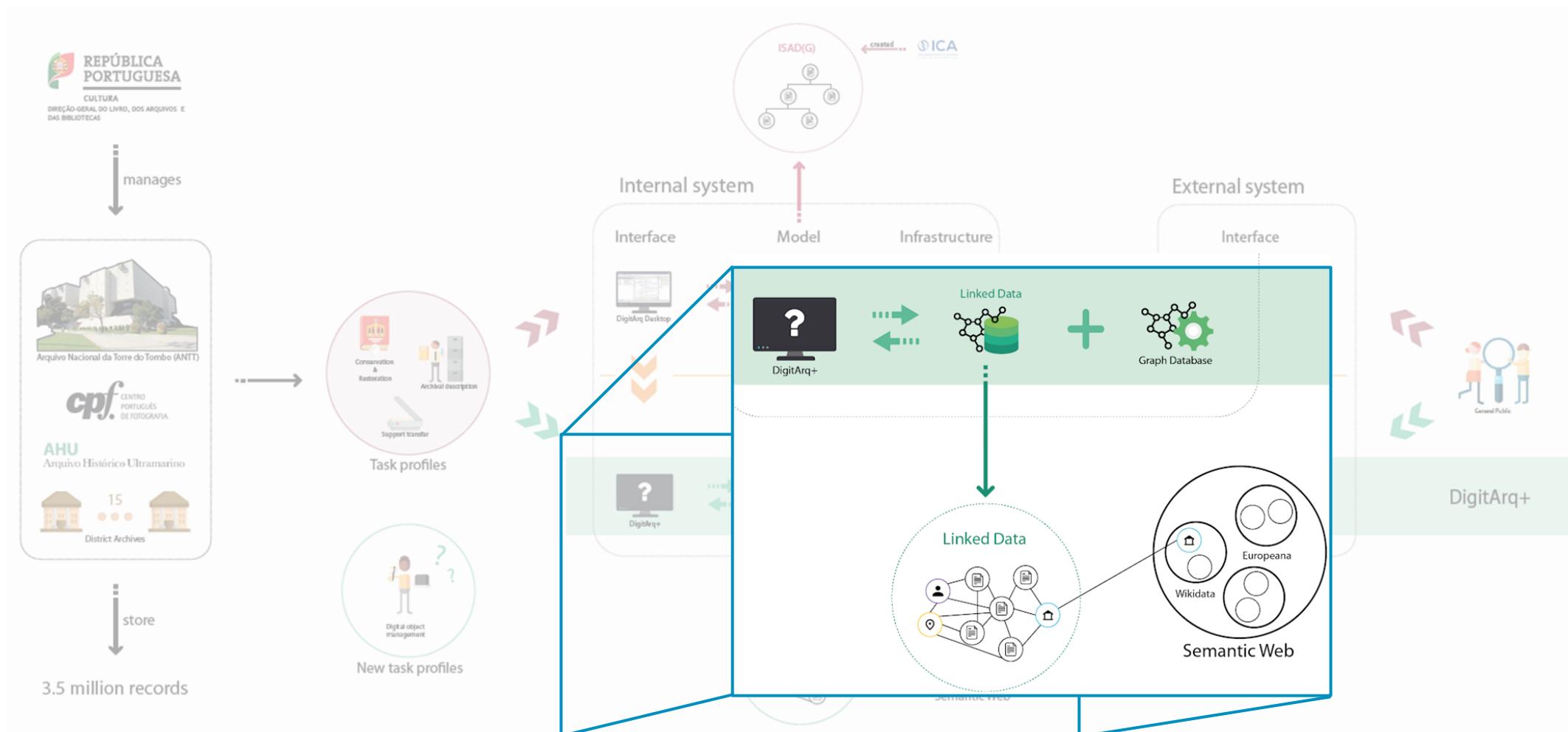
From DigitalArq to DigitalArq+



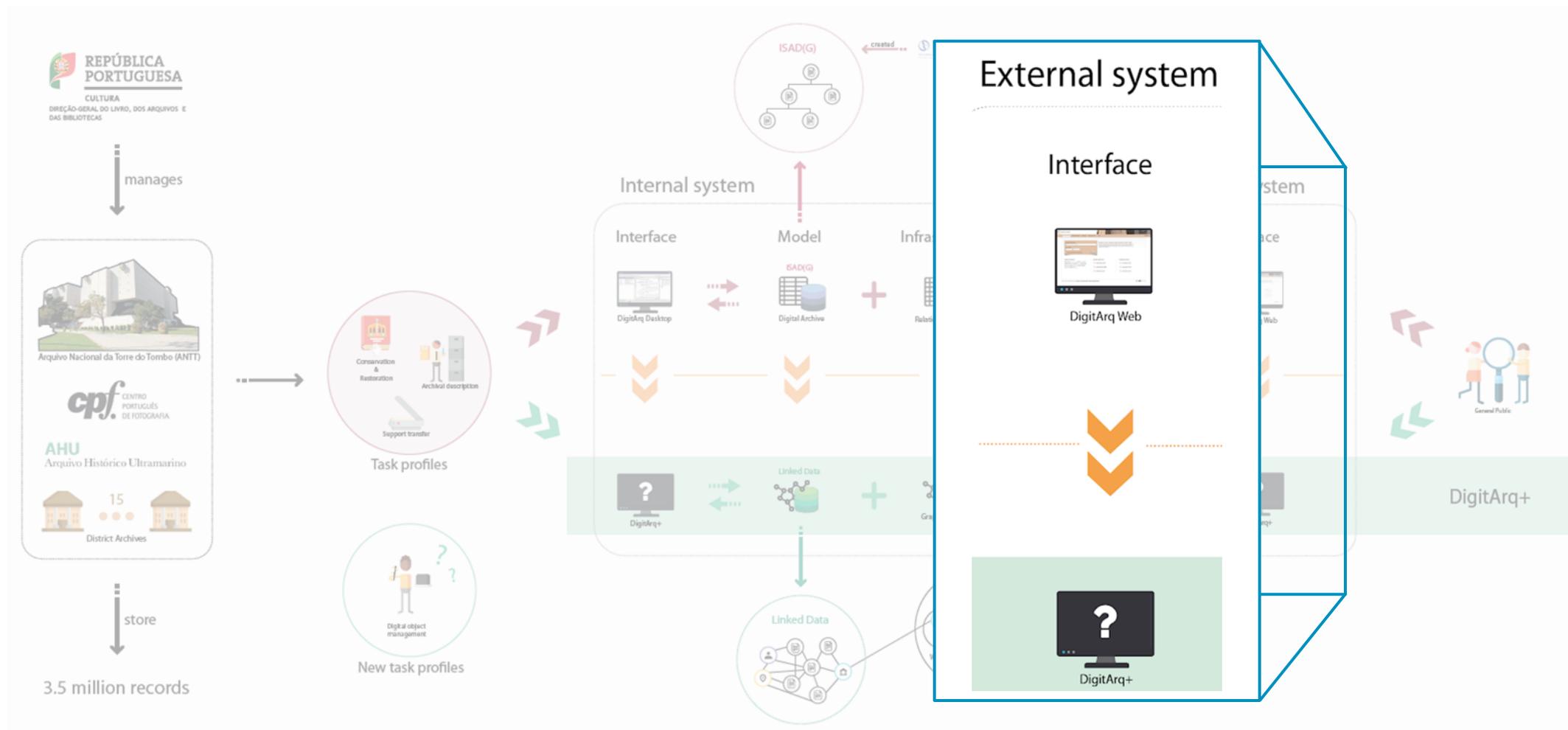
From Digitalarq to Digitalarq+



From Digitalarq to Digitalarq+



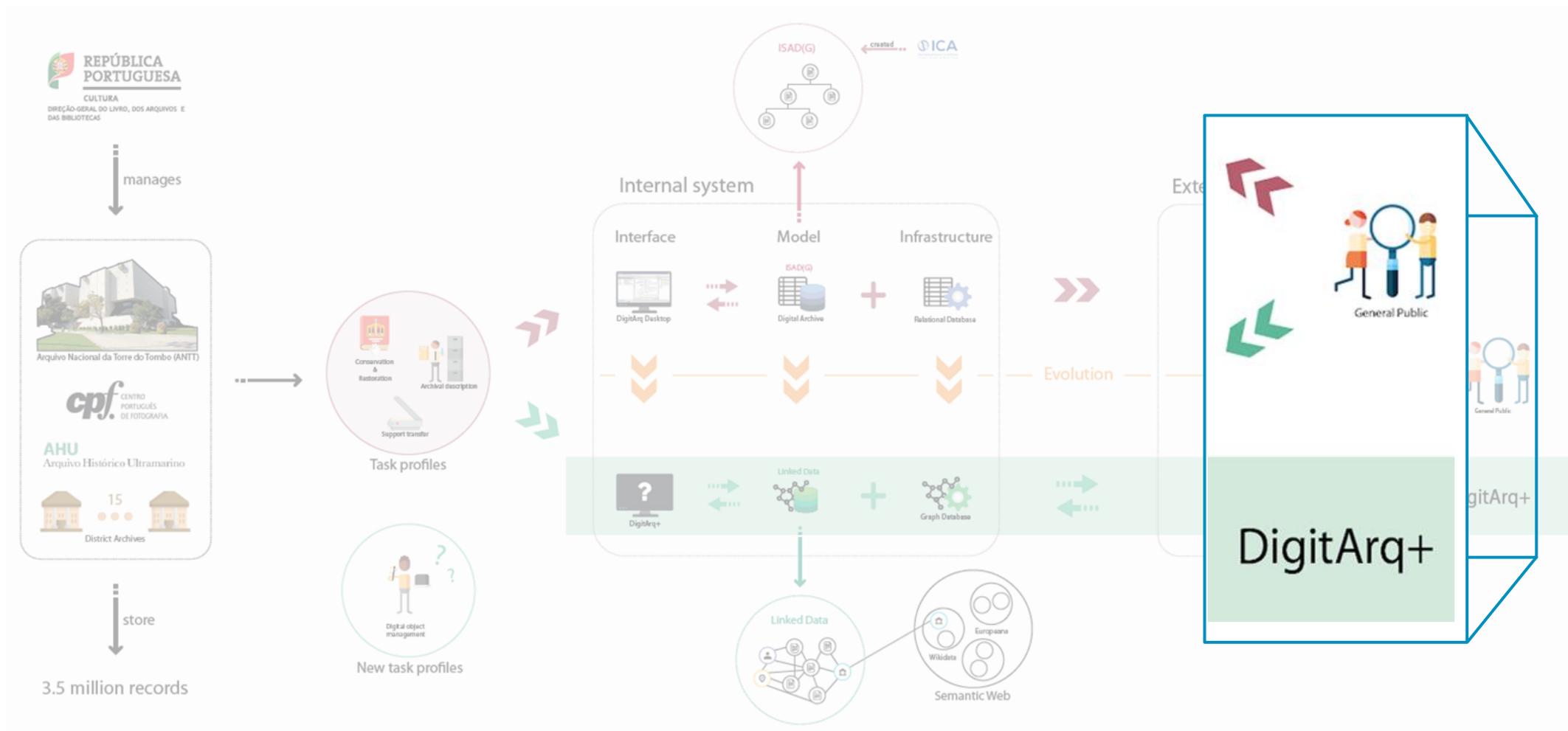
From Digitalarq to Digitalarq+



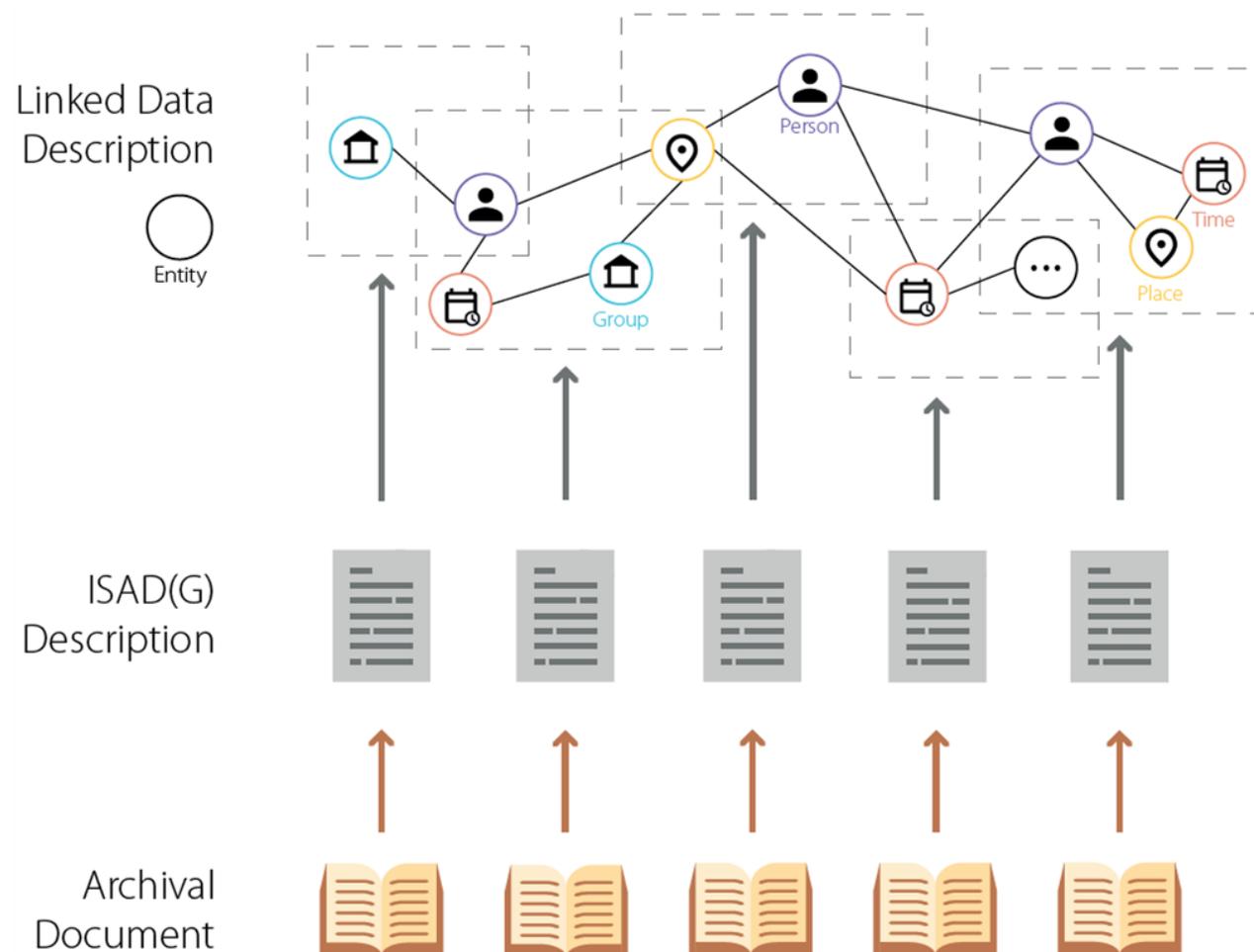
From Digitalarq to Digitalarq+

The screenshot shows the Digitalarq+ website interface. At the top, there is a header for the Portuguese Republic and the National Archive of Torre do Tombo (DGLAB). Below this is a navigation bar with options like 'SIMPLE SEARCH', 'ADVANCED SEARCH', and 'ONLINE SERVICES'. A search box is prominently displayed with a 'SEARCH' button. The main content area includes a welcome message and three columns of document recommendations: 'FREQUENT SEARCHES', 'MOST VIEWED DOCUMENTS', and 'RECENT DOCUMENTS'. Annotations include a blue box labeled 'External system' pointing to the top navigation and search area, and a green box labeled 'DigitArq+' pointing to the bottom section of the page. A diagram on the right illustrates the integration of an external system with the main portal.

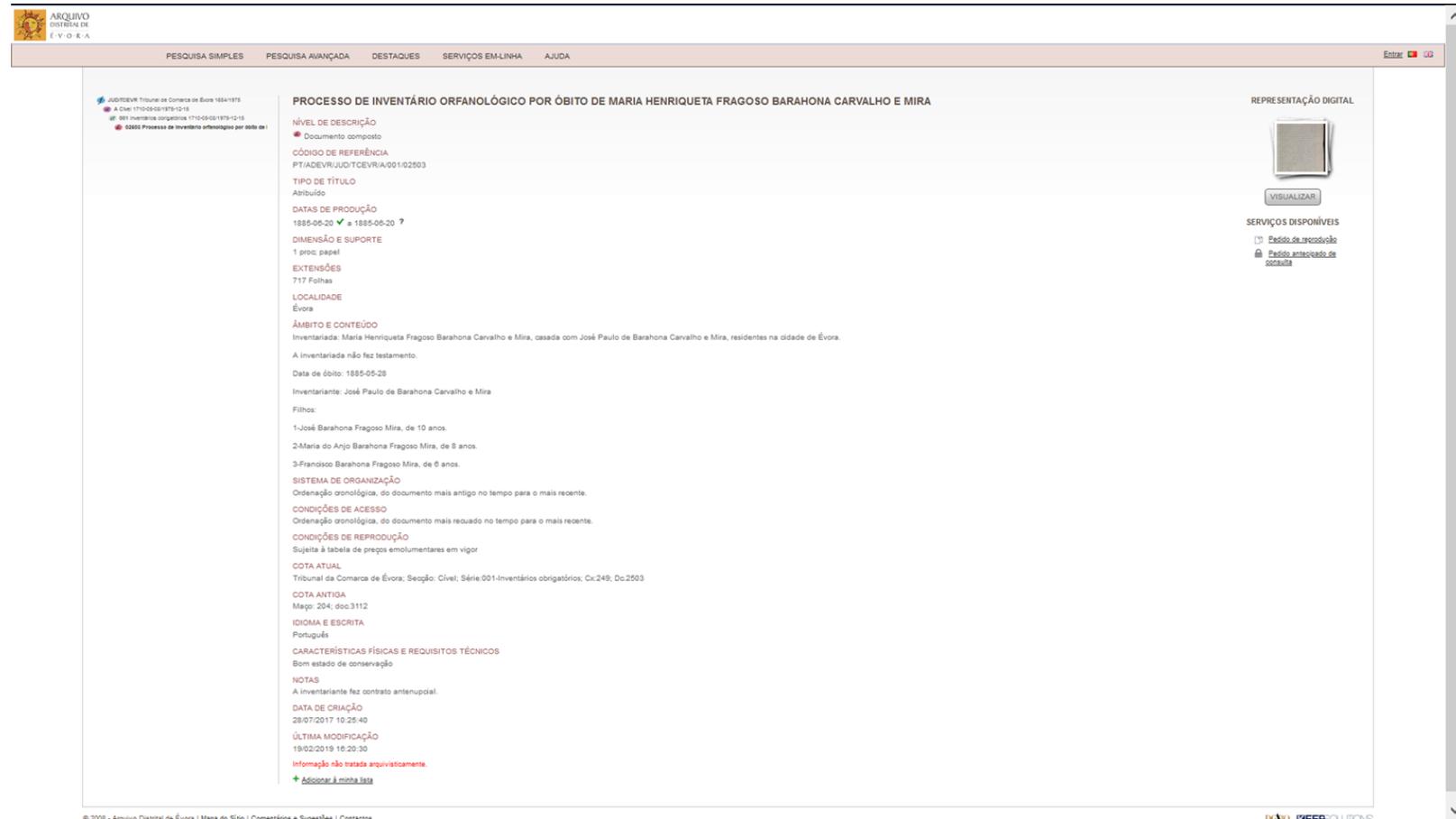
From Digitalarq to Digitalarq+



Proposed evolution of archival description



Atomization of an ISAD(G) record in linked data



ARQUIVO HISTÓRICO DE ÉVORA

PESQUISA SIMPLES PESQUISA AVANÇADA DESTAQUES SERVIÇOS EM-LINHA AJUDA

ENTRADA 133

PROCESSO DE INVENTÁRIO ORFANOLÓGICO POR ÓBITO DE MARIA HENRIQUETA FRAGOSO BARAHONA CARVALHO E MIRA

NÍVEL DE DESCRIÇÃO
Documento composto

CÓDIGO DE REFERÊNCIA
PT/ADEV/AR/JUD/TCEVRIA/001/02503

TIPO DE TÍTULO
Atribuído

DATAS DE PRODUÇÃO
1885-06-20 a 1885-06-20

DIMENSÃO E SUPORTE
1 proc. papel

EXTENSÕES
717 Folhas

LOCALIDADE
Évora

ÂMBITO E CONTEÚDO
Inventariada: Maria Henriqueta Fragoso Barahona Carvalho e Mira, casada com José Paulo de Barahona Carvalho e Mira, residentes na cidade de Évora.
A inventariada não fez testamento.
Data de óbito: 1885-05-28
Inventariante: José Paulo de Barahona Carvalho e Mira
Filhos:
1-José Barahona Fragoso Mira, de 10 anos.
2-Maria do Anjo Barahona Fragoso Mira, de 8 anos.
3-Francoiso Barahona Fragoso Mira, de 6 anos.

SISTEMA DE ORGANIZAÇÃO
Ordenação cronológica, do documento mais antigo no tempo para o mais recente.

CONDIÇÕES DE ACESSO
Ordenação cronológica, do documento mais recuado no tempo para o mais recente.

CONDIÇÕES DE REPRODUÇÃO
Sujeita à tabela de preços emolumentares em vigor

COTA ATUAL
Tribunal da Comarca de Évora; Secção: Cível; Série:001-Inventários obrigatórios; Cr:249; Dc:2503

COTA ANTIGA
Maço: 204; doc:3112

IDIOMA E ESCRITA
Português

CARACTERÍSTICAS FÍSICAS E REQUISITOS TÉCNICOS
Bom estado de conservação

NOTAS
A inventariante fez contrato antenupcial.

DATA DE CRIAÇÃO
28/07/2017 10:25:40

ÚLTIMA MODIFICAÇÃO
19/02/2019 16:20:30

Informação não tratada arquivisticamente.

[Adicionar à minha lista](#)

REPRESENTAÇÃO DIGITAL
VIZUALIZAR

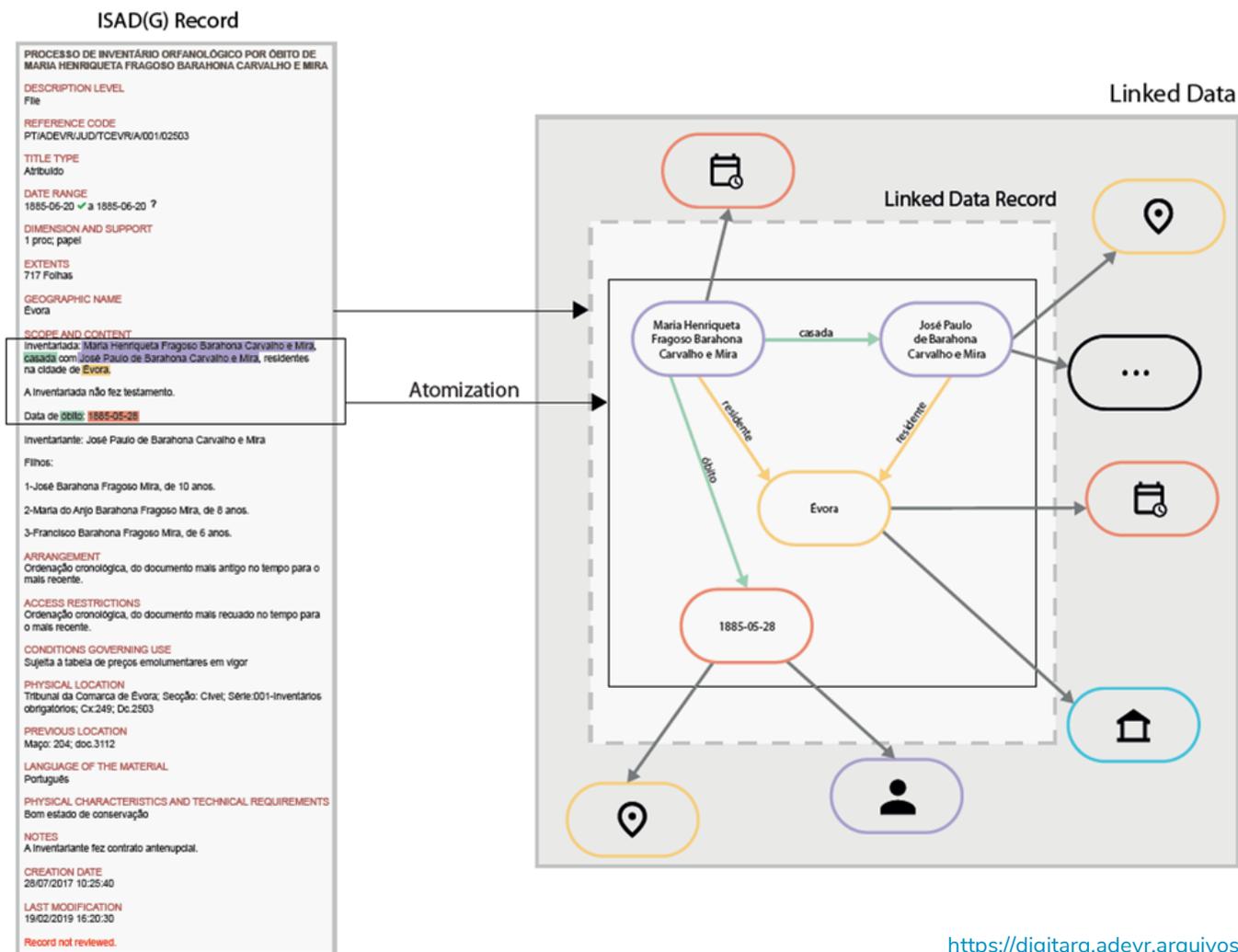
SERVIÇOS DISPONÍVEIS
Pedido de reprodução
Pedido antecipado de consulta

© 2008 - Arquivo Distrital de Évora | Mapa do Sítio | Contactos e Localização | Privacidade

INESCTEC



Atomization of an ISAD(G) record in linked data



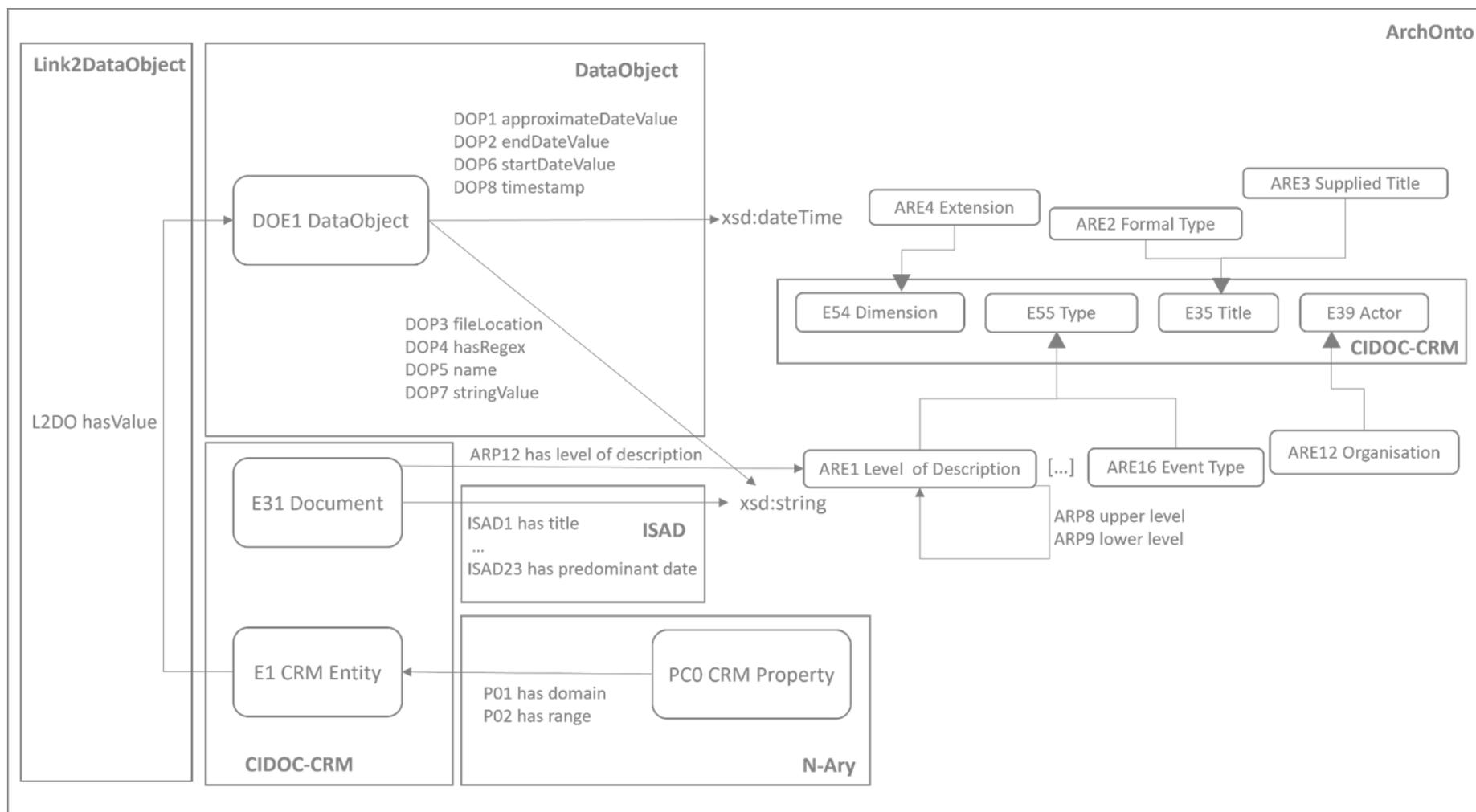
ArchOnto

An extension of CIDOC CRM for
archives

ArchOnto

- Is composed by 5 different ontologies:
 - CIDOC CRM
 - DataObject
 - N-ary
 - ISAD Ontology
 - Link2DataObject
- Use the prefixes:
 - ARE – Archival Entity, ex.: ARE1 Level of Description.
 - ARP – Archival Property, ex.: APR12 has level of description.

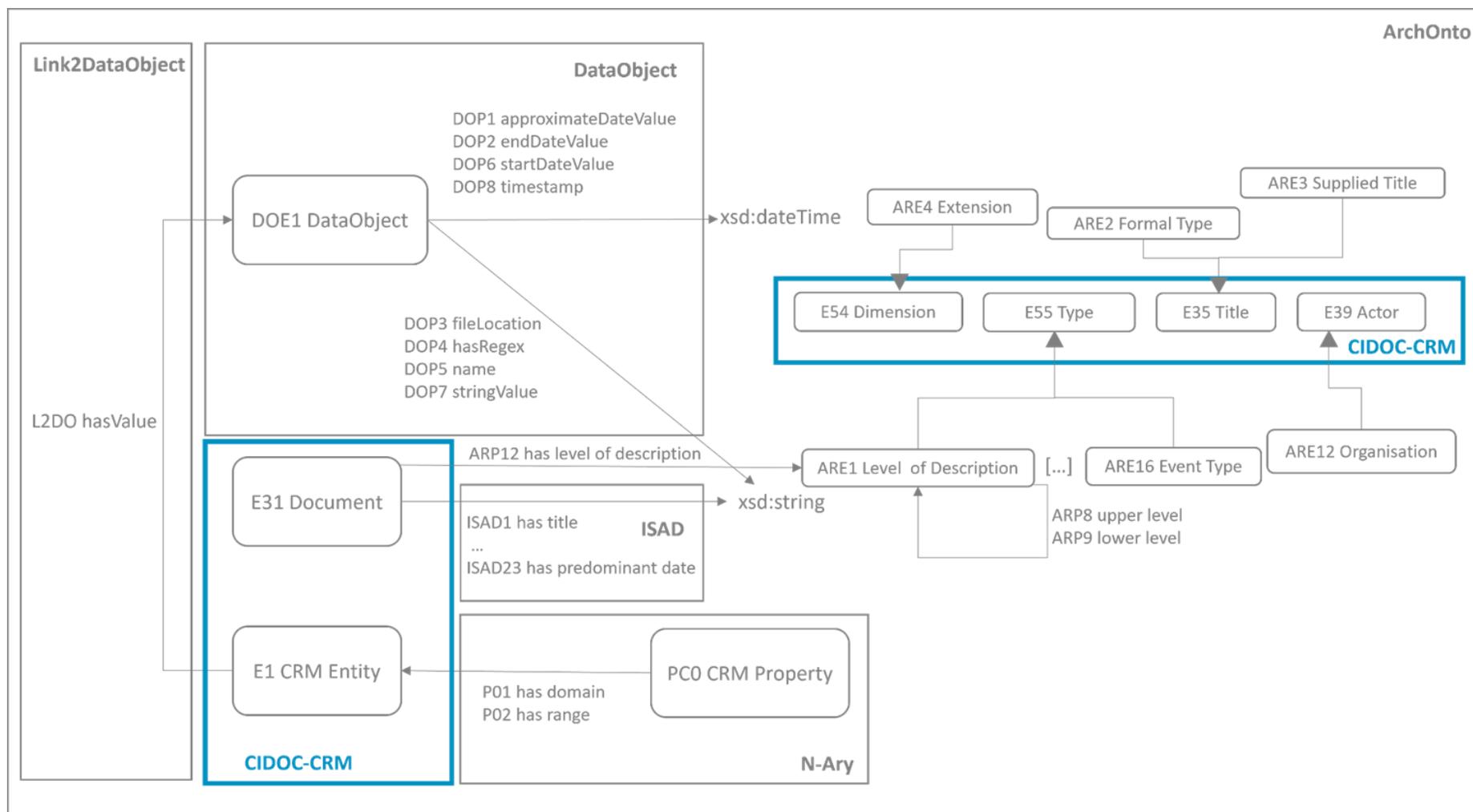
ArchOnto



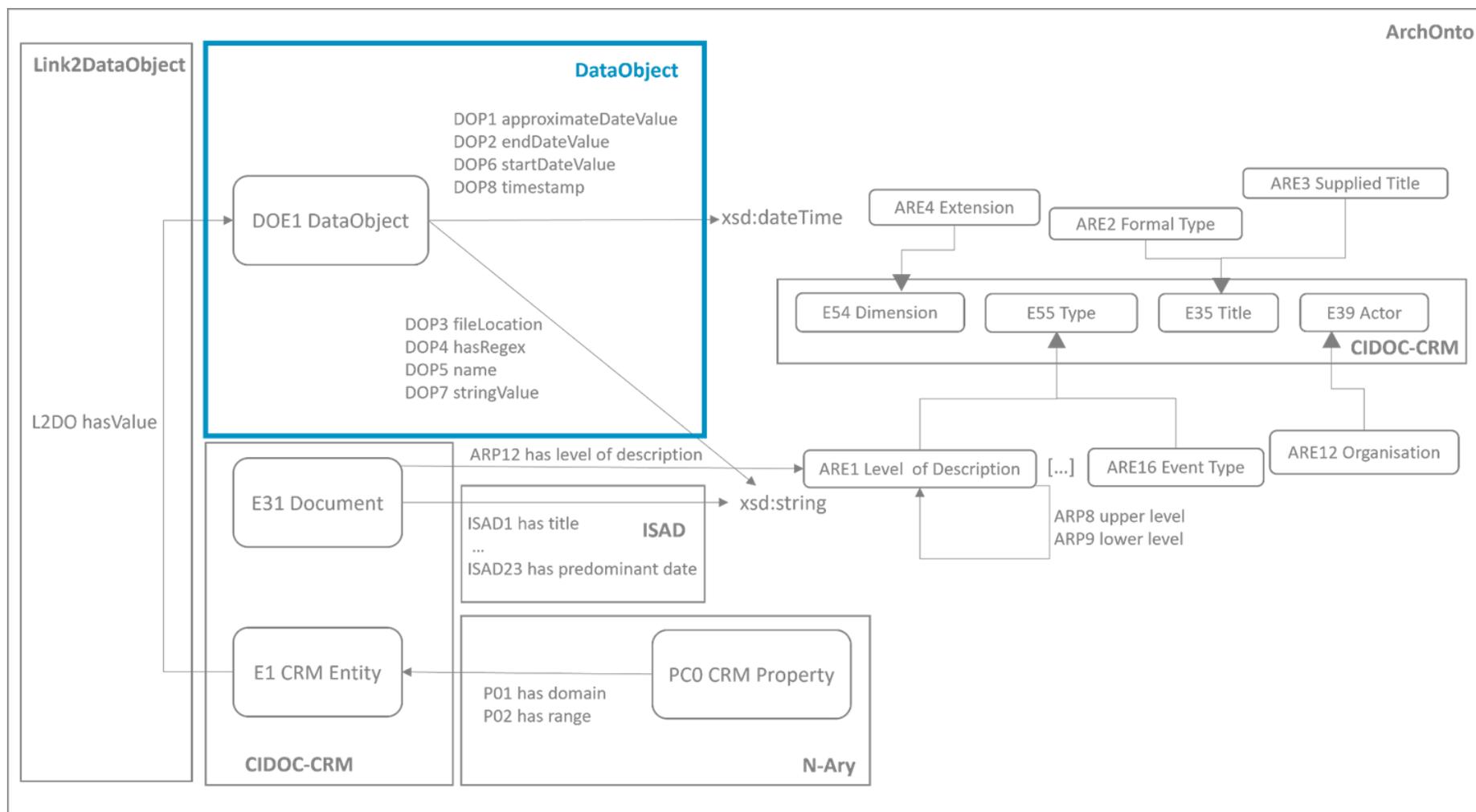
ArchOnto - Controlled vocabularies

Class	Vocabulary	Example Values
ARE1 Level of Description	Level of description	Fonds; Series; Section; File; Item
ARE2 Formal Title ARE3 Supplied Title	Title Type	Formal, Supplied
ARE5 Identifier Type	Identifier of collective person/group	PT; VCT; AGH01; 161016; ADLSB; 600084892; PT-LiBN
ARE6 Date Type ARE9 Date Certainty	Type of time period	Exact dates; Inferred dates; Predominant dates
ARE7 Name Type	Type of name of collective person/group	Authorized form of name; Another form of the name; Parallel name form
ARE8 Role Type	Role played	Producer; Material Author; Recipient
ARE11 Documentary Typology	Documentary Typology	Certificate; Income book; Patent
ARE13 Subject Type	Subject	Education; Science; Law; Management
ARE14 Place Type	Type of jurisdictional entity	Ocean; Archipelago; Mountain range; Country; District
ARE15 Acquisition Type	Transfer of Custody / Acquisition Identifier	Purchase; Giving; Donation; Deposit; Swap; Legacy; Reintegration; Transfer
ARE16 Event Type	Event Type	Evaluation; Expertise; Financial management
E56 Language	Language Identifier	Portuguese; Latin; French; Greek
E57 Material	Support	Paper; Parchment; Photosensitive film
E58 Measurement Unit	Measurement Unit	Centimeter; Gram; Byte; Minute; Pack
E98 Currency	Currency	Euro; Dollar; Kwanza

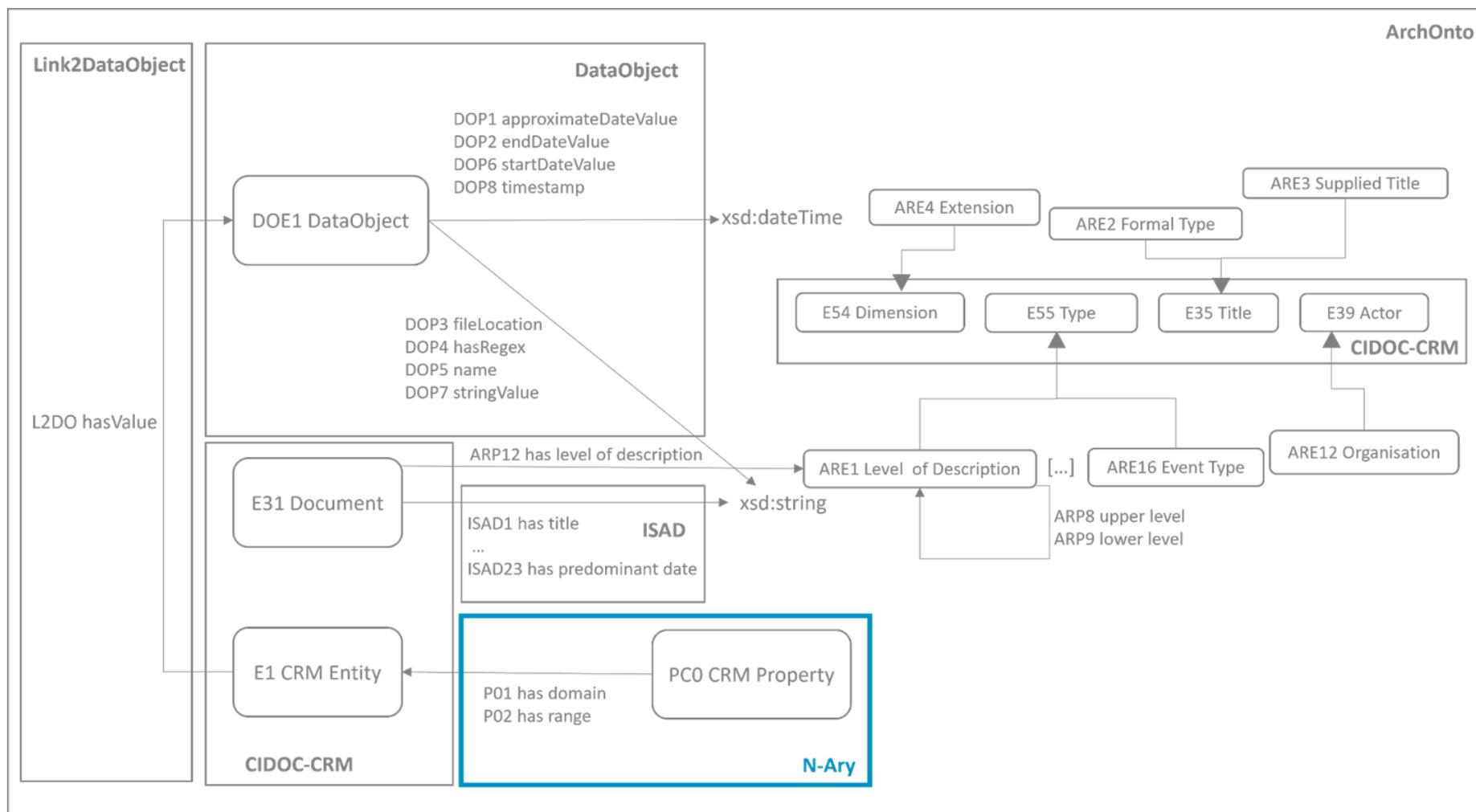
ArchOnto - CIDOC CRM



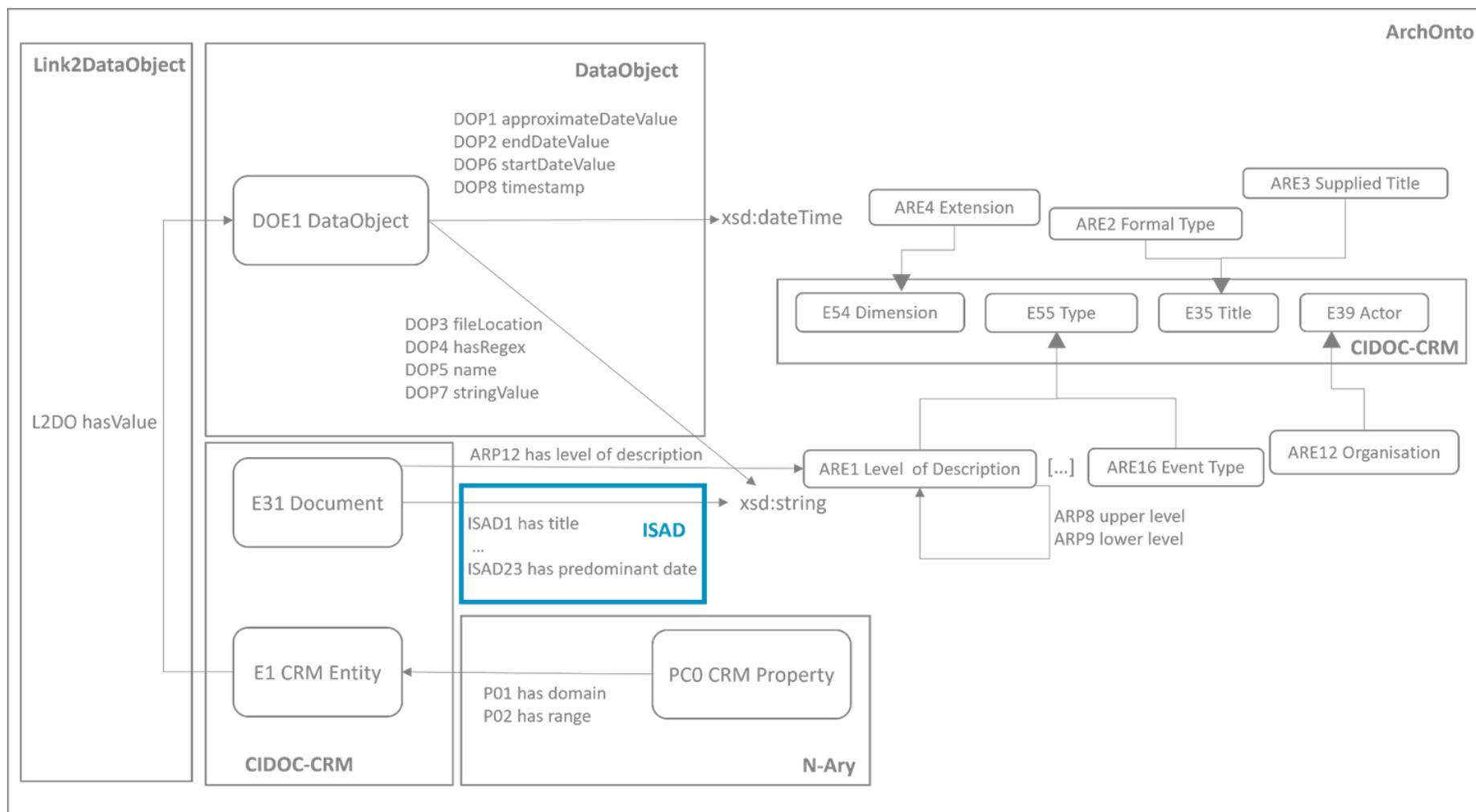
ArchOnto - DataObject



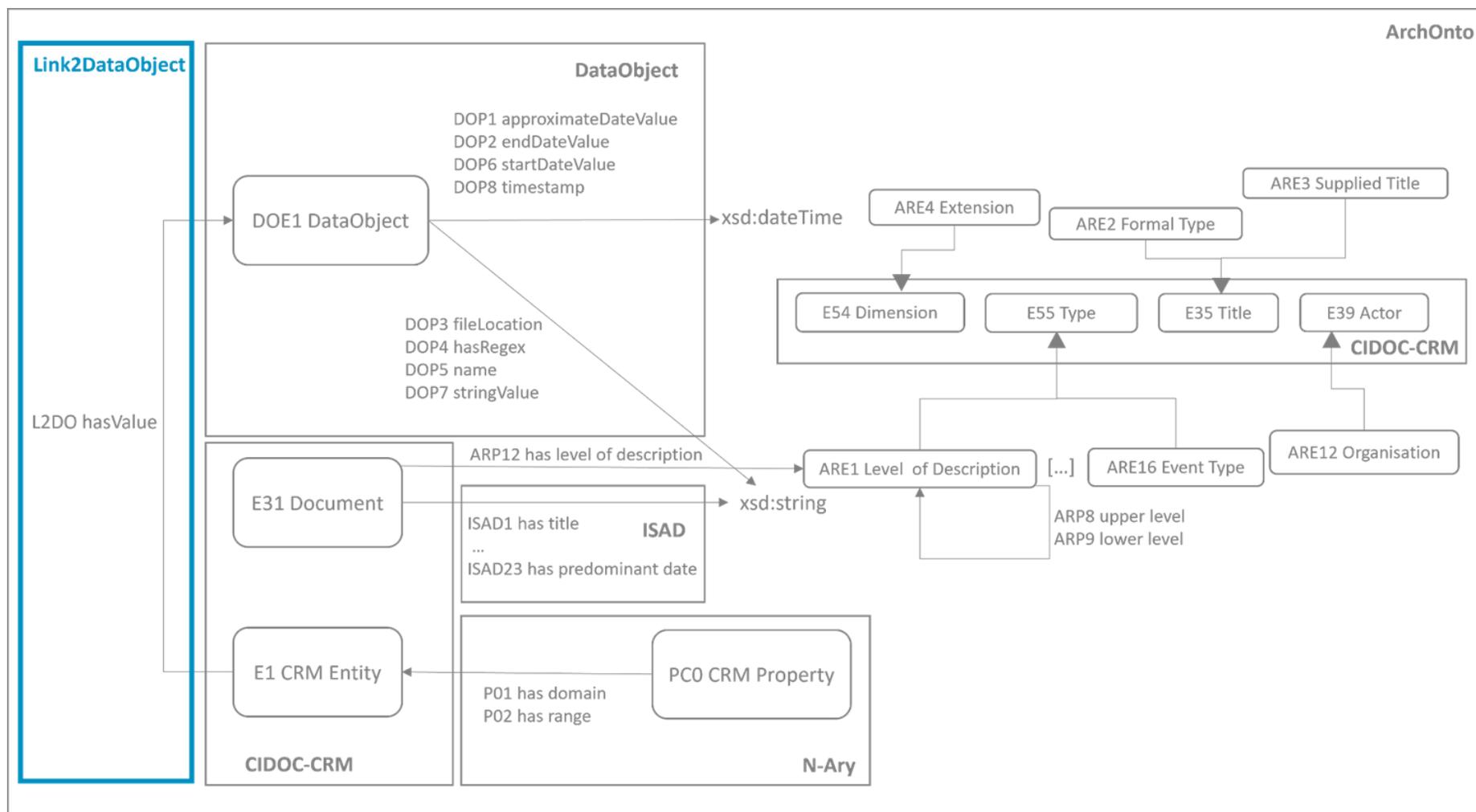
ArchOnto - N-ary



ArchOnto - ISAD Ontology



ArchOnto - Link2DataObject



EPISA Platform

An archival management system
based on linked data technologies

EPISA Platform

The EPISA Platform is the software infrastructure developed to support archival records management based on linked data technologies.

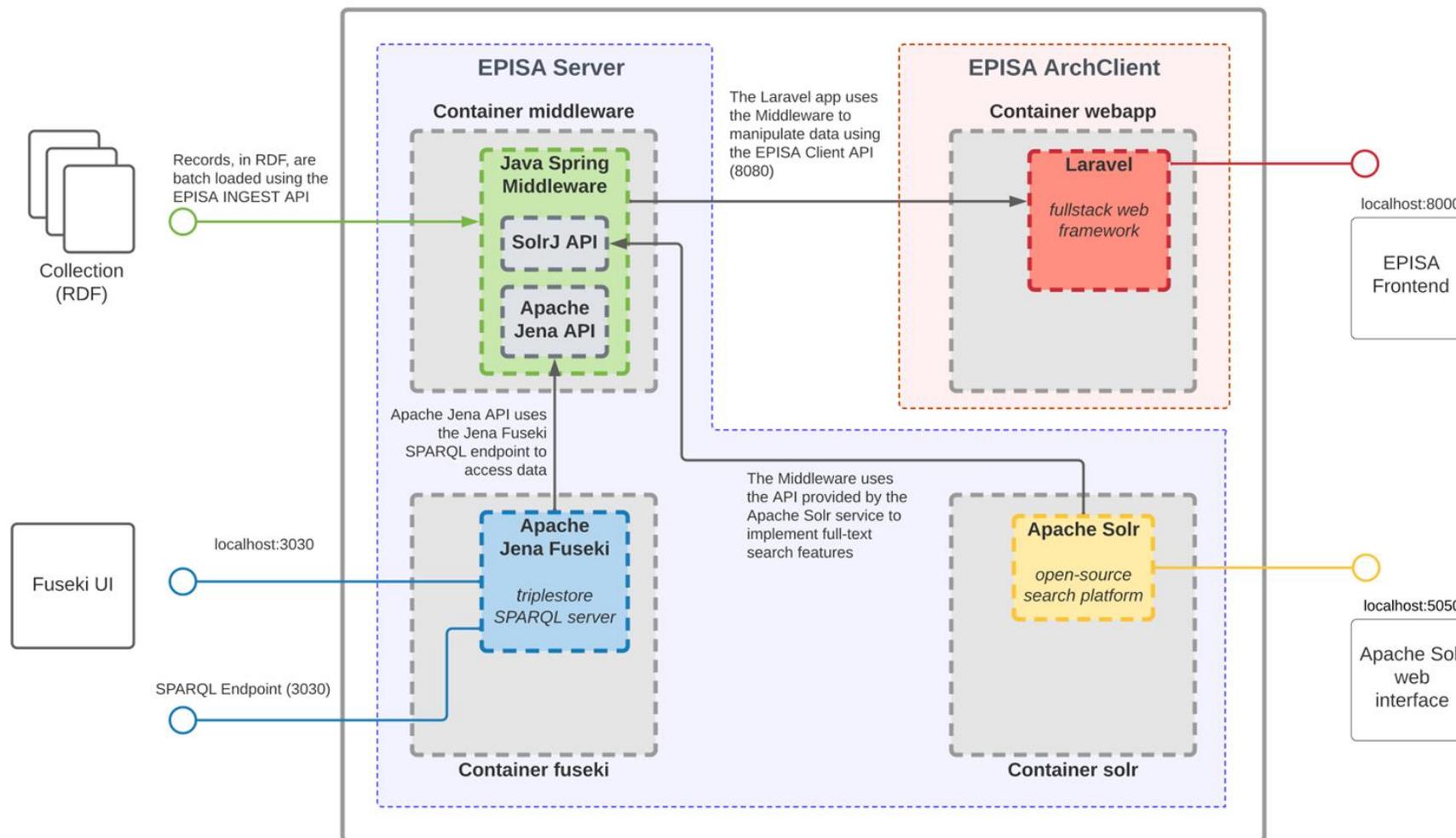
Supports storage, creation, search, access, and navigation over archival records and related entities.

Is based on open-source native linked data technologies.

Organized in two main components:

- **EPISA Server** is responsible for storing, reasoning, managing access, and providing an effective search mechanism over the archival data.
- **EPISA ArchClient** is a web application providing a graphical user interface for archivists to access, manage and describe collections of archival records.

EPISA Platform Docker Environment



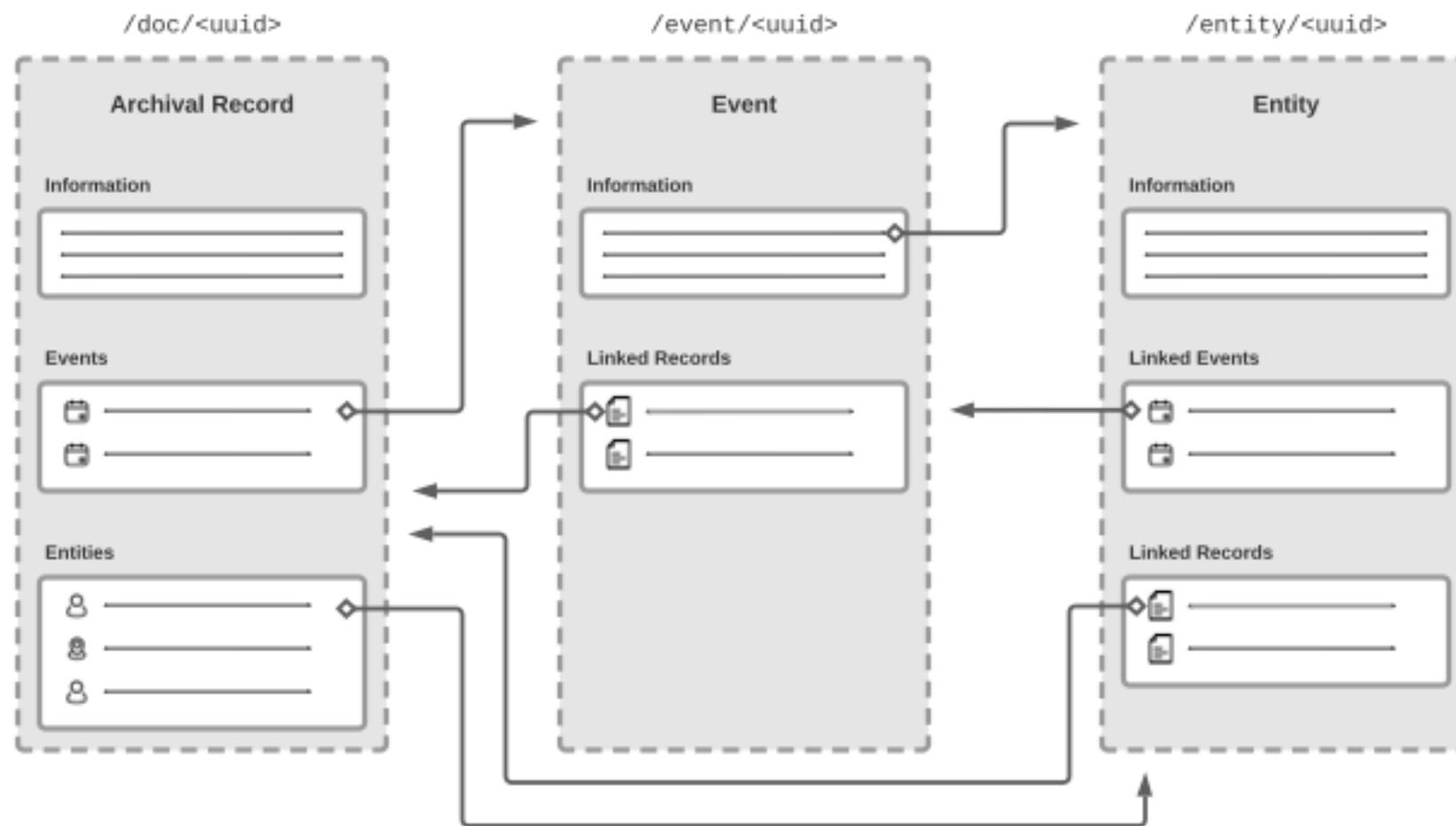
User Interface Abstractions

Higher-level abstractions need to be defined at the user interface level.

The abstractions defined are:

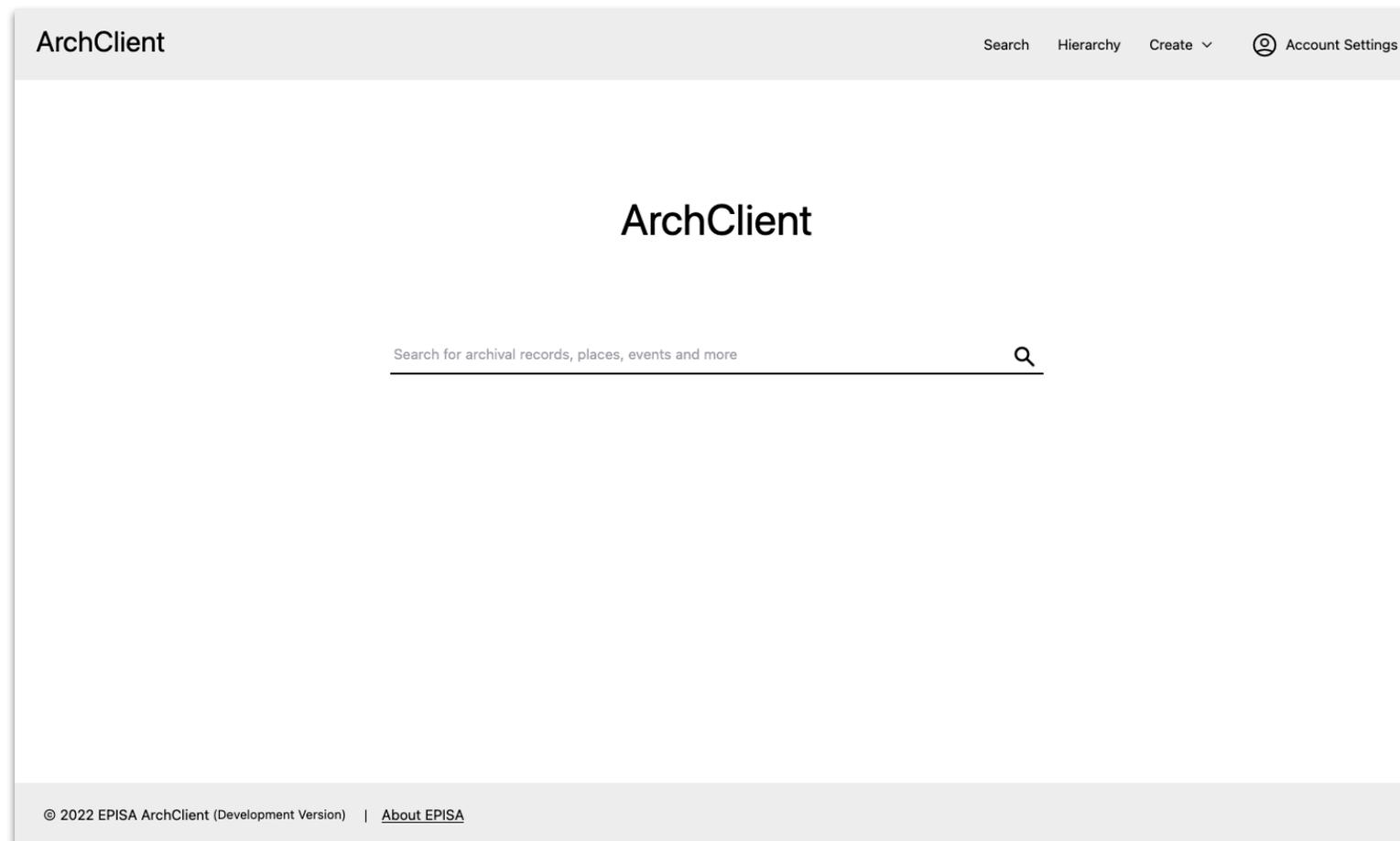
- **Records**, representing the collection of archival documents (e.g., fonds, collections, records).
- **Entities**, representing concepts (i.e., persons, places, organizations) mentioned in the records.
- **Events**, allowing the creation of complex structures linking records and entities — e.g., birth and death events, marriages, places of domicile.

Central Information Concepts at the UI Level





EPISA ArchClient Homepage



Hierarchical Navigation

ArchClient
Search Hierarchy Create ▾  Account Settings

Hierarchical Navigation

- > *it* "Um aquário de 2,40x1,00 m"

- > *fi* Inventário por óbito de Manuel Joaquim Ferreira Lino

- > *fo* Mosteiro de São Bernardo de Portalegre

- > *iu* Registo de passaportes deferidos

- ▼ *fo* Juízo da Índia e Mina
 - > *fi* Autos cíveis de petição e justificação para protesto de avaria grossa e rectificação do mesmo em que é autor Acácio José Correia,...

 - > *fi* Acção cível de fretes em que é autor António Henriques Leal e réu António Esteves Costa

 - > *it* Extrato do jornal a bordo do navio sueco "Joanes de Gotemburgo"

 - > *it* Procuração de Vasco Salter Mascarenhas e Mendonça dando poderes a João Manuel da Costa

 - > *sc* Cartório do escrivão Lino José de Almeida Lobo da Torre do Vale

 - ▼ *sf* Juízo das Justificações Ultramarinas
 - ▼ *fi* Autos de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de...

This document does not have descendants

 - > *iu* Maço 5

 - > *sr* Oriente

- > *fo* Desembargo do Paço

- > *fo* Mosteiro de São Bernardo de Portalegre

Search by Keywords (archival records results)

ArchClient
Search Hierarchy Create v Account Settings

2 results in search for "carta"

Archival Records (2)
Entities (0)

it

Title: Carta náutica, de João Teixeira Albernaz I

Identifier: PT/TT/CRT/198

it

Title: Carta de Antão Santos da Cunha para o Sub-secretário de Estado

Identifier: PT/ADCTB/PSS/APASC/00001

© 2022 EPISA ArchClient (Development Version) | [About EPISA](#)

Search by Keywords (entities results)

ArchClient
Search Hierarchy Create v Account Settings

🔍

6 results in search for "Maria"

Archival Records (1)
Entities (5)

👤 Person
Name: [Maria Emilia Cabral Da Silva](#)

👤 Person
Name: [Maria Luísa Barroso Da Câmara](#)

👤 Person
Name: [Maria De Brito Rocha E Melo](#)

👤 Person
Name: [Ludgero Maria de Lima e Quina](#)

👤 Person
Name: [Maria Luísa De Brito Rocha E Melo](#)

© 2022 EPISA ArchClient (Development Version) | [About EPISA](#)

Single Record Description

ArchClient
Search Hierarchy Create v
 Account Settings

fo Juízo da Índia e Mina

sf Juízo das Justificações...

ff Autos de sentença de José...

Autos de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de Manuel Leal e de Rita Mariana, naturais e moradores no distrito de Vila da Praia, Ilha Terceira (Açores)

<http://www.episa.inesctec.pt/archonto/#document7>

Record Description

References

Linked Data

View Source

Description Level ^

Description Level	File
-------------------	------

Identifiers ^

Identifier Type	Identifier
ReferenceCode	PT/TT/JIM-JJU/003/0005/00008
Identifier Type	Identifier
PhysicalLocation	Feitos Findos, Juízo da Índia e Mina, Justificações Ultramarinas, Ilhas, mç. 5, n.º 8
Identifier Type	Identifier
PreviousLocation	Feitos Findos, Fundos Geral, Letra J mç 4158

Titles ^

Title Type	Title
NamedIndividual	Autos de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de Manuel Leal e de Rita Mariana, naturais e moradores no distrito

Edit

Single Record References

ArchClient
Search Hierarchy Create v
 Account Settings

fo Juízo da Índia e Mina

sf Juízo das Justificações...

fi Autos de sentença de José...

Autos de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de Manuel Leal e de Rita Mariana, naturais e moradores no distrito de Vila da Praia, Ilha Terceira (Açores)

<http://www.episa.inesctec.pt/archonto/#document7>

Edit

Record Description
References
Linked Data
View Source

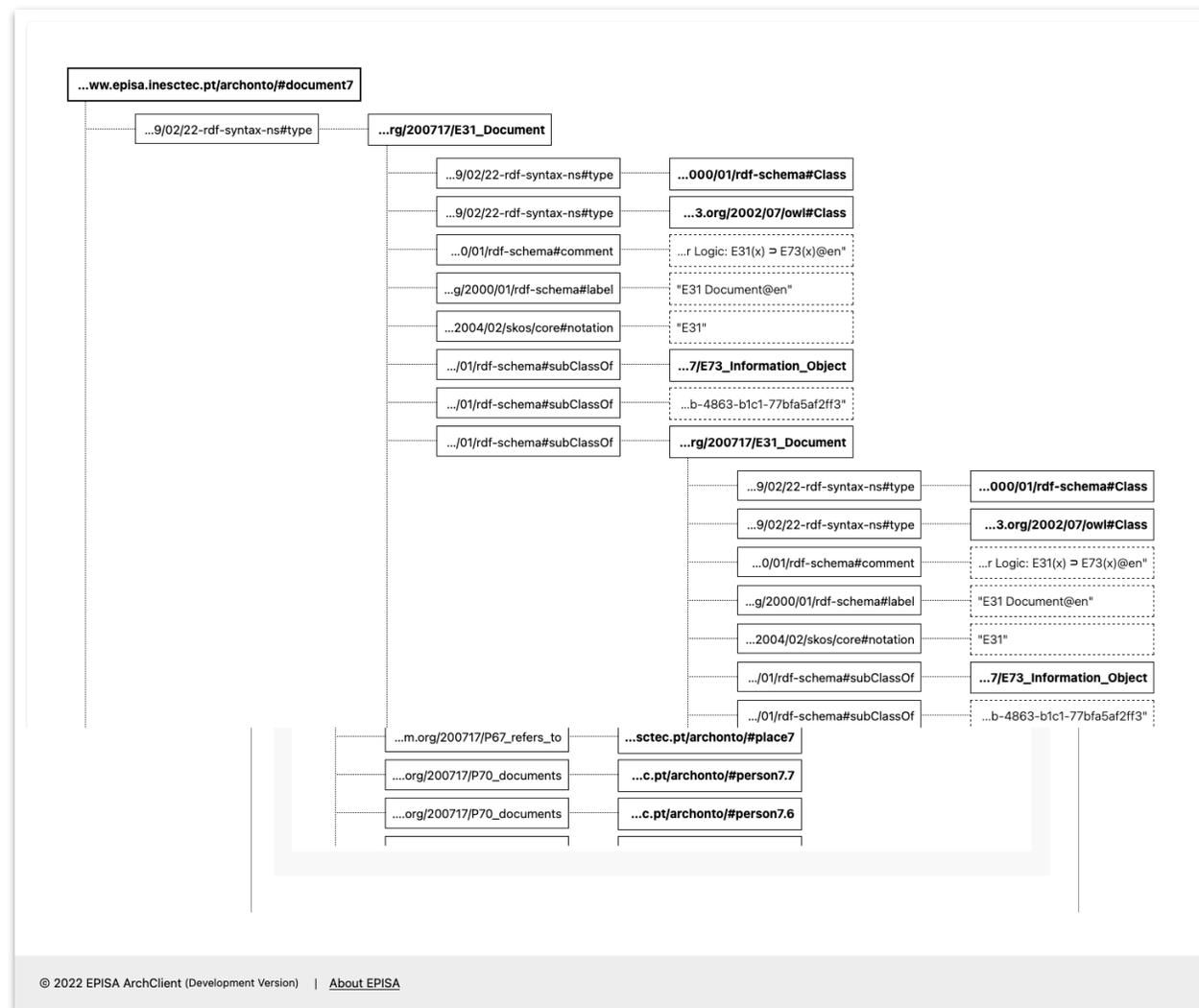
Add new reference ^

Type * Select Type v

Existing references ^

Relationship	Target Node	
refers to	Ilha Terceira	×
documents	Rita Mariana	×
documents	Manuel Leal	×
documents	José Borges Leal	×
documents	Dionísio José	×
documents	Jacinto Borges Leal	×
documents	Mariana de Jesus	×

View Record ArchOnto Source



View Entity of Person Type

ArchClient
Search Hierarchy Create ▼ Account Settings

👤

Jerónima da Cruz

<http://www.episa.inesctec.pt/archonto/#person21> 🔗

Entity Description
Linked Data
View Source

Linked Documents ^

Relationship	Document Name
refers to	Processo de Jerónima da Cruz
refers to	Culpas de Jerónima da Cruz mulher estrangeira do Reino de Castela, pobre, presa no cárcere desta Santa Inquisição de Évora e é cristã-velha

Linked Entities ^

Relationship	Target Node
has current or former residence	Évora
was born	Untitled Event
was present at	Untitled Event
was brought into existence by	Untitled Event
was present at	Untitled Event
participated in	Untitled Event

Development Challenges

- Limited maturity of linked data technologies, i.e. no standard off-the-shelf solutions.
- Lack of tools and resources focused on linked data authoring platforms.
- Lack of examples or guidelines for user interaction in authoring linked data.
- Complex underlying linked data model.

ENTITY AND PROPERTY INFERENCE FOR SEMANTIC ARCHIVES

55th CIDOC CRM and 48th FRBR CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
Sérgio Nunes

Follow us:

<http://episa.inesctec.pt>

<https://purl.org/episa/archonto>

