Anna Clements | euroCRIS & University of St Andrews
@AnnaCClements | ORCID: 0000-0003-2895-1310
Dom Fripp | Jisc
@Domicus | ORCID: 0000-0001-5352-4666
Jan Dvorak | euroCRIS & Czech Technical University
ORCID: 0000-0001-8485-152X

Research Data Shared Service (RDSS)

CRIS2018

Content

- Introducing Jisc
- Jisc Research Data Services Context
- RDSS context and vision
- RDSS timeline and progress
- RDSS at St Andrews
- Data Model & CERIFication Project

Link on https://www.jisc.ac.uk/rd/projects/research-data-shared-service

Introducing Jisc

Jisc is the UK **higher, further education** and **skills** sectors' **not-for-profit** organisation for **digital services** and **solutions**

Jisc

We…

Provide trusted **advice** and **practical assistance** for universities, colleges and learning providers

Operate **shared digital infrastructure** and **services**

Negotiate **sector-wide deals and conditions** with IT vendors and commercial publishers

RDSS: How and Why?

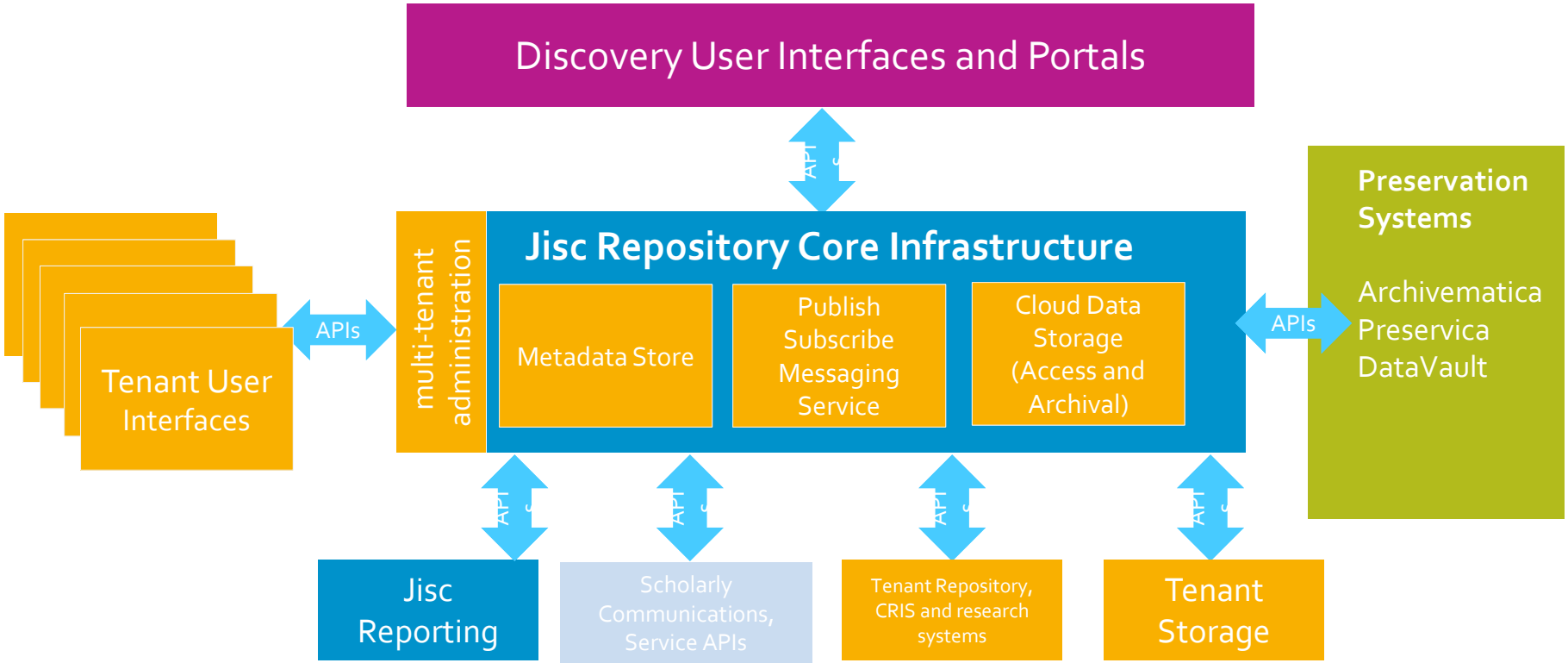
- Drivers
- More than £5 million investment over 2 years
- Open access

- Sector defined requirements
  - "R@R" co-design
  - Over half the HEI sector involved

RDSS: The Challenge

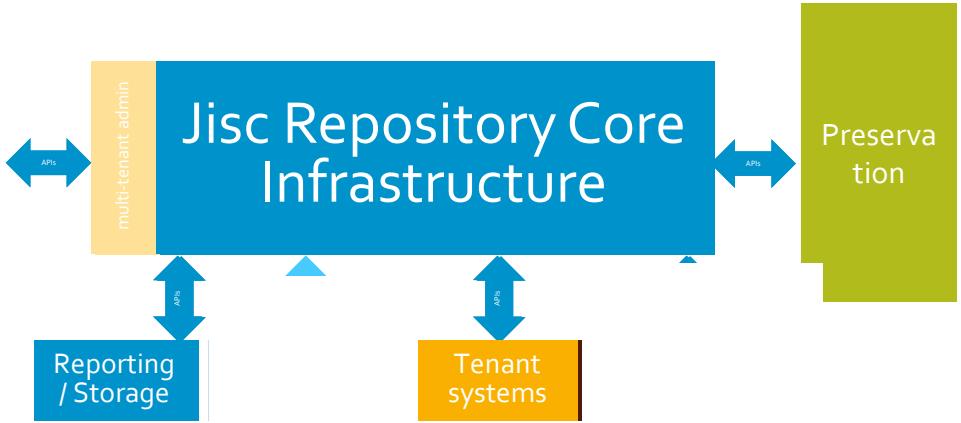
- To meet the sector's requirements we need
  - **True multi tenant**
  - **Multi content**
  - **Shared data model**
  - Interoperable
  - Sustainable
  - Much improved user experience

RDSS: The Vision

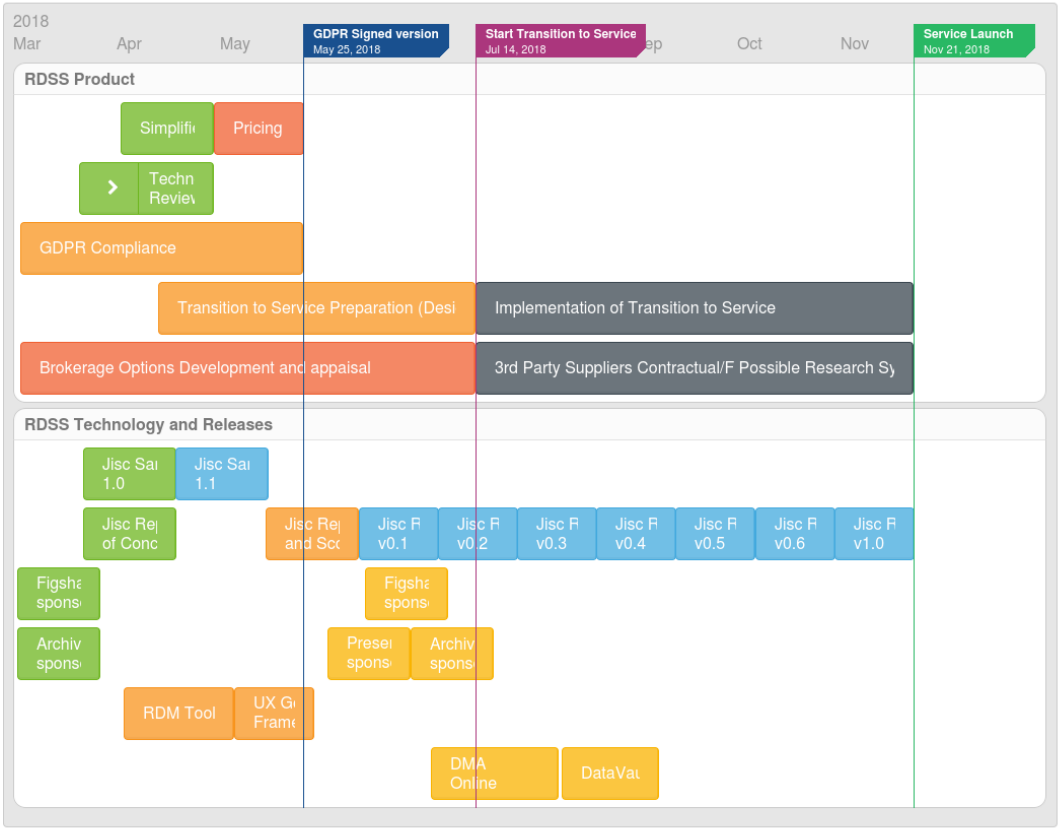


RDSS: Where we are now

- Core Architecture
  - multitenant database
  - Interoperability layer
- Data model
- Proof of concept front end API
- Initial Front end design

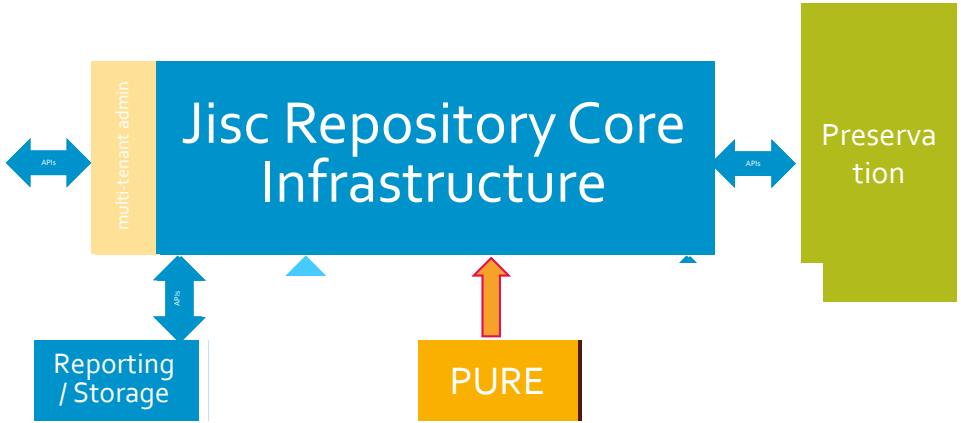


RDSS: Roadmap



RDSS: St Andrews pilot institution

- Pure is our data catalogue & repository
- RDSS pulling metadata and files from API into preservation systems
- We are trialling both Archivematica & Preservica
- Nothing passed back to Pure … yet

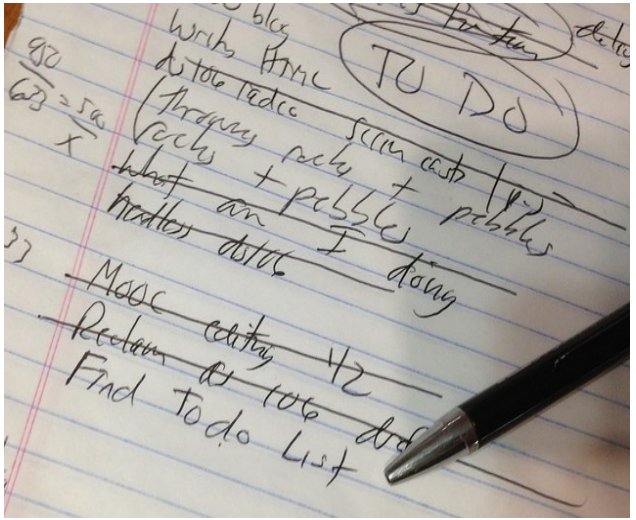


Goal for St Andrews: sustainable digital preservation

- Integrate with our existing systems particularly, Pure - to keep single interface for researchers and rekeying of metadata and transfer of data to a minimum
- Provide a preservation platform/service - integrated with Pure; two-way – preservation status back into Pure
- Solution that is flexible e.g. loosely coupled integrations - based on standards, to ensure we can swap systems in/out easily
- Solution that works for other digital content e.g. university records, building plans, e-theses, digitised special collections

RDSS: Priorities

- First priority is research data
  - Research output (Article/Thesis etc.)
  - Research data
  - Research software/code
  - Provenance metadata (method)
- But also…..
  - Preservation systems tailored for multiple digital objects and data types
  - Use cases and pilots for objects beyond research data



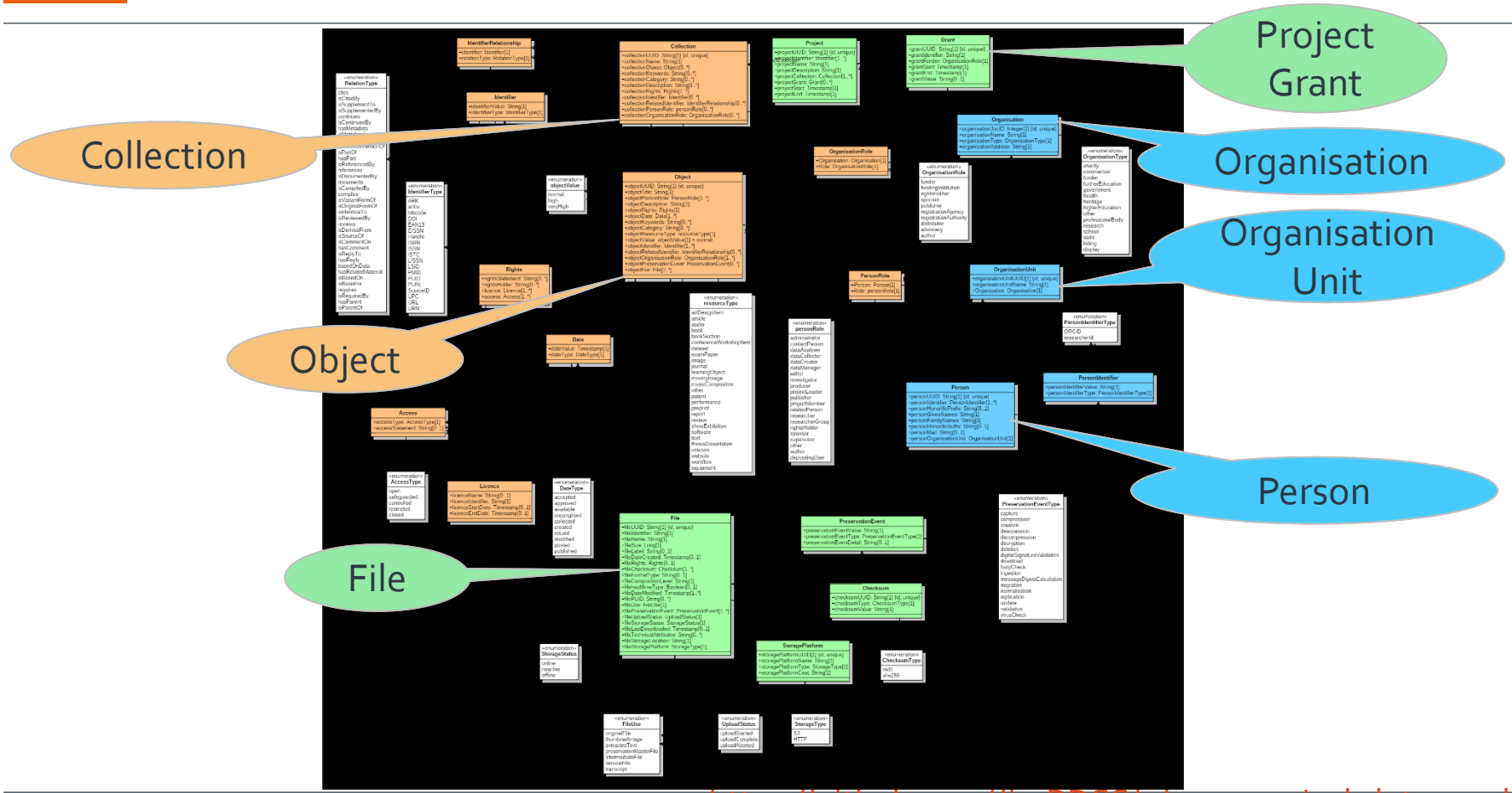
RDSS: Modelling interoperability

RDSS: CERIFication project

Project goals

- RDSS logical data model mapped to CERIF logical data model, including full documentation.
- Specific use cases of RDSS, or related services, mapped to CERIF-XML and accompanying guidelines for use.
- CERIF model feedback to euroCRIS and consideration in enhancements to the standard.
- Engagement via workshop/webinar(s) to disseminate outcomes from project.

RDSS: Canonical Data Model



https://github.com/JiscRDSS/rdss-canonical-data-model

RDSS: Data Model CERIFication

RDSS class	CERIF entity
Person	Person
Organisation	OrgUnit
Project	Project
Grant	Funding
Collection	Service
Object	ResultPublication ResultProduct ResultPatent Event Equipment
File	Medium

Mapping example - CDM to CERIF

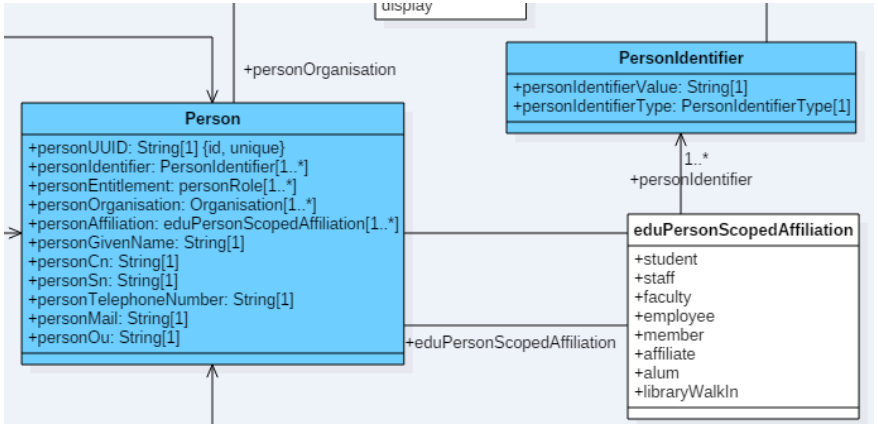
RDSS canonical data model version 2.0 mapping to CERIF

2018-01-24 Czech Technical University

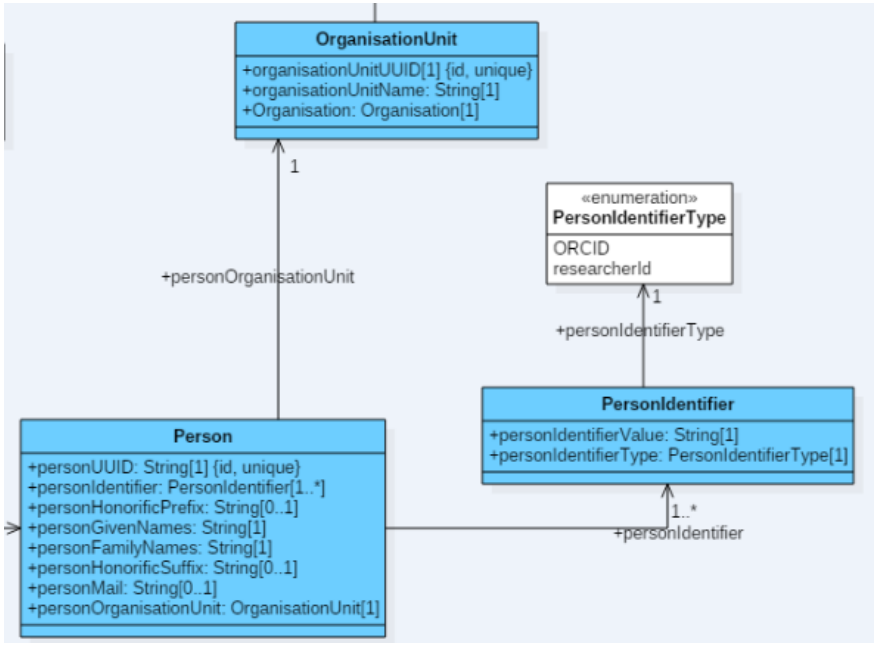
RDSS model element		-->	*CERIF model element*		*Notes*
Class	**Object**		**Entity**	**ResultProduct**	for the following resourceTypes: artDesignItem, audio, dataset, image, movingImage, musicComposition, software, website, workflow, unknown, other (if not textual)
Property	objectUUID		Attribute	ResultProductId	
Property	objectTitle		ML Attribute	ResultProduct.Name	multilingual; language defaults to English
Property	objectPersonRole : PersonRole [1..*]		Foreign key	ResultProduct_Person.ResultProductId	link; the type of the relationship is expressed using the RDSS Object Collection Person Roles class scheme (based on the personRole enumeration)
Property	objectDescription		ML Attribute	ResultProduct.Description	multilingual; language defaults to English
Property	objectRights : Rights [1..*]		Foreign key	ResultProduct_Class.ResultProductId	probably a unary classification if rights sets can be somehow pre-packaged
Property	objectDate : Date [1..*]		Foreign key	ResultProduct_Class.ResultProductId	unary classifications of ResultPublication by status; use the startDate and endDate to indicate validity
Property	objectKeywords : String [0..*]		ML Attribute	ResultProduct.Keywords	multilingual; language defaults to English; join them all in a semicolon separated list
Property	objectCategory : String [0..*]		Foreign key	ResultProduct_Class.ResultProductId	unary classifications by subject
Property	objectResourceType : resourceType		Foreign key	ResultProduct_Class.ResultProductId	unary classifications of ResultPublication using the RDSS Resource Type class scheme
Property	objectValue : objectValue		Foreign key	ResultProduct_Class.ResultProductId	unary classifications of ResultPublication using the RDSS Object Values class scheme
Property	objectIdentifier : Identifier [1..*] without relationType		Entity	FederatedIdentifier (fixed to this ResultProduct)	identifier type expressed by classifying the FederatedIdentifier using the RDSS Identifier Types class scheme
Property	objectRelatedIdentifier : Identifier [0..*] with relationType				?
Property	objectOrganisationRole : OrganisationRole [1..*]		Foreign key	ResultProduct_OrgUnit.ResultProductId	link; the type of the relationship is expressed using the RDSS Object Collection Organisation Roles class scheme (based on the OrganisationRole enumeration)
Property	objectPreservationEvent : PreservationEvent [0..*]		*Foreign key*	*ResultProduct_Event.ResultProductId*	? - if full detail needed use Event; if one needs to expess just the states use a unary classification with startDate & endDate
Property	objectFile : File[0..*]		Foreign key	ResultProduct_Medium	link; the type of the relationship is "File"

An example of data model economy

CDM 2.0.0

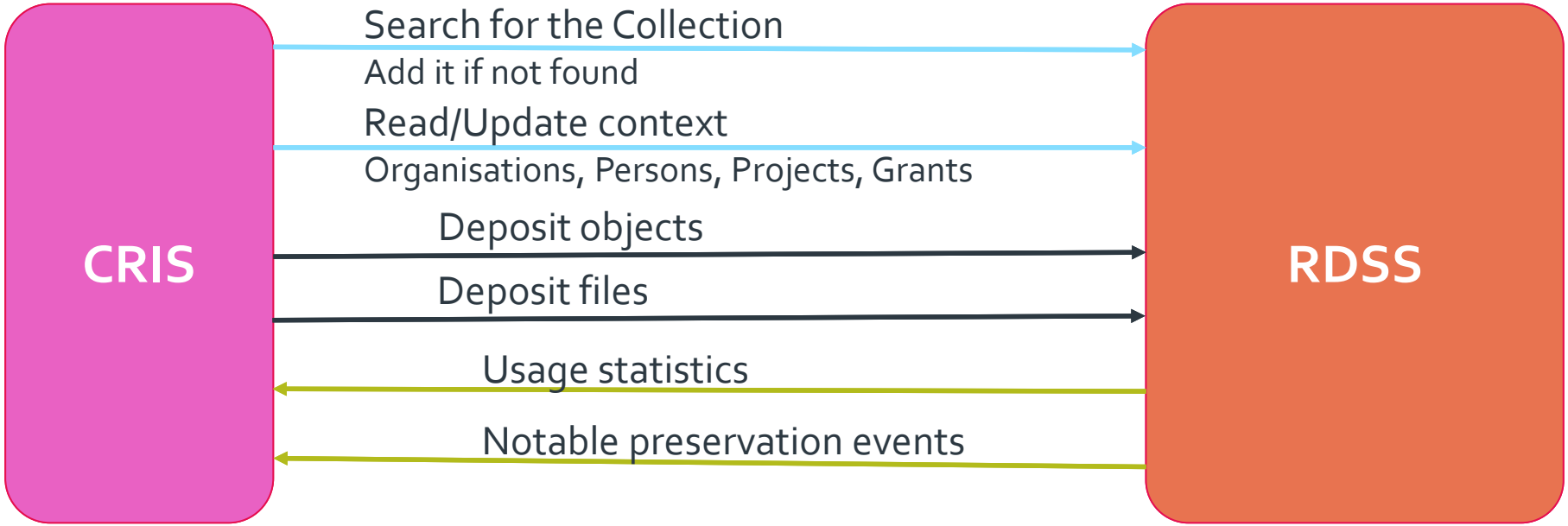


CDM 3.0.0



Alignment with CERIF resulted in handling session (authentication) metadata elsewhere and keeping the CDM about core metadata fields to aid interoperability with CRISes.

RDSS Information Interchange Use Cases



Thank you!

Anna Clements akc@st-andrews.ac.uk @annakclements
Assistant Director Library Services (Digital Research), University of St Andrews

Jan Dvorak jan.dvorak.2@cvut.cz
CRIS specialist, Computing and Information Centre, Czech Technical University

Dom Fripp Dom.Fripp@jisc.ac.uk @Domicus
Senior curation metadata developer, Jisc