University of St Andrews

University Library

Implementation of the OpenAIRE CERIF-XML guidelines in Pure

a marriage made in heaven …

Anna Clements | euroCRIS Board & University of St Andrew, UK @AnnaKClements | ORCID: 0000-0003-2895-1310

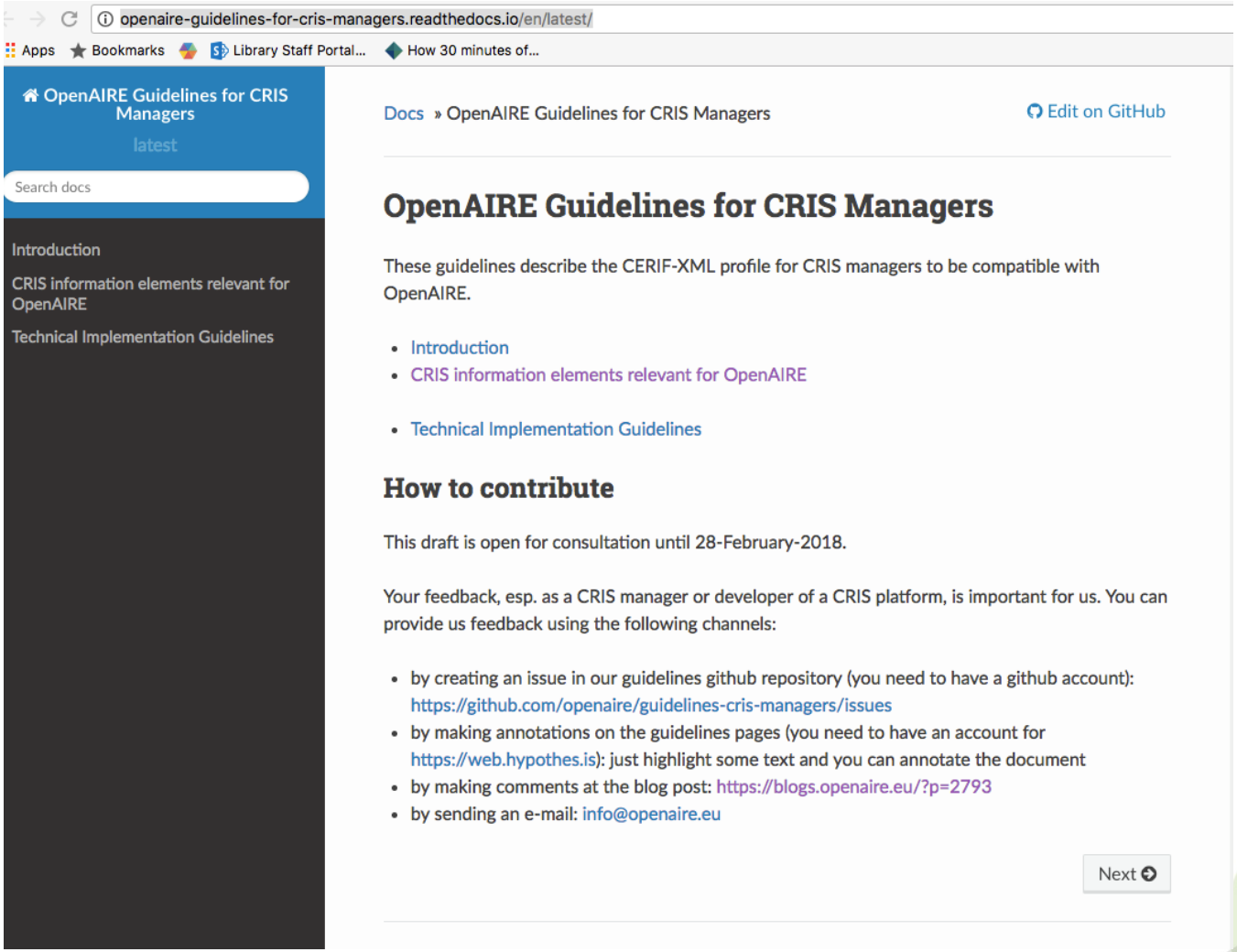
Alex Kujath | Product Manager - Pure, Elsevier

Jan Dvorak | CERIF TG, euroCRIS & Czech Technical University

University of St Andrews

University Library

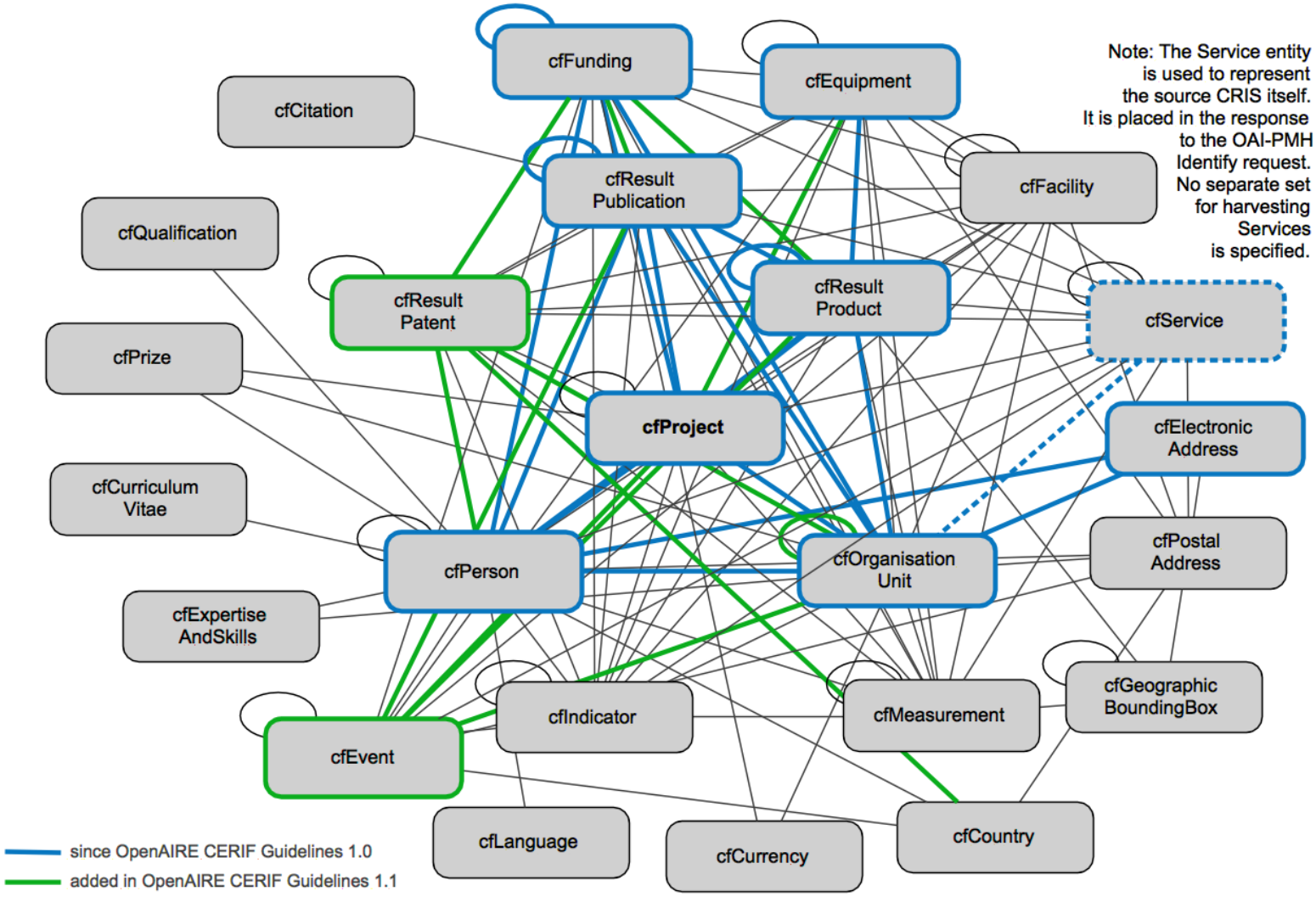
OpenAIRE CERIF-XML guidelines



University of St Andrews

University Library

CERIF elements



University of St Andrews

University Library

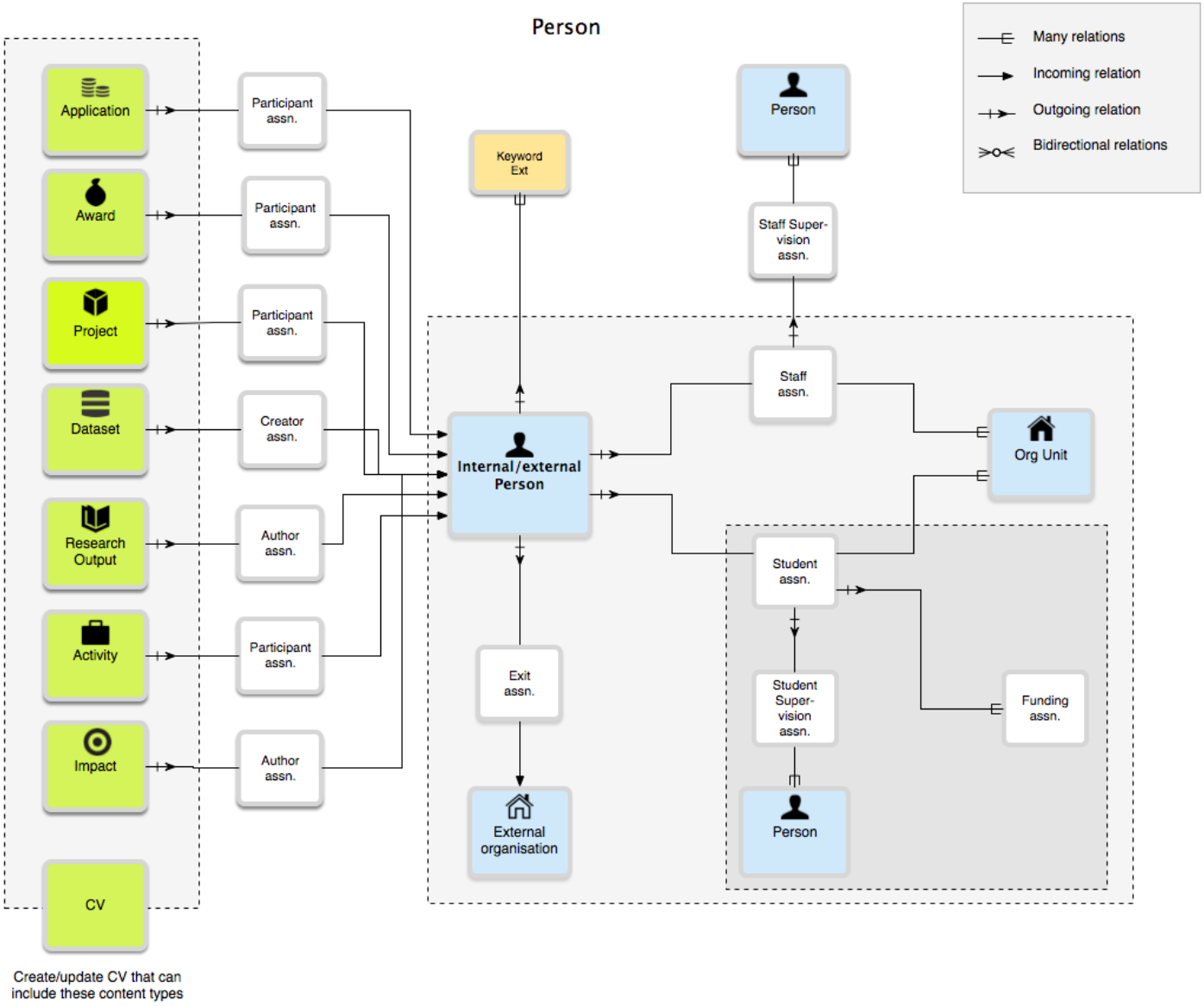
Developing specification: mapping content types

OpenAire CERF-XML
Publications
Products
Patent
Person
OrgUnit
Project
Funding
Service
Equipment
Event

University of St Andrews

University Library

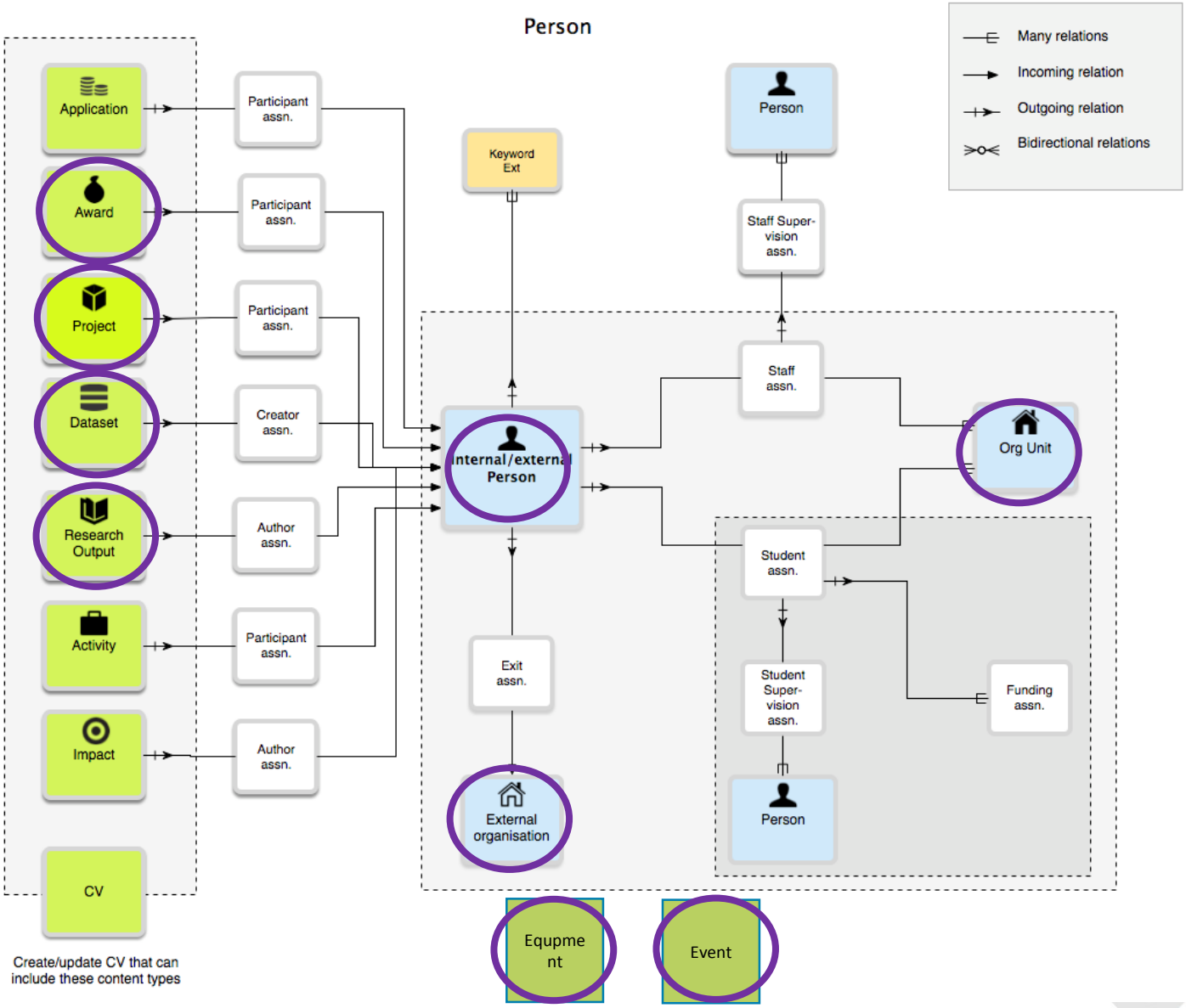
Pure data model



University of St Andrews

University Library

Pure data model



University of St Andrews

University Library

Developing specification: mapping content types

OpenAire CERF-XML	Pure 5.11 - suggested content types
Publications	Research Output
Products	Dataset
Patent	Template within Research Output
Person	Person, External person
OrgUnit	Organisational unit, External organisation, Publisher
Project	Project
Funding	Award
Service	??
Equipment	Equipment/facility
Event	Event

University of St Andrews

University Library

UK Pure User Group

Meet three times a year

How many members – 40+

Special interest groups:
- REF2021
- Repositories
- Research data
- Interoperability
- Theses

Short-life working groups:
- GDPR
- OpenAire

User-led

OpenAIRE Short Life Working Group

Aim : Provide mapping of Pure data model to OpenAire CERIF-XML guidelines and for CRIS Managers

Elsevier have committed to implementing these guidelines in PURE for the October 2018 (v 5.13). To meet this deadline, Elsevier have asked for input from euroCRIS and Pure Customers to help develop the specification mapping the data held in Pure to the new OpenAire CERIF-XML guidelines. A short life working group (SLWG) will be setup to deliver this specification.

SLWG Members

- Anna Clements (St Andrews, Chair)
- Federica Fina (St Andrews)
- Patricia Herterich (Birmingham)
- Kara Jones (Bath)
- Nykhola Strong (Aberdeen)
- Anne Thorst Melbye (South Denmark)
- Bill Worthington (Hertfordshire)
- Jan Dvorak (euroCRIS & Czech Technical University)
- Alex Kujath (Elsevier)

University of St Andrews

University Library

Project timeline 2018

Feb	Pure dev version made available [5.11] with access to API, accompanying documentation and complete dataset for OpenAire mapping	Elsevier /Alex Kujath (AK)
Feb	Provide spreadsheet format of OpenAire CERIF-XML guidelines	euroCRIS/Jan Dvorak (JD)
Mar	Establish SLWG	euroCRIS/Anna Clements (AC)
Apr - Jun	Develop specification	SLWG
July	Implement specification	Elsevier
Aug	Feedback on implementation	SLWG
Sep	Test/Mature product	Elsevier
Oct	Release 5.13	Elsevier

University of St Andrews

University Library

Developing specification: use of Pure web service

Pure API 511

This is the Pure Web Service. Listed below are all available endpoints, along with a short description.

In order to use the Pure Web Service, you must enter an API key. These are generated in the Administrator tab of Pure, and issued with a given set of available endpoints.
To enter your API key and begin your use, press the 'Authorize' button to at the top of this page.
You are then presented with two options for entering the API key: Use the API key in query format.Use the API key in a header.
For further documentation, see API Documentation.

A new version of the API is released with each major version of Pure, and remains available for one year. This version is no longer available in Pure 5.15.
The old web service is deprecated, but still available here, and it will no longer be available in Pure 5.13.

Created by Elsevier

activities	Show/Hide	List Operations	Expand Operations
applications	Show/Hide	List Operations	Expand Operations
author-collaborations	Show/Hide	List Operations	Expand Operations
awards	Show/Hide	List Operations	Expand Operations
changes	Show/Hide	List Operations	Expand Operations
classification-schemes	Show/Hide	List Operations	Expand Operations
concepts	Show/Hide	List Operations	Expand Operations
curricula-vitae	Show/Hide	List Operations	Expand Operations
datasets	Show/Hide	List Operations	Expand Operations

GET	/datasets	Lists all datasets
POST	/datasets	Complex operation for datasets
GET	/datasets/{id}	Get dataset

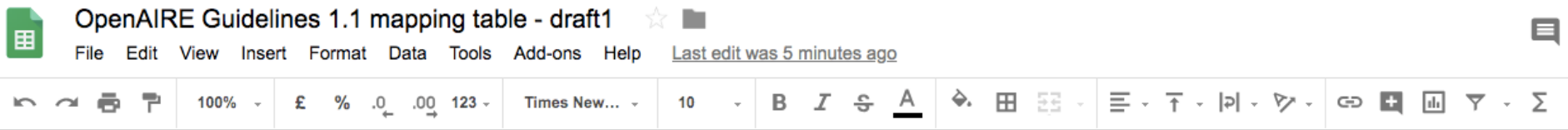
Implementation Notes
Get dataset with specific ID (path parameter).

Response Content Type application/xml

Parameters

Parameter	Value	Description	Parameter Type	Data Type
id	(required)	ID of the desired dataset	path	string

Developing specification: google sheets : orgunit



	A	B	C	D	E
1	**This is a draft table for ma**	**pping your CRIS onto the OpenAIRE Guidelines for CRIS Managers 1.1 CERIF profile**			
2	**The XML Schema**	**for the OpenAIRE CERIF profile 1.1**			
3	**XPath** (purely structural XML elements in *italic*)	**Documentation**	**Cardinality**	**Populate from**	Mapping Notes
16	OrgUnit/@id		exactly one	UUID	
17	OrgUnit/Acronym	The acronym of the organisation unit	optional	/organisational-units/organisationalUnit/nameVariants type="Short name"	
18	OrgUnit/Name	The name of the organisation unit	optional, possibly repeated if in different languages	/external-organisations/externalOrganisation/name | /organisational-units/organisationalUnit/name	There is no postal address in the schema. In some external orgs in St Andrews Pure we have the full postal address in the name field. In others it is split out into separate address fields.
19	OrgUnit/Identifier	An identifier of the organisation unit	optional, possibly repeated	/external-organisations/externalOrganisation uuid | /external-organisations/pureId | /organisational-units/organisationalUnit/addresses/address id	Should be Ringgold or FundRef ID any other shared identifier. Ask Alex what plans are to add orgids to Pure. Alex: In pratical terms this looks like it will be the URL of the website of the organistions, see: https://github.com/openaire/guidelines-cris-managers/blob/master/samples/openaire_cerif_xml_example_orgunits.xml
20	OrgUnit/ElectronicAddress	An electronic address associated with the organisation unit	optional, possibly repeated	/external-organisations/externalOrganisation/email	For OrgUnit - email appears in UI but can't find in API
21	*OrgUnit/PartOf*	Chain up to the larger unit that encompasses this unit	optional	/organisational-units/organisationalUnit	Relationships can be seen for Organisational Units in the PURE UI, I couldn't find them exposed in the API. We also have to agree how much of the hierarcy should be included. For external organisations we don't have concept of hierarchy in Pure; often the full postal address is in the Name field. Sometimes this is split out into separate address fields .. but this is not a hierarchy - Update: PH found them now and added path accordinlgy

University of St Andrews

University Library

Publication/date

OpenAIRE Guidelines 1.1 mapping table - draft1

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

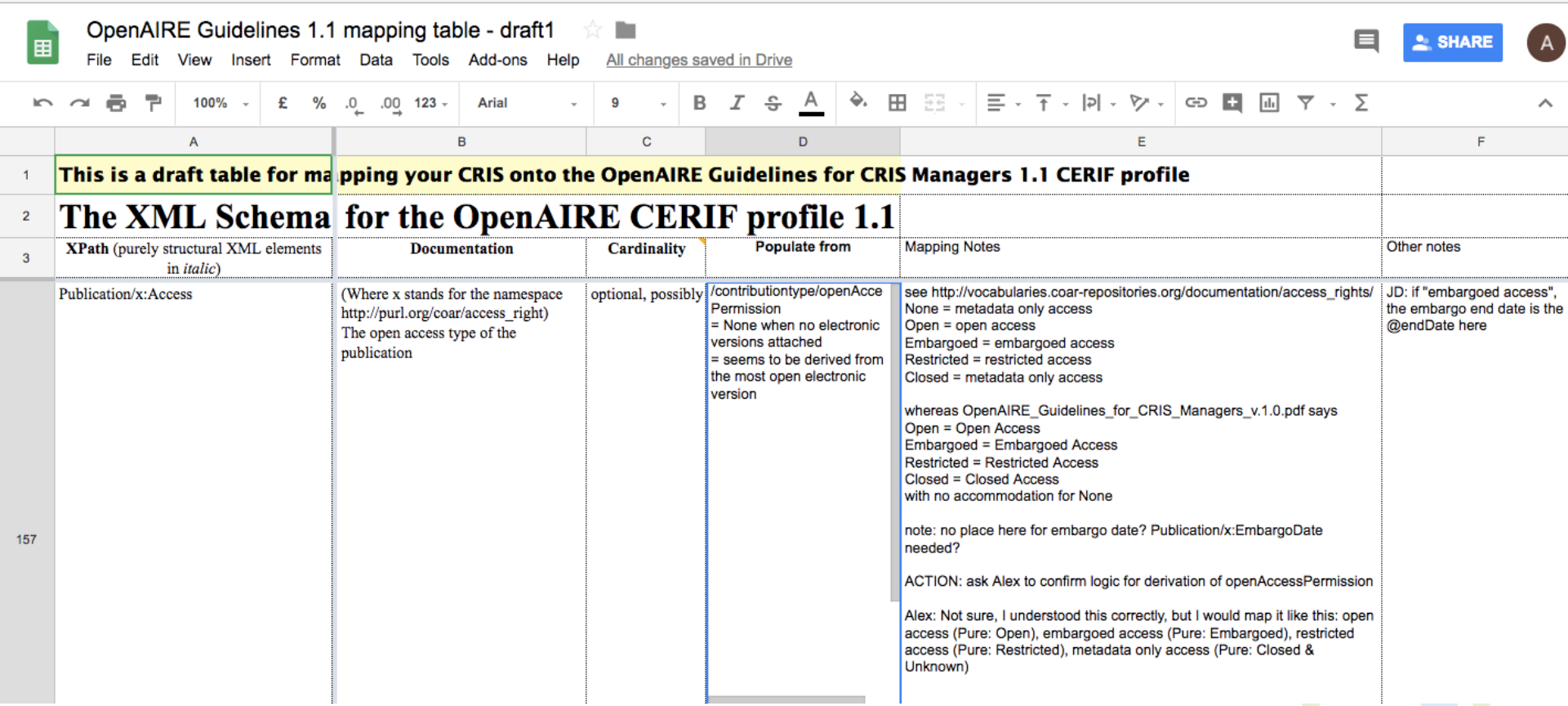
SHARE

	A	B	C	D	E
1	**This is a draft table for ma**	**pping your CRIS onto the OpenAIRE Guidelines for CRIS Managers 1.1 CERIF profile**			
2	**The XML Schema**	**for the OpenAIRE CERIF profile 1.1**			
3	**XPath** (purely structural XML elements in *italic*)	**Documentation**	**Cardinality**	**Populate from**	Mapping Notes
106	Publication/PublicationDate	The date the publication appeared	optional	/contributiontype/publicationStatuses/publicationStatus/publicationDate/year /contributiontype/publicationStatuses/publicationStatus/publicationDate/month /contributiontype/publicationStatuses/publicationStatus/publicationDate/day	we might only reveal this element if /publicationStatus = published or e-pub ahead of print? this works for published journal articles but what about other cases? e.g. pre-prints, post-prints where pure or a local repository is the publication medium? should we look to the visiblity/'available on date' of an open pre-publication manuscript in the way REF visibility is calculated? (ugh!) **Agreed** just rely on same visibility rules as for API
107	Publication/Number	The number of the publication (e.g. Article Number)	optional	/contributiontype/articleNumber	
108	Publication/Volume	The volume of the publishing channel where this publication appeared	optional	/contributiontype/volume	
109	Publication/Issue	The issue of the publishing channel where this publication appeared	optional	/contributiontype/journalNumber	
110	Publication/Edition	The edition of the publication	optional		
111	Publication/StartPage	The page where this publication starts, in case the publishing channel or containing publication has numbered pages	optional	/contributiontype/pages the first integer when the content takes the form integer[separator-nonintegers]integer eg 100-101, 100 - 101, 100:101 etc.	
112	Publication/EndPage	The page where this publication ends, in case the publishing channel or containing publication has numbered pages	optional	/contributiontype/pages the second integer when the content takes the form integer[separator-nonintegers]integer eg 100-101, 100 - 101, 100:101 etc.	
113	Publication/DOI		optional	/contributiontype/electronicVersions/electronicVersion	

University of St Andrews

University Library

Publication/access



OpenAIRE Guidelines 1.1 mapping table - draft1

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

SHARE

	A	B	C	D	E	F
1	**This is a draft table for ma**	**pping your CRIS onto the OpenAIRE Guidelines for CRIS Managers 1.1 CERIF profile**				
2	**The XML Schema**	**for the OpenAIRE CERIF profile 1.1**				
3	**XPath** (purely structural XML elements in *italic*)	**Documentation**	**Cardinality**	**Populate from**	Mapping Notes	Other notes
157	Publication/x:Access	(Where x stands for the namespace http://purl.org/coar/access_right) The open access type of the publication	optional, possibly	/contributiontype/openAcce Permission = None when no electronic versions attached = seems to be derived from the most open electronic version	see http://vocabularies.coar-repositories.org/documentation/access_rights/ None = metadata only access Open = open access Embargoed = embargoed access Restricted = restricted access Closed = metadata only access whereas OpenAIRE_Guidelines_for_CRIS_Managers_v.1.0.pdf says Open = Open Access Embargoed = Embargoed Access Restricted = Restricted Access Closed = Closed Access with no accommodation for None note: no place here for embargo date? Publication/x:EmbargoDate needed? ACTION: ask Alex to confirm logic for derivation of openAccessPermission Alex: Not sure, I understood this correctly, but I would map it like this: open access (Pure: Open), embargoed access (Pure: Embargoed), restricted access (Pure: Restricted), metadata only access (Pure: Closed & Unknown)	JD: if "embargoed access", the embargo end date is the @endDate here

University of St Andrews

University Library

Publication/type mapping

Pure research outputs for Openaire SLWG.xlsx

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

SHARE

1	**Types of research output**			Versions			
2	Base URI: /dk/atira/pure/researchoutput/researchou			dk.atira.pure.core:core:5.11.0-1			
3	Classifications:			dk.atira.pure.templates:templates:5.11.0-1			
4				dk.atira.pure.base-uk:base:5.11.0-1			
5				dk.atira.pure.base-uk.cust-uk-unifiedprojectdemo:cust:5.11.0-1			
6							
7	PURE 5.11.0-1			COAR Resource Type Vocabulary Version 2.0			
8	Type	SubType		**Term**	**Code/URL**	**Definition**	
9	Contribution to journal						
10		Article	y	journal article	http://purl.org/coar/resource_type/c_6501	An article on a particular topic and published in a journal issue. (adapted from fabio)	
11		Meeting abstract	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
12		Letter	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
13		Comment/debate	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
14		Book/Film/Article review	y	review	http://purl.org/coar/resource_type/c_efa0	A review of others' published work.	
15		Literature review	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
16		Editorial	y	editorial	http://purl.org/coar/resource_type/c_b239	A brief essay expressing the opinion or position of the chief editor(s) of a (academic) journal with respect to a political, social, cultural, or professional issue. Adapted from ODLIS	
17		Special issue	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
18		Review article	y				
19		Short survey	y	contribution to journal	http://purl.org/coar/resource_type/c_3e5a	A contribution to a journal denotes a work published in a journal. If applicable sub-terms should be chosen.	
20		Conference article	y	journal article	http://purl.org/coar/resource_type/c_6501	An article on a particular topic and published in a journal issue. (adapted from fabio)	
21	Contribution to Book/Report types						
22		Chapter (peer-reviewed)	y	book part	http://purl.org/coar/resource_type/c_3248	A defined chapter or section of a book, usually with a separate title or number.	
23		Chapter	y	book part	http://purl.org/coar/resource_type/c_3249	A defined chapter or section of a book, usually with a separate title or number.	
24		Entry for encyclopedia/diction	y	book part	http://purl.org/coar/resource_type/c_3250	A defined chapter or section of a book, usually with a separate title or number.	
25		Conference contribution	y	conference paper	http://purl.org/coar/resource_type/c_5794	A conference paper that is submitted to a conference and presented to the audience. The conference paper is published in proceedings.	
26		Foreword/postscript	y	book part	http://purl.org/coar/resource_type/c_3252	A defined chapter or section of a book, usually with a separate title or number.	

University of St Andrews

Publication/type mapping: queries

University Library

34	Contribution to specialist publication types					
35		Article	?	periodical	http://purl.org/coar/resource_type/c_2659	A periodical is a publication issued on a regular and ongoing basis as a series of issues. (Adapted from fabio)
36		Featured article	?	periodical	http://purl.org/coar/resource_type/c_2660	A periodical is a publication issued on a regular and ongoing basis as a series of issues. (Adapted from fabio)
37		Book/Film/Article review	y	review	http://purl.org/coar/resource_type/c_efa0	A review of others' published work.
38		Editorial	y	editorial	http://purl.org/coar/resource_type/c_b239	A brief essay expressing the opinion or position of the chief editor(s) of a (academic) journal with respect to a political, social, cultural, or professional issue. Adapted from ODLIS
39		Letter	?	periodical	http://purl.org/coar/resource_type/c_2663	A periodical is a publication issued on a regular and ongoing basis as a series of issues. (Adapted from fabio)
40		Special issue	?	periodical	http://purl.org/coar/resource_type/c_2664	A periodical is a publication issued on a regular and ongoing basis as a series of issues. (Adapted from fabio)
49	Non-textual form types					
50		Software	y	software	http://purl.org/coar/resource_type/c_5ce6	A computer program in source code (text) or compiled form.
51		Data set/Database	y	dataset	http://purl.org/coar/resource_type/c_ddb1	A collection of related facts and data encoded in a defined structure. (adapted from fabio; DataCite)
52		Digital or Visual Products	y	image	http://purl.org/coar/resource_type/c_c513	A visual representation other than text, including all types of moving image and still image.
53		Web publication/site	y	website	http://purl.org/coar/resource_type/c_7ad9	A website, also written as web site or simply site, is a set of related web pages typically served from a single w domain. A website is hosted on at least one web server, accessible via a network such as the Internet or a privat area network through an Internet address known as a uniform resource locator (URL). All publicly accessible websites collectively constitute the World Wide Web.
54		Artefact	?			
55		Exhibition	?			
56		Performance	?			
57		Composition	?			
58		Design	?			
64	Memorandum/exposition					
65		Memorandum	y	memorandum	http://purl.org/coar/resource_type/c_18wz	It is a note, document or other communication that helps the memory by recording events or observations on a A memorandum can have only a certain number of formats; it may have a format specific to an office or institu
66		Question & Answer/hearing	?			
67	Contribution to memorandum/exposition					
68		Memorandum contribution	y	memorandum	http://purl.org/coar/resource_type/c_18wz	It is a note, document or other communication that helps the memory by recording events or observations on a A memorandum can have only a certain number of formats; it may have a format specific to an office or institu
69		Question & Answer/hearing co	?			

University of St Andrews

University Library

Implement specification

- Implementation starts with the mapping from Pure to CERIF XML
  - Create a consistent mapping, making sure all mandatory fields can always be provided
  - Complete with optional fields
- Provide a mapping in a form, that it can be read an examined by humans (google sheet)
- Mapping will be examined and checked for consistency
- Implementation according to documented mapping
- Feedback to SLWG with a Pure system exposing the first version
- Gather feedback and discuss
- Implement changes and improvements
- Release to all Pure customers, planned with Pure 5.13.0, October 2018

University of St Andrews

University Library

Implementation of the OpenAIRE CERIF-XML guidelines in Pure

a marriage made in heaven …

Anna Clements | euroCRIS Board & University of St Andrew, UK @AnnaKClements | ORCID: 0000-0003-2895-1310

Alex Kujath | Product Manager - Pure, Elsevier

Jan Dvorak | CERIF TG, euroCRIS & Czech Technical University