A CERIF description of an OA Policy to ease monitoring compliance

Table of contents

Authors: Keith Jeffery (Keith G Jeffery Consultants), Anne Asserson (Anne Asserson Consulting), Laurent REMY (IS4RI), Valérie BRASSE (IS4RI)

Introduction

PASTEUR4OA (Open Access Policy Alignment Strategies for European Union Research) aims to support the European Commission's Recommendation to Member States of July 2012 that they develop and implement policies to ensure Open Access (OA) to all outputs from publicly-funded research.

PASTEUR4OA is helping develop and/or reinforce open access strategies and policies at the national level and facilitate their coordination among all Member States. It is building a network of centres of expertise in Member States that will develop a coordinated and collaborative programme of activities in support of policymaking at the national level under the direction of project partners. It aims at documenting and analysing OA policies, and then at encouraging policies to strengthen the degree of openness and coverage.

In order to do this – and regarding the heterogeneity of policies and their evolution through time – it is necessary to have a canonical representation of each policy. Using CERIF to describe an OA policy requirements will thus facilitate the monitoring of the compliance to the policy, especially for institutions using a CERIF-based CRIS.

This document will first introduce a schema for Open Access policies that has been developed, and make a focus on its formalisation using the CERIF data model. A second part will describe the modules developed to build a demonstration system that includes OA policies expressed in CERIF.

I) Policy schema mapped to CERIF

1) Open Access policy schema

Introduction

The purpose of PASTEUR4OA is to be able to document and analyse OA policies and then to encourage policies to strengthen the degree of openness and coverage. In order to do this – and regarding the heterogeneity of policies and their evolution through time – it is necessary to have a canonical representation of each policy.

Jacobs N. (2014, December 1) presented in Jisc scholarly communications a common schema for OA policies. The proposed schema introduced the problem of the variety of policies as follows.

Funders' and institutions' OA policies documented, for example in ROARMAP (http://roarmap.eprints.org/) and Sherpa/Juliet (http://www.sherpa.ac.uk/index.html), are very varied. While there are likely to be good business and/or strategic reasons for this variety, it comes at the price of increased complexity for those asked to comply with those policies, and those who seek to advise them or provide other information services to them.

It would not be practical at this time to ask those issuing policies to converge on a single policy type, though that might arise through other means. Instead, a practical response to the problems generated by this variety would be to ask those issuing policies to express them using consistent language.

Jisc (https://www.jisc.ac.uk/open-access), SHERPA services and ROARMAP have jointly developed a schema for OA Policies essentially in the form of a questionnaire with values of elements recorded against questions.

Overview of the schema

The schema is divided in five sections, each one containing several fields. The sections are described in the proposed schema as following.

1. AGENCY: The 'agency' is the organisation (funder, institution or other) issuing the OA policy. If a policy varies within an agency, then each of those variations should be represented by a separate instance of the schema. For example, if different OA policies apply at a sub-institutional level or at the research-funder level (where national research council OA policies vary, then the schema should be completed separately for each sub-institutional policy / each research council OA policy).
2. POLICY: This section gives general parametric information about the OA policy (links, dates, and constituency).
3. REPOSITORY REQUIREMENTS: This section is about the policy requirements on (1) depositing items into repositories, and on (2) making the deposited items OA.
4. OA PUBLISHING REQUIREMENTS: This section is about items formally published in Open Access form. Where journal articles are concerned, this is commonly referred to as 'Gold' Open Access. Some publishers produce Open Access books and these, too, are part of the focus here.
5. OTHER CONDITIONS: This section enables agencies to add further information not captured by the schema.

An OA policy may have a number of different areas of application with different conditions. For example, a general policy covering books, journal articles and other outputs could have different requirements for each type of output. Another example might be where different discipline areas (eg,

humanities, life sciences) might have different allowable embargo periods. In these cases, each variant should be considered as a separate policy, and should be represented by a separate instance of the schema.

2) Mapping the schema to CERIF

This OA policy schema has been formalised using the CERIF data model. The logic of this choice was driven by three points. Firstly, CERIF is an EU Recommendation to member states. Secondly, CERIF is a data model in use in more than 43 countries. Lastly, CERIF is used by OpenAIRE and other large EC-funded projects.

The methodology used was based on the schema document. Each row has been mapped to a CERIF attribute within an entity (or several entities), allowing for multilingual entities (storing textual representations) and linking entities (to relate concepts of role and temporal validity). This information has been added to the schema document as additional columns. Apart from that, the figure 1 represents the diagram realised to clearly represent the relevant CERIF entities used.

Figure 1: CERIF entities used for the OA policy schema

An extensive usage of the semantic level has been done to follow the ethos of the schema document; indeed, this could have been done almost exclusively. Some elements have been represented using the CERIF linking relation technology. This technology could be applied to more elements but that would involve expanding the model (for example to link result publication to electronic address).

A few concepts are implied in the OA policy schema and do not directly appear as fields of the schema. These hidden concepts have been represented in the diagram. One of these concepts is the fact that a policy is linked to an organisation. This concept has been represented using the cfOrganisationUnit_ResultPublication (cfOrgUnit_ResPubl) link entity. The organisation unit (cfOrgUnit) represents the organisation and the result publication (cfResPubl) represents the policy.

Another concept is the availability of several variations for a policy. This concept is present in the description of the five sections of the schema. These variations are represented by the cfResultPublication_ResultPublication (cfResPubl_ResPubl) link. A specific classification can be used within this link to categorise the role of the link between two policies.

The full schema with mapping can be found in appendix 1.

II) Building a demonstration platform

1) Integrating CERIF

CERIF can be integrated as is in any relational database management system. The model integrates several mechanisms that help solve common problems in databases.

The first one gives the model the ability to link different entities. These links, or relations, can be either single valued (meaning that an instance of an entity A can only be linked to one instance of an entity B), like a postal address references only one country, or multi-valued, meaning that an instance of entity A can be linked to several instances of entity B. In this last case, a new relation entity is created between both entities to store the many links between instances.

The second mechanism also concerns the relation entities. This mechanism adds beginning and ending dates to each relation entities to allow the model to deal with the evolution of the relations. For example a person can be linked to an organisation unit for a certain period, leave this organisation unit and come back again, which results in a new link between these two entities.

The third mechanism is the multilingual capacity. For each textual information, CERIF provides the ability to store different values, each one corresponding to the value of the information in a specific language.

The last one is the semantic layer. This layer has two goals. First, it is used to enrich the description of entities with keywords coming from thesaurus or ontologies. The semantic layer can store these ontologies, and their terms can then be linked to the entities in the model. Second, the semantic layer is used in relation entities to indicate the meaning of the relation, or the role of the entities within the relation (for example: a person is director of an organisation unit). The semantic layer itself has relation entities so it is possible to relate terms in one terminological/classification scheme (in a particular language) to another. This is useful when interoperating across repositories with different vocabularies.

As the CERIF data model has been integrated in a platform that already provides several of these mechanisms, the model has been derived so that it can take advantage of the built-in capacities of the platform. The demonstration platform for Pasteur4OA is integrated within a Liferay platform, which provides built-in mechanisms for multilingual capacity and thesaurus management. The model has been derived to integrate the Liferay multilingual capacity, as it is close enough to the multilingual mechanism from CERIF. Using the built-in mechanism allows developers to take advantage of internal services of Liferay that already exists. It is also a very convenient way to provide users with a homogeneous look and feel all over the platform. The built-in thesaurus management has not been used as it does not provides as much possibilities as the semantic layer from CERIF.

Apart from this technical implementation, to fully integrate the OA policy schema within a CERIF database, a particular attention has been paid to the semantic layer. The classification schemes and terms that have been imported in the instance of CERIF give it the particular frame that allows users to manage OA policies. Fortunately, the OA policy schema comes with a full list of terms for each criterion in which user is expected to select a value. The classification schemes and terms just reflect these lists. Additional terms have also been integrated to manage the roles within relation-entities. This has been done by adding terms within CERIF semantics version 1.5.

2) Overlaying CERIF

CERIF is a model that can store any kind of data concerning research. It contains several mechanisms that gives this model a good level of completion.

This level of completion makes it difficult to be handled by people that are not aware of the model and its mechanisms. Hiding the complexity of CERIF can ease the presentation of data to end-users, allowing them to manipulate concepts that belong to their own world.

Another advantage of overlaying CERIF is the ability to separate concepts that will be stored within the same entity in CERIF. For example, if I want to store scientific institutions to organise researchers, and companies that are distributors of equipment used by the researchers, both concepts will be stored within the cfOrganisationUnit entity, but it does not make sense to present them in the same lists, or in the same way. By overlaying CERIF, developers can distinguished these two concepts and present them separately to end-users. In the focus of Pasteur4OA, a policy is built upon a questionnaire, and the data will be stored in several entities, mainly cfMeasurement and cfClassification, depending on whether user is free to enter any value, or if he has to select a value from a list. Sometimes, the user can even do both (the country or region in which the agency is based can be either a country selected from ISO 3166-1 list, or a region that the user may fill). By overlaying CERIF, developers will be able to mix data within a page, regardless of their storage.

Overlaying CERIF also means that developers can build several layers of services to access the data. A service based on CERIF can be provided so that data can be shared with external consumers using a common frame. Internal software can use a service based on the additional layer so that the same data can be shared using internal concepts, understood by people within the boundaries of the RI.

Building an additional layer means creating a new “virtual” schema that will be used to present data. The figure 2 represent this “virtual” schema for Pasteur4OA.

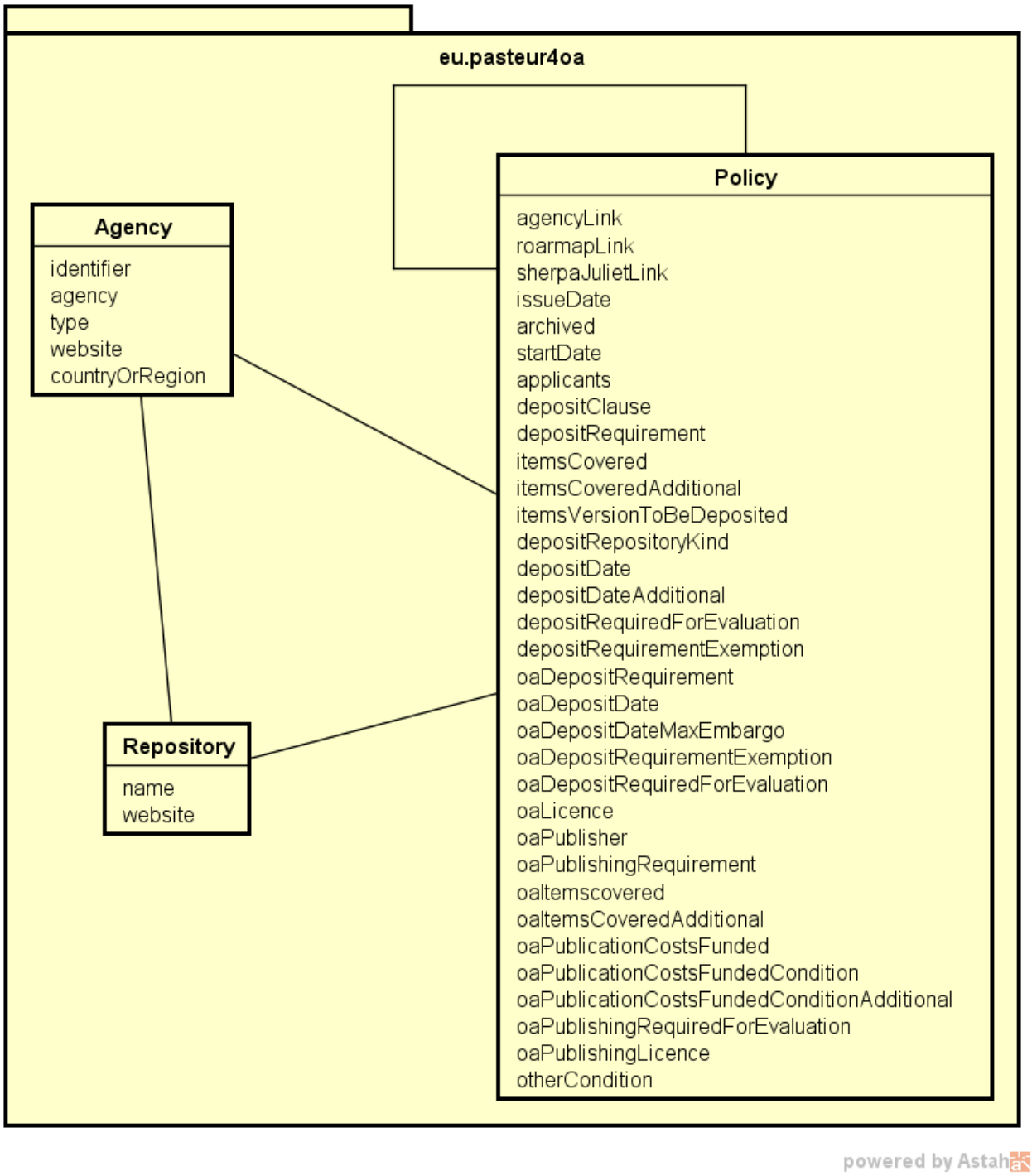

Figure 2: second layer schema for Pasteur4OA

Conclusion

The demonstration platform is currently being tested before being published in mid-2016 at the following address: http://pasteur4oa.is4ri.com. This platform stores ROARMAP data using CERIF, and provides some reports about these OA policies.

The work of analysis and development synthesized in this document proved that it is possible to have a homogeneous way to describe OA policies. This description can be expressed in a software using the CERIF model as a data model standard.

Using the schema and mapping proposed for CERIF, the documentation and analysis of OA policies could be facilitated, helping Pasteur4OA Knowledge Net reach its goals, so that they can help Member States develop or reinforce their strategies about OA policies.

References

Jacobs, N. (2014, December 1). A schema for OA policies [Web log post]. Retrieved from https://scholarlycommunications.jiscinvolve.org/wp/2014/12/01/a-schema-for-oa-policies/.

Appendix

1) OA policy schema with CERIF mapping

#	Field	Contents	Mand.	Guidelines	CERIF entity	CERIF attribute
Section 1: agency						
1	Name of agency issuing the policy	String	Y	Name of the agency (funder, institution or other) issuing the policy.	cfOrgUnitName	cfName
2	Identifier for agency	If funder, then FundRef ID If institution, then ISNI, Ringgold identifier, etc	N	An identifier for the agency, if known	cfFedId	cfFedId
3	Type of agency	funder / institution / both / other	Y	Specify whether the agency type issuing the policy is a research funder, an institution (research performing organisation), both, or other.	cfOrgUnit_Class cfClass *cfClassScheme*	cfTerm
4	Link to agency	URL	Y	Link to the agency website.	cfOrgUnit_EAddr cfEAddr *cfClass* *cfClassScheme*	cfEAddrId
5	Country or region in which agency is based	If region, then string. If country, then ISO 3166-1 code	Y	Name of the country or region where the agency is based. If it is a country insert ISO 3166-1 code, which is an international standard defining the codes for the representation of names of countries and their subdivisions.	cfOrgUnit_Class cfClass *cfClassScheme* or cfOrgUnit_Meas cfMeas *cfClass* *cfClassScheme*	cfTerm or cfValJudgeText
Section 2: policy						
6	Link to policy on agency site	URL	Y	Link to policy on agency website.	cfResPubl	cfURI
7	Link to policy in ROARMAP	URL	N	Link to policy in ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies - http://roarmap.eprints.org/	cfFedId *cfFedId_Srv* *cfSrv*	cfFedId

					cfClass *cfClassScheme* *cfSrv_Class*	
8	Link to funder policy in Sherpa/Juliet	URL	N	Link to funder policy in SHERPA/Juliet: Research funders' open access policies - http://www.sherpa.ac.uk/juliet/	cfFedId *cfFedId_Srv* *cfSrv* *cfClass* *cfClassScheme* *cfSrv_Class*	cfFedId
9	Date the policy was issued	Date	Y	Date the policy was issued.	cfResPubl	cfResPublDate
10	Is the policy current or archived?	Current/archived	Y	Indicate whether the policy is currently in place or it is an older version that has been archived. This field is to be changed when a new version is adopted.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
11	Policy start date	Date	Y	Date from when compliance with the policy must begin.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfStartDate
12	To whom does the policy apply?	String	N	Indicate to whom within the agency the policy applies, e.g. research students, researchers, all members of an institution, grant holders.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfValJudgeText
Section 3A: repository requirements - deposit						
13	Is there a deposit clause in the OA policy?	Y / N	Y	Indicate whether the OA policy makes explicit reference to depositing in a repository	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
14	Is depositing the item a requirement or a recommendation ?	Required Recommended Not specified	N	Required means you must deposit. Recommended means you are encouraged to deposit. This field can be qualified by the answer to field 23 (i.e. exemptions are possible).	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm

15	What kinds of items does the policy cover?	Book Book chapter Book edited Conference Paper/Proceeding/Abstract Journal Article/Review Manual/Guide Monograph Policy briefing report Technical Report Technical Standard Thesis Consultancy Report Working paper Other	N	Identify which items the policy covers (e.g. articles, books). More than one item type may be selected. If the policy varies according to item type -- e.g. where different OA embargo periods are allowed, or different licences specified -- then this should be considered a separate policy and a new schema should be filled in for each variant.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
16	More information about kinds of items covered	String	N	Further define item types specified in Field 15, if necessary.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfValJudgeText
17	Which version of the item is to be deposited?	1, 2, 3, 4	N	Specify whether the deposit is to be: 1. Author's manuscript pre-peer-review 2. Author's manuscript post-peer-review 3. Published version 4. Other Note, for articles, the options here should be interpreted as followed in relation to the NISO/ALPSP vocabulary http://www.niso.org/publications/rp/RP-8-2008.pdf [PDF]: 1.Author's manuscript pre-peer-review = "AO" or "SMUR" 2. Author's manuscript post-peer-review = "AM"	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm

				3. Published version = "VoR", "CVoR" or "EVoR" 4. Other		
18	Where to deposit	Institutional repository Funder or subject repository National repository Any repository Other Not specified	N	Indicate the repository in which the item is to be deposited.	cfResPubl_Srv cfSrv *cfClass* *cfClassScheme* *cfSrv_Class*	cfTerm
19	If deposit is not in the institutional repository, then which repository?	String	N	If deposit is external to the institution, indicate in which repository the item is to be deposited	cfResPubl_Srv cfSrv *cfClass* *cfClassScheme* *cfSrv_Class*	cfName cfURI
20	When to deposit	A+ x P+ x O Y Z	N	The codes refer to the date specified in the OA policy when the item is to be deposited. A + x - Immediately upon (or within x months of) date of acceptance. P+x = Immediately upon (or within x months of) date of publication. O - other = at another date not listed here, please add details in field 19. Y - as early as possible - the item is to be deposited at the earliest possible opportunity. Z - unspecified = the policy does not specify when the item is to be deposited.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
21	When to deposit if not covered by Field 19	String	N	If you have selected 'O- other' in Field 20, then specify when the item should be deposited.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfValJudgeText
22	Is depositing the item required for eligibility in	Yes No Not specified	N	Note whether the agency requires an item to be deposited into a repository, as a condition for research performance evaluation or assessment.	cfResPubl_Class cfClass cfClassTerm	cfTerm

	research performance eval uation or assessment?				*cfClassScheme*	
23	If there is a requirement to deposit the item, are exemptions allowed?	Yes unconditionally Yes conditionally No Not specified	N	If under some conditions, an item might be exempted from the deposit policy, indicate what kind of item and under what condition (if any). An exemption is sometimes also referred to as a waiver or an opt-out.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
Section 3B: repository requirements - OA						
24	Is making the deposited item OA a requirement or a recommendation ?	Required Recommended Not specified	N	Making the deposited item available free for all online. Required means you must make the deposited item OA. Recommended means you are encouraged to make the deposited item OA. This field can be qualified by the answer to field 27 (i.e. exemptions are possible).	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
25	When to make the deposited item OA?	A+ x P+ x O Y Z	N	The codes refer to the date specified in the OA policy when the deposited item is to be made OA (i.e., available free for all online). A + x - Immediately upon (or within x months of) date of acceptance. P+ x = Immediately upon (or within x months of) date of publication. O - other = at another date not listed here. Y - as early as possible - the item is to be made available at the earliest possible opportunity. Z - unspecified = the policy does not specify when the item is to be deposited.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
26	Maximum Allowable	String	N	If you have selected 'O- other' in Field 25, then specify what is the longest publisher embargo period that the policy allows.	cfResPubl_Meas cfMeas *cfClass*	cfValJudgeText

	Publisher Embargo Length				*cfClassScheme*	
27	If there is a requirement to make the deposited item OA, are exemptions allowed?	Yes unconditionally Yes conditionally No Not specified	N	If under some conditions, an item might be exempted from the OA clause of the policy, indicate what kind of item and under what condition (if any). An exemption is sometimes also referred to as a waiver or an opt-out.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
28	Is making the deposited item OA a condition for evaluation?	Yes No Not specified	N	Note whether the agency requires a deposited item to be made OA, as a condition for research performance evaluation or assessment.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
29	What licence should be used for making the item OA?	URL String Not specified	N	Provide the URL or description of the licence for the OA item. Multiple fields can be added so long as they are alternative allowable licences for all the item types covered by the policy. If some licences are specified for particular item types, then that would need to be expressed in different policies.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfValJudgeText
Section 4: OA publishing requirements						
30	Does the policy specify that the item is to be published OA through a publisher?	Yes No	Y	Explain whether the OA policy specifies formal OA publication. OA publication means the item (journal, book, etc) is to be published with a publisher who makes it OA (available free for all) immediately upon publication (sometimes for publication fee).	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
31	Is the formal OA publishing policy a requirement or a recommendation ?	Required Recommended Not specified	N	Does the OA policy require (i.e., you must) or recommend (i.e., you are encouraged) that the item (journal, book, etc) be published with a publisher who makes it OA (available free for all) immediately upon publication (sometimes for publication fee).	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm

32	What kinds of items does the policy cover?	Book Book chapter Book edited Conference Paper/Proceeding/Abstract Journal Article/Review Manual/Guide Monograph Policy briefing report Technical Report Technical Standard Thesis Consultancy Report Working paper Other	N	Identify which items the policy covers (e.g. articles, books). More than one item type may be selected. If the policy varies according to item type -- e.g. where different licences specified -- then this should be considered a separate policy and a new schema should be filled in for each variant.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
33	More information about kinds of items covered	String	N	Further define item types specified in Field 32, if necessary.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfVaLJudgeText
34	Does the agency fund OA publication costs?	Yes No Not specified	N	Funding might be through block grants, allowances in project budgets, dedicated institutional funds, or other methods	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
35	Are there conditions on use of agency OA publication funds?	None Maximum contribution per publication charge Prohibition on using the charge for hybrid journals Other	N	Explain which conditions apply for the use of the specific OA funds. Choose more than one option if that is applicable. Further qualifications for conditions on use are available in Field 36, and Field 37.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
36	Other conditions on use of agency OA publication funds	String	N	If Field 34 is "Other", then give further information about the conditions on the use of the agency's OA publication fund.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfVaLJudgeText

37	Does an item have to be published OA through a publisher for eligibility in research performance evaluation or assessment?	Yes No Not specified	N	Note whether the agency requires that the item (journal, book, etc) be published with a publisher who makes it OA (available free for all) immediately upon publication (sometimes for publication fee) in order to be eligible for performance review or research assessment.	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
38	If there is a requirement to publish the item OA through a publisher, are exemptions allowed?	Yes unconditionally Yes conditionally No Not specified	N	If under some conditions, an item might be exempted from a requirement to publish with an OA publisher, indicate what kind of item and under condition (if any). An exemption is sometimes also referred to as a waiver or an opt-out	cfResPubl_Class cfClass cfClassTerm *cfClassScheme*	cfTerm
39	Licenses	URL String Not specified	N	Provide the URL or description of the licence for the OA item. Multiple fields can be added so long as they are alternative allowable licences for all the item types covered by the policy. If some licences are specified for particular item types, then that would need to be expressed in different policies.	cfResPubl_Meas cfMeas *cfClass* *cfClassScheme*	cfValJudgeText
Section 5: Other						
40	Are there any other policy conditions not covered above?	String	N	Other conditions or comments of which those covered by the policy should be aware.	cfResPubl_Meas cfMeas cfClass cfClassScheme	cfValJudgeText