

機関リポジトリ推進委員会
メタデータ検討タスクフォース
平成 29 年 3 月 3 日

メタデータスキーマに関する国際動向調査（まとめ）

1 概要

機関リポジトリ推進委員会メタデータ検討タスクフォースでは、junii2 の改訂作業を開始するに当たり、国際的に広く利用されているメタデータスキーマについて調査を行った。調査の目的は、研究データ、助成情報、識別子等、現在の junii2 では十分に対応できていない情報の標準的な記述方法を把握し、相互運用性に優れたメタデータスキーマの設計に資すること、ひいては日本の機関リポジトリのメタデータの国際的な流通性を高めることにある。

2 調査対象

今回は以下の 4 種類のスキーマについて調査を行った。

	記述対象	主な利用地域	管理主体
OpenAIRE	文献、研究データ、研究情報 (research information)	欧州	OpenAIRE プロジェクト (Horizon2020 助成)
DataCite	研究データ	世界	DataCite
RIOXX	文献	英国	RCUK、HEFCE
CERIF/XML	研究情報	欧州	euroCRIS

3 調査結果

3.1 OpenAIRE

1) 概要

- OpenAIRE は、EU が助成した研究成果の OA 化を支援するプロジェクト。Horizon 2020 では助成を受けた研究論文の OA 化が義務化されており、OpenAIRE ではリポジトリから収集した論文情報と Horizon 2020 の助成プロジェクト情報をマッチングすることにより、OA 方針の実施状況をモニタリングしている。助成を受けた者は可能な限り速やかに出版社版か著者最終版をリポジトリに登録しなければならない。さらに、論文の妥当性検証に必要な研究データも論文と同時に登録することを目指さなければならないとされている。リポジトリへの登録は Gold OA の場合は出版時に、Green OA の場合は 6 ヶ月（社会科学及び人文科学の論文は 12 ヶ月）以内に登録する必要がある。
- 研究データについてはオープンリサーチデータ・パイロットが提唱され、パイロットへの参加を選択すると、論文の妥当性検証に必要な研究データとそれに関連するメタデータ及び直接論文に貢献しないものの収集されたデータや生データについても公開する必要がある。ただし、個人情報など公開にそぐわないデータは非公開とすることができる。
- OpenAIRE ではリポジトリ向け、データアーカイブ向け、CRIS 向けのメタデータガイドラインを作成しており、ガイドラインに準拠することで OA 化のモニタリングが可能となる。

2) メタデータの特徴

学術雑誌論文等の OA モニタリングに関係するメタデータ要素には以下がある。

要素名	Project Identifier
dc	dc:relation
繰り返し	0-1
説明	info:eu-repo の名前空間を使用してプロジェクト ID を記述する。プロジェクト ID がある場合は必須。 文法 info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectID/[Jurisdiction]/[ProjectName]/[ProjectAcronym]
例	<dc:relation> info:eu-repo/grantAgreement/EC/FP7/244909/EU/Making Capabilities Work/WorkAble </dc:relation>

要素名	Access Level
dc	dc:rights
繰り返し	1
説明	info:eu-repo の名前空間を使用してアクセスレベルを記述する。値は以下のいずれかとなる <ul style="list-style-type: none"> • info:eu-repo/semantics/closedAccess • info:eu-repo/semantics/embargoedAccess • info:eu-repo/semantics/restrictedAccess • info:eu-repo/semantics/openAccess
例	<dc:rights>info:eu-repo/semantics/openAccess</dc:rights>

要素名	Embargo End Date
dc	dc:date
繰り返し	0-1
説明	info:eu-repo の名前空間を使用してエンバーゴ終了日を記述する。Access Level が info:eu-repo/semantics/embargoedAccess の場合必須。日付は YYYY-MM-DD の形式で入力する。
例	<dc:date>info:eu-repo/date/embargoEnd/2015-12-31</dc:date>

研究データについては以下の要素がある。

要素名	Dataset Reference
dc	dc:relation

繰り返し	0-N
説明	関連するデータセットへの識別子。 info:eu-repo/semantics/dataset/<scheme>/<identifier>の形式で記述。 <scheme>の部分は統制語彙から選択する
例	"<dc:relation> info:eu-repo/semantics/dataset/doi/10.1234/789.1 </dc:relation>"

3) 参考文献

- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
<https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020>
- H2020 Programme AGA – Annotated Model Grant Agreement Version 2.1.1 1 July 2016
p.363-365.
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf
- OpenAIRE Guidelines for Literature Repositories
<https://guidelines.openaire.eu/en/latest/literature/index.html>

3.2 DataCite

1) 概要

- DataCite が開発している、研究データを記述するためのメタデータスキーマのデファクトスタンダード。引用等のために研究データを識別することが主目的である。特定の分野を対象としない汎用的な性格を持つ。
- 最新バージョンは 4.0 (2016 年 9 月リリース)
 - <https://schema.labs.datacite.org/meta/kernel-4.0/>
 - JaLC の研究データ用メタデータスキーマはバージョン 3.1 準拠
https://japanlinkcenter.org/top/doc/JaLC_tech_meta_lab_data.pdf
https://japanlinkcenter.org/top/doc/JaLC_tech_rd_guideline_ja.pdf
 - OpenAIRE Guideline for Data Archives はバージョン 3.0 準拠
<https://guidelines.readthedocs.io/en/latest/data/index.html>
- バージョン 3.1→4.0 での変更点 (一部)
 - creator/contributor
 - ✧ creator/contributor で複数の nameIdentifier が入力可能に
 - ✧ creatorName/contributorName にサブプロパティ familyName と givenName が追加
 - 助成情報
 - ✧ contributorType =Funder が廃止
 - ✧ 任意項目として独立したプロパティ FundingReference が追加 (サブプロパティ: funderName、funderIdentifier、awardNumber、awardURI、awardTitle)

- 他の変更点については以下を参照。

https://schema.labs.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf#page=5

2) メタデータの特徴

- 必須プロパティの少なさ（汎用的な用途のため）
 - 必須 (5) : Identifier、Creator、Title、Publisher、PublicationYear
 - 推奨 (7) : Subject、Contributor、Date、ResourceType、RelatedIdentifier、Description、GeoLocation
 - 任意 (6) : Language、AlternateIdentifier、Size、Format、Version、Rights
- 豊富に用意された寄与者タイプ (contributorType)
 - 全 21 種類 (ContactPerson、DataCollector、DataCurator、DataManager、Distributor、Editor、HostingInstitution、Producer、ProjectLeader、ProjectManager、ProjectMember、RegistrationAgency、RegistrationAuthority、RelatedPerson、Researcher、ResearchGroup、RightsHolder、Sponsor、Supervisor、WorkPackageLeader、Other)
 - ※Funder はバージョン 4.0 で廃止された
- OA ステータス
 - バージョン 4.0 では該当するプロパティはない。
- 助成情報の入力方法
 - バージョン 3.1 までは contributor (contributorType="Funder") に記述していたが、バージョン 4.0 では FundingReference という独立したプロパティが用意されている。5 個のサブプロパティのうち awardNumber、awardURI、awardTitle が実質的に追加されたと言える。

```

<fundingReferences>
  <fundingReference>
    <funderName>European Commission</funderName>
    <funderIdentifier funderIdentifierType="Crossref Funder ID">http://doi.org/10.13039/501100000780</funderIdentifier>
    <awardNumber awardURI="http://cordis.europa.eu/project/rcn/100180_en.html">282625</awardNumber>
    <awardTitle>
      MOTivational strength of ecosystem services and alternative ways to express the value of BIOdiversity
    </awardTitle>
  </fundingReference>
  <fundingReference>
    <funderName>European Commission</funderName>
    <funderIdentifier funderIdentifierType="Crossref Funder ID">http://doi.org/10.13039/501100000780</funderIdentifier>
    <awardNumber awardURI="http://cordis.europa.eu/project/rcn/100603_en.html">284382</awardNumber>
    <awardTitle>
      Institutionalizing global genetic-resource commons. Global Strategies for accessing and using essential public knowledge assets in the life science
    </awardTitle>
  </fundingReference>
</fundingReferences>

```

図：助成情報の記入例

<https://schema.labs.datacite.org/meta/kernel-4.0/example/datacite-example-fundingReference-v4.0.xml>

3) 関連情報

- オープンサイエンスにおけるメタデータの検討

http://www.nii.ac.jp/csi/openforum2016/track/pdf/20160526_IR01-7_oozono.pdf

3.3 RIOXX

1) 概要

- 英国研究会議 (RCUK) は 2013 年 4 月 8 日に「Policy on Open Access and Supporting Guidance」を公表し、RUCK の助成を受けた研究成果の OA 化を義務付けた。この方針に従い、助成を受けた研究成果が OA 化されていることをモニタリング可能にするため、EDINA が、英国情報システム合同委員会 (JISC) の助成を受け、RCUK および高等教育助成会議 (HEFCE) と共同でメタデータ仕様を策定した。
- 最新バージョンは RIOXX 2.0 (2015 年 1 月 22 日リリース)
- 2017 年 1 月現在、イギリス国内 67 機関で採用されている。採用機関のメタデータの妥当性チェックの結果が RIOXX のウェブサイトで公表されている。 (<http://rioxx.net/implementation/>)

RCUK の OA 方針

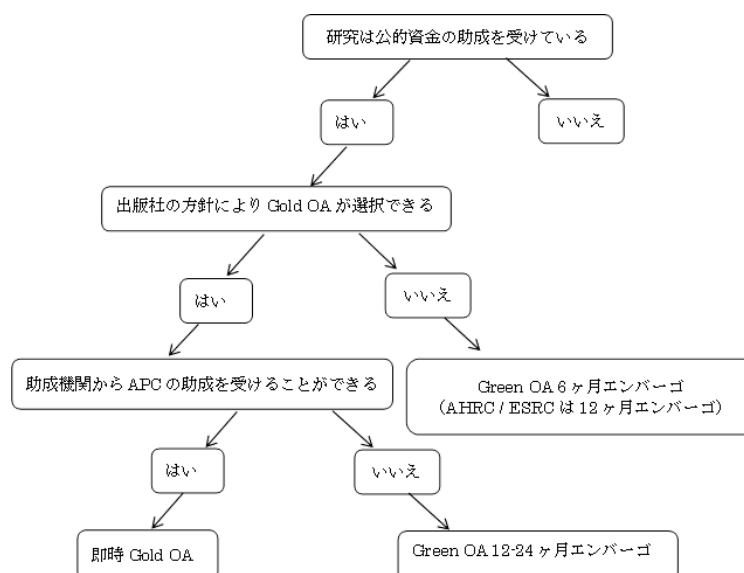
公的資金を投入し投資した研究を社会に還元する方法のひとつとして、ピアレビューを受けた研究成果の即時で制限のない無料の OA 化を求めている。

OA 対象

2013 年 4 月 1 日以降の RCUK の助成を受けたピアレビュー論文、会議録を対象とする。

OA の方法

Gold OA と Green OA の両方を支持するが再利用の機会を最大にするためエンバーゴなしでの OA 化が望ましい。しかし、出版社の方針によりエンバーゴが必要な場合は下図に従い、エンバーゴを設けることを認めている。なお、RCUK では、APC 支払のための補助金を研究機関に対して支出している。



ライセンス

Gold OA の場合は CC-BY での公開を求める。Green OA の場合は CC-BY-NC のような非営利での再

利用に制限のないライセンスを求める。

リポジトリについて

リポジトリは OA と助成情報のメタデータをアグリゲートや検索可能な形式で保持しなければならない。

2) メタデータの特徴

OA 化のモニタリングに関連するメタデータ

助成機関情報、エンバーゴ期間、APC 支払有無、アクセプト日に係わるメタデータを用意することで、助成を受けた論文が RCUK の求める期間で OA 化されているか確認できるようにしている。以下に関連する要素を表で示す。

①助成機関情報

要素名	rioxxterms:project
属性	funder_name funder_id (ISNI か DOI を推奨)
繰り返し	0~1
説明	値には助成機関のプロジェクト ID を記入する。助成機関名と助成 ID は属性で記入する。助成 ID は HTTP URI の形式。
例	<pre><rioxxterms:project funder_name="Engineering and Physical Sciences Research Council" funder_id="http://isni.org/isni/0000000403948681" > EP/K023195/1 </rioxxterms:project></pre>

②エンバーゴ期間

要素名	ali:free_to_read
属性	start_date end_date
繰り返し	0~1
説明	コンテンツが一定の期間、あるいはエンバーゴ期間後に無料で認証なしにアクセスできる場合に記入する。属性は YYYY-MM-DD の形式で記入する。OA のステータスを表現するために使用する。値はなし。
例	<pre><ali:free_to_read start_date="2013-03-28" end_date="2014-04-30"> <ali:free_to_read start_date="2013-03-28"> <ali:free_to_read></pre>

③APC 支払有無

要素名	rioxxterms:apc
繰り返し	0~1
説明	APC 支払有無を統制語から選択する。値は paid, partially waived, fully waived, not charged, not required, unknown

※OA 方針では APC 支払有無等により求めるエンバーゴ期間が異なるため用意された要素

④アクセプト日

要素名	dcterms:dateAccepted
繰り返し	1
説明	アクセプトされた日付を YYYY-MM-DD の形式で記入する。

3) 参考文献

- RCUK Policy on Open Access
<http://www.rcuk.ac.uk/research/openaccess/policy/>
- RCUK RIOXX Application Profile Version 2.0 Final
<http://riox.net/profiles/v2-0-final/>
- Paul Walk (2016). RIOXX . Insights 29(2).
<http://insights.uksg.org/articles/10.1629/uksg.303/>

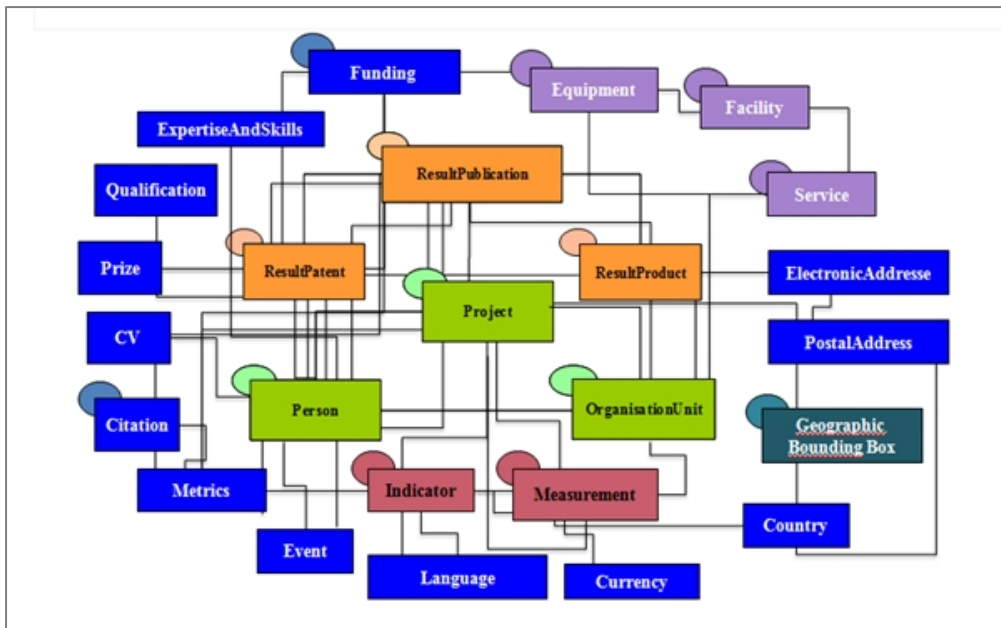
3.4 CERIF (Common European Research Information Format)

1) 概要

- 欧州の euroCRIS が開発する、研究情報 (research information) を管理するためのデータモデル (Entity-Relationship モデル)
 - <http://www.eurocris.org/cerif/main-features-cerif>
- CRIS (Current Research Information System) で利用されている
 - Thomson Converis
 - Elsevier Pure
 - CINECA DSpace-CRIS (香港大学、euroCRIS 等)
- 最新バージョンは 2013 年の 1.6 (Entity 数 293、Relationship 数 665)
 - <http://www.eurocris.org/cerif/feature-tour/cerif-16>
 - <http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/documentation/MInfo.html>
- 交換フォーマットは CERIF/XML
 - http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.6/CERIF_1.6_2.xsd

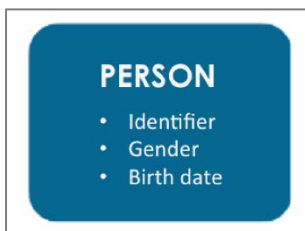
2) メタデータの特徴

①Entity-Relationship モデルの採用



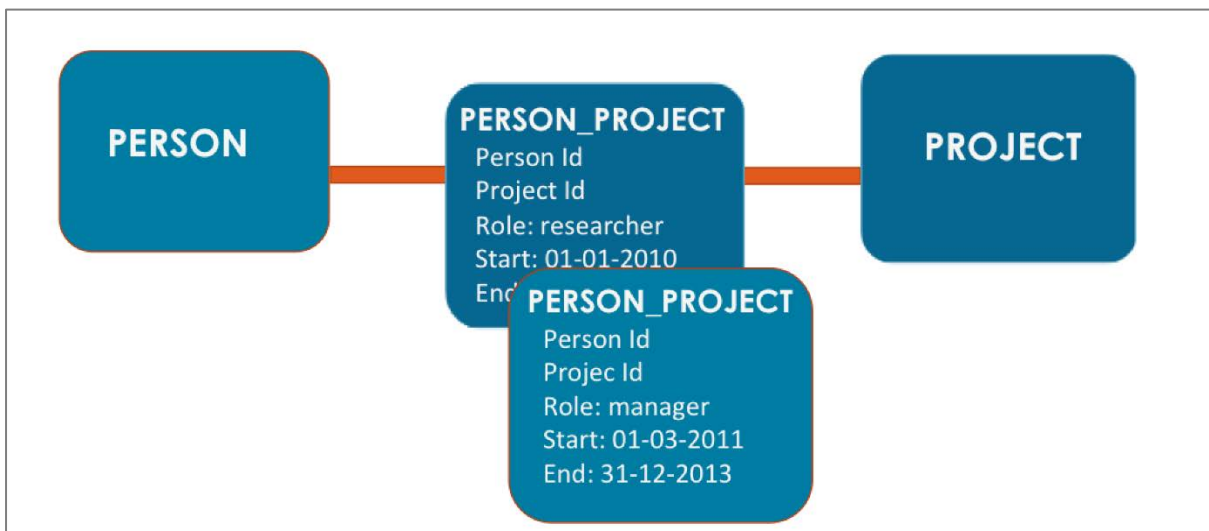
②Entity

例えば cfPers エンティティ（下図）には名前を持たせず、cfPersName に記述するなど細かく分けられている。



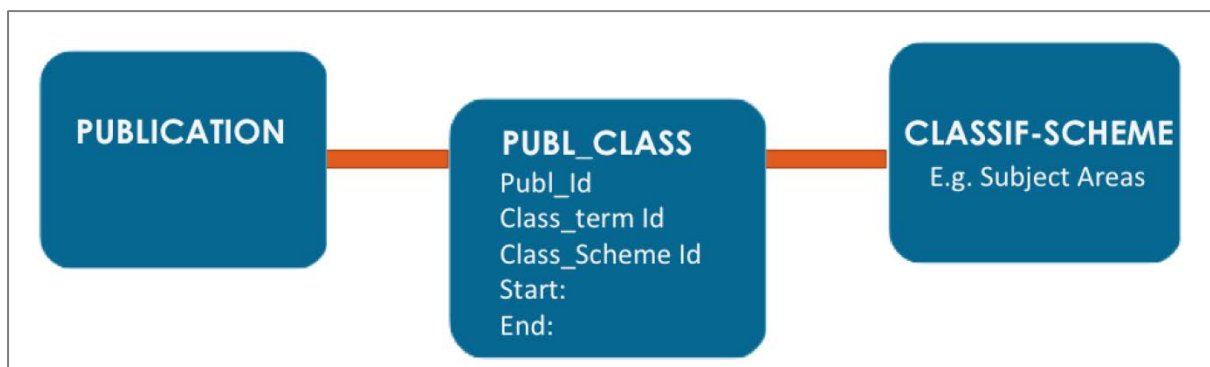
③Relationship (Linking Entities)

各 Entity に ID を持たせ、Entity 間の関係を記述する Entity を用意している（関係の種類だけ Entity がある）。それぞれの関係について、Start、End を記述できる。



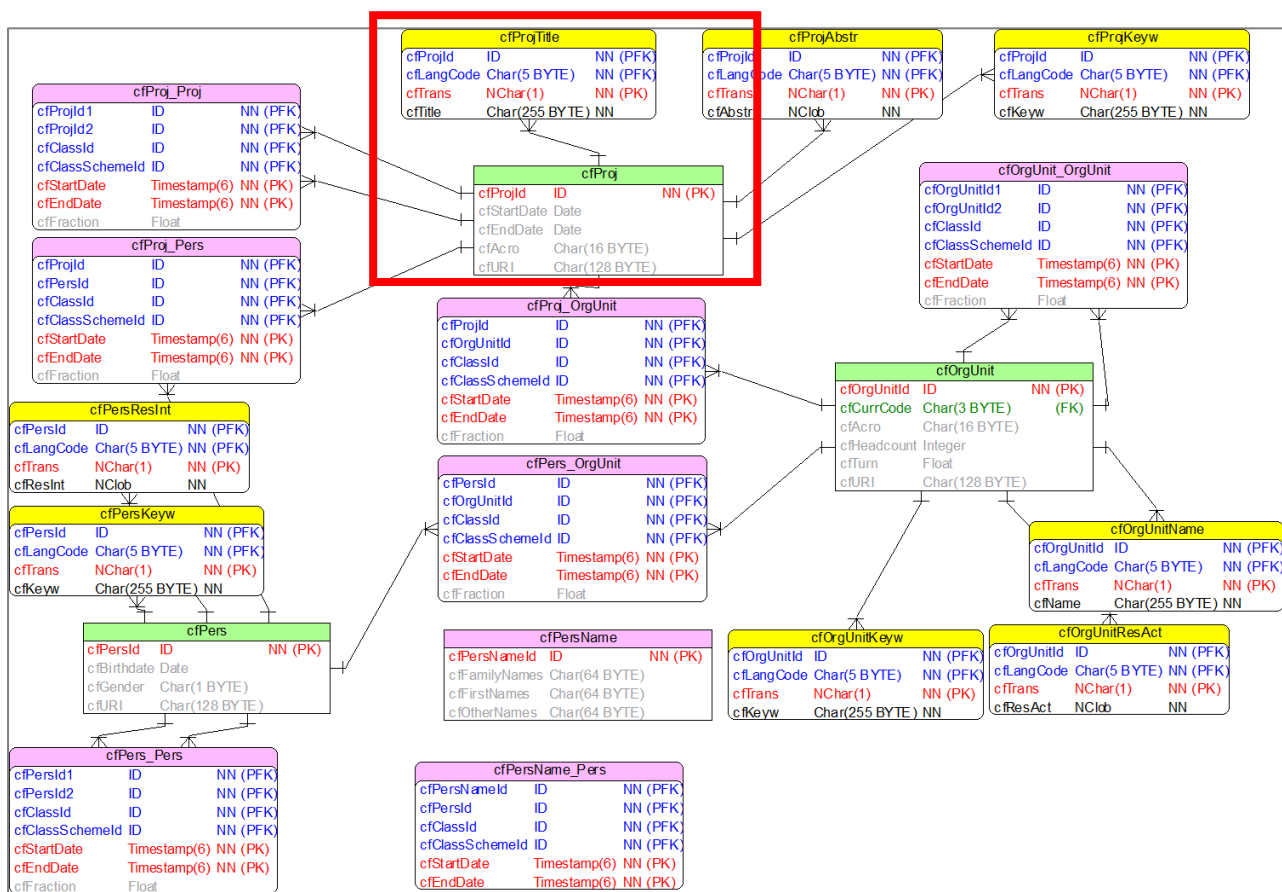
④Class

統制語等の体系は cfClassScheme で管理して、*_Class で記述する。



⑤多言語対応 (cfLangCode、cfTrans 属性)

例えば、ある cfProj に対して複数の cfProjTitle を関係付け、各 cfProjTitle において記述言語と翻訳有無を指定可能。



3) OpenAIRE Guideline for CRIS Managers との関連

OpenAIRE では CERIF/XML をベースにしたガイドラインを作成し、CRIS からのメタデータ収集に対応している (ただし、2016 年夏時点で実際のガイドライン対応例はまだ EKT の 1 つと聞いている)。同ガイドラインでは CERIF のサブセットとして以下の 8 つの Entity に対応しており、オープンサイエ

ンス対応としてはこれらが主要な項目と考えられる。各 Entity の詳細はガイドラインを参照のこと。

<http://doi.org/10.5281/zenodo.17065>

Publication	cfResPubl	Open Access Type (Closed Access、Embargoed Access、Open Access、Restricted Access) を指定
Product/Dataset	cfResProd	〃
Person	cfPers	
Organisation	cfOrgUnit	
Project	cfProj	Open Access Requirements として OA mandated、OA not mandated を指定
Funding	cfFund	〃
Equipment	cfEquip	機器等について記述 (equipment/devices that are used for the generation of data sets)
Service	cfSrv	メタデータ提供元について記述