

B2 : 次世代リポジトリシステム BoF: How should we develop a next generation repository system?

座長 :

林 正治 (国立情報学研究所オープンサイエンス基盤研究センター)

本セッションの目的

- 「オープンアクセスリポジトリ連合（COAR）による次世代リポジトリ（NGR）についての機能要件および技術勧告が公開され、実装に向けた国際的な議論が活発化している。本セッションでは、機関や分野別リポジトリ運用者を集い、NGRが提案する11の機能の意義や実現可能性、さらに我々が必要とする12つ目の機能について議論する。」

COARのNGRに限らず
次世代リポジトリに必要な機能を
議論したいと思っています

本日の内容

- 林 正治（国立情報学研究所オープンサイエンス基盤研究センター）
 - ・イントロダクション
- 八塚 茂（科学技術振興機構 バイオサイエンスデータベースセンター）
 - ・オープンで信頼できるリポジトリの条件
- 石井 真史（物質・材料研究機構）
 - ・マテリアルズ・インフォマティクスのためのデータ駆動型リポジトリ構築に向けて
- ディスカッション

発表 20分（質疑含む）

イントロダクションとして 2つの次世代リポジトリ像を紹介

オープンアクセスリポジトリ連合（COAR）の Next Generation Repositories（NGR）

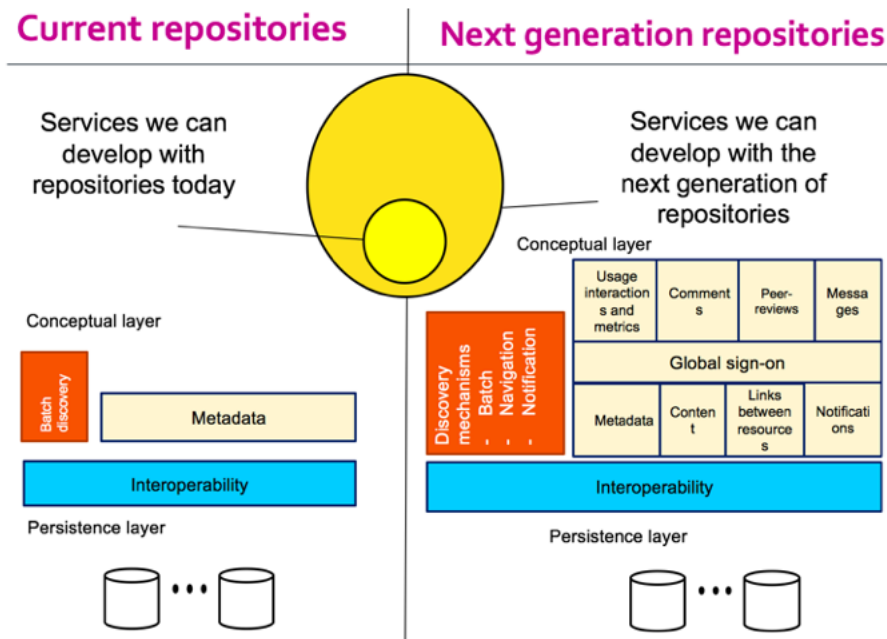


COAR NGR WGの経緯

- 2016年4月
 - COAR Next Generation Repositories WG設立
- 2017年2月
 - 「Introduction, Rationale and User Stories」公開
- 2017年3月
 - COAR 2017 Annual Meeting
- 2017年11月
 - 「Behaviours and Technical Recommendations」公開
- 2018年3月
 - COAR 2018 Annual Meeting

COAR NGRのビジョン

- 「リポジトリを、分散型でグローバルにネットワーク化された学術コミュニケーションのインフラストラクチャの基礎として位置付け、その上に付加価値サービスを積み重ね、それにより（商業出版社に支配された）既存のシステムを、より研究中心的で革新的な、学術コミュニティによって共同管理されたシステムに、変えていくこと」



技術的なビジョン

- ウェブのアイデアを踏襲し、リソース中心のインフラ構築
- デポジットプロセスを単純化等、ユーザワークロードの省力化

COAR NGRの特徴

- 多様な学術リソースを管理、アクセスを提供
 - 論文、データセット、プロジェクト、著者、組織など
- リソース中心のインフラストラクチャを提供
 - すべての学術リソースにURLを付与
- リポジトリネットワーク
 - 個々のリポジトリの垣根を越えて、学術リソースによるネットワーク形成を実現
- 機械処理対応
 - リポジトリは機械処理に対応し、機械が学術リソースを適切に処理できるよう支援
- アクティブなリポジトリ
 - ユーザアクティビティ、コメント、アノテーションなど積極的に情報を集めて発信、外部サービスとも積極的に連携

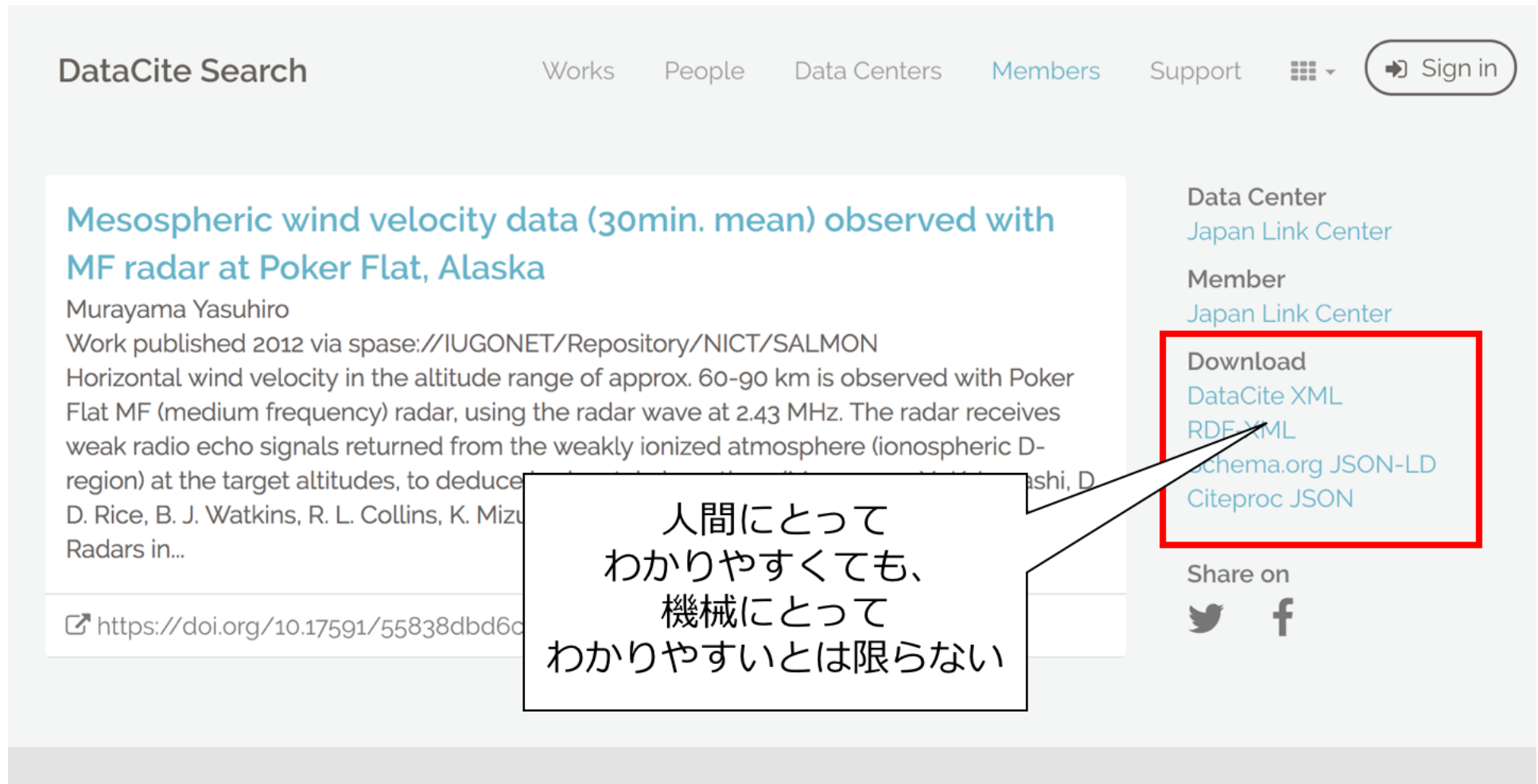
振る舞いと対応技術

1. 識別子の公開 (Exposing Identifiers)
 - **Signposting**
2. リソースレベルでのライセンスの明示 (Declaring Licenses at a Resource Level)
 - Creative Commons Licenses, **Signposting**
3. ナビゲーションによる発見 (Discovery through Navigation)
 - **Signposting**
4. リソースとの相互作用 (アノテーション, コメント, レビュー) (Interacting with Resources (Annotation, Commentary and Review))
 - Activity Streams 2.0, Web Annotation Model and Protocol, International Image Interoperability Framework
5. リソースの転送 (Resource Transfer)
 - ResourceSync, SWORD, IPFS
6. バッチディスカバリ (Batch Discovery)
 - ResourceSync, **Signposting**, Sitemaps
7. アクティビティの収集・公開 (Collecting and Exposing Activities)
 - Activity Streams 2.0, Linked Data Notifications, ResourceSync, **Signposting**, Webmention, WebSub
8. ユーザの識別 (Identification of Users)
 - ORCID, Social Network Identities, WebID
9. ユーザの認証 (Authentication of Users)
 - HTTP Signatures, OpenID Connect, WebID/TLS
10. 標準化された利用状況メトリクスの公開 (Exposing Standardized Usage Metrics)
 - COUNTER, SUSHI, ETag
11. リソースの保存 (Preserving Resources)

Signposting

- 学術リソースを機械処理させるための手法の一つ
- HTTPヘッダー内にTyped Linksを埋め込むことで学術リソースの意味を機械に理解させる
- 定義されているパターン
 - Author : 作者の発見を支援
 - Bibliographic Metadata : 書誌メタデータの発見を支援
 - Identifier : 識別子の発見を支援
 - Publication Boundary : 出版物の構成リソースの発見を支援
 - Resource Type : リソースのセマンティックタイプを伝達

DataCite Searchでの実装例



DataCite Search Works People Data Centers Members Support Sign in

Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska

Murayama Yasuhiro

Work published 2012 via spase://IUGONET/Repository/NICT/SALMON

Horizontal wind velocity in the altitude range of approx. 60-90 km is observed with Poker Flat MF (medium frequency) radar, using the radar wave at 2.43 MHz. The radar receives weak radio echo signals returned from the weakly ionized atmosphere (ionospheric D-region) at the target altitudes, to deduce

D. Rice, B. J. Watkins, R. L. Collins, K. Mizu Radars in...


<https://doi.org/10.17591/55838dbd6c0ad>

Data Center
Japan Link Center

Member
Japan Link Center

Download

- DataCite XML
- RDF/XML
- Schema.org JSON-LD
- Citeproc JSON

Share on  

人間にとって
わかりやすくても、
機械にとって
わかりやすいとは限らない

Ref: <https://search.datacite.org/works/10.17591/55838dbd6c0ad>

HTTPヘッダーの取得

```
$ curl -I "https://search.datacite.org/works/10.17591/55838dbd6c0ad"
```

```
HTTP/2 200
```

```
date: Mon, 11 Jun 2018 23:17:59 GMT
```

```
content-type: text/html;charset=utf-8
```

```
content-length: 14926
```

```
set-cookie:
```

```
AWSALB=2hUUGI3GMQDB+ZyXheGm3KMxl5hJUDZw8jXuM1d6FVJ6BD2fxpKopT2Yu1/5MudGxcHlvWEB3qGi5Lwzngmift6MHB82/dsvaXO  
hnItP0ztuTvzr/Ln90UnaEeNd; Expires=Mon, 18 Jun 2018 23:17:56 GMT; Path=/  
status: 200 OK
```

```
link: <https://doi.org/10.17591/55838dbd6c0ad> ; rel="identifier", <https://doi.org/10.17591/55838dbd6c0ad> ;  
rel="describedby" ; type="application/vnd.datacite.datacite+xml", <https://doi.org/10.17591/55838dbd6c0ad> ;  
rel="describedby" ; type="application/vnd.citationstyles.csl+json", <https://doi.org/10.17591/55838dbd6c0ad> ;  
rel="describedby" ; type="application/x-bibtex"
```

```
access-control-allow-origin: *
```

```
x-xss-protection: 1; mode=block
```

```
x-content-type-options: nosniff
```

```
x-frame-options: SAMEORIGIN
```

```
x-powered-by: Phusion Passenger 5.3.1
```

```
server: nginx/1.14.0 + Phusion Passenger 5.3.1
```

```
access-control-allow-headers: Content-Type,Accept,Accept-Encoding,Origin,User-Agent,Cache-Control,Keep-Alive
```

HTTPヘッダーの中身

```
$ curl -I "https://search.datacite.org/works/10.17591/55838dbd6c0ad"
```

```
HTTP/2 200
```

```
date: Mon, 11 Jun 2018 23:17:59 GMT
```

```
content-type: text/html;charset=utf-8
```

```
content-length: 14926
```

```
set-cookie:
```

```
AWSALB=2hUUGI3GMQDB+ZYXheGm3KMxl5hJUDZw8jXuM1d6FVJ6BD2fxpK  
hnItP0ztuTvzr/Ln90UnaEeNd; Expires=Mon, 18 Jun 2018 23:17:56 GMT; Pat
```

```
status: 200 OK
```

```
link: <https://doi.org/10.17591/55838dbd6c0ad> ; rel="identifier", <https://doi.org/10.17591/55838dbd6c0ad> ; rel="describedby" ;  
type="application/vnd.datacite.datacite+xml" <https://doi.org/10.17591/55838dbd6c0ad> ; rel="describedby" ;
```

```
link:  
<https://doi.org/10.17591/55838dbd6c0ad> ; rel="identifier",  
<https://doi.org/10.17591/55838dbd6c0ad> ; rel="describedby" ; type="application/vnd.datacite.datacite+xml",  
<https://doi.org/10.17591/55838dbd6c0ad> ; rel="describedby" ; type="application/vnd.citationstyles.csl+json",  
<https://doi.org/10.17591/55838dbd6c0ad> ; rel="describedby" ; type="application/x-bibtex"
```

```
x-content-type-options: nosniff
```

```
x-frame-options: SAMEORIGIN
```

```
x-powered-by: Phusion Passenger 5.3.1
```

```
server: nginx/1.14.0 + Phusion Passenger 5.3.1
```

```
access-control-allow-headers: Content-Type,Accept,Accept-Encoding,Origin,User-Agent,Cache-Control,Keep-Alive
```

識別子の提供

書誌メタデータの提供

Citation Style Language(CSL)の取得

```
$ curl -LH "Accept: application/vnd.citationstyles.csl+json" https://doi.org/10.17591/55838dbd6c0ad
```

```
{
  "type": "article",
  "id": "https://doi.org/10.17591/55838dbd6c0ad",
  "language": "en",
  "author": [
    { "family": "Yasuhiro", "given": "Murayama" }
  ],
  "issued": {
    "date-parts": [[2012]]
  },
  "abstract": "Horizontal wind velocity in the altitude range of approx. 60-90 km is observed with Poker Flat MF (medium frequency) radar, using the radar wave at 2.43 MHz. The radar receives weak radio echo signals returned from the weakly ionized atmosphere (ionospheric D-region) at the target altitudes, to deduce horizontal air motions (Murayama, Y., K. Igarashi, D. D. Rice, B. J. Watkins, R. L. Collins, K. Mizutani, Y. Saito, and S. Kainuma, Medium Frequency Radars in Japan and Alaska for Upper Atmosphere Observations, IEICE Trans., E83-B, pp.1996-2003, 2000). Poker Flat MF radar has been constructed as part of Japan-US joint research program of Arctic middle & upper atmosphere (¥"Alaska Project¥") in collaboration between National Institute of Information and Communications Technology, Japan (formerly Communications Research Laboratory), and Geophysical Institute, University of Alaska Fairbanks.",
  "DOI": "10.17591/55838dbd6c0ad",
  "publisher": "spase://IUGONET/Repository/NICT/SALMON",
  "title": "Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska"
```

処理例

```
In [1]: import requests
import json
from citeproc import Citation, CitationItem, CitationStylesStyle, CitationStylesBibliography, formatter
from citeproc.source.json import CiteProcJSON
r = requests.get('https://doi.org/10.17591/55838dbd6c0ad', headers={'Accept' : 'application/vnd.citationstyles.csl+json'})
csl_data = json.loads("["+r.text+"]")
source = CiteProcJSON(csl_data)
style = CitationStylesStyle('apa.csl', validate=False)
bib = CitationStylesBibliography(style, source, formatter.plain)
bib.register(Citation([CitationItem('https://doi.org/10.17591/55838dbd6c0ad')]))
print(''.join(bib.bibliography()[0]))
style = CitationStylesStyle('elsevier-harvard.csl', validate=False)
bib = CitationStylesBibliography(style, source, formatter.plain)
bib.register(Citation([CitationItem('https://doi.org/10.17591/55838dbd6c0ad')]))
print(''.join(bib.bibliography()[0]))
style = CitationStylesStyle('ieee.csl', validate=False)
bib = CitationStylesBibliography(style, source, formatter.plain)
bib.register(Citation([CitationItem('https://doi.org/10.17591/55838dbd6c0ad')]))
print(''.join(bib.bibliography()[0]))
```

Yasuhiro, M.. (2012). Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska. spase://IUGONET/Repository/NICT/SALMON. <http://doi.org/10.17591/55838dbd6c0ad>

Yasuhiro, M., 2012. Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska.. doi:10.17591/55838dbd6c0ad [1]M. Yasuhiro, "Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska" . spase://IUGONET/Repository/NICT/SALMON, 2012.

Signpostingにより
機械が情報を取得することが容易化
リポジトリの可能性は広がる・・・

COAR NGRのまとめ

- リポジトリ間の相互運用性の向上
 - リソース中心なインフラをベースとした分散型リポジトリネットワークを構築
- 付加価値サービス登場奨励
 - リポジトリの価値を高め、学術リソース流通を活性化
- 商業出版者に支配された学術コミュニケーションシステムの変革
 - 学術コミュニティ主導の学術コミュニケーションシステムの実現

学術リソースを中心とした
機械可読なリポジトリネットワークの構築
学術コミュニケーションの変革

次期JAIRO Cloud開発共同タスクフォース報告： 次世代リポジトリの機能に関する提案

2017年9月、次期JAIRO Cloud開発のために
JPCOARとNIIとが共同でタスクフォースを設置

J P C O A R
オープンアクセスリポジトリ推進協会



報告書の入手先：

<http://id.nii.ac.jp/1458/00000091/>

現状認識

• 日本の状況

- 2005年の千葉大学に始まり、800近い機関がリポジトリを構築
 - 500機関以上がJAIRO Cloudを利用
- コンテンツ（本文あり）の数は200万件超
 - 紀要論文が約半数、つづいて学術雑誌論文、学位論文などが続く
- 機関リポジトリの構築目的
 - グリーンOAの推進と灰色文献の保存公開
- OAポリシーはゆるやかに浸透

• 世界の動向

- COARの次世代リポジトリ
 - リポジトリネットワークによる学術コミュニケーションの変革
- プレプリントサーバの台頭、プレプリントの重要性の増加
 - COSによるプレプリントサーバ、arXivの拡大
- 既存の出版社に依存しない学術論文プラットフォームの登場
 - ウェルカム財団によるWellcome Open Researchなど
- 機関リポジトリの利用拡大
 - 論文のエビデンスデータ、デジタルアーカイブ、電子化された教材の共有

課題

1. セルフアーカイブの推進
 - セルフアーカイブの実施はいまいち、プレプリントサーバの利用は上昇
2. グリーンOAの評価と見直しが必要
 - グリーンOAへの投資効果の検証
3. 機関リポジトリのビジビリティ向上、ブランド化が必要
 - 研究者は機関リポジトリが見えていない
4. OA方針の評価、モニタリングが必要
 - OA方針の策定はゆるやかに進んでいる
5. 日本の機関リポジトリの国際化が必要
 - 国際発信の視点が必要
6. 機関リポジトリネットワークを用いた付加価値サービスの実現
 - 現状は統合検索システムJAIROのみ
7. 機関リポジトリの利用拡大
 - データリポジトリ、デジタル・アーカイブ、教材リポジトリ対応が必要
8. 機関リポジトリの現状把握が必要
 - 利用統計、評価状況の評価
9. 現状は世界的な学術流通の生態系から孤立、外部システムやサービスとの連携が必要
 - サービス連携が必要

共同TFの視点

課題解決への検討事項

1. グリーンOAの検証・評価・見直し
2. メタデータ入力等の自動化
 - メタデータ自動入力、著作権ポリシ確認の自動化
3. 新たな学術コミュニケーションのインフラとして
機関リポジトリを活用するための機能の実現
 - プレプリント、統計分析、版管理、真正性の担保など
4. 機関リポジトリの価値を高めるためのアクション
 - 研究データ、デジタルアーカイブ、教育コンテンツ

まずは、グリーンOAの検証・評価
そして、メタデータ入力等の自動化による作業負荷の軽減
機関リポジトリの活用・価値向上

まとめ

• COARの NGR

- 学術リソースを中心とした機械可読なりポジトリネットワークの構築→付加価値サービスの実現→学術コミュニケーションの変革

• 共同TFのNGR

- グリーンOAの評価→メタデータ入力の自動化等の作業負荷軽減→機関リポジトリの活用・価値向上へ向けた取り組み

研究データを見据え、次世代リポジトリには
どういった機能が求められるのか？
より具体的な議論に落とし込んでいく必要がある