

趣旨説明

西岡 千文（京都大学附属図書館）

北村 由美（京都大学附属図書館）

2021年度京都大学図書館機構講演会「オープンデータとしての学術論文」



京都大学
KYOTO UNIVERSITY



学術論文とオープンデータ

オープンデータ

制度面ならびに技術面における再利用性が確保された情報公開の枠組み

大向一輝. (2015). 学術情報流通とオープンデータ (< 特集> オープンデータ). 情報の科学と技術, 65(12), 503-508.

- 制度面 ← ライセンス等
 - オープンアクセス運動で主に取り組まれてきている
- 技術面 ← 機械可読性、相互運用性等
 - オープンアクセス運動で言及されてはいるが、制度面での要件ほど浸透してはいない
 - BOAI10では、リポジトリでPDFファイルを登録するだけでなく、XML等の機械可読フォーマットへ変換するツールも提供するべきであるとされている

関連する動向

- 第五期国立国会図書館科学技術情報整備基本計画策定に向けての提言―「人と機械が読む時代」の知識基盤の確立に向けて―(国立国会図書館)

<https://doi.org/10.11501/11631622>

- Digital Strategy for the Library of Congress FY2019-2023 (米国議会図書館)

<https://www.loc.gov/digital-strategy/>

- Technical Guidance and Requirements (Plan S)

<https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>

メタデータ	標準的な相互運用性のあるフォーマットで提供する	義務
引用データ	I4OCの標準に沿ってアクセス可能にする	強く推奨
本文	JATS XML等の標準的な機械可読性のある形式で(補足テキストとデータを含む)出版物の全文をダウンロードできるようにする	強く推奨

学術論文の機械可読性 [1/2]

- メタデータ(タイトル、著者、出版年等)
 - CrossrefやJaLCによって、JSONやRDFで提供されている
 - ORCID等の識別子の普及により一意性の保証
 - 助成情報など、対象となるメタデータが広がっている
- 引用データ
 - I4OC(Initiative for Open Citations)の取り組みによって、引用データの制度面ならびに技術面の再利用性が高まっている
 - 制度面 → Crossref上の引用データを公開するよう出版者に働きかけ
 - 技術面 → オープン・サイテーションの原則
 - STM分野の学術論文が主な対象であり、人文社会学系分野の文献については課題がある

学術論文の機械可読性 [2/2]

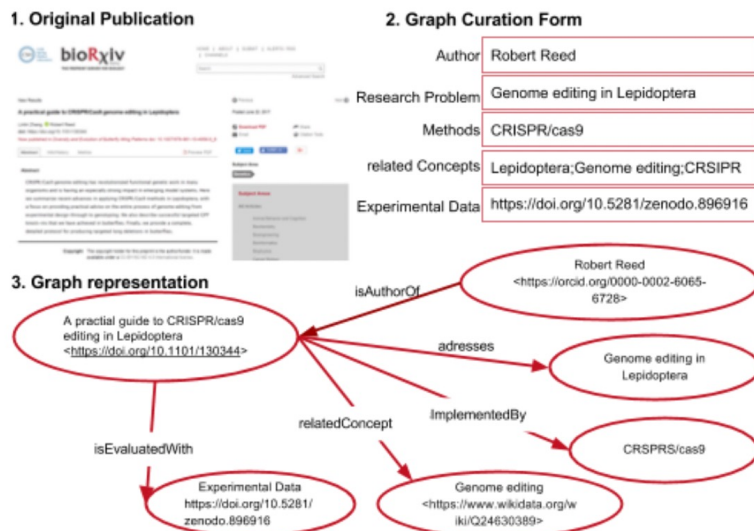
• 本文

• 本文テキスト

- JATS XML等の形式等で論文を公開
- Elsevier OA CC-BY Corpus: 40,001件の論文のデータセット
<https://doi.org/10.17632/zm33cdndxs.2>
- CLARIN ERICは22の学術的テキストのコーパスを提供
<https://www.clarin.eu/resource-families/corpora-academic-texts>

• 本文コンテンツの構造化

- Open Research Knowledge Graph (ORKG)



論文が扱っている研究課題、使用した手法、使用したデータセット等の情報を構造化して表現

セマンティック・ウェブ

メタデータの活用により、ウェブ情報の「意味」(semantics)をコンピュータが検知できるようにして、情報収集・利用の高度な自動化を図る技術

「セマンティックウェブ」. 日本図書館情報学会用語辞典編集委員会(編). 図書館情報学用語辞典第5版.

- Webページは主にHTMLで記述されており、Webブラウザが人間に情報を表示することを主な目的としている
→ 機械が読み込んで意味のあるデータとして活用するには不十分
- セマンティック・ウェブでは、Webページの情報がそれぞれ何を表しているかを統一された形式で記述することで、機械による情報の取得、比較、統合などを可能にする

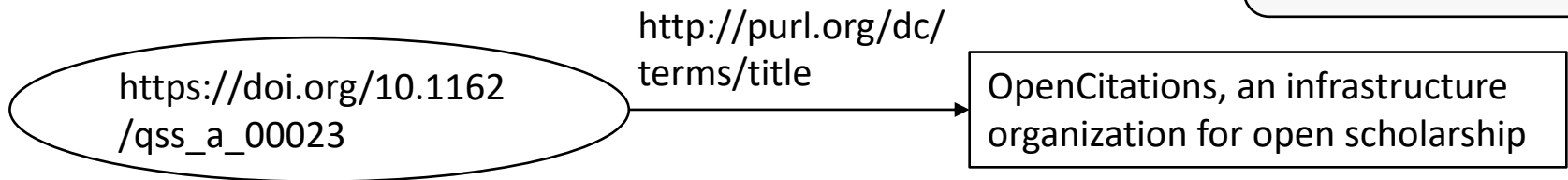
RDF [1/2]

RDFはリソースの関係を主語・述語・目的語という
3つの要素(トリプル)で表現

https://doi.org/10.1162/qss_a_00023には“OpenCitations, an infrastructure organization for open scholarship”というタイトルがあります。

主語 (Subject)	<u>リソース</u>	https://doi.org/10.1162/qss_a_00023
述語 (Predicate)	<u>プロパティ</u>	http://purl.org/dc/terms/title
目的語 (Object)	プロパティの値	OpenCitations, an infrastructure organization for open scholarship

曖昧さの解消の
ためにURIを参照

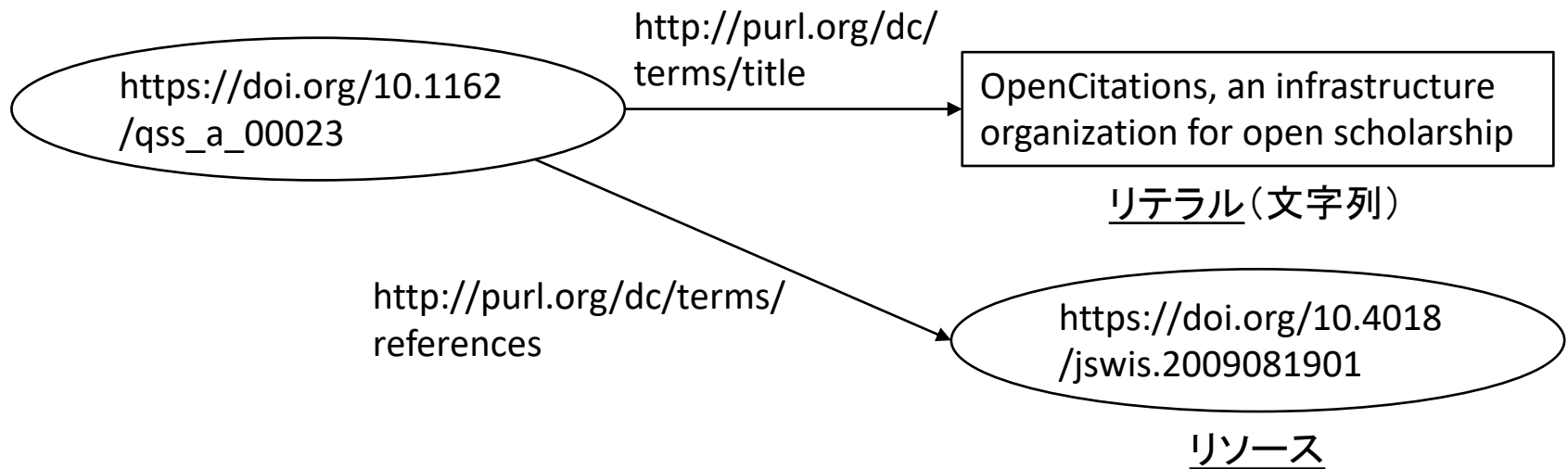


参考: 神崎正英. RDF -- リソース表現のフレームワーク. The Web KANZAKI.
<https://www.kanzaki.com/docs/sw/rdf-model.html>

RDF [2/2]

https://doi.org/10.1162/qss_a_00023は<https://doi.org/10.4018/jswis.2009081901>を引用しています。

主語 (Subject)	<u>リソース</u>	https://doi.org/10.1162/qss_a_00023
述語 (Predicate)	<u>プロパティ</u>	http://purl.org/dc/terms/references
目的語 (Object)	<u>プロパティの値</u>	https://doi.org/10.4018/jswis.2009081901



参考: 神崎正英. RDF -- リソース表現のフレームワーク. The Web KANZAKI.
<https://www.kanzaki.com/docs/sw/rdf-model.html>

プログラム [1/3]

OpenCitations: crossing one billion citations and beyond

Silvio Peroni (I4OC共同設立者/ボローニャ大学)

- OpenCitations: オープン・サイテーションの情報基盤の提供を行っている組織 (c.f. Initiative for Open Citations (I4OC) はオープン・サイテーションを促進するイニシアティブ)
- オープン・サイテーションの原則
- オープン・サイテーションの経緯
- OpenCitationsの活動と今後の計画

オープン・サイテーションの実践1: 京都大学図書館機構における

紀要論文のオープン・サイテーションに向けた取り組み

西岡千文 (京都大学)

- 京都大学学術情報リポジトリ (KURENAI) で公開されている紀要論文を対象としてオープン・サイテーションを試行
- 引用データの組織化と判明した課題
- 引用データの利活用 → KURENAIでの表示

プログラム [2/3]

オープン・サイテーションの実践2: Building the Humanities Citation Index (HuCI): why and how?

Matteo Romanello (スイス連邦工科大学ローザンヌ校)

Giovanni Colavizza (アムステルダム大学)

Silvio Peroni (I4OC共同設立者/ボローニャ大学)

- 人文学分野の引用索引の課題
- 人文学分野の引用索引が備えるべき特徴
- 図書館が果たすべき役割
- Humanities Citation Index (HuCI) の構成

プログラム [3/3]

Introduction to the Open Research Knowledge Graph

Lars Vogt, Sören Auer, Allard Oelen, Vitalis Wiens (ドイツ国立科学技術図書館)

- 学術情報流通が数世紀の間、それほど変化していないという指摘と課題
- ORKGにおける学術情報の表現と利用
- デモ: ORKGへの研究成果の登録、ORKGを利用した研究成果の比較、研究成果の可視化

What can University Libraries do for Open Research Knowledge Graph?

Sören Auer (ドイツ国立科学技術図書館)

- ORKGの重要性
- ORKGにおける図書館の役割
- 人文社会学におけるORKGの実装例
- ORKGの最新の機能
- 今後の計画

質疑応答

- 当日(2/15)の16～17時にリアルタイムで【第三部：質疑応答、ディスカッション】を開催
- 質問受付
 - 2/10(木)17時までは質問フォームにて受付
https://docs.google.com/forms/d/162xqAEBmprOKEdfg5zRc_2P_U_T5rHD9D5nQv_3wMPA
 - 当日も直接質問可能

講演を通して、将来の学術情報流通と図書館が担う役割について考えるきっかけとなればと思います。

