オープン・サイテーションとは？
ー欧米で進む引用データのオープン化、日本における現状と機関リポジトリへの影響ー

京都大学図書館機構オープンアクセス推進事業プロジェクトチーム

共同出展：京都大学東南アジア地域研究研究所・東南アジア研究の国際共同研究「東南アジア地域研究資料のオープン・サイエンス化に向けたとりくみ」

1. オープン・サイテーション

学術出版物の引用データは、研究評価、研究プロセスの理解、図書館の蔵書形成等、様々な目的で利用されている。しかし、引用データは従来は複雑なライセンスにより保護され自由なアクセスが難しい状況であった。この状況を打開するために、2017年4月に、学術機関と出版社によって、I4OC（Initiative for Open Citations）がオープン・サイテーション、すなわち引用データのオープン化を推進する国際的なイニシアティブとして設立された。I4OCは、機械可読なフォーマットでの引用データのオープン化を推進することで、引用データの利用可能性を高めることを目的としている。具体的には、Crossrefにメタデータとして登録された各文献の引用文献リスト、すなわち引用データをオープンにすることを推進している。I4OCでは、以下の3つの条件を満たす引用データが、オープン・サイテーションであるとしている。

- **構造的（Structured）**：RDF等の機械可読なフォーマットで表現されていること（図1）
- **分離可能（Separable）**：引用データが記載されている引用元文献にアクセスしなくても、引用データを入手可能であること
- **オープン（Open）**：無償でアクセス可能であり、再利用に際して制限がないこと

Crossref DOIが付与された約1億件の出版物のうち、2018年9月時点で24.20%の文献の引用データがオープンになっている。I4OCの活動によって公開された引用データは、CrossrefメタデータAPIあるいはI4OCの関連組織であるOpenCitationsが公開しているCOCI（OpenCitations Index of Crossref Open DOI-to-DOI references）というデータセットから取得可能である。

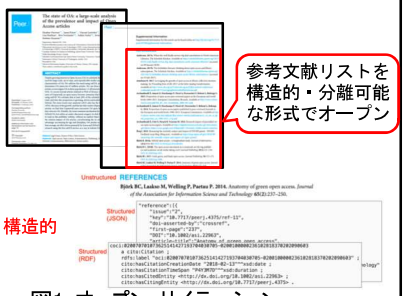

図1：オープン・サイテーション

2. 日本の学術出版物におけるオープン・サイテーションの現状

調査方法. 日本の学術出版物を「ジャパンリンクセンター（JaLC）によって付与されたDOIをもつ出版物」であると仮定する。I4OCは現在のところCrossref DOIをもつ文献のみを対象としているため、JaLC経由で登録されたCrossref DOIをもつ文献2,049,891件（JaLCメタデータ全文献のうち32.18%）を調査対象とする。COCIデータセットとCrossrefメタデータAPIを利用することで、調査対象の文献のオープン・サイテーション状態を下記のうちいずれかに分類する。

- **オープン**：引用文献リストをオープンにしている文献
- **クローズド**：引用文献リストを意図的に非公開にしている文献（出版者が引用文献リストを非公開に設定している）
- **未整理・非存在**：文献の引用文献リストが組織化されていない（未整理）あるいは引用文献が存在しない（非存在）の文献

調査結果. 主要な調査結果を下記に示す。

- 世界の学術出版物と比較すると日本の学術出版物では、引用データをオープン・クローズドにしている文献の割合が低く、未整理・非存在である文献の割合が高い（表1）。
- 化学・化学工業を筆頭にSTM（科学・医学・工学）分野の文献でオープン・サイテーションが進展している（表2）。対して、歴史・地理、哲学・宗教等の人文学系分野においては、引用データの未整理を原因として、遅れている。
- オープン・サイテーションが進展している学術誌はSTM分野に属する英文誌である。これらの学術誌のうち多くは出版プラットフォームとして商業出版社のプラットフォームを利用している。
- 過去10年間に出版された文献では、概ね40%以上の文献がオープン・サイテーションになっている（図2）。しかし、クローズドにしている文献の割合も増加傾向が観察される。過去の文献では、引用データが組織化されていないことが推察される。

表1：日本と世界の学術出版物におけるオープン・サイテーションの現状

	全文献件数	オープン		クローズド		未整理・非存在	
日本 [1]	2,049,891	386,632	(18.86)	148,130	(7.23)	1,515,129	(73.91)
世界 [2]	99,848,571	24,178,446	(24.22)	16,589,545	(16.61)	59,080,580	(59.17)

[1] JaLC メタデータ，Crossref メタデータ，unpaywall データセット全てに収録されている文献（≈JaLC 経由で登録された Crossref DOI をもつ文献）.
[2] Crossref メタデータかつ unpaywall データセットに収録されている文献.

表2：分類別のオープン・サイテーションの状況

	全文献件数	オープン		クローズド		未整理・非存在	
政治・法律・行政	4,735.00	99.50	(2.10)	55.00	(1.16)	4,580.50	(96.74)
経済	10,179.00	1,534.00	(15.07)	1,675.00	(16.46)	6,970.00	(68.47)
社会・労働	10,758.50	1,446.00	(13.44)	1,367.00	(12.71)	7,945.50	(73.85)
教育	17,621.00	3,234.00	(18.35)	1,991.50	(11.30)	12,395.50	(70.35)
歴史・地理	14,740.00	192.00	(1.30)	990.00	(6.72)	13,558.00	(91.98)
哲学・宗教	15,104.00	1.00	(0.01)	124.00	(0.82)	14,979.00	(99.17)
芸術・言語・文学	6,162.00	595.00	(9.66)	605.00	(9.82)	4,962.00	(80.53)
科学技術	164,957.75	28,172.00	(17.08)	11,250.25	(6.82)	125,535.50	(76.10)
建設工学・機械工学 [1]	248,957.25	34,152.50	(13.72)	26,390.75	(10.60)	188,414.00	(75.68)
化学・化学工業 [2]	364,817.75	95,301.50	(26.12)	27,415.25	(7.51	242,101.00	(66.36)
生物学・農林水産	192,386.00	35,434.00	(18.42)	17,268.00	(8.98)	139,684.00	(72.61)
心理学・医学・薬学 [3]	407,761.75	68,203.50	(16.73)	21,191.25	(5.20)	318,367.00	(78.08)
分類不明	591,711.00	118,267.00	(19.99)	37,807.00	(6.39)	435,637.00	(73.62)
合計	2,049,891.00	386,632.00	(18.86)	148,130.00	(7.23)	1,515,129.00	(73.91)

[1] 建設工学・建設業，機械工学・工業，運輸工学，電気工学・電気機械工業，原子力工学・工業が含まれる.
[2] 化学・化学工業，繊維工学・工業，食品工学・工業，金属工学・鉱山工学，印写工学，その他の工学・工業が含まれる.
[3] 人類学，心理学，医学，薬学が含まれる.

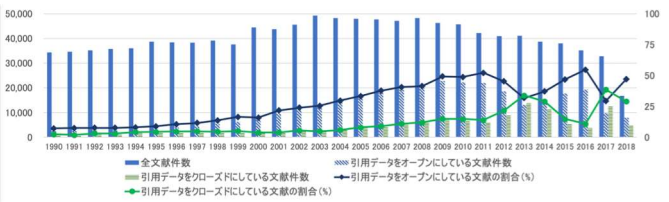

図2：出版年別のオープン・サイテーションの状況

3. 機関リポジトリへの展開に対する期待

大学などの研究機関が発行する学術雑誌である紀要は、特に人文社会学系分野において重要な研究成果の公表の場であり、機関リポジトリがプライマリな公開場所として利用されていることが多い。商業出版社のプラットフォームでは各文献のページにて引用文献や被引用文献へのリンクが付与されている場合が多く、科学的知見の検証可能性の向上に寄与している。対して、多くの機関リポジトリでは引用文献・被引用文献へのリンクは付与されておらず、同様の機能が望まれる。このような機能を検討するにあたっては引用データがオープンになっていることが望ましい。

調査対象の文献のうち、計6,572件の文献が大学の機関リポジトリで公開されており、そのうち33.70%が引用データをオープンにしているが、18.69%クローズドにしている。これらはいずれもオープンアクセスであることから、オープン・サイテーションの認知の向上が求められる。47.61%が未整理・非存在に分類されることから、引用データの組織化の支援が求められる。

オープン・サイテーションによって、大学図書館はコンテンツの利活用を把握・評価ができるだけではなく、コンテンツの引用文献・被引用文献のリンクによって利用者に対して知の体系を提示することができ、新たな研究支援につなげることが期待される。