

## 小規模データの長期保存とDOI登録：国立国会図書館 Web アーカイブ利用の提案

# 村山 泰啓 [1]; 能勢 正仁 [2]; 今井 弘二 [1]; 国武 学 [1]; 家森 俊彦 [3]; 渡邊 堯 [4]; 徳原 直子 [5]

[1] 情報通信研究機構; [2] 京大・理 地磁気センター; [3] 京大・理・地磁気センター; [4] WDS-IPO/NICT; [5] 国立国会図書館

### Proposal: Long-Term Preservation and Minting DOIs for Small Datasets by Web Archiving of National Diet Library of Japan

# Yasuhiro Murayama[1]; Masahito Nose[2]; Koji Imai[1]; Manabu Kunitake[1]; Toshihiko Iyemori[3]; Takashi Watanabe[4]; Naoko Tokuhara[5]

[1] NICT; [2] DACGSM, Kyoto Univ.; [3] WDC for Geomagnetism, Kyoto Univ.; [4] WDS-IPO/NICT; [5] National Diet Library

Open Data and Open Science are increasingly becoming hot topics, in parallel to establishing ICSU-WDS (2008), G8 Open Data Charter (2013), deployment of RDA (2013), and so forth, in addition to development of Open Access of journal articles. National guiding principle of open science has been released by the Cabinet Office (March 2015). In this context, long term preservation of research data is increasingly important and is being required. Minting DOI (digital object identifier) to datasets are also essential tool for discovery, reuse and citation (and bibliometrics) of datasets. This paper discusses possible use of WARP (Web Archiving Project) which is carried out by National Diet Library of Japan (NDL). NDL's mission includes long term preservation of electronic material collected from web sites of public institutions and universities, without definition of expiration in NDL by nature. Small-size datasets linked from lab. pages therein are often retrieved and preserved in WARP also, which enables data provides and users to browse, refer to, link to, and even mint DOI to, the datasets preserved in the WARP project for a long term. Possible collaboration with such an archival institution like NDL will be helpful and useful for research performing bodies like labs in research institutes and universities, where closer cooperation of their internal library/data management to NDL will be also important.

ジャーナルのオープンアクセス化と並行して研究データオープン化の議論が進む中、データの利用やデータ生成者の評価が重要な問題となってきている。ICSU-WDS 設立 (2008 年)、G8 オープンデータ憲章 (2013 年) などの節目を経てオープンサイエンスや科学と社会の関わりを視野に入れた議論が活性化している。2015 年 3 月には内閣府・総合科学技術イノベーション会議から、我が国のオープンサイエンスの基本方針が発表されている [内閣府、2015]。また研究不正への対応の視点から、一定期間の研究データの保存がも求められる [文部科学省、2014] など、学術研究活動をめぐって従来と異なる研究者の対応が不可欠になりつつある。これは国内だけでなく、国際的な議論が先導しているものも多く、研究コミュニティによって今後は本格的な対応が必要となると考えられる。

こうした対応の中で、まず研究機関で求められている対応の一つとして、上述の研究データの一定期間の保存、がある。例えば 5 年など一定期間の保存を義務化する動きがあるが、その後は法人文書のように破棄してよいのか、課題がある。ICSU は WDC システムを使って 1957-58 年から半世紀以上にわたって地磁気、電離層、宇宙線、など重要な科学データセットの保存事業を推進してきた。この前例をみても、重要な研究データには一研究者の個人資産ではなく人類の共有知的資産であるべきものはすくなくない。太陽地球系物理学においても STEP、S-RAMP をはじめとして超高層や磁気圏、太陽データ等の整理・保存を試みてきている。

さらに、2016 年日本で開催された G7 科学技術大臣会合で合意されたように、G7 国はオープンサイエンスについて前向きに取り組むことが求められている。データは発見可能で再利用可能なものであり、またデータ生成者は論文著者と同様にそれ以上に、重要な科学的成果をもたらした功績者として評価されることが重要である。

これを可能とするツールの一つとして、データに DOI (Digital Object Identifier) を付与して、文献中で引用する、という「データ引用 (data citation)」がある [DataCite, 2015; 村山・林、2014 など]。データセットに関する DOI を国際機関に登録することで、論文の DOI と同様に簡便な参照、引用が可能になり、また被引用度の算出などが飛躍的に容易となる。データ生成者・寄与者への評価尺度の開発が進展することが望まれている。

一方、DOI は persistent identifier (恒久的識別子) と呼ばれるように、登録後は削除しないことがのぞましい。出版論文を考えれば、論文本体のファイル (PDF など) も、その DOI も、削除されたり参照不可能になることは想定されていない。データセットと DOI はともに長期保存されることが、本来であれば必要である。しかし、近年は研究機関・大学等の組織やミッションの長期的な将来が不透明になりがちであり、過去のプロジェクトで生成されたデータを保存し続けると断言するのは容易なことではない。

一方、国立国会図書館 (以下、NDL) は法に基づき収集した資料の保存を行い、情報の廃棄を前提としていない。電子情報時代の今日、NDL は印刷資料だけでなく、電子情報資料の収集事業も開始している。その一つが、WARP (Web

Archiving Project) である。公的機関や許可を得た機関の外部公開 Web サイトを一定期間ごとにダウンロードして、過去の Web 情報が NDL の WARP サイトから参照可能となっている。WARP は本来、Web 上の文書にあたる情報の収集が本務であるが、Web 上の情報は種々雑多であるため、例えばアスキー形式で書かれた研究データが研究室のサイトからリンクされていれば、これもしばしば保存される。いったん NDL により保存されれば、これはその時点の Web 情報として今後も長期的に外部から参照可能である。

こうしたデータ保存事業を研究現場の情報管理と接続するためには、大学図書館などは今後、学術情報・成果の管理・発出の上で、ますます重要な役割を担うことが期待される。研究機関は、限られた研究資金や運営資金のなかで、今後、国内外でますます求められていくであろう学術データの保存・管理のために、情報保存を業務とするこうした外部機関との連携を重視し、学術研究が生み出す人類資産として、論文および研究データの取り扱いについて検討できるようにすると期待される。

#### 参考文献

内閣府、<http://www8.cao.go.jp/cstp/sonota/openscience/>、2015 年 3 月 30 日閲覧

村山泰啓, 林和弘, 2014, 146, p.12-17 : <http://hdl.handle.net/11035/2972>

DataCite、例えば <https://www.datacite.org/services/cite-your-data.html>、2015 年 8 月 4 日閲覧

文部科学省、2015、[http://www.mext.go.jp/b\\_menu/houdou/26/08/1351568.htm](http://www.mext.go.jp/b_menu/houdou/26/08/1351568.htm)、2016 年 8 月 17 日閲覧