

観測データから生成された2次データに関するDOI引用法の新提案

村山 泰啓 [1]; 能勢 正仁 [2]; 小山 幸伸 [3]; 武田 英明 [4]; 是津 耕司 [5]; 木下 武也 [5]; 家森 俊彦 [6]; 渡邊 堯 [7]
[1] 情報通信研究機構; [2] 京大・理 地磁気センター; [3] 融合研究センター
; [4] NII; [5] NICT; [6] 京大・理・地磁気センター; [7] WDS-IPO/NICT

Proposal of A New DOI Citation Method for Dataset Generated from Other Datasets

Yasuhiro Murayama[1]; Masahito Nose[2]; Yukinobu KOYAMA[3]; Hideaki Takeda[4]; Koji Zettsu[5]; Takenari Kinoshita[5]; Toshihiko Iyemori[6]; Takashi Watanabe[7]
[1] NICT; [2] DACGSM, Kyoto Univ.; [3] TRIC; [4] NII; [5] NICT; [6] WDC for Geomagnetism, Kyoto Univ.; [7] WDS-IPO/NICT

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)

It is really important to appropriately evaluate works by researchers and contributors of data creation and management (those personnel are called 'data producers' hereafter in the present paper). If shared data is used for other researchers' studies and publications without any credit or citation, the data producers may keep their data only for internal use. Practices for journal articles include proper credit and citation, which enable proper evaluation of the cited works according to citation number, and/or appropriate metrics. Similar mechanism and eco-cycle for data citation are necessary.

When a dataset is created from other datasets, referring the original ones and their data producers is an important practice. For example, the Dst index is frequently used and analyzed in geomagnetic research fields, as a measure of global-scale geomagnetic activity. The Dst index is the dataset produced from geomagnetic observation data at the world-wide four observatories, Hermanus (South Africa) of SANS, Kakioka (Japan) of Japan Meteorological Agency, Honolulu (Hawaii) and San Juan (Puerto Rico) both of US Geological Survey [Sugiura and Kamei, 1991]. The Dst index is the secondary data (SD), reduced from the original data (OD) of the four observatories. Observing/obtaining OD requires highly scientific/engineering expertise, which also have been developed and operated by researchers/scientists in many universities/institutes. Evaluating research products and their use is increasingly important to support the foundational work such as the continuous experiments and observations.

Experiments and practices using DOI (Digital Object Identifier) for data citation are now underway. DOI is one of persistent identifiers used for digital objects, which has been minted to published articles in past. DOI minting to data is increasingly recognized important for citation, accessibility, discoverability and reuse of data.

In the present paper, a new method is proposed for measuring the indirect use of OD when SD is cited, with DOIs. A DOI Registration Agency (RA) defines a metadata (MD) format (schema) for attribution information or metadata. Once DOIs of ODs are published as part of MD of SD, provenance of SD can be tracked via the OD-DOIs.

Metadata schema defined by DOI-RAs such as JaLC (Japan Link Center) and DataCite provides an element of 'relation', which can describe identifiers of related datasets and articles, specifying a type of relation. This element can be used to provide links between SD and ODs. An appropriate 'search tool software' will be a powerful mechanism to give information for metrics of indirect uses of ODs.

On the other hand, this kind of tool can become really useful only when a code of this kind of MD description is widely shared in the community. In 2014 IAGA (International Association of Geomagnetism and Aeronomy) has established a new task force for data citation of geomagnetic data. Such an academic coordination authority will work effectively for internationally shared practices among scientists. It is expected to enable a thorough world-wide search of data use and citation about datasets of hundreds of the world-wide geomagnetic observatories and their secondary datasets (indices such as Dst, Kp, K, and Ap and even more of various SDs). Also for other data products as well as empirical models and so on, the proposed mechanism is expected to be useful and beneficial for stakeholders including observatories, data centers, scientists, users, funders, and so on.

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]。

一方、オープンデータ (研究データ共有) を推進する場合、データ生成・整備に労力をかけた研究者・担当者 (ここでは「データ生成者」と呼ぶ) の活動評価は大変重要である。共有されたデータから他人が研究成果を出すだけでは、デー

タ生成者は報われずデータを囲い込む等が心配される。論文の場合は被引用数などで論文や著者の評価となる。同様にデータやデータ生成者が評価される仕組みが必要と考えられる [DataCite, 2015 ; 村山・林、2014]。

例えばある観測データから作られたデータ、複数のデータを統合・処理して得られるデータ等においては、元のデータの生成者を参照して評価するしくみは明確でない。例えば、地球電磁気学分野の数多くの論文で用いられる Dst 指数は、比較的の低緯度に分散した Hermanus (南アフリカ・南アフリカ国立宇宙機関)、柿岡 (日本・気象庁)、Honolulu (米・米国地質調査所)、San Juan (プエルトリコ・米国地質調査所) の 4 地点の地磁気観測データを解析・処理して得られる地磁気活動度の指標となる指数データの 1 つである [Sugiura and Kamei, 1991]。4 観測所から得るような元データを 1 次データ、そこから生成されるデータを 2 次データ、と仮に呼ぶことにする。1 次データ取得・整備はしばしば高度に理学的・工学的な活動で研究者自身が実施するケースも少なくない。今後、基盤的な実験・観測活動の資金的・人的資源確保が困難になるなかで、研究成果物全般の評価はますます重要といえる。

現在、DOI (Digital Object Identifier) をデータに付与して引用に役立てる取り組みが実施されつつある。DOI は恒久的識別子の 1 つであり、従来は主に論文に付与されてきた。別の形の研究成果物であるデータへの DOI 付与、その引用、検索等の促進が重要と言われている [DataCite, 2015 ; 村山・林、2014]。

2 次データ引用における 1 次データの間接的利用をカウントするために、本論文では以下のような DOI を用いた方式を提案したい。DOI 登録機関では、データの付帯情報を書き込むメタデータ (MD) の形式が定められる。1 次データへのアクセスを可能になるように 1 次データの DOI を 2 次データの MD に書き込んで公開し、これを利用する。これはデータのプロベナンス (provenance ; 典拠/来歴情報) を明示して利用することに他ならない。

JaLC (ジャパンリンクセンター) や DataCite といった DOI 登録機関の定めるメタデータには関係 (relation) を書くための項目が用意されていて、関係する他のデータや文献を識別子で指定し、また関係の種類も指定することができる。この項目を 1 次データ・2 次データの関係記述に利用できる。また、1 次・2 次データの MD を読み込む検索ツールを用意して、2 次データが引用された際に 1 次データを検索し利用数を計量する機能が有用となるだろう。

一方、こうした MD 記述の作法は、国際的に広く共有されてはじめて上記のツールが有用となる。地磁気分野では幸い 2014 年、IAGA (国際地磁気・高層物理学会) にデータ引用に関するタスクフォースが設置された。こうした場で国際標準的な方法が認証されれば、広く地球上各地の地磁気データとその 2 次データ (地磁気だけでも Dst、Kp、K、Ap 等の指数が一般的に用いられる) が、互いにどのような利活用をされているか、が一目でわかり、また機関の評価などでも活用可能になる。こうした活動は測地学分野でも重要性が指摘する声があり、さまざまな研究分野で有用性があると考えられる。

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

Sugiura, M. and T. Kamei, IAGA Bulletin. 40, 1991.