

S001-12

Zoom meeting A : 11/2 PM1 (13:45-15:30)

14:45-15:00

データ引用に向けた科学データリポジトリの開発: 名古屋大・宇地研・統合データサイエンスセンターでの活動

#堀 智昭¹⁾, 三好 由純¹⁾, 能勢 正仁²⁾, 田 采祐³⁾, 中村 紗都子⁴⁾, 今城 峻⁵⁾, 北原 理弘⁶⁾, 小路 真史¹⁾, 前田 麻代¹⁾, 梅村 宜生²⁾, 瀬川 朋紀¹⁾, 塩川 和夫⁷⁾, 篠原 育⁸⁾, 栗田 怜¹⁾, 今田 晋亮⁵⁾, 増田 智⁹⁾

¹⁾名大 ISEE, ²⁾名大・宇地研, ³⁾名大 ISEE 研, ⁴⁾京大生存研, ⁵⁾名大・ISEE, ⁶⁾東北大・理・地球物理, ⁷⁾名大宇地研, ⁸⁾宇宙研 / 宇宙機構, ⁹⁾名大 STE 研

Development of science data repository for data citation: CIDAS/ISEE activities

#Tomoaki Hori¹⁾, Yoshizumi Miyoshi¹⁾, Masahito Nose²⁾, Chae-Woo Jun³⁾, Satoko Nakamura⁴⁾, Shun Imajo⁵⁾, Masahiro Kitahara⁶⁾, Masafumi Shoji¹⁾, Asayo Maeda¹⁾, Norio Umemura²⁾, Tomonori Segawa¹⁾, Kazuo Shiokawa⁷⁾, Iku Shinohara⁸⁾, Satoshi Kurita¹⁾, Shinsuke Imada⁵⁾, Satoshi Masuda⁹⁾

¹⁾ISEE, Nagoya Univ., ²⁾ISEE, Nagoya Univ., ³⁾ISEE, Nagoya Univ., ⁴⁾RISH, Kyoto Univ., ⁵⁾ISEE, Nagoya Univ., ⁶⁾Dept. Geophys., Grad. Sch. Sci., Tohoku Univ., ⁷⁾ISEE, Nagoya Univ., ⁸⁾ISAS/JAXA, ⁹⁾STEL, Nagoya Univ.

The Center for Integrated Data Science (CIDAS) in the Institute for Space-Earth Environmental Research (ISEE), Nagoya University has conducted scientific research projects, and also developed and provided research databases and infrastructures for the science community to promote science for space and solar-terrestrial systems. Recently ISEE has become a member of the Japan Link Center (JaLC), capable of assigning digital object identifiers (DOIs) to scientific data obtained or archived by the institute. As a kind of pilot project for DOI minting in CIDAS, we have just started to work on space plasma, field, other related observational data yielded by the Exploration of energization and Radiation in Geospace (ERG) project. The ERG-science center (ERG-SC) /CIDAS that has been co-operated by the Institute of Space and Astronautical Science / Japan Aerospace Exploration Agency (ISAS/JAXA), archives the ERG project data and have made them available to the science community. After some discussion and trials, a pre-existing scheme, which was originally developed by ISAS to generate metadata in the JaLC schema and a landing page in HTML, is slightly modified and then employed in our workflow to prepare a JaLC metadata file (necessary to register a DOI) and landing page for each of all publicly available scientific data products from the ERG project. A DOI assigned for an ERG data product refers (through its landing page) to the entire data files for the particular data product archived in the ERG-SC data repository. This way of minting DOIs thus enables the author of scientific articles to cite specific datasets in their papers but does not allow for citing a specific data file of the particular data version which was used in the articles. To redeem the gap, ERG-SC requests the data users to describe the version number of each dataset in the acknowledgment section of their journal papers. Thus, a combination of the DOI and version number described in an article provides us with a comprehensive way of data citation uniquely linking the article and its used data up to the level of individual data file. This is a working scheme that we have developed so far and hopefully would be found useful by the research community to consider the best practice of data citation supporting open science.

名古屋大学宇宙地球環境研究所(ISEE)・統合データサイエンスセンター(CIDAS)では、独自の研究プロジェクトを推進すると同時に、宇宙・太陽地球系システムを研究するための科学データベース及び各種インフラストラクチャを開発して科学コミュニティに提供している。このほど ISEE はジャパンリンクセンター(JaLC)の正会員となり、CIDAS が実作業を担当することで、研究所で取得・アーカイブしている科学データに独自のデジタルオブジェクト識別子(Digital object identifier: DOI)を付与することができるようになった。そのための CIDAS でのパイロット的な試みとして、ジオスペース探査プロジェクト(ERG プロジェクト)の宇宙プラズマ・電磁場データやその他の関連観測データへの DOI 付与を開始した。この ERG プロジェクトデータは、日本宇宙航空研究開発機構の宇宙科学研究所(JAXA/ISAS)と ISEE とで共同運用する ERG サイエンスセンター(ERG-SC)のよってアーカイブされ、ERG-SC データリポジトリより科学コミュニティに向けて公開されている。DOI 付与・管理に関する所内での議論・試験を経て、元々 JAXA/ISAS で開発された手法を改変することで ISEE でのニーズに適合させることで、JaLC での DOI 登録に必要ないわゆる JaLC スキーマのメタデータと、DOI のリンク先となるランディングページ(HTML ファイル)を同時生成する仕組みを開発した。この仕組みを用いて、ERG プロジェクトの公開データプロダクト全種類について、DOI 登録とランディングページ準備を進めている。DOI 付与の粒度としては、単一の科学データプロダクトに属するデータファイル全体に対して、1つの DOI が割り振るようにしている。これにより、科学論文中で特定のデータプロダクト全体を引用できるようになるが、実際の研究に使用される、そのデータセット内の特定のデータファイルをユニークに引用することはできないという問題が生じる。これは、同一データセット内の同一日時データのデータであっても、データ較正や修正・修復の過程を経て、複数のバージョンが存在し得るためである。この問題を解決して研究の源泉である元データの追跡性を担保するため、ERG-SC では、ERG プロジェクトデータを科学論文中で使用の際に、謝辞のところに各データセットのバージョン番号を記載するように要請している。このように DOI を用いた使用データプロダクトの特定と、データセットのバージョン番号の記録を合わせることで、個々のデータファイルのレベルまで、科学論文と源泉データを紐付けるデータ引用を確立させることができる。これが現時点で我々が開発・

推進しているデータ引用スキームであるが、この試みが、これからのデータ引用、ひいてはオープンサイエンスを推進していく上でのベストプラクティスに関する議論をしていく際の一助となればと考える次第である。