

## C. C4.0 に基づいた環境放射能データの DOI を付与しての出版と、

### PHP・PostgreSQL を用いた出版データ検索サイトの構築

谷田直子<sup>a</sup>

筑波大学アイソトープ環境動態研究センター(契約技術職員)

〒305-8573 茨城県つくば市天王台 1-1-1

#### 概要

各科学者が手元に持つ環境放射能データにデジタルオブジェクト識別子(DOI)を付与し出版を開始した。さらにそのデータセットを串刺し検索できるシステムを構築し公開している。

本報告書では、筆者が取り組んだ内容を「DOI 登録出版」、「データベースの構築とプログラミング」に分けて解説する。また当該ワークフローを学外研究機関との協業プロジェクトにおいて横展開する取り組みについても触れる。

**キーワード:** 環境放射能、データベース、東京電力福島第一原子力発電所、Web サイト構築

#### 1. はじめに

東日本大震災から 10 年が経ち、その記憶が薄れつつある。そのような情勢において、東京電力福島第一原子力発電所の事故(以降福島事故)により環境中に放出された放射能についてのデータを収集し、できる限り 1 か所に集めた状態で自由に利用できる形で後世に残し伝えることは、学術的観点において意義深いことである。

従来、実験や調査などの研究過程で得られるデータは科学者自身が手元に保存し、論文執筆に利用した後は埋もれてしまいがちであった。筑波大学など 3 大学 3 研究機関が参加する放射能環境動態・影響評価ネットワーク共同研究拠点・共同研究(ERAN)の事業のひとつとして、福島事故に由来する環境試料のカタログやデータをアーカイブし出版する事業が 2019 年から始まった。この事業を担う筑波大学アイソトープ環境動態研究センターでは、これらのデータを広く活用できるような体制を整えることを喫緊の課題のひとつと認識し、データアーカイブワーキンググループを立ち上げた。

ワーキンググループは以下の 2 つを進めることとした。ひとつは、複数の同じタイプのデータをまとめたデータセットに DOI を付与して出版し、クリエイティブコモンズ 4.0<sup>1</sup>のライセンスのもとに永続的に公開することだ。もうひとつは ERAN の当初目標

には無かったが、当センターで出版したデータセットに加え、国内外の研究施設や省庁の持つデータセットも合わせて、時間空間情報と試料の種類・核種などをキーとしてすべてのデータを串刺し検索できる仕組みを構築し、世界中どこからでも誰でも検索してその結果を使うことができるようにすることである。

筆者は当該ワーキンググループを率いる青山道夫客員教授の指導の下、未経験ながらもプログラミングやデータベース構築などを含む全工程に携わることとなった。

#### 2. DOI 登録出版

##### 2.1 DOI とは

DOI とは Digital Object Identifier の頭文字を並べたもので、その名の通り、インターネット上にあるコンテンツに与えられた識別子であり、ブラウザで入力して検索すると、URI に変換され、DOI に紐づけられたコンテンツに移動することができる。URL と似ているが、URL は可変であり消滅することもある。しかし DOI は不変であり、コンテンツへの永続的なアクセスを提供する。また出版者にはコンテンツへの永続的なアクセスを提供する義務が課せられている<sup>2</sup>。

##### 2.2 DOI の構造

DOI は出版者(この場合は当センター)に特有の「プレフィクス」と、個々のコンテンツに与えられた「サフィックス」で構成される。

プレフィクス                      サフィックス  
10.34355/CRiED.U.Tsukuba.00001

図 1. DOI 例

プレフィクスである「10.34355」は DOI の登録機関(Registration Agency)であるジャパンリンクセンター(JaLC)から付与された、当センターに特有の識別子である。また、当センターではデータセットの主要著者の所属機関あるいはデータ提供機関に応じてサ

<sup>a</sup> otani.naoko.fu@un.tsukuba.ac.jp

<sup>1</sup> <https://creativecommons.org/licenses/by/4.0/deed.ja>

<sup>2</sup>

[https://www.doi.org/doi\\_handbook/translations/japanese/doi\\_handbook/1\\_Introduction.html](https://www.doi.org/doi_handbook/translations/japanese/doi_handbook/1_Introduction.html)

フィックスの前半部分を用意し、下5桁はすべてのデータセットを連番で管理をしている。

いくつかの例を示す。

10.34355/Forest.Agency.00049  
林野庁のプロジェクトのデータ

10.34355/Fukushima.NIES.00023  
国立環境研究所福島支所所属研究者が投稿したデータ

10.34355/IAEA.RAS5084.00091  
IAEA の地域プロジェクトの参加研究者が投稿したデータ

これにはデータの投稿者に敬意を表す意味合いと、一見しただけでどこのデータかがわかる利点もある。

## 2.3 DOI 出版までの流れ

学会発表等での周知や直接の呼びかけに応じた科学者に情報の提供を依頼し、集まったデータを点検する。データの形式や内容については著者の意向を最優先とするが、串刺し検索に必要ないくつかの情報(時間情報、位置情報など)が必須項目として欠けている場合は提供を促す。また、使用される単位は国際単位系 SI<sup>3</sup>とし、これが使用されていない場合は使用と数値の換算を依頼し、国際単位系に合わせたデータセット構成としている。

また、出版されるデータセットには、データそのもののほかに、ris ファイル、metadata ファイル、format\_description ファイルも含まれる。ris ファイルとは書誌情報を含んだ標準化されたテキストファイルで、多くの書誌情報引用プログラム、ソフトウェアが採用している。metadata ファイルには ris ファイルに含まれる情報のほか、位置情報などのデータセットの全容が記されている。format\_description ファイルには、data ファイルの各カラムの説明が記されている。

いずれも国内外で広く活用されることを念頭に置き、すべて英語で情報提供している。

## 2.4 データセットの Web ページ

DOI 出版が完了し DOI を管理するサーバーに登録されると、DOI を検索窓から入力することで、登録時に指定したデータセットをダウンロード可能な Web サイトへ飛ぶことができるようになる。Web サイトは先述の metadata ファイルの内容に加え、図表(Graphical Abstract)の掲載を推奨している。

また、データが多く必要とされる人の手により届きやすくなるよう、Web サイトのソースコードには「仕掛け」が施されている。ソースコードに Google 構造化データ<sup>4</sup>を埋め込むことで、Google が提供する Dataset Search<sup>5</sup>の検索対象となるため、これを利用している。metadata ファイルに含まれる情報を JSON 形式でマークアップし、html ファイルの header タグ内に記載している。このマークアップは Google のマークアップ支援ツールで生成することもできるが、テキストエディタで書くこともできる。

## 2.5 DOI 出版後の動向

出版後のデータセット利用状況の把握のために、ダウンロード数記録プログラムを利用している。このプログラムは Web 上で配布されている php で書かれた無償プログラムをダウンロードしカスタマイズしたものをサーバーにアップロードした。データセットがダウンロードされるたびに、カウントする仕組みになっている。これを月初めに確認しダウンロードし、月次の統計を取っている。

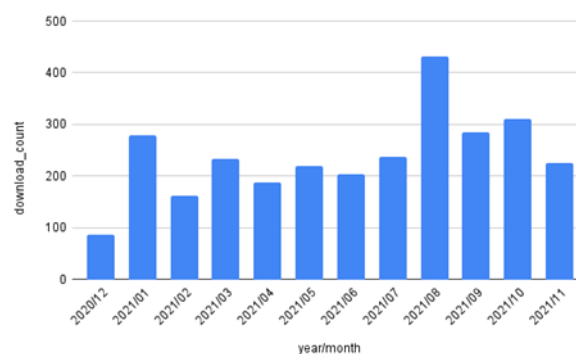


図2. データセットダウンロード数の推移

データセット出版数の増加に伴い、図2に示されるように、ダウンロード数も増加する傾向にある。

また、データセットリリース直後はアクセス数が多い傾向にあり、ダウンロード数と比例する。内容別に見ると、福島事故の影響が強い海域で採取された魚についてのデータセットは、関心度の高さを反映してかコンスタントにダウンロードがある。福島事故について書かれた論文の情報を集めたデータセットもダウンロード数が多く、利便性が評価されている傾向が見える。

またデータセット1件当たりの月ごとのダウンロード数を見ると、どのデータセットも増減はあるが継続的にダウンロードされていて、本データセットのサイトの認知度が上がっていると言える。

<sup>3</sup> <https://unit.aist.go.jp/nmij/library/units/si/>

<sup>4</sup> <https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data?hl=ja>

<sup>5</sup> <https://datasetsearch.research.google.com/>

### 3. データベースの構築とプログラミング

#### 3.1 データベースとは

データベースとは、検索と蓄積を整然と行うことができるように整理された情報の集まりを示す。身近なところでは、銀行の口座情報などが挙げられる。ここでは、個々の情報を「レコード」、固有の目的のもと収集されたレコードの集まりを「データセット」とし、データベースにはデータセットの枠を取り除いたレコードを格納する。これによりデータセットの枠を超えて、レコードを検索することができるようになる。このデータセットの枠を超えた検索を、特に「串刺し検索」と呼ぶことが多い。

#### 3.2 データベースを扱う技術

データベースを扱うソフトウェアとして PostgreSQL を採用した。このソフトウェアはリレーショナル型データベースソフトのひとつで、無償で利用できることもあり 30 年以上にわたり世界中で広く利用されている。

PostgreSQL を動かすサーバーは筑波大学の専用レンタルサーバーを利用しており、サーバー OS は使用事例がネットで多く見られることから CentOS8 を選択した。

#### 3.3 Web サイトの技術

データベースへのアクセス窓口となる Web サイトも PostgreSQL と同じサーバーに設置している。HTTP サーバーは Apache をインストールした。Web ページはユーザーの入力に応じてデータベースを検索し、そして出力結果を表示するために、動的に生成することが必要となる。このためのプログラミング言語は PHP を選択した。ネットワーク監視システムは Zabbix を導入している。

いずれも、オープンソースであったり GPLv2 ライセンスで提供されていたりなどして、無償で利用できるソフトウェアを使用している。

#### 3.4 Web サイトの構築と運用

Web サイトは 2 本立てで運営している。ひとつは個々にデータセットをダウンロードできるサイトであり、こちらはアイソトープ環境動態研究センターの公式ウェブサイトが置かれているサーバーに置いている。もうひとつの検索サイトを前述の専用レンタルサーバーに置いている。

いずれの Web サイトもトップページ以外は、エディターで html を手書きして作成している。これはシンプルな構造にすることで、素早く動作することを目的としている。またブラウザに依存しない基本

的な技術のみを使うことで、どのような環境に置いても、想定した動作をすることが期待できる。

#### 3.5 個々にデータセットをダウンロードできるサイト

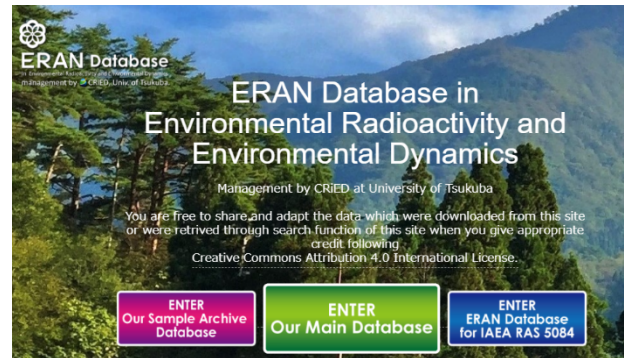
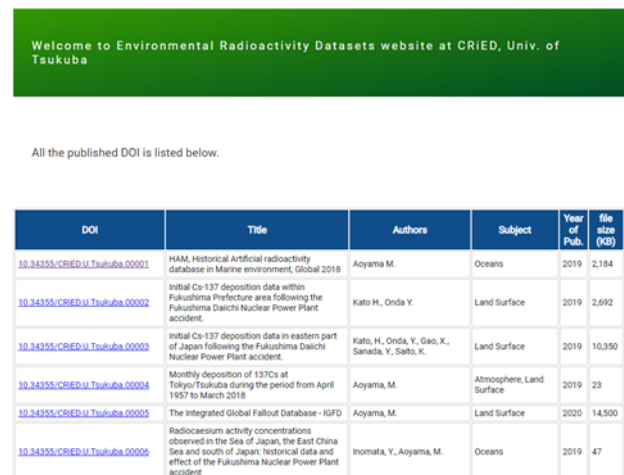


図 3. データベース Web サイト(抜粋)

図 3 の Web サイトが入口となり、現在運用している 3 つのデータベースへ誘導している。中央の「Main Database」は環境放射能のデータを扱うデータセットのトップページ<sup>7</sup>に導く。左の「Sample Archive Database」は、ERAN に所属する機関が保有する環境試料データセット出版のトップページ<sup>8</sup>に導く。右の「ERAN Database for IAEA RAS 5084」は国際原子力機関(IAEA)のプロジェクトである RAS5084 (Assessing and Improving Soil and Water Quality to Minimize Land Degradation and Enhance Crop Productivity Using Nuclear Techniques)に関連するデータセット出版のトップページ<sup>9</sup>に導く。

いずれも、データセットの DOI、タイトル、著者、分野、出版年、データセットのサイズが表形式で一覧できる。



DOI	Title	Authors	Subject	Year of Pub.	file size (KB)
10.34355/CRIED.U.Tsukuba.00001	HAM, Historical Artificial radioactivity database in Marine environment, Global 2018	Aoyama M.	Oceans	2019	2,184
10.34355/CRIED.U.Tsukuba.00002	Initial Co-137 deposition data within Fukushima Prefecture area following the Fukushima Daiichi Nuclear Power Plant accident.	Kato H., Onda Y.	Land Surface	2019	2,692
10.34355/CRIED.U.Tsukuba.00003	Initial Co-137 deposition data in eastern part of Japan following the Fukushima Daiichi Nuclear Power Plant accident.	Kato, H., Onda, Y., Gao, X., Sanada, Y., Saito, K.	Land Surface	2019	10,350
10.34355/CRIED.U.Tsukuba.00004	Monthly deposition of 137Cs at Tokyo/Tsukuba during the period from April 1957 to March 2018.	Aoyama, M.	Atmosphere, Land Surface	2019	23
10.34355/CRIED.U.Tsukuba.00005	The Integrated Global Fallout Database - IGFD	Aoyama, M.	Land Surface	2020	14,500
10.34355/CRIED.U.Tsukuba.00006	Radiocaesium activity concentrations observed in the Sea of Japan, the East China Sea and south of Japan: historical data and effect of the Fukushima Nuclear Power Plant accident	Inomata, Y., Aoyama, M.	Oceans	2019	47

図 4. リスト表示されたデータセット

図 4 の「DOI」の項目をクリックすると図 5 のように個別のページが表示される。

<sup>6</sup> <https://www.ied.tsukuba.ac.jp/database/index.html>

<sup>7</sup> <https://www.ied.tsukuba.ac.jp/database/databaselist.html>

<sup>8</sup> <https://www.ied.tsukuba.ac.jp/database/sampledb.html>

<sup>9</sup> <https://www.ied.tsukuba.ac.jp/database/iaeadb>

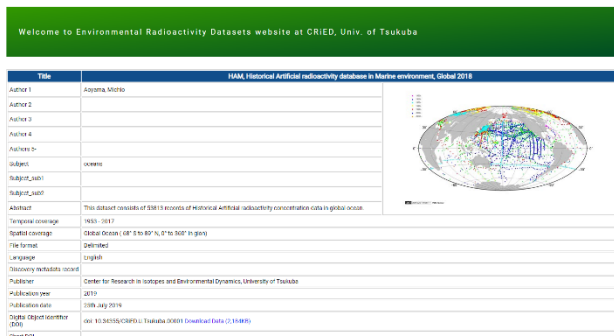


図 5. データセット個別ページの一例

「Download Data」をクリックするとデータセット一式がダウンロードされる。

### 3.6 データ検索サイト

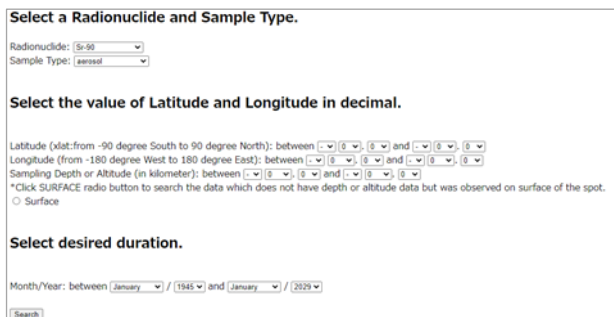


図 6. データ検索サイト

「Main Database」をクリックすると表示されるリストページ<sup>7</sup>には、緯度経度情報の入力方法が異なる 2 つの方法によるデータ検索システムへのリンクが張られている。そのひとつである「Selecting latitude and longitude by drop down list」をクリックすると図 6 のようなデータ検索サイトに導かれる。ここで、検索したい核種、対象となる試料タイプ(Water、Soil、Aerosol など)、緯度経度、高度(深度)、期間をリストから選択し、「Search」から検索する。すると図 7 のような画面が表示される。もう一方は、緯度経度は地図上で場所を選択することで検索できるようになっている。

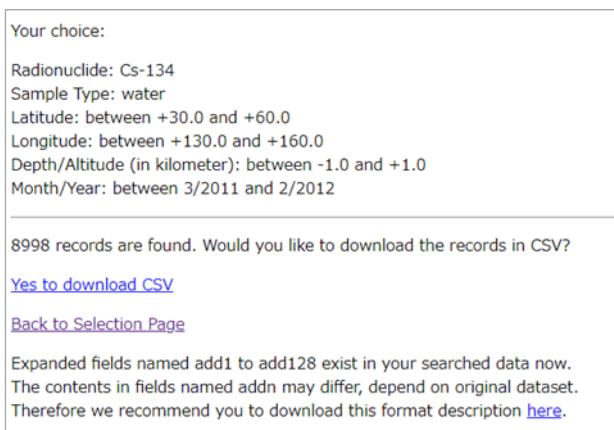


図 7. データ検索結果表示

検索結果画面では、検索条件に加えて、何件のデータがヒットしたかが示される。検索結果は CSV ファイルでダウンロードができる。

## 4. 今後の展望

データベース事業を開始してから 3 年が経ち、扱うデータセットが広がりを見せている。

IAEA の海洋に関する放射能データを扱うデータベース「MARIS」<sup>10</sup> のデータも取り込み、串刺し検索ができるように組み込んでいる。また、国内の省庁、研究機関、企業(東京電力など)が Web で公開している環境放射能データについても有用であると認識したため、当データベースに情報を取り込むこととした。これらの放射能データは日本語であったり、PDF であったりして、汎用性が低い状態であり、著作権の法律と当該 Web ページに記載されている使用の条件に従って、必要に応じて個別に利用許可を得て、出版にこぎつけた。これらの情報のデジタル化と出版については、データ内に「公開済みのデータを、許可を得て DOI を付与して出版した」旨を書き添えている。

さらに、林野庁や IAEA のプロジェクトに関わるデータ出版の取り扱いが始まるなど、開始当時の取り組みが横展開を見せている。

2021 年よりモデル計算データについても出版を検討していたが、データセットによっては数テラバイトのデータサイズとなることから、どのようにデータを提供するかなど検討すべき事項が多くあった。しかし、出版して残すことは必要との判断からモデル計算データを持っている研究者等との議論を開始し、本格的にデータ出版を進めることとした。

データの配布は他のデータと同様に Web 経由で配布することとし、データの受付は申し出を受け、物理 HDD を送付し、送り返してもらう方法を想定している。データの配布については専用のデータサーバーを立ち上げ、実用的な速度が出るかのテストを並行して進めており、実用的であれば、オンラインでの配布を開始する予定である。

<sup>10</sup> <https://maris.iaea.org/home>

# Publication of environmental radioactivity data with DOI under C.C4.0 licence, and development of data retrieval Web service system using PHP / PostgreSQL

Naoko TANITA<sup>a)</sup>

Center for Research in Isotopes and Environmental Dynamics, University of Tsukuba,  
1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan

Our project started providing the digital object identifier (DOI) to the environmental radioactivity data that each scientist had. Furthermore, we have developed and released a Web system that allows people to cross-search the data set. In this report, the contents of the author's efforts are divided into "DOI registration and publication" and "database development and programming". It also touches on efforts to expand this workflow in collaborative projects with other research institutes.

**Keywords:** Environmental radioactivity, database, TEPCO FDNPP1 accident, Website development

---

<sup>a)</sup> otani.naoko.fu@un.tsukuba.ac.jp