

第1回 SPARC Japan セミナー2018

「データ利活用ポリシーと研究者・ライブラリアンの役割」

国立環境研究所データポリシーと データ公開に向けた試み

白井 知子

(国立環境研究所 地球環境研究センター)

講演要旨



国立環境研究所では、幅広い分野の環境研究を学際的かつ総合的に進めている。環境研究は多岐に渡るため、使われるデータも分野ごとに特徴があり、データを公開することへの関心や懸念も異なる。そのような状況の下、研究所は、平成29年4月、「国立環境研究所データの公開に関する基本方針（データポリシー）」を策定、公開した。データポリシーの紹介および策定に至る背景、研究所におけるデータ公開に向けた試みについて紹介する。



白井 知子

国立環境研究所地球環境研究センター 主任研究員。1999年東京大学理学系研究科博士課程修了。同年宇宙航空研究開発機構地球観測研究部門にて、上空大気の航空機観測や、宇宙からの地球観測センサ等を担当。2001年～2002年、米カリフォルニア大学アーバイン校に滞在し、越境大気汚染等について共同研究を行う。2004年以降、国立環境研究所にて、温室効果ガスの排出・吸収状況と、その大気中での挙動についての研究を行う。2011年からは、地球環境データのデータベース化やデータ利活用の推進にも携わっている。2018年4月～9月、企画部研究企画主幹兼務。

国立環境研究所（NIES）は、国立研究開発法人として既にデータポリシーを持っている数少ない研究所です。本日はその策定の経緯と論点、現場の例、NIES においてデータを公開していくときの課題について述べたいと思います。

自己紹介・個人的動機

私は大気化学の研究をしており、最初は真空装置を使って微量成分の分析をしていました。大気球実験で成層圏の空気を採って分析など、だんだん内容が広がってきて、データ解析の魅力に取りつかれました。その後、旧宇宙開発事業団（NASDA、現 JAXA）に就職し、地球観測部門で、航空機観測実験をしたり、衛星のプロジェクトに関わったり、アメリカの大学に滞在

して共同研究をしたりしていました。

主に観測に携わっていた際、大量に得られるデータを使い尽くせないところにフラストレーションを覚え、だんだん解析や計算などを自分でやるようになってきて、両方ずっと続けることは大変なので、今は主に解析や計算をしているという状況です。

一方で、私は「地球環境データ統合解析推進室」というところに所属があり、主に観測により得られたデータのデータベースを作って発信していくという業務も行っています。また、4月から半年間、企画部も兼務しており、今回ご紹介するデータポリシーは企画部中心に作られたものです。

個人的動機としては、観測とモデルと両方やっていると、観測の方は観測で忙しいし、論文も書かなけれ

ばいけないし、次の観測のプロポーザルも書かなければいけないということで、とにかく余裕がないです。一方、モデル研究者は、使いやすいデータはないかといつもサーチしています。どういうものが使いやすいデータかというと、欠損がなく、グリッド化されているものです。観測をしている人は、他人のためにデータ整備や発信をするのはどうも後回しになってしまいますし、時空間欠損がない観測というのは非常にまれです。かといって、モデラーがこういったデータに手を付けて整備するかというと、世界を見回すと、整備されたデータセットが見つかることが多いので、一番整理されているデータがいつも使われることになり、日が当たらないデータはずっと日が当たらないという状況になります。そのような背景から、そこをつなぐことに私は非常にやりがいを感じています。

NIES としてのデータポリシー

NIES はつくば市にあり、九つの研究実施部門と三つの管理部門があります。

(<http://www.nies.go.jp/sosiki/index.html>)

私がいるのは研究実施部門の一つである地球環境研究センターです。今回、ご紹介する、研究所のデータポリシーは主に管理部門が中心になって作ったというところが一つポイントです。

策定のきっかけは、ある研究成果に係る全ての情報について、総務部に開示要求が来たことでした。開示要求については、法人文書に関しては情報公開法で決められていますが、研究データに関しては何ををもって「全ての情報」とするかの定義がはっきりしていません。実験ノートに至るところから全てのデータだと言いはじめるときりがありませんし、担当部門で対応に困ったことがあり、同様のケースに備えるとともに、研究データの公開を促進するためにも、研究データの公開について定めたデータポリシーを策定しておくことが非常に大事であるということから、策定への流れになりました。

策定の経緯は、平成 28 年 9 月に理事から企画部に

データポリシー作成の指示が来て、10 月に企画部で素案を作成し、11 月にオープンサイエンスとの切り分けについて、オープンサイエンスを別途進めている環境情報部と相談し、整理しました。12 月に所内幹部で議論した後、平成 29 年 2 月に方針および原案を所内に周知し、意見を募りました。その際、データの帰属・利用条件に意見が集中しましたが、これについては後で詳しくご紹介します。3 月に弁護士に確認し、4 月にデータポリシー策定・公開となりました。

ポイントは、半年ぐらいで作られたということです。やる気になればできるという言い方もできますし、そうやって作られたデータポリシーなので、先ほどの内閣府のデータポリシーガイドラインのご発表を聞きながら、NIES のデータポリシーがどの項目をカバーしているかをチェックしてみたのですが、カバーしているものもあれば、全くしていないものもあり、部分的にカバーしているものもあるという状況です。

企画部、環境情報部、総務部で意見交換して決まったことは、「データポリシーはふわっとしたもので、NIES の姿勢を示す所信表明のようなものとする」「研究不正の観点でデータを公開するのは異なり、外部で積極的にデータを利用してもらうための公開を対象とする」「データの技術的な方法ではなく、理念を示すことが主眼」「細かな規定は別に定める。最小限必要なことを記載する」「オープンサイエンス推進のために情報部で検討している内容は、論文等を公開する機関リポジトリをはじめ、扱う範囲は幅広いので、オープンサイエンスの方向性とデータポリシー策定は方向性が同じということで、矛盾していなければ、各々議論を進めてもよい」ということです。そして、オープンサイエンスについては別途ワーキンググループを設けて検討することになりました。オープンサイエンスの方が進みが遅いので、もし一緒に進めていたら、データポリシーができるのはもう少し遅れていた気がします。

データポリシーの素案策定のときに参考にした文書やポリシーは、2015 年の JST 科学技術情報委員会の

「データシェアリングのあり方に関する提言」など、図1に示した通りです。機関全体としてのポリシーはなくても、「〇〇センター」というレベルではポリシーがあるところが比較的あり、そのようなところの個別データポリシーなども参考にしながら論点を絞っていき、素案を作りました。

データポリシー素案作成の論点は、例えば「1.目的」では、「積極的にデータ公開し、国の研究開発成果の最大化に資する方針でよいか」でした。決まった方針は、「政府の基本方針はデータが自由に利用できること。NIESのデータは税金で作成したもので、自由に使うことが大原則。ただし、状況により商用利用や改変への制限はあり得る」でした。

「2.公開するデータの範囲」についての主な論点は、「公開するデータの範囲は適当か」でした。これは、策定のきっかけとして、「公開範囲に制限を設けないと対応に困る」ということがありましたので、「公開するデータの範囲にやや制限を加える」という方針で進めています。

「3.データの品質・管理」についての論点は、「品質という曖昧な言い方でよいか。精度、真度、利用しやすさなどに言及するか」「トレーサビリティについて言及するか」で、基本方針は、言及するということにし、弁護士のアドバイスで「非保証」を入れました。

「4.データの帰属・利用条件」は、所内で意見を募った際に、一番関心が高かったところです。論点として、「知的財産権に言及するだけでよい」「他機関と

の共同研究が行われたケースについて、言及する必要があるか」「データを利用する際に、引用元（NIES）を示すことを義務として記載するか。その記載例を示すか」「データの商業利用について、利用料を課すなどの条件を加えるか」について話し合いました。

まずは、知的財産権についてですが、データというのは著作物とは違って、著作権がはっきりと定められておらず、単純なデータの羅列は著作権法の保護対象ではありません。体系的に整理されたデータベースであって、創作性が認められるものは保護対象となるのですが、過去の裁判例でも著作権が認められたり認められなかったりと、曖昧なので、そこに触れはじめるのが記載が煩雑になるということで、著作権についてはあえて触れていません。

また、公開したデータがこちらの意図しない形で二次利用されたり、商業利用されたりすることを避ける観点から、「無断転載や二次配布は、別に定める場合を除き認めない」と、基本的にはフェイルセーフの方針を取っています。ただ実際には、公開中のデータの中には「積極的に配布していいよ」「公開していいよ」というものもたくさんあり、そのようなものについては、別途「いいですよ」と規定で定めています。基本は「認めない」となっているということが特徴かと思えます。また、データ利用時には、「出典明記はお願いするが、通常は金銭的保証は求めない」という方針で進めました。

帰属については、議論が非常に分かれるところではあったのですが、「職務上取得したデータの権利は、研究所（研究所に所属する個人を含み得る）に属する」ということにしています。要するに研究者個人に付くのではなく、あくまでも研究所に帰属するという形で記述しています。

「5.データの公開期間」については、「データ公開期間に関する項目は必要か」「公開の開始時期の遅延、公開中止の理由として挙げた例文は適切か」が論点になりました。ここについても弁護士からアドバイスをもらって、「合理的な理由のある場合には」という文

データポリシー素案作成 step0					
参考にした既存の文書・ポリシー等（表の項目は一部）					
機関	JST科学技術情報委員会	日本学術会議情報学委員会国際サイエンスデータ分科会	水産研究・教育機構	産業技術総合研究所地質調査総合センター	農業環境技術研究所農産物環境情報センター
表題	データシェアリングのあり方に関する提言	オープンデータに関する権利と義務	環境調査データポリシー	研究成果情報データポリシー	サイトおよびデータ利用ポリシー
設定時期	2015年4月	2014年9月30日	2016年3月2日 2016年4月1日改正		GOSATデータポリシー 2008年11月10日 2009年4月16日改訂 2015年3月13日改訂
データの範囲	シェアリング範囲		データの定義	用語の定義（データの範囲）	サイトへのリンク
提供機関の権利		データに関する権利	データの帰属	保護される権利・帰属第三者に帰属する権利	
提供機関の義務	品質保証方法	データに関する義務	データの取得、保管、管理、整備		データ配布に関わる使命
データ提供期間	データ整備ルール			公開の開始・終了	格納データを予告なく変更 データ配布時期
⋮	⋮	⋮	⋮	⋮	⋮

(図1)

言を入れました。

こうしてできた国立環境研究所のデータポリシーは以下から公開されています。

(https://www.nies.go.jp/kihon/kitei/kt_datapolicy.pdf)

このポリシーを策定したのは管理部門ですが、データの公開は実質的には各研究部門・研究者が中心となって行われています。その際、データ公開ポリシー（ライセンス含む）はデータごとに定めています。研究データ公開に関して、機関共通で実質的効力を持つポリシーを細かく定めてしまうと、多様なデータがある機関の場合は、そのポリシーを決めるだけでも非常に大変です。今の NIES のデータポリシーは、各データ提供者の意思で別途詳細を定められるようになっていので、「情報公開法との切り分けを明確にする」「データ公開を促進するベースを提供する」という目的を素早く達成することができたと思っています。

ただし、一方で、初めてデータを公開する担当者から「こういうデータを公開してもよいのか」「公開の際、どこに許可を取る必要があるか」「このようなデータポリシーで公開したいが、管理部門に確認・判断・決裁が必要か」といった問い合わせが寄せられたり、「データポリシーは誰が決めるのですか」という質問があつたりするほど、データポリシーに関する意識がそれほど高くない現場もあります。実際にこのデータポリシーを運用していくに当たっては、現場ごとに必要なデータポリシーを追加で定められるような、要するに、所全体の大枠のデータポリシーに基づいて、各

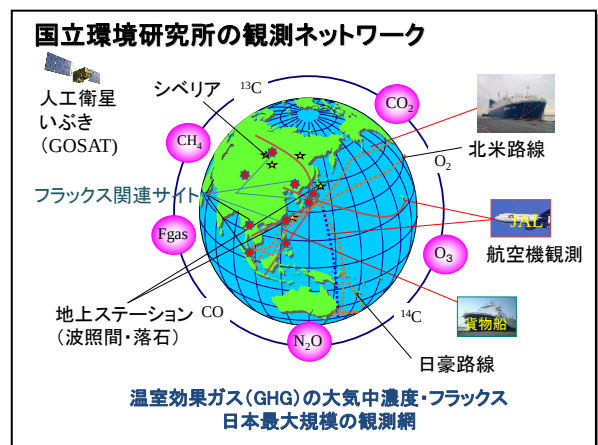
データの細かいところを定められるような教育、支援が必要という状況です。

例として、地球環境データベースは 2016 年 9 月から DOI 付きでデータを公開しており、利用規約は、クリエイティブ・コモンズ・ライセンスまたは独自で作成した規約からの選択としています。独自ライセンスの例（図 2）では、「配布の条件」で「有料で販売してはならない」としたい人はそうしますし、「利用者は、再配布に際し、再配布により本データセットを受け取る者に対し、本データセットが本利用規約の下で提供されていることを通知しなければならない」とすることもできます。「出典の明記」では、作成者が出典の記載を削除するように要求できるようにもできます。もちろん「引用の書式」も独自で決められます。研究者がクリエイティブ・コモンズなどをあまり使わない一番の理由は、不特定多数に向けてデータの再配布や成果物の公表をしたときには報告してほしい、何かデータセットに問題があった場合は報告してほしいといった、データ利用者からの利用報告や問題報告、フィードバックが欲しいというケースがほとんどです。そのあたりをうまく拾っていきけるようなポリシーが必要なのではないかと認識しています。

このように、大枠のデータポリシーの他に、現場で細かいことをデータごとに決めていくということをご紹介しました。

利用規約の例：地球環境データベース (GED)	
2016年9月からDOI付きでデータを公開。利用規約はCCライセンスまたは独自で作成。	
許諾の種類	利用者は、本データセットの内容を改変することなく第三者に再配布すること、および、本データセットをもとにして新たに知的生産物（以下、派生物）を作成して公表することができる。
配布の条件	本データセットを第三者に再配布する場合、 有料で販売してはならない 。 本データセットを本利用規約に基づいて第三者に再配布した場合、再配布されたデータセットにも本利用規約が適用される。 利用者は、再配布に際し、再配布により本データセットを受け取る者に対し、本データセットの原配布元のURL、本データセットが本利用規約のもとで提供されていること、および、本データセットは作成者によって随時更新されていることを通知しなければならない。
最新版の利用	利用者は、特別の理由がない限り、本データセットの最新版を利用しなければならない。
出典の明記	利用者は、派生物に、その派生物が本データセットを利用して二次的に作成されたものであることを明記しなければならない。 ただし作成者は、利用者に、派生物から出典の記載を削除するように要求することができる。その場合、利用者は、合理的に実施可能な範囲で、派生物から出典の記載を削除しなければならない。
引用の書式	本データセットを引用する場合は、原則として、以下のように表記しなければならない。 向井ら(2014)、波照間島における大気 CO ₂ 濃度の連続観測データセット、Ver. xx ("1)、国立環境研究所 地球環境研究センター、 DOI:10.1255/ged.0001.001 (参照: YYYY/MM/DD("2")) ("1")バージョン番号はファイル名に記載されている。 ("2")参照年月日は、ファイルをダウンロードした日付を記載すること。 ただし、掲載誌の投稿規程等により引用の書式が定められている場合は、その掲載誌の規程に従うものとする。
利用報告	再配布、および成果物の公表が、 不特定多数に向けて行われた場合 、利用者は データ提供機関 にそのことを速やかに報告しなければならない。
問題報告	本データセットの内容について、不備や誤り、またはその疑いのある記述を発見した場合、利用者は、その情報を速やかに データ提供機関 に報告しなければならない。
無保証及び免責	作成者およびデータ提供機関は、本利用規約に基づく本データセットの利用に関して一切の保証をしない。 利用者は、派生物の内容について、作成者またはデータ提供機関が何らかの保証を与えているかのような表現を、作成者に断りなく行ってはならない。 本データセットの利用によって何らかの損害が生じた場合、いかなる損害であってもすべて利用者が責任を負うものとする。

(図 2)



(図 3)

現場の例：地球環境研究センター

現場の例として、私が所属する地球環境研究センターの取り組みについてご紹介したいと思います。地球環境研究センターは 1990 年に発足しました。人工衛星、船舶観測、航空機観測、地上ステーションなど、日本最大の温室効果ガスの観測網を持っており、そのデータの発信を主に地球環境データベースで行っています（図 3）。

2010 年代前半までは、大気中温室効果ガス濃度や、海や陸域におけるフラックスといったデータは、アメリカなどに本拠地がある国際的なデータベースを通じての提供や、個別にリクエストをもらった研究者からの提供がほとんどでした。

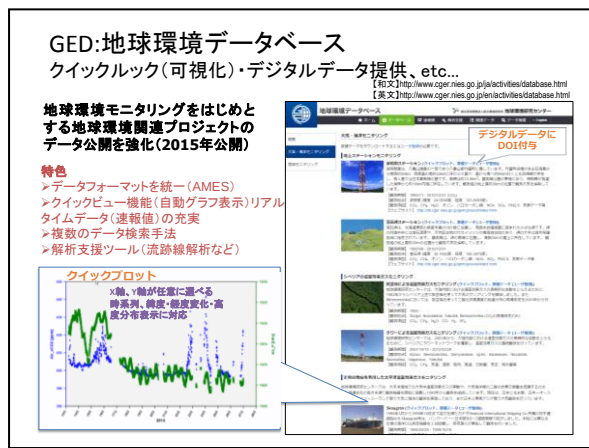
2011 年からデータ統合・解析システム（DIAS）の 2 期に参加させていただき、このときにメタデータのみ DIAS に投入し、公開したのですが、実データをデータ

ベースから提供ということはしていませんでした。ということで、データ公開用に地球環境データベースを構築し、2015 年から公開しました（図 4）。

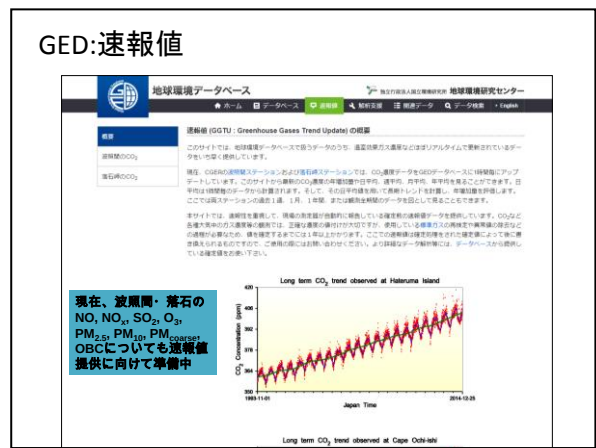
地球環境データベースでは、地上ステーションや船などで観測したデータをダウンロードできるようになっており、ユーザー登録が必要なデータもありますが、クイックプロットができたり、データフォーマットが統一されていたり、今まで個別に研究者が抱えていたデータを一括して提供できるようなものになっています。

例えばデータ検索なども、地点検索、メタデータ検索など、いろいろな検索ができるようになっていたりします（図 5）。検索については、現在、改良中です。

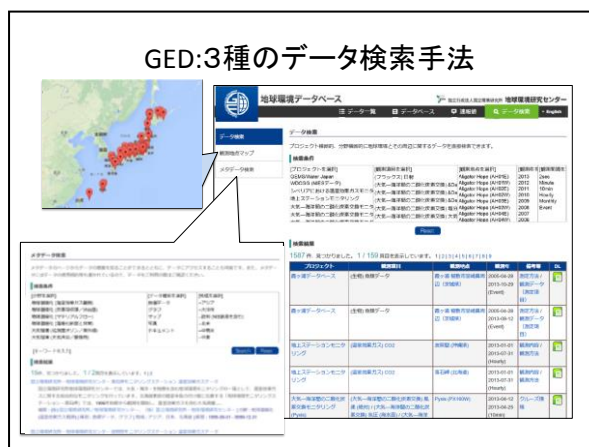
また、速報値といって、そのページに行けば今日の値が表示されているというような、沖縄の波照間島と北海道の落石岬で観測している今の大気中の CO_2 の



(図 4)



(図 6)



(図 5)



(図 7)

濃度がリアルタイムで見られるようなサイトがあります（図 6）。それらの値を別のサイトで表示したいという場合、API を提供しています（図 7）。

また、トレンド解析、トラジェクトリ、フットプリントのような、一般の方は使うかどうかは分かりませんが、学生や研究者がよく使う解析ツールも提供しています。これらは DIAS から公開しています（図 8）。

情報提供サイトも作っています。例えば、アジア太平洋の温室効果ガス関係のモニタリングサイトをまとめて、検索して一括表示できるようなものも作成しています（図 9）。

研究データに DOI を付与する流れができていることは皆さんも十分ご存じだと思うのですが、NIES でもその試みを進めています。国立情報学研究所の武田先生を中心に進めてくださった、研究データに DOI を付与するプロジェクト「研究データへの DOI 登録

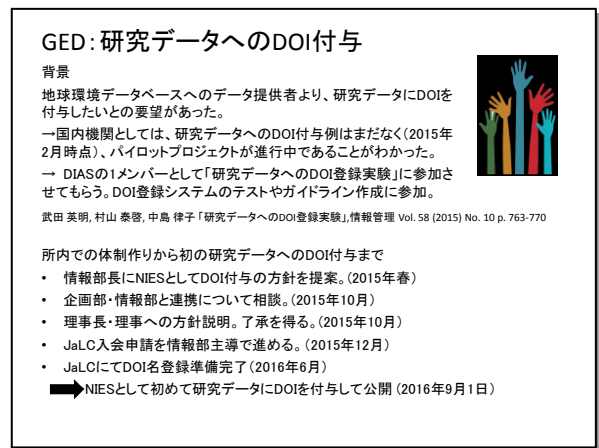
実験」に 2015 年から参加させていただきました。研究所の中で、NIES として研究データに DOI を付与する方針を提案し、2016 年 9 月から、NIES として研究データに DOI を付与して公開を始めました（図 10）。

その後も順調に、地上ステーションのデータや、モデルで計算した二酸化炭素フラックスマップ、富士山頂で観測した大気中二酸化炭素濃度、民間航空機で観測している二酸化炭素濃度、オゾン層衛星観測センサのデータも、DOI 付きで公開しました（図 11）。

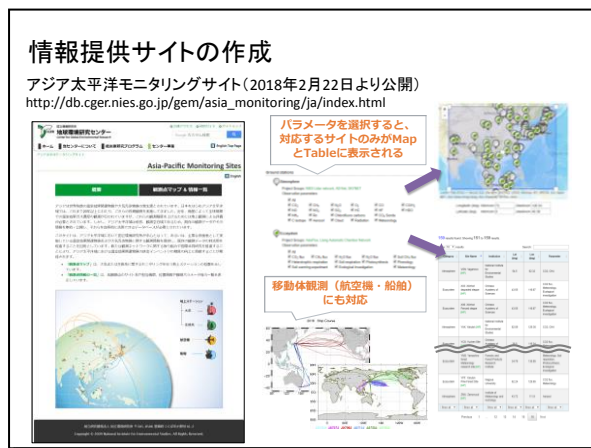
このようなデータセットはどんどん使ってほしいデータであり、DOI を付与した公開が、データを提供することが評価されるということにつながっていくといいなと思ってやっている面もあります。また、DOI を付けることでメタデータをきっちり作ってもらえるようになったという他のメリットもあります。ですから、率先して DOI を付与した研究データ公開を進めているところです。



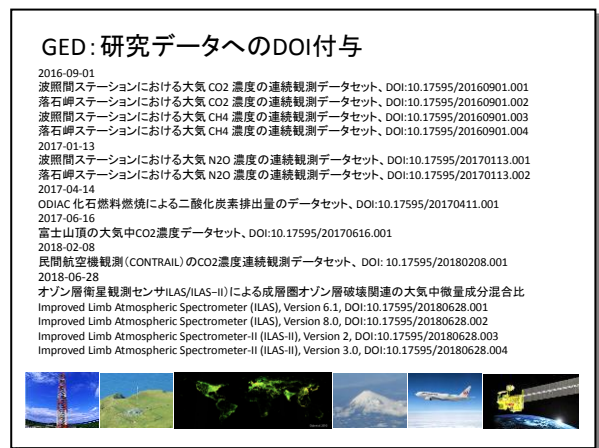
(図 8)



(図 10)



(図 9)

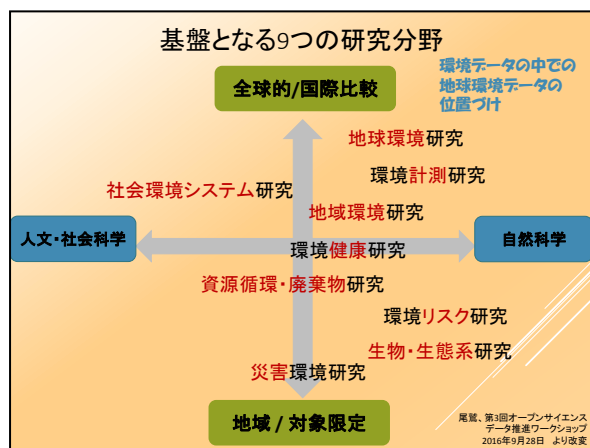


(図 11)

NIES におけるオープンデータの課題

NIES には非常にたくさんの分野があり、いろいろな研究をしています。基盤となる九つの研究分野として、地球環境研究、環境計測研究、地球環境研究、環境健康研究、資源循環・廃棄物研究、環境リスク研究、生物・生態系研究、災害環境研究、社会環境システム研究があります。これらは、全球的で国際的な研究か、地域的で対象が限定された研究かという縦軸と、人文・社会科学系か、自然科学系かという横軸のマトリックスで分類することができます（図 12）。その中でどこに位置するかによって、データが大きく異なります（図 13）。自然科学系はモニタリングや実験をしますが、人文・社会科学系はアンケートを取ったりするので、データそのものがどう出てくるかも違い、気になることが全く違ってきます（図 14）。

地球環境分野の現場観測であれば、バージョン管

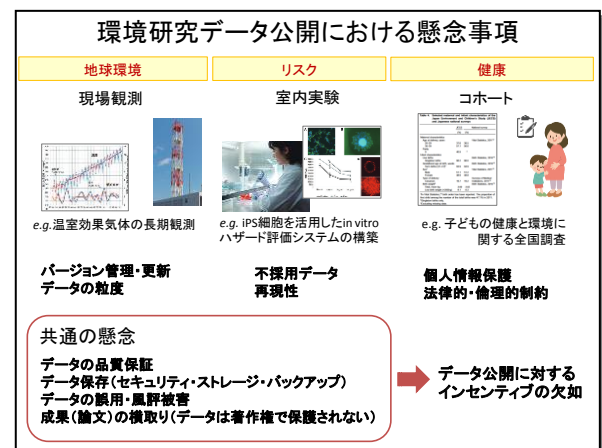


(図 12)

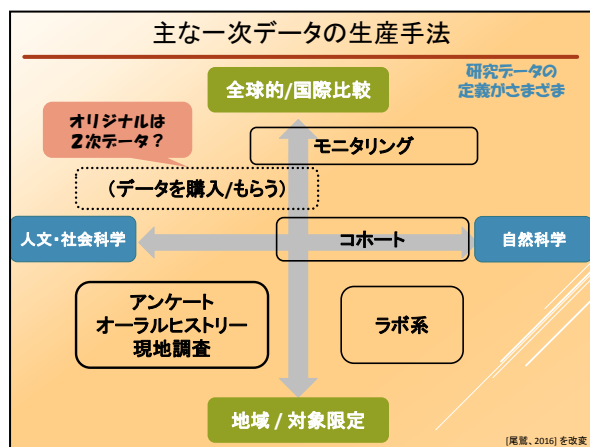
理・更新、データの粒度が問題になりますが、室内実験の場合は、失敗したデータはどうするのか、再現性をどうするのかということが問題になってきます。健康分野のコホートでは、個人情報保護や法律的・倫理的な制約が問題になります。

その一方で共通の懸念もあり、それはデータのクオリティコントロール、データ保存（セキュリティ・ストレージ・バックアップ）の問題、データの誤用や風評被害の問題、成果の横取りです。これらがデータ公開に対するインセンティブの欠如につながっています。

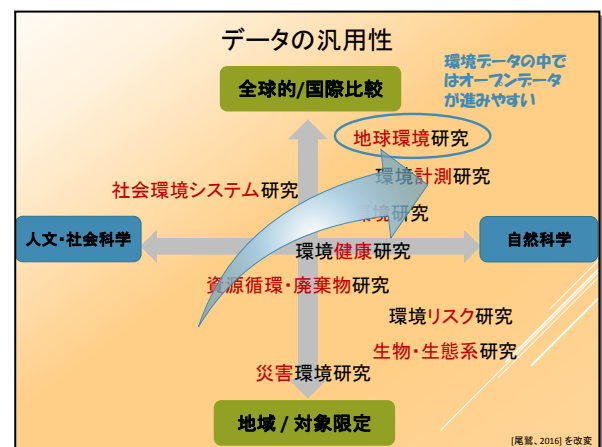
データの汎用性については、私が今携わっている地球環境データは、環境データの中では非常に全球的で、国際的にデータを使い合っていくような分野であるということ、また、自然科学系なので、ある程度データのフォーマットなどがそろっていることが多いこともあり、環境データの中ではオープンデータが進みやすい、モチベ



(図 14)



(図 13)



(図 15)

ーションが湧きやすい部類に入ります（図 15）。ところが、研究所の他の分野はもう少しハードルが高いことが多いので、研究所全体でなかなか一緒に進めないというのが現状です。

「研究現場におけるオープンデータに向けた課題」をまとめてみました。一つ目として、動機不足があります。そもそもデータをなぜ公開する必要があるのか分からない、成果を出してからデータを公開したいのに、なぜ今、データを出さなければいけないのかという疑念から、ずっとデータを抱えてしまいます。これについては、Schmidt ら（2016 年）、池内ら（2017 年）による、アンケートの結果で、データを公開する際の問題として「自分の成果を出してからデータを公開したい」という割合が非常に多いということが分かっています。メリットがまだ分からない、未知の恐怖のようなものがあります。

二つ目は、知識不足です。学術コミュニケーションについての知識不足があり、メタデータ、著作権、DOI、研究データマネジメントなど言葉はよく聞けけれど、一体何をすればいいのかが全然分からないということもインセンティブを下げてしまいます。そのあたりは図書館員の皆さんに役割を発揮していただける部分ではないかと思っています。

三つ目は、リソース不足です。データを公開したいがリソースが足りない、これは皆さん深刻だと思うのですが、ストレージ容量や IT 人材、時間がありません。セキュリティ対策、バックアップといったデータ

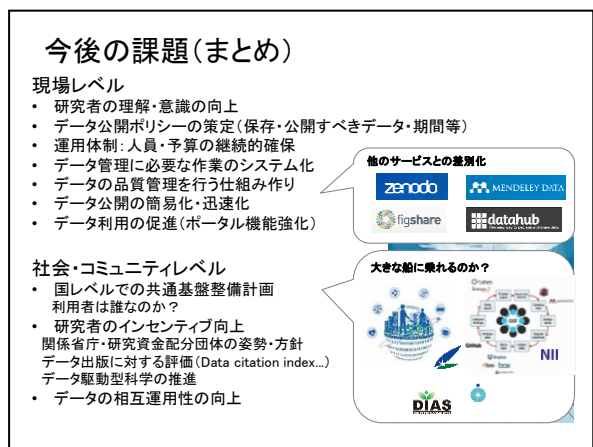
管理への支援不足、そのようなことを考えていくと、なかなか手が動きませんし、やりたいと思って一人ではできないということがあります。

このような課題が現場では非常に大きいです。これらを全部クリアしたとしても、四つ目に、公開したデータが有効に利活用されているかという課題があります。例えば、データの見つけやすさ、アクセスの容易さ、データの品質情報・メタデータが明示されているか、データ利用条件の簡素化・無償提供をしているか、そのようなポイントが幾つかあると思うのですが、データを公開したことで満足してしまって、本当にそれが有効に活用されているか、本当はもっと使ってもらえるのではないかといいところまではいっていないのが現状ではないかと思います。

今後の課題のまとめです（図 16）。現場レベルでは、一つ目は研究者の理解・意識の向上です。二つ目は、データ公開ポリシーの策定です。まだ NIES もあまりやっていないものとして、保存や公開すべきデータ、期間等についても決めた方が良いでしょう。三つ目は、運用体制で、人員・予算の継続的確保です。これが非常に難しいということがあります。四つ目は、データ管理に必要な作業のシステム化です。これは NIES でも今行っていますが、非常に地道な作業です。五つ目は、データの品質管理を行う仕組みづくりです。これは高度な話です。六つ目は、データ公開の簡易化・迅速化です。これはシステム化とも関係しています。七つ目は、データ利用の促進です。

これらの課題を考えると、他の民間のサービスなどいろいろあり、スピードや手軽さでは競えないことがあるので、そのようなところも考えながら研究所の現場はやっています。

社会・コミュニティレベルの課題としては、国レベルの共通基盤整備計画がいろいろあるのですが、利用者が誰なのかがあまりはっきりしていないものもあり、大きな船に乗れるのかどうかに関心事の一つではあります。また、研究者のインセンティブ向上には、ファンディングエージェンシーや関係省庁の姿勢や方針が



(図 16)

非常に効いてきます。データ出版に対する評価がもう少し上がれば少し変わってくると思います。また、ビッグサイエンスや AI、ビッグデータ解析などの連携がどんどん進んでいくと、もう少しデータを整備しようという気持ちになっていくと思うので、そのようなデータ駆動型科学の推進があるのではないかと思います。そして、データの相互運用性の向上があります。

データポリシーは作ってあると非常に有用だと思います。それがデータを使っていくきっかけになることは確かだということをメッセージとして残したいと思っています。

●林賢紀 ありがとうございます。それでは質問を幾つかお受けしたいと思います。

●フロア 1 東京大学附属図書館の図書館員です。データポリシーの策定の経緯から実際の使われ方までご説明いただきまして、どうもありがとうございました。ポリシーの策定から1年半経過して、その間に内閣府から、「国立研究開発法人におけるデータポリシー策定のためのガイドライン」が出ています。1年半運用されてみて、あるいはガイドラインがこのように国から出た状態で、NIES のデータポリシーに関して、改訂の必要性は認識されていらっしゃるか、あるいはその運用の中で研究者からどのような意見が出ているかということを伺えればと思います。

●白井 とてもポイントを突いたご質問だと思います。

最初の、内閣府のガイドラインを見て改訂の必要性が感じられるかですが、それは非常にあるとは思っています。先ほど、私も最初に NISTEP の林さんのご発表を聞きながらチェックを入れたりしたという話をしたと思うのですが、全く触れていない部分などもあります。ただ、やはり内閣府としても別に全部チェック

しろという意味で作られたわけではないということなので、もちろん全部を網羅するための検討をしなければいけないとは思ってはいません。一方、ガイドラインができたことで研究所での議論はしやすくなったというのはあるので、優先順位を付けて、少しずつ検討していない部分を検討して、データポリシーを改訂していくことはもちろん考えているところです。

それから、データポリシーに関する研究者からの影響ですが、あまり聞こえてきてはいません。このデータポリシーは本当にふわつとしたものなので、その先の現場で作成する細かいポリシーやライセンスについては、もちろん議論が白熱することも現場ではあるのですが、研究所ポリシーについては特に利害が反するところがないというのもありまして、議論にはなりません。平和なポリシーです（笑）。

ただ、内閣府でこれについて発表させていただいたときに、科学技術振興機構（JST）の方から共同研究のときのことをいろいろ聞かれて、そのあたりについてはもう少し書き込んでおいた方が組織としていいのではないかという認識は持っています。

●フロア 2 横浜国立大学（SPARC Japan 運営委員）の深貝です。社会科学系がベースで、思想研究の側なので、自然系の方の感覚を伺いたいと思ひまして。白井さんのお話の最後の方で、研究者がデータをなかなか出しながらないという、当然といえば当然の話を頂きました。データポリシーを使ってなるべく公開に向けて促して、やがては赤池さんの話の Society 5.0 まで行こうという遠大な話です。実験系の方だと、善しあしはともかく、データを隠したいとなりがちだというのはそうだと思うのです。特に資金をどこから出してもらったかを含めた上で、特許や発明などいろいろなことをしていますから。それでもなおデータをオープンにしてほしいということを今やっているわけですが、観測系の方ですと、観測データが多くなることでようやく分かることがありそうで、傾向的变化という場合に、わずかな観測ではとても足りないから多く集める

必要があるわけです。そのような意味では、地球環境の分野では、データを出すことにもっと積極的であっていいのではないかと思いますのですが、どうでしょうか。

それから、自分がたまたま非常に珍しいデータを観測してしまった場合はやはり隠したくなりそうですが、研究者はどう振る舞うのでしょうか。

●白井 1 番目の質問に関しては、地球環境データのようなものはオープンデータと親和性が高く、究極をいっているのは天文の分野で、データを使い合わないとサイエンスできないということで、分野内だけです。ものすごくオープンデータが進んでいるようです。

地球環境の場合はその中間というか、使い合った方がいい分野とそうではない分野がまだあるのですが、やはりインセンティブとしては非常に高いですし、皆さん、研究の過程で他の人の作ったデータセットのお世話に必ずつなっていることが多い分野なので、そのような意味では、データをオープンにしていくことへの認識は高い分野だと思っています。ただし、認識は高いけれども、実際に手を動かすかというのはまた話は別というのはあります。これが 1 問目の答えになります。

2 番目の珍しいデータに関しては、戦略的なオープン・クローズの判断になると思います。珍しいものはやはり自分で最初にしゃぶり尽くしたいというのが研究者の本能的な感覚だと思いますので、すごいネタになりそうなものであれば、しばらく公開しないという方針を戦略的に決めるのはある話だと思います。それは特に悪いことだとは思っていないのですが、それによって大発見が遅れてしまうとなるともったいないので、2 年などそれなりのエンバゴ期間が決めてあるのがいいと個人的には思っています。