第 4 回 SPARC Japan セミナー2014

「グリーンコンテンツの拡大のために我々はなにをすべきか？」

開会／概要説明

三角　太郎
（千葉大学附属図書館）

三角　太郎
1992年東北大学理学部卒業（専攻は地質学）。1998年宇部工業高等専門学校に採用後、山口大学、山形大学を経て2014年4月より現職。図書系の業務は一通り経験し、機関リポジトリとの出会いは2005年山口大学にて。以後リポジトリと携わる。機関リポジトリ推進委員会国際連携ワーキンググループメンバー。大学図書館コンソーシアム連合（JUSTICE）作業部会委員。

本日は、年度末のお忙しい中をお集まりいただきありがとうございます。第 4 回 SPARC Japan セミナー 2014 のテーマは、「グリーンコンテンツの拡大のために我々はなにをすべきか？」です。今回は国立極地研究所の南山さん、学術資源リポジトリ協議会の堀井さん、そして、私、千葉大学の三角の 3 名が企画ワーキンググループとして、今回のセミナーは企画いたしました。

本日の趣旨は、オープン化推進機関・図書館関係者・学術研究者などを交えて、今後の多様な学術情報流通に関する議論を行い、日本の将来における学術資源のオープン化のイメージを広く共有するとともに、リポジトリ等による学術資源管理の具体的な道筋を探ることです。

学術資源のオープン化

学術研究活動の中で生成された学術資源のオープン化は、学術研究の促進目的ばかりでなく、社会的要請からも求められてきています。現状は、学術研究分野（理工・人文など）や機関・組織（大学・博物館・自治体など）が異なると、オープン化に対する認知やモチベーション・普及状況などに大きな差異があります。また、多くの機関・組織では、一つの機関内で生成された学術資源が複数のリポジトリや各部局のデータベースなどのシステムに散在しています。そのような中で、組織的なオープン化の推進が必要ではないかと考え、今回のテーマを設定しました。

普通、グリーンコンテンツとは、ゴールドに対するグリーンのことだと思われるでしょうが、本セミナーでは、「学術研究機関が機関リポジトリ等において公開・発信するオープンな学術コンテンツであり、研究データ、博物館資料に関するメタデータおよび資料画像データ等も視野に入れた多様なコンテンツ」として定義させてください。

IRDB による分析

図 1 は、今、機関リポジトリにどのようなコンテンツが入っているか、IRBD で分析したものです。通常、グリーンコンテンツといわれているものはこの緑色の部分です。圧倒的に論文に寄っているのが現状ですが、

そうではなく、データや教材、研究報告書などをもっと増やしていきたいと考えています。

図 2 は、今の表を円グラフにしたものです。圧倒的に紀要が多く、今回特にターゲットにしたいデータやデータセットは約 4%しかないのが現状です。

二つのアプローチ

今日は、「コンテンツの多様化のために、我々はなにをすべきか」を考え、みんなで頑張ろうという決起集会です。今日の会で何かまとめようというよりは、皆さんに問題提起をしていただこうと思っています。

コンテンツへのアプローチは大きく二つに分けられます。それは、「最初に研究ありき」と「最初にデータありき」です。前者は、最終的な研究成果物である論文があり、その添付物としての研究データがあるというものです。後者は、博物館標本の画像や計測データ、観測データなど、まずデータがあって、それらがみな論文になっているとは限らないというものです。

「最初に研究ありき」は、図書館員には非常に分かりやすいアプローチです。ただ、いつまで論文が最終的な研究成果物であるかが分からなくなってきている状態だと思います。今は、研究データ流通の構造そのものが大きく変わろうとしているからです。

武田先生には、「機関リポジトリと DOI - JaLC における DOI 付与について - 」と題して、今、研究データ流通がどのように変わろうとしているのか、その中で DOI などの識別子はどのような位置付けになっていくのかについてご講演いただきます。

「最初にデータありき」は、代表的なものとして博物館が挙げられます。こちらに関しては、山下先生に「大学博物館における学術資料情報のオープン化に関する取組み」と題してご講演いただきます。

機関リポジトリの現状

データ・データセットの日本の機関リポジトリ中のデータ件数は 5 万 2,421 件で、そのうち千葉大のコンテンツが 5 万 2,223 件と、実は千葉大学のコンテンツが 99.6%を占めているのです（図 3)。2008～2009 年度の CSI 事業で、千葉大学は集中的に研究データに取り組んだことがあり、このような形になっています（図 4)。

ただ、このプロジェクトはあまり広がらなかったのが実情です。千葉大学も、その当時、5 万件入れた後

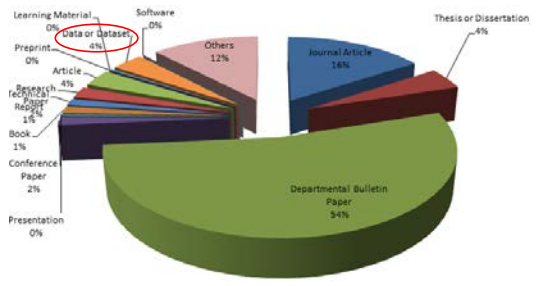
IRDBによる分析

	author（著者版）	publisher（出版者版）	ETD	none（その他）	計
Journal Article（学術雑誌論文）	28,916	164,385	0	495	193,796
Thesis or Dissertation（学位論文）	20,311	14,247	11,526	0	46,084
Departmental Bulletin Paper（紀要論文）	34,025	504,885	0	1,143	540,053
Conference Paper（会議発表論文）	2,489	16,086	0	111	18,686
Presentation（会議発表用資料）	2,396	2,037	0	258	4,691
Book（図書）	700	6,502	0	4,956	12,158
Technical Report（テクニカルレポート）	958	9,985	0	45	10,988
Research Paper（研究報告書）	3,695	18,333	0	578	22,606
Article（一般雑誌記事）	4,541	16,094	0	38	20,673
Preprint（プレプリント）	278	25	0	14	317
Learning Material（教材）	626	1,886	0	538	3,050
Data or Dataset（データ・データベース）	27	51,954	0	71	52,052
Software（ソフトウェア）	23	5	0	0	28
Others（その他）	8,759	77,995	0	13,991	100,745

2015.2

（図 1）

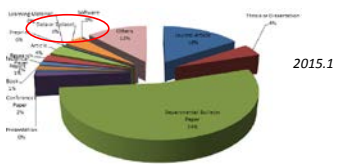
IRDBによる分析



国内機関リポジトリのコンテンツタイプ　2015.1

（図 2）

あらためて機関リポジトリの現状はどうなっているのか



2015.1

日本の機関リポジトリ中のデータ　52,421件
そのうち千葉大学のコンテンツ　52,223件
99.6%

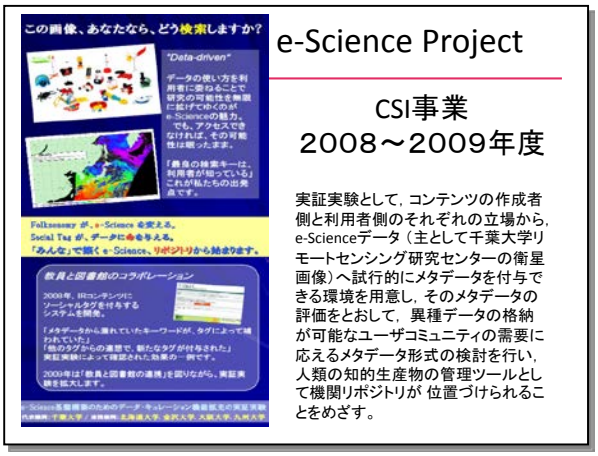
（図 3）

はほとんど増えていません。いろいろな事情がありますが、少し取り組む時期が早かったのではないかと思います。特に多いのが萩庭標本という植物標本です（図 5）。これが 5 万件に近くあります。

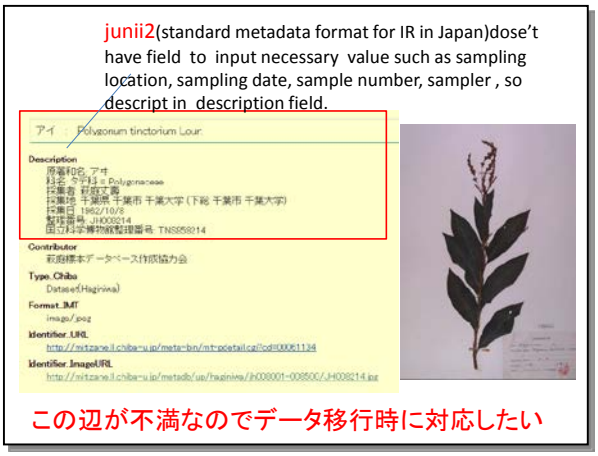
私が千葉大学に異動してきたのは昨年 4 月で、このコレクションの存在を知り、これは非常にいい試みだと思いました。一方でかなり不満もあります。植物標本は採集者・採集地・採集日などが非常に重要ですが、それらを入れる項目が junii2 にないため、Description に全て放り込んでいる状態で、採集日や整理番号から検索することができないからです（図 6）。ですから、植物標本のデータベースとしては、あまり適切ではありません。このあたりは今後のデータ移行時に対応したいと考えています。

ところが、学内のデータ移行についてのディスカッションの際、上司に「こういうものをリポジトリに載せる意味はないのではないか。ZIP か何かで固めてどこかに置いておけばそれで済むのではないか」と言われたのです。正直言ってむっときました。一応、私も反論しましたが、通じたかどうかは分かりません。

私は、自分がなぜむっときたのかと考えてみました。私は図書館員ですが、大学では地質学古生物学を専攻していて、有孔虫という化石を研究していました（図 7）。顕微鏡で一個一個ピックアップして、図録を参照しながら、種を同定するまではいかなくても、学生でも当たりぐらいは付けていました。そのような経験があったので、標本の写真がオープンになって、記載事項などの検索もできるようになると非常に有効なツールになるのではないかと思ったのでした。恐らく地質学以外でも、オープンにすると有効なリソースはたくさんあるのではないでしょうか。



（図 4）



（図 6）

（図 5）

（図 7）

図書館はなにをすべきか？

それを考えると、図書館員はリポジトリの可能性を本当は理解していないのではないかと思うようになりました。リポジトリはほとんどの場合、図書館員が構築していますが、本当に使っているのは図書館員ではありません。また、論文にこだわり過ぎなのではないのか、データや標本などにもっと取り組むべきではないのかという疑問も持ちました。ささやかな私の経験から考えても、リポジトリはもっと有用なツールになるポテンシャルがあると思います。

もう一つの疑問は、図書館員は研究者マインドが分かっているのかということです。それを考えると、自分に関係のない分野の先生などをむっとさせたことがあるのではないかと思いはじめました。恐らく、欧米のサブジェクトライブラリアンは研究者マインドが分かっている人たちではないかと思います。それに対して、日本はサブジェクトライブラリアンがいないので、その辺がウイークポイントになっていくのではないでしょうか。日本でサブジェクトライブラリアンは少し難しいと思いますが、せめて理系出身の図書館員などへ、このようなプロジェクトへの参画を求めたりできないのかと思いはじめました。

図書館はなにをすべきかについては、まずは現状を押さえるということで、南山さんに「図書館によるデータ管理への道筋」と題してご講演いただきたいと思います。