

# 電子ジャーナルによる情報発信と管理について\*

高橋 洋成

(Tel Aviv University)

tjonah@khh.biglobe.ne.jp

## 1 はじめに

学術論文の掲載された雑誌が公刊されるまでには、編集者による査読、紙面の構成、著者による校正といったプロセスが必要になり、数ヶ月、場合によっては数年の時間を要する。だが、インターネットの普及に伴い、こうしたプロセスの煩雑さとコストを削減すべく、学術情報をオンラインで配信する団体が増えている。

オンライン配信される学術情報は、大まかに二種類に分けられる。一つは、査読のプロセスを簡略化し、速報性を重視するもの。物理学、数学、生物学などの分野では、早くからプレプリントサーバと呼ばれる論文登録システムを用いている<sup>1</sup>。ここに投稿された論文は誰もが自由に閲覧し、コメントを残すことができる。さらに、論文の著者は寄せられたコメントをもとに内容を推敲し、改めて査読付き学術雑誌に投稿することもできる。オンライン配信のもう一つの特徴は、査読を通過した後、紙面構成から最終校正の途中の論文であっても、出版前に公開できることである。冊子体は出版後の送付・流通にも時間がかかるが、オンライン配信は流通のコストを省くことができる。

本稿では、日本における学術情報のオンライン化の現状を概観し、次いで電子ジャーナル“Studies in Ethiopian Languages”がどのような特徴を持つかを述べる。

## 2 日本における電子ジャーナル

### 2.1 J-STAGE

学術情報の電子化について、日本では文部科学省の指導のもと、いくつかのプロジェクトが進められている。ひとつの例として、独立行政法人・国立情報

---

\*本調査は平成 22～25 年度科学研究費基盤研究 (B)「変容するエチオピア諸言語の静態と動態に関する総合的研究、ならびにデータベース構築」代表：柘植洋一（金沢大学）（課題番号 22401046）によるものである。

<sup>1</sup>例えば、<http://arxiv.org/>

学研究所（NII）による、論文・図書・雑誌を検索するデータベース・サービス CiNii<sup>2</sup> が挙げられる。これは、大学図書館等が所蔵する論文・図書の目録検索サービス NACSIS Webcat<sup>3</sup> を、より総合的なデータベースとして拡張した後継システムである。

日本における学術情報の電子化のもう一つの例は、独立行政法人・科学技術振興機構（JST）による電子ジャーナル・サービス（J-STAGE）である<sup>4</sup>。これは、電子ジャーナルについてのノウハウを持たない学術団体でも容易に情報をオンライン発信できるよう、1998年に立ち上げられたプロジェクトである。JSTによる審査をパスした団体は、その発信・流通システムを無償で利用できる。

J-STAGEについては、文部科学省の研究環境基盤部会・学術情報基盤作業部会（第41回）の資料が閲覧可能である<sup>5</sup>。それによると、J-STAGEには大きく次のような機能がある。

- 海外データベースと連携した引用文献へのリンク
- 海外データベースと連携した被引用文献へのリンク
- 利用統計、全文検索
- 剽窃の検知

また、J-STAGEのウェブサイトの目次では、電子ジャーナルが次のように分類されている。この分類は便宜的なものだが、日本における電子ジャーナルの現状と性格を把握する上で有用であろう。

## 学会・協会系

学術団体が発行している学会誌や協会誌の電子版。

## 商業出版系

出版社が発行している専門誌の電子版。

## 無料電子ジャーナル

学会誌や専門誌のうち、会員以外にも公開されているもの。

## オンライン・ジャーナル

冊子体のない雑誌。

---

<sup>2</sup><http://ci.nii.ac.jp/>

<sup>3</sup><http://webcat.nii.ac.jp/>

<sup>4</sup><http://ejournal.jp/>

<sup>5</sup>[http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/002-1/siryu/\\_icsFiles/afieldfile/2011/07/19/1308123-1.pdf](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/002-1/siryu/_icsFiles/afieldfile/2011/07/19/1308123-1.pdf)

## 電子アーカイブ

過去に出版された学会誌や専門誌を電子化したもの。

## 総合サービス

各種の電子ジャーナルについての総合的な情報や検索システムを提供するもの。

## ビデオジャーナル

動画として提供されるもの。

こうした種々の性格を持つ情報発信サービスを総合して、本稿では「電子ジャーナル」と呼んでいる。

## 2.2 導入事例 – 日本言語学会

2011年6月に行われた研究環境基盤部会・学術情報基盤作業部会（第40回）において、日本言語学会における電子ジャーナル導入の経緯が簡潔に報告された。その資料は文部科学省の研究環境基盤部会・学術情報基盤作業部会のウェブページで閲覧できる<sup>6</sup>。

電子ジャーナル事業を行っている JST は、国内で発行された学術雑誌を電子アーカイブ化する Journal@rchive というプロジェクトも同時に運営している<sup>7</sup>。日本言語学会は2006年に J-STAGE の審査をパスし、Journal@rchive の助力のもと学会誌『言語研究』の第1号（1939年）から第100号（1991年）までをオンライン公開した<sup>8</sup>。また、第101号（1992年）からは日本言語学会が独自に電子化を企画し、学会のウェブサイトから論文ファイルを直接ダウンロードできるようになった<sup>9</sup>。

当該資料における日本言語学会の電子ジャーナルについての見解を整理しておく。ここには、学術情報の電子化に伴うメリットと、将来に向けて解決すべき点が簡潔にまとめられている。

- 現在は紙冊子と電子版を並行して刊行している。紙冊子は会員のみに配布する。電子版は刊行から一年を経たものを、制限を設けずに公開する（4-11「電子ジャーナル化」）。
- 著作権は基本的に原著者に帰属するが、ウェブサイトで公開するため2006年に「日本言語学会著作物取扱規程」を定めた。この規程は、制定された

<sup>6</sup>[http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/002-1/siryu/attach/1307336.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/002-1/siryu/attach/1307336.htm)

<sup>7</sup><http://www.journalarchive.jst.go.jp/japanese/>

<sup>8</sup><http://www.journalarchive.jst.go.jp/japanese/jnltop-ja.php?cdjournal=gengo1939>

<sup>9</sup>[http://www3.nacos.com/lstj/modules/documents/index.php?cat\\_id=23](http://www3.nacos.com/lstj/modules/documents/index.php?cat_id=23)

期日以前に『言語研究』に掲載された著作物に対しても、遡って適用される(4-10「著作権」)。

- 未だ模索中の段階だが、英語論文を増やすために、現在刊行している和文号とは別に、投稿を会員に限定しない英文号を年1回電子ジャーナルのみで発行することが考えられる。電子ジャーナルであるから印刷経費がかからない(6-5「投稿資格について」)。
- 日本国内で出版された言語学関係の文献資料の電子化が遅れている。これには著作権の問題を伴うが、早急に推進する必要がある(6-5「文献資料の電子化」)。

### 3 電子ジャーナル “Studies in Ethiopian Languages”

#### 3.1 背景

SIL International が運営する Ethnologue によれば、エチオピアで使われている言語は少なくとも 90 を数える<sup>10</sup>。2001 年に文部科学省科学研究費の助成を受け、アジスアベバ大学の協力のもとにエチオピアの諸言語を記述するプロジェクトが発足した。その報告書は 2004 年以降、Cushitic-Omotc Studies としてほぼ毎年刊行された。

こうした研究成果を国内外の研究者に周知し、議論を深める場が必要である。それゆえ、これらの報告書をもとにした新たな学術誌の立ち上げを 2011 年に決定し、名前を Studies in Ethiopian Languages と改めた。発行のスピードと柔軟性、および利用者の利便性を考慮し、この学術誌を電子ジャーナルとして提供する。

#### 3.2 提供ホスト

電子ジャーナルを配信する場として、現在は山口大学のウェブサーバ上にウェブサイト上を構築した。ここでは 2004 年から 2010 年までの Cushitic-Omotc Studies、および 2011 年からの Studies in Ethiopian Languages の各号に掲載された論文が PDF 形式で配布される。また、今後は現地調査で収集された音声や動画データなどを提供していくことも可能である。

同ウェブサイトに関連し、エチオピア諸言語の分類、話者人口、使用文字および言語特徴を掲載したウェブページを随時追加していく予定である。これらの言語データは同大学における地理情報システム (GIS) と連携しており、言語特徴の分布や変化などを視覚的に追うことができる。

<sup>10</sup>[http://www.ethnologue.com/show\\_country.asp?name=ET](http://www.ethnologue.com/show_country.asp?name=ET)

### 3.3 URI

冊子体のない電子ジャーナルで考慮すべき点の一つに、ネットワーク上のアドレスの永続性問題が挙げられる。

通常、論文を参照または引用するには著者名・論文題名・ジャーナル名・発行年などの書誌情報が必要であるが、電子ジャーナルの大きなメリットは、リンク機能を用いて論文を容易に参照できることにある。そのために、各論文ファイルには統一資源識別子 (URI) というネットワーク上のアドレスが割り当てられる<sup>11</sup>。裏を返せば、将来にわたってリンクが適切に機能するためには、論文ファイルに割り当てられた URI が変更されないようにしなければならない。電子ジャーナルの提供を開始する時点で、URI の永続性を保証する仕組みを考える必要がある。

さて、ウェブサイトを通じて配布される PDF ファイルの URI は、次のような各成分から構成される。

http : //ds22n.cc.yamaguchi-u.ac.jp/~abesha/SEL/pub/2010/article.pdf  
scheme authority path

本件において Scheme 成分は変化しないため、考えるべきは Authority 成分と Path 成分の構成となる。

まず Authority 成分について、電子ジャーナル Studies in Ethiopian Languages では Online Computer Library Center (OCLC) が提供している PURL サービスを利用する<sup>12</sup>。

http : //purl.org/JAEL/SEL/pub/2010/article1.pdf  
scheme authority path

上記 URI において、//purl.org/JAEL/SEL/ は //ds22n.cc.yamaguchi-u.ac.jp/~abesha/SEL/ へのショートカットとして設定されている。つまり、PURL リンクへのアクセスは、山口大学のウェブサーバへのアクセスとして転送される。論文の PDF ファイルにリンクする場合、あるいは冊子体にアドレスを掲載する場合には、PURL リンクの URI を用いるようにする。そうすれば、後にウェブサーバの移転やディレクトリ構造の

<sup>11</sup>http:... から始まるウェブページアドレスや、mailto:... で始まるメールアドレス、あるいは urn:isbn:... で表される ISBN など、従来それぞれの分野で用いられてきた識別子をインターネット上で統一的に扱うための規格が URI である。また、URI に用いることのできる文字種制限を大幅に緩和した IRI (Internationalized Resource Identifier) もあるが、本稿では慣習的に使われる URI の用語を用いる。

<sup>12</sup>http://purl.org/。OCLC は、コンピュータ技術を基盤とし、各地の大学や研究機関における情報の共有化と図書館コストの削減を目的として、1967 年にアメリカで設立された非営利団体である。現在、約 170 か国の 2,600 の図書館と研究機関が参加している。オンライン蔵書目録 (OPAC) としては最大の WorldCat を運営している (http://www.worldcat.org/)。

変更によりファイル実体を示す URI が変わったとしても、PURL リンクの転送先を修正するだけで済み、論文利用者が情報を見失わずに済む<sup>13</sup>。

また、URI の Path 成分については次のような規則を設ける。

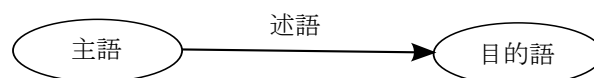
/ SEL / pub / 2011 / Takahashi - 2011 - Hamar .pdf  
ジャーナル名    ジャンル    発行年    著者名    発行年    キーワード

ジャンルについては、pub が論文である他、今後 video や audio といったメディア配信用ディレクトリが追加されていく。発行年をディレクトリ名だけでなくファイル名でも繰り返しているのは、ファイルをダウンロードしたときに見失わないようにする目的と、同名のファイルを誤って上書きしてしまう可能性を減らす目的がある。キーワードについては任意とする。

### 3.4 学術情報のリンク

電子ジャーナルの大きなメリットは、リンクによって参照論文を入手しやすいことにある。さらに、リンクに「意味」を与えることで、論文がどのような参照関係にあるのかを示すことができる。その一例として、ここでは XHTML+RDFa を考えてみたい。

まず、RDF とは「主語」「述語」「目的語」という三つ組によってリンクを意味付けする規格である。ここで使われる各用語は言語学的な意味ではない。RDF では、リンクの起点リソースを主語、終点リソースを目的語、リンクの種類を述語と呼ぶ。述語に用いられる URI のことを「語彙」とも呼ぶ。



この三つ組を基本として、様々な「述語関係」を構築したものを RDF グラフと呼ぶ。目的語は、別の三つ組では主語として機能しうる。

さて、RDFa とは XHTML にいくつかの属性を追加し、XHTML データから直接 RDF グラフを抽出するための規格である<sup>14</sup>。RDFa で用いられる属性を表 1 に列挙しておくが、これらを RDF トリプルを構成するように XHTML データに埋め込んでいく。

次に挙げる例は、PDF ファイルのダウンロードフォームを構成する XHTML 断片である。表の行を示す tr 要素に about 属性が、表のセルを示す td 要素

<sup>13</sup>PURL サービスでは転送に用いる HTTP ステータスコードを選択することができる。301 コードは恒久的移転、302 コードは一時的転送、303 コードは参考リソースを示す。セマンティック・ウェブとの関連では、電子ジャーナルは情報リソースであるため、非情報リソースを示す 303 コードではなく 302 コードで転送するのが望ましい。

<sup>14</sup>概要については <http://www.w3.org/TR/2008/NOTE-xhtml-rdfa-primer-20081014/> を参照。

|          |                                 |
|----------|---------------------------------|
| about    | 主語リソースを示す URI (CURIE)。          |
| src      | about を同じだが、ウェブブラウザ上で埋め込み表示される。 |
| rel      | 目的語リソースを示す述語の CURIE リスト。        |
| rev      | 目的語リソースを示す逆方向の述語の CURIE リスト。    |
| property | 目的語リテラルに向かう述語を示す CURIE リスト。     |
| resource | 目的語リソースを示す URI (SafeCURIE)。     |
| href     | resource と同じだが、ウェブブラウザ上でクリック可能。 |
| content  | 目的語リテラル。                        |
| typeof   | リソースのデータ型を示す CURIE リスト。         |
| datatype | リテラルのデータ型を示す CURIE。             |

表 1: RDFa 属性の一覧

に property 属性と datatype 属性が指定されている。RDFa では xmlns 属性を URI のショートカットとして転用する<sup>15</sup>。

```
<tr xmlns:bib="http://bibtexml.sf.net/"
    xml:lang="ja"
    about="http://purl.org/JAEL/SEL/pub/2008/Takahashi-2009b.pdf">
  <td property="bib:title" datatype="">エチオピア言語の Web サイト構築</td>
  <td property="bib:author" datatype="">高橋 洋成</td>
  <td property="bib:pages" datatype="">5-10</td>
</tr>
```

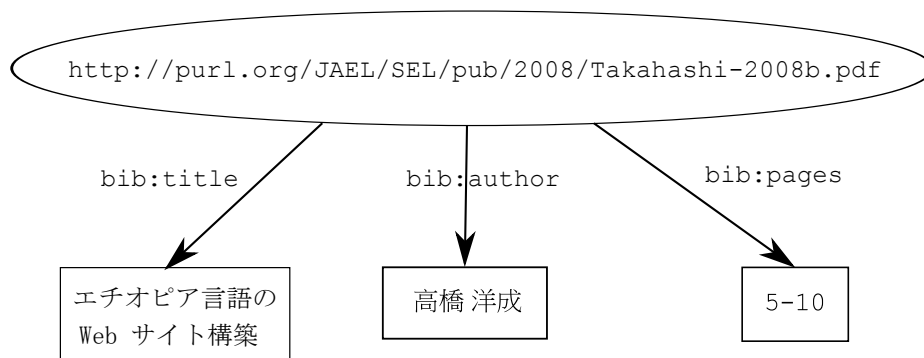
上記 XHTML データで使われている RDFa 属性とその意味は次の通りである。

- about 属性が主語リソースである PDF ファイルの URI を示す。
- property 属性が、目的語リテラルとのリンクの種類を示す。ここでは bib:title、bib:author、bib:pages の 3 つが表されている<sup>16</sup>。
- 空である datatype 属性は、要素内容が目的語リテラルであることを示す。

ここから次のような RDF グラフが構築される。XHTML で記述されたデータに RDFa 属性を埋め込むことで、著者・題名・ページ数などの書誌情報を曖昧さなく機械的に抽出することが可能になり、論文データをまとめてデータベース化するなどの用途に応用できる。

<sup>15</sup>CURIE と呼ばれる。RDFa 1.1 では prefix 属性を用いる。

<sup>16</sup>BibTeX 由来の BibTeXXML 語彙については高橋 (2007) を参照。



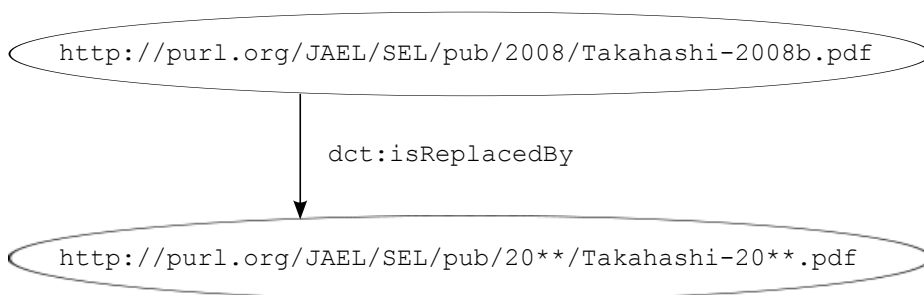
また、RDFaによるリンクの別例として、学術情報の関係を示す次のXHTML断片を見てみたい。

```

<tr ....
  about="http://purl.org/JAEL/SEL/pub/2008/Takahashi-2009b.pdf"
  xmlns:dct="http://purl.org/dc/terms/"
  rel="dct:isReplacedBy"
  resource="http://purl.org/JAEL/SEL/pub/2010/Takahashi-2011b.pdf">
  ....
</tr>

```

これは、高橋 (2009b) の論文が高橋 (2011b) の論文によって置き換えられたことを示す。dct:isReplacedBy というリンクの種類は、Dublin Core<sup>17</sup> の拡張語彙 DCMI Metadata Terms で定義されている<sup>18</sup>。このXHTMLデータから抽出されるRDFグラフは次のようになる。



このリンクを利用し、学術情報の利用者に対してその信頼性や代替・更新情報を提示することができる。学術情報の関係を示すのに利用可能な語彙の例を表2に挙げておく<sup>19</sup>。

<sup>17</sup>情報に関する情報 (メタ情報) の記述を共通化するための語彙セットであり、ISO 15836 として標準化されている。

<sup>18</sup><http://dublincore.org/documents/domain-range/>

<sup>19</sup>ここに挙げたのは、Dublin Core の基本語彙 relation の拡張語彙として定義されたものである。



|                    |                           |
|--------------------|---------------------------|
| dct:isVersionOf    | リソースは他のリソースの1バージョンである。    |
| dct:hasVersion     | リソースは別バージョンを持つ            |
| dct:isReplacedBy   | リソースは別リソースで置き換えられる。       |
| dct:replaces       | リソースは別リソースを置き換える。         |
| dct:isRequiredBy   | リソースは別リソースで必要とされる。        |
| dct:requires       | リソースは別リソースを必要とする。         |
| dct:isPartOf       | リソースは別リソースの一部である。         |
| dct:hasPart        | リソースは別リソースの一部を含む。         |
| dct:isReferencedBy | リソースは別リソースで参照、引用されている。    |
| dct:references     | リソースは別リソースを参照、引用する。       |
| dct:isFormatOf     | リソースは同じ内容を別形式で提供されたものである。 |
| dct:hasFormat      | リソースは同じ内容を別形式でも提供する。      |
| dct:conformsTo     | リソースが準拠している標準を示す。         |

表 2: 情報の関係を示す語彙の一覧

メタ情報を表現するには、本節で述べた RDFa 以外にも Microdata が近年注目を集めている。ウェブの検索サービスを提供する各社が RDFa や Microdata のサポートを表明している<sup>20</sup>。

### 3.5 著作権

現在のところ、電子ジャーナル *Studies in Ethiopian Languages* で配信される学術情報は制限なく自由にダウンロードできる。今後、一部の情報については会員制にし、会員間の検討を経た上で情報公開されるケースも考えられよう。動画ファイルの場合は肖像権など、クリアしなければならない法的な問題もある。

論文などの学術情報の剽窃を検知するシステムを用意する予定はない。従来の冊子体における学術情報と同様、利用者の良心に委ねられる。ただ、ウェブ検索エンジンに拾われる情報が引用の範囲内であるか、それとも転載と見なすべきであるのかなど、従来とは異なる問題が生じうる可能性もある。

## 4 おわりに

学術情報を素早く発信し、種々のコストを削減できる電子ジャーナルの必要性は、今後ますます大きくなるだろう。本稿では日本における電子ジャーナルの種類と配信状況を概観するとともに、著作権や会員の権利などのクリアしなけれ

<sup>20</sup>検索サービスを提供する各社が Schema.org を立ち上げ、検索性を向上させる語彙の取りまとめを行っている。

ばならない問題点についても整理した。また、新ジャーナル *Studies in Ethiopian Languages* を立ち上げた際に検討された個々の技術論が、今後導入される学術情報のオンライン配信の一助になることを願う。

#### 【参照文献】

- 杉井学 2006 「GIS を用いた言語分布地図の作成と地理的条件との相関解析」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2006）』 9-14.
- 杉井学 2008 「GIS サーバの機能増強と検索システムの構築」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2007）』 9-11.
- 杉井学 2009 「属性情報の地図分布を解析する GIS 構築」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2008）』 1-4.
- 高橋洋成 2008 「XML を利用した文献データベースの構築」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2007）』 13-29.
- 高橋洋成 2009 「エチオピア言語の Web サイト構築」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2008）』 5-10.
- 高橋洋成 2010 「XHTML を利用した語彙データの格納形式」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2009）』 1-19.
- 高橋洋成 2011 「エチオピア言語研究の Web サイト構築」 乾秀行（編）『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築（Cushitic-Omotc Studies 2009）』 1-7.