

ポータルの試験実装とメタデータ仕様

麻生和彦

KAZUHIKO ASOU

東京大学

THE UNIVERSITY OF TOKYO

行木孝夫

TAKAO NAMIKI

北海道大学

HOKKAIDO UNIVERSITY

Abstract

World Digital Mathematics Library に関連する情報を報告し、数学会および各教室の紀要等に関する情報を統合したポータルサイトの構築に関する考察を行い、Hokkaido Mathematical Journal の電子化に関して述べる。WDML に関しては麻生が執筆し、それ以外を行木が作成した。

1 WDML に関する報告

2005 年 4 月 15 日 (金) 9:00 から 17 日 (日) 12:00 にかけて MSRI で WDML に関するワークショップが開催された。本節はその報告である。

1.1 概要

正式名称は”Digitizing Older Mathematical Journals”であり、開催場所は Mathematical Sciences Research Institute (Berkeley, USA) である。主催者は Philippe Tondeur (イリノイ大学、元 NSF メンバー)、David Eisenbud (MSRI)、John Ewing (AMS) となっている。

参加総数 38 名、年齢は 50 歳前後でほぼ全員がそれぞれの所属での担当責任者であった。国別にみるとアメリカ、イギリス、ドイツ、フランス、日本 他であり、アジアからは麻生と行木である。所属別には大学 (MSRI, UCB, Cornell, Cambridge, 北大, …)、研究所 (MSRI)、団体 (AMS, EMS, JMS, LMS, CMS, SIAM, …)、出版社 (Springer, Euclid, …) となり、専門別には数学者、司書、出版社、経済学者などであった。

1.2 ワークショップの様子

会議の進行は、通常の講演形式でなくセッションごとに話題提供者がテーマに関するトピックスや問題点を提案しその後、ほとんどの時間を全員で議論する形式であった。内容は、WDML プロジェクトをどう実現していくか、問題点をどう解決していくかなど具体的な議論が中心。参加者は、既に何度か同じメンバーで会議を行っているようで初めての参加者は少なかった (日本以外いなかった)。

1.3 WDML プロジェクトとは

World Digital Mathematics Library の略称であり、デジタイズに限らず web of literature を目指している。各大学/研究機関で行われている電子ジャーナルのオンライン公開プロジェクトを統一運用化を目指すものである。

具体的な目標は、過去に発行された紙媒体のすべてのジャーナルに対しスキャナーを使って電子化を行いオンライン上で公開し、公開されたジャーナルに含まれるそれぞれの論文ごとにメタデータ(書誌情報)を作成すること、作成されたメタデータをオンライン上で自動収集できるようにし、収集されたメタデータを元にして数学分野の代表的なレビュー雑誌 Mathematical Reviews(MathSciNet)や Zentralblatt Math から論文を参照でき、さらにリファレンスからも参照できるようにすることである。

これを実現するには各プロジェクトによって微妙に違うメタデータを標準化する必要があり、ベースとなるのは、Dublin Core の Metadata Initial 's element set であろう。メタデータの交換方式は、OAI-PMH を採用する。分類(classification)は MathSciNet の分類(MSC)を利用し、国際的に利用可能にするためには文字コードに Unicode を採用する必要がある。オープンでフリーなアクセス環境を提供することが理想だが、アクセスにかかるコスト負担は問題である。

進行中のプロジェクトはフランスの NUMDAM(WDML の試験実装 mini-DML)、ドイツでは Emani(中国、韓国 KISTI が参加済) アメリカでは EUCILID などである。

1.4 ポータル

ユーザに活用してもらうために電子化されたジャーナルのリンクを集約したサイトが必要である。MathSciNet, Zentralblatt Math が DML のポータルサイトであり、オンラインで公開されているジャーナルに対してすべてリンクを作成する。ジャーナルがオンライン公開された後、ポータルサイトに反映するまでの時間は重要である。Google Scholar との違いは何かといえば、分類があるかどうかである。

1.5 権利関係

出版社の権利、著者の権利の明確化が必要だが、どこのプロジェクトも困っている。過去の文献に対する権利の扱い、国ごとの copyright の違いなど、個別対応になる。出版社の利益を守るためには公開までの期間を設定(moving wall)したほうがよい。ポータルサイトの役目として権利所有リストの保守もある。

1.6 日本ではどうするか

ジャーナルの電子化は、現在の研究遂行のためだけでなく、未来の研究者や社会への説明責任のために必要なものである。日本の数学界の数年後を考えると早急にジャーナルの電子化を進める必要がある。

電子化と OAI-PMH 対応を独自で行う例は、北大の「数学の海」プロジェクトでの HMJ にある。電子化したジャーナルに関して日本独自のポータルサイト jpDML(仮称) を立ち上げることも必要。日本でのジャーナル電子化をコーディネートするのはどこか。日本数学会か数学コンソーシアムか。

数学者だけではマンパワーの問題で不可能な時期が来ると思われるので、NII のリポジトリ事業、JST の J-STAGE との連携を考えるべきであろう。次回の WDML ワークショップには、図書館側の立場からも参加を要請するべきである。

2 ポータルの構築

数学会は欧文誌(紀要), 「数学」, 「数学通信」など、多様な形態の定期刊行物を発行し、各数学教室においても紀要、講究録、公開講座の資料など多くの出版物を持つ。

通常の流通経路を除き、従来はその一部が web ページとして公開され、サーチエンジンやリンク等を経由して利用されていると考えられる。問題は、これらの個別コンテンツ各々について WDML をはじめとするポータルに組み込む手段と手続きであろう。

本稿では、ポータルとしてこれらを統合する技術の一部を紹介し、WDML への前段階となるであろう国内でのポータル構築に関して考察する。

2.1 基礎技術

出版物の電子化は論文等の単位によって PDF 等のファイルに収め、HTML によって web から公開するものという認識が一般的であろう。実際、いくつかの紀要はこのように公開されている。つまり、TCP/IP 上の HTTP アクセスによる公開を基礎とすることになる。

このような基礎の上に何を構築すればその価値を向上させられるだろうか。一つのアイデアは何らかのリンク集を作成し、その上に Google 等を用いた簡易検索サービスを構築することであろう。場合によってはリンク先サイトと協力してデザイン等を統一するなどの共同作業も可能である。多くのポータルサイトはこのように作成されている。

しかしながら、このような人力に頼ったポータルの維持はそこに参加するサイトが増えるほど困難になる。

加えて、ポータルを作成した後からは多くのアイデアが出るものである。特に分野を特定しない電子化された刊行物から特定の分野の論文や記事を取り出して再構成したいなどの発想は誰でも抱くものであろう。

このような発想に対応しようと思えばその作業は膨大なものになる。このような作業こそ計算機に任せるべきである。幸い、同様の問題は web 一般においても重要視されており、多くの技術が提案され、実装され、普及している。Wiki, blog などの隆盛はその一部であろう。

誤解を恐れずに述べることにすれば、電子図書館に関連して有効な技術が存在する。各地に散在する書誌情報(メタデータ)を同時にアクセスすることで検索を一元的に実行するプロトコル Z39.50 は早くから実装されてきた。現在では Open Archives Initiative (OAI) が提唱するメタデータ交換

プロトコル (OAI-PMH, RFC2119) の実装が増え、いくつかのコンテンツ管理システムはこれをサポートしている。EPrints, DSpace などはその代表である。

2.2 保存

ポータルを支えるものは利用者には見えない保存に関する機能である。入力, 保存と維持にかかるコストに応じ、JSTAGE や EUCLID など既存のジャーナルプラットフォームを利用するか、機関リポジトリ等を利用するか、独自に行うかという選択になる。技術的な蓄積を重視するなら後者を選択し、即効性を期待するなら前者である。機関リポジトリはこれから成長するプラットフォームであろうから、中間的な性格を持つ。第一候補は既に実績のある EUCLID や JSTAGE である。第二候補として千葉大や北大に例を見る機関リポジトリが続く。

独自の電子化はリスクが高い。理由を以下に述べる。ハードウェア、OS, web サーバの維持を含むシステム管理、更新やチェックを含む web コンテンツ管理、入力に関わる教育、OCR 作業などを考えればコンテンツ管理システム (CMS) は必須となる。Web ページの設計についてさえも HTML を書いてディレクトリ構造を維持できるような支援員の確保は至難になりつつある現在、これらの業務をこなす人材を確保することは難しい。

ただし、独自の電子化をあきらめた場合には時代の変化へ追従することも不可能になる。ここは組織としての政策判断が必要であろう。

将来的に何が必要になるか不明であるならば、WDMML の仕様変化への対応も考慮し、内部で維持するメタデータは可能なかぎり詳細にする必要がある。内部メタデータを適切に管理できる CMS はリポジトリ構築ツールであり、DSpace, GNU EPrints 等は紀要や予稿の保存には最適な CMS といえる。WDMML に必要な OAI-PMH 対応は結果としてついてくる。

2.3 公開

ポータルとして利用者側に必要な機能は、各紀要ごと分野ごとの一覧、検索、新着通知、MR へのリンク (リファレンス、本文) などであろう。これらは公開に関する機能と考えられる。

主な対象読者は研究者だが、公開した瞬間に研究者だけが読むとは限らない。従って、読者階層別に分ける機能も必要である。

総合報告などはそれだけを横断的に取り出すなど、公開手段としては可能な限り体系的に公開したい。しかしながらカテゴリ分けは労力が大きいので、これを人手で解決しようとすれば問題があらうし、現在の新理論も後世では教育資料になるかもしれない。

現実的な解決法として、当面は MSC の後二桁で対応する。MSC は後二桁の 00 から 06 の範囲は全分野共通になっており、階層化を可能にしている。数論の例を挙げておく。

```
11-xx Number theory
  11-00 General reference works
  11-01 Instructional exposition
  11-02 Research exposition
  11-03 Historical
  11-04 Explicit machine computation and programs
  11-06 Proceedings, conferences, collections, etc.
  11Axx Elementary number theory
  11Bxx Sequences and sets
```

```

11Cxx Polynomials and matrices
11Dxx Diophantine equations
11Exx Forms and linear algebraic groups
11Fxx Discontinuous groups and automorphic forms
11Gxx Arithmetic algebraic geometry (Diophantine geometry)
11Hxx Geometry of numbers
11Jxx Diophantine approximation, transcendental number theory
11Kxx Probabilistic theory: distribution modulo  $i$ ; metric theory of algorithms
11Lxx Exponential sums and character sums
11Mxx Zeta and  $L$ -functions: analytic theory
11Nxx Multiplicative number theory
11Pxx Additive number theory; partitions
11Rxx Algebraic number theory: global fields
11Sxx Algebraic number theory: local and  $p$ -adic fields
11Txx Finite fields and commutative rings (number-theoretic aspects)
11Uxx Connections with logic
11Yxx Computational number theory
11Z05 Miscellaneous applications of number theory

```

従って、後二桁に 00 から 06 までの値を含む MSC を指定したコンテンツは公開時に別扱いするという処理も可能である。通常の新着順などの表示とは別に、想定する利用者として高校生から研究者、評価機関までを想定した処理が考えられよう。

公開講座、サーベイ、新着論文、教材 (OCW) などによる分類も有効であると思われる。

2.4 実装へ向けて

これらの実装に当たってはメタデータの収集による実現を考えよう。当面は OAI-PMH によるメタデータ収集が現実的である。メタデータ形式は oai_dc に準拠し、内部構造は Euclid に従えば比較的簡単に実現できる。内部メタデータについては EPrints なり DSpace なりの内部メタデータに従えば最低限の内容は確保できる。

独自に電子化する場合は実質的に保存サーバが公開サーバを兼ね、開発ツール、web ページ等をミラーすることになるであろう。

EUCLID および独自電子化を行った Hokkaido Mathematical Journal からハーベストしたメタデータを用い、簡単なポータルの実装をした web ページを図 1 に示す。

最も右側に各ジャーナルへのリンク情報を示している。

上部には発行年による統合インデックスへのリンクを置き、JavaScript によってクリックは不要なインターフェースとしている。発行年を指定すれば MSC インデックスによって分類が表示される。それぞれのリンクによって論文のリストを表示する。

分野を特定しないジャーナルを集約し、再構成することで比較的扱いやすいインターフェースになっていると思われる。

2.5 メタデータの例

2.5.1 メタデータの仕様 Project EUCLID – OAI_DC

Project Euclid は数学関連ジャーナルのポータルとしてコーネル大学にて運営されている。OAI-PMH によって提供するメタデータは下記のような仕様であり、subject は MSC と自由キーワードとなっている。identifier には URI および Euclid 内部表現符合、書誌情報を持たせている。

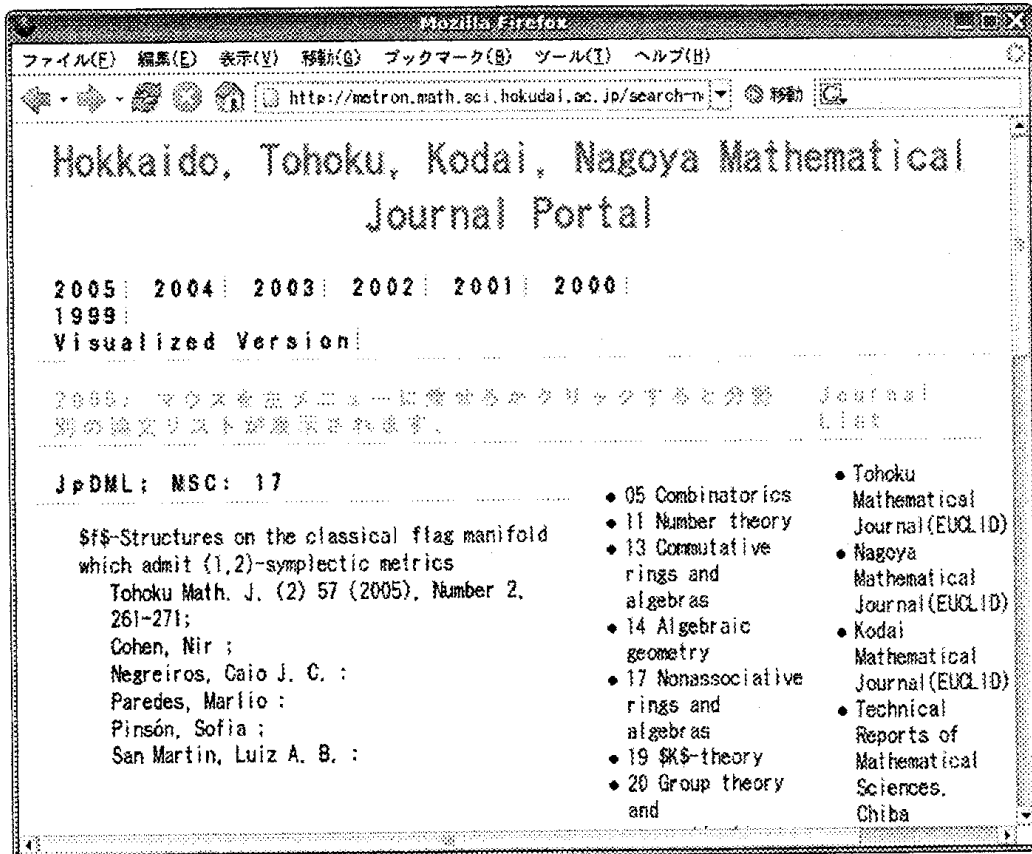


図 1: スクリーンショット

title: The distribution of general final state random variables for
 stochastic epidemic models
 creator: Ball, Frank
 creator: O'Neill, Philip
 subject: 92D30 (MSC2000)
 subject: 60K99 (MSC2000)
 subject: Epidemic
 subject: final size
 subject: final severity
 subject: Gontcharoff polynomial
 publisher: Applied Probability Trust
 date: 1999-06 (Issued)
 type: text
 format: application/pdf
 identifier: <http://ProjectEuclid.org/getRecord?id=euclid.jap/1029349824>
 identifier: euclid.jap/1029349824
 identifier: Citation: J. Appl. Probab. 36 (1999), no. 2, 473-491
 language: en
 rights: Copyright 1999 Applied Probability Trust
 description: In this paper we introduce the notion of general final
 state random variables for generalized epidemic models. These random
 variables are defined as sums over all ultimately infected
 individuals of random quantities of interest associated with an
 individual; examples include final severity. By exploiting a
 construction originally due to Sellke (1983), exact results
 concerning the final size and general final state random variables
 are obtained in terms of Gontcharoff polynomials. In particular, our
 approach highlights the way in which these polynomials arise via
 simple probabilistic arguments. For ease of exposition we focus
 initially upon the single-population case before extending our
 arguments to multi-population epidemics and other variants of our
 basic model.

2.5.2 メタデータの仕様 千葉大学リポジトリ - OALDC

機関リポジトリの例として千葉大学附属図書館から例を採った。subject は NDC(日本十進分類)と自由キーワードであり、日本語著者名を追っている点に注目するべきである。identifier には

title: Dualities for the Domany-Kinzel model
 creator: Katori, Makoto
 creator: Konno, Norio
 creator: Sudbury, Aidan
 creator: Tanemura, Hideki
 creator: 香取, 眞理
 type: 雑誌掲載論文
 type: 研究成果-論文
 identifier: <http://mitizane.ll.chiba-u.jp/metadb/up/C0000050988/MR2054582.dvi>
 identifier: BibliographicCitation:info:ofi/fmt:kev:mtx:ctx:Journal of
 Theoretical Probability Vol.17 no.1 page.131-144 (2004)
 language: eng
 relation: hasVersion:<http://dx.doi.org/10.1023/B:JOTP.0000020478.24536.26>

3 Hokkaido Mathematical Journal の電子化

Hokkaido Mathematical Journal は 1972 年から Journal of Faculty of Science, Hokkaido University, Ser. 1 の後続誌として刊行されている。

この電子化を独自に行うに当たり次の点に留意した。

- メタデータハーベスティングへの対応

- 登録作業等の軽減

これを実現するためにコンテンツ管理には EPrints 2.3 を利用した。OAI-PMH によるメタデータハーベスティングは実装されており、多少の調整を施すだけでよい。

結果として、web ページ等の作成はトップページ、共通スタイルファイル等を作成しておくだけで済み、特に技術的知識を要求せずに論文の登録をできるようになっている。

交換先とは IP アドレスにより公開する。販売先とは代理店を経由して IP アドレスを取得する。IP アドレスの管理は簡単ではない。特にクラス C 数十個で登録する場合はミスが生じやすく、今後の問題である。

新規に発行する 34 巻以降については著者から著作権委譲の確認をとり、33 巻以前についてはまず告示で対応する予定である。最新巻から 3 年より古い巻に関しては Open Access とすることになっている。

内部メタデータについては EPrints の内部形式を用い、以下のように設計した。

```
<eprintsdata>
<record >
  <field name="datestamp">2005-05-16</field>
  <field name="type">article</field>
  <field name="creators" id="">
    <part name="given">Rikio</part><part name="family">Yoneda</part>
  </field>
  <field name="title">Multiplication operators, integration
operators and companion operators on weighted Bloch space</field>
  <field name="ispublished">pub</field>
  <field name="subjects">30D55(MSC2000)</field>
  <field name="subjects">QA</field>
  <field name="full_text_status"></field>
  <field name="keywords">Bloch space, boundedness, compactness.</field>
  <field name="abstract"></field>
  <field name="date_sub">2003-01-07</field>
  <field name="date_issue">2005-02-00</field>
  <field name="date_effective">2005-02-00</field>
  <field name="publication">Hokkaido Mathematical Journal</field>
  <field name="volume">34</field>
  <field name="number">1</field>
  <field name="publisher">Department of Mathematics, Faculty of
Science, Hokkaido University; Sapporo, Japan</field>
  <field name="pagerange">135-147</field>
  <field name="pages"></field>
  <field name="refereed">TRUE</field>
  <field name="issn">0385-4035</field>
  <field name="fileinfo">
pdf;http://hmj.math.sci.hokudai.ac.jp/secure/00000015/01/302.pdf
  </field>
  <field name="referencetext"></field>
</record>
</eprintsdata>
```

この内部形式を OALDC へ変換し、メタデータハーベスティングへ対応する。

4 おわりに

世界的な電子化関連の動向に追随することは容易ではないが、必要であることも事実である。日々提唱される新技術から本質を見極めて対応しなければならない。

参 考 文 献

- [1] 数理解析研究所講究録 1446 電子情報交換に関する最近の話題
- [2] Stephen Pinfield, Mike Gardner and John MacColl, *Setting up an institutional e-print archive*, Ariadne. Issue 31, March-April, 2002,
(邦訳: <http://www.nii.ac.jp/metadata/oai-pmh/eprints/>)
- [3] Budapest Open Access Initiative, *A Guide to Institutional Repository Software*,
<http://www.soros.org/openaccess/software/>
(邦訳: http://www.nii.ac.jp/metadata/oai-pmh/osi/_ir/_software/_guide.html)