



WDMLプロジェクトの解説

麻生 和彦

東京大学 大学院数理科学研究科

asou@ms.u-tokyo.ac.jp



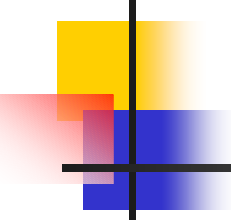
本日のお話

1. WDMLワークショップ
2. WDMLとは？
3. 実現の方法
4. 実現に向けての課題
5. 当面の目標



WDMLワークショップ

- 正式名称
“Digitizing Older Mathematical Journals”
- 開催日時
2005年4月15日(金)～17日(日)
- 開催場所
Mathematical Sciences Research Institute
(Berkeley, USA)
- 主催者
Philippe Tondeur(イリノイ大学、元NSFメンバー),
David Eisenbud (MSRI),
John Ewing (AMS)



WDMLワークショップ WDMLプロジェクト

- **World Digital Mathematics Library**の略称
- 2003～2004年に行われたDMLプロジェクトの発展形
- オンライン公開されているジャーナルの相互運用
Web of literature (NOT digitize)
- 参考文献
“The Digital Mathematics Library”
Allyn Jackson, Notices of the AMS,
Vol. 50 (8), PP 918-923, September 2003
<http://www.wdml.org/publications/comm-jackson.pdf>

WDMLワークショップ

参加者

- 参加総数 38名
- 年齢構成は、50歳前後が中心。
- 参加者のほぼ全員がそれぞれの所属での担当責任者。
- 国別：アメリカ、イギリス、ドイツ、フランス、日本 他

- 所属別：
 - 大学 (MSRI, UCB, Cornell, Cambridge, 北大, ...)
 - 研究所 (MSRI)
 - 団体 (AMS, EMS, MSJ, LMS, CMS, SIAM, ...)
 - 出版社 (Springer, Euclid, ...) 他

- 専門別：数学者、司書、出版社、経済学者 他

WDMLワークショップ

感想

- ジャーナルのオンライン公開は、現在の研究者のためだけでなく**未来の研究者や社会貢献のため**に必要
- 司書(図書館員)の積極的な参加
- MathSciNetを中心とした相互運用
→電子化されていない日本のジャーナルの価値が低下していくことへの危機感

「早急にジャーナルの電子化、オンライン公開を進めなければ、数年後の日本は...」



WDMLとは？

1. 電子化 (PDF + OCR)
2. メタデータ
3. オンライン公開
4. 相互運用 (OAI-PMH)

WDMLとは？

1. 電子化

これまでに発行されたすべてのジャーナルを
スキャナーを使いPDF化する

- 文字認識アプリケーション(OCR)を使って検索可能なPDFファイル(テキスト情報を付加)を作成 #あくまで全文検索用
- OCRで数式もTeX形式(MathML形式他)に変換できる！ #鈴木氏@九大の研究成果

WDMLとは？

2. メタデータ

論文ごとにメタデータ(書誌情報)を作成

- メタデータに含まれる情報
 - ジャーナル名、号、巻、ページ、...
 - タイトル、著者、キーワード、分類、...
 - 本文(PDFファイル)のURL

WDMLとは？

3. オンライン公開

論文のPDFファイル、メタデータを
オンライン公開

- PDFファイルの公開は、Webサーバを利用
- メタデータの公開は、メタデータを交換するプロトコルOAI-PMHを利用
- OAI-PMH対応の公開アプリケーションで代表的なものは、ePrints, Dspaceなどがある。

WDMLとは？

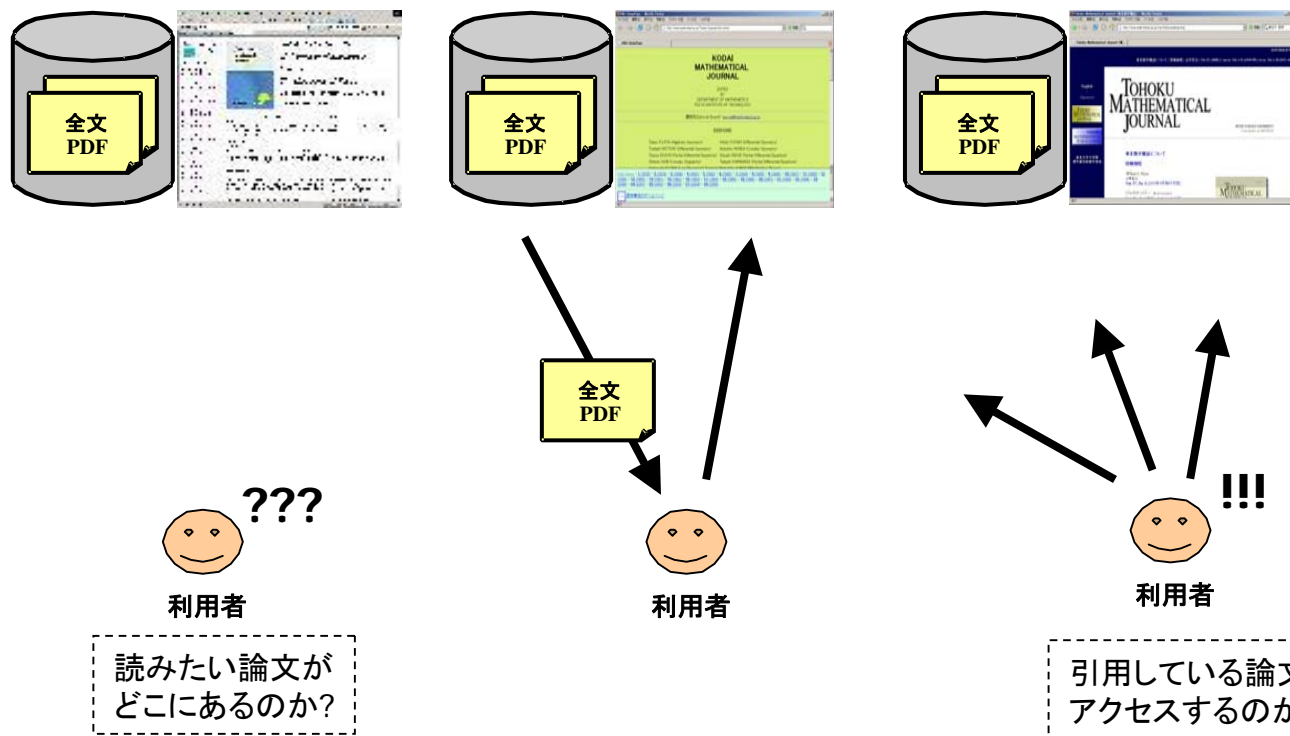
4. 相互参照

OAI-PMHを使ってメタデータを自動収集

収集したメタデータを使って、

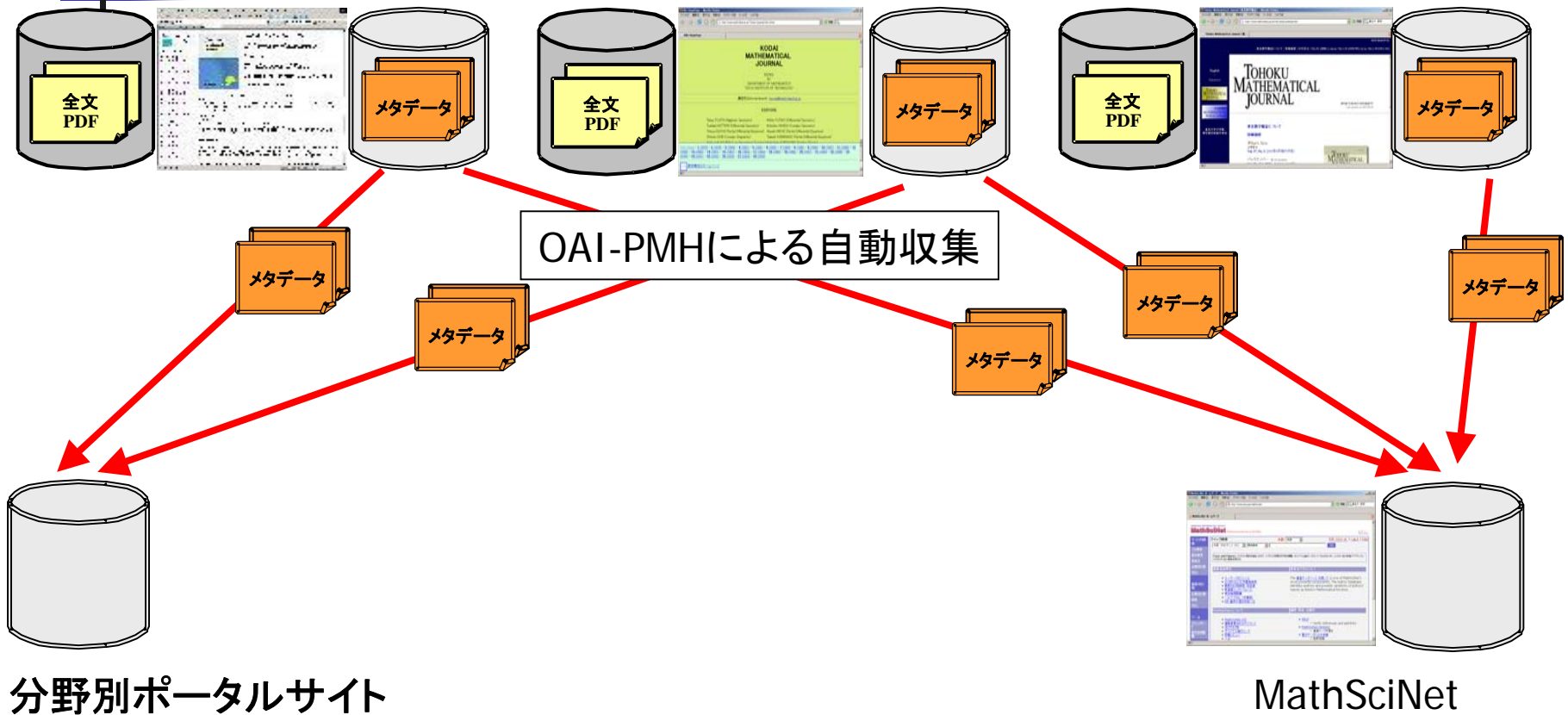
- Mathematical Reviews (MathSciNet)や Zentralblatt Mathから論文のPDFファイルへの直接リンクを自動作成
- 各論文の引用文献から論文のPDFファイルへの直接リンクを自動作成
- 分野別や個人別などさまざまなポータルサイト(リンク集)も簡単に自動作成可能

従来のオンライン公開



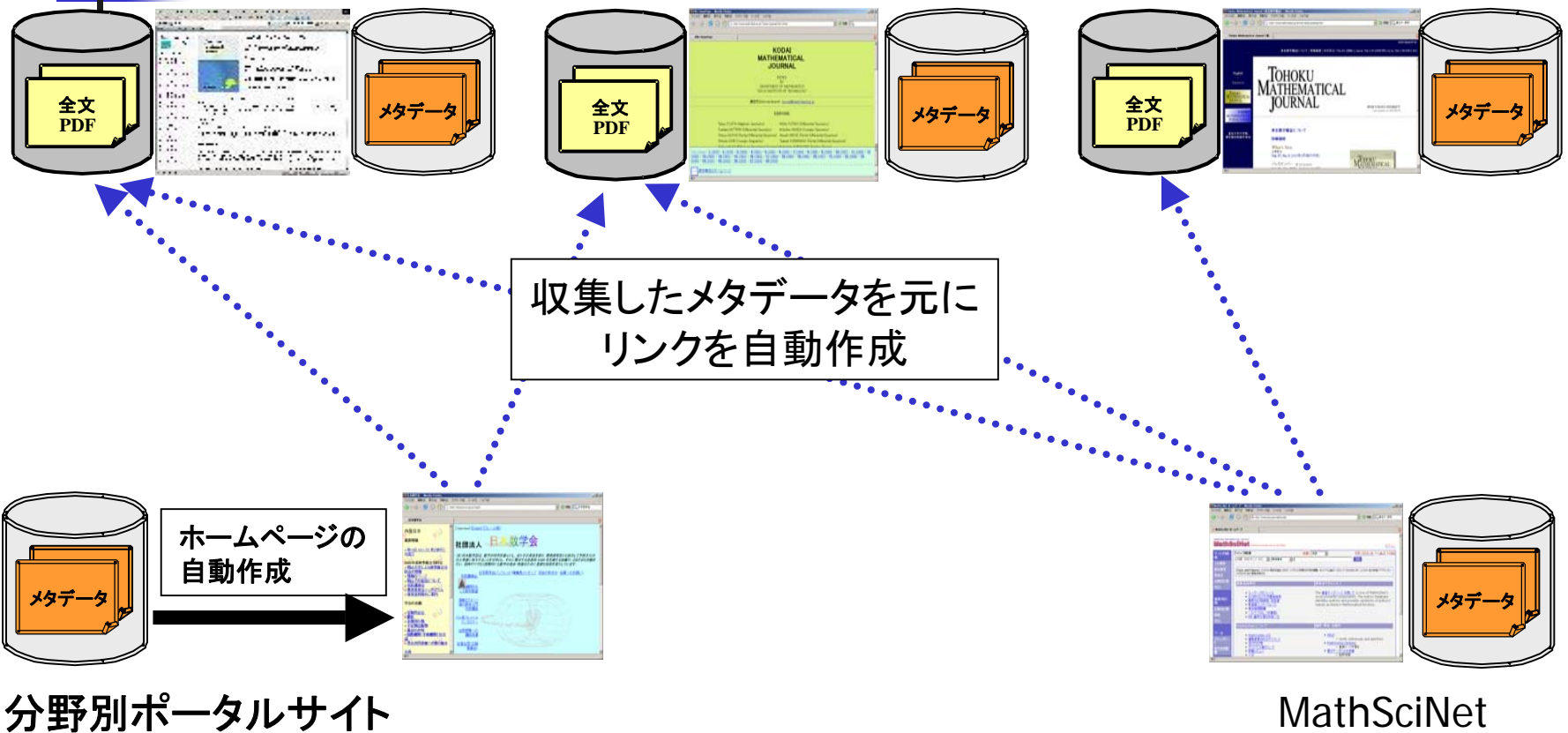
WDMLでのオンライン公開 ①

- OAI-PMHによるメタデータの自動収集 -

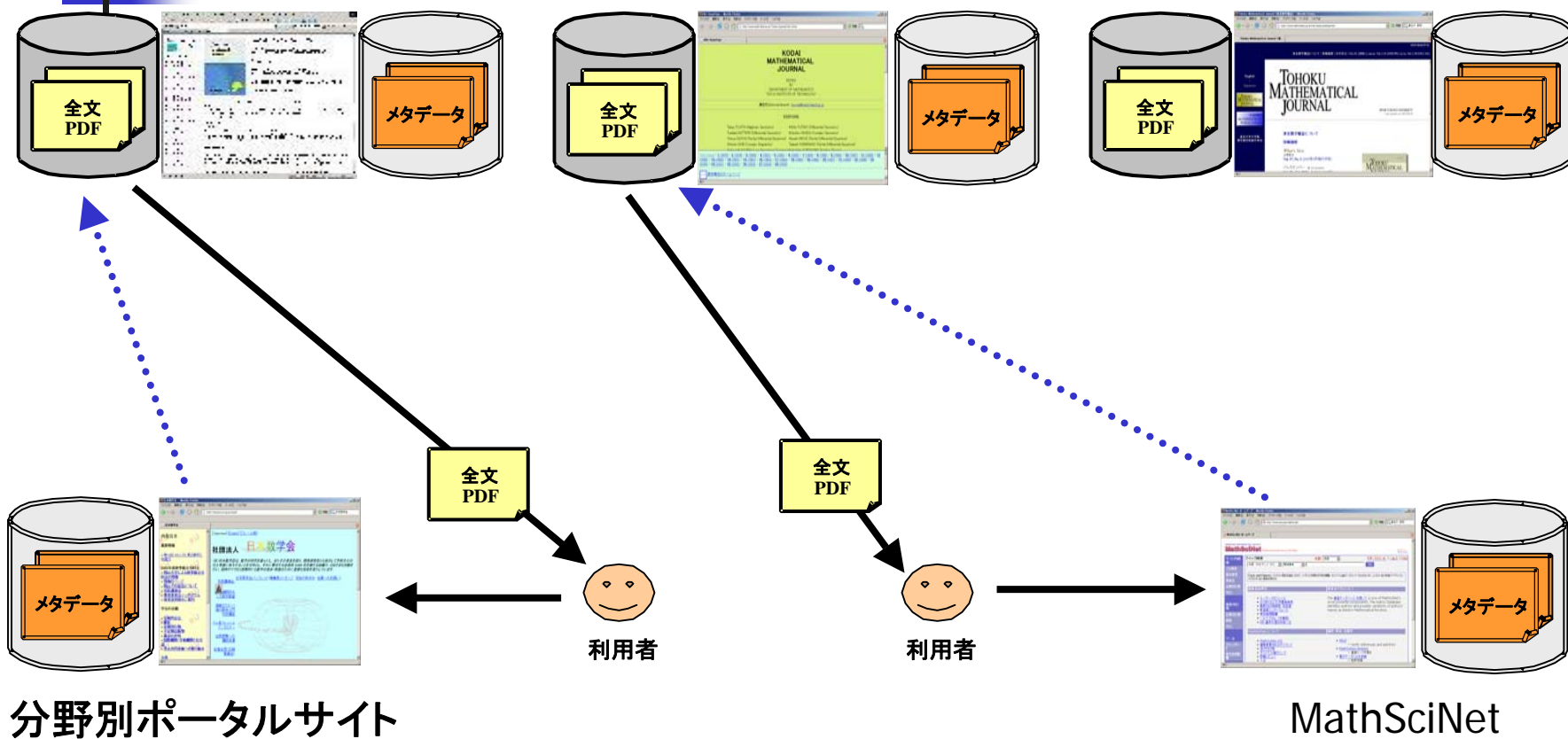


WDMLでのオンライン公開 ②

- リンクの自動生成 -



WDMLでのオンライン公開 ③



分野別ポータルサイト

MathSciNet

2005年9月21日(水)

日本の数学文献の電子化について



現在進行中のプロジェクト

- NUMDAM: フランス
#WDMLの試験実装mini-DML
- Emani: ヨーロッパ
#中国、韓国KISTIが参加済
- Goettingen: ドイツ
- EUCILID: アメリカ
- 数学の海: 北大
- ...



WDML実現の方法

1. 冊子の収集
2. 電子化とメタデータ作成
3. オンライン公開と相互運用



WDML実現の方法

冊子の収集

- これまでに発行されたジャーナルすべてを収集
- ただし、スキャナーで読み取るために裁断可能であること
- もし、一部しかない場合は...

WDML実現の方法

電子化とメタデータ作成

- スキャナーを用意し自前で電子化
 - 必要機材: 両面スキャナ(要シートフィーダ)
 - 必要アプリ: OCRソフト, PDF作成ソフト
- 業者に委託
 - # 30冊(18,000頁)で約60~100万円

WDML実現の方法

オンライン公開と相互運用

- 教室独自で行う
- EUCLIDを利用
#国内の数大学が既に利用
- J-STAGE(科学技術振興機構 JST)を利用
#OAI-PMHへの対応は検討中
- 各大学図書館の機関リポジトリを利用
#正式に稼動している大学は少ない




WDML実現に向けての課題

- PDFファイル作成時の解像度
- メタデータの標準化
#項目、文字コード、数式の表現形式
- 著作権などの取り扱い
- 電子化、オンライン公開にかかる費用
- 公開用サーバの運用管理
- ...



WDML実現に向けての目標

- 日本全体
- 大学・研究機関・分科会



WDML実現に向けての目標

日本全体

- 啓蒙活動
- 実験プロジェクトの立ち上げ
#メタデータの自動収集、ポータルサイトの作成...
- ガイドラインの作成、配布
 - ジャーナルの電子化ガイドライン
PDF作成時の解像度、メタデータの項目...
 - 著作権などの権利関連対策ガイドライン
- 国内プロジェクトとの連携
J-STAGE(JST), 機関リポジトリ(NII), ...
- WDMLプロジェクトへの積極的な参加



WDML実現に向けての目標

大学・研究機関・分科会

**いますぐ、ジャーナルを
集めてください!!!**

電子化とメタデータの作成準備を進めてください!!!

組織内の合意、予算の確保、...