大学図書館の抱える

紙とデジタル
ジレンマと挑戦、連携

メディアセンター本部
入江　伸

ジレンマのはなし

紙とデジタルのジレンマ

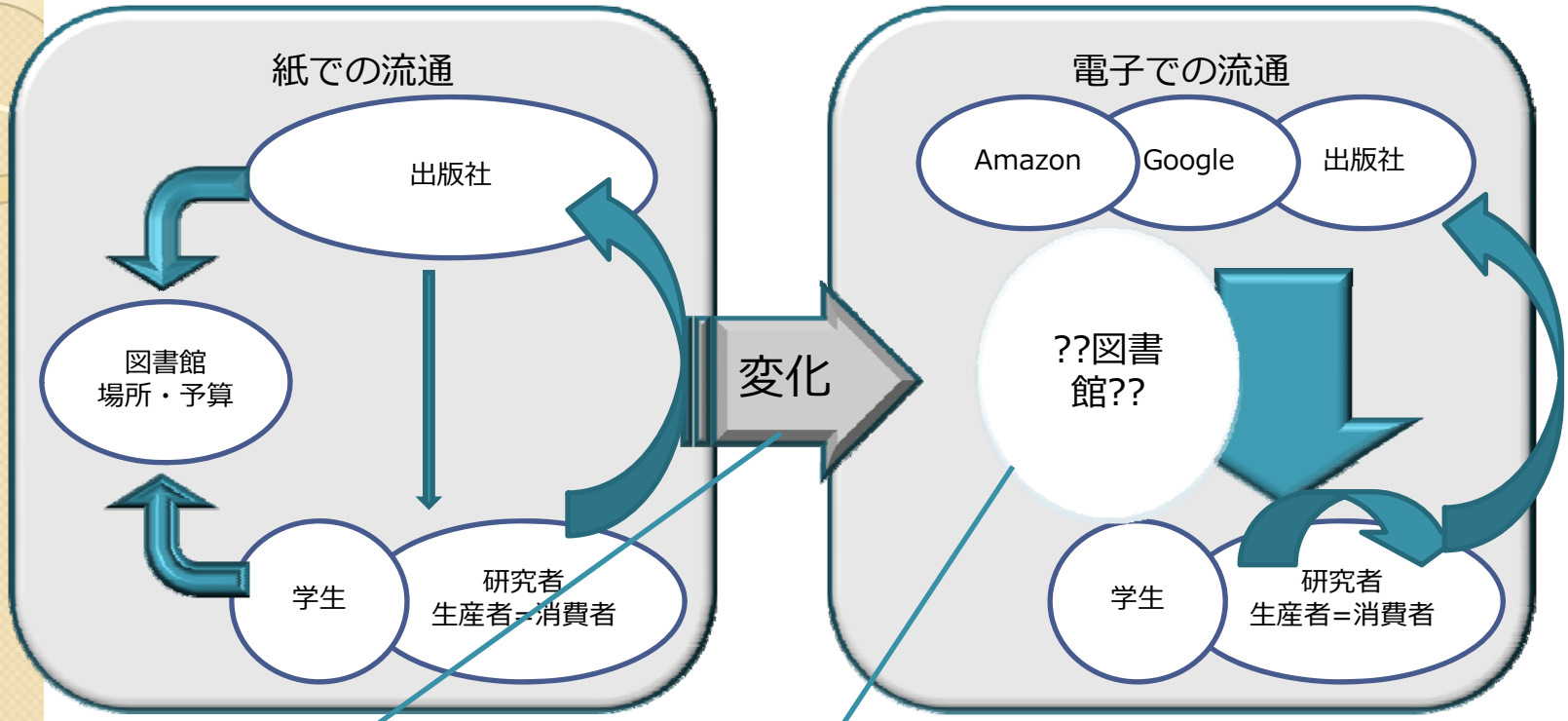
- 著作権のジレンマ
- デジタルデータ保存のジレンマ
- 紙とデジタル両立を運用しないといけないというジレンマ
  - ハイブリッド運営としてはコストとしては夢物語
  - 紙とデジタルのバランスと変化をどうするか？
  - 大胆なモデル改革が必要

大学図書館では存立基盤の問題

- 紙をベースとした運営モデル
  - 紙を前提とした著作権の制限
  - 書庫、予算、組織も紙のサービスが前提
- デジタルを前提として運営モデルへの転換？
  - デジタルデータをサービスするための特権？
    - インターネットと紙でのサービス格差
  - デジタルデータの信頼性確保？
  - デジタルデータを保存するための設備、予算、技術？
    - 紙の保存を前提としたデジタル化
    - ボーンデジタルデータの保存は？
- 電子図書館はどこへ行ったのか？
  - 90年代電子図書館の限界は　社会モデルとの連携がなかったこと

挑戦のはなし

アカデミック情報流通での図書館の役割の変化

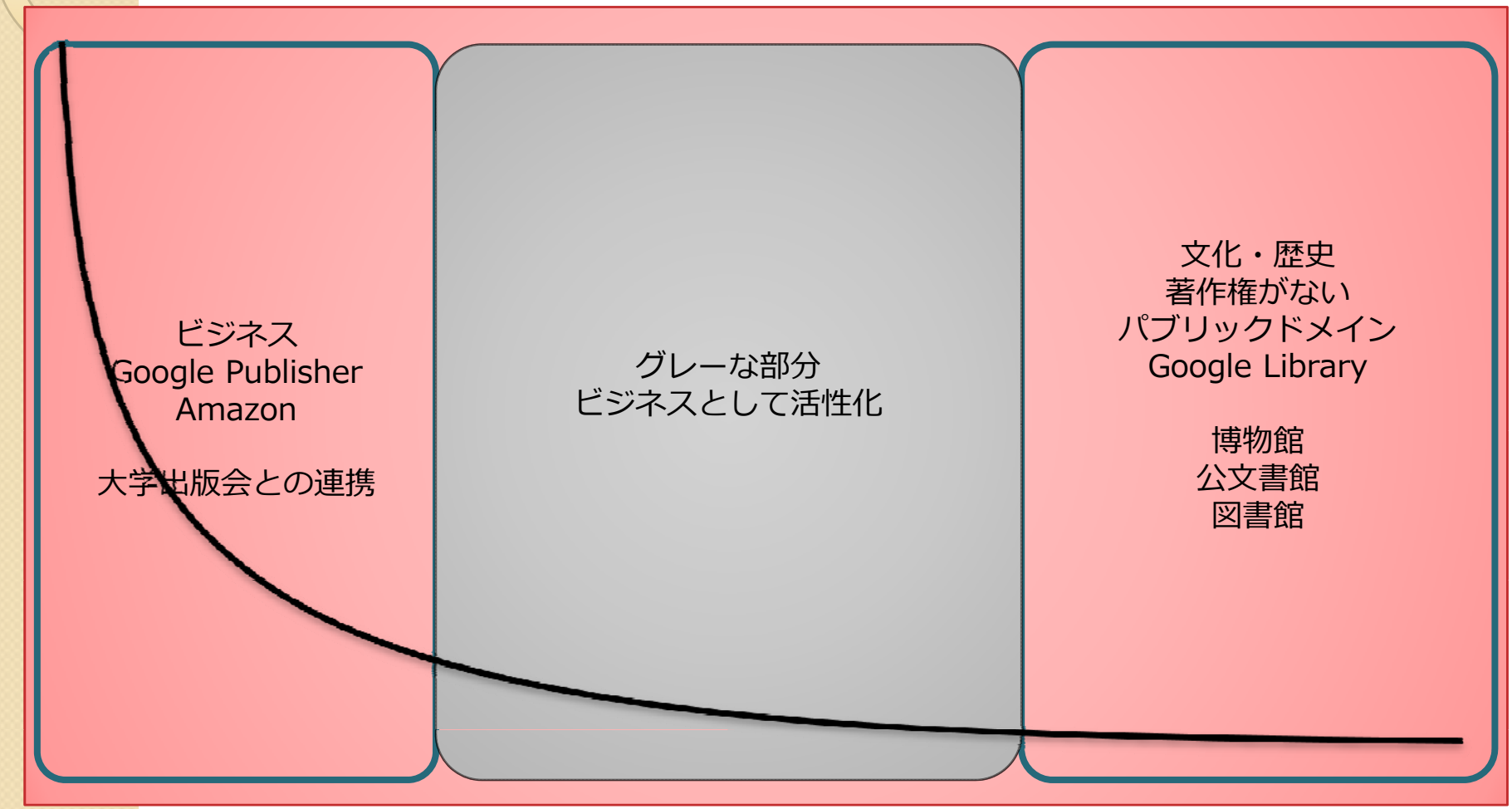


時間と空間からの自由
生産者と消費者の直結
Web One stop
対象資料、メディアの拡大

図書館はどういう役割を果たせるのか？
何を要求されるのか？
データ、サービスをすべて見直す必要
図書館特殊ではなく、流通データを活用し付加価値
蔵書だけではなく、多様な資料・メディアを対象

デジタイジングへの挑戦

long tail モデルでの分担



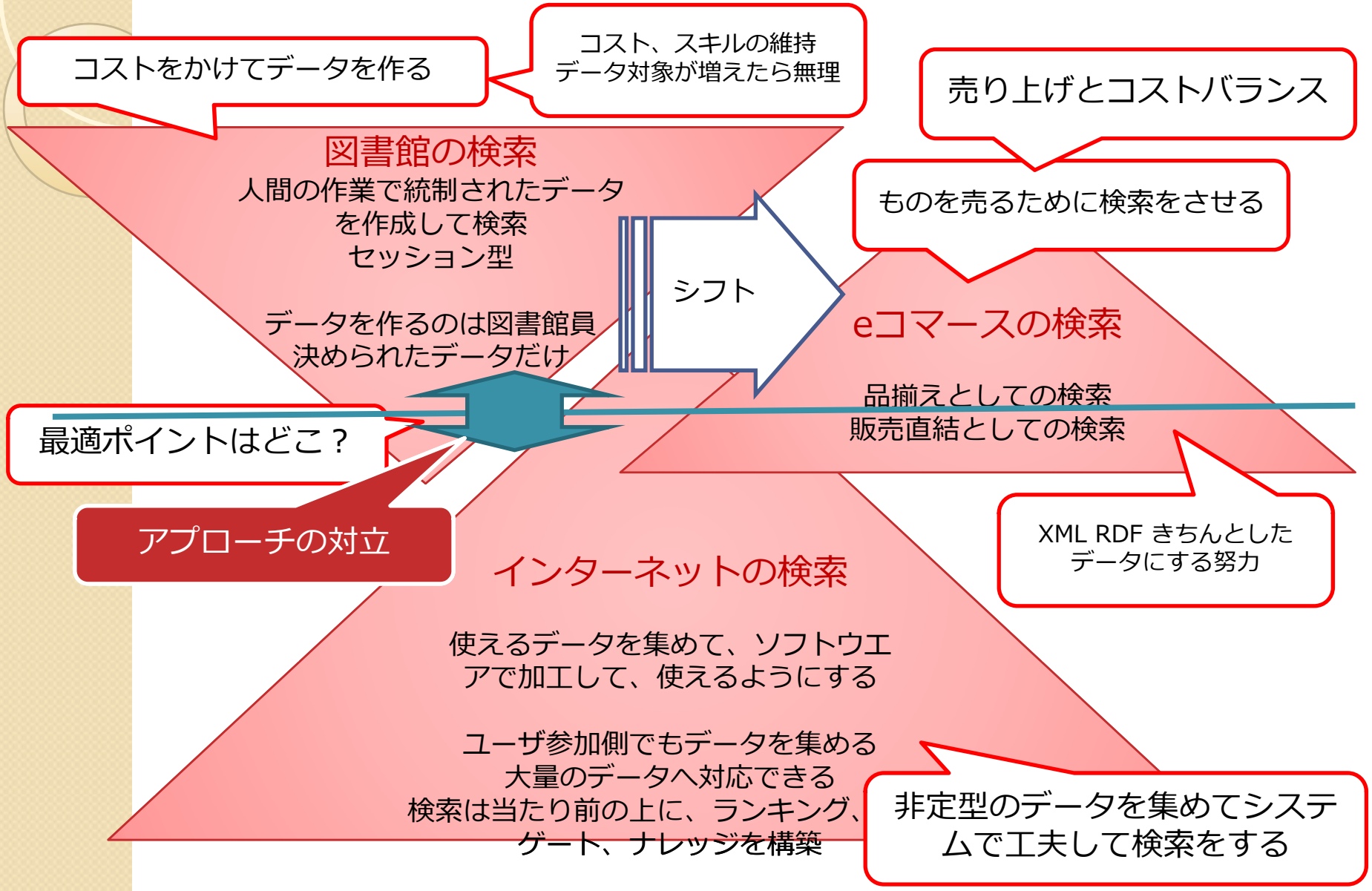
大学図書館としての全文検索への挑戦

- 多様なデータ形式からデータを収集して、きめ細やかで高度な検索を実現する。
- 統合検索、全文検索がインターネットへの架け橋
- MARCは一つのフォーマットであって依存しない
- 基本技術
  - クロール、XML、高速エンジン、言語処理技術、ナビゲーション、アクセスログ解析、表示コントロール、運用コントロール、タギング
- eコマースなどで鍛えられたエンジンがほしい

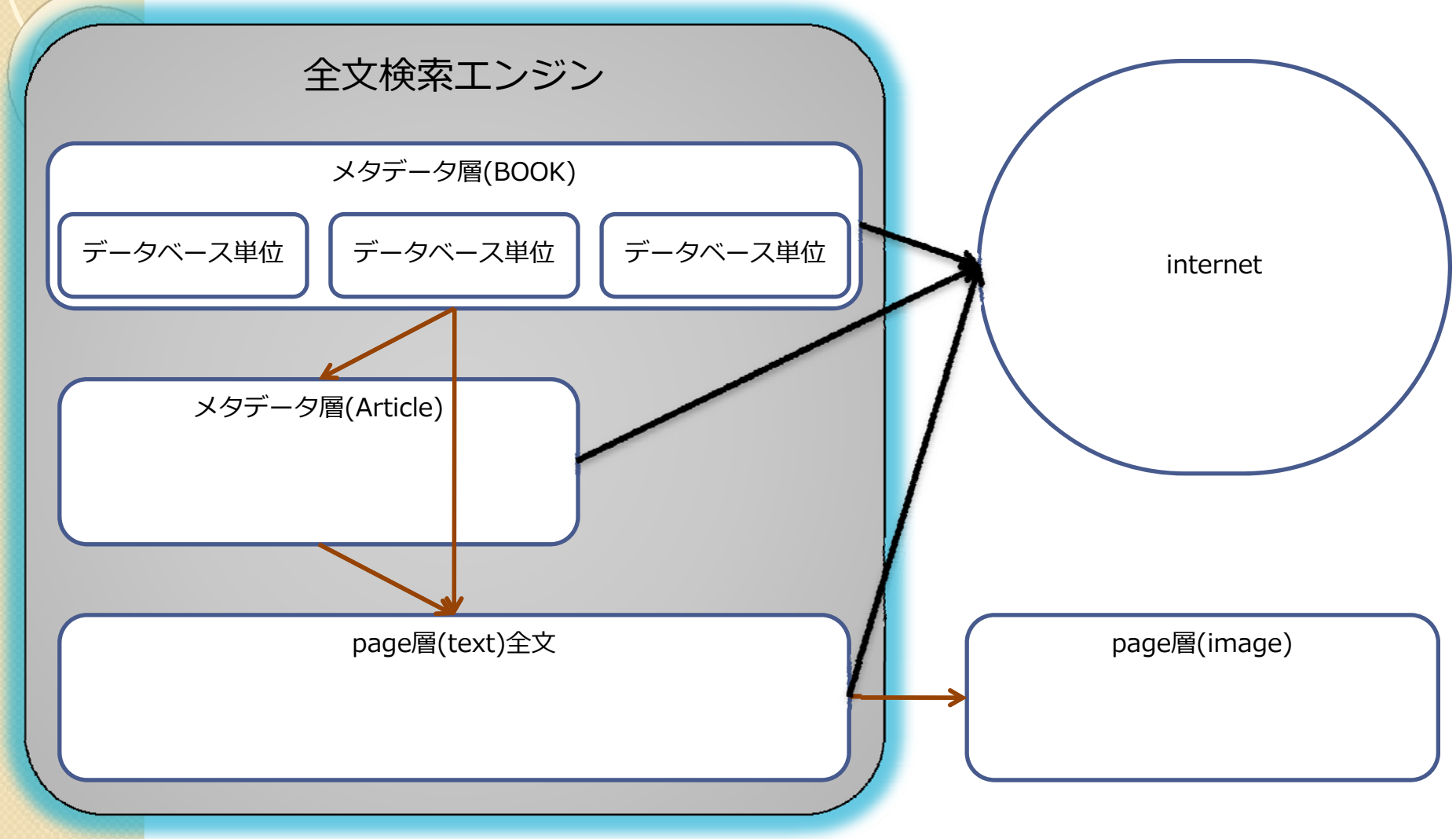
目録データ作成と利用モデルの変化が起きている

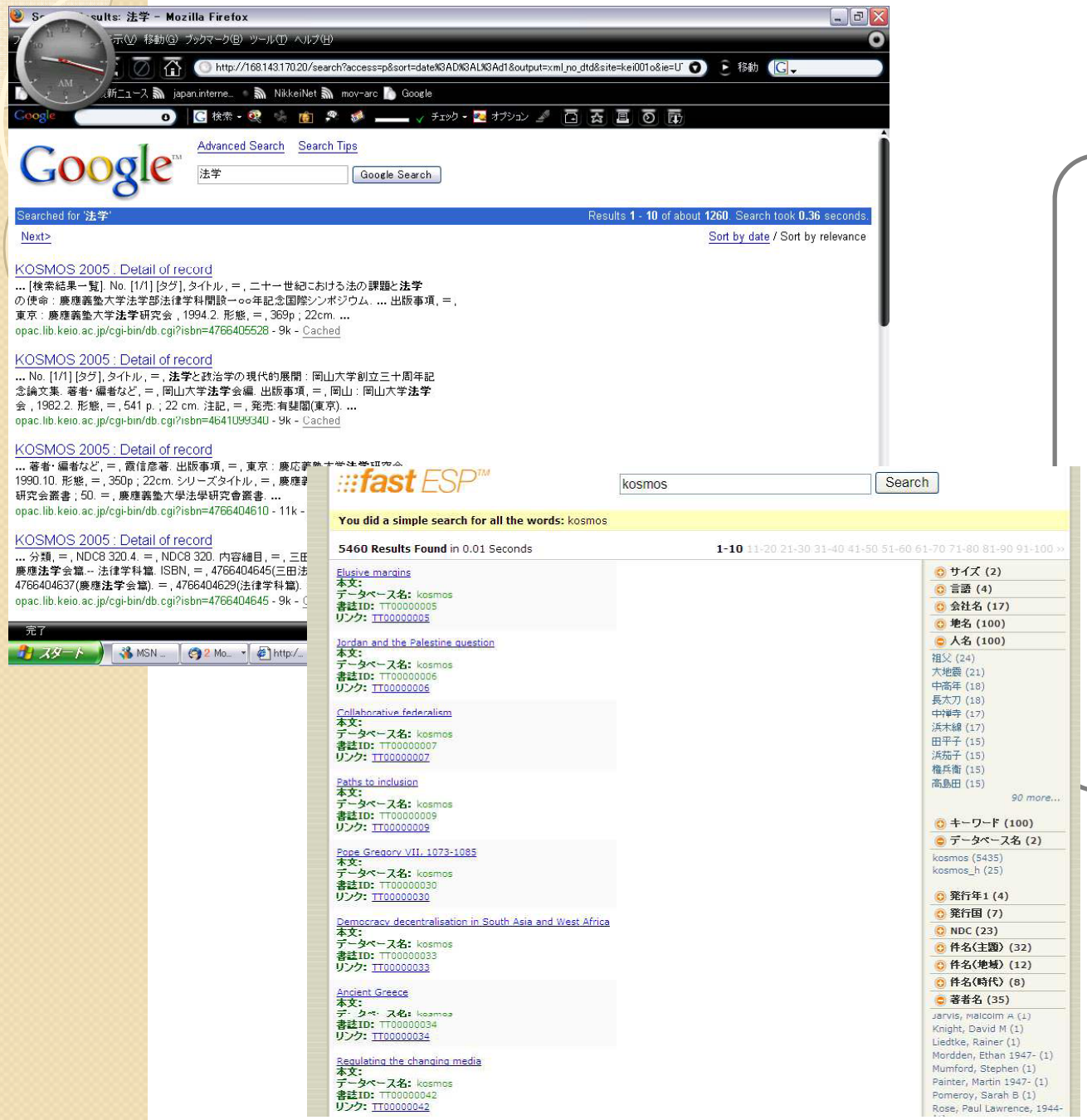
- Shard catalog モデルからData exchange モデルへの移行
- Shard catalog モデル
  - 複数の図書館がオンラインを利用して共同で目録を作成する
  - コミュニティーは図書館が中心
  - オンラインネットワーク、単一データベース技術
- Data exchangeモデル
  - 情報発信者がインターネットでのデータ交換を利用してデータを交換する
  - 目録作成と利用が図書館コミュニティーを超えてきている
  - ユニークな番号をどうするか（ISBN)
  - だれがデータをつくるのか。(情報発信者)
  - データをだれが集めるのか。(コミュニティーできめる)
  - データを集める方法は。(クロール、ハーベスト・・・)
  - データフォーマットは。(流通単位、番号単位、インターネット標準)
  - 図書館はそれにどういう付加価値をつけるか。
  - ポイントは、著者名と件名のコントロールとコミュニティー毎のタギング

図書館 VS インターネット - eコマース



全文検索実験でのデータ構造





ナビゲーション機能
高速検索
表示順番コントロール

メタデータ構造
DB名、コード、件名を前面に

MARCは目録カード印刷用フォーマットです

- MODSフォーマットをベースに検索用として設計(MARCからの脱皮)
  - KOSMOSはMARCからMODSへ変換
  - KOARA デジギャラはMODSで設計され透過です
- データ部分を以下の３つに構成(データの目的を明確化)
  - 検索結果のナビゲーションのために、これまで冷遇されてきた、データベース名、コード、件名を前面に出して利用する(KOSMOS ではコード系が整備されていない)
  - 目録の記述部分（表示部分）を明確にし、それ以外は検索のためのデータとして、インデックス部分(ヨミ　ローマ字　その他)として追い込んでしまう。
- 分かち、ヨミ、ローマ字の明確化
  - 目的を明確にして設計する
  - 分かちは今回の実験では設定していません。
  - ヨミはソートで使用することがあるために主要なタグに設定しました。
  - ローマ字は検索用インデックスと考え、インデックス部分に設定しました。
- 大きな失敗
  - ４４０をタイトルにして２４５と一緒にしてしまったので、シリーズ典拠を作れなかった。

コードは重要になります

漢字分かちは古典的な検索システムでしか必要ないでしょう

物理単位レコードの工夫

<db name="**kosmos**">
<book id="**TT00000005**">
<meta type="**bib**">
<udate>**2008-04-01**</udate>
<reserve>**1**</reserve>
<Type_of_record>**a**</Type_of_record>
<Bibliograhic_level>**m**</Bibliograhic_level>
<date1>**1998**</date1>
<country>**ja**</country>
<langage>**jpn**</langage>
<classification_NDC>**349**</classification_NDC>
<subject_topic>**地方財政**</subject_topic>
<subject_topic>**地方自治**</subject_topic>
<isbn>**4641280029**</isbn>
<line>**地方分権のための地方財政改革：吉田和男著：東京：有斐閣，1998.8：東京：有斐閣，1998.8：viii, 229 p.；19 cm.**</line>
<title>**地方分権のための地方財政改革**</title>
<relatedItem_title>**有斐閣選書；**</relatedItem_title>
<auther>**吉田和男著**</auther>
<name_>**吉田, 和男 (1948-)**</name_>
<origin_dateIssued>**1998.8**</origin_dateIssued>
<origin_placeTerm>**東京：**</origin_placeTerm>
<origin_publisher>**有斐閣，**</origin_publisher>
<phyD_extent>**viii, 229 p.；**</phyD_extent>
<Discription>**補助金制度と地方交付税制度の問題点.カバー；**</Discription>
<sort_title>**チホウ ブンケン ノタメノ チホウ ザイセイ カイカク**</sort_title>
<index_>**ヨシダ, カズオ(1948-)；チホウ ブンケン ノタメノ チホウ ザイセイ カイカク**</index_>
<index_>**ヨシダ カズオ チョ；トウキョウ：ユウヒカク，1998.8；ユウヒカク センショ；517；チホウ ザイセイ；チホウ ジチ；Yoshida, Kazuo,1948-；Chiho bunken notameno chiho zaisei kaikaku**</index_>
<index_>**Yoshida Kazuo cho；Tokyo：Yuhikaku，1998.8；Yuhikaku sensho；517；Chiho zaisei；Chiho jichi；**</index_>
</meta>

Type_of_record Bibliograhic_level
date1 date2 country langage title_uniform_h
phyDescription_marccategory phyDescription_marcsmd
classification_NDC subject_authority_topic
subject_authority_geographic subject_authority_genre
subject_authority_temporal

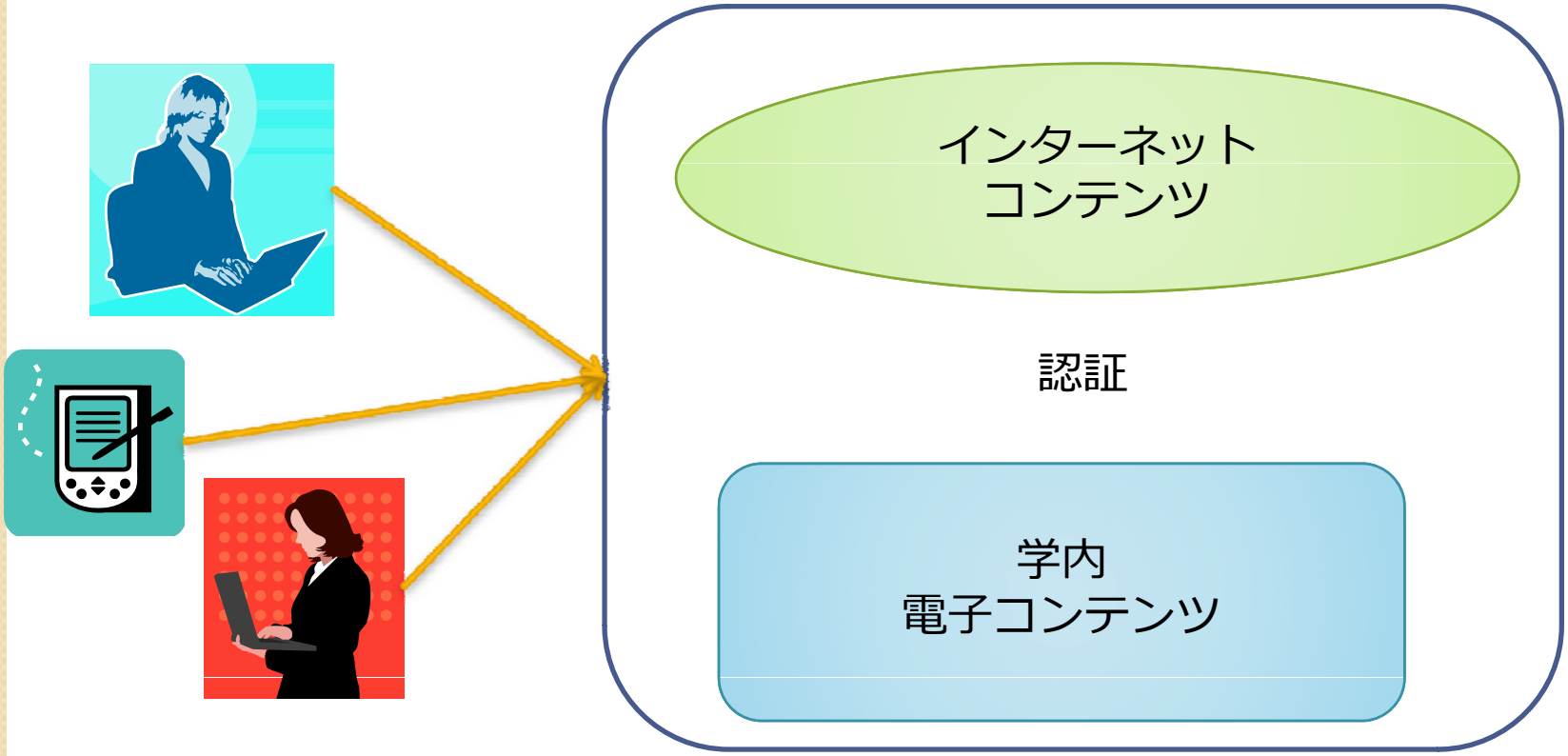
ナビゲーション用　これが重要
DB名、媒体、資料形態、年、国、言語
時間、空間、著者、分類、サブジェクト

表示用

インデックス、
ソート用

流通を促進するための著作権への対応

- 著作権を認証と契約で制限できないか？　そもそも学術コンテンツは流通したい
- 認証を使ったアクセス空間とアクセス管理とコンテンツの認証により信頼性確保
- 公開リスクの回避とフェアーユース条件の確保
- パグリックドメインと認証コンテンツの区分け
- 無料コンテンツと有料コンテンツの区分け
- 学術コンテンツの活性化



貧困なデジタルデータ保存
私の机の周りの30テラ

NAS

eSata

Blu-ray

DVD

Tape

UDO

50T Sataの大容量ストレージ

図書館が扱うデータ量は増加する
400万冊の目録データ　　4G
10万冊の全文テキスト　　80G
10万冊の全ページ画像　30,000G
書庫問題からストレージ問題へ

・作業フローに合わせた複数のバックアップ
・撮影した人、加工した人が作業の区切りでDVDへ複写
　DVD のバックアップでは限界(1T が 250枚)
・システム的なバックアップを併用
　HD to HD , HD to Tape

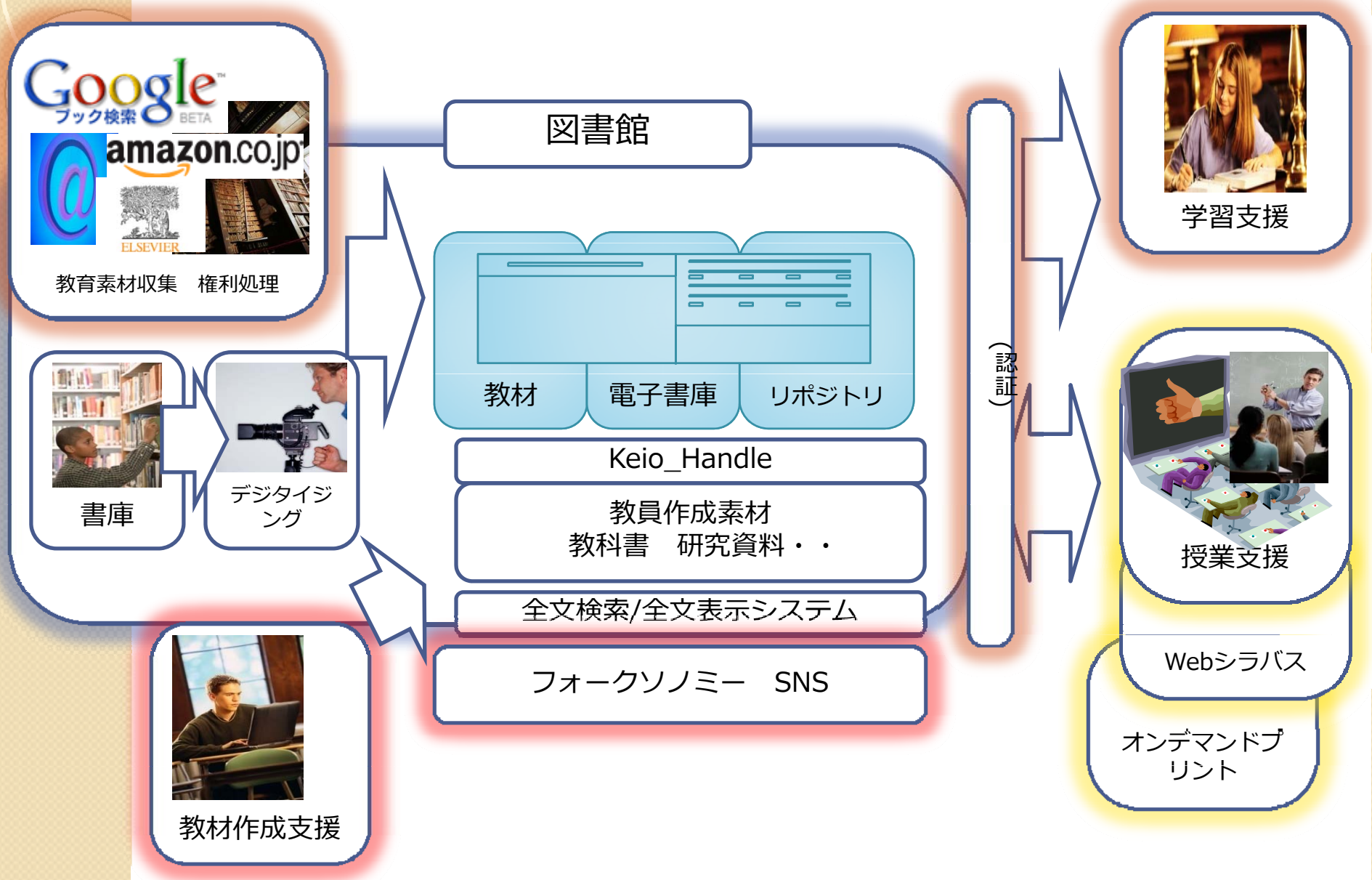
・磁気テープだとバックアップ後の中身を見るのが大変

・画像保存はTIFFやめてraw　1枚の画像が 200M->50M
・ハイビジョン画像からの撤退

・デジタルデータはマイグレーションが前提となる
・問題はマイグレーションの間隔とコスト
・DVDは 3年でマイグレーションが必要 DVDをどこに置くか
・ソフトウエア、ドライブ、媒体の寿命を考えて１０年に一度にできないか？　10年もつ媒体で最適コストは？

まとめ

日本の大学図書館巻き返し戦略？

- Library standard から Internet standard へ
  - Library standard を持っていない
  - Library の技術力の低下
- 高度なIT技術の活用
- 大学の教育研究ニーズへ直結し、認証を前提としたアクセス空間へ高品質なコンテンツをサービスモデル（電子図書館）をつくる

電子図書館イメージ
デジタル資料を使った図書館の学習・教育活動支援