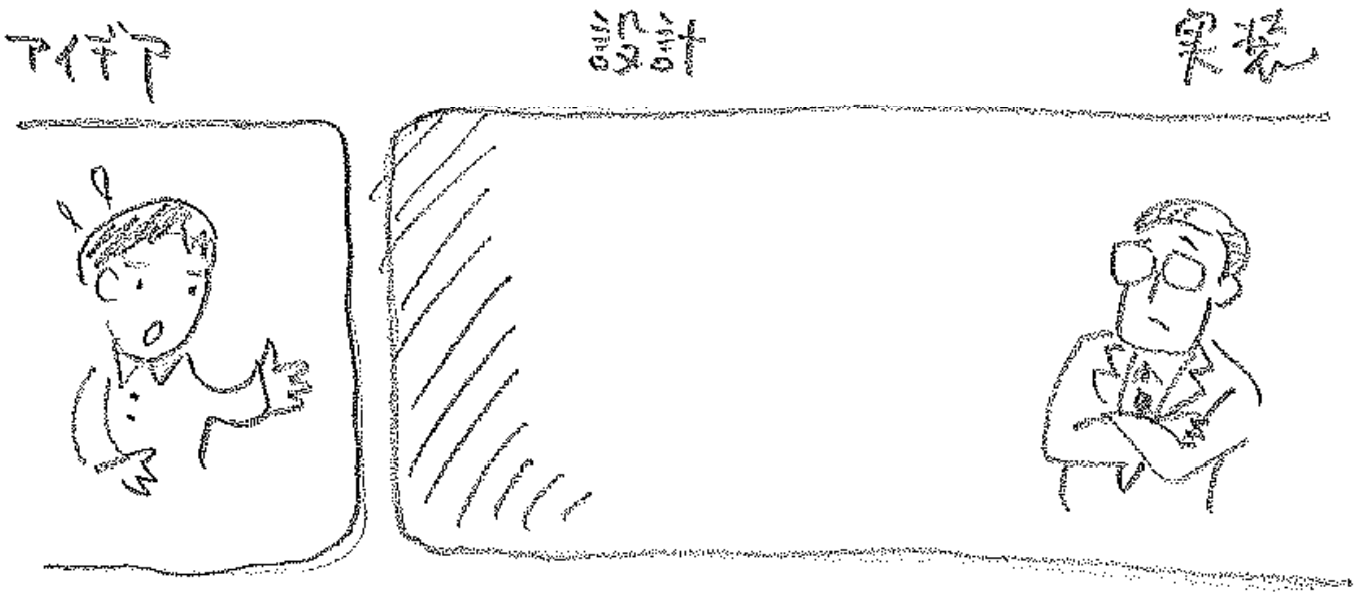
講義

流通する学術情報コンテンツ

山本哲也
名古屋大学　情報連携統括本部情報推進部

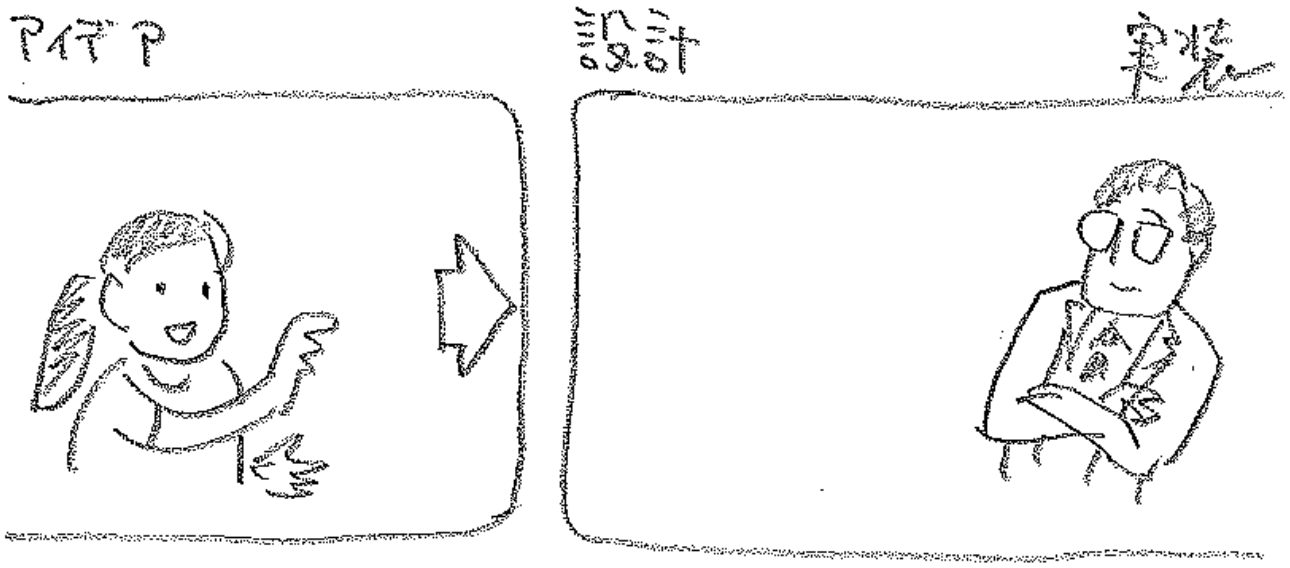
平成23年度学術ポータル担当者研修
2011.8.3（名古屋大学）
2011.8.24（NII）

今より少し技術寄りな言葉で
やりたいことが表現できるようになりたい



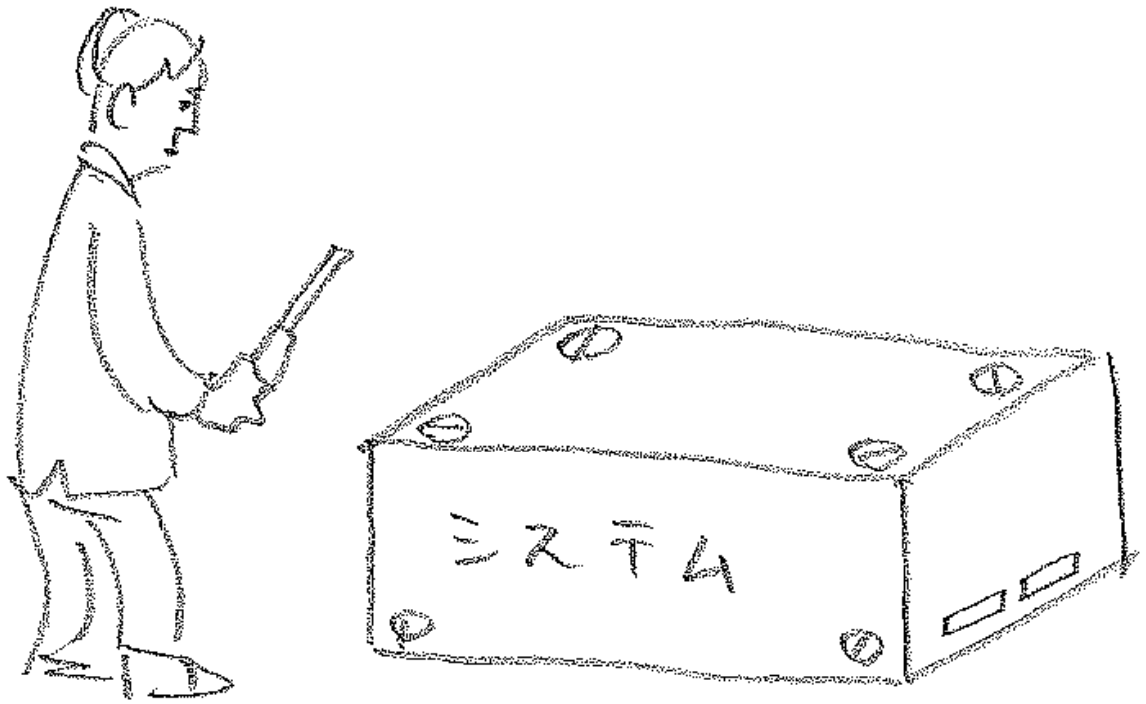
「何々な感じのサービスをつくりたい。
詳細は、よしなに…」

コントロールを（ちょっとでも）取り戻す

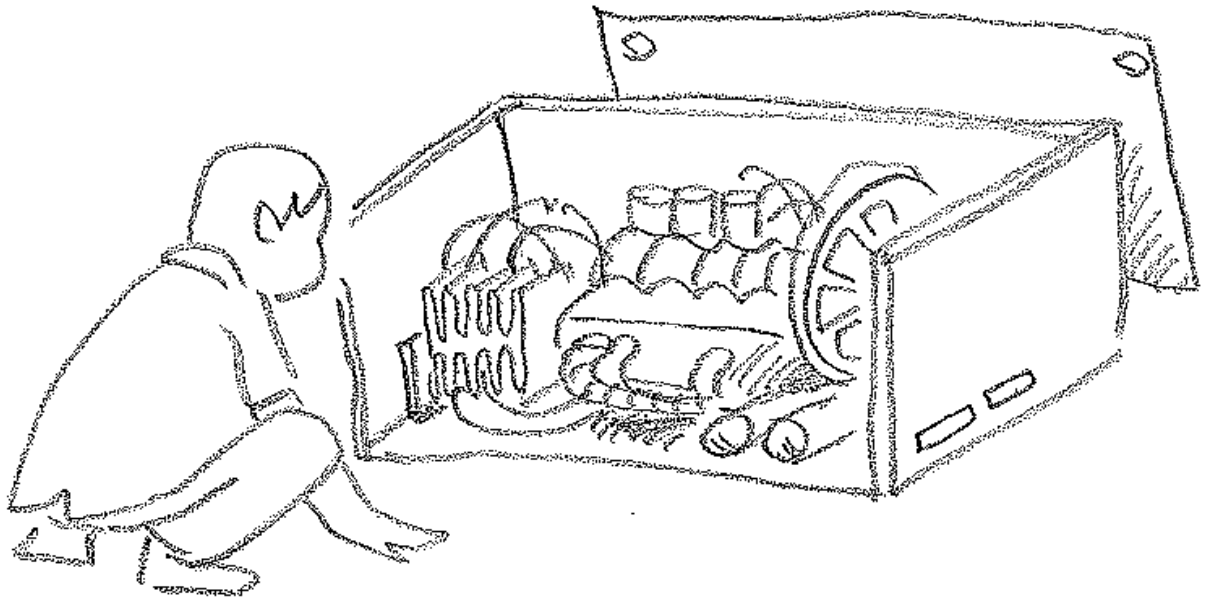


「何々なサービスをつくりたいので、
〇〇を使って〇〇を扱えるようにしたい」
※知ったかぶりは危険だけど

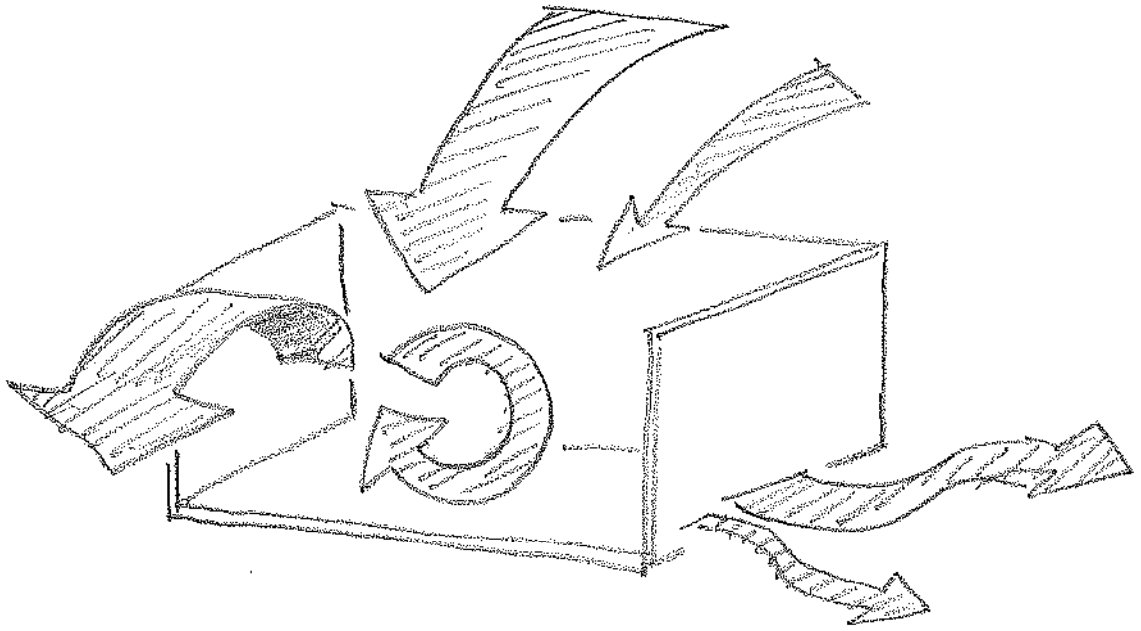
じゃあ、手元にある適当なシステムを
よく調べてみようか



「・・・」



さしあたりメカ的な詳細ははぶいて、主な
「データ」がどうなっているのかという視線のみに
切り替えてみる



「コンテンツ」と「データ」

ここではほぼ同じような意味に扱います。

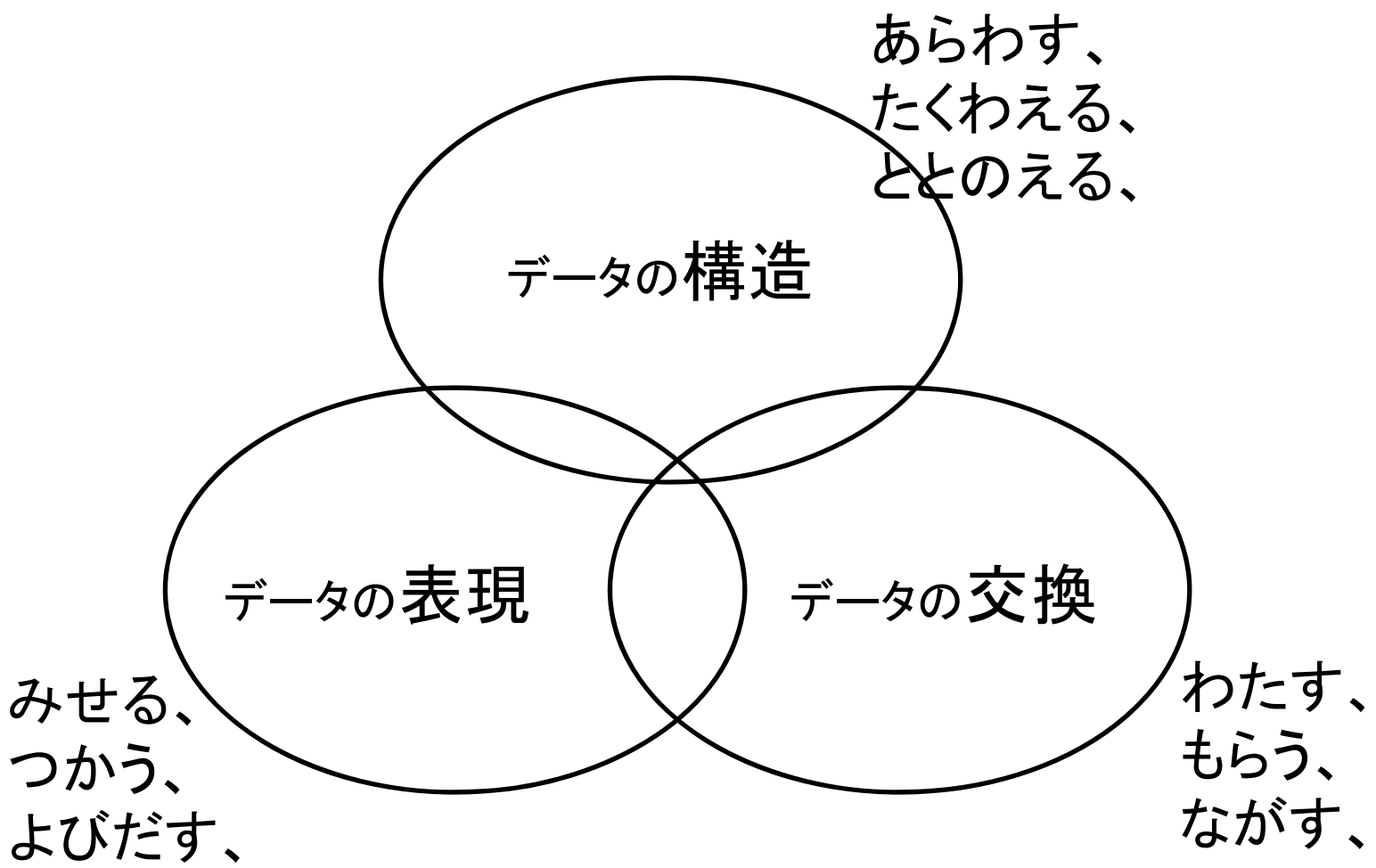
コンテンツが、実際に利用して価値のあるモノ、
それにくっつくデータ（メタデータ）が、コンテンツ
の効果的な利用を実現するためのもの、と
いった傾向（でいいですか？）

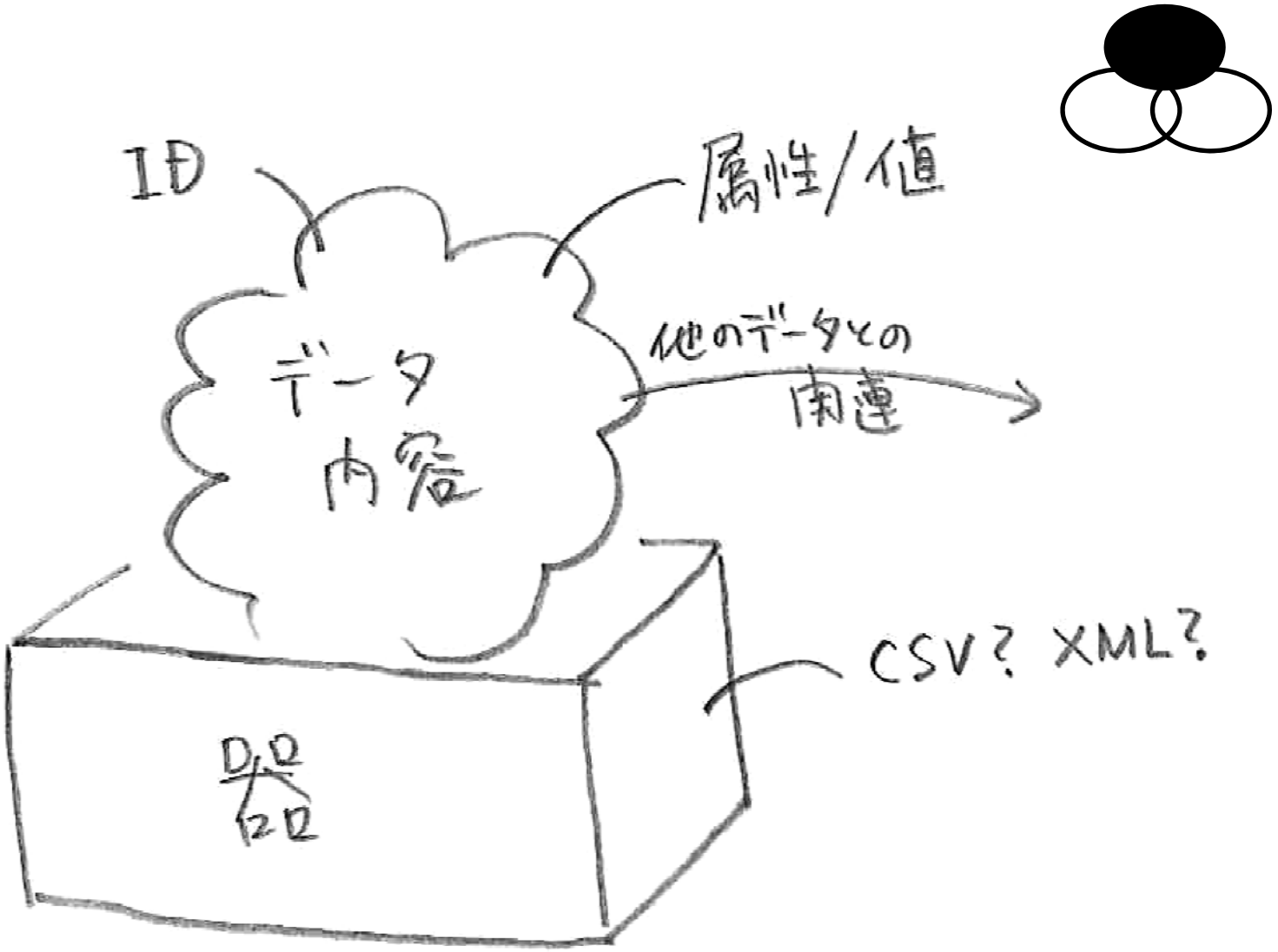
データの扱いに関わる技術用語を思いつくままに挙げてみるだけでも、容易に手に負えません

CSV, DTD, Dublin Core, HANDLE/DOI, HTTP, java, junii2, METS, Nグラム, OAI-PMH, OpenURL, perl, RDB, REST, RSS/ATOM, rsync, ruby, SOAP, SQL, TCP/IP, UNICODE, URI, URL, WEB-API, WEBクローラ, XML, XMLスキーマ, XMLスタイルシート, Z39.50, エンコーディング, クロスウォーク, サーバーサイド/クライアントサイド, スクレイピング, パーマリンク, ファイルシステム, 正規表現, 転置インデックス, 漢字統合インデックス, 文字セット, …

整理する手がかりを作ってみました→ → →

「データ」を理解する三つの側面




ID
属性/値
データ
内容
他のデータとの
関連
CSV? XML?
DB

データそのものと、データの「器」を峻別
できること。

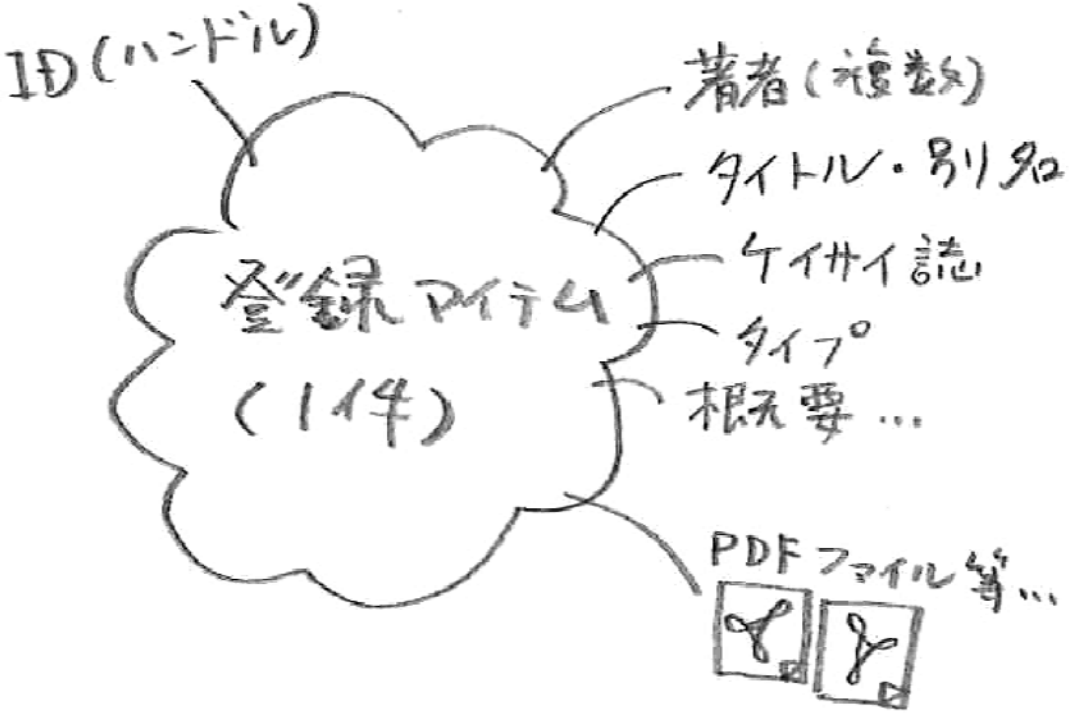
データ「一件分」にどのような属性があって、
それらがどういう値を持つか。値は単数か
複数か。これによって、適した「器」は異なる。

データ「一件分」を、用途にあわせてどう
デザインするかは、経験のいる仕事。
全部でどんな種類の「データ」を使うかも、同様。

どんな属性を標準的に揃えておいてくれ、という
要請が、junii2などのメタデータルール。このルール
は、データの「交換」において威力を発揮する。

他のデータへのつながりを表現しておくのも、
その用途を豊かにするため重要。
「データ」の固有の識別子は何か、と考えると、
けっこう難しい。Webを視野に入れるなら、URI
(URL)で記述できるか考える。後述する、
パーマリンクという考え方が重要であることも
同時にわかるはず。

（例）機関リポジトリが一般的に使う、論文一件分の「データ」内容はどうなっているか

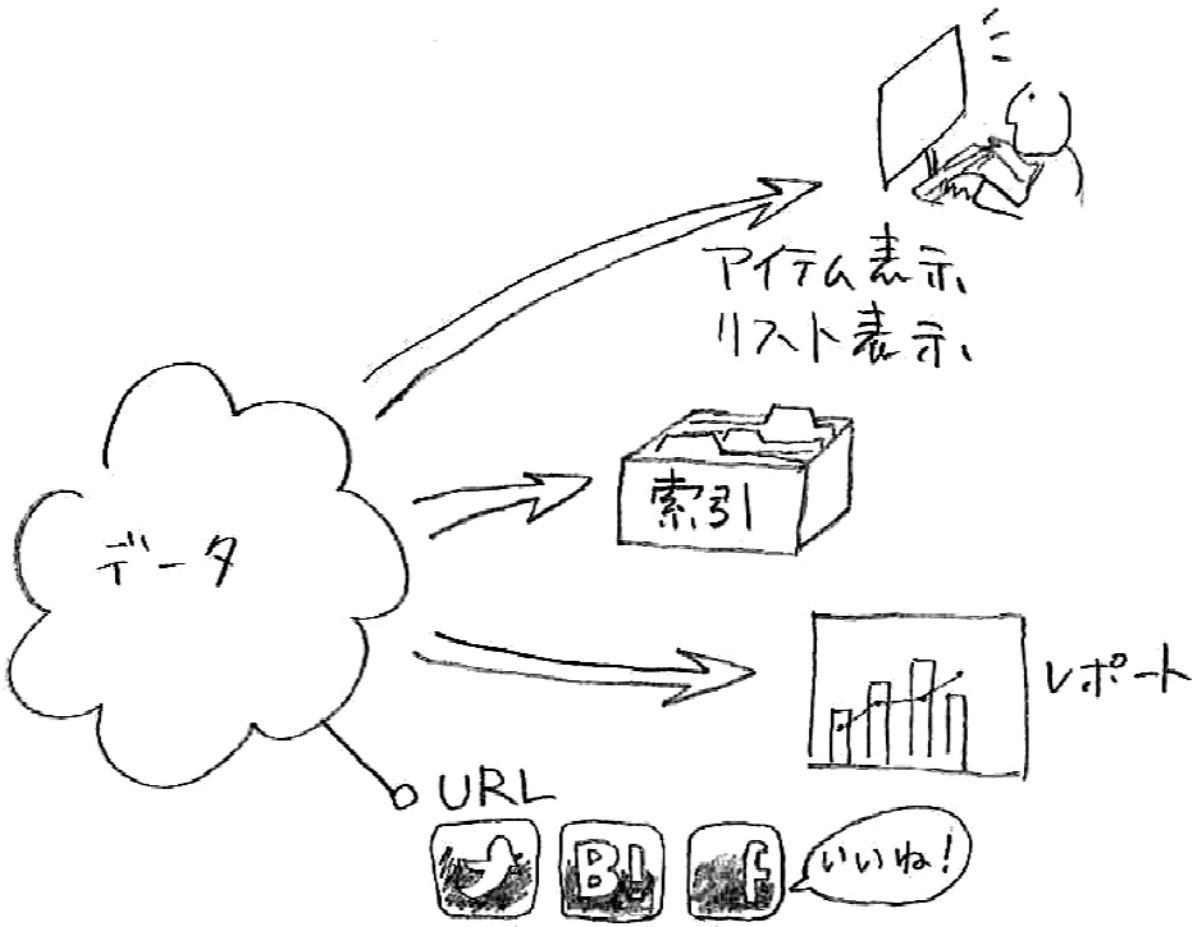


これを内部的にはPostgreSQLに格納している、
OAI-PMH上ではXMLで記述されている、というのは
「器」レベルの話。

ひとつの「データ」が色々な表現形式に変化する。

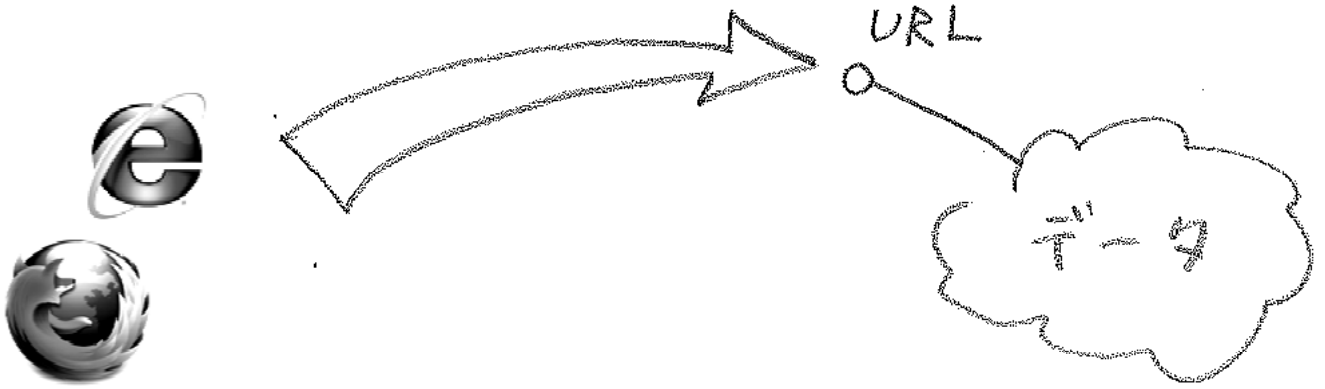
新しい「データ」を持ち出さなくても、新しい表現形式に落とすことはできる。

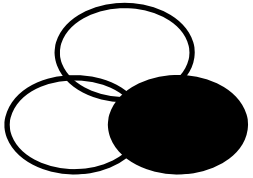
データのどの属性をどんなルールで整形すれば実現できるか、説明できれば開発はできる。



Webにおいては、データの表現はブラウザ上で
（特定のURLのもとで）行われることが多い。
このURLがいつも一定に保たれることを保証
することも、データを「表現」する上で非常に重要。
端的には、ブックマークできるかどうか。

※特定アイテムへのブックマークができないWeb
コンテンツは、意識して探してみると結構あるものです。





システムから他のシステムにデータを
渡す場面：単純なのも、複雑なのも

「定期的にA学部から受け取るExcelファイルを
検索システムに投入してるんだー」

「システム内の全部のメタデータを誰でも再利用
できるように、標準的なDublinCore属性セットを
持ったデータをOAI-PMHで持ってけるように
してるんだー」

どっちも、立派な「データの交換」です

JAIROに（OAIStarに）メタデータをまとめて渡せば、
統合検索をやってくれる！

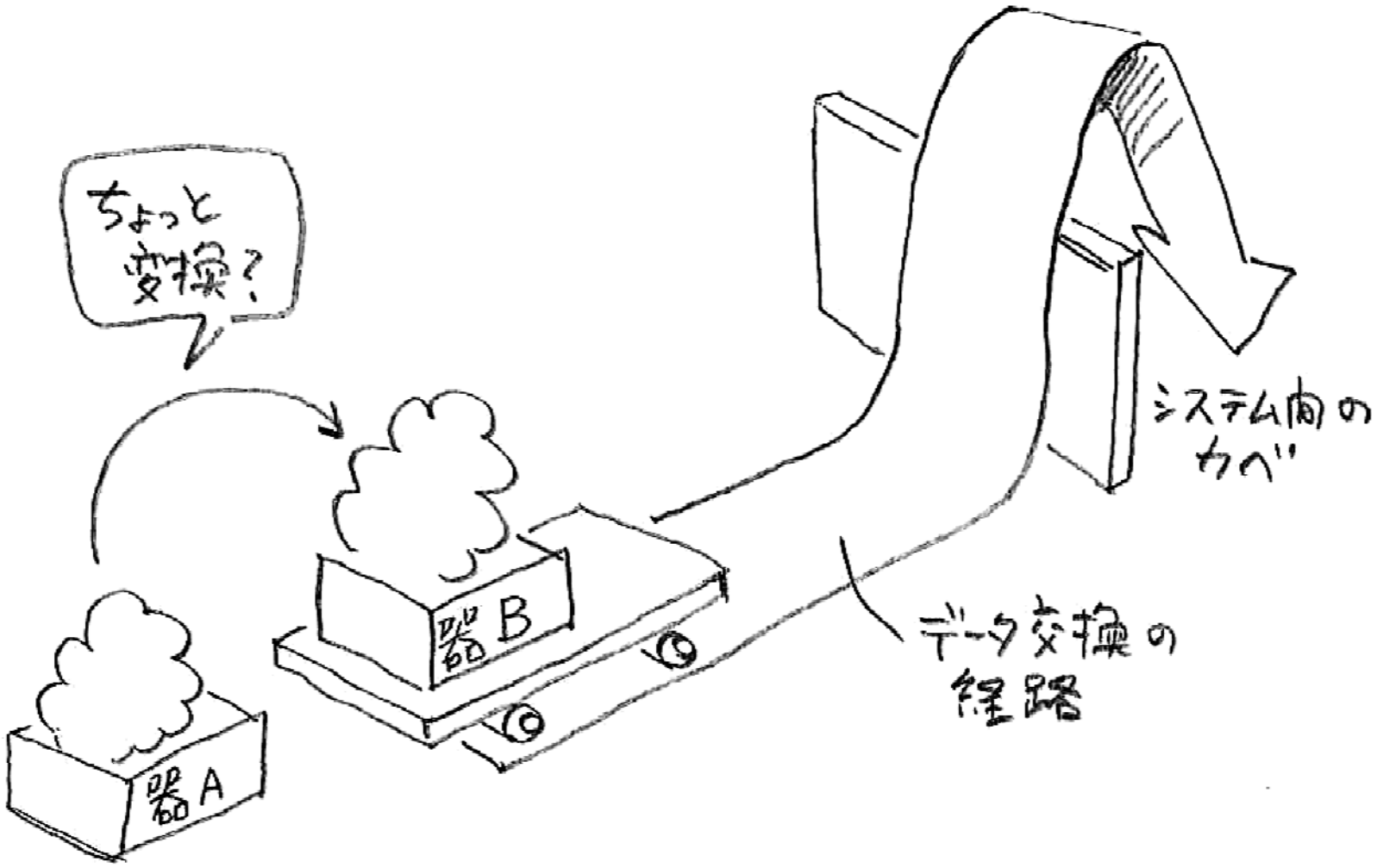
RSSリーダー(Google Readerとか）にフィードを
渡せば、Webサイトの最新情報を楽にチェック
してもらえる！

反対に、よそから取得できるデータをこちらで
活用できれば、新しい価値にならないか？
「よそから」じゃなくても、自機関に二つ以上の
「つながりそうなのに、つながってない」システムが
あるなら…

データの交換は、「器の乗り換え」

データの内容そのものは変わらないかも知れないが、たとえば内部データベースという「器」からXMLという「器」に盛りなおされて、交換のための経路上を移動する。
いや、移動先がデータ内容に注文をつけている場合は、乗り換えのタイミングでデータ内容の変換も行っておく必要があるかもしれない。

経路：OAI-PMH、RSS、WEB-API、rsync、フロッピ～、…

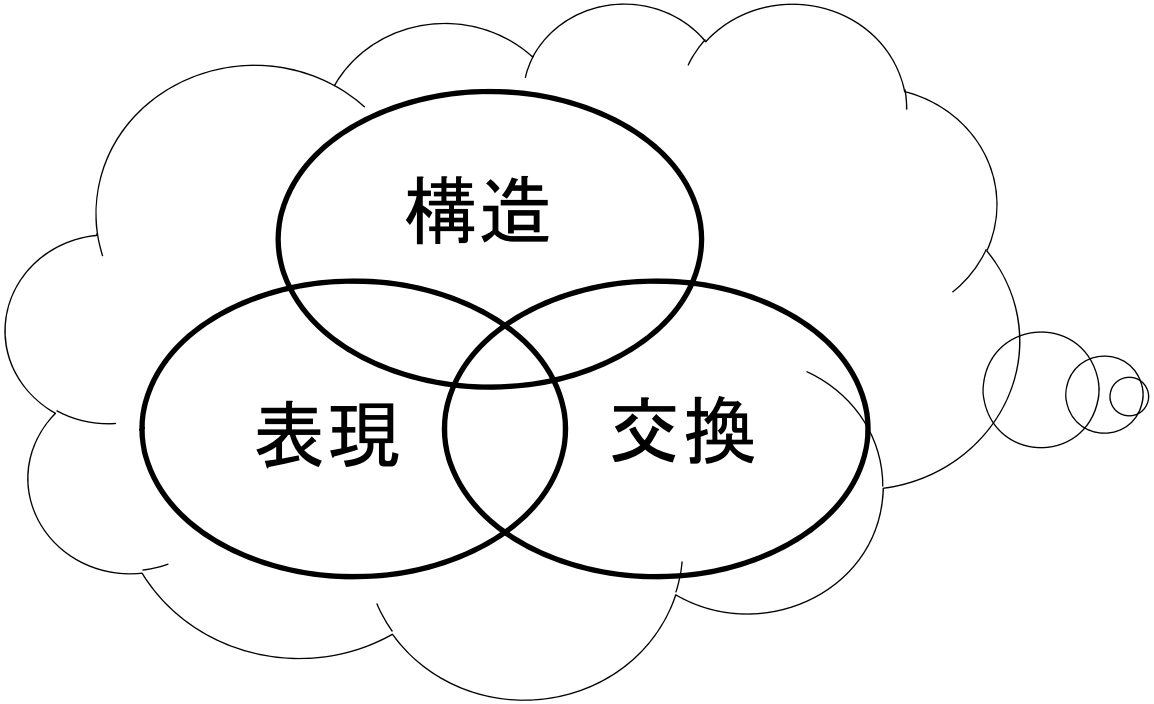

ちょっと
交換？
器A
器B
システム向の
カベ
データ交換の
経路

プログラミングを皆ができるようになる必要は
ないかもしれないけど、もしも、ちょっとできる人が
いるとして、この種のデータ変換っていうのは
割と取り組みやすい分野ですよ。
ー昔前とくらべて、格段に強力なことが、簡単に
できるようになってきています。

データ変換ってのはほぼ例外なくカスタムメイド
な仕様で行うものですし、素早い対応も求められ
がちですし。

「システム糊づけ職人」になってみたくない？

（まとめ）
データ（コンテンツ）流通に関わる技術的知識を
整理しておくための枠組みに



【参考】ここで扱わなかった（主な）もの

〇セキュリティに関わる要素
　SSL、サーバー証明書、統合認証、Cookie…

〇パーソナライズに関わる要素
　ポータルサイト、ソーシャルサービス…

〇コンテンツそのものになるデータの
　処理ノウハウ（PDF、画像データ、音声・動画）