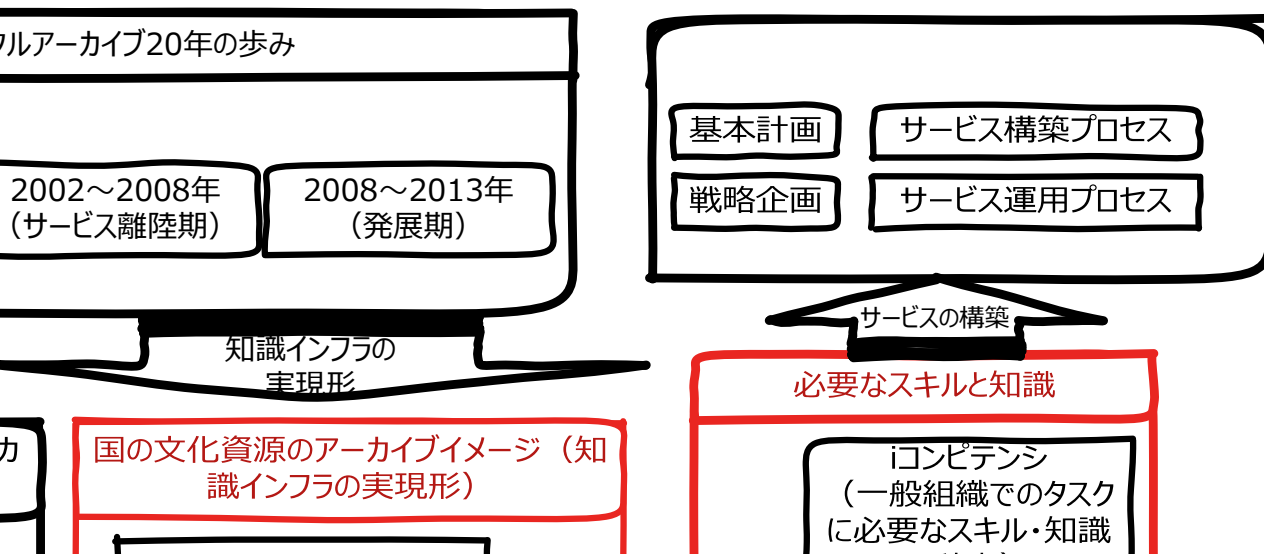


作成者
元国立国会図書館 専門調査員 中山正樹

ナショナルアーカイブの方向性と、それを担う人材の育成



電子図書館20年の歩み

NDLは1992年、21世紀初頭に
関西学術文化研究都市の一角に
関西館を設置するために、具体的
な構想を取りまとめました。
関西館の予定する機能が、電子図
書館的な機能であったこと、また国
の産業構造審議会情報産業部会
が公共部門の情報化を積極的に
進めるべきとの提案を行った

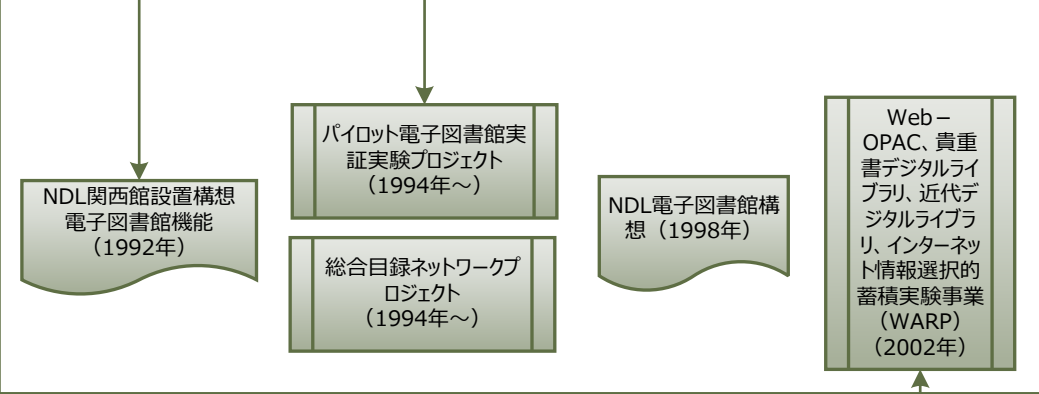
21世紀の高度情報社会において、
地球規模の知的財産を誰でも容易
に利用できるように、地球上に広く分
散して個々に収集・蓄積されている知
的資源を、空間的・時間的制約を越
えてアクセス可能とする環境を提供す
るための実証実験

「国のデジタルアーカイブ構想」、
「ジャパン・ウェブ・アーカイブ構
想」の実現を、「国立デジタル・
アーカイブ・ポータル構想」を一
層推進すること

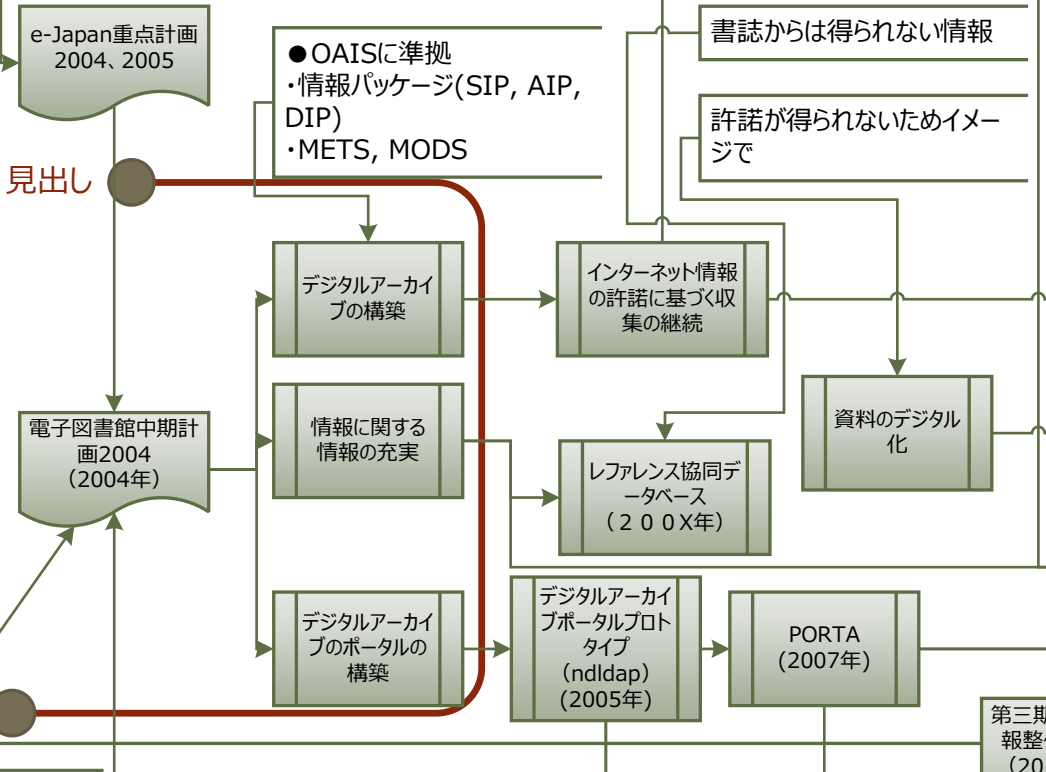
日本中のウェブページを悉皆的に
収集し、将来にわたって利用を保
証すること

- 情報のグラフ化
- Wikipedia、件名典拠・人名
典拠との関係づけ
-

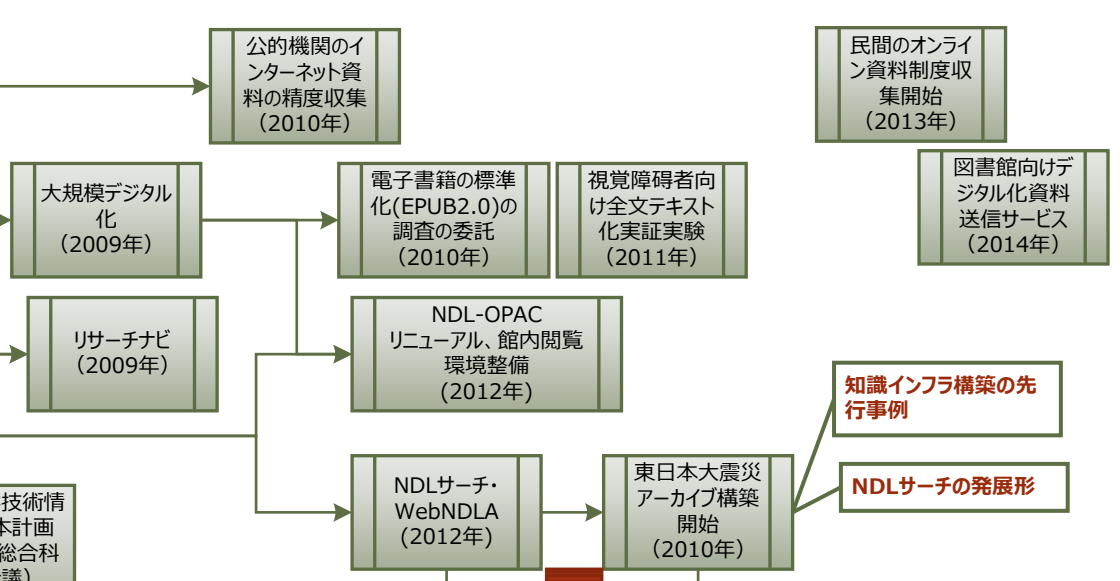
1994～2002年（揺籃期・始動期）



2002～2008年（サービス離陸期）



2008～2013年（発展期）



実証実験の成果を踏まえて

デジタルコンテンツを広汎な利用者に提供
するために、当館が国のデジタルアーカイブ
の重要な拠点となるということ、また国内外
の多様な利用者層の需要に応じ、日本の
デジタル情報全体へのナビゲーションを行う
総合サイトを構築し、利用者がワンストップ
で利用できるようにする

「個別図書館サービスの横
断的利用が可能になるよ
うなサービスの提供を目指
す」、「同じ分野、同じ利
用者層をターゲットにした
複数の専門情報サイトが
連携して、利用者がワン
ストップで利用できるように
する」

- SOA指向、OSS (Lin
ux, Apache, Xoops,
Wiki, Blog, RSS,
MySQL, PHP, Dspace,
chasen, GETA等)
- CMS利用を推奨
メタデータの自動付与、内
容の構造化
- 連携レベルの想定
・SQLレベル
・OAI-PMHレベル
・Webページレベル
・Webサービス (SOAP)
- 標準プロトコル
OAI-PMH, RSS, SRU,
SRW等) の実装

- 拡張容易性、障害時
運用継続性、環境変更
容易性、直感的操作性の
確保 (ベンダーに依存し
ないパッケージ、OSSの適用)
- 可能な限り、先進技術
の適用を目指す。(VMw
areによる仮想サーバ環
境の適用)
- 永続的識別子
■メタデータ記述要素・記
述規則 (シンタックス、セマ
ンティックの共通化)
標準メタデータとして、DC
ベースのDCNDL+α (DC
NDL_Porta) を適用

「知識インフラ」とは、情報資源
を統合して検索、抽出するこ
とが可能な基盤で、国内の各機
関が保有する情報を意味的に
関連づけて知識として集約し、
新たな知識の創造を促進し、
知識の集積・流通・活用と創
造のサイクル構築を目指すもの

「文献から研究データまで
の学術情報全体を統合し
て検索・抽出が可能なシ
ステム (「知識インフラ」)
の展開を図る」

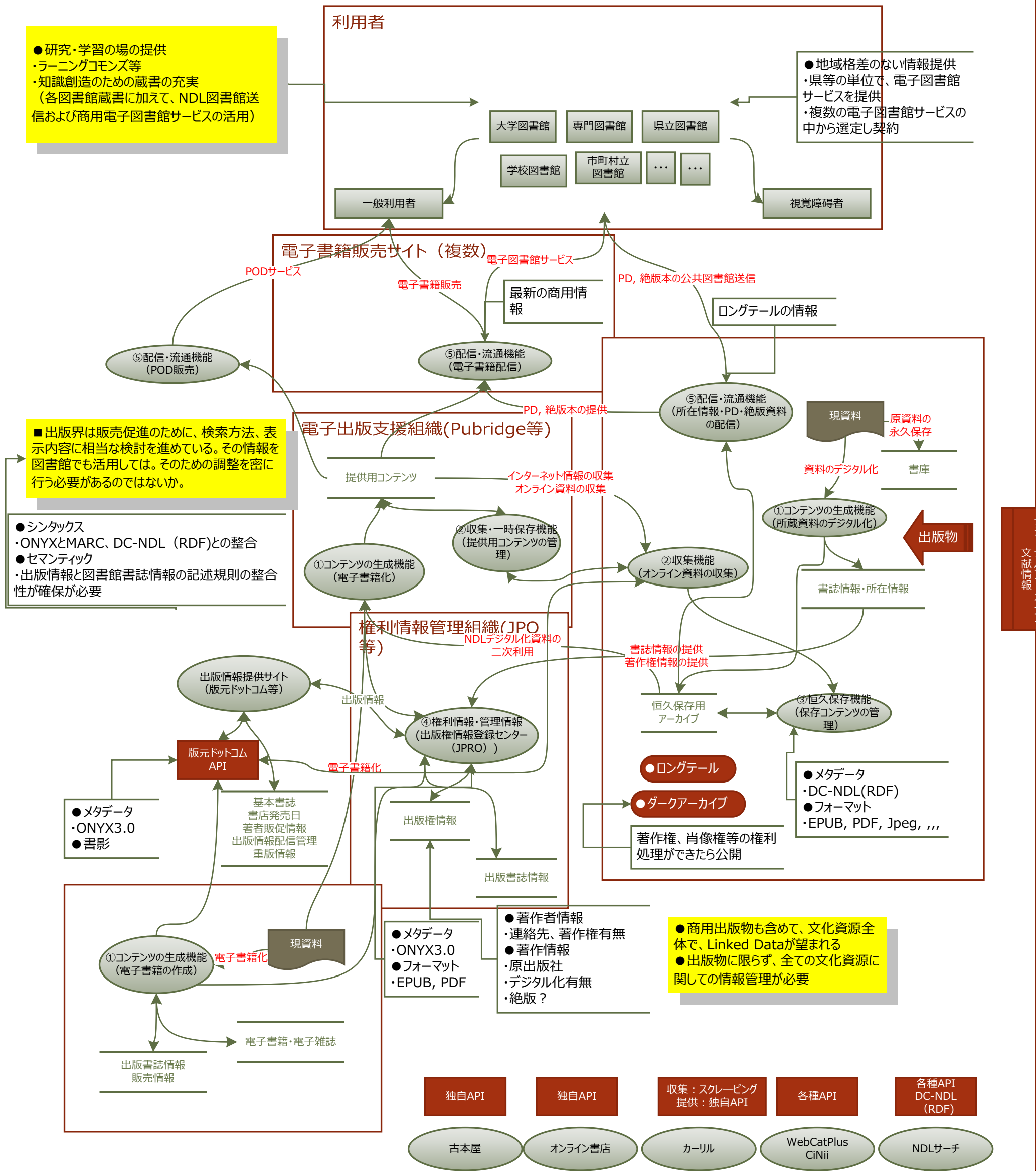
- 言語差異の吸収
日中韓英翻訳機能
- 外部API、メタデータ仕
様の詳細仕様公開
・WebNDLAのSPARQL
公開

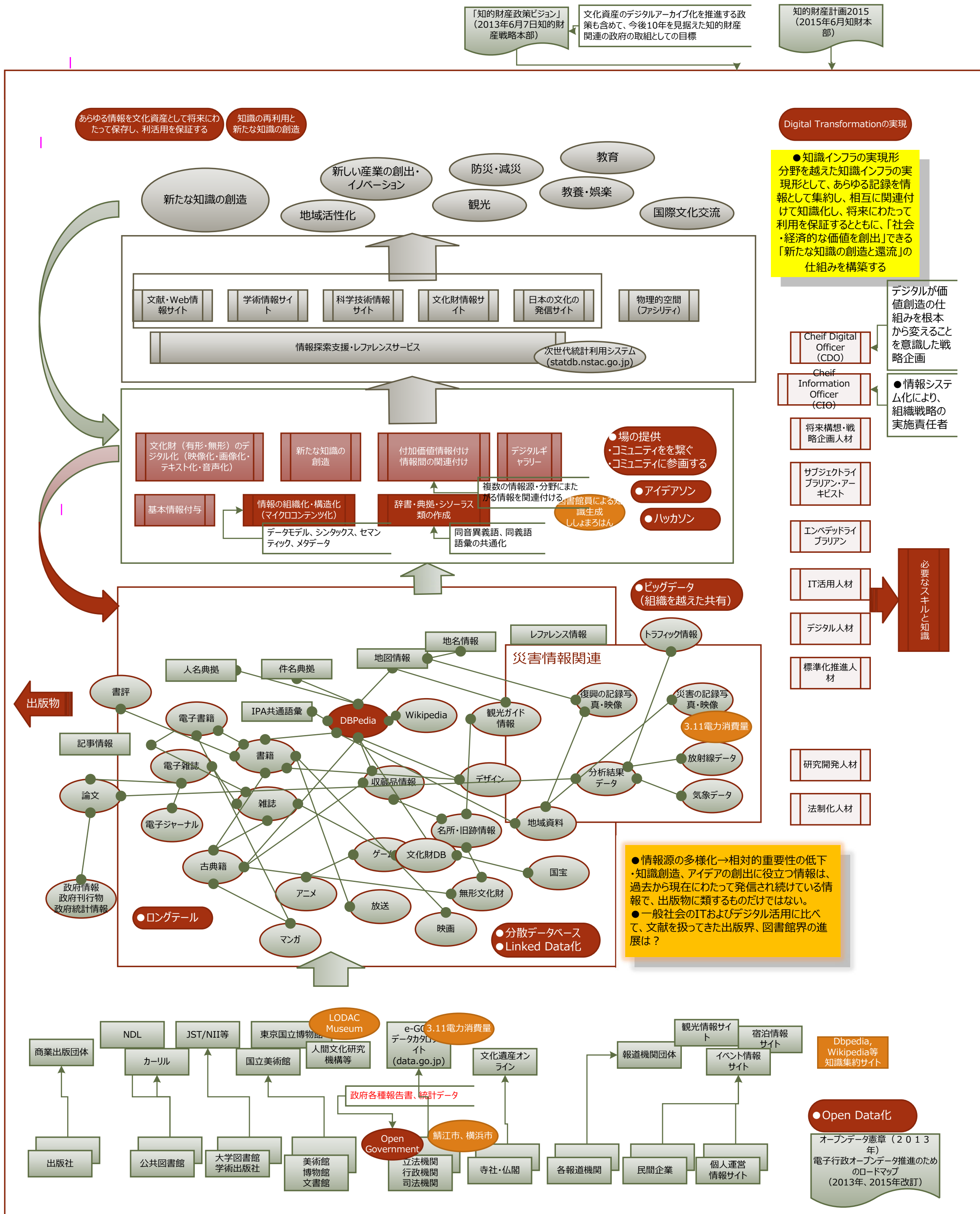
- 大震災に関連する災害現象その
もの、災害前・災害直後・復興の過
程、災害時の対応、他地域・次世
代への教訓等のあらゆる記録を後世
に残す
- 今後の防災に生かすため、関係府
省、各種震災関連情報の保有機関
と協力して分担収集・保存し、一元
的に検索・閲覧できるように

- 媒体、形式を問わず
- 分散処理システム(Hadoop,
GlusterFS)、分散ファイルシステム

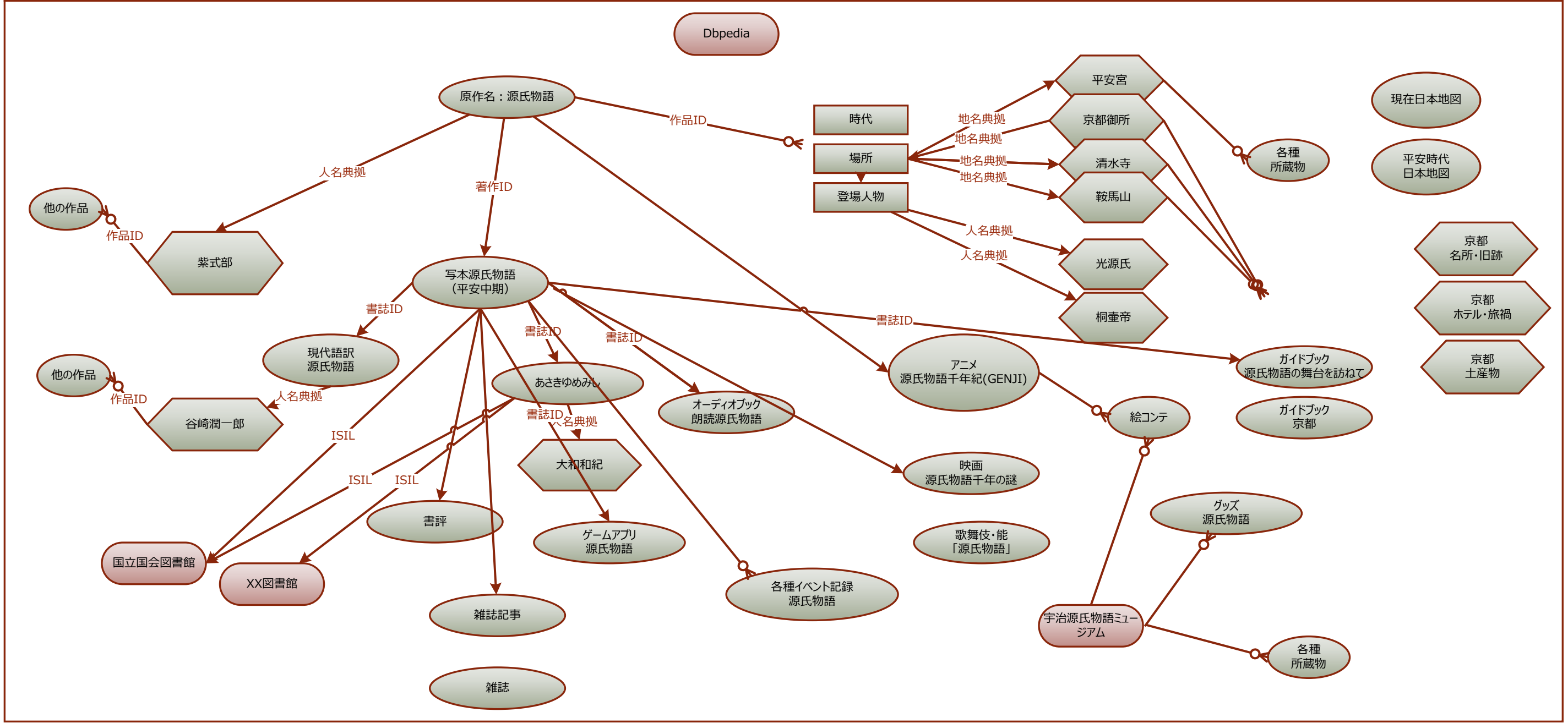
文化財のナショナルアーカイブ

電子書籍のナショナルアーカイブイメージ (DFD)





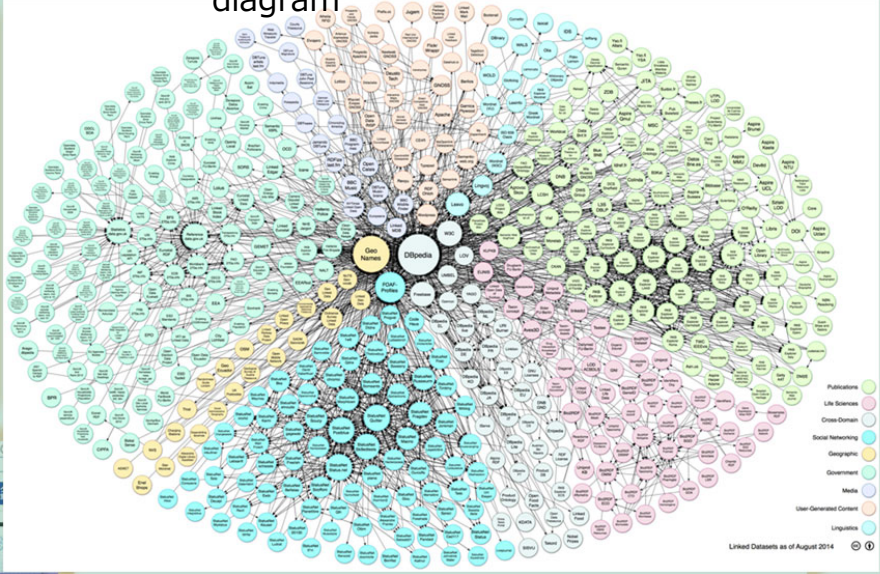
関係付けるべきインスタンス





The Linking Open cloud diagram

<http://lod-cloud.net/>



About: 源氏物語

An Entity of Type: Thing, from Named Graph: <http://ja.dbpedia.org>, within Data Space: ja.dbpedia.org

『源氏物語』（げんじものがたり）は、平安時代中期に成立した日本の長編物語、小説である。文獻初出は長保3年（1001年）で、このころには相当な部分までが成立していたと思われる。

Raw Data -

CSV
N-Triples
N3 (Turtle)
RDF/JSON
RDF/XML
Atom (OData)
JSON (OData)
JSON (Microdata)
HTML (Microdata)
JSON-LD

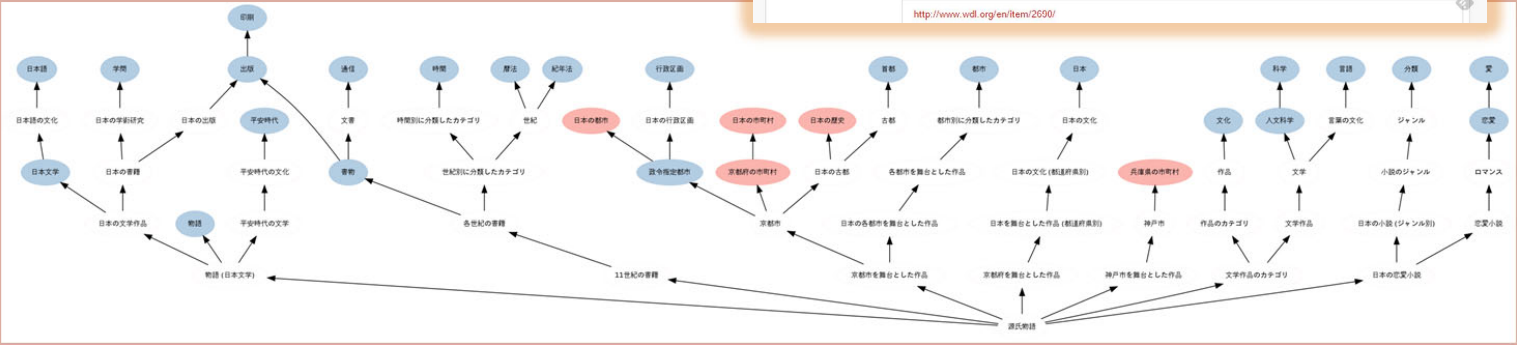
Value

『源氏物語』（げんじものがたり）は、平安時代中期に成立した日本の長編物語、小説である。文獻初出は長保3年（1001年）で、このころには相当な部分までが成立していたと思われる。

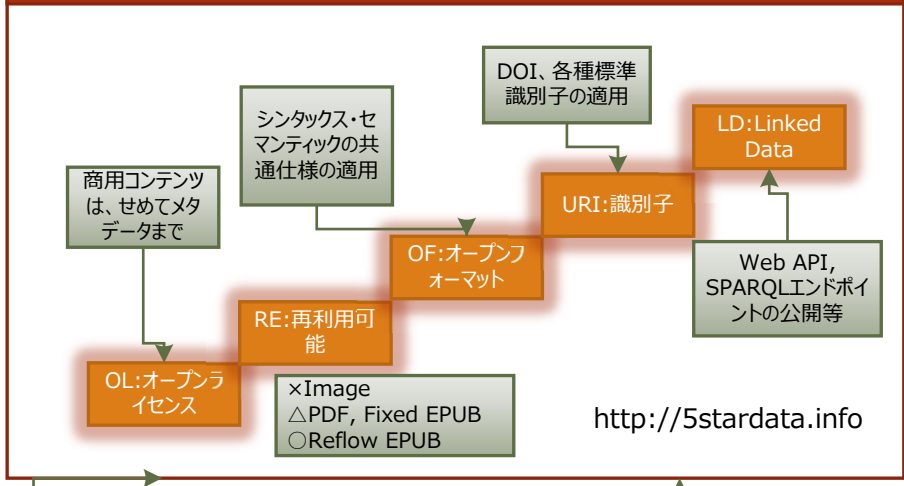
http://commons.wikimedia.org/wiki/Special:FilePath/Ch5_wakamurasaki.jpg?width=300

dbpedia-owl:wikiPageExternalLink

<http://www.uji-genji.jp/>
<http://www.is2.or.jp/>
<http://www.soinet.or.jp/~eshibuya/>
<http://www.wd.org/en/item/2690/>



Open Dataの定義



LOD化に必要なITスキルは？

- 例
- NDLサーチ
- WebNDLA
- CiNii
- Dbpedia

- コンセプト (NII大向先生)
- ウェブの方法論をデータ共有に活用
- ・識別子とリンク
- ・自由な参加
- 多様なデータの表現形式を一本化
- ・テキスト・表形式・ツリー形式,,,
- 既存のデータベースを活用する
- ・現在のウェブサービスは構造化されている
- ・概念の精緻化よりも今あるデータの関連付けを
- ・Web of DocumentからWeb of Dataへ

- LODの原則 (NII大向先生)
- Linked Data = ウェブ + RDF
- ・あらゆるデータの識別子としてURIを使用
- ・識別子にはHTTP URIを使用し、参照やアクセスを可能にする
- ・URIにアクセスされた際には、有用な情報を標準的なフォーマットで提供する
- ・データには他の情報源における関連情報へのリンクを含め、ウェブ上での情報発見を支援する

