

Challenges for Linked Data in Japan

Linked Dataにおける日本の課題

武田英明 Hideaki Takeda

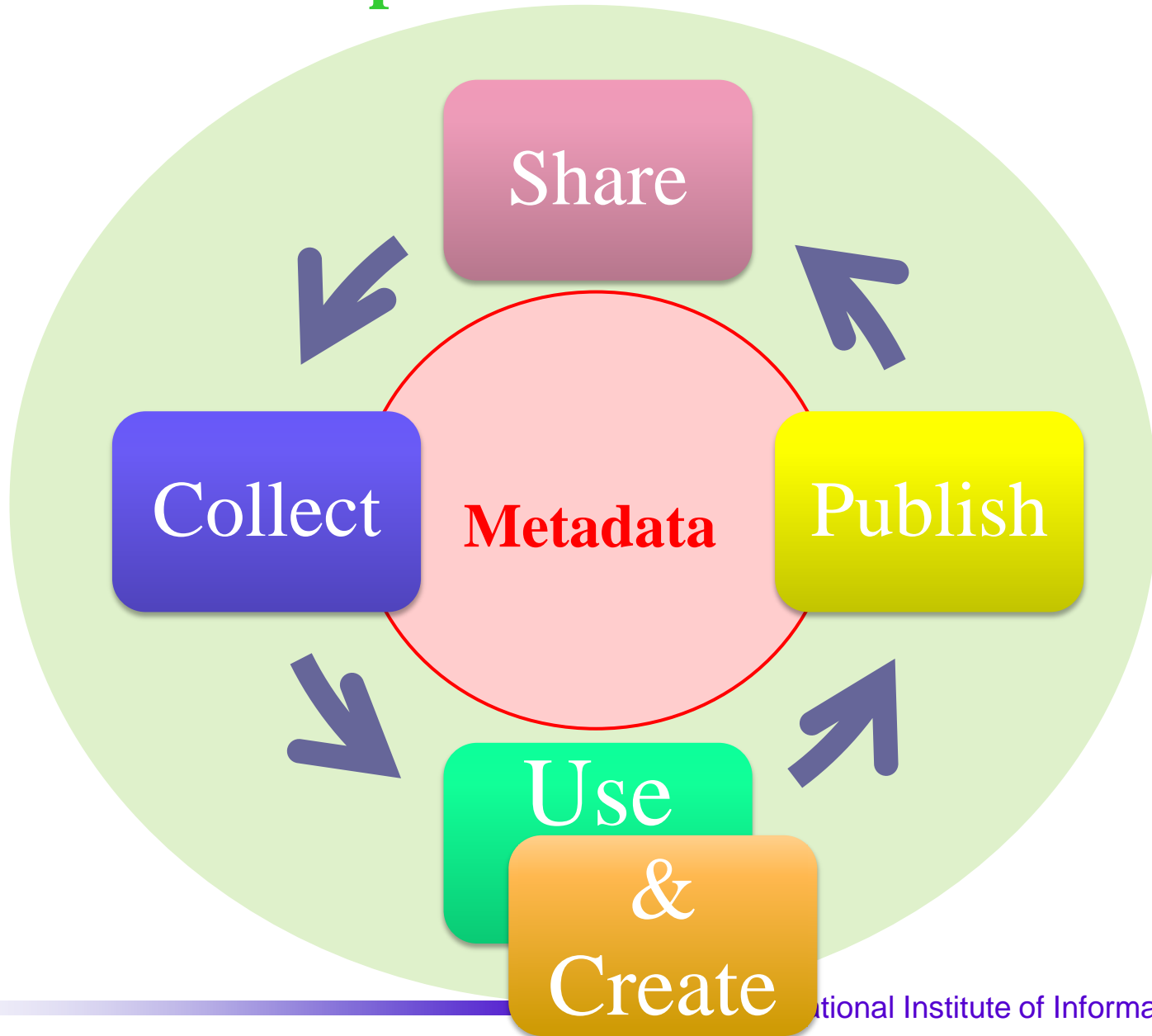
国立情報学研究所

National Institute of Informatics

Outline

- Information Cycle
/情報サイクルとLinked Data
- Challenges for Linked Data in Japan
/Linked Dataにおける日本の課題
- LODAC Museum

Metadata is the platform of Information Cycle



Linked Data

- Linked Dataとは “*Web of Data*”
 - RDFで公開されるデータ
 - 外部から参照可能
- Linked Dataのための4条件
 - 事柄の名前にURIを使うこと
 - ◆ *すべてのモノ, コトにURIを!*
 - 名前の参照がHTTP URIでできること
 - ◆ *DOIとかいったURNは使わないでね*
 - URIを参照したときに関連情報が手に入るように
 - ◆ *理解可能なデータを提供してね.*
 - 外部へのリンクも含めよう
 - ◆ *Webのようにリンクでつながるデータを作ろう*

Linked Data, TBL, <http://www.w3.org/DesignIssues/LinkedData.html>

Importance of data in public sector as Linked Data

Linked Dataにおける公共的組織のデータの重要性

- In principle, it should be shared / 本来共有すべき情報
- It is the basic knowledge of our society / 我々の社会の知識基盤
- Data in public sector
 - Library / 図書館
 - Museum / 美術館・博物館
 - Archive / 文書館
 - Government / 政府

Challenges for Linked Data in Japan

Linked Dataの日本での課題

- Lack of Culture of Sharing
/ 共有文化の欠如
- Immature Community for Linked Data
/ Linked Dataコミュニティの未発達
- Lack of Central Data Set
/ 中心的データの欠如
- Difficulty of Multi-lingual Data
/ 日本語の取り扱い

Lack of Culture of Sharing

共有文化の欠如

- Lack of consensus for public / 公共性に関する意識の薄さ
 - Culture of sharing and openness / 公開と共有の文化
 - Public often just means “governmental” (“官”)
 - ◆ Governmental vs. Private (“民”)
 - ◆ Obscurantism
 - 由らしむべし知らしむべからず
 - Advocate the culture of sharing
/ 共有の文化をもっと浸透させよう
 - ◆ This kind of symposium / 例えばこの種の会議

Immature Community for Linked Data

Linked Dataコミュニティの未発達

- Linked Data can be realized only by collaboration with technology-people and culture-people
/ 文理共同の必要性
 - Community in technology-people / 技術系の人
 - Community in culture-people / 文化の人
 - Cross-discipline Community bridging technology- and culture-people / 両者を交えたコミュニティ
- TAKE ACTION! / 行動しよう
 - LODAC project

Lack of Central Data Set

中心的データの欠如

- Dbpedia is the core of Linked Data
/ DbpediaはLinked Dataの中心
 - Easy to join LOD by linking Dbpedia
/ とにかく DbpediaにつなげばLODに入れる
- We are preparing dbpedia in Japan!
/ 日本語版dbpediaを準備中！

Difficulty of Multi-lingual Data

日本語の問題

- Use of non-English Languages
英語外の言語を使うときの問題
- Coding / コード
IRIを使えば文字コードとしては原理的にはOKだが、特殊文字や各種ツールの対応が問題になることがある
 - IRI (Internationalized Resource Identifier)
But, some exceptions
 - ◆ Special characters
 - ◆ Tools

Difficulty of Multi-lingual Data

日本語の問題

- Metadata by Japanese or by English / メタデータは日本語? 英語?
 - Use Japanese 日本語なら
 - ◆ Pros / 利点
 - Easy to understand / (日本人には) 理解しやすい
 - No modification of the original data (easy to publish) / 翻訳しないので簡単にデータが作れる
 - ◆ Cons / 欠点
 - Mix of English and Japanese is confusing / 日英混交
 - Difficulty to understand for non-Japanese / 日本人のみ理解
 - Use English / 英語
 - ◆ Pros / 利点
 - Internationally interoperable / 国際的に流通
 - Clean code / コードがクリーンで扱いやすい
 - ◆ Cons / 欠点
 - Translation is needed (may cause ambiguity) / 翻訳が必要
- **Decide it from the viewpoint of users! / 目的でわけよう**

LODAC Project

- Open Social Semantic Web Platform for Academic Resources
学術リソースのためのオープン・ソーシャル・セマンティックWeb基盤の構築
 - Providing platforms for Linked Data
/ Linked Dataの基盤をつくる
 - Practicing data accumulation and publishing
/ 実際にデータを集め利用可能にする
- Interested Areas / 現在興味をもっている分野
 - Museum information / 美術館・博物館情報
 - Geographical information, especially geographical names
/ 地理・地名情報
 - Local information / ローカルな情報
 - ...

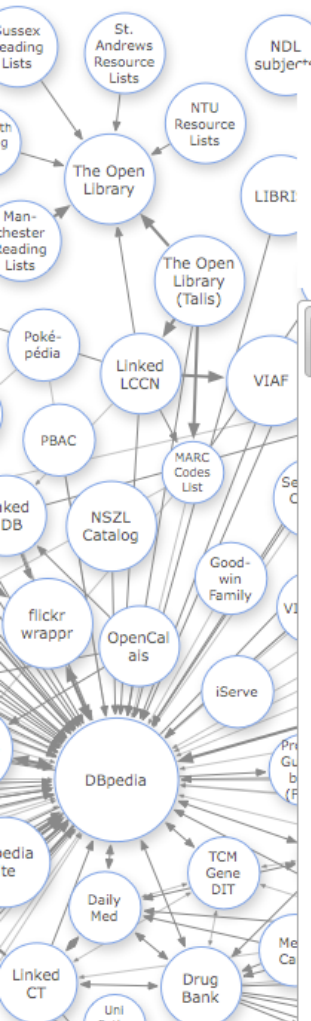
LODAC-Museum (tentative)

LODAC-Museum (仮)

- The state-of-the-art of museum information in Japan / 日本の美術館・博物館情報の現状
 - Distributed / 分散
 - ◆ Self maintained / 各館で維持
 - ◆ Isolated / 孤立
 - Opaque / 透明性がない
 - ◆ Self designed / 各館それぞれで設計
 - ◆ Messy / ばらばら
- Aggregating and associating museum information / 美術館・博物館情報を集めて関係づける
 - LODAC-Museum (tentative) / LODAC-Museum (仮)

下村観山

<http://lod.ac/id/359>

 別形式: [turtle/n3](#) | [xml](#) | [json](#) | [trix](#)


データ

参考1

参考2

<code>lodac:creates</code>	http://lod.ac/id/43625
<code>lodac:creates</code>	http://lod.ac/id/42903
<code>lodac:creates</code>	http://lod.ac/id/20755
<code>lodac:creates</code>	http://lod.ac/id/42995
<code>lodac:creates</code>	http://lod.ac/id/43350
<code>lodac:creates</code>	http://lod.ac/id/3762
<code>lodac:creates</code>	http://lod.ac/id/82734
<code>lodac:creates</code>	http://lod.ac/id/24768
<code>lodac:creates</code>	http://lod.ac/id/20029
<code>lodac:creates</code>	http://lod.ac/id/26732

<code>lodac:creates</code>	http://lod.ac/id/42974
----------------------------	---

<code>lodac:creates</code>	http://lod.ac/id/4588
----------------------------	---

<code>lodac:creates</code>	http://lod.ac/id/34400
----------------------------	---

<code>lodac:creates</code>	http://lod.ac/id/20128
----------------------------	---

<code>skos:prefLabel</code>	下村観山
-----------------------------	------

<code>rdf:type</code>	foaf:Person
-----------------------	-----------------------------

<code>dc:references</code>	http://dbpedia.jp/resource/下村観山
----------------------------	---

<code>dc:references</code>	http://lod.ac/ref/359
----------------------------	---

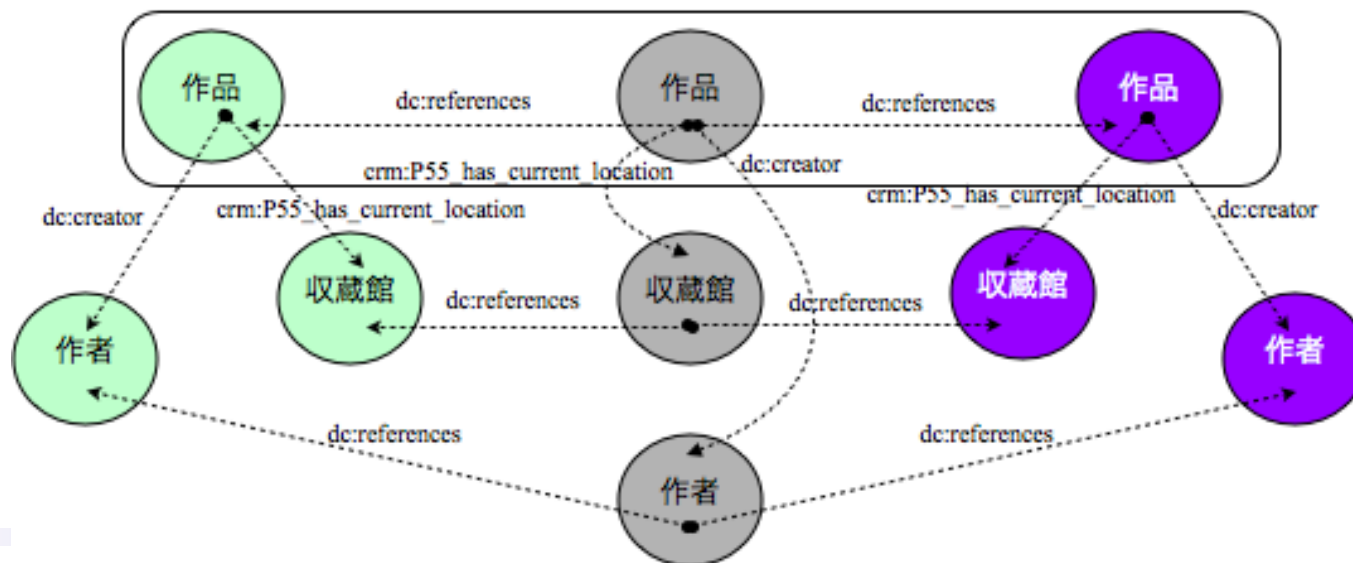
<code>foaf:name</code>	下村観山
------------------------	------

Metadata design / メタデータ設計

- Work – Creator – Museum/ 作品 – 作者 – 収蔵館
- Interoperability is more considered than correctness in the domain / 専門的な正確さより使いやすさを優先、
 - DC > DCTerm > FOAF > iCal > SKOS > NDLSH > RDA > CIDOC CRM
- The simple is the best
なるべく構造も簡単に
 - Keep it flat as long as possible

Integration Policy 統合方針

- How to integrate data from different sources / 異なる情報源から情報の統合
 - sharing of responsibility / 責任の分担
 - ◆ Each source is responsible for its data / データ源は各々管理
 - Identifying IDs for data and managing data with the IDs
 - ◆ LODAC is only responsible for integration / IDでひも付け
 - Assigning original IDs and associating other IDs to them



Data sources

- Thesaurus / 横断的美術データ
 - 日本美術シソーラスDB絵画編
 - 国指定文化財データベース
 - 文化遺産オンライン
- Collection / 個別コレクションデータ
 - 国立美術館所蔵作品総合目録検索システム(国立国際美術館, 京都国立近代美術館, 東京国立近代美術館)
 - 国立西洋美術館
 - 京都国立博物館
 - 奈良国立博物館
 - 福島県立美術館
- そのほか
 - 日本語版DBPedia Lite
 - 国土交通省国土計画局GIS
 - 栃木県立美術館
 - 秋田県立近代美術館
 - 岩手県立美術館
 - 徳島県立近代美術館
 - 山梨県立美術館
 - 東京都現代美術館
 - 香川県立東山魁夷せとうち美術館

data

情報源	情報種別	データ数
国立美術館(西美を除く 3 館)	作品	25180
国立西洋美術館	作品	4373
京都国立博物館	作品	5819
奈良国立博物館	作品	431
福島県立美術館	作品	20
栃木県立美術館	作品	32
秋田県立近代美術館	作品	22
岩手県立美術館	作品	1558
徳島県立近代美術館	作品	18482
山梨県立美術館	作品	262
東京都現代美術館	作品	5416
香川県立東山魁夷せとうち美術館	作品	266
日本美術シソーラス DB	作品	3800
日本美術シソーラス DB	人物	1332
日本美術シソーラス DB	グループ	289
日本美術シソーラス DB	施設情報	648
文化遺産オンライン	施設情報	915
国土交通省国土計画局 GIS	施設情報	25000
国指定文化財データベース	作品	10115
合計		103096

Data resources:529,449
(1,915,586 triples)

Summary

- Linked Data is the new way for sharing information
/ Linked Dataは情報共有の新しい方法
- Metadata is getting more important
/ Metadataはもっと重要になっている
- Time is now for action
/ いま行動するチャンス