

# Study Support and Integration of Cultural Information Resources with Linked Data

Tetsuro, KAMURA

Doctoral Student

School of Multidisciplinary Sciences, Department of Informatics  
The Graduate University for Advanced Studies(SOKENDAI)

Fumihiro, KATO (National Institute of Informatics)

Toru, TAKAHASHI (ATR-Promotions.Inc)

Hiroshi, UEDA (ATR-Promotions.Inc)

Ikki, OHMUKAI (National Institute of Informatics)

Hideaki, TAKEDA (National Institute of Informatics)

# Agenda

## Introduction

- Next Generation Web
- Linked Data

## Approach

- Gathering
- Standardization
- Integration & Association
- Publishing & Sharing

## Applications

- Yokohama Art Spot
- Photo BURARI<sub>(LODAC version)</sub>

# Introduction

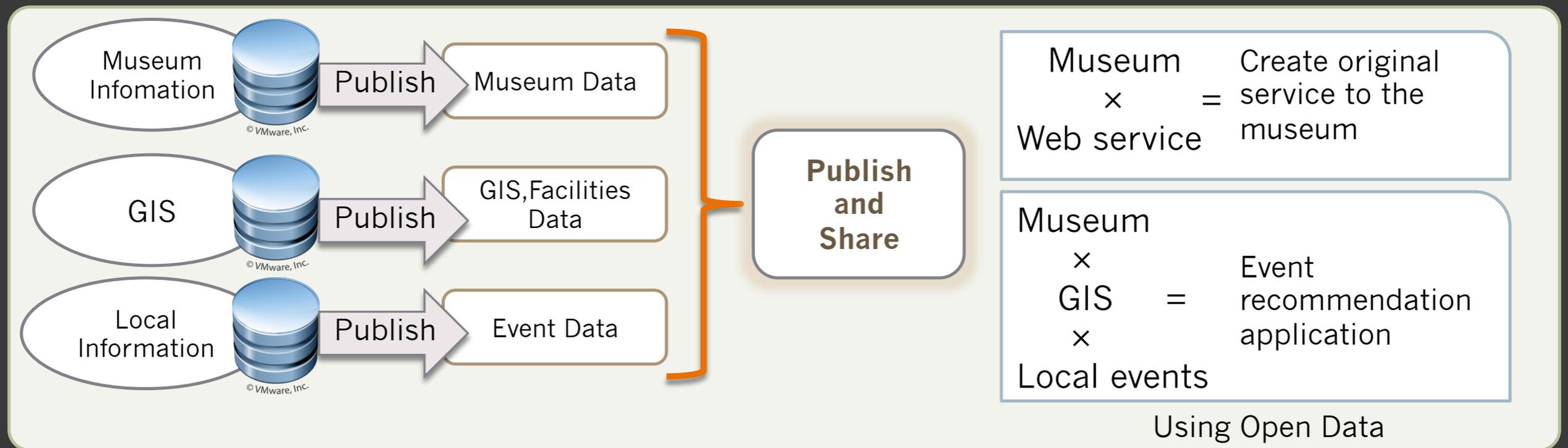
Up until now, lots of Japanese museums have been built **Online DB** and **Digitized collection**.



Each museum has developed **separate** collection management system with **original metadata**.

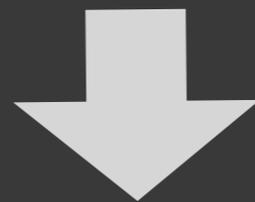
It is difficult to retrieve relevant information by searching **multiple** museum information.

If, we could be **search and use** the multiple information **on the Web....**

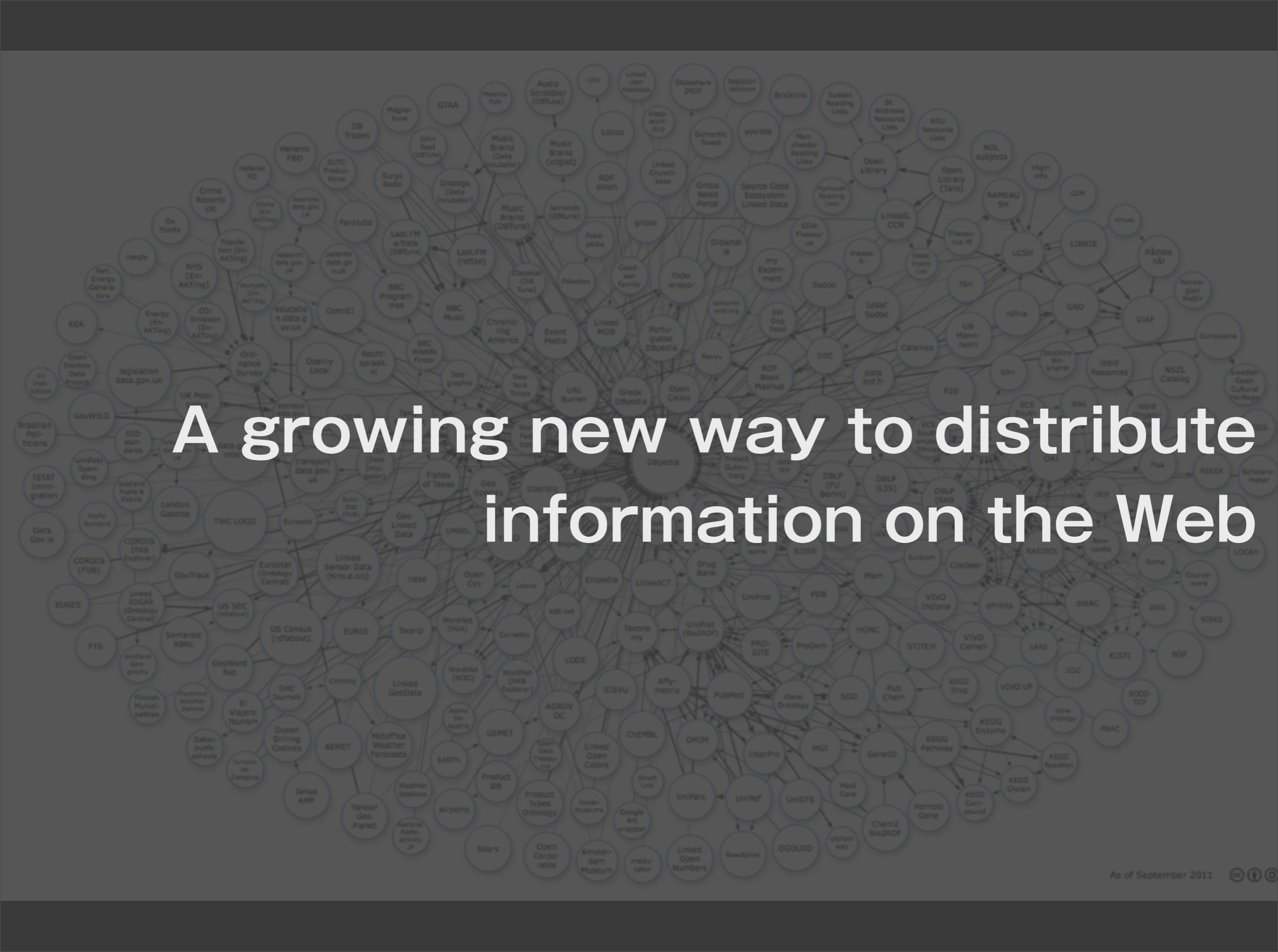


We can find new frontier, new study and create Web service so on...

The Museum field in particular.



**We challenge to solve Japanese Arts and Culture fields propositions with next generation Web technology**

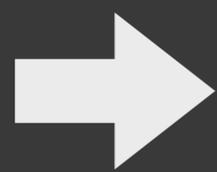


# A growing new way to distribute information on the Web

# Next Generation of Distribute Information

Existing Web = **Web of Document**

ex) PDF, HTML, Image format **information**.

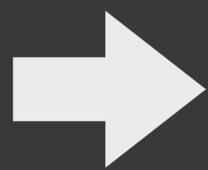


Already **processed** data.

if, want to use data, you have to extract data from pdf data or strip HTML Tags from HTML data.

Next Generation Web = **Web of Data**

ex) SPARQL Endpoint, RDF format **data**.



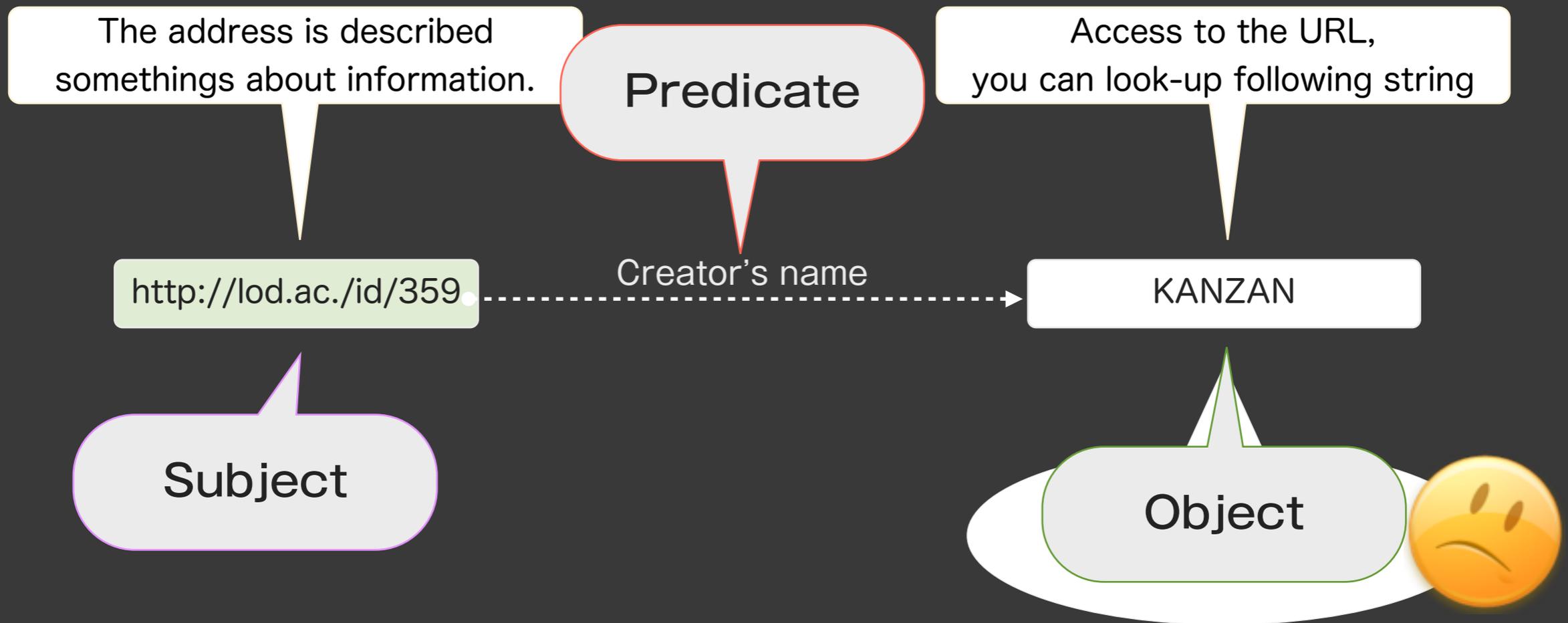
Directly refer to **open data**.

Available to use RAW data immediately.

The platform called...

# Linked Data

# Basic Structure of the Linked Data

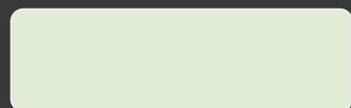
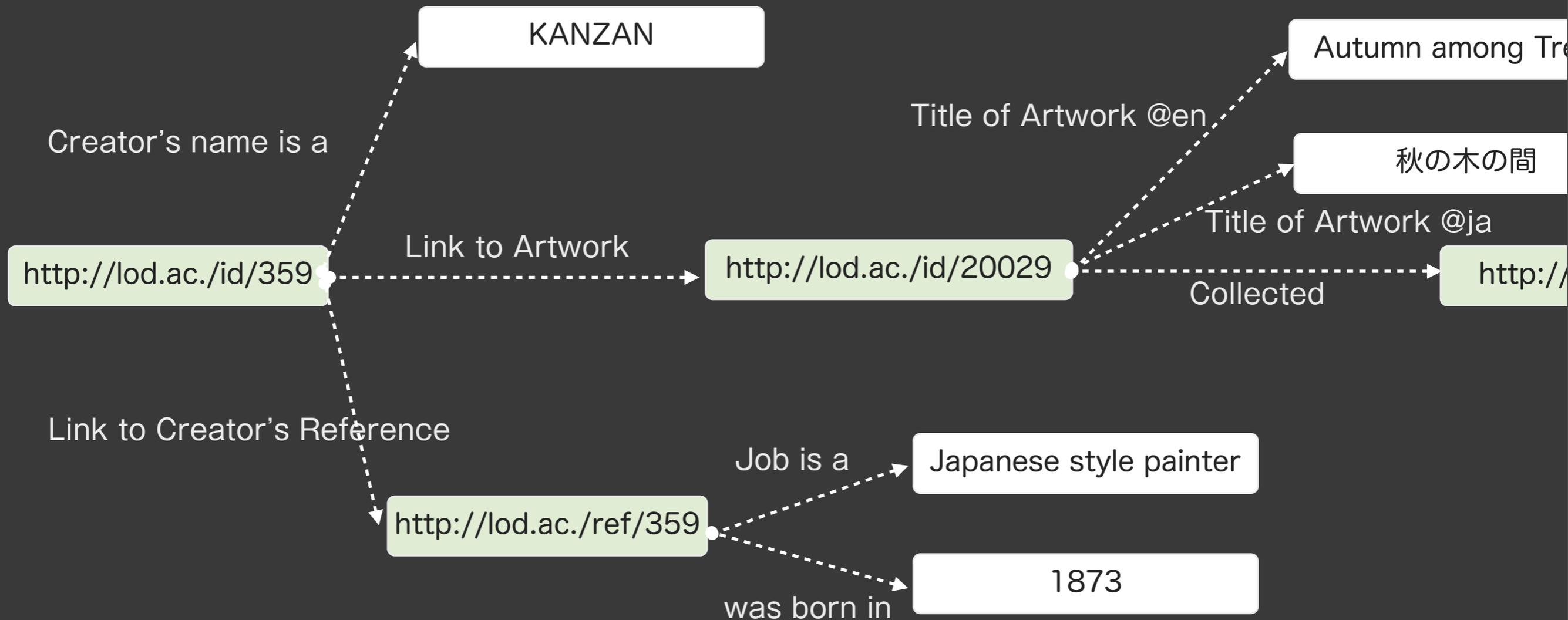


Understand “`http://lod.ac./id/359`” is described about “creator’s name” “KANZAN”

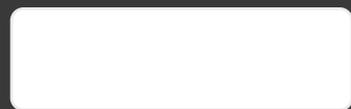
The Linked Data consist of **the three parts of resource.**

This structure is called RDF model. (Resource Description Framework) <sup>10</sup>

# Linking Data

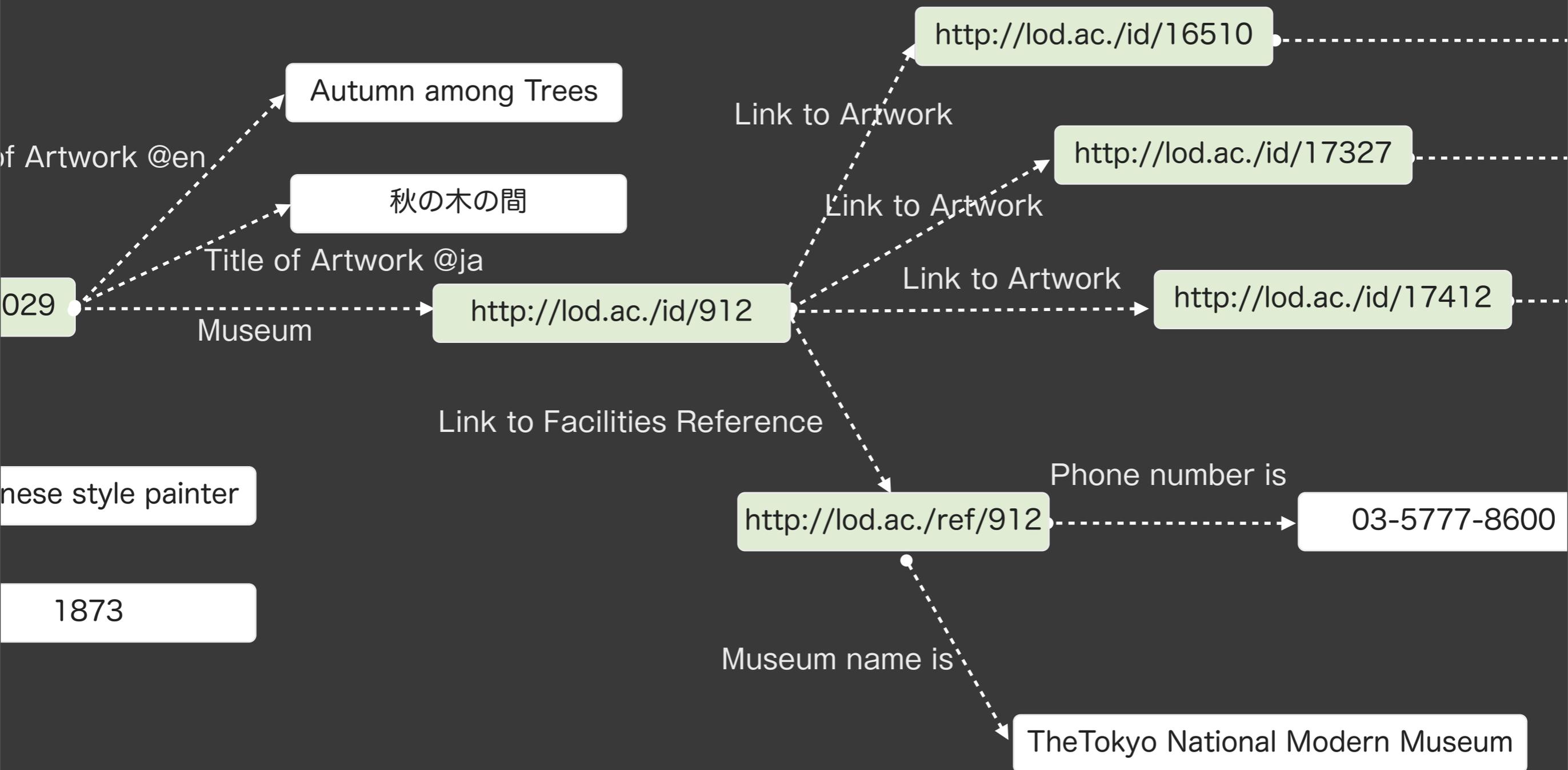


Link node, Contains other information links.



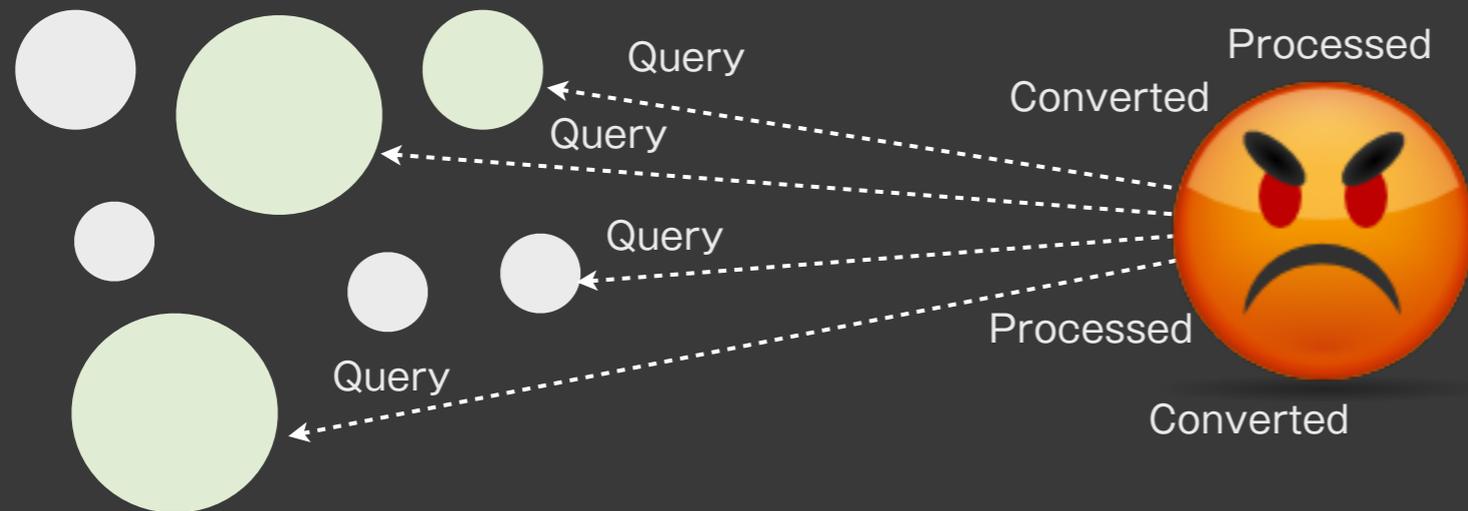
String node. Represent string information,(string,number,date )

# Linked Data represents information as node and arc labeled directed graph



# If, user wants look-up data.

## Current Web VS Linked Data

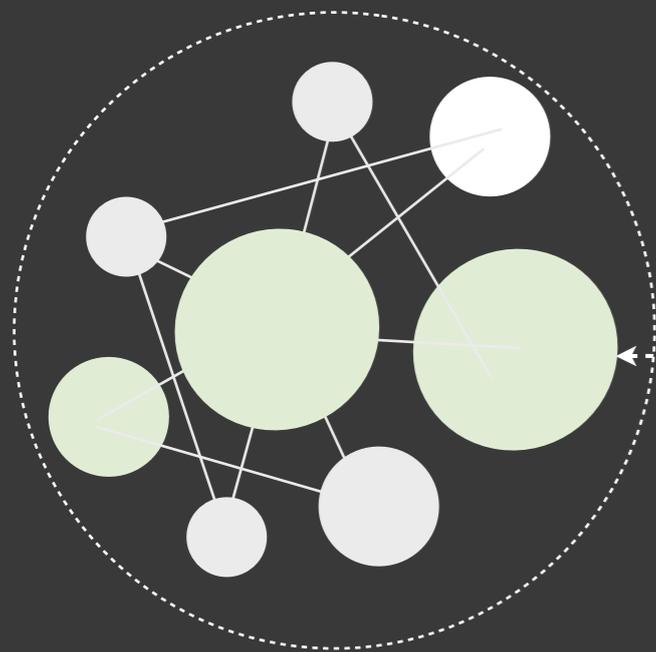


**Current**

Search and extract data with several websites every time.

Distribute Information

Irritated User

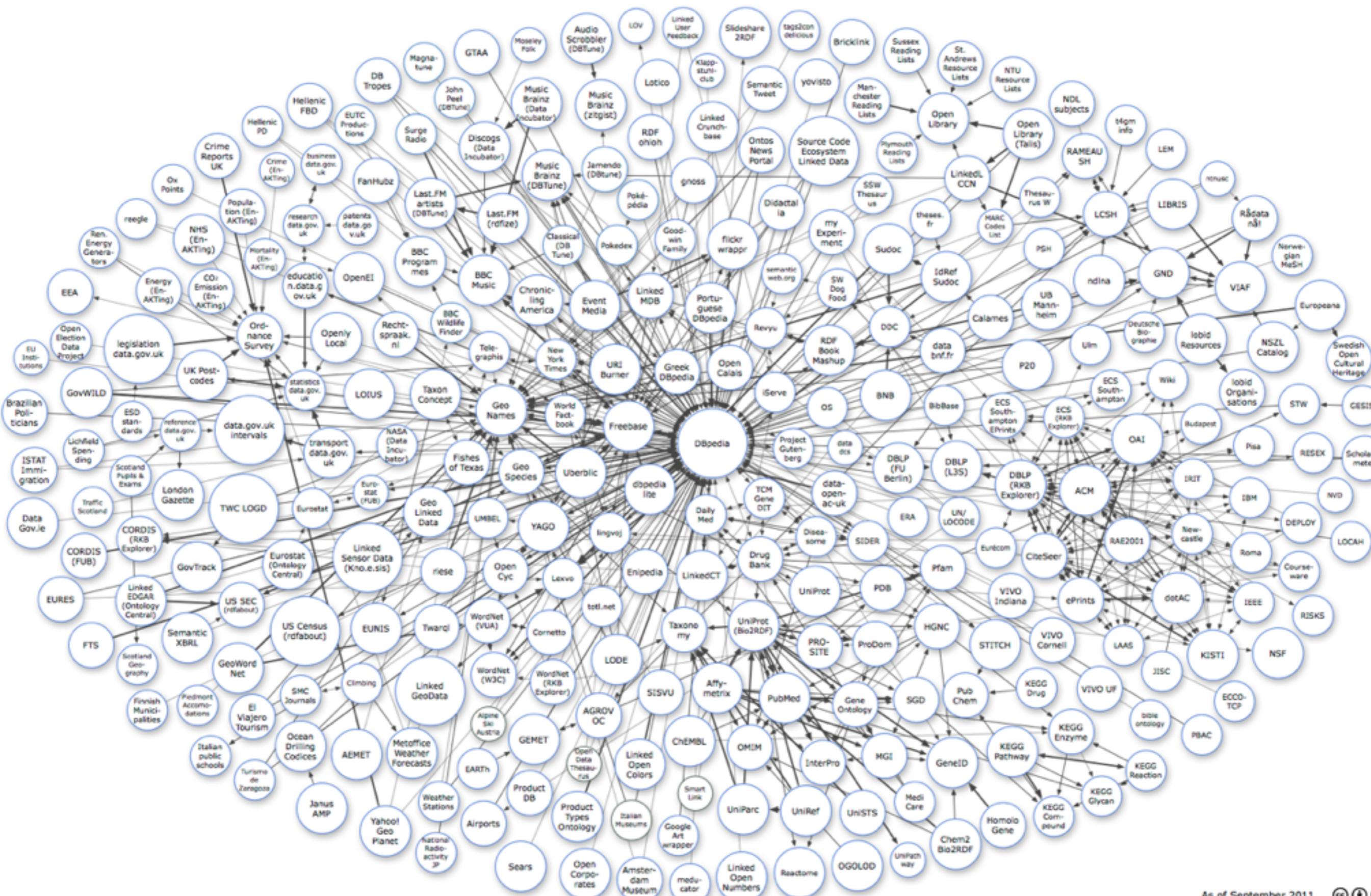


**Linked Data**

Querying integrated data.

Integrate Information

Happy User



## We applied Linked Data to a Japanese museum information.



The screenshot displays the LODAC Museum website. On the left, a network diagram shows various linked data sources such as 'The Open Library', 'St. Andrews Resource Lists', 'NDL subjects', 'NTU Resource Lists', 'LIBRIS', 'The Open Library (Talis)', 'Linked LCCN', 'VIAF', 'PBAC', 'NSZL Catalog', 'Goodwin Family', 'OpenCalais', 'iServe', 'flickr wrapper', 'NSZL Catalog', 'Goodwin Family', 'Vivi', 'Sem Cru Be', 'MAAC Codes List', 'Pokepedia', 'Man-hesher reading Lists', 'Listex reading Lists', and 'iServe'. The main content area features the 'LODAC Museum' logo, a search bar with a 'Search' button, and a list of examples in Japanese:

例:

- [藤田朋治](#)
- [下村観山](#)
- [横浜美術館](#)
- [三重県立美術館](#)
- [地獄草紙](#)

The footer includes the text 'LODAC Project powered by OWLIM'.

# Approach

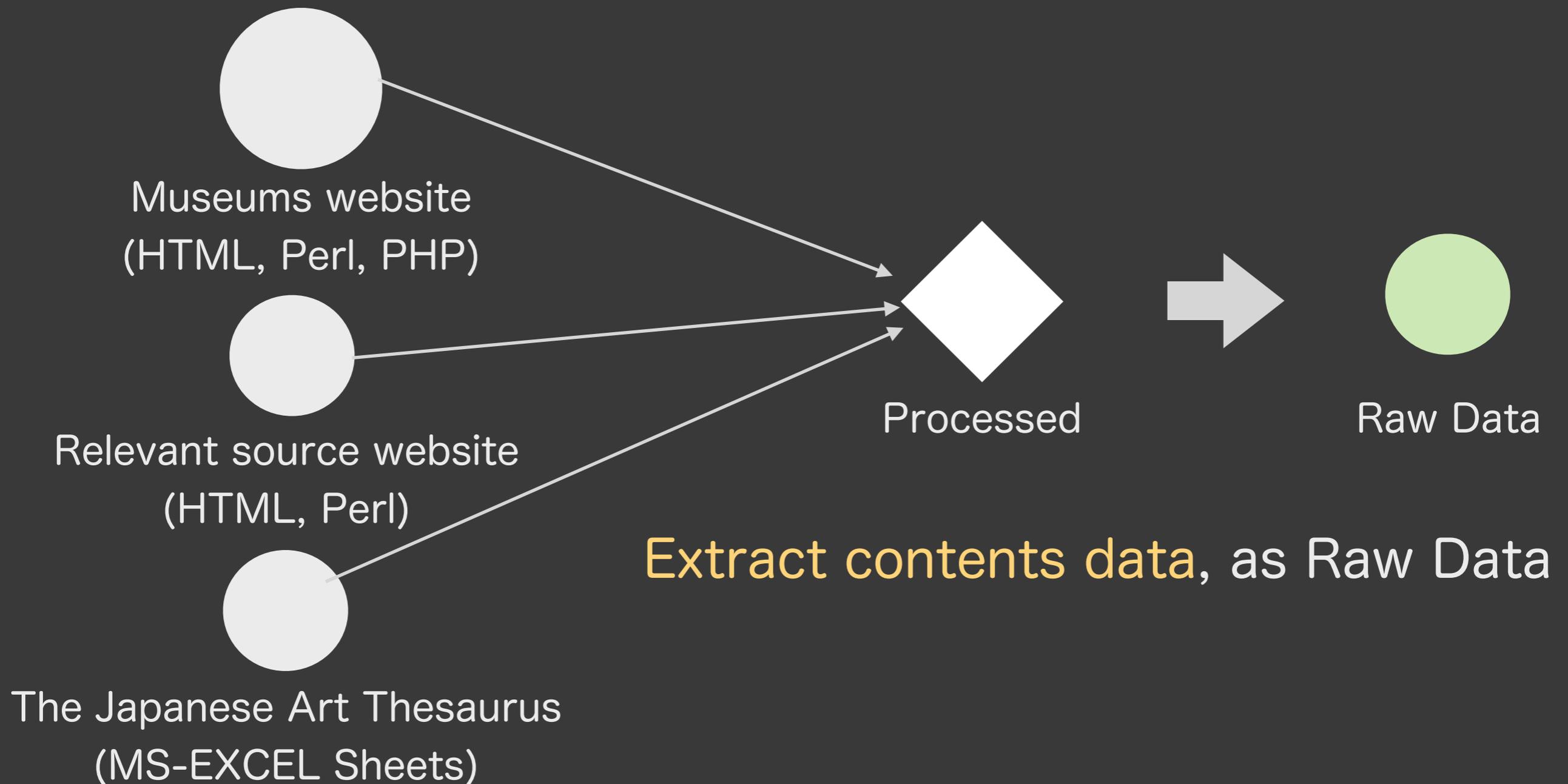
# Gathering data

Museums Source	Uses Data	Amount of Data
Catalog of the collections of 3 National Art Museum	Artwork	25,180
National Museum of Western Art		4,373
Kyoto National Museum		5,819
Nara National Museum		431
Fukushima Pref. Art Museum		20
Tochigi Pref. Art Museum		32
Akita Pref. Art Museum		22
Iwate Pref. Art Museum		1,588
Tokushima Pref. Art Museum		18,482
Yamanashi Pref. Art Museum		5,416
Kagawa Higashitama Kaii Setouchi Art Museum		5,416
Yokohama Art Museum		6,286

Relevant Sources	Use Data	Amount of Data
Database for National Treasure & Important Cultural Property of National Designated	Artwork	915
Cultural Heritage Online	Facilities	648
DBpedia Japanese (Referred to DBpedia)	WikiPedia	-
GIS data National and Regional Planning Bureau	Geographical	
	Facilities	
The Japanese Art Thesaurus	Artwork	266
	Creator	3,800
	Association for Arts	1,332
	Facilities	289
Overall		109,382

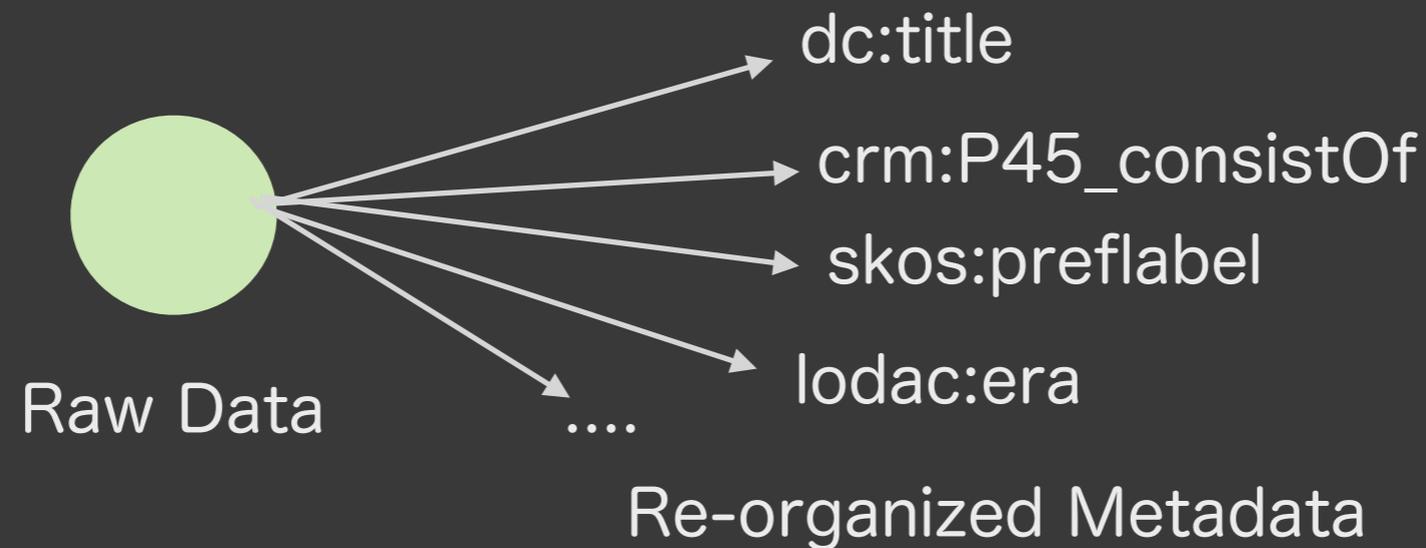
Covers a wide range of content types as already structured concept. Contains several metadata such as creator name, work title, era, owner, current location, facilities.

# Scraping and processing sources



# Standardization of data

Re-organized common metadata.



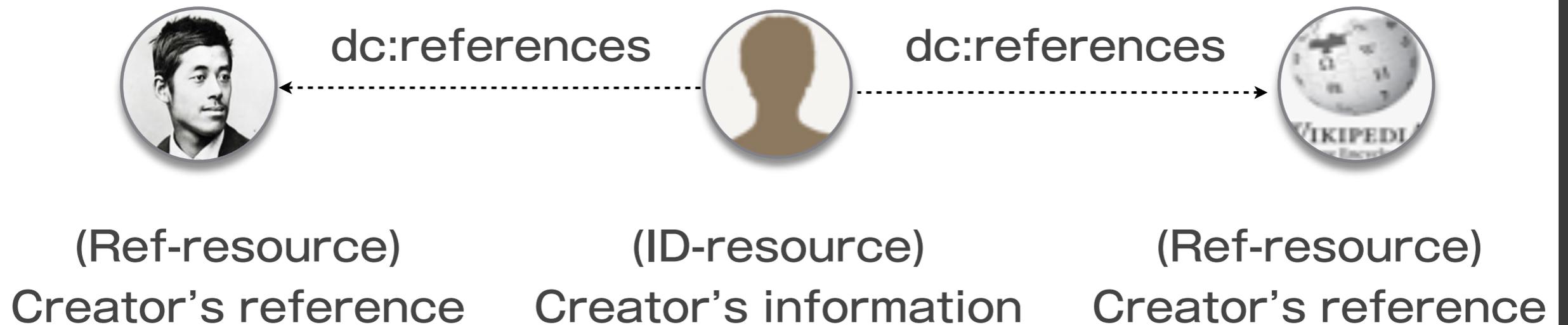
## Current organized policies

- Use **existing** metadata. (Use string as data only)
- **Define own metadata.**

<b>Prefix</b>	<b>Metadata Name</b>
crm	CIDOC-CRM
dc11	Dublin Core 1.1
dc	DCMI Terms
skos	Simple Knowledge Organization System
rdfs	Resource Description Frame Work Schema
foaf	Friend of a Friend
rda2	Resource Description and Access
lodac	LODAC Project

# Integrating Data

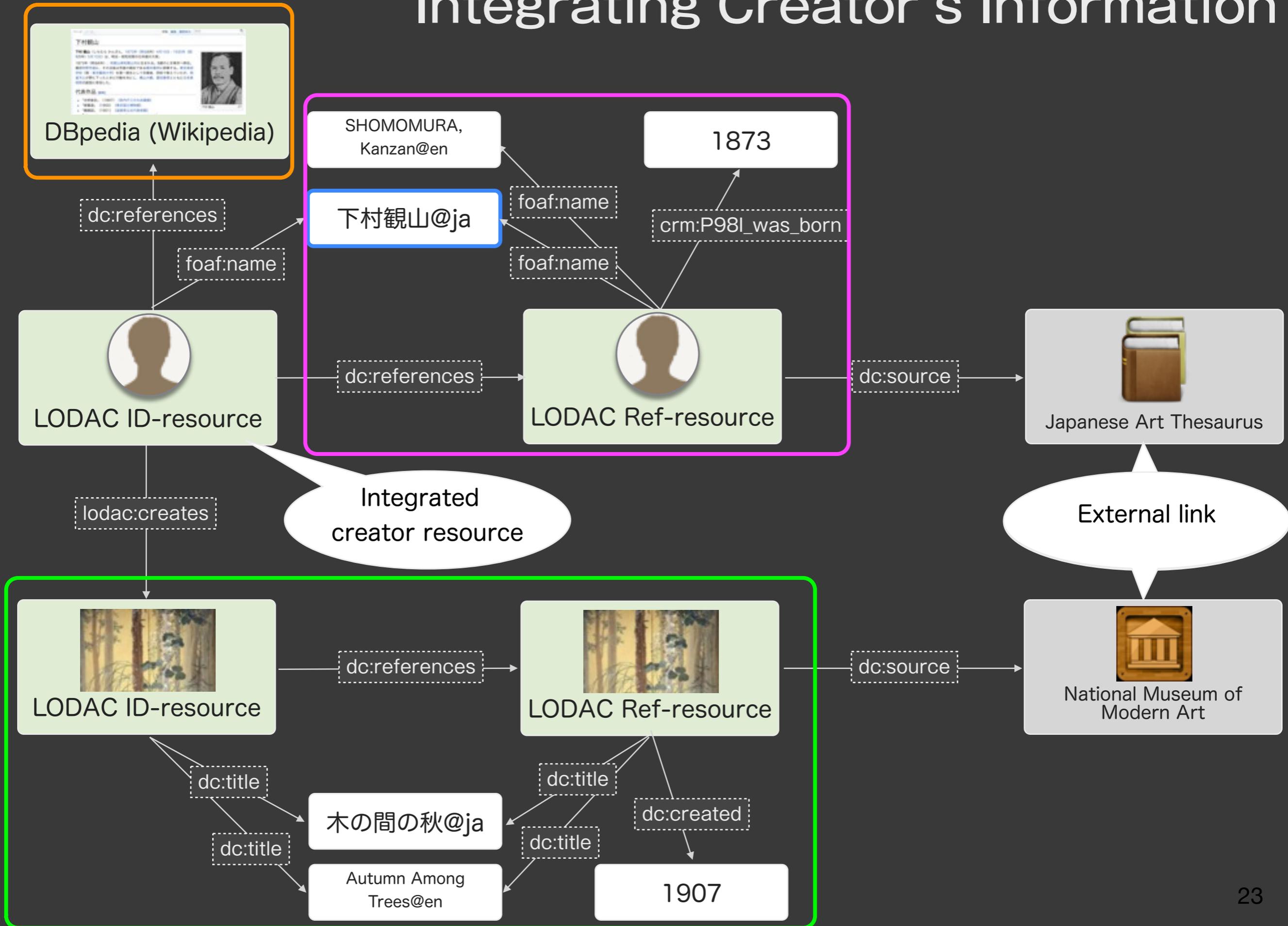
## Integration Data



Generated from  
RAW data to RDF

Generate RDF and  
assigned LODAC ID

# Integrating Creator's Information



# Associating data

## Associate Creator and Artwork

A. Japanese Art Thesaurus - **1,332** creators

B. All of artwork - **61,861** titles

## Using string match method

Matching KEY

- A. Creator of artwork
- B. Creator's Name

Integrate Item	Source	Amount of Data	Integration Data
Facilities	A.Japanese Art Thesaurus	648	77
	B.Cultural Heritage Online	915	
Title of important cultural properties	A.Japanese Art Thesaurus (Art work)	3,800	74
	B.DB for National Treasure (Art work)	10,115	
Creator information and Work Title	A.Japanese Art Thesaurus (Creator)	1,332	15,020
	B.All of art work (Work title string)	61,861	
Creator name	A.Japanese Art Thesaurus (Creator)	1,332	615
	B.All of art work title(using creator name)	61,861	

# Publishing & Sharing

We build a Linked Data infrastructure for  
for the museum information

The screenshot displays the LODAC Museum website. At the top left is the LODAC logo. A navigation bar includes links for Home, SPARQL Endpoint, Location, Apps, Site Info, Blog, and Wiki. On the left side, a network diagram shows various data sources like 'The Open Library', 'Linked LCCN', 'VIAF', and 'MARC Codes List' interconnected. The main content area features the text 'LODAC Museum' with a pink star icon containing the Japanese characters 'alpha版' (alpha version). Below this is a search bar with a 'Search' button. A list of examples is provided:

例:

- [藤田嗣治](#)
- [下村観山](#)
- [横浜美術館](#)
- [三重県立美術館](#)
- [地獄草紙](#)

At the bottom right, it says 'LOD.AC Project powered by OWLIM'.

# Publish data as RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:ns0="http://purl.org/dc/terms/"
xmlns:ns1="http://xmlns.com/foaf/0.1/"
xmlns:ns2="http://lod.ac/ns/lodac#"
xmlns:ns3="http://www.w3.org/2000/01/rdf-schema#"
xmlns:ns4="http://www.w3.org/2004/02/skos/core#">
  <rdf:Description rdf:about="http://lod.ac/id/359.json">
    <ns0:title>JSON representation for http://lod.ac/id/359</ns0:title>
    <ns1:primaryTopic>
      <ns1:Person rdf:about="http://lod.ac/id/359">
        <ns2:creates rdf:resource="http://lod.ac/id/20029"/>
        <ns0:references rdf:resource="http://dbpedia.jp/resource/%E4%B8%8B%E6%9D%91%E8%A6%B3%E5%B1%B1"/>
        <ns0:references rdf:resource="http://lod.ac/ref/359"/>
        <ns3:label xml:lang="ja">下村観山</ns3:label>
        <ns4:prefLabel xml:lang="ja">下村観山</ns4:prefLabel>
        <ns1:name xml:lang="ja">下村観山</ns1:name>
      </ns1:Person>
    </ns1:primaryTopic>
  </rdf:Description>
</rdf:RDF>
```

ID-resource URI  
(Own address)

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

External link  
DBpedia Japanese

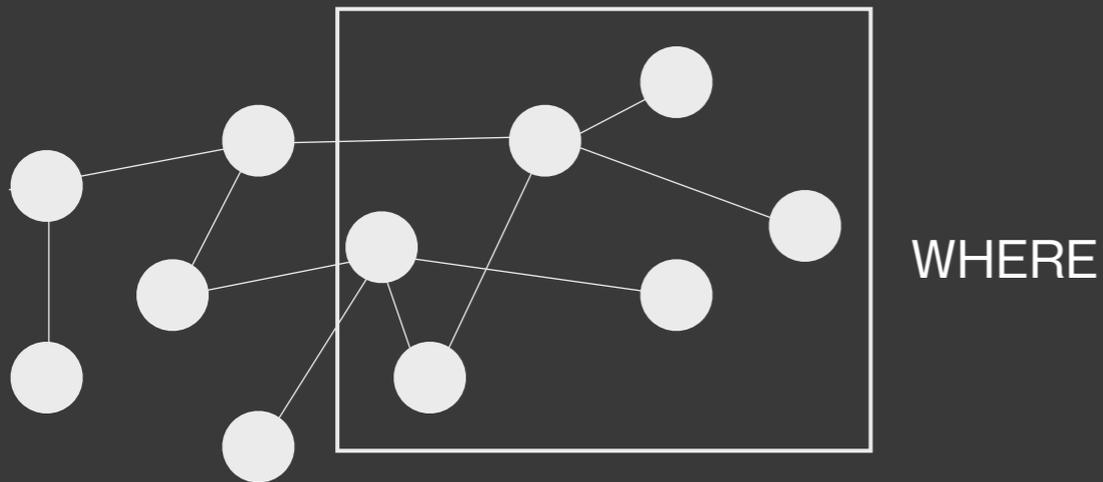
Ref-resource URI

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

# SPARQL Query

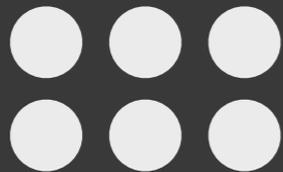
SPARQL query language is widely used for querying RDF data.

How many duplicate titles?



Pull an artwork resources out of the RDF dataset

An artwork resources



SELECT

Pulled data, count duplicate work title.

# TOP20's Duplicate Titles

## SPARQL Query Results

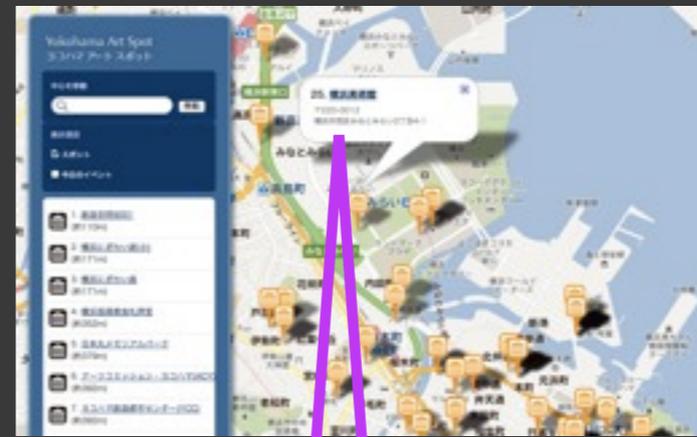
workName	countName
[無題] @ja	425
無題 @ja	357
作品 @ja	207
裸婦 @ja	190
風景 @ja	153
木造阿彌陀如来坐像 @ja	141
(題名なし) @ja	134
自画像 @ja	125
山水図 @ja	122
木造薬師如来坐像 @ja	120
木造十一面観音立像 @ja	119
静物 @ja	101
ヌード @ja	94
花 @ja	79
裸婦(らふ) @ja	76
木造地藏菩薩立像 @ja	70
木造聖観音立像 @ja	67
壺 @ja	65
梵鐘 @ja	62
木造毘沙門天立像 @ja	56

# Analyzed Technique and Medium of the Artworks

label	label	label	label
その他 @ja	一面 @ja	金属・板シルクスクリーン @ja	etching, aquatint @en
写真 @ja	一本 @ja	paper and label seal @en	lithograph and off-set @en
水墨画 @ja	彩牋墨書:一帖 @ja	紙・既成ラベル @ja	オフセット、リトグラフ @ja
絵馬 @ja	インタリオ @ja	stencil, paper @en	紙・墨摺(木版) @ja
木版画 @ja	エナメル、ボール・板 @ja	孔版、紙 @ja	紙・墨拓摺 @ja
板目木版 @ja	エナメル、板、その他 @ja	drypoint, aquatint @en	screenprint @en
卵テンペラ @ja	エンコスティック・紙 @ja	ドライポイント、アクアチント @ja	スクリーンプリント @ja
テンペラ @ja	オフセット @ja	墨・紙 @ja	linocut @en
gouache @en	紙本金地著色:八曲屏風一双 @ja	silk screen @en	リノカット @ja
グワッシュ @ja	紙本金地著色:六面 @ja	lithograph, off set, paper @en	siklscreen, paper @en
12 @ja	紙本金地著色:六曲屏風一双 @ja	オフセット・リトグラフ @ja	intaglio on paper @en
パステル @ja	紙本彩色:一幅 @ja	lithograph, off set @en	インタリオ・紙 @ja
油絵 @ja	紙本墨画:一幅 @ja	紙オフセット・リトグラフ @ja	relief process, silkscreen,paper @en
クレヨン @ja	紙本墨画:一巻 @ja	lithograpf, off set, paper @en	凸版,シルクスクリーン,紙 @ja
素描 @ja	紙本墨画:額装 @ja	紙、オフセット、リトグラフ @ja	color woodcut, framed @en
墨 @ja	紙本墨画:十六幅 @ja	lithograph, framed @en	木版手彩色 @ja
鉛筆 @ja	紙本墨画:七賢ノ間貼付十六枚 @ja	silkscreen @en	lithograph, embossing @en
インク @ja	紙本墨画:二幅 @ja	リトグラフ(手彩色) @ja	リトグラフ,空押(エンボス) @ja
水彩 @ja	紙本墨画:八幅 @ja	metal print @en	リトグラフ,空押(エンボ) @ja

# Applications

# YOKOHAMA Art Spot



## Facilities

### 横浜美術館

桜木町駅(JR)下車、動く歩道を利用徒歩10分。みなとみらい駅(みなとみらい線)下車、「美術館口」を出て徒歩3分。



## Local

イベント

収蔵品リスト

ジャンルで絞り込み

- すべて
- 日本画
- 油彩画(日本)
- 油彩画(外国)
- 版画(日本)
- 版画(外国)
- 水彩・素描(日本)
- 水彩・素描(外国)
- 彫刻(日本)
- 彫刻(外国)
- 工芸(日本)
- 工芸(外国)
- 写真
- 写真(日本)
- 写真(外国)
- 作品外資料

#### 童児図

恩地孝四郎

1923 油彩画(日本) 油彩、カンヴァス/59.0x44.5cm

## Artwork

#### ジャンル

芸術・文化

#### 営業時間または休業日

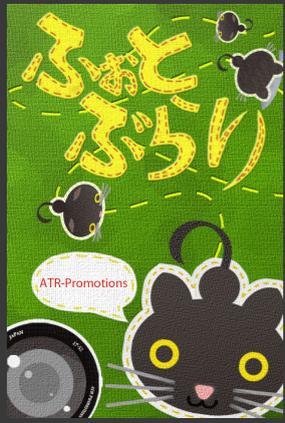
毎週木曜日(祝日を除く)、木曜日に祝日開館した場合はその翌日、年末年始

#### 住所

〒220-0012  
横浜市西区みなとみらい3丁目4-1

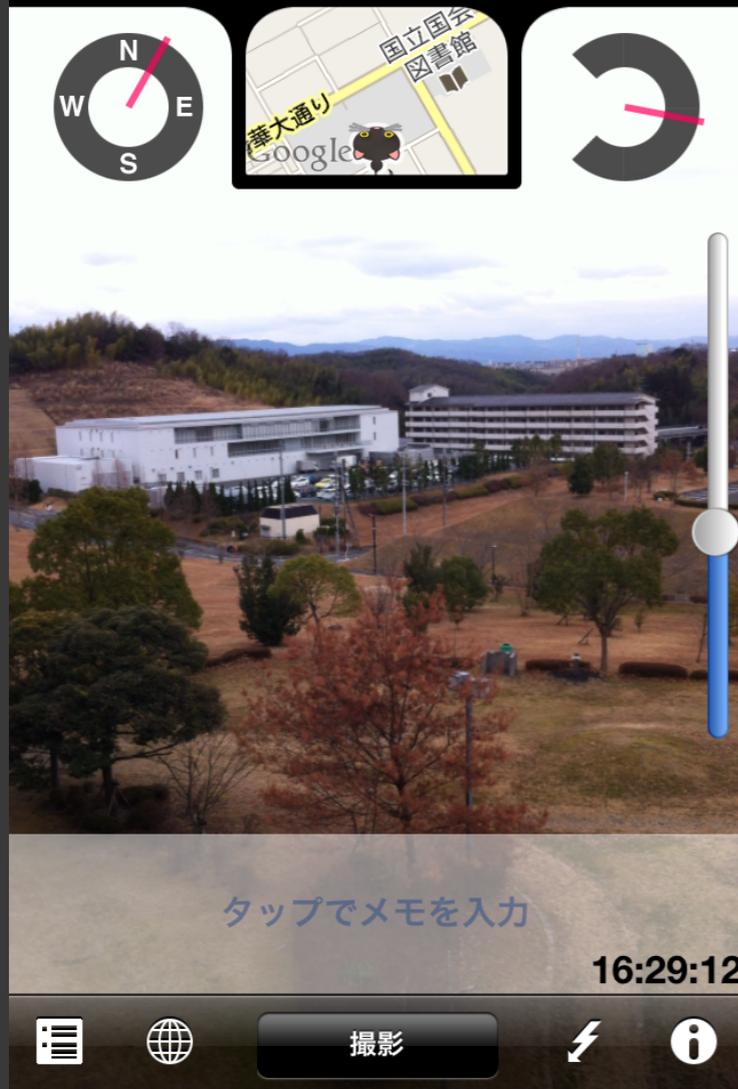


## GIS



# Photo BURARI (LODAC.Ver)

(C)ATR-Promotions,Inc



GIS and Facilities information  
through the SPARQL

# Summary

- **Organizing**

We tried to integrating distributed information as Linked Data. In consequence, approximately 11 million information available for common platform.

- **Publishing**

We published an RDF data on a LODAC Museum website. These are everybody can use for free!

- **Using**

Currently, the two applications use LODAC Museum's Data. We are more consider how to use these resources.  
(We have a plan to use for the purpose of study)

# LODAC Project

Linked Open Data for Academia

<http://lod.ac>



# Typical Issue

- **Fluctuation of description**

It is serious to integrate data. We need good reference database for abbreviation,

- **Updating resources**

We need find out how to give them regular or on-demand updates.

- **Easy participation**

We need to provide easier ways to participate in Linked Data such as importing collection data from other format data.

- **Authority information**

We need the technical term thesaurus.

16,605,601

# LODAC Project

Linked Open Data for Academia

1. Provide platform for **Linked Data**.
2. Use **academic** resources on the Web.
3. **Accumulate** data and **distribute**.

First target field is...

- **Complex structure** of information
- **An unexplored** field of Web technology
- **Contribution** to the academic and society



## 横浜美術館

<http://lod.ac/id/3172>

formats ▼

- turtle/n3
- xml
- json
- trix

Export RDF data

データ 参照1 参照2 参照3 参照4  
参照5 参照6

URI:<http://lod.ac/id/3172>

タイプ (rdf:type)	<a href="#">foaf:Organization</a>
ラベル (rdfs:label)	横浜美術館@ja

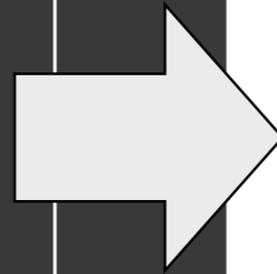
# Ref-resource

## Scraped Data

(HTML Source(Art work))



Creator  
Title  
Location



## Reference resource

(Ref-resource)

Assigned a unique ID



Convert into RDF format

Can not change for any resources context as scraped data.  
Any context (without metadata) have authority over each  
museums curator.

# ID-resource

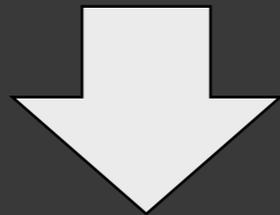
Step 1.

Based on the  
Japanese Art Thesaurus



Use Several Source  
Artwork  
Creator  
Facilities  
....

Generate RDF



Step 2.

Associating the same  
data by a string match

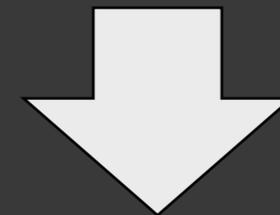
Matching based data to all of resources



=



Identified as RDF



## LODAC ID resource

(ID-resource)

Assigned a unique ID



# 下村観山

http://lod.ac/id/359

ID-resource  
(Integration data)

formats ▾

データ

参照1

参照2

参照3

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

タイプ  
(rdf:type) foaf:Person

ラベル  
(rdfs:label) 下村観山@ja

名前  
(foaf:name) 下村観山@ja

参照 URI  
(dc:references) [dbpedia.jp:下村観山](http://dbpedia.jp/下村観山)

参照 URI  
(dc:references) [lod.ac:ref:359](http://lod.ac/ref/359)

ラベル  
(skos:prefLabel) 下村観山@ja

作品 URI  
(lodac:creates) [lodacid:20029](http://lod.ac/id/20029)

作品 URI  
(lodac:creates) [lodacid:20128](http://lod.ac/id/20128)

下村観山

下村 観山（しもむら かんざん、1873年（明治6年）4月10日 - 1938年（昭和13年）5月10日）は、明治 - 昭和初期の日本の画家。

1873年（明治6年）、東京都神田区山手町に生まれる。本郷のとき東京府立第一中学校に入学し、その後東京府立第一中学校に転学する。東京美術学校（現・東京藝術大学）を第一専攻として卒業後、同校で教えたが、同校で学んだ時に下村の行動を共にし、横山大観、藤田鳴鶴とともに日本美術院の創設に参加した。

代表作品 [画像](#)

- 「光琳遺稿」（1897）（京内庁工の丸内蔵庫）
- 「静寂道」（1900）（東京国立博物館）
- 「無常観」（1901）（京内庁工の丸内蔵庫）
- 「タイオゼニス」（1903）（東京国立近代美術館）
- 「木の影の秋」（1907）（東京国立近代美術館）
- 「大津御幸」（1908）（東京国立近代美術館）
- 「無常観」（1912）（東京国立博物館）
- 「白鷺」（フエッコ）（1914）（東京国立博物館）
- 「無常観」（よるぼろし）（1915）（東京国立博物館）（重要文化財）
- 「静寂」（1916）（東京国立博物館）
- 「無常」（1921）（東京国立博物館）
- 「静寂観」（1922）（京内庁工の丸内蔵庫）

DBpedia  
(WikiPedia)

```

rdf:type      lodac:PersonReference
rdfs:label    下村観山 @ja
dc:source     http://www.tulips.tokuba.ac.jp/ja1/mokul/index.html
foaf:name     下村観山 @ja
foaf:name     SHIMOMURA, Kanzan Gen
foaf:name     しもむら・かんざん @ja-rrkt
skos:prefLabel 下村観山 @ja
dc:isReferencedBy  lodacid:359
lodac:creates lodac:ref:359
lodac:creates lodac:ref:4588
lodac:area    1622番 @ja

```

Ref-resource  
(Japanese Art  
Thesaurus)

Autumn among Trees  
<http://lod.ac/id/20029>

formats ▾

URI: <http://lod.ac/id/20029>

タイプ  
(rdf:type) lodac:Work

ラベル  
(rdfs:label) 木の影の秋@ja

ラベル  
(rdfs:label) Autumn among Trees@en

参照 URI  
(skos:references) [lodac:ref:283435](http://lod.ac/ref/283435)

Ref-resource  
(Art Work from  
each Museum)

# SPARQL Endpoint

## SPARQL Endpoint

### Query

- prefixes -

```
1 PREFIX dc: <http://purl.org/dc/terms/>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX lodac: <http://lod.ac/ns/lodac#>
4 PREFIX lodacid: <http://lod.ac/id/>
5
6 SELECT ?work ?label
7 WHERE {
8   lodacid:3172 lodac:isProviderOf ?work .
9   ?work rdfs:label ?label;
10  dc:references ?workRef .
11  ?workRef dc:medium "紙本着色"@ja .
12 }
13 LIMIT 100
```

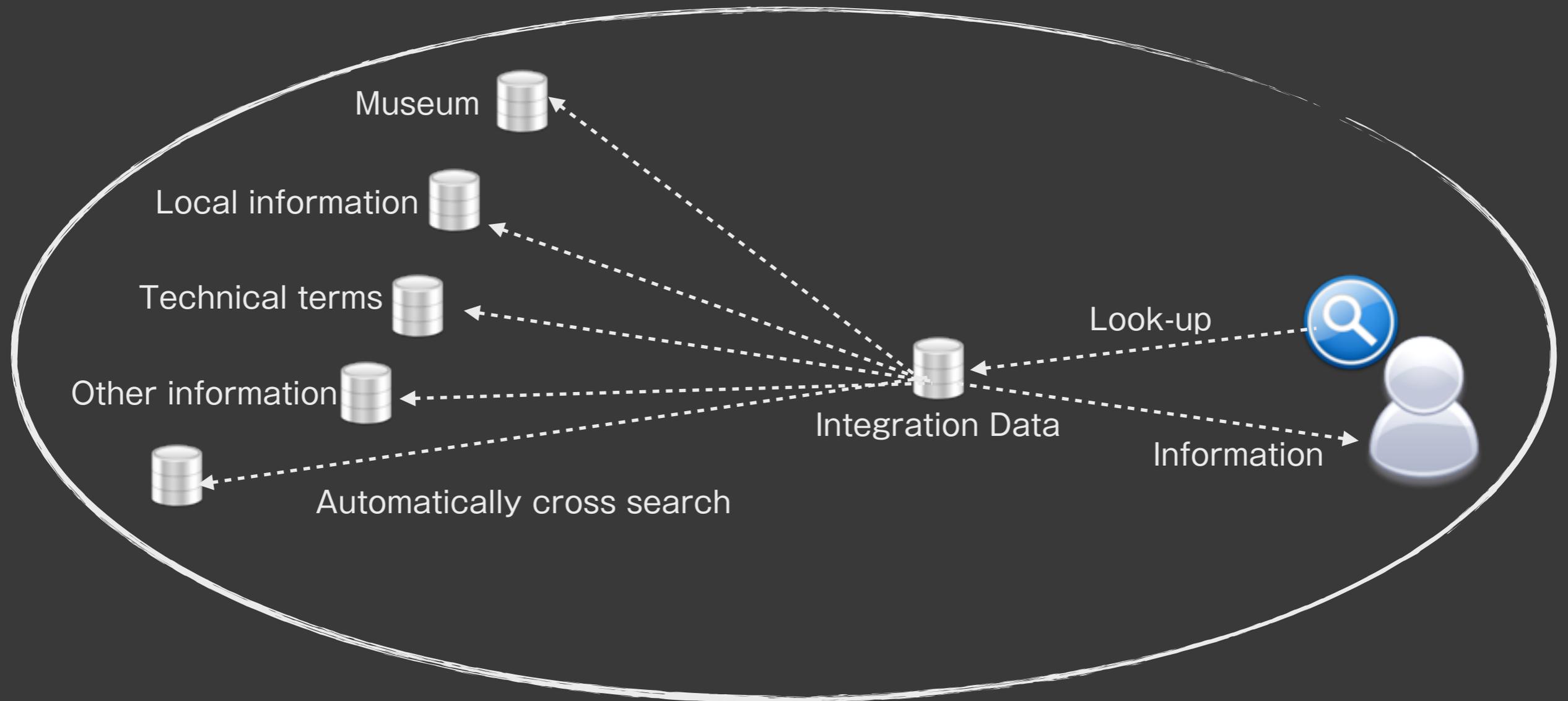
送信

リセット

### Options

Output Format

Show query URI



If, could be **search and use** the multiple information **on the Web....**

Home SPARQL Endpoint Location Apps Site Info Blog Wiki

**LODAC**

下村観山  
<http://lod.ac/id/359>

formats ▾

データ 参照1 参照2 参照3

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

タイプ (rdf:type)	<a href="#">foaf:Person</a>
ラベル (rdfs:label)	下村観山@ja
名前 (foaf:name)	下村観山@ja
参照 URI (dc:references)	<a href="#">dbpedia:下村観山</a>
参照 URI (dc:references)	<a href="#">lodacref:359</a>
ラベル (skos:prefLabel)	下村観山@ja
作品 URI (lodac:creates)	<a href="#">lodacid:20029</a>

ID-resource Tab

Link to an artwork ID-resource

Home SPARQL Endpoint Location Apps Site Info Blog Wiki

**LODAC**

下村観山  
<http://lod.ac/id/359>

formats ▾

参照1 参照2 参照3

URI: <http://lod.ac/ref/359>

タイプ (rdf:type)	<a href="#">lodac:PersonReference</a>
ラベル (rdfs:label)	下村観山@ja
出典元 (dc:source)	<a href="http://www.tulips.tsukuba.ac.jp/jart/mokuji/index.html">http://www.tulips.tsukuba.ac.jp/jart/mokuji/index.html</a>
名前 (foaf:name)	下村観山@ja
名前 (foaf:name)	SHIMOMURA, Kanzan@en
名前 (foaf:name)	しもむら・かんざん@ja-hrkt
ラベル (skos:prefLabel)	下村観山@ja

Ref resource Tab

# “Web of Data” thinking

- **Collect**    Aggregate data and metadata.
- **Use**        Use data for oneself .
- **Publish**    Publish through the web services.
- **Share**     Share information in society.



**New distribute information platform**

That's called...