



小倉百人一首LODと みんなde翻刻の連携

高橋 菜奈子(かるたLOD)
2021年2月15日 みんなde翻刻サミット

自己紹介

- 東北大学大学院文学研究科(国史学)修了
- 大学図書館員(東北大、新潟大、宮城教育大、一橋大)
 - 目録・機関リポジトリ・和漢古書のデジタルアーカイブ・ウェブサイトなどを担当
- 国立情報学研究所(2010.4～2016.3)
 - 学術情報流通(メタデータとオープンアクセスなど)の仕事
- 大学管理職(千葉大、東京学芸大学)
 - 現在は学校図書館連携にも取り組んでいる
- オープンデータに関わる活動
 - 2012～2014 NACSIS-CAT(大学図書館目録)のオープンデータ化
 - 2016～2020 LOD (Linked Open Data)チャレンジに毎年参加

LODチャレンジへの参加

- LODチャレンジとは
 - 2011年から実施されているLODのコンテスト
 - アイデア・アプリケーション・データセット・分析可視化・基盤技術を作品として応募
 - 研究者・技術者・行政職・図書館員・学生など様々な参加者
 - <https://2020.lodc.jp/>



LOD challenge

2016 郷土かるたLOD

- アイデア部門優秀賞を受賞
 - <http://idea.linkdata.org/idea/idea1s2085i>

2017 小倉百人一首LOD

- データセット部門最優秀賞を受賞
 - <http://idea.linkdata.org/idea/idea1s2398i>

2018 IIIF対応

- オープンサイエンス賞・LinkData賞を受賞
 - <http://idea.linkdata.org/idea/idea1s2623i>

2019 多言語対応

- 国際化LOD賞を受賞
 - <http://idea.linkdata.org/idea/idea1s2854i>

2020「みんなで翻刻de小倉百人一首LOD」

- プロGRESS賞を受賞
 - <http://idea.linkdata.org/idea/idea1s2928i>

かるたLODは、新しい技術動向を試してみる「遊び場」

本日の話題

- LODとは
- 小倉百人一首LODとは
- みんなで翻刻de小倉百人一首LOD
 - 制作過程
 - データセットの紹介
 - みんなで翻刻への期待



LODとは

Linked Open Dataとは

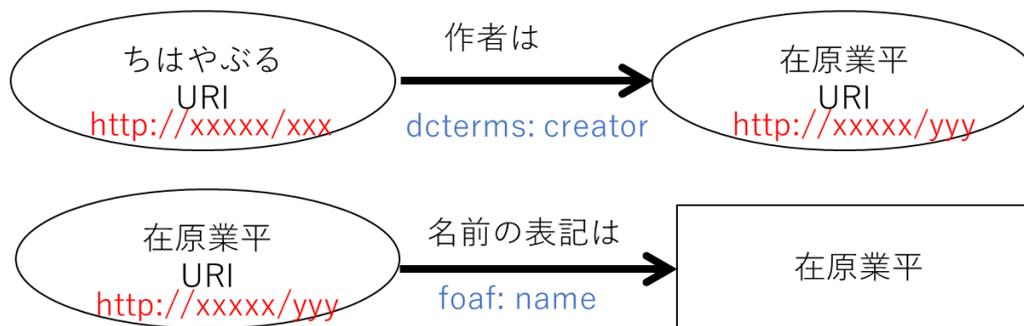
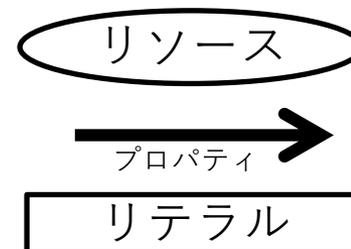
- オープンデータ
 - 自由な再利用可能なライセンスの明示されたデータ
 - 機械可読のデータ
- リンクトデータ
 - リソースとリソースがリンクされたデータ
- 5つの星
 - ***** LOD
 - **** RDF
 - *** 非独占
 - ** 構造化
 - * ライセンス



RDF (Resource Description Framework)

- 3つのトリプルで表現する

- リソースはURIを指定する必要がある
- プロパティで関係を表す(語彙を探すor独自定義)
- リテラルはURIを持たない文字列



- Linked Data ^

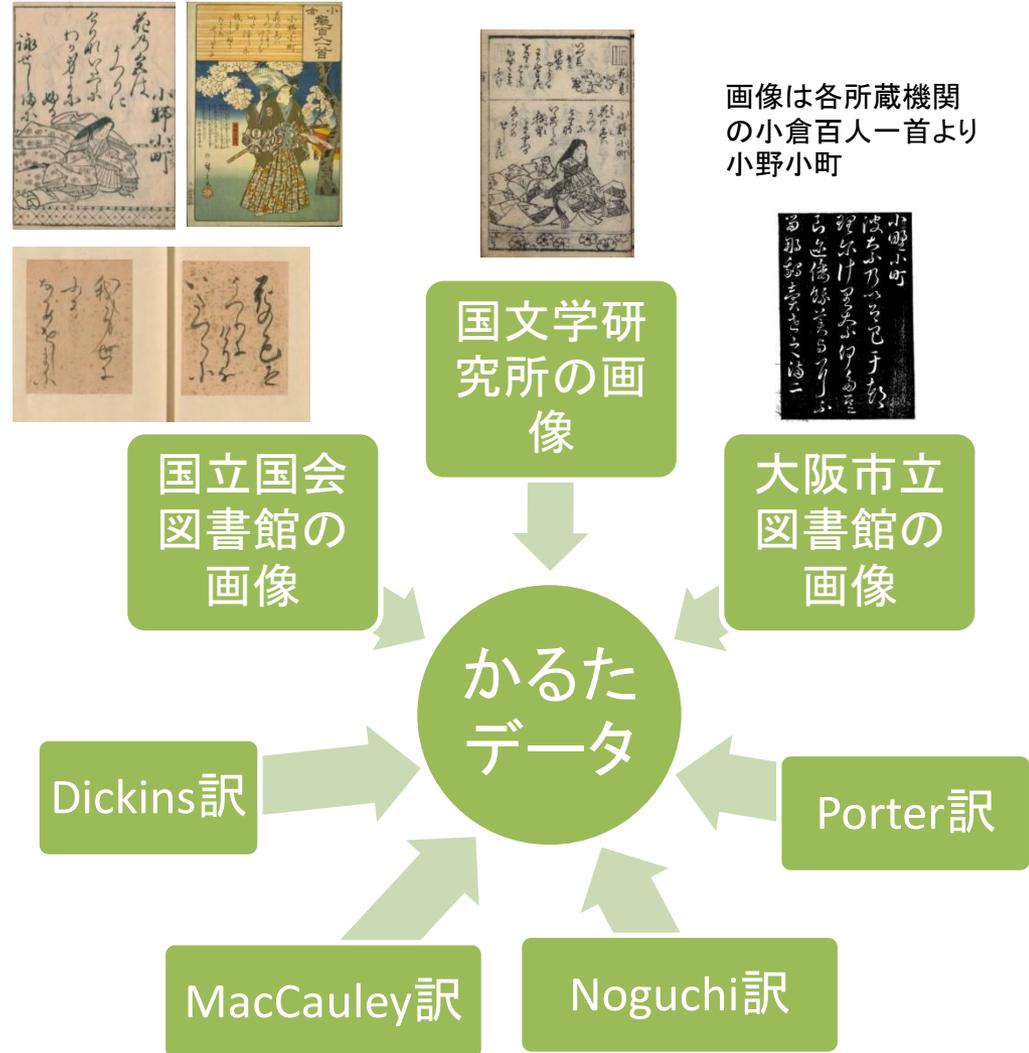




小倉百人一首LOD

小倉百人一首LODとは

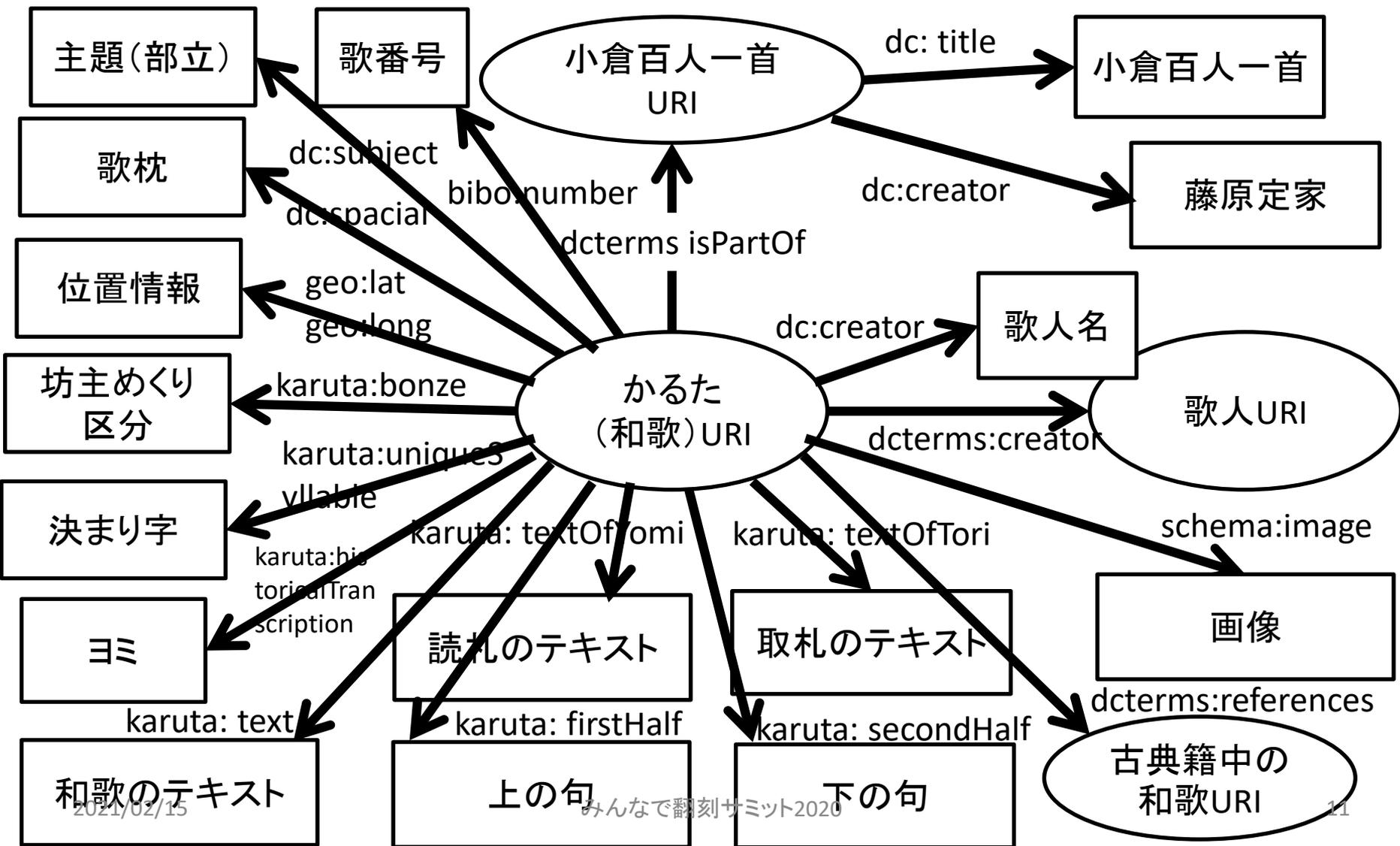
- かるたのLOD化
(モデルと語彙の設計)
 - 各地の図書館で公開された古典籍画像のオープンデータをリンク
 - 英語翻訳作品もリンク
- バラバラのオープンデータをつなぐ
- 情報の組織化・LOD化によって、活用の可能性を示す
- 文化情報資源のLODの普及



小倉百人一首かるたデータ

- 和歌をリソースとして作成
- ゲームに必要な情報を整理し、語彙を定義
 - 独自定義の語彙：読札のテキスト、取札のテキスト、読札の画像、取札の画像、決まり字、坊主めくり用の区分
- 和歌集としての必要な情報も収録
 - 独自定義の語彙：上の句、下の句
 - 歌枕として詠まれた地名と緯度・経度
 - 歌番号、出展の選集と部立て
- 正確な漢字・かな遣いの表記
 - 競技かるた用の読札に準拠
- 古典籍画像へのリンク
 - かるたデータの和歌から各古典籍の和歌にリンク

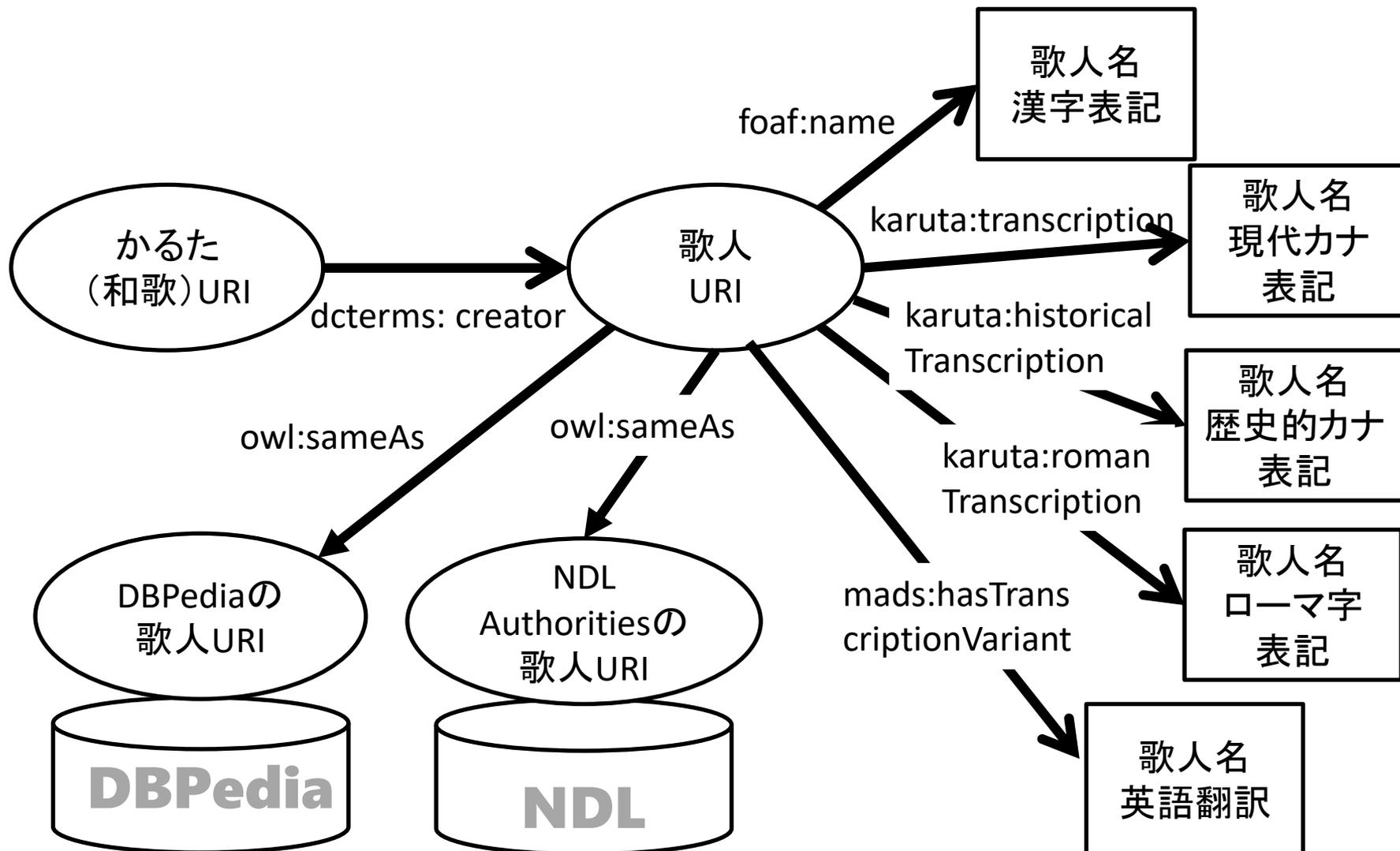
小倉百人一首かるたのグラフ



小倉百人一首歌人データ

- 歌人をリソースとして作成
- かるたデータからリンクする典拠の役割
- 基本情報として名前表記をデータ化
 - 漢字表記・歴史的かな遣い、現代かな遣い、ローマ字表記、英語翻訳名
- DBPedia、NDL Authoritiesにリンク
 - 人物の詳細な情報は外部リソースにゆだねる

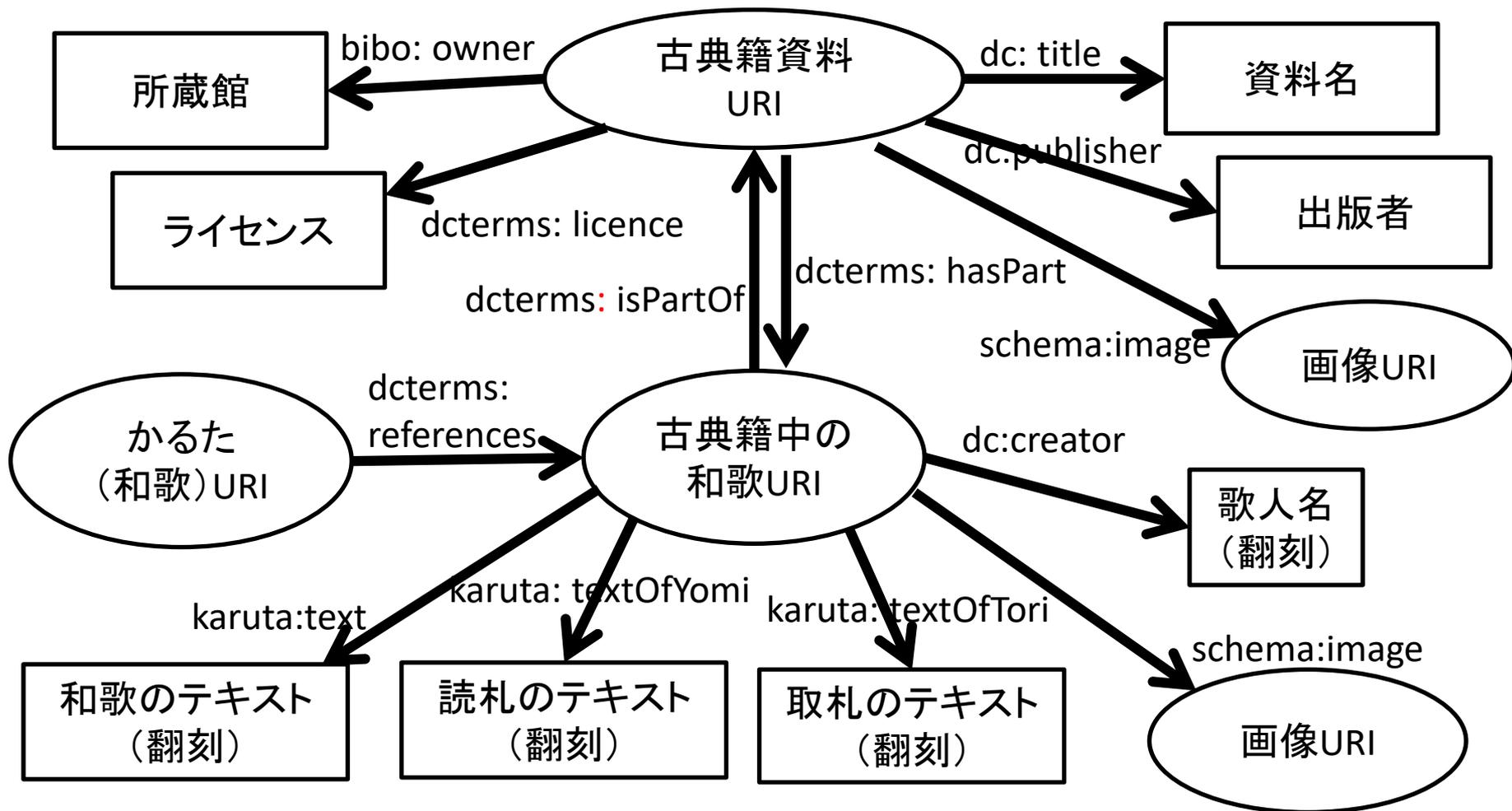
歌人のグラフ



古典籍画像データ

- オープンデータ画像が利用できる古典籍を資料単位でリソースとして作成
 - タイトルごとにリソースを設定
 - 書誌事項、所蔵館、ライセンスの情報を整理
 - 収録する和歌にリンク
- 各古典籍ごとに和歌をリソースとして作成
 - 現代かな遣い、歴史的かな遣いを忠実に翻刻
 - 公開されている画像へのリンク(DOI, URL)
 - かるたデータの和歌からのリンク

古典籍のグラフ



RDFの生成

- LinkData.orgのツールを使って、EXCELからRDFを生成
 - 表形式で作成すれば、Turtleに変換してくれる
 - 表にIdentifierを記入しておけば、URIを生成してくれる

Identifier	Karuta:text	dc:creator
ogura_017	ちはやぶる神代も聞かず...	在原業平朝臣
ogura_018	住の江の岸に寄る浪...	藤原敏行朝臣
ogura_019	難波潟短き葦のふしのまも...	伊勢



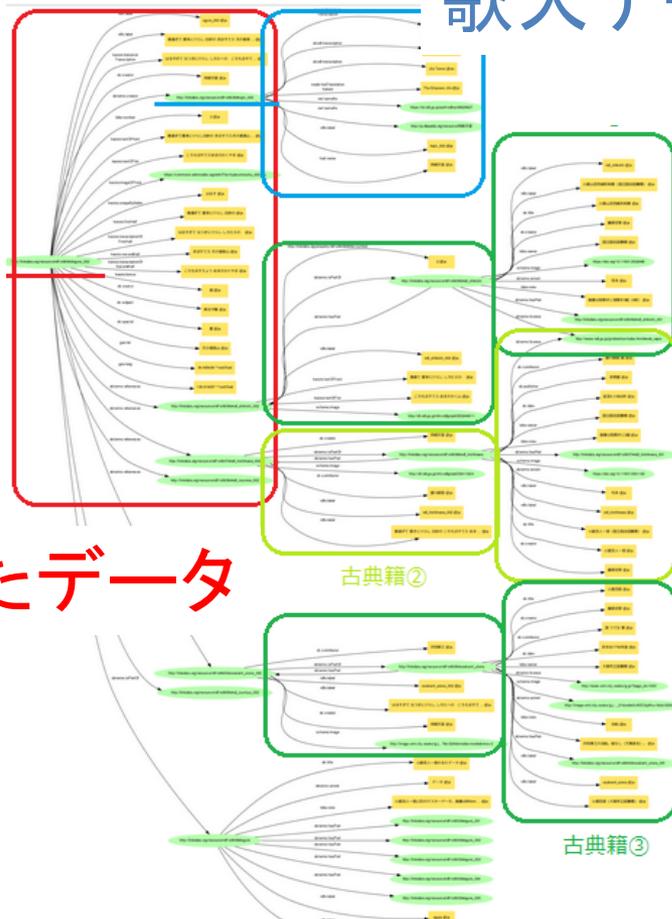
※ Open Refineを使うなど、他の手段でもLODは作れる。

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix dcterms: <http://purl.org/dc/elements/1.1/> .
@prefix karuta: <http://karutalod.web.fc2.com/vocabulary/> .

<http://linkdata.org/resource/rdf1s6834#ogura_017>
  karutas:text "ちはやぶる 神代も聞かず 龍田川 から紅に 水くくるとは"@ja ;
  dc:creator "在原業平朝臣"@ja ;
  dcterms:creator <http://linkdata.org/resource/rdf1s6833i#kajin_017> ;
  bibo:number "17"@ja ;
  karuta:textOfYomi "ちはやぶる神代も聞かず龍田川 から紅に水くくるとは"@ja ;
  karuta:textOfTori "からくれなゐにみつくとは"@ja ;
  karuta:imageOfYomi <https://commons.wikimedia.org/wiki/File:Hyakuninisshu_017.jpg> ;
  karuta:uniqueSyllable "ちは"@ja ;
  dc:source "古今集"@ja ;
  dc:subject "秋"@ja ;
```

小倉百人一首のデータモデル

歌人データ



公開済データ (LinkData.org)

<http://user.linkdata.org/user/tnanako/work>

かるたLODのウェブサイト

<http://karutalod.web.fc2.com/ogura.html>

Namespace

<https://github.com/tnanako/karutalod>

LODチャレンジ2017「小倉百人一首LOD」

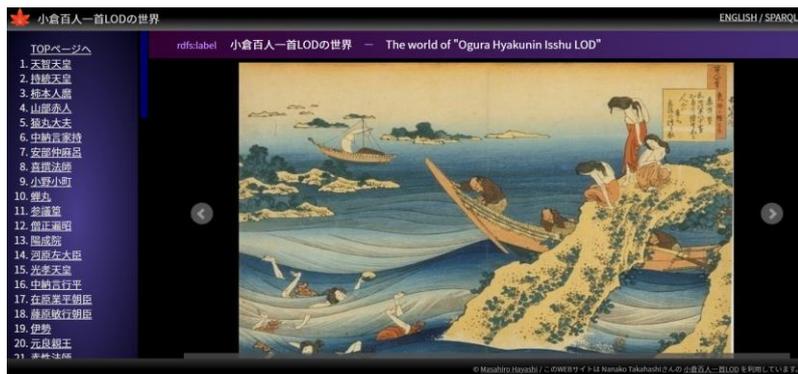
<http://idea.linkdata.org/idea/idea1s2398i>

かるたデータ

各古典籍のデータ

小倉百人一首LODのアプリ

- 小倉百人一首LODの世界 (by 林正洋氏)



- <https://www.mirko.jp/ogura/>
- SPARQL Endpointの設置

- 小倉百人一首クイズ (by uedayou)



- <https://nyakunin-issu.uedayou.net/>
- 合成音声により読み上げ

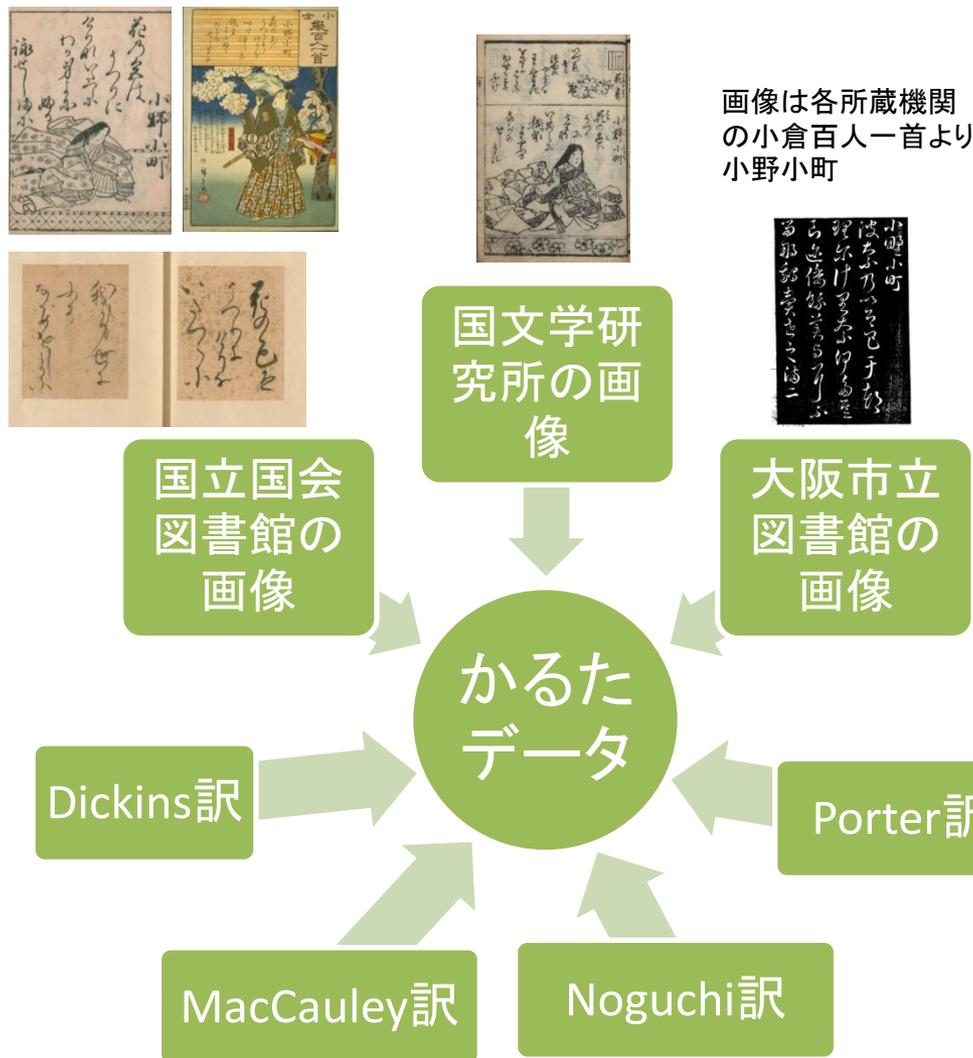
オープンデータを公開すると、誰かが利活用してくれる。
誰かがアプリケーションも作ってくれる(他力本願)!



みんなで翻刻DE小倉百人一首LOD

小倉百人一首LODの課題

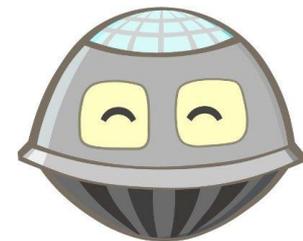
- かるたのLOD化(モデルと語彙の設計)
- 各地の図書館で公開された古典籍画像のオープンデータをリンク
- 英語翻訳作品もリンク



古典籍の翻刻に
時間がかかる点が課題
→「みんなで翻刻」との
協働へ

みんなで翻刻とは

- 市民参加型翻刻プラットフォーム
 - 多数の人々が翻刻に参加し、歴史資料の解読
 - 約7200人により1250万文字の史料が翻刻済(2020.12)
 - 2017年に京都大学古地震研究会によって、東京大学地震研究所が所蔵する地震関連史料の翻刻プロジェクトとして開始
- 2019年7月のリニューアル
 - 翻刻対象資料・対象機関の拡大
 - IIF (International Image Interoperability Framework)に対応した画像の取込みが可能に
 - AIによるくずし字解読機能の搭載
- <https://honkoku.org/>



連携の概要

①みんなで翻刻に
コレクションを登録



対象画像
リスト作成



IIIF画像の
取込み

②クラウドソーシングによる翻刻

画像は
国立国会図書館所蔵『百人一首絵抄』
小野小町

テキスト
整形



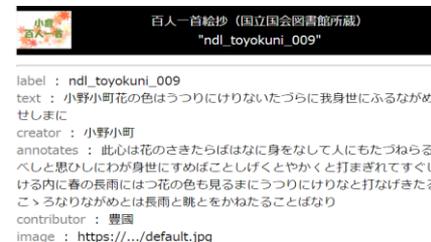
③翻刻データをキュレーション
(入カールの相違を吸収)

RDF生成



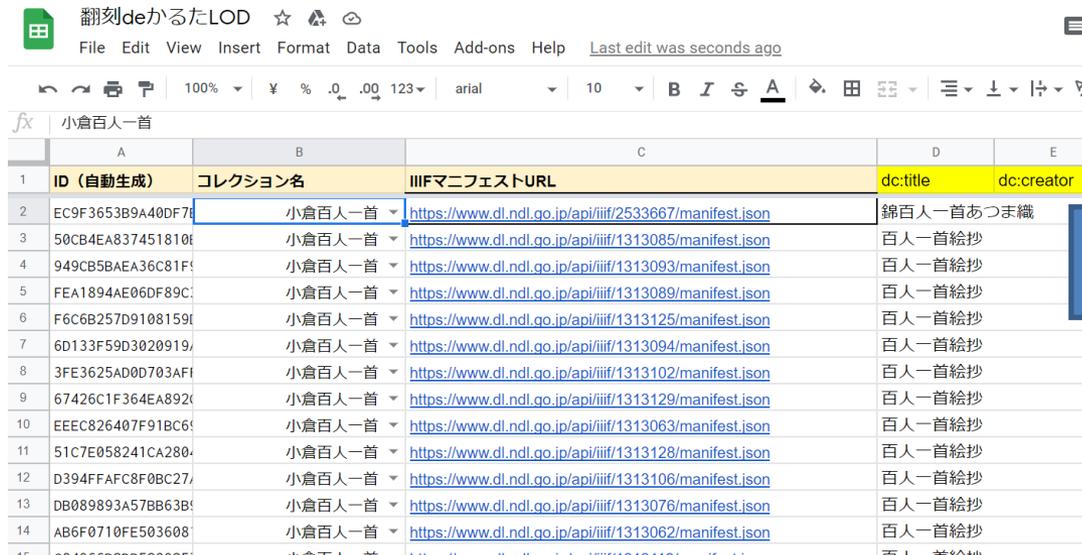
小倉百人
一首LOD

④小倉百人一首LODのデータ作成



①コレクションを登録

- 「翻刻deかるたLOD」プロジェクトの作成
- コレクション登録用のスプレッドシートにIIIFマニフェストURLを記入して登録依頼



ID (自動生成)	コレクション名	IIIFマニフェストURL	dc:title	dc:creator
EC9F3653B9A40DF7	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/2533667/manifest.json	錦百人一首あつま織	
50CB4EA837451810	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313085/manifest.json	百人一首絵抄	
949CB5BAEA36C81F	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313093/manifest.json	百人一首絵抄	
FEA1894AE06DF89C	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313089/manifest.json	百人一首絵抄	
F6C6B257D9108159	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313125/manifest.json	百人一首絵抄	
6D133F59D3020919	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313094/manifest.json	百人一首絵抄	
3FE3625AD0D703AF	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313102/manifest.json	百人一首絵抄	
67426C1F364EA892	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313129/manifest.json	百人一首絵抄	
EEEE826407F91BC6	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313063/manifest.json	百人一首絵抄	
51C7E058241CA280	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313128/manifest.json	百人一首絵抄	
D394FFAFC8F0BC27	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313106/manifest.json	百人一首絵抄	
DB089893A57BB63B	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313076/manifest.json	百人一首絵抄	
AB6F0710FE503608	小倉百人一首	https://www.dl.ndl.go.jp/api/iiif/1313062/manifest.json	百人一首絵抄	



プロジェクト まなぶ 案内 使い

翻刻deかるたLOD

古典籍に含まれる「かるた」を翻刻するプロジェクトです。翻刻したデータをもとにLOD (Linked Open Data) を作成し、公開します。

所蔵機関：国立国会図書館、国文学研究資料館、京都大学ほか

103/105点の翻刻が完了

② みんなで翻刻

- 2020/6/4登録直後から、次々と「翻刻完了」に

The image shows two overlapping screenshots. The top screenshot is a digital manuscript viewer for '百人一首絵抄 九 小野小町'. It displays the text of the poem and includes a '凡例' (Legend) section with red annotations: '【かるた上の句】' (top line) and '【かるた下の句】' (bottom line). The bottom screenshot is a social media post from '翻刻deかるたLOD' showing three book covers with red '翻刻完了' (Completed Transcription) stamps. The covers are for '百人一首絵抄 十六 中納言行平', '六十六番 前大僧正行尊', and '六十二番 清少納言'. Each cover also indicates '1/1コマが翻刻' (1/1 page transcribed).

- まとめてダンプデータを出力(2020/8/11)

③ 翻刻データのキュレーション

- みんなで翻刻の翻刻ガイドライン
「見たままを翻刻し、原文の姿をできるかぎり残す」



【左丁 本文】
 在原 竜(龍)田河
 業平 から紅
 朝臣 に
 千早振
 水くゝ 神代も
 ると
 は きかす

ダンプデータ

在原 竜(龍)田河 業平 から紅 朝臣 に 千早振水くゝ
 神代も ると は きかす

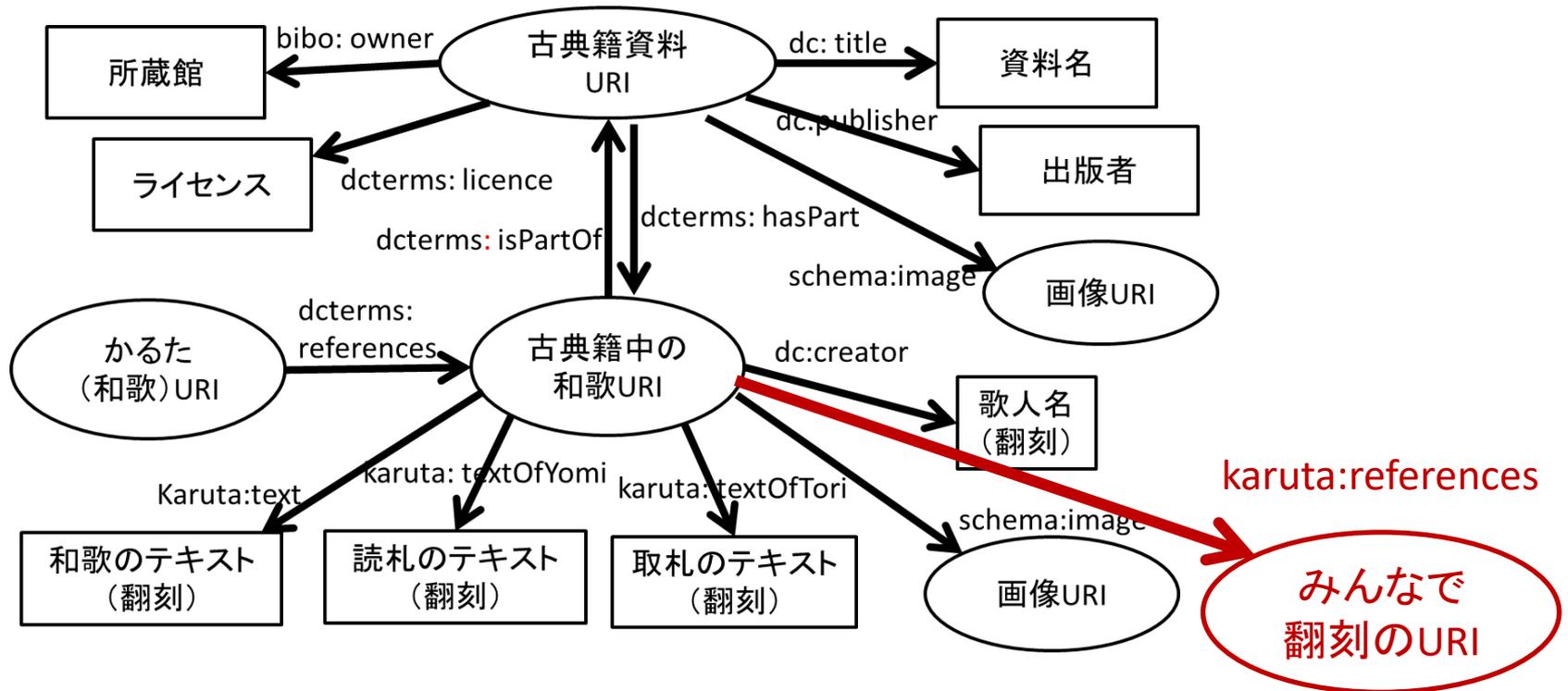


- キュレーションによって、テキストを整形
- 語順の調整・空白のトルツメ
- それ以外のルールは踏襲

千早振神代もきかす竜(龍)田河から紅に水くゝるとは

④小倉百人一首LODに登録

- 「原文の姿」を残すために、「みんなで翻刻」へのリンクを追加



作成したデータセット①

百人一首絵抄(国立国会図書館所蔵)

- <http://linkdata.org/work/rdf1s8473i>



みんなで翻刻

小倉百人一首LOD

みんな CSV, EXCEL, RDF(Turtle)がダウンロードできる

LinkData.orgのUI

小倉百人一首LOD

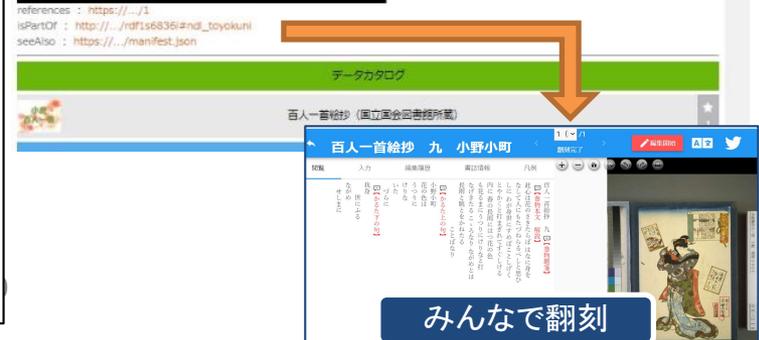
- EXCEL



- RDF

```
<http://linkdata.org/resource/rdf1s8473#ndl_toyokuni_009>
<http://www.w3.org/2000/01/rdf-schema#label> "ndl_toyokuni_009"@ja ;
karuta:text "花の色はうつりにけりないたづらに我身世にふるながめせしまに"@ja ;
dc:creator "小野小町"@ja ;
karuta:annotates "此心は花のさきたらばはなに身をなして人にもたづねらるべしと思ひしにわが身世にすめばことしげくとやかくと打まぎれてすぐしける内に春の長雨にはつ花の色も見るまにうつりにけりなと打なげきたるころなりながめとは長雨と眺とをかねたることばなり"@ja ;
dc:contributor "豊國"@ja ;
schema:image
<https://www.dl.ndl.go.jp/api/iiif/1313090/R0000001/full/full/0/default.jpg> ;
karuta:references
<https://honkoku.org/app/#/transcription/28A8A2168084A5CA030C516DE0E43A2C/1/> ;
dcterms:isPartOf <http://linkdata.org/resource/rdf1s6836#ndl_toyokuni> ;
rdfs:seeAlso <https://www.dl.ndl.go.jp/api/iiif/1313090/manifest.json> .
```

http://linkdata.org/resource/rdf1s8473#ndl_toyokuni_009

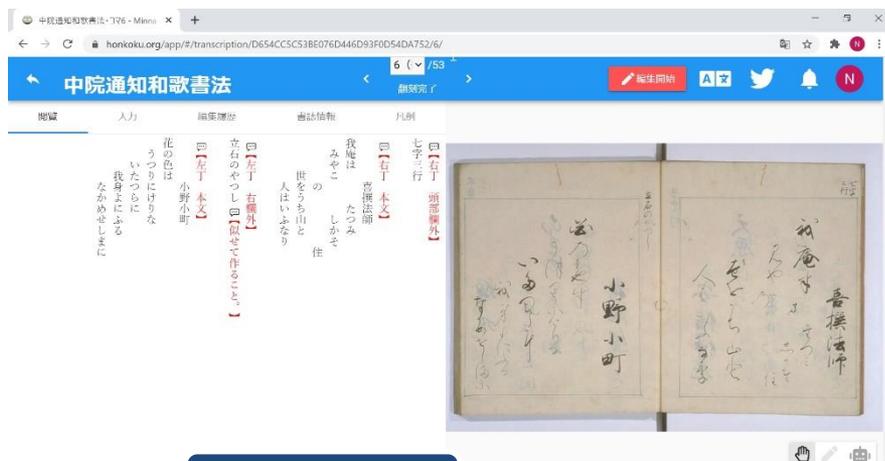


みんなて翻刻

作成したデータセット②

中院通知和歌書法(京都大学附属図書館所蔵)

- <http://linkdata.org/work/rdf1s8475i>

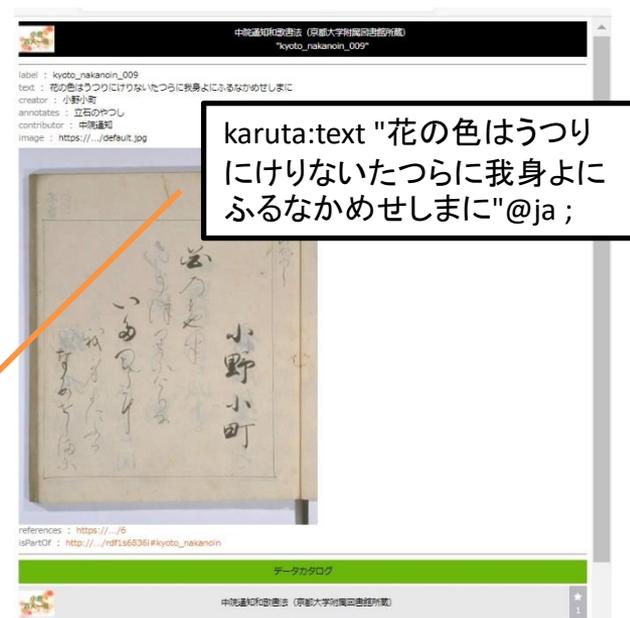


みんなで翻刻



小倉百人一首LOD

http://linkdata.org/resource/rdf1s8475i#kyoto_nakanoin_009



「かるたLOD」でデータキュレーション
語順の調整・空白のトルツメ

作成したデータセット③

錦百人一首あつま織 (国立国会図書館所蔵)

- <http://linkdata.org/work/rdf1s8472i>



http://linkdata.org/resource/rdf1s8472i#ndl_azuma_009



小倉百人一首LOD

```
schema:image
<https://www.dl.ndl.go.jp/api/iiif/2533667/R0000015/500,1000,2900,4100/full/0/default.jpg> ;
```

「かるたLOD」でデータキュレーション
IIIFを使って、特定のかるとの画像を切出し

作成したデータセット④

寿賀多百人一首小倉錦 (国文学研究資料館所蔵)

- <http://linkdata.org/work/rdf1s8474i>

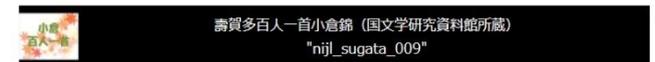


みんなで翻刻

小倉百人一首LOD

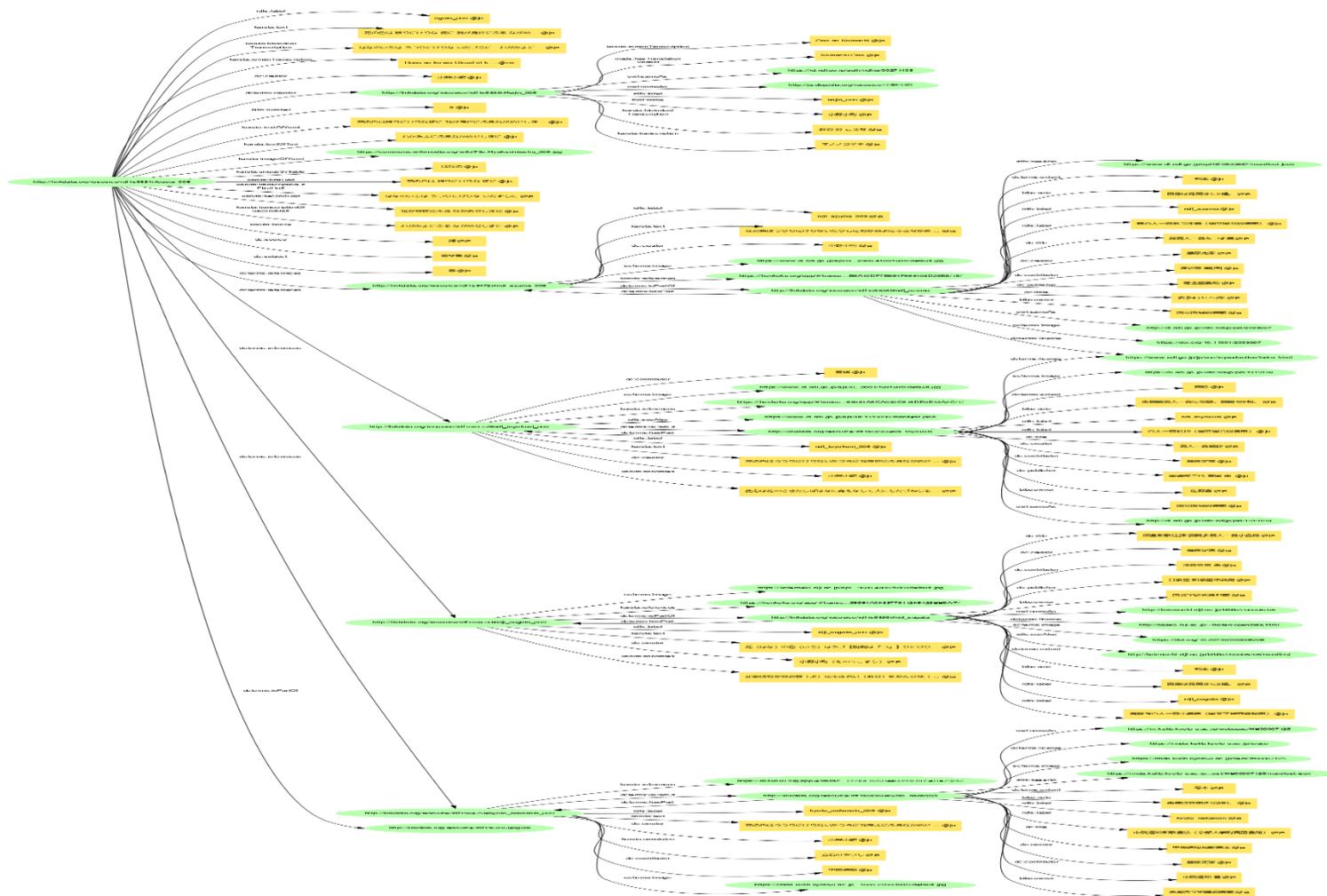
「みんなで翻刻」のおかげで、注釈 (karuta:annotetes) のデータを追加

http://linkdata.org/resource/rdf1s8474i#nijl_sugata_009



label : nijl_sugata_009
text : 花 (はな) の色 (いろ) はうづ【別刷は「つ】りにけりないたづらに我身 (わがみ) よにふるながめせしみに
creator : 小野小町 (をのゝこまち)
annotates : 此哥はわか身の世 (よ) にふる老い (おひ) をかんじ花 (はな) によそへて思ひをいひのへたり小町が古今にて第 (たい) 一の歌とすかの集に後 (のち) に入れたる哥也此哥にうらおもての説 (せつ) あり表 (おもて) は花のさきたらば花に身 (み) をなさんと思ひしも世 (よ) にすめばことしけくてやかくやうちまぎれ過たる花も見ぬ花なれうつりにけりなど察 (さつ) していふ也義 (うら) の説 (せつ) は身 (み) のおとろへもわれとはしらぬものなり只 (たゞ) 花のおとろへを見てわが身もかくうつりてこそあらめとおもひやる義 (ぎ) 也
image : https://.../default.jpg

みんなで翻刻で作成した 「小野小町」データセット



みんなで翻刻への期待

- ① コレクションの中の配列
 - 小倉百人一首には順番がある
 - メタデータで解決すればよいのかもしれない。
- ② 「翻刻完了」≠校閲完了
 - クラウドソーシングの本質的な課題
 - ポリシーの問題なので、利用側が校閲すればよい。
- ③ ダンプデータを簡単に入手できる仕組み
 - ダウンロードAPIはあるが、翻刻データの活用には不可欠
 - と考えていたら、できてました！
 - <https://www.dheq.org/honkoku/karuta.html>

かるたを翻刻してくださった
「みんな」に感謝！

<https://honkoku.org/app/#/projects/karuta/info>

引き続き、翻刻をよろしくお願いします！

おまけ

- 東京学芸大学「学びと遊びの歴史」を翻刻！
プロジェクト開始

<https://honkoku.org/app/#/projects/gakugei/info>

