

Wikidata と連携する Linked Data の構築

亀田 堯宙

2021年1月15日 NDLデジカフェ



大学共同利用機関法人 人間文化研究機構

国立歴史民俗博物館

National Museum of Japanese History



国立歴史民俗博物館

総合資料学の創成

INTEGRATED STUDIES OF CULTURAL AND RESEARCH RESOURCES
National Museum of Japanese History

<http://www.metaresource.jp/>



khirin

Knowledgebase of
Historical Resources in Institutes

今日のお話

自己紹介

国立歴史民俗博物館 特任助教 亀田 堯宙
専門は情報学（**Linked Data**、自然言語処理）
→人文系を中心とした知識の共有に応用

- 知識の欠片をどうやって共有してつなぎ合わせるか
- **Wikipedia → Wikidata** / 文書のウェブ → データのウェブ(**Linked Data**)
- 研究データと**Wikidata** ～信頼性・永続性・協働～
- 研究**Linked Data**とのつなげ方（歴博の**khirin**を例として）
- **Wikidata**の編集の仕方
- **2021年**の共有知

今日のお話

自己紹介

国立歴史民俗博物館 特任助教 亀田 堯宙
専門は情報学（**Linked Data**、自然言語処理）
→人文系を中心とした知識の共有に応用

- 知識の欠片をどうやって共有してつなぎ合わせるか
- **Wikipedia → Wikidata / 文書のウェブ → データのウェブ(Linked Data)**
- 研究データとWikidata ～信頼性・永続性・協働～
- 研究Linked Dataとのつなげ方（歴博のkhirinを例として）
- Wikidataの編集の仕方
- 2021年の共有知

Douglas Adams

From Wikipedia, the free encyclopedia

For other people named Douglas Adams, see [Douglas Adams \(disambiguation\)](#).

Douglas Noel Adams (11 March 1952 – 11 May 2001) was an English author, screenwriter, essayist, humorist, satirist and dramatist. Adams was author of *The Hitchhiker's Guide to the Galaxy*, which originated in 1978 as a BBC radio comedy before developing into a "trilogy" of five books that sold more than 15 million copies in his lifetime and generated a television series, several stage plays, comics, a video game, and in 2005 a feature film. Adams's contribution to UK radio is commemorated in The Radio Academy's Hall of Fame.^[2]

Adams also wrote *Dirk Gently's Holistic Detective Agency* (1987) and *The Long Dark Tea-Time of the Soul* (1988), and co-wrote *The Meaning of Liff* (1983), *The Deeper Meaning of Liff* (1990), and *Last Chance to See* (1990). He wrote two stories for the television series *Doctor Who*, co-wrote *City of Death*, and served as script editor for its seventeenth season in 1979. He co-wrote the Monty Python sketch "Patient Abuse" which appeared in the final episode of *Monty Python's Flying Circus*. A posthumous collection of his selected works, including the first publication of his final (unfinished) novel, was published as *The Salmon of Doubt* in 2002.

Adams was an advocate for environmentalism and conservation, a lover of fast cars,^[3] technological innovation and the Apple Macintosh, and a self-proclaimed "radical atheist".

Contents [hide]

- 1 Early life
- 1.1 Education
- 2 Career
- 2.1 Writing
- 2.1.1 *Doctor Who*
- 2.1.2 *The Hitchhiker's Guide to the Galaxy*
- 2.1.3 *Dirk Gently* series
- 2.2 Music
- 2.3 Film

Douglas Adams



Born	Douglas Noel Adams <div>11 March 1952</div> <div>Cambridge, Cambridgeshire, England</div>
Died	11 May 2001 (aged 49) <div>Montecito, California, US</div>
Resting place	Highgate Cemetery, London, England
Occupation	Writer
Alma mater	St John's College, Cambridge
Genre	Science fiction, comedy, satire
Notable	<i>The Hitchhiker's Guide to the</i>

Wikidata

- 「構造化」
- 基本は

主語

Douglas Adams

述語

Educated at

目的語

St John's College

全部にIDを振る

The image shows a Wikidata item page for **Douglas Adams** (Q42). The page is annotated with labels for its various components:

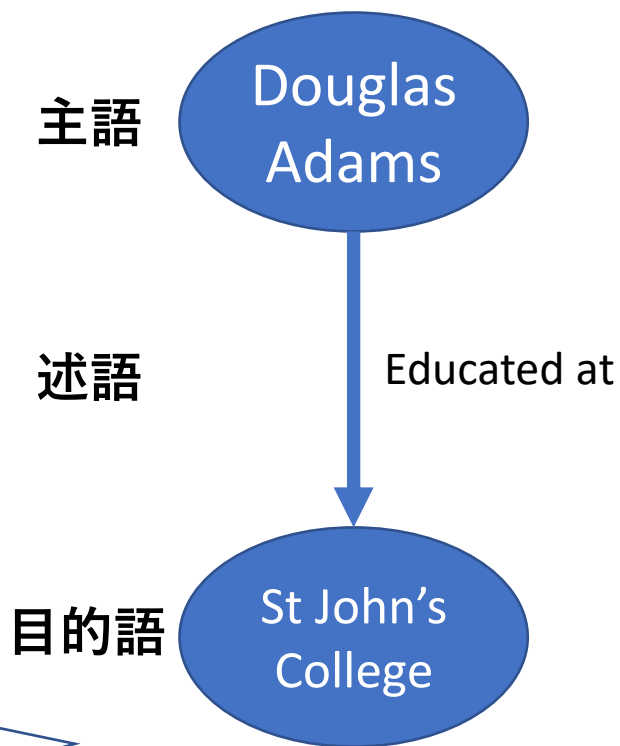
- label**: Points to the main title "Douglas Adams".
- item identifier**: Points to the ID "(Q42)".
- description**: Points to the text "English writer and humorist" and the aliases "Douglas Noël Adams | Douglas Noel Adams".
- aliases**: Points to the text "Douglas Noël Adams | Douglas Noel Adams".
- property**: Points to the "educated at" statement.
- value**: Points to the value "St John's College".
- qualifiers**: Points to the table of qualifiers for the "educated at" statement, including "end time", "academic major", "academic degree", and "start time".
- rank**: Points to the "rank" column in the statements table.
- statement group**: Points to the "statement group" column in the statements table.
- opened references**: Points to the "2 references" section, which is expanded to show details like "stated in", "reference URL", "original language of work", "retrieved", "publisher", and "title".
- collapsed reference**: Points to the "0 references" section, which is collapsed.

The "educated at" statement is highlighted with a green border, and the "2 references" section is highlighted with a pink border. The "0 references" section is also highlighted with a pink border.

文書のウェブからデータのウェブへ


- 小風さんや橋本さんの対象とされているのは主に「文書」
 - たまに他の文書と共通していたり比較可能だったりする語彙や値は登場するが、地の文がメイン
- **Linked Data** の世界では、一文の意味内容すら3つの要素に刻まれる
RDF: 右のような知識の表現方法
Linked Data: それぞれのHTTP URI = URL で人間が読めるHTMLと機械が読めるRDF/XML や JSON-LD などで情報を提供する枠組み
Wikidata: Linked Data の1つ

全部にURIを振る



集計や可視化が簡単に！









SPARQLという検索式を書くと、条件を付けた検索ができる

 Wikidata Query Service


例

ヘルプ

その他のツール



```
1 #defaultView:Map
2 SELECT *
3 WHERE
4 {
5   ?item rdfs:label ?label;
6     wdt:P463 wd:Q104447152;
7     wdt:P6375 ?o6375.
8   OPTIONAL { ?item wdt:P625 ?geo }
9   FILTER (lang(?label) = 'ja')
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "ja,[AUTO_LANGUAGE]". }
11 }
```



今日のお話

自己紹介

国立歴史民俗博物館 特任助教 亀田 堯宙
専門は情報学（**Linked Data**、自然言語処理）
→人文系を中心とした知識の共有に応用

- 知識の欠片をどうやって共有してつなぎ合わせるか
- **Wikipedia → Wikidata** / 文書のウェブ → データのウェブ(Linked Data)
- **研究データとWikidata ～信頼性・永続性・協働～**
- **研究Linked Dataとのつなげ方**（歴博のkhirinを例として）
- **Wikidataの編集の仕方**
- **2021年の共有知**

総合資料学 情報基盤システム khirin

<https://khirin-lid.rekihaku.ac.jp>



二又鉄鉾 (23347162)



タイトル	二又鉄鉾 (23347162)
タイプ	国立歴史民俗博物館・歴史民俗調査カード (考古)
コレクション・シリーズ	国立歴史民俗博物館・歴史民俗調査カード (考古) セット 歴史カード・考古・公開用(2件) から探す
ライセンス	https://creativecommons.org/licenses/by-nc-sa/4.0/
記述対象	二又鉄鉾
タイプ	歴史カード (考古) 23347162

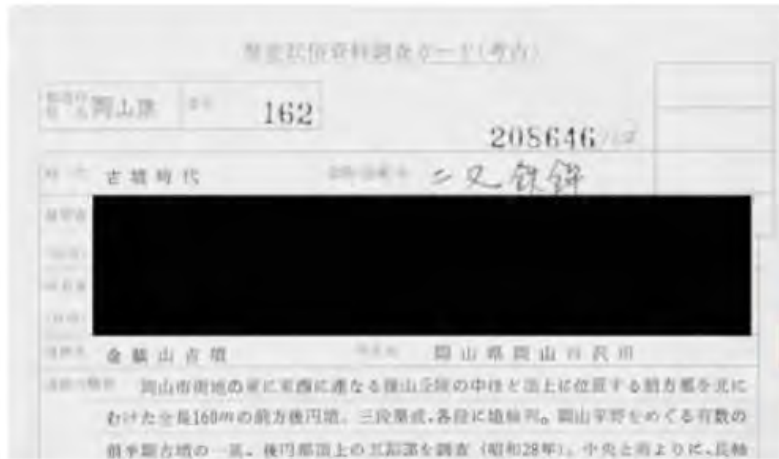
発見地名	金蔵山古墳
遺跡の調査者	(財) 倉敷考古館 (西谷真治、鎌木義昌)
遺跡の説明	岡山市街地の東に東西に連なる鑊山丘陵の中ほど頂上に位置する前方部を北にもけた三段築成、各段に短輪列。岡山平野をめぐる有数の前半期古墳の一基。後円部頂上の年。中央と南より、長軸に直交する板石積造穴式石室（中央石室、南石室）あり。方形地輪列が確認された。中央石室東側に副葬品を入れた小竪穴式石室（副石室）発見され、鉄器多数がその中に入っていた。
遺跡の位置	岡山市浜田岡山島岡山市浜田
記述対象のタイプ名	国立歴史民俗博物館・歴史民俗調査カード (考古) 記述対象
記述対象の所蔵者名	(財) 倉敷考古館
記述対象の所蔵者の所在地	岡山県倉敷市中央一丁目3-13

目的の1つ：地域文化資源所在情報DB



地域文化資源のLinked Data 化の例

二又鉄鉾が発見された金蔵山古墳



歴民カード
(例: カード番号 23347162 のカード)

歴民カードに書かれている二又鉄鉾



地理情報



二又鉄鉾を所有・所蔵している
(財)倉敷考古館

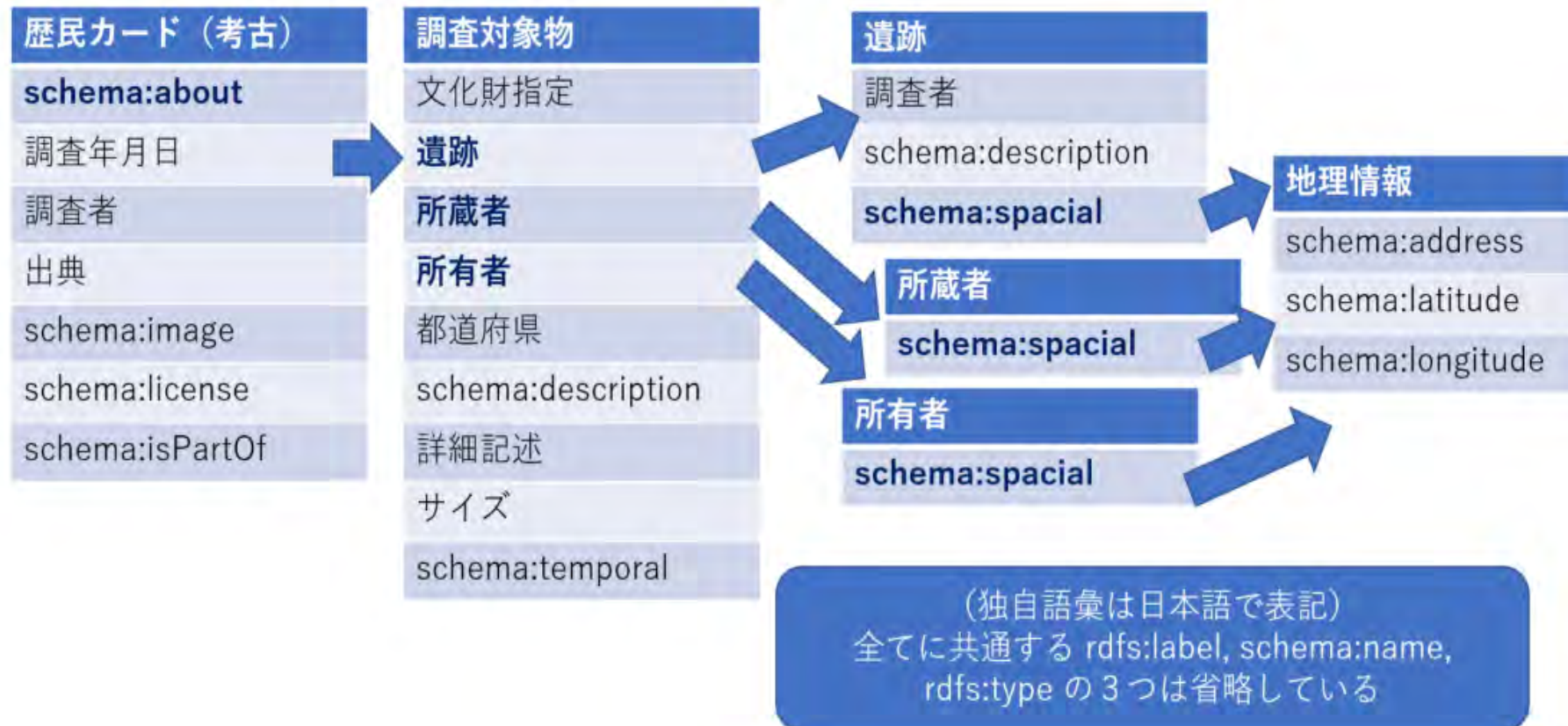


地理情報



Linked Data 化の例

- Schema.org を中心にして設計



Linked Data 化の例

長期保存

資料

個別概念

まとめや相関の把握

共通概念

歴民カード（考古）

schema:about

調査年月日

調査者

調査対象物

文化財指定

遺跡

所蔵者

遺跡

調査者

schema:description

schema:spacial

地理情報

schema:address

schema:latitude

schema:longitude

所蔵者

schema:spacial

所有者

schema:spacial

所有者

府県

schema:description

記述

ズ

schema:temporal

都道府県	岡山県	番号	162
調査番号	208646	162	
時代	古墳時代	遺物(遺構)名	二又鉄鉋
保管者			
住所			
所有者			
住所			
遺跡名	金蔵山古墳	所在地	岡山県岡山市沢田
遺跡の概要 岡山市街地の東に東西に連なる操山丘陵の中ほど頂上に位置する前方部を北にむけた全長160mの前方後円墳。三段築成、各段に埴輪列。岡山平野をめぐる有数の前半期古墳の一基、後円部頂上の主体部を調査（昭和28年）。中央と南よりに、長軸に直交する板石積竪穴式石室（中央石室、南石室）あり、それぞれの石室をめぐる方形埴輪列が確認された。中央石室東側に副葬品を入れた小竪穴式石室（副石室）があり、埴輪質合子4個が発見され、鉄器多数がその中に入っていた。			
発掘調査者	（財）倉敷考古館	（西谷真治、鎌木義昌）	
参考文献	西谷真治、鎌木義昌「金蔵山古墳」	（財）倉敷考古館	
調査日	昭和34年8月30日		

情報管理の協働
(モデル化の負担の低減、
「共通」性の担保)

Linked Data 化の例

WIKIDATA

国立歴史民俗博物館 (Q3330690)

千葉県佐倉市にある日本の歴史、民俗学、考古学について総合的に研究・展示する歴史博物館
国立歴史民俗博物館 | 人間文化研究機構国立歴史民俗博物館 | れきはく

▼ 他言語の表示

言語	ラベル	説明	別名
日本語	国立歴史民俗博物館	千葉県佐倉市にある日本の歴史、民俗学、考古学について総合的に研究・展示する歴史博物館	国立歴史民俗博物館 人間文化研究機構国立歴史民俗博物館 れきはく
英語	National Museum of Japanese History	museum in Japan	Kokuritsu Rekishi Minzoku Haku...
韓国語	국립역사민속박물관	説明はありません	Rekihaku

入力済みの全言語

文

分類

博物館

▼ 0件の情報源

+ 情報源を追加

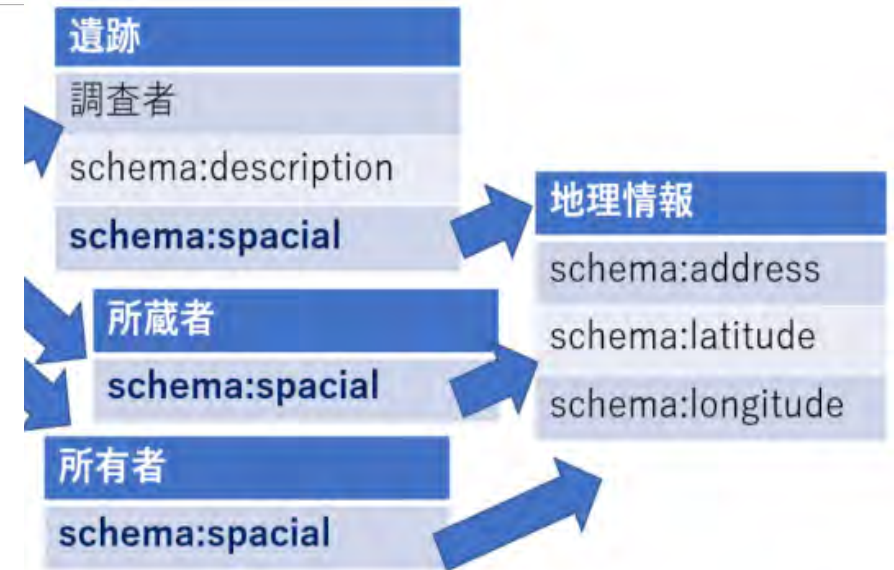
+ 追加

画像

研究データを横断する概念にWikidataを使う

まとめや関連の把握

共通概念



研究データを横断する
概念にWikidata を使う

[+ 追加](#)

同等のプロパティ

<https://schema.org/geo>[編集](#)

▼ 0件の情報源

[+ 情報源を追加](#)<http://www.w3.org/2006/vcard/ns#geo>[編集](#)

▼ 0件の情報源

[+ 情報源を追加](#)<https://geogratis.gc.ca/services/geolocation/en/Point>[編集](#)

▼ 0件の情報源

[+ 情報源を追加](#)[+ 追加](#)

外部下位プロパティ

<https://schema.org/latitude>[編集](#)

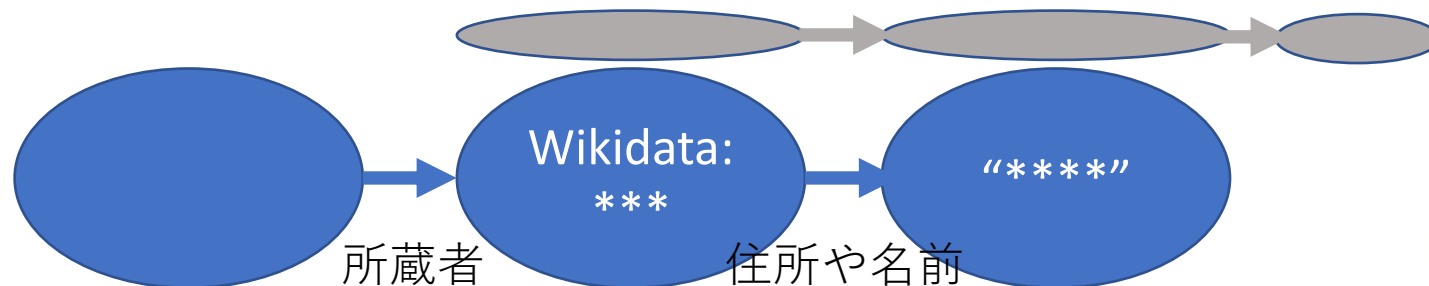
▼ 0件の情報源

<https://schema.org/longitude>[編集](#)

Wikidata Property と
Schema.org 等是对应がある

信頼性と永続性

- Wikipedia と同様、Wikidata も信頼性の問題を抱えている。研究データ側が情報源になることで、信頼性に寄与／他のデータで使える
- Wikidata 側の変更などに備えて、「のりしろ」を書くのが良い（Wikidata にはつなげるが、その先の具体的な値も研究データ側で保持する）



位置する行政区画

佐倉市

▼ 1件の情報源

GRID ID grid.471895.6

位置座標



35°43'28.19"N, 140°13'8.69"E

▼ 1件の情報源

移入元のウィキメディア・プロジェクト 英語版ウィキペディア

所属グループ

全国歴史民俗系博物館協議会

▼ 1件の情報源

出典URL <https://rekimin.com/pla>

Wikidata の編集の仕方

- 手動と自動（バッチ）がある
- Excel などでもコマンドを準備して投稿が可能
- 新規項目の重複チェックは OpenRefine および目視



はじめは手動で編集して経験を積む
& バッチの投稿資格を得る必要がある

さいごに～2021年の共有知～

- 「つながる」データから「つなげた」データへ
- そのハブとしてのWikidata
 - 組織や人の情報
 - 古い書誌の情報
 - 論文の情報→OpenCitation
- 「のりしろ」のあるデータで、信頼性と永続性にも気配りを
- みんなでWikidataを使いあいましょう！