

DH ・ 技術要素 ・ RDF / Linked Data

RDFとリンクトデータ入門

データを線でつなぐ

DH入門 / 技術要素シリーズ

中村

※実験的な取り組みです（構成・図・AI音声合成を含む）。内容をご確認・ご注意のうえご利用ください

この動画について

- ✓ **クリエイティブ・コモンズ**のオープン教材を参照し、独自に構成した解説です
- ✓ スライド・図は新規作成、ナレーションは**本人声のAI音声合成**
- ✓ これは**実験的な取り組み**です。内容は**ご確認・ご注意のうえ**ご利用ください
- ✓ 誤りに気づいたら概要欄からご指摘ください。出典・ライセンスは末尾と概要欄に記載しています

この回のゴール

データを「つながる形」で表す考え方をつかむ

- ✓ RDFの**三つ組（主語・述語・目的語）**でデータを表す考え方を説明できる
- ✓ **URI** がなぜ必要かを説明できる
- ✓ データを「つなぐ」 = **リンクトデータ**の発想を説明できる
- ✓ **スパークル**で問い合わせる流れのイメージを持てる

前提知識は特にありません。データを「つなぐ」考え方を、一から見ていきます。

今日の流れ

- ✓ データを「つなぐ」とは（三つ組）
- ✓ 同じものを指す — URI
- ✓ つないで広げる — リンクトデータ
- ✓ 問い合わせる — スパークル
- ✓ 始め方・学ぶには

データを「つなぐ」とは

まずは RDF の三つ組から

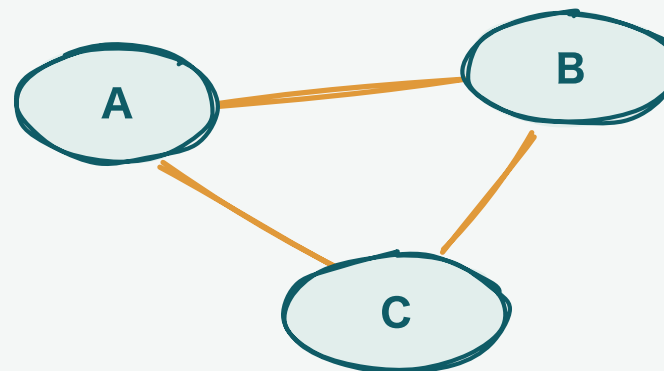
表のデータは「関係」が見えにくい

表：行と列



つながりが見えない

グラフ：線でつながる



行と列に収めると、**項目どうしのつながり**が表に残りにくい

書きたいのは「関係」

- ✓ 知りたいのは、しばしば「AとBが**どう関係するか**」
- ✓ 「この人物は、この場所で生まれた」「この作品は、この人が書いた」
- ✓ そこで、データを**関係のかたまり**として書く方法が要る
- ✓ それが **RDF** (アールディーエフ) の考え方

RDF = 主語・述語・目的語の三つ組



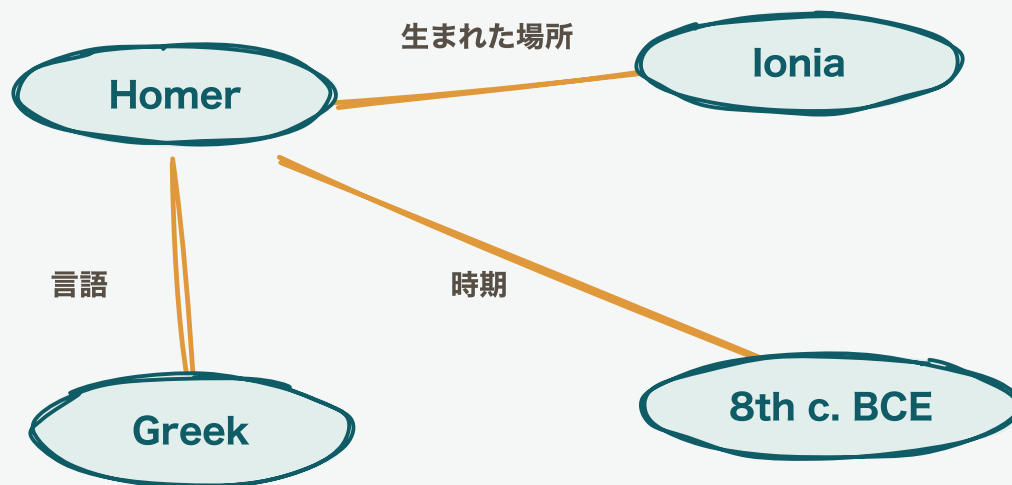
ひとつの「関係」を、**主語・述語・目的語**の三つ組で書く

たとえば、こう書く



「**ホメロス** — 生まれた場所 — **イオニア**」のように書きます

つなぐと「グラフ（網）」になる



三つ組がいくつもつながって「網」になる

三つ組を重ねると、点（もの）と線（関係）の網ができていく

ここまでのポイント

- ✓ データの中身は、しばしば「もの」と「もの」の**関係**
- ✓ RDFは、その関係を**主語・述語・目的語の三つ組**で書く
- ✓ 三つ組を重ねると、つながった**グラフ**になる

でも「ホメロス」と書くだけで、世界の誰もが同じ人を思い浮かべる？ → URI へ

同じものを指す — URI

言葉は曖昧

- ✓ 「ホメロス」と書いても、**叙事詩人**か、店の名前か、はたまた別人か
- ✓ 「金沢」も、地名なのか、人の姓なのか
- ✓ 人間は文脈で察するが、データをつなぐには、それでは困る
- ✓ **「どれを指すか」**を取り違えない名前が要る

URI=世界共通の「名札」

世界共通の名札 (URI)

Homer

<http://example.org/id/Homer>

どの「ホメロス」かを、一意に決められる

ユースールアイで、どの「ホメロス」かを世界で一意に決められる

三つ組を URI で書く

```
<.../id/Homer> <.../birthPlace> <.../place/Ionia> .
```

主語・述語・目的語を、それぞれ **URI** で指す（中身は「ホメロス／生地／イオニア」）

述語も「共通の語彙」を使う



同じ述語（共通語彙）だから、かみ合う

よく使う述語は、schema.org や [ダブリンコア](https://www.dublincore.org/) など既存の語彙を借りる

ここまでのポイント

- ✓ 名前の曖昧さを避けるため、ものに **URI** という名札をつける
- ✓ 主語・述語・目的語を URI で書けば、取り違えが起きにくい
- ✓ 述語も **共通の語彙** を使うと、別のデータと**かみ合う**

URIと共通語彙がそろると、いよいよ「つなぐ」ことができる

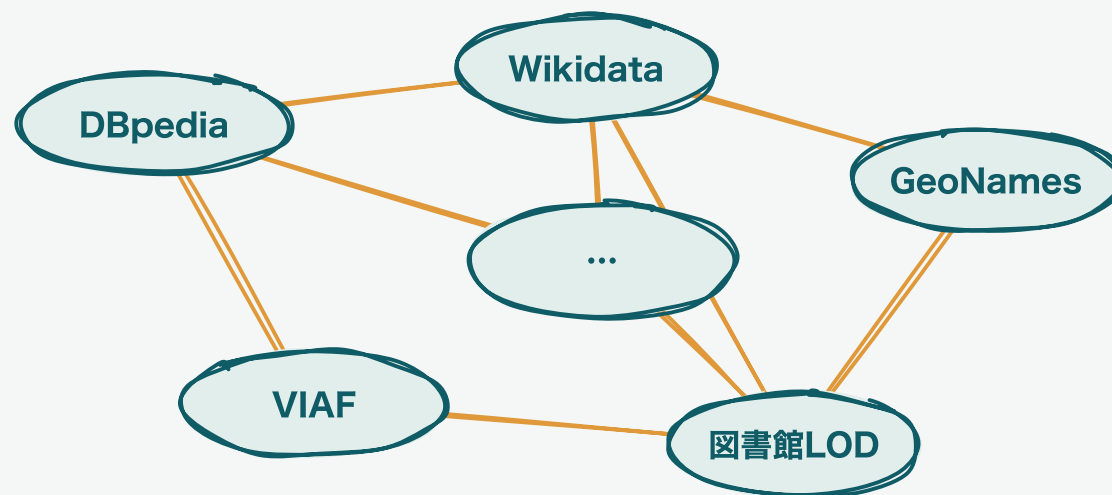
つないで広げる — リンクトデータ

リンクトデータ = 外のデータへリンクを張る



自分のデータの「もの」を、**外部の同じもの**へ URI で結びつける

LOD=開かれてつながる「雲」



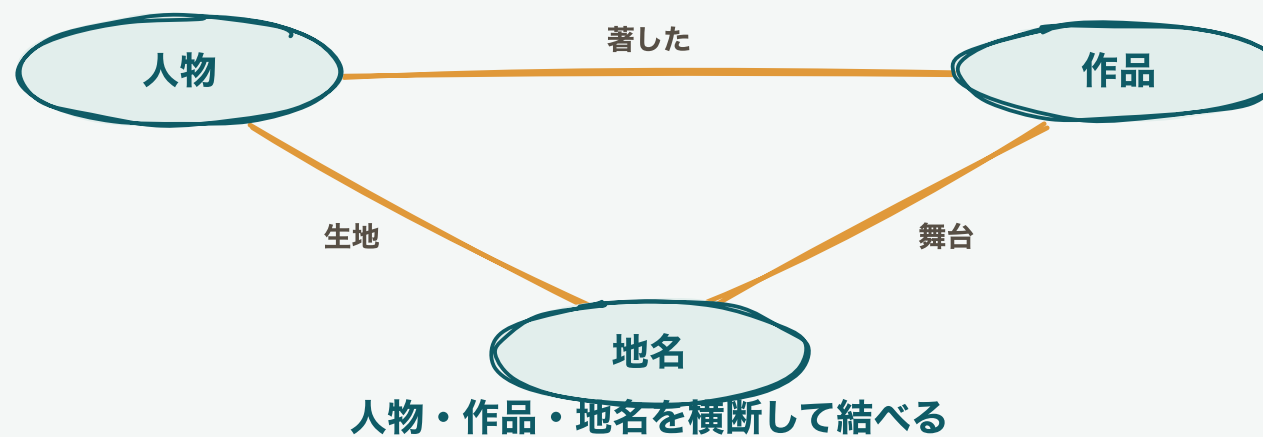
公開データが相互につながる「雲」(LOD)

エルオーディー (リンクト・オープン・データ)。公開データが相互につながっていく

例：Wikidata という結節点

- ✓ **ウィキデータ** は、誰でも編集できる巨大なデータのハブ
- ✓ 多くのデータが、この URI を「共通の待ち合わせ場所」として参照する
- ✓ **ディービーペディア**（ウィキペディア由来）なども、大きな結節点
- ✓ 自分のデータを、こうしたハブに**つなぐ**と一気に広がる

DHでの活用



人物・作品・地名を横断して結び、**機関をまたいで**たどれるようにする

問い合わせる — スパークル

SPARQL=グラフに質問する言語

```
SELECT ?person WHERE {  
  ?person birthPlace Ionia .  
}
```

?person は「探したいもの」。 = 「イオニアで生まれた人を、全部あげて」

一度に、まとめて引き出せる

- ✓ 三つ組のうち、知りたい部分を **?変数** にして問い合わせる
- ✓ 「条件に合う三つ組」を、グラフ全体から**機械的に集めて**返す
- ✓ 手で一件ずつ探すのではなく、**問いを一度書けば**答えがそろう
- ✓ 大量のデータでも、同じ問いで**くり返し**実行できる

考えてみよう

あなたの扱うデータを、**三つ組**にすると何になりますか？

「誰が — 何を — どうした」… ここで少し、動画を止めて考えてみてください。

ここまでのポイント

- ✓ リンクトデータは、自分のデータを**外のデータへ結ぶ**
- ✓ 公開データがつながり、巨大な**LODの雲**になっている
- ✓ **スパークル**を使えば、その網に問いを投げて答えを引き出せる

どうつなぐかにも「判断」が入る

- ✓ どの語彙を使うか、何と何を「同じ」とみなすかには、**解釈**が入る
- ✓ つなぎ方を変えれば、見えてくる関係も変わる
- ✓ データを「どう構造化するか」にも、同じように**解釈**が入る
- ✓ 線でつなぐことは、機械的なようであり、すぐれて**人文学的な営み**

始め方・学ぶには

- ✓ まず触れる：**ウィキデータ**で、ものの URI と三つ組を眺めてみる
- ✓ 体系的に：**Programming Historian** 「Intro to Linked Open Data」
- ✓ DH向け：**Aalto SeCo** の Linked Data 教材
- ✓ 試す：公開された **スパークル・エンドポイント** に、簡単な問いを投げしてみる

まとめ

- ✓ RDFは、データを**主語・述語・目的語の三つ組**で表す
- ✓ **URI** と**共通語彙**で、取り違えず・かみ合う形にする
- ✓ リンクで結ぶと**LODの雲**になり、**スパークル**で引き出せる
- ✓ どうつなぐかにも判断が入る、**解釈をともなう営み**

テキストに「タグを付ける」TEIから、データを「線でつなぐ」RDFへ。地続きの考え方です

出典・ライセンス

本動画は、以下のオープンライセンス教材を参照して作成しました。

- ✓ Introduction to the Principles of Linked Open Data / J. Blaney (Programming Historian) — CC BY 4.0
- ✓ Linked Data Technologies for Cultural Heritage and DH / SeCo, Aalto University — CC BY 4.0

スライド・図は中村による新規作成（概念を参照し、表現は新たに構成）。

ご清聴ありがとうございました