

コミュニティ生成データを典拠 として用いた目録作成の試み： ビデオゲーム目録を事例として

TP&Dフォーラム

2021-11-27

福田一史（大阪国際工科専門職大学）

関連リンク

- <https://scrapbox.io/fukudakz/TP&Dフォーラム2021>

TP&Dフォーラム2021

参考リンク

Wikidata

- [Wikidataのゲームリソース数](#)
- [Wikidataのメディア芸術データベースID登録数](#)

Wikipedia

- [ファミリーコンピュータのゲームタイトル一覧](#)

RCGSコレクション

- [総トリプル数カウント](#)
- [タイプ一覧と各々のレコード数](#)
- [Wikidataにマッピング可能な作品リソースのプロパティ数 \(降順\)](#)

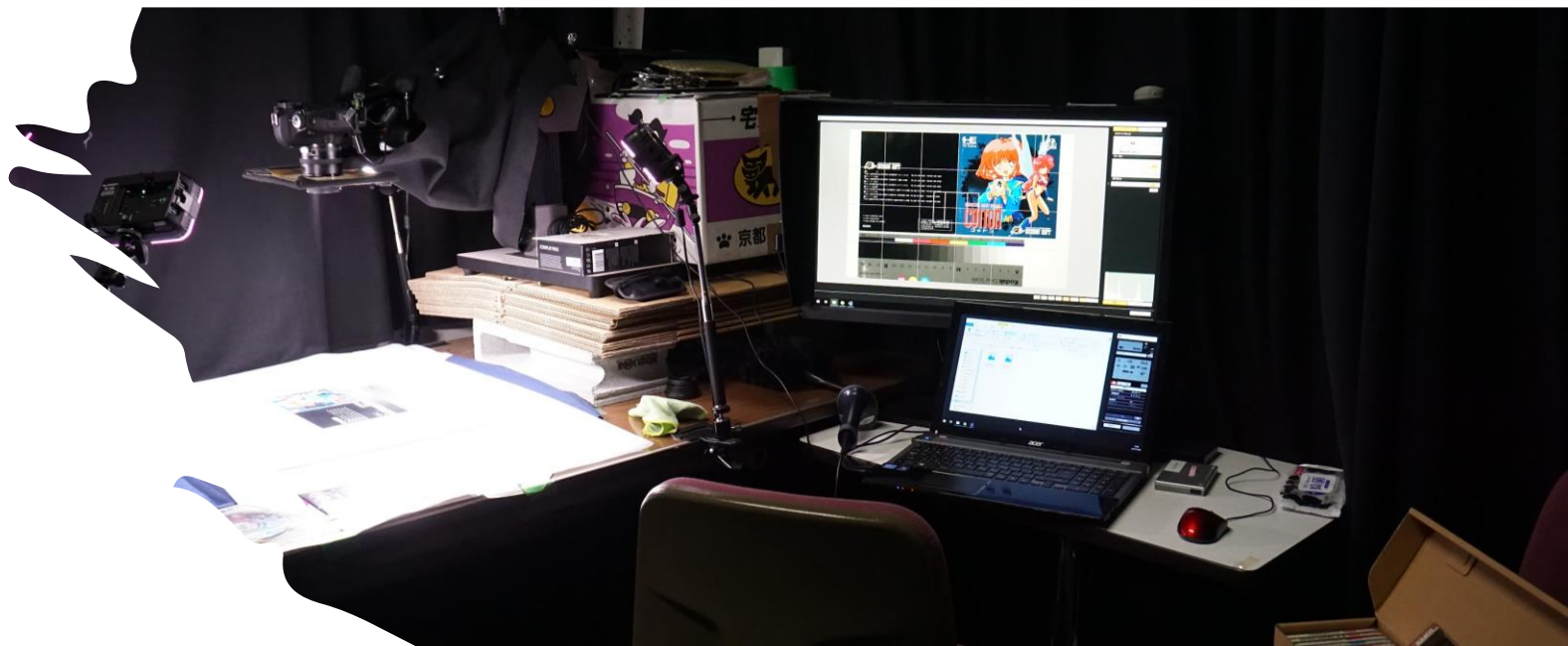
NDL Authorities: SPARQL Endpoint

- [NDL Authoritiesの著作一覧](#)



これまでの 研究活動

- ゲームアーカイブの構築 (1999～)
 - ゲームアーカイブプロジェクト
 - 立命館大学ゲーム研究センター
- ビデオゲーム文化・産業の研究 (2006～)
 - 経営学・科学技術社会論
- 目録・メタデータ (2015～)
 - メディア芸術データベース
 - RCGSコレクション



研究的関心

- ゲームを中心とするポピュラー文化研究のための知的基盤構築
- ビデオゲームのアーカイブ構築とその活用
 - データモデリング
 - 書誌データ作成の方法論構築
 - アーカイブの活用（教育利用・データ利用）

1. 研究の背景

ウェブの普及を通じたユーザ指向の目録への転換

研究の背景

- コンピュータとウェブの発展により、情報の整理や組織化も変化が生じている
 - 例えば、カード・冊子の目録、コンピュータ目録、オンライン目録といった目録の形式の変化
- オンラインでの情報公開の普及に伴い、セマンティックウェブやLinked Dataへの注目が高まる（渡辺 2004, 橋詰 2015, Smith-Yoshimura 2020）

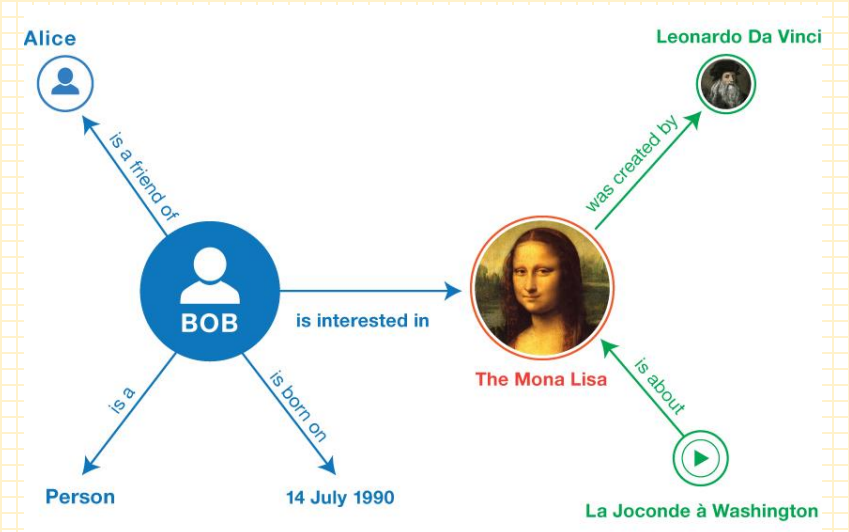


図. Informal graph of the sample triples
<https://www.w3.org/TR/rdf11-primer/> (Accessed 2021-11-25)

研究の背景

- ここでは、図書館コミュニティが生成する膨大な知的リソースの公開と、ウェブオリエンテッドなリソースへの接続が期待されている。
- 一方で、図書館コミュニティにおける目録作成は、ルールドリブンであり、純化が進展しているようにも感じられる。
 - そうして作成したデータを別の人たちがつなぐという役割分担？

研究の背景

- FRBR／IFLA LRMの台頭
- 図書館資料記述のための概念モデルであり、RDAやNCR2018などといった形で社会的・技術的な実装が進んでいる
 - 研究は進められているものの（橋爪 2007, 田辺・原田 2009, 谷口 2009）、実際にどのようにこれらのデータを作るかは議論不十分？
- そもそも90年代以前は目録作成では簡易化が議論されていたにもかかわらず、FRBRなどのモデルはユーザ指向（和中 2002）でありリッチ化が進んだ。
 - 理論的整理のみならずデータ作成の方法論にも注目する必要がある

研究の背景

- 情報のリッチ化への転換は、**コンピュータとウェブへの期待が示唆**されていると捉える事もできる
 - 機械学習などによる**リソースの自動生成**や**オンラインコミュニティとの協働**など
- オンラインコミュニティとの協働や、データ作成への多様なアクターの参加というアプローチは近年の図書館の動向にもフィットするものだと言える（2. で詳述）
- そもそもユーザ指向のデータを作るためには本来的に多くの人々が論争に参加し、様々な論点整理と方法論の試行ならびにそこで得られた知見の共有が必要だと想定される
 - 目録作成を「開かれた」ものへ転換していくことを検討していきたい

研究の背景

- 本研究では、「ビデオゲームの所蔵目録」の事例を通じて、このような環境下における目録作成の新しいアプローチ、オンラインコミュニティにより生成されたデータ（以下、コミュニティ生成データと呼ぶ）活用の方法論について検討を行っていく。
 - オンラインコミュニティ = Wikipediaなどのサービスとそのユーザ群で形成される共同体

2. 先行研究

ビデオゲームの記述とオンラインコミュニティにより生成されるデータ

Wikipedia関連サービスと図書館コミュニティ

- **Wikipedia**とその関連サービスの**図書館での活用**については多くの実践や研究蓄積がある
 - 図書館におけるWikipediaのサポート機会の調査 (Bartholmei et al. 2016)
 - 高等教育における活用事例を集めた論集 (Bridges et al. 2021)
 - カナダの音楽シーンのプロモーションのためのWikipedia編集イベントの事例と、そこでのLinked Dataの有効性への着目 (Allison-Cassin and Scott 2018)
 - OCLCと大英図書館のWikipedian in Residenceが主導した、バーチャル国際典拠ファイル (VIAF) の典拠データを人名入りのWikipedia記事と統合するプロジェクトの事例研究 (Klein and Kyrios 2013)
 - VIAFとWikipediaの相互リンク作成実践 (Gray 2013) - [サンプル](#)
 - Wikimedia Commonsのサービス「GLAM-WIKI」の活用事例報告 (嘉村 2020)

Wikipedia関連サービスと図書館コミュニティ

- Wikipediaには数多くの姉妹プロジェクトが存在する。目録・書誌データという観点ではWikidataは親和性が高い。

ウィキペディアは非営利団体であるウィキメディア財団によって運営されています。並びに以下の姉妹プロジェクトも運営しています。

 コモンズ ファイルの集積	 MediaWiki ウィキソフトウェアの開発
 メタウィキ 全プロジェクトの議論	 ウィキブックス 教科書や解説書
 ウィキデータ フリー知識ベース	 ウィキニュース 自由なニュース
 ウィキクォート 引用句集	 ウィキソース 著作権フリー文書
 ウィキスピーシーズ 生物種のディレクトリ	 ウィキバーシティ 学習支援
 ウィキボヤージュ フリー旅行ガイド	 ウィクシヨナリー 多機能辞典

図. Wikipediaの姉妹サービス（Wikipediaトップページより）

Wikidata

- Wikidataは「人間とコンピューターの双方が平等に参照・編集できるフリーかつオープンな知識データベースサイト」であり、「ウィキメディアの姉妹プロジェクト群（ウィキペディアやウィキボヤージュ、ウィクショナリーなど）の構造化データを集中的に保管・管理」することを目的とするサービス
 - “Wikidata” n.d. Accessed 2021-10-28.
<https://www.wikidata.org/>



図. Wikidataロゴ

Wikidataの機能的特徴

1. あらゆる概念を対象とし得るデータモデルと、それに基づくURIの発行（大向 2019）
2. Wikipediaの各言語記事のハブとして用いられることによる言語間リンクの維持コスト低減
3. 外部のデータセットの識別子のハブ（大向 2020）

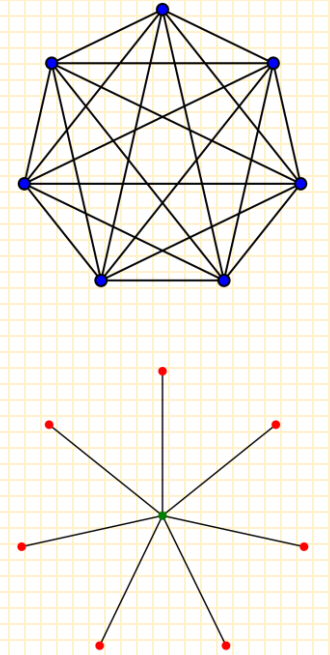


図. Wikidataの言語リンクの構造
(大向 2019, p.560 より)

Wikidataの機能的特徴

- オンラインコミュニティによりデータ作成が進められる。
- そのため注目されやすい概念とそうでないものとの、情報の密度に落差がある。
- ただし、多くのWikidataのリソースはWikipediaの記事にリンクを持っており、その文書の記述によってカバーすることができる (Erxleben et al. 2014) 。

ビデオゲームのコミュニティ生成データ

- グローバルな受容を背景に、インターネットでは活発なビデオゲームに関するデータの生成とコミュニティ形成が展開
- ビデオゲーム文化における知識生成性の高さには注目が集まっている（McGonigal 2011）。
 - 「シリアスゲーム」「ゲーミフィケーション」などといったキーワードでビデオゲームの社会応用として注目される。

WoWiki

EXPLORE ▼ WOW INFO ▼ MORE WOW INFO ▼ MOVIE INFO ▼ COMMUNITY ▼

106,202
PAGES



図. WOWWikiトップページヘッダー

Wikidataのビデオゲームデータ

- コミュニティ型サービスにおいてビデオゲームはデータの充実が期待される分野
- Wikidataのビデオゲーム (Q7889) リソース登録数
 - 44,544件 (2021-11-01時点)
- Wikipediaのゲームタイトル一覧では、おおよその網羅率が確認できる

4月 [編集]

- 4月4日 [アーマード・コア3](#) (フロム・ソフトウェア)
- 4月4日 [実戦パチスロ必勝法!Sammy's Collection](#) (サミー)
- 4月4日 [ノックアウトキング 2002](#) (エレクトロニック・アーツ・スクウェア)
- 4月4日 [信長の野望・嵐世記](#) (コーエー)
- 4月4日 [リリーのアトリエ プラス ~ザールブルグの錬金術士3~](#) (ガスト)
- 4月11日 [スタントGP](#) (エム・ティー・オー)
- 4月18日 [熱チュー!プロ野球2002](#) (ナムコ)
- 4月18日 [ときめきメモリアル2 ミュージックビデオクリップ サーカスで逢いましょう](#) (コナミ)
- 4月25日 [ガレリアンズ:アッシュ](#) (エンターブレイン)
- 4月25日 [機甲兵団 J-PHOENIX バースタクティス](#) (タカラ)
- 4月25日 [機甲武装Gブレイカー 第三次クラウドディア大戦](#) (サンライズインタラクティブ)
- 4月25日 [サーヴィランス 監視者](#) (ソニー・コンピュータエンタテインメント)
- 4月25日 [最強の囲碁2](#) (アンバランス)
- 4月25日 [最終電車](#) (ヴィジット)
- 4月25日 [SHINE](#) (サクセス)
- 4月25日 [ジェネレーションオブカオスNEXT](#) (アイディアファクトリー)
- 4月25日 [SIMPLE2000シリーズVol.4 THEダブル麻雀パズル](#) (ディースリー・パブリッシャー)
- 4月25日 [SPY HUNTER](#) (ミッドウェイ・ゲームス)
- 4月25日 [絶体絶命都市](#) (アイレムソフトウェアエンジニアリング)
- 4月25日 [必殺パチンコステーションV3 出動!ミニスカポリス](#) (サン電子)
- 4月25日 [日米間プロ野球 FINAL LEAGUE](#) (スクウェア)
- 4月25日 [ワールドサッカーウイニングイレブン6](#) (コナミ)

図. WikipediaのPlayStation2 2002年4月発売タイトル一覧

Wikidataのビデオゲームデータ

- URI参照、リソースAPI、MediaWiki API、SPARQLエンドポイントといったデータ提供により、目録へのデータ援用性は高い

その他のコミュニティ生成データ

- [IGDB](#)

- ゲーム動画プラットフォーム「Twitch」の関連サービス
- 複数の実体とその関連で記述される精度の高いデータが提供される
- <https://api-docs.igdb.com/>

- [MobyGames](#)

- 最も古いゲームデータ提供サービス、登録ゲーム28万件以上
- XML形式のAPIによるデータ提供

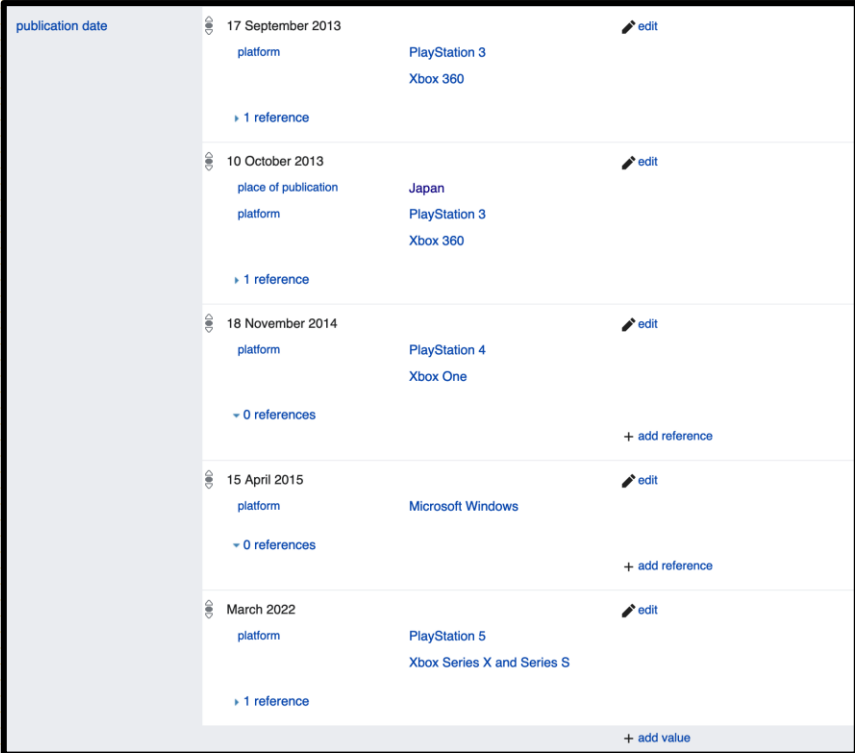
- [Steam](#)、[GiantBomb](#)、[thegamesdb.net](#)など...



図. ゲームサービスロゴ

ビデオゲームのコミュニティ生成データ

- これらのサービスのデータは、ゲーム作品（Work）を主要な対象として、リッチなデータを提供する
- 一方、ゲームのパッケージやリリースなどのデータはほとんど記録されないか、限定的
 - Wikidataの事例 ([Grand Theft Auto V](#))



publication date	platform	edit
17 September 2013	PlayStation 3 Xbox 360	edit
↳ 1 reference		
10 October 2013	Japan PlayStation 3 Xbox 360	edit
↳ 1 reference		
18 November 2014	PlayStation 4 Xbox One	edit
↳ 0 references		
+ add reference		
15 April 2015	Microsoft Windows	edit
↳ 0 references		
+ add reference		
March 2022	PlayStation 5 Xbox Series X and Series S	edit
↳ 1 reference		
+ add value		

図. GTA5の出版日

ビデオゲームのコミュニティ生成データ

- MobyGamesに登録されるビデオゲームとその開発スタッフのデータを分析し、社会的な距離感がイノベーションにもたらす影響（Vaan et al. 2015）。
- メディア芸術データベース・MobyGames・IGDBを接続したデータセットによる日本のビデオゲームの歴史的発展の可視化（Mühleder et al. 2020）
- その他にも関連研究があるなど（Therrien et al. 2021, 原 et al. 2021）、研究ニーズと言う観点からコミュニティ生成データへの期待は高い

ビデオゲーム記述のデータモデル

- 2010年代から以下など研究蓄積が進む（福田 2018）。
 - FRBRの適用と作品間のねじれた関連（McDonough et al. 2010）
 - ゲームミュージアムのためのメタデータスキーマ（Lee et al. 2013）
 - 専用データモデルの提案（Jett et al. 2016）
 - IFLA図書館参照モデルのビデオゲーム記述適応モデルの開発（Fukuda and Mihara 2018）
 - オンライン目録における実装（福田 2019, 福田2021）
 - 複数のビデオゲームデータセットを統合するためのオントロジー・語彙の開発（Hoffman 2019）

ビデオゲーム記述のデータモデル

- FRBR／IFLA図書館参照モデルのアプローチはある程度有効
- ただし、そこで定義される実体がそもそもコンピュータ・プログラムのようなものを想定したものではないこともあり、それらモデルの実体の**解釈や配置に「ずれ」が生じえる**点が課題
- これらの概念モデルと相互運用可能でユーザ・開発者・研究者など**ドメインの情報要求を満たす専用モデルの開発が必要**

ビデオゲーム記述に関する検討

- 筑波大学
 - WikipediaやDBpediaを用いたメディア芸術書誌データの組織化研究（カブリン et al. 2013, 濱田 et al. 2017, 大石 et al. 2018, 内海 et al. 2019）
- diggr.link
 - ライプツィヒ大学図書館らのビデオゲームのデータ接続プロジェクト「diggr.link」（Freybe et al. 2019, Huffman 2019）
- [Japanese Visual Media Graph](https://jvmedia-graph.org/)
 - アニメ・ビジュアルノベルに焦点化した複数データセットの接続・変換を行うStuttgart大学のプロジェクト（Kiryakos and Pfeffer 2021）

3. 研究の目的・方法

研究の目的・方法

- 作品データの作成における膨大な作業量が課題
- ビデオゲームの領域では、複数のサービスで作品のデータセット生成体制が確立しており、継続的なデータの登録やメンテナンスも期待できる
- 本研究では、コミュニティ生成データを典拠とする、データ接続による目録作成を試行する。

研究の目的・方法

- 本方法論で想定されるメリット
 1. 膨大なリソースとそれらの関連が**低コスト**に記録可能
 2. ともしれば主観的な記録になりやすい「作品」などの抽象度の高い作品のデータについてコミュニティでの検証を経た**データセットの精度における妥当性の高さ**
 3. 外部データセットとの接続が活発であり、LODとして高価値なデータに
 4. コミュニティ生成データの欠陥や不備を、資料に基づきデータ提供することで、所蔵機関からコミュニティへの寄与が期待できる

研究の目的・方法

- 立命館大学ゲーム研究センター（Ritsumeikan Center for Game Studies: RCGS）の所蔵資料をサンプルとする
 - ビデオゲーム記述モデルと目録作成実践（福田 2021）
 - Omeka Sを用いたウェブサービス「[RCGSコレクション](#)」

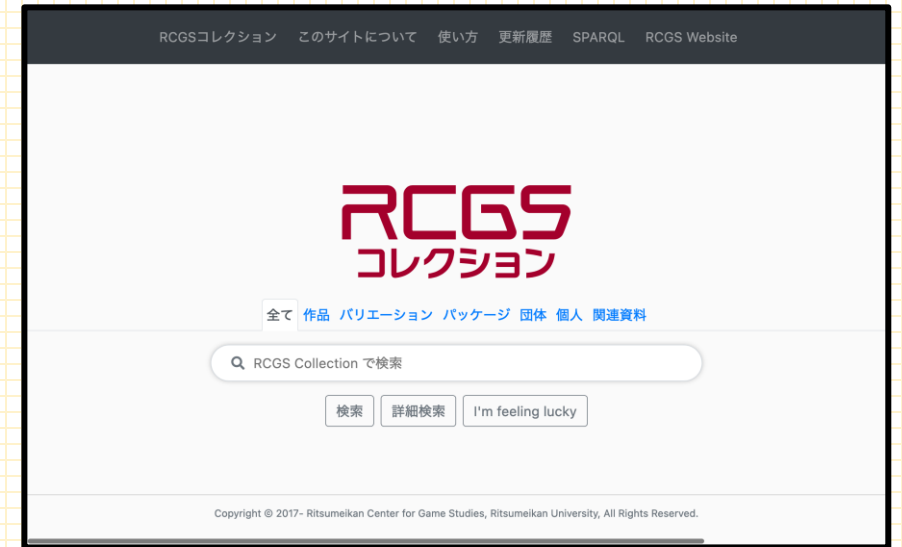


図. RCGSコレクショントップページ

種別	件数
個別資料：ゲームパッケージ	11283
個別資料：関連資料	7565

表. RCGSの所蔵資料数

研究の目的・方法

- 本研究ではWikidataを典拠として選定
- 選定要因は下記：
 1. RCGSが所蔵するビデオゲームのうち大部分を占める、**日本で流通したゲーム作品**の登録が多い
 2. 識別子のハブとして成長しているため**Linked Dataとして機能が高い、もしくは他の典拠への転換も比較的容易**
 3. RCGSコレクションはメディア芸術データベースとの連携を想定しており、アニメやマンガなど**その他のメディア芸術もデータ作成の対象である**

研究の目的・方法

- 以下の課題から、Wikidataの目録データ活用
の課題を検討する
 1. 網羅性：カバー率
 2. 粒度：記述単位の統一性
 3. 記録の完全性：記録の厚みとその平準化

4. メタデータモデルとデータ作成

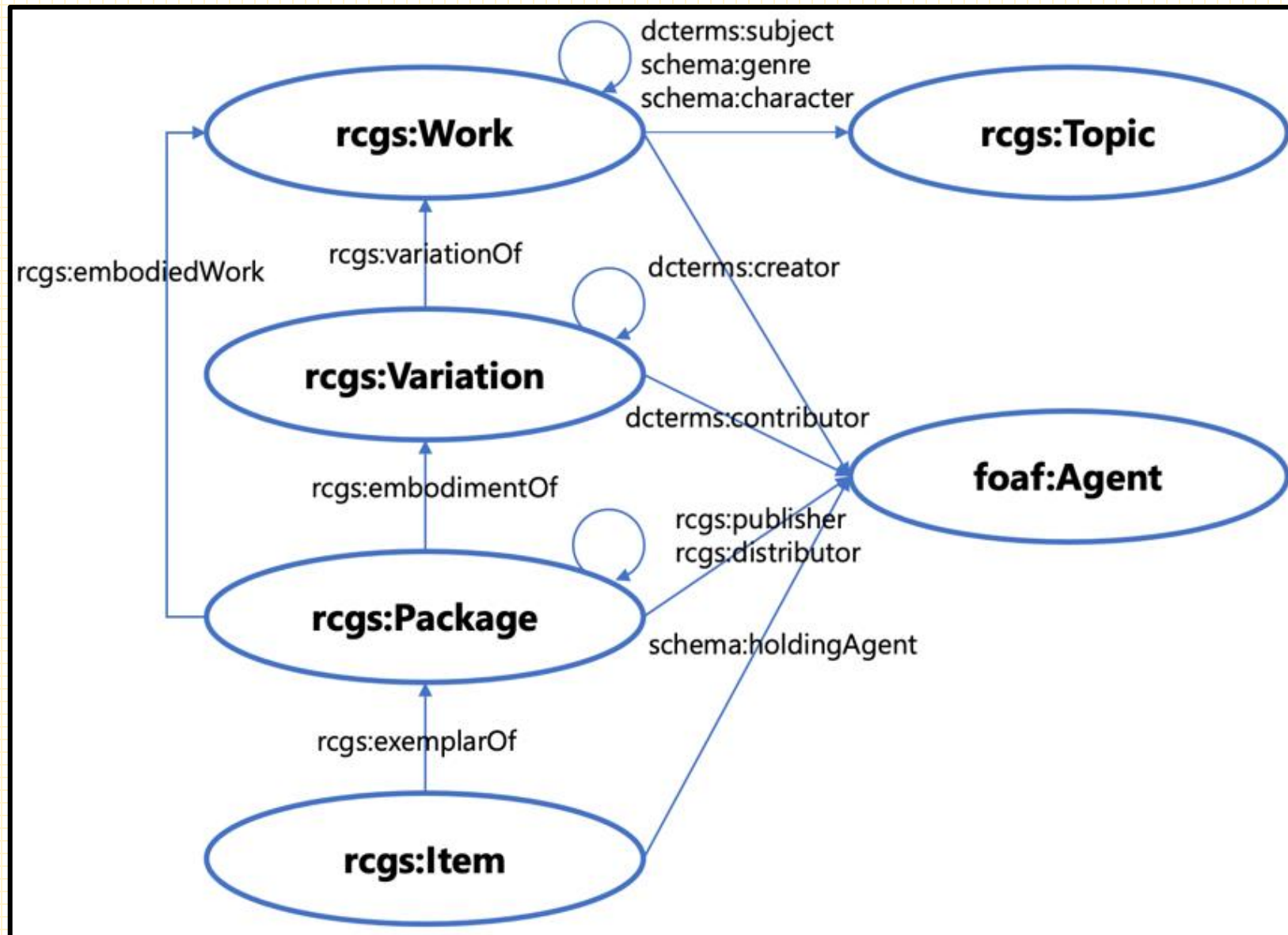


図. RCGSのビデオゲーム記述モデル

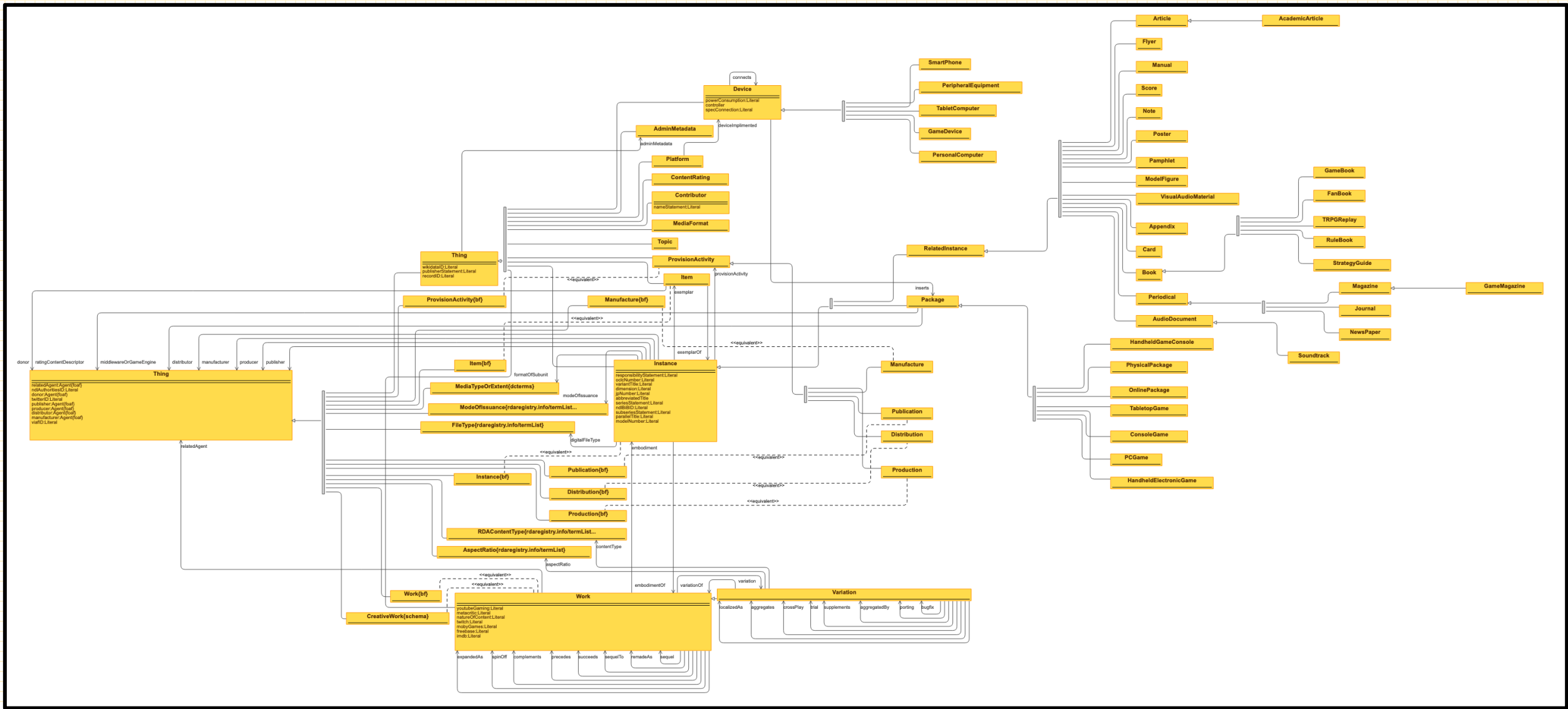


図. RCGSビデオゲーム語彙のグラフ by OWLGrEd

- 前頁：データモデル概要
- 右図は「Dark Souls 3」をモデルで記述した事例
- 作品リソースはWikidataとWikipediaのリソースとのマッチングで生成
- 2年間の期間中、16名のRCGSスタッフによりデータ作成が進められた

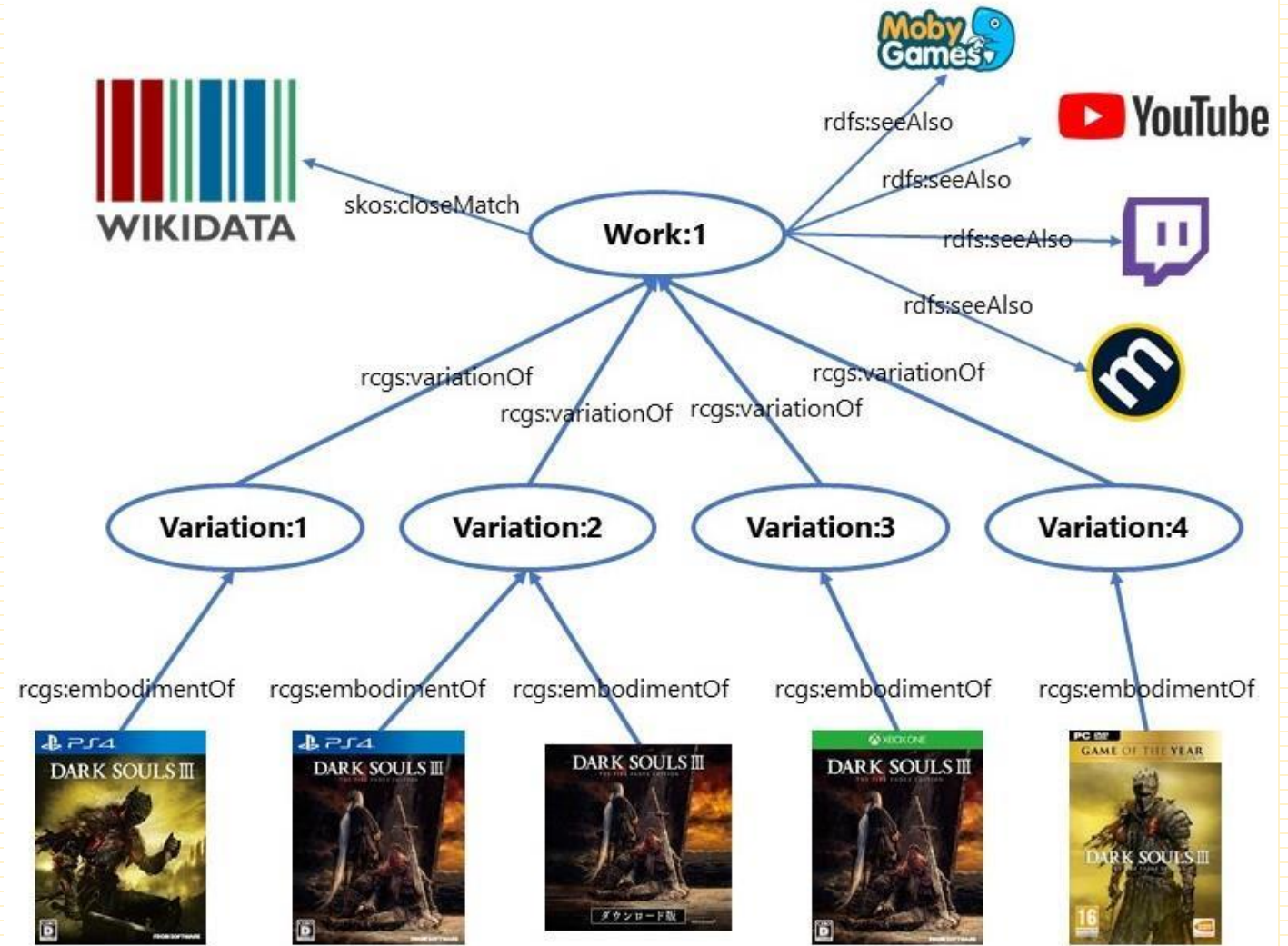


図. “Dark Souls 3”のデータモデルへの適用

5.

目録作成を通じた課題検討の結果

網羅性

- 本目録作成で**7,680件の作品リソース**が作成された
- Wikidataにリンクされた作品リソース数は「**4,087件（53%）**」
- 下図の通り作業中の一部を除き大部分のリソースの関連が確立

クラス	A. リソース数	B. 上位クラス関連数	B/A
作品	7,680	-	-
バリエーション	8,615	8,609	0.999
パッケージ	9,569	8,610	0.899
個別資料	11,283	11,283	1

表. ビデオゲームの各クラスリソース数と関連数

粒度

- Wikipediaには**記事の粒度に関する明確な基準がない**。
- 特に作品という抽象性の高い実体を記録するためには、定義・基準に基づく粒度の統一が図られることが好ましい。
- しかし、日本語のWikipediaのガイドラインによると、その粒度の決定は**各編集者に委任**されており、[英語版](#)でも同様である。
- 実際にリメイクをオリジナルと別リソースとして記録する群と、一つのリソース（記事）で記録する群に分かれる。

記録の完全性

- 本モデルでは**作品に56件のプロパティ**が定義される。
 - [記述セットプロファイル](#)より
- これらのデータが充実することで、特にゲームファンや開発者・研究者などゲーム内容に強く興味を抱くユーザの要求に対応したデータとして活用可能性が高まると想定される。
- 次頁の図はWikidataにマッピングされたリソース4,087件のプロパティの記録数をプロットしたものである。
- 40件以上の記録があるリソースは111件 (**2.71%**) であるなど、**ごく一部のリソースに記録が集中**していることが観察される。

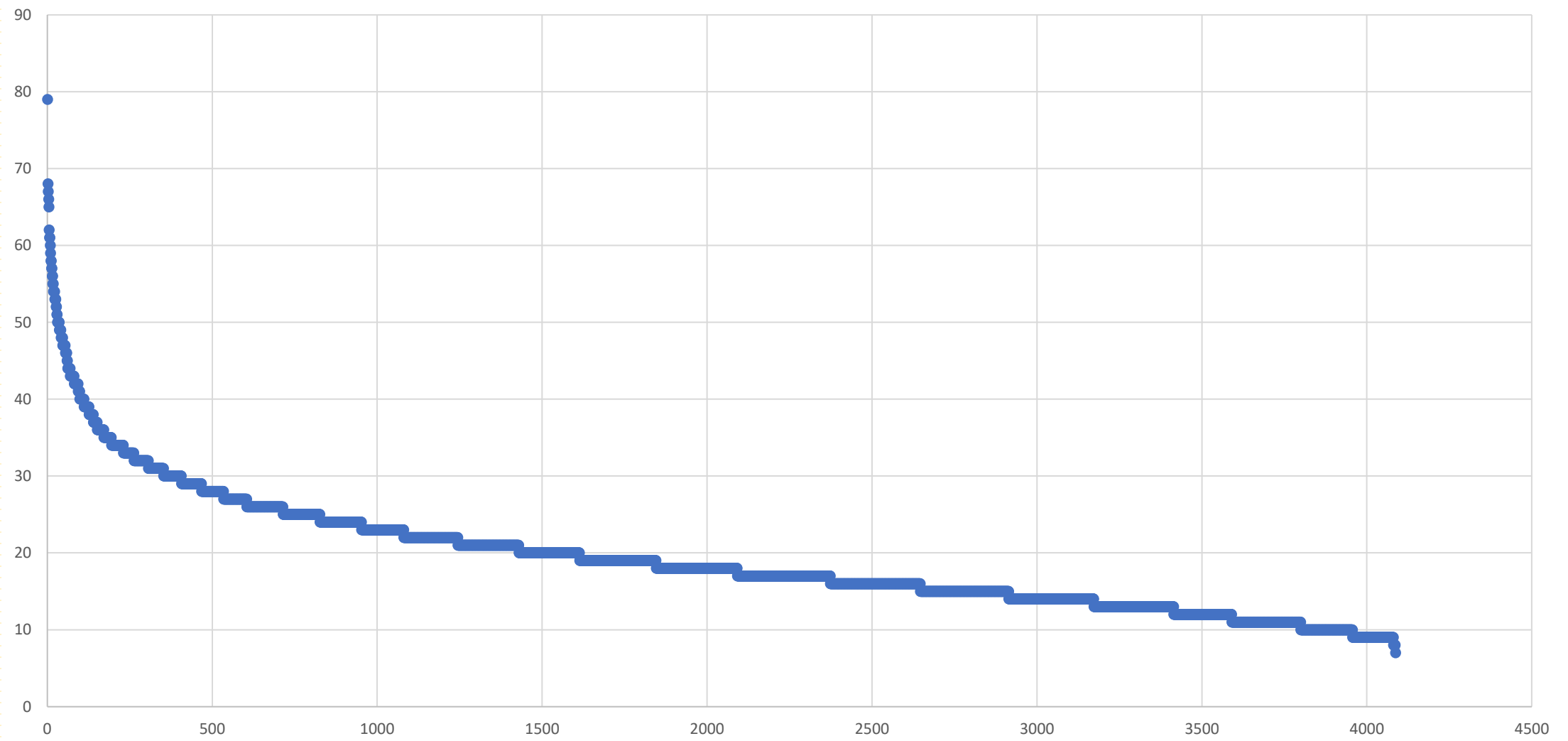


図. RCGSコレクションの作品リソースのプロパティ記録数

検証を通じた評価

- 作品の網羅率は53%。
- まず、**コスト**という観点から**有用性**は否定できない。またデータに一定の信頼性は担保され、ユーザの認識にある程度合致したものにはなることが期待できる。
- カタロガーによると、Wikidataを用いた作品リソースの作成作業は、そうでない作品のリソース作成と比べて、**単位あたり作業時間が短かった**。また、ビデオゲームや当該作品にそれほど詳しくないカタロガーでも作業可能。

検証を通じた評価

- 粒度については、ガイドラインの段階から課題となる。
- ただし、Wikidataのリソースは、外部リソースのIDを複数含むものであるため、他のデータセットにおける記述単位との「ずれ（記述単位や粒度の違い）」に関する**将来的な検証性を有している**という観点は、肯定的に評価できる。
- ただし、IGDBとMobyGamesはカバー率が高いもしくはWikidataで記録できなかった群をカバー出来る可能性があるため、これら**サービスに登録されたゲーム作品データとの比較や補完的な活用の可能性**を改めて検討すべき。

検証を通じた評価

- 記録の厚みに明らかに差が生じることが明らかになった。
- Wikidataではリソース種別ごとにプロパティのリストを共有するなど、スキーマを共通化するための施策は展開されているが、あくまでWikidataの基本設計から**拡張性ならびに自由度は高い**。
- 今後データ登録が進めば、この問題は徐々に解決していく可能性はあるが、当面は**他のデータセットの記録を補完的に利用**して充実を図ることを、検討していくべき。

考察

考察

- 明確な識別対象が定義される実体（Expression, Manifestation, Item）と、作品（Work）やシリーズやフランチャイズやキャラクターという抽象度が高く解釈の余地が大きい実体を区分し、**データ作成を図書館コミュニティとウェブコミュニティで分業・協業**と位置づける
- そうすることで高機能的な書誌データの構築が進められるのではないか、という**アプローチの提案**

考察

- 目録作成にかかるコストの増大化や、特定のドメインの専門性という観点から、図書館コミュニティだけで高精度・高機能な書誌データを構築するということは難しい？
 - ルールオリエンテッドなデータセットを構築しても、外部のリソースと接続しなければLODとしての価値が低くなる可能性も
- また、カタログガーが特定のドメインに詳しいとは限らない。
- さらに、特定のドメインの専門家であっても、国際化やジャンルの複雑化と細分化が進むなか、あらゆるゲームに詳しいという人は存在し難い。

考察

- 本実践では、総じて方法論の有効性が確認された
 - 低コストでの膨大なデータ登録、主観的になりがちな抽象性の高い実体のデータの妥当性の担保、外部リソースのID登録によるLinked Dataとしての価値向上など
 - 特に構造化データの提供による調査研究活用に高い有効性が想定される
 - REST API、SPARQLエンドポイント、ダンプデータ公開など

考察

- オンラインコミュニティへの貢献までは検証が不十分だった
 - 一方で、すでにWikidataとメディア芸術データベースのデータ接続が進んでいる。プロパティの登録と、メディア芸術データベースのリソースIDのWikidataへの登録（14,874）など。
- 本目録作成を通じて、Wikidataに登録されていない作品リソースも多く確認された（47%）。今後、このデータのコミュニティサービスへのデータ登録を検討していきたい。

考察：課題

- 一方で本方法論の**課題**も幾つか明らかになった。
- これらはWikidataのデータセットを単体として使うことによる課題であるという側面がある。網羅率・粒度・記録、さらには統制語彙の不完全性など。
- IDが記録されるIGDBやMobyGamesなどの**他のデータセットを補完的に用いることである程度の解決**が見込まれる。
- 今後、これらを統合的／並列的に運用するための方法論を検討する必要性がある。

考察：モデル

- 本研究の観点からすると、FRBR/LRMの本質的な課題として、「正しい著作リスト」の存在を仮定した点、またそれを前提とするモデルの統合性・権威性が指摘される。
- 解釈可能性の高いWorkのデータセットはユーザの認識に沿った内容あることが好ましいだろう。そうするのであれば、複数の著作典拠との関連という新しいデータモデルのあり方が示唆される。

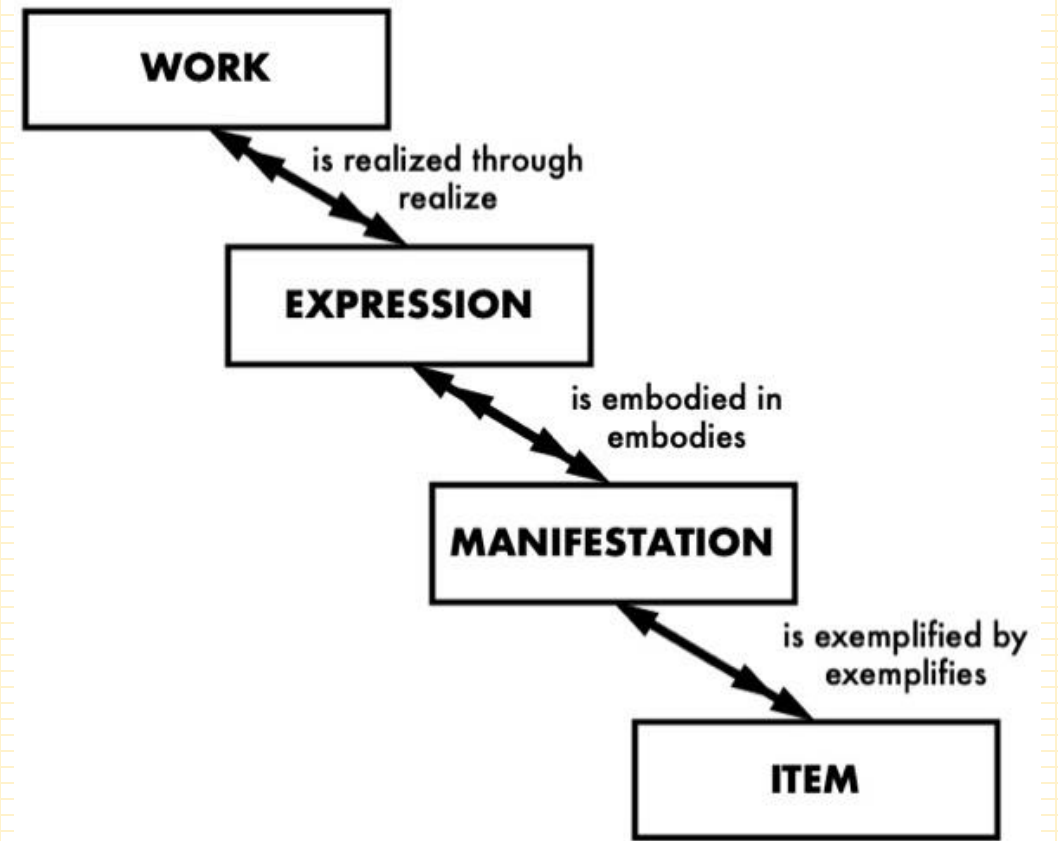


図. WEMI関連におけるモデル拡張

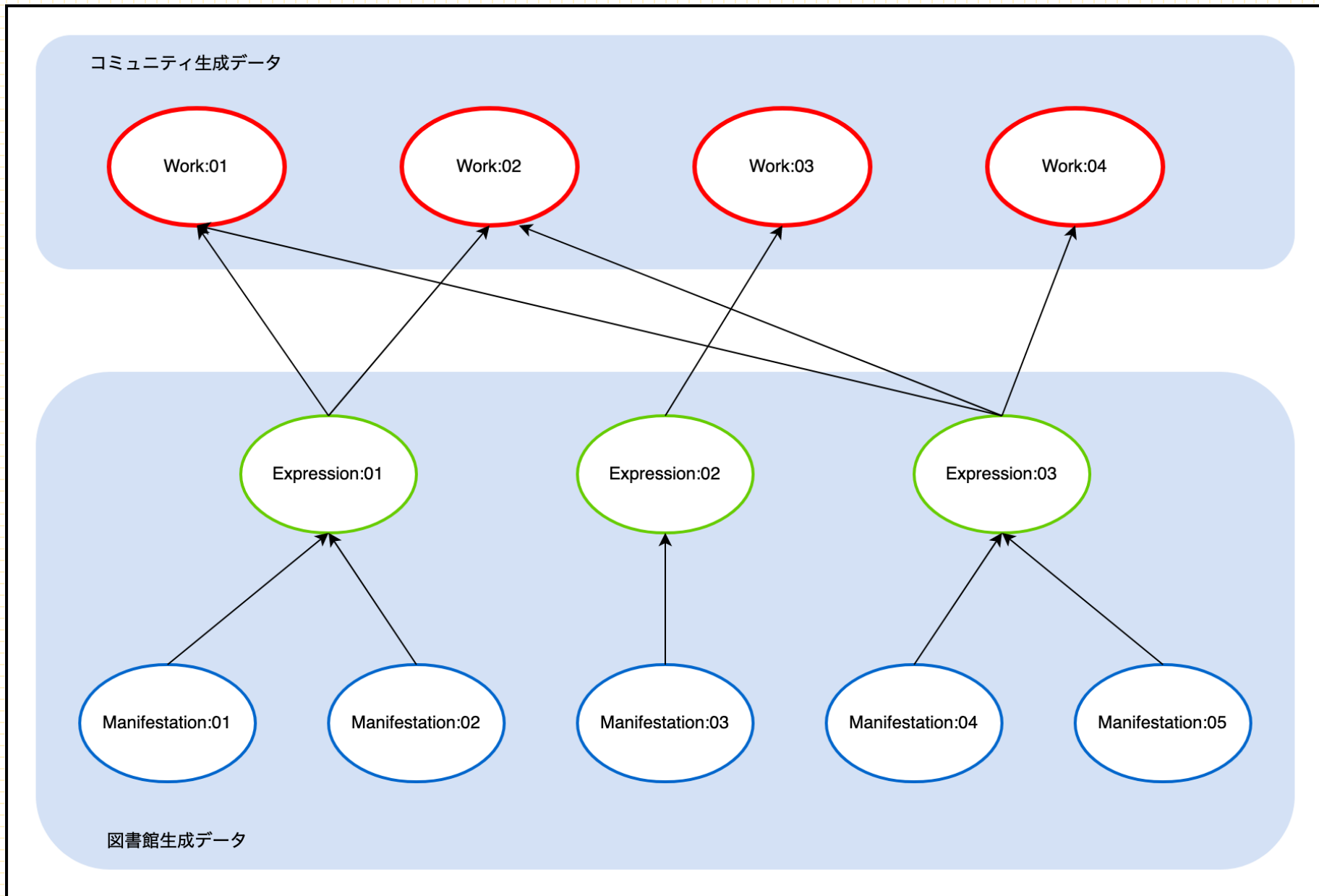


図. 拡張モデルのイメージ

結論と今後の課題

結論と今後の課題

- コミュニティ生成データの有効性と課題が確認できた。
- ただし、これらを活用するためには、ウェブコミュニティとの協業のためのFRBR/IFLA図書館参照モデルの拡張もしくは解釈の可能性が示唆された。
- 本方法論を通じたデータの充実による、専門家ニーズへの対応が可能に。
 - ただし、SPARQLによるアクセスなどでしかデータを活用できないのでは効果が限定的。
 - 各部文書の整備と、構造を簡易化した目録サービスによるアクセス性の向上なども並行して検討していきたい。

結論と今後の課題

- データセット間のリンク生成が本質的なボトルネックとなるが、これらの接続の自動化なども検討していくべき。