

Creating a unified ontology for the Japanese Visual Media Graph

Funded by



Deutsche
Forschungsgemeinschaft

German Research Foundation

Magnus Pfeffer and Zoltan Kacsuk
“Ontologies for narrative and fiction” workshop
University of Groningen
3-4 July 2023

Outline

1. Key ideas and characteristics of the Japanese Visual Media Graph (JVMG) project
2. Overview of source databases
3. Data integration approach
4. Developing a unified ontology for the JVMG knowledge graph:
A pragmatic approach

Key ideas and characteristics of the Japanese Visual Media Graph (JVMG) project

Introducing the JVMG project

- Databases by enthusiast communities are often the **most up-to-date online resource for checking information**



- Project aim: Make these databases available for **large-scale quantitative research, in collaboration with the communities**
 - Funding: German research foundation, 6 years duration

Key characteristics of the JVMG project



COLLABORATION WITH COMMUNITIES

We work with diverse fan and enthusiast communities to make their data available to researchers. We respect their wishes and conditions for the use of the data.



SUITABILITY FOR RESEARCH

The needs of researchers drive all aspects of the development of the graph database, from the choice of data sources and the data model to specific representation details.



OPEN DEVELOPMENT

We document the process of data integration and the associated research in an open manner. You will find regular updates on the [project blog](#).

Overview of source databases

Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database (VNDB):** Focused on visual novel games only
- **Anime Characters Database (ACDB):** Focus on one aspect of the domain

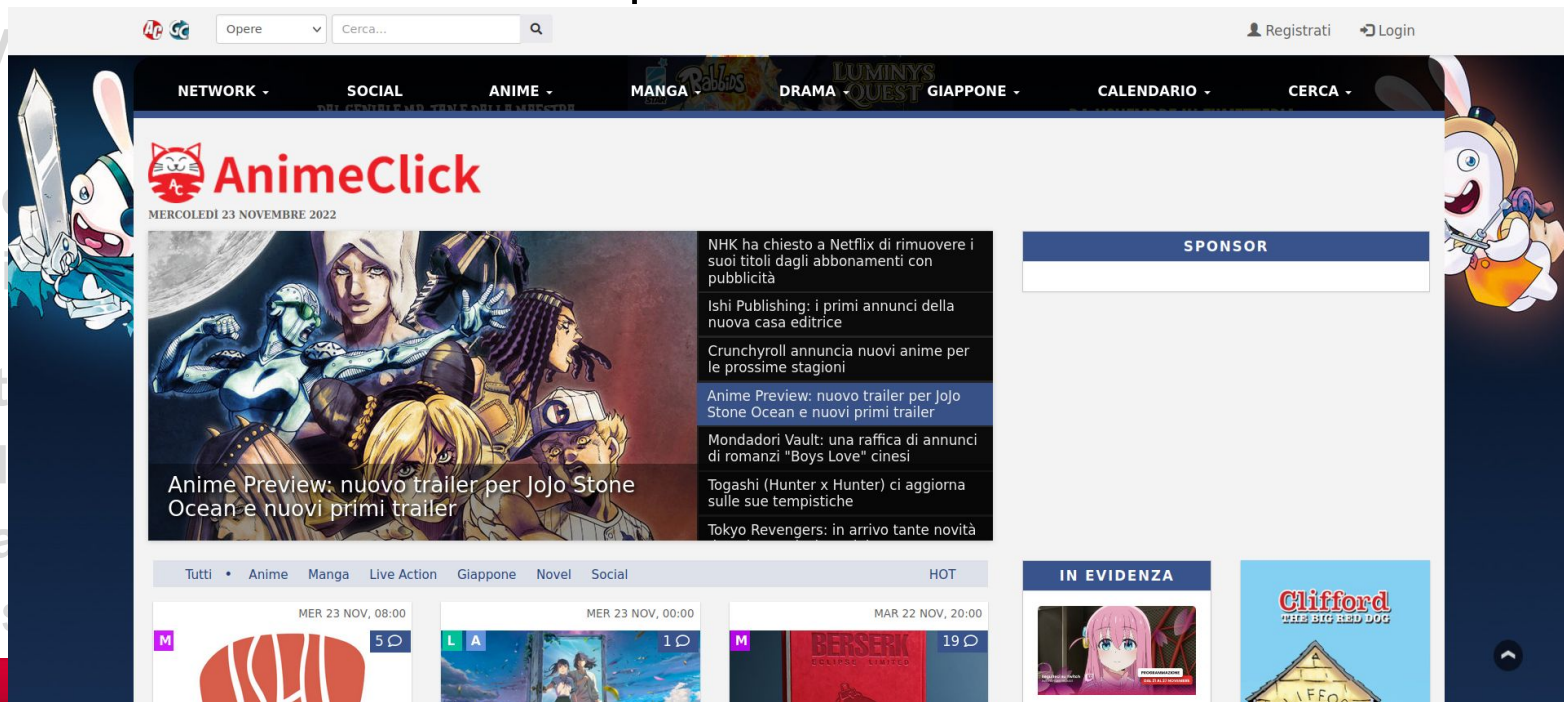
Other databases:

- **Wikidata:** Not focused on Japanese visual media
- **Media-Arts Database:** Collects information on manga, animation, games and media art from institutions, creators and publishers

Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture

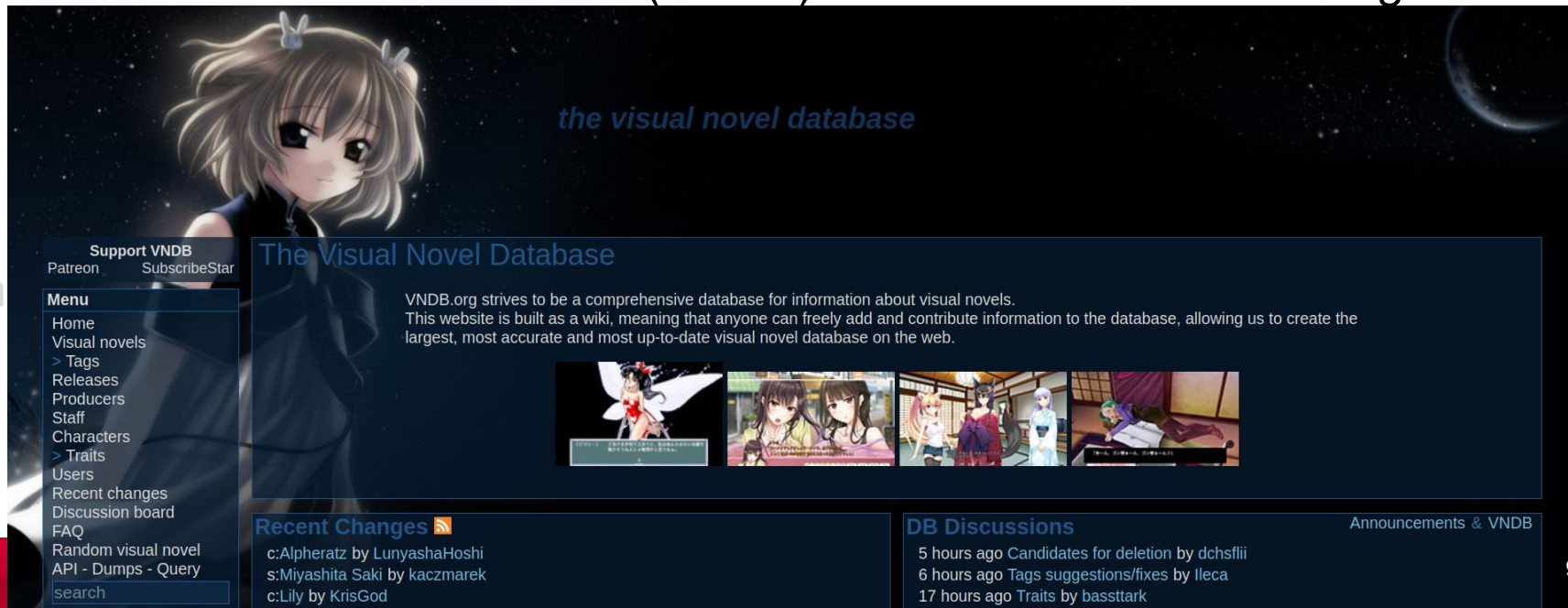


The screenshot shows the AnimeClick website interface. At the top, there is a search bar and navigation links for 'Registrati' and 'Login'. The main navigation bar includes categories: NETWORK, SOCIAL, ANIME, MANGA, DRAMA, GIAPPONE, CALENDARIO, and CERCA. The page title is 'AnimeClick' with a date 'MERCOLEDI 23 NOVEMBRE 2022'. The main content area features a large anime preview for 'Jojo Stone Ocean' with the text 'Anime Preview: nuovo trailer per Jojo Stone Ocean e nuovi primi trailer'. To the right, there is a 'SPONSOR' section. Below the main content, there are sections for 'Tutti' (Anime, Manga, Live Action, Giappone, Novel, Social) and 'HOT' (Berserk, etc.). At the bottom, there are sections for 'IN EVIDENZA' (Clifford the Big Red Dog) and a 'GAMES' section.

Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database (VNDB):** Focused on visual novel games



the visual novel database

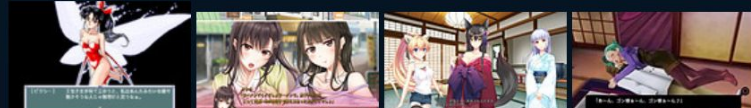
Support VNDB
Patreon SubscribeStar


Menu

- Home
- Visual novels
- > Tags
- Releases
- Producers
- Staff
- Characters
- > Traits
- Users
- Recent changes
- Discussion board
- FAQ
- Random visual novel
- API - Dumps - Query
- search

The Visual Novel Database

VNDB.org strives to be a comprehensive database for information about visual novels. This website is built as a wiki, meaning that anyone can freely add and contribute information to the database, allowing us to create the largest, most accurate and most up-to-date visual novel database on the web.



Recent Changes 

- c:Alpheratz by LunyashaHoshi
- s:Miyashita Saki by kaczmarek
- c:Lily by KrisGod

DB Discussions Announcements & VNDB

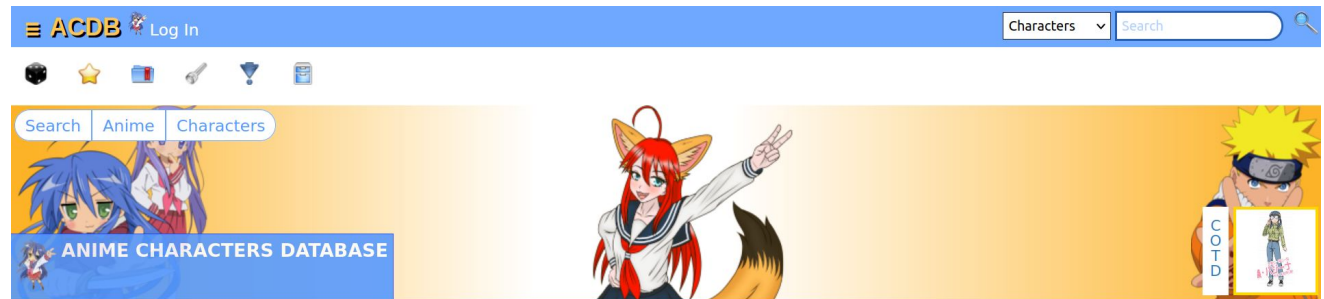
- 5 hours ago Candidates for deletion by dchsflii
- 6 hours ago Tags suggestions/fixes by lleca
- 17 hours ago Traits by basstark

Oth

Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database (VNDB):** Focused on visual novel games only
- **Anime Characters Database (ACDB):** Focus on one aspect of the domain



Other databases:

- **Wikidata:** Not
- **Media-Arts Da**
games and me

WELCOME TO ANIME CHARACTERS DATABASE

The visual search engine for animated characters. We index characters by eye color, hair color, hair length, age, gender, and animal ears.

Project page Discussion Read View source View history Search Wikidata



The screenshot shows the Wikidata homepage with a central banner. The banner text reads: 'ウィキデータへようこそ' (Welcome to Wikidata), '100,469,930 件の項目を持つ、誰でも編集可能なフリーの知識ベースです。' (A free knowledge base with 100,469,930 items, editable by anyone), and 'はじめに・井戸端・コミュニティ・ポータル・ヘルプ' (Getting started, Village pump, Community, Portal, Help). Below the banner is a network diagram with nodes labeled '自由' (Free), '多言語' (Multilingual), '共同作業' (Collaboration), and '連続' (Continuity). A search bar is at the top right. The left sidebar contains navigation links like 'Main page', 'Community portal', and 'Tools'. At the bottom, there are two boxes: 'ようこそ!' (Welcome!) and 'データについて' (About data).

ウィキデータへようこそ

100,469,930 件の項目を持つ、誰でも編集可能なフリーの知識ベースです。

はじめに・井戸端・コミュニティ・ポータル・ヘルプ

Want to help translate? Translate the missing messages.

ようこそ!

ウィキデータは、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す。

データについて

素晴らしいウィキデータの世界は初めてですか？ すぐに基本を身につけられるコンテンツを通してデータ・リテラシーを高めましょう。

and culture
ial novel games

aspect of the

Other databases:

- **Wikidata:** Not focused on Japanese visual media
- **Media-Arts Database:** Collects information on manga, animation, games and media art from institutions, creators and publishers



メディア芸術データベース

ベータ版

すべて

マンガ

アニメーション

ゲーム

メディアアート

まとめ

日本語

English

簡体中文

한국어

James

the

Other databases:

- Wikidata: Not focused on Japanese visual media
- **Media-Arts Database:** Collects information on manga, animation, games and media art from institutions, creators and publishers

Entity and concept numbers

Enthusiast community	Works and media				Company	Characters	Work properties	Character properties	Involved people
ACDB	Work					Character	Work Tag	Character Tag	People
	10.207					107.369	1.088	4.051	5.557
AnimeClick	Animation Work	Comic Work				Character			Staff
	9.491	11.762				102.143			39.604
VNDB			Visual Novel	Release	Producer	Character	Tag	Trait	Staff
			28.190	71.349	10.394	90.077	2.585	2.777	21.164

Entity and concept numbers

Database	Works and media					Characters	
Wikidata	Anime titles	Manga series	Video game	Light novel & LN series		Anime character	Manga character
	4.467	13.871	47.192	867		3.788	2.990
Media-Arts Database	Anime titles	Anime items	Game items	Manga book series	Manga magazine issues		
	12.085	~135.000	~61.000	133.779	170.670		

Recap: what are we working with?

Heterogeneous data sources:

- Collected data
- **Granularity**
- **Ontologies**
- Database structures

Data integration approach

Data Integration: The RDF Way

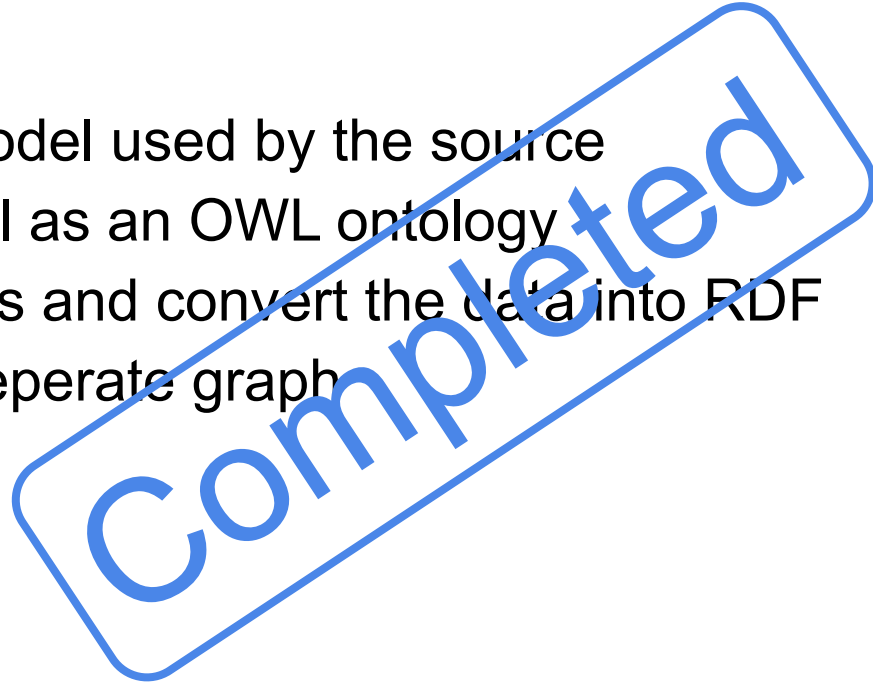
For each data source

- Analyse the data model used by the source
- Formalize the model as an OWL ontology
- Mint URIs for entities and convert the data into RDF
- Ingest data into a separate graph

Data Integration: The RDF Way

For each data source

- Analyse the data model used by the source
- Formalize the model as an OWL ontology
- Mint URIs for entities and convert the data into RDF
- Ingest data into a separate graph



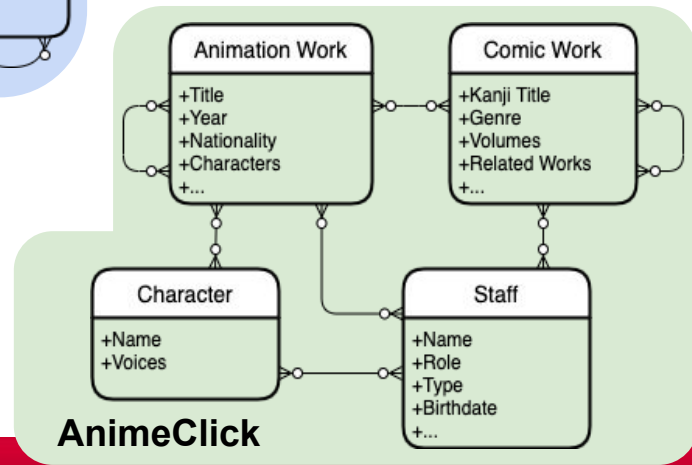
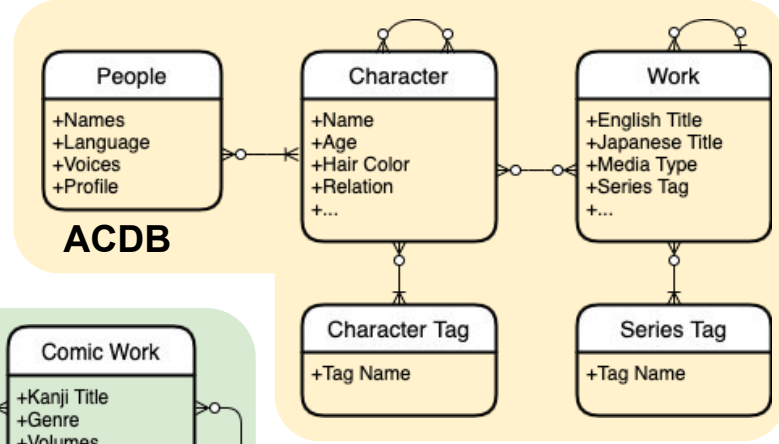
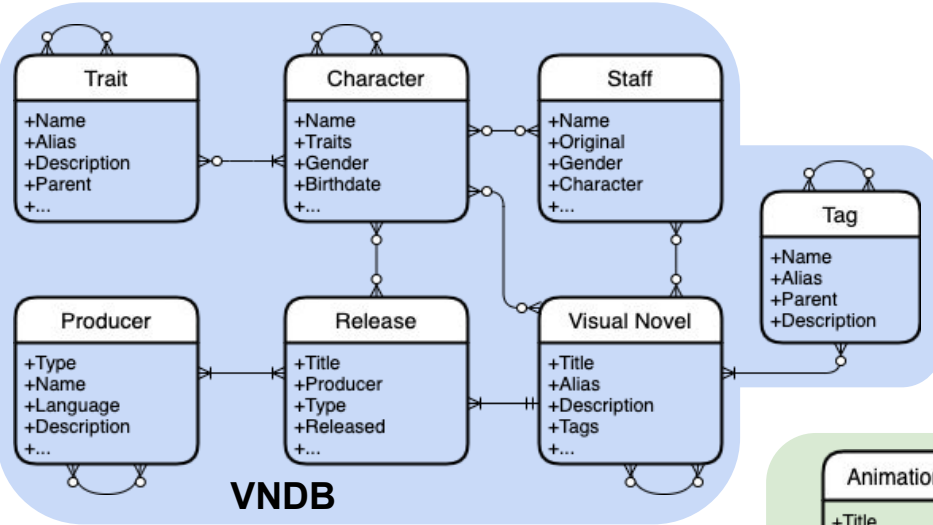
Data Integration: Analysis

- What are the core entities across all sources?
- How are they described?
- What are their relationships?

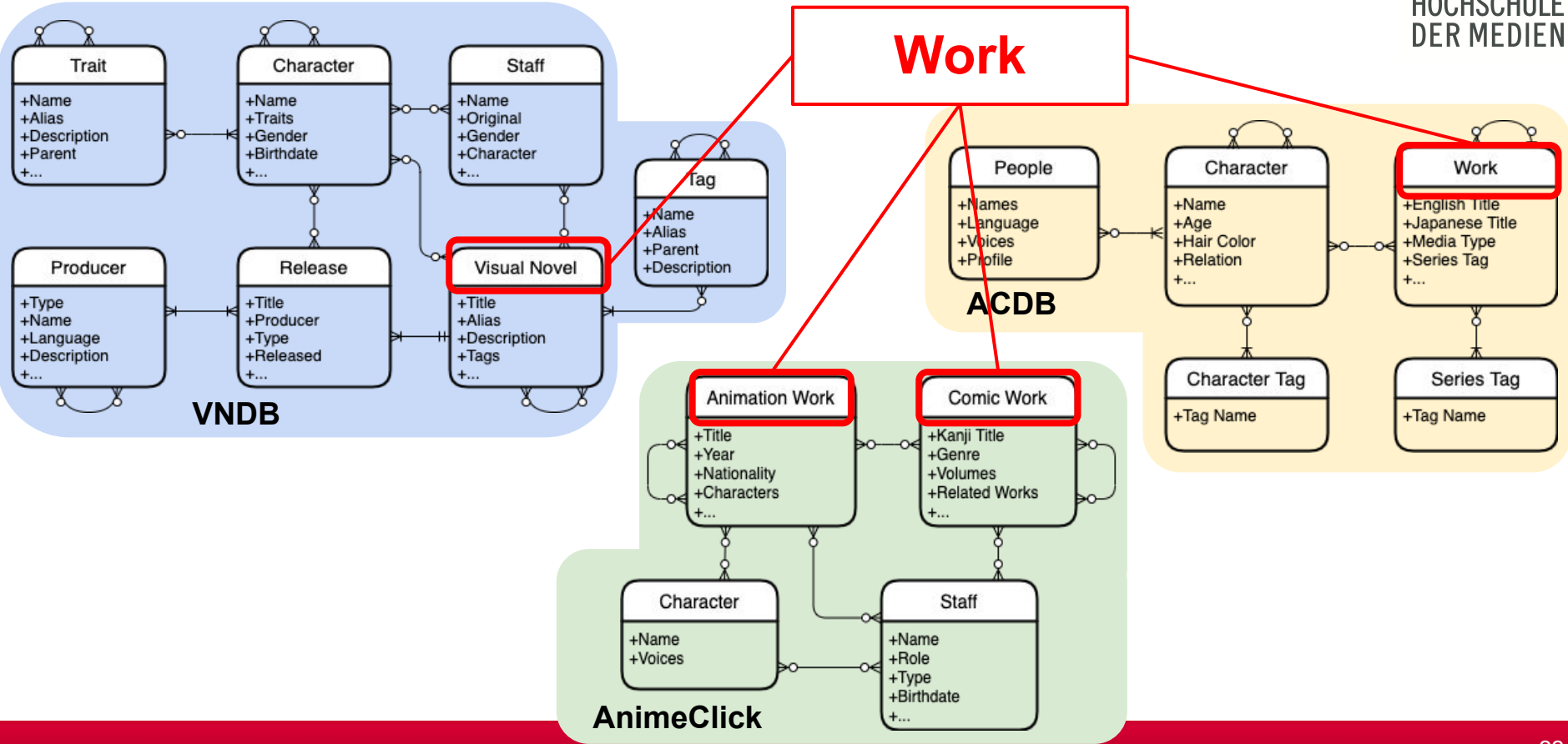
→ Can we create a unified view of the domain?

Developing a unified ontology for the JVMG knowledge graph: A pragmatic approach

Examining the ontologies of the sources

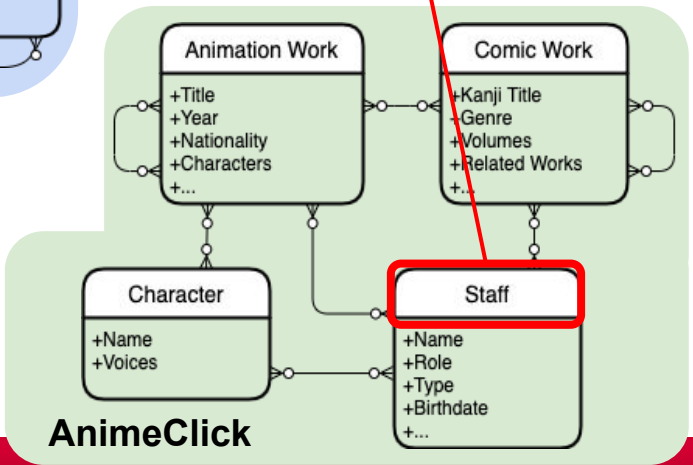
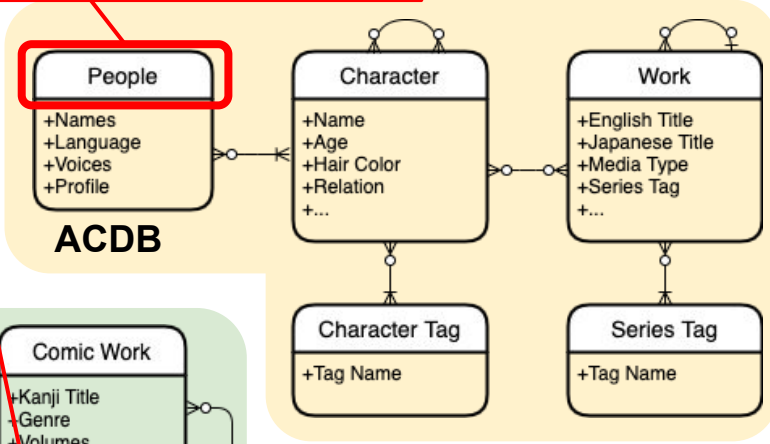
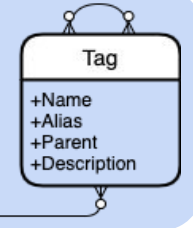
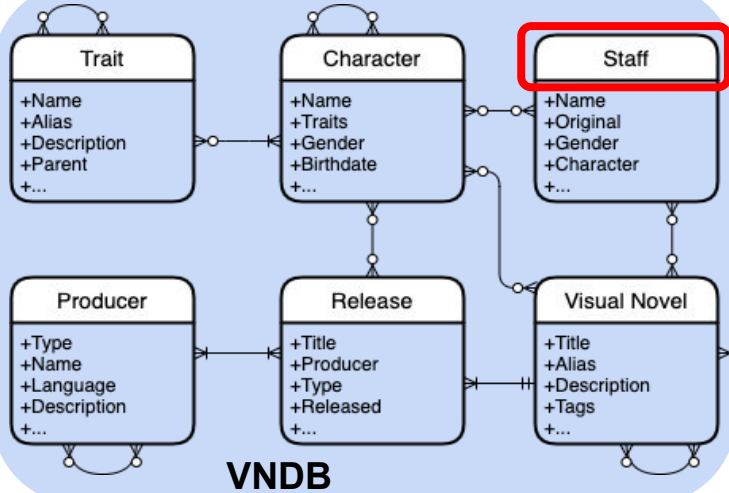


Examining the ontologies of the sources

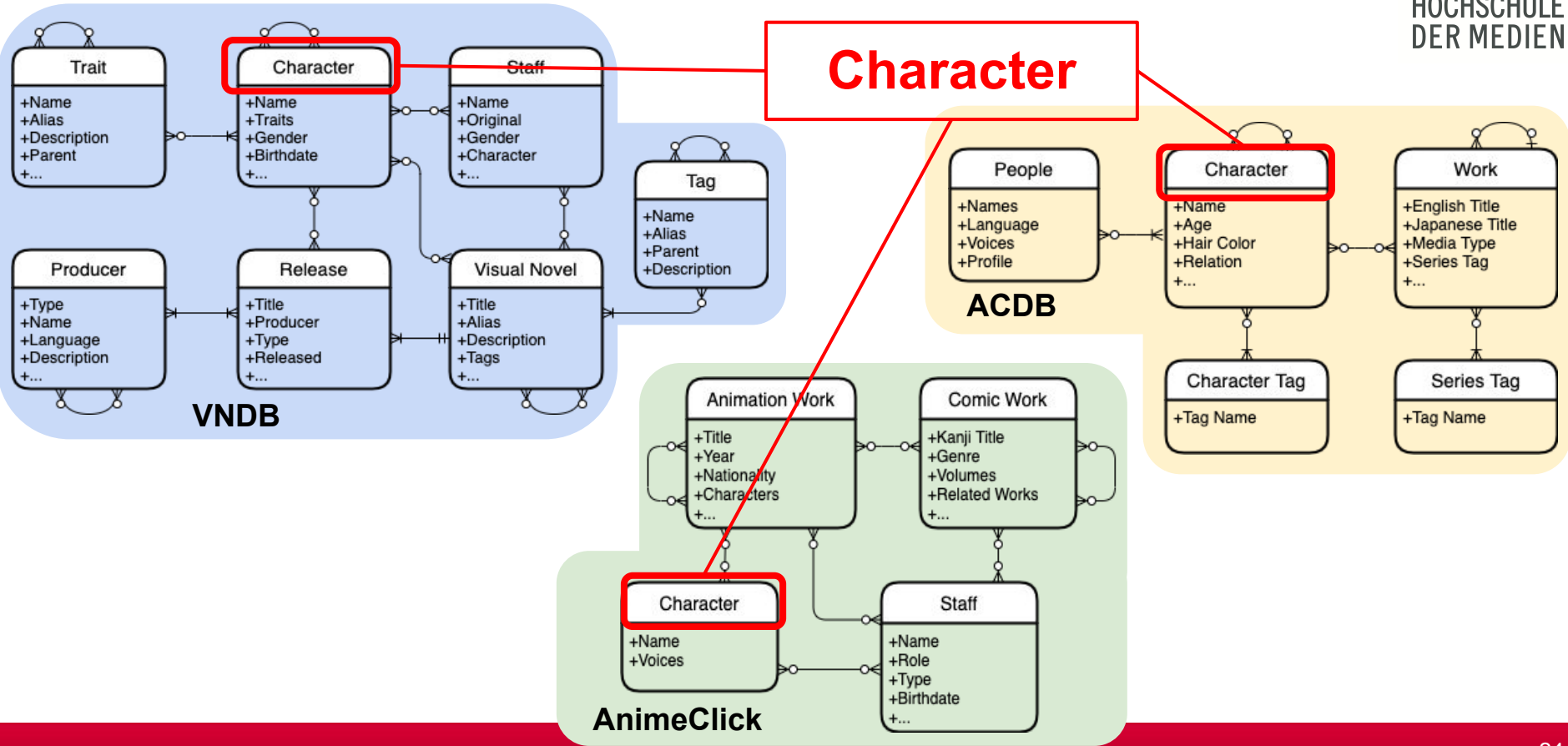


Examining the ontologies of the sources

Person (creators)



Examining the ontologies of the sources



Needs of the JVMG ontology

- Has to be **feasible**
- Has to fit the **characteristics of the domain**
- Has to fit the **needs of the researchers**
- Has to be **extendable**
- Potentially compatible with FRBRoo

Two phase – two layer approach

- Phase one: The **Core JVMG Ontology**
 - Small number of core entities to enable traversing the knowledge graph between its subgraphs
 - Needs to enable the connection of all subgraph elements to this layer
- Phase Two: The **Extended JVMG Ontology**
 - The full expression of the JVMG ontology
 - Not all subgraph elements will be connectable to this layer
- (Phase Three: Further extensibility)
 - Extending the JVMG ontology with more granular ontologies

Phase one: The Core JVMG Ontology

- **Work**

- Abstract entity describing a specific expression, like a manga or anime or game

- **Person**

- Literally a 1:1 representation of real people

Matching: Process

- Process starts with media works entities
 - Character or people entities are ambiguous on their own, but less so if connected to a work
- Matching is based on Japanese title (kanji form) **only**
 - This data point is available for all entities
 - Quality assessment has shown very accurate entries
 - Other information (e.g. dates) might not be consistent

Matching: Challenges

- Media works granularity of representation
 - TV series can be documented as a single entry or multiple seasons
 - Extra episodes can be part of the series or separate
 - Multi-part movies can be single or multiple entries
- Current workflow
 - Document such 1:n or n:1 matching relations
 - For the final knowledge graph, a decision has to be made for the representation

Phase one: The Core JVMG Ontology

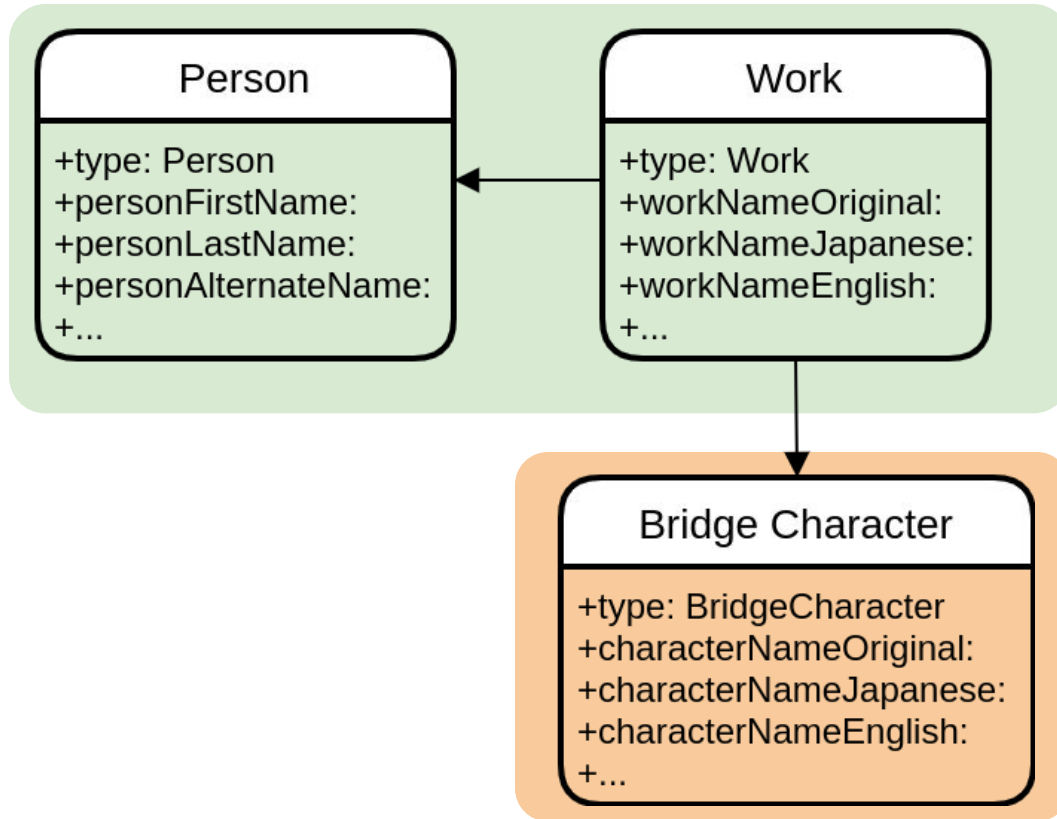
- **Observation**

- Characters in our data sources can be both specific for a given work or generic for a franchise
 - E.g. Naruto (Boruto TV Anime) ↔ Naruto (Naruto franchise)

- **Crutch: Bridge Character**

- Technical ontology element to enable the connection of both (meta-) **Characters** and **Realized Characters** without needing to distinguish between them

Phase one: The Core JVMG Ontology



The Extended JVMG Ontology: Works

- **Franchise**
 - Mediamixes, shared universes
- **Work** (already part of the Core JVMG Ontology)
- **Release**
 - Physical (CD, DVD, Bluray, LD, VHS, ...), broadcast or digital (Download, Stream) form of publication units

The Extended JVMG Ontology: Persons

- **Person** (already part of the Core JVMG Ontology)
- **Organization**
 - Juridical persons, like companies or agencies, NGOs, loose groups of doujin artists, etc.
- **Descriptor**
 - Abstract concepts that are used to describe works or characters, like tags, traits, keywords

The Extended JVMG Ontology: Descriptors

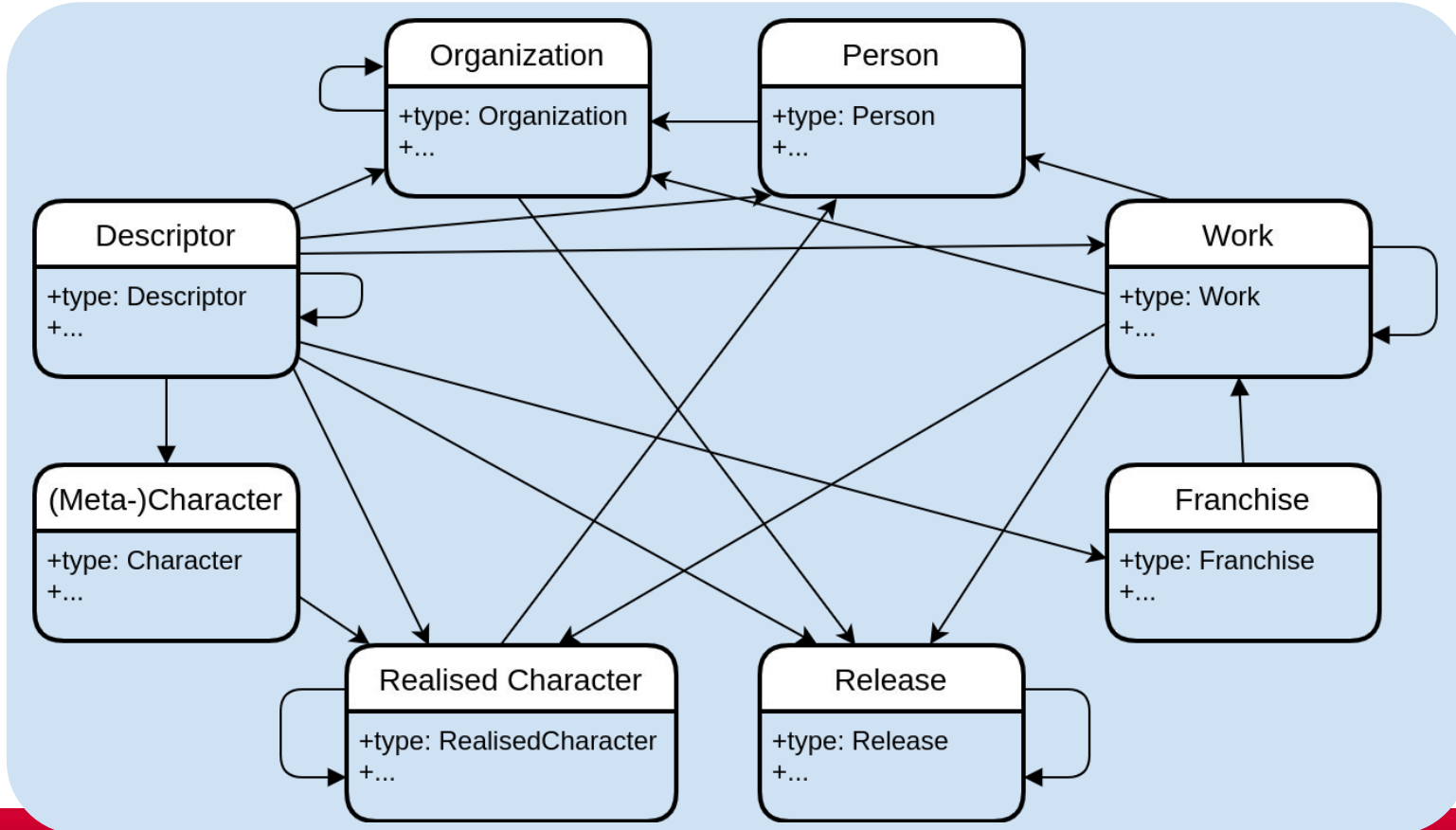
- **Descriptor**

- Abstract concepts that are used to describe works or characters
 - tags, keywords, genre for works
 - tags, traits for characters
- Organized as multiple concept hierarchies

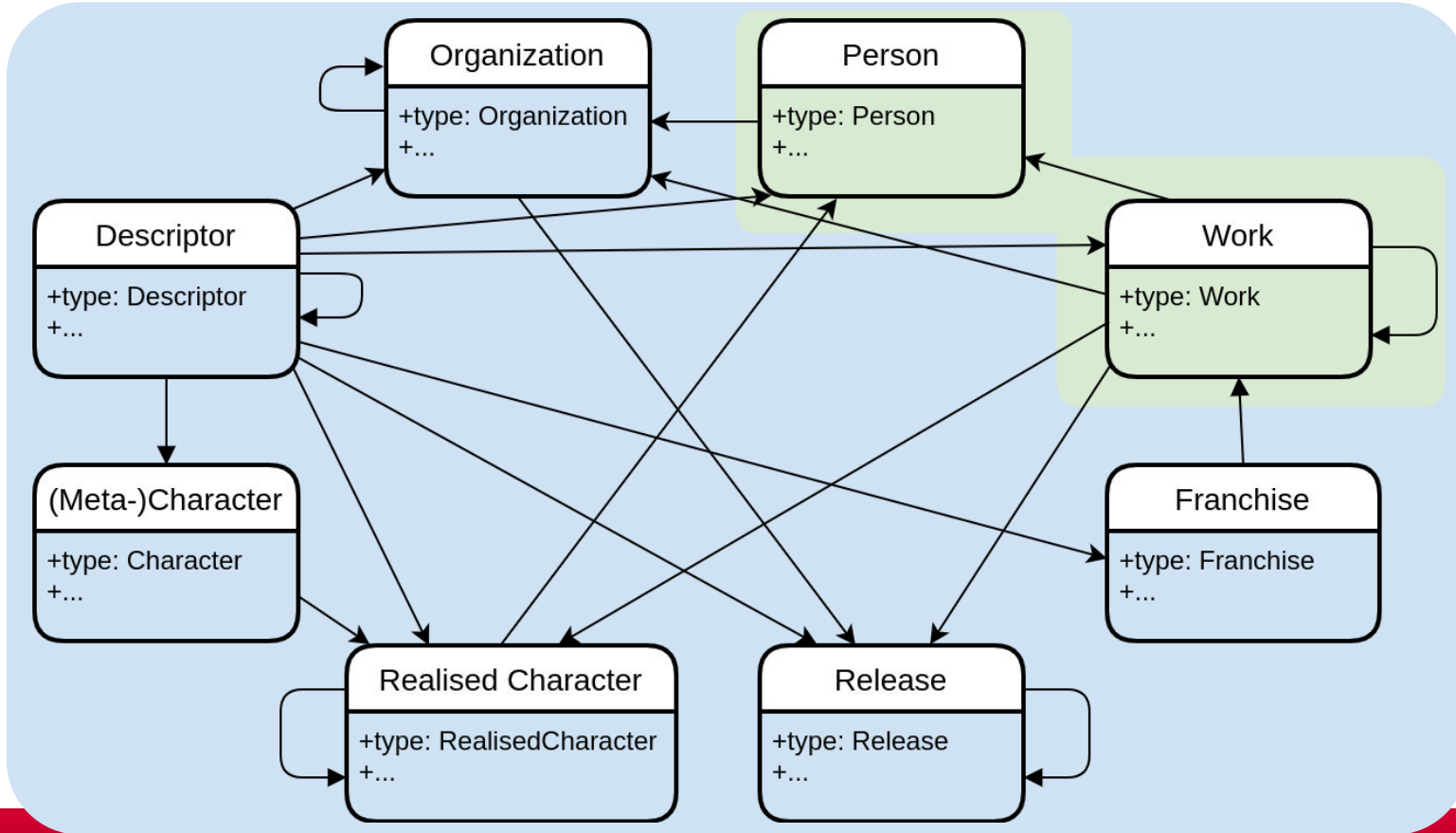
The Extended JVMG Ontology: Character

- **Character**
 - Abstract character that combines/connects all instantiated versions of it
- **Realized Character**
 - Distinct recognisable representation of a (visual) character in a specific work
- **Bridge Character (NOT part of the Extended JVMG Ontology)**

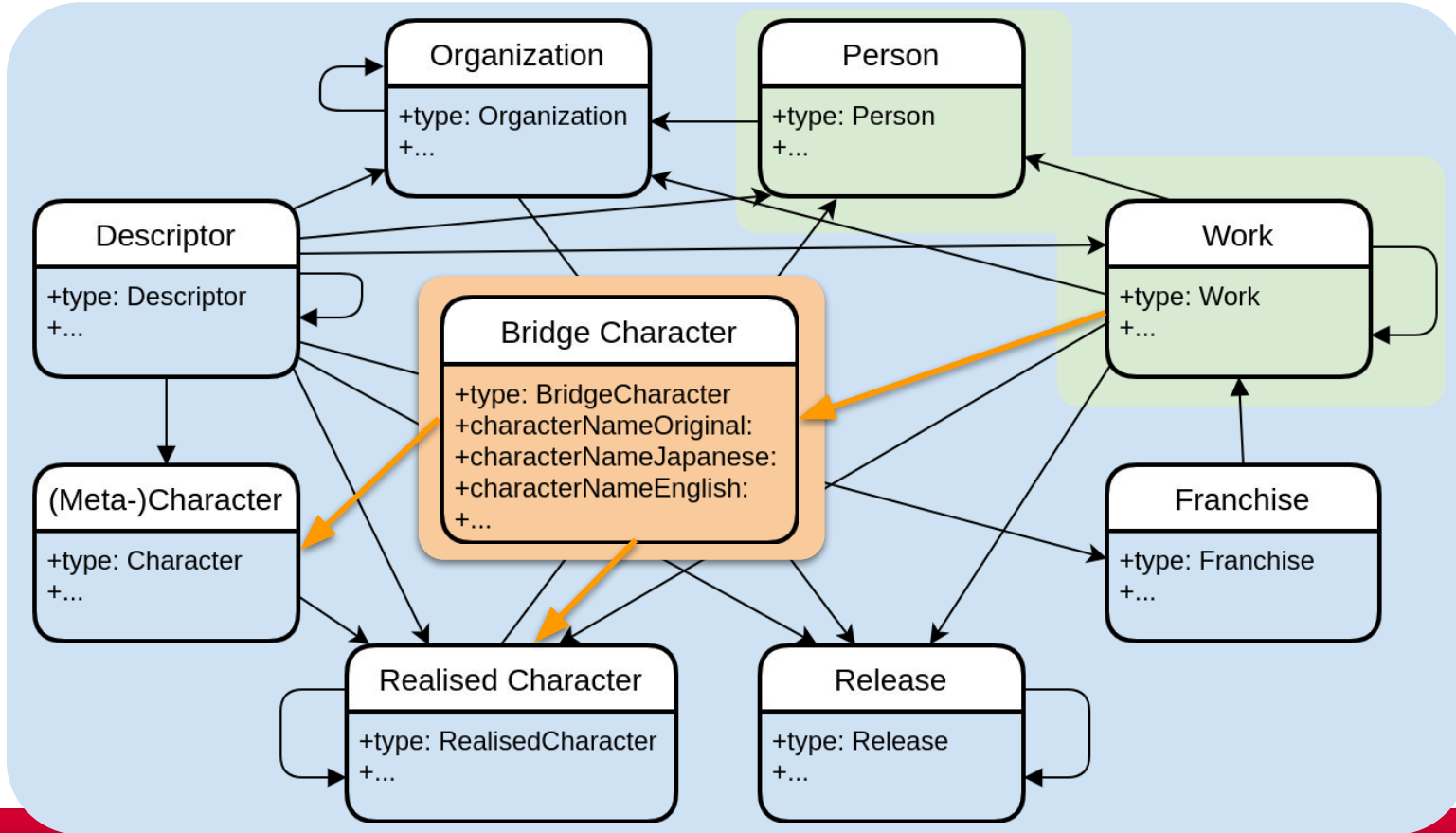
Phase two: The Extended JVMG Ontology



Phase two: The Extended JVMG Ontology



Phase two: The Extended JVMG Ontology



Summary

- **Heterogeneous** data sources
- Need to **create connections** between the subgraphs:
 - **Core JVMG Ontology**
 - Entities on this level should have corresponding data points in each subgraph if possible
- The **Extended JVMG Ontology** provides a more complete model:
 - Not all entities have corresponding data points in all subgraphs
 - Relationships are modelled with the minimal possible required relationships for all possible relationships to be then inferrable

Thank you for your attention!

Get in touch at: pfeffer@hdm-stuttgart.de

Visit our project website:

<https://jvmg.iuk.hdm-stuttgart.de/>

Visit the JVMG database:

<https://mediagraph.link/>