Introducing the Japanese Visual Media Graph

Magnus Pfeffer

Deutsche Forschungsgemeinschaft
German Research Foundation

Outline of the presentation

HOCHSCHULE DER MEDIEN

1. **Key characteristics** and **data sources** of the JVMG project
2. Measuring **data quality**
3. **Data integration** and **matching**
4. **Legal harmonization** of the data
5. The JVMG **web frontend**
6. Future directions
7. The **Tiny Use Case** workflow methodology

Introducing the JVMG project

HOCHSCHULE DER MEDIEN

- 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

The current JVMG project team

HOCHSCHULE DER MEDIEN

- Project partners
  - **Magnus Pfeffer**, Hochschule der Medien Stuttgart
  - **Martin Roth**, Hochschule der Medien Stuttgart / Ritsumeikan University Kyoto
  - Martin Hennig, IZEW Tübingen
  - Olaf Brandt, Universitätsbibliothek Tübingen

- Project team members
  - Zoltan Kacsuk
  - Tobias Malmsheimer
  - Christopher Zysik
  - Saskia Dreßler

Key characteristics of the JVMG project

HOCHSCHULE DER MEDIEN

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.

Source databases

HOCHSCHULE DER MEDIEN

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

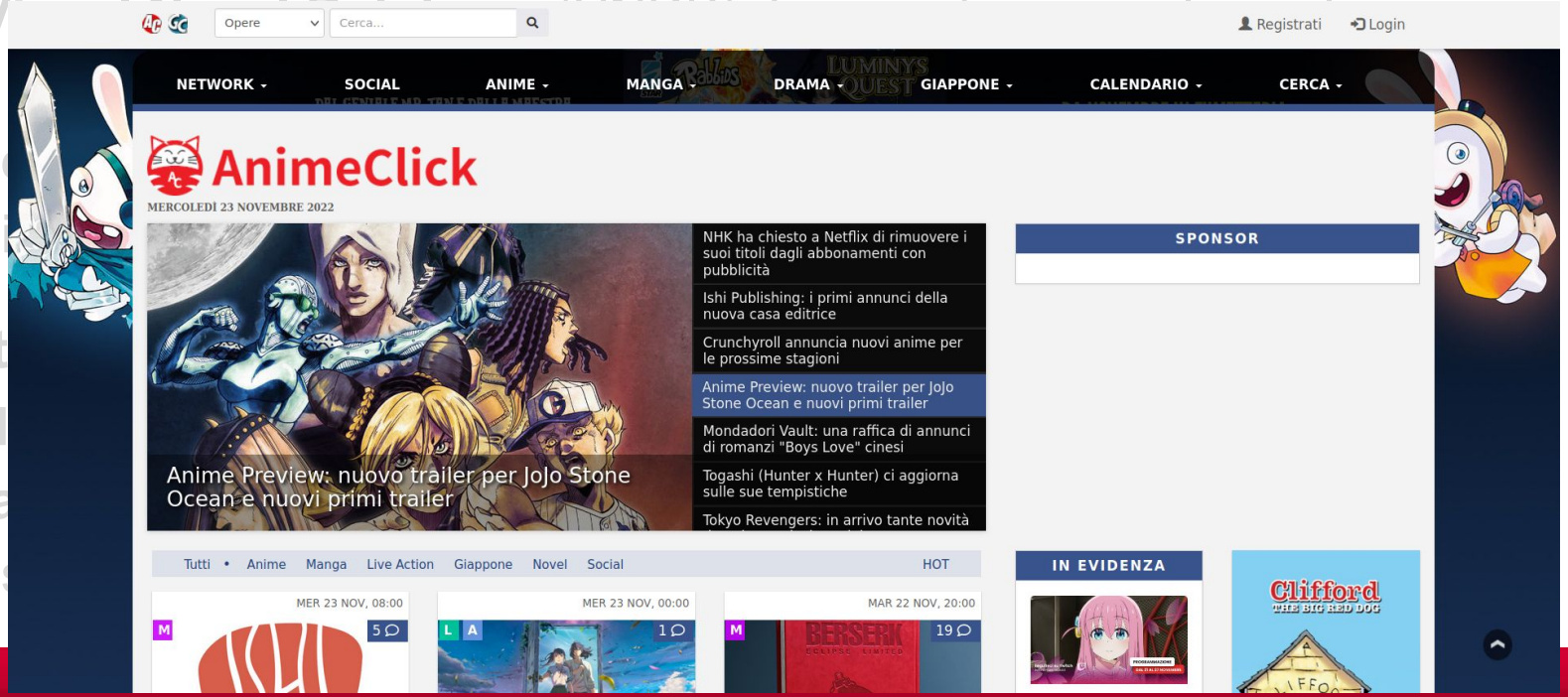
Other databases:

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

Source databases

Fan/enthusiast community databases:

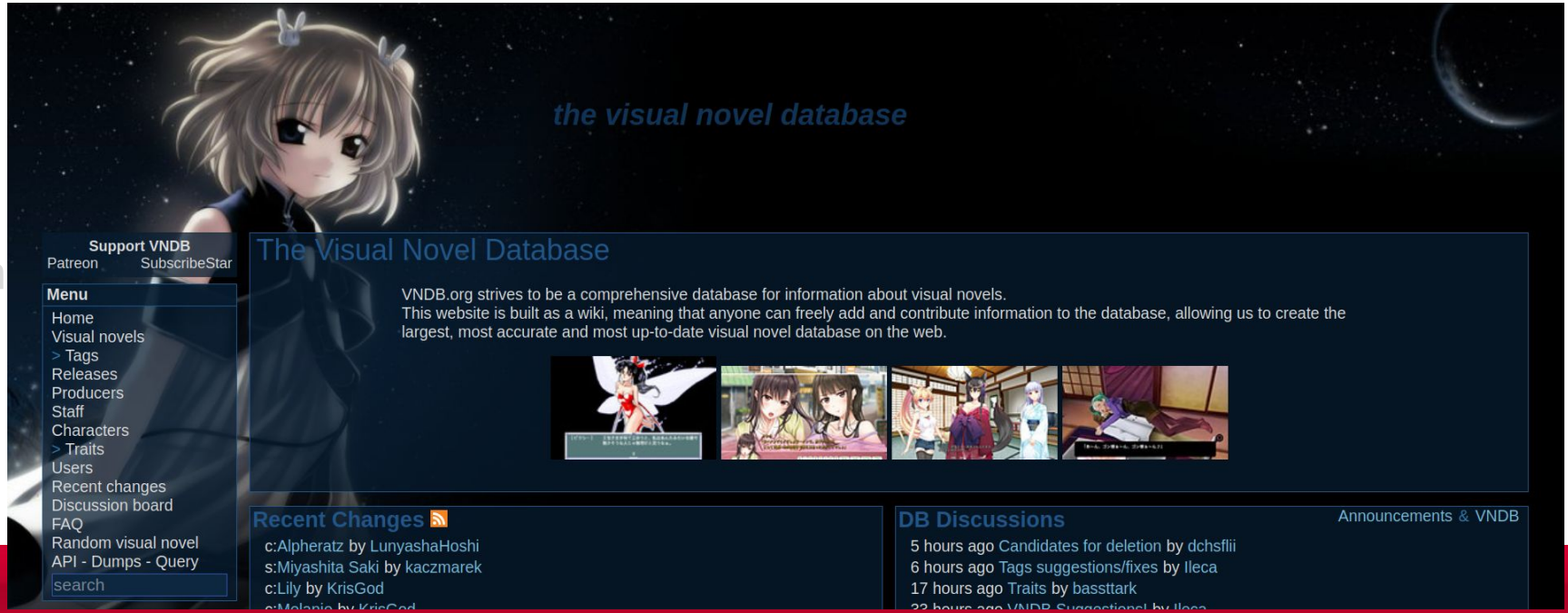
- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The V** [illegible] (VNDB): Focus on visual novels only
- **Anime** [illegible] domai
- **Other dat** [illegible]
- **Wikid**
- **Media** [illegible] games



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



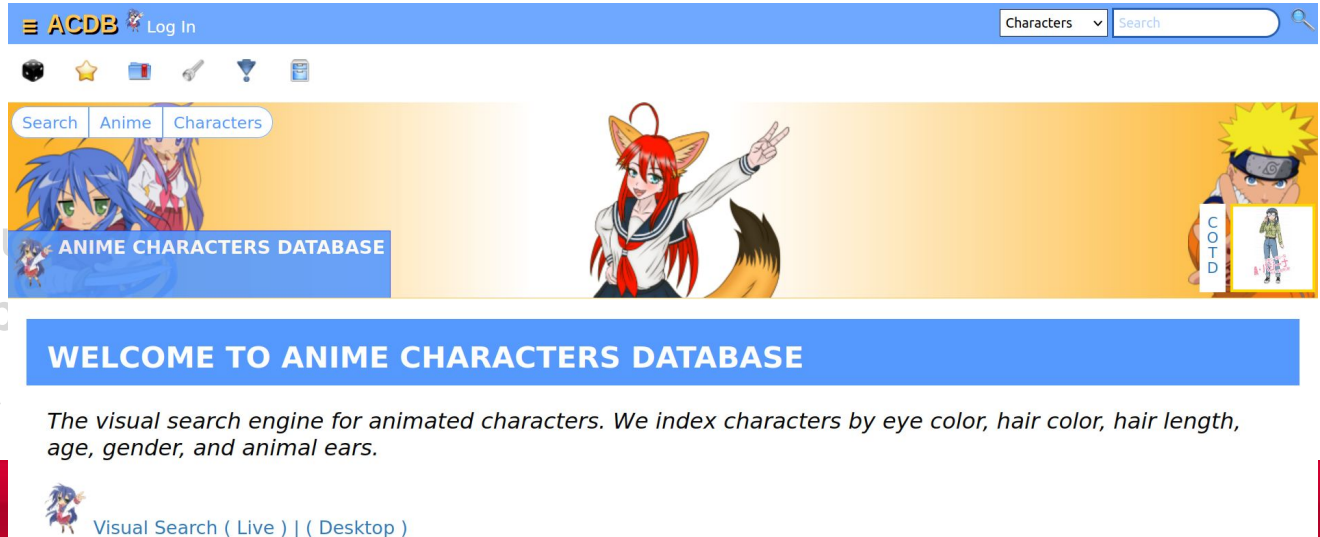
Source databases

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 focu
- **Media-Arts Datab** games and media



HOCHSCHULE DER MEDIEN

nd 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 in Japan

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

Entity and concept numbers

HOCHSCHULE DER MEDIEN

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	Anime 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 & manga 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		

HOCHSCHULE
DER MEDIEN

Measuring data quality

Assessing data accuracy

HOCHSCHULE DER MEDIEN

- **Random sample** of anime (or visual novel) titles
- **Sample sizes determined** so that statistical estimates can be drawn for the population parameters
- **Manual checking** of sample elements against ground truth or official websites, etc.

Accuracy main findings

HOCHSCHULE DER MEDIEN

- **Enthusiast community databases** have a **very high accuracy**
  - Most errors are only **typographical errors**
- **Wikidata** suffers from a variety of problems
  - Large number of problems with the **title contents**
  - Largest number of **missing data**
  - **Miscategorization** errors
- **Media-Arts Database** errors are mostly due to the **automatic processing** of the material and excess information added to the title field
  - **OCR** results need to be checked and edited manually
  - A lot of **disambiguation** information added to titles

HOCHSCHULE
DER MEDIEN

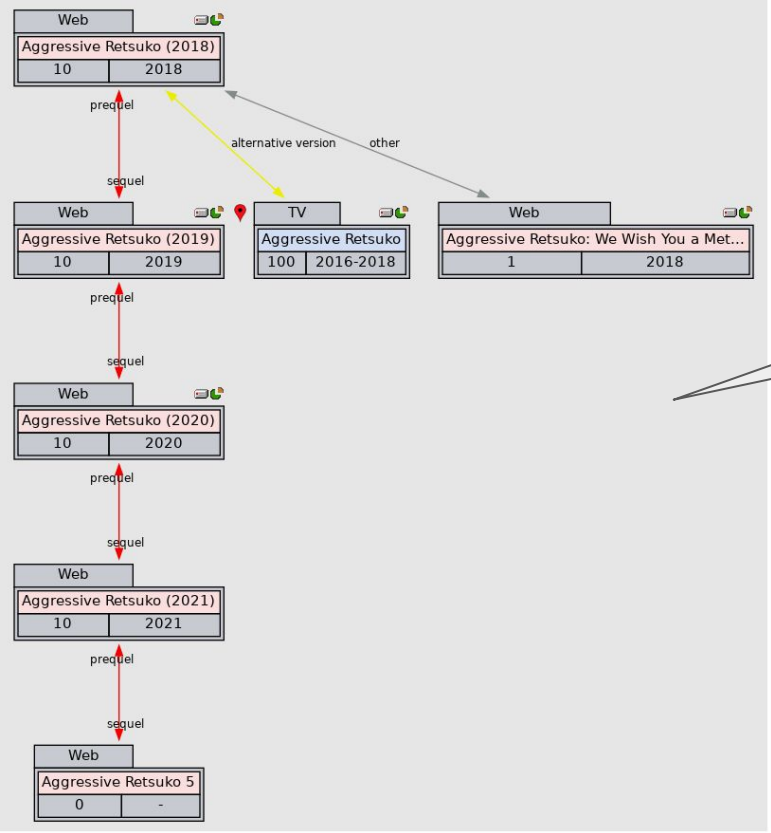
Data integration and matching

Data integration

HOCHSCHULE DER MEDIEN

- Data is **cleaned and preprocessed**
- All data transformed into **RDF** form
- **Unified JVMG ontology** integrating the individual ontologies
- **Connections** (e.g. matching) between the databases
- **Can be connected** to other linked data sources

Matching: Granularity example



AniDB:
4 web seasons with 10 episodes each
1 announced(?) web season
1 TV series with 100 (short) episodes
1 web christmas special

Image source: anidb.net

Matching: Granularity example

AnimeClick:
4 web seasons with 10 episodes each
1 TV series with 100 (short) episodes
1 web christmas special

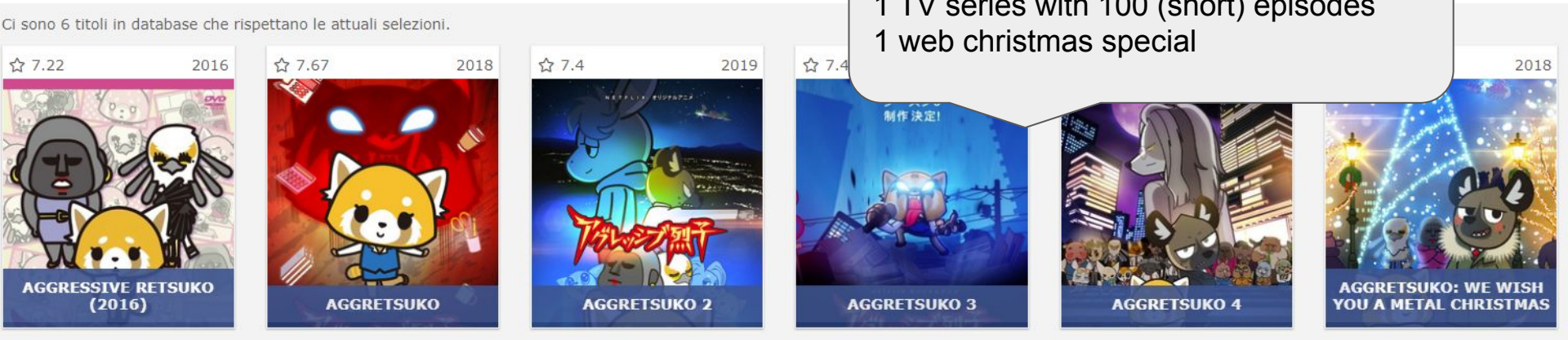


Image source: www.animeclick.it

Matching: Granularity example

Anime Characters Database:
1 TV Anime (no episode information)
1 Web ONA (no episode information)

Image source: www.animecharactersdatabase.com

Matching: Granularity example

HOCHSCHULE DER MEDIEN

	AniDB	AnimeClick	Anime Characters Database
4 web seasons	Yes	Yes	Yes
1 announced(?) web season	Yes	No	Maybe?
1 web christmas special	Yes	Yes	Maybe?
1 TV series	Yes	Yes	Yes

HOCHSCHULE
DER MEDIEN

Legal integration of the data

Problems and challenges

HOCHSCHULE DER MEDIEN

- **Licensing practices** of the communities
  - **Lack of awareness** of or disregard for copyright issues
  - Varying and often **incompatible licenses**

- **Concerns** of the communities
  - **Wholesale copying** of their work
  - **Traffic subverted** from their sites
  - **Lack of acknowledgment** of their work

Problems and challenges

HOCHSCHULE DER MEDIEN

- **Licensing needs** of the JVMG project
  - The license has to be **open, allowing all kind of research**
  - Need to find the **lowest common denominator**
  - Have to cover **most jurisdictions**

Overview of JVMG project data sources

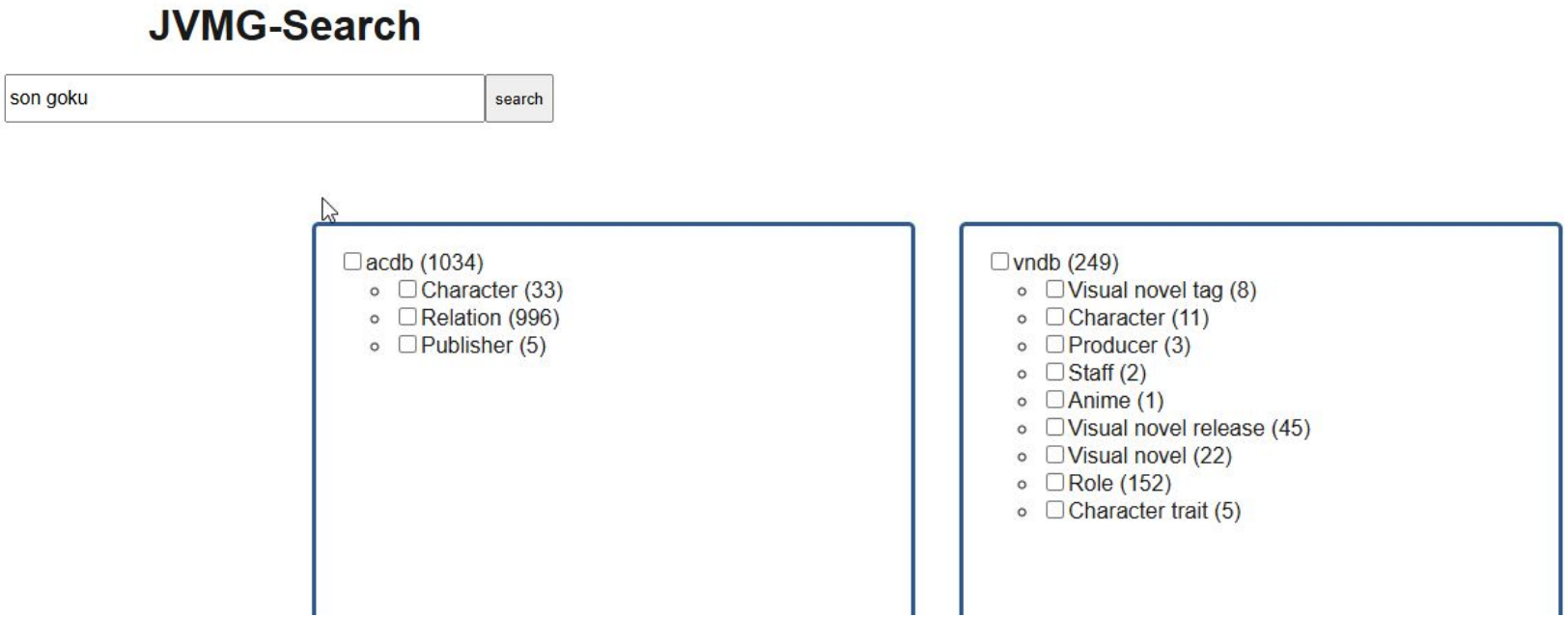
HOCHSCHULE DER MEDIEN

Data source	License	Compatibility with the CC BY-NC-SA 4.0 license
Anime Characters Database	-	CC BY-NC-SA 4.0 license provided for the JVMG project by **individual agreement** for the parts used in each case
AnimeClick	-	
The Visual Novel Database	ODbL	
Media-Arts Database	CC BY 4.0	**yes**
Wikidata	CC0	**yes**
AniDB (publicly available anime titles data dump only)	CC BY-NC-SA 4.0	**identical**

HOCHSCHULE
DER MEDIEN

The JVMG web frontend

Web frontend search:
results by subgraph and entity types



Web frontend presentation of the data

Hair Color (Hair)

Resource: http://mediagraph.link/vndb/trait/2

Graph: vndb hide

Property	Value
Alias en	Hair Colour en
Applicable en	false
Default spoil status en	0 en
Group en	Hair en
Name en	Hair Color en
Searchable en	false
Trait parent en	Hair en
is Trait parent en of 15	Black (Hair) en Blond (Hair) en Blue (Hair) en Brown (Hair) en Cyan (Hair) en Dyed (Hair) en Green (Hair) en Grey (Hair) en Multicolored (Hair) en Orange (Hair) en Pink (Hair) en Red (Hair) en Teal (Hair) en Violet (Hair) en White (Hair) en
VNDB link en	https://vndb.org/i2

Shinpi no Hou

Resource: http://mediagraph.link/jvmg/mediawork/A6132

Graph: http://mediagraph.link/graph/cluster hide

Property	Value
Source URI en	Shinpi no Hou x-jat 神秘の法
label	Shinpi no Hou x-jat
type	Anime en

Shinpi no Hou

Resource: http://mediagraph.link/anidb/aid/9081

Graph: anidb hide

Property	Value
AniDB link en	https://anidb.net/anime/9081
Official title en	神秘の法 ja
Primary title en	Shinpi no Hou x-jat
Synonym en	Happy Science Movie 6 en Happy Science Movie 6 x-unk The Mystical Laws en
label	Shinpi no Hou x-jat
type	Creative Work en

Graph: http://mediagraph.link/graph/cluster hide

Property	Value
is Source URI en of	Shinpi no Hou x-jat

神秘の法

Resource: http://mediagraph.link/madb/id/C412802

Graph: http://mediagraph.link/graph/cluster hide

Property	Value
is Source URI en of	Shinpi no Hou x-jat

Graph: madb hide

Property	Value
actor	【Sho Shishimaru】獅子丸 翔 （KOYASU TAKEHITO） ／ 【Leika Chan/Theta】張 麗華/シータ （AYUMI FUJIMURA） ／ 【Konohana@no@Sakuya@Bime】木花開耶姫 （RYOKA YUZUKI） ／ 【Utica】ユチカ （SHIN@ICHIRO MIKI） ／ 【General】ジェネラル （HIROSHI YANAKA） ／ 【Urano】浦野 （HIROSHI TSUCHIDA） ／ 【Julia】ジュリア （SATSUKI Y
additionalType	AN
contributor	[原作]大川隆法 ／ [脚本]「神秘の法」 シナリオプロジェクト ／ [Director]今掛勇 ／ [Character Designers]佐藤陵 ／ [Character Designers]須田正己 ／ [Art Director]川口正明 ／ [Music]水澤有一（MIZ MUSIC）
datePublished	2012-10-06

Web frontend intro page

Inspired by the DraCor (dracor.org) and GOLEM (golemlab.eu) projects

HOCHSCHULE DER MEDIEN

Overview

Anime Characters Database Graph

Character A character found in ACDB.	131029
Publisher Within ACDB, Publisher refers to a title or series to which a character belongs, and not an entity that is engaged in a publishing role.	12381
Tag A class for descriptive details about an anime or manga.	331
Character Tag A class for descriptive details about a character	1240
Individual A real-world individual associated with a character.	9308

AnimeClick Graph

Character Character of a manga or anime	57579
Author Author of an anime or manga	59502
Staff A person or group involved in the creation or publication of a creative work.	187017
Episode Single episode of an anime	212396
Creative work Creative work includes visual novels and animes	36604
Tag A class for descriptive details about an anime or manga.	2885

Media Arts Database Graph

Agent	69119
Animation TV program	172160
Animation video package	65919
Game package	49684
Manga book	380226
Manga other	9286
Media Art exhibition or performance	16670

Visual Novel Database Graph

Anime Indicates the anime of a visual novel	1831
Visual novel A specific visual novel.	45612
Character A visual novel character.	126578
Staff An individual staff member that has contributed to the creation of a visual novel.	33267
Visual novel release Physical or online release of a visual novel	107961
Character trait A descriptive character trait found in VNDB.	3218

HOCHSCHULE
DER MEDIEN

Future directions

Currently in development

HOCHSCHULE DER MEDIEN

- Functions
  - **New search interface** with assistance for **complex queries**
  - Data **export functions** for working with the data in other tools (Pandas, Spreadsheets, Gephi, etc.)
- Data
  - Adding data from **tvtropes.org**
  - **Updating** to the newest version of the Media Arts Database data
  - **Integration** of all data in the shared JVMG graph
- Services
  - Online **documentation** and **tutorials** for typical use cases
  - In-person **JVMG Labs** for researchers in Stuttgart and Kyoto

Possible future directions

- **End of the current project phase** beginning of 2026
  - Further development with own resources or new partners?
- Integrating further data sources
  - Music and audio drama CD publications
  - **Light Novels**
  - Computer games
  - Further international communities
- Deepening of the descriptions
  - Media contents
  - Character attributes
  - **More granular breakdown of works** (For example: individual manga chapters or anime episodes)

HOCHSCHULE DER MEDIEN

The Tiny Use Case workflow methodology

General idea for the Tiny Use Case (TUC)

HOCHSCHULE DER MEDIEN

- Pioneered by the **diggr** (Databased Infrastructure for Global Games Culture Research) research project team
- Inspiration from **agile** software development principles
  - Cycle of continuous incremental innovations and assessments
- Each TUC is **3-4 months** long
- Each TUC confronts the data with a **specific question/aim**

The purpose of TUCs

HOCHSCHULE DER MEDIEN

- Exploring of the **data's usability and adequacy for research**
  - **Limits**, **quality** issues and **fit**
- **Empowering** the researcher
  - Needs of the researchers in relation to the **frontend**
- Finding a **common language**
  - Bridging disciplinary boundaries between library and computer science on the one hand, and humanities and social science on the other

Example TUCs with the JVMG knowledge graph

HOCHSCHULE DER MEDIEN

1. Testing one of the points from Hiroki Azuma's **"Otaku: Japan's Database Animals"**
2. **Exploring recurring patterns in character creation** in visual novel games
3. **Examining the concept of media mix** by looking at networks of co-appearing characters
4. **Census of characters** in Japanese visual media

Thank you for your attention!

HOCHSCHULE DER MEDIEN

Get in touch with us:

Email: pfeffer@hdm-stuttgart.de

Website: https://mediagraph.link/