



**Designing the Metadata
Model for the Aggregation of
MAG Metadata**

**- the Development of
Media Art Database**

Dr. Tetsuya Mihara

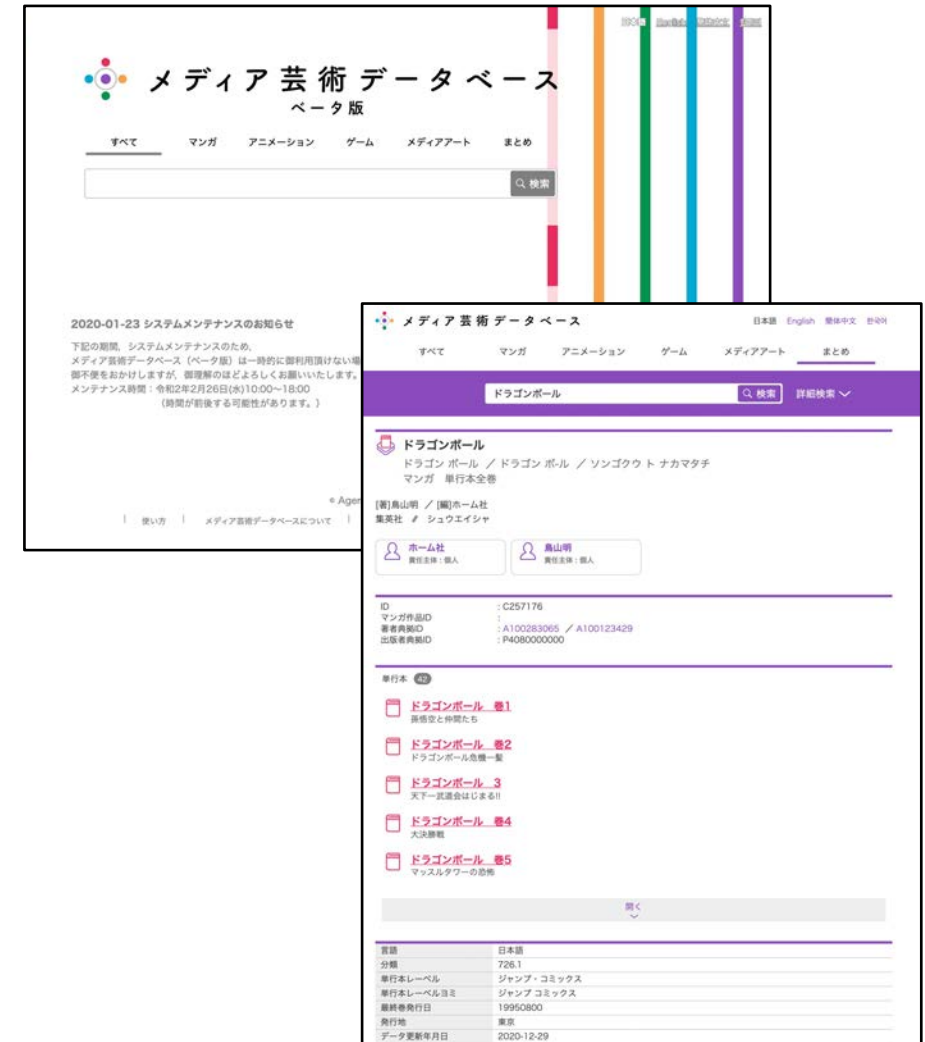
Self Introduction: Tetsuya Mihara

- Independent IT consultant, engineer
- Manga editor
- Researcher
 - Former position: Assistant professor ,Faculty of Library, Information and Media Science, University of Tsukuba
- Join the **Media-Art Database project** since 2013
 - as a specialist about **MAG metadata** and **information system**
- Research Interest: **Metadata for Manga**
 - **Linked Open Data, semantic Web, digital archives, digital humanities**

Media Art Database (メディア芸術データベース, MADB)

<https://mediaarts-db.bunka.go.jp/>

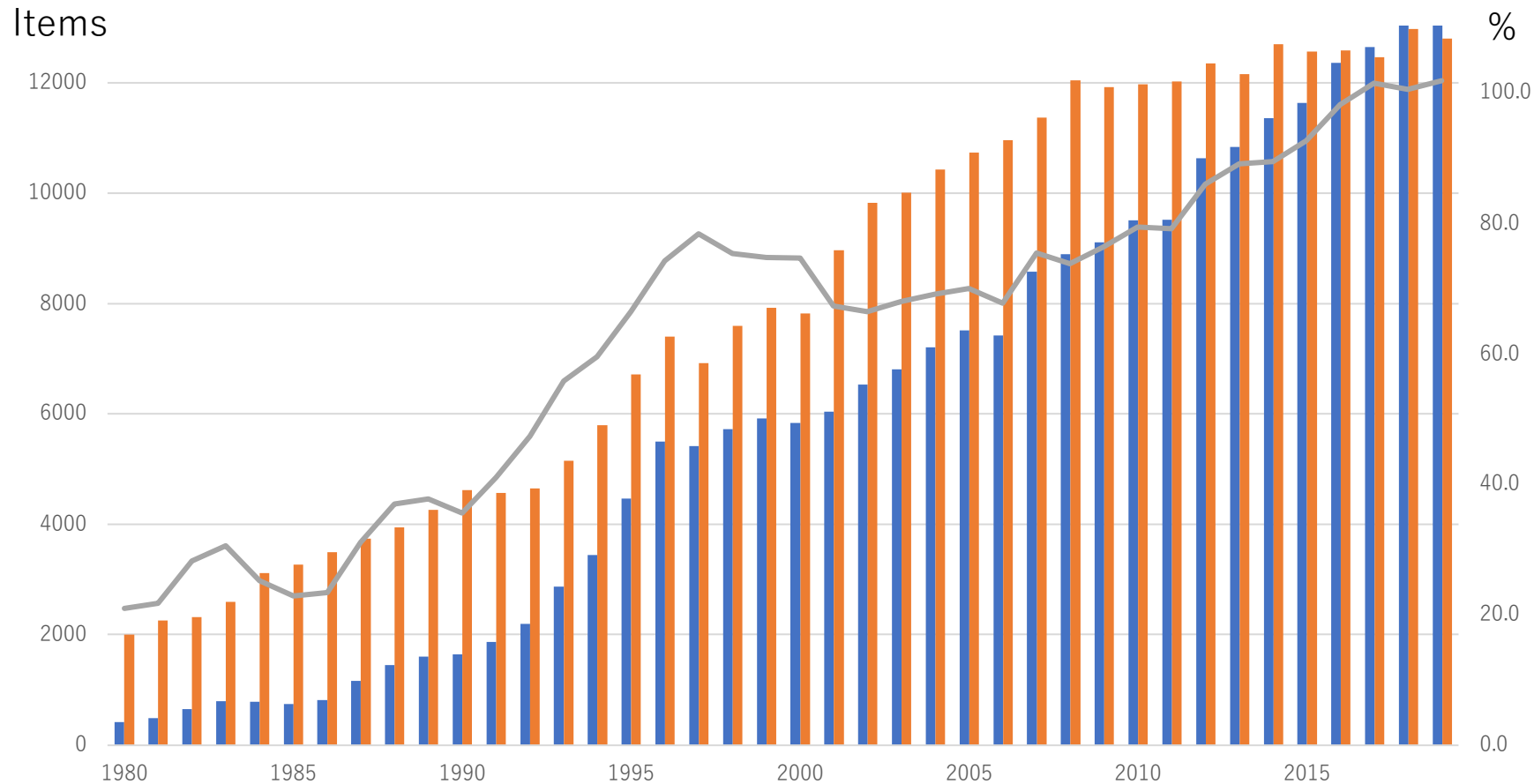
- The **unique national database** for Media Arts produced in Japan
- Consists of 4 category for representative media of Media Art
 - Manga
 - Animation
 - Video game
 - New media art
- Collect data to **cover information of Media Art produced in Japan**
 - Bibliographies about over **480 thousand Manga books and magazines, 9 thousand Anime series, 48 thousand video game titles** are recorded



The Challenge of MADB Project

- Media Art resources are **less collected and archived**
 - They were **not considered worth collecting archiving formally until 25~30 years ago**
- The Concept of “Media Art” is **quite vague and diverse**
 - **The term is defined by the law** “Basic Act for the Promotion of Culture and the Arts” (文化芸術振興基本法) enforced in 2001
 - nearly equal to “**MAG/ACG + pop culture + pop art + digital art**”
- **Connect collections** about Media Art resources
 - Database development started to **gather cataloging records of libraries and museums** which have Media Art collections

The ratio of collected manga book by National Diet Library (NDL) : 1980~2019



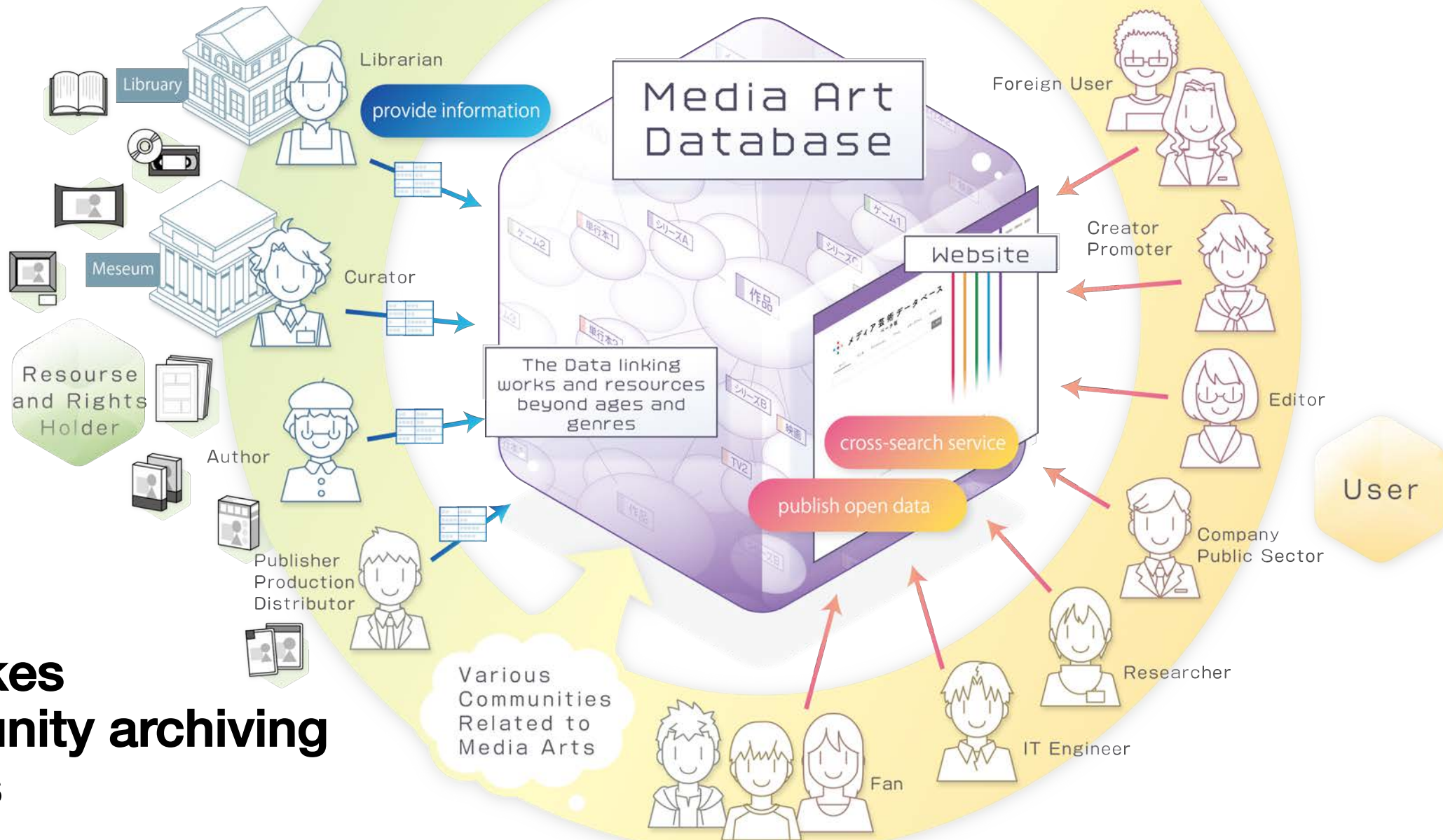
■ Number of manga book published in each year : Data from 出版科学研究所「出版指標年報」(1980~2019)

■ Number of Collected manga book by NDL : counted books with NDC「726.1 漫画, 劇画, 諷刺画」

■ The ratio : $\frac{\text{blue}}{\text{orange}} \times 100$

- About 60% manga books are collected by national memory institution
- Increase the ratio to over 80% to add collection of special library of manga
- Manga is the better case, others are worse...

Linking Archives and Communities of Media Art by the Data



**MADB makes
the community archiving
Media-Arts**

Item-centric VS Content-centric

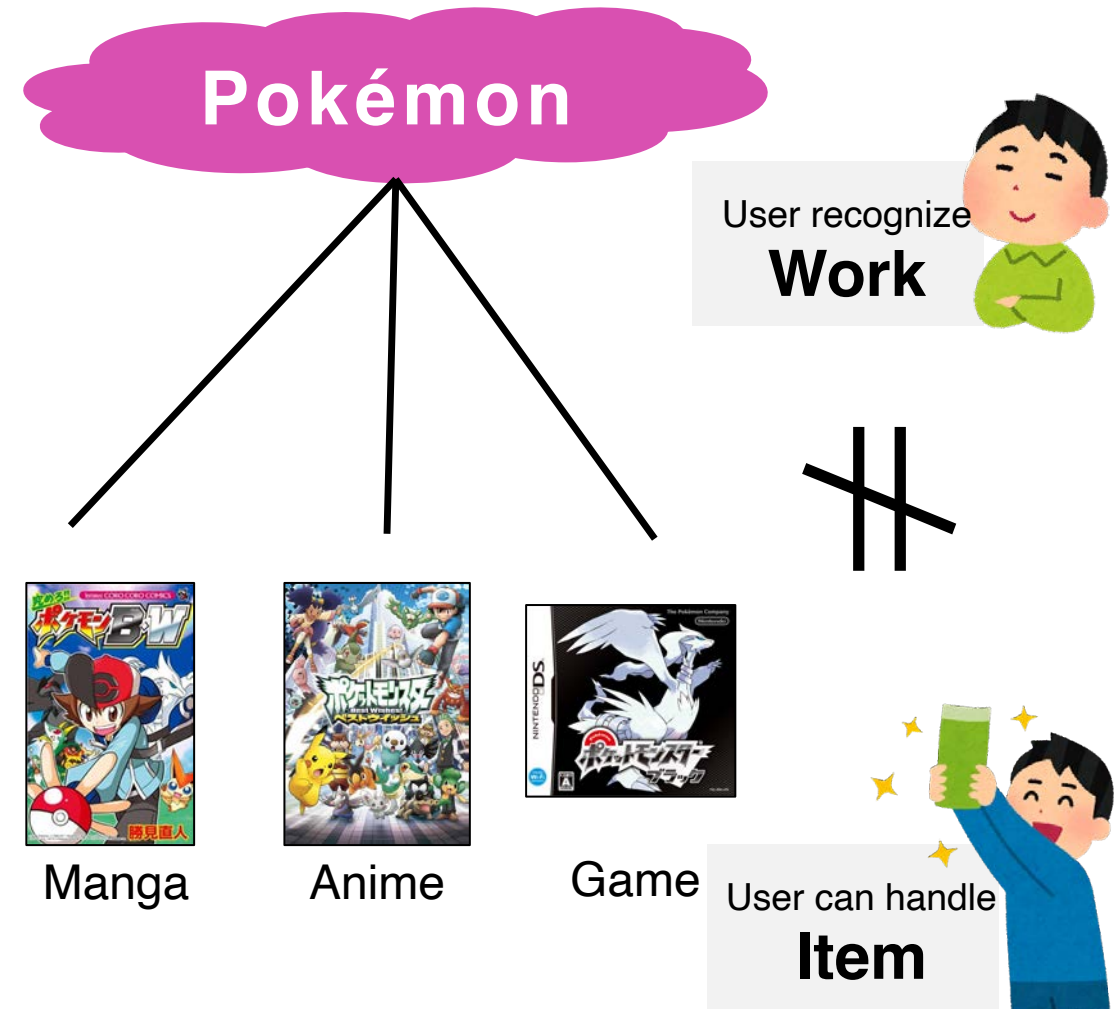
- **Item-centric**

- The data from collection is mainly **based on physical object**

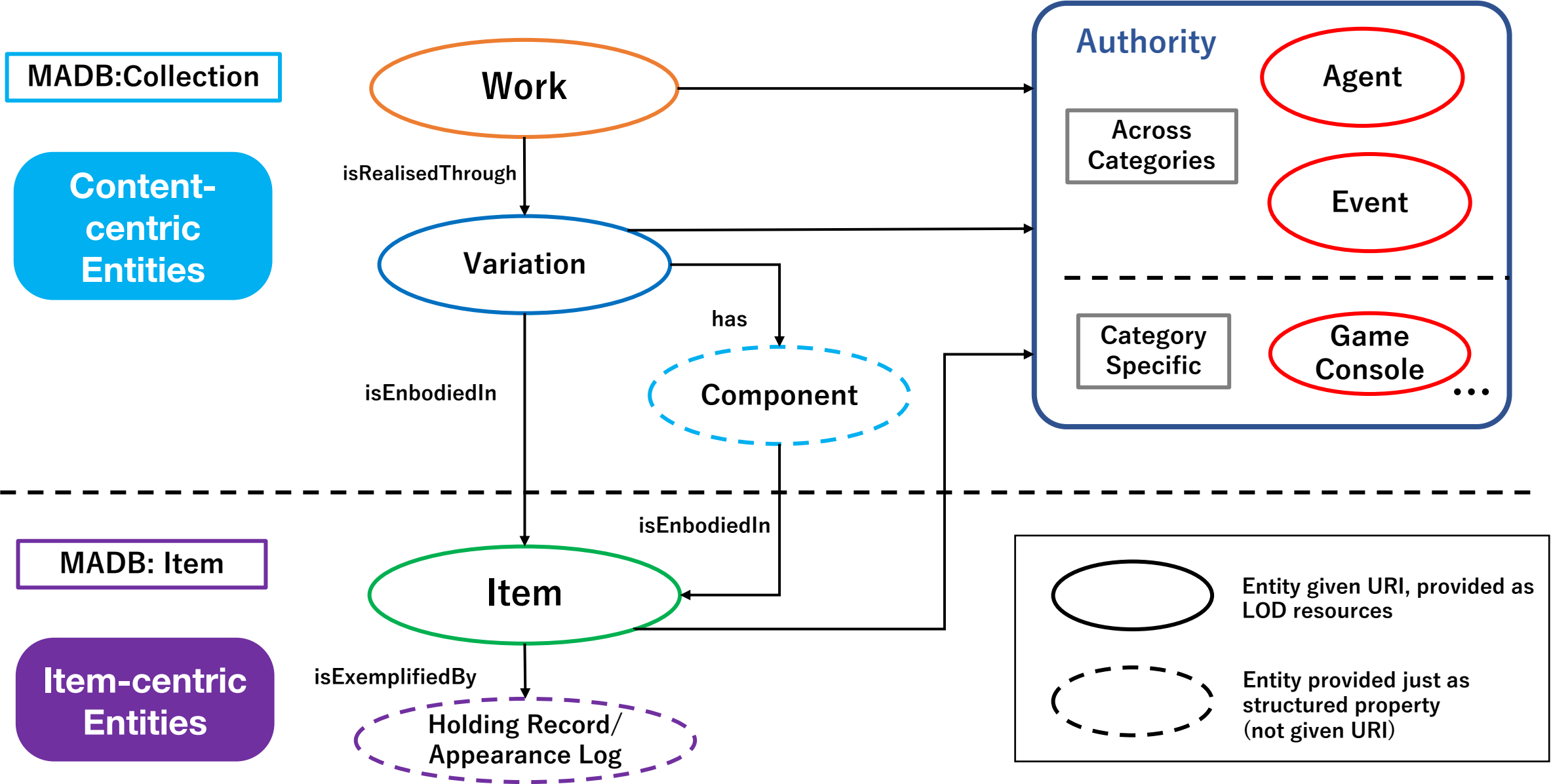
- **Content-centric**

- Common interests of users focus on the content, narrative, and intellectual topics
- Most typical: **Work**
 - FRBR, BIBFRAME...
 - e.g. **articles of Wikipedia** about Media Arts tend to be created based on work

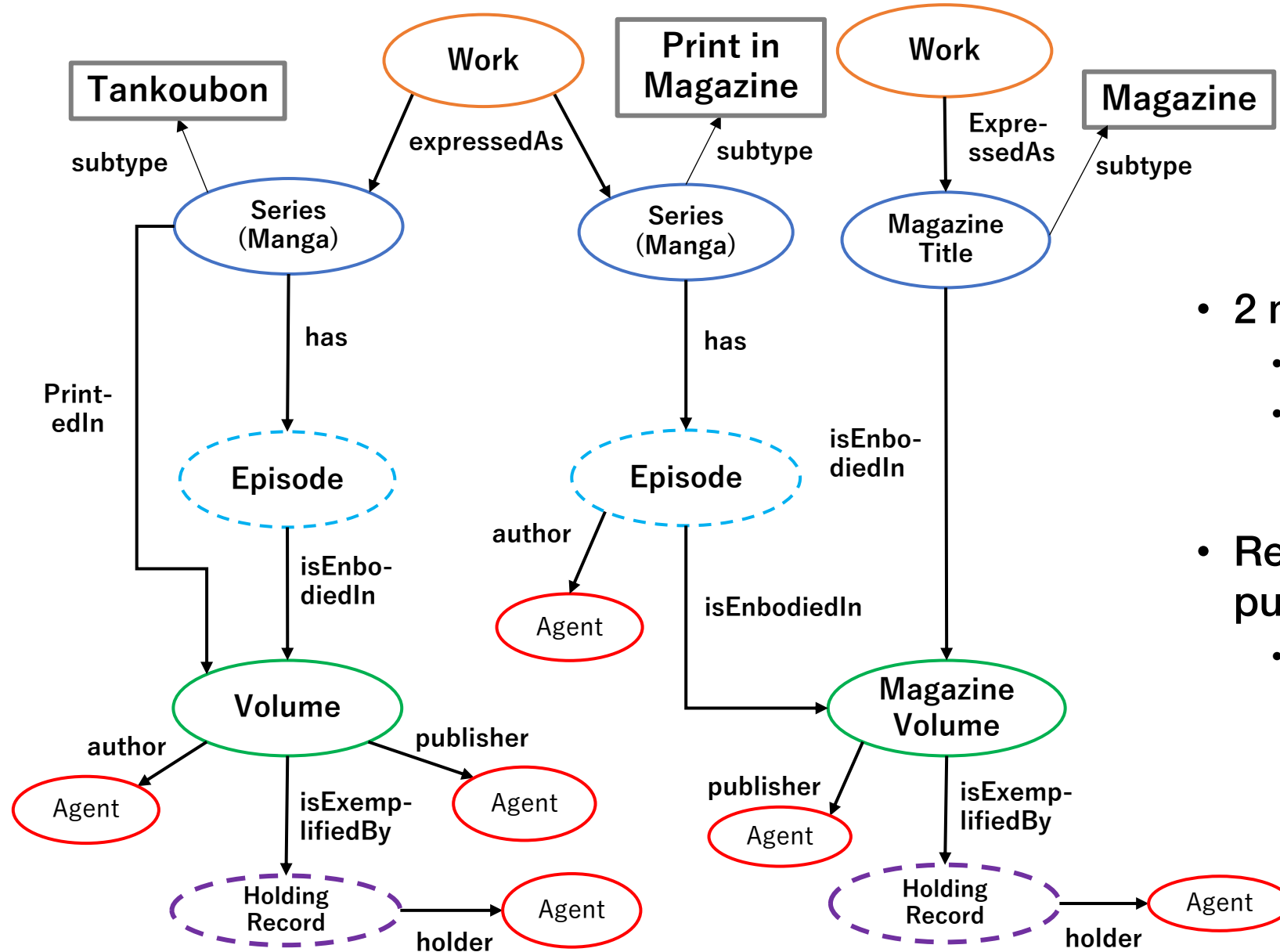
- The data model of MADB has both aspect and link them



Structure of the MADB Abstract Metadata Model

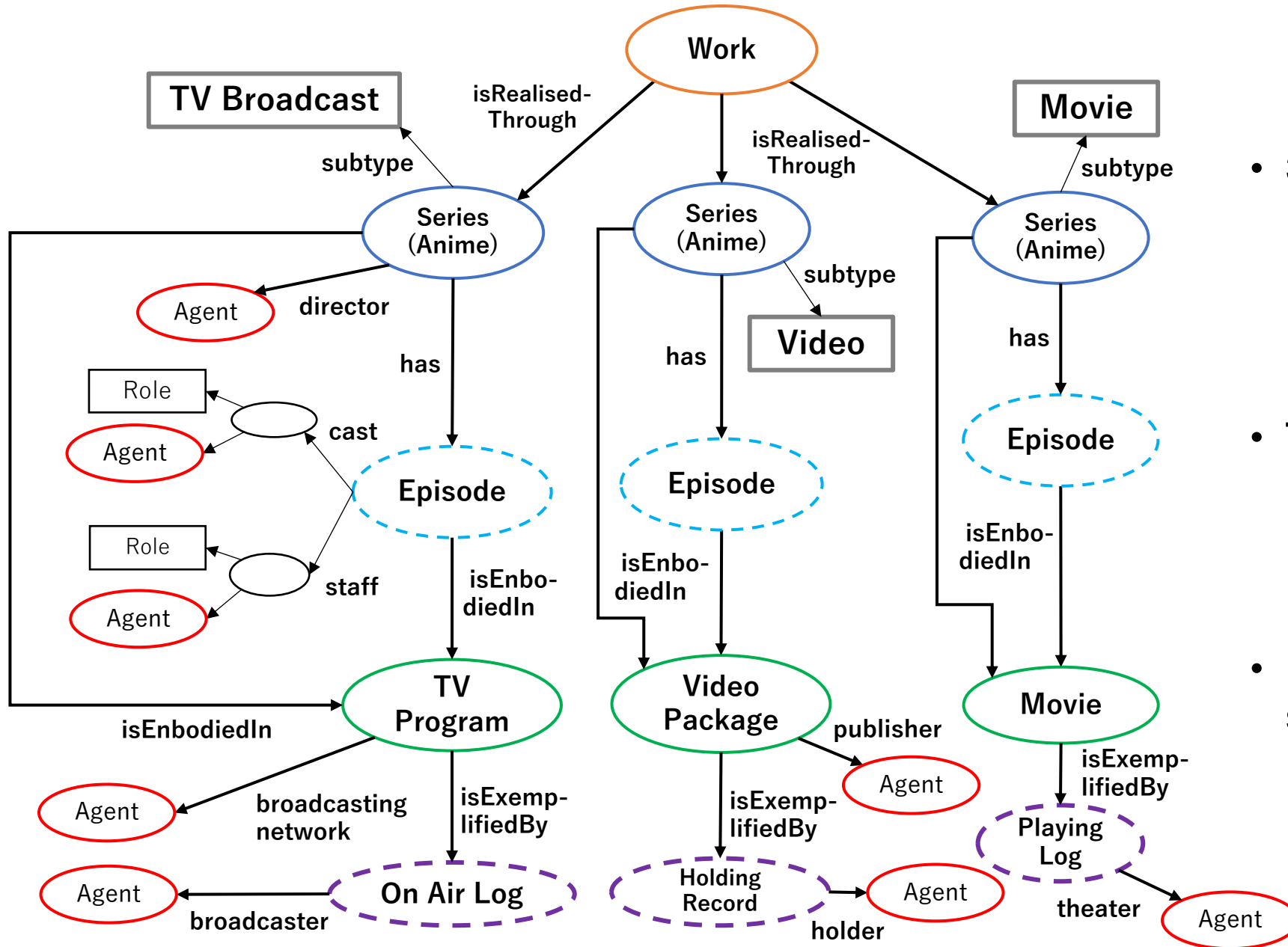


Data Model: Manga



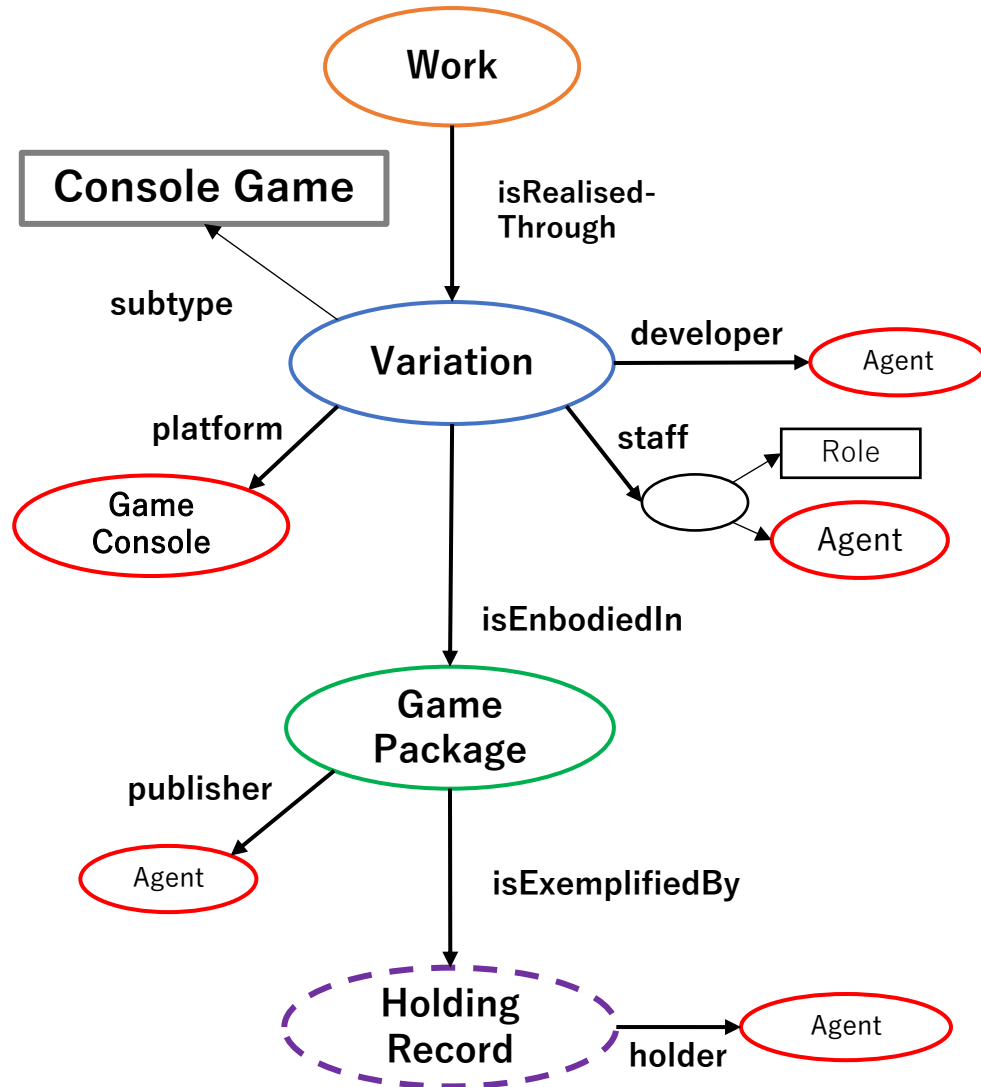
- 2 major materials:
 - Comic book (Tankoubon/paperback)
 - Magazine
- Represent the common publication pattern of manga
 - Almost manga works published in a magazine by a episode with other works first, then republished a book to collect some episode

Data Model: Anime



- 3 major media:
 - TV broadcasting
 - Video package
 - Movie
- Temporal object (TV broadcasting / movie) VS Item (video package)
- Its publication pattern is similar to manga
 - Temporal object is republished by item

Data Model: Videogame



- **Mainly focused on console game**
 - Unfortunately MADB has **less** the data of **arcade videogames** and **PC games**
- **Simpler than manga and anime**
- A work of videogame has **many variation for each platform** (NES, PlayStation etc...)
 - **Variation** \equiv **program**
 - **Some variations looks almost same, others are not**
 - new graphics, add/reduce contents...

Discussion

- **Metadata for Content-centric entities**
 - **Need the vocabulary**
 - to describe the **relation of items by their content** (copy, remake, remaster, repacking etc...)
 - **Collecting Web resources generated by users**
- **Digital Item (born digital)**
 - **Their appearance is more various** than physical object
 - **Easy to get metadata** because it is well developed for the distribution on the web
- **Data Development**
 - **Data aggregation** is always complex
 - **Identification entities** from various data source is sometime difficult