

# Combining metadata on works and characters from heterogeneous sources on anime and fanfiction in the cooperation between the GOLEM and JVMG projects

Xiaoyan Yang,<sup>1</sup> Zoltan Kacsuk,<sup>2</sup> Federico Pianzola,<sup>1</sup> Magnus Pfeffer<sup>2</sup>

<sup>1</sup> University of Groningen

<sup>2</sup> Stuttgart Media University

## Abstract

In the 21st century, the proliferation of digital and social media has markedly expedited the transmission of cultural phenomena, leading to an increasingly rapid exchange of information among diverse communities. Contrasting with the past, where cultural consumption experiences were largely isolated and confined to individual or small, real-world groups (such as reading physical books or attending concerts in drama houses), contemporary practices have evolved to embrace instantaneous and extensive communication through the internet. This paradigm shift has given rise to the formation of online fan communities [1][2][3] where people easily reach out to others who share the same interests with them for certain cultural products.

These communities proactively establish extensive, high-quality databases and inventories encompassing both original and fan-created works, alongside archives dedicated to the latter (refer to [4][5] for examples) with various data organization strategies. These strategies are designed to aid in the identification of works possessing specific attributes or bearing similarities to others in certain respects [6][7]. Numerous studies have highlighted the congruence between these methods and the principles and practices of library and information science, while also acknowledging the distinctive characteristics inherent in these information behaviors [8][9]. These online platforms are invaluable for research not only into fan information behavior [10][11], but also for broader inquiries related to fandoms [4][10], data sources [12], and fan-produced content [4,6]. Additionally, they offer insights into reading practices, communities, and the evolution of literature [3][13][14].

The collaboration between the Graphs and Ontologies for Literary Evolution Models (GOLEM)<sup>1</sup> project (2023-2027), receiving funding from the European Commission (ERC StG), and the Japanese Visual Media Graph (JVMG)<sup>2</sup> project, supported by the German Research Foundation's (Deutsche Forschungsgemeinschaft, DFG) e-Research Technologies program, is fundamentally rooted in the way both projects build on data and archives aggregated by online fan communities. These resources are instrumental in constructing research infrastructures that cater to scholars focusing on popular narrative and fiction, and Japanese visual media, respectively. Both initiatives are engaged in the development of knowledge graphs within their individual realms. Given the shared research interests and the potential offered by linked open data, the teams from both projects are collaborating to interconnect their datasets. This cooperative endeavor aims to facilitate research inquiries that intersect the various concerns of these two domains.

The GOLEM project collects stories in six different languages (English, Spanish, Italian, Chinese, Korean and Indonesian) from various leading fanfiction platforms (e.g., Archive of Our Own<sup>3</sup>, Fanfiction.net<sup>4</sup>) and original fiction platforms (e.g., Wattpad<sup>5</sup>, Webnovel<sup>6</sup>) including the texts of stories, the metadata describing them, and the data about reader responses (comments/reviews), together with the canon information from Wikipedia and fan databases like Fandom.com<sup>7</sup>. The

---

<sup>1</sup> <https://golemlab.eu/>

<sup>2</sup> <https://jvmg.iuk.hdm-stuttgart.de/> and <https://mediagraph.link/>

<sup>3</sup> <https://archiveofourown.org/>

<sup>4</sup> <https://www.fanfiction.net/>

<sup>5</sup> <https://www.wattpad.com/>

<sup>6</sup> <https://www.webnovel.com/zh>

<sup>7</sup> <https://www.fandom.com/>

JVMG project, in collaboration with online fan communities, has created a knowledge graph on Japanese visual media, including anime, manga, computer games, and is working to further develop it to enable large-scale quantitative research on these domains.

The integration of metadata between the GOLEM and JVMG projects is beneficial for both of our projects. Primarily, such a connection fosters a comprehensive, cross-cultural understanding of narrative trends and audience engagement with a special media format: Japanese anime. As one vigorous flow in the popular culture, Japanese anime has transcended geographical boundaries, resonating with audiences across Asia, and the West [15], and fan works related to Japanese animation and manga occupy a significant part of the GOLEM database which makes it an extremely valuable object for the GOLEM project to study narrative, cultural evolution, and cultural transmission. By linking the GOLEM database to the JVMG knowledge graph, the GOLEM project can get easy access to the data about anime production, and anime character traits which will contribute to the further improvement of the GOLEM ontology and characterization analysis. Conversely, with the fanfiction creation data in the GOLEM database, the JVMG project can extend to track and compare the influence of Japanese anime across cultures and retroactively deepen its understanding of core elements that lead to the popularity of anime. The interconnection of these databases enables a multidimensional approach to studying fan communities, allowing for a more nuanced understanding of how different cultures interact with, interpret, and influence various narrative forms.

Ontology development is a central task for both projects to enable data integration and harmonization from different sources. With the aim to enhance interoperability and explicitness, the GOLEM project ensures its ontology is compliant with CIDOC-CRM<sup>8</sup> and LRMoo<sup>9</sup> standards. While the JVMG project employs a more bottom-up strategy to develop a unified ontology to harmonize the multi-sourced data in their knowledge graph. Both of these two teams face the difficulties of processing multi-dimensional data using non-uniform metadata as well as bridging the research interests of the researchers working on this domain with unstructured information embedded in the data. In this presentation we will focus on the way the two projects handle these challenges on the one hand and the way the two different approaches can complement each other on the other hand. Furthermore, we illustrate the way we have achieved interoperability between the two data sources on an example use case featuring source and fan work publication dates.

## References:

- [1] Jenkins, Henry, 2012. *Textual poachers: Television fans and participatory culture*. Routledge.
- [2] Hellekson, Karen, and Kristina Busse, 2006. *Fan fiction and fan communities in the age of the internet: new essays*. McFarland.
- [3] Pianzola, Federico, 2021. "Digital Social Reading: Sharing Fiction in the 21st Century."
- [4] De Kosnik, Abigail, 2021. *Rogue archives: Digital cultural memory and media fandom*. Cambridge, MA: MIT Press.
- [5] Pfeffer, Magnus, and Martin Roth, 2020. Japanese Visual Media Graph: Providing researchers with data from enthusiast communities. In: 2019 Proceedings of the International Conference on Dublin Core and Metadata Applications, eds. Koraljka Golub et al. Seoul: Dublin Core Metadata Initiative (DCMI), 136-41.
- [6] Johnson, Shannon Fay, 2014. Fan Fiction Metadata Creation and Utilization within Fan Fiction Archives: Three Primary Models. *Transformative Works and Cultures* 17, <http://dx.doi.org/10.3983/twc.2014.0578>.
- [7] Price, Ludi, 2019. Fandom, Folksonomies and Creativity: the case of the Archive of Our Own. In: *The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization*. Baden-Baden: Ergon Verlag, 11-37.
- [8] Price, Ludi, and Lyn Robinson, 2021. Tag analysis as a tool for investigating information behaviour: comparing fan-tagging on Tumblr, Archive of Our Own and Etsy. *Journal of Documentation* 77(2), 320-358.
- [9] Bullard, Julia, 2019. Curated folksonomies: three implementations of structure through human judgment. *Knowledge Organization* 45(8), 643-652.
- [10] Price, Ludi and Lyn Robinson, 2017. 'Being in a knowledge space': Information behaviour of cult media fan communities. *Journal of Information Science* 43(5), 649-64, <https://doi.org/10.1177/0165551516658821>.

---

<sup>8</sup> <https://www.cidoc-crm.org/>

<sup>9</sup> <https://www.cidoc-crm.org/frbroo/Issue/ID-360-lrmoo>

- [11] Hill, Heather, and Jen J. L. Pecoskie, 2017. Information activities as serious leisure within the fanfiction community. *Journal of Documentation* 73(5), 843-857, <https://doi.org/10.1108/JD-04-2016-0045>.
- [12] Kiryakos, Senan, and Magnus Pfeffer, 2021. Exploring the research utility of fan-created data in the Japanese visual media domain. *International Conference on Asian Digital Libraries*. Cham: Springer International Publishing, 210-218.
- [13] Pianzola, Federico, 2020. Linked-Potter: An example of ontology for the study of the evolution of literature and reading communities. In: *JADH2020 Proceedings of the 10th Conference of the Japanese Association of Digital Humanities "A New Decade in Digital Scholarship: Microcosms and Hubs"*, ed. B. Hodošček. Osaka: Graduate School of Language and Culture, Osaka University, 28-32, <https://jadh2020.lang.osaka-u.ac.jp/programme/longpaper/pianzola-linked.html> (Data and code: <https://github.com/fedormyskin/Linked-Potter>).
- [14] Pianzola, Federico, Alberto Acerbi, and Simone Rebora, 2020. Cultural Accumulation and Improvement in Online Fan Fiction. *CEUR Workshop Proceedings* 2723, 2-11.
- [15] Suan Stevie, 2021. *Anime's identity: performativity and form beyond Japan*. U of Minnesota Press.