

The impossible quest for the complete list of all anime

Magnus Pfeffer, Hideyuki Ōtsubo, Zoltan Kacsuk and Martin Roth
Mechademia 2023: Aftermath
Mechademia x Kyoto International Manga Museum
Kyoto, 27–29 May 2023

[1]

Hello, and welcome to our presentation on the impossible quest for the complete list of all anime.

[2]

My name is Zoltan Kacsuk, and to make this presentation more seamless, I will be explaining our work on behalf of all four co-authors: namely, our colleague Hideyuki Ōtsubo, Executive Director of the Japan Animation Creators Association, and also representing the Anime Tokusatsu Archive Centre, as well as the anime data collection leg of the Media Arts Database; furthermore, from the Japanese Visual Media Graph project – JVMG in the following – we have Magnus Pfeffer (professor at Stuttgart Media University), Martin Roth (associate professor at Ritsumeikan University) and myself as project researcher at Stuttgart Media University.

[3]

Our collaboration on today’s topic grew out of the work between the Media Arts Database and the JVMG. We both share an interest in constructing a list of all anime, that is as complete as possible, but approach this undertaking in somewhat different ways. We further contrast our own work with that of Masahiro Haraguchi and the List Production Committee, which is in part the inspiration for this presentation’s title due to its almost heroic nature, and which is not only the forerunner of all such attempts at creating a complete list of all anime, but also its most likely candidate for a gold standard, should one exist. Through introducing these three approaches and comparing some of their most important dimensions, we then proceed to draw out and highlight the shared theoretical questions that underpin and partly undermine all such attempts at constructing a complete list of all anime. Finally, we offer three answers to the question of “why even bother trying to construct such a list”, and connect our discussion to larger trends in arts and humanities research. One important caveat before we proceed, as will hopefully become clear, all three of these approaches actually collect data that are supersets of all Japanese animation and thus of what would be commonly understood under the term anime in the context of anime studies.

[4]

So, starting with the introduction of the three approaches, we will discuss them in chronological order of appearance.

[5]

Thus, the first of our three, is what is known as Haraguchi Complete Data, which is compiled by Masahiro Haraguchi and now supported by the work of the volunteers of the List Production Committee. Haraguchi wanted to create a complete list – which later grew into a database – of all animation credits information, with a scope of animation broadcast on TV, screened in theaters, and released on direct to video formats in Japan. What sets this undertaking truly apart is its strict emphasis on collecting ground truth for all the information included in the database: most often in the form of recordings of opening and closing credits information, but in some cases even archival production materials from the studios themselves. The scale of the ground truth that has been collected by Haraguchi and the List Production Committee is truly breath-taking, and has been the subject of preliminary research by Hideyuki Ōtsubo and his colleagues; but we hope that this

presentation, among other things, helps raise awareness of this work and possibly spark further research into this undertaking and the people behind it.

[6]

Although the database itself is not openly available, Haraguchi has published the data he has collected annually since 1984, and has even compiled retrospective summary publications for TV and theatrical anime releases respectively.

[7]

Next up, we have the Media Arts Database, a project of the Agency for Cultural Affairs of the Japanese Ministry of Education, Culture, Sports, Science and Technology that was initiated in 2010. It collects information on four domains: manga, animation, video games and media art in Japan. The data sources vary among the four domains, but they all prioritize institutional sources (such as libraries and official associations) and the publishers themselves. The data is made publicly available in RDF format with a very open Creative Commons license.

[8]

The data on anime in the Media Arts Database is compiled based on a range of data sources, including sources based on Haraguchi Perfect Data (highlighted in red), but also various other information aggregation services and companies. The present slide only provides a rough list of examples for the sources used for the various categories, you can find a very detailed breakdown on the MADB website regarding the origin of the different parts of the data.

[9]

To illustrate just how complex the composition of the data from the various sources in the MADB is, you can see on this slide how different parts of the data are compiled from the range of sources. This diagram also highlights how certain parts of the data are still very much a work in progress, you can see that there are incomplete, unpublished and unexamined parts of the map laid out here.

[10]

Finally, we have the Japanese Visual Media Graph, which is a research initiative funded by the German Research Foundation, building and maintaining a knowledge graph about popular Japanese visual media, such as anime, manga and video games for researchers working with these domains. The project integrates data from collaborating online enthusiast communities as well as other sources that have compatible open licenses such as Wikidata and the Media Arts Database. All data is made available under a Creative Commons license in RDF format.

[11]

The list of anime titles in the JVMG is a product of merging the data from the various sources. This is done in an automated fashion with human supervision and manual matching where necessary. The aim is to create a homogeneous list with no duplicates and no stubs (that is entries with no information). The more data sources are integrated into the JVMG the larger the coverage should become, however, this also means that there will be over-coverage based on the definition of anime a given researcher might want to employ. The solution for handling this is the possibility of filtering the data according to various traits.

[12]

Comparing the three approaches we can see that they have distinct institutional backgrounds (private, state, and academic), corresponding to different goals (namely archiving, promotion of Japanese culture, and research) and dependency structures for their ongoing operation. Although the

work of Masahiro Haraguchi and the List Production Committee has the longest history, it also seems to be the most precarious in its reliance on the volunteer work of a select few individuals with no institutional background or financial support.

[13]

Regarding data sources, Haraguchi Perfect Data is quite unique in the way that it works almost exclusively from ground truth sources, whereas Media Arts Database and the JVMG rely on data sources that are deemed trustworthy as opposed to engaging with the source material itself. The latter two projects are similar in that they both employ RDF and make their data available with open licenses. Haraguchi Perfect Data, although not open in the strict sense, is regularly published and even more importantly is incorporated into the Media Arts Database (and thus the JVMG as well), and is therefore released openly albeit not directly.

[14]

The scope of the data collection of the three projects is where the differences really come to the fore. Haraguchi Perfect Data has a well defined aim of collecting all credits information on animation broadcast on TV, screened in theaters and released on various straight to video formats in Japan. Since ground truth is collected for all works directly in these three domains, the likelihood of something missing from the data is very low. At the same time, due to its scope, this data does include some titles that would not be considered anime by most definitions, such as Disney animated feature films. The Media Arts Database has a slightly different intended scope, namely, collecting all Japanese animation works, including those that are published on other channels such as original net animation (although this part of the collection is currently only up to date until 2018), as well as potentially all animation works that have been released in Japan. As a result it is much harder to guarantee that all titles are included in the data that would fall under its scope. Furthermore, similar to the Haraguchi Perfect Data there is a definite presence of titles in the database that would not be considered anime by most definitions, such as the already mentioned original Disney animation works, that were also released in Japan. Finally, the Japanese Visual Media Graph, has the widest scope, as it collects all data that is in at least one of its source databases. In this case, however, as explained earlier over-coverage is not a bug, but rather a feature, as ideally the data can be filtered to better approximate the definition of anime a given researcher would want to consider. Since the JVMG aims to service all domain specific research needs it cannot allow itself to take a stance on what should be excluded from the coverage.

[15]

Which leads us into the inevitable theoretical considerations that both underpin and undermine these attempts at creating complete lists of anime.

[16]

There are two important questions that need to be addressed when constructing such databases. The first one revolves around the problem of what should be considered anime, and the second is the problem of how to best define the unit of recording, that is, should we record the franchises, the works, or even the individual episodes of a series.

[17]

Due to the limits of the current presentation we will only discuss the first of these problems.

[18]

Anime can be seen as pointing towards a particular style. But, how would we go about delineating its boundaries, which stylistic elements would be must have markers for something to be considered anime? And whose definition should be taken into account?

Anime can also be understood as Japanese animation. In this case should claymation works also be considered anime, if not, then we are once again back at the style side of the problem.

But even if we accept the purely geographic version of the interpretation of what is anime, we are still tripped up by a host of questions: what exactly does Japanese animation mean? Does it depend on the publishing company being Japanese, or it having at least its headquarters in Japan? Or whether the work was created primarily for the Japanese market? Or the amount of Japanese creative labor or otherwise involved in its production? And so on.

[19]

These questions are, of course, not new, indeed, they have been discussed at length by anime scholars like Nobuyuki Tsugata and Stevie Suan. The reason we need to foreground these questions here once again, is because the endeavor of trying to construct complete lists of anime not only means that decisions have to be made regarding such delineations, but also that the lists themselves will act as position takings in these debates around what constitutes anime.

Maybe if we go back far enough in time we can find a unity of the above listed dimensions: style, production, country of origin, target market, and so on. But at some point this unity started to break down, and for a long time now anime functions more as a discursive construct than anything else, that is mobilized in various contexts, by different actors to articulate a range of positions and to pursue all kinds of aims. We must therefore be conscious of the fact that our projects too will inevitably be positioned within this wider context of the many possible meanings of anime.

[20]

Underpinning and further complicating these fundamental questions in relation to what is anime and how to define the elementary unit of our data collection we find a wealth of technical considerations and challenges that emerge from the actual practices of creating, distributing and consuming anime. Again due to time constraints we will only highlight one such issue to provide a sense of the complexity of the topic.

[21]

As with most creative industries, anime production and distribution is not necessarily a well delineated process with a clear endpoint, but can rather involve retakes, remasters, edits, special editions, and so on, leading to a proliferation of versions,

[22]

which can be hard to capture in all their detail as well as to compare and verify, leaving us researchers with a difficult task when trying to pin down our objects of study.

[23]

But why do we even need a complete list of anime, especially, if as we have just ascertained, it is almost impossible to find a common ground for defining anime satisfyingly for everyone?

[24]

The three projects highlighted in this talk each have somewhat different answers to this question. For Haraguchi and his team of volunteers, recording is an end in itself, the value of which does not come into question. For the Media Arts Database the underlying goals of cultural policy and promotion provide the foundations of this work. And as already stated the Japanese Visual Media

Graph is interested in opening up new avenues of research for anime and other areas of Japanese popular visual culture. In closing we would like to narrow in on this last aim and what it means in the context of the transforming arts and humanities.

[25]

Anime research, not unlike a large part of arts and humanities undertakings, often focuses on major works – based on critical acclaim, popularity or both. The anime industry and its output, however, is far bigger; in a sense we are engaging with just the tip of the iceberg. In fact, we could ask the question, when we talk about anime, are we talking about just the canon of major works or anime in general? And even more importantly, how can we understand the larger underlying body of all anime works and its relation to the major works discussed in detail? Without a list of all anime it is very hard to know what the actual population of all anime works that should be considered is. If we have access to such a list, we can utilize a host of different approaches to gain better insights about anime in general, not just analyzing the whole population with various automated tools, but also taking samples that are more representative, which can then be analyzed in more detail by humans, machines or both. Having access to an enumeration of the whole population also enables more rigorous longitudinal approaches to answer questions about the changes and developments in anime over its history.

[26]

This new approach of looking at the previously un-analyzed mass of works is, of course, not specific to research on anime, but has been making an impact across the arts and humanities. For example, two well known representatives of this approach are Franco Moretti's work in literary studies, *Distant Reading*, and Lev Manovich's *Cultural Analytics* in media studies, both calling for an engagement with the wealth of literary and visual materials now available on a previously unimaginable scale, and both advocating for the incorporation of new digital approaches to processing amounts of material that would be unfeasible to examine not only for individual researchers, but even whole research teams. In this same vein of research, our colleague, Martin Roth has already pioneered this approach in relation to the field of Japanese video games.

[27]

One obvious question that comes to the fore, however, is “what if the reason nobody has examined these neglected titles in detail, is because they are simply irrelevant?” To this question we offer two answers here. First, knowing something exists and ignoring it knowingly, is definitely different from not knowing that it exists at all. And in this respect having a more or less complete overview of the whole population of anime is essential to delineate parts that would seem to be redundant to analyze. Second, and more importantly, sometimes it is the very parts that seem to be plain, coarse and undeserving of analysis that are actually the keys towards new understanding in a particular domain. For example, in the related field of manga studies, we have the famous example of Gō Itō's work. In his book *Tezuka Is Dead*, he takes as his starting point the lack of attention afforded to Gangan manga by manga critics. This blind spot in manga criticism, and the seemingly incorrect page compositions, that could be dismissed as just simply bad craftsmanship, become the key to unlocking a new understanding of manga. Another reason Itō's work is such a good example here, is because he does not rely on large-scale quantitative analysis to reach his findings and present his theory. As many authors have stressed in various fields, the possibilities opened up by digital tools and archives are not meant to displace traditional ways of working in the arts and humanities, rather they are meant to complement them. In a similar way, we are not advocating for an either-or question of analyzing major works versus the mass of unexamined anime, but rather want to draw attention to the new horizons being opened up for research that were previously inaccessible.

[28]

Thus, this is why working on what a complete list of all anime could or should be is so relevant today, even if we can never fully agree on its contents and will never be able to create it in an absolutely watertight fashion. As you all know, we are standing on the brink of yet another huge revolution, one that will fundamentally change how we work with digital materials. Indeed, previously one of the big questions many research support groups were struggling with was how to enable the arts and humanities researchers to be able to deal with the large wealth of digital data now at our disposal. AI assistants might make this task for humanities researchers much easier and enable getting over the hurdle of having to learn to code in order to be able to fully engage with large databases and knowledge graphs. Once the questions of the researchers can be formulated in everyday language and the AI will take care of translating it into program code to execute, a new era of research will begin. And like Major Kusanagi merged with the Puppet Master we too working in tandem with our AI assistants will survey the vast landscape of anime before us, and ask: “where do we go now, what questions shall we ask next?”

[29]

Thank you for your attention, and please get in touch with us, if you are interested in our work!