Zoltan Kacsuk

Using Fan-Compiled Metadata for Anime, Manga and Video Game Research

Revisiting Azuma's "Otaku: Japan's Database Animals" Twenty Years On[1]

Wikipedia is perhaps the most well-known testament to the efficacy of volunteer contributions adding up to creating a truly robust and powerful resource. Beyond such general projects like Wikipedia, there are a vast number of specialized communities that have compiled all kinds of wikis and databases on their foci of interest. The metadata databases[2] compiled by fan or enthusiast communities[3] represent a so far mostly untapped source of high-quality granular data for researchers to work on. The present chapter aims to both showcase the possibilities opened up by utilizing such databases for academic research, while at the same time addressing one of the key questions in relation to the validity of research conducted on such resources, namely that of data incompleteness. Data on characters from mainly Japanese anime and visual novel games[4] are employed in a case study examining one of the claims from Hiroki Azuma's *Otaku: Japan's Database Animals* (2009 [2001]) to unpack these points.

1 The present chapter incorporates materials from blogposts the author has written for the Japanese Visual Media Graph Project Website (https://jvmg.iuk.hdm-stuttgart.de/), and also builds on the presentation and feedback from the JADH2020 conference "A New Decade in Digital Scholarship: Microcosms and Hubs" (20–22 November 2020).

2 Although these databases often also contain summaries, pictures, reviews and so on, they are primarily made up of descriptive metadata relating to, for example, studios, franchises, individual works, and characters. For the purpose of the present discussion, these databases will be referred to as metadata databases to emphasize this aspect.

3 The expressions "fan community" and "enthusiast community" are used as synonyms in the present chapter. For a detailed elaboration on the ways finer distinctions can be drawn between these concepts, see for example Nicholas Abercrombie and Brian Longhurst (1998) or Kacsuk (2016).

4 Visual novel is a predominantly PC game genre originating in Japan and oriented towards mostly telling stories of love and sexuality – although works with other foci, like horror, have also achieved prominence – featuring still images and written text and involving little or no choice on the part of the players.

Published in: Martin Roth, Hiroshi Yoshida, Martin Picard (eds.): Japan's contemporary media culture between local and global. Content, Practice and Theory. Heidelberg: CrossAsia-eBooks, 2021. DOI: https://doi.org/10.11588/crossasia.971.c12881

Metadata from enthusiast communities among the data sources in video game studies

Video game studies, in the narrow sense, is the field defined by the humanities- and social-science-informed questions related to video games, their creators, publishers, and players, and their impact on culture and society. However, there is also a wider field of video games research, a large part of which is actively concerned with the understanding of player behaviour and game design from the perspective of monetizing video games and game play experiences. There is no clear boundary between the more academic research of video game studies defined in the narrow sense, and the research on games and players conducted from a business and industry perspective. Indeed, there are large areas of overlapping interest between the two,[5] albeit certain frameworks and methods can be more readily associated with the one or the other approach.

One distinguishing feature of the business side of video games research is the access it enjoys to in-game generated server-side data, which is a key input for many game analytics (El-Nasr, Drachen, and Canossa 2013), but which is not often made available to researchers external to the companies concerned (for an example of this, see Bonenfant, Richert, and Deslauriers 2017). There is, however, one special subfield of academic inquiry that also builds heavily on processing in-game data, namely the research on games and education – referred to as edugames, serious games, or game science – where the games and game play themselves are often experiments, with researchers thus having full access to all in-game data (see De Freitas 2018).

There are, of course, various ways to capture in-game data, for example, researchers can take advantage of "open client-side user interface[s]" that "allow customizations and modifications by the user community" (Lukacs 2014, 408). Furthermore, quite a few games offer a range of summary statistics for players to be able to track their own progress and to compare it to others', which can then be used for research purposes as well. Some publishers and games go even further, and have extensive application programming interfaces (APIs) providing a wealth of mineable data that services and researchers can actively build on. Just playing a game can also be a way of accessing a host of in-game data, not to mention the experience of the gameplay itself, and for game-centric research this is often a key ingredient (Aarseth 2004; Myers 2014).

Traditional data collection methods and sources are typically more prominent in academic research (Lankoski and Björk 2015; Lukacs 2014; Myers 2014),

5 In Japan, for example, there is a particularly strong connection between the games industry and academic research (Picard and Pelletier-Gagnon 2015).

although surveys, interviews, and focus groups can play a significant part in business-side research as well (El-Nasr, Drachen, and Canossa 2013). Industry data can also be an important source of quantitative information, not only on distribution and sales figures, but also player demographics (Lukacs 2014). Experimental setups are a common feature – beyond the already mentioned domain of games and education – of the wide-ranging research on the impact of video games on social behaviour (cf. Greitemeyer and Mügge 2014). Capturing user experience research-type biometric data, such as eye movement or pulse information, also occurs in a range of video game research contexts (see for example Bamparopoulos et al. 2016; De Freitas 2018; El-Nasr, Drachen, and Canossa 2013).

Another important group of materials employed in video game studies research are the various documents generated around the games by a host of different actors spanning professional journalists and media content creators to player communities. Documents, here, refer to all forms of data recorded on forums, blogs, videos, podcasts, discussion and comment threads, reviews, articles, how-to guides, wikis, game information databases, and so on (cf. Sköld et al. 2015). Among these, the archival and database-building work of enthusiast communities or dedicated individuals merits special attention for the way it can provide access to both source materials and well-structured, granular metadata (Picard and Pelletier-Gagnon 2015).[6]

This type of archival and database-building work is, of course, not unique to video games. Media fans were compiling data on their favourite works long before the internet became the most important channel for communication (Jenkins 1992; Okada 1996; Yoshimoto 2009). Naturally, with the advent of the internet, this process of collecting and cataloguing information by individual fans and enthusiast communities has only become easier and amplified in its scale. The level of detail afforded by these fan-compiled data sources has also not gone unnoticed in academic research (Hills 2002; Picard and Pelletier-Gagnon 2015). Although various online databases created by enthusiast communities have become the go-to resource for checking information on hard-to-find media texts and artefacts, their use for large-scale quantitative research has

6 Some databases also act as repositories for hard-to-find works, while others only record the descriptive metadata used to identify, find, and describe them. For simple search and retrieval functionality very rudimentary cataloguing systems could suffice. Nevertheless, a large number of these projects demonstrate elaborate and well-structured ontologies (in the information science sense of the word) mirroring professional cataloguing and classification systems.

yet to become widespread (for an excellent example, however, see Utsch et al. 2017).

A number of researchers have already started the work of creating robust metadata ontologies and infrastructures for facilitating research on various aspects of video games (e.g. Bamparopoulos et al. 2016; Fukuda, Mihara, and Oishi 2020; Lee, Clarke, and Perti 2015). In line with this research direction, the diggr (Databased Infrastructure for Global Games Culture Research) research project has been a pioneer in incorporating data compiled by enthusiast communities (Hoffmann, Freybe, and Mühleder 2017) in its work on building "a data driven research infrastructure" for researchers working on Japanese video games (Freybe, Rämisch, and Hoffmann 2019, 14). Continuing in the same direction, harnessing the power of fan-created databases for academic research is precisely the aim of the Japanese Visual Media Graph (JVMG) project.[7] By learning from and building on the experiences and content produced and aggregated by various enthusiast communities, the project aims to create an integrated database and query tools primarily for academic researchers working on Japanese visual media. The research project showcased in the present chapter is one of the first experimental use-cases aimed at demonstrating the way this type of data can be utilized for academic research.

Has there been a shift in the way characters in Japanese manga, anime, and video games are produced?

Hiroki Azuma's *Dōbutsu ka suru posutomodan: Otaku kara mita Nihon shakai* (Animalizing postmodern: Japanese society as seen from otaku), published in 2001, has been one of the most influential treatises not only on Japanese otaku, but also on the production and consumption paradigm defining Japanese anime, manga, light novels, and video games in late modernity. The book's impact on the discourse around otaku and the domains just enumerated is truly international, thanks, in part, to the English translation, which was published in 2009 as *Otaku: Japan's Database Animals*.[8] With almost twenty years since the original publication in Japanese, and more than ten years since the English translation was released, the concepts and frameworks outlined by Azuma have become cornerstones of this scholarly discourse (see for example

7 Both the diggr project and the JVMG project have been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). The author of the present chapter is a member of the JVMG project.

8 All quotes in the following are from this English edition.

Itō 2005; Lamarre 2009; Suan 2017). However, even though the book's line of argument contains a number of potentially testable statements, to date these have not been subjected to any large-scale empirical test. This is probably due, in part, to the inclination towards case-study-based research approaches in the concerned fields; however, the lack of available data that is both large and granular enough to use for such an analysis has also not been conducive to the emergence of quantitative examinations of this sort.

In his book, Azuma positions otaku culture as originating in Japan's postwar reception (or, more precisely, "domestication") of subcultural works from the United States; as such, it represents a move towards the postmodernization of culture. In this sense, the sensibilities and production/consumption patterns found in otaku culture foreshadow changes beyond the realm of the subcultural. He goes on to point out that two of the hallmarks of postmodernization found in otaku culture are its embracing of derivative works on the one hand, and the substitution with fiction (for an earlier generation) or the complete letting go of *grand narratives*[9] (for a newer generation) on the other hand.

This generational shift is exemplified by the emergence of *chara-moe*, an emotional reaction felt by the consumers to characters and/or their specific elements, such as cat ears, glasses, and so on. As a result, "a database for *moe*-elements that generates the characters has been established" (Azuma 2009 [2001], 47). For Azuma, the metaphor of the database[10] captures the space of imagination shared by creators and consumers – the lines between the two already blurred by the prominence given to derivative works – that gives rise to new works, settings, and, most importantly, characters, but which is devoid of any structuring grand narrative.

> As a result, many of the otaku characters created in recent years are connected to many characters across individual works, rather than emerging from a single author or a work. (Ibid., 49)

9 The term "grand narrative" was most famously used by Jean-François Lyotard in his work *The Postmodern Condition: A Report on Knowledge* 1984 [1979], in which he pointed out that the major driving conceptual frameworks – the grand narratives – of modernity (e.g. Enlightenment, Marxism, etc.) have started to lose their appeal.

10 For a discussion of the significance of the database in Azuma's work, see Schäfer and Roth (2012). Furthermore, it is worth noting here that among the various works discussed in Azuma's book, he also brings up the example of the Japanese website TINAMI (the name originally stands for "The Information NAvigator of Manga Artists on the Internet"), which at the time served as a sort of search engine for illustrators, manga artists, and similar creators active on the internet (from the Japanese Wikipedia article: https://ja.wikipedia.org/wiki/TINAMI).

Although Azuma's arguments in the book go far beyond a simple description of the changes characterizing consumption/production patterns in otaku culture with the emergence of database consumption, this point, in relation to its impact on the way characters become increasingly derivative of each other, is one of the most well defined and potentially testable in relation to the central notion of the database. The following quote illustrates what characters being increasingly connected to other characters across works actually means in Azuma's view.

> I believe that it is more appropriate to use the image of the database to grasp this current situation. The emergence of Ayanami Rei did not influence many authors so much as change the rules of the *moe*-elements sustaining otaku culture. As a result, even those authors who were not deliberately thinking of *Evangelion* unconsciously began to produce characters closely resembling Rei, using newly registered *moe*-elements (quiet personality, blue hair, white skin, mysterious power). Such a model is close to the reality of the late 1990s. Beyond Rei, characters emerging in otaku works were not unique to individual works but were immediately broken into *moe*-elements and recorded by consumers, and then the elements reemerged later as material for creating new characters. Therefore, each time a popular character appeared, the *moe*-element database changed accordingly, and as a result, in the next season there were heated battles among the new generation of characters featuring new *moe*-elements. (Ibid., 51–52)

Let us then formulate a testable hypothesis based on this idea of how characters become increasingly connected as they draw on each other, or rather the database elements into which they are broken down into. If Azuma is correct in that such a shift did take place in the way characters become more dependent on characters that came before them, as opposed to the potentially more diverse creative input of creators that supposedly characterized previous eras, then we should find that the portion of new characters with shared traits should increase over time especially in the case of works aimed at the otaku market.[11]

11 The original analysis had two hypotheses, see the blogposts for details: https://jvmg.iuk.hdm-stuttgart.de/2020/10/28/tiny-use-case-2-can-we-test-one-of-the-points-from-hiroki-azumas-otaku-japans-database-animals-with-the-jvmg-database-part-5-summary-and-lessons-learned/.

Data sources and operationalization of the hypothesis

To try and test this hypothesis, data from The Visual Novel Database[12] (VNDB) and Anime Characters Database[13] (ACDB) was employed,[14] as these databases both have a significant number of characters listed (over 79,000 and 100,000, respectively) and a relatively large number of traits describing them. There are, however, important differences between the two datasets. VNDB only focuses on visual novels, whereas ACDB collects data on a wide range of characters from various media (although predominantly focusing on visual novels and anime). Furthermore, VNDB has a very rich and rigorously structured ontology of traits, which, however, lacks a core set of featured traits that would be expected to be available in relation to all characters. In contrast, ACDB features a hybrid system for describing characters, which, on the one hand, supports a closed ontology for eight flagship traits (e.g. hair and eye colour) that are part of each character's factsheet, and, on the other hand, provides the opportunity for free-form tagging of characters with user-created labels.[15] For the purpose of this study, traits relating to sexual activity were excluded from the VNDB dataset, as those would have unnecessarily inflated shared trait counts. Gender was also excluded for technical reasons from both data sources.

Considering the structure of the data, the concepts from the hypothesis were operationalized in the following ways. First, to define characters with shared traits, it is important to consider what Azuma and, by extension, the hypothesis is referring to. Instead of taking any two characters that share at least one trait, it is better to set some kind of minimum threshold for characters to be considered characters with shared traits, in order to capture the change in character creation/consumption practices described by Azuma. We can think

12 https://vndb.org/.

13 https://www.animecharactersdatabase.com/.

14 Both of these database projects were started in 2007 by their respective lead developers, and both have grown a community of contributors around themselves. VNDB is strictly non-commercial, while ACDB features advertising on its website to cover operational costs. VNDB is exclusively focused on providing information about visual novels. ACDB is centred around characters, mostly from Japanese visual media, and also features small games and interactive features for its users. The data in both databases is added to and edited by the users.

15 The differences between the two databases also become apparent when trying to assign a date of first appearance to the characters featured in them. For further details, see https://jvmg.iuk.hdm-stuttgart.de/2020/09/16/tiny-use-case-2-can-we-test-one-of-the-points-from-hiroki-azumas-otaku-japans-database-animals-with-the-jvmg-database-part-2-descriptive-statistics/.

of the increase in the number of characters with shared traits as a result of some popular trait combination appearing – as in the Ayanami Rei example above – and that group of traits being replicated in other subsequent works. The possibility exists, of course, that such correspondences happen by chance, but those should be uniformly present throughout the data and should not have an impact on longitudinal trends. The cut-off point of a minimum number of five shared traits was selected for characters to be considered characters with shared traits. Second, to operationalize change over time, and to try and capture the phenomenon of certain templates becoming suddenly popular, the data was segmented – for ease of analysis – according to calendar years. Thus, characters with shared traits were counted only among characters with shared first appearance dates.

Furthermore, in order to allow for a better comparison between the VNDB and ACDB data, the latter was separated into two datasets for visual novels and other works, resulting in a relatively even split (see Figure 1 below), which were treated as independent datasets for this analysis (even though creative influence can and those travel between the realms of visual novels and other types of works). Since Azuma's book is about Japanese works, characters in the ACDB dataset that belong to media types that are clearly non-Japanese (e.g. "Western animation") or have no media type information were also disregarded for this analysis.

A first look at the data: Descriptive statistics

Beyond the differences in the temporal range of the three datasets (1: VNDB characters, 2: ACDB visual novel characters, and 3: ACDB other characters) the distributions of the *number of characters* by year, in Figure 1, below, all follow a similar trend. The number of characters recorded for each year demonstrates a mostly growing tendency, which then plateaus before going into a steady decline. Since it is highly unlikely that such a drop should have occurred in the number of new characters appearing per year in the second half of the 2010s, these distributions clearly signal that the datasets are not complete.[16]

16 The snapshots of the data used in the present analysis are from early 2020, thus it would seem logical that information on 2020 and potentially also on 2019 are not yet as complete as for earlier years. There is no available explanation for the drop in numbers of recorded characters prior to 2019. It could be due to a temporal lag in the works making their way to the non-Japanese audiences responsible for these databases, or it could be a general decreasing interest in these types of databases, or some other reason.

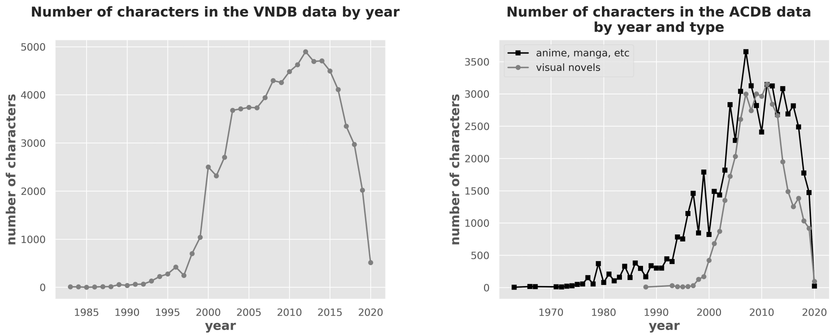

Figure 1. Number of characters in the datasets by year.

Examining the distribution of *average number of traits* per character by years, in Figure 2, below, for the two ACDB datasets, there is a somewhat similar trend to the preceding one in Figure 1. A mostly upward climb is followed by a clear drop in the average number of traits recorded per character after 2012, mirroring – although not exactly overlapping with – the drop in the number of characters. The decline in the VNDB data is less pronounced, and disregarding 2019–2020 could be seen as merely a plateauing around twenty traits per character on average.

The larger number of average number of traits per character in the VNDB data is a result of the difference between the ontologies and trait structures in the two databases. The difference in numbers between the two ACDB datasets, however, clearly signals that visual novels for the most part (all the way up till 2015) receive more attention when it comes to adding character traits to the database. This fact, along with the uneven completion of the datasets in relation to character trait information, both point towards the significance of the focus of attention of contributors in building these types of databases and their records.

Figure 3, below, showing the distribution of the *average number of characters traits are shared with*, once again follows the trend seen in the above figures. This is most likely a result of the way both the number of characters and the average number of traits per character potentially impact the number of characters with shared traits per year. The more characters there are in a given year the higher the number of potential characters with shared traits, and the higher the average number of traits per year the easier it becomes to reach the threshold of five shared traits between pairs of characters set in this analysis.

To see whether one or both of these two values do indeed, in some way, determine the changes in the *average number of characters traits are shared with*, or whether there is an element of temporal change as predicted by the

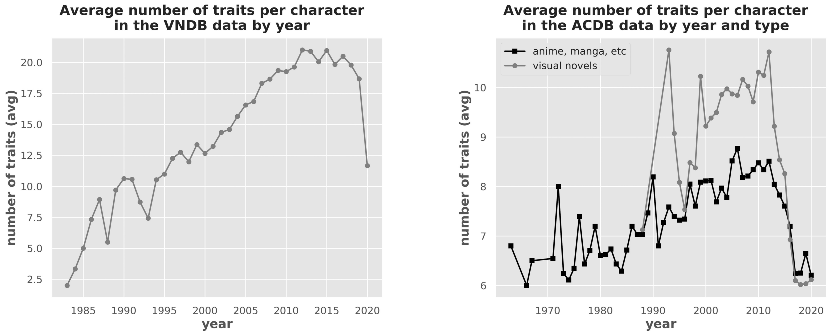

Figure 2. Average number of traits in the datasets by year.

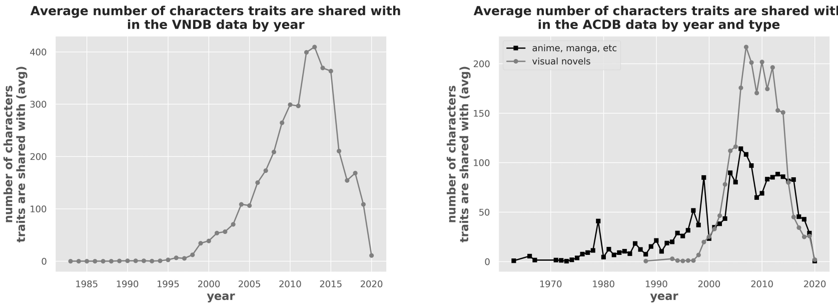

Figure 3. Average number of characters traits are shared with in the datasets by year.

hypothesis, regression analysis was conducted on the three datasets. Before discussing the results of the regression analysis, however, the central problem of working with incomplete data needs to be addressed.

Working with incomplete data

Incomplete or missing data is an almost pervasive characteristic of most empirical research in the humanities and social sciences. As such, it is a very common approach to just ignore its presence, as it goes without saying that the data is incomplete. A more reflexive stance is to explicitly draw attention to the problems that the incompleteness of the data presents. In some cases, there is no immediate way to handle the missing data; nevertheless, analysing the available data can still yield new insights, and is therefore deemed a better option than discarding the data altogether, while acknowledging the limitations of this approach (cf. Roth forthcoming). Wherever possible, addressing the missing data problem with various tools in one or more phases of the research process, from research design through data gathering and processing to data analysis, is, of course, the preferred approach from a methodological perspective.

Using metadata for the type of research presented in this chapter comes with its own set of special challenges in handling missing data. Crucially, some of the common ways of dealing with missing data, such as imputation and weighting, cannot be employed. Using only complete items is also problematic since there is no way of defining completeness. How many traits should a character have to be considered satisfactorily described? Of course, if only considering a closed set of given traits, such as that found in ACDB without the tag information, then completeness could be defined. Nevertheless, the approach of fixing the number of traits employed in the regression analysis below, can be seen as potentially analogous to working with only complete cases.

When working with statistical models, for example in regression analysis in this case, the distribution of the missing data is also a key factor in relation to the potential validity and bias of the results. Looking at the data in the present analysis, the *average number of traits* can be thought of as a proxy for attention being paid to the completion of the database, since it is probably safe to assume that the level of detail in describing characters is less a function of the characters themselves and has more to do with the interest that builders of the databases have in them. As we saw above, examining the data from this angle, it is very plausible that these datasets have been completed unevenly with regards to the publication years of the characters. And, as already noted above, visual novel characters on average receive more attention than other characters in the ACDB database for the majority of publication years. Based on these two points, it is already certain that the incompleteness of the data is not random in any of the three datasets. There are statistical tools for addressing missing data, and even non-random missing data (Little and Rubin 2020), especially for survey data, however, their complexity is mostly beyond the scope of specialist researchers.

One further possibility for dealing with non-randomly incomplete data that is not necessarily discussed as part of standard solutions, most likely because it is more reliant on innovative approaches to the question being studied[17] rather than generalizable rules and guidelines, is the incorporation

17 This approach is not specific to dealing with incomplete data, but is rather employed more generally to solve the problem of researching social phenomena, which are often hard to observe directly. One of the most famous examples of this type of thinking is found in Durkheim's *Suicide* (2002 [1897]).

of assumptions that can help reposition the results in a way that allows for conclusions even in the face of potential bias.[18] The present problem will serve as an example to illustrate this approach. In the section discussing the formulation of the hypothesis to be tested, the quote from Azuma's book ended with:

> [E]ach time a **popular** character appeared, the *moe*-element database changed accordingly, and as a result, in the next season there were heated battles among the new generation of characters featuring new *moe*-elements (Ibid., 52, my emphasis added in bold)

Thus, if we expect to see an increase in the portion of characters with shared traits over time in line with Azuma's proposition, we can also take on board this point about the mechanism behind the proliferation of certain character templates, namely popularity. Researching Wikipedia, Royal and Kapila found that it "reflects the viewpoints, *interests*, and emphases of the people who use it" (2008, 146, my emphasis). It is probably safe to assume that this finding generalizes over to other online resources, such as our datasets, compiled by communities on a voluntary basis. Thus, unless we have a strong reason to believe that characters conforming to popular templates are shunned for some reason among the users who compiled the data, then if popularity is a driving factor for the phenomenon described by Azuma, we can also expect a higher level of interest directed at these very characters (even though, most likely, not all of the newer versions of the same template will enjoy identical levels of success). This would imply that (1) these types of characters should be more prevalent among characters with higher recorded trait counts, and (2) that they should also be more likely to be in the database in the first place. We will return to these two assumptions in our interpretation of the regression analyses' results in the following section.

Finally, although this technique is again not specific to the problem of incomplete data, by examining multiple sources, in our case two different databases and three different datasets, we can increase the validity of our findings. Although there is no guarantee for the same bias not being present across all three datasets, its likelihood is definitely smaller than there being some form of bias due to missing data in just one of the datasets.

18 It is important that the bias inherent in our models to be discussed in the next section cannot be overcome purely by adding assumptions. Nevertheless, we are not interested in the exact figures, but rather the existence or lack of certain trends, and this approach can help increase our confidence in interpreting our results.

A deeper dive into the data: Regression analysis

Regression analysis revolves around trying to estimate the relationship between the dependent variable (for which the observed changes in its values are to be explained) and the independent variables (also called explanatory variables, since they are used to explain the changes in the dependent variable's values). In this case, the *average number of characters traits are shared with* is the dependent variable, and the *number of characters*, the *average number of traits* and the publication *year*[19] (as a proxy for temporal change) serve as possible explanatory independent variables. If there are multiple possible relationships between the independent variables and the dependent variable, for example due to the large number of possible explanatory variables that can be included in the model, the process of regression analysis involves comparing the different possible models and selecting the best performing one.

In order to capture potential non-linear relationships between the dependent variable and the independent variables, squared terms of the *number of characters* and *average number of traits* variables were also included in the regression model building process. Furthermore, to account for potential interactions between the independent variables, interaction terms were also created between them,[20] and introduced in the model selection process. For each dataset, only years with at least thirty characters were used and continuous ranges of years were selected in every case. The Python statsmodels package was employed to build and evaluate the regression models. Figures 4–5 and Tables 1–3, below, contain the results for the best performing models for each dataset examined.

Based on the regression analyses for both ACDB datasets the interaction term between the *number of characters* and the *average number of traits*, and for the VNDB dataset the interaction term between the squared versions of these variables proved to be the single best explanatory variable for the changes in the values of the *average number of characters traits are shared with* variable. This result points to there being no actual temporal effect at play in the observed changes in the number of characters with shared traits, they are merely a function of the number of new characters and the average number of traits recorded for those characters. Thus, based on these results the hypothesis being tested was found to be unsubstantiated.

19 Transformed for ease of interpretation to integers, starting from one for the first year in the range and then incrementally growing by one for each year.

20 Below interaction terms are written with a capital "X" between the names of the two interacting variables.

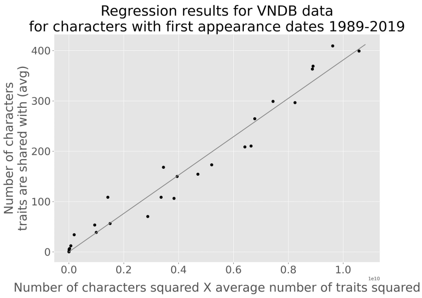

Figure 4. Best regression model for characters in the VNDB data.

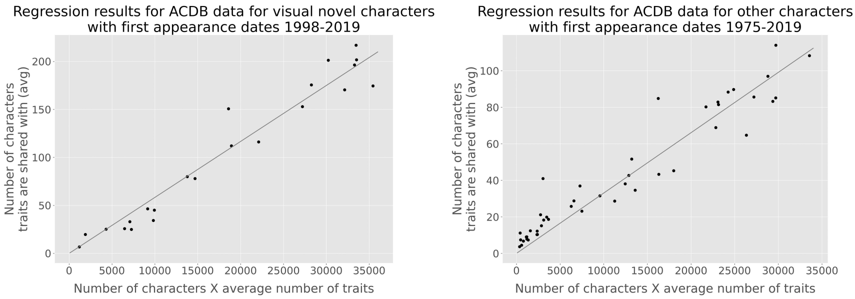

Figure 5. Best regression models for visual novel characters (left) and other characters (right) in the ACDB data.

Table 1. Regression results for best model for characters in the VNDB data for 1989–2019.

<table>
<tr><th colspan="6">OLS regression results for best model for characters in the VNDB data for 1989–2019</th></tr>
<tr><td colspan="6">Regression model</td></tr>
<tr><td colspan="2">Variables</td><td>Coefficient</td><td>Standard error</td><td>z-value</td><td>p-value</td></tr>
<tr><td colspan="2">number of characters squared X average number of traits squared</td><td>3.813e-08</td><td>1.11e-09</td><td>34.271</td><td>0.000</td></tr>
<tr><td colspan="6">Goodness of fit</td></tr>
<tr><td colspan="2">Adjusted R-squared:</td><td colspan="4">0.983</td></tr>
<tr><td colspan="6">Error terms</td></tr>
<tr><td>Skew:</td><td>0.049</td><td>Kurtosis:</td><td>2.754</td><td>Durbin-Watson:</td><td>1.196</td></tr>
<tr><td colspan="3">Heteroscedasticity robust covariance employed:</td><td colspan="3">HC3</td></tr>
</table>

Table 2. Regression results for best model for visual novel characters in ACDB data for 1998–2019.

OLS regression results for best model for visual novel characters in ACDB data for 1998–2019					
Regression model					
Variables		**Coefficient**	**Standard error**	**z-value**	**p-value**
number of characters X average number of traits		0.0058	0.000	27.940	0.000
Goodness of fit					
Adjusted R-squared:		0.982			
Error terms					
Skew:	0.742	**Kurtosis:**	3.699	**Durbin-Watson:**	1.793
Heteroscedasticity robust covariance employed:			HC3		

Table 3. Regression results for best model for other characters in ACDB data for 1975–2019.

OLS regression results for best model for other characters in ACDB data for 1975–2019					
Regression model					
Variables		**Coefficient**	**Standard error**	**z-value**	**p-value**
number of characters X average number of traits		0.0033	0.000	28.952	0.000
Goodness of fit					
Adjusted R-squared:		0.961			
Error terms					
Skew:	0.235	**Kurtosis:**	4.623	**Durbin-Watson:**	1.289
Heteroscedasticity robust covariance employed:			HC3		

As tempting as it would be, finishing the analysis at this point would unfortunately be premature for two main reasons. On the one hand, as already explained in the previous section (see the second assumption there), one of our expectations is that characters that are created based on popular templates are more likely to be included in these databases compiled by enthusiast communities. It is, however, crucial to our analysis whether this effect has some kind of temporal bias to it or not. If this increased likelihood of these types of characters appearing in the data is uniform across all publication years then it is potentially not a problem. In this case, if the temporal change we are looking

Table 4. Correlation between base explanatory variables for the VNDB data.

	Year	***Number of characters***	***Average number of traits***
Year	1.00	0.83	**0.95**
Number of characters	0.83	1.00	0.91
Average number of traits	**0.95**	0.91	1.00

for exists, we should find its presence regardless of the level of attention being paid to different years of publication by the community, since within each cohort of characters the over-representation of these characters should follow the same degree, and thus not impact their relative proportion vis-à-vis the proportions found for other years. If, however, this effect of popular template derived characters having a higher chance of being included in the data is not constant over time, then all sorts of problems can arise. One potential concern might be that perhaps the earlier years suffer from an over-representation of these types of characters in the data. This would make detecting a temporal trend of increasing proportions highly problematic.

On the other hand, there is another reason why it would be perhaps too early to accept the findings as is at this point. Namely, there is a high level of correlation between the independent variables *average number of traits* and *year* in the VNDB dataset (see Table 4), thus it is hard to ascertain how much of the relationship between the *average number of characters traits are shared with* and the composite independent variable (the interaction term between the squared value of the *number of characters* and the squared value of the *average number of traits*) is a result of the change in the *average number of traits* or due to a potential temporal effect.

Fortunately, we can employ the same decomposition method to address both of these problems. To solve the problem of the high correlation between the temporal variable and that of *average number of traits,* we can fix the latter variable at given levels and thereby eliminate its effect from the subsets of data to be analysed, since it becomes a constant value in each case. To approach the problem of the impact of the higher likelihood of popular template following characters being included in the data potentially not being constant over time we can leverage our first assumption, that these types of characters should be more prevalent among characters with higher recorded trait counts. Assuming that the level of attention indicated by the number of traits is more or less constant over publication years, we can again compare the subsets of the data formed by fixing the value of the *average number of traits* variable at different levels.

Table 5, below, summarizes the regression analysis results for each subset of data pertaining to various levels of number of traits for the three datasets.

For each subset to be analysed only characters with the corresponding number of traits were included from their respective datasets. Table 5 only contains information on the single best performing explanatory variables[21] for each data subset to keep the discussion concise as much as possible. Due to the reduced number of characters in each data subset the temporal ranges of these subsets differ from the ranges indicated for the original regression results on the full datasets. For the VNDB subsets, the temporal range varies between 1998–2019 and 2000–2019. In the case of ACDB visual novel characters, the range narrows from 2000–2016 to 2003–2014 with the increasing of the number of traits. The range for ACDB other characters changes in a similar fashion from 1992–2018 towards 2005–2014. The reason for the lower number of subsets analysed for the ACDB data is due to the lower average number of traits contained in those datasets (see Figure 2, above). Data subsets, where a temporal effect (which is positive in each case) is part of the best performing variable, are indicated by the grey background of the cells.

Interpreting results for the ACDB datasets is more straightforward. There is only one subset (corresponding to nine traits) for both datasets where the best performing variable was the interaction term between the *number of characters* and the *year* variables. Other than that, it is basically the *number of characters* variable – which in all cases except for one (see table footnotes) was also the best performing model – for every subset of data. Considering the fact that the regression results for the analyses conducted on the whole datasets yielded the best explanatory variables *number of characters* X *average number of traits* in both of these cases, it seems quite persuasive that the decomposition of the datasets into their subsets according to numbers of traits simply led to the *average number of traits* variable's effect being controlled for, and thus left the *number of characters* as the sole explanatory variable in the relationship. In other words, there was no temporal effect being covered up by the interaction term in these datasets.

The results of the VNDB data are slightly more ambiguous. For number of traits equal to eight through thirteen, an interaction term including the temporal variable *year* proved to be the best performing variable and model for these data subsets. This would seem to confirm the suspicion that the effect of the *average number of traits* variable in the interaction term of the best performing model for the whole VNDB dataset had indeed covered up the presence of a temporal effect at play in the data.

21 In almost every case the single best explanatory variable is also the only variable in the best performing model, due to the high level of multicollinearity between the different explanatory variables.

Table 5. Single best explanatory variables for subsets of the data according to number of traits.

Number of traits	VNDB characters	ACDB visual novel characters	ACDB other characters
5	No model*	-	-
6	number of characters squared**	number of characters	number of characters
7	number of characters squared**	number of characters	number of characters
8	number of characters squared X year	number of characters†	number of characters
9	number of characters squared X year	number of characters X year††	number of characters X year
10	number of characters squared X year	number of characters	number of characters
11	number of characters squared X year	number of characters OR no model†††	number of characters‡
12	number of characters squared X year	number of characters	number of characters
13	number of characters squared X year	number of characters OR number of characters squared	number of characters
14	number of characters squared	number of characters	number of characters
15	number of characters	-	-
16	number of characters	-	-
17	number of characters OR number of characters squared	-	-
18	number of characters OR number of characters squared	-	-

* No model provided satisfactory values to be accepted, due to the large number of years with zero characters with shared traits.
** Only slightly better model results than for the variable *number of characters squared X year*.
† *Number of characters* is the best explanatory variable, but the normality of error terms is violated to such a degree that *number of characters X year* is perhaps a better explanatory variable in this case. There is, however, one obvious outlier year (2014), which if removed *number of characters* becomes the clearly best performing single explanatory variable.
†† There is one obvious outlier year (2014), which if removed *number of characters* becomes the clearly best performing single explanatory variable.
††† All possible models indicate the presence of strong autocorrelation among the error terms.
‡ The actual best performing model features two explanatory variables: *number of characters* and *year*.

Table 5. Single best explanatory variables for subsets of the data according to number of traits *(cont.)*.

Number of traits	VNDB characters	ACDB visual novel characters	ACDB other characters
19	number of characters	-	-
20	number of characters	-	-
21	number of characters	-	-
22	number of characters	-	-
23	number of characters	-	-

On the other hand, moving upwards from fourteen in the number of traits *number of characters* becomes the dominant explanatory variable. Since the temporal range of the data subsets shrinks from 1998–2019 to 2000–2019 as the number of traits increases, one possible explanation might be that for the shorter range of years the temporal effect can no longer be identified. To check for this, the regression analyses were also performed for the lower number of traits data subsets with the years 1998–1999 removed, which yielded the same results as those indicated in the table above. Thus, it is less likely that the lack of a temporal effect in the range of data subsets with higher number of traits is due only to the lack of information for the years 1998–1999 in those subsets.[22]

Recalling the first assumption that characters based on popular templates have a higher propensity to receive a more detailed treatment in the data we can offer the following interpretation of the results. The higher up we are in the number of traits the more likely we are to encounter these type of characters, and so the more plausible it is that we are indeed checking for the actual phenomenon described by Azuma. In this way, the presence of the temporal effect in the lower range of numbers of character traits versus its disappearance in the higher range of trait numbers points to the hypothesis being tested not holding up for the VNDB dataset either. The presence of the temporal effect in the lower range of number of traits data subsets is, however, consistent enough to warrant further investigation in the future.[23]

22 Another possible question is whether this pattern is just a result of the cut-off point being set at five shared traits between two characters for them to be considered characters with shared traits. To control for this, the VNDB data was also analysed according to this decomposition by number of traits for a higher threshold of seven shared traits needed to consider two characters to be characters with shared traits. The resulting pattern again followed the one in Table 5, thus the results do not seem to be specific to the threshold of five shared traits.

23 One could even argue that the presence of the temporal effect in the lower range of number of character traits is precisely the evidence confirming Azuma's claim, since

What if character production in Japanese manga, anime, and video games has been database-like all along?

Based on the results from the regression analysis, it seems that the hypothesis we started out from has not been substantiated. Before continuing to the theoretical implications of these results, let us first consider some of the potential limits to the findings. First, and most importantly, although the regression analysis on the whole VNDB data had a temporal range of 1989–2019, the regression analyses on the data decomposed by number of traits only had a publication year range of 1998–2019 at the most. Since Azuma is talking about a change that had occurred in recent years, from the point of view of 2001, when the book was originally published, it would be quite acceptable to argue that the decomposed data only captures the state of affairs after the shift had happened. Therefore, if it did not find anything it is perhaps due to this limited temporal range in the analysis. The same argument could be leveraged against all of the analyses conducted on the ACDB visual novel characters dataset as well. As for the ACDB other characters dataset, although it does indeed span a potentially long enough timeframe for at least the overall dataset and a number of the data subsets, it is mostly made up of data on anime characters, and thus does not fit Azuma's argument completely, as that concerns otaku specific media, such as visual novels.

Having established some of the potential limitations to the validity of the findings it is time to examine what all of this means for reconsidering Azuma's work and the way characters are produced. Azuma is mostly concerned with practices of consumption and the way those are changing in late modernity, but he nevertheless addresses the production side as well. Depending on our reading he is either implying a corresponding paradigm shift happening on the side of production, or at the very least points out the emergence of the proliferation of templated character creation.

One of the implications of the present analysis is that there was maybe no paradigm shift going on on the production side of character creation (either for otaku-like characters or characters in more general, as demonstrated by the results of the ACDB other characters dataset), or to address the narrower

---

it could be taken to imply that the effect of replicating templates, which based on our results has always been present for the more popular characters, has over time become increasingly present for less popular characters as well. This interpretation, however, would overlook the fact that Azuma never states that this effect has always been present for the more popular characters. Thus, we would once again end up with an argument, where Azuma's point about the temporal change in relation to character production practices would prove to be at least partially unsubstantiated.

interpretation, that the templated creation of characters is far less of a new phenomenon. Changing the optics through which we look at this question, it suddenly becomes more than obvious that an equally strong argument could be made for why the production side has always operated in the manner Azuma describes. We could, of course, argue that inspiration is always part influence, and character creation has always drawn on preceding works, but we can make an even stronger case for the tendency to draw on a pool of available character templates in the case of Japanese manga and anime. Going back to the fountainhead of both modern story manga and limited animation (or television anime) in Japan, Osamu Tezuka, we find that it would not be hard to put forth an argument that this mode of production, namely relying on the elements of "the database," is how things were set up to be done from the very beginning. Firstly, Tezuka is famous for starring the same cast of characters in various works as different characters, which as far as templates go is the end point along the continuum from original to copy in the direction of re-use. Secondly, the limited animation techniques adopted by Tezuka for *Astro Boy*, which set the model for most television anime to follow, also heavily rely on the re-use of character elements.

> When it comes to animating characters, it is true that limited animation tends to move as little of the figure as possible and to reuse as much of the figure as possible. With faces, for instance, the eyebrows, eyes, or the mouth may move but nothing else; and drawings of the face seen from a couple different angles are used again and again. Likewise with the animation of bodies, the legs and arms may move, but nothing else. Limited animation tends toward the production a series of cel copies of the same body or face, and minor additions are made to them as you use them. The best way to assure maximum reuse of figures and bits of figures is to develop a cel bank, so you can piece together different scenes and different movements by assembling elements already drawn. The cel bank prepares the way for a relation to characters based on assembly—it forms the basis for the overlap between animation and garage kits and models (self-assembled characters) as well as an overlap between cel animation and the customizable characters of many videos games. […] The cel bank provides the assembly diagrams for taking apart and piecing together animated life forms. (Lamarre 2009, 192)

Following on from Lamarre's description of the cel bank, it is quite easy to imagine how the materiality of the animation process would inspire a similar approach in the act of character creation both as a matter of economic necessity (saving money by using already available elements from the cel bank) and as a model for creativity. Although I will not attempt to provide a well-researched argument that this approach to character creation in Japanese anime and manga has, indeed, been the dominant form throughout their history, I hope to have provided a few convincing pointers for how such a position would be quite plausible to consider.

However, the fact that Azuma might have only projected the changes he discusses in relation to consumption practices on to the production side without there being any substantial changes there in reality, in no way poses a significant challenge to his book's overall argument, as it hinges on his discussion of consumption practices. Nevertheless, potentially amending his argument in this way helps better draw out its connections and indebtedness to Toshio Okada's 1996 book *Otaku-gaku Nyumon* (Introduction to Otakuology), which is conspicuously absent from Azuma's book's references, even though it is highly unlikely that he did not read it, especially in light of his explicit use of the generational model of otaku introduced by Okada.

In *Otaku-gaku Nyumon* Okada, also focusing on the consumption side, explained how early otaku would take note of and catalog the differences in the television anime series they enjoyed, in a way attempting to reverse engineer the production process. They started to understand the connections between the end credits and the changes in the looks of the characters, the animation style, or the structure of the story. To rephrase this, in Azuma's vocabulary these early otaku were invested in understanding the underlying structure of the "database" that underpinned their favourite shows.

Thus, by positing that the production side has to a certain degree always followed this model of relying on the "database" of character elements, the corresponding argument would be that it is only now that the consumption side has caught up to it,[24] and thereby also made these tendencies more explicit on the production side as well. This might seem like a minor shift in emphasis, but it would definitely require a stronger acknowledgement of Okada's work in relation to highlighting the way otaku have from the start been engaged with the database aspect of the production side of anime. Which is very much in line with the way Lamarre treats the question of the connection between Okada's and Azuma's work, through the image of the exploded projection, which serves as a central metaphor in his book, *The Anime Machine* for the way Okada and studio GAINAX approach anime:

24 At least for a wider audience since Okada's argument is precisely about how this form of knowledge accumulation and sharing has been going on for a long time within the circles of the most dedicated fans. Similar to Jenkins (1992), he points out how the home video recorder greatly catalyzed this process of close reading of favourite media texts. And the mass adoption of internet use, of course, made the sharing and accessibility of these types of information for increasingly wider audiences even easier.

> In fact, I would go so far as to say that the underlying structure in Azuma, which he calls database structure, is actually exploded projection. (Lamarre 2009, 260)

Furthermore, this would also mean that the relevance of the ACDB other characters dataset for evaluating the initial hypothesis should be reevaluated. If otaku have been working on deciphering the "database" behind their favourite anime even before the rise of so-called otaku specific media products, then it seems quite relevant to also consider the lack or existence of the temporal trend we were looking for in the ACDB other characters dataset as well. Thus, the fact that there was no temporal effect to be found in that dataset either, further strengthens the argument in relation to this proposed amendment to Azuma's argument. This last point is also a testament to the way theoretical and empirical investigations mutually rely on each other to make sense of the phenomena in question.

Summary

The metadata databases compiled by fan or enthusiast communities, while holding certain challenges, such as the incompleteness of their data, are a rich and granular resource that can be harnessed for academic research. The present chapter has demonstrated this applicability by testing one of the points from Azuma's book *Otaku: Japan's Database Animals* (2009 [2001]) on two of the databases processed within the framework of the Japanese Visual Media Graph project. By having found no strong evidence in the data to support the hypothesis that the creation of new characters has become increasingly reliant on popular templates, we have stumbled upon an equally interesting position. Namely, the possibility that the production side of Japanese anime and manga has always operated in a manner congruent with Azuma's database description. Should this indeed be the case, it would mean only a minor adjustment to the book's overall argument. However, it would help bring to the fore Azuma's indebtedness to Okada's work (1996), further substantiating the connection between the two authors' positions already highlighted by researchers like Lamarre (2009).

ORCID®

Zoltan Kacsuk https://orcid.org/0000–0002-6733–3294

References

Aarseth, Espen. 2004. "Playing Research: Methodological approaches to game analysis." In *Game Approaches /Spil-veje. Papers from spilforskning.dk Conference, august 28.-29. 2003.* Denmark: Spilforskning.dk.

Abercrombie, Nicholas and Brian Longhurst. 1998. *Audiences: A Sociological Theory of Performance and Imagination*. London: Sage.

Azuma, Hiroki. 2009 [2001]. *Otaku: Japan's Database Animals*. Introduction and translation by Jonathan E. Abel and Shion Kono. Minneapolis, MN: University Of Minnesota Press.

Bamparopoulos, Giorgos, Evdokimos Konstantinidis, Charalampos Bratsas, and Panagiotis D. Bamidis. 2016. "Towards Exergaming Commons: Composing the Exergame Ontology for Publishing Open Game Data," *Journal of Biomedical Semantics* 7 (4), 1–15. https://doi.org/10.1186/s13326-016-0046-4.

Bonenfant, Maude, Fabien Richert and Patrick Deslauriers. 2017. "Using Big Data Tools and Techniques to Study a Gamer Community: Technical, Epistemological, and Ethical Problems," *Loading... The Journal of the Canadian Game Studies Association* 10 (16), 87–108.

de Freitas, Sara. 2018. "Are Games Effective Learning Tools? A Review of Educational Games," *Educational Technology & Society* 21 (2), 74–84.

Durkheim, Emile. 2002 [1897]. *Suicide: A Study in Sociology*. Second Edition. London: Routledge.

El-Nasr, Magy Seif, Anders Drachen, and Alessandro Canossa, eds. 2013. *Game Analytics: Maximizing the Value of Player Data*. London: Springer.

Freybe, Konstantin, Florian Rämisch, and Tracy Hoffmann. 2019. "With Small Steps to the Big Picture: A Method and Tool Negotiation Workflow." In *Proceedings of the Twin Talks Workshop at DHN 2019. Proceedings of the Twin Talks Workshop at DHN 2019, co-located with Digital Humanities in the Nordic Countries (DHN 2019)* (*CEUR Vol-2365*), edited by Steven Krauwer and Darja Fišer, 13–24. Aachen: CEUR-WS.org. http://ceur-ws.org/Vol-2365/03-TwinTalks-DHN2019_paper_3.pdf

Fukuda, Kazufumi, Tetsuya Mihara, and Kosuke Oishi. 2020. "A Development of LOD Dataset for Description of Video Game Resources and the Research Utilization." *Proceedings of 10th Annual Conference, Digital Games Research Association JAPAN*, 31–34. Japan: Digital Games Research Association Japan.

Greitemeyer, Tobias, and Dirk O. Mügge. 2014. "Video Games Do Affect Social Outcomes: A Meta-Analytic Review of the Effects of Violent and Prosocial Video Game Play." *Personality and Social Psychology Bulletin* 40 (5), 578–589.

Hills, Matt. 2002. *Fan Cultures*. London: Routledge.

Hoffmann, Tracy, Konstantin Freybe, and Peter Mühleder. 2017. "Workflows zur datenbasierten Videospielforschung. Am Beispiel der populären Videospielserie Metal Gear Solid." In *INFORMATIK 2017*, edited by Maximilian Eibl and Martin Gaedke, 1113–1124. Bonn: Gesellschaft für Informatik.

Itō, Gō. 2005. *Tezuka izu deddo. Hirakareta manga hyōgenron e* [Tezuka is Dead: Opening up theoretical approaches to manga representation]. Tokyo: NTT Publishing.

Jenkins, Henry. 1992. *Textual Poachers: Television Fans & Participatory Culture*. London: Routledge.

Kacsuk, Zoltan. 2016. *From Geek to Otaku Culture and Back Again: The Role of Subcultural Clusters in the International Dissemination of Anime-Manga Culture* *as seen through Hungarian Producers*. Ph.D. dissertation. Kyoto: Kyoto Seika University.

Lamarre, Thomas. 2009. *The Anime Machine: A Media Theory of Animation*. Minneapolis, MN: University Of Minnesota Press.

Lankoski, Petri, and Staffan Björk, eds. 2015. *Game Research Methods: An Overview*. Pittsburgh, PA: ETC Press.

Lee, Jin Ha, Clarke, Rachel Ivy, and Andrew Perti. 2015. "Empirical Evaluation of Metadata for Video Games and Interactive Media," *Journal of the Association for Information Science and Technology* 66 (12), 2609–2625.

Little, Roderick J. A. and Donald B. Rubin. 2020. *Statistical Analysis with Missing Data, Third Edition*. Hoboken, NJ : Wiley.

Lukacs, Andras. 2014. "Sociology." In *The Routledge Companion to Video Game Studies*, edited by Mark J. P. Wolf and Bernard Perron, 407–414. New York: Routledge.

Lyotard, Jean-François. 1984 [1979]. *The Postmodern Condition: A Report on Knowledge*. Translated by Geoff Bennington and Brian Massumi, foreword by Fredric Jameson, MN: University Of Minnesota Press.

Myers, David. 2014. "Research." In *The Routledge Companion to Video Game Studies*, edited by Mark J. P. Wolf and Bernard Perron, 331–338. New York: Routledge.

Okada, Toshio. 1996. *Otakugaku nyūmon* [Introduction to otakuology]. Tokyo: Ōta.

Picard, Martin, and Jérémie Pelletier-Gagnon. 2015. "Introduction: Geemu, Media Mix, and the State of Japanese Video Game Studies." *Kinephanos* 5 (December 2015), 1–19.

Roth, Martin. Forthcoming. *Capturing Cultural Complexity: Metadata Pathways to Console Videogame Spatialization*.

Royal, Cindy, and Deepina Kapila. 2009. "What's on Wikipedia, and What's Not…?: Assessing Completeness of Information," *Social Science Computer Review* 27 (1), 138–148.

Schäfer, Fabian, and Martin Roth. 2012. "Otaku, Subjectivity and Databases: Hiroki Azuma's *Otaku: Japan's Database Animals*." *Digital Culture & Education* 4 (2), 208–211.

Sköld, Olle, Suellen Adams, J. Tuomas Harviainen, and Isto Huvila. 2015. "Studying Games from the Viewpoint of Information." In *Game Research Methods: An Overview*, edited by Petri Lankoski and Staffan Björk, 57–73. Pittsburgh, PA: ETC Press.

Suan, Stevie. 2017. "Anime's Performativity: Diversity through Conventionality in a Global Media-Form," *animation: an interdisciplinary journal* 12 (1), 62–79.

Utsch, Sofia, Luiza C. Bragança, Pedro Ramos, Pedro Caldeira, and Joao Tenorio. 2017. "Queer Identities in Video Games: Data Visualization for a Quantitative Analysis of Representation." In *Proceedings of SBGames 2017*, 847–854. Curitiba, PR: Sociedade Brasileira de Computação.

Yoshimoto, Taimatsu. 2009. *Otaku no kigen* [The origin of otaku]. Tokyo: NTT.

Comment by Shunsuke Mukae

The academic use of fan community works is often criticized. Some say it is a form of exploitation or treating them with a lack of respect. The uproar over the analysis of pixiv submissions in 2017 is a typical example. But it is also unavoidable in the current data-driven climate. This essay tries to show a way forward for such future development, using Azuma's famous thesis as an auxiliary line. The careful explanation of the deficiencies and biases in the data shows the author's sincerity and makes the essay more credible.

Yet, is it appropriate to reinterpret Azuma's argument, which is limited to the Japanese situation, using the English-based Visual Novel Database (VNDB) where even characters in non-Japanese games are registered?

And does Tezuka's reuse of characters in his works and the example of cell banks in the anime production support the author's claim that the anime industry has been using character templates? In this case, it seems more appropriate to compare this with another medium, such as film.

Thus, there are questions in this essay that remain unsolved. Despite this, the author's attempt to bridge Azuma's theory of consumption with that of production is fascinating. Regardless of the validity of the conclusions, the essay is worth referring to as a case study of the academic use of fan community works.

Comment by Peter Mühleder

This chapter provides valuable insights into the possibilities and problems that arise when working with fan-created digital resources. In terms of the study of pop culture, these resources can provide important insight into fan practices, but – as the author shows – also facilitate a "distant reading" of trends and developments in a specific cultural field. But these opportunities do come with a price. Online data, even structured data from databases like that used in this case, is often incomplete, messy, and incompatible. Data cleaning, integration, and linking are laborious and difficult tasks that rarely get the attention they deserve. Therefore, the technical and methodological rigour that the author openly describes and employs to deal with these issues is impressive and acts as a great example of how to tackle such matters.