



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The culture of the very rich and very poor

Do museum digital collections tell us anything about Jewish culture?

Citation for published version:

Kizhner, I, Terras, M, Afanaseva, J, Pusenkova, D, Sherer, M & Skorinkin, D 2022, The culture of the very rich and very poor: Do museum digital collections tell us anything about Jewish culture? in G Zaagsma, DS Ben Ezra, M Rürup, M Margolis & AS Levi (eds), *Jewish Studies in the Digital Age*. Studies in Digital History and Hermeneutics, vol. 5, De Gruyter Oldenbourg, pp. 43-64, #DHJewish - Jewish Studies in the Digital Age, Luxembourg, Luxembourg, 12/01/21. <https://doi.org/10.1515/9783110744828-003>

Digital Object Identifier (DOI):

[10.1515/9783110744828-003](https://doi.org/10.1515/9783110744828-003)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Jewish Studies in the Digital Age

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Inna Kizhner, Melissa Terras, Julia Afanasieva,
Diana Pusenkova, Maria Sherer, and Daniil Skorinkin

The Culture of the Very Rich and Very Poor: Do Digital Museum Collections Tell us Anything about Jewish Culture?

Abstract: Digital approaches to Jewish Studies allow collecting data from multiple sources. This enlarges the picture, makes our understanding of historical experiences more complete, and triggers further research questions. This chapter compares samples of artworks from digital online collections of the Metropolitan Museum of Art in New York and the Russian State Catalogue of Museum Collections in the Russian Federation. We show that the difference in their temporal and geographical coverage of Jewish artworks/historical documents often results from tagging, naming and understanding of what is deemed Jewish. The collective decisions made in these datasets, by many individuals, and differing institutions about cataloguing, indexing, and returning searches on Jewish culture in a digital age inform what parts of Jewish culture are accessible. The technological and data-led decisions become a part of multiple layers of decisions that inform how primary sources are formed in a digital age.

Keywords: online museum collections, cultural heritage, indexing, cataloguing, bias, Metropolitan Museum of Art, Russian State Catalogue of Museum Collections

1 Introduction

Data-driven research in Jewish Studies has been used to analyze historical Hebrew newspapers,¹ manuscripts,² how Jewish content appears in cultural heritage aggre-

1 O. Soffer et al., “Computational Analysis of Historical Hebrew Newspapers: Proof of Concept,” *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.

2 M. Zhitomirsky-Geffet and Gila Prebor, “SageBook: Toward a Cross-Generational Social Network for the Jewish Sages’ Prosopography,” *Digital Scholarship in the Humanities* 34, no. 3 (September 2019): 676–95; G. Prebor, Maayan Zhitomirsky-Geffet, and Yitzchak Miller, “A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types,” *Digital Scholarship in the Humanities* 35, no. 2 (June 2020): 441–58.

gators such as Europeana,³ and to evidence the politics of digitization as revealed through textual collections.⁴ However, analysis has not yet focused on how Jewish culture is represented within datasets of museum objects. The aim of this chapter is to present the results of an exploratory study to establish a methodology that lets us understand Jewish culture (or other minority cultures) as represented within digital collections in different parts of the world. The comparison of search results for the Metropolitan Museum of Art in the United States and those for the State Catalogue of Museum Collections of the Russian Federation will allow us to understand the effect that a variety of collecting and cataloging approaches may have on research at the crossroads of Digital Humanities and Jewish Studies. This will also help us to examine how the status of a minority population within different societies has affected the institutional and collections' response to their culture. In doing so, we can see a diversity of cataloging approaches for different cultural environments and the effects of data-driven analysis when studying minority cultures.

Bias in cultural heritage data and representations of human knowledge in mass-digitized collections have been widely covered in recent literature.⁵ Such work also relates to epistemic complications of data collecting, data cleaning, and model training,⁶ and the difficulties imposed by “the infrastructures of knowledge-making.”⁷ Apart from selection and exclusion bias,⁸ these “infrastruc-

3 Dov Winer, “Judaica Europeana: An Infrastructure for Aggregating Jewish Content,” *Judaica Librarianship* 18 (2014): 88–115.

4 G. Zaagsma, “Digital History and the Politics of Digitization,” Paper presented at *Digital Archive and Canon Workshop*, March 10, 2021, accessed May 11, 2021, https://www.digitales-archiv-und-kanon.de/contributions/Zaagsma_en.pdf.

5 See, for example, B.H. Daru et al., “Widespread Sampling Biases in Herbaria Revealed from Large-Scale Digitization,” *New Phytologist* 217 (2018): 939–55; N.B. Thylstrup, *The Politics of Mass Digitization* (Boston: MIT Press, 2019); K. Bode, “Why You Can’t Model Away Bias,” *Modern Language Quarterly* 81, no. 1 (2020): 95–124; A. Liu, “Toward a Diversity Stack: Digital Humanities and Diversity as Technical Problem,” *PMLA/Publications of the Modern Language Association of America* 135, no. 1 (2020): 130–51; S. Bagga and A. Piper, “Measuring the Effect of Bias in Training Data for Literary Classification,” *Proceedings of LaTeCH-CLfL 2020*, Barcelona, Spain, December 12, 2020, 74–84.

6 A. Bechmann and G.C. Bowker, “Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media,” *Big Data & Society* 6, no. 1 (2019): 1–11.

7 B. Mak, “Archaeology of a Digitization,” *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–26, 1519, cited in Bode, “Why You Can’t Model Away Bias,” 2.

8 Maria Economou, “Heritage in the Digital Age,” in *A Companion to Heritage Studies*, ed. William Logan, Máiréad Nic Craith, and Ullrich Kockel (Chichester, UK: Wiley, 2016), 215–28; Bode, “Why You Can’t Model Away Bias”; T. Hauswedell et al., “Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives

tures of knowledge-making” include misrepresentations caused by metadata conventions,⁹ cataloging approaches,¹⁰ classification principles,¹¹ and linguistic issues.¹² How things are named, classified and cataloged, and what contexts are selected to denote the meaning of artworks and documents strongly influences whether users are able to find minority cultures in the wealth of digitized cultural heritage we have. Perspectives and surrounding contexts may vary in different cultural environments as they tend to reflect the values of catalogers during a historical period. Only by understanding how cataloging principles and linguistic approaches influence the results of data retrieval, can we develop a methodology for studying minority cultures in contexts that combine big data approaches and epistemic dependence on the historical perceptions of data. We need to make explicit the principles that govern how information infrastructures shape our perceptions of Jewish culture and other cultures. We are always at risk of losing what we do not count and losing what we do not help users to see.¹³ On the other hand, whatever classification instruments and cataloging principles we employ, some cultural objects tend to be always left behind. This happens because multiple “sieves” or instruments have different users in mind.¹⁴ This means that due to their coding cultures and functionality they tend to exclude some objects, and therefore potential users, along the way. These lost objects and associated cultural contexts may be important regarding minority cultures where the risk of losing contexts is especially high.¹⁵

of Historical Newspapers,” *Archival Science* 20 (2020): 139–65; Zaagsma, “Digital History and the Politics of Digitization.”

9 M.V. Fernandez, “The Coloniality of Metadata: A Critical Data Analysis of the Archive of Early American Images at the John Carter Brown Library,” PhD thesis, University of Texas, 2018; Zaagsma, “Digital History and the Politics of Digitization.”

10 N.J. Bingham and H. Byrne. “Archival Strategies for Contemporary Collecting in a World of Big Data: Challenges and Opportunities with Curating the UK Web Archive,” *Big Data & Society*, January 2021.

11 G.C. Bowker and S.L. Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999); K. Cotter et al., “‘Reach the Right People’: The Politics of ‘Interests’ in Facebook’s Classification System for Ad Targeting,” *Big Data & Society*, January 2021.

12 J. Aguilera, “Another Word for ‘Illegal Alien’ at the Library of Congress: Contentious,” *The New York Times*, July 22, 2016, accessed April 25, 2021, <https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html>.

13 G.C. Bowker, “Biodiversity Datadiversity,” *Social Studies of Science* 30, no. 5 (2000): 643–83.

14 J. Likhter, personal communication, May 7, 2021.

15 J. Likhter, personal communication, May 7, 2021.

2 Methods

Our sources of data for analysis were the entirety of the online digital collections of the Metropolitan Museum of Art¹⁶ in New York and the State Catalogue of the Museum Collections in the Russian Federation.¹⁷ Our choice of online digital collections from two different countries and different contexts in cultural heritage was determined by our aim to see the difference (if any) in search results for “jew,” “jewish,” and related search terms in different settings. We sought to know whether and how the representation of objects reflected political, social, and epistemic attitudes in these different parts of the world, and between a leading Western institution, and a national catalog, to amplify difference in approach to test if this methodology would yield results. There is no national catalog in the USA against which to compare the Russian catalog, so the Metropolitan Museum of Art was chosen given it “collects, studies, conserves, and presents significant works of art across all times and cultures in order to connect people to creativity, knowledge, and ideas.”¹⁸ Both collections detail a significant number of objects published online, are accessed by large numbers of users, and allow similar access to metadata, which facilitated comparison. The online digital collection of the Metropolitan Museum of Art included over 400,000 objects in 2017¹⁹ and it attracted 8 million visitors in 2018.²⁰ The Russian State Catalogue includes images for over 24 million museum objects at the time of writing, which makes it one of the largest national aggregators of cultural heritage across the world. It was built for inventory purposes, representing a third of Russian museum collections covering almost all state museums across the country. This makes it a representative dataset to study Jewish culture in the Russian environment. The comparison of results for the Metropolitan Museum of Art and those for the Russian State Catalogue will allow us to examine how the status of a minority population within different societies and contexts has affected the institutional and collections’ response to culture in societies with different political and collections traditions. We did not compare the results for the Russian State Catalogue with those from

16 <https://www.metmuseum.org/>, accessed April 25, 2021.

17 <https://goskatalog.ru/portal/>, accessed April 25, 2021.

18 Metropolitan Museum of Art, “About the Met,” 2021, accessed June 25, 2021, <https://www.metmuseum.org/about-the-met>.

19 T. Navarette and E. Villaespesa, “Digital Heritage Consumption: The Case of the Metropolitan Museum of Art,” *magazén* 1, no. 2 (December 2020).

20 Metropolitan Museum of Art, “Met Museum Sets New Attendance Record with More Than 7.35 Million Visitors,” 2018, accessed May 12, 2021, <https://www.metmuseum.org/press/news/2018/met-museum-sets-new-attendance-record>.

a large international aggregator of cultural heritage images, such as Google Arts and Culture or Europeana Collections. This is because we wanted to focus on the epistemic path dependence, and the political and social attitudes that national and institutional digital infrastructures maintain in producing knowledge. The content of aggregators can be studied at a later date so that we see a combined approach to representations of Jewish culture or other minority cultures.

We used the collections' search engines for museum objects matching "jew," "jewish," "hebrew," and "yiddish" search terms. Russian search terms were "еврей" as a lemma because the Catalogue's search results can return results for parts of words and this is important for a language with numerous word forms, such as Russian. Other search terms were "иврит" for "hebrew," "идиш" for "Yiddish," and "древнееврей (ский)" for "Old Jewish," which is an equivalent to "hebrew" for religious, biblical, or literary contexts. We did not use the lemma "иудей" or "Judaic" as opposed to "Christian," although this search term returned over 500 results for the Russian State Catalogue and two results for the Metropolitan Museum of Art's online digital collection. Neither did we use results for "Judaism," although this search term returned 200 results for the Metropolitan Museum of Art and 24 results for the Russian State Catalogue.

As of October 2020, our search yielded approximately 900 results for the Metropolitan Museum of Art's digital collection and about 6,300 results for the Russian State Catalogue (see Tables 1 and 2 for a detailed breakdown of results). To get the results that indeed influence the perception of Jewish culture, we used only the "artworks with images" option on the websites. We used a sample (10%) of the results from the Russian State Catalogue which we made more representative by conducting a random search in the 12 categories that were offered by the Catalogue's interface, such as "paintings," "sculpture," "graphics," and "rare books." After filtering out duplicates and irrelevant results, we were left with 407 records for the Metropolitan Museum of Art, while the sample from the Russian State Catalogue yielded 404 records after excluding duplicates. We tabulated the metadata and data from textual descriptions for the results of the search. The fields included the name and culture of artists, types of objects, dates or time periods, and geographical descriptions. We also included data on whether the artwork was or had been a part of an exhibition in the Jewish Museum in New York or any other results for an exhibition that contextualized search results. This could only be done for the Metropolitan Museum of Art as the Russian State Catalogue does not record this information. We preferred to include the geography that was not directly related to the place where an artwork was produced because it enabled us to contextualize the Jewish culture showing the breadth and depth of linkages. The textual data for the Metropolitan Museum of Art were a combination of metadata, extensive historical contexts revealed in textual descrip-

tions, and exhibition histories that placed an artwork in a variety of cultures (Figure 1). The information for the Russian State Catalogue was obtained from the titles (which were sometimes quite extensive), the date, and the place where an artwork was produced (Figure 2). In this way, the interface of the two digital collections structured, directed, and limited our results (see Section 3 for analysis). Biographical details for the artists whose artworks were returned following our search in the Metropolitan Museum of Art online digital collection were obtained from Wikipedia. The maps showing the distribution of search results and their historical contexts across space were produced using Adobe Illustrator. We obtained contemporary maps through Yandex Maps service.²¹ Maps of historical places, such as Mesopotamia, were more difficult to produce due to the uncertainty about their borders. Uncertain borders, derived from popular sources, such as Wikipedia, are shown as blurred lines to reflect this.

Table 1: The number of records and metadata returned by the search engine of the Metropolitan Museum of Art’s online digital collection when using “jewish,” “jew,” “hebrew,” and “yiddish” as search terms in autumn 2020.

Search terms	Number of records with metadata and images returned by the search engine
“jewish”	496
“jew”	210
“hebrew”	194
“yiddish”	4

Table 2: The number of records and metadata returned by the search engine of the Russian State Catalogue’s online digital collection when using “jewish,” “jew,” “hebrew,” “yiddish,” and “old jewish” as search terms in autumn 2020.

Search terms	Number of records with metadata and images returned by the search engine
“еврей” as a part of the word for “jew” and “jewish”	4,687
“идиш” for “yiddish”	1,510
“иврит” for “hebrew”	96
“древнееврей(ский)” for “old jewish”	33

²¹ Accessed April 25, 2021, <https://yandex.ru/maps/?ll=92.852572%2C33.461352&z=2>.

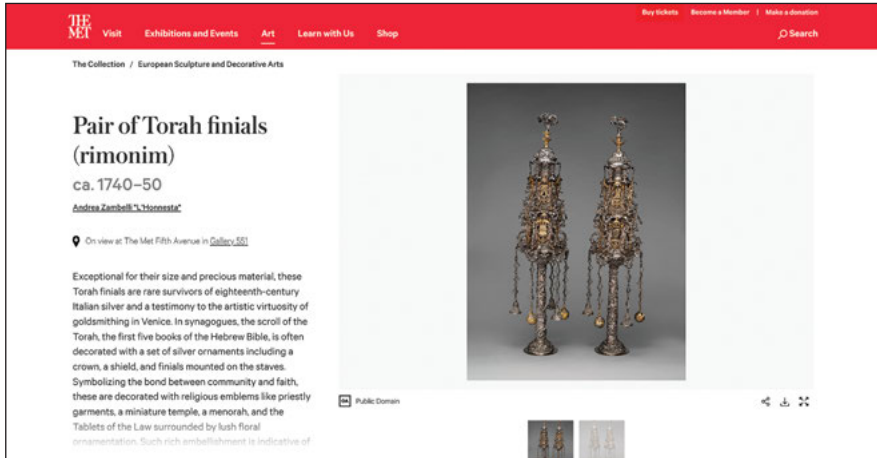


Figure 1: An example of a record returned by the search engine of the Metropolitan Museum of Art’s online digital collection when using “hebrew” as a search term. Images, metadata, and a textual description showing a pair of Torah finials. The Metropolitan Museum of Art, New York. Walter and Leonore Annenberg Acquisitions Endowment Fund, 2016.

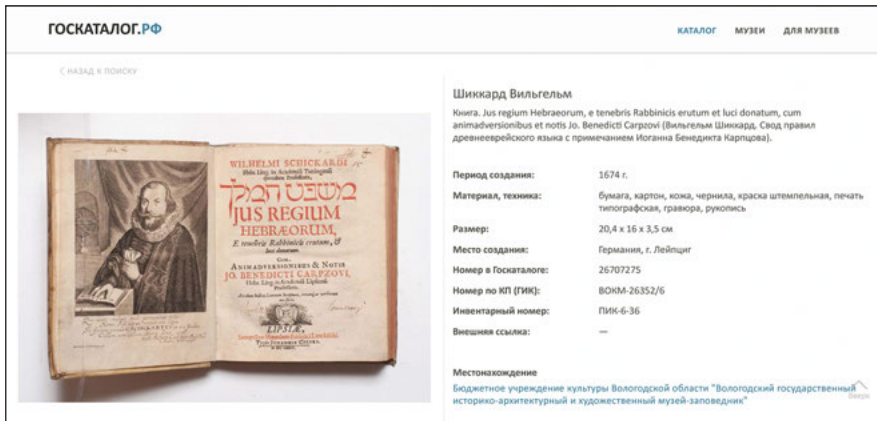


Figure 2: An example of a record returned by the search engine of the Russian State Catalogue’s online digital collection when using “old jewish” as a search term. An image with metadata showing Wilhelm Schickard’s book with rules for Old Jewish language. Vologda State Museum of History, Architecture and Art.

3 Results

3.1 Geographical Distribution of Artworks

Figure 3 shows the geographical distribution of artworks in our sample for the search conducted on the Metropolitan Museum of Art website. We can see a wide geographical coverage. The artworks related to the USA dominate in the sample but we can also see artworks related to Europe, Russia, Egypt, Israel, Northern Africa, and Mesopotamia. Figure 4 shows the results for the geographical distribution in our sample for the search conducted on the website of the Russian State Catalogue. The artworks in the sample are almost exclusively produced in, published in, or related to places within Russia or countries that used to be a part of the Russian Empire or the Soviet Union. This representation is unrelated or weakly related to either historical or contemporary cultures but rather shows Jews as a separate ethnicity in Russian society. Conversely, in the representation of the Metropolitan Museum of Art, Jewish culture is embedded in its relations with neighboring countries and cultures. What is more important, the Metropolitan Museum of Art's online digital collection represents Jewish culture as a part of the Western canon²² through its relation to biblical geography and the nations that interacted and influenced Jewish culture from the foundations of their society. This evidence supports a wide historical span and is augmented with the evidence on 20th century artists closer to the newer end of the canon (Figure 5). The results for the distribution of the Metropolitan Museum of Art artworks differ dramatically from the results of the Russian State Catalogue. This happens because contexts for “Jewish” and “Hebrew” differ in the two online digital collections and the words that are used as search terms, as a result, often mean different things. This is connected with different histories of Jews, different cultural environments, different attitudes to reconstructing histories, different cataloging approaches and different meanings of the search terms in the two countries. To a large extent, the difference is determined by what is deemed Jewish by catalogers and institutions in the two countries. It is also explained by the structure of the interfaces of the two digital collections and by how (whether) the artworks are contextualized through exhibitions or titles (see Section 2).

²² N. Frye, *The Great Code: The Bible and Literature* (San Diego: Harcourt Brace Jovanovich, 1982); H. Bloom, *The Western Canon* (New York: Harcourt Brace, 1994).

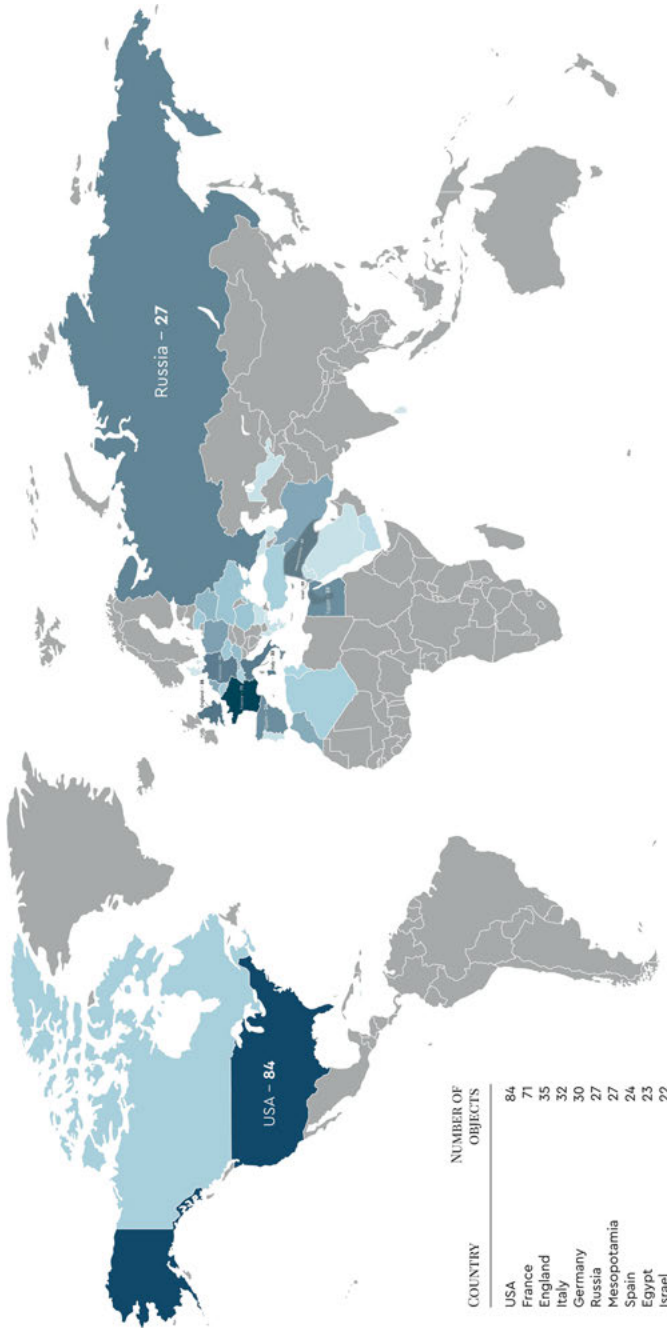


Figure 3: Geographical distribution of the records related to Jewish heritage in the Metropolitan Museum of Art, New York. Map: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Danil Skorinkin.

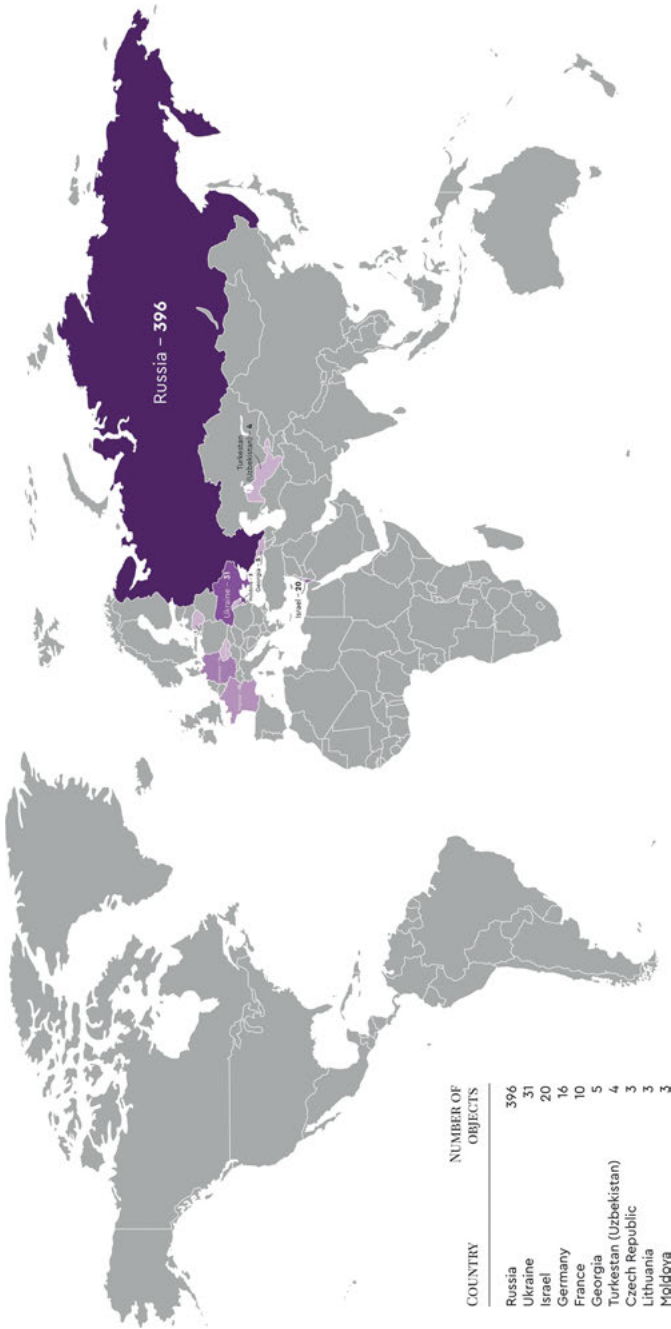


Figure 4: Geographical distribution of the records related to Jewish heritage in the State Catalogue of Russian Museum Collections. Map: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.

3.2 Distribution of Artworks Across Time

Figure 5 shows the difference in the distribution of artworks across time for the two samples. It demonstrates a pronounced peak in early 20th-century artworks from the Metropolitan Museum of Art, and an even distribution going back to the fourth millennium BC. Combined with a significant number of avant-garde Jewish artists in this dataset (see Section 3.3), the results show a span across time and space that ranges from the “exemplary ancient” to “model moderns.”²³ This happens because the Metropolitan Museum of Art has a metadata field that mentions all the places where artworks were exhibited. If such a place has “jewish” in its title, such as the Jewish Museum in New York, the artwork will be returned as a search result by the Metropolitan Museum of Art search engine. As a result, artworks are linked to the exhibitions at the Jewish Museum in New York in the period between 1956 and 2018,²⁴ and the Biblical Archaeological Exhibition at the University of Wisconsin-Madison Department of Hebrew and Semitic Studies (1975).²⁵ The links to the “Colmar Treasure” exhibition at the Metropolitan Museum of Art in 2019²⁶ show the possessions of a Jewish community in medieval Germany. The results linked to such exhibitions relate to 157 search results or about 40% of the dataset. Tagging systems and classification approaches mean that artworks from Assyrian, Babylonian, Sumerian, medieval, and contemporary cultures are included in the artworks returned from our queries and are therefore present in the dataset and analysis. Conversely, the Russian State Catalogue does not have this functionality and, consequently, the temporal and geographical spans of the sample are much narrower due to a lack of information and context. Figure 5 shows that artworks related to Jewish identity in the Russian State Catalogue date back to the 17th century and have little relation to biblical contexts or to the part of the European cultural canon that is associated with the history of Mesopotamia and ancient civilizations. The trough in the temporal distribution in the 1940s–1960s in the artworks from the Russian State Catalogue, which is most

²³ Yuri Slezkine, *The Jewish Century* (Princeton, NJ: Princeton University Press, 2002).

²⁴ See, for example, Jewish Museum, New York, “Russian Jewish Artists in a Century of Change, 1890–1990,” September 21, 1995–January 28, 1996.

²⁵ University of Wisconsin-Madison Department of Hebrew and Semitic Studies, “The Book and the Spade” (Biblical Archaeological Exhibition), April 13–May 4, 1975.

²⁶ Metropolitan Museum of Art, “The Colmar Treasure: A Medieval Jewish Legacy,” July 22, 2019–January 12, 2020, accessed April 30, 2021, <https://www.metmuseum.org/exhibitions/listings/2019/colmar-treasure-medieval-jewish-legacy>.

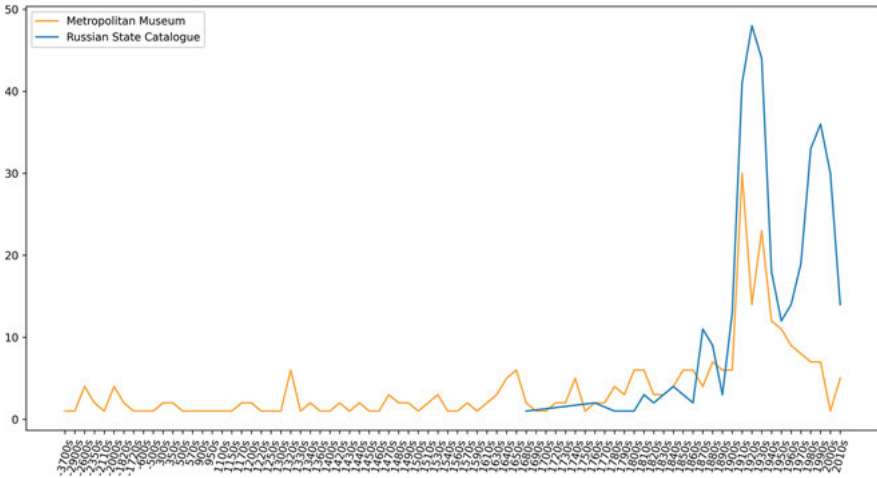


Figure 5: Temporal distribution of collections related to Jewish Studies in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Daniil Skorinkin, Inna Kizhner, Melissa Terras, Julia Afanasieva, Diana Pusenkova, and Maria Sherer.

pronounced directly after World War II, is explained by the famous anti-Jewish campaigns in the Soviet Union at the end of the 1940s.²⁷

3.3 Distribution of Artworks by Topics and Types of Art

As shown in Figure 6, Jewish identity in the sample of search results returned by the Metropolitan Museum of Art’s search engine is linked to contemporary artists, biblical contexts, and Mesopotamia (ancient civilizations). The perception of “Jewishness” is broadened by the inclusion of relations between Jews and Muslims or Sephardic Jews. On the other hand, Figure 6 shows that Jewish identity in the sample of search results returned by the Russian State Catalogue relates to high Yiddish culture and is linked to Jewish (Yiddish) theater, literature, and music. This is because the Russian State Catalogue returns numerous results for the Jewish State Theatre (over 30% of the search results for “Jewish” as a search term – over 1,500 records – where plays were performed in Yiddish). Other results from the sample are linked to books translated from (into) Yiddish and musical scores where texts were written in Yiddish.

²⁷ See, for example, B. Pinkus, *The Jews of the Soviet Union: The History of a National Minority* (Cambridge: Cambridge University Press, 1988).

Figure 7 shows the distribution of records by type of art for the most frequent types in the Metropolitan Museum of Art and Russian State Catalogue. We can see the prevalence of books for the Russian State Catalogue and the dominance of paintings for the Metropolitan Museum of Art. This cannot be explained by the fact that the Russian museums and, consequently, the Russian State Catalogue collect books rather than paintings. Indeed, paintings exceed books in the collection of the Metropolitan Museum of Art but there are twice as many paintings as books in its online digital collection.²⁸ However, the dominance of paintings over books from this online digital collection is much more pronounced in the artworks related to Jewish identity in our sample. The share of books related to Jewish artworks in the sample of search results from the Russian State Catalogue is much greater than the share of all books in the Russian State Catalogue.²⁹ This happens because texts accompanying images in the State Catalogue are descriptive and rarely contextualize artworks, while books come with extensive titles, quite often with the note “translated from Jewish (Yiddish).” It means that biblical paintings come without an explanation of who is depicted and what biblical story inspired the artwork. In addition, in the Russian State Catalogue textual descriptions that are not a part of the title are not included as a field that is used for retrieval. It means that even if Judith, the Jewish heroine, is presented as such in the textual description, the image and metadata will not be returned as search results for the terms “Jewish” or “Hebrew.” Conversely, the Metropolitan Museum of Art’s search engine retrieves artworks if the search term is mentioned in any of the metadata fields displayed to the user. This explains why “paintings” is the most frequent category for the topical distribution of search results for the Metropolitan Museum of Art. A third of the paintings from the Metropolitan Museum of Art in our sample involve biblical and Christian themes and contexts. Two-thirds of the 93 paintings in the sample are 20th-century artworks, and half of them are by American Jewish artists, such as Max Weber (1881–1961). A quarter of them are by French Jewish artists, such as Chaim Soutine (1893–1943) or Mark Chagall (1887–1985). On the other hand, famous contemporary artists in the Russian State Catalogue are rarely retrieved as Jewish. While the Metropolitan Museum of Art’s search engine retrieves 11 of Chagall’s paintings tagged as Jewish in data annotations (19% of 59 Chagall works in the online digital collection), the Russian State Catalogue’s search engine produces four works tagged as Jewish or 2% of 185 records related to Chagall in this online digital collection). In the search results for the Metropolitan Museum of Art,

²⁸ Metropolitan Museum of Art, “Explore the Collection,” 2021, accessed June 25, 2021, <https://www.metmuseum.org/art/collection>.

²⁹ Russian State Catalogue (The State Catalogue of Museum Collections of the Russian Federation), 2021, accessed June 25, 2021, <https://goskatalog.ru/portal/#/>.

a significant proportion of contemporary paintings (42 records out of 57 artworks, or about 70% of the sample) was produced by artists who were born in the Russian Empire, such as Mark Chagall or Max Weber. This group also includes artists whose parents emigrated from the Russian Empire at the end of the 19th and the beginning of the 20th century. The textual descriptions for these records could have mentioned Yiddish, since these painters were from families where the Yiddish language and culture were a part of their life and a significant source of influence.³⁰ However, users can rarely retrieve the word Yiddish in the curatorial texts and data annotations of the Metropolitan Museum of Art's online digital collection. The "Yiddish" search term returns four artworks or about 1% of all search results for the sample of search results for the Metropolitan Museum of Art. This means that until recently Yiddish meant "wide association with marginality, mutability, or obsolescence,"³¹ and the museum community preferred to focus on other representations of Jewish culture. Conversely, searching for "Yiddish" in the Russian State Catalogue produces 1,510 results or 24% of results for the three search terms. However, nearly all the results (95% or 1,414 objects out of 1,510 results for this search term) are retrieved from the Regional Museum of the Jewish Autonomy in the Far East, reducing Yiddish culture to the idea of Jewish territorialism.³²

3.4 Linguistic Differences Related to Search Results

One of the results reflecting the difference in linguistic approaches is the use of "Old Jewish" (древнееврейский) to write about "Hebrew" and the use of "Jewish" (еврейский) to write about "Yiddish" in the 20th century in the results from the Russian State Catalogue. This observation agrees with literature showing that "Jewish" was used in Russian/Soviet publication standards to denote "Yiddish"³³ and with literature on the eradication of Hebrew publishing in the 1920s in

30 See, for example, B. Harshav and M. Chagall, *Marc Chagall and His Times: A Documentary Narrative* (Palo Alto, CA: Stanford University Press, 2004); J. Shandler, *Yiddish: Biography of a Language* (Oxford: Oxford University Press, 2020).

31 J. Shandler, *Adventures in Yiddishland: Postvernacular Language and Culture* (Berkeley: University of California Press, 2006).

32 Ellen Eisenberg, *Jewish Agricultural Colonies in New Jersey, 1882–1920* (Syracuse, NY: Syracuse University Press, 1995).

33 L. Kogan and S. Loesov, "Old Jewish Language," in *World Languages: Semitic Languages. The Akkadian Language. Northwest Semitic Languages*, ed. A. Belova, L. Kogan, S. Loesov, and O. Romanova. Russian Academy of Sciences. Institute of Linguistics (Moscow: Academia, 2009), 296–375 [Russian].

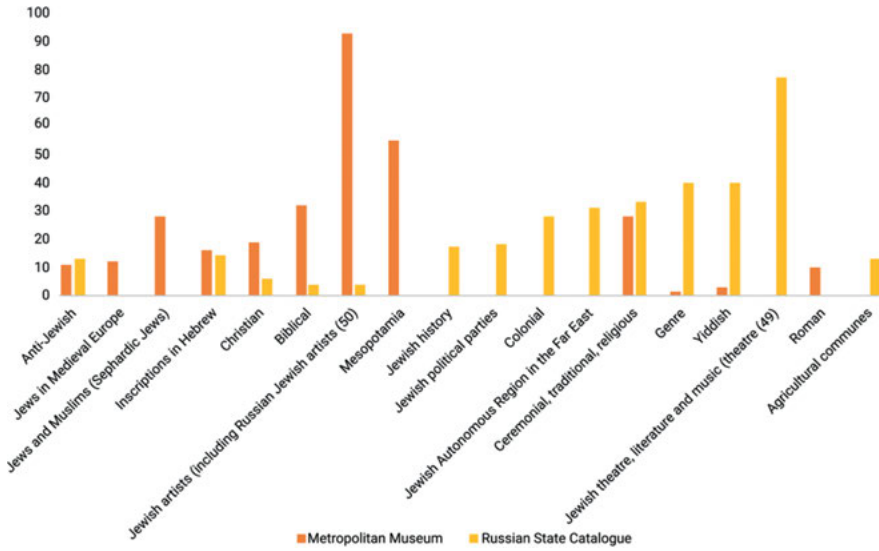


Figure 6: Distribution of records by topics for the most frequent types in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.

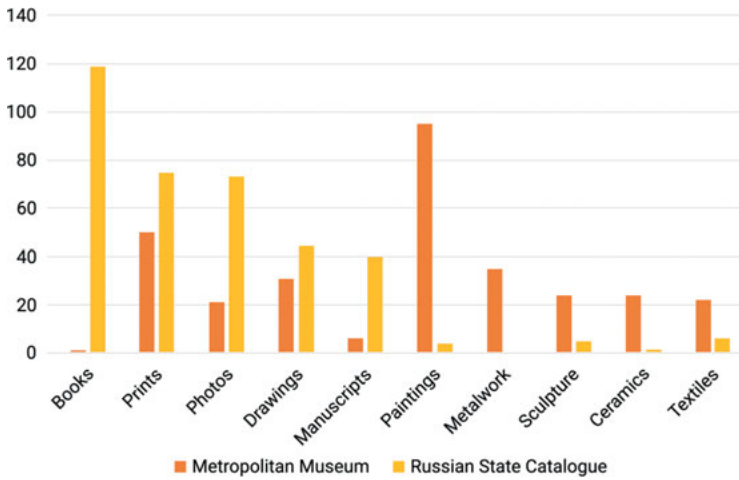


Figure 7: Distribution of records by type of art for the most frequent types in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.

Russia.³⁴ The latest item described with the expression “translated into Old Jewish” in the Russian State Catalogue is a book of Chekhov’s stories published in 1918. The author of the earliest book in our dataset that uses “Jewish,” meaning “Yiddish,” includes a caveat, “written in Jewish-German dialect.” The book was published in 1913. This means that the search for “Jewish” for the results dating back to the 20th century from the Russian State Catalogue will almost always return results that mean “Yiddish,” and they do not include biblical contexts, religious contexts, and the contexts of ancient civilizations. Such confusion of words and terminology related to the circumstances and environments of Jewish culture definitely influenced what was deemed Jewish, and what aspects of Jewish life were emphasized.

Different contexts in which the results of the search are displayed imply that they reflect different attitudes and, probably, different understandings of the Jewish identity in the two digital collections. The contexts in the Russian State Catalogue are very much about the Yiddish speaking population of the Russian Empire and the Soviet Union, using “the language of the Jewish street,”³⁵ while the contexts of the Metropolitan Museum of Art are very much about Jewish identity as a part of the European cultural canon and post-colonial perception of Jews and Muslims or Jews in Northern Africa, such as in the exhibition “The Cairo Geniza: Jews & Muslims in the Mediterranean World 800–1500” about Jews in Egypt.³⁶ However, sometimes the results related to this part of Jewish identity are mixed with the representation of colonialism and orientalism.³⁷ In this sense, the results from the Metropolitan Museum of Art are in line with the results from the Russian State Catalogue displaying colonial photography from Georgia, Dagestan, and Uzbekistan at the turn of the 20th century.

Our results show different ways of representing Jewish identity in metadata and collection management systems. In this way, Russian cataloging approaches or cataloging voices show differences in cultural or social attitudes that date back to the very first years after the Russian Revolution. They are very much about the formal policy and representation of Jewishness through the Yiddish language and literature. In the case of the Metropolitan Museum of Art, Jewish identity is given a broader context that encompasses biblical culture, ancient civilizations, medieval European culture, Northern Africa, and an age of modernity showing contemporary Jewish art in Europe and the United States (mostly in New York).

34 A. Blyum, “Hebrew Publications and the Soviet Censor in the 1920s,” *East European Jewish Affairs* 23, no. 1 (1993): 91–99.

35 Blyum, “Hebrew Publications and the Soviet Censor in the 1920s,” 98.

36 Jewish Museum, New York. “The Cairo Geniza: Jews and Muslims in the Mediterranean World 800–1500.” January 1, 1997–October 12, 1997.

37 Edward W. Said, *Orientalism* (New York: Random House, 1978).

4 Conclusion

This chapter showed that comparing metadata and catalog searches in this manner provides a methodology that lets us understand the representation of Jewish culture (or other minority cultures) by digital collections in different parts of the world, and in different contexts. Our results demonstrate that digital cultural heritage searches are complicated by a variety of cataloging voices and classification principles. The evidence that supports the argument of this chapter demonstrates how the layers of debates on “what is Jewish?” intertwine through collecting, tagging, linking, and publishing artworks online in very different contexts. Generalizations that can follow such analysis at scale vary widely, depending on political, cultural, and social perspectives. These perspectives are reflected in tagging principles, word usage, and classification approaches. What constitutes datasets under analysis, what texts, links and associating concepts are involved in producing contexts, determines our conclusions. Generalizations that emerge as a result produce contrasting, absolutely different representations of objects under study as shown in this chapter. One cannot but agree with Andrew Piper’s argument about our “failure to generalize *well*” and about the crisis of reproducibility in the humanities based on “case-driven research.”³⁸ Our chapter argues, however, that even analysis at scale does not contribute to better generalizations. To a certain extent, this happens because of a variety of cataloging conventions. Of course, the difference in results between the two chosen sources can also be assigned to different collections practices, since this exploratory research is comparing the curated collection of a large institution (the Metropolitan Museum of Art) with the portal to the collections of an entire nation (the State Catalogue of Museum Collections of the Russian Federation). Further work is needed on collections practices and processes, and how these approaches may have affected individual or national collections at different points of time, and we stress that the choice of which collections to compare is a crucial part of this design methodology.

The existence of multiple views and multiple approaches to museum annotation is not in itself a problem if our aim is to show that the truth is constituted by numerous contexts and does not depend on singular political and historical circumstances of building a collection. What is important, however, is making these circumstances and their influence explicit. In doing so we can demonstrate

³⁸ A. Piper, *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data* (Cambridge: Cambridge University Press, 2020), 4–5.

how data models at the foundation of information systems depend on “ethical and political values, modulated by local administrative procedures.”³⁹

One of the ways of avoiding the complications of subjective interpretations and consequent difference in data models is reducing classifications to a list of categories from pre-developed ontologies as it is done in sciences. However, literature on biological databases in recent years has been devoted to the importance of contexts, the relevance of time and space, the historiography (provenance) of data diversity,⁴⁰ and semantic relations of synonyms and overlapping taxonomic classes.⁴¹ What holds for biological data relates to social and cultural data where the contexts of data production are highly relevant. The complexity of metadata standardization in this case is augmented by the necessity to have standards reflecting a variety of types of data content. While Iconclass,⁴² for example, gives a wide range of classes related to Jewish culture, they are mostly about religion, rituals, and religious institutions.⁴³ Participation in politics, Jewish welfare, the histories of Jewish territorialism, lay Yiddish and Hebrew culture, literature, and theater are difficult to catalog using the Iconclass perspective. In addition, the contexts of data provenance (how, when, and where data were produced) influence cataloging approaches and, consequently, affect the compatibility of data produced under different conditions. The layers of intertwined perspectives demonstrated in this chapter show how much work is to be done to achieve the compatibility of digital infrastructures for minority cultures.

What we have here is not only data produced in different spatial and temporal contexts. These are also data produced in cultural environments that were very different from (and sometimes hostile to) those of a minority culture. Such dispersal of attitudes and perspectives does not make easier the work of standardizing the approaches to cataloging Jewish culture. Future work may concentrate on how to approach the “overly optimistic” task of standardizing data structures and formats⁴⁴ for minority cultures. A possible problem here is how to select features to build classes if the size of a sample is too small or if a sample is not

³⁹ Bowker and Star, *Sorting Things Out*, 321.

⁴⁰ See, for example, Bowker, “Biodiversity Datadiversity”; L.M. Schriml et al., “COVID-19 Pandemic Reveals the Peril of Ignoring Metadata Standards,” *Scientific Data* 7 (2020): article 188.

⁴¹ Beckett W. Sterner, Nico M. Franz, and J. Witteveen, “Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies,” *History and Philosophy of the Life Sciences* 42, no. 1 (2020): article 8.

⁴² Iconclass, 2009, accessed April 25, 2021, <http://www.iconclass.nl/home>.

⁴³ Accessed April 25, 2021, http://www.iconclass.org/rkd/1/?q=jewish&q_s=1; accessed April 25, 2021, <http://www.iconclass.org/rkd/12A/>.

⁴⁴ Bowker, “Biodiversity Datadiversity,” 661n4, cited in Claire Waterton, “From Field to Fantasy: Classifying Nature, Constructing Europe,” *Social Studies of Science* 32 (2002): 177–204

representative enough.⁴⁵ Possible directions for complicated tasks of building classifications for the analysis of minority cultures is an exciting research area that can provide foundations for finding, sustaining, and disseminating cultures in today's world of statistical dominance and historical generalizations that privilege empires rather than local data.⁴⁶ This, however, does not solve the problem of the authoritative voices behind these classifications and the issues of applying standard classifications by curators educated within a certain school of thought or research community. Principles that will guide data models will rely on ethical norms, educational institutions, academic research, reading lists, and users to whom cataloging voices will be addressed. These agents will determine the data models and the accessibility of cultures we will see in the near future.

What is clear from this research is that the technological and data-led decisions made about cataloging, indexing, and returning searches on Jewish culture in a digital age inform what parts of Jewish culture are *accessible* in a digital age. The choices made in these platforms and data structures form online Jewish historical culture, as reflected in, supported by, and delivered from galleries, libraries, archives, and museums. The collective decisions made in these datasets, by many individuals, and differing institutions, often over long timeframes, will affect how users can find and navigate Jewish culture for years to come. We therefore suggest that the data structures which underpin content management systems are worthy of future study, comparison, and critique, when understanding how minority cultures are delivered to online users of cultural heritage.

45 L.I. Kuncheva et al., "Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale," August 28, 2020, accessed May 12, 2021, <https://arxiv.org/pdf/2008.12025.pdf>.

46 G.C. Bowker, Foreword, in *All Data Are Local: Thinking Critically in a Data-Driven Society*, ed. Yanni Alexander Loukissas (Cambridge, MA: MIT Press, 2019); Yanni Alexander Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society* (Cambridge, MA: MIT Press, 2019).

References

- Aguilera, J. "Another Word for 'Illegal Alien' at the Library of Congress: Contentious." *The New York Times*, July 22, 2016. Accessed April 25, 2021. <https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html>.
- Bagga, S. and A. Piper. "Measuring the Effect of Bias in Training Data for Literary Classification." *Proceedings of LaTeCH-CLfL 2020*, 74–84. Barcelona, Spain, December 12, 2020.
- Bechmann, A. and G.C. Bowker. "Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media." *Big Data & Society* 6, no. 1 (2019): 1–11.
- Bingham, N.J. and H. Byrne. "Archival Strategies for Contemporary Collecting in a World of Big Data: Challenges and Opportunities with Curating the UK Web Archive." *Big Data & Society*, January 2021.
- Bloom, H. *The Western Canon*. New York: Harcourt Brace, 1994.
- Blyum, A. "Hebrew Publications and the Soviet Censor in the 1920s." *East European Jewish Affairs* 23, no. 1 (1993): 91–99.
- Bode, K. "Why You Can't Model Away Bias." *Modern Language Quarterly* 81, no. 1 (2020): 95–124.
- Bowker, G.C. "Biodiversity Datadiversity." *Social Studies of Science* 30, no. 5 (2000): 643–83.
- Bowker, G.C. Foreword. In *All Data Are Local: Thinking Critically in a Data-Driven Society*, edited by Yanni Alexander Loukissas. Cambridge, MA: MIT Press, 2019.
- Bowker, G.C. and S.L. Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 1999.
- Cotter, K., M. Medeiros, C. Pak, and K. Thorson. "'Reach the Right People': The Politics of 'Interests' in Facebook's Classification System for Ad Targeting." *Big Data & Society*, January 2021.
- Daru, B.H., D.S. Park, R.B. Primack, C.G. Willis, D.S. Barrington, T.J.S. Whitfield, T.G. Seidler, et al. "Widespread Sampling Biases in Herbaria Revealed from Large-Scale Digitization." *New Phytologist* 217 (2018): 939–55.
- Economou, Maria. "Heritage in the Digital Age." In *A Companion to Heritage Studies*, edited by William Logan, Máiréad Nic Craith, and Ullrich Kockel, 215–28. Chichester, UK: Wiley, 2016.
- Eisenberg, Ellen. *Jewish Agricultural Colonies in New Jersey, 1882–1920*. Syracuse, NY: Syracuse University Press, 1995.
- Fernandez, M.V. "The Coloniality of Metadata: A Critical Data Analysis of the Archive of Early American Images at the John Carter Brown Library." PhD thesis, University of Texas, 2018.
- Frye, N. *The Great Code: The Bible and Literature*. San Diego: Harcourt Brace Jovanovich, 1982.
- Harshav, B. and M. Chagall. *Marc Chagall and His Times: A Documentary Narrative*. Palo Alto, CA: Stanford University Press, 2004.
- Hauswedell, T., J. Nyhan, M. Beals, M. Terras, and E. Bell. "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives of Historical Newspapers." *Archival Science* 20 (2020): 139–65.
- Iconclass. 2009. Accessed April 25, 2021. <http://www.iconclass.nl/home>.

- Jewish Museum, New York. "The Cairo Geniza: Jews and Muslims in the Mediterranean World 800–1500." January 1, 1997–October 12, 1997.
- Jewish Museum, New York. "Russian Jewish Artists in a Century of Change, 1890–1990." September 21, 1995–January 28, 1996.
- Kogan, L. and S. Loesov. "Old Jewish Language." In *World Languages: Semitic Languages. The Akkadian Language. Northwest Semitic Languages*, edited by A. Belova, L. Kogan, S. Loesov, and O. Romanova, 296–375. Russian Academy of Sciences. Institute of Linguistics. Moscow: Academia, 2009. [Russian]
- Kuncheva, L.I., C.E. Matthews, A. Arnaiz-Gonzalez, and J.J. Rodriguez. "Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale." August 28, 2020. Accessed May 12, 2021. <https://arxiv.org/pdf/2008.12025.pdf>.
- Likhter, J. Personal communication, May 7, 2021.
- Liu, A. "Toward a Diversity Stack: Digital Humanities and Diversity as Technical Problem." *PMLA/Publications of the Modern Language Association of America* 135, no. 1 (2020): 130–51.
- Loukissas, Yanni Alexander. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press, 2019.
- Mak, B. "Archaeology of a Digitization." *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–26.
- Metropolitan Museum of Art. "About the Met." 2021. Accessed June 25, 2021. <https://www.metmuseum.org/about-the-met>.
- Metropolitan Museum of Art. "The Colmar Treasure: A Medieval Jewish Legacy". July 22, 2019 – January 12, 2020. Accessed April 30, 2021. <https://www.metmuseum.org/exhibitions/listings/2019/colmar-treasure-medieval-jewish-legacy>.
- Metropolitan Museum of Art. "Explore the Collection." 2021. Accessed June 25, 2021. <https://www.metmuseum.org/art/collection>.
- Metropolitan Museum of Art. "Met Museum Sets New Attendance Record with More Than 7.35 Million Visitors." 2018. Accessed May 12, 2021. <https://www.metmuseum.org/press/news/2018/met-museum-sets-new-attendance-record>.
- Navarette, T. and E. Villaespesa. "Digital Heritage Consumption: The Case of the Metropolitan Museum of Art." *magazén* 1, no. 2 (December 2020).
- Pinkus, B. *The Jews of the Soviet Union. The History of a National Minority*. Cambridge: Cambridge University Press, 1988.
- Piper, A. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge: Cambridge University Press, 2020.
- Prebor, G., Maayan Zhitomirsky-Geffet, and Yitzchak Miller. "A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types." *Digital Scholarship in the Humanities* 35, no. 2 (June 2020): 441–58.
- Russian State Catalogue (The State Catalogue of Museum Collections of the Russian Federation). 2021. Accessed June 25, 2021. <https://goskatalog.ru/portal/#/>.
- Said, Edward W. *Orientalism*. New York: Random House, 1978.
- Schriml, L.M., M. Chuvochina, N. Davies, et al. "COVID-19 Pandemic Reveals the Peril of Ignoring Metadata Standards." *Scientific Data* 7 (2020): article 188.
- Shandler, J. *Adventures in Yiddishland: Postvernacular Language and Culture*. Berkeley: University of California Press, 2006.
- Shandler, J. *Yiddish: Biography of a Language*. Oxford: Oxford University Press, 2020.

- Slezkine, Yuri. *The Jewish Century*. Princeton, NJ: Princeton University Press, 2002.
- Soffer, O., Z. Segal, N. Greidinger, S. Rusinek, and V. Silber-Varod. "Computational Analysis of Historical Hebrew Newspapers: Proof of Concept." *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.
- Sterner, Beckett W., Nico M. Franz, and J. Witteveen. "Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies." *History and Philosophy of the Life Sciences* 42, no. 1 (2020): article 8.
- Thylstrup, N.B. *The Politics of Mass Digitization*. Boston: MIT Press, 2019.
- University of Wisconsin-Madison Department of Hebrew and Semitic Studies. "The Book and the Spade" (Biblical Archaeological Exhibition). April 13–May 4, 1975.
- Waterton, Claire. "From Field to Fantasy: Classifying Nature, Constructing Europe." *Social Studies of Science* 32 (2002): 177–204.
- Winer, Dov. "Judaica Europeana: An Infrastructure for Aggregating Jewish Content." *Judaica Librarianship* 18 (2014): 88–115.
- Zaagsma, G. "Digital History and the Politics of Digitization." Paper presented at *Digital Archive and Canon Workshop*, March 10, 2021. Accessed May 11, 2021. https://www.digitales-archiv-und-kanon.de/contributions/Zaagsma_en.pdf.
- Zhitomirsky-Geffet, M. and Gila Prebor. "SageBook: Toward a Cross-Generational Social Network for the Jewish Sages' Prosopography." *Digital Scholarship in the Humanities* 34, no. 3 (September 2019): 676–95.