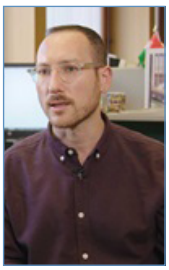
Научная статья

УДК 811.512.1:002.2:004
https://doi.org/10.20913/1815-3186-2022-1-22-28

Making Siberian Turkic Printed Collections Visible

Michael J. Erdman



Erdman Michael James,

British Library,
96 Euston Rd, London,
NW1 2DB, Great Britain,
PhD, Curator, Turkish
and Turkic Collections,

ORCID: *0000-0002-7924-9719*
e-mail: *Michael.Erdman@bl.uk*

Abstract. In the age of big data, small collections are often overlooked. Too numerically few to cause waves, the issues and problems that affect them are rarely addressed by cataloguing institutions. In the present paper, the author focuses on one such collection – holdings of Siberian Turkic materials at the British Library – to explore how homogenizing trends impact the cataloguing of these materials. The author concentrates particularly on aspects related to MARC21 requirements and authority files, and how these work to occlude the unique printing histories of Siberia's indigenous Turkic languages. The author concludes that these trends may be overcome by treating the collections qualitatively through the creation of collection guides, addressing the inconsistencies in MARC21 and Library of Congress authority files, and outreach to and engagement with contemporary Siberian Turkic communities.

Keywords: Siberian Turkic languages, catalogs, bibliographic systems

Citation: Erdman M. J. Making Siberian Turkic Printed Collections Visible. Bibliosphere. 2022. № 1. P. 22–28. https://doi.org/10.20913/1815-3186-2022-1-22-28.

Received 20.09.2021
Revised 18.10.2021
Accepted 04.02.2022

Как сделать видимыми сибирские тюркские печатные коллекции

М. З. Эрдман

Эрдман Михаил Залманович,
Британская библиотека,
96 Euston Rd, Лондон, NW1 2DB,
Великобритания,
PhD, Куратор турецких и тюркских коллекций,
ORCID: *0000-0002-7924-9719*
e-mail: *Michael.Erdman@bl.uk*

Аннотация. В эру больших данных часто упускаются из виду небольшие коллекции. Малочисленные институты, которые занимаются их созданием, редко обращаются к подобной проблематике, так что публикаций, посвященных таким коллекциям, немного. В статье речь идет о коллекции сибирских тюркоязычных материалов, которые хранятся в Британской библиотеке, и объясняется, как эффекты трендов по гармонизации (усреднению) влияют на каталогизацию этих материалов. Автор концентрируется на аспектах, связанных с требованиями системы MARC21 и авторитетных файлов, и на том, как эти аспекты работают, но не останавливается на уникальности изданий на коренных тюркских языках Сибири. Делается вывод, что эти тенденции могут быть преодолены в результате качественного обращения с коллекциями, создания справочников по коллекциям, устранения несоответствий в архивах MARC21 и Библиотеки Конгресса, а также взаимодействия с современными сибирскими тюркскими сообществами.

Ключевые слова: сибирские тюркские языки, каталоги, библиографические системы

Для цитирования: Эрдман М. З. Как сделать видимыми сибирские тюркские печатные коллекции // Библиосфера. 2022. № 1. С. 22–28. https://doi.org/10.20913/1815-3186-2022-1-22-28.

Статья поступила в редакцию 20.09.2021
Получена после доработки 18.10.2021
Принята для публикации 04.02.2022

Introduction

Large digital corpuses of metadata have become the goal of many collecting institutions around the world. Scholars from all disciplines now eagerly conduct analyses of huge volumes of data. They search, crunch, convert, and categorize massive numbers of words and figures in order to support hypotheses about broad, numerically-large societies. Yet there is still value to be found in small collections. It is not just, to paraphrase the title of Ernst Schumacher's 1973 work, that "small is beautiful." Small can also force us to stop and question content on its own terms; to refrain from quantitative, statistical analysis in favour of qualitative study. Small can be agile; small can be complex; small can be unique in ways that big data does not allow for.

In their seminal 2012 paper "Decolonization is not a Metaphor", Tuck and Yang explore, in part, how big data can often render statistics and analyses of Indigenous groups inconsequential. It buries them in footnotes and asterisks without ever attempting disaggregation or, better yet, understanding (Tuck, Yang, 2012, p. 22–23). In this paper, I will look at similar dynamics within library metadata on printed materials in Siberian Turkic lects[1]. Given my employment at the British Library (BL), the collections of such materials held by the BL will be the main thrust of the piece, but not its sole source of information. In doing so, I will attempt to untangle the ways in which categorization, cataloguing, and curation all hide the complexity of Siberian Turkic publishing cultures. Moreover, I will seek out possible solutions to these problems. To start, however, I will outline the languages and publishing histories that are at the core of this paper.

A Linguistic Overview

What are the Siberian Turkic languages? Two different answers might be offered. One is from a strictly linguistic point of view, bringing together the Northeastern Turkic languages. These are generally divided between Sakha (Yakut), the dominant North Siberian language, and four different sub-groups of South Siberian lects: Sayan (including Tuvan, the most common language of the group), Yenisei, Chulym, and Altai (Johanson, 1999, p. 83). As Schönig explores (1999b), the South Siberian subgroup exhibits considerable diversity in both linguistic characteristics and history. Each one is distinguished by particular diachronic phonetic changes, as well as 19th- and 20th-century attempts at the creation, or lack thereof, of literary languages heavily managed by the Orthodox Church or the Soviet state. Unlike groups such as Oghuz or Kipchak Turkic, these languages are brought together more by geographic and areal features than a clear, linear pattern of descent (Schönig, 1999a, p. 78–79) (Figure).

An alternative view is to include all Turkic languages spoken in Siberia. The region's exact boundaries are far from certain, as different authorities have defined it in various ways over the centuries. The name is of Tatar origin and was first used in the 16th century. The crudest means of defining the region is geographically, bounding it by the Ural Mountains (west), the Pacific Ocean (east), the Arctic Ocean (north), and the Kazakh and Mongolian Steppes (south) (Naumov, 2006, p. 3). Russian annexation of the region occurred piecemeal, which means that different parts came under different administrative structures at different times, with a unified Siberian province created only in 1708, only to be divided and reunited numerous times over the coming century and a half (Naumov, 2006, p. 98). The division is more than just administrative, as political, economic, and social differentiation has continued through the 20th and 21st centuries, making it difficult to treat the entire landmass as whole. Even when it comes to Turkology, to do so would result in considering West Siberian Tatar in this study while leaving some South Siberian language communities out. For this reason, I will stick to the linguistic definition of Siberian Turkic when exploring printed materials.

Each member of the Northeastern Turkic language grouping has a unique history of linguistic and orthographic standardization. Sakha has had a Cyrillic-based writing system created by O. N. Bötlingk in 1851 (Zakharova, 2014, p. 6), a bespoke system based on the International Phonetic Alphabet divised by Sakha intellectual Semyon Novgorodov, and the Uniform Alphabet from 1927 to the late 1930s (Stachowski, Menz, 1999, p. 421). The literary language, by contrast, was produced through a combination of state intervention and the efforts of prominent authors in the early part of the 20th century, many of whom wrote in the dialect associated with the region around Yakutsk/-D'okuuskai (Ferguson, 2020, p. 139).

While Tuvan, by contrast, did not have an indigenous writing system prior to 20th century, the Old Mongolian and Tibetan languages and scripts were widespread among feudal and religious elites (Bicheldei, 2010, p. 213–214). Although not formally part of the Soviet Union until 1944, the Tuvan People's Republic adopted a Latin alphabet based on the Uniform Alphabet in 1930, to be replaced, in 1941 and again in 1943, by the Cyrillic version that is in use today (Saaia, Badarch, 2018, p. 139–140).

For the remaining lects and ethnic groups, the history of literacy and national languages is both shorter and faster-paced. Missionaries used the Cyrillic alphabet for Shor publications in the 19th

[1] Throughout the paper, the author makes use of the "lect", a generic identifier for a linguistic system of communication that avoids the thorny issue of deciding whether an object is a language or a dialect. Both are lects, and the division between the two is largely irrelevant to the discussion that follows.

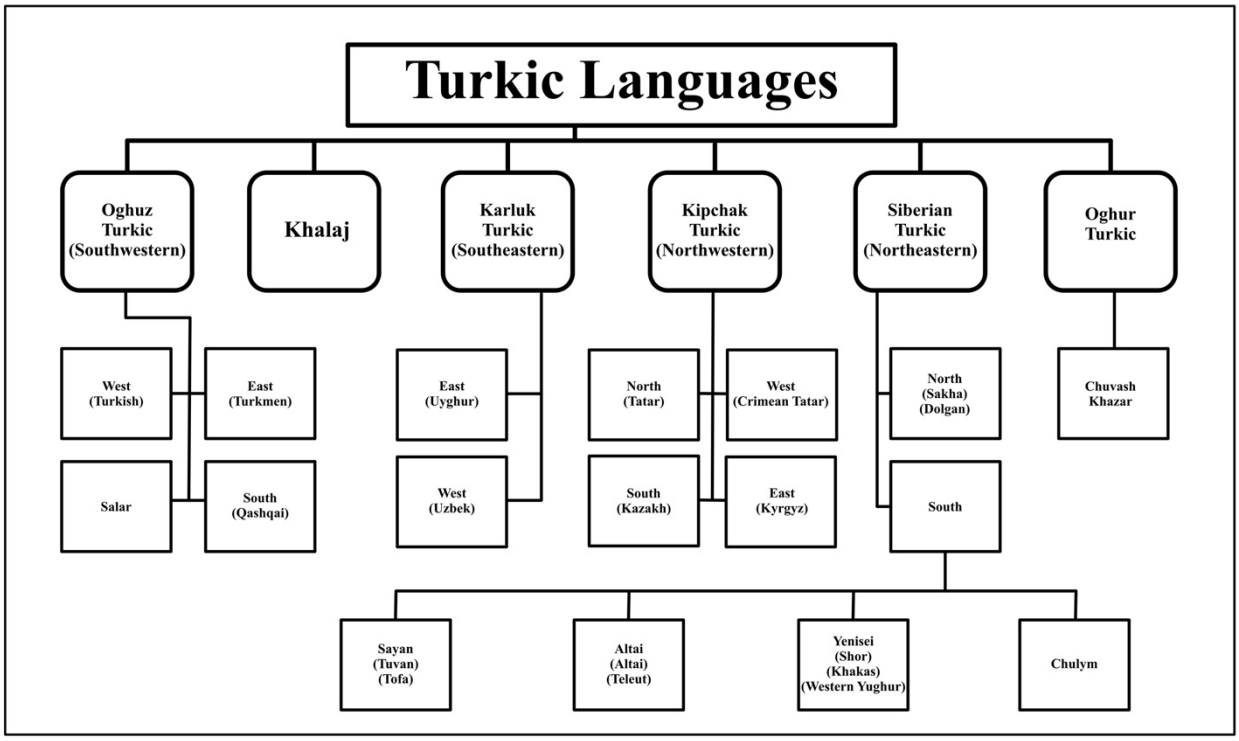


Fig. Genetic relationships between the Turkic languages
Рис. Генетические связи между тюркскими языками

century, and small groups of Shor students were schooled in the Orthodox Church pedagogical tradition, as well as in written Teleut, in the 19th century (Esipova, 2017, p. 65). It was in the 1920s that the Altai (Oirots), Khakas, and Shor all saw the creation of Latin-based orthographies, national literary standards, and state-sanctioned national identities (Borina, 2015, p. 98). The decision to create three distinct national languages was eventually adopted at the 1931 All-Siberia Conference of Turco-Tatar Peoples. Suitable dialects were selected and the preparation of Latin-script literature was undertaken throughout the 1930s, only to have the Latin alphabet dropped in favour of the Cyrillic one in 1938-39 (Borina, 2015, p. 99).

A Brief History of Siberian Turkic Publishing

For Siberian Turkic-speaking communities in which Islam was common, either one of the literary lects subsumed under the umbrella of Chagatai, or Persian dominated literary expression. Arabic was often the language of jurisprudence and ritual practice in the pre-Soviet period (DeWeese, 2019, p. 132–133). Original production in Turkic lects, including both Chagatai and other languages, became an increasingly crucial component of book culture among Tatar communities as well as other linguistic groups in the 18th and 19th centuries (DeWeese, 2019, p. 147–148). This was altered radically by the events of 1917.

Publishing did occur in some Siberian Turkic languages prior to the February and October Revolutions; these were largely to materials dedicated to proselytization, linguistics, or ethnology. Published Sakha translations of Orthodox creed go back to as early as 1812 (Rufova, Sivtceva-Maksimova, 2020, p. 352), albeit in a script identified as "Russian" rather than Sakha (Zakharova, 2014, p. 5). Similarly, publishing was undertaken in Tuva in old Mongolian up until the 1920s, with a particular focus on Buddhist tracts (Bicheldei, 2010, p. 214). The Bolshevik seizure of power in 1917, and subsequent Civil War, changed that. To start, communities with an established written tradition witnessed the transformation of their publishing realm into an ideological and propagandistic weapon on behalf of the Bolsheviks (see, for example, Esipova, 2017, p. 65). The Tatar press was among the most active in this respect, owing to the community's burgeoning pre-Revolutionary publishing industry. While non-Bolshevik elements did manage to operate within this structure for a while, these individuals were eventually ousted, replaced by more trustworthy apparatchiks, even if not of the titular nationality (Lenkart, 2017, p. 659–660).

Most of the other Siberian Turkic languages, apart from Sakha and Tuvan, did not have considerable publishing histories prior to the 1920s. These communities witnessed the birth of mass publication under the Soviet, or more precisely, Stalinist regime. The creation of the various Turkic alphabets and spread of printed materials went hand-in-hand with literacy campaigns. Many items produced during this period were pedagogical in nature, as well as propagandistic (see Bicheldei, 2010; Saaia, Badarch, 2018).

But the process of creating appropriate alphabets and literary standards meant that publications in languages such as Shor, Altai, and Khakas during this period were relatively few; not necessarily targeted at mass consumption; and often translations from Russian (Borina, 2015, p. 99; Esipova, 2017, p. 65). It is only after 1945, when the Cyrillic alphabet was firmly entrenched, that widespread publication of literary and non-fiction works really commenced. Even this, however, was uneven across the various linguistic groups, especially after the suppression of Shor-language instruction between 1943 (five years after the dissolution of the Gorno-Shor National Region) and 1989 (Esipova, 2017, p. 66). Others, such as Sakha and Tuvan, continued to be the official languages of semi-autonomous regions, and thus languages of education, literature, and statecraft.

Accounting for Siberian Turkic Works

Works of such rarity are highly prized by Euro-Atlantic institutions. But how are they reflected in the catalogues? As with the rest of its contemporary printed holdings, the British Library imposes MARC21 guidelines on the creation of metadata about them, and requires staff to use Library of Congress (LOC) authorities for subjects and authors. Much the same is found in other Anglophone institutions.

When processing texts in Siberian Turkic languages according to these guidelines, however, various issues arise. One of four language codes can be applied: SAH for Sakha (known as "Yakut"); TYV for Tuvan (known as "Tuvinian"); ALT for Gorno-Altai (also called "Southern Altai" or, before 1948, Oirot); and TUT, a generic "Altaic Other" code for all other languages. Siberian Tatar is put together with the larger Kazan Tatar group under TAT2. The TUT code includes languages such as Shor, Khakas, Salar, Teleut, and Western Yugur, but it also covers Western Turkic lects such as Gagauz and Karaim; historic Old Turkic; and non-Turkic languages such as Evenki, Khitan, and Oroqen. Discoverability is therefore a complicated affair.

Sakha, Tuvan and Altai books, by virtue of their separate language codes, are easy to identify in the catalogue. When searching in the BL's cataloguing software (known as ALEPH), we pull up 22 hits in Altai, 43 records with the TYV code, and 76 with SAH. Some of these items are non-Turkic works that contain extracts in Tuvan, Sakha or Altai, whether as sample texts for linguistic or literary studies, or as broader items with one or two chapters in the language. Some are translations into Russian or other languages that have their original language marked with a language code. Similarly, there are undoubtedly some Altai, Tuvan or Sakha titles in the collections that have catalogue records lacking the appropriate language codes, or that do not have electronic records at all. Nonetheless, we still gain some insight into the relative sizes of the collections, their timeframes, and their subject weightings through a language code search. The same exercise conducted on the front-end catalogue (the one seen and used by researchers, known as Explore), yields much more mixed results, as users cannot search by language codes. If a user searches for the word "Yakut", they come up with only 12 results identified as being in the Sakha language. Similarly, a search of the word "Tuvinian" produces only 16 results in the language, implying that the back-end catalogue (the one used by Library staff and cataloguers, known as ALEPH) is more reliable for complete data about the size of the collections.

The situation becomes much more complex when we come to the TUT group. As we cannot search for language codes on Explore, I have used the term "Altaic" instead. It produces a total of 8 hits tagged with the language code TUT (Altaic (Other)), a paltry number for a catch-all "other" category. Clearly, Explore is abysmal in connecting users with materials produced in smaller linguistic communities. The intervention of someone else with access to ALEPH is necessary. But this is where things go off the rails. The catalogue contains some 560 items that are tagged with the TUT language code. The discrepancies between Explore and ALEPH speak volumes to the impotence of Explore, but they also expose a far deeper problem with the miscellaneous language code, and how smaller language groupings are handled in computerized cataloguing.

Of the records that bear the TUT coding, very few relate to Siberian Turkic languages. A rough count of the records shows that 14 works contained text in Khakas, four each in Shor and Teleut, and two in Dolgan. TUT is largely composed of works in languages with relatively prominent publishing histories, such as Gagauz (70 items). The same can be said of scholarly works on now-dormant Turkic languages such as Gök Turkic, Khwarezmian and old Uyghur (29 in total). How, then, is a user expected to find details on the small collections held in these languages, when language coding itself provides an unstable at best, non-serviceable at worst, mechanism for isolating them?

Decoding the Catalogue

Catalogues, obviously, are built on more than just language codes and page counts. Cataloguers at the British Library, and at most other institutions around the world, are also expected to provide information on authors and subjects, often based on the Library of Congress's authority files. The LOC's authorities provide good coverage for a wide variety of authors operating in the Soviet and post-Soviet spaces, as well as subjects relating to the Turkic peoples of the Russian Federation. The strength

[2] *MARC Code List for Languages: Part 1: Name Sequence.* URL: https://www.loc.gov/marc/languages/language_name.html (accessed 01.04.2021).

of Sovietology in the United States during the Cold War, and the continuation of Eurasian Studies after 1991, has ensured healthy representation in the pantheon of LOC authorities. A number of university libraries are actively engaged in this effort, especially the University of Illinois at Urbana; the University of Chicago; Columbia University; University of Wisconsin at Madison; and the University of California at Berkeley; as well as staff at the LOC itself. Further emphasis is given thanks to the expansion of agencies providing materials from these regions, such as MIPP and EastView[3], the latter of which creates metadata on new publications that appears in the Online Computer Library Center (OCLC)'s union catalogue, WorldCat.

Some institutions have sought to address the newly-independent Turkic republics of the Caucasus and Central Asia as separate spheres of study, linked both to the regions south and north of them. But Turkic librarianship and area studies in the US (excluding Ottoman and Turkish Studies) remain firmly linked to the Russian, or Slavonic, sphere. Turkic Studies continue to be dealt with in the realm of Eurasian Studies, as evidenced in the Association for Slavonic, East European and Eurasian Studies and SLAVLIB[4]. This is at odds with the British tradition of bundling Caucasian, Central Asian and Siberian Turkic librarianship together with Middle Eastern librarianship. This establishes linguistic and religious ties as paramount, rather than political and economic ones, and mirrors Oriental Studies departments at various universities.

What does this mean for authorities? As these rely heavily on American institutional input, they are generally skewed towards a Russian or Russo-centric bias. Where authors have published, or been published, in Russian as well as their native languages, Russian names are often those that enter the authority database. This is not an unusual course of events for minority, regional, non-official, or suppressed languages (see Hughes, 2019). Authors, editors, compilers, and others who work in Siberian Turkic languages thereby appear first and foremost according to their Russian names. In some cases, the distinction between Russian and Turkic monikers is minimal or non-existent, but there are many cases of non-negligible divergence. This is becoming more apparent as many individuals in Turkophone communities opt to use indigenous patronymics, group names, and naming patterns, alongside or instead of their official Russian patronymics and surnames (see Harrison, 1999; Madieva, Tayeva 2014, p. 4799–4800). While Turkic names become more common in practice, it is their Russian counterparts which continue to dominate in the authority fields.

Why does this matter? For one, just based on personal experience and anecdotal evidence, the number of cataloguers and curators comfortable with Russian far exceeds those working with Turkic languages. There is a tendency, then, for metadata that is easily available through systems such as OCLC to be skewed towards the more familiar Russian version than its Turkic (or, indeed, Mongolic, Tungusic, or Caucasian) forms. The trend is changing, but only very slowly. MIPP, for example, continues to employ Russian titles and names on its website for works entirely in Siberian Turkic languages.

EastView, perhaps the largest provider of Turkic materials from the former Soviet Union to Anglo-American libraries, does include more original-language metadata. This, however, is neither uniform nor entirely reliable. Original-language names will be provided in transliteration, but these are not always accompanied by original-script entries, especially when such information would require the use of characters not found on a keyboard provided in the standard Microsoft Office packages. The result is often records with titles in transliterated versions of the original language as well as their Russian equivalents, alongside the author's name in a form that doesn't necessarily accord with LOC authorities or the original spelling. The ability, then, for a speaker of a Siberian Turkic language to access such works according to their own usage of their native language is impeded. So too is the aggregation of works according to titles and authors. This makes it exceptionally hard to determine just how widespread Siberian Turkic materials are in Anglo-American collections, and where they can be found.

Subject authorities should mitigate this, at least in theory. The titles might not reflect what is on the book, and the names don't always match Indigenous orthographies or LOC authorities. But simple tags such as "Shor literature" or "Yakut poetry" should help us get around such blockages. In a way, they do, especially with respect to literary pieces. Even when translated into Russian or other languages, poetry and prose can still be linked back towards a broader corpus. Problems arise, however, when dealing with non-fiction items, where categories such "Yakut politicians" or "Shor cooking" are non-existent within the LOC's subject authorities database. The cataloguer then needs to construct complex subject categories which may or may not be picked up

[3] MIPP International, whose name does not appear to be an acronym, and EastView are both distributors, connecting libraries and other consumers with publishers and in-country distributors of printed and other materials from the former Soviet Union, Eastern Europe, and other regions. MIPP (http://www.mippbooks.com) is currently based in Vilnius, Lithuania and was founded in 1991. EastView (https://www.eastview.com), founded in 1989, is based in Minneapolis, USA.

[4] The Association for Slavonic, East European and Eurasian Studies (ASEEES; https://www.aseees.org) was founded in 1948 in the United States and currently boasts more than 3000 members worldwide. SLAVLIB (https://mailman.yale.edu/mailman/listinfo/slavlib) is a listserv domiciled on the Yale University servers. It allows the sharing of professional information, requests for assistance, and best practices among librarians, archivists, scholars and others who have an interest in collections relating to Slavic Studies, Eastern Europe and the former Soviet Union.

by Boolean searches, and which may or may not cause considerable mixing of Indigenous and foreign language collections. This is, in effect, the "dispersal" of Indigenous collections that Sutherland and Purcell warn about (Purcell, Sutherland, 2021). A holistic approach to Indigenous cultural production is no longer possible, and neither is a recategorization of Indigenous production according to Indigenous taxonomies or epistemologies.

Of course, dispersal is only a concern provided that subjects are listed in the first place. As many works in Siberian Turkic languages are catalogued by individuals without appropriate linguistic and cultural knowledge, subjects are often only added when a Russian synopsis of the work is available. Alternatively, all works are catalogued as "literature" or "fiction" by some cataloguers, greatly skewing the perceived composition of these collections. Whatever the case may be, it should now be apparent that comprehensive cataloguing of these items is a long way off.

Reimagining the Archive

To what end do we expend such energy? Speakers of the Siberian Turkic languages are few and far between in Western Europe and the Americas. Few scholars in the Anglo-American sphere engage with these linguistic communities beyond their linguistic characteristics. Sakha, Tuvan, Shor, Western Yughur and Chulym sources are rarely employed in disciplines other than linguistics. There is a pressing need for adequate cataloguing and curation of such materials in the regions in which the languages are spoken; in large diasporic centres such as Moscow or St. Petersburg; and in scholarly nodes such as Istanbul or Ankara (Yılmaz, 2018). But the urgency appears to wane outside of the former Soviet Union and Turkey. The metadata created regarding collections of Siberian Turkic materials in, for example, the BL is likely only to be used by highly-specialized scholarly or academic researchers physically present in the Library's vicinity. Matching it to broader categorizations of language collections, or orthographic and linguistic norms, does not necessarily enhance the experiences of the Library's users as a whole.

Such calculations, however, create a false analogy between the pre-Internet world of card and printed catalogues, and contemporary online databases and open access metadata. Some Siberian Turkic languages are used heavily in digital media and online forums. Sakha in particular (Ferguson, 2020, p. 134), but also Tuvan and other languages, flourish in new virtual spaces as intracommunity and intergenerational mechanisms of communication (Ferguson, 2020, p. 137–138). The presence of online data regarding bibliographic materials in the Siberian Turkic languages helps to build a broader space for these linguistic communities. It encourages the spread of publication data and histories not always available from local archival institutions or libraries. And, importantly, it strengthens the hand of those seeking to grow IT tools for using these languages online. Keyboards for Sakha, Tuvan, Altai, and many other languages are not readily available to either PC or Mac users, despite the existence of Unicode characters for the alphabets employed by each of these communities. Whether as part of a *Microsoft* or *Apple* product, or as third-party and open source software, the routinization of typing original script entries with bespoke Siberian Turkic keyboards will help ensure the wider appearance of letters particular to the languages in library catalogues, and the authorities that flow from them. Although small, adding cataloguer and curatorial demand to the voices of the language users can only help to further the cause of IT for all.

Beyond practical considerations, there are also issues of equity and social justice. To come back to Tuck and Yang's point, enhanced visibility creates new possibilities of understanding collections, publishing histories, and the value of language for communities around the world. Appropriate cataloguing is an initial step in this direction, one that needs to be complemented by contextualized, qualitative descriptions of the works held. A catalogue that reflects the orthographic systems and names of Sakha, Shor, Altai, Tuvan or others helps reverse the tendency to make such minority writers invisible within the broader sphere of Soviet or Russian production. An initial step that I have taken, therefore, is to correct and enhance as much metadata as possible, including the usage of original-script cataloguing. But in a Library such as the British Library, with over 170 million items, a collection of 100 or 200 books is unlikely to be noticed in aggregate, machine-based analyses of holdings. To paraphrase Lenin, what is to be done?

Although not particularly popular in the age of big data, my proposition is the creation of interactive and linked collection guides: documents intended to provide context as well as direct users towards individual items. While not perfect finding tools, they have, and do, create an awareness about access and history. The presence of a hub, including explanations of languages and histories; lists of notable works; examples of text, audio, and video; and links to other institutions, collections, and works creates a new landing site for those interested in Siberian Turkic collections and cultural production. It takes the onus away from the catalogue to perform this function, and recognizes that visibility comes through display and demonstration, rather than data alone. It also acts as a platform for hosting and disseminating multilingual content, shifting the languages and cultures from objects of study to means of communication and connection.

In recent years, blog posts, whether hosted by BL or off-site, have been especially helpful in this

regard. The pieces are hyperlinked to catalogue records, ensuring that the catalogue continues to receive traffic, and that the metadata on these works is still used. But the surrounding text, occasionally linked to works by scholars – especially those from the linguistic communities discussed by the guides – or the catalogues of other institutions, also provides valuable context about the history of the publishing culture and of the languages. Most importantly, such text is easily searchable using Google, Yandex or other search engines, thus proving to be much more powerful in attracting potential or real users to the books and periodicals in question.

Conclusion

The tumultuous history of writing and publishing in the Siberian Turkic languages is relatively short. It is, nonetheless, of great value to the communities in which these languages are spoken, written, and read. As collectors and cataloguers, we have an obligation to rectify the issues that render these complex and rich histories invisible within catalogues and big data. The process I've proposed is only in its beginning stages. It is my hope that, in several years, when more metadata is available, and more contextualizing pieces online, we will begin to see the fruits of these efforts. We will have proven that small is not only beautiful, but visible as well.

References

Bicheldei KA (2010) 80 years of Tuvan writing: formation, development, prospects. *Novye issledovaniya Tuvy* 4: 210–229. (In Russ.).

Borina LS (2015) The graphics reform in the USSR in the 1920s-1930s as a factor to form ethnic consciousness of the Turkic-speaking ethnic groups of the Sayano-Altai. *Vestnik Tomskogo gosudarstvennogo Universiteta* 395: 97–101. (In Russ.).

DeWeese D (2019) Persian and Turkic from Kazan to Tobolsk: literary frontiers in Muslim Inner Asia. *The Persianate world: the frontiers of a Eurasian Lingua Franca.* Berkeley: Univ. of California Press, pp. 131–155.

Esipova AV (2017) National education and Shor writing. *Rossiiskaya tyurkologiya* 1/2: 64–73. (In Russ.).

Ferguson J (2020) Don't write it with "h"! Standardization, responsibility and territorialization when writing Sakha online. *Responsibility and language practices in place.* Helsinki: Fin. Lit. Soc., pp. 131–151.

Harrison KD (1999) Name practices and ethnic identity in Tuva. *Proceedings of the Chicago Linguistics Society* 35 (2): 69–79.

Hughes HK (2019) Cataloguing Kurdistan: imagining liberated geographies. *Ethical questions in name authority control.* Sacramento: Libr. Juice Press, pp. 99–107.

Johanson L (1999) The history of Turkic. *The Turkic languages.* London: Routledge, pp. 81–125.

Lenkart J (2017) Russian revolutions in print. *Slavic Review* 76 (3): 655–663.

Madieva G and Tayeva RM (2014) The evolution of the Kazakh surnames: 'for' and 'against'. *Procedia: Social and Behavioural Sciences* 116: 4798–4802.

Naumov IV (2006). The history of Siberia. London: Routledge.

Purcell A and Sutherland T (2021) A weapon and a tool: decolonizing description and embracing redescription as a liberatory archival practice. *The International Journal of Information, Diversity & Inclusion* 5 (1): 60–78. DOI: 10.33137/ijidi.v5i1.34669.

Rufova ES and Sivtceva–Maksimova PV (2020) Historiographic analysis of the role of Christian culture in the formation of the Yakut writing system. *Tarih Kültür ve Sanat Araştırmaları Dergisi* 9 (2): 346–354.

Saaia OM and Badarch B (2018) Tuvan orthography in the period of the Tuvan People 's Republic. *Novye issledovaniya Tuvy* 3: 138–150. (In Russ.).

Schönig C (1999a) The internal division of modern Turkic and its historical implications. *Acta Orientalia Academiae Scientiarum Hungaricae* 52 (1): 63–95.

Schönig C (1999b) South Siberian Turkic. *The Turkic languages.* London: Routledge, pp. 403–416.

Stachowski M and Menz A (1999) Yakut. *The Turkic languages.* London: Routledge, pp. 417–433.

Tuck E and Yang WK (2012) Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society* 1 (1): 1–40.

Yılmaz KI (2018) Study and teaching of the Turkic languages of the Siberian group at Ankara University. *Novye issledovaniya Tuvy* 1: 122–142. (In Russ.).

Zakharova TV (2014) Ehe history of the Yakut alphabet. *Severo–Vostochnyi gumanitarnyi vestnik* 1: 5–11. (In Russ.).