NEH Application Number: HD-248405-16
“Historical Demography and Population Behavior among Muslims in Russian Central Eurasia, 1828-1918: The Case of Kazan’ City.”
Institution: Trustees of Indiana University
PI: Edward J. Lazzerini

“Raising the Dead” by Employing the Big Data in Imperial Russian Confessional Registers

A White Paper
Table of Contents

Abstract

The Project: A Synoptic Guide

Appendices

Metrical Book Field Analysis

Table Listing Categories of Data Sought by Muslim Metrical Books

Map of Old and New Tatar Suburbs, Kazan City

Project Team

Bibliography

Painting of the Old Tatar Suburb in Kazan City along the Eastern Shore of Lake Qaban
Abstract

Unique in conception and implementation, this project seeks to combine qualitative and quantitative scholarship in a comprehensive effort to achieve unprecedented results for its cultural and historical area of attention—Central Eurasia as encompassed by the Russian Empire down to 1917. The project will gradually create, first of all, a unique digital research resource based on virtually untapped records in the confessional registers, or metrical books (метрические книги),1 from which wide-ranging scholarship will result about the substantial Turkic-speaking Muslim communities of the Empire. Secondly, it will address a consequential scholarly question: During the final ninety years of the Empire, when state and society were experiencing the growing pressures of early stage modernization as well as repeated disruptions brought about by war, revolution, and popular unrest, how did its Muslim population fare demographically, socially, and economically? Thirdly, the records will be accessible to all interested in tracing ancestry within the specific population group.

Support from the National Endowment for the Humanities, through its Digital Humanities Start-up Program, sustained the first phase of the project whereby we created and programed the database to store non-aggregated data on persons, families, and households. Seeking proof-of-concept for the project, we gathered 56,751 discrete pieces of data relating to births, marriages, divorces, and deaths for Muslim inhabitants of Kazan, the third largest city in Russia, as organized around seven mosques and their congregations (mahalles). By adhering to the Intermediate Data Structure (IDS) format, we will take advantage of its open, scalable, and extendable

1 The image above is of a page from a metrical book.
character while proceeding beyond the bounds of our particular dataset to contribute to the common interface for the history of the life course of Eurasia as a whole.

This project embraces complex statistical and networking analysis of hitherto unexamined longitudinal data from the metrical books, in conjunction with narrative and other data-intensive sources, so as to open up unique opportunities to make Imperial Russian demographic and social history take center stage, beginning with its Muslim populations. More than usual for the field of Russian studies, our open-source project is about large numbers of humans in a time and place not our own, but whose very identity in multiple ways is revealed as if they were “raised from the dead” to be spoken to directly and individually. Because of the nature of the data in the primary source, we are already able to follow the life paths of tens of thousands of individuals, and with time potentially several million.

The more common demographic information about such issues as fertility, causes of death, infant mortality, number of divorces, and identities of local elites is especially valuable in a period of increasing social, economic, political, and familial stress that characterized Russia as it headed into the early twentieth century. Deeper revelations about our subjects will result from our team’s “data mining” ability to draw more sophisticated conclusions about (1) social and economic networking beyond the extended families and mahalles, linking identifiable persons across generations, social classes, and professions; (2) the level of economic well-being in both families and parishes, whereby we will be able to assess the success of Islamic society in Russia as compared to its past and to other confessional groups, discern whether Islamic society was able to deliver high levels of sustainable socio-economic circumstances as a result of Russia’s pursuit of industrialization, and whether the quality of life of the group or sub-groups was measurably positive; (3) the “value” of marriage over time in terms of dowries provided by the brides’ families as well as the implications of a growing divorce rate—how that affected familial wealth and opportunity for advancement, the status of women, and family structure—and how the causes for divorce might have changed and how women increasingly launched these processes; and (4) mobility, both spatial and social that will reveal the shifting contexts of networking, the level of economic well-being, and the value and stability of marriage.

The Project: A Synoptic Guide

By decree of the Holy Synod in 1722, the Russian Empire began to require Russian Orthodox priests to compile annual records of births, marriages, and deaths—but not divorces, because they were not theologically recognized—for their parishes. Implementation, however, was slow and limited, requiring additional decrees in 1779, 1802, 1812, and 1824 to institutionalize the practice. Even then,
carelessness and failure to comply remained widespread, prompting further decrees in 1886, 1889, 1890, and 1903.

At various times during the later eighteenth and nineteenth centuries, the same requirement was applied to Evangelical-Lutherans (1764), Roman Catholics (1826), Talmudic and Karaim Jews (1835), Russian Schismatics (1874, under the authority of local police, as were pagans), Baptists (1879), and Old Believers and Sectarians (1905). Muslims, the subject of this major research project, witnessed their imams adding this obligation to their duties between 1828 and 1832, and subsequently in 1872 in the Transcaucasus region.

Initially, how records were kept was left to the respective confessional authorities, but in 1838 the process was regularized, and thenceforth the government produced printed and bound registers (метрические книги, often described in English as “metrical books”) on an annual basis appropriate to each confession. The annual "book" was divided into the three or four sections—divorce was added where recognized—for the categories of data that authorities desired collected. Over the length of its history, this practice produced a massive amount of social information not found elsewhere; much of it has survived in archives of the Russian Federation (national and regional) and those of now independent states once territorially part of the empire. Although remarkably significant, practically none of the data—save from an occasional Orthodox Church—have been examined or subjected to organization and analysis.

---

2 The photo above is of a number of bound metrical books.
History is filled with those who have lived and died; statistically, virtually all humans have disappeared without any trace, their names forever lost. Without names, so the Italian scholar and novelist, Umberto Eco, frequently reminds us, nothing can have meaning. At the simplest level, it helps to know, for example, whom the great republican orator and consul, Marcus Tullius Cicero (106-43 b.c.), had in mind when he launched into the first of four stunning orations before the Roman Senate in 63 b.c. with the words “Quo usque tandem abutere, Catalina, patientia nostra?” (“How long, O Catalina, will you try our patience?”) So, also, it helps to know to whom the great Latin epic poem by Vergil was addressed, when it opens with “Arma virumque cano, Troiae qui primus ab oris Italiam.” (“I sing of arms and of the man, who first from the shores of Troy came to Italy”) Naming and knowing are such integral parts of memory and history that biography is one of the earliest genres of historical writing and remains the most popular form of recollection by amateur and professional alike. That historians are tempted to limit their horizons to the attractions of great men, and occasionally women, merely helps make the point that the temptations to treat history as the stories of “great men”—whether as monarchs or priests, warriors or inventors, saints or sinners, lovers or clowns—are seductive and, dare I suggest, may be at least partially genetically coded in our species. Proof can be seen in the simple word cloud below of the 25 most famous people in history as selected and ranked by one electronic source. The least among these has had more than 500,000 books written about them.

On occasion, and certainly more so in the past half century, some historians have resisted these temptations by focusing on the less dramatic, even mundane patterns of human communal life to produce accounts that offer the usually nameless some recognition. The more than ample number of books examining colonial New England towns, the outpouring of French studies following March Bloch, Lucien Febvre, Fernand Braudel, and the many other
advocates of the Annales school, and even German attempts spearheaded by Reinhard Koselleck to apply the methods of Begriffsgeschichte (history of concepts) to unravelling conceptual meanings at popular and elite levels, have all contributed to a paradigmatic challenge to “normal history” and to lifting, a bit, the veils that otherwise hide the faces of the past. So, too, have studies that draw upon various vital statistics and related evidence from confessional records and cadasters (a classic example being David Herlihy’s Tuscans and their Families that excavates the Florentine cadastre of 1427).

Of greater significance to the potential for illuminating long-term trends among large populations are the number of team-driven projects that have been amassing huge data-sets for major parts of the Eurasian continent. Most are territorially limited (e.g., to eastern Belgium, northern Italy, southern Sweden, northeastern Japan, or northeast China). One, the so-titled “Eurasian Population and Family History Project,” however, involves the work of some twenty scholars from a variety of countries and disciplines seeking to analyze roughly two million longitudinal individual-level records largely from the eighteenth and nineteenth centuries to make explicit comparisons of rural populations at the extreme eastern and western ends of the vast land mass. [Notice that Central Eurasia is not included in this effort.]

The comparisons are primarily of patterns of demographic responses to economic conditions in a variety of contexts, offering a challenge to the Malthusian paradigm’s application to regions outside as well as inside of Europe.

---

3 The Annales school gives importance to both the longue durée (long-term historical structures rather than events) and the apparently ephemeral lives of marginal peoples—slaves, peasants, serfs, the urban poor—to reveal their contributions to the wealth and power of their respective masters and societies.

4 Cadasters are registers of property showing the extent, value, and ownership of land for taxation.

5 Longitudinal: (of research or data) involving information about an individual or group gathered over a long period of time.

6 The basic principle of the Malthusian paradigm is that the race between population growth and development of resources, especially food supply, is highly unequal, and that poverty and stagnation are the inevitable outcome. That being the case, population growth must be curbed, but only by the moral restraint implied in postponement of marriage. In recent decades, this thesis has been severely criticized as untenable for three reasons: (1) it is erroneous because science and technology have nullified the argument that overpopulation is the cause of poverty; (2) it is inadequate for ignoring variables other than the economic; and (3) it is dangerous for treating symptoms rather than the disease.
My intention to this point is not to dismiss some historical methodologies as valueless or less valuable than others, but to reflect a bit on the limitations of all methodologies and to appreciate, nevertheless, how these limitations may be accounted for by different approaches and may still add to the accumulated riches of discovery. If we juxtapose the contradictory approaches, say, of the “great man” theory and the *longue durée* as they affect the possibility of knowing who lived in the past and how, then we have methodologies as much at odds to historians as Einstein’s theory of relativity and Bohr’s Quantum Mechanics are at odds to physicists. In both cases, the two approaches are so antithetical as to require, it would seem, dismissal of the other in the arena; in both cases, the antithesis is rooted in a desire to see all or to see little at all, to see the huge or numerically huge or to see the tiny or numerically miniscule. Like King Arthur’s knights in search of the Holy Grail, physicists lust after a unified theory of everything that will contain the extremes by filling in everything in between. Like Arthur’s knights, historians would feel vindicated if they could see more than just the faces of the few mighty, famous, or infamous, but not so many as to end up with a mob that emotionally as well as physically becomes faceless. Where, then, are the faces from the past?

If we turn our attention to Turkic Central Eurasia, those parts, at least, that fell under the dominion of the Russian Empire between the mid-sixteenth and early twentieth centuries, we may have an answer. We certainly have a range of sources that separately can reveal some of the features of some of the faces we seek; most of these sources are Russian and the product of a large and expanding state, but a smaller number, growing with time, are indigenous to Turkic peoples, who are also Islamic by confession, and include:

<table>
<thead>
<tr>
<th>Russian Sources</th>
<th>Turkic Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Метрические книги (Metrical Books)</td>
<td>Grave stones and mausolea</td>
</tr>
<tr>
<td>Дворянские родословные книги (Noble Genealogical Books)</td>
<td>Stelae</td>
</tr>
<tr>
<td>Памятные книжки (Memorial Books)</td>
<td>Tamgi (clan or tribal symbols)</td>
</tr>
<tr>
<td>Межевые книги (Survey Books)</td>
<td>Şecereler (genealogies)</td>
</tr>
<tr>
<td>Russian Sources</td>
<td>Turkic Sources</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Формулярные списки (Official Lists)</td>
<td>Awqaf contracts (inalienable religious endowments)</td>
</tr>
<tr>
<td>Документы на учащихся (Students Records)</td>
<td></td>
</tr>
<tr>
<td>Алфавитные списки студентов (Alphabetic Lists of Students)</td>
<td>Tarihler (chronicles)</td>
</tr>
<tr>
<td>Куровые ведомости (Church Rolls)</td>
<td>Bio-Bibliographical Compendia</td>
</tr>
<tr>
<td>Послужные списки монашествующих (Monastic Service Records)</td>
<td></td>
</tr>
<tr>
<td>Ведомости об иностранцах, принявших русское подданство (Rolls of Foreigners Who have Accepted Russian citizenship)</td>
<td></td>
</tr>
<tr>
<td>Обывательские книги (Registers of Urban Dwellers)</td>
<td></td>
</tr>
<tr>
<td>Алфавит жителям города на год (Annual Register of Urban Dwellers)</td>
<td></td>
</tr>
<tr>
<td>Адресные книги (Address Books)</td>
<td></td>
</tr>
<tr>
<td>Домовые книги (Dwelling Registers)</td>
<td></td>
</tr>
<tr>
<td>Исповедные ведомости (Confessional Rolls)</td>
<td></td>
</tr>
<tr>
<td>Брачные обыски (Marriage Registers)</td>
<td></td>
</tr>
<tr>
<td>Венчальные памяти (Marriage Registers)</td>
<td></td>
</tr>
<tr>
<td>Кладбищенские книги (Burial Records)</td>
<td></td>
</tr>
<tr>
<td>Ревизские сказки (Revision Lists)</td>
<td></td>
</tr>
<tr>
<td>Переписи населения (Population Censuses)</td>
<td></td>
</tr>
<tr>
<td>Посемейные списки (Family Lists)</td>
<td></td>
</tr>
<tr>
<td>Книги для записывания раскольников (Books for Recording Old Believers)</td>
<td></td>
</tr>
<tr>
<td>Описи дворовым людям (Lists of Estate Dwellers)</td>
<td></td>
</tr>
<tr>
<td>Свидетельство об отпуске на волю крепостных людей (Testimony of Voluntary Absense of Serfs)</td>
<td></td>
</tr>
<tr>
<td>Книги о разделе крестьян (Records of Allotment of Peasants)</td>
<td></td>
</tr>
<tr>
<td>Паспорта (Passports)</td>
<td></td>
</tr>
</tbody>
</table>
These serve to provide information about the once living, but typically from among the elite, even when they are locally produced for local consumption. Historians or sociologists have spent little time approaching these for the information they contain.

Increasingly from the early eighteenth century, administrators ruling the expanding multi-ethnic and multi-confessional Russian Empire sought ever more effective ways to manage their tasks. Cadastral records (писцовые книги), especially of the Middle Volga region, and genealogical books (родословные книги), primarily used to ensure social status from and in government service, were some of the earliest such vehicles for recording information about people, Russian and not, but they too inherently brought attention to those of means and status. The late imperial practice of publishing annual “memorial books” (памятные книжки, somewhat like the modern telephone book, but without the telephone numbers) in many provinces added one more source of such information. Revisions (ревизии) and ultimately real

---

7 Compiled from the late 15th to the late 17th centuries by revenue agents (писты) and revenue clerks (подьячие) in the service of the Chancellery of Service Lands (Поместный Приказ), cadasters amounted to regular surveys of the lands of the Muscovite state. Therein one can find descriptions of individual villages, the number of peasants, quantity of arable land, whether slash-and-burn farming was practiced, the amount of meadow lands, and data on forests. Some 700 Tatar settlements are mentioned in the 17th-century cadasters for the Kazan region.
censuses (the first, though not until 1897) accumulated individual-level data on a large scale, but used them to categorize people according to habitation, gender, ethnicity, confession, schooling, or livelihood. Unimaginable numbers of individual records (карточки) disappeared into boxes that slowly gathered dust in grand depositories, somewhat like the site in which the Ark was stored at the end of the film “Indiana Jones and the Lost Ark.” No one paid any attention to their preservation, with the result that almost all have disappeared.

Fortunately, one source of immense richness in both quantitative and qualitative information about real people—the confessional registers (метрические книги)—is extant and accessible. It has remained virtually unexplored, save for a few small projects on rather restricted areas of the empire, and a major general study of the source by D. H. Antonov and I. A. Antonova. In the late 1820s, as part of its increasing interest in data gathering about every aspect of its diverse population and its activities, the Russian government began to require maintenance of annual metrical books by the Muslim clerisy. These books were to record the vital statistics relating to births, deaths, marriages, and divorces within the purview of the lowest institutional level of the faithful, the mosque and its mahalle (community), and to transfer these books to bureaucratic offices each year. While the extent of each record depended on many factors, including the dedication and diligence of the local imam, examination of Muslim metrical books reveals remarkably comprehensive data, particularly on the pages devoted to marriages, wherein wide information concerning the bride and groom, their respective families, the witnesses to their marriage, and gifted dowries abound with every year. In each of the books, we are introduced to “real” people, with names and relations.

From the metrical books we can gather absolutely unique data about the most basic level of society that, once compiled into a massive longitudinal dataset, will enable us to peer into several thousand Turkic communities within the Russian Empire (in the nineteenth century, between 10 and 12 million people) for a period of 75 years (nearly four generations). We will be able to identify all community

---

8 Periodical "revisions" of the population were carried out in 1719, 1744-45, 1763, 1782, 1795, 1811, 1815, 1833, 1850, and 1857 for purposes of collecting the poll tax (подушная подать). Since only males were taxed, these compilations tell us virtually nothing of women. The 1795 revision appears to be the most complete and best analyzed.

9 Метрические книги России XVIII-начала XX в. (Москва, 2006).
members, begin to think about the lives of mostly ordinary people, and enjoy the possibility of contemplating otherwise unapproachable topics:

- Life under and through multiple pressures;
- Mortality and living standard;
- Evidence of social mobility;
- The influence of kinship on social and demographic behavior;
- How kinship networks and household context influenced such social demographic outcomes as employment, marriage, reproduction, and survivorship;
- Test the assumption that kinship becomes less influential with the trajectory of modernization, especially economic commercialization and state penetration;
- Fate and fortune;
- Reconstructing and analyzing life histories from longitudinal data;
- Demography of resettlement;
- Gender relations.

Our project is designed to be long-term and to fall into a number of discrete phases identified primarily by the territory covered:

- Phase 1: Mosques & mahalle within the city of Kazan'
- Phase 2: Mosques & mahalle within the remainder of Kazan' Province
- Phase 3: Mosques & mahalle in Orenburg and Ufa Provinces
- Phase 4: Mosques & mahalle in the remainder of the Volga-Kama region (Viatka, Perm, & Tobol’sk Provinces)
- Phase 5: Mosques & mahalle in Nizhegorod and Tambov Oblasts
- Phase 6: Mosques & mahalle in Samara, Saratov, and Astrakhan Provinces
- Phase 7: Mosques & mahalle in the Caucasus and Tavrida Province
- Phase 8: Mosques & mahalle in Siberia
- Phase 9: Remaining identifiable mosques & mahalle

With funding from the Digital Humanities Start-Up Program under the auspices of the National Endowment for the Humanities, we began Phase 1 in April 2016 with a small team including several historians in the Russian Federation who are native speakers of both Russian and Tatar—the latter being the language of the records—and who are skilled in working with the Old Tatar (Iske Tatar) that prior to the
1920s was written in Arabic script. We gradually added other team members, one to create the database that would house what the confessional registers provided, one to provide English versions of the data through application of machine translation, one to develop the project website and begin creation of data visualizations, and a fourth to assist with the identification of repositories of the extant confessional registers in various archives and libraries scattered across the Russian Federation and parts of the former Soviet Union.

During this first phase, we engaged in the following activities:

- Arranged for access to the metrical books housed in the National Archives (NA RT) of the Republic of Tatarstan;
- Created Excel files for collecting birth, marriage, divorce, and death data for seven historic mosques in Kazan’ City;
- Identified data categories and established consistent lexicons in Tatar, Russian, and English based on the data records;
- Created the database for storing all information collected from the confessional registers;
- As the data for each mosque arrived to the team PI, they were cleaned and then passed on to our computer-translation specialist to run the raw data through Apertium software and create translations to English of data; in both Tatar and English forms, the data was then uploaded to the database based upon each of the four categories;
- Explored the Russian archives for documentation on the history of the confessional registers and for information as to repositories in provinces beyond that of Kazan’; engaged with the National Archive of the Republic of Bashkortostan where the records of the Orenburg Muslim Spiritual Administration are preserved, including the metrical books from all mosques under its supervision. We have begun creating an Excel file of these records;
- Ran initial analyses of data collected;
- Created visualizations of the data based upon queries;
- Prepared four semi-annual reports for NEH;
- Created an open-source website to make all data available from the confessional registers along with a variety of texts and charts to support the
project. Still off-line, the site may be reached at http://mosquedata.teamernst.net;

- Creation of **Attendant Resources**: As we noted in semi-annual reports, the US branch of our team continues to develop a set of indices, inventories, and a variety of visualizations focused on historical and current images, charts, tables, and maps to supplement and extend the broader impact of our project. All those listed below are either **completed** or **underway**, and copies may be requested from the PI:

  - **Images**
    - Selections of Pages from Metrical Books [Completed]
    - Historic and current images of mosques [Completed]
    - Historic heads of mosques [Underway]
    - Plans of Land Allotments in Villages of Kazan’ Province [Underway]

  - **Maps**
    - Of Kazan’ City
      - 1768 [Completed]
      - 1884 [Completed]
      - 1887 [Completed]
      - 1889 [Completed, with Tatar labels]
      - 1910 [Completed]
      - 1914 (With demarcation of the Old Tatar Suburb) [Completed]
    - Of the Old Tatar Suburb with its 10 Mosques [Completed]
    - Of the territory administered by the Orenburg Muslim Spiritual Authority [Completed]
    - Of the location of each mosque in Kazan’, with geographic coordinates [Completed]
    - Of the regions within Kazan’ Province [Completed]

  - **Indices**
    - Index of Imperial Russian Occupational Terms [Completed]
    - Index of Imperial Russian Weights and Measures [Completed]
    - Index of Russian Familial Terms [Completed]
    - Index of Russian Geographical Terms [Completed]
    - Index of Russian Religious Terms [Completed]
    - Index of Russian Laws Concerning Metrical Books for all Confessions within the Empire [Underway]

  - **Inventories**
    - Inventory of Mosques in Kazan’ City, with relevant information [Completed]
• The Number of Mahalle with Mosques in the Russian Empire, 1883 [Completed]
• Data Records from Metrical Books for Kazan’ City Mosques [Completed]
• Inventory of Extant Metrical Books associated with Mosques from across the Russian Empire [Underway]
• Snapshot inventory of Mosques and Mahalles on the Territory of the Russian Empire in the year 1886, with data on population by gender and number of clerics [Underway; this is a large and data-intensive inventory that includes information on all extant metrical book]
• Inventory of Populated Sites in Kazan’ Province, 1859 [Underway]
• Inventory of Villages in Kazan’ Province, 1910-1914 [Underway]
• Inventory of Regulations Affecting Muslims in Matters relating to Mixed Marriages and the Religious Upbringing of Children from such Marriages [Underway]
• Builders & Patrons of Mosques [Underway]

• Tables, Graphs & Charts
  • Number of Mahalles in the Russian Empire, 1886 [Completed]
  • Table Listing Categories of Data Sought by Muslim Metrical Books [Completed]
  • Kazan’ City Population by Confession [Completed]
  • A Tatar Lexicon for use with Muslim Metrical Books [Underway]
  • Comparative Chart of Religious Institutions—Orthodox Churches, Roman Catholic Churches, Muslim Mosques, & Jewish Synagogues [Underway]
  • Administrative Divisions within Kazan Province, as of 1802 [Completed]

• Bibliography [Completed]

Future Tasks

• Bring the project web presence to online status after expanding it linguistically to encompass Tatar, Russian, and English;
• Bring to a conclusion the work on the mosques within Kazan City, then move to those within the remainder of Kazan Province;
• Create and test an extraction process that allows our data collectors to input data from the records directly into the database, avoiding the need for intermediary Excel files. Moreover, we hope to train our transcribers to not only transcribe the records word for word, but to identify “atoms” of
information, such as bride and groom, during the transcription process. For details, see “Metrical Book Field Analysis” in Appendices;

- Create and test the usability of a computer interface for capturing the data directly from the metrical books without need for a human intermediary;
- Parse familyinfo column for births as preparation for analysis;
- Trace child bearing over time (longitudinal) to build lineages;
- Begin seeking international collaborators to open up this project to other confessional communities within the Russian Empire, e.g., Lamaists, Roman Catholics, Old Believers, Jews, etc.
- Seek future funding from federal and private agencies to continue and expand this project.
- Attached is a document—“Metrical Book Field Analysis”—composed during the history of the project that captures many of the problems and difficulties we confronted and the resolutions we produced or proposed for latter attention.
This document contains an analysis which breaks down natural language fields in the metrical books into distinct fields which can be considered discrete units of actionable, analyzable information that is of interest to the broader. Designing means to extract these fields directly from the data is high priority, as field identification is probably the largest current bottleneck to downstream research. This analysis is being conducted with the intent of asking the transcribers in Tatarstan to manually pick out the information under these fields during the process of transcription. This will add somewhat to their work burden, but it saves a large amount of work downstream, and also will increase the possibility of automating the extraction at some point in the future. For the purposes of easing the process of pulling out this data for the transcribers, we may build an online interface that is directly attached to a database. This document will help with the design of that database.

I am going to start off this analysis, but as the person with the least “domain expertise” there’s a high probability that I’ll either miss a vital piece of information or misunderstand a field. Prof. Lazzerini: I think you will play the final arbiter as to whether or not we have captured all the information we’d like to mine from the metrical books. It is extremely important that this analysis is as complete and as "final" as possible at this stage in the project, as "finding out" that we are missing a field long after we “finalize” this schema will be extremely problematic (in the sense that we’d have to “go back” to everything that has been entered and “do it again”). I think after we’ve made this document we should also run it by our Kazan colleagues to do a spot check.

**Birth Records**

1. Birthplace
   a. English: Where Born
   b. Tatar: Кайда туган
   c. Malic Notes: it’s difficult for me to ascertain if there’s a "standardized" way this field can be broken down. For the most part, this field seems to indicate the owner of the house in which the child was born. The most common response seems to be "his/her own home," according to Ilnar's translations. However, some data in this field seems to have more fine-grained information, such as Захарьевский урамында Зөһәрә Апанаева йортында, which I believe means "Zohre Ananeav's house on Zaxar'evskiy Street," and "Кул буенда атасы йортында," or "By the lake at the father’s
house.” Thus, we can either leave this field as it is or try break it down to subfields such as:

1) Owner of Birth Household (Zohre Ananaev or father)
2) Location of Household (Zaxar’evskiy Street, "by the lake")—I am unsure how fine grained this subfield could get. Based on what you’ve seen, are there instances where the Imam includes the village name? The county? Etc.? If so, we could anticipate this but also include subfields such as "Village" "County" etc. We’ll have to make this call based on what we see in the data.

2. Family Info
   a. English: Name of Child's Father and Mother; Grandfather's Name on Both Sides; Family Name
   b. Tatar: Баланың ата анасының һәм һәр ике тарафта улан бабасының исеме, әгәр дә булса фамилиясе, каю таифәдән идеке
   c. Malic Notes: this is definitely a natural language field we want to break down into discrete units. Here are some that I've found and believe should be standardized.
   d. Proposed subfields:
      1) Father Name
      2) Father Origin (most entries say where the father is from) (optional)
      3) Father Class (urban commoner, merchant) (optional)
      4) Paternal Grandfather Name (most entries say "father, son of X," where X would be the paternal grandfather)
      5) Mother Name (to the best of my knowledge, most entries do not include information about the class or the origin of the mother, but I could be mistaken)
      6) Maternal Grandfather Name (like with the father, entries usually say "mothername, daughter of X")
      7) Others?

3. BirthDate
   a. English: Birth Day and Month
   b. Tatar: Каю айда туган, айның ничәнче көне
   c. Malic Notes: I believe this field is already discrete and doesn't need to have anything pulled from it.

4. ChildName
   a. English: Name of Child
   b. Tatar: Баланың исеме
   c. Malic Notes: I believe no need for discretization.

Death Records

1. BurialPlace
   b. English: Where Buried
   c. Tatar: Кайда жирәнгән
d. Malic Notes: I believe there is no need for discretization

2. CauseDeath
   a. English: cause of death
   b. Tatar: Нинди авыру
   c. Malic Notes: Although there is a lot of variation in how this information is recorded (e.g., the same cause of death could be written differently), I think for the most part this field needs no discretization by the transcribers. However, if the transcribers can come up with a standardization that maps different "variations" of the cause of death to a canonical vocabulary, we could include a subfield that is "cause of death – standardized."

3. AgeDeath
   a. English: age of death
   b. Tatar: Мәетнең яше
   c. Malic Notes: Although this field seems “already discretized,” I think it merits potential discretization into units of time. This should be relatively easy and painless for the transcribers to do and will help us sort by age enormously. Otherwise, we’d have to develop code to identify if a number in this field is months, years, days, etc., and there is variation in how this information is recorded (4 aylik vs. 4 ay) that is easy for the transcribers to spot but more troublesome to account for via code.
   d. Proposed Subfields:
      1) Years
      2) Months
      3) Days
      4) Others? Weeks? I didn't spot other unit of times in a quick glance.

4. FamilyInfo
   a. English: Names of Father and Grandfather
   b. Tatar: Ата вә бабасының исеме
   c. Malic Notes: should be discretized
   d. Proposed Subfields:
      1) Father Name
      2) Father Origin (optional)
      3) Father class (optional)
      4) Grandfather name (optional? It seems like in some fields this information is not supplied)
      5) Others?

5. DeathDate
   a. Malic Notes: no need for discretization

6. NameDeceased
   a. Malic Notes: no need for discretization

Marriage Records
Professor Lazzerini: the marriage and divorce records are probably the toughest for me to analyze. They include very large, complex natural language fields that have a huge amount of information in them. I might easily miss or misunderstand something. I think these records will be the ones you'll have to look most closely at and bring your expertise to bear on.

1. ImamInfo
   a. English: Which imam recited the prayers
   b. Tatar: Никах укыган имам
   c. Malic notes: definitely need to discretize
   d. Proposed Subfields:
      1) Imam Name
      2) Imam Father
      3) Imam Mosque (field seems to indicate which mosque the imam is from)
      4) Amount Paid (seems like in some instances the information includes how much the imam was paid, commonly 25 "kopec" according to Ilnar's translations)
      5) Others?

2. DowryInfo
   a. English: The amount of the dowry, how much was handed over and how much was not; and who accepted the dowry handed over and who accepted the dowry to be delivered at a later date, and his signature
   b. Tatar: Жөмлә мәһәр күпме, нә кадәр вә бирелгән вә нә кадәр бирелмәгән. Вә бирелмәгән мәһәр кем зәммәсендә вә бирелгән мәһәрне кем алган, аның имзасы
   c. Malic notes: Need to discretize. Very hard for me to parse, will need help on this one. This seems to me to be a field including all information about the dowry being handed over by the groom's family. Here are some proposals:
      1) Dowry Handed Over
         a) Dowry Amount Handed Over(an integer)
         b) Dowry Handed Over Unit (according to Ilnar's translation, I see "coins" and I also see "mahr of coins" - not sure what that means.
            Are there other forms of dowry? Is it always money?)
         c) Repeatable?
      2) Dowry Debt
         a) Dowry Amount Debt (integer) - for any dowry promised, but not delivered at the wedding
         b) Dowry Debt Unit
         c) Repeatable?
            MALIC NOTE: can the dowry be "split" further into different types? We'll need an answer to this question to determine if this
is a "repeatable" field. E.g dowry handed over, 10 coins and 1 sheep

3) Handed Over Recipient Name?
   a) Handed Over Recipient Origin?
   b) Handed Over Recipient Class?
   c) Handed Over Recipient Father?
   d) Repeatable field? Multiple recipients?

4) Debt Recipient Name?
   a) Debt Recipient Origin?
   b) Debt Recipient Class?
   c) Debt Recipient Father?
   d) Repeatable?

5) Envoy? (Ilnar's translations of this field often has the word "envoy" in it. I'm unsure who this is. If it's the person giving the dowry? Receiving it?)
   a) Envoy Origin?
   b) Envoy Class?
   c) Envoy Father?
   d) Repeatable?

6) Witness (people who signed off on the dowry transaction)
   a) Witness Name
   b) Witness Father
   c) Witness Class
   d) Witness Origin
   e) REPEATABLE FIELD: there is often more than one witness=

3. WitnessInfo
   a. English Name: [Names of nobles or of authorized persons and witnesses who bring the marriage contract to the wedding ceremony and consent to the marriage]
   b. Tatar Name: Никакта гакыд йөрткән асыллары яки вәкилләренең вә шәнитләренең исемнәре вә куллары. Ихтыярлары илә эйтсәләр тәгъликъ сурәте
   c. Malic notes: Definite discretize. I interpret this field to be members of the bride's family who sign off on the marriage. There are often multiple witnesses to the "signing off" of the marriage
   d. Proposed subfields:
      1) Envoy
         b) Envoy Name
         c) Envoy Class
         d) Envoy Father
         e) Envoy Origin
         f) Repeatable?
      2) Witness (people who witnessed the consent of the bride's family)
4. PeopleInfo (information about the bride and the groom)
   a. English Name: [Names of the groom's and bride's fathers and grandfathers, with family names if they exist; to which social group do they belong; and which marriage ritual did they follow]
   b. Tatar Name: Ирнең, кызның ата вә бабаларының исемнәре, булса фамилияләре. Каю таифадән идекләре. Вә ћәр берәм-берәмнәрәнән вә ћәр икесенең ничәнче никахләры.
   c. Malic notes: definite discretize. Important observation: this field distinctly says "last name, if exists." If a person can have a last name, that means that we'll have to add an optionall "last name" field to every person subfield above (witnesses, fathers, mothers, envoys, etc.). IMPORTANT NOTE: it seems that several clerks put the AGE of the groom and bride in this field, even though there is a separate field for bridge and groom age. I suggest we keep the ages separate.
   d. Proposed Subfields:
      1) Groom
         b) Groom Name
         c) Groom Surname (optional)
         d) Groom Father
            (1)Groom Father Class?
            (2)Groom Father Origin?
         d) Groom Class
         e) Groom Origin
         f) NO GROOM AGE: if groom age is presented here, enter in other dedicated field
      2) Bride
         a) Bride Name
         b) Bride Surname (optional)
         c) Bride Father
            (1)Bride Father Class?
            (2)Bride Father Origin?
         d) Bride Class?
         e) Bride Origin?
         f) Virgin? (boolean True/False or "no data" - this is mentioned in many fields. Record if interested?)

5. AgesDate
   a. English Name: Age of groom and bride on date of marriage
   b. Tatar Name: Никах булган ай ћәм кән ир илә кызның яшьләре
c. Malic notes: If the ages of the bride and groom are mentioned in PeopleInfo, record it here anyway

d. Proposed Subfields:
   1) Day of Month
   2) Month
   3) Groom Age (years)
   4) Bride Age (years)

**Divorce Records—Also a difficult field to parse down. Ilnar’s translator seems weakest on this data so it's the most difficult for me to figure out what's going on in these fields.**

1. ImamInfo
   b. English Name: [Which imam... was present at the declaration of талак and which at that of хәлегь, and which imam conveyed the agreement]
   c. Tatar Name: Кәю имам хөзүрнәда талак булган хәлегъ сүрэтендә каю имам гәкыд ыйырткән
   d. Malic notes: need to discretize
   e. Proposed subfields:
      1) Talak Imam?
         (b) Imam name, imam father name, imam mosque
      3) Xeleg Imam?
         (a) Imam name, imam father name, imam mosque
      3) Agreement conveying imam?
         (a) Imam name, imam father name, imam mosque

2. DayMonth
   a. English name: [Day and month on which the талак or хәлегь was pronounced]
   b. Tatar name: Кәю айда Һәм айның ничәнче көнендә талак яки хәлегъ булган
   c. Proposed subfields:
      1) Day
      2) Month

3. WitnessInfo
   a. English name: [The names of witnesses and elders and of those who authenticated the event; from where do these people come?]
   b. Tatar name: Шаһитларның Һәм аталарының исемнәрә вә куллары. Кайдағы кешеләр идәк
   c. Proposed subfields
      1) Witness (MULTIPLE)
         b) Witness name, witness father name, witness origin, witness class

4. SepReason
   a. English name:[What were the reasons for the талак or the хәлегъ, and what were the circumstances surrounding the separation?]
b. Tatar name: Ни сәбәптән талак яки хәлегь булган ничәнче мәртәбә фәрекатләре

c. Malic Notes: I'm honestly not sure where to start here. It's very complicated. We could try to make a curated, finite vocabulary of the most common divorce reasons. This field, more than any other field, is a full-blown natural language “story”

d. Proposed subfields
   1) Reasons????
   2) Witness? (there seem to often be witnesses in this subfield)
      a) Witness name, witness father name, witness origin, witness class

5. People Involved

   a. English name: [The names and surnames of the husband and wife who wish to separate, and of their parents. Who signed the agreement for the хәлегь? What was provided for the divorce as recompense for bridewealth? Who signed the agreement for the recompense of the bridewealth?

b. Tatar name: Фәрәикат улынган ир илә хатын һәм ата вә бабаларның исемнәре вә фамилияләре. Хәлегь сурәтенә әкәд итүчәләрнең имзалары. Бәдәл хәлегь нинди шәйләр иде. Бәдәл хәлегь ни алган кемсәнән имзасы

c. Malic notes: another complicated field, but with SOME information that is easily identifiable

d. Proposed subfields:
   1) Divorcing husband
      b) Name, father name, origin, class,
   2) Divorcing wife
      a) Name, father name, origin, class
   3) Agreement signer?
   4) Recompense? Number, units? Multiple?
   5) Recompense witness?
      a) Name, father name, origin, class
<table>
<thead>
<tr>
<th>Births</th>
<th>Marriages</th>
<th>Divorces</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s gender</td>
<td>Ages of bride and groom on the date of marriage</td>
<td>The names and surnames of the husband and wife who wish to separate</td>
<td>Decedent’s gender</td>
</tr>
<tr>
<td>Child’s given name</td>
<td>Names of the groom’s father and grandfather</td>
<td>The names of the husband’s and wife’s parents</td>
<td>Decedent’s given name</td>
</tr>
<tr>
<td>Child’s birth day and month</td>
<td>Names of the bride’s father and grandfather</td>
<td>The name of the signatory to the agreement for the хəлегь [announcement of separation]</td>
<td>Decedent’s date of death</td>
</tr>
<tr>
<td>Names of child’s father and mother</td>
<td>Family names of the bride and groom if they exist</td>
<td>Recompense offered for the bride wealth (dowery)</td>
<td>Names of decedent’s father and grandfather</td>
</tr>
<tr>
<td>Names of child’s paternal grandfathers</td>
<td>Social class to which each family belongs</td>
<td>Guarantor of the bride wealth</td>
<td>Decedent’s age at death</td>
</tr>
<tr>
<td>Names of child’s maternal grandfathers</td>
<td>Marriage ritual that was followed</td>
<td>The reasons for the тазак or the хəлегь? [Announcements of separation]</td>
<td>Decedent’s cause of death</td>
</tr>
<tr>
<td>Family surname</td>
<td>Names of notables or of authorized persons and witnesses who bear the marriage contact and consent to the marriage</td>
<td>The circumstances surrounding the separation of the couple</td>
<td>Decedent’s burial site</td>
</tr>
<tr>
<td>Social class to which the child’s family is attached</td>
<td>The amount of the dowry, how much was delivered and how much was not at the time of the marriage</td>
<td>The names of the witnesses, elders, and those who authenticated the separation</td>
<td></td>
</tr>
<tr>
<td>Place where the child was born (at home; else-where)</td>
<td>Who accepted the part of the dowry to be delivered at a later date, and his signature</td>
<td>Domicile of the parties</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The name of the imam who conveyed the agreement</td>
</tr>
</tbody>
</table>
Map of Old (northeast) and New (southwest) Tatar Suburbs, Kazan City
Project Team

**Principal Investigator**

Dr. Edward J. Lazzerini  
Director, Sinor Research Institute for Inner Asian Studies  
Professor, Department of Central Eurasian Studies  
Indiana University—Bloomington

Dr. Rafiyya Gimazova [Project Transcription]  
Research Associate  
Marjani Institute of History  
Tatarstan Academy of Sciences  
Russian Federation

Dr. Elmira Salakhova [Project Transcription]  
Research Associate  
Marjani Institute of History  
Tatarstan Academy of Sciences  
Russian Federation

Dr. Rashid Kadyrov [Project Transcription]  
Research Associate  
Institute for the Tatar Encyclopedia  
Kazan, Tatarstan  
Russian Federation

Ilnar Salimzianov [Machine Translation]  
Kazan, Tatarstan  
Russian Federation

David Ernst [Web Developer & Data Analyst]  
Internet-Minded Design & Development (IMDD)  
Bloomington, Indiana

Vincent Malic [Database Creator & Management]  
School of Informatics  
Indiana University—Bloomington

**Consultants**

Dr. George Alter [Network Analyst]  
Research Professor, ICPSR  
Professor of History  
University of Michigan

Dr. Barbara A. Anderson [Demography Consultant]  
Department of Sociology  
Research Professor, Population Studies Center  
University of Michigan—Ann Arbor

Stephanie Dickinson [Technical Consultant]  
Senior Consultant and Managing Director  
Indiana Statistical Consulting Center  
Indiana University—Bloomington

Jonathan North Washington [Technical/Linguistic Consultant]  
Assistant Professor  
Department of Linguistics  
Swarthmore College

Dr. Leyla Almazova [Cultural Consultant]  
Institute for Oriental Studies  
Kazan Federal University  
Kazan, Tatarstan  
Russian Federation

Dr. Rozaliya Garipova [Historical/Cultural Consultant]  
Mellon Post-Doctoral Fellow  
Department of Religious Studies  
University of Pennsylvania

Dr. Paul Werth [Historical/Cultural Consultant]  
Professor of History  
University of Nevada
Bibliography

The Russian Metrical Book and Social History


Bel’kova, Anna. “Metricheskie knigi pervoi poloviny XIX veka tiumenskogo dukhovnogo pravleniia kak zhann delyovoi pis’mennosti.”


Chernovskaia, V.V. “Metricheskie tetradi Jaroslavskoi sobornoi mecheti, kak istochnik izucheniiia makhalli (1907-1917 gg.).” Unpublished paper.


__________. “Raising the Dead: New Sources for the Study of Turco-Islamic Central Eurasia.” Presentation to the Indiana University Islamic Studies Workshop, 2012.

__________. “Rewriting the History of Russia’s Islamic Orient with Big-Data: A Synoptic Overview.” Presentation to the annual Central Eurasian Studies Society Conference, 2012.


_______. “Metrika iaznalyrnda gaile tornyshynyng chaglylyshi.” In press.

_______. “Muslim Parish Records Written in the Tatar Language in the National Archives of the Republic of Tatarstan.” Presented to the Thirty-First International Conference of the European Association of Librarians of the Middle East, Kazan, 22-24 June 2009.


Segal', I. “Neskoli'ko slov o metricheskoi registratsii na Kavkaze.” Kavkaz, No. 95 (13 April 1905).


