A multi-dimensional ontology-based analysis of the Censorship of Hebrew Manuscripts

Gila Prebor <gila_dot_prebor_at_biu_dot_ac_dot_il>, Department of Information Science Bar-Ilan University, Ramat Gan, 52900, Israel

Maayan Zhitomirsky-Geffet <Maayan_dot_Zhitomirsky-Geffet_at_biu_dot_ac_dot_il>, Department of Information Science Bar-Ilan University, Ramat Gan, 52900, Israel

Isaac Miller <Isaac_dot_Miller_at_biu_dot_ac_dot_il>, Department of Information Science Bar-Ilan University, Ramat Gan, 52900, Israel

Abstract

In this study, we present the first large-scale quantitative analysis of a corpus of censored historical Hebrew manuscripts that have survived through the ages. A new multi-dimensional ontology-based approach was applied to explore the geographic, temporal, actor- and subject-based distribution of censorship events. We adopted an ontology-based approach to apply statistical analysis on the metadata of censored Hebrew manuscripts for estimating the scope and quantifying the extent of the known facts on the censorship activity and its various characteristics over the years. In addition, we revealed some previously unknown phenomena and trends. Particularly, we analysed the relationship of censorship on other types of events in manuscripts' lifecycle and compared the distribution of censored vs. non-censored manuscripts in different dimensions. We also devised a set of rules to complete the missing locations of over 50% of censorship events, which has substantially changed the big picture of spatial distribution of censorship activity. From the temporal perspective our findings demonstrate that censorship was conducted in "waves" and there was a decrease in the creation of new manuscripts in periods of high censorship activity. Certain subjects, such as Kabbalah and Philosophy were censored significantly more than others, and the locations and script types' distribution in censored manuscripts differs from the non-censored manuscripts.

Introduction

Our study aims to investigate the historical phenomenon of censorship of Hebrew manuscripts performed by the Catholic Church in Italy[1] during the 16th-18th centuries (1550-1775). Thus far, most of the studies on the censorship of Jewish manuscripts and books during this period were done in a qualitative and focused manner [Baruchson-Arbib 2007] [Phillips 2016]. [1]

Recently, quantitative methods have been embraced by scholars for the analysis of Big Data in the Humanities [Kaplan and de Lenardo 2017] [Lei Zeng 2017]. Accordingly, in this paper we present the first (to the best of our knowledge) large-scale quantitative analysis of historical Hebrew manuscript data. For the purpose of this study, we identified 5,185 censored manuscripts in the catalogue of the Department of Manuscripts and the Institute of Microfilmed Hebrew Manuscripts in the National Library of Israel ((http://web.nli.org.il)), which represent the largest collection of Hebrew manuscript metadata in the world. [2]

In Ethington (2007) the author argues that "the past is the set of all places made by human action. … Knowledge of the past, therefore, is literally cartographic: a mapping of the places of history indexed to the coordinates of spacetime" [Ethington 2007]. Inspired by the multi-perspective framework of Ethington's theory of visual representation of the past we propose a new multi-dimensional methodology for the large-scale quantitative analysis of historical manuscript data. Particularly, we examined the temporal, geographic, actor- and subject-based dimensions of the censored manuscripts [3]

and also compared the statistical distribution of censored manuscripts on these dimensions to the non-censored manuscript corpus. The proposed methodology is open-ended, as it is not limited to the dimensions specified above and can be extended with more dimensions which might be relevant for a given data collection.

The availability of a relatively large censorship database on one hand, and the ability to apply to it modern analysis and visualization techniques on the other hand, has the potential to discover new insights into censorship activity over the years. This includes both revealing unknown phenomena or trends as well as bringing new (mainly quantitative) perspectives to known facts.

Thus, the research questions posed in this research are:

How are censored manuscripts distributed in various dimensions mentioned above? More specifically we explored:

1. When were the peaks, inclines and declines in censorship activity and can they be explained by historical knowledge of the corresponding periods?
2. In what cities did censorship take place?
3. What major geographic transitions can be observed from the large-scale quantitative analysis of censorship data and what is known about them in historical literature?
4. Who were the most dominant censors?
5. Did censors work in collaboration and to what extent?
6. What was the rate of repetitive censorship, when a certain manuscript was censored more than once?
7. Finally, what were the main subjects of the censored manuscripts?

What are the differences between censored manuscripts in Italy in the 16th-18th centuries vs. non-censored manuscripts in the collection in various dimensions? These differences can shed light on special characteristics of the censored corpus which distinguish it from the rest of the manuscripts. Particularly, as part of the spatial analysis we investigated the following phenomena:

1. Whether there is a difference in the distribution of censored vs. non-censored manuscript script types and whether this distribution changes over time?
2. Whether the distribution of censored manuscripts among their current storage locations differs from the overall distribution of storage locations of the dataset?
3. In addition, from the subject analysis perspective we examined: Whether the distributions of censored vs. non-censored manuscripts differ according to subjects?

Historical Background

In the context of the Counter-Reformation, during the 16th-18th centuries, the Catholic Church closely supervised written and printed literature. Papal bulls were issued against proscribed books, and indices were drawn up containing lists of prohibited books (*Index Librorum Prohibitorum*), which also included Hebrew books. In addition to the indices that were written for literature in general, some indices listed Hebrew books only (see Prebor 2003; Raz-Krakotzkin 2007; Francesconi 2012; Francesconi 2016; Phillips 2016; Van Boxel 2016). [2]

The techniques of censorship directed towards Hebrew literature evolved concurrently with, and as part of the development of church censorship as a whole, and were formulated by the same bodies. Italian Jews who wished to save their books operated in two ways: on the one hand, internal censorship of Hebrew printed books and, on the other hand, by lobbying the Pope and Church leaders to ease the instructions for confiscating books and forbidding the printing of the Talmud.

The Jewish lobbying of the Pope and Church leaders to ease directives regarding confiscation of manuscripts and books led to the operation of external censorship [Baruchson-Arbib 2007]. One of the concessions won by Italian Jews

was the introduction of the censorship of manuscripts and printed books possessed by the Jewish public, as an alternative to confiscation and wholesale destruction of virtually all the books they possessed. The term *purification* (*Ziquq*) is the commonly used term in this context, parallel to the Latin term *expurgatio* [Sonne 1942] [Benayahu 1971]. To carry out the expurgation, the Church put censors at the disposal of the communities, most of them apostates and experts in the Hebrew language, and the communities undertook to pay their fees. The involvement of the Jews in the purification work arose from the fundamental approach of the Church that the Jews were responsible for the "forbidden literature" in their possession.

The owners of Hebrew books were required in the name of the Inquisition to hand them over to the censors, who kept them in the local office. Concealing books was punished by their confiscation and a heavy fine, and in certain cases, also by the confiscation of property and imprisonment. The books collected were examined by the censors, who destroyed the forbidden books and punished those who kept them. The books of a controversial nature were expurgated and returned to their owners with the approval of the censor.

9

The censorship procedure entailed erasing or replacing all the paragraphs which, in the view of the censors, contravened the principles of Christianity, its customs and rites, or those paragraphs which contained blasphemy, heresy or errors. When the expurgation of the book was completed, the censor wrote a short approval note. The approval note was written for the most part on the final page of the book and sometimes on the title page. Proof of the activity of the various censors can be found in a large number of manuscripts and printed works which have survived until the present day, and in which ink erasures, emendations, signs of removed pages and the signature of the censor can be observed.

10

In the next section we review literature on computational methods for digitization and large-scale quantitative analysis of historical corpora and present several case studies of this research paradigm.

11

Related work

In the past decade computational methods were proposed and implemented for historical data analysis, e.g. Geographic Information Systems and spatial analysis for historical collections (see Gregory and Healye 2007; Ayers 2010; Moretti 2005; Knowles and Hillier 2008), temporal and subject-based analysis and visualisation of historical and cultural heritage collections (see Glinka et al. 2017; Chandna et al. 2016).

12

In this section we review several projects applying quantitative analysis of historical data on literary resources, such as hand-written manuscripts and printed books. The Schoenberg Institute of Manuscript Studies, maintained by the University of Pennsylvania Libraries, presents the multi-faceted analysis of the Schoenberg Database of Manuscripts (SDBM) (https://sdbm.library.upenn.edu/). The database contains over 200,000 records representing 90,000 hand-written manuscripts from diverse collections and institutions (Van Hooland and Verborgh, 2014). The SDBM's website displays statistics on manuscript distribution over time and space as well as by subject, authorship and provenance based on historical sales lists. The Stanford project Republic of Letters presents quantitative analysis and visualisation of Voltaire's letters, distribution of letters and writers displayed on a map and timeline, as well as their correspondence network [3].

13

A closely related project which aimed to organize and quantitatively analyse the data on Jewish historical manuscripts is *SfarData* (http://sfardata.nli.org.il/). *SfarData* is the codicological data-base of the Hebrew Palaeography Project which locates all the medieval codices written in Hebrew script, which contain explicit production dates (until 1540) or at least scribe names (some 6000 codices). The project aims to study and document the codices' visual and measurable material features and scribal practices in situ; to classify these features and practices in order to build a historical typology of the hand-produced Hebrew book. This, in turn, provides users of Hebrew manuscripts with a tool for identifying the production region and assessing the period of the studied manuscripts [Beit-Arié 2018].

14

Several projects investigate printed book distribution over time and space. The Footprints project maintains a database of printed Jewish books (https://footprints.ccnmtl.columbia.edu/). The goal is to consolidate and analyse scattered information on Jewish books (in Hebrew, in other Jewish languages, and on books in Latin and non-Jewish vernaculars

15

with Judaica content) in order to build up a composite view of the movement of Jewish texts and ideas from place to place and across time. The French Book Trade in Enlightenment Europe (FBTEE) project [Burrows and Curran 2012] (http://fbtee.uws.edu.au/stn/) uses database technology to map the French book trade across late-Enlightenment Europe, between 1769 and 1794. It investigates best-selling texts and authors; reading tastes across Europe; changing patterns of demand over time; and networks of exchange in the print-trade. The project tracks the movement of around 400,000 copies of 4,000 books across Europe. It details, where possible, the exact editions of these works, the routes by which they travelled and the locations of the clients that bought or sold them.

Method

The catalogue's data structure

The catalogue of the Department of Manuscripts and the Institute of Microfilmed Hebrew Manuscripts in the National Library of Israel, lists most of the Hebrew manuscripts in existence in the world, whether held in public or in private collections. The catalogue provides rich data about these manuscripts, including references to censorship (http://web.nli.org.il/sites/NLI/English/collections/manuscripts/Pages/default.aspx). [16]

The manuscript catalogue was exported as an XML file on January 31, 2016, and contained a total of 163,248 records. First of all, we excluded 30,418 Genizah manuscript records (18.6% of all the records). The Genizah manuscripts are fragments of codicological units that were separated from their original manuscripts (due in part to damage or deterioration of codices) and were preserved in designated storage units called Genizah (from Hebrew "גניזה", the verbal noun signifying the act of storing/archiving). Most of them are from the famous Cairo Genizah and some are from the European Genizah. These fragments are not relevant to our study because they have different characteristics than codex manuscripts and since Genizah research is a separate field of study [Richler 1990]. [17]

Next, in order to create a dedicated censored manuscript collection, we automatically searched the catalogue's XML file for records with the word stem "censor" (both in English and in Hebrew). In almost all of the cases, a proper "censor" stem was found in a dedicated MARC field, leaving only the cases where some morphological variations of the stem "censor" were found for manual inspection. The obtained dataset was further narrowed by filtering out some tens of manuscripts that were not censored in Italy and/or were censored later than 1775. The final dataset contains 5,185 censored manuscripts. [18]

The structure of the catalogue follows a fundamental observation about the formation of medieval codices. The word "manuscript" has multiple meanings including: scrolls and codices, entire volumes (these can include several works copied by different scribes either at the same time or in different periods), a fragment of a few folios or a single folio. In our research we analyzed only codices. In every codex we found three fundamental layers: (1) the entire manuscript – the Bibliographic Unit (BU) that refers to the physical entity comprising the entire codex with its shelf-mark in the collection (generally corresponding to microfilm numbers). (2) Codicological Units (CU) – Some of the bibliographic units include a number of codicological units, namely a number of different parts of the manuscripts that were produced in different locations and/or at different time periods, either with or without a common theme. (3) Paleographical Units – The codicological units can also contain a number of paleographical units, and fragments written by different hands [Sirat 2002] [Beit-Arié 2018]. [19]

The catalogue database's basic unit, the record, describes a single CU. When the entire manuscript comprises a single CU (i.e. the codex is a single manuscript produced in one place and time), then the BU and the CU overlap and we obtain a *unified codex* [Beit-Arié 2018]. The record's data, in this case, refers to this manuscript as a whole (the majority, more than 90% of the manuscripts in the catalogue are of such a unified nature). The catalogue does not have, however, an independent notion (e.g. a special record type) of BU. Rather, when a codex comprises several CUs (termed as an *assembled codex* - multi-CU), the catalogue contains a special record, an *index record* that describes and lists the CUs of this codex BU. This set of CUs (excluding the index record) comprises a multi-CU BU with the shelf-number as its unique ID. In contrast, when there is a series of records that share a shelf number, but there is no index record with the same shelf number, we did not construct a BU for them and treated each record as an independent BU [20]

that contains a single CU. The index records themselves were not counted while calculating the various statistics of this study as in most cases they only summarize the data already contained in the individual CU records. However, we did use the data contained in index records when we built the study dataset, mainly as a means to verify the completeness of the data gathered from the BUs individual records (e.g. in some cases, the cataloguers specified the normalized full censor name only in the index record, while in the individual records they only gave a partial name).

In many cases it is difficult to decide when exactly in time a multi-CU BU (a codex) was bound. For example, if two censor signatures appear in a single BU, one dated to 1590 and the other dated to 1600, then one possibility is that this codex was bound before censorship (i.e. before 1590 (and censored together, while another possibility is that different CUs in this codex were censored in different settings/occasions only to be later bound together. Also, in some cases when the censor signed the manuscript, he also wrote the name of the place where the censorship took place (e.g. Mantua, Lugo). Following the previous example, by comparing the 1590's place of censorship to the 1600's place of censorship one may learn about manuscript transitions between people's hands and places. We manually checked all the BUs that contain two CUs or more. For each such BU, we used several inference rules to determine whether the signatures that the codex contains refer to all the manuscripts (CUs) in this BU or not. For instance, when the signature(s) appears at the end of the codex we assumed that the signature(s) refer(s) to the entire BU. If different types of script were used in different CUs of the BU, this can be an indication for distinct manuscripts that were censored separately.

Based on the above observations, BUs can be logically divided into two types of virtual entities: *Whole Censored Unit* and *Non-Censored Unit*. A *Whole Censored Unit* (WCU) groups CUs (of the given BU(that underwent censorship as a whole textual unit. The rest of the CUs in a BU, which are non-censored CUs are grouped under a *Non-Censored Unit* (NCU). We, thus, split partially or wholly censored BUs to one or multiple WCUs and a single NCU (see Figure 1).
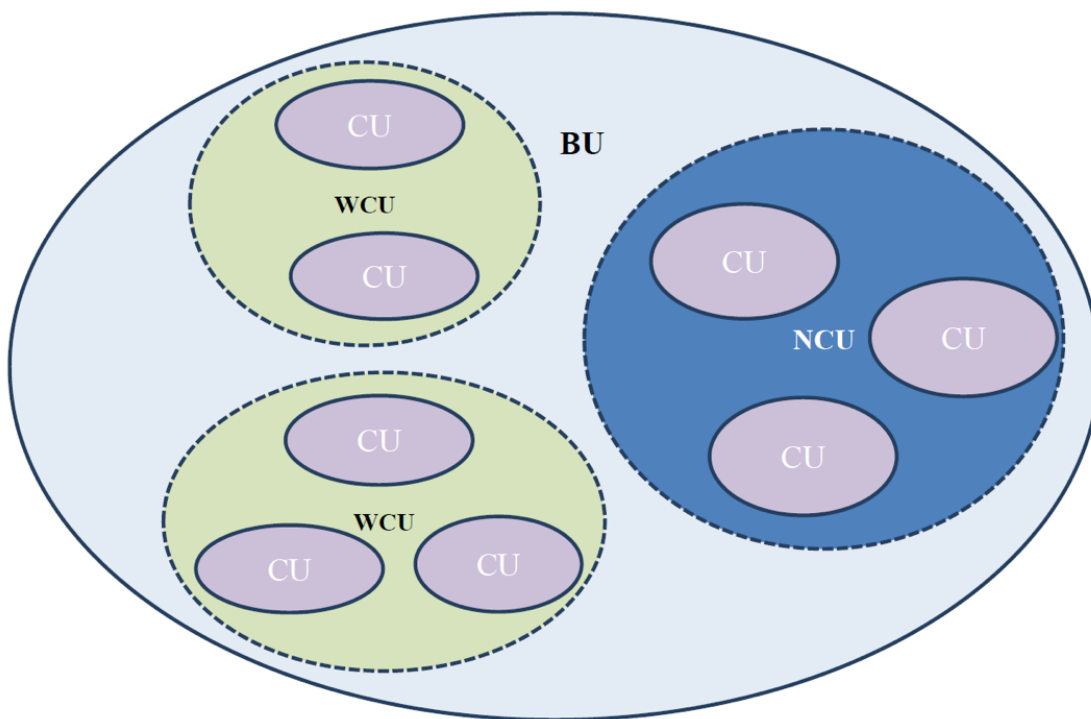

Figure 1. The structure of the study's dataset. WCUs and NCU have a hyphenated borderline to highlight that they are virtual entities created for the purposes of this study.

We found that out of the 586 examined BUs, 428 were censored as a single WCU. Out of the remaining 158 BUs, 144 BUs have undergone censorship as a single WCU and were bound together with a non-censored unit at a later time. In 13 cases, we identified two separate WCU units, and in only one case did we split a BU into three distinct WCUs as we could identify that each one of them had been censored at different occasions and only later bound together into a single physical codex.

The proposed data model

In this study we adopted the event-based modelling approach presented in Zhitomirsky-Geffet and Prebor 2016 and defined three types of events relevant to the goals of this research:

1. Manuscript creation event (CRE) – representing an individual CU's creation including the creation date and place, the script type and subject/s of the manuscript;
2. Manuscript censorship event (CE) – representing a single censorship action on a single WCU encapsulating the date and location of the censorship action, and the involved censor/s name, life/activity period and location/s (if known); there can be multiple CEs associated with the same WCU at different periods and locations, each CE might be associated with multiple censors involved in it, and, apparently, the same censor can take part in different CEs;
3. Manuscript storage event (SE) – representing the information on current storage (an institution and its geographic location) of an individual BU.
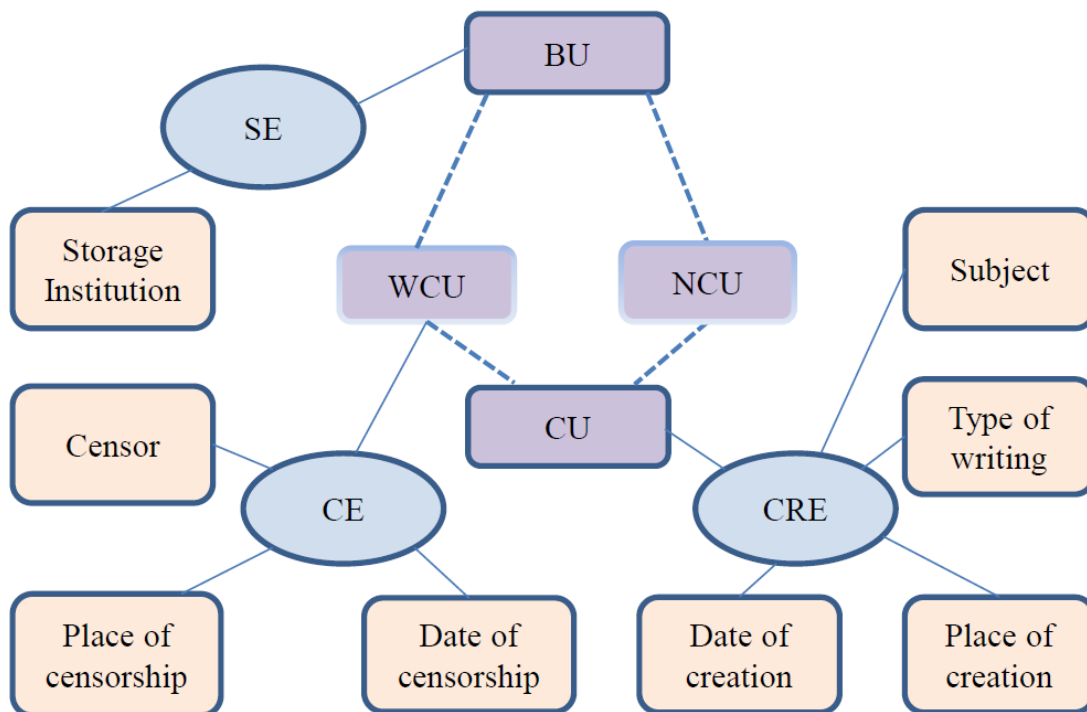
Figure 2. The study's event-based data model. The blue ellipses represent the event types (Storage Event - SE, Creation Event - CRE and Censorship Event - CE) and the rectangles are their main attributes. The hyphenated lines represent *part-of* relationships between the data entities, according to the dataset structure depicted in Figure 1 - CU is part-of a WCU or an NCU, and WCU and NCU are parts of a BU.

According to the attributes of these types of events (represented as rectangles in Figure 2), the manuscript dataset was divided into subsets, such as censored (WCUs) vs. non-censored (NCUs) manuscripts. These subsets' data were analysed separately in different dimensions, then the findings were compared and differences between the subsets were analysed.

Data analysis methodology

First, the corpus of manuscripts underwent a process of data cleaning and normalization. This process included decomposition of catalogue fields with mixed information into its defined atomic entities (e.g. people names, events, dates and places), standardization of a variety of date formats in the catalogue (e.g. specific Hebrew or Gregorian dates, ranges of years, centuries), and names of places (which appear in different languages and without the name of the country or geographic coordinates). Missing information was completed as well, using existing authority files and

ontologies, such as VIAF (https://viaf.org/) and Geonames (http://www.geonames.org/). As part of this place normalization process we also obtained the coordinates of all the mentioned places in the collection which were extracted from the abovementioned authority files. The data normalization process was performed semi-automatically. It was based on identifying lexical patterns for each ontological entity, and further manual correction and completion by two experts in manuscript studies. For instance, for extraction of the censorship event's location from the following string in the catalogue: censor signature: "LUGO 1611 CAMILLO JAGHEL", the matching pattern was <CENSOR * [PLACE] [DATE] [NAME]>, thus LUGO was retrieved and then looked up in the Geonames' "1000cities" list.

Next, we performed a systematic event-based analysis of the manuscript corpus. In addition to the descriptive statistical analysis we propose a new multi-dimensional methodology for a large-scale quantitative analysis of the historical manuscript data. The following types of variables were investigated for each of the dimensions:

> For the temporal dimension: we examined the date of CRE (a CU creation event) and the date(s) of CE(s) (a WCU's censorship events).

> For the spatial dimension: we used the place of CRE, the place(s) of CE(s), the current place of the corresponding BU storage, and the CU's type of script (the scribe's writing type is determined by its country of origin (Beit-Arié, 2018,) which points out to the place of writing of the manuscript – although it should be considered that sometimes writers moved to other countries while still keeping their original script type.

> For the actor dimension: an individual censor activity was examined in terms of the number of CEs in which the censor was involved.

> For the subject dimension: we analysed subjects assigned to a CU.

Based on the above dimensions and variables a three-phase analysis has been performed:

1. Descriptive statistics and distribution of censored manuscript collection (WCUs) in multiple dimensions;
2. Comparative analysis of the distribution of the two sub-collections of manuscripts, WCUs (censored manuscripts) vs. NCUs (non-censored manuscripts) in various dimensions.

Results

Descriptive statistics of the censored manuscript corpus

Only 5,185 CUs have censorship information recorded for them in the catalogue. These 5,185 CUs are grouped in 2,466 WCUs, while the other 118,921 non-censored CUs are grouped in 85,562 NCUs (as shown in Table 1). The total of 124,106 CUs are associated with 87,848 BUs. As explained before, there are cases when a single BU was (manually) "split" into a WCU and an NCU or cases when a BU was "split" into more than a single WCU (a total of 158 of both cases). Note however, that the vast majority of cases (87,690) are indeed cases where a BU consists of either a single WCU or a single NCU. Moreover, note that 79,415 BUs (90% of all the BUs in the study corpus) are unified codices, that is, BUs that comprise a single CU.

Entity	Number of items
BUs	87,848
WCUs	2,466
CUs in WCUs	5,185
NCUs	85,562
CUs in NCUs	118,921
CUs (total)	124,106

Table 1. Table 1: Counts of the study's main data entities

The 2,466 WCUs are associated with 3,914 instances of Censorship Events (CEs) (see Table 2). The number of CEs is

greater than the number of WCUs since in many cases, a WCU has been censored more than once, thus having more than one CE associated with it as discussed below. From Table 2 it can be noticed that in addition to the 3,190 CE instances where the censor responsible for the censorship was identified, in 724 cases, there is a clear notion of censorship, but no censor was documented in the manuscript's record.

Out of the 3,914 CEs, we filtered out 64 CEs: 56 CEs did not occur in Italy (and also didn't occur in the study period), and 8 Italian CEs were excluded since they occurred after 1775. The resulted dataset of Italian censorship events in the period of 1550-1775 consists, therefore, of 3,850 CEs (in 2,408 WCUs), 677 (17.6%) of which are un-named cases (see Tables 2 and 3). The earliest identified evidence of censorship in a manuscript in Italy is from 1554 and only two censorship events of Italian identified censorship have a recorded date after 1775 (in 1802 and in 1833), which fall outside our study period of 1550-1775.

	Named censorship			Un-named censorship			Total number of CEs		
	Italian	non-Italian	All	In Italy	Not in Italy	All	Italian	non-Italian	All
Dated (total)	2,518	8	2,526	391	36	427	2,909	44	2,953
Between 1550-1775	2,516	0	2,516	385	0	385	2,901	0	2,901
After 1775	2	8	10	6	36	42	8	44	52
With no date (total)	657	7	664	292	5	297	949	12	961
Dated between 1550-1775	657	0	657	292	0	292	949	0	949
After 1775	0	7	7	0	5	5	0	12	12
All periods (total)	3,175	15	3,190	683	41	724	3,858	56	3,914
Between 1550-1775 (total)	3,173	0	3,173	677	0	677	3,850	0	3,850
After 1775	2	15	17	6	41	47	8	56	64

Table 2. Table 2: Censorship events' (CEs) descriptive statistics (Italy and non-Italy cases) with named and un-named CE cases in the period of 1550-1775 and later.

From Table 3 we observe that out of the 677 CEs in Italy that have no censor name associated with them, in 513 cases a censorship signature was blurry and thus could not be identified. In 151 cases a clear sign of censorship text erasures is present in the manuscript, but without a signature, and in 13 cases the cataloguers identified self-censorship actions.

Type of un-named CE	# of cases in the entire dataset	# of cases that are Italian CEs in the range 1550-1775
Un-identified signature	560 (77.3%)	513 (75.8%)
Censorship without a signature	151 (20.9%)	151 (22.3%)
Internal self-censorship	13 (1.8%)	13 (1.9%)
Total	724	677

Table 3. Table 3: The distribution of different cases of un-named censorship.

A rather interesting finding from analyzing the censorship events of the 2,408 WCUs in our corpus is the number of re-occurring censorship actions, that is, censorship of manuscripts that have been already censored. In other words, these are the cases where the same WCU (a single or a bounded set of CUs) has been submitted for approval to the same or a different censor at various times and locations. Of the 2,408 WCUs, 992 WCUs (41%) were censored more than once (see the distribution in Table 4). For example, the WCU - PARIS BN Velins 908 (identified by the shelf number F 73304) has been censored on four different occasions by four different censors, according to this manuscript's record in the catalogue. It bears Luigi da Bologna's signature from 1599, Camillo Jagel's signature from 1623, Renato de Modena's from 1626 and Girolamo da Durazzano's from 1640. WCU - Moscow - Russian State Library, Ms. Guenzburg 147 (F 6831), on the other hand, records Girolamo da Durazzano's signature, twice: once from 1640 and then from 1641, which suggests that this manuscript was censored in two consecutive years by the same censor (this manuscript has an earlier third signature by Renato de Modena from 1626).

33

The fact that Hebrew printed books have undergone repeated censorship is well known as is mentioned, for example, in a study by Baruchson-Arbib and Prebor (2007). What we are showing in our study is that a) the repeated censorship occurs in handwritten books as well, and more importantly b) that its scope is quite substantial: in total, for the 2,408 censored manuscripts (the WCUs) in our study, compared to an expected 2,408 censorship events (CEs) (one for each), 3,850 distinct CEs are registered for these manuscripts, i.e. 1,442 additional, repeated CEs (see Table 4). This finding may indicate that the repeated censorship phenomenon cannot be explained only by sporadic actions by local censors, and that much more weight should be attributed to other, more general, explanations that were given to that phenomenon.

34

Several directions were suggested to explain this phenomenon. One explanation for this phenomenon is possibly that Church authorities feared that the Jews "corrected" their manuscripts after the censorship and hence re-checking is required. However, the substantial amount of renewed censorship must also be attributed to factors relating to the Church's conduct on one hand, and the conduct of local inquisition administrations, on the other hand. Firstly, the act of censorship of Hebrew texts did not receive an official and formal authorization. As a direct consequence, manuscripts which were approved by an official censor did not gain the official recognition of the Church. Also, the Church's policy regarding the censorship of Hebrew books was seldom changed as new Popes came into office and the rules by which censors should examine the text did also change and a stricter expurgation policy was adopted in the course of time [Raz-Krakotzkin 2007]. On a local inquisition basis, censors apparently doubted their predecessors' work for negligence or even deliberate eye-closing facing the monetary benefits from the Jews who paid for their job.

35

The findings in our study show that re-occurring censorship is a phenomenon that spread across time and across places. There are cases where repetitive censorship took place in the same place, while in other cases, the repeating censorship occurred in different place(s), and usually by different censor(s). Table 4 summarizes the repetitive censorship statistics. More detailed investigation, however, which will be dedicated to this phenomenon, could produce more insight. It may identify the censors who are responsible for repeated censorship more than others, it may find whether the repeated censorship happened on specific years or places, it might supply some characteristic as to the time elapsed between CEs of the same manuscript (e.g. is re-occurring censorship more common after a long time since previous censorship or maybe shorter elapsed time is more common).

36

# of CEs on the same WCU	Total # of WCUs of this case	# of WCUs that have been censored more than once	Total # of CEs on the same WCU	# of repeating CEs
1	1,416	0	1,416	0
2	663	663	1,326	663
3	238	238	714	476
4	67	67	268	201
5	20	20	100	80
6	2	2	12	10
7	2	2	14	12
Total	2,408	992	3,850	1,442

Table 4. Table 4: Number of CEs per WCU

By lexical processing the censorship recording in the catalogue, we were able to automatically identify the cases where more than one censor was responsible for a censorship mission. Part of these cases were identified by having the censors' signatures on the same page with a single mention of the year of the censorship (e.g. Alessandro Scipione 1597 (275b); Dominico Irosolomitano (275b)) like in the example of the campaign of expurgation by a censorship committee that was in Mantua in 1595-1597 (see below). Other cases were identified by their unique form of the censor's approval text which explicitly mentions the multiple censors as in the following example: "Subsignavi ego Guidus menotius ... Jac Geraldini ... 28 Marti 1556". [4]

Temporal analysis of the dataset

Censorship Events Dates

We first looked at the years when the censorship actually took place and have used for this analysis the subset of the 2,901 precisely dated CEs out of the whole set of 3,850 CEs in the database. We have plotted a 5-year bin-based histogram of 2,901 dated CEs (see Figure 3). The immediate observation from inspecting the histogram in Figure 3 is that although the recorded range of manuscript censorship in Italy is 222 years long (1554-1775), the majority of the censorship events took place during a much shorter period. The first substantial narrower period corresponds to the years 1589-1602 (with 36.0% of the total dated CEs) and the second substantial period corresponds to the years 1607-1629 (with 49.1% of the total dated CEs).
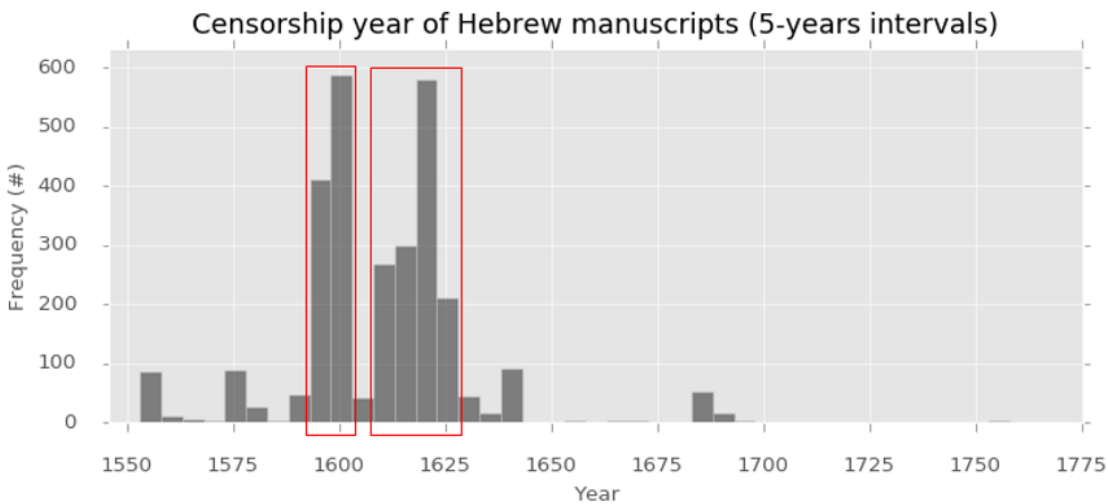

Figure 3. Number of the dated CEs (1550-1775) (2,901 cases).

This wave-like distribution of CEs along the study time period supports the historical knowledge we have about censorship activity in Italy. Much of the censorship was performed in accordance with the instructions of the church and local leaders and was the result of the activity of a number of censors who were active in certain years. The first big wave corresponds to the years 1589-1602 (with 36% of the total dated CEs). This reflects the big campaign of expurgation by a censorship committee that was in Mantua in 1595-1597. On August, 27th the bishop of Mantua appointed a committee of censors of the Jewish books in Mantua. Domenico Irosolimitano, determined by our analysis as the most dominant censor in the corpus, was the head of the Censorship Committee in Mantuain 1595-1597, and the author of *Sefer Ha-Ziquq* (*The Book of Purification*) [Baruchson-Arbib 2007]. Sefer Ha-Ziquq is an index of forbidden Hebrew books specifying the corrections required for each of them. This censorship committee was active in the years 1595-1597 and censored hundreds of books and manuscripts [Baruchson-Arbib 1993] [Baruchson-Arbib 2007]. The second wave in 1607-1629 (with 49.2% of the total dated CEs) corresponds to the activity of two prolific censors: Camillo Jagel da Correggio and Giovanni Domenico Carretto. Camillo Jagel was active from 1611-1629 [Popper 1899]. He began his activity as a censor in 1611 in Lugo (Segra, 1977-1980), which we quantitatively verified by detecting 26 censorship signatures with his signature with the year 1611 and with the place named Lugo. There are an additional 50 censorship cases with Camillo Jagel's signature dated to 1611, but lacking any place reference. We inferred that these cases are based in Lugo as well. In 1613-1614 Jagel was in charge of expurgating Hebrew books in Modena and according to Francesconi's (2012) findings he also wrote censorship rules, lists of books to be censored and lists of passages to be erased [Francesconi 2012]. After that, in 1619-1620, Jagel was active in Ancona [Popper 1899].

Giovanni Domenico Carretto served as a censor in 1607-1628, mainly in 1617-1619 when he was appointed to the position of censor of Mantua [Popper 1899]. These two waves can be decomposed into a few micro-waves as shown by our findings (Figure 11).

Regarding the research question of the periods of censorship activity, the findings discussed above show that the censorship activity is not uniformly spread over time. There are periods of increased activity and periods of a "calmer" nature. Whereas there is a direct correlation between the first wave of censorship (mainly 1595-1597 in Mantua) to the censorship committee appointed to at that same time in Mantua, there is no direct historical event that can be attributed to the second censorship wave. It is, however, known that during different Popes' ruling periods, an inclination to force a stricter censorship policy or, on the contrary, to mitigate such activity was expressed [Raz-Krakotzkin 2007]. In a broader perspective, it will be interesting to compare the trend of Hebrew literature censorship activity to the Church censorship of prohibited books in general. It might be discovered that "lower" activity periods are associated both with Jewish and general non-Jewish texts alike and can be attributed to different Papal periods.

Creation Events' Dates

Overall, Hebrew manuscript creation during the last millennium was influenced by many factors. Examination of the distribution of the entire dated CRE set in Italy in Figure 4 (the grey bars), shows major changes in the rate of manuscripts creation over the years. This dataset comprises 3,213 CUs with a known creation place – Italy, and with a known precise or approximate creation time. In cases when only an approximate date (range of dates rather than a specific year) of the CRE was provided in the catalogue, the CRE was assigned a random year within the given range.
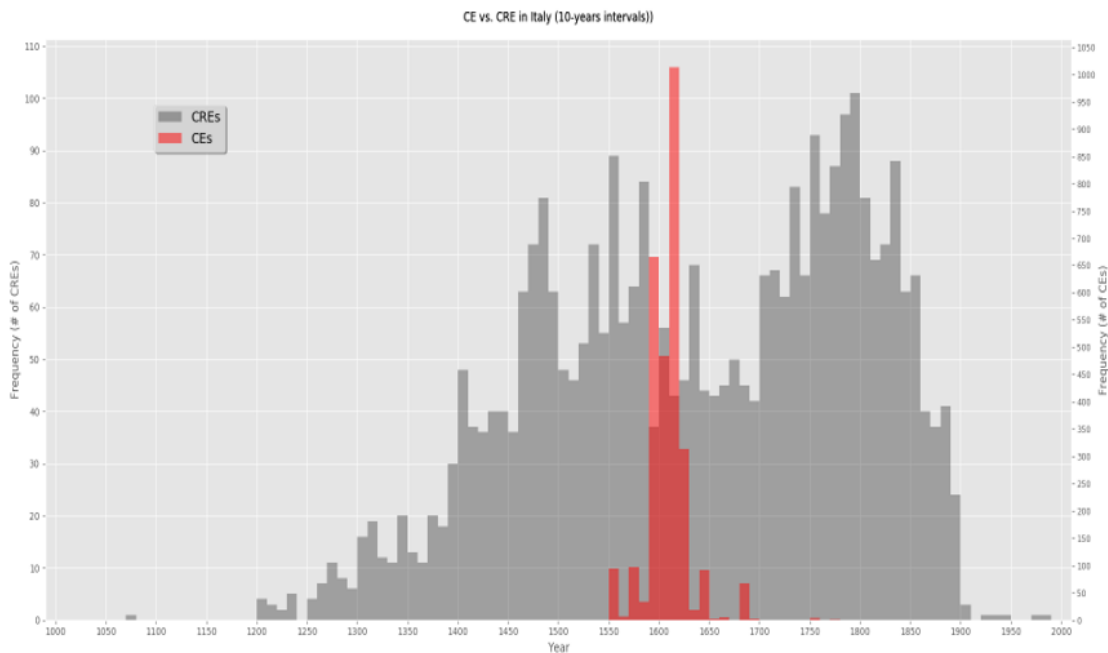
Figure 4. Counts of newly created manuscripts in Italy (grey bars) and counts of manuscript censorship actions (red bars) (1200-1775) – dated only CREs and CEs.

There is a sharp rise in the rate of manuscript creation in Italy in the 14th century compared to earlier times. Then, during the 15th century the manuscript creation rate is doubled. But in the 16th century we see lower rates compared to the previous century with a growing rate during the 1st half of the 16th century and a rather sharp decline in the 2nd half of the century continuing in the 17th century with a lowest rate point around 1700. From about that point, manuscript creation rates start to rise again with a pick of manuscript creation rate around 1800. From the middle of the 19th century we see an expected decline in creating new manuscripts.

43

In order to investigate a possible relationship between the different periods of high and low manuscript creation rates and the periods of censorship activities, we plotted the censorship activity histogram on top of the manuscript creation event histogram on the same time axis (see the red bars in Figure 4). It can be noticed that the first declining period in CREs that starts around 1475 is not correlated with censorship activity as this have started only 75 years later, this decline is probably attributable to the invention of the printing press.

44

However, the second manuscript creation declining period which starts in 1590, right after the beginning of substantial censorship activity, might be influenced by substantial censorship activity in the closely preceding and following time periods (Figure 4). During the period of high censorship activity (the 41-year period between 1589 and 1629), Hebrew manuscript creation level in Italy was relatively lower compared to all other times in the period of 1375-1775.

45

Spatial analysis of the dataset

Censorship Events Locations

When signing the censorship action, the censors, in most cases, added the date of the signature (mostly the year and in few cases the full date), but they very rarely (only in 71 cases, less than 2%) specified the place where this censorship had been performed and signed.

46

In order to complete the missing information on CEs locations we used ontological relationships between events, censors, time periods and locations and the information from the Pooper's list of Italian censors (Popper, 1889). In this list, Pooper recorded all the censors known to him in alphabetical order. For each censor he noted the years and places of his activity. For example, the censor Alessandro Scipione was active in Mantua in the years 1589-1590 and 1593-1598; Camillo Jagel was in Lugo in 1611, in Modena-Reggio in 1613-1614 and in Ancona in 1619-1620. We also used a

47

newer study on the censor Domenico Irosolimitano that gave us more information on his activity as a censor [Prebor 2007]. Thus, given a CE's censor name and date we could infer its location from the corresponding location in the Pooper's list. As a result, we obtained places of censorship for 2,077 (53.9%) out of all the CEs in the corpus. Eight of the inferred places did not appear in the original list of 12 places that have been recorded in the catalogue: Ferrara, Milano, Monreale, Turin, Venice, Modena-Reggio, Ancona and the Papal States. In 1452–1796, 1814–1859 Modena and Reggio were part of the Duchy of Modena and Reggio. The place name "Papal States", refers to a whole region which was at that time under the Pope's rule. In different periods this region included different parts of Northern Italy. Figure 5 displays the CEs geographic distribution on a map.

Figure 5. Places of censorship events (explicitly mentioned in the catalogue and inferred) on Italy's map (2,077 CEs).

As can be seen in Figure 5 the distribution of places of censorship events are spread out in various places in the Italian peninsula, especially in the north. This uneven distribution, which as stated above was dependant on the individual censor involved, also supports the historical knowledge we have about censorship activity in Italy. Much of the censorship was performed in accordance with the instructions of the church and local leaders and was the result of the activity of a number of censors who were active in certain places. The two places with the highest censorship events are Mantua and Modena.

Script type analysis

Another aspect of spatial analysis can be learned from the script types of the censored manuscripts. Hebrew manuscripts have been written in many scripts that varied from region to region, each with its typical script type. The Sephardic script, for example, is the typical script used by Sephardic scribes who worked within the Jewish community

in the Iberian Peninsula. Since censored manuscripts were in the possession of the Jewish community in Italy, analyzing the distribution of their script types may provide new (quantitatively) perspectives of the extent of the transition of manuscripts from outside regions into Italy as well as on the immigration of non-Italian Jewish scribes who continued to write with the script type of their homeland and education [Beit-Arié 2018].

Five main branches of script types were proposed in the literature: Eastern script (which includes the Yemenite script and Persian script subtypes); Sephardic script; Ashkenazic script; Italian script and Byzantine script [Beit-Arié 2018] [Richler 1990]. The catalogue named more than 70 distinct Hebrew script types. As a first step of the analysis, we have compiled a list of 17 script type classes with an additional *Miscellaneous* category for the rest of the scripts: Ashkenazic Script, Italian Script, Eastern Script, Sephardic Script, Yemenite Script, Western Script, Byzantine Script, Karaite Script, Persian Script, Provencal Script, Samaritan Script, Ashkenazic-Italian Script, Sephardic-Eastern Script, Sephardic-Provencal Script, Sephardic-Italian Script, Sephardic-Western Script, Typewriter and Miscellaneous. Next, we classified all the manuscripts into seven unified script classes (*ustypes* for short). For instance, the Sephardic unified script class was assigned to all manuscripts with variants of the Sephardic Script (e.g. Sephardic-Eastern script, Sephardic-Western script and Sephardic- Provencal script).

We then examined the distribution of the seven ustypes amongst 5,015 censored manuscripts and compared it to their distribution among a set of 72,233 non-censored manuscripts (overall, 77,248 manuscripts were created before 1700 as almost no censorship took place afterwards). The result of this comparison is depicted in Figure 6.[5]
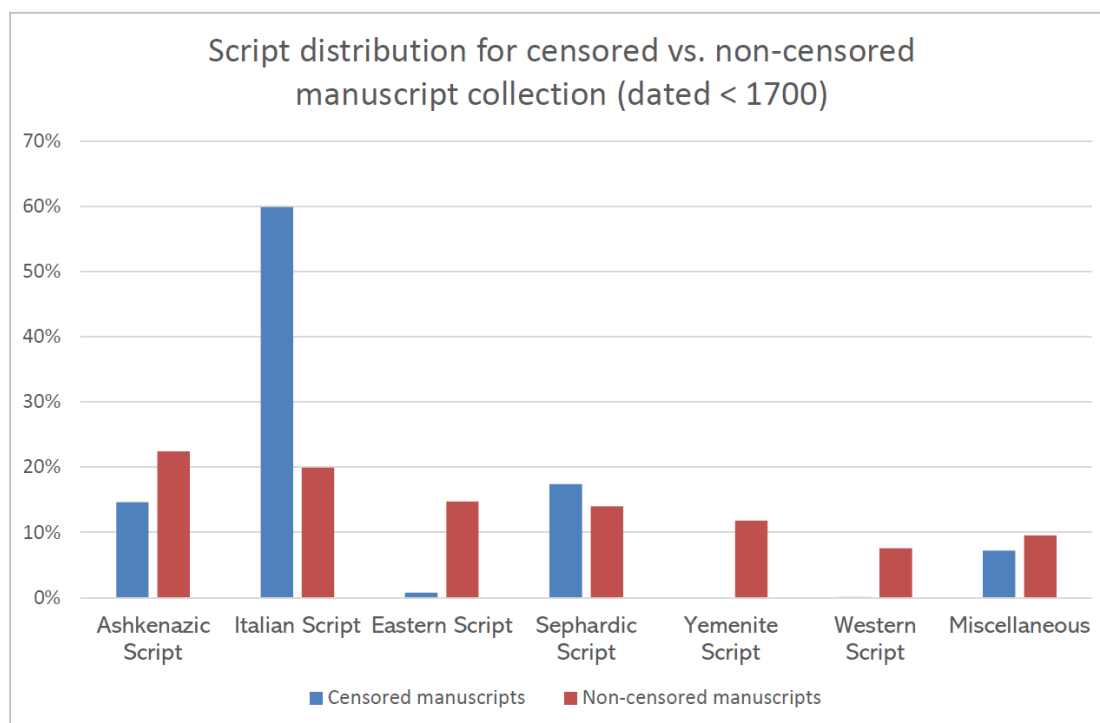

Figure 6. Unified script types distribution for 77,248 manuscripts created before 1700).

It can be immediately observed that the Italian script type is the dominant one amongst the censored manuscripts (59.9%) compared to this script occurrence percentage (only 19.9%) amongst the non-censored manuscript created during the same time period. There was a significant difference in script type distribution between the censored and non-censored corpora, $\chi2(6)=118.21$, $p<0.0001$. This implies that a major part of the manuscripts in the possession of Italian Jews at that time were probably written by this community, namely, written in Hebrew with an Italian script type. This finding also strengthens the results of Beit-Arié that the Italian script was the most frequent one (55%) used in the manuscripts written in Italy before 1540 [Beit-Arié 2018]. These results can be explained by Italy being the centre of the Catholic Church on the one hand, as well as being a prominent centre of rich Jewish life and culture, on the other hand.

The presence of manuscripts written in other script types can be explained by manuscripts brought/bought by Italian Jews at different times as well by the use of these writing scripts by newcomer Jews. We show that 17.4% of the censored manuscripts were written in the Sephardic script which is an indication of the (mainly forced) immigration of Sephardic Jews into this region. This script type is found in only 14.0% of the non-censored scripts which strengthen the indication of a major transition of Sephardic Jews into Italy, compared to all other regions. Another well recognized fact confirmed from this comparative analysis is regarding Yemenite manuscripts. From Figure 6, one can infer that almost no manuscript written in Yemen arrived in Italy before 1700, as there are no censored manuscripts found that were written in the Yemenite script type. Thus, through analysis of script types we can indirectly identify migrations (or the lack of migrations) of Jews (with scribes among them) and partly answer research question number 1c (other parts of this paper discuss other types of geographic migrations, those of the manuscripts themselves, this time) see also in: Prebor, Zhitomirsky-Geffet & Miller (2019).

Analysis of the locations of manuscripts' storage events

Another indication of the migration of manuscripts (albeit, through a later time span) is their current location. The original Hebrew manuscripts that are documented in the catalogue of the Department of Manuscripts and the Institute of Microfilmed Hebrew Manuscripts in the National Library of Israel belong primarily to public collections held by many public institutions (mainly academic or municipal/governmental libraries) or to private collections all over the world. As part of our research we wanted to shed light on the distribution of censored manuscripts in these institutions. We found that the Hebrew manuscript corpus is spread across 1,133 different institutions around the globe (see Figure 7). 37.6% of them are stored in Israel's libraries (the majority are in the National Library of Israel in Jerusalem), 16.3% - in the United States and 15.7% are located in Russia.
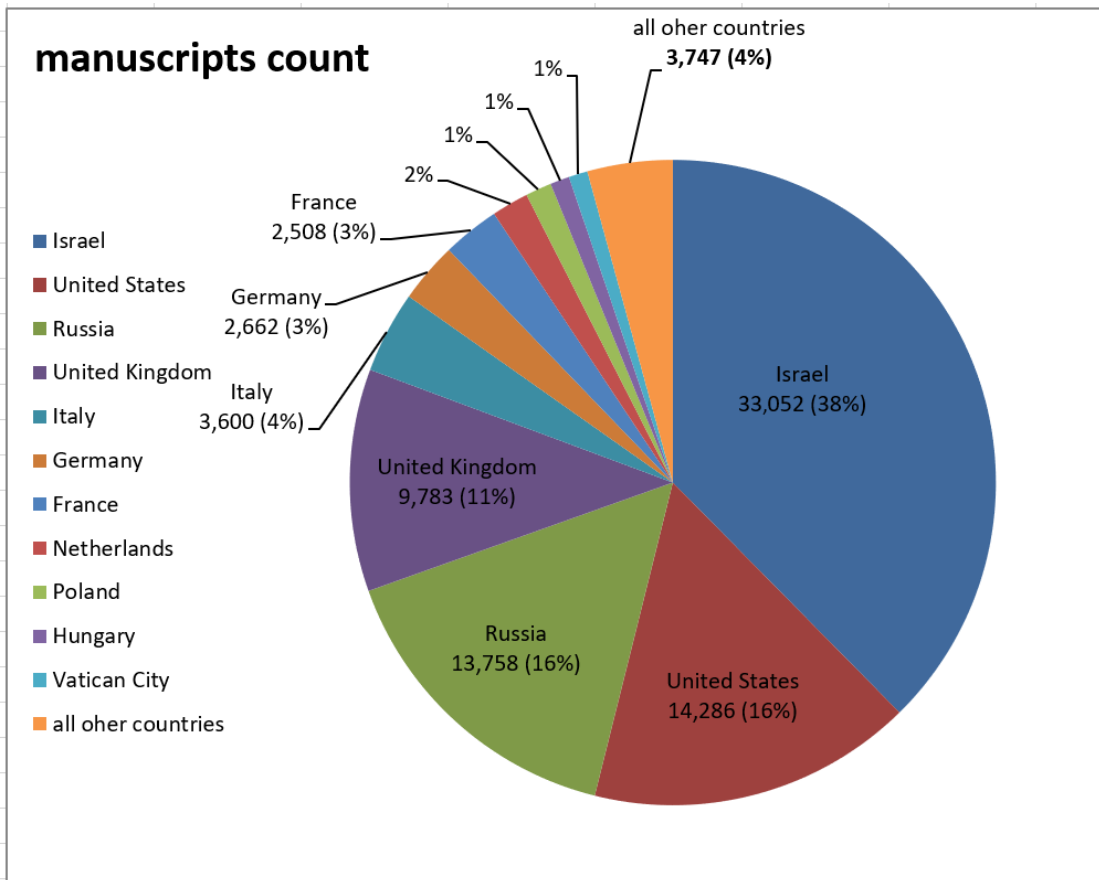

Figure 7. Number of manuscripts (BUs) in countries' libraries.

However, the picture for the censored manuscripts is quite different as can be viewed in Figure 8.
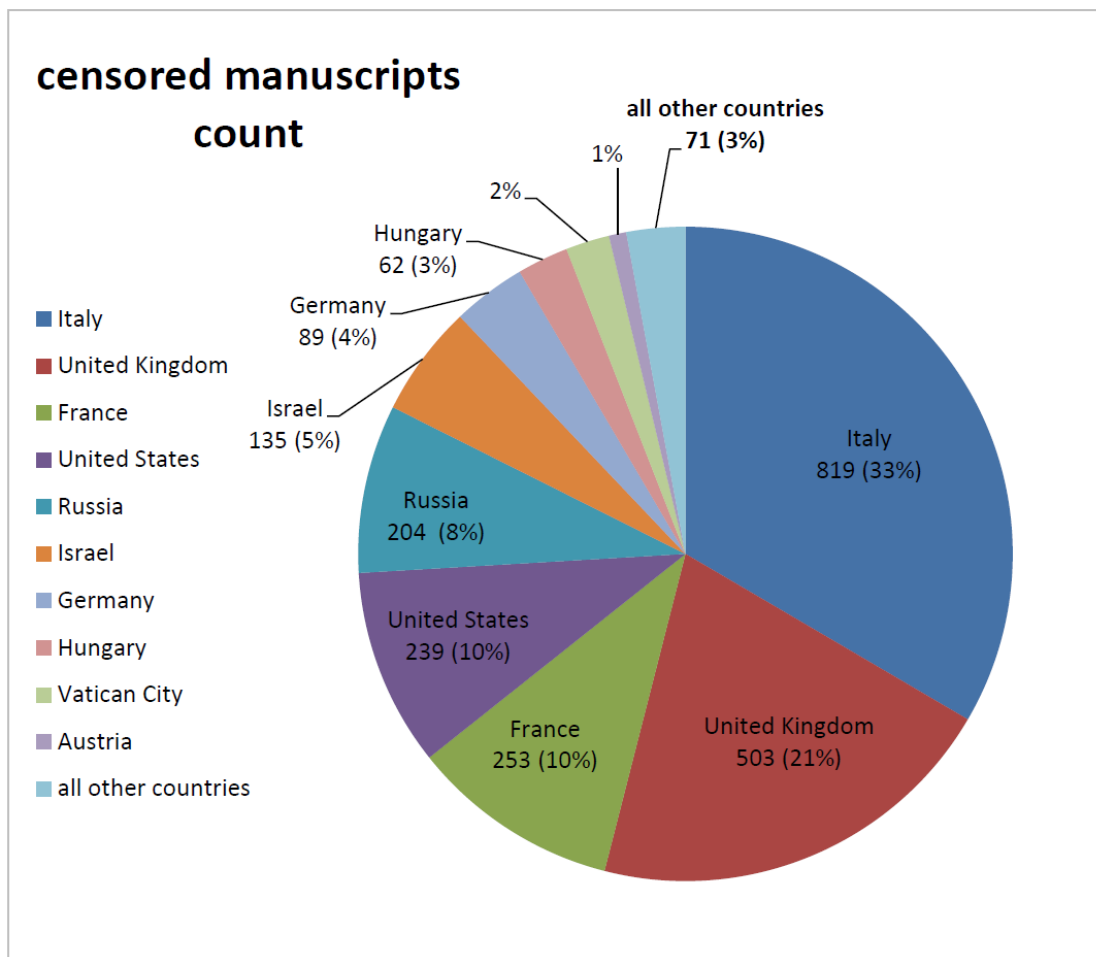
Figure 8. Number of censored manuscripts (BUs) in countries' libraries for all 2,449 censored BUs (including those censored after 1775 in non-Italian regions).

The largest portion of the censored manuscripts is still stored in Italy and the Vatican City: 35.6%. This portion is much higher than the portion of Hebrew manuscripts in general found in Italy and Vatican City out of all Hebrew manuscripts (5%). This is another indication of the censored manuscripts being primarily written in Italy, as well as an indication of a relatively high retention rate (i.e. archiving) of material of this nature in a close proximity to where it was originally written/possessed.

56

Yet, the majority of censored manuscripts were transferred over the centuries to institutions outside Italy as shown in Figure 8. Following Italy are the United Kingdom (20.5%) and France (10.3%). Israel is only sixth with a 5.5% share.

57

Subject-related analysis of the manuscripts

Catalogue records include subjects from the Library of Congress Subject Headings (LCSH) (http://id.loc.gov/authorities/subjects.html). Each manuscript record was assigned multiple subjects which were sometimes too specific.

58

In order to determine what were the main subjects of the censored manuscripts and whether there is a difference in the distribution of censored vs. non-censored manuscripts among subjects, we created a list of 13 general topics. The list of 13 subjects that generalize and group together some hundreds of subjects from the catalogue are: Popular Beliefs and Magic, Ancient Rabbinic Biblical Exegesis, Non-Legalistic Exegesis and Homilies, Science, Philosophy and Ethics, History, Mishna and Talmud Commentaries, Jewish Legal Texts, Liturgy, Bible and Commentaries, Kabbalah and Mysticism, Letters and Belles Lettres, Grammar. The Miscellaneous category includes Karaite literature, Jewish apocrypha, list of names, polemic, Gematria. For each censored manuscript we asked an expert researcher to assign subjects from the above list (thus, a single manuscript may have been assigned more than one subject).

59

In order to compare the distribution of the subjects in the censored manuscript corpus to that in the non-censored manuscript corpus, a random sample of 857 non-censored manuscripts written before 1700 was subject-annotated by our expert as well (with subjects from the above list only). For this analysis, we used a subset of 5,046 censored CUs that were created before 1700. The results are depicted in Figure 9.
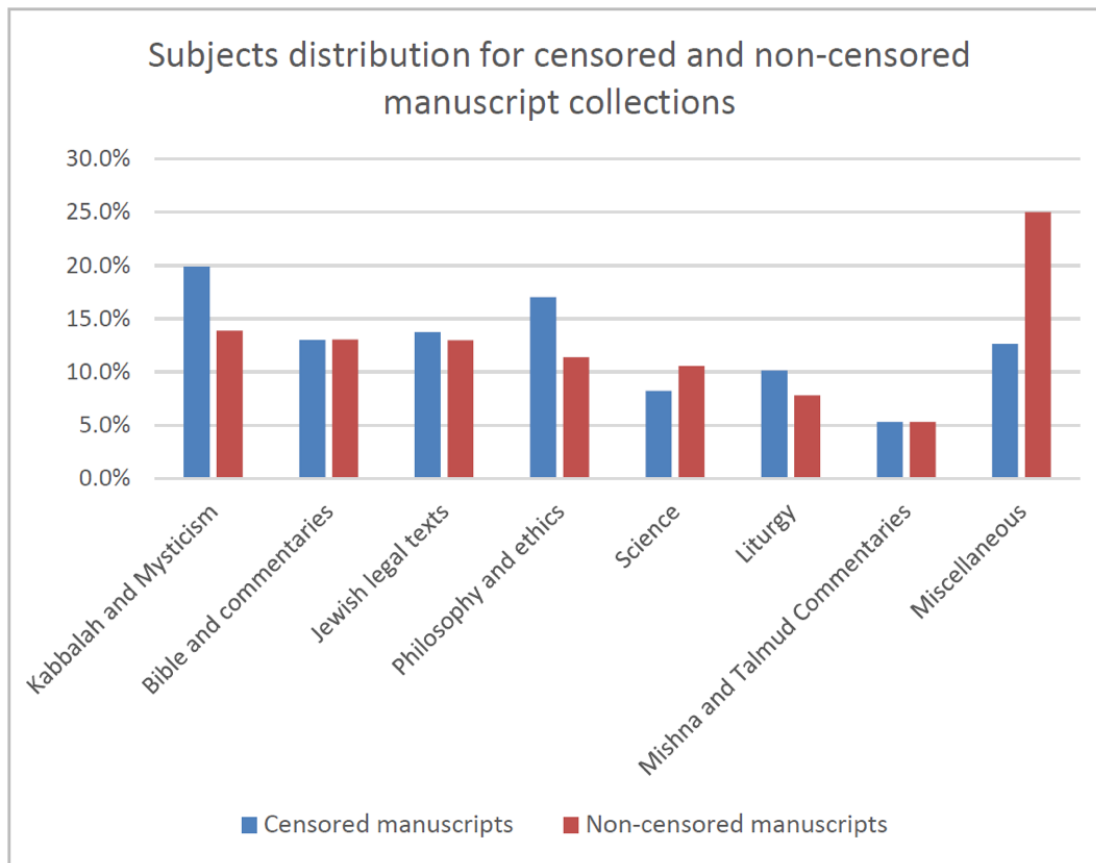
Figure 9. Top subject distribution in the censored and non-censored manuscript sample created before 1700, for 5,046 censored CUs and 857 non-censored CUs.

We found significant differences in subject distribution for censored vs. non-censored manuscripts, $\chi^2(7)=14.525$, $p<0.05$. The three major differences are in the *Kabbalah and Mysticism*, *Philosophy and Ethics*, and *Liturgy* that were censored significantly more than other subjects. Subjects with percentages of 5% and lower are included in Miscellaneous in Figure 9.

The subject analysis of the censored manuscripts might shed light on the goals of the censorship activity. We found that the most censored subject was *Kabbalah and Mysticism*, 20% of the censored manuscripts belong to this subject (compared to only 14% of the non-censored manuscripts). This is not surprising since in the 16th century Kabbalah study became more widespread in Jewish culture and a debate arose over the printing of Kabbalistic literature. In this period, Kabbalah also attracted the attention of Christian scholars. Several Kabbalistic works were translated into Latin, and the interest and preoccupation with Kabbalah became wider. The main reason why Christian scholars expressed interest in Kabbalah is the similarity they saw in some of the foundations of the Kabbalah and the principles of the Christian faith. It is impossible to ignore the historical context of the debate over Kabbalah literature in the Jewish world in the period of the highest Catholic censorship activity when each body acted out of its own motivations and goals [Raz-Krakotzkin 2007] [Weiss 2016]. The second most censored subject is Philosophy and Ethics (17% of the censored corpus compared to only 11.4% in the non-censored corpus). The books of philosophy that deal with the core of faith and content that can harm Christian faith, such as Rabbi Joseph Albo's *Sefer ha-Ikkarim* (*The Book of Principles*), which has the most comprehensive censorship requirements in Sefer Ha-Ziquq [Prebor 2003]. The third subject with a significant difference between the percentage of censored and non-censored manuscripts is Liturgy. The ecclesiastical

supervision of liturgical books, especially in the Ashkenazi *mahzorim* (festival prayer books), was massive because they contained anti-Christian piyyutim (liturgical poems) and many expressions condemning the nations and the Christian faith [Benayahu 1971] [Raz-Krakotzkin 2007]. The prayer books and the *mahzorim* were the most popular category in the libraries of the Jews of Mantua [Baruchson-Arbib 1993]. The percentages of the top-censored subjects partially correlate with the numbers of the required corrections for various subjects specified in *Sefer Ha-Ziquq*. The highest number of corrections (24 on average) in *Sefer Ha-Ziquq* is required for the Liturgy subject and Kabballah is ranked third (with 16 corrections required on average) following the ancient rabbinic biblical exegesis and non-legalistic exegesis subject (Prebor, 2003). In contrast, we found more "neutral" subjects that did not catch the censors' attention, subjects with "non-problematic" content. This includes Karaite literature, apocrypha, lists of names, polemics, gematria, letters and belles-lettres, science, popular beliefs and magic, and history.

Analysis of the individual censor activity

The National Library cataloguers have identified 65 censors that worked in Italy between 1554 and 1775. Some of the censors were very active with hundreds of signed censorship actions, while others only appeared in 1, 2 or 3 CEs (censorship events). Figure 10 displays the counts of CEs for the top-20 censors (out of the 65) that were involved in at least 15 CEs.
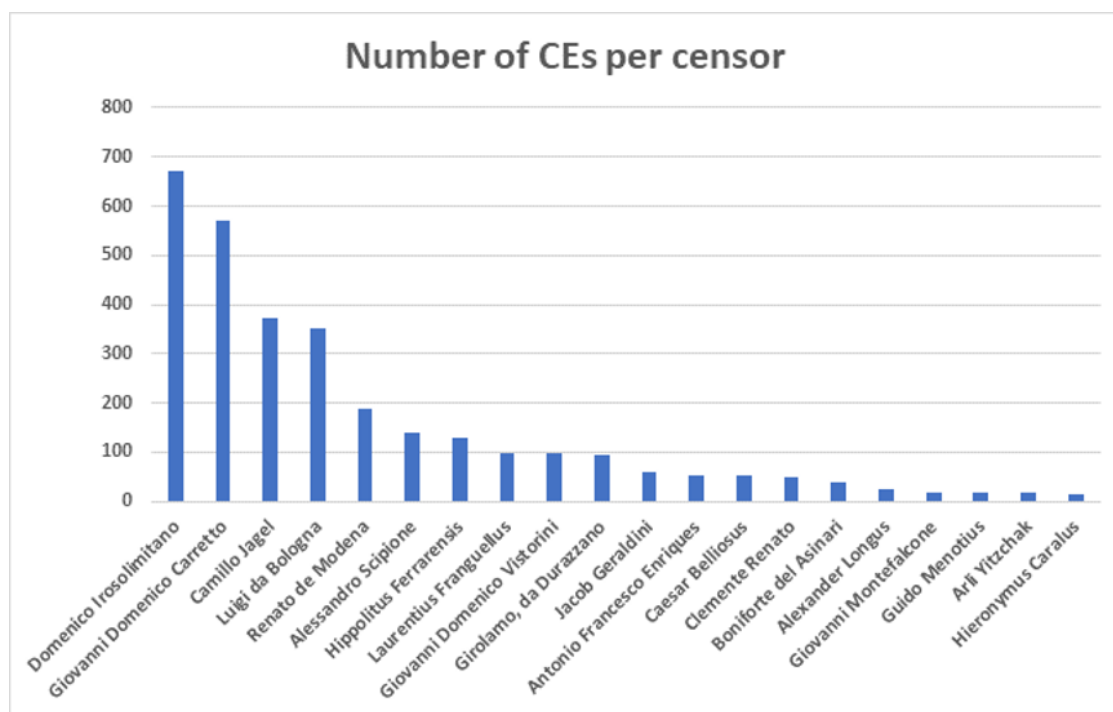
Figure 10. Number of CEs per censor (only for censors involved in at least 15 CEs).
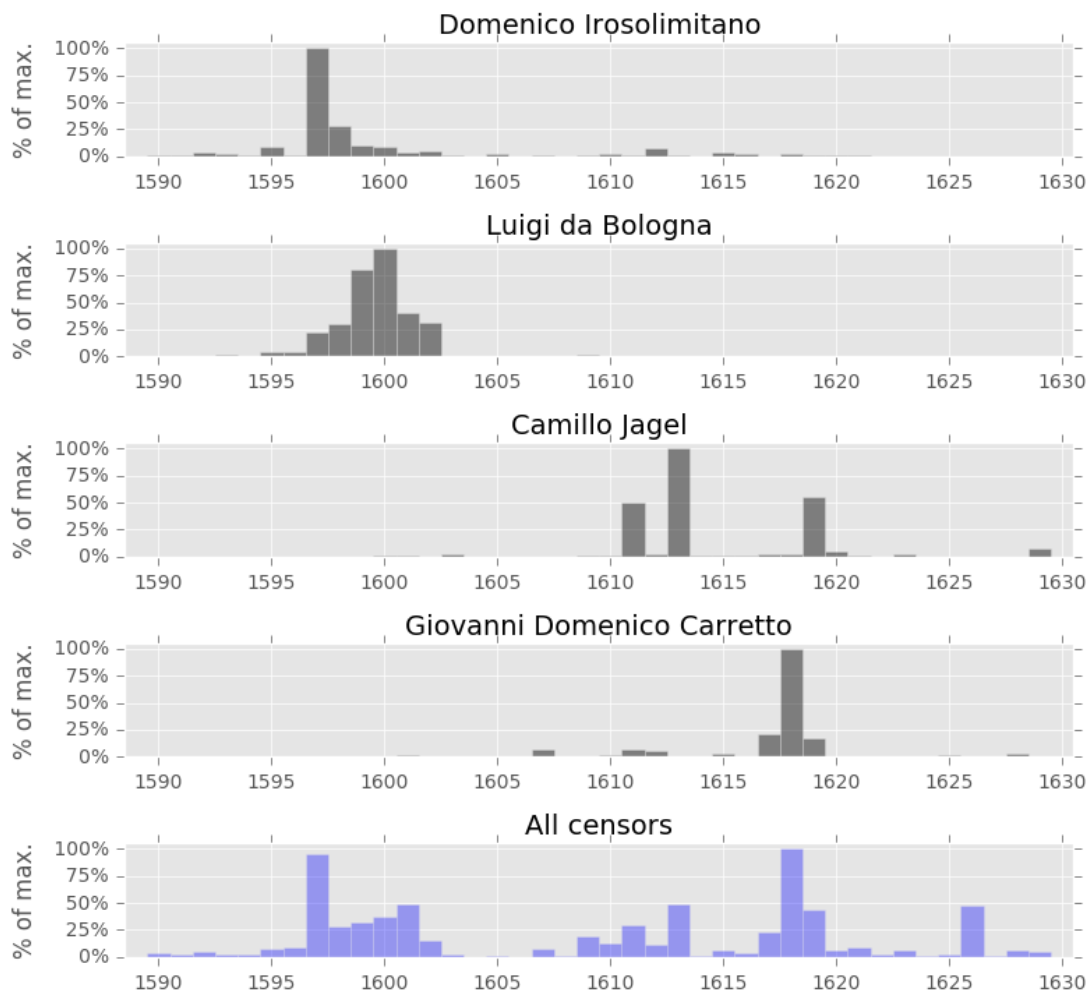
Figure 11. Distribution of CEs for the top-4 censors (and all censors) 1590-1630.

As shown in Figure 10 the most active censor was Domenico Irosolimitano. Figure 11 displays his (and three additional most dominant censors') activity on a timeline. Domenico Irosolimitano acted mainly from 1595 to 1602 (the pick year is 1597) and then again in 1612. His signature is found in 671 manuscripts (about 19% of all CEs) although only 378 (56.3%) of them has a date of censorship. Domenico Irosolimitano was not the first censor to act in Italy. He was preceded by several censors (Caesar Belliosus, Jacob Geraldini, Guido Menotius, Hieronymus Caralus and Laurentius Franguellus) that worked between 1554 and 1589, but each of these censors is responsible to only few dozens of CEs.

The censors' frames in Figure 11 are presented in chronological order of their activity. Thus, the next censor (chronologically) is Luigi da Bologna who acted mainly between 1597 and 1602. Unlike Domenico Irosolimitano who has quite an exceptional pick year, the CEs of Luigi da Bologna are more evenly spread along its active years and are virtually all dated (326 out of 351). The third censor, Camillo Jagel, acted mainly in three years: 1611, 1613 (the pick year) and 1619. Camillo Jagel was responsible for 371 CEs (356 of them are dated). The fourth censor in Figure 11, Giovanni Domenico Carretto, acted mainly in 1617-1619 (the pick year - 1618). He is the second most active censor after Domenico Irosolimitano with 572 CEs (552 of them are dated).

The last frame in Figure 11 shows the overall censorship activity distribution over time. As can be observed, the CEs are not evenly distributed along the timeline. There were four main micro-waves of censorship activity in 1595-1602, 1608-1613, 1617-1619 and around 1626. These micro-waves generally match the activity periods of the four most dominant censors discussed above.

To complete the censor-related CEs analysis we performed a detailed breakdown of the CEs over time for the top eight

places (including "the Papal States") that have at least 30 CEs. The result is shown in Figure 12 (these places refer to 96.7% of the CEs with known or inferred locations: 2,009 out of 2,077). The figure also demonstrates the relative part of CEs that are associated with each of the most prominent censors (Domenico Irosolimitano, Camillo Jagel, Luigi da Bologna and Giovanni Domenico Carretto, represented with different colors).
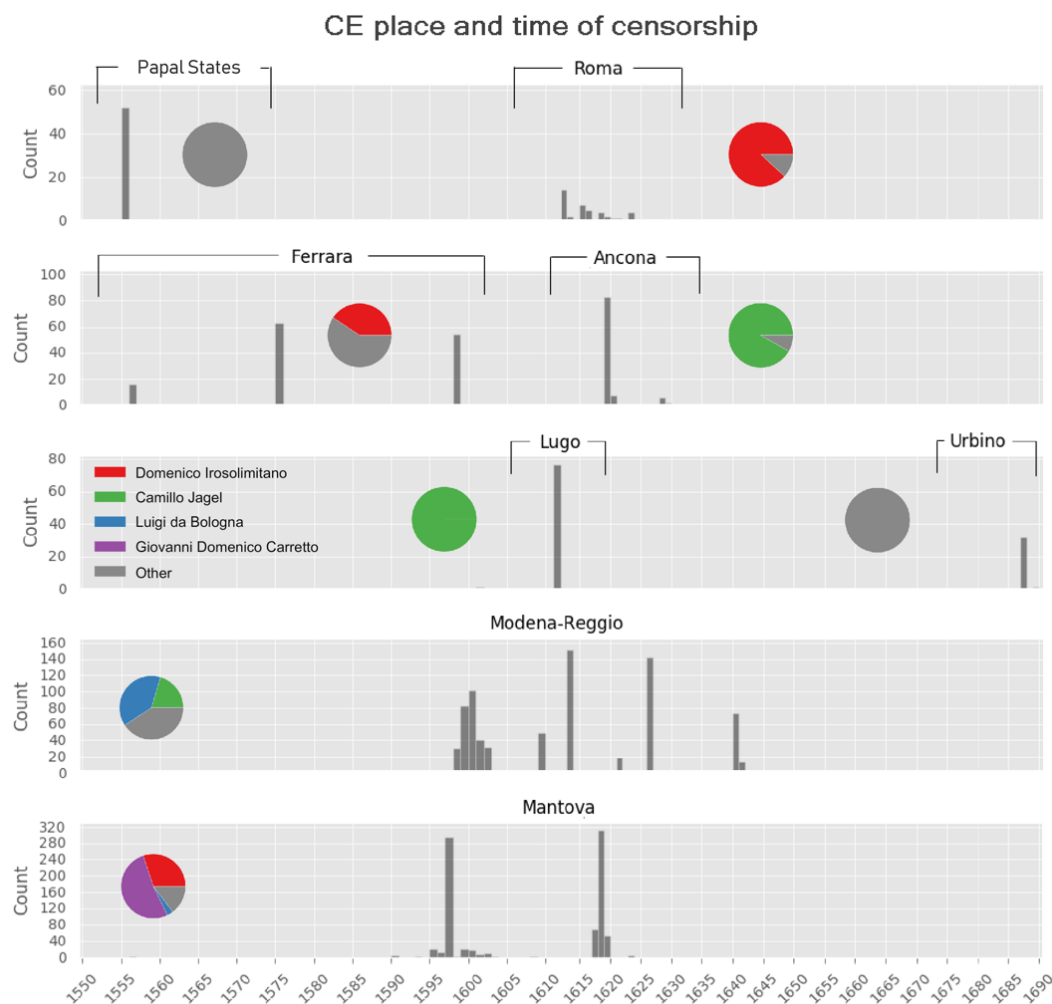

Figure 12. Distribution of CEs in top eight places (years: 1550-1690) (2,009 CEs), the colors correspond to the most dominant censors involved in each place.

From Figure 12 it can be observed that in some places the censorship took place during a very short period (sometimes of one or two years), e.g. in Lugo, Ancona, Urbino and the Papal States. Censors' movements in time and geographic locations can be traced as well, e.g. Domenico Irosolimitano (denoted by the red color in the chart circles) moved from Mantua to Ferrara and then to Rome, and Camillo Jagel (marked by the green color) moved from Modena-Reggio to Lugo and later to Ancona. Also, it can be noticed that Rome's CEs are mainly attributed to Domenico Irosolimitano, who was also a dominant censor in Ferrara and Mantua. Camillo Jagel is responsible for virtually all censorship in Lugo and Ancona. Modena-Reggio CEs are mainly executed by Luigi da Bologna with some noticeable contribution of Camillo Jagel (the late CEs in the 1640's are Girolamo da Durazzano's). Finally, Mantua's censorship was performed by Domenico Irosolimitano (in the early period until 1603), while more than half of the CEs in 1617-1619 are attributed to Giovanni Domenico Carretto. The CEs in the Papal States were almost entirely signed by Caesar Belliosus and in Urbino all CEs were signed by Antonio Francesco Enriques (not in the list of the top four censors).

Conclusion

This study presents the first large-scale quantitative analysis of the corpus of censored historical Hebrew manuscripts

that survived through the centuries, particularly focusing on Catholic Church censorship in Italy. The primarily methodological contributions of this study include the following:

1. Definition of the generic data model of the Manuscript catalogue records.
2. A new multi-dimensional ontology-based approach for historical data analysis from the geographic, temporal, actor- and subject-based perspectives.
3. Utilizing previous studies results and ontological relationships for automated rule-based inference of over 50% of missing locations of censorship events.
4. Comparative analysis of censored vs. non-censored manuscript corpora.
5. Qualitative examination of the quantitative results in light of the extant literature and their historical and cultural context.

We found that the two centres of the most massive censorship in the history of Hebrew manuscripts were Mantua with 832 CEs and Modena-Reggio with 741 CEs. Also, rule-based inference helped discover eight censorship locations which did not appear in the catalogue. The time periods of the most intensive censorship activity in Italy were 1589-1602 and 1607-1629. The most active censors in Italy were Domenico Irosolimitano (671 CEs) and Camillo Jagel (572 CEs).

The distribution of the censored manuscript corpus differed from that of the non-censored corpus in several dimensions, e.g. by subjects, script types and storage locations. Furthermore, the following trends in censorship activity were discovered by our analysis:

1. A few big and micro-waves of censorship during the time period of the study;
2. Geographic mobility of manuscripts, their creators and censors over time (the script types' distribution reflects the dominance of the Italian script along with a significant transition of scribes from Spain);
3. Current locations of the manuscripts (interestingly, as opposed to the entire corpus, censored manuscripts were less geographically mobile and are still mostly located in Italian libraries);
4. Censorship influence on the Jewish literature in terms of applying more restrains on the "problematic subjects" (e.g. Kabbalah and Philosophy) and indirectly reducing new manuscript creation (our findings show a substantial retention in writing new manuscript during the pick periods of censorship activity).
5. Repetitive censorship as a phenomenon (41% of the censored manuscript in the corpus were censored more than once, while only 2% of the censored manuscripts were collaboratively censored by two or more censors at the same time and location).

The above results show the power of a large-scale statistical analysis of the well-structured ontological data of the manuscripts. Firstly, the perspective of known phenomena was reinforced by a concrete number, for example, the migrations of Sephardic scribes to Italy. Also, the quantitative findings help determine whether a phenomenon is negligible as in the case of the small number of mutual co-censorship or substantial as in the case of repeating censorship. Secondly, new, previously not identified, insights were discovered. For example the relative restraint in creating manuscripts of "problematic subjects" manuscripts during periods of intense censorship.

This research has also some limitations, since Hebrew manuscripts are hosted in many different libraries around the globe, the metadata recorded in the catalogue is based mostly on microfilms or digitized images, and only in a few cases the cataloguers inspected the codices themselves. In addition, the catalogue of the Department of Manuscripts and Institute of Microfilmed Hebrew Manuscripts was created during a 70-year period by many different people. A large part of the catalogue was copied from catalogues of manuscripts in various libraries in other countries and contains only partial information. The rules of cataloging have changed over time. Therefore, there might be some mistakes in the catalogue data, but according to our arbitrary inspection, their proportion is relatively small and have no significant impact on the big picture of the studied phenomena.

In future work we plan to analyse more ontological event types, and extend the methodology for automated discovery and correction of catalogue mistakes as well as the completion of missing data.

Acknowledgements

Notes

[1] The geographical terms in this paper represent geographical and cultural areas, but not necessarily political entities. For example, by Italy we mean the territory of modern Italy, and similarly when mentioning other geographical terms. [Beit-Arié 2018, 75].

[2] For studies on the impact of the index in Italy and on the impact of censorship on the Italian Jewish communities see: Francesconi 2012, p. 154, n.2-3; Francesconi 2016, p. 102, n.7-8.

[3] http://republicofletters.stanford.edu/publications/voltaire/letters/

[4] In English (from Latin): "I signed Guidus menotius ... Jac Geraldini ... 28 Mars 1556".

[5] The two subsets of censored and non-censored manuscripts contain slightly less manuscripts than all the dated manuscripts before 1700 as few manuscripts lack script type information (e.g. of the 5,046 dated manuscripts created before 1700 in the 1st subset, 31 manuscripts are not included as they lack script style indication).

Works Cited

Ayers 2010 Ayers, E. L. "Turning towards Place, Space, and Time". In D. J. Bodenhamer, J. Corrigan and T. M. Harris (eds), *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, Bloomington (2010), pp. 1-13.

Baruchson-Arbib 1993 Baruchson-Arbib, S. *Books and Readers*, Bar-Ilan University Press, Ramat-Gun. (1993).

Baruchson-Arbib 2007 .Baruchson-Arbib, S. and Prebor, G. "Sefer Ha-Ziquq: (An Index of forbidden Hebrew books: The Book's use and its influence on Hebrew Printing", *La Bibliofilia*, CIX (1) (2007): 3-31.

Beit-Arié 2018 Beit-Arié, M. *Hebrew Codicology. Historical and Comparative Typology of Medieval Hebrew Codices based on the Documentation of the Extant Dated Manuscripts until 1540 from a Quantitative Approach*, Preprint internet English version 0.2 (November 2018). Retrieved June 05, 2019 from: https://web.nli.org.il/sites/NLI/English/collections/manuscripts/hebrewcodicology/Pages/default.aspx.

Benayahu 1971 Benayahu, M. *Haskama VeReshut in Venetian Printed Books*, Ben-Zvi Institute Mossad Harav Kook, Jerusalem. (1971).

Burrows and Curran 2012 Burrows, S. and Curran, M. "The French Book Trade in Enlightenment Europe Project and the STN Database". *Journal of Digital Humanities*, 1 (3) (2012).

Chandna et al. 2016 Chandna, S., Rindone, F., Dachsbacher, C., and Stotzka, R. "Quantitative exploration of large medieval manuscripts data for the codicological research". In *Large Data Analysis and Visualization (LDAV), 2016 IEEE 6th Symposium*, October 2016: 20-28.

Ethington 2007 Ethington, P. J. "Placing the past: 'Groundwork' for a spatial theory of history", *Rethinking History*, 11 (4) (2007): 465-493.

Francesconi 2012 Francesconi, F. "This passage can also be read differently…: How Jews and Christians censored Hebrew texts in early modern Modena", *Jewish History*, 26 (1-2) (2012): 139-160.

Francesconi 2016 Francesconi, F. "Illustrious Rabbis Facing the Italian Inquisition: Accommodating Censorship in Seventeenth-Century Italy". In *Jewish Books and their Readers*, Brill (2016), pp. 100-121.

Glinka et al. 2017 Glinka, K., Pietsch, C. and Dörk M., "Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections", *Digital Humanities Quaterly*, 11 (2) (2017).

Gregory and Healey 2007 Gregory, I. N. and Healey, R. G. "Historical GIS: structuring, mapping and analysing geographies of the past", *Progress in Human Geography*, 31(5) (2007): 638-653.

Kaplan and de Lenardo 2017 Kaplan, F., and di Lenardo, I. "Big Data of the Past", *Frontiers in Digital Humanities*, 4, 12 (2017).

Knowles and Hillier 2008 Knowles, A. K and Hillier, A. (Eds.). *Placing history: how maps, spatial data, and GIS are*

changing historical scholarship. Redlands, ESRI Press. (2008).

Lei Zeng 2017 Lei Zeng, M. "Smart Data for digital humanities", *Journal of Information and Documentation*, 2 (1), (2017): 1-2.

Moretti 2005 Moretti, F. *Graphs, maps, trees: abstract models for a literary history*. London, Verso. (2005).

Phillips 2016 Phillips, A. E. "Censorship of hebrew books on sixteenth century Italy: a review of a decade of english and french language scholarship", *Bibliofilia*, 118 (3) (2016): 409-425.

Popper 1899 Popper, W. *The Censorship of Hebrew Books*. The Knickerbocker press, New York. (1899).

Prebor 2003 Prebor, G. *Sepher Ha-Ziquq by Domenico Yerushalmi (1555-1621) and its Influence on Hebrew Printing*. Ph.D. thesis, Bar-Ilan University (2003). (In Hebrew).

Prebor 2007 Prebor, G. "From Jerusalem to Venice: The Life of Domenico Yerushalmi, his writings and his work as a Censor". *Pe'amim*, 111-112 (2007): 215-242. (In Hebrew).

Prebor et al. 2019 Prebor, G., Zhitomirsky-Geffet, M. and Miller, Y. (2019). "A new analytic framework for prediction of migration patterns and locations of historical manuscripts based on their script types". *Digital Scholarship in the Humanities*.

Raz-Krakotzkin 2007 Raz-Krakotzkin, A. *The Censor, the Editor, and the Text: The Catholic Church and the Shaping of the Jewish Canon in the Sixteenth Century*. University of Pennsylvania Press, Philadelphia. (2007).

Richler 1990 Richler, B. *Hebrew manuscripts: a treasured legacy*, Ofeq Institute, Cleveland. (1990).

Sirat 2002 Sirat, C. *Hebrew manuscripts of the Middle Ages*. New York: Cambridge University Press, Cambridge. (2002).

Sonne 1942 Sonne, I. "Expurgation of Hebrew Books", *Bulletin of the New York Public Library*, XLVI,(1942): 975-1015.

Van Hooland and Verborgh 2014 Van Hooland, S. and Verborgh, R. *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*. Neal-Schuman, Chicago. (2014).

Weiss 2016 Weiss, J. *A Kabbalistic Christian Messiah in the Renaissance*. Hakibbutz Hameuhad, Bnei Brak. (2016). (In Hebrew).

Zhitomirsky-Geffet and Prebor 2016 Zhitomirsky-Geffet, M., and Prebor, G. "Toward an Ontopedia for Historical Hebrew Manuscripts". *Frontiers in Digital Humanities*, 3 (3) (2016).

van Boxel 2016 van Boxel, P. *Jewish books in Christian hands: theology, exegesis and conversion under Gregory XIII (1572-1585)*. Biblioteca apostolica vaticana, (2016).