

METADATA, DIGITIZATION AND THE CAIRO GENIZAH: ISSUES IN USER ACCESS AND RETRIEVAL

Seth Jerchow and Heidi G. Lerner

Description: The contents of the Cairo Geniza amount to well over 2,000,000 fragments, scattered in collections worldwide, and covering a vast range of topics from the religious to the secular, and the literary to the documentary. Furthermore, the languages and orthographies in which the fragments are written include, among others, Hebrew, varieties of Arabic and Aramaic, Greek Persian, Coptic, and even Chinese. Another distinctive feature of Geniza materials is that individual fragments of a single manuscript are often dispersed among different institutions. Since the late nineteenth century, various approaches have been used to catalogue Geniza materials. Topical catalogues first appeared in the 1890s, and the 1920s saw the emergence of catalogues by collection. By the 1990s various electronic projects were being developed by institutions such as Cambridge University's Taylor-Schechter Research Unit, Princeton University, and the Saul Lieberman Institute of Talmudic Research. To date, however, no comprehensive catalogue of the Geniza as a unit has been produced. The University of Pennsylvania Library, in collaboration with the Taylor-Schechter Genizah Research Unit, has embarked on a project, whose goal is, via digitization and Internet technology, to implement an online catalogue and virtual repository of the Cairo Geniza, to maximize access to these materials and transcend institutional boundaries and their geographic limitations. The authors will discuss the advantages that Penn's collection presents in initiating such a project, and will address problematic issues of cataloguing methods and controls, including formatting, authorities, access and retrieval, and cross-joining of records of matching fragments.

Seth Jerchow is the public services librarian at the Center for Jewish Studies Library of the University of Pennsylvania. From 1993 to 1999, he served as research associate to Special Collections in the Library of the Jewish Theological Seminary. Seth received his doctor of letters degree from the Università degli Studi, Florence, Italy, and is completing a second doctorate in linguistics at the Albert-Ludwigs-Universität in Freiburg, Germany. His research interests include: linguistics and grammar, Hebrew and Judaica bibliography, codicology, paleography, history of the book, information technology, and Cairo Genizah studies. He has spoken and written on a wide variety of topics, including Judeo-Italian and Jewish languages, linguistics, history of printing, and Jewish liturgy.

Heidi Lerner received an MLS from the Graduate School of Library and Archive Studies at the Hebrew University in Jerusalem in 1982. Since 1990, she has served as the Hebraica-Judaica cataloger at the Stanford University Libraries. Ms. Lerner has spoken at several AJL conventions on various topics as well as at the past conference of the Association for Jewish Studies. She currently serves as chair of the AJL Cataloging Committee, and also is the coordinator for the NACO Hebraica Funnel within the Library of Congress' Program for Cooperative Cataloging.

Premise

The University of Pennsylvania and the Taylor-Schechter Genizah Research Unit at Cambridge University in England have embarked recently on a project to digitize their joint holdings. One specific goal of this collaboration is to develop and implement a working online catalogue for the University of Pennsylvania's collection of Genizah fragments, which will provide the foundation for a global electronic and online catalogue of the entire Cairo Genizah. Since the primary goal of this project is to make fragments accessible to as wide a group of scholars and researchers as possible, the project staffs at University of Pennsylvania and Cambridge have developed preliminary guidelines intended for use of standardized descriptive metadata. In this paper we shall discuss the issues and difficulties specific to Genizah cataloguing; how an online catalogue can facilitate this ambitious task, and why MARC tagging was adopted for this purpose.

Introduction

The Cairo Genizah is the most significant repository of primary documents regarding Jewish studies to have come to light in the Modern Era. The tradition of setting aside volumes containing sacred Hebrew texts rather than destroying or disposing of them is an ancient one, and can be found in practically every Jewish Community. Yet very few genizot have survived, for an obvious reason: the contents of genizot (denoting a "hidden" place [for storage], such as a closet or aron) are typically buried. The Genizah in the Ben Ezra synagogue in Fustat (founded 643 C. E.¹ - originally a Byzantine fortress, surrendered to Muslim Arabs in 641; this eastern district predates the founding of the modern Egyptian capital) is unique for a number of reasons:

1. It survived, due to the fact that the majority of the fragments were never removed for burial. Worn-out volumes and leaves were deposited in a second floor chamber located behind the women's gallery (in later periods this entrance was closed off and the chamber was accessible only through an exterior passageway). Remarkably, it witnessed and survived fires and acts of vandalism.
2. The quantity of materials, amounting at an estimated 220,000+ fragments.
3. The time span its contents cover. We know that the Ben Ezra Synagogue, or *Kanīsat al-Yerūshalmīyin* (or *al-Shāmiyīn*), was ordered destroyed by the Shiite caliphate in the 11th century, and rebuilt about 1040 (one Muslim source states that the Coptic Patriarch was forced to sell the church of St. Michael to the Jewish community in 882²). The fragments extend from the 8th or 9th century (and even earlier, as we count the palimpsests) up through the 19th, with large concentrations of materials dating from the 10th through the 15th centuries.
4. The importance of the Fustat community. The Ben Ezra Synagogue was *the* center of the Egyptian and indeed Mediterranean Jewish world from the time the Fatimid period (969-1171), and home to the Egyptian *nagid*. Prior to this, the *Kanīsat al-Shāmiyīn* was the seat of the

Palestinian Jewish community, one of two Rabbanite communities (the other being Babylonian) which coexisted with a Karaite community. It is virtually the only source to document a living *Minhag Yerushalmi*.

Throughout much of the 19th century, various collectors gained limited access into the Genizah. Some items were given as gifts by the community; others made their way into the *suq*. In 1896, Solomon Schechter, then Reader of Rabbinic Literature at Cambridge University, became aware of the Genizah's potential importance for Jewish studies. With the intellectual and pecuniary support of Charles Taylor, Master of St. John's College at Cambridge, the balance of the fragments – today estimated at around 140,000+ – were acquired from the Cairo Jewish community and brought to Cambridge.

Problems Inherent to Genizah Studies

In its present state[s], the Cairo Genizah presents, depending on one's proclivity, a cataloguer's paradise or a cataloguer's nightmare. As stated, Cambridge, at 140,000+, holds the largest collection of fragments in the world. But what about the remaining 80,000 to 100,000?

- **England:**
 - Cambridge, University Library - >140,000; Westminster College - ± 2,000.
 - Manchester, John Rylands University Library - ± 10,000.
 - Oxford, Bodleian Library - ± 5,000.
 - London, British Library - ± 5,000.
 - Birmingham, Selly Oak Colleges, Mingana and Mittwoch Collections - ± 40.
- **United States:**
 - New York, JTS - ± 40,000.
 - Philadelphia, University of Pennsylvania - > 487
 - Cincinnati, HUC - ± 250.
 - Washington, D.C., Smithsonian (various) – 114.
- **France:**
 - Paris, Alliance Israélite Universelle - ± 4,000; Jack Mosseri Collection - ± 4,000.
 - Strasbourg, Bibliothèque nationale et universitaire - ± 1000.
- **Austria:** Vienna, Österreichische Nationalbibliothek, Rainer Collection - ± 150.
- **Hungary:** Budapest, Academy of Sciences - ± 650
- **Russia:** St. Petersburg, National Library of Russia: Antonin Collection - ± 1200; Firkovich Collection – several thousand.
- **Ukraine:** Kiev, Academy of Sciences, Abraham Harkavy Collection – several dozen.
- **Israel:** JNUL – ± 300.

The Genizah, in its “original” state, was not a collection as much as a completely disorganized and unattended mass of discarded materials, subject to asystematic perusings and plunderings. Its contents are better described as presently “scattered”, rather than “distributed”. Individual leaves from a given particular manuscript, or fragments of individual leaves are now dispersed among different institutions. From the

time of the first divisions of fragments into personal and institutional holdings, collections have been sold, institutions have come and gone, two world wars have been fought, and maps, states, governments and ideologies have changed. All these events have complicated understatedly any inventorial assessment.

Catalogue Typology

Over the past century and a quarter, many catalogues for the Genizah have been produced. One catalogue type is organized around a local collection. As early as 1886, Adolf Neubauer's *Catalogue of the Hebrew Manuscripts in the Bodleian Library* [reissued in 1994 with addenda and corrigenda by Malachi Beit-Arye] included entries on the Bodleian's collection of Cairoene fragments. Another, perhaps the most exasperating, although [personally] engaging example is the one published in 1921 for the Elkan Nathan Adler collection (now at JTS). Some of the smaller collections, such as that of the Smithsonian (*Fragments from the Cairo Genizah in the Freer Collection*, edited by Richard Gottheil and William H. Worrell, New York-London, 1927), produced catalogues of great value. The Freer Catalogue treats its collection as an integral and publishable entity, and contains extensive descriptions, photographic facsimiles, transcriptions, and translations of its 52 fragments. Benzion Halper's *Descriptive Catalogue of Genizah Fragments in Philadelphia*, (Dropsie College for Hebrew and Cognate Learning, Philadelphia, 1924), provides neither facsimiles nor transcriptions. However, it is organized topically. The 487 fragments it describes are now housed at the University of Pennsylvania's Center for Judaic Studies Library. As such, it provides an important base for the Penn/Cambridge project.

Another catalogue type is organized solely by a single topic or genre. Some are oriented on specific genres within local collections such as Lewis-Gibson on the Syriac palimpsest fragments at Cambridge. Neil Danzig's 1997 catalogue of JTS Rabbinic fragments, while oriented upon the holdings of one local collection, exhaustively provides cross-matches and concordances. The best results transcend borders: As early as 1901, the *Facsimiles of the Fragments Hitherto Recovered of the Book of Ecclesiasticus in Hebrew*, edited by Solomon Schechter³, was published containing 60 photographic facsimiles of all extant *Ben Sira* fragments. This edition, although sparse on physical description, gathers in one volume fragments from the Taylor-Schechter, E. N. Adler, British Museum, Lewis-Gibson, Bodleian, Consistoire Israélite (Paris), and Gaster collections, identifies and collates from what had originally been four separate codices, and provides [for that time] exhaustive bibliographies⁴. The late Michael Klein's catalogues of Palestinian Targumic fragments assess: fragments within the context of the Targumin; reconstructions of the original codices; their distribution throughout the current collections.

Another noteworthy endeavor is the CD-ROM distributed by the Saul Lieberman Institute Database of Talmudic Versions. This CD-ROM is not restricted to Genizah fragments, but includes all manuscripts and early printed versions of the Talmud. However, it does include information regarding all Talmudic Genizah materials, and is subject to periodical updates. It also represents the first major multi-tiered and searchable electronic catalogue of the Genizah.

Cataloguing Issues⁵

To date, all catalogues have shared the disadvantage that plagues printed catalogues: the contained data (or “descriptive metadata”, as it were) are *static*. The advantage first presented by an online catalogue is its *dynamic* relation to the data. Data can be virtually input and distributed, as well as updated as needed. It also provides and facilitates an exponentially greater power of search capabilities. This type of catalogue is best adapted for the handling of Genizah materials. Data may be entered locally, it can be centrally stored, and the relation between materials and locations approaches an ideal level of recursivity previously unattainable.

However, in addition to the concept of static vs. dynamic data there is another concept relative to the *typology* of data, i.e. of information. Scholars and special collections librarians often refer to collections of many now defunct Jewish libraries. These former collections (e.g. Sassoon Collection), have well documented printed catalogues (e.g. *Ohel Dawid*, London, 1932), whose holdings were broken up at various points in time and redistributed into many collections (the London Beth Din and Beth Hamidrash Library, whose collection of manuscripts was obtained from the estate of Rabbi Solomon Hirschell in 1842, and described in the erroneously titled *Catalogue of the Hebrew Manuscripts in the Jews’ College, London*, Oxford, 1886, by Adolf Neubauer, is undergoing a similar fate of redistribution). Although these original collections are no longer intact, the volumes (for the most part...) remain whole and these catalogue still offer a valid point of reference.

This is not the case for the Cairo Genizah. As long as it existed and functioned as a *genizah*, it could be considered, only in the most generous of descriptions, a repository. During its existence as a unity no efforts at cataloguing its contents were ever made. We have already discussed the random and international scattering of individual codices, leaves, and pieces of leaves. One fragment can have had during its history as a holding item plural identifiers, and the same may apply to its relatives. By necessity, each entry produced by the Penn/Cambridge collaboration will include its respective Halper reference (containing, *inter alia*, important provenance information, e.g. [Cyrus] Adler, Amram, Sulzberger, etc.).

Conceptually, Genizah fragments contain two types of information: *intrinsic* and *extrinsic*⁶. The Genizah’s unusual distribution renders necessary an expansion of the definition “extrinsic data.” Any information regarding cross-matches, whether intra- or extra-institutional, textual, codicological, is data extrinsic to the fragment itself.⁷

Genizah – Ideal for testing MARC

As noted earlier, a key element in the project is to provide access to digitized images of the Geniza fragments. Scholars will view multiple, related documents, simultaneously, on one screen. These fragments, can be moved around, and mixed and match, unifying them in a way never before possible. The descriptive data of the individual fragments, will, via,

a web-based search-interface, lead the scholars to the graphic display of the scanned fragments.

Metadata has become widely discussed in the library, scholarly, computing and publishing communities. Information professionals in particular are excited about its potential to improve access to electronic materials. But any institution that begins a project utilizing metadata to describe its resources should very carefully develop its strategy so as to realistically address the technical, organizational, and human challenges involved in such a project. Successfully using metadata requires careful collaboration and planning between individuals and among institutions. Many digitization projects are discovering that it is expedient to integrate metadata into existing library systems and take advantage of well-defined standards of organizing information.

As most of us already know, Machine Readable Cataloging (MARC) originated in the 1960s as a means of exchanging library-cataloging records. It is made up of a data structure and encoding procedure that implements national and international standards. MARC is the encoding format most commonly used by libraries in North America and Europe for the storage, retrieval and exchange of bibliographic information. And significantly for this project, the MARC format provides a means for integrating descriptive metadata into existing systems.

Up until the present, the manuscript community has not on the whole embraced the MARC standard for its cataloging purposes. In codicology, the traditional methods for locating manuscripts have been printed catalogs. With the advent of the web and electronic technologies such as Text Encoding Initiative (TEI) or the European project known as Manuscript Access through Standards for Electronic Record or MASTER, it is getting easier to put manuscript metadata onto the Internet.

The University of Pennsylvania and Taylor-Schechter Genizah Research Unit jointly felt that a key component of the project is to integrate the fully cataloged bibliographic records for the digitized images into Penn's local catalog Franklin that runs on Voyager.

As much as possible, the project aims to adapt MARC encoding procedures to the cataloging of these fragments, and to provide cataloging that is compatible with AACR2, 2nd ed., rev. The resultant records will provide bibliographic control over the fragments, which, owing to their unique linguistic, religious, intellectual, historical and literary value, require precise and detailed identification. The appropriate MARC tags also allow linking from the online catalog record to the digitized fragment. *The Library of Congress Subject Headings* is intended to provide controlled subject access. Personal, corporate and title headings provide a unique challenge. For the most part, the records will use the authorized headings that already exist in the *National Authority File* or headings that have been created according to AACR2 guidelines.

The initial phase of the project focuses on the digitization of the fragments and the creation of online catalog records for the individual fragments held by the Center for Judaic Studies at the University of Pennsylvania. The aim of this project is to create a

searchable web-based database, which will allow scholars to locate and identify individual fragments by title, author, institution, language, physical characteristics, subject, or bibliographic history. A template is being developed that provides the descriptive elements of individual Genizah fragments. These elements were defined and then mapped to the corresponding MARC tags. Again though, the complexities of cataloging these items have to be emphasized. Most of these fragments are incomplete documents, with their mates scattered among many different institutions and collections, or even in different volumes within the same collection. They very often lack a title or colophon, and individual volumes may include multiple and even unrelated texts.

A main entry (100, 110, or 130) field is provided in as many cases as possible. The fragments cover a variety of types of material including literary fragments, liturgical works, biblical and rabbinic texts and their related commentaries, and other philosophical, scientific, and linguistic writings. Also included in the collections are a number of legal documents, communal and commercial records, educational documents, and private letters. For those works in which there is an identifiable author, a personal name heading goes into the 100 field. Ideally the heading will match the form established in the National Authority File. If no heading exists, the author's name should be formatted in accordance with current AACR2 cataloging standards. Alternatively, a fragment emanating from an administrative or communal body, institution or synagogue will have its issuing body recorded in a 110 field. Those fragments containing liturgical, biblical or rabbinic texts will have a uniform title with indication of part, version, language, translator, etc.

For those fragments in which there is a formal title, the title proper [245 field] will reflect the exact wording, which appears on it, or is extracted from one of the appropriate printed descriptions of the collection. If the fragment does not have an identifiable title, the title statement is provided by the cataloger based on an existing description of the fragment or from direct examination of it. If the manuscript contains several unique items bound together, a title is provided that represents the themes found in the group of items.

Many of the fragments are works that are or include translations or are parts of a larger work. When author or other main entry is included in a record for a fragment, a uniform title will be provided in the 240 field.

Alternative titles as might be extracted from various existing printed catalogs such as B. Halper's *Descriptive Catalogue of Genizah Fragments in Philadelphia* can be recorded in a 246 field.

The location and date that the fragment was copied is recorded in the 260 field. The collation as recorded by number of folios; and the measurements, given in centimeters are noted in the 300 field.

Codicological and paleographical features are crucial in providing the most precise description and identification of the fragments. One of the challenges currently facing the development of the template is the limit of the MARC 21 Bibliographic Format to encode

these unique characteristics. And the encoding of these features is what ultimately needs to be done to enable the most precise locating and identification of the fragments. Information such as the material of the fragment, the number of columns, blank sides, the condition of the fragment, binding and other unique characteristics are recorded as notes in a series of 500 fields. The 340 field contains the physical description for an item that has special conservation or storage needs and can possibly be adapted to the template to record the physical material of fragments and how the text was entered onto it. The 500 fields will also be used to provide detailed descriptions for those fragments which contain more than one work, or whose author and title are unidentifiable. Very often a detailed overview of the contents of the fragments needs to be provided with precise listings of the various passages. Passages from anonymous texts are often quoted. Provenance/acquisition information providing details on where the original of the digitized fragment is held is recorded in the 541 field.

The contents of the Geniza include almost anything written in Hebrew script, i.e., Hebrew, Judeo-Arabic, Ladino, Judeo-Greek, Jewish Aramaic, Judeo-Persian, and Yiddish. There are also fragments that are written in non-Hebraic scripts and in non-Hebraic languages such as Arabic in Arabic script, Coptic, Ethiopic, Syriac and even Chinese. Many of these fragments are/or include translations. Information on the language/dialect of the item, and details such as the type of script, vocalization and other linguistic and calligraphic details are currently recorded in the 546 field with the corresponding MARC language codes appearing in the 041. It is hoped that a more exact method for indexing these paleographic elements can be developed.

Another key element in identifying the Geniza fragments is a detailing of its bibliographic history. The 510 field has been adopted to provide the location of other bibliographical or archival data such as when, or where the item is cited, published or reproduced in facsimile. Producers, i.e., the contributors to the database; and end-users, i.e., scholars and researchers need to know such things as copyright and reproduction information. The 540 field displays the terms governing the use or reproduction of the described materials. The 544 field gives the location of related Genizah fragments by provenance, i.e., name of the collection and/or name of the institution where they are held.

The 580 field provides information that can link part of one incomplete manuscript to its mate(s) in another collection or collections. This field also links individual collections (as subsets) to the superset of the Cairo Genizah.

The 581 field has been developed to provide scholars with bibliographic information on published descriptions of the collection, or to lead researchers to articles or monographs that are based on research that emanates from the collection.

Topical identification of the fragments is crucial. This can include identification of periods in which the original text was written such as tannaitic, geonic, etc.; genres of literary texts such as piyyutim, responsa, midrash, aggada, or philosophical tracts; subject matter of the documents; identification of persons cited or involved in personal or

commercial transactions; religious, rabbinic, biblical and liturgical works and their related commentaries. This information is to be recorded in the relevant 600, 610, 630, 650 or 651 fields.

The 700 fields provide an adequate means of giving access to the many additional titles, people, and institutions that may be identified with a fragment. The 700, 710, 730 and 740 fields are used to record added entries for people or corporate bodies that are partially responsible for the document, as well as texts that co-exist, are related, or are included alongside the work that is being described. These can include an additional author, translator, commentator, editor or signatory, court, synagogue, or school. Fragments frequently contain multiple literary entities such as piyyutim and identifying each individual work is imperative for researchers and scholars. These will be recorded by title (or opening refrain) in the 740 or 730 fields. Often fragments will contain text that covers more than one book of the bible or tractates of Talmud, and these additional books and chapters will be recorded in the 730 field as added uniform title entries. The 787 field can be used to provide a link to other collections whose relationship has been described in the 581 field.

One of the major goals of this project is to utilize these online records as the basis for searching and retrieval of the digitized documents themselves. The 856 field will record the electronic location of the digitized fragment. Ultimately this field, can also provide an electronic link to digitized images of related fragments that may be found in other collections.

In a collection as large and diverse linguistically and bibliographically as the Cairo Geniza, it is obvious that an author or institution will have multiple forms of their names. These forms differ both within the documents themselves as well as the way they are cited in descriptive catalogs, etc. While no overall authority control currently exists as part of the initial project, the project has begun exploring the idea of creating a full authority file from the headings list. This could be accomplished by participation in an already existing cooperative program such as the Name Authority Cooperative (NACO) and its affiliated Hebrew and Arabic funnels.

Languages and Character Sets

One additional technology to render more feasible the cataloguing of the Genizah is the advent and expansion of Unicode. The Genizah includes an array of languages including, Hebrew, Jewish Aramaic and Christian Syriac, Arabic, Greek, Persian, Coptic and even Chinese. A number of languages are written both in native and Hebrew orthographies. Unicode virtually resolves all issues of language and character set:

1. “Plain text” Unicode contains well over 65,000 characters, happily exceeding in scope the linguistic demands of Genizah cataloguing.
2. Among these are characters used for romanization and transliteration.
3. Vocalization and prosodic pointing is included (*nikkudot* and *te’amim*)

4. Unicode characters have built-in bi-directional coding; Hebrew, Arabic, etc. automatically output as right-to-left. Additionally, Arabic and Syriac characters have built-in coding for positional shaping.
5. Unicode is increasingly gaining recognition as a major international standard. RAMBI and other on-line Israeli catalogues have already migrated their platforms to Unicode encoded versions. Systems such as Voyager are actively developing Unicode versions as well.

Although not within the current scope of the project, transcriptions of the fragments potentially could be incorporated (Princeton's Geniza Project – <http://www.princeton.edu/~geniza/> – is engaged in placing Goitein's transcriptions online). Links could be made available from within the individual records to fully searchable, online transcriptions of the fragments in both plain and marked-up text format. Traditionally, because of technical considerations, manuscript texts of this type have been transcribed or transliterated into Latin characters, sometimes with the addition of diacritics. However, with the implementation of Unicode, the potential to encode non-Roman language characteristics in texts is greatly improved. Unicode could also be used to enable the display of the diverse scripts and character sets of the Genizah fragments, within the descriptive records as well as in authority records for headings generated by the database.

You have seen a brief overview of the elements that are to be used in describing the digitized fragments. Of course other types of information or metadata are required to “define” or encapsulate a digital collection. Producers and end-users need to have recorded such things as digitization information, hardware and software requirements, and so on. The team from the University of Pennsylvania and Taylor-Schechter Genizah Research Institute includes librarians, software and application developers, and scholars. As project requirements and needs are looked at and discussed; reviewing, testing and refining results will continue throughout.

Conclusion: Virtual Collection

A question could be posed at this point: “Isn't the goal of this project simply one of expediency?” I would answer resoundingly “no.” Expediency is certainly a noteworthy objective. The fact remains, however, that the printed catalogues represent only a portion, albeit a majority one, of examined materials. Much still remains unexamined, unclassified, crated – even unconserved. Even as we proceed with the first stages of this project, we find it necessary to correct prior Halper entries.

Ultimately, all fragments must be described intrinsically and matched, indeed inserted into their extrinsic contexts of text, history and society. Digitization and dynamic cataloguing is our most economic means of *reunifying* Genizah materials, “under a *virtual* roof”. We will ultimately be able to measure its success when images of unclassified fragments can be identified and matched. At that point, this endeavor will have succeeded, not in the recreation of a cemetery, but in restoring life to the past.

¹ Unless otherwise noted, all dates are *C.E.*

² Reif p. 4; if this source is reputable, it is still uncertain if the church in question became the actual Ben Ezra synagogue [*ibid.*].

³ Oxford, University Press – Cambridge, University Press, 1901.

⁴ It was later discovered that the E. N. Adler collection contained a fragment from a fifth codex (ENA 3597, as well as a fragment from a Hebrew paraphrase of *Ben Sira* (ENA 3053). These were published in: *The Newly Discovered original Hebrew of Ben Sira (Ecclesiasticus XXXII, 16-XXXIV, 1), the fifth manuscript, and A prosodic version of Ben Sira (Ecclesiasticus XXII, 22-XXIII, 9)*. Ed. Joseph Marcus. Philadelphia, The Dropsie College for Hebrew and Cognate Learning, 1931.

⁵ For past overviews see: Jay Rovner. "The computerized Genizah cataloguing project of the Jewish Theological Seminary of America; its History, Current Status, and Future Prospects, with Some General Considerations of Bibliographic Control of Genizah Fragments". *Shofar* 8,4 (1990) 37-58; Ezra Chwat. "Danzig's Catalogue of Halakhah and Midrash Fragments in the E.N. Adler Collection and its Usage as a Research Tool". *Jewish Quarterly Review* 90,3-4 (2000) 405-415.

⁶ For the purposes of this paper, see the following definition: "The Dublin Core concentrates on describing intrinsic properties of the object. Intrinsic data refer to the properties of the work that could be discovered by having the work in hand, such as its intellectual content and physical form. This is distinguished from extrinsic data, which describe the context in which the work is used. For example, the "Subject" element is intrinsic data, while transaction information such as cost and access considerations are extrinsic data. The focus on intrinsic data in no way demeans the importance of other varieties of data, but simply reflects the need to keep the scope of deliberations narrowly focused." Dublin Core' is shorthand for the Dublin Metadata Core Element Set which is a core list of metadata elements agreed at the OCLC/NCSA Metadata Workshop in March 1995. The workshop report forms the documentation for the Dublin Core element set. (Stuart Weibel, Jean Miller, Ron Daniel. OCLC/NCSA metadata workshop report. OCLC, March 1995).

⁷ Such types of information lend themselves particularly well to set analysis for schematizing the information/data types and their relations. Any uniform catalogue model of the Cairo Genizah must express such relations:

INTRINSIC (data • •) = INTRINSIC (data • •);
EXTRINSIC (•x • [•y • •n...]);
• (•x • [•y • •n...]) = (•x • [•y • •n...]);
• INTRINSIC • • EXTRINSIC •n...

In this model, all types of intrinsic information are an element of fragment •, identical with the statement that: •x contains a subset of textual and codicological information, which in turn [potentially] intersects with another • or other •n... Therefore: 1. •x as a subset of •y or ••n... implies that •x is an element of •y or ••n... ; 2. the relation of any intrinsic to its extrinsics forms a union. The Penn-Cambridge project's template effectively expresses this unique relation between intrinsic and extrinsic elements.