Digitizing 140,000 Pages of Yiddish Linguistics: Strategies and Lessons Learned

Michelle Chesner
Norman E. Alexander Librarian for Jewish Studies
Columbia University

Association of Jewish Libraries Proceedings
June 2019
Background

"Data ages like wine, software ages like fish"
(attributed to Andrew Hankinson)

American Philosophical Society Grant:

URIEL WEINREICH, Columbia University

HUMANITIES

The grantee, who is serving as the director of the Language and Culture Atlas of Ashkenazic Jewry, wished to explore the possibilities of utilizing electronic data-processing machinery for the storage, filing, sorting, and searching of the accumulating field materials of the atlas.

Association of Jewish Libraries Proceedings
June 2019
Description of Project

Project: Language and Culture Atlas of Ashkenazic Jewry

Proposer: Professor Uriel Weinreich, Chairman, Department of Linguistics

Planner: Uriel Weinreich and Marvin I. Herzog (research assistant)

Computer: Mendel Hoffman

Dates: Beginning February 27, 1961 to January 31, 1963

Magnitude: Number of cards - 200,000
Machines used: 026, 402, 519, 682, 077, 650
Number of hours: about 500

Results: An atlas of Yiddish dialects and of Jewish folk-culture patterns of Central and Eastern Europe, for correlations with language and culture boundaries in the non-Jewish societies of the same territories.

Publications: The Atlas itself and numerous reports and monographs on special subjects.

Method: Frequency distributions, symbol tables, profile system tabulations.
Data Collection Formats

Sound:
- Reel to reel audio

Transcription:
- Blue books
- Answer sheets
- Printouts
Reverse engineering old media

And reprocessing a ~300 box archival collection
Decisions, decisions

• “Scope creep”
• Fitting within internal infrastructure
  – Data v. image as data
Integrated into existing image viewer
Decisions, decisions

• “Scope creep”

• Fitting within internal infrastructure
  – Data v. image as data

• Information overload (no “custom” option)
Dedicated **Libguide** for all the additional details

---

**LCAAJ Digital Archive User Guide: Introduction**

This is a guide to using the Digital Archive to the Language and Culture Atlas of Ashkenazic Jewry (dil.library.columbia.edu/caaj)

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Overview of the Collection</th>
<th>Questionnaire</th>
<th>Interviewee (Location) Data</th>
<th>Answer Data</th>
<th>Audio Tapes and Recordings</th>
<th>Using the Data</th>
<th>Appendices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glossary</td>
<td>Bibliography of Works citing the LCAAJ</td>
<td>Acknowledgements</td>
<td>Copyright Statement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Introduction**

This is a supplementary guide to the **Digitized data for the Language and Culture Atlas of Ashkenazic Jewry** which makes available online the linguistic and ethnographic information about Yiddish, gathered by the LCAAJ team so that researchers could map or otherwise systematically analyze it. (A history and fuller description of the LCAAJ project can be accessed by clicking here.)

For information on how to search the Digitized LCAAJ, see Searching for Answers to Questions.

Before beginning your search, we strongly suggest that you look at the page providing an overview of the collection, and look at the important scope note below.

**What the digitized LCAAJ website does:**

- Provides access to the images of the pages with answers given in interviews in the form of Blue Books, Answer Sheets, and Printouts
  - Further information about these materials can be found in this guide under Answer Data
- Provides filters to allow one to limit their search by format or location/region
- Provides detailed metadata for those answers to allow ease of searching
- Provides persistent URLs and IIF manifests to allow use of the images in other contexts

**What the digitized LCAAJ website DOES NOT do:**

- Provides an interface for digitally mapping data (for that, we recommend taking a look at the LCAAJResearch page, as well as joining our community of LCAAJ scholars by sending us an email at lcaaj@library.columbia.edu)
- Provide test-searchable answers to the questions (for that, see LCAAJResearch for access to downloadable PDFs of the printouts with OCR, and to read more about our plans on gathering that information as a community).
Decisions, decisions

• “Scope creep”
• Fitting within internal infrastructure
  – Data v. image as data
• Information overload (no “custom” option)
• Ongoing work
  – Divisional priorities (DLST v. DS)
Future Plans (pending..)

• Time-based media
• **Creative reuse** (and feedback)
• Matching audio to transcript
  – Time-based media project
• Archives as window into the world of early Yiddish in the US
• Classroom potential
• Exhibit