Building for the Future
AI and Machine Learning

Thomas Padilla
Internet Archive

Jarrow Ladies Fire Brigade, 1916
Tyne & Wear Archives & Museums
How can libraries strategically collaborate to address these challenges?
collections as data (_, I, II)

illustration by adam ferriss
We are excited to announce the release of the Vancouver Statement on Collections-as-Data (in English, in Spanish, in French, in Arabic). This statement is an update to the Santa Barbara Statement on Collections as Data and is a direct product of the international working event Collections as Data: State of the Field and Future Directions (~ 60 participants, 18 countries).
Responsible Operations: Data Science, Machine Learning, and AI in Libraries

Thomas Padilla
Practitioner Researcher in Residence
Challenges are organized across seven areas of investigation:

1. Committing to Responsible Operations
2. Description and Discovery
3. Shared Methods and Data
4. Machine-Actionable Collections
5. Workforce Development
6. Data Science Services
7. Sustaining Interprofessional and Interdisciplinary Collaboration

Organizations can use Responsible Operations to make a case for addressing challenges, and the recommendations provide an excellent starting place for discussion and action.
How can libraries strategically collaborate to address these challenges?
Augmenting and Adapting Services

2. Realizing Collection Access and Computational Use
3. Determining and exercising our Agency
### Academic Library Staffing & Services

<table>
<thead>
<tr>
<th>90s</th>
<th>00s</th>
<th>10s</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emphasis on technical services staffing</strong></td>
<td><strong>Emphasis on IR and public services staffing</strong></td>
<td><strong>Emphasis on research services staffing</strong></td>
</tr>
<tr>
<td>Realize opportunity of the web v.1</td>
<td>Realize opportunity of the web v.2</td>
<td>Support engagement with networked content and workflows v.1</td>
</tr>
</tbody>
</table>
20s ...

New ability to scale staffing capabilities across functional roles

Support engagement with networked content and workflows v.2
Current Tiers of Service

Tier 1

Library supports ...

Intro to Python, Databases, Visualization, etc.

Tier 2

Tier 3

Centralized High Performance Computing

Engineering support for advanced users
## Potential Tiers of Service

<table>
<thead>
<tr>
<th>Tier 1</th>
<th>Tier 1.5</th>
<th>Tier 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library supports ...</td>
<td>Library + LLM ...</td>
<td>Centralized High Performance Computing</td>
</tr>
<tr>
<td>Intro to Python, Databases, Visualization, etc.</td>
<td>Next gen. info literacy, Intermediate Python, Advanced webscraping, Intermediate visualization, etc.</td>
<td>Engineering support for advanced users</td>
</tr>
</tbody>
</table>
As the functional work of generating code becomes easier ... it suggests a role for libraries in investing more deeply in a long core area of work ... fostering fundamental literacies - information, data, algorithms, and more.

Libraries will need to collaborate on development of next generation AI-responsive information literacy instruction. They already are at the individual organization level (e.g., Carnegie Mellon University). Collective work (e.g., framework, standards updates) around the bend.
1. Augmenting and Adapting Services
   - Realizing Collection Access and Computational Use
3. Determining and exercising our Agency
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
backlog(s)
<table>
<thead>
<tr>
<th>Code</th>
<th>Language</th>
<th>Country</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>Madagascar</td>
<td>9,398,700</td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>Tunisia</td>
<td>9,308,000</td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td>Rwanda</td>
<td>9,306,800</td>
<td></td>
</tr>
<tr>
<td>ZU</td>
<td>South Africa</td>
<td>9,142,000</td>
<td></td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>9,000,000</td>
<td></td>
</tr>
<tr>
<td>SW</td>
<td>Sweden</td>
<td>9,000,000</td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>Italy</td>
<td>8,974,000</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>Ethiopia</td>
<td>8,920,000</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>Afghanistan</td>
<td>8,206,000</td>
<td></td>
</tr>
<tr>
<td>KZ</td>
<td>Kazakhstan</td>
<td>8,000,000</td>
<td></td>
</tr>
<tr>
<td>ILC</td>
<td>Philippines</td>
<td>8,000,000</td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>Russia</td>
<td>8,000,000</td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>Nigeria</td>
<td>7,611,900</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>Yemen</td>
<td>7,600,000</td>
<td></td>
</tr>
<tr>
<td>UY</td>
<td>China</td>
<td>7,595,512</td>
<td></td>
</tr>
<tr>
<td>HA</td>
<td>Haiti</td>
<td>7,372,000</td>
<td></td>
</tr>
<tr>
<td>AZ</td>
<td>Azerbaijan</td>
<td>7,059,000</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>Italy</td>
<td>7,047,400</td>
<td></td>
</tr>
<tr>
<td>KH</td>
<td>Cambodia</td>
<td>7,039,200</td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>Afghanistan</td>
<td>7,000,000</td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td>Ghana</td>
<td>7,000,000</td>
<td></td>
</tr>
<tr>
<td>HI</td>
<td>Philippines</td>
<td>7,000,000</td>
<td></td>
</tr>
<tr>
<td>KU</td>
<td>Turkey</td>
<td>7,000,000</td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>Zimbabwe</td>
<td>7,000,000</td>
<td></td>
</tr>
</tbody>
</table>
Introducing speech-to-text, text-to-speech, and more for 1,100+ languages

May 22, 2023 • 7 minute read
Opportunity: resolve “no preservation without use” tension for materials that are otherwise inaccessible due to resource limitations of collecting institutions (e.g., capacity, linguistic, etc.)

Challenges: finding a way to join or make use of efforts like Meta’s to identify a solution to collection access for collections in less commonly known languages (“less common” a relative, sometimes problematic framing).
Longtail collections as data
Current LLM focus; collections as data
Potential LLM focus; collections as data
Opportunities: library community driven LLMs - more representative data, more representative functionality, and richer research and learning possibilities.

Challenges: strategic collaboration that supports collection integration for LLM development from less well-resourced organizations.
1. Augmenting and Adapting Services
2. Realizing Collection Access and Computational Use
   Determining and exercising our Agency
Looking towards a brighter future: the potentiality of AI and digital transformations to library spaces

Responsible Operations: Data Science, Machine Learning, and AI in Libraries

Thomas Padilla
Practitioner Researcher in Residence
“Who or what orgs have particular agency to bring this work forward?”

- Lorcan Dempsey
library
associations...

archive
associations ...

museum
associations ...
Call for Expressions of Interest in Joint ARL/CNI Task Force on Scenario Planning for AI and Machine-Learning Futures—Deadline Friday, October 20

Cynthia Hudson Vitale & Clifford A. Lynch | cvitale@arl.org & clifford@cni.org | October 4, 2023
Opportunities: advance library objectives at service, policy, and legislative levels.

Challenges: Determination of particular library community agency relative to AI and ML. Consensus developing. Collective action needed.