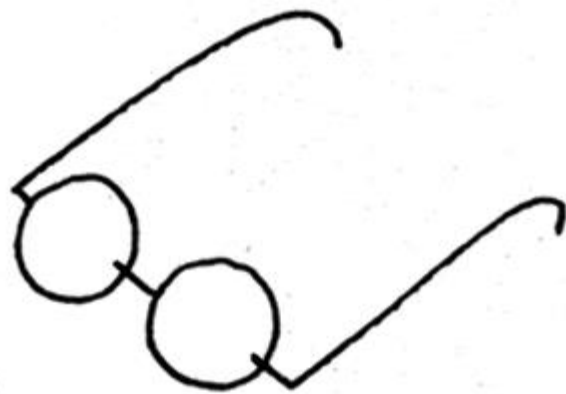


Building for the Future AI and Machine Learning

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

Vancouver Statement on Collections-as-Data; La Déclaration de Vancouver relative aux collections-en-tant-que- données; بيان فانكوفر بشأن التجميعات كبيانات

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

90s

Emphasis on
technical services
staffing

**Realize
opportunity of the
web v.1**

00s

Emphasis on
IR and public
services
staffing

**Realize
opportunity
of the web v.2**

10s

Emphasis on
research services
staffing

**Support
engagement with
networked content
and workflows v.1**

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 ?



Tier 3

Centralized High
Performance
Computing

Engineering
support for
advanced users

Potential Tiers of Service

Tier 1

Library supports ...

Intro to Python,
Databases,
Visualization, etc.

Tier 1.5

Library + LLM ...

Next gen. info
literacy, Intermediate
Python, Advanced
webscraping,
Intermediate
visualization, etc.

Tier 3

Centralized High
Performance
Computing

Engineering
support for
advanced users

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)

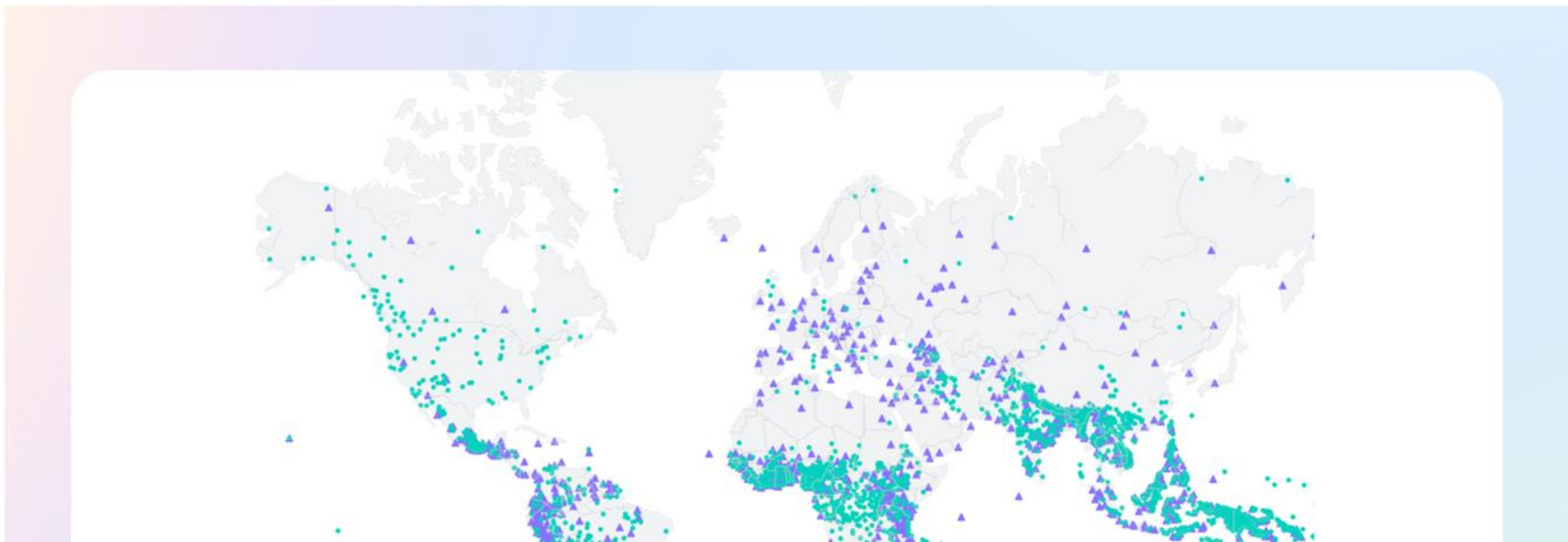


84	MALAGASY [MEX]	Madagascar	9,398,700
85	ARABIC, TUNISIAN SPOKEN [AEB]	Tunisia	9,308,000
86	RWANDA [RUA]	Rwanda	9,306,800
87	ZULU [ZUU]	South Africa	9,142,000
88	BULGARIAN [BLG]	Bulgaria	9,000,000
89	SWEDISH [SWD]	Sweden	9,000,000
90	LOMBARD [LMO]	Italy	8,974,000
91	OROMO, WEST-CENTRAL [GAZ]	Ethiopia	8,920,000
92	PASHTO, SOUTHERN [PBT]	Afghanistan	8,206,000
93	KAZAKH [KAZ]	Kazakhstan	8,000,000
94	ILOCANO [ILO]	Philippines	8,000,000
95	TATAR [TTR]	Russia	8,000,000
96	FULFULDE, NIGERIAN [FUV]	Nigeria	7,611,000
97	ARABIC, SANAANI SPOKEN [AYN]	Yemen	7,600,000
98	UYGHUR [UIG]	China	7,595,512
99	HAITIAN CREOLE FRENCH [HAT]	Haiti	7,372,000
100	AZERBAIJANI, NORTH [AZE]	Azerbaijan	7,059,000
101	NAPOLETANO-CALABRESE [NPL]	Italy	7,047,400
102	KHMER, CENTRAL [KMR]	Cambodia	7,039,200
103	FARSI, EASTERN [PRS]	Afghanistan	7,000,000
104	AKAN [TWS]	Ghana	7,000,000
105	HILIGAYNON [HIL]	Philippines	7,000,000
106	KURMANJI [KUR]	Turkey	7,000,000
107	SHONA [SHD]	Zimbabwe	7,000,000

Natural Language Processing

Introducing speech-to-text, text-to-speech, and more for 1,100+ languages

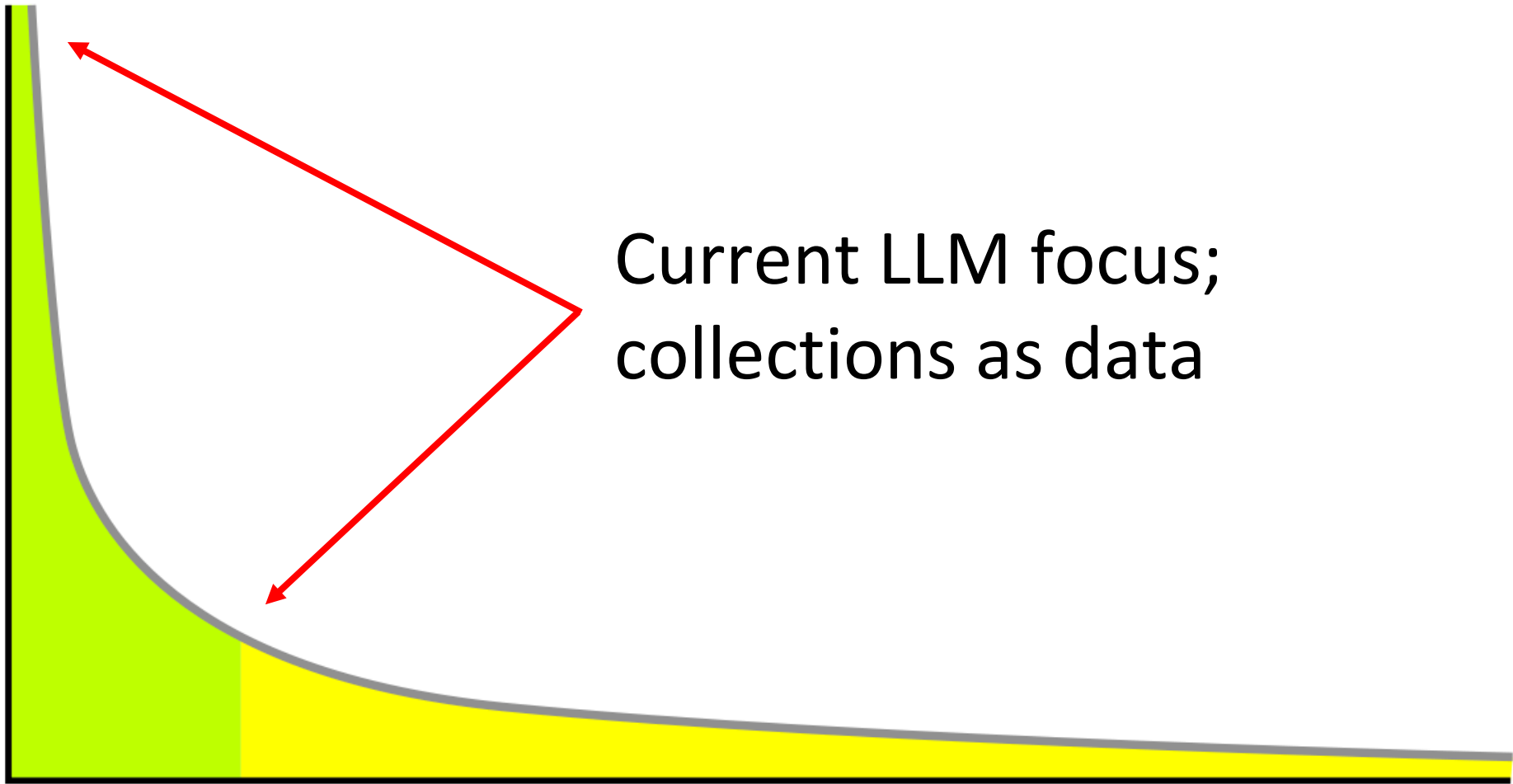
May 22, 2023 • ⌚ 7 minute read



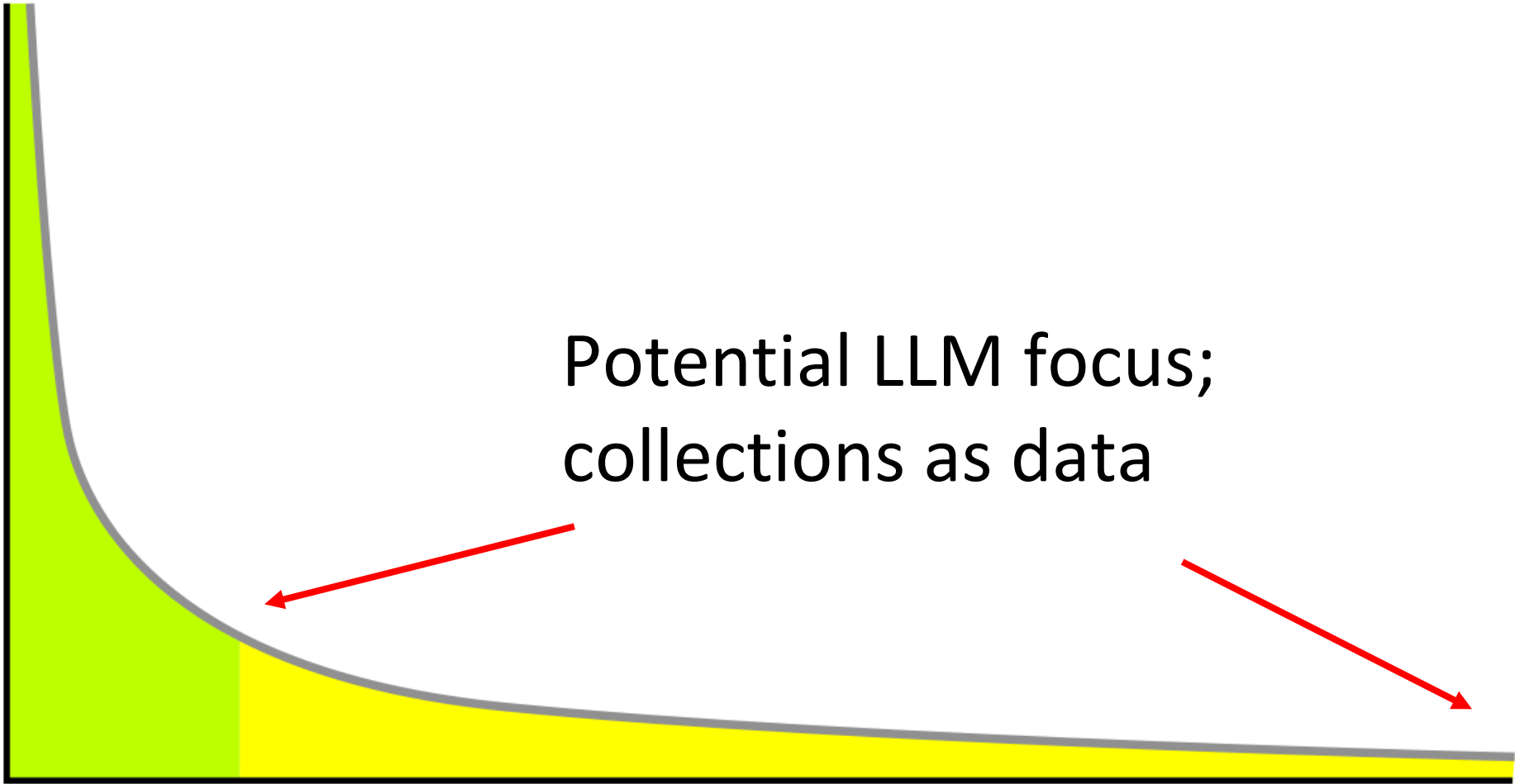
Opportunity: resolve “no preservation without use” tension for materials that are otherwise inaccessible due to resource limitations of collecting institutions (e.g., capacity, linguistic, etc.)

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



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

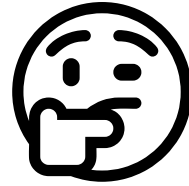
Thomas Padilla

Practitioner Researcher in Residence

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 ...



[Home](#) / [News](#) / [ARL News](#)

[/ Call for Expressions of Interest in Joint ARL/CNI Task Force on Scenario Planning for AI and Machine-L](#)

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.

