Los Alamos NATIONAL LABORATORY

Indexing LANL's Historical Collections in Less Than 1,000 Years

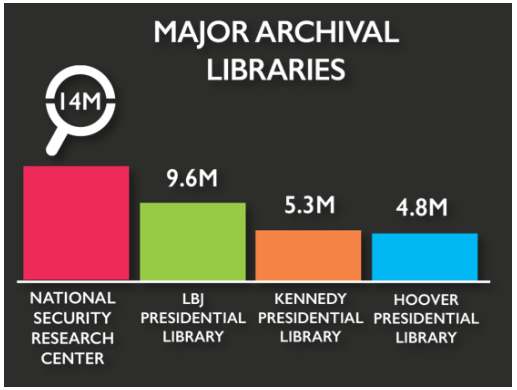
Julie Maze

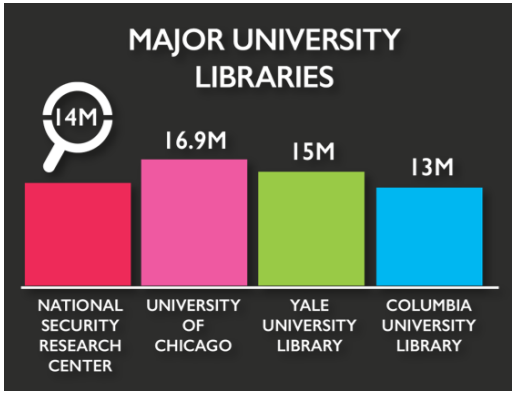
June 1, 2022

LA-UR-22-24853

National Security Research Center

Thousands of years to catalog/index





TITAN
ON THE RED

Requirements

- ⇢ *Auto-metadata extraction*
- ⇢ *Aid self-guided search and discovery*

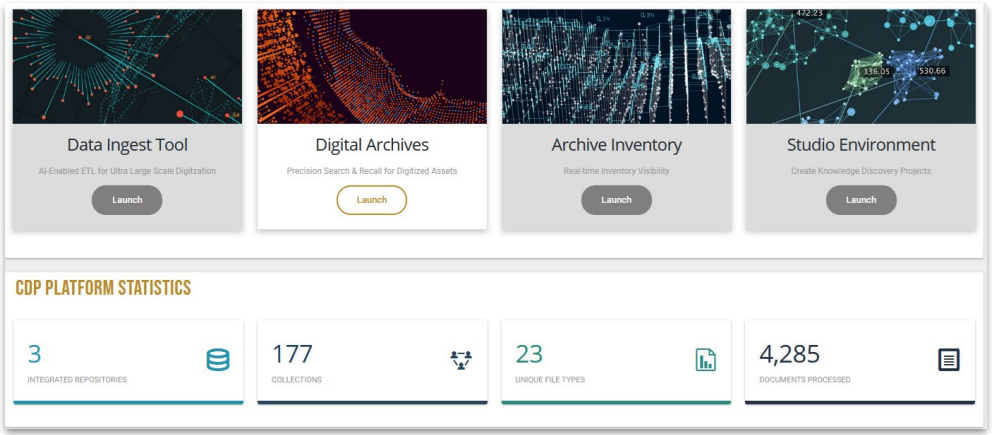
Key capabilities

- ⇢ *Recognize text*
- ⇢ *Extract metadata*
- ⇢ *Natural language processing*
- ⇢ *Present relevant result*
    - *search terms*
    - *relational concepts*

TITAN ON THE RED

6 month pilot

- ⇢ *connected 4 repositories*
- ⇢ *tested permission persistence*
- ⇢ *tested formats and fonts*
- ⇢ *created small taxonomy*



Los Alamos NATIONAL LABORATORY

Cataloging / Indexing

- *Currently using **fully manual** process*
- *Process 10-30 minutes per doc*
- *1.5 FTEs*
- *Total number of **digitized files growing** rapidly*
- *14M documents **not yet** digitized*

Current fully manual process		
	digital	physical
Quantity	**2.4 M**	**14 M**
Rate/month *1.5 FTEs*	486	486
Years to complete	**412**	**2,400**

Cataloging / Indexing

- *Currently using **fully manual** process*
- *Process 10-30 minutes per doc*
- *1.5 FTEs*
- *Total number of **digitized files growing** rapidly*
- *14M documents **not yet** digitized*

Current fully manual process		
	digital	physical
Quantity	**2.4 M**	**14 M**
Rate/month *1.5 FTEs*	486	486
Years to complete	**412**	**2,400**

If we don't change it will take 2,800+ years to catalog

Los Alamos NATIONAL LABORATORY

History is the foundation for the future

Divider: last nuclear test 1992

Knowledge transfer

Current geopolitics

Los Alamos NATIONAL LABORATORY

What it takes

Task	LANL Config	Cost
Identify repositories	Archive, Shared Drives, Sharepoint Sites	n/a
Build ontologies	3 FTEs	$500K/yr
Software	2 dev, 1 enterprise license	$2.5M/yr
Infrastructure	1PB encrypted storage 0.6FTE	$500K equip $130K FTEs
Connect repositories	1 FTE	$300K
QA Testing	TBD	TBD

Can you afford thousands of years?