

# Curators to the Rescue!

Using ABBYY FineReader PDF Software to Make  
Accessible Legacy Documents and Datasets

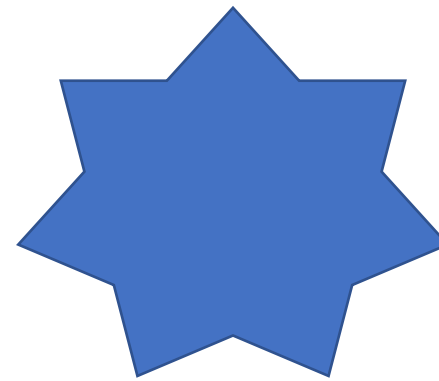
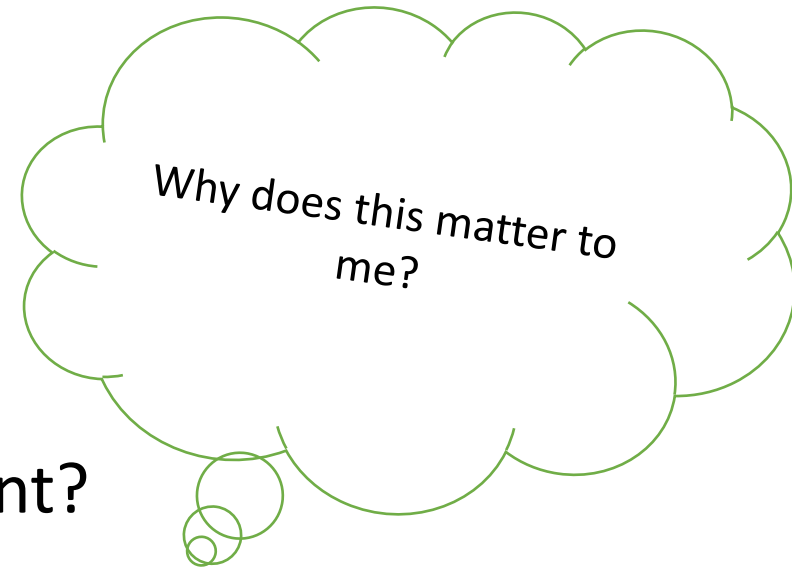
Peyton Tvrdy, Data Curation and Data Management Fellow at the  
National Transportation Library, [0000-0002-9720-4725](tel:0000-0002-9720-4725)





# Today's Objectives

- What is Legacy Data?
- What is ABBYY FineReader?
- Why is Data Accessibility Important?
- VIUS/TIUS Rescue
- Demonstration of the software





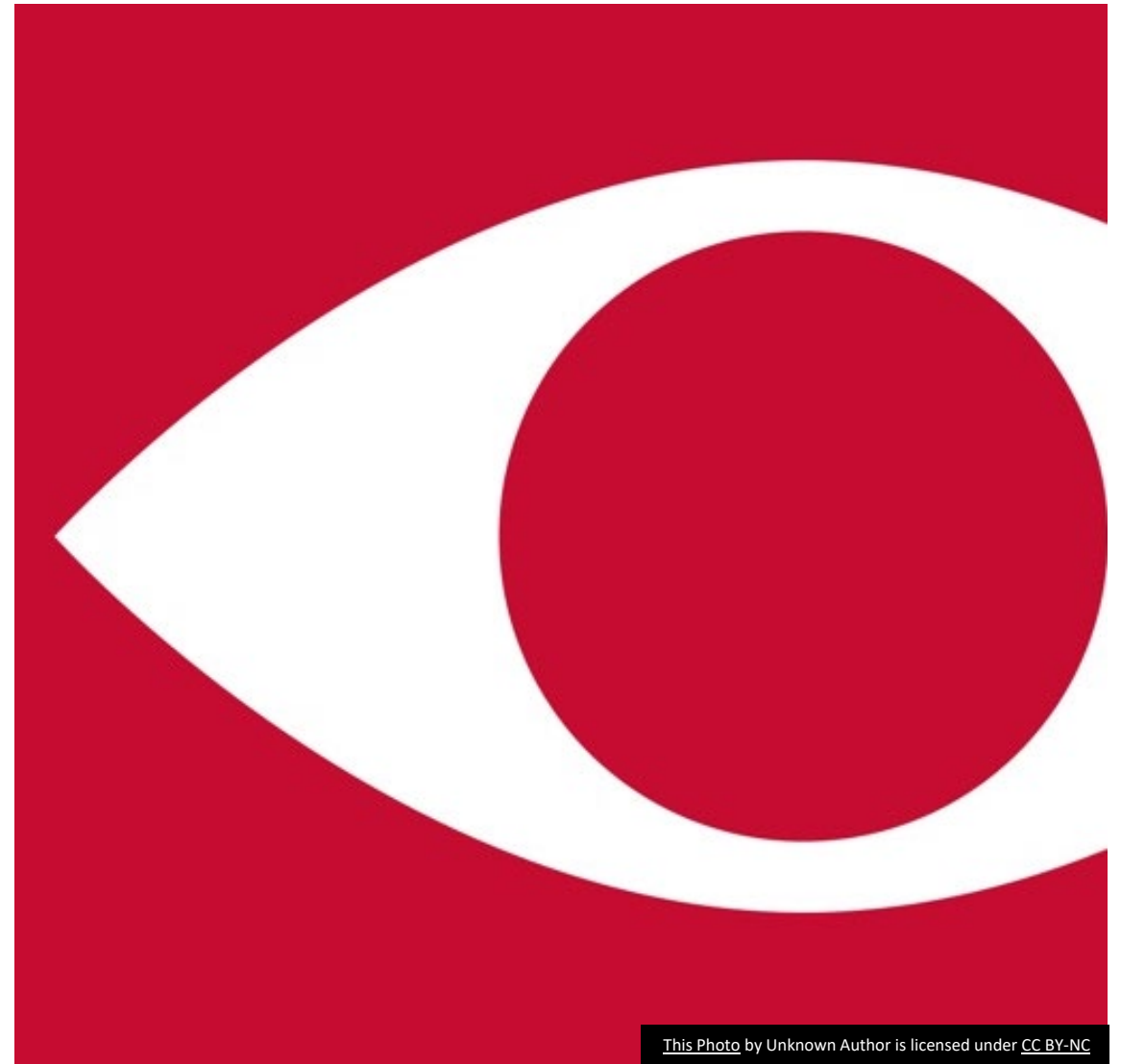
# What is “Legacy Data”?

- **Legacy Data:** data that is outdated due to its system, format, any surrounding technology, and relevance.
- Could have historical, cultural, or other uses to researchers today
- Examples:
  - Old print only reports and data tables (*VIUS Project*)
  - Graphs and source data stored on floppy disks
  - Raw data collection only on paper
  - Database that is not accessible on modern computers/software
  - *Anything that is not accessible to the public in a digital form that should be*

# What is ABBYY FineReader and How Does it Help Me?

---

- ABBYY FineReader PDF software is an **Optical Character Recognition** program that can edit PDFs, embed machine-readable text layers into image-only PDFs, export PDFs into various formats, and even add annotations.
- Benefits:
  - Makes PDFs searchable and accessible
  - Can be read by accessibility software
  - Can extract items such as data tables in tabular formats such as Excel
  - Fix scanning issues such as skewed pages
  - Significantly reduces human effort in extracting text and data



This Photo by Unknown Author is licensed under [CC BY-NC](#)

## Why is Data Accessibility Important for Legacy Data?

- Improve the quality and reliability of legacy data as opposed to manually transcribing, leading to more accurate analysis and future use
- Enable new insights and comparisons that were not possible easily before
- Unlocks and new world of use and application
- Maximizes value of research output

Item	Reported figure based on sample (1)	Sampling variability <sup>1</sup> (2)
BODY TYPE		
Pickups	68	1.2
Panel, walk-ins	7.5	1.1
Platform and cattle rack	14	0.8
Van	4.5	0.2
Refrigerator	1.7	0.2
Dump	2.4	0.1
Tank	1.6	0.2
Cement mixers and "not reported"	0.3	—
BODY SIZE		
Light	77	1
Medium	7.2	0.3
Light-heavy	9	0.4
Heavy-heavy	6.8	0.5
MAJOR USE		
Agriculture	32.7	1.5
Personal	29.9	2
For hire	3.6	0.4
Contract construction	6.4	1
Manufacturing	3.3	0.5

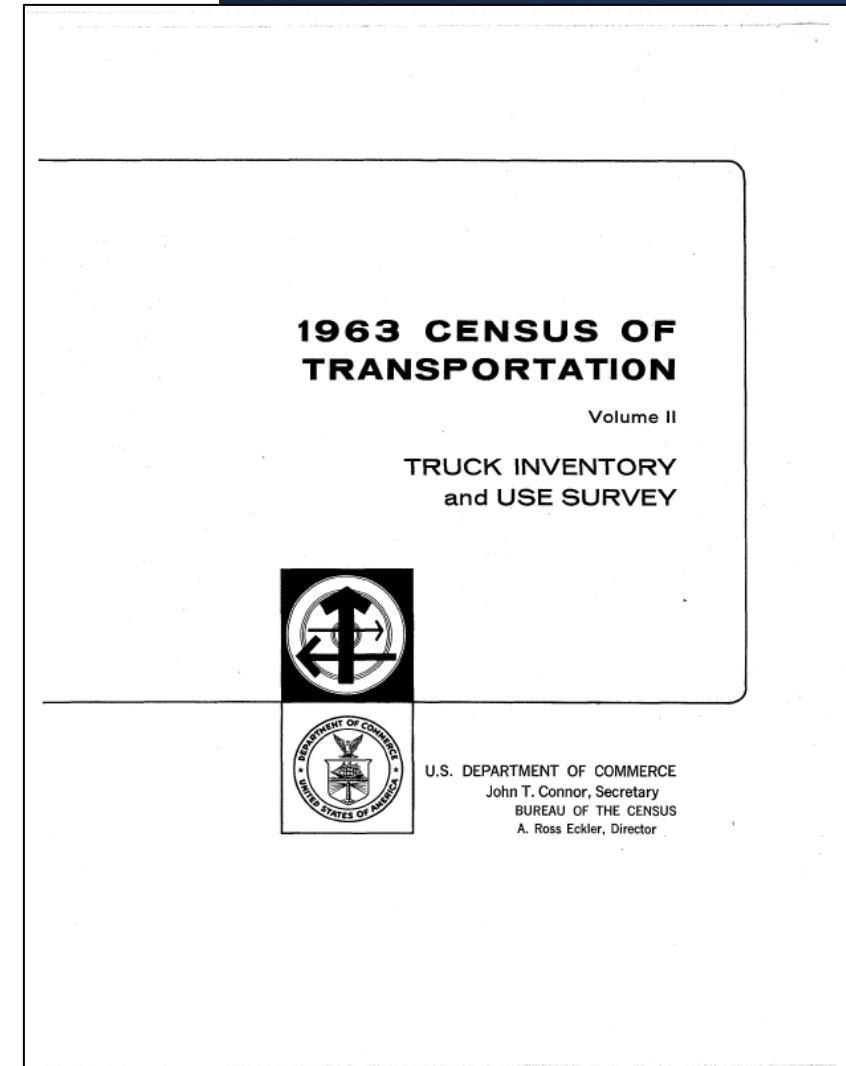


# Section II

ABBYY FineReader Use Case and Application  
The Truck Inventory and  
Use Survey Project

# TIUS, VIUS, and ABBYY Timeline and Overview

- Vehicle Inventory and Use Survey (VIUS) was restarted in 2021 after dissolution in 2002
- Originally known as the Truck Inventory and Use Survey (TIUS), changed in 1997
- Project focused on 1963, 1967, and 1972
- Output
  - 1387 new data tables
  - 156 cleaned, corrected, searchable PDFs
  - New workflow for extracting and correcting PDF text and tables





# Using ABBYY to get Results



- 
- Recognize:
    - Tables
    - Pictures
    - Text
  - Higher accuracy than manual, human effort, but still requires review and correction by professional
  - Software flags potential errors in blue highlight
  - Distinguish columns, rows, and cell contents, but human adjustment is needed to get correct format
  - Still errors, especially with tables. Requires human effort and good eye





How  
It's Made

