

Titled Data Tracking System

Jeffrey Shen

Supervised by Carla Medalia, Jon Morgan, and Emily Dich

Economic Reimbursable Surveys Division (ERD)



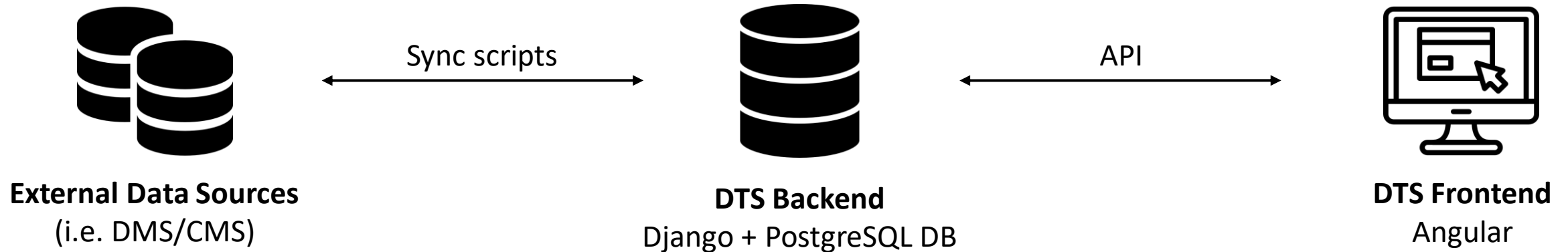
Motivation—Data Stewardship

Census is statutorily authorized to work with protected data, including administrative data from across the Federal government, so there should be a way to track and audit this data use.

- Data providers trust us with their data, want to know where their data lives, who is using it, and how it is being used
- **Solution:** Data Tracking System (DTS)
 1. Enterprise-wide logging of data file use (completed April 2022)
 2. Web app (In progress)
 - a. Aggregates information from disparate systems: usage from Phase 1, permissions metadata from Data Management System (DMS), etc.
 - b. Generates reports, dashboards, and alerts on data use
 - c. Shows if file use is consistent with permissions
 3. Detecting file lineage (Previous CDF)



DTS Architecture



Backend

- Updated scripts to sync the data from DMS
- Modified DTS database models
- Added API endpoints

Frontend

- Conducted user research
- Developed features for different stakeholders



Stakeholder: Information Owner

- Many types of users: i.e. project contacts, division chief, information owner
- Information Owner: manage permissions for datasets, responsible for ensuring compliance
 - **Example workflows:** generating quarterly reports on usage, seeing compliance issues, managing dataset metadata
- **Solution:** Developing features for info owners to assess the datasets they own for any inconsistencies or issues



Dataset Title [DMS ID: 0000000]

Description: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

[Overview](#) [Projects](#) [Users](#) [Servers](#) [Issues](#) [Activity](#) [Alerts](#)

Dataset

DMS ID: 0000000

Source: DMS

Type: Statistical

Safeguards:

- Title 13

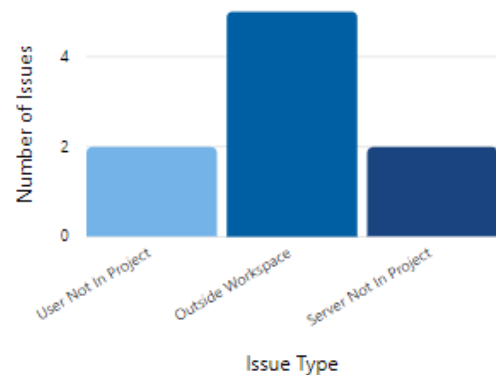
Info Owner

- John B Smith
Email: john.b.smith@census.gov

Activity

[Week](#) [Month](#) [Six Months](#)

Open Issues



Project Summary

[By Project Type](#) [By Project Status](#)

Projects who have used this dataset: 22

- Internal: 19
- External: 3

Projects approved to use this dataset: 50

- Internal: 46
- External: 4

User Summary

[By User Type](#) [By User Status](#)

Users who have used this dataset: 22

- Contractor: 11
- Employee: 11
- N/A: 0

Users approved to use this dataset: 819

- Contractor: 459
- Employee: 355
- N/A: 5

- Shows dataset metadata
- High level overview of activity, issues, and who is using dataset

Disclaimer: screenshot includes fake data

Dataset Activity									
Datetime ↑↓ ▾	Description ↑↓ ▾	Dataset Issue? ↑↓ ▾	Event Issue? ↑↓ ▾	Server Issue? ↑↓ ▾	User Issue? ↑↓ ▾	Project ↑↓ ▾	Server ↑↓ ▾	User ↑↓ ▾	Path ↑↓ ▾
1/28/2022 2:38 AM	delete	False	False	False	False	000000	Server A	user001	/path/000000/a/b.xml
1/18/2022 10:12 PM	write	False	False	False	False	000000	Server B	user002	/path/000000/a/b.csv
2/4/2022 6:39 AM	delete	False	False	False	False	000134	Server A	user003	/path/0000134/x/y/z.xlsx
1/10/2022 3:42 AM	write	False	False	False	False	000002	Server C	user001	/path/to/project/000002/a/b.txt
2/4/2022 1:36 AM	read	False	True	False	True	000001	Server D	user007	/work/000001/c/d.xml
1/26/2022 1:34 AM	create	False	False	False	False	000000	Server D	user003	/path/000000/a/b.xml
1/12/2022 1:38 PM	delete	False	False	False	False	000000	Server B	user008	/path/to/000000/file.txt

The page also aggregates detailed information about a dataset from DMS, usage logs, and other sources

Disclaimer: screenshot includes fake data



Deliverables

- Created features for info owners, project contacts, and started page for detailed tracking of issues
- Made reusable Angular components and established visual frontend patterns
- Developed and improved API (Django REST framework)
- Wrote API documentation (using Swagger)
- Developed test cases

Next Steps

- Improving and adding features for more stakeholders
- Continued collaboration with DMS to improve data quality