

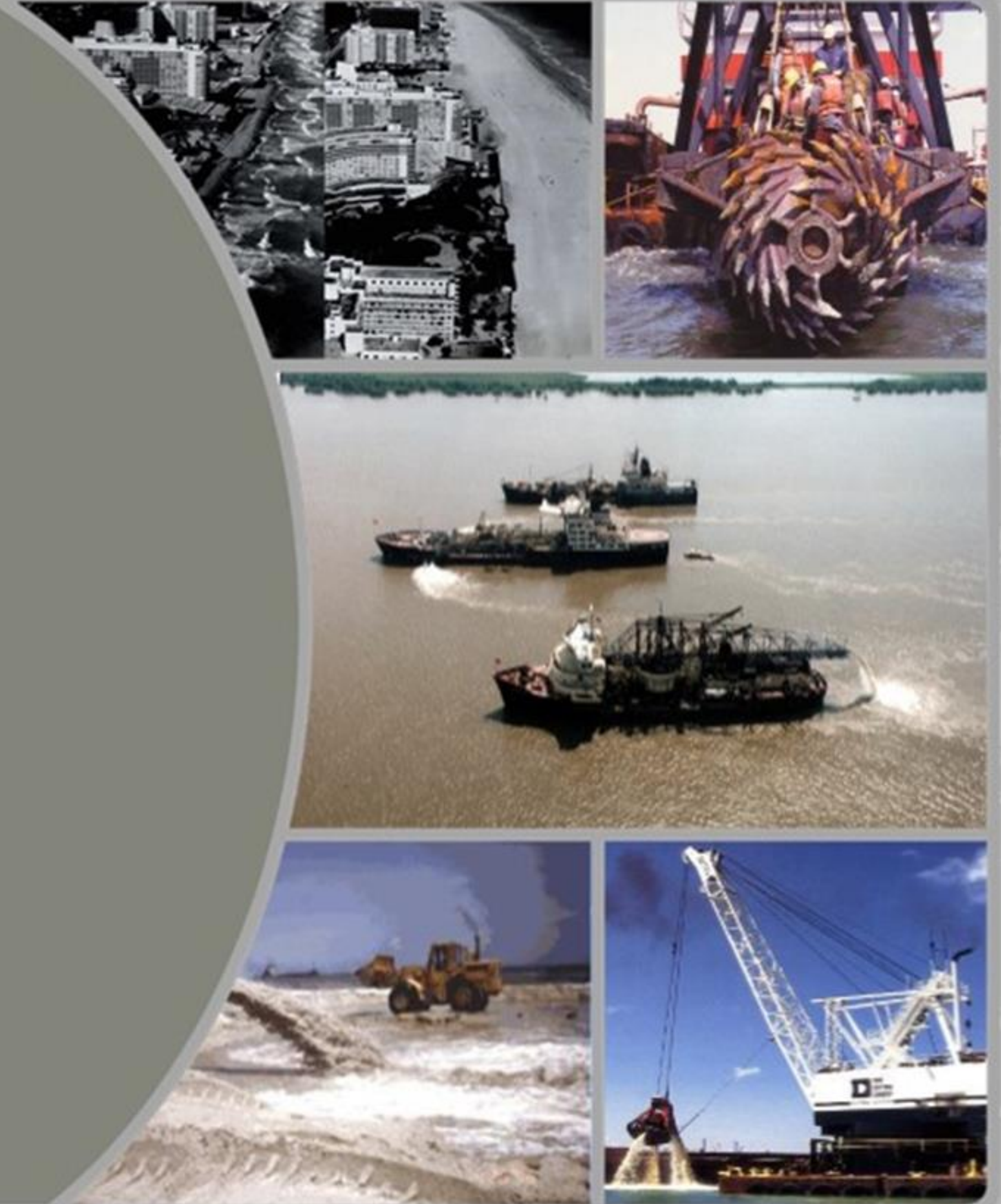
Dredging Data Extracts: A Century of USACE Dredging Activities

Brandan Scully, PE, PhD

Coastal & Hydraulics Lab

U.S. Army Engineer Research & Development Center

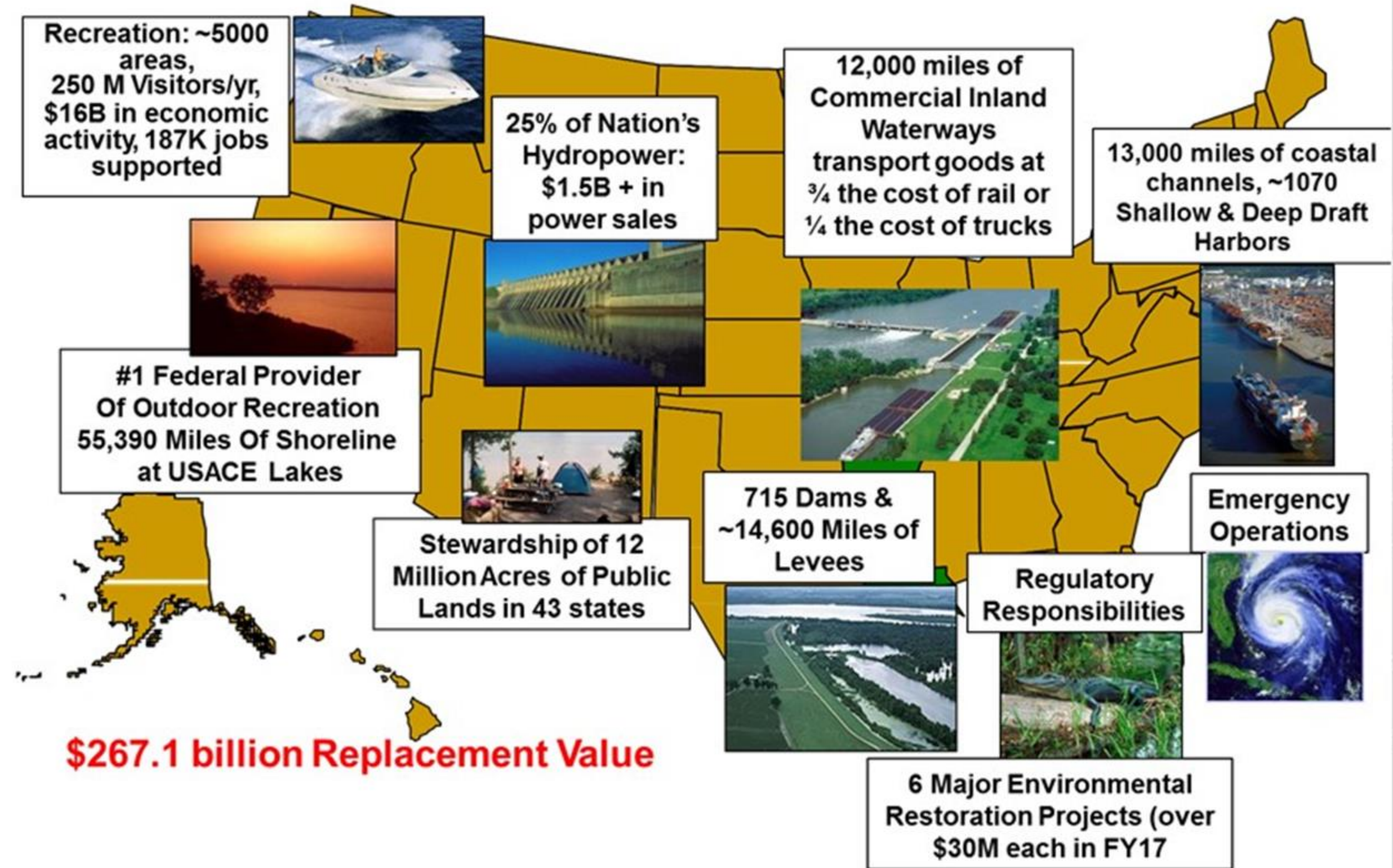
July 18, 2018



Innovative solutions for a safer, better world

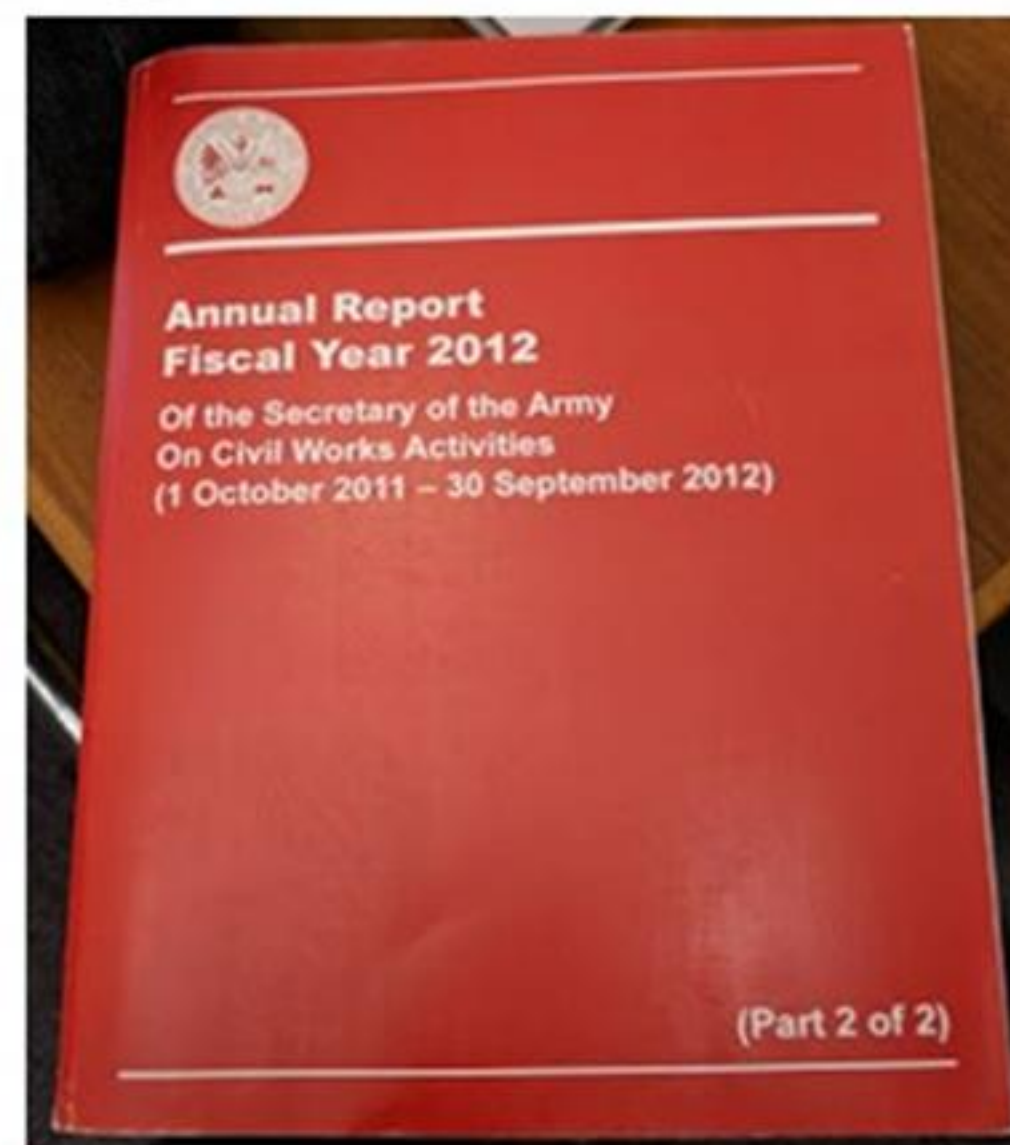
Outline

- USACE Civil Works Reporting
- Project Need
- Project Objectives
- Process
- Results
- Quality Control
- Applications
- Website
- Status & Future Efforts



USACE Civil Works Reporting

- USACE Civil Works (and civil works reporting) began in 1824
 - General Survey act of 1824
 - “the surveys, plans, and estimates, for each, when completed, to be laid before Congress”
 - Act for Improvements of the Ohio and Mississippi Rivers of 1824
 - “to be laid before Congress... after the commencement of each session, a statement of the proceedings under this act”
- USACE has a diverse and evolving mission portfolio
- USACE has a range of historical influences on Civil Works
 - Addition of Territories and the formation of States
 - Wars
 - Bear and Bull economies
 - Environmental Regulations (to name a few)

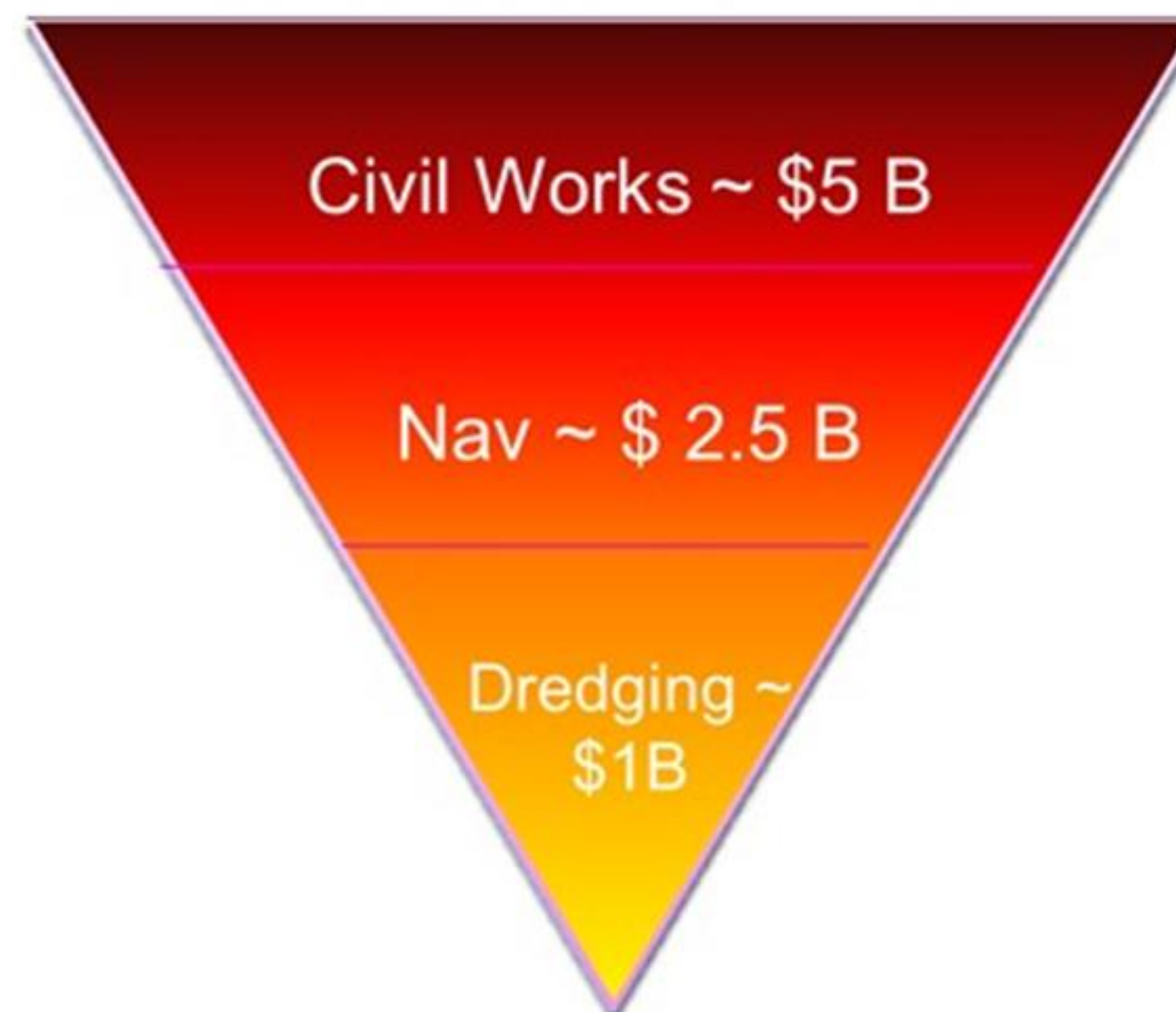


The Atlantic-Pacific Interoceanic Canal Study Commission, appointed on April 18, 1965, in compliance with Public Law 88-609, is studying the feasibility of a sea level canal in the American Isthmus, whether it should be built by nuclear or conventional means, the best location, and the estimated cost. The law limits the study time to June 30, 1968, and the cost to \$17,500,000. The Commission appointed the Chief of Engineers as the Agent for Engineering Feasibility, with the responsibility of coordinating the efforts of the Corps of Engineers, the Atomic Energy Commission, the Panama Canal Co. and other government agencies working on this phase of the study.

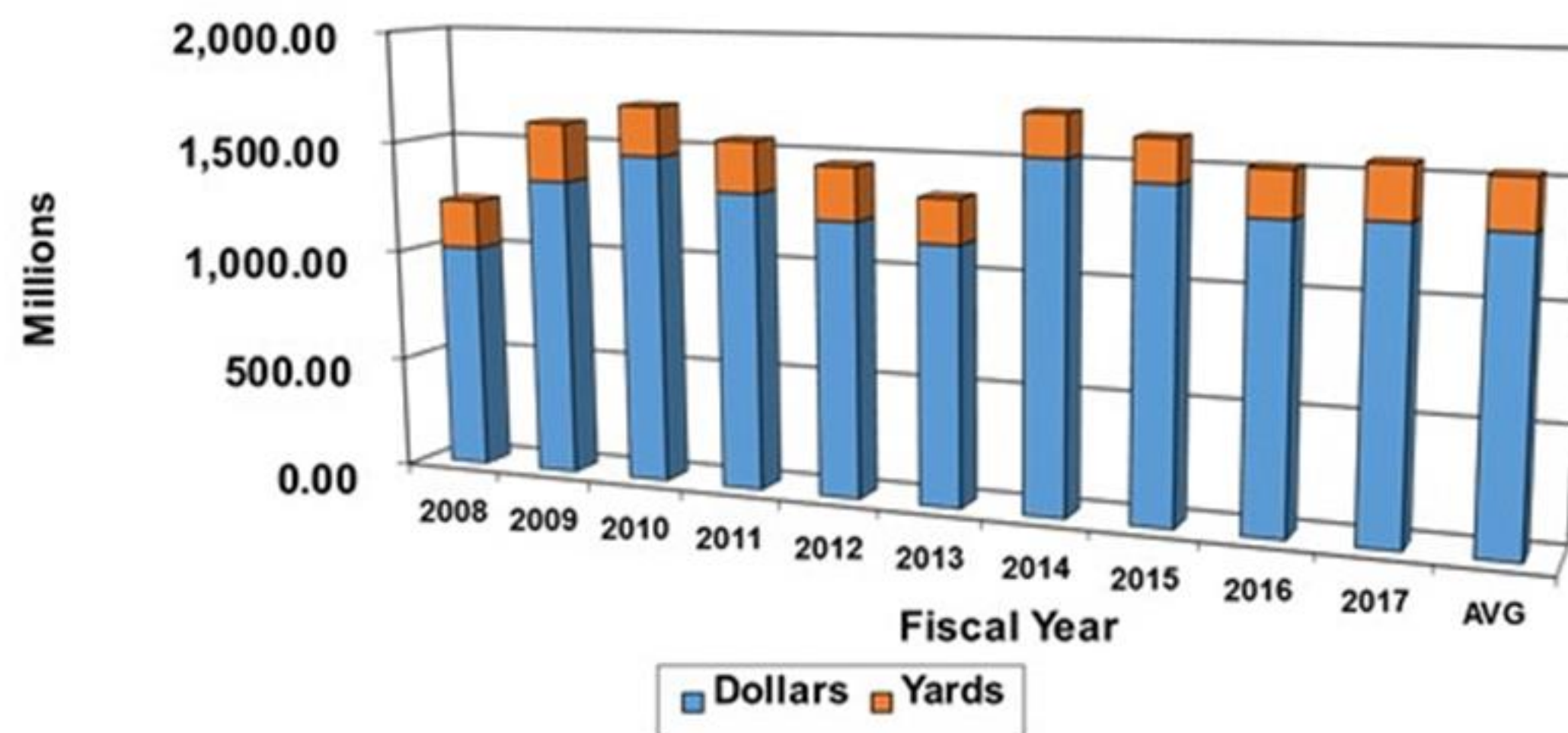
1966 Chief's Report

Project Need

- Increasing need to identify navigation operational efficiencies
 - Constrained operating environment
 - Increasing demands on aging infrastructure footprint
- Incomplete operating picture through time
 - Modern knowledge management ca. 1990's
 - Generational knowledge ca. 1960's
 - Functioning projects from the 19th century
 - Chief's reports are a comprehensive summary
- Routine uses of dredging data
 - Sediment budgets
 - Project design validations
 - Sedimentation analyses
 - Event driven impacts to sediment transport



Total Dredging FY 2008-2017



Project Objectives

- Demonstrate the use of Optical Character Recognition (OCR) and Natural Language Processing (NLP) to identify and liberate useful data from the corpus of annual reports.
- Identify and develop useful insights and data products from extracted data.
- Deliver extracted data and data model to users who will inform the direction of data use and future model development.

Annual reports 1848-2012:
~31 l.f. of shelf space in ERDC library
>100,000 pages
>30 GB of pdf documents
~1GB of text data



Process: Workflow

- Outline target data to extract for dredging and placement

- Date
- Division
- Project
- Location
- Operation Type
- Volume/Weight
- Unit of measure
- Cost
- Work type

- Model the relevant information

- Test the model

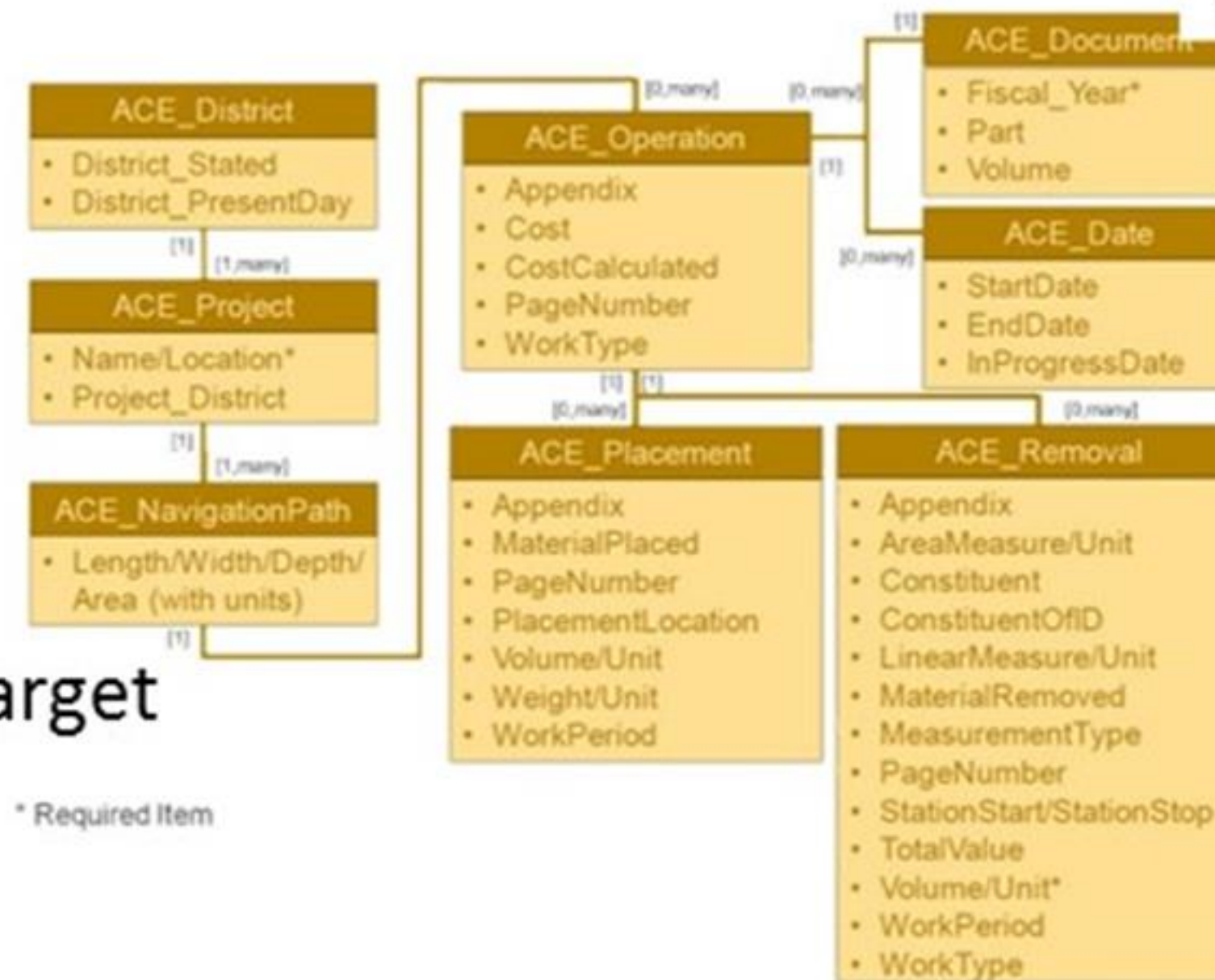
- Validate Results

- Identify error modes

- Improve the model and/or the target

- Scale model deployment

- Publish



sure, of about 11½ feet above or below the mean lake level prevailing at the time.

The cost of completed new work is \$483,000, exclusive of amounts expended on previous projects. The average annual maintenance cost during the past 5 years was \$25,659.

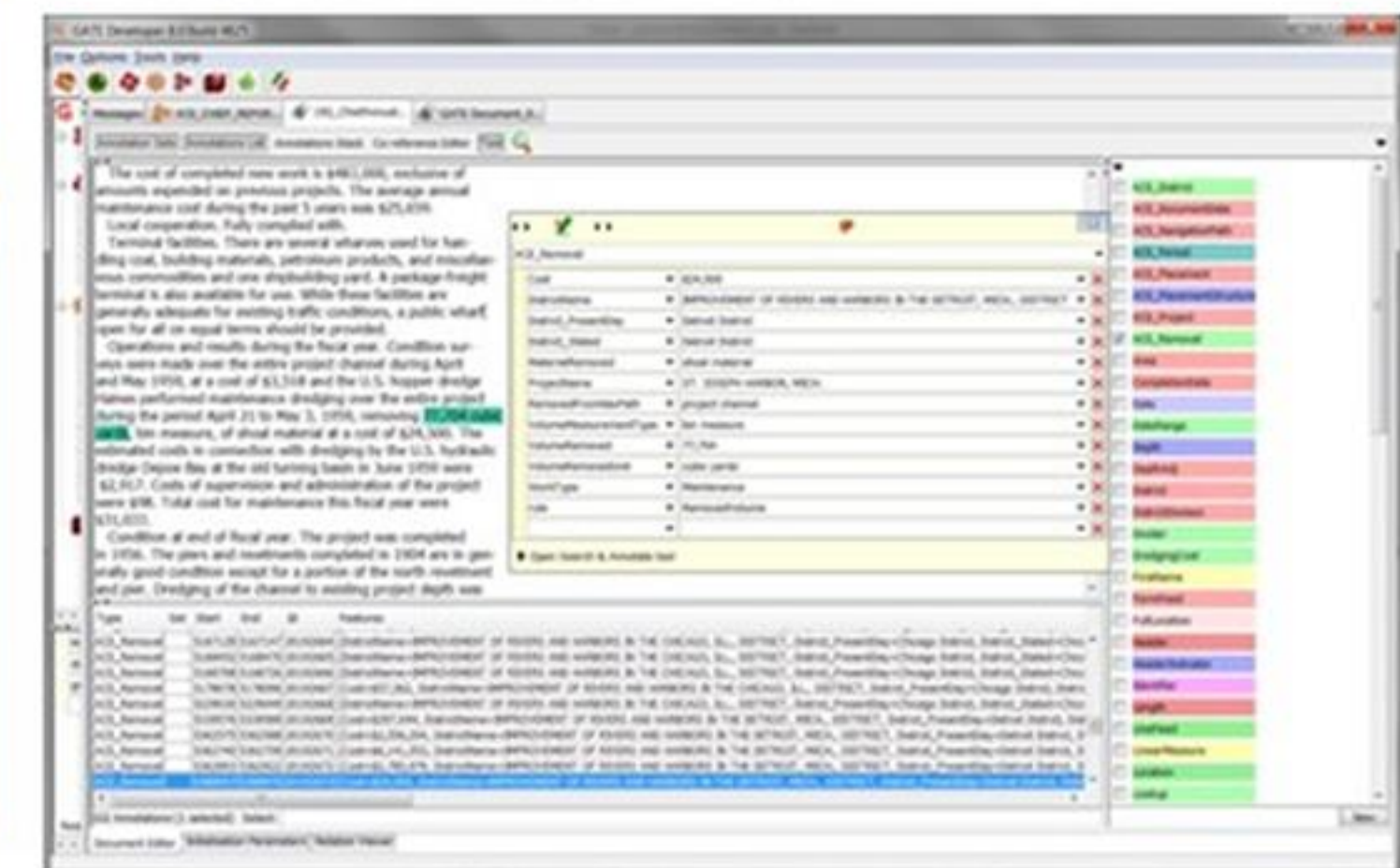
Local cooperation. Fully complied with.

Terminal facilities. There are several wharves used for handling coal, building materials, petroleum products, and miscellaneous commodities and one shipbuilding yard. A package-freight terminal is also available for use. While these facilities are generally adequate for existing traffic conditions, a public wharf open for all on equal terms should be provided.

Operations and results during the fiscal year. Condition surveys were made over the entire project channel during April and May 1959, at a cost of \$3,518 and the U.S. hopper dredge Haines performed maintenance dredging over the entire project during the period April 21 to May 3, 1959, removing 77,704 cubic yards, bin measure of shoal material at a cost of \$24,500. The estimated costs in connection with dredging by the U.S. hydraulic dredge Depoe Bay at the old turning basin in June 1959 were \$2,917. Costs of supervision and administration of the project were \$98. Total cost for maintenance this fiscal year were \$31,033.

Condition at end of fiscal year. The project was completed

GATE Interactive Test Environment & Automated Extraction



Results

- Operations, 1908-2000

- 41,774 records
- 15 fields

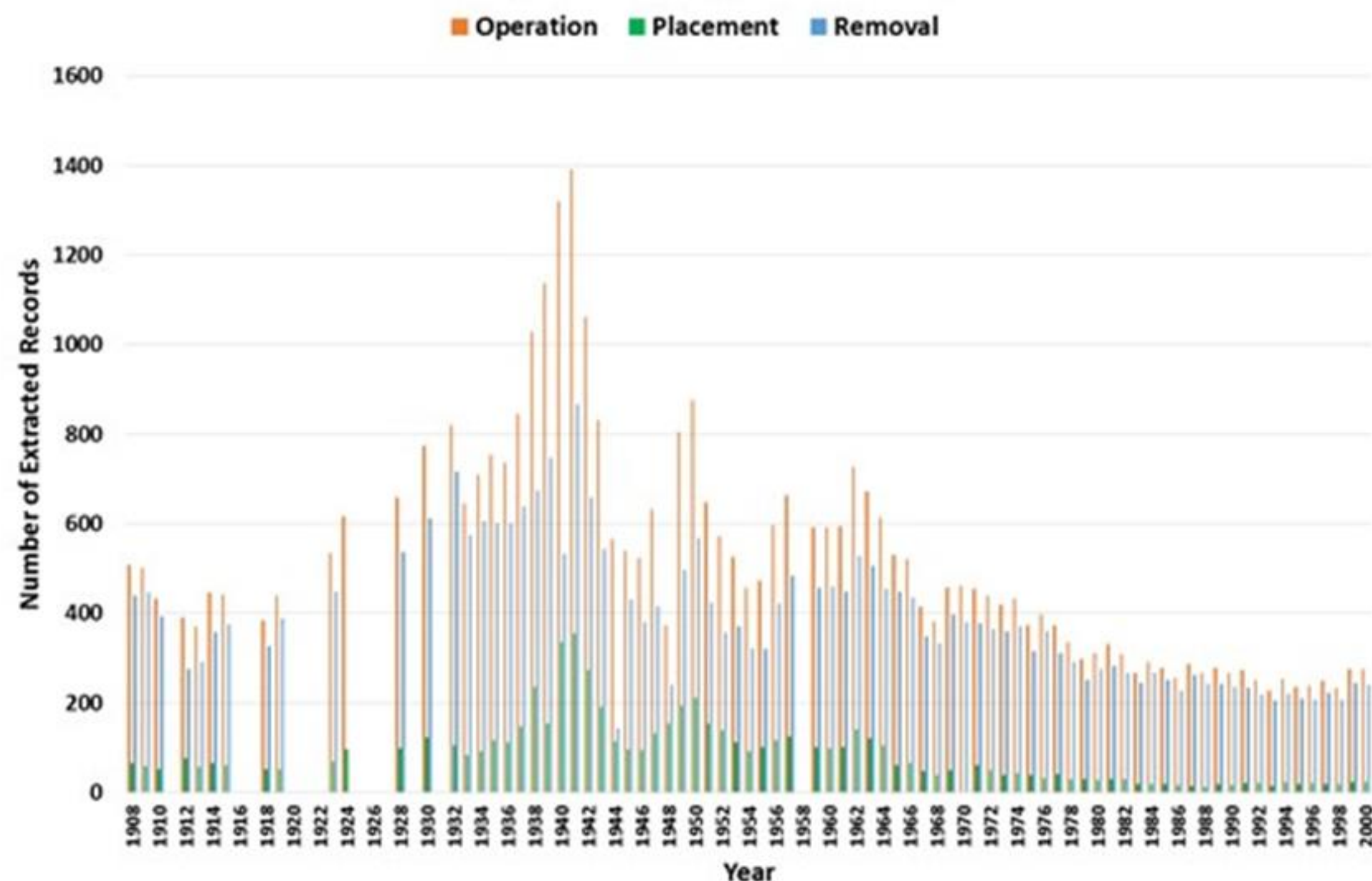
- Placements 1908-2000

- 6,692 records
- 24 fields

- Removals 1908-2000

- 31,248 records
- 27 fields

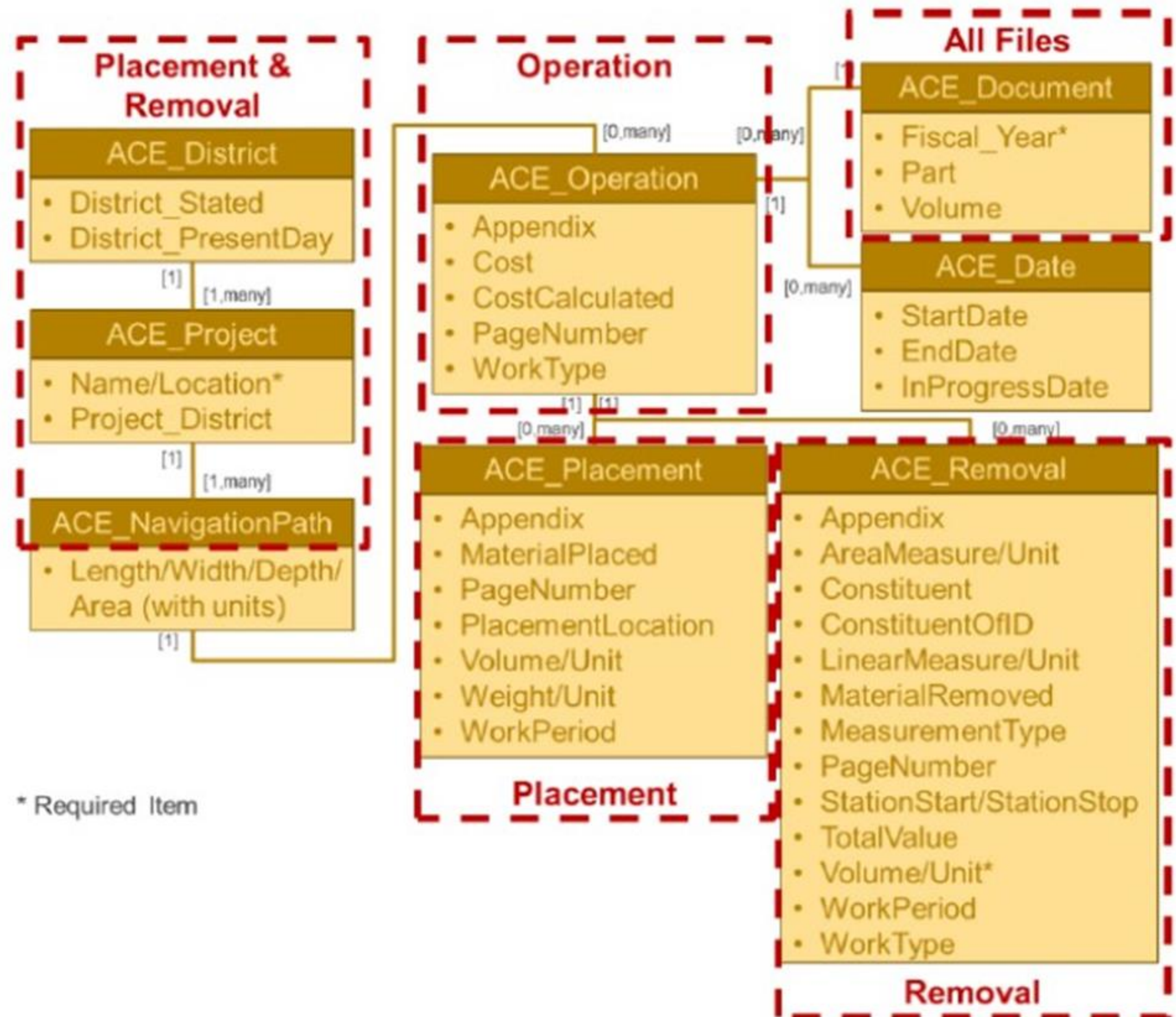
Removal, Placement, and Operation Records Extracted from the Annual Report of the Chief of Engineers US Army, on Civil Works, 1908-2000



	A	B	C	D	E	F	G	H	I	J	K
	ActionType	Stated District	Present Day District	ProjectName	NavigationPath	Volume	VolumeUnit	Measurement Type	Material	Work Type	Cost
1	Removal	New England District	New England District	ROCKLAND HARBOR, MAINE	14-foot channel	4,500	cubic yards		material		
2	Removal	New England District	New England District	ROCKLAND HARBOR, MAINE	HARBOR	3,100	cubic yards		ledge rock		
3	Removal	New England District	New England District	KENNEBEC RIVER, MAINE	27-foot channel	26,183	cubic yards		material		
4	Removal	New England District	New England District	NEWBURYPORT HARBOR, MA	12-foot channel and outer harbor	35,694	cubic yards		ordinary mate		
5	Removal	New England District	New England District	SALEM HARBOR, MASS	outer harbor	308,453	cubic yards		material		
6	Removal	New England District	New England District	BOSTON HARBOR, MASS	anchorage	1,797,816	cubic yards		material		
7	Removal	New England District	New England District	BOSTON HARBOR, MASS	anchorage	4,381,558	cubic yards		material		
8	Removal	New England District	New England District	WEYMOUTH FORE RIVER, MA	Rock Island Cove	1,065,402	cubic yards		material		
9	Removal	New England District	New England District	TOWN RIVER, MASS	turning basin	19,200	cubic yards		ledge and oth		
10	Removal	New England District	New England District	SCITUATE HARBOR, MASS	anchorage	74,900	cubic yards		material		
11	Removal	New England District	New England District	DUXBURY HARBOR, MASS	existing chan- nel	7,000	cubic yards		material		
12	Removal	New England District	New England District	CAPE COD CANAL, MASS	Hog Island Channel, Cleveland Le	176,440	cubic yards				\$135,498
13	Removal	New England District	New England District	HARBOR OF REFUGE, NANTUC	15-foot channel	70,547	cubic yards		material		
14	Removal	New England District	New England District	HYANNIS HARBOR, MASS	HARBOR	73,000	cubic yards		material		
15	Removal	New England District	New England District	BULLOCKS POINT COVE, RHOD	rubblestone dike	183,210	cubic yards		material		
16	Removal	New England District	New England District	POINT JUDITH HARBOR OF RE	15-foot entrance channel	62,534	cubic yards		material		
17	Removal	New England District	New England District	POINT JUDITH HARBOR OF RE	15-foot entrance channel	62,534	cubic yards		material		

Results: Extracted Data

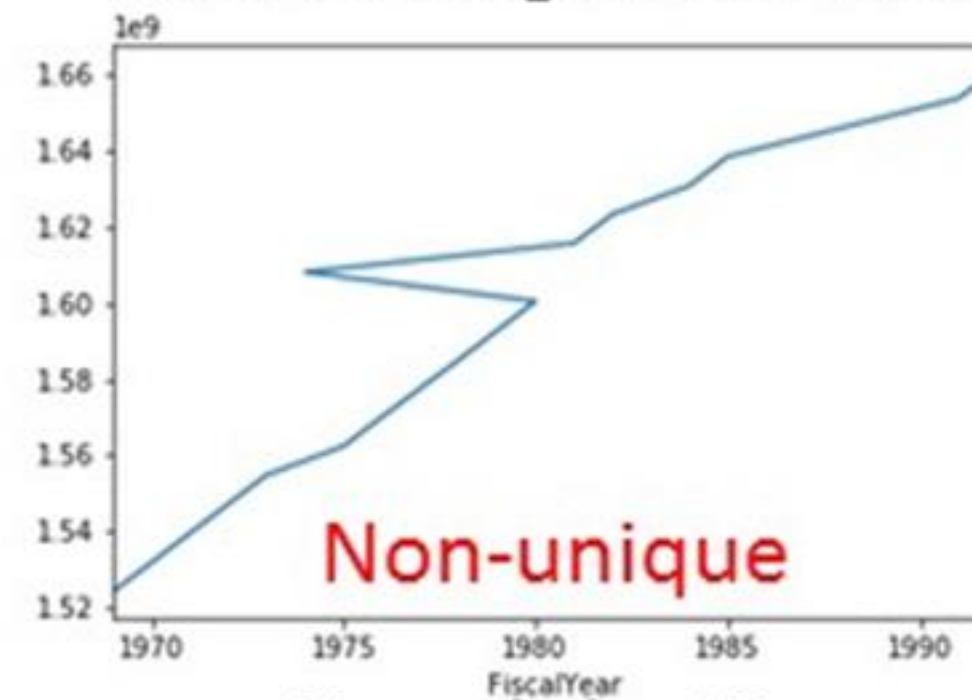
- Operations – tasks where dredging, placement, or a combination of the two tasks are identified
- Placements - tasks where material is intentionally moved to a location
- Removals - tasks where material is intentionally removed from a location



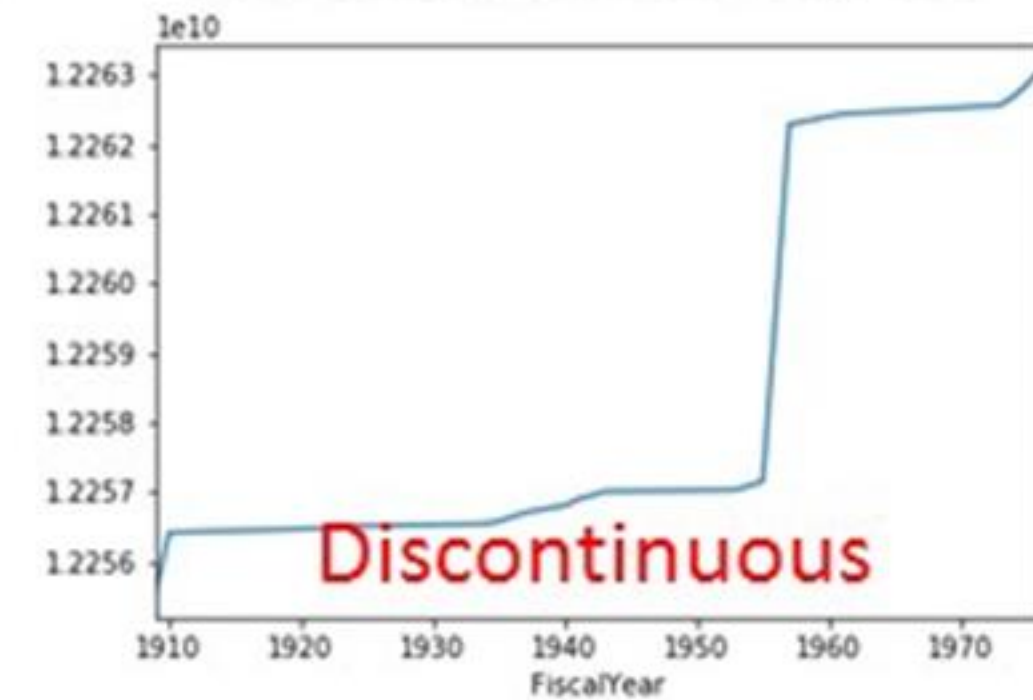
Quality Control: Data Validation

- Malformed Numbers
 - “2.078,221” is clearly wrong
 - Numbers beginning with 0
- Outlier Detection
 - Removals over 50,000,000 or less than 10
 - Placements over 10,000,000 or less than 10
- Aggregation
 - Sum volumes and operation record counts by year
 - Limit decadal volume differences of annual totals to 10x
- Data Gaps
- Visual Analysis
- Name Field Acrobatics
 - Uniqueify, Spell Check, Fuzzy String Matching

Baton Rouge Harbor, LA

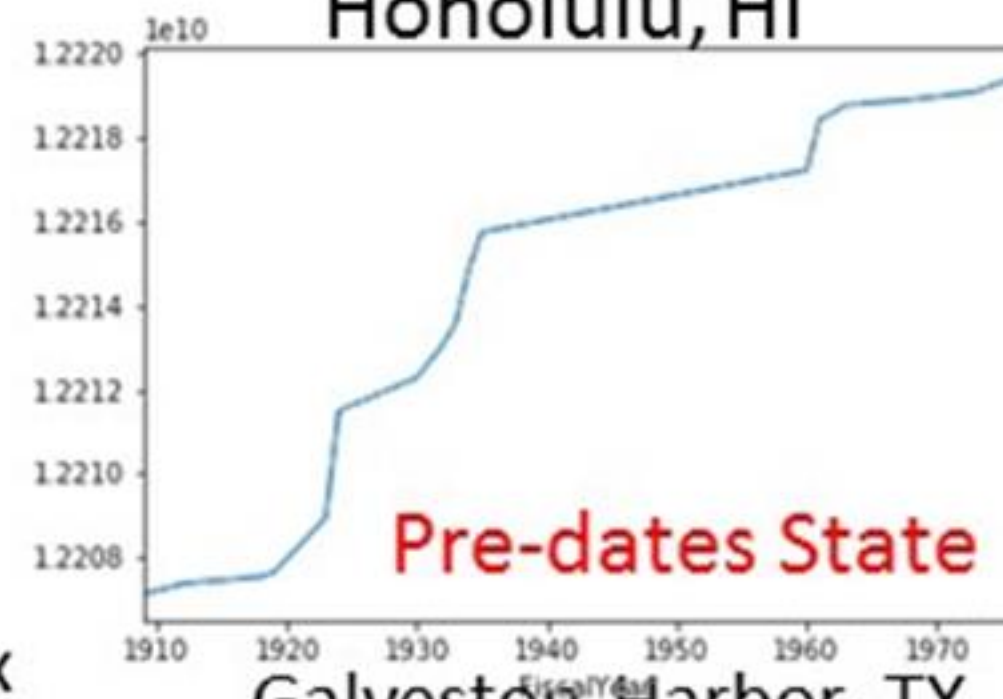


Housatonic River, CT

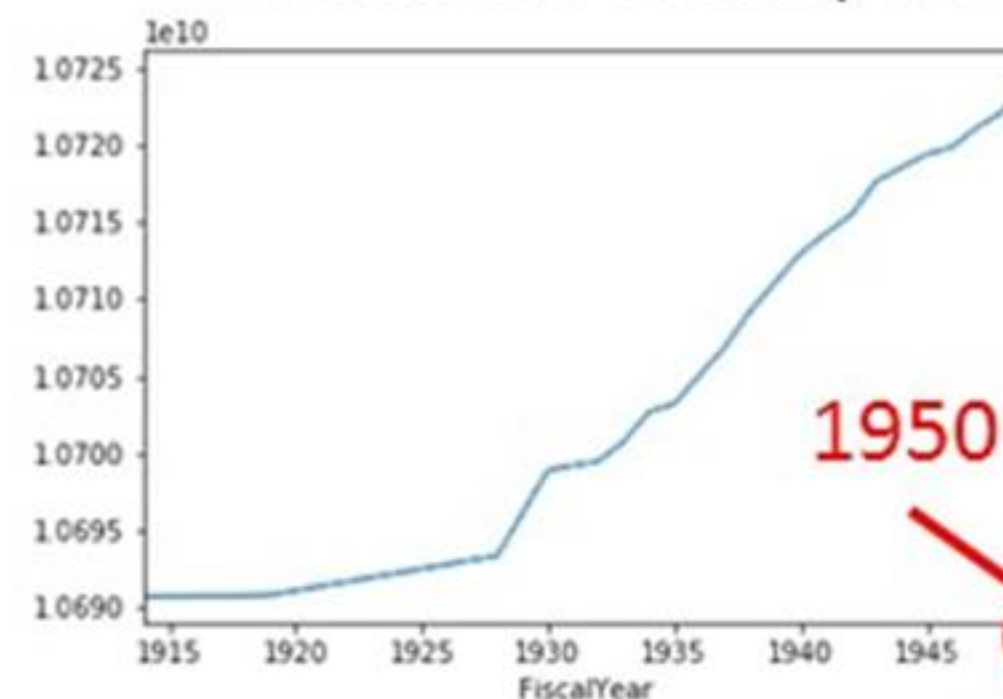


Investigate results
when they don't line
up with expectations.

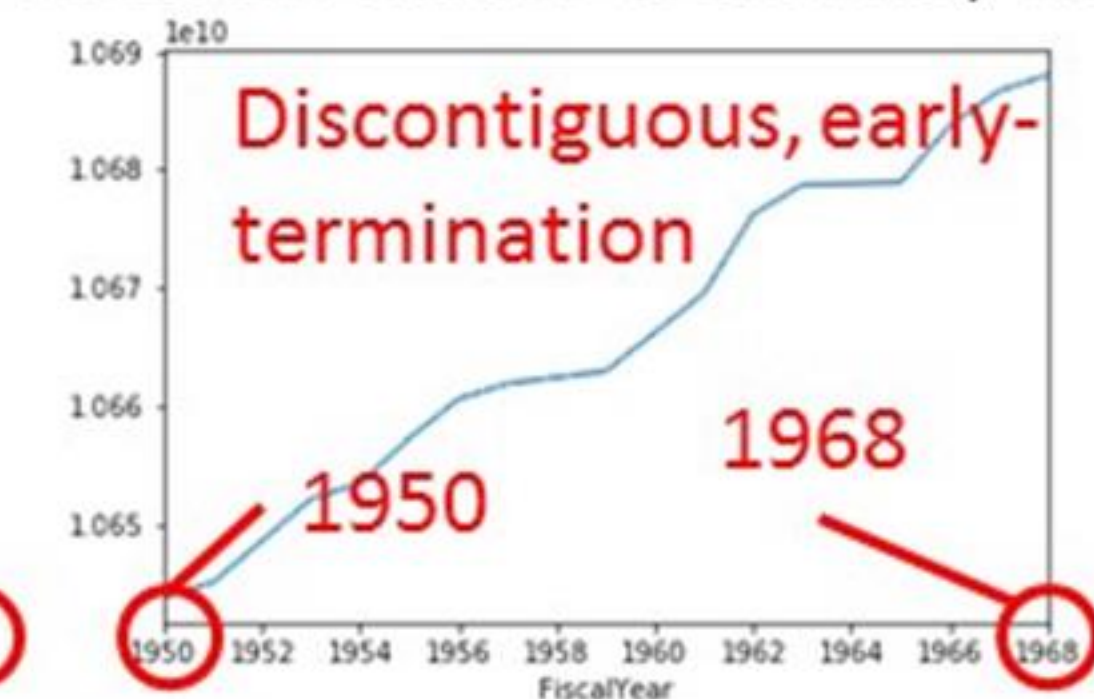
Honolulu, HI



Galveston Harbor, TX



Galveston Harbor & Channel, TX



Quality Control: What could possibly go wrong?

- Pre-OCR Source Material Errors

- incorrect words
- misspelled words
- improper characters
- introduction or omission of whitespace
- chipped ink
- hand annotation

Pre-OCR: target extracted as "2.078,221 cubic yards"

Agitation dredging: Four pump barges were operated in the Vicksburg district moving a total of 3,863,504 cubic yards of material (2;078,221 cubic yards at Duckport to Delta Point (598-602)¹ 1,538,688 cubic yards at Racetrack Towhead (606-612), and 246,595 cubic yards at Shipland Point (548), at a total cost of \$296,071.78.

- Post-OCR Data Extraction Errors

- Interpretation of stray marks as text
- Character misinterpretation
- Binding errors

Post-OCR: target extracted as "385bi" and "075 cubic yards"

Tacoma, Grays Harbor, and Olympia, Wash., revised 1963.)
Operations and results during fiscal year. Regular funds: Maintenance, hired labor: Channel condition surveys were made. U.S. hopper dredge *Pacific*, October 5, 1965, to March 7, 1966, removed 488,415 cubic yards from Sand Island shoal, and 385,075 cubic yards from Crossover channel, a total of 873,490 cubic yards, bin measurement. Maintenance, contract: Pipeline dredge *Robert Gray*, leased from Port of Grays Harbor Commission, removed 790,690 cubic yards of material from Chehalis River and north channel, July to December; and 30,000 cubic

Quality Control: What could possibly go wrong?

• Modeling Errors

- Precision Errors – model extracts correct information incorrectly
- Recall Errors – model fails to extract valid information

• Administrivia

- Addition, deletion, renaming
 - Projects
 - Districts
 - Divisions
 - Territories
 - States
- Multi-referencing
- Project Transfer

Recall Error: unexpected tabular arrangement

RIVERS AND HARBORS — DETROIT, MICH., DISTRICT 1361

sure, of about 1½ feet above or below the mean lake level prevailing at the time.

The cost of completed new work is \$483,000, exclusive of amounts expended on previous projects. The average annual maintenance cost during the past 5 years was \$25,659.

Local cooperation. Fully complied with.

Terminal facilities. There are several wharves used for handling coal, building materials, petroleum products, and miscellaneous commodities and one shipbuilding yard. A package-freight terminal is also available for use. While these facilities are generally adequate for existing traffic conditions, a public wharf open for all on equal terms should be provided.

Operations and results during the fiscal year. Condition surveys were made over the entire project channel during April and May 1959, at a cost of \$3,518 and the U.S. hopper dredge Haines performed maintenance dredging over the entire project during the period April 21 to May 3, 1959, removing 17,704 cubic yards, in measure of shoal material at a cost of \$24,500. The estimated costs in connection with dredging by the U.S. hydraulic dredge Depoe Bay at the old turning basin in June 1959 were \$2,917. Costs of supervision and administration of the project were \$98. Total cost for maintenance this fiscal year were \$31,033.

Condition at end of fiscal year. The project was completed

Annual Report Of The Chief Of Engineers, 1959, Volume 2

TABLE 15-1 GALVESTON 1972 DREDGING OPERATIONS

Section	Project	Description	Period	Cubic yards of material
2.	Beaumont Island Harbor, Texas	Entrance Channel (U.S. hopper dredge A. Mackenzie)	Aug 9 - Sept 19, 1971	394,382
		Entrance Channel (U.S. hopper dredge A. Mackenzie)	Jun 5 - 21 & Jun 26 - 30, 1972	346,000
		Brownsville Channel	Jul 17 - Aug 8, 1971	692,846
		Brownsville Channel	Jun 26 - 30, 1972	100,000
6.	Corpus Christi Ship Channel, Texas (Maintenance)	Entrance Channel and Bay Channel (U.S. hopper dredge A. Mackenzie)	Jul 1 - 25, 1971	564,855
		Bay Channel and Corpus Christi Turning Basin	Jul 1 - Sep 25, 1971	4,275,430
		La Quinta Channel and Basin w/new work	Jun 14 - 30, 1972	340,000
		La Quinta Channel and Basin w/maintenance	Jun 14 - 30, 1972	340,000
8.	Corpus Christi Ship Channel, Texas (New Work)	Entrance Channel (U.S. hopper dredge A. Mackenzie)	Oct 29 - Nov 28, 1971	306,657
	Freeport Harbor, Texas	Entrance Channel (U.S. hopper dredge A. Mackenzie)	Apr 17 - May 14, 1972	854,558
		Inside Channels and Basins	Oct 11 - Dec 2, 1971	587,570
9.	Galveston Harbor and Channel, Texas	Entrance, Outer Bar & Inner Bar Channels (U.S. hopper dredge A. Mackenzie)	Nov 29, 1971 - Mar 19, 1972	3,150,027
		Bolivar Roads to Pier "B"	Jul 1 - Oct 28, 1971	3,292,623
10.	Gulf Intracoastal Waterway Between Apalachee Bay, Fla., and Mexican Border (Galveston District)	Port Mansfield Entrance Channel (U.S. hopper dredge A. Mackenzie)	May 15 - 31, 1972; Jun 1 - 4 & Jun 21 - 25, 1972	278,769
		Channel to Victoria (Including Channel to Seadrift)	Jul 1 - 21, 1971	16,557
10.	Gulf Intracoastal Waterway Between Apalachee Bay, Fla., and Mexican Border (Galveston District) (cont'd)	Arroyo Colorado to Mud Flats and Channel to Port Mansfield	Completed prior Fiscal Year (Jun 30, 1971)	23,653
		Freeport Harbor to Mile 558	Oct 18 - Nov 30, 1971	224,454
		Channel to Port Mansfield	Jul 1 - 15, 1971	420,066
		Older Lakes to Colorado River	Aug 11 - Dec 5, 1971	842,699

Annual Report Of The Chief Of Engineers, 1972, Volume 2

Applications

- Sedimentation Analyses

- Zakikhani et al. (2011)
- Tate et al. (2014)

- Project Design Validation

- Zakikhani et al. (2011)

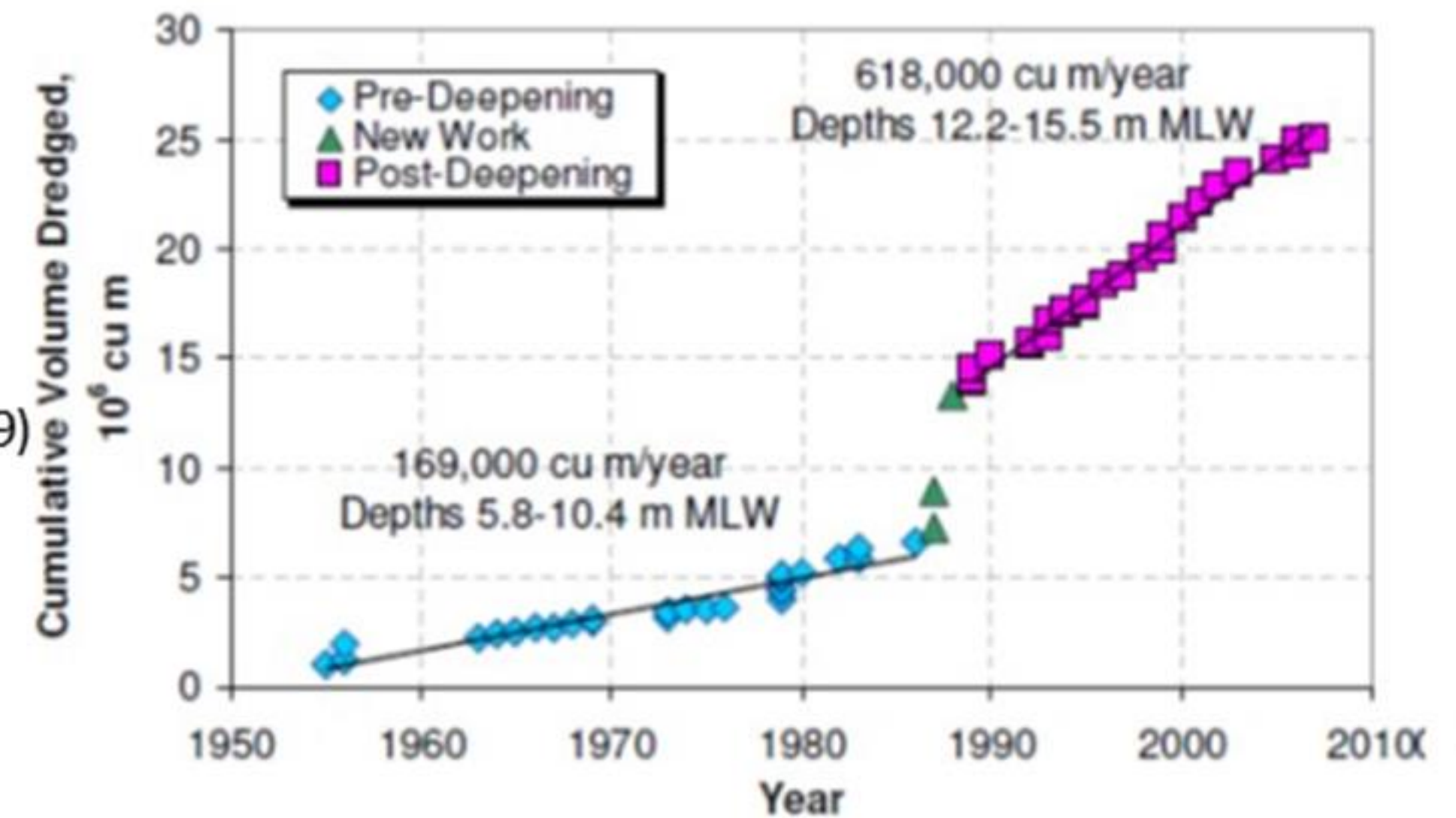
- Long Term Sediment Budgets

- Patterson (1983)
- Rosati (2005)

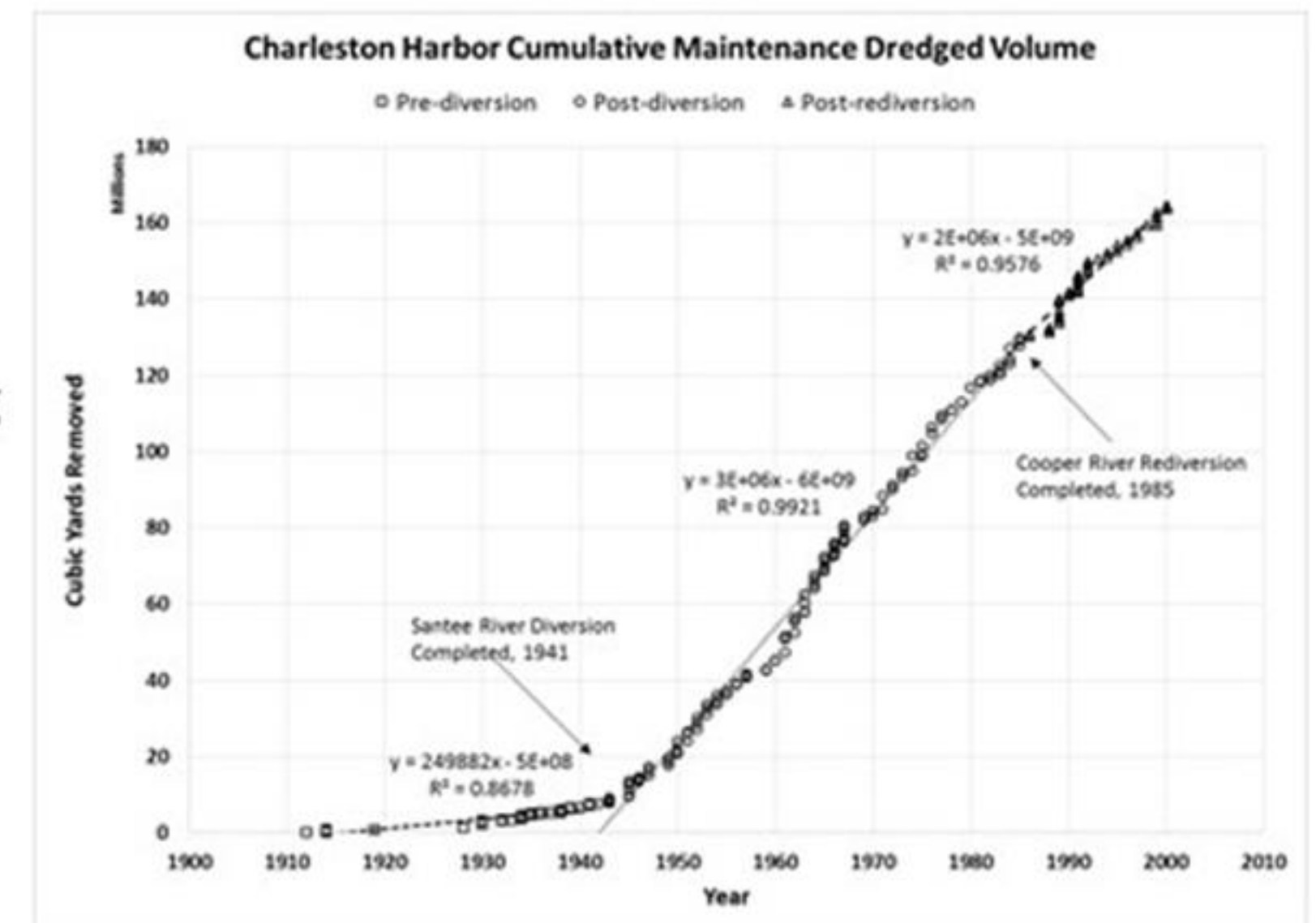
- Event Driven Impacts

- Rosati & Kraus (2009)
- Andrew Morang et al. (2012)

Sedimentation rate change following deepening and widening (Rosati & Kraus, 2009)

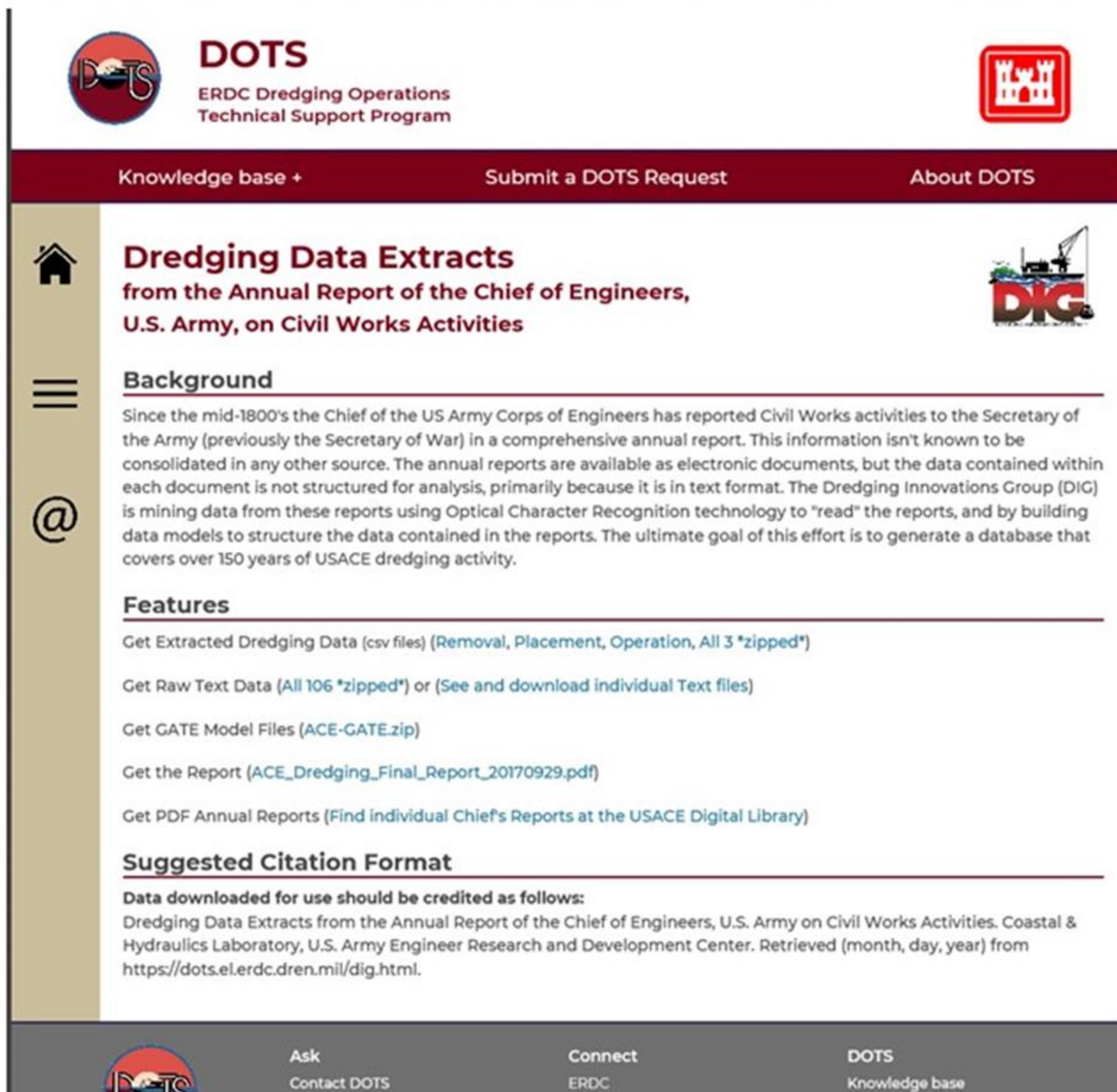


Sedimentation rate change following hydrologic routing change, derived from CR extracted data.



Website

- Project documentation at <https://dots.el.erdc.dren.mil/dig.html>
- Available for download:
 - Extracted dredging records
 - Extracted text data
 - GATE NLP model
 - Contractor metadata report
 - Link to PDF reports at USACE Digital Library



The screenshot displays the DOTS (Dredging Operations Technical Support Program) website. The header includes the DOTS logo, the text "ERDC Dredging Operations Technical Support Program", and a red castle icon. The navigation bar contains links for "Knowledge base +", "Submit a DOTS Request", and "About DOTS". The main content area features a sidebar with a home icon, a menu icon, and an @ icon. The main text area is titled "Dredging Data Extracts from the Annual Report of the Chief of Engineers, U.S. Army, on Civil Works Activities" and includes a "Background" section, a "Features" section with links to download data, and a "Suggested Citation Format" section. A "DIG" logo is visible in the top right corner of the content area.

DOTS
ERDC Dredging Operations
Technical Support Program

Knowledge base + Submit a DOTS Request About DOTS

Dredging Data Extracts
from the Annual Report of the Chief of Engineers,
U.S. Army, on Civil Works Activities

Background

Since the mid-1800's the Chief of the US Army Corps of Engineers has reported Civil Works activities to the Secretary of the Army (previously the Secretary of War) in a comprehensive annual report. This information isn't known to be consolidated in any other source. The annual reports are available as electronic documents, but the data contained within each document is not structured for analysis, primarily because it is in text format. The Dredging Innovations Group (DIG) is mining data from these reports using Optical Character Recognition technology to "read" the reports, and by building data models to structure the data contained in the reports. The ultimate goal of this effort is to generate a database that covers over 150 years of USACE dredging activity.

Features

Get Extracted Dredging Data (csv files) ([Removal](#), [Placement](#), [Operation](#), All 3 *zipped*)

Get Raw Text Data (All 106 *zipped*) or ([See and download individual Text files](#))

Get GATE Model Files ([ACE-GATE.zip](#))

Get the Report ([ACE_Dredging_Final_Report_20170929.pdf](#))

Get PDF Annual Reports ([Find individual Chief's Reports at the USACE Digital Library](#))

Suggested Citation Format

Data downloaded for use should be credited as follows:
Dredging Data Extracts from the Annual Report of the Chief of Engineers, U.S. Army on Civil Works Activities. Coastal & Hydraulics Laboratory, U.S. Army Engineer Research and Development Center. Retrieved (month, day, year) from <https://dots.el.erdc.dren.mil/dig.html>.

Ask
Contact DOTS

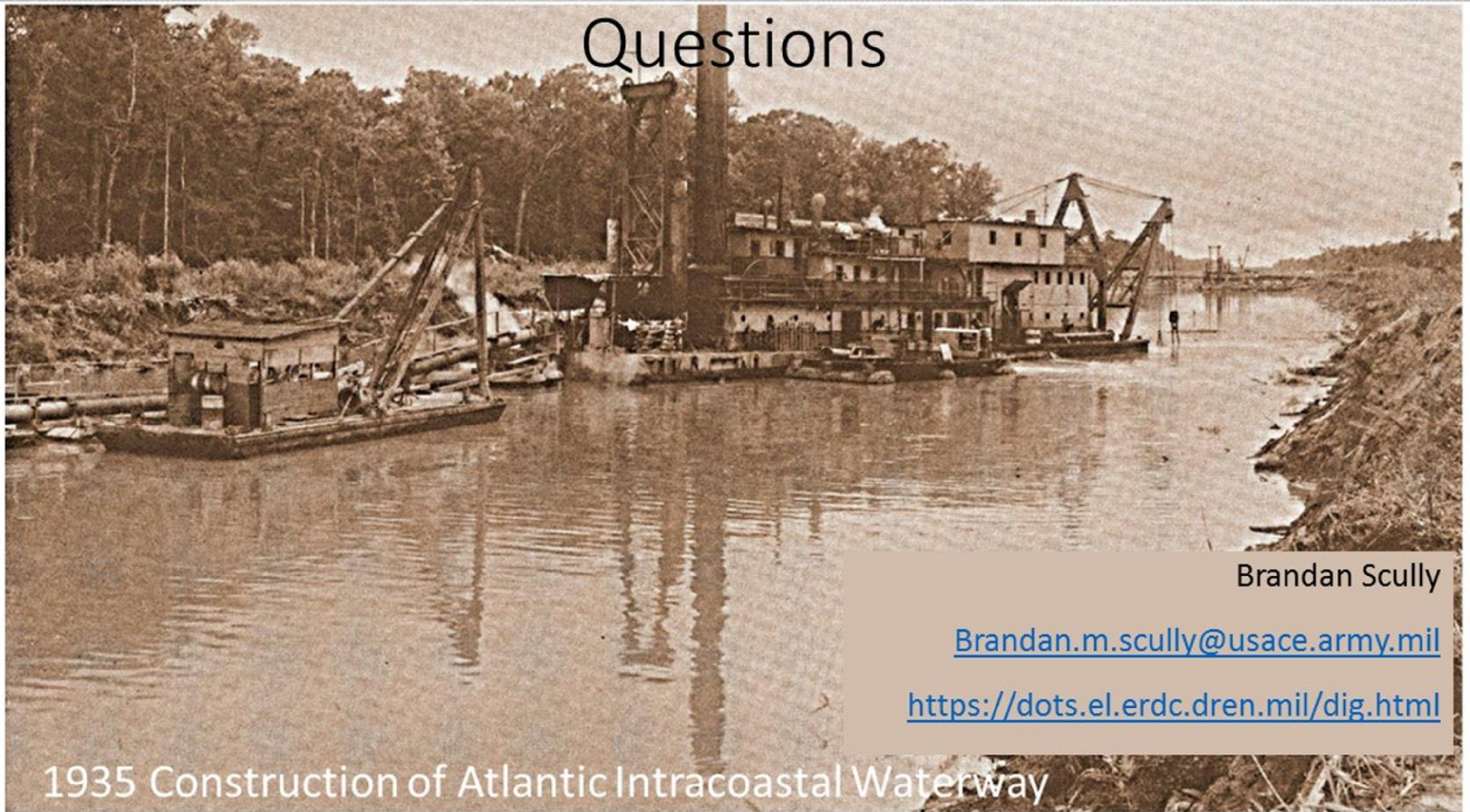
Connect
ERDC

DOTS
Knowledge base

Status & Future Efforts

- Site is live – go play with the data
- Continuous improvement on data quality
- Expansion of data gap coverage
 - Goal is comprehensive dataset
- Improvement/expansion of model competence
 - We appreciate feedback. See the ask on last slide!
- Query and visualization interface
- Integration with contemporary databases

Questions



Brandan Scully

Brandan.m.scully@usace.army.mil

<https://dots.el.erdc.dren.mil/dig.html>

1935 Construction of Atlantic Intracoastal Waterway

Our Ask

- Help us improve:
 - What would be interesting topics to extract from the corpus?
 - Do you have data that you think we missed?
 - Did we find data that conflicts with your local information?
 - Are you interested in expanding the GATE model to extract new features?
- Thanks!

Brandan Scully

Brandan.m.scully@usace.army.mil

<https://dots.el.erdc.dren.mil/dig.html>