

Archival & Discovery of Tsunami Events on Marigram Microfilm Rolls

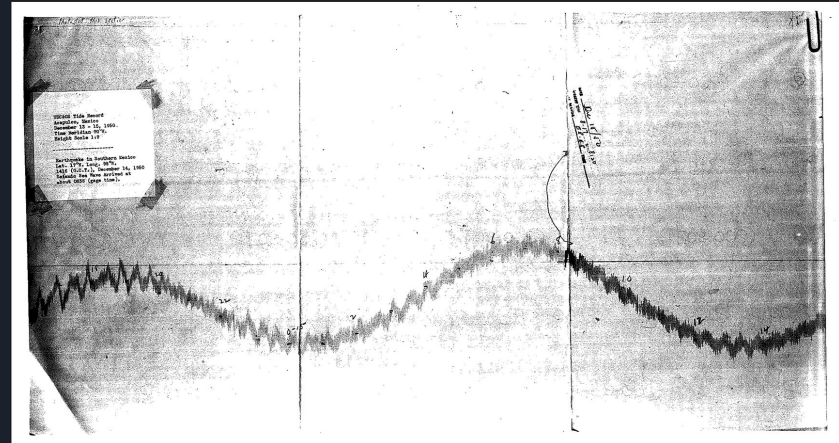
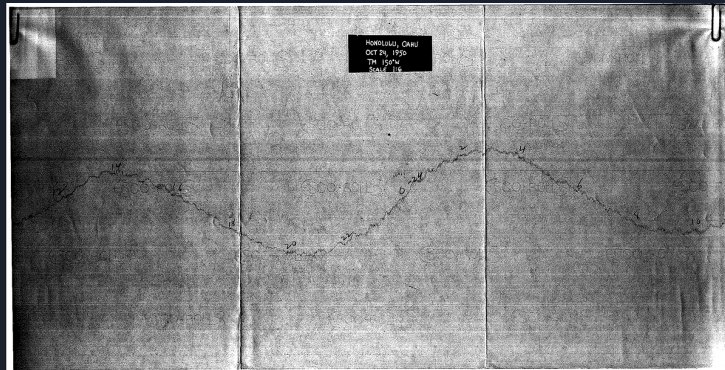
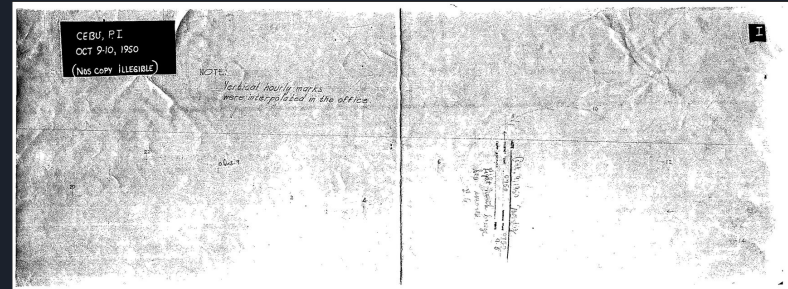
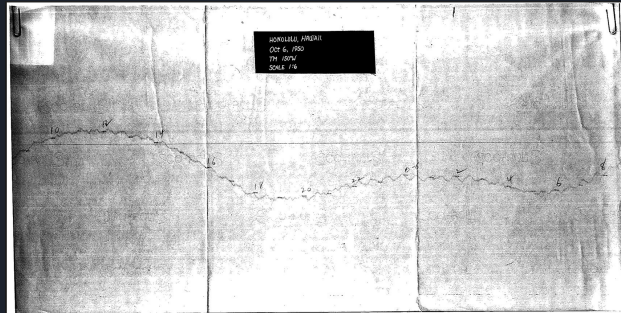
Aaron Sweeney, NOAA
Erik Radio, CU Boulder
Moukthika Gunapaneedu, CU Boulder (Data
Science)



Project Description

- 3700 Marigrams held by National Centers for Environmental Information
- Prior to 1994, tsunami data only resided on marigrams
- Move beyond digitization to data curation
- Provide important tsunami data which can aid in predictive modeling

Marigrams





Process

- Data Preprocessing
 - Deduplication,
 - Image correction (OpenCV, PIL)
- Metadata Extraction
 - Tesseract OCR
 - Data matched against NOAA/NCEI station names
 - Regular expression for data reconciliation



Impacts

- 3,959 marigrams cataloged
- Synthesis of three distinct fields
 - Oceanography (Marigram history and contextualization)
 - Data science (data extraction and processing)
 - Metadata/Information science (documentation, data curation)