Accenture and Arolsen Archives: challenges handout
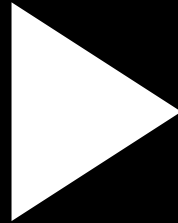
Hertie School | Data4Good Festival

20 January 2025

#everynamecounts

https://www.youtube.com/watch?v=iGrOozqLqtM

We leverage AI for Every Name Counts

The Arolsen Archives contains the largest database of documents related to Nazi persecution (~110M) and has the goal of making them universally accessible.

We used a patented AI for Document Processing Application relying on:

- Deep Learning
- OCR
- Fast Name Matching
- Geolocation and Translation

Further Links:

- **https://www.accenture.com/us-en/case-studies/public-service/arolsen-archives**
- **https://everynamecounts.arolsen-archives.org/en/**

2.7M

Documents classified to date, with 1.5M documents with data extracted
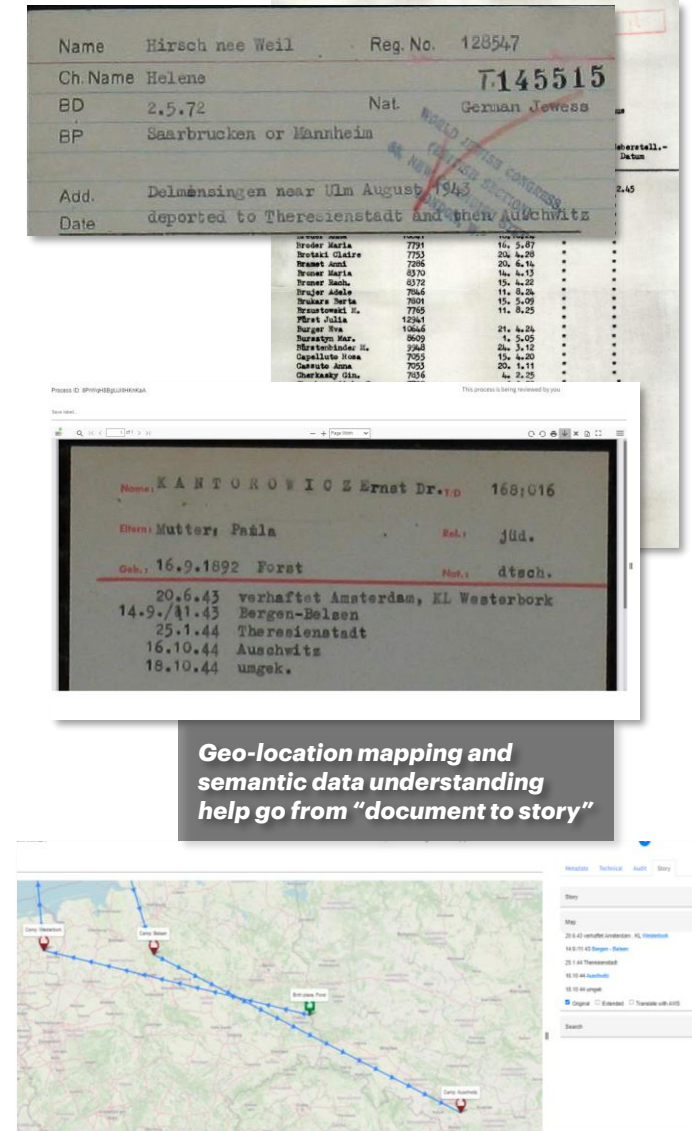
82%

Straight-through processing with OCR and AI

40X

Increased speed in documents indexed per hour with human validation, for remaining 18%

76 Years

Saved of manual processing time

Geo-location mapping and semantic data understanding help go from "document to story"

The dataset contains (raw) textual data of the digitalised tracing cards

> ☐! Please read, sign, and comply with the data usage agreement

Subset of **digitalised tracing cards**, including:

- Names

- Family members

- Nationality

- Religion

- Location record (incl. tagged geolocations)

- OCR Confidence score

Tips:

- Use the Column Description document to better understand each variable.

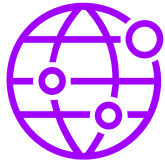- You are working with a randomly sampled subset.

Reminder: Use this data respectfully, as it contains sensitive information of victims of NS persecution!

Choose your challenge or create your own!

Identifying Patterns in Imprisonment Locations During the Holocaust

Connect fragmented or partial data about a person's imprisonment or transport route. Create a visualization tool that uses historical data from multiple tracing cards to map patterns of movement for victims, potentially revealing new historical insights

Potential Outcomes

- A geographic visualization that connects various tracing card records, showing the routes victims of the Holocaust took.
- AI tools that predict missing locations based on other available data, providing suggestions to aid volunteers in completing records.
- Insights into the transportation networks and incarceration practices used during the Holocaust.

Matching Historical Data with Contemporary Identities

Link historical records with modern data and stories shared by victims. Create a system to detect and match name variations or similar data to modern resources.

Potential Outcomes

- A data matching tool that uses machine learning to detect similarities in names, even with misspellings or variations, linking historical victims to living descendants.
- A database of stories and related cards.
- Tools that aid genealogical researchers, historians, and volunteers working on Holocaust data to connect modern identities with the past
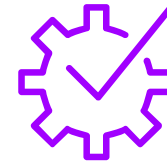
Automating the Detection of Inconsistencies in Holocaust Documentation

Develop an AI system that flags potential discrepancies across different tracing card records (e.g., different spellings of names, conflicting birth dates, or varied imprisonment locations) and suggests possible corrections.

Potential Outcomes

- A machine learning system that flags potential inconsistencies across records, suggesting areas that need human review.
- Insights into common data errors and inconsistencies, and how to improve future digitization efforts.

Join our partner talk on
Data & AI in the public
sector and the EU AI Act

We look forward to
hear about your
projects.