

Harnessing the Power of OpenAleph

Simon Wörpel & Alex Ștefănescu

Data and Research Center, 25.05.2025



Who are we?

Simon Wörpel (simon@dataresearchcenter.org)

- Director of Technology
- Investigative Data Journalist, ex-CORRECTIV

Alex Ștefănescu (alex@dataresearchcenter.org)

- Senior Developer
- Tech activist with ApTI, ex-OCCRP

What is OpenAleph?

Let's address the elephant in the room:

- Soft fork of the OCCRP Aleph tool
- **We commit to the commons**
- Open-source
- Community-oriented
- [Join us on Discourse](#)

What is OpenAleph?

Let's address the elephant in the room:

- Why did we fork?
 - Taking investigative tooling in new directions
 - **Committing to open-source**
 - Experiments, plug-ins
 - OpenAleph and Aleph remain compatible

What can OpenAleph do?

- We updated the user documentation
- It now includes video tutorials
- We will publish more in-depth guides and "follow along" workflows

openaleph.org/docs/

What can OpenAleph do?

Processing and extraction

- Different file formats, e-mail archives
- Extract interesting names automatically (as Mentions)
- OCR
- Making PDFs safe to browse inside OpenAleph

What can OpenAleph do?

Structured data

- Common schema: [FollowTheMoney](#)
- Map tabular data into entities
- Stream data into and out of OpenAleph
- Allows us to build a **Library of Commons**
- **Cross-reference**

What can OpenAleph do?

Search through data

- Flexible search (supports Elasticsearch syntax)
- Filter results (by language, dataset, file type etc.)
- Automatically uses transliterations of names

There's more?

Additional tooling

- An improved [command-line client](#)
- [ftmq](#) to query FollowTheMoney data from different stores directly
- [nomenklatura](#) (by OpenSanctions) to dedupe and store FollowTheMoney a different way

Where do we go from here?

New features

- Making audio & video files searchable
- Providing transcripts
- Geo-coding data, displaying distance
- Putting a face to the name through WikiData

OpenAleph

Making Audio and Video Files Searchable

We kept hoping that audio transcription tools would eventually deliver faithful textual representations of audio streams. We kept hoping that ✨AI✨ would solve this problem.

Published at 5/22/2025, 12:00:00 PM

OpenAleph

development

AI Transcriptions in OpenAleph: the First Steps

We need OpenAleph to be able to transcribe audio and video files, and we're turning to AI to help us do it. We're in store for a bumpy ride.

Published at 5/22/2025, 12:00:00 PM

Where do we go from here?

Housekeeping

- Improve maintenance, stability,
- Refactoring for a better developer experience
- Break it apart and built it up again - iteratively

Where do we go from here?

Work in progress

- Improved cross-referencing
- New FollowTheMoney schemas for instant messaging and social media scraping
- Improved UI

Where do we go from here?

Thinking big

- "Pocket OpenAleph"
- Federating the Library of Commons
- AI-powered data summarization, categorization

What do you want from OpenAleph?

Join the conversation

darc.social

- Join the conversation
- Suggest new features
- How are you using OpenAleph in your workflow?

Stay up to date

openaleph.org/blog

- New features
- Technical behind-the-scenes

darc.li/news

- Newsletter

Thanks!

<https://darc.li/dh25-openaleph>

