Funded by
the European Union

SUPPORTED BY NGI

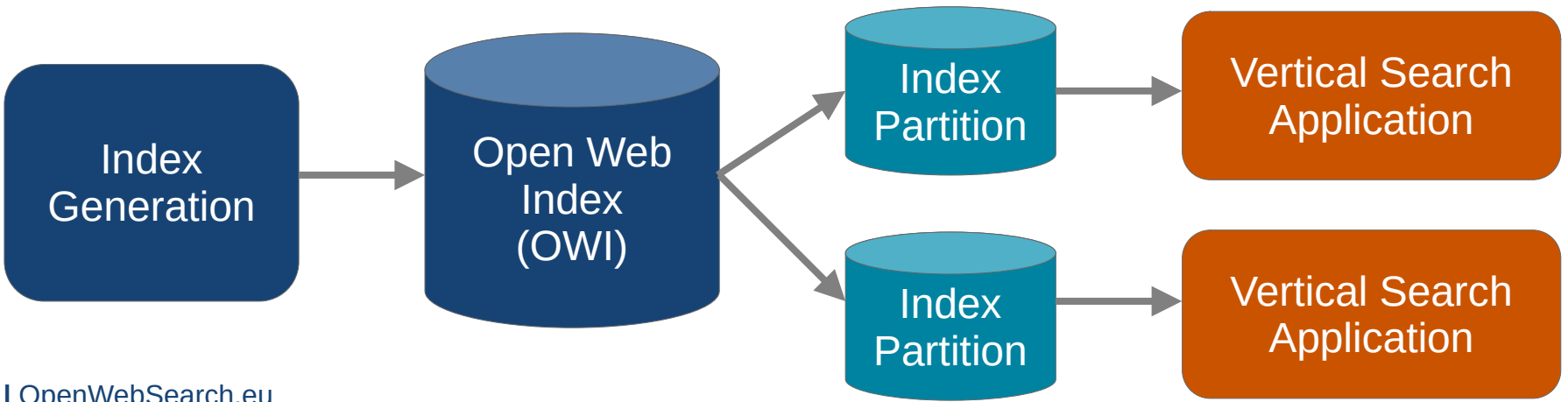
Conceptual Design and Implementation of a Prototype Search Application using the Open Web Index

Alexander Nussbaumer, Rohit Kaushik,
Gijs Hendriksen, Sebastian Gürtl, Christian Gütl

5th International Open Search Symposium (OSSYM)
4-6 October 2023

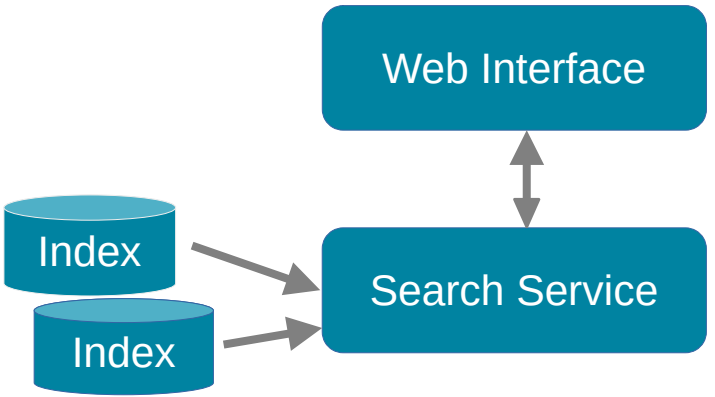
Vertical Search Engines in OWS

- Vertical Search Engines serve special purposes, topics, or domains, e.g. product search
- Search applications are a key concept in OpenWebSearch.eu
- Based on Open Web Index and index partitions



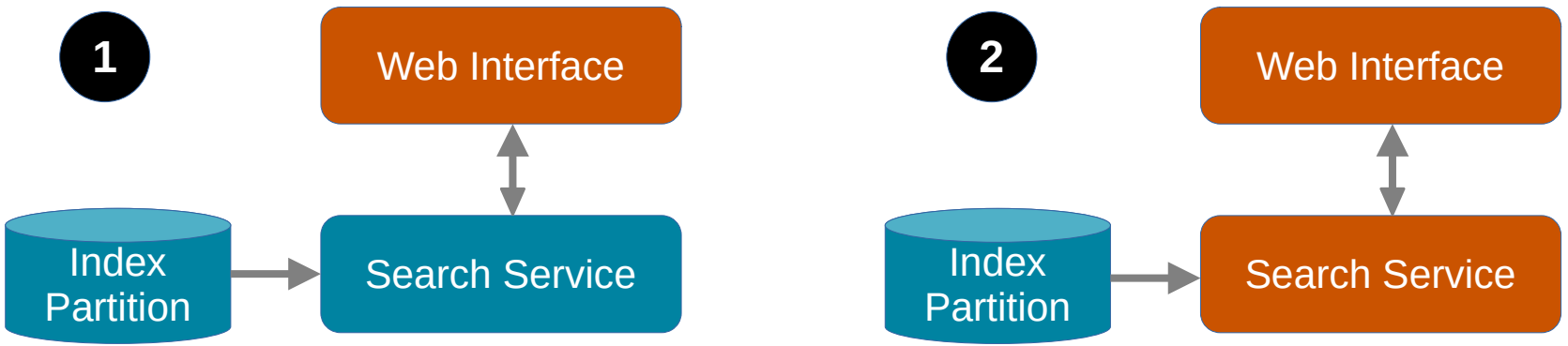
Prototype Search Application

- Minimal implementation of a OWS vertical search engine
  - Index partitions in reasonable size on a local server
  - Search service that searchers the index and provides REST API
  - Web interface that allows end-users specific search

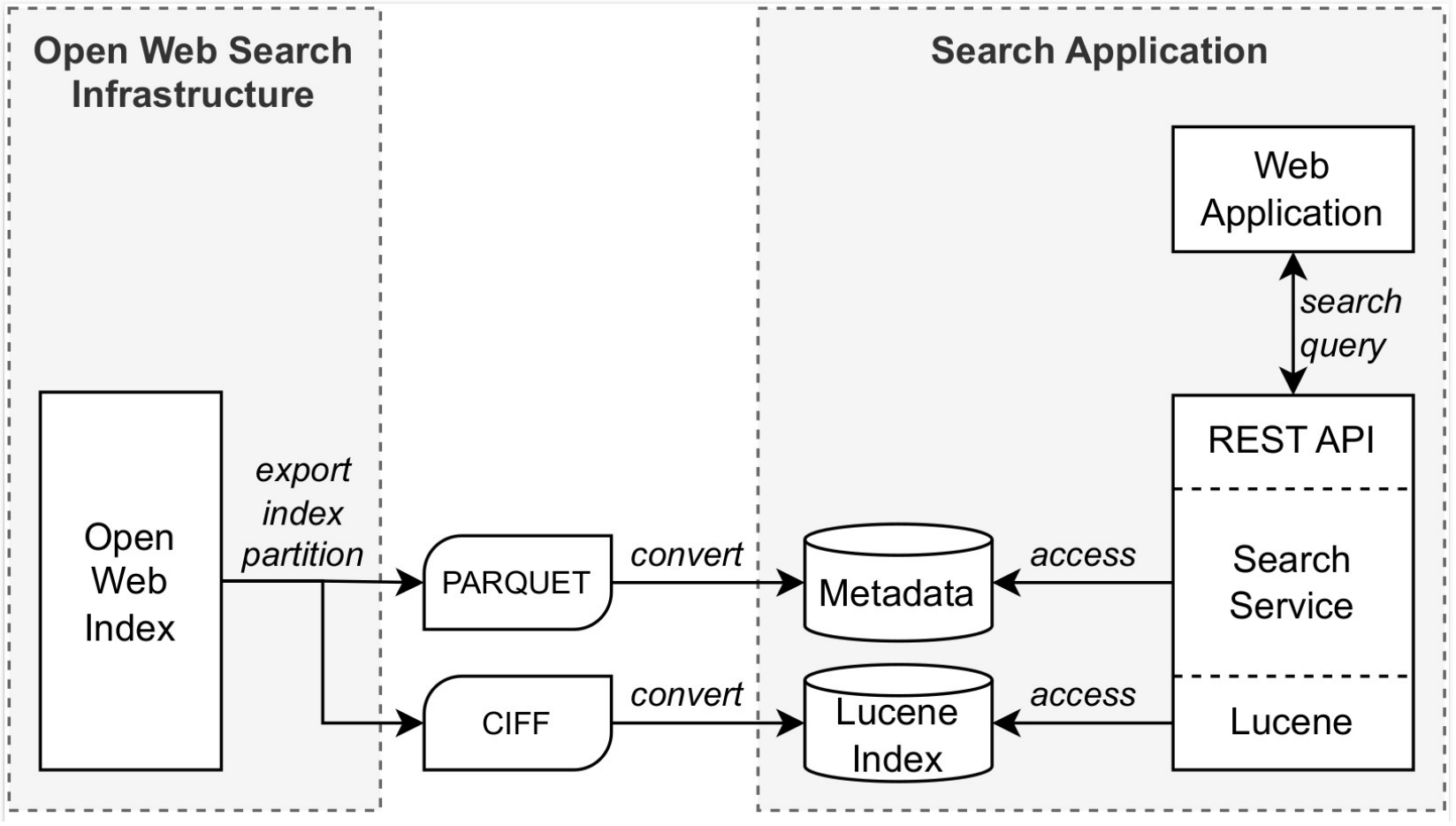


Prototype Search Application

- Guides the development of a OWS search application
  - Using the search service and development of a web interface (1)
  - Using the concept and development of the whole application (2)



Software Architecture



Index Partition: CIFF and PARQUET

- CIFF: Index exchange format; inverted index
- Parquet: Tabluar format

CIFF

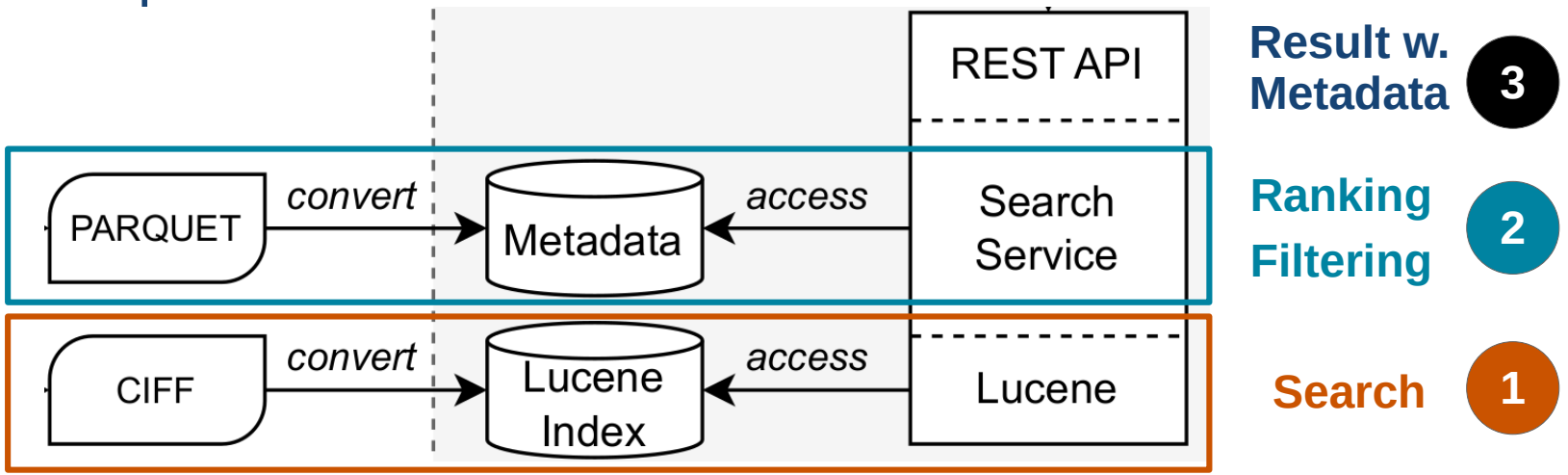
Term	List of (Document, Frequency)
Term 1	(docID A, 1) (docID B, 3) (docID C, 1)
Term 2	(docID D,2)
Term 3	(docID E, 4) (docID F, 2)
Term 4	(docID H, 2) (docID I, 1)

Parquet

Document	Lang	Full text	WARC date	...
docID A	en	... text ...	2023-07-01	
docID B	de	... text ...	2023-07-01	
docID C	de	... text ...	2023-07-01	
docID D	en	... text ...	2023-07-01	

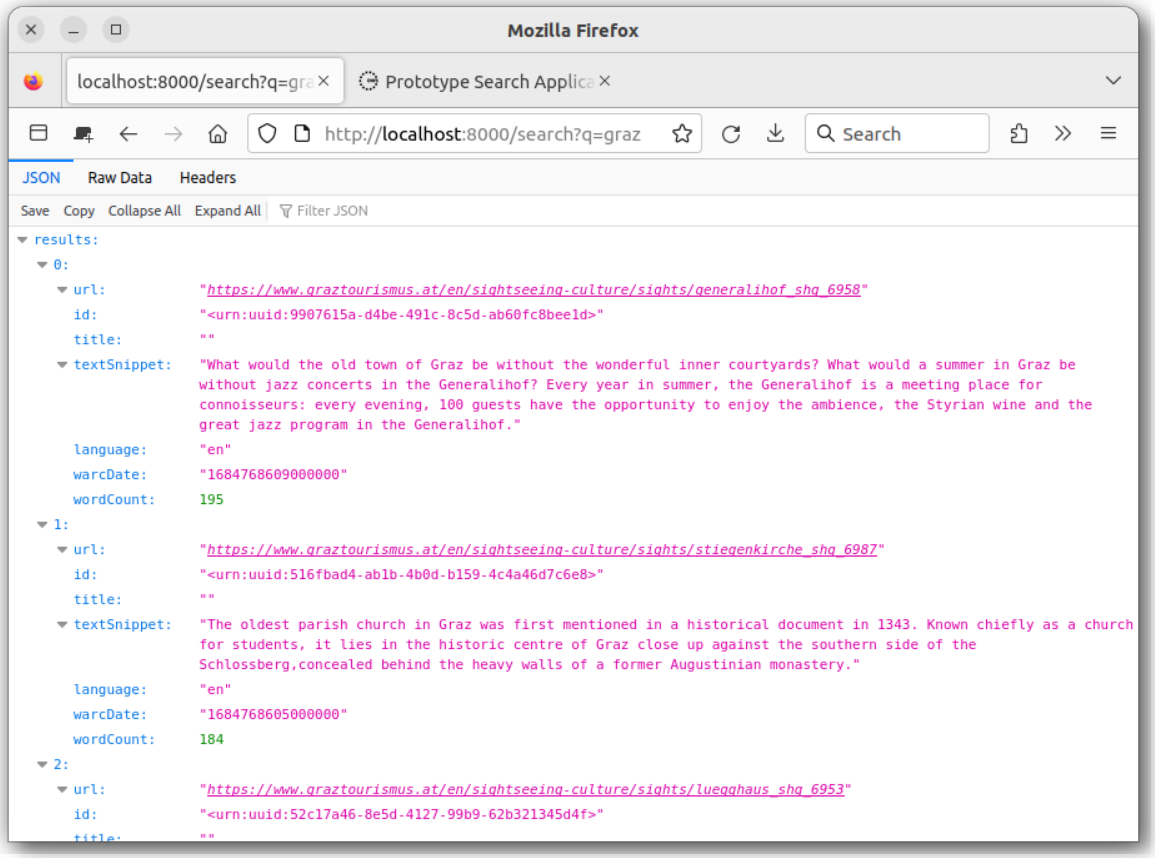
Search Process

- CIFF file is converted to a Lucene index that can be searched with the Lucene library
- Search result is filtered and ranked using metadata coming from the Parquet file



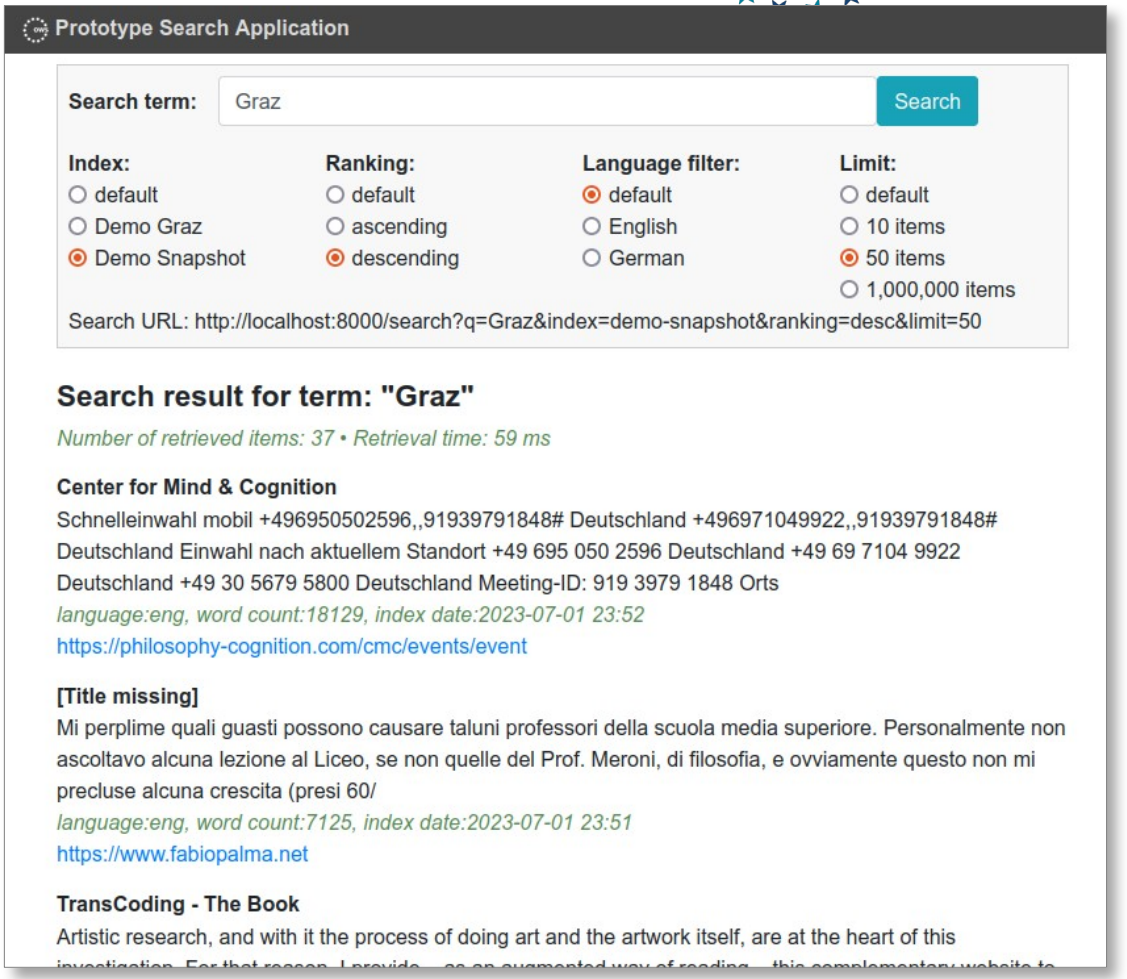
REST API

- End point:
  - host:port/search?...
- Request parameters
  - Search term, Index, ranking, language, limit
- Response data
  - URL, title, text snippet, WARC date, word count



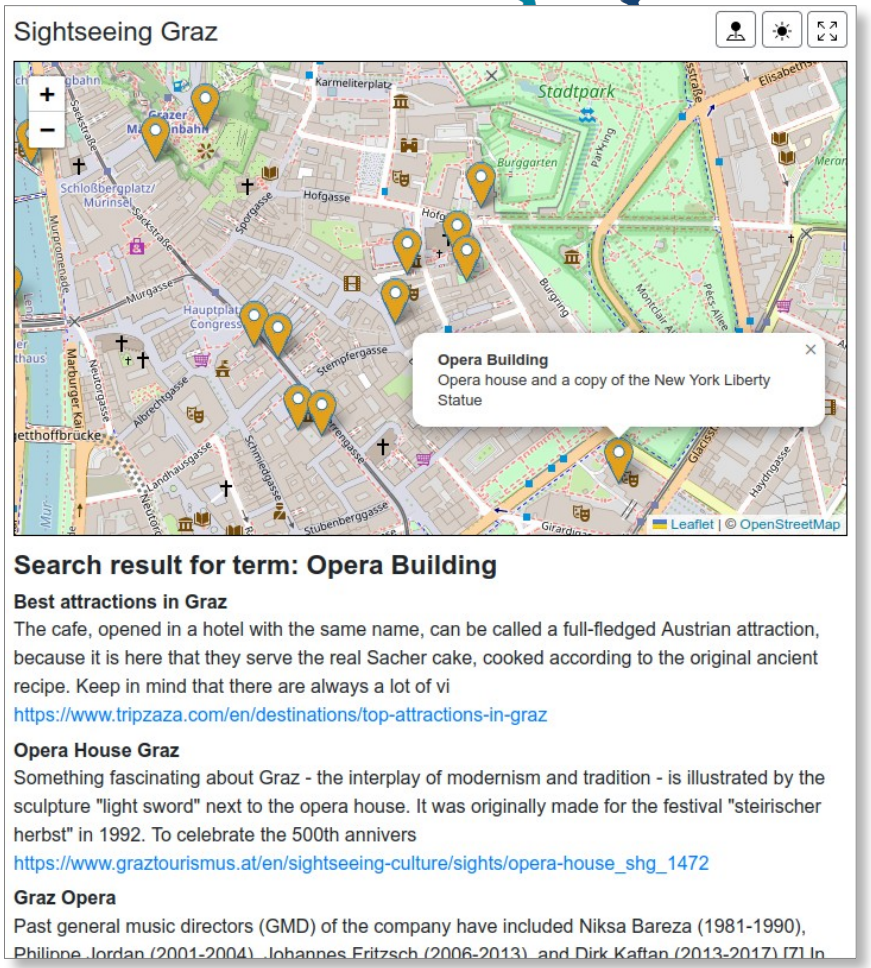
Demonstration 1: Basic Search

- Simple web interface to the Search Service
- Operates two indices
- Performs ranking and filtering
- Displays metadata



Demonstration 2: Sightseeing Application

- Integration into existing Sightseeing application Spots
- Clicking on a sightseeing marker triggers a search with the title as search term



Create You Own Search Application!

- Code available on GitLab in public area
- Follow instructions in Readme file to download and start service
- Try out REST API in your browser
- Try out Web interface in your web server
- Use the code to create you own application

https://opencode.it4i.eu/openwebsearcheu-public/prototype-search-application

Next Steps

- CIFF and Parquet file Management
  - Alignment CIFF and Parquet files with OWS infrastructure
  - Merging multiple CIFF files and related Parquet files
  - Performance tests with large index partitions
- Development environment for creation search application
  - Depending on feature requests from Hackathons, OWS Applications, 3rd party applications, etc.
- Ethical and legal implications

Any Questions?

Contact:

- Alexander Nussbaumer <alexander.nussbaumer@tugraz.at>
- Sebastian Gürtl <sguertl@student.tugraz.at>
- Graz University of Technology, Austria