

Federated Content Search Past, Present and Future

Erik Körner

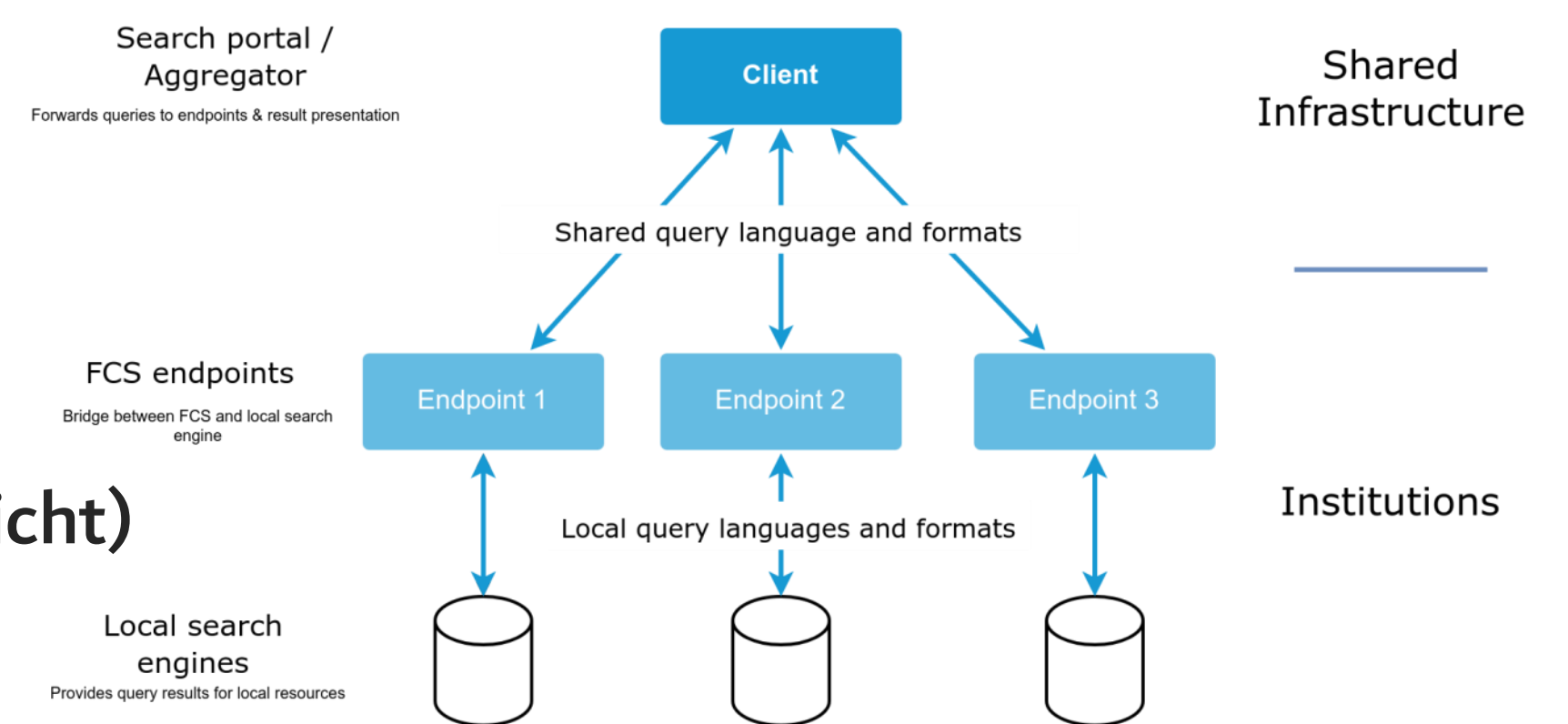
Saxon Academy of Sciences and Humanities in Leipzig
koerner@saw-leipzig.de



The Federated Content Search Infrastructure

- Started in 2011 @ CLARIN, specification currently in version 2.0
- 209 resources in 61 languages from 20 institutions (12 countries)
- No central index → Federated Search over distributed resources, results are aggregated in search portal “CLARIN Content Search”
- Integration with Metadata Search (VLO) and Text Processing (Weblicht)

→ Improves visibility of resources and institutions, enables easier search for textual evidence across various sources



Recent Changes

- FCS SRU Aggregator
 - Active maintenance and development
 - VLO Bridge; new deployment
 - UI & UX improvements
- FCS Awesome List
- Maven Archetype for Endpoint project development
- Python SRU/FCS libraries (translation of Java reference libraries)
- Example Endpoint Implementations for Korp, CQI and NoSketchEngine



Plans & Future Work

- Authentication & Authorization support
- Major rewrite of FCS Aggregator UI
 - Visualizations for new Data Views like TEI, ...
 - Support for special fonts and text scripts
- FCS specification updates
- FCS educational materials, e.g. tutorials, introduction slides, user and developer guides, ...

German NFDI Text+ Project

Text+ is a consortium in Germany's National Research Data Infrastructure (NFDI) working on the utilization of text- and language-based research data in a distributed research environment.



Integration of Lexical Resources (LR)

- Lexical Resources (Dictionaries, WordNets, ...) pose challenges: structurally different to text corpora
- Active work in German **Text+** project
- First specification draft ([10.5281/zenodo.7849753](https://zenodo.org/record/105281/files/zenodo.7849753))
 - LexCQL query language
 - LexHits Data View + advanced Data View (WIP)
- Prototype Endpoints & Aggregator: fcs.text-plus.org



Integration of AAI for search in restricted resources

- FCS specification extension
- Authentication & Authorization support in Aggregator, Endpoint → Prototypes

Contributions

- Alternative Aggregator UI as web component



www.clarin.eu

clarin@clarin.eu

[@CLARIN.ERIC](https://twitter.com/CLARIN.ERIC)

github.com/clarin-eric

CLARIN ERIC was established in 2012 and received ESFRI Landmark status in 2016

Acknowledgements



Project number
460033370

Funded by



Part of
nfdi
Nationale Forschungsdaten Infrastruktur

Deutsche Forschungsgemeinschaft
German Research Foundation

This presentation was created in the context of the work of the association German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal Republic of Germany and the 16 federal states, and the consortium Text+ is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 460033370. The authors would like to thank for the funding and support. Furthermore, thanks also include all institutions and actors who are committed to the association and its goals.