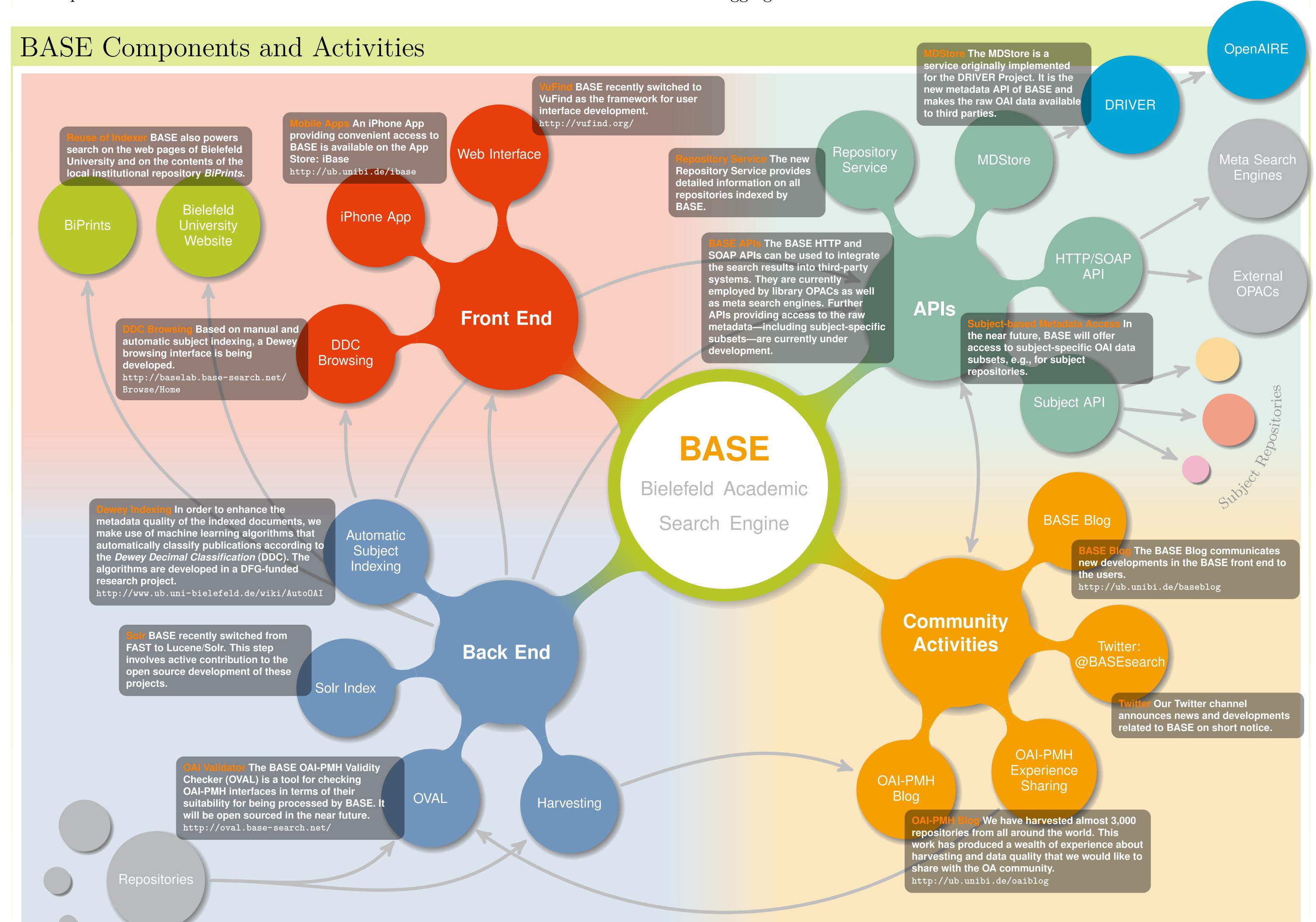
## Approaching Next-Generation OAI Service Providers: The BASE Case

Friedrich Summann, Dirk Pieper, Bernd Fehling, Renata Mitrenga, Sebastian Wolf, Marek Imialek, Mathias Lösch, & Wolfram Horstmann base.ub@uni-bielefeld.de

## Contribution

OAI service providers are a well-established backbone of scholarly communication. However, the development of the academic web does not stand still. Aggregating metadata and making them searchable via unified interfaces is performed by many search engines and not unique anymore. Rather, a service provider could exploit the potential of repositories by supporting value-added services such as enrichment, refinement or data linking. This involves service providers becoming data providers themselves by offering the aggregated—and ideally enhanced—data via APIs, so that the retrieval of documents in repositories can be improved and the development of further services will be fueled. Current activities in the context

of BASE focus on the redesign of the architecture for an OAI service provider to fulfill these novel requirements. Having been run by Bielefeld University Library since 2004, BASE is a registered service provider that collects, normalizes, and indexes data using OAI-PMH and provides one of the world's most complete aggregations of OAI publication data. BASE is providing data interfaces and index access for other service providers. Moreover, we are developing (DFG funded) machine learning-based text categorization to automatically apply subject indexing of OAI DC records for being able to provide access to subject-specific subsets of the aggregation.



## BASE Fact Sheet



- > 28 million documents indexed
- from > 1,800 document sources
- from > 70 countries
- sources selected for academic
   Open Access contents
- crawling and indexing of selected academic Web sites

## The BASE Team

Dirk Pieper project coordinator for search engine technologies

Friedrich Summann head of IT department at Bielefeld UL

Bernd Fehling senior developer

Renata Mitrenga junior developer

Sebastian Wolf department for electronic services

Marek Imialek APIs, MDStore, DRIVER, OpenAIRE (FP7)

Mathias Lösch text mining, automatic subject indexing project (DFG)

Wolfram Horstmann head of projects at Bielefeld University Library

www.base-search.net





