PADERBORN UNIVERSITY

GERBIL PG

Summer Semester 2023

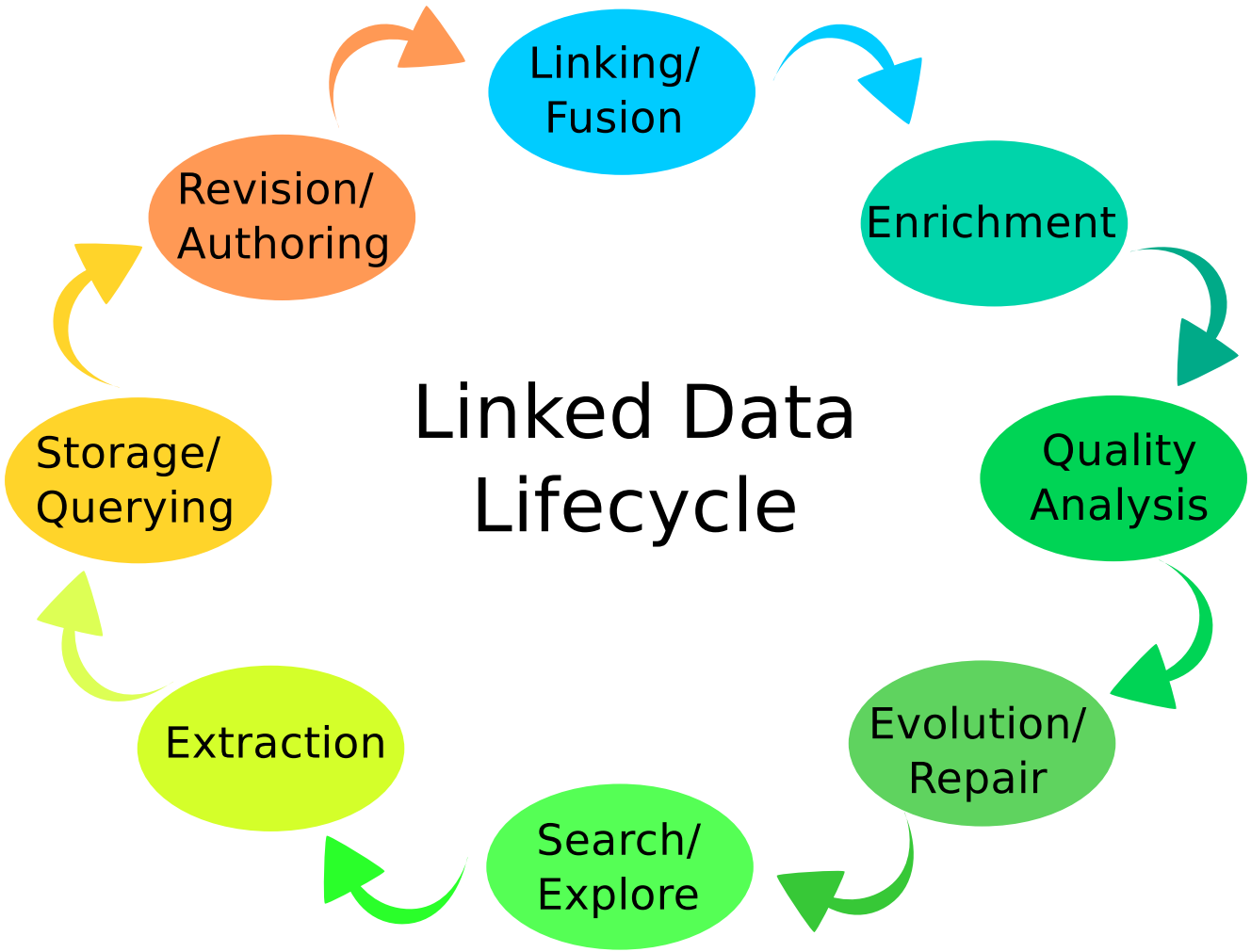
Michael Röder

DICE

Data Science Group
Paderborn University

July 17, 2023

Linked Data Lifecycle



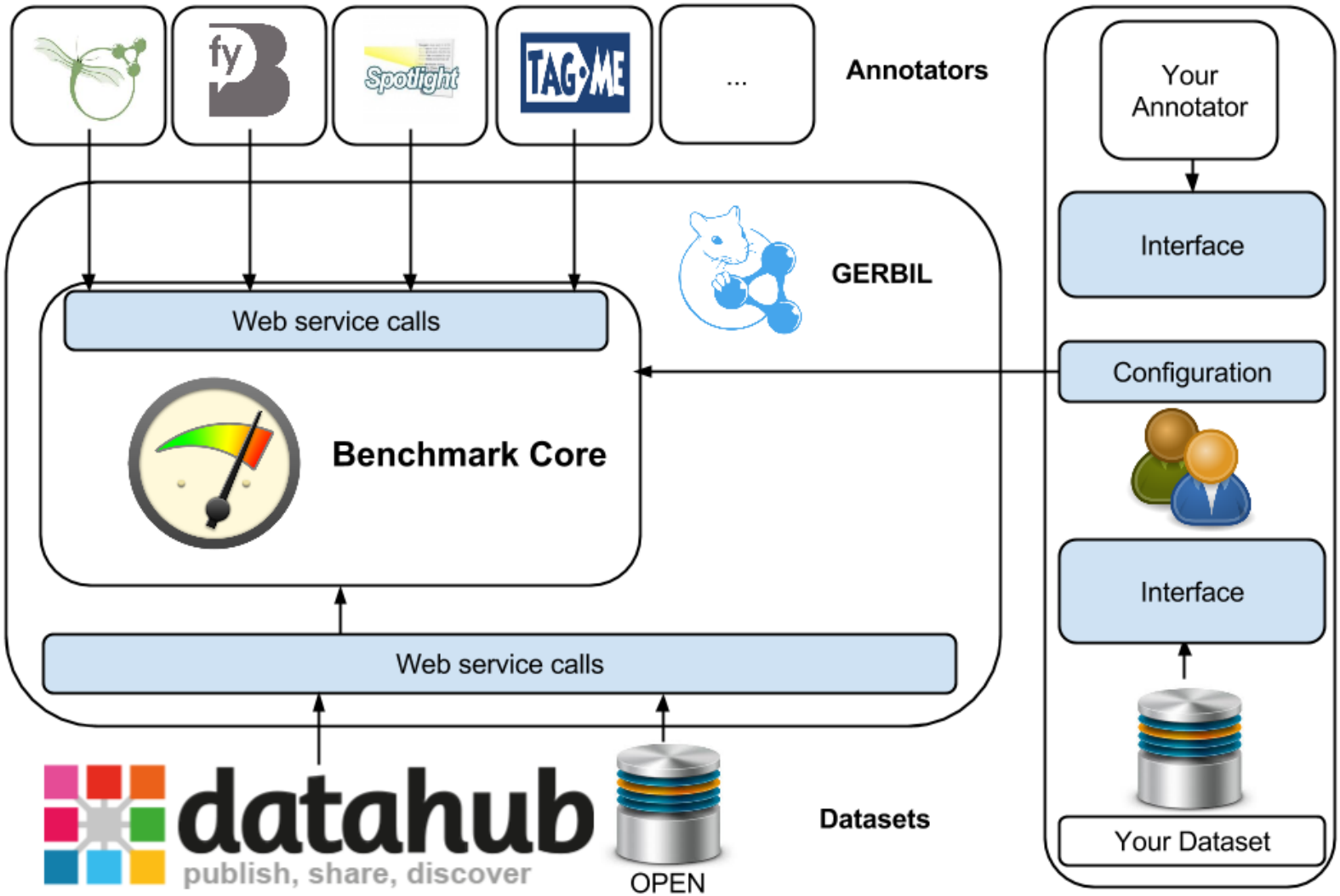
Benchmarking

Benchmarking? What is that?

Very briefly: Evaluate a system in a controlled environment and measure its effectiveness and efficiency.

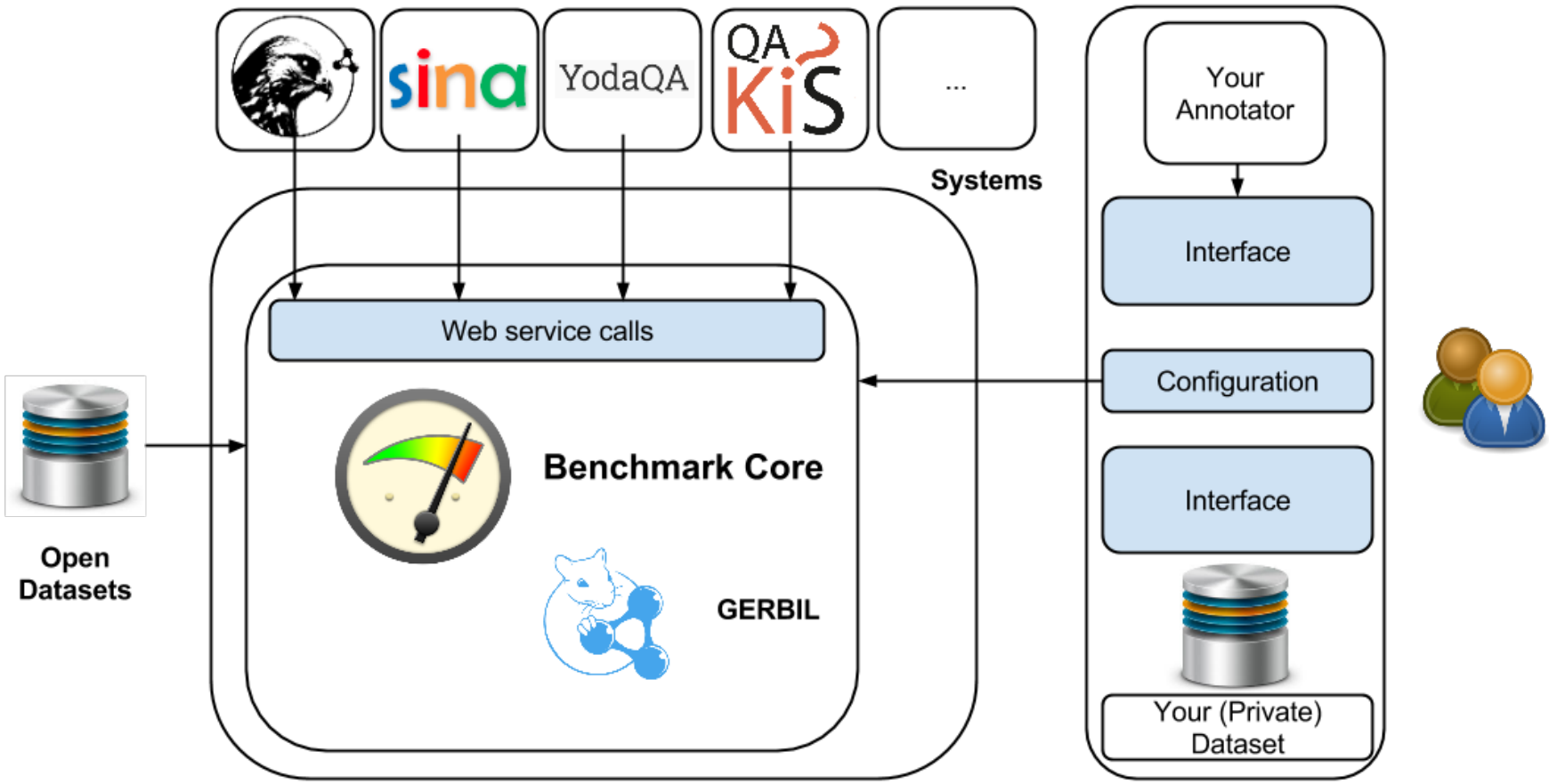
GERBIL

Knowledge Extraction



GERBIL

Question Answering



GERBIL

Variants

Domain	Input	Output
Knowledge Extraction	Text	Structured data
Question Answering	Questions	Answers
Fact Checking	Facts	Veracity values
Machine Translation	Text	Text

GERBIL

Variants

Domain	Input	Output
Knowledge Extraction	Text	Structured data
Question Answering	Questions	Answers
Fact Checking	Facts	Veracity values
Machine Translation	Text	Text
Knowledge Graph Embedding (Link Predication)	Graph	Embeddings

GERBIL

Goals

Summary

- Problem: Development and maintenance of new GERBIL instances for different domains

GERBIL

Goals

Summary

- Problem: Development and maintenance of new GERBIL instances for different domains
- Solution: Extend GERBIL to become a light-weight benchmarking platform

GERBIL

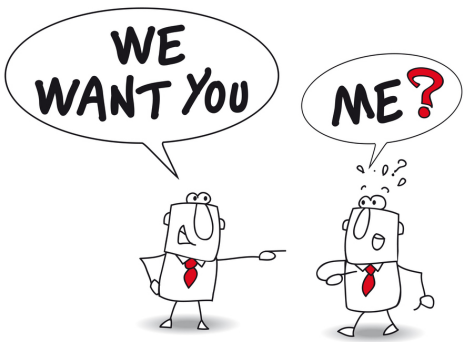
Goals

Summary

- Problem: Development and maintenance of new GERBIL instances for different domains
- Solution: Extend GERBIL to become a light-weight benchmarking platform

- Develop GERBIL 2.0
  - Basic platform implementation
  - Variants only extent the platform as needed
  - Usage of patterns like Dependency Injection
- Implement a new variant based on the new platform
  - Task related to Knowledge Graph Embeddings

We want



- Java / Maven
- Software development experience is beneficial
- Prior knowledge of RDF is beneficial
- Experience with git is beneficial
- Motivation to develop software

We Offer

- Real software: Participate in a "real" software development project
- Code Review: Regular code reviews from experienced developers
- Evaluation: Understand the importance of evaluation
- Master theses: Topics can be extended accordingly
- Publications: at top conferences (ISWC, ESWC, WWW)

GERBIL

Further Information

Publications:

- M. Röder, R. Usbeck, and A. Ngomo: GERBIL – Benchmarking Named Entity Recognition and Linking consistently. 2018
- R. Usbeck, M. Röder, M. Hoffmann, F. Conrads, J. Huthmann, A. Ngomo, C. Demmler, and C. Unger: Benchmarking Question Answering Systems. 2019

Github projects:

- `https://github.com/dice-group/gerbil`

Technologies:

- Java / Maven
- Knowledge about RDF can be helpful