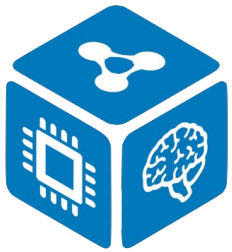
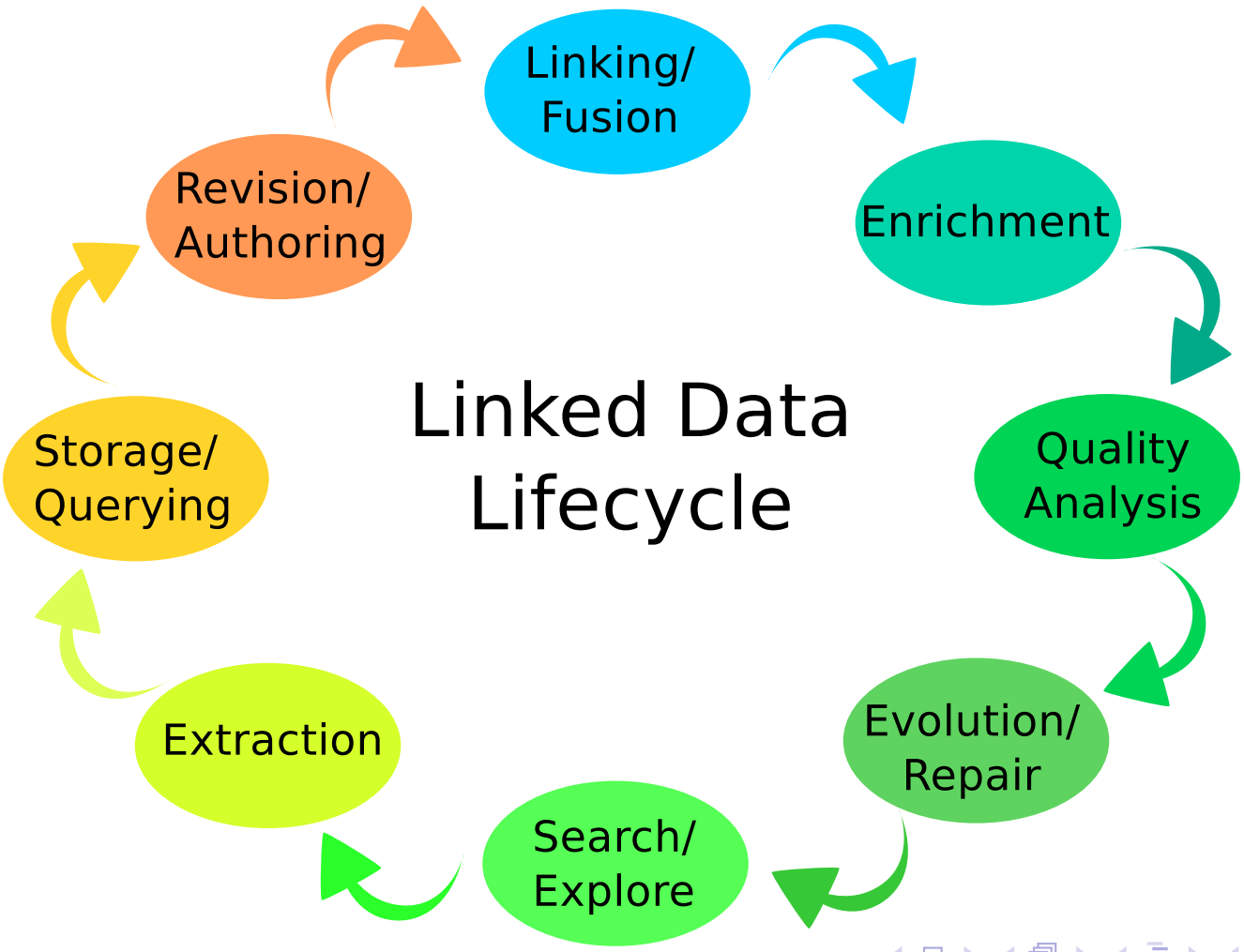
Project Group
Data Science Suite V

Group: Data Science
Prof. Dr. Axel Ngonga
Tutor: Michael Röder



DICE – Data Science Group, University Paderborn, Germany

July 13, 2020

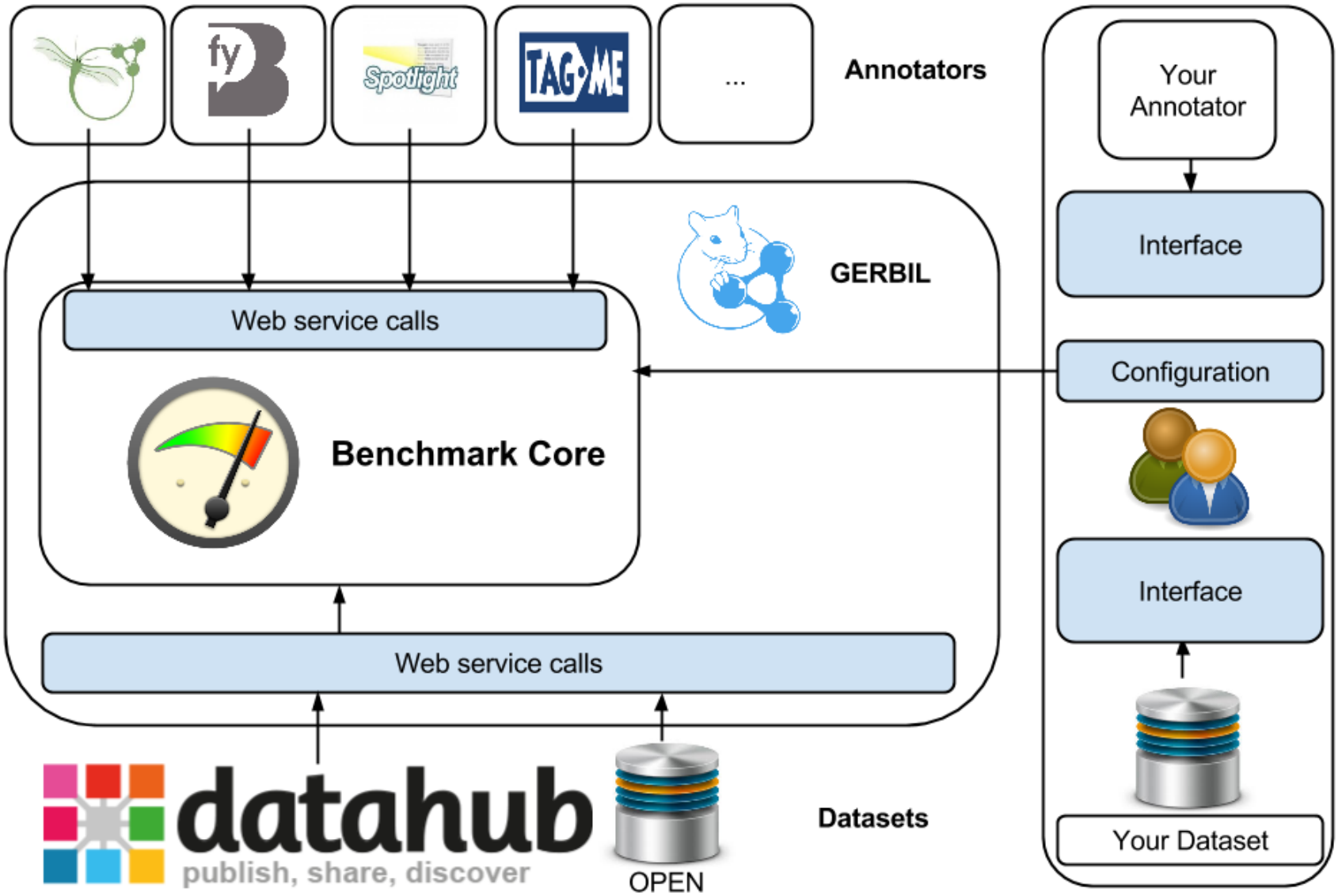

Linking/
Fusion
Enrichment
Quality
Analysis
Evolution/
Repair
Search/
Explore
Extraction
Storage/
Querying
Revision/
Authoring
Linked Data
Lifecycle

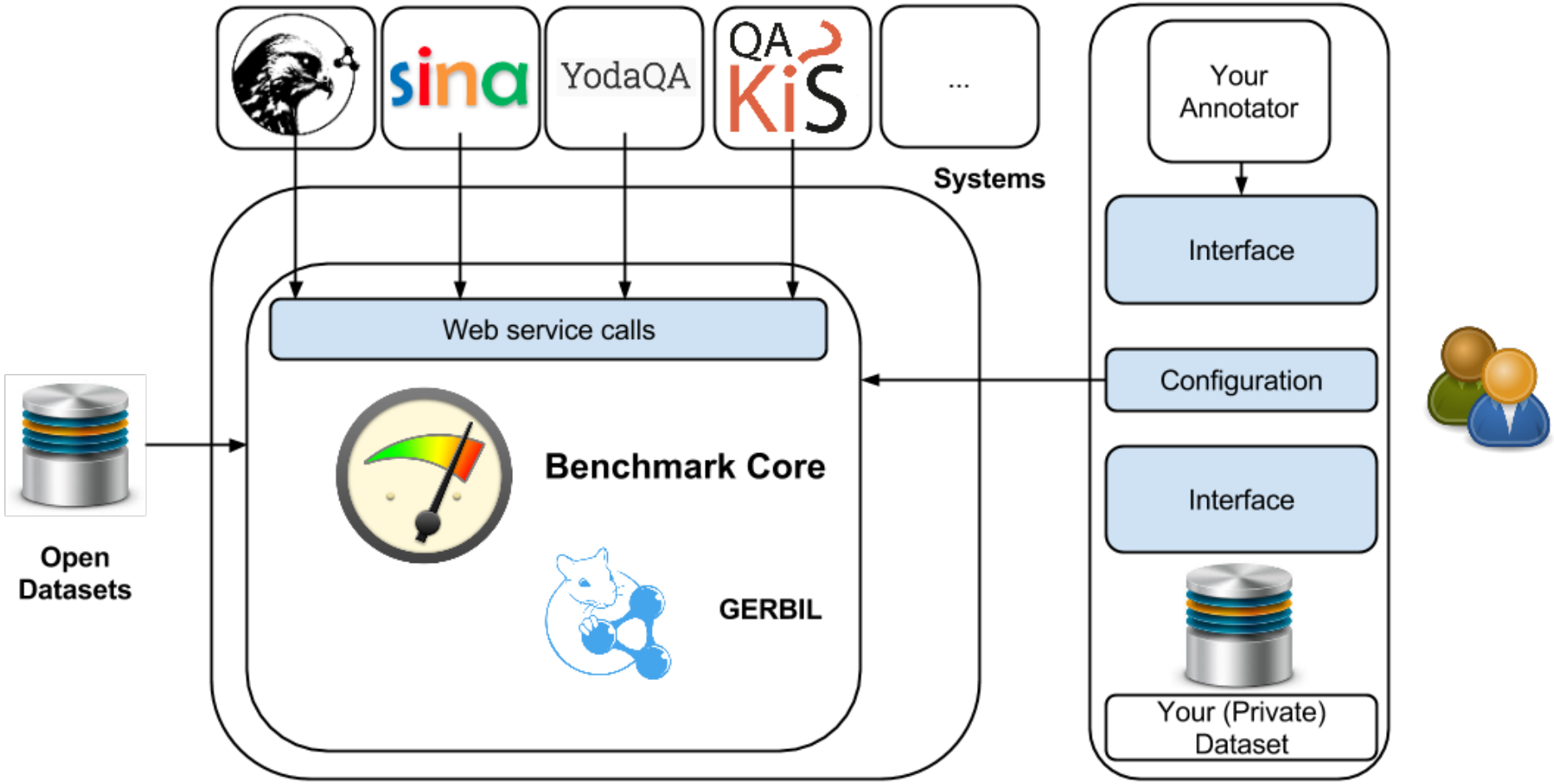
Section 1

Topics

Benchmarking? What is that?

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


Annotators
...
Your Annotator
Interface
GERBIL
Web service calls
Configuration
Benchmark Core
Interface
Web service calls
datahub
publish, share, discover
OPEN
Datasets
Your Dataset


sina
YodaQA
QA KiS
...
Systems
Your Annotator
Interface
Web service calls
Configuration
Benchmark Core
Interface
Open Datasets
GERBIL
Your (Private) Dataset

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

Summary

- Problem: Development and maintenance of new GERBIL instances for different domains
- Solution: Extend GERBIL to become a light-weight benchmarking platform
- Goal: Develop GERBIL 2.0

Further information:

- 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:

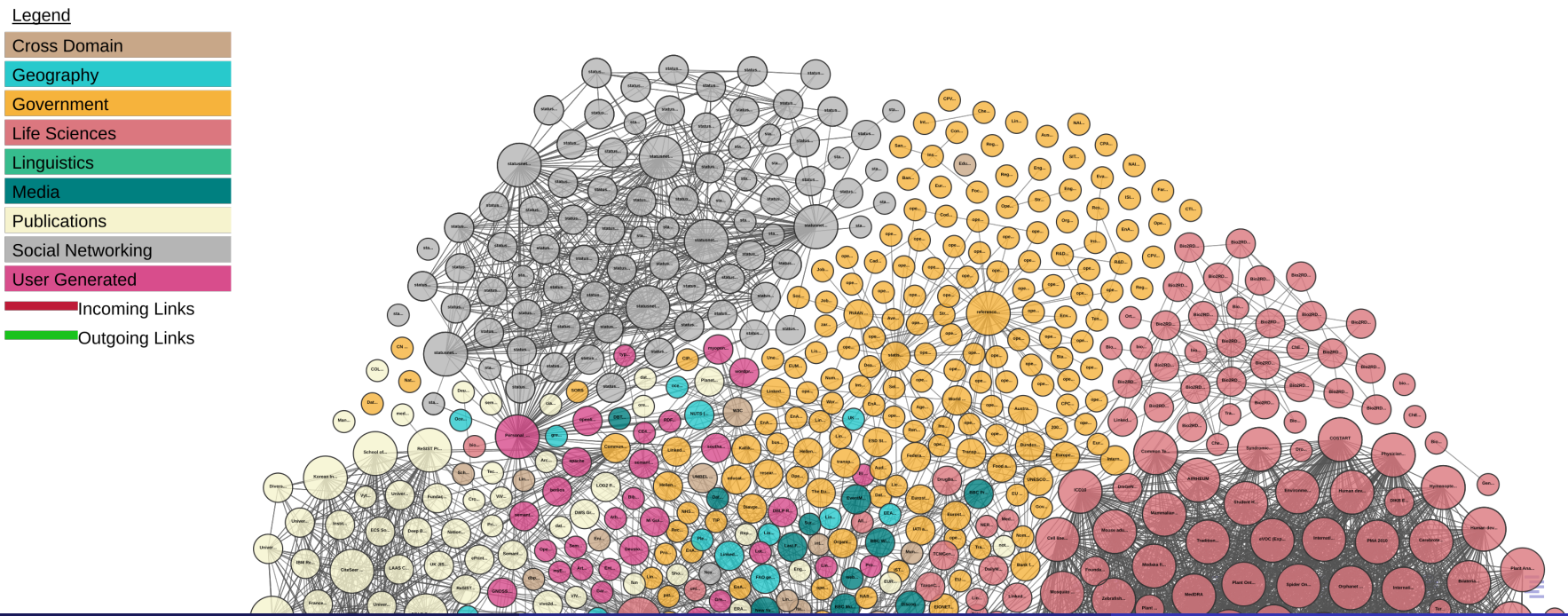
- `https://github.com/dice-group/gerbil`

Technologies:

- Java / Maven
- UML
- RDF

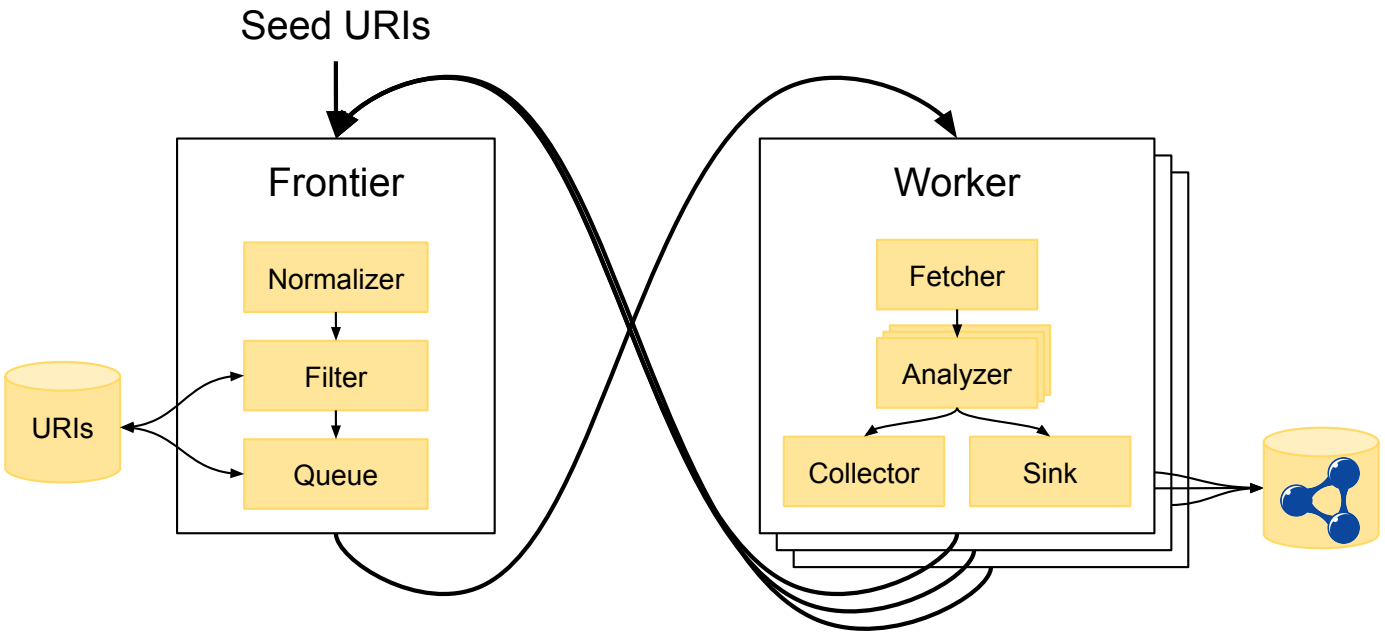
Summary

- Problem: We need to improve crawlers for gathering data from the Data Web.
- Solution: Create a benchmark for such crawlers.
- Goal: Improve our Data Web crawler benchamrk.



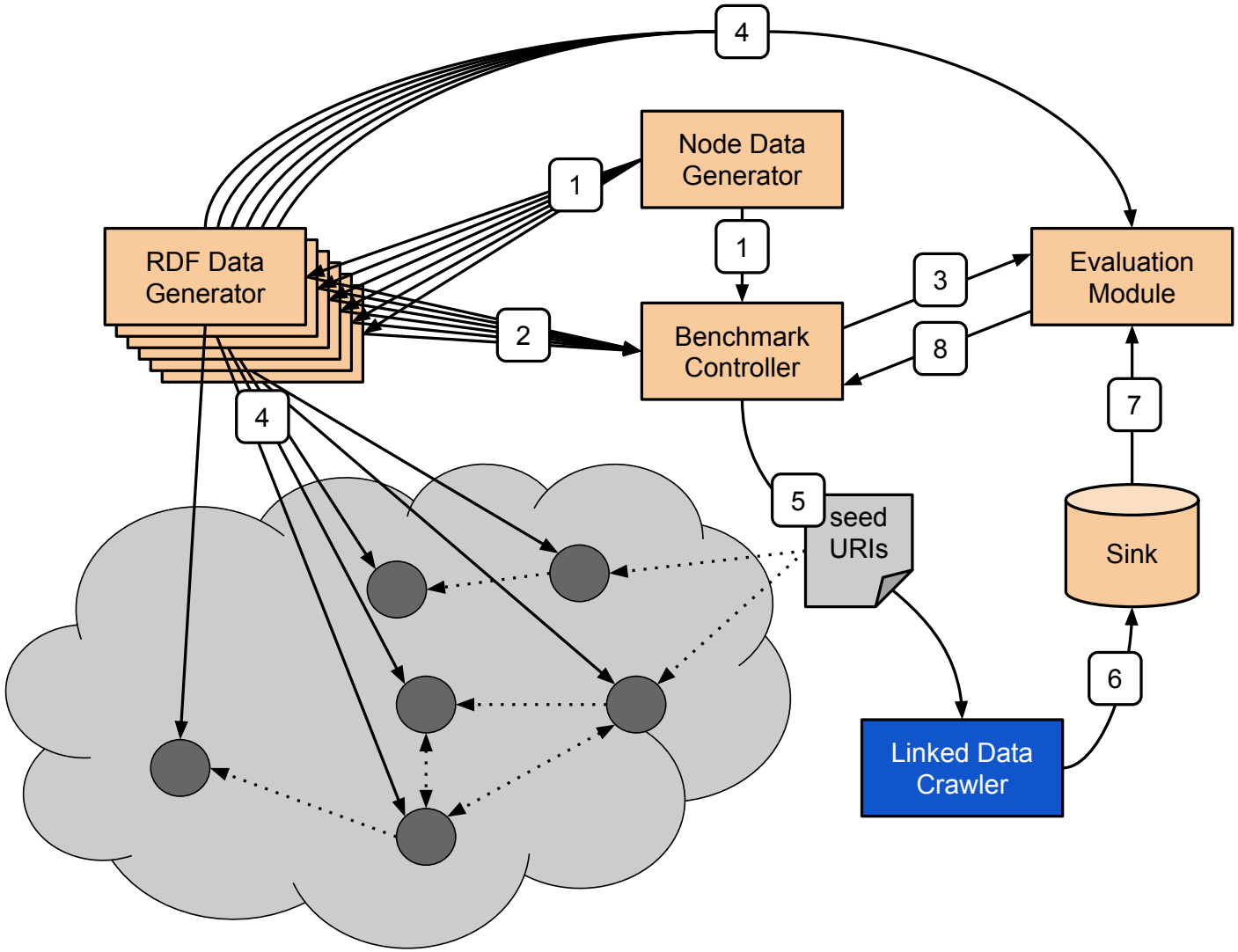
Summary

- Problem: We need to improve crawlers for gathering data from the Data Web.
- Solution: Create a benchmark for such crawlers.
- Goal: Improve our Data Web crawler benchamrk.



Topics

Benchmarking – ORCA



Further information:

- Paper on arxiv.org
  `https://arxiv.org/abs/1912.08026`

Github projects:

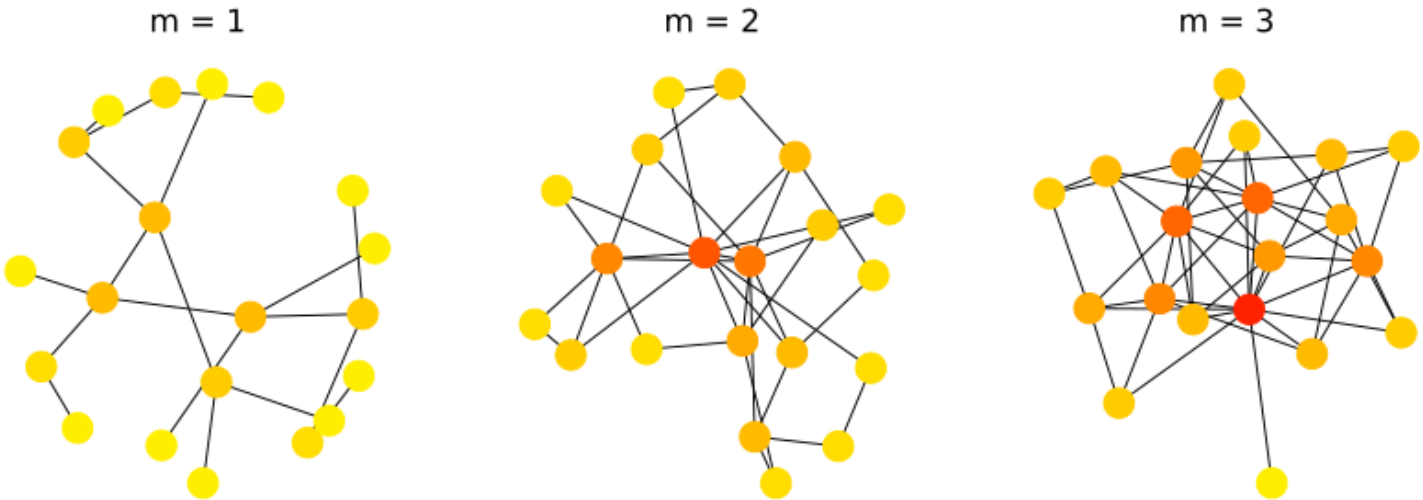
- `https://github.com/dice-group/orca`
- `https://github.com/dice-group/Squirrel`

Technologies:

- Graph theory
- RDF, RDFa, Microformat, microdata, JSON-LD
- Java / Maven
- Docker

Summary

- Problem: We would like to be able to scale knowledge graphs.
- Solution: Create an algorithm that is able to mimic real-world graphs.
- Goal: Improve this library with respect to runtime and functionalities.



Further information:

- Master thesis
  `https://hobbitdata.informatik.uni-leipzig.de/teaching/projectgroups/Thesis-Final-lemming.pdf`

Github projects:

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

Technologies:

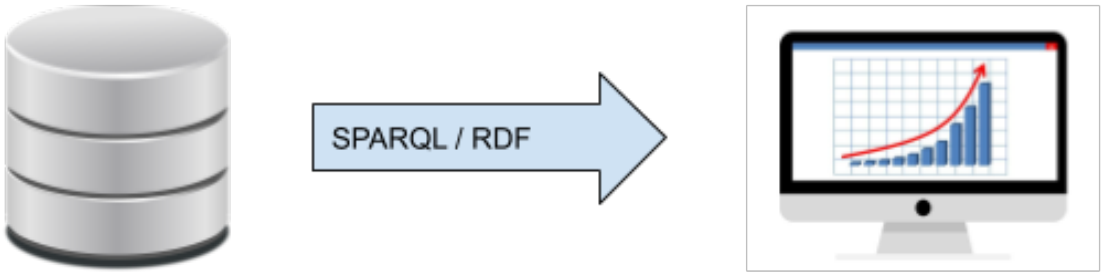
- Graph theory
- RDF
- Java / Maven

Summary

- Problem: Application server is needed to transform data from RDF to JSON
- Solution: Let the UI work directly with RDF
- Goal: Implement a prototype based on templates

Summary

- Problem: Application server is needed to transform data from RDF to JSON
- Solution: Let the UI work directly with RDF
- Goal: Implement a prototype based on templates



Further information:

- Our DICE group website already relies on the template-based translation of RDF to static pages.
  `https://github.com/dice-group/dice-website`

Technologies:

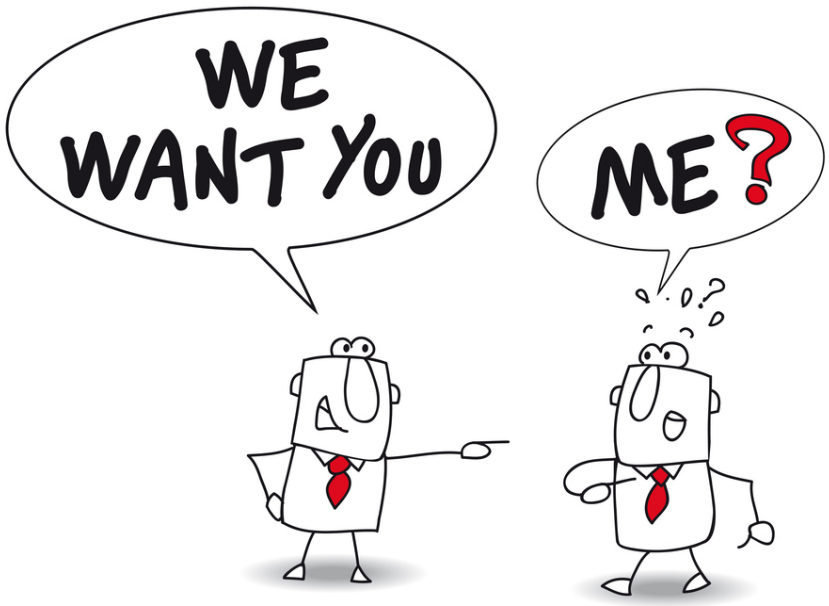
- RDF, SPARQL
- JavaScript

Section 2

Summary

We want you

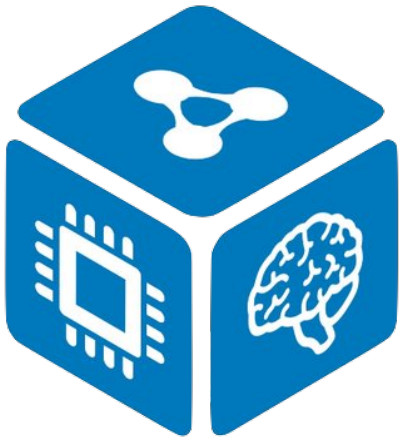
- Create new software: Develop new software and research prototypes.
- Enhance code: Improve existing solutions.
- Participate: Bring your own ideas in.



We Offer

- Running software: Open-source, industry-grade solutions
- Real data: Billions of facts from Wikipedia, bio-medicine, etc.
- Expert tutors, who developed the core software
- Follow-up: Topics can be extended to MSc thesis
- Publications at top conferences (ISWC, ESWC, WWW)

That's all Folks!



Thank you!

Topics:

- Benchmarking
  - GERBIL
  - ORCA
  - Lemming
- Generic RDF UI templates

The topics are subject to change.
More information at
`https://dice-research.org`