

TIRA Integrated Research Architecture

Maik Fröbe

Friedrich-Schiller-Universität Jena [\[webis.de\]](http://webis.de)

TIRA Integrated Research Architecture [TIRA]

- ❑ TIRA handles software submissions in shared tasks.
(“software submission” as opposed to “run submission”)
- ❑ TIRA is used in different labs since 2012, especially in PAN and Touché at CLEF, and is actively developed by the Webis Group.*

Objectives:

1. Support replicability and reproducibility in scientific research.
2. Blinded evaluation.
3. Processing of non-public or confidential documents.

* Gollub et al. (DEXA 2012). [TIRA: Configuring, Executing, and Disseminating Information Retrieval Experiments.](#)
Potthast et al. (2019). [TIRA Integrated Research Architecture.](#)
Fröbe et al. (SIGIR 2023). [The Information Retrieval Experiment Platform.](#)

TIRA Integrated Research Architecture [TIRA]

- ❑ **Replicability:** same input and same method $\leadsto$ same output
- ❑ **Reproducibility:** similar input and equivalent method $\leadsto$ comparable output

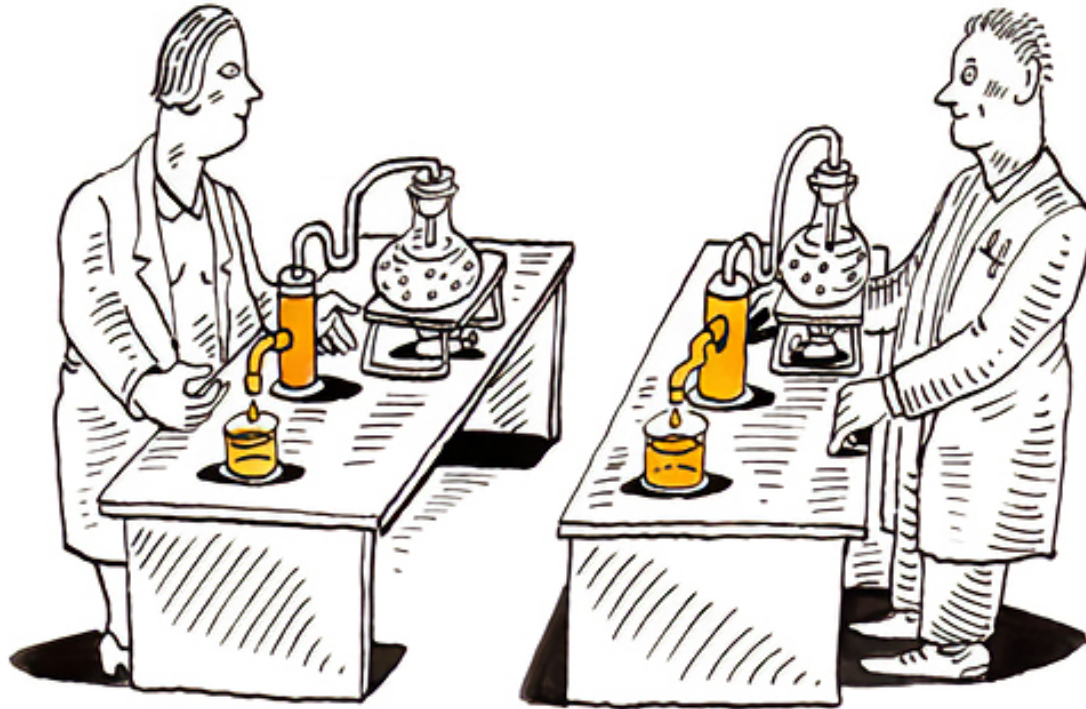
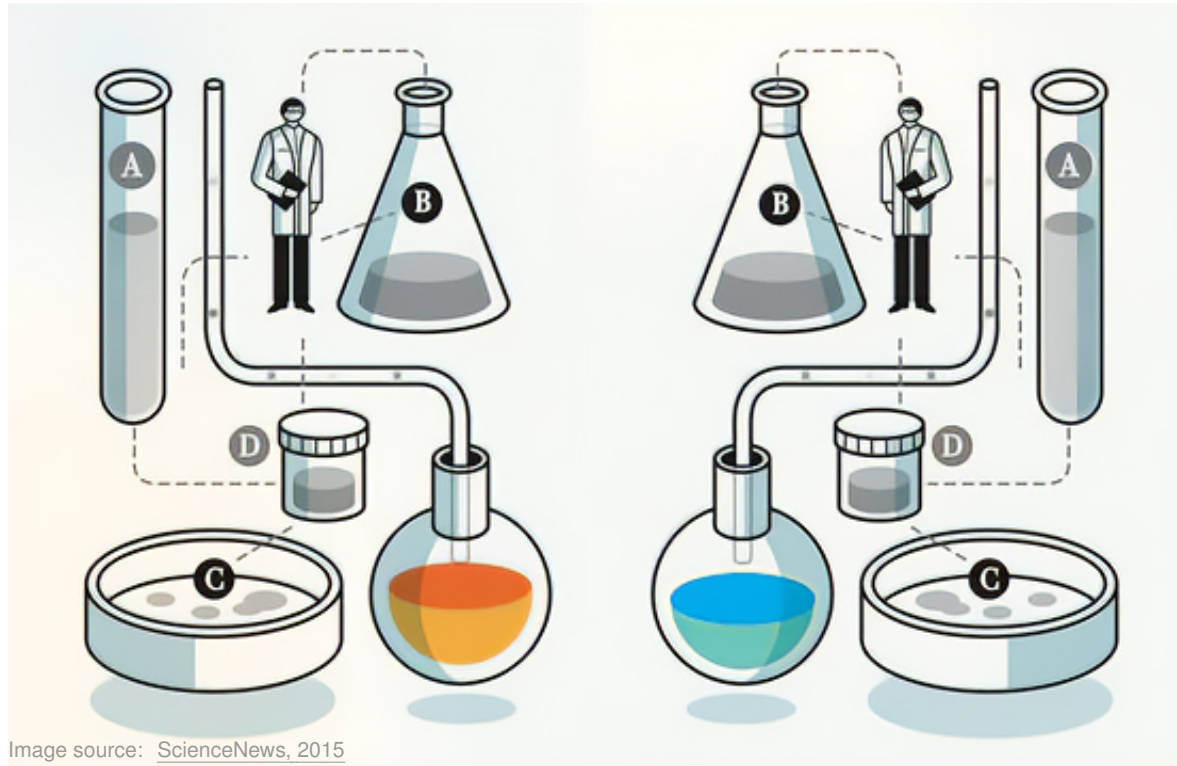


Image source: [F1000Research.com](https://www.f1000research.com), 2014

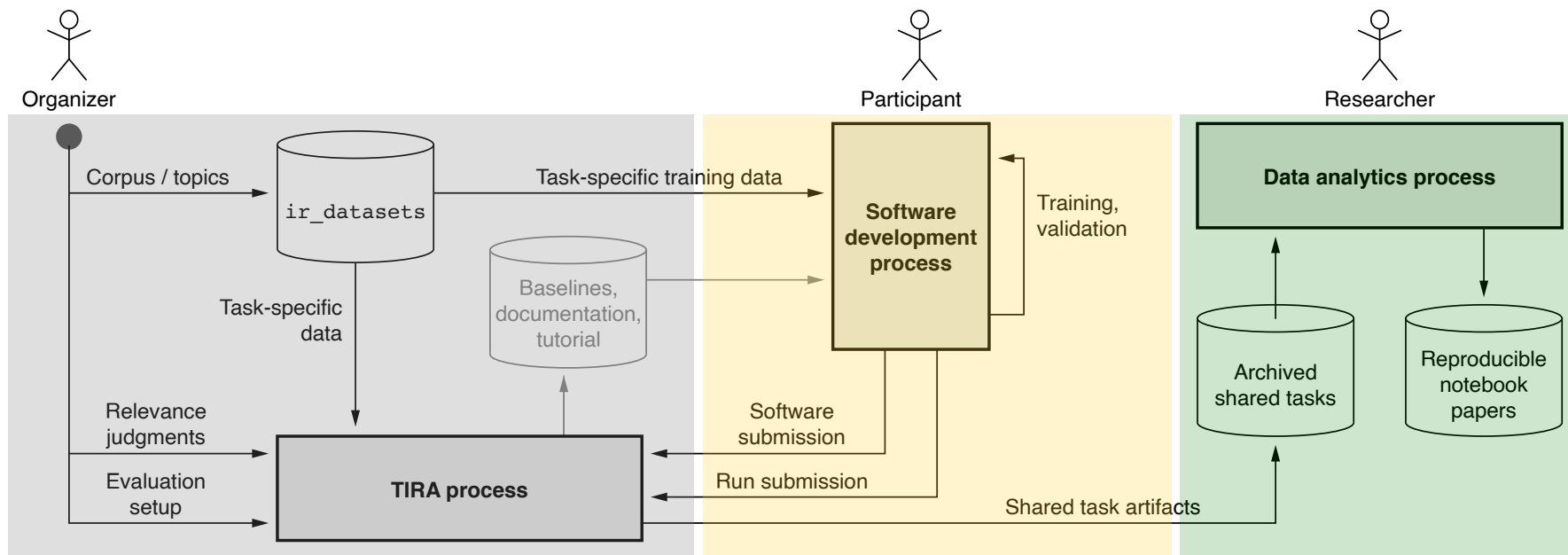
TIRA Integrated Research Architecture [TIRA]

- ❑ **Replicability:** same input and same method $\leadsto$ same output
- ❑ **Reproducibility:** similar input and equivalent method $\leadsto$ comparable output



TIRA Integrated Research Architecture [TIRA]

End-to-End Shared Task-Support



Example TIREx :

- ❑ 50 transferable retrieval models.
- ❑ 32 exchangeable benchmarks.
- ❑ Metadata (hardware and experiment configuration) tracking.

* Fröbe et al. (ECIR 2023). [Continuous Integration for Reproducible Shared Tasks with TIRA.io.](#)

AI Benchmarking

Keypoints for Services and Research

- ❑ Efficiency/Resource measurements, e.g., LSR-Benchmark, ANN-Benchmark.¹
- ❑ Retrieval Experiments with expensive neural models via corpus subsampling.²
- ❑ LLMs as a judge, e.g., for RAG via TREC AutoJudge^{3,4,5}
- ❑ Benchmarking interactive systems.^{6,7} [user simulation]

¹ Hagen et al. (SIGIR 2025). [TIREx Tracker: The Information Retrieval Experiment Tracker.](#)

² Fröbe et al. (ECIR 2025). [Corpus Subsampling: Estimating the Effectiveness of Neural Retrieval Models on Large Corpora.](#)

³ Faggioli et al. (ICTIR 2023) [Perspectives on Large Language Models for Relevance Judgment.](#)

⁴ Kiesel et al. (CLEF 2024). [Who Will Evaluate the Evaluators? Exploring the Gen-IR User Simulation Space.](#)

⁵ [TREC AutoJudge 2026.](#)

⁶ Schaer et al. (Sim4IA 2025) [SIGIR 2025 Second Workshop on Simulations for Information Access.](#)

⁷ Aliannejadi et al. (TREC iKAT 2025) [TREC Interactive Knowledge Assistance Track.](#)

Thank you for your attention!

