enRichMyData

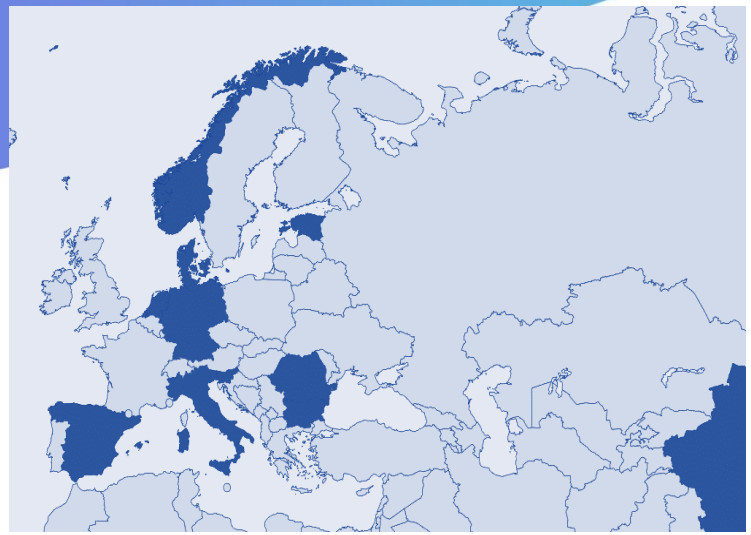
Enabling Data Enrichment Pipelines for AI-driven Business Products and Services

Irena Pavlova

GATE Big Data and AI Centre of Excellence

enRichMyData project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070284

enRichMyData Project



HORIZON-CL4-2021-DATA-01-03: Technologies for data management (IA)

2022-2025

SINTEF
PHILIPS
BOSCH
JOT INTERNET MEDIA
ES GROUP
expert.ai
ontotext
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
Jožef Stefan Institute Ljubljana, Slovenia
University of Copenhagen
GATE
BGRIMM TECHNOLOGY GROUP

enRichMyData Vision and Objectives

Vision

Create a novel paradigm for building rich, high-quality, valuable, and FAIR-compliant datasets to feed downstream Big Data and AI applications in the context of **Data-sharing Ecosystems**

Objectives

- Improve data discovery and profiling of potentially valuable data
- Improve wrapping of data sources in different formats
- Simplify cleaning, linking and extension of data
- Simplify annotation and classification of textual data
- Support the management of data enrichment pipelines
- Support data streaming in data enrichment pipelines
- Monitor and reduce energy consumption for executing data enrichment pipelines

Main challenges targeted by enRichMyData

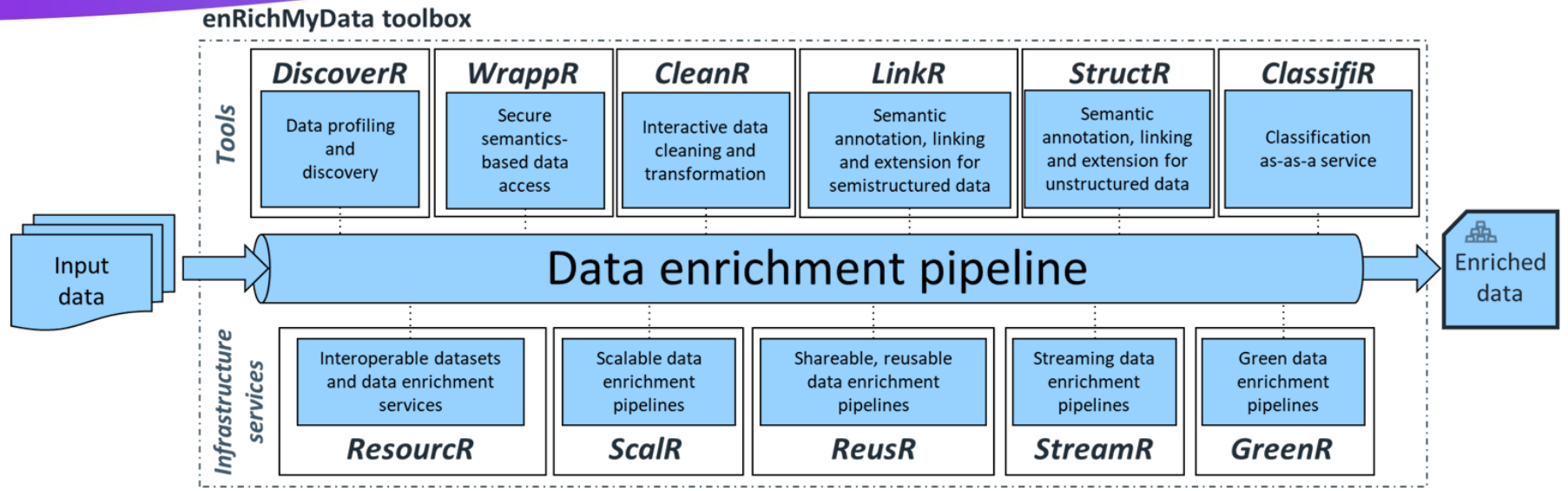
Related to data enrichment

- Lack of holistic end-to-end support for data enrichment lifecycle
- Steep learning curve for performing data enrichment tasks
- Lack of Machine Learning with Humans-In-The-Loop paradigm

Related to infrastructure

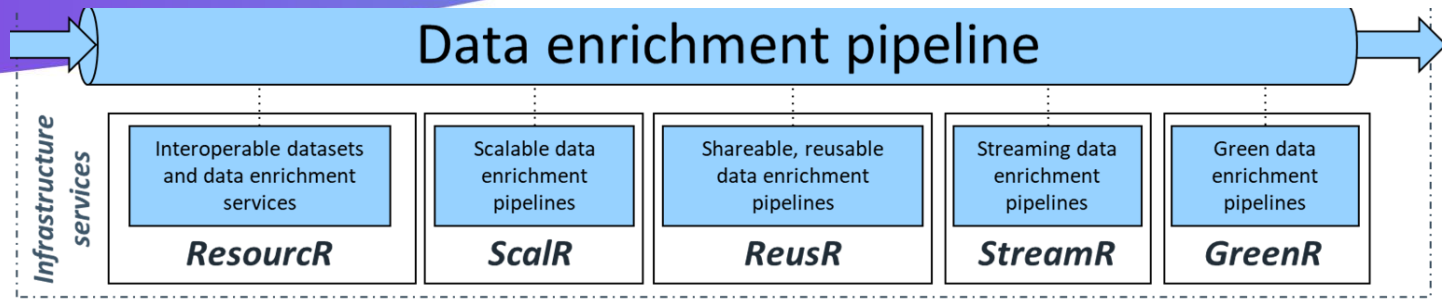
- Limited support for scalable execution of data enrichment pipelines,
- Little repeatability and reusability of data enrichment pipelines,
- Low sustainability due to inefficient use of resources and high CO2 footprint.

enRichMyData Toolbox and Business Cases



(***Digital Marketing***) Marketing data enrichment for smart-bidding optimization
(***Manufacturing***) AI-based Welding Analytics
(***Predictive Maintenance***) Smart Maintenance of Medical Imaging Systems
(***Public Procurement***) European Register of Entities from Known Actions
(***Innovation Ecosystems***) Innovation Knowledge Graph
(***Mineral Processing***) Mineral Processing Optimization

enRichMyData Tools addressing infrastructure challenges



ScalR provides horizontal scalability of data enrichment pipelines and support for management of the different procedures associated with the **execution of data enrichment pipelines flexibly on heterogeneous computing infrastructures**.

StreamR provides for streaming support in data enrichment pipelines. It pipes data streams from/to appropriate endpoints and **ensures high throughput**, providing a configurable set of tools for **setting up custom streams for new applications.**

GreenR supports monitoring of data enrichment pipelines in terms of their environmental impact. It monitors the carbon footprint of the various components in the pipeline and provides the results to a **dashboard to log and modulate the pipelines' heavy computations environmental impact.**

enRichMyData

https://enrichmydata.eu

https://www.linkedin.com/company/enrichmydata

https://twitter.com/enrichmydata

irena.pavlova@gate-ai.eu

enRichMyData project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070284