



enRichMyData

@EBDVF2022

Dumitru Roman
SINTEF AS
Oslo, Norway

- Enabling Data Enrichment Pipelines for AI-driven Business Products and Services
- HORIZON-CL4-2021-DATA-01-03
- 2022-2025
- Budget: approx. € 6M

Data Enrichment: **putting data in context, typically add more data to an input data set (structured / semi-structured / texts), improving data quality**

- Representative operations part of data enrichment: reconciliation/linking, extension/augmentation, cleaning and preparation, extraction of entities and relations, classification, profiling





enRichMyData



SPECIFIC NEEDS

Lack of dedicated software tools and infrastructure for supporting data enrichment processes required to build rich and high-quality data needed for successful implementation of Big Data analytics and AI applications.

TARGET GROUPS

Data/Software industry: Commercial data platform vendors, software vendors, system/data integrators, software development companies, commercial data providers, software practitioners.

Open Source Software (OSS) communities: Software practitioners freelancers, DevOps/DataOps.

Research and education: Researchers (computer scientists, data scientists) and students.

Others: Data Ecosystems, Policy makers, Entrepreneurs, Society at large.

EXPECTED RESULTS

Robust and practical software toolbox (enRichMyData toolbox) to support data enrichment processes, consisting of

6 software tools (DiscoverR, WrappR, CleanR, LinkR, StructR, ClassifiR) for data enrichment operations and

5 supporting infrastructural services (ResourcR, ScalR, ReusR, StreamR, GreenR), validated in

6 business products (MarkEn, AIWelA, ServiceEn, EUREKA, InnoGraph, IndEn).

Join the session “Fueling Industrial AI with Data pipelines” today at 4pm to learn more about the project!

Contact: dumitru.roman@sintef.no