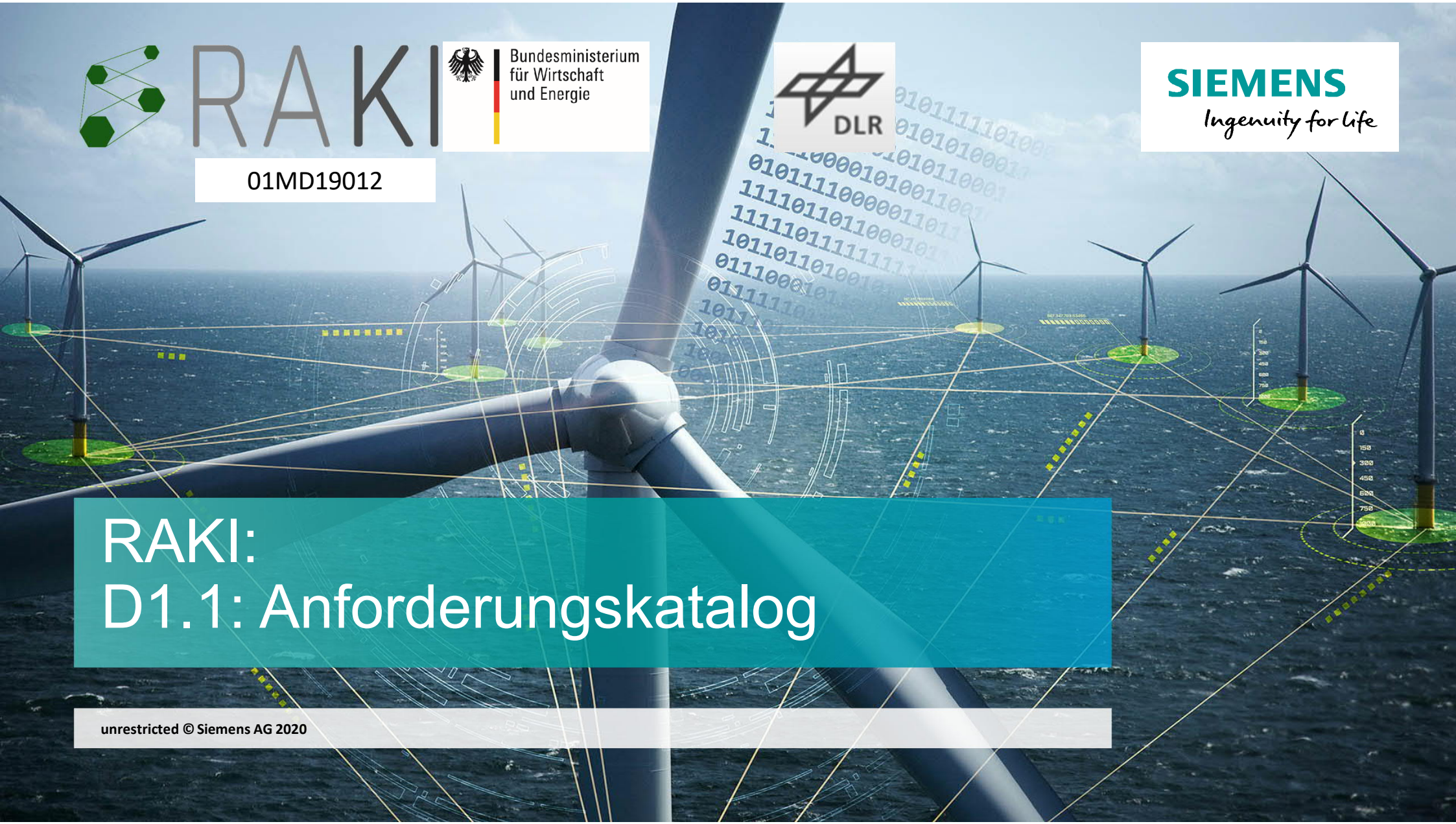

RAKI:
D1.1: Anforderungskatalog

RAKI Requirement Overview

Bundesministerium für Wirtschaft und Energie

RAKI

SIEMENS

Classes of Requirements

- Data Acquisition and Semantic Modeling (1)
- Verbalization (2)
- Scalable Machine Learning on Semantic Data (3)
- Framework and APIs (4)

Data Acquisition and Semantic Modeling (I)

Bundesministerium für Wirtschaft und Energie

RAKI

SIEMENS

1.1 Data Preparation

1.1.1 Identification and removal of redundant information

1.1.2 Merge of data from different sources / files

- Synchronization

1.1.3 Annotation of classes (e.g. detectable anomalies)

- Definition of rules describing the classes
- Applying the formulated rules on the data to label classes
- Calculation of value-added data required for the desired task
- Annotation based on video-data

1.2 Semantic Modelling

1.2.1 common ontology for all use cases

- Identification of concepts and attributes using data and expert knowledge
- Modeling the ontology and creating documentation e.g. by provisioning a data lexicon

1.2.2 ontology provision, e.g. file-based or via an RDF store

1.3 Data Transformation

Afterwards, the preprocessed data have to be converted to RDF and provided in a RDF store.

1.3.1 proper tooling for the data mapping for

- CSV to RDF, SQL to RDF
- Conversion to RDF
- Storage of RDF data

Verbalization

2.1 Fluency: flüssiges lesen

2.2 Correctness: Minimum an Fehlern im Text

2.3 Adequacy: dem Kontext angemessen

3.1 Metrics

Based on the use case problem, algorithms should be able to give different weights to measures such as precision, recall or accuracy.

3.1.1 Precision, Recall, Accuracy and F1-Score:

$$\text{Precision: } P = \frac{TP}{TP + FP}$$

$$\text{Recall: } R = \frac{TP}{TP + FN}$$

$$\text{Accuracy: } A = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{F1 Score: } F1 = \frac{2\,P}{P\,R}$$

3.1.2 Runtime, Queries per Second and Query Mixes per Hour

- **Runtime**: The runtime of algorithms or queries is a central metric for evaluating the performance of a system. It measures the time that passes between starting and finishing a task.
  maximum average runtime threshold: 2ms

- **Queries per Seconds (QpS)**: Is a metric that measures at what speed a system can answer queries that are sent to it. Typically, average QpS are measured in a stress test scenario where a mix of queries is sent for a certain period of time.
  minimum average QpS threshold: 500

- **Query Mixes per Hour (QMpH)**: is a metric that measures for a fixed query mix how often a system can answer all those queries within one hour. Compared to QpS, a single long-running query has a stronger influence on the QMpH value.
  minimum average QMpH threshold: 7000

SIEMENS

3.2 Input

- RDF format and an OWL ontology.
- Data need access via API or SPARQL endpoint.
- Algorithm needs to be configurable

3.3 Types of Machine Learning / Output

- ML component solves classification questions
- ML component solves regression questions.
- Output: 1 to n OWL axioms
- Output with confidence scores (→ 3.1 metrics for quality of the results)

3.4 Runtime

- Enable rapid reactions in production processes
- Runtime should be limited to a reasonable length
- Guaranteed maximal reaction time

3.5 Scalability and Transferability

- Large amounts of data
- Handle more complex functionalities
- Processes in larger production plants.
- Transferable to slightly modified production scenarios, e.g. generalizable to process industry.

Framework and APIs (I)

Bundesministerium für Wirtschaft und Energie

RAKI

SIEMENS

4.1 Microservice-based architecture

- Decomposition of “business” functionality to re-use services
- functionality available as web service using a REST API / OpenAPI file including data schema.
- Data management is decoupled from a service, they are state-less
- ETL pipelines on-top of the microservices, existing frameworks need to be used.

4.2 Continuous Delivery and Deployment

- Service as a Docker images.
- Build pipelines to create the Docker images
- Docker images be available for all partners
- Proper orchestration for the microservices deployed as Docker containers
- Automated testing and benchmarking

Framework and APIs (II)

Bundesministerium für Wirtschaft und Energie

RAKI

SIEMENS

4.3 Data Format

- RDF (Resource Description Framework) is the data format to work on.

  → good to read in chunks.
- SPARQL as query language on RDF
- Ontologies in RDFS and / or OWL.

4.4 Human-in-the-Loop Accessibility

- Reachability of domain expert,
- Response guarantee,
- Synchronous or asynchronous interaction
- Expert communication and involvement: consistent with the GDPR.