CREXDATA

Critical Action Planning over Extreme-Scale Data

D3.1 - Initial Report on System Architecture, Integration and Released Software Stacks

Version 1.0

Documentation Information

Contract Number	**101092749**
Project Website	https://crexdata.eu/
Contractual Deadline	M18, 06.2024
Dissemination Level	Public
Nature	Report
Author	Ralf Klinkenberg (Altair RapidMiner (ARM))
Contributors	Dimitris Zissis (UAEGEAN), Giannis Spiliopoulos (UAEGEAN), Elias Xidias (UAEGEAN), Arnau Montagud (BSC), Thalia Diniaco (BSC) Miguel Ponce de León (BSC), Beatriz Eguzkitza (BSC), Jordi Roca (HYDS), Manolis Kaliorakis (Kpler), Georgios Grigoropoulos (Kpler), Antonios Deligiannakis (TUC), Jens Pottebaum (UPB), Marcel Ebel (UPB)
Reviewer	Antonios Deligiannakis (TUC)
Keywords	System Architecture, System Integration, Use Cases, Data Integration, Data Processing Workflows, Visual Workflow Design, Distributed Data Stream Processing

CREXDATA has received funding from the European Union's Horizon Europe programme under grant agreement number 101092749.

Change Log

Version	Author	Date	Description Change
V0.10	Ralf Klinkenberg (ARM)	23/05/2024	Document creation, initial draft document incl. table of content (Toc)
V0.20	Ralf Klinkenberg (ARM)	06/06/2024	ToC updated, content about initial system architecture added
V0.21	Ralf Klinkenberg (ARM)	12/06/2024	Content about system architecture and system integration and use cases added
V0.30	Ralf Klinkenberg (ARM)	20/06/2024	Content about system architecture and system integration added.
V0.31	Ralf Klinkenberg (ARM)	24/06/2024	Figures 4 and 5 corrected, content about system architecture and system integration updated
V0.32	Ralf Klinkenberg (ARM)	25/06/2024	Content about API for system integration added
V0.33	Ralf Klinkenberg (ARM)	27/06/2024	Content about API for system integration added
V0.34	Ralf Klinkenberg (ARM)	28/06/2024	Content federated data processing requirements added, section about use case scenarios and data requirements updated
V0.35	Ralf Klinkenberg (ARM)	29/06/2024	Content about federated data processing requirements and section about use case scenarios and data requirements updated
V0.36	Ralf Klinkenberg (ARM)	07/07/2024	Content about federated data processing requirements and section about use case scenarios and data requirements updated
V0.37	Ralf Klinkenberg (ARM)	08/07/2024	Content about use case scenarios and data requirements updated
V0.90	Ralf Klinkenberg (ARM)	10/07/2024	Final version submitted to the coordinator
V1.0	Antonios Deligiannakis (TUC)	10/07/2024	Final Version

Contents

List of Figures

List of Tables

Executive Summary

The vision of CREXDATA is to develop a generic platform for real-time critical situation management including flexible action planning and agile decision making over data of extreme scale and complexity. CREXDATA develops the algorithmic apparatus, software architectures and tools for federated predictive analytics and forecasting under uncertainty. The envisioned framework boosts proactive decision making providing highly accurate and transparent short- and long-term forecasts to end-users, explainable via advanced visual analytics and accurate, real-time, off and on-site augmented reality facilities.

This document describes the initial CREXDATA system architecture, the system components, the system integration approach, and the initial version of the released software stacks of the CREXDATA project. The initial system architecture of the prototype being conceptualized and developed in the CREXDATA project and to be utilized by the CREXDATA pilots and simulators includes a graphical tool for designing processing workflows. This graphical design tool interacts with and influences the system architecture. The system architecture must handle multi-modal data fusion from dispersed sources and of different types (extreme scale data streams, time series data, text data, image data, video data, weather data, etc.) and includes the overall system integration.

This document also describes the use cases, their scenarios, and requirements in relation to the system components that are involved to realize those scenarios as data processing workflows. Hence this documents also describes the use cases and their scenarios, the system components involved, and workflow level aspects. The CREXDATA system is adopted per use case in terms of demonstrator systems. This includes communication interfaces to use case domain-specific systems, implementation of custom data fusion operators to be mapped to generic ones, as well as configuration of CREXDATA components and visualization schemes in case specific User Interfaces (UIs). The CREXDATA use cases include:

- weather-induced emergencies, with pilots in Dortmund, Austria and Finland, as well as an initial application scenario on pluvial flooding
- health, with two application scenarios on epidemiology and multiscale lung infection
- maritime, with two application scenarios on Collision Forecasting and Hazardous Weather Rerouting

In these use case scenarios, data streams are produced and collected over inherently distributed, networked architectures composed of several sites, including edge devices operating in the field, relay nodes and one or more clouds or computer clusters. Continuously accumulating all the extreme scale streaming data at a single site/cloud for executing analytics is severely suboptimal resource-wise, and it compromises the real-time nature of the target applications. Therefore, novel algorithms and frameworks for federated learning and forecasting solutions will optimize the use of the underlying hardware resources while delivering analytics as a service. CREXDATA will exploit in-situ processing capabilities of on-site and remote devices and will distributively allocate the computational load in-network exploiting the capacity of each available resource.

The deliverable supports alignment of interfaces between the integrated systems and use case-specific demonstrators, mapping of terminologies across use cases, and preparation of cross-impact evaluation. As simulators are use case specific, these are covered as an extension of CREXDATA system interfaces. This document reflects the information available at this point of the project and will be updated in future versions.

1 Introduction

The project CREXDATA deals with planning and decision making based on data of extreme scale and complexity. Three use cases of real-time critical situation management will be carried out involving numerous datasets of very different nature. Those datasets, that could be either input or output of algorithms, software tools and hardware devices, will be integrated in a Prediction-as-a-Service (PaaS) system as outcome of the project. Use cases are:

- The Weather Emergencies Use Case is devoted to improving situational awareness in weather emergency situations, so that informed decisions are taken by civil protection avoiding disaster impacts. Two levels of data sources are considered: (i) local environment data including stationary sensors systems, as well as mobile robotic sensor platforms deployed in case of an incident and (ii) global environment data referring to current and forecasted weather data issued by Copernicus covering specific fields as forest fires, health, etc.
- The Health Crisis Use Case considers two different scales: (i) epidemiological models will be used to build digital twins of the movement and infection of populations and (ii) mechanistic multiscale models will simulate different treatments of patients as drug administration or the use of mechanical ventilators in severe patients with collapsed lungs among other interventions.
- The Maritime Use Case aims to develop a vessel routing and route forecasting solutions in emergency situations that performs for all vessels of a fleet simultaneously (instead of on-demand request per vessels).

This document describes the initial CREXDATA system architecture, the system components, the system integration approach, and the initial version of the released software stacks of the CREXDATA project. This document serves to capture all aspects of the use cases and their scenarios in relation to the system components that are involved to realize those scenarios as data processing workflows. Hence this documents also describes the use cases and their scenarios, the system components involved, and workflow level aspects.

Within the CREXDATA project, work package WP3 defines and implements the system architecture that materializes the application scenarios on which the CREXDATA prototype will be based. The objectives are:

- Define an architecture for extreme-scale analytics that enables the seamless interaction and interconnection between its data processing components.
- Enable the graphical design of data processing workflows, in order to facilitate the data analyst in specifying and correctly programming the processing that the data analyst wishes to perform on the data.
- Enable different visual analytic tools and interfaces to be easily plugged in.
- Materialize the architecture into a working prototype.

Deliverable D2.1 provides a detailed description of the use cases, their scenarios, their data, and their requirements. For each use case, WP2 develops or enhances the simulators to be integrated in demonstrators and to be used in pilots of CREXDATA. The system provided through WP3 (covering elements of WP4-5) is adopted per use case in terms of demonstrators. This includes communication interfaces to use case specific systems (like ARGOS in T2.1), implementation of custom data fusion operators (WP3) to be mapped to generic ones, as well as configuration of CREXDATA components and visualization schemes in case specific user interfaces (UIs) (especially for augmented reality (AR), see WP5).

The system architecture must handle multi-modal data fusion from dispersed sources and of different types (extreme scale data streams, time series data, text data, image data, video data, weather data, etc.). In the use case scenarios of this project, data streams are produced and collected over inherently distributed, networked architectures composed of several sites, including edge devices operating in the field, relay nodes and one or more clouds or computer clusters. Continuously accumulating all the extreme scale streaming data at a single site/cloud for executing analytics is severely suboptimal resource-wise, and it compromises the real-time nature of the target applications. Therefore, novel algorithms and frameworks for federated learning and forecasting solutions will optimize the use of the underlying hardware resources while delivering analytics as a service. CREXDATA will exploit in-situ processing capabilities of on-site and remote devices and will distributively allocate the computational load in-network exploiting the capacity of each available resource.

This document will reflect the information available at this point of the project and will be updated in future versions.

1.1 Purpose of this Document

This document describes the initial CREXDATA system architecture, the system components, the system integration approach, and the initial version of the released software stacks of the CREXDATA project. The initial system architecture of the prototype being conceptualized and developed in the CREXDATA project and to be utilized by the CREXDATA pilots and simulators includes a graphical tool for designing processing workflows. This graphical design tool interacts with and influences the system architecture. The system architecture must handle multi-modal data fusion from dispersed sources and of different types (extreme scale data streams, time series data, text data, image data, video data, weather data, etc.) and includes the overall system integration.

This document serves to capture all aspects of the use cases and their scenarios in relation to the system components that are involved to realize those scenarios as data processing workflows. Hence this documents also describes the use cases and their scenarios, the system components involved, and workflow level aspects.

The CREXDATA system is adopted per use case in terms of demonstrator systems. This includes communication interfaces to use case domain-specific systems, implementation of custom data fusion operators to be mapped to generic ones, as well as configuration of CREXDATA components and visualization schemes in case specific User Interfaces (UIs).

The CREXDATA use cases include:

- weather-induced emergencies, with pilots in Dortmund, Austria and Finland, as well as an initial application scenario on pluvial flooding
- health, with two application scenarios on epidemiology and multiscale lung infection
- maritime, with two application scenarios on Collision Forecasting and Hazardous Weather Rerouting

The deliverable supports alignment of interfaces between the integrated systems and use-case-specific demonstrators, mapping of terminologies across use cases, and preparation of cross-impact evaluation, especially in terms of consolidation of application program interfaces (APIs). As simulators are use case specific, these are covered as an extension of CREXDATA system interfaces.

1.2 Relation to Other Project Documents

[GA] Grant Agreement (no. 101092749) with its

[DoA] Description of Actions (DoA, part of the [GA])

[CA] Consortium Agreement

[D1.1] Deliverable D1.1 Quality Assurance Plan

[D1.2] Deliverable D1.2 Data Management Plan

[D1.3] Deliverable D1.3 Ethics Manual

[D2.1] Deliverable D2.1 Data Handling and Use Case Scenario Specifications

1.3 Contribution and Structure of this Document

This document describes the initial system architecture, the system components, the system integration approach, and the initial version of the released software stacks of the CREXDATA project. The document first describes the initial CREXDATA system architecture, system components, system integration approach and application programming interfaces (APIs) and then the three CREXDATA use cases, their requirements for the system architecture (e.g. with regards to various data types, data formats, and data volumes to integrated by the system), the use-case-specific system components for the use cases (e.g. various simulators), and the data processing flow and integration aspects per us case. At the end, the document provides a summarizing overview of the initial software stacks of the CREXDATA system and its components.

1.4 Target Audience

- CREXDATA technology partners ("developers")
- CREXDATA use case partners
  - (end users)
  - technology partners providing demonstrator components
- Third party researchers in system- and application-oriented research domains, especially
  - software architecture and system design
  - system design for extreme-scale data stream processing and predictive analytics
  - software architecture and system design for distributed federated data stream processing and federated machine learning
  - handling of weather emergencies

- handling of health crisis
- maritime event detecting and handling
- security

1.5 Glossary

Abbreviation	Expression	Explanation
UC	*Use Case*	Applications of CREXDATA technology resp. the CREXDATA system in real-world scenarios. Within the project three use cases are defined: weather induced emergencies, health, and maritime.
	Pilot (site)	Conceptual term to describe a set of stakeholders within their context like spatial environment, equipment, data sources etc. For each Use Case, several Pilot (sites) can be specified (for instance, Dortmund and Austria in the emergency use case).
	Application Scenario	Procedural and structural description of potential uptake of CREXDATA technologies in Use Cases (for instance, flooding and forest fires in the emergency case).
	CREXDATA System	Output of WP3, integrating technologies created in WP4 and WP5 without use case-specific customizations. It includes customization and configuration functionality, esp. through graphical workflow management.
	Demonstrator (System)	Technical system based on the CREXDATA system, which is customized and configured for specific Use Cases, Pilots and/or Application Scenarios. The Demonstrator might include additional components both as data sources and sinks (for instance, legacy systems of end users or the ARGOS system in the emergency use case).
	(Data Processing) Workflow	A Data Processing Workflow (or short Workflow) describes the steps to be performed from data ingestion from the data source systems (e.g. sensors, IoT and edge devices), data processing, data analysis, model generation with machine learning (if applicable), predictions and forecasting, to model deployment and visualization.

Abbreviation	Expression	Explanation
		Workflows can be designed and edited visually in the Workflow designer/editor.
	(Data Processing) Operator	A step in a Data Processing Workflow is called an Operator. Operators are functional modules or building blocks.

2 Initial CREXDATA System Architecture

This chapter describes the conceptual system architecture and its layers and components (Section 2.1) and the more technology-focused system integration architecture (Section 2.2), technologies and platforms chosen on which the CREXDATA system will be implemented, the APIs (Application Programming Interfaces) for the seamless interconnection of all components and their integration (Section 2.3), the graphical workflow flow designer (Section 2.4) and key components to be integrated (like for example use-case-specific simulators) (Section 2.5).

2.1 CREXDATA Concept and Reference System Architecture for Use Case Demonstrators

An initial system architecture is provided in the Description of Action (DoA) of the CREXDATA project [DoA, part B, p.16]. The core CREXDATA system is developed and configured through an integrated development environment (IDE). It comprises application logics of all technologies that are subject of WP3 to WP5. Figure 1 provides an overview of the major elements of the CREXDATA system and a general view on interfaces (input and output).

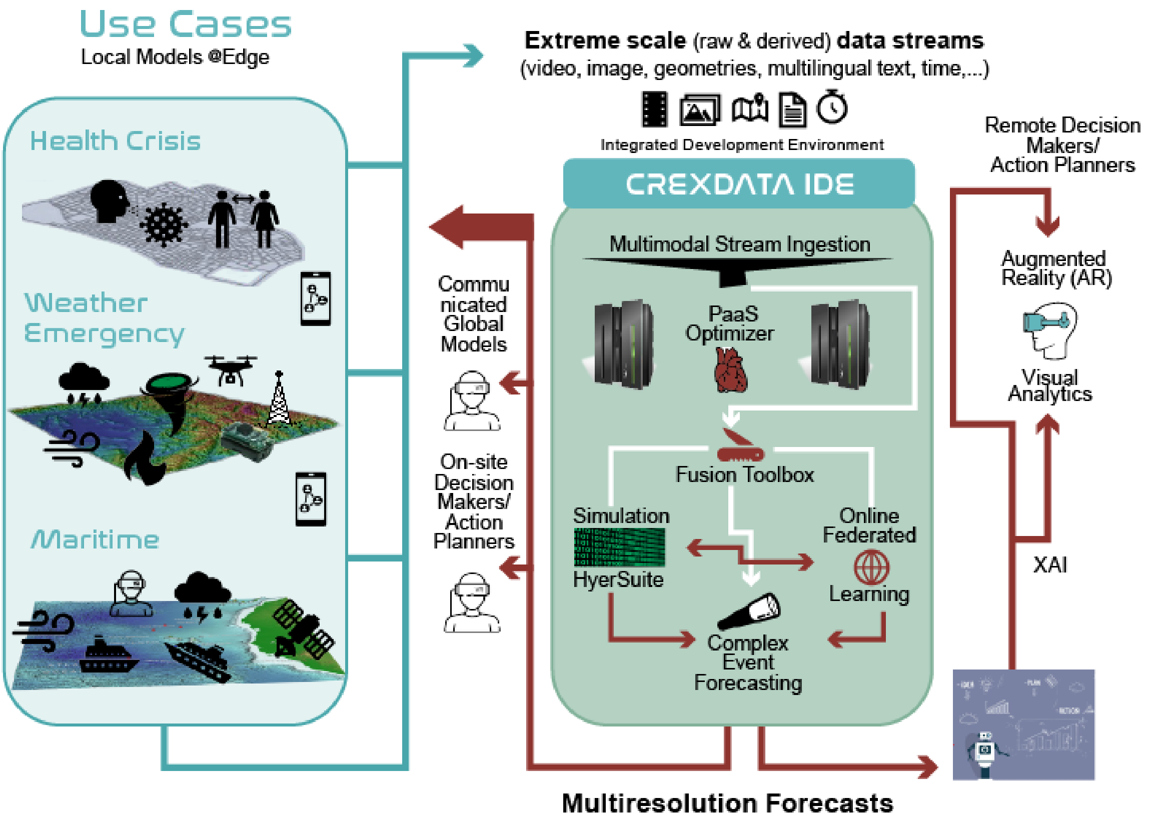

Figure 1: CREXDATA Concept Overview (www.crexdata.eu, based on [DoA, part B, p.16]).

The core CREXDATA system is integrated in work package WP3. The resulting system itself is made available for integration in demonstrators per use case. Thus, it can be understood as a sub-system of these demonstrator systems. Demonstrator system architectures are described per use case in Sections 3.1.2 (weather emergency case), 3.2.3 (health crisis) and 3.3.4 (maritime use case).

The following layers can be identified in the CREXDATA core system architecture:

Multimodal Data Stream Processing

The CREXDATA system consumes data from a variety of data sources. For training of models etc., access to data sets is required. For real-time data processing, streaming data from various potentially distributed data sources needs to be handled and integrated. Therefore, technical solutions like Apache Kafka or RabbitMQ, general data formats like JSON and XML as well as domain specific data schemas need to be considered. Use case-specific data sources are identified in Section 3 and detailed in deliverable D1.2 [D1.2]. In CREXDATA, these requirements will be mainly implemented by the Altair RapidMiner Data Science and AI Platform via data access modules (RapidMiner operators, data connectors, and project-specific extensions) in combination with use-case-specific systems (e.g. ARGOS in the CREXDATA weather emergency use case), leveraging Kafka clusters for the interaction between system components and for data streaming between system components, Altair RapidMiner AI Studio (desktop) with CREXDATA-specific extensions for visually designing the data processing workflows, which are then executed on distributed servers and compute clusters (leveraging e.g. Altair RapidMiner AI Hub (server or cloud instances)) as well as distributed and deployed on IoT and edge devices (using Altair RapidMiner Real-Time Scoring Agents (RTSA) on the edge and on IoT devices). Key components of the multimodal data stream processing system architecture:

- Multimodal Stream Ingestion Modeler (offline): Graphical editor for visual data processing workflow design and modelling, operators from data stream ingestion through data analysis to visualization and forecasting are needed. Therefore, extensions need to be implemented for data sources (per use case), for demonstrator system elements (like algorithms implemented in ARGOS in the CREXDATA emergency use case), and especially for CREXDATA algorithms developed in WP4/WP5.
- Prediction-as-a-Service (PaaS) Optimizer: For testing and evaluating the PaaS algorithms, workflows need to be specified that include various computational elements distributed across technical systems (hardware resources, processors). There needs to be an opportunity to optimize computations across available resources (e. g., from drones to base stations to cloud services in the emergency use case).
- Fusion Toolbox: The toolbox needs to be capable to execute entire data processing workflows and models specified by the Multimodal Stream Ingestion modeler in real-time based on streaming data.

The definition of operators (functional modules and data connectors) is both driven by use cases (data sources, demonstrator elements) and technology developments (WP4-WP5). Even though there is no algorithmic interdependency of PaaS optimizations with other CREXDATA technologies, workflows should be distributed across different computational devices/services to enable distributed data processing (e.g. for federated machine learning) and distributed use case scenarios.

Simulation HyperSuite

The Simulation HyperSuite is created with the intention of a domain-specific integration of relevant simulators. Thus, it can be seen as a frame for use case-specific simulator sub-systems. The Simulator HyperSuite is developed in work package WP2, providing data to algorithms developed in work package WP4-WP5. Details are provided in Section 2.5.

Machine Learning

Detailed use cases of Machine Learning (ML) typically require available data to learn from. Regarding stakeholder requirements elicitation, a guiding question is "Are there situations in which you think that data is available, but so far there is no way to make sense out of it?". For training models, there needs to be an option to create or acquire labelled data sets (ground truth).

- Online Federated Machine Learning: "online" and "federated" are two attributes that focus the solution space with regards to ML in CREXDATA. Sequential steps can help requirements elicitation, from general relevancy of ML to online ML, and then even federated ML.
- Interactive Learning for simulation exploration: Simulators are part of demonstrators in all three use cases. In contrast to ML models, simulation models are transparently known but need to be guided towards an optimum to identify best fitting parameter sets. In CREXDATA, ML-based algorithms are integrated to guide and reduce the efforts for such an exploration of the parameter space. Thus, relevant simulation types need to be identified and the principles of exploration need to be determined.

These ML algorithms are not dependent on other algorithms in WP4. Interactive learning regarding simulations needs to be aligned with Simulation HyperSuite developments to cover relevant simulations.

Complex Event Processing

Complex Event Processing (CEP) includes both, Complex Event Recognition (CER) and Complex Event Forecasting (CEF). Based on an input data stream, events are recognized and forecasted based on, for instance, logical expressions. An event might mean a single point in time (e. g., threshold at river gauge reached) or it might have a duration (e. g., person endangered by rising river gauge). Event patterns are either learned from data (cf. ML) or defined based on semantics of a domain of interest (i.e., the three use cases).

- CEF: Event patterns are modelled by logical formulae and transition systems. Simple events represent events close to data sources, while complex events hold as soon as relevant simple events occur. Inputs to event forecasting need to be provided as multi-variable time-series data, for instance, through a single Kafka topic.
- Text Mining: In the context of CREXDATA, text mining is understood as a synonym for Natural Language Processing (NLP). For instance, RapidMiner includes a Text Processing operator. The intention is to learn from social media streams or to detect events in social media streams. As a rough differentiation, both identification of highly relevant postings and analysis across a large number of postings are relevant. While in the first case, input for CEF might be generated, in the latter one this might be coupled with simulation exploration (verifying simulated futures).

CEF could be based on data streams created by other CREXDATA components (e. g., online federated learning and esp. text mining), but does not need such interdependent setups.

Explainability and Visualization

Explainable Artificial Intelligence (Explainable AI (XAI)) and Visual Analytics require models and data as input to extend them with explainability and visualization layers. Thus, these algorithms are strongly dependent on ML and EP algorithms.

- Explainable AI: A focus should be set to ML models as input, independent from specific use cases (online, federated, interactive). The intention is not to implement own ML models, but to extend existing ones.
- Visual Analytics: Prerequisites are similar to XAI. For Visual Analytics, event streams (recognized and/or forecasted) are of similar interest like ML models. There is a direct interconnection with XAI models.
- Uncertainty Visualization: Situational awareness in decision-making situations includes an understanding of uncertainty in available data resp. information. For human actors, visualization of uncertainty semantics is required. Such semantics can be based, for instance, on established concepts of data/information quality dimensions, criteria and indicators (DQ/IQ).

Algorithms adding explainability and visual features require ML and/or CEP models as a basis. Therefore, requirements need to target the cascade from data sources through ML/CEP to XAI/Visual Analytics.

Augmented Reality

Augmented Reality (AR) in CREXDATA shall be tested by Head-Mounted Displays (HMDs) like HoloLens 2. AR provides very special means to visualize data and information. It is based on the principle that reality is enriched by superimposed information (i. e., not only virtual objects like in Virtual Reality/VR). Therefore, specific use cases need to involve stakeholders close to an environment or to objects of interests. Information is visualized with reference to such real-world perceptions. AR is implemented in terms of AR applications (e. g., using the development environment Unity3D). Interfaces to other services (i. e., algorithms) need to be incorporated into such standalone applications.

AR is relevant in very specific situations where real objects are of interest. In CREXDATA, AR is meant to be one possible User Interface (UI) to access results from data processing pipelines. Thus, AR might be added as a sink in graphical workflows, visualizing results from ML, CEP, XAI and Visual Analytics.

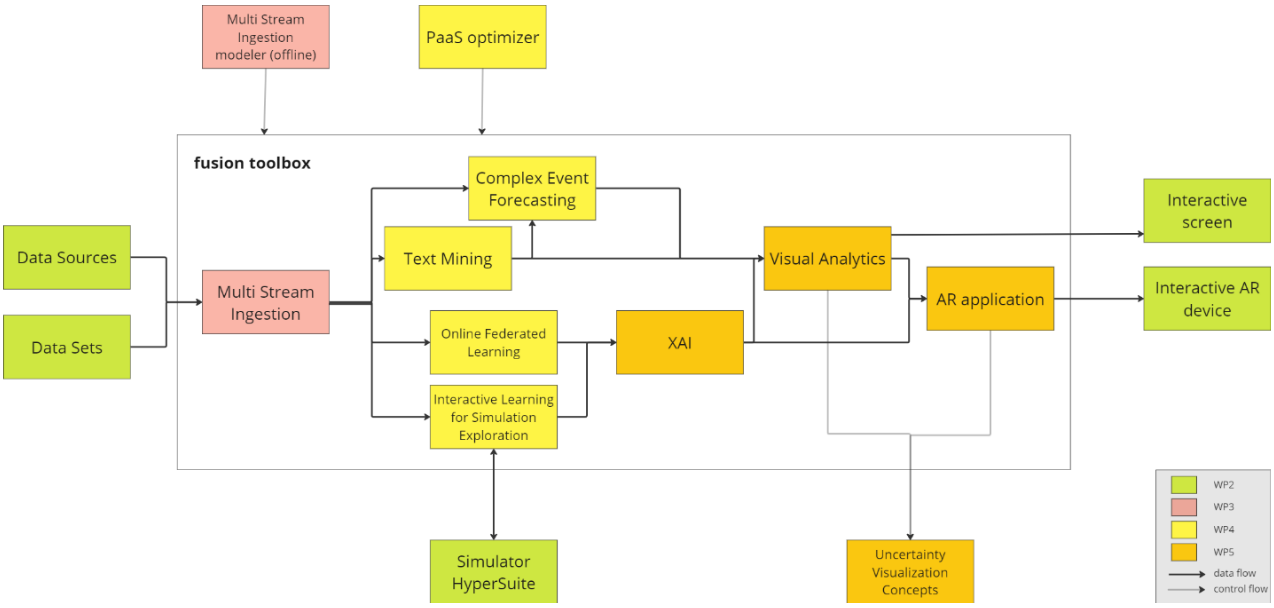

Figure 2: Elements of the CREXDATA system (to be integrated in demonstrator systems).

The following diagram visualizes the work package (WP) level interaction in CREXDATA:

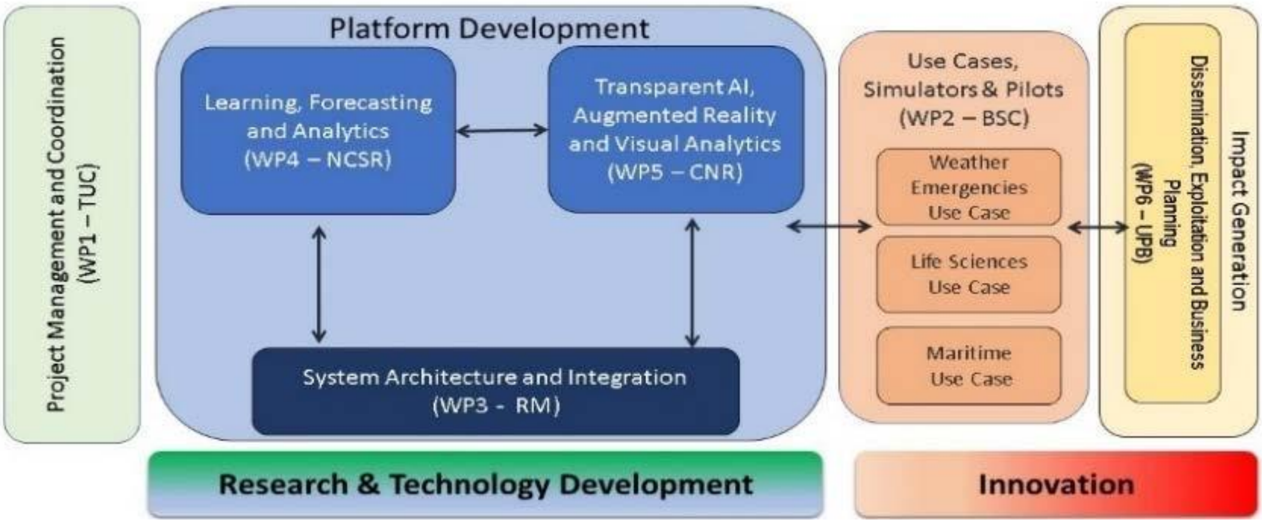

Figure 3: CREXDATA system architecture and integration with respect to work packages (WP) and WP level interaction (from [DoA] CREXDATA Project Proposal, part B, page 30, figure 3).

2.2 CREXDATA System Integration Architecture

This section describes the CREXDATA system integration architecture and the technologies and platforms chosen on which the CREXDATA system will be implemented.

The main objectives for the CREXDATA architecture are

- enable graphical data processing workflow design and automated deployment
    - low code / no code workflow development through visual workflow design (ease of use by visual data processing workflow design without the need to code)
    - platform-agnostic workflows for supported streaming platforms (ease of use by abstraction)
    - automated deployment of the visually designed workflows on the execution platforms, including servers, clusters, cloud, edge and IoT devices (ease of use via automated distribution and execution)
- extensible design to integrate project- and use-case-specific components and tools
    - pluggable connections to data sources and sinks
    - ingest data from components into workflow
    - export data out of workflow (from intermediary or final step)
- enable extreme scale data analytics
    - supported backends (for stream execution or batch execution) are scalable and support distribution.

The CREXDATA system needs to be able to integrate data in heterogeneous formats from distributed data sources, both static data as well as continuous data streams, process the data either on the edge or IoT device or on a central server, compute clusters, or cloud or distributed and in parallel across different system components (e.g. for distributed data processing and distributed machine learning). Hence, we chose underlying platforms that enable the CREXDATA system to fulfil all these requirements.

- For the visual design of data processing workflows, we chose and extend the visual data processing workflow designer Altair RapidMiner AI Studio.
- For central data processing, we chose and extend the server version of the Altair RapidMiner data science platform, i.e. Altair RapidMiner AI Hub.
- For distributed parallel data processing, we leverage a special version of Altair RapidMiner AI Hub optimized for low system resource requirements, low latency, and high throughput, namely Altair RapidMiner Real-Time Scoring Agents (RTSA), which can be deployed on distributed IoT devices, edge devices, remote servers, and distributed compute clusters.
- For the system integration, the system components use
    - web services for sending requests and responses (control flow) and
    - data streams and event streams (data flow).
- For the web services, Application Programming Interfaces (APIs) specify the exact format of the supported web services requests and responses in JSON format (see Section 2.3 for more details).
- For the event streams and data streams, we leverage the open-source frameworks Kafka and Flink.
- For the simulations in the various use case scenarios, domain- and use-case-specific simulators are developed and used (see Sections 2.4 and 3 for more details).

For the integration and interaction of the system components, we use web services as Application Programming Interface (API). The web service requests and responses and their JSON formats are described in more detail in Section 2.3.

For exchanging data between the system components, we use Kafka data/event streams. Data sources (data producers) write their data (e.g. sensor data from IoT devices or the other edge devices) into Kafka streams, from which other system components can read these data streams (see Figures 4 and 5 below) and write their responses to other Kafka streams (e.g. events detected or predicted by automated complex event detection and prediction (CEP) or weather or flood forecasts or maritime vessel collision predictions, etc.).

The following diagrams (Figures 4 and 5) show the overall technical CREXDATA system architecture and system integration approach leveraging JSON-based Web Service APIs (Application Programming Interfaces) and Kafka data/event streams and how the CREXDATA system components interact (via web service requests and responses) and exchange data (by writing data to and reading data from Kafka streams) and thereby how users can define data processing workflows using the CREXDATA system.

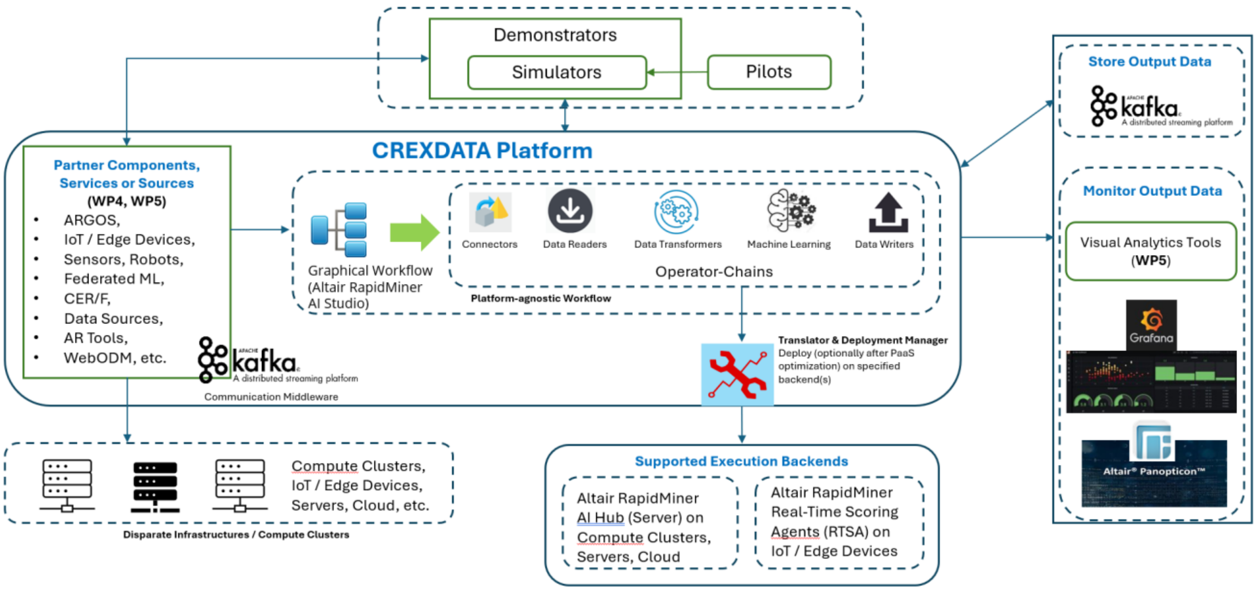

Figure 4: Technical CREXDATA system architecture and system integration approach via web services (control flow) and Kafka clusters (data flow, e.g. for sensor data streams from IoT devices and edge devices, as well as for stream of analysis results and predictions).

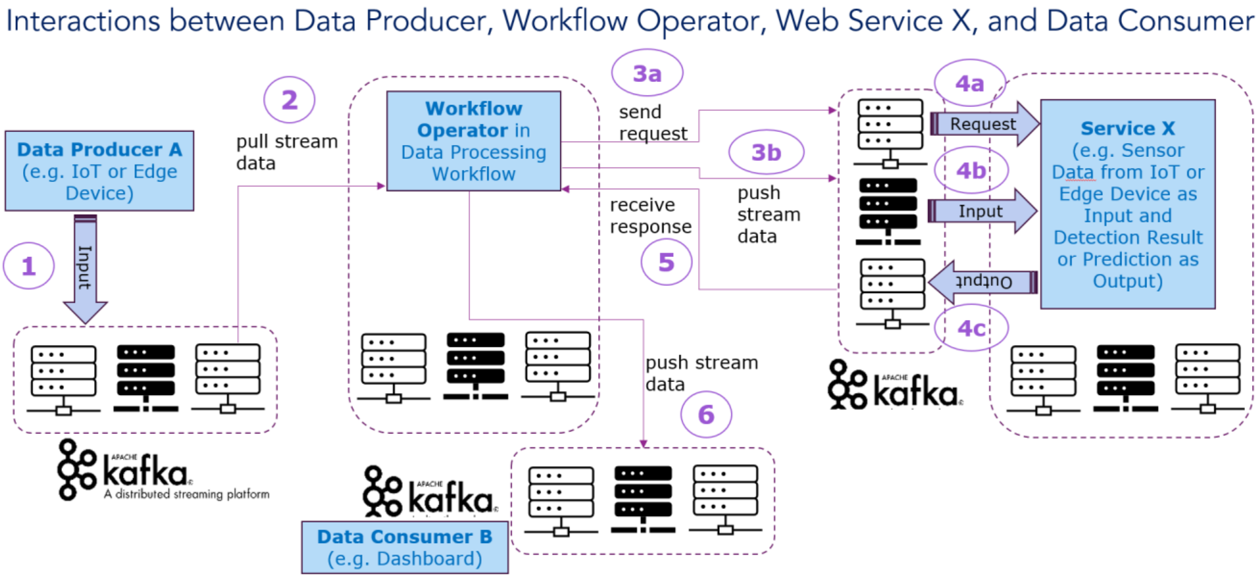

Figure 5: Interactions between data producers, workflow operator, web service, and data consumer: process control flow and data flows via web services and Kafka data/event streams.

The following two sections provide more details about the APIs (Section 2.4) and how the users can define data processing workflows using the visual workflow editor of the CREXDATA system (Section 2.3).

2.3 Graphical Workflow Designer and Data Processing Platform

This section describes graphical workflow flow designer of the CREXDATA platform, a graphical tool that enables non-expert programmers and workflow designers to visually specify complex data processing workflows. The graphical workflow designer is implemented as an extension of the graphical RapidMiner workflow designer (Altair RapidMiner AI Studio). The workflow designer supports the integration and fusion of heterogeneous data from dispersed data sources. In CREXDATA, we

- specify new and customized data processing and data fusion operators, where operators are functional modules or process steps represented as graphical elements (boxes) within the graphically represented data processing workflow, and
- specify processing tasks as data processing workflows (using operators), and these
- designed workflows are automatically translated to code that will run in the underlying CREXDATA architecture.

Altair RapidMiner is a unified data analytics and artificial intelligence (AI) software platform for all stages of data analytics – from data ingestion from heterogeneous data sources, data integration and fusion, data (pre-)processing, machine learning and modelling, model validation to model operationalization (deployment), and orchestration of data processing workflows, which can handle static data as well as continuous data streams and which can be executed in distributed processing nodes on servers, clusters, clouds, desktops, IoT and edge devices (see Figures 6 to 11 on the following pages).

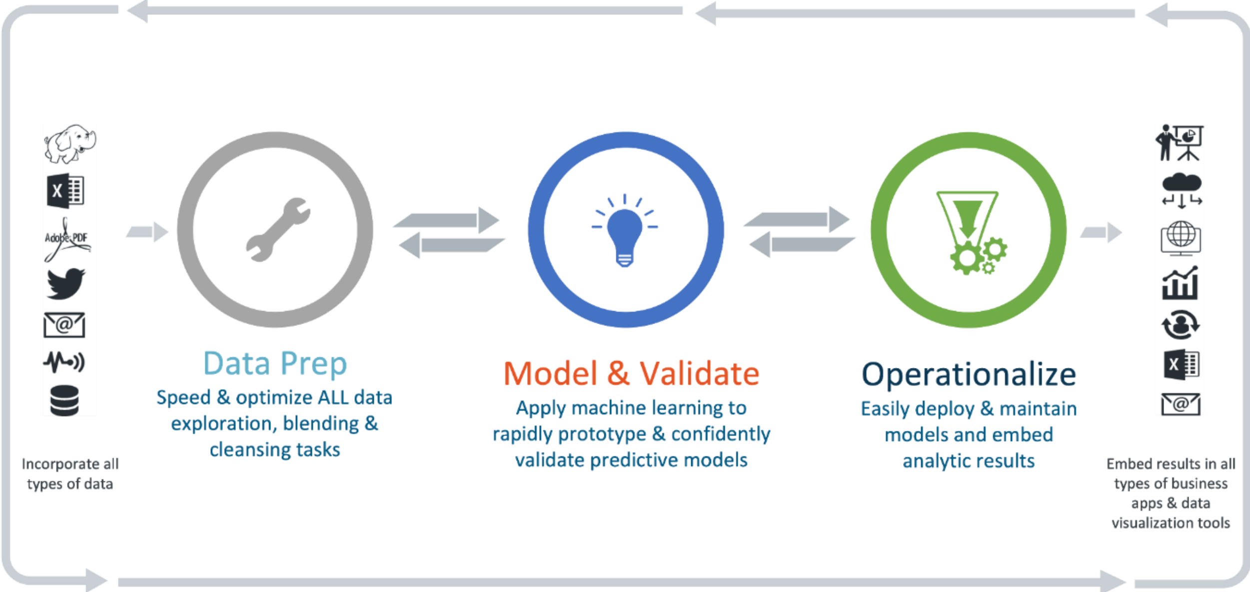

Figure 6: Altair RapidMiner platform as unified data platform from data integration from heterogeneous data sources, data (pre-)processing, modelling with machine learning, and model validation to model deployment (operationalization).

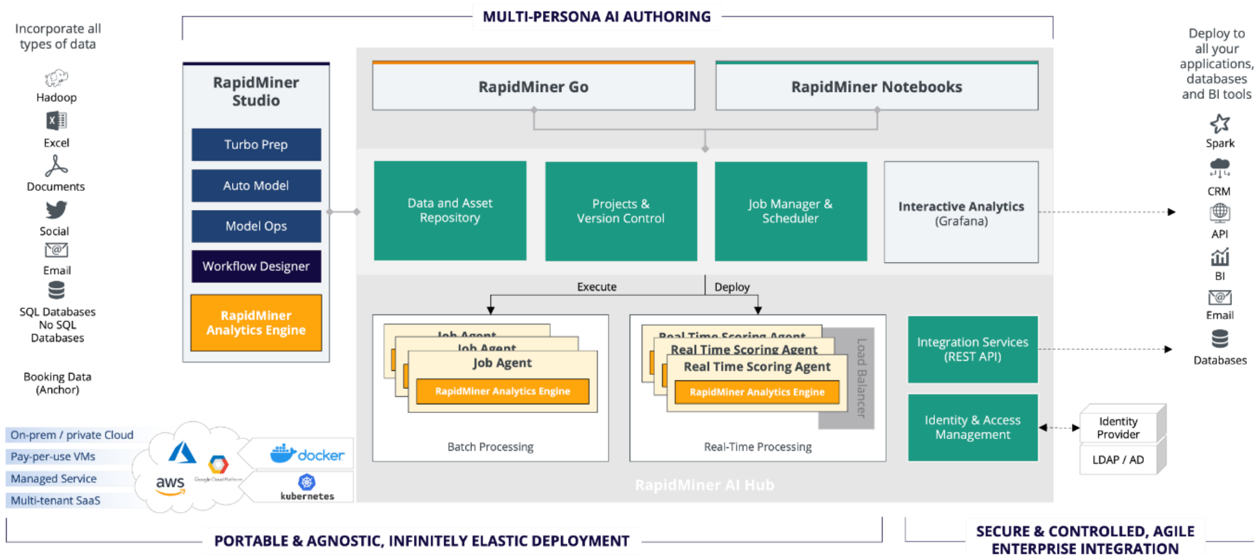

Figure 7: The RapidMiner platform overview including the visual workflow designer Altair RapidMiner AI Studio, the server Altair RapidMiner AI Hub, and the Altair RapidMiner Real-Time Scoring Agents (RTSA) for the process execution on the edge or on IoT devices.

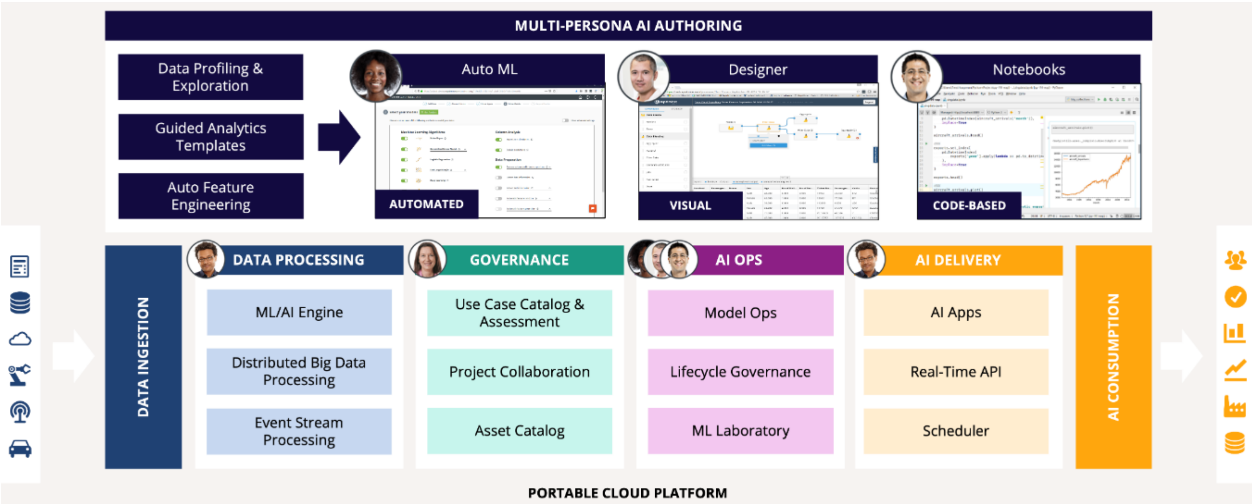

Figure 8: RapidMiner cloud platform architecture and server architecture

Altair RapidMiner AI Studio is a visual editor for designing complete data analytics workflows (see Figure 9). AI Studio supports a rich collection of over 100 machine learning (ML) algorithms and statistical functions out-of-the-box and is extensible using a plugin-based framework allowing to flexibly extend the platform and to integrate other tools and ML libraries. Process workflows are visually designed in AI Studio and can then be executed anywhere, locally on the desktop (in AI Studio), remotely on servers (on AI Hun) or in the cloud (on AI Hub) or on compute clusters or distributed on IoT or edge devices (on an RTSA).

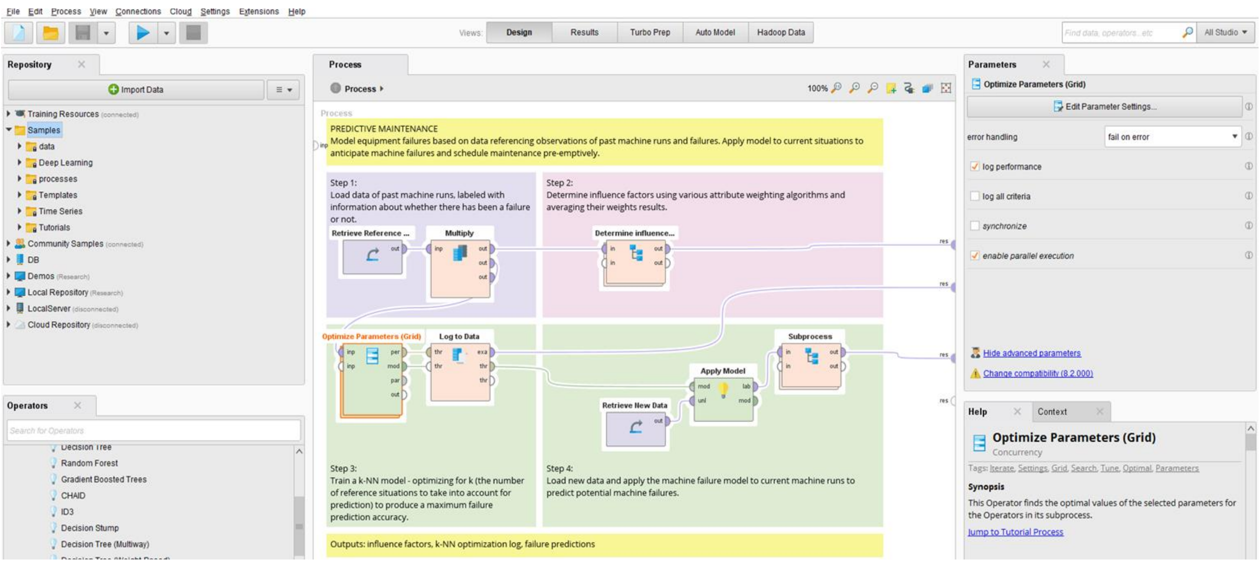

Figure 9: Graphical User Interface (GUI) of RapidMiner AI Studio for graphically designing data processing workflows with an exemplary visually defined data processing workflow (here for predictive maintenance, i.e. for predicting and avoiding machine failures before they occur).

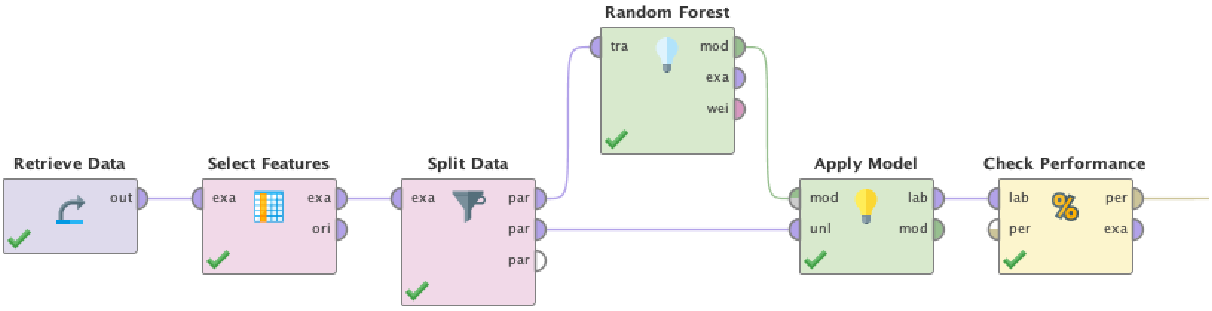

Figure 10: Exemplary RapidMiner data processing workflow for training, application, and testing (validation) of a machine learned model (ML model).

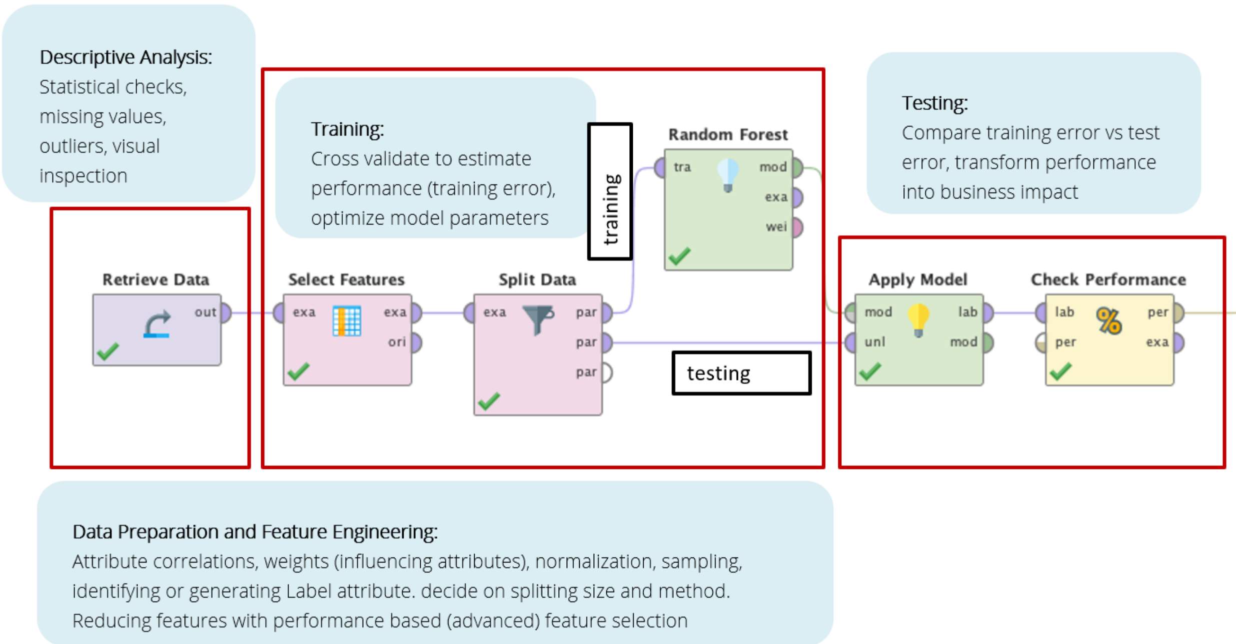

Figure 11: Exemplary RapidMiner data processing workflow for training, application, and testing (validation) of a machine learned model (ML model) with explanations for the process steps.

Kafka is leveraged as a communication middleware to integrate external components like IoT and edge devices, sensors, event emitting servers, etc. Kafka data streams (topics) are used to pass data between the systems and their components (i.e. to read and write streaming data). The following figure shows an example streaming analysis workflow to integrate external components via Kafka topics, i.e. for a fictitious ‘critical action planning’ scenario: plan a rescue action by interpreting current weather in context of historical risks:

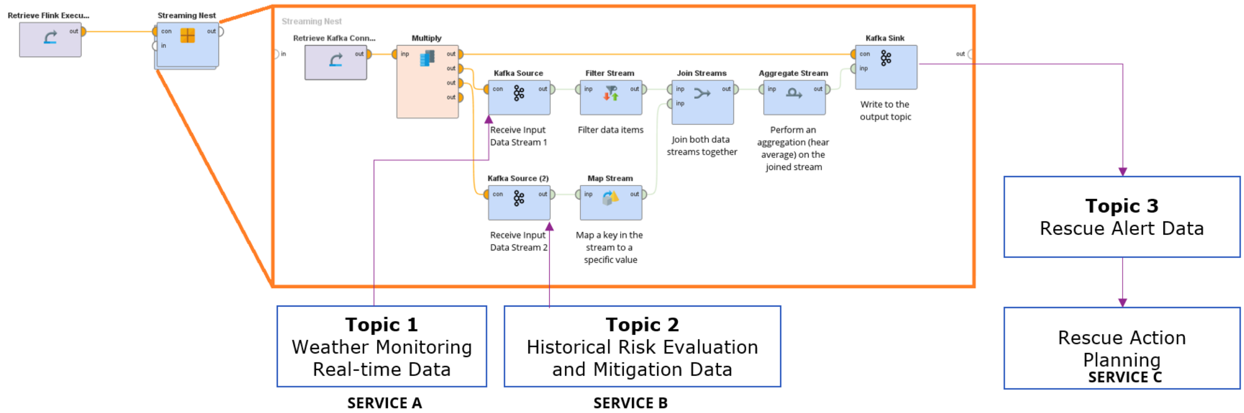

Figure 12: Example streaming analysis workflow to integrate external components via Kafka topics, i.e. for a fictitious 'critical action planning' scenario: plan a rescue action by interpreting current weather in context of historical risks.

The RapidMiner-based CREXDATA system architecture offers different approaches for executing data stream processing workflows on the CREXDATA platform. One option, depicted on the left side in Figure 13, is via "Streaming Nest" operator, which translates the workflow to and runs them on big data streaming platforms like Flink and Spark Streams, e.g. for big data stream processing where visual workflows are translated to Flink code and run in parallel on big data clusters, or another options, depicted on the right side in Figure 13, is via a "Federation Nest" operator for running big data stream processing workflows on remote Real-Time Scoring Agents (RTSA), e.g. on servers, clusters, clouds, but also on edge nodes or IoT devices, e.g. for federated data processing or federated machine learning, without the need to install Flink on these devices and without the need to translate visual workflows to Flink code and thereby allowing to execute any RapidMiner operator on the RTSA(s). For CREXDATA, the latter is the preferred approach, because it does not require a translation of the workflows to Flink, it does not require Flink on the IoT devices and edge nodes, and it offers a broader range of operators (functional modules) across all platforms (i.e. all RapidMiner operators).

In case of the "Streaming Nest" operator, which was developed in the EU-funded project INFORE, the (sub)process (workflow) within the streaming nest is automatically translated to Flink code on distributed to and executed in parallel on big data clusters running Flink. This requires that every machine or device has Flink installed, which may be to resource-intensive for IoT and edge devices. It also only works for RapidMiner operators, for which automated translations to Flink exist.

In case of the "Federation Nest" operator, which is developed in the CREXDATA project, the data processing does not require Flink or a translation of workflows to Flink code. Instead, the (sub)process (workflow) with the federation next is distributed to RapidMiner Real-Time Scoring Agents (RTSA), which require only limited resources on the target machines and devices, and which can execute every RapidMiner operator and process.

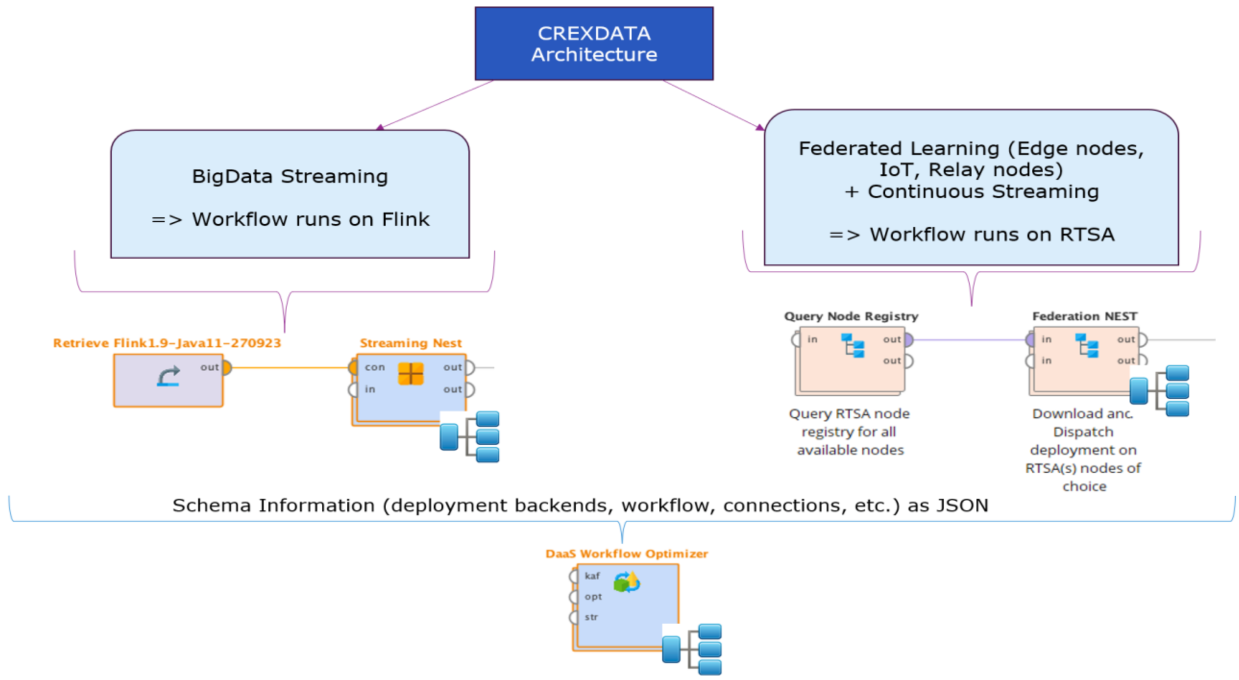

Figure 13: Different approaches for executing data stream processing workflows on the CREXDATA platform: via a "Streaming Nest" operator on Flink (left), e.g. for big data stream processing where visual workflows are translated to Flink code, or via a "Federation Nest" operator for running big data stream processing workflows on remote Real-Time Scoring Agents (RTSA), e.g. on edge notes or IoT devices, e.g. for federated data processing or federated machine learning, without the need to translate visual workflows to Flink code and thereby allowing to execute any RapidMiner operator on the RTSA(s). For CREXDATA, the latter is the preferred approach, because it does not require a translation of the workflows to Flink, it does not require Flink on the IoT devices and edge nodes, and it offers a broader range of operators (functional modules) across all platforms (i.e. all RapidMiner operators).

For the CREXDATA system to be able to distribute the execution of data processing workflows across servers, clusters, clouds, IoT and edge devices, etc., every compute node of the distributed network to be used for data processing needs to be registered in a node registry (see operator "Query Node Registry" in Figure 13).

Figure 14 shows the general interaction of the CREXDATA platform components and the workload distribution on various Real-Time Scoring Agents (RTSA) at different sites:

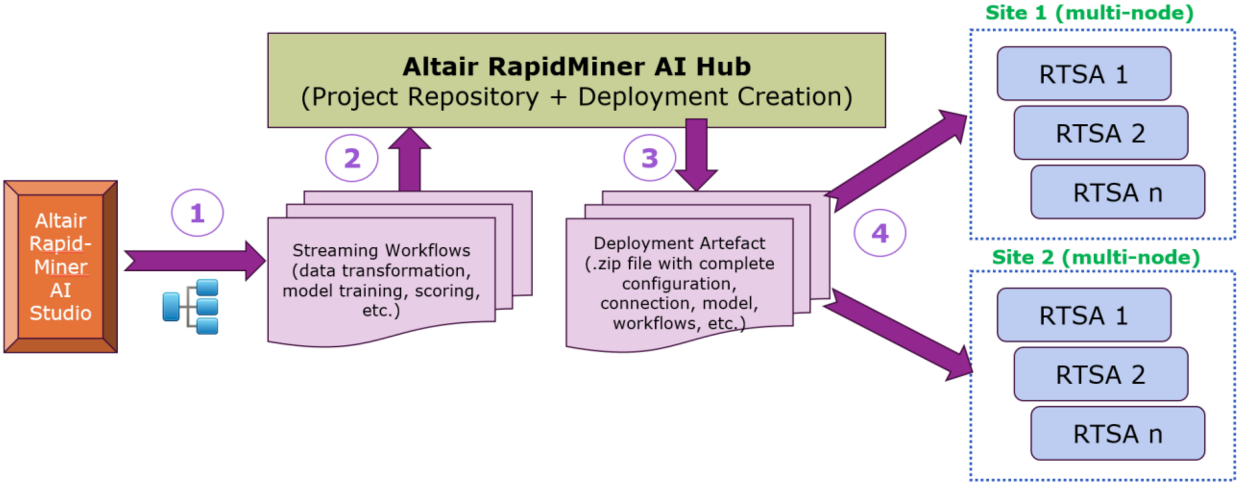

Figure 14: General interaction between the visual workflow designer (AI Studio), the data processing server (AI Hub), and numerous execution nodes (RTSA 1..n) at different sites, supporting non-federation and federation use cases (e.g. executing processes on the edge, on IoT devices, on compute clusters, or in the cloud/fog).

This interaction flow is used for the federated data processing use case with distributed IoT and edge devices. Altair RapidMiner AI Hub automates the deployment creation and dispatches the deployment artefacts to the different sites, where these artefacts are executed on Altair RapidMiner Real-Time Scoring Agents (RTSAs). The RTSAs are preconfigured environments to support workflow execution, making connections between Kafka data sources and Kafka data sinks, and can optionally run Python code (scripts) if needed.

Figure 15 shows the RTSA infrastructure including the management of the execution nodes (data processing nodes) via the query node registry and automated workflow deployments:

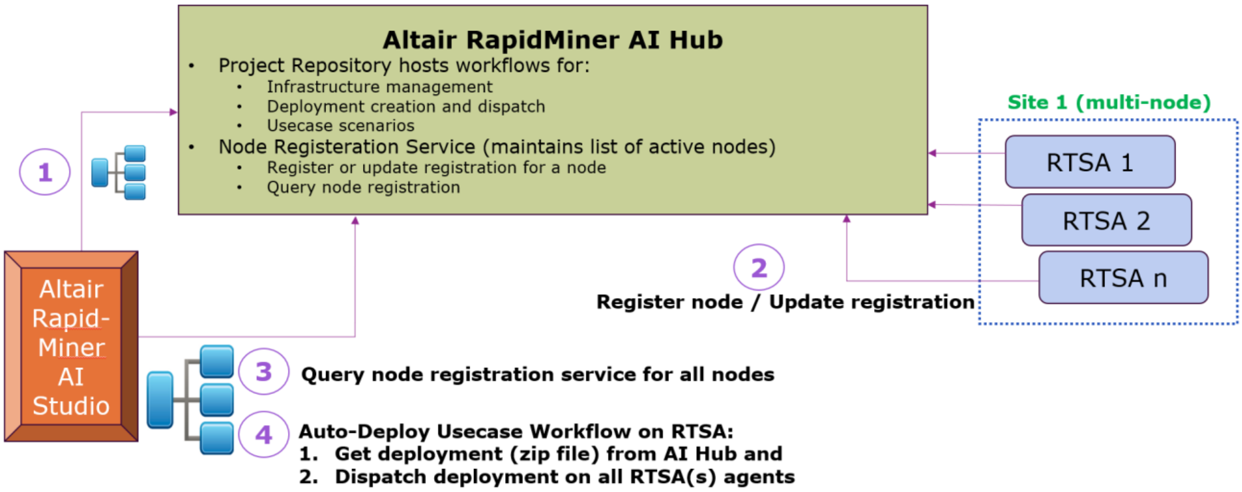

Figure 15: Real-Time Scoring Agent (RTSA) infrastructure incl. node management and node registry and automated workflow deployments on distributed RTSAs.

Infrastructure management workflows are exposed as AI Hub endpoints like shown in the screenshot of the AI Hub web service endpoint management user interface:

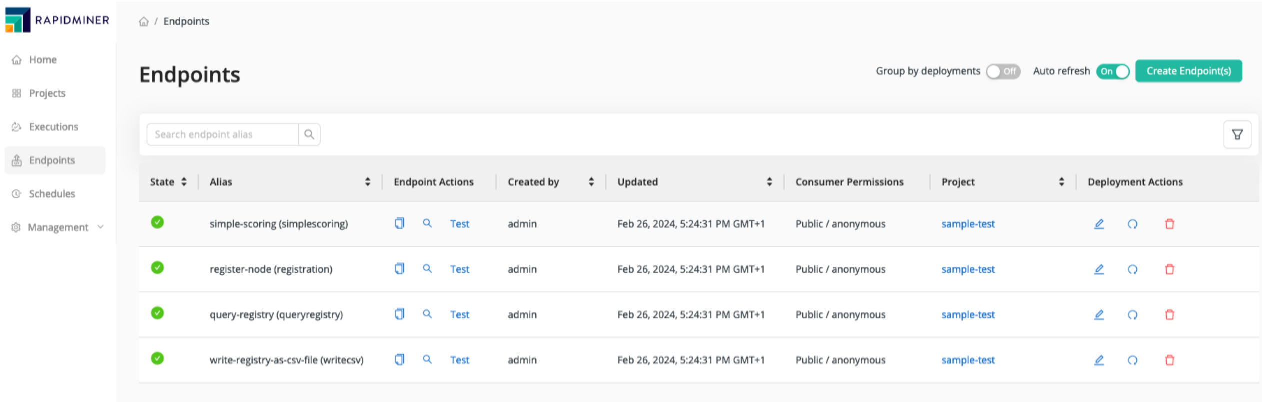

Figure 16: Infrastructure management workflows are exposed as AI Hub endpoints (screenshot of the AI Hub web service endpoint management user interface).

Each AI Hub web service is provided by an underlying RapidMiner process (workflow). Accordingly, each CREXDATA infrastructure management web service implemented as an AI Hub web service endpoint has an underlying infrastructure management workflow, which can also be visually designed like any other data processing workflow using AI Studio.

The following three figures (Figures 17 to 19) provide examples of such CREXDATA infrastructure management workflows and use-case-oriented data processing workflows, which were all visually designed in AI Studio.

For the deployment creation and dispatch workflow, the following blueprint method shown in Figure 17 is used.

The available data providers (e.g. IoT and edge devices with sensor data) and compute resources (compute nodes) are registered in a node registry. The user can filter nodes of interest and deploy use case workflows on them (to be customized for the respective application use case).

Figure 18 shows the “Filter Nodes” operator as part of an example workflow and the user interface of the “Filter Nodes” operator for defining the filters for filtering the list of available nodes received from the note registry. Then the operator “Deploy Workflow on Node” automatically deploys the workflow to be executed on the selected nodes.

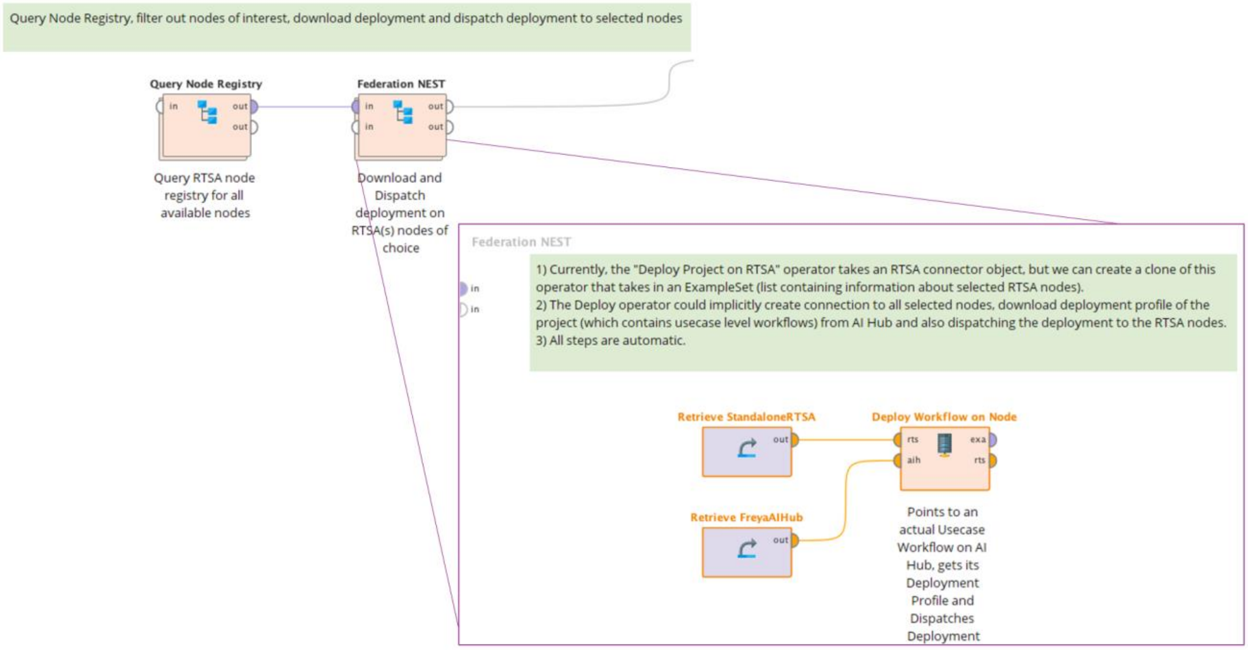

Figure 17: Deployment creation and dispatch workflow (blueprint method): The first subprocess queries the RTSA node registry for all available nodes and filters out the nodes of interest. The second subprocess, i.e. the "Federation Nest" operator and its inner process dispatch the deployment to these nodes.

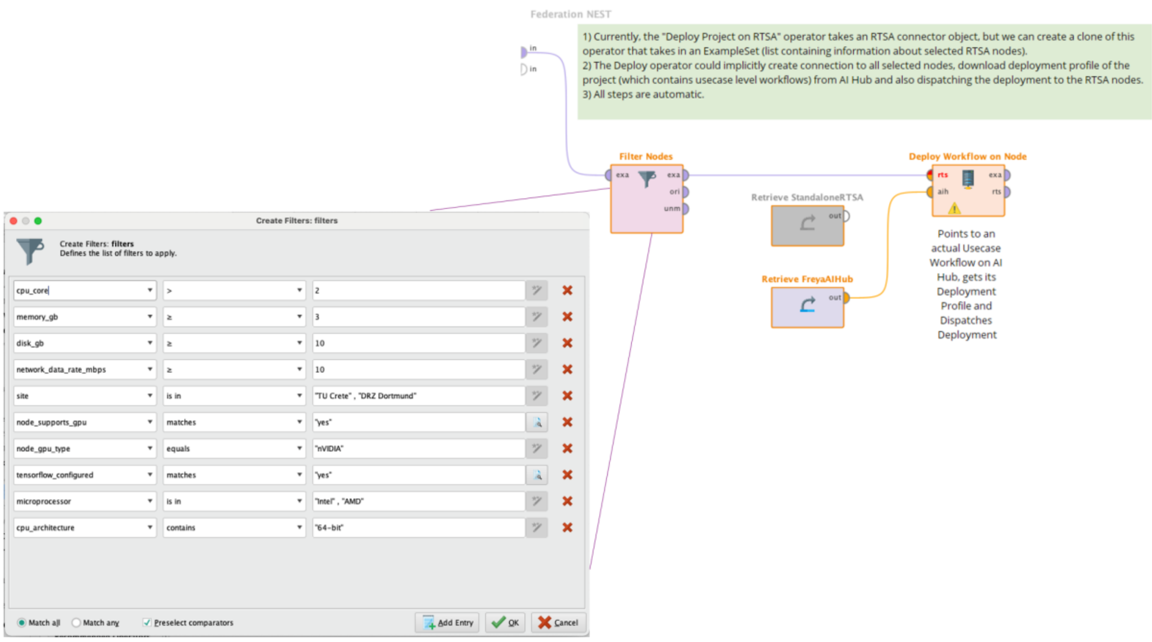

Figure 18: The user can filter nodes of interest and deploy use case workflows on them (i.e. workflows for the respective application use case). An example of such a use case workflow is shown in Figure 19.

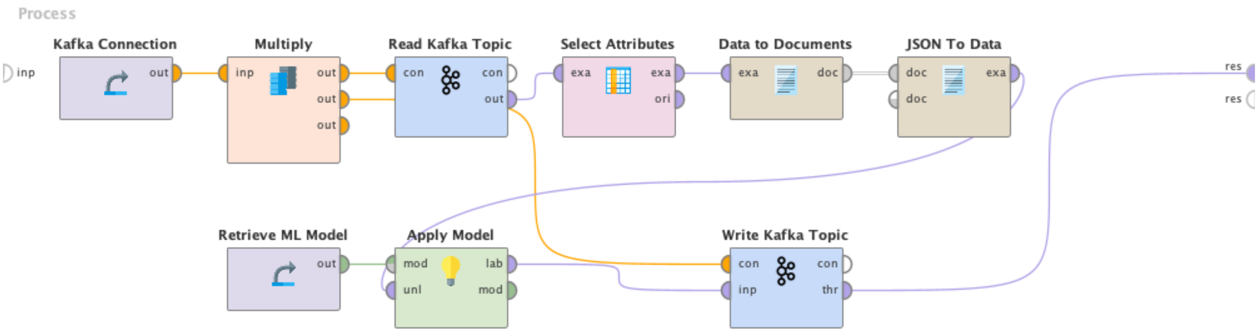

Figure 19: Use case workflow example for executing processes on RTSA node(s). This example workflow reads input data from a Kafka data stream (Kafka topic) using a pre-defined "Kafka Connection" and the "Read Kafka" operator, converts the data from JSON text format to a data format, retrieves a machine learned model (ML Model) and applies this model to the input data read from the Kafka data stream. The prediction result of this ML model is then written to the Kafka output data stream (Kafka topic) using the "Write Kafka" operator. This workflow is an example of a workflow that could be distributed to and executed on different machines and devices (e.g. IoT and edge devices or compute clusters or servers or cloud computers).

2.4 CREXDATA APIs for the System Component Integration

The previous section described the technologies and platforms chosen on which the CREXDATA system will be implemented. This section describes the APIs (Application Programming Interfaces) for the seamless interconnection of all components and their integration in more detail. For the integration, the CREXDATA system uses web services (in JSON format) and Kafka data streams (Kafka topics).

Since the CREXDATA use case scenarios involve continuous data streams, the CREXDATA platform needs to be able to not only handle static data or batch data, but also continuous data streams. Hence, the use-case-specific system components to be integrated into to the CREXDATA platform also need to be stream-enabled. A blueprint for the API of each component is defined leveraging web services for the control flow and Kafka for the data (stream) exchange and for the control flow, using Kafka data streams (Kafka topics) for input data, output data, and supported requests for each component (in JSON format).

For the CREXDATA project, we followed the following approach to map a use case scenario to a data processing workflow:

- Use case scenario description as a sequence of steps involving system components:
  - Scenario tells a story as a sequence of steps.
  - Scenario involves 1 or more system components.
- Component to scenario mapping:
  - Component developers assist in deciding which scenario suits them.
  - Use case leaders help in this mapping exercise.
- Implement scenario as visual RapidMiner workflow:
  - Scenario defines interaction sequence.

- Component API is stream-enabled (i.e. provides a standard mechanism to interface via web services and Kafka), so that it can be integrated into the CREXDATA platform (via web services and Kafka).
- Use-case-specific operators are implemented in RapidMiner extensions.

A CREXDATA reference API was used to stream-enable the use-case-specific system components and to define the web service interfaces (APIs). The reference API leverages web services, which process requests in JSON format, and Kafka data streams (Kafka topics). The following figures and paragraphs explain the structure of the API and its key parameters using a simple calculator as an example for illustration.

Following this CREXDATA system component reference API structure, APIs are defined for all use cases and for all components to be integrated.

The basic structure of the CREXDATA system component reference API describes the possible web service request(s), i.e. the method names (Request Topic) and their signatures, i.e. parameters, inputs (Input Data Topic), and outputs (Output Response Topic). Figure 21 shows this basic structure for a simple calculator web service example for summing up input data values.

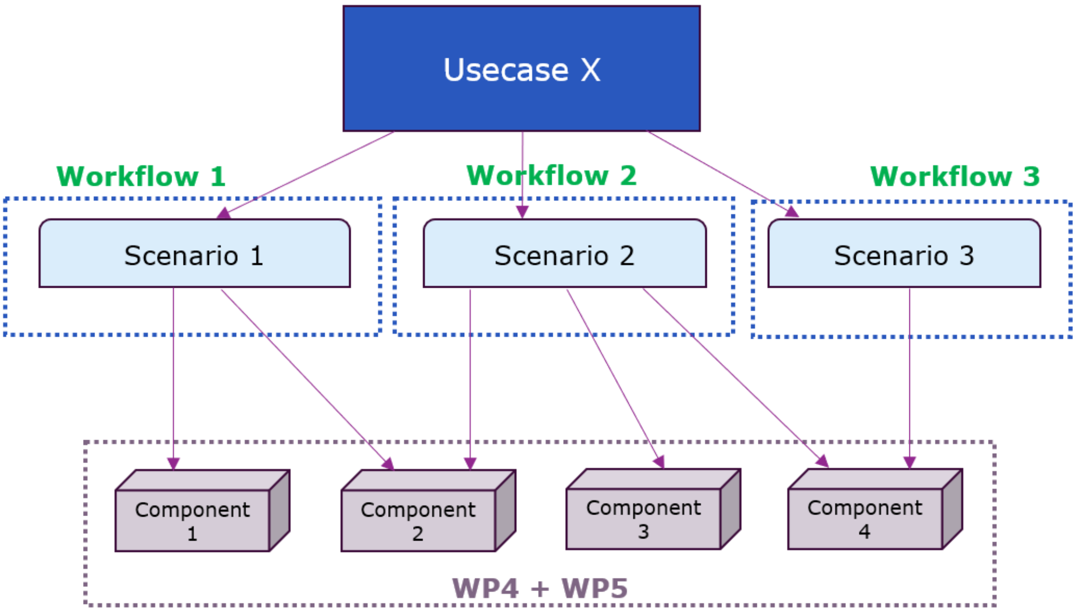

Figure 20: Approach to map use case scenarios to process steps and system components

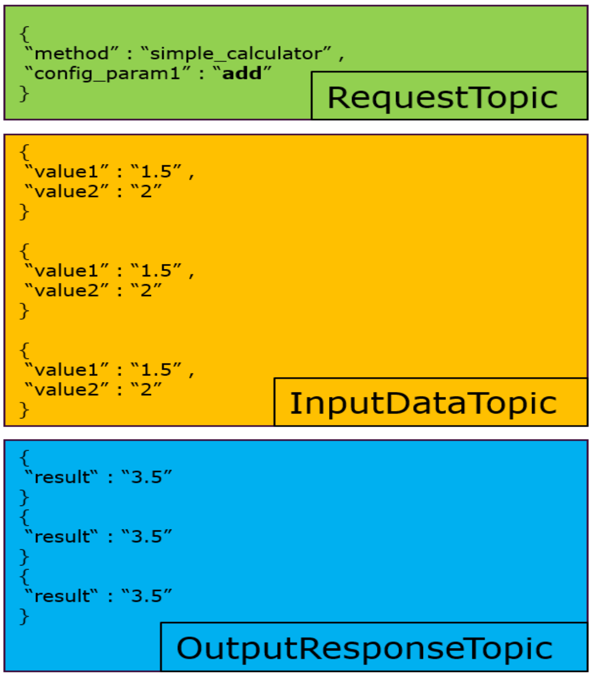

Figure 21: Basic structure of the CREXDATA system component API using web services (with data in JSON format) and Kafka data streams (Kafka Topics), here illustrated with a simple calculator example.

The CREXDATA system component API is implemented via web services (in JSON format) using Kafka data streams (Kafka topics) for web service requests, input data, and output data streams:

- Implementation of a basic reference service (job), here with
  - 2 Input Kafka Topics: RequestTopic, InputDataTopic
  - 1 Output Kafka Topic: OutputResponseTopic
- The service implements an API and exposes its signature:
  - Example Signature:
    - string simple calculator(string methodNameConfig)
      - methodNameConfig : **add, subtract, multiply, divide**
  - Behaviour:
    - request : read request in JSON format from RequestTopic
    - input : read JSON data to be used for this request from InputDataTopic
    - output : write result in JSON format to OutputResponseTopic
- Implementation of a client of this API to test the API method defined above:
  - Write JSON to Kafka topics using a Kafka console script or using the "Write Kafka Topic" operator in the visual process design GUI (RapidMiner AI Studio)

  - Read the method output using a Kafka console script or the “Read Kafka Topic” operator in the visual process design GUI (RapidMiner AI Studio)

In order to allow web services and input data topics with multiple input data streams (e.g. multiple sensors providing one sensor measurement data stream each), the above API is refined with a unique identifier for the data streams (“datasetKey”), so that each input data stream can be identified uniquely.

The refined CREXDATA system component API uses data stream identifiers (i.e. a key called “datasetKey”) to support multiple data streams from different workflows:

- Different workflows write data on the same input data topic.
- To differentiate input data (events) coming in from different workflows:
  - **datasetKey**: add a key called “**datasetKey**” in both, request and input data (JSON payload).
- Behaviour:
  - A request is only processed for data that have matching “datasetKey”.
  - This allows for multiple workflows to use the same backend service, where each workflow submits an add request for a different dataset, i.e. using different dataset keys.

In order to enable data processing workflows not only on single values from data streams, but also on time windows from data streams, the CREXDATA system component API was further refined to support time windows on the input data, either requiring a minimum number of events (“minEvents”) (data points) or specified waiting time (“waitingTime”) to complete a time window and to start processing the web service request:

- Some web services may need to process requests for a batch of data (time window).
- Hence, if less or more data comes in via the input data stream, the input data can be collected in batches (time windows) for a specified time or requiring a specified minimum number of events (data points).
- Collecting input data events by windowing by either requiring a specified minimum number of events per time window (“minEvents”) or covering a “waitingTime” span:
  - minEvents: add a key called “minEvents” to the request to require a specified minimum number of events (data points) per time window (batch of data).
  - waitingTime: add a key called “waitingTime” to the request to collect all events (data points) within the specified waiting time into one time window (batch of data).
- Behaviour:
  - Collect events based on number of events or time passed parameters before computing the web service request (here the request “add”).
  - Since the windowing may block the response from the web service (due to the waiting time until the specified time has passed or until the specified minimum number of events (data points) could be collected from the input data stream), appropriate values should be used to avoid undesired web service blockages.
  - In case of concurrent requests from other workflows, these other requests should be delegated to other task managers.

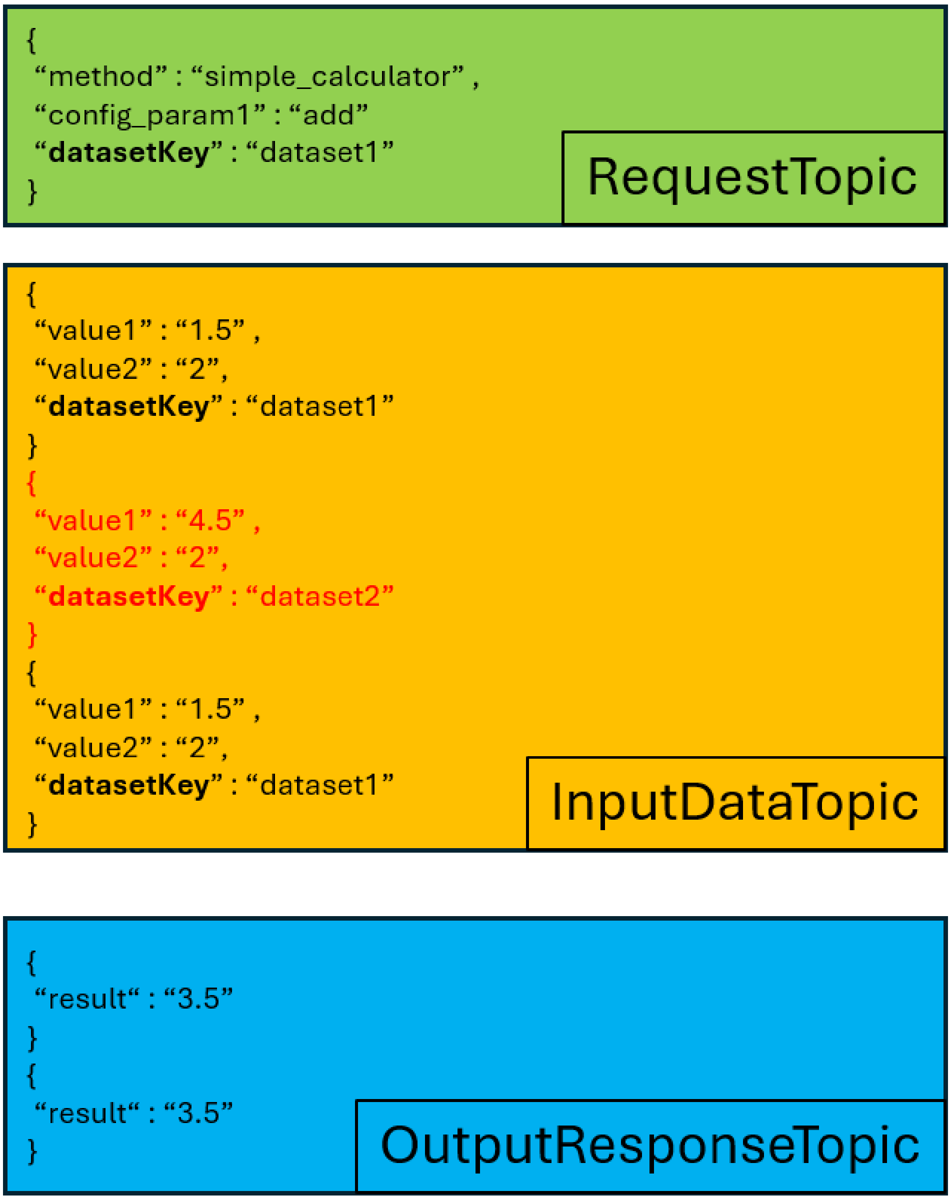

Figure 22: Refined CREXDATA system component API allowing multiple input data streams using data stream identifiers (i.e. a key called "datasetKey"). In this example only input data values from the data stream with the "datasetKey" value "dataset1" are considered for the summation of the simple calculator web service.

Figure 23 shows an example of a summation process for a time window of (at least) three events ("minEvents" : "3"), i.e. waiting until three input data values have been received from the input data stream (InputDataTopic), before computing the sum of these three input data values and writing the sum to the output data stream (OutputResponseTopic):

○

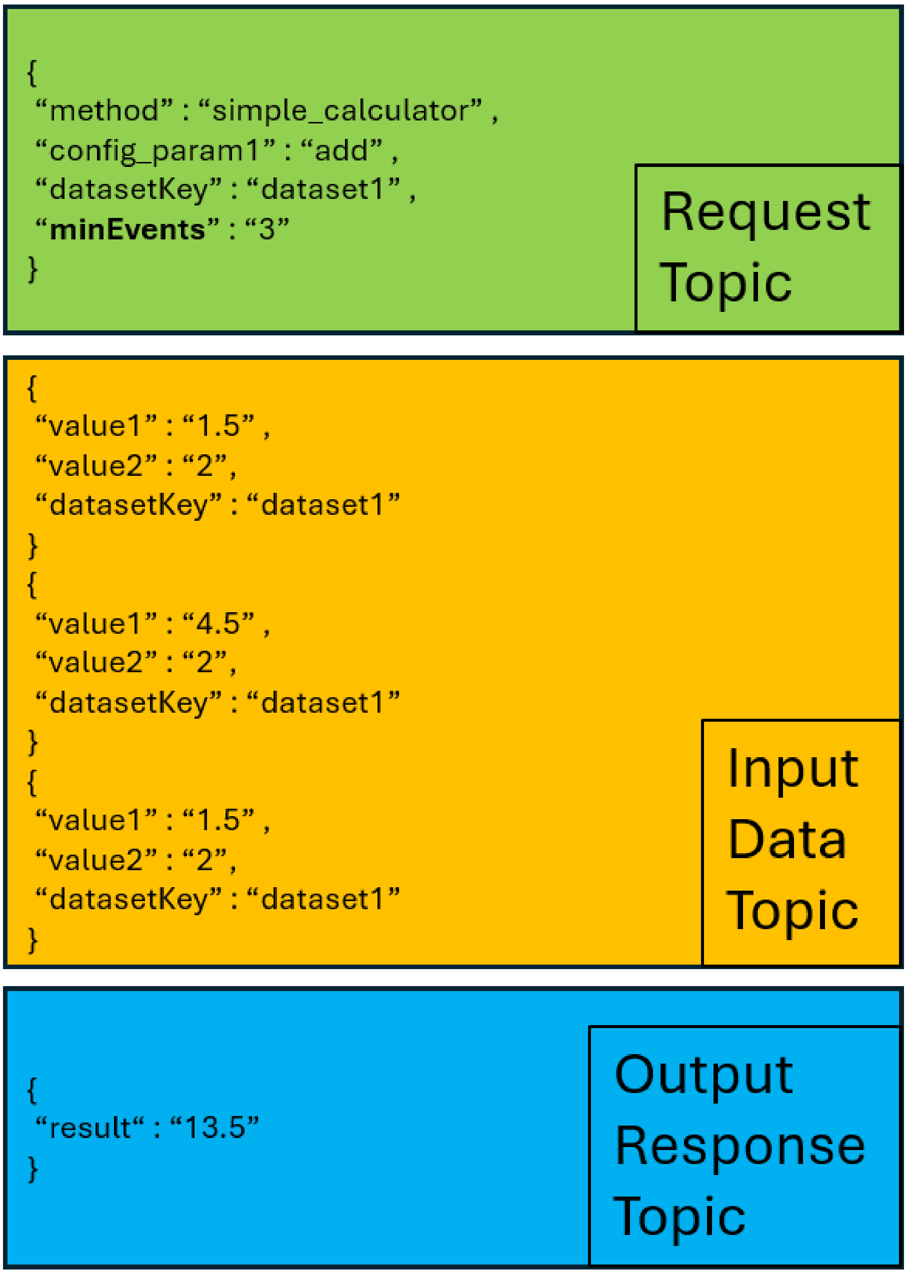

Figure 23: Refined CREXDATA system component API supporting time windows on the input data streams (InputDataTopic), either requiring a minimum number of events ("minEvents") (data points) or specified waiting time ("waitingTime") to complete a time window, to start processing the web service request, and to writing the result to the output data stream (OutputResponseTopic).

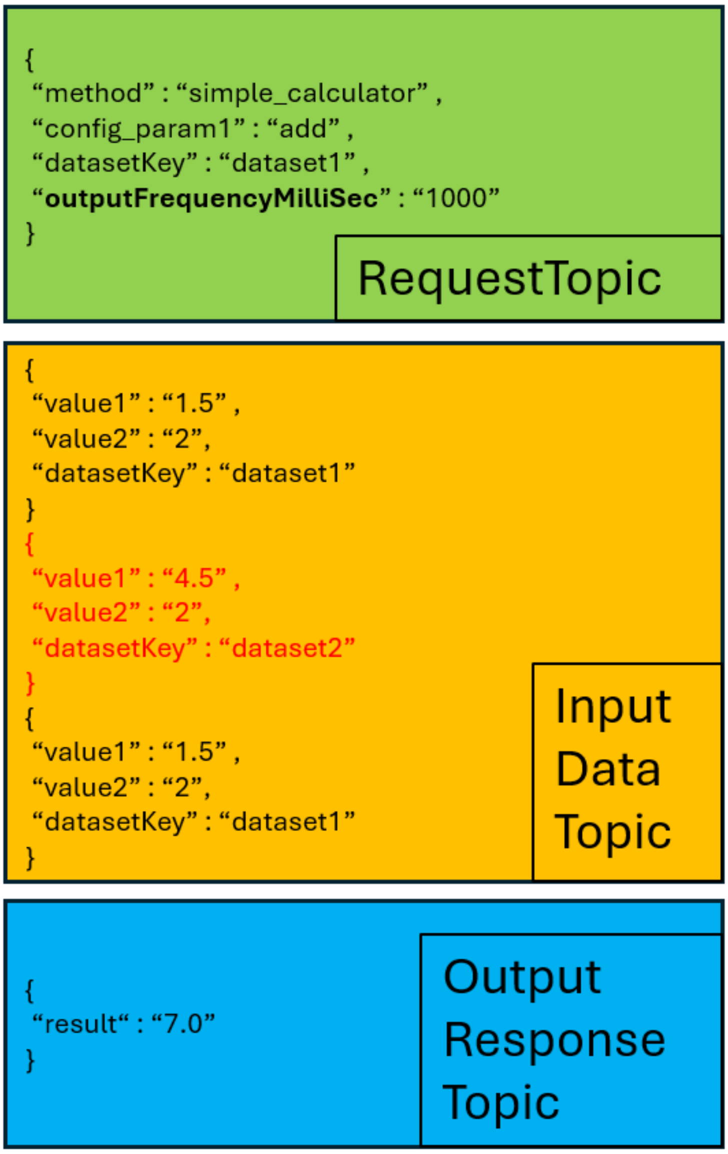

Figure 24: Refined CREXDATA system component API allowing to limit the output frequency of a web service via the parameter "outputFrequency" specified in milliseconds.

In order to avoid that web services produce output data *too frequently*, e.g. too much output data to handle for other components, the output frequency of a web service can be limited, i.e. specifying the frequency of the output generation in milliseconds between two consecutive outputs using the request parameter "outputFrequencyMiliSec":

- A web service may produce output data *too frequently*. In this case:
- Control the frequency of generating output data:
    - **outputFrequency**: add a key called "**outputFrequency**" to limit the frequency with which output (response) events are generated.
- Behaviour:

- A service may expose method configuration on the frequency of its output.

This could be suitable for services which would otherwise generate output data too frequently, especially if the output data is too large.The CREXDATA system component API was further refined to allow several clients (CREXDATA system nodes or web services) to use the same datasets (e.g. input data streams for their process inputs and output data streams for their process outputs (Kafka response topic)), request identifiers ("requestID") can be used to identify which web service output value in the output data stream corresponds to which web service request:

- In cases where different clients (CREXDATA system nodes or web services) use the same input and output dataset(s) and where most configuration parameters for the web service requests are also the same, it is necessary to be able to differentiate which web service outputs belong to which of these different web service requests.
  - Adding a **requestID**: Instead of adding all web service request parameters to the output data stream, just a key called "**requestID**" is added in the request and in the output event(s).
  - This requires each client to do book-keeping of its requestID(s) in a separate file.

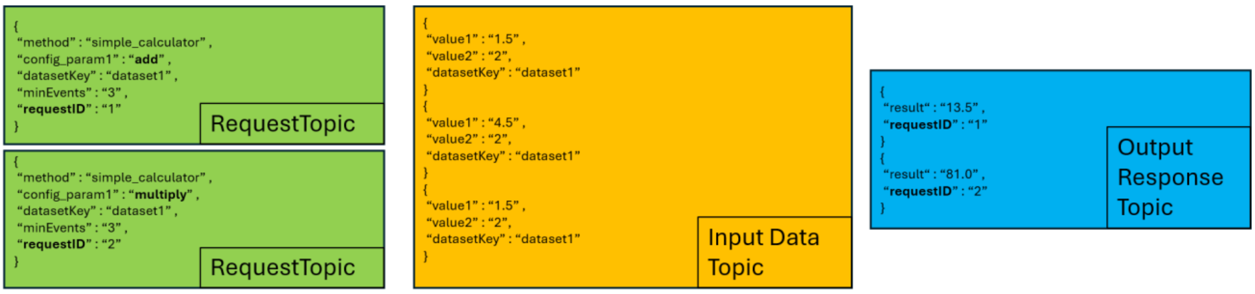

Figure 25: Refined CREXDATA system component API allowing multiple clients to use the same input and output data streams, leveraging request identifiers ("requestID") to specify which web service output value corresponds to which web service request (client).

The data in the web service input data streams and output data streams must be formatted in the following JSON format to the processable by the Altair RapidMiner Data Streaming extension:

Data type and format:

JSON formats are supported where the format of each datapoint is non-hierarchical. JSON objects should look like the following:

```
{key1: value1, key2: value2, key3: value3}
```

Concrete example:

```
"{
    "time": "02/19/2019 06:07:14",
```

```
        "StockID": "ForexALLNoExpiry",
        "price": "110.11",
        "streamID": "7040"
    }"
```

Importing data from the system components / Altair RapidMiner AI Studio into Kafka:

Altair RapidMiner has strong data integration capabilities (Extract Transform Load (ETL). Altair RapidMiner can read in data from a variety of file formats (like CSV, XML, JSON, Excel, text files, PDF documents, etc.), reshape them (header adjustment, etc.) and transform them into the RapidMiner format. Datasets in RapidMiner are called ExampleSet (e.g. training examples for a machine learning algorithm). RapidMiner can then insert an ExampleSet directly into a Kafka topic (data stream) in the supported JSON format (shown above). For each CREXDATA system component to be integrated into the CREXDATA platform (and hence with Altair RapidMiner), a sample of their input data is used to implement an ETL pipeline (automatable data processing workflow) to retrieve data (which can also read files from folders, etc.) and to populate data in a Kafka topic (e.g. output response data stream).

For the CREXDATA project and its use cases, additional data connectors are implemented as RapidMiner extensions to support the required additional data formats (e.g. for images and videos as well as for the use-case-specific simulators and system components).

Note:

- This standardization of the JSON data formats for the integration of the components is necessary for the data streaming extension operators to do their job in a common way.
- Once a component/simulator/etc. receives output from a RapidMiner workflow, it can internally of course arrange the flat JSON into a complex hierarchical format as required internally.

2.5 Domain-Specific Simulators

Simulation models and tools are developed in task T2.4 specifically per Use Case, as they usually represent domain specific phenomena. From a technical perspective, simulators are components which are part of the Demonstrator systems developed per Use Case. They are integrated with the CREXDATA system based on the overall system architecture of WP3, incorporating the research outcomes of T2.1-T2.3. Simulation models and tools developed within CREXDATA will be made available open source.

2.5.1 Simulator for the Weather Emergency Use Case

For the weather emergency case, no integrated solution for the different types of "simulation" is available. By nature, very different influence factors need to be considered in an emergency. In general, a separation is made between own and foreign situation. Resources are deployed (own situation) to cope with natural and man-made events evolving into an emergency to be mitigated (foreign situation). Different approaches of simulation can be recognised, but also very different understandings of the term "simulation". In the CREXDATA Simulation HyperSuite for emergency management, the following elements will be integrated:

- weather related simulation: Typically, in meteorology there is no simulation as such, but nowcasting and forecasting. As a very specific phenomenon, fires depend on a trigger event that starts the fire (see, e. g., the Propagator forest fire simulation model [10]). Similarly, there could be flood simulators modelling the actual flow of water in urban surroundings or even the diffusion of water into buildings.
- weather data archives: Additionally, there are large archives accessible through data services to load satellite images and forecasts from a certain point in time and location in the past like (for instance, flooding events in Germany in 2021). So, instead of simulating a certain weather condition, it might be relevant to just select a weather situation in the past. This complies with typical behaviours of experience-based decision-making in the domain of emergency management.
- robotic simulation models: Physical robots can be extended by kind of digital twins in a virtual environment. By simulating environmental effects in such environments like Gazebo), "what-if" scenarios can be tested in such a safe space. By incorporating parameters representing the actual environment like wind speed relevant for UAVs measuring flood levels, action planning regarding deployment and routing of drones can be supported.
- domain specific simulation models: In requirements elicitation sessions, mainly two types of simulation models were indicated: a) people movements and b) spread of hazardous goods. People movement is relevant both for understanding a situation which is not controlled by Public Protection and Disaster Relief (PPDR) organisations, as well as situations in which evacuation is conducted under control of authorities. Spread of hazardous goods refers to information that is required to plan actions especially regarding spatial parameters. For instance, a contamination might be transferred through the air or, in case of a flooding, through water.
- event injection: There are several approaches in the field of emergency simulation used to train emergency management staff where scenario editors are used to "inject" incident data into simulated or real-time data streams. In case data from the field is missing, it could be injected through such kind of a tool. If this is done, corresponding uncertainty needs to be considered.

For such an integration, a core component is required that provides functionality for the orchestration and configuration of multi-domain services like those mentioned above.

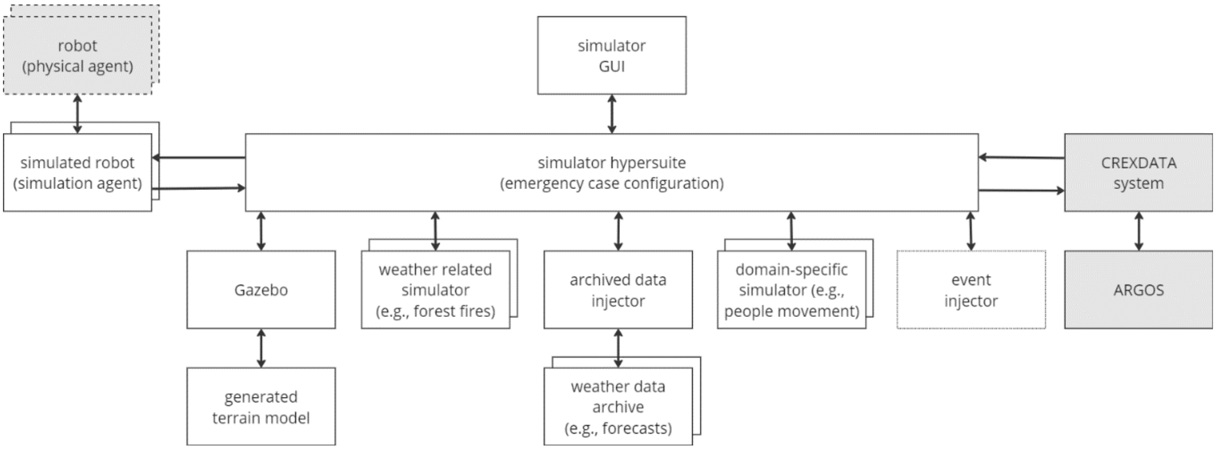

Figure 26: Architecture of the simulator system for the weather emergency case (Source: CREXDATA Deliverable D2.1).

2.5.2 Simulators for the Health Emergency Use Case

The health case is separated into two scenarios, each requiring specific types of simulation models. For the epidemiological scenario, the MMCACovid19 simulator is adopted [11]. For the multiscale lung infection scenario, the Alya [12] and PhysiBoSS simulators are coupled [13, 14].

2.5.2.1 Simulator for the Epidemiology Scenario

The *MMCACovid19* simulator is a flexible package written in Julia language that allows simulating the spread of a disease in a metapopulation with multiple agent types. The package implements a general Susceptible-Exposed-Infected-Recovered (SEIR) model and simulates an epidemic process using the Micro-Markov-Chain-Approach. Moreover, the simulator incorporates a wide range of data, including population demographics, epidemiological parameters, population mobility, healthcare system capacity, and intervention strategies, including confinements, mobility reduction and vaccination. Importantly, MMCACovid19 can be extended beyond COVID-19 to encompass general epidemiology by performing a calibration of the epidemiological parameters. Overall, the MMCACovid19 simulator serves as a powerful tool for decision-makers, public health officials, and researchers, enabling them to simulate, analyse, and optimize strategies in response to health crises like the COVID-19 pandemic. By providing valuable insights, it supports evidence-based decision-making and aids in mitigating the impact of the crisis on the population and healthcare systems.

emews-mmca-covid19 is HPC-based high-throughput model exploration workflow designed for running iterative calibration of epidemiological models and design of optimal interventions based on EMEWS, which provides a sophisticated platform for large-scale model exploration and optimization in HPC infrastructures. By combining it with the MMCACovid19 simulator, users can leverage the capabilities of both tools to enhance the understanding of the spreading patterns of an epidemiological outbreak and inform decision-making processes.

2.5.2.2 Simulator for the Multiscale Lung Infection Scenario

The Lung infection scenario is based on the coupling of two simulators, an organ-level simulator Alya and the cell-level simulator PhysiBoSS. This coupling will allow for a mechanistic multiscale model that will encompass the pulmonary alveoli sacks' structure and cells, the vascular and lymphatic system around these, the airflow that comes in these sacks and the virus infection and cells' interactions.

Technically, PhysiBoSS simulates cell-agents with mechanistic models as well as their interactions and Alya will simulate the air pressure arriving to the lungs and the perfusion of the vascular and lymphatic vessels.

2.5.3 Simulator for the Maritime Use Case

The simulator will present and simulate a high-level overview of the Maritime Use Case test cases for the end user operators. The simulated events will be presented to the end users via a graphical user interface (GUI) for evaluation and testing. In this use case, the simulator will not be open source.

- MT (Kpler) will create streams of simulated vessel positions, events, and weather conditions in order to simulate and optimise the navigation of a vessel under certain conditions (e.g., traffic, weather conditions).
- MT (Kpler) will develop a simulator system for the needs of the Maritime Use Case. The system that will be developed will be in position to use synthetic simulated data and simulate specific scenarios of emergency events.

3 Use Case Specifications, Data Sources, Use-Case-Specific Requirements for the CREXDATA System Architecture, and Use Case Demonstrator Architectures

This section provides an overview of the real-word CREXDATA application use cases, i.e. their initial specifications and resulting requirements for the CREXDATA system architecture and its capability to handle large amounts of multi-modal data of various types from a variety of data sources including extremely large data streams and its capability for federated real-time data processing of extremely large data streams.

This section summarizes the parts of CREXDATA deliverable D2.1 relevant to the understanding of the use cases, their scenarios, and their requirements with regards to the data sources (variety, volumes, frequencies, etc.), data processing capabilities (multi-modal data integration, real-time data stream processing, federated data processing, federated machine learning, etc.), the overall CREXDATA system architecture, the use case demonstrator architectures, as well as use-case-specific simulators, systems, and system components. For more details on these topics see CREXDATA Deliverable D2.1. If you have already read CREXDATA deliverable D2.1, you can skip this section.

Emergency management and critical action planning call for timely and accurate decision making in several, diverse applications with the goal to optimize economic, societal, or environmental impacts. In the maritime domain, critical situations may occur due to imminent harsh weather conditions, vessel collisions, groundings, piracy events and a multitude of other hazardous situations at sea. Civil protection authorities face emergency situations of vegetation fire outbursts or sudden floods due to rapid weather-induced events as direct effects of climate change. Critical action planning is also of great importance in the life sciences domain. The recent world-wide health crisis of the COVID-19 pandemic demonstrated the need for governments and health agencies to make timely decisions to mitigate the impact (a) at a macroscopic level for the evolution of the COVID-19 pandemic at various levels of spatiotemporal resolutions, and (b) at a microscopic level, studying viral evolution for forecasting emerging mutations of clinical relevance [DoA, part B, page 3].

3.1 Weather Emergency Use Case

Weather induced emergencies are characterized by underlying weather phenomena, their evolution in time and space as well as their impact on the environment including people, nature and infrastructure. Large-scale data services are used for data logs, current satellite images and forecasts (like Copernicus services EFAS, EFFIS and EDO). Stationary sensor systems (like weather radars, automatic weather stations and river gauges) are used to verify satellite data, to trigger alerts in cases of critical values and to enable nowcasts. For similar purposes, multi-lingual text messages from social network sites are gathered and interpreted. Weather-related impact databases are only partially available as a source to train machine learning algorithms. Typically, weather information needs to be scaled from high-resolution to regional and local settings. Mobile robotic sensor platforms (Unmanned Ground and Aerial Vehicles/UGVs and UAVs) with their local viewpoint can complement these data sources with different types of cameras, laser scanners, radar and other sensors. The hydrological impact is specifically challenging with respect to specific terrain like in Austria and urban environment like in Dortmund; fire is based on environmental conditions but dependent on trigger/cause of fire. Operations responding to large-scale emergencies are coordinated in (mobile) control rooms with the need of situational awareness, even though they are not fighting the emergency face-to-face. They perform action planning and take decisions based on command posts closer to the site, and operational forces acting at a scene and being responsible for data collection (assisted by robots). Continuous monitoring is required, based on regular data updates or specifically dispatched reconnaissance robots. This can be complemented by human-generated content provided in terms of multi-lingual text messages, photos and videos via social media. Therefore, high data volumes of very different spatial and temporal resolutions and types need to be analyzed, while robotic platforms collect large amounts of data in high frequency [DoA, part B, pp. 10-11].

The CREXDATA Use Case will actually deploy different types of mobile robotic sensor systems, utilize visualization concepts (including collaborative AR) from WP5 and implement them into the extended ARGOS system (provided to the control center staff) as well as the extended robotics situational awareness system (in the command car “RobLW”) and T5.4 on-site AR tools (both provided to commanders in the field) [DoA, p.8].

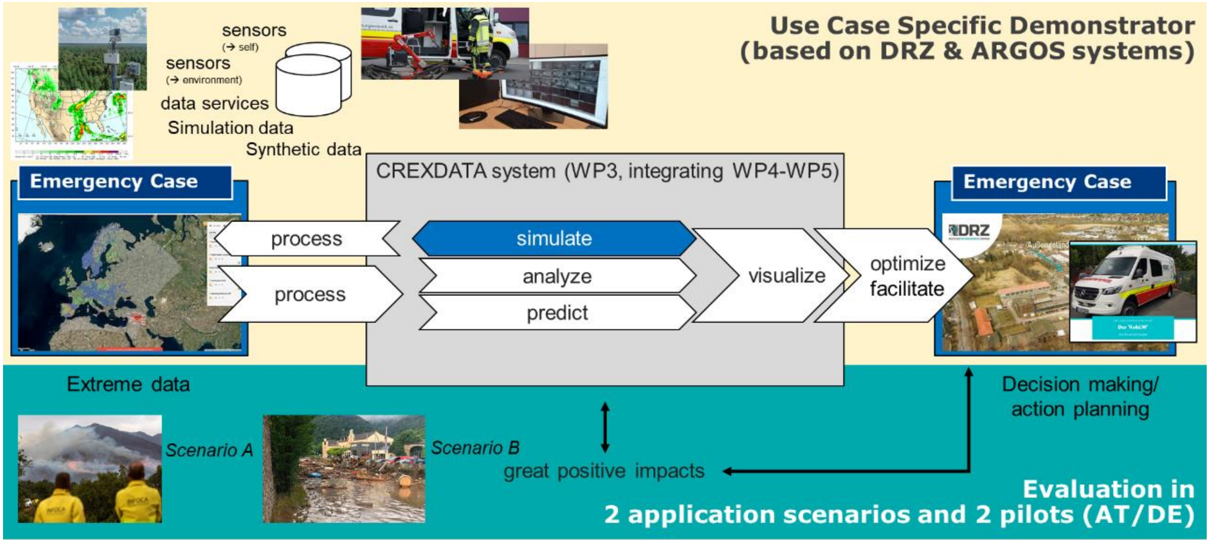

Figure 27: Impact oriented research within CREXDATA in the emergency use case (Figure from CREXDATA Deliverable D2.1).

The evaluation will be performed in two kinds of settings (see Figure 28):

a) at the German Rescue Robotics Center, the existing indoor and outdoor test bed will be used. For this test bed, exact terrain information, building information models etc. are available for reproducible evaluation settings. Both UAVs and UGVs can be operated within the area.
b) two pilots will be setup to evaluate the system in relevant environments in an urban area (Dortmund, DE) and an alpine landscape (Innsbruck, AT).

So, in total four field trials are scheduled. Different levels of decision-making from local fire brigades (FDDO) through technical relief units (DCNA) to national stakeholders with links to the EU Civil Protection Mechanism (MoIFI) are considered. They take decisions with different functional, spatial and temporal responsibility [DoA, p.8].

The purpose of this use case is to **improve situational awareness** significantly so that **informed decisions** are taken by civil protection **considering ranked future worlds** with **explicit uncertainties** avoiding disaster **impacts**. Use case validation is performed in reproducible test bed scenarios and in 4 field trials [DoA, part B, p. 11].

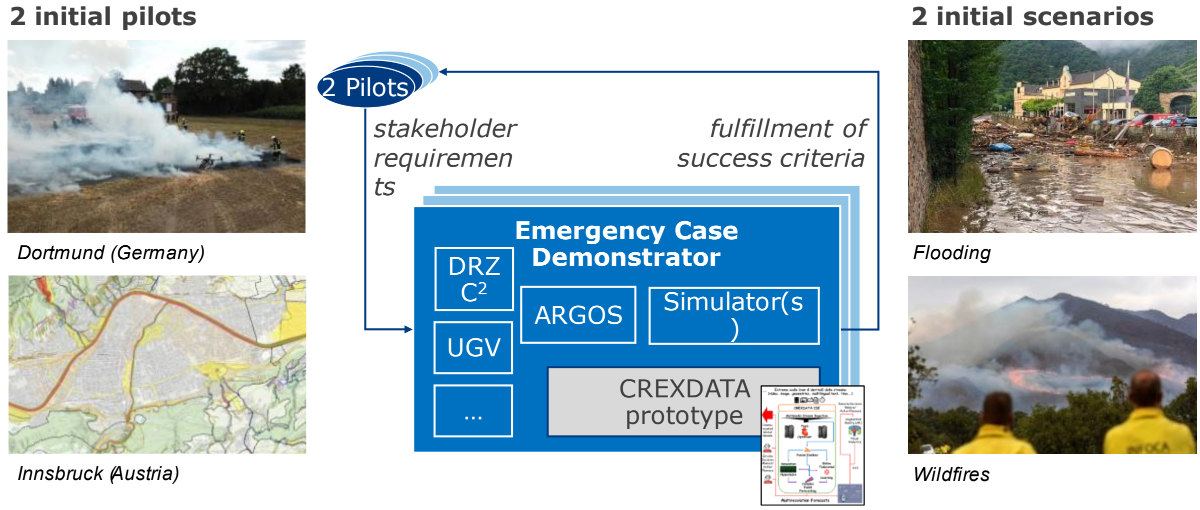

Figure 28: Demonstrator, pilot sites and application scenarios for the emergency case (Figure from CREXDATA Deliverable D2.1).

3.1.1 General Considerations on the Use Case Scenarios

In the first period of the project towards the first series of field trials, application scenarios are focused on hydrological weather phenomena, namely fluvial floodings. By aligning pilot sites to a common scenario setting, configurations of the system and the use of data sets can be managed much more consistently. Data sets from Dortmund, Austria and Finland can be used focusing on similar types of events to be forecasted, features to be learned and explained as well as simulations to be guided. At the same time, domain specific models will be setup in a way that the inclusion of further application scenarios like forest fires is well prepared. For instance, event types like "Vulnerable people endangered" remain relevant while data sources and the underlying simple event type model need to be extended. Additionally, stakeholders and their demonstrator contributions remain the same.

3.1.1.1 Flooding

Flooding is observed from indications of heavy rainfall to monitoring of affected points of interests. Due to heavy rainfall, the gauge of a river might rise or drainage systems might be overloaded. In general, this can be anticipated based on satellite images and forecasts. In case of fluvial flooding, the actual river gauge at a certain point in time at a very certain location needs verification. The impact of a rising gauge may depend on many influencing factors: terrain, bridges or walls channelling the water, congested tubes changing the prepared flow of water, materials in the river (mud, wooden material, debris, cars). Therefore, action planning depends on current data acquired by stationary sensors (e.g., river gauge sensors) or mobile equipment that is specifically dispatched (e.g., drones recording material at the water surface). Typically, sensor data itself is (a) too voluminous and frequent to be analyzed by human operators and (b) too specific/technical to be helpful to operational forces [DoA, part B, p. 11].

3.1.1.2 Forest- and Wildfires

Forest fires are observed from critical drought situation (based on EDO) through fire spreading (with fixed observation systems) to monitoring of the fire zone (with mobile robots). To effectively fight large fires, it is necessary to detect them at an early stage to be able to immediately initiate suppression measures. The next step is to get qualified personnel and resources to the right place in the shortest possible time. For this purpose, continuous reconnaissance of the fire spread is the basis for an effective extinguishing operation. Prediction and quick response to situational events are crucial capabilities. This network can only function effectively if the emergency forces and command can rely on an optimal IT-supported information supply with prognostics [DoA, part B, p. 11].

3.1.2 System Architecture for Demonstrator

The CREXDATA demonstrator system architecture can be read in terms of rows and columns (Figure 29). In rows, a layered IT system architecture goes from (Graphical) User Interface ((G)UI, top layer) to data sources (bottom layer). The CREXDATA system is integrated as a middle layer, interfacing different types of IT sub-systems to be found in columns. These columns indicate fully functional systems. Such sub-systems of the demonstrator system aremainly domain-specific systems with interfaces to the CREXDATA system (cf. Section 2.1). The composed system can be viewed as a common information space [16]. Key components are ARGOS provided by HYDS, robotics provided by DRZ and AR system developed in WP5.

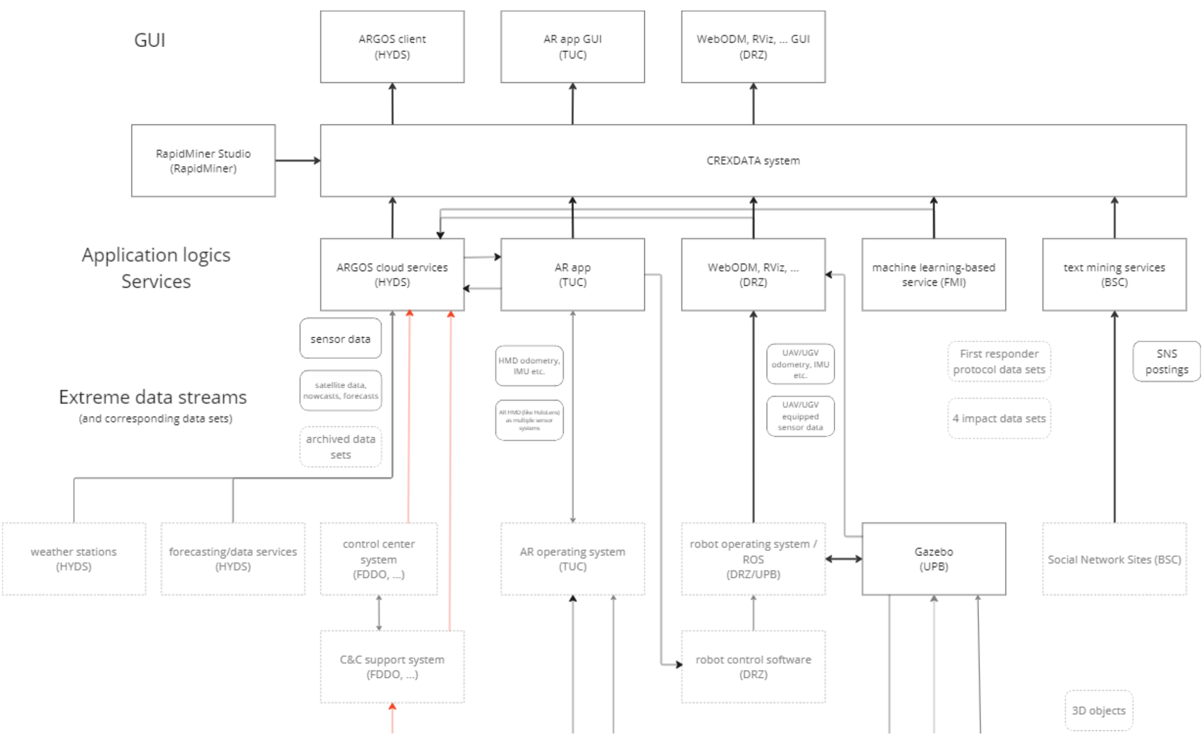

Figure 29: Demonstrator System Architecture for the emergency use case: core system interlinked with application logics and data sources of ARGOS, AR, robotic platforms and FMI services feeding corresponding user interfaces (CREXDATA Deliverable D2.1).

At the UI layer, the system's core GUI is adopted from ARGOS. The initial view as an entry point is assumed to be a geo-based situational map with layers and links to special applications. Thus, geo-referenced data can be super-imposed or marked within layers on top of geographical maps. Additional views need to be available, like green-to-red scale listing of critical events, sortable by priorities, time stamps etc. Besides the ARGOS UI, special applications are required either as extensions to resp. configurations of existing software or as newly developed apps. For visualizing sensor data from robotic systems, solutions like WebODM and RViz/RVizWeb should be adopted. With regards to AR, apps need to be developed for operational staff using development environments like Unity3D. It is essential to include Head-Mounted Displays (HMDs) as See-Through devices on-site. Interaction is implemented, for instance, to perform queries to the system or visualize data processed by CREXDATA algorithms. Application logics are not limited to CREXDATA algorithms but benefit from existing solutions. This includes information processing for layer-based visualisation, to be prepared for an uptake of visual analytics and uncertainty visualizations from WP5. To enable modelling of entire data processing pipelines, not only WP4 functionality needs to be implemented in terms of RapidMiner operators. Similarly, operators need to be made available to connect, for instance, to weather-related services. With regards to robotics, functionality should support analysing sensor data from multi-sensor systems, routing, etc.

Table 1: Demonstrator components extending the CREXDATA system (Weather Emergency Use Case / Dortmund) (Table from CREXDATA Deliverable D2.1).

Component	Description
ARGOS [DoA, part B, p.9]	• multi-hazard early warning system developed by HYDS • data fusion • mapping of geographical data sources (as layers in the geo-service) • within mathematical models for specific topics (like height of snow on streets based on precipitation and ground temperature gradients) • ROS framework is applied using custom scripts to incorporate sensor data from robotic platforms
DRZ	• RobLW ("DRZ C2"/"Lagebildsystem"): mobile robot control and mission command post, equipped for robot operation as well as collecting data from robots and data processing to support situation awareness and decision making • Ground Robots (UGV): various platforms with variable payload modules, selected depending on the mission (application and tasks); e.g., mid-size Telemax equipped for navigation and mapping • Aerial Robots (UAV): various commercial platforms, typically equipped with cameras; additional payload depends on the mission
FMI	• gradient boosting machine learning approach, used in various impact forecasting products and in previous research projects like SILVA
UPB	• Gazebo • RViz • event injection (cf. training support systems) • simulation configuration • data fusion

3.1.2.1 ARGOS system

ARGOS incorporates all processes required to manage weather-induced hazards, harmonising data, products, warnings, impact and protocols in one integrated solution. Core functionality is highlighted as hydrometeorological forecast, early-warning detection, exposure and vulnerability, impact forecasts, management protocols and dissemination. ARGOS has been designed from ground up to seamlessly integrate any source of information useful for operative management. It supports authorities in defining new rules of warning decision flows. Integrated data services subsume

a) services that are activated in case of a large emergency
b) general services: EFAS, EFFIS, EDO, etc.
c) sensor systems

Besides meteorological data, data from 112 calls, traffic cameras or social networks can be integrated (cf. Figure 30).

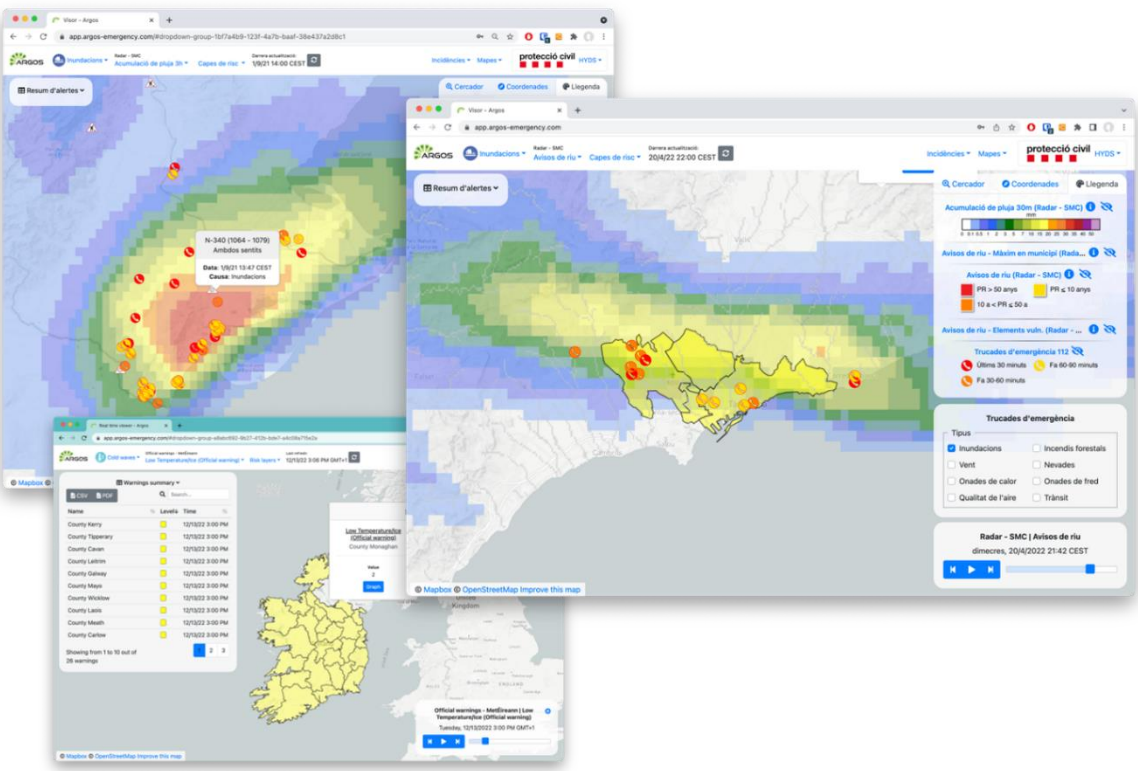

Figure 30: ARGOS service examples: radar data, forecasts, threshold visualization, geo-located emergency calls, warnings summaries, playback function etc. (Figure from CREXDATA Deliverable D2.1).

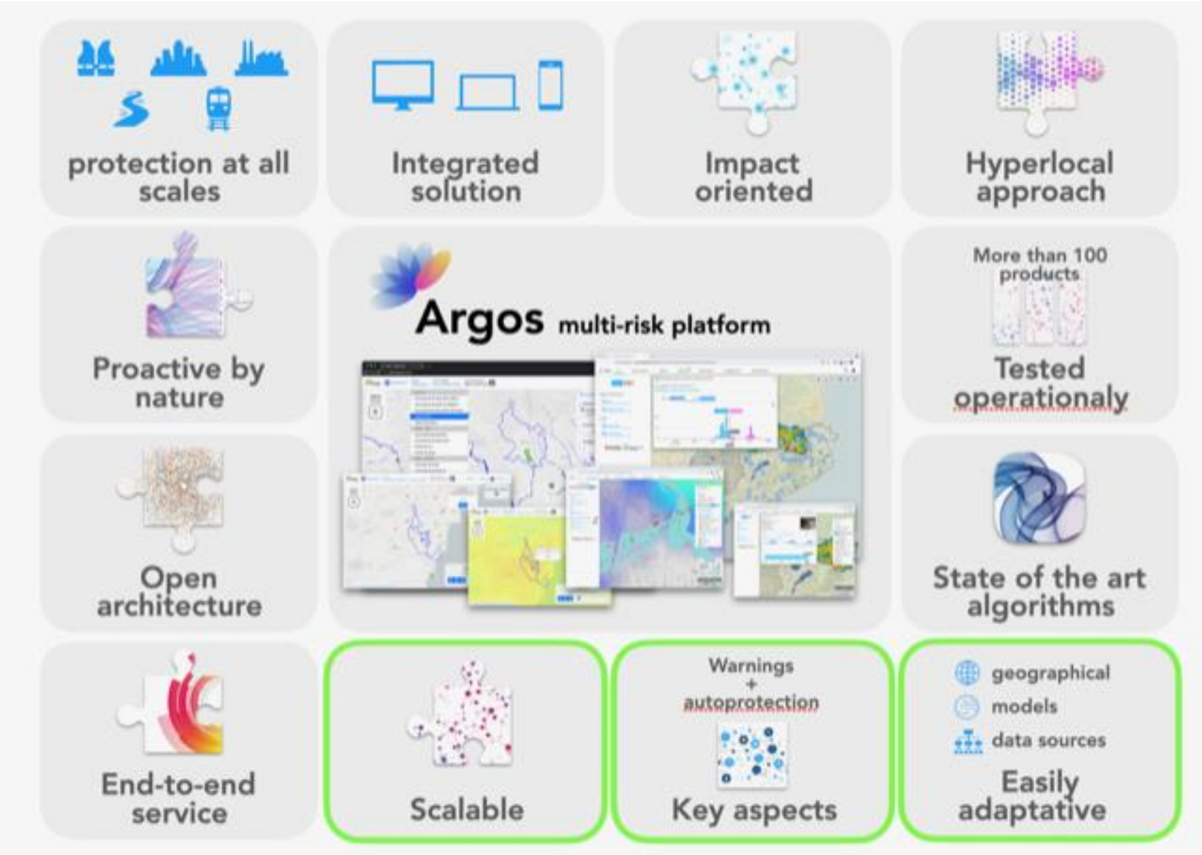

Figure 31: ARGOS features (https://www.hyds.es/argos/). (Figure from CREXDATA Deliverable D2.1).

ARGOS is built on top of an open architecture (see Figure 31) based on self-dependent modules, so that adding new weather data and sensors, building new products and warnings, growing sets of critical points, redefining protocols and expanding dissemination channels is supported. Due to its modular structure, communication with external services or platforms is inherently flexible in ARGOS. Real-time data is collected through available machine-to-machine interfaces, web map services (WMS) for geo-structured data or raw file transfer through sFTP servers. On the other side, Argos generated data (warnings, related products, registered values, etc.) can also be pulled by other systems using its own API (Application Programming Interface, available in https://api.argos-city.com/). Communication interfaces will be adapted or extended to integrate ARGOS with the overall CREXDATA system.

The cloud-based ARGOS architecture enables collaboration between different operational authorities in width (e.g., between federal districts) and depth (i.e., on different levels of organisational responsibility). ARGOS is designed as a family of products to support different types of stakeholders (Argos City, Argos Site, Argos Flow, Argos Hydro). To do so, the following characteristics were enhanced from the very beginning on its construction (see all features in Figure 31):

- Easy adaptation: on new places, on new organisations with similar needs but different procedures.
- Impact oriented: not only weather monitoring but automatic activation in vulnerable elements based on previous knowledge
- Integrated solution: Based on the cloud, the system is available from any device with internet connection.

3.1.2.2 Robotics system

In CREXDATA, different types of robots are foreseen for experiments[1]. In general, Unmanned Ground Vehicles (UGVs) and Unmanned Aerial Vehicles (UAVs) are differentiated. In case of research setups, main building blocks of such robotic systems are basic platforms, communication hardware and payload like sensor systems or, for instance, robotic arms/manipulators (Figure 32). The Robot Operating System (ROS) is implemented on the platform, ensuring standard interfaces to, for instance, sensors. Commercial robots (esp. UAVs) are integrated systems not necessarily changeable. Robots are operated using a robot operator interface (base station). In the CREXDATA setup based on DRZ capacities, robots are deployed by operational staff, with a perspective to expert units entitled “Robotic Task Force” (RTF). Currently, the DRZ operates a special command vehicle called “RobLW” (robot command verhicle)[2] which is equipped with both command and research tools. It is a prototype for a vehicle with robot integration, combined with a joint utilisation concept between DRZ and Dortmund Fire Brigade. Therefore, it is operated by mixed crews from DRZ/project and fire brigade for certain scenarios. Within the RobLW, special information systems are used to process robotic sensor system data. They are subsumed under the term “DRZ Command & Control (C2) system” (German: “Lagebildsystem”). Two specific software products are WebODM and RViz. WebODM enables visualization of maps, point clouds, DEMs and 3D models from aerial images. RViz/RVizWeb provides functionality for 3D visualization of sensor data, robot model, and other 3D data in a combined view.

[1] https://rettungsrobotik.de/en/testing-facility/the-robotic-systems-on-an-overview

[2] https://rettungsrobotik.de/en/testing-facility/the-robotic-command-vehicle

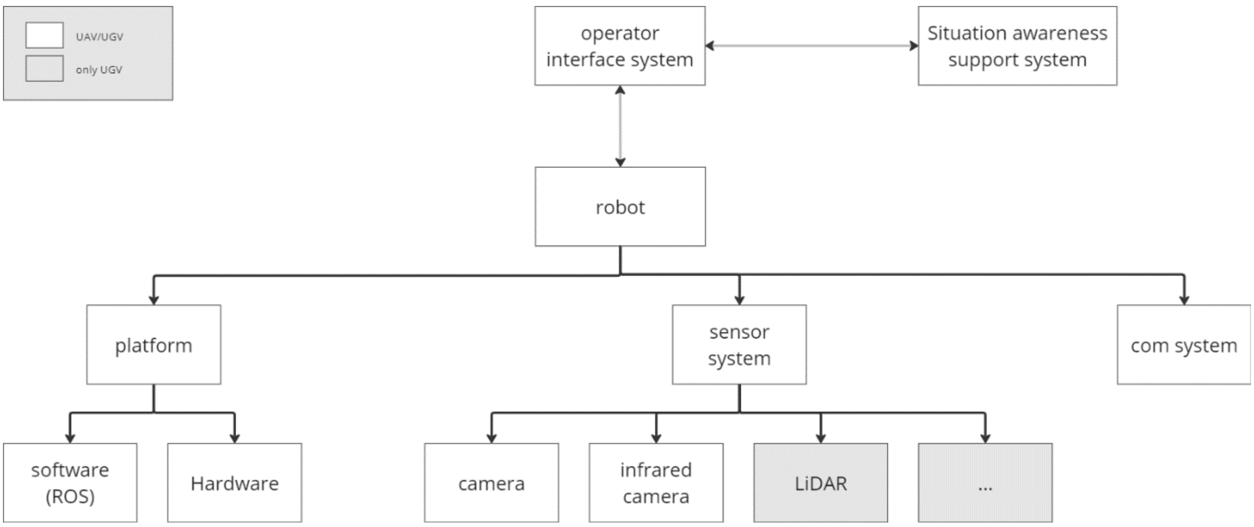

Figure 32: Architecture of the robotic demonstrator sub-system (Figure from CREXDATA Deliverable D2.1).

The RobLW is equipped with a powerful server for the creation of 3D models, two operator PCs and two 43'' screens. For power supply and communication network setup, a 4 kW power generator, blue light/ Tetra radio, 2.4 and 5.8 GHZ receiver for drones, WLAN, 2 TETRA MRT and 4 TETRA HRT are available. A small workbench is built in with two fixed workstations plus an additional workstation for researcher/leader. Additionally, a flexible roof panel enables easy integration of various antennas.

As an option, even under-water robots could be used in CREXDATA (see, for instance, requirements in the Austrian pilot site). In a parallel project, DRZ builds a test setting for such kind of robots. So this would enable use cases like:

- detection of physical or chemical contamination in the water that floods an urban area,
- selection of sensors to see under water (from in water or above water),
- 3D mapping and/or object detection under water,
- creation of realistic dataset for 3D mapping and object detection under water.

3.1.2.3 FMI service

FMI develops impact-based forecasts which are derived from the available impact data. The tools are developed in close cooperation between FMI, MoIFI and the Rescue Department of Helsinki, and possibly other pilot site partners like FDDO outside of Finland. There will be a testing period during which the new products are piloted in real-time and the user experiences are collected to develop and enhance the machine learning-based model.

The model behind the tools developed by FMI for the Finnish showcase are utilizing machine learning (gradient boosting method). The Finnish showcase will be utilizing the ARGOS system in data and model output distribution as well as for utilizing the data provided by other pilots to extend the area of operation of the ML-based service of FMI. The same CREXDATA platform as the Germany and Austria pilots will be used in Finland as well. Eventually, the data produced and shared by FMI is available for the use of the technologies of WP3-5, for instance for input for explainable AI (XAI) of T5.1. Within the limits of available data, the gradient boosting machine learning model is used as a base to forecast the number of ambulance units needed in both Helsinki and Dortmund areas in a weather emergency or to create early warning tools for flooding events. The gradient boosting machine learning method is used to model the impact data, for example the number of emergency operations or road traffic accidents. The method graph is presented in the Figure 29.

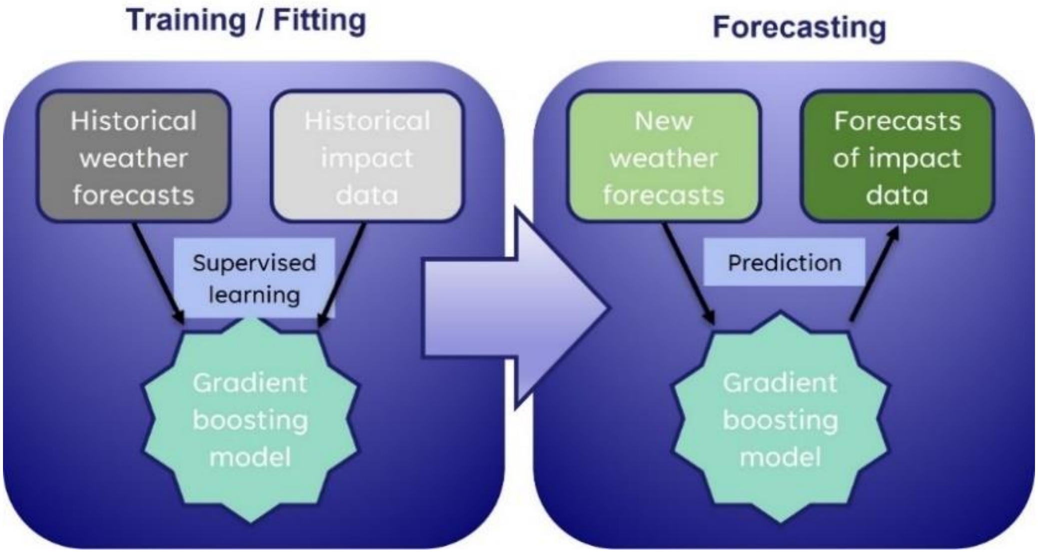

Figure 33: Use of gradient boosting machine learning in training and forecasting of weather impacts (Figure from CREXDATA Deliverable D2.1).

The ML method has been used in various impact forecasting products successfully, mainly in Finland, but also in the USA to model river streamflows. The method has been used also in previous research projects, for instance in a national level SILVA-project (2020-2023), where national level impact products were created (Figure 34). The model has been found reliable and accurate in several different applications; thus, it was chosen also for the main method in the Finnish showcase and as an input for T5.1.

For the modelling, the materials are collected spatially and temporally: local compilation and six-hour temporal compilation are used to unify the number of cases in Uusimaa and Helsinki areas. For the time series, a gradient boosting model is fitted using mainly surface weather parameters from the ECMWF HRES weather model from the Uusimaa and Helsinki area to explain the impact data. As additional information, information about the time of day and the season of each moment of time is included into the model. After matching, the provincial quantile limit values were derived for each data and month to illustrate the number of cases: the familiar green-yellow-orange-red colour coding is used in the visualization to distinguish the classifications determined by the quantile limit values from each other. The visualization is done by compiling the forecasts into five-day forecast maps by using rolling 24-hour sums.

In the CREXDATA project, new forecasts of new datasets will be implemented, such as the numbers of ambulance operations mentioned before. In addition to that, we will refine the resolution of the old products from the county level to the municipality level. The possibility to include estimation on uncertainty of the impact forecast is exploded by utilizing instead of ECMWF HRES numerical weather prediction model, the limited and higher resolution ensemble model, MEPS. For the European context the possibilities to model the river streamflow data for early warning of the flooding events are explored.

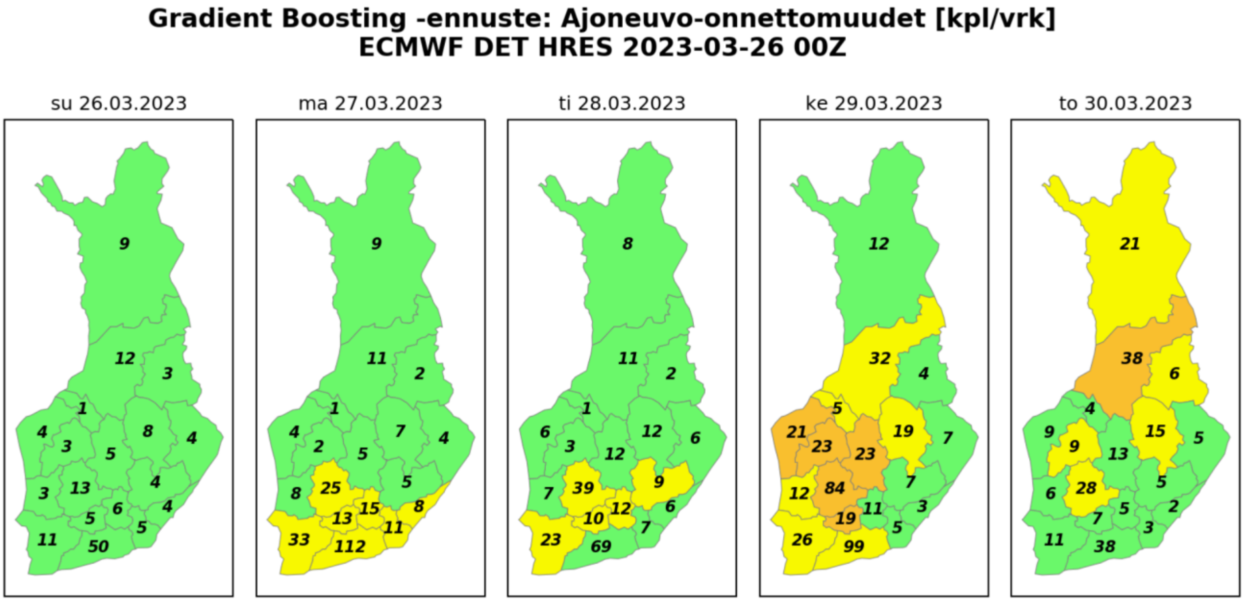

Figure 34: A 5-day outlook of gradient boosting forecast of car accidents per day per municipality. The colors (green, yellow, amber, red) represent the severity of the impact (Figure from CREXDATA Deliverable D2.1).

3.1.3 Data sources

The Weather Emergencies use case is based on data sources that can be categorized into local and global environment (see Figure 27). Data about the local environment is gathered through stationary sensors systems which are extended by mobile robotic sensors platforms in case of an incident. Local data requires context metadata like position, orientation and accuracy of sensors. Temporal effects are induced by sample rate (of sensors), pre-processing at the edge (e.g., creating geo-referenced point clouds) and communication channels (with frequency and bandwidth). Decision-making is informed by processed data formats, but explainability for humans often requires raw data analysis (e.g., video) and visualization. Position, point/field of view and annotations from AR devices are new data sources. At the level of global environment data, Copernicus services provide both TBs of data archives and GBs of current and forecasted weather data covering various natural phenomena and their effects on forest fires, health etc. Both temporal and spatial resolution vary. Sensor systems are increasingly applied in terms of forest observation systems (even with drone extensions), river gauge installations and publicly accessible weather stations. Data is either pushed in certain time intervals so that it can be processed message-based, or data is retrieved on request through data services and APIs. In Austria, for example, there is the ehyd database providing pre-filtered data (https://ehyd.gv.at/). Raw data is accessed through national, European and global weather services. In case of Austria, the corresponding partner would be Geosphere. Multi-lingual and multi-modal social media extend that broad range of data sources. Social media networks are expanded to global scale; for an emergency, the identification of relevant assertions in terms of location and content are significant challenges [DoA, part B, p. 11].

Table 2: Local environment data sources in the Weather Emergency Use Case [DoA, part B, pp.11-12] (Table from CREXDATA Deliverable D2.1).

Data sources	Data Source Volume (Vol), Velocity (Vel), Veracity (Ver), Characteristics/Description (Ch/D)
Data sources (local environment)	
Robot localization system	Ch/D: Position, orientation, velocity and acceleration of a robot, derived by fusing data from various sensors. E.g., odometry is often obtained by fusing data from wheel encoders, IMUs, and gyroscopes. Can be enhanced or substituted by visual odometry (extracting camera movement from consecutive images). Global pose (I.e., relative to a known reference frame) is typically obtained by combining data from GNSS receivers (e.g., GPS), magnetometers, odometry and matching range sensor data (e.g., from a lidar) to a known map. Pose data is transferred using designated ROS message types, e.g., geometry_msgs/Pose3D. Ver: Localization accuracy depends on chosen sensors and algorithms. Vel: Typically, between 10 and 100Hz. Vol: A single pose message is <1kB, overall data rate depends on required update frequency.
RGB video stream per camera (H-264, MPEG-4)	Vol: 1920 x 1080 (full HD) or 720 x 480 (standard definition). Vel: Common video stream data rates (e.g., 30fps). Ch/D: FPV Camera, wide-angle camera. Ver: If the connection is poor, a live stream may be blurred or stop for a short time.
Thermal video stream (H-264, MPEG-4)	Ch/D: Images with temperature data, e.g., derived from near-infrared spectrum. Vol: Common thermal camera models offer resolutions up to 640x480 pixels. Vel: Common video stream data rates. Ver: If the connection is poor, a live stream may be blurred or stop for a short time.
3D environment model	Ch/D: Map generated from robotic sensors. Commonly derived by combining localization data (see above) with range measurements (from cameras or lidars). Some solutions build the map “live” (Simultaneous Localization and Mapping – SLAM). Others build the model offline by processing collected data (most often images) in batch, which is much slower but often more accurate (e.g., using WebODM). Vol: Depending on the size, sparsity and type of the model (e.g., showing only the physical structure, or also including confidence, colors, temperatures, etc.). A model can easily reach several hundred MBs in size. Vel: Typically <10Hz. Ver: Depends on chosen sensors, algorithms and environmental conditions (e.g. ambient lightning, disturbance from smoke).
AR (HoloLens 2) (MPEG4)	Vol: 1920x1080px. Vel: 30fps (through miracast). Ch/D: Video stream, incl. annotations of AR users in the scene.

Table 3: Global environment data sources in the Weather Emergency Use Case [DoA, part B, pp.11-12] (Table from CREXDATA Deliverable D2.1).

Data sources	Data Source Volume (Vol), Velocity (Vel), Veracity (Ver), Characteristics/Description (Ch/D)
Data sources (global environment)	
EFAS	Vol: ~1Gb/day. Vel: Twice per day. Regular internet velocities. Ver: Probabilistic flow forecasts. Ch/D: Temporal series, distributed fields, deterministic and probabilistic
EFFIS	Vol: ~1Gb/day. Vel: Twice per day. Regular internet velocities. Ch/D: Distributed fields.
ECMWF forecasts	Vol: ~20Gb/day. Vel: 4 times /day. Regular internet velocities. Ver: Deterministic and probabilistic forecasts. Ch/D: Distributed fields.
Weather sensors	Vol: ~0.2Gb/day. Vel: ~Every 10 min. Regular internet velocities. Ch/D: Temporal series.
Weather radar	Vol: ~5Gb/day. Vel: ~Every 10 min. Regular internet velocities. Ch/D: Probabilistic data.

3.1.4 Use Case Analysis

The application scenario is setup based on various parameters to allow for a wide variety of sub-scenarios. Examples are fluvial vs. pluvial flooding, seasons, cascade effects (like debris/mudflow), prepared vs. escalating evolution of the situation as well as mainly affecting people vs. objects. Initially, the focus will be on pluvial flooding. So, the assumption is that the strain a heavy rain causes on urban drainage systems is too heavy. Figure 35 provides an overview of the generic scene elements that are scheduled for application scenarios: Each scenario shall include a river that passes a locality with potentially endangered inhabitants. This might be urban or smaller scale. It includes vulnerable buildings and people, as well as critical infrastructure (like energy or transport). In case of an emergency, the emergency management structure is either built up step by step (in case of unforeseeable events) or it is anticipated and well prepared (e. g., in case of precise weather forecasts). Assuming a large-scale weather induced emergency, the structure (see small box in Figure 35) includes the high-level strategic command in a crisis management room (led by the mayor in Germany, supported by a high-level fire officer like the director of a fire department). Search and Rescue, fire protection and technical relief are coordinated by a tactical command unit (mobile Command & Control / C^2 post, operating from a large command truck or bus). Technologies like AR and robots are operated close to the operation by lower-level commanders (C level, in small command cars like vans) and sub-ordinated operators.

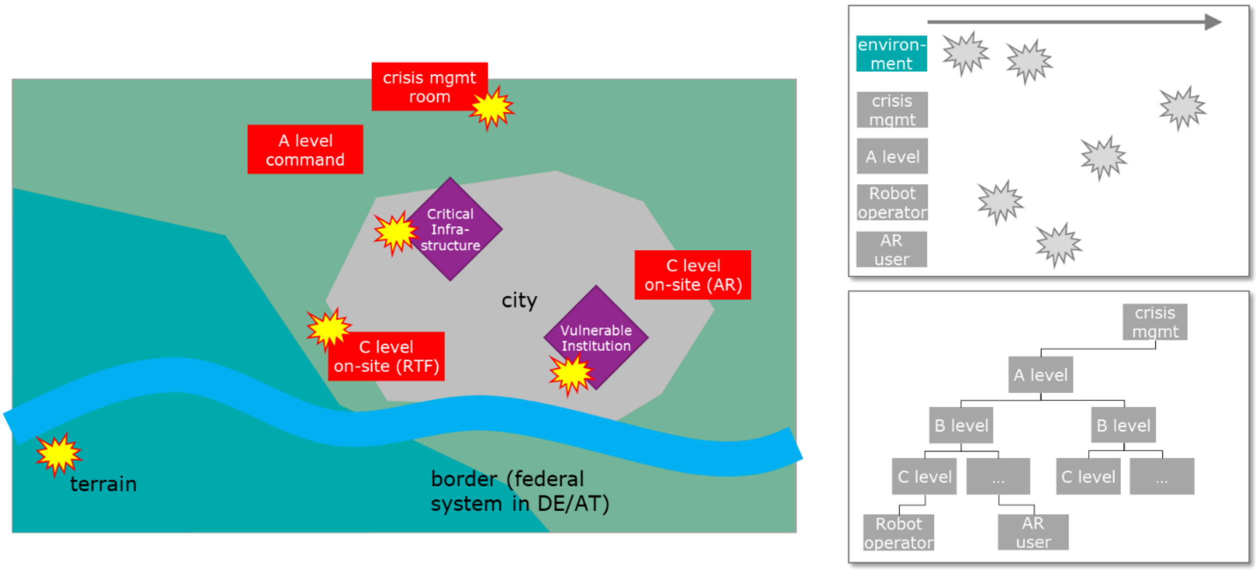

Figure 35: Elements of the spatial environment and basics of a command hierarchy (Figure from CREXDATA Deliverable D2.1).

Figure 35 hints at different views within an emergency situation: the spatial view which is typically represented by geo-services and map-based user interfaces, an event stream view representing changes of the situation, and a structural view focusing on resources deployed by emergency response organizations. Figure 36 provides a more tangible insight into the operational environments that frame use cases of the CREXDATA demonstrator system. AR and robots are operated close to the actual incident by operational responders. They are led by low-level commanders (entitled “C level commanders” at FDDO). Mid-level commanders (B level) are not focused on initially. For the A level command post, a staff room is equipped where four to six or even more fire officers cooperate with regards to pre-defined tasks (like ICT, map, supply, press, etc.). Different settings are feasible: a room setup within a command truck (entitled “ELW3”), or a stationary environment at fire station 1 in Dortmund. For the sake of clarity, the first is called “A level staff room” and the latter is called “crisis management room”. For a large-scale disaster, the mayor of Dortmund activates and leads a crisis management team, being super-ordinated to the structures of FDDO. In all these staff environments, tools like the ARGOS system are relevant. The control center dispatches resources to the operation and tracks status changes (at station, departing to incident, active in an operation etc.) of resources[3].

Figure 37 presents a swim-lane visualization of the evolution of such an event. Focusing on decisions (swim lane in the middle), a distinction is made to events that occur in the environment (e.g., change of weather forecasts, threshold reached at river gauge) and events that are triggered by own actions (e.g., a human is rescued, or a drone reached the operational position).

[3] GPS is not available due to GDPR issues.

Figure 36: Environments for use cases of the CREXDATA weather emergency application scenario (Figure from CREXDATA Deliverable D2.1).

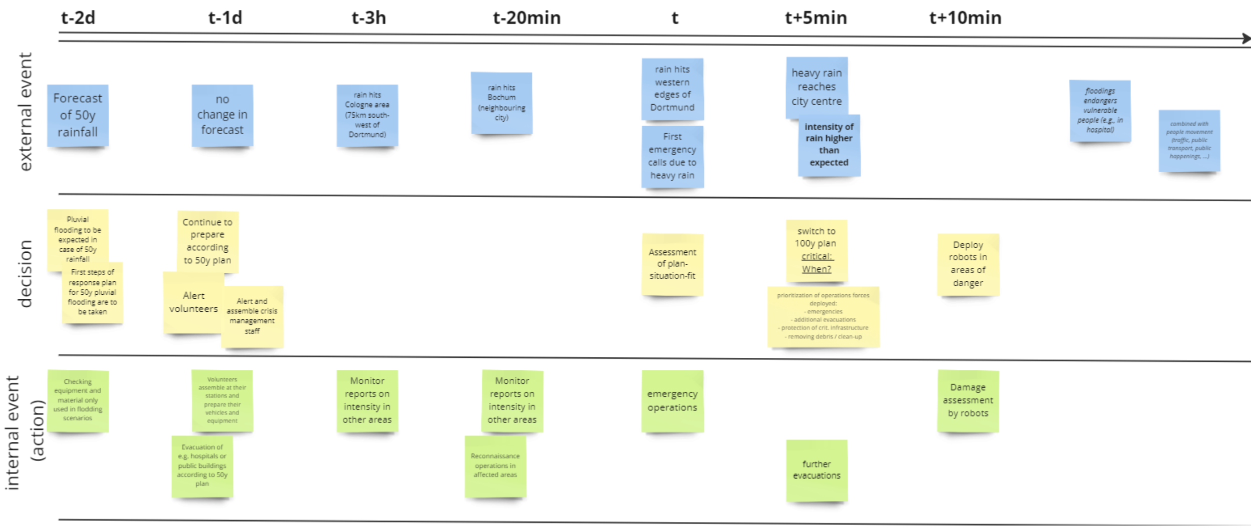

Figure 37: Elements of the temporal evolution of the scenario (Figure from CREXDATA Deliverable D2.1).

Specific sub-scenarios within this overarching application scenario are provided in the appendix of CREXDATA deliverable D2.1. These sub-scenarios are designed in a way that they can be either a) combined with each other for integrated test scenarios or b) used to derive specific test cases. While test scenarios would be relevant for field trials, test cases are required to setup lab-scale test environments (for instance, at DRZ and at UPB). Figure 38 presents an overview of the initial set of sub-scenarios.

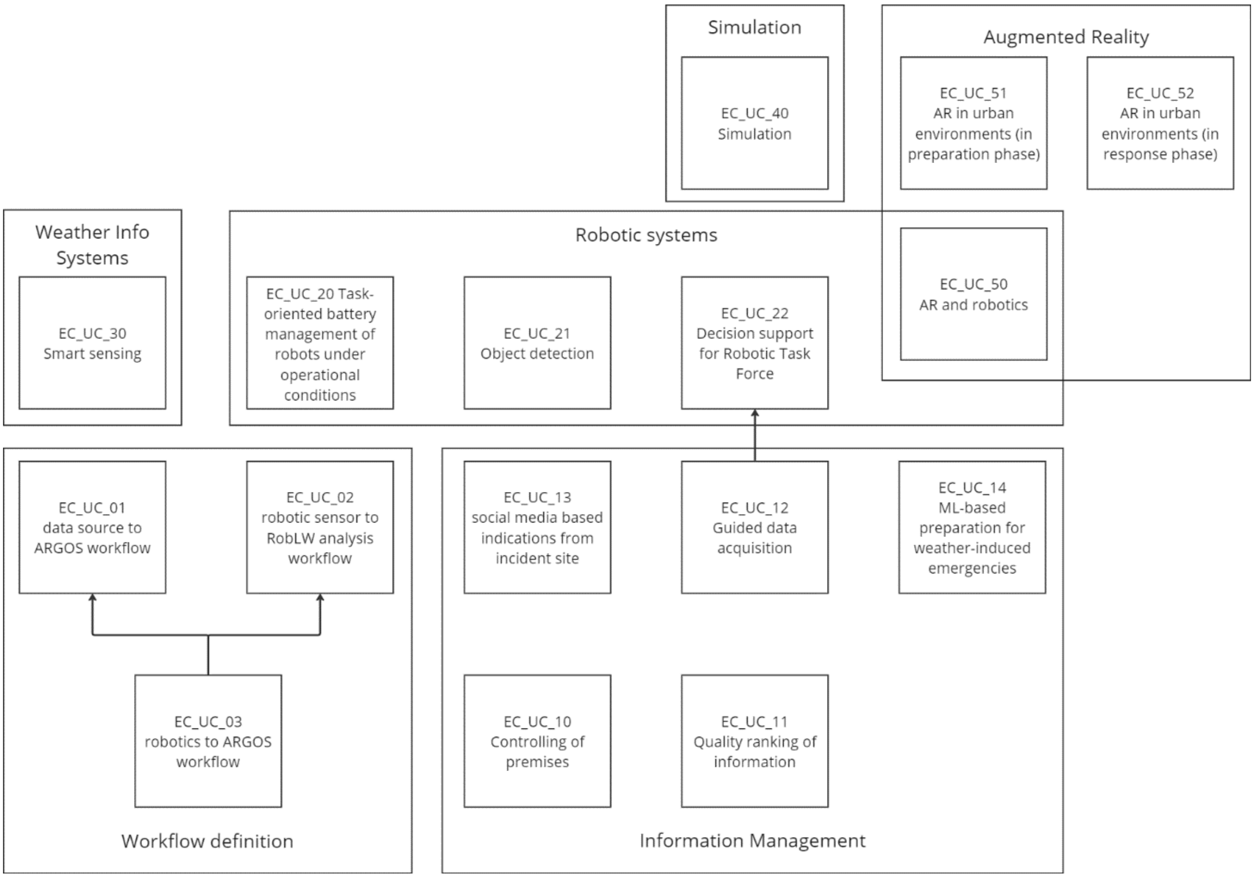

Figure 38: Overview of application sub-scenarios (use case narratives) (Figure from CREXDATA Deliverable D2.1).

As an initial restriction, there will not be any sub-scenario where AR is used in staff room environments. That would have to be combined with physical object, so it is out of scope at least for the initial phase of CREXDATA.

3.1.5 Mapping of Sub-Scenarios to WP3-WP5 Technologies

Table 4 on the next page shows which emergency use case sub-scenarios use which technologies developed in the CREXDATA work packages WP3 to WP5.

Table 4: Uptake of technologies in the Weather Emergency Use Case (Table from CREXDATA Deliverable D2.1).

Specific Use Case / Technologies Used	1	2	3	4	
T3.2 Graphical Workflow Specification	Em_UC_01	Em_UC_02	Em_UC_03	all others	
T4.1 Complex Event Forecasting	Em_UC_10	Em_UC_12			
T4.5 Text Mining for Event Extraction	Em_UC_13				
T4.2 Interactive Learning for Simulation Exploration	Em_UC_40 **	Em_UC_12 **			
T4.3 Federated Machine Learning	Em_UC_20	Em_UC_21	Em_UC_30	Em_UC_12	Em_UC_14
T4.4 Optimized Distributed "Analytics as a Service"	Em_UC_01 *	Em_UC_02 *	Em_UC_03 *		
T5.1 Explainable AI	Em_UC_11	Em_UC_14	Em_UC_20	Em_UC_22	Em_UC_30
T5.2 Visual Analytics Supporting XAI	Em_UC_11	Em_UC_14	Em_UC_20	Em_UC_22	Em_UC_30
T5.3 Visual Analytics for Decision Making under Uncertainty	Em_UC_10	Em_UC_11	Em_UC_22	Em_UC_30	
T5.4 Augmented Reality in the Field	Em_UC_50	Em_UC_51	Em_UC_52		
T5.5 Uncertainty Visualization in Augmented Reality	Em_UC_50	Em_UC_51	Em_UC_52	Em_UC_11	
* to be detailed based on workflow descriptions					

Specific Use Case / Technologies Used	1	2	3	4	
** to be detailed based on simulator selection					

Field trials are conducted within real or at least realistic settings of stakeholders in Dortmund, Germany, and Austria. Details are provided in the following Sections 3.1.6, 3.1.7 and 3.1.8.

3.1.6 Evaluation Scenario and Data Source for Weather Emergency Use Case Pilot Site in Dortmund, Germany

The pilot site in Dortmund will test the CREXDATA system architecture and the weather emergency demonstrator architecture with a pluvial flooding scenario in the city of Dortmund.

3.1.6.1 Application Scenario

The application scenario is introduced in Section 3.1.4 as a preparation of use case analysis. In this Section, the fundamental scenario is adopted and transferred to the specific environment of the city of Dortmund in Germany. A pluvial flooding is assumed to happen in specific areas of Dortmund. All aforementioned use cases are assumed to be relevant in this application scenario. Thus, crisis management and high-level emergency management of FDDO are required. The emergency is extreme both in terms of spatial and temporal extensions. The scenario is drafted based on experiences made during the extreme weather event in western Germany, especially in the Ahr region and around Erftstadt 2021. A detailed report of 325 pages is available [18] (cf. [19, 20]). Some sub-scenarios with corresponding use cases are setup in small-scale reproduction in the test-bed at DRZ.

Figure 39: Application scenario centered around Dortmund main station (Figure from CREXDATA Deliverable D2.1).

For the overall situation, risk maps are available that were created in standard preparation activities[4]. Figure 39 presents some insights into these maps assuming a flooding with 100 years return period, which are available in the control center and in the A level command post. The RobLW is deployed in an operational section that coordinates the use of UAVs and UGVs.

3.1.6.2 Available Data

The following data sources (cf. [D1.2, Section 3.1]) are available to provide data either continuously or in real-time during a test setup or in a field trial:

- Unmanned Vehicles[5] with robotic platform and payload in terms of sensor systems (off-the-shelf and experimental setups, see Section 0; available at DRZ and planned at UPB)
    a) UGVs
    b) UAVs
    c) Under-water robot (planned)
- RobLW[6]
    a) Applications for 3D model creation
    b) Communication networks
- AR devices (available at TUC and UPB)
    a) Microsoft HoloLens 2
    b) Tablets, smartphones
- Test bed
    a) Robot localization system (DRZ)
    b) Video observation system (UPB)

In desktop-like conditions, different devices are used to experiment with UIs. This subsumes laptops, tablets and smartphones, but also multi-touch displays and tables.

In the course of the project, it might be relevant to incorporate further equipment of FDDO. For instance, there is a specialized Analytical Task Force (ATF) operating a wide ranging of measuring equipment [21].

The following datasets are available (see details in [D1.2, Section 2.1.2]):

- Emergency Cases FDDO 2020-2021
- Damage clearance tasks in Finland
- Warnings for flooding in Dortmund
- flight data from Erftstadt (flood event 2021), available at DRZ (owned by the pilot)[7]

---

[4] https://geoweb1.digistadtdo.de/doris_gdi/mapapps4/resources/apps/starkregengefahrenkartetn100/index.html?lang=de&vm=2D&s=10000&r=0&c=393280.24263935274%2C5708048.11168041

[5] URL https://rettungsrobotik.de/en/testing-facility/the-robotic-systems-on-an-overview

[6] URL https://rettungsrobotik.de/en/testing-facility/the-robotic-command-vehicle

[7] cf. https://www.youtube.com/watch?v=BIq9P9NHbT0

3.1.7 Evaluation Scenario and Data Source for Weather Emergency Use Case Pilot Site in Austria

For Innsbruck, the coordination in the event of an operation is compliant with the introduction of Section 3.1.4. The emergency dispatch center is led by “Leitstelle Tirol”, they alert the emergency organization (fire department). Depending on the size of the event, there is also an official command (crisis management on municipality level) in addition to the fire department operational command. These two commands work closely together. In the case of pluvial flooding, as for example in the district of Amras, there was only one operational command at the fire department. This coordinates the disaster case. In addition, there is an operation site management and various operation teams on site in the disaster area. An exceptional situation in Innsbruck is that there is the Landeswarnzentrale Tirol, LWZ (in which the Leitstelle Tirol is integrated), which functions as an official coordinating body. That means that the Leitstelle Tirol and the fire departments provide data to the LWZ. Thus, they collect the data and sends the warning/alarm or all-clear to the population if requested by the authorities. Furthermore, the LWZ has a drone competence center, which can send drones to the disaster area on request and send aerial images, thermal images, etc. via live stream to the official geo-information-based situation management system KATGIS (provided by Geosphere Austria, the Federal Geological Service).

3.1.7.1 Application Scenario

In the case of the Austrian pilot site, two application scenarios are under discussion, in which the fundamental scenario (introduced in Section 3.1.4) is adopted and transferred:

- pluvial flooding due to a heavy rainfall in the city of Innsbruck, i. e. Amras district
- fluvial flooding due to the Danube river in Lower Austria, i. e. Tulln an der Donau,

whereby the focus and implementation, as well as the exchange with stakeholders, are currently being placed on the use case in Innsbruck (Figure 40). If it is deemed necessary during the project, further discussions and activities will be made towards the fluvial scenario in Lower Austria.

In general, the application scenario in the city of Innsbruck aims to (i) increase the situational awareness of key stakeholders, i.e. emergency response teams and decision makers (see Figure 40) and (ii) expand the possibilities of technology use in an urban emergency in an alpine environment. Due to the topographical conditions, the city of Innsbruck must also expect cascading debris flows during heavy rain events. In addition, due to the narrowness of the valley, there is little space to drain the water. Thus, in the event of local heavy precipitation, the city's drainage system is most likely overloaded within a very short time. The time component depends on various factors, such as the amount and duration of precipitation, the mobilization of loose material from the slopes and other obstacles, but also on the time of year – e.g. leaves in autumn. In addition, such thunderstorm cells in alpine areas are often accompanied by hail which can cause blockage of drainage systems. In combination with a supra-regional, regional heavy precipitation event, the Inn River can also cause flooding in the city of Innsbruck and turn the disaster operation into a major event. Additional remedial measures are required here, e.g. with mobile flood protection and retention basins. But in this application scenario, we will mainly focus only on pluvial flooding, as these have been severely increasing in recent years and pose extreme challenges to emergency services. I.e. in a very short time, city districts can be completely under water, buildings have to be evacuated because their stability is no longer guaranteed, and people are in danger. In Innsbruck, debris flows can be triggered as secondary processes and endanger people and infrastructure. The action of the emergency forces is required in the shortest possible time. This application scenario is planned to be processed on the basis of the heavy precipitation event of July 2, 2016, which severely flooded the district of Amras (see pictures in Figure 40).

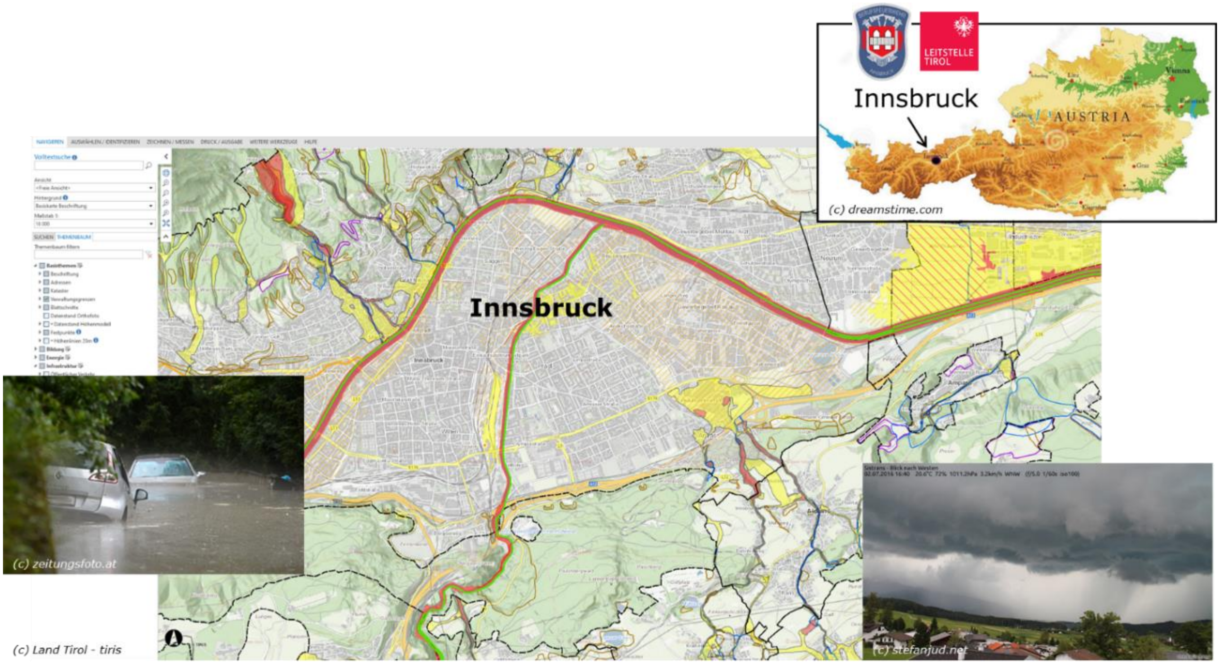

Figure 40: Spatial environment of the application scenario in Innsbruck (https://maps.tirol.gv.at/) (Figure from CREXDATA Deliverable D2.1).

3.1.7.2 Available Data

The following data sources (cf. [D1.2, Section 3.1]) are available to provide data either continuously or in real-time during a test setup or in a field trial. First indications after an initial workshop with third parties in Innsbruck are:

- Professional Fire Brigade Innsbruck:
  - a) Deployment protocols
  - b) Radio communication data (incl. location data)
- ECC Tyrol (Leitstelle Tirol):
  - a) Emergency call data (except from the police)
- National Warning Center Tyrol (Landeswarnzentrale Tirol) [optional, not yet sure - still in discussion]:
  - a) Deployment protocols
  - b) Data collected and provided for the key stakeholders in the geoinformation-based systems (i) webGIS tiris OEI - Local operation information, and (ii) katGIS (Digital situation recording and situation information)

- IKB: Innsbrucker Kommunalbetriebe AG (municipal infrastructure service company) [optional, not yet sure - still in discussion]:
    a) Water level and (surface) runoff data of sensors in relation to the drainage system
    b) Supply bottlenecks (energy, electricity, etc.)
- Geosphere Austria [optional, not yet sure - still in discussion]:
    a) Meteorological raw data (precipitation, river and groundwater level, river flow rate of sensors)
    b) Rain radar data
    c) Weather forecasts incl. thunderstorm cells (national to local)
    d) Heavy rainfall forecasts for pre-defined, small areas
    e) INCA (Integrated Nowcasting through Comprehensive Analysis) data

The availability of continuous data (and especially in real-time) on the part of the stakeholders still needs to be clarified, as well as whether access to the local IT system is possible at all. This is also very critical from the point of view of data protection. In general, as discussions are still ongoing with the stakeholders, other or additional datasets may arise as the project progresses. Nevertheless, it should also be noted here that in the case of Innsbruck these data might not be available.

The following datasets (cf. [D1.2, Section 3.2]) might be made available:

- Professional Fire Brigade Innsbruck:
    a) Deployment protocols
    b) Radio communication data (incl. location data)
    c) Operational documentation during/after the event (pictures, etc.) if available
- ECC Tyrol (Leitstelle Tirol):
    a) Emergency call data (except from the police)
- National Warning Center Tyrol (Landeswarnzentrale Tirol) [optional, not yet sure - still in discussion]:
    a) Deployment protocols
    b) Operational documentation during/after the event
    c) Drone recordings during/after the event (aerial photos, thermal images, terrain maps, etc.)
    d) Data collected and provided for the key stakeholders in the geoinformation-based systems (i) webGIS tiris OEI - Local operation information, and (ii) katGIS (Digital situation recording and situation information)
- IKB: Innsbrucker Kommunalbetriebe AG (municipal infrastructure service company) [optional, not yet sure - still in discussion]:
    a) Water level, (surface) runoff data of sensors in relation to the drainage system
    b) Supply bottlenecks (energy, electricity, etc.)
- Geosphere Austria [optional, not yet sure - still in discussion]:
    a) Meteorological raw data (precipitation, river and groundwater level, river flow rate of sensors)

b) Rain radar data
c) Weather forecasts incl. thunderstorm cells (national to local)
d) Heavy rainfall forecasts for pre-defined, small areas
e) INCA (Integrated Nowcasting through Comprehensive Analysis) data

- Department of Bridge and Hydraulic Engineering (City of Innsbruck) [optional, still in discussion]:
  a) Hazard / susceptibility maps

However, as discussions are still ongoing with the stakeholders, other or additional datasets may arise as the project progresses.

3.1.7.3 Stakeholder Requirements

A first workshop was held with the stakeholders of the Innsbruck application scenario on June 6, 2023, and discussions were initiated to which extent the technology from CREXDATA can be used and what the requirements are. These are as follows:

- Weather forecasting in different time scales (depending on the type of event): In the case of forecasted pluvial floods, and early prediction is deemed necessary (lead time >>1h).
- Forecasting of cascading events such as debris flows.
- Forecasting of critical events with enough lead time, of e.g. overloading of the canal/drainage system, instability of buildings, flooding of underground constructions (parking lots, underpasses)
- Facilitate and simplify communication among stakeholders in the event of an incident.
- Use of AR during pluvial floods: Visualization of predicted water level at different time intervals, trapped people in floating cars, etc.
- Use of underwater robots and/or drones during pluvial floods: Detection of trapped, buried people in floating cars, or of hazardous materials in water (chemical hazards), etc.

However, as discussions are still ongoing with the stakeholders, other or additional requirements may arise as the project progresses.

3.1.8 Involvement of Finnish Experts

The involvement of Finnish experts, including Finnish Meteorological Institute (partner, FMI), The Ministry of Interior Finland (partner, MoIFI) and the Rescue Department of Helsinki (external stakeholder) is built around showcasing the use of machine learning in weather-related impact forecast and early warning tool development. The aim of the showcase is to negotiate with the data owners to find, utilize and make new datasets open within the project consortium as well as openly for everyone. The Finnish showcase focuses not only on one, but on several weather hazards which have a large impact on Finnish emergency management in different seasons. The aim is to use statistical modelling and machine learning to produce forecasting products that describe the forecasts in a useful and concrete way for the end users in Finland and possibly also in other pilot locations, such as in Dortmund.

3.1.8.1 Application Scenario

Currently FMI is providing for instance following services for end users where machine learning can be utilized and where it can give a significant support to the impact estimations of forecasters or emergency managers:

- Severe Weather Warnings: FMI issues alerts for hazardous weather events, enabling proactive measures and emergency planning.
- Early Warning Systems: Collaboration with civil protection to develop systems that detect and forecast (the impacts of) extreme weather events, including prompt evacuation and response coordination.
- Specialized Forecasts: Tailored forecasts for specific sectors like emergency management assist in decision-making, minimizing risks and optimizing operations.
- Data Dissemination: Collecting, analysing, and disseminating weather- or weather impact-related data, enabling risk assessments, emergency planning, and response coordination.

The scenarios of the Finnish showcase are built rather on the impacts of different weather hazards than one specific weather hazard. In Finland a variety of weather-related hazards are experienced throughout the year. The autumn and winter season are dominated by strong windstorms with strong winds, and occasionally also with heavy snowfall conditions. These windstorms and falling trees cause lot of clearance tasks for the rescue department and inconvenience for public in form of power outages, damage to infrastructure, hazardous road conditions, and threat to human lives. One of the tools developed in the project is specifically tailored for forecasting the number of clearance tasks of upcoming windstorms. During heavy snowfall cases, the tool forecasting the number of road traffic accidents helps the rescue department to plan their resource management in a case of difficult winter storm and extreme road weather conditions (Figure 41).

During the summer months, Finland experiences in increasing frequency extreme heatwaves and dry weather conditions, which increase the risk of forest fires as well. Forest fires cause threat to people and infrastructure. The preparedness of the rescue department for forest fires can be increased by providing a tool that is estimating directly the number of wildfire fighting events on the Helsinki and Uusimaa region instead of traditional weather forecast predicting weather conditions and leaving space for individual interpretation.

The impacts of weather-related hazards are often consequences of combination of meteorological, environmental and infrastructural factors. The Finnish ML-based service can be taken up for instance in the complex event forecasting of T4.1. The impacts of the complex events may be difficult to grasp and understood solely by human brain, and thus machine learning algorithms are excellent aid for instance in the data analysis, recognizing complex patterns in weather data, and detecting early signs of specific events. ML service can be developed to create more accurate impact forecasts and enhancement of weather warnings or early warning systems, which often are missing the estimation of weather hazard impacts. The information and analysis of FMI’s ML service can be also utilized in T5.2 (Visual analytics supporting XAI) in simplifying and interpretation of complex and vast weather data. The outcomes of T5.3 (Visual analytics for decision making under uncertainty) can be possibly tested as an extension of the ML service tools to visualize the uncertainties of the forecasts in the understandable way, which is currently missing in the existing ML based tools of FMI.

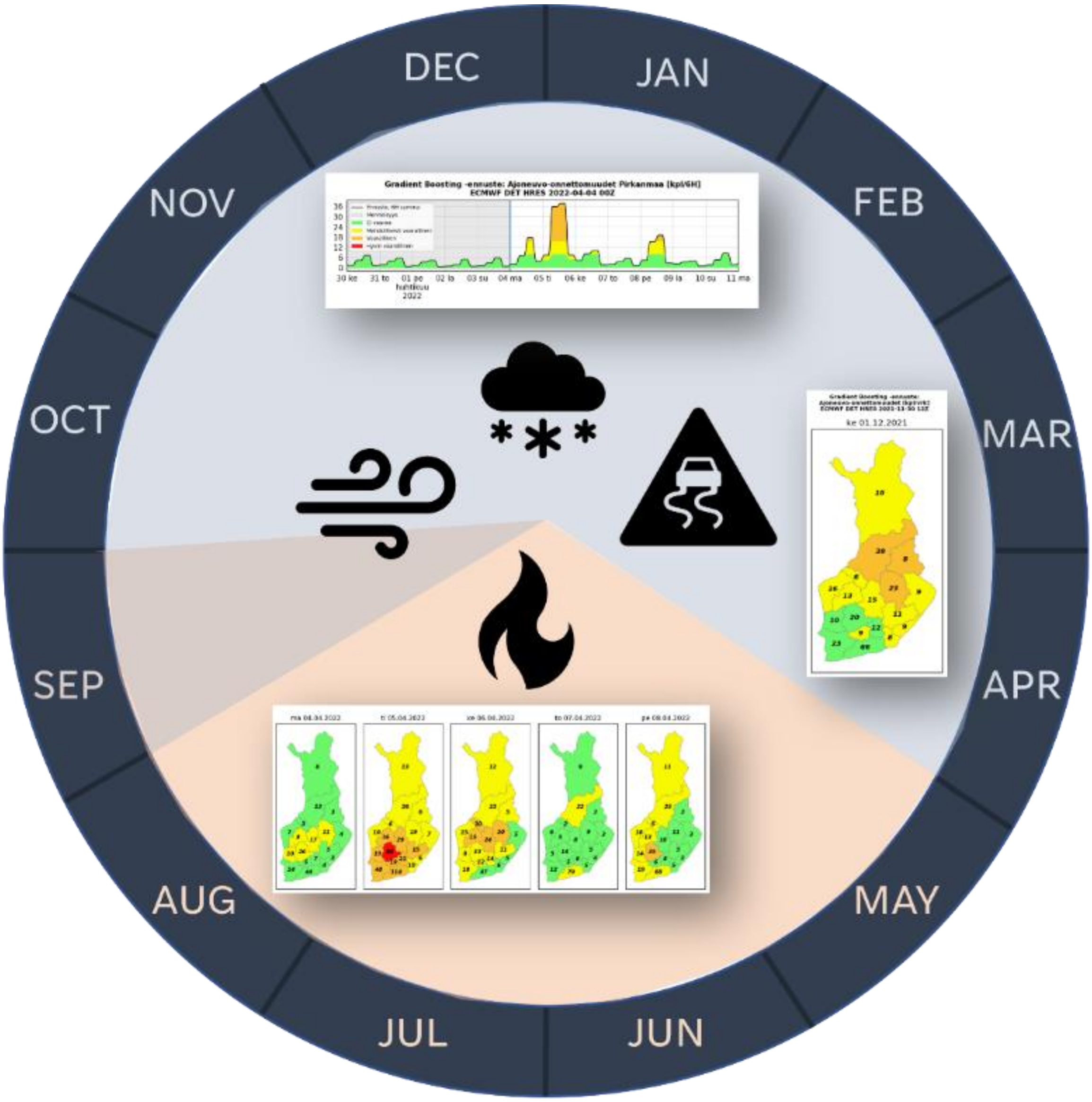

Figure 41: The seasonal occurrence of the main weather hazards in Finland (forest fires in summer, wind- and snowhazards in winter). The developed impact forecast tools aim to address various impact variables induced by these hazards (Figure from CREXDATA Deliverable D2.1).

3.1.8.2 Available Data

In a Finnish national level SILVA-project (2020-2023), a comprehensive weather-related impact database was created. Of 13 collected impact datasets, four were made openly available. In CREXDATA we demonstrate the use of these four datasets presented in Table 5 with their volume, temporal and geographical coverage, and resolution. The advantage of these datasets especially from the machine learning perspective is that the time series lengths are long and the accuracy of both spatial and temporal resolution of the impacts is sufficient for training the model that requires large amounts of data to perform well. Additionally, currently the ECMWF HRES weather model is used, and later MEPS limited area ensemble weather prediction model is tested as the predictor data representing the weather-dependency of the impacts. In CREXDATA, the possibility of including new datasets and making them openly available is also being explored. The plan is also to utilize for instance the number of ambulance operations between 2007 and 2023. The open impact data can be made available to other partners also through ARGOS system, as well as selected meteorological data described more in details in D1.2.

Table 5: Impact datasets used for training and validating the gradient boosting machine learning method that are openly available. The ambulance operation dataset is a new dataset and the use is being explored in the project (not openly available) (Table from CREXDATA Deliverable D2.1).

DATASETS	VOLUME	TIME SERIES LENGTH	COVERAGE	RESOLUTION
Wind damage clearance	130k	22 yrs	National	• Municipal • 1 hour
Wildfire fighting events	63k	22 yrs	National	• Municipal • 1 hour
Traffic accident clearance	281k	22 yrs	National	• Municipal • 1 hour
Road traffic accidents	1.1M	24 yrs	National	• Municipal • 1 hour
Ambulance operations	~1M	16 yrs	Helsinki region	• Accurate coordinates • 1 second

3.2 Health Use Case

This use case will assess WP4 abilities to enable critical action planning and intervention by providing efficient parameter exploration forecasting and effective interventions in the modelling of epidemics and drug treatment optimization in COVID-19 infection. Additionally, this task will specify the requirements and scenarios that will be used by the novel tools developed in T2.5. The evaluation will be performed on two scenarios. In the epidemics scenario, we will use epidemiological compartmental models to build a digital twin for COVID-19 transmission that identifies efficient policies and enables accountability. In the drug treatment scenario, we will use multiscale mechanistic models to build a digital twin of a drug assay in COVID-19 patients that identifies the best treatment for each patient and condition [DoA, p.8].

3.2.1 Stakeholders

Epidemiological simulations in a health crisis have the potential to benefit a wide range of stakeholders. Government and public health agencies can leverage these simulations to make informed decisions and effectively allocate resources, implement effective Non-Pharmaceutical Interventions and design optimal vaccination campaigns. By understanding the potential spread of diseases and evaluating the impact of different intervention strategies, policymakers can develop evidence-based policies and guidelines. Healthcare providers and hospitals can utilize the simulations to assess the strain on healthcare systems, such as the demand for hospital beds, ICUs, ventilators, and the healthcare workforce. This information aids in resource planning, capacity management, and optimizing healthcare delivery to meet the needs of the affected population.

Emergency management and disaster response agencies also find value in epidemiological simulations. These simulations assist in understanding potential scenarios, predicting resource requirements, and strategizing response plans. By incorporating simulation results into their preparedness efforts, these agencies can develop response protocols, coordinate multi-agency efforts, and ensure effective coordination and implementation of response measures. Additionally, researchers and academia benefit from epidemiological simulations as they provide a tool for testing hypotheses, exploring different scenarios, and analysing the potential impact of interventions. By contributing to scientific knowledge, simulations inform research directions and support the development of evidence-based guidelines for mitigating health crises.

Each of the identified stakeholders is interested in application scenarios for forecasting the evolution of new potential epidemics, detecting new outbreaks or wave and finding efficient intervention to reduce the impact from different perspectives. Likewise, they are also interested in novel, optimised drug treatments that provide alterative clinical care pathways for COVID-19 patients.

The list of functionalities proposed will address the specific end-user requirements regarding the healthcare system. All developed functionalities will potentially target one or more of the following key healthcare system managers, public health decision-makers and data scientist/IT service provider needs.

We have taken advantage of the EBI's Competency Hub developed by us in a recent project to characterise some of these personas in more detail. The PerMedCoE competency framework defines a series of competencies required of professionals in the field of computational personalised medicine. A competency is an observable ability of any professional, integrating multiple components such as knowledge, skills and behaviours. The competencies that an individual might need to fulfil a particular role are listed in the reference profiles, which can be used to guide career choices[8].

[8] This competency framework is available at: https://competency.ebi.ac.uk/framework/permedcoe/ 2.1 and https://permedcoe.eu/deliverables/ (Deliverable 4.2).

The PerMedCoE competency profile builds on the work done in several related initiatives, as we took inspiration from the competency profiles of BioExcel[9], CINECA[10] and ISCB[11] to create an initial draft that was updated with feedback from experts from the community. All the profiles are freely accessible through the EMBL-EBI Competency Hub[12]. Additionally, the Competency Hub allows users to create their own profile on the site and to compare it with the existing profiles, which can inform about career development options.

Scenario 1: Designing Effective NPIs and Vaccination Campaigns for Controlling a Disease Outbreak using epidemiological modelling

In this scenario, a region is facing a sudden outbreak of a highly contagious infectious disease. The objective is to use epidemiological simulations to evaluate potential scenarios under different assumptions such as different reproduction numbers and fatality rates when those parameters are still unknown or hard to estimate due to the absence of data. In the second stage, optimization-via-simulation can be used for the design and implement effective Non-Pharmaceutical Interventions (NPIs) and vaccination campaigns to control the disease spread and minimize its impact on the population.

Scenario 2: Finding optimised drug treatments and alternative clinical care pathways for COVID-19 patients using multiscale modelling

In this scenario, a patient has been identified as being infected by SARS-CoV-2 and has been taking in charge by a hospital. The objective is to have a digital twin of the clinical care pathway using a multiscale model to propose clinical interventions of drugs and NPIs (such as mechanical ventilation) that allows the patient to have a healthy status. For this, first we will need to couple two simulators (Alya and PhysiBoSS) and fit different parameters to clinically-relevant variables. Second, we will use optimization-via-simulation to design and implement patient-specific, effective combinations of interventions that heal the patients.

3.2.2 Available Data

As an initial assumption, no real-time data sources are available. Available data sets are described per health sub-scenario in the following Sections 3.2.2.1 and 3.2.2.2 on the following two pages.

[9] https://competency.ebi.ac.uk/framework/bioexcel/2.0

[10] https://competency.ebi.ac.uk/framework/cineca/1.0

[11] https://competency.ebi.ac.uk/framework/iscb/3.0

[12] https://competency.ebi.ac.uk/framework/permedcoe/2.1

3.2.2.1 Scenario 1: Epidemiological Modelling

The following datasets (cf. [D1.2, Section 4.2) are available on Zenodo. The open datasets consist of COVID-19 case reports and population mobility patterns in the form of origin destination matrices, both reported on a daily basis:

- COVID19 Flow-Maps GeoLayers dataset: Geographic layers on which the different data records are geo-referenced (e.g., mobility, COVID-19 cases). The different layers can be grouped into those that cover the whole territory of pain (e.g., municipalities) and those that are restricted to a specific region (Table1). Among those that cover the full territory of Spain, the record accounts for the first four levels of administrative division, that is, autonomous communities, provinces, municipalities and districts [22].
- COVID19 Flow-Maps Daily Cases Reports: This repository contains COVID-19 data for Spain, including daily cases at the level of autonomous communities as well as provinces, and higher spatial resolution for several autonomous communities (eight out of the nineteen autonomous communities publish reports with local daily COVID-19 cases at the level of municipalities or Basic Health Areas). Each record has an identifier, the associated date, the corresponding identifier of the layer and code of the region and a set of COVID-19 related fields, which include the number of new cases (daily incidence) and total cases. The dataset includes case reports for a time period of approximately two years [23].
- COVID19 Flow-Maps Daily-Mobility for Spain: This data-set contains daily aggregations of the hourly data provided by MITMA, aggregated at different levels of spatial resolution. The dataset includes Origin-Destination matrix for the mobility layer, with hourly resolution Each entry has a date and time period (the range between two consecutive hours), the origin and destination zones and the number of trips from a origin to a destination. Origin and destination zones correspond to geometries from the MITMA mobility layer and internal trips (same layer of origin and destination) are also reported. Additionally, it also includes a data record containing the trips per person matrix on each mobility area on a daily basis. This indicator reports population-based daily mobility behaviour. For each date and zone from the MITMA mobility layer, the indicator reports how many persons have performed 0, 1, 2 or more than 2 trips. While the indicator does not provide the destination of the trips, it accounts for the fractions of people performing at least one trip or none, as well as the estimated total population in that zone for the given date, considering as population those persons who stay overnight in the zone on that date [24].
- COVID19 Flow-Maps Population data. Daily population and trips per person data from Spain 2020-2022. This data record contains daily population records based on a study conducted by the MITMA, that analysed the mobility and distribution of the population in Spain from February 14th 2020 to May 9th 2021. The study is based on a sample of more than 13 million anonymised mobile phone lines provided by a single mobile operator whose subscribers are evenly distributed. Data provided by MITMA is related to the layer mitma_mov. For the rest of the layers, the population was estimated using the population grid from GEOSTAT[13] [25].

13 https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat

3.2.2.2 Scenario 2: Multiscale Lung Infection Modelling

- Anonymized patient omics molecular data. Publicly available pseudo-anonymized raw experimental data from patients. This dataset bundles different studies that will be the input used to analyze and personalize our multiscale models. It potentially consists of transcriptomics, genomics, copy number variations and proteomics data.
- Anonymized patient pulmonary 3D positional data. Publicly available pseudo-anonymized image data from patients. This dataset bundles different studies that will be the input used to have complex 3Dal setups for our multiscale models. Once analysed, this dataset will have positional data for each of the alveoli of the patients.

3.2.3 System Architecture for Demonstrator

Table 6: Demonstrator components extending the CREXDATA system (Health Use Case) (Table from CREXDATA Deliverable D2.1).

Component	Description
Epidemiological Scenario	
T2.4 – Simulation and Tools	• Creation of a synthetic mobility dataset for evaluation of different scenarios • User interface development. The end-users of the user interface will be able to view the forecast the evolution of the epidemic process under different scenarios • Simulation Framework: Integration of population mobility data with the epidemiological models. The architecture will give the capability to simulate/replicate a test scenario using both, real and synthetic data of the COVID-19 pandemic in Spain, with potential extensions to other countries.
T3.2 Graphical Workflow Specification	• Graphical tools, for instance using different operators in RapidMiner, will be developed to allow a non-expert programmer to specify complex data processing workflows to enable the fusion of different spatiotemporal data sources georeferenced to different territorial units.
T4.1 Complex Event Forecasting	• NCSR/BSC will develop novel algorithms to forecast new outbreak hot-spots based times series of cases, population mobility
T4.2 Interactive Learning for Sim. Exploration	• Extreme-scale model exploration will be combined with interactive learning approaches to explore the large space of epidemiological parameters, as well as, optimal interventions.
T4.3 Federated Machine Learning	• Federated Machine Learning (FML) offers a privacy-preserving approach to calibrating epidemiological parameters in a pandemic scenario. It allows multiple data owners to collaborate and train a shared machine learning model without sharing their sensitive data. The calibrated epidemiological parameters obtained through

Component	Description
	FML will be validated using a time series of COVID-19 cases reported at different levels of spatial aggregations. Robust statistical methods can be applied to assess the accuracy and uncertainty of the calibrated parameters.
T5.3 Visual Analytics for Decision Making under Uncertainty	• Creates visual representations of the epidemic model outputs. This can include interactive maps, charts, graphs, and dashboards that depict the spread of the disease over time, hotspots of infection, and the effectiveness of interventions.
Multiscale lung infection Scenario	
T2.4 – Simulation and Tools	• Simulators: Coupling organ-level with cell-level simulation tool • Study the use of surrogate models for parts of the algorithms. • Use model exploration to fit some of the parameters using clinical data or desired simulated behaviours. • Prepare a set of design variables that control the treatment of patients (drug and mechanical interventions) that will be inspected using interactive learning.
T3.2 Graphical Workflow Specification	• Graphical tools, for instance using different operators in RapidMiner, will be developed to allow a non-expert user to use and browse complex workflows that simulate patient treatments using clinical data and bedside variables.
T4.1 Complex Event Forecasting	• NCSR and BSC will develop novel algorithms to forecast at early times the outcomes of a lung infection simulation.
T4.2 Interactive Learning for Sim. Exploration	• Extreme-scale model exploration will be combined with interactive learning approaches to explore the large space of potential clinical interventions and simulate the patient's clinical care pathway until recovery.
T5.3 Visual Analytics for Decision Making under Uncertainty	We will create visual representations of different simulations of the multiscale infection model. For instance, we will study the usefulness of GUIS and dashboards to ease the exploration of the parameter sensitivity analyses and their effect on model outputs.

Table 7: Uptake of technologies in the Health Use Case (Table from CREXDATA Deliverable D2.1).

Specific "use cases"	Parameters calibration and optimal intervention design	Forecasting of outbreak hotspots	Fusion of different spatiotemporal data sources	Forecasting of lung infection dynamics	Interactive learning of COVID19 patients clinical care pathway
T2.4 Simulation and Tools	X	X		X	X
T3.2 Graphical Workflow Specification			X		X
T4.1 Complex Event Forecasting	X	X		X	X
T4.2 Interactive Learning for Simulation Exploration	X				X
T4.3 Federated Machine Learning	X				
T4.4 Optimized Distributed "Analytics as a Service"					
T5.1 Explainable AI					
T5.2 Visual Analytics supporting XAI					
T5.3 Visual Analytics for Decision Making under Uncertainty	X	X		X	X
T5.4 Augmented reality at the field					
T5.5 Uncertainty Visualization in Augmented Reality					

3.3 Maritime Use Case

The scenarios will be validated using created streams of data and in sea trial experiments with numerous vessels with several levels of autonomy as defined by the International Maritime Organization (IMO) in cooperation with the SMARTMOVE Lab of the University of the Aegean (UoA). Sea trials under realistic conditions guarantee an application-related real-world assessment. Data collection and sea trials will be conducted during dedicated experiments at the Aegean University sea testbed and during the Aegean Ro-boat Races planned to take place annually in the summer periods starting in July 2023. The sea testbed of the University of Aegean is located on the island of Syros in the Aegean Sea and offers proximity to open sea testing grounds. It covers sea and land utilizing 100% WiFi coverage and is capable of hosting and deploy several types of UVs: air, land, sea and subsurface. A running prototype version of the IoT-Voyage Data Streamer – VDS was deployed at the 1st Aegean Ro-boat Race in July 2023. The sea trials are used to collect real world datasets from onboard the competing vessels, which will be used to test components in real world maritime conditions.

3.3.1 Stakeholders

The following key stakeholder groups are identified in the context of the Maritime Use Case. Each stakeholder has specific interests directly linked to aspects regarding maritime safety, or is accountable for ensuring safe maritime operations, safe navigation as well as passenger, crew and cargo safety and is part of the decision making and accountability chain in the context of hazardous maritime events detection, forecasting and mitigation.

Table 8: Key stakeholder groups & roles of the Maritime Use Case (Table from D2.1).

Key stakeholder group	Key stakeholder roles
Action planner	• Vessel pilot & Vessel crew • VTS operator • Port/coastal authorities • Remote operator
Decision maker	• Ship deck officers • VTS operator • Port/coastal authorities • Fleet managers • Vessel owners • Remote operator
System administrator	• IT / Data Science staff of the maritime service provider • IT / Data Science staff of the port authority • IT / Data Science staff of the vessel owners • Fleet managers • Insurance companies
Workflow designer	• IT / Data Science staff of the maritime service provider • IT / Data Science staff of the port authority • IT / Data Science staff of the vessel owners • Fleet managers • Insurance companies

Each of the aforementioned stakeholders are interested in application scenarios for global vessel safety, tracking and management from his/her own perspective. The envisaged functionalities address the specific end-user requirements regarding maritime safety. All developed functionalities will potentially target one or more of the following key maritime users and data scientist/IT service provider needs. Resulting functionalities that are listed below are in line with the importance evaluation results of the user requirements survey contacted by MT (Kpler) (see service importance evaluation Table 13):

- Vessel's route analysis and/or prediction.
- Early warnings of possible collisions and near-real time collision avoidance.
- Early warnings of possible intrusion of sea areas with hazardous weather conditions.
- Rerouting and mitigation actions for collision mitigation
- Rerouting and avoidance of sea areas with forecasted hazardous weather conditions.

The key personas involved directly in the detection and management of a maritime emergency are identified in Table 11.

3.3.2 Application scenario

In the context of the Maritime Use Case two application scenarios will be examined:

- collision forecasting and rerouting
- hazardous weather rerouting

Both application scenarios that are part of the pilot will be evaluated in sea trial experiments using the vessel of high autonomy (test vessel) of the SMARTMOVE Lab of the University of Aegean. Regarding the collision forecasting application scenario, a collision event will be simulated at the sea test bed with the test vessel forecasting the imminent collision event in a short-term time horizon of under 15 minutes according to the KPIs. In a subsequent step a rerouting option will be provided in order to mitigate the collision event. Regarding the hazardous weather rerouting, using simulated and synthetic data streams, first a sea area will be designated as an area with hazardous weather conditions. The test vessel will be provided with information regarding its imminent approach to the area with hazardous weather conditions along with rerouting instructions. The demonstration of both events during sea trial experiments at the University of Aegean Sea test bed will validate the pilot of the maritime use case under realistic conditions and guarantee an application-related real-world assessment.

Table 9 and Table 10 present the high-level usage scenarios for the CREXDATA Maritime Use Case incorporating the identified key stakeholders and considering the results of the user requirements survey (see: Section 3.3.5 Stakeholder Requirements). For each of the scenarios, apart from the actor, the overview and the detailed description, the main benefits and challenges are provided.

Table 9: Collision forecasting and rerouting high level scenario description (Table from CREXDATA Deliverable D2.1).

Attribute	Description
ID	Maritime_UC01
Name	Collision Forecasting and Rerouting

Attribute	**Description**
Short description	Forecasting of probable collision events among vessel in a short time prediction horizon of under 15 minutes. In case of a collision event detection, the proposed service informs the relevant actors of the collision event detection and provides rerouting suggestions for the mitigation of the collision event. The actors decide either to follow the service's instructions or correct the proposed rerouting suggestion based on their own local view
Author	MT (Kpler)
Last update	09.06.2023
Actors	• Vessel pilot • Vessel crew • VTS operator • Ship deck officers
Additional Actors	• Vessels of different levels of autonomy
Actors interested in the outcomes	• Port/coastal authorities • Fleet managers • Vessel owners • Insurance companies
Detailed scenario	The pilot of vessel A is alerted of a possible collision event with vessel B, as their current routes will intersect. The pilot evaluates the emergency of the forecasted event and is provided with a set of alternative routes for vessel A to follow in order to avoid colliding with vessel B. Based on the experience of the vessel pilot, he/she may opt to accept the proposed route, correct the proposed suggestion or follow an entirely different route in order to avoid collision with vessel B
Benefits	• Route monitoring and collision event forecasting for vessel traffic increasing safe navigation and efficiency of maritime operations • Improved route predictions of the vessel traffic through fusion of local and global data streams • Automation of collision event detection forecast and mitigation steps through automated rerouting suggestions. • Increased situational awareness • Early identification of possible collision events providing comfortable response time windows and informed decision support • Exploration of collision mitigation alternatives

Attribute	Description
	• Vessel crew and VTS operator work effort alleviation during traffic monitoring, decision making and action planning for vessel collision events
Challenges	• Accuracy of vessel path prediction in short term time horizons • Accuracy of path planning and rerouting for collision event mitigation • Near real-time response of the system for collision event detection • Near real-time fusion of different stream inputs from local and global data sources • Quantification of prediction and solution uncertainty • Real time re-evaluation of vessel route after collision event detection and continuous monitoring of the involved vessels' routes • Small response time path planning methods for rerouting • Real-time extraction of extreme-scale situational data

Table 10: Hazardous weather rerouting high level scenario description (Table from CREXDATA Deliverable D2.1).

Attribute	Description
ID	Maritime_UC02
Name	Hazardous weather rerouting
Short description	Rerouting of vessels in order to avoid sea areas with forecasted hazardous weather conditions. Based on weather forecast updates vessels are monitored and alerted on changes of the weather conditions along their route and provided with rerouting instructions
Author	MT (Kpler)
Last update	09.06.2023
Actors	• Vessel pilot • Vessel crew • VTS operator • Ship deck officers
Additional Actors	• Vessels of different levels of autonomy
Actors interested in the outcomes	• Port/coastal authorities • Fleet managers

Attribute	Description
	• Vessel owners • Insurance companies
Detailed scenario	The crew of a vessel starts their journey and plan their route according to their initial weather forecast. As weather dynamically changes over the journey, the vessel crew receives hourly updates in case of weather conditions influencing the passage safety through specific sea areas. In case of changes affecting the safe passage, an alert with automatic rerouting suggestion is generated by the system alleviating the vessel crew from the task of continuously monitoring the weather conditions and updating the vessel route.
Benefits	• Route monitoring and rerouting according to updated weather forecasts for global vessel traffic, increasing safe navigation and efficiency of maritime operations. • Improved rerouting of the vessel traffic according to forecasted weather conditions through fusion of local and global data streams. • Automation of the weather monitoring during a vessel's journey • Safe navigation due to avoidance of hazardous weather areas. • Increased situational awareness and early identification of hazardous weather conditions that give operators time to plan an alternative route. • Informed decision making • Effective fleet intelligence and management based on future weather forecast at global scale • Vessel operators are alleviated from weather forecast monitoring tasks
Challenges	• Accuracy of rerouting as a function of weather forecast uncertainty • Reliability on longer vessel routes with duration exceeding the weather forecast time range • Fusion of different stream inputs from local and global data sources • Quantification of prediction and solution uncertainty

3.3.3 Available Data

The following data sources will be collected during the annual Aegean Ro-Boat races organized by the University of the Aegean (UoA). Local environment data sources refer to sensors mounted on the UoA vessel. Additionally, the MT (Kpler) IoT-Voyage Data Streamer – VDS will be the only additional sensor mounted on both the UoA vessel and the RoBoat Race participating vessels during the races. Data from the RoBoat Race will be made available for batch processing to the CREXDATA partners after the end of the annual RoBoat Race (the partners could be in position to simulate the timeseries data in streaming mode, if applicable per dataset type). All the data sources listed here could be potentially used for the needs of the Maritime Use Case (see appendix of Deliverable D2.1).

The following datasets (cf. [D1.2, Section 5.2]) are available on Zenodo. The open datasets consist of AIS related data and available for all partners to work on. Additional data will be generated and provided through the annual Aegean Ro-Boat Races organized by the University of Aegean that will be publicly released on Zenodo:

- Single Ground Based AIS Receiver Vessel Tracking Dataset: This dataset published by MT (Kpler), contains all decoded messages collected within a 24h period (starting from 29/02/2020 10PM UTC) from a single receiver located near the port of Piraeus (Greece). All vessels' identifiers such as IMO and MMSI have been anonymized and no down-sampling procedure, filtering or cleaning has been applied.
- Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance: This dataset contains ships' information collected through the Automatic Identification System, integrated with a set of complementary data having spatial and temporal dimensions aligned. The dataset contains four categories of data: Navigation data, vessel-oriented data, geographic data, and environmental data. It covers a time span of six months, from October 1st, 2015 to March 31st, 2016 and provides ships positions within the Celtic Sea, the Channel and Bay of Biscay (France). The dataset is proposed with predefined integration and querying principles for relational databases. These rely on the widespread and free relational database management system PostgreSQL, with the adjunction of the PostGIS extension, for the treatment of all spatial features proposed in the dataset.
- The Piraeus AIS Dataset for Large-scale Maritime Data Analytics: The AIS dataset (coming from MT's (Kpler's) receiver) comes along with spatially and temporally correlated data about the vessels and the area of interest, including weather information. It covers a time span of over 2.5 years, from May 9th, 2017 to December 26th, 2019 and provides anonymized vessel positions within the wider area of the port of Piraeus (Greece), one of the busiest ports in Europe and worldwide. The dataset consists of over 244 million AIS records, an average of more than 10,000 records per hour, which makes it an ideal input for large-scale mobility data processing and analytics purposes.
- Hellenic Trench AIS Data: Data from Automatic Identification System (AIS) transmissions received from both satellite and terrestrial receivers of the Marine Traffic network (www.marinetraffic.com) for one year (31 July 2015 to 31 July 2016) along the Hellenic Trench, the core habitat of the eastern Mediterranean.

3.3.4 System Architecture for Demonstrator

Table 11: Demonstrator components extending the CREXDATA system (Maritime Use Case) (Table from CREXDATA Deliverable D2.1).

Component	Description
T2.4 – Simulation and Tools	• Creation of a synthetic dataset of simulated AIS data. • User interface development. The end-users of the user interface will be able to view the forecast motion of vessels in the future of each predicted route and potential mitigation actions • Simulation Framework: Integration of synthetic collision data with the VR interface. The architecture will give the capability to simulate/replicate a test scene using real historical and synthetic data. • Define paradigms for interactive exploration of the model behaviours using synthetic simulation data using VR
T3.2 Graphical Workflow Specification	• Integration of the maritime use case applications with the CREXDATA platform • Integration with the RapidMiner graphical workflow. Extension from INFORE with new operators: Fusion operator, forecasting operator and rerouting operator • Integration of existing RapidMiner operators from INFORE: Maritime Event Detector, Fusion • Development of new Kafka streams and related operators
T3.3 – System Integration and Released Software Stacks	• Development of software prototype
T4.1 Complex Event Forecasting	• MT (Kpler) / UoA will develop in-house models for route and collision forecasting • MT (Kpler) /UoA will develop an in-house solution for hazardous weather rerouting
T4.4 Optimized Distributed “Analytics as a Service”	• AKKA distributed tool to run fusion and route prediction models • Technical details and goals will be clarified at a later stage. Potential provision of AKKA performance analytics for resource optimization
T5.3 Visual Analytics for Decision Making under Uncertainty	• The pilot will support TUC for visual Analytics to facilitate decision making for the collision and weather rerouting under uncertainty • Technical details and goals will be clarified at a later stage.

Component	Description
T5.4 Augmented reality at the field	• The pilot will support TUC for an AR/VR prototype on the maritime use case • Technical details and goals will be clarified at a later stage.
T5.5 Uncertainty Visualization in Augmented Reality	• The pilot will support TUC for uncertainty visualization in augmented reality on the maritime use case • Technical details and goals will be clarified at a later stage.

Table 12: Uptake of technologies in the Maritime Use Case (Table from CREXDATA Deliverable D2.1).

Specific "use cases"	Collision Forecasting	Hazardous Weather Rerouting
T2.4 Simulation and Tools	X	X
T3.2 Graphical Workflow Specification	X	X
T4.1 Complex Event Forecasting	X[1]	X[1]
T4.2 Interactive Learning for Simulation Exploration		
T4.3 Federated Machine Learning		
T4.4 Optimized Distributed "Analytics as a Service"	X[2]	X[2]
T4.5 Text Mining for Event Extraction	Not relevant	Not relevant
T5.1 Explainable AI		
T5.2 Visual Analytics supporting XAI		
T5.3 Visual Analytics for Decision Making under Uncertainty	(X)[3]	
T5.4 Augmented reality at the field	(X)[3]	
T5.5 Uncertainty Visualization in Augmented Reality	(X)[3]	

[1] MT's (Kpler's) models for route and collision prediction will be developed
[2] Akka distributed framework will be used to run fusion and route prediction models
[3] support of potential TUC contribution

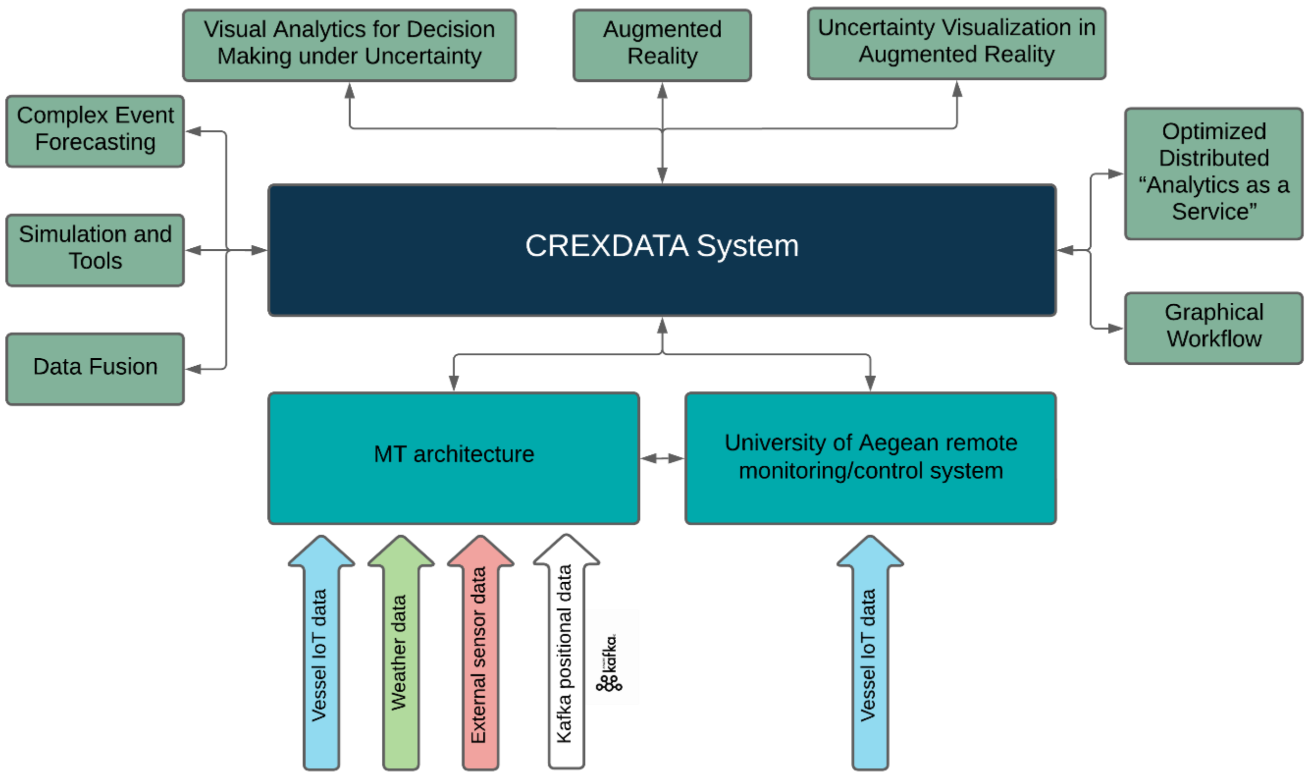

Figure 42: High-level system architecture for the Maritime Use Case interlinked with the CREXDATA system and its components (Source: CREXDATA Deliverable D2.1).

Figure 43 presents the MarineTraffic (Kpler) system architecture supporting the deployment of the Maritime Use Case pilot.

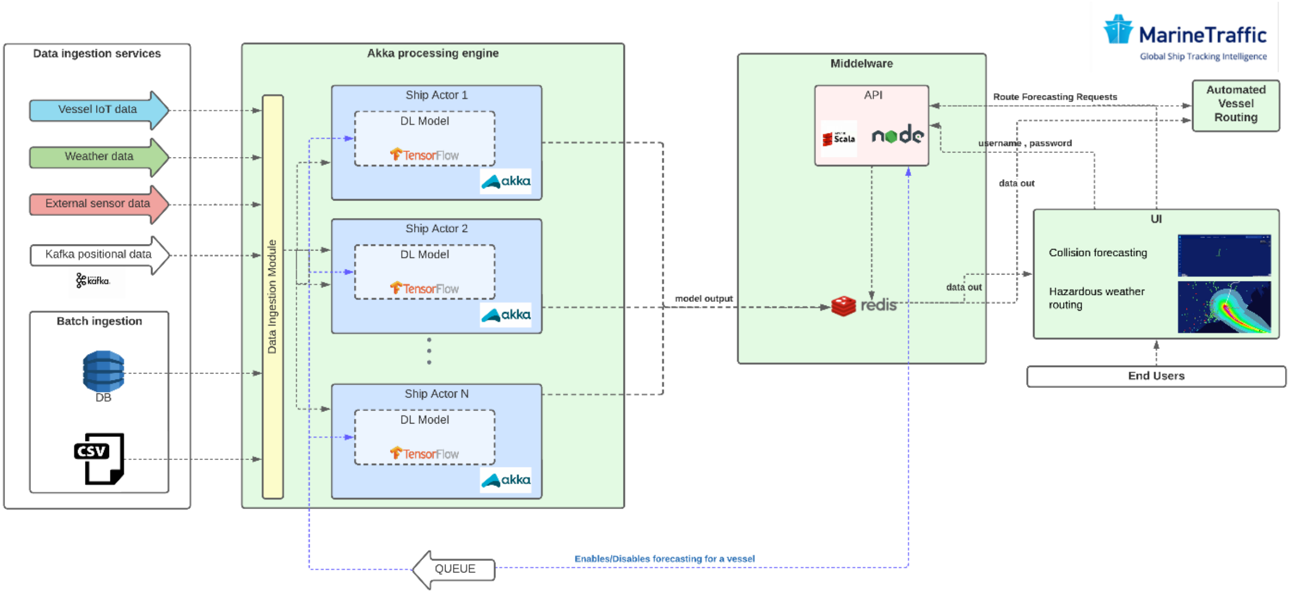

Figure 43: MarineTraffic (Kpler) system architecture for the Maritime Use Case (Source: CREXDATA Deliverable D2.1)

3.3.5 Stakeholder Requirements

CREXDATA deliverable D2.1 describes the stakeholder requirements for the maritime use case collected in a survey. With respect to the importance of forecasting the defined hazardous maritime events, maximum acceptable forecasting latency and the received information from the corresponding forecasting service, results for the respective requirement with the maximum number of consensus votes are presented in Table 13 on the next page.

Table 13: Maritime Use Case User Requirements (Table from CREXDATA Deliverable D2.1).

Requirement	Collision forecasting MAR_1	Hazardous weather routing MAR_2
service importance	very important	very important
maximum acceptable latency	minute latency	hour latency
provided information features to the end-users	• ETA to conflict point • prediction confidence • rerouting information with path suggestion	• ETA to destination port • prediction confidence • rerouting information with path suggestion
suggestions for possible courses of action	yes	yes
frequency of updating mitigation actions	every minute	every hour
data sources / data sets for development	• real-time / streaming AIS data • real-time/Streaming IoT vessel data • historical AIS data	• real-time / streaming weather data • historical weather data
data sources/ data sets for evaluation	• Sea trial/experimental data	• Historical vessel data

In the context of facilitating informed decision making by the end-users during hazardous maritime event, the survey defines the respective requirements for the development of such services. In the context of the CREXDATA platform according to the respective questionnaire participants the development of such services should be based on specific data features that may be available through different data sources.

3.4 Federated Data Processing and Federated Machine Learning

In addition to the requirements of the three use cases described in the three previous sections 3.1 to 3.3, there are also requirements from an algorithmic and technological perspective: The CREXDATA platform and streaming architecture need to be able to support federated data processing and federated machine learning as well:

- Federated machine learning deals with training models or applying models at the edges of the network, i.e. distributed across multiple IoT or edge devices, some of which may have limited resources regarding the amount of available memory and compute power.
- A streaming process can be split up by the optimizer component of the CREXDATA platform to be placed for execution on multiple compute clusters (which could be located at different locations) and/or on multiple edge devices. Federated data processing and federated machine learning process data in a distributed way, i.e. without storing or passing all of the data to a central location, but by processing it locally and only passing on the results, e.g. only relevant observations and events, aggregates, extracted features, or machine learned models.
- The above requires an interface between the CREXDATA platform and the federated data processing and federated machine learning services, which exposes the methods that (1) can be queried to get the number of nodes and a list of nodes available, and/or (2) provide configuration parameters in order to trigger training of a model using a particular algorithm and parameters on the federated data processing resources (compute nodes, federated machine learning algorithm and its parameters, etc.).
- The query service may also be used by the optimizer component to get statistics or related information about the different compute clusters and nodes supported by the cluster for federated data processing and federated machine learning purposes.
- The outcomes of the training or scoring action taking place on the federated data processing backend service is to be consumed just as in other streaming workflow operators.
- The federated data processing backend could be packaged as a bunch of docker containers with pre-configured python environment and packages installed, so that the workflow operators do not have to invoke installation and configuration methods or pass scripts. Dockerized backends also enable more portability and scalability.

For the communication between the CREXDATA platform, the CREXDATA optimizer, and the federated data processing and federated machine learning services, JSON-based web services and Kafka-based event and data streams are used.

In order to integrate the distributed data processing and the distributed machine learning into CREXDATA data processing workflows, a nested operator, i.e. a RapidMiner operator with a subprocess, is implemented in the CREXDATA project, named “Edge Nest”, and made available as a RapidMiner Extension. The subprocess of the Edge Nest operator is the part of the data processing workflow that is to be distributed to and executed on the edge devices. Hence it is possible to design data processing workflows that combine distributed parallel execution of data processing on multiple edge and IoT devices (via the Edge Nest operator) with non-distributed centrally executed data processing, e.g. for collecting and merging the results of the distributed data processing and federated machine learning.

The location, where a particular data processing workflow or a subprocess of it are to be run, i.e. the IP address of the node to execute the data processing workflow or subprocess thereof may be changed by the CREXDATA optimizer.

The CREXDATA platform maintains a node registry for managing a list of the IP addresses and port numbers of all available data processing nodes (compute nodes) including IoT and edge devices, compute clusters, servers, etc.

The list of nodes to be used for federated processing can be a subset of the list of all available nodes, i.e. the CREXDATA platform allows to filter the list of nodes, e.g. by selecting or deselecting nodes in the list individually or by filtering nodes by node type, category of nodes, group of nodes, or use case application id.

The CREXDATA optimizer component can adjust, where which part of the data processing workflow (i.e. operator) will be run. The optimizer only needs to change the IP addresses and port numbers of the nodes to specify where to execute the particular part of the data processing workflow (i.e. operator). The CREXDATA process execution engine then takes these IP addresses and port numbers and runs the operators on the corresponding nodes, i.e. IoT and edge devices, servers, and clusters, respectively.

The integration from the edge to the central CREXDATA system may also include other systems and components. Example for the weather emergency use case: Instead of a sending its data directly to the CREXDATA system, an edge device like a drone taking arial video footage, photos, etc. can also connect to a use-case-specific system (e.g. ARGOS by HYDS), which in turn connects to the CREXDATA system.

4 Initial Software Stacks

This section summarizes the software tools and software stacks used and developed within the CREXDATA project. More detailed descriptions of how these tools are used within the CREXDATA system are available in Section 2 (system architecture and system components and system integration approach) and Section 3 (domain- and use-case-specific components like for example domain-specific simulators).

4.1 Data Source Systems and Data Collection

The data source systems of the different use cases are described in Sections 2.5 and 3. For the weather emergency use case, for example, the system ARGOS of CREXDATA project partner HYDS integrates the data needed for this use case (weather data, events, sensor data, IoT and edge device data, etc.) and connects to the CREXDATA platform via Kafka, in order to provide the data to the CREXDATA platform and to get predictions from the CREXDATA platform in return.

4.2 Data Stream Connection: Kafka

Kafka is used as middleware for the communication and data transfer between the CREXDATA system components. The use of Kafka and web services for the integration of the CREXDATA system is described in Section 2.

4.3 Data Stream Infusion and Processing: Altair RapidMiner

The Altair RapidMiner data science software platform is used as underlying platform for data integration and fusion, data stream processing, federated data processing, etc., which is extended with CREXDATA-specific data connectors and modules (extensions) to form the core of the CREXDATA system.

Altair RapidMiner AI Studio (with CREXDATA-specific extensions) as is used as the visual data processing workflow editor for code-free visual process workflow design.

Altair RapidMiner AI Hub is used as central server and optionally for further distributed servers or cloud servers.

Altair RapidMiner Real-Time Scoring Agents (RTSA) are used for data processing on the edge and IoT devices.

4.4 Domain-Specific Simulators

Use-case-specific simulators and CREXDATA system components are described in Section 2.5 and in Section 3.

4.5 Source Code Repository for CREXDATA Open-Source Software

The source code of the open-source components of the CREXDATA system as well as RapidMiner processes and system configurations files are shared and published via the following Bitbucket repository:

https://bitbucket.org/crexdata/

Software source code, RapidMiner process files (i.e. specifications of the data processing workflows for the various use cases), system configuration files, and documents shared via this repository will be continuously updated during the duration of the CREXDATA project (and most likely beyond).

5 CREXDATA Objectives Addressed by the CREXDATA System Architecture

The CREXDATA architecture described in Sections 1 and 2 addresses the CREXDATA objective MO1.1 by supporting multi-modal data ingestion and fusion and by developing corresponding algorithms and techniques in the CREXDATA project that are incorporated in as operators into the CREXDATA system, so that they can be easily used in the visual data workflow design in the graphical user interface of the CREXDATA workflow editor. The visual workflow designer based on Altair RapidMiner AI Studio with the extensions developed in the CREXDATA project including the data connectors for the various data sources required by the use cases composes a data ingestion and fusion toolbox for bringing together and aligning data of multiple modalities and formats including for example time series data, sensor data, weather data, text data, image data, video data, JSON, geometries, raster, graph, spatiotemporal data, etc., for which the extension mechanism provides a flexibly configurable way to accommodate the various use case scenarios with their respective requirements.

The CREXDATA system components, like the underlying data integration and data science platform Altair RapidMiner and Kafka as a middleware for connecting the system components and for their communication and data exchange ensure the capability for processing extreme-scale data streams in real-time and support federated data processing and federated machine learning not only on computer clusters, servers, and in the cloud, but also on IoT and edge devices. These underlying system components together with the above-mentioned visual workflow designer and its pluggable extensions provide the envisioned outcome: a Pluggable Extreme-scale Data Ingestion and Fusion Toolbox in the CREXDATA IDE (Integrated Development Environment) with cross-library support and the ability to plug-in custom User Defined Functions (UDFs).

The CREXDATA platform integrates use-case-specific systems (e.g. ARGOS by HYDS in the weather emergency use case) to integrate use-case-specific data sources (e.g. sensors and edge devices), use case specific simulators (see Sections 2 and 3), as well as use-case-specific functionalities and capabilities via the extension (plugin) mechanism (e.g. for algorithms and techniques developed in the CREXDATA work packages WP3, WP4, and WP5).

6 Conclusions

In this report we presented the initial CREXDATA system architecture, along with its system components, the system integration approach, and the initial version of the released software stacks of the CREXDATA project. This is all motivated by all aspects of the use cases and their scenarios. The initial system architecture of the prototype being conceptualized and developed in the CREXDATA project includes a graphical tool for designing processing workflows. This graphical design tool interacts with and influences the system architecture, which must handle multi-modal data fusion from dispersed sources and of different types (extreme scale data streams, time series data, text data, image data, video data, weather data, etc.). The overall system integration relies on our developed graphical design tool.

In the following months, the initial system prototype and software components will be further developed (and documented in Deliverable D3.2). The objective remains to ease the specification of complex data processing workflows for non-expert programmers, while utilizing technology stacks to the maximum possible extent. In the next reporting period, we also aim to collect feedback for the further enhancement of the CREXDATA architecture from different stakeholders (use case and technical partners in the project).

7 Acronyms and Abbreviations

Each term should be bulleted with a definition.

Below is an initial list that should be adapted to the given deliverable.

- AI – Artificial intelligence
- AI Hub – Altair RapidMiner AI Hub – server version of the Altair RapidMiner data science software platform
- AIS – Automatic Identification System
- AI Studio – Altair RapidMiner Studio – desktop version of the Altair RapidMiner data science software including a visual editor for designing data processing workflows
- API – Application Programming Interface
- AR – Augmented Reality
- ATF – Analytical Task Force
- C2 – Command & Control
- CA – Consortium Agreement
- CEF – Complex Event Forecasting
- CER – Complex Event Recognition
- CEP – Complex Event Processing
- D – deliverable
- DoA – Description of Action (Annex 1 of the Grant Agreement)
- DQ – Data Quality
- DWD – Deutscher Wetterdienst (German Weather Service)
- EB – Executive Board
- EC – European Commission
- ECC – Emergency Control Center
- ECMWF –European Centre for Medium-Range Weather Forecasts
- EFAS – European Flood Awareness System
- EFFIS – European Forest Fire Information System
- EDO – European Drought Observatory
- EMSA – European Maritime Safety Agency
- ERCC – Emergency Response Coordination Centre
- EMS – Emergency Management System (Copernicus)
- ETA – Estimated Time of Arrival
- EUCPM – European Union Civil Protection Mechanism
- GA – General Assembly / Grant Agreement
- GDPR – General Data Protection Regulation
- GPS – Global Positioning System
- GUI – Graphical User Interface
- HMD – Head-Mounted Display
- HMI – Human Machine Interface
- HPC – High Performance Computing
- HRES – High-Resolution Forecast
- ICU – Inertial Control Unit
- ID – (unique) identifier
- IDE – Integrated Development Environment

- IMO – International Maritime Organization
- IoT – Internet-of-Things
- IPR – Intellectual Property Right
- IQ – Information Quality
- IT – Information Technology
- KPI – Key Performance Indicator
- M – Month
- ML – Machine Learning
- MMSI – Maritime Mobile Service Identity
- MoSCoW – Must have, Should have, Could have, Won't have
- MS – Milestone
- MT – Maritime Traffic, now a subsidiary of Kepler
- NIST – National Institute of Standards and Technology (US)
- NLP – Natural Language Processing
- NPI – Non-Pharmaceutical Interventions
- PaaS – Prediction-as-a-Service
- PM – Person Month (or Project Manager, depending on the context)
- PPDR – Public Protection and Disaster Relief
- ReqIF – Requirements Interchange Format
- RM – RapidMiner – Altair RapidMiner – data ingestion, data processing, and data science platform underlying the CREXDATA system
- RM AI Hub – Altair RapidMiner AI Hub – server version of the Altair RapidMiner data science software platform
- RM AI Studio – Altair RapidMiner AI Studio – desktop version of the Altair RapidMiner data science software featuring a graphical user interface (GUI) including visual data processing workflow designer/editor
- RobLW – Command car of a special robotic emergency response unit
- ROS – Robot Operating System
- RTF – Robotic Task Force
- RTSA – Altair RapidMiner Real-Time Scoring Agent, a light-weight version of Altair RapidMiner AI Hub, optimized for low resource consumption, low latency, and high throughput and hence deployable on IoT and edge devices
- Rviz – ROS Visualizer
- SEIR – Susceptible-Exposed-Infectious-Recovered
- SysML – Systems Modelling Language
- T – Task
- TRL – Technology Readiness Level
- UAV – Unmanned Aerial Vehicle
- UC – Use Case
- UGV – Unmanned Ground Vehicle
- UI – User Interface
- UML – Unified Modelling Language
- URL / URI – Uniform Resource Locator / Identifier
- UV – Unmanned Vehicle
- VDS – Voyage Data Streamer
- VTS – Vessel Traffic Service
- VR – Virtual Reality
- WebODM – Web Open Drone Map

- WMS – Web Map Service
- WP – Work Package
- WPL – Work Package Leader
- XAI – eXplainable Artificial Intelligence

8 References

[1] Pruitt J, Grudin J (2003) Personas: Practice and theory. In: Arnowitz J, Chalmers A, Swack T et al. (eds) Proceedings of the 2003 conference on Designing for user experiences - DUX '03. ACM Press, New York, New York, USA, pp 1–15

[2] Rupp C (2014) Requirements-Engineering und -Management: Aus der Praxis von klassisch bis agil, 6., aktualisierte und erweiterte Auflage. Hanser, München

[3] Gräßler I, Oleff C (2022) Systems Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg

[4] OMG (2016) Requirements Interchange Format (ReqIF), 1.2th edn.

[5] (2019) ReqIF Implementation Guide: Referring to ReqIF1.2

[6] OMG (2019) Systems Modeling Language, 1.6th edn.

[7] Vallejo P, Mazo R, Jaramillo C et al. (2020) Towards a new template for the specification of requirements in semi-structured natural language. JSERD 8:3. https://doi.org/10.5753/jserd.2020.473

[8] Pottebaum J, Artikis A, Marterer R et al. (2012) User-Oriented Evaluation of Event-Based Decision Support Systems. In: 2012 IEEE 24th International Conference on Tools with Artificial Intelligence. IEEE, pp 162–169

[9] Yin RK (2018) Case study research and applications: Design and methods, Sixth edition. SAGE, Los Angeles, London, New Dehli, Singapore, Washington DC, Melbourne

[10] Trucchia A, D'Andrea M, Baghino F et al. (2020) PROPAGATOR: An Operational Cellular-Automata Based Wildfire Simulator. Fire 3:26. https://doi.org/10.3390/fire3030026

[11] Arenas A, Cota W, Gómez-Gardeñes J et al. (2020) Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions. Phys Rev X 10. https://doi.org/10.1103/PhysRevX.10.041055

[12] Vázquez M, Houzeaux G, Koric S et al. (2016) Alya: Multiphysics engineering simulation toward exascale. Journal of Computational Science 14:15–27. https://doi.org/10.1016/j.jocs.2015.12.007

[13] Ponce-de-Leon M, Montagud A, Noel V et al. (2022) PhysiBoSS 2.0: a sustainable integration of stochastic Boolean and agent-based modelling frameworks

[14] Letort G, Montagud A, Stoll G et al. (2019) PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics 35:1188–1196. https://doi.org/10.1093/bioinformatics/bty766

[15] Zurich Insurance Group Ltd (2023) Three common types of floods explained. https://www.zurich.com/en/knowledge/topics/flood-and-water-damage/three-common-types-of-flood. Accessed 12 Jun 2023

[16] Pottebaum J, Schafer C, Kuhnert M et al. (2016) Common information space for collaborative emergency management. In: 2016 IEEE Symposium on Technologies for Homeland Security (HST). IEEE, pp 1–6

[17] Kruijff-Korbayova I, Grafe R, Heidemann N et al. (2021) German Rescue Robotics Center (DRZ): A Holistic Approach for Robotic Systems Assisting in Emergency Response. In: 2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). IEEE, pp 138–145

[18] (2022) Starkregenereignis „Bernd“ 2021: Bericht zur Einsatznachbereitung, Berlin

[19] Szönyi M, Roezer V, Deubelli T et al. (2022) PERC floods following „Bernd“, Zürich

[20] (2022) Die Flutkatastrophe im Juli 2021 in Deutschland: Ein Jahr danach: Aufarbeitung und erste Lehren für die Zukunft. DKKV-Schriftenreihe, Bonn

[21] BBK Referat III.2 (2019) Die Analytische Task Force: Informationen zu Leistungsspektrum und Anforderungswegen, 3.0th edn., Bonn
[22] Miguel Ponce-de-Leon, Javier del Valle, José María Fernández et al. (2021) COVID19 Flow-Maps GeoLayers dataset. Zenodo
[23] Miguel Ponce-de-Leon, Javier del Valle, José María Fernández et al. (2021) COVID19 Flow-Maps Daily Cases Reports. Zenodo
[24] Miguel Ponce-de-Leon, Javier del Valle, José María Fernández et al. (2021) COVID19 Flow-Maps Daily-Mobility for Spain. Zenodo
[25] Miguel Ponce-de-Leon, Javier del Valle, José María Fernández et al. (2021) COVID19 Flow-Maps Population data. Zenodo
[26] Batini C, Scannapieco M (2006) Data quality: Concepts, methodologies and techniques. Data-centric systems and applications. Springer, Berlin, Heidelberg
[27] Eppler MJ (2006) Managing Information Quality. Springer Berlin Heidelberg, Berlin, Heidelberg
[28] Cimolino U (2014) Analyse der Einsatzerfahrungen und Entwicklung von Optimierungsmöglichkeiten bei der Bekämpfung von Vegetationsbränden in Deutschland
[29] Gizikis A, O'Brien T, Gomez Susaeta I et al. (2017) Guidelines to increase the benefit of social media in emergencies: EmerGent Deliverable 7.3, Paderborn
[30] Lorini V, Castillo C, Dottori F et al. (2019) Integrating Social Media into a Pan-European Flood Awareness System: A Multilingual Approach. In: Franco Z, González JJ, Canós JH (eds) Proceedings ofthe 16th ISCRAM Conference
[31] Havas C, Resch B, Francalanci C et al. (2017) E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time. Sensors (Basel) 17. https://doi.org/10.3390/s17122766
[32] Erat O, Isop WA, Kalkofen D et al. (2018) Drone-Augmented Human Vision: Exocentric Control for Drones Exploring Hidden Areas. IEEE Trans Vis Comput Graph 24:1437–1446. https://doi.org/10.1109/TVCG.2018.2794058
[33] Chen L, Takashima K, Fujita K et al. (2021) PinpointFly: An Egocentric Position-control Drone Interface using Mobile AR. In: Kitamura Y, Quigley A, Isbister K et al. (eds) Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp 1–13
[34] Suzuki R, Karim A, Xia T et al. (2022) Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In: Barbosa S, Lampe C, Appert C et al. (eds) CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp 1–33
[35] Gräßler I, Pottebaum J, Scholle P (2018) Influence Factors for Innovation in Digital Self-Preparedness Services and Tools. International Journal of Information Systems for Crisis Response and Management 10:20–37. https://doi.org/10.4018/ijiscram.2018010102
[36] Chowell G (2017) Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A Primer for parameter uncertainty, identifiability, and forecasts. Infect Dis Model 2:379–398. https://doi.org/10.1016/j.idm.2017.08.001
[37] Ponce-de-Leon M, Montagud A, Akasiadis C et al. (2022) Optimizing Dosage-Specific Treatments in a Multi-Scale Model of a Tumor Growth. Front Mol Biosci 9:836794. https://doi.org/10.3389/fmolb.2022.836794
[38] Akasiadis C, Ponce-de-Leon M, Montagud A et al. (2022) Parallel model exploration for tumor treatment simulations. Computational Intelligence 38:1379–1401. https://doi.org/10.1111/coin.12515

[39] Richardson RA, Wright DW, Edeling W et al. (2020) EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing. JORS 8:11. https://doi.org/10.5334/jors.303
[40] Rong H, Teixeira AP, Guedes Soares C (2019) Ship trajectory uncertainty prediction based on a Gaussian Process model. Ocean Engineering 182:499–511. https://doi.org/10.1016/j.oceaneng.2019.04.024
[41] Zhang W, Deng Y, Du L et al. (2022) A method of performing real-time ship conflict probability ranking in open waters based on AIS data. Ocean Engineering 255:111480. https://doi.org/10.1016/j.oceaneng.2022.111480
[42] Abebe M, Noh Y, Kang Y-J et al. (2022) Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. Ocean Engineering 256:111527. https://doi.org/10.1016/j.oceaneng.2022.111527
[43] Lyu H, Hao Z, Li J et al. (2023) Ship Autonomous Collision-Avoidance Strategies—A Comprehensive Review. JMSE 11:830. https://doi.org/10.3390/jmse11040830
[44] Walther L, Rizvanolli A, Wendebourg M et al. (2016) Modeling and Optimization Algorithms in Ship Weather Routing. International Journal of e-Navigation and Maritime Economy 4:31–45. https://doi.org/10.1016/j.enavi.2016.06.004
[45] Rawson A, Brito M, Sabeur Z et al. (2021) A machine learning approach for monitoring ship safety in extreme weather events. Safety Science 141:105336. https://doi.org/10.1016/j.ssci.2021.105336
[46] Vodas M, Bereta K, Kladis D et al. (2021) Online Distributed Maritime Event Detection & Forecasting over Big Vessel Tracking Data. In: 2021 IEEE International Conference on Big Data (Big Data). IEEE, pp 2052–2057
[47] Vouros GA, Vlachou A, Santipantakis G et al. (2018) Increasing Maritime Situation Awareness via Trajectory Detection, Enrichment and Recognition of Events. In: R. Luaces M, Karimipour F (eds) Web and Wireless Geographical Information Systems, vol 10819. Springer International Publishing, Cham, pp 130–140
[48] Xiao Z, Fu X, Zhang L et al. (2020) Traffic Pattern Mining and Forecasting Technologies in Maritime Traffic Service Networks: A Comprehensive Survey. IEEE Trans Intell Transport Syst 21:1796–1825. https://doi.org/10.1109/TITS.2019.2908191
[49] Chou C-C, Wang C-N, Hsu H-P (2022) A novel quantitative and qualitative model for forecasting the navigational risks of Maritime Autonomous Surface Ships. Ocean Engineering 248:110852. https://doi.org/10.1016/j.oceaneng.2022.110852
[50] Popov AN, Kondratiev AI, Smirnov IO (2018) The algorithm for fast forecasting of the collision danger degree with ships and surface objects in the e-navigation area. In: 19th Annual General Assembly – AGA 2018 International Association of Maritime Universities (IAMU)
[51] Xiao Z, Fu X, Zhang L et al. (2017) Data-driven multi-agent system for maritime traffic safety management. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp 1–6
[52] Jie W, Yao-Tian F (2008) Risk analysis based on the ship collision modeling and forecasting system. In: 2008 IEEE International Conference on Systems, Man and Cybernetics. IEEE, pp 1517–1521
[53] Zinchenko S, Nosov P, Mateichuk V et al. (2019) Automatic collision avoidance system with many targets, including maneuvering ones. Bul.Kar.Univ "Phys" Ser 96:69–79. https://doi.org/10.31489/2019ph4%2F69-79

[54] ARTUSI E (2021) Ship path planning based on Deep Reinforcement Learning and weather forecast. In: 2021 22nd IEEE International Conference on Mobile Data Management (MDM). IEEE, pp 258–260
[55] Zis TP, Psaraftis HN, Ding L (2020) Ship weather routing: A taxonomy and survey. Ocean Engineering 213:107697. https://doi.org/10.1016/j.oceaneng.2020.107697
[56] Shin YW, Abebe M, Noh Y et al. (2020) Near-Optimal Weather Routing by Using Improved A* Algorithm. Applied Sciences 10:6010. https://doi.org/10.3390/app10176010
[57] Grifoll M, Borén C, Castells-Sanabra M (2022) A comprehensive ship weather routing system using CMEMS products and A* algorithm. Ocean Engineering 255:111427. https://doi.org/10.1016/j.oceaneng.2022.111427
[58] Vettor R, Guedes Soares C (2016) Development of a ship weather routing system. Ocean Engineering 123:1–14. https://doi.org/10.1016/j.oceaneng.2016.06.035
[59] Vettor R, Szlapczynska J, Szlapczynski R et al. (2020) Towards Improving Optimised Ship Weather Routing. Polish Maritime Research 27:60–69. https://doi.org/10.2478/pomr-2020-0007
[60] Frydenberg S, Nordby K, Eikenes JO (2018) Exploring designs of augmented reality systems for ship bridges in arctic waters. In: Human Factors: RINA - International Conference on Human Factors
[61] Ostendorp M-C, Lenk JC, Lüdtke A (2015) Smart Glasses to Support Maritime Pilots in Harbor Maneuvers. Procedia Manufacturing 3:2840–2847. https://doi.org/10.1016/j.promfg.2015.07.775
[62] Takenaka M, Nishizaki C, Okazaki T (2019) Development of Ship Collision Prevention Device with Augmented Reality Toolkit. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, pp 4290–4295