

Automated Metadata Acquisition in Energy Research:

BPMN-driven Workflows at the KIT Energy Lab

Jan Martin Reckel, Malte Holzhäuer, Tobias Moser, Peter Moster, Simon Waczowicz, Veit Hagenmeyer

Institute for Automation and Applied Informatics (IAI)

iai.kit.edu



Agenda

1. Motivation
2. BPMN: A Visual and Executable Solution
3. Workflow Design and Implementation
4. Summary and Outlook

1. Motivation

- Energy research generates **massive, complex** datasets
- **KIT Energy Lab**: Research facility with large quantities of **devices, sensors** and other **data sources**
- **Manual** metadata definition, management and documentation is time-consuming and error-prone
- Increasing demand for implementation of FAIR principles
- Need for automatic and systematic metadata capturing and management



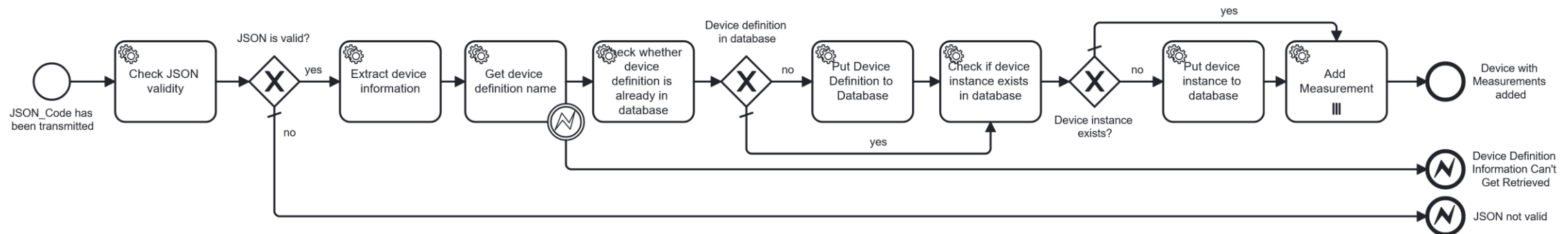
www.elab.kit.edu

2. BPMN: A Visual and Executable Solution

BPMN: Business Process Model and Notation

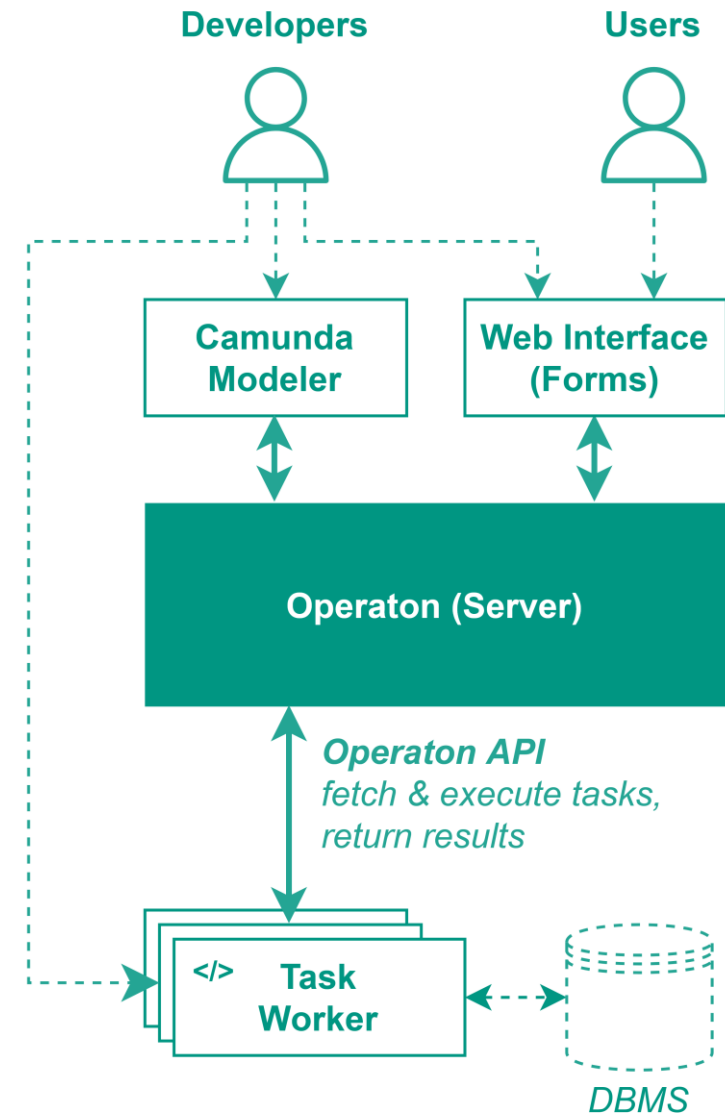
- Process modelling language and de facto standard for process modelling
- Widely adopted in industry and some research domains
- Multi-purpose:
 - Visual modelling and documentation
 - Machine-executable: Every process is code
 - Code files allow **version control**, **testing** and **CI/CD** using established tools

→ **Abstraction** of technicalities between **domain scientists**, **data stewards** and **other stakeholders**



3. Workflow Design

- Usage of **Operaton** BPMN engine, a Camunda 7 fork
- Process design using **Camunda Modeler**
- Execution of tasks with **external task worker approach**
 - Parallelization and scaling
 - Statistical evaluations
- All processes are traceable
- User tasks can provide **customized web forms** for users
- BPMN processes for metadata acquisition and ingestion into any DBMS
 - **Web forms:** user input + validation
 - **Task workers:** data validation, DBMS interactions
 - Different paths and end events



3. Workflow Design

The screenshot displays the Camunda Modeler interface with a BPMN workflow titled "create_or_update_device.l". The workflow starts with a start event "JSON_Code has been transmitted", followed by a task "Check JSON validity". A decision gateway "JSON is valid?" branches the flow: if "no", it leads to an end event "JSON not valid"; if "yes", it proceeds to "Extract device information", then "Get device definition name", and "Check whether device definition is already in database". Another decision gateway "Device definition in database" branches: if "no", it leads to "Put Device Definition to Database"; if "yes", it leads to "Check if device instance exists in database". A third decision gateway "Device instance exists?" branches: if "no", it leads to "Put device instance to database"; if "yes", it leads to "Add Measurement". The "Add Measurement" task leads to an end event "Device with Measurements added". There are also two error events: "Device Definition Information Can't Get Retrieved" and "JSON not valid".

On the right side, the "PROCESS" configuration panel is visible for "Add Device with Measurements to MetaDB".

PROCESS
Add Device with Measurements to MetaDB

General

Name: Add Device with Measurements to MetaDB

ID: Process_Add_Device_To_MetaDB

This maps to the process definition key.

Version tag: 0.15

Executable

Documentation

History cleanup

Time to live: 15

Tasklist

Candidate starter

External task

Process variables (16)

- > device_class
- > device_definition_docref
- > device_definition_in_database
- > device_definition_name
- > device_definition_uuid
- > device_instance_building
- > device_instance_description
- > device_instance_exists_in_database
- > device_instance_influx_measurement
- > device_instance_measurements

XML Camunda 7.24 5.45.0 Update

3. Workflow Design

The screenshot displays a web browser window with the URL `elab-operator.cloud.elab2.kit.edu/operaton/app/elab-tasklist/default/#/instance/26ff0238-3ef2-11f1-8910-da02411c6161?defid=cf0bffe-3cc5-11f1-8bf7-32127acbc77`. The page title is "Process Instance" and it shows a "Running" status. On the left, a form titled "Zu bearbeitendes Enum auswählen" (Select enum to be processed) is visible. It includes a dropdown menu for "Aktion auswählen*" (Action to select) with the option "Einen bestehenden Eintrag ändern" (Change an existing entry) selected. Below this is a section for "Zu bearbeitendes Element" (Element to be processed) with a list of "Enum-Definition" items: "place", "place", and "validity". A "Submit" button is at the bottom of the form. On the right, a BPMN diagram illustrates the workflow. The process starts with a start event, followed by a task "Enum-Definitionen aus Datenbank laden" (Load enum definitions from database), then a task "Zu bearbeitendes Enum auswählen" (Select enum to be processed), which is highlighted in green. This is followed by "Enum-Werte aus Datenbank laden" (Load enum values from database), "Enum-Werte bearbeiten" (Edit enum values), and "Angaben prüfen" (Check details). A decision gateway "Admin-Genehmigung benötigt" (Admin approval required) follows. If "Nein" (No), it leads to another decision gateway "Bearbeitungsaktion" (Processing action). If "Ja" (Yes), it leads to a task "Vorgang durch Administrator prüfen" (Check process by administrator). The "Bearbeitungsaktion" gateway has two paths: "INSERT" leading to "Neuen Enum hinzufügen" (Add new enum) and "UPDATE" leading to "Bestehendes Enum aktualisieren" (Update existing enum). Both paths lead to end events: "Neuer Enum wurde hinzugefügt" (New enum added) and "Bestehendes Enum aktualisiert" (Existing enum updated). The "Vorgang durch Administrator prüfen" task leads to a decision gateway "Admin-Genehmigung liegt vor" (Admin approval present). If "Nein", it leads to an end event "Vorgang wurde nicht genehmigt" (Process not approved). If "Ja", it leads to the "Admin-Genehmigung benötigt" gateway.

4. Summary and Outlook

- **BPMN successfully used for managing metadata at KIT Energy Lab**
- **Enforcement of data quality** through process design
- Further process implementations planned
- Exchange with related projects
- Use for BPMN also for organizational aspects, e.g. data management plan
- **Seeking collaboration for future projects**



Thank you for your attention!

