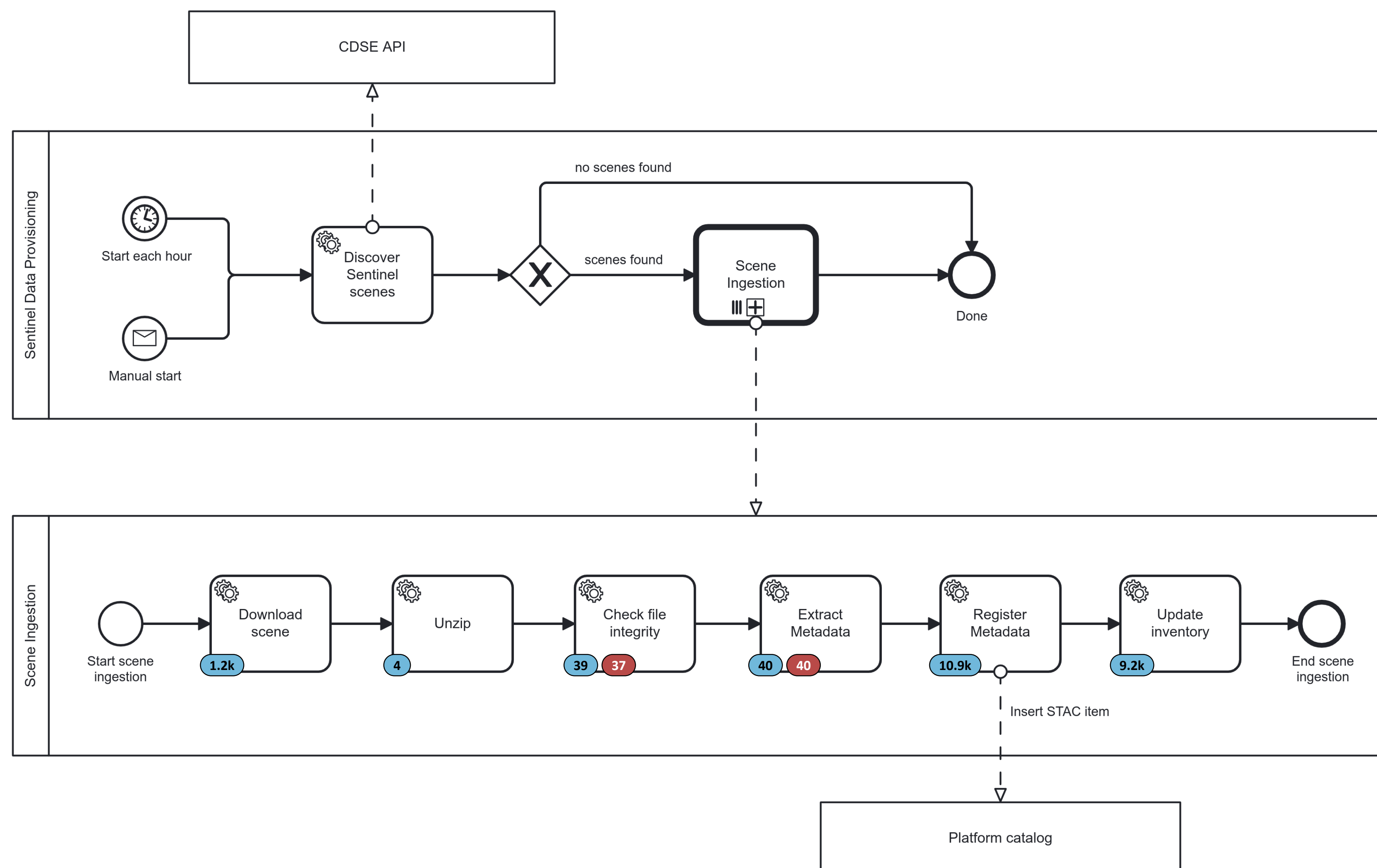


# Leveraging BPMN for Earth Observation Data Provisioning Workflows

Mario Winkler, Jonas Eberle, Jonas Müller  
German Aerospace Center, DLR



BPMN model representing the provisioning workflow for Sentinel data at DLR's *terrabyte* platform. For the operator, the current status of all running workflow instances is displayed directly in the diagram by the BPMN tooling.

## BPMN

The Business Process Model and Notation (BPMN) is a standardized graphical modeling language, widely used to define, document and implement processes. We choose BPMN to model our EO data workflows because in contrast to text-based approaches like CWL, it directly provides a visual framework which enhances the visibility of the process during the design, development and operational phases, easily bridging the gap between stakeholders, requirements, technical implementation and operational tasks.

### Basic Notation

Several core symbols are used to visually represent different elements of a business process. **Events** are depicted as circles and indicate something that happens, such as the start, end, or an intermediate occurrence in a process. **Activities** represent tasks or work that needs to be performed and are shown as rounded rectangles.

**Gateways**, symbolized by diamonds, control the flow of a process by branching, merging, or making decisions based on conditions. **Sequence flows** are solid arrows that connect these elements and show the order in which activities and events occur.

Additionally, **pools and lanes** are used to organize different participants or processes within the diagram, improving the readability.

## Common Use Cases and BPMN Patterns

This section provides an overview of BPMN constructs facilitate the implementation of EO data provisioning workflows for common use cases like

- On-demand, batch-oriented provisioning of historic data
- Continuous or near real-time provisioning of new data sets
- Continuous or on-demand consistency verification and ensuring completeness

### Continuous execution

A process can have a **timer start event** which represents a time cycle definition. When a process is deployed, a corresponding timer is scheduled and when its due, the process is executed automatically. This way a continuous data discovery and ingestion can be achieved, keeping up with new data at the source by dynamically calculating a search time window.

### Manual execution

A process can also be started by a **message**, identified by a unique name. When the BPMN engine receives a message, it can correlate it to the start event of the process if the name matches. Messages are encoded in JSON and

can contain input parameters to the workflow. This way the discovery step can be provided with a specific filter expression matching data e.g. published in a certain time interval or having specific criteria.

### Dynamic concurrency

BPMN makes it easy to model and coordinate many concurrent tasks with the concept of a **multi-instance activity**. This kind of activity is executed for each element of a given list (e.g. a set of STAC items resulting from a catalog search) sequentially or in parallel. Only when all input elements are completed, the activity is left and the workflow continues with the next step.

### Encapsulation, modularization and reuse

BPMN provides the possibility to encapsulate a sequence of tasks into a **subprocess** to modularize and build reusable components. In EO workflows, often the same sequence of tasks need to be applied repetitive to a data package (a single scene) or a metadata object (STAC item).

### Scalability with external tasks

Activities can be configured to be handled externally. When the BPMN engine encounters such activity, it provides this unit of work as a job, which can be polled and acquired by independent worker processes. The benefit of this pattern is that the entities performing the actual work are independent of the workflow engine. This allows the worker to be deployed separately and to be scaled to match the current work load. In our EO data workflows, all activities are defined as external tasks and implemented as individual Python packages.

## Conclusion

BPMN is well suited to implement the data provisioning workflows of EO data and are already successfully operated on the DLR's Earth Observation exploitation platform *terrabyte*. BPMN is also supported by the ESA EOEPKA+ project as part of the Resource Registration building block's Harvester component.

