

Creating Rich Metadata for Collaborative Research: Case Studies and Challenges

Christian Backe
DFKI Robotics Innovation Center Bremen
NFDI4Ing Conference 2023-09-27

Creating Rich Metadata for Collaborative Research: Case Studies and Challenges

1. Data management in RoBivaL and DeeperSense
2. Metadata workflow elements

1. Data management in RoBivaL and DeeperSense



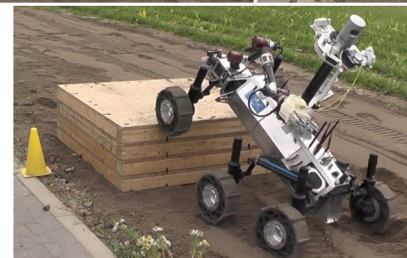
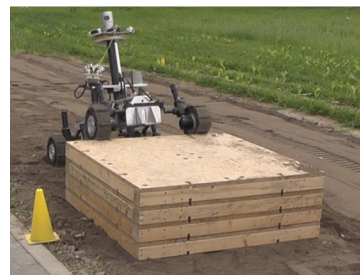
2 x 2 Robots

- Agriculture: BoniRob, NaioOz
- Space: ARTEMIS, SherpaTT



6 Experiments

- Obstacle avoidance
- Repeated rollover
- Sill crossing
- Straight travel
- Tensile force
- Turn around

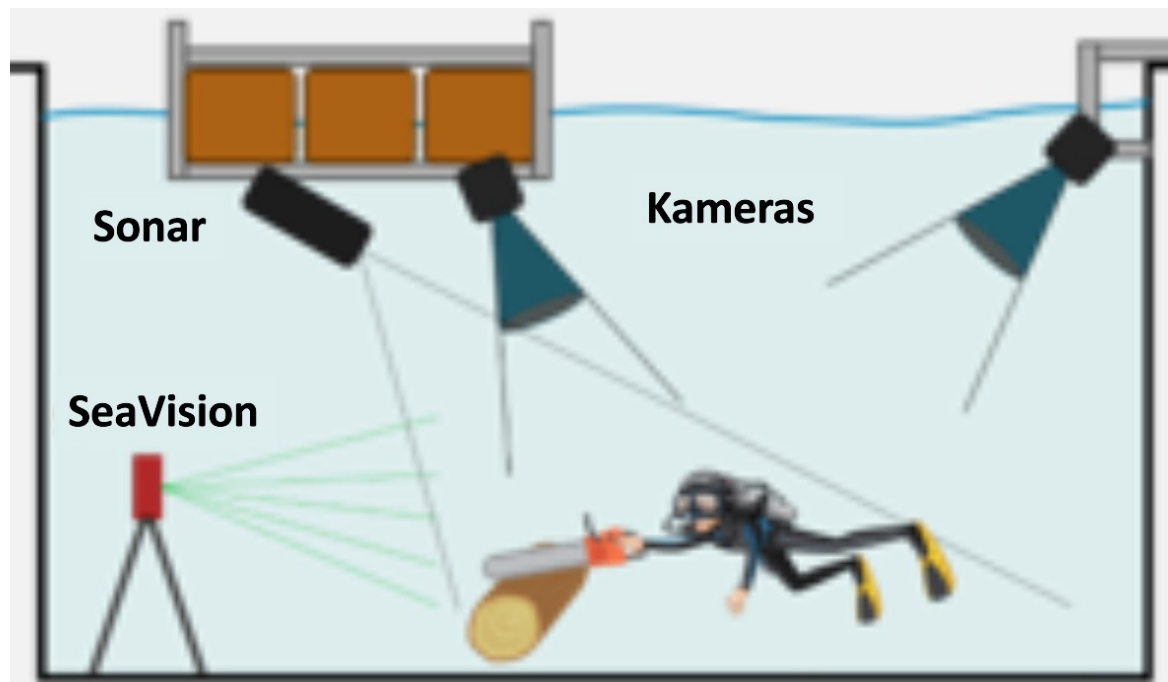


3 x 2 Soil conditions

- Moisture: Dry, Moist, Wet
- Density: Compacted, Tilled



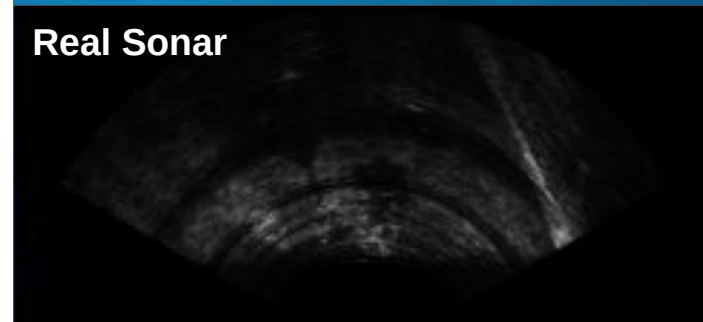
Improve diver monitoring with Sonar-to-Camera translation



Real Camera

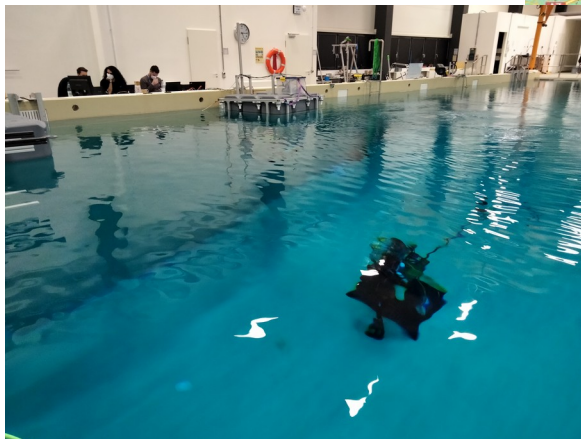


Real Sonar



Predicted from Sonar





Maritime Exploration Hall



Field Locations



Images: © DFKI, Bilal Wehbe / Christian Backe
Map: © Google / GeoBasis-DE/BKG



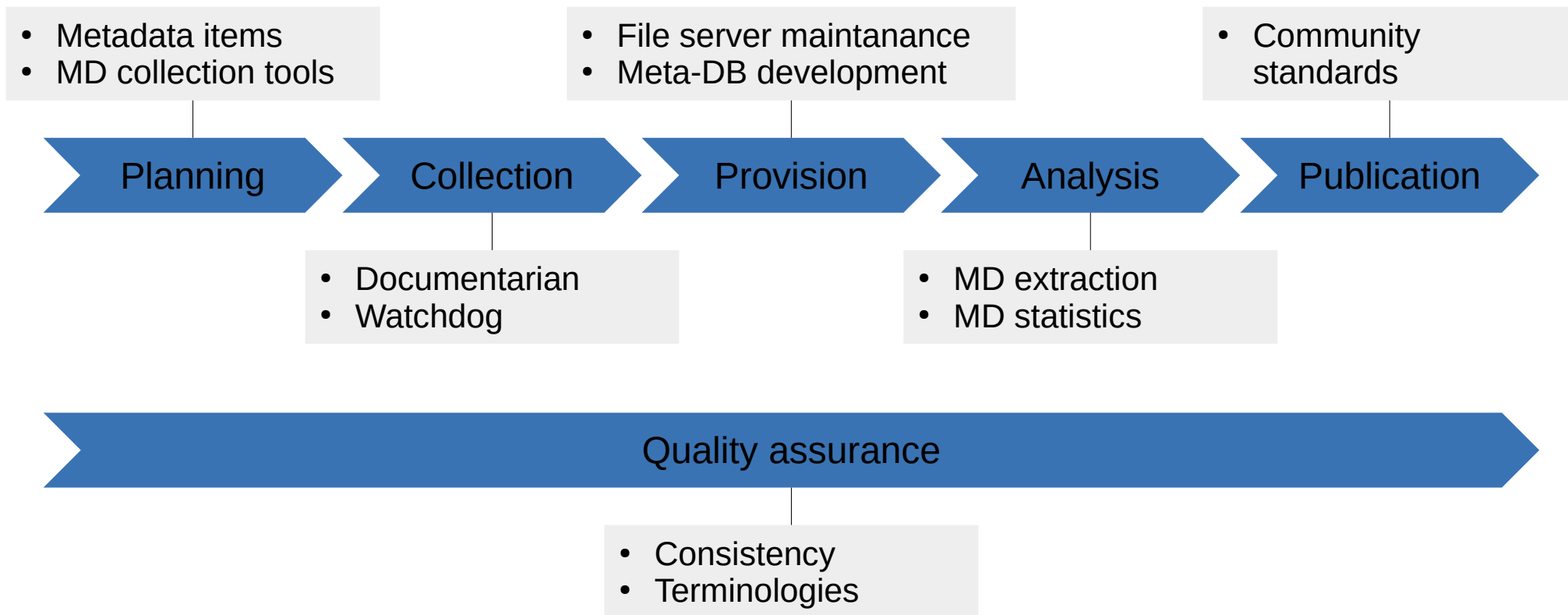
Tank Wash Basin

Starnberg Lake

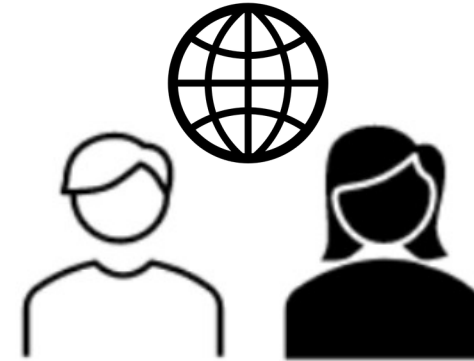
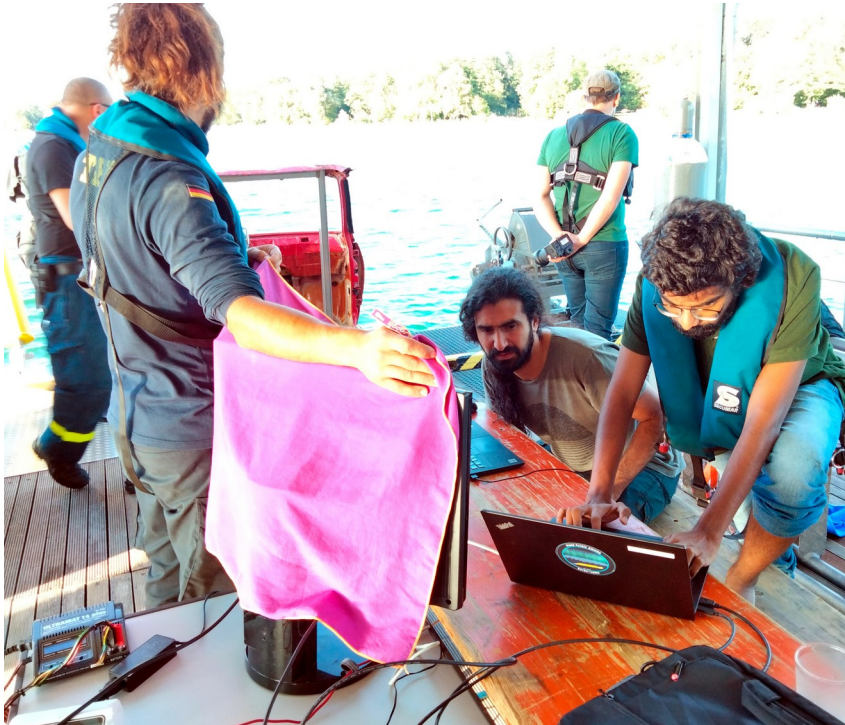


		
Data	GPS, IMU, Forces, Laserscan, Camera, Moisture meter, Penetrometer	Camera, Sonar
Metadata	Experiment parameters, Run interval, Robot description, Data type	Location description, Scene description, Sample time, Sensor configuration


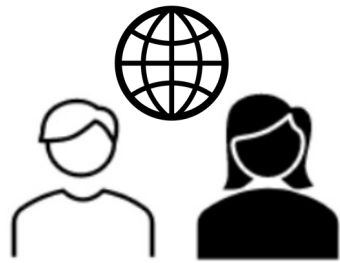
Data management tasks



Research team vs. Data re-users



Different requirements and practices

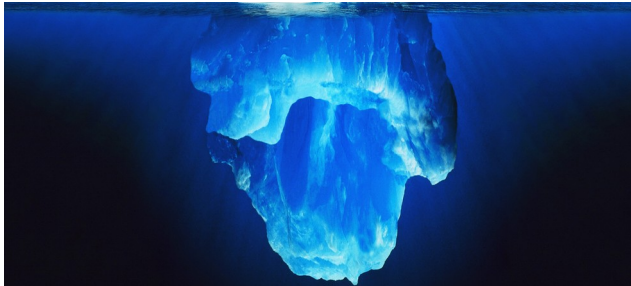

	
Common knowledge	Detailed, formal, explicit metadata
Distributed information	Coherent information
Ad-hoc communication	Terminologies, Ontologies

An iceberg floating in a blue ocean under a clear sky. The small tip of the iceberg is above the water line, while the much larger, more complex structure is submerged below. The water is a deep blue, and the sky is a lighter blue with some wispy clouds.

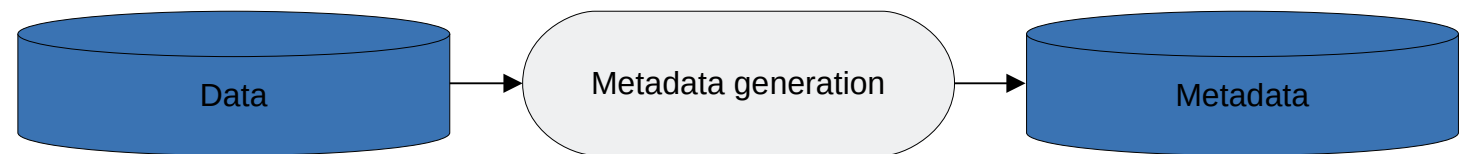
Consumer MD

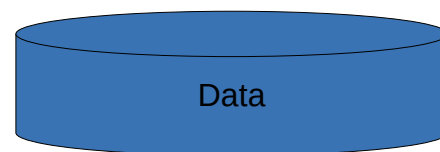
Precursor MD

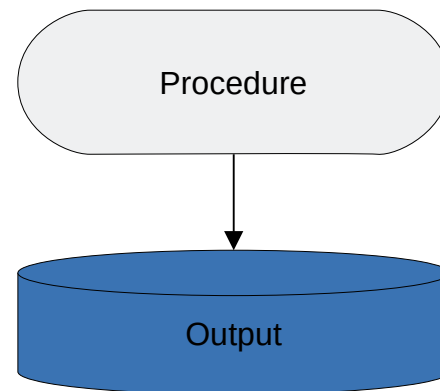
Additional effort

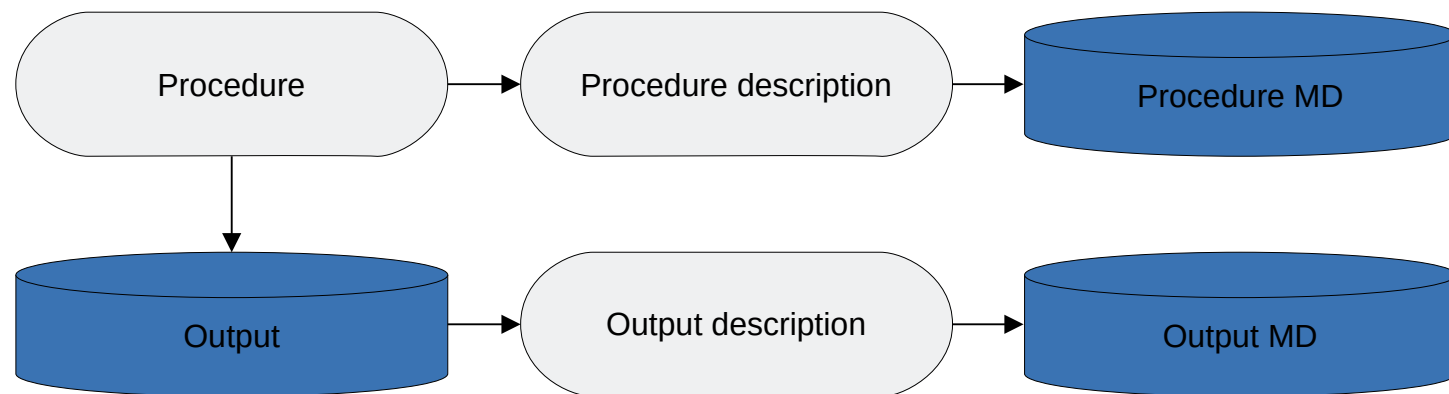
	
<ul style="list-style-type: none"> • Describe all processing phases • Describe actual and possible applications 	<ul style="list-style-type: none"> • Terminology / Ontology • Use case matching
<ul style="list-style-type: none"> • Data consistency checks • Error logs / handling 	<ul style="list-style-type: none"> • Data constraints • Data quality

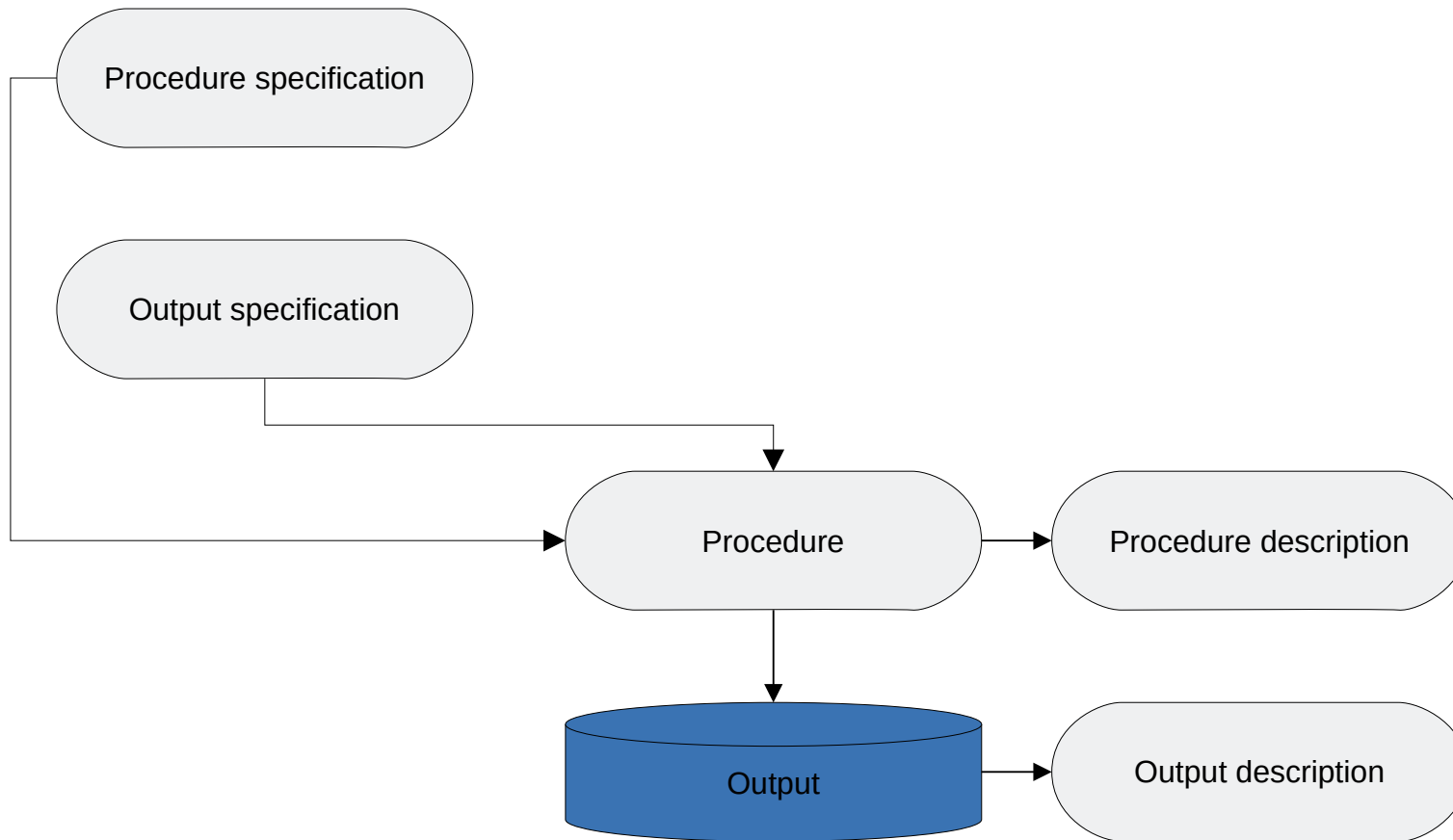
2. Metadata workflow elements

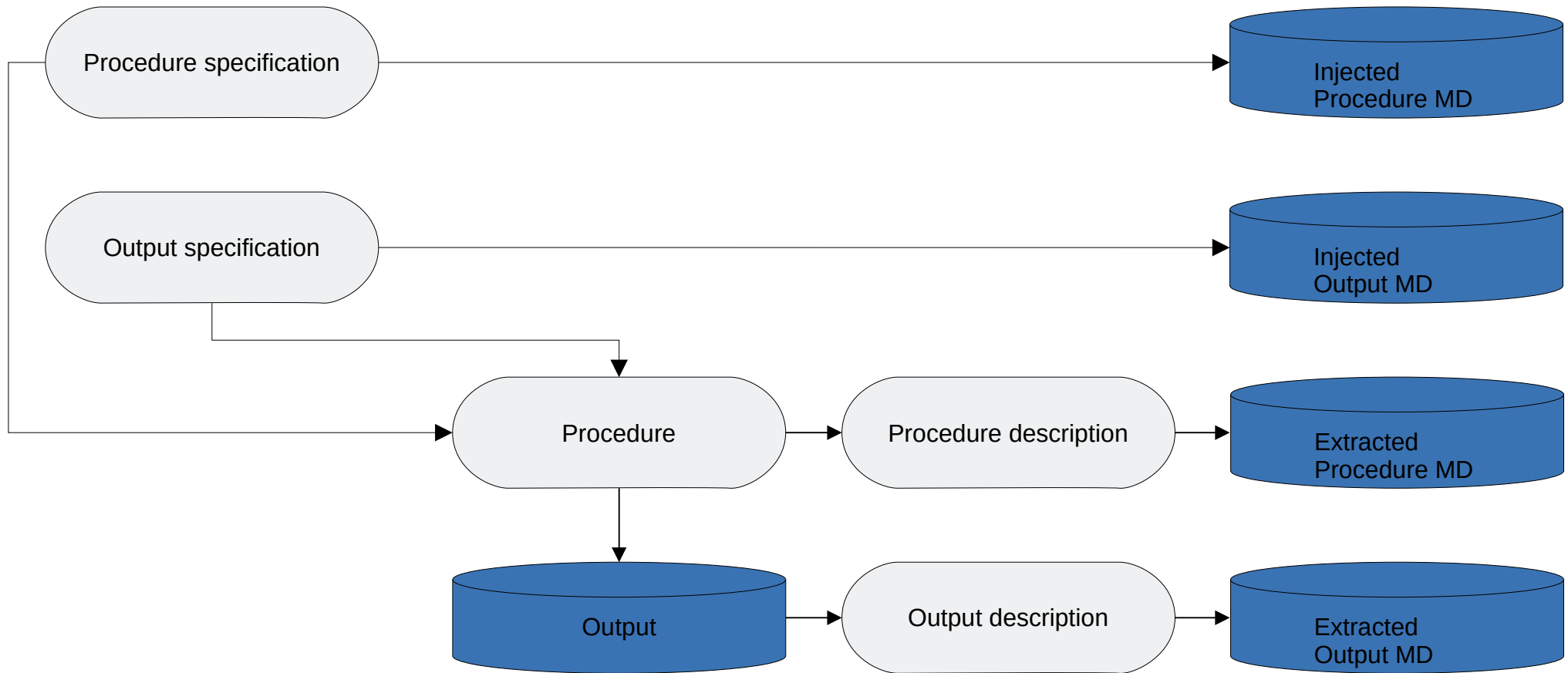




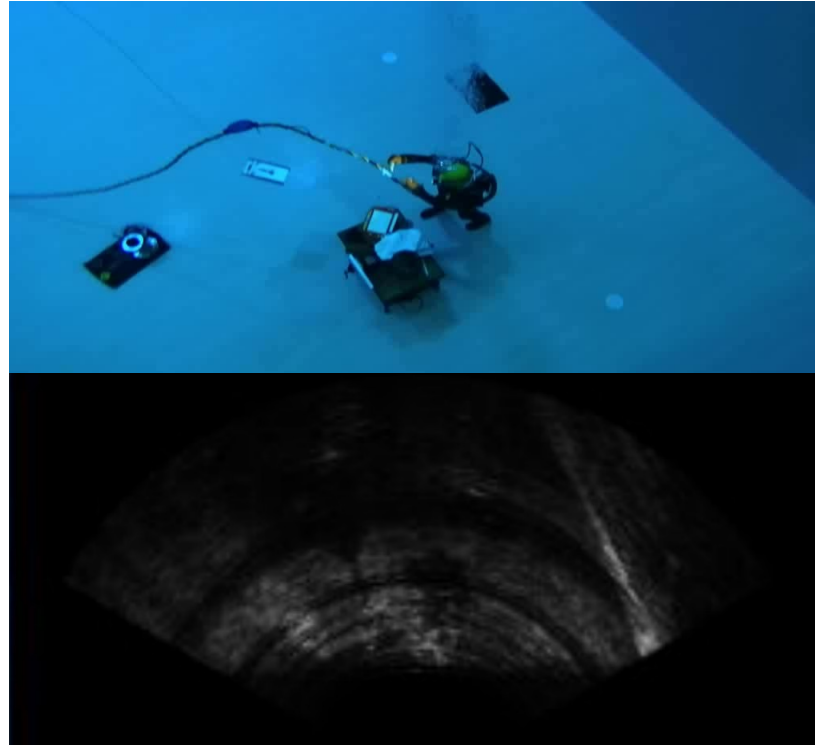








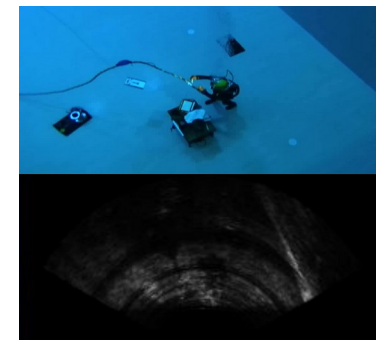
Procedure: Capture camera and sonar images of a diver



Procedure: Capture camera and sonar images of a diver

```
rosbag record --duration=10m --output-name=zed-2023-04-26-12-13-00 /camera/left_image /camera/right_image
rosbag record --duration=10m --output-name=oculus-2023-04-26-12-13-00 /sonar/image
```

Injected Procedure MD	Extracted Procedure MD
<ul style="list-style-type: none"> Middleware (ROS 2.0) Start time (Duration) 	<ul style="list-style-type: none"> Scene description Event documentation



Output: Logfiles with raw data

```
zed-2023-04-26-12-13-00.bag
oculus-2023-04-26-12-13-00.bag
```

Injected Output MD	Extracted Output MD
<ul style="list-style-type: none"> Raw data structure ("topic") File name (Identifier) 	<ul style="list-style-type: none"> Number of recorded samples ("messages") File size

Metadata4Ing “Processing Step”

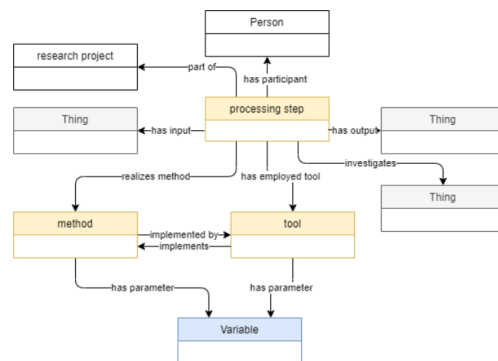
Applicable?

3. Metadata4Ing: Description

[back to ToC](#)

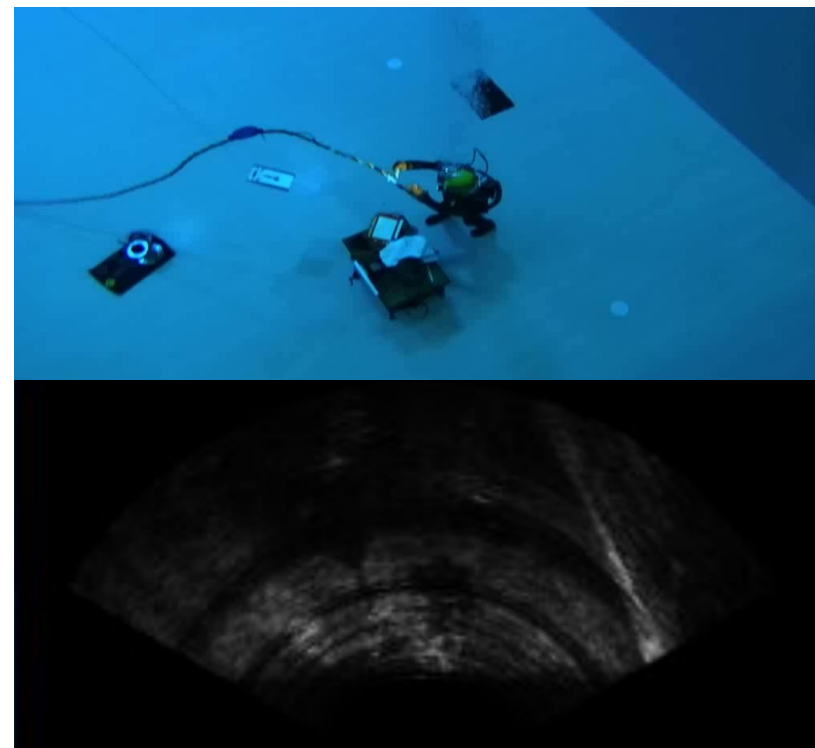
3.1. General process model

One of the main goals of the Metadata4Ing ontology is to enable researchers to document the provenance of data and material objects created or modified during research processes. Metadata4Ing accomplishes this with the help of a generalized process model centered around the class *processing_step*. The aforementioned data and material objects are described as output of the *processing_step*. Other relevant information like, e.g., the methods or tools used in a research process, are described in separate classes that can be linked to *processing_step*. A series of *processing_steps* can be used to represent complex research processes (cf. [Section "Specifying complex setups and processes by using composition"](#)). Metadata4Ing can therefore be seen as a system of building blocks that can be referred to by *processing_step* and in their totality enable a complete description of the provenance of a dataset or material object.



3.2. Processing step as central element

The central class of Metadata4Ing is *processing_step* that can be used to describe all kinds of processes and does not distinguish between different types of processing steps. The pattern formed by *processing_step* and its admissible object properties is therefore considered universal for all kinds of processes. A distinction between processes is achieved by adding further information as described in the following subsections.



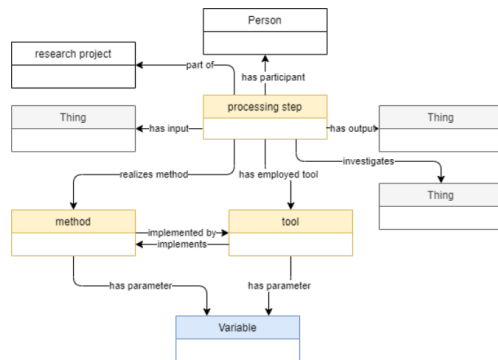
Metadata4Ing “Processing Step”

3. Metadata4Ing: Description

[back to ToC](#)

3.1. General process model

One of the main goals of the Metadata4Ing ontology is to enable researchers to document the provenance of data and material objects created or modified during research processes. Metadata4Ing accomplishes this with the help of a generalized process model centered around the class *processing_step*. The aforementioned data and material objects are described as output of the *processing_step*. Other relevant information like, e.g., the methods or tools used in a research process, are described in separate classes that can be linked to *processing_step*. A series of *processing_steps* can be used to represent complex research processes (cf. [Section "Specifying complex setups and processes by using composition"](#)). Metadata4Ing can therefore be seen as a system of building blocks that can be referred to by *processing_step* and in their totality enable a complete description of the provenance of a dataset or material object.



3.2. Processing step as central element

The central class of Metadata4Ing is *processing_step* that can be used to describe all kinds of processes and does not distinguish between different types of processing steps. The pattern formed by *processing_step* and its admissible object properties is therefore considered universal for all kinds of processes. A distinction between processes is achieved by adding further information as described in the following subsections.

Processing step attributes	
start time	...
end time	...
starts / ends with	...
has employed tool	...
has runtime assignment	...
investigates	...
realizes method	...
usage instruction	...
has input	...
has output	...

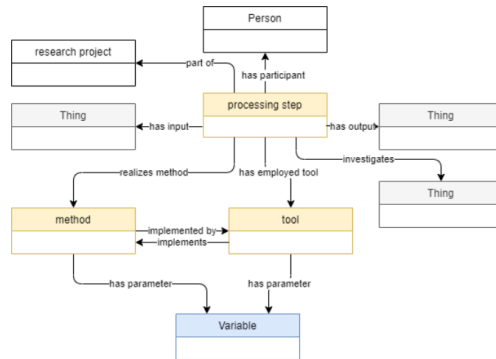
Metadata4Ing “Processing Step”: Capture camera and sonar images of a diver

3. Metadata4Ing: Description

[back to ToC](#)

3.1. General process model

One of the main goals of the Metadata4Ing ontology is to enable researchers to document the provenance of data and material objects created or modified during research processes. Metadata4Ing accomplishes this with the help of a generalized process model centered around the class *processing_step*. The aforementioned data and material objects are described as output of the *processing_step*. Other relevant information like, e.g., the methods or tools used in a research process, are described in separate classes that can be linked to *processing_step*. A series of *processing_steps* can be used to represent complex research processes (cf. [Section "Specifying complex setups and processes by using composition"](#)). Metadata4Ing can therefore be seen as a system of building blocks that can be referred to by *processing_step* and in their totality enable a complete description of the provenance of a dataset or material object.

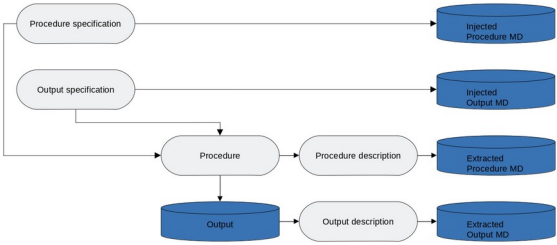
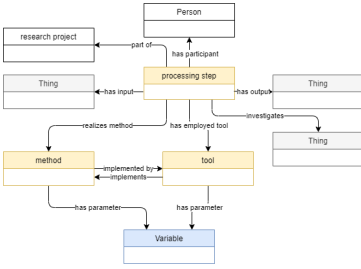


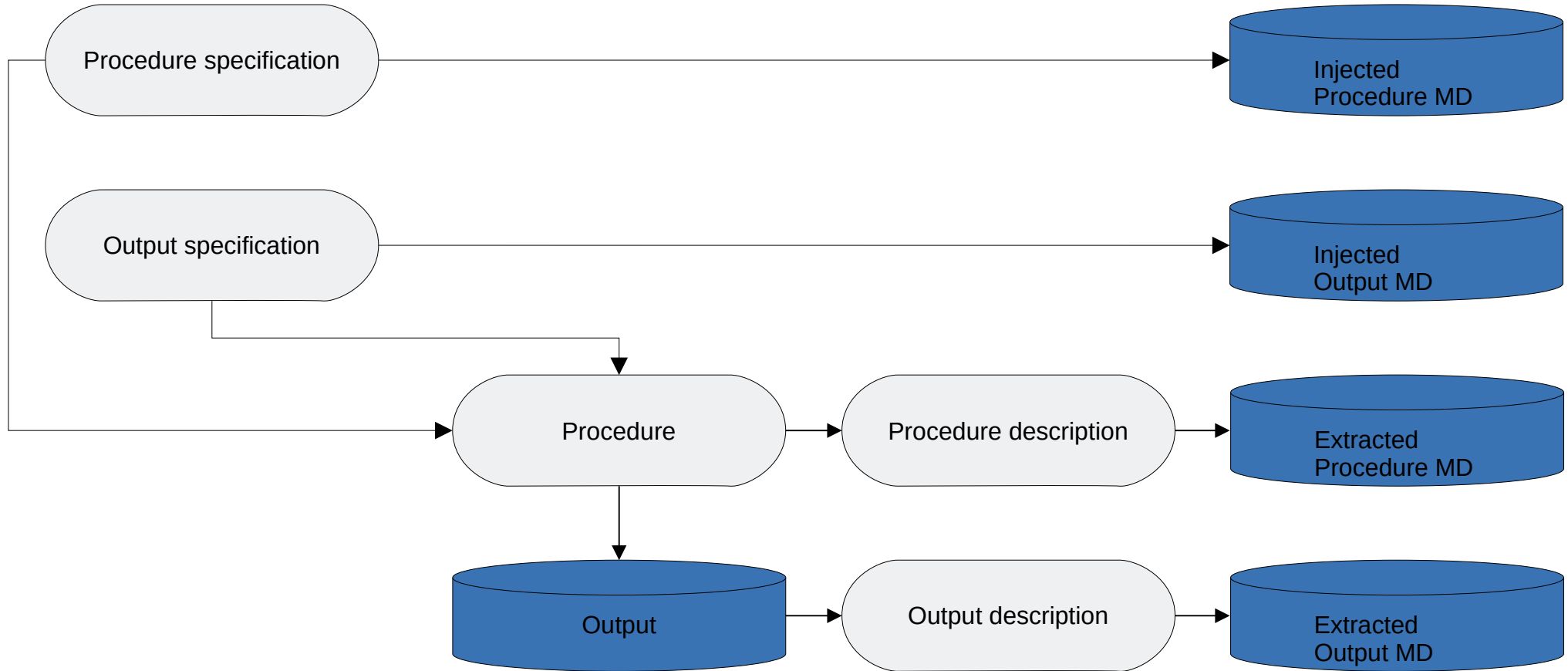
3.2. Processing step as central element

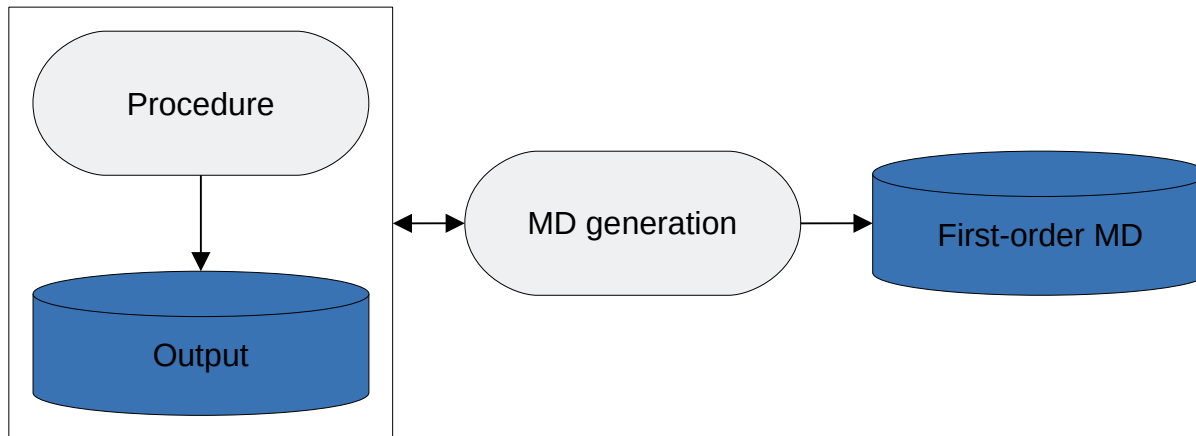
The central class of Metadata4Ing is *processing_step* that can be used to describe all kinds of processes and does not distinguish between different types of processing steps. The pattern formed by *processing_step* and its admissible object properties is therefore considered universal for all kinds of processes. A distinction between processes is achieved by adding further information as described in the following subsections.

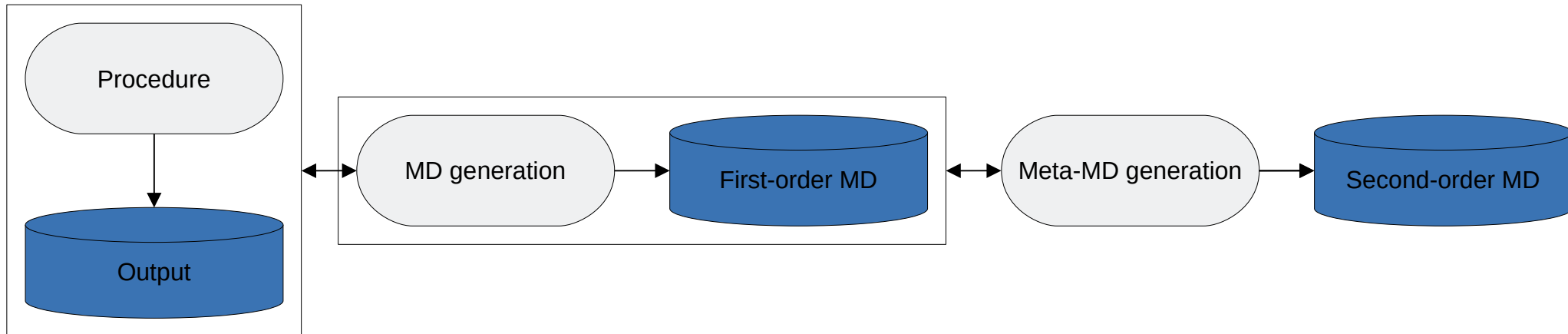
start time	2023-04-26T12:13:00+02:00
end time	2023-04-26T12:23:00+02:00
starts / ends with	? (i.e. substep)
has employed tool	ROS 2.0
has runtime assignment	<ul style="list-style-type: none"> duration=10m --output-name=zed-2023-04-26-12-13-00 /camera/left_image /camera/right_image
investigates	<ul style="list-style-type: none"> Technical divers under water? Sonar-to-camera translation?
realizes method	... ?
usage instruction	<ul style="list-style-type: none"> roslaunch record --duration=10m --output-name=zed-2023-04-26-12-13-00 /camera/left_image /camera/right_image
has input	None
has output	<ul style="list-style-type: none"> zed-2023-04-26-12-13-00.bag oculus-2023-04-26-12-13-00.bag

PO-IE vs. M4I

	
Production-/Workflow-oriented	Consumer-oriented
MD creation before/after data	MD communication at the end
Types of MD with any attributes	Fixed MD attributes







Examples of higher order metadata

Base data	Metadata	Meta-Metadata	Meta-Meta-Metadata
5.3 m/s	2023-09-27 09:37:51	%Y-%m-%d %H:%M:%S	ISO 8601
camera.mp4	metadata.json	schema.json	https://json-schema.org

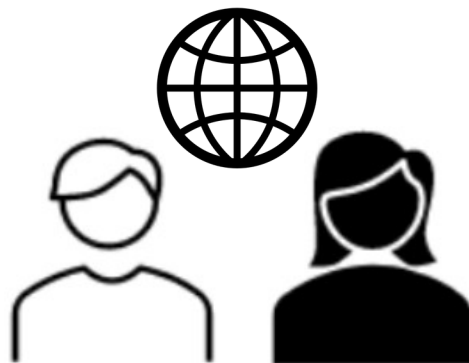
Different purposes of higher order metadata

Base data	Metadata	Meta-Metadata	Meta-Meta-Metadata
5.3 m/s	2023-09-27 09:37:51	%Y-%m-%d %H:%M:%S	ISO 8601
camera.mp4	metadata.json	schema.json	https://json-schema.org
<ul style="list-style-type: none"> Production for own primary purpose 	<ul style="list-style-type: none"> Precursor MD => Consumer MD 	<ul style="list-style-type: none"> Design choices 	<ul style="list-style-type: none"> Usage, without contribution



Different purposes of higher order metadata

Base data	Metadata	Meta-Metadata	Meta-Meta-Metadata
5.3 m/s	2023-09-27 09:37:51	%Y-%m-%d %H:%M:%S	ISO 8601
camera.mp4	metadata.json	schema.json	https://json-schema.org
<ul style="list-style-type: none"> “If they are not there, the data is not FAIR.” 			<ul style="list-style-type: none"> Reference is sufficient (required if possible)



Different purposes of higher order metadata

Base data	Metadata	Meta-Metadata	Meta-Meta-Metadata
5.3 m/s	2023-09-27 09:37:51	%Y-%m-%d %H:%M:%S	ISO 8601
camera.mp4	metadata.json	schema.json	https://json-schema.org
<ul style="list-style-type: none"> Use cases, Requirements 			<ul style="list-style-type: none"> Contribute + Lead



Takeaways

- Broad spectrum of use cases, even at a single institute
- Data managers are intermediaries between research teams, data re-users and RDM community
- Consumer-MD requires lots of Precursor-MD (“Iceberg”)
- MD creation needs different tools than MD communication (@M4I)
- MD creation is recursive (“Metadata is data”)
- How do we avoid “scope explosion”? How is (meta)data management different from general knowledge management?

Christian Backe
DFKI RIC Bremen
christian.backe@dfki.de
robotik.dfki-bremen.de

The author would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713.

This work was supported by the project RoBivaL, funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWE) under grant number 50RP2150.

This work was supported by the project DeeperSense that received funding from the European Commission. Program H2020-ICT-2020-2 ICT-47-2020 Project Number: 101016958.

The responsibility for the content of this presentation lies with the author.