Are my Data Sufficient?
On the Development of Quality Metrics

EGEOS GmbH: Jann Wendt, Daniel Wehner

Thanks for the support to the questionnaire and supply of data

Based on the delivered data several key challenges were identified:

- Structure and management of data

- Quality assessment
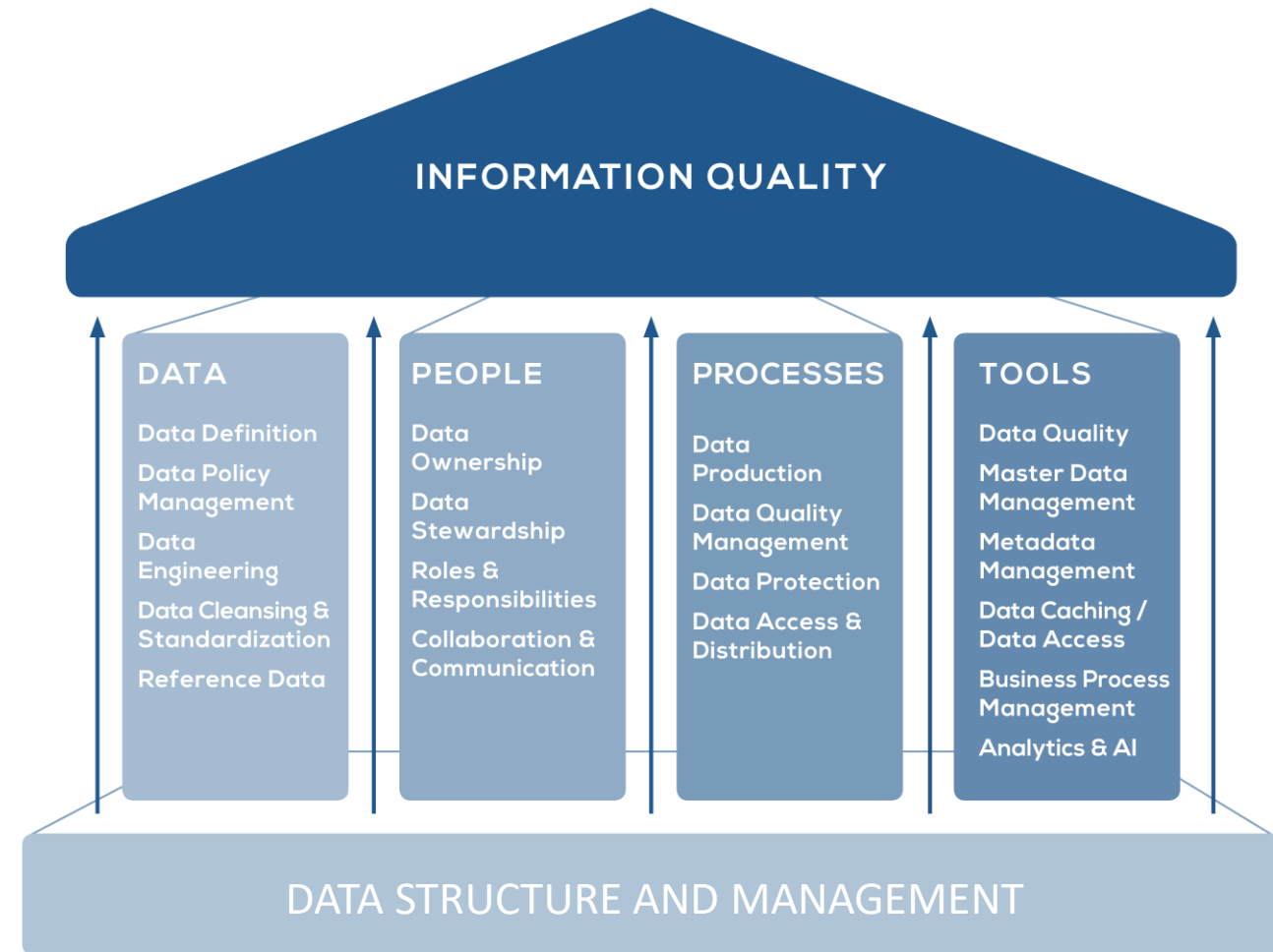
- Interpretation and analysis
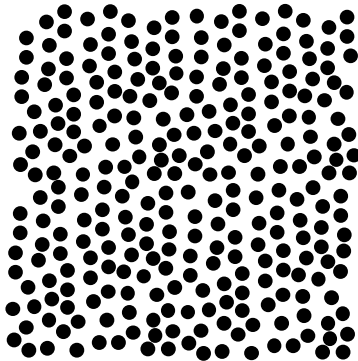
Data structure and management

- solid foundation enables all further steps

- several necessities to reach high a level of information quality

Why data quality metrics?

- Option 1: planning based on „standards"

- Option 2: review of measured datasets

INFORMATION QUALITY

DATA	PEOPLE	PROCESSES	TOOLS
Data Definition	Data Ownership	Data Production	Data Quality
Data Policy Management	Data Stewardship	Data Quality Management	Master Data Management
Data Engineering	Roles & Responsibilities	Data Protection	Metadata Management
Data Cleansing & Standardization	Collaboration & Communication	Data Access & Distribution	Data Caching / Data Access
Reference Data			Business Process Management
			Analytics & AI

DATA STRUCTURE AND MANAGEMENT

**Point cloud
(raw or processed data)**

**Raster
(grid or image data)**

- entire dataset

- full information

- large data sizes

- **required to compute data quality parameter**
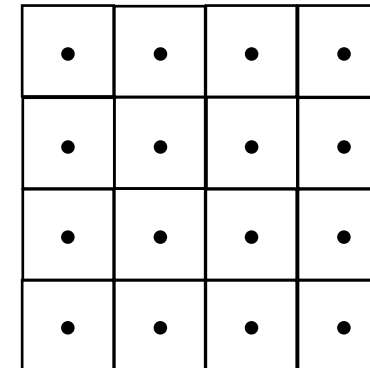
- reduced data

- information loss (e.g. color values in images)
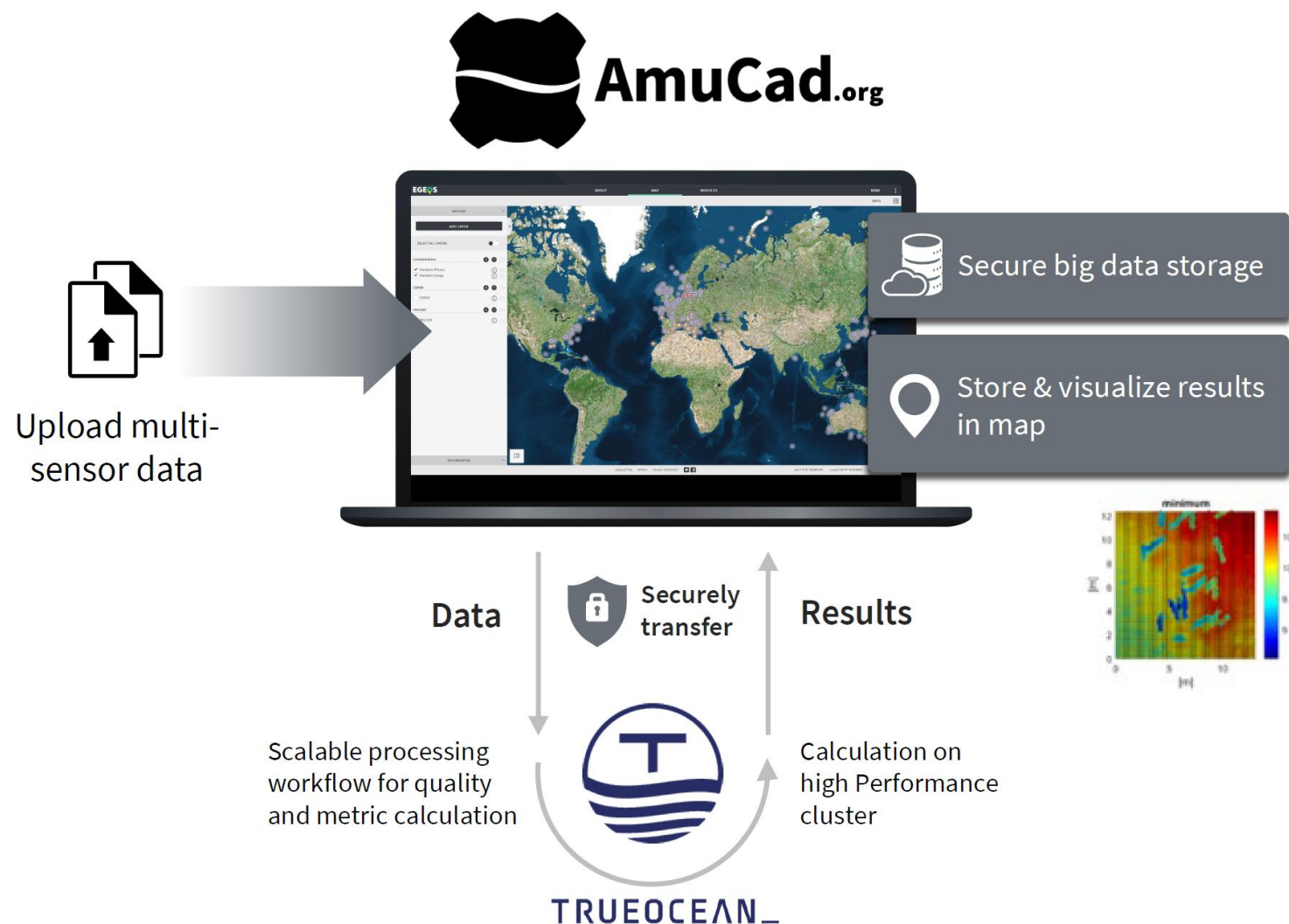
- reduced data sizes

Software-as-a-Service-Platforms

The software developed in the BASTA project will be integrated into the Ammunition Cadastre Sea platform (AmuCad.org). For the underlying large-scale data processing the technologies of TrueOcean will be used to enable:

- Data management and storage

- Data processing

- Data interpretation

- Data visualization

AmuCad.org under: https://www.AmuCad.org

TrueOcean under: https://www.trueocean.io

AmuCad.org

Upload multi-sensor data

Secure big data storage

Store & visualize results in map

Data   Securely transfer   Results

Scalable processing workflow for quality and metric calculation

Calculation on high Performance cluster

TRUEOCEAN_

EGEOS

| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Challenges

Data and metadata standards

- FAIR principles are suggested

Data upload and management concepts

Big Data
- Storage solutions
- Processing solutions

Heterogeneity
- variety of measurement systems
- multiple file types

GO FAIR

FAIR Principles

https://www.go-fair.org/

- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

EGEOS

| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Target Definition

- Defining the target object for determination of quality parameters with e.g.:
  - size
  - burial depth
  - magnetic properties

Data Upload | Data Storage | Preprocessing | Target | **Quality metrics** | Postprocessing | Artificial Intelligence | Visualization

Qualitative ⟶ **Quantitative**
(subjective)　　　　　　　　　　　　　　　　(objective)

Questionnaire to experts

Future Workshops

=

Updated Guideline

AmuCad.org

Software

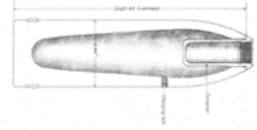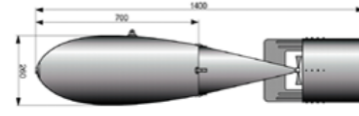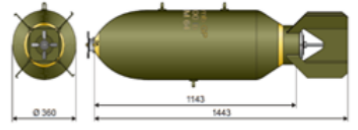| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Questionnaire - Infos

- Four different types/sizes of munition
  - 155mm shell,
  - GP250,
  - GP500,
  - Bottom Mine MK VI
- 41 parameters were defined for five sensors:
  - Side Scan Sonar (9),
  - Multibeam Echosounder (8),
  - Magnetics (14),
  - Sub-bottom Profiler (10),
  - [Electromagnetics (4)]

Reference Ob	
155mm shell BL Mark VII (nation: GB)	GP 250 lb MK IV (nation: GB)
0.15	0.26
34.9	82
10.5	30
41, 12, 5, 2.5, 2, 1, 0.5	75, 23, 10, 5, 3, 1, 0.5
0.4 -- 2	6 -- 20
1	2

bject	
GP 500 lb M64 (nation: USA)	Bottom mine, Influence, MK VI (nation: GB)
0.36	0.5
158.9	385
68.1	431
300, 91, 40, 20, 12, 5, 2.5	1900, 725, 350, 195, 120, 55, 30
8 -- 80	200 -- 1200
2	3

available under: https://www.basta-munition.eu/software

| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Questionnaire - Evaluation

- 10 Institutions participated until now

- Responses:
  - 66.8 % with yes
  - 17.6 % too small values
  - 16.6 % too large values

- 12 of 41 parameters with total consensus (answers: yes)

- 14 of 41 parameter with no consensus (answers: yes, too small, too large)

- **Next steps:** workshops on specified survey methods (spring 2021)
  - people who filled out the questionnaire will be contacted by BASTA, but everyone is welcome

available under: https://www.basta-munition.eu/software

| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Data Point Spacing
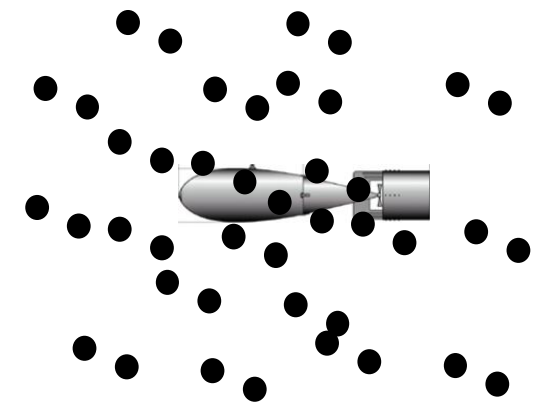
MBES
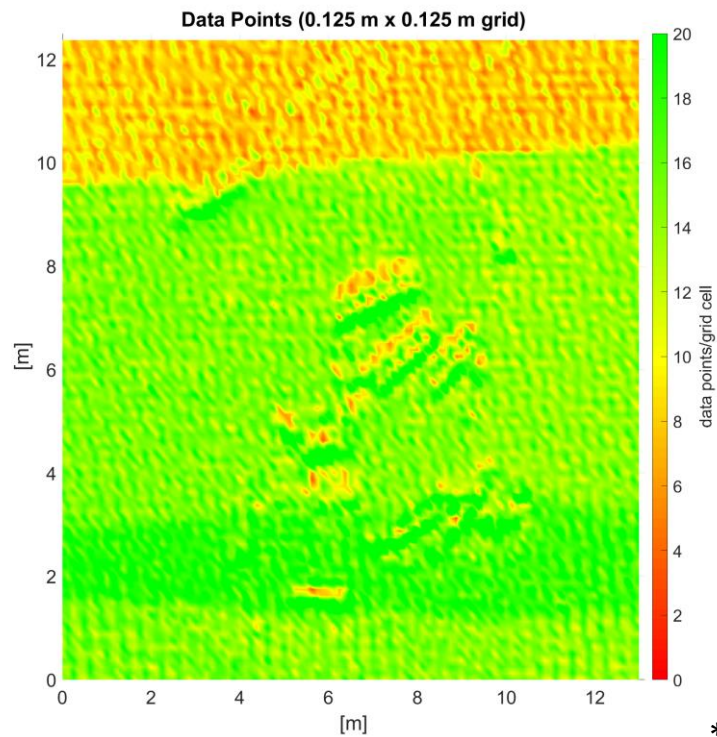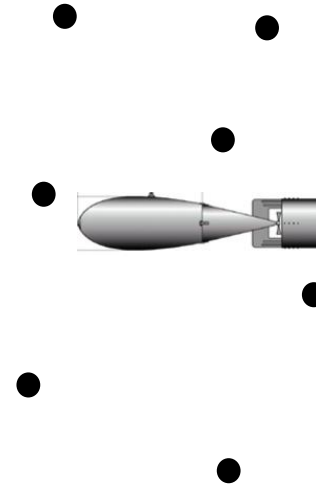
Data Points (0.125 m x 0.125 m grid)

bad

good

*arbitrary colorscale

EGEOS

Footprint

MBES

bad

good

Footprint - mean (0.125 m x 0.125 m grid)

beam footprint

40 m water depth

beam footprint

10 m water depth

*arbitrary colorscale
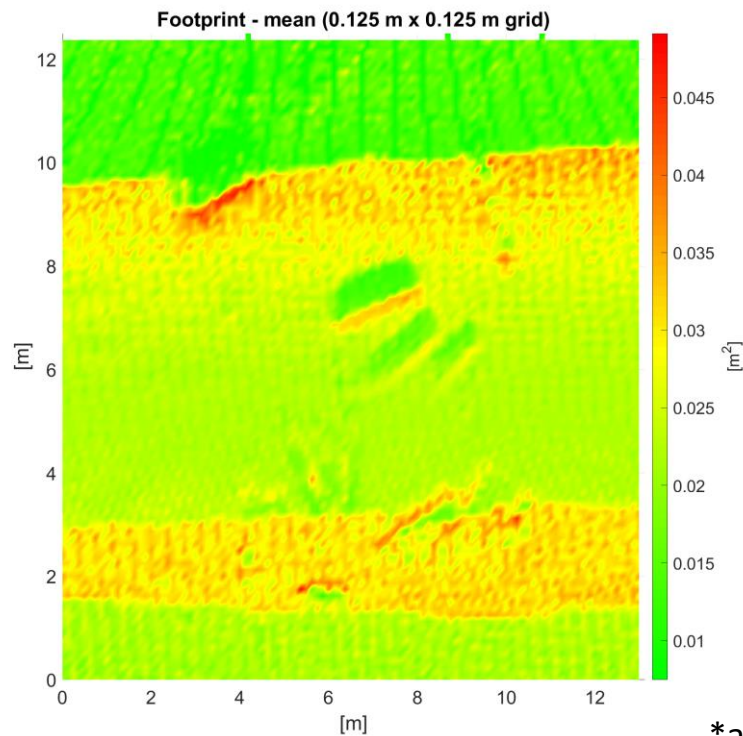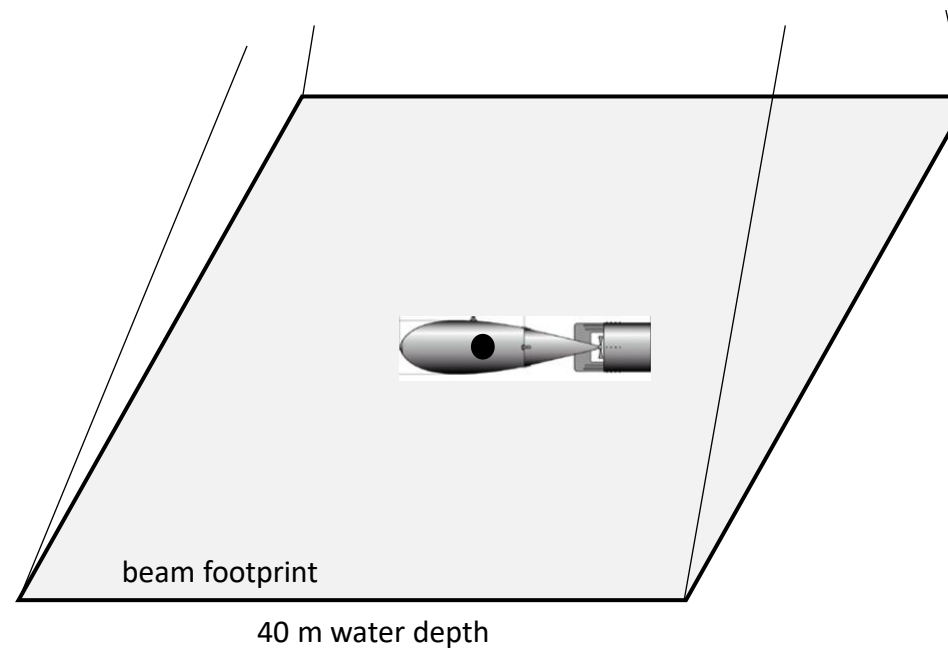
| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

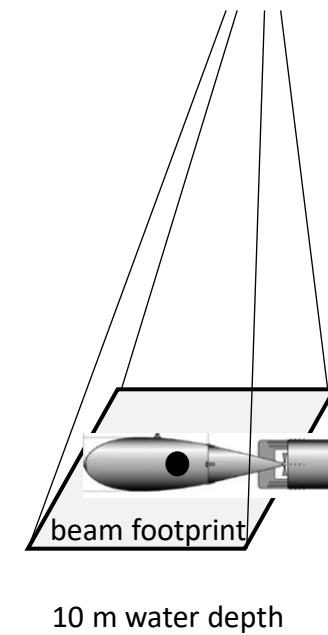## Horizontal Uncertainty

MBES

Uncertainty horizontal (0.125 m x 0.125 m grid)

*arbitrary colorscale
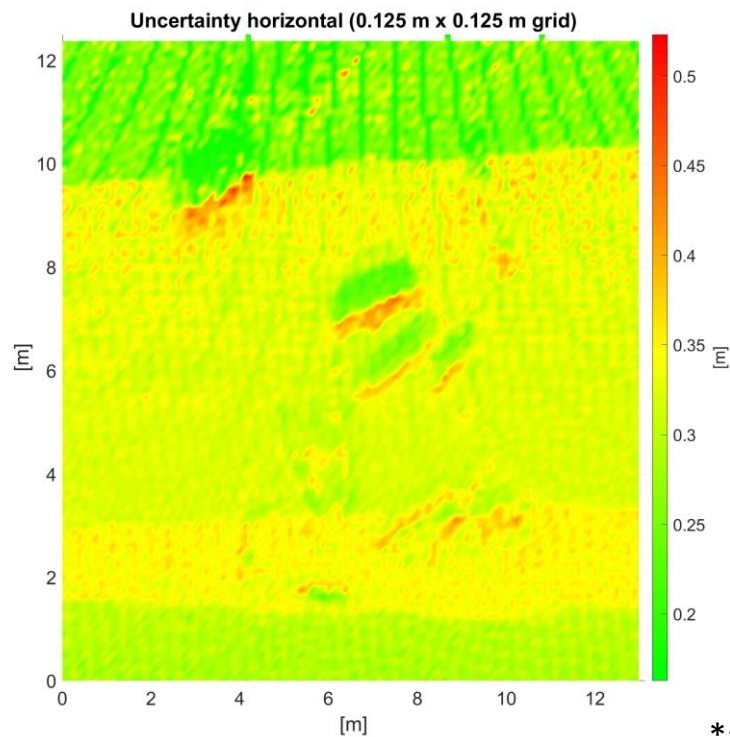
bad
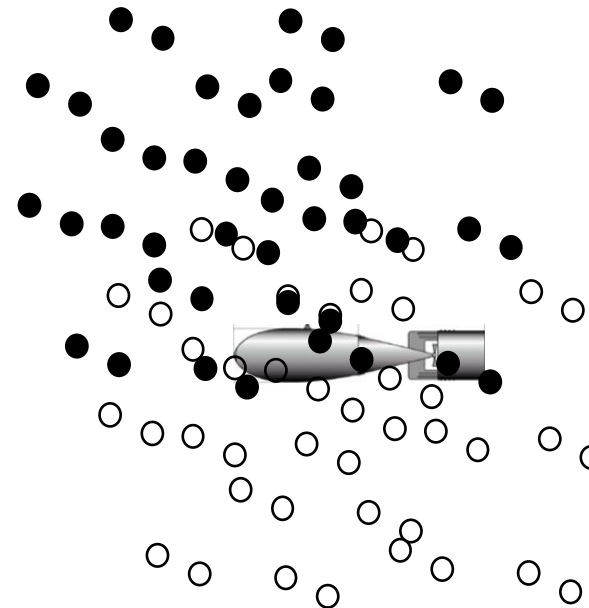
good

● measured

○ true position

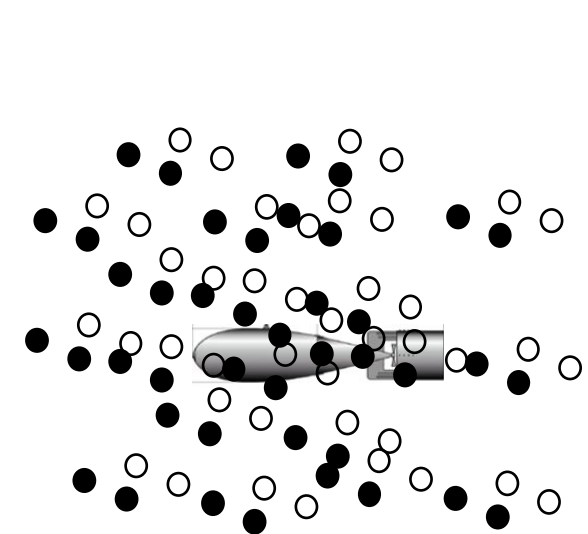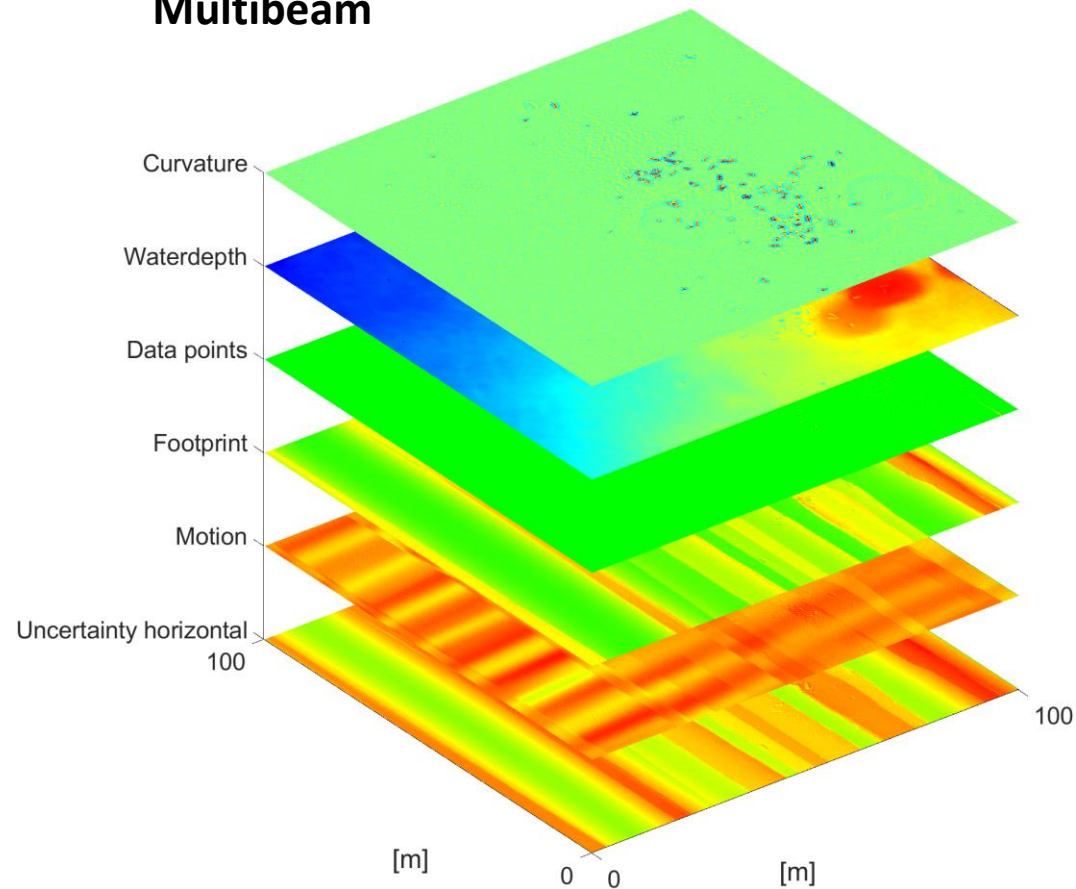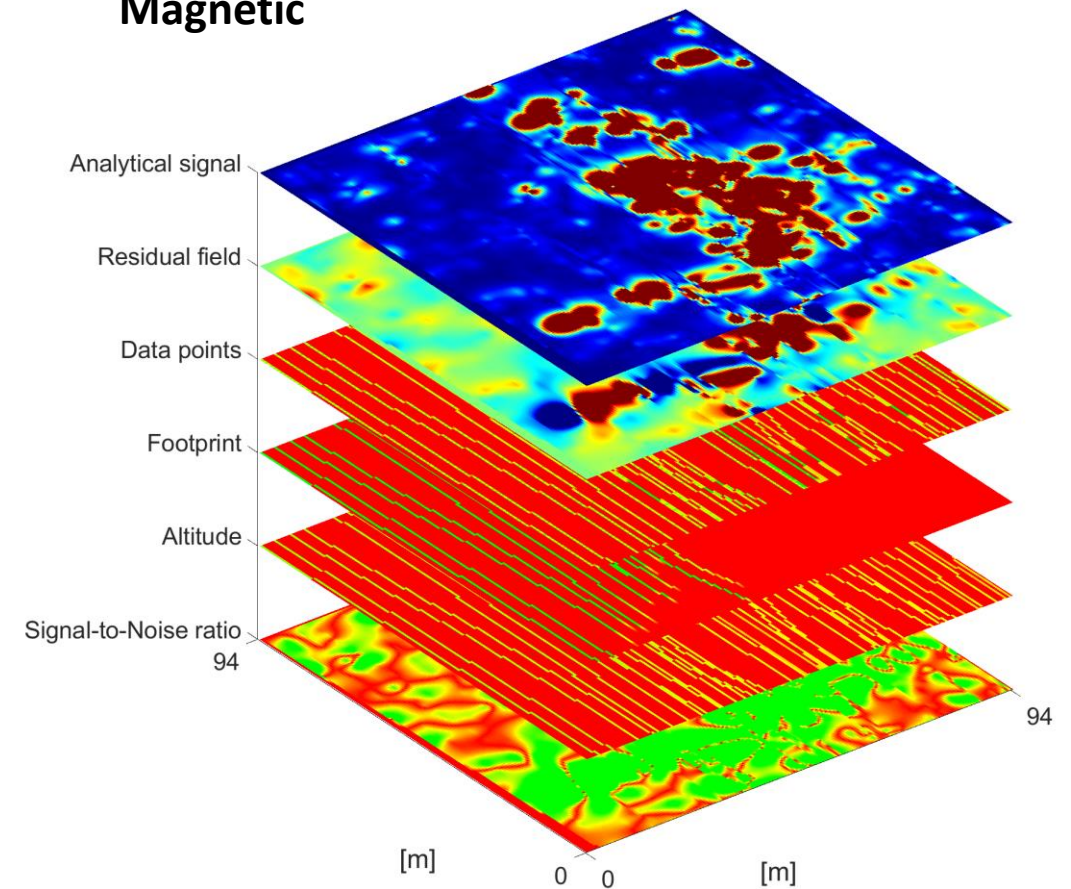Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization

Multibeam

Magnetic

*arbitrary colorscale
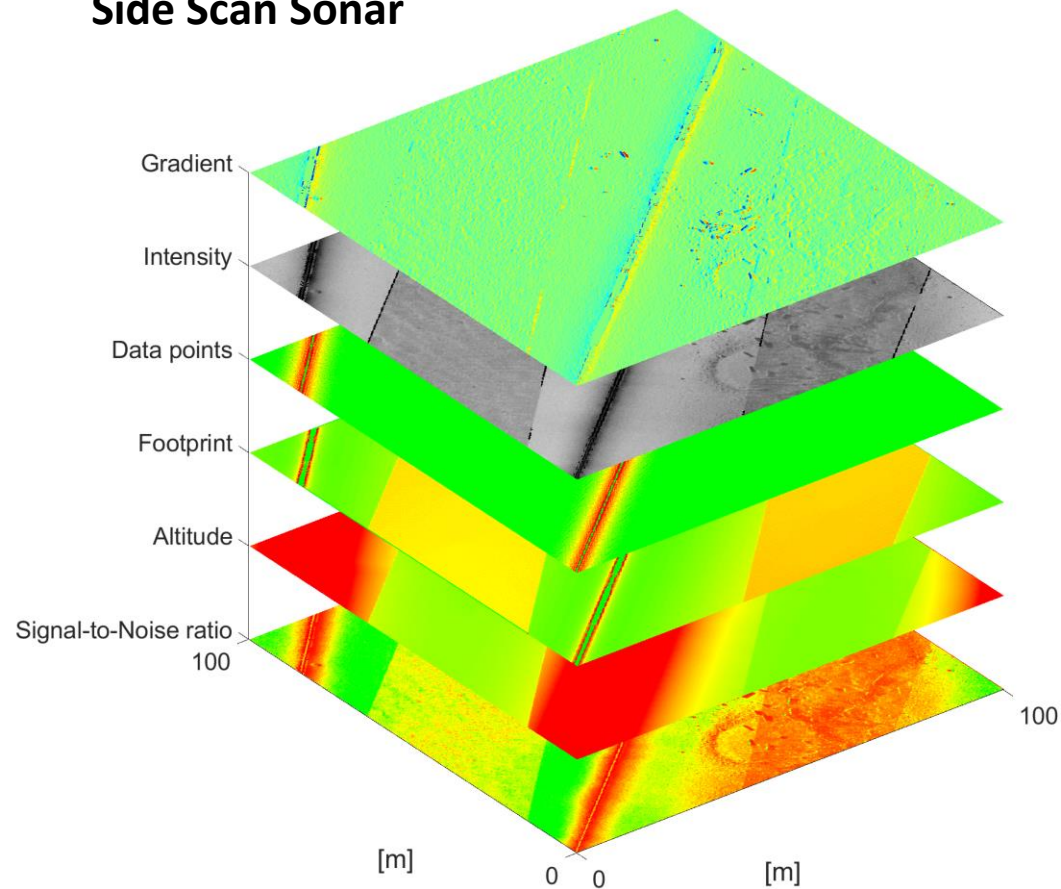
Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization

Side Scan Sonar

Subbottom Profiler

*arbitrary colorscale

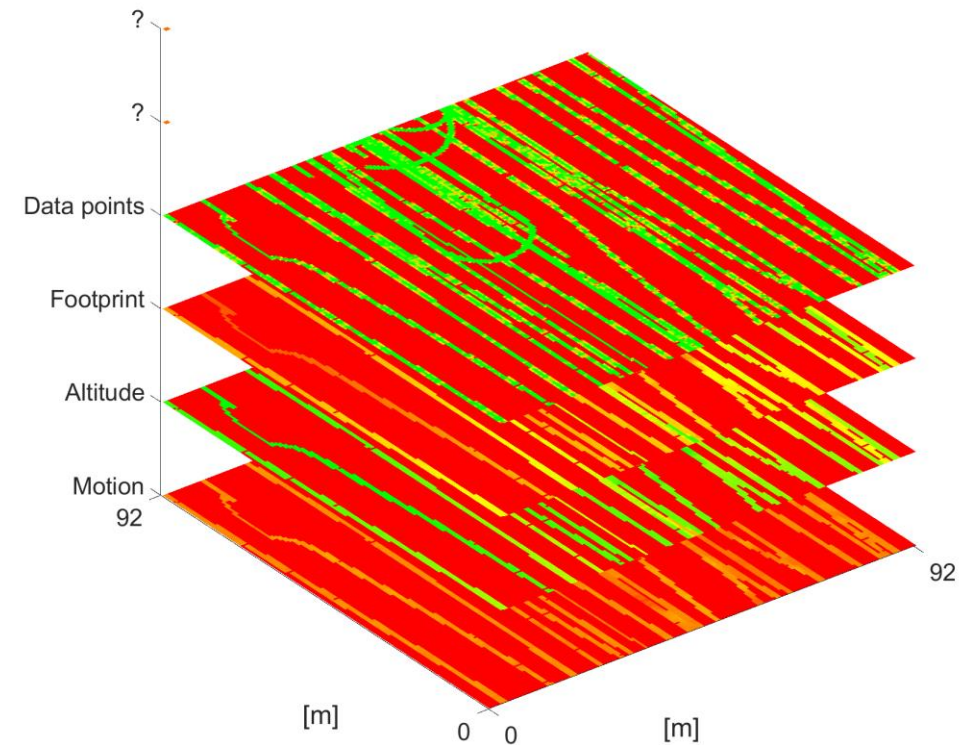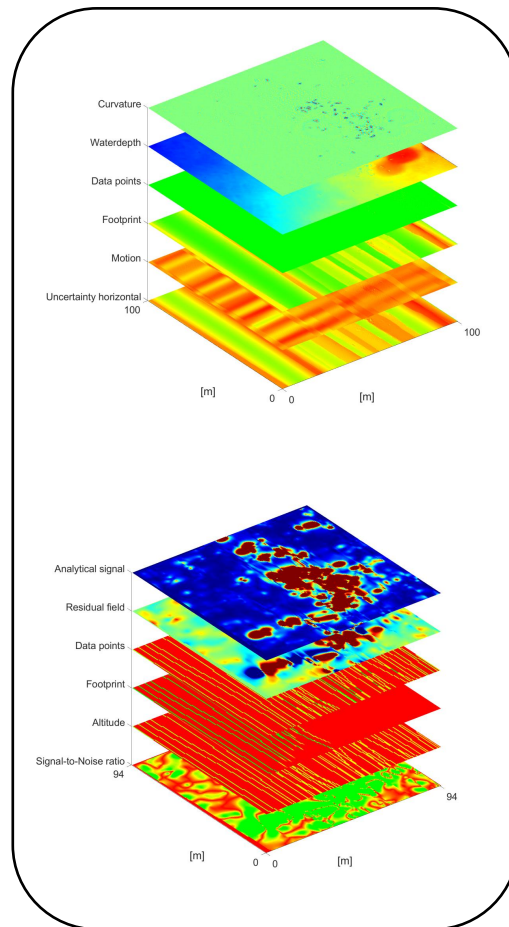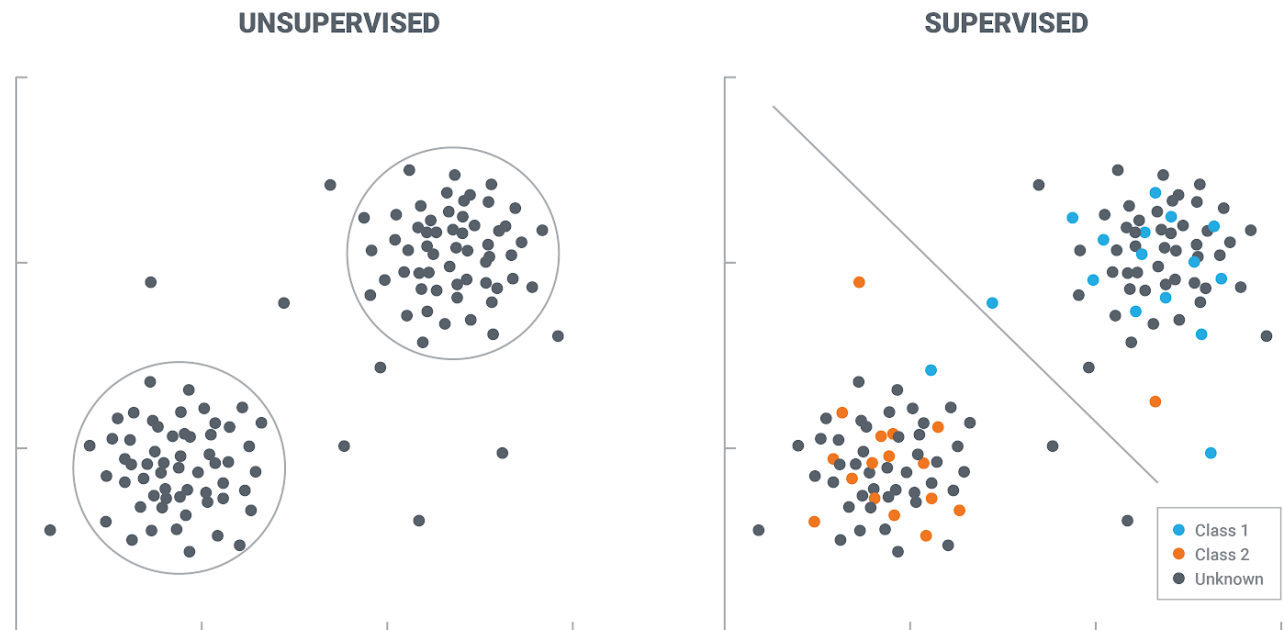| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Verified target list (data labels) of high importance for success

UNSUPERVISED

SUPERVISED

Class 1
Class 2
Unknown

Remember: garbage in, garbage out!

Data Pipeline

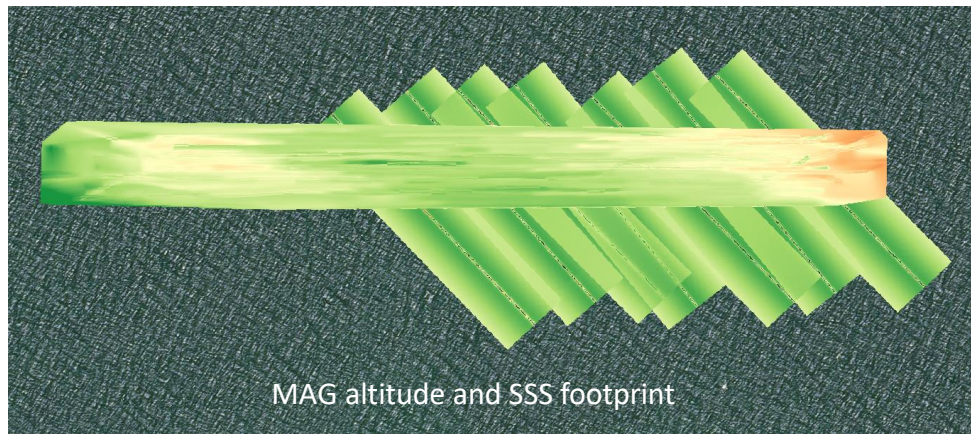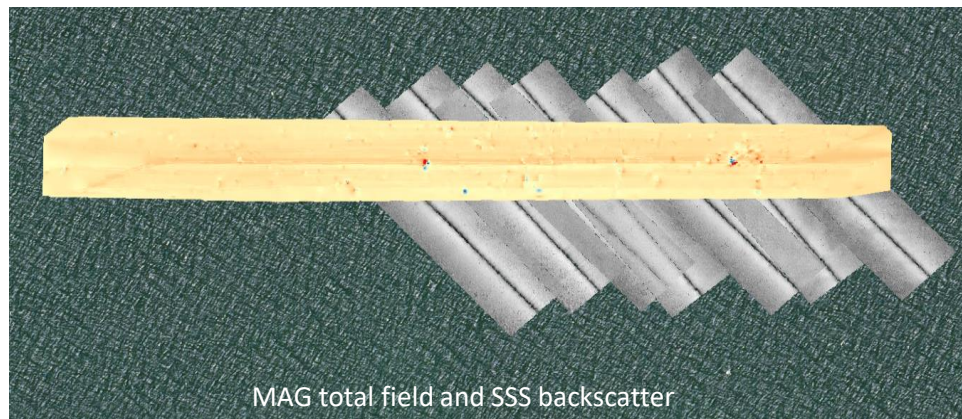| Data Upload | Data Storage | Preprocessing | Target | Quality metrics | Postprocessing | Artificial Intelligence | Visualization |

Quality Parameters

MAG altitude and SSS footprint

Data

MAG total field and SSS backscatter

EGEOS
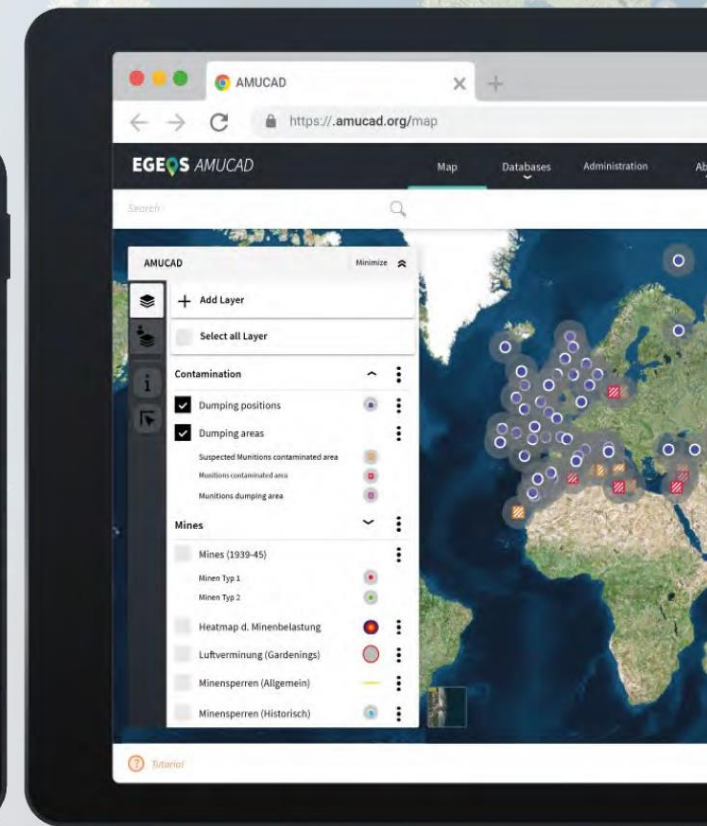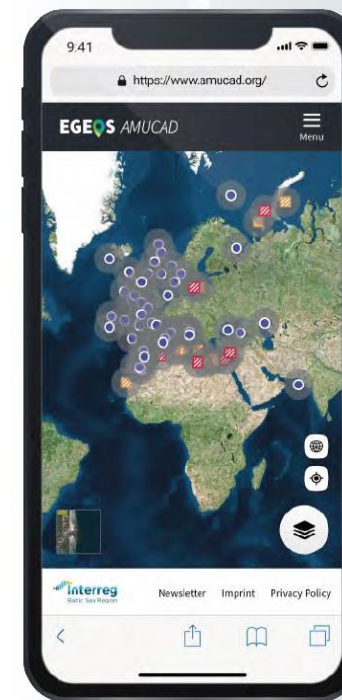
Next steps (short-term):

- finishing the implementation of Big Data processing pipelines for calculation of quality measures

- starting first real-world tests of quality measures on ongoing and/or finished infrastructure projects

Next steps (medium-term):

- conception of the User-Interface (workshops)

- development of analytical algorithms/AI

- technical implementation of all functionalities into the AmuCad.org platform