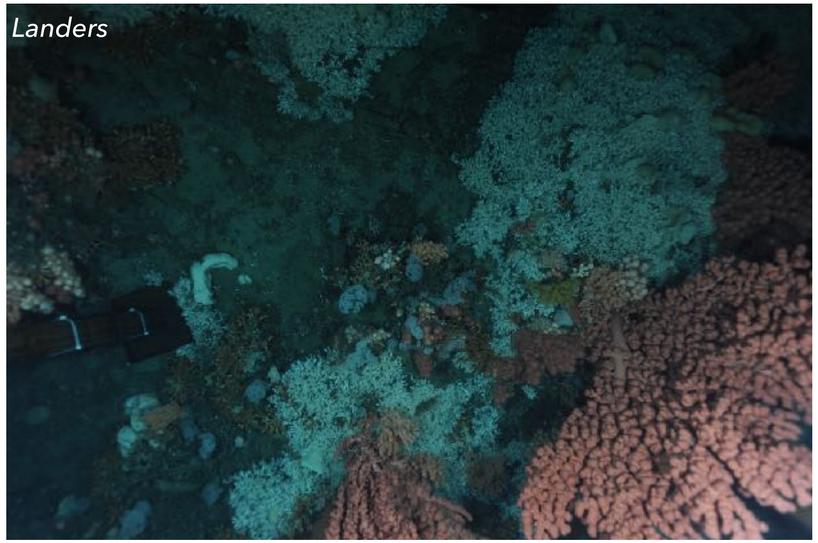
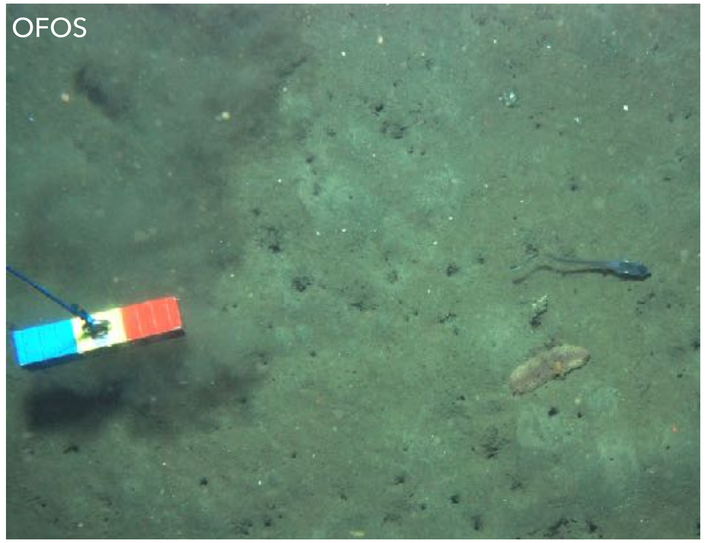

# Making marine image data FAIR with iFDOs

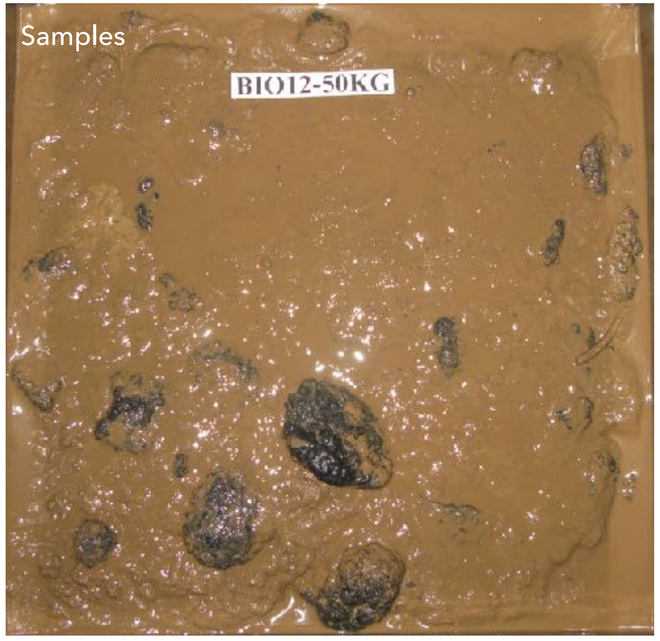
Timm Schoening

# Marine image data



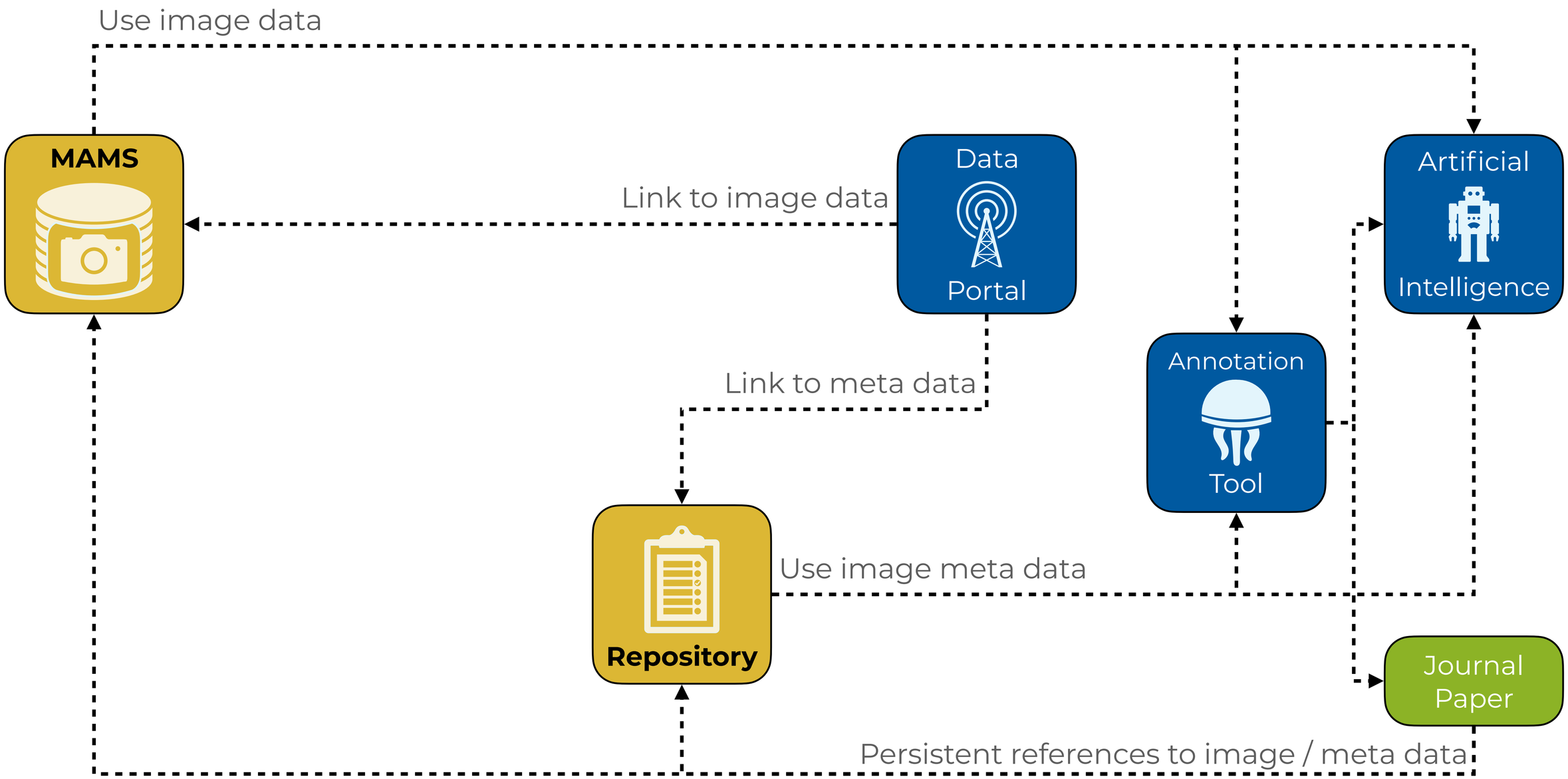



# Marine image data management

**State-of-the-art marine image “data management”**

- Data transfer from sea to shore happens by the scientists that have no incentive to upload / register data centrally
- Mobile hard disks are the sad state-of-the-art for media storage, mobility of scientists can lead to data disappearance
- Metadata standards for images and workflows that link data and existing tools are generally lacking
- As image data is not published alongside scientific articles, it can neither be reused nor the initial study be repeated

# What do we need?



# What is FAIR?

# FAIR

# Findable, Accessible, Interoperable, Reusable

## The FAIR Guiding Principles

**To be Findable:**

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

https://doi.org/10.1038/sdata.2016.18

# What is FAIR?

| FAIR principle | FAIR Metrics ID | Indicator | Priority |
|---|---|---|---|
| F1 | RDA-F1-01M | Metadata is identified by a persistend identifier | Essential |
| F1 | RDA-F1-01D | Data is identified by a persistent identifier | Essential |
| F1 | RDA-F1-02M | Metadata is identified by a globally unique identifier | Essential |
| F1 | RDA-F1-02D | Data is identified by a globally unique identifier | Essential |
| F2 | RDA-F2-01M | Rich metadata is provided to allow discovery | Essential |
| F3 | RDA-F3-01M | Metadata includes the indentifier for the data | Essential |
| F4 | RDA-F4-01M | Metadata is offered in such a way that it can be harvested and indexed | Essential |
| A1 | RDA-A1-01M | Metadata contains information to enable the user to get access to the data | Important |
| A1 | RDA-A1-02M | Metadata can be accessed manually | Essential |
| A1 | RDA-A1-02D | Data can be accessed manually | Essential |
| A1 | RDA-A1-03M | Metadata identifier resolves to a metadata record | Essential |
| ... | ... | **+26 more!** | ... |
| R1 | RDA-R1.3-01M | Metadata complies with a community standard | Essential |
| R1 | RDA-R1.3-01D | Data complies with a community standard | Essential |
| R1 | RDA-R1.3-03M | Metadata uses a machine-understandable standard | Essential |
| R1 | RDA-R1.3-02D | Data uses a machine-understandable standard | Important |

https://doi.org/10.15497/rda00050

# What is FAIR?

| FAIR principle | FAIR Metrics ID | Indicator | Priority |
|---|---|---|---|
| F1 | RDA-F1-01M | Metadata is identified by a persistend identifier | Essential |
| F1 | RDA-F1-01D | Data is identified by a persistent identifier | Essential |
| F1 | RDA-F1-02M | Metadata is identified by a globally unique identifier | Essential |
| F1 | RDA-F1-02D | Data is identified by a globally unique identifier | Essential |
| F2 | RDA-F2-01M | Rich metadata is provided to allow discovery | Essential |
| F3 | RDA-F3-01M | Metadata includes the indentifier for the data | Essential |
| F4 | RDA-F4-01M | Metadata is offered in such a way that it can be harvested and indexed | Essential |
| A1 | RDA-A1-01M | Metadata contains information to enable the user to get access to the data | Important |
| A1 | RDA-A1-02M | Metadata can be accessed manually | Essential |
| A1 | RDA-A1-02D | Data can be accessed manually | Essential |
| A1 | RDA-A1-03M | Metadata identifier resolves to a metadata record | Essential |
| **...** | **...** | **+26 more!** | **...** |
| **R1** | **RDA-R1.3-01M** | **Metadata complies with a community standard** | **Essential** |
| R1 | RDA-R1.3-01D | Data complies with a community standard | Essential |
| R1 | RDA-R1.3-03M | Metadata uses a machine-understandable standard | Essential |
| R1 | RDA-R1.3-02D | Data uses a machine-understandable standard | Important |

https://doi.org/10.15497/rda00050

# iFDOs: image FAIR Digital Objects

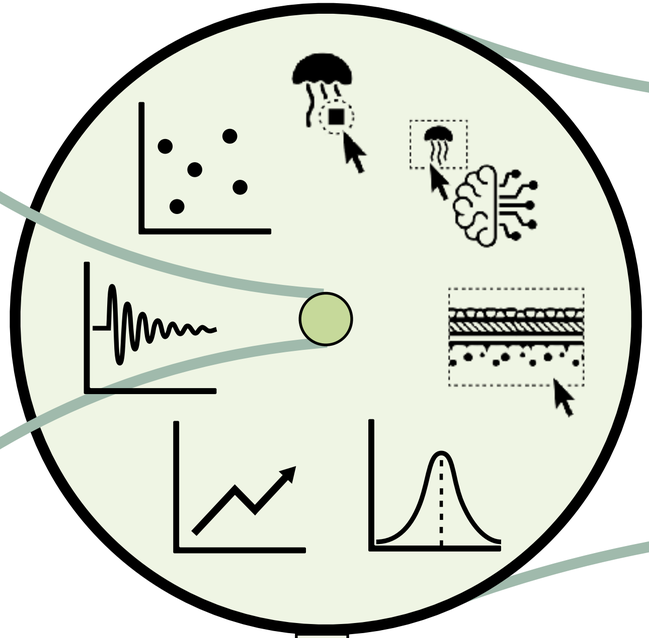
## iFDO core

Persiste dentifier
Li se
Equ ent
Rese hers

## iFDO capture


Acquisition parameters
Camera extrinsic / intrinsics
Acquisition plan
Acquisition execution
...

## iFDO content



Ann tions
Parti counts
Pixe etrics
MPEG7 scriptors

## iFDO

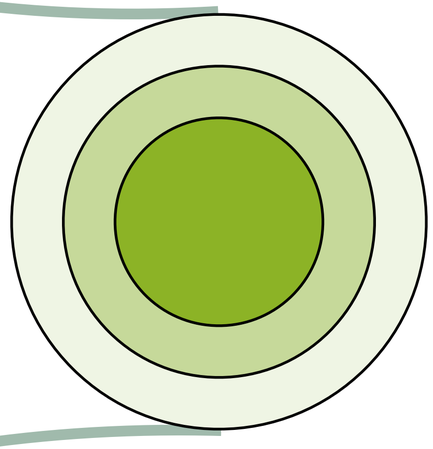
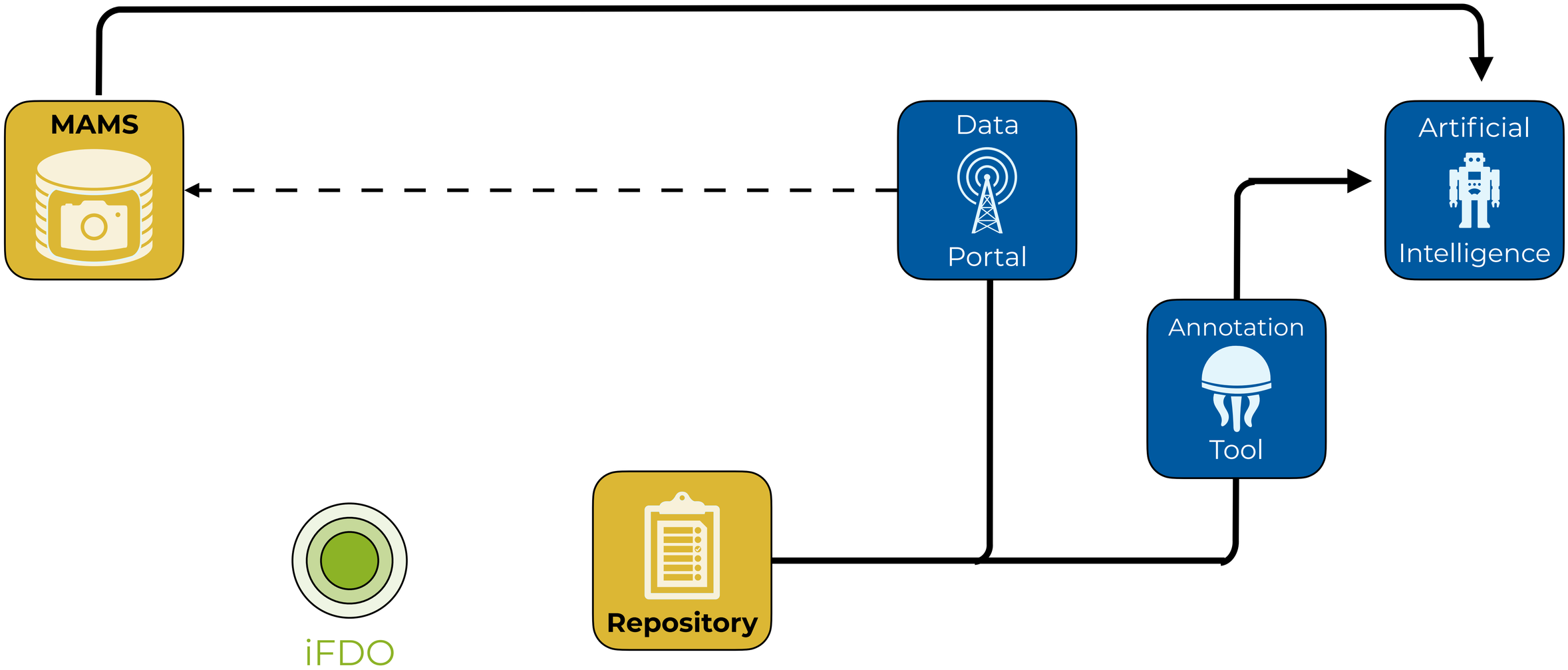


Image data itself is not part of an iFDO file!

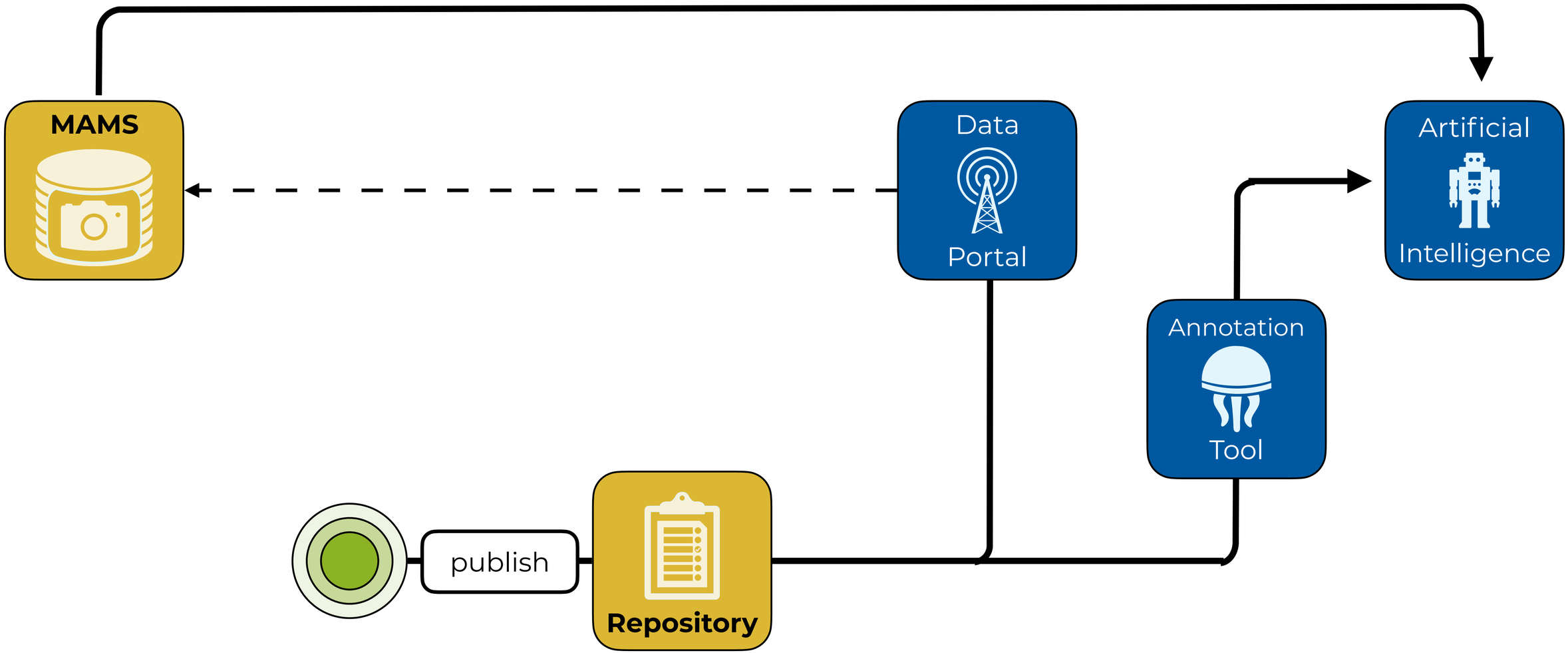
Makes also the pixel & semantic content of images “accessible”!

https://marine-imaging.com/fair/ifdos/iFDO-overview/

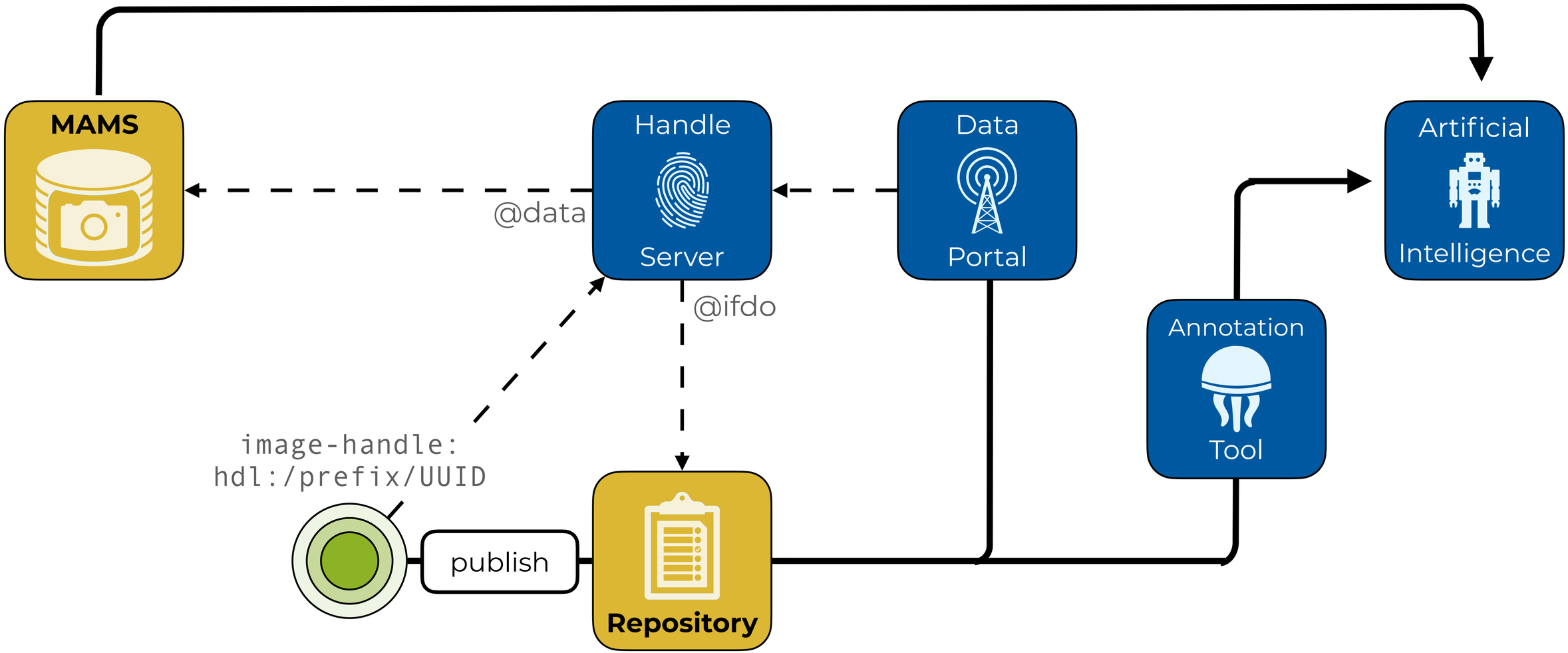
# What do we need?



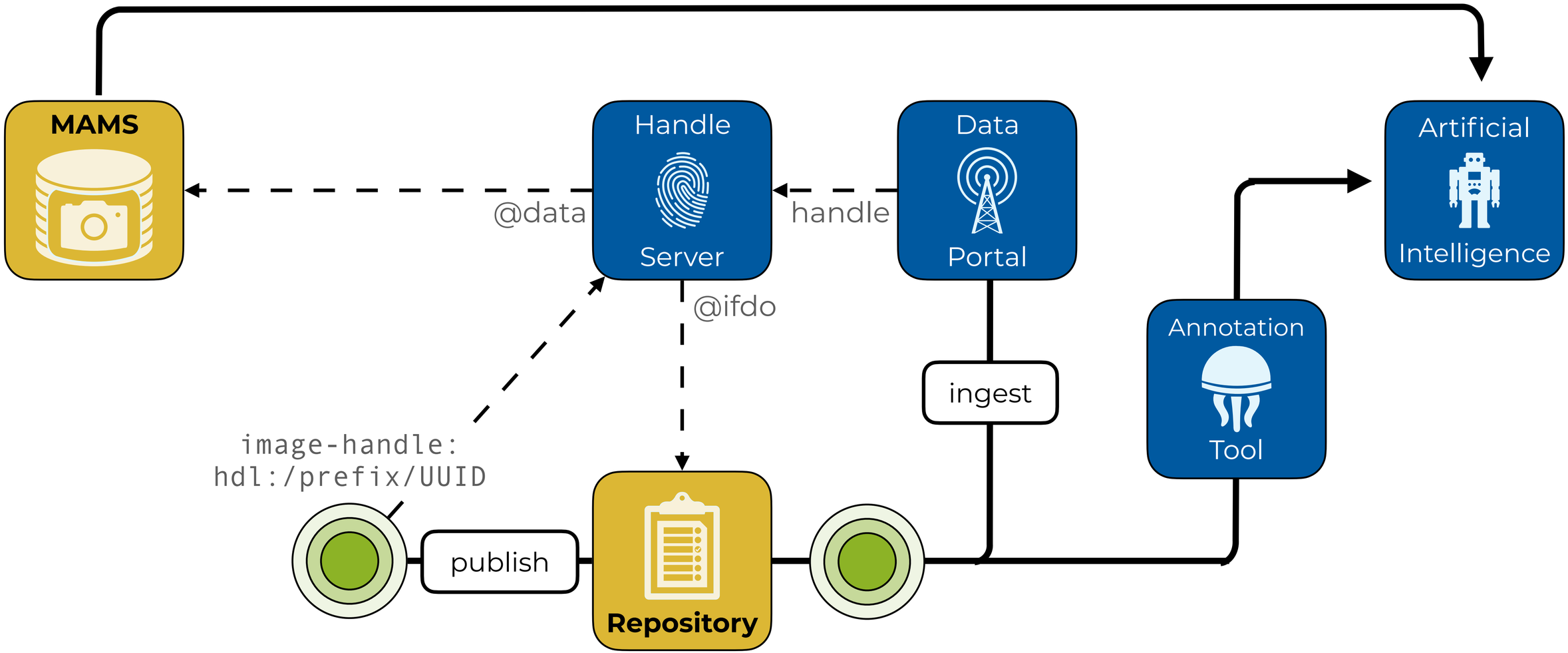
# What do we need?



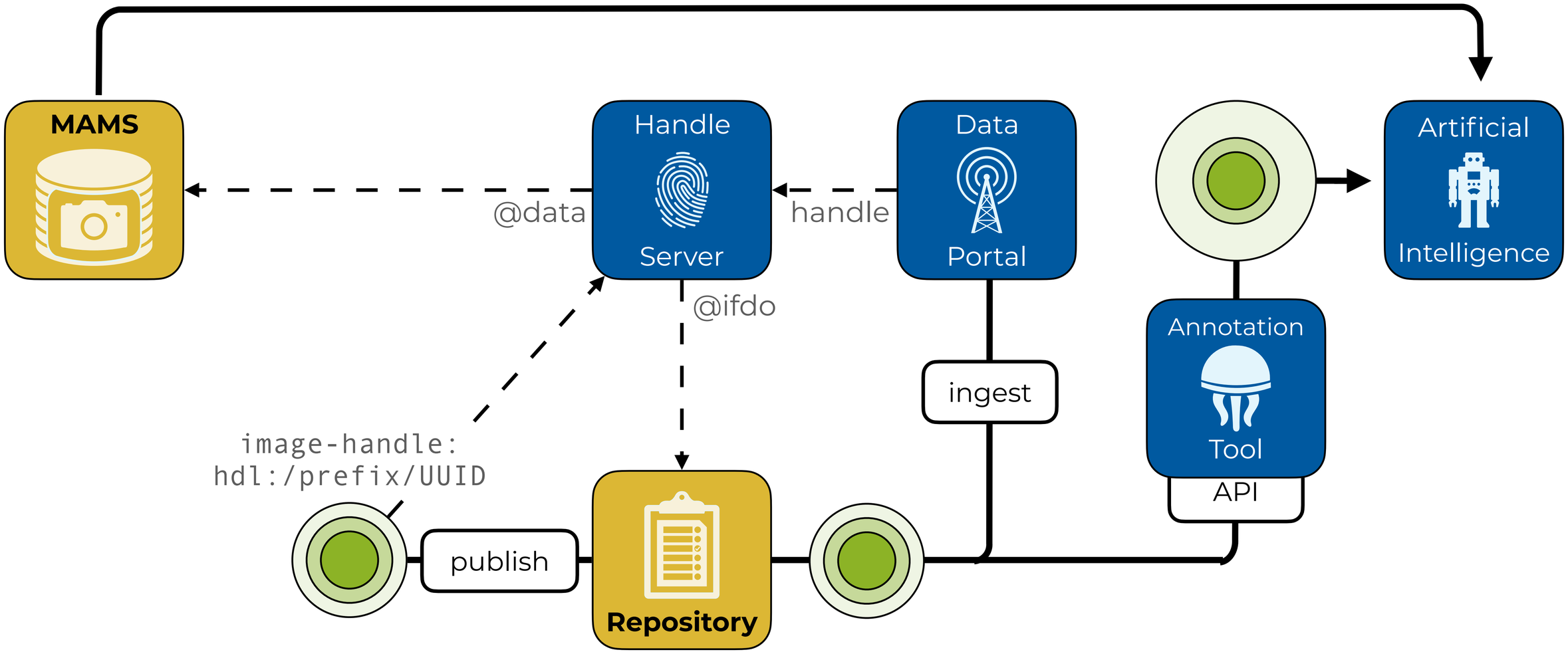
# What do we need?



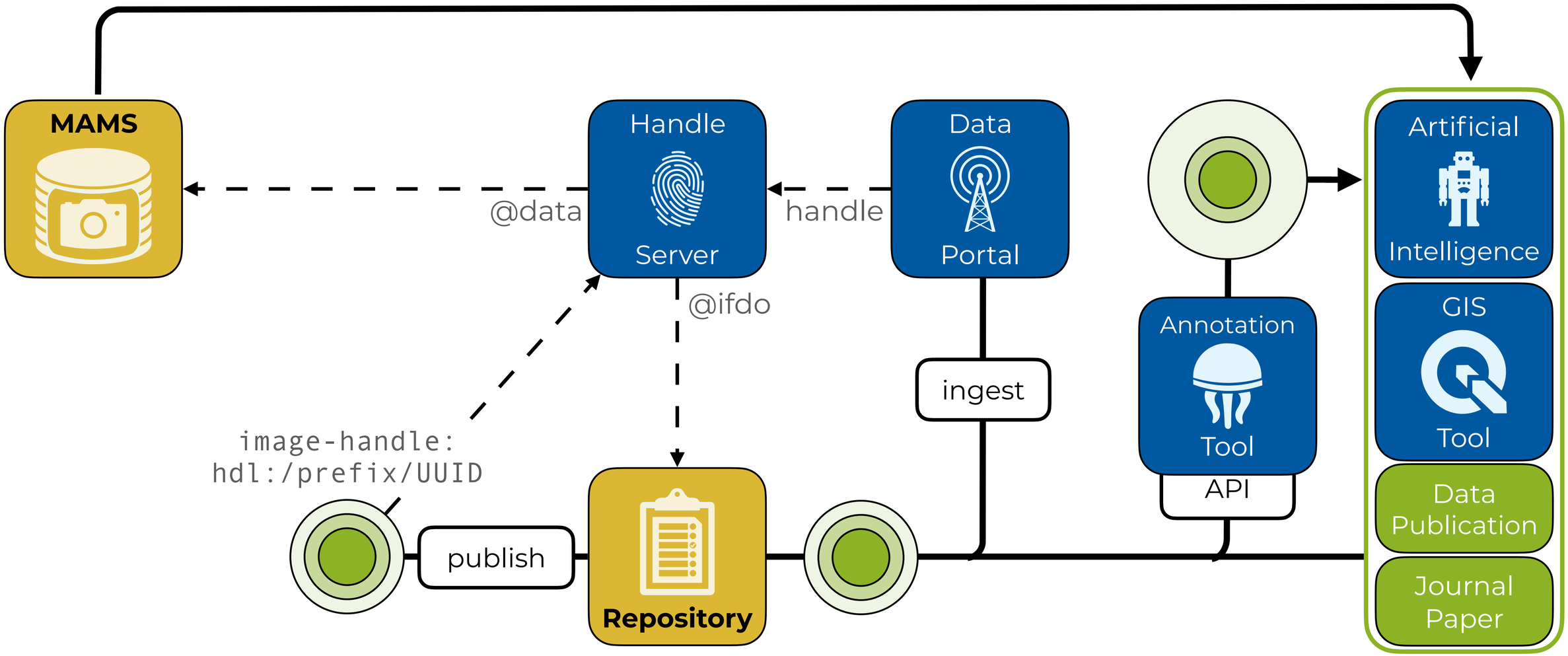
# What do we need?



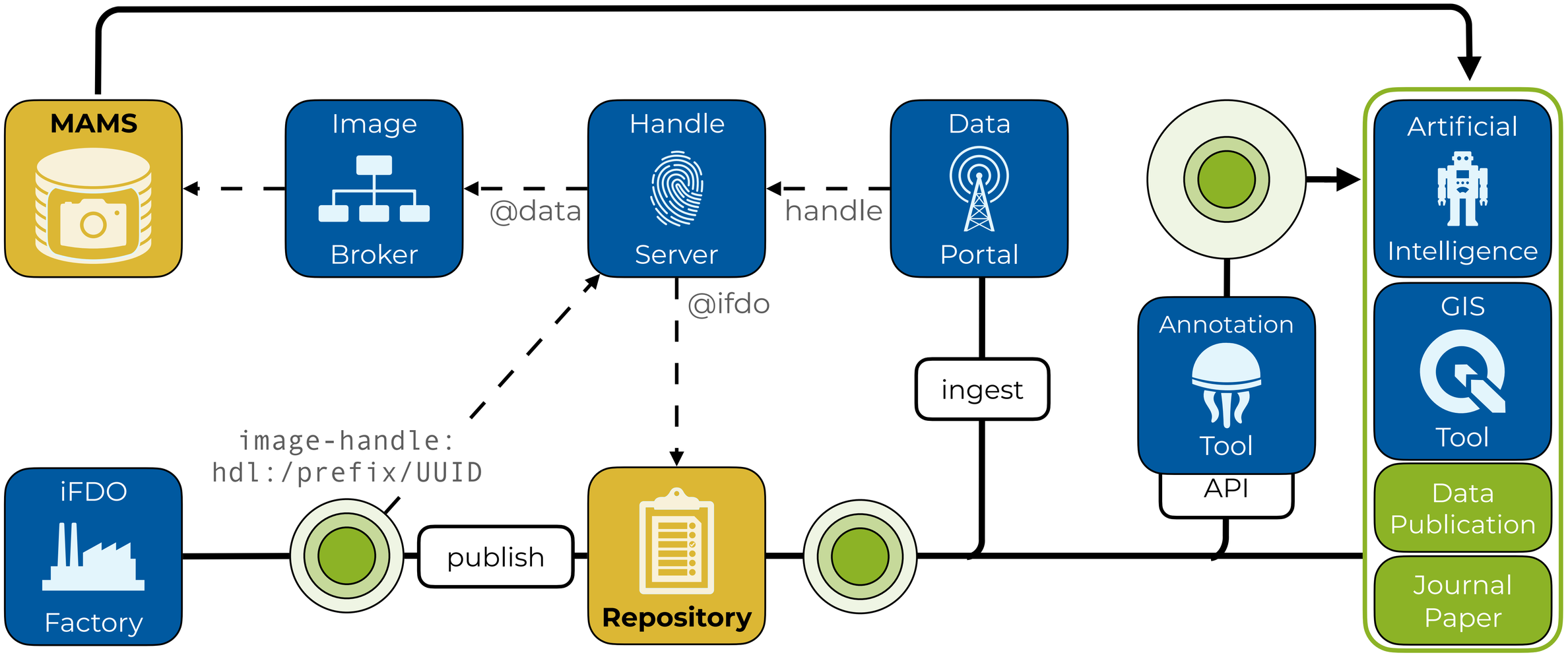
# What do we need?



# What do we need?



# What do we need?



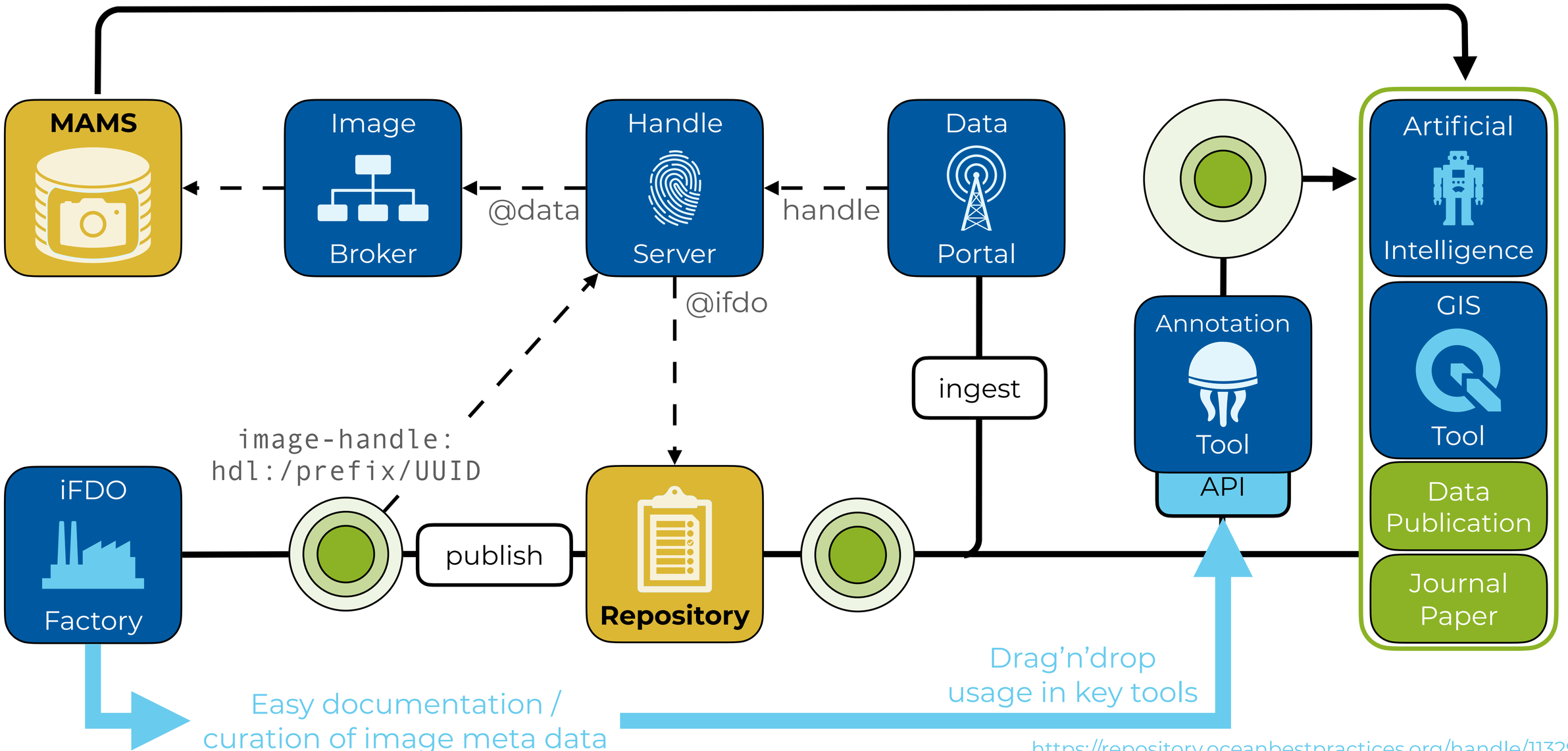
# Why should researchers bother?

| FAIR principle | FAIR Metrics ID | Indicator | Priority |
|---|---|---|---|
| F1 | RDA-F1-01M | Metadata is identified by a persistend identifier | Essential |
| F1 | RDA-F1-01D | Data is identified by a persistent identifier | Essential |
| F1 | RDA-F1-02M | Metadata is identified by a globally unique identifier | Essential |
| F1 | RDA-F1-02D | Data is identified by a globally unique identifier | Essential |
| F2 | RDA-F2-01M | Rich metadata is provided to allow discovery | Essential |
| F3 | RDA-F3-01M | Metadata includes the indentifier for the data | Essential |
| F4 | RDA-F4-01M | Metadata is offered in such a way that it can be harvested and indexed | Essential |
| A1 | RDA-A1-01M | Metadata contains information to enable the user to get access to the data | Important |
| A1 | RDA-A1-02M | Metadata can be accessed manually | Essential |
| A1 | RDA-A1-02D | Data can be accessed manually | Essential |
| A1 | RDA-A1-03M | Metadata identifier resolves to a metadata record | Essential |
| **...** | **...** | **+26 more!** | **...** |
| **R1** | **RDA-R1.3-01M** | **Metadata complies with a community standard** | **Essential** |
| R1 | RDA-R1.3-01D | Data complies with a community standard | Essential |
| R1 | RDA-R1.3-03M | Metadata uses a machine-understandable standard | Essential |
| R1 | RDA-R1.3-02D | Data uses a machine-understandable standard | Important |

https://doi.org/10.15497/rda00050

# They should not!

# Why should researchers bother?




https://repository.oceanbestpractices.org/handle/11329/1781
https://repository.oceanbestpractices.org/handle/11329/1782

# The road ahead

| Indicator | Implementation |
|---|---|
| Metadata is identified by a persistent identifier | iFDO: image-set-handle [@default] |
| Data is identified by a persistent identifier | iFDO: image-set-handle@data |
| Metadata is identified by a globally unique identifier | UUID4 |
| Data is identified by a globally unique identifier | UUID4 |
| Rich metadata is provided to allow discovery | iFDO format definition |
| Metadata includes the indentifier for the data | iFDO: image-set-uuid |
| **Metadata is offered in such a way that it can be harvested and indexed** | **OSIS-API & OAI-PMH service** |
| Metadata can be accessed manually | iFDO: image-set-handle → OSIS |
| Data can be accessed manually | iFDO: image-set-handle@data → MAMS |
| Metadata identifier resolves to a metadata record | iFDO: image-set-handle → OSIS |
| Data identifier resolves to a digital object | iFDO: image-set-handle@data → MAMS |
| Metadata is acccessed through standardised protocol | HTTPS |
| Data is accessible through standardised protocol | HTTPS, NFS, SMB |
| Metadata is accessible through a free access protocol | HTTPS |
| Metadata is guaranteed to remain available after data is no longer available | Handle server + H5AI server |
| Plurality of accurate and relevant attributes are provided to allow reuse | iFDO content and capture fields |
| Metadata includes information about the license under which the data can be reused | iFDO core fields |
| Metadata complies with a community standard | iFDO standard |
| Data complies with a communty standard | JPG, PNG, MOV, AVI, ... |
| Metadata is expressed in compliance with a machine-understanable community standard | YAML |

https://marine-imaging.com/fair/FAIR-marine-images/

- **Implement all FAIR infrastructure tools (converter, catalogue, iFDO ontology, …)**
- **Address data provenance**
- **Establish automated handle registration workflow**
- **Automate all possible steps in Elements**
- **Implement iFDO creation into imaging platforms (ROVs, AUVs, drones, …)**

https://marine-imaging.com/fair/roadmap/

- **Code & governance for iFDOs & tools:**
  https://gitlab.hzdr.de/datahub/marehub/ag-videosimages
- **iFDO Documentation:**
  https://marine-imaging.com/fair
- **Nature Scientific Data paper on iFDOs:**
  in press - UPDATE: https://doi.org/10.1038/s41597-022-01491-3
- **OceanBestPractice SOPs on iFDOs:**
  https://repository.oceanbestpractices.org/handle/11329/1781
  https://repository.oceanbestpractices.org/handle/11329/1781