THE BIG PICTURE
Benthic Imagery Analysis Group

BIAP progress in the digital realm: AI, Data flows and Catalogues

Kerry Howell
kerry.howell@plymouth.ac.uk

UNIVERSITY OF PLYMOUTH
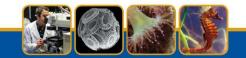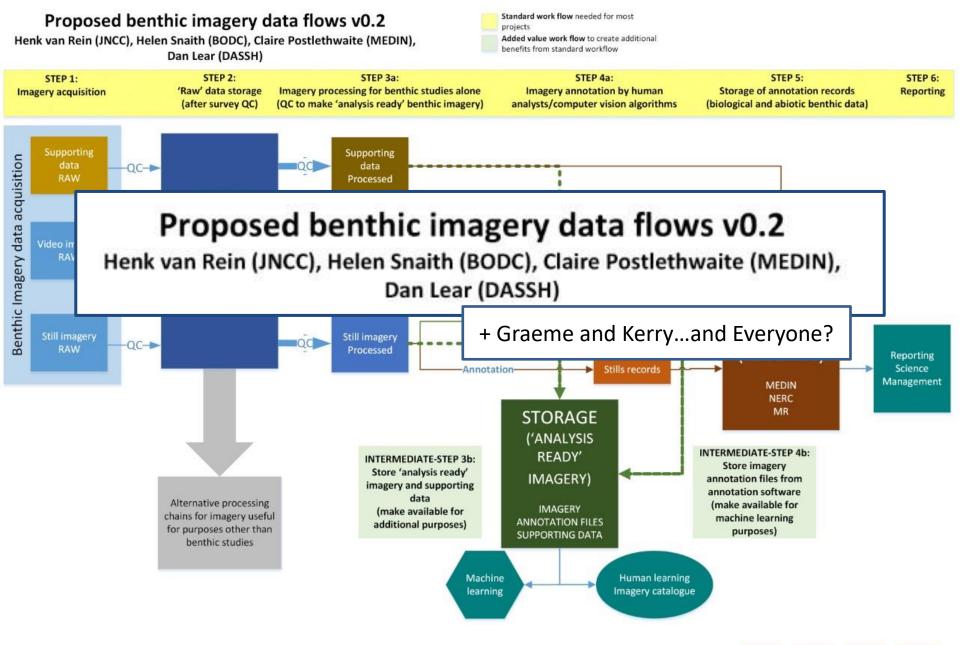
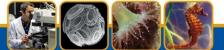## Graeme Duncan
graeme.duncan@jncc.gov.uk

JNCC

Dan Lear
dble@mba.ac.uk

THE MARINE BIOLOGICAL ASSOCIATION

Proposed benthic imagery data flows v0.2

Henk van Rein (JNCC), Helen Snaith (BODC), Claire Postlethwaite (MEDIN),
Dan Lear (DASSH)

Standard work flow needed for most projects

Added value work flow to create additional benefits from standard workflow

| STEP 1: Imagery acquisition | STEP 2: 'Raw' data storage (after survey QC) | STEP 3a: Imagery processing for benthic studies alone (QC to make 'analysis ready' benthic imagery) | STEP 4a: Imagery annotation by human analysts/computer vision algorithms | STEP 5: Storage of annotation records (biological and abiotic benthic data) | STEP 6: Reporting |

Benthic Imagery data acquisition

Supporting data RAW — QC →

Video imagery RAW

Still imagery RAW — QC →

Supporting data Processed

Still imagery Processed

Alternative processing chains for imagery useful for purposes other than benthic studies

INTERMEDIATE-STEP 3b: Store 'analysis ready' imagery and supporting data (make available for additional purposes)

STORAGE ('ANALYSIS READY' IMAGERY)

IMAGERY ANNOTATION FILES SUPPORTING DATA

Annotation

Stills records

MEDIN NERC MR

INTERMEDIATE-STEP 4b: Store imagery annotation files from annotation software (make available for machine learning purposes)

Reporting Science Management

Machine learning

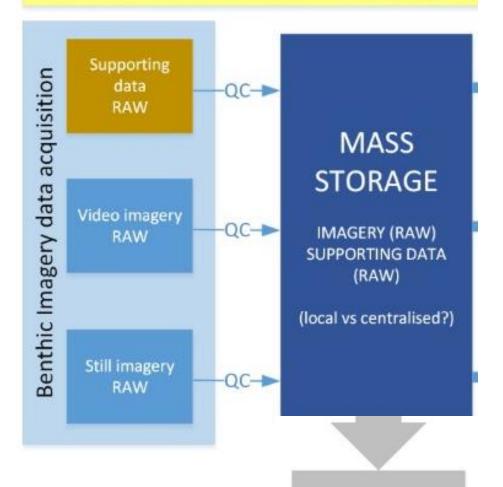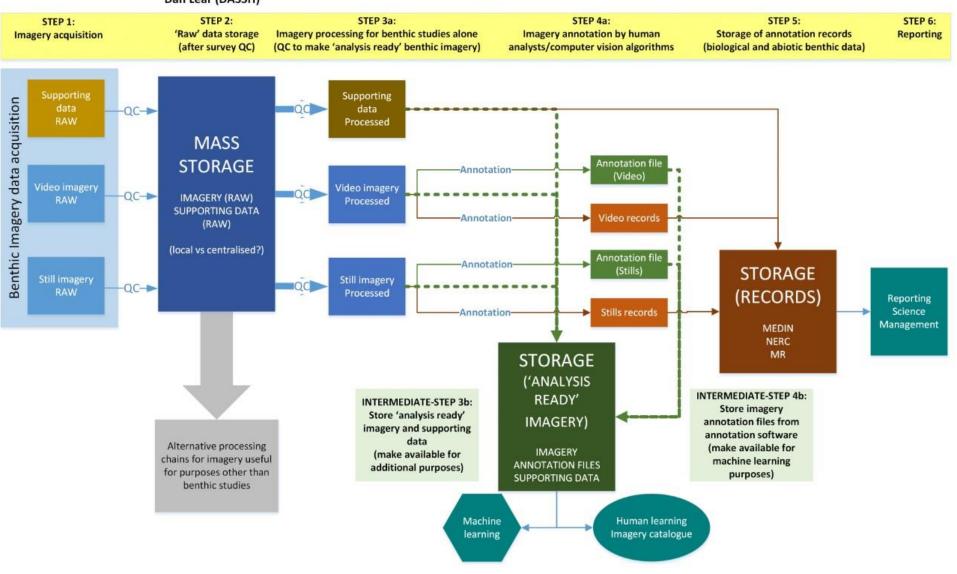Human learning Imagery catalogue

+ Graeme and Kerry...and Everyone?

Proposed benthic imagery data flows v0.2

Henk van Rein (JNCC), Helen Snaith (BODC), Claire Postlethwaite (MEDIN), Dan Lear (DASSH)

Standard work flow needed for most projects

Added value work flow to create additional benefits from standard workflow

STEP 1: Imagery acquisition

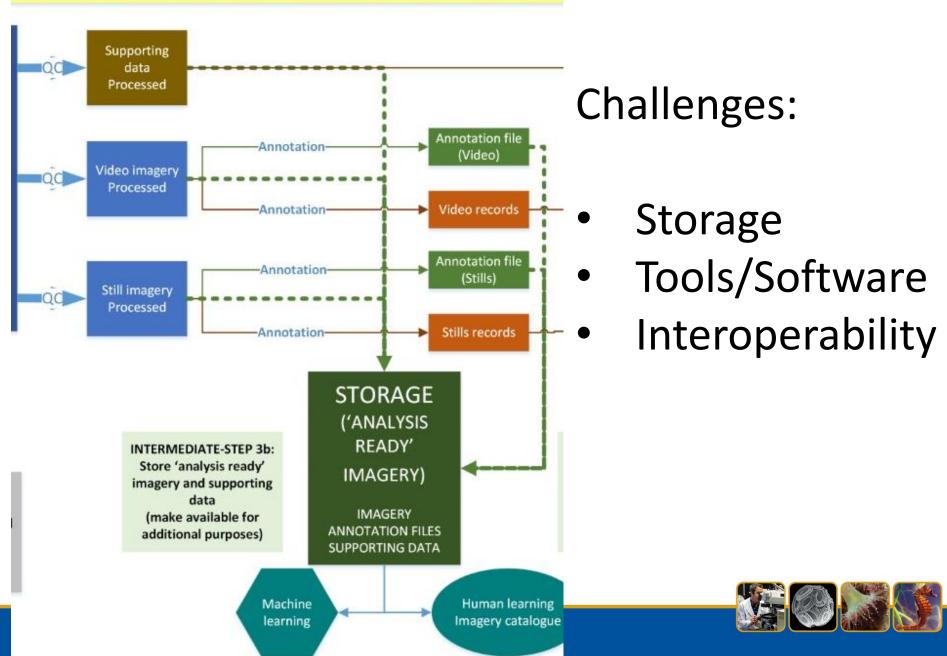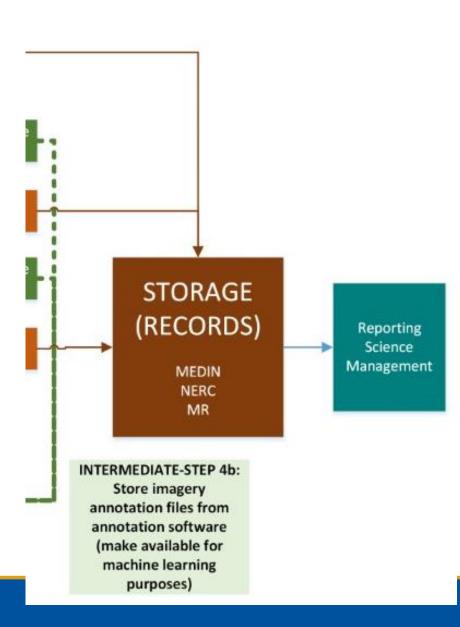STEP 2: 'Raw' data storage (after survey QC)

STEP 3a: Imagery processing for benthic studies alone (QC to make 'analysis ready' benthic imagery)

STEP 4a: Imagery annotation by human analysts/computer vision algorithms

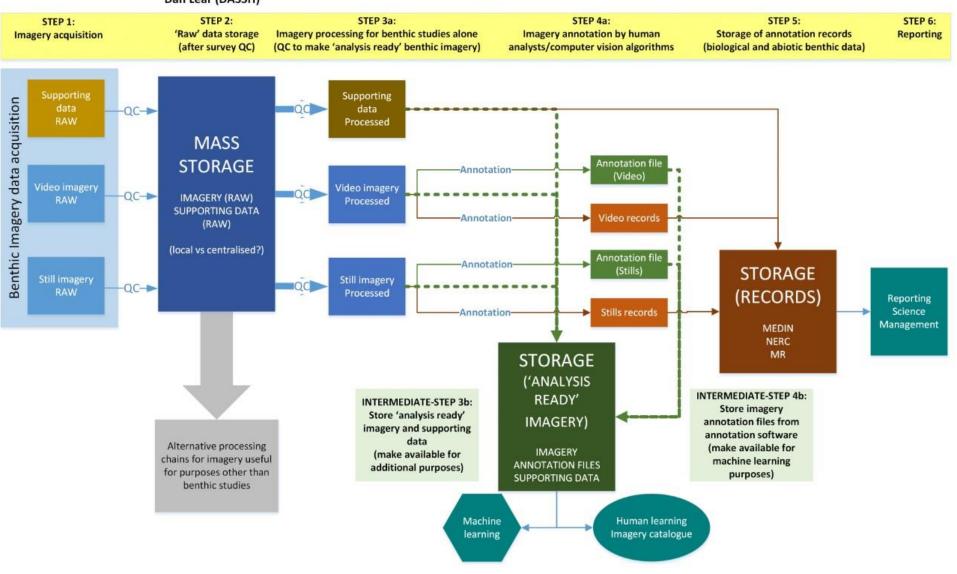STEP 5: Storage of annotation records (biological and abiotic benthic data)

STEP 6: Reporting

Benthic Imagery data acquisition

Supporting data RAW

Video imagery RAW

Still imagery RAW

QC

MASS STORAGE

IMAGERY (RAW) SUPPORTING DATA (RAW)

(local vs centralised?)

QC

Supporting data Processed

Video imagery Processed

Still imagery Processed

Annotation

Annotation file (Video)

Annotation

Video records

Annotation

Annotation file (Stills)

Annotation

Stills records

STORAGE (RECORDS)

MEDIN NERC MR

Reporting Science Management

Alternative processing chains for imagery useful for purposes other than benthic studies

INTERMEDIATE-STEP 3b: Store 'analysis ready' imagery and supporting data (make available for additional purposes)

STORAGE ('ANALYSIS READY' IMAGERY)

IMAGERY ANNOTATION FILES SUPPORTING DATA

INTERMEDIATE-STEP 4b: Store imagery annotation files from annotation software (make available for machine learning purposes)

Machine learning

Human learning Imagery catalogue

STEP 1: Imagery acquisition	STEP 2: 'Raw' data storage (after survey QC)	

Benthic Imagery data acquisition

Supporting data RAW — QC→

Video imagery RAW — QC→

Still imagery RAW — QC→

MASS STORAGE

IMAGERY (RAW)
SUPPORTING DATA (RAW)

(local vs centralised?)

Alternative processing chains for imagery useful for purposes other than benthic studies

Challenges:

- Storage
- Standards
- QC processes
- Provenance

STEP 3a:
Imagery processing for benthic studies alone
(QC to make 'analysis ready' benthic imagery)

STEP 4a:
Imagery annotation by human
analysts/computer vision algorithms

Supporting data Processed

Video imagery Processed

Still imagery Processed

Annotation

Annotation

Annotation

Annotation

Annotation file (Video)

Video records

Annotation file (Stills)

Stills records

INTERMEDIATE-STEP 3b:
Store 'analysis ready' imagery and supporting data
(make available for additional purposes)

STORAGE
('ANALYSIS READY' IMAGERY)

IMAGERY
ANNOTATION FILES
SUPPORTING DATA

Machine learning

Human learning
Imagery catalogue

Challenges:

- Storage
- Tools/Software
- Interoperability

STEP 5:
Storage of annotation records
(biological and abiotic benthic data)

STEP 6:
Reporting

STORAGE
(RECORDS)

MEDIN
NERC
MR

Reporting
Science
Management

INTERMEDIATE-STEP 4b:
Store imagery
annotation files from
annotation software
(make available for
machine learning
purposes)

Challenges:

- Re-aggregation
- Interpretation

Proposed benthic imagery data flows v0.2

Henk van Rein (JNCC), Helen Snaith (BODC), Claire Postlethwaite (MEDIN), Dan Lear (DASSH)

Standard work flow needed for most projects

Added value work flow to create additional benefits from standard workflow

| STEP 1:
Imagery acquisition | STEP 2:
'Raw' data storage
(after survey QC) | STEP 3a:
Imagery processing for benthic studies alone
(QC to make 'analysis ready' benthic imagery) | STEP 4a:
Imagery annotation by human
analysts/computer vision algorithms | STEP 5:
Storage of annotation records
(biological and abiotic benthic data) | STEP 6:
Reporting |

Benthic Imagery data acquisition

Supporting data RAW

Video imagery RAW

Still imagery RAW

QC

MASS STORAGE

IMAGERY (RAW)
SUPPORTING DATA
(RAW)

(local vs centralised?)

QC

Supporting data Processed

Video imagery Processed

Still imagery Processed

Annotation

Annotation file (Video)

Video records

Annotation file (Stills)

Stills records

STORAGE (RECORDS)

MEDIN
NERC
MR

Reporting Science Management

Alternative processing chains for imagery useful for purposes other than benthic studies

INTERMEDIATE-STEP 3b:
Store 'analysis ready' imagery and supporting data
(make available for additional purposes)

STORAGE ('ANALYSIS READY' IMAGERY)

IMAGERY
ANNOTATION FILES
SUPPORTING DATA

INTERMEDIATE-STEP 4b:
Store imagery annotation files from annotation software (make available for machine learning purposes)

Machine learning

Human learning Imagery catalogue

Data Landscape

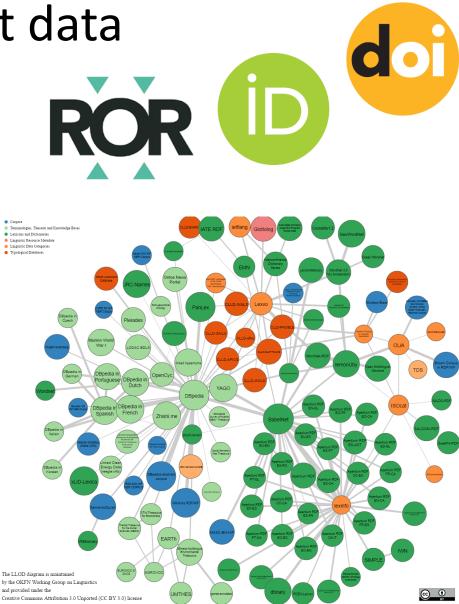| National	International
UKMMAS/MSCC Framework	EMODnet
	OBIS
MEDIN	ICES
Data Archive Centres	OSPAR
Data Standards	
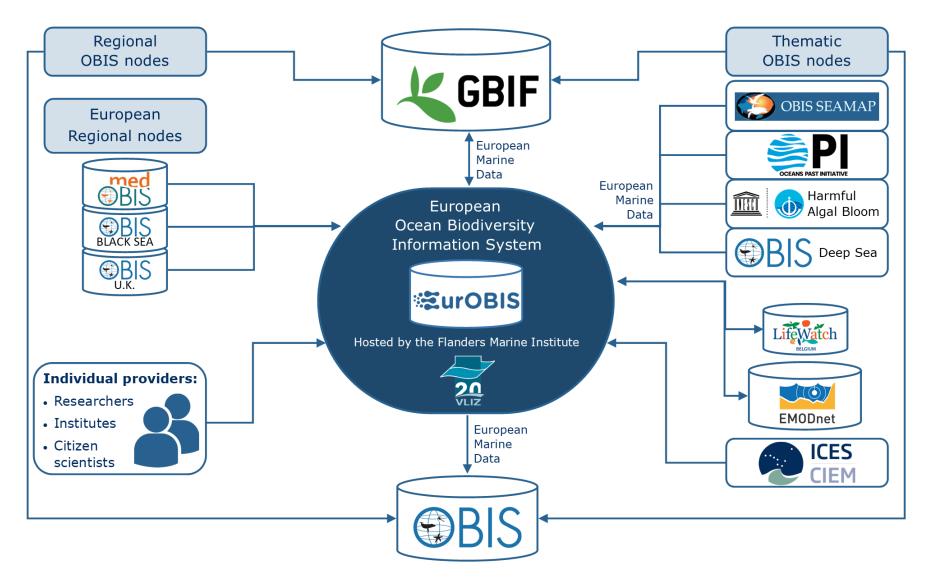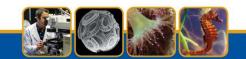
Record storage (Step 5)

Re-aggregation of split data

High quality metadata
Consistent use of PID's

Technically possible but still challenging!

Being explored within MEDIN

Reviewing Action Plan Activities

Are they still valid?

Is anything missing?

Who wants to be involved?

What are the quick wins/high value targets?

And…..

Benthic imagery data flows

Review imagery data culture in UK and explore possible incentives to share imagery data	Coordination/ Review	Range of options for incentives for sharing of imagery data across UK	L
Consider what data should be shared and how it could be archived	Project	Data archive flows	M
Identify current data-sharing flows and gaps for benthic imagery	Project	Understanding current data-sharing flows and gaps	M
Target data-sharing gaps for imagery, explore potential data-sharing options products and recommend routes for organisations to follow in order to make their data available	Project	Options and recommendations to address data sharing gaps	M

Benthic imagery data archives and catalogues

Create a pathway for annotated imagery to be shared with machine learning community (via common image library?)	Project	Open access to imagery library (for machine learning algorithms)	H
Develop online common image library/catalogue to collate and share reference collections across the UK	Project	Benthic imagery library	H
Explore options for ownership/management of central/common image library of reference collection and, potentially, other annotated imagery	Coordination	Range of options for implementation of Benthic imagery library	M
Develop confidence assessment approach for image library reference collections	Project	Image library confidence assessments	L
Identify knowledge gaps in reference collections and address	Project	Complete reference collections	L

Machine-learning approaches for benthic imagery

Define data format specifications necessary for machine learning algorithms to work, as well as export formats for biologists to use (e.g. csv files)	Project	Data format requirements for machine learning algorithms	**H**
Review image annotation software and machine learning approaches for benthic imagery analysis purposes to explore differences in user interface, open-source vs contracted, point vs polygon/boxes, stills vs video	Project	Review of current annotation and machine learning approaches	**H**
Carry out a machine learning workshop for key participants in UK machine learning community (biologists and computer engineers) to conceptualise works flows for benthic biology imagery and facilitate collaboration	Workshop	Machine learning workshop for benthic ecologists	**M**
Scope out international expertise in machine learning and explore options for how to collaborate	Coordination	List of international contacts with expertise in machine learning using benthic imagery	**M**
Define targets/tasks for machine learning algorithms to be tested on, such as identifying indicator taxa in benthic imagery	Project	List of targets/tasks for machine learning algorithms	**H**
Explore potential uses and value of image annotation and machine learning approaches to support Quality Assurance procedures, training, development of reference collections	Project	Role of annotation and machine learning approaches in QA	**M**
Explore use of google imagery to substitute machine learning needs	Project	Use of google images for machine learning applications	**M**
Develop confidence assessment approaches for machine learning outputs	Project	Machine learning confidence assessments	**L**