

Session 1: Long-term monitoring of hard-bottom marine communities ARMS Workflow - development, current status, functionalities



Antonio José Sáenz-Albanés | LW-ERIC ICT-Core Operations Coordinator

1. Forewords about ICT presentations

Two sections in each ICT presentation

Associated Workflow

Explain technical aspects related to the current deployment

Tesseract Architecture

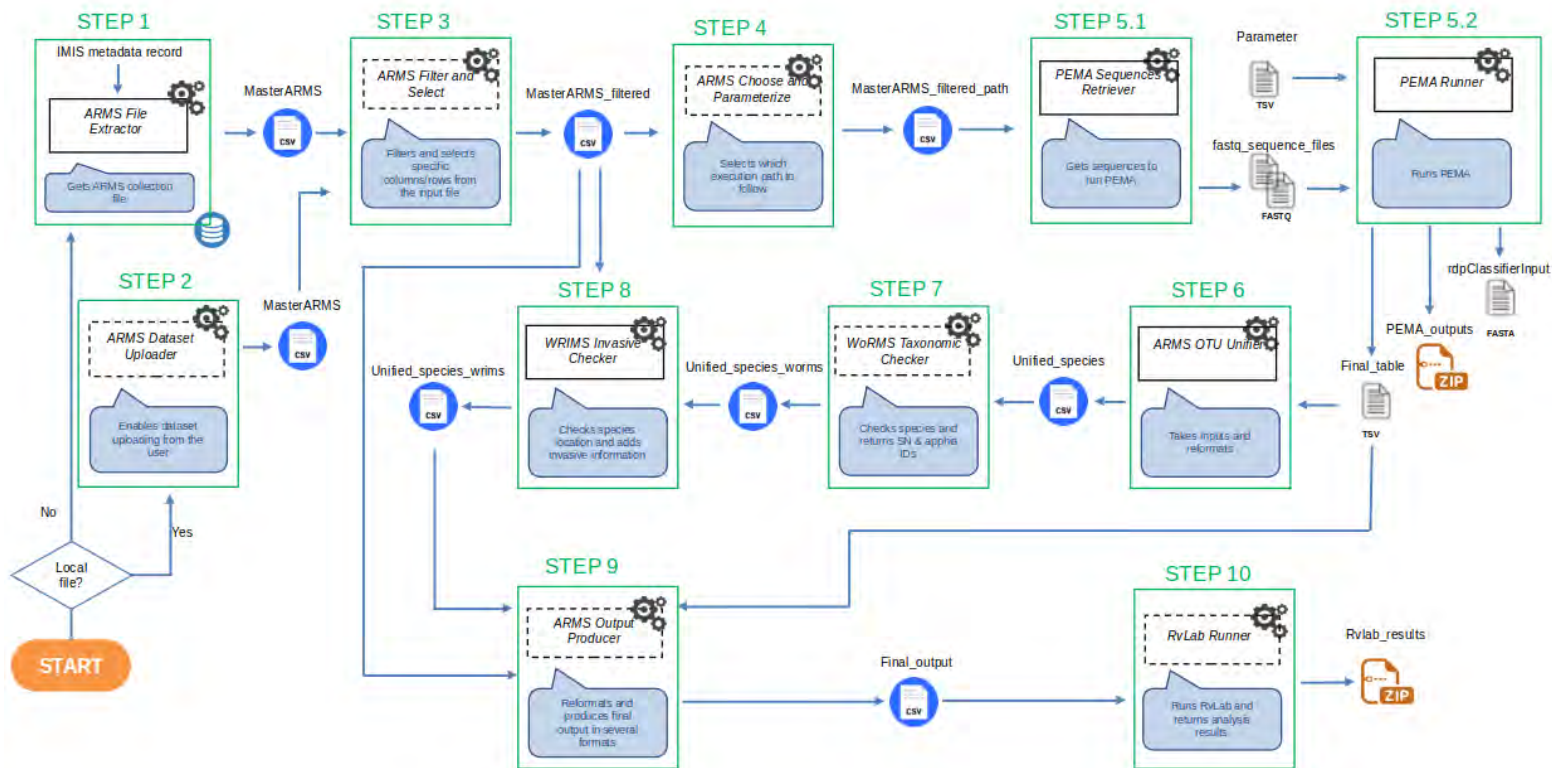
Explain technical aspects of the architecture **tested** in the development of the workflows

2. Current deployment of ARMS workflow

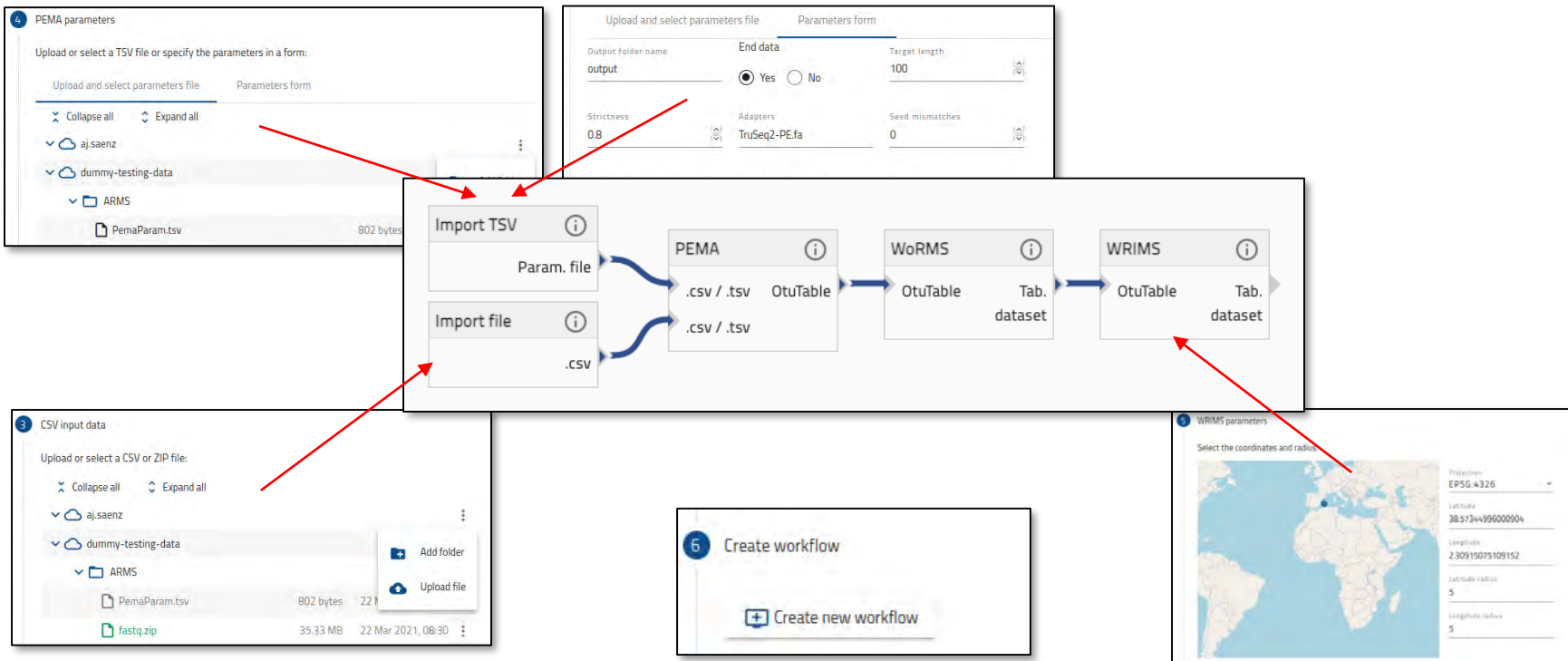
Associated Workflow

Explain technical aspects
related to the current
deployment

ARMS Validation Case



ARMS UI Current Implementation

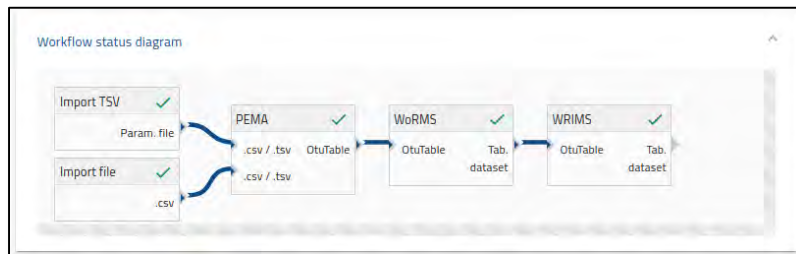


The image displays a workflow in the ARMS UI, consisting of six numbered steps:

- PEMA parameters:** A form for uploading a TSV file or specifying parameters. It includes a file browser showing 'PemaParam.tsv' (802 bytes) and a 'Parameters form' with fields for 'Output folder name' (output), 'End data' (Yes/No), 'Target length' (100), 'Strictness' (0.8), 'Adapters' (TruSeq2-PE.fa), and 'Seed mismatches' (0).
- CSV input data:** A form for uploading a CSV or ZIP file. It shows a file browser with 'PemaParam.tsv' (802 bytes) and 'fastq.zip' (35.33 MB).
- Workflow diagram:** A central flowchart showing the data processing pipeline:
 - 'Import TSV' (Param. file) and 'Import file' (.csv) feed into 'PEMA' (.csv / .tsv OtuTable).
 - 'PEMA' feeds into 'WoRMS' (OtuTable Tab. dataset).
 - 'WoRMS' feeds into 'WRIMS' (OtuTable Tab. dataset).
- WRIMS parameters:** A form for selecting coordinates and radius. It includes a map and fields for 'Projection' (EPSG:4326), 'Latitude' (38.5734+996000904), 'Longitude' (2.30915075109152), 'Latitude radius' (5), and 'Longitude radius' (5).
- Create workflow:** A button labeled 'Create new workflow'.

Red arrows indicate the flow of data and parameters between these steps: from the file browser in step 1 to the 'Import TSV' node; from the 'Parameters form' in step 1 to the 'PEMA' node; from the file browser in step 2 to the 'Import file' node; from the 'PEMA' node to the 'WoRMS' node; from the 'WoRMS' node to the 'WRIMS' node; and from the 'WRIMS parameters' form to the 'WRIMS' node.

ARMS UI Current Implementation



```

    < 0
    name: ComponentImportFile1
    createdAt
    updatedAt
    params
    inputs
    < result
    message
    < files
    < 0
    scheme: minio://
    resource: minio://971683ea8b7c475999df4a88989bc7b4/ComponentImportFile1/fastq.zip
    < log
    scheme: minio://
    resource: minio://971683ea8b7c475999df4a88989bc7b4/ComponentImportFile1/log.txt
    status: DONE
    > 1
  
```

Workflow output files

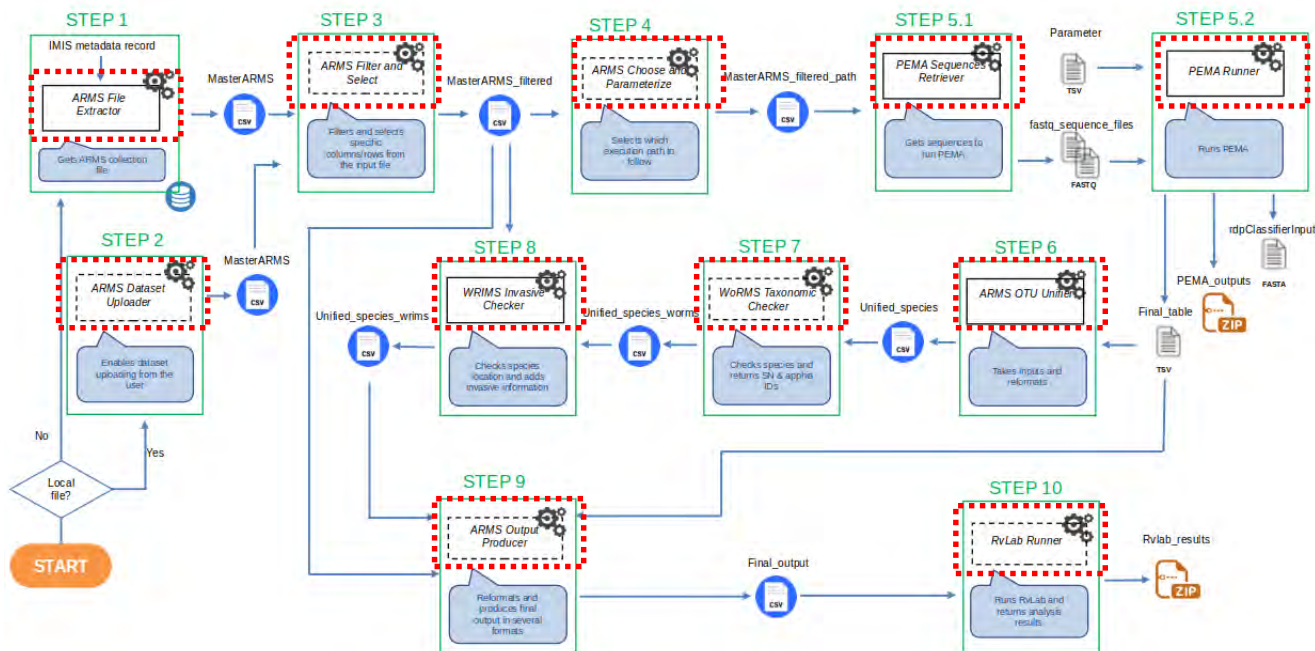
✖ Collapse all ⤴ Expand all

☑ 971683ea8b7c475999df4a88989bc7b4

- ☑ ComponentImportFile1
 - 📄 fastq.zip 35.33 MB 23 Apr 2021, 16:44
 - 📄 log.txt 363 bytes 23 Apr 2021, 16:44
- ☑ ComponentImportTSV1
 - 📄 log.txt 372 bytes 23 Apr 2021, 16:44
 - 📄 out.tsv 800 bytes 23 Apr 2021, 16:44
- ☑ ComponentPEMA1
 - 📄 final_table.tsv 9.11 KB 23 Apr 2021, 16:46
 - 📄 log.txt 7.68 MB 23 Apr 2021, 16:46
- ☑ ComponentWRIMS1
 - 📄 classification.csv 43 bytes 23 Apr 2021, 16:47
 - 📄 log.txt 1.04 KB 23 Apr 2021, 16:47
- ☑ ComponentWoRMS1
 - 📄 log.txt 9.8 KB 23 Apr 2021, 16:47
 - 📄 otu_table.csv 1.06 KB 23 Apr 2021, 16:47

Download Delete

ARMS Wrappers Implementation and Generalization



A “wrapper” provides two aspects:

- An interface to the workflow engine
- An execution environment and implementation

One of our goals is to generalize them so that we can reuse



ARMS File Extractor

Label: ARMS File Extractor

Description: Retrieves MasterARMS.csv

Input: ARMS URL

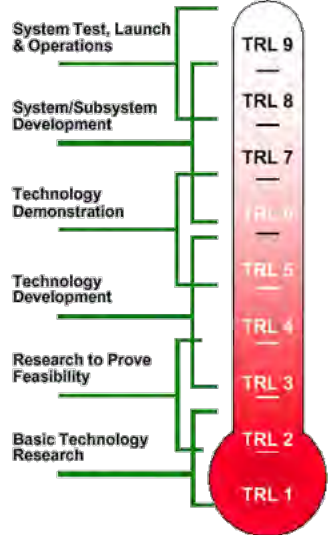
Output: MasterARMS (csv)

Metadata Harmonized: YES

TRL: 5

GUI Integration: YES

Technology: R / Python



Technology Readiness Level

PEMA Sequences Retriever

Label: PEMA Sequences Retriever

Description: Retrieves Sequences fasta files for PEMA

Input: MasterARMS (csv)

Output: fastq_sequence_files (fastq)

Metadata Harmonized: YES

TRL: 4

GUI Integration: YES

Technology: R / Python

PEMA Seq Retriever	MasterARMS.csv+	
	Parameter.tsv	fastq sequence files

PEMA Runner

Label: PEMA Runner

Description: Executes PEMA code and produces results

Input: fastq_sequence_files (fastq)

Output: Final_table (tsv),
rdpClassifierInput (fasta),
PEMA_outputs (zip)

Metadata Harmonized: YES

TRL: 5

GUI Integration: YES

Technology: Docker / R / Python

PEMA Runner	fastq sequence files	
	Parameter.tsv	final_table.tsv
		rdpClassifierInput.fasta
		PEMA_outputs.zip

ARMS OTU Unifier

Label: ARMS File Extractor

Description: Takes input CSV, TSV and XLS and unifies them

Input: Final_table (tsv)

Output: Unified_species (csv)

Metadata Harmonized: YES

TRL: 2

GUI Integration: NO

Technology: R / Python

Unify OUT

final_table.tsv |

unified_species.csv



WoRMS Taxon Check

Label: WoRMS Taxon Check

Description: Species names checked by WoRMS

Input: Unified_species (csv)

Output: Unified_species_worms (csv)

Metadata Harmonized: YES

TRL: 3

GUI Integration: YES

Technology: R / Python

WoRMS taxon check

unified_species.csv+

unified_species.csv



WRIMS Invasive Check

Label: WRIMS Invasive Check

Description: Checks Species location and returns invasive

Input: Unified_species_worms (csv), MasterARMS (csv)

Output: MasterARMS (csv)

Metadata Harmonized: YES

TRL: 2

GUI Integration: NO

Technology: R / Python

WRIMS check	unified_species.csv+	
	MasterARMS.csv+	wrims_output.csv

Produce output

Label: Produce output

Description: Produces final output in various formats

Input: MasterARMS_filtered (csv),

Unified_species_wrims (csv),

Final_table (tsv)

Output: Final_output (csv)

Metadata Harmonized: YES

TRL: 3

GUI Integration: YES

Technology: R / Python

Produce output	MasterARMS.csv	
	unified_species.csv	
	final_table.tsv	final_output



Run RvLab

Label: Run RvLab

Description: Runs RvLab and returns analysis results

Input: Final_output (csv)

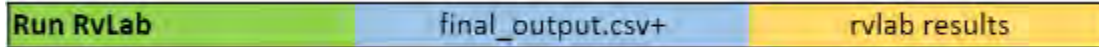
Output: Rvlab_results.zip

Metadata Harmonized: YES

TRL: 3

GUI Integration: NO

Technology: Docker / R / Python



3. Security and Distributed Heterogeneous Executions

Tesseract Architecture

Explain technical aspects of the architecture **tested** in the development of the workflows



Partial Restricted Access

Services:

- Single Sign On
- Workflow UI
- Authorization
- Monitorization

Cloud Services:

- Workflow Engine
- Distributed Storage
- Native Execution
- Distributed Execution
- Remote Service Invocation
- Monitorization

Secured Channel

Secured Channel

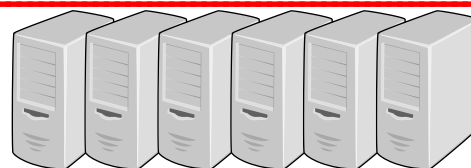
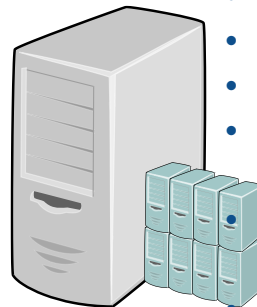
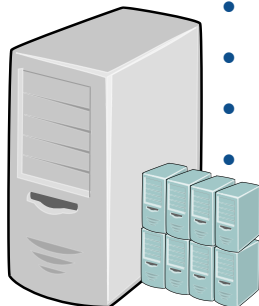
Advanced Perimeter Security

Cloud-HPC Gateway

HPC



User Browser



Thanks!

aj.saenz@lifewatch.eu