



# SeaDataCloud

## WP9 - CDI Replication: Status of the new ingestion process

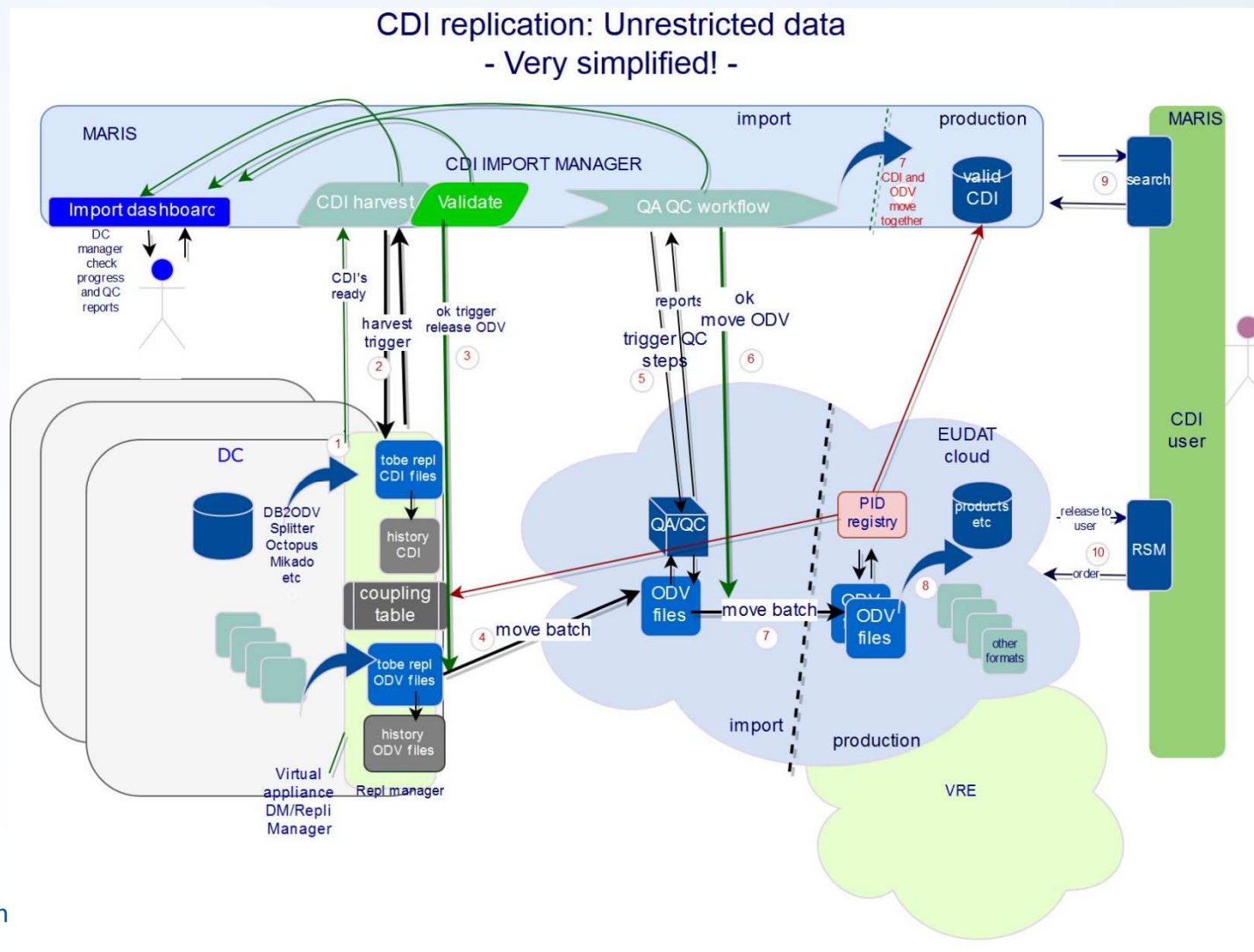
Peter Thijsse on behalf of WP9 team, GA Barcelona  
8/9 Nov 2018

[sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) – [www.seadatanet.org](http://www.seadatanet.org)

## Main components

- Replication manager(s)
- Import manager
- EUDATs HTTP API
- B2SAFE (Cloud storage)
- Authentication B2ACCESS => B2SAFE local accounts

# Overview **unrestricted** dataflow



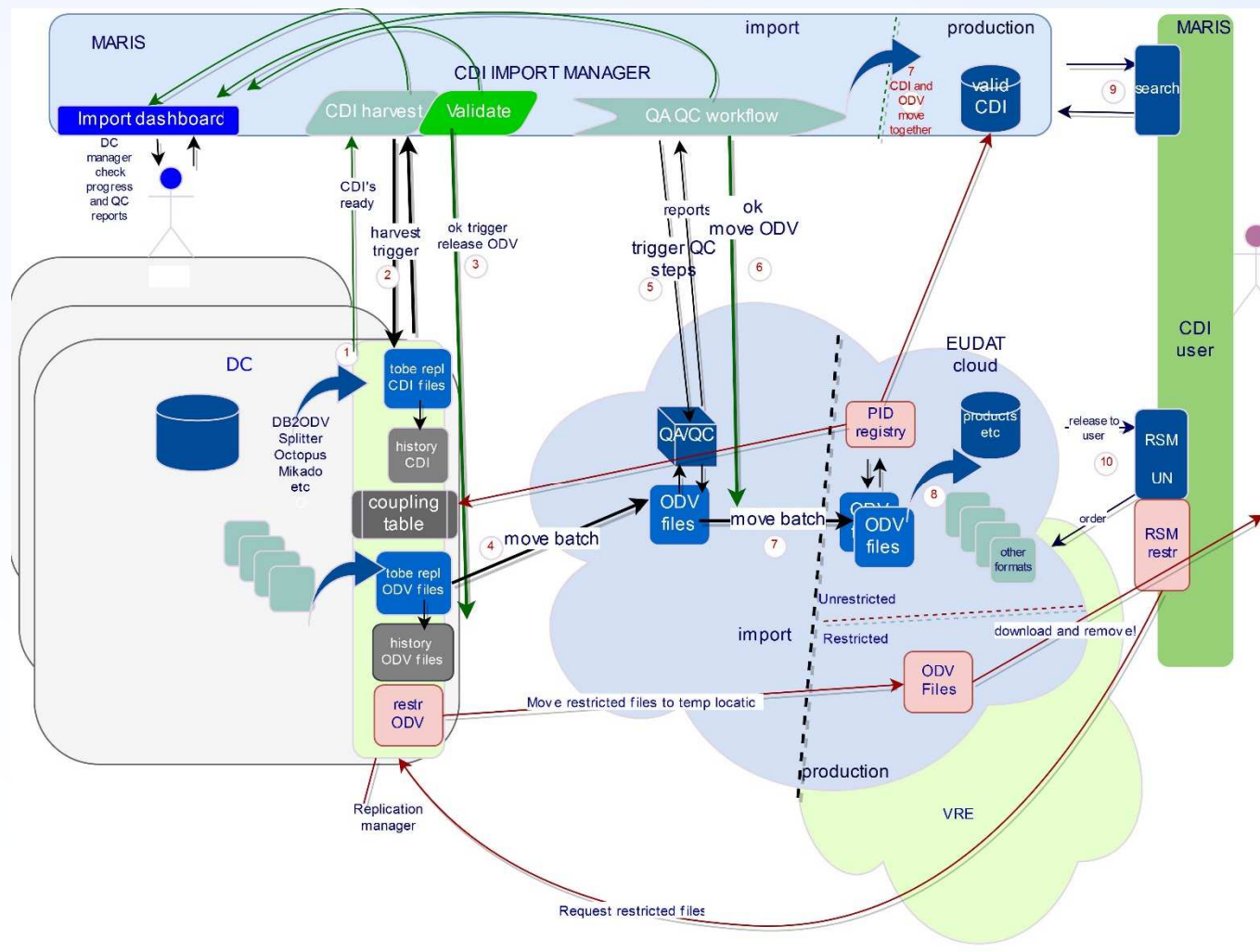
# Remarks unrestricted

The main solution is designed for Unrestricted (UN/LS) data where the process of QA/QC and file format conversions in the cloud are involved.

Important steps:

- The CDI import manager is guiding the process of CDI and ODV release and QA/QC towards production.
- At data center there is production of CDI and ODV (or one other accepted exchange data format) using the stack of tools produced at IFREMER: DB2ODV, Splitter, Mikado, Octopus etc.
- Replication Manager at the Data Center is triggered by import manager to:
  - Move a batch of CDI files to CDI central server
  - Follow up requests to move a batch of ODV files triggered by import manager. One batch at a time handled by the system.
- After moving a batch it is archived by the RM (history of full situation).
- The EUDAT cloud consists of a area for Import and an area of Production. During import QA/QC processes will be triggered by the CDI import manager after which reports flow back. After an ok the datasets will be made available in production. Again moved after a trigger from the import manager.
- When a dataset moves to production the PID is reported back to import manager and DC (coupling table)
- The DC data manager (and masters) can follow progress and if necessary take actions to get to a next stage, or e.g. discard the set and release a new updated batch.

# Overview **restricted** dataflow



## Remarks restricted

Replication Manager at the Data Center is for restricted data triggered by RSM to:

- Follow up requests for releasing and moving a set of restricted data
  - Only when allowed by the DC via confirmation in the RSM
  - Directly to a “user section” in a restricted part of the cloud where it will be available to the user to download it for limited time.
  - After download it will be removed.

Replication of RS data is a sensitive step:

- Trust is needed from the DC in the secure part of the cloud to temporarily store the data
- But, working with the same “trigger and move” process/protocols for each release of data makes it simpler
- The set will be removed in the cloud once downloaded, so it is ready for the next batch.

## 6 steps in the unrestricted data ingestion process

1. Authentication
2. Publishing metadata
3. Checking metadata
4. Uploading a databatch
5. Unzip and run data checks
6. Move validated metadata and data in production



# Step 1: Authentication

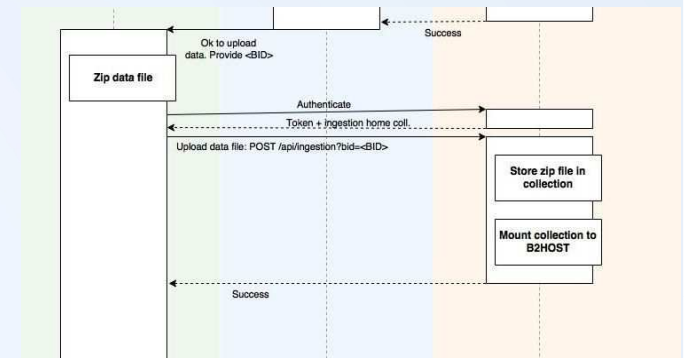
The problem:

RM needs to access the Import Manager and the EUDAT B2STAGE HTTP-API through a software (RM).

Data center contacts needs to access the IM dashboard.

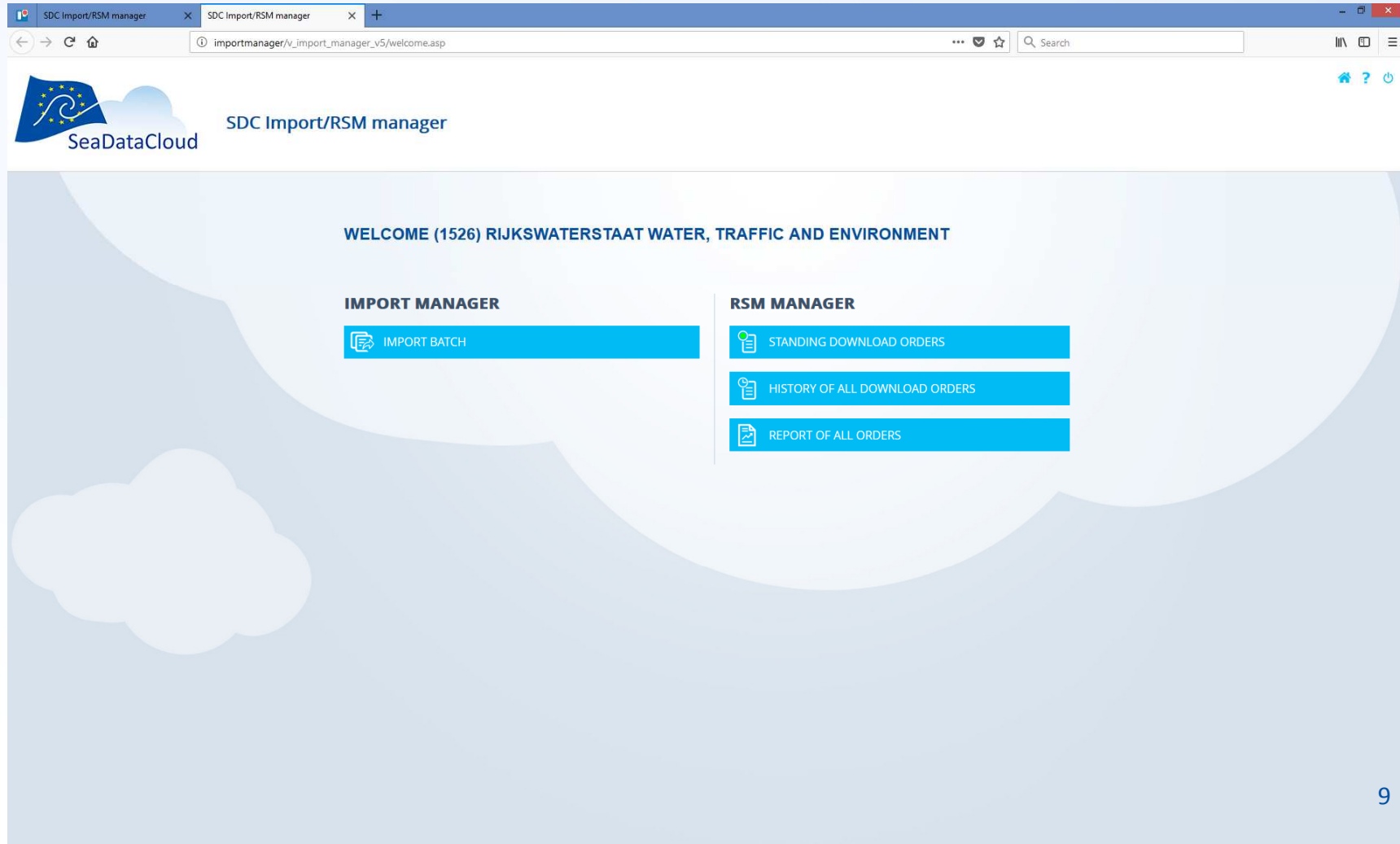
The solution

1. Every RM registers **with a B2SAFE local account**. MARIS prepared neutral “@sdn-taskmanager.org” addresses. These are forwarded to the RSM contacts marineID’s.
2. DC will configure the RM with the credentials
3. RM calls /auth/**b2safe**proxy with credentials
4. Data center contacts linked to the IM via their MarineID can contact the IM dashboard





## Login to the Import Manager for process overview



The screenshot shows a web browser window with the URL `importmanager/v_import_manager_v5/welcome.asp`. The page header includes the SeaDataCloud logo and the text "SDC Import/RSM manager". The main content area displays a welcome message: "WELCOME (1526) RIJKSWATERSTAAT WATER, TRAFFIC AND ENVIRONMENT". Below this, there are two main sections: "IMPORT MANAGER" and "RSM MANAGER".

**IMPORT MANAGER**

- IMPORT BATCH

**RSM MANAGER**

- STANDING DOWNLOAD ORDERS
- HISTORY OF ALL DOWNLOAD ORDERS
- REPORT OF ALL ORDERS

The page has a light blue background with a large white cloud graphic on the left side.

## Step 2: Publishing the CDI metadata

### The problem

- RM needs to notify the Import Manager a new batch is ready
- IM will download when ready

### The solution

1. DC manager puts one or more batches in publish-directory
2. RM triggers
3. IM downloads the CDI metadata batch
4. One batch at a time in the process



# View of RM dashboard

SDC Import/RSM manager

SDC Import/RSM manager

Replication Manager

+


← → ↻ 🏠

83.163.127.252:8080/rm/

⋮ 📄 ☆ 🔍 Search

☰ 📄 ☰

## Replication Manager Dashboard



Summary | Batches History | LOCAL\_CDI\_IDs

2018/05/23 16:52

### Maintenance

Logs

External resources

UPDATE

Coupling Table Checks

LAUNCH CHECKER

Database

BACKUP

### Configuration

RM configuration is valid

#### Automatic parameters

parameter	value	status
objectdb_path	C:\Program Files\Apache Software Foundation\Tomcat 8.5\webapps\rm\db\batchentity.odb	OK
harvestPath_path	C:\Program Files\Apache Software Foundation\Tomcat 8.5\webapps\rm\HARVEST	OK

#### Custom parameters

parameter	value	status
edmo_code	1526	OK
test_mode	true	OK
proxy_host	null	null
proxy_port	-1	null
dataTranslationFile_path	dataTranslationFile_path	OK



# RM: Batch received by IM - Now indicates batchnumber

Import Batch

Replication Manager

83.163.127.252:8080/rm/BatchesHistory

Search

Download

Print

Fullscreen

Menu

<input type="checkbox"/>	2018-06-14 16:59:48	demo12.zip
<input type="checkbox"/>	2018-06-14 16:59:52	demo13.zip
<input type="checkbox"/>	2018-06-14 16:59:55	demo14.zip
<input type="checkbox"/>	2018-06-14 17:00:00	demo15.zip
<input type="checkbox"/>	2018-06-14 16:59:09	demo6.zip
<input type="checkbox"/>	2018-06-14 16:59:24	demo7.zip
<input type="checkbox"/>	2018-06-14 16:59:28	demo8.zip
<input type="checkbox"/>	2018-06-14 16:59:33	demo9.zip

### Current batch

name	batch_number	request_id	sent date	Batch global status	Batch CDIs status	Batch Data status	CDIs files
demo4	325	939cb7b3-5da5-46c6-95a5-dbed02ef610d	2018-06-19 17:15:00	[INGESTION_PENDING] Ingestion started - in process	[ARCHIVED] CDIs batch harvested and archived	[IN_QUEUE_ZONE] Moved to Data queue directory	demo4 (15 files)

### Batches in queue

name	batch_number	request_id	sent date	Batch global status	Batch CDIs status	Batch Data status	CDIs files
demo5			2018-06-20 10:35:35	[IN_QUEUE_READY] In queue, CDIs and data ready in queue	[CHECK_READY] Check is done	[IN_QUEUE_ZONE] Moved to Data queue directory	demo5 (15 files)

### Batches cancelled

name	batch number	request id	sent date	Batch global status	Batch CDIs status	Batch Data status	CDIs files
------	--------------	------------	-----------	---------------------	-------------------	-------------------	------------

## Step 3: Checking the CDI metadata

### The problem

- Import Manager to check the CDI XML and its content
- RM manager to check also visually

### The solution

1. Unzip the batch
2. Run sequence of metadata checks (same as current)
3. All logs added to IM CMS
4. Metadata loaded to import database
5. RM notified visual check in import is needed
6. RM manager decides which files continue to next steps

# RM: Notification calls datamanager to action

Import Batch
Replication Manager
83.163.127.252:8080/rm/BatchesHistory
Search

<input type="checkbox"/>	2018-06-14 16:59:48	demo12.zip
<input type="checkbox"/>	2018-06-14 16:59:52	demo13.zip
<input type="checkbox"/>	2018-06-14 16:59:55	demo14.zip
<input type="checkbox"/>	2018-06-14 17:00:00	demo15.zip
<input type="checkbox"/>	2018-06-14 16:59:09	demo6.zip
<input type="checkbox"/>	2018-06-14 16:59:24	demo7.zip
<input type="checkbox"/>	2018-06-14 16:59:28	demo8.zip
<input type="checkbox"/>	2018-06-14 16:59:33	demo9.zip

### Current batch

name	batch_number	request_id	sent date	Batch global status	Batch CDIs status	Batch Data status	CDIs files
demo4	325	939cb7b3-5da5-46c6-95e5-dbed02ef610d	2018-06-14 16:59:48	<a href="#">CMS [HUMAN_ACTIVITY_METADATA]</a> Human action concerning metadata is required on IM CMS	[ARCHIVED] CDIs batch harvested and archived	[IN_QUEUE_ZONE] Moved to Data queue directory	<a href="#">demo4</a> (15 files)

### Batches in queue

name	batch_number	request_id	sent date	Batch global status	Batch CDIs status	Batch Data status	CDIs files
demo5			2018-06-20 10:35:35	[IN_QUEUE_READY] In queue, CDIs and data ready in queue	[CHECK_READY] Check is done	[IN_QUEUE_ZONE] Moved to Data queue directory	<a href="#">demo5</a> (15 files)

### Batches cancelled

CDIs



# IM: Ready to visually check logs and interface (only human action)

Import Batch

Replication Manager

importmanager.seadatanet.org/v\_import\_manager\_v5/content.asp?screen=4

Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

Free search

SEARCH CLEAR

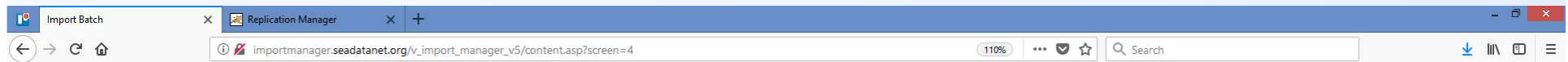
Found 19 Show (1-10)

Key	Input_date	Last_update	Batch_test	Batch_status	Metadata_files_count	Metadata_files_ok	N_code	Edit
325	19/6/2018 17:16 PM	20/6/2018 10:51 AM	True	(2300) Wait for human visual check in import database Import_database	15	15	<a href="#">link</a>	
324	19/6/2018 15:18 PM	19/6/2018 15:37 PM	True	(8920) Batch in production; Ready for a new batch	15	13	-	
323	15/6/2018 13:33 PM	19/6/2018 15:12 PM	True	(2180) Removed from import database	15	15	-	
322	15/6/2018 12:22 PM	15/6/2018 13:06 PM	True	(8920) Batch in production; Ready for a new batch	15	15	-	
317	11/6/2018 13:21 PM	11/6/2018 13:57 PM	True	(8920) Batch in production; Ready for a new batch	100	80	-	
316	11/6/2018 11:55 AM	11/6/2018 12:38 PM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
314	11/6/2018 11:37 AM	11/6/2018 11:53 AM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
313	11/6/2018 10:20 AM	11/6/2018 11:36 AM	True	(8920) Batch in production; Ready for a new batch	26	23	-	
311	7/6/2018 10:32 AM	11/6/2018 10:19 AM	True	(8920) Batch in production; Ready for a new batch	3	3	-	





# IM: Status progress and logs XML validation available



Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

## IMPORT BATCH

Free search

SEARCH CLEAR

CURRENT ACTION STATUS OVERVIEW METADATA LOGFILES

Results Found 19 Show 1 Update Clear

Key:325 Created:19/6/2018 17:16:07 Last update:20/6/2018 10:41:06

Import log files

```
2018-06-20 10:40:08
=====
-- start of program 6/20/2018 10:40:08
=====
2018-06-20 10:40:08 Batch.number = 325
2018-06-20 10:40:08 Batch.Edmo_code = 1526
2018-06-20 10:40:08 Batch.Directory = e:\Internet\seadatanet\CDI_V5_bestanden_aangeleverd\1526\1526_325
2018-06-20 10:40:08 Batch.files = 0
2018-06-20 10:40:08 Batch.delete = False
2018-06-20 10:40:08
2018-06-20 10:40:13 =====
2018-06-20 10:40:13 Number of records for duplicate check = 15
2018-06-20 10:40:13 =====
2018-06-20 10:40:14
2018-06-20 10:40:14 -----
2018-06-20 10:40:14
2018-06-20 10:40:14 Files Inserted = '15'
2018-06-20 10:40:14 Files error = '0'
2018-06-20 10:40:14 Batch duplicates (identifiers) = '0'
2018-06-20 10:40:14 Potential Batch duplicates = '0'
2018-06-20 10:40:14 Potential Batch Production duplicates = '0'
2018-06-20 10:40:14
```

# IM: Visual check - Individual records can be removed


Import Batch

Seadatanet V5

Replication Manager

cdi\_import\_test.maris.nl/v\_cdi\_v5/welcome.php?query=325&query\_code={F581FF35-B152-405D-8426-420428B2FF70}

Search



PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

PETER THIJSSE

DELETE RECORDS

NEW SEARCH

SEARCH RESULTS

SUMMARY

TIMESERIES

Filter search

You searched for:

Batch number:

325

Reset all

INPUT FIELDS

Free search ?

Date search ?

From

yyyymmdd

To

yyyymmdd

Geographic search ?

North

West

East

South

SEARCH

RESET

SEA REGIONS

North Sea (15)

Northeast Atlantic Ocean (40W) (15)

Western Scheldt (14)

DISCIPLINE (P08)

Physical oceanography (15)

PARAMETER GROUP (P03)

Water column temperature a... (15)

DISCOVERY PARAMETER (P02)

20

100

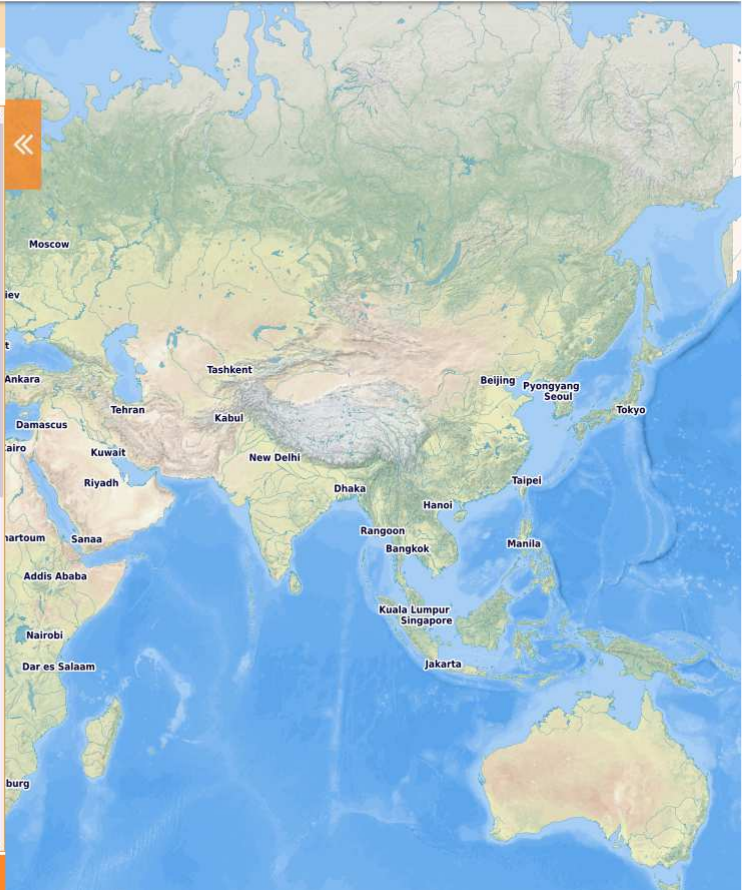
1000

found 15 | show (1 - 15) | Prev | Next

Dataset name	Country	Start date	Instrument / gear type
<input type="checkbox"/> Measurements at Bath in 1913 by RIKZMON_TEMP using THER MMTR	Netherlands	19130101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1919 by RIKZMON_TEMP using THER MMTR	Netherlands	19190101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1925 by RIKZMON_TEMP using THER MMTR	Netherlands	19250101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1931 by RIKZMON_TEMP using THER MMTR	Netherlands	19310101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1937 by RIKZMON_TEMP using THER MMTR	Netherlands	19370101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1943 by RIKZMON_TEMP using THER MMTR	Netherlands	19430101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1949 by RIKZMON_TEMP using THER MMTR	Netherlands	19490101	water temperature sensor
<input type="checkbox"/> Measurements at Bath in 1955 by RIKZMON_TEMP using THER MMTR	Netherlands	19550101	water temperature sensor

EXPORT RESULT

SAVE QUERY





## IM: RM manager indicates check is done (go on or cancel batch)

Import Batch | Seadatanet V5 | Replication Manager | +

importmanager.seadatanet.org/v\_import\_manager\_v5/content.asp?screen=4

Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

### IMPORT BATCH

Free search

SEARCH CLEAR

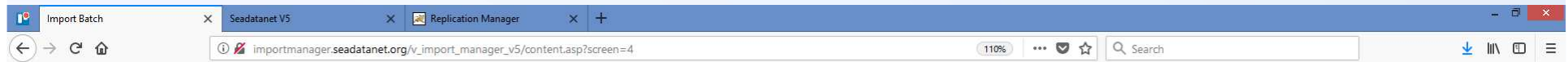
CURRENT ACTION | STATUS OVERVIEW | METADATA LOGFILES

Results Found 19 Show 1 Update Clear

Key:325 Created:19/6/2018 17:16:07 Last update:20/6/2018 10:51:05

Start_api_request_n_code	3814
Batch_status	(2300) Wait for human visual check in import database Import_database
Batch_status	<input type="radio"/> Remove from import database
	<input checked="" type="radio"/> Ready with human visual check in import database
Batch_test	1
Metadata_path_name	e:\internet\seadatanet\CDI_V5_bestanden_aangeleverd\1526\1526_325\
Metadata_zip_file_name	demo4.zip
Metadata_files_count	15
Metadata_files_ok	15
Metadata_files_error	0
Metadata_duplicates_identifiers	0
Metadata_duplicates_batch	0
Metadata_duplicates_production	0
Metadata_duplicates_batch_log	
Metadata_duplicates_production_log	

# IM: Confirmed and programmes take over again












Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

## IMPORT BATCH

Free search

SEARCH CLEAR

Key -	Input_date -	Last_update -	Batch_test -	Batch_status -	Metadata_files_count -	Metadata_files_ok -	N_code -	Edit
325	19/6/2018 17:16 PM	20/6/2018 11:05 AM	True	(2320) Ready with human visual check in import database	15	15	-	
324	19/6/2018 15:18 PM	19/6/2018 15:37 PM	True	(8920) Batch in production; Ready for a new batch	15	13	-	
323	15/6/2018 13:33 PM	19/6/2018 15:12 PM	True	(2180) Removed from import database	15	15	-	
322	15/6/2018 12:22 PM	15/6/2018 13:06 PM	True	(8920) Batch in production; Ready for a new batch	15	15	-	
317	11/6/2018 13:21 PM	11/6/2018 13:57 PM	True	(8920) Batch in production; Ready for a new batch	100	80	-	
316	11/6/2018 11:55 AM	11/6/2018 12:38 PM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
314	11/6/2018 11:37 AM	11/6/2018 11:53 AM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
313	11/6/2018 10:20 AM	11/6/2018 11:36 AM	True	(8920) Batch in production; Ready for a new batch	26	23	-	
311	7/6/2018 10:32 AM	11/6/2018 10:19 AM	True	(8920) Batch in production; Ready for a new batch	3	3	-	

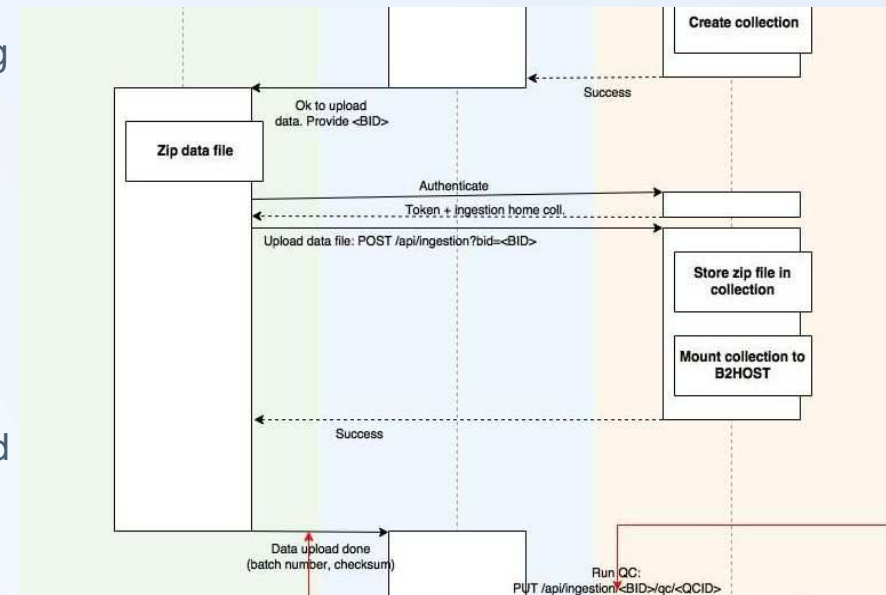
## Step 4: Uploading a databatch (Unrestricted only!)

### The problem

IM requests to RM to upload specified datafiles (matching CDI's with datafiles). RM has one zip file with data (!) to upload

### The solution

1. RM holds files in a folder
2. RM receives from IM a Batch ID (BID), CDI\_identifiers, and formats to use for upload
3. IM creates BID and creates the virtual folder on cloud
4. RM uploads the zip file to the folder in cloud
5. HTTP API confirms upload completed to RM
6. RM notifies Import Manager that upload is ready.





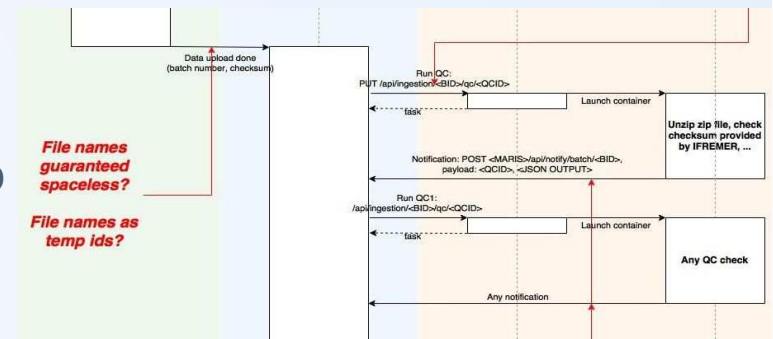
## Step 5: Unzip and run datachecks

## The problem

QC developer wants to run an action (unzip/quality check) on every uploaded batch

## The solution

1. QC Developer writes a program to test each file of a specific folder
  1. QC developers pushes the container for this binary to EUDAT HUB
  2. QC identifier available in the end (QCID)
2. Import Manager calls PUT on /api/ingestion/BID/qc/QCID
3. HTTP API runs an asynchronous task T for QCID mounting BID folder
4. T ends and calls POST on MARIS API with:
  1. The JSON output from the task/JAVA code incl QCID
5. Next action



## Technical QC implementation status

- Checks implemented:
  - Compare checksum of file
  - Compare length of file
  - Unzip file
  - Compare nr of files
  - Check length of files > 0
  - Check for each CDI\_identifier if requested files and formats exist
- All errors reported back to IM, logs are user and shown in CMS.
- Erroneous records removed from batch, but batch moves forward as long as ok records exist.
- No more records, or batch cancelled by data manager, start again.



Followed by a sequence of additional “QC checks”

- ODV quality checks
- Conversion to other formats (if not available yet)
- P01 Extraction
- Format syntax check

Facilitated by Octopus, plus additional programmes



IM: Unzip and checks are done, left-over files ready to be moved.  
RM manager should check logs and confirm

Import Batch

Replication Manager

importmanager.seadatanet.org/v\_import\_manager\_v5/content.asp?screen=4

110%

Search

Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

IMPORT BATCH

Free search

SEARCH CLEAR

Found 20 Show (1-10)

Key	Input_date	Last_update	Batch_test	Batch_status	Metadata_files_count	Metadata_files_ok	N_code	Edit
326	20/6/2018 12:28 PM	20/6/2018 13:26 PM	True	(7000) Wait for human approval go to production	15	14	-	
325	19/6/2018 17:16 PM	20/6/2018 12:28 PM	True	(2180) Removed from import database	15	14	-	
324	19/6/2018 15:18 PM	19/6/2018 15:37 PM	True	(8920) Batch in production; Ready for a new batch	15	13	-	
323	15/6/2018 13:33 PM	19/6/2018 15:12 PM	True	(2180) Removed from import database	15	15	-	
322	15/6/2018 12:22 PM	15/6/2018 13:06 PM	True	(8920) Batch in production; Ready for a new batch	15	15	-	
317	11/6/2018 13:21 PM	11/6/2018 13:57 PM	True	(8920) Batch in production; Ready for a new batch	100	80	-	
316	11/6/2018 11:55 AM	11/6/2018 12:38 PM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
314	11/6/2018 11:37 AM	11/6/2018 11:53 AM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
313	11/6/2018 10:20 AM	11/6/2018 11:36 AM	True	(8920) Batch in production; Ready for a new batch	26	23	-	

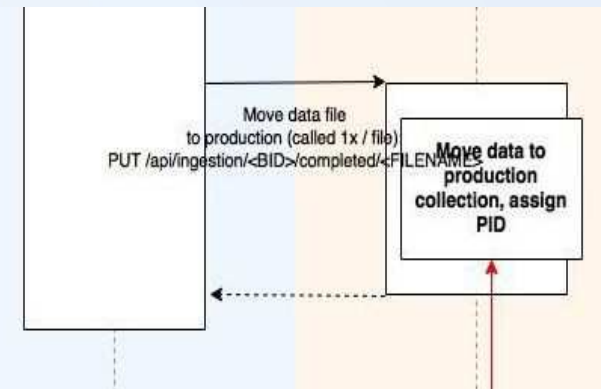
## Step 6: Move validated data in production

The problem

- Import manager has a matrix of Quality checks for the batch
- All batch records that passed all quality checks is to move in production

The solution

1. IM calls HTTP API with the filename for each approved file
2. HTTP API asynchronously triggers the ingestion process of IRODS
3. iRODS moves the collection to Production
4. Automatic irules are triggered for each file to generate the PID
5. HTTP API supplies IM a JSON with filenames and their PIDs (per format)





# IM: Human approval to move to production

Import Batch

Replication Manager

importmanager.seadatanet.org/vv\_import\_manager\_v5/content.asp?screen=4

110%

Search

Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

IMPORT BATCH

Free search

SEARCH

CLEAR

CURRENT ACTION

STATUS OVERVIEW

METADATA LOGFILES

Results

Found 20

Show 1

Update

Clear

Key:326 Created:20/6/2018 12:28:53 Last update:20/6/2018 13:26:57

Start\_api\_request\_n\_code

4849

Batch\_status

(7000) Wait for human approval go to production

Batch\_status

Remove from import database

Ready human approval go to production

Batch\_test

1

Metadata\_path\_name

e:\internet\seadatanet\CDI\_V5\_bestanden\_aangeleverd\1526\1526\_326\

Metadata\_zip\_file\_name

demo5.zip

Metadata\_files\_count

15

Metadata\_files\_ok

14

Metadata\_files\_error

1

Metadata\_duplicates\_identifiers

0

Metadata\_duplicates\_batch

0

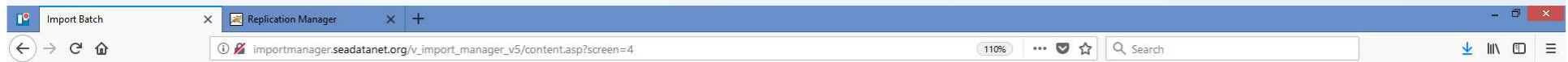
Metadata\_duplicates\_production

0

Metadata\_duplicates\_batch\_log

Metadata\_duplicates\_production\_log

# IM: Ready to go for final steps













Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

## IMPORT BATCH

Free search

SEARCH CLEAR

Key -	Input_date -	Last_update -	Batch_test -	Batch_status -	Metadata_files_count -	Metadata_files_ok -	N_code -	Edit
326	20/6/2018 12:28 PM	20/6/2018 13:46 PM	True	(7020) Ready human approval go to production	15	14	-	
325	19/6/2018 17:16 PM	20/6/2018 12:28 PM	True	(2180) Removed from import database	15	14	-	
324	19/6/2018 15:18 PM	19/6/2018 15:37 PM	True	(8920) Batch in production; Ready for a new batch	15	13	-	
323	15/6/2018 13:33 PM	19/6/2018 15:12 PM	True	(2180) Removed from import database	15	15	-	
322	15/6/2018 12:22 PM	15/6/2018 13:06 PM	True	(8920) Batch in production; Ready for a new batch	15	15	-	
317	11/6/2018 13:21 PM	11/6/2018 13:57 PM	True	(8920) Batch in production; Ready for a new batch	100	80	-	
316	11/6/2018 11:55 AM	11/6/2018 12:38 PM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
314	11/6/2018 11:37 AM	11/6/2018 11:53 AM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
313	11/6/2018 10:20 AM	11/6/2018 11:36 AM	True	(8920) Batch in production; Ready for a new batch	26	23	-	
311	7/6/2018 10:32 AM	11/6/2018 10:19 AM	True	(8920) Batch in production; Ready for a new batch	3	3	-	



# IM: Success! Ready for a new batch! (if next one in queue it automatically continues)

Import Batch

Replication Manager

Seadatanet V5

importmanager.seadatanet.org/v\_import\_manager\_v5/content.asp?screen=4

110%

Search

Login code (1526) Rijkswaterstaat Water, Traffic and Environment

Import Batch

## IMPORT BATCH

Free search

SEARCH CLEAR

Found 20 Show (1-10)

Key	Input_date	Last_update	Batch_test	Batch_status	Metadata_files_count	Metadata_files_ok	N_code	Edit
326	20/6/2018 12:28 PM	20/6/2018 14:00 PM	True	(8920) Batch in production; Ready for a new batch	15	14	-	
325	19/6/2018 17:16 PM	20/6/2018 12:28 PM	True	(2180) Removed from import database	15	14	-	
324	19/6/2018 15:18 PM	19/6/2018 15:37 PM	True	(8920) Batch in production; Ready for a new batch	15	13	-	
323	15/6/2018 13:33 PM	19/6/2018 15:12 PM	True	(2180) Removed from import database	15	15	-	
322	15/6/2018 12:22 PM	15/6/2018 13:06 PM	True	(8920) Batch in production; Ready for a new batch	15	15	-	
317	11/6/2018 13:21 PM	11/6/2018 13:57 PM	True	(8920) Batch in production; Ready for a new batch	100	80	-	
316	11/6/2018 11:55 AM	11/6/2018 12:38 PM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
314	11/6/2018 11:37 AM	11/6/2018 11:53 AM	True	(8920) Batch in production; Ready for a new batch	2	1	-	
313	11/6/2018 10:20 AM	11/6/2018 11:36 AM	True	(8920) Batch in production; Ready for a new batch	26	23	-	



# Batch active in interface (example record – in test)


Import Batch

Replication Manager

Seadatanet V5

cdi\_front\_test/vr\_cdi\_v5/welcome.php

Search



PAN-EUROPEAN INFRASTRUCTURE FOR  
OCEAN & MARINE DATA MANAGEMENT

LOGIN

DATASET BASKET 0

NEW SEARCH

SEARCH RESULTS

SUMMARY

DETAILS

Results | show (1 - 1) | Prev | Next

WHAT?

Data set name

Discipline

Parameter groups

Discovery parameter

P01

GEMET-INSPIRE themes

Abstract

Data format

Data set creation date

WHERE?

Map

Measurements at Bath in 1955 by RIKZMON\_TEMP using THERMMTR

Physical oceanography

Water column temperature and salinity

Temperature of the water column

Julian Date (chronological)


Temperature of the water body by in-situ thermometer

Oceanographic geographical features

BATH-1955

Ocean Data View ASCII input **Version 0.3**

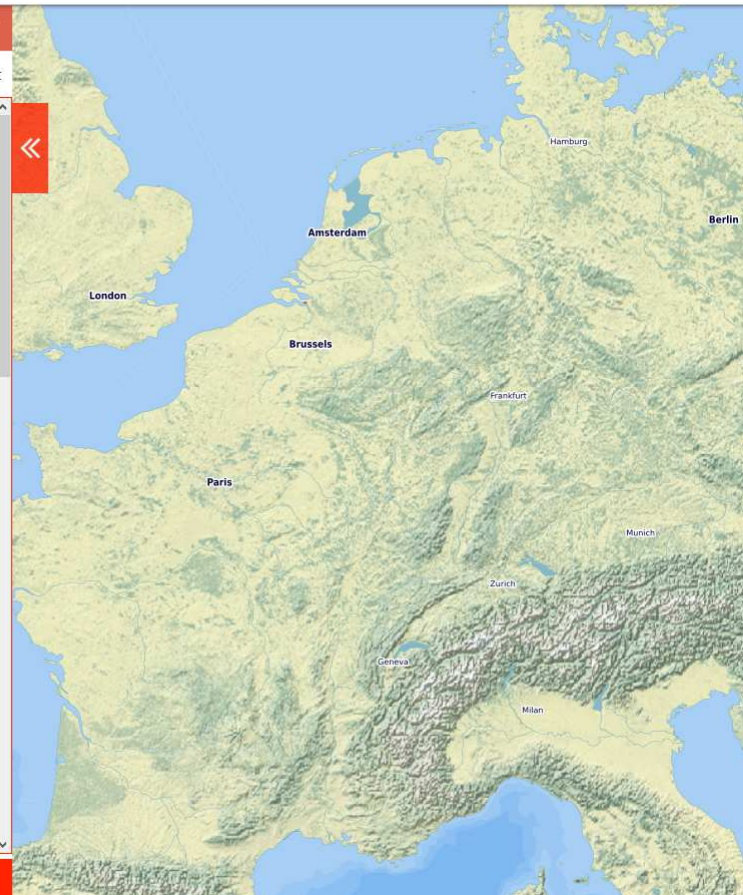
20140306



ADD TO BASKET

XML

PRINT





## Restricted data (Recently finalised)

- ✓ CDI metadata same flow
- ✓ Data files extracted at same time as metadata. Copy stored (for each version)
- ✓ Files will be temporarily released to cloud directory of user upon request RSM
- ✓ When released: Added all to same order-zipfile in the cloud

## Other recent actions completed

- ✓ Updates of RM and IM software
- ✓ Development version still at CINECA
- ✓ More stable test (pre-production) environment with TEST and IMPORT-PUBLIC versions at CSC environment.
- ✓ Risk analysis and completed calls/error responses to make the developed system processes full proof

## Ongoing actions – Replication manager

- Check minimum RAM: Now 4GB, ok for 2 installations side by side (test and production)
- Large file upload test (split if too large)
- Large batches test
- Manual/documentation update
- Consistency/integrity checks - what if files in order/request not available => Monitoring plan
- First stable version release to 5 test partners – expected around 20 November

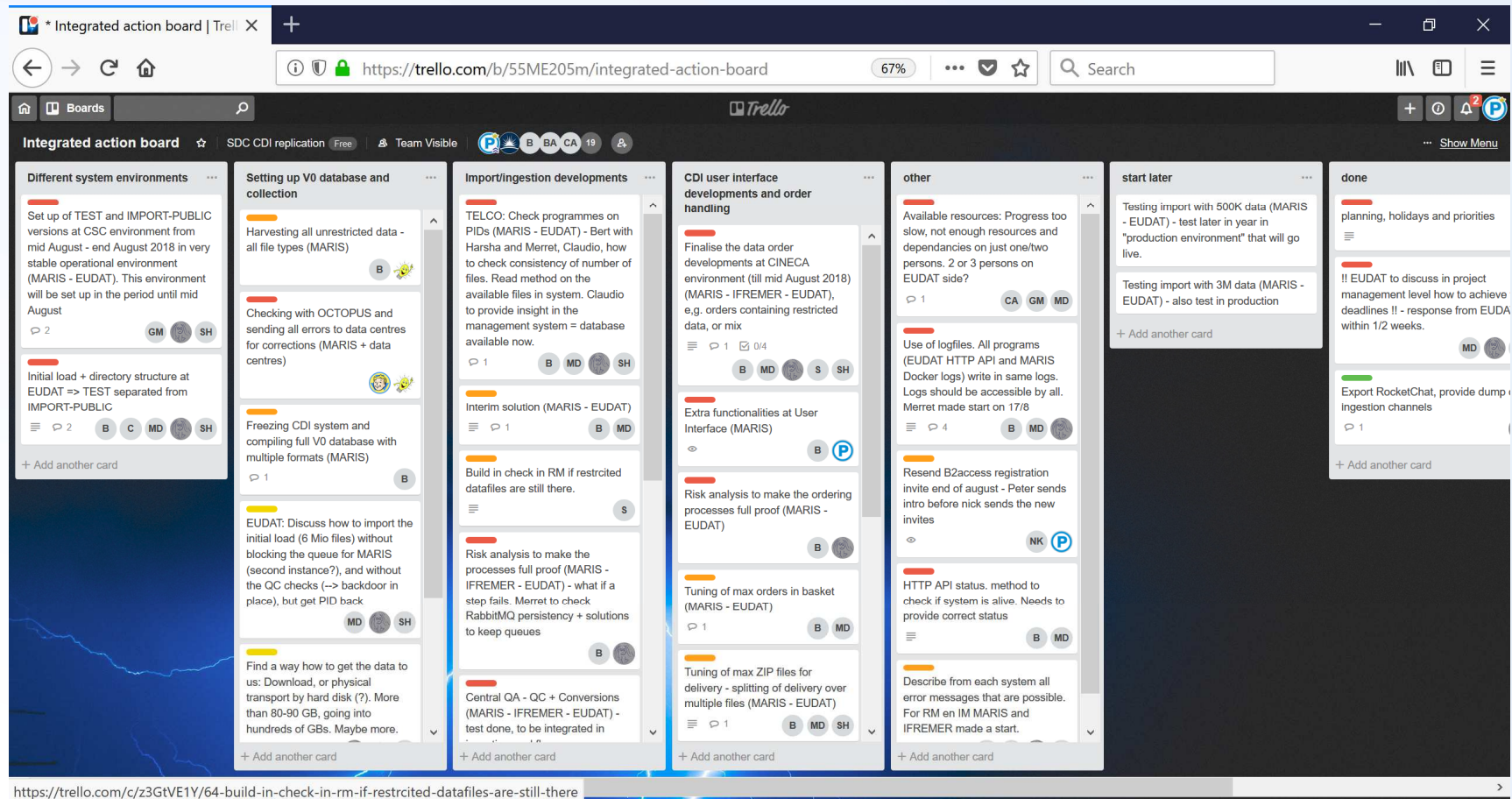
## Ongoing actions – EUDAT cloud

- Stability improvement
- Authorisation hick-ups
- Test with workers – now 5 (ingestion and orders combined?), is this sufficient?
- Test and production environment:
  - RM and IM API calls have a switch Test/Prod, needs balancer:
    - If test => go to test cloud
    - If prod => go to production
- Performance of ingestion and order handling optimization
- Initial load of files (500K first, then 5M files)
- Consistency/integrity checks - what if files in order/request not available => Monitoring plan

## Ongoing actions - Import Manager

- Test vs Production calls testing
- Consistency/integrity checks - what if files in order/request not available => Monitoring plan
- Integration tests of ingestion and order workflows

# In more detail on our action board



The screenshot shows a Trello board titled "Integrated action board" with a dark theme. The board is organized into several columns, each representing a different stage or category of work. The columns are:

- Different system environments**: Contains cards about setting up TEST and IMPORT-PUBLIC versions at CSC environment from mid August - end August 2018, and initial load + directory structure at EUDAT => TEST separated from IMPORT-PUBLIC.
- Setting up V0 database and collection**: Contains cards about harvesting all unrestricted data - all file types (MARIS), checking with OCTOPUS and sending all errors to data centres for corrections (MARIS + data centres), freezing CDI system and compiling full V0 database with multiple formats (MARIS), and discussing how to import the initial load (6 Mio files) without blocking the queue for MARIS (second instance?), and without the QC checks (-> backdoor in place), but get PID back.
- Import/ingestion developments**: Contains cards about TELCO: Check programmes on PIDs (MARIS - EUDAT) - Bert with Harsha and Merret, Claudio, how to check consistency of number of files. Read method on the available files in system. Claudio to provide insight in the management system = database available now, interim solution (MARIS - EUDAT), build in check in RM if restricted datafiles are still there, risk analysis to make the processes full proof (MARIS - IFREMER - EUDAT) - what if a step fails. Merret to check RabbitMQ persistency + solutions to keep queues, and central QA - QC + Conversions (MARIS - IFREMER - EUDAT) - test done, to be integrated in.
- CDI user interface developments and order handling**: Contains cards about finalising the data order developments at CINECA environment (till mid August 2018) (MARIS - IFREMER - EUDAT), e.g. orders containing restricted data, or mix, extra functionalities at User Interface (MARIS), risk analysis to make the ordering processes full proof (MARIS - EUDAT), tuning of max orders in basket (MARIS - EUDAT), and tuning of max ZIP files for delivery - splitting of delivery over multiple files (MARIS - EUDAT).
- other**: Contains cards about available resources: Progress too slow, not enough resources and dependencies on just one/two persons. 2 or 3 persons on EUDAT side?, use of logfiles. All programs (EUDAT HTTP API and MARIS Docker logs) write in same logs. Logs should be accessible by all. Merret made start on 17/8, resend B2access registration invite end of august - Peter sends intro before nick sends the new invites, HTTP API status. method to check if system is alive. Needs to provide correct status, and describe from each system all error messages that are possible. For RM en IM MARIS and IFREMER made a start.
- start later**: Contains cards about testing import with 500K data (MARIS - EUDAT) - test later in year in "production environment" that will go live, and testing import with 3M data (MARIS - EUDAT) - also test in production.
- done**: Contains cards about planning, holidays and priorities, EUDAT to discuss in project management level how to achieve deadlines !! - response from EUDA within 1/2 weeks, and export RocketChat, provide dump + ingestion channels.

The board also shows a search bar at the top right, a "Show Menu" button, and a "67%" progress indicator. The URL in the browser is <https://trello.com/b/55ME205m/integrated-action-board>.



Plus:

Create documentation

Further planning towards mid March:

- See implementation plan



- END -