



dcf-shiny – R Shiny application in support to Data Collection Frameworks

Blondel E. ¹, A. Bennici ², A. Ellenbroek ¹, M. Assante ³, A. Dell’Amico ³, P. Pagano ³

¹ UN-FAO, NFISI, Rome

² UN-FAO, NFIS, Rome

³ ISTI-CNR, Pisa

Contact: **Emmanuel Blondel**

Geographic Information Systems and R Expert, NFISI

✉ Emmanuel.Blondel@fao.org

Context & Challenges

- Increasing need of FAO data calls for collecting and/or collating datasets
- Support the active participation of
 - country members to submit data to regional and/or global contexts
 - regional organizations under FAO partnerships, eg. FIRMS Partnership and global data initiatives such as the FIRMS Tuna Atlas

Objective

- Builds a generic web-application to support Data Collection frameworks, configurable to specific application contexts
 - **Regional Data Base (RDB) contexts**
Country members submit country data for collation at regional levels
Use cases: WECAFC-FIRMS, RECOFI, FCWC
 - **Global Data Base contexts**

RFMOs submit tuna datasets to FAO
Use case: Global Tuna Atlas

Country members submit country data to FAO
Use cases: FAO questionnaires?

- Compound by
 - R components developed by NFISI/NFISS
 - Infrastructure components developed by i-Marine/D4science CNR partner
- Packaged into a simple user R shiny web-application
- Deployed using highly advanced D4Science 'ShinyProxy' e-infra enabling:
 - Standard user authentication
 - User/Roles management & customization to specific contexts, eg
 - RDB: country data managers + regional data manager + system admin
 - GTA: tRFMO data managers + GTA data manager + system admin

Overview

- Modular approach with **core generic modules** for:
 - **Users management**
 - Manage context-specific users properties: country, organization, etc
 - **Data calls management**
 - Create data calls on periodical basis (eg. Yearly) to collate data
 - **Data validation & pre-submission**
 - Test datasets validity, based on:
 - Standard data format specifications, eg. FIRMS data exchange format
 - CWP standards, other international standards (eg. ISO) and related best practices
 - Data call characteristics
 - Pre-submit the valid datasets
 - **Access to data submissions**
 - **My submissions** – for the data submitter
 - **All submissions** – for the data manager
 - **Data availability:** data submission indicators per task / reporting entity over time

Overview

- Modular approach with **thematic plugins** for:
 - **Data merge & upload**
 - Merge & Upload pre-submitted datasets into a database system
 - **Data publication & dissemination**
 - Publish & disseminate datasets based on FAIR data principles
 - Enabling and feeding thematic atlases, eg. web map viewers

Pilot use cases:

- Regional Data Bases:
 - WECAFIS: <https://www.fao.org/wecafc/data/wecafis/en>
 - RECOFI PILOT
- Global Data Base: FIRMS Global Tuna Atlas

- **Simple configuration with:**
 - Description of the target context: eg. WECAFC-FIRMS DCF
 - Reporting entities definition (name, list of possible values)
 - Setting of data tasks covered by the DCF:
 - Basic description, and references
 - Link to the Data Structure Definition to use for data validation
 - Configuration of software to interact with:
 - Data base (for data calls and data upload management)
 - Spatial Data Infrastructure – SDI (for data publication & enabling of geospatial atlases)



Core modules




Home dashboard


- Profile information: name / roles assigned
- Administration panel (for admin/manager only)

Welcome Emmanuel Blondel!










My profile



ASSIGNED ROLES
regional data manager
system administrator





ASSIGNED REPORTING ENTITIES

 Barbados[BRB]	 Martinique[MTQ]
 Cuba[CUB]	 Nicaragua[NIC]
 Dominica[DMA]	 Saint Lucia[LCA]
 Guadeloupe[GLP]	 Suriname[SUR]
 French Guyana[GUF]	


Current datacalls

Administration

Users

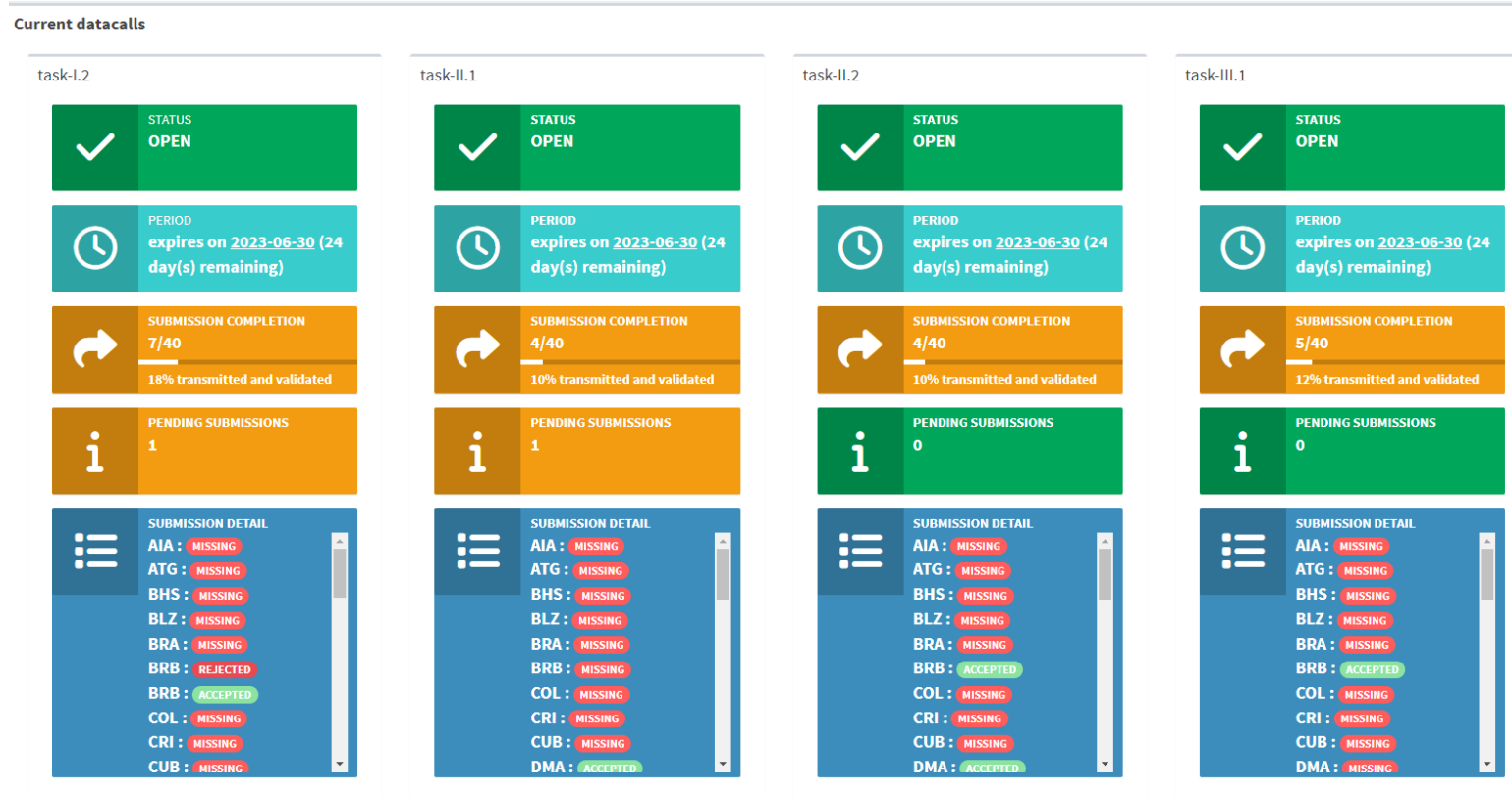
 <p>USERS 30</p>	<p>0</p> <p>Users without roles</p>
 <p>USERS IN DB 7</p>	<p>1</p> <p>SUBMITTER NOT ASSOCIATED TO A REPORTING ENTITY</p>

Reporting entities

 <p>REPORTING ENTITIES 40</p>	<p>31</p> <p>REPORTING ENTITIES NOT ASSIGNED TO A SUBMITTER</p>
---	--

Home dashboard

- Data calls summary:
 - For a data submitter: view over his/her current submissions status (by task)
 - For a data manager: view over all data submitters' submissions (by task)
(see below example)



Users management

- Master list of users / roles managed in VRE
- Extendable in app with reporting entities (eg. country, flagstates, organization) for finer control over data validation and submission processes:

Examples: Extend data validation with extra validation rules:

- flagstates data submission (1 country manager could report data for 1 or more flagstates)
- regional data controls (restrained by source organization)

Data calls management

- Restricted access to main data manager & system admin
- A data call is created in association to a DCF **data task**, open for a given **time period**. Dates (start/end) and status (open/closed) can be modified.
- Expired data calls are automatically closed (they can be re-opened and extended if needed)
- When a data call is created/updated, all target data managers are notified and invited to validate and submit their data.

Data validation & submission

- Service for data submitters to test data validation, and to submit data in the context of a data call (specific to a given data task)
- “Wizard” (step by step) approach
- Data validation:
 - Can be used as preparatory work, even if no data call is opened
 - If a data call is open, data submitter can submit its data to data manager
 - One data validation/submission per data task (data call) and reporting entity
 - Two types of validation:
 - Conformity with standards, including data exchange format (CWP RH/ FIRMS data exchange draft format), international codelists/classifications
 - Consistency with data call specifications:
 - Temporal extent of data expected: one year (N-2, N-1), historical series, etc
 - Check over reporting entities included in the dataset



Data validation & submission

Home

Welcome to the Data validation and Submission module

Within this module you you will be able to:

- Check the validity of your data against standards (CWP, FIRMS, etc)
- Submit your data as part of data calls

If you are ready, click on 'Start'

Start

Data validation & submission

1-Select your data

Task

Please select a task ▼

Reporting entity

Please select a reporting entity ▼

1-Select your data

Task

Nominal catches [task-I.2] ▼

✓ A data call is currently open for this task

Reporting entity

Barbados ▼

Data format

Simplified ▼

File to test

Choose file test_task_I_2_S_V_S.csv

Upload complete

Next

1. User selects a data task

User is informed if a data call is open (in which case data submission will be open)

2. User selects a reported entity (among entities for which he/she is assigned to submit data)

3. User selects a data format (simplified/generic)

4. User selects file and go next

Data validation & submission

User can preview dataset

1-Select your data

2-Preview

Please verify if data displayed correspond to the data to send. If it is, please click 'Next' to submit this file, otherwise click 'Previous' to select a new file.

Search:

fishingfleet ▾ flagstate ▾ year ▾ period ▾ geographic_identifier ▾ geographic_coordinates ▾ fleet_segment ▾ gear_type ▾ fishing_mode ▾ species ▾ school_type ▾

All

All

All

All

All

All

All

All

All

All

1	FRA	2011	Q1	WCA						FFV
2	TTO	2011	Q2	WCA						FFV
3	SUR	2011	Q3	WCA						FFV
4	BRB	2011	Q4	WCA						FFV
5	BRB	2012	Q1	WCA						FFV

Showing 1 to 5 of 40 entries

Previous

1

2

3

4

5

...

8

Next

Previous

Next

Data validation & submission

Conformity with standards is checked
including errors /warnings regarding
data structure & codelists

1-Select your data

2-Preview

3-Conformity with standards

✓ Data is valid

Congratulations! Your data passed the validation step. You can see below the data analysis details and click 'next'

Task ID: task-l.2

Task Name: Nominal catches

Format : simplified

Date of Report : 2022-10-28

File : test_task_l_2_S_V_S.csv

Approved for Upload : Yes

VALIDITY SUMMARY

Readable Dataset	PASSED	✓
Structure of Dataset	PASSED	✓
No missing Values	PASSED	✓
Conformity with Standards	PASSED	✓
Valid Reporting entity	PASSED	✓
Valid Dates	PASSED	✓
Skipped Information	WARNING	⚠

DETAILS

CSV

Search:

	Type	Rule	Row	Column	Category	Message
	All	All	All	All	All	All
1	INFO	I01	-	fishingfleet	Column not used	'fishingfleet' is not a mandatory column and will be skipped
2	INFO	I01	-	geographic_coordinates	Column not used	'geographic_coordinates' is not a mandatory column and will be skipped
3	INFO	I01	-	fleet_segment	Column not used	'fleet_segment' is not a mandatory column and will be skipped
4	INFO	I01	-	gear_type	Column not used	'gear_type' is not a mandatory column and will be skipped
5	INFO	I01	-	fishing_mode	Column not used	'fishing_mode' is not a mandatory column and will be skipped

Showing 1 to 5 of 7 entries

Previous 1 2 Next

Data validation & submission

Consistency with data call is checked

1-Select your data 2-Preview 3-Conformity with standards 4-Consistency with data Call

✖ Data is not consistent with the current data call

Oops something seems to be incorrect in your data. Please see below the data analysis details and click 'finish' to return to home page and try again.

VALIDITY SUMMARY

SE01	Consistency with data call period	PASSED	✓
SE02	Presence of mandatory years	PASSED	✓
SW01	Presence of historical serie	NOT TESTED	⊘
SE03	Consistency with reporting entities	FAILED	✖

DETAILS

CSV Search:

Type	Rule	Row	Column	Category	Message
			All	All	All
1	ERROR	SE03	-	flagstate	unique reporting entity
4 reporting entities are detected in the dataset (FRA,TTO,SUR,BRB) and only 'BRB' should be include					

Showing 1 to 1 of 1 entries Previous 1 Next

Download Report

Finish

Data validation & submission

1-Select your data

2-Preview

3-Conformity with standards

4-Consistency with data Call

✓ Data is ready to be submitted

Congratulations! Your data is consistent with the current data call. You can see below the data analysis details and click 'next'

VALIDITY SUMMARY

SE01	Consistency with data call period	PASSED	✓
SE02	Presence of mandatory years	PASSED	✓
SW01	Presence of historical serie	NOT TESTED	⊘
SE03	Consistency with reporting entities	PASSED	✓

DETAILS

No error detected

Download Report

Next

Data validation & submission

1-Select your data

2-Preview

3-Conformity with standards

4-Consistency with data Call

5-Send Data

You are going to send your data to the manager.

Validity reports (conformity with standards, consistency with data call) will be attached to the submission

You may also add notes to the submission here below. Once ready, click on 'Send' to proceed with the submission

Submission notes

Send

In case a data submission has already been done, user is informed if the previous submission should be overwritten

Caution a submission was already deposited for this datacall

Would you do really overwrite your precedent submission by this one ?

Click 'Update' to send and overwrite your precedent submission or click 'Cancel' to return to menu without sending your submission.

Yes

Cancel

Data validation - Reference data repository

- Based on the CWP Reference Harmonization standard:
 - digital implementation guidelines
 - Builds on a framework of **digital resources** and **data exchange format technical specifications** fostered by the Fisheries Data Interoperability adhoc working group (fdiwg) and its collaborative environment: <https://github.com/fdiwg>

Data validation - Reference data repository

- Resources
 - Dictionary of terms & definitions used in data structures and underlying resources
 - Codelists (global, regional)
 - Mappings
- Publicly available
 - Simple Github adhoc repositories (<https://github.com/fdiwg>) used for **operational** exploitation of reference data in all CWP pilot use cases (GTA, WECAFC-FIRMS RDB, Calipseo, yearbooks)
 - No fancy technology – keep it simple
 - Standard CSV file resource pushed upon needs

My submissions dashboard

- **As data submitter, I can :**
 - access and browse **my** data submissions in a dashboard
 - delete a data submission
- All data files are uploaded in my workspace, and not visible to other data submitters. They are shared with data managers in charge of the target database (eg. RDB manager)

My submissions

Task
Nominal catches [task-I.2]

Inspect your submissions for data call :
task-I.2 (2022-10-21/2022-12-24)



Refresh

Submission of entities 1 / 1

REMAINING ENTITIES 0

REJECTED SUBMIS... 0

Search:

Submission ID	Data call ID	Task ID	Flag	Reporting entity	Creation time	Last modified by	Last modification time	Status	Actions
All	All	All	All	All	All	All	All	All	All
ca98f21e-421c-48ba-9868-9ac39ba65d9b	4	task-I.2	ψ	BRB	2022-10-21T09:45:36Z	emmanuel.blondel	2022-10-21T09:45:36Z	SUBMITTED	 

Showing 1 to 1 of 1 entries

My submissions dashboard

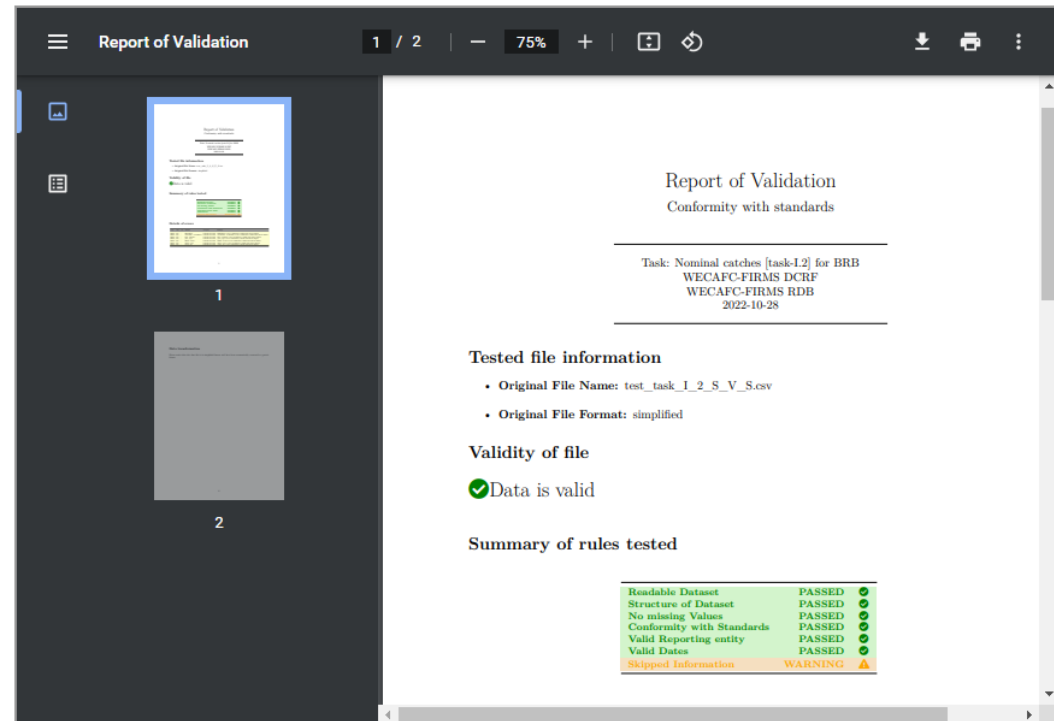
- Each data file (including metadata) attached to a data submission can be accessed and displayed including:
 - Source original data
 - Transformed data (into generic format if provided as simplified)
 - Standards conformity report
 - Data call consistency report
 - Metadata (as Dublin Core)

Show submitted item :

Please select a item ▼

Show submitted item :

datacall-4_task-task-
I.2_for_BRB_report_standard_conformity.pdf ▼



Report of Validation 1 / 2 | 75% + | [Icons]

Report of Validation
Conformity with standards

Task: Nominal catches [task-I.2] for BRB
WECAFC-FIRMS DCRF
WECAFC-FIRMS RDB
2022-10-28

Tested file information

- Original File Name: test_task_I_2_S_V_S.csv
- Original File Format: simplified

Validity of file

✓ Data is valid

Summary of rules tested

Readable Dataset	PASSED	✓
Structure of Dataset	PASSED	✓
No missing Values	PASSED	✓
Conformity with Standards	PASSED	✓
Valid Reporting entity	PASSED	✓
Valid Dates	PASSED	✓
Skipped Information	WARNING	⚠

All submissions dashboard







- As data manager, I can:
 - access and browse all data submissions in a dashboard
 - approve or reject a data submission
 - Access overall % of data call completion
 - send reminder to data submitters that still didn't send their data, on individual basis or to all

All submissions




























Task: Nominal catches [task-1.2]

Inspect submissions for data call: task-1.2 (2022-10-21/2022-12-24) [OPENED]

Refresh

 SUBMISSIONS 4/40 10% of completion	 PENDING SUBMISSIONS 3	 MISSING SUBMISSIONS 36 Click to send Reminder
 ACCEPTED SUBMISSIONS 2/40 5% of completion	 DUPLICATE SUBMISSIONS 1	 DAYS REMAINING 57 days

Search:

Submission ID	Data call ID	Data call Folder	Task ID	Flag	Reporting entity	Temporal extent	Submitter	Creation time	Last modified by	Last modification time	Status	Actions
4391523f-64e0-4bfe-89bb-54db34da9d9a	4	datacall-4_task-task-1.2_for_BRB	task-1.2		BRB	2011-01-01/2020-12-31	alexandre-ben	2022-10-24T09:11:40Z	alexandre-ben	2022-10-24T09:11:40Z	ACCEPTED	  
5545059a-cf45-419c-926e-d5ca99be6c92	4	datacall-4_task-task-1.2_for_GLP	task-1.2		GLP	2011-01-01/2020-12-31	alexandre-ben	2022-10-24T13:50:21Z	alexandre-ben	2022-10-24T13:50:21Z	ACCEPTED	  
ca98f21e-421c-48ba-9868-9ac39ba65d9b	4	datacall-4_task-task-1.2_for_BRB	task-1.2		BRB	2011-01-01/2020-12-31	emmanuel.blondel	2022-10-21T09:45:36Z	emmanuel.blondel	2022-10-21T09:45:36Z	SUBMITTED	   
53cbf294-fb75-4800-9d1c-534f0cdcfe04	4	datacall-4_task-task-1.2_for_GUF	task-1.2		GUF	2011-01-01/2020-12-31	alexandre-ben	2022-10-24T13:52:36Z	alexandre-ben	2022-10-24T13:52:36Z	SUBMITTED	   
c637e6c0-9848-4725-9d38-5c4d39978787	4	datacall-4_task-task-1.2_for_MTQ	task-1.2		MTQ	2011-01-01/2020-12-31	alexandre-ben	2022-10-26T08:10:49Z	alexandre-ben	2022-10-26T08:10:49Z	SUBMITTED	   
	4		task-1.2		AIA						MISSING	
	4		task-1.2		ATG						MISSING	

Data availability

- Present coverage indicators by task and reporting entity over time
- Examples of indicators:
 - Covered period - **summary by task / reporting entity**

Summary By Task

Compute summary ☒ Limit to entities with data

Statistic Covered period

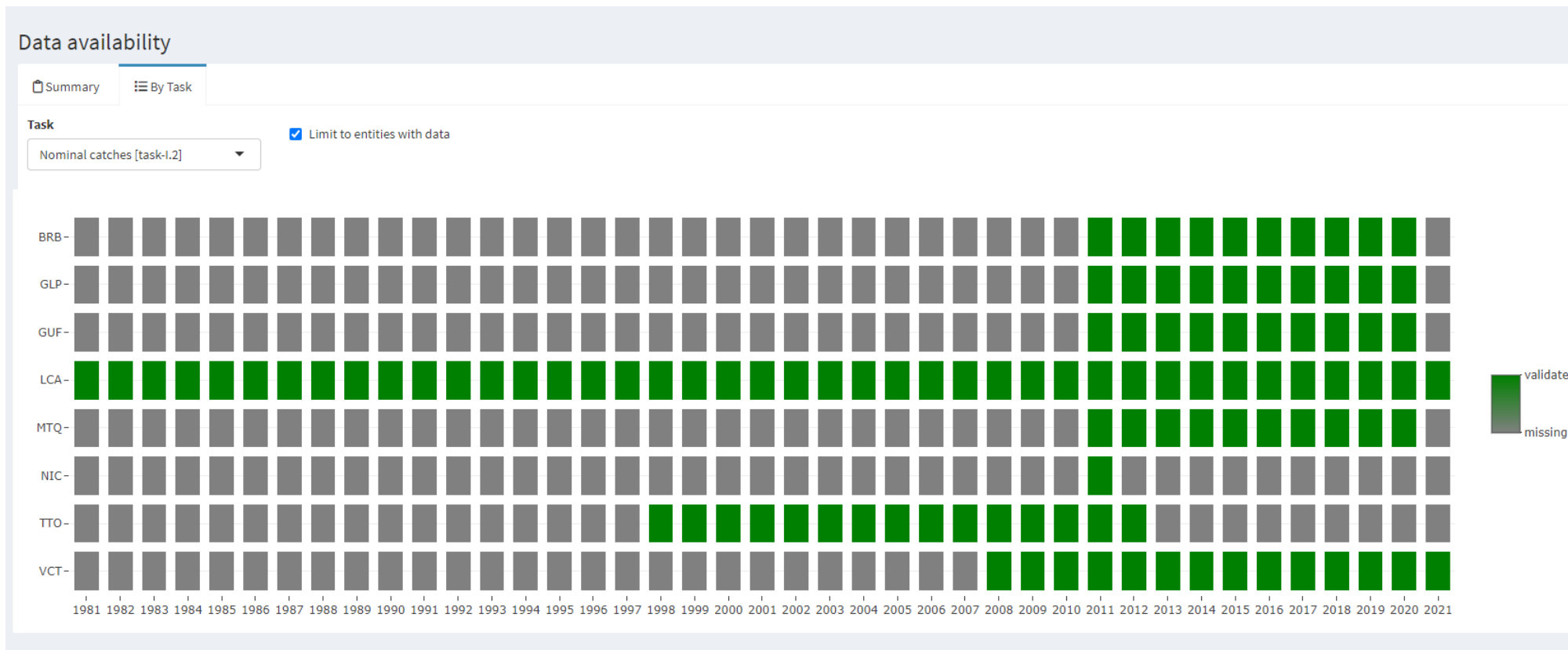
Download summary

	task-I.2	task-II.1	task-II.2	task-III.1
BRB -	first:2011- last:2020	(no data)	first:2015- last:2020	first:2015- last:2020
DMA -	(no data)	first:2013- last:2020	first:2013- last:2020	(no data)
GLP -	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020
GUF -	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020
LCA -	first:1981- last:2021	first:1991- last:1991	(no data)	(no data)
MTQ -	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020	first:2011- last:2020
NIC -	first:2011- last:2011	(no data)	(no data)	(no data)
SUR -	(no data)	first:2017- last:2021	first:2017- last:2021	(no data)
TTO -	first:1998- last:2012	first:2014- last:2014	first:2014- last:2014	first:2020- last:2020
VCT -	first:2008- last:2021	first:2008- last:2021	(no data)	(no data)

data
no data

Data availability

- Present coverage indicators by task and reporting entity over time
- Examples of indicators:
 - Covered period - **details by task**



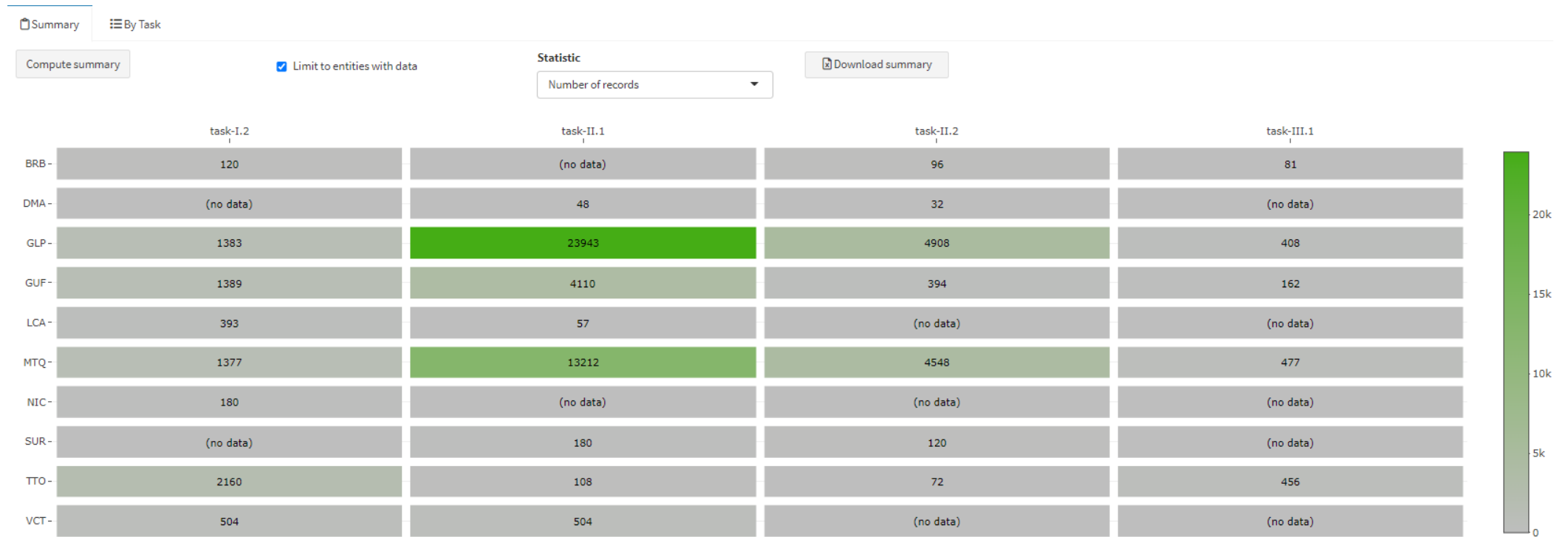
Data availability

- Present coverage indicators by task and reporting entity over time
- Examples of indicators:
 - Number of years covered by data



Data availability

- Present coverage indicators by task and reporting entity over time
- Examples of indicators:
 - Number of records

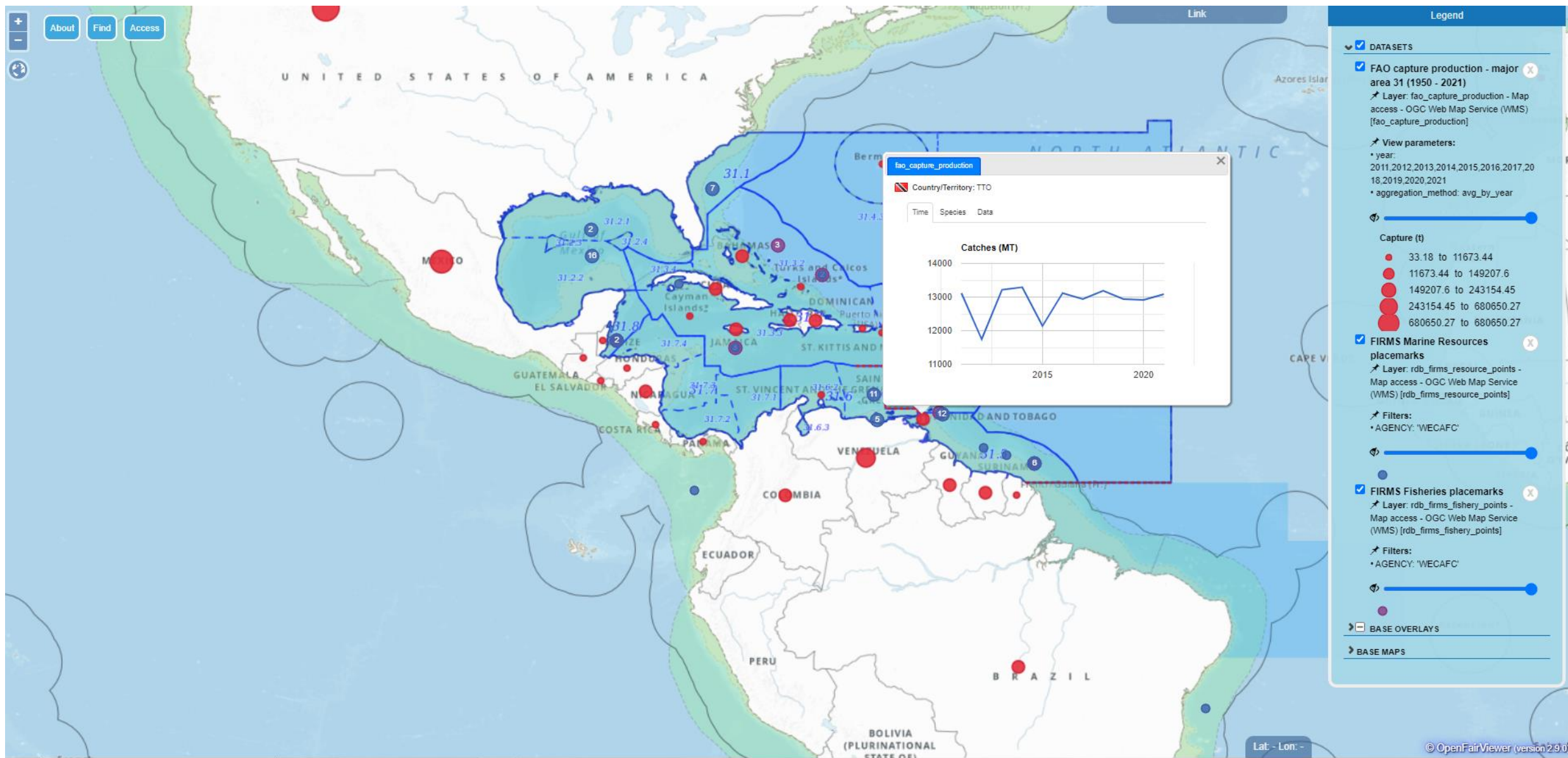




Thematic plugins

WECAFC-FIRMS fisheries atlas plugins

- Compiles country datasets on data task basis, into a regional dataset (timeseries) that progressively expands over the years
- Triggers a [*geoflow*](#) to publish spatialized dataset
 - Described by standard ISO 19115/19139 dataset metadata
 - Structurally described by standard ISO 19110/19139 structure metadata
 - Backed by OGC data services enabling rich multi-dimensional filtering
- Feeds the WECAFC-FIRMS map viewer application



FIRMS Global Tuna Atlas plugin

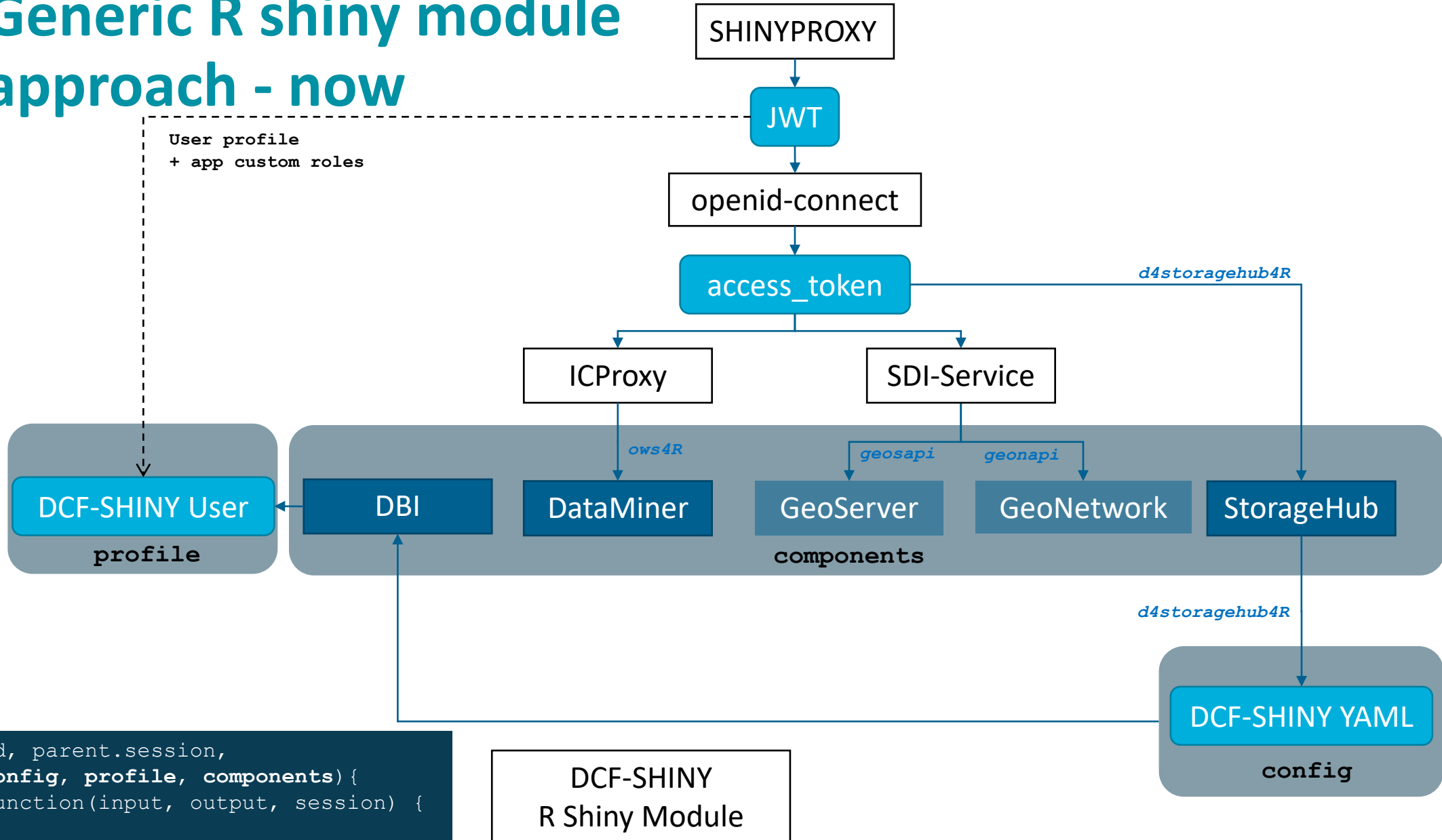
- Pilot plugin to be drafted to facilitate submission by tRFMOS for yearly updates of the FIRMS Global Tuna atlas
- Triggers a *geoflow* to publish spatialized dataset
 - Described by standard ISO 19115/19139 dataset metadata
 - Structurally described by standard ISO 19110/19139 structure metadata
 - Backed by OGC data services enabling rich multi-dimensional filtering
- Feeds the FIRMS Global Tuna Atlas map viewer application

Perspectives

- **Further integration with metadata**
 - Access to reference metadata (e.g. for codelists)
 - Submission of metadata information together with datasets
- **New modules**
 - Data entry (ongoing development)
 - Facilitate the data entry according to DCRF templates
 - On-the-fly data validation
 - Data task administration with
 - capacity to create/validate new tasks vs. draft data exchange standard
 - capacity to register data tasks for use in data calls
- **Use in other contexts**
 - At regional level, other RFBs/RFMOs: RECOFI? FCWC? IOTC?
 - At global level: FIRMS factsheets? FAO questionnaires?

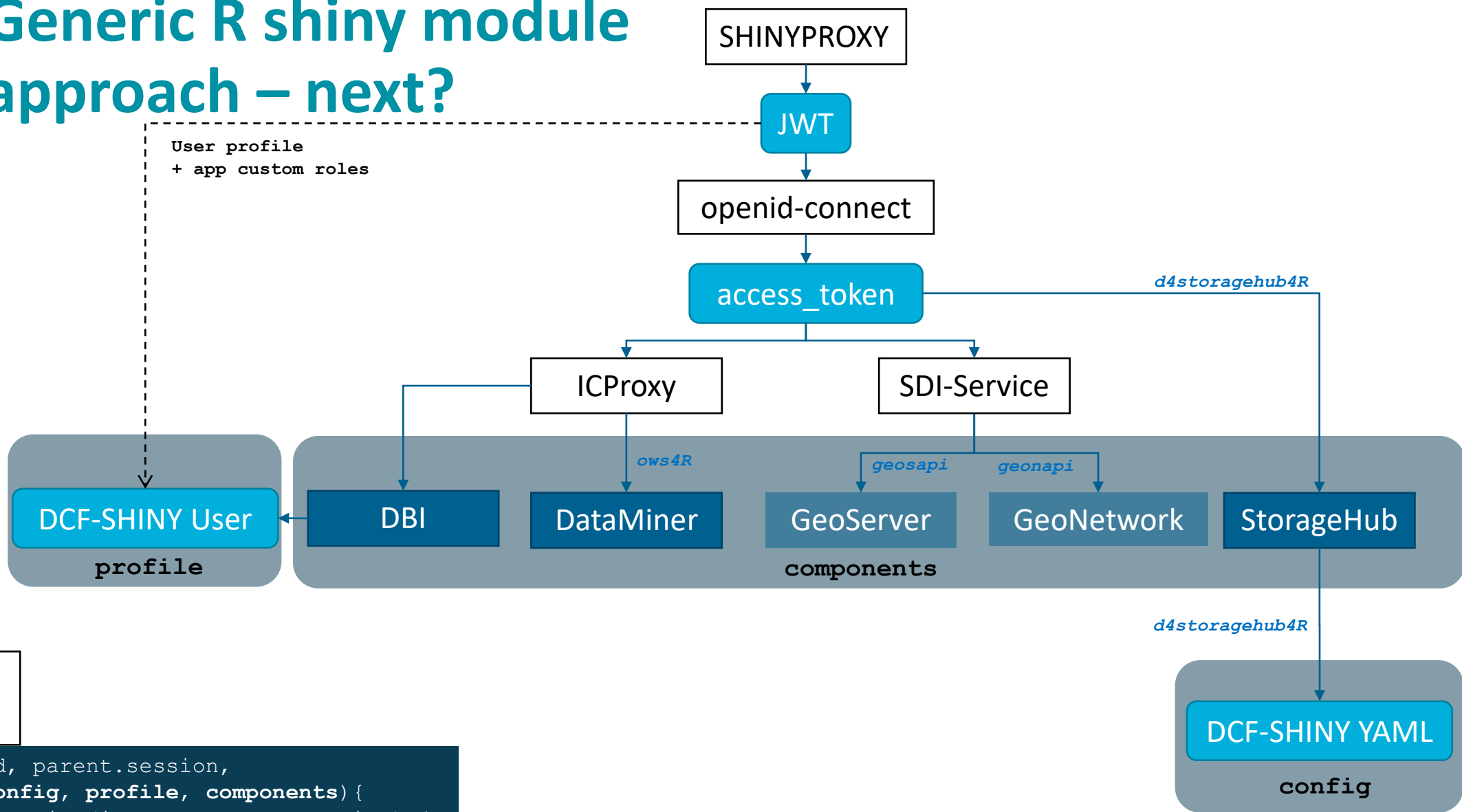


Generic R shiny module approach - now



```
module <- function(id, parent.session,
                   config, profile, components){
  moduleServer(id, function(input, output, session) {
    ...
  })
}
```

Generic R shiny module approach – next?



DCF-SHINY
R Shiny Module

```

module <- function(id, parent.session,
                   config, profile, components){
  moduleServer(id, function(input, output, session) {
    ...
  })
}

```