27-29 May 2024

imdis

International conference on **M**arine **D**ata and **I**nformation **S**ystems

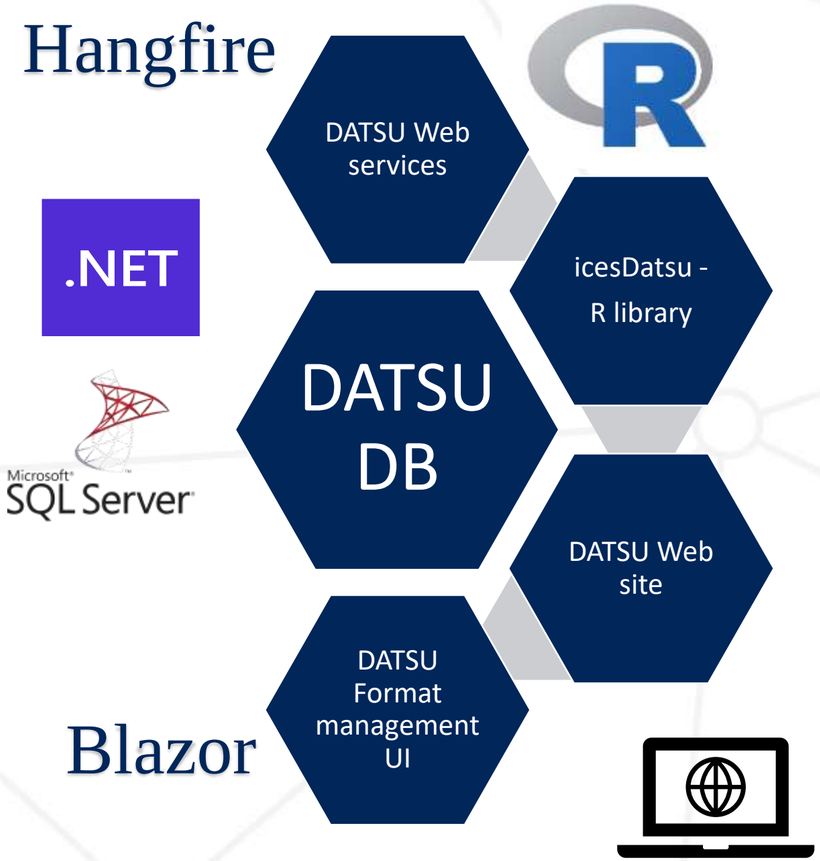
SeaDataNet AISBL · Havforskningsinstituttet · Ifremer · MARIS · MEDIN · National Oceanography Centre · INGV · iMagine · IODE · eosc Blue-Cloud2026

Streamlining Data Submission: A Guide to Ensuring Data Quality and Compliance with ICES Data Screening Utility (DATSU)

Laura Andreea Petre
Carlos Pinto
Anna Osypchuk
Colin Millar

Introduction to DATSU

- DATSU is the main data validation tool used by the ICES data centre.
- Designed to streamline and enhance data validation before uploading to ICES databases.
- Based on an SQL Server database (DATSU DB) with a user-friendly interface (DATSU UI).
- Utilizes web services for data submission and validation.



History

Early 2000s: Initial development with key components: DATSU DB, ACCESS-based UI, and DATSU VB6 DLL.

2007: Extension to include SQL rules in screening reports.

2016: Reprogrammed to run as .NET web services with asynchronous runs for improved performance.

2019: Introduction of the R library (icesDATSU) for local data checks.

2023: UI migrated to a C# .NET and Blazor-based web application with enhanced functionalities.

Components of DATSU

DATSU DB:

- Stores format definitions and screened data.
- Central repository for data validation rules and vocabularies.

DATSU Management User Interface (DATSU UI):

- Allows data managers to define data formats and quality checks.
- Provides an intuitive interface for managing data validation rules.

DATSU API (webservices):

- Web Interface for data submission and validation.
- Can be called synchronously or asynchronously to check data files.
- Supports integration with other applications and programming environments.

DATSU Web

- Allows data submitters to manually upload files for screening
- Allows viewing the screening report
- Exposes format information, data types, and vocabularies

icesDATSU – R Library

- Provides functions to interact with DATSU web services.
- Allows experts to incorporate DATSU checks into their data pre-processing workflow.
- Supports local file checking before submission.

Data validation and submission process

ICES Data managers

Data submitters

External server

Define new format on DATSU Format management UI

Check file locally with icesDATSU R library

Upload file for screening on DATSU Web

Direct HTTP POST call to DATSU Web service

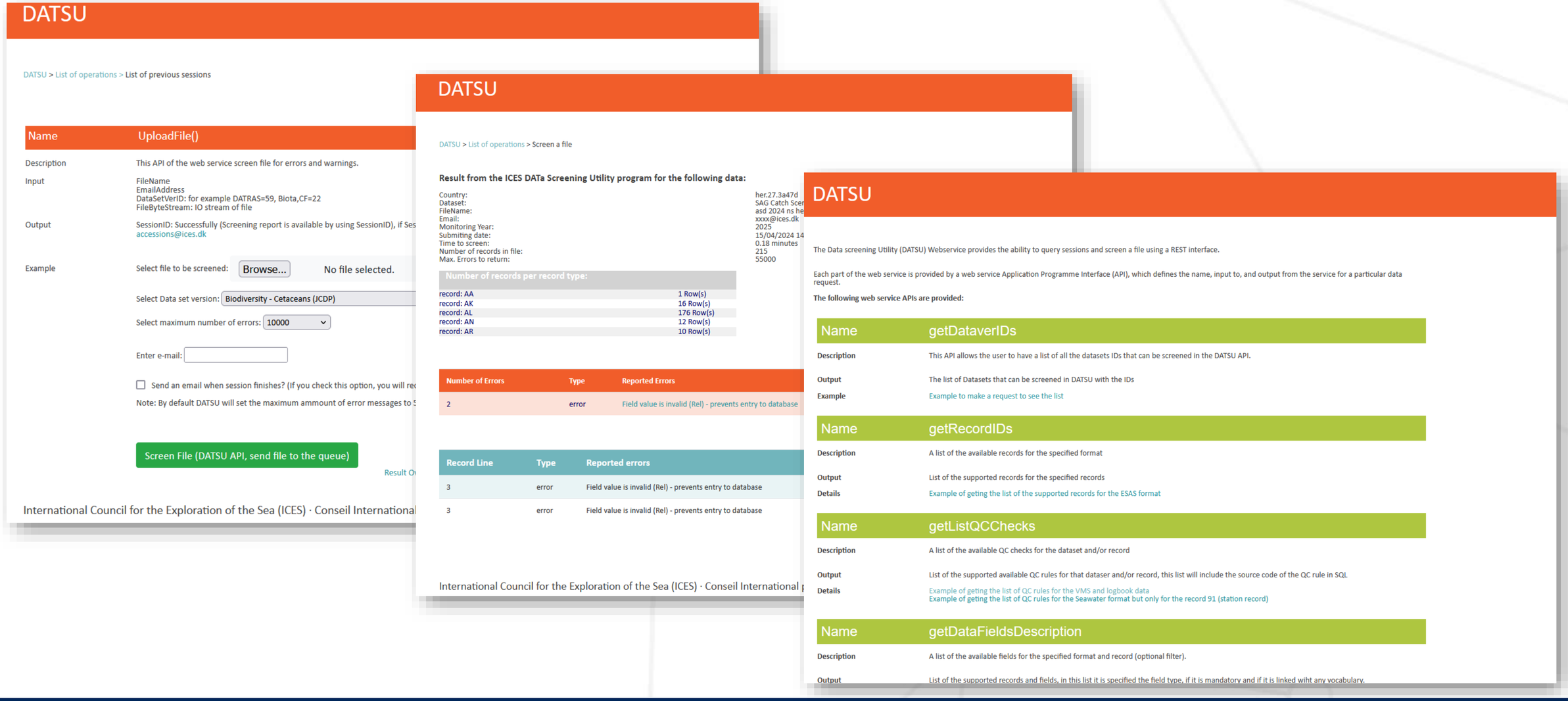
No errors

Request to upload the data to the ICES Database

Correct errors

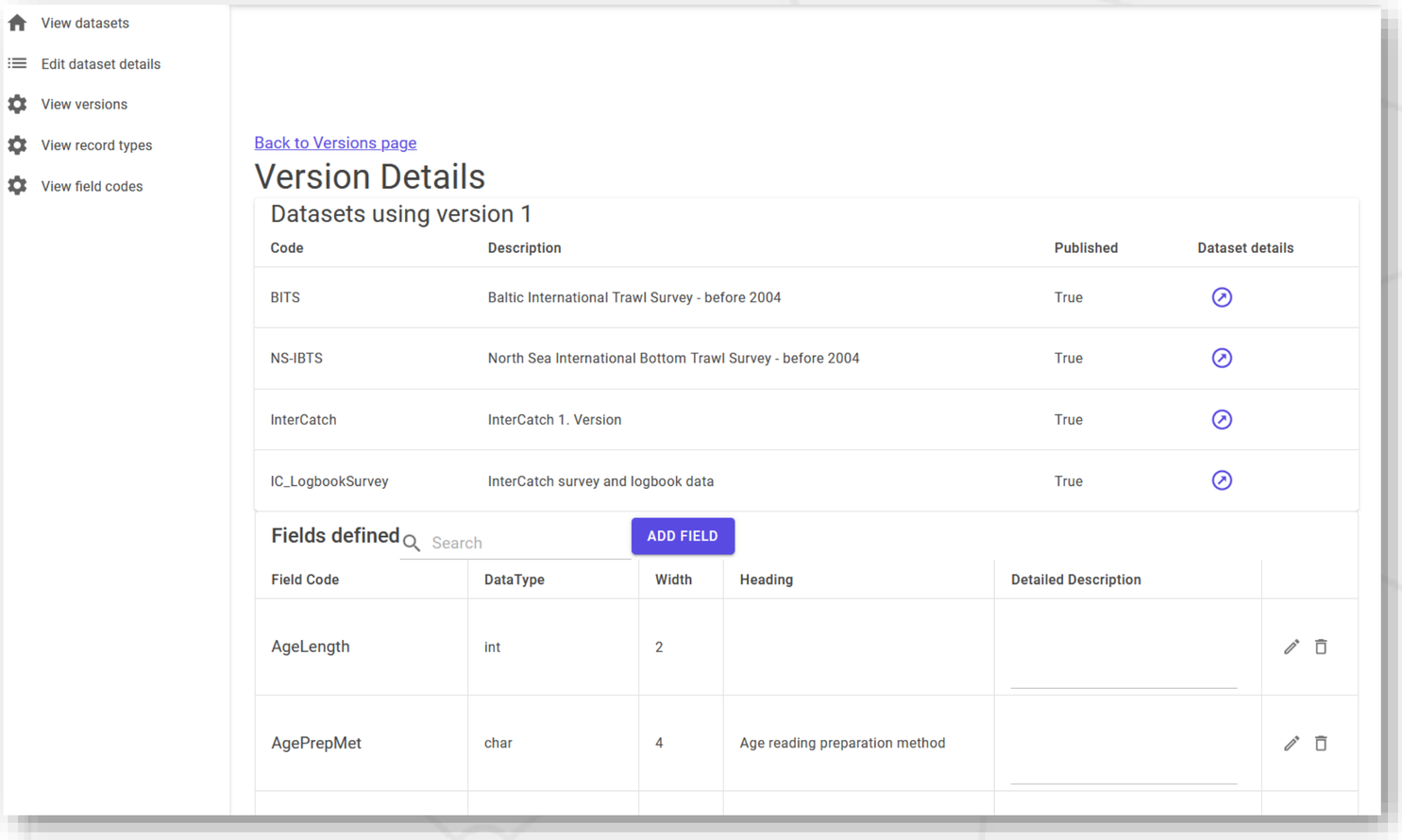
Science for sustainable seas

Data validation process cont.



Format management DATSU UI Web app

- Implemented using C# .NET and the Blazor framework.
- Focus on enhanced format management, improved user experience, and better validation for format changes.
- An automatic synchronization with the ICES Vocabulary Server was implemented using the Vocab Web API.
- Redundant features, such as species lists and reports, were omitted from the new version.



Benefits & limitations

- Exposes format information, data types, and vocabularies online and via web services.
- Allows for flexible data submission and screening methods.
- Generates comprehensive reports accessible through web pages and services.
- Shows the source code for quality checks.
- R package facilitates quick local data checks.

- Screening queue can be blocked by large files, causing delays.
- DATSU is heavily based on shared functionality and definitions for the format which can be both an advantage and a disadvantage.

Future development

Major change involving using message queues in DATSU

- Will avoid long waiting time and parallelize the process, by each dataset having its own queue and processor.
- In the future move to microservice architecture by separating the methods for running different types of checks.
- This will be implemented using RabbitMQ and MassTransit

Functionality extension

- Allow the fields to be reported in another order in the text files if the user reports the headings.
- Accept JSON file format in submissions.
- Add logging for each change made using the DATSU UI by data managers.

Thank you for your attention.
Any questions?

27-29 May 2024

International conference on Marine Data and Information Systems

SeaDataNet AISBL · Havforskningsinstituttet · Ifremer · MARIS · MEDIN · National Oceanography Centre · INGV · iMagine · ICDE · eosc Blue-Cloud2026