ESGF Errata Service for CMIP6

Errare humanum est ~ To err is human

Guillaume Levavasseur
Atef Ben Nasser
Mark Greenslade
Sébastien Denvil

Plan:
- Overview
- Progress & accomplishments
- Short demo
- Roadmap & developments

December 5th, 2018
The "Errata Service" before CMIP6

Houston?
We've had a problem...
The "Errata Service" before CMIP6

Houston? We've had a problem...

- Centralized issue management
- Standardized issue template
- Easy access to timely information
- Version tracking with related issues
The Errata Service for CMIP6 and more

Houston? We've a problem...

The Errata Service's success depends on only one condition...

... any kind of error or mistake leads to a version change of a dataset and, conversely, **any change of a dataset should be justified by an issue.**

Check [errata.es-doc.org](http://errata.es-doc.org)
The Errata Service in a nutshell

Errata Web-Service
- Authorization via GitHub Orgs & Teams
- Validation based on JSON Schemas
- Database Storage

Command-Line Interface
- Authentication via Credentials Setup
- Local Template
- Validation based on JSON Schemas

Errata Front-End

Handle Service
- RabbitMQ Consumer
- Issue Identifier Recording
- Persistent Identifier Storage

Stakeholders
- ESGF User Requesting
- Data Provider - Issue Evaluation
- Data(Node) Manager - Data Identification
- ESGF User Requesting

Overview | Progress & Accomplishments | Short demo | Roadmap & developments

https://errata.es-doc.org/
What's new?

Choose how to manage your issues!

Exercice during CMIP6 Data Challenges
Available since June 2018

Errata Web-Service
Authorization via GitHub ORGS & TEAMS
Validation based on JSON schemas
Database storage
Handling service
Persistent IDentifier storage
RabbitMQ consumer
Issue identifier recording

ESGF User Request
Data provider - issue evaluation
Data(Node) Manager - Data identification
ESGF User reporting

Command-Line Interface
Authentication via credentials setup
Local template
Validation based on JSON schemas

Errata Front-End
Issue lookup
Issue viewer
Version tracker
Authentication via GitHub login
Web forms
Browser based validation

Overview | Progress & Accomplishments | Short demo | Roadmap & developments
What's new?

ESGF Errata Service

As part of the ES-DOC ecosystem, the Errata Service centralizes timely information about known issues of ESGF data.

Proper handling of errata information in the ESGF publication workflow has major impact on the quality of data. Version changes should be documented and justified by explaining what was updated, retracted and/or removed. This leads to the unspoken rule that any publication of a newer version of a dataset needs to have a valid motivation, which we will refer to as an issue.

Consequently, the publication of a new version of a dataset, as well as the unpublishation of a dataset version, has to be motivated by an issue and conversely.

**Note**

The issue registration should be executed prior to the publication process and is ought to be mandatory for additional versions, version removal or retraction.
What's happened since our last F2F?

CMIP6 Data Challenges provided a lot of useful feedbacks!
Many thanks to all DC participants!

We've been in production since

But...

✓ a minor issue related to PID ingestion due to password change on IPSL RabbitMQ instance.
✓ some minor bugs related to front-tend display features.
What's happened since the ESGF opening to CMIP6?

Of the 21478 CMIP6 datasets (without replicas) currently published on ESGF, **5403 datasets affected by at least one issue or 25% of the current CMIP6 archive.**

20 CMIP6 issues currently declared on the ESGF Errata Service

Fortunately, most of the issues are related to metadata and/or coordinates.
"What else?"

**WE PLAN IT...**
- Improve error messages from the server: *What's wrong in my form/template, why I'm getting this error?*
- Add PID lookup to the command-line client: `> esgissue check`
- Communicate best practices over the world! *(Please, we need you again...)*

**WE THINK ABOUT IT...**
- Add support of "info" errata about no published datasets.
- A notification system to allow users to subscribe to an "issue flux".

**WE DREAM ABOUT IT...**
- Errata statistics endpoint
- More integration of the Errata Service into CoG:
  - How third-party services can add useful info into Solr metadata?
  - Green/red flag before download based on asynchronous requests to the Errata Service API?
Thank you for your attention!

› Developers:
  - Atef BEN NASSER (CNRS/IPSL - abennasser@ipsl.fr)
  - Mark GREENSLADE (SU/IPSL - momipsl@ipsl.fr)
  - Guillaume LEVAVASSEUR (SU/IPSL - glipsl@ipsl.fr)

› Contributions:
  - Sébastien DENVIL (CNRS/IPSL)
  - Merret BUURMAN (DKRZ)
  - Katharina BERGER (DKRZ)

› Acknowledgments:
  - Martina STOCKHAUSE (DKRZ)
  - Tobias WEIGEL (DKRZ)
  - Ag STEPHENS (CEDA)
  - Alan IWI (CEDA)

https://errata.es-doc.org and click on "Docs" button