



ESGF Errata Service for CMIP6/Cordex

Errare humanum est ~ To err is human

Atef Ben Nasser – IPSL/CNRS
Guillaume Levavasseur – IPSL/SU
Mark Greenslade – IPSL

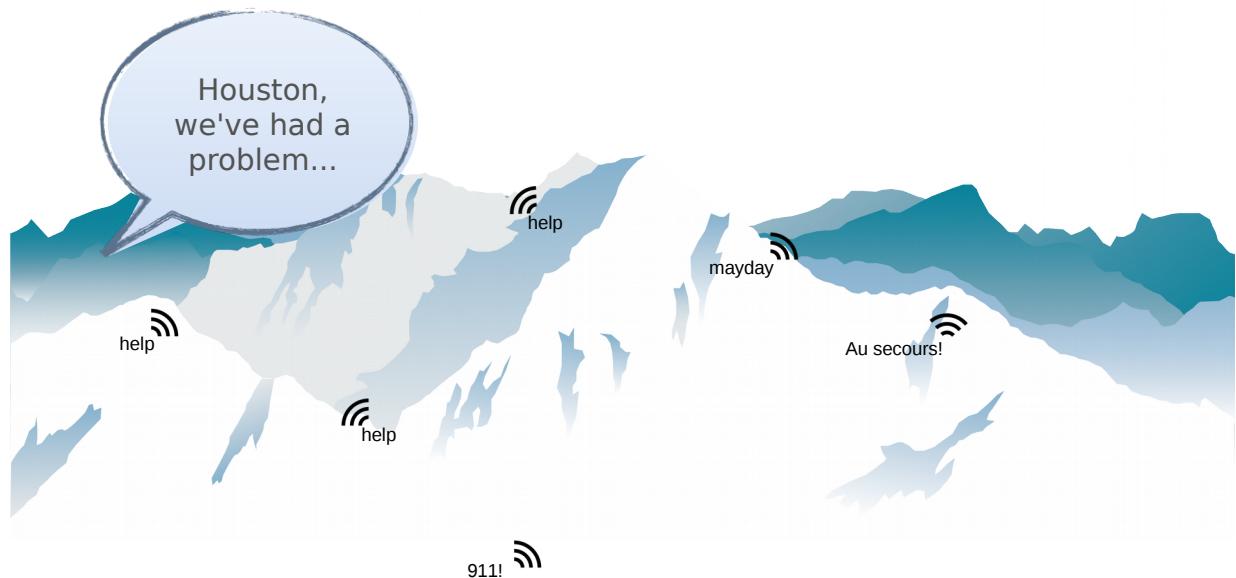


Motivation



Compromised FAIR

The collection of issues before CMIP6



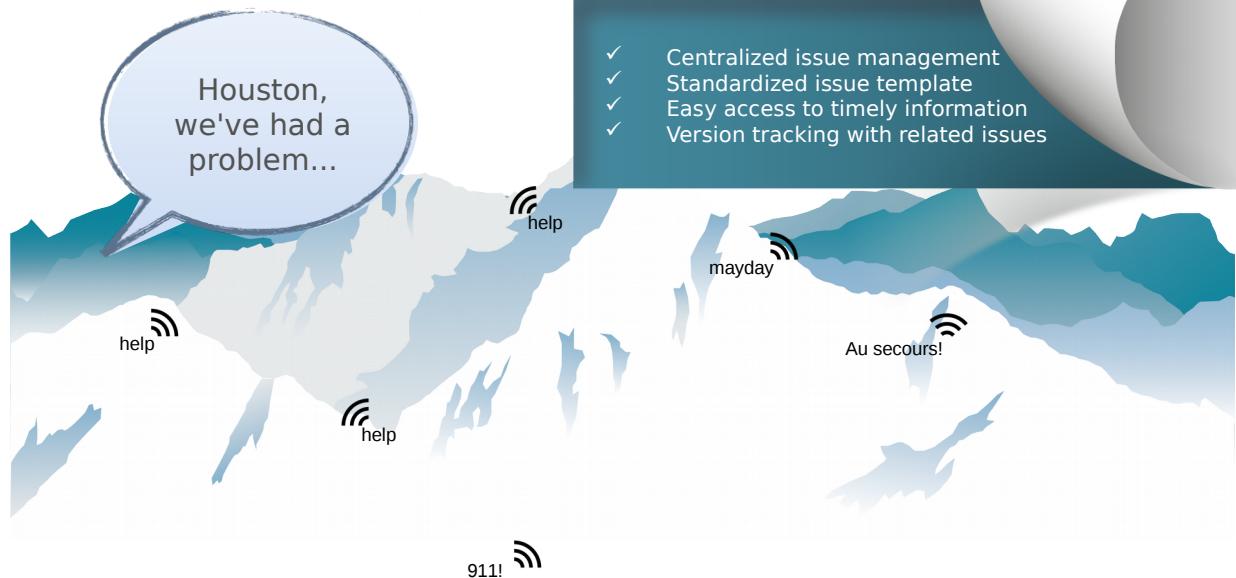


Motivation



Proposed fixes

The collection of issues before CMIP6





Motivation

Proposed fixes



The idea :

- A bug tracker à la GitHub, BitBucket or JIRA...

The requirements :

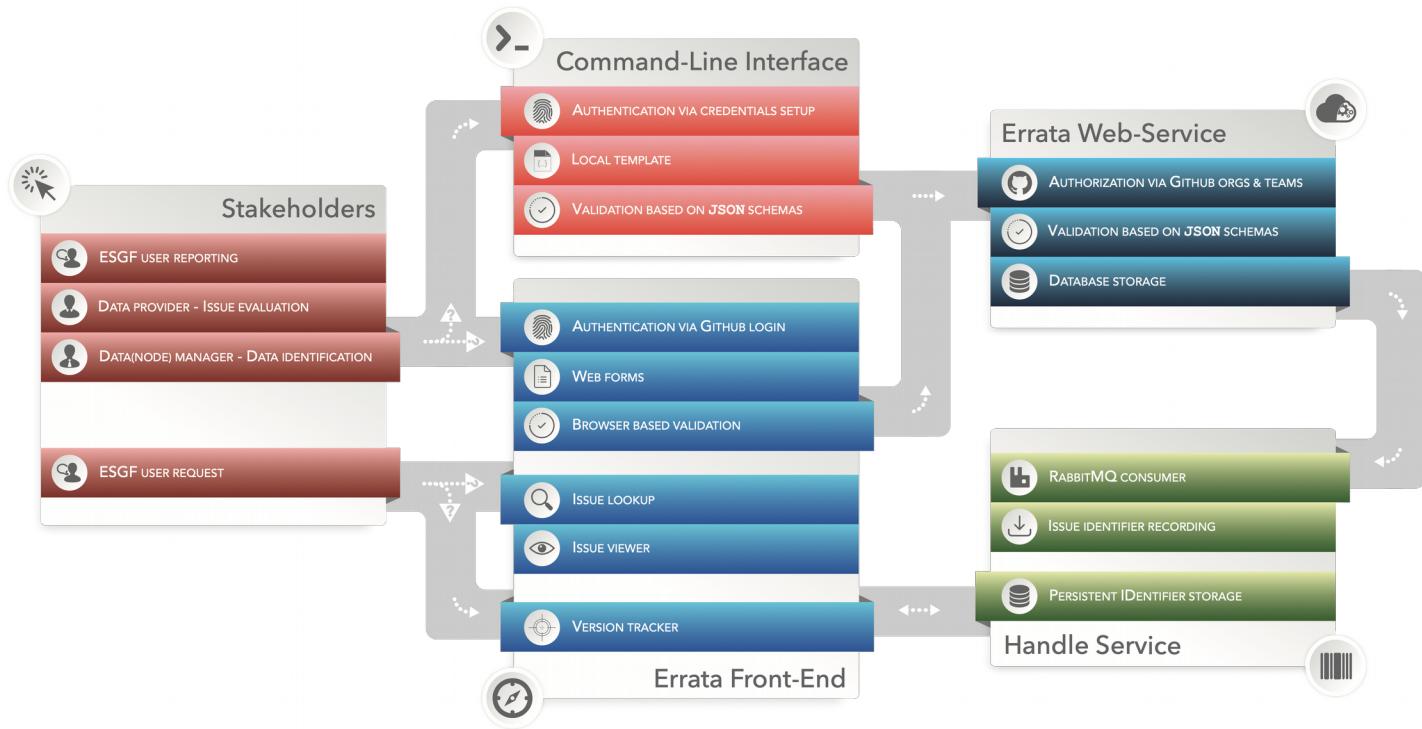
- Requires authentication
- Requires authorization
- Has a unique title
- Has an elaborate description
- Includes a non void list of affected datasets
- Is part of a supported project
- Includes an optional list of materials (figures...)
- Includes an optional list of external URLs.



The esdoc errata service



In a nutshell





The esdoc errata service



es-doc
Earth System Documentation

In practice : the old way (still supported, will always be)

- 1) Have a GitHub account
- 2) Generate a personal access token from your account settings
- 3) The ES-DOC errata client only requires read privileges to your orgs memberships (i.e. token scope = read:org)
- 4) Ask ES-DOC errata admins for proper team affiliation :
 - i. Errata publication team
 - ii. The institute on behalf of which you publish errata
- 5) Your GitHub username and personal access token will be required at least once during the first interaction between the errata client and server.

The diagram illustrates the process of generating a personal access token on GitHub. On the left, a screenshot of the GitHub user profile page shows the 'Settings' option highlighted. An arrow points from this 'Settings' link to the 'Personal access tokens' section on the right. The right-hand panel is titled 'Personal access tokens' and contains a list of generated tokens. A specific token, 'cdc8a93eb6cb9067305fa1c59b6ad4f3e5e1abeb', is selected, and its details are shown in a modal window. The modal includes fields for 'Token description' (set to 'esgissue') and 'What's this token for?' (with 'read:org' checked). At the bottom of the modal, there are checkboxes for 'admin:org' (unchecked), 'write:org' (unchecked), and 'read:org' (checked).

Personal access tokens

Generate new token | Revoke all

Tokens you have generated that can be used to access the GitHub API.

Make sure to copy your new personal access token now. You won't be able to see it again!

cdc8a93eb6cb9067305fa1c59b6ad4f3e5e1abeb

Edit | Delete

Token description
esgissue

What's this token for?

admin:org Full control of orgs and teams

write:org Read and write org and team membership

read:org Read org and team membership

Signed in as glevava

Your profile | Your stars | Explore | Integrations | Help | Settings | Sign out

Developer settings | OAuth applications | Integrations | Personal access tokens

INSTITUT PIERRE-SIMON LAPLACE



The esdoc errata service



In practice : the old way (still supported, will always be)

Step 2 - One issue = two files

- The issue manager works with both JSON and TXT files. This allows the data provider in charge of ESGF issues to manage one or several templates gathering the issues locally.

```
# Edit your issue following our JSON template (with only 4 mandatory attributes)
> vi myissue.json
```

```
{
  "title": "My Issue Title",
  "description": "This is a test description, void of meaning.",
  "project": "cmip5",
  "severity": "high",
  "materials": ["http://www.my-errata.org/images/issue.jpg"],
  "url": "http://www.myissue-landing-page.org/"
}
```

```
# Edit the list of affected dataset ids WITH version number
> vi mydsets.txt
```

```
cmip5.output1.IPSL.IPSL-CM5A-MR.historical.mon.land.Lmon.r1i1p1#20111111
cmip5.output1.IPSL.IPSL-CM5A-LR.historical.3hr.atmos.Lmon.r3i2p3#20121212
cmip5.output1.IPSL.IPSL-CM5A-LR.historical.6hr.atmos.Lmon.r2i5p4#20121212
[...]
```

- Each issue will be validated against the appropriate JSON schema to ensure template conformance, security and issue quality.



The esdoc errata service

In practice : the old way (still supported, will always be)

Step 3 - ES-DOC issue publishing

```
# Issue creation is the first step in the issue life cycle
> esgissue create --issue myissue.json --dsets mydsets.txt
Username: glevava
Token: 16b015c38b9a7129475bcb190f1fa3fcfa6c022a8
INFO Credentials were successfully saved.
INFO Validating issue...
INFO Validation Result: SUCCESSFUL
INFO Requesting issue #cd6ebba4-3eda-4e9a-9882-17a370c49419 creation from errata service...
INFO Updating fields of payload after remote issue creation...
INFO Issue json schema has been updated, persisting in file...
INFO Issue file has been created successfully!
```

```
# On success the issue UID, the status, the creation and update dates are appended to the local
issue file
```

```
{
  "uid": "cd6ebba4-3eda-4e9a-9882-17a370c49419",
  "title": "My Issue Title",
  "description": "This is a test description, void of meaning.",
  "project": "cmip5",
  "dateCreated": "2016-12-06 15:23:02",
  "dateUpdated": "2016-12-06 15:23:02",
  "materials": ["http://www.my-errata.org/images/issue.jpg"],
  "severity": "high",
  "status": "new",
  "url": "http://www.myissue-landing-page.org/"
}
```



The esdoc errata service



es-doc
Earth System Documentation

In practice : the regular way

- 1) Go to <https://errata.es-doc.org>
- 2) Press the login button
- 3) Sign into your GitHub account
- 4) Press the create button
- 5) Fill in the form with respect to the constraints
- 6) Enjoy your newly created issue



The esdoc errata service

In practice : the old way (still supported, will always be)

Dataset Errata - Search v0.8.0.0 [Support](#) [Docs](#) [Search](#) [Login](#)

Experiment ID:	Institution ID:	Source ID:	Variable ID:	Severity:	Status:
...

4. Filtered Issues = 313. << < > >> 25 / page

Institute	Title	Created	Updated	Closed	Severity	Status
CONSORTIUM	Publication of different versions of the same variab ...	2021-03-05	2021-03-05	--	Low	Resolved
THU	Scaling errors of Amon.clt in THU/CIESM	2021-03-04	--	--	Medium	Low
CMI	Extend CMCC-ESM2 piControl experiment	2021-03-04	2021-03-09	--	Low	Resolved

Sign in to GitHub
to continue to esdoc-errata-publisher

User email address

Password [Forgot password?](#)

[Sign in](#)

New to GitHub? [Create an account.](#)

Dataset Errata - Search v0.8.0.0 [Support](#) [Docs](#) [Search](#) [Create](#)

Project:	Experiment ID:	Institution ID:	Source ID:	Variable ID:	Severity:	Status:
CMP6

Total issues = 314. Filtered Issues = 313. << < > >> 25 / page

#	Institute	Title	Created	Updated	Closed	Severity	Status
1	HAMMOZ-CONSORTIUM	Publication of different versions of the same variab ...	2021-03-05	2021-03-05	--	Low	Resolved
2	THU	Scaling errors of Amon.clt in THU/CIESM	2021-03-04	--	--	Medium	Low
3	CMI	Extend CMCC-ESM2 piControl experiment	2021-03-04	2021-03-09	--	Low	Resolved

es-doc
Earth System Documentation

Dataset Errata - Create Issue v0.8.0.0 [Support](#) [Docs](#) [Save](#)

New Issue

Project *

Title *

Description *

Severity *

Links

Materials / Images

Datasets *



The esdoc errata service

In practice : either way you choose to operate

es-doc

Earth System Documentation

Errata Search v0.2.8.0

[Support](#)

Project:	Severity:	Status:	Institute:	Model:	Experiment:	Variable:
CMIP5	*	*	*	*	*	*
Total Issues = 202. Filtered Issues = 57.						
#	Institute	Title	Created	Updated	Closed	Severity
1	IPSL	My Issue Title	2016-12-06	2016-12-06	--	High
2	BADC	Test issue title - 120b27f7-029a-45ea-85e8-15a8efcd60 ...	2016-11-05	--	--	Medium
3	DKRZ	Test issue title - 19c7fa72-fbc5-4eb4-b41b-1c28b23eb4 ...	2016-11-05	--	2016-11-09	High
4	BADC	Test issue title - 2da439d5-5cd3-44bc-950e-ee594627c1 ...	2016-11-05	--	--	Medium
5	DKRZ	Test issue title - 16717c3a-3a8e-41cd-b46c-a5b0a913c5 ...	2016-11-04	--	2016-11-08	Critical
6	BADC	Test issue title - 175df12d-eecf-49c8-af97-d9b811d922 ...	2016-11-03	--	--	Medium
7	DKRZ	Test issue title - 6f9a158a-15ea-4f12-874c-9c47b55641 ...	2016-11-02	--	--	Low
8	DKRZ	Test issue title - 3d45ccb8-410e-4a67-bac0-2d9aa17b62 ...	2016-11-02	--	2016-11-06	High
9	DKRZ	Test issue title - a6ad64dc-693e-47aa-bb2b-97519de9d1 ...	2016-11-02	--	2016-11-06	Medium
10	BADC	Test issue title - b392d71d-4021-45fe-ba9f-eedde3240d ...	2016-11-01	--	2016-11-05	Critical
11	IPSL	Test issue title - 75e035bb-7b20-4ab4-a72a-eb68c4d612 ...	2016-11-01	--	--	Low
12	IPSL	Test issue title - 0879a5c2-edfc-46d3-9fd4-4c0583ac46 ...	2016-10-31	--	--	Critical
13	DKRZ	Test issue title - 2956beca-bcd4-40fd-a143-86fed49f28 ...	2016-10-30	--	2016-11-03	Low
25 / page						



The esdoc errata service

In practice : either way you choose to operate

es-doc

Earth System Documentation

Errata Viewer v0.2.8.0

Support

CMIP5 - IPSL - My Issue Title

DETAILS	MATERIALS (1)	AFFECTED DATASETS (3)
Project		CMIP5
Institute		IPSL
Title		My Issue Title
Description		This is a test description, void of meaning.
Affected Experiments		historical
Affected Models		IPSL-CM5A-MR
Affected Variables		lmon
Landing Page		http://errata.ipsl.upmc.fr/issue/2
Severity	High	
Status	New	
Created By		Guillaume @ 2016-12-06 15:23:02
Last Updated By		Guillaume @ 2016-12-06 15:23:02
Closed By		--
ESG-F Errata Identifier		cd6ebba4-3eda-4e9a-9882-17a370c49419



The esdoc errata service

In practice : either way you choose to operate

CMIP5 - IPSL - My Issue Title

[DETAILS](#)

[MATERIALS \(1\)](#)

[AFFECTED DATASETS \(3\)](#)

ESG-F Dataset Identifier

- 1 cmip5.output1.IPSL.IPSL-CM5A-MR.historical.mon.land.Lmon.r1i1p1#20111119
- 2 cmip5.output1.IPSL.IPSL-CM5A-LR.historical.3hr.atmos.Lmon.r3i2p3#20121212
- 3 cmip5.output1.IPSL.IPSL-CM5A-LR.historical.6hr.atmos.Lmon.r2i5p4#20121212

Affected Datasets = 3

Errata v0.2.8.0 © 2016 ES-DOC



The esdoc errata service

In numbers



es-doc
Earth System Documentation

- Of the 4,592,953 distinct CMIP6 datasets (no replicas) currently published on ESGF 169416 datasets are affected by at least one issue which makes 3.7 % of the current archive
- 314 CMIP6 issues are registered in the ES-DOC Errata Service.

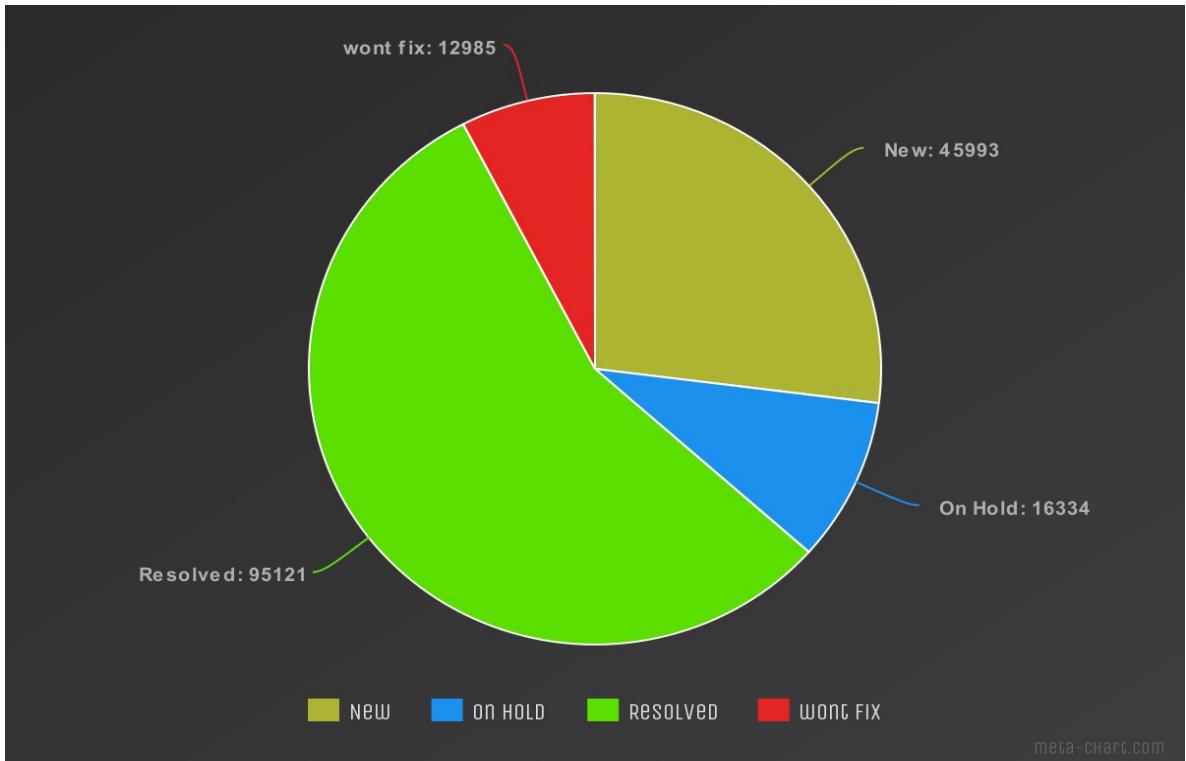


The esdoc errata service



es-doc
Earth System Documentation

In numbers



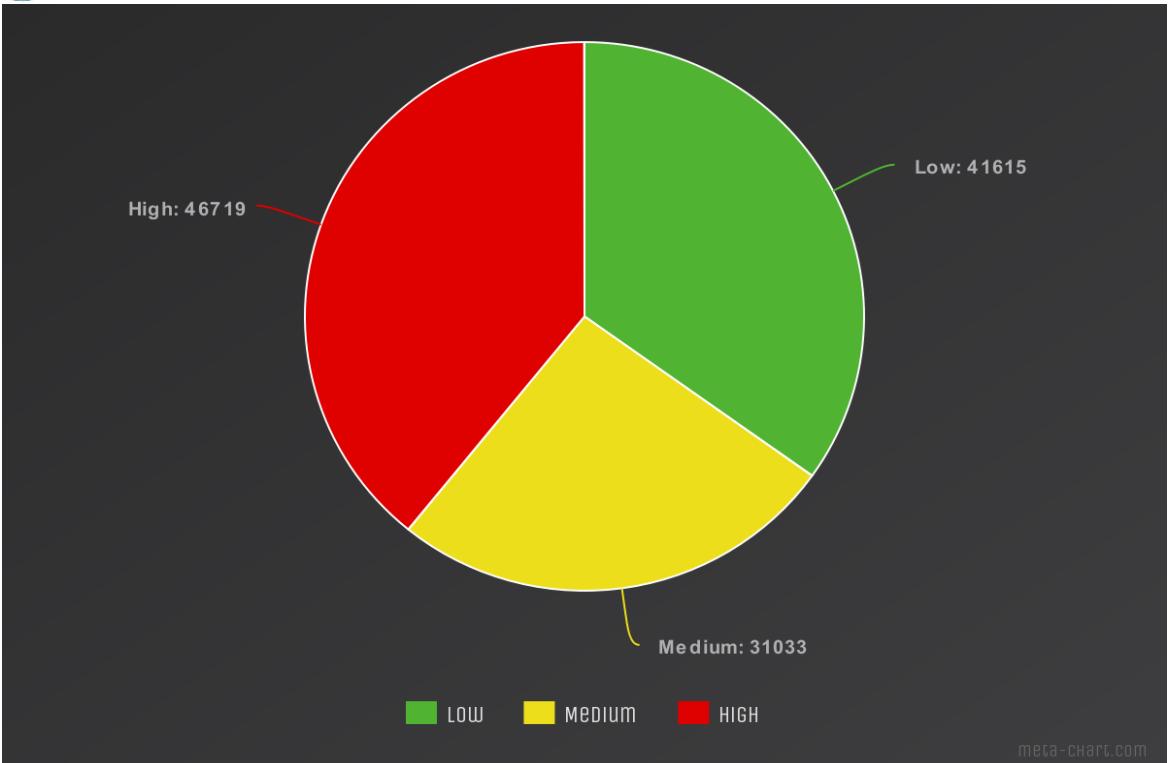


The esdoc errata service



es-doc
Earth System Documentation

In numbers





The esdoc errata service

Persistent Identifiers (PID) integration

- Created with every published CMIP6 dataset/file
- Hosts a variety of metadata
- Includes a predecessor/successor version chaining
- Example :

<https://handle-esgf.dkrz.de/1p/hdl:21.14100/e0f0906c-1546-3319-b981-411e89747dc0>

```
{  
    "u""VERSION_NUMBER": "u""20190311",  
    "u""REPLACED_BY": "u""hdl:21.14100/cce8a435-1a43-3fd4-9081-1743fb41624e",  
    "u""UNPUBLISHED_ORIGINALS": "u""<locations><location publishedOn=\\"2020-07-23T12:24:29.496+00:00\\"  
host=\\"vesg.ipsl.upmc.fr\\" /></locations>",  
    "u""HAS_PARTS": "u""hdl:21.14100/2aa2ea36-950e-4e00-a05e-c2a8a4d6e835",  
    "u""URL": "u""https://handle-esgf.dkrz.de/1p/21.14100/e0f0906c-1546-3319-b981-411e89747dc0",  
    "u""_REMOVED_HOSTS": "u""vesg.ipsl.upmc.fr##2020-07-23T12:24:29.496+00:00",  
    "u""DRS_ID": "u""CMIP6.GeoMIP.IPSL.IPSL-CM6A-LR.G6solar.r1i1p1f1.3hr.clr.gr",  
    "u""FIXED_CONTENT": "u""TRUE",  
    "u""REPLICA_NODE": "u""<locations><location publishedOn=\\"2019-05-27T05:31:30.335+00:00\\"  
host=\\"aims3.llnl.gov\\" /></locations>",  
    "u""HS_ADMININ": "u""{u'index': 200, u'handle': u'21.14100/ADMINLIST', u'permissions':  
u'111111111111'}",  
    "u""AGGREGATION_LEVEL": "u""DATASET",  
    "u""ERRATA_IDS": "u""1852e557-3ba4-8333-b222-6c42b6a944ce",  
    "u""_ADDED_ERRATA_IDS": "u""1852e557-3ba4-8333-b222-6c42b6a944ce##2020-12-  
18T16:34:18.706+00:00;1852e557-3ba4-8333-b222-6c42b6a944ce##2020-12-18T16:34:18.706+00:00;1852e557-  
3ba4-8333-b222-6c42b6a944ce##2020-12-18T16:34:18.706+00:00"  
}
```



The esdoc errata service

Persistent Identifiers (PID) integration



es-doc
Earth System Documentation



es-doc

Earth System Documentation

Dataset Errata - PID v0.8.0.0

Support

Docs

Get version history of a (list of) file(s) or dataset(s):

CMIP6.GeoMIP.IPSL.IPSL-CM6A-LR.G6solar.r1i1p1f1.3hr.clt.gr#20190311

OR

Choose File

Search

21.14100/e0f0906c-1546-3319-b981-411e89747dc0

#	Dataset / File	Issue
1	CMIP6.GeoMIP.IPSL.IPSL-CM6A-LR.G6solar.r1i1p1f1.3hr.clt.gr#20190311 Queried Version First	
2	CMIP6.GeoMIP.IPSL.IPSL-CM6A-LR.G6solar.r1i1p1f1.3hr.clt.gr#20200723 Latest	



Perspectives

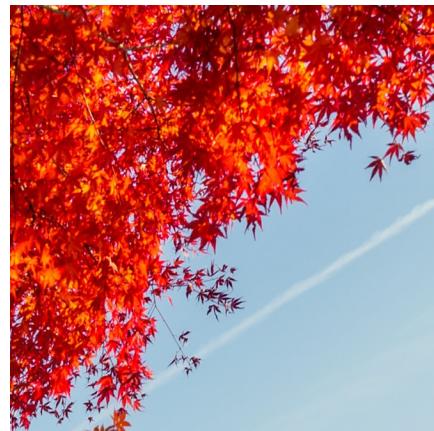
Some things in the pipeline, some things are just ideas for a distant future

More open contributions, not limited to data providers but still monitored by them to ensure quality

More integration with ESGF stack for instant visual feedback on the errata state of data

A notification service for post-usage case

Much more interactions with the community (that's you)





MERCI - THANK YOU

Institut Pierre-Simon Laplace (IPSL)