ESGF Errata Service for CMIP6

Guillaume Levavasseur[1]
Atef Bennasser[1]
Mark Greenslade[1]
Sébastien Denvil[1]

December 6th, 2016

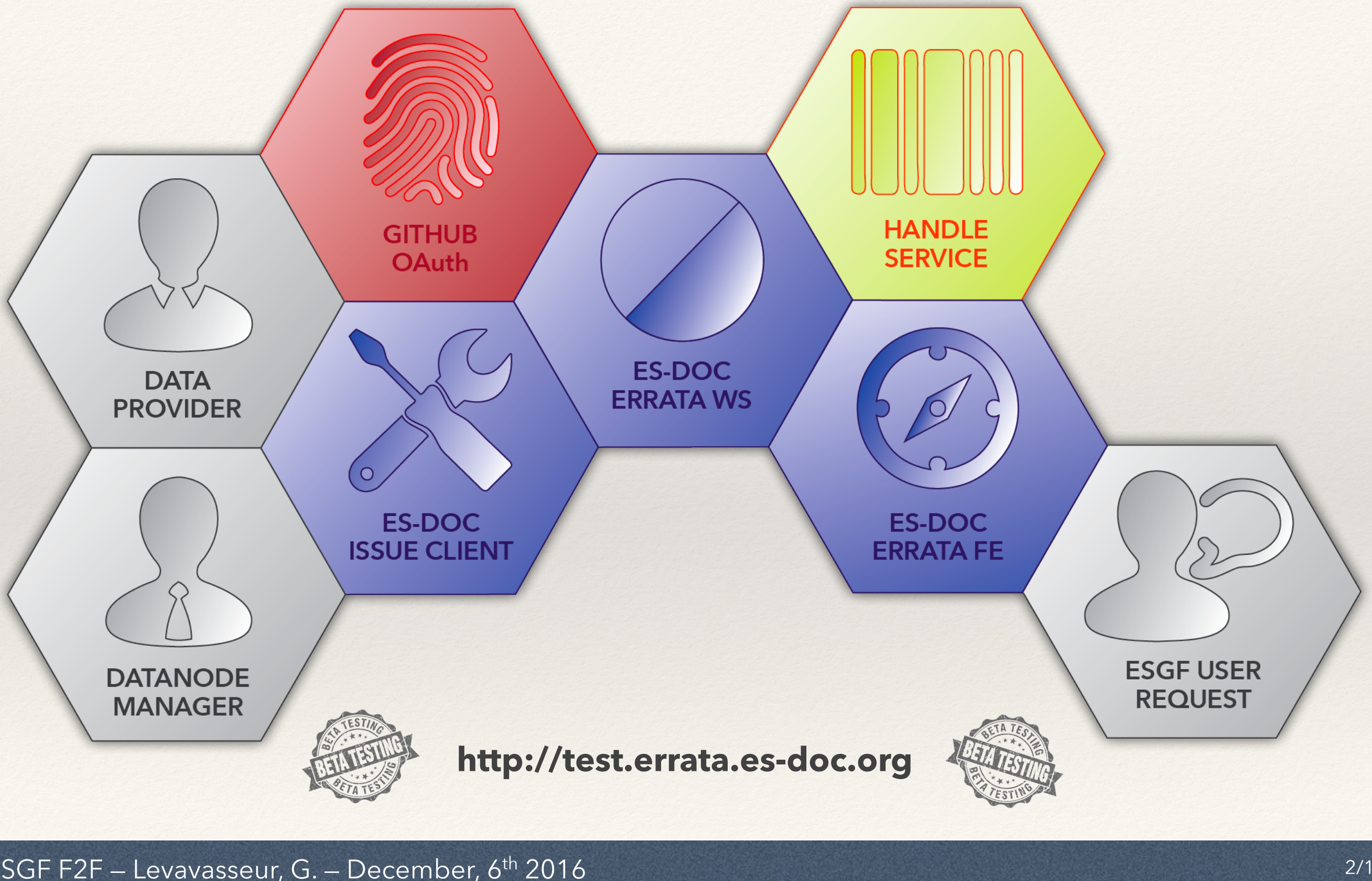
Plan:

- Overview and contributors
- Progress and accomplishments
- Incoming roadmap and developments
- Short demo

(1) The Institute Pierre Simon Laplace (IPSL) is a research federation including 9 laboratories, 1,400 researchers especially focused on climate and environment.

The Errata Service in a nutshell

branded by es-doc Earth System Documentation *and in collaboration with* DKRZ

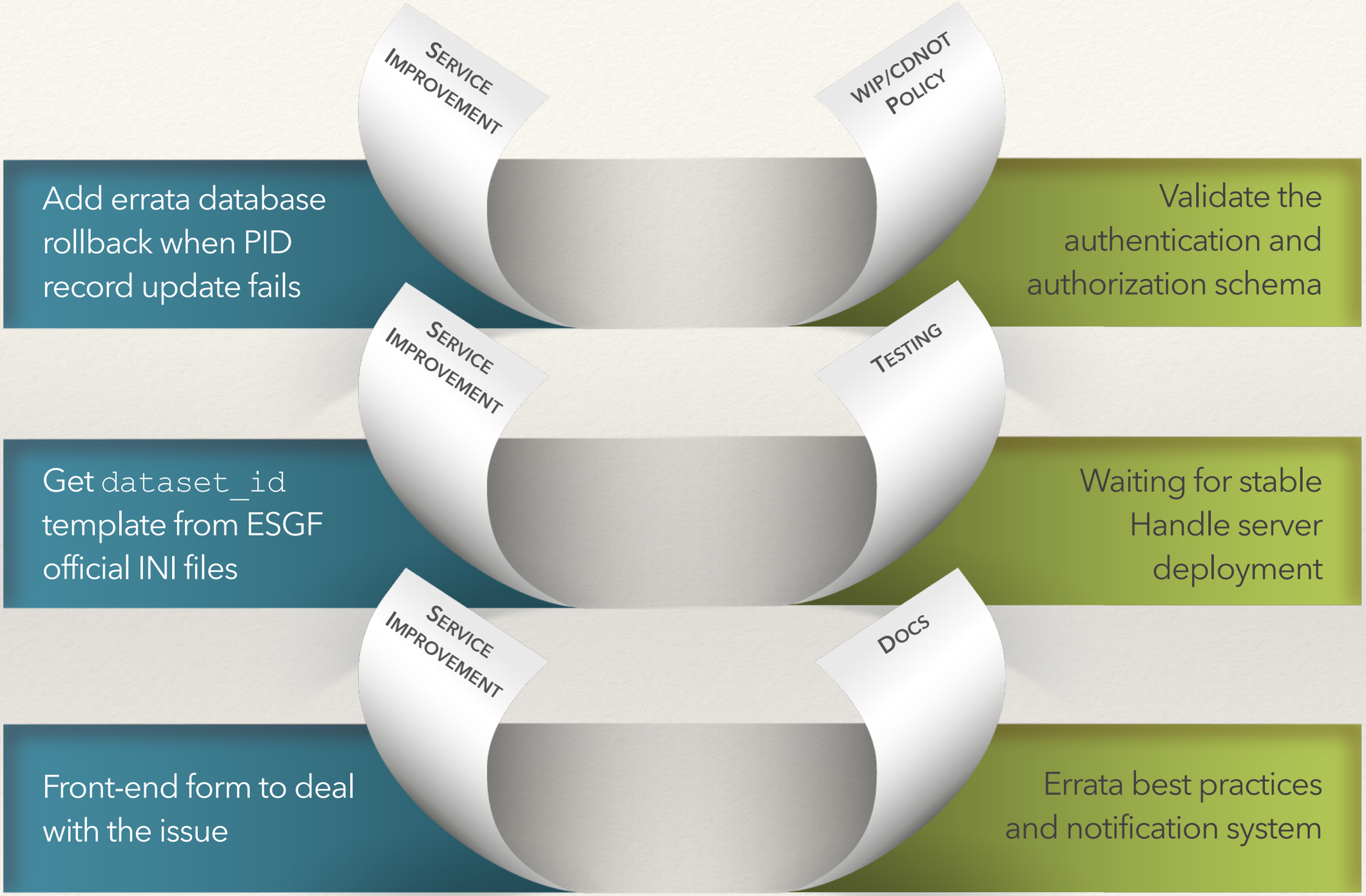


http://test.errata.es-doc.org

From a proof-of-concept to an ESGF service

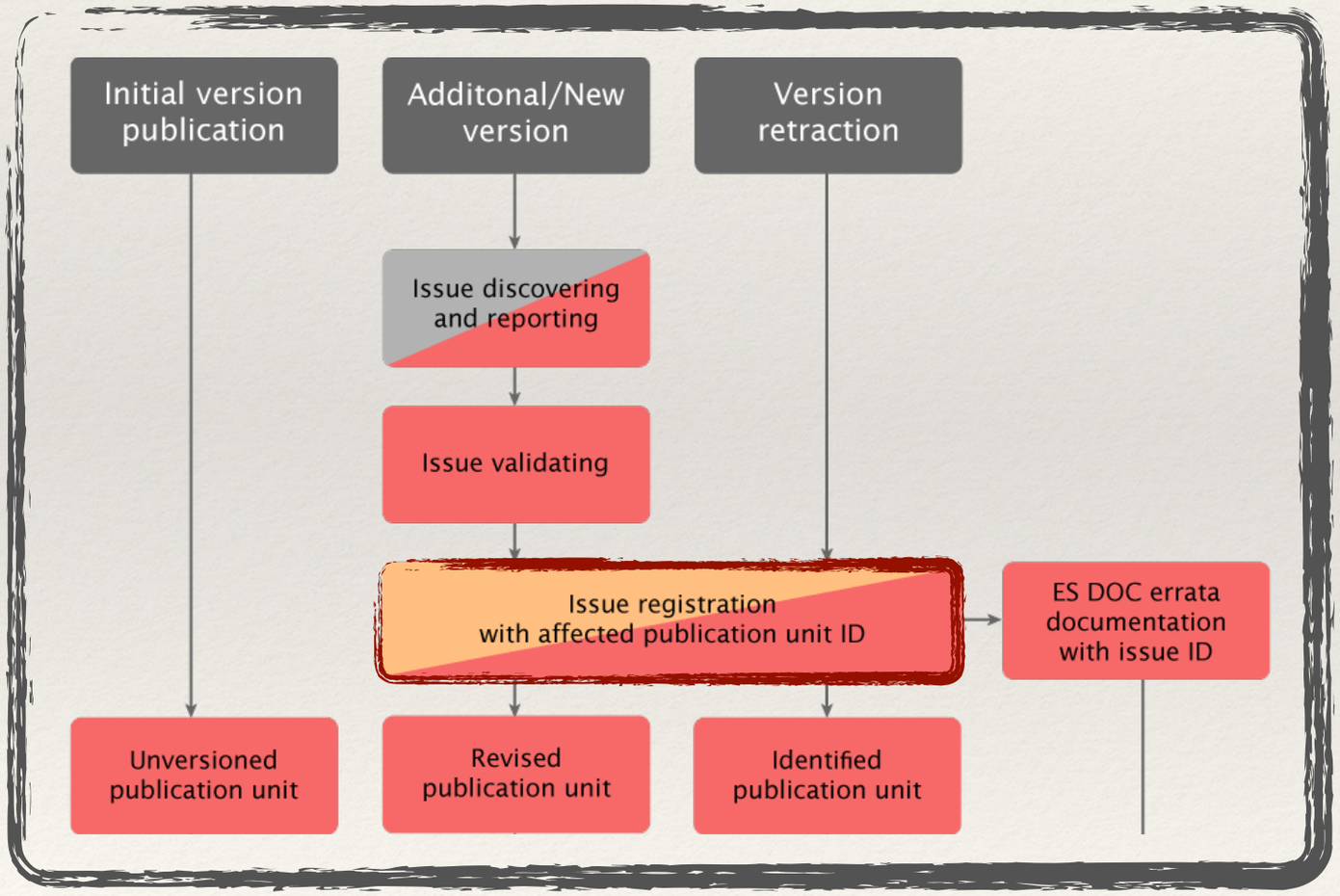
CONSTRAINTS	α-PHASE DELIVERABLES
Issue registration is manually triggered and cannot be made mandatory.	✓ Ability to create and close an issue through an efficient and simple workflow ✓ Centralized issue management ✓ Lightweight client available through PyPI ✓ Updated documentation
The risk of user mistakes must be anticipated.	✓ Ability to update and retrieve an issue ✓ Structured issue template ensures service quality and security ✓ Unchanged or value-controlled attributes
An issue provides a sensitive message about climate model quality that must not be misused.	✓ Issue validation is mirrored on client and server sides ✓ Authentication and authorization are delegated to GitHub OAuth ✓ Publisher permissions set at the modeling groups/institute level
Users can query about corrections applied to a dataset/file or some facets	✓ Exploit PIDs to support the link between dataset version and issue ✓ An algorithm deduces affected files from issues at the dataset level ✓ Responsive issue tracking system deployed and branded by ES-DOC to provide timely information about known issues

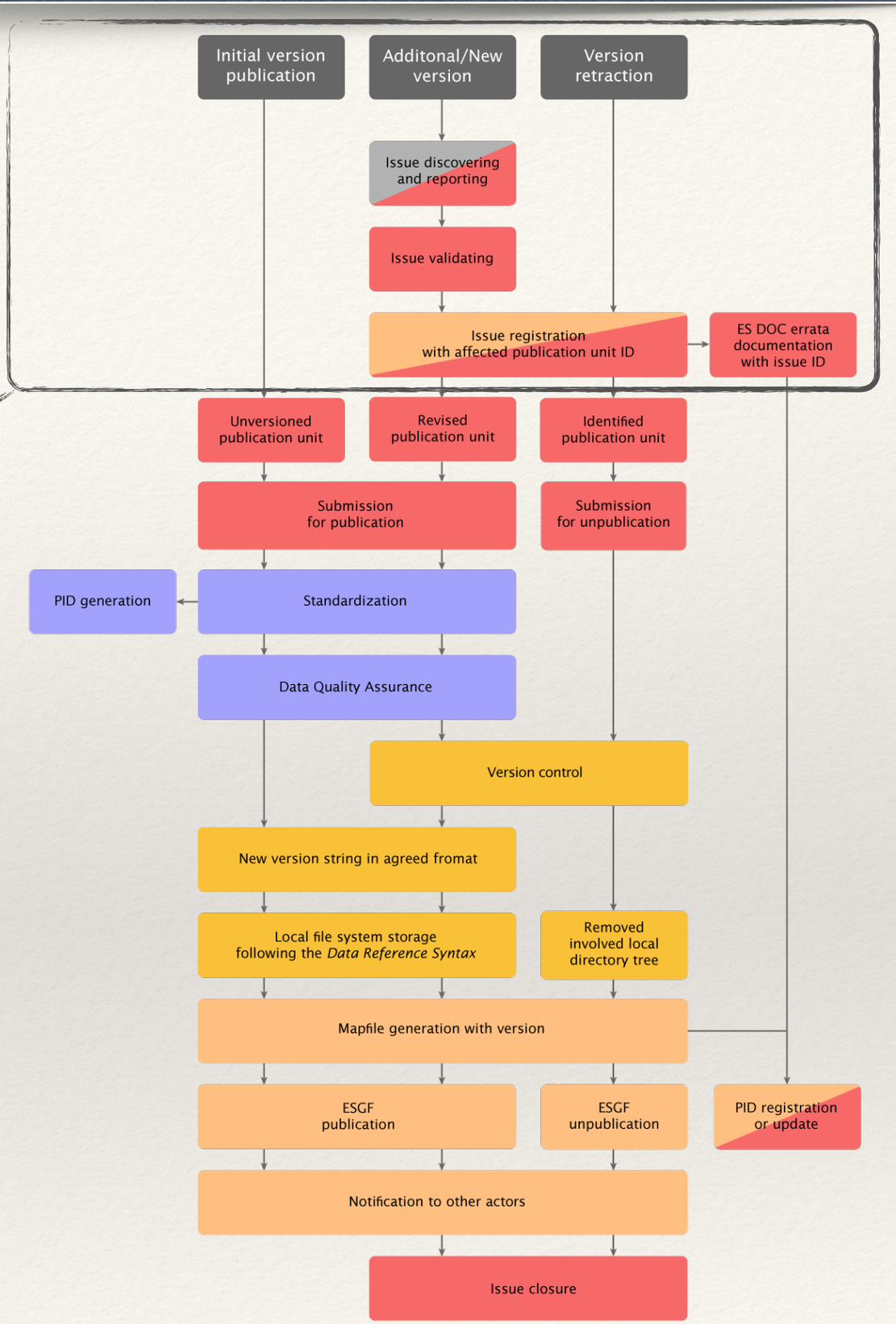
Progress towards beta release



Step 1 - When to register my issue?

Any kind of error or mistake leads to a version change of a dataset and, conversely, any change of a dataset (values, metadata, etc.) should be justified by an issue.





Step 2 - How to authenticate with my GitHub account?

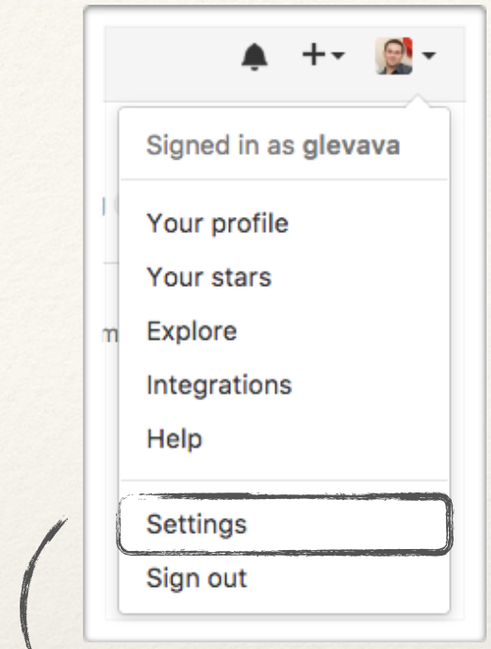
a. Create a GitHub account

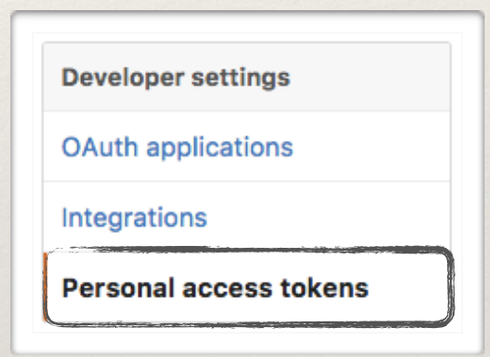
b. Generate a personal access token from your account settings
(see: *https://help.github.com/articles/creating-an-access-token-for-command-line-use/*)

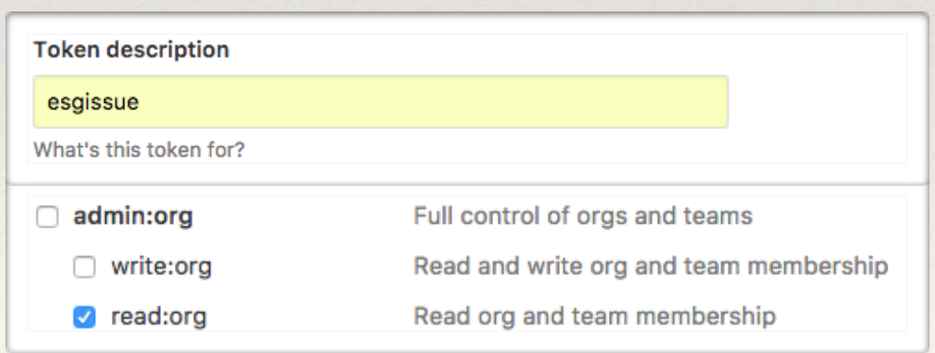
c. The ES-DOC errata client only need to read your organization membership (i.e., token scope = **`read:org`**).

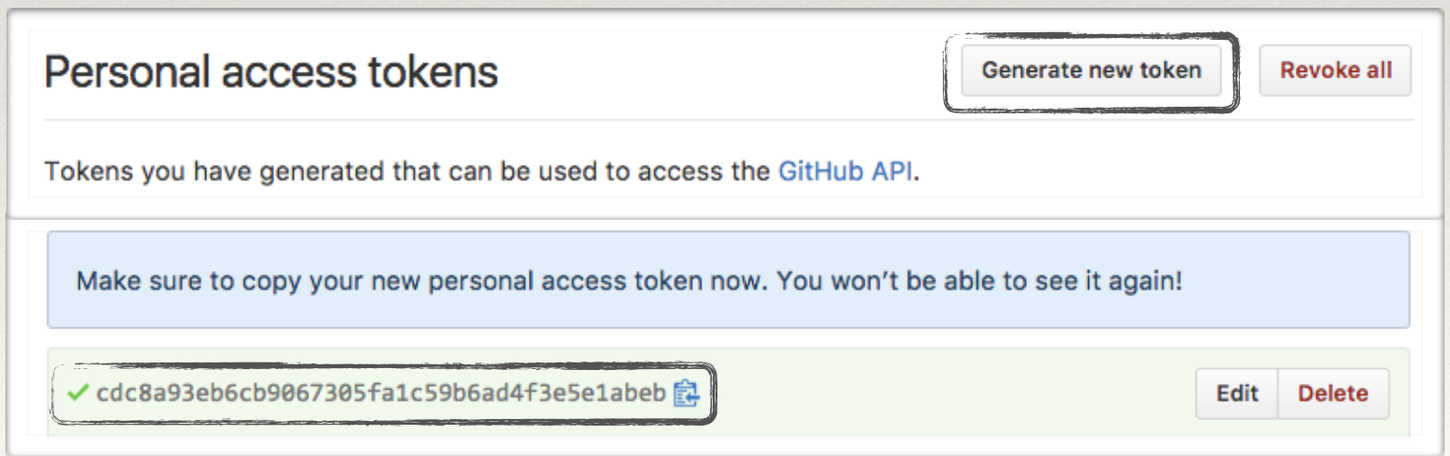
d. Ask authorization to ES-DOC errata administrators to be part of:
- The **`es-doc`** GitHub organization (may change depending on the WIP),
- Your **`errata-publication-<institute>`** team.

e. Your GitHub username and personal access token will be required (at least) for the first use of the ES-DOC errata client.









Step 3 - 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#20111119
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.

Step 4 - 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: 16b015c38b9a7129475bcb190f1fa3fca6c022a8
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/"
}
```

Step 4 - ES-DOC issue publishing

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. << < Page 1 of 3 > >> 25 / page

#	Institute	Title	Created	Updated	Closed	Severity	Status
1	IPSL	My Issue Title	2016-12-06	2016-12-06	--	High	New
2	BADC	Test issue title - 120b27f7-029a-45ea-85e8-15a8efcd60 ...	2016-11-05	--	--	Medium	Wont Fix
3	DKRZ	Test issue title - 19c7fa72-fbc5-4eb4-b41b-1c28b23eb4 ...	2016-11-05	--	2016-11-09	High	Wont Fix
4	BADC	Test issue title - 2da439d5-5cd3-448c-950e-ee594627c1 ...	2016-11-05	--	--	Medium	Resolved
5	DKRZ	Test issue title - 16717c3a-3a8e-41cd-b46c-a5b0a913c5 ...	2016-11-04	--	2016-11-08	Critical	On Hold
6	BADC	Test issue title - 175df12d-eecf-49c8-af97-d9b811d922 ...	2016-11-03	--	--	Medium	New
7	DKRZ	Test issue title - 6f9a158a-15ea-4f12-874c-9c47b55641 ...	2016-11-02	--	--	Low	New
8	DKRZ	Test issue title - 3d45ccb8-410e-4a67-bac0-2d9aa17b62 ...	2016-11-02	--	2016-11-06	High	New
9	DKRZ	Test issue title - a6ad64dc-693e-47aa-bb2b-97519de9d1 ...	2016-11-02	--	2016-11-06	Medium	Wont Fix
10	BADC	Test issue title - b392d71d-4021-45fe-ba9f-eedde3240d ...	2016-11-01	--	2016-11-05	Critical	Resolved
11	IPSL	Test issue title - 75e035bb-7b20-4ab4-a72a-eb68c4d612 ...	2016-11-01	--	--	Low	Wont Fix
12	IPSL	Test issue title - 0879a5c2-edfc-46d3-9fd4-4c0583ac46 ...	2016-10-31	--	--	Critical	New
13	DKRZ	Test issue title - 2956beca-bcd4-40fd-a143-86fed49f28 ...	2016-10-30	--	2016-11-03	Low	Resolved

Step 4 - ES-DOC issue publishing

es-doc
Earth System Documentation

Errata Search v0.2.8.0 Support

Project: CMIP5 | Severity: * | Status: * | Institute: * | Model: * | Experiment: * | Variable: *

Status: * / New / On Hold / Resolved / Wont Fix

Total Issues = 202. Filtered Issues = 57. << < Page 1 of 3 > >> 25 / page

#	Institute	Title	Created	Updated	Closed	Severity	Status
1	IPSL	My Issue Title	2016-12-06	2016-12-06	--	High	New
2	BADC	Test issue title - 120b27f7-029a-45ea-85e8-15a8efcd60 ...	2016-11-05	--	--	Medium	Wont Fix
3	DKRZ	Test issue title - 19c7fa72-fbc5-4eb4-b41b-1c28b23eb4 ...	2016-11-05	--	2016-11-09	High	Wont Fix
4	BADC	Test issue title - 2da439d5-5cd3-448c-950e-ee594627c1 ...	2016-11-05	--	--	Medium	Resolved
5	DKRZ	Test issue title - 16717c3a-3a8e-41cd-b46c-a5b0a913c5 ...	2016-11-04	--	2016-11-08	Critical	On Hold
6	BADC	Test issue title - 175df12d-eecf-49c8-af97-d9b811d922 ...	2016-11-03	--	--	Medium	New
7	DKRZ	Test issue title - 6f9a158a-15ea-4f12-874c-9c47b55641 ...	2016-11-02	--	--	Low	New
8	DKRZ	Test issue title - 3d45ccb8-410e-4a67-bac0-2d9aa17b62 ...	2016-11-02	--	2016-11-06	High	New
9	DKRZ	Test issue title - a6ad64dc-693e-47aa-bb2b-97519de9d1 ...	2016-11-02	--	2016-11-06	Medium	Wont Fix
10	BADC	Test issue title - b392d71d-4021-45fe-ba9f-eedde3240d ...	2016-11-01	--	2016-11-05	Critical	Resolved
11	IPSL	Test issue title - 75e035bb-7b20-4ab4-a72a-eb68c4d612 ...	2016-11-01	--	--	Low	Wont Fix
12	IPSL	Test issue title - 0879a5c2-edfc-46d3-9fd4-4c0583ac46 ...	2016-10-31	--	--	Critical	New
13	DKRZ	Test issue title - 2956beca-bcd4-40fd-a143-86fed49f28 ...	2016-10-30	--	2016-11-03	Low	Resolved

Step 4 - ES-DOC issue publishing

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

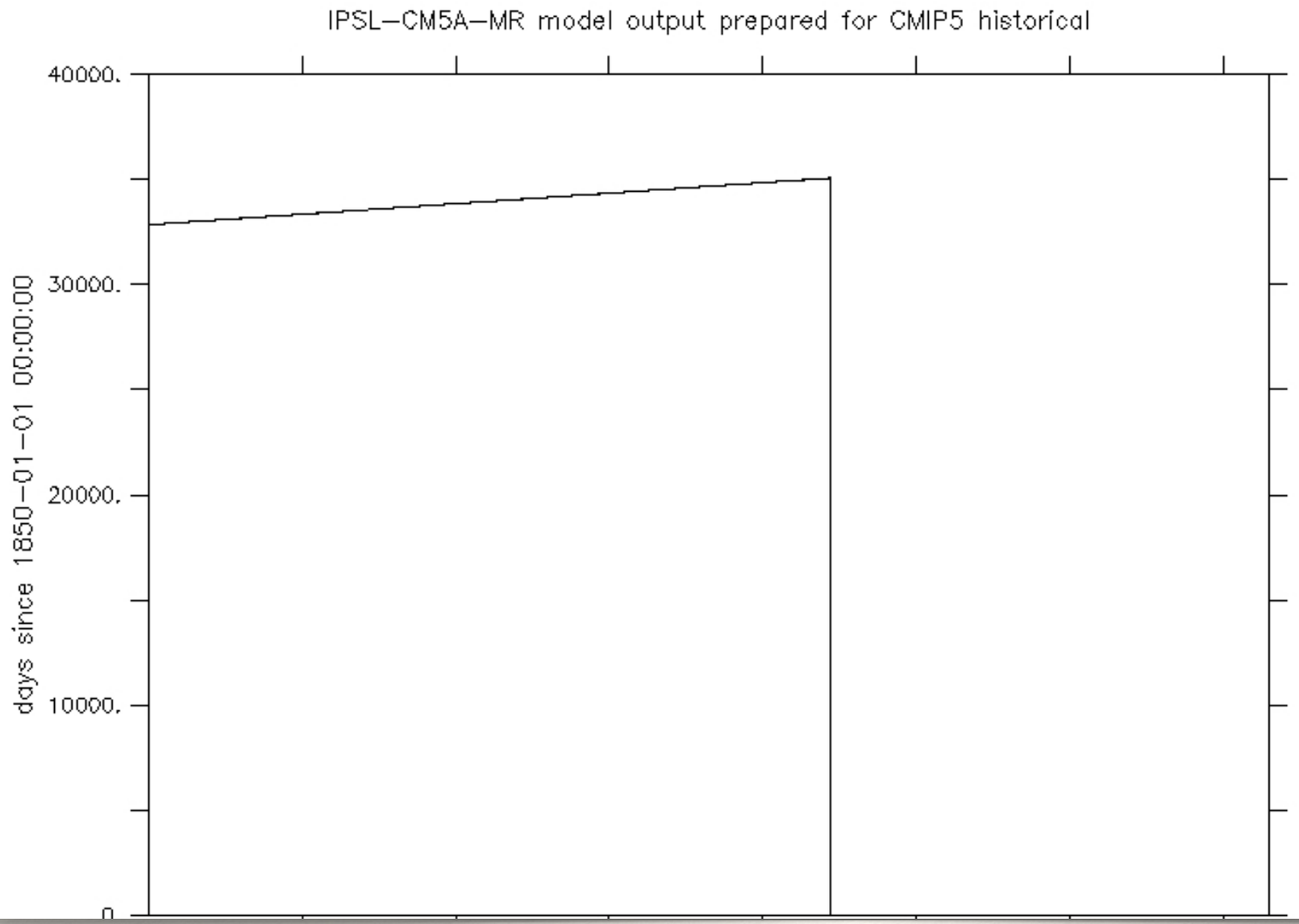
Step 4 - ES-DOC issue publishing

es-doc
Earth System Documentation

Errata Viewer v0.2.8.0 Support

CMIP5 - IPSL - My Issue Title

DETAILS | MATERIALS (1) | AFFECTED DATASETS (3)



Step 4 - ES-DOC issue publishing

es-doc
Earth System Documentation

Errata Viewer v0.2.8.0 Support

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

Step 4 - ES-DOC issue publishing

```
# Some issue information can be updated by just editing your local issue file again
```

```
{
  […]
  "description": "This is a new description, but now it's updated.",
  "status": "onhold",
  […]
}
```

```
> esgissue update --issue myissue.json --dsets mydsets.txt
INFO Validating issue...
INFO Validation Result: SUCCESSFUL
INFO Update issue #cd6ebba4-3eda-4e9a-9882-17a370c49419
INFO Issue has been updated successfully!

# On success the update date is changed accordingly into the local issue file
```

Safeguards requirements:

- The issue id, title, creation and update dates have to be unchanged,
- The issue description cannot be empty,
- The issue severity must be one of "`low`", "`medium`", "`high`" or "`critical`",
- Optional URLs must be valid and reachable,
- The issue status must be one of "`new`", "`onhold`", "`wontfix`" or "`resolved`" and cannot be changed back to "`new`",
- The submitted dataset ids must be properly formatted.

Step 4 - ES-DOC issue publishing

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 new description, but now it's updated.
Affected Experiments	historical
Affected Models	IPSL-CM5A-MR
Affected Variables	lmon
Landing Page	http://errata.ipsl.upmc.fr/issue/2
Severity	**High**
Status	**On Hold**
Created By	Guillaume @ 2016-12-06 15:23:02
Last Updated By	Guillaume @ 2016-12-06 15:23:05
Closed By	--
ESG-F Errata Identifier	cd6ebba4-3eda-4e9a-9882-17a370c49419

Step 4 - ES-DOC issue publishing

```
# Close your issue is just a shortcut to update the status to "wontfix" or "resolved"
> esgissue close --issue myissue.json --dsets mydsets.txt
INFO Validating issue...
INFO Validation Result: SUCCESSFUL
INFO Closing issue #cd6ebba4-3eda-4e9a-9882-17a370c49419
Issue status does not allow direct closing. Please change it to either (w)ontfix/
(r)esolved: r
INFO Update issue #cd6ebba4-3eda-4e9a-9882-17a370c49419
INFO Issue has been updated successfully!
INFO Issue has been closed successfully!

# On success the status is updated and the closure date is appended to the local issue
file
```

```
{
  "uid": "cd6ebba4-3eda-4e9a-9882-17a370c49419",
  "title": "My Issue Title",
  "description": "This is a new description, but now it's updated.",
  "project": "cmip5",
  "dateCreated": "2016-12-02 08:24:15",
  "dateUpdated": "2016-12-06 15:35:27",
  "dateClosed": "2016-12-06 15:35:27",
  "materials": ["http://www.my-errata.org/images/issue.jpg"],
  "severity": "high",
  "status": "resolved",
  "url": "http://www.myissue-landing-page.org/"
}
```

Step 4 - ES-DOC issue publishing

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 new description, but now it's updated.
Affected Experiments	historical
Affected Models	IPSL-CM5A-MR
Affected Variables	lmon
Landing Page	http://errata.ipsl.upmc.fr/issue/2
Severity	**High**
Status	**Resolved**
Created By	Guillaume @ 2016-12-06 15:23:02
Last Updated By	Guillaume @ 2016-12-06 15:33:27
Closed By	Guillaume @ 2016-12-06 15:35:27
ESG-F Errata Identifier	cd6ebba4-3eda-4e9a-9882-17a370c49419

Step 4 - ES-DOC issue publishing

```
# At any time your issue files can be retrieved from the server
> esgissue retrieve --id cd6ebba4-3eda-4e9a-9882-17a370c49419
INFO processing id cd6ebba4-3eda-4e9a-9882-17a370c49419
INFO Issue has been downloaded.
> ls ./issues
issue_cd6ebba4-3eda-4e9a-9882-17a370c49419.json
> ls ./dsets
dset_cd6ebba4-3eda-4e9a-9882-17a370c49419.txt
```

- Output folders for local issue files and datasets list can be respectively set with "`--issue`" and "`--dsets`" flags.
- One or several issue id(s) can be retrieved.
- If no issue id, default is to retrieve all errata issues of your modeling group/institute.

Step 5 - ES-DOC PID request

```
# Edit the list of PIDs you want to submit
> vi pid-test.txt
```

```
21.14100/d9053480-0e0d-11e6-a148-3e1d05defe78
21.14100/37043d8e-ac5e-3843-a019-c03017cc68aa
21.14100/e0560a9d-2227-3175-b943-fc26c427a923
[…]
```

Step 5 - ES-DOC PID request

es-doc
Earth System Documentation

Errata PID Lookup v0.2.8.0 Support

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

file://pid-test.txt OR Choose File Search

21.14100/d9053480-0e0d-11e6-a148-3e1d05defe78

#	Dataset / File	Issue
1	**cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20010101**	
2	**cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20020101**	
3	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20040101	
4	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20030101	

21.14100/37043d8e-ac5e-3843-a019-c03017cc68aa

#	Dataset / File	Issue
1	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20010101	
2	**cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20020101**	
3	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20030101	
4	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20040101	

21.14100/e0560a9d-2227-3175-b943-fc26c427a923

#	Dataset / File	Issue
1	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20010101	
2	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20020101	
3	**cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20030101**	
4	cmip5.output1.MPI-M.MPI-DUMMY.atef.test.dataset.ABCD#20040101	

Thank you for your attention.

- *Developers:*
  - *Guillaume Levavasseur (UPMC/IPSL - glipsl@ipsl.juisseu.fr)*
  - *Atef Bennasser (UPMC/IPSL - abennasser@ipsl.jussieu.fr)*
  - *Mark Greenslade (UPMC/IPSL - momipsl@ipsl.jussieu.fr)*
- *Contributors:*
  - Sébastien Denvil (CNRS/IPSL)
  - Merret Buurman (DKRZ)
  - Katharina Berger (DKRZ)
- *Acknowledgements to:*
  - Martina Stockhause (DKRZ)
  - Tobias Weigel (DKRZ)
  - Ag Stephens (CEDA)
  - Alan Iwi (CEDA)

"*CMIP6 Errata System*" WIP paper revised (pending for review)

"*ES-DOC and ES-DOC Services*" talk on Day 2

https://github.com/ES-DOC/esdoc-errata-client

http://esdoc-errata-client.readthedocs.io