

Register your own data

Tobias G.W. Verhulst

STCE – Royal Meteorological Institute of Belgium

PITHIA-NRF Innovation Day 3, Warsaw, 2024 June 12



The PITHIA-NRF project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101007599



Should You Register Your Data Collection?

Should you register your space weather data or models in the PITHIA-NRF e-Science center?

Should You Register Your Data Collection?

Should you register your space weather data or models in the PITHIA-NRF e-Science center?

Yes!

Should You Register Your Data Collection?

Should you register your space weather data or models in the PITHIA-NRF e-Science center?

Yes!

- All data/models related to ionosphere/thermosphere/plasmasphere (or space weather in general) are in principle welcome.
- It will help your data/models to become more FAIR, and allow integration with other data-sets into workflows.

Should You Register Your Data Collection?

Should you register your space weather data or models in the PITHIA-NRF e-Science center?

Yes!

- All data/models related to ionosphere/thermosphere/plasmasphere (or space weather in general) are in principle welcome.
- It will help your data/models to become more FAIR, and allow integration with other data-sets into workflows.

Two questions:

- 1 What to register,
- 2 How to do that.

When *not* to Register Something?

There is one point to consider for registering your space data/models:

There is no point registering your data collection if it is part of an already registered one.

When *not* to Register Something?

There is one point to consider for registering your space data/models:

There is no point registering your data collection if it is part of an already registered one.

But consider:

- Partial overlap between registered collections is fine.
- Different levels of data quality can be registered in parallel (e.g., reprocessed or cleaned data vs. original/raw data).
- If you provide a different data format or additional access method, it should be registered.

Additional Considerations

Data collections can be registered with different types of access provided: direct download, link to a repository, API, etc.

API is preferred, but not required. You can register different access types to the same data/model.

Additional Considerations

Data collections can be registered with different types of access provided: direct download, link to a repository, API, etc.

API is preferred, but not required. You can register different access types to the same data/model.



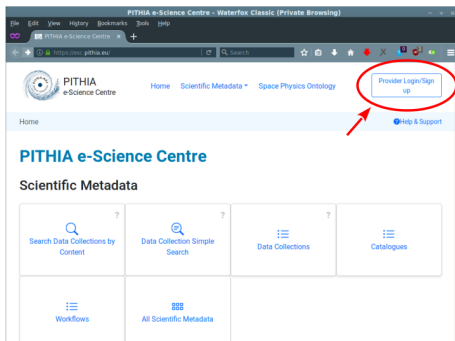
What about registering existing data repositories in the e-Science Centre, rather than separate instruments?

There are some advantages to registering data repositories rather than individual instruments: less clutter in search, consistent data access & format...

But, sending data to some central repository does not preclude also registering it in the e-Science Center.

Becoming a Data Collection Owner

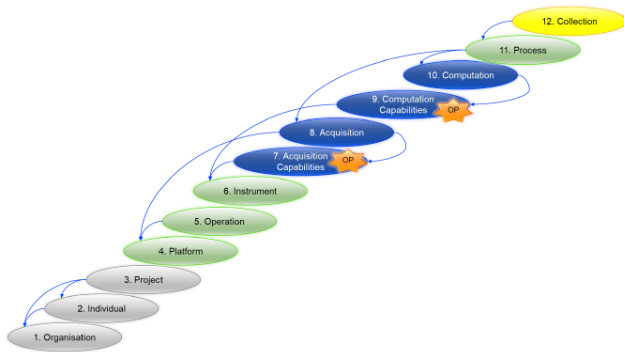
Accessing data and models through the e-Science Centre is entirely open, and does not require users to have an account. To register data *an account is needed*.



You can use an existing account (e.g. your institutional account) or create one, and someone from the consortium will approve your account.

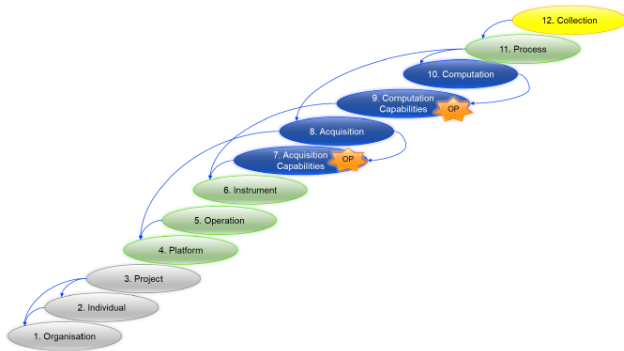
12 Steps of Data Registration

The 12 step registration process may seem daunting at first...



12 Steps of Data Registration

The 12 step registration process may seem daunting at first...



Main point is *reuse of resources*: e.g. your organisation needs to be registered only once for all your data. Also, not all steps apply in all cases.

Each step requires the creation of an XML file with meta-data, but this can now be done interactively.

What needs to be registered?

- 1 Organisation
- 2 Individual
- 3 Project
- 4 Platform
- 5 Operation
- 6 Instrument
- 7 AcquisitionCapabilities
- 8 Acquisition
- 9 ComputationCapabilities
- 10 Computation
- 11 CompositeProcess
- 12 DataCollection

What needs to be registered?

- 1 Organisation
- 2 Individual
- 3 Project
- 4 Platform
- 5 Operation
- 6 Instrument
- 7 AcquisitionCapabilities
- 8 Acquisition
- 9 ComputationCapabilities
- 10 Computation
- 11 CompositeProcess
- 12 DataCollection

You want to register step 12, but you can't just start there...

What needs to be registered?

- 1 Organisation
- 2 Individual
- 3 Project
- 4 Platform
- 5 Operation
- 6 Instrument
- 7 AcquisitionCapabilities
- 8 Acquisition
- 9 ComputationCapabilities
- 10 Computation
- 11 CompositeProcess
- 12 DataCollection

You want to register step 12, but you can't just start there...

Often you don't need all 12 steps; those you need you might be able to reuse old registrations.

What needs to be registered?

- 1 Organisation
- 2 Individual
- 3 Project
- 4 Platform
- 5 Operation
- 6 Instrument
- 7 AcquisitionCapabilities
- 8 Acquisition
- 9 ComputationCapabilities
- 10 Computation
- 11 CompositeProcess
- 12 DataCollection

You want to register step 12, but you can't just start there...

Often you don't need all 12 steps; those you need you might be able to reuse old registrations.

Steps 1 – 3 are “administrative”; will be reusable many times.

What needs to be registered?

- 1 Organisation
- 2 Individual
- 3 Project
- 4 Platform
- 5 Operation
- 6 Instrument
- 7 AcquisitionCapabilities
- 8 Acquisition
- 9 ComputationCapabilities
- 10 Computation
- 11 CompositeProcess
- 12 DataCollection

You want to register step 12, but you can't just start there...

Often you don't need all 12 steps; those you need you might be able to reuse old registrations.

Steps 1 – 3 are “administrative”; will be reusable many times.

Steps 4 to 8 identify a particular instrument, operating at a specific location. 8 is your specific instrument, the rest may be reusable.

Similar for 9 to 11 for computations (models).

The more data & models you register, the less work it is (relatively speaking)!

Summary of Most Important Best Practices

- Reuse resources as much as possible.
- When multiple access methods exist, register them all.
- Partial overlap of datasets is fine;
- Only if nothing new is added should a collection *not* be registered.
- Consider providing an API in the future if there isn't one yet.
- Register now, update later.
- Make sure to update your registration if something changes!

The end!

Questions?