Landing Page Demo for ESGF files and datasets

*2015/06/18, Merret Buurman (DKRZ), Tobias Weigel (DKRZ)*

## Abstract

The landing page demo is a Pyramid-based prototype of a landing page for ESGF data entities (files, datasets etc.) which are uniquely identified by a Handle. The user gets to see the landing page if he/she resolves a handle, i.e. if he/she "asks" the global Handle System for information about the data entity referenced by that handle. The Handle System redirects the user to the demo landing page which presents some relevant information to the user, for example about newer versions, older versions, contained files (in the case of datasets) and containing datasets (in the case of files). There is also a placeholder for information gathered from the ESGF catalog or the QC database.

## Reaching the landing page

In this demo case, the handles for ESGF files are their tracking ids (from the netcdf header) preceded by the prefix "10876.test/". Resolving a handle is easy – it is done by visiting the website [http://hdl.handle.net](http://hdl.handle.net/) and entering the handle (for example 10876.test/ f05e5f1e-f011-11e4-8220-5404a60d96b5) in the text field. Alternatively, the handle can be added to the URL: <http://hdl.handle.net/10876.test/f05e5f1e-f011-11e4-8220-5404a60d96b5>. All handles with the prefix 10876.test are then redirected to the landing page demo located at[[1]](#footnote-2) <https://handle8.dkrz.de/landingpagedemo/>.

Please note that if the user resolves a handle using a client that prefers a non-standard mime type (i.e. any mime type that is not one of the standard mime types preferred by common web browsers, such as "text/html"), the client is redirected directly to the download page of the data. This way, users can use tools such as wget in scripts to directly download the netcdf file if an “Accept: application/x-netcdf” is explicitly given. CNRI will deploy a new version of the proxy server hdl.handle.net in the near future which can also deal with cases where there is no explicit Accept header given.

## Information contained in the landing page

The landing page demo shows the user some basic information about the data entity described by/referred to by its handle record. This may include its location, creation date, check sum, etc ... Furthermore, the demo verifies if there are older or newer versions of the dataset and displays this information on the page. If a newer version is listed, the demonstrator checks whether that version is replaced by an even newer version and so on (the number of newer versions that are looked up is limited to 20). All older/newer versions are displayed as links. The user can click on them to reach their landing pages and view information about them.

The demo also checks for parent and child data entities. Files are children of datasets, so for each file, the containing dataset is displayed and for each dataset, the contained files are specified. Again, the user can click on them to reach their landing pages and view information about them. In principle, this mechanism may be extended to cover higher aggregation levels up to the DOI level. Further information, such as metadata included in the ESGF catalog or the QC data base, could be included in future versions.

If a user wishes to download a file, he/she can follow the URL(s) displayed on the landing page.

## Required handle values

The demonstrator uses of handle values of specific types to combine the information described above. The following is a list of possible values the demo currently understands. This should be further discussed and extended to eventually become a CMIP6-wide agreement on standardized Handle records.

parent or PARENT

If aggregating data entities should be displayed, they have to be specified in the handle record. None, one or several PARENT values are accepted. If none is given, there is no PARENT section. There is no recursive lookup of parents, i.e. only the next aggregation level is considered. Parents do not have to be of any specific type. Each parent handle value should contain one handle.

children or CHILDREN

If contained data entities should be displayed, they have to be specified in the handle record. None, one or several CHILDREN values are accepted. If none is given, there is no CHILDREN section. There is no recursive lookup of children. Children do not have to be of any specific type. Each children handle value should contain a list of handles: [handle1, handle2, handle3].

replacedBy or replaced\_by or isReplacedBy

None or one newer version is accepted. The existence of even newer versions (recursive newer versions) are checked up to a limit, this can be switched off in the configuration. The existence of each handle in the Handle System is checked and displayed. Each replacedBy handle value should contain one handle.

replaces

None, one or several older versions are accepted. The existence of the handle is checked. There is no recursive checking of older version. Each replaces handle value should contain one handle.

aggregationType or aggregation\_type

Aggregation type specifies if the data entity is a file, a dataset or another type. If no aggregation level is given, it is referred to as a ‘data entity’. If more than one type is specified, it is given in parentheses. Each aggregationType handle value should contain one type as string.

creation\_date or creationDate

The creationDate handle value should contain one date as a string. If several creation dates are given, they are all displayed on the page.

URL

None, one or several URL values are accepted. Each URL handle value should contain one URL.

10320/loc

This handle is required to enable content negotiation, redirecting either to the landing page or the actual entity.

# Appendix I: First test file collection

A set of 10 files has been generated to test the framework. Each file had a global attribute “tracking\_id” set to a handle. For example:

tracking\_id = "hdl:999999/f05dc00e-f011-11e4-8220-5404a60d96b5"

Because the files were created in parallel to the implementation, the domain ID in the files (“999999”) is incorrect and needs to be replaced with “10876.test”. The ten landing pages for the files are:

“Version 1”
<http://hdl.handle.net/10876.test/f05e5f1e-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f05d326a-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f05ca8f4-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f05c1876-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f05dc00e-f011-11e4-8220-5404a60d96b5>

“Version 2”
<http://hdl.handle.net/10876.test/f05eefb0-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f060bb7e-f011-11e4-8220-5404a60d96b5> <http://hdl.handle.net/10876.test/f0600a58-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f05f7caa-f011-11e4-8220-5404a60d96b5>
<http://hdl.handle.net/10876.test/f0615066-f011-11e4-8220-5404a60d96b5>

There are two additional virtual dataset handles to illustrate other features of the landing page:
<http://hdl.handle.net/10876.test/49634b69-6662-4a52-9175-45f296dc9578>
<http://hdl.handle.net/10876.test/ca9e9abd-e66e-413e-ab29-6c26fe00b859>

For each file handle it is possible to get a JSON representation of the file handle metadata, e.g.
<http://hdl.handle.net/api/handles/10876.test/f0600a58-f011-11e4-8220-5404a60d96b5>

When the user goes to the landing page of a “Version 1” file they are notified that a newer version exists.

1. The demo page can also be reached directly via https://handle8.dkrz.de/landingpagedemo/10876.test/f05e5f1e-f011-11e4-8220-5404a60d96b5 [↑](#footnote-ref-2)