CEOS

Working Group on Information Systems and Services

Data Stewardship Interest Group

Data Purge Alert Procedure
Version 1.0

Doc. Ref.:   CEOS.WGISS.DSIG.TN
Date: 18 April 2016
Issue: 1.0

Introduction

All organizations holding EO space data have the responsibility to assess the relative value of their data holdings and to preserve them for the long term. However, occasionally, an organization must make the decision to "purge" one or more datasets that could be important to help meeting the mission requirements of another agency. A Data Purge action permanently and irrecoverably removes all copies of an EO dataset held in an organization.

The Committee on Earth Observation Satellites (CEOS), an international framework for coordinating all spaceborne Earth observation missions, realizes that it is a responsibility of all organizations holding EO space data to assess the relative value of their holdings and to preserve them for the long term.

CEOS has established a "Purge Alert Procedure" service to help to ensure the long-term preservation of valuable Earth observation data. This initiative enables data archive managers to:

1. Advise other archives of Earth observation data holdings scheduled to be purged;
2. Offer these data to other archive centers.

The CEOS Working Group on Information Systems and Services (WGISS) suggests that during the data holdings Appraisal and long-term preservation processes, it is helpful to share information and findings with other data managers. It is crucial to inform other Organizations well in advance (NOAA suggests at least 12 months but we would not declare a specific period) before deciding to purge any data sets. <u>A Data Appraisal of the data set is performed during the initial phase of the data life cycle and it will provide an initial conception of whether the data set should be preserved and kept accessible and usable for the long term. The circulation of this data holdings assessment will help in the data purging decision.</u>

Thus, working as a community will be beneficial, as wiser decisions, regarding purging and transferring archival data, can be attained.
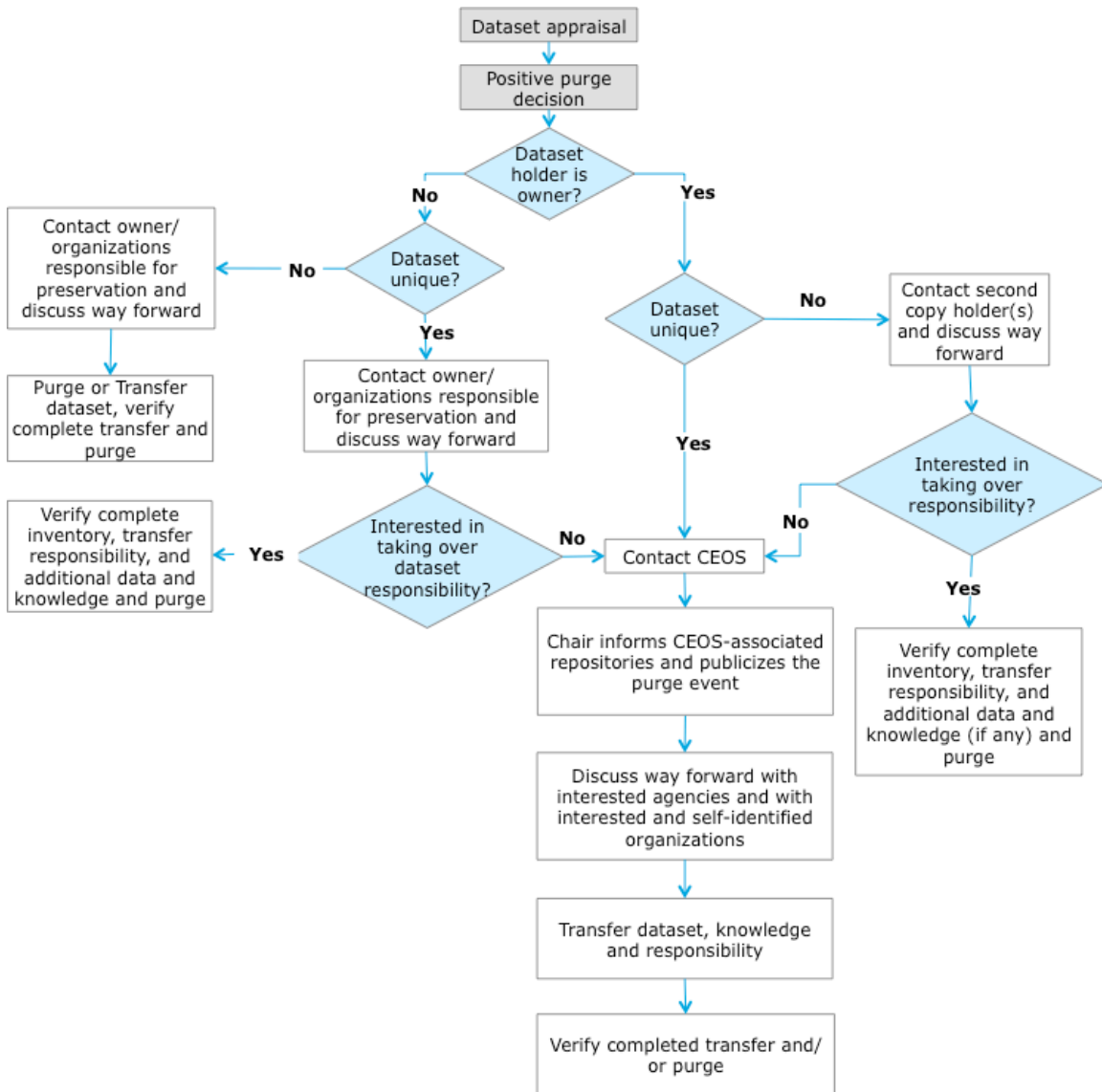
Data Purge Alert Procedure

The "Data Purge Alert" procedure aims at preventing, or at least minimizing, the loss of EO space data. Organizations intending, for whatever reason, to purge an EO dataset should apply the procedure (before purging the data) in order to inform other organizations, with the goal to trigger a possible transfer of preservation responsibility to another interested entity.

CEOS might support in some cases initiation of dialogue between a "Purging organization" and entities interested in taking over data responsibility.

Data Purge Alert Diagram

The following diagram depicts the Data Purge Alert Flow chart. The procedure described below varies depending on dataset uniqueness and purging organization responsibility versus the dataset (e.g. ownership).

The input to the Data Purge Alert Procedure is the Dataset Appraisal and a positive purge decision.

The **Dataset Appraisal** consists in performing an assessment of an EO space data set under evaluation (for example in case of possible discontinuation of preservation) through answering to a set of categorized questions extracted from the USGS EROS Appraisal Online Tool (http://eros.usgs.gov/government/ratool/) or from the CEOS website http://wgiss.ceos.org/archive/index.html.

A non-exhaustive list of questions should cover:
- Mission relevancy;
- General policy (E.g. ISO Standard);
- Physical properties (Media);
- Metadata;
- Data Records information (E.g. Size, Temporal and spatial coverage etc.)
- User community needs;
- Cost / benefit analysis.

When performing the Dataset Appraisal, an appraisal report should be produced in order to properly document the different steps performed in the appraisal procedure and the final results. The appraisal report provides all information useful to eventually make a **purge decision**. In the case of a positive purge decision the appraisal report is provided to other data holders and archive owners interested in taking over the responsibility of the Data set, together with a description of the data set, when applying the data alert procedure.

The procedure varies slightly depending on dataset uniqueness and purging organization responsibility versus the dataset (e.g. ownership).

The different control point's options are:

- Path 1- Dataset Holder is owner and Dataset is unique:
  - Contact CEOS WGISS Chair
  - Chair informs CEOS-associated repositories and publicizes the purge event
  - Discuss way forward with interested agencies and with interested and self-identified organizations
  - Transfer dataset, knowledge and responsibility
  - Verify completed transfer and/or purge

- Path 2 - Dataset Holder is owner and Dataset is not unique:
  - Contact second copy holder(s) and discuss way forward
  - If the second copy holder(s) is not interested in taking over responsibility then the path 1 is followed
  - If the second copy holder(s) is interested in taking over responsibility then verify complete inventory, transfer responsibility, and additional data and knowledge (if any) and purge.

- Path 3 - Dataset Holder is not owner and Dataset is unique:
  - Contact owner/ organization responsible for preservation and discuss way forward

o If the owner/ organization responsible for preservation is interested in taking over dataset responsibility then transfer dataset, together with the knowledge, verify complete transfer and purge.
o If the owner and preservation responsible are not interested in taking over dataset responsibility then the Path 1 shall be followed.

➢ Path 4 - Dataset Holder is not owner and Dataset is not unique:
o Contact owner/ organization responsible for preservation and discuss way forward
o Transfer dataset, verify complete transfer and/or purge

Data Purge Alert Information

Before contacting other organizations, the initiator should provide the following information:

1. *Data Records description* as a minimum in terms of: mission/sensors characteristics, products specifications, temporal coverage, geographical coverage, dataset size, media of storage, archiving format, Intellectual Property Rights IPRs, user community information, tools to work with the data, ownership and access policies.
2. *List of available associated knowledge* (e.g. tools, information).
3. Results and documentation of a *Data Record Appraisal* (e.g. http://eros.usgs.gov/government/ratool/).
4. Overview of constraints and conditions if any.
5. Recommendations concerning Data Records handling.
6. Point of Contact.

As stated in the CEOS PID Best Practices, if the Data Records Collection has a PID assigned to, the relocated data set series keeps the same PID but the PID metadata are updated to reflect the new location. A new landing page should be created and maintained by the new host. If the new host does not use PIDs and cannot maintain the metadata and landing page, a "tombstone" landing page must be created by one of the institutions, with information on the new data host. All PID management details should be specified in the purge alert agreement between the two institutions, e.g. who should update the PID metadata, commitment of persistence by the new host, is a tombstone page required, etc.

Data Purge Alert Response & Implementation

Organizations contacted, in the frame of the Data Purge Alert procedure, (e.g. dataset owners, copy holders or entities interested to take on the responsibility for the preservation of a dataset) should respond to the "Data Purge Alert" within **three months** to start conducting negotiations/assessment with the alert initiator.

For activating the procedure please contact WGISS Chair about intention to purge when no interested entity has been directly identified, through an email (see http://ceos.org/ourwork/workinggroups/wgiss/contacts/) with the following subject: <mission> Dataset to be purged and including a description of the case.

The WGISS Chair will inform all CEOS Agencies Principals and publish a ″Purge Alert News″ on CEOS and WGISS web pages. The interested agencies should respond within **three months** to start conducting negotiations/assessment with the alert initiator. The CEOS will establish contact between "Purging organization" and "Interested Entities" which responded to the announcement.

Contact Point

For any information on the Data Purge Alert procedure please contact Mirko Albani (Mirko.Albani@esa.int).
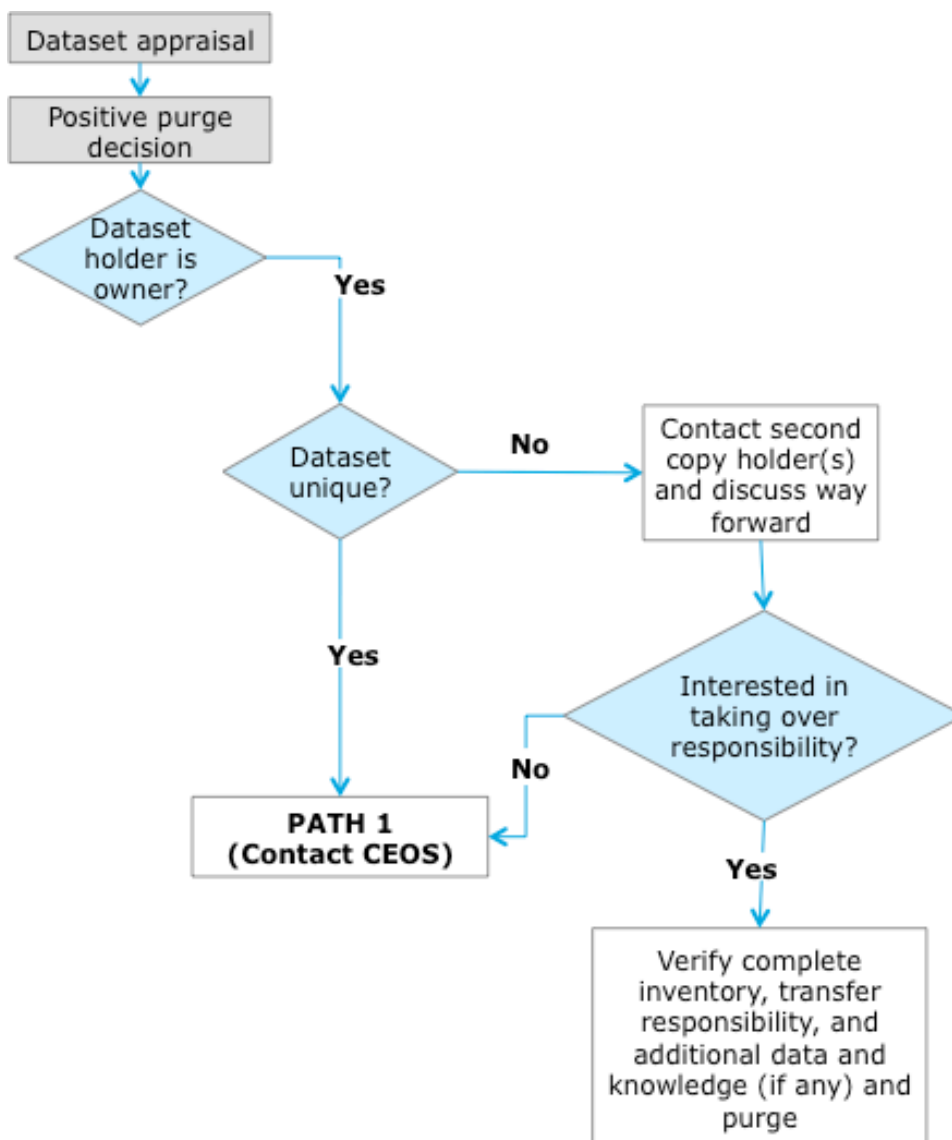
Use Cases - PATH 1

Agency A performs a Dataset Appraisal and as a result of this activity a positive purge decision is taken. Agency A is the Dataset owner and the Dataset is unique.

Agency A contacts the CEOS WGISS Chair in order to inform all possible interested agencies. In case of Agency B being interested in taking over dataset responsibility, then the Agency A transfers the responsibility, dataset and knowledge, and then proceeds to purge them.
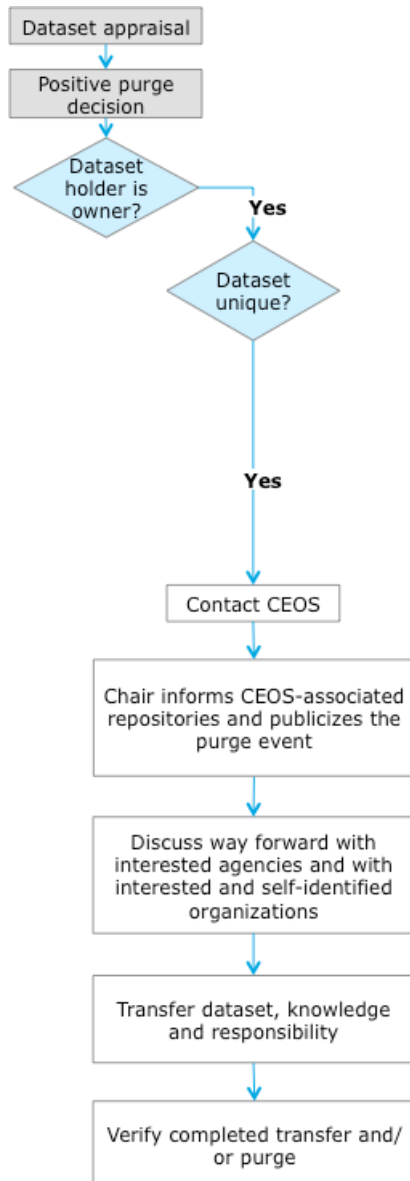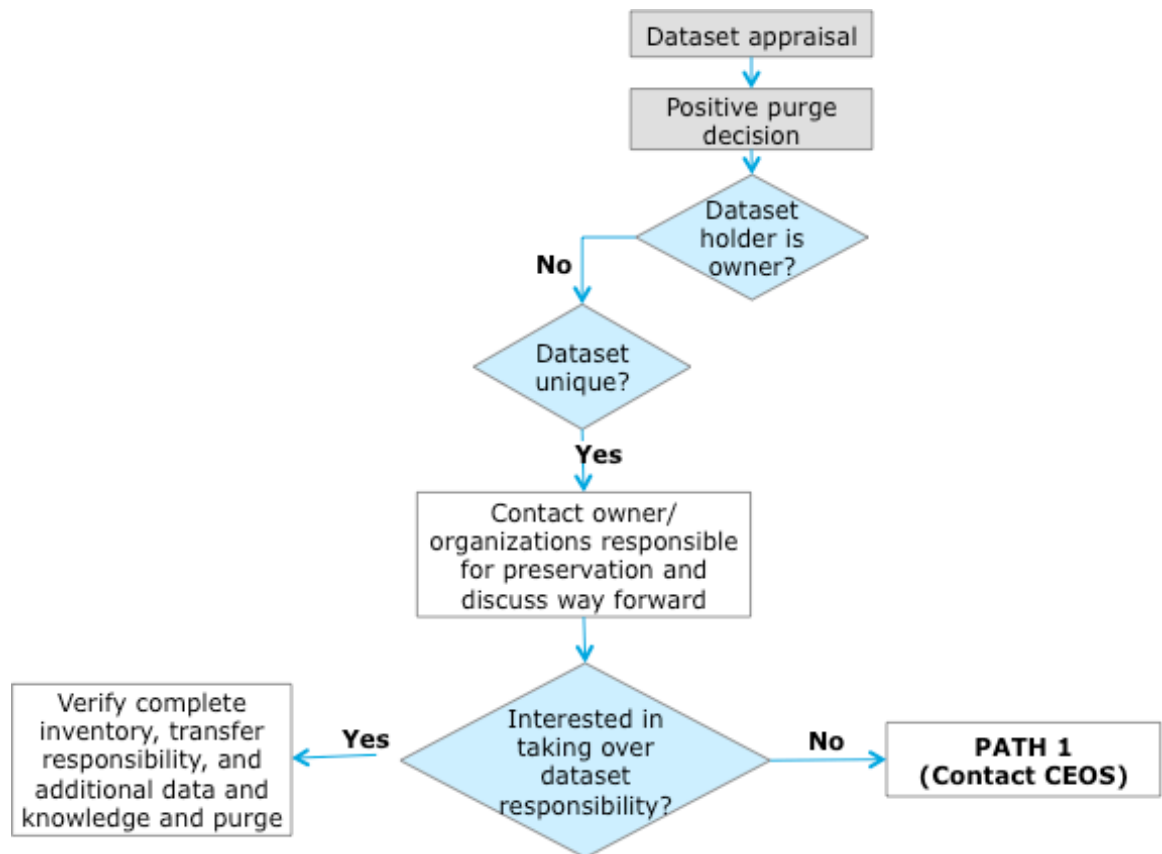
Use Cases - PATH 2

Agency A performs a Dataset Appraisal and as a result of this activity a positive purge decision is taken. Agency A is the Dataset owner but the Dataset is not unique. Agency A contacts Agency B, who is the second copy holder(s), and a way forward is discussed. If the Agency B is interested in taking over ownership responsibility the Agency A transfers the knowledge, if additional associated information is available, and missing from Agency B, and then proceeds to purge the Dataset. If Agency B is not interested in taking over ownership responsibility then Agency A contacts the CEOS WGISS Chair in order to inform all interested agencies. If any other Agency is interested to take over Dataset responsibility then Agency A transfers the responsibility, dataset and knowledge before purging them.

Use Cases - PATH 3

Agency A performs a Dataset Appraisal and as a result of this activity a positive purge decision is taken. Agency A is not the Dataset owner, but the Dataset is unique. Agency A contacts the Preservation Responsible/Dataset owner (Agency B) who could be interested to take over dataset responsibility. If this is the case then Agency A transfers the Dataset and knowledge, and verifies complete data transfer before purging them. However, if Agency B is not interested in taking over responsibility, then Agency A contacts the CEOS WGISS Chair in order to inform all interested agencies. If any other Agency is interested to take over the dataset responsibility then the Agency A transfers the responsibility, dataset and knowledge, and verify complete data transfer before purging them.

Use Cases - PATH 4

Agency A performs a Dataset Appraisal and as a result of this activity a positive purge decision is taken. Agency A is not the Dataset owner and the Dataset is not unique.

Agency A contacts the owner of the Dataset and the Preservation Responsible in order to discuss the way forward. Agency A transfers the Dataset (Data Records and Knowledge), verifies complete transfer and/or purges them.