DataCite DOI Minting UNIX Command File Usage

Script Name: datacite_json_doi_template.com

Example Usage:

datacite_json_doi_template.com *xml_filename [DATACITE/TEST_DATACITE] [YES/NO]*

Getting started: To begin using the DataCite DOI minting script and the SPASE editor update script, the run directory should have the following files:

```
curl_check.com
datacite_doi_name_list.com
datacite_doi_name_list_all.com
datacite_json_doi_info_all.com
datacite_json_doi_template.com
doi_check.com
github_person_mission.com
person_mission.com
smwg_person.com
spase_editor_update.com
spase_xml_tab.com
spase_xml_tab_wrapper.com
vxo_validation_check.com
spase_editor_update.sed
spase_editor_update.tab
```

And all '.com' command files must be executable as denoted by an asterisk '*' when listed by using the 'ls -laF' command:

```
$ ls -laF *.com

-rwxr-xr-x  1 astropooch  staff    969 Aug 22 10:49 curl_check.com*
-rwxr-xr-x  1 astropooch  staff   8138 Sep 11 15:02 datacite_doi_name_list.com*
-rwxr-xr-x  1 astropooch  staff    204 Sep 12 08:21 datacite_doi_name_list_all.com*
-rwxr-xr-x  1 astropooch  staff   3162 Sep 12 07:41 datacite_json_doi_info_all.com*
-rwxr-xr-x  1 astropooch  staff  16504 Sep 12 08:14 datacite_json_doi_template.com*
-rwxr-xr-x  1 astropooch  staff    968 Aug 22 10:56 doi_check.com*
-rwxr-xr-x  1 astropooch  staff   1823 Aug 22 10:17 github_person_mission.com*
-rwxr-xr-x  1 astropooch  staff   1025 Aug 22 10:17 person_mission.com*
-rwxr-xr-x  1 astropooch  staff   1242 Sep 11 16:02 smwg_person.com*
-rwxr-xr-x  1 astropooch  staff  12657 Sep 12 08:37 spase_editor_update.com*
-rwxr-xr-x  1 astropooch  staff   3500 Aug 22 10:17 spase_xml_tab.com*
-rwxr-xr-x  1 astropooch  staff    604 Aug 22 10:49 spase_xml_tab_wrapper.com*
-rwxr-xr-x  1 astropooch  staff   1590 Aug 22 10:56 vxo_validation_check.com*
```

Note that all of these command files are UNIX scripts. An alternate implementation of these scripts places them in a separate directory. If this approach is chosen, please ensure that the UNIX PATH environment variable is set to include the directory where the command files are located. The PATH variable can be listed as follows:

```
$ echo $PATH

/Library/Frameworks/Python.framework/Versions/3.7/bin:/opt/local/bin:/opt/local/sbin:/usr/local/bin:/usr/bin:/bin:/usr/
sbin:/sbin:/opt/X11/bin:/Users/astropooch/bin:/Applications/cdf/cdf37_0-dist/bin:/Applications/harris/idl87/bin:.
```

Each valid path directory is separated by colons. The UNIX command file that executes will be the one found in the first directory (from left to right) that contains the file so please be careful if one has multiple copies of a command file. One can use the UNIX 'which' command to verify that the desired command file is actually the one being executed (see the example command file 'which.com').

We suggest creating a directory such as we have done named **/Users/astropooch/bin** to place the UNIX command files. If one chooses this method storage for these files, just replace **astropooch** with your local username and add the chosen directory path to the PATH environment variable by issuing the following command in the BASH login file found under the computer home directory (for instance, the .bash_login file or the .bash_rc file):

```
$ declare -x PATH=$PATH:/Users/astropooch/bin
```

Some UNIX implementations may use an alternative command for updating the UNIX PATH environment variable.

Note that the spase_editor_update.sed 'stream editor' file and spase_editor_update.tab tab-separated table file would be the only files remaining in the run directory if the UNIX command files are stored elsewhere. These two files are required to be present for the spase_xml_tab.com command file to run to full completion. We will return to the utility of these non-command files below.

Prior to running **datacite_json_doi_template.com**, one must issue Git commands in order to clone the SPASE XML repositories required to perform the DOI minting process and the subsequent update to the SPASE description to add the DOI related metadata text.

For instance, the Git repository containing the SPASE description to receive the newly minted DOI must be cloned as well as any other Gits that contain metadata (i.e., any referenced Person, Repository, or Instrument SPASE description) that are referenced within the SPASE description in need of the DOI. For the present example, we only need two Gits: NASA and SMWG. The Git commands and outputs generated are as follows:

```
$ git clone git@github.com:hpde/NASA

Cloning into 'NASA'...
remote: Enumerating objects: 77272, done.
remote: Counting objects: 100% (7282/7282), done.
remote: Compressing objects: 100% (1706/1706), done.
remote: Total 77272 (delta 2609), reused 7058 (delta 2434), pack-reused 69990
Receiving objects: 100% (77272/77272), 17.83 MiB | 7.46 MiB/s, done.
```

```
Resolving deltas: 100% (32553/32553), done.
Updating files: 100% (2924/2924), done.

$ git clone git@github.com:hpde/SMWG

Cloning into 'SMWG'...
remote: Enumerating objects: 60554, done.
remote: Counting objects: 100% (2596/2596), done.
remote: Compressing objects: 100% (805/805), done.
remote: Total 60554 (delta 1731), reused 2434 (delta 1619), pack-reused 57958
Receiving objects: 100% (60554/60554), 9.07 MiB | 6.24 MiB/s, done.
Resolving deltas: 100% (48916/48916), done.
Updating files: 100% (9712/9712), done.
```

The run directory now has two subdirectories: NASA and SMWG that contain the full, up to date SPASE content for these two Gits.

We can now issue the **datacite_json_doi_template.com** command to receive a newly minted DOI from DataCite and subsequently update the SPASE description for which the DOI was desired. For this example, we will act on the SPASE description for magnetic field data in RTN coordinates at 0.092 s time resolution from the Wind spacecraft. The SPASE description file name is NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml, which is the first parameter passed to **datacite_json_doi_template.com** in the following example command:

datacite_json_doi_template.com NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml DATACITE YES

The second parameter, set to 'DATACITE', directs the command file to request a real DOI from DataCite and not a DOI from the DataCite DOI testing service. Test DOIs are returned if the second parameter is set equal to 'TEST_DATACITE'. The third and last parameter passed to the command file, here set to 'YES', prompts the script to test the integrity of any URLs that are listed in the Wind magnetic field SPASE description. The alternative setting, 'NO', results in no testing of the URLs, which is useful in order to speed up the DOI minting process when the URLs are known to be valid. The output from the execution of the command is shown in tiny 4-pt font format in the last page of this document.

Brief description of the utilities of the spase editor update.tab and spase editor update.sed files:

The **spase_editor_update.tab** file contains a table of information with one line per SPASE description. The table has the following nine column headings: **Num, Metadata File Name, SPASE Resource ID from XML description, SPASE Revise, SPASE Review, Metadata Contact, Revision Note, Revised SPASE Resource ID, Review Notes.** For the Example shown in this document, the following line with nine tab-separated text fields was appended:

1) Num: **1368**
2) Metadata File Name: **NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml**
3) SPASE Resource ID from XML: **NASA/NumericalData/Wind/MFI/RTN/PT0.092S**
4) SPASE Revise: **AK**
5) SPASE Review: **AK**

6) Metadata Contact: **Andriy.Koval**
7) Revision Note: **Added DOI and PublicationInfo minted by AK, metadata versioned up to SPASE 2.6.0, reviewed by AK 20230912**
8) Revised SPASE Resource ID: **[blank text field]**
9) Review Notes: **[blank text field]**

The **spase_editor_update.sed** file is used to make routine updates to SPASE descriptions that are known to be required. For instance, all instances of:

<RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF</RepositoryID>

need to be replaced by the following:

<RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF/CDAWeb</RepositoryID>

The **spase_editor_update.sed** file will be increase in size as such new blanket changes are required.

Other Output from running the datacite json doi template.com:

The following directories and files are also produced when executing **datacite_json_doi_template.com:**

DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml: The updated version of the SPASE description including newly minted DOI information. This version of the xml would be ingested into the NASA git upon passing inspection. See the change log listed on the last page below to identify what metadata content has been updated to aid the inspection and validation step prior to pushing the updated metadata to the git.

datacite_doi_name_list_all.tab (Author/Creator information):

SPASE_PERSON_RESOURCE_ID FULL_NAME FIRST_NAME AUTHORS_GIVEN_NAME FAMILY_NAME INITIAL_NAME AFFILIATION_NAME AFFILIATION_ROR MISSION_ACRONYM NAME_PATTERN

SMWG/Person/Andriy.Koval Andriy Koval Andriy Andriy Koval A. Koval Goddard Planetary Physics Institute, University of Maryland, Baltimore County WIND N_N
SMWG/Person/Adam.Szabo Adam Szabo Adam Adam Szabo A. Szabo NASA Goddard Space Flight Center WIND N_N

datacite_json_doi_template_all.tab (summary of metadata from DOI minting):

NEW 10.48322/VGSK-P788 10.48322 vgsk-p788 WI_H4-RTN_MFI
NASA/NumericalData/Wind/MFI/RTN/PT0.092SDATACITE/JSON/WIND/datacite_json_doi_template_wi_h4-rtn_mfi.json

DATACITE/datacite_curl_put_wind_20230912.com (the DOI minting curl command, password deliberately replaced):

```
curl -ksSX POST -H 'Content-Type: application/vnd.api+json' -u 'heliophy.spdf:XXX___PASSWORD___XXX' -d
@DATACITE/JSON/WIND/datacite_json_doi_template_wi_h4-rtn_mfi.json https://api.datacite.org/dois
```

DATACITE/JSON/WIND/datacite_json_doi_template_wi_h4-rtn_mfi.json (JSON file posted to DataCite for minting the DOI):


```
{
  "data": {
    "type": "dois",
    "attributes": {
      "doi": "",
      "prefix": "10.48322",
      "suffix": "",
      "url": "https://hpde.io/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.html",
      "types": {
        "ris": "DATA",
        "bibtex": "misc",
        "citeproc": "dataset",
        "schemaOrg": "Dataset",
        "resourceTypeGeneral": "Dataset"
      },
      "creators": [
        {
          "name": "Koval, Andriy",
          "nameType": "Personal",
          "givenName": "Andriy",
          "familyName": "Koval",
          "affiliation": [
            {
              "name": "Goddard Planetary Physics Institute, University of Maryland, Baltimore County"
            }
          ],
          "nameIdentifiers": []
        },
        {
          "name": "Szabo, Adam",
          "nameType": "Personal",
          "givenName": "Adam",
          "familyName": "Szabo",
          "affiliation": [
            {
              "name": "NASA Goddard Space Flight Center"
            }
          ],
          "nameIdentifiers": []
        }
      ],
      "titles": [
        {
          "lang": "en",
          "title": "Wind Magnetic Field Investigation (MFI) Full Resolution Data in RTN Coordinates",
          "titleType": null
        }
      ],
```

```
      "publisher": "NASA Space Physics Data Facility",
      "subjects": [
        {
          "subject": "MagneticField"
        }
      ],
      "contributors": [],
      "dates": [],
      "publicationYear": 2023,
      "language": "en",
      "identifiers": [
        {
          "identifier": "spase://NASA/NumericalData/Wind/MFI/RTN/PT0.092S",
          "identifierType": "SPASE"
        }
      ],
      "sizes": [],
      "formats": [
        {
          "subject": "CDF"
        },
        {
          "subject": "CSV"
        },
        {
          "subject": "Text"
        }
      ],
      "rightsList": [
        {
          "rights": "Creative Commons Zero v1.0 Universal",
          "rightsUri": "https://creativecommons.org/publicdomain/zero/1.0/legalcode",
          "schemeUri": "https://spdx.org/licenses/",
          "rightsIdentifier": "cc0-1.0",
          "rightsIdentifierScheme": "SPDX"
        }
      ],
      "descriptions": [
        {
          "lang": "en",
          "description": "This data product contains Wind MFI full resolution magnetic field magnitude and RTN Cartesian components. Time resolution is
typically 0.046s or 0.092s from launch through 1997, and is 0.092s thereafter. Final Version 5 data are accessible to within about 3 months of current
date, newly defined Version 4 data (with final Bz offsets and not-yet-final spacecraft position vectors) between 3 months and about 2 weeks of current
date, and newly defined Version 3 data (with the most recently determined Bz offset value, not yet final for the Version 3 interval) from 2 weeks to about
2 days of current.",
          "descriptionType": "Abstract"
        }
      ],
      "geoLocations": [],
      "fundingReferences": [],
      "relatedIdentifiers": [],
      "schemaVersion": "http://datacite.org/schema/kernel-4",
      "providerId": "nasasmd",
      "clientId": "heliophy.spdf",
      "agency": "datacite",
      "state": "draft"
    }
  }
}
```

Output text from the command: datacite_json_doi_template.com NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml DATACITE YES

```
################################################################################################################################################################################################################################################################

Start Date: Tue Sep 12 12:42:52 PDT 2023

################################################################################################################################################################################################################################################################

Old File:   DOI/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml
New File:   NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

SPASE_EDITOR_PROCESS_UPDATE        NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

AK          2,3c2,3
AK          < <Spase xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.spase-group.org/data/schema" xsi:schemaLocation="http://www.spase-group.org/data/schema http://www.spase-group.org/data/schema/spase-2_5_0.xsd">
AK          <    <Version>2.5.0</Version>
AK          ---
AK          > <Spase xmlns="http://www.spase-group.org/data/schema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.spase-group.org/data/schema http://www.spase-group.org/data/schema/spase-2_6_0.xsd">
AK          >    <Version>2.6.0</Version>
AK          8,9c8,9
AK          <          <DOI></DOI>
AK          <          <ReleaseDate>2023-07-11T00:00:00</ReleaseDate>
AK          ---
AK          >          <DOI>10.48322/gmt4-1397</DOI>
AK          >          <ReleaseDate>2023-09-12T12:34:56.789</ReleaseDate>
AK          12,13c12,17
AK          <                <ReleaseDate>2023-07-11T00:00:00</ReleaseDate>
AK          <                <Note>Initial SPASE XML resource description submission, metadata generated by AK</Note>
AK          ---
AK          >                <ReleaseDate>2023-09-12T12:34:56.789</ReleaseDate>
AK          >                <Note>Only known prior ReleaseDate of the metadata</Note>
AK          >             </RevisionEvent>
AK          >             <RevisionEvent>
AK          >                <ReleaseDate>2023-09-12T12:34:56.789</ReleaseDate>
AK          >                <Note>Added DOI and PublicationInfo minted by AK, metadata versioned up to SPASE 2.6.0, reviewed by AK 20230912</Note>
AK          17a22,26
AK          >          <PublicationInfo>
AK          >             <Authors>Koval, Andriy; Szabo, Adam</Authors>
AK          >             <PublicationDate>2023-01-01T00:00:00</PublicationDate>
AK          >             <PublishedBy>NASA Space Physics Data Facility</PublishedBy>
AK          >          </PublicationInfo>
AK          33c42
AK          <          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF</RepositoryID>
AK          ---
AK          >          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF/CDAWeb</RepositoryID>
AK          50c59
AK          <          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF</RepositoryID>
AK          ---
AK          >          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF/CDAWeb</RepositoryID>
AK          62c71
AK          <          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF</RepositoryID>
AK          ---
AK          >          <RepositoryID>spase://SMWG/Repository/NASA/GSFC/SPDF/CDAWeb</RepositoryID>

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

SPASE_EDITOR_URL_CHECK NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        HTTP/1.1 200 OK          HTTPS          https://wind.nasa.gov
SPASE_EDITOR_URL_CHECK NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml FTP        FTP Server not verified  FTPS           ftps://spdf.gsfc.nasa.gov/pub/data/wind/mfi/mfi_h4-rtn/
SPASE_EDITOR_URL_CHECK NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        HTTP/1.1 200 OK          HTTPS          https://spdf.gsfc.nasa.gov/pub/data/wind/mfi/mfi_h4-rtn/
SPASE_EDITOR_URL_CHECK NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        HTTP/1.1 200 OK          HTTPS          https://cdaweb.gsfc.nasa.gov
SPASE_EDITOR_URL_CHECK NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        HTTP/1.1 200 OK          HTTPS          https://cdaweb.gsfc.nasa.gov/hapi

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

SPASE_EDITOR_ID_CHECK  NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        SMWG      InstrumentID   SMWG/Instrument/Wind/MFI
SPASE_EDITOR_ID_CHECK  NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        SMWG      PersonID       SMWG/Person/Adam.Szabo
SPASE_EDITOR_ID_CHECK  NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        SMWG      PersonID       SMWG/Person/Andriy.Koval
SPASE_EDITOR_ID_CHECK  NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml   DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml OOO        SMWG      RepositoryID   SMWG/Repository/NASA/GSFC/SPDF/CDAWeb

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

SPASE_EDITOR_DOI_CHECK DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml NASA/NumericalData/Wind/MFI/RTN/PT0.092S               10.48322/gmt4-1397       NON

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

SPASE Validation Check: DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml

SPASE OK, Yataaaaa !!!: DOI/DATACITE/NASA/NumericalData/Wind/MFI/RTN/PT0.092S.xml

dyld[59121]: missing symbol called

+--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+

################################################################################################################################################################################################################################################################

Stop Time:  Tue Sep 12 12:42:57 PDT 2023

################################################################################################################################################################################################################################################################
```