

Supplementary Material for:

On the Utility of Identification Schemes for Digital Earth Science Data: An Assessment and Recommendations

Ruth E. Duerr¹, Robert R. Downs², Curt Tilmes³, Bruce Barkstrom⁴, W. Christopher Lenhardt⁵, Joe Glassy⁶, Luis E. Bermudez⁷, Peter Slaughter⁸

¹*University of Colorado at Boulder, National Snow and Ice Data Center, Boulder CO 80309-0449, email: rduerr@nsidc.org*

²*Columbia University, Center for International Earth Science Information Network (CIESIN), 202 Geoscience, Lamont-Doherty Earth Obs., Palisades, NY, 10964, USA, email: rdowns@ciesin.columbia.edu*

³*NASA Goddard Space Flight Center, Greenbelt MD 20771, email: Curt.Tilmes@nasa.gov*

⁴*15 Josie Lane, Asheville, NC 28804 email: brbarkstrom@gmail.com*

⁵*Environmental Sciences Division, Oak Ridge National Laboratory email: lenhardtc@ornl.gov*

⁶*Lupine Logic, Inc., Missoula, Montana email: um.glassy@gmail.com*

⁷*Open Geospatial Consortium (OGC), 483B Carlisle Drive, Herndon, VA 20170, email: lbermudez@opengeospatial.org*

⁸*University of California at Santa Barbara, Earth Research Institute, Santa Barbara, CA 93106-3060, email: peter@eri.ucsb.edu*

The FOO Example

An illustration of what can you do when you use DOIs as Unique Locators data sets and UUIDs as Unique Identifiers for data files or granules.

NASA launches on 2000-01-01 the "FOO" instrument. It captures 1 granule per month of data, so in the first year 12 granules are captured.

The processing flow has these data sets or Earth Science Data Types (ESDT) in NASA parlance:

FOOL0 - 1 Month of Level 0 data, the data captured from the instrument

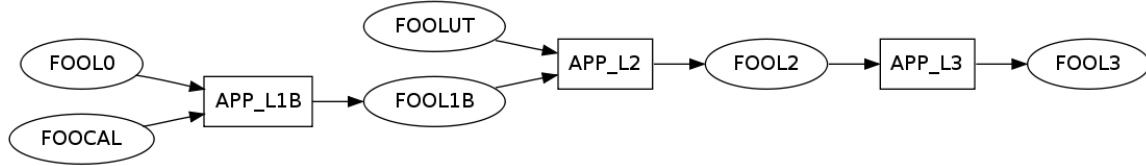
FOOCAL - Calibration data produced by the CAL team used to calibrate the data

FOOL1B - 1 Month of Level 1 calibrated data

FOOL2 - 1 Month of Level 2 data with some geophysical parameters retrieved.

FOOL3 - 1 year of gridded data.

We have 3 programs that process the data to successively higher levels as depicted in the following data flow diagram: APP_L1B, APP_L2, and APP_L3.



Our LOCALGRANULEIDs look like this:

- <ESDT>.<granulenum> for Level 0
- <ESDT>.v<collection> for calibration data and
- <ESDT>.v<collection>.<granule number> for the others

Initial Processing for Year 1:

We start from these files:

FOOL0.1
FOOL0.2
...
FOOL0.12

FOOCAL.1 (we've only calibrated once)

We run FOOL1BP 12 times, inputting each of the L0 granules and the calibration file to produce these files:

FOOL1B.v1.1
FOOL1B.v1.2
...
FOOL1B.v1.12

Then we run FOOL2P on each of those, producing the level 2 data:

FOOL2.v1.1
FOOL2.v1.2
...
FOOL2.v1.12

Finally, we run FOOL3P once, reading each Level 2 file for the year, producing a single file:

FOOL3.v1.1

The First Reprocessing

Now the calibration guys do their magic and come up with a better FOOCAL.2, so we reprocess all of the Level 0 data and all of the higher level products creating collection 2, producing 25 more files:

FOOL1B.v2.1

FOOL1B.v2.2

...

FOOL1B.v2.12

FOOL2.v2.1

FOOL2.v2.2

...

FOOL2.v2.12

FOOL3.v2.1

All of those files go into the archive.

Now we are looking for some good identifiers.

Data Set Identifiers

We're proposing DOI to be used to identify "ESDT+Collection", so lets assign some:

doi:10.9999/US/FOOL0 => FOOL0

doi:10.9999/US/FOOL1B.v1 => FOOL1B, collection 1

doi:10.9999/US/FOOL1B.v2 => FOOL1B, collection 2

doi:10.9999/US/FOOL2.v1 => FOOL2, collection 1

doi:10.9999/US/FOOL2.v2 => FOOL2, collection 2

doi:10.9999/US/FOOL3.v1 => FOOL3, collection 1

doi:10.9999/US/FOOL3.v2 => FOOL3, collection 2

Since these are "Open Data Sets", those identifiers aren't sufficient to determine the precise set of granules used (right now, they map to 12 granules, but next month they will map to 13 granules after the next Level 0 data file is acquired and processed).

What they are useful for is referring to the general set of granules, and locating

them and even (which not perfect), citing them.

For example, you can put "<http://dx.doi.org/10.9999/US/FOOL2.v1>" into your browser and it will take you right to the web page at the archive with collection level metadata about that ESDT and information on how to download the data, etc.

If you write a paper using the L2 data, you can add "doi:10.9999/US/FOOL2.v2" into your data citation and someone else can follow it later back to that data.

Granule Identifiers

As the data processing group received the L0 data and produced the higher level products, they tagged them with UUIDs to keep track of them, also to distinguish their filenames, they decided to append each Localgranuleid with the UUID.

Here's the whole list of files so far:

ESDT: FOOL0.001
Shortname: FOOL0
Longname: "Monthly Level 0 data from FOO"
DOI: 10.9999/US/FOOL0

FOOL0.01.fa9cf2e0-b60b-43b7-baee-d18cc185b407
FOOL0.02.8a29012d-be8d-4af1-a158-783cfdcfc7fc
FOOL0.03.90463be2-a1a4-4d63-ad70-2f1f3c09798e
FOOL0.04.c121f001-6851-433f-9400-8e3acaa0229a
FOOL0.05.2558aa9a-ce8c-47df-ac05-0fc982568462
FOOL0.06.7223358c-f92d-42d0-bff4-1cd6140d4a89
FOOL0.07.e2e145cd-899e-483f-989a-fbf1017a3df8
FOOL0.08.d0e94ea8-08b3-4bdf-af3f-7af72f2f9221
FOOL0.09.ddc94379-0086-4817-9595-5fbe378c5a29
FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7
FOOL0.11.27d94c01-51ed-45c4-8c50-e19ca7f20882
FOOL0.12.0af9c3a5-6ee5-4435-8437-d96ebfa36625

ESDT: FOOCAL.001
Shortname: FOOCAL
Longname: "Calibration data for FOO"
DOI: 10.9999/US/FOOCAL

FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

ESDT: FOOL1B.001
Shortname: FOOL1B
Longname: "Level 1B, monthly calibrated radiances from FOO"
Collection: 1
DOI: 10.9999/US/FOOL1B.v1

FOOL1B.v1.01.974b98ed-5d3e-4296-a80d-002fecc764e7
FOOL1B.v1.02.a24dce74-f83f-4b31-973f-db4e4334e3de
FOOL1B.v1.03.802e3d7a-86e5-4e11-b84d-58d85626cbd6
FOOL1B.v1.04.5aa1de0c-688e-413d-8ec2-20ee083a9d58
FOOL1B.v1.05.c70c2dd6-6c2a-495a-976c-cca87361948a
FOOL1B.v1.06.9d75dc70-27d9-4456-a6e7-a4f5754fef77
FOOL1B.v1.07.70aaf233-087a-426b-8cb3-34d80328b2f1
FOOL1B.v1.08.29a11b08-f55b-4bfc-ae0b-10b21eefc20b
FOOL1B.v1.09.6260b254-de84-42bd-a73a-07b08a84dfbf
FOOL1B.v1.10.4b93b41a-f812-47fa-aab5-2db8bcd4282
FOOL1B.v1.11.bdc6dd75-d76a-42c7-a29d-5f88bfd99eb8
FOOL1B.v1.12.4840f38e-81c2-4440-bd7b-8d846e88debe

ESDT: FOOL2.001
Shortname: FOOL2
Longname: "Level 2 data, monthly geophysical parameter from FOO"
Collection: 1
DOI: 10.9999/US/FOOL2.v1

FOOL2.v1.01.b9149c24-9356-4fcd-baa0-50cee9e17ca2
FOOL2.v1.02.8356d66b-167e-4410-81b3-e05d89052ea7
FOOL2.v1.03.e7293f65-5aa1-476b-8e5b-ebe04c4a552c
FOOL2.v1.04.84024c84-a563-429e-bc5a-00967565f2ea
FOOL2.v1.05.c8501462-1690-431e-8477-a452792dadce
FOOL2.v1.06.c92b1dff-f3f6-4c6a-ad84-9f8ce0b55a1b
FOOL2.v1.07.51a3bc66-fbc9-4393-8ee1-b257d59aea17
FOOL2.v1.08.c154f53c-3d7b-4a19-8384-7f0e11192651
FOOL2.v1.09.f7843584-1a32-46e5-9cb0-404aa2ed903d
FOOL2.v1.10.dc298ee5-c88b-471a-8d7f-390ea0eaff23
FOOL2.v1.11.b8782b3f-d228-4029-87c9-4d2ae3104d4f
FOOL2.v1.12.44032e38-5248-4b3b-917c-c1e088f43578

ESDT: FOOL3.001
Shortname: FOOL3
Longname: "Annual gridded data from FOO"
Collection: 1
DOI: 10.9999/US/FOOL3.v1

FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4

ESDT: FOOL1B.002
Shortname: FOOL1B
Longname: "Level 1B, monthly calibrated radiances from FOO"
Collection: 2
DOI: 10.9999/US/FOOL1B.v2

FOOL1B.v2.01.d615b4f6-5e35-49f0-834a-ee199db7597c
FOOL1B.v2.02.6c1a5a3b-55ab-4b53-8659-982d591cc744
FOOL1B.v2.03.58575454-3a4e-46af-8cb5-27c6aa4321cc
FOOL1B.v2.04.09124f68-3f14-446b-a1aa-d7f57e7f1603
FOOL1B.v2.05.6091c13e-5ea4-44c3-92cc-f20823249421
FOOL1B.v2.06.fcf1bb4c-51ea-464e-9935-b7653354ef73
FOOL1B.v2.07.424aa11d-fdfb-4a63-9b1b-0a386d23b1fa
FOOL1B.v2.08.e3660e2e-4248-43ea-b91c-64799f3a1e74
FOOL1B.v2.09.9ca12548-2a6b-47e4-9f7d-732013984ec1
FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e
FOOL1B.v2.11.6cfec73e-bf2e-42cb-a427-2d30694f43e8
FOOL1B.v2.12.5cb6c2d9-386c-4d87-a103-56f0e459a26f

ESDT: FOOL2.002
Shortname: FOOL2
Longname: "Level 2 data, monthly geophysical parameter from FOO"
Collection: 2
DOI: 10.9999/US/FOOL2.v2

FOOL2.v2.01.bba34792-f256-4c54-81dd-9977e432c204
FOOL2.v2.02.2fd12da6-a3e2-4e50-8140-3ac645882419
FOOL2.v2.03.29bda893-765d-476d-851b-8b9acd7f140e
FOOL2.v2.04.57509ddb-3d40-4d60-8204-da4b99867fc7
FOOL2.v2.05.0e8604fa-fb4e-4cfb-b412-5364ca12cf14
FOOL2.v2.06.0eb26b4e-b718-41c5-bbf8-c83d3d79c233

FOOL2.v2.07.43079ea6-43b5-4622-b492-bcdb824a818e
FOOL2.v2.08.590fd64c-ec12-44a5-9b14-0042d19ed3dc
FOOL2.v2.09.226173b9-4ef7-49e8-8b9e-701b892a8f57
FOOL2.v2.10.533b2a95-d57f-4f75-9b7d-914d3d220310
FOOL2.v2.11.af235d11-777c-4bf1-a5e6-15273a5e5d80
FOOL2.v2.12.bdc9dc33-38bd-403c-991e-48dcd4762ca7

ESDT: FOOL3.002
Shortname: FOOL3
Longname: "Annual gridded data from FOO"
Collection: 2
DOI: 10.9999/US/FOOL3.v2

FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768

Data is Cited in a Publication

Alice compares the FOO collection 1 Level 3 data to the collection 2 Level 3 data and publishes a paper discussing the differences between the two versions.

She cites the data, including in her two references the two DOIs to the level 3 data:

doi:10.9999/US/FOOL3.v1 and doi:10.9999/US/FOOL3.v2

The publisher assigns her paper its own DOI:
doi:10.8888/1

After that, 2 more months of L0 data come in and get added to the database:

FOOL0.13.76e9b680-2d06-4a55-ad2c-79b533ce86ca
FOOL0.14.f6bf9378-4215-41b5-9d3e-96dc0c2e7eeb

The production team process the new L0 data to create new L1B and L2 data in collection 2 (they don't even bother extending collection 1, since it used the old, obsolete calibration).

They produce these 4 new granules:

FOOL1B.v2.13.e36a5d4d-1687-4947-bca8-2312c5f6eb8f
FOOL1B.v2.14.eefa4113-200c-4a97-b5c6-3bd4b695cbc5

FOOL2.v2.13.f8f9564d-cc2a-4760-b1bc-13f1ef5cbdcb

FOOL2.v2.14.4814ed46-0e41-4e3f-8f73-33d0cd2ef0bc

Not only that, the team discovers that the L0 data from month 10 were corrupt and they obtain corrected data, replacing

FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7

file with a corrected file:

FOOL0.10.3adf6dea-06af-478f-8216-2bbcdb0caad2

They also re-run that granule through L1B, L2, and L3, replacing these files:

FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e

FOOL2.v2.10.533b2a95-d57f-4f75-9b7d-914d3d220310

FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768

with these new ones:

FOOL1B.v2.10.b8f1e287-2d92-4340-9a71-5da5af191f3b

FOOL2.v2.10.6e58a410-60e7-4956-aeaf-37f76a16b171

FOOL3.v2.01.2a365058-fb52-4559-ab4b-085cb5ac0b73

Now, the official archive is distributing the new files, but we've still got a paper out there that was developed using the old, bad data. Note that using simply the DOI for the collection, as a whole, is not sufficient to indicate that fact.

Including Provenance in the Archive

Here's the whole database so far, starting to include some rudimentary provenance:

Paper: doi:10.8888/1

Author: Alice

References: doi:10.9999/US/FOOL3.v1
 doi:10.9999/US/FOOL3.v2

Granules actually used:

FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4

FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768

FOOL0.001

Shortname: FOOL0

Longname: "Monthly Level 0 data from FOO"

DOI: 10.9999/US/FOOL0

FOOL0.01.fa9cf2e0-b60b-43b7-baee-d18cc185b407

FOOL0.02.8a29012d-be8d-4af1-a158-783cfdcfc7fc

FOOL0.03.90463be2-a1a4-4d63-ad70-2f1f3c09798e

FOOL0.04.c121f001-6851-433f-9400-8e3acaa0229a

FOOL0.05.2558aa9a-ce8c-47df-ac05-0fc982568462

FOOL0.06.7223358c-f92d-42d0-bff4-1cd6140d4a89

FOOL0.07.e2e145cd-899e-483f-989a-fbf1017a3df8

FOOL0.08.d0e94ea8-08b3-4bdf-af3f-7af72f2f9221

FOOL0.09.ddc94379-0086-4817-9595-5fbe378c5a29

FOOL0.10.3adf6dea-06af-478f-8216-2bbcdb0caad2 (new updated file)

FOOL0.11.27d94c01-51ed-45c4-8c50-e19ca7f20882

FOOL0.12.0af9c3a5-6ee5-4435-8437-d96ebfa36625

FOOL0.13.76e9b680-2d06-4a55-ad2c-79b533ce86ca

FOOL0.14.f6bf9378-4215-41b5-9d3e-96dc0c2e7eeb

FOOCAL.001

Shortname: FOOCAL

Longname: "Calibration data for FOO"

DOI: 10.9999/US/FOOCAL

FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.001

Shortname: FOOL1B

Longname: "Level 1B, monthly calibrated radiances from FOO"

Collection: 1

DOI: 10.9999/US/FOOL1B.v1

FOOL1B.v1.01.974b98ed-5d3e-4296-a80d-002fecc764e7

Used: FOOL0.01.fa9cf2e0-b60b-43b7-baee-d18cc185b407

FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.02.a24dce74-f83f-4b31-973f-db4e4334e3de

Used: FOOL0.02.8a29012d-be8d-4af1-a158-783cfdcfc7fc

FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.03.802e3d7a-86e5-4e11-b84d-58d85626cbd6
Used: FOOL0.03.90463be2-a1a4-4d63-ad70-2f1f3c09798e,
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.04.5aa1de0c-688e-413d-8ec2-20ee083a9d58
Used: FOOL0.04.c121f001-6851-433f-9400-8e3acaa0229a
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.05.c70c2dd6-6c2a-495a-976c-cca87361948a
Used: FOOL0.05.2558aa9a-ce8c-47df-ac05-0fc982568462
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.06.9d75dc70-27d9-4456-a6e7-a4f5754fef77
Used: FOOL0.06.7223358c-f92d-42d0-bff4-1cd6140d4a89
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.07.70aaf233-087a-426b-8cb3-34d80328b2f1
Used: FOOL0.07.e2e145cd-899e-483f-989a-fbf1017a3df8
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.08.29a11b08-f55b-4bfc-ae0b-10b21eefc20b
Used: FOOL0.08.d0e94ea8-08b3-4bdf-af3f-7af72f2f9221
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.09.6260b254-de84-42bd-a73a-07b08a84dfbf
Used: FOOL0.09.ddc94379-0086-4817-9595-5fbe378c5a29
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.10.4b93b41a-f812-47fa-aab5-2db8bcd4282
Used: FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7 (old, corrupt file)
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.11.bdc6dd75-d76a-42c7-a29d-5f88bfd99eb8
Used: FOOL0.11.27d94c01-51ed-45c4-8c50-e19ca7f20882
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL1B.v1.12.4840f38e-81c2-4440-bd7b-8d846e88debe
Used: FOOL0.12.0af9c3a5-6ee5-4435-8437-d96ebfa36625
FOOCAL.1.85f7c7ca-c67d-446a-bd24-f24c8451a57c

FOOL2.001

Shortname: FOOL2

Longname: "Level 2 data, monthly geophysical parameter from FOO"

Collection: 1

DOI: 10.9999/US/FOOL2.v1

FOOL2.v1.01.b9149c24-9356-4fcd-baa0-50cee9e17ca2

Used: FOOL1B.v1.01.974b98ed-5d3e-4296-a80d-002fecc764e7

FOOL2.v1.02.8356d66b-167e-4410-81b3-e05d89052ea7

Used: FOOL1B.v1.02.a24dce74-f83f-4b31-973f-db4e4334e3de

FOOL2.v1.03.e7293f65-5aa1-476b-8e5b-ebe04c4a552c

Used: FOOL1B.v1.03.802e3d7a-86e5-4e11-b84d-58d85626cbd6

FOOL2.v1.04.84024c84-a563-429e-bc5a-00967565f2ea

Used: FOOL1B.v1.04.5aa1de0c-688e-413d-8ec2-20ee083a9d58

FOOL2.v1.05.c8501462-1690-431e-8477-a452792dadce

Used: FOOL1B.v1.05.c70c2dd6-6c2a-495a-976c-cca87361948a

FOOL2.v1.06.c92b1dff-f3f6-4c6a-ad84-9f8ce0b55a1b

Used: FOOL1B.v1.06.9d75dc70-27d9-4456-a6e7-a4f5754fef77

FOOL2.v1.07.51a3bc66-fbc9-4393-8ee1-b257d59aea17

Used: FOOL1B.v1.07.70aaf233-087a-426b-8cb3-34d80328b2f1

FOOL2.v1.08.c154f53c-3d7b-4a19-8384-7f0e11192651

Used: FOOL1B.v1.08.29a11b08-f55b-4bfc-ae0b-10b21eefc20b

FOOL2.v1.09.f7843584-1a32-46e5-9cb0-404aa2ed903d

Used: FOOL1B.v1.09.6260b254-de84-42bd-a73a-07b08a84dfbf

FOOL2.v1.10.dc298ee5-c88b-471a-8d7f-390ea0eaff23

Used: FOOL1B.v1.10.4b93b41a-f812-47fa-aab5-2db8bcd4282

FOOL2.v1.11.b8782b3f-d228-4029-87c9-4d2ae3104d4f

Used: FOOL1B.v1.11.bdc6dd75-d76a-42c7-a29d-5f88bfd99eb8

FOOL2.v1.12.44032e38-5248-4b3b-917c-c1e088f43578

Used: FOOL1B.v1.12.4840f38e-81c2-4440-bd7b-8d846e88debe

FOOL3.001

Shortname: FOOL3

Longname: "Annual gridded data from FOO"

Collection: 1

DOI: 10.9999/US/FOOL3.v1

FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4

Used:

FOOL2.v1.01.b9149c24-9356-4fcd-baa0-50cee9e17ca2

FOOL2.v1.02.8356d66b-167e-4410-81b3-e05d89052ea7

FOOL2.v1.03.e7293f65-5aa1-476b-8e5b-ebe04c4a552c

FOOL2.v1.04.84024c84-a563-429e-bc5a-00967565f2ea

FOOL2.v1.05.c8501462-1690-431e-8477-a452792dadce

FOOL2.v1.06.c92b1dff-f3f6-4c6a-ad84-9f8ce0b55a1b

FOOL2.v1.07.51a3bc66-fbc9-4393-8ee1-b257d59aea17

FOOL2.v1.08.c154f53c-3d7b-4a19-8384-7f0e11192651

FOOL2.v1.09.f7843584-1a32-46e5-9cb0-404aa2ed903d

FOOL2.v1.10.dc298ee5-c88b-471a-8d7f-390ea0eaff23

FOOL2.v1.11.b8782b3f-d228-4029-87c9-4d2ae3104d4f

FOOL2.v1.12.44032e38-5248-4b3b-917c-c1e088f43578

FOOL1B.002

Shortname: FOOL1B

Longname: "Level 1B, monthly calibrated radiances from FOO"

Collection: 2

DOI: 10.9999/US/FOOL1B.v2

FOOL1B.v2.01.d615b4f6-5e35-49f0-834a-ee199db7597c

Used: FOOL0.01.fa9cf2e0-b60b-43b7-baee-d18cc185b407

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.02.6c1a5a3b-55ab-4b53-8659-982d591cc744

Used: FOOL0.02.8a29012d-be8d-4af1-a158-783cfdcf7fc

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.03.58575454-3a4e-46af-8cb5-27c6aa4321cc

Used: FOOL0.03.90463be2-a1a4-4d63-ad70-2f1f3c09798e,

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.04.09124f68-3f14-446b-a1aa-d7f57e7f1603

Used: FOOL0.04.c121f001-6851-433f-9400-8e3acaa0229a

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.05.6091c13e-5ea4-44c3-92cc-f20823249421
Used: FOOL0.05.2558aa9a-ce8c-47df-ac05-0fc982568462
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.06.fcf1bb4c-51ea-464e-9935-b7653354ef73
Used: FOOL0.06.7223358c-f92d-42d0-bff4-1cd6140d4a89
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.07.424aa11d-fdfb-4a63-9b1b-0a386d23b1fa
Used: FOOL0.07.e2e145cd-899e-483f-989a-fbf1017a3df8
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.08.e3660e2e-4248-43ea-b91c-64799f3a1e74
Used: FOOL0.08.d0e94ea8-08b3-4bdf-af3f-7af72f2f9221
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.09.9ca12548-2a6b-47e4-9f7d-732013984ec1
Used: FOOL0.09.ddc94379-0086-4817-9595-5fbe378c5a29
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.10.b8f1e287-2d92-4340-9a71-5da5af191f3b (new file)
Used: FOOL0.10.3adf6dea-06af-478f-8216-2bbcd0caad (new file)
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.11.6cfec73e-bf2e-42cb-a427-2d30694f43e8
Used: FOOL0.11.27d94c01-51ed-45c4-8c50-e19ca7f20882
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.12.5cb6c2d9-386c-4d87-a103-56f0e459a26f
Used: FOOL0.12.0af9c3a5-6ee5-4435-8437-d96ebfa36625
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.13.e36a5d4d-1687-4947-bca8-2312c5f6eb8f
Used: FOOL0.13.76e9b680-2d06-4a55-ad2c-79b533ce86ca
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL1B.v2.14.eefa4113-200c-4a97-b5c6-3bd4b695cbc5
Used: FOOL0.14.f6bf9378-4215-41b5-9d3e-96dc0c2e7eeb
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL2.002

Shortname: FOOL2

Longname: "Level 2 data, monthly geophysical parameter from FOO"

Collection: 2

DOI: 10.9999/US/FOOL2.v2

FOOL2.v2.01.bba34792-f256-4c54-81dd-9977e432c204

Used: FOOL1B.v2.01.d615b4f6-5e35-49f0-834a-ee199db7597c

FOOL2.v2.02.2fd12da6-a3e2-4e50-8140-3ac645882419

Used: FOOL1B.v2.02.6c1a5a3b-55ab-4b53-8659-982d591cc744

FOOL2.v2.03.29bda893-765d-476d-851b-8b9acd7f140e

Used: FOOL1B.v2.03.58575454-3a4e-46af-8cb5-27c6aa4321cc

FOOL2.v2.04.57509ddb-3d40-4d60-8204-da4b99867fc7

Used: FOOL1B.v2.04.09124f68-3f14-446b-a1aa-d7f57e7f1603

FOOL2.v2.05.0e8604fa-fb4e-4cfb-b412-5364ca12cf14

Used: FOOL1B.v2.05.6091c13e-5ea4-44c3-92cc-f20823249421

FOOL2.v2.06.0eb26b4e-b718-41c5-bbf8-c83d3d79c233

Used: FOOL1B.v2.06.fcf1bb4c-51ea-464e-9935-b7653354ef73

FOOL2.v2.07.43079ea6-43b5-4622-b492-bcdb824a818e

Used: FOOL1B.v2.07.424aa11d-fdfb-4a63-9b1b-0a386d23b1fa

FOOL2.v2.08.590fd64c-ec12-44a5-9b14-0042d19ed3dc

Used: FOOL1B.v2.08.e3660e2e-4248-43ea-b91c-64799f3a1e74

FOOL2.v2.09.226173b9-4ef7-49e8-8b9e-701b892a8f57

Used: FOOL1B.v2.09.9ca12548-2a6b-47e4-9f7d-732013984ec1

FOOL2.v2.10.6e58a410-60e7-4956-aeaf-37f76a16b171 (new file)

Used: FOOL1B.v2.10.b8f1e287-2d92-4340-9a71-5da5af191f3b

FOOL2.v2.11.af235d11-777c-4bf1-a5e6-15273a5e5d80

Used: FOOL1B.v2.11.6cfec73e-bf2e-42cb-a427-2d30694f43e8

FOOL2.v2.12.bdc9dc33-38bd-403c-991e-48dcd4762ca7

Used: FOOL1B.v2.12.5cb6c2d9-386c-4d87-a103-56f0e459a26f

FOOL2.v2.13.f8f9564d-cc2a-4760-b1bc-13f1ef5cbdcb

Used: FOOL1B.v2.13.e36a5d4d-1687-4947-bca8-2312c5f6eb8f

FOOL2.v2.14.4814ed46-0e41-4e3f-8f73-33d0cd2ef0bc
Used: FOOL1B.v2.14.eefa4113-200c-4a97-b5c6-3bd4b695cbc5

FOOL3.002
Shortname: FOOL3
Longname: "Annual gridded data from FOO"
Collection: 2
DOI: 10.9999/US/FOOL3.v2

FOOL3.v2.01.2a365058-fb52-4559-ab4b-085cb5ac0b73
Used:
FOOL2.v2.01.bba34792-f256-4c54-81dd-9977e432c204
FOOL2.v2.02.2fd12da6-a3e2-4e50-8140-3ac645882419
FOOL2.v2.03.29bda893-765d-476d-851b-8b9acd7f140e
FOOL2.v2.04.57509ddb-3d40-4d60-8204-da4b99867fc7
FOOL2.v2.05.0e8604fa-fb4e-4cfb-b412-5364ca12cf14
FOOL2.v2.06.0eb26b4e-b718-41c5-bbf8-c83d3d79c233
FOOL2.v2.07.43079ea6-43b5-4622-b492-bcdb824a818e
FOOL2.v2.08.590fd64c-ec12-44a5-9b14-0042d19ed3dc
FOOL2.v2.09.226173b9-4ef7-49e8-8b9e-701b892a8f57
FOOL2.v2.10.6e58a410-60e7-4956-aeaf-37f76a16b171 (new file)
FOOL2.v2.11.af235d11-777c-4bf1-a5e6-15273a5e5d80
FOOL2.v2.12.bdc9dc33-38bd-403c-991e-48dcd4762ca7

orphaned:

FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7 (corrupt L0 data)

FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e (deleted)
Used: FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7 (old, corrupt file)
FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL2.v2.10.533b2a95-d57f-4f75-9b7d-914d3d220310 (deleted)
Used: FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e (old, deleted)

FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768 (deleted)
Used:
FOOL2.v2.01.bba34792-f256-4c54-81dd-9977e432c204
FOOL2.v2.02.2fd12da6-a3e2-4e50-8140-3ac645882419
FOOL2.v2.03.29bda893-765d-476d-851b-8b9acd7f140e

FOOL2.v2.04.57509ddb-3d40-4d60-8204-da4b99867fc7
FOOL2.v2.05.0e8604fa-fb4e-4cfb-b412-5364ca12cf14
FOOL2.v2.06.0eb26b4e-b718-41c5-bbf8-c83d3d79c233
FOOL2.v2.07.43079ea6-43b5-4622-b492-bcdb824a818e
FOOL2.v2.08.590fd64c-ec12-44a5-9b14-0042d19ed3dc
FOOL2.v2.09.226173b9-4ef7-49e8-8b9e-701b892a8f57
FOOL2.v2.10.533b2a95-d57f-4f75-9b7d-914d3d220310
FOOL2.v2.11.af235d11-777c-4bf1-a5e6-15273a5e5d80
FOOL2.v2.12.bdc9dc33-38bd-403c-991e-48dcd4762ca7

The archive has chosen to keep the Level 0 file, since published research was based on it. The bad L1B, L2, L3 data were deleted, but since we have complete provenance, we maintain the ability to recreate them.

Note that the one of the granules used directly in Alice's research is among the deleted data. You can't order it directly from the archive any more.

You could however ask the archive about "FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768" and it could tell you how it was made, and how those files were made and so on and you could reproduce an equivalent granule.

Results Replication

Bob has been reading Alice's paper, and notices a weird bump in the graphs around month 10 (due to the corrupt data though he doesn't know that).

He tries to reproduce her process, following the methodology she documented with her paper.

Following the cited data DOIs referenced in her paper, he downloads
FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4 and
FOOL3.v2.01.2a365058-fb52-4559-ab4b-085cb5ac0b73

and performs the analysis. Unsurprisingly, since he is using the revised data, he gets different answers from Alice.

He gives her a call, and they work out that she actually used

FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4
FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768

Ok, says Bob, I'll just go back to the archive and get those files. They aren't available, but he does see that since the metadata and provenance were preserved, he can get the software and reproduce those granules.

He gets the old L0 file and calibration file and produces some new files:

FOOL1B.v2.10.c911b994-91fb-4d5c-b9e1-642c0a9c46a3

Used: FOOL0.10.0b337185-82af-4662-89b0-419bfd3e5db7 (old, corrupt file)

FOOCAL.2.d2a14052-f426-4d2e-a506-ec052fdb69d4

FOOL2.v2.10.2c09ed89-57cf-40ed-910b-16c1aafcd947

Used: FOOL1B.v2.10.c911b994-91fb-4d5c-b9e1-642c0a9c46a3

FOOL3.v2.01.52562fbd-5969-4572-a757-47ff3f92dda4

Used:

FOOL2.v2.01.bba34792-f256-4c54-81dd-9977e432c204

FOOL2.v2.02.2fd12da6-a3e2-4e50-8140-3ac645882419

FOOL2.v2.03.29bda893-765d-476d-851b-8b9acd7f140e

FOOL2.v2.04.57509ddb-3d40-4d60-8204-da4b99867fc7

FOOL2.v2.05.0e8604fa-fb4e-4cfb-b412-5364ca12cf14

FOOL2.v2.06.0eb26b4e-b718-41c5-bbf8-c83d3d79c233

FOOL2.v2.07.43079ea6-43b5-4622-b492-bcdb824a818e

FOOL2.v2.08.590fd64c-ec12-44a5-9b14-0042d19ed3dc

FOOL2.v2.09.226173b9-4ef7-49e8-8b9e-701b892a8f57

FOOL2.v2.10.2c09ed89-57cf-40ed-910b-16c1aafcd947

FOOL2.v2.11.af235d11-777c-4bf1-a5e6-15273a5e5d80

FOOL2.v2.12.bdc9dc33-38bd-403c-991e-48dcd4762ca7

(For now, let's assume he can perfectly reproduce the environment and run the processing in an identical manner.)

Note that each of the files Bob produces has a different UUID, and different provenance.

If, for example, you asked about

FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e

the provenance would say that the main data processing system produced that file on a certain date/time on a certain host, while the file Bob made:

FOOL1B.v2.10.c911b994-91fb-4d5c-b9e1-642c0a9c46a3

was made by Bob, at an entirely different time on a different host.

The key thing though is that they each used the same software and the same input files and were run in the same way.

In this world, let's accept that

FOOL1B.v2.10.2f269e5e-cce7-41e4-8a83-baad1e087c8e
is scientifically equivalent to
FOOL1B.v2.10.c911b994-91fb-4d5c-b9e1-642c0a9c46a3

Similarly:

FOOL2.v2.10.2c09ed89-57cf-40ed-910b-16c1aafcd947
is scientifically equivalent to
FOOL2.v2.10.533b2a95-d57f-4f75-9b7d-914d3d220310
and
FOOL3.v2.01.52562fbd-5969-4572-a757-47ff3f92dda4
is scientifically equivalent to
FOOL3.v2.01.07aa9ae3-9c3e-4508-b027-890dae11b768

So, Bob can now use these two granules:

FOOL3.v1.01.d08925b3-3eb3-407d-8db1-f5e0d101a0a4
FOOL3.v2.01.52562fbd-5969-4572-a757-47ff3f92dda4

and follow Alice's methodology and reproduce the research in her paper.

Summary

So far, this simple scenario has:

1. Shown how DOI works well to identify and locate "ESDT+Collection".
2. Show how DOI doesn't precisely identify sets of granules.
3. Show how UUID can be used to unambiguously refer to individual granules.