

NDACC & IRWG Archive Changes & Licences & DOI

- Talk Ian & Bavo on HDF changes
- Data_Source attribute
 - How do we IRWG use this attribute
- Creative Commons licensing
- DOI
- Who are our data users? New answer: ECMWF

NDACC & IRWG Archive Changes & Licenses & DOI

NDACC & IRWG Archive Changes & Licenses & DOI

“Because of the discussions that we had in the NDACC SC meeting last year and that we will pursue further during this year's SC meeting, it may be good that you also have a discussion about **DOI** and especially about **data licensing** with the IRWG”

It would be really useful to get the opinion from the IRWG about their choice...

‘Grouping’ of the data for a DOI”

1. For the DOI, my personal preference would be to have it per species/per site/per processing version
2. One DOI per yearly snapshot of entire NDACC archive
3. But there may be different options ?

For the data license, we proposed at the NDACC SC meeting to have only a few options

PI would have choice between:

- (1) CC0 ->(mandatory choice for US Gov't data – much like how satellite data is available and used)
- (2) CC-BY-SA-4.0
- (3) CC-BY-NC-SA

Can the IRWG choose only one of the three or do we need all three ?

Can we all agree with these three options ?

<https://creativecommons.org/share-your-work/licensing-types-examples/>

NDACC & IRWG Archive Changes & Licenses & DOI

We would like to be in a position to make a decision at the next SC (Oct 2019)

It would be really useful to get the opinion from the IRWG about their choice...

For the data license, we proposed at the NDACC SC meeting to have only a few options

PI would have choice between:

1. CC0
 - a. mandatory choice for US Gov't data
 - b. much like how satellite data is available and used)
2. CC-BY-SA-4.0
3. CC-BY-NC-SA

Can the IRWG choose only one of the three or do we need all three ?





Can we all agree with these three options ?

<https://creativecommons.org/share-your-work/licensing-types-examples/>

NDACC & IRWG Archive Changes & Licenses & DOI

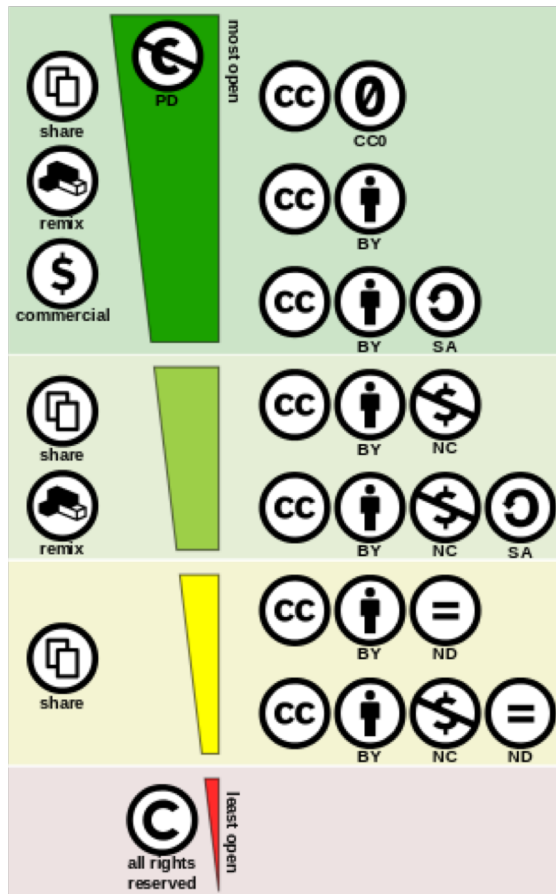
Creative Commons (CC) licence

- **Creative Commons** is a non-profit organisation proposing legal and flexible solutions to give users the right to access, share, and use content and data.
- CC licences is one of several public **open** licences placing few restrictions (with the 'copyright' being the most restricting).
- Several types of CC licences exist, the different CC licences condition the terms of distribution and are based on four conditions:

Icon	Licence Right	Description
	BY	Attribution: allows to copy, distribute, <i>make</i> derivative work, remix work if credits are given to the author or licensor (in manner specified)
	SA	Share-alike: allow to <i>distribute</i> derivate work under an identical licence ("not more restrictive").
	NC	Non-commercial: allows to copy, distribute, make derivative work, remix only for non-commercial purposes.
	ND	No derivative works: allows to copy, distribute, and display but not to make derivate work or remix work.

NDACC & IRWG Archive Changes & Licenses & DOI

CC licence spectrum



most open

- Commonly used are 7 CC licences (combination of conditions)
- Most open: **CC0** (public domain waiver)
- Least open: „all rights reserved“ (copyright owner hold all rights for own use)
 - **CC-BY-NC-ND**: share only, credits required, non commercial use, no modification
 - **CC BY-NC**: credits required, allows to share and remix, non-commercial only
 - **CC BY-SA**: credits required, allows to share and remix, same legal conditions
 - **CC BY 4.0** (Attribution 4.0 International): latest version of CC licences, applicable to most jurisdictions: allows to
 - to copy and redistribute the material in any medium or format,
 - Remix, transform, and build upon the material for any purpose, even commercially.

least open

NDACC & IRWG Archive Changes & Licenses & DOI

- A paradigm shift in that large data ingesters may be the largest 'data users' – not our nearest neighbor science researchers
 - *They wish to take, manipulate, assimilate, repackage, post under another brand, sell...*
- The more data users who are dependent on your data the better...
- The loop between funders recognizing value (they will pay for) being a large ingester (rather than e.g. publications) is not closed
- We did not previous to this decide to be data producers alone
- Most groups are already stressed
 - Limited funding, limited personnel, limited science production, limited ability for improvements
 - Without an operational commitment (certainly an unfunded / non-recognized, or appreciated one)
- IRWG needs to articulate this back to the SC (and others.)

NDACC & IRWG Archive Changes & Licenses & DOI

DATA_SOURCE

Reserved Keywords	Purpose
	No keyword: Describes a dataset obtained from a normal/default analysis
vNNN	Describes a dataset obtained from an algorithm defined only by its version number <i>NNN</i>
DEFAULT	Describes a dataset obtained from a normal/default analysis, but some alternate version exists
HOURLY	Describes a dataset with a time granularity of 1 hour
DAILY	Describes a dataset with a time granularity of 24 hours
WEEKLY	Describes a dataset with a time granularity of 1 week
MONTHLY	Describes a dataset with a time granularity of 1 month
YEARLY	Describes a dataset with a time granularity of 1 year
SUNRISE	Describes a dataset obtained from measurements made at sunrise
SUNSET	Describes a dataset obtained from measurements made at sunset
SUN	Describes a dataset obtained from measurements tracking the sun
MOON	Describes a dataset obtained from measurements tracking the moon
LIMB	Describes a dataset obtained from measurements made in a limb geometry
NADIR	Describes a dataset obtained from measurements made in a nadir geometry
ZENITH	Describes a dataset obtained from measurements made in a zenith geometry
CALVAL	Describes a dataset with a spatio-temporal granularity tailored to calibration/validation
CLIM	Describes a dataset with a spatio-temporal granularity tailored to climatology and trends
HIRES	Describes a dataset with the highest possible spatial and/or temporal resolution
LOWRES	Describes a dataset whose spatial and/or temporal resolution has been voluntarily degraded
GRIDDED	Describes a dataset gridded or re-gridded onto a standardized grid
HOMOGEN	Describes a re-processed dataset as a result of an homogenization effort
INHOUSE	Describes in-house-processed data
CENTRLZD	Describes centrally-processed data
CTRL	Describes a dataset produced from a "control run"

Will have a 'fixed' set of keywords but adopted at GEOMS level. But we define by working Groups.

Description of variant in IRWG metadata file at archive.

RT data may be a variant

We plan to not have many variants!

For instance:

Next Standard = HIT16.01

NDACC & IRWG Archive Changes & Licenses & DOI

DATA_SOURCE

DATA_SOURCE = ftir.ch4_ncar001

DATA_FILE_VERSION = 001

New DATA_SOURCE = ftir.ch4_ncar001_HIT16.01 <- now contains variant tag

New full data version in this file: DATA_SOURCE. DATA_FILE_VERSION :

ftir.ch4_ncar001_HIT16.01. 001

DATA_SOURCE *composed of reserved names (at GEOMS) but defined by WG*

DATA_FILE_VERSION *reserved by PI for minor updates / error corrections*

NDACC & IRWG Archive Changes & Licenses & DOI

NDACC & IRWG Archive Changes & Licenses & DOI