



OPEN DATA & OPEN KNOWLEDGE Workshop

Encoding heterogeneous in-situ data

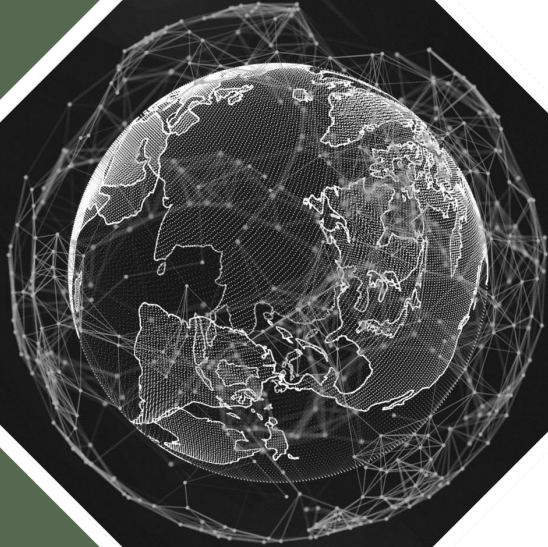
Raphaël Jolivet

Research Engineer



O.I.E (Observation, Impacts, Energy) @ Sophia Antipolis

- Assessment of **solar resources** from observation of earth
- Life cycle assessment of **environmental impacts** of energy systems



Common challenges for in-situ data

- **Various providers** of data
- Heterogeneous data : **format, protocol, timezone, units, ...**
- Lack of embedded **meta-data**
- Various policies (licenses)

Main motivation : what you get

Index of /aftp/data/radiation/solrad/abq/2002/

Name	Last modified
Parent Directory	
abq02032.dat	2015-02-05 15:41
abq02033.dat	2015-02-05 15:47
abq02034.dat	2015-02-05 15:47
abq02035.dat	2015-02-05 15:47
abq02036.dat	2015-02-05 15:47
abq02037.dat	2015-02-05 15:41
abq02038.dat	2015-02-05 15:41
abq02039.dat	2015-02-05 15:41
abq02040.dat	2015-02-05 15:41
abq02041.dat	2015-02-05 15:41
abq02042.dat	2015-02-05 15:41
abq02043.dat	2015-02-05 15:41
abq02044.dat	2015-02-05 15:41
abq02045.dat	2015-02-05 15:41
abq02046.dat	2015-02-05 15:41

Albuquerque																					
35.03796	-106.62211	1617	-7	version	1																
2023	2	1	2	0	0	0.000	89.21	12.4	0	0.8	0	15.2	0	0.2	0	43.0	0	0.504	0.000	0.526	0.032
2023	2	1	2	0	1	0.017	89.36	10.5	0	0.7	0	13.5	0	0.2	0	43.0	0	0.378	0.251	0.421	0.000
2023	2	1	2	0	2	0.033	89.50	9.1	0	0.1	0	12.1	0	0.2	0	43.0	0	0.504	0.126	0.421	0.000
2023	2	1	2	0	3	0.050	89.64	7.6	0	0.0	0	10.6	0	0.2	0	43.0	0	0.504	0.000	0.421	0.015
2023	2	1	2	0	4	0.067	89.77	6.3	0	0.0	0	9.1	0	0.1	0	43.0	0	0.378	0.000	0.526	0.024
2023	2	1	2	0	5	0.083	89.90	4.9	0	0.0	0	7.4	0	0.1	0	43.0	0	0.504	0.000	0.421	0.009
2023	2	1	2	0	6	0.100	90.00	3.3	0	0.0	0	5.9	0	0.0	0	43.0	0	0.378	0.000	0.421	0.000
2023	2	1	2	0	7	0.117	90.98	2.0	0	0.0	0	4.8	0	0.0	0	43.0	0	0.504	0.000	0.210	0.000
2023	2	1	2	0	8	0.133	91.16	0.8	0	0.0	0	3.0	0	0.0	0	43.0	0	0.378	0.000	0.316	0.000
2023	2	1	2	0	9	0.150	91.34	0.0	0	-0.1	0	2.9	0	0.0	0	43.0	0	0.378	0.126	0.316	0.000
2023	2	1	2	0	10	0.167	91.53	-0.4	0	-0.6	0	2.4	0	0.0	0	43.0	0	0.000	0.251	0.105	0.000
2023	2	1	2	0	11	0.183	91.71	-1.0	0	-0.8	0	1.9	0	0.0	0	43.0	0	0.378	0.000	0.210	0.000
2023	2	1	2	0	12	0.200	91.89	-1.3	0	-0.8	0	1.8	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	13	0.217	92.07	-1.3	0	-0.8	0	1.6	0	0.0	0	43.0	0	0.000	0.000	0.210	0.000
2023	2	1	2	0	14	0.233	92.26	-1.3	0	-0.4	0	1.3	0	0.0	0	43.0	0	0.000	0.377	0.210	0.000
2023	2	1	2	0	15	0.250	92.44	-1.3	0	0.0	0	1.1	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	16	0.267	92.62	-1.3	0	0.0	0	1.1	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	17	0.283	92.81	-1.3	0	0.0	0	1.1	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	18	0.300	92.99	-1.3	0	0.0	0	0.8	0	0.0	0	43.0	0	0.000	0.000	0.105	0.000
2023	2	1	2	0	19	0.317	93.17	-1.3	0	0.0	0	0.7	0	0.0	0	43.0	0	0.000	0.126	0.105	0.000
2023	2	1	2	0	20	0.333	93.36	-1.3	0	0.6	0	0.4	0	0.0	0	43.0	0	0.000	0.377	0.105	0.000
2023	2	1	2	0	21	0.350	93.54	-1.3	0	0.8	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	22	0.367	93.72	-1.3	0	0.8	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	23	0.383	93.91	-1.3	0	0.8	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	24	0.400	94.09	-1.3	0	0.8	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	25	0.417	94.28	-1.3	0	0.2	0	0.4	0	0.0	0	43.0	0	0.000	0.251	0.000	0.000
2023	2	1	2	0	26	0.433	94.46	-1.4	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	27	0.450	94.65	-1.8	0	0.0	0	0.4	0	0.0	0	43.0	0	0.126	0.000	0.000	0.000
2023	2	1	2	0	28	0.467	94.84	-1.8	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	29	0.483	95.02	-1.6	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	30	0.500	95.21	-1.4	0	0.0	0	0.4	0	0.0	0	43.0	0	0.126	0.000	0.000	0.000
2023	2	1	2	0	31	0.517	95.40	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	32	0.533	95.58	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	33	0.550	95.77	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.126	0.000	0.000	0.000
2023	2	1	2	0	34	0.567	95.95	-1.4	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	35	0.583	96.14	-1.6	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	36	0.600	96.33	-1.7	0	0.0	0	0.4	0	0.0	0	43.0	0	0.126	0.000	0.000	0.000
2023	2	1	2	0	37	0.617	96.52	-1.6	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	38	0.633	96.70	-1.4	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	39	0.650	96.89	-1.4	0	0.0	0	0.4	0	0.0	0	43.0	0	0.252	0.000	0.000	0.000
2023	2	1	2	0	40	0.667	97.08	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.126	0.000	0.000	0.000
2023	2	1	2	0	41	0.683	97.27	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	42	0.700	97.46	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.000	0.000	0.000	0.000
2023	2	1	2	0	43	0.717	97.64	-1.3	0	0.0	0	0.4	0	0.0	0	43.0	0	0.000	0.126	0.000	0.000

- Mean of Access: **FTP**, HTTP

- Format: *.dat, csv, ..

Main motivation

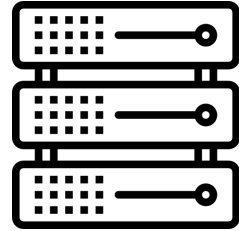
	Input : heterogenous
Format	X Mixed (csv, zip, dat, ...)
Granularity	X Mixed (daily, monthly, ...)
Access	X Mixed (http, ftp, ...)
Timezone	X Mixed (local , utc)
Units	X Mixed, implicit
Metadata	X Missing
Standard	X None
Compression	X Optional / external (zip)

- A **methodology** following **FAIR** principles
- A draft **standard** for **in-situ data**, based on **other standards**
- An open source **library** : *libinsitu*, implementing it
- **Deployed** “live” for our field of research (solar energy)

Practical data workflow



Raw data from data logger



Open standard encoding and access protocol



Agnostic access



Dissemination search & discovery



REUSABLE



“Metadata and data should be well-described”

NetCDF format with CF Conventions



ACCESSIBLE



“User needs to know how data and metadata can be accessed”

Thredds Data Server



INTER-OPERABLE



“Data need to interoperate with applications”

Open Source Web Clients



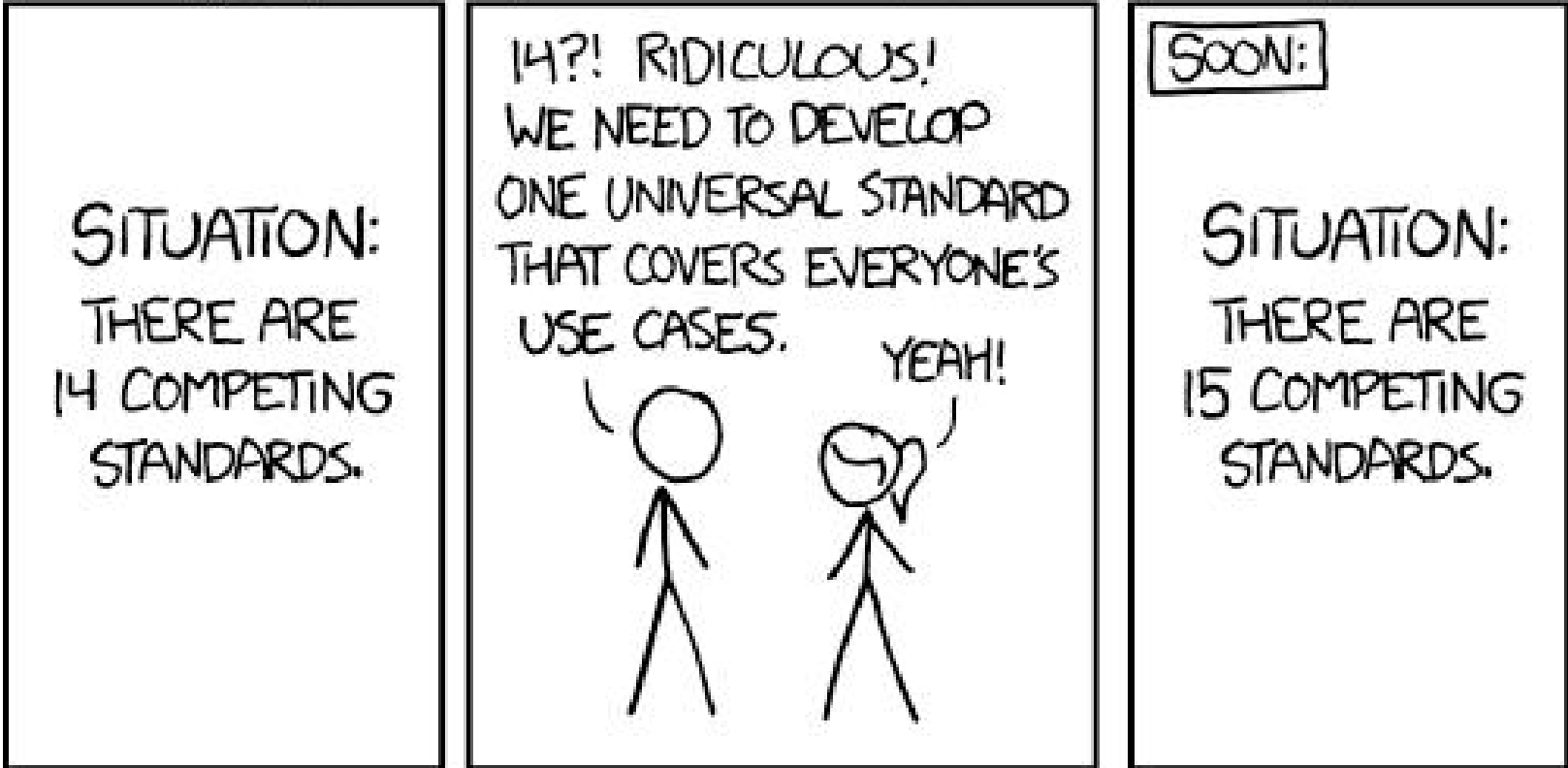
FINDABLE



“Metadata and data should be easy to find”

GEO Portal and GEO Knowledge Hub

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)



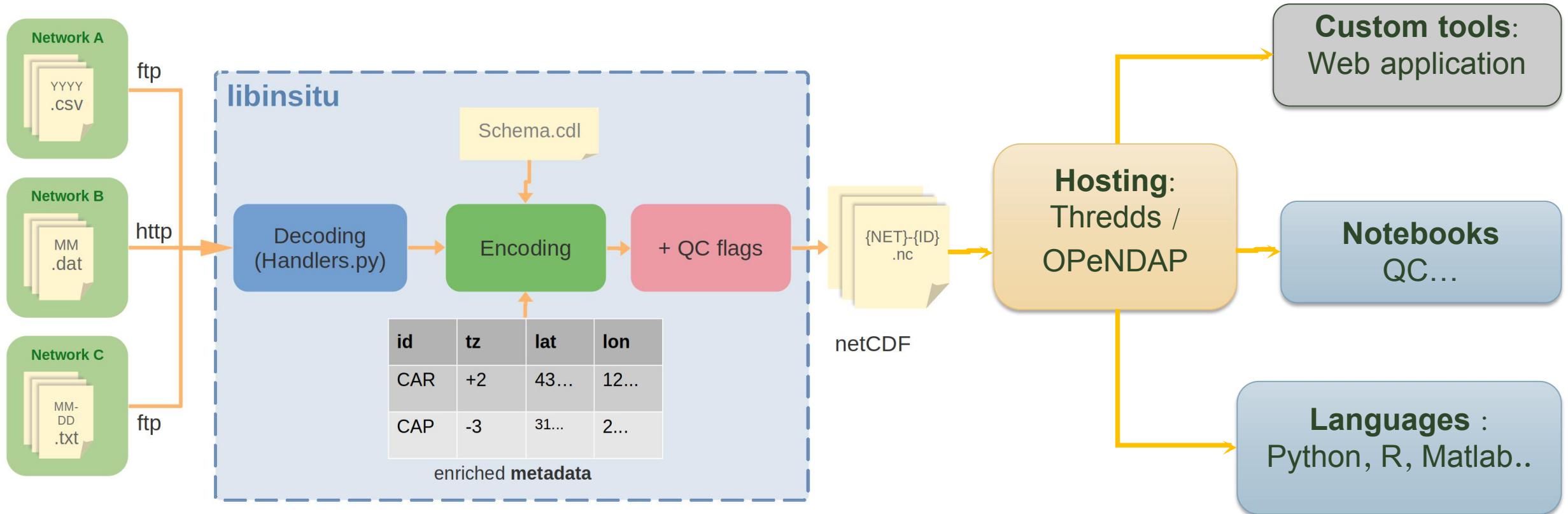
Proposed standard for in situ data
DRAFT

CF conventions
Semantic of metadata
Standard names for **physical** variables

NetCDF
Efficient binary storage of matrices
Embedded metadata
Remote access (OpenDAP)
Wide support

ACDD (ESIP)
Attribute Convention For Data Discovery
Global Attributes
[Findable]

Encoding pipeline



Solution

	Input : heterogenous	Output : standardized
Format	X Mixed (csv, zip, dat, ...)	V NetCDF
Granularity	X Mixed (daily, monthly, ...)	V Single file per station
Access	X Mixed (http, ftp, ...)	V TDS & OpenDAP (random subset)
Timezone	X Mixed (local , utc)	V UTC , explicit
Units	X Mixed, implicit	V SI , explicit
Metadata	X Missing	V Embedded
Standard	X None	V CF conventions
Compression	X Optional / external (zip)	V Native (~10% ratio)



Demo

Binder : <http://bit.ly/3X6e6lG>



Take aways

- **Libinsitu** :
Use it... **It's open, free, documented and exemplified**
- **Replicable** for any domain
- **Tested and implemented** by Marine Radioactivity Information System (**MARIS**) community
- **New breath** for in-situ measurement networks towards **standard and interoperable practices**

Documentation

<https://libinsitu.readthedocs.io>

Source code

<https://git.sophia.mines-paristech.fr/oie/libinsitu/>

Methodology

<https://doi.org/10.23646/AC2M-8504>

Standard

<https://libinsitu.readthedocs.io/en/latest/conventions.html>

Mailing list

solar-insitu@groupe.mines-paristech.fr

<https://groupe.minesparis.psl.eu/www/info/solar-insitu>

CONTACT DETAILS



EMAIL ADDRESS

raphael.jolivet@minesparis.psl.eu