

DMP & DOI for MASER service collections and associated projects

A Prototype for FAIR data publishing

Cecconi B. and MASER & PADC teams — 2023-05-10 (IVOA-Bologna)

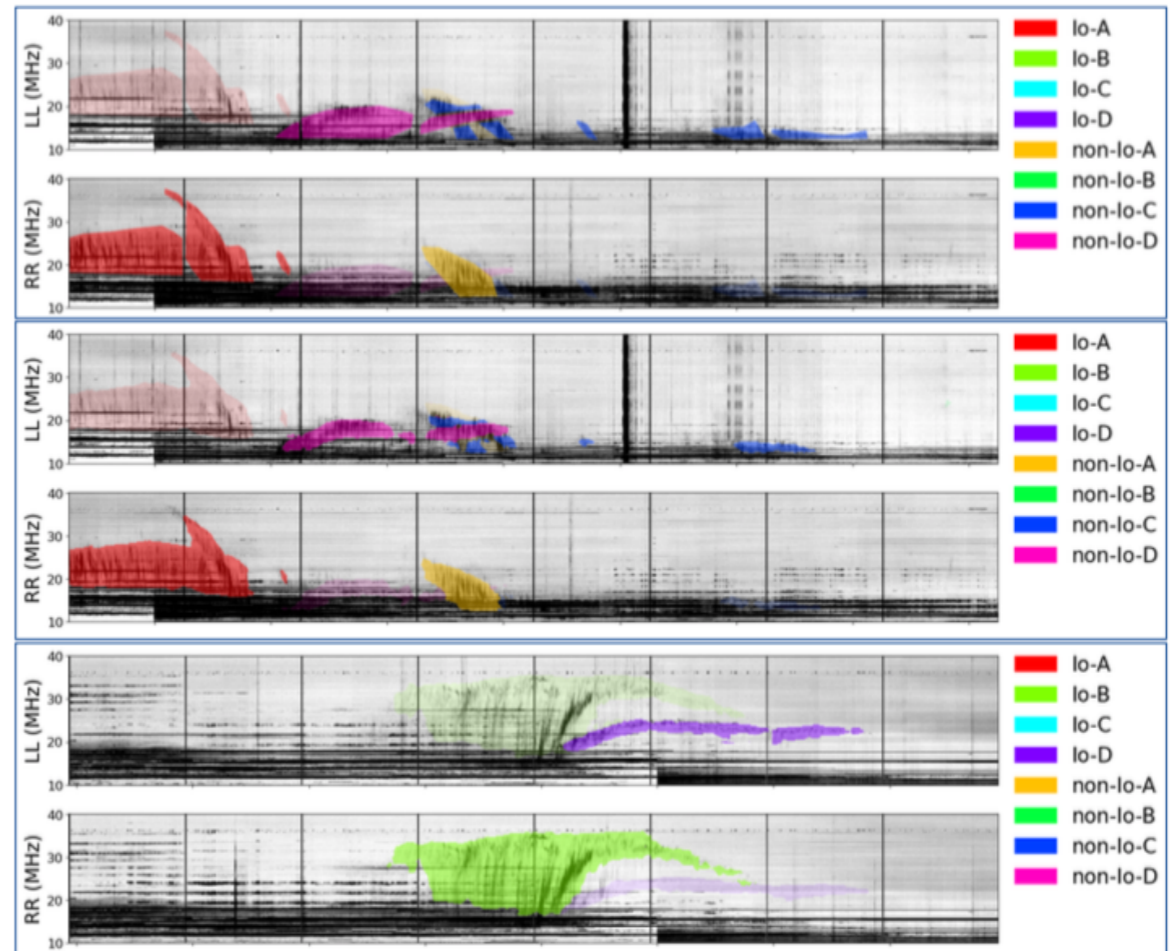
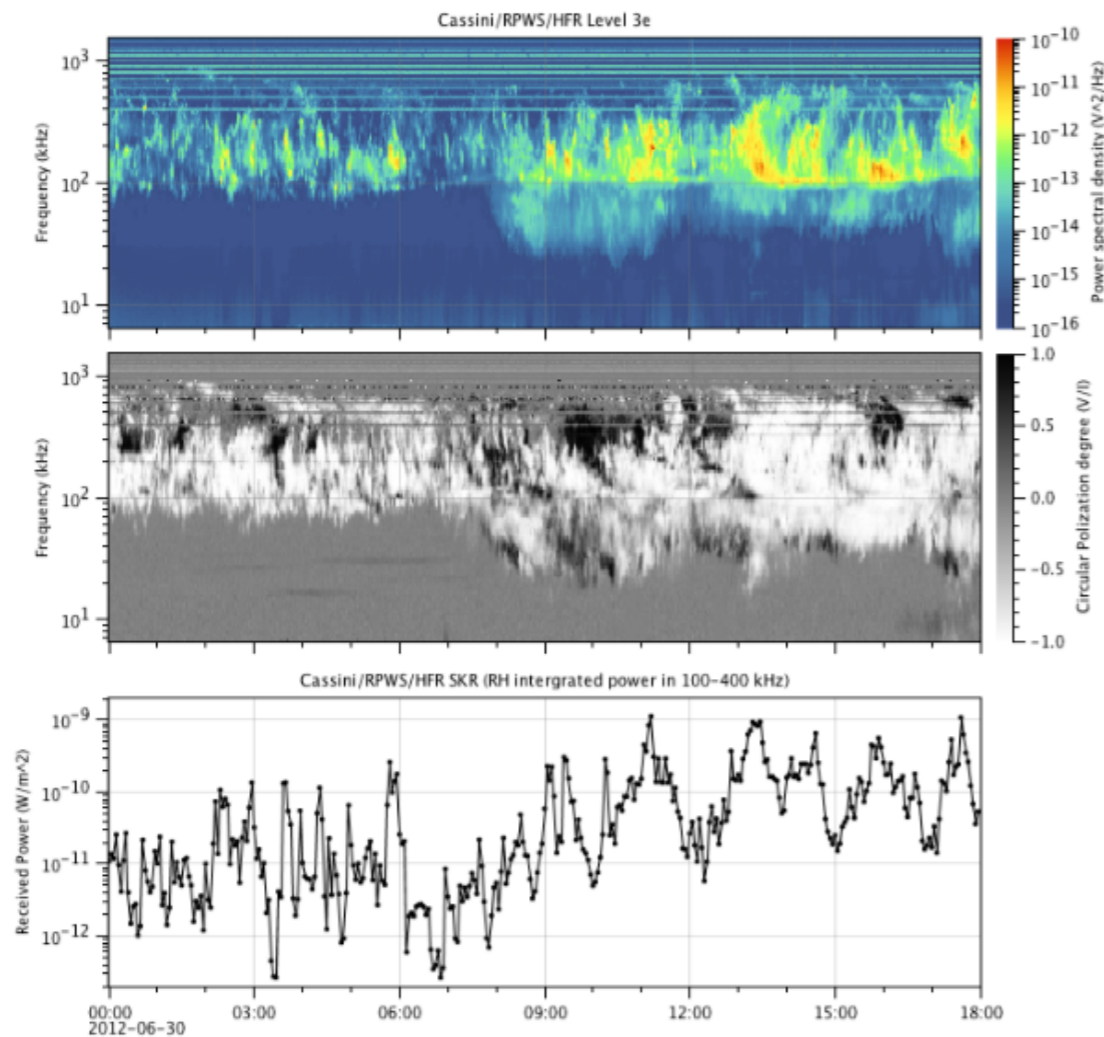
The Europlanet-2024 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.



MASER service

Measurement, Analyses, Simulations of Emissions in the Radio range

- *Science-ready and Open science* Toolbox for low frequency radio astronomy



MASER

Why and how

- Users needs:
 - **discovery** of datasets
 - online access for **visualisation**
 - python library for **programmatic access**
 - **annotation and sharing** of event/feature catalogues
 - **hosting** datasets
- VO Protocols / Standards:
 - data discovery: TAP (ObsTAP + EPNTAP)
 - data streaming: das2 (heliophysics community: IHDEA)
 - data files in FITS or CDF
- New formats:
 - catalogues: TFCat (Time-Frequency Catalogue) + FT-MOC (soon)
- Support to other projects (e.g.: Solar-Orbiter/RPW, Nançay/NDA)

MASER

Dataset typology

<https://maser.lesia.obspm.fr/publications/doi>

- **Data Collections** [22 collections]
(instrument + data processing level)
Data Types:
 - raw measurement (low level)
 - calibrated (intermediate)
 - derived (high level)
- **Collections** par instrument [3 collections]
(with links to collections per processing level)
- **Documentation** (e.g.: format specification) [14 documents]
- **Catalogues** (TFCat format) [12 collections]
- **Supplementary Material** (excluding catalogues) [4 collections]
- **Other** [3 datasets]

MASER

Architecture / data management

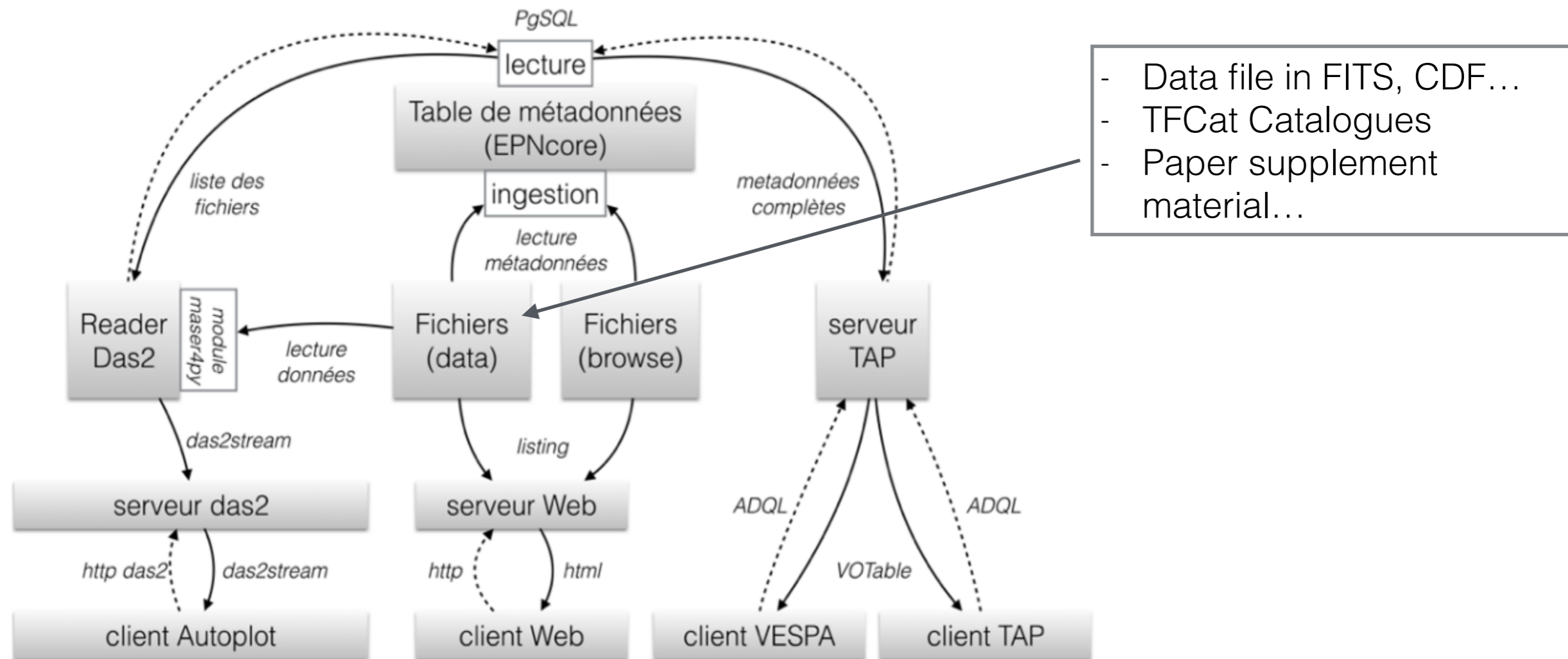


Figure 5. Ecosystème MASER : relation entre les différentes interfaces mises en place.

**Access
(stream)**

**Access
(web)**

Discovery

DMP + DOI (+landing page) per collection

MASER / DMP

Data Management Plans

- **Why?**
 - define the structure of collections
 - define interoperable interfaces
 - empowering teams
 - plan storage needs
- **What?**
 - describe collections and interfaces (EPN-TAP, das2)
 - responsibilities: scientific content maintenance, storage maintenance, VO interfaces maintenance
- Template created by *Open Science WG* of ObsParis, updated by the Europlanet/VESPA team.
 - ObsParis DMP Guide: <https://doi.org/10.25935/1mh3-nn37>
 - ObsParis DMP Template: <https://doi.org/10.25935/X859-TH79>
 - EPN-VESPA version: <https://voparis-wiki.obspm.fr/display/VES/Individual+DMP+of+EPN2024RI+collections>

MASER / DMP

Data Management Plans

- Data Collection Description
 - Data Collection Title
 - Persons/Entities
 - Keywords
 - Collection Details
- Making the Data FAIR
 - Findable
 - Accessible
 - Interoperable
 - Re-usable
- Responsibilities and resources
- Archiving and preservation
- Ethical aspects
- Other issues

*incl. ORCID & ROR
e.g., IVOA-UAT
collection + EPN-TAP + das2*

*URL/protocol + PID
URL/protocol
standards
licence (data, code)*

MASER / DMP

Data Management Plans

Data Collection Description

Data Collection Title

STEREO/Waves L3 DF Data Collection

Persons/Entities

Data collection persons

Name	Identifier URI (from ORCID, ROR or Re3Data)	Affiliation	Affiliation URI (from ROR or Re3Data)	Creator	Contribution Type
Cecconi, Baptiste	0000-0001-7915-5571	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	yes	Project Member
Bonnin, Xavier	0000-0003-4217-7333	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	yes	Developer
Nguyen, Quynh Nhu	0000-0003-1948-5426	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	yes	Developer
Maksimovic, Milan	0000-0001-6172-5062	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	yes	Project Leader
Krupar, Vratislav	0000-0001-6185-3945	NASA Goddard Space Flight Center, Greenbelt, MD, USA	https://ror.org/0171mag52	yes	Developer
Florence Henry	0000-0002-4016-8241	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	no	Developer

EPNcore table persons

Name	Identifier URI (from ORCID, ROR or Re3Data)	Affiliation	Affiliation URI (from ROR or Re3Data)	Contribution Type
Cecconi, Baptiste	0000-0001-7915-5571	LESIA, CNRS, Observatoire de Paris	https://ror.org/02eptjh02	Data Manager

Keywords

Keywords	Vocabulary
radio-astronomy	IVOA-UAT
space-plasmas	IVOA-UAT
solar-radio-emissions	IVOA-UAT

Collection Details

This section contains the main description of the data collection (see Column A). Column B contains a description of the additional metadata tables following the IVOA standards.

	A. Data Collection	B. IVOA metadata
Description	The STEREO/Waves L3 DF collections contain Direction-Finding (or <u>goniopolarimetric</u>) derived data from the LFR (Low Frequency Receiver) and HFR (High Frequency Receiver) receivers of the Waves experiments embarked <u>onboard</u> the STEREO-A and STEREO-B spacecraft. A dedicated collection is available for each receiver and spacecraft pairs.	The <u>EPNcore</u> table serves a <u>catalogue</u> of the products with <u>metadata</u> according to the <u>EPNcore</u> specification. The table is accessible through TAP.
Research Data Objects Types	Derived data files (L3)	<u>Relational database</u> table
Formats	CDF files	Compliant with EPN-TAP TAP results are served in <u>VOTable</u> format.
Storage	The data files are currently stored on a storage space managed by DIO in Meudon (<u>swaves</u> NFS share, on <u>filer-m2</u> server). A copy of the files will be made available at SPDF (NASA) and CDPP (CNES).	Database stored in server: http://voparis-tap-maser.obspm.fr No backup (but scripts for importing <u>metadata</u> are stored)
Sizing	LFR files: 3.4 MB per day. (CDF: 1.8 MB; PNG: 0.3 MB; SAV: 1.3 MB) HFR files: 36 MB per day. (CDF: 15 MB; PNG: 1.0MB; SAV: 20 MB)	16495 rows (at the time of writing)
Product Types	<u>dynamic spectra</u>	Same as data collection
Methodology	The data processing is described in https://doi.org/10.1029/2011JA017333	The source code of the import script is managed with <u>git</u> and available here: https://voparis-gitlab.obspm.fr/vespa/dachs/services/padc/voparis-tap-maser/stereo_waves
Naming convention	The collections naming convention are: st<<s>>_wav_<<r>>_l3_df_v<<v>> The individual data file naming convention is: st<<s>>_l3_wav_<<r>>_<<d>>_v<<v>>_cdf with: - <<s>> = A or B (spacecraft id);	<ul style="list-style-type: none"> table name: <u>stereo_waves.epn_core</u> granule unique identifiers: <u>product_id</u> (see <u>column A</u>)

MASER / DMP / DOI

Persistent identifier (DOI)

- ObsParis can mint DOI with Datacite.
- One DOI per collection/dataset/document.
 - Landing Page content: *title, citation, abstract, link to data, description, acknowledgments, references*
 - Web-semantic annotation (schema.org)
- Current status: **manual process for**
 - creation / maintenance of DOIs (on Datacite portal)
 - creation / maintenance of Landing Page (SPIP)
 - creation / maintenance of annotations (JSON-LD)Only two persons authorised.

MASER / DMP / DOI

Persistent identifier (DOI)

STEREO/Waves/LFR-HFR L3 DF Data Collection V1.0

Monday 13 February 2023, by Baptiste Cecconi, Xavier Bonnin

This collection contains the daily direction finding (aka goniopolarimetric) Level 3 datasets for the Waves/LFR and Waves/HFR receivers onboard the STEREO-A and STEREO-B spacecraft.

- ▶ DOI: <https://doi.org/10.25935/4tak-5225>
- ▶ Publisher: PADC/MASER
- ▶ License: [CC-BY 4.0](#)
- ▶ Citation: Krupar, V., Q.N. Nguyen, X. Bonnin, B. Cecconi & M. Maksimovic (2022). STEREO/Waves/LFR-HFR L3 DF Data Collection (Version 1.0) [Data set]. PADC. <https://doi.org/10.25935/4TAK-5225>

Link to data repository

- ▶ Direct access to data:
 - STEREO-A/Waves HFR: [sta_wav_hfr_l3_df_v01](#)
 - STEREO-A/Waves LFR: [sta_wav_lfr_l3_df_v01](#)
 - STEREO-B/Waves HFR: [stb_wav_hfr_l3_df_v01](#)
 - STEREO-B/Waves LFR: [stb_wav_lfr_l3_df_v01](#)

Description

The data collection contains daily Level 3 direction finding (aka goniopolarimetric) CDF files, derived from the Low Frequency Receiver (LFR) and High Frequency Receiver (HFR) of STEREO-A/Waves and STEREO-B/Waves spacecraft. This data collection has been produced following the goniopolarimetric inversion described in Krupar et al (2012).

Content

The collection is composed of daily CDF files, which variables are described in the table below. They are available in the `cdf/` directory, with a directory structure based in the year and month of observation.

CDF Variables descriptions

Name	Type	Length	Unit	Description
Epoch	CDF_TT2000	(N)	ns	Epoch encoded as Terrestrial Time on rotating Earth geoid, ns since J2000
SENSOR_CONFIG	CDF_BYTE	(N,319,3)		Dual channel receiver antenna configuration
TIMING	CDF_FLOAT	(N,319,3)	s	Temporal offset of each antenna channel configuration
QUALITY_FLAG	CDF_INT	(N)		High level information about the quality of the electric field measurements
INTERPOL_FLAG	CDF_INT	(N,319)		Flag that indicates if the current record is real or interpolated
PSD_V2	CDF_FLOAT	(N,319,3,2)	V ² /Hz	Power spectral density before 1% background subtraction and antenna calibration
PSD_FLUX	CDF_FLOAT	(N,319)	W/m ² /Hz	Power spectral density after 1% background subtraction and antenna calibration
PSD_SFU	CDF_FLOAT	(N,319)	sfu	Power spectral density after 1% background subtraction and antenna calibration in solar radio flux unit normalized to 1 au
STOKES_I	CDF_FLOAT	(N,319)	W/m ² /Hz	Radio flux density

WAVE_AZIMUTH_HCI	CDF_FLOAT	(N,319)	deg	Wave vector azimuth in the Heliocentric Inertial (HCI) system
WAVE_COLATITUDE_HEE	CDF_FLOAT	(N,319)	deg	Wave vector colatitude in the Heliocentric Earth Ecliptic (HEE) system
WAVE_AZIMUTH_HEE	CDF_FLOAT	(N,319)	deg	Wave vector azimuth in the Heliocentric Earth Ecliptic (HEE) system
WAVE_COLATITUDE_HEEQ	CDF_FLOAT	(N,319)	deg	Wave vector colatitude in the Heliocentric Earth Equatorial (HEEQ) system
WAVE_AZIMUTH_HEEQ	CDF_FLOAT	(N,319)	deg	Wave vector azimuth in the Heliocentric Earth Equatorial (HEEQ) system
SC_POS_HCI	CDF_FLOAT	(N,3)	km	STEREO spacecraft position in the Heliocentric Inertial (HCI) system in km
SC_POS_HEE	CDF_FLOAT	(N,3)	km	STEREO spacecraft position in the Heliocentric Earth Ecliptic (HEE) system in km
SC_POS_HEEQ	CDF_FLOAT	(N,3)	km	STEREO spacecraft position in the Heliocentric Earth Equatorial (HEEQ) system in km

The file naming convention is: `stX_l3_wav_RRR_yyyymmdd_vVV.cdf`

- ▶ X: a 1-letter code for the spacecraft in use (a=STEREO-A; b=STEREO-B);
- ▶ RRR: receiver in use (lfr or hfr).
- ▶ yyyy: Year in 4 digits
- ▶ mm: Month in 2 digits (0-padded)
- ▶ dd: Day in 2 digits (0-padded)
- ▶ vv: version number (2 digits)

Each dataset also contains quicklooks, in the `png/` directory, with the same directory structure. Raw IDL Saveset files are also available in the `sav/` directory.

Coverage and sampling

- ▶ Time range:
 - 2006 to now for STEREO-A
 - 2006 to 2014 for STEREO-B
- ▶ Time resolution: 30 seconds to 2 minutes.
- ▶ Spectral range: 2.5 kHz to 15 MHz

Acknowledgements

The STEREO/Waves L3 DF data collection has been produced by V Krupar (NASA/GSFC), QN Nguyen X Bonnin, B Cecconi and M Maksimovic (LESIA/Observatoire de Paris-PSL), with the support of CNRS and CNES.

BC was also supported by PADC and EPN2024-RI. The Europlanet 2024 Research Infrastructure (EPN2024-RI) project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.

Contact


Any question or request should be addressed to contact.maser@obspm.fr

References

- ▶ Krupar, V. et al. 2012 Goniopolarimetric inversion using SVD: An application to type III radio bursts observed by STEREO. J. Geophys. Res. 117. [doi:10.1029/2011ja017333](https://doi.org/10.1029/2011ja017333)

MASER / DMP / DOI



Persistent identifier (DOI)

About Support INIST.OBSPARIS ▾

Observatoire de Paris / DOIs

10.25935/4tak-5225


[Update DOI \(Form\)](#)
[Update DOI \(File Upload\)](#)

Findable
Metadata Export
[DataCite XML](#)
[DataCite JSON](#)
[Schema.org JSON-LD](#)
[BibTeX](#)
DOI created
September 1, 2022, 10:10:54 UTC
DOI registered
October 14, 2022, 10:17:29 UTC
DOI last updated
February 13, 2023, 16:19:19 UTC
Metadata version 
9
Schema 
[DataCite Metadata Schema 4](#)

URL


<https://maser.lesia.obspm.fr/publications/doi/stereo-waves-lfr-hfr-l3-df-data.html>

Metadata

Summary View 


STEREO/Waves/LFR-HFR L3 DF Data Collection

Collection
Vratislav Krupar, Quynh Nhu Nguyen, Xavier Bonnin, Baptiste Cecconi, Milan Maksimovic,
Version 1.0 of Collection published 2022 via PADC
This collection is contains the daily direction finding (aka goniopolarimetric) Level 3 datasets for the Waves/LFR and Waves/HFR receivers onboard the STEREO-A and STEREO-B spacecraft.



<https://doi.org/10.25935/4tak-5225>

Citation

APA 

Krupar, V., Nguyen, Q. N., Bonnin, X., Cecconi, B., & Maksimovic, M. (2022). *STEREO/Waves/LFR-HFR L3 DF Data Collection (Version 1.0)*. PADC. <https://doi.org/10.25935/4TAK-5225>

schema.org JSON-LD: <https://api.datacite.org/dois/application/vnd.schemaorg.ld+json/10.25935/4tak-5225>

MASER / DMP / DOI

What works and what doesn't

- Not difficult to set up. Adequate template/tools.
Note:
 - no « name resolver » for ORCID
- Maintenance and versioning:
 - lack of time et persons (mostly because it is manual process)
- **How to Improve?**
 - use an MA-DMP (machine-actionable DMP), but it must adapted to our template.
=> ongoing RDA works, but not mature enough
 - Use of *Recherche.data.gouv* for DOI /Landing page management (French national repository based on Dataverse)
=> requires institutional decision...
=> needs training with Dataverse (use and development)

DOI in publications — side topic

Historically for journal article citation

- When not used, strange things can happen... even with A&A...
Example: <https://doi.org/10.1051/0004-6361/202245293>
 - their ref 5: Carley, E. P., et al. 2019, Nat. Commun., 10, 2276.
 - No DOI, but a Google Scholar link...
- What about data citation?
Various publisher policies for open data. Examples:
 - **AGU journals, Elsevier Journals (like PSS):**
 - data is cited the same way as papers
 - requirement to cite to progenitor data
 - resulting data products must be open access
 - full citation including DOI.
 - **A&A:**
 - data policy implies CDS is the default repository
 - data citation with footnotes or acknowledgments (may include DOI)
 - not listed in references => issue with tracking of reuse?
 - no DOI in PDF version => how to cite data in repositories?

DOI in publications — side topic

Historically for journal article citation

- When not used, strange things can happen... even with A&A...
Example: <https://doi.org/10.1051/0004-6361/202245293>
 - their ref 5: [Carley, E. P.](#), et al. 2019, Nat. Commun., 10, 2276.
 - No DOI, but a Google Scholar link...

- What about data citation?
Various publisher policies for open data. Examples:

- **AGU journals, Elsevier Journals (like PSS):**
 - data is cited the same way as papers
 - requirement to cite to progenitor data
 - resulting data products must be open access
 - full citation including DOI.
- **A&A:**
 - data policy implies CDS is the default repository
 - data citation with footnotes or acknowledgments (may include DOI)
 - not listed in references => issue with tracking of reuse?
 - no DOI in PDF version => how to cite data in repositories?

A&A paper example (cites paper without DOI):

[Novel hybrids of adaptive neuro-fuzzy inference system \(ANFIS\) with several metaheuristic algorithms for spatial susceptibility assessment of seismic-induced ...](#)

[H Moayedi, M Mehrabi, B Kalantar... - ... , Natural Hazards and ...](#), 2019 - Taylor & Francis

Strong ground motions usually trigger lots of slope failures in the affected area. In this work, we analyse the occurrence likelihood of earthquake-triggered landslide by employing the ensembles of adaptive neuro-fuzzy inference systems (ANFIS) with four well-known metaheuristics techniques, namely particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimization (ACO), and differential evolution (DE) algorithms. Twelve landslide conditioning factors namely, elevation, slope degree, lithology, peak ground ...

☆ Enregistrer Citer Cité 69 fois Autres articles Les 3 versions Web of Science: 70

DOI in publications — side topic

Historically for journal article citation

- When not used, strange things can happen... even with A&A...
Example: <https://doi.org/10.1051/0004-6361/202245293>
 - their ref 5: Carley, E. P., et al. 2019, Nat. Commun., 10, 2276.
 - No DOI, but a Google Scholar link...

- What about data citation?
Various publisher policies for open data. Examples:

- **AGU journals, Elsevier Journals (like PSS):**
 - data is cited the same way as papers
 - requirement to cite to progenitor data
 - resulting data products must be open access
 - full citation including DOI.

Fogg, A. R., Jackman, C., Waters, J. E., Bonnin, X., Lamy, L., Cecconi, B., et al. (2021). *Bursts of auroral kilometric radiation individually selected from wind/waves data*. version 1.0.

- **A&A:**
 - data policy implies CDS is the default repository
 - data citation with footnotes or acknowledgments (may include DOI)
 - not listed in references => issue with tracking of reuse?
 - no DOI in PDF version => how to cite data in repositories?

A&A paper example (cites paper without DOI):

Novel hybrids of adaptive neuro-fuzzy inference system (ANFIS) with several metaheuristic algorithms for spatial susceptibility assessment of seismic-induced ...

H Moayedi, M Mehrabi, B Kalantar... - ... , Natural Hazards and ..., 2019 - Taylor & Francis

Strong ground motions usually trigger lots of slope failures in the affected area. In this work, we analyse the occurrence likelihood of earthquake-triggered landslide by employing the ensembles of adaptive neuro-fuzzy inference systems (ANFIS) with four well-known metaheuristics techniques, namely particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimization (ACO), and differential evolution (DE) algorithms. Twelve landslide conditioning factors namely, elevation, slope degree, lithology, peak ground ...

☆ Enregistrer Citer Cité 69 fois Autres articles Les 3 versions Web of Science: 70

PSS paper example (cites data with DOI):

Cecconi, Baptiste, Zarka, Philippe, 2019. Cassini Rpws Jupiter Encounter Calibrated Dataset v1.0 [dataset]. PADC. <https://doi.org/10.25935/H98J-MA66>.

JGR paper example (cites data, but no DOI):

A&A paper example (cites data with DOI in footnote):

Example: <https://doi.org/10.1051/0004-6361/202244914>

DOI in publications – side topic

Historically for journal article citation

- When not used, strange things can happen... even with A&A...
Example: <https://doi.org/10.1051/0004-6361/202245293>
 - their ref 5: **Carley, E. P.**, et al. 2019, Nat. Commun., 10, 2276.
 - No DOI, but a Google Scholar link...

- What about data citation?
Various publisher policies for open data. Examples:

- **AGU journals, Elsevier Journals (like PSS):**
 - data is cited the same way as papers
 - requirement to cite to progenitor data
 - resulting data products must be open access
 - full citation including DOI.

Fogg, A. R., Jackman, C., Waters, J. E., Bonnin, X., Lamy, L., Cecconi, B., et al. (2021). *Bursts of auroral kilometric radiation individually selected from wind/waves data*. version 1.0.

- **A&A:**
 - data policy implies CDS is the default repository
 - data citation with footnotes or acknowledgments (may include DOI)
 - not listed in references => issue with tracking of reuse?
 - no DOI in PDF version => how to cite data in repositories?

A&A paper example (cites paper without DOI):

Novel hybrids of adaptive neuro-fuzzy inference system (ANFIS) with several metaheuristic algorithms for spatial susceptibility assessment of seismic-induced ...

H Moayed, M Mehrabi, B Kalantar... - ... , Natural Hazards and ..., 2019 - Taylor & Francis

Strong ground motions usually trigger lots of slope failures in the affected area. In this work, we analyse the occurrence likelihood of earthquake-triggered landslide by employing the ensembles of adaptive neuro-fuzzy inference systems (ANFIS) with four well-known metaheuristics techniques, namely particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimization (ACO), and differential evolution (DE) algorithms. Twelve landslide conditioning factors namely, elevation, slope degree, lithology, peak ground ...

☆ Enregistrer Citer Cité 69 fois Autres articles Les 3 versions Web of Science: 70

PSS paper example (cites data with DOI):

Cecconi, Baptiste, Zarka, Philippe, 2019. Cassini Rpws Jupiter Encounter Calibrated Dataset v1.0 [dataset]. PADC. <https://doi.org/10.25935/H98J-MA66>.

JGR paper example (cites data, but no DOI):

A&A paper example (cites data with DOI in footnote):

Example: <https://doi.org/10.1051/0004-6361/202244914>

=> burden on Author to check references (even more for data, when DOI is not shown by default)

MASER / DMP / DOI

Where there is room for improvement

- No ORCID « name resolver » => manual process.
- Maintenance & versioning:
 - lacking time and people (manual work at this point)
- Use an external repository for managing landing pages and PIDs:
 - Zenodo, EUDAT/B2Share, Dataverse... (+Recherche.data.gouv in France)
- Various publisher policies for open data.
 - **Possible improvements**
 - data must be cited the same way as papers
 - requirement to cite to progenitor data
 - resulting data products must be open access
 - **full citation including DOI should be mandatory**