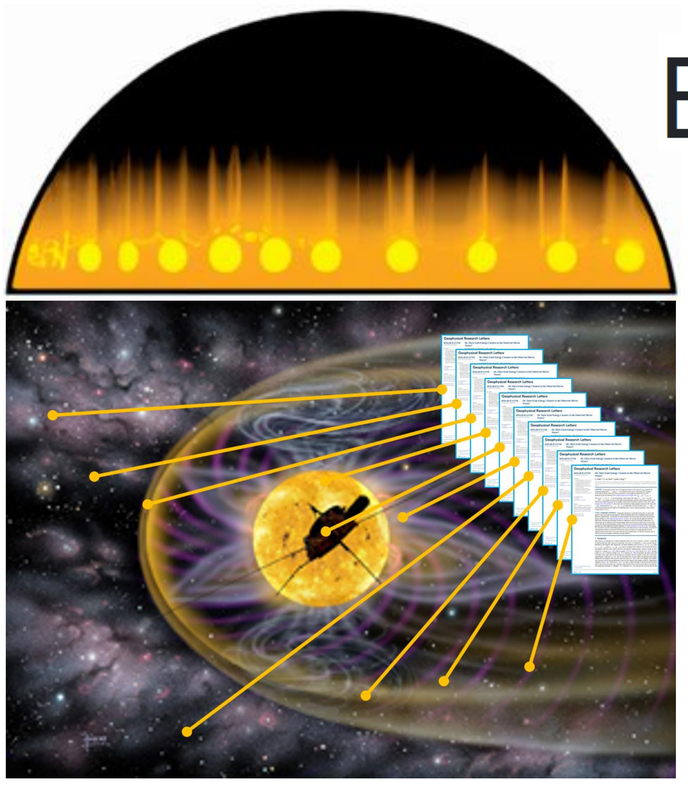
BibHelioTech

Vincent Génot, Axel Dablanc, Richard Hitier, Camille de Salabert, Sabine Barreaux, Pascal Cuxac, Guillaume Cabanac, Dominica Leung, Nicolas Aunai

Abstract ID: 689
Abstract Title: BibHelioTech: cataloguing and documenting all heliospheric events ?
Session: CD07
Display time: Poster II, Wednesday 14:00 - Friday 12:00
Authors in attendance time: Thursday 10:15-11:45, Friday 10:15-11:45



BibHelioTech

Vincent Génot, Axel Dablanc, IRAP (Toulouse)
Richard Hitier, SAPIE (Limoux)
Camille de Salabert, Sabine Barreaux, Pascal Cuxac, Inist-CNRS (Nancy)
Guillaume Cabanac, IRIT (Toulouse)
Dominica Leung, CNES (Toulouse)
Nicolas Aunai, LPP (Palaiseau)

DOI 10.5281/zenodo.7746571

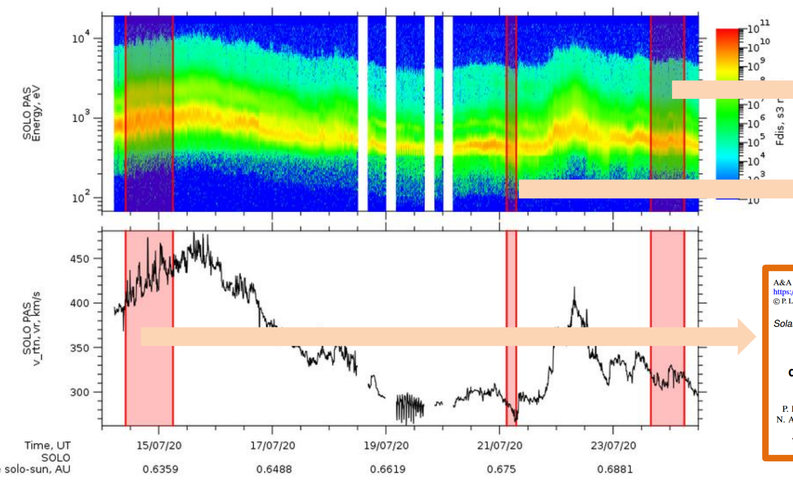
From a corpus of heliophysics articles using data from space missions, BibHelioTech carries out automated textual detections on the observed events, the satellites/instruments used, the spatial regions and the physical processes concerned, in order to link these entities with the publications from which they are extracted, in catalogs usable by the discipline's data analysis tools. This strong and systematized link between data and publications, non-existent to date, 1/ will increase the data analysis experience by immersing the researcher in the bibliographic context of their case study, 2/ will significantly improve the reproducibility of published results, and 3/ will facilitate the reuse of these catalogs in new statistical and comparative studies.

publi data

BibHelioTech is on **GitHub**

Use case

- Visualization of Solar Orbiter PAS data in **AMDA** in July 2020
- The **3 intervals** in red correspond to those studied in the **3 articles** on the right
- **BibHelioTech** will produce catalogs, bridging intervals and articles, and readily integrable in tools like **AMDA** and **Sciqlop**



A&A 656, A12 (2021) https://doi.org/10.1051/0004-6361/202140915 © R. Kieokaew et al. 2021 — Astronomy & Astrophysics, Special issue
Solar Orbiter First Results (Cruise Phase)
Solar Orbiter observations of the Kelvin-Helmholtz waves in the solar wind
R. Kieokaew[1], B. Lavraud[1,2], Y. Yang[3], W. H. Matthaeus[4], D. Ruffolo[5], J. E. Stawarz[6], S. Aizawa[1], C. Foullon[7], V. Génot[1], R. F. Pinto[1,8], N. Fargette[1], P. Louarn[1], A. Rouillard[1], A. Fedorov[1], E. Penou[1], C. J. Owen[9], T. S. Horbury[6], H. O'Brien[6], V. Evans[6], and V. Angelini[6]

A&A 656, A34 (2021) https://doi.org/10.1051/0004-6361/202140998 © ESO 2022 — Astronomy & Astrophysics, Special issue
Solar Orbiter First Results (Cruise Phase)
Simulations of radio-wave anisotropic scattering to interpret type III radio burst data from Solar Orbiter, Parker Solar Probe, STEREO, and Wind
S. Musset[1,2], M. Maksimovic[3], E. Kontar[2], V. Krupar[4,5], N. Chrysaphi[1,2], X. Bonnin[3], A. Vecchio[3,6], B. Cecconi[3], A. Zaslavsky[3], K. Issautier[3], S. D. Bale[7,8,9], and M. Pulupa[7]

A&A 656, A36 (2021) https://doi.org/10.1051/0004-6361/202141095 © P. Louarn et al. 2021 — Astronomy & Astrophysics, Special issue
Solar Orbiter First Results (Cruise Phase)
Multiscale views of an Alfvénic slow solar wind: 3D velocity distribution functions observed by the Proton-Alpha Sensor of Solar Orbiter
P. Louarn[1], A. Fedorov[1], L. Prech[2], C. J. Owen[3], R. Bruno[4], S. Livi[5], B. Lavraud[1,14], A. P. Rouillard[1], V. Génot[1], N. André[1], G. Fruit[1], V. Réville[1], R. Kieokaew[1], I. Plotnikov[1], E. Penou[1], A. Barthe[6], D. Khataria[1], M. Berthomier[7], R. D'Amicis[4], L. Sorriso-Valvo[15,17], F. Allegrini[5], J. Raines[8], D. Verscharen[3,16], V. Fortunato[9], G. Mele[10], T. S. Horbury[11], H. O'brien[11], V. Evans[11], V. Angelini[11], M. Maksimovic[12], J. C. Kasper[8], and S. D. Bale[13]

Prototype pipeline and admin website

Upload / exe / database interface

Upload new File — Parcourir... Aucun fichier sélectionné. Upload
Upload from URL — Upload

#	Title	del	pdf	cat	run	status
1	ark_67375_80W-JTG4N0JL-N	del	pdf	cat	run	finished 0:04:25
2	A51606C002834CBC0D0FEE52E896579C22E80384	del	pdf	cat	run	finished 0:03:38
3	ark_67375_80W-FX54V24C-8	del	pdf	cat	run	
4	ark_67375_80W-RTQRJRCK-8	del	pdf	cat	run	
5	ark_67375_80W-NGMLHKT9-3	del	pdf	cat	run	finished 0:02:54

Example catalogue in HPEvent format

```
# Name: 1010022017ja024061_bibheliotech_V1;
# Creation Date: 2023-06-14T13:55:44.858821;
# Description: Catalogue of events resulting from the HelioNER code (Dablanc & Génot, "https://github.com/ADablanc/BibHelioTech.git") on the paper
"https://doi.org/10.1002/2017ja024061". The two first columns are the start/stop times of the event. the third column is the DOI of the paper, the fourth column is
the mission that observed the event with the list of instruments (1 or more) listed in the fifth column. The sixth column is the most probable region of space where
the observation took place (SPASE ObservedRegions term);
# Parameter 1: id:column1; name:DOI; size:1; type:char;
# Parameter 2: id:column2; name:SATS; size:1; type:char;
# Parameter 3: id:column3; name:INSTS; size:1; type:char;
# Parameter 4: id:column4; name:REGS; size:1; type:char;
# Parameter 5: id:column5; name:D; size:1; type:int;
# Parameter 6: id:column6; name:R; size:1; type:int;
# Parameter 7: id:column7; name:SO; size:1; type:int;
# Parameter 8: id:column8; name:occur_sat; size:1; type:int;
# Parameter 9: id:column9; name:nb_durations; size:1; type:int;
# Parameter 10: id:column10; name:conf; size:1; type:float;
2015-09-20T06:00:00.000 2015-09-20T06:00:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 316 1 56 131 108 0.0008907430375465104
2015-09-20T06:10:00.000 2015-09-20T06:10:59.999 https://doi.org/10.1002/2017ja024061 "OMNI" "" "Heliosphere.NearEarth" 580 1 2 131 108 0.04577742699289661
2015-09-20T06:20:00.000 2015-09-20T06:20:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 624 1 56 131 108 0.0017589356184462737
2015-09-20T06:20:00.000 2015-09-20T06:40:59.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 2405 1 53 131 108 0.007162961087697875
2015-09-20T06:20:00.000 2015-09-20T06:40:59.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 2492 1 53 131 108 0.007422078598978422
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 7162 1 56 131 108 0.02018829631300034
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 21935 1 56 131 108 0.0618305333182997
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 21757 1 53 131 108 0.0648002263555271
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "THEMIS-E" "FGM" "Earth.Magnetosheath" 1784 1 14 131 108 0.020115007328898413
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "THEMIS-E" "FGM" "Earth.Magnetosheath" 3558 1 14 131 108 0.04011726237456309
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "THEMIS-E" "FGM" "Earth.Magnetosphere.Main" 6651 1 14 131 108 0.07499154357875747
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1002/2017ja024061 "THEMIS-E" "FGM" "Earth.Magnetosheath" 18646 1 14 131 108 0.21023790731762318
2015-09-20T06:33:50.000 2015-09-20T06:34:15.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 6442 1 53 131 108 0.01918660928355497
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 10666 1 56 131 108 0.030065396324275566
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 18340 1 56 131 108 0.05169692186266772
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 8988 1 53 131 108 0.026769519441259253
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 20095 1 53 131 108 0.059850188381409 06
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1002/2017ja024061 "THEMIS-E" "FGM" "Earth.Magnetosheath" 10652 1 14 131 108 0.12010373210057503
2015-09-20T06:37:15.000 2015-09-20T06:37:40.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 6706 1 53 131 108 0.019972896903992496
2015-09-20T06:37:15.000 2015-09-20T06:37:40.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 7509 1 53 131 108 0.022364521749489955
2015-09-20T06:37:55.000 2015-09-20T06:38:20.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 6445 1 53 131 108 0.019195544370150853
2015-09-20T06:38:30.000 2015-09-20T06:42:00.999 https://doi.org/10.1002/2017ja024061 "MMS" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 32156 1 56 131 108 0.090641560495999 97
2015-09-20T06:38:30.000 2015-09-20T06:40:10.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 27107 1 53 131 108 0.0807344641181814
2015-09-20T06:38:30.000 2015-09-20T06:40:10.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 28529 1 53 131 108 0.08496969516462897
2015-09-20T06:38:30.000 2015-09-20T06:42:00.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 29756 1 53 131 108 0.08862414558234426
2015-09-20T06:38:30.000 2015-09-20T06:42:00.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 30677 1 53 131 108 0.09136721716727 97
2015-09-20T06:38:30.000 2015-09-20T06:39:40.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 30219 1 53 131 108 0.09000312728030856
2015-09-20T06:38:30.000 2015-09-20T06:39:40.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 31996 1 53 131 108 0.09529567690726869
2015-09-20T06:38:30.000 2015-09-20T06:40:10.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 31601 1 53 131 108 0.09411922383881104
2015-09-20T06:38:30.000 2015-09-20T06:40:10.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 31671 1 53 131 108 0.09432770919271492
2015-09-20T06:38:30.000 2015-09-20T06:40:10.999 https://doi.org/10.1002/2017ja024061 "MMS4" "DIS,HPCA,FGM,EDP,DES,FPI" "Earth.Magnetosphere" 31750 1 53 131 108 0.09456299980640646
```

Catalogues per mission

Available Missions:

ACE 110 | AIM 1 | AMPTE 2 | BepiColombo 20 | CCE 1 | Cassini 101 | Cluster 68 | Cluster-1 11 | Cluster-3 3 | DMSP 4 | Dawn 8 | GOES 55 | GONG 8 | Galileo 2 | Geotail 86 | Helios 11 | Hinode 8 | IMAGE 85 | IMP 2 | IMP-6 2 | IMP-8 1 | ISEE 6 | ISEE-1 1 | Interball 1 | Interball-Tail 1 | L'Aquila 1 | MESSENGER 62 | MEX 1 | MMS 118 | MMS1 1 | MMS4 53 | Mariner-10 17 | MarsScienceLaboratory 1 | NOAA 1 | NSO 1 | OMNI 3 | PSP 104 | PVO 3 | Pioneer 6 | Pioneer-11 1 | RHESSI 4 | SOHO 9 | SPSG 1 | STEREO 6 | STEREO-A 33 | SolarOrbiter 210 | THEMIS-E 26 | Thule 5 | Ulysses 16 | VEX 3 | Vega 1 | Voyager 106 | Voyager-2 31 | Wind 94 | Yohkoh 1

210 events for SolarOrbiter Download

start	stop	doi	instrument(s)	region
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855 CR	RPW,MAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPW,MAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPW,MAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPW,MAG,SWA,PAS	Heliosphere.Inner

ISTEX

Le plus vaste réservoir d'archives scientifiques au service de la recherche française
27,9 M de documents — 10 249 revues — 437 306 ebooks — publiés de 1473 à 2022

https://demo.istex.fr/
https://api.istex.fr/

- Large article database
- PDF, TXT, enriched TXT, XML
- API for direct access

Example search for 'magnetopause' in ISTEX

```
https://api.istex.fr/document/?q=magnetopause&facet=corpusName[*]&size=10
&sortBy=publicationDate[desc]&output=*&stats
```

Requête | Réponse brute complète

Corpus: wiley 8869, elsevier 2389, springer-journals 1322, springer-ebooks 1193, iop 220, nature 192, edp-sciences 106

Résultats : 14399 (180 ms) — 1/ 1440

Effect of the solar wind density on the evolution of normal and inverse cor…
Context. The evolution of magnetised coronal mass ejections (CMEs) and their interaction with the background solar wind leading to deflection, deformation, and erosion is still largely unclear as there is very little observational data available. Even so, this evolution is very…
Fulltext | Metadata | Annexes | Enrichments
edp-sciences | research-article | Astronomy & Astrophysi… | ark:/67375/80W-97JN0… | Score : 9.333 | Mots : 11912 | Publication : 2019

What is inside ?

- OCR for efficient (*but costly*) PDF2TXT transformation
- GROBID for DOI recognition
- **SUTime** library for time pattern recognition
- **SPASE** dictionary for spatial regions
- Lists of missions + instruments
- Interval – mission association is done with a proximity criterion

What are the performances ?

- Several minutes / article (*too much !*)
- Good recognition of time patterns
- Fair interval – mission association
- Confidence index helps sorting intervals

What are the plans ?

- Get rid of OCR, GROBID and use ISTEX metadata and file formats
- Massive analysis with **ISTEX API**
- AI to get article / observations summary
- Integration in **AMDA** / **Sciqlop**

European Space Weather Week
20-24 November, 2023
Toulouse, France