



HAL
open science

BibHelioTech

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

► **To cite this version:**

Vincent Génot, Axel Dablanc, Richard Hitier, Camille de Salabert, Sabine Barreaux, et al.. BibHelioTech. European Space Weather Week, Nov 2023, Toulouse, France. hal-04285006

HAL Id: hal-04285006

<https://hal.science/hal-04285006v1>

Submitted on 14 Nov 2023

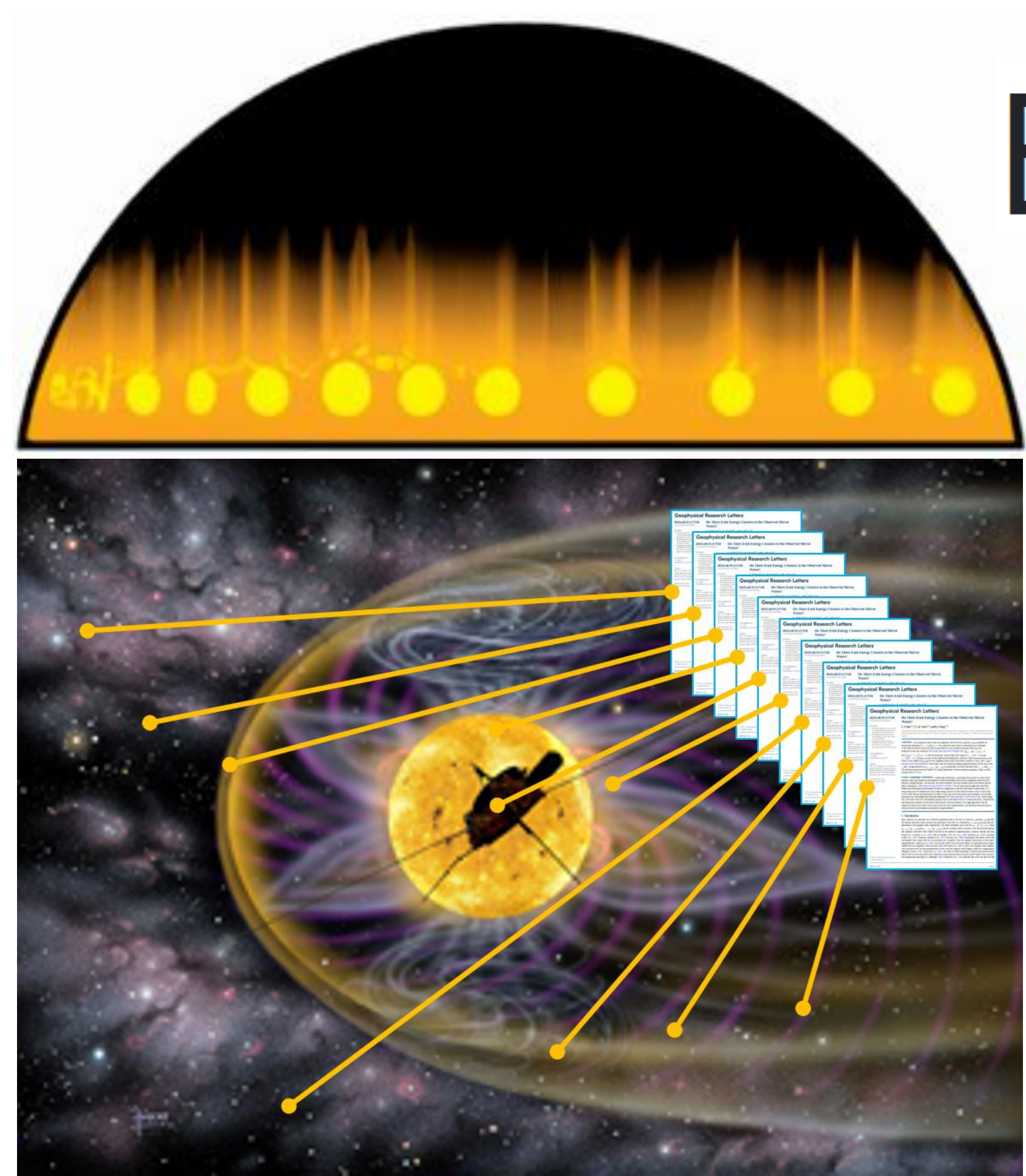
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

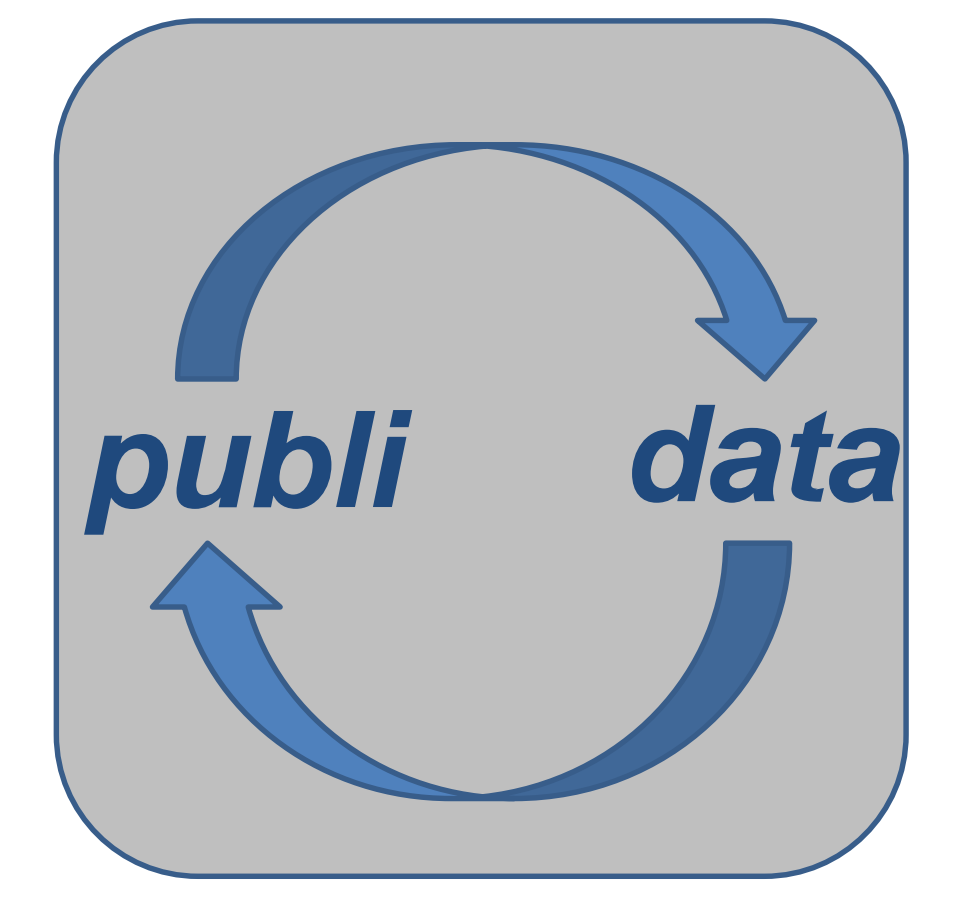
Abstract ID: 689
 Abstract Title: BibHelioTech: cataloguing and documenting all heliospheric events ?
 Session: CD07
 Display time: Poster II, Wednesday 14:00 - Friday 12:00
 Authors in attendance time: Thursday 10:15-11:45, Friday 10:15-11:45



BibHelioTech

Vincent G not, Axel Dablan, IRAP (Toulouse)
 Richard Hiti r, 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

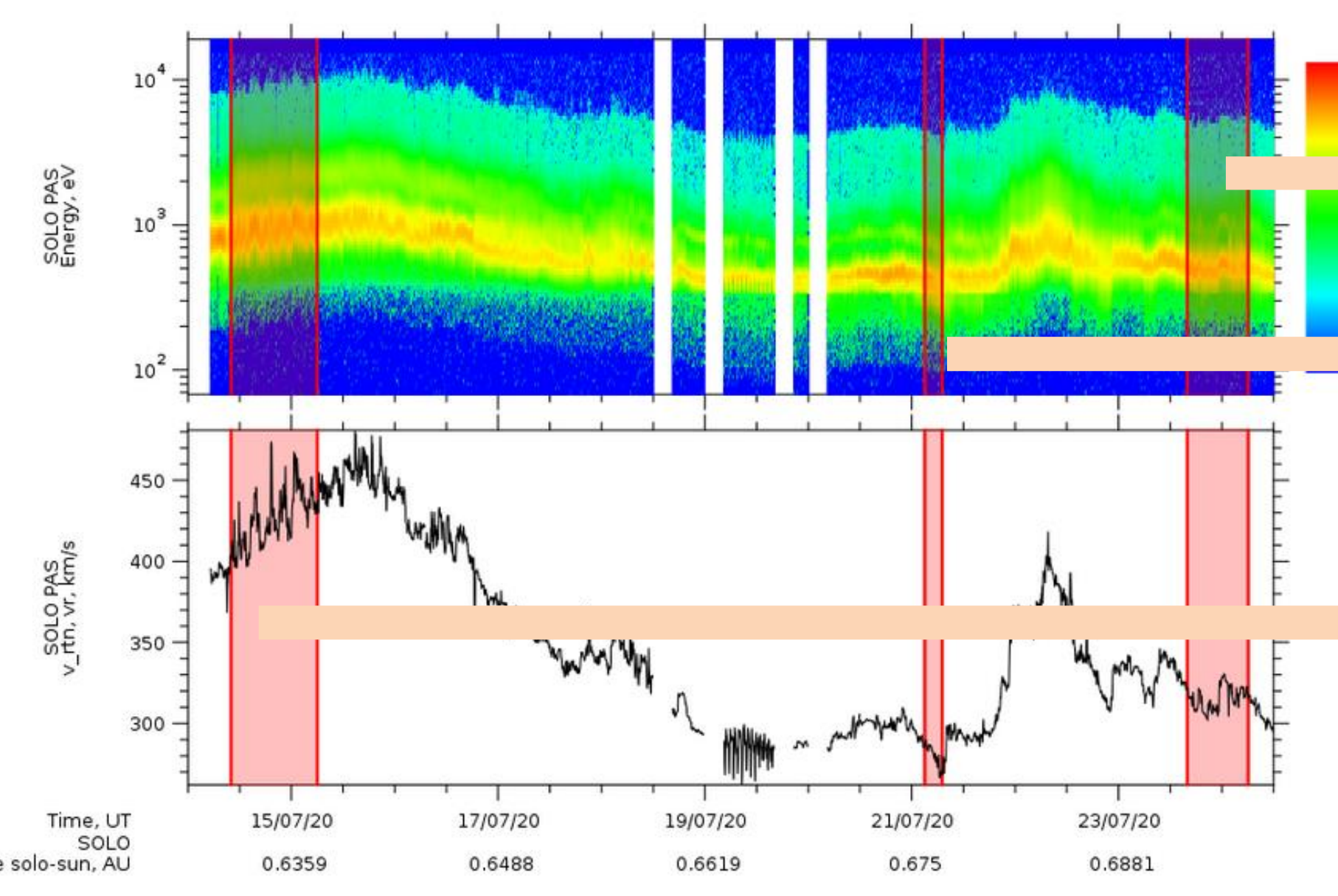


BibHelioTech is on **GitHub**

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.

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**



AAA 656, A12 (2021)
<https://doi.org/10.1051/0004-6361/202148015>
 R. Knolke et al. 2021
Solar Orbiter observations of the Kelvin-Helmholtz waves in the solar wind
 R. Knolke¹, B. Lavraud², Y. Yang³, W. H. Matthaeus⁴, D. Raffalli⁵, J. E. Sturrock⁶, S. Aizawa⁷, C. Foullet⁸, V. Gloc⁹, R. F. Pinto¹⁰, N. Fargnoli¹¹, P. Louarn¹², A. Rouillard¹³, A. Fokov¹⁴, E. Pennu¹⁵, C. J. Owen¹⁶, T. S. Horbury¹⁷, H. O'Brien¹⁸, V. Evans¹⁹, and V. Angelini²⁰

AAA 656, A14 (2021)
<https://doi.org/10.1051/0004-6361/202148096>
 R. Knolke et al. 2021
Simulations of radio-wave anisotropic scattering to interpret type III radio burst data from Solar Orbiter, Parker Solar Probe, STEREO, and Wind
 S. Mous^{1,2}, M. Maksimovic^{3,4}, E. Kontar^{5,6}, V. Krupar^{7,8}, N. Chryssoulis^{9,10}, X. Bonin¹¹, A. Vecchia^{12,13}, B. Cecconi¹⁴, A. Zarnecki¹⁵, K. Isomura¹⁶, S. D. Bale^{17,18}, and M. Pilipp¹⁹

AAA 656, A16 (2021)
<https://doi.org/10.1051/0004-6361/202144096>
 P. Louarn et al. 2021
Multiscale views of an Alfvénic slow solar wind: 3D velocity distribution functions observed by the Proton-Alpha Sensor of Solar Orbiter
 P. Louarn¹, A. Fokov², L. Prech³, C. J. Owen⁴, R. Bruno⁵, S. Liu⁶, B. Lavraud^{7,8}, A. P. Rouillard⁹, V. Gloc¹⁰, N. Andri s¹¹, G. Fritz¹², V. Gloc¹³, R. Knolke¹⁴, J. Fontana¹⁵, E. Pennu¹⁶, A. Barthel¹⁷, D. Knaack¹⁸, M. Bartholom w¹⁹, R. D'Amicis²⁰, L. Serre-Vahia²¹, F. Albertini²², E. Rauso²³, D. Verschuur²⁴, V. Fortuna²⁵, G. Miki s²⁶, Y. S. Horbury²⁷, H. O'Brien²⁸, V. Evans²⁹, V. Angelini³⁰, M. Maksimovic³¹, J. C. Kasper³², and S. D. Bale³³



Prototype pipeline and admin website

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

Upload / exe / database interface

#	Title	del	pdf	cat	run	status
1	ark_67375_80W-JTG4NOJL-N	del	pdf	cat	run	finished 00:42:25
2	A51606C002834C8C0D0FE52E896579C2E80384	del	pdf	cat	run	finished 00:3:38
3	ark_67375_80W-FXS4V24C-8	del	pdf	cat	run	
4	ark_67375_80W-RTQRJCK-8	del	pdf	cat	run	
5	ark_67375_80W-NGMLHKT9-3	del	pdf	cat	run	finished 00:2:54

Example catalogue in HPEvent format

```

# Name: 1010022017a024061_bibhelio_tech_V1;
# Creation Date: 2023-06-14T13:55:44.855Z;
# Description: Catalogue of events resulting from the HelioNER code (Dablan & G not, "https://github.com/Dablan/BibHelioTech.git") on the paper "https://doi.org/10.1051/0004-6361/202148015". 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 observation took place (OBSERVE 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:INST; size:1; type:char;
# Parameter 4: id:column4; name:REG; 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; 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/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 316 1 56 131 108 0.0008907430375465104
2015-09-20T06:10:00.000 2015-09-20T06:10:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "" "Heliosphere.NearEarth" 580 1 2 131 108 0.045774269928661
2015-09-20T06:20:00.000 2015-09-20T06:20:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 624 1 56 131 108 0.0017589356184462737
2015-09-20T06:30:00.000 2015-09-20T06:30:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 2405 1 53 131 108 0.00742961087697875
2015-09-20T06:40:00.000 2015-09-20T06:40:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 2492 1 53 131 108 0.00742961087697875
2015-09-20T06:50:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 7162 1 56 131 108 0.0201829631300034
2015-09-20T06:59:00.000 2015-09-20T06:59:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 21935 1 56 131 108 0.0618305333182297
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 21751 1 53 131 108 0.064800223555271
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "THEMIS-E" "FGM" "Earth.Magnetosphere" 1784 1 14 131 108 0.02011500732898413
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "THEMIS-E" "FGM" "Earth.Magnetosphere" 3558 1 14 131 108 0.04011726237456309
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "THEMIS-E" "FGM" "Earth.Magnetosphere.Main" 6651 1 14 131 108 0.07499154357878747
2015-09-20T06:30:00.000 2015-09-20T06:50:59.999 https://doi.org/10.1051/0004-6361/202148015 "THEMIS-E" "FGM" "Earth.Magnetosphere" 18646 1 14 131 108 0.2102379071276218
2015-09-20T06:33:00.000 2015-09-20T06:34:15.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 6442 1 53 131 108 0.0191866028355497
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 10666 1 56 131 108 0.03065396324275566
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 10340 1 56 131 108 0.051694992186246772
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 8988 1 53 131 108 0.02476951941259253
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 20095 1 53 131 108 0.05985018838140906
2015-09-20T06:34:00.000 2015-09-20T06:34:59.999 https://doi.org/10.1051/0004-6361/202148015 "THEMIS-E" "FGM" "Earth.Magnetosphere" 10652 1 14 131 108 0.12010373210057503
2015-09-20T06:37:15.000 2015-09-20T06:37:40.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 6706 1 53 131 108 0.01907289603992446
2015-09-20T06:37:15.000 2015-09-20T06:37:40.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 7509 1 53 131 108 0.0223645421749489955
2015-09-20T06:37:15.000 2015-09-20T06:38:20.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 6445 1 53 131 108 0.01919544370150853
2015-09-20T06:38:00.000 2015-09-20T06:42:00.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 32156 1 56 131 108 0.05064156049199997
2015-09-20T06:38:00.000 2015-09-20T06:40:10.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 27107 1 53 131 108 0.0307344661151814
2015-09-20T06:38:00.000 2015-09-20T06:40:10.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 28529 1 53 131 108 0.0849696516462897
2015-09-20T06:38:00.000 2015-09-20T06:42:00.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 29756 1 53 131 108 0.0886241458234426
2015-09-20T06:38:00.000 2015-09-20T06:42:00.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 30677 1 53 131 108 0.0913672171672797
2015-09-20T06:38:00.000 2015-09-20T06:39:40.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 30219 1 53 131 108 0.0900321278603056
2015-09-20T06:38:00.000 2015-09-20T06:39:40.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 31896 1 53 131 108 0.0952567490726689
2015-09-20T06:38:00.000 2015-09-20T06:40:10.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 31601 1 53 131 108 0.0941392283881104
2015-09-20T06:38:00.000 2015-09-20T06:40:10.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 31671 1 53 131 108 0.09432770919271492
2015-09-20T06:38:00.000 2015-09-20T06:40:10.999 https://doi.org/10.1051/0004-6361/202148015 "MMS" "DIS,HPCA,FGM,EDP,DES,FFI" "Earth.Magnetosphere" 31750 1 53 131 108 0.09456299980460466
  
```

ISTEX

Le plus vaste r servoir d'archives scientifiques au service de la recherche fran aise

27,9 M de documents 10 249 revues
 437 306 ebooks publi s de 1473   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

Request: [https://api.istex.fr/document/?q=magnetopause&facet=corpusName\[*\]&size=10&sort=publ1|cat|tondate|desc|dotpuce\[*\]&stats](https://api.istex.fr/document/?q=magnetopause&facet=corpusName[*]&size=10&sort=publ1|cat|tondate|desc|dotpuce[*]&stats)

Response brute complete

Corpus: 14399 (144)

Effect of the solar wind density on the evolution of normal and inverse cor...

- ### 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**

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 LAquila 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	RPWMAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPWMAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPWMAG,SWA,PAS	Heliosphere.Inner
2020-02-10T00:04:00.000	2020-02-10T00:04:59.999	https://doi.org/10.1051/0004-6361/202140855	RPWMAG,SWA,PAS	Heliosphere.Inner

