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
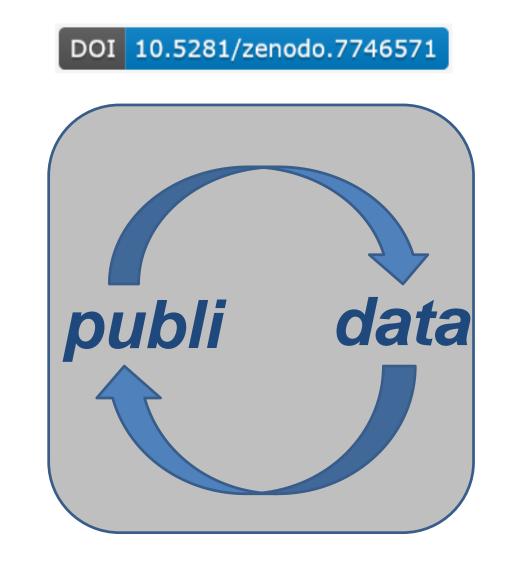
BibHelioTech

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.
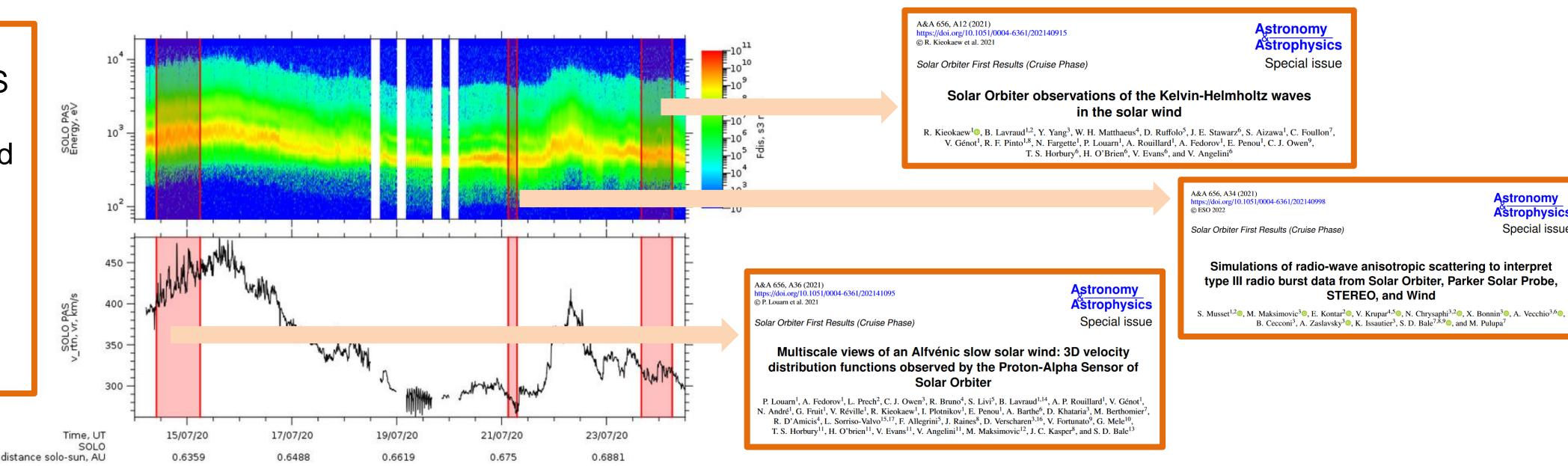
publi → data

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**
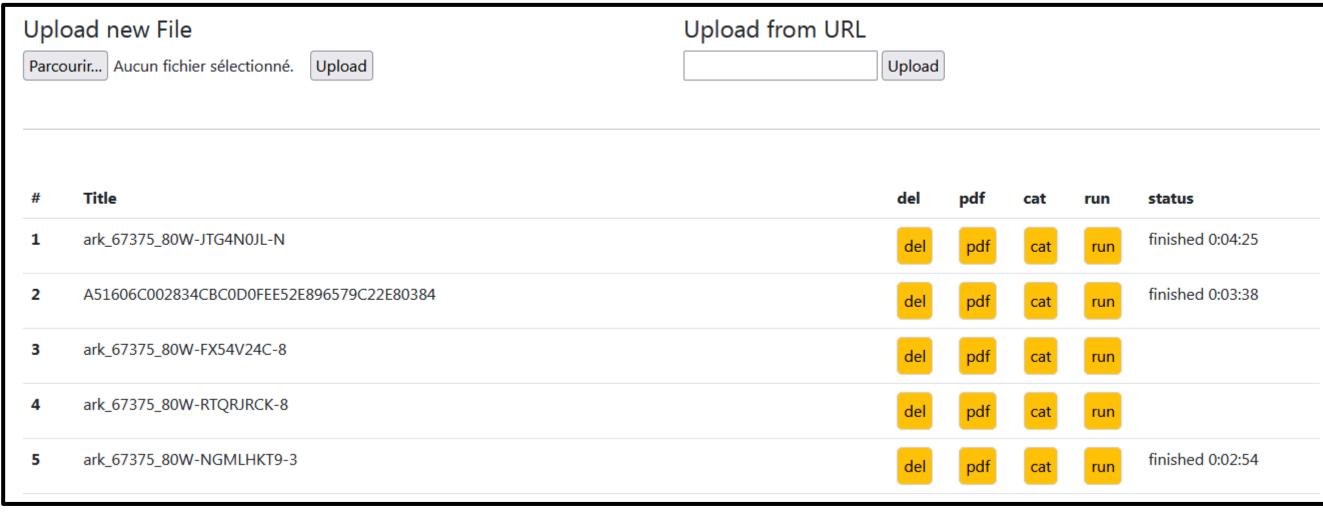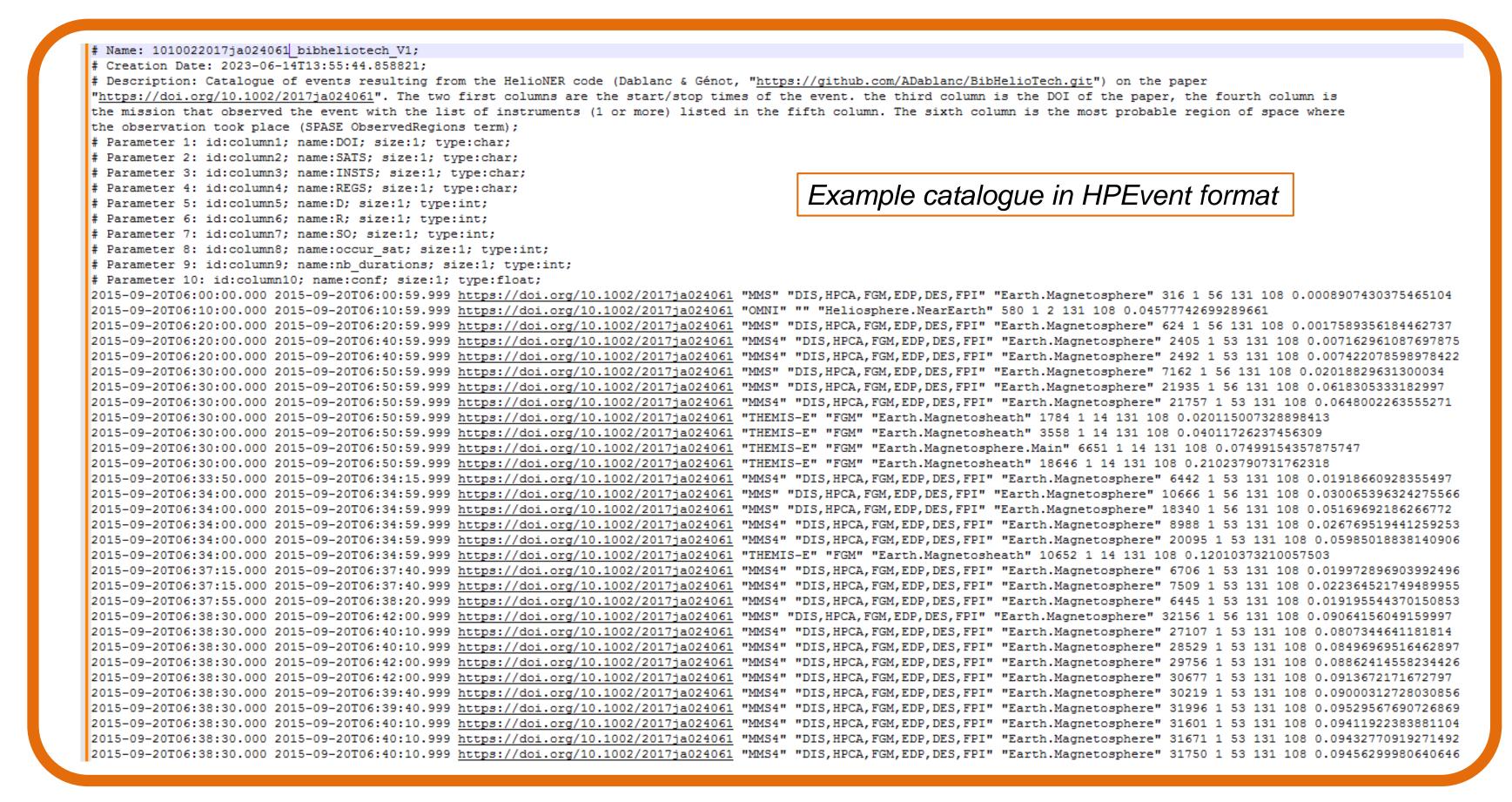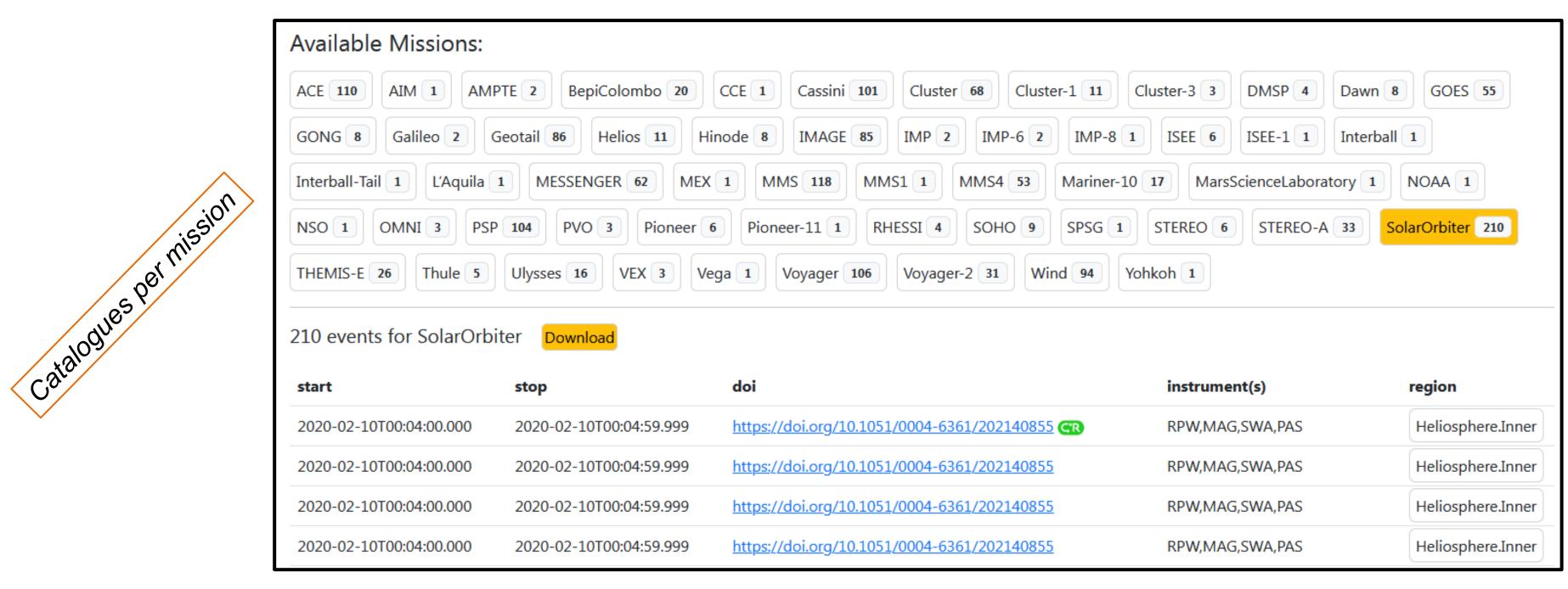

A&A 656, A12 (2021)
Astronomy & Astrophysics — Special issue
Solar Orbiter First Results (Cruise Phase)
Solar Orbiter observations of the Kelvin-Helmholtz waves in the solar wind

A&A 656, A36 (2021)
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

A&A 656, A35 (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

Prototype pipeline and admin website

Upload / exe / database interface

#	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

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 1	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 1	ISEE-1 1	Interball 1
Interball-Tail 1	L'Aquila 1	MESSENGER 62	MEX 1	MMS 118	MMS1 1	MMS4 54	Mariner-10 17	MarsScienceLaboratory 1	NOAA 1
NSO 1	OMNI 3	PSP 104	PVO 3	Pioneer 6	Pioneer-11 1	RHESSI 3	SOHO 9	SPSG 1	STEREO 6	STEREO-A 31	**SolarOrbiter 210**
THEMIS-E 26	Thule 5	Ulysses 16	VEX 1	Vega 1	Voyager 1	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	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
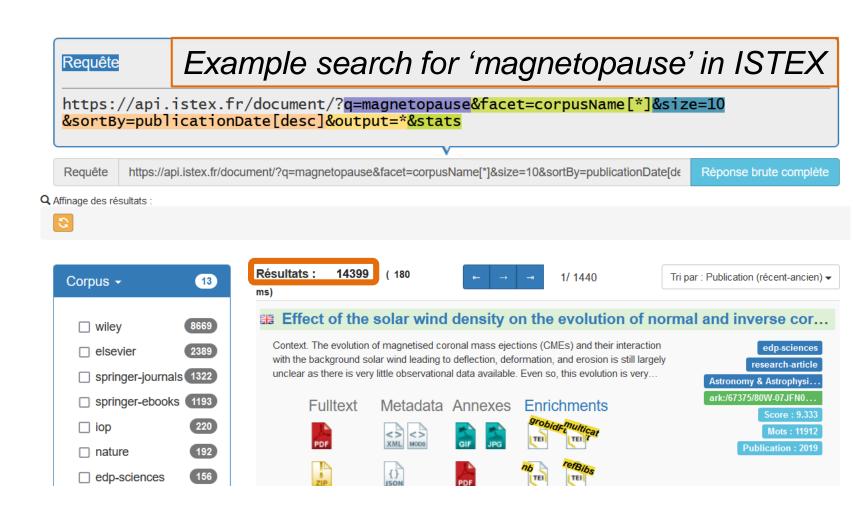
https://demo.istex.fr/
https://api.istex.fr/

- 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

Résultats : 14399

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**

cnes · CNRS · UNIVERSITÉ TOULOUSE III PAUL SABATIER · irap · SAPIE · irit · LPP · COPP · Inist