Data archiving for the next generation of solar high-resolution telescopes

Nazaret Bello González

Kiepenheuer-Institut für Sonnenphysik, Freiburg (Germany)

ASTERICS European Data Provider Forum and Training Event, Heidelberg, 15 June 2016

Data archiving for the next generation of solar high-resolution telescopes

Nazaret Bello González

Kiepenheuer-Institut für Sonnenphysik, Freiburg (Germany)

Scientist at the Experimental Solar Physics group at KIS
Head of the *Center for Advanced Solar Spectro-polarimetric Data Analysis* (CASSDA project)

Outline

- Space-based vs. ground-based solar observatories
- Next generation ground-based solar observatories
  - (US) DKIST & (European) EST 4m-class telescopes
- Challenges for data search from ground-based observations
- CASSDA & SOLARNET projects
- The European SVO - A prototype

Space-based vs. ground-based observatories

- Space-based (solar) telescopes provide with science-ready data to the scientific community

⇨ This translates into a high scientific output (backed up by a high number of publications)

- This is still not the case for ground-based solar observatories..

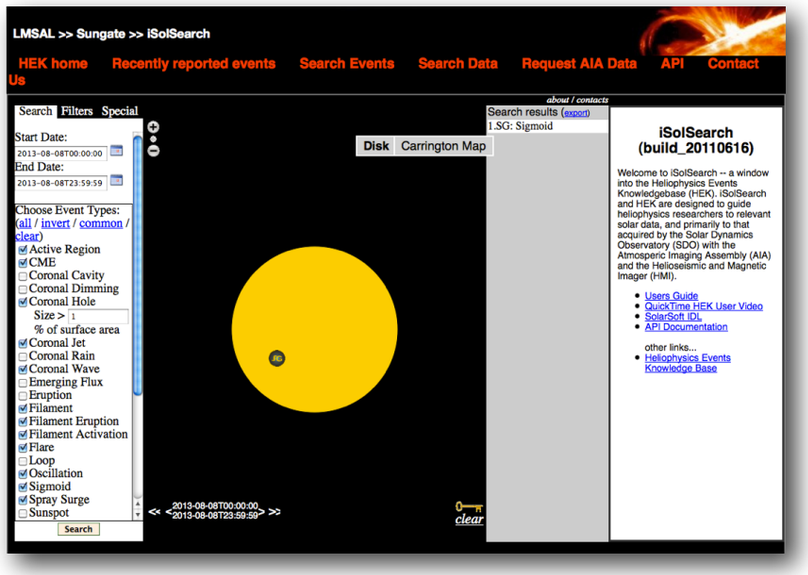
Space-based vs. ground-based observatories

- There is a need to make data discovery easy & to distribute data, especially for ground-based observations

Data archives, e.g., VSO

Browsing tools, e.g., (j)Helioviewer

Event searching, e.g., HEK



González, KIS

Next generation 4m-class ground-based solar telescopes

Nazaret Bello González, KIS

Next generation large solar telescopes

- Next generation ground-based solar observatories
  - (US) DKIST & (European) EST 4m-class telescopes

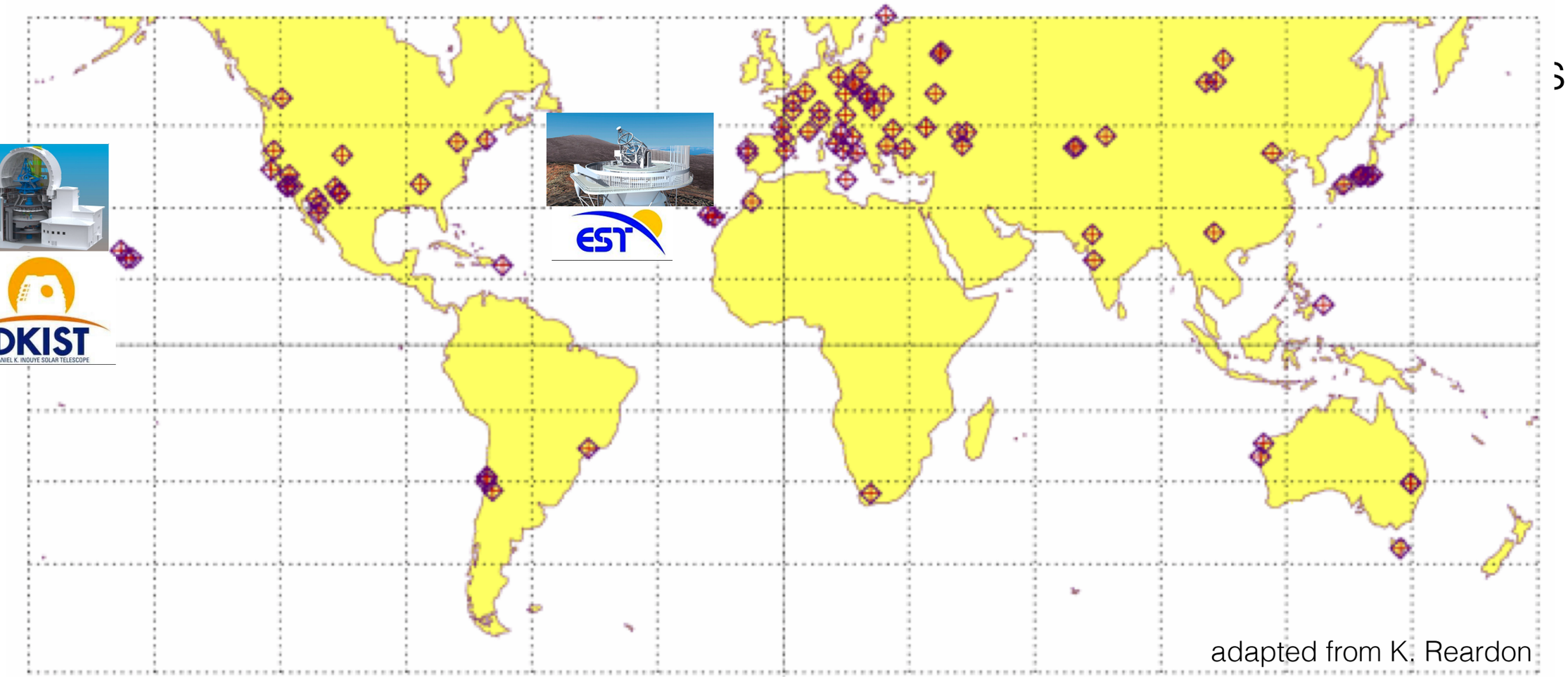
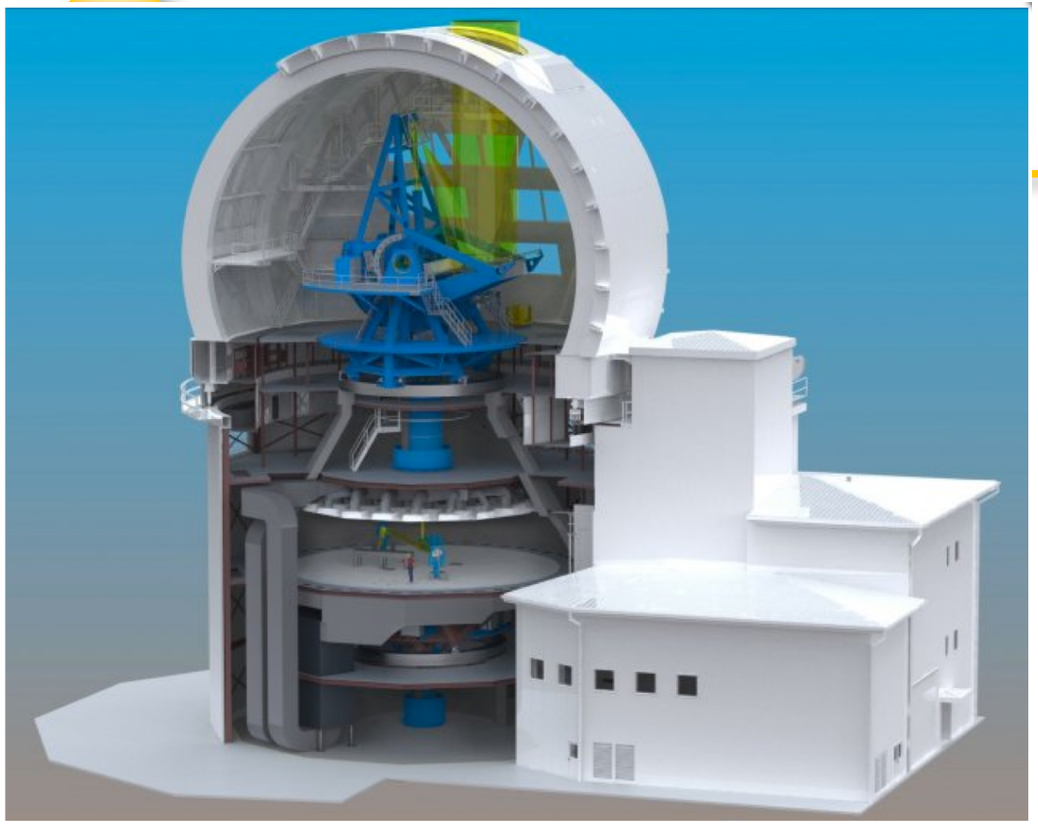


Fig. 1: The distribution of the observatories around the globe.

Nazaret Bello González, KIS

Daniel K. Inouye Solar Telescope

DKIST

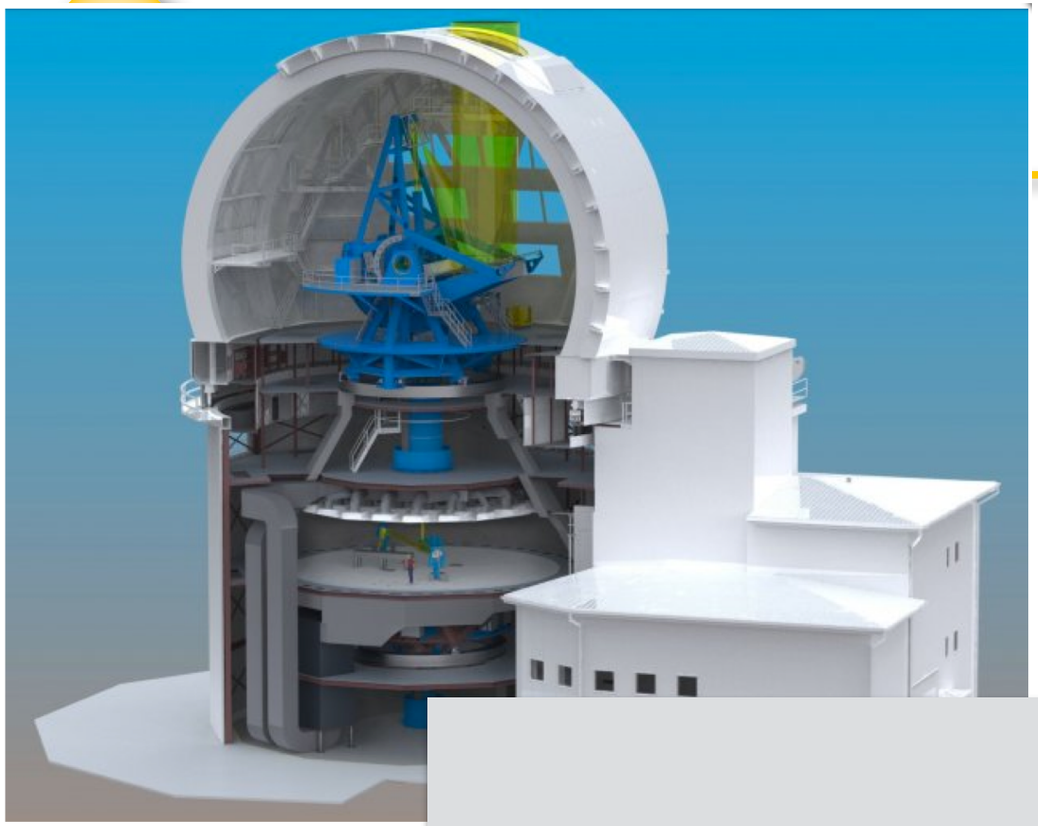


Haleakala Obs., Maui

DKIST is a 4m solar telescope facility under construction by the (US) National Solar Observatory atop the Haleakala volcano in Maui

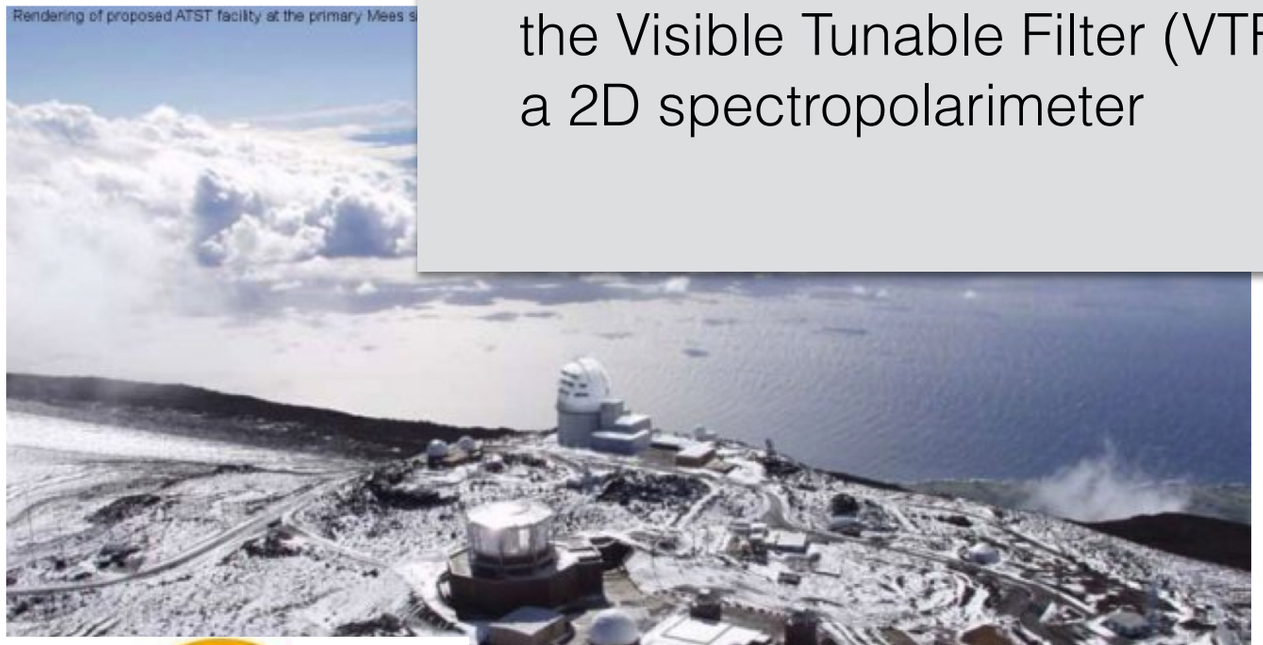
On completion around 2018, it will be the world's largest solar telescope

DKIST



KIS (Germany) is contributing to the DKIST project with the Visible Tunable Filter (VTF) post-focus instrument, a 2D spectropolarimeter

Haleakala Obs



DKIST is a 4m solar telescope facility under construction by the (US) National Solar Observatory atop the Haleakala volcano in Maui

On completion around 2018, it will be the world's largest solar telescope

New generation solar telescopes - Data volume

- The upcoming ground-based large solar telescopes + new post-focus instrumentation (detectors) ➡ vast increase of the data volume

Example: expected DKIST data stream

DKIST Instrument	Detectors	Hourly Data Volume
VBI	1 x 4096 x 4096 detector, 30 fps	**6.5 TB/hour**
DL		
Cry		
AO		**0.3 TB/hour**
		12 TB/hour
		~50 TB/day

~ 20 x increased data volume/hour!

K. Reardon

European Solar Telescope

Nazaret Bello González, KIS

KIS
Kiepenheuer-Institut für Sonnenphysik

EST

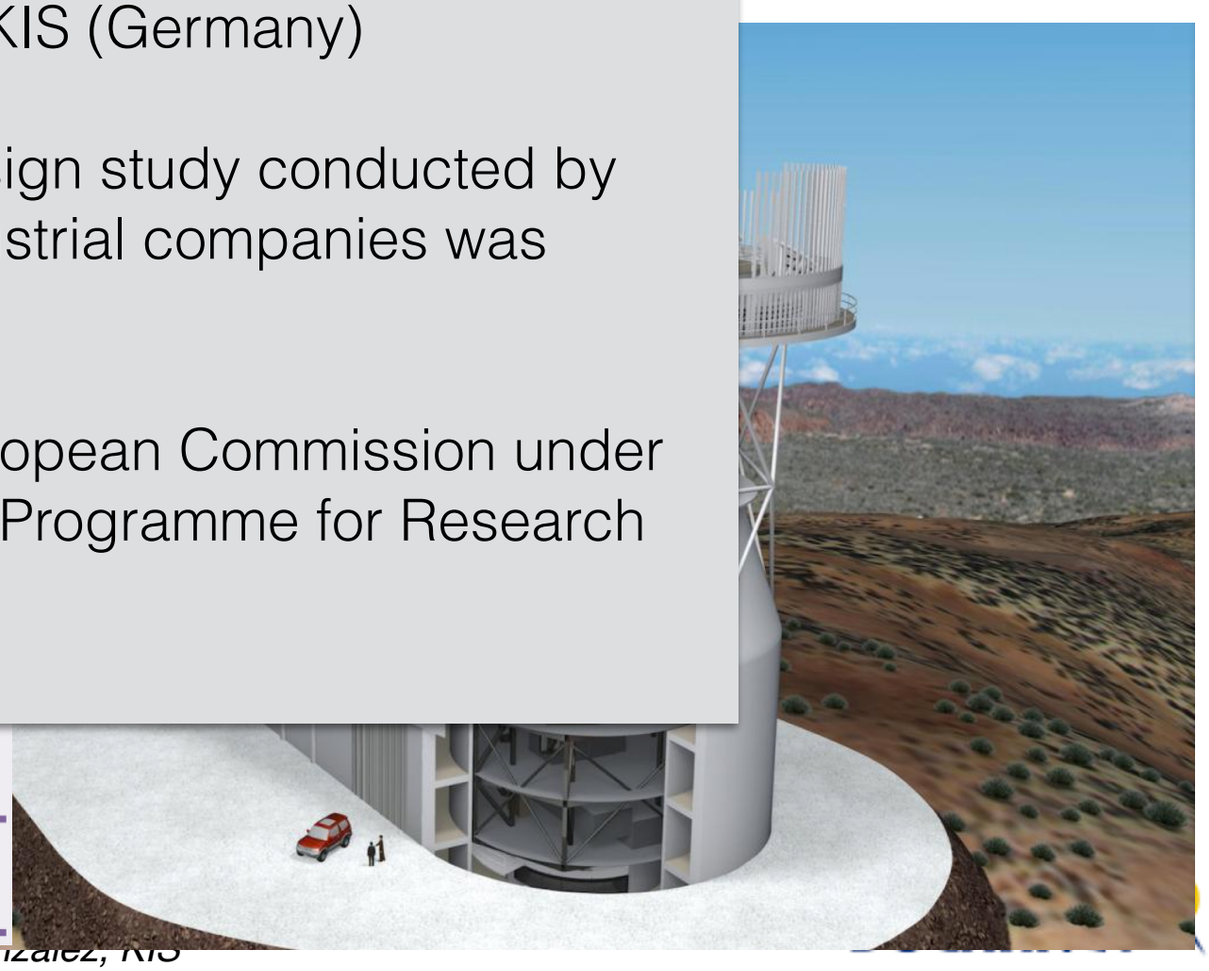
Institute		Location	
IGAM	Institutsbereich Geophysik, Astrophysik und Meteorologie		Graz
HVO	Hvar Observatory		Hvar
AIASCR	Astronomical Institute AS CR		Ondrejov
THEMIS	THEMIS S.L.,[n		
KIS	Kiepenheuer-In		
UniDeb	Heliophysical C		
INAF	Istituto Naziona		
UU	Ultrecht Univers		
ITA	Institute of Theo		
IA UWr	Astronomical In		
AISAS	Astronomical In		
IAC	Instituto de Astr		
SU	The Institute for		
IRSOL	Istituto Ricerche		
UCL-MSSL	University Colle		

The *European Solar Telescope* (EST) is a 4m solar telescope for high-resolution solar observations to be located in the Canary Islands

The EST project engages 15 European institutions lead by the IAC (Spain) and KIS (Germany)

The (3 years) conceptual design study conducted by research institutions and industrial companies was finalised in May 2011

It was co-financed by the European Commission under the EU's Seventh Framework Programme for Research (FP7)



EST

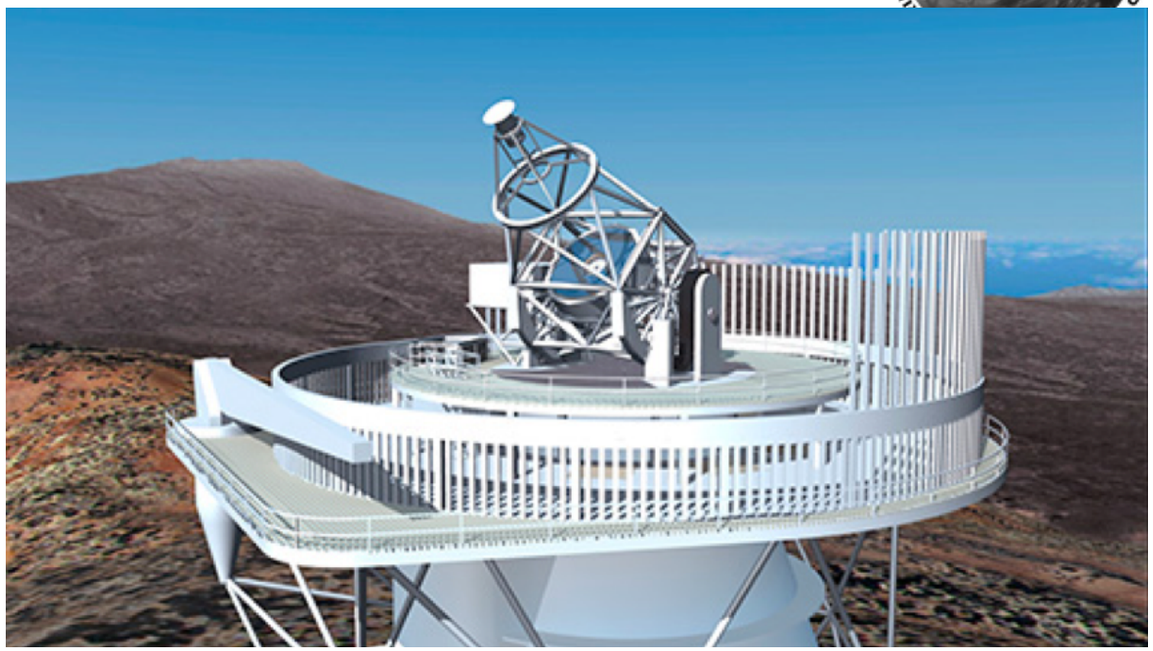
Canaries, Spain

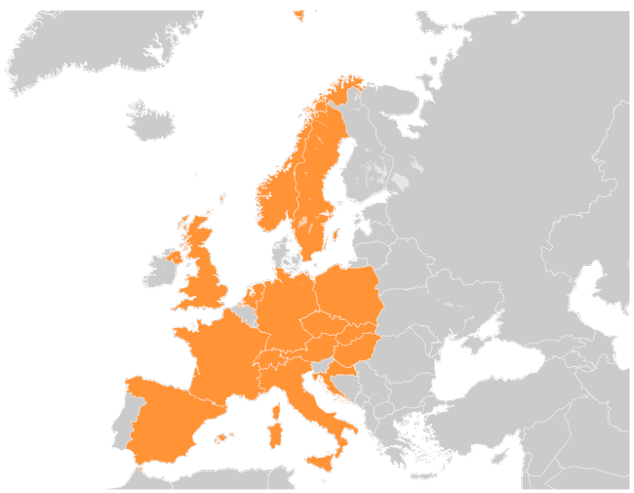
Nazaret Bello González, KIS

KIS
Kiepenheuer-Institut für Sonnenphysik

EST

Institute		Location	
IGAM	Institutsbereich Geophysik, Astrophysik und Meteorologie		Graz
HVO	Hvar Observatory		Hvar
AIASCR	Astronomical Institute AS CR		Ondrejov
THEMIS	THEMIS S.L.,[note 1] INSU-CNRS, CNR		Paris
KIS	Kiepenheuer-Institut für Sonnenphysik		Freiburg
UniDeb	Heliophysical Observatory Debrecen		Debrecen
INAF	Istituto Nazionale di Astrofisica		Rome
UU	Ultrecht University, Sterrekundig Instituut		Utrecht
ITA	Institute of Theoretical Astrophysics		Oslo
IA UWr	Astronomical Institute of the Wroclaw University		Wroclaw
AISAS	Astronomical Institute of the Slovak, Academy of Sciencees		Tatranská Lomnica
IAC	Instituto de Astrofísica de Canarias		La Laguna
SU	The Institute for Solar Physics		Stockholm
IRSOL	Istituto Ricerche Solari		Locarno
UCL-MSSL	University College London - MSSL		London



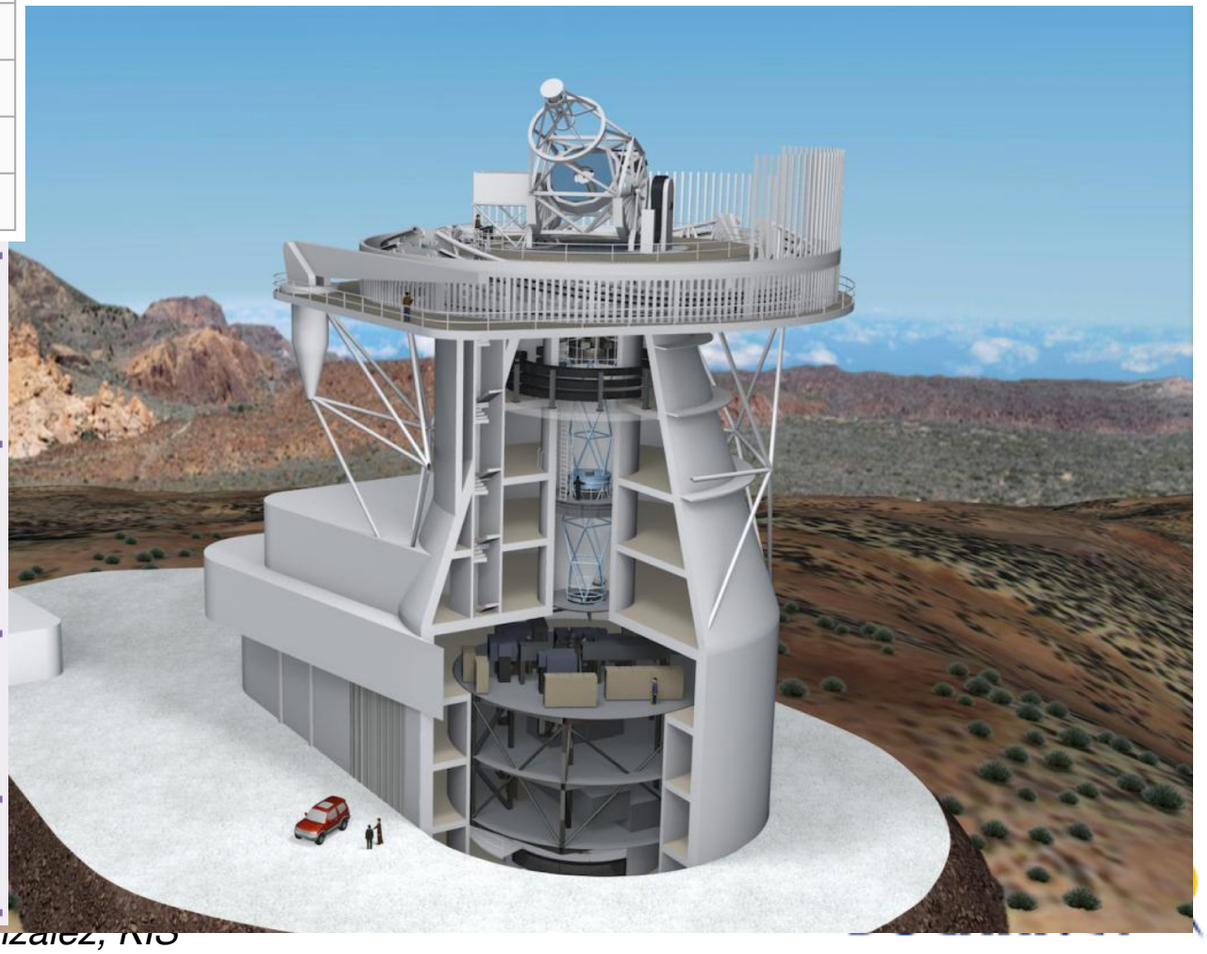


TIMELINE
- ESFRI Roadmap entry: 2016
- Preparation phase: 2011-2019
- Construction phase: 2019-2025
- Operation start: 2026

ESTIMATED COSTS
- Capital value: Not Available
- Preparation: 10 M€
- Construction: 200 M€
- Operation: 9 M€/year

HEADQUARTERS
Instituto de Astrofísica de Canarias
Canary Islands
Spain

WEBSITE
http://www.est-east.eu/



EST

Canaries, Spain

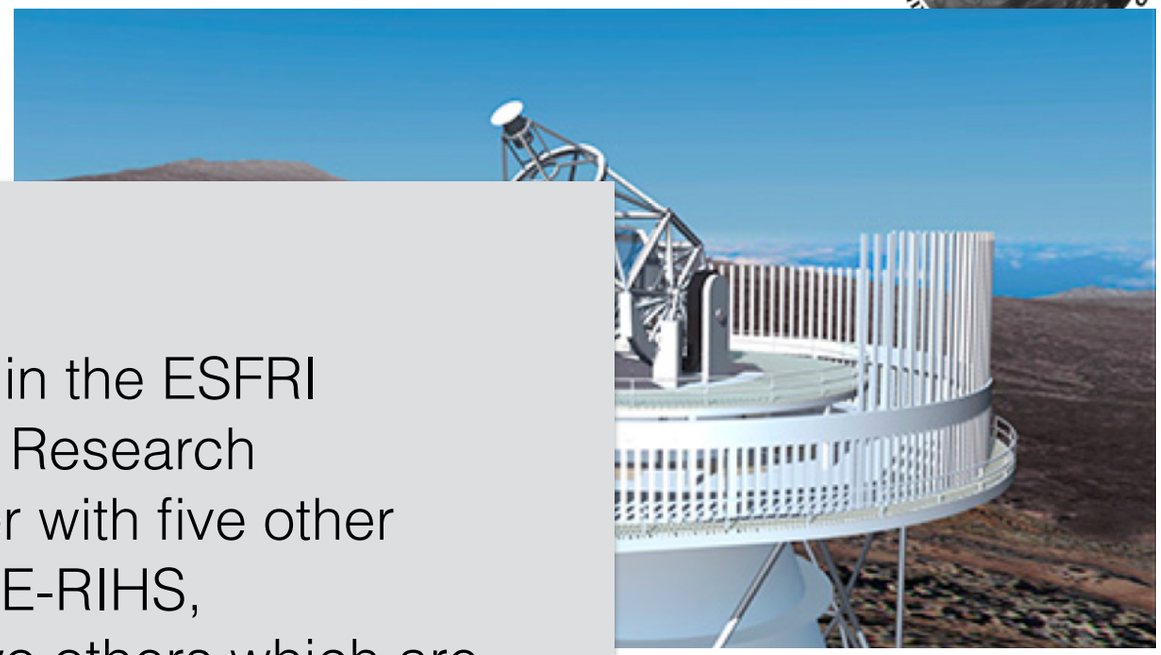
Nazaret Bello González, KIS

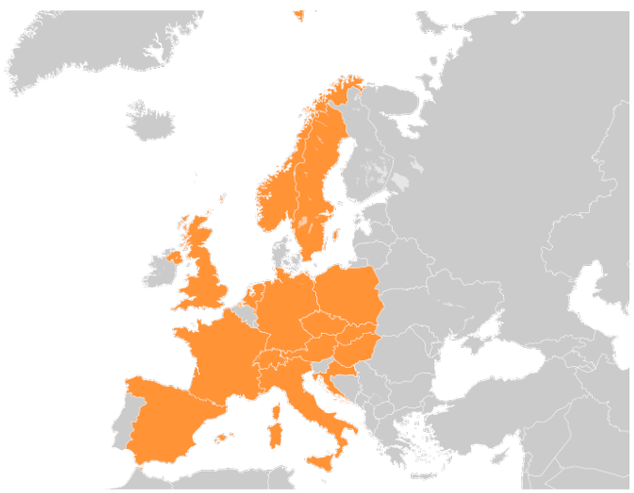
KIS
Kiepenheuer-Institut für Sonnenphysik

EST

Institute		Location	
IGAM	Institutsbereich Geophysik, Astrophysik und Meteorologie		Graz
HVO	Hvar Observatory		Hvar
AIASCR	Astron		
THEMIS	THEMI		
KIS	Kiepen		
UniDeb	Heliop		
INAF	Istituto		
UU	Ultrech		
ITA	Institut		
IA UWr	Astron		
AISAS	Astron		
IAC	Institut		
SU	The Ins		
IRSOL	Istituto		
UCL-MSSL	Univers		

In March 2016, EST was included in the ESFRI (Forum for a European Strategy in Research Inrastructures) route map, together with five other projects (ACTRIS, DANUBIUS-RI, E-RIHS, EMPHASIS Y KM3NeT 2.0) and two others which are considered to be emblematic (CERN LHC Y ESRF EBS)



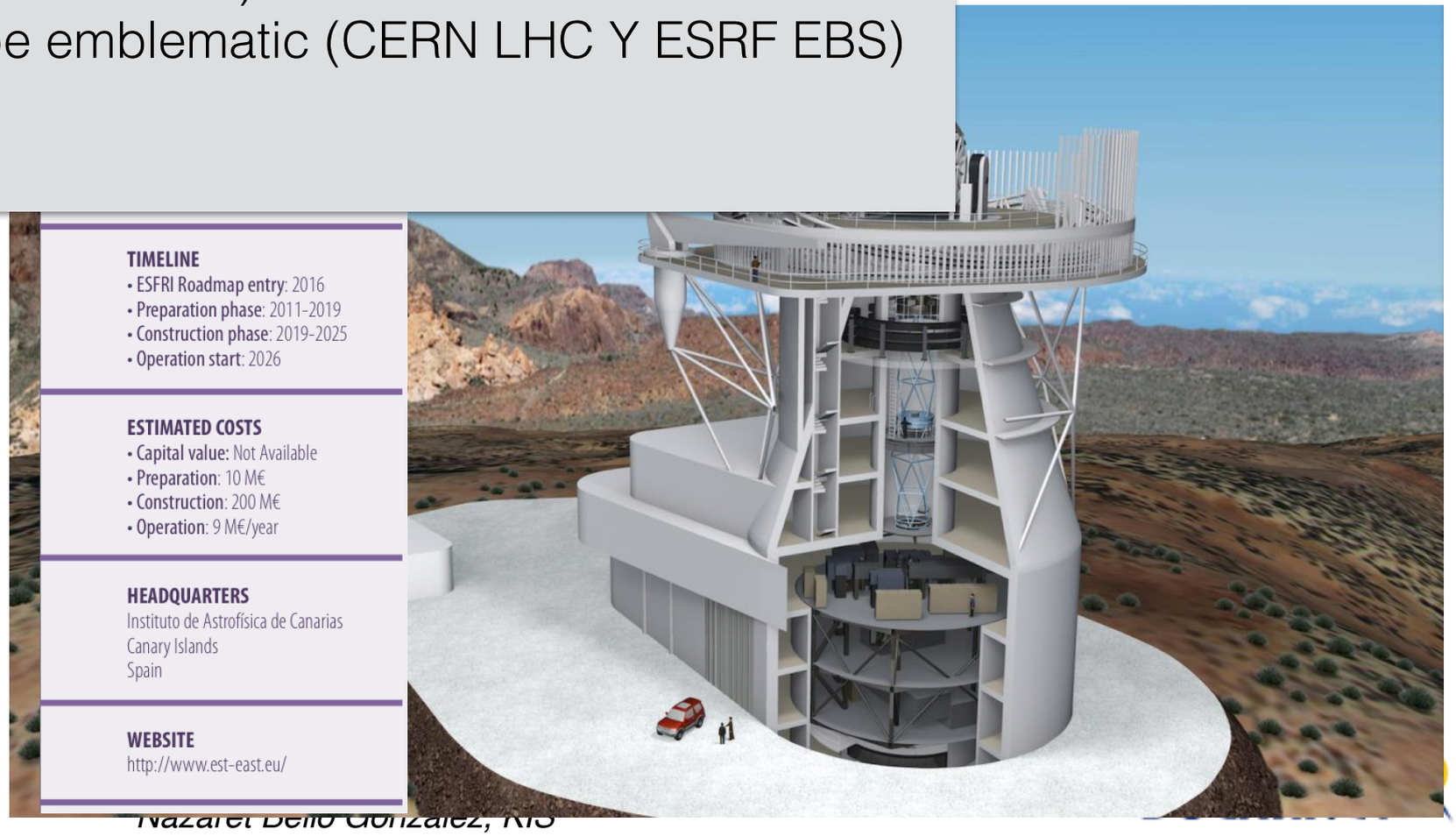


TIMELINE
- **ESFRI Roadmap entry**: 2016
- **Preparation phase**: 2011-2019
- **Construction phase**: 2019-2025
- **Operation start**: 2026

ESTIMATED COSTS
- **Capital value:** Not Available
- **Preparation**: 10 M€
- **Construction**: 200 M€
- **Operation**: 9 M€/year

HEADQUARTERS
Instituto de Astrofísica de Canarias
Canary Islands
Spain

WEBSITE
http://www.est-east.eu/



Canaries, Spain

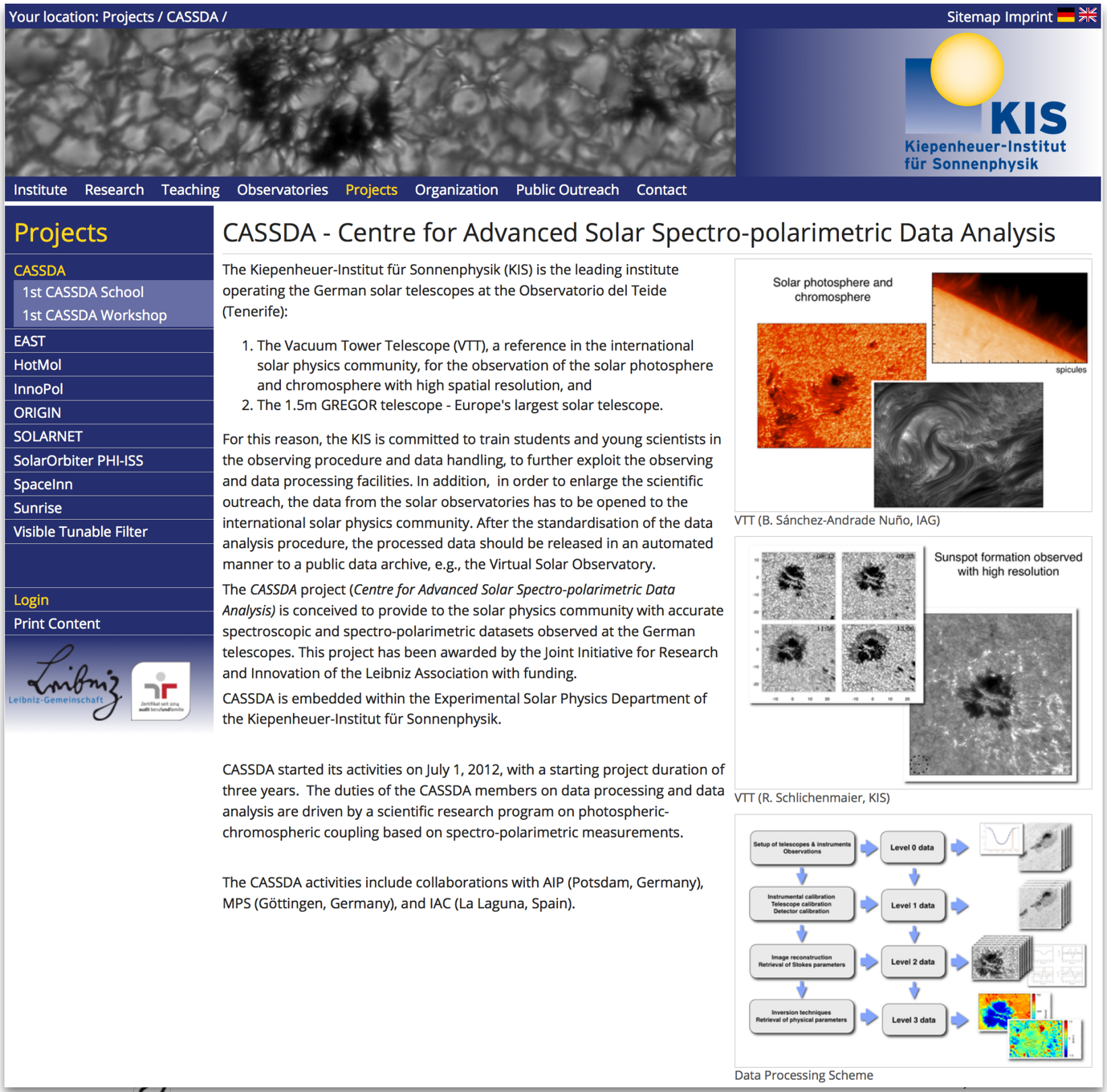
Nazaret Bello González, KIS

The EST project is driving other projects..

CASSDA and SOLARNET projects

CASSDA Project

http://www.solarnet-east.eu/



CASSDA@KIS

Development of data pipelines for the German solar observatories

Startup grant funded by the Leibniz Association

Precursor of the future *European Data Centre* project at KIS

SOLARNET Project

http://www.solarnet-east.eu/

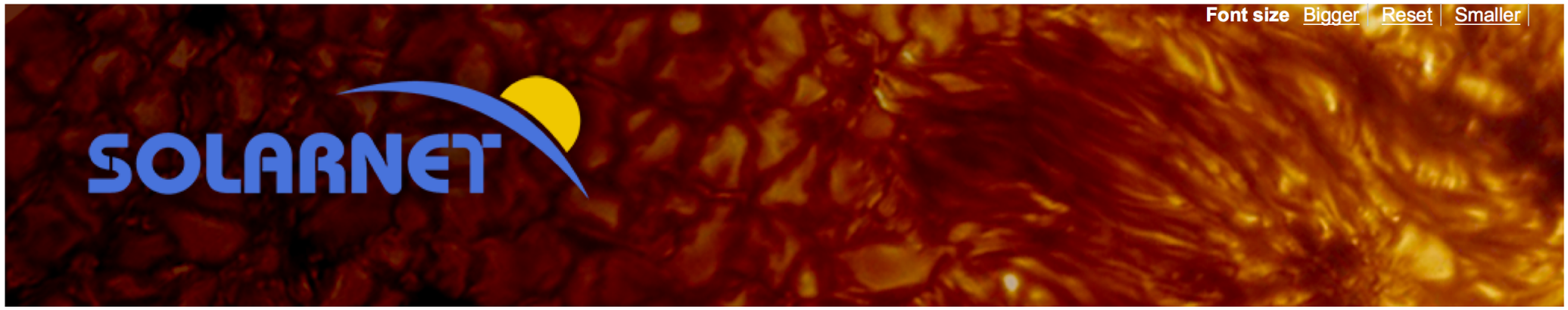
SOLARNET integrates the major research infrastructures in high-res solar physics

Funded by the FP7 since 2013

Getting Ready for EST (GREST) project funded by HORIZON2020

SOLARNET Project

http://www.solarnet-east.eu/



HIGHLIGHTS

Login Form

User Name

Joint Research Activities

The following joint research activities will be carried out to improve the service provided by owners/operators of research infrastructures for research on solar physics:

WP50. Tools for Innovative Data Handling: Pipelines, Databases & SVO

Develop data-reduction pipelines for the most important European ground-based high resolution solar instruments. Enhancement of observational procedures for increased productivity and easier coobserving and combination of data. The pipelines will produce data and meta-data fulfilling the requirements of a Solar Virtual Observatory (SVO). A SVO archive prototype will be implemented.

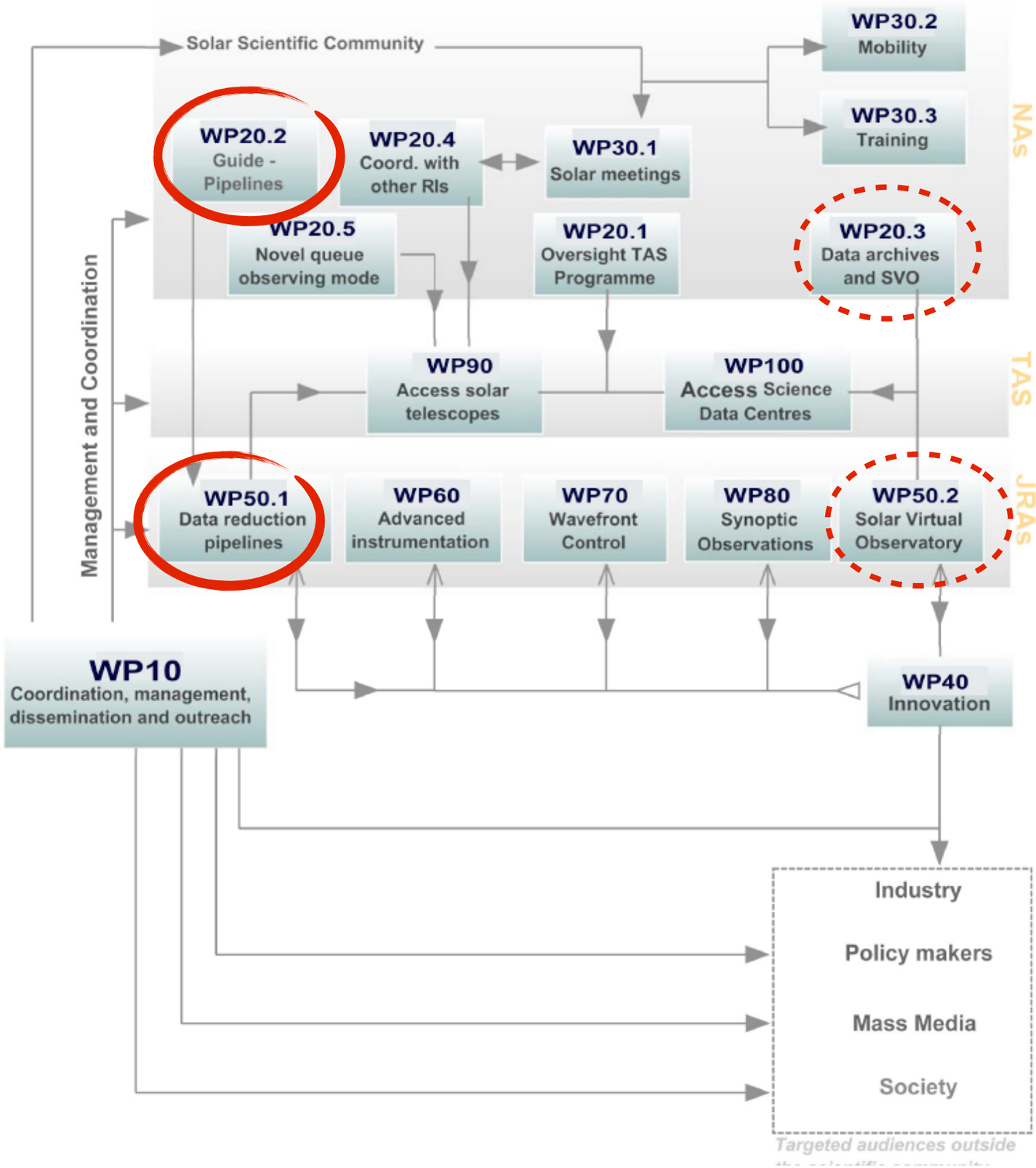
WP60. Advanced Instrumentation Development

Development of instrumentation to improve the existing solar telescopes and with possible application to the future large aperture solar telescopes. The instrumentation developments included in this WP are the following: large diameter FPIs (100 to 300mm), image slicer for 2D spectroscopy, microlens-fed spectrograph and Fast Imaging Polarimeter.

WP60.2 Image Slicers for 2D spectroscopy

SOLARNET Project

http://www.solarnet-east.eu/



WP20.3. Data archives and Solar Virtual Observatory.

The object of this task is to establish the rules for an adequate dissemination of data to the community [with a long list of actions]

- To identify the metadata needed in order to determine whether a dataset is relevant to a specific scientific investigation

- To define a subset of the available metadata that should be available to form search criteria in a virtual observatory/archive

A group composed by experienced partners (UiO –chair–, ROB, UCL-MSSL, INAF, IGAM, AISAS, HVAR, AIP, UPS, UWR, NSO and CfA-SAO) has been formed with the above goals

Challenges in data search from ground-based solar observations

Nazaret Bello González, KIS

Based on the

Document on Standards for Data Archiving and VO

SOLARNET WP20.3

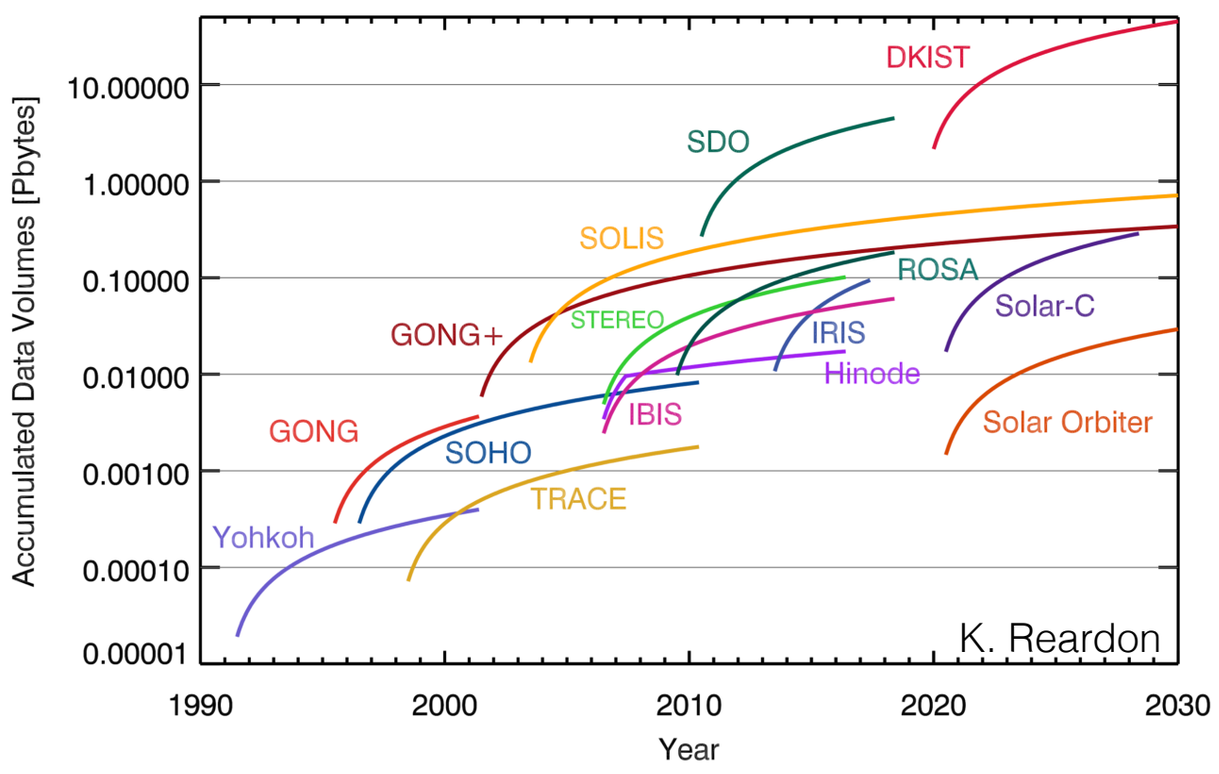
http://sdc.uio.no/open/solarnet-20.3/

Ground-based solar observations

Traditionally, solar observation archives and VOs have been used primarily to locate data from data sets that researchers have already known existed, namely from space-based solar observatories

Ground-based solar observations

However, the number of data sets available has grown, and will continue to grow as an increasing amount of data flow from ground-based observations are made and will be available



Nazaret Bello González, KIS

Ground-based solar observations

The use of multi-instrument analysis of solar phenomena has grown over the last decade, but the ability of SVOs to locate multi-instrument observations has **not** grown with it

Ground-based solar observations

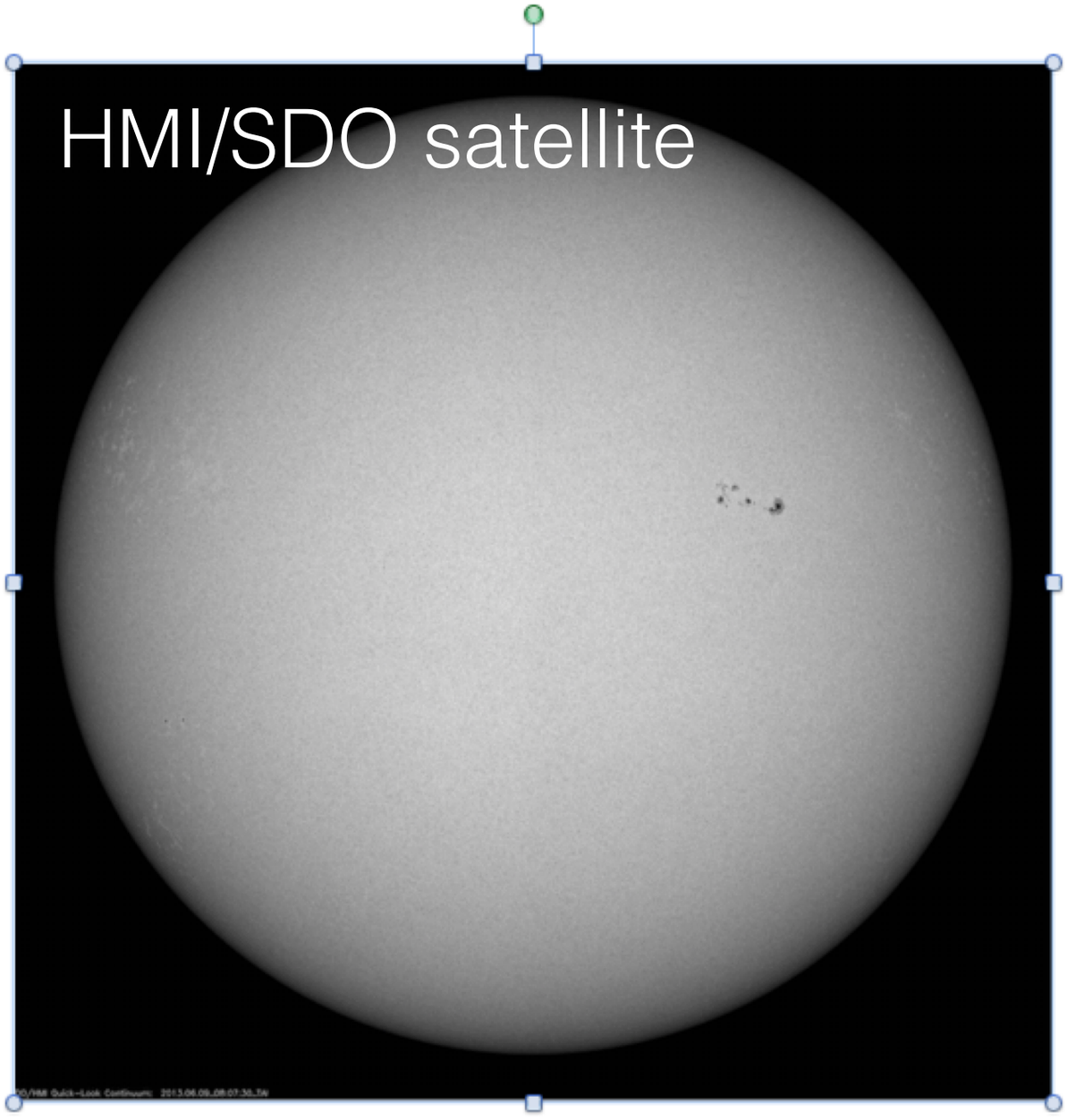
An ideal Solar Virtual Observatory (SVO) should be able to find sets of successful observations matching a hypothetical ideal observation proposal:

joint observations of specific targets/events from multiple instruments

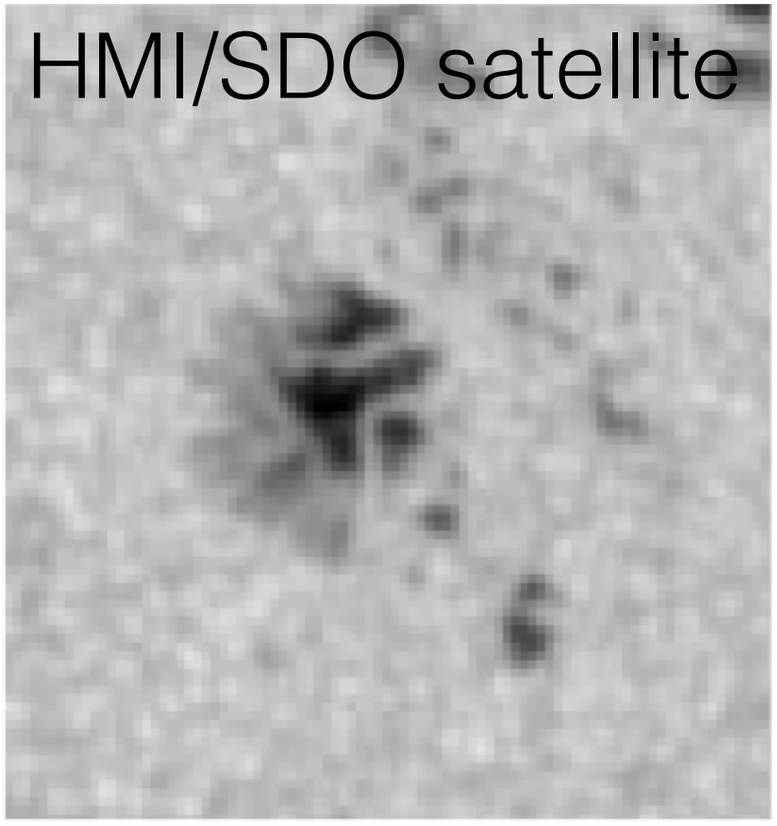
Such a scenario may even involve observations that do not overlap in time, e.g. solar disc observations of events vs. *in situ* observations of particles/shocks/interactions at a later time

Ground-based solar observations

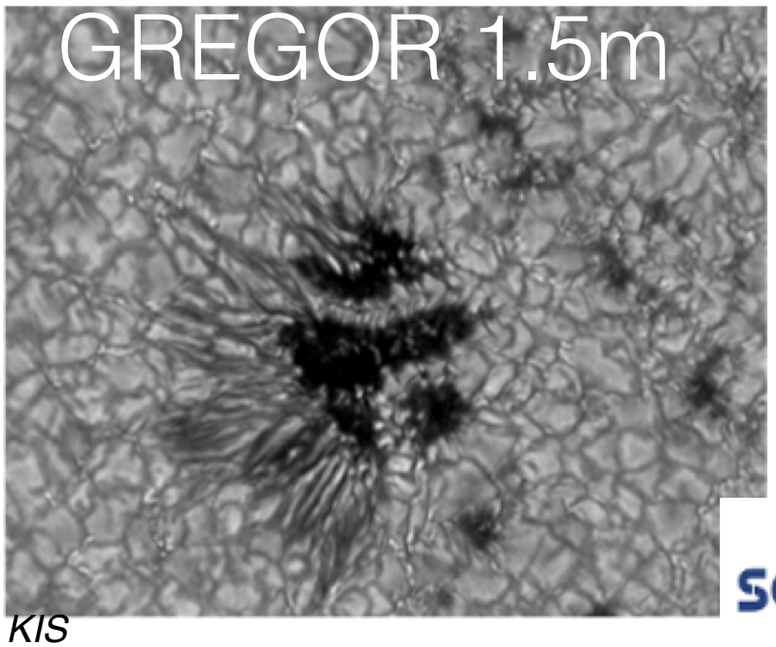
Characteristics of ground-based solar observations:



R. Schlichenmaier (KIS)



..



g

Ground-based solar observations

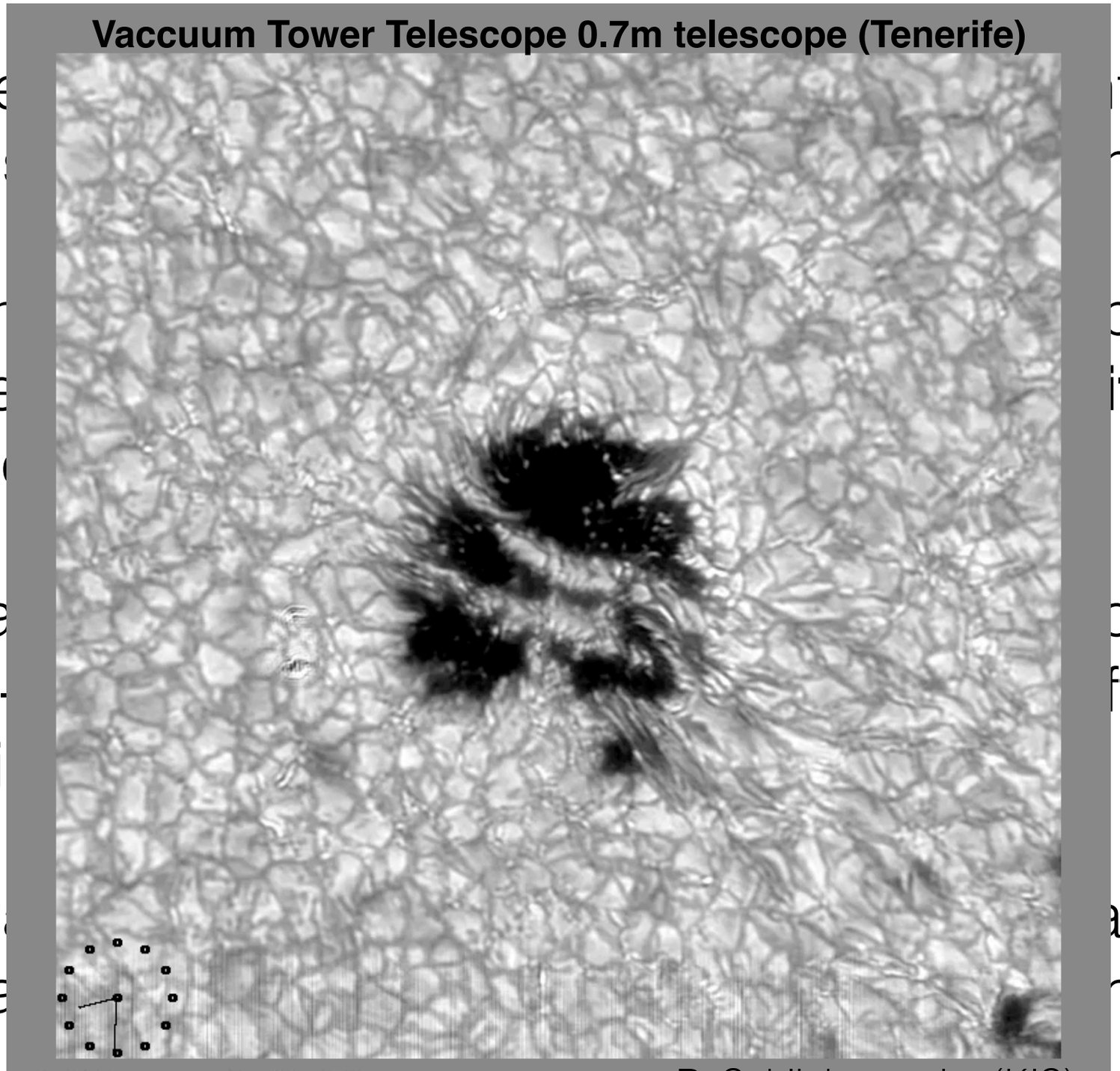
Characteristics of ground-based solar observations:

Limite t: quiet
Sun, nences,...

Seeir pped
frame ity/
succ

Versa bserving
runs - fficulty in
unifyi

Upgr a
chara nent



Ground-based solar observations

Characteristics of ground-based solar observations:

- Limited FoV (non-full disc) - Target dependent: quiet Sun, sunspots, pores, plages, faculae, prominences,…

- *Seeing* conditions, cadence variations, # dropped frames, polarimetric accuracy, etc. as a **quality/success** parameters

- Versatile observing modes: non-standarised observing runs - novel science (multi wavelength,..) - difficulty in unifying data pipelines

- Upgrade of instrumentation - changes in data characteristics for a given (upgraded) instrument

The ideal SVO

The ideal SVO should address:

1. Efficient presentation of search results
2. Visualisation: quick-look and movies, using external existing websites
3. Type of observations, targets and events must be identified
4. Instrument specific criteria: ideally, the archive should extract generic parameters matching specific criteria

Defining the output of data-pipelines

In order to fulfil the “vision” of an ideal SVO, it is necessary to ensure that the data to be served contains the necessary metadata

Standards for data archiving and SVO for ground-based observations

Nazaret Bello González, KIS

Standards for Data Archiving and SVO

GOAL:

Establish standards for adequate data dissemination

DIFFICULTIES:

(1) The standards should include clear descriptions of **data quality/success** criteria that may be used to select data

(2) It should be possible to develop **generic tools** for data visualisation and analysis

Standards for Data Archiving and SVO

There are numerous projects with SVO-like characteristics and goals (e.g. VSO, HELIO, Helioviewer, and the Hinode archive)

SOLARNET tries to take advantage of the lessons learned by these

In addition, we also try to maximise the ability of a future SVO to take advantage of existing utilities

Standards for Data Archiving and SVO

However, inherent differences and inconsistencies between existing pipeline outputs, services, and SVOs make this very hard through a simple merging of the existing practices

Instead, we concentrate on using the most generally accepted FITS standards (e.g. the World Coordinate System) to describe the physical aspects of the data as accurately and exhaustively as possible

Standards for Data Archiving and SVO

There must be translation routines that make legacy data compatible with SVOs that depend on SOLARNET recommendations

This would be done *without* modifying the original files, but rather by “synthesising” a set of metadata compatible with the SOLARNET recommendations, in order to ingest the data in an SVO

Standards for Data Archiving and SVO

For new pipelines, there must be a convergence towards:

1. Including as much as possible of the metadata of SOLARNET recommendations

2. Excluding superfluous and sometimes ill-defined keywords

3. Using only the SOLARNET definitions of all SOLARNET keywords

Standards for Data Archiving and SVO

SOLARNET recommendations on File Format:

Based on common practice in the solar community, SOLARNET highly recommends using the **FITS** file format for data dissemination

SOLARNET addresses how to include the metadata content through keywords inside FITS files, but *that does not preclude the use of other file formats*

As long as the requirements for the metadata information content are met, an automated translation between the representations in different formats is achievable

Standards for Data Archiving and SVO

SOLARNET recommendations on *File Format*:

1. File name recommendations
2. FITS File Header and Data Units (HDUs)
3. Storing data in a single file or in separate files

Standards for Data Archiving and SVO

SOLARNET recommendations on the
physical description of observational data:

1. Use of World Coordinate System (WCS) and related keyword
2. WCS positional keywords and relative radial velocity
3. Exposure time, binning
4. Instrument/data characteristics etc. (wavelength band, resolution power, use of adaptive optics,..
5. Quality aspects
6. Data statistics
7. Missing and saturated pixels, spikes/cosmic rays

Standards for Data Archiving and SVO

Quality aspects:

Many quality aspects of ground-based observations change rapidly, even from one exposure to the next

Until now, there has been little effort in order to characterise quality aspects of ground-based observations in a manner that is *consistent* between different telescopes, and even between different setups at the same telescope

Example: In FITS files from ESO (European Southern Observatory), the keyword **PSF_FWHM** is used to give the full width at half maximum in arc seconds for the point spread function. However, this quantity is generally not available for solar observations. Some adaptive optics systems, however, may record parameters like the atmospheric coherence length r0. If available, the value of r0 should be stored in the keyword **ATMOS_R0**

Standards for Data Archiving and SVO

SOLARNET recommendations on metadata on
the origin of the observations

1. From where to how, PROJECT to SETTINGS/OBS_MODE
2. Who and why? (“free text”)
   - OBSERVER
   - PLANNER
   - CAMPAIGN

Standards for Data Archiving and SVO

SOLARNET recommendations on

- Grouping
- Pipeline processing information applied to the data
- Integrity and administrative information
- Reporting of events detected by the pipeline/ spacecraft
- Other keywords/rules

The European SVO - A prototype

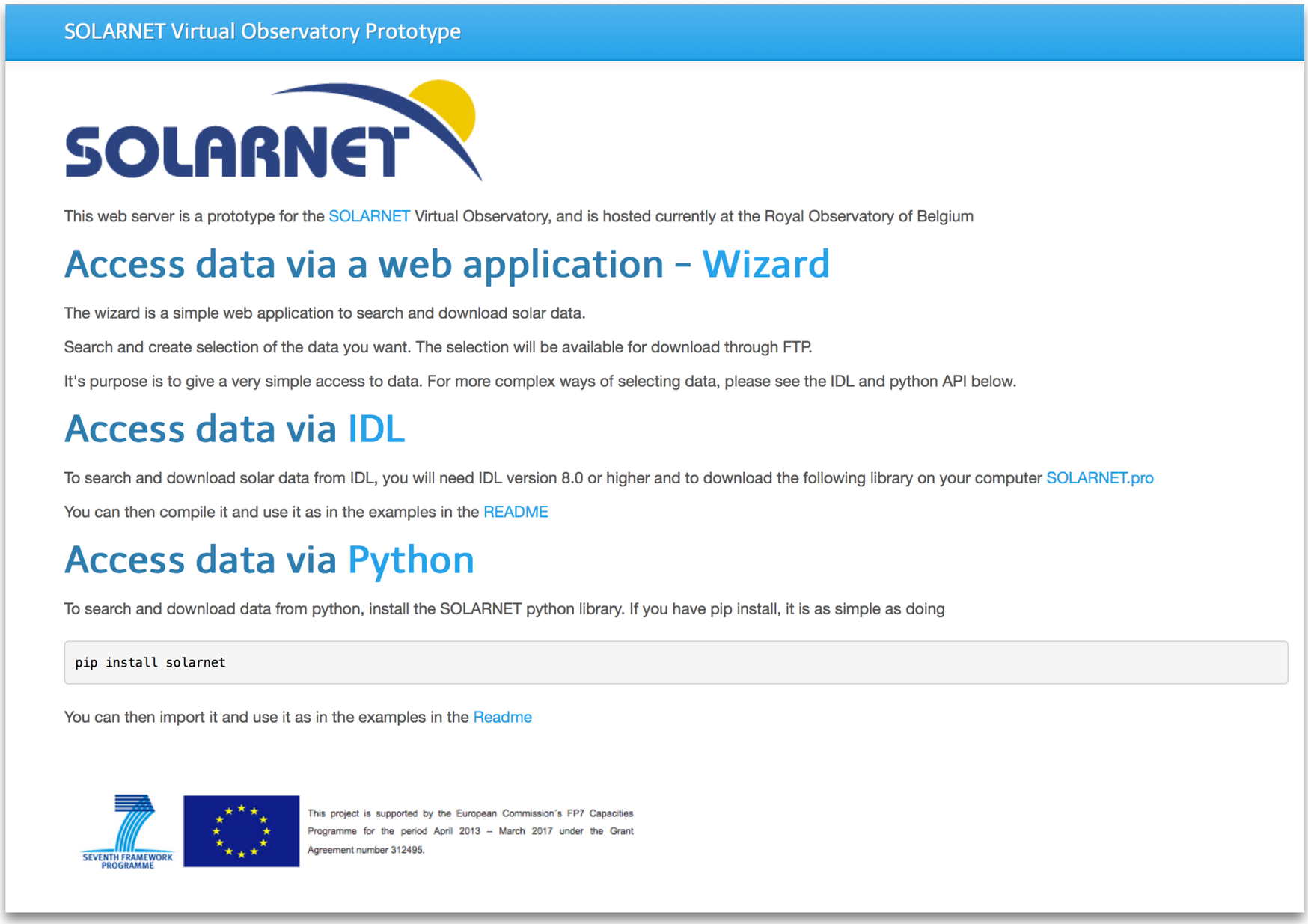
Nazaret Bello González, KIS

The European SVO - A prototype

The SOLARNET VO needs to, from a user point of view, enable optimal data retrieval to enable scientific exploration. It should interfaces with the (US) VSO and offer data from ground-based and space-based observatories

The European SVO - A prototype

http://solarnet.oma.be/



SOLARNET Virtual Observatory Prototype

SOLARNET

This web server is a prototype for the SOLARNET Virtual Observatory, and is hosted currently at the Royal Observatory of Belgium

Access data via a web application – Wizard

The wizard is a simple web application to search and download solar data.

Search and create selection of the data you want. The selection will be available for download through FTP.

It's purpose is to give a very simple access to data. For more complex ways of selecting data, please see the IDL and python API below.

Access data via IDL

To search and download solar data from IDL, you will need IDL version 8.0 or higher and to download the following library on your computer SOLARNET.pro

You can then compile it and use it as in the examples in the README

Access data via Python

To search and download data from python, install the SOLARNET python library. If you have pip install, it is as simple as doing

```
pip install solarnet
```

You can then import it and use it as in the examples in the Readme

SEVENTH FRAMEWORK PROGRAMME

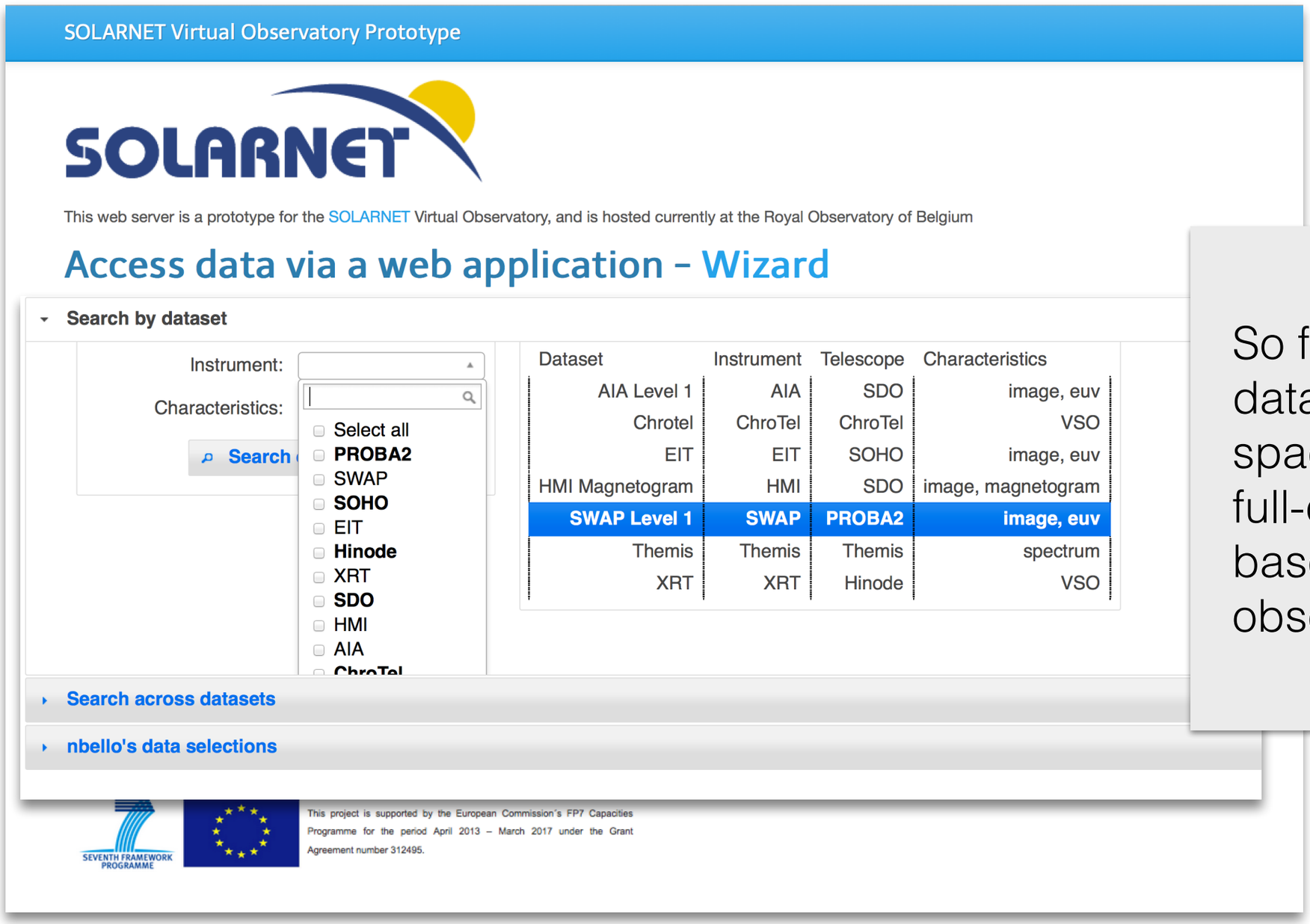
This project is supported by the European Commission´s FP7 Capacities Programme for the period April 2013 – March 2017 under the Grant Agreement number 312495.

SOLARNET WP lead by the Royal Observatory of Belgium (Brussels)

Nazaret Bello González, KIS

The European SVO - A prototype

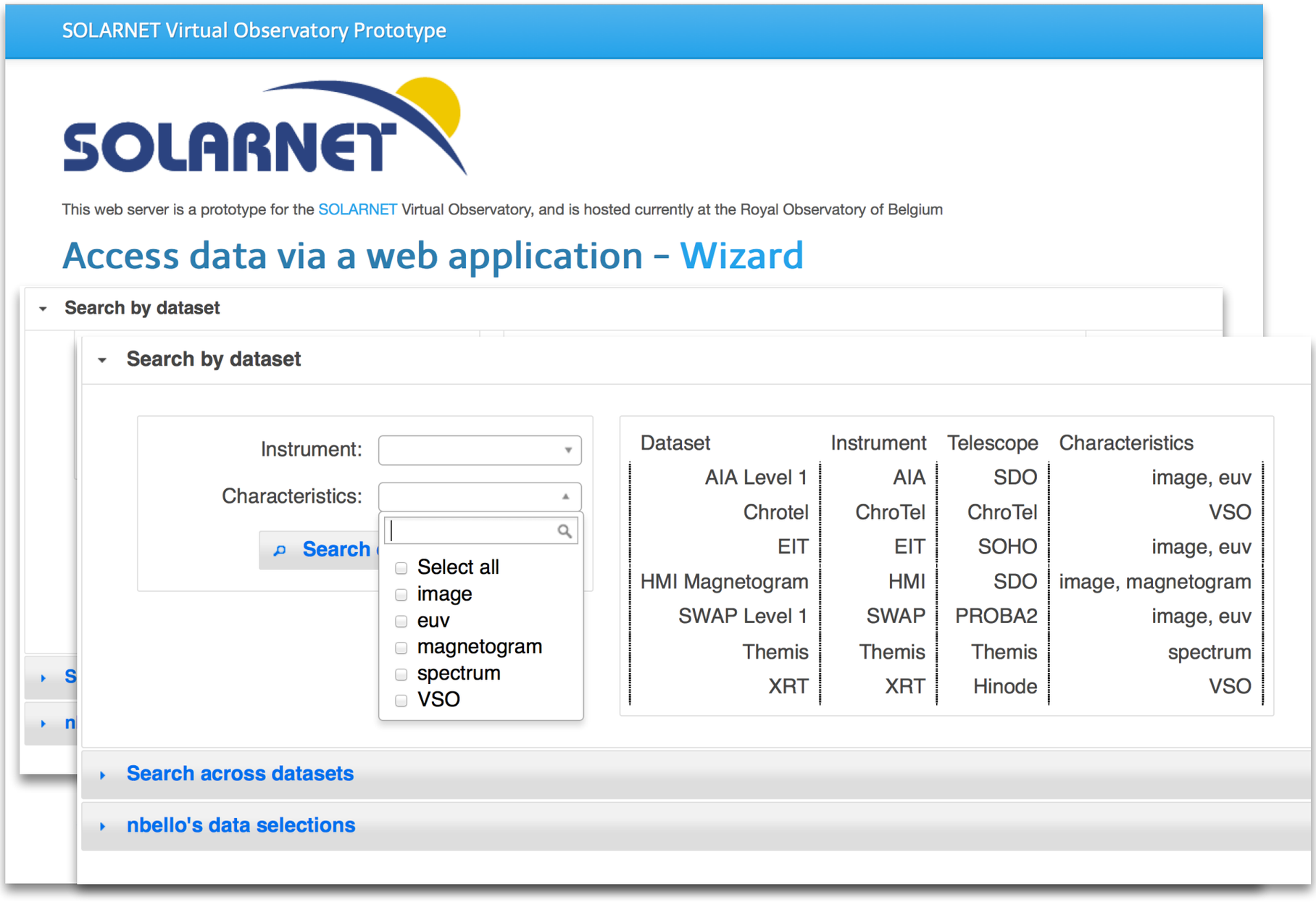
http://solarnet.oma.be/



So far contains data from space-based and full-disc ground-based observatories

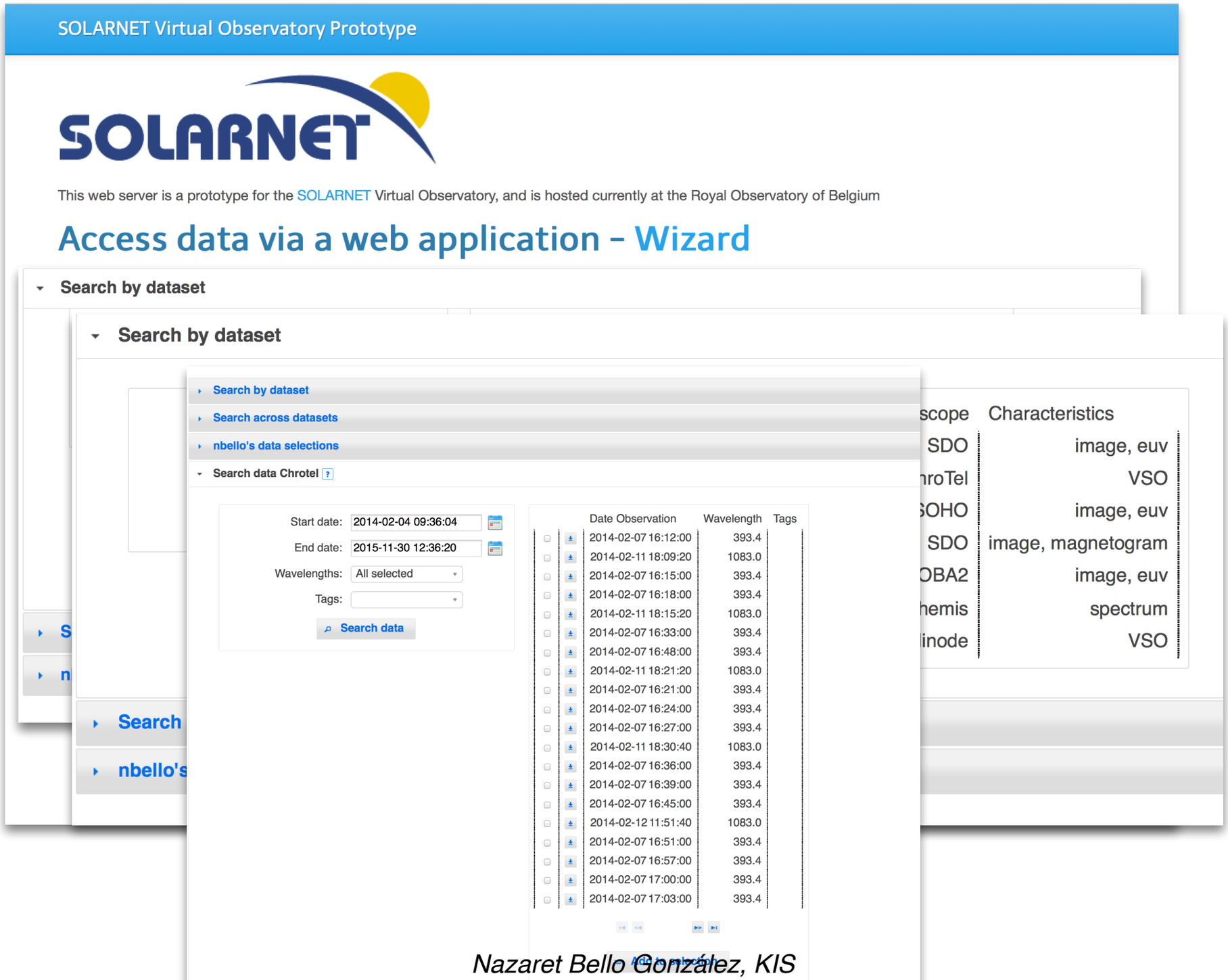
The European SVO - A prototype

http://solarnet.oma.be/



The European SVO - A prototype

http://solarnet.oma.be/



Nazaret Bello González, KIS

Outlook

How can EST be integrated in and profit from ASTERICS?