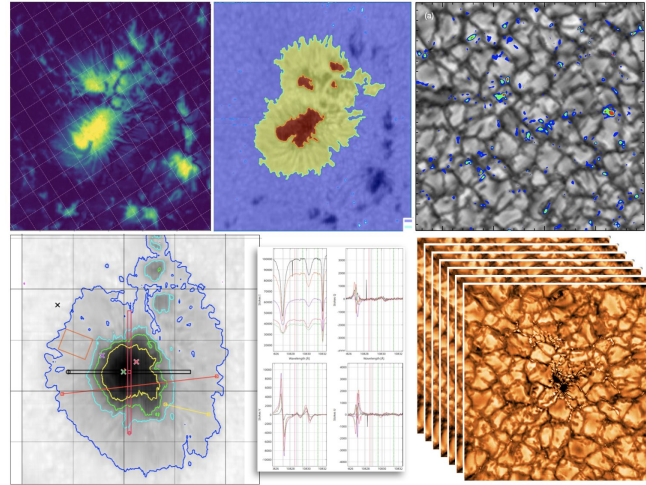# The KIS Science Data Centre (SDC)

The SDC Team

J. Beck, N. Bello González, S. Berdyugina, A. Bührer, P. Caligari, A. Diercke, V. Gangadharan,

I. Gorbachev, A. Gorobets, M. Günter, A. Hochmuth, L. Hohl, P. Kehusmaa, M. Knobloch, M. Franz*,

S. Patel*, C. Schaffer* & T. Yakobchuk

Leibniz-Institut für Sonnenphysik (KIS), Freiburg

# KIS solar observatories and large instrumentation

- Observatorio del Teide, Tenerife:

- **GREGOR:** 1.5 m telescope
- **VTT**: 70 cm Vacuum Tower Telescope,
- **ChroTel:** 10 cm full-disc Chromospheric Telescope

- 4 m Daniel K. Inouye Telescope, **DKIST** (NSO, Maui)

KIS contribution to DKIST:
Visible Tunable Filter (**VTF**): spectro-polarimetric imager

DKIST/DC and SDC/KIS plan to expand
their collaboration (LoI signed in 2018)
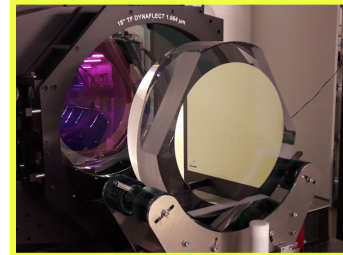
- 4 m European Solar Telescope, **EST** ESFRI infrastructure to be build in the Canary Islands.
KIS as co-founder of **European Solar Telescope Foundation** (July 2023)

KIS planned contribution to EST:
Participation in developing the **EST Data Center** & a VTF-like instrument (**TIS**)

VTF Etalon plate

The KIS Science Data Centre – Introduction

The Science Data Centre (SDC, https://sdc.leibniz-kis.de) is a service launched by the Leibniz-Institut für Sonnenphysik (KIS) with a two-fold main purpose:

1. To provide a common platform for the solar community to
   STORE   ACCESS   ANALYSE   ARCHIVE   DISSEMINATE
   solar data produced by a heterogeneous group of scientific instruments.

2. The development of
   - new diagnostic tools, e.g. stochastic analysis of fluctuations in physical parameters
   - data science, e.g., research on statistical properties from solar data all over the archive
   - high-level data products, e.g., automatic inversions run over all archived data

The KIS Science Data Centre – Collaborations

SDC Cooperation with EU H2020 & DE projects

SOLARNET
- Coordination of the WP5 *Towards a European Science Data Centre*
- Participation in the Virtual Access Programme (WP10) with the SDC archive

ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research infrastructures

ESCAPE Collaboration Agreement is intended to bring all ESFRI projects data like EST into a common framework and towards the European Open Science Cloud (EOSC)
- Participation on WP4 integrating ground-based solar data in the Astronomical VO
- Participation on the ESFRI Science Analysis Platform (WP5) including high-level ground-based solar data tools (ESCAPE H2020 ended in Jan 2023)

PRE·EST
- Coordinator of the EST Data Center concept design

PUNCH 4NFDI
- Representing the solar community as participant in the German PUNCH4NFDI Consortium of the particle, astro-, astroparticle, hadron and nuclear physics community in Germany

PUNCH aims to develop a common German infrastructure that brings together all German data centers and data in this broad field into a common framework (implementing FAIR principles, common data lake,…), similarly to the ESCAPE goals at European level

AG Meeting 2023 - EScience, Machine Learning and Virtual Observatory
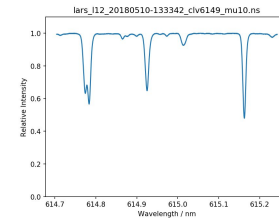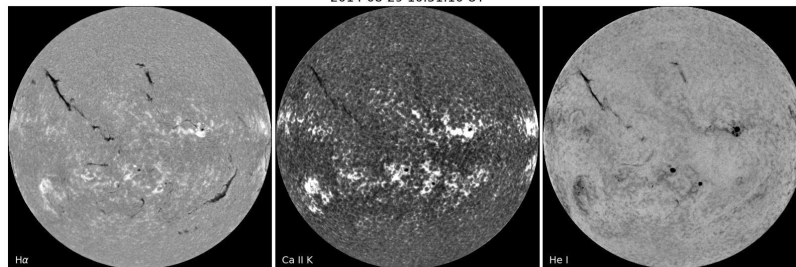
SDC data archive and dissemination

Current Structure of the Data Archive

o Central archive for 3 instruments: GRIS/GREGOR, LARS/VTT & ChroTel
o BBI integration planned for this year
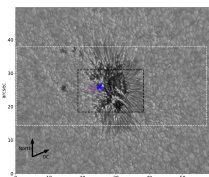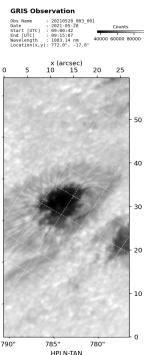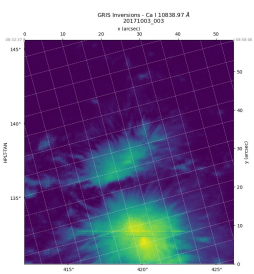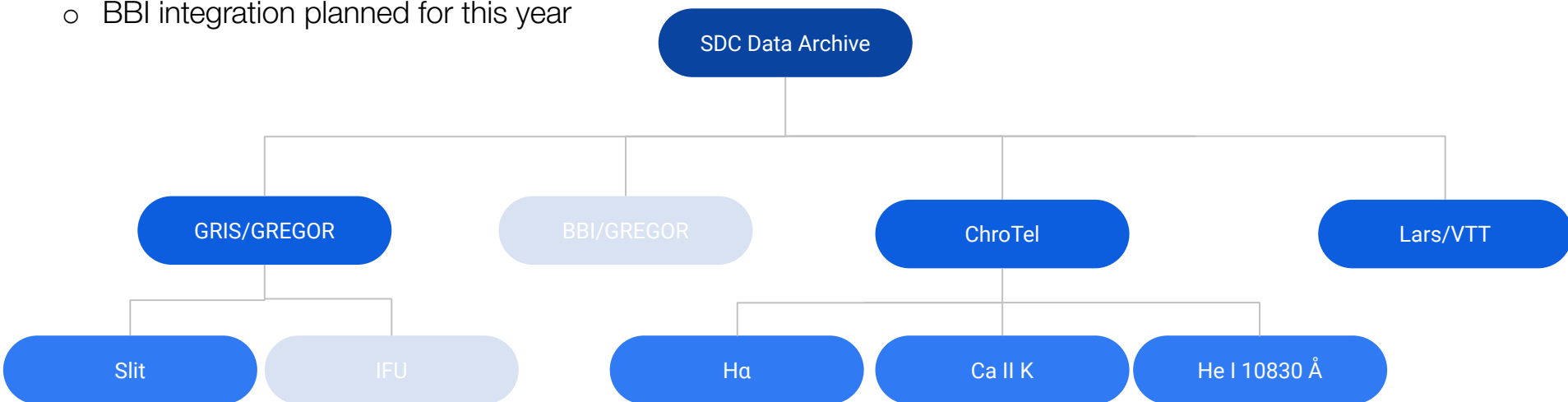
Lindner+2023

KIS Science Data Centre Archive

- The SDC archive web allows to query data through date, wavelength, heliocentric angle, target,..
- archive.sdc.leibniz-kis.de: web interface for data preview and download
- FITS data compliant with SOLARNET and IVOA* metadata standards
- 204k observations
- Access to calibration resources (GitLab)
- Help Centre
- Data availability calendar
- Jupyter notebooks launched via ⊗ binderhub * service allow to explore GRIS datasets on the cloud directly from the SDC archive page (for GRIS data)
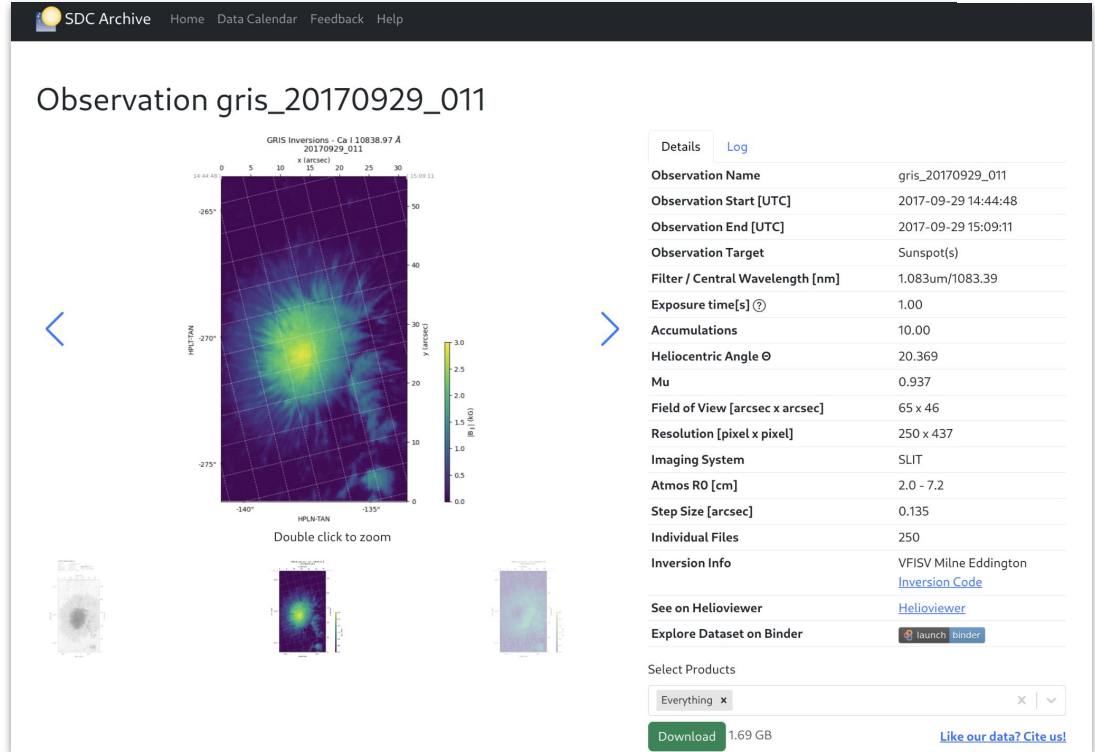- Link to access the selected target on Jhelioviewer

*BinderHub is a cloud-based service that allows users access to a sharable, interactive, and reproducible computing environment to execute and modify a chosen workflow or create their own workflow.

SDC archive dissemination – Data publicly available
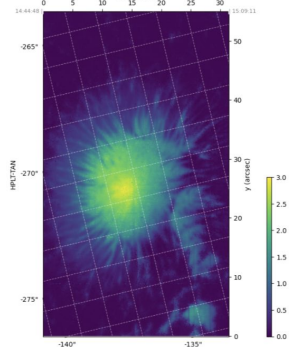
GRIS-slit/GREGOR instrument:

- o GREGOR Infrared Spectropolarimeter: Collados+2012; operated in the 1.56μ and 1.08μ (HeI) spectral windows

- o Data available from 2014-onwards

- o Upcoming soon: GRIS-IFU data

SDC archive dissemination – Data publicly available

GRIS-slit/GREGOR instrument:

o **Science-ready (Stokes parameters) data**:
data calibrated applying M. Collados' (GRIS PI, IAC) calibration routines

o **VFISV (Borrero+2011) M-E inversion results**:
Magnetic field vector and LOS velocities

o **HMI context information**: location of target on HMI intensity maps included

Science-ready data

Stokes I   Stokes Q   Stokes U   Stokes V
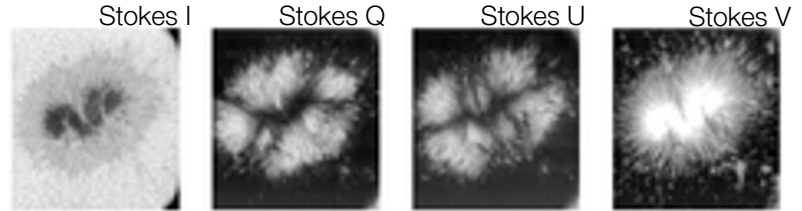
M-E inversion results
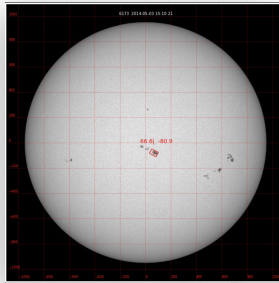
Vertical component of the magnetic field
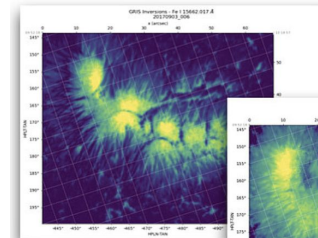
Horizontal component of the magnetic field

LOS velocities

Context info
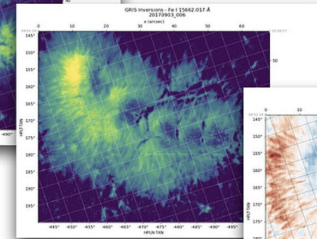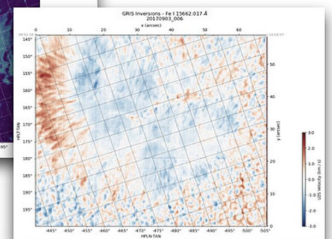
SDC archive dissemination – Data publicly available

LARS/VTT instrument:
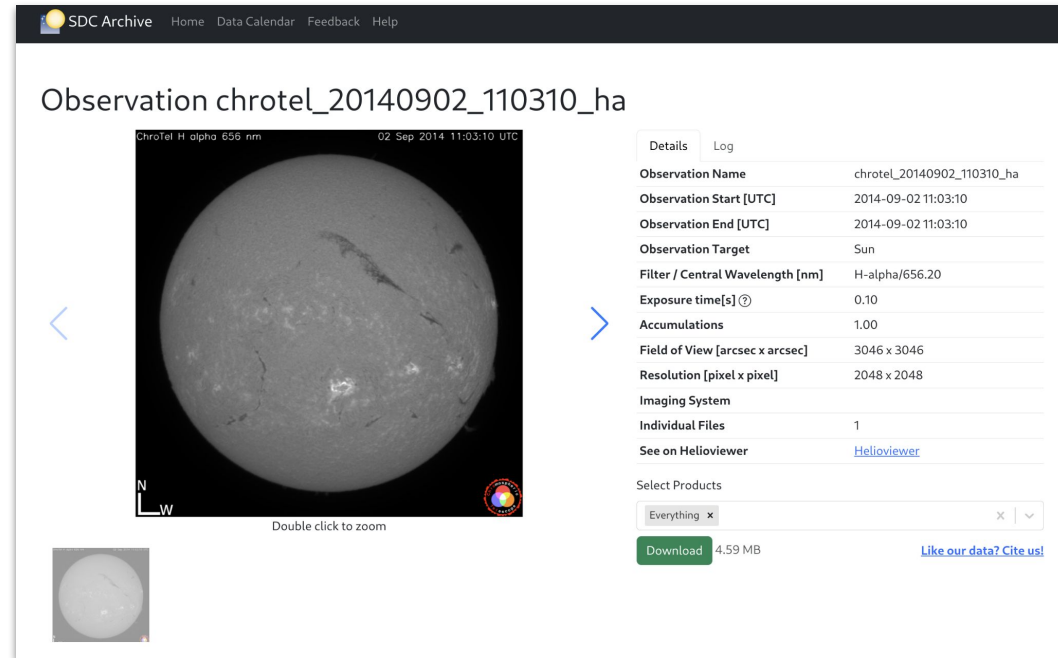
- Laser Absolute Reference Spectrograph:
  Doerr (2016) PhDT; Löhner-Böttcher+2017;

- Single-mode fiber fed instrument for precision solar spectroscopy in applications that require a high signal-to-noise ratio and an accurate wavelength calibration @520nm and 630nm spectral windows

- Science-ready (spectroscopic) data:
  data calibrated applying
  Doerr+Löhner-Böttcher's calibration routines

- Limited datasets: 2016-2018

- HMI context information: location of target on HMI intensity maps included
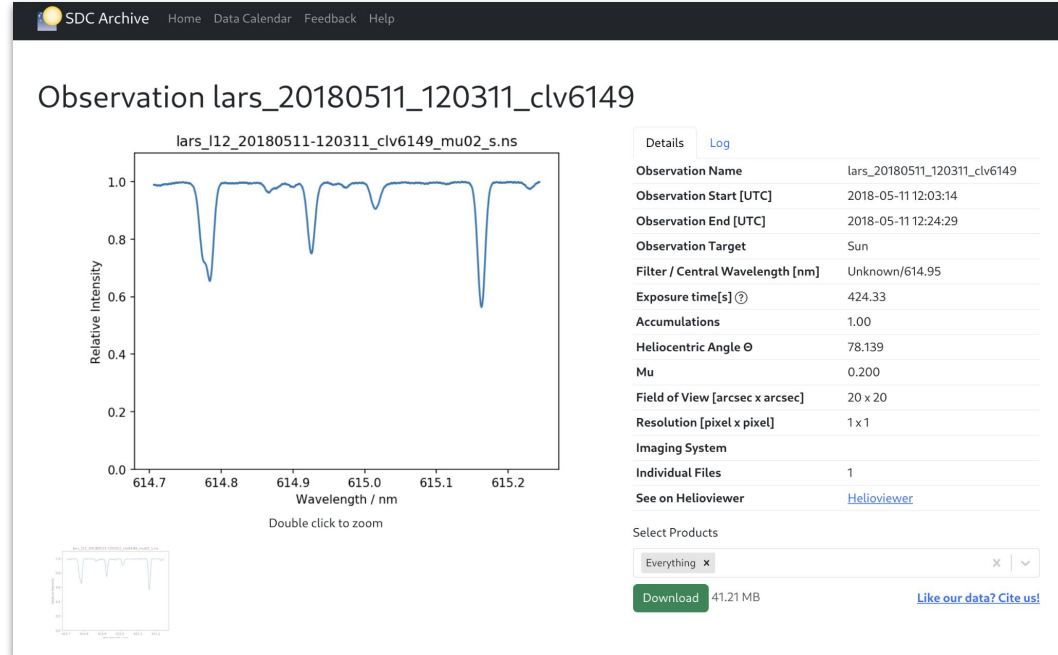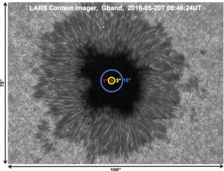
SDC archive dissemination – Data publicly available

ChroTel data:

- The Chromospheric Telescope:
  (Kentischer+2008; Halbgewachs+2008; Bethge+2011) is a full-disc imager of the solar chromosphere in CaII K 393nm, Hα 656nm, and HeI1083nm

- Science-ready data:
  data calibrated applying Ch. Bethge's calibration routines

- Continuous datasets from 2012-2020

see Poster Nr. 30: *Counter-streaming flows during a solar filament eruption with ChroTel data*

SDC archive dissemination – Data access

o SDC archive web – allows for browsing through data

o APIs:
An application programming interface (API) has been set up that enables programmatic access to the data archive for external users. Currently, the access is provided via a RESTHeart API connected to the MongoDB database.

The service is accessible at https://restheart.sdc.leibniz-kis.de.

Applications can be built against this API for easy access to our archive.
In collaboration with Aperio Software Ltd.
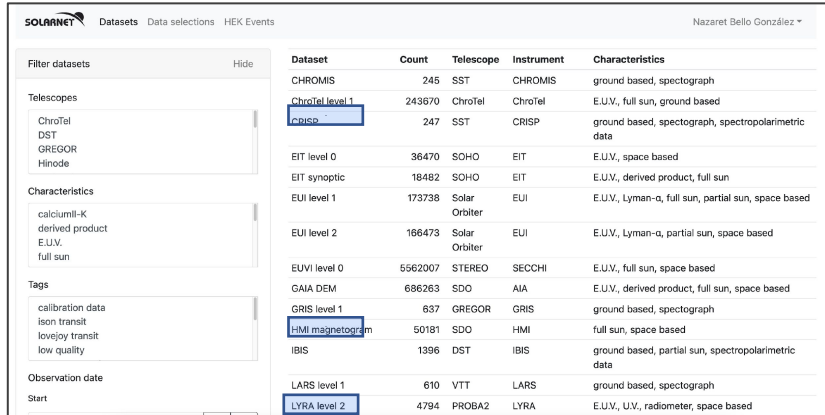
o EPN-TAP:
The SDC archive data is also accessible via EPN-TAP services at the end-point:
http://dachs.sdc.leibniz-kis.de/tap.

Our TAP end-point is registered with the IVOA Registry of Registries and is searchable via the various VO networks/clients (e.g TOPCAT, Aladin, ESAP) extensively used by the broader Astronomical community

SDC archive dissemination – Data access

○ Via the SOLARNET Virtual Observatory

The SDC archive data can be accessed via the SOLARNET VO which offers the possibility to cross-search data of various solar observatories

○ Via the ESCAPE ESAP platform

The SDC archive data can be now also accessed by the broader astro-particle and astronomical community through the ESFRI Science Analysis Platform developed within the framework of the ESCAPE H2020 project

NOTE: We are shifting the access of GRIS data to the EPN-TAP service at the moment -> GRIS Level 2 data will become available in the SVO soon

NOTE: All activities within ESCAPE H2020 and beyond are intended to bring all ESFRI projects data – including EST-- into a common framework and towards the European Open Science Cloud (EOSC)

SDC Back-end

adapted from C. Schaffer (KIS)

https://archive.sdc.leibniz-kis.de/
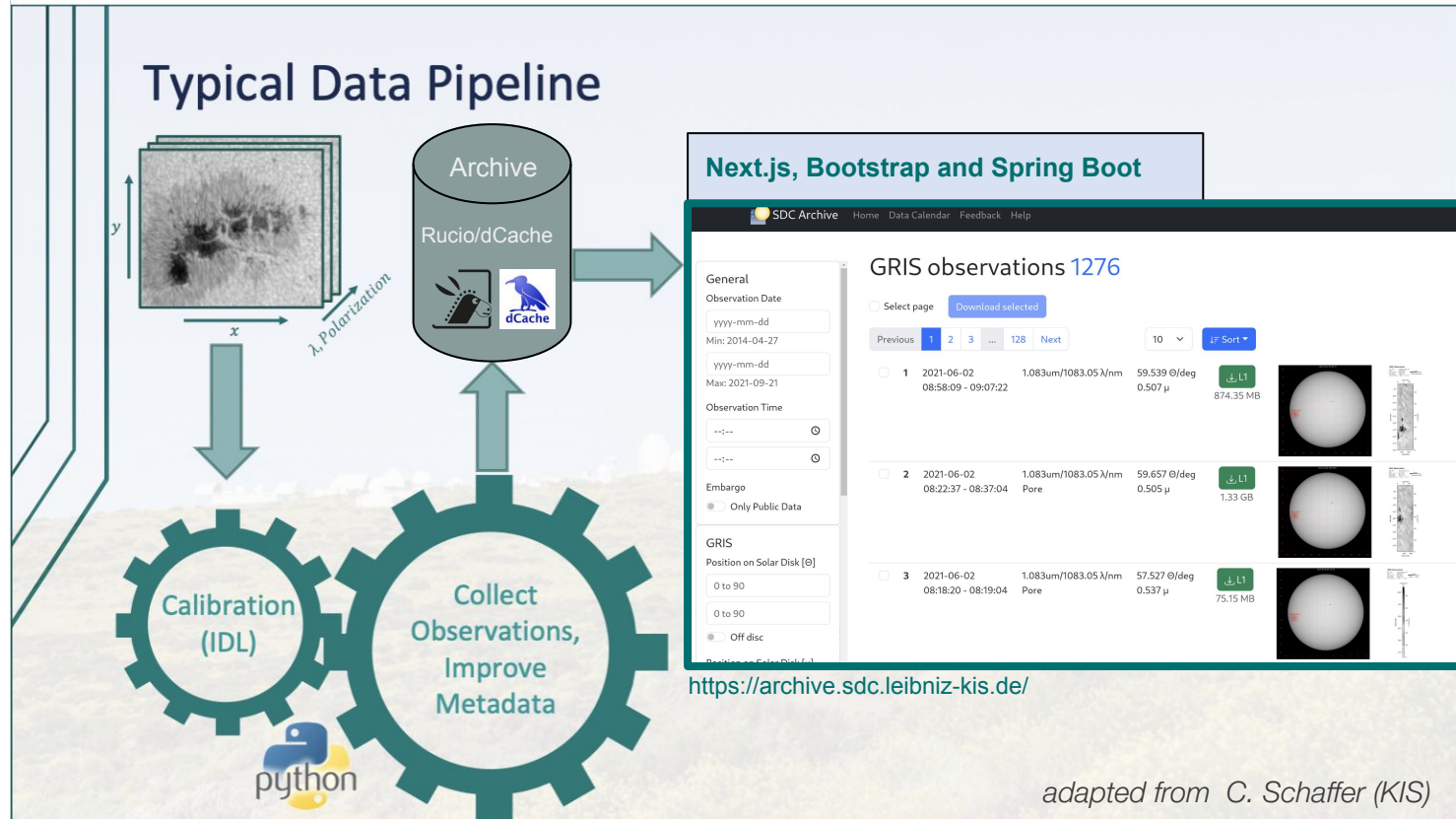
SDC Back-end

o SDC scales to multiple sites
  - OT already is an edge (in a sense)
  - Could have more sites similar to KIS

o One data lake managed by Rucio

o Standardized Software & Pipelines

o Open Source!

o Requirements are similar throughout astronomy

o Metadata standards implemented

Metadata

gris_observations

_id: ObjectID(name_uid)
metadata: dict
header: dict (IVOA)
tarsize_l1data: int
data_quality: int
campaign_id: str
embargo_until: datetime
links: dict
type1: list of DBRef
...
typeN: list of DBRef
lastModified: datetime

gris_data

_id: str (DID)
filename: str
length: int
metadata: dict
date: datetime
type: str
header: dict (opt, SOLARNET)
md5: str
mimeType: str
uploadDate: datetime

gris_auxdata (GridFS)

_id: ObjectID
filename: str
length: int
chunkSize: int
metadata: dict
date: datetime
type: str
header: dict (opt)
md5: str
mimeType: str
uploadDate: datetime

Grouping: *observations*

- kept in MongoDB
- Corresponds to Rucio *container*
- Header: IVOA
- Used for VOs, e.g.: https://solarnet.oma.be
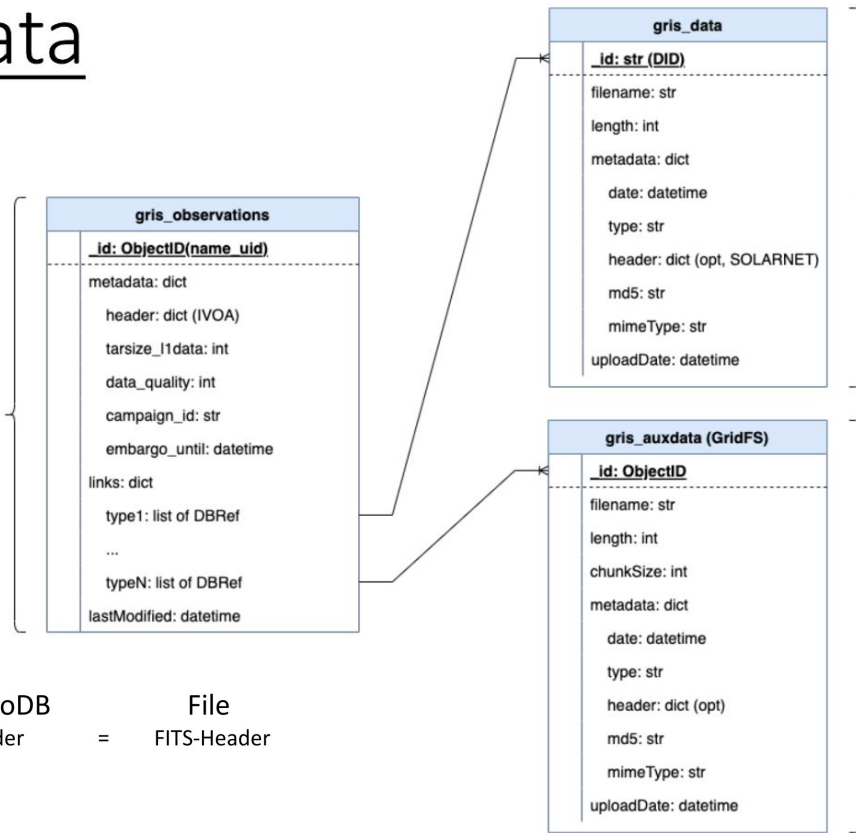
Science-ready data

- kept in Rucio (file)
- Versioned!
- FITS-files[2] (contain SOLARNET conform header[1])
- Subject to embargoes
- Header: extended FITS header
- Eventually associated to DOIs

Volatile data
(previews, log-files, calibration files, etc.)

- Kept in MongoDB / GridFS
- **No** science-ready data
- Used on: https://archive.sdc.leibniz-kis.de
- One (latest) version, only!

Rucio DID   MongoDB   File
               Header   =   FITS-Header

Rucio MD plugin
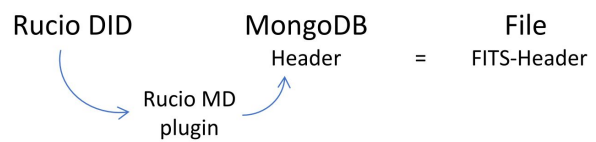
[1] https://arxiv.org/abs/2011.12139.
[2] https://fits.gsfc.nasa.gov/fits_documentation.html

SDC GRIS Data Exploration and Analysis – GRISview

GRISView - GRIS data explorer and analysis tool

o **Explore** GRIS data from the archive

o **Analysis** of **Stokes profiles and maps** and/or the **inversion results** including the fitted profiles

o **Figures** created with GRISView can be **saved**, re-worked and downloaded

o Written in **Python**

o **GUI based** on PyQt framework and PyQtGraph package

o **Open-source** and uses community-developed open-source packages AstroPy and SunPy

o Support for **all GRIS data** products: science-ready and inversion results

o **Fast, interactive** and **user-friendly**

o **Planned**: support for other instruments

Contact: yakobchuk@leibniz-kis.de
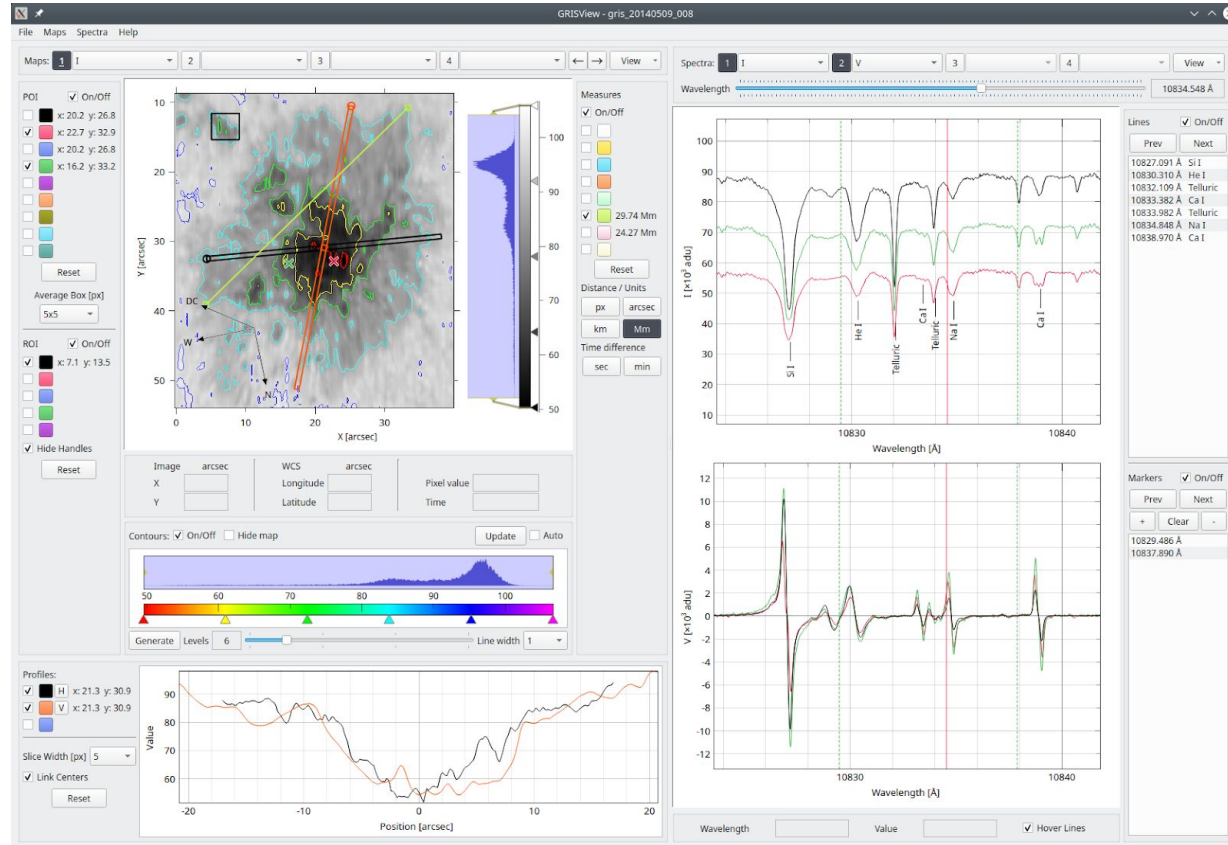
SDC Data Analysis

Stochastic diagnostic of fluctuations of observables
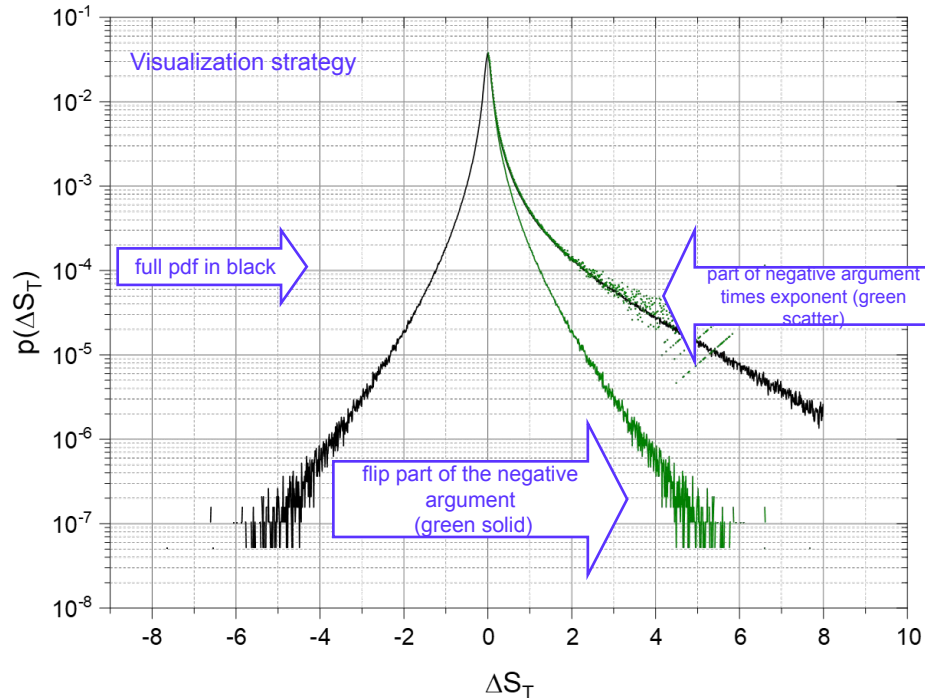
Figure1. Experimental verification of DFT.

$$\frac{p(\delta s_T > 0)}{p(\delta s_T < 0)} = \exp(|\delta s_T|) \quad (DFT)$$

- **Fluctuations** sequences are characterized by a single parameter δs_T – stochastic entropy production

- Positive and negative entropy production are connected via an exact relation known as *detailed fluctuation theorem* (DFT). The case of the quiet Sun HMI data is explored

- Stochastic entropy productions reveals new properties of the observables ☐ See Poster by Gorobets & Berdyugina

Contact: gorobets@leibniz-kis.de

Gorobets, A.Y., Berdyugina, S.V., 2019, MNRAS Lett., 483, L69

14/09/2023

New ChroTel data products

Before   After

○ **Amelioration of ChroTel datasets.** Shen+2018 made use of Zernike polynomials to approximate intensity variations in ChroTel images introduced by the Lyot filters. This led to an evident improvement of the datasets affected by these intensity variations.

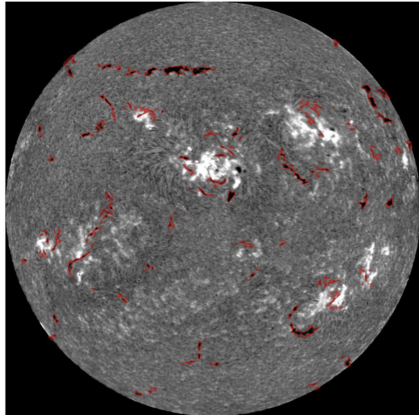o These techniques will be applied to all archived ChroTel data actually off-line and will be included in the SDC archive
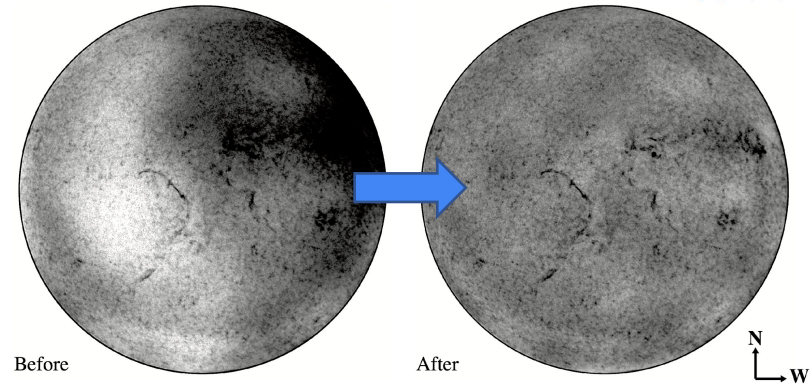
ChroTel 2013-01-14 12:36:10 UT

ChroTel 2018-11-10 14:48:10 UT
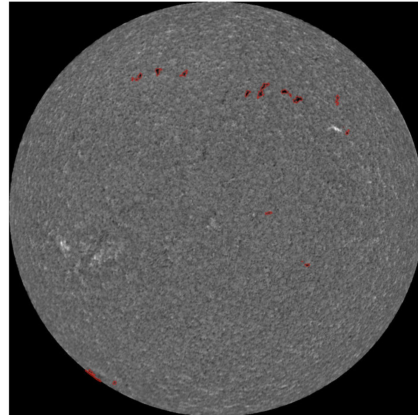
○ **ChroTel high-level data products.** Diercke+2023 (in prep.) have successfully applied segmentation techniques with DNNs to detect chromospheric filaments in H-alpha I images.

These high-level data products will become available to the scientific community through the SDC archive access paths

Contact: diercke@leibniz-kis.de

High-level data products:
Inversions using NNs – First tests

- GRIS (M-E) inverted Stokes profiles are used to train a Convolutional Neural Network (CNN)

- as first basic tests to infer the magnetic field parameters and LOS velocities from data with NNs trained on inversion results

- CNN performs on the entire 2D map at once

original image · norm. pred. · denorm. pred. · correlation orig. - pred.

Contact: gorbachev@leibniz-kis.de
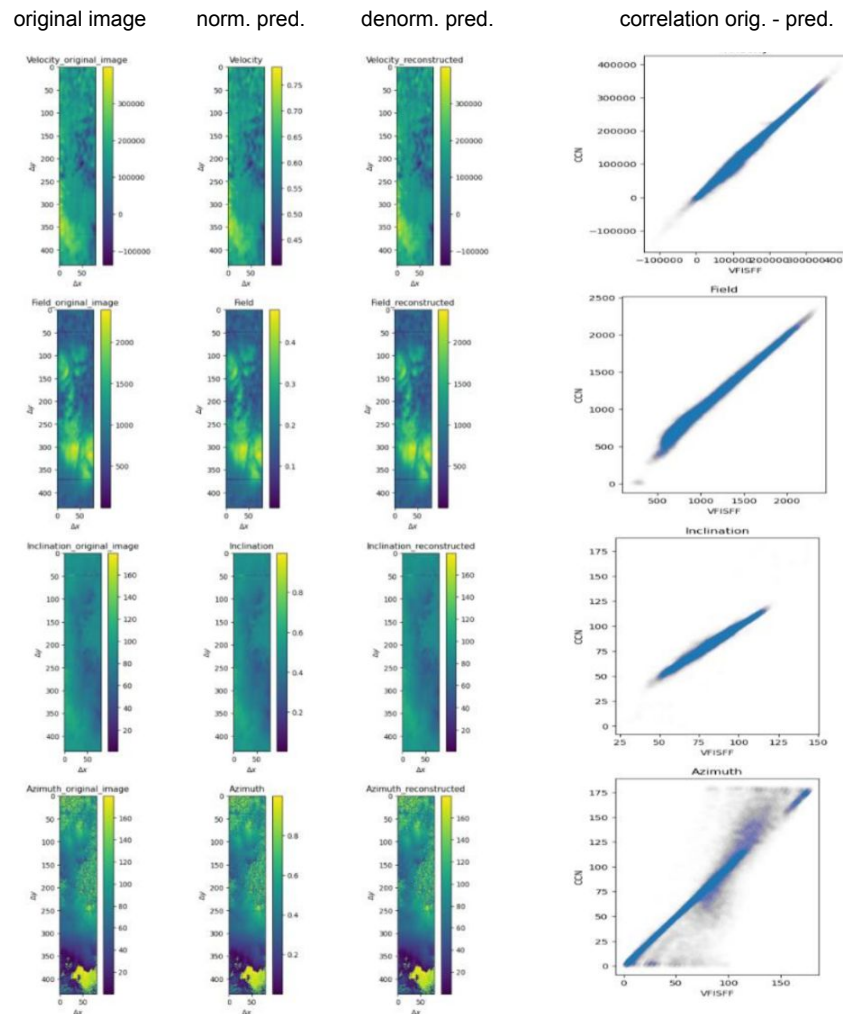
Conclusion and call to data providers

The KIS Science Data Centre

On 15 Nov 2021, SDC platform was launched as a service for the solar community
https://sdc.leibniz-kis.de

During the last years, we have been building a solid infrastructure foundation for future needs meaning flexibility and scalability for large data amounts, software development and computing power

We warmly welcome your feedback and suggestions on how to improve the platform

The KIS Science Data Centre

On 15   ommunity
https:

During   ion for
future   ware
devel

We wa   he
platfo

SDC call to Data Providers:

Those interested in the storage and dissemination of solar
instrument/observatory/simulations data, please, contact us:

sdc-team@leibniz-kis.de

Planned data providers:
DKIST data and Ondrejov Observatory 'solar patrol' flare data