Unlocking Efficiency : Python-powered DOI Creation.

NRF
National Research Foundation

SARAO
South African Radio Astronomy Observatory

Sydil & Zanele
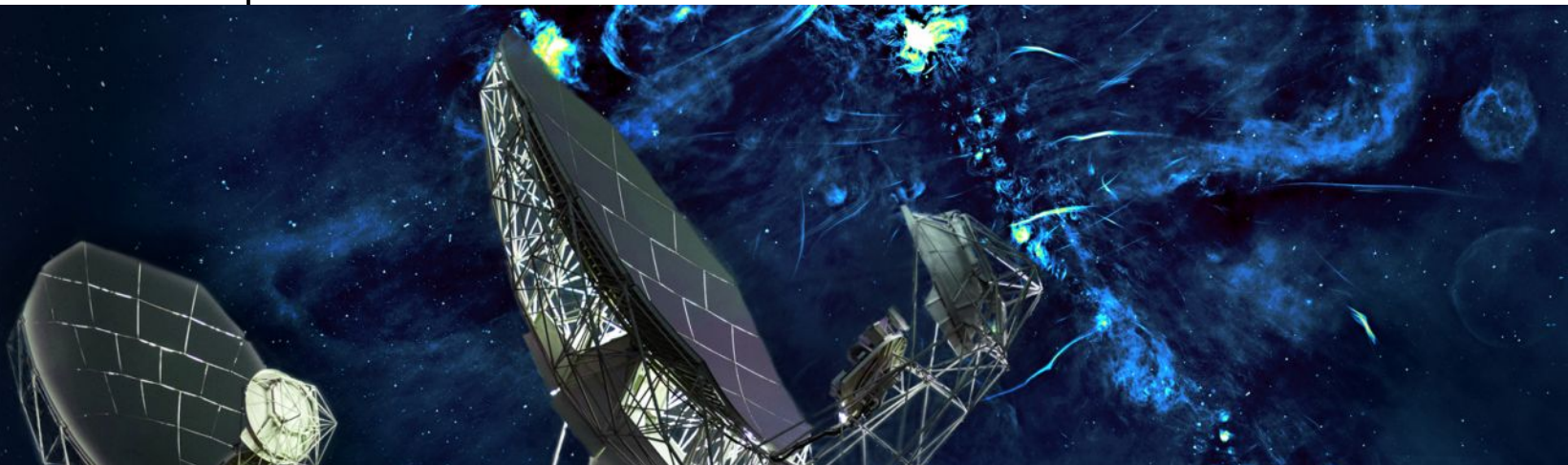
Introduction

Zanele & Sydil (Software Engineers)

Organisation

- **South African Radio Astronomy Observatory (SARAO)**

- **SARAO** is funded by the **NRF**

- **Mission:** to collect data using MeerKAT telescope for astronomers to analyse and help understand the universe better. They do this by publishing scientific papers.

Software Team

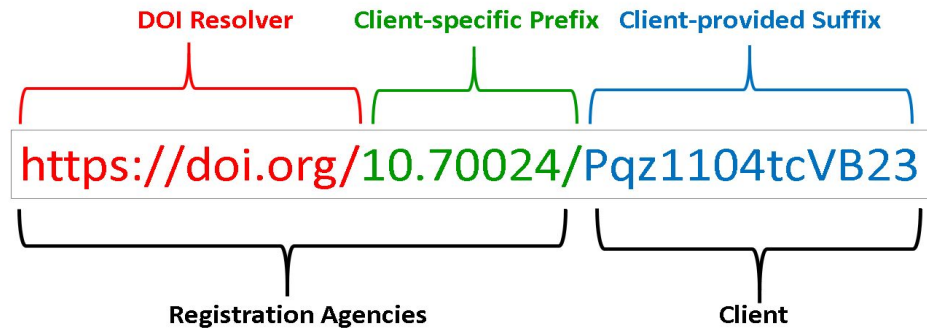- **Responsibilities:** Controlling, Monitoring and Supporting the MeerKAT telescope.

Software Team, 2023

Pixels

- A subteam of the **Software Team**
- Responsible for creating different applications to link users with the telescope data

- **One** of our applications is a **Digital repository** to store **digital assets**
- These Digital assets each have a **Digital Object Identifier (DOI)**

DOI

- **DOI** is an alphanumeric string to locate the object persistently

DOI Resolver   Client-specific Prefix   Client-provided Suffix

https://doi.org/10.70024/Pqz1104tcVB23

Registration Agencies   Client

(Novacescu, 2018)

- **DOIs** differ from normal urls which might change location or get lost

Application (Digital repository)

Datacite

- The agency we use to create DOIs
- It generates the DOI with **mandatory fields**
- {Title, Authors, Abstract, Data location}

Application (Digital repository)

client.uid:whno.ljncxe

Pages ▾   Support   Sign In

▤ Works   ♟ People   🏛 Organizations   🗄 Repositories

18 Works

Beamformer Coherency Commissioning Tests in 1K-mode at L-band

Marisa Geyer, Sarah Buchner & Maciej Serylak

Content published 2021 in SARAO Repository

Similar to an earlier commissioning report (M2600-0000-010) on the coherency of the MeerKAT Tied-Array Beam (TAB) in the 4096 frequency channel (4K) mode, this report investigates the level of coherency of the TAB in the 1024 frequency channel (1K) mode at L-band frequencies (856 MHz to 1712 MHz). The 1K-mode is designed specifically for high precision pulsar timing work, exchanging frequency resolution for higher timing resolution, and is the preferred mode for most Meertime1 pulsar observations. In the 1K folded mode used here, the Pulsar Timing User Supplied Equipment (PTUSE) provides 8 second folded profiles, with 1024 phase bins per profile, similar to the 4K folded mode. The data quality of the TAB is monitored while observing a well-known pulsar and changing the number of antennas used to form the TAB. The signal-to-noise-ratio (S/N) of the pulsar observation is expected to increase linearly with an increasing number of antennas. Perfect antenna addition would produce a coherency value (α) of 1.0. The primary target used for these tests is PSR J1056−6258, as per the 4K report (M2600-0000-010), along with a couple of observations using PSR J1644−4559. The TAB coherency specification is described in R.C.SP.23 CBF TA beamforming efficiency as follows: "The Correlator Beamformer (CBF), when requested to produce one or more TA beams, shall maintain, with an efficiency factor of >= 0.95, constant S/N per TA produced." We find that this specification is met by achieving a mean coherency factor of hαi = 0.99±0.01 and a minimum α value using PSR J1056−6258 of 0.95±0.03.

DOI registered February 19, 2021 via DataCite.

ⓒ ⓘ

Report   Physical sciences   English

doi https://doi.org/10.48479/6ty4-jv90

Overview of MeerKAT L-band AND U-band Nonlinearity

Marcel Gouws

Creators & Contributors ⓘ

☐ Cotton, William	3
☐ Camilo, Fernando	3
☐ Legodi, Letjatji	2
☐ Heywood, I	2
☐ Bester, H	2
☐ Smirnov, O	2
☐ Condon, James	2
☐ Randriamampandry, Solohery	1
☐ Moloko, Malebo	1
☐ Rudnick, Lawrence	1

Publication Year

☐ 2023	1
☐ 2022	7
☐ 2021	10

Work Type

☐ Dataset	14
☐ Report	4

FEEDB

Datacite

DEMO

Manual Process

- Multiple Authors ??

Problems

- Tedious to input Authors
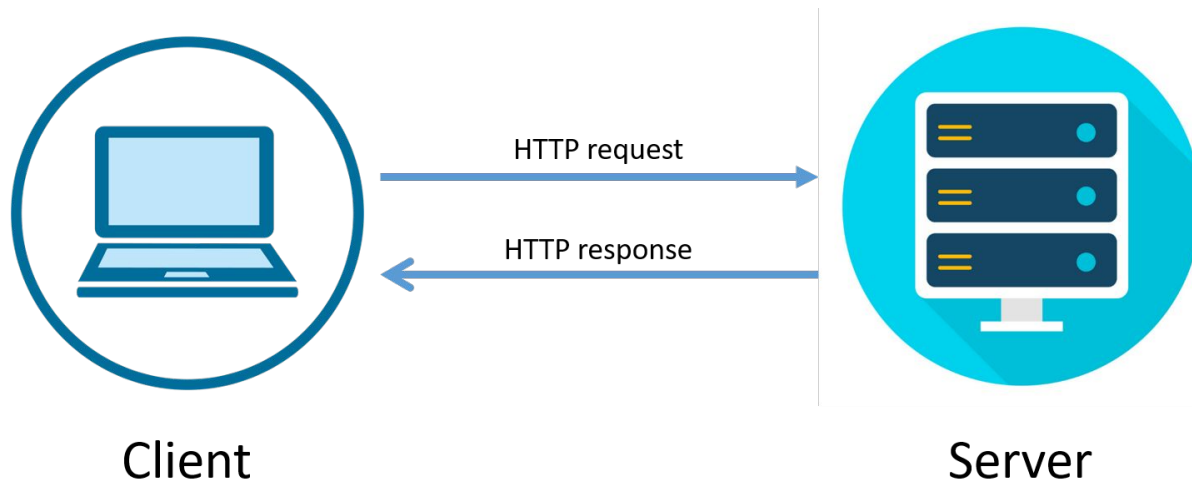
- Prone to errors

- Time Consuming

Solution

Automate the process of generating a DOI using a **Python Script**

Method

- Python's **Request library** : A set of rules to allow browsers to request web pages from servers

HTTP request

HTTP response

Client

Server

Python Class

- Metadata
- Authorization

- create_doi
- update_doi
- register_doi
- findable_doi
- hide_doi
- getInfo
- delete_doi

Python Code

```python
class Dois():
    """Short su...

    Generating of D...    all associated updates such as changing the states of
    the DOI from draft to reg...tered to being findable.


    Parameters
    ----------
    password_file : file
        A path to where the password file is stored.
    yaml_data : type
        A file containing the metadata of the data/ report release.
    doi : type
        Digital Object Identifie received for the data/ report release.
    """

    def __init__(self, password_file ,yaml_data, doi,  url   "https://api.datacite.org/dois", headers = {'Content-type': 'application/vnd.api
        """
        Initialising the objects's attributes that w... be used throughout
        the instance call.
        """

        self.doi = doi
        self.url = url
        self.username, self.pas...ord   self.pswd(password_file)
        self.headers = headers
        self.data_dict = sel...t_yaml_to_dict(yaml_data)


    def create_doi(self):
        """
        Creates a doi using a yaml file, which is converted into a dictionary
        and then send into a post request  that creates the DOI.
        """

        resp = requests.post(self.url, headers=self.headers, data=json.dumps(self.data_dict), auth=(self.password, self.username))
        suffix = resp.json()['data']['attributes']['doi']
        if resp.status_code == 201 or resp.status_code == 200 :
            logger.info(f"DOI {suffix} was created")
        else:
```
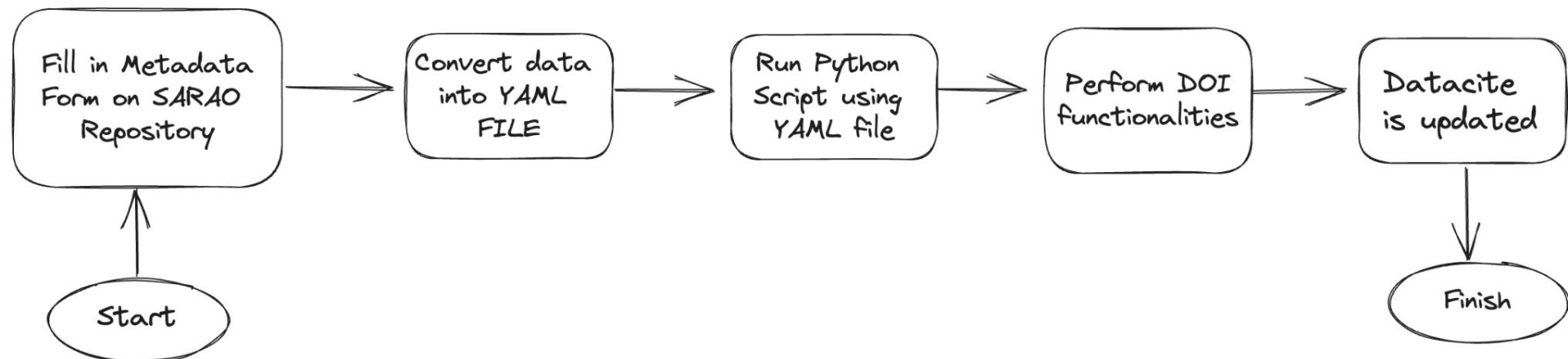
Work Flow

Python Script

DEMO

Examples

- **Automatic web page creation**

https://archive-gw-1.kat.ac.za/public/repository/10.48479/7epd-w356/index.html

NRF | SARAO
National Research Foundation | South African Radio Astronomy Observatory

MeerKAT Galaxy Cluster Legacy Survey Data Release 1 (MGCLS DR1)
https://doi.org/10.48479/7epd-w356

The first data release of the MeerKAT Galaxy Cluster Legacy Survey (MGCLS) consists of the uncalibrated visibilities, a set of continuum imaging products, and several source catalogues. All clusters have Stokes-I products, and approximately 40% have Stokes-Q and U products as well. For full details, including caveats for usage, see the survey overview and DR1 paper (Knowles et al., 2021).

When using any of the below products, please cite Knowles et al. (2021) and include the following Observatory acknowledgement: *"MGCLS data products were provided by the South African Radio Astronomy Observatory and the MGCLS team and were derived from observations with the MeerKAT radio telescope. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation."*

The DR1 products:

- ***Visibilities***: These can be accessed through the MeerKAT archive, under proposal ID SSV-20180624-FC-01.

- ***Basic imaging products***: Full field of view, with no primary beam correction, available at full resolution only (~7-8"). Consists of a single 16-plane cube with integrated intensity, spectral index, and 14 frequency planes. Useful for full-field searches and source finding.
  Access: Basic products

Examples

- **Data Transfer from CEPH storage to s3 bucket**

HOW AWS S3 BUCKETS WORK

User — Upload/Download an Object — Internet — Amazon S3
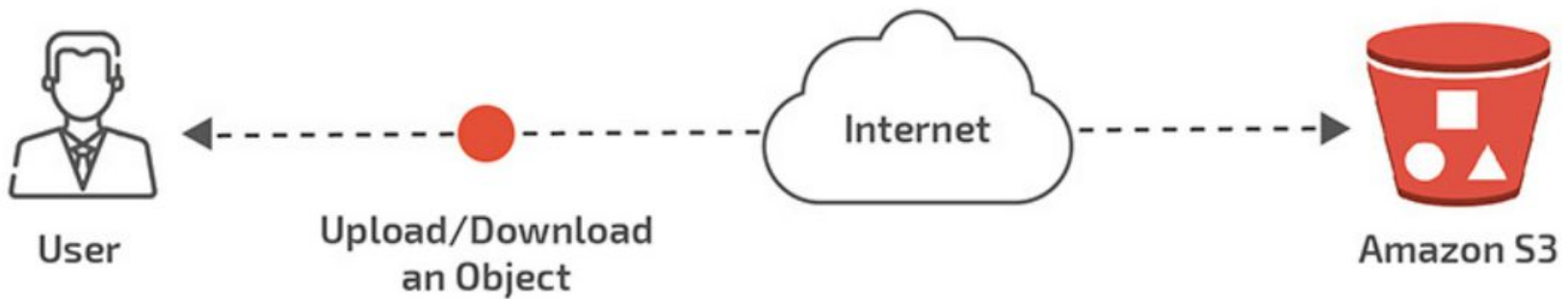
Conclusion

- Leveraged the power of Python to **automate** our process

- **Simple** and **Adaptable**

- **High level language**, thus allows developers to focus on solving problems rather than dealing with system-specific intricacies

- Large **standard libraries** for different types of tasks

References

- http://www.sarao.ac.za/
- https://archive-gw-1.kat.ac.za/public/repository/10.48479/7epd-w356/index.html