Implementing CKAN with Datacite DOI

Olatunbosun Obileye
ORCID: 0000-0002-1200-0994

Hafeez Adepoju
ORCID: 0000-0003-3516-4294

Presented during Datacite Members Meeting for EMEA
20 October 2020

What we will cover…

- A little insight into CKAN and Datacite at IITA
  - IITA's dream
  - Journey to…
- IITA data repository
  - Why Datacite
  - CKAN design
  - Architecture
  - Integration with APIs
- Features of IITA's CKAN
- Doing more with Datacite…
  - Offerings
  - DOI minting
  - Integration of breeding database (Cassavabase) into CKAN
- Impact
- Credits
- Discussions

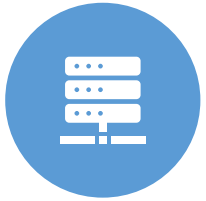
A little insight into CKAN and Datacite at IITA

IITA's dream

IITA wished to have an institutional data repository that conforms to open access/open data.

Repository should be trusted with long-term access.

The repository must meet FAIR data repository expectations.



The repository should be acceptable and integratable to future CGIAR global data platform(s).

Must be accessible globally without any restriction.

Conforms to CGIAR agreed metadata standard.

Journey to…

Trusted Repository

- Sustainability
- Security
- Long-term preservation
- Get credit for data use
  - Visible
  - Findable - Accessible
  - Permanent Identifier.. Datacite DOI

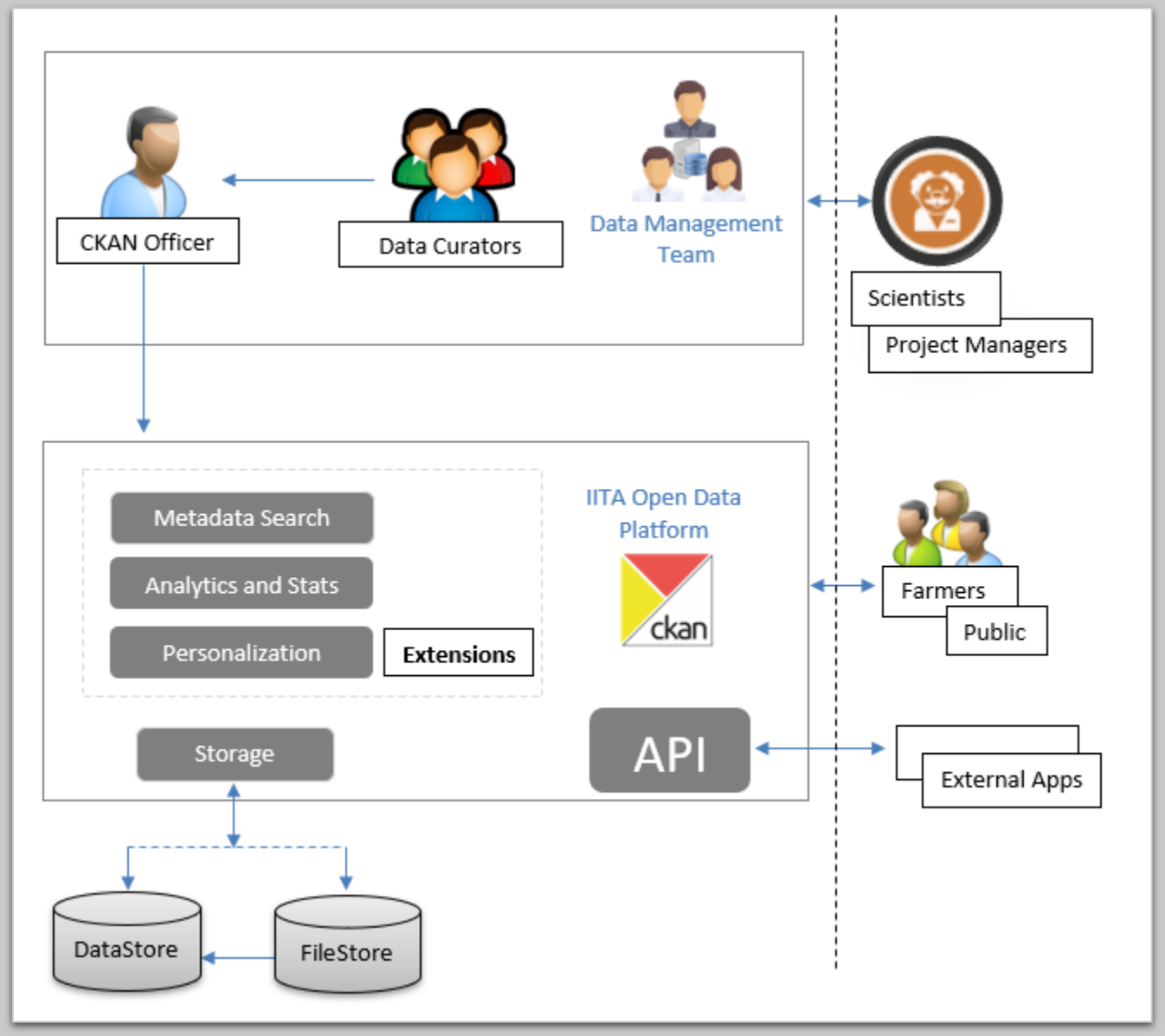
IITA Data Repository

CKAN is the platform!
Datacite DOI is the PID!

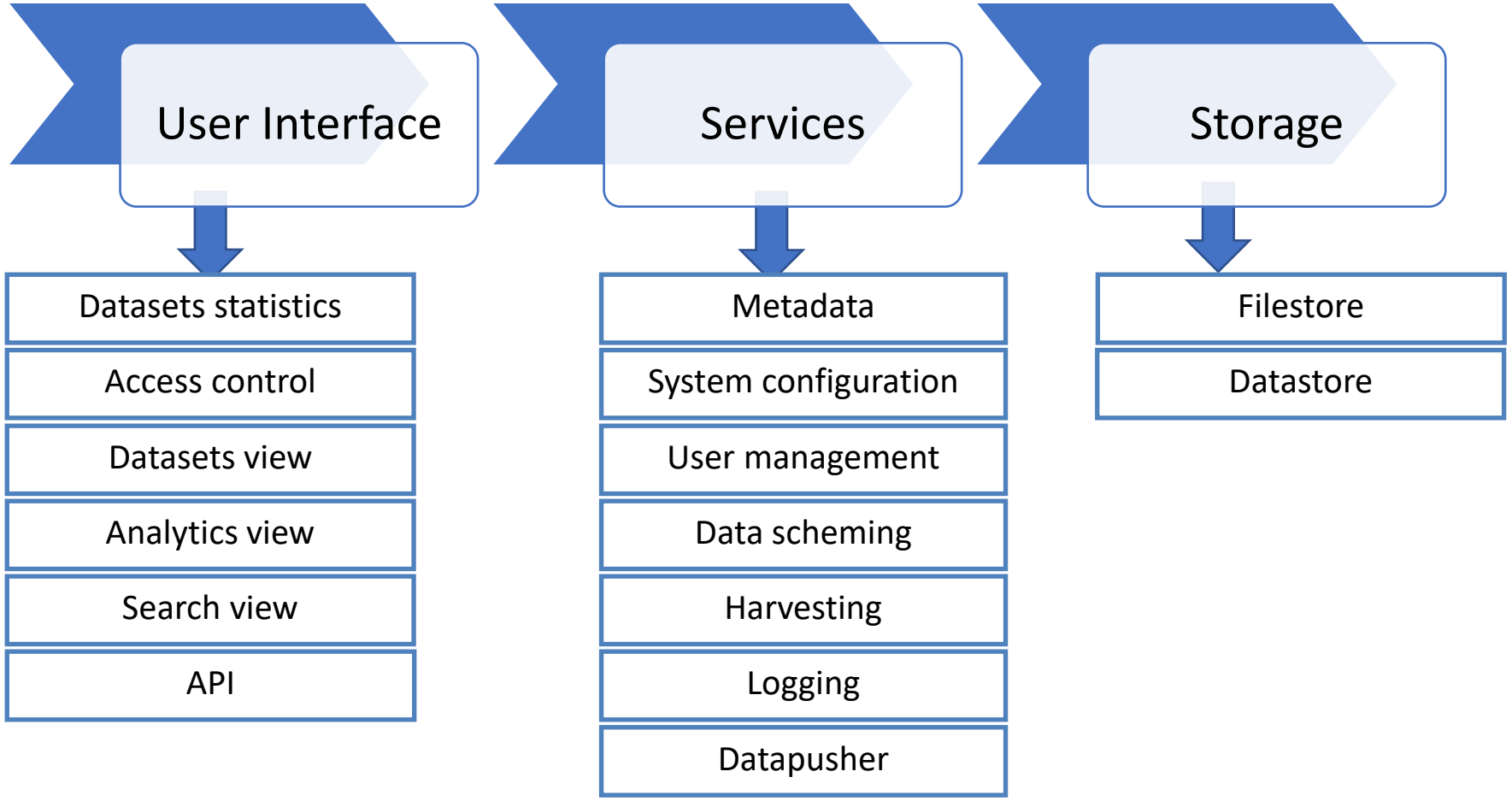
Why Datacite?

- Support data citation API’s
- Increase data visibility with 3rd party partnerships and integration
- Data use tracker
- Availability of citation formatter through http://citation.crosscite.org
- It has a supportive community
- It is easy to manage

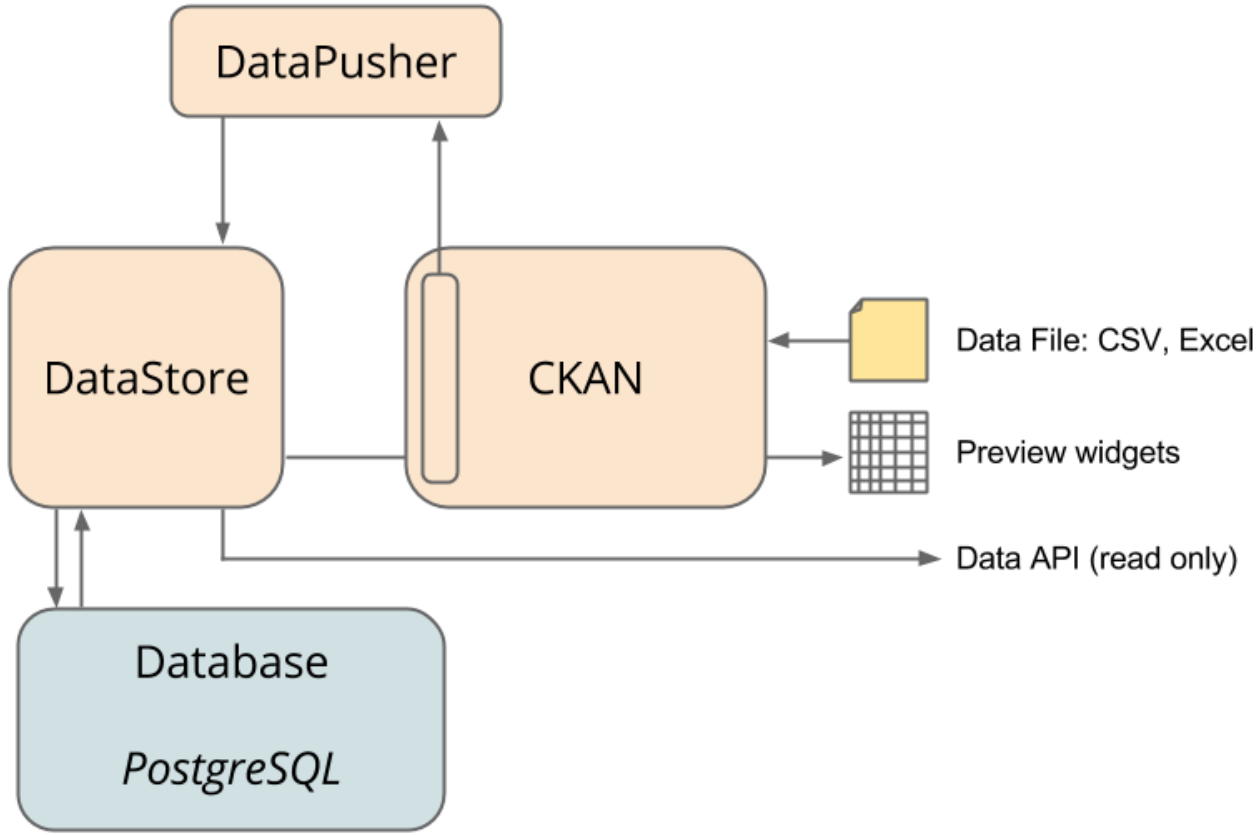
IITA's CKAN design overview



IITA's Open Data (CKAN) Architecture



DataStore + DataPusher



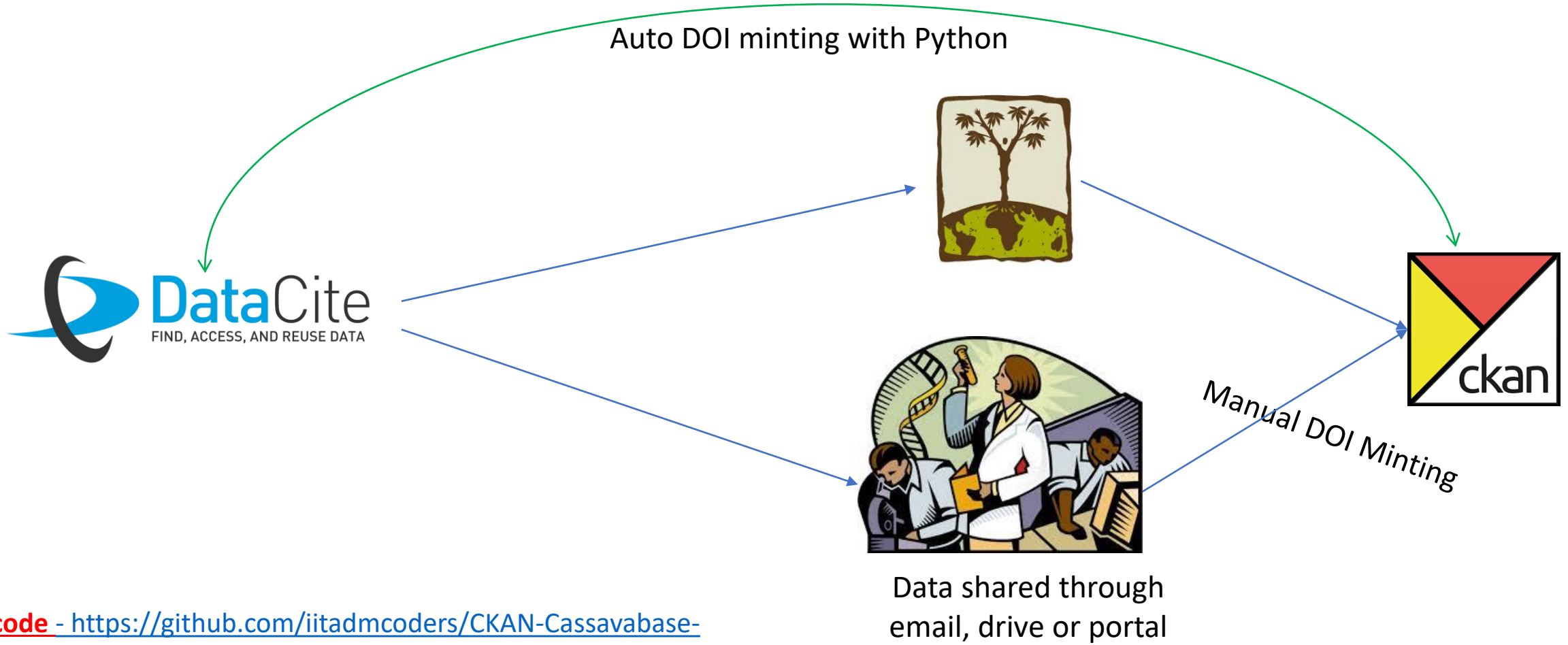
DataStore and DataPusher relationship

CKAN and Datacite DOI Integration

How it works

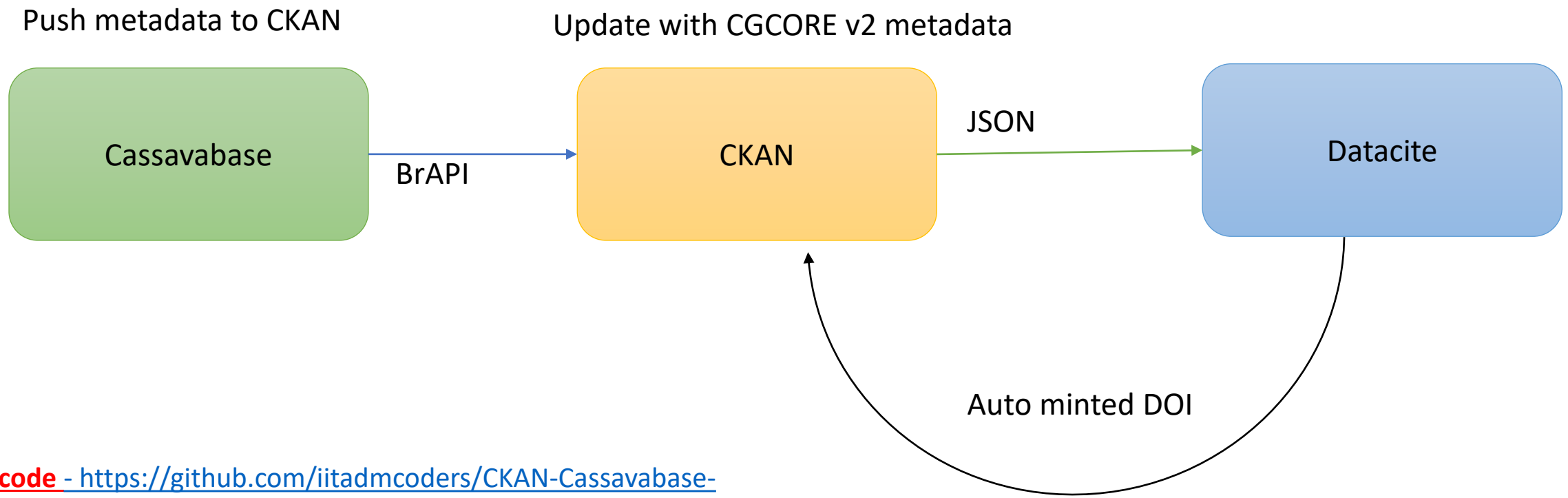
- IITA adopts 2 methods of DOI minting
  - Auto-generation using python and Datacite API
  - Manual from Fabrica
- Auto-generation is for data coming from other research databases like Cassavabase
- Manual is for data shared through any of:
  - email,
  - data submission app or
  - shared-drive

DOI Minting



Get the code - https://github.com/iitadmcoders/CKAN-Cassavabase-integration/blob/main/ckanext/ckanmusabase/controller.py

Auto minting of doi using python

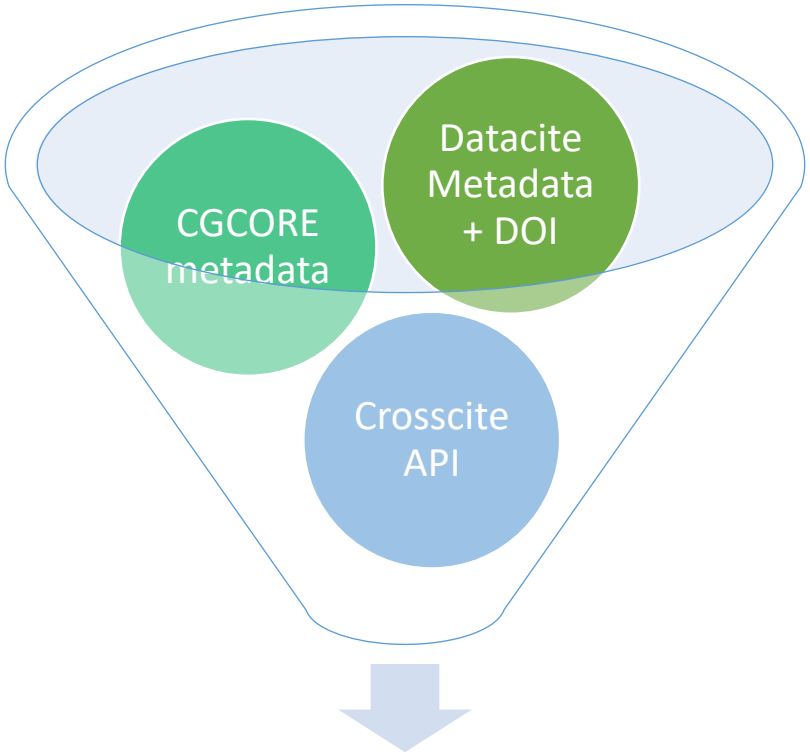


Get the code - https://github.com/iitadmcoders/CKAN-Cassavabase-integration/blob/main/ckanext/ckanmusabase/controller.py

Data citation from Datacite doi… How it works

- IITA uses CGCORE v2 metadata standard
- It is merged with Datacite metadata and the minted doi
- This is transmitted to Crosscite
- The output citation styles is posted to CKAN

Data citation generation



Multi-styled Data Citation

Features of IITA's Data Repository, CKAN

- IITA central institutional data repository features
  - Conforms to global FAIR and trusted repository requirements
  - PID – Datacite doi
  - AGROVOC
  - Tags
  - CGcore metadata v2
  - 10 Data citation styles/standards
  - Embargo period settings
  - Integration with Breeding database (Cassavabase)
  - Integration with DSpace
  - Integration with GARDIAN and other Big Data Platforms
  - Available in Google Data-search
  - Google Analytics
  - Data visualization
  - Data submission platform – beta stage

IITA's Institutional Data Repository URL – http://data.iita.org

Doing more with Datacite doi …

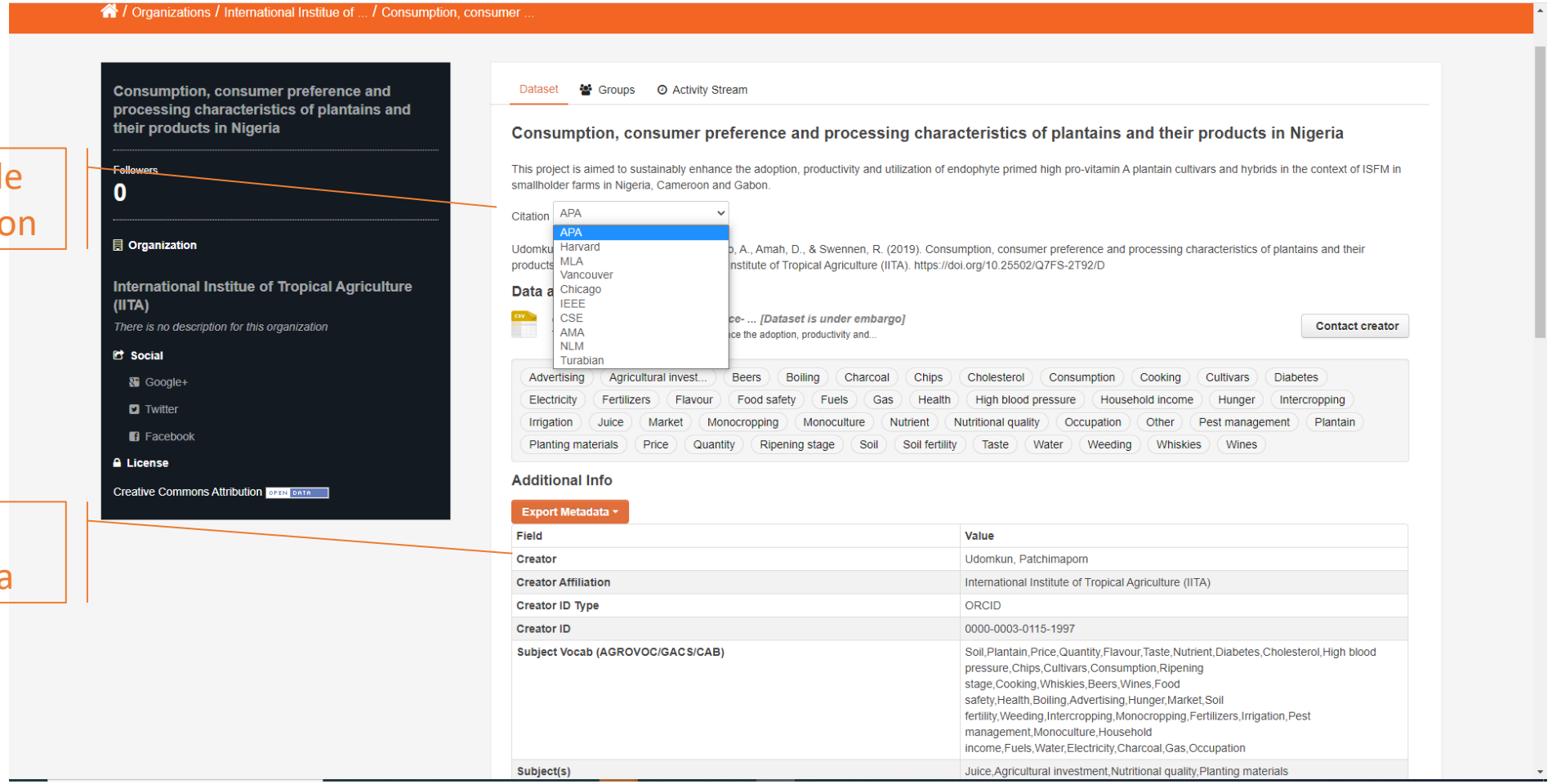
Doing more with Datacite offerings

- Datacite DOI facilitated multi-style data citation
- Integration with customized metadata for CGIAR, CGcore metadata, achieved seamlessly
- Setting out plans to have multirepository doi infrastructure
- Research databases without DOI, data citation, and CGCore metadata were integrated into CKAN
  - Started with Cassavabase
  - Musabase near completion

Doing more with Datacite DOI

Multi-style data citation

CGcore metadata

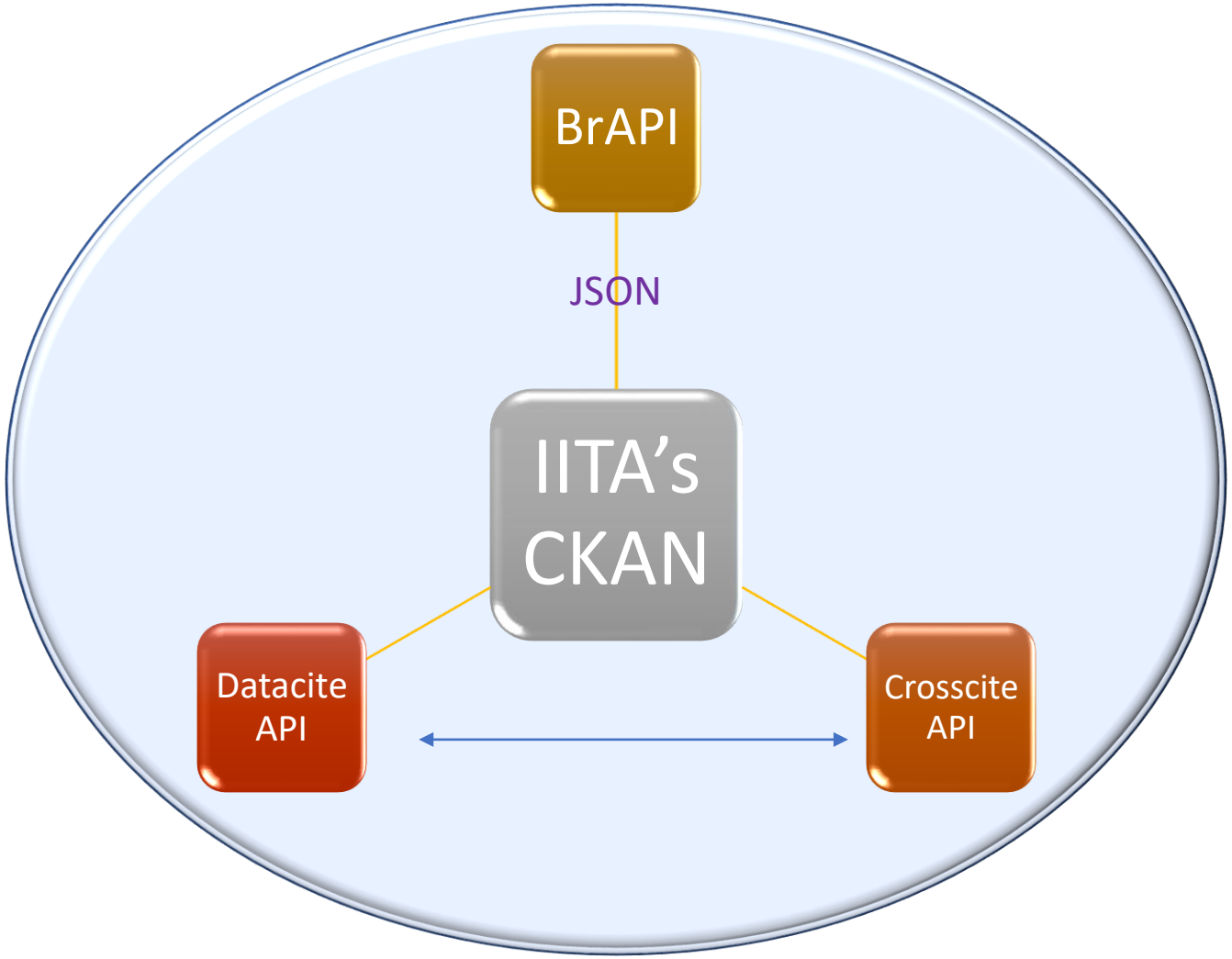


Doing more with Datacite doi – Cassavabase integration with CKAN

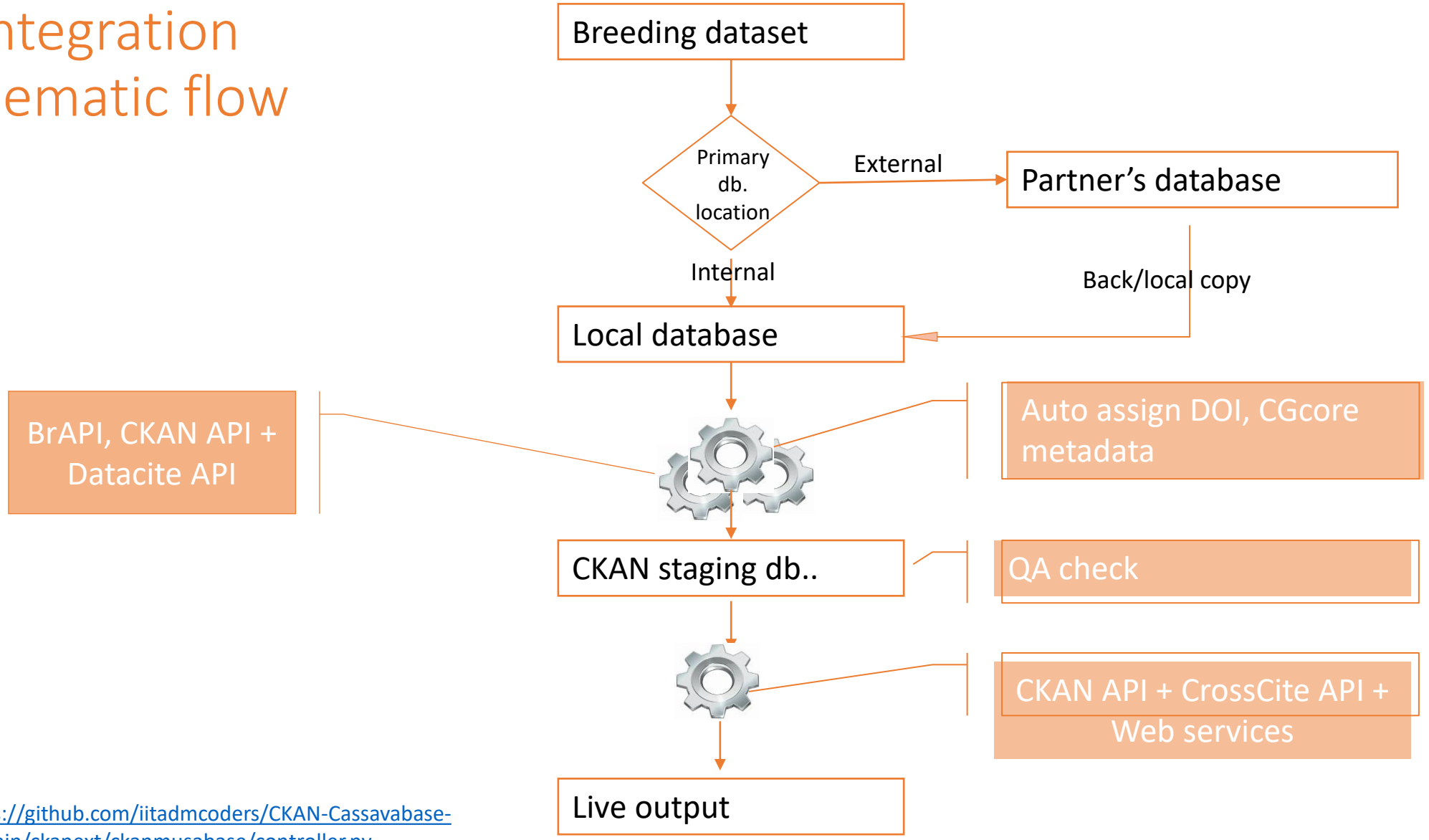
How it works

- Use unique identifier to identify IITA datasets in breeding database (studyDbId & programDbId)
- Define rest state as completed projects using certain parameters
- Create script to crawl breeding database to:
  - Identify IITA's data
  - Data in rest state call up
  - Crawl db at off-peak with schedule
- Auto-review data and update with necessary doi, tags and CGcore metadata v2
- Data quality auto check (PII, ethical concerns, anonymization, license type)
- Final publishing

CKAN integration with API’s

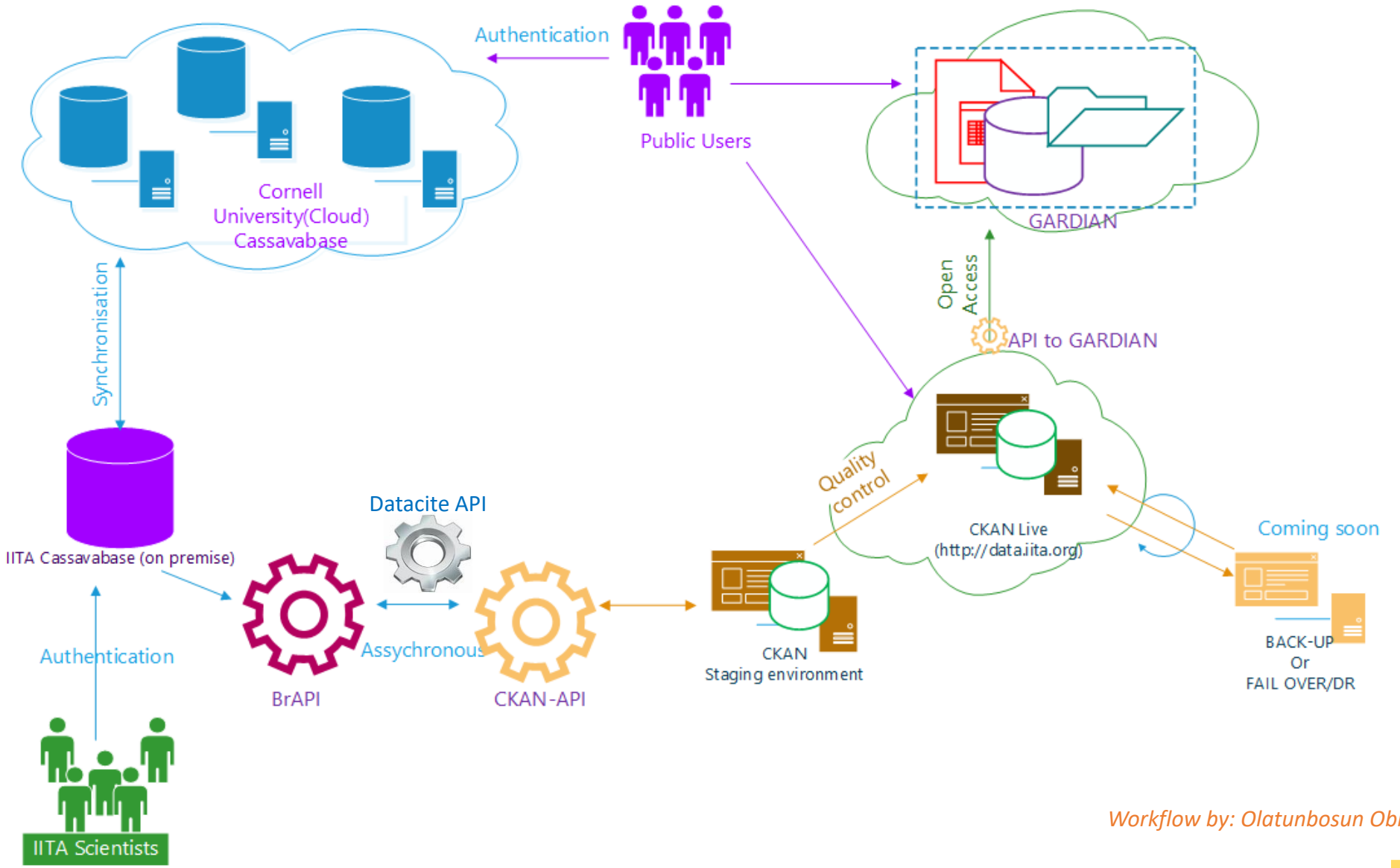


Integration schematic flow



Get the code - https://github.com/iitadmcoders/CKAN-Cassavabase-integration/blob/main/ckanext/ckanmusabase/controller.py

CASSAVABASE TO CKAN DATA INTEGRATION
IITA BIGDATA BUILD-UP



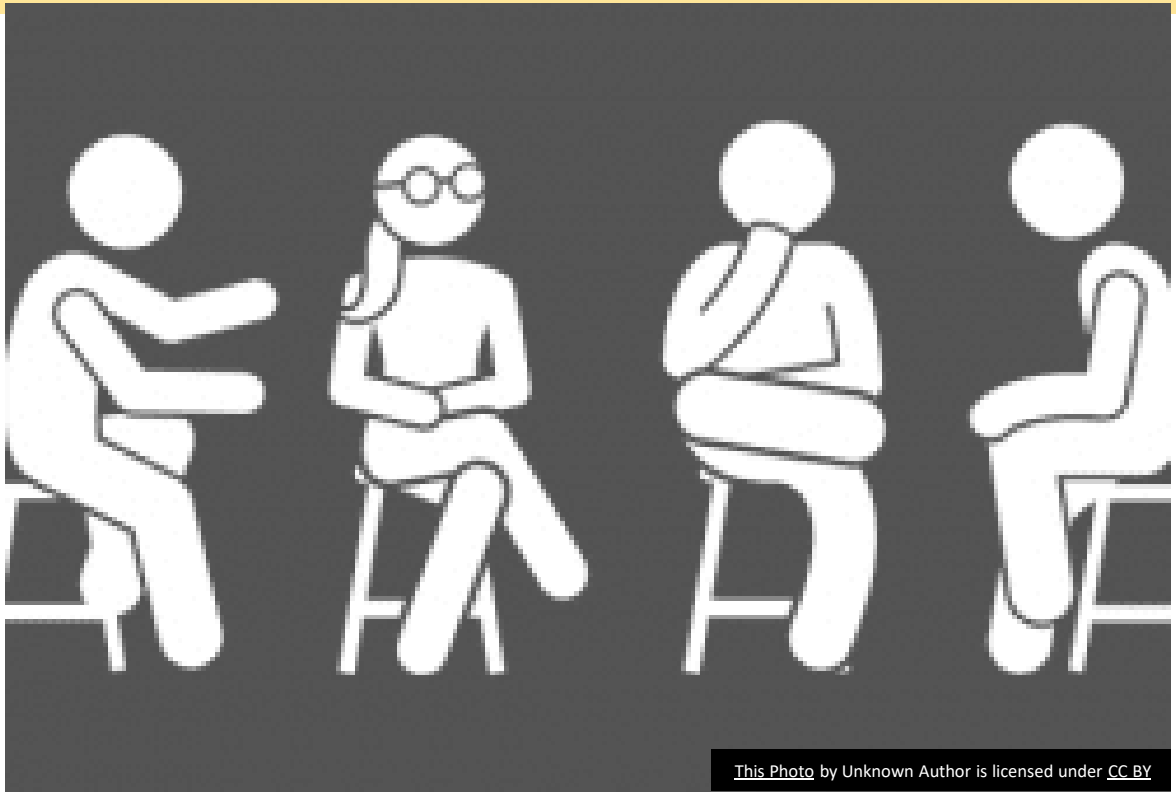
Workflow by: Olatunbosun Obileye & Abduljelil Olalekan

Impact

- Better acceptance from researchers/scientists
- Increased research data visibility within and outside CGIAR – GARDIAN
- Compliance with FAIR data principle
- Improved partnership
- Enhanced data management practices in the institution
- More funding

Credits

- Martin Mueller
- Peter Kulakow
- Katherine Lopez
- Tonny Omwansa
- Olatunbosun Obileye
- Hafeez Adepoju
- Phanuel Ayuka
- Caroline Owuor
- Lekan Anifowose
- Afolabi Agbona
- Peteti Prasad
- Lukas Mueller
- Cornell University
- USDA
- BTI
- IITA



Thank you