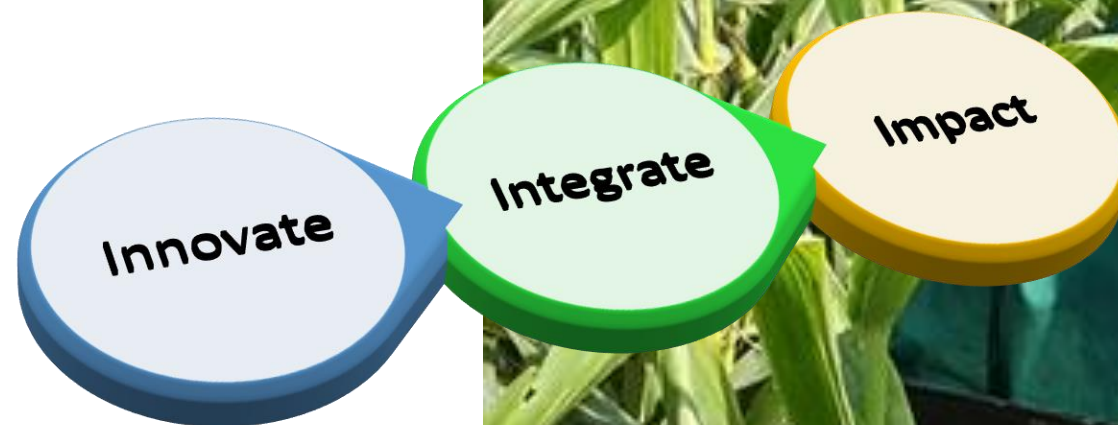


# Implementing CKAN with Datacite DOI



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

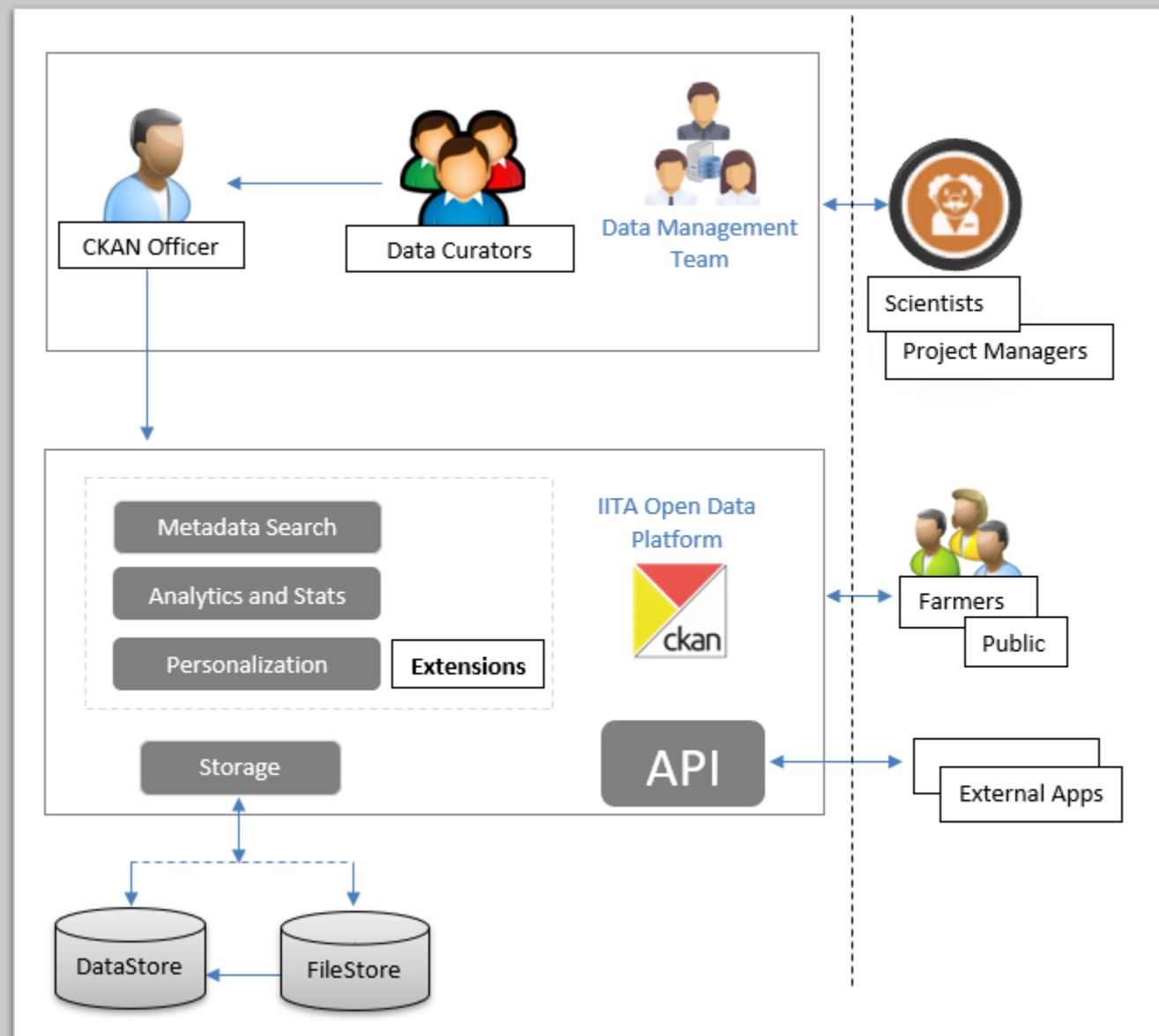
# IITA Data Repository

CKAN is the platform!  
Datacite DOI is the PID!

# Why Datacite?

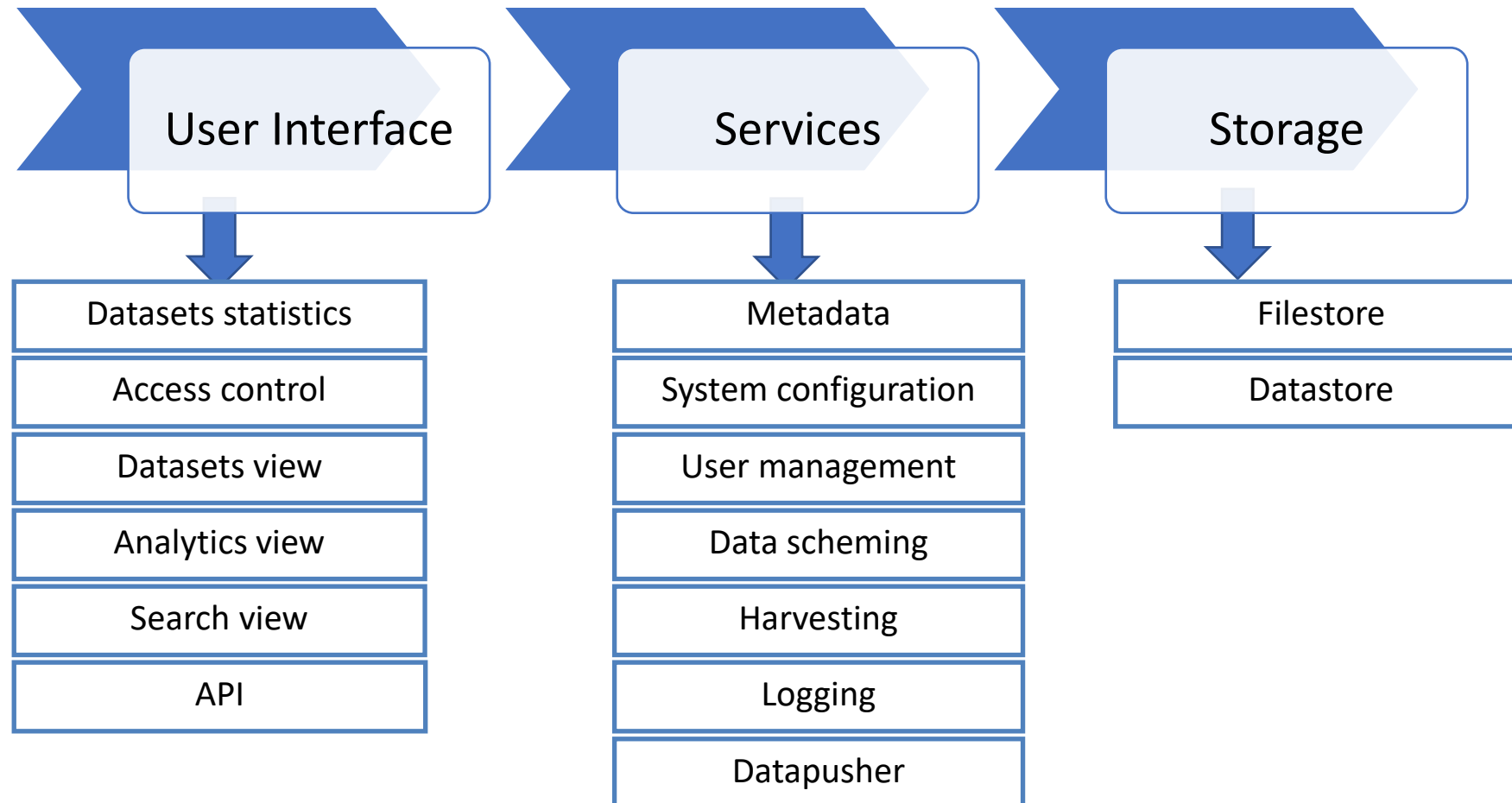
- Support data citation API's
- Increase data visibility with 3<sup>rd</sup> party partnerships and integration
- Data use tracker
- Availability of citation formatter through <http://citation.crosscite.org>
- It has a supportive community
- It is easy to manage

# IITA's CKAN design overview

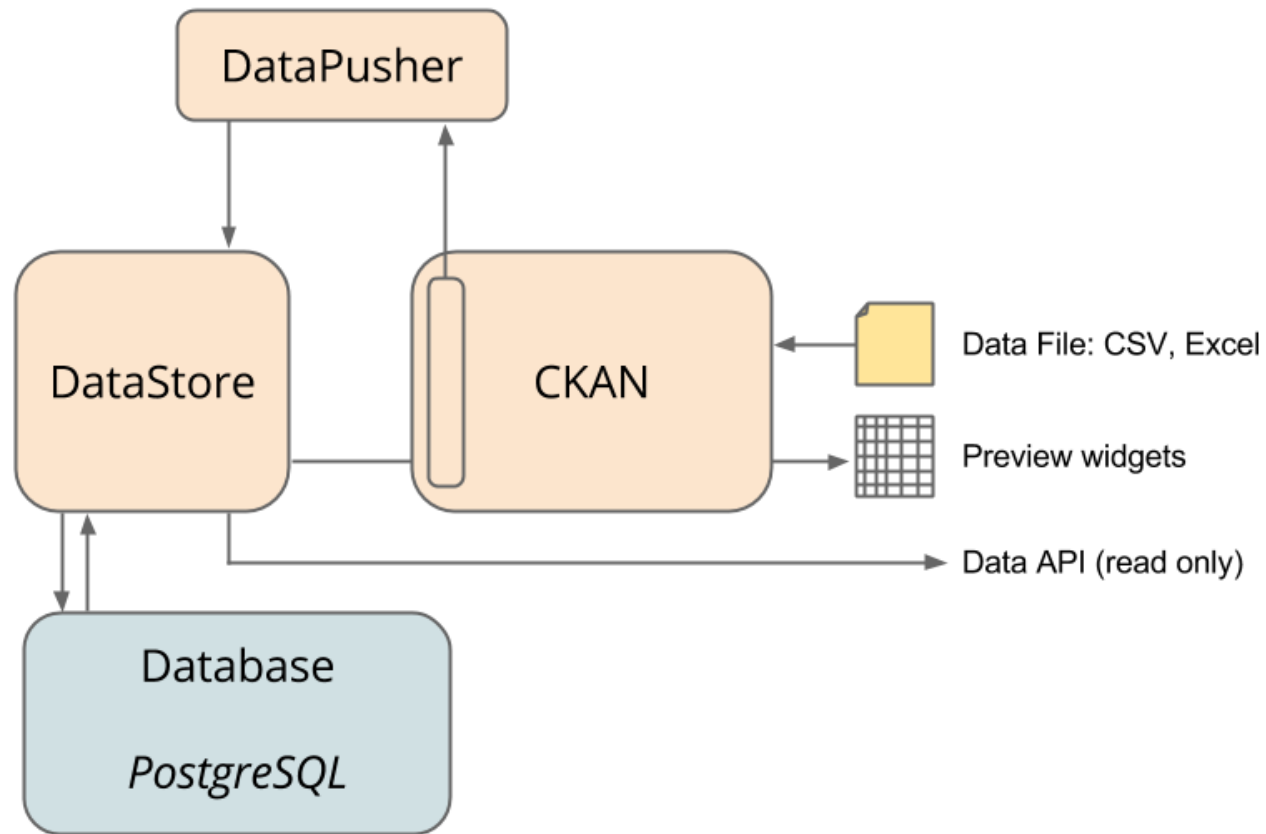




# IITA's Open Data (CKAN) Architecture



# DataStore + DataPusher



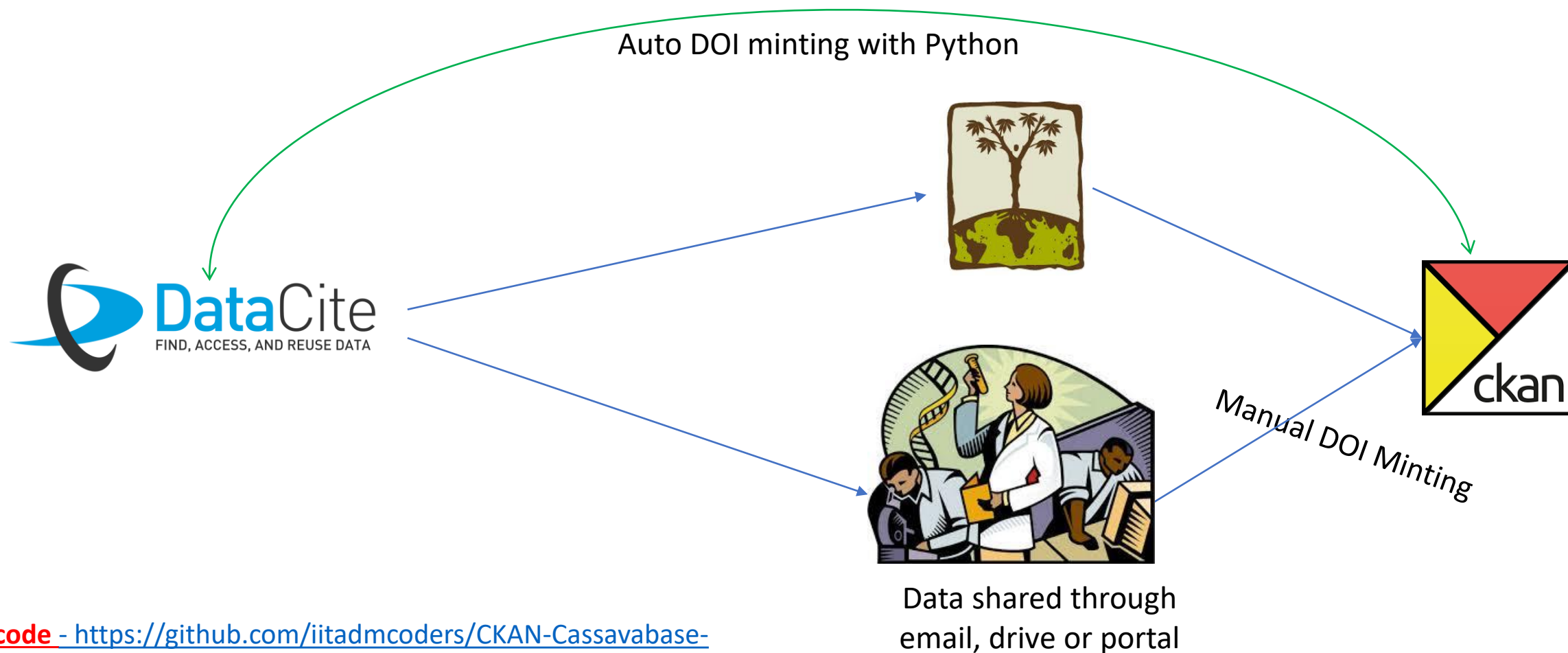
DataStore  
and  
DataPusher  
relationship

# CKAN and Datacite DOI Integration

## How it works

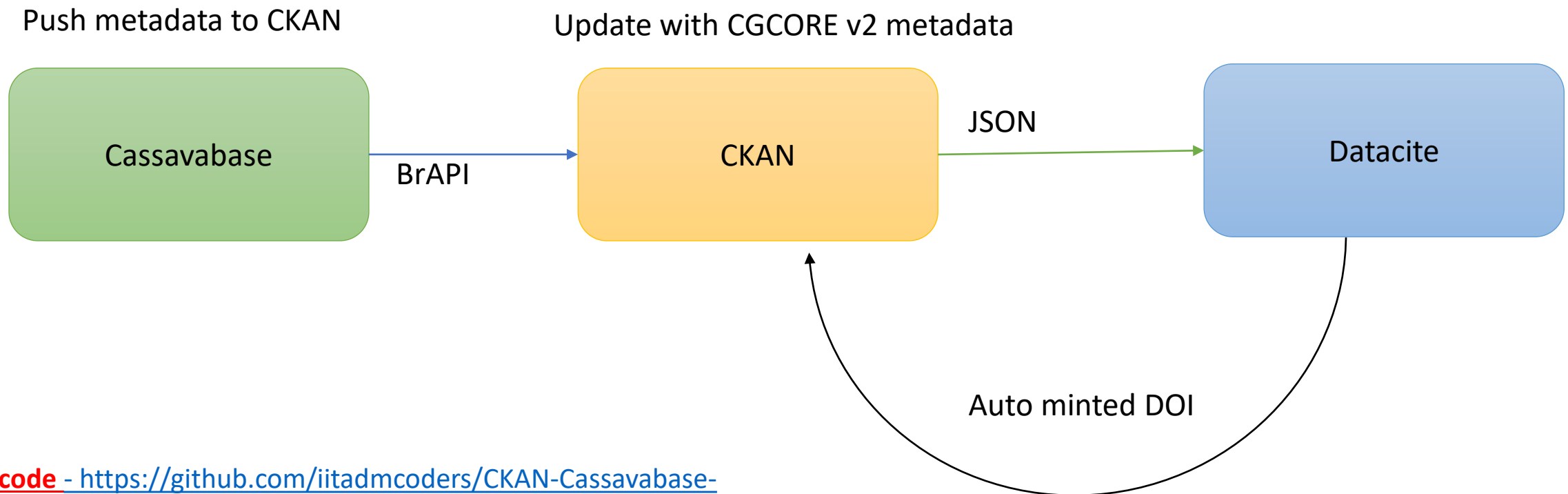
- IITA adopts 2 methods of DOI minting
  - Auto-generation using python and Datacite API
  - Manual from Fabrica
- Auto-generation is for data coming from other research databases like Cassavabase
- Manual is for data shared through any of:
  - email,
  - data submission app or
  - shared-drive

# DOI Minting



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

# Auto minting of doi using python

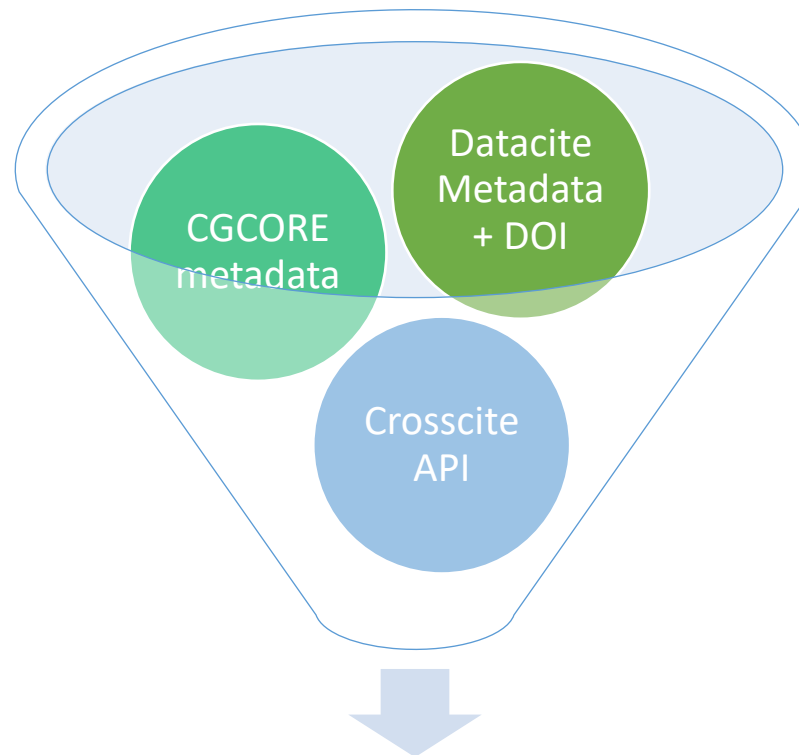


**Get the code** - <https://github.com/iitadmcoders/CKAN-Cassavabase-integration/blob/main/ckanext/ckanmusabase/controller.py>

# 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

[Home](#) / [Organizations](#) / [International Institute of ...](#) / [Consumption, consumer ...](#)

**Consumption, consumer preference and processing characteristics of plantains and their products in Nigeria**

Followers  
**0**

**Organization**

**International Institute of Tropical Agriculture (IITA)**

*There is no description for this organization*

**Social**

Google+

Twitter

Facebook

**License**

Creative Commons Attribution [OPEN DATA](#)

**Dataset** Groups Activity Stream

**Consumption, consumer preference and processing characteristics of plantains and their products in Nigeria**

This project is aimed to sustainably enhance the adoption, productivity and utilization of endophyte primed high pro-vitamin A plantain cultivars and hybrids in the context of ISFM in smallholder farms in Nigeria, Cameroon and Gabon.

Citation **APA**

Udomkun, A., Amah, D., & Swennen, R. (2019). Consumption, consumer preference and processing characteristics of plantains and their products in Nigeria, Cameroon and Gabon. International Institute of Tropical Agriculture (IITA). <https://doi.org/10.25502/Q7FS-2T92/D>

**Data**

CSV

APA

Harvard

MLA

Vancouver

Chicago

IEEE

CSE

AMA

NLM

Turabian

Advertising Agricultural invest... Beers Boiling Charcoal Chips Cholesterol Consumption Cooking Cultivars Diabetes Electricity Fertilizers Flavour Food safety Fuels Gas Health High blood pressure Household income Hunger Intercropping Irrigation Juice Market Monocropping Monoculture Nutrient Nutritional quality Occupation Other Pest management Plantain Planting materials Price Quantity Ripening stage Soil Soil fertility Taste Water Weeding Whiskies Wines

**Additional Info**

**Export Metadata**

Field	Value
Creator	Udomkun, Patchimaporn
Creator Affiliation	International Institute of Tropical Agriculture (IITA)
Creator ID Type	ORCID
Creator ID	0000-0003-0115-1997
Subject Vocab (AGROVOC/GACS/CAB)	Soil, Plantain, Price, Quantity, Flavour, Taste, Nutrient, Diabetes, Cholesterol, High blood pressure, Chips, Cultivars, Consumption, Ripening stage, Cooking, Whiskies, Beers, Wines, Food safety, Health, Boiling, Advertising, Hunger, Market, Soil fertility, Weeding, Intercropping, Monocropping, Fertilizers, Irrigation, Pest management, Monoculture, Household income, Fuels, Water, Electricity, Charcoal, Gas, Occupation
Subject(s)	Juice, Agricultural investment, Nutritional quality, Planting materials

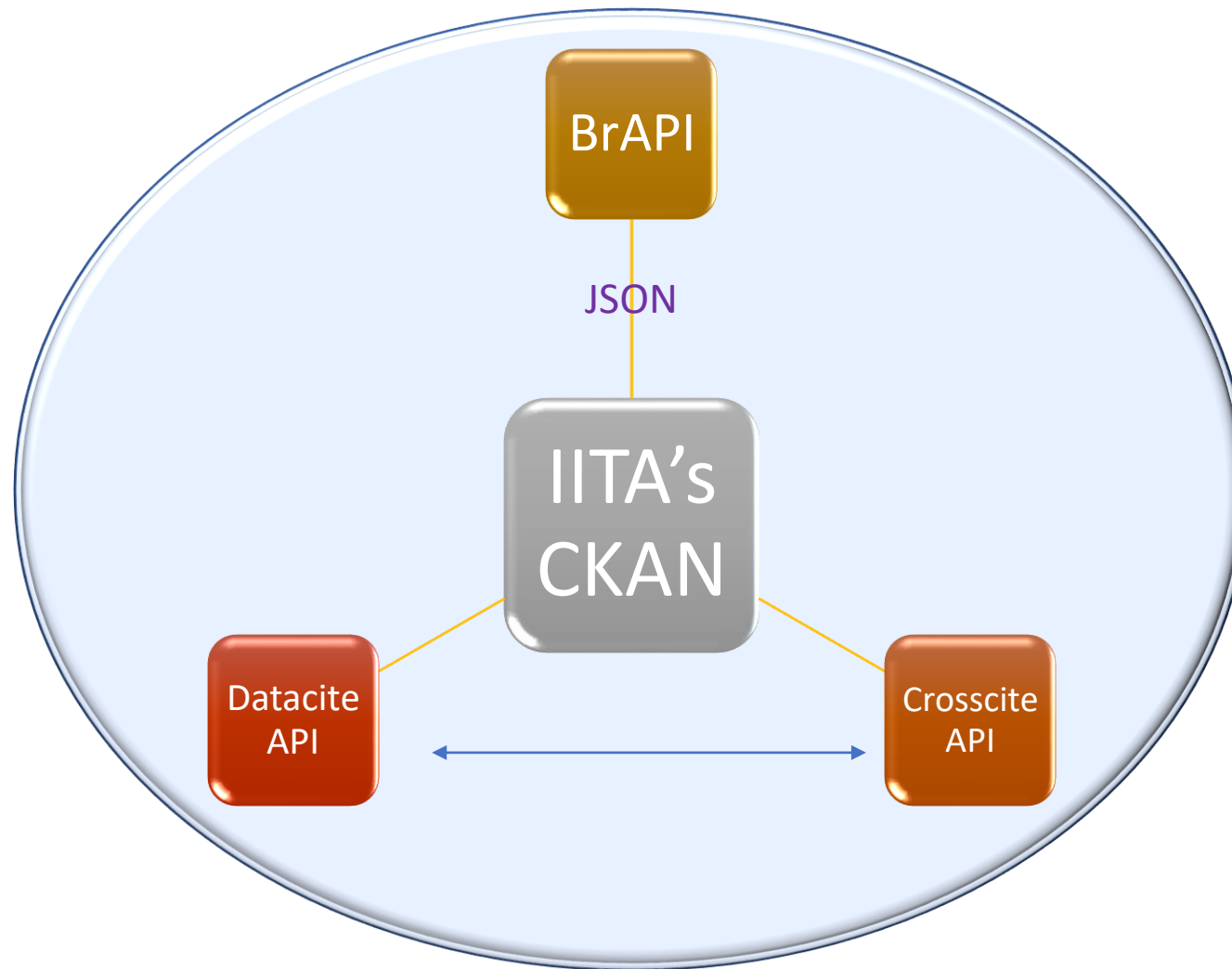
Contact creator

# Doing more with Datacite doi – Cassavabase integration with CKAN

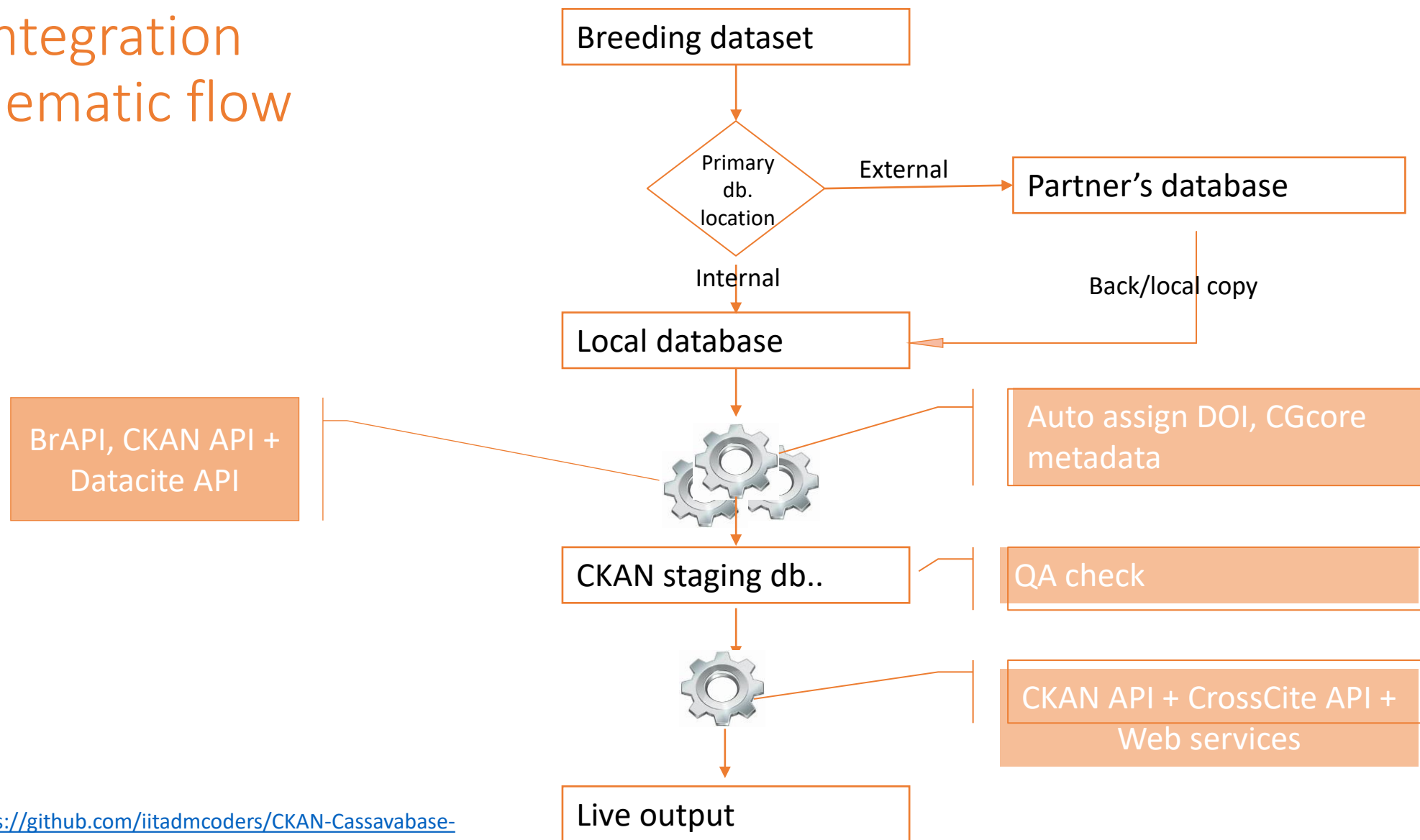
## How it works

- 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



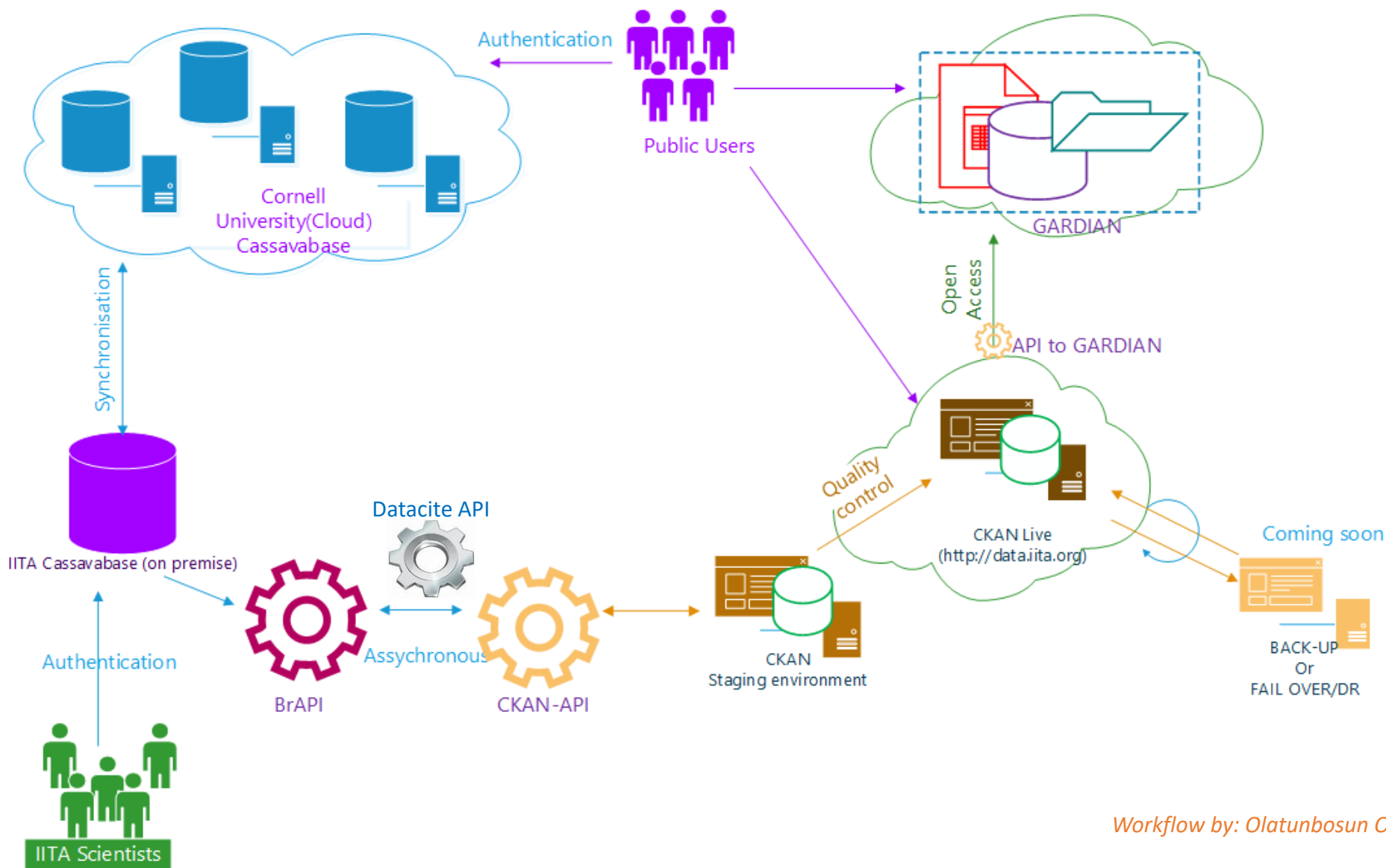
# Integration schematic flow



**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

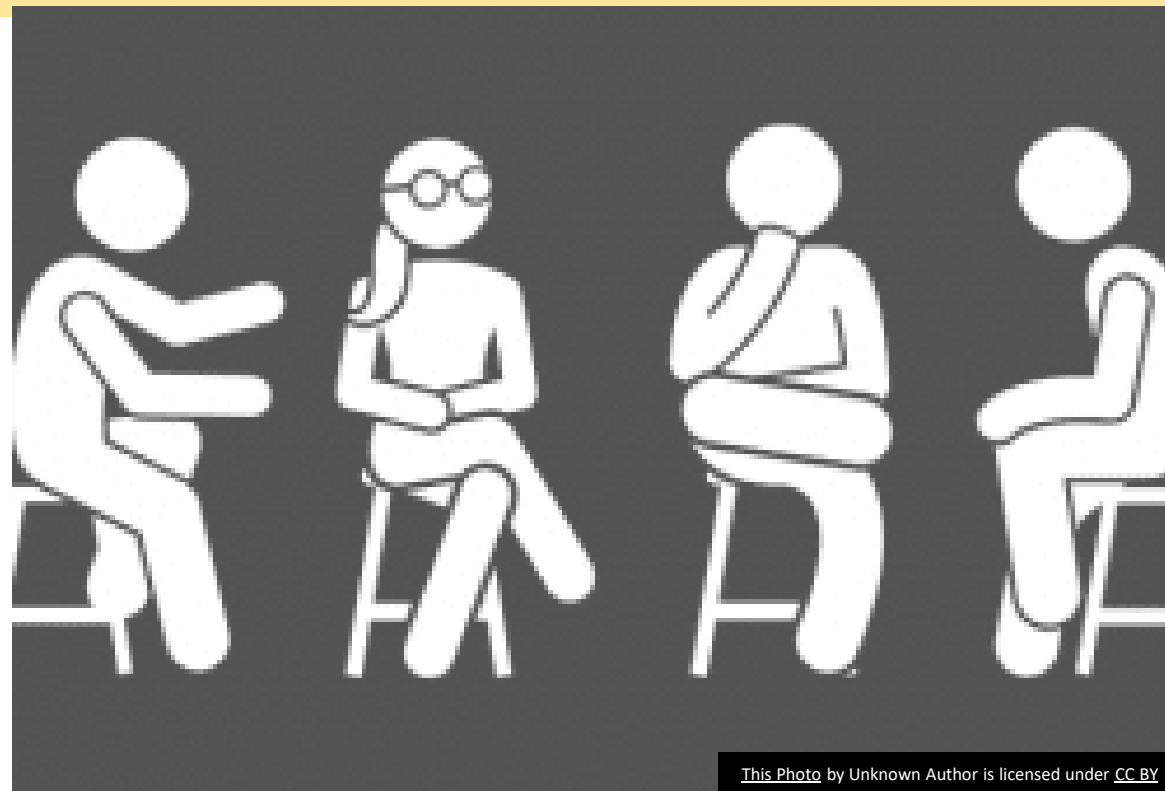
# Impact

- 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
- Phaniel Ayuka
- Caroline Owuor

- Lekan Anifowose
- Afolabi Agbona
- Peteti Prasad
- Lukas Mueller
- Cornell University
- USDA
- BTI
- IITA



# Thank you

