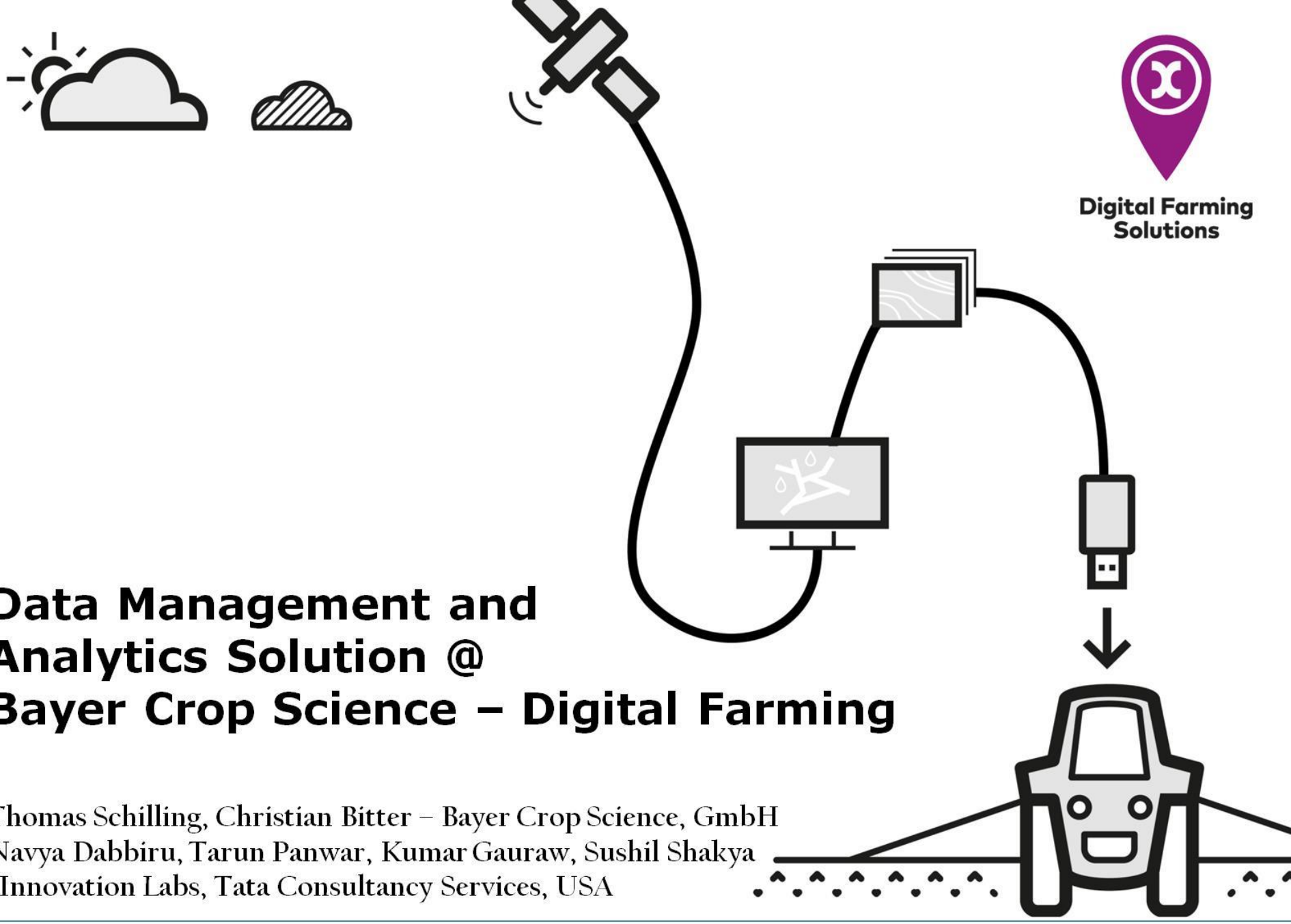


Data Management and Analytics Solution @ Bayer Crop Science – Digital Farming

Thomas Schilling, Christian Bitter – Bayer Crop Science, GmbH
Navya Dabbiru, Tarun Panwar, Kumar Gauraw, Sushil Shakya
-Innovation Labs, Tata Consultancy Services, USA

AGENDA

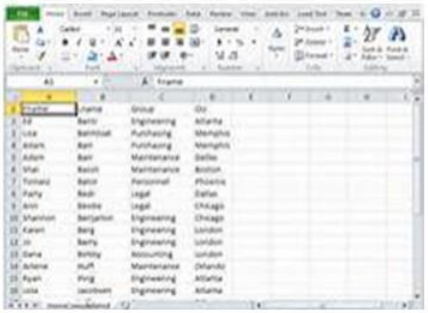
- **Motivation**
- **Our Solutions**
  - iRODS – Digital Farming Data Management Platform
  - iRODS – R-Client
  - iRODS – iMetaExploreR

Motivation - Our Data is Diverse

Digital Farming Solutions

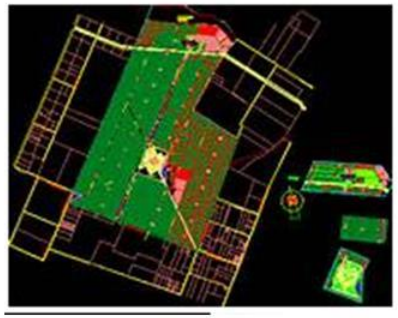
- Digital Farming produces large data sets mostly as diverse flat files that are spread across geographies.
- Typically these files are like

CSV files



JSON files

PRJ files

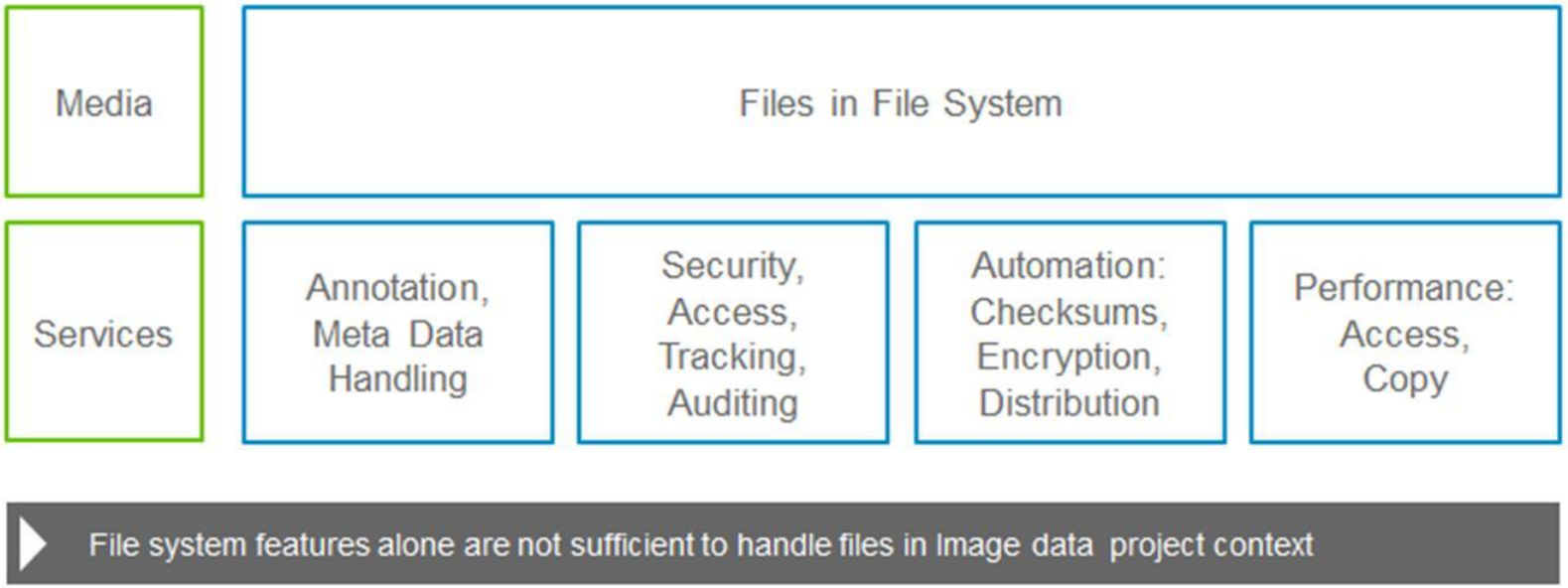


Shape files (Images)

Motivation - Information Management

Digital Farming Solutions

- Image Data Processing creates a whole bunch of data files in different formats. Content description, context and access restrictions (eg security, confidentiality) per file are difficult to handle across systems
- Storing this information in directory path and file name will no longer work for huge data platforms



Motivation – Data Management and Analysis

Digital Farming Solutions

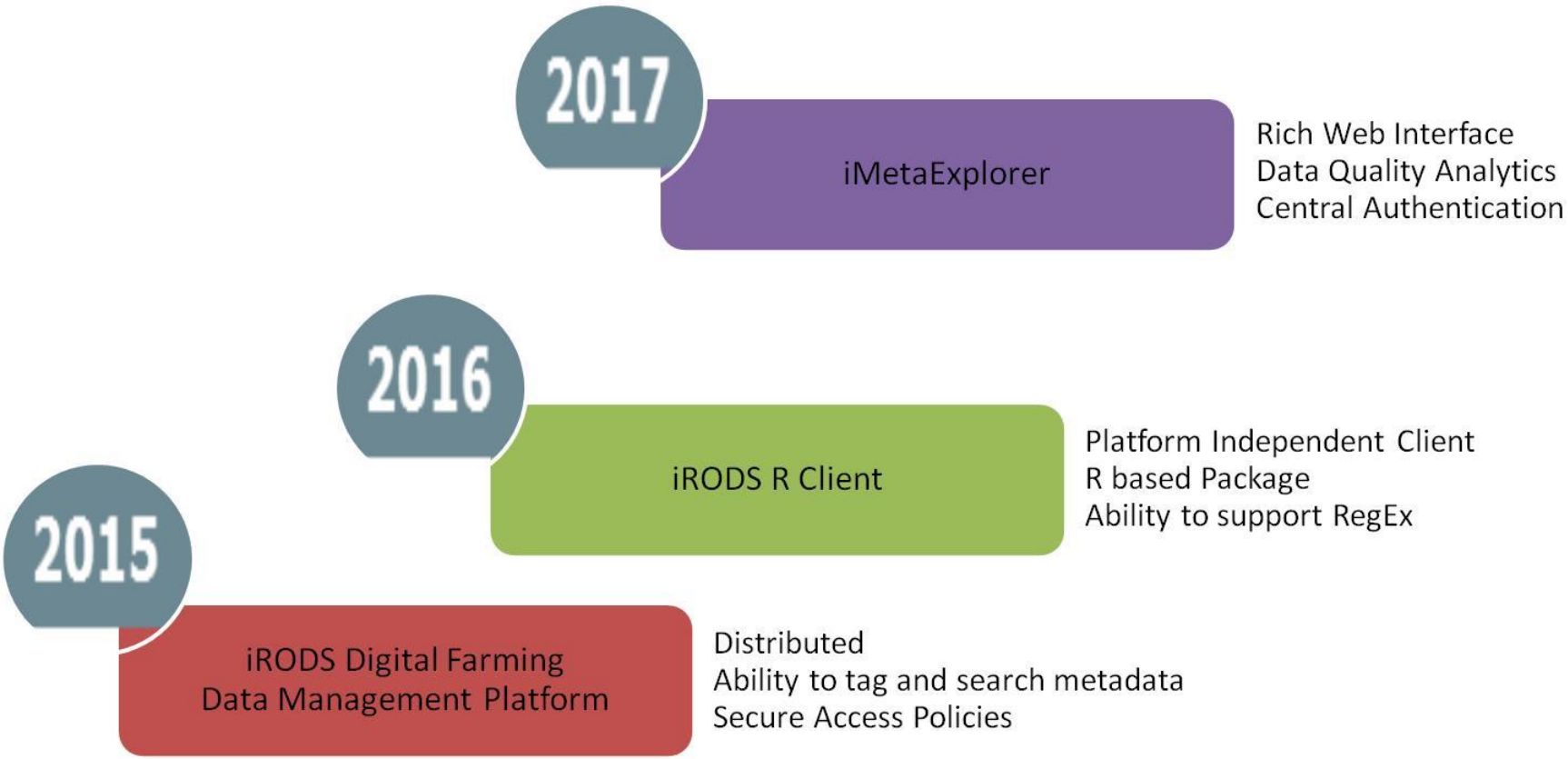
- Applications in Digital Farming tied up with these big data.
- Need to manage unstructured heterogeneous data.
- Data distributed across geographies.

iRODS , a powerful big data management system is a one stop solution for storing such unstructured distributed data and tag metadata to it.

- R-Scripts being used as a basis for any analysis. Major activities of any analysis is to pull in the data that is relevant to the analysis.
- So, reading and writing data to iRODS by passing an IRODSpath /uri from R-scripts, is a major requirement for data scientists, analysts.

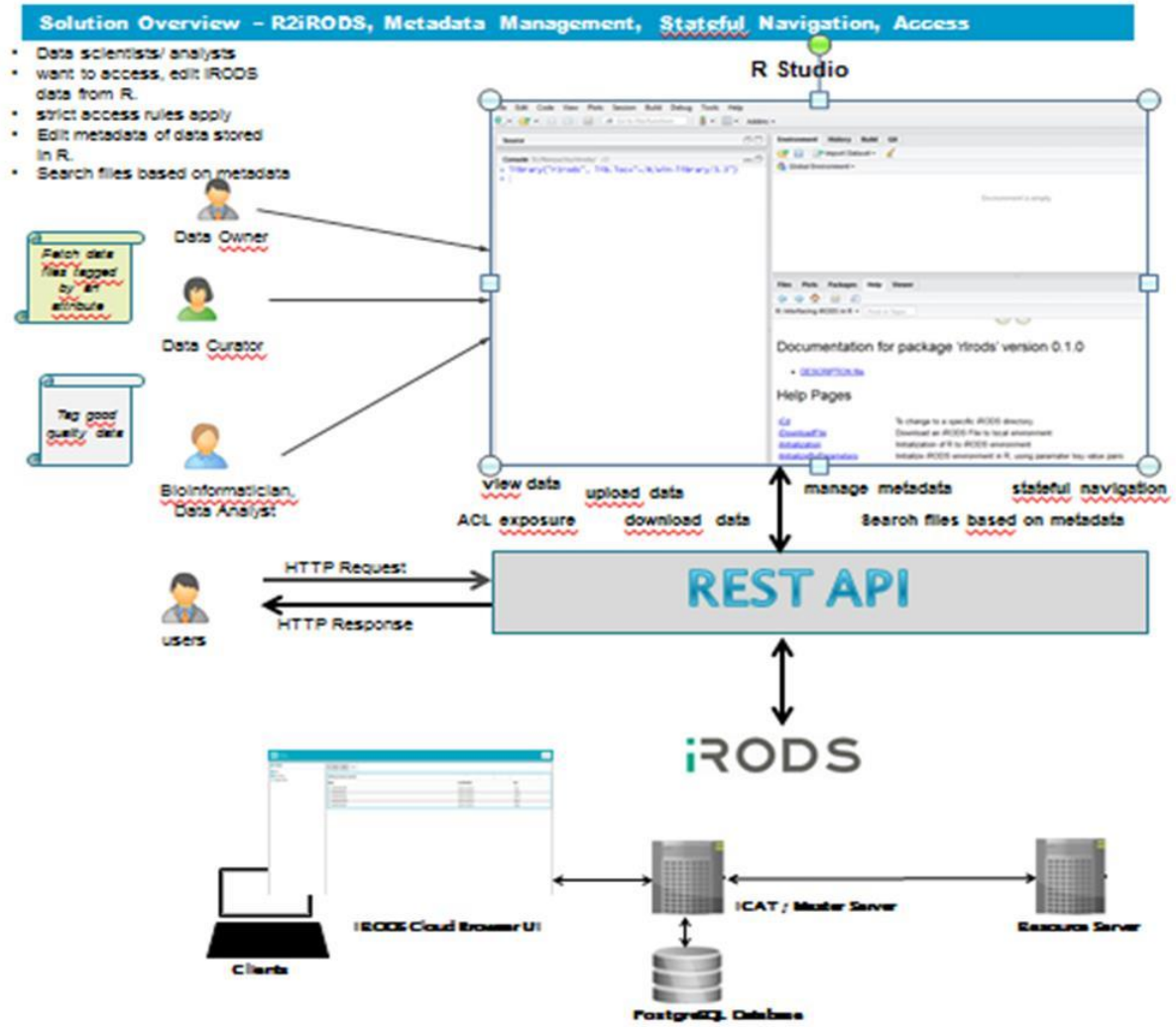
Our Solutions so far

Digital Farming Solutions



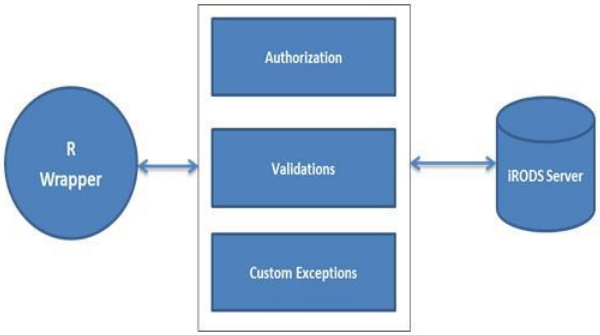
R-Client for iRODS (RStudio)

Digital Farming Solutions



Key Features of Implementation

- Distributed data management system with efficient data storage, data provenance.
- Stateful navigation of data from R.
- Store and retrieve data and metadata to and from distributed data storage.
- Metadata operations from R.
- Business-defined access rights.
- Configurable on different servers.
- Customized REST interface to iRODS server.
- Packaged application, based on R 3.3.1
- Scalable interface with powerful data view, upload, download, metadata management and search features.



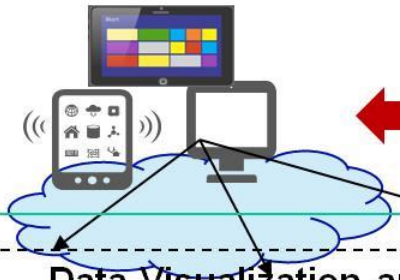
R-Client Functions

Digital Farming Solutions

iAddMetadata	iAddMetadata
iCd	To change to a specific iRODS directory
iDeleteAllMetadata	iDeleteAllMetadata
iDeleteAllMetadataByAttributeName	iDeleteAllMetadataByAttributeName
iDeleteSpecificMetadata	iDeleteSpecificMetadata
iDownloadFile	Download an iRODS File to local environment
iInitialization	Initialization of R to iRODS environment
iInitializeByParameters	Initialize iRODS environment in R, using paramater key value pairs
iList	Lists the iRODS directory struture
iListMetadata	iListMetadata
iMkdir	Create a directory at iRODSPath
iPullAllMetadata	iPullAllMetadata
iPullCustomMetadata	iPullCustomMetadata
iPwd	To view current iRODS directory
iRm	Delete a collection or data object at iRODSPath
iUpdateMetadata	iUpdateMetadata
iUpdateSpecificMetadata	iUpdateMetadata
iUploadFile	Upload file to iRODS server, using paramater key value pairs

iMetaExploreR – Shiny based Web Interface

Digital Farming Solutions



Data Visualization and Web Interface

Services

Rule Engine

Visualization

Geo Mapping

Metadata Associations

Cloud tagging

Search

Analytical Layer

Analytical Models

R

Java

iRODS – Data Management Framework

AWS

Master DATA

Analytical DATA

iMetaExploreR - Features

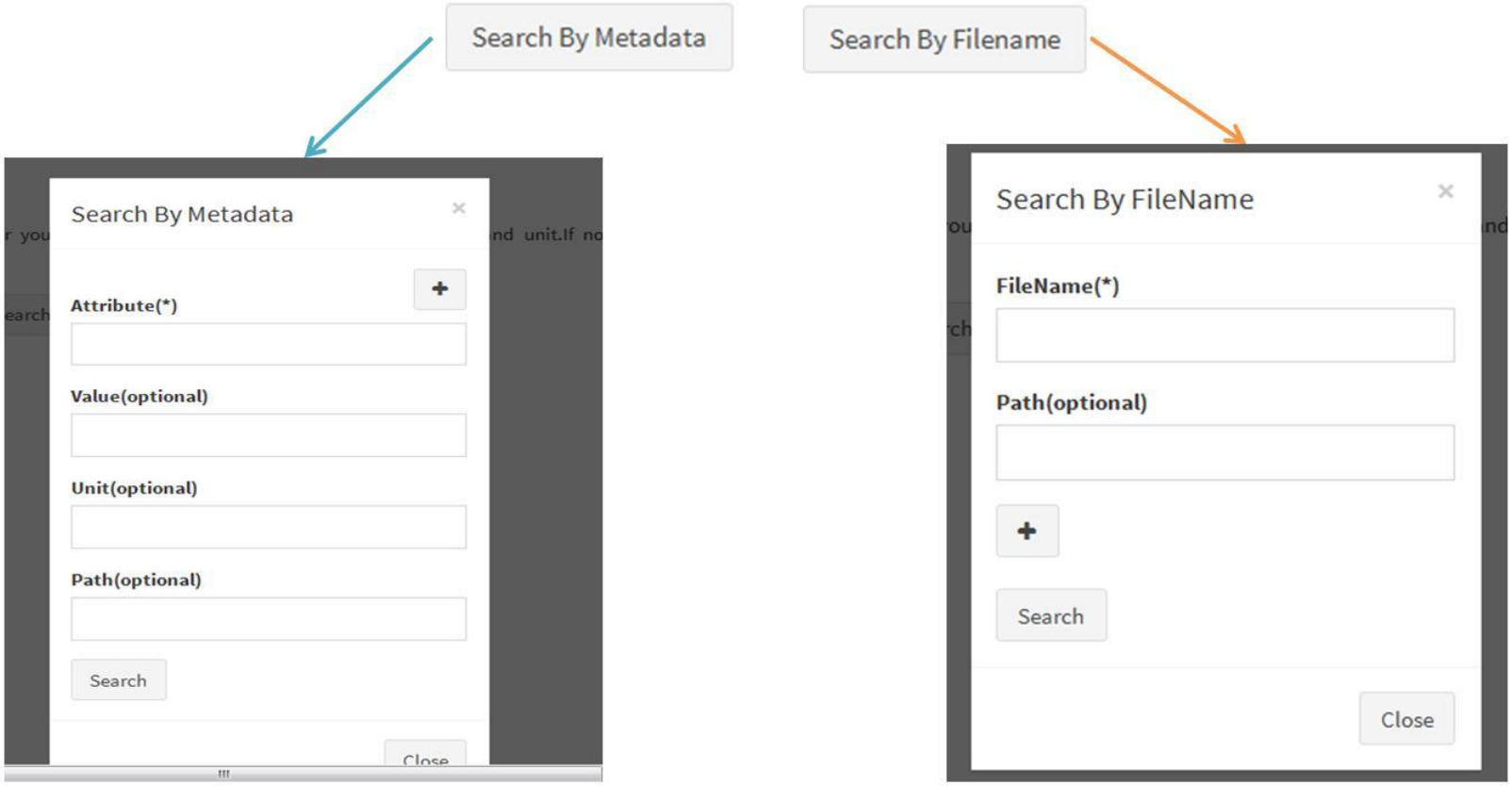
Digital Farming Solutions

- Log in
  - Fixed Users/ Roles
  - Central Authentication Service
- Easy Interaction with (file-based) Data Navigation friendly
  - File /Folder Explorer – Normal View, Tree View
  - Upload/Download files and folders
  - Create, Move, copy, rename, delete files and folders
  - File / Folder Bookmarking
  - Geo Metadata access
- Meta-data Operation
  - View
  - add, edit, delete
  - Bulk operations
- Interactive Data Exploration via Visual Analytics
  - Meta-data Cloud tagging
  - Key-Value Interaction Graph
- Search Functionality
  - File based
  - Meta-data based

iMetaExploreR - Search

Digital Farming Solutions

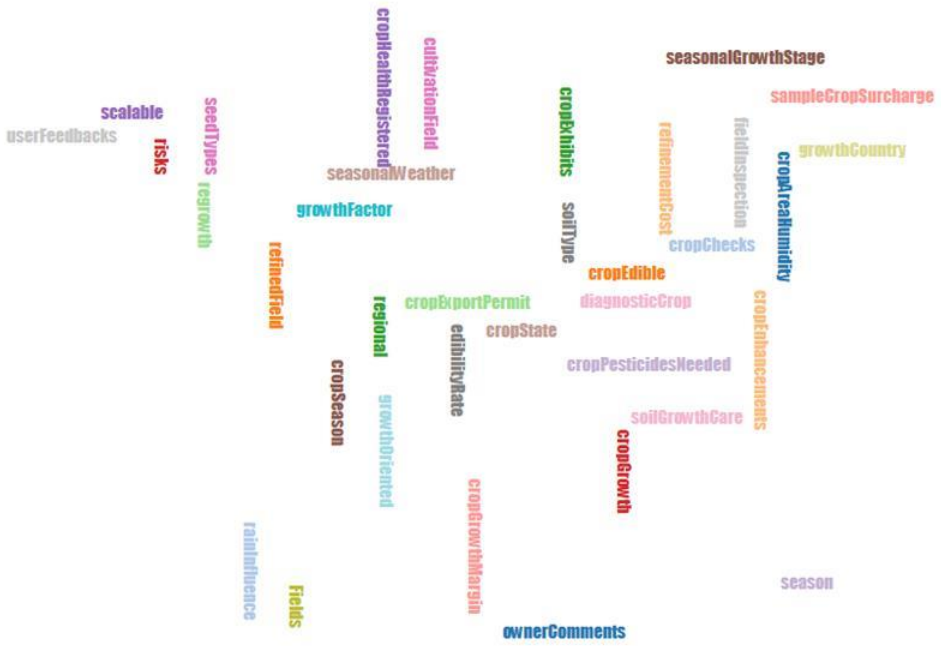
- Powerful search features from IRODS available in Front-End, API or Command Line Interface



iMetaExploreR – Meta Cloud Tagging

Digital Farming Solutions

- iRODS Meta-data of current search context shown as frequency/ word cloud
- Ask questions like
  - "what is most prominent meta-datum"
  - "What met-datum is rarely provided"
  - "Are there spelling variations in my meta-data"



iMetaExploreR – Geo Metadata Tagging

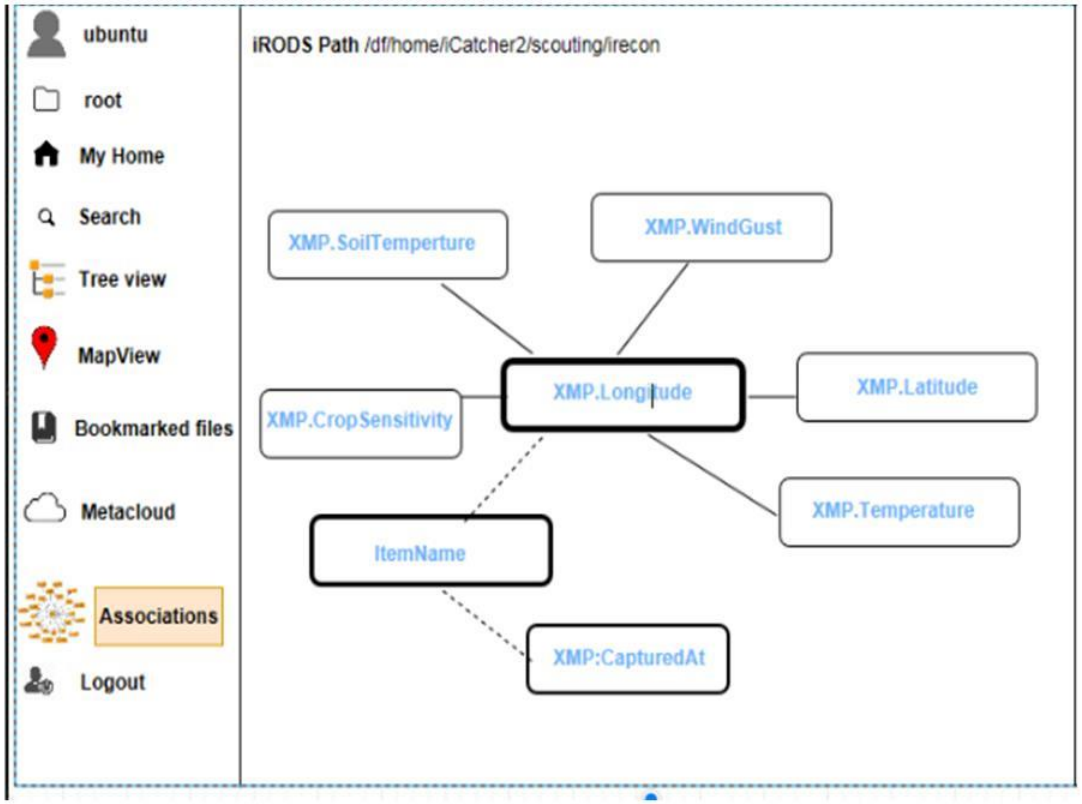
Digital Farming Solutions

- iRODS Meta-data of current search context geo-location shown
- Ask questions like
  - "From where is data provided"
  - "Who provides data"
  - "What locations are missing"
  - "Why does my data not show up on the map?"

iMetaExploreR – Network Association Graphs

Digital Farming Solutions

- iRODS Meta-data of current search context associations shown
- Explore Meta-data and ask
  - "which meta-data co-occurs?"
  - "what if my file misses a link between meta-data"
  - "what are meta-data islands"
  - "what topics do these islands represent"



Future Topics

Digital Farming Solutions

R-iRODS Package

- Internal CRAN Packaging.
- R Studio Integration – add ins support.

iMetaExploreR

- Central Authentication.
- Easy regex based metadata search from Shiny Web Application.
- Fully extensible support for data and visual analytics.

THANK YOU

Digital Farming Solutions