

TERRA-REF Data Processing Infrastructure

Maxwell Burnette

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
mburnet2@illinois.edu

Rob Kooper

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
kooper@illinois.edu

J.D. Maloney

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
malone12@illinois.edu

Gareth S. Rohde

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
srohde@illinois.edu

Jeffrey A. Terstriep

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
jefft@illinois.edu

Craig Willis

National Center for Supercomputing
Applications, University of Illinois at
Urbana-Champaign
willis8@illinois.edu

Noah Fahlgren

Donald Danforth Plant Science Center
nfahlgren@danforthcenter.org

Todd Mockler

Donald Danforth Plant Science Center
tmockler@danforthcenter.org

Maria Newcomb

University of Arizona, School of Plant
Sciences, Maricopa Agricultural Center
newcombm@email.arizona.edu

Vasit Sagan

Department of Earth and Atmospheric
Sciences, Saint Louis University
Vasit.Sagan@slu.edu

Pedro Andrade-Sanchez

University of Arizona, Maricopa
Agricultural Center
pandrade@email.arizona.edu

Nadia Shakoor

Donald Danforth Plant Science Center
nshakoor@danforthcenter.org

Paheding Sidike

Department of Earth and Atmospheric
Sciences, Saint Louis University
sidike.paheding@slu.edu

Rick Ward

University of Arizona, School of Plant
Sciences, Maricopa Agricultural Center
rickw@email.arizona.edu

David LeBauer

University of Illinois at Urbana-
Champaign
dlebauer@illinois.edu

ABSTRACT

The Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF) provides a data and computation pipeline responsible for collecting, transferring, processing and distributing large volumes of crop sensing and genomic data from genetically informative germplasm sets. The primary source of these data is a field scanner system built over an experimental field at the University of Arizona Maricopa Agricultural Center. The scanner

uses several different sensors to observe the field at a dense collection frequency with high resolution. These sensors include RGB stereo, thermal, pulse-amplitude modulated chlorophyll fluorescence, imaging spectrometer cameras, a 3D laser scanner, and environmental monitors. In addition, data from sensors mounted on tractors, UAVs, an indoor controlled-environment facility, and manually collected measurements are integrated into the pipeline. Up to two TB of data per day are collected and transferred to the National Center for Supercomputing Applications at the University of Illinois (NCSA) where they are processed.

In this paper we describe the technical architecture for the TERRA-REF data and computing pipeline. This modular and scalable pipeline provides a suite of components to convert raw imagery to standard formats, geospatially subset data, and identify biophysical and physiological plant features related to crop productivity, resource use, and stress tolerance. Derived data products are uploaded to the Clowder content management system and the BETYdb traits and yields database for querying, supporting research at an experimental plot level. All software is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

PEARC '18, July 22–26, 2018, Pittsburgh, PA, USA
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

978-1-4503-6446-1/18/07...\$15.00

<https://doi.org/10.1145/3219104.3219152>

open source² under a BSD 3-clause or similar license and the data products are open access (currently for evaluation with a full release in fall 2019). In addition, we provide computing environments in which users can explore data and develop new tools. The goal of this system is to enable scientists to evaluate and use data, create new algorithms, and advance the science of digital agriculture and crop improvement.

CCS CONCEPTS

• **Computer systems organization** → **Architectures**

KEYWORDS

Phenotyping, Container-based analysis environments, open science

ACM Reference Format:

Maxwell Burnette, Rob Kooper, J.D. Maloney, Gareth S. Rohde, Jeffrey A. Terstriep, Craig Willis, Noah Fahlgren, Todd Mockler, Maria Newcomb, Vasisht Sagan, Pedro Andrade-Sanchez, Nadia Shakoor, Paheding Sidike, Rick Ward, and David LeBauer. 2018. TERRA-REF Data Processing Infrastructure. In *PEARC'18: Practice & Experience in Advanced Research Computing, July 22-26, 2018, Pittsburgh, PA, USA*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3219104.3219152>

1 INTRODUCTION

Automated sensing is increasingly important in the domain of agriculture. In the context of crop improvement, automated collection of plant trait measurements helps scientists link observable traits with underlying genetic and environmental drivers. New sensors and sensing platforms have enabled increased spatial and temporal frequency of observation and increased diversity, resolution, and volume of data collection. These data streams require modern high-performance computing (HPC) systems capable of efficiently processing and organizing massive amounts of data. Researchers are interested in leveraging these new tools and capabilities. However, much further investment by academic and commercial research programs, both in physical hardware and in software and technical expertise, is required.

The goal of the Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF) is to provide an open reference dataset, open source software, and accessible computing to enable researchers to evaluate and leverage emerging sensing and genomic technologies. Our sensing platforms include a 30-ton custom-built robotic scanning system constructed over an experimental field near the Maricopa Agricultural Center at the University of Arizona (MAC), a controlled-environment platform at the Danforth Center in St. Louis, and aerial, ground, and handheld sensing devices. The data pipeline transfers large quantities of data to the National Center for Supercomputing Applications at the University of Illinois (NCSA), where the computing pipeline

processes and publishes data, and provides multiple interfaces for data access and analysis. Our pipeline can process and publish data within forty-eight hours of collection and data are made available for cloud-based analysis and algorithm development.

One of the most powerful assets of these data is the value they add to genomic information available for phenotyped germplasm [5]. In the case of TERRA-REF, over 350 accessions of bioenergy sorghum germplasm are being re-sequenced, and this same germplasm is grown and phenotyped in the two primary platforms described here.

Here we describe the complete pipeline involved in the collection, storage, transfer, dissemination, and analysis of these data. Section 2 describes the data collection platforms, including the robotic scanner. Section 3 describes the automated data transfer and computing pipeline. Section 4 describes the interfaces for accessing data; these include GUIs, APIs, R and Python utilities, and the cloud-based 'workbench' with pre-configured containerized development environments that provides users with direct access to the file system and databases. Section 5 describes future development and potential applications for the TERRA-REF data and computing pipeline.

2 DATA COLLECTION

The pipeline was developed to handle data from diverse sensors and sensing platforms. Our sensor platforms include a high-resolution gantry system, a conveyor and cabinet-based imaging facility, UAVs, tractors, and hand-held instruments. These platforms represent trade-offs among factors such as cost, weight, resolution, and measurable area.

2.1 High Resolution Field Scanner

The primary source of TERRA-REF data is the LemnaTec Field Scanalyzer (LemnaTec GmbH, Aachen, Germany, Figure 1), an automated gantry-mounted robotic scanning system with a suite of plant and environmental sensors that scan a 200 m x 20 m (1 acre) field at the MAC field station in Arizona. The field scanner system moves a sensor box along three axes to collect high-resolution imagery from the experimental plots underneath. The sensor box contains RGB, thermal, hyperspectral and active fluorescence cameras as well as environmental sensors (Table 1).



Figure 1: Scanalyzer system (with sensors in white box) operating on 200m steel rails in Maricopa, Arizona.

² <https://github.com/terraref>

Movement of the gantry is controlled by pre-scripted instruction sets which vary based on the crop being grown, the sensors being used, and the time of day/season. A human operator triggers these scan scripts to support particular experimental foci. Not every sensor is operating for every scan, so records are kept of the scanning schedule in order to inform which data products are available at a particular time. Table 1 has detailed descriptions of several of these sensors and their data specifications. Sensors are calibrated and validated using a set of control target objects and panels.

Table 1: Sensors Mounted in Sensor Box

Sensor	Range/Resolution
Hyperspectral imagers	380-1000 nm
	900-2500 nm
Thermal infrared	-40 to 150 C
Downwelling radiance spectrometer	337-824 nm
Laser scanners	0.3-0.9 mm
Stereo cameras	3-band RGB
Environmental sensors	Wind, Air Temperature, Humidity, Light, Pressure, Precipitation, CO2
PSII fluorescence	680-900 nm
NDVI	633 nm / 800 nm
PRI	531 nm / 569 nm

As data is collected, it is staged on a 65 TB cache server for transfer from the field in Arizona to computers at NCSA in Illinois. Section 3 describes this data transfer pipeline.

2.2 Controlled-Environment Facility

An indoor controlled-environment phenotyping system, a LemnaTec Greenhouse Scanalyzer 3D-HT (LemnaTec GmbH, Aachen, Germany), at the Donald Danforth Plant Science Center in St. Louis, Missouri provides an automated conveyor belt system that moves individual plants through a series of sensor boxes with RGB and near-infrared imagers. The data is transferred to NCSA as part of the same pipeline and processed by PlantCV software [6].

2.3 Hand and Other Measurements

A variety of more conventional measurements are hand-collected and made available alongside sensor-derived measurements. Hand measurements are taken directly by scientists in the field. These measurements are used for derived data validation and for training and testing new algorithms. Hand measurements include plant physiological traits related to yield, resource use, and stress tolerance. These measurements are

recorded and entered into the BETYdb traits database alongside measurement protocols and experimental metadata (see Section 4.2 for more details).

Unmanned aerial vehicles (UAVs) include both fixed and rotary wing drones whereas phenotracors are modified sprayer tractors that move sensors through the field [1]. These platforms collect data at the field sites in Arizona and Ashland, Kansas. Unlike the field scanner, these platforms more closely reflect the cost and capabilities of systems that are typically used in an agricultural setting. They collect aerial imagery and proximate sensing measurements at lower resolution but are capable of covering larger areas and being moved around to different fields. In Arizona, georeferenced orthomosaics from UAVs help validate plot boundaries for the Field Scanner plots.

3 DATA TRANSFER AND PROCESSING

Raw data and metadata is collected on computers attached to sensors in the field. It is then transferred for processing. Here we will focus on the large volume data stream from the field scanner. Other data sources follow similar pipelines with varying levels of automation.

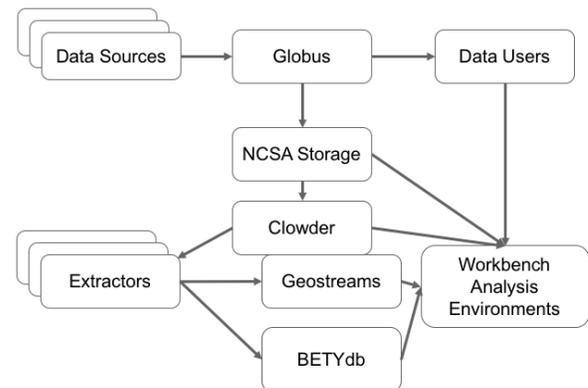


Figure 2: Simplified representation of data flow from source to long-term storage to extractor and analysis pipelines.

Data collected by the field scanner is transferred to a local 65 TB cache server where it is transferred to NCSA within twenty four hours of collection. At typical data collection rates, roughly two months of data can accumulate so that the pipeline can withstand hardware and internet failures between the sensors and the computing pipeline.

3.1 High-Throughput Low Latency Data Transfer Pipeline

TERRA-REF uses Globus to transfer data to NCSA. Globus uses a GridFTP framework to schedule transfers between two Globus endpoints with built-in checksum verification, batch transferring, automatic pause/resume/retry loops, and an API to query transfer status [7]. To support the large data volume

coming from the field scanner, there is a dedicated 1 GbE connection between the cache server and the outside world via Internet2 [14].

As incoming entries appear on the cache server FTP log, a lightweight service bundles the new files into Globus transfers and initiates them. At the same time, an HTTPS notification is sent to a sister service at NCSA with a secure API to tell the receiving end of the pipeline that new files are on the way. Under normal operations, files are typically transferred to the NCSA storage condo within hours of collection.

The receiver service at NCSA will monitor Globus transfers from MAC until the status is complete. When the Globus transfer is complete the data is uploaded to the Clowder data management system (see 4.1) and appended to the appropriate processing queues (see 3.2). It is also after this successful transfer that data is backed up onto Blue Waters Nearline tape-based storage for long-term preservation.

3.2 Data Processing Architecture

The field scanner system generates data in raw formats, primarily BIN and XML files written alongside JSON metadata. The metadata files describe the sensor, sensor platform, collection methodology, and variable parameters such as time, location and scan script. The data processing pipeline extracts spatial and temporal metadata and uses these to georeference data streams. The output of this pipeline is a suite of files and database records that follow open geospatial consortium (OGC) format standards and Climate and Forecast (CF) conventions. These formats support scalable computing and community standards and conventions where they exist in order to support modularity and integration with existing pipelines, software, and databases.

TERRA-REF uses a suite of custom processing modules to convert data formats and derive new data and traits for downstream use. In the Clowder framework these modules are called extractors because they extract or derive data and metadata from incoming sources. Extractors are designed to operate on a single file or dataset and accomplish a task such as converting from BIN to GeoTIFF or translating from JSON scanner-based coordinates to geographic coordinates for georeferencing. Data generated by an extractor can be posted back into Clowder, which can trigger additional extractors in a workflow.

In order to reduce the amount of duplicated code and minimize resulting errors, most TERRA-REF extractor containers are based on a shared base container pre-built with PyClowder [8] and Terrautils [2] Python libraries. By using a shared base, updates to these libraries only need to be performed once on that base and all extractors benefit. Containers in this context are lightweight virtual images similar to a virtual machine that have all dependencies and libraries installed so there is no need to maintain or install them on every server or computer that runs the extractors.

The PyClowder, Terrautils and science package Python libraries contain utility functions for interfacing with TERRA-REF databases and pipelines: PyClowder for Clowder

downloading and uploading, Terrautils for trait generation and TERRA metadata manipulation, and sensor-specific science packages for interacting with sensor data directly with minimal overhead. Any new code specific to our pipeline that can be generalized to be shared across multiple extractors is considered for generalization and inclusion in the Terrautils or science libraries.

Because the extractors are designed to be atomic in operation and work independently of other extractors, it is possible to deploy many instances of a given extractor to distribute processing workload for an incoming set of data across many worker nodes. TERRA-REF uses a RabbitMQ message bus to notify the cluster of distributed extractors that a new job is available for processing and allows extractors to claim and process jobs from the message stack in parallel for efficient throughput. If resources are limited, the size of these extractor clusters will vary depending on the current length of the queue and the requested needs of end users.

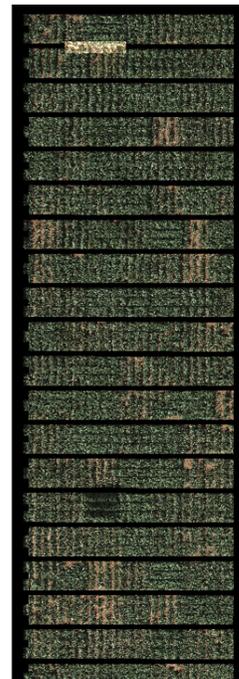


Figure 3: A partial image of a field-level mosaic composed of thousands of images from the stereo RGB camera.

We use Kubernetes³ to manage and scale extractors across available computing resources. Kubernetes is an open-source container orchestration system that facilitates deployment and management of many such extractor nodes in an efficient manner. The approach used by TERRA-REF is to package extractor code and libraries into Ubuntu-based Docker containers⁴, allowing them to be deployed immediately in any

³ <https://kubernetes.io/>

⁴ <https://hub.docker.com/u/terraref/>

Docker-supporting environment without the need to install complex dependencies or monitor local software changes. By using an architecture leveraging Docker and Kubernetes, extractor code can be tested and packaged prior to deployment and the actual production deployment process is very fast.

With Kubernetes it is easy to scale a particular extractor as well as to manage the entire suite of deployed extractors. For example, if a particularly data-intensive scan is performed during the week and a small number of extractor instances is too slow to process all the resulting datasets, additional instances can be deployed in order to meet the additional demand until the backlog is reduced.

3.3 Logging and Provenance

With the complex chain of data collection, transfer, cleaning, upload to Clowder, and distribution through extraction pipelines, a robust plan is needed to track datasets through this process and make the process of debugging and issue resolution easier. TERRA-REF utilizes several occasionally redundant systems for logging and provenance tracking.

TERRA-REF utilizes the Check_MK monitoring system to track and send notifications regarding the health and status of virtual machines and critical systems in the pipeline, including the main public-facing Clowder and BETYdb portals. If there are issues with these systems, administrators will be notified and can respond quickly.

A Postgres database is utilized underneath services on both the MAC cache server and the NCSA receiving endpoint to track Globus transfers. Because entries are added to the database immediately upon discovery and statuses are updated with each critical step (initialization of transfer, completion of transfer, notification to NCSA API, upload into Clowder), the system can effectively enter a holding pattern if, for example, Globus is temporarily unreachable. When connections are restored the service will go back and begin handling incomplete transfers at the next appropriate stage.

A centralized Logstash instance atop an Elasticsearch database is used to collect a variety of system and extractor logging messages that can be queried and visualized using Kibana (together collectively referred to as the ELK stack). Logging messages include relevant information such as dataset identifiers, error text, generated filenames, and progress updates. It is possible to filter these messages to focus on the journey of one specific dataset through several extractors—by timestamp, for example. Simple wrapper methods for writing to the Logstash instance are included in Terrautils so that all collaborators in TERRA-REF can benefit from this resource.

In addition to logs, metadata is inserted into Clowder after processing steps so that files and datasets have a record of when various extractors were triggered and any failures or resulting notable outputs that resulted. This way, users who download data from Clowder will also see a history of what was done to those data and any related outputs and derived properties that were created.

3.4 Standardization and Quality Assurance

As the first step in uploading datasets into the Clowder data management system (see 4.1), the JSON contents of the corresponding metadata files are read, standardized, and cleaned. The cleaning step attempts to convert the available metadata fields and properties into CF standards⁵, provide consistent names and units, and precompute some basic information about the context of the capture (for example, associated plot IDs or status markers from the experiments being conducted). Extraneous fields are omitted from the cleaning process. The cleaned metadata is then indexed in a Clowder-specific Elasticsearch index for search and retrieval by metadata.

Currently in addition to CF standards for climate and forecasting metadata, OGC standards⁶ are used for geospatial data to ensure consistent handling of location across data sources. These standards help to ensure interoperability with outside environments that consume environmental data such as GeoTIFF imagery. TERRA-REF has future plans to support ICASA⁷ and BrAPP⁸ standards for further interoperability with scientific communities.

To assess the usability of sensor data for downstream analysis, a suite of data quality metrics (QM) is also incorporated into the data processing pipeline. The QM will be generated either from raw or the processed data by extractors and reported to the corresponding JSON metadata. The output of QM indicates whether an input data lies in the confidence interval in terms of image quality considering various aspects, such as blur effect [4] and focus measures [12], which can be subject to optical artifacts, scanning mechanism and weather conditions. The results of this module essentially provide quantitative summary associated to the dataset quality and uncertainty estimates.

The trait database BETYdb [9,13] restricts trait data to valid ranges and exerts database level constraints on metadata validity and consistency. A dynamic web visualization interface⁹ supports quality control as well as summaries of data availability.

4 DATA DISSEMINATION AND ANALYSIS

The goal of TERRA-REF is not only to collect and process data, but also to make it available to the scientific community. Access to the data for further research is a critical component, and the data volumes represented in this pipeline make it impractical to download and analyze significant amounts of our full-resolution sensor data products. To that end, we provide a variety of web, API, and scripting interfaces that enable users to find and combine different types of data and metadata.

Globus not only supports our data transfer pipeline, it also allows end-users to browse and transfer files located on our file system. While these files have geospatial metadata, they generally lack sensor, provenance, and experimental context. We

⁵ <http://cfconventions.org/>

⁶ <http://www.opengis.org/standards>

⁷ https://dssat.net/data/standards_v2

⁸ <https://brapi.org/>

⁹ <http://traitvis.workbench.terraref.org/>

use a suite of databases to store metadata and the output of downstream analyses.

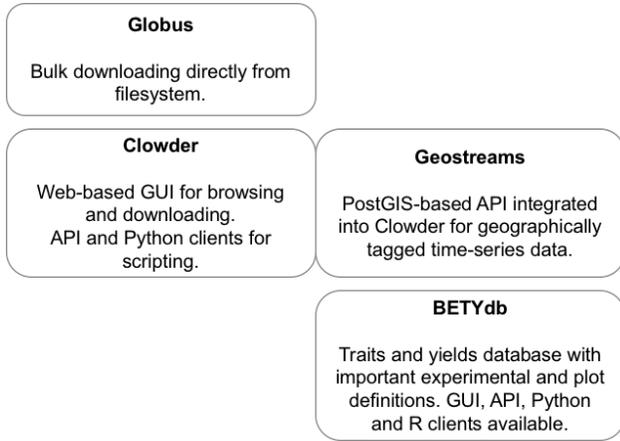


Figure 4: Summary of four primary interfaces in TERRA-REF project for data access.

4.1 Clowder Frontend for Management

TERRA-REF uses Clowder as a web frontend for browsing and downloading sensor data. Clowder is open-source research data management software developed at NCSA [10,11]. Clowder supports in-depth permissions models and user groups and allows users to annotate file metadata as well as create and publish data collections¹⁰. Clowder has two database backends: a Mongo database for storage of unstructured metadata and a PostGIS database for geospatial data. Both of these databases store pointers to files on the filesystem and to metadata in other databases. These databases can be accessed via the web interface or via an API that supports querying, retrieval, and interaction with a RabbitMQ message bus for triggering data processing extractors as data arrive and are generated.

When data arrives at NCSA, it is uploaded into the Clowder database. Collections of sensor data are nested in a hierarchy by processing level, sensor, date, and time. Clowder automatically sends messages to RabbitMQ based on events defined by the file types and associated metadata that arrive or on other criteria so that relevant extractors begin work on the newly created datasets without manual intervention. As downstream data products are generated and posted into Clowder, new extractors are triggered in a similar fashion. In this way, Clowder and RabbitMQ provide a flexible and lightweight framework for automating the workflow.

At the dataset level users can view metadata, download files, check process event history, and trigger additional process executions. Clowder has a search interface¹¹ that supports finding data by metadata field, location, or name.

¹⁰ <https://clowder.ncsa.illinois.edu/>

¹¹ <https://terraref.ncsa.illinois.edu/clowder/>

It is also here that users can search through time series data using the experimental plot-based Geostreams API to view traits over time, filtered by location, and contribute new files and datasets. This resource can be used, for example, to quickly identify all images overlapping a particular plot between particular times and any traits derived from those images

4.2 Trait Data Access Interfaces

Plant and plot level traits and agronomic and experimental metadata are stored in a BETYdb PostGIS database [9,12]. BETYdb provides web, API, R, and Python interfaces for searching and downloading files of trait data and agronomic metadata. Section 4.1 describes the Clowder-associated Geostreams for additional time-series data.

BETYdb provides time series of trait data aggregated to the level of plant (for the indoor system) or plot and sub-plot (for field data). BETYdb also includes experimental metadata such as agronomic inputs and experimental design, including the boundaries and identities of plant varieties in each plot. Extractors post newly derived traits to BETYdb tagged with appropriate metadata and citation information. Automated constraints implemented in the API and the database constrain data to valid ranges and ensure consistency of metadata records across tables. TERRA-REF has contributed to the R trait package that facilitates access to data from TERRA-REF and other projects that leverage the BETYdb software (e.g. pecanproject.org) and other plant trait databases [3].

4.3 Cloud Analysis Environments

To avoid asking users to download large volumes of data, TERRA-REF also offers an innovative option for working with Clowder and BETYdb data: The National Data Service at NCSA has developed Labs Workbench¹² as an open-source web platform for working directly with data in place.

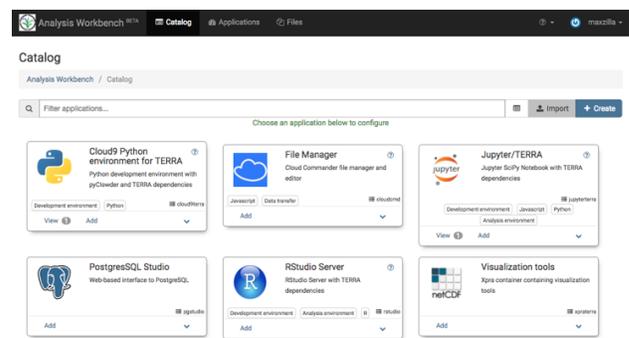


Figure 5: A selection of containers with prebuilt tool suites are available to Labs Workbench users to launch.

Users with access to Labs Workbench can select from a list of preconfigured container images that are loaded with libraries,

¹² <http://www.nationaldataservice.org/platform/workbench.html>

packages, and software to interpret and work with data formats present in the TERRA-REF pipeline. The entire TERRA-REF file system folder structure is mounted into each container so users have access to petabytes of data instantly.

Labs Workbench supports GUI-based software such as Quantum GIS in the browser as well as the more common Jupyter Notebooks and RStudio. This is a powerful resource not just for scientists, but also for collaborators and developers who want to develop software with TERRA-REF data but cannot easily scale their resources for testing and development.

Containers can be launched and shut down at will and new applications can easily be added to the Labs Workbench catalog as new needs arise from the collaborative network.

5 Conclusions

TERRA-REF is an ambitious effort spanning the collection, transfer, processing, distribution and analysis of an unprecedented agricultural dataset. With a broad offering of data sources and sensors at several distinct locations and extremely high spatial and temporal resolutions, this project aims to enable new breakthroughs in the area of crop improvement and precision agriculture. By making our data open access and our software open source, TERRA-REF aims to catalyze new advances and collaborations across agricultural, biological and engineering disciplines in both public and private sectors.

The data processing architecture created for TERRA-REF focuses simultaneously on reliability, speed, reproducibility and accountability alongside modularity, interoperability and usability throughout the pipeline. Development of TERRA-REF is ongoing and open to contributions and new applications. Moving forward we will add new data products, algorithms, and exchange formats to support the diverse needs of this emerging research community.

ACKNOWLEDGEMENTS

The work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000598.

REFERENCES

- [1] Andrade-Sanchez Pedro, Gore Michael A., Heun John T., Thorp Kelly R., Carmo-Silva A. Elizabete, French Andrew N., Salvucci Michael E., White Jeffrey W. Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology* 41, 68-79. 2013.
- [2] Max Burnette, Craig Willis, David LeBauer, Nick Heyek, Wei Qin, & Solmaz Hajmohammadi. (2018, March 16). terraref/terrautils: alpha release (Version 1.1.0-alpha). 2018.
- [3] Scott Chamberlain, Zachary Foster, Ignasi Bartomeus, David LeBauer and David Harris (NA). traits: Species Trait Data from Around the Web. R package version 0.2.0.9415. <https://github.com/ropensci/traits>
- [4] Crete, Frederique, Thierry Dolmiere, Patricia Ladret, and Marina Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. *Human vision and electronic imaging XII*, vol. 6492, p. 64920I. International Society for Optics and Photonics, 2007.
- [5] Robert T. Furbank, Mark Tester. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, Volume 16, Issue 12, 2011.
- [6] Gehan MA, Fahlgren N, Abbasi A, Berry JC, Callen ST, Chavez L, Doust AN, Feldman MJ, Gilbert KB, Hodge JG, Hoyer JS, Lin A, Liu S, Lizárraga C, Lorence A, Miller M, Platon E, Tessman M, Sax T. PlantCV v2: Image analysis software for high-throughput plant phenotyping. 2017.
- [7] I. Foster. Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2005.
- [8] Rob Kooper, Max Burnette, Sandeep Satheesan, Bing Zhang, Todd Nicholson, Indira Gutierrez, ... Ward Poelmans. (2018, March 23). Python wrappers for use with the Clowder system (Version 2.0.2). Zenodo. <http://doi.org/10.5281/zenodo.1206210>
- [9] LeBauer D, Kooper R, Mulrooney P, Rohde S, Wang D, Long SP, Dietze MC. BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. *GCB Bioenergy*. 2017 Jan 1;10(1):61-71.
- [10] Luigi Marini; Rob Kooper; Indira Gutierrez; Constantinos-Sophocleous; Max Burnette; Michal Onrejek; mfelarca; Brock Angelo; qqmyers; Eugene; Jong Lee; Will Hennessy; jansz; Bing Zhang; Yibo Guo; Gregory Jansen; yxcmax; tenczar; Varun Kethineedi; nmayudu; mkavinashkumar; wpoely86; Chris Navarro; mpitel. (2018, March 12). ncsa/clowder v1.3.5 (Version v1.3.5). Zenodo. <http://doi.org/10.5281/zenodo.1196565>
- [11] Luigi Marini, Indira Gutierrez-Polo, Rob Kooper, Sandeep Puthanveetil Satheesan, Maxwell Burnette, Jong Lee, Todd Nicholson, Yan Zhao, and Kenton McHenry. 2018. Clowder: Open Source Data Management for Long Tail Data. In PEARC '18: Practice and Experience in Advanced Research Computing, July 22–26, 2018, Pittsburgh, PA, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3219104.3219159>
- [12] Pertuz S, Puig D, Garcia MA. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*. 46(5):1415-32, May 2013.
- [13] Scott Rohde, Carl Crott, David LeBauer, Patrick Mulrooney, Rob Kooper, Jeremy Kembal, Jimmy Chen, Andrew Shirk, Zhengqi Yang, Max Burnette, MarvinJ, phenolphtalein, Uday Saraf, Michael Dietze, Chris Black. PecanProject/bety: BETYdb 4.20. Zenodo; 2018. doi: 10.5281/zenodo.1199667
- [14] Angela Stark. Arizona's Sun Corridor Network Announces Two 100 Gigabit Connections to Internet2. *Internet2*. Apr 08, 2015. <https://www.internet2.edu/news/detail/8109/>