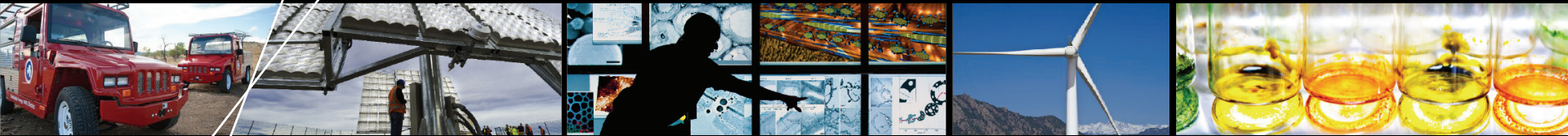


Designing and Developing a Collaborative Data Infrastructure for DuraMat



Robert R. White & Kristin Munch

2017-05-23

DuraMAT / BAPVC Workshop - Stanford

Agenda of talk

- What?
- Why?
- How?
- Where?
- When?





Energy Materials Network

U.S. Department of Energy



ElectroCat
Electrocatalysis Consortium

LightMAT



Lightweight Materials Consortium



HydroGEN
Advanced Water Splitting Materials



ChemCatBio
Chemical Catalysis for Bioenergy



DuraMAT
Durable Module Materials Consortium



CaloriCool™
CALORIC MATERIALS CONSORTIUM

Why do we need a DataHub?

Current Common Accessible Technology

- Drop Box, or similar software
- Email
- Sharepoint or similar multi-user document version sites

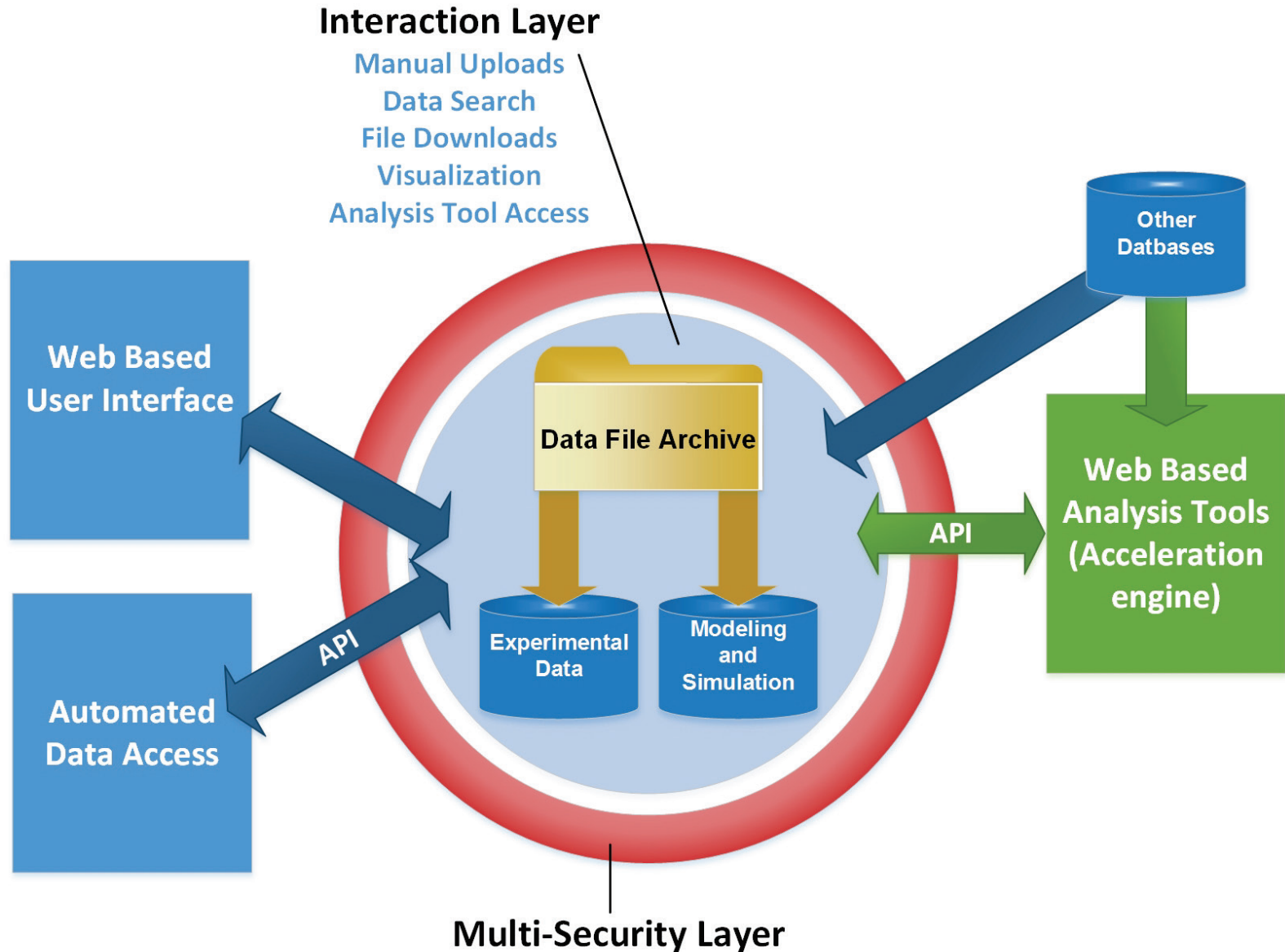
The newer software systems must be science-enabling and provide:

- Accessibility
- Security
- Longevity
- Reusability



- **Manage research data.**
- ***Accessibility to historical and time-series research data.***
- **Aggregation of metadata and data into a searchable database.**
- **Facilitate access to advanced data tools and analysis. (Acceleration Engine)**
- **Ease of access to consortium and partners.**
- **Secure the data at levels from proprietary, to embargoed, to publically available.**

Criteria Driven Design



Evaluation Criteria

- **Deployment (installation, prerequisites, upgrades)**
- **Cost (open source?, fees, maintenance)**
- **Industry standard programming languages and backend databases**
- **File Support**
- **Datasets (File grouping, Identifiers, QC)**
- **Data Analytics (plots, customizable, Site specific analytics, preview data)**
- **Data Organization (structure, methods)**
- **Storage (limits, locations, backups)**
- **Metadata (customizable, schemas, required fields)**
- **User Access (customizable, restricted vs. public, access methods)**
- **Framework Programming Language**
- **Application Program Interface (API)**
- **Import Data (Limits, customizable, methods, curation)**
- **Export data (limits, Single or Multiple, methods, specify target)**
- **Extensibility (plugins, customizable, plugin programming language)**
- **Version Control (Datasets, archives, toolsets/plugins)**
- **Search (faceted, browsing options, open access)**

Platforms Evaluated

- Granta
- DSpace
- Materials Commons
- Fedora Commons
- CKAN
- DKAN
- Socrata
- Materials Data Facility
- Dataverse
- Digital Commons
- Hub Zero
- Citrine
- NIST Materials Data Curation System
- NREL Research Data System

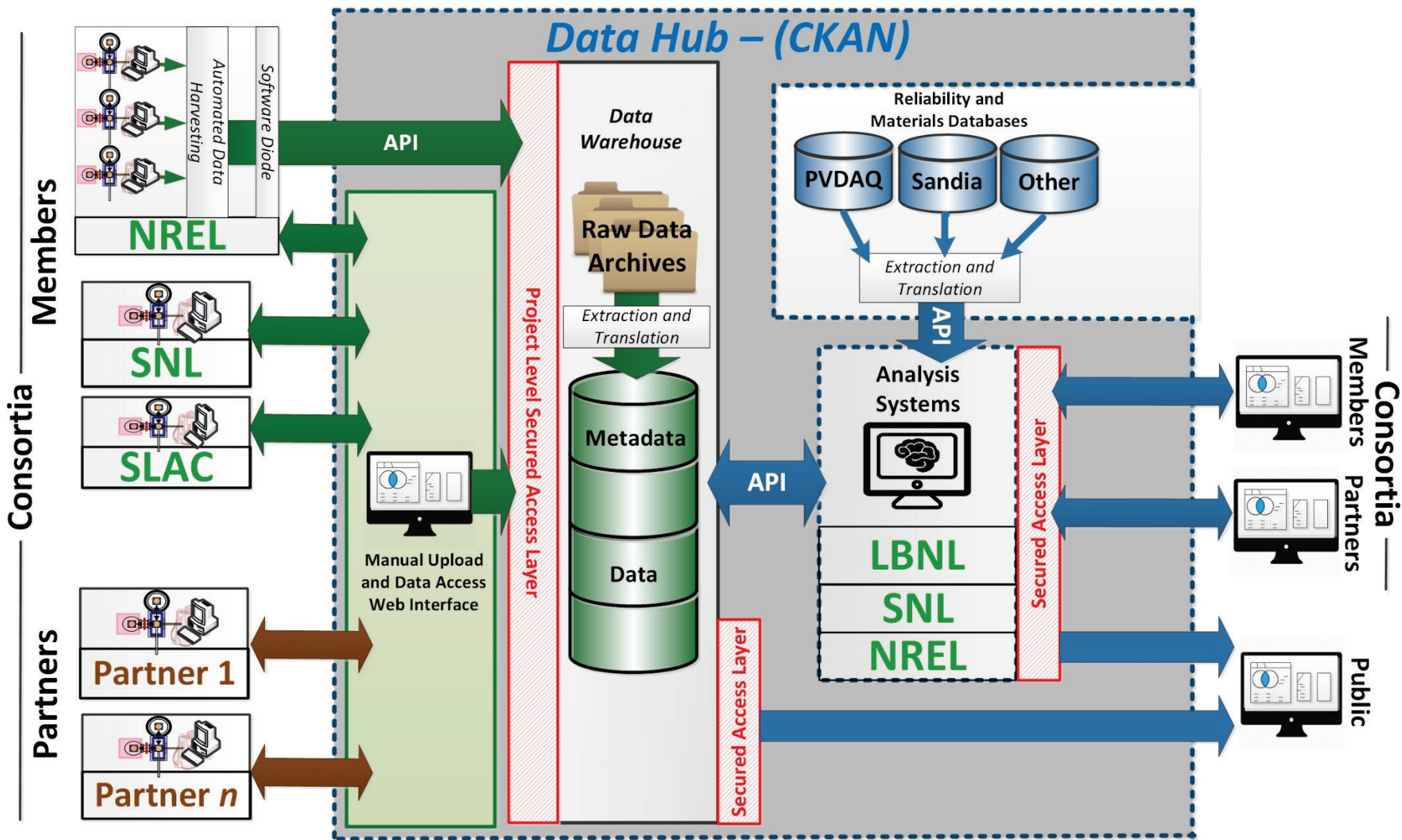
CKAN

- Very large user base
- Simple Interface
- Slow to deploy, but easily customizable
- Easy linked to Outside Web tools
- Built in simple visualization, expanded by custom plugins
- API

HubZero

- Large user base
- Complex interface
- Fast Deployed, harder to customize
- More difficult to link to outside tools
- No built in visualization, but plugins possible
- API

DuraMAT Framework



Welcome - DuraMat | x

ckantest.shadow.org

Apps Google Gmail NREL:theSOURCE FOM STM Campus - Flo Astronomy Collaborations Other bookmarks

Robert White 0

DuraMAT

Durable Module Materials Consortium

Datasets Organizations Groups About Search


Search data

E.g. environment

Popular tags reliability nrel ASU

Welcome to the DataHub

This is the site supporting data management and access for the collaboration members and the public.



DuraMat Data Hub statistics

4	3	0	0
datasets	organizations	groups	related items

Apollo Tests

Test org 1

ASU reliability data PRL Arizona Sites

Several sets of data provided for data hub

ASU reliability data PRL New York Sites

ASU reliability data PRL Arizona Sites

Followers

0

[+ Follow](#)

Organization



Apollo Tests

Test org 1 [read more](#)

Social

[Google+](#)

[Twitter](#)

[Facebook](#)



[Dataset](#) [Groups](#) [Activity Stream](#) [Related](#)

[Manage](#)

ASU reliability data PRL Arizona Sites

Several sets of data provided for data hub

Data and Resources

- 
AZ16 PRL Field data
[Explore](#)
- 
AZ11 PRL Field Data
[Explore](#)
- 
AZ10 PRL Field Data
[Explore](#)
- 
AZ Module Information
[Explore](#)
- 
PRL Field Database - Composite
 Combined original version of PRL test sites. Includes New York and Arizona...
 [Explore](#)

[ASU](#) [Arizona](#) [reliability](#) [site_data](#)

Additional Info

Data Source Metadata

Field	Value
Author	Johnathan Trinastic
State	active
Last Updated	May 17 2017 20:20

DJ_review8

[Manage](#) [Go to resource](#) [Data API](#)

URL: <http://ckantest.shadow.org/dataset/14655401-5a74-4cf2-8acf-7dddc8e96b22/resource/5ede61d4-48ae-4dbc-be4d-191afed062b1/d...>

Data gathered by hand via literature searches for deployed modules and arrays and failure mechanisms

Data Explorer

[Embed](#)

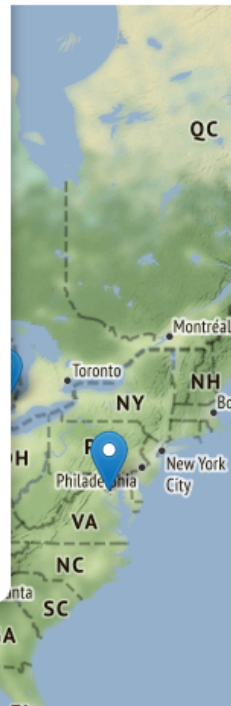
Add Filter

Grid Graph Map 1000 records << 1 - 100 >>

Search data ... Go » Filters



Climate3: Hot & Humid
Mounting: 1-axis tracker
Begin.Year: 1987
Precipitation: fully humid
Module Type1: mono-Si
Module Type2: mono-Si
Manufacturer: ARCO
System size (kW): null
Country: USA
Longitude: -97.75
Main Climates: Warm temperature
Model: AUSTIN PV300
Load:
No.modules: null
Reference: 10 year PV review_Rosenthal_NM_PVSC_1993.pdf
State/Town: TX, Austin
Years: 4
Climate2_PVSEC2012: Hot & Humid
Module Type3: x-Si
Rd+Pmax+: 1.70
_id: 13
Imax+:
Isc+:



Latitude / Longitude fields
 GeoJSON field
Latitude field
Longitude field
Update
 Auto zoom to features
 Cluster markers

Conclusions

- **CKAN was chosen as the principle framework of the system and provides science enabling capability to meet the 70% solution.**
- **CKAN will need some customization to meet the needs of DuraMat**
- **The Prototype deployment of the basic DuraMat DataHub is underway and we are testing data sets against it.**
- **We expect the basic hub to be up sometime during the summer.**