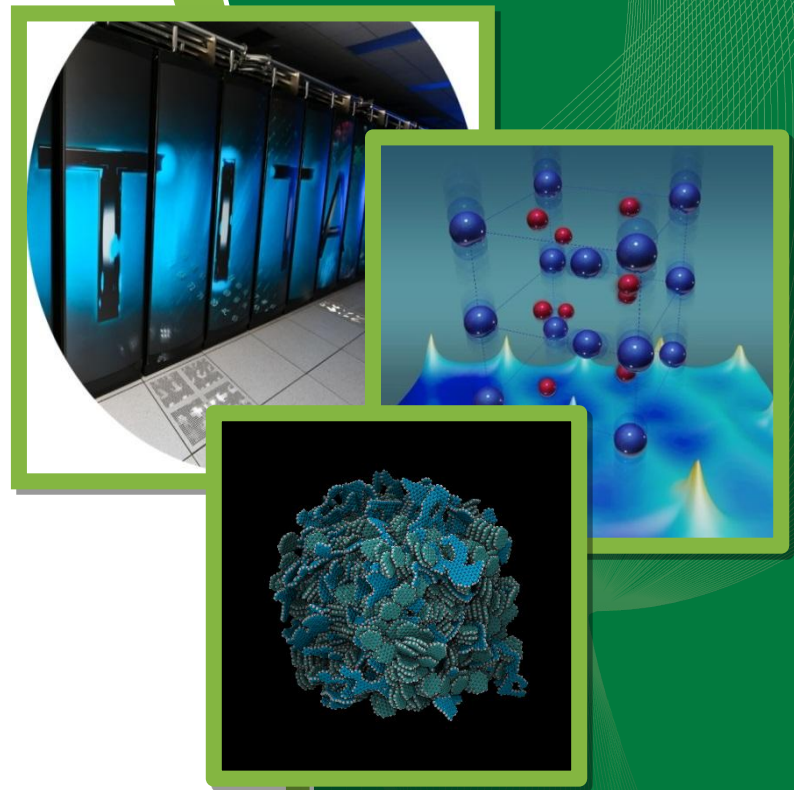# The New Energy Science and Technology Software Center
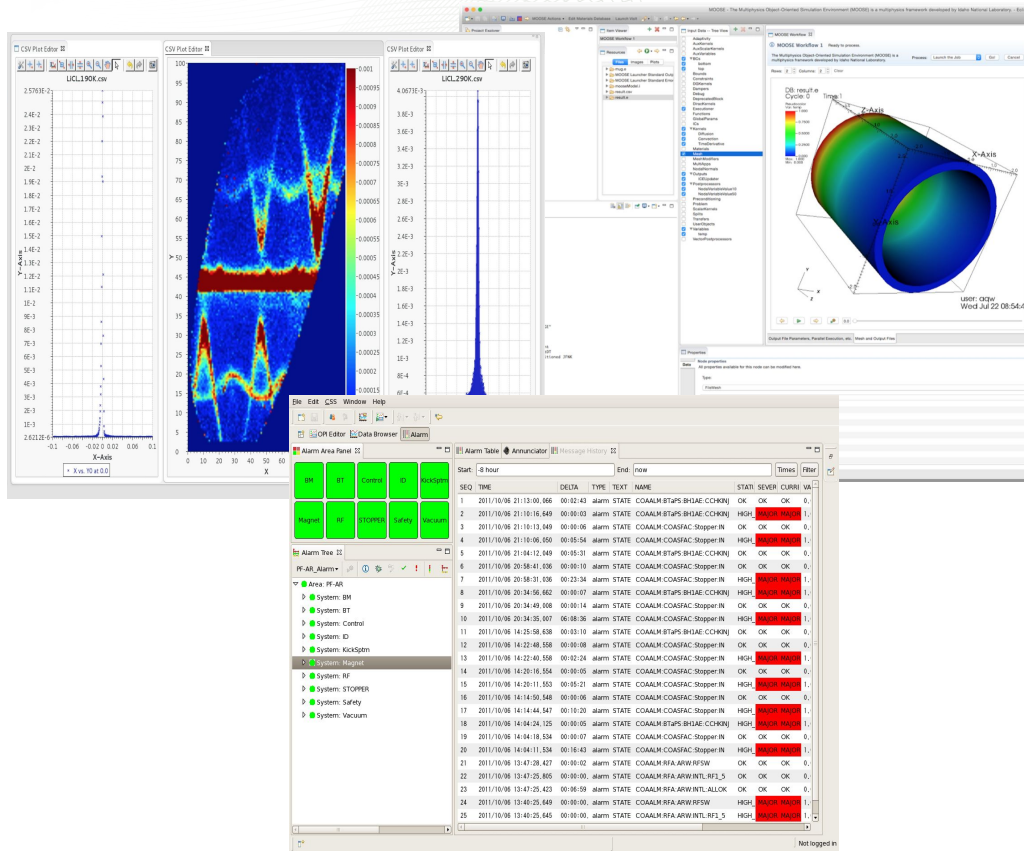
Jay Jay Billings, ORNL & OSTI
with contributions from the rest of the DOE Code Team

20170503
@jayjaybillings

OAK RIDGE
National Laboratory

Don't worry: We're not making you stop using GitHub!
(or any other repo service)

OAK RIDGE
National Laboratory

Tracking DOE Scientific Software

OAK RIDGE
National Laboratory

+ System Software, etc.

Think Scientific Software - not business software

OAK RIDGE
National Laboratory

What is the ESTSC?

The Energy Science and Technology Software Center (ESTSC)...

- is the DOE's centralized software management facility
- licenses and distributes scientific and technical software
- was moved to OSTI in 1991
- originally started at ANL in 1960 by Margaret Butler
- has >3517 packages

OAK RIDGE
National Laboratory

What is the ESTSC?

U.S. Department of Energy

E·S·T·S·C

Energy Science & Technology
Software Center

Thank you for visiting the Energy Science and Technology Software Center!

We have recently redesigned our software submission and dissemination processes, and we hope you will like the new websites.

To Submit Software, please use the software submission form 241.4 on E-Link.

To Search for Software, please visit Scitech Connect, and use the "Software" tab to limit searches to software only.

To Order Software, please submit an email to estsc@osti.gov.

For additional information about DOE scientific and technical software, refer to http://www.osti.gov /stip/241.4.

Please contact us by Email if you have any questions or feedback.

Some links on this page may take you to non-federal websites. Their policies may differ from this site.

U.S. DEPARTMENT OF **ENERGY** | Office of Science | Office of Scientific and Technical Information   Science.gov   WORLDWIDE SCIENCE.ORG

OAK RIDGE National Laboratory

Two Big Problems

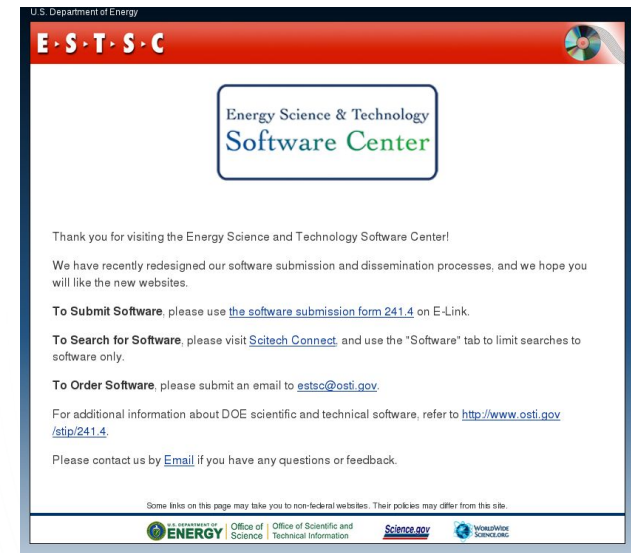1. Not a modern product; has a poor user experience.

   Literally still uses CDs. ⟶

   This is the <u>entire</u> product. ⟶

2. Misses most modern DOE software.

   ⟶ What is a *"release?"*

ESTSC misses many "in development" projects

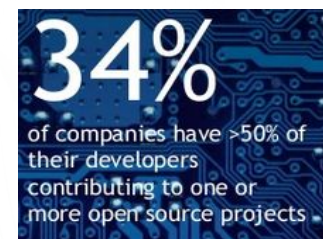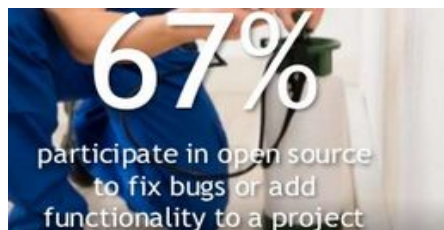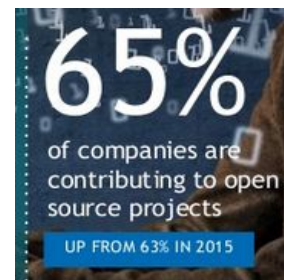**We expect there are between 5x-10x more codes in the wild than those already available in the ESTSC.**

There are more than 900,000 mentions of "Department of Energy" on GitHub <u>alone</u>.

That's roughly 1000-10000 codes.

OAK RIDGE
National Laboratory

What's important for a "modern" ESTSC?

OAK RIDGE
National Laboratory

Modern Software is Open Source

According to the 2016 Future of Open Source Survey...
(1313 Companies Participating)

>65% leverage OSS to speed application development

>55% leverage OSS for production infrastructure

65% of companies are contributing to open source projects
UP FROM 63% IN 2015

67% participate in open source to fix bugs or add functionality to a project

59% participate in open source to gain competitive edge

67% actively encourage developers to engage in and contribute to open source projects

34% of companies have >50% of their developers contributing to one or more open source projects

https://www.blackducksoftware.com/2016-future-of-open-source

OAK RIDGE
National Laboratory

Modern Software is Collaboratively Developed

Either through open *social coding* sites like...
- **GitHub**
- **BitBucket**
- **Sourceforge.net**

Or through "Software Foundations" like...
- **The Eclipse Foundation**
- **The Apache Software Foundation**
- **The Linux Foundation**

OAK RIDGE
National Laboratory

Modern Software is a Social Activity

c.f. - "Hackathons"

AT&T Hackathon

ORNL Hackathon

OAK RIDGE
National Laboratory

True for Science?

ORNL receives Exascale Computing Project awards to develop next-gen applications

All ECP projects are multi-institution projects that depend on open source in one way or another.

EXASCALE COMPUTING PROJECT

Exascale Computing Project (ECP) Awards $39.8 million for Application Development

OAK RIDGE
National Laboratory

What would a "modern" ESTSC look like?

OAK RIDGE
National Laboratory

New Name and Branding

DOE(Code){}

The DOE depends on code, which these **super nerdy** "Constructor" logos acknowledge and celebrate.

DOE (Code){}

OAK RIDGE
National Laboratory

A "Modern" Software Experience

DOECode will...
- **Focus on enabling social coding**
- **Integrate researchers, data, etc. *with* the Software**
- **Be open source and developed by the whole DOE community**
  - **OSTI, ORNL, etc.**

Integration with Open Communities

Social Media Integration

Open Web API

Private DOE-only Repositories

OSTI-Hosted Repositories

Strong Community Backing

OAK RIDGE
National Laboratory

Strong Community Backing

We have engaged the DOE Research Community through Requirements Teams

- "Core Requirements"
- OSTI – Technical
- OSTI – Policy
- DOE – GC
- Labs – Software
- Labs – HPC
- Labs – Data
- DOE ASCR

- Students
- Post Graduates
- Labs – Ops
- Labs – STI Managers
- Exascale Computing Program
- DOE Other Offices
- University Partners

Teams are roughly 4-5 people (sometimes more) and a set of interviews.

OAK RIDGE
National Laboratory

Strong Community Backing - We're on GitHub!
https://github.com/doecode

Strong Community Backing - Events

We're hitting the road!

Talks this year
- ICSTI TACC Workshop
- Supercomputing 2016
- SIAM CSE 2017 Spring Conference in Atlanta
- DOE ASCR
- STIP
- CENDI

Currently submitting to several more!

OAK RIDGE
National Laboratory

Leveraging - not replacing! - existing services

A profile on DOECode will combine info plus...

Your repo location from, e.g. -

Your papers from, e.g. -

Your data from, e.g. -

(Among others!)

Integrations we're thinking about

What if your code already has a DOI, i.e. - from Zenodo?

What about code.gov, the OMB's Open Source pilot?

OAK RIDGE
National Laboratory

Some cool requests thus far

- Automated metadata ingest
- DOIs for software!
- Code Scanning - Performance, Quality, Security
- Automatic provisioning of containers on clouds
- Metrics and Analytics Dashboard
- Social Media Features for Authors and Users
- Portfolios and Lists
- "Success story" blog posts
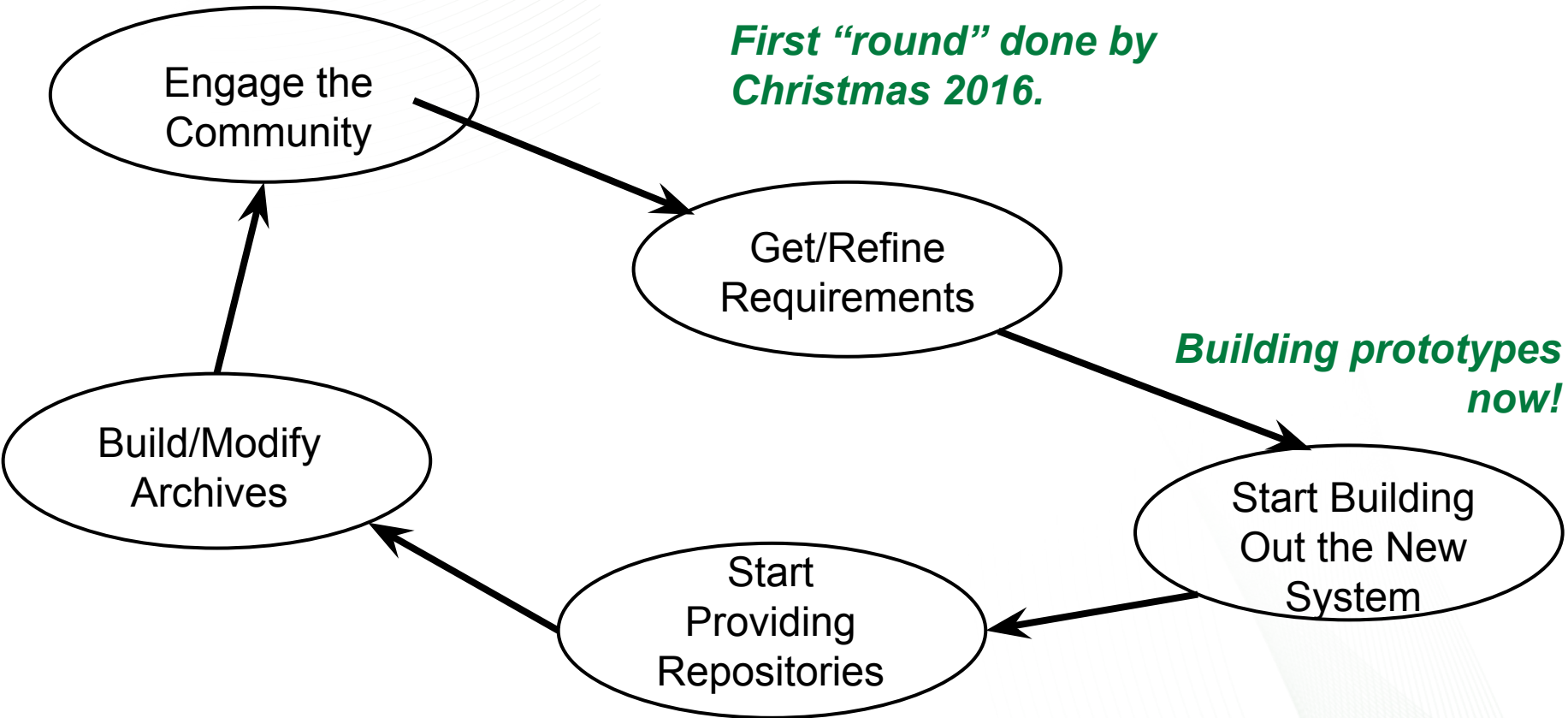- Long-Term Download Support for old versions

Path Forward

OAK RIDGE
National Laboratory

2yr development plan with many iterations!

Iteratively developing this product with community support!

First "round" done by Christmas 2016.

Building prototypes now!

Engage the Community → Get/Refine Requirements → Start Building Out the New System → Start Providing Repositories → Build/Modify Archives → Engage the Community
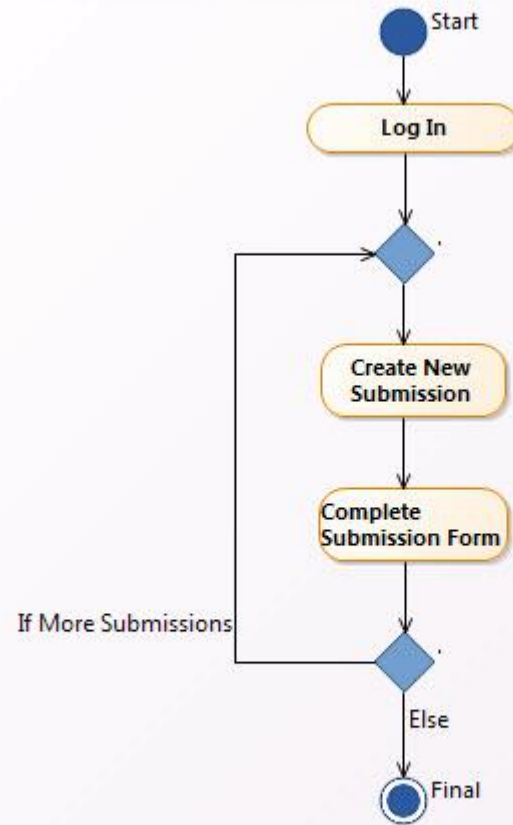
OAK RIDGE
National Laboratory

What does this process look like?

1. The policy teams ask a question and answers it with a model.
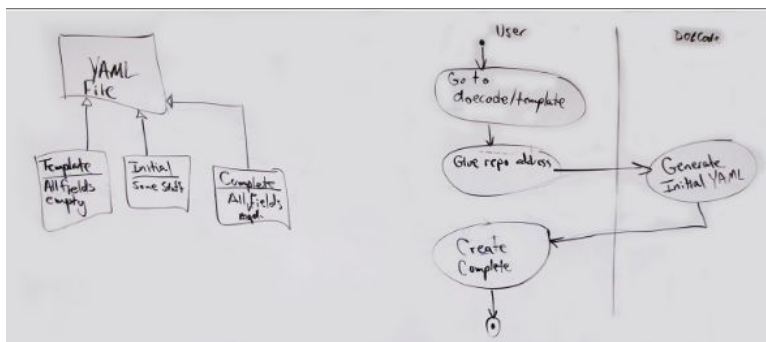2. The tech team responds with more questions, models, wireframes or code.
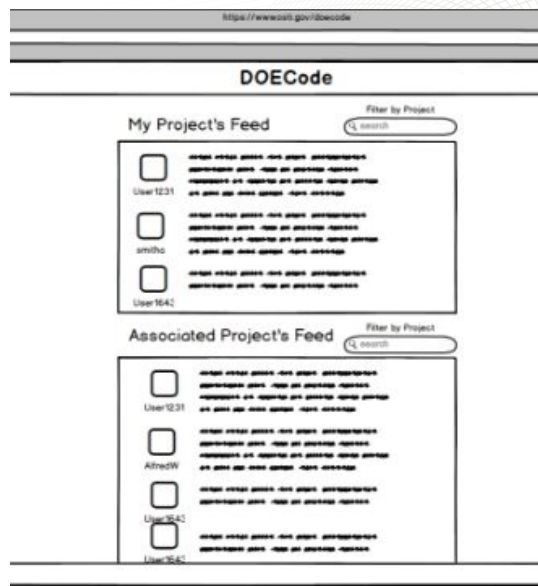
Managing Requirements, pt 1

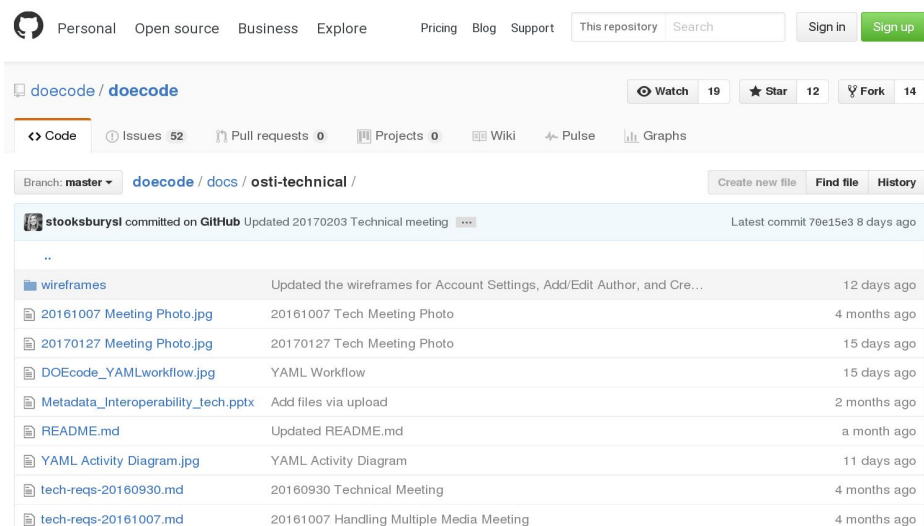Includes high-level requirements and features captured as...

- activity/workflow diagrams
- wireframes
- metadata schemas

Managing Requirements, pt 2

Strong engagement in detailed analysis and tracking from team AND community.

Label	Version
Definition	The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format.
Scheme	n/a
Property	dcterms:hasVersion / dcterms:isVersionOf
Usage	Potential use for multiple versions is the standard rdf:Alt container (or rdf:Seq or rdf:Bag) as a blank node containing the ordered or unordered versions
Cardinality	0 - n

Example:

```
<dcterms:hasVersion>
<rdf:Alt rdf:about="http://www.github.com/Software">
    <rdf:_1 rdf:resource="github.com/Version2">
    <rdf:_2 rdf:resource="github.com/Version3">
    <rdf:_3 rdf:resource="github.com/Version4">
```

Metadata Overview

Critical Updates:

- Used "Software Fields" spreadsheet to map local fields to all established controlled vocabularies
- Met with stakeholders/teams to verify/discuss proposed mappings
- Drafted Metadata Application Profile, Data Map, and YAML files
- Appended initial "Software Fields" spreadsheet with controlled vocabulary mappings
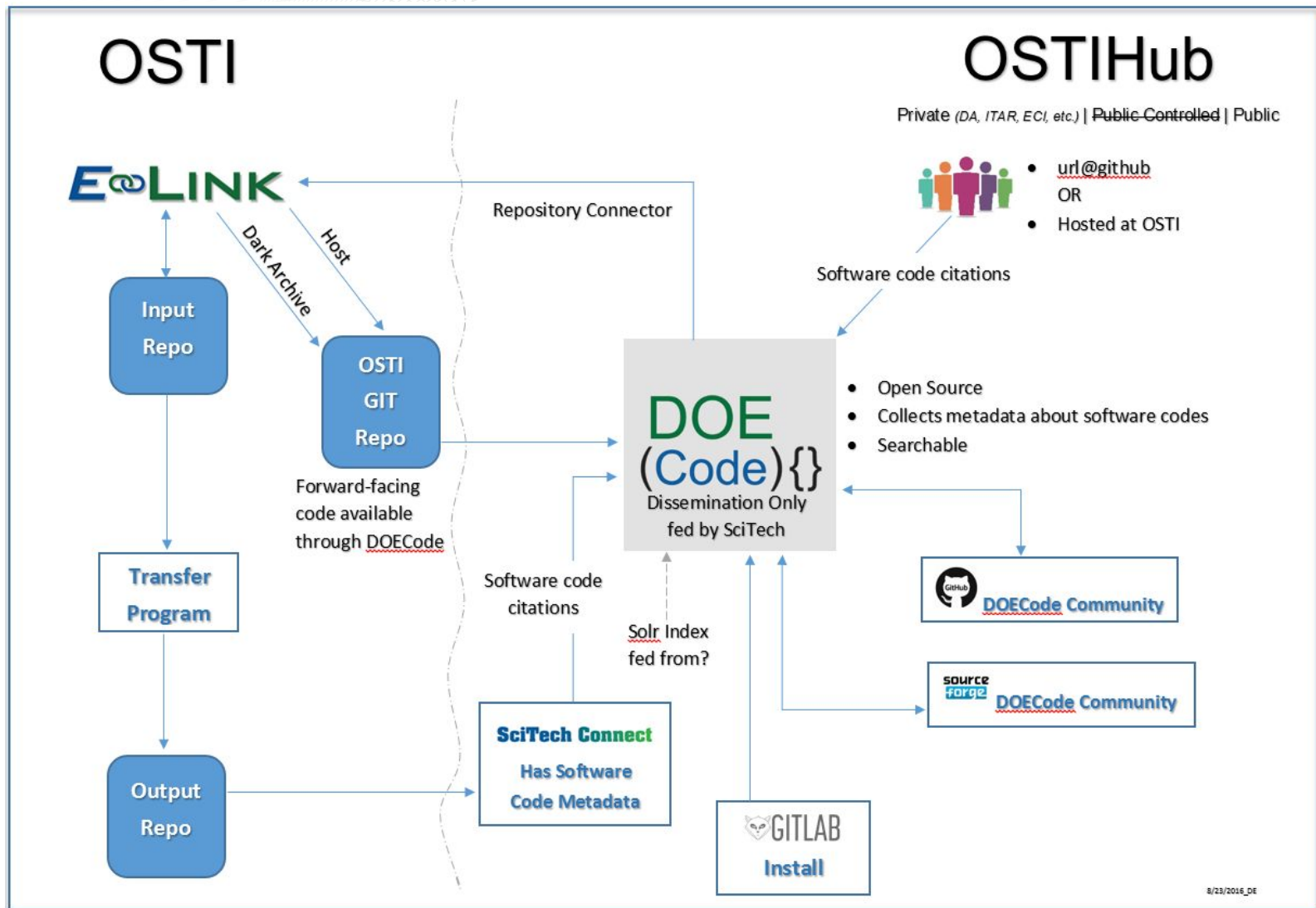
Schemas used:

- ADMS, ORG, SKOS, FOAF, Dublin Core, Schema.org

OAK RIDGE
National Laboratory
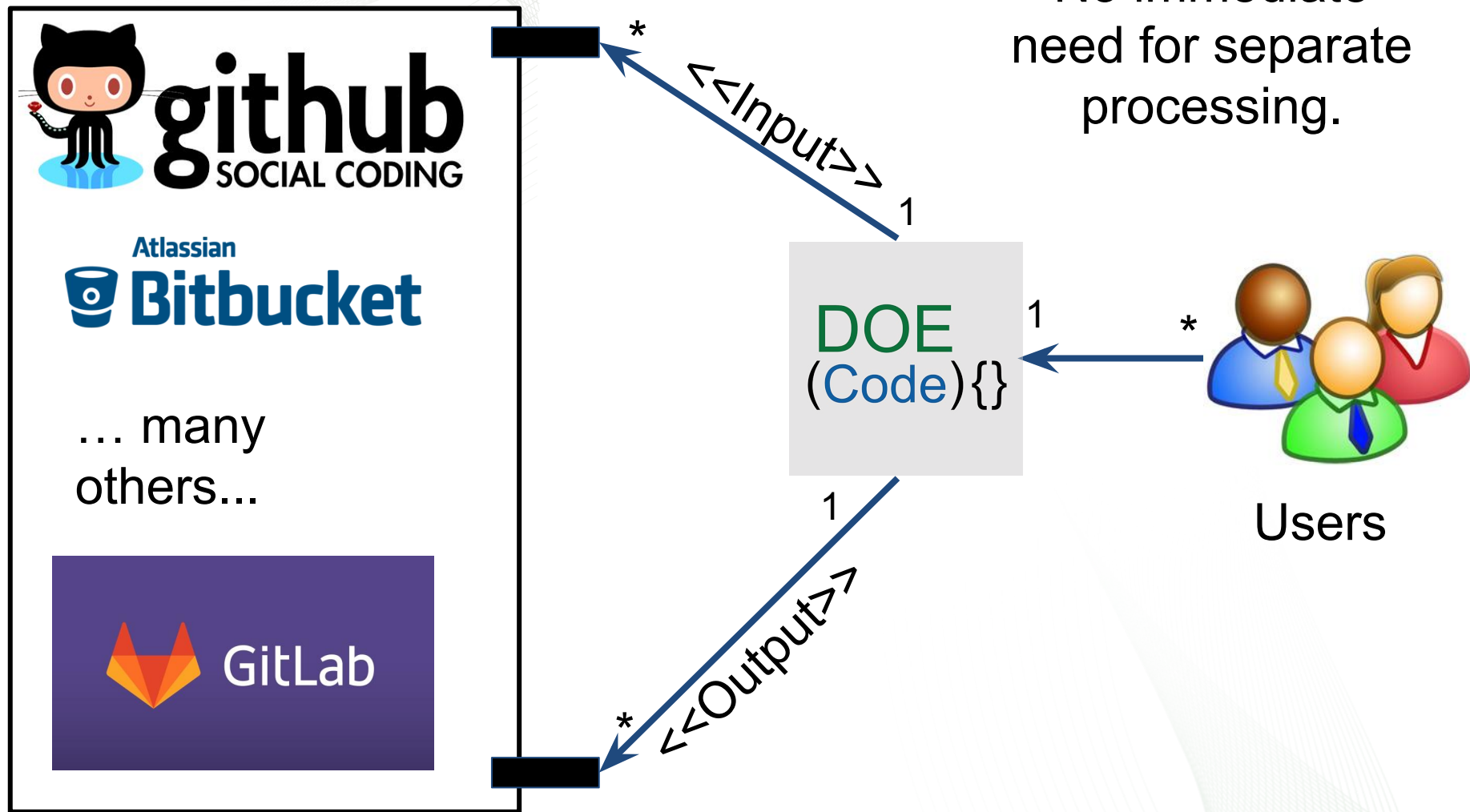
Metadata Overview, pt 2

This will give OSTI **<u>best-in-class metadata:</u>**
- Make the data migratable/interoperable with other systems
- Make OSTI compliant with best practices described by:
  - ISO 23081-2:2009E
    - "Standardized metadata are an essential prerequisite for information system interoperability."
  - ISO/TC 46/SC11N800R1
    - "…discover and then analyse any existing relevant schemas to see if any can be implemented without further change"
    - "If possible, do not introduce any new elements"
  - The Government Linked Data Working Group (W3C).
    - "It is best practice to use or extend an existing vocabulary before creating a new vocabulary."

OAK RIDGE
National Laboratory

Deployment Architecture for OSTI

Deployment Architecture for ORNL

No immediate need for separate processing.

github SOCIAL CODING

Atlassian Bitbucket

… many others...

GitLab

* <<Input>> 1

DOE (Code){}

1 *

Users

1 <<Output>> *

OAK RIDGE National Laboratory

Other Important Things

Training:
- Can't "drop" this on the community
- Need tutorials for each role, and submitters
- Might need training for deployment

Alpha is "Unlimited" Access only
- Not addressing limited access restrictions until the beta (at least)

OAK RIDGE
National Laboratory

How can you help?

Tell us what YOU want from DOECode!
Join a requirements team!
Engage at the GitHub site!

Website: https://www.osti.gov/doecode/

GitHub: https://github.com/doecode

Email: doecode@osti.gov

Twitter: @ostigov or @jayjaybillings

OAK RIDGE
National Laboratory

We are very grateful for the support of the Oak Ridge Leadership Computing Facility + IDEAS!

Demo

OAK RIDGE
National Laboratory

Questions?

DOE(Code){}

Thank You!

Additional Authors from OSTI: Chris Augustus, Kim Buckner, Daphne Evans, Neal Ensor, Joy Fender, Susie Faust, Darel Finkbeiner, Judy Gilmore, Lorrie Johnson, Peter Lincoln, Mark Martin, Josh Nelson, Rebel Powell, Carly Robinson, Crystal Sherline, Andrew Smith, Shelby Stooksbury, Lance Vowell and others.

Please reach out by email!
Jay: billingsjj@ornl.gov
Mark Martin: martinm@osti.gov

Or social media!
Twitter Handles:
@jayjaybillings
@OSTIgov

OAK RIDGE
National Laboratory