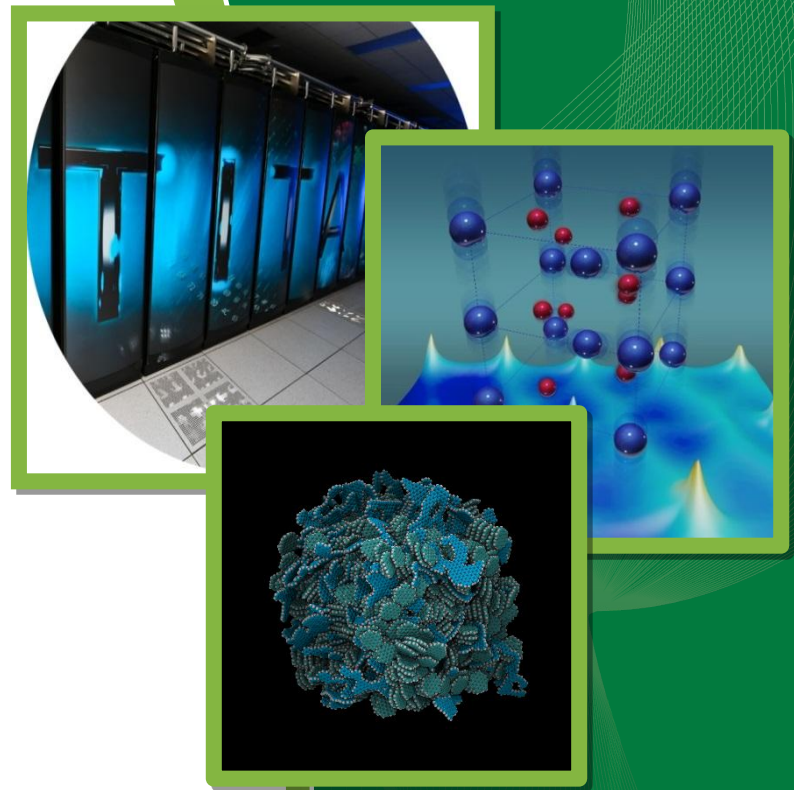


The New Energy Science and Technology Software Center

Jay Jay Billings, ORNL & OSTI
with contributions from the rest of
the DOE Code Team

20170221

@jayjaybillings



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

Office of Scientific and Technical Information

“OSTI”

What is OSTI?

“...the Department of Energy (DOE) office that collects, preserves, and disseminates DOE-sponsored research and development (R&D) results that are the outcomes of R&D projects or other funded activities at DOE labs and facilities nationwide and grantees at universities and other institutions.”



OSTI HQ in Oak Ridge, TN

OSTI Products



Energy Science & Technology
Software Center

DOE Data Explorer

DOE Data ID Service



DOE R&D
Accomplishments

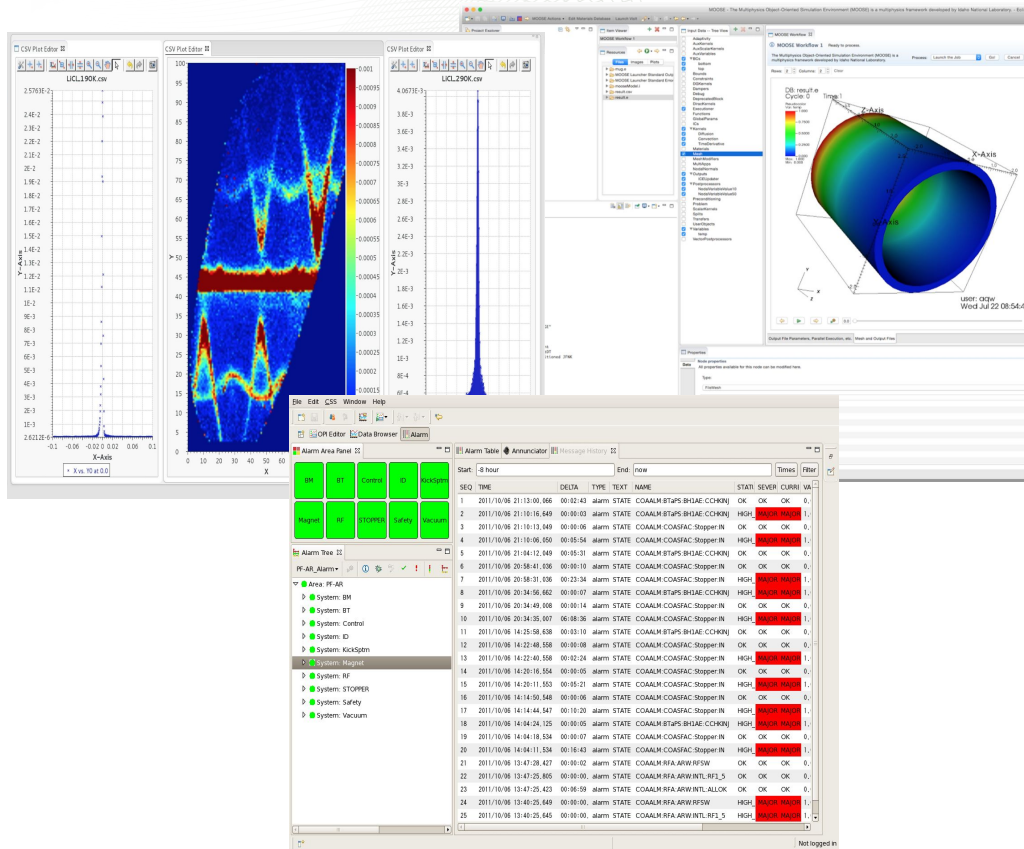


SciTech Connect



DOepatents

Tracking DOE Scientific Software



+ System Software, etc.

Think Scientific Software - not business software

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



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](http://science.gov)  **WORLDWIDE SCIENCE.ORG**

Two Big Problems

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

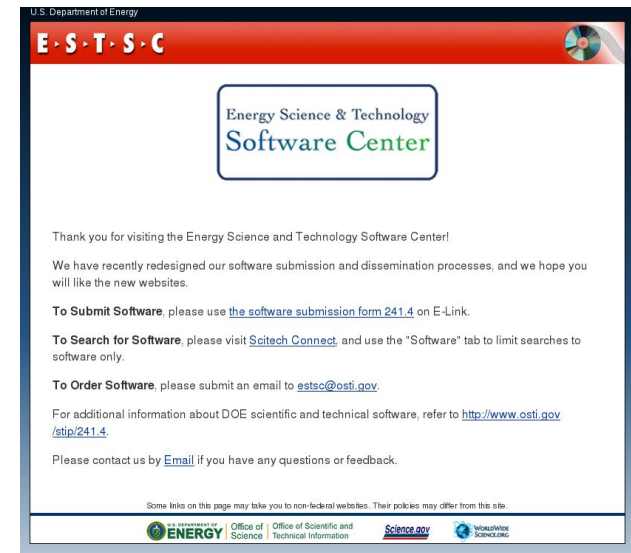
Literally still uses CDs. →



This is the entire 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.

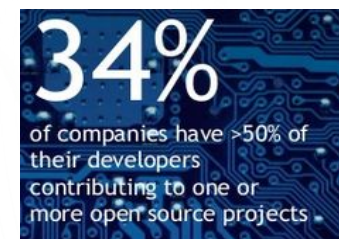
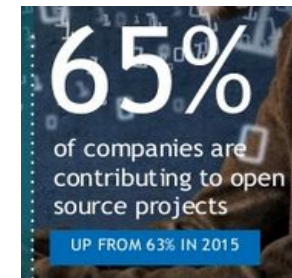
There are more than 900,000 mentions of “Department of Energy” on GitHub alone.

That’s roughly 1000-10000 codes.

What's important for a “modern” ESTSC?

Modern Software is Open Source

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



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

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**

Modern Software is a Social Activity

c.f. - “Hackathons”



AT&T Hackathon

ORNL Hackathon



True for Science?

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



All 22 ECP projects are multi-institution projects that depend on open source in one way or another.

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

What would a “modern” ESTSC look like?

New Name and Branding

DOE(Code){}

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

DOE
(Code){}

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

Strong Community Backing

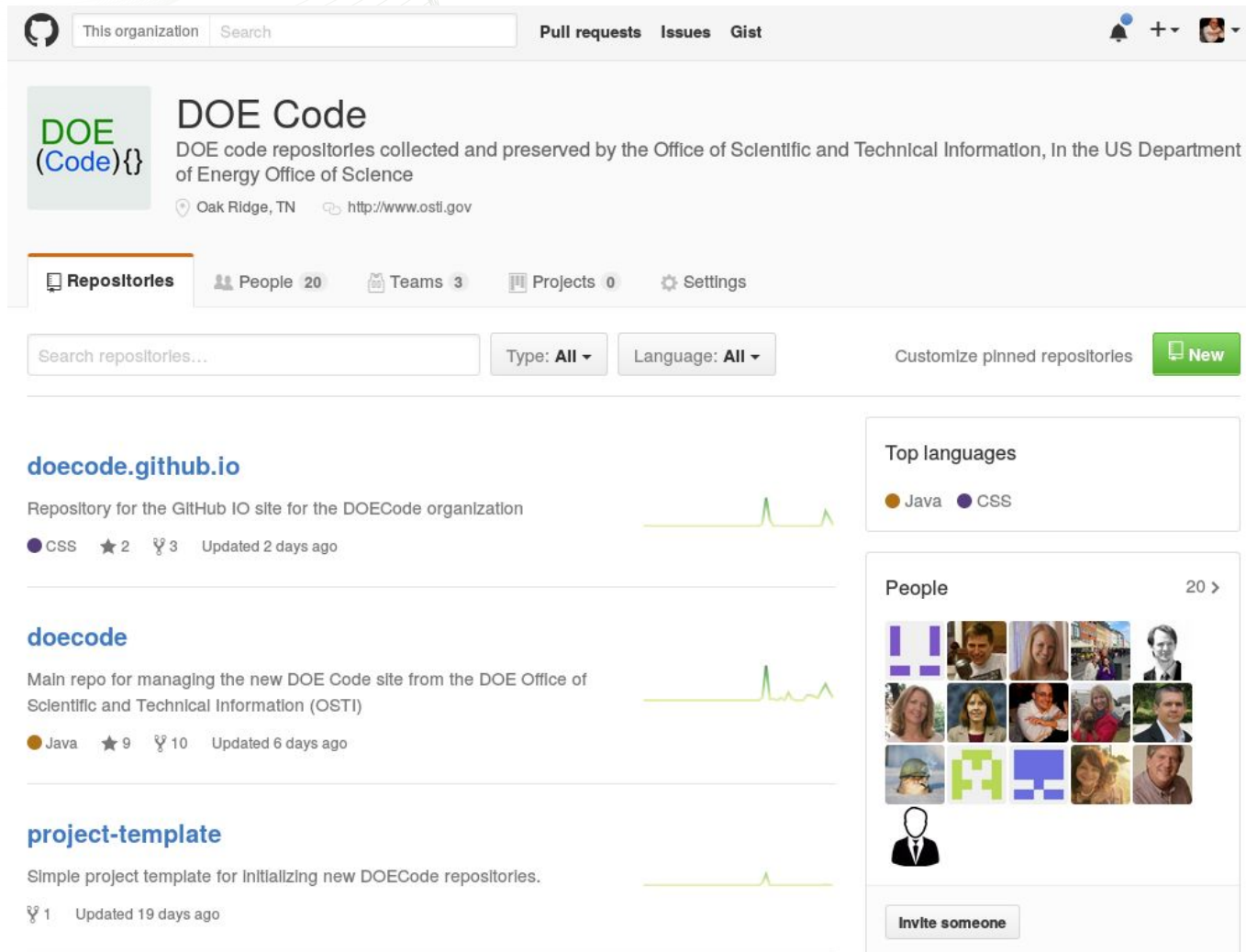
We have engaged the DOE Research Community through Requirements Teams

- “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.

Strong Community Backing - We're on GitHub!

<https://github.com/doecode>



The screenshot shows the GitHub organization page for DOE Code. At the top, there's a navigation bar with 'Pull requests', 'Issues', and 'Gist'. Below this, the organization's profile is displayed with the 'DOE (Code){}' logo, the name 'DOE Code', and a description: 'DOE code repositories collected and preserved by the Office of Scientific and Technical Information, In the US Department of Energy Office of Science'. The location is 'Oak Ridge, TN' and the website is 'http://www.osti.gov'.

The main section is titled 'Repositories' and includes a search bar, filters for 'Type: All' and 'Language: All', and a 'New' button. Three repositories are listed:

- doecode.github.io**: Repository for the GitHub IO site for the DOECode organization. It has 2 stars, 3 forks, and was updated 2 days ago. The primary language is CSS.
- doecode**: Main repo for managing the new DOE Code site from the DOE Office of Scientific and Technical Information (OSTI). It has 9 stars, 10 forks, and was updated 6 days ago. The primary language is Java.
- project-template**: Simple project template for initializing new DOECode repositories. It has 1 fork and was updated 19 days ago.

On the right side, there are two sections: 'Top languages' showing Java and CSS, and 'People' showing a grid of 20 team members with an 'Invite someone' button at the bottom.

Strong Community Backing - Events

We're hitting the road!

Accepted Conference Events

- ICSTI TACC Workshop
- Supercomputing 2016
- SIAM CSE 2017 Spring
Conference in Atlanta

Currently submitting to several more!

Leveraging - not replacing! - existing services

A profile on DOECode will combine info plus...

Your repo location from, e.g. -



Your papers from, e.g. -

SciTech Connect

Your data from, e.g. -

DOE Data Explorer

(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?

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

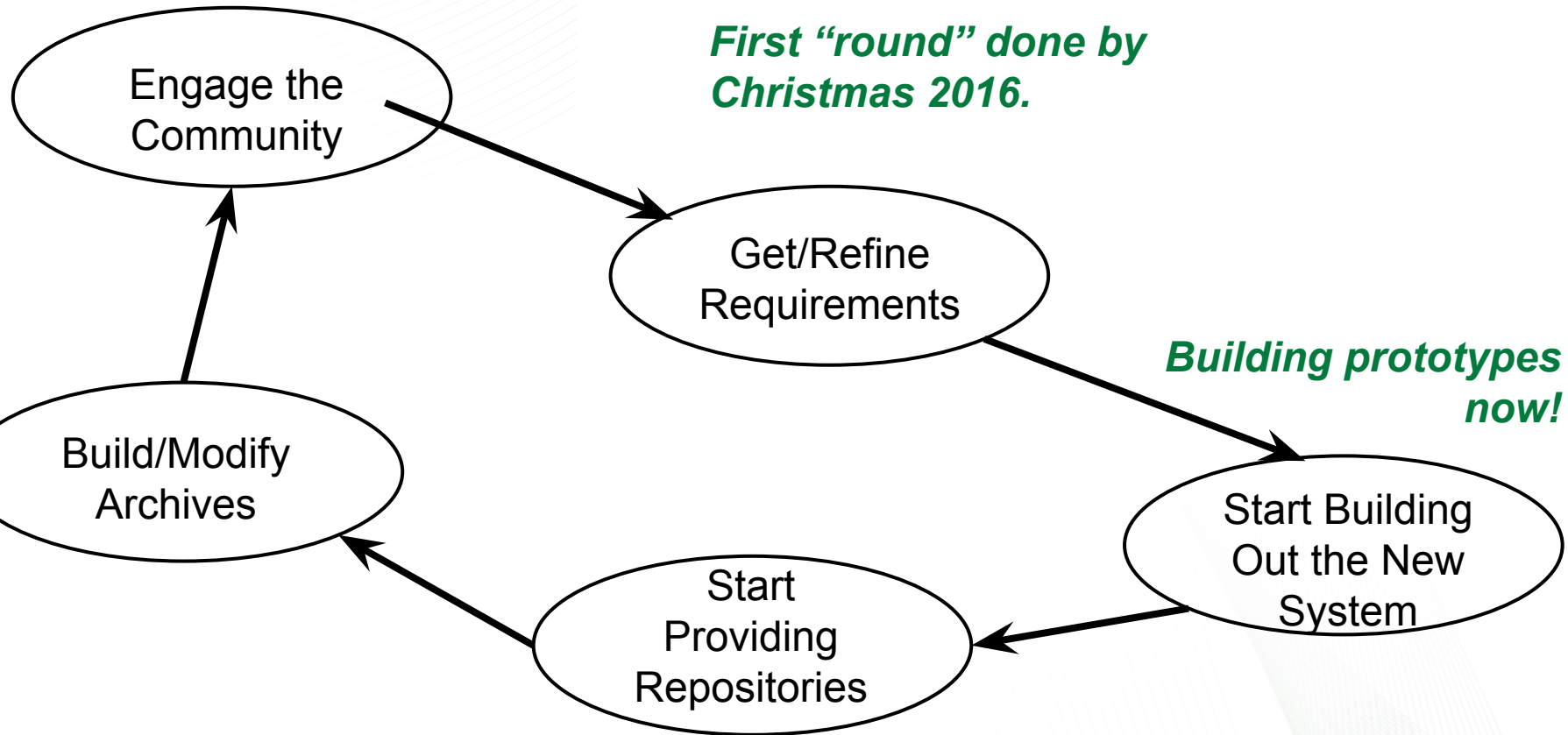
Path Forward

*Iteratively developing
this product with
community support!*

2yr development plan with many iterations!

*First “round” done by
Christmas 2016.*

*Building prototypes
now!*



Current Technical and Policy Questions

Example Technical Questions

- Should we use web frameworks?
- Security concerns?
- UI Wireframes to drive design

Example Policy Questions

- What must change?
- What ought to change?
- What should our metadata look like?
- What's the new process for submitting software?

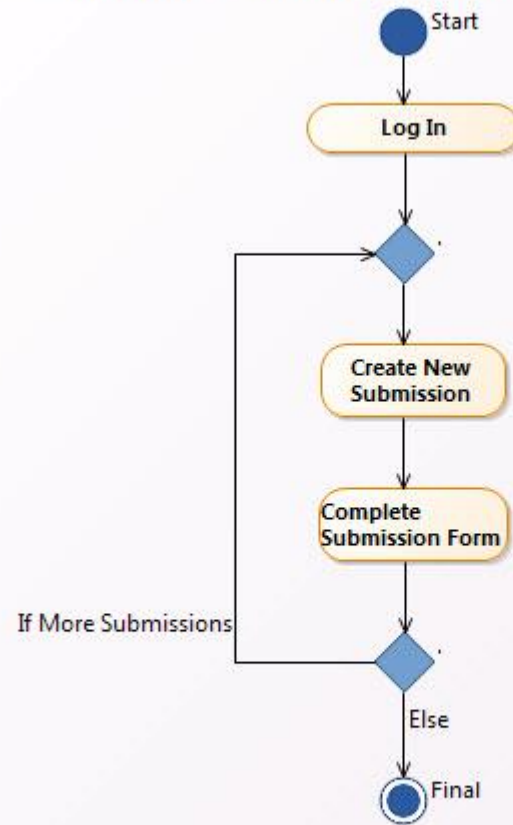
Both: What do users want?

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 or wireframes.

The screenshot shows a web browser window with the URL <https://www.oost.gov/doi/code>. The page title is "DOE Code". The main content area displays a submission form for "ACHILLES: Heat Transfer in PWR Core During LOCA Reflood Phase". The form includes fields for Publication Date, OSTI Identifier, Report Number(s), DOE Contract Number, Software Revision, Software Package Number, Software Package Contents, Software CPU, Open Source, Source Code Availability, Other Software Info, Research Org, Sponsoring Org, Contributing Org, and Country of Publication. The form is partially filled out with details about the ACHILLES project, including its purpose, test facility, and software information.

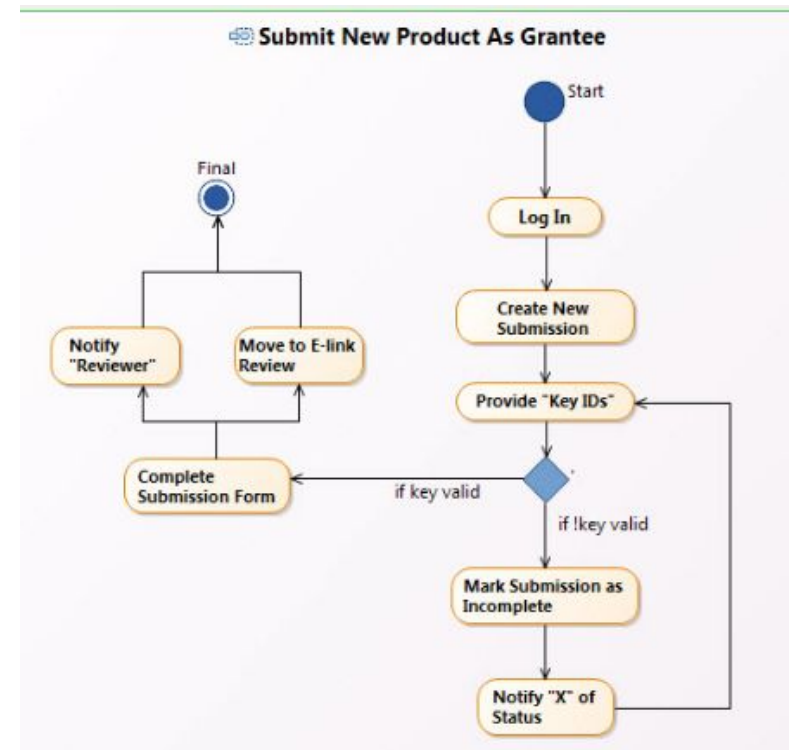
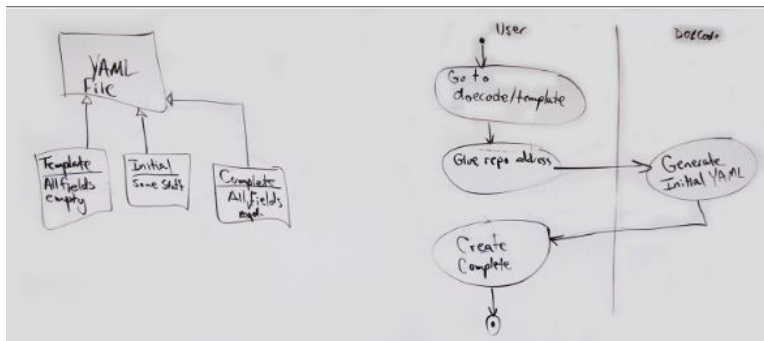
Lab STI POC Submission(s) Through Web API



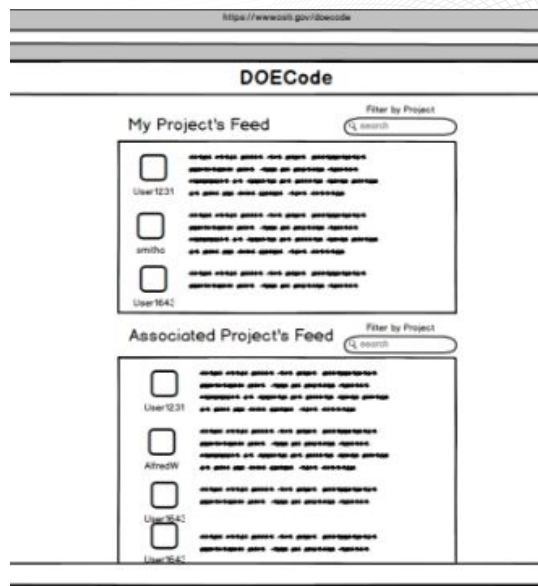
Managing Requirements, pt 1

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

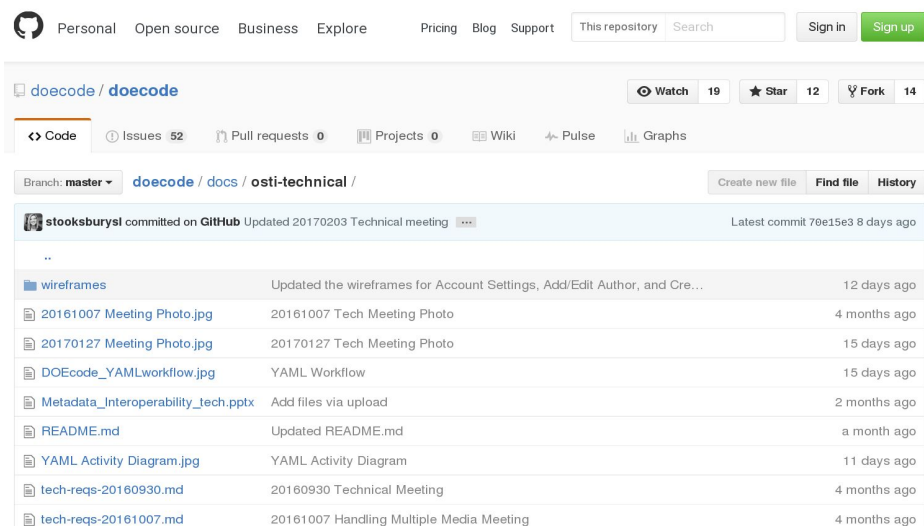
- activity/workflow diagrams
- wireframes
- metadata schemas



Managing Requirements, pt 2



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

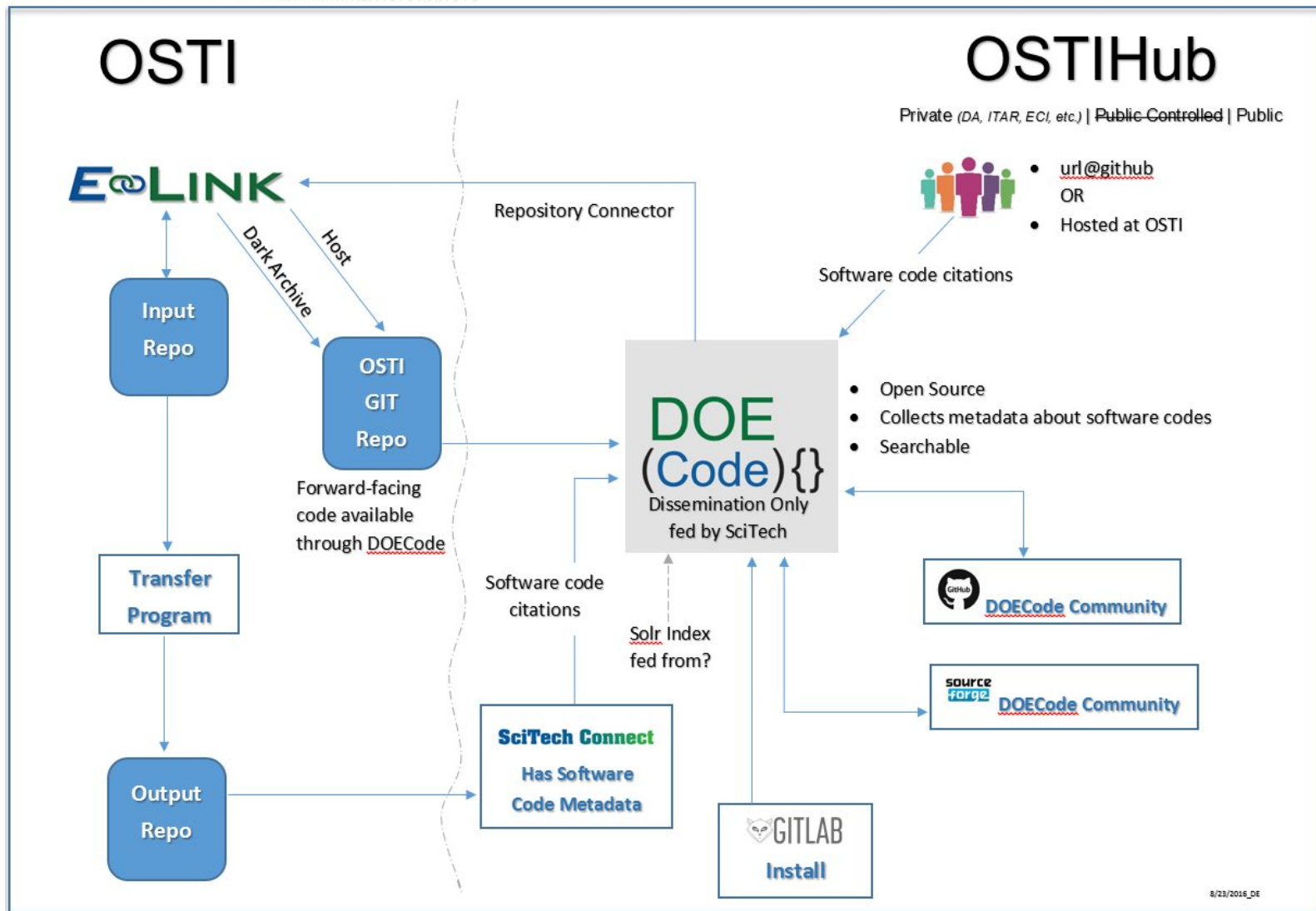


Label	Version
Definition	The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format.
Scheme	n/a
Property	dcterms:hasVersion / dcterms:isVersionOf
Usage	Potential use for multiple versions 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

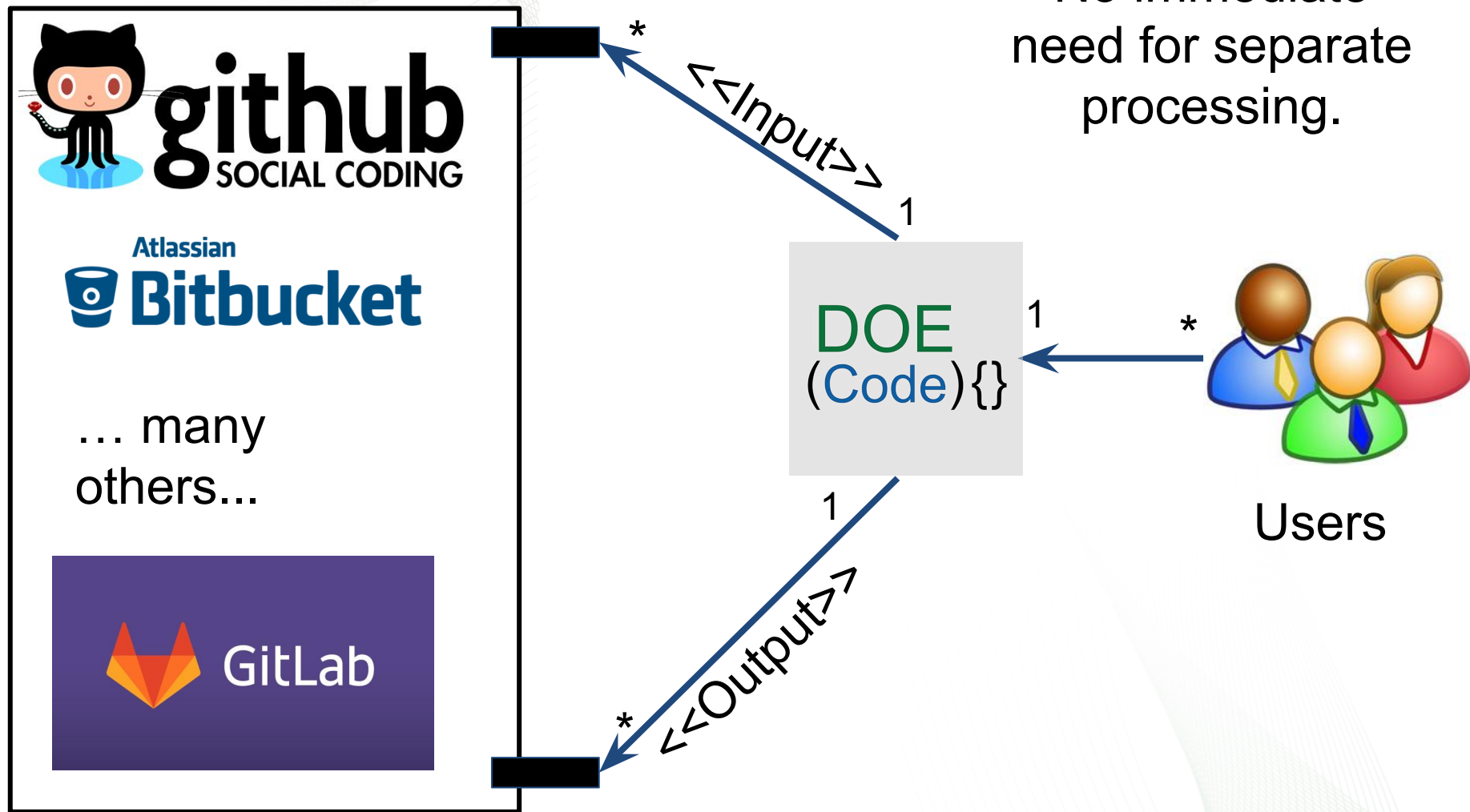
Example:

```
<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">
```

Deployment Architecture for OSTI



Deployment Architecture for ORNL



Metadata Overview

Work done:

- 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

Metadata Overview, pt 2

This will give OSTI **best-in-class metadata**:

- 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.”

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

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



IDEAS
productivity

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

Abstract

The U.S. Department of Energy's Office of Scientific and Technical Information (OSTI) is charged with cataloging and preserving the software artifacts, among other things, produced by the Department's researchers. The fast pace at which the software world changes - including the rise and dominance of Open Source Software - has presented OSTI with the opportunity to update and expand its Software Center. These updates will provide a platform for the research community that combines social coding with OSTI's vast database of papers, authors, data, and legacy software to provide an integrated, dynamic environment for scientific software developers. This talk will describe the on-going development effort around that platform, as well as the requirements that define it, and how OSTI is engaging the open research community at large in its development. This talk will also present a description of scientific software in a broader context to clarify further why such a platform is needed and the extremely great benefits expected from its deployment.