



The Information Management Code Registry: *Software Solutions for Information Management Needs*

**Colin Smith and Kristin Vanderbilt
Earth Science Information Partners (ESIP)
Summer Meeting (July 17, 2019)**

bit.ly/imcr-notes

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric circles, and the lines are thin and grey. The overall structure is organic and sprawling.

Overview

- ◎ Introduction
 - ◎ Report
 - ◎ Activities
- 
- A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes having concentric circles. The diagram is positioned in the lower right area of the slide.



“

Central to enhancing the use and value of Earth Science data are good information management practices, and the Information Management Code Registry works to this end by facilitating discovery and use of software solutions for information management needs.

Large research groups

Have IM expertise



Small
research
groups

Need IM expertise



Scope



IM software

Providing software for IM tasks whether run from a users local machine or accessed via web-services.



Open source

Emphasizing freely accessible and community supported software while not excluding proprietary sources.



Small teams

Focused on the needs of individuals and small research teams working in the “Earth Sciences” domain.

Goals



Accelerate IM tasks

Simplify discovery and use through a comprehensive registry searchable by task, language, and science domain, returning high level information for fitness of use assessments.



Facilitate development

Highlight new opportunities by identifying coverage gaps, openly discussing and recording community ideas, and organizing hackathons.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The overall structure is organic and sprawling.

Report

- ◎ Implementation
 - ◎ Curation
 - ◎ Discovery
 - ◎ Maintenance
 - ◎ Engagement
- 
- A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes having concentric circles. The diagram is positioned in the lower right area of the slide.

Implementation

The IMCR is implemented in OntoSoft, which provides a robust and rich science software ontology, and human friendly interface to search and discovery.



EARTHCUBE

<http://www.ontosoft.org/index.html>

Implementation

Identify

Locate - unique identifier

Understand

Relate - domain knowledge

Trust - quality and ratings

Do Research

Experiment - run with data

Compose - run with software

Cite - scientific publications

Execute

Access - download

Install - execution requirements

Run - testing execution

Get Support

Discuss - community support

Update

Track - evolution

Contribute - evolution

<http://ontosoft-earthcube.github.io/ontosoft/ontosoft%20ontology/v1.0.1/doc/>

Implementation

Identify

Locate - unique identifier

Understand

Relate - domain knowledge

Trust - quality and ratings

Do Research

Experiment - run with data

Compose - run with software

Cite - scientific publications

Execute

Access - download

Install - execution requirements

Run - testing execution

Get Support

Discuss - community support

Update

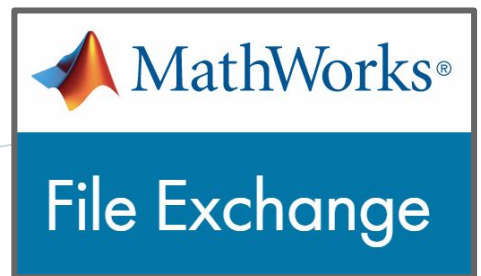
Track - evolution

Contribute - evolution

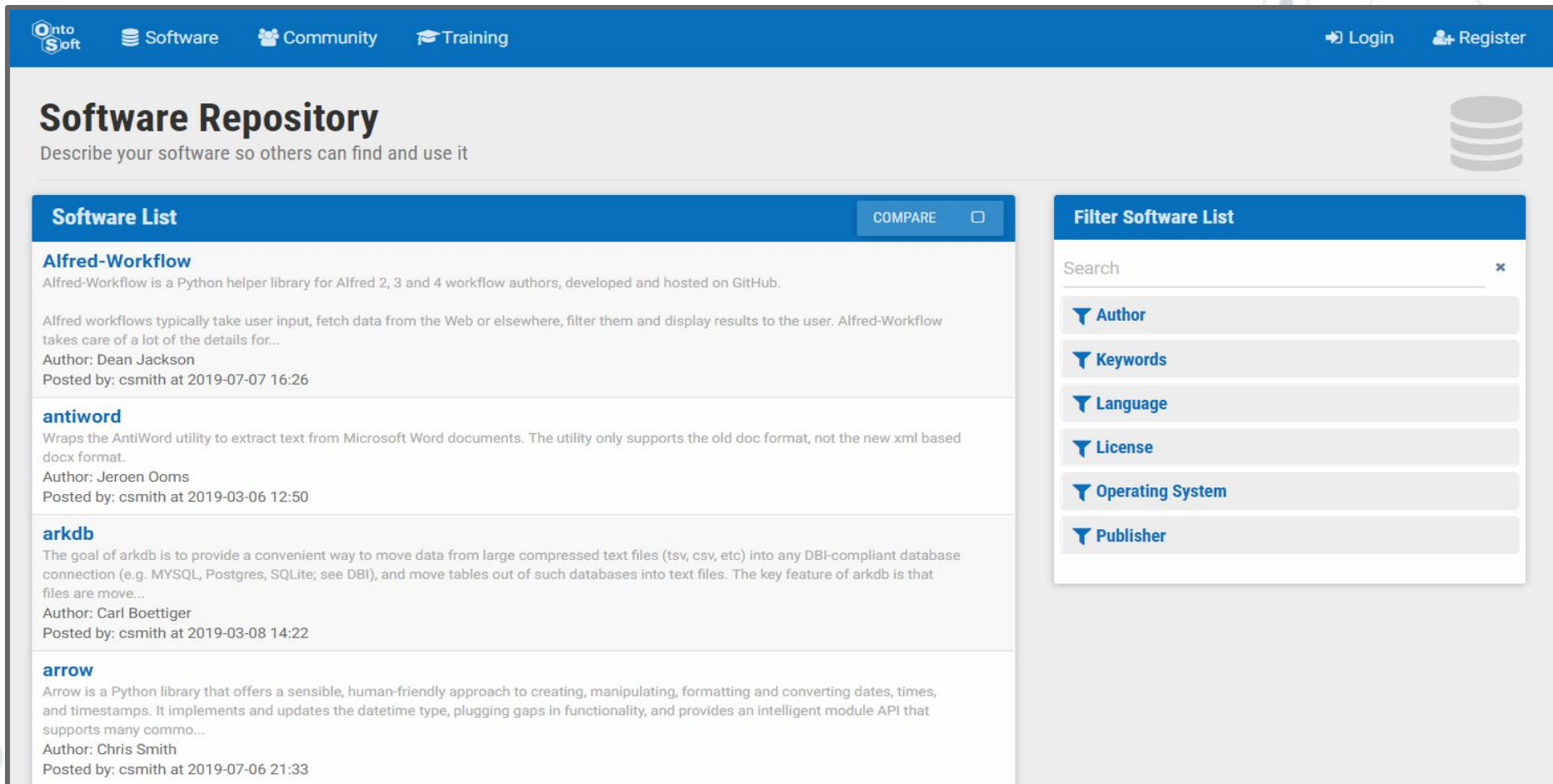
<http://ontosoft-earthcube.github.io/ontosoft/ontosoft%20ontology/v1.0.1/doc/>

Curation

An ongoing process involving manual discovery and metadata entry primarily done by IMCR maintainers, though anyone is welcome to contribute.



Curation



The screenshot displays the OntoSoft Software Repository website. The header features the OntoSoft logo and navigation links for Software, Community, and Training, along with Login and Register buttons. The main content area is titled "Software Repository" and includes a description: "Describe your software so others can find and use it". A "Software List" section shows a table of software entries, each with a title, description, author, and posting date. A "Filter Software List" sidebar on the right allows users to search and filter the list by various criteria.

Software List COMPARE

| Software List |
|---|
| Alfred-Workflow Alfred-Workflow is a Python helper library for Alfred 2, 3 and 4 workflow authors, developed and hosted on GitHub. Alfred workflows typically take user input, fetch data from the Web or elsewhere, filter them and display results to the user. Alfred-Workflow takes care of a lot of the details for... Author: Dean Jackson Posted by: csmith at 2019-07-07 16:26 |
| antiword Wraps the AntiWord utility to extract text from Microsoft Word documents. The utility only supports the old doc format, not the new xml based docx format. Author: Jeroen Ooms Posted by: csmith at 2019-03-06 12:50 |
| arkdb The goal of arkdb is to provide a convenient way to move data from large compressed text files (tsv, csv, etc) into any DBI-compliant database connection (e.g. MYSQL, Postgres, SQLite; see DBI), and move tables out of such databases into text files. The key feature of arkdb is that files are move... Author: Carl Boettiger Posted by: csmith at 2019-03-08 14:22 |
| arrow Arrow is a Python library that offers a sensible, human-friendly approach to creating, manipulating, formatting and converting dates, times, and timestamps. It implements and updates the datetime type, plugging gaps in functionality, and provides an intelligent module API that supports many commo... Author: Chris Smith Posted by: csmith at 2019-07-06 21:33 |

Filter Software List

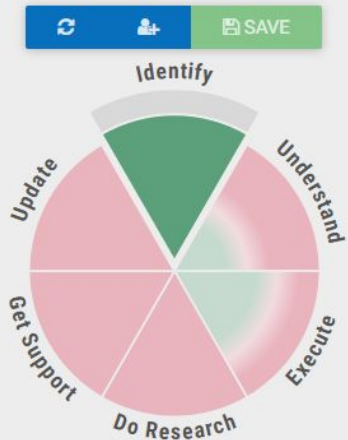
Search ×

- Author
- Keywords
- Language
- License
- Operating System
- Publisher

<http://imcr.ontosoft.org/#list>

Curation

arrow » Edit » Identify » **LOCATE**



Locate
unique description

Important

Optional

What is the software called ?

arrow

Last edited by csmith at 2019-07-06 21:40

What is a short description for this software ?

Arrow is a Python library that offers a sensible, human-friendly approach to creating, manipulating, formatting and converting dates, times, and timestamps. It implements and updates the datetime type, plugging gaps in functionality, and provides an intelligent module API that supports many common creation scenarios. Simply put, it helps you work with dates and times with fewer imports and a lot less code.

Last edited by csmith at 2019-07-06 21:40

What are general categories (keywords, labels) for this software ?

integration

Last edited by csmith at 2019-07-06 21:40

transformation

Last edited by csmith at 2019-07-06 21:40

Is there a project website for the software ?

<https://arrow.readthedocs.io/en/latest/>

Last edited by csmith at 2019-07-06 21:40

<http://imcr.ontosoft.org/#list>

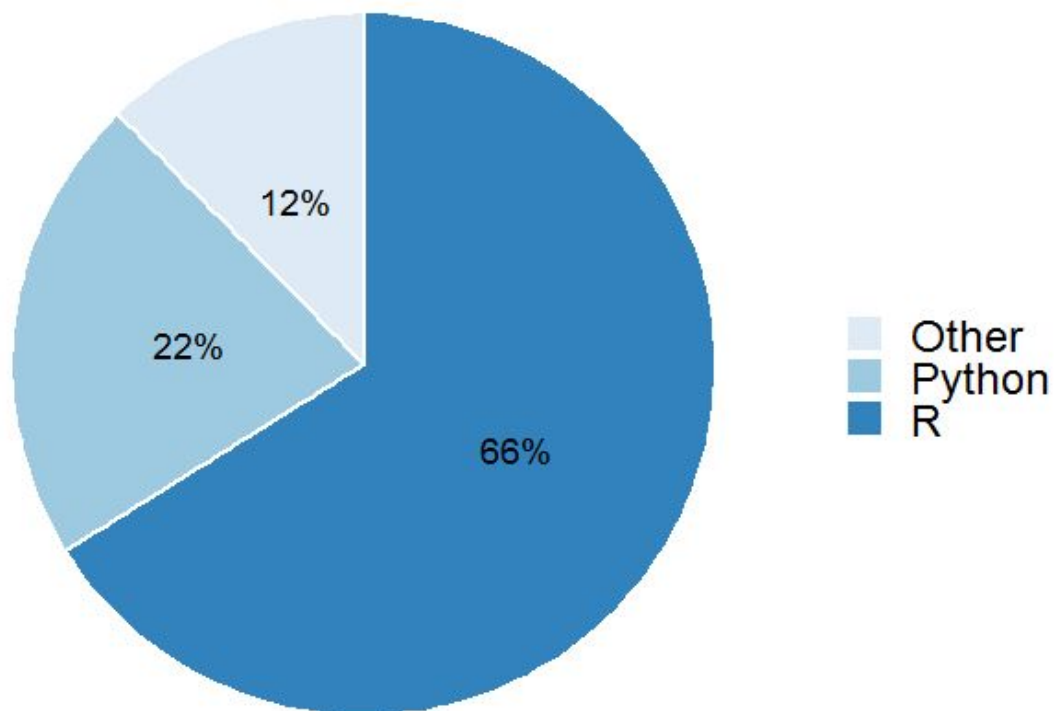
Curation

The IMCR now contains **183** software libraries primarily focused on R and Python and spanning a broad range of IM tasks.

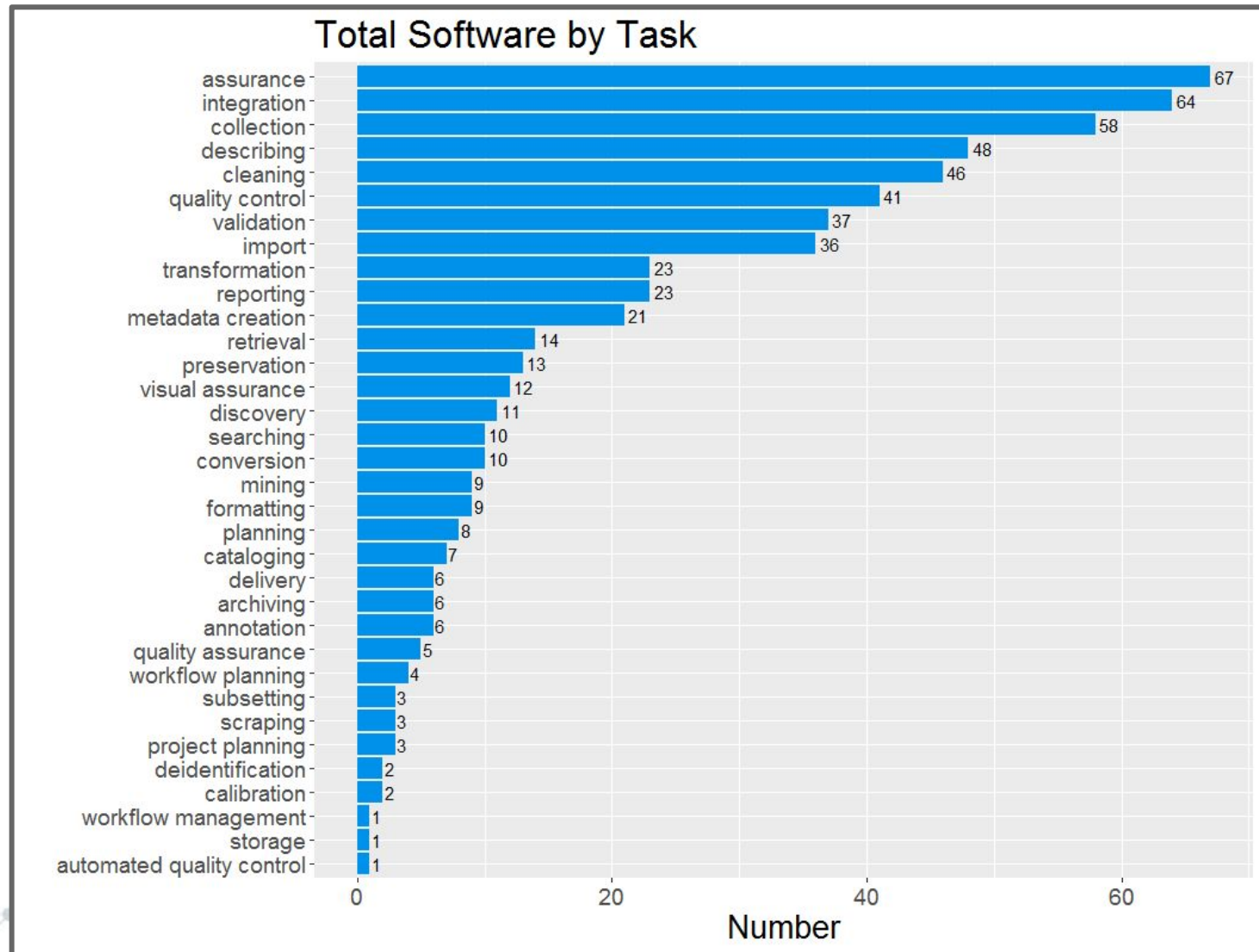


Curation

Implementation Languages



Curation

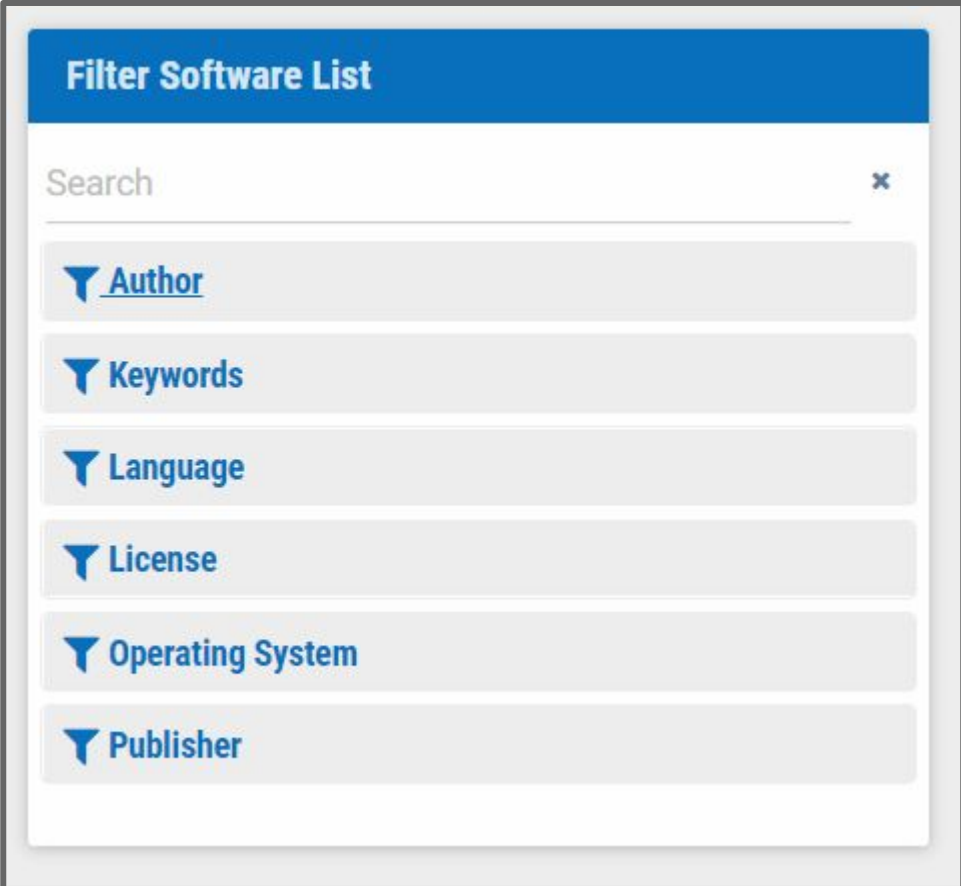


Curation



Discovery

OntoSoft provides an simple and effective search interface for discovering software by core attributes.



Filter Software List

Search x

▼ Author

▼ Keywords

▼ Language

▼ License

▼ Operating System

▼ Publisher

<http://imcr.ontosoft.org/#list>

Discovery

Keyword searches are enhanced with a controlled vocabulary organized around the DataONE research life-cycle and common IM tasks therein.

planning ▼

data management planning

project planning

workflow planning

collection ►

assurance ►

describing ►

preservation ►

discovery ►

integration ►

<http://vocab.lternet.edu/vocab/registry/>

Discovery

We enable discovery by science discipline using a section of the LTER controlled vocabulary.



- disciplines ▼
 - genomics
 - morphometrics
 - demography
 - geography ►
 - geology ►
 - history ►
 - hydrology
 - limnology ►
 - meteorology ►
 - oceanography
 - agriculture ►
 - biogeochemistry ►
 - biology ►
 - chemistry ►
 - conservation
 - ecology ►
 - biodiversity ►

<http://vocab.lternet.edu/vocab/vocab/index.php>

Maintenance

Automated maintenance of software metadata ensures content is current, accurate, and is maintained with minimal effort.



Engagement

Highlighting gaps in coverage, openly discussing and recording ideas, and organizing hackathons will develop new and useful tools.



Summary



IM software

Providing software for IM tasks whether run from a users local machine or accessed via web-services.



Open source

Emphasizing freely accessible and community supported software while not excluding proprietary sources.



Small teams

Focused on the needs of individuals and small research teams working in the “Earth Sciences” domain.

Summary



Implementation

Complete.



Curation

Ongoing manual process.



Discovery

Simple search augmented by controlled vocabularies.



Maintenance

Automation will ensure accuracy and reduce effort.



Engagement

Creating new tools as a community.



Activities

- Delineating potential
- Facilitating development
- Test-driving search and discovery
- Handling non-generalized code

bit.ly/imcr-notes

Delineating potential

You just heard the scope of our plans for IMCR ...

- ◎ How else can this resource be utilized?
- ◎ Where are the collaborative opportunities?
- ◎ What aspects of our plans need reconsideration?

Facilitating development

We plan on facilitating software development by exposing gaps and needs then organizing hackathons ...

- What info can be mined from the metadata?
- What are effective ways in gathering community ideas?
- How can we support hackathons?
- What other ways can we facilitate development?

Test-driving search and discovery

Review the IMCR controlled vocabulary and use it to search the portal for software of interest.

- ⦿ Do you find what you expect?
- ⦿ Do the search fields support the content you'd like to search on?
- ⦿ Comments on the vocab structure?
- ⦿ What terms should be added or removed?

Handling non-generalized code

Sharing non-generalized code, or code that is not apart of a library can be useful.

- ◎ How can we support this type of software?
- ◎ Is there utility or is it just clutter?
- ◎ How could it be implemented?

Thanks!

Stop by the IMCR Wiki for onboarding materials (bit.ly/IMCRwiki)

NSF grants #1565103 and #1629233



EARTH CUBE

Slide Deck Credits

Special thanks to all the people who made and released these awesome resources for free:

- ◎ Simple line icons by Mirko Monti
- ◎ E-commerce icons by Virgil Pana
- ◎ Streamline iconset by Webalys
- ◎ Presentation template by SlidesCarnival
- ◎ Photographs by Unsplash & Death to the Stock Photo
(license)