

# Towards Building a CyberInfrastructure for Facilitating the Assessment, Dissemination, Discovery, & Reuse of Software and Data Products

Ritu Arora, Wayne State University, [ritu@wayne.edu](mailto:ritu@wayne.edu) (NSF award # 2314202 & 2037661);

Subhashini Sivagnanam, San Diego Supercomputer Center, [sivagnan@sdsc.edu](mailto:sivagnan@sdsc.edu) (NSF award # 2037656)

## Overview

The overarching goal of our project is to develop a software infrastructure for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products. As a preliminary step towards meeting this goal, we are prototyping: (1) **iTracker**, (2) **CompChecker**, (3) **Catalog**, and (4) **Select-A-License tool**.

The project has resulted in two prototypes of a catalog of scientific software and data products that can help the community in discovering these products while also helping the project leaders in publicly disseminating their products. Following are the URLs to these prototypes:

**Opuntia:** <https://opuntia.online>

**Software Discovery Platform (SDP):** <https://sdf.sdsc.edu>

Opuntia (pronounced as "up-un-ehia") is a software infrastructure for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products. It consists of a catalog of software and data products, and tools named as iTracker, CompChecker, and Select-A-License.

The Opuntia project is under active research and development, and new features are added to the public release iteratively.

**Catalog**

The catalog of products facilitates their discovery, dissemination, and reuse. A product can be added to the Opuntia catalog after logging in to this website and filling the form on the "Add Product" page. The login can be done either by using the credentials associated with an account created on the Opuntia website, or by clicking the link for "Login using CILogon" and choosing one of the existing accounts such as Google, GitHub, ORCID, or the home organization (if the organization is a part of the

**iTracker**

iTracker helps in tracking the user-defined metrics for evaluating software and data products deployed and used on different platforms and computing environments. While some types of metrics are dynamically tracked and updated in iTracker's database, others are currently updated statically. For the statically updated metrics, at the time of editing their products' information in the catalog, the product owners can specify the name of the metric of interest, and the measured value. They can also provide the

**CompChecker**

CompChecker is a tool that can assist with checking the legal and technical interoperability of the products. Currently, it can assist with comparing any two cataloged products or licenses directly and provide suggestions on whether the products or licenses being compared can be mixed with each other or not. Specific clauses of licenses, such as those around sublicensing, commercialization, warranties, modifications, and reuse are compared with each other to provide guidance. In the next phase of research and

**Select-A-License**

Select-A-License is a decision-support tool that can help in selecting appropriate licenses for software and data products. It is designed to help in selecting a license from amongst the 65 open-source licenses in the Opuntia database. The output of this tool is driven by the responses it receives to a set of questions it presents. For convenience, this tool also provides the text that can be used to prepare the license agreement for each license that it suggests.

Screenshot of a prototype of a catalog - Opuntia landing page

**Software Discovery Portal**  
Portal for Product Discovery, Dissemination, & Reuse

Search feature is available

Discovery Catalog

The catalog will facilitate the discovery, dissemination, and reuse of NSF funded products, and help in reporting on the evaluation metrics as defined by the product owners. This can be important in measuring the short-term and long-term impact of the products, and hence, in facilitating their evaluation, adoption, and sustainability. The catalog will showcase the parent-child relationship between products registered in the catalog and establish their provenance trail. The information on product licenses and software stacks will be displayed for each product which can help identify the talent required for building a new product based on the existing software stack.

License Compatibility Checker

The license compatibility checker, **CompChecker**, can help you in checking the compatibility of the different software licenses. You can select any two licenses from the page and find out if they can

Screenshot of a prototype of the catalog - SDP landing page

Search feature is available

**Product Details**

Product name (Required) Product description (Required)

Product Type (Required)

License (Required) Keywords describing product (Required) Product website URL Product download URL

Product install instructions URL Product software stack Other products that use this product How to cite the product

Related Products How to acknowledge the product

DOI number

Funding agency award no (Required) Funded project title PI Name (Required) PI Email (Required)

GitHubuser/GitHubrepo GILabuser/GILabrepo UsageMetric file URL Google Analytics URL

Metric Usage

Add Additional Metric

Submit

## Metrics Gathering

**Automatic** collection of metrics from publicly accessible GitHub repositories, GitLab repositories, and Google Analytics reports (as exported file) is supported. We have also developed an approach for making the code self-reportable on its use. Additional work is needed for ensuring that user-consent is obtained from the user before the software starts collecting and reporting its own metrics.

Feature for providing manually defined and editable set of metrics is also supported.

## Software Interoperability

Can the selected software or data products interoperate with each other? Explored two aspects to find answers to this question:

- Technical compatibility:** categorized software products (e.g., middleware, libraries, storage management software, and web applications), and analyzed how the different software products could be deemed compatible or incompatible based on the level of data exchange that can be supported between them directly or through adaptations
- Legal compatibility:** automated license comparison

Software Stack: Javascript, Ruby, Lua, NGINX, Apache

Other products that use this product: Azure OnDemand, GCP OnDemand

Citation: Hudak et al., (2018). Open OnDemand: A web-based client portal for HPC centers. Journal of Open Source Software, 3(25), 622. <https://doi.org/10.21105/joss.00622>

Acknowledgement: This material is based upon work supported by the National Science Foundation under grant numbers 1534949 and 1835725, and under active development by a team from the Ohio Supercomputer Center, U. of Buffalo CCR, and Virginia Tech

NSF Award No: 1534949, 1835725

Funded Project Title: Frameworks: Software NSCI-Open OnDemand 2.0: Advancing Accessibility and Scalability for Computational Science through Leveraged Software Cyberinfrastructure

PI Name: David Hudak

PI Email: [dhudak@osc.edu](mailto:dhudak@osc.edu)

Related Products: Open XDMoD, ColdFront

Last Update:

DOI No: <https://doi.org/10.21105/joss.00622>

**Usage Metric**

Metric Name	Count
github_commits	6828
github_contributors	49
github_forks	67
github_stars	172

Relationships between different products are captured

Screenshot showing a portion of the catalog page for Open OnDemand software product

The metrics shown below are automatically updated when these get updated in GitHub

You can use the **Select-A-License tool** for deciding on the appropriate licenses for your software and data products. To begin using this tool, you can click on the "Choose Your License" button and respond to the questions posed. Based on your responses, appropriate license/s meeting your criteria will be displayed, and the displayed license name will be hyperlinked to the license text for your convenience.

**Choose Your License**

What is your product type?

Software Data

Do you want to release the data with copyright/s and/or related rights?

Yes No

Do you own the copyright and related rights on the data and all the parts within it?

Yes No

Are others allowed to make derivative products using your data?

Yes No

Do you require that the products derived from your data are shared under a compatible license?

Yes No

Are others allowed to use your data for commercial purposes?

Yes No

We recommend the following option/s for the license. Please feel free to click on the license name/s to see the license text.

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

More Details - <https://creativecommons.org/licenses/>

The text of the suggested license

Screenshot showing the Select-A-License tool in action

**Software-Software Comparison**

Below is the list of products that are currently in the Opuntia catalog. You could compare them for legal-interoperability by selecting and clicking on the check compatibility button.

Software 1: Basil

Software 2: Greyfish

Check Compatibility

**License-License Comparison**

Below is the list of licenses that are currently in the Opuntia catalog. You could compare them for legal-interoperability by selecting and clicking on the check compatibility button.

License 1: --- License ---

License 2: --- License ---

Check Compatibility

**License Compatibility Result**

Software with license BSD-3-Clause, is compatible with the software having license LGPL-3.0-only. After combining the software, your final license should be the leading license or the most restrictive one out of the two. In this case, your final license could be 'LGPL-3.0-only'.

**Additional Results of License Comparison**

'Y' in 2nd and 3rd columns shown below denotes 'Yes' in response to the questions in the property column. The blank values indicate that the property is either Not Applicable or license does not have clear coverage of the property.

Property	BSD-3-Clause	LGPL-3.0-only
Are you allowed to use the product released under this license or reproduce it?	Y	Y
Are you allowed to distribute the product released under this license?	Y	Y
Are you allowed to modify or merge the product released under this license?	Y	Y
Are you allowed to merge the product released under this license with other software or data product?	Y	Y
Is sublicensing of the product permitted?	Y	
Is commercial use of the product permitted?	Y	Y
Are you allowed to freely use the product authors' patent?		Y

Screenshot showing CompChecker in action

## Conclusion & Future Work

The project can help in quantifying the societal impacts of the NSF-funded software and data products and in supporting their dissemination, discovery, and reuse. The project can be scaled to support the products funded by other funding agencies too.

## Acknowledgement

We are grateful to the **National Science Foundation** for funding this project. We are also grateful to **ACCESS** for providing us with the allocation on **Jetstream2** cloud computing platform on which Opuntia (<https://opuntia.online>) is running. We use **CILogon** for supporting federated logins and are grateful to the CILogon team for their services.