

# Open Core Data

Quick Overview



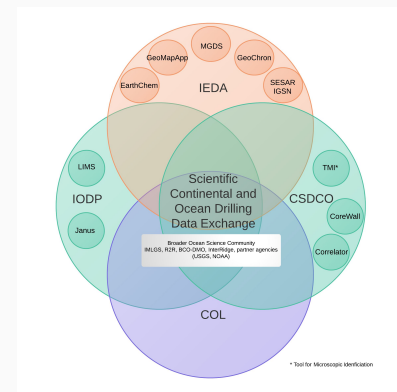
# Open Core Data

<http://opencoredata.org>

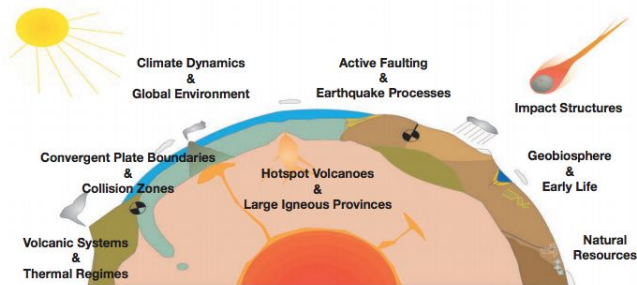
*Open Core Data is an infrastructure focused on making data from scientific continental and ocean drilling projects semantically discoverable, persistent, citable, and approachable to maximize their utility to present and future geoscience researchers.*

The specific benefits that Open Core Data will bring to the community include:

- Integrating data management systems and services from multiple facilities, adding scientific value and economies of scale;
- Improving scientific drilling data discoverability and reuse through integration with evolving data infrastructures, augmenting existing domain-specific data systems (e.g. Neotoma, MagIC, EarthChem, PBDB, dbSEABED, GPLates) with scientific drilling data;
- Capturing and integrating PI-generated, post-moratorium scientific drilling data;
- Providing standards-based interoperability for tools to visualize and analyze scientific drilling data;
- Promoting and facilitating a Geoscience community of practice in data publication and citation;
- Providing a scalable resource that other communities and facilities could employ in the future (e.g. ANDRILL, ICDP, MGG-funded marine core repositories).

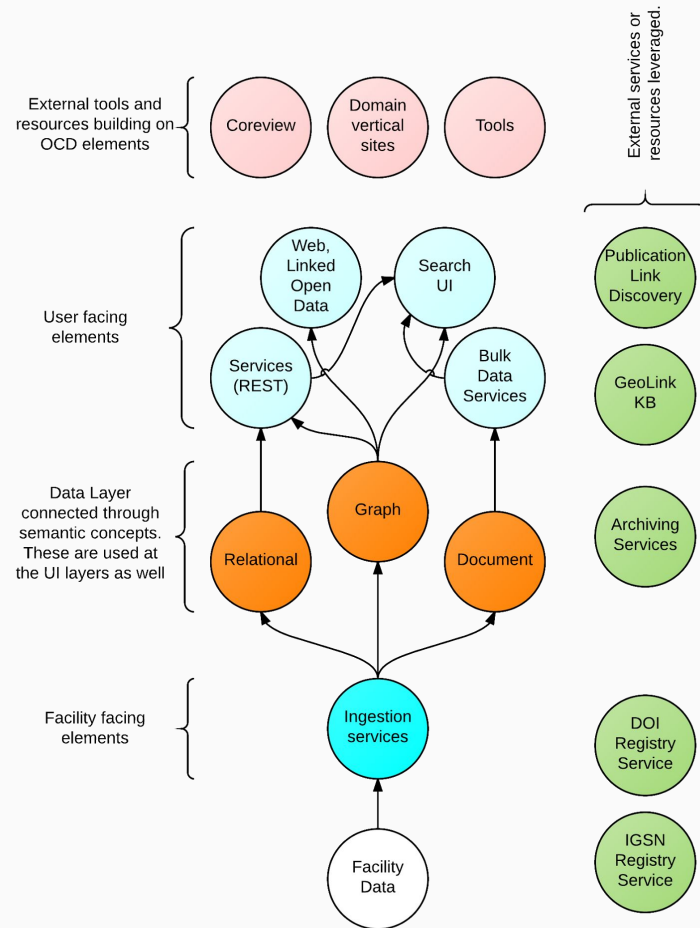


# Quick Overview



the **thrill to drill**

- Developed in discussion with IEDA, CSDCO and JRSO
- Layer over facility data to enable value added functionality
- Supported as a supplement to IEDA for this phase
- NSF GeoInformatics proposal in now



# Open Core Data: Status

- Phase 1 development started in 2015 with a supplement to the Interdisciplinary Earth Data Alliance (IEDA)
- Current state shown at [opencoredata.org](http://opencoredata.org). All code is open source, available at: [github.com/OpenCoreData](https://github.com/OpenCoreData).
- NSF has informed the PIs that the Open Core Data Geoinformatics proposal will be funded in full.
- Initial work focuses on 4 major themes:

## ***Migration***

Initial work on moving data from JRSO and CSDCO holdings

~ 20K datasets with associated metadata so far

## ***Patterns and Models***

Exposing data and metadata in standards-based methods

Using multiple formats to maximize human and machine access to data sets

Examples:  
Schema.org  
RDF (GeoLink and others)  
CSV for the Web  
JSON-LD

## ***Access***

Focus on both human and machine access. Integrating citable data (via DOIs) into science tools like iPython and others.

## ***Discovery***

*Enhanced semantics* utilizing output from GeoLink (EarthCube Building Block) and other vocabularies

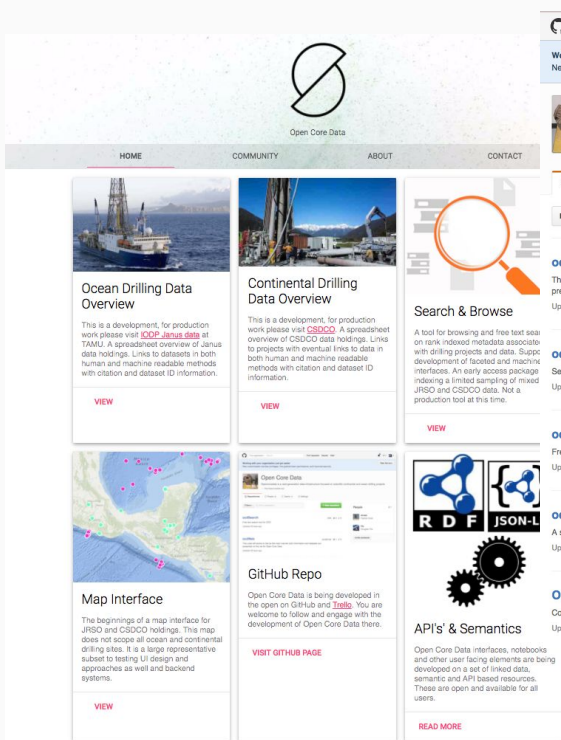
*Linked Open Data* structures for machine indexing

*Provenance and Citation* enhancement utilizing EarthCube and ESIP Federation outputs

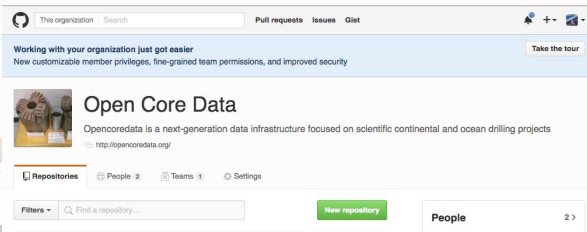
# Open Core Data: Tour

A quick tour of Open Core Data with a look at the Linked Open Data, API and notebook plans.

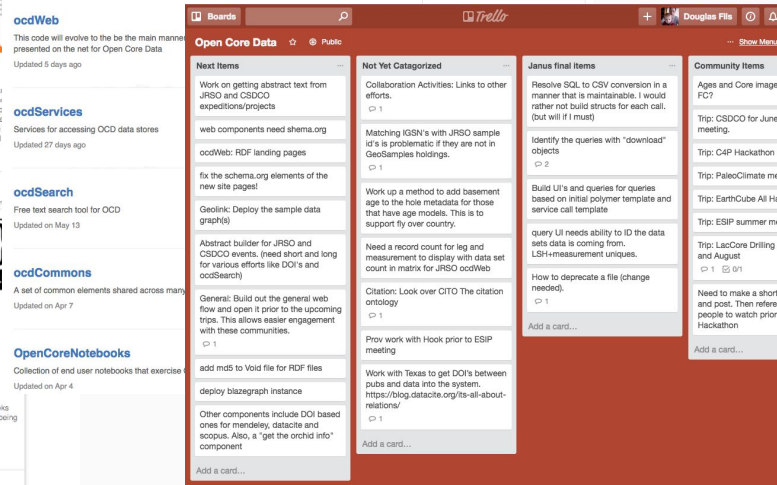
<http://opencoredata.org/> <https://trello.com/b/dHxNEnCN/open-core-data> <https://github.com/OpenCoreData>



The Open Core Data website features a navigation bar with links to HOME, COMMUNITY, ABOUT, and CONTACT. The main content area is divided into several sections: Ocean Drilling Data Overview, Continental Drilling Data Overview, Search & Browse, Map Interface, and GitHub Repo. Each section includes a brief description and a 'VIEW' button. The Map Interface section highlights the beginnings of a map interface for JRSO and CSDCO holdings. The GitHub Repo section mentions that Open Core Data is being developed in the open on GitHub and Trello.



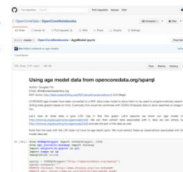
The Open Core Data GitHub repository page shows the repository name, description, and a list of repositories. The description states: "Opencoredata is a next-generation data infrastructure focused on scientific continental and ocean drilling projects". The page also includes a search bar and a list of people.



The Open Core Data Trello board is organized into several columns: Next Items, Not Yet Categorized, Janus final Items, and Community Items. Each column contains a list of tasks and items, such as "Work on getting abstract text from JRSO and CSDCO expedition/projects", "Matching GSN's with JRSO sample", and "Resolve SQL to CSV conversion in a manner that is maintainable". The board also includes a search bar and a list of people.

## Notebooks

As part of the effort to enable more machine access to Open Core Data holdings we will develop a set of notebooks showing examples of using the APIs. Linked data and semantics that are part of Open Core Data. The goal then is to foster connections and development of further examples of their use in tools like Python, R, Matlab, and other working environments scientist use. This repository ([OpenCoreDataNotebooks](https://github.com/OpenCoreData/OpenCoreDataNotebooks)) is just the start and we will strive to build connections and develop examples that can be shared here elsewhere.

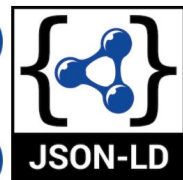


## API's



As the OGD APIs are developed they will be described using the swagger approach. Swagger (ref <https://swagger.io/>) is a method to describe RESTful API's in a way that is both human and machine readable. It also provides for interactive documentation, SDK's for client generation and discoverability. A listing of current API's can be found at <http://opencoredata.org/community/swagger-ui/>. As now API's are developed they automatically show up here with documentation pulled from the source code itself.

## Semantics



Open Core Data is incorporating several Linked Open Data (<https://www.w3.org/standards/semanticweb/data>) patterns and semantic approaches. This includes the development of supporting SKOS based vocabularies and integration with larger ontologies like those in development by the EarthCube Geologic Building Block (<http://www.eearthlink.org/>). Additionally Open Core Data hosts some of its