

Open Core Data: Connecting scientific drilling data to scientists and community data resources

Transformative Data Infrastructure for Integrating and Accessing Scientific Drilling and Coring Data (opencoredata.org)

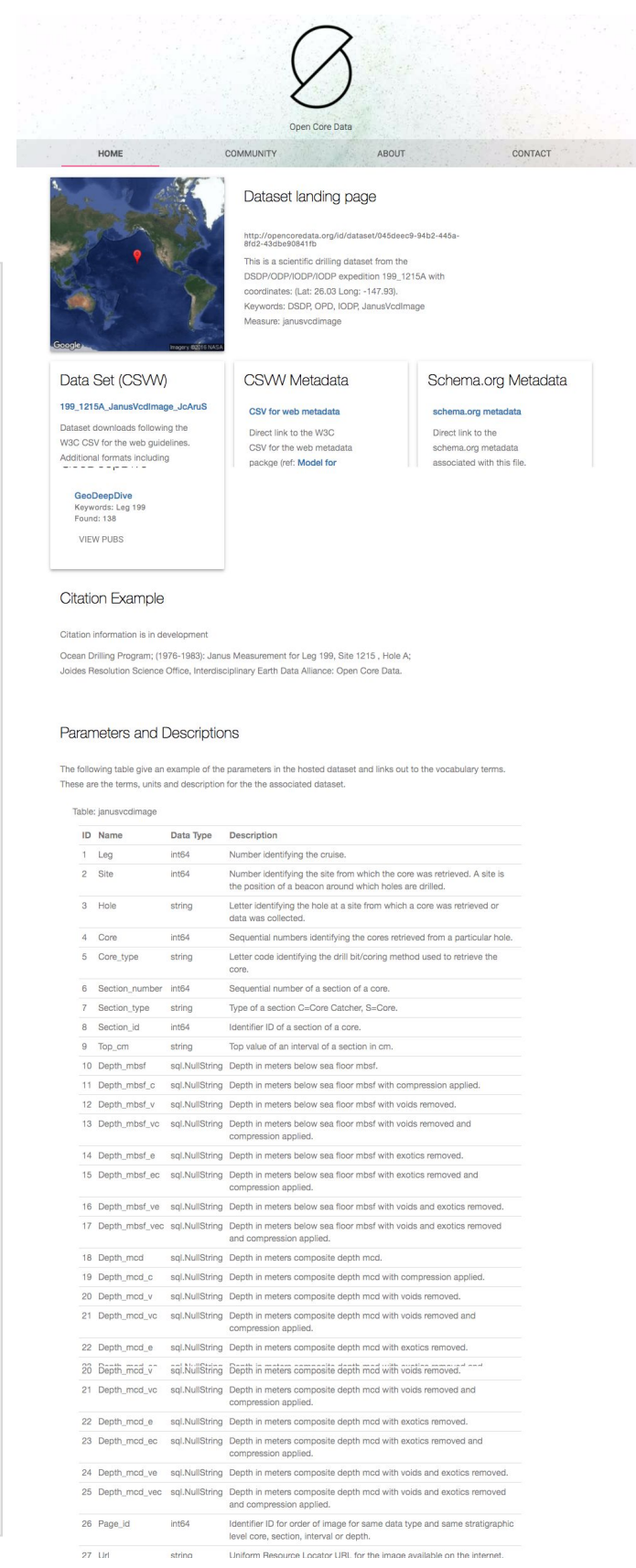
IN53C-1898

Description & Vocabularies

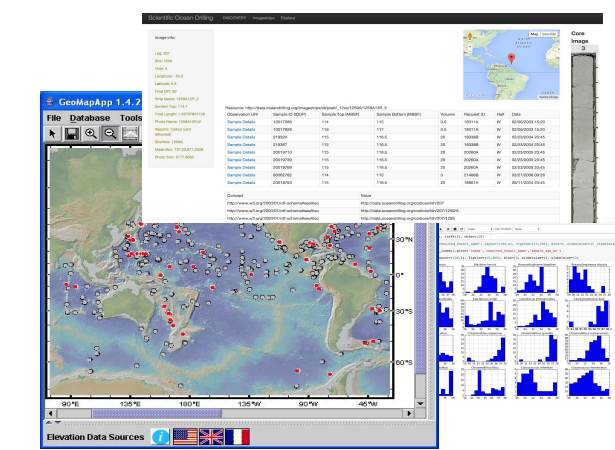


Data sets and resources inside Open Core Data are annotated with information detailing:

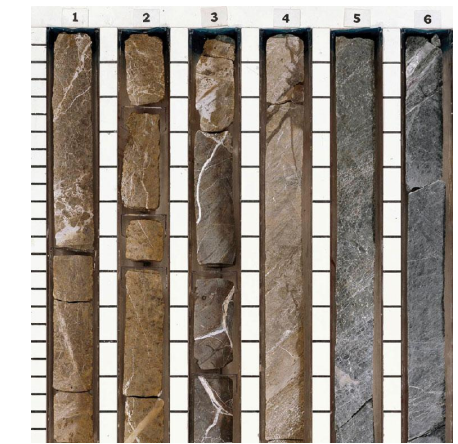
- Licensing information
- Connections to community vocabularies like GeoLink and others for defined classes and concepts
- Where necessary defining location vocabularies
- Connections to associated resources such as geologic stages from CSIRO and data types defined by IODP and CSDCO



Persistent Identifiers



Datasets follow DataCite and IEDA metadata profiles and are assigned DOIs. DOIs are exposed through APIs and also schema.org/Dataset and other methods.



Samples are identified through the International Geo Sample Number (IGSN)



People are identified by Orcid identifiers where individuals have used them and they are associated with resources inside Open Core Data

Provenance data

Notional approach based on Prov-ES



Platform Input Data Software Output Data

Information about input data sources such as CSDCO, Janus/IODP or LIMS/IODP are associated with software and output dataset results.

This information is associated whenever possible with persistent identifiers.

This information is passed along to downstream users who can then extend the provenance chain with actions and actors.

Douglas Fils
Data Manager
Ocean Leadership / Scientific
Drilling
dfils@oceanleadership.org

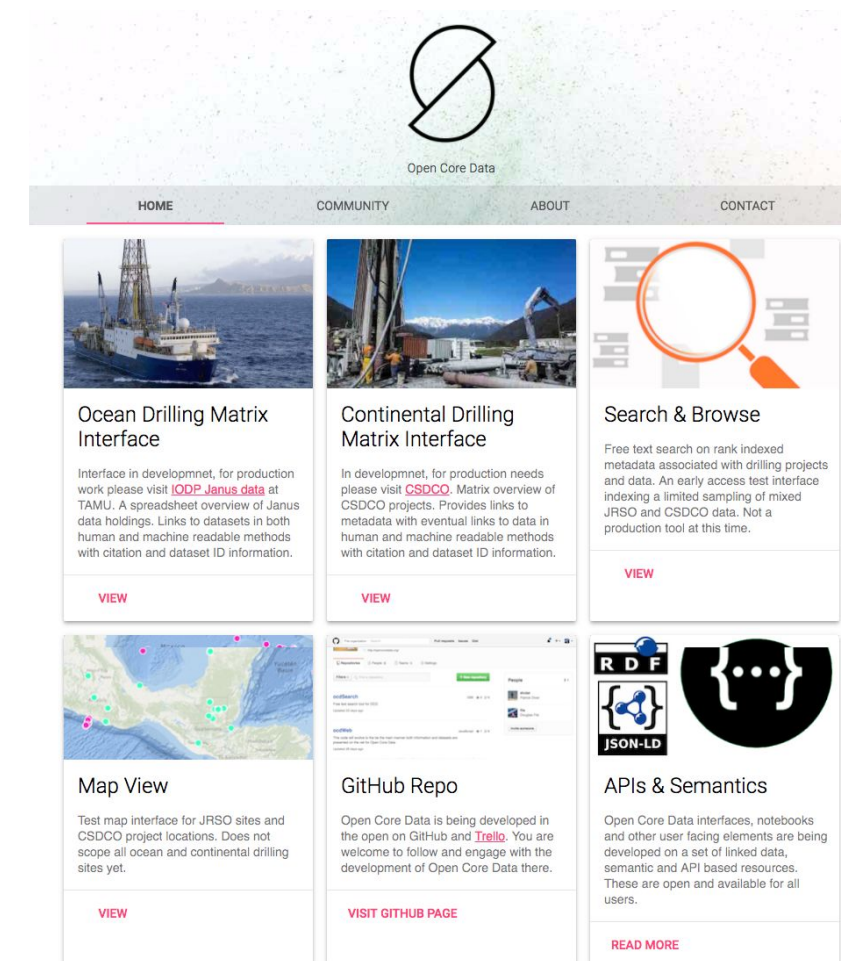
ORCID ID



Anders Noren
Director
Continental Science Drilling
Coordination Office (CSDCO)
noren021@umn.edu

Kerstin Lehnert
Director
Interdisciplinary Earth Data
Alliance (IEDA)
lehnert@ldeo.columbia.edu

Discovery & Access: Data Distribution Approaches



Web browsing and forms interfaces to access CSV files to support existing workflows.



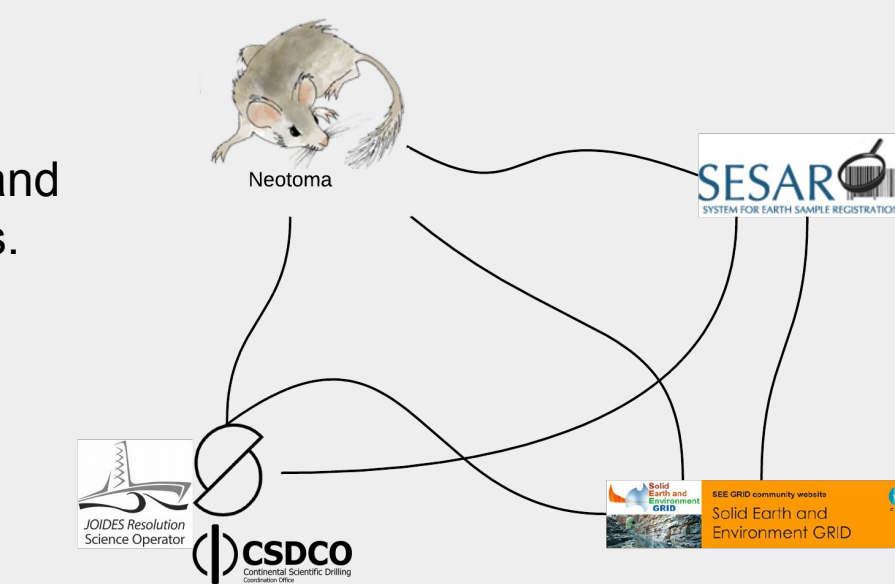
Semantic Web

- SPARQL 1.1 query options
- Linked Open Data patterns for crawling
- Embedded schema.org and W3C CSV for the Web metadata (JSON-LD)

All data in Open Core Data is available in RDF. Bulk download of graphs is available and the graphs are queryable via SPARQL.

Leveraging connections based on shared concepts in vocabularies like GeoLink or SeeGrid geologic time and persistent identifiers like IGSNs or DOIs.

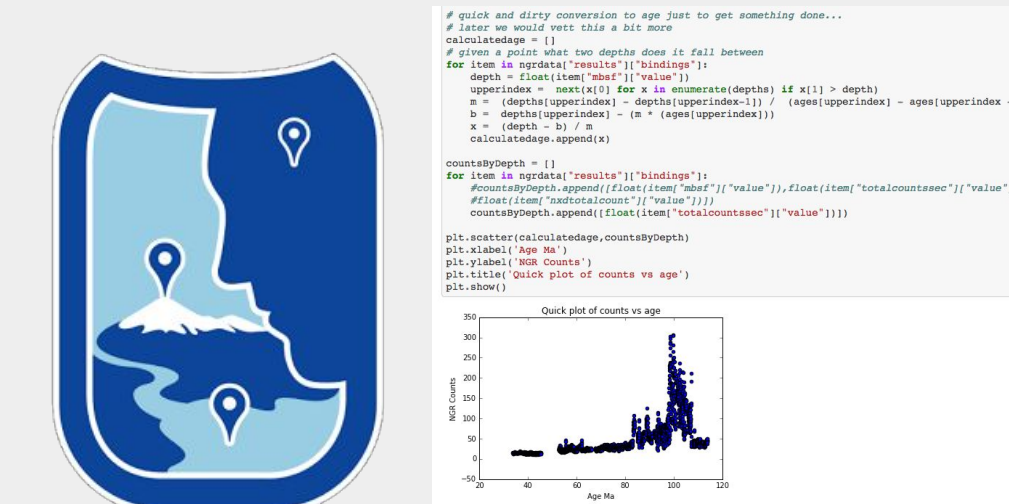
Use these connections to support development of provenance and citation approaches.



Application Program Interfaces

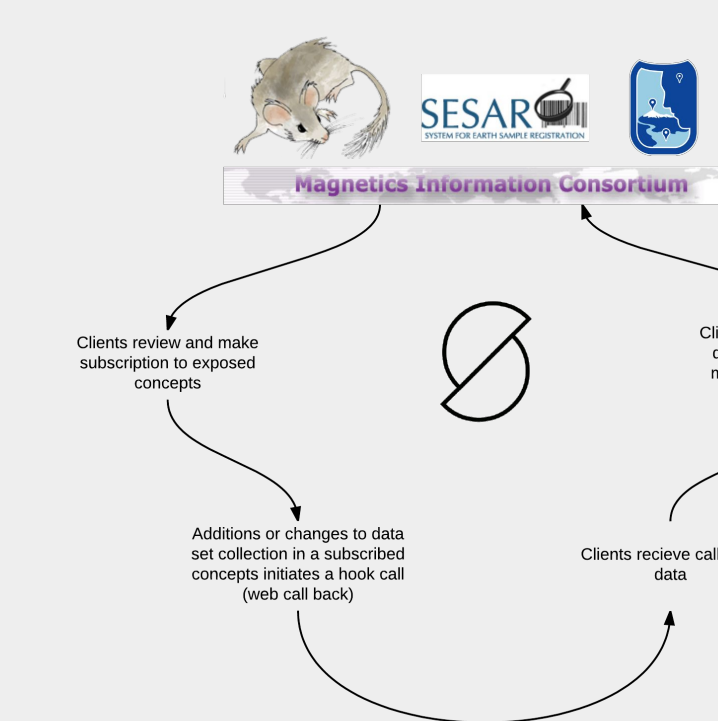
- RESTful services described via Swagger <http://opencoredata.org/common/swagger-ui/>
- Exposing services via gRPC is also under exploration to facilitate more structured approach based around ProtoBuf to precisely define services and autogenerate client stubs

APIs support interaction with remote tools such as iPython/Jupyter or R. Also, Open Core Data is working with the NSF funded Fly Over Country to support exposing data in their outreach application (Android/iOS/Web).

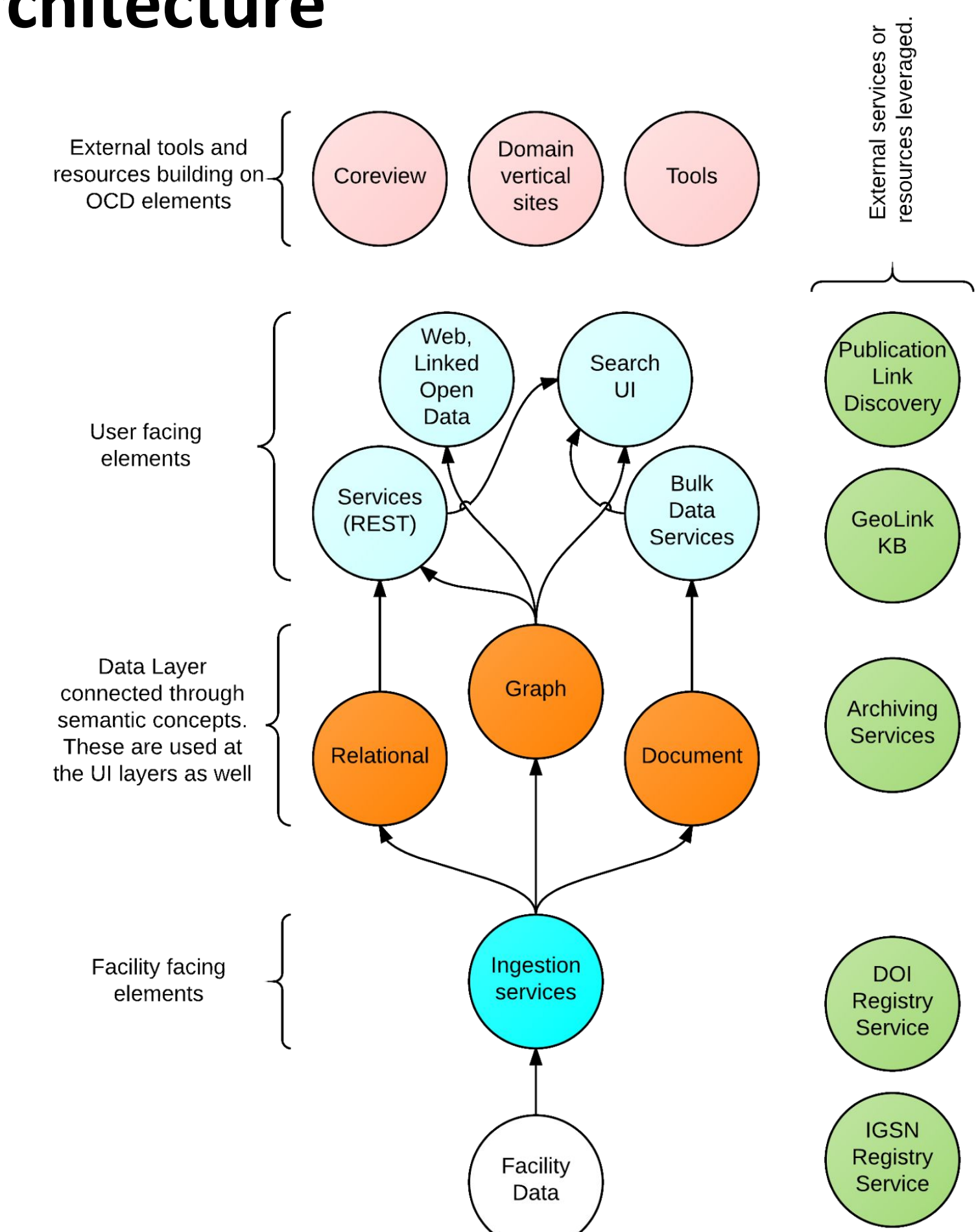


Programmable Web

- Support push notification to community data repositories based on tags and using web hooks style subscription approaches
- Expose data for the above notifications based on defined concepts such as:
 - Measurement
 - Parameter
 - Person
- Difference management via MD5 allows quick checks for datasets that are new or deprecated

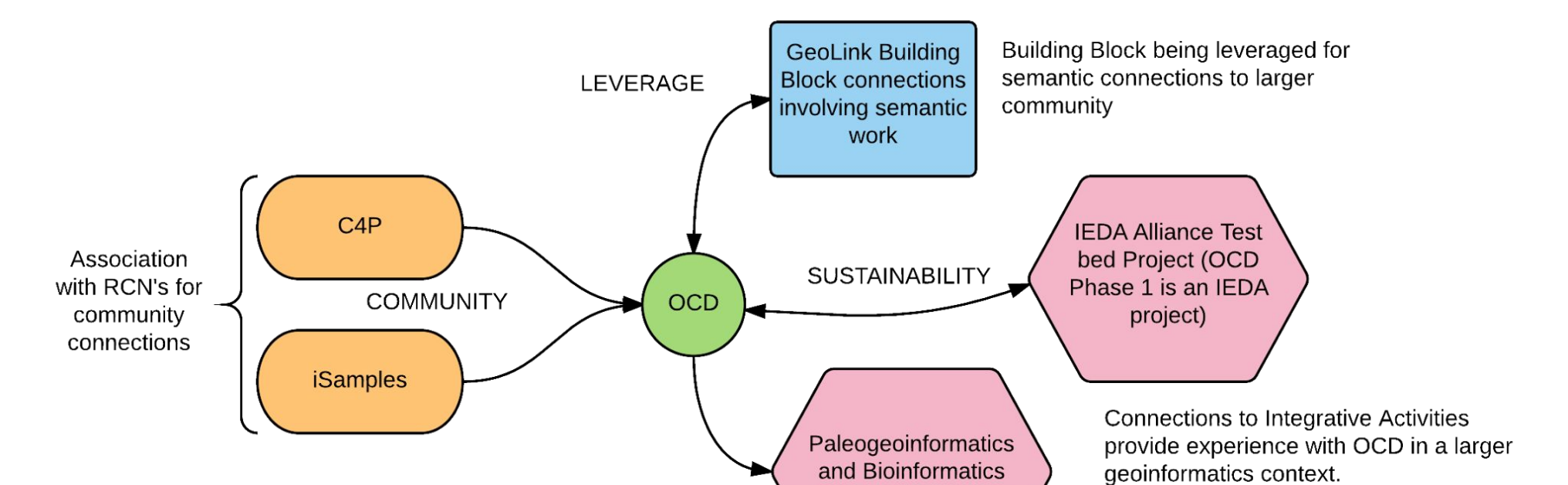


Architecture



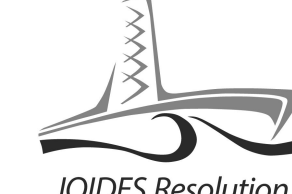
EarthCube Connections

- C4P: <http://earthcube.org/group/c4p>
- GeoLink: <http://geolink.org>
- iSamples: <http://earthcube.org/group/isamples>
- Paleogeoinformatics and Bioinformatics Cyberinfrastructure: <http://earthcube.org/group/paleogeoinformatics-bioinformatics-cyberinfrastructure>
- Alliance Testbed Project: <http://earthcube.org/group/alliance-testbed-project-atp>



Acknowledgments

Open Core Data development is funded by NSF: EAR-1550917, EAR-1550887, EAR-1550787



Further information

A copy of this poster is at: <https://github.com/OpenCoreData/ocdDocumentation>

All code for effort is at Github: <https://github.com/OpenCoreData>

Session Number and Title:
IN53C: Innovative Tools and Services to Enable Data Use across Broad User Communities II Posters