

**U.S. Geological Survey (USGS)
Community for Data Integration (CDI)
Request for Proposals (RFP)**

For Fiscal Year 2016

Issue Date: September 9, 2015

Table of Contents

Overview	3
Eligibility.....	3
Estimated Available Funds.....	3
Estimated Schedule for Submission, Review, and Awards.....	3
Application Process.....	3
Project Reporting.....	4
Introduction to the Request for Proposals	4
Evaluation Criteria for the Statement of Interest and Full Proposal.....	5
Appendix A – Statement of Interest (SOI) Guidance.....	6
Appendix B – Invited Full Proposal Guidance.....	7
Full Proposal Template	7
A. Cover Sheet Page Instructions (max. 1 page).....	7
B. General Public Summary (max. 1 page).....	8
C. Proposal Narrative (max. 7 pages).....	8
D. Appendices	9
Budget Form.....	10
Data Management Planning Form.....	10
Appendix C – Example Cover Sheet	11
Appendix D – CDI Science Support Framework (SSF).....	12
Appendix F – CDI Coordinators	15

Overview

This document describes the CDI Request for Proposals (RFP) process for fiscal year 2016 (FY16). The CDI RFP consists of a two-stage process: Statements of Interest (SOI) and Full Proposals. Contact cdi@usgs.gov with any questions about the CDI RFP.

Eligibility

Any USGS Mission Area, Region, Program, Center, Office, or duty station and their partner(s) are encouraged to apply. All proposals must specify a **USGS Federal employee** as lead Principal Investigator (PI). USGS personnel may be involved in more than one prospective or existing proposal, but may be the lead PI on only *one* proposal.

Estimated Available Funds

Funding for CDI projects varies from year to year and is directly influenced by the overall USGS budget. Since fiscal year 2009, CDI projects have been funded at a total of about \$500,000 (or less) each year and it is anticipated that FY16 will be similar to previous years. Applicants can request funding up to \$50,000 maximum and projects must include a minimum of 30% in matching funds.

Estimated Schedule for Submission, Review, and Awards

Submission deadline for Statement of Interest.....	October 9, 2015
SOI Voting Closing Session.....	November 3, 2015
Applicants Notified and Full Proposals Requested.....	November 20, 2015
Invited Full Proposals Due.....	January 22, 2016
Final Projects Reviewed and Identified.....	February 19, 2016
Funded Projects Announced.....	March 9, 2016

Application Process

- 1. Submit Statement of Interest** – Interested parties may submit a 2-page SOI for initial evaluation. An SOI template for guidance can be found in [Appendix A – Statement of Interest \(SOI\) Guidance](#). SOIs must be submitted via the online proposal management system at https://my.usgs.gov/CDI_RFP (*sign in using AD credentials*). Project work must be completed by *September 30, 2016*.
- 2. Evaluate Statements of Interest** – All SOIs will be reviewed by the CDI Community. The SOIs will be considered according to the [Evaluation Criteria for the Statement of Interest and Full Proposal](#). Note that the CDI Executive Sponsor may select proposals that support Bureau emerging priorities and a percentage of CDI Funds may be awarded to proposals that address those topics.
- 3. Invitation to Submit Full Proposal** – Based on the SOI evaluation, selected SOI applicants will be invited to submit a Full Proposal that will consist of a cover sheet, general public summary, proposal narrative, and budget justification. Applicants must also submit a [Budget Form](#) and [Data Management Planning Form](#). Supporting documents must include curriculum vitae (CVs) for Principal Investigator(s) and may include CVs for other collaborators or letters of support. View the Full Proposal format in [Appendix B – Invited Full Proposal Guidance](#).
- 4. Full Proposal Review Process** – Full Proposals will be evaluated according to the [Evaluation Criteria for the Statement of Interest and Full Proposal](#). Proposals will be reviewed by a panel consisting of a professional peer group that is knowledgeable in data management, information technology, and other relevant disciplines in the context of the CDI. All efforts will be made to include broad disciplinary representation and expertise. Reviewers may consult with subject experts outside of the Review Panel as needed. Reviewers will be asked to sign a Conflict of Interest Statement and Certification to ensure objectivity of the evaluation. Proposals will be scored by reviewers and ranked through a peer consensus process. Recommendations by the Review Panel will be presented to the CDI Executive Sponsor for final selection.

Project Reporting

All CDI funded projects will be required to provide an informal mid-year briefing to the *CDI Facilitators* to communicate the status of the project. Projects must also contribute to the CDI Annual Report, which will be compiled in late 2016. At that time, project leads must provide a brief report describing the project purpose and objectives, as well as the benefits and deliverables with links to products or publications and a visual representation of project accomplishments. The project leads must also update and submit their *Data Management Planning Forms*.

Introduction to the Request for Proposals

The CDI seeks to build and share knowledge about topics such as data integration, data handling and stewardship, scientific computing, and approaches for knowledge delivery. The main goal of CDI funding is to improve our collective knowledge about how to create better, longer-lasting and more accessible science products by leveraging the tools, methods and datasets available to the Earth and biological science communities. The CDI places high value on innovative projects that, in the near-term, produce new and reusable ideas, methods or tools that have an impact beyond a single Program, Center, Region, or Mission Area. CDI project proposals will be evaluated based on the following guiding principles:

- Focus on targeted efforts that yield near-term benefits to Earth and biological science
- Leverage existing capabilities and data
- Implement and demonstrate innovative solutions (e.g. methodologies, tools, or integration concepts) that could be used or replicated by others at scales from project to enterprise
- Preserve, expose, and improve access to Earth and biological science data, models, and other outputs
- Develop, organize, and share knowledge and best practices in data integration

CDI Science Support Framework

Project proposals must also relate to elements of the *CDI Science Support Framework* (SSF), which categorizes and relates the activities and processes through which research data flows, and upon which the CDI operates. These elements include Data Management, Knowledge Management, the stages of the *Science Data Lifecycle Model*, Applications, Web services, Semantics, Information, Data assets, and Communities of Practice (See *Appendix D – CDI Science Support Framework (SSF)*).

Examples of projects that relate to the goals and Science Support Framework

- Delivery of an immediate benefit to solve an existing data integration challenge, such as methods for blending datasets, or best practices for alignment/assimilation of data at different scales
- Creation of innovative environments, tools, data stores, or services that enable discovery and usage of USGS data. This includes design patterns, management approaches, or products like web services or other software that can be used by other data publishers
- Development of standards or best practices for data management through community consensus building, such as convening a workshop and writing a white paper
- Development of a general ontology or tools for tagging data in support of standards and environments to facilitate discovery, understanding, and integration
- Testing or application of the aforementioned to a new, real-world problem to demonstrate and document strengths and issues for the purpose of feedback and improvement
- Exploitation of advanced or emerging technologies or approaches that enable new forms of USGS scientific knowledge creation or communication, such as developing mobile computing applications for rigorous data collection, or establishment of scientific policies or protocols around the novel component
- Development of innovative practices, methods, and strategies to better exploit collected data resources, such as data mining, parallel processing, large-scale data analysis, or scientific computing techniques and to improve data sharing, facilitate data preservation, and encourage lifecycle data management
- Development of vehicles to communicate or share knowledge, such as a committee to propose protocols/standards, workshops, online or in-person training course/materials, white paper, etc.
- Any of these or similar topics that have not been addressed otherwise within the agency

Proposal Concepts that should *not* be submitted to the CDI

The CDI does not seek to supplant traditional natural science research or to fill a funding gap on a project supported elsewhere. Examples of topics that are a poor fit for CDI funding include:

- Supporting the collection of new data or field research.
- Monitoring, assessment, or dataset creation projects. Although the CDI may fund the creation of some broadly-usable (“foundational”) data content, this is normally considered out of scope.
- Projects focused exclusively or even predominantly on non-CDI research topics, such as cartographic science, hydrologic processes, ecosystem response, or climate change studies.
- Projects that would normally be funded by individual Program Areas or other proposal processes such as the *John Wesley Powell Center for Analysis and Synthesis*, *Center of Excellence for Geographic Information Science (CEGIS)*, *Innovation Center for Earth Sciences (ICES)*, and *Office of Organizational and Employee Development (OED)*.

Examples of past CDI Projects: <http://www.usgs.gov/cdi/products-publications.html>.

Evaluation Criteria for the Statement of Interest and Full Proposal

Both the SOIs and Full Proposals will be evaluated based on the following six criteria. SOIs will only be expected to provide a concise statement in each of the criteria while Full Proposals must provide more detail. The evaluation weights (percentages) will only apply to the Full Proposal evaluation. For instructions to fill in SOIs, see [Appendix A – Statement of Interest \(SOI\) Guidance](#); for Full Proposals, see [Appendix B – Invited Full Proposal Guidance](#).

Scope (25%)

Evaluation will be based on whether the proposal adequately demonstrates the need for the effort/activity, how much the proposal contributes to the guiding principles and element(s) of the CDI Science Support Framework, and whether the effort has potential impact beyond a single Program, Center, Mission Area, or Region. CDI projects will also be evaluated on anticipated return on investment (e.g. cost savings, code utilization, publications, operational efficiencies, etc.).

Technical Approach (25%)

Evaluation will be based on the reasonableness of the technical approach applied to the problem and whether the approach is innovative or employs a proven, reliable technique that is appropriate to the problem.

Project Experience and Collaboration (25%)

Evaluation will be based on the appropriateness of the experience, special qualifications, and skills possessed for successful completion of the proposed project. Evaluation will also consider whether the inclusion of interdisciplinary or cross-Mission Area/Region collaboration and partnerships has been pursued where appropriate.

Sustainability (15%)

Evaluation will be based on how well the proposal describes the intended sustainability of the project deliverables (products, tools, services, metadata) for long-term access, reusability, and potential for integration. All products resulting from CDI projects must comply with the new [Office of Science Quality and Integrity Instructional Memoranda](#) on data management. These products must be freely shared and made available, without charge or restriction, to the CDI, the broader USGS community, and beyond as appropriate. Starting this year, software products developed with CDI funding must, at a minimum, be maintained in a [USGS Stash Repository](#).

Budget Justification (5%)

Evaluation will be based on whether the budget is at or below \$50,000 and meets the minimum 30% in-kind match. The budget should also include travel to one CDI-sponsored event. Evaluation will also consider whether justification of salaries and contractor costs, travel, and equipment/publication costs are appropriate to project needs and the work hours proposed are reasonable within the timeframe. Projects with contractor support must describe how the contract work will be managed and documented to ensure that products are USGS property.

Timeline (5%)

Evaluation will be based on clear presentation of the project phases and milestones described in the technical

approach and the feasibility of the proposed workload given the project duration.

Appendix A – Statement of Interest (SOI) Guidance

The SOI document will be **two pages maximum**, and prepared using Times New Roman 11 point font with one-inch margins, and saved in Portable Document Format (**PDF**). The first page is Administrative and Budget information, and the project description is on the second page. Each SOI must be separately registered within the CDI RFP online proposal management system (https://my.usgs.gov/CDI_RFP). Please respond to all questions in the proposal management system or the submittal will be considered incomplete and invalid.

SECTION 1. PROJECT ADMINISTRATIVE INFORMATION (½ page)

- Indicate which element(s) of the CDI Science Support Framework the proposal relates to
- Project title
- Name of lead USGS cost center requesting funding
- Name of USGS Principal Investigator, *ORCID ID*, mailing address, telephone, fax, and email
- Names and contact information for additional Principal Investigators or collaborators
- Short description of the project (two or three sentences)
- List of anticipated deliverables from the project

SECTION 2. ESTIMATED BUDGET (½ page)

Budget Category	Federal Funding “Requested”	Matching Funds “Proposed”
1. PERSONNEL (SALARIES including benefits):		
Federal Personnel Total:	\$	\$
Contract/Collaborator Personnel Total:	\$	\$
Total Salaries:	\$0	\$0
2. TRAVEL EXPENSES:		
Travel Total (Per Diem, Airfare, Mileage/Shuttle) x # of Trips:	\$	\$
Other Expenses (e.g. Registration Fees):	\$	\$
Total Travel Expenses:	\$0	\$0
3. OTHER DIRECT COSTS: (itemize)		
Equipment (including software, hardware, purchases/rentals):	\$	\$
Publication Costs:	\$	\$
Office Supplies, Training, Other Expenses (specify):	\$	\$
Total Other Direct Costs:	\$0	\$0
Total Direct Costs:	\$0	\$0
Indirect Costs (%):	\$0	\$0
GRAND TOTAL:	\$0	\$0

**Note: Travel must be included for at least one representative to attend a CDI event. Travel cannot include data field collection.*

SECTION 3. PROJECT SUMMARY (1 page)

Please provide a brief narrative summary of the project based on the goals and the Science Support Framework in context of the *Evaluation Criteria for the Statement of Interest and Full Proposal* listed in the RFP.

Appendix B – Invited Full Proposal Guidance

Proposals must be submitted through the online proposal management system at https://my.usgs.gov/CDI_RFP. In addition to submitting the documents listed below, please complete any questions that appear within the online proposal management system.

Proposal Structure: Applicants must submit three separate documents to the online proposal management system.

- ✓ **A Full Proposal, single PDF document** (not to exceed 10MB) with:
 - Cover Sheet Page (max. 1 page) (See [Appendix C – Example Cover Sheet](#))
 - General Public Summary (not to exceed 200 words; submitted on a separate page within the proposal)
 - Proposal Narrative (max. 7 pages)
 - Budget Justification
 - Appendices (e.g. CVs – max. 2 pages each, letters of support – max. 1 page each)
- ✓ **A Budget Form** using the MS Excel template: <https://my.usgs.gov/confluence/display/cdi/CDI+FY16+RFP+Forms>
- ✓ **A Data Management Planning Form, single PDF document** (using the MS Word template): <https://my.usgs.gov/confluence/display/cdi/CDI+FY16+RFP+Forms>

Proposals should be formatted to standard letter size (8.5” W by 11” L). All proposals should be no more than 7 pages, single-spaced, not including the cover sheet, the general public summary, and appendices. Narrative (body) text must be rendered in Times New Roman 11 point font, excluding headings which must be formatted bold and 12 point. All pages following the cover sheet (including appendices) must be numbered. Failure to follow the stated guidelines may reflect negatively on the proposal.

All graphics, photos, illustrations, tables, graphs, and charts must be embedded directly in the proposal document and be specifically referenced at least once in the narrative (body) of the proposal. All graphics must be accompanied by a caption that describes the graphic. These count towards the total number of pages allotted.

Full Proposal Template

A. Cover Sheet Page Instructions (max. 1 page)

The cover sheet will aid reviewers and the review process by allowing them to easily distinguish between proposals and see each proposal’s basic elements at a glance. See [Appendix C – Example Cover Sheet](#).

CDI SSF Element(s): Indicate which element(s) of the CDI SSF the proposal covers.

Project Title: Include a descriptive title of proposed project.

Principal Investigator(s): List the lead USGS Principal Investigator with affiliation, mailing address, phone, fax, and email, as well as the same contact information for any co-Principal Investigators.

Fiscal Contact: Provide the lead USGS fiscal contact with affiliation, mailing address, phone, fax, & email.

Collaborators: Provide the names, organizational affiliation, mailing address, telephone number and email address for all project personnel involved (other than the Principal Investigator).

Abstract: Provide a 200-250 word abstract that briefly summarizes the project’s value or importance in the context of research science and data management/integration and its relevance to the CDI SSF. Include statements regarding assumptions or hypotheses that will be tested or data management/integration challenges that will be addressed. Describe any results, outcomes or products that will be generated by the project and their value or application to research science or data management/integration. Key sections

from the full proposal that must be summarized are 1) Scope; 2) Technical Approach; 3) Project Experience and Collaboration; 4) Sustainability; and 5) Expected Products/Outcomes.

Total funding amount requested: Provide the total project funding requested from CDI, not to exceed \$50,000.

Total in-kind funding: Provide the total in-kind funding from other sources. There must be a minimum of 30% in-kind match.

Expected Product(s) Generated: Include a basic list of the types of products that will be generated as a result of the project, including both ancillary and final deliverables, e.g. mobile application, fact sheet, GIS shape file, GIS data layer, desktop data entry application, online data entry application, online data cleaning application, USGS Blog article or press release, etc.

B. General Public Summary (max. 1 page)

Provide a synopsis of the overall project that is written for a general public audience, does not exceed 200 words, and is suitable for sharing on public Web sites and other outreach methods. Key points to include: Why is the project important? Why should the public care? How will the results of the project improve aspects of the goals of CDI and how do they resonate with stakeholders?

The general public summary should be submitted on a separate page within the proposal PDF document.

C. Proposal Narrative (max. 7 pages)

The main body of the proposal should consist of six sections which will be evaluated by the Review Panel. Below is a description of each section. For the evaluation criteria, see [Evaluation Criteria for the Statement of Interest and Full Proposal](#).

- **Scope (25%)**
Clearly outline the problem and the need for the effort/activity, identifying who will benefit from the outcome. Describe how the effort contributes to the guiding principles and element(s) of the CDI Science Support Framework. Discuss the impact the project will have on the achievement of these goals and its potential impact and return on investment beyond a single Program/Center, Mission Area, or Region. Proposals may include letters of support from USGS or outside partners indicating a clear need for this effort.
- **Technical Approach (25%)**
Outline the steps, methodologies, technologies, and resources to be utilized in implementing the project. This includes facilities, computational/analytic platforms and tools, hardware/software, and other equipment supporting the project and/or its products. Indicate the project implementation approach in adequate detail with clear milestones described. The approach must make reasonable sense to implement and solve this problem.
- **Project Experience and Collaboration (25%)**
Identify specific individual(s) roles, qualifications and skills represented in the project team and how they will contribute to the success of the proposed project. To promote integration and diffusion of knowledge across inter-disciplinary or cross-Mission Areas/Regions, seek to include collaboration/partnerships where appropriate. Proposals require brief CVs for the Principal Investigator(s) while optional CVs may be included for each team member as appendices.
- **Sustainability (15%)**
Describe the extent to which project results, products, and the data/metadata created will continue or

be sustained after the performance period, e.g. metadata creation resource established, Web presence, infrastructure support secured, data management plan implemented, or other sustainable measures. All products resulting from CDI projects must comply with the new *Office of Science Quality and Integrity Instructional Memoranda* on data management. These products must be freely shared and made available, without charge or restriction, to the CDI, the broader USGS community, and beyond as appropriate. Software products developed with CDI funding must at a minimum be maintained in a *USGS Stash Repository*. Full Proposal submissions require a *Data Management Planning Form* describing all deliverables from the projects including any digital/electronic products, metadata, and intended publications. Submissions may also include Memoranda of Understanding (MOU) and/or letters of support indicating commitment to the longevity of the project.

- **Budget Justification (5%)**

All proposals may not exceed \$50,000 in requested funding and must include a minimum of 30% in-kind match within the overall budget. The lead PI(s) must work with their Administrative Officer (AO) to ensure an accurate budget and funding management responsibilities before submission. Proposals utilizing USGS contracting staff must include in the Budget Justification statement a confirmation from the Contracting Officer's Representative (COR) that there is an available contract to complete the project. All CDI funds will be transferred to a USGS cost center through a change of allocation. The USGS cost center may then provide sub-awards to other collaborating organizations/cost centers.

Full Proposals must submit a *Budget Form* and include the Budget Justification statement to explain project costs in the following categories:

- *Personnel (Salaries including benefits)*: Include estimates (by hours) and rate of compensation proposed for each named individual or category (e.g., graduate student). Ensure that the identified personnel and their affiliations are clearly listed. Projects with contractor support must describe how the contract work will be managed and documented to ensure that products are USGS property.
 - *Travel Expenses*: Specify travel requirements for project meetings, and/or conference attendance. Itemize estimated travel costs to show the number of trips required, destinations, the number of travelers and per diem rates, cost of transportation (e.g. vehicle rental), and miscellaneous expenses for each trip. Travel must be included for at least one representative to attend a CDI-sponsored event. Travel cannot include data field collection.
 - *Other Direct Costs*: Itemize any proposed permanent equipment acquisitions (\$5,000 or more) and show each estimated cost. Explain costs including publication costs, office supplies, training, etc.
 - *Indirect Costs (Overhead)*: Provide indirect cost rate and amount approved for each institution.
- **Timeline (5%)**

Provide an estimated timeline describing major and minor project phases, milestones, and deadlines as applicable and including any relevant procurement deadlines. Assume that funding will be awarded no sooner than March 31, 2016 and reference specific months or dates within FY16 or relative to *time from date of award* (e.g. 3 weeks after award date). Project work must demonstrate completion by *September 30, 2016*. Recognizing the USGS publication process may take additional time beyond the end of the project, please indicate the anticipated publication date for any USGS publications resulting from the project.

D. Appendices

- Required: CV(s) of Principal Investigator(s) that highlights relevance to the proposed work (max. 2 pages each)
- Optional: CV(s) of other collaborator(s) that highlights relevance to the proposed work (max. 2 pages)

each)

- Optional: Letters of support (max. 1 page each)
- Optional: Any other materials relevant to the explanation of the project

Budget Form

Applicants are required to use the Budget Form Template (in MS Excel format) (<https://my.usgs.gov/confluence/display/cdi/CDI+FY16+RFP+Forms>). Include Personnel, Travel Expenses, and Other Direct Costs, separating the CDI funds from the in-kind match as indicated in the example below:

Budget Category	Federal Funding “Requested”	Matching Funds “Proposed”
1. PERSONNEL (SALARIES including benefits):		
Personnel :		
John Doe, 150 hrs at \$65/hr (PI, Assistant Deputy Director)	\$	\$9,750
Jane Smith, 125 hrs at \$57/hr (GIS Manager, data reviewer)	\$7,125	\$
Student, 100 hrs at \$37/hr (data miner)	\$1,850	\$1,850
Contract Personnel :		
Bob Thomas, XYZ Inc., 125 hrs at \$48/hr (database developer)	\$19,000	\$
Total Salaries:	\$27,975	\$11,600
2. TRAVEL EXPENSES:		
Trip 1 (GIS Conference, 2 days, 4 travelers)		
Per Diem:	\$	\$228
Transportation (Airfare + Mileage/Shuttle):	\$	\$472
Other expenses (e.g. Registration fees):	\$	\$450
Trip 2 (CDI Workshop, 2 days, 2 travelers)		
Per Diem:	\$142	\$
Transportation (Airfare + Mileage/Shuttle):	\$1,000	\$
Other expenses (e.g. Registration fees):	\$386	\$
Total Travel Expenses:	\$1,528	\$1,150
3. OTHER DIRECT COSTS: (itemize)		
Equipment (inc. software, hardware, purchases/rentals):	\$	\$5,000
Publication Costs:	\$5,000	\$
Office supplies:	\$	\$
Training:	\$	\$
Other expenses (specify):	\$	\$
Total Other Direct Costs:	\$5,000	\$5,000
Total Direct Costs:	\$34,503	\$17,750
Indirect Costs (24%):	\$8,281	\$0
GRAND TOTAL:	\$42,784	\$17,750

Data Management Planning Form

Applicants are required to use the Data Management Planning Form Template and submit in a PDF format (<https://my.usgs.gov/confluence/display/cdi/CDI+FY16+RFP+Forms>). As required in the template, include the following categories of information: Digital/Electronic Products, Data/Metadata, and Publications.

Appendix C – Example Cover Sheet

CDI SSF Element(s): Semantics, Data

Title: A Semantic Standard and Tool for Integrating Mineral Deposit Data from Rock, Soil, Water, and Plant Analyses.

Principal Investigator(s):

Graham Granule, USGS Oregon Water Science Center, 2130 SW 5th Ave, Portland, OR 97201. Ph (503) 251-3200, Fax (503) 251-3333, Email ggranule@usgs.gov
Sandy Lake, Utah State University, Mineral Deposits Laboratory, 0900 Old Main Hill Rm 164, Logan, Utah 84322-0900. Ph (435) 797-1189, Fax (435) 797-1100, Email slake@usu.edu, orcid.org/0000-0002-9999-9999

Fiscal Contact:

Sharon Thrift, USGS Oregon Water Science Center, 2130 SW 5th Ave, Portland, OR 97201. Ph (503) 251-3201, Fax (503) 251-3333, Email sthrift@usgs.gov

Collaborators:

Elliott Smith, USGS Oregon Water Science Center, 2130 SW 5th Ave, Portland, OR 97201. Ph (503) 251-3202, Fax (503) 251-3333, Email esmith@usgs.gov
Sherry Thomas, Oregon State University, Department of Lichens, 555 SW Jefferson Way, Corvallis, OR 97331. Ph (541) 737-1002, Fax (541) 737-1000, Email sthomas@oregonstate.edu, orcid.org/0000-0002-9999-9999

Abstract:

[200-250 word block of text here.]

Total funding amount requested: \$25,000

Total in-kind funding: \$15,000

Expected Product(s) Generated:

Mobile application, GIS Data Layers, USGS Open File Report

Appendix D – CDI Science Support Framework (SSF)

The Community for Data Integration (CDI) represents a dynamic aggregation of multiple communities of practice, focused on the advancement of scientific data and information management and integration capabilities across the USGS and external organizations.

Since 2009, CDI has funded a variety of projects that support the overarching goal of data integration. USGS and other researchers conduct monitoring, assessment, and research activities that generate data assets. Through the application of business, computational, and analytic processes and technologies, these data assets are converted into information that contributes to our understanding of the Earth’s physical and biological systems. This is the context within which data management and integration occur and where the CDI operates (Fig. 1).

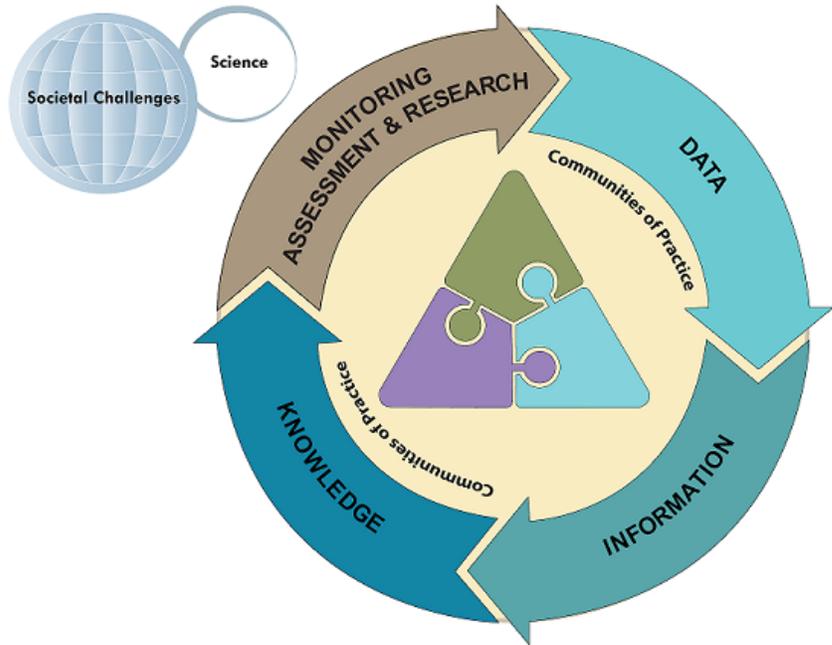


Figure 1: Overview of CDI Operational Context

 <p>Communities of Practice</p>	<p>Communities of practice include scientists, the CDI as a whole, CDI Working Groups, external partners, and the human network of scientific domain collaborators.</p>
 <p>Computational Tools & Services</p>	<p>Computational tools and services include applications, Web services, data discovery tools, models, semantic services and tools, infrastructure, data brokers, and visualization tools.</p>
 <p>Management, Policy & Standards</p>	<p>Management, policy, and standards include data stewardship, the implementation of the Science Data Lifecycle, knowledge management, data standards, governance, and policy.</p>
 <p>Data & Information Assets</p>	<p>Data and information assets include persistent archives, data registries, catalogs, data, metadata, derived information products, knowledge bases, and vocabularies/ontologies.</p>

The CDI SSF (Fig. 2) provides a conceptual architecture that illustrates how the CDI contributes to Bureau-level data integration efforts; and defines how current and future CDI projects fit within the *framework*.

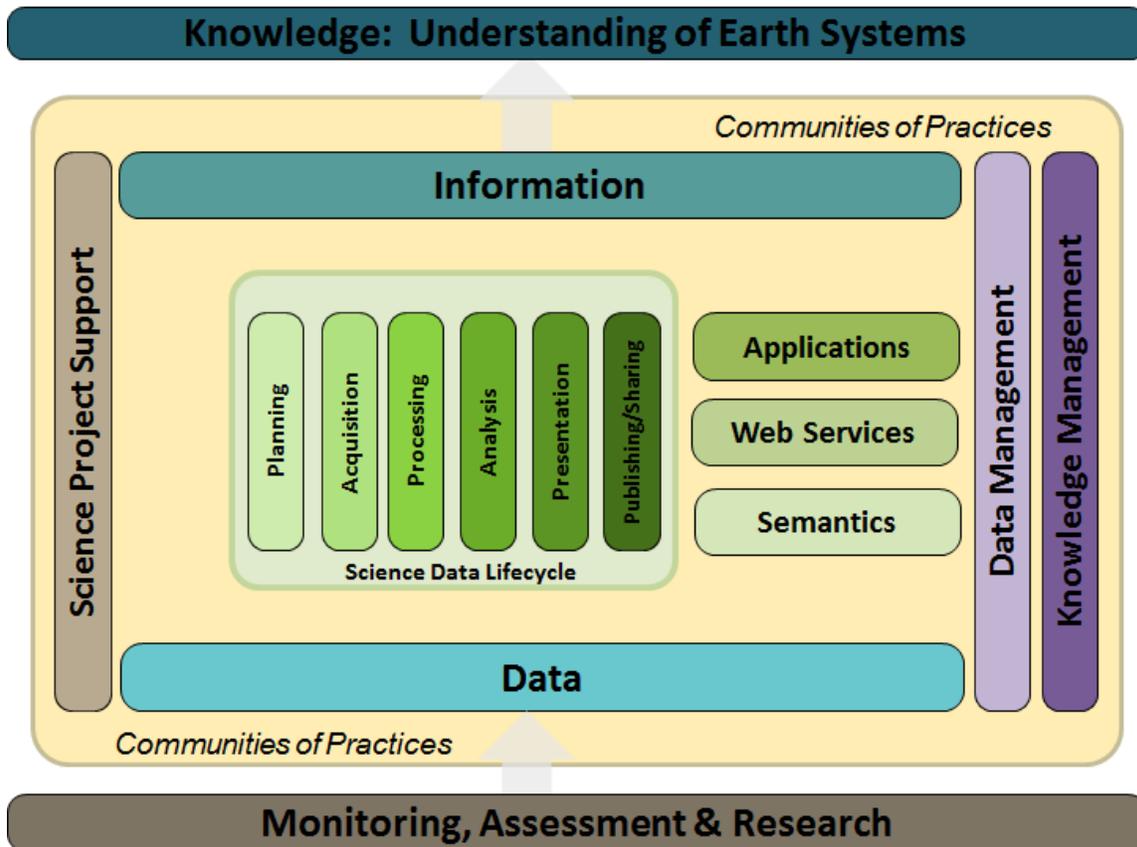


Figure 2: CDI Science Support

USGS Data Assets Flow through the CDI Science Support Framework.

USGS data assets flow vertically through the SSF from a base of monitoring, assessment, and research through the Science Data Lifecycle, applications, Web services, and semantics. The assets are transformed into information products that benefit from data and knowledge management and also increase knowledge and understanding of the Earth's physical and biological systems. Data assets flow horizontally through the SSF from science

project support to data and knowledge management.

The horizontal elements in the SSF represent the “what” of the CDI: products and tools that contribute to the advancement of scientific data and lead to the development of knowledge and understanding of the Earth’s systems.

The vertical elements in the SSF represent the “how” of the CDI: the processes, the implementation of standards and best practices, and the interactions among people, data,

and technology used to achieve data integration.

Individual Framework element descriptions:

Science Inputs (brown elements)

Monitoring, Assessment, & Research: USGS scientists conduct monitoring, assessment, and research that generate data assets. Through the application of business, computational, and analytical processes and technologies, these assets are

converted into information products that can be shared with other researchers, stakeholders, and citizens to increase our knowledge and understanding of the Earth's physical and biological systems.

Science Project Support:

Successful science projects encompass a range of activities represented in the Data Lifecycle. At each step in the cycle, researchers and data stewards rely on an array of sophisticated tools and services for data, information and knowledge discovery, acquisition, integration, management, and sharing.

Communities of Practice (tan element)

Communities of practice are the foundation for CDI and all its products – the communities of people working towards the goal of advancing scientific data and information management and data integration across the USGS.

Data & Information Assets (blue elements)

USGS assets include **Data** (e.g., raw data, databases, and linked open data (RDF¹)); **Information** or derived/interpreted information products (e.g., published or shared maps, reports, datasets); and **Knowledge** of all types and in all forms — recorded, organized, and preserved in the form of artifacts. Knowledge can be improved, shared across groups,

organizations, and domains, and reused to support learning and research.

Computational Tools & Services (green elements)

Science Data Lifecycle include tools and services that move data through the lifecycle, human and machine interactions, and interactions with data through technology.

Detailed descriptions of the Science Data Lifecycle:

- **Planning** – A documented sequence of intended actions to identify and secure resources and gather, maintain, secure, and utilize data assets.
- **Acquisition** – The series of actions for collecting or adding to data assets.
- **Processing** – A series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.
- **Analysis** – A series of actions and methods performed on data that help describe facts, detect patterns, develop explanations, and test hypotheses.
- **Preservation** – Actions and procedures to keep data for some period of time; to set data aside for future use.
- **Publishing/Sharing** – To prepare and issue, or to

disseminate data or information products.

Semantics convert raw data into data that can be interpreted by machines: Machine Readable Metadata, Semantic Mediation for Data Integration & Discovery, Ontologies/Vocabularies, and World Wide Web Consortium Standards.

Web Services include machine to machine data exchange, SOAP,² REST,³ SPARQL⁴ EndPoints, and other protocols and services.

Applications include human readable data services and user interfaces to data driven applications.

Management, Policy, & Standards (purple elements)

Data Management includes data and metadata standards and policies and occurs in all phases of the Data Lifecycle from scientific research to finished information products.

Knowledge Management involves the creation, standardized documentation, and organization of knowledge using tools such as SKOS⁵ Vocabularies and information modeling, resulting in the formation of knowledge bases.

¹ Resource Description Framework
² Simple Object Access Protocol
³ REpresentational State Transfer
⁴ SPARQL Protocol and RDF Query Language
⁵ Simple Knowledge Organization System

Appendix F – CDI Coordinators

CDI Executive Sponsors

Kevin Gallagher, Associate Director, USGS Core Science Systems

Cheryl Morris Director, USGS Core Science Analytics, Synthesis and Libraries Program

CDI Facilitators

Madison Langseth (mlangseth@usgs.gov)

CDI Citizen Science Working

Group <https://my.usgs.gov/confluence/display/cdi/Citizen+Science+Working+Group>

David Govoni – Lead

CDI Data Management Working Group

<https://my.usgs.gov/confluence/display/cdi/Data+Management+Working+Group>

Heather Henkel – co Lead

Vivian Hutchison – co Lead

John Faundeen

Steve Tessler

CDI Earth-Science Themes Working Group

<https://my.usgs.gov/confluence/display/cdi/Earth-Science+Themes+Working+Group>

Roland Viger – Lead

CDI Semantic Technologies Working Group

<https://my.usgs.gov/confluence/display/cdi/Semantic+Web+Working+Group>

Janice Gordon – co Lead

Fran Lightsom – co Lead

CDI Tech Stack Working Group

<https://my.usgs.gov/confluence/display/cdi/Technology+Stack+Working+Group>

Daniella Birch - Lead

CDI Connected Devices Working Group

<https://my.usgs.gov/confluence/display/cdi/Connected+Devices+Working+Group>

Lance Everette – co Lead

Tim Kern – co Lead