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

For Fiscal Year 2021

Issue Date: September 9, 2020

Table of Contents

Overview

This document describes the CDI Request for Proposals (RFP) process for fiscal year 2021 (FY21). 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. Additional resources for the RFP are available on the proposals section of the CDI wiki: *https://my.usgs.gov/confluence/x/j4xbK*.

New in FY21

- **Theme for the FY21 Request for Proposals:** This year, the CDI executive sponsors are encouraging proposals that address increased capacity for Integrated Predictive Science related to fire science and coastal resilience. See more detail in the section *FY21 Topical Emphasis.*
- **Collaboration Forum:** A channel is set up on the CDI Microsoft Teams space to discuss proposal ideas and look for collaborations and advice before submission.
- **Travel guidance:** We require the budget to include travel funds to attend the 2021 CDI Workshop, but full proposals must include a backup plan for all travel funds if travel is restricted in 2021 due to COVID-19.
- **Microsoft Word Submission Template:** Submissions must use the .docx template available under the Downloads section of the 2021 Proposals Wiki page.
- **Co-PIs and Collaborators** identified on the cover page are limited to four.
- **Full proposal budgets** must be reviewed and approved by an appropriate Center staff member.

Eligibility

Personnel from 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 $400,000 – $500,000 each year and it is anticipated that FY21 will be similar to previous years. Applicants can request funding up to $50,000 maximum.

Distribution of Funds

In FY21, CDI will distribute funds to the lead USGS Science Center only, and the lead Science Center will be in charge of further distributions to other Centers or external partners. This practice was initiated in FY18 to expedite the distribution of funds.

Estimated Schedule for Submission, Review, and Awards

This is a tentative schedule, please check the schedule posted at the *2021 Proposals page* for any changes.

RFP Information Session……..……..September 23, 2020, 3 pm ET
Submission Deadline for Statement of Interest (SOI)…...……..…….....October 16, 2020, 5 pm ET
Comment Period Begins ……...……..……...October 20, 2020
SOI Lightning Presentation Session…….……October 22, 2020, 2 pm ET
SOI Voting Opens……...…..…….……..…….....October 28, 2020
SOI Voting Closing……...……..…….....November 13, 2020, 11:59 pm ET
Applicants Notified and Full Proposals Requested……..…….……...Early December 2020
Invited Full Proposals Due…….…….....January 28, 2021
Funded Projects Announced…….……March 2021
Awarded Funds must be Spent……..…… ……...September 30, 2021

Application Process

1. Submit Statement of Interest and Present Lightning Talk

Submit a statement of interest using the new SOI .docx template, available on the FY21 Proposals page under the Downloads section. SOI guidance can be found in *Appendix A – Statement of Interest (SOI) Guidance*. In addition, submitters will present their proposals to the community in a one-minute, one-slide lightning-talk format on October 22, 2020, at 2 pm ET. The purpose of the lightning talk session is to help submitters receive suggestions, gain support from the community, and help the community efficiently learn about the breadth of the SOIs. To aid the development of statements of interest, submitters may use the *FY21 RFP Microsoft Teams channel* to seek partners or other feedback.

2. Evaluate Statements of Interest

SOIs will be reviewed by the CDI Community. The SOIs will be considered according to *the Evaluation Criteria for the Statements of Interest and Full Proposals*.

3. Invitation to Submit Full Proposal

Based on the SOI evaluation, selected SOI applicants will be invited to submit a full proposal. Note that the CDI Executive Sponsors may also advance SOIs that support current USGS priorities or RFP themes.

4. Full Proposal Review Process

Full Proposals will be evaluated according to the *Evaluation Criteria for the Statements of Interest and Full Proposals.* Proposals will be reviewed by a panel consisting of a professional peer group that is knowledgeable in USGS-related scientific disciplines, data management, information technology, and other relevant topics in the context of the CDI. Recommendations by the Review Panel will be presented to the CDI Executive Sponsors for final selection.

Project Reporting

A representative from each CDI funded project will be required to attend the 2021 CDI Workshop (which may be virtual in 2021) and to provide a separate informal mid-year briefing to the CDI facilitators to communicate the status of the project. Projects must also contribute to the CDI Project Report, which will be compiled in early 2022. At that time, project leads must provide a brief written report describing the project accomplishments, benefits, and deliverables with links to products or publications.

Examples of previous project reports can be seen at

- FY2018 CDI Funded Project Report
- FY2017 CDI Funded Project Report

Description of the Request for Proposals

The CDI builds and shares knowledge about topics such as data integration, data 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

FY21 Topical Emphasis: Fire Science and Coastal Resilience

Each year, the CDI accepts statements of interest on any topic that follow the CDI guiding principles, noting that the CDI Executive Sponsors may select proposals that support the current Bureau priorities. This year, we continue to support a focus outlined in the *USGS Director's Science Planning Strategy* to "[embrace] an integrated and predictive capability that accounts for complex natural system interactions; anticipates the likelihood and consequences of evolving threats and hazards; and helps guide resilient adaptation and mitigation efforts."

In FY21, the CDI executive sponsors are encouraging proposals that address increased capacity for Integrated Predictive Science related to fire science and coastal resilience and protection.

Examples of relevant projects would be those that

- advance the capability to link existing data and new data rapidly collected by advanced new sensor technologies;
- integrate data from across disciplines;
- develop and apply methods of co-producing decision analysis capabilities by including decision making practitioners on project teams;
- integrate models to evaluate post-fire or post-coastal change conditions and highlight effective recovery options.

The CDI will accept submissions on any topic that follow the CDI guiding principles. Of total funding, CDI hopes to award approximately 50% of funding to qualified projects that are associated with the stated themes, depending on the number of relevant submissions received. Project teams in the focus areas may be asked to work with the CDI facilitators to leverage related projects' work and progress.

Head to the *FY21 RFP Collaboration channel on Microsoft Teams* for more discussion of this year's topical emphasis. More information about the CDI Microsoft Team can be found here.

Unique Aspects of the CDI Request for Proposals Process

The CDI proposals process has several aspects that may be unfamiliar for first-time participants. The selection process incorporates community-involvement and multiple ways for submitters to communicate about their statements of interest.

Gaining Community Support for Statements of Interest

All CDI members have the ability to comment and vote on the submitted Statements of Interest. Therefore, it is in the submitter's interest to promote their idea to the community, and to communicate the value added to the CDI by the proposed activities. This process may be puzzling to participants who are not accustomed to any promotion of proposals beyond submitting files to a review panel. However, the CDI views this process as an opportunity to practice and improve plain-language communication and the ability to articulate the value of proposed activities. The community comment period also harnesses the expertise of the CDI community to make suggestions and improve on project ideas. These unique aspects help the CDI toward its goal of supporting the most useful and innovative ideas. In Phase 2 of the proposals process, after the community has given its input, a formal review panel evaluates the full proposals, similar to more typical proposal processes.

Statement of Interest Lightning Presentation

In past years, the community has expressed that it is difficult to absorb the large number of ideas that are submitted in Phase 1 of the proposals process. Since FY 2018, to assist in acquainting the CDI community with the submitted statements of interest, we host an **SOI Lightning Presentation Session** at the beginning of the commenting and voting period. Each submitter will have the chance to present their SOI idea in an online one-minute lightning presentation. To simplify the process, slides follow a strict template with project title, Lead PI name and contact information, and a single image. Submitters have the option of pre-recording a one-minute presentation as an audio file, or they can present live.

CDI Science Support Framework

Project proposals must 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 C – 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 integrating datasets, or best practices for alignment or assimilation of data at different scales particularly with respect to Bureau priorities. (e.g., *Integrating Disparate Spatial Datasets from Local to National Scale for Open-Access Web-Based Visualization and Analysis: A Case Study Compiling U.S. Landslide Inventories* and *CDI Risk Map*)
- **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 (e.g., *MetadataWizard* and *sbtools: An R package for ScienceBase*)
- **Development of standards or best practices for data management** through community consensus building, such as convening a workshop and writing a white paper (e.g., *Content Specifications to Enable USGS Transition to ISO Metadata Standard*)

- **Development of a general ontology or tools for tagging data** in support of standards and environments to facilitate discovery, understanding, and integration (e.g., *Use of Controlled Vocabularies in USGS Information Applications: Requirements Analysis for Automated Processes and Services (Bureau-wide Application)*)
- **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 (e.g., *Evaluating a new open-source, standards-based framework for web portal development in the geosciences*)
- **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 (e.g., *Knowledge Extraction Algorithms (KEA): Turning Literature Into Data* and *ICE! Ice Jam Hazard Mobile-Friendly Website*)
- **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 (e.g., *Hunting Invasive Species with HTCondor: High Throughput Computing for Big Data and Next Generation Sequencing* and *Mapping Land-Use, Hazard Vulnerability and Habitat Suitability Using Deep Neural Networks*)
- **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. (e.g., *Mapping Land-Use, Hazard Vulnerability and Habitat Suitability Using Deep Neural Networks*; *Building a Roadmap for Making Data FAIR in the U.S. Geological Survey* and *Data Management Training Clearinghouse*)

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 that would normally be funded by individual Program Areas.
- Projects that would normally be funded by other proposal processes such as the *John Wesley Powell Center for Analysis and Synthesis.*

Examples of successful past CDI statements of interest and full proposals:
https://my.usgs.gov/confluence/x/WgdEJ

All past CDI Projects:
https://my.usgs.gov/confluence/x/SgtAK

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 on submitting 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. Evaluation will consider 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.

Project Experience and Collaboration (25%)

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

Sustainability, Outreach, and Communication (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, as well as the plan for communicating the value of the products during and after the project period. All products resulting from CDI projects must comply with the *U.S. Geological Survey Manual Chapters* on data management (SM 502.6-502.9). 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 be uploaded to an appropriate code repository at the close of the funding period and follow the latest relevant USGS Instructional Memoranda (*IM OSQI 2019-01, Review and Approval of Software for Release*).

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. Travel cannot include field data collection. Travel funds of $2,000 must be included for at least one representative to attend the 2021 CDI Workshop in Shepherdstown, WV. In the full proposal phase, a secondary plan for spending travel funds should be specified (for example, additional salary or training) due to the possible travel restrictions regarding COVID-19. Evaluation will 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 and the feasibility of the proposed workload given the short project duration. Although notification of award may come earlier, assume that funding will be awarded no sooner than May 31, 2021 and reference specific months or dates within FY21 or relative to time from date of award (e.g., 3 weeks after award date). **The timeline must demonstrate reasonable completion and complete use of funds by September 30, 2021.** Recognizing the USGS publication process may take additional time beyond the end of the project, indicate the anticipated publication date for any USGS publications resulting from the project.

Appendix A – Statement of Interest (SOI) Guidance

1) Prepare a Statement of Interest document

Use the template on *the 2021 Proposals page Downloads section*.

For the fillable form template to work, use the desktop app for Microsoft Word.

The following fields are required:

- **Title:** Include a descriptive title of the proposed project (140 character limit)

- **Principal Investigator (PI) Name:** List the lead USGS Principal Investigator (First Name Last Name)
- **PI Email address**
- **PI Organization:** Sub-Unit within USGS. Example: USGS Woods Hole Coastal and Marine Science Center
- **PI Mission Area:** Enter "Other" if Mission Area is not applicable.
- **PI Region:** Enter "Other" if Region is not applicable
- **PI City and State:** Example: Denver, CO

- **Project Description:** This short (1-2 sentence) description will be used in the voting system. The text should transmit the essence of the benefits of the project. (300 character limit)
- **List of anticipated deliverables from the project:** Include a list of the types of products that will be generated as a result of the project using the vocabulary: data release, mobile application, presentation, software, source code, web application, web link, web service (300 character limit)
- **CDI Science Support Framework Element(s):** Indicate up to three element(s) of the CDI SSF that the proposal covers (See *Appendix C - CDI Science Support Framework (SSF)*).

- **Collaborator Information:** Fill out the table for your top four co-PIs or collaborators.

- **Project Narrative:** Use headers to identify each section and describe the evaluation criteria: Scope; Technical Approach; Project Experience and Collaboration; Sustainability, Outreach, and Communication; Budget Justification; Timeline. **In the Scope section, include a concise statement of the alignment of the project to the FY21 RFP topical theme if relevant**. (3000 character limit)

- **Estimated Budget:** Enter estimated values for the categories Personnel, Travel Expenses, and Other Direct Costs. Enter the total direct costs and the indirect costs due to relevant overhead. Enter the grand total. In Phase 1, it is suggested, but not required, that your budget be reviewed by a Center budget analyst, so that you have an accurate picture of how far the available funds will go.

2) Submit Your Statement of Interest

Submit your file as a .docx as an email attachment To: gs_cdi@usgs.gov, CC: lhsu@usgs.gov. Subject: FY21 CDI Statement of Interest Submission before the due date of October 16, 2020, 5 pm Eastern Time.

Appendix B – Invited Full Proposal Guidance

Proposals must be prepared using the Full Proposal template, which will become available after we evaluate the use of the Statement of Interest template. The template will include sections for the proposal narrative (4-7 pages), appendices (CVs and letters of support), budget form, and data management planning form.

Appendix C – 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).

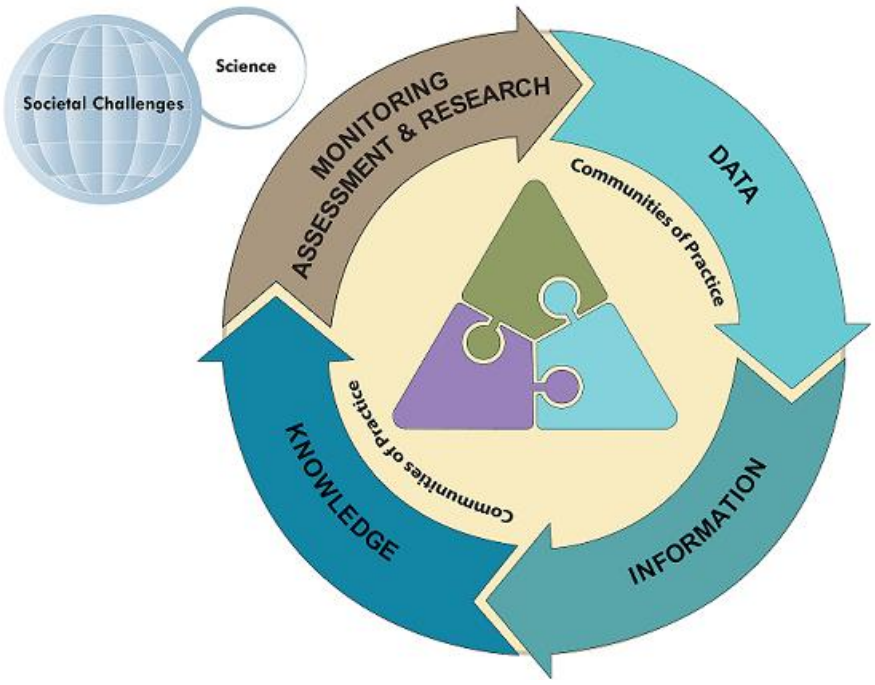

Figure 1: Overview of CDI Operational Context

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

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.

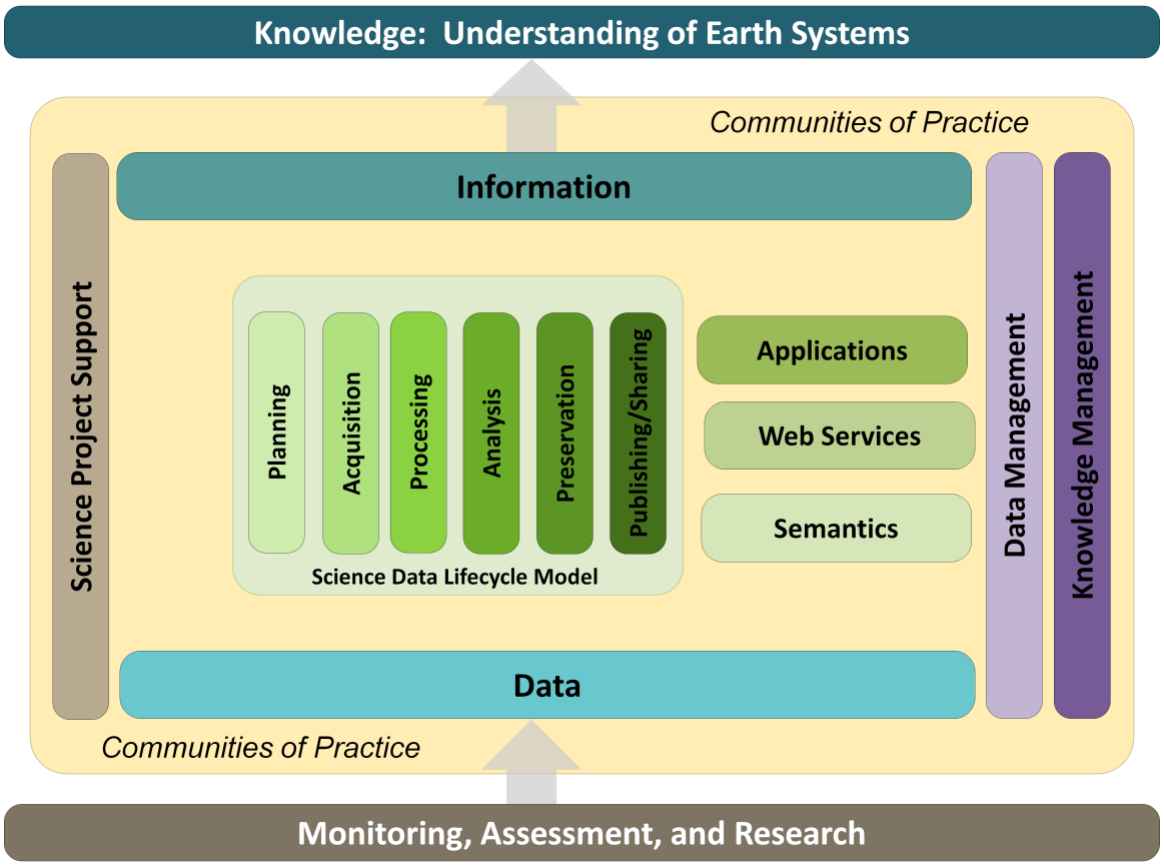

Figure 2: CDI Science Support Framework

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 (RDF[1])); **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] SPARQL[4] 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 SKOS[5] 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 D – CDI Sponsors and Staff

We encourage proposers to get in touch with relevant CDI contacts to discuss their proposals.

CDI Executive Sponsors
Kevin Gallagher, Associate Director, USGS Core Science Systems
Tim Quinn, Chief, Office of Enterprise Information, USGS
Cheryl Morris, Director, USGS Science, Analytics, and Synthesis Program

CDI Facilitators
Leslie Hsu (lhsu@usgs.gov)
Amanda Liford (aliford@usgs.gov)
Grace Donovan (gdonovan@usgs.gov)

CDI Collaboration Area leads may have specific suggestions or contacts for their focus topic. All CDI Collaboration Area pages can be accessed at *https://my.usgs.gov/confluence/x/yhv1I*.

Appendix E - Additional Instructions for Project Products

Use the specific wording below to acknowledge funding in CDI publications and products:

This work was supported by funding from the USGS Community for Data Integration (CDI).