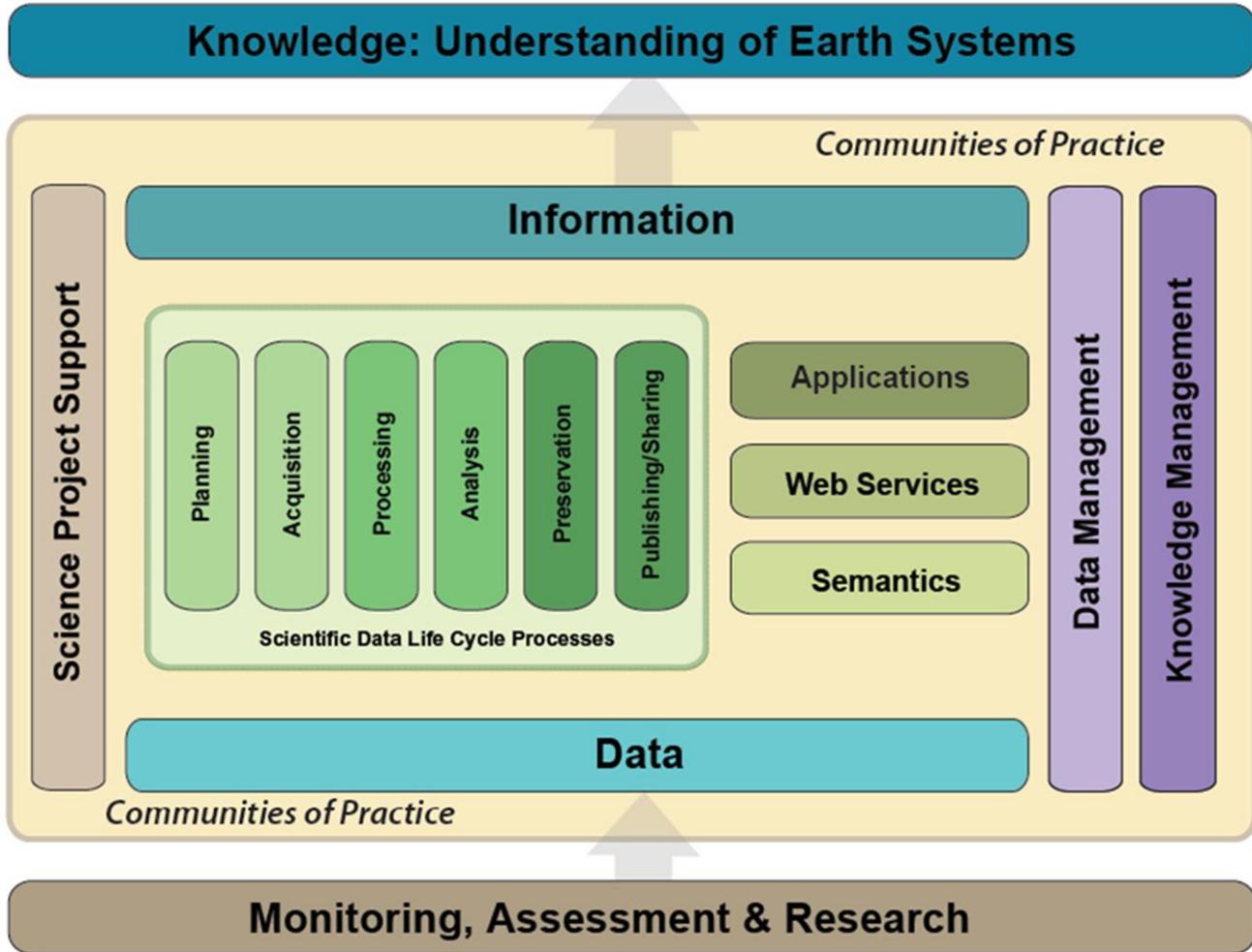


## U.S. Geological Survey Community for Data Integration (CDI) Science Support Framework (SSF)

The CDI SSF 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.



**USGS Data Assets Flow through the CDI Science Support Framework.** USGS data assets flow vertically through the SSF from a base of monitoring, assessment, & research through the Scientific Data Life Cycle (SDLC) processes, 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 and through science projects to data and knowledge management, too.

**The horizontal elements in the SSF represent the “what” of the CDI:** products and tools, the things that contribute to the advancement of scientific data and that 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 generates 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 SDLC. 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<sup>1</sup>)); **Information** or

derived/interpreted information products in the broad sense (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 various artifacts. Knowledge can then be improved; shared across groups, organizations, and domains; and reused to support individual and group learning and research.

**Computational Tools & Services (green elements)**

**Scientific Data Life Cycle processes** include tools and services that move data through the SDLC, human and machine interactions, and interactions with data through technology.

**Detailed descriptions of SDLC Processes:**

- **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; and

- **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,<sup>2</sup> REST,<sup>3</sup> SPARQL<sup>4</sup> 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 Life Cycle from scientific research to finished information products.

**Knowledge Management** involves the creation, standardized documentation, and organization of knowledge using tools such as SKOS<sup>5</sup> Vocabularies and information modeling, resulting in the formation of knowledge bases.

<sup>1</sup> Resource Description Framework  
<sup>2</sup> Simple Object Access Protocol  
<sup>3</sup> REpresentational State Transfer  
<sup>4</sup> SPARQL Protocol and RDF Query Language  
<sup>5</sup> Simple Knowledge Organization Systems