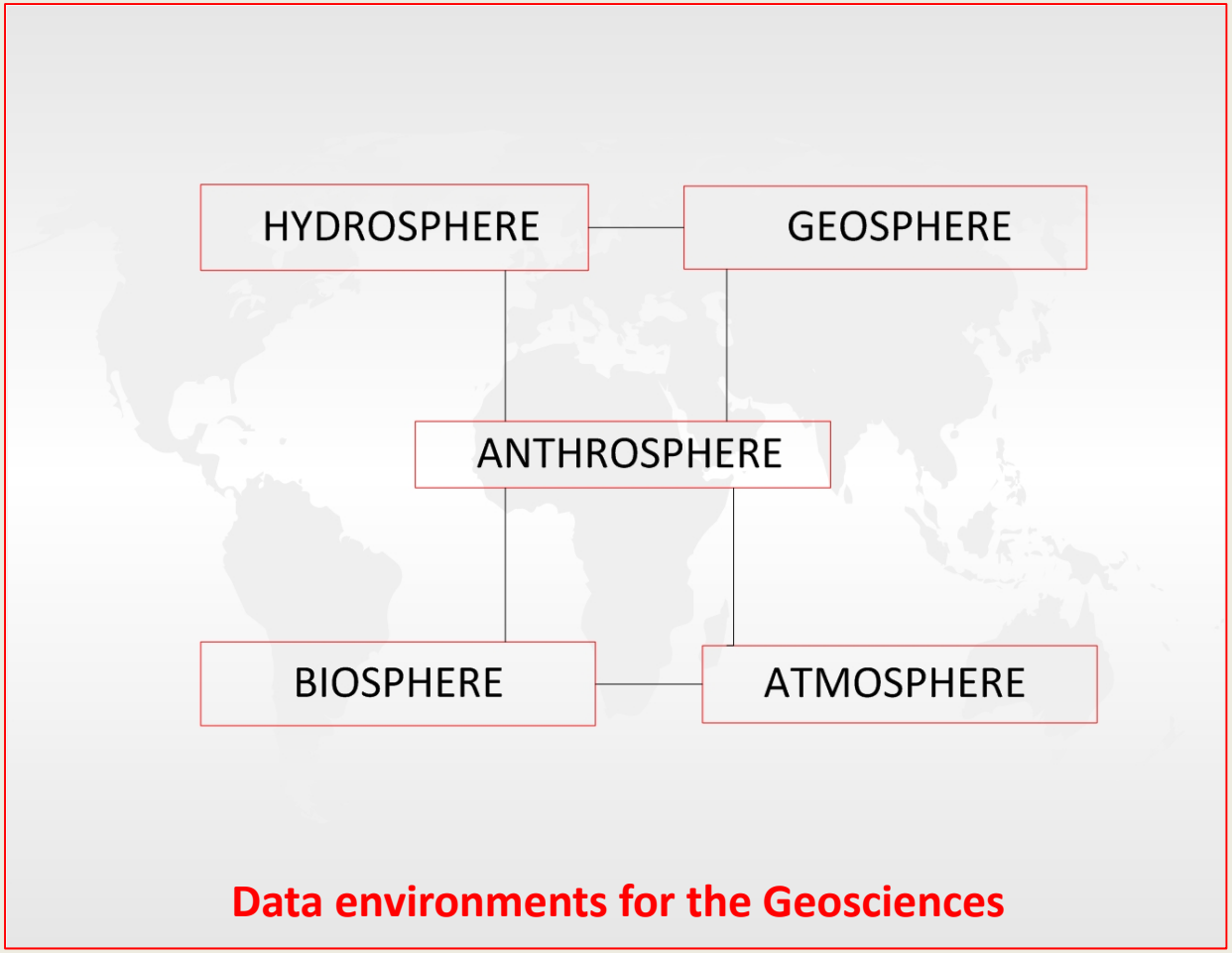
Integrative view of earth as a system of systems: driver for EarthCube



Use cases to reflect data heterogeneity and are globally distributed

Data publication and access

Collection

Diversity

Interoperability

Trust

Security

Education and training

Usability

Preservation and Sustainability

Commercial exploitation

BUILDING BLOCKS OF AN INFRASTRUCTURE TO ADDRESS THESE TASKS THROUGH

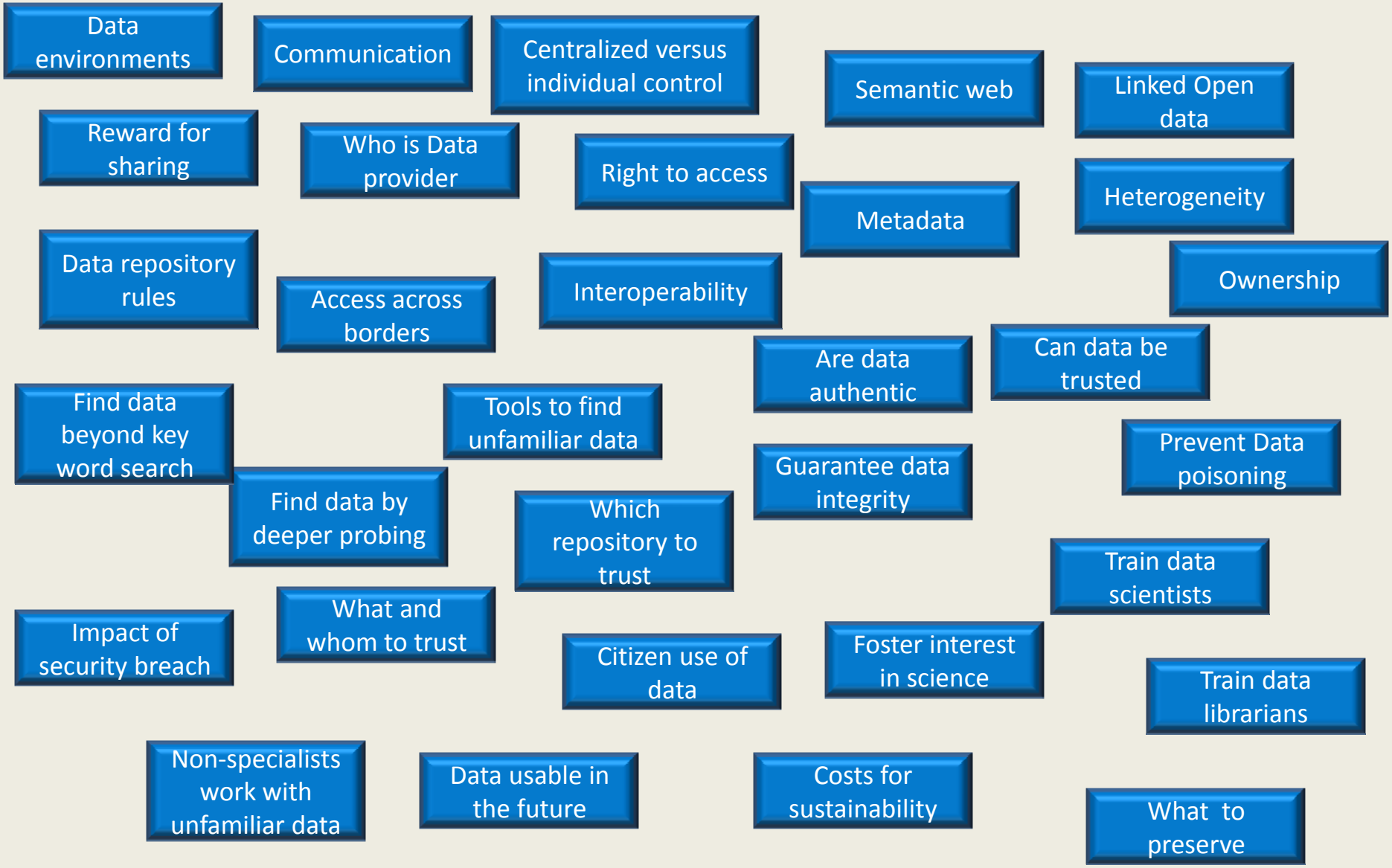
- SOCIAL
- ENGINEERING
- SCIENCE

RESEARCH &DEPLOYMENT

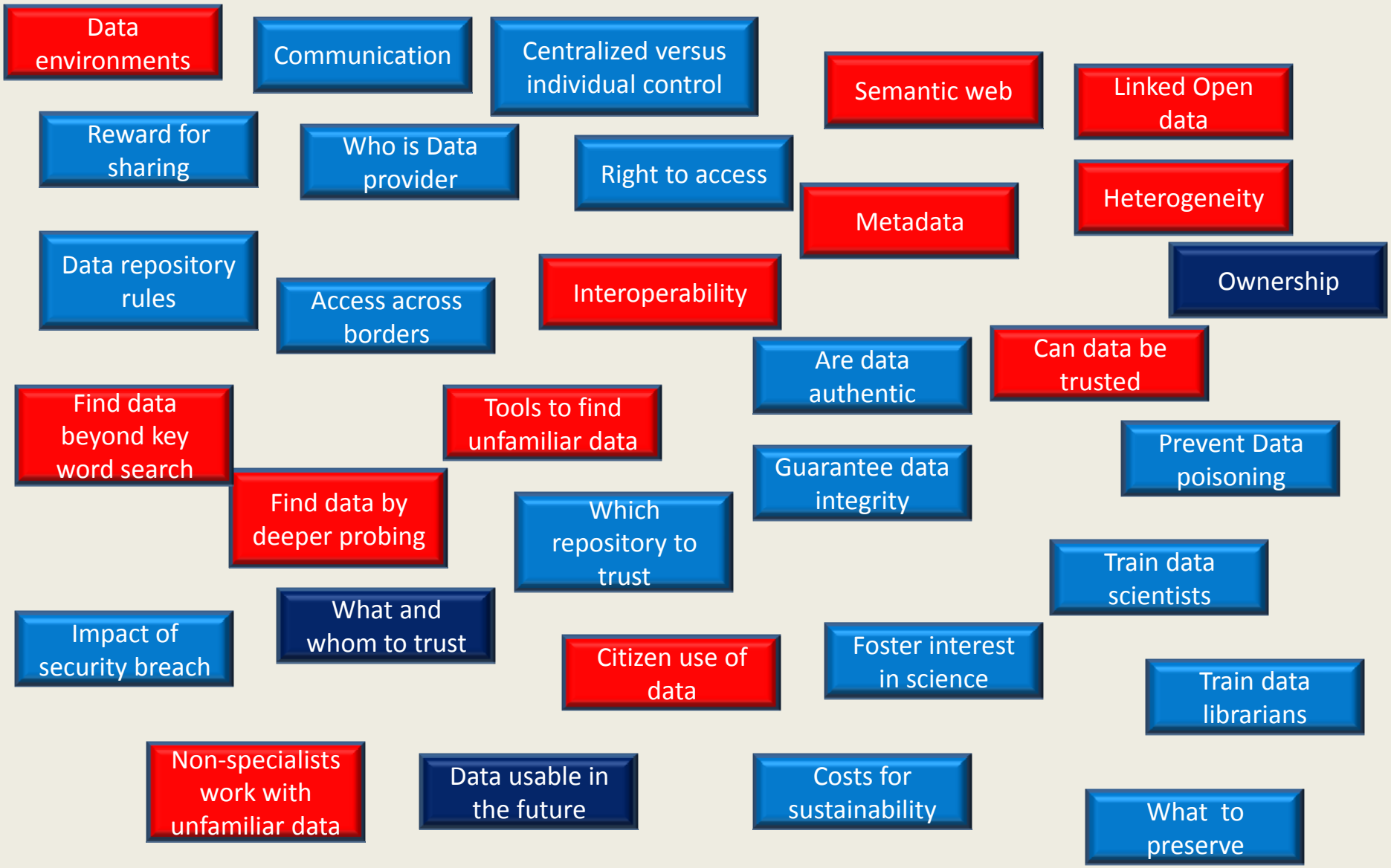
http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf)

- Data publication and access: enable individuals and large data centers to **publish data**, as well as enabling access through semantic technologies. These technologies must be coupled with a **reward system** to encourage data sharing.
- Collection: community endorsed **simple metadata** requirements for data providers to facilitate access to both curated as well non curated data
- Diversity: provide capabilities to access and discover globally distributed **heterogeneous resources**
- Interoperability: provide capabilities (semantic as well as syntactic) to understand the content of databases, as well as support discovery of resources **beyond keyword searches**
- Trust: provide innovative algorithms to **enhance users trust of data** they are unfamiliar with, as well as management of resources to enhance trust in the source of the data
- Security: New technologies to **enhance data integrity**, as well as capabilities to detect data biased by a group or individual
- Education and training: New capabilities to generate **interest in geoscience** through training of data scientists and librarians, as well as providing capabilities to bring information to a class room setting
- Usability: Provide new technologies to **enable non-specialists to use data**, as well as scientists to work with unfamiliar data
- Preservation and Sustainability: Develop capabilities to **preserve current data**, as well as legacy data
- Commercial exploitation: Utilize commercial developments to **enable revenue generation** for sustainability of resources

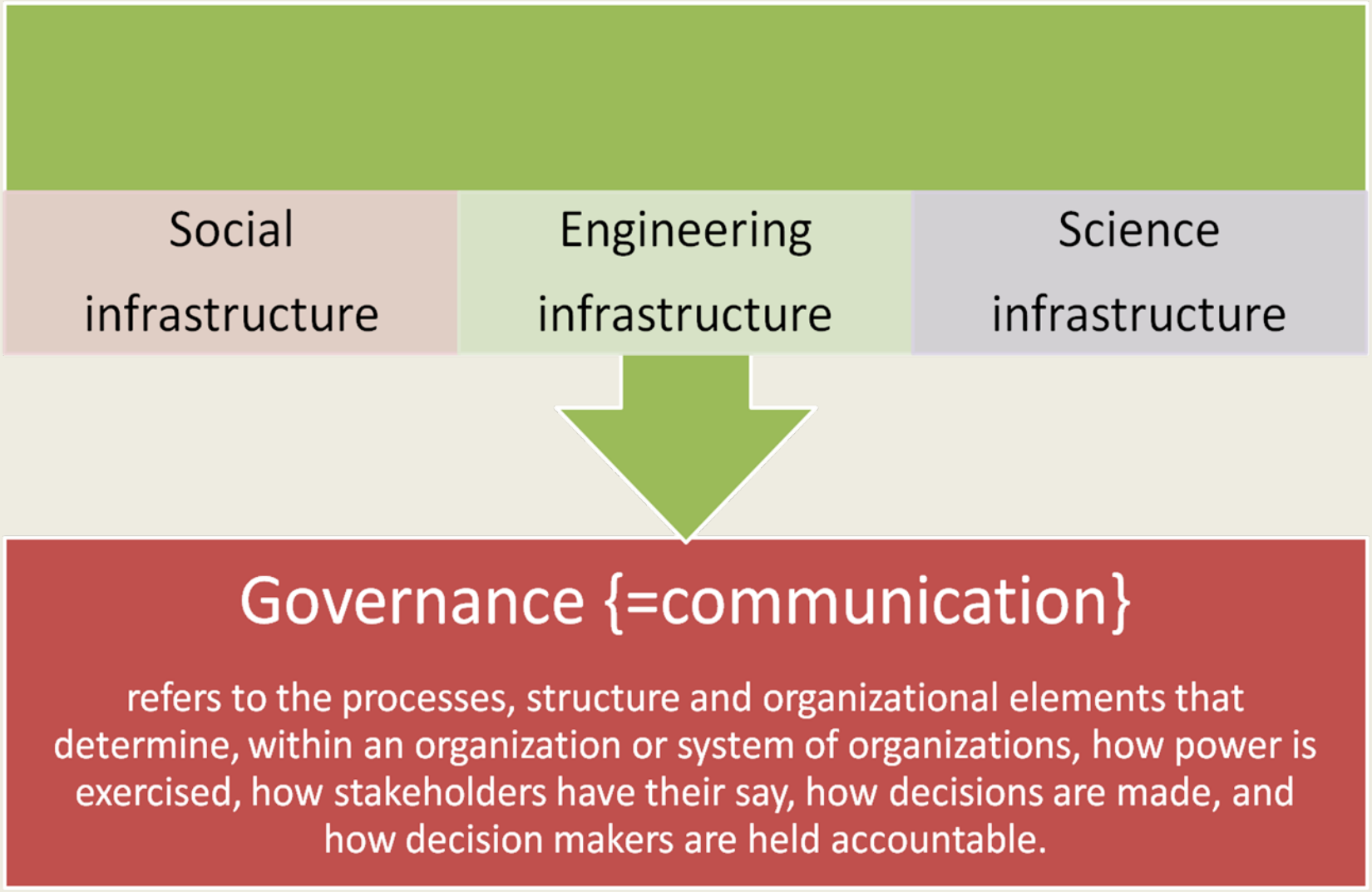
Challenges that require a systems approach to building an infrastructure: Social, Scientific and Engineering components



Semantic Capabilities



Three recognized infrastructures within the EarthCube enterprise are required to meet the staged development of building blocks.



SOCIAL INFRASTRUCTURE
Communication, Governance, Education, Trust

- Communication: barriers between contrasting disciplines
- Governance- centralized or distributed
- Backgrounds and data-sharing cultures in the scientific community
- Data producers be rewarded for publishing data
- Learn from the wisdom of crowds about what and whom to trust
- Enable citizens to utilize these benefits for sensible investigations
- Foster the training of more data scientists and data librarians

ENGINEERING INFRASTRUCTURE

- Layered architecture: interoperability mechanisms between the community resources and the research environment
- Brokering: connecting disparate systems
- Workflows to manage complex computations and enable selection of models appropriate for their data, configure them with appropriate parameters, and execute them efficiently
- Interop: ability to assess whether a given dataset is “fit for use” in a particular scientific context
- Web Services
- Data Grid
- Linked Open Data
- Web; Semantic web
- Ontology

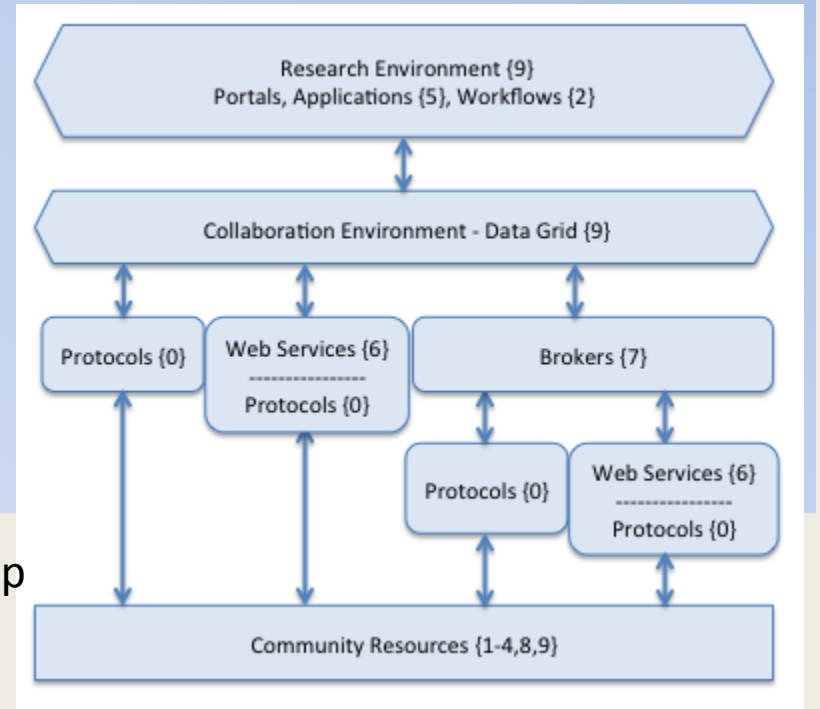
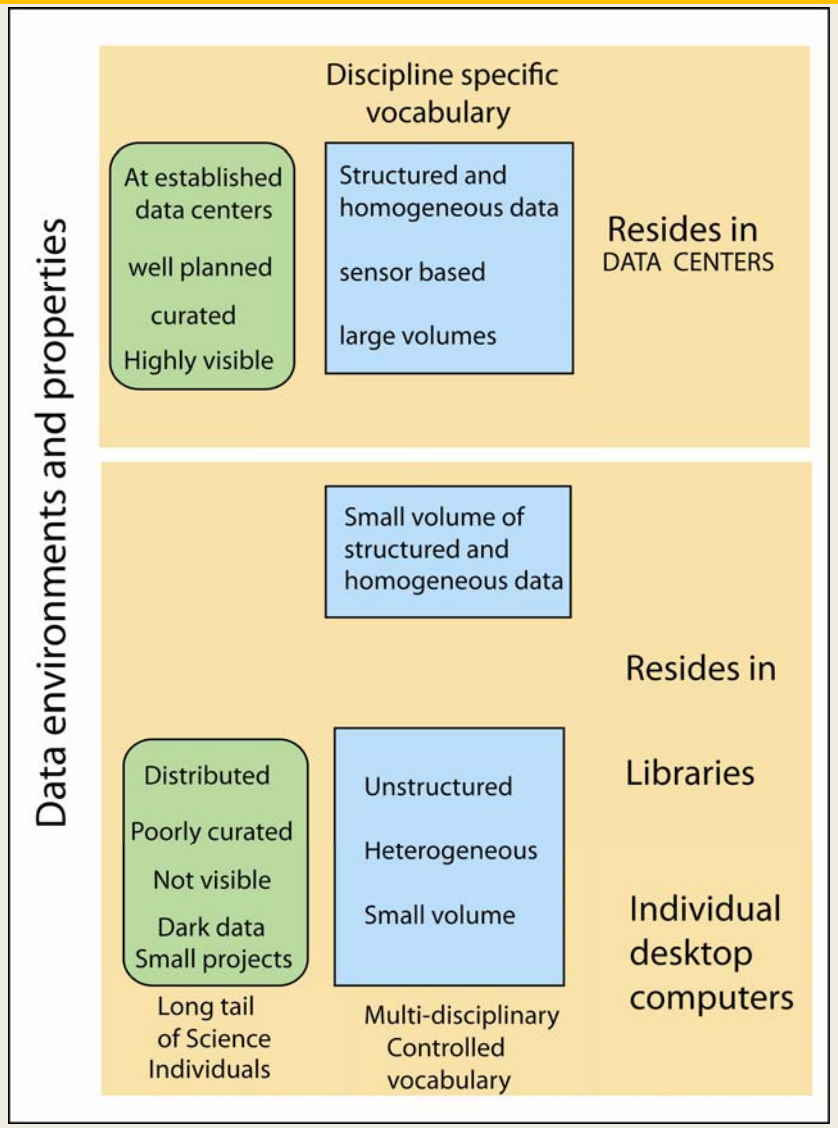
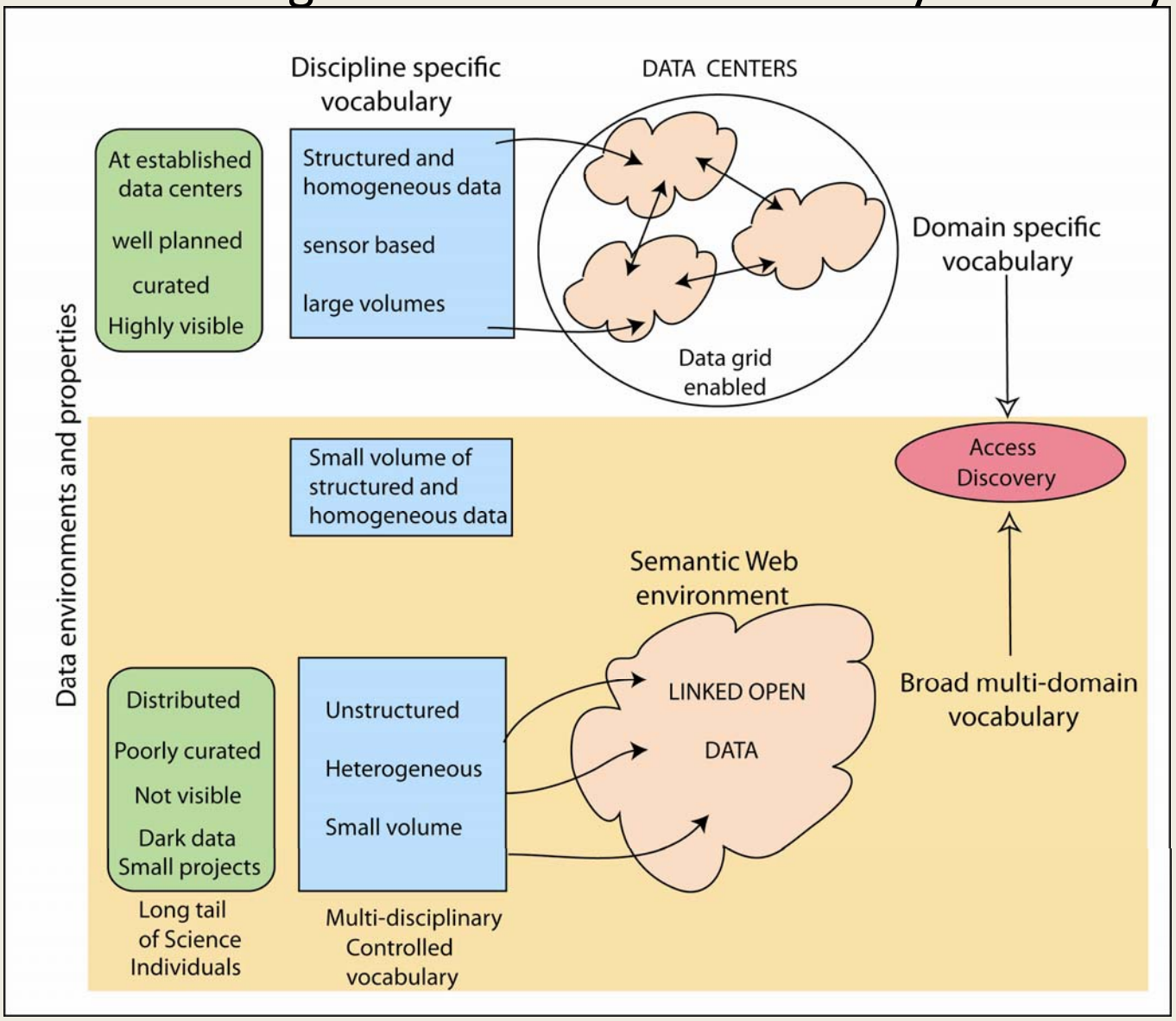


Figure from Layered Architecture Roadmap

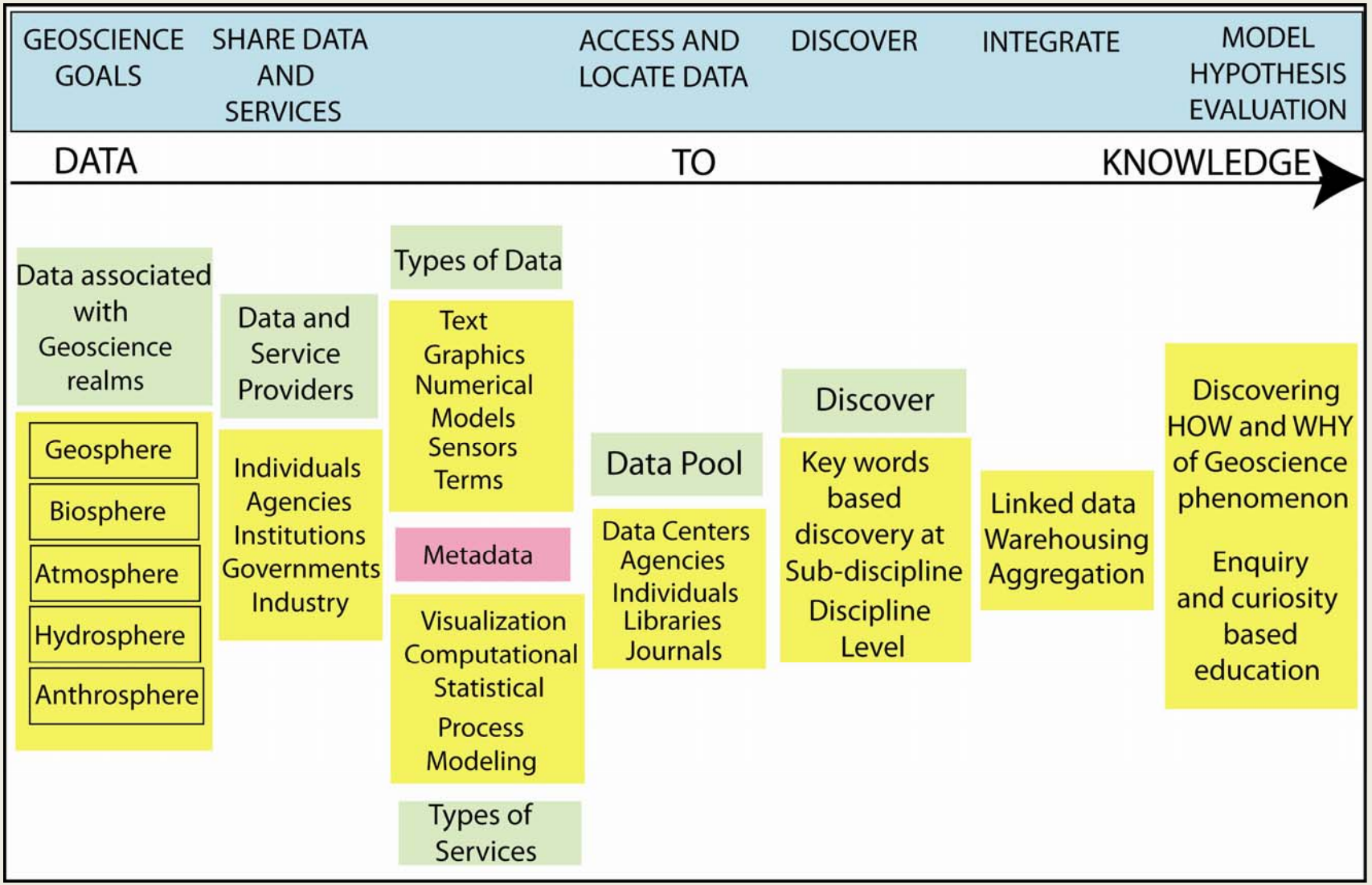
SCIENCE INFRASTRUCTURE
Semantics, Semantic Web, Vocabulary, Metadata, Heterogeneity, Quality, Re-use, Environment, Sharing of data



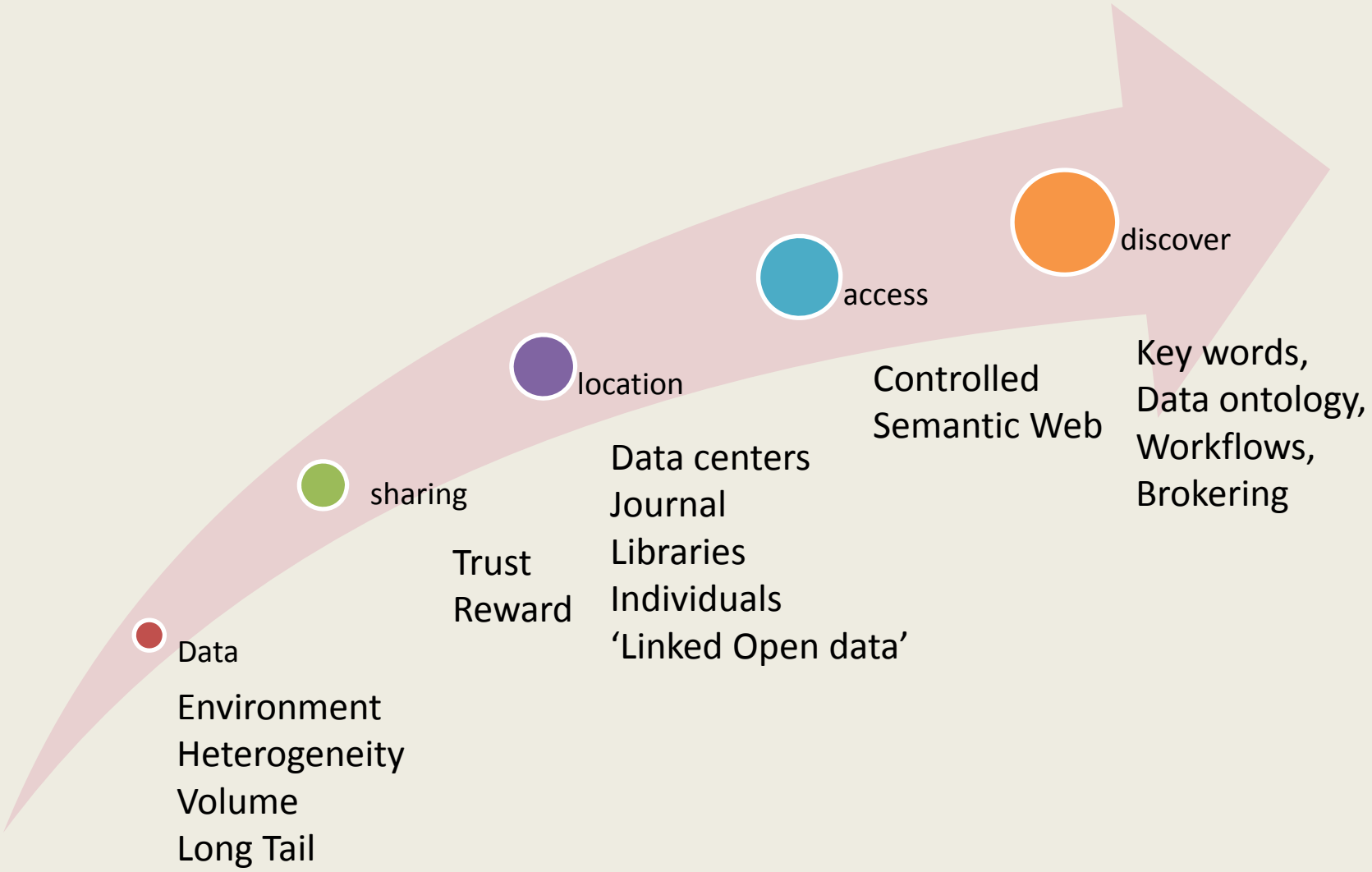
Deploying three infrastructures (social, engineering, science)together to provide an integrative view of earth as a system of systems

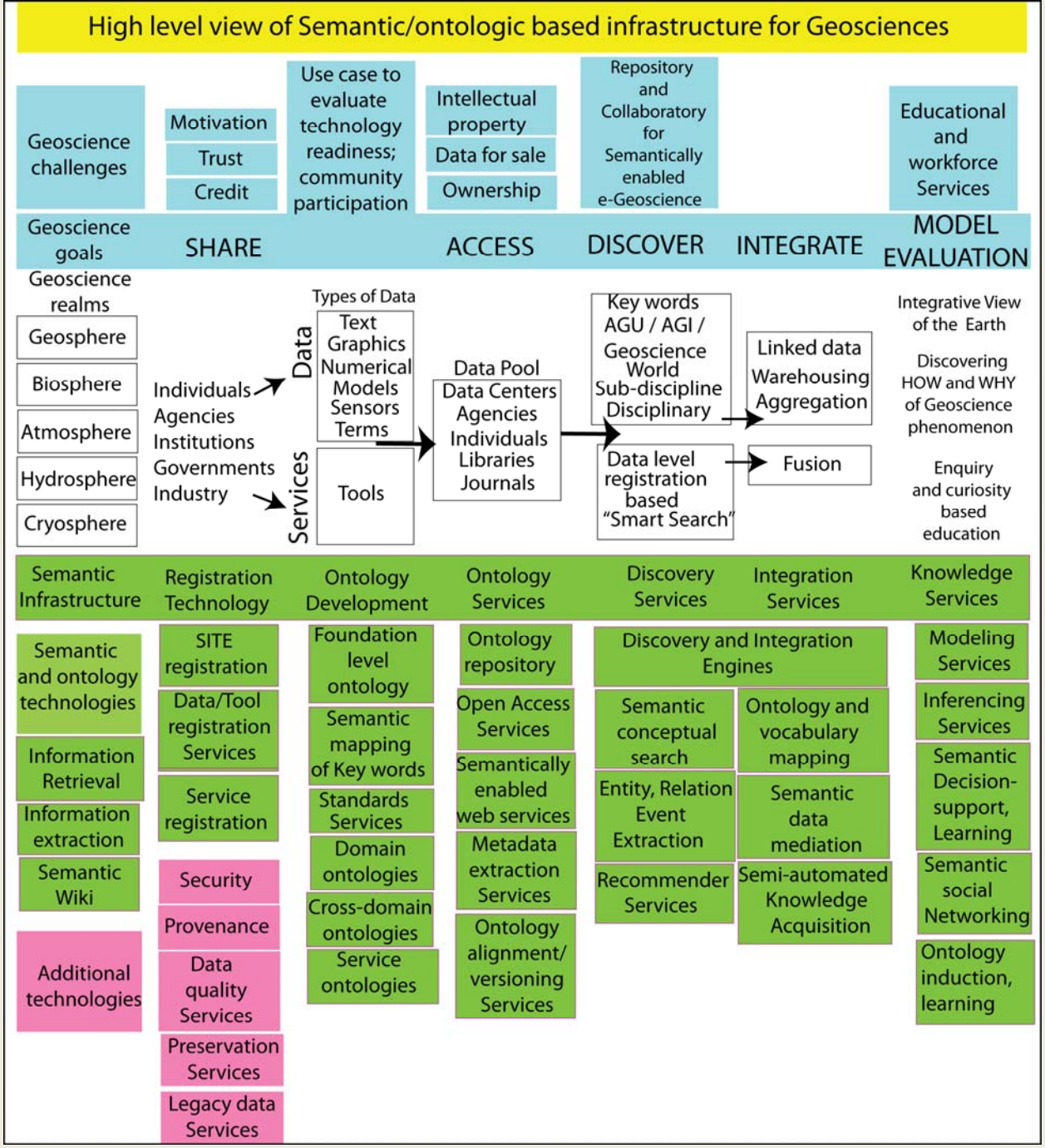


STAGES ASSOCIATED WITH DATA TO KNOWLEDGE TRANSFORMATION



Outline for assembling building blocks for EarthCube Infrastructure




High level view of Semantic/ontologic based infrastructure for Geosciences
Geoscience challenges
Motivation
Trust
Credit
Use case to evaluate technology readiness; community participation
Intellectual property
Data for sale
Ownership
Repository and Collaboratory for Semantically enabled e-Geoscience
Educational and workforce Services
Geoscience goals
SHARE
ACCESS
DISCOVER
INTEGRATE
MODEL EVALUATION
Geoscience realms
Geosphere
Biosphere
Atmosphere
Hydrosphere
Cryosphere
Individuals
Agencies
Institutions
Governments
Industry
Types of Data
Data
Text
Graphics
Numerical
Models
Sensors
Terms
Services
Tools
Data Pool
Data Centers
Agencies
Individuals
Libraries
Journals
Key words
AGU / AGI / Geoscience World
Sub-discipline
Disciplinary
Data level registration based "Smart Search"
Linked data
Warehousing
Aggregation
Fusion
Integrative View of the Earth
Discovering HOW and WHY of Geoscience phenomenon
Enquiry and curiosity based education
Semantic Infrastructure
Registration Technology
Ontology Development
Ontology Services
Discovery Services
Integration Services
Knowledge Services
Semantic and ontology technologies
Information Retrieval
Information extraction
Semantic Wiki
Additional technologies
SITE registration
Data/Tool registration Services
Service registration
Security
Provenance
Data quality Services
Preservation Services
Legacy data Services
Foundation level ontology
Semantic mapping of Key words
Standards Services
Domain ontologies
Cross-domain ontologies
Service ontologies
Ontology repository
Open Access Services
Semantically enabled web services
Metadata extraction Services
Ontology alignment/ versioning Services
Discovery and Integration Engines
Semantic conceptual search
Entity, Relation Event Extraction
Recommender Services
Ontology and vocabulary mapping
Semantic data mediation
Semi-automated Knowledge Acquisition
Modeling Services
Inferencing Services
Semantic Decision-support, Learning
Semantic social Networking
Ontology induction, learning

Summary

- Infrastructures development to be co-ordinate through stages
- Individual scientists right to data be protected
- Use cases to reflect social (cultural) and scientific environments
- Semantic capabilities for discovery to go beyond vocabulary to data level ontologies
- EarthCube will succeed through changes in social environment