Global Scientific Data Infrastructures: Empowering the Multidisciplinary/Interdisciplinary Science

Costantino Thanos

ISTI - CNR

GRDI2020
A Vision for Global Research Data Infrastructures

Multidisciplinary Science

By Multidisciplinary Science we intend a Science in which the approaches adopted to resolve a problem draw appropriately from multiple disciplines in order to redefine the problem outside of the normal boundaries and identify solutions based on a new understanding of complex situations

GRDI2020
A Vision for Global Research Data Infrastructures

Interdisciplinary Science

By Interdisciplinary Science we intend a Science in which the approaches adopted to resolve a research problem involve the connection and integration of expertise belonging to different disciplines

GRDI2020
A Vision for Global Research Data Infrastructures

Scientific Data Infrastructures

The next generation of Scientific Data Infrastructures should face two main challenges:

- to effectively and efficiently support multidisciplinary science
- to effectively and efficiently support interdisciplinary science

GRDI2020
A Vision for Global Research Data Infrastructures

Scientific Data Infrastructures

Scientific Data Infrastructures can be defined as managed digital data-networked environments consisting of services and tools that support the full life cycle of scientific data (capture, collection, curation, documentation, analysis, visualization, preservation, and publication) for the benefit of different communities of researchers involved in data-intensive multidisciplinary/interdisciplinary activities

GRDI2020
A Vision for Global Research Data Infrastructures

Behavioral Barriers

Multidisciplinary Research

Long established tradition of highly focused researchers cultivating a protective (and thus restrictive) boundary around their area of expertise.

Interdisciplinary Research

Most researchers engaged in interdisciplinary research are trained in traditional disciplines and so they have to learn to appreciate different perspectives and methods.

GRDI2020
A Vision for Global Research Data Infrastructures

Technological Barriers

Multidisciplinary Research

Several barriers (syntactical, semantic, and pragmatic) must be overcome when moving data between disciplines.

Interdisciplinary Research

There is a need to integrate data and activities that are based on different ontological foundations.

Information/Knowledge Characteristics

The characteristics of knowledge that drive problem solving within a discipline actually hinder problem solving and knowledge creation across disciplines.

Knowledge can be described as:

- localized
- embedded and
- invested

GRDI2020
A Vision for Global Research Data Infrastructures

Information/Knowledge Characteristics

Knowledge is localized around particular problems faced by a discipline.

Knowledge is embedded in the technologies, methods, and rules of thumb used by individuals in a given discipline

Knowledge is invested in methods, ways of doing things, and successes that demonstrate the value of the knowledge developed

Information/Knowledge Characteristics

This specialization of knowledge in practice makes working across discipline boundaries and accommodating the knowledge developed in another discipline especially difficult.

Information/Knowledge Boundaries

The information objects/knowledge when moving between communities of practice or disciplines have to cross a number of knowledge boundaries:

- Syntactic boundary
- Semantic boundary
- Pragmatic boundary

Syntactic Boundary

A syntactic boundary is constituted by the different syntax of the languages used by the communities/disciplines in order to interact between them.

One shared and stable syntax across a given boundary can guarantee an accurate communication between two communities/disciplines.

Alternatevily, a function that maps the syntax used by one community/discipline into a semantically equivalent syntax used by the other community/discipline is sufficient to overcome the syntactic boundary.

GRDI2020
A Vision for Global Research Data Infrastructures

Semantic Boundary

A semantic boundary can arise even if a common syntax is present due to the fact that interpretations are often different.

A shared and stable ontology across a given boundary could allow the interacting communities/disciplines to share the meaning of the exchanged information objects.

Particular attention must be paid to the challenges of conveyed meaning and the possible interpretations by individuals; context specific aspects of creating and transferring knowledge must be considered.

Pragmatic Boundary

A pragmatic boundary arises when a community/discipline is trying to influence or transform the knowledge created by another community/discipline.

A shared syntax and meaning are not always sufficient to permit the cooperation between communities/disciplines. In fact, some rules that discipline the activities of the cooperating communities/disciplines can hider the exploitation of the exchanged information.

Compatibility of policies, quality, etc. established by cooperating communities/disciplines can allow to overcome pragmatic boundaries.

Boundary Objects

A useful means of representing and transforming knowledge to resolve the consequences that exist at a given boundary is the boundary object.

"… both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual site-use. Like a blackboard, a boundary object "sits in the middle of a group of actors with divergent viewpoints……."

(S. Star)

Boundary Objects

The concept of a boundary object, developed by Star, describes information objects that are shared and shareable across different problem solving contexts.

A boundary object in order to overcome a syntactic, semantic, pragmatic boundary should establish a shared syntax, a shared means for representing and specifying differences, and a shared means for representing and specifying dependencies.

Boundary Objects

At a syntactic boundary:

a boundary object must establish a shared (meta) data model, a shared data language, a shared taxonomy, etc.

At a semantic boundary:

a boundary object must establish a shared ontology, a shared methodology, etc.

At a pragmatic boundary:

a boundary object must establish a shared policy framework, a shared quality framework, etc.

Discipline-specific boundary objects

Communities of practice belonging to the same discipline in order to be able to work together must create a consistent set of boundary objects at syntactic, semantic, and pragmatic boundaries.

In the next future discipline-specific boundary objects will be developed:

(metadata models, data models, data languages, taxonomies, ontologies, policies, etc.

In many disciplines there is currently underway a major effort towards the definition of discipline-specific boundary objects.

Discipline-specific Data Infrastructures

A discipline-specific data infrastructure supporting cooperation among the different communities of practice of a discipline must efficiently implement a consistent set of discipline-specific boundary objects.

Multidisciplinary Boundary Objects

It is much more problematic the definition of boundary objects between different disciplines.

We envision that in order to enable multidisciplinary/interdisciplinary research new methods and techniques must be developed which implement a mediation function between boundary objects of different disciplines.

By mediation function we mean a function able to map a boundary object defined by a discipline into a semantically equivalent boundary object defined by another discipline.

Mediation

A data infrastructure must allow the "data flow" between cooperating "communities of practice" to cross syntactic and semantic boundaries without distortions.

To make this happen a mediation function must be implemented by a software device (mediator) capable of allowing data to be exchanged according to syntactic, structural, and semantic matching.

GRDI2020
A Vision for Global Research Data Infrastructures

Multidisciplinary Data Infrastructures

The future data infrastructures should effectively support multidisciplinary/interdisciplinary research by developing a new mediation technology

We envision that one of the most important features of these data infrastructures will be the mediation software.

GRDI2020
A Vision for Global Research Data Infrastructures

Mediation

The ultimate aim should be the definition and implementation of an "integrated mediation framework" capable of providing means to handle and resolve all kinds of heterogeneities and inconsistencies that may hamper the effective usage of the resources of an information infrastructure