

MEDIATION: THE TECHNOLOGICAL FOUNDATION OF MODERN SCIENCE

Costantino Thanos

Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR)
 Email: Costantino.Thanos@isti.cnr.it

ABSTRACT

Modern science is increasingly data-intensive, multidisciplinary, and network-centric. There is an emerging consensus among the members of the academic research community that the practices of this new science paradigm should be congruent with “open science”. This entails that the bonanza of research data, the wide availability of algorithms, data tools, and data services produced by the members of the research community must be discoverable, understandable, and usable by overcoming all kinds of heterogeneity and logical inconsistencies. The main concept for coping with the many dimensions of heterogeneity and logical inconsistency is mediation. Mediation is achieved by mediators or brokers. These are software modules that exploit encoded knowledge about certain datasets, data services, and user needs in order to implement an intermediary service. A mediating environment is an environment that provides a core set of intermediary services. Mediation should be a distinct functionality of future research data infrastructures. This paper surveys the different levels of interoperability, i.e., exchangeability, compatibility, and usability, their properties and relationships, mediation concepts, functions, and intermediary services. The current interoperability landscape is also illustrated. Finally, the paper advocates the need for mediating environments to be supported by future research data infrastructures and envisions that one of the most important features of future research data infrastructures will be mediation software.

Keywords: Exchangeability, Compatibility, Usability, Interoperability, Mediation

1 INTRODUCTION

A new science paradigm is emerging characterized by: the availability of publicly network-accessible vast volumes of curated scientific data, i.e., *data intensive science* (Hey, Tansley, & Tolle, 2009); the drawing from multiple scientific disciplines in order to find solutions to difficult problems based on a new understanding of complex situations, i.e., *multidisciplinary science*; and the increasingly global collaborations of scientists and of shared resources, i.e., *science globalism*. In addition, the scientific computational environment is characterized by the number of computer systems, information repositories, scientific applications, and users multiplying at an explosive rate.

These characteristics of science entail that large volumes of scientific datasets, tools, and services should be moved across scientific disciplines. There are several technological barriers that must be overcome in order to effectively and efficiently support data movement. In particular, there is the risk of interpreting data descriptions in different ways caused by the loss of the interpretative context. This can lead to a phenomenon called “ontological drift” as the intended meaning becomes distorted as information moves across semantic boundaries (semantic distortion) (Bannon & Bodker, 1997).

In a distributed science environment, different service providers use different independent vocabularies or ontologies to describe data tools/services, for example, data mining/visualization/analysis tools. This makes it hard to achieve data tool/service discovery. The data bonanza and the large availability of data tools and services demand more *mediation* in order to allow heterogeneous parties to communicate and interoperate. In fact, the new science paradigm very much depends on the ability to *reconcile* information from multiple sources and to make geographically and institutionally separated research teams interoperable.

Interoperability is an extremely complex and evolving problem. Although researchers have been struggling with interoperability for many years, it is often not clear what principles or key results have been established (Paepcke, Chang, Garcia-Molina, & Winograd, 1998). There is a wide range of views as to what “interoperability” means

(Wegner, 1996; Park & Ram, 2004; Sciore, Siegel, & Rosenthal, 1994). It means different things to different people. Interoperability intended as the ability of two entities to work together very much depends on the working context in which these two entities are embedded (research infrastructures, digital libraries, Web services, control and command systems, etc.) and the nature of the interoperable entities (people, software components, organizations, etc.). This leads to a wide spectrum of definitions of interoperability. At one end of the spectrum, there are definitions that focus mainly on technological aspects while at the other end the focus is on legal and political aspects (Gasser & Palfrey, 2007). As a consequence, interoperability is a multifaceted complex concept. Due to this inherent complexity and multifaceted nature, interoperability has been often misunderstood. Simple data exchangeability has been confused with interoperability; several forms of service compatibility (composability, replaceability, etc.) have also been confused with interoperability. Furthermore, when addressing interoperability between two entities, the fact that often they belong to two different organizations that have their own policies has been ignored. The fact that true interoperability between two entities can occur only within a shared policy framework between two organizations (trusted organizations) has not been addressed in the literature.

We think that the development of mediating environments is of paramount importance. These should be implemented by research data infrastructures, able to provide a number of intermediary services in order to link data, knowledge, tools, and scientists by overcoming the different kinds of heterogeneities and logical inconsistencies.

The goal of this paper is to (i) present a broad and formal introduction to the issues of interoperability; (ii) define formally the different levels of interoperability and their properties; (iii) identify the existing relationships among the different levels of interoperability; (iv) suggest, for each level of interoperability, the appropriate mediation technologies to be adopted in order to accomplish the desired level of interoperability; and (v) stimulate research activity towards an extension of the binary model of interoperability in order to embrace multilateral modes of interoperation as they are more appropriate for modern multidisciplinary scientific environments.

The paper is organized in the following way. Section 2 gives some basic definitions: it describes the different aspects and properties of data exchangeability, discusses the different aspects and properties of service/policy/behavior compatibility, and discusses the data/service usability problem and its conceptual foundation. Section 3 illustrates the relationships among exchangeability, compatibility, usability, and interoperability. Section 4 discusses some extensions to the binary model of interoperability. Section 5 addresses the mediation technology able to bridge the several information heterogeneities (syntactic, structural, semantic) as well as inconsistencies of logical representations of functionality, policy, and behavior. Section 6 describes the interoperability landscape. Section 7 discusses the role of standards in achieving interoperability. Section 8 summarizes the main points to be taken into consideration when addressing the pressing need for achieving exchangeability, compatibility, and usability of data and services.

2 BASIC DEFINITIONS

From the IEEE definition of interoperability: *The ability of two or more systems or components to exchange information and to use the information that has been exchanged.* It follows that in order to achieve interoperability between two entities (author, user) three conditions must be satisfied:

- The two entities must be able to exchange *meaningful* information (*exchangeability*);
- The two entities must be able to exchange *logically consistent* information (when the exchanged information is a description of functionality, policy, or behavior) (*compatibility*);
- The user entity must be able to use the exchanged information in order to perform a set of tasks that depend on the utilization of this information (*usability*).

In essence, this definition entails that two entities (author, user) are interoperable only when:

- The data produced by the author entity are *understandable* and *usable* by the user entity;
- The descriptions of a functionality of a service/policy/behavior produced by an author entity are *logically consistent* with the user needs and *usable* by the user entity.

2.1 Data exchangeability

Data exchangeability is the process of taking data structured under a schema, called the source schema, and transforming it into data structured under another schema, called the target schema (Kolaitis, 2005).

The Data Heterogeneity Problem: during the data exchange process between author and user entities, different sources of heterogeneity can be encountered depending on (Papakonstantinou, Garcia-Molina, & Widom, 1995):

- How data are requested by the user entity;
- The use of different terminologies;
- How data are represented;
- The semantic meaning of data;
- How data are actually transported over a network.

Therefore, there are three types of heterogeneity to be overcome in order to achieve a *meaningful* exchange of data: First, heterogeneity between the data/query languages adopted by the author and the user entities. When this heterogeneity is resolved we say that *syntactic exchangeability* between the two entities has been achieved. Second, heterogeneity between the data models adopted by the author and the user entities for representing data. When this heterogeneity is resolved we say that *structural exchangeability* between the two entities has been achieved. Third, heterogeneity between the “semantic universe of discourses” of the author and user entities (differences in granularity, differences in scope, temporal differences, synonyms, homonyms, etc.). When this heterogeneity is resolved we say that *semantic exchangeability* between the two entities has been achieved (Heiler, 1995; March, Henver, & Ram, 2000).

These three levels of exchangeability, i.e., syntactic, structural, and semantic, allow a *meaningful exchange* of data between two interoperating entities (author, user), i.e., they make the exchanged data *understandable* by the user entity (Kolaitis, 2005). An author entity, in order to make data understandable by a user entity, must endow it with appropriate metadata information (provenance, context, quality, uncertainty, etc.). However, exchangeability does not guarantee that the exchanged data are also *usable* by the user entity, and thus it doesn't guarantee that the two entities are interoperable.

Properties of Exchangeability (Wang, 2005): exchangeability enjoys the following two properties:

Asymmetry

If the data D_A produced by entity A are meaningful for entity B, this does not imply that the data D_B produced by entity B are meaningful for entity A.

Transitivity

If the D_A produced by entity A are meaningful for entity B and the data D_B produced by entity B are meaningful for entity C, this implies that the data D_A are also meaningful for entity C.

2.2 Compatibility

Different types of mismatches on the functional, behavioral, policy, and business logics levels can arise (Stollberg, et al., 2006):

The Functional Mismatching Problem: mismatches on the functional level arise when the functionality provided by a service provider does not precisely match with the one requested by a user.

The Behavioral Mismatching Problem: mismatches on the behavioral level can arise during the consumption of a service S by a requester R. For example, the requester R expects an acknowledgment while the service S waits for the next input; in this case the interaction process between R and S runs into a deadlock situation.

The Business Logics Mismatching Problem: mismatches on the business logics level can arise when services that provide complementary functionalities execute interaction protocols whose respective behaviors do not match.

The Logical Inconsistency Problem: in the above situations the interacting entities can be software systems, data organizations, service providers and users, etc. In all these cases the exchanged information could be, for example, a description of the functionality offered by a service provided by a software component or a service provider, or

descriptions of service requirements of a service user, or a description of a policy adopted by an organization, for example, a data policy adopted by a discipline-specific data center or a description of a user behavior, etc. These descriptions are usually expressed in some logic-based languages. Syntactic and semantic heterogeneities can also arise when functionality, policy, or behavior are described by using different logic-based/knowledge representation languages and/or when semantic conflicts arise from differences in implicit meanings, perspectives, and assumptions.

In addition, a *logical inconsistency* problem can arise. In fact, when the exchanged information specifies the functionality of services to be composed or the functionality of services provided by a service provider to meet the service needs of a service user or describes policies of cooperating organizations or user/software components behaviors, some inconsistencies may arise between the logical relationships of these descriptions. This means that the logical relationships between the functionality/policy/behavior descriptions do not share a logical framework. In this case, we have functional inconsistency between two service descriptions, policy inconsistency between the policies of two organizations, etc. When these inconsistencies are resolved, we say that two services are compatible, i.e., a logic compatibility between two services has been established; or that the policies of two organizations are compatible, i.e., a logic compatibility between their policies has been established; or two user behaviors are compatible. The syntactic and semantic exchangeability together with the logical consistency guarantee that two services are compatible, i.e., are composable, replaceable, etc., two policies are compatible as well as two behaviors.

A data tool/service provider, in order to enable the checking of the compatibility with another service or the service needs of a user, must provide a two level description of the service provided (Keller, et al., 2005; Stollberg, et al., 2006): (i) a first level that describes the static characteristics of the service also called abstract capabilities - the abstract capabilities of a service describe only what a published service can provide, not under which circumstances a concrete service can actually be provided and (ii) a second level that describes the dynamic characteristics of the service also called contracting capabilities - the contracting capabilities describe what input information is required for providing a concrete service, what conditions it must fulfill (i.e., service pre conditions), and what conditions the information delivered fulfills depending on the input given (i.e., post conditions). However, compatibility does not guarantee that, for example, two compatible services can actually be composed and a concrete service provided. In fact, a concrete service can be provided only if the requirements of the operational environment, i.e., the deployment capabilities where the service will be hosted, are met. Therefore, compatibility does not guarantee interoperability.

Properties of Compatibility (Wang, 2005): compatibility enjoys the following two properties:

Asymmetry

If a functionality/policy/behavior F_A of entity A logically implies a functionality/policy/behavior F_B of entity B, this does not mean that F_B implies F_A .

Transitivity

If a functionality/policy/behavior F_A of entity A logically implies a functionality/policy/behavior F_B of entity B, and F_B logically implies a functionality/policy/behavior F_C of entity C, then F_A implies F_C .

2.3 Usability

By data usability we mean the ease of using data that is produced by a data author for legitimate scientific research by a data user. We use the term data reusability to mean the easy use of data collected for one purpose, to study a new problem (Zimmerman, 2003). This term denotes the reutilization of existing data sets in significantly different contexts. According to Davis (1989) usability has two determinants: *perceived usefulness* and *perceived ease of use*. By perceived usefulness we mean the degree to which a user believes that using a particular data set/data tool/data service produced by a data author or a data tool/service provider would enhance her/his job performance. By perceived ease of use we mean the degree to which a user believes that using a particular dataset/data tool/data service would be free of effort.

The Usage Inconsistency Problem: usage inconsistency occurs when the gap between perceived usefulness and perceived ease of use is wide. This gap hampers the accomplishment of the user's goal as she/he is unable to effectively and easily use the exchanged data.

Possible causes for usage inconsistency are:

- Data quality mismatching;

- Data-incomplete mismatching;
- Data abstraction mismatching;
- Lack of data tool/service metadata;
- Data service deployment capabilities mismatching;
- Other.

Quality mismatching: data quality has a number of specific dimensions. A dimension or characteristic captures a specific facet of quality. The more commonly referenced dimensions include: *accuracy, completeness, consistency, currency, timeliness, and volatility*. For specific categories of data and for specific scientific disciplines, it may be appropriate to have specific sets of dimensions. Quality mismatching occurs when the set of quality dimensions associated with the exported data is not the one expected by the user entity.

Data-incomplete mismatching occurs when the exported data lack some useful information to enable the user entity to fully exploit the received data.

Data abstraction mismatching occurs when the level of data abstraction (spatial, temporal, graphical, etc.) created by an author entity does not meet the expected level of abstraction by the user entity.

Lack of data tool/service metadata occurs when the exported data tool/service is not endowed with appropriate metadata.

Service deployment capabilities mismatching occurs when the deployment capabilities, i.e., the capabilities that describe the hosting operational environment where the exchanged service will be hosted do not allow the running of this service.

Usability is accomplished when perceived usefulness and perceived ease of use of data/tool/service are tightly linked.

We think that usability is a *relational concept* (Osterlund & Carlile, 2005). This means that a data set/data tool/data service produced by an author/provider entity in order to be usable by a user entity must be endowed with some auxiliary information that takes into account the characteristics of the *usability relation* established between the two entities. Several kinds of usability relations can be established between two entities. For example, a *confirmation relation* is established when the user entity tries to find a confirmation of some scientific expectation by gathering enough evidence from a data set produced by the author entity. Another kind of usability relation is the *reproduction/verification relation* that is established when the user entity tries to reproduce/verify a scientific result by using a data set produced by the author entity. One more kind of usability relation is the *discovery relation* that is established when the user entity tries to discover new insights from a data set produced by the author entity. Therefore, an author/provider entity in order to make a data set/data tool/data service usable by a user entity must complement it with appropriate metadata information. The properties of the metadata information (provenance, context, quality, uncertainty, functionality, etc.) heavily depend on the *usability relation* established between the author and user entities. Thus, if a data set is to be used by different user entities, different metadata information must be provided to these diverse entities depending on the characteristics of the *usability relations* that link the author entity with them. For example, for one user entity it could be enough to know who, where, and when a data set was produced; for another it could be important to know how this data set was produced.

3 RELATIONSHIPS BETWEEN EXCHANGEABILITY, COMPATIBILITY, USABILITY, AND INTEROPERABILITY

The essence of interoperability is the ability of two entities to work together. It is implemented by three consecutive actions: exchanging meaningful information, making this information compatible with the user's needs, and making it usable by the user entity. The first action, i.e., exchangeability, is responsible for making the exchanged information understandable. It consists, as already said, of three levels: syntactic, structural, and semantic, which together contribute only to making the exchanged information understandable by the user entity. They do not guarantee that it is also usable. Therefore, exchangeability is a necessary but not sufficient condition for achieving interoperability between two entities. When the sole exchangeability is sufficient to enable the user entity to perform a set of tasks based on the exchanged information, we can talk about *basic* interoperability between author and user entities.

The second action, i.e., compatibility, is applicable when the exchanged information describes the functionality of a data tool/service, a policy, a user behavior, etc. It is responsible for guaranteeing consistency between logical

descriptions of data tool/service functionality, policy, behavior, and user need. Obviously, the exchanged logical descriptions must be understandable. Therefore, exchangeability is a necessary but not sufficient condition for assuring compatibility. Compatibility is a *weaker* concept than interoperability as it only guarantees logical consistency between two descriptions.

The third action, i.e., usability, is responsible for making the exchanged information usable. This means bridging the gap between perceived usefulness and perceived ease of use as the essence of usability is the *ability* of the user entity to easily and effectively *use* the information received from the author entity. Exchangeability and compatibility are necessary but not sufficient conditions for guaranteeing usability. *When exchangeability, compatibility, and usability are assured, we say that interoperability between two entities has been accomplished.* Based on the above considerations, the following relationships hold between exchangeability, compatibility, usability and interoperability (see Figure 1):

- Usability implies exchangeability, but the reverse is not true.
- Compatibility implies exchangeability, but the reverse is not true.
- Usability implies compatibility, but the reverse is not true.

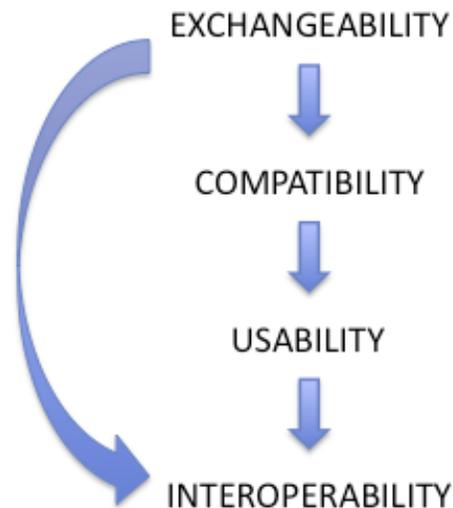


Figure 1. Hierarchy of different interoperability layers

Sometimes interoperability and compatibility between two entities can be characterized by the type of tasks the user entity is enabled to perform on the exchanged information. For example, if the user entity applies a preservation action on the exchanged data, we characterize this type of interoperability as *temporal interoperability* as it guarantees access to the exchanged data over time. Another type of interoperability occurs when the user entity is obliged to observe security, integrity, confidentiality/privacy, etc. constraints when performing tasks on the exchanged information. In this case we characterize this type of interoperability as *secure interoperability*. When the user entity (usually a middleware component) is searching for a provider entity (another middleware component) that provides a compatible functionality, we characterize this type of compatibility as *functional compatibility*. When the user entity (usually an organization), in order to perform some tasks, needs to check the compatibility of its behavior/policy with those of a provider entity (another organization), then we characterize these types of compatibility as *behavioral compatibility* or *organizational compatibility*, respectively. Therefore, temporal, secure, behavioral, functional, organizational, etc. are specializations of the interoperability and compatibility concepts.

4 EXTENSIONS TO THE BINARY MODEL OF INTEROPERABILITY

The IEEE definition of interoperability and the author-user model that is based on this definition and used to describe the interoperability problem have been criticized as being inadequate to describe all the facets of the interoperability problem. In essence, these are the main criticisms:

Asymmetric roles of author/user

This criticism regards the fact that often the process of the information exchange is not a one-way flow (author → user) but it may be bi-directional (author ↔ user). Unfortunately, the *asymmetry* property of exchangeability does not permit the adoption of a bi-directional model for describing the interoperability problem. We always have to decompose the bi-directional model in two unidirectional models.

Bilateral versus multilateral or direct versus indirect interoperability models

This criticism points out the fact that the author-user model conveys the idea that interoperability is a binary problem, i.e., it only regards the ability of two entities to work together. In a networked scientific environment, more than two entities may be involved in carrying out a task, and therefore the scope of interoperability is wider than that of the binary model. Below we illustrate two scenarios of multilateral interoperability.

SCENARIO 1

A Decision Maker (entity A) asks a Service Provider (entity B) to produce a report containing some statistics about some economic activities; the Service Provider collects data from a Data Author (entity C), compiles the report, and sends it to the Policy Maker (see Figure 2). In this scenario, in addition to the necessary “direct” functional compatibility between A and B and the interoperability between B and C, there is a need for an “indirect” interoperability between A and C. This indirect interoperability is inferred by the *transitivity* property of exchangeability and compatibility.

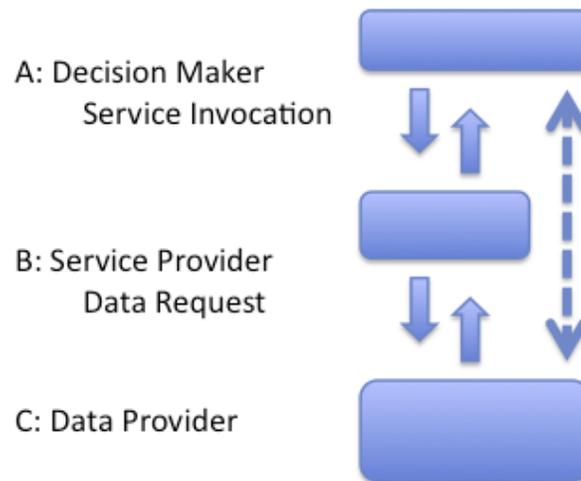


Figure 2. Example of indirect interoperability

SCENARIO 2

In this hypothetical scenario, entity A in order to perform a task invokes a service from entity B and requests some data from entity C; entity B in order to perform the service requested by entity A requests some data from entity D (see Figure 3). In order to implement this scenario, there must be achieved: a direct functional compatibility between A and B entities, a direct interoperability between A and C entities, a direct interoperability between B and D, and an indirect interoperability between A and D entities. The indirect interoperability between A and D is inferred by the “transitivity” property of exchangeability and compatibility while an indirect interoperability between B and C and C and D cannot be inferred by the fact that each of them is interoperable with entity A.

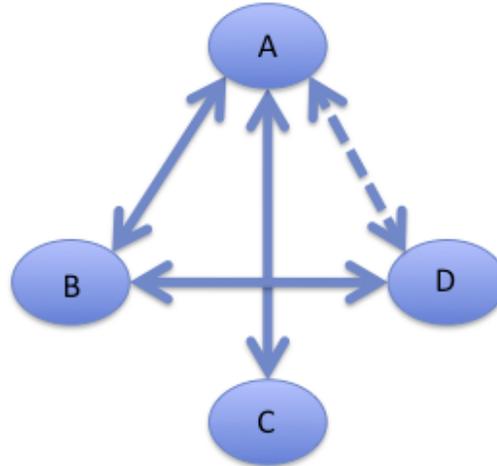


Figure 3. Example of limited indirect interoperability

These two scenarios demonstrate that actually a “multilateral” interoperability is achieved only when a transitivity property holds among the entities wishing to work together.

Data-centrism of the interoperability definition

This criticism considers the definition of interoperability proposed by IEEE as data-centric, and thus it does not adequately reflect the fact that the object of interoperability can be not only data but also services, policies, behaviors, etc. The objection is founded in the sense that this definition of interoperability has contributed to confusing interoperability with data exchangeability. In addition, the concept of compatibility as a weaker form of interoperability is not at all taken into consideration. Therefore, a more general and complete definition of the interoperability concept must be formulated.

In an effort to contributing to the formulation of a new definition of interoperability we propose the following: *Interoperability is the ability of two or more entities to work together by exchanging data and services and using them within the context of an agreed quality and policy framework.*

5 MEDIATION

5.1 Mediation concepts and functions

The main concept enabling the “meaningful” exchange of data is *mediation* (Kahng & McLeod, 1998; Huhns & Singh, 1998; Ludascher, Gupta, & Martone, 2001; Wiederhold & Genesereth, 1997). This concept has been used to cope with many dimensions of heterogeneity, i.e., data language syntaxes, data models, and data semantics as well as logical inconsistencies of functionality/policy/behavior representations. The mediation concept is implemented within a *mediating environment* that enables the establishment of exchangeability and/or compatibility of resources by resolving heterogeneities and logical inconsistencies. In particular, the mediating environment should support four main mediation levels (Spalazzese, Inverardi, & Issarny, 2009):

- *Mediation of data structures:* this permits data to be exchanged according to syntactic, structural, and semantic matching;
- *Mediation of functionalities:* this makes it possible to overcome logical inconsistencies among representations of service functionality;
- *Mediation of policies:* this makes it possible to resolve inconsistencies between policies;
- *Mediation of protocols:* this makes it possible to overcome behavioral mismatches among protocols run by interacting parties.

A mediating environment, in order to support the above mediation levels, should support a mediation schema (Ullman, 1997) capturing user requirements and providing a core set of *intermediary services* between this schema and the distributed information resources. The core set of intermediary services should include (Wiederhold, 1994):

- *Data Discovering*: refers to the action of quickly and accurately finding data that support specific research requirements;
- *Service Discovering*: refers to the action of locating data tools/services that fulfill a research goal;
- *Mapping*: refers to how data structures, properties, and relationships are mapped from one representation scheme to another one, equivalent from the semantic point of view;
- *Matching*: refers to the action of verifying whether two strings/patterns match or whether semantically heterogeneous data match (*matching*);
- *Integration*: refers to the action of combining data residing at different sources and providing the consumer entity with a unified view of these data;
- *Consistency checking*: refers to the action of checking whether the logical relationships between the static and dynamic descriptions of functionality/policy/behavior share a logical framework;
- *Optimization*: refers to the action of optimizing access strategies to provide small response times or low cost;
- *Resolution*: refers to the action of resolving domain terminology and ontology differences and also to the action of resolving scope mismatching;
- *Pruning*: refers to the action of pruning data ranked low in quality or relevance;
- *Summarizing*: refers to the action of producing statistical summarization into higher level objects as defined by the consumer model.

5.2 Automated mediation

The mediation paradigm depends on models: models of resources and models of user needs. *Automated mediation* basically focuses on matching the information resources of the author entity to the user entity needs (Rahm & Bernstein, 2001). It heavily relies on *adequate modeling* of both the exchanged information and the user needs. In essence, the intermediary functions must translate languages, data structures, logical representations, and concepts between two systems. The effectiveness, efficiency, and computational complexity of the intermediary function very much depend on the characteristics of the information models (expressiveness, levels of abstraction, semantic completeness, reasoning mechanisms, etc.) and languages adopted by the author-user entities. Ideally, they must provide a *framework for semantics and reasoning*. Therefore, the interoperable entities must adopt formally defined and scientifically sound information models and ontologies. They constitute the conceptual (semantic) and syntactic basis for data languages.

An information model is a representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse. It can provide sharable, stable, and organized structure of information requirements for the domain context. Several formal information models and languages have been defined and developed for representing, organizing, and exchanging information objects (for example, RDF, XML, etc.). Several *discipline-specific standard models* have been proposed and developed for representing discipline-specific descriptive information (discipline-specific metadata models) that greatly support the mediation process (Haslhofer & Klas, 2010). *Logic-based and ontology-based models and languages* have been defined for specifying behavior, functionality, and policy (for example, OWL-S).

An important role in the mediation process is played by *ontologies*. Several domain-specific ontologies are being developed (gene ontology, sequence ontology, cell type ontology, biomedical ontology, CIDOC, etc.). Ontologies were initially developed by the artificial intelligence community to facilitate knowledge sharing and reuse. An ontology consists of a set of concepts, axioms, and relationships that describes a domain of interest. Ontologies have been extensively used to support all the intermediary functions because they provide an explicit and machine-understandable conceptualization of a domain.

Thus, *automated mediation* relies on:

- Adequate modeling of structural, formatting, and encoding constraints of the author entity information resources;
- Adequate modeling of data descriptive information (metadata);
- Adequate modeling of the user entity needs;

- Formal domain-specific ontologies;
- Abstract models and languages for policy specification;
- Abstract models and languages for functionality specification;
- Formally defined transfer and message exchange protocols;
- The definition of a matching relationship between the author information resources and the user models

Automated mediation entails making the semantics of the exchanged information explicit. This requires the involvement of people (users, designers, and developers) who intuitively associate semantics with data and procedure names, type definitions, type hierarchies, etc. Other semantic information is implicit in application code, in text and diagrams, and in local “oral tradition”. Making semantics explicit in metadata would allow people to detect mismatched assumptions and to create mappings to overcome them. However, making the necessary semantic information explicit can be extraordinary difficult for several reasons (Heiler, 1995):

- Discovering semantic information and resolving mismatches requires the application of human intelligence and judgment;
- Few tools are available to help, except to discover simple name matches;
- Documenting the semantics of legacy systems is an enormous task;
- Interesting semantic information is context-dependent so documenters will need to understand the planned applications and guess at the unplanned ones;
- The meanings of names and values may change over time;
- The resulting metadata must be managed and will require similar agreements about the semantics of the terminology used in the documentation.

Addressing these points would require extensions in at least two areas: discovery of semantics and representation and management of mappings. Understanding data and software can never be fully automated. However, advances in knowledge technologies could ease this task.

5.3 Evaluating mediation approaches

There are many different mediation approaches that operate under different assumptions. It is therefore important to understand tradeoffs among them (Paepcke, Chang, Garcia-Molina & Winograd, 1998). We propose the following important criteria for evaluating the tradeoffs made by any given approach:

- Degree of participant party’s autonomy
- Cost of the mediating environment
- Scalability/openness of the mediating environment
- Complexity of the mediating environment
- Ease of using the mediating environment

These criteria do not provide quantitative measurements but rather useful guidelines for comparing different approaches and understanding their strong and weak points.

Participant party’s autonomy

This criterion refers to the amount of compliance to global rules that is required of each participating party in an interoperable federated/distributed information system. Higher autonomy is better because it provides more local control over implementation and operation of the participating party and also because it makes it easier to include legacy systems as a participating party. Limitation in autonomy may affect many aspects of a participating party. However, high autonomy can lead to solutions that only allow interoperation at the lowest common denominator of functionality.

Cost of the mediating environment

The implementation cost of a mediating environment that supports a given mediation solution is another aspect to be considered. It includes the cost of the software development and maintenance of a core set of intermediary services that altogether accomplish the chosen mediation solution. The mediating environment is an infrastructural service, and as such its cost should be shared among many users. Therefore, these costs can be very difficult to assess. Local costs related to installing and operating an intermediary service are easier to assess.

Scalability/openness of the mediating environment

This criterion concerns the scalability of the mediating environment, i.e., whether the core set of intermediary services implementing a mediation solution is easily extendable by adding new intermediary services to accommodate the requirements of new parties that join the interoperable distributed/federated system. In addition, the incremental cost of enabling interoperability when new parties are joining the system should be considered.

Complexity of the mediating environment

This criterion refers to the computational complexity of the core set of intermediary services that accomplish a mediation solution provided by a data infrastructure.

Ease of using the mediating environment

This criterion refers to the complexity of interacting with the mediating environment at run time. For example, simple query interfaces might make the interaction easy.

These evaluation criteria are interrelated in complex ways. In order to select a particular mediation solution, one must weigh how well that solution satisfies these criteria.

5.4 Research data infrastructures and mediation

Mediation should be a distinct functionality of the research data infrastructures (Thanos, 2013; Kim, 1999). They must support a mediating environment that should:

- Provide a core set of *intermediary services* that make the holdings of discipline-specific repositories and data centers, data archives, data service providers, discipline-specific research infrastructures, domain-specific communities of data authors, and data users discoverable, understandable, and (re)usable, thus making all these entities interoperable within an agreed policy framework.
- Support the creation, operation, and maintenance of *mediators* (Wiederhold, 1992), sometimes also called *brokers* (Nativi, Craglia, & Pearlman, 2013). A mediator is a software module that exploits encoded knowledge about certain sets or subsets of data, data services, and user needs in order to implement an intermediary service. A core set of mediators should include: data discovery mediator, service discovery mediator, mapping mediator, matching mediator, consistency checking mediator, data integration mediator, and policy mediator.
- Maintain metadata registries, data dictionaries, data inventories, and data tool/service registries.

A mediating environment, in essence, should support a two-phase mediation process: the first phase providing assistance in locating and understanding resource capabilities; the second phase focusing on matching identified resources to user needs. The ultimate aim should be the definition and implementation of an *integrated mediating environment* capable of providing means to handle and resolve all kinds of heterogeneities and inconsistencies that might hamper the effective usage of the resources of research data infrastructures. We envision that one of the most important features of the future research data infrastructures will be the mediation software.

6 THE INTEROPERABILITY LANDSCAPE

6.1 Semantic exchangeability

Previous work in data interoperability has focused mainly on semantic exchangeability. This research can be categorized in three broad areas (Park & Ram, 2004):

Mapping-based approach

This approach requires a thorough and explicit description of the semantics and structure of all information sources, i.e., the definition of all data schemas. It creates mappings between semantically related data schemas. Schema mappings are specifications that describe the relationships between data schemas at a high level. These specifications are typically given in a logical formalism that captures the interaction between schemas at a logical level without spelling out implementation details relevant to the physical level. The vision of this approach is to

allow a specification-based interaction among the interacting parties without the help of mediators. Semantic exchangeability solutions based on this approach rate a high degree of autonomy because of their strict separation of data description from the implementation. However, they suffer from the complexity of explicitly representing the semantics of information sources. Schema mappings are widely used in all data management applications that involve data sharing or data transformations. In particular, schema mappings are essential building blocks in information integration and data exchange. There are obvious similarities, but also clear differences, between data integration and data exchange. In both frameworks, schema mappings are used to specify the relationships between the schemas involved. In data integration, the goal is to synthesize data from different sources into a unified view under a global schema; this view is virtual, in that the data remain in the sources and are accessed by the users symbolically via the global schema. In data exchange, the goal is to take a given source instance and transform it to a target instance such that it satisfies the specifications of the schema mapping; unlike data integration, this target instance is a materialized instance, not a virtual view.

Intermediary-based approach

This approach is based on the use of intermediary mechanisms, e.g., mediating connectors, brokers, agents, ontologies, etc. (Sciore, Siegel & Rosenthal, 1994; McLeod & Si, 1995). Such intermediary mechanisms may have domain-specific knowledge, mapping knowledge, or rules specifically developed for coordinating various autonomous data sources. These mechanisms use ontologies to share standardized vocabulary or protocols to communicate with each other. The EuroGEOSS research project has introduced a brokering approach that interconnects heterogeneous disciplinary and domain service buses, avoiding the imposition of any federated or common specification (Nativi, Craglia, & Pearlman, 2013). A broker is a software module that implements a number of functionalities, including semantic discovery, resource tagging, and clustering of results, quality control, etc. Recently, this approach has been adopted by the GEOSS Common Infrastructure (GCI). The main drawback of the mediation-based approach is scalability. When a new party is added, a corresponding mediation facility needs to be built as well. More generally, if this approach is adopted to allow n kinds of parties to exchange information sources with m other kinds, $n \times m$ mediation facilities must be constructed.

Query-oriented approach

This approach is based on interoperable languages, most of which are either declarative logic-based or extended SQL (Czejdo, Rusinkiewics & Embley, 1987; Krishnamurthy, Litwin & Kent, 1991). They are capable of formulating queries spanning several databases. In order to resolve semantic conflicts over data structure and data semantics, it is desirable to have high-order expressions that can range over both data and metadata. The drawback of this approach is that it places too heavy a burden on users by requiring them to understand each of the underlying local databases. This approach typically requires users to engage in the detection and resolution of semantic conflicts because it provides little or no support for identifying semantic conflicts.

6.2 Mediation levels and techniques

Heterogeneity is an inherent characteristic of open and networking environments such as the Internet. Therefore, mediation techniques for handling and resolving mismatches that hamper interoperability of Web resources have been studied extensively in the past (Stollberg, Cimpian, Mocan, & Fensel, 2006). Here, we consider three main mediation levels:

Data level mediation

At this level, the most common type of mismatch occurs due to usage of different terminologies by entities wanting to exchange information. Within ontology-based environments such as the Semantic Web, the mismatch results from the usage of heterogeneous ontologies as the terminological basis for information descriptions. Such types of mismatches can be handled on a semantic level by so-called ontology integration techniques. Another type of heterogeneity is the representational heterogeneity of formats and transfer protocols. A suitable way of resolving such heterogeneities is to lift the data from the syntactic to a semantic level on the basis of ontologies and then resolve the mismatches at this level (Moran & Mocan, 2005).

The main mediation techniques for the data level are:

- *Ontology mapping*
This involves the creation of a set of rules and axioms that precisely define how terms from one ontology relate with terms from the other ontology. The rules and axioms are expressed using a mapping language. Ontology mapping refers to mapping definitions only.
- *Ontology alignment*
This bridges the involved ontologies in a mutual agreement. In this technique one of the involved ontologies has to be altered in order to allow their alignment in the overlapping parts.
- *Ontology merging*
This results in a new ontology that replaces the original ontologies. The merging can be done either by unification or by intersection.

Functional level mediation

At this level, heterogeneity arises when the functionality provided by a service does not precisely match with the one requested by a user. In order to determine the compatibility of a service with a given request, complex reasoning procedures are required (Keller, Lara, Lausen, Polleres, & Fensel, 2005). The main technique proposed for functional level mediation is based on Δ -relations, which denote the explicit logical relationship between functional descriptions of services and goals. Functional descriptions can be defined as conditions in pre- and post-states in some first order logic derivate that provide a black box description of normal runs of a service.

Process level mediation

Mismatches at this level occur during service consumption or interaction. For example, during the consumption of a service S by a requester R , R expects an acknowledgement while S waits for the next input; in this case, the interaction between R and S runs into a deadlock situation. Mismatches at the process level can occur in every interaction a service is involved in. These heterogeneities can be resolved by inspecting the individual processes of the entities that interact and trying to establish a valid process for interaction on the basis of pre-defined mediation operations on processes.

6.3 Legal interoperability

In a federated information system, each party should define its own set of formal semantic policies that enhance the authorization, obligation, and trust processes that permit regulated access and use of data and services (data policies). Semantic policies are described by policy representation and specification languages. Logic compatibility among the local data policies is assured by a mediation facility operated by the system that detects and resolves conflicts among the policies adopted by the local parties.

Legal interoperability means that the legal rights, terms, and conditions of databases from two or more sources are compatible and the data may be combined by any user without compromising the legal rights of any of the data sources used. When substantial amounts of statutorily protected data are combined from two or more data sources, the new resulting database often will have to respect the most restrictive conditions regulating any of the sources used. Legal interoperability for data is essential for the data reuse. However, interoperability is not only the result of technological development but is also shaped by the legal and regulatory system. General laws such as intellectual property law, competition law, consumer protection law, and legal provisions that specifically address interoperability issues have an impact on the interoperability landscape. The same body of law can either be used to achieve higher levels of interoperability or to hinder it.

So far, while the scientific community has focused primarily on the technological aspects of data interoperability, the legal community concentrated on proprietary protection and restrictive licensing of data and information in the commercial sector, rather than enabling common-use in the government and academic sectors. It is therefore both necessary and timely to bring together key stakeholders in both the scientific and legal communities in order to obtain a better understanding of the ways in which the public law of intellectual property and private law of contracts and licenses affect scientific data interoperability and data sharing.

Recently, an *RDA/CODATA Legal Interoperability IG* has been established. It aims at defining the legal interoperability of research data and articulating why it is important for data exchangeability and reuse. The group

will analyze some case studies and establish best practices through which the legal interoperability of research data can be achieved and adopted by stakeholders.

6.4 European interoperability framework (EIF)

The EIF (<http://ec.europa.eu/isa>) addresses interoperability in the very specific context of providing European public services. Although the provision of European public services almost always involves exchanging data between ICT systems, interoperability is a wider concept and encompasses the ability of organisations to work together towards mutually beneficial and commonly agreed goals. The following definition is used in the EIF:

Interoperability, within the context of European public service delivery, is the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the exchange of data between their respective ICT systems.

This framework was issued under the Interoperable Delivery of European e-Government Services to public Administrations, Businesses and Citizens program (IDABC). It consists of a set of recommendations that specify how administrations, businesses, and citizens communicate with each other within the EU and across member states borders:

- (1) *Twelve underlying principles*: these principles illustrate the context in which European public services are established and implemented;
- (2) *The conceptual model for public services*: it helps develop a common vocabulary and understanding about the main elements of a public service; it emphasizes a building-block approach, allowing for the interconnection and reusability of service components when building new services; and it is sufficiently generic to be applicable at any level of government that provides public services;
- (3) *Four layers of interoperability*: in order to implement the conceptual model, four levels of interoperability are suggested: *legal interoperability* that allows the alignment of legislation so that exchanged data is accorded proper legal weight; *organizational interoperability* that allows the coordination of processes in which different organizations achieve a previously agreed and mutually beneficial goal; *semantic interoperability* that ensures that the precise meaning and formats of exchanged information are preserved and understood by all parties; and *technical interoperability* that entails formalized technical specifications be agreed upon when establishing European public services;
- (4) *Interoperability agreements*: cooperation amongst public administrations at different levels of interoperability should be formalized in interoperability agreements, containing sufficient detail to provide a European public service whilst providing each organization with autonomy;
- (5) *The governance of interoperability*: Interoperability governance covers the ownership, definition, development, maintenance, monitoring, promoting, and implementing of interoperability frameworks in the context of multiple organizations working together to provide (public) services. It is a high-level function providing leadership, organizational structures, and processes to ensure that the interoperability frameworks sustain and extend the organizations' strategies and objectives.

6.5 DCAT application profile for data portals in Europe

In order to improve interoperability in European e-government systems, a document has been prepared by the European Commission's Interoperability for European Public Administrations program. The purpose of this document (https://joinup.ec.europa.eu/sites/default/files/87/4d/c8/DCAT-AP_Final_v1.00.pdf) is to define an application profile that can be used for the exchange of descriptions of data sets among data portals.

Models for describing datasets

Asset description metadata schema (ADMS): ADMS is a vocabulary to describe interoperability assets (resources such as specifications, schemas, code lists, and software tools that facilitate interoperability) making it possible for interested parties to discover and re-use those assets.

CERIF for datasets: CERIF is a European Union recommendation that defines a data model and XML interchange format for interoperability of research information. The overall aim of CERIF for datasets is to develop a framework for incorporating metadata into CERIF such that research organizations and researchers can better discover and make use of existing datasets, wherever they may be held.

CKAN dataset schema: CKAN is a Web-based open source data management system for the distribution of data maintained by the Open Knowledge Foundation. The dataset is the central domain object in the CKAN domain model.

INSPIRE metadata schema: INSPIRE is a directive of the European Parliament and of the council aiming to establish a EU-wide spatial data infrastructure to give cross-border access to information that can be used to support EU environmental policies as well as other policies or activities having an impact on the environment. In order to ensure cross-border interoperability of data infrastructures, INSPIRE sets out a framework based on common specifications for metadata, data, network services, data and service sharing, monitoring, and reporting. Such specifications consist of a set of implementing rules. The INSPIRE metadata implementing rules include rules for the description of datasets.

Statistical data and metadata exchange (SDMX): SDMX is an initiative to foster standards for the exchange of statistical information. The specifications include an information model, XML formats and schemas, and an UN/EDIFACT format.

Vocabulary of interlinked datasets (VoID): VoID is an RDF vocabulary for expressing metadata about RDF datasets. It is intended as a bridge between the publishers and users of RDF data, with applications ranging from data discovery to cataloguing and archiving of datasets.

7 STANDARDS

One of the oldest approaches to achieving interoperability among heterogeneous parties is to agree on a standard that achieves a limited amount of homogeneity among them. The role of standards in increasing data understandability and reusability is crucial. Standards come about in different ways. A number of standards were created by committees that convened because a large and diverse enough community agreed that a standard was needed. Sometimes one product gains enough market share that it becomes a de facto standard by virtue of its broad deployment. Other times, government organizations can help a standard gain wide acceptance. The success or failure of standards, and the design philosophies underlying standardization efforts are very often determined more by social and business decisions than by technical merits.

One drawback of standards is that they are difficult to agree on and therefore often end up being complex combinations of features that reflect the interest of many disparate parties. Another fundamental problem is that a standard by its very nature infringes on the autonomy of the single parties. With a single standard, parties are no longer free to introduce local optimizations or to satisfy the preferences of different groups (Paepcke, Chang, Garcia-Molina, & Winograd, 1998).

Standardization activities characterize the different phases of the scientific data life-cycle. Several activities aim at defining and developing standards to represent scientific data, i.e., standard data models; standards for querying data collections/databases, i.e., standard query languages; standards for modeling domain-specific metadata information, i.e., metadata standards; standards for identifying data, i.e., data identification standards; standards for creating a common understanding of a domain-specific data collection, i.e., standard domain-specific ontologies/taxonomies and lexicons; and standards for facilitating the transfer of data between domains, i.e., standard transportation protocols, etc.

A big effort has been devoted to creating metadata standards for different research communities. Metadata standards vary in terms of their specificity, structure, and maturity largely because each standard has been developed on the basis of the needs of a particular user community. Given the plethora of standards that now exist, some attention should be directed to creating crosswalks or maps between the different standards.

Standardization is particularly important for the reuse of data *across distance* (Zimmerman, 2003), where the use of data outside their original context implies distance. The word distance is subject to a variety of interpretations. Most commonly, distance is used to refer to something outside the local sphere of activity. An example of this definition is the space between the assumptions and methods of one discipline and another. Distance can also exist within a community for reasons such as personal or institutional status, subspecialty, or epistemological view. Additionally, the word distance can be defined in a temporal sense. For example, there can be a time lag between the original data collection and reuse.

Standards are important because they can help to span all kinds of distance (spatial, temporal, cultural, etc.) as they have the capability to transform local knowledge into public knowledge and thus avoid epistemological differences due to distance that can lead to different interpretations of the same data.

8 CONCLUDING REMARKS

The present data-intensive multi-disciplinary science era is characterized by a data bonanza, an increasing availability of data tools and services and intensive interactions between globally distributed research teams. *This abundance demands more mediation.* There is a need for research data infrastructures that support mediating environments enabling researchers to interoperate. Data and services, in order to be discoverable, understandable, assessable, and usable, must be formally described.

The use of purpose-oriented metadata models is of paramount importance to achieve data exchangeability and usability. Data is incomprehensible and hence useless unless there is a detailed and formal description of how and when it was gathered and how the derived data was produced. Metadata information is also needed for describing the functionality of a data tool/service in order to make it discoverable and verify its compatibility with the researcher needs.

Unfortunately, while a big effort has been devoted to creating metadata standards for scientific data, little effort has been devoted to creating metadata information for data tools and services. Currently, we don't have metadata models for describing the functionality, for example, of a data mining service, or a data visualization service, or a data analysis service. The development of such metadata models is of paramount importance in order to make data tools and services discoverable and usable as open science entails open access not only to data but also to scientific analyses, data tools, and services.

Finally, policies, adopted by research organizations that allow for security, privacy, authorization, obligation, etc., must also be formally specified in a machine understandable way. The principles of open scientific data and open science can be widely accepted only if realized within an *integrated science policy framework* to be implemented and enforced by mediating infrastructures.

9 REFERENCES

- Bannon, L. & Bodker, S. (1997) Constructing Common information Spaces. *Proceedings of the Fifth European Conference on Computer-Supported Cooperative Work (ECSCW'97)*, Lancaster, UK.
- Czejdo, B., Rusinkiewics, M., & Embley, D. (1987) An Approach to Schema Integration and Query Formulation in Federated Database Systems. *Proc. of the 3rd International Conference on Data Engineering*, Los Angeles, CA, Feb 3-5). 477-484
- Davis, F. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*/September 1989.
- Gasser, U. & Palfrey, J. (2007) *Breaking Down Digital Barriers: When and How ICT Interoperability Drives Innovation*, Report of The Berkman Center for Internet & Society at Harvard Law School.
- Haslhofer, B. & Klas, W. (2010) A Survey of Techniques for Achieving Metadata Interoperability. *ACM Computing Survey* 42(2).

- Heiler, S. (1995) Semantic Interoperability. *ACM Computing Surveys* 27(2).
- Hey, T., Tansley, S., & Tolle, K. (Eds) (2009) *The Fourth Paradigm: Data Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.
- Huhns, M. & Singh M. (1998) Managing Heterogeneous Transaction Workflows with Cooperating Agents. In *Agent Technology: Foundations, Applications, and Markets*, Springer, Berlin, Germany, pp. 219-239.
- Kahng, J. & McLeod, D. (1998) Dynamic Classificational Ontologies: Mediation of Information Sharing in Cooperative Federated Database Systems. In *Cooperative Information Systems: Trends and Directions*, Academic Press, San Diego, CA, pp179-203.
- Keller, U., Lara, U., Lausen, H., Polleres, A., & Fensel, D. (2005) Automatic Location of Services. *Proceedings of the 2nd European Semantic Web Conference (ESWC)*, Heraklion, Greece.
- Kim, K. (1999) *The Development of Interoperable Mediation System Architecture: An Ontology and Mediator-based Approach*. Master's Thesis. Dept. of Management Information Systems, The University of Arizona, Tucson, AZ.
- Kolaitis, P. G. (2005) Schema Mappings, Data Exchange, and Metadata Management. *Proceedings of the Twenty-fourth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, Baltimore, Maryland, USA: ACM, pp 61-75.
- Krishnamurthy, R., Litwin, W., & Kent, N. (1991) Language Features for Interoperability of Databases with Schematic Discrepancies. *Proceedings of the 1991 ACM SIGMOD International Conference on Management of Data*, Denver, CO, May 29-31.
- Ludascher, B., Gupta, A., & Martone, M. (2001) Model-based Mediation with Domain Maps. *Proceedings of the 17th International Conference Data Engineering*, Heidelberg, Germany.
- McLeod, D. & Si, A. (1995) The Design and Experimental Evaluation of an Information Discovery Mechanism for Networks of Autonomous Database Systems. *Proceedings of the 11th IEEE International Conference on Data Engineering*, Taipei, Taiwan, March 6-10, pp 15-24.
- March, S., Herver, A., & Ram, S. (2000) Research Commentary: An Agenda for Information Technology Research in Heterogeneous and Distributed Environments. *Inform. Syst. Res.* 11(4), pp 327-341.
- Moran, M. & Mocan, A. (2005) Towards Translating between XML and WSML based on mappings between XML Schema and an equivalent WSMO Ontology. *Proceedings of the WIW 2005 Workshop on WSMO Implementations*, Innsbruck, Austria.
- Nativi, S., Craglia, M., & Pearlman, J. (2013) Earth Science Infrastructures Interoperability: The Brokering Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6(3).
- Osterlund, C. & Carlile, P. (2005) Relations in Practice: Sorting Through Practice Theories on Knowledge Sharing in Complex Organizations. In *The Information Society*.
- Paepcke, A., Chang, C-C., Garcia-Molina, H., & Winograd, T. (1998) Interoperability for Digital Libraries: Problems and Directions. Retrieved September 5, 2014 from the World Wide Web: <http://ilpubs.stanford.edu:8090/317/1/1998-24.pdf>
- Papakonstantinou, Y., Garcia-Molina, H., & Widom, J. (1995) Object Exchange Across Heterogeneous Information Sources. *Proceedings of 11th International Conference on Data Engineering (ICDE)*, Taipei, Taiwan.
- Park, J. & Ram, S. (2004) Information Systems Interoperability: What Lies Beneath? *ACM TOIS* 22(4), pp 595-532.

- Rahm, E. & Bernstein, P. A. (2001) A survey of Approaches to Automatic Schema Matching. *The VLDB Journal - The International Journal on Very Large Data Bases* 10(4), pp 334-350.
- Sciore, E., Siegel, M., & Rosenthal, A. (1994) Using Semantic Values to Facilitate Interoperability Among Heterogeneous Information Systems. *ACM Trans. Database Syst.* 19(2), pp 254-290.
- Spalazzese, R., Inverardi, P., & Issarny, V. (2009) A Formalization of Mediating Connectors: Towards on the Fly Interoperability. *Proc. Joint Working Conference IEEE/IFIP Conference on Software Architecture (WICSA/ECSA)*.
- Stollberg, M., Cimpian, E., Mocan, A., & Fensel, D. (2006) A Semantic Web Mediation Architecture. *Proceedings of the 1st Canadian Semantic Web Working Symposium (CSWWS)*.
- Thanos, C. (2013) A Vision for Global Research Data Infrastructures. *Data Science Journal* 12, pp 71-89.
- Ullman, J. (1997) Information Integration Using Logical Views. *Proceedings of the International Conference on Database Theory*, Delphi-Greece.
- Wang, X. (2005) On DRM Interoperability and Compatibility, Workshop on Digital Rights Management Impact on Consumer Communications. *Consumer Communications and Networking Conference*, Las Vegas, NV, USA
- Wegner, P. (1996) Interoperability. *ACM Computing Surveys* 28(1).
- Wiederhold, G. & Genesereth, M. (1997) The Conceptual Basis for Mediation Services. *IEEE Expert* 12(5).
- Wiederhold, G. (1994) Interoperation, Mediation, and Ontologies. *Proc. International Symposium on Fifth Generation Computer Systems (FGCS94)*, Workshop on Heterogeneous Cooperative Knowledge-Bases, Vol. W3, ICOT, Tokyo, Japan.
- Wiederhold, G. (1992) Mediators in the Architecture of Future Information Systems. *IEEE Computer*, pp 33-49.
- Zimmerman, A. (2003) *Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists*. Retrieved September 2, 2014 from the World Wide Web: <http://deepblue.lib.umich.edu/handle/2027.42/39373>.

(Article history: Received 24 April 2014, Accepted 3 September 2014, Available online 11 September 2014)