Standards for Databases on the Grid

Susan Malaika
IBM Silicon Valley Laboratory
San Jose, CA
malaika@us.ibm.com

Andrew Eisenberg
IBM, Westford, MA 01886
andrew.eisenberg@us.ibm.com

Jim Melton
Oracle, Sandy, UT 84093
jim.melton@acm.org

Guest Column Introduction

Relational databases supported applications in a centralized environment in the 1960's and 1970's. They progressed to a client/server environment in the 1980's. The 1990's saw application servers with a multi-tiered architecture, in most cases supported by an RDBMS. Most recently we have seen the emergence of XML, XML storage in DBMS's, navigation within an XML document via XPath, and the XQuery query language for XML.

In this article, Susan provides an introduction to the Grid and describes how databases will be used in this new environment. The Global Grid Forum (GGF) is producing technical specification to enable both Relational and XML databases to be located, accessed, and replicated in this environment. They make use of a variety of existing an emerging database, file, networking, and web services standards. Susan is a Senior Technical Staff Member at IBM. She is a member of the GGF DAIS (Data Access and Integration Services) working group.

Jim Melton and Andrew Eisenberg

1. Introduction

"GRID TECHNOLOGIES are infrastructure, which doesn't make for compelling reading—after all, when infrastructure works, it is as invisible as the sewage system." So said Ian Foster in an article on grid computing in Scientific American [22]. The objective of the sewage infrastructure is to manage waste disposal without requiring families to run their own sewage processing plants or to be familiar with disposal details. The same is true for electricity, wireless, and water infrastructures. The word 'grid' in grid computing was adopted from the electrical power grid usage, representing a widely available shared resource accessible through standard interfaces.

The goal of grid technology is to provide shared computing resources, without consumers knowing the details of the computing components, where they are located, or how to manage, fix or upgrade them. The grid infrastructure connects hardware, storage, and software across high-speed networks providing flexible and secure access to applications and data. Currently, grid technology is used in scientific communities mostly at universities and research laboratories to perform major computations and analyses. However the use of grids and associated technology is likely to proliferate into other communities.

In order to support any significant infrastructure over heterogeneous resources, standardization of the interfaces between the infrastructure, the managed resources and their consumers is essential. In 1998, a group was formed, which two years later became the Global Grid Forum (GGF), to standardize grid interfaces [3].

2. Global Grid Forum (GGF)

The GGF is a young organization compared with ANSI (1918) [1], ISO (1946) [24], IETF (1986) [23], and OMG (1989) [26]. It is even younger than recently formed organizations such as Oasis (1993) [26] and W3C (1994) [29]. GGF models itself on IETF (Internet Engineering Task Force) which is open to all interested individuals who pay an annual or conference fee in order to participate in working groups at face to face meetings and on mailing lists. The objective of the groups is to produce documents describing best practices, technical specifications, user experiences, and implementation guidelines for distributed and grid computing environments, with a particular emphasis on interoperability and technology reuse. The focus of GGF is on the production of technical specifications that become grid recommendations after a review process.

2.1 GGF Process

There is no formal voting. The general approach for disputed topics is that the group has to reach "rough consensus". Occasionally, when there is a lot of discussion, a chairperson will use humming (a procedure adopted from the IETF where the degree

of loudness of the participants' hum indicates the group's overall opinion) or a show of hands to get a sense of the viewpoint of the meeting. On a mailing list, a chairperson may ask for e-mails taking a position, in order to reach consensus.

There are four categories of GGF documents: Informational Documents, Experimental Documents, Community Practice Documents, and Recommendations Documents. Technical specifications go through the recommendation track in three stages, which are reached as follows:

- *Proposed Recommendation:* A *Proposed Recommendation* must be submitted by a GGF working group.
- *Draft Recommendation:* Comments must be resolved within a 60-day public comment period after submitting a *Proposed Recommendation* for it to reach *Draft Recommendation*. At the time of writing there are two GGF *Proposed Recommendations* in their 60-day public comment period.
- *Grid Recommendation:* At least two interoperable implementations must be demonstrated within 24 months of completing a *Draft Recommendation* in order for a document to reach *Grid Recommendation* status. No GGF documents have reached this stage yet.

2.2 GGF Organization

The GGF is organized into seven topic areas:

- GGF **APME** (Applications Programming Models and Environments Area) [4]
- GGF **ARCH** (Architecture) [5]
- GGF **DATA** (Data) [11]
- GGF **ISP** (Information Systems and Performance) [14]
- GGF **P2P** (Peer to Peer) [19]
- GGF **SEC** (Security) [20]
- GGF **SRM** (Scheduling and Resource Management) [21]

Within each of these areas, there are GGF Working Groups and GGF Research Groups. Typically, each GGF Working group is focused on producing one technical specification, whereas GGF Research Groups produce a variety of documents and run workshops. At the time of writing there are about 50 groups overall across all areas. New groups are created at the three major GGF meetings each year. The most recent GGF meeting was GGF8, which took place in Seattle in June 2003, with about 700 participants. A number of working groups hold conference calls and face-to-face meeting in addition to the three major meetings. The meeting format is likely to change in 2004 in response to the needs of the working groups. For more information on the GGF, see [25].

2.3 Architectural Context

In the GGF Architecture Area, the Open Grid Services Architecture Working Group (**OGSA-WG**) [15] is gaining the attention of other GGF groups, including groups in the GGF Data Area. The objective of the **OGSA-WG** is to identify the following:

- The components that make up the infrastructure for a distributed computing environment for the grid. The components are described as services. A service describes a network-enabled entity that provides a particular function such as accessing a file or tracing or logging changes.
- The basic mechanisms, expressed as Web Services, which must be supported by grid components in order to form an integrated distributed grid system, such as a method of event notification. The mechanisms are defined in a separate specification produced by the Open Grid Services Infrastructure Working Group (**OGSI-WG**) [17]

The OGSI mechanisms concern services that have state: long term information that is maintained across client requests. Conventions for performing a number of service related activities appear in the OGSI specification. Examples include ways of:

- Referring to an instance of a service (a handle)
- Forming and referring to collections of service instances as a whole (useful when creating a registry of available services)
- Starting up a service instance (a factory)
- Accessing a service's state (which is expressed in XML and is often referred to as service data)
- Being notified when the state of a service changes
- Managing service lifetimes (which is very important in a distributed fault tolerant environment)
- Defining Web services interfaces with inheritance

OGSI compliant grid components must expose Web Services interfaces that conform to some of the constructs in the OGSI specification, e.g., factory. An objective of these interface descriptions is that they do not impose a programming language, a programming model or a particular operating system, because they focus on communication protocols and not on APIs. For more information see [28].

3. Data-related Groups in GGF

Topics that concern data in the grid environment can be classified into the followed broad categories:

- ***Virtualizing*** data in the grid
- ***Managing*** data in the grid
- ***Integrating*** data within the grid infrastructure

Virtualizing data means applications are able to discover, access and update data irrespective of the format of the data and where the data is located in the grid. Diverse features and interfaces play a part in data virtualization, including caching, replication, data placement, schema mappings, and format mappings. Some standards activities are underway in support of data virtualization including the **DAIS-WG** [8], the **OREP-WG** [18], and the **DFDL-WG** [7].

Managing data means making sure data is available to applications with appropriate performance characteristics in the grid environment. Thus, data resources must play their part in the overall quality of service and provisioning mechanisms of a scalable grid system. Little standardization work has been done in this area yet, although initial possibilities are described in the "SRM" section in this article.

Integrating data with the grid infrastructure means making sure that data management systems support the mechanisms of the grid. Examples include ensuring that data resource systems play their part in carrying necessary context information, delivering appropriate tracing, monitoring, debugging, accounting information, so they form an integral part of the grid. Little database-specific standardization work has been done yet in this area, although the beginning of such activities are described in the "Database Management Characteristics" section in this article.

Active working groups in the GGF Data Area include **DAIS-WG**, **OREP-WG**, **DFDL-WG**, **GRIDftp-WG** [13] and the very recently formed **GFS-WG**. Birds of a Feather session take place in order to determine whether new groups are needed, e.g., **DAMO-RG** [10]. The **RG** suffix denotes a Research group. The remainder of this section will describe some of these groups.

3.1 DAIS-WG

The **DAIS-WG** (Data Access and Integration Services Working Group) has been in existence for about a year. Its objective is to produce a specification [9] that describes a service-based interface for accessing and integrating data in existing relational and XML databases on the grid.

Features are included that are very useful in grid applications, and which can be implemented outside the database systems, without modifying the database system itself, if they are not already supported. So far the requirements have been collected from the scientific communities in the main. These features include:

- Naming results for subsequent use
- Multiple result formats
- Chunking of large quantities of data
- Asynchronous delivery of results
- Delivery of results to a third party

The DAIS specification has evolved over the last few months in response to comments. Interest in the specification is growing, as demonstrated by the increasing numbers of participants at both the face-to-face working meetings and the specification status presentations (over 60 participants).
DAIS is the first specification of a resource manager trying to adhere to the principles and infrastructure described in OGSI.

The DAIS specification as submitted to GGF8 includes the following constructs in the model:

- **Data Resource Manager**: represents the database management system.
- **Data Resource**: represents a database (tables for a relational database or collections for an XML database)
- **Data Access Session**: represents the relationship between a client and a data resource
- **Data Request**: contains the SQL, XPath or XQuery request to be executed
- **Data Set**: represents the output result format

The **Data Resource Manager**, **Data Resource** and **Data Access Session** are all **OGSI** compliant services. The **Data Request** and the **Data Set** are data formats. Here is a small sample scenario:

- Create **Data Resource Manager** service instance
- Create **Data Resource** service instance
- Create **Data Access Session** service instance
    - Iterate
        - Issue **Data request** delivering output to third party
    - EndIterate

The **DAIS-WG** operates in terms of topics, to enable individuals or small groups to focus on specific sections of the specification and to explore related areas. These topics include:

- *The DAIS model*: which continues to be the source of very animated discussions and thus the model outlined briefly above is subject to change.
- *Transformations*: which enables the transformations of results. There is discussion whether *Transformations* have a place in the DAIS specification.
- *Stored Procedures:* which describes how parameters and result sets should be handled.
- *Security:* which describes how database and grid security interact. Much of this effort is dependent on other standards activities for Web Services. It should be noted that all GGF specifications must include a security section
- *Transaction:* which describes how database and transactions interact in the grid environment. Like security, much of this effort is dependent on other standards and coordination activities for Web Services
- *Metadata:* much of the section entitled "Related Standards Activities outside the GGF" concerns metadata issues for **DAIS**

Multiple drafts of the **DAIS** specification have been produced:

- The first draft focused on various data delivery styles
- The second draft focused on conformance with OGSI
- The third draft emphasized the underlying model for grid data access
- The fourth draft will be submitted to GGF9 in October 2003

3.2 OREP-WG

The recently-formed **OREP (**OGSA Replication Services) group started out as the Replication research group exploring data replication technologies for the grid.

The original interest of the research group was in replicating large files ranging in size from gigabytes to petabytes. With the formation of the **OREP-WG,** specifications for the interfaces for data replication will be produced that apply to files and databases. The first **OREP** specification will focus on defining the interfaces for a replica location service (RLS). The RLS is an essential ingredient for data virtualization. It should be possible for applications to access data with the desired performance characteristics, without knowledge of the location of the data. The RLS keeps track of where copies of files and databases are, and their freshness.

In the scientific community, most of the data is read only (scientists don't update the results of their experiments, they create new result sets). In the commercial database world, data is can be modified (withdrawing money from a bank account typically causes the current balance to be changed). Thus, it is often necessary to keep feeding changes into replicas to keep them synchronized with the master copy. In some systems, there is no recognized master copy. In this case applications can update any copy, and changes have to be fed between all replicas.

An area of intersection of the **OREP-WG** with the **DAIS-WG** is the way data is moved. In the DAIS specification, there is a desire to describe the movement of results of queries to third parties. The mechanisms for moving results to third parties may also be appropriate for moving data between replicas to keep them synchronized. Thus, Data Movement may become a separate topic for the OREP-WG. No proposed recommendations have been submitted to GGF yet from **OREP-WG** as the group was formed recently.

3.3 DFDL-WG (pronounced Daffodil)

The recently formed **DFDL** (Data Format and Description Language) working group's objective is to define a general language for describing and labeling the structure of data streams. A data stream is a sequence of bits that can represent text or binary data. The labeling of structures allows the attachment of information for better understanding the encoding and meaning of the bits in the structure, e.g., decimal, big endian, miles per hour, or department number. The motivation for the notation is the large bodies of data that will never be annotated as XML, although **DFDL** itself will be an XML based language. Thus, **DFDL** will make it possible to manipulate arbitrary un-annotated data streams programmatically without pre-planning.

The **DFDL** group members are reviewing related existing languages such as BinX (Binary XML) Although most of the interest in **DFDL** is from the scientific and data preservation communities, **DFDL** may be of interest to the relational database community in a number of ways. For example:

- To describe the layout and content of a database load format
- To describe a compact stream representation of a relational result set

An area of intersection with the **DAIS-WG** is the possible use or recognition of data streams described by **DFDL** as an output format for queries.

An initial draft document of the ***DFDL*** specification was presented at GGF8.

3.4 GridFTP-WG

The established GridFTP group concentrates on providing a specification for a reliable file transfer protocol, e.g., re-startable, that is built on the ftp protocol defined in the IETF. GridFTP is likely to play a part in any Data Movement standardization activity, mentioned in section 3.2.

3.5 GFS-WG

The **GFS** (Grid File Systems) group will have its first meeting at GGF9 in October 2003. At the time of writing, the group is preparing its charter.

4. Other GGF groups with impact on the GGF DATA area

In this section, we review some of the standards activities in the GGF, outside the data area that are relevant to the standardization work for databases in a grid environment.

In section 2, we described how the **OGSA-WG** and **OGSI-WG** groups in the **ARCH** (Architecture) area provide a context for many of the database activities.

4.1 SRM

The **SRM** (Scheduling and Resource Management) area is particularly relevant to data on the grid. To deploy major grid systems, it is necessary to adjust service behaviors according to overall system, application or organizational goals, e.g., the interactions of the company's vice president with the grid must receive a response within 0.5 second. To support such requirements, a specification proposing an agreement-based model was submitted to the **GRAAP-WG** (Grid Resource Allocation Agreement Protocol Working Group) [12] in GGF8.

Agreements form the basis for negotiation between management applications (applications that monitor and manage grid applications) and management services with which the management applications negotiate agreements. The role of a management service need not be separate. It can be adopted by an application or by a resource manager.

A re-negotiation might take place, for example if a grid application or resource is unable to satisfy the response time goal for vice presidents, which had been agreed. The re-negotiation may imply other re-negotiations take place with various service managers, applications or resource managers in attempt to achieve the required goals.

It is likely that database systems in a grid environment would plug into an agreement framework, and database system-specific agreement notations will be defined, e.g., data freshness for vice presidents must be within 5 minutes.

4.2 SEC

Activities in the **SEC** (Security) area are also relevant to databases. Securing data is an important requirement and is especially critical in shared environments such as the grid. Key features to be considered when securing access to data include:

- **Establishing the identity** (e.g., the person) that is accessing the data in order to enforce access policies based on that identity.
- **Ensuring data isolation** in grid environments, where resources are shared within a data center, and multiple services access shared data from different organizations. An environment that hosts multiple grid services must ensure that an application will not be able to access data associated with another application executing within the same environment.
- **Enforcing data privacy** such that only users who have rights to view particular data (e.g., medical records) can view such data.
- **Supporting identity federation** particularly when federating data from different sources, when it is likely that these data sources belong to different security domains.

In order for these requirements to be met across a variety of data sources that may use different security mechanisms, it is important to adopt interfaces that allow interoperable and extensible security architectures. The **OGSA-SEC-WG** (OGSA Security Working group) at GGF, is examining the Web Services security for use in a grid environment. Given OGSA is based on Web Services, Web Services security is equally applicable to grid environments. Web Services security specifications and infrastructure are being defined in the industry. These are some relevant industry initiatives (where the **WS** prefix denotes Web Services):

- **WS-Security** to provide message level security
- **WS-Policy** to specify policies for accessing and modifying data sources and services
- **WS-Trust** to identify and authenticate users in a trusted way
- **WS-Federation** to manage and broker trust relationships including federated identity

Additional extensions such as intrusion detection are being defined in the **OGSA-SEC-WG** [16]. As mentioned in the DAIS section earlier, there is a security topic in the **DAIS-WG** that is likely to involve collaboration with **OGSA-SEC-WG.** In summary, secure access to data will be based on standards and technologies defined in the relevant web services and grid services groups, together with data access mechanisms supported by data management systems.

5. Related Standards Activities outside the GGF

In this section, we review some of the standards outside the GGF that are of interest to the standardization work for databases in a grid environment.

The grid community is defining unified service-based interfaces. Some of the components already have standards associated with them, but are specific to a particular programming language, or operating environment, or notation. It is not surprising that interactions are beginning to take place between the members of the GGF and other standards groups that focus on a particular domain.

Examples of relationships between the GGF and other groups include the collaboration with the Distributed Management Task Force (DMTF) group that defines a model for real world managed resources, including a model for managing databases [2]. The DTMF, with collaboration from GGF, will very likely provide an XML-based interface definition for real world managed resources. Unification of management interfaces is important for supporting systems that can manage themselves end-to-end in heterogeneous grids.

Another area where there is collaboration concerns some of the W3C groups, such as in defining WSDL 1.2 in order to incorporate features useful in the grid environment that are currently defined in **OGSI**.

Over time, it is anticipated that there will be more cooperation as many of the concepts and interfaces being introduced for grid computing are useful in a general distributed computing setting. The strategy of the GGF is to develop relationships with other standards bodies, to identify opportunities for cooperation and possible areas of overlap.

For the remainder of this section we'll focus on some of the database-related standards that relate to constructs in the **DAIS** specification as submitted to GGF8. As amply illustrated in the following sections, there is an absence of agreed formats, such as XML, for representing basic information concerning relational databases and content. These formats are needed in many environments beyond grid and distributed systems environments, such as in development tools, and administration tools.

5.1 Database Management System Characteristics

The DAIS **Data Resource Manager** has service data that corresponds to any executing resource manager, For example:

- ***ExternalResourceManager:*** product name, version and vendor name
- ***PhysicalProperties***: physical constructs; includes space used, buffers
- ***SoftwareCapabilities***: capabilities of the software; includes software version, service packs, patch level.
- ***FeaturesInstalled***: description of installed features; includes product enhancements, features, extra functionalities

In order to express this data in an OGSI compliant way, XML schema definitions are required. The intention of the DAIS group is to collaborate with CGS-WG (CIM Based Grid Schema Working Group) [6] and the DMTF to define service data for DAIS ***Data Resource Managers.***

5.2 Database Characteristics

The DAIS specification refers to ***Data Resources.*** ***Data Resources*** are of two kinds:

- ***Relational Data Resources:*** where the content is stored in relational tables. Results of a SQL request are returned in a result set.
- ***XML Data Resources:*** where the content is stored in XML collections. Results of an XQuery requests are returned as a sequence of items.

To represent the structural aspects of ***Relational Data Resources*** as service data, descriptions of databases, tables, columns, data types, constraints, triggers, stored procedures etc are required in XML. The SQL standards do not provide such descriptions in XML. However, the following two standards may assist in providing such representations:

- SQL Information Schema – Part 11 (SQL: 2003): It provides descriptions of over 60 relational views of database catalog information. These views are comprehensive and include views for representing: COLUMNS, REFERENTIAL_CONSTRAINTS,

ROUTINES, SQL_FEATURES, SQL_LANGUAGES, and TABLES.

- SQL/XML – Part 14 (SQL 2003): which provides a default XML tagging for relational data. To see the XML Schema defined by this part of SQL go to: http://standards.iso.org/iso/9075/2002/12/

By combining the default XML tagging to represent the relevant views, the appropriate service data can be created. Currently, the DAIS specification has simple service data structures to represent constructs such as tables, based on the availability of the information through existing database interfaces such as ODBC and JDBC. For example, here is a snippet of a schema from the DAIS specification to represent service data for a relational table definition:

```
<xs:complexType name="tableDefinition">
<xs:sequence>
    <xs:element name="column"
    type="dais:columnDefinition"
    minOccurs="1" maxOccurs="unbounded"/>
    <xs:element name="primaryKey"
    type="dais:primaryKeyDefinition"
    minOccurs="0" maxOccurs="1"/>
    <xs:element name="foreignKey"
    type="dais:foreignKeyDefinition"
    minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="key"
    type="dais:foreignKeyDefinition"
    minOccurs="0" maxOccurs="unbounded"/>
</xs:sequence>
    <xs:attribute name="name" use="required"
    type="xs:string"/>
</xs:complexType>
```

To represent the structural aspects of ***XML Data Resources*** as service data, descriptions of collections and associated schemas are required. There is no standardization work in the area of XML collection descriptions, so the DAIS specification currently includes minimal service data in this area. Here is a snippet of a schema to describe an XML Collection:

```
<xs:complexType name="CollectionType">
  <xs:sequence>
    <xs:element name="collection"
    type="CollectionType" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
     <xs:attribute name="name" type="xs:Name"
    use="required"/>
</xs:complexType>
<xs:complexType name="CollectionSchemaType">
<xs:sequence>
    <xs:any namespace="http://www.w3.org/
    2001/XMLSchema"/>
 </xs:sequence>
    <xs:attribute name="collectionName"
    type="xs:Name" use="required"/>
</xs:complexType>
```

5.3 Database Content

For representing relational database content, in the context of relational tables, the ISO: SQL/XML – Part 14 (SQL 2003) specification contains a schema that can be used. See:

http://standards.iso.org/iso/9075/2002/12/

There is no agreed notation for representing XML documents in the context of the collections they are stored in. The DAIS specification has not attempted such a definition yet, although it may required in future iterations of the DAIS specification.

5.4 Query Results

There are no agreed XML schemas to represent the results of a relational query in XML, using a single schema that is independent of table and column names, outside the context of a particular programming environment. One example is the WebRowset that is part of JCP JSR 114. WebRowset is planned for inclusion in JDK 1.5. Mentioning the WebRowset schema in the DAIS specification caused much discussion. It was assumed that the schema itself was dependent on Java, and hence it was thought that the DAIS specification had introduced dependencies on the Java environment. Similar concerns arose for ADO.NET whose format is also mentioned in the specification, and which does not have a published schema through a standards community yet.

For both ADO.NET and WebRowset, the DAIS specification is exploring the support of the associated schemas without introducing dependencies on the corresponding programming environments. For the WebRowset schema file, see http://java.sun.com/xml/ns/jdbc/webrowset.xsd

There is no agreed format for representing an XPath or XQuery sequence result in XML yet. It is possible that a W3C XQuery group or the JCP JSR 225 group might define a suitable notation. (JSR 225 is XQJ: an XQuery API for Java.)

5.5 Query Requests

The DAIS specification includes a data request document for client applications to submit one or more queries. Query languages of interest are SQL, XPath, and XQuery.

Currently there is no agreed definition of an XML representation of SQL requests. Thus, we see a variety of XML formats for SQL requests, with diverse notations to represent parameters, outside the grid community already. These notations do not interoperate. Here is an example of a representation without parameters from the DAIS specification [9]:

```
<sqlQueryStatement name="select">
    <expression>
    select * from employee where salary &gt;
    50000 fetch first 5 rows only
    </expression>
<webRowSetStream name="dbanswer"/>
</sqlQueryStatement>
```

The W3C's XML Query Working Group is preparing a specification for an XML syntax for the XQuery language; this XML syntax is known as XQueryX. The notation will be very helpful to the DAIS specification work.

6. Directions for Grid Standards

These are some directions for grid standards in general:

- ***OGSA***: An increasing number of OGSA-related working groups are likely to appear in GGF, as the architecture gains recognition and momentum.
- ***Agreements and policies****:* Standards relating to end-to-end grid management, quality of service and provisioning will gain impetus
- ***Existing standards***: In general the grid community will continue to work with existing systems and standards, and to introduce new components and interfaces only where absolutely necessary. Reliance on standards that are not widely implemented or difficult to implement will continue to be avoided.
- ***Collaboration***: Collaboration with other standards groups will continue to increase.
- ***Increased Interactions with Web Services activities***: Context for Transactions and Security and Quality of Service are examples of features that will be required to create integrated and manageable grid environments. These kinds of features are likely to adopted from the Web services activities in these areas rather than be invented solely for the grid.

These are some likely directions in the area of data and grid computing:

- ***Files and Databases:*** Requirements are emerging to treat file data and database data as a whole, so the features and interfaces that are relevant to data in general can be described and developed in a unified way. For example, it is likely that the DAIS specification will describe ***Data Resources*** that represent files and files systems.
- ***Adapted database and file systems:*** The standards activities so far have focused on enabling existing databases, file systems and large bodies of data to the grid. In the future, interfaces in support of systems that are adapted to the grid environment are likely to be standardized.
- ***More groups in the GGF DATA area:*** Technologies that are common to all types of data include metadata, annotations, federation, discovery, movement, lineage. It is likely that new groups will form to standardize these areas that are particularly important to grid applications.
- ***Collaboration:*** So far, there has not been strong collaboration between the standards groups in the GGF data area with non-GGF standards groups concerned with data. There is much potential for collaboration as illustrated in section 5 in this article.

7. Thanks

Many thanks to Mario Antonioletti, Andrew Eisenberg, Shannon Hastings, Amy Krause, Steve Langella, Allen Luniewski, James Magowan, Jim Melton, Nataraj Nagaratnam, Inderpal Narang, Mary Ratcliff, Ellen Stokes, Martin Westhead, and Kathryn Zeidenstein for comments on this article.

8. Web References

[1] American National Standards Institute (ANSI): http://www.ansi.org/
[2] DMTF and GGF Liaison work http://www.dmtf.org/download/about/GGF-DMTFWorkRegister.pdf
[3] Global Grid Forum (GGF) http://www.ggf.org/
[4] GGF **APME** (Applications Programming Models and Environments) Area: http://www.ggf.org/7_APM/APS.htm
[5] GGF **ARCH** (Architecture) Area: http://www.ggf.org/5_ARCH/arch.htm
[6] GGF **CGS-WG** (CIM Based Grid Schema Working Group) http://www.isi.edu/~flon/cgs-wg/
[7] GGF **DFDL-WG** (data format definition language) http://forge.grihdforum.org/projects/dfdl-wg/
[8] GGF **DAIS-WG** (data access and integration) http://www.gridforum.org/6_DATA/dais.htm
[9] GGF **DAIS specification** as submitted to GGF8 http://forge.gridforum.org/docman2/ViewProperties.php?group_id=90&document_content_id=232
[10] GGF **DAMO-RG** (data models research group in formation) http://www.zib.de/ggf/data/damo/dm.html

[11] GGF **DATA** (Data) Area: http://www.gridforum.org/6_DATA/data.htm
[12] GGF **GRAAP-WG** (Grid Resource Allocation Agreement Protocol Working Group) https://forge.gridforum.org/projects/srm
[13] GGF **GRIDFTP-WG** (grid-ftp) http://www.gridforum.org/6_DATA/gridftp.htm
[14] GGF **ISP** (Information Systems and Performance) Area: http://www.ggf.org/1_GIS/GIS.htm
[15] GGF **OGSA-WG** (Open Grid Services Architecture) http://www.ggf.org/ogsa-wg/
[16] GGF **OGSA-SEC-WG** (ogsa-security) https://forge.gridforum.org/projects/ogsa-sec-wg
[17] GGF **OGSI-WG** (Open Grid Services Infrastructure) http://www.gridforum.org/ogsi-wg/
[18] GGF **OREP-WG** (ogsa replication Services) http://www.isi.edu/~annc/orep.htm
[19] GGF **P2P** (peer to peer) Area: http://www.ggf.org/4_GP/P2P.htm
[20] GGF **SEC** (security) Area: http://www.ggf.org/2_SEC/SEC.htm
[21] GGF **SRM** (Scheduling and Resource Management) Area: http://www.ggf.org/3_SRM/srm.htm
[22] The Grid: Computing Without Boundaries, Ian Foster, April, 2003, Scientific American http://www.sciam.com/
[23] Internet Engineering Task Force (IETF): http://www.ietf.org/
[24] International Standards Organization (ISO) http://www.iso.ch/
[25] Standards for Grid Computing: Global Grid Forum, Charlie Catlett, Journal of Grid Computing, Volume 1, No 1, 2003 http://www.kluweronline.com/issn/1570-7873/
[26] Object Management Group (OMG): http://www.omg.org/
[27] Oasis: http://www.oasis-open.org/
[28] The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, I. Foster, C. Kesselman, J. Nick, S. Tuecke http://www.globus.org/research/papers/ogsa.pdf
[29] World Wide Web Consortium (W3C): http://www.w3.org/