

Universität Koblenz-Landau

Seminar Semantic Grid

WS 04/05

Ideen und Prinzipien des Semantic Grid

Prof. Dr. Steffen Staab
Betreuung des Seminars:
Dipl.-Inform. Bernhard Tausch

von

Sascha Stumm
Matrikelnr.: 200210202
sstum@uni-koblenz.de
11.02.2005

Inhaltsverzeichnis

1	Semantic Grid eine Vision des E-Science	3
1.1	Grid Computing	3
1.2	E-Science	3
1.3	Semantic Grid als zukunftsweisende Technologie	4
2	Die Entstehung von Semantic Grid	4
2.1	Begriffliche Erklärung	4
2.2	Entstehung des Wissenschaftsbereiches	4
2.3	Probleme im Grid Computing	5
2.4	Die Idee des Semantic Grid	5
3	Semantic Web	6
3.1	Nutzung von Semantic Web Technologien	6
3.2	RDF Metadaten	7
3.3	OWL	7
3.4	Semantic Web Services	8
4	Service orientierte Sichtweise	8
4.1	Web-Services	9
4.2	Agentenbasierte Sichtweise	10
5	Architektur des Semantic Grid	11
5.1	Knowledge Grid	11
5.2	Knowledge Generating Services	12
5.3	Knowledge Based Services	12
5.4	Grid Knowledge Services	12
6	MyGrid	13
6.1	Programmbeschreibung	13
6.2	Aufbau	13
7	Aktueller Status	15
7.1	Forschungsstand	15

1 Semantic Grid eine Vision des E-Science

1.1 Grid Computing

Mitte der 90er Jahre gab es große Anstrengungen im Bereich des Grid Computing. Die in einem Grid enthaltenen Computer und Instrumente teilen sich nicht nur das Wissen und Informationen, sondern auch ihre Rechenleistung und Ressourcen. Dabei stand die Kombination und Bereitstellung von Ressourcen zur Steigerung der Rechenleistung, durch den Verbund von Computern und die Bildung von Clustersystemen im Vordergrund. Aus dieser Sicht können Computer, wissenschaftliche Instrumente, Ressourcen und Daten als Werkzeuge bzw. Dienstleistungen angesehen werden, welche über ein Netzwerk zur Verfügung gestellt sind.

Um den rasant wachsenden Bereich des Grid Computing mit seinen Millionen von Komponenten weiter skalierbar halten zu können, müssen neue Technologien entwickelt werden. Das eigentliche Problem, welches dem Ansatz des Grid- Computing unterliegt, ist die Koordination der Ressourcenteilung, sowie die Bildung von dynamischen, multiinstitutionellen virtuellen Organisationen.

1.2 E-Science

e-Science offers a promising vision of how computer and communication technology can support and enhance the scientific process. It does this by enabling scientists to generate, analyse, share and discuss their insights, experiments and results in a more effective manner[10]

Die Vision des E-Science bietet hierfür einen zukunftsweisenden Ansatz, wie Computer und Kommunikationstechnologie den wissenschaftlichen Prozess unterstützen und steigern kann. Man möchte es den Wissenschaftlern ermöglicht, Analysen zu generieren, diese zu teilen, zu diskutieren und wieder zu verwenden. Experimente und Resultate sollen in einem kurzen zeitlichen Rahmen realisierbar sein, für die man vorher Jahrzehnte benötigt hat. Es existiert jedoch eine große Lücke zwischen den Anstrengungen und der eigentlichen Vision des E-Science. Man verfolgt das Ziel

- der nahlosen Automatisierung
- der Kombination von Berechnungsressourcen
- einfache Bedienbarkeit
- Bildung von flexiblen Kollaborationen

auf einer globalen Ebene

1.3 Semantic Grid als zukunftsweisende Technologie

Um diese Lücke zu schließen und die Idee des E-Science voll auszuschöpfen zu können, werden in vielen Bereichen der Wissenschaft neue zukunftsweisende Technologien entwickelt. Ein sehr viel versprechender Ansatz bietet der des Semantic Grid. Es wird als eine Erweiterung des bestehenden Bereiches des Grid Computing angesehen, in dem Informationen, Services und Dienstleistungen wohl definiert existieren. D.h. es wird eine Verbesserung der Zusammenarbeit zwischen Mensch und Computer ermöglicht.

Um den Ansatz ausreichend fördern zu können, wurde der Forschungsbereich des Semantic Grid in das E-Science Programm aufgenommen. Es ist Teil des mit 500 Millionen US-Dollar finanzierten 5 Jahresplanes des UK E-Science Programms, welches darüber hinaus 100 weitere separate Projekte finanziert.

2 Die Entstehung von Semantic Grid

2.1 Begriffliche Erklärung

Der Begriff wurde 2001 von David De Roure, Nicholas R. Jennings und Nigel R. Shadbolt in 'Research Agenda for the Semantic Grid' und 'The Semantic Grid: A Future e-Science Infrastructure' zum ersten Mal eingeführt. Es existiert jedoch keine vollständige Übereinkunft darüber, wie man den Begriff des Semantic Grid genau eingrenzt. Einige Forschungsgruppen legen den Schwerpunkt auf die Konvergenz von Semantic Web und Grid Technologien. Dabei konzentriert man sich hauptsächlich auf die Integration dieser bestehenden Techniken in den Bereich des Grid Computing. Einige fassen den Begriff breiter und binden Techniken von Web Services mit ein. Trotz der verschiedenen Sichtweisen haben alle das Ziel dem Grid eine Bedeutung zuzuordnen.

The Semantic Grid is an extension of the current Grid in which information and services are given well-defined and explicitly represented meaning, better enabling computers and people to work in cooperation[10]

2.2 Entstehung des Wissenschaftsbereiches

Durch die Forschung an den Schnittpunkten zwischen den Teilbereichen des Semantic Web, Grid und Webservices stellte man fest, daß viele Überschneidungen, dieser zu Beginn disjunkten Technologien, existieren. So greift das Semantic Web auf Grid Technologie zurück um die notwendigen Berechnungsressourcen für seine Inhalte zu koordinieren und zusammenzuschließen.

2 Die Entstehung von Semantic Grid

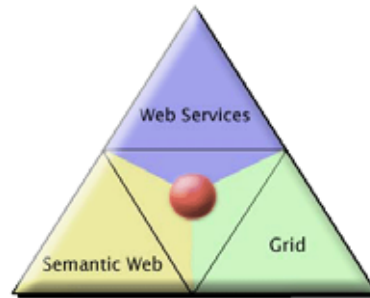


Abbildung 1: Entstehung des Wissenschaftsbereiches

2.3 Probleme im Grid Computing

Als größtes Problem des Grid Computing, wird die rasante Entwicklung der Grid Technologien angesehen. Es existiert eine riesige Menge von verfügbaren Ressourcen, was für den Benutzer zu einer reduzierten Transparenz und schweren Interpretation der Ressourcen führt. Das mächtige Angebot macht es dem Anwender fast unmöglich die richtige Ressource zu finden und für seine Zwecke zu nutzen.

Wir benötigen somit eine intelligente Koordination der Ressourcenteilung und die Übernahme der Interpretation von Ressourcen durch das Grid selber. In den Bereichen von Job Control, Accessmanagements oder der Medizin, ist das automatisierte Finden von Ressourcen und Services, um den menschlichen Aufwand zu reduzieren und eine effektivere Verwendung der Ressourcen zu gewährleisten, besonders gefordert.

2.4 Die Idee des Semantic Grid

Wir gehen aus von den Funktionen des Grid, erweitern diese um die Technologie des Semantic Web und binden Kernfunktionen der Webservices mit ein.

Wir haben die Existenz von Informationen, Services und Dienstleistungen. Dabei verfolgen wir die Absicht, eine bessere und effektivere Zusammenarbeit von Computer und Personen zu erschaffen. Das Semantic Grid wird Charakterisiert, als ein offenes System, bestehend aus Usern, Softwarekomponenten, Ressourcen, Berechnungseinheiten, Applikationen und wissenschaftlichen Instrumenten, wie Analysewerkzeugen, Teleskope, Mikroskope usw.. Diese Ressourcen kommen und gehen auf einer kontinuierlichen Basis, d.h. Ressourcen werden der Öffentlichkeit oder Benutzergruppen bereitgestellt und wieder vom Markt genommen. Es entstehen virtuelle Zusammenschlüsse von Ressourcen und Organisationen. Was wir benötigen ist eine Sicht mit dem Ziel der Vereinigung der Vorteile von Grid, Semantic Web und Web Services.

Es existieren Dienstleistungen, die lokal nicht Verfügbar sind, zu teuer in der Anschaffung oder nur für eine einmalige Benutzung benötigt werden. In diesem Fall kann man zur Deckung dieser Ressourcenbedürfnisse zu den Dienstleitungen eines Service Grids greifen. Damit man die jeweiligen Ressourcen besser und effektiver finden, nutzen und verteilen kann, benötigt das Grid eine Semantik. Das System muß in der Lage sein, in einem zeitlich vertretbaren Rahmen, ein potentiell großes Volumen von verteilten Inhalten speichern und verarbeiten zu können. Es muß die Möglichkeit bestehen, Ressourcen zu lokalisieren und eine Verbindung mit ihnen auszuhandeln. Des Weiteren benötigt man eine Verarbeitung von Aufträgen, die bei einer bestimmten Ressource anfallen, sowie eine On Demand geplante Nutzung von Ressourcen.

3 Semantic Web

3.1 Nutzung von Semantic Web Technologien

In den letzten Jahren wurden schon viele Ideen des Ansatzes in die Praxis umgesetzt und die Forschungsgemeinde des Semantic Grid wird immer größer. Im Einzelnen wurde das Semantic Web Resource Description Framework (RDF) und die Web Ontologie (OWL) vom W3C zertifiziert und als Standard freigegeben. Werkzeuge, welche diese Dienste unterstützen, sind mit immer größer werdender Anzahl verfügbar.

Um die Vorgaben der Semantic Grid Vision effektiv Nutzen zu können greifen wir auf die bestehende Technologie des Semantic Web zurück.

Hier stehen uns Resource Description Framework (RDF) für die Beschreibung der Metadaten, sowie OWL und OWL-S für die Erstellung von Ontologien zur Verfügung. Wir nutzen damit Semantic Web Technologie für

- Klassifikation von Berechnungs- und Datenressourcen
- Job Control
- Workflowbeschreibung
- Authentifikation
- Accounting
- Accessmanagement

innerhalb des Knowledge Managements

3.2 RDF Metadaten

Um Informationen über Ressourcen für weitere Zwecke interpretierbar zu gestalten, ist die Verwendung von XML als Beschreibungssprache nicht mehr ausreichend. Das Resource Description Framework (RDF) wurde vom World Wide Web Konsortium (W3C) entwickelt. Es nutzt XML als Grundlage und bietet eine einheitliche Konvention für die semantische Beschreibung von Inhalten, Ressourcen und Dienstleistungen. Metadaten im Semantic Grid, welche mittels Services verbunden sind, weisen darauf hin, wie man auf den Service selber zugreifen kann und welche Vereinbarungen getroffen werden müssen, um den Dienst zu Nutzen.

Die Beschreibung einer Ressource ist für eine automatisierte Suche und Entdeckung sowie Selektion von großer Bedeutung. Die Wahl der Ressource ist abhängig von den Metadaten die existieren, sowie der Klassifizierung der Services, basierend auf der Funktionalität die sie anbieten. In einer Architektur in der Services veränderlich sind, kommen und gehen und Konfigurationen sich auflösen und reorganisieren, ist die Semantik von großer Bedeutung. Die Fähigkeit Datenanfragen an völlig verschiedene Informationsspeicher zu stellen und den Inhalt in einem neuen Sachverhalt zu verwenden, benötigt eine Interoperabilität von

- Informationen
- Fachsprachen
- Terminologien

welche in unterschiedlichen Bereichen verwendet werden. Das Namensproblem wird durch existierende Standards in den verschiedenen Bereichen, wie z.B. in der Medizin erleichtert. Damit semantische Beziehungen effizient erstellt werden können, greift man auf Ontologiesprachen zurück.

3.3 OWL

Die Web Ontologie Language (OWL) wurde auf Basis der Darpa Agent Markup Language (DAML) und der Ontology Inference Layer (OIL) entwickelt. Diese bauen wiederum auf RDF und XML auf. Beide Sprachen DAML (amerikanische Entwicklung) und OIL (europäische Pendant) wurden wegen ihrer gleichen Entwicklungsrichtungen und Ziele zu DAML+OIL zusammengefasst.

Ontologien legen die Erweiterungen von Begriffen und Termen und ihre Beziehungen zwischen ihnen fest. Sie beschreiben

- Objekte
- Prozesse

- Ressourcen
- Dienstleistungen

Man kann somit sagen, daß Ontologien die Basis von Metadaten bilden. Semantic Grid benötigt Ontologien als fundamentalen Baustein für die spätere Beschreibung von Ressourcen, und angebotenen Dienstleistungen. Technologien wie UDDI und WSDL bietet uns eine eingeschränkte Möglichkeit der mechanisierten Serviceerkennung, Kombination, Konfiguration und automatisierten Verhandlungsmechanismen. Die Grid-Research-Group verfolgt das Ziel, Grid Nutzern und Entwicklern bei der Integration von Semantic-Web-Technologien zu helfen und die Brücke zu schlagen für den Erfolg dieses Ansatzes.

3.4 Semantic Web Services

Semantic Web Services stellt eine neue Basis für automatisierte Abläufe im World Wide Web dar. Durch die Integration von Semantic Web Techniken in Web Services, wird für Software Agenten die Basis geschaffen, Web Services automatisch zu finden, sie zu kombinieren und auszuführen.

Man will damit komplexere Sachverhalte, welche bisher nur durch die manuelle Ausführung realisierbar waren, autonom und anpassungsfähig gestalten. Die Web ontology Language for Services (OWL-S) , welche auf OWL aufbaut, bietet hierfür die Grundlage zur Beschreibung von Semantic Web-Services. Einen alternativen Ansatz bietet das Web Services Modelling Framework (WSMF), um Annotationen für Semantic Web Services zu entwickeln.

Die bisherigen Forschungsergebnisse im Grid kann man als semi-automatisiert bezeichnen, da die Nutzung und Aushandlung von Dienstleistungen, sowie die Komposition im Grid meist noch manuell und nicht flexibel realisiert werden. In diesem Bereich existiert noch sehr viel Forschungsbedarf, um das Ziel des vollständig autonomen Ablaufes zu realisieren.

4 Service orientierte Sichtweise

Jeder Service hat einen Inhaber oder eine Gruppe von Besitzern. Er ist verantwortlich für das Angebot, welches er zum Verbrauch anbietet. Er legt die Konditionen fest, wie auf Service zugegriffen wird. Die Beziehung zwischen Inhaber und Konsument wird in einem Service Vertrag festgehalten. Um eine vollautomatisierte Aushandlung und Komposition der Ressourcen und Dienstleistungen zu gewährleisten, wird diese Serviceorientierte Sichtweise für das Semantic Grid benötigt.

Semantic Grid as a service-oriented architecture in which entities provide services to one another under various forms of contract.

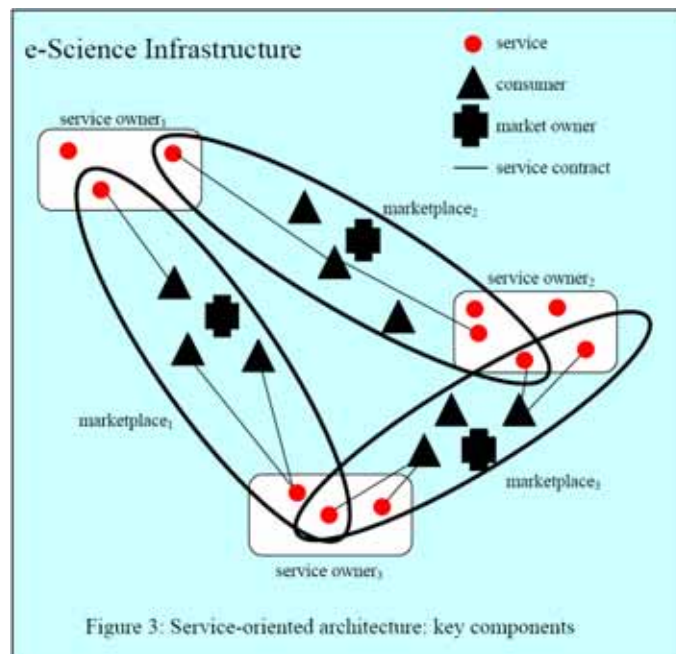


Abbildung 2: Service-oriented architecture[4]

4.1 Web-Services

Web Services bieten Ressourcen, Software und weiteren Komponenten, die Möglichkeit miteinander zu kommunizieren und sich zusammenzuschließen. Sie verbinden, kombinieren Computer und andere Ressourcen über das Web, um Daten neu zu organisieren oder auszutauschen. Dabei bilden sie lose miteinander verbundene und wieder verwendbare Softwarekompositionen. Genau diese Eigenschaften lassen sich auf den Bereich des Semantic Grid übertragen. Web Services implementieren Applikationen die den Status managen, Dienstleitungen und andere Stateful Ressourcen finden, untersuchen und integrieren. Hierbei wird auf Web Services Resource Framework (WSRF) zurückgegriffen.

The WS-Resource Framework (WSRF) defines uniform mechanisms for defining, inspecting, and managing remote state, a crucial concern in many settings.[7]

Jeder Service kann einem Besitzer oder einer Gruppe zugeordnet werden. Bietet Dienstleistungen sozusagen als Verbrauchsmaterial für Konsumenten an. Der Besitzer legt die Bedingungen fest, wie auf den Service zugegriffen werden kann. Ist der Services universal erreichbar, frei für jeden zugreifbar, handelt es sich um Re-Seller-Ressourcen, oder ist es eine Kombination von Services. Festgelegt wird es als Service Vertrag mit Konditionen wie Zustimmung oder Ablehnung, Preis, was liefert der Service usw.

4.2 Agentenbasierte Sichtweise

Die Idee der autonomen Verhandlungen muß in das Konzept von Grid Services integriert werden. Services sind nicht immer erreichbar, werden gerade durch andere blockiert oder sind für das Experiment ungeeignet. Man benötigt Verhandlungsmechanismen, um von beiden Seiten akzeptierbare Vereinbarungen zu schließen. Es muß sich über die Bedingungen und Konditionen der Dienstleistungen geeinigt werden. Dieser Punkt wird aktuell in vielen Communities von Grid und Web diskutiert.

Dienstleistungsinhaber und Konsumenten werden als autonome Agenten betrachtet, die miteinander interagieren. Jeder Inhaber eines Services besitzt mehrere Agenten, welche in seinem Interesse handeln. Beim Inhaber kann es sich dabei um einen Menschen oder eine künstliche Intelligenz handeln. Die Agenten regeln den Zugriff auf den Service für den sie verantwortlich sind und stellen sicher, daß die Bedingungen für die Vereinbarungen eingehalten werden. Dies setzt eine Planung der lokalen Aktivitäten voraus.

Diese Techniken auf den Bereich des Semantic Grid zu übertragen ist jedoch nicht

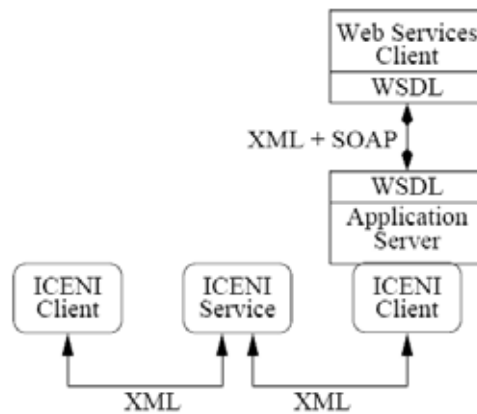


Abbildung 3: Interaction with web services[1]

so einfach. Verschiedene Agenten verwenden verschiedene Sprachen für die Interaktion. Die ersten Agenten orientierten Interaktionen werden sich auf dem Knowledge Level entwickeln d.h. zu welchem Zeitpunkt und von wem können Dienstleistungen angefragt werden. Dabei soll der Status der Ressourcen mittels eines XML Dokumentes definiert und der Zugriff auf diesen Status durch ein in WDSL definiertes Interface erfolgen.

5 Architektur des Semantic Grid

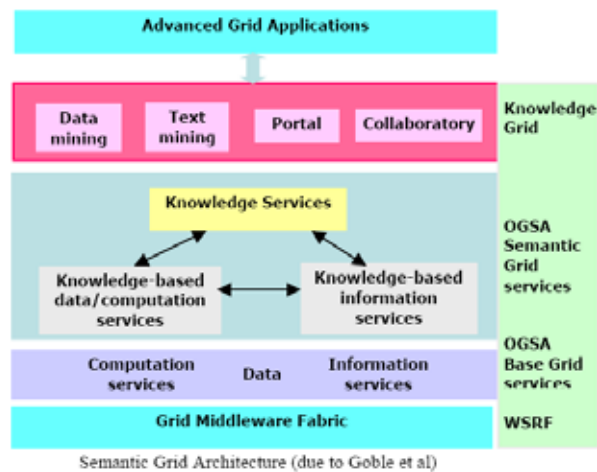


Abbildung 4: Semantic Grid Architecture[10]

Für zukünftige Grid Applikationen spielt das High Level Knowledge eine entscheidende Rolle. Grid Applikationen sind verbunden mit den repräsentativen Bereichen des Datamining, der Knowledge Discovery, der Text und Multimedia Content Analyse.

5.1 Knowledge Grid

Das Knowledge Grid besteht aus mehreren Zusammenstellungen von

- Informationen
- Expertisen
- Erkenntnissen

die aggregiert werden können, um ein wissenschaftliches Problem zu analysieren. Es besteht desweiteren aus Individuen und Gruppen des Business, der Medizin sowie Bereichen der Wissenschaft. Man stellt ihnen

- Workflow Mechanismen
- Datenbestände
- Notizen
- digitale Archive

zur Verfügung zur Unterstützung ihrer Forschungsarbeit.

5.2 Knowledge Generating Services

Knowledge Services identifizieren Muster, empfehlen, bestimmen Aktionen und publizieren Resultate, die für verschiedenen Personen oder Gruppen von Interesse sind.

5.3 Knowledge Based Services

Knowledge-Aware, Knowledge-Based, or Knowledge-Assisted Grid services, verteilen Grid-Ressourcen, die auf den Knowledge Bereich zugeifen. Desweiteren finden wir auf dieser Ebene

- Problemlösungsanwendungen
- Semantic Based Service Discovery
- Ressourcen Brokering
- Semantic Data Integration
- Workflow Kompositionsplanung

5.4 Grid Knowledge Services

Grid knowledge services sind Dienstleistungen und Technologien für das global verteilte Knowledge Management, welches von Netzwerken, Grid Services und Grid Applikationen verwendet wird. Sie beinhalten Sprachen wie OWL und OWL-S

6 MyGrid

6.1 Programmbeschreibung

MyGrid stellt ein Middleware Konzept für *in silico Experimente* der Biologie dar. In den Bereichen des Life Science werden traditionsgemäß Datenbankanfragen und analytischen Werkzeuge mittels komplizierten Scripten miteinander verbunden, um bestehende Inkompatibilitäten zu lösen. In verschiedenen Situationen werden oft Inhalte durch Copy und Paste Verfahren manuell eingefügt oder auf den neuesten Stand gebracht. *In silico Experimenten* werden normalerweise ohne eine

- Wiederverwendbarkeit der Resultate
- ihrer Herkunft
- ihrer Methoden

durchgeführt.



Abbildung 5: MyGrid functions[9]

6.2 Aufbau

Das MyGrid Projekt hat eine komplette Suite von Middlewarebestandteilen entwickelt, um datenintensive *in silico Experimente* der Biologie durchzuführen.

Die Middleware wurde entwickelt, für

6 MyGrid

- das Zusammensetzen das von in silico Experimenten
- das Finden und anpassen von anderen bestehenden Resultaten
- die Steuerung des Experimentellen Ablaufes
- mittels Applikation Workflows und Web services
- das Semantisches suchen
- Semantic workflow Kompositionen
- Semantische Integration von Resultaten

Man kann durch die Integration der MyGrid Middleware die Arbeit von Wochen auf Stunden reduzieren.

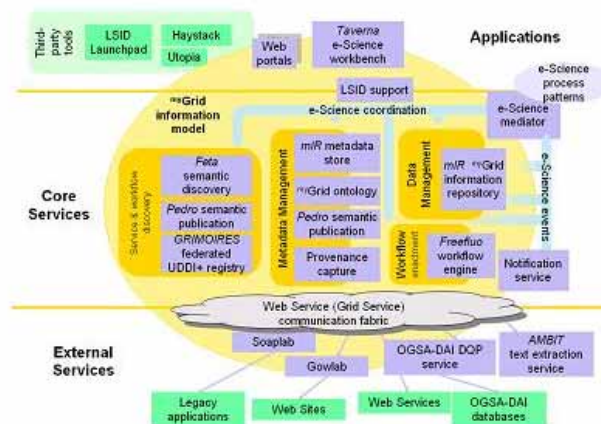


Abbildung 6: MyGrid architecture[9]

MyGrid verwendet RDF, um die Herkunft von Workflows und Services darzustellen und zu integrieren. Zu diesem Zwecke nutzt es die DAMIL+OIL Ontologie für die Klassifikationen von Services.

7 Aktueller Status

7.1 Forschungsstand

Die Grid Community ist international und besteht aus Wissenschaftlern verschiedenster Gebiete. Sie kommen aus den Bereichen der Anwendungsentwickler, Service Provider, welche das die Basis für das Grid und Ressourcen bilden, sind Softwarespezialisten, die Middleware Lösungen anbieten und konfigurieren.

Im Moment sind Grid Middleware Lösungen noch schwer zu konfigurieren anzupassen und zu nutzen.

Beide Gebiete Semantic Web und Semantic Grid befinden sich noch im Entwicklungsstadium und ihre Funktionalität ist noch nicht voll ausgereift. Der aktuelle Status des Semantic Grid ist der, des Semantic Web vor etwa 3-4 Jahren.

Whilst the basic metadata infrastructure already exists in the shape of RDF, metadata issues have not been fully addressed in current grid deployments[4]

Nach einer Aussage von Jim Hendler ist der Forschungsstand der US Semantic Grid Community hinter ihrem Gegenstück im UK. Besonders im Bezug auf der Ansicht des Grids als Service orientierte Umgebung.

Im United Kingdom werden Semantic Grid Projekte innerhalb des E-Science Programms finanziert. China arbeitet an einem eigenen Semantic Grid Forschungsplan. Es existieren nur wenige Personen die mehrere Wissenschaftsgebiete abdecken und somit eine Brücke für die Zusammenführung der Teilbereiche schlagen können. Um Personen und der Industrie die Wichtigkeit dieser Technologie zu zeigen, benötigt man sogenannte Killerapplikationen. Erste in dieser Richtung werden in Gebieten des Life Science und der Astronomie erwartet.

Literatur

- [1] Steven Newhouse, Anthony Mayer, Nathalie Furmento, Stephen McGough, James Stanton and John Darlington. *Laying the Foundations for the Semantic Grid*, 2001.
- [2] Carole Goble, David De Roure. *The Grid: An Application of the Semantic Web*, 2002.
- [3] Carole Goble, David De Roure. *The Semantic Grid: Myth Busting and Bridge Building*, 2004.
- [4] David De Roure, Nicholas Jennings, Nigel Shadbolt. *A Future e-Science Infrastructure*, 2001.
- [5] Marije Geldof. *The Semantic Grid will Semantic Web and Grid go hand in hand?*, 2004.
- [6] Semantic Grid. <http://www.semanticgrid.org>.
- [7] Ian Foster, Nicholas R. Jennings, Carl Kesselman. *Brain Meets Brawn: Why Grid and Agents Need Each Other*, 2004.
- [8] Michelle Bachler, Simon Buckingham Shum, Jessica Chen-Burger, Jeff Dalton, David De Roure, Marc Eisenstadt, Jiri Komzak, Danus Michaelides, Kevin Page, Stephen Potter, Nigel Shadbolt, Austin Tate. *Collaborative Tools in the Semantic Grid*, 2003.
- [9] Mygrid. <http://www.mygrid.org.uk>.
- [10] David De Roure, Nicholas R. Jennings, Nigel R. Shadbolt. *The Semantic Grid: Past, Present and Future*, 2004.
- [11] Shalil Majithia, David W. Walker, W.A. Gray. *Automated Composition of Semantic Grid Services*, 2003.