



IFLA  
2005  
OSLO

## World Library and Information Congress: 71th IFLA General Conference and Council

### "Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

*Conference Programme:*

<http://www.ifla.org/IV/ifla71/Programme.htm>

August 8, 2005

**Code Number:**

**007-G**

**Meeting:**

**101 Science and Technology Libraries**

### **CRIS und Open Access = Der Weg zu Forschungswissen im GRID**

**Keith G Jeffery**

Director IT,

CCLRC Rutherford Appleton Laboratory, Chilton, UK

E-mail: [kgj@rl.ac.uk](mailto:kgj@rl.ac.uk)

#### ***Abstract***

*Die Verknüpfung von CRIS (Current Research Information Systems) und OA (Open Access) verbindet Systeme für die Organisation von Forschung und Entwicklung mit Systemen für freien Zugang zu wissenschaftlichen Publikationen – dem sichtbarsten Ergebnis von Forschung und Entwicklung – im Rahmen der entstehenden europäischen GRID-Infrastruktur. Die Diskussion über Open Access verläuft sehr intensiv zwischen den beiden Polen „grün“ (Selbstarchivierung an den Forschungseinrichtungen) und „gold“ (durch den Autor bzw. die Institution finanzierte Veröffentlichung) als konkurrierende und sich ergänzende Ansätze. Die großen Verlage experimentieren mit „gold“-Dienstleistungen, während die „grünen“ Selbstarchivierungsverfahren schnell an Bedeutung gewinnen. GRIDs, vor allem seit den NGG (Next Generation GRID) Reports, sind zu einer Vision für die europäische IT-„Struktur“ geworden und werden nun unter der Federführung der EC DG INFSO F2 stufenweise implementiert, um einen einfach zu nutzenden Zugang zu Information und Rechenleistung anzubieten. CRISs bieten einen Hintergrund für die Evaluation des wissenschaftlichen Publizierens und für das Verständnis des wissenschaftlichen Publikationsprozesses selbst. CRISs bieten außerdem eine Organisationsplattform für Forschung und Entwicklung in sämtlichen Institutionen, von den Fördereinrichtungen über die nationale Forschungseinrichtun-*

*gen bis hin zu Universitäten. Zudem stellen sie eine Plattform für den Austausch von Forschungs- und Entwicklungsergebnissen zur Verfügung.*

## **1. Einleitung**

Wir leben in einer Welt, die von Wettbewerb und Bewertung geprägt ist. Der wissenschaftliche Bereich ist davon nicht ausgenommen. Die meisten Staaten evaluieren die Qualität ihrer Universitäten und Forschungseinrichtungen auf der Grundlage ihrer meßbaren Größen – und treffen auf dieser Datengrundlage Entscheidungen über ihre Finanzierung. Eine dieser messbaren Größen ist die Anzahl der wissenschaftlichen Veröffentlichungen.

CRISs bieten die notwendigen Informationen, um Forschung und Entwicklung zu organisieren – ebenso wie den Technologietransfer und die Wertschöpfung durch Forschung und Entwicklung. In der Welt von F&E werden CRISs vor allem in Institutionen der Forschungsförderung / -finanzierung genutzt, in Forschungseinrichtungen und Universitäten. Bislang werden die CRIS-Systeme hauptsächlich genutzt, um das Forschungsumfeld – Projekte, Wissenschaftler, Organisationen, technische Ausstattung, Finanzierung – zu beschreiben, aber nicht direkt um den Output zu messen (Patente, Produkte, Veröffentlichungen).

Die Veröffentlichung von Forschungsergebnissen (oder der Transfer von Daten oder Wissen) war eine der wesentlichen Stützen der Entwicklung der technischen Kultur der Menschheit über Jahrtausende – angefangen bei den Höhlenmalereien mit Tierdarstellungen (gemalt vielleicht, um die Jäger auszubilden oder für die Jagd zu begeistern) bis hin zu den Zeichnungen und Texten Leonardo da Vincis (frühe Forschungs- und Projektberichte). Die Verfügbarkeit preiswerter Drucktechniken ermöglichte ein explosives Anwachsen des Schriftguts; dies erforderte Maßnahmen zur Qualitätsmessung. Wissenschaftliche Gesellschaften rezensierten die Veröffentlichungen, häufig schon wenn sie in Manuskriptform vorlagen und dann im kleinen Kreis vorgetragen wurden. Daraus entwickelte sich das gegenwärtig übliche Peer-review-Verfahren. Heutzutage kann ein elektronischer Preprint mit Hyperlinks und mit ausführbarem Programmcode sowie angehängten Datenblättern von jedem begutachtet werden, indem er ihn mit Anmerkungen in elektronischer Form versieht.

Der Prozess, wissenschaftliche Ideen aus dem Bewusstsein des Forschers zu konkretisieren und der damit verbundene der Prozess des Aufzeichnens (nicht nur der Ideen, sondern auch der damit zusammenhängenden experimentellen Ergebnisse und Beobachtungen) bewahrt das Ergebnis der Arbeit über die Lebenszeit des Forschers hinaus und macht es für andere nachvollziehbar und verfügbar. Einige Wissenschaftler behaupten, dass dieses bewahrte aufgezeichnete Gedächtnis das wesentliche Unterscheidungskriterium des Menschen gegenüber den übrigen Lebewesen ist.

Um einen ganzheitlichen Blick auf die wissenschaftliche Forschung zu bekommen – nicht nur auf die Ergebnisse (von denen die offensichtlichsten die Veröffentlichungen sind) , sondern auch auf die Zusammenhänge – , wollen wir CRISs und die Publikationsdatenbanken zusammenbringen. Genauer gesagt – und um die Vorteile der GRIDs-Technologie zu nutzen – wollen wir CRISs und Open-Access-Dokumentenspeicher zusammenbringen, um Forschung auf einer elektronischen Grundlage zu ermöglichen. Dieses Konzept erlaubt es dem Endnutzer, sich des elektronischen Zugriffs auf vorhergehende Forschungsarbeiten, auf den Forschungskontext , auf die Experimente und Daten bzw. die eingesetzte Software zu bedienen. Wenn nötig kann der Nutzer aus der Entfernung Experimente kontrollieren und wieder-

holen und neue Berichte verfassen, die auf der soliden Grundlage leicht zugänglichen elektronischen Materials erstellt werden. Dieses Material bildet den Korpus des Forschungswissens.

Parallel dazu kann der Wissenschaftsmanager die Forschungsergebnisse im Zusammenhang analysieren und muß sich nicht allein auf z.B. das bloße Zählen von Publikationen verlassen. Außerdem können private Wirtschaftsunternehmen die Daten der CRIS-Systeme nutzen, um neue Möglichkeiten der Wertschöpfung zu nutzen.

Das WWW ermöglicht – über Suchmaschinen – einfachen Zugang zu Informationen, wenngleich oft mit zweifelhafter Relevanz (precision) oder Vollständigkeit (recall). Die GRID-Technologie – ursprünglich im Bereich des Metacomputing entstanden (Verbindung von Supercomputern zur Simulation eines noch leistungsfähigeren Rechners) – hat sich inzwischen weiterentwickelt und integriert Web-Services in OGSA (Open GRID Services Architecture). Die GRID-Technologie geht dadurch nahtlos über ins WWW und erweitert es um Rechenkapazität. Das hinter GRIDs in Europa stehende Konzept (das nicht unbedingt das Gleiche wie in Nordamerika ist) besteht darin, dass der Endnutzer eine Anfrage stellt, das GRID-System diese übersetzt und dann dem Nutzer einen „Vertrag“ vorschlägt (der finanzielle Vergütungen und/oder die Übertragung von Rechten beinhaltet). Falls der Endnutzer zustimmt, führt das GRID-System die Anfrage aus, indem es Agenten und Makler benutzt, die zwischen verschiedenen Quellen und Ressourcen vermitteln.

Der wissenschaftliche Prozeß kann als ein Workflow zur Aufzeichnung von Ergebnissen in verschiedenen Stadien betrachtet werden, von der ursprünglichen Idee über Projektanträge, Zwischenberichte bis zu den endgültigen Ergebnisveröffentlichungen – zusammen mit den gewonnenen Daten, der eingesetzten Software und Querverweisen auf andere Arbeiten.

Dies kann in einer Art Gleichung zusammengefasst werden: CRIS + Open-Access = der Weg zu Forschungswissen im GRID. Die weiteren Teile dieses Berichts untersuchen das Forschungswissen, CRISs, Open Access, GRIDs und fassen die Ergebnisse schließlich zusammen.

## **2. Forschungswissen**

Im Bereich von Forschung, Entwicklung und Innovation besteht das geistige Eigentum (IP – intellectual property) aus Produkten, Patenten und Veröffentlichungen (im weitesten Sinne: jede außerhalb des Bewusstseins abgespeicherte Form menschlichen Verstands). Während konventionelle wissenschaftliche Veröffentlichungen („weiße“ Literatur) den größten Anteil des sichtbaren geistigen Eigentums ausmachen, besteht der „unter der Wasseroberfläche liegende Teil des Eisbergs“ aus der „grauen“ Literatur. Dieser repräsentiert gewöhnlich das eigentliche „know how“ oder die eigentliche Wissensbasis. [JE99], [JeAsRe00]. Dies beruht auch auf juristischen Überlegungen: viele Organisationen schützen ihre geistigen Eigentumsrechte mit Patenten oder Vorab-Veröffentlichungen; dem Urheberrecht und Datenbankrecht stehen Informationsfreiheit und Datenschutz gegenüber. Innovation, Technologietransfer, Wertschöpfung und Lebensqualität sind die Hauptziele von Forschung & Entwicklung; deshalb investieren Regierungen, Wirtschaftsunternehmen, Wohltätigkeitsorganisationen und sogar Einzelpersonen in Forschung und Entwicklung. Ein Großteil der Technologie, die wir heute nutzen, ist das Ergebnis von Forschung und Entwicklung früherer Zeiten und auch die

Lebensqualität, die wir heute genießen, resultiert aus Forschung und Entwicklung etwa in den Bereichen Medizin, Erziehung, Umwelt. Das ist das geistige Eigentum.

Jede Organisation muß ihr geistiges Eigentum für ihr wirtschaftliches Überleben nutzen (dies gilt auch für das Geschäft von Forschung und Entwicklung selbst) – ebenso wie für Zwecke der Öffentlichkeitsarbeit und des Marketings. Das bedeutet, dass eine Organisation wissen muß, welches geistige Eigentum sie besitzt. Sie muß es ordnen, verwalten und wirtschaftlichen Nutzen daraus ziehen – nicht zuletzt um Investitionsentscheidungen für weitere F&E-Projekte in der Zukunft zu treffen, die weiteres geistiges Eigentum produzieren. Die Wissensbasis besteht nicht nur aus der „weißen“ Literatur, sondern auch aus dem „Eisberg“ der grauen Literatur, die das Know-how der Organisation in Forschungsberichten, Gebrauchsanleitungen, Schulungsmaterialien usw. enthält. Weiterhin befindet sich das geistige Eigentum zunehmend in Datensammlungen (z.B. Ergebnisse klinischer Drogentests), Datenbanken (z.B. Kundendatenbanken) und in Softwareprogrammen (die Geschäftsabläufe eines Unternehmens beinhalten). All das kann zum Gegenstand einer genauen Untersuchung durch die Wirtschaftsprüfung oder eine Anfrage im Rahmen des „Freedom of Information“-Aktes werden und deshalb besteht hier ein großer Anreiz für eine Organisation, den Bestand ihres geistigen Eigentums möglichst gut zu verwalten.

Im F&E-Bereich bewerten die meisten Finanzierungsbehörden den Output – das geistige Eigentum – einer von ihr finanzierten Organisation, z.B. einer Universität oder eines Forschungsinstituts – und sie machen zukünftige Finanzierungsentscheidungen - zumindest teilweise – von dieser Bewertung abhängig. Deshalb sind Aufzeichnung, Pflege und Management des geistigen Eigentums sehr wichtig für Forschungsorganisationen. In ähnlicher Weise bestimmt für moderne Wirtschaftsunternehmen die Qualität ihres geistigen Eigentums den Erfolg und das künftige finanzielle Engagement ihrer Aktionäre. Daraus leiten wir folgende Forderung ab: es müssen Systeme und eine Umgebung bereitgestellt werden, die es den Organisationen ermöglichen, ihr geistiges Eigentum effektiv zu managen, indem sie das geistige Eigentum (z.B. graue Literatur) und die organisatorische Geschäftsstruktur und -ziele zusammenbringen.

### **3. CRISs (Current Research Information Systems)**

Die geschichtliche Entwicklung von CRISs (Current Research Information Systems) gründet im Bereich des Information Retrieval (IR). Die wichtigsten Gründe dafür waren, dass die Struktur der aufgezeichneten Informationen üblicherweise analog zum bibliographischen Schema auf Katalogkarten in Bibliotheken war (in der Form: Titel–Autor–Datum–Schlagwort–Standort).

Als Einzellösungen waren solche Systeme ausreichend, bis weitere Anforderungen wie statistische Analysen, Integration mit Daten aus anderen Datenbankmanagementsystemen, flexibles Reporting (inkl. Integration in Büroanwendungen) und Umgang mit Multimedia (bzw. Hypermedia) wichtig wurden. Diese Erfordernisse entstanden seit der Mitte der 1970er Jahre. Die Notwendigkeit, Zugang zu Informationen in sehr heterogenen und geographisch verteilten CRISs (und anderen Datenbanksystemen) zu bekommen, stellten die Unzulänglichkeiten dieser Systeme sehr deutlich heraus: es fehlte ihnen ein grundlegendes theoretisches Konzept (um eine strukturelle Abstimmung zu ermöglichen); es fehlten ihnen gemeinsame

Standards der Datenspeicherung (um vernetztes Arbeiten zu ermöglichen), und es fehlten ihnen Standardschnittstellen (um die Kommunikation mit Standard- Büroanwendungen sowohl für Eingabe und Aktualisierung als auch für Informationsaufbereitung und Reporting zu ermöglichen). Schlimmer noch, es fehlten ihnen die Metadaten, um solche Zusammenarbeit „on the fly“ zu ermöglichen.

CERIF bietet ein ausführliches Datenmodell für F&E-Informationen, das die Zustimmung von Vertretern europäischer Staaten erfahren hat (sowohl von Staaten der Europäischen Union als auch weiterer europäischer Staaten). Die ursprünglichen CERIF-Empfehlungen (1991) wurden als Grundlage für das ERGO-Pilotprojekt genutzt. Das neue CERIF-Datenmodell wurde von einer Arbeitsgruppe zur Überarbeitung von CERIF erstellt [CERIF]. CERIF 2000 bietet ein Datenmodell zur Verwaltung von Projekten, Organisationen, Personen, Finanzierung, Veranstaltungen, Ausstattungen, Räumlichkeiten und allen Beziehungen zwischen ihnen. Es bietet eine große Flexibilität, die nicht nur hierarchische (1:n) Beziehungen, sondern auch n:m (also many to many) Beziehungen erlaubt. Weiterhin sind die Beziehungen rollengesteuert und zeitlich definiert, dies ermöglicht ein sehr reichhaltiges Datenmodell. Das Modell berücksichtigt die Existenz von Patenten, Produkten und Publikationen – es setzt aber die Existenz von Systemen voraus, die diese Patente, Produkte und Publikationen managen. CERIF 2000 definiert also CRISs in einer formellen Weise und bietet so eine stabile Plattform für den Aufbau von CRISs, für den Datenaustausch zwischen CRISs-Systemen und für das Angebot von Metadaten, um CRIS-Inhalte zu beschreiben.

Die Europäische Kommission übergab die Pflege und Weiterentwicklung von CERIF an EUROCRIS; seitdem haben die neueren Versionen von CERIF die Anforderungen der Endnutzer in einer strukturierten Form berücksichtigt. Heute enthält CERIF ein ausführliches Set von Entitäten und Attributen für die bibliographischen Daten. Die wesentlichen Anforderungen, die zu diesem Modell geführt haben sind dargestellt in [Jee99], [JeAsRe00], [AsJe04].

#### **4. Open Access (OA)**

Das WWW [W3C] ermöglicht preisgünstiges und unkompliziertes e-Publishing. Dies hat zu einem explosiven Wachstum von institutionellen (und fachspezifischen, z.B. [ArXiv]) Dokumentspeichern geführt. Die Open Archives Initiative (OAI) nutzt den Dublin Core-Metadatenstandard (DC) und entwickelte ein Harvesting-Protokoll (OAI-PMH) ein, um die Dokumentspeicher zu verknüpfen.

Die beiden größten Herausforderungen für die weitere Entwicklung des WWW sind das Semantic Web (um das Netz intuitiv verständlich zu machen) und das Web of Trust (um das WWW sicher zu machen). Das Semantic Web entsteht gerade, größtenteils durch (a) stärker formalisierte Datenstrukturen, die geeignet sind für die Anwendung der Prädikatenlogik der 1. Ordnung, die zumeist strukturierte Metadaten voraussetzt, und (b) durch den Gebrauch von Ontologien, um die Bedeutung von Begriffen und die logischen Beziehungen zwischen ihnen zu definieren – als unterstützende assoziative Metadaten. Das Web of Trust bedeutet, dass das Material geschützt ist gegen Mißbrauch und dass die Organisation, die die Informationen besitzt, sicher sein kann, dass sie in einer mit der Unternehmensethik übereinstimmenden Weise genutzt werden. Dies wird erreicht durch assoziative restriktive Metadaten, die mit der Ursprungsinformation verknüpft sind.

Unglücklicherweise ist DC nicht formalisiert, so dass es von Computern zwar gelesen, aber nicht verstanden werden kann. In gewisser Weise stellt es also einen Rückschritt gegenüber MARC dar. Es fehlen Eigenschaften des Semantic Web und des Web of Trust. Dies begrenzt die Brauchbarkeit von DC für die automatische Verarbeitung, in der formalisierte Metadaten eine absolute Vorbedingung darstellen. Metadaten sind Daten über Daten; es gibt eine Klassifikation (ursprünglich 1998, wieder veröffentlicht in [Je00]), die die Trennung von Metadatenarten ermöglichte und die korrekte logische Verarbeitung erleichterte. Die Anwendung der Klassifizierung von Metadaten in CRISs wurde gezeigt in [JeLoAs02]. Eine formalisierte Fassung von Dublin Core wurde stufenweise entwickelt in [Je99], [JeAsRe00] und [AsJe04].

Der Begriff des Open-Access-Datenspeichers beinhaltet, dass der Zugriff für den Endnutzer frei ist. Es gibt zwei Hauptmodelle für Open-Access: das Modell „gold“ wird vor allem von Verlegern vertreten und verpflichtet den Autor (oder die Institution, der der Autor angehört), für die Publikation zu bezahlen. Die Alternative „grün“ macht es erforderlich, dass der Autor seinen Text in einem institutionellen oder fachspezifischen Dokumentenspeicher zeitnah zur Veröffentlichung ablegt, so dass das Material für den Online-Zugriff frei zugänglich ist. Es gibt einige Copyright-Probleme mit bestimmten Verlegern, obwohl die Mehrheit inzwischen freien Zugriff nach unterschiedlich langen Embargozeiträumen zum Publikationsdatum erlaubt. Natürlich gestattet der „grüne“ Open-Access ebenfalls die Speicherung von Preprints oder „grauer“ Literatur; dies erfordert eine klare Unterscheidung zu den durch ein Expertengremium begutachteten Veröffentlichungen. Inzwischen ermöglichen die meisten Verlage den Online-Zugriff auf ihre Publikationen im Abonnement; das Problem besteht darin, dass der Forscher Zugriff zu verschiedenen Plattformen benötigt, die alle verschiedene Oberflächen (inkl. Login ID und Passwort) haben. Für viele ist dieser Aufwand zu hoch – sie nutzen einfach Google (oder Google Scholar) und bedenken nicht den relativ schlechten Recall (Vollständigkeit) bzw. die relativ schlechte Relevanz (Präzision) ihrer Suche.

## 5. GRIDs

### 5.1.Hintergrund

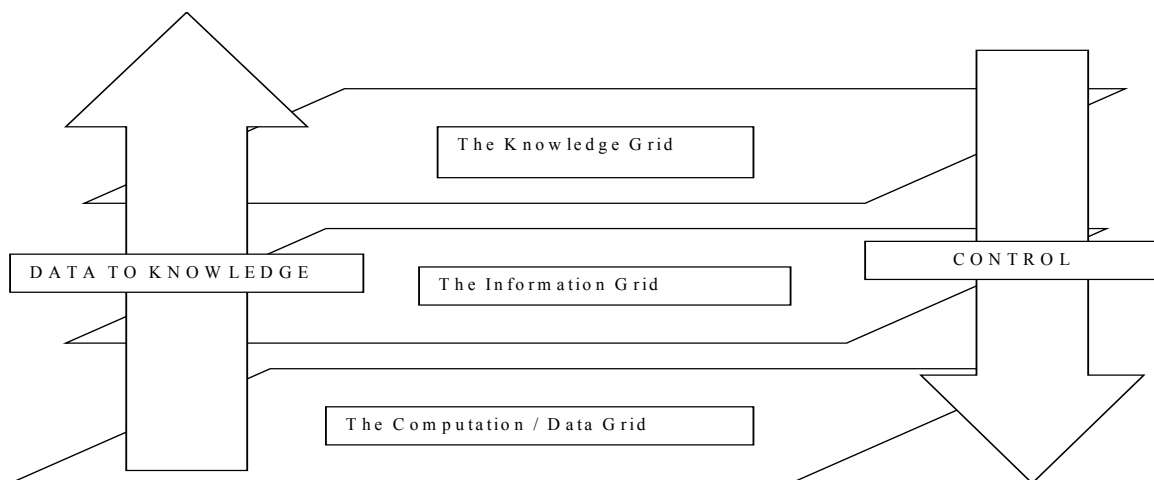


Abbildung 1: Die 3-Schichten-Architektur des GRID

In den Jahren 1998/99 war die Forschergruppe des UK Research Council mit verschiedenen IT-basierten Problemen beschäftigt. Ihr Streben nach wissenschaftlichen Entdeckungen beinhaltete fortgeschrittene Genforschung, die Erklärung des Klimawandels, ozeanographische Studien, die Messung der Umweltverschmutzung und das Erstellen von Modellen, exakte Materialwissenschaft, Studien zu Verbrennungsprozessen, fortgeschrittene Ingenieurwissenschaften, Entwicklung neuer Medikamente und Datenverarbeitung und Simulation im Bereich der Hochenergiephysik. Sie benötigten mehr Rechenleistung, mehr Speicherkapazität, bessere Analyse und Visualisierung – alles unterstützt von leicht zu bedienenden Software-Tools, die über eine intuitive Benutzeroberfläche verfügen. Der Autor wurde gebeten, eine integrierende IT-Architektur vorzuschlagen.

Die vorgeschlagene Architektur besteht aus drei Schichten (Abb. 1). Das Rechner-/Daten-GRID besteht aus Supercomputern, großen Servern, gewaltigen Datenspeichern und spezialisierten Anlagen und Einrichtungen (z.B. für VR - Virtual Reality), die alle durch Hochgeschwindigkeitsnetze verbunden sind, sie bilden die unterste Schicht der Architektur. Die Hauptfunktionen sind die Berechnung der Lastverteilung, Algorithmenverteilung, Auflösung der Adressen von Datenquellen, Sicherheit, Abgleich und Nachrichtenumleitung. Das Informations-GRID liegt über dem Rechner-/Daten-GRID und verteilt den einheitlichen Zugriff auf die unterschiedlichen Informationsquellen, hauptsächlich durch den Gebrauch von Metadaten und Middleware. Die oberste Schicht ist schließlich das Wissens-GRID, das die Entdeckung von Wissen in Datenbanken ausnutzt, um Wissen zu generieren und das die Verkörperung von Wissen in Lehrbüchern, begutachteten Publikationen und grauer Literatur ermöglicht, wobei die graue Literatur vor allem mit wissenschaftlichen Primärdaten verknüpft ist, um die wissenschaftlichen Behauptungen zu untermauern.

Parallel zu den ersten Überlegungen zu GRIDs in Großbritannien haben [FOKe98] eine Aufsatzsammlung veröffentlicht, die als die GRID-Bibel bekannt geworden ist. Der wesentliche Gedanke darin ist, Supercomputer zu vernetzen, um eine höhere Rechenleistung zu erzielen – die Metacomputer-Technologie. Der wichtigste Beitrag dazu besteht in den Systemen und Protokollen für die Planung der Zuteilung der Rechenleistung. Das GRID korrespondiert mit der untersten Schicht (Rechner-/Daten-Schicht) der von Großbritannien vorgeschlagenen GRID-Architektur.

## 5.2 Die GRID-Architektur

Die hinter GRIDs steckende Idee besteht darin, eine IT-Umgebung bereitzustellen, die mit dem Anwender zusammenarbeitet, um die notwendigen Bedingungen für die benötigten Dienste zu bestimmen und dann diese Erfordernisse über heterogene Umgebungen von Datenspeichern, Prozessorleistung, speziellen Einrichtungen für die Anzeige und Datensammlungssystem für den Endnutzer einheitlich erscheinen zu lassen.

Die außerhalb von GRIDs liegenden Bestandteile (Abb. 2) sind: a) Nutzer: Menschen oder andere, technische Systeme; b) Quellen: Daten, Informationen, Software; c) Ressourcen: z.B. Computer, Sensoren, Detektoren, Visualisierungen oder Virtual-Reality- (VR)-Einrichtungen. Jeder dieser drei Hauptbestandteile ist im GRID dauerhaft und aktiv repräsentiert durch: 1) Metadaten: sie beschreiben die externen Komponenten und sie verändern sich mit den durch äußere Ereignisse veränderten Umstände; 2) einem Agenten: er handelt im Namen der exter-

nen Ressourcen, die er innerhalb der GRID-Umgebung repräsentiert. Schließlich gibt es eine Komponente, die als Vermittler zwischen den einzelnen Agenten auftritt. Diese sind Makler, die – als Computerprogramme – genauso handeln wie menschliche Makler, indem sie Verträge und Abmachungen zwischen den Agenten aushandeln. Daraus wird deutlich, dass die Hauptkomponenten die Metadaten, die Agenten und die Makler sind.

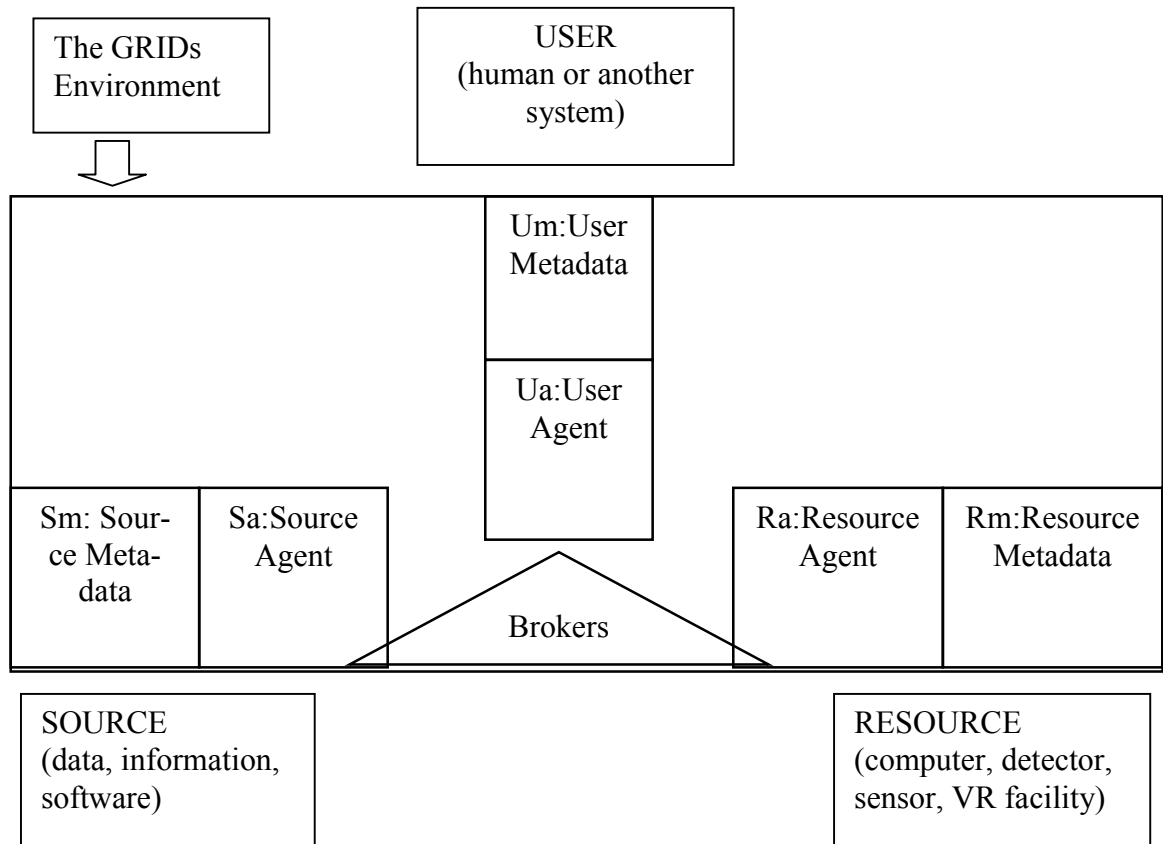


Abbildung 2: Die Bestandteile des GRIDs

### 5.3 Ambient Computing

Das Konzept des Ambient Computing besteht darin, dass die Computerumgebung **immer** präsent und erreichbar ist. Das Konzept des Pervasive Computing bedeutet, dass die Rechnerumgebung **überall** erreichbar ist und in allem steckt. Das Konzept des Mobile Computing schließlich beinhaltet, dass der Endnutzer selbst dann vernetzt ist, wenn er **unterwegs** ist. Im allgemeinen Sprachgebrauch umfaßt der Begriff des Ambient Computing sowohl Pervasive als auch Mobile Computing.

Eine typische Geräteausstattung dafür besteht etwa aus folgenden Komponenten: a) ein Headset mit Kopfhörern und Mikrophon für die verbale Kommunikation, das Set ist kabellos über Bluetooth-Technologie verbunden mit b) einem PDA (Personal Digital Assistant) mit einem kleinen Bildschirm, Zahlen/Buchstabentastatur (wie bei einem Telefon), GSM/GPRS-Verbindung (Handy) für Sprach- und Datenübertragung, drahtlose LAN-Anbindung und An-



schlüsse für Sensoren (um alle Daten in der Nähe des Nutzers aufzeichnen zu können), die ebenfalls über Bluetooth-Technologie angeschlossen sind an c) ein optionales Notebook, das in einem Rucksack mitgeführt (und bei Bedarf ausgepackt) werden kann und über einen normalen Bildschirm, Tastatur, große Festplatte und GSM/GPRS-Anschluss verfügt, ebenso wie drahtloses LAN, kabelgebundenes LAN und Wähltelefon.

Der Nutzer wird vielleicht nur die Komponenten a) und b) (oder vielleicht nur b) mit eingebautem Lautsprecher und Mikrophon) in seiner privaten oder beruflichen Umgebung als Handy und Organizer sowie zur Unterhaltung benutzen – mit oder ohne Verbindung zum Server und der IT-Umgebung seiner Heimatstation. Zum herkömmlichen Arbeiten mit Tastatur und Bildschirm wird wohl das Notebook benutzt werden, vielleicht ohne PDA. Beide können eingesetzt werden in Verbindung mit Software zur Validierung und Kalibrierung der Datensammlung auf dem Notebook und Sensoren, die an den PDA angeschlossen sind.

Eine solche Geräteausstattung ist sicherlich sinnvoll für einen „Road Warrior“ (Handelsreisenden), für Notfalldienste wie Feuerwehrleute oder Rettungssanitäter, für Geschäftsleute, Industriemanager, für die Logistikdienstleistungsbranche (Warenhäuser, Transport- und Speditionsgewerbe), für Wissenschaftler bei Feldversuchen ... ebenso wie für Freizeitaktivitäten wie Bergwandern, den Besuch einer Kunstgalerie, die Suche nach einem Restaurant oder den Besuch eines archäologischen Ausstellungsgeländes. Das Konzept besteht darin, den Zugang zu einer GRID-Umgebung überall, zu jeder Zeit und an jedem Ort zu ermöglichen. Das Konzept des Ambient Computing gibt die Möglichkeiten von GRIDs dem Menschen „in die Hand“. In Verbindung mit CRISs und Open-Access-Datenspeichern versorgt es den Wissenschaftler, Wissenschaftsmanager oder Unternehmer mit leichtem Zugang zu dem benötigten Wissen.

## **6. Zusammenfassung**

So wurde das GRID-Konzept – ursprünglich ein Metacomputing-Konzept, d.h. das Verbinden von Supercomputern [FOKe98] – erweitert (erste interne Berichte 1999, veröffentlicht in [Je01]) zu einer ausgefeilten, verteilten Computerumgebung, die – als GRID-Services – das W3C-Konzept der Web-Services mit den Vorstellungen des Semantic Web und des Web of Trust zusammenbringt. Durch das Ambient Computing ist diese Umgebung überall zugänglich und ermöglicht den Anschluss beliebiger Peripheriegeräte [Je04].

Diese Computerumgebung stellt die Plattform dar, in der die Unterschiede zwischen weißer und grauer Literatur aufgehoben werden - von Veröffentlichungen, die den Prozeß des peer-reviews durchlaufen haben, über annotierte Preprints bis hin zu Forschungsberichten und Handbüchern mit formalisierten Metadaten, die die automatische Verarbeitung in Open-Access-Dokumentspeichern ermöglichen. Diese sind verbunden mit Datenbeständen und geeigneter Software und – unter Verwendung der CRIS-Technology [EuroCris] – mit Personen, Organisationen, Projekten, Patenten, Veröffentlichungen, Veranstaltungen, Anlagen und Geräten. Die Umgebung wird vervollständigt durch angegliederte Rechenleistung, spezielle Ausgabegeräte (z.B. Virtual Reality (VR)) und die dynamische Kontrolle von Detektoren und Geräten, die Daten sammeln. Dies unterstützt den gesamten Forschungsprozeß als einen zusammenhängenden Arbeitsablauf.

Beim CCLRC haben wir aufgebaut:

(a) einen institutseigenen Open-Access-Dokumentspeicher mit über 20.000 Einträgen in einem formalisierten Dublin-Core-Format, um eine automatische Verarbeitung und Zusammenarbeit zu ermöglichen; das Datenmodell stützt sich auf das Konzept des [IFLA FRBR]-Modells

(b) einen gemeinsamen Datenspeicher, der das CERIF-Datenmodell verwendet (erweitert, um internationale Geschäftsprozesse genauso zu managen wie CRIS-Anforderungen

(c) eine GRID-Umgebung mit Ambient-Computer-Zugang, der eine einfache Zusammenarbeit erlaubt und einen einfachen Nutzerzugriff auf Informationen und Computerressourcen.

Zur Zeit integrieren wir diese Systeme gleichzeitig mit einer Umstrukturierung der innerbetrieblichen Geschäftsabläufe, um die e-Research-Umgebung für die Zukunft vorzubereiten. Das wird eine effektive und effiziente Basis für das F&E-Management des Betriebs sein.

## **Danksagung**

Ein Großteil der Arbeit an CRISs wurde in Zusammenarbeit mit Kollegen von EuroCRIS geleistet, für deren Beiträge ich mich bedanke. Die Hauptarbeit an CERIF und Informationen über die Verknüpfung von Publikationen habe ich gemeinsam mit Anne Asserson (Universität Bergen) geleistet. Die meiste Arbeit an den Open-Access-Dokumentspeichern habe ich mit meinem eigenen Team am CCLRC geleistet – besonders Matthew Mascord, Catherine Jones, Brian Matthews und Heather Weaver –, das eine ständige Quelle der Inspiration und neuer Ideen war. Die Arbeit an GRIDs hat außerdem Nutzen gezogen aus Diskussionen mit der britischen e-Science-Community und weiter gefaßten Diskussionen im europäischen Kontext, hauptsächlich durch [ERCIM] und die [NGG] Gruppe. Die Zusammenfassung und dieser Aufsatz liegen in meiner Verantwortung

## **Literatur**

[ArXiv] [www.arxiv.org](http://www.arxiv.org)

[AsJeLo02] Asserson,A; Jeffery,K.G; Lopatenko,A: ‘CERIF: Past, Present and Future’ in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6<sup>th</sup> International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 33-40 2002 (available under [www.eurocris.org](http://www.eurocris.org) )

[AsJe04] Asserson, A; Jeffery, K.G.; ‘Research Output Publications and CRIS’ in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 (available under [www.eurocris.org](http://www.eurocris.org) )

[Be99] Berners-Lee,T; ‘Weaving the Web’ 256 pp Harper, San Francisco September 1999 ISBN 0062515861

[CERIF] <http://www.cordis.lu/cerif/>

[DC] [http://purl.oclc.org/metadata/dublin\\_core/](http://purl.oclc.org/metadata/dublin_core/)

[ERCIM] <http://www.ercim.org>

- [ERGO] <http://www.cordis.lu/ergo/>
- [EuroCRIS] <http://www.eurocris.org>
- [FoKe98] Foster I and Kesselman C (Eds). The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufman 1998
- [IFLA] <http://www.ifla.org/>
- [IFLA FRBR] <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- (Je99) Jeffery, K G: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 <http://www.konbib.nl/greynet/frame4.htm>
- [JeAsRe00] Jeffery K.; Asserson A.; Revheim J; (2000) CRIS, Grey Literature and the Knowledge Society, Proceedings CRIS-2000, Helsinki [ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery\\_fulltext.pdf](ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery_fulltext.pdf)
- [Je00] Jeffery, K.G., 2000, 'Metadata': in Brinkkemper, J; Lindencrona, E; Solvberg, A: 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.
- [Je01] Jeffery, K.G.; 'GRIDs: Next Generation Technologies for the Internet' Invited Keynote Presentation (Abstract in Proceedings) Eds Wang, Y; Patel, S; Johnston, R.H; OOIS2001 Conference, Calgary August 2001, page 1, Springer, ISBN 1-85233-546-7
- [JeLoAs02] Jeffery, K.G; Lopatenko, A; Asserson, A: 'Comparative Study of Metadata for Scientific Information: The Place of CERIF in CRISs and Scientific Repositories' in Adamczak, W & Nase, A (Eds): Proceedings CRIS2002 6<sup>th</sup> International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 77-86 (available under [www.eurocris.org](http://www.eurocris.org) )
- [Je04] Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino, E; Christodoulakis, S; Plexousakis, D; Christophies, V; Koubarakis, M; Bohm, K; Ferrari, E (Eds) Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004
- [Je04a] Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 (available under [www.eurocris.org](http://www.eurocris.org) )
- [MARC] <http://minos.bl.uk/services/bsds/nbs/marc/commarcm.html>
- [NGG] [www.cordis.lu/ist/grids](http://www.cordis.lu/ist/grids)
- [OAI] [www.openarchives.org](http://www.openarchives.org)
- [W3C] [www.w3.org](http://www.w3.org)
- [W3Cmetadata] <http://www.w3.org/Metadata/>