



Arbeitspaket 2: Blaupausen und Beratung

Aufbaukonzepte für neue Community-Grids, Aktivitäten und Ergebnisse der Fachberaterteams¹

Deliverable	2.2.2 Aufbaukonzepte Community-Grids
Autoren	Arbeitspaket 2: Blaupausen und Beratung
Editoren	I. Agapov, F. Schlünzen
Datum	22-11-2010
Dokument Version	1.0.0

A: Status des Dokuments

Deliverable 2.2.2, Version 1.0.0, Release

B: Bezug zum Projektplan

Basierend auf den Blaupausen und Materialien werden Konzepte für den Aufbau neuer Community-Grids entwickelt.

C: Abstract

Die Fachberater-Teams sollen den Aufbau neuer Community-Grids unterstützen, und neuen User-Communities den Zugang zu der Grid-Infrastruktur erleichtern. Die Aufbaukonzepte für die neuen

¹ This work is created by the WissGrid project. The project is funded by the German Federal Ministry of Education and Research (BMBF).

Communities werden vorgestellt, sowie die Aktivitäten der Fachberater dokumentiert.

D: Änderungen

Version	Date	Name	Brief summary
0.1.0	14.01.2010	F.Schlünzen	Working Draft Creation
0.1.1	28.09.2010	I. Agapov	Überarbeitung
0.1.2	12.11.2010	T. Rathmann	Überarbeitung
0.1.3	22.11.2010	I. Agapov	Überarbeitung

E:

Inhaltsverzeichnis

1	Overview	4
2	Aufbaukonzepte	5
2.1	Compute-Grid	6
2.2	Daten-Grid	6
2.3	Prozesse	7
2.4	Software-Entwicklung, Adaptierung von existierender Software	7
2.5	Testumgebungen	7
2.5.1	Die D-Grid-Referenzinstallation	8
2.5.2	dech VO	8
2.6	Deployment	8
2.7	Support	9
2.7.1	WissGrid Support	9
2.7.2	Community Support	9
3	Aktivitäten und Ergebnisse	10
3.1	Forschung mit Photonen	10
3.1.1	Virtuelle Organisationen	10
3.1.2	Authentifizierung	11
3.1.3	Prototypische Anwendungen	12
3.1.4	Beratungs-Aktivitäten - Photon Science	17
3.2	Sozialwissenschaften	19
3.3	Biostatistik und Epidemiologie	19
3.3.1	Beschreibung	19
3.3.2	Beratungs-Aktivitäten - Biostatistik und Epidemiologie	19

1 Overview

Große Forschungscommunities wie die Hochenergiephysik, welche die LHC Experimente betreibt, haben meistens Grid-ähnliche Lösungen für das verteilte Datenmanagement und Rechnen entwickelt, weil beides für die Datenanalyse essenziell ist.

In kleinen Forschungscommunities und Communities, in denen traditionell kein großer Bedarf an Rechen- und Speicherkapazität besteht, fehlen solche Erfahrungen in der Regel. Forschungslandschaft und Bedarf an IT-Ressourcen können sich aber schnell ändern.

Erstens, dank moderner Technologien wie z.B. bei Experimenten mit Synchrotronstrahlung, ist die Menge der entstehenden primären Forschungsdaten so groß geworden, dass sie nicht mehr lokal an den Mess-PC's oder Laptops verwaltet werden können, sondern in Datenzentren aufbewahrt werden müssen.

Zweitens, mehr und mehr Forschung wird im Rahmen von internationalen Kollaborationen durchgeführt. Die Forschungsdaten sollen weltweit für mehrere Nutzergruppen zugänglich sein. Dies lässt sich nur innerhalb von Grid-Infrastrukturen bzw. virtuellen Forschungsumgebungen vernünftig lösen.

WissGrid stellt Expertise für Aufbau von neuen Community-Grids bereit und bietet entsprechende Beratung an. In diesem Dokument werden generische Aufbaukonzepte präsentiert, die nicht Community-spezifisch sind und in einem breiten Spektrum von Umgebungen umgesetzt werden können.

Im Kapitel 2 werden typische Anwendungsszenarien und die Fragen erörtert, die vor dem Aufbau der Infrastruktur geklärt werden müssen. Verfahren zum Aufbau von Community-Grids werden diskutiert. Im Kapitel 3 werden Aktivitäten und Ergebnisse der Fachberaterteams am Beispiel Photonphysik, sowie Epidemiologie und Sozialwissenschaften vorgestellt.

2 Aufbaukonzepte

Der Grund, warum lokales Rechnen und Datenmanagement nicht genügen und eine Grid-Lösung erwünscht ist, muss klar verstanden werden. Es kann, z.B., Mangel an Speicherkapazität, Mangel an Rechenkapazität oder entstehende wissenschaftliche Kollaborationen sein. Die Ziele der Umsetzung von Grid-Infrastruktur oder Virtueller Forschungsumgebung müssen definiert werden. Das Ziel kann z.B. weltweiter Zugriff auf die Daten oder schnelleres Rechnen sein. Die folgende Liste kann ein Startpunkt für die Aufstellung eigener Ziele sein:

- **Erforderliche Rechenkapazität:**
Reichen die lokalen Rechenressourcen für die aktuellen Forschungsaufgaben? Wenn nicht, wie viele Ressourcen fehlen (in FLOPS [1], SPEC [2], etc.)?
Wie viel wird die Beschaffung der entsprechenden Ressourcen kosten?
Gibt es Partnerinstitute, die Rechenkapazität teilen können?
- **Software:**
Wird überwiegend Standardsoftware (wie R [3]) oder spezifische Software (wie LHC ATLAS Software) in der Community verwendet?
Wie aufwändig ist es, diese Software zu unterstützen?
Sind Lizenzbestimmungen zu beachten?
- **Datenmengen, Anforderungen an Speicher:**
Wie viele Daten sind zu speichern?
Wie hoch ist die gewünschte Zugriffsgeschwindigkeit?
Welche Latenz ist zu tolerieren?
- **Datenzugriff:** Sollen die Daten direkt von der Anwendungssoftware gelesen werden?
Soll auf die Daten über das Web, bzw. via FTP zugegriffen werden können?
Wie oft werden die Daten gelesen/geschrieben?
Ist der Zugriff konsequent oder random?
- **Datenmanagement:** Sind übliche Metadaten in der Community definiert?
Existieren schon standardmäßige Verfahren und Policies? Sind community-spezifische Datenformate zu beachten?
- **Sicherheitsanforderungen:**
Sind die Anwendungen und Daten sicherheitskritisch?
Wie viel können Datenverluste und andere Sicherheitsvorfälle eigentlich dem Nutzer kosten?
- **Organisatorische Rahmenbedingungen:**
Müssen bestimmte Regelungen (Land, DFG, EU Projekte) in Sachen Sicherheit, Nutzerverwaltung usw. eingehalten werden?
- **Personal und Finanzierung**

Je nachdem, ob verteiltes Datenmanagement oder effiziente Nutzung von Rechenkapazität im Mittelpunkt steht, wird von einem Daten-Grid oder einem Compute-Grid gesprochen.

2.1 Compute-Grid

Beim Compute-Grid befindet sich im Fokus die Möglichkeit, Rechenjobs zwischen verschiedenen Rechenzentren zu verteilen. Dies wird von Grid-Middlewares wie Globus oder gLite unterstützt. Der Hauptanwendungsfall ist die Kooperation mehrerer Rechenzentren zum Zwecke der Lastverteilung. Am meisten profitieren davon jedoch kleinere Forschungsgruppen, die selber keine vollständige Recheninfrastruktur aufbauen können.

Das Grid ist nur dann nützlich, wenn dort Community-spezifische Anwendungssoftware zur Verfügung steht. Allein Rechenkapazität zu teilen reicht nicht. Dann könnten sich kleinere Forschungsgruppen theoretisch auch Rechenzeit bei kommerziellen Anbietern kaufen.

Es ist wichtig, dass im Compute-Grid ein Rechenzentrum die Führungsrolle übernimmt. Ein Beispiel für ein gut geführtes Compute-Grid ist das LHC Grid [4]. Dort leitet das CERN als Tier-0 Zentrum Betrieb und Entwicklung. Große Tier-1 Zentren leisten dabei bedeutende Beiträge. Kleinere universitäre Forschungsgruppen profitieren stark vom Grid, weil sie dort physikalische Analysen ohne zu großen administrativen Aufwand laufen lassen können. Grid-Infrastrukturen, in denen kein Partner eine führende Rolle übernimmt und die Hauptverantwortung für den Betrieb trägt, scheinen viel schwieriger aufzubauen und zu betreiben zu sein.

Compute-Grids können auch eingesetzt werden, um spezifische Hardware, wie z.B. GPU's, zur Verfügung zu stellen. Solche Systeme sind aber noch wenig verbreitet.

Die Wirtschaftlichkeit des Compute-Grid wurde in letzter Zeit durch sinkende Hardware-Preise und entstehende Cloud-Computing- und Virtualisierungstechnologien begrenzt. Wikipedia zufolge [1] betrug der Preis pro GFLOP im Jahr 2000, als das Grid-Konzept etwa entstanden ist, um \$ 1000, im Jahr 2010 aber nur noch etwa \$ 0.10 - \$ 0.50. Es wird erwartet, dass die Anzahl der Cores pro Prozessor weiter steigen und der Bedarf an föderiertem Computing im Nicht-Hochleistungsbereich entsprechend zurückgehen wird.

2.2 Daten-Grid

Ein Daten-Grid ist eine Infrastruktur, in der eine oder mehrere Nutzer-Gruppen ihre Daten verteilt managen und auf die Daten zugreifen können. Dabei kann die Infrastruktur auch zur Datenverarbeitung genutzt werden, diese steht aber nicht im Vordergrund. Beispiele für solche Infrastrukturen sind:

- Klimadatenarchive
- Archive für medizinische Daten
- Archiv für Primärdaten aus Synchrotron-Strahlenquellen

Beim Aufbau solcher Infrastrukturen ist Folgendes zu beachten:

- Bei großen Datenmengen muss geprüft werden, ob Netzwerk-Bandbreite, Storage-Throughput usw. ausreichen. Beispielsweise können die Daten von Experimenten wie LHC [4] oder XFEL [5, 6] nur in einer spezifischen Grid-Topologie und Netzwerk-Infrastruktur gespeichert werden.

- Sollen die Daten in lokalen Rechenzentren repliziert werden? Bei solchen föderierten Lösungen ist es wichtig, dass die lokalen Sites eine ähnliche Infrastruktur besitzen (Metadaten-Datenbanken, Zugriffsprotokolle, AAI usw. unterstützen).
- Der Zugriff soll benutzerfreundlich sein, z.B. über ein Web-Portal.

2.3 Prozesse

In vielen Communities sind die IT-Prozesse nicht genügend formalisiert und standardisiert, um sie auf einer gemeinsamen Plattform wie einem Grid ausführen zu können. Dies muss geschehen, bevor von einer Grid-Lösung gesprochen werden kann. Solche Prozesse sind

- Nutzer-Management
- Software-Management
- Sicherheits-Policies

Die genannten Prozesse sind nicht nur von der Funktionalität her Voraussetzung für den Grid-Betrieb, sondern bergen auch viele Risiken – der Aufwand, manche IT-Prozesse zu betreiben, wird oft nicht rechtzeitig erkannt bzw. unterschätzt. Grid-enabled Workflow Engines wie GWES [7] können dabei helfen, Policies und Workflows technisch in Grid umzusetzen.

2.4 Software-Entwicklung, Adaptierung von existierender Software

Es kommt sehr selten vor, dass kein Bedarf an Software-Entwicklung besteht.

- Community-Grids können sich in existierenden Software-Stacks bedienen (z.B. iRODS [9], Fedora [10], Globus [11]). Es besteht aber fast immer Bedarf an Nutzer-Interface-Entwicklung (wie z.B. Web-Portale).
- Für spezifische Anwendungsfälle kann es einen großen Bedarf an Entwicklungen geben.

Bei geringem Bedarf an Entwicklung (z.B., wenn nur ein einfaches Web-Interface zum Daten-Archiv und einfache Workflows benötigt werden), können existierende Lösungen relativ mühelos adaptiert werden. Hier können neue Communities Hilfe beispielsweise bei WissGrid-Experten suchen. Die Entwicklung komplexer Software soll aber innerhalb der Communities geschehen.

2.5 Testumgebungen

Um die Funktionalität von Grid-Infrastrukturen besser zu verstehen und den Entwicklungs- und Betriebsaufwand schätzen zu können, brauchen neue Communities eventuell eine geeignete Testumgebung. WissGrid selbst hat keine Testumgebung, es können aber unter bestimmten Voraussetzungen die Referenzinstallation des D-Grid oder die dech VO des DECH-Verbundes für Tests genutzt werden. Eventuell kann die neue Community auch das Testgrid der betreuenden Community nutzen. Um Zugang zu solchen Testumgebungen zu bekommen, sollten sich neue Communities an das Fachberater-Team wenden.

2.5.1 Die D-Grid-Referenzinstallation

Das D-Grid ist ein Zusammenschluss einer Vielzahl von Sites. Die Referenzinstallation bildet ein Site mit allen unterstützten Software-Komponenten in Form eines Prototypen (D-Grid Standard-Site) ab.

Den Prototypen können Nutzer und Entwickler von D-Grid-Projekten zum Testen nutzen. Hierzu ist lediglich eine D-Grid-Registrierung erforderlich, weil dieselbe Gridmap-Datei wie im D-Grid verwendet wird. Die virtuellen Organisationen des D-Grid sind auch im Prototypen vorhanden. Außerdem dient die D-Grid-Referenzinstallation dem D-Grid zur Reproduktion von Fehlern in einem definierten Umfeld und als Testplattform für neue offiziell bereitgestellte Software-Versionen.

Die D-Grid-Referenzinstallation besteht nämlich nicht nur in dem Prototypen, sondern bietet auch Software an. Unter der Webadresse

<http://dgiref.d-grid.de>

werden Download und Dokumentation der unterstützten Software angeboten:

- Die drei Grid-Middlewares des D-Grid : Globus [11], [12] und [13]
- dCache [14] und OGSA-DAI [15] als Datengrid-Software
- Die Betriebssysteme Scientific Linux und OpenSUSE Linux
- Cluster-Software

Die Referenzinstallation wird auch künftig zur Verfügung stehen. Sie wird ab Ende 2010 nicht mehr wie zuvor zweimal jährlich, sondern nur noch bei Bedarf (Security-Fixes) erneuert. Bevor ein Release zur neuen Referenzinstallation erklärt wird, wird dieses auf einer von der Referenzinstallation getrennten Plattform getestet. Beta-Releases dienen nur noch zum Testen von Security-Fixes, so dass das stabile Release im Wesentlichen so bleibt, wie es ist.

2.5.2 dech VO

Die DECH-Region besteht aus den Staaten Deutschland und der Schweiz, die sich im Rahmen der internationalen Großprojekte EGEE und EGI zu einem Regionalverband zusammengeschlossen haben und eng zusammenarbeiten. Die dech VO ist ein Testbed für deutsche und schweizer ROC (Regional Operation Center) Mitglieder. Eine Community, die hier testen will, sollte international ausgerichtet sein. Die dech VO eignet sich auch, um erste Erfahrungen im Grid-Umfeld zu sammeln, und ist vergleichsweise gut dokumentiert mit einfachen [Tutorials](#) und [Support-Materialien](#).

Die schweizer Seite von DECH bietet darüber hinaus an, Federated IDs (Shibboleth) zu verwenden, um daraus Short-Lived User Certificates (SLCS) zu generieren. Diese Zertifikate haben eine Lebensdauer von 11 Tagen und ersparen die Generierung von persönlichen Grid-Zertifikaten und die damit verbundenen Probleme.

2.6 Deployment

Die produktive Umsetzung von Grid-Infrastrukturen soll außerhalb von WissGrid geschehen.

2.7 Support

Der Support durch WissGrids Fachberater-Team ist eingebettet in eine vernetzte Support Struktur (für mehr Infos sieh auch WissGrid Deliverable 2.1.8 [16] "Blaupausen für den Aufbau einer technischen Support-Infrastruktur").

2.7.1 WissGrid Support

WissGrid stellt den neuen Communities eine zentrale Anlaufstelle für Support-Anfragen bereit. Primär können Anfragen an die E-Mail-Adresse fachberater@wissgrid.de geschickt werden. Abonnenten dieser Mailing-Liste sind die Fachberater-Teams.

WissGrid ist als Support-Unit in den D-Grid User Support (DGUS) eingebunden und somit auch über das Helpdesk des D-Grid erreichbar. Tickets können im NGI-DE-Portal [17] geöffnet werden. Beim Aufruf des NGI-DE-Portals wird ein persönliches Zertifikat abgefragt. Wer noch kein Zertifikat hat, kann das Portal aber trotzdem nutzen. Für den Zugang zum NGI-DE-Portal ohne Zertifikat benötigt man eine Nutzerkennung und ein Passwort. Um diese zu bekommen, muss ein Webformular ausgefüllt werden.

Bei rein nationalen Community-Grids werden die Anfragen in der Regel innerhalb von WissGrid, DGUS oder den Community-eigenen Support-Strukturen behandelt. Einige Communities wie zum Beispiel die Photonenphysik sind hingegen grundsätzlich international ausgerichtet. Instrumente und Services der verschiedenen Einrichtungen wie zum Beispiel EMBL [18] werden von Wissenschaftlern weltweit genutzt. Ihre Support-Anfragen können daher nicht immer im (nationalen) Rahmen von WissGrid oder DGUS abgehandelt werden.

2.7.2 Community Support

Die Community-Grids sollen mittelfristig ihre eigenen Support-Strukturen aufbauen, mit Unterstützung von WissGrid. Der Community-spezifische Support ist essenziell, denn Applikations- oder Experiment-spezifische Probleme können nur innerhalb der Community selbst gelöst werden. Insbesondere wird auf diese Weise auch gewährleistet, dass Anfragen an entsprechende Bulletin-Boards und Ticketing-Systeme zielgerichtet weitergeleitet werden können.

3 Aktivitäten und Ergebnisse

3.1 Forschung mit Photonen

Die Photonphysik Communities haben bislang kaum Erfahrungen mit dem Grid. Grundsätzlich neigen diese Communities dazu, lokale Infrastrukturen zu nutzen, so unzugänglich diese auch sein mögen. Durch den Aufbau neuer Lichtquellen und Entwicklungen in der Detektor-Technologie werden die alten Probleme – wie Mangel an Speicherkapazität und ineffizientes Daten-Management –, bereits jetzt dramatisch verschärft.

Bislang lag das Daten-Management weitgehend in der Hand der Anwender oder Experimentatoren. Nach erfolgtem Experiment werden die Daten entweder über das Internet in eines oder mehrere Heimatinstitute transferiert oder auf portable Festplatten kopiert und verschickt. Lokale Kopien werden nach einer nicht genau definierten Zeit gelöscht. Es existieren mit anderen Worten keine Strukturen um die langfristige Archivierung von Daten auch nur ansatzweise zu gewährleisten. Es gibt keine konsistente Erfassung von Metadaten und keine ausreichende Computing-Infrastruktur, um eine zeitnahe Analyse zu garantieren.

In den kommenden Monaten und Jahren werden die Datenraten dramatisch ansteigen und können ein Petabyte pro Woche überschreiten. Um mit den Entwicklungen Schritt zu halten und eine langfristige Verfügbarkeit der Daten auch im Sinne einer Nachnutzung zu ermöglichen, haben das EMBL [18] und das European XFEL Consortium [5] begonnen, Möglichkeiten zu eruieren. Eine Grid-basierte Lösung bietet viele Vorteile. Neben einer konsistenten und erprobten Daten-Managementstruktur erlaubt es die nahtlose Integration von Computing Infrastrukturen, die die Möglichkeiten für die beteiligten Wissenschaftler deutlich verbessert. Daher haben sowohl EMBL als auch der European XFEL starkes Interesse an der Beratung und Unterstützung durch WissGrid.

Im Folgenden sollen einige der grundsätzlichen Überlegungen skizziert und Aktivitäten der Fachberater dokumentiert werden.

3.1.1 Virtuelle Organisationen

EMBL hatte einige Anwender, die reine Bio-Informatik Anwendungen auf dem Grid realisieren wollten. Im Wesentlichen handelt es sich um Sequence-Alignment, 2D- und 3D-Strukturvorhersagen und Generierung von Struktur-Templates biologischer Makromoleküle. Innerhalb der biomed-VO existiert bereits eine Software-Umgebung, die praktisch alle Voraussetzungen mitbringt, um solche Anwendungen mit minimalem Aufwand im Grid zu realisieren. Daher haben wir in diesem Fall die biomed-VO, die von DESY auch unterstützt wird, als Einstiegspunkt empfohlen. Grid-Zertifikate wurden über DESY-RA erstellt. EMBL hat mittlerweile diese Applikationen im Grid in Produktion genommen.

Generell stellt sich aber die Frage, wie die VOs zu organisieren sind. Es würde sich zum Beispiel anbieten, VOs spezifisch für wissenschaftliche Disziplinen einzurichten. EMBL zum Beispiel bietet Experimente im Umfeld von Kleinwinkelstreuung (SAXS - Small Angle X-ray Scattering) und Protein-Kristallographie (MX - Macromolecular X-ray crystallography) an. Diese Experimente erfordern sehr unterschiedliche Analysis-Umgebungen. Andererseits finden sich zunehmend Anwendungen, die beide Disziplinen miteinander kombinieren, die also von einem einheitlichen Framework profitieren würden. European XFEL hingegen ist primär daran interessiert das Datenmanagement im Grid zu

realisieren. In diesem Fall hätte eine Facility-spezifische VO ihre Vorteile.

Um den Einstieg zu erleichtern, bieten wir verschiedene Test-Umgebungen an, die zum Teil spezifisch für die Photonenphysik eingerichtet wurden:

- DECH VO
Ist ein Testbed für deutsche und schweizer ROC (Regional Operation Center) Mitglieder. Es eignet sich um erste Erfahrungen im Grid-Umfeld zu sammeln.
- DESYVO
Als lokale VO bietet DESY diese VO generell als Testumgebung an. Als Betreiber der VO kann DESY insbesondere in den ersten Phasen sehr zeitnahen Support bei Problemen anbieten und die lokalen Computing und Storage Elemente sind unmittelbar ins lokale Netzwerk eingebunden, was bisweilen hilft die initialen Frustrationen zu minimieren.
Insbesondere für die Photonenphysik hat DESY ebenfalls eine AMGA-Instanz [19] mit Oracle als Backend implementiert. Dies dient als Testumgebung für FLASH und später XFEL um die automatische Registrierung und Annotierung der experimentellen Daten zu testen. Angesichts der grossen Datenmengen und Anzahl der Dateien, die pro Sekunde generiert werden, sind Performance-Test hilfreich um sich langfristig für eine Lösung zu entscheiden.
- XFEL VO
Für die Belange der European XFEL wurde eine spezifische VO eingerichtet. Bislang gab es wenig Aktivitäten in diesem Umfeld.
- PS VO
VOs definieren nicht nur in gewissem Maße den Einstiegspunkt sondern bieten auch eine Sichtbarkeit nach aussen. DESY wird daher eine generische VO für die Photonenphysik einrichten und betreiben. Die Details müssen allerdings noch geklärt werden.

3.1.2 Authentifizierung

In der Regel erfolgt die Authentifizierung im Grid über persönliche Grid-Zertifikate. Authorisierungen können z.B. über GridMaps oder Rollenbasiert vergeben werden. Alternative Authentifizierungs-Instanzen mit OpenID oder Shibboleth/SAML sind bislang nicht sehr verbreitet im Grid. Die Akzeptanz für Grid-Zertifikate ist allerdings in einigen Communities nicht sehr ausgeprägt. Im TextGrid wird eine Shibboleth/SAML basierte Authentifizierung bevorzugt. C3Grid vermeidet die Verwendung von Zertifikaten und setzt im Rahmen Earth System Grids (ESG) auf eine OpenID basierte Lösung [21]. In einigen Diskussionen mit vielen verschiedenen Vertretern der Photonenphysik zeigen sich zwei konträre Standpunkte. Die einen vertreten die Auffassung, dass kein Weg um Zertifikate herumführt und das generell eine sinnvolle Angelegenheit ist. Die anderen finden die Notwendigkeit Grid-Zertifikate zu verwenden komplett inakzeptabel.

Die Lichtquellen am DESY, FLASH, European XFEL, Petra3 und DORIS, werden rund 5000 Anwender pro Jahr zu bedienen haben, von denen rund 80% jedes Jahr neu hinzukommen. Es wären also jedes Jahr rund 4000 Grid Zertifikate zu generieren, was mit erheblichem Aufwand verbunden wäre. Zur Zeit diskutierte Alternativen basieren auf Shibboleth und OpenID, die beide ihre Vor- und Nachteile mit sich bringen. OpenID ist einfach zu implementieren und cross-Site Trusts einfach zu realisieren. OpenID hat allerdings kaum Mechanismen, eine einmal aufgebaute Authentifizierungskette wieder abzubauen. Shibboleth bietet Federated IDs. Das ist aber in einem internationalen Umfeld nicht einfach zu realisieren, da es kaum Mechanismen gibt, Trusts über nationale Grenzen hinweg aufzubauen.

ESRF, PSI und das Helmholtzzentrum Berlin versuchen zur Zeit einen Shibboleth-basierten Prototyp zu implementieren, der die nötigen Voraussetzungen erfüllt. DESY verfolgt diese Bemühungen und ist durch ein pan-Europäisches Projekt namens [PaN-data](#) direkt involviert.

3.1.3 Prototypische Anwendungen

Wie erwähnt, gibt es eine Vielzahl sehr unterschiedlicher Anwendungen, die für die Photonophysik von Interesse sind. Nicht alle diese Anwendungen lassen sich sinnvollerweise im Grid realisieren. Eine der Aufgaben ist es also, Anwendungen zu identifizieren, die sich gut eignen (das sind nicht-graphische nicht-echtzeit Anwendungen, die relativ viel Rechenkapazität brauchen und keine Lizenzbeschränkungen haben). DESY hat mit EMBL und European XFEL einige Applikationen ausgewählt, die als prototypische Anwendungen implementiert werden sollen. Diese können dann wiederum als Demonstrations-Projekte und zur Dokumentation dienen. Im Rahmen der Gespräche mit Vertretern der Photonophysik haben sich, neben dem Datamanagement, einige Anwendungen herauskristalliert, die im weiteren beschrieben werden sollen.

Generischer Prototyp Anhand der Materialien-Sammlung (D2.1.1 [20]) und den daraus extrahierten Architektur-Konzepten wurde ein generisches Konzept für die Photon Sciences erstellt (siehe Abb. 1). In Gesprächen mit den beteiligten Partner aus den Photon Science Communities hat sich gezeigt, dass diese generische Architektur als Test-Plattform den zur Zeit formulierten Anforderungen genügen kann. Erfahrungen mit dem Testbed in Kooperation mit ESRF und PSI haben gezeigt, dass nur die konkrete Implementierung einer Grid-Infrastruktur und den Anwendungen Aufschluss über das optimale Modell geben wird.

Gespräche mit EMBL legen nahe, dass mittelfristig eine Erweiterung der Architektur um einige Komponenten wünschenswert werden könnten (z.B. EDNA Workflow Engine). EMBL strebt an die verschiedenen Instrumente und Prozesse in eine einheitliche, integrierte Umgebung einzubetten, so dass Informationen und Daten aus so verschiedenen Bereichen wie Nasslabor, Kristallisation oder Kleinwinkelstreuung verarbeitet werden können, wobei insbesondere die Remote-Steuerung von Instrumenten zunehmend an Bedeutung gewinnt. Dies wird für den Aufbau des Prototypen noch nicht unmittelbar von Interesse sein, allerdings sollte die Infrastruktur hinreichend modular aufgebaut sein, so dass eine spätere Integration physikalischer Instrumente und von Schnittstellen zu den wichtigsten Datenbanken problemlos implementiert werden könnte.

European XFEL Datamanagement Der European XFEL befindet sich zur Zeit in der Konstruktionsphase, die IT-Infrastruktur wurde jedoch noch nicht komplett definiert. Die groben Rahmenbedingungen wurden in einem Computing Technical Design Report (C-TDR) dokumentiert. Darin sind die Anforderungen an das Datenmanagement basierend auf Erfahrungen von FLASH und LCLS dargelegt. Die Computing-Anforderungen hängen stark von den avisierten Experimenten und den Analyse-Methoden ab. Diese lassen sich zur Zeit nicht abschliessend abschätzen, so dass der Fokus auf dem Datenmanagement und der Archivierung liegt. Der European XFEL beabsichtigt das Datenmanagement in eine Grid-Infrastruktur einzubetten und dCache als Storage-Backend zu nutzen. Authentifizierung und Authorisierung sollen zumindest anfänglich auf X.509 PKI basieren.

Da der XFEL noch im Aufbau begriffen ist, fallen zur Zeit keine experimentellen Daten an. Allerdings werden Daten bereits am LCLS und insbesondere in Simulationen des Strahlungsfeldes im

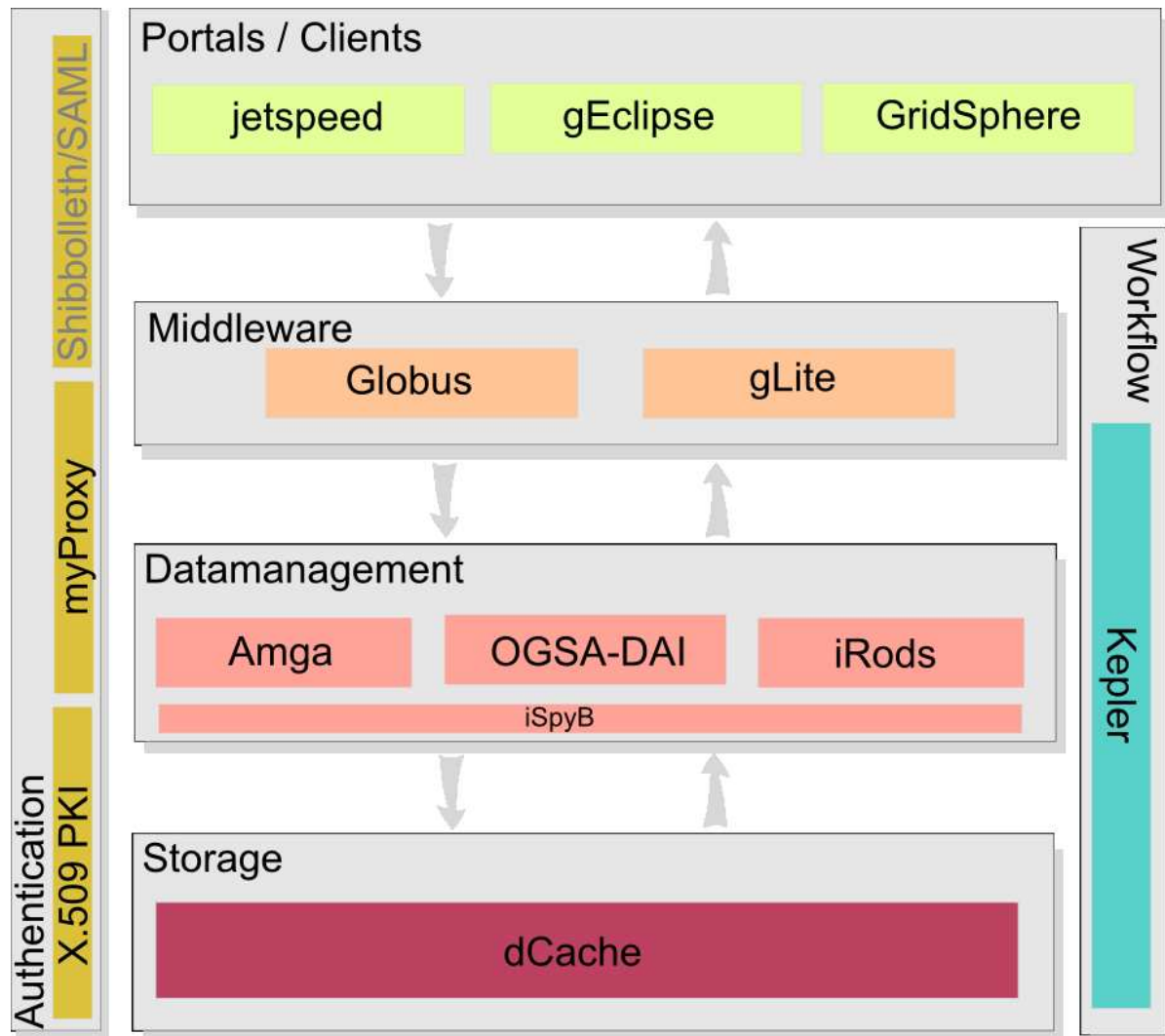


Abbildung 1: Generische Architektur-Konzept für Photon Sciences

XFEL. Diese Simulationsdaten sind Grundlage für weitergehende Analysen, die den Bau des XFEL beeinflussen. Die Simulationsdaten sowie alle Metadaten sollen daher als Public Domain bereitgestellt werden. Sekundäre Simulationen müssen mit den Original-Daten verknüpft und annotiert werden.

Die Simulationsdaten liegen in Dateien von 50-250GB Größe pro Datei vor. Das Dateiformat basiert auf Fortran Stream I/O, die dazugehörigen Metadaten sind in den Header-Records der Dateien selbst abgelegt. Die Dateigrößen und Formate erschweren die Handhabung erheblich. Daher sollen die Daten in HDF5 [22] umgewandelt werden, die Metadaten in einer Metadatenbank abgelegt werden. Die Daten sollen über ein Portal lokalisierbar und zugreifbar gemacht werden.

Da XFEL erwägt das Data Acquisition (DAQ) System auf dem Glassfish Application Server aufzubauen erlauben java-basierte Portal-Lösungen (wie z.B. Liferay [25]) eine einfache Integration von Datenmanagement und DAQ. Ebenso werden zunächst keine Workflow-Frameworks benötigt. Daher lässt sich das Architektur-Konzept in diesem Falle auf einige wenige Komponenten reduzieren, die sich aber vollständig aus dem generischen Photon Science Konzept ableiten lassen (Abb. 2).

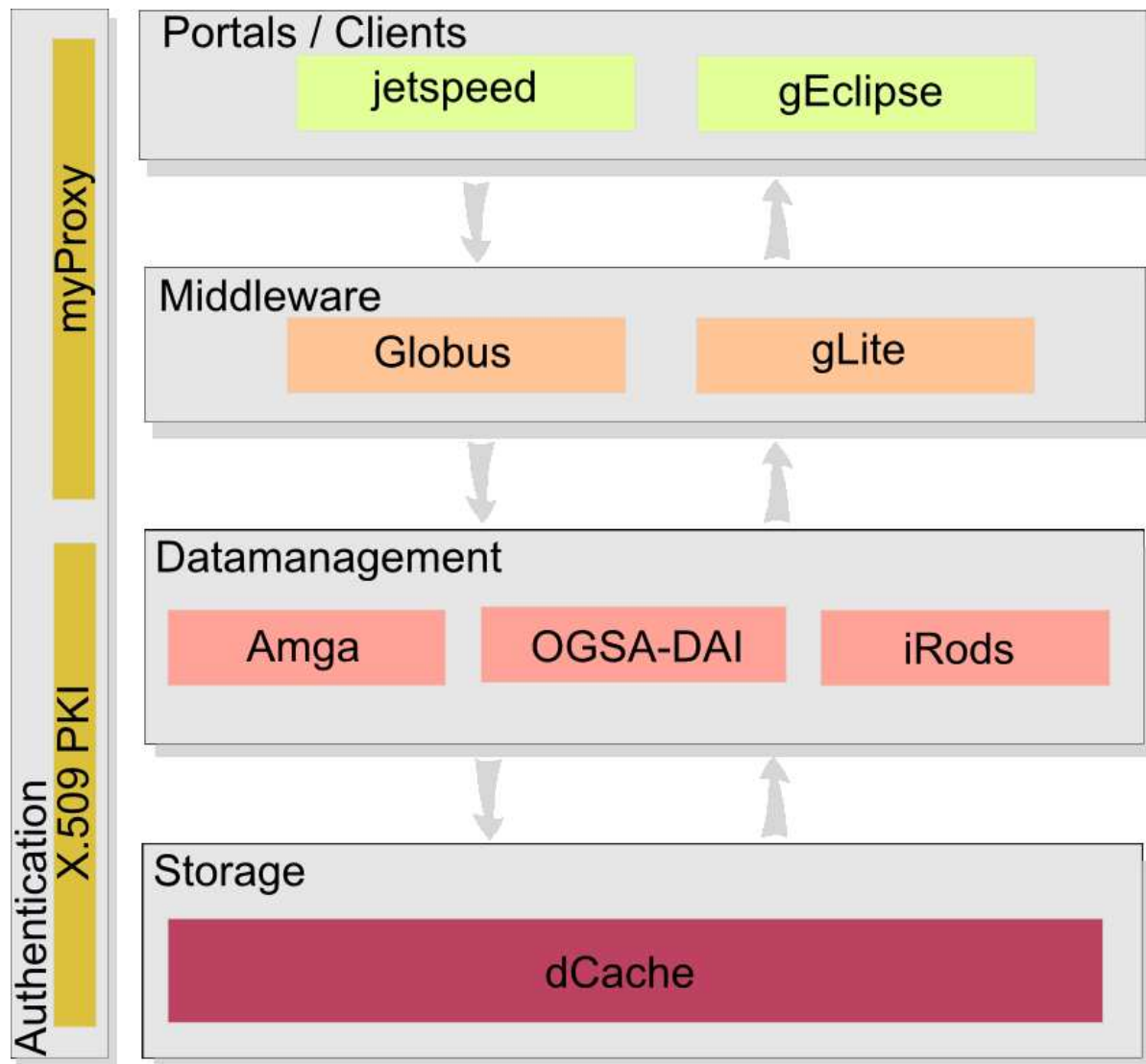


Abbildung 2: Reduziertes Architektur-Konzept für European XFEL

EMBL Daten- und Compute-Grid Im Rahmen des EU-Projektes ESRFUp [26] wurde untersucht, inwieweit das Grid sich für Photon Science spezifische Anwendungen nutzen ließe. Diese Evaluierung fiel durchweg negativ aus, was im wesentlichen auf vier Gründe zurückzuführen ist:

- Datenmanagement: Datentransfer-Tests wurden auf den sequentiellen Transfer vieler sehr kleiner Dateien beschränkt. Es ist nicht überraschend, dass Latenzen sich in diesem Fall negativ bemerkbar machen.
- Computing: natürlich profitiert nicht jede Anwendung automatisch von der Migration ins Grid. Eine Evaluierung der Grid-Tauglichkeit von Anwendungen wurde aber unterlassen.
- Grid-Architektur: die ESRFUp Grid-Infrastruktur hat sich ausschliesslich auf EGEE und gLite als Middleware beschränkt ohne deren Konzepte oder mögliche Alternativen auf ihre Eignung für das spezifische Anwendungsfeld zu prüfen.
- Community: ESRFUp hatte die individuellen Wissenschaftler, die die Experimente ausführen

als Zielgruppe gewählt. Aufgrund der Heterogenität dieser Communities und deren bisweilen recht geringen IT-Grundkenntnisse war diese keine gute Wahl, da die angebotene Grid-Infrastruktur die Anwender schlicht überfordert und frustriert zurück gelassen hat.

Die Materialiensammlung und die daraus konsensierten Blueprints und Konzepte erlauben diese Fehler zu vermeiden. Konsequenterweise wurden für die Photon Science Anwendungen und insbesondere für das EMBL ganz andere Konzepte vorgeschlagen:

- Community: wähle die Beamline- und Facility-Betreiber als Zielgruppe, die eine angepasste Grid-Infrastruktur als Service anbieten. Dies reduziert die Komplexität erheblich und vereinfacht die Definition der benötigten Standards.
- Grid-Architektur: verwende Architektur-Konzepte, die auf die Bedürfnisse der Anwender zugeschnitten ist.
- Computing: eine sorgfältige Auswahl der Anwendungen mindert die Frustration der Service-Anbieter wie der Anwender.
- Datenmanagement: die Verwendung von standardisierten Container-Formaten und die Integration von existierenden Metadaten-Engines hilft Throughput zu erhöhen, und erlaubt die Definition von standardisierten Workflows. De facto Standards für Datenformate in den Photon Science sind HDF5 basiert (wie HDF5 oder Nexus) oder CIF (wie imageCIF) basiert.

- Kristallisation

EMBL betreibt eine Kristallisations-Anlage, die prinzipiell als Service allen Anwendern weltweit zur Verfügung steht. Diese Kristallisations-Anlage ist in der Lage tausende von Kristallisations-Ansätzen parallel zu verarbeiten. Insgesamt werden einige Millionen dieser Versuche parallel durchgeführt. Von jedem Versuch werden täglich typisch drei fotografische Aufnahmen gemacht, die die Grundlage für die Beurteilung der Versuche sind. Täglich sind also einige Millionen dieser Aufnahmen von weltweit verteilten Gruppen zu analysieren.

EMBL möchte zwei Aspekte dieser Analyse im Grid-Umfeld realisieren. Zum einen ist dies die automatische Zustellung der Bilddaten zu den einzelnen Gruppen zum anderen ist die automatisierte Analyse der Daten ein Anliegen.

Die Zustellung der Bilddaten ist technisch unproblematisch. Für die Analyse existieren zur Zeit noch keine entsprechende Algorithmen, was den Vorteil hat, das diese Algorithmen von vornherein mit Blick auf die Porabilität und Grid-Eignung implementiert werden können.



Abbildung 3: Kristallisations-Roboter am EMBL

- AutoRickshaw

EMBL-Hamburg hat eine Software-Suite zur automatischen Kristall-Strukturbestimmung entwickelt. Diese Suite namens [Auto-Rickshaw](#) wird als Service für die Anwender angeboten und kann remote über einen Web-Server gestartet werden [28]. Die Analyse läuft zur Zeit auf einem lokalen 64-core Cluster. Die Verfügbarkeit des Clusters und die Performance sind nicht

mehr ausreichend. Eine Erweiterung des Clusters wäre eine Option, andererseits ist diese Anwendung sehr gut ins Grid portierbar und soll daher prototypisch implementiert werden.

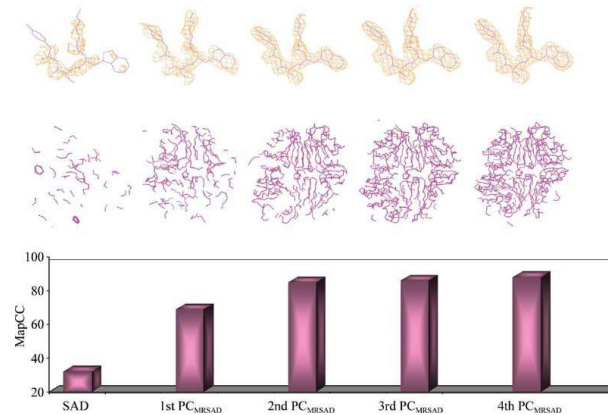


Abbildung 4: Illustration der iterativen Strukturbestimmung in Auto-Rickshaw [29]

Die Analyse basiert auf einem iterativen Prozess, der unter anderem viele Struktur-Templates parallel zur Strukturbestimmung an die experimentellen Daten anpasst. Dieser Prozess lässt sich relativ einfach in eine Vielzahl individueller Einzelprozesse zerlegen, die sich in einen wohl definierten Workflow abbilden lassen.

Diese Klasse von Anwendungen erfordert die Einbindung einer Vielzahl von Datenbanken und Informationsservices (z.B. PDB [31], NDB [32], EMDB [33], EBI [34] oder NCBI [35]). Für einige dieser Datenbanken wie z.B. PDB gibt es eine Java/WSDL-API, die eine Integration in ein Portal wie Liferay vereinfacht, für andere, wie NDB, müsste die Möglichkeiten einer Integration evaluiert werden.

3.1.4 Beratungs-Aktivitäten - Photon Science

Die Wiss-Grid Aktivitäten in den Photon Science Communities beinhalten Arbeitstreffen, Teilnahme an und Organisation von Workshops sowie prototypische Softwareentwicklung. Die nachfolgende Tabelle enthält eine unvollständige Liste der WissGrid-Beratungsaktivitäten.

Datum	Aktion	Beschreibung
2010.09.22-24	ESRF Workshop Grenoble	Teilnahme am mxCUBE/EDNA/ISPYB workshop am ESRF Grenoble; diskussion über potentielle Föderation von Metadatenkatalogen (z.B. ICAT) zwischen ESRF, DESY, Diamond
2010.09.02	Vortrag am HASY-LAB	Vortrag zum Thema Datenmanagement von I. Agapov
2010.07.06	Workshop	Organisieren eines Workshops zum Thema Datenmanagement für Photon Science am DESY mit Representation von Wissgrid und wichtigen Vertretern von Experimentalgruppen. https://indico.desy.de/conferenceDisplay.py?confId=3183
2010.04.21	Meeting mit EMBL	Zugang zu DESY dCache Instanz bereitgestellt. Autoindexing software und Workflows diskutiert.
2010.04.19	Vortrag M. Lindner	Vortrag von M. Lindner am DESY zum Thema <i>iRODS</i> . URL: http://www.desy.de/dvsem/SS10/lindner_talk.pdf

Datum	Aktion	Beschreibung
2010.03.03	PETRAIII meeting	Meeting mit PETRAIII beamline supervisors, um Archivierungsverfahren zu definieren
2010.02.28	CFEL meeting	Teilnahme am CFEL (Centre for Free Electron Laser Science) IT meeting
2010.02.18	Meeting DESY Bibliothek	Diskussion über Integration von Desy Bibliothek Services (Inspire) ins Photon Science Archiv
2010.01.15	Meeting mit dCache Team	Aufsetzen einer dCache Test-Instanz. Bereitstellung von 20TByte. Abklärung der Sicherheitsmechanismen und Anforderungen.
2010.01.15	Meeting mit XFEL	Diskussion der Architektur-Konzepte. XFEL erwägt Glassfish Application Server für das Data Aquisition (DAQ) System einzusetzen. Würde sich gut in das prototypische Framework einfügen. Diskussion der dCache-Instanz.
2010.01.14	Sample Data von JCSG	Das Joint Center for Structural Genomics (JCSG) verfügt über ein (closed) Repository mit MX Rohdaten. Wir haben jetzt Zugang zu den Daten, die als Muster für verschiedenste Aspekte der Archivierung, Datamanagement und Analyse dienen werden. URL: http://www.jcsg.org/
2010.01.14	Meeting mit dCache Team	Für die Implementierung eines Prototypen wird eine SRM-Lösung und physikalischer Speicherplatz benötigt. Die Einrichtung einer dedizierten Photon Science dCache Instanz mit einigen 10TByte Plattenplatz und Tape-Backend wurde verabredet.
2010.01.12	Meeting mit EMBL	Vorstellung der generischen Architektur-Konzepte. Diskussion der zu implementierenden Services, Archivierung und Compute-Grid. Aufsetzen eines Prototypen für SAXS und MX vereinbart. Einrichten einer dCache Instanz diskutiert.
2010.01.11	Vortrag S. Kindermann	Vortrag von S. Kindermann am DESY zum Thema <i>Grid Infrastructures for Climate Data Handling</i> . URL: http://www.desy.de/dvsem/WS0910/
2009.12.10	Meeting mit EMBL	Eingrenzung der Prototyp-Applikationen.
2009.12.04	Meeting mit XFEL	Diskussion zu Datamanagement. Konkrete Anwendung seitens XFEL um massive Simulations-Daten für die Weiterverwertung anbieten zu können. Aufsetzen eines Prototypen mit Portal und dCache als Backend wurden vereinbart.
2009.11.23	Vortrag J. Ludwig	WissGrid Vortrag von J. Ludwig am DESY zum Thema LZA. Beteiligung u.a. von HASYLAB, XFEL und EMBL (alle Photonen). Anschliessendes Meeting mit EMBL. URL: http://www.desy.de/dvsem/WS0910/
2009.11.13	Meeting mit XFEL	Diskussion zu Remote Instruments im Grid und LZA.
2009.11.11	Meeting mit EMBL	Diskussion zu iSpyB, Datenformaten (nexus, hdf5) und (Meta-) Daten-Standardisierung.
2009.10.15	Meeting mit XFEL	Diskussion um WissGrid und Roscoe. Diskussion über Netzwerk-Infrastruktur im Grid-Kontext.
2009.10.07	Meeting mit EMBL	Allgemeine Erörterungen zu Grid, WissGrid und Roscoe

3.2 Sozialwissenschaften

Nach einer vorangegangenen allgemeinen Information über die Grid-Umgebung und Wiss-Grid Aktivitäten wurde im Rahmen der WissGrid-AP3-Begutachtung der Anwendungsfall der Sozialwissenschaften besprochen und vorgestellt. Im Anschluss wurden im Rahmen eines Workshops die Fachwissenschaftlichen Anforderungen an eine virtuelle Arbeitsumgebung für SOEB erarbeitet. Weiterhin wurde WissGrid im Februar 2010 damit beauftragt, eine Expertise für (VirtAug) zur Grid Nutzung zu erstellen.

3.3 Biostatistik und Epidemiologie

3.3.1 Beschreibung

Die Biostatistik wird in diesem Dokument von der Gruppe von Prof. Beißbarth aus der Universitätsmedizin Göttingen repräsentiert. In der Biostatistik werden die im Rahmen klinischer Studien erhobenen Daten ausgewertet, um z.B. Korrelationen zwischen Krankheitsbildern und genetischen Merkmalen zu erforschen. Die Basis dafür sind anonymisierte bzw. pseudonymisierte Daten zu Genetik, Biomaterial, Bildmaterial und Krankheitsbildern, auf denen z.B. Mustererkennungsalgorithmen und statistische Analysen im Grid durchgeführt werden. Ein typisches Szenario ist hierbei der Erhalt bestimmter in einer Studiendatenbank vorgehaltenen Daten und die anschließende Auswertung mithilfe von statistischen Entwicklungsumgebungen wie R oder SAS.

Die Leiterin der Genetischen Epidemiologie der Universitätsmedizin Göttingen, Prof. Bickeböller, ist zugleich 1. Vizepräsidentin der GMDS (Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie). Zudem ist Prof. Bickeböller Mitglied im GMDS-Fachausschuss für Epidemiologie. In der Epidemiologie werden Ursachen und Folgen sowie Verbreitung von Krankheitsbildern in Populationen untersucht. Dabei werden vornehmlich statistische Methoden angewendet. Daher besteht eine große methodische Nähe zur Biostatistik.

3.3.2 Beratungs-Aktivitäten - Biostatistik und Epidemiologie

Datum	Aktion	Beschreibung
2010.10.20	Abschlusspräsentation der IAESTE-Studentin	Die Ergebnisse des IAESTE-Praktikums wurden vorgestellt.
2010.10.05	Finales Angebot von Grid-Ressourcen	Für ein zukünftiges Projekt wurde mit einem Support Letter ein Angebot von Grid-Ressourcen erstellt, welches die Nutzung von Grid-Ressourcen und Schulungsmaßnahmen umfasst.
2010.09.28	Treffen mit A. Leha	Arbeitstreffen zur Diskussion der bisherigen Ergebnisse des IAESTE-Praktikums. Die weiteren Schritte wurden festgelegt.
2010.09.06	Gemeinsame Publikation mit der Genetischen Epidemiologie im Rahmen der GMDS-Tagung 2010.	In der Publikation werden die Ergebnisse der Zusammenarbeit mit der Genetischen Epidemiologie beschrieben und die Vorteile einer Einbindung in eine virtuelle Forschungsumgebung herausgearbeitet. http://www.gmds2010.de/cms/wordpress/wp-content/uploads/fohlen/227.pdf
2010.08.23	Treffen mit A. Leha	Abstimmung zum Start des gemeinsamen Projekts. Konkrete Anwendungsfälle der Gruppe von Prof. Reißbarth wurden ausgetauscht. Anschließend erfolgt eine gemeinsame Betreuung der IAESTE-Studentin aus China.
2010.05.05	Treffen mit Prof. Reißbarth, A. Leha	Besprechung gemeinsamer Aktivitäten in Bezug auf die Nutzung von Grid Computing in der Gruppe von Prof. Reißbarth. Initiierung eines gemeinsamen Arbeitsprojekts mit einer IAESTE-Studentin aus China.
2010.04.01	Treffen mit M. Sohns	Informationsaustausch über einen Anwendungsfall aus der Genetischen Epidemiologie. Dieser wurde lokal an der UMG umgesetzt und in einer Publikation beschrieben. Darin wurden die Vorteile einer Einbindung in eine virtuelle Forschungsumgebung herausgearbeitet.
2010.02.25	Treffen mit der gesamten Gruppe von Prof. Reißbarth	Beiderseitige Vorstellungsrunde: Beschreibung der Anwendungsgebiete der Gruppe von Prof. Reißbarth, konkreten Anwendungen und Anforderungen in Bezug auf Rechenleistung und Datenmanagement. Seitens Grid-Gruppe erfolgte eine Darstellung der Möglichkeiten der Nutzung von virtuellen Forschungsumgebungen mittels Grid Computing.
2009.10.05	Angebot an Prof. Bickeböller gesendet	Ein Support Letter über das Angebot von Grid-Ressourcen für ein zukünftiges Projekt wurde erstellt und an Prof. Bickeböller gesendet. Das Angebot umfasst ebenso einen Workshop als Schulungsmaßnahme, um die angebotenen Grid-Ressourcen optimal für das Projekt einsetzen zu können. Der angebotene Workshop umfasst eine zweitägige Schulung durch Spezialisten, die mit der angebotenen Rechenverbund-Infrastruktur vertraut sind und bereits Erfahrung in der Schulung von Nutzern aus dem Umfeld der Biomedizin haben. Die angebotenen Rechenressourcen werden aus dem deutschlandweiten Rechenverbund D-Grid bereitgestellt und umfassen die darin für die biomedizinische Community nutzbaren Ressourcen.
2009.09.25	Treffen mit Prof. Reißbarth	Information über WissGrid-Aktivitäten vermittelt. Im Gegenzug wurden anhand eines Fragebogens Informationen über die Anforderungen aus der Biostatistik ausgetauscht.

Literatur

- [1] <http://en.wikipedia.org/wiki/FLOPS>
- [2] Standard Performance Evaluation Corporation <http://www.spec.org/>
- [3] <http://www.r-project.org/>
- [4] Worldwide LHC Computing Grid <http://lcg.web.cern.ch/lcg/>
- [5] The European XFEL
<http://www.xfel.eu/>
- [6] The European XFEL Computing Technical Design Report (TDR)
http://www.xfel.eu/sites/site_xfel-gmbh/content/e63617/e68670/european-xfel-computing-tdr-eng.pdf
- [7] GWES Grid Workflow <http://www.gridworkflow.org/snips/gridworkflow/space/GWES>
- [8] Kepler workflow
<https://kepler-project.org>
- [9] iRods
<https://www.irods.org/>
- [10] Fedora Commons
<http://fedora-commons.org/>
- [11] Globus middleware
<http://www.globus.org/toolkit/>
- [12] gLite middleware
<http://glite.web.cern.ch/glite/>, <http://en.wikipedia.org/wiki/GLite>
- [13] unicore middleware
<http://unicore.eu>,
- [14] dCache
<http://www.dcache.org/>
- [15] OGSA-DAI
<http://www.ogsadai.org.uk/>
- [16] Blaupausen für den Aufbau einer technischen Support-Infrastruktur
<http://www.wissgrid.de/publikationen/deliverables/wp2/Support-Infrastruktur.pdf>
- [17] NGI-DE-Portal
<https://helpdesk.ngi-de.eu/>
- [18] EMBL: European Molecular Biology Laboratory, Outstation Hamburg
<http://www.embl-hamburg.de/>
- [19] AMGA
amga.web.cern.ch/amga/

- [20] Community-Grids - Überblick und Report <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2>.
- [21] F. Siebenlist, R. Ananthkrishnan, D. E. Bernholdt, L. Cinquini, I. T. Foster, D. E. Middleton, N. Miller, D. N. Williams: Earth System Grid Authentication Infrastructure: Integrating Local Authentication, OpenID and PKI, *The 2009 TeraGrid Conference*, www.teragrid.org/tg09/files/tg09_submission_79.pdf
- [22] HDF5 data model
<http://www.hdfgroup.org/HDF5/>
- [23] GridSphere
<http://www.gridsphere.org>
- [24] Jetspeed
<http://portals.apache.org/jetspeed-2/>
- [25] Liferay portal
<http://www.liferay.com/>
- [26] ESRF-Up WP11, EGEE feasibility study
<http://www.esrf.eu/Infrastructure/Computing/Grid>
- [27] GridShib: Bridging SAML/Shibboleth and X.509 PKI for campus and grid interoperability
<http://gridshib.globus.org/>
- [28] Auto-Rickshaw: The EMBL-HH Automated Crystal Structure Determination Platform
<http://www.embl-hamburg.de/Auto-Rickshaw/>
- [29] Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA. Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica D*61:449-457, 2005.
- [30] Information System for Protein CrystallographY Beamlines (ISPYB)
<http://www.esrf.eu/UsersAndScience/Experiments/MX/Software/ispvb>
- [31] PDB: The Protein DataBase
<http://www.pdb.org/>
- [32] NDB: The Nucleic acid DataBase
<http://ndbserver.rutgers.edu/>
- [33] The Electron Microscopy DataBase
www.ebi.ac.uk/pdbe/emdb/
- [34] EBI: European Bioinformatics Institute
<http://www.ebi.ac.uk/>
- [35] NCBI: The National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/>
- [36] GlassFish: open source application server implementing Java EE
<https://glassfish.dev.java.net/>
- [37] PaNdata: Photon and Neutron Data Infrastructure
http://www.pan-data.net/Main_Page