



### Arbeitspaket 2: Blaupausen und Beratung

## Aktivitäten von WissGrid-Fachberater-Teams<sup>1</sup>

Deliverable	2.2.1 Aktivitäten von WissGrid-Fachberater-Teams
Autoren	Arbeitspaket 2: Blaupausen und Beratung
Editoren	Joachim Schultes, Stephan Sandvoss
Datum	01-10-2010
Dokument Version	1.0.0

### A: Status des Dokuments

Deliverable 2.2.1, Version 1.0.0, Release

**Hinweis:** Bei diesem Dokument handelt es sich um eine zentrale Dokumentation der Fachberater Aktivitäten im WissGrid Projekt und wird daher fortlaufend aktualisiert und erweitert.

### B: Bezug zum Projektplan

Dieses Deliverable ist eine erste Übersicht über die in der Startphase des WissGrid in Bezug auf die vorhandenen Aktivitäten und Verbindungen.

---

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

**B: Änderungen**

<b>Version</b>	<b>Date</b>	<b>Name</b>	<b>Brief summary</b>
0.0.0	08.4.2010	S. Sandvoss	Erstellung des Arbeitsdokumentes
0.0.1	21.4.2010	J. Schultes	Überarbeitung der Gliederung
0.0.2	30.4.2010	J. Schultes	Definition der Kernaufgaben der FB
0.1.0	04.5.2010	J. Schultes	Einpflegen der FB Inhalte für PS
0.1.1	14.5.2010	J. Schultes	Einpflegen der FB Inhalte für SW
0.1.2	24.5.2010	J. Schultes	Einpflegen der FB Inhalte für Klimafolgenforschung und Altertumswissenschaften
0.2.0	31.5.2010	J. Schultes	Beendigung der ersten Version für die Überarbeitung im AP2
0.2.1	01.10.2010	T. Rathmann	Korrekturen
1.0.0	01.11.2010	H. Enke	Release

# Inhaltsverzeichnis

<b>1</b>	<b>Kernaufgabe der Fachberater</b>	<b>4</b>
<b>2</b>	<b>WissGrid-Fachberatertätigkeiten</b>	<b>6</b>
2.1	Photon Science . . . . .	6
2.2	Sozialwissenschaften (VirtAug) . . . . .	9
2.3	Alttertumswissenschaften . . . . .	9
2.4	Radio . . . . .	10
2.5	Klimafolgenforschung . . . . .	10
<b>3</b>	<b>Zusammenfassung</b>	<b>11</b>

# Kapitel 1

## Kernaufgabe der Fachberater

Das WissGrid-Projekt sieht sich als Bindeglied zwischen der ersten Idee einer Nutzung des Grids für die betreffenden akademischen Communities und des bereits existierenden Basis-Supports des DGI für den Betrieb, das primär Unterstützung auf der Ebene von Betriebssystemen, Basistools und Middleware anbietet.

WissGrid unterstützt daher die Konzeptionsphase und den Aufbauprozess neuer Communities nicht nur passiv durch Bereitstellung von Blaupausen, sondern zusätzlich durch die Einführung eines Fachberater-Teams. Dessen Kern besteht aus Vertretern der fünf bereits etablierten akademischen Grid-Communities, die damit eine allgemeine Schnittstelle zu den Experten innerhalb der beteiligten Communities bilden.

WissGrid AP2 hat einen Teil der im Grid-Betrieb gemachten Erfahrungen in Form von Deliverables dokumentiert [D-2.1.1, ..., D-2.1.8]. Diese so genannten *Blaupausen* dienen als Grundlage für die Beratung. Die Fortführung der nachhaltigen Dokumentation im Rahmen des WissGrid Projektes wird dann aus den Erfahrungen der Beratungstätigkeiten erfolgen.

Kernaufgabe des Fachberater-Teams ist die Weitergabe des Knowhows an die potentiellen neuen Communities in der Konzeptions- und Aufbauphase hinsichtlich einer maßgeschneiderten Infrastruktur in Anlehnung der einzelnen Aspekte der Blaupausen:

- Innere Organisationsstruktur [D-2.1.3]
  - Leitung
  - Arbeitsgruppen
  - Kommunikation
  - Repository
- Definition eines Betriebmodells (zusammen mit dem AP1)
- Kosten-Nutzen-Abschätzung [D-2.1.4]
  - Gridifizierbarkeit von existierenden Strukturen
  - CPU- und/oder datenlastige Aufgaben

- Dienste und CG-Architekturkonzepte
  - Herausbildung einer Security-Struktur (Zertifikatsmanagement, RAs)
  - Bildung einer Virtuellen Organisation (VO)
  - Auswahl einer geeigneten Middleware
  - Integration eines Kerns von Community-Ressourcen und/oder
  - Implementierung von essentiellen CG-Diensten
- Software-Lifecycle-Management [D-2.1.6]
  - Ermittlung und Umgang mit Anforderungen
  - Organisation und Vorgehensmodelle
  - Umgang mit Risiken
  - Testumgebungen, Testverfahren und Qualitätsmanagement
  - Lebenszyklus nach der Fertigstellung und Nachhaltigkeit
- Interaktion mit der Community
- disziplinbezogene Supportstruktur [D-2.1.8]

Dabei unterstützt das Fachberater-Team die neuen Communities sowohl auf elektronischen Wege über eine eigens dafür eingerichtete Mailingliste [fachberater@wissgrid.de](mailto:fachberater@wissgrid.de) und über das Trouble-Ticket-System des D-Grid User Support [NGI-DE-Portal], als auch nach Absprache persönlich per Telefon und Arbeitstreffen.

Weiterhin unterstützt das Team im Anschluss an die Aufbauphase neue Communities bei der Orientierung innerhalb der bestehenden Support- und Kommunikationsstrukturen, um einen reibungslosen Übergang in die Benutzung des Grids zu ermöglichen.

## Kapitel 2

# WissGrid-Fachberatertätigkeiten

Im letzten Jahr wurden bereits fünf neue Communities bei ihren Bemühungen unterstützt, zu evaluieren ob eine Möglichkeit der Nutzung des Grids sinnvoll ist bzw. die ersten Schritte in die Nutzung des Grids zu beschreiten und in die bestehende Struktur zu integrieren. Die nachfolgende Auflistung dokumentiert die wesentlichen Aktivitäten der Fachberater innerhalb der einzelnen Communities anhand von WissGrid internen Berichten des ersten Jahres. Diese Dokumentation wird in Zukunft weiter geführt und jährlich veröffentlicht werden.

### 2.1 Photon Science

**Organisationen** DESY/PETRAIII, European XFEL GmbH, EMBL, GKSS, CFEL

**Aktivität** Sammlung der Anforderungen für Grid-Nutzung, Beratung zur Entwicklung des Datenmanagementsystems sowie allgemeine Beratung

**Zeitraum** seit Dezember 2009

**Fachberater** IA, FS

HASYLAB, EMBL und GKSS sind die Hauptbetreiber der Experiment-Stationen (Beamlines) an der neuen Synchrotronlichtquelle „PETRAIII“ sowie dem Freien Elektronen Laser „FLASH“. Die Institute stellen die Beamlines und Strahlzeit einer internationalen, akademischen Anwender-Community als Service zur Verfügung. Die Lichtquellen werden ab Ende 2010 jährlich mehrere PByte von Daten produzieren, die entsprechend archiviert werden sollen. DESY stellt benötigte Hardware zur Verfügung. Die Software, insbesondere das Web Portal für remote Zugriff auf Messdaten, muss noch erstellt werden. Zusammen mit HASYLAB entwickeln wir die Architektur des Datenmanagementsystems und Software für ein Grid-Portal.

Die European XFEL GmbH baut zurzeit die XFEL Lichtquelle. Diese wird voraussichtlich 2014 im Betrieb genommen. Sie wird noch deutlich größere Datenmengen erzeugen mit Peak-Raten von bis zu 8 Pbyte/Woche. Eur. XFEL benötigt ähnliche Infrastruktur wie die anderen Synchrotron-

Gruppen, hat aber mehr Zeit für Forschung und Entwicklung. Zusammen mit XFEL GmbH werden entsprechende Lösungen und Architekturkonzepte für ein verteiltes Datenmanagementsystem evaluiert.

CFEL nutzt intensiv Freie Elektronen Laser wie FLASH und LCLS, aber auch konventionelle Synchrotrons wie SLS/PSI. Der Großteil der Daten wird im Ausland generiert. Zum Beispiel am LCLS typisch 250TByte pro Woche. Die Daten müssen also auf die lokalen Ressourcen transferiert werden, und zudem an andere Institute repliziert werden.

Während diverser Treffen mit den einzelnen Organisationen wurde eine erste Bestandsaufnahme der bestehenden Infra- und Organisationsstruktur vorgenommen. Außerdem wurden allgemeine Informationen zum Grid ausgetauscht und auf potentielle Beschränkungen überprüft. Dabei kristallisierten sich folgende Organisationsabhängige Kernfragen heraus:

**European X-ray Free Electron Laser GmbH - XFEL** Für XFEL stand im Fokus die Netzwerk-Infrastruktur. Thema unter anderem war die Möglichkeit Bandbreiten und Latenzen zu garantieren. Obwohl Remote Steuerung von Instrumenten und Experimenten wird in zunehmenden Maße gefordert wird, ist dies für XFEL mittelfristig kein Thema. Die generelle Möglichkeit wurde jedoch diskutiert, genau wie Fragen zur LZA und über Anwendungsfälle. Weiterhin wurden Diskussionen zu Architektur-Konzepte geführt. XFEL erwägt Glassfish Application Server für das Data Acquisition (DAQ) System einzusetzen, welche sich gut in das prototypische Framework einfügen würden.

Bezüglich der Entwicklung des Datenmanagementsystems gibt es bereits eine konkrete Anwendung. Damit die massiven Simulationsdaten für die Weiterverwendung angeboten werden können wurde die Implementation eines Prototypen mit Portal und dCache als Backend vereinbart. Ob und wie Fedora Commons und iRods eingebettet wird, bleibt noch zu klären.

**European Molecular Biology Laboratory - EMBL** Für EMBL bestand die Augenmerk auf die Datenverwaltung im Grid. Aus diesem Grund gab es gemeinsame Diskussionen über iSpy (Metadata-Engine für Protein-Kristallographie), Datenformate (nexus, hdf5) und (Meta-)Daten-Standardisierung. Die notwendige Einbettung in den Grid-Kontext und die Erweiterung auf andere Photon-Science Felder sind offene Fragen. Da es in der Photon science bisher kaum Daten-Standards gibt, entwickeln sich existierende Formate wie Nexus und HDF5 zu de facto Standards. HDF5 wird voraussichtlich ebenfalls von einer HGF-Initiative (PNI-HDRI=HighDataRateInitiative) Standardisiert werden. Weiterhin wurde ein generisches Architektur-Konzept vorgestellt und erörtert und über die zu implementierenden Services diskutiert. Um die genutzten Applikationen (SHELX) gridfähig zu machen werden diese mittelfristig von EMBL mit einem Workflow versehen um parametrische Jobs zu realisieren. Zusätzlich wird eine Gridfähige Version der SHELX-Gui notwendig evaluiert.

Für Testzwecke wurde nach einer ersten Diskussion das Aufsetzen eines Prototypen für SAXS und MX im Rahmen von dCache vereinbart.

**Hamburger Synchrotron Laboratory - HASYLAB** HASYLAB betreibt eine Vielzahl verschiedener Experimente an drei verschiedenen Strahlungsquellen, DORIS, FLASH und PetraIII, die in den kommenden Jahren noch erweitert und ergänzt werden sollen. Jedes Jahr führen einige tausend, wechselnde Anwender Experimente an den Strahlungsquellen durch. Für die Anwender gibt es eine User-Management Plattform (<http://door.desy.de>). Es gibt allerdings bislang kein Datenmanagement. Angesichts der Bestrebungen ein Datenarchiv für Photon Science Daten aufzubauen,

ist ein Datenmanagement-System, dass in irgendeiner Form mit dem User-Management verknüpft ist, unabdingbar. Die offensichtlichen Probleme liegen in der organisierten Migration der Daten in das Archiv, das Erfassen und Management der Meta-Daten, Definition und Standardisierung von Datenformaten, oder der Authorisierung von Zugriffsrechten. Wir haben begonnen in Zusammenarbeit mit HASYLAB ein Web-Portal zu entwickeln, dass die wesentlichen Probleme lösen könnte. Das Web-Portal erlaubt die Authentifizierung mit verschiedenen Credentials, wie Kerberos, dem DOOR-Account und Grid-Zertifikaten. Aufgrund der internationalen Bemühungen eine European Authentication and Authorization (EAA) Infrastruktur aufzubauen, wird die Integration von Shibboleth-Credentials ebenfalls erfolgen. Eine Federation dieser Datenbanken mit anderen Institutionen ist noch eine offene Frage. Das Grid-Portal ist zunächst an DOOR gebunden, kann aber leicht auf andere IdP-Systeme gebunden werden, wie zum Beispiel SMIS, dass voraussichtlich vom EMBL eingesetzt werden soll.

**Gesellschaft für Kernenergieverwertung in Schiffbau und Schifffahrt mbH - GKSS** GKSS betreibt bereits eine Tomographie-Beamline am DORIS Speicherring. Ende 2010 wird eine weitere Beamline an PetralIII hinzukommen. Kurz gefasst handelt es sich um eine 3D Gigapixel-Kamera mit einer Auflösung von einigen Nanometern. Individuelle Datensätze haben eine Größe von typisch 10-80GB, die vollständig im Speicher gehalten werden müssen. Aufgrund der Analyse-Algorithmen muss jeder Pixel des 3D-Volumens durch jeden Core eines Compute-clustern gehen, so dass schnelle Interconnects essentiell sind. Die während der Analyse generierten Daten sind mindestens ebenso groß wie die experimentellen Roh-Daten. Grid-Computing wird daher zur Zeit nicht in Betracht gezogen. Datamanagement und die Bereitstellung von lokalen Compute-Ressourcen für die Anwender ist hingegen ein Problem. Lösungen die Daten organisiert zu archivieren werden zur Zeit gemeinsam entwickelt und getestet. Möglichkeiten die Daten für die Anwender zu replizieren werden noch diskutiert. Aufgrund der limitierten Netzwerk-Bandbreite auf der User-Seite werden Daten vorwiegend händisch auf externe USB-Platten kopiert, was aber an der PetralIII-Beamline nicht mehr skaliert.

**Centre for Free Electron Laser - CFEL** Die CFEL-Experimente generieren massive Datenmengen an verschiedenen Beamlines im In- und Ausland. Pump-Probe Experimente mit biologischen Nanokristallen generieren Datensätze die aus Millionen von Einzelaufnahmen (typisch 16MB groß) bestehen. Analyse, Transfer und Replizierung der Daten bereiten massive Probleme. Bislang wurde das Datamanagement ausschliesslich händisch realisiert. Die Daten wurden auf einem singulären Fileserver ohne Redundanz gehostet. Transfer und Replikation von Daten wird in Laborbüchern protokolliert. Um ein transparentes, skalierendes Datenmanagement zu realisieren, beraten wir CFEL in der Bereitstellung und Realisierung von Archivierungsmöglichkeiten. Die Datenanalyse ist zur Zeit nicht zeitnah zu realisieren, ein Problem, dass sich mit dem weiteren Ausbau von LCLS und XFEL massiv verstärken wird. Massiv parallele Algorithmen werden daher von CFEL zur Zeit entwickelt. DESY stellt als Entwicklungsplattform unter anderem einen DUAL-GPU-Server mit 256 Cores zur Verfügung, der akkumuliert Daten mit 20GByte/s parsen kann. Um die kollaborative Entwicklung zu unterstützen planen wir diesen Server (und weitere Multi-Core Systeme) als Grid-Ressourcen bereitzustellen.

**Übergreifende Aktivitäten** Da die verschiedenen Institute und Gruppe überwiegend sehr ähnliche Anforderungen und Probleme zu lösen haben, wäre eine übergreifende Koordinierung der verschiedenen Aktivitäten sehr hilfreich. Wir werden daher Mitte Juli einen Workshop mit dem Thema *Datamanagement for Photon Sciences* organisieren.



**Aktivitäten in nationalen und Internationalen Projekten** Es gibt zur Zeit drei verschiedene Photon Science Projekte, die sich auf nationaler wie europäischer Ebene mit Aspekten des Data- und User-Managements befassen. PNI-HDRI ist eine Initiative der Photonen, Neutronen und Ionen-Forschungseinrichtungen der Helmholtz-Gemeinschaft. Ziel des Projektes sind ein HGF-weites Daten- und User-Management sowie die Standardisierung von Datenformaten und die gemeinsame Entwicklung von HPC-Anwendungen. IRUVX-PP/EuroFEL ist ein ESFRI-Projekt, das in eine europäische Organisation der FEL-Forschungseinrichtungen übergehen soll. Eines der Ziele ist die Schaffung der erwähnten European Authentication and Authorization Infrastruktur. Das dritte Projekt, PaNdata, ist ein Zusammenschluss der führenden Photonen und Neutronen-Laboratorien in Europa. Dieses FP7-geförderte Projekt hat insbesondere die Schaffung von verbindlichen User- und Data-Policies zum Ziel, und damit verbundenen Datenformat-Standards und User-Management. Da alle drei Projekte verschiedene Aspekte derselben Probleme angehen, wie auch die Aktivitäten aus WissGrid heraus, bemühen wir uns um eine Koordination der verschiedenen Entwicklungen, um zu vermeiden, dass mehrere verschiedene, inkompatible Lösungen parallel implementiert werden, und nach Möglichkeit die eigenen Entwicklungen wie das Datamanagement-Portal mit vertretbarem Aufwand integriert werden können.

## 2.2 Sozialwissenschaften (VirtAug)

**Organisationen** SOFI/SOEB

**Aktivität** Beratung

**Zeitraum** seit November 2009

**Fachberater** IA FS JL

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.

## 2.3 Altertumswissenschaften

**Organisationen** Deutsches Archäologisches Institut

**Aktivität** Information und Unterstützung in der Antragsstellung

**Zeitraum** seit Februar 2010

**Fachberater** HN, JL In einem ersten Treffen wurden allgemeine Informationen die über Grid-Umgebung und die Aktivitäten des WissGrid Projektes angesprochen.

## 2.4 Radio

**Organisationen** LOFAR/GLOW

**Aktivität**

**Zeitraum**

**Fachberater** HE

## 2.5 Klimafolgenforschung

**Organisationen** Servicegruppe Anpassung (SGA) bis Ende 2009, Climate Service Center (CSC) seit Anfang 2010

**Aktivität** Beratung bei der Extraktion und Nutzung der Klimadaten, die vom DKRZ bereits im Grid bereitgestellt werden.

**Zeitraum** seit 2009

**Fachberater** TR

Da das Hauptinteresse der Klimafolgenforschung im Grid die Nachnutzung von Archivdaten ist, überschneidet sich die Fachberatertätigkeit mit den Aufgaben des WissGrid-AP3 „Langzeitarchivierung von Forschungsdaten“. Die Anforderungen der Klimafolgenforschung wurden in je einen Gespräch mit der SGA und dem CRC aufgenommen und sind in die WissGrid-Deliverables „Generische Langzeitarchivierungs-Architektur für D-Grid“ und „WissGrid-Spezifikation: Langzeitarchivierungsdienste“ eingeflossen.

Das CSC und früher die SGA bereiten Daten aus der Klimaforschung so auf, dass diese in der Klimafolgenforschung verwendet werden können. Durch den ständigen Kontakt zu Wissenschaftlern der Klimafolgenforschung hat das CSC einen guten Überblick, welche Erwartungen die Klimafolgenforschung an das Grid hat. Die Klimafolgenforschung hat nicht vor, ein eigenes Community-Grid aufzubauen, sondern möchte das vorhandene Daten-Grid mit nutzen. Da sich die Forschungsschwerpunkte von denen der Klimaforschung unterscheiden, benötigt die Klimafolgenforschung vor allem andere Werkzeuge. Der WissGrid-Fachberater nimmt die Wünsche der neuen Community auf und berät bezüglich ihrer Realisierbarkeit. Ein Werkzeug zur Formatkonvertierung ist sogar durch WissGrid im Rahmen von AP3 realisiert worden.

## Kapitel 3

# Zusammenfassung

Bisher ist das Fachberaterteam in fünf neuen wissenschaftlichen Bereichen tätig geworden und hat bereits auch schon wichtige Hilfestellung im Aufbau einer Grid-Community geleistet. Weiterhin konnte durch die Beratung der Sozialwissenschaften gezeigt werden, dass im WissGrid-Projekt ein hohes Maß an Grid-Fachkenntnis vorhanden ist. In der Folge wurde WissGrid beauftragt, eine Expertise zur Grid-Nutzung für die Sozialwissenschaften zu erstellen. Zusätzlich zu den hier dokumentierten Aktivitäten finden ständig weitere Diskussionen und Beratungen fachübergreifend statt, deren vollständige Aufzählung den Rahmen dieses Dokuments sprengen würde.

# Literaturverzeichnis

- [D-2.1.1] *Community-Grids - Überblick und Report* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.1.pdf>
- [D-2.1.3] *Community-Grids - Organisationsstrukturen* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.3.pdf>
- [D-2.1.4] *Kosten Nutzen Analyse CG Aufbau* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.4.pdf>
- [D-2.1.5] *Community-Grids - Gridarchitekturen im Ueberblick* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.5.pdf>
- [D-2.1.6] *Software Lifecycle Management im Grid* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.6.pdf>
- [D-2.1.8] *Support Infrastrukturen* <http://www.wissgrid.de/publikationen/deliverables/wp2/D-2.1.8.pdf>
- [NGI-DE-Portal] *NGI-DE Grid User Support (Helpdesk-System)* <https://helpdesk.ngi-de.eu/>