

Erfolgskontrollbericht

Zuwendungsempfänger: Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung Bremerhaven	Förderkennzeichen: 01 G09005F
Vorhabenbezeichnung: WissGrid – Grid für die Wissenschaft	
Laufzeit des Vorhabens: 01.05.2009 – 30.04.2012	Berichtszeitraum: 01.05.2009 – 30.04.2012

Summary

Based on experiences of five academic communities in the first phase of D-Grid, WissGrid aimed in helping new scientific communities to use grid technology. AWI contributed with its experience in developing the community grid infrastructure for the climate community C3Grid. This includes aspects of sustainability and financing for AP1 as well as technical features for AP2/AP3.

In the blueprints, we presented architectural considerations and described grid services from C3Grid which can be reused by new communities. Our consulting activities were focused on the climate impact research community and biodiversity/research on habitat transformations.

AWI's expertise in data long term archiving resulted in the contribution to documents in AP3 describing general considerations of grid repositories and services. For the widely used data format NetCDF, a new JHOVE2 module was implemented, which extracts and validates the metadata from the file headers. Furthermore, a user friendly Fedora frontend was developed to support users in storing the complex metadata for their data. The developed software is freely available and can be reused by other communities.

1. Beitrag des Ergebnisses zu den förderpolitischen Zielen, z. B. des Förderprogramms - (ggf. unter Angabe des Schwerpunkts) - soweit dies möglich ist -,

Die Arbeit im Projekt WissGrid diente dazu, die Erfahrungen aus unterschiedlichen fachspezifischen Grid-Projekten zusammenzufassen und zu verallgemeinern. Auf dieser Grundlage wurden Blaupausen erarbeitet, die neuen akademischen Communities den Einstieg in das Grid erleichtern sollen. Die gesammelte Expertise wurde insbesondere bei der Fachberatung neuer Wissenschaftsdisziplinen genutzt. Insofern leistete WissGrid einen Beitrag zur Verbreitung der Technologie für die collaborative Zusammenarbeit. Vom AWI wurden insbesondere die Communities der Klimafolgenforschung und der Biodiversität/Habitatsforschung angesprochen.

2. Das **wissenschaftlich-technische Ergebnis** des FE-Auftrags, die erreichten Nebenergebnisse und die gesammelten wesentlichen Erfahrungen,

Das AWI lieferte in allen drei Arbeitspaketen Beiträge zu WissGrid. Die Ergebnisse reichen dabei von konzeptionellen Arbeiten über Beratungsaktivitäten bis hin zu konkreten technischen Implementierungen. Im Folgenden sollen die wichtigsten Ergebnisse zu den einzelnen Arbeitspaketen dargestellt werden.

AP1

Gemeinsam mit den anderen Partnern wurden im AP1 Vorstellungen entwickelt, wie sich Infrastrukturen für die wissenschaftliche Arbeit nachhaltig betreiben lassen. Dabei verlagerte sich der Fokus im Lauf des Projekts von den reinen Grid-Umgebungen hin zu virtuellen Forschungsumgebungen, deren technologische Basis nicht festgeschrieben ist. Diese Verallgemeinerung erlaubt es, flexibler auf die schnelle Weiterentwicklung von Technologien zu reagieren und ist damit perspektivisch besser aufgestellt. Das AWI konnte dabei auf seine Erfahrungen bei der Ausarbeitung eines Betriebsmodells für C3Grid zurückgreifen, wo durch *Letter of Intend* die wichtigsten Partner bereits ihre Bereitschaft bekundet haben, auch nach Projektende den Betrieb von C3Grid zu gewährleisten. Die Klimaforschung besitzt mit dem Deutschen Klimarechenzentrum bereits eine Forschungsplattform, die von allen Mitgliedern der Community genutzt werden kann. Da das AWI Gesellschafter des DKRZ ist, konnten wir die dabei gewonnenen Erkenntnisse auch in AP1 einbringen und auf Vorteile, aber auch Schwächen des Betriebs- und Finanzierungsmodells hinweisen. All diese Erfahrungen flossen in die vom AP1 erstellten Dokumente ein. Aktiv beteiligte sich das AWI bei der Vorbereitung und Durchführung des Workshops „Förderung virtueller Forschungsumgebungen“ am 16.09.2010 in Bonn sowie des Nachhaltigkeitsworkshops im Rahmen des Digital Humanities Festakts in Göttingen.

AP2

Für das AP2 wurde zunächst der *Status Quo* der „alten“ Communities festgehalten. Dabei brachten wir unsere Erkenntnisse aus dem Projekt „*Collaborative Climate Community Data and Processing Grid – C3Grid*“¹ ein, wo für die deutsche Klimacommunity eine Grid-Infrastruktur zum einheitlichen Zugriff auf verteilte Daten und für ein verteiltes Processing aufgebaut wurde. Ebenso konnte die Expertise aus dem Projekt „*Nutzung von kurzlebigen Zertifikaten in portalbasierten Grids – GapSLC*“ genutzt werden. Die Analyse reichte dabei von technischen bis hin zu organisatorischen Aspekten. In den Blaupausen wurden die unterschiedlichen Ansätze

¹ www.c3grid.de

und Realisierungen der verschiedenen „alten“ Communities dann zusammengefasst und für neue Communities nutzbar gemacht. Im Bereich Sicherheit wurde ein zusätzliches Deliverable gemeinsam mit den Partnern erstellt, um diesen wichtigen Teilbereich nochmals genauer zu beleuchten. Außerdem wurden gemeinsam mit GapSLC der 5. D-Grid Security Workshop organisiert und durchgeführt.

Ein großer Schwerpunkt der Arbeiten lag auf der Fachberatung, wo die vorhandene Expertise für die Lösung konkreter Probleme neuer Communities genutzt werden konnte. In einem Fachberater-Workshop am 19.01.2011 wurden neue Communities angesprochen, was auf sehr gute Resonanz stieß. Jedoch wurde auch deutlich, dass die Spezifika der einzelnen Forschungsdisziplinen so unterschiedlich sind, dass Einzelberatungen weitaus effektiver sind.

Die vom AWI durchgeführte Fachberatung umfasste folgende Bereiche:

a) Klimafolgenforschung

Die Klimafolgenforschung zielt darauf, die Auswirkungen des Klimawandels möglichst konkret für verschiedene Bereiche der Gesellschaft zu prognostizieren. Sie steht damit an der Schnittstelle zwischen der eigentlichen Klimaforschung und all jenen, die an den Ergebnissen der Klimaforschung interessiert sind. Dabei muss ein extrem breites Spektrum von Interessenten abgedeckt werden. Exemplarisch seien hier Mediziner, Städteplaner und Landwirte genannt, aber auch die Tourismusbranche und Versicherungen benötigen die Informationen aus der Klimaforschung.

Das AWI arbeitete dabei eng mit dem Climate Service Center² in Hamburg und dem Klimabüro für Polargebiete und Meeresspiegelanstieg³ in Bremerhaven zusammen. Dabei ergab die Analyse der Anforderungen, dass keine neue Grid-Infrastruktur aufgebaut werden muss, sondern die bereits vorhandenen Komponenten von C3Grid nachgenutzt werden und durch spezifische Elemente erweitert werden können. Als ein Beispiel dafür wurde gemeinsam mit dem Klimabüro für Polargebiete und Meeresspiegelanstieg ein Portal konzipiert und implementiert, über das Interessierte umfangreiche Informationen zum Meereis bekommen. Der Datenteil mit prozessierten Satellitendaten des Instituts für Umweltphysik Bremen verwendet dabei genau die gleichen Architekturkomponenten wie C3Grid, um Meereisdaten für den Nutzer bereitzustellen. Dabei erweist es sich immer wieder als unumgänglich, dass die Infrastrukturkomponenten erweitert werden müssen, da den Nutzer außerhalb der Klimacommunity viele Standards nicht bekannt sind, sie z.B. mit den üblichen Formaten von NetCDF oder GRIB nicht arbeiten können oder Hilfe bei der Interpretation der Ergebnisse benötigen. Die begonnenen Arbeiten werden jetzt im Rahmen von Helmholtz-Klimainitiative REKLIM (Regionale Klimaänderungen) weitergeführt.

b) Biodiversität/Habitatsforschung

Thematisch ebenfalls eng mit der Klimaforschung verbunden ist die Habitatsforschung, bei der untersucht wird, wie sich der Lebensraum von Pflanzen und Tieren ändert. Sie ist eine

² www.climate-service-center.de

³ www.klimabuero-polarmeer.de

noch relativ junge Forschungsrichtung, wo erst in letzter Zeit die Datenmengen stark anwachsen. Hier wurde durch unseren WissGrid-Mitarbeiter José Mejía grundlegende Beratung für den Bereich Datenhaltung, Daten Viewer und Metadaten geleistet.

In dieser Community werden Daten in Form von Rasterbildern sowie Vektordaten und Attributtabelle zu Studiengebieten gesammelt. Diese lassen sich mit geringem Aufwand über GIS⁴-Tools prozessieren und als Webservices anbieten. Um die Zusammenarbeit der Wissenschaftler mit solchen Webservices zu ermöglichen, hat unsere Fachberatung eine Web-Infrastruktur konzipiert und die wesentlichen Dienste für ihre Realisierung implementiert: einen Web-GIS Viewer, einen Metadaten Geoportal und einen Kartenkatalog.

Verschiedene Tools des ESRI ArcGIS Viewer für Flex 2.4 wurden für die Online Visualisierung der Webservices, die Datenanalyse und die Anzeige der im ArcGIS Geoportal 1.1 verwalteten Metadaten nach dem Inspire Profil⁵ erweitert. Die leichte Konfiguration des Web-GIS Viewer vereinfacht die Integration mehrerer Webservices in ein sogenanntes Kartenprojekt, das der Community über einen HTTP Server zur Verfügung gestellt wird. Schließlich unterstützt der Kartenkatalog die Wissenschaftler bei der räumlichen und textuellen Suche der im Internet vorhandenen Kartenprojekte.

c) Biomedizin: DFG-Projekt LABIMI

Doch nicht nur für „benachbarte“ Communities wurde Fachberatung angeboten. Über WissGrid wurde der Kontakt zu Göttinger Kollegen aus dem Projekt LABIMI⁶ hergestellt, die sich zur Nutzung von Repositorien bei uns informierten und Hilfestellung bei der Evaluierung des am AWI entwickelten Federico als Repository für biomedizinische Forschungsdaten erhielten.

Die Fachberateraktivitäten erwiesen sich für die Außenwirkung von WissGrid als sehr fruchtbar. Einige Ergebnisse aus AP2 flossen und fließen auch direkt wieder zurück in das aktuell laufende Projekt „C3Grid – INAD - Towards an infrastructure for general access to climate data“, das auf den Vorarbeiten des D-Gridprojekts C3Grid aufbaut und jetzt durch das BMBF-Referat „Globaler Wandel“ gefördert wird. Dazu gehört z.B. der im Deliverable D2.1.6. definierte Software-Management-Prozess.

AP3

Am Beginn standen vor allem konzeptionelle Arbeiten, die vor allem in den Deliverables zur Spezifikation der LZA-Architektur, der Dienste und der Repositories niedergelegt wurden. Sie bildeten die Grundlage für die Zwischenbegutachtung, die erfolgreich nach einem Jahr absolviert werden konnte.

Daneben wurden umfangreiche technische Implementierungen vorgenommen.

⁴ Geoinformationssystem

⁵ Inspire Metadatenprofil <http://inspire.jrc.ec.europa.eu/>

⁶ <http://www.labimi-f.de>

a) Tools für Jhove2

Jhove2 ist ein Framework zur Analyse komplexer digitaler Objekte und umfasst Komponenten zur Identifizierung, Validierung und Metadatenextraktion. Für die Anwendung in der Klimawissenschaft wurde ein eigenes Modul JANEME entwickelt, mit dem Dateien in den beiden weit verbreiteten Datenformaten GRIB und NetCDF verarbeitet werden können. Es erlaubt neben der Validierung der Formate vor allem die Extraktion der in den Headern enthaltenen Metadaten, die in ein vorhandenes Metadatenmodell eines LZA-Repositorys eingetragen werden können, so dass der manuelle Aufwand für die Erstellung von Metadaten minimiert werden kann. Es werden unterschiedliche Metadatenprofile unterstützt, die zudem auch noch individuell angepasst werden können. In einer individuell konfigurierbaren Validierungskette werden die gefundenen Metadatenelemente auch gleich validiert, um auf Fehler aufmerksam zu machen, die eine spätere Weiterverarbeitung erschweren würden.

JANEME zeichnet sich durch folgende Eigenschaften aus:

- Plattformunabhängig – funktioniert unter den Betriebssystemen Linux, Solaris und Windows XP/Vista/7
- Einfache Integration neuer Vokabularien für Variablennamen und Attributwerte
- Flexibel anpassbare Validierung mit Wiederverwendung von vordefinierten Validatoren
- Unterstützung von mehreren (auch komplexen) Metadatenprofilen

JANEME ist unter <http://aforge.awi.de/gf/project/jhove2/frs> verfügbar. Eine ausführliche Dokumentation ist enthalten. Das Modul wird als „3rd-party module development“ der offiziellen Jhove2-Distribution der internationalen Community bereitgestellt⁷.

Um die Flexibilität des gewählten Ansatzes für die Implementierung von JANEME zu demonstrieren, wurde ein analoges Modul für das FITS-Format⁸ bereitgestellt, das vor allem in der Astronomie verwendet wird. Damit konnte gezeigt werden, dass es relativ einfach ist, die Anpassungen an neue Communities vorzunehmen. Das Modul ist frei verfügbar⁹.

b) Federico

Innerhalb von AP3 wurde Fedora als Repository diskutiert. Am AWI wurde dazu ein AJAX Frontend Federico (*Fedora Enabled Repository with Cocoon*) konzipiert und entwickelt, um Fedora für die Ablage von Metadaten nutzen zu können. Federico ist dabei ein komplexes System mit mehreren Komponenten:

⁷ https://bitbucket.org/jhove2/main/wiki/Community_Activities

⁸ Wells, D. C.; Greisen, E. W.; Harten, R. H.: "FITS - a Flexible Image Transport System". *Astronomy and Astrophysics Supp. Ser.* 44: 363–370. June 1981.
<http://adsabs.harvard.edu/abs/1981A%26AS...44..363W>

⁹ <http://aforge.awi.de/gf/project/jhove2fits/>

- Nutzerverwaltung mit verschiedenen Authentifizierungsmechanismen (LDAP, OpenID) und der Möglichkeit, den Nutzern unterschiedliche Rollen zuteilen zu können
- Schnittstelle zur Ablage von XML-Files, wobei die Metadatenprofile jeweils angepasst werden können. Validierung der Metadaten.
- Bereitstellung der Metadaten über OAI-PMH. Damit können die verwalteten Metadaten im Grid verfügbar gemacht werden, so dass das Repository in einer Föderation integriert werden kann.
- Suchfunktionalität: Hier ist insbesondere die Möglichkeit der nutzerfreundlichen Navigation mittels Karte zu erwähnen, in dem die Nutzer das sie interessierende Gebiet anwählen können und dann alle dazu gehörigen Datensammlungen angezeigt bekommen.

Federico ist einschließlich der Dokumentation unter <http://aforge.awi.de/gf/project/federico/> verfügbar.

3. **Erfindungen/Schutzrechtsanmeldungen** und erteilte Schutzrechte die vom AN oder von am Auftrag Beteiligten gemacht oder in Anspruch genommen wurden, ggf. auch deren standortbezogene Verwertung (Lizenzen u.a.),

Im Rahmen des Projekts wurden keine Schutzrechtsanmeldungen vorgenommen.

Die neu entwickelten Tools stehen unter der Open Source Lizenz allen Interessierten zur Verfügung.

4. die evtl. **wirtschaftlichen Erfolgsaussichten** nach Auftragsende (mit Zeithorizont) - z.B. auch funktionale/wirtschaftliche Vorteile gegenüber Konkurrenzlösungen, Nutzen für verschiedene Anwendergruppen/-industrien am Standort Deutschland, Umsetzungs- und Transferstrategien (Angaben, soweit die Art des Auftrags dies zulässt),

Das Projekt zielte auf die akademischen Communities, insofern ist keine wirtschaftliche Verwertung der Ergebnisse zu erwarten. Allerdings wurden in Hinsicht auf die Nachhaltigkeit der Entwicklung wichtige Arbeiten im AP1 Betriebsmodelle gemacht. Da die entstandenen Infrastrukturen nicht dauerhaft über Projektmittel gefördert werden können, wurden Modelle erarbeitet, die eine nachhaltige Finanzierung des Grids ermöglichen können. Allerdings sind dafür Änderungen in der Förderstruktur notwendig, die eine Umlenkung von Mitteln aus der Förderung von fachspezifischen Projekten in geeignete Organisationsformen der Infrastrukturbetreiber vorsehen.

5. die evtl. **wissenschaftlichen und/oder technischen Erfolgsaussichten** nach Auftragsende (mit Zeithorizont) - u.a. wie die geplanten Ergebnisse in anderer Weise (z.B. für öffentliche Aufgaben, Datenbanken, Netzwerke, Transferstellen etc.) genutzt werden können. Dabei ist auch eine etwaige Zusammenarbeit mit anderen Einrichtungen, Firmen, Netzwerken, Forschungsstellen u.a. einzubeziehen,

Das Meereisportal wird gemeinsam mit dem Klimabüro für Polargebiete und Meeresspiegelanstieg im Rahmen von REKLIM weiter betrieben. An einer Ausweitung der Datenbasis wird gearbeitet.

JANEME und Federico sollen im Rahmen des Dateningests für Modelldaten am AWI weitergenutzt werden. JANEME ist als Community-Beitrag auf den offiziellen Jhove2-Seiten verlinkt und steht damit der internationalen Community zur Verfügung.

6. die evtl. **wissenschaftliche und wirtschaftliche Anschlussfähigkeit** für eine mögliche notwendige nächste Phase bzw. die nächsten innovatorischen Schritte,

Weiterentwicklungen der Komponenten sind durch die generische Struktur möglich.

7. Arbeiten, die zu keiner Lösung geführt haben,

-

8. Präsentationsmöglichkeiten für mögliche Nutzer - z.B. Anwenderkonferenzen (Angaben, soweit die Art des Auftrags dies zulässt),

siehe Anhang A

9. die Einhaltung der Kosten- und Zeitplanung.

Die Zeit- und Kostenplanung wurden jeweils im Wesentlichen eingehalten. Verschiebungen in AP1 betrafen das Gesamtprojekt und wurden vom Projektmanagement mit dem Projektträger abgestimmt.

Anhang A

Die Ergebnisse der Arbeiten in WissGrid wurden wie folgt präsentiert:

Datum	Autoren/Titel	Veranstaltung
17.09.2009	B. Fritsch: „Verwendung von Metadaten im Grid am Beispiel von C3Grid“	DGI Metadaten Workshop, SUB Göttingen
16.09.2010	B. Fritsch: „Use Case Klimaforschung“	Workshop „Förderung virtueller Forschungsumgebungen“, Bonn
19.09.2010	J. Mejía: „JaNeME: J-a NetCDF Metadata Extractor“	JHOVE2 Tutorial -ipress 2010, Wien
29./30.09.2010	H. Enke, B. Fritsch, C. Grimme, F. Schlünzen: „Betrachtungen zur akademischen Sicherheitsinfrastruktur im D-Grid“	Security Workshop, Göttingen
10.11.2010	J. Mejía: „JaNeME“	DKRZ Datenseminar, Hamburg
19.01.2011	B. Fritsch: „Gemeinsame Standards, Verfahren und Arbeitsabläufe nutzen“	Workshop „Virtuelle Forschungsumgebungen aufbauen - mit D-Grid“, Göttingen
31.01.2011	B. Fritsch, J. Mejía: „Data management in Climate Research“	DESY Computing Seminar (DVSEM), Hamburg
22.06.2011	H. Neuroth, B. Fritsch: „Vernetzte Forschungsumgebungen in Grid-Projekten“	DINI Workshop Virtuelle Forschungsumgebungen, Berlin
21.09.2011	S. Pinkernell, B. Fritsch: „Kurzlebige Zertifikate in einem Gridportal-Szenario“	Workshop Grid-Technologie für den Entwurf technischer Systeme (Grid4TS), Dresden
29.11.2011	J. Mejía: „JANEME und Federico“	AG Informationssysteme, Bremerhaven
19.-21.03.2012	B. Fritsch, C. Grimme: „C3Grid – Schritte von der Anschubfinanzierung zum nachhaltigen Betrieb“	D-Grid Ergebnisworkshop
19.-21.03.2012	S. Kindermann, F. Schintke, B. Fritsch: „C3Grid als Werkzeug für das Datenmanagement in der Klimaforschung“	D-Grid Ergebnisworkshop

Anhang B:

Das AWI war in allen drei Arbeitspaketen beteiligt und liefert daher Beiträge zur Mehrzahl der Deliverables:

Deliverable	Bezeichnung
AP1	
D1.0	Fördermodelle
D1.2	Draft Betriebsmodell: Überlegungen zu Zweck, Form und Inhalt zum Betrieb Virtueller Forschungsumgebungen
D1.3	Expertenrunde „Förderung Virtueller Forschungsumgebungen“
D1.4	Konzept eines Betriebsmodells für Virtuelle Forschungsumgebungen
D1.5	Aspekte der Implementierung eines Betriebsmodells für Virtuelle Forschungsumgebungen
D1.6	Erfahrungsaustausch und Anpassungen: Workshop „Nachhaltigkeit Virtueller Forschungsumgebungen“
D1.7	Literaturrecherche
AP2	
D2.1.1	Community-Grids - Überblick und Report
D2.1.2	Evaluation, Dokumentation und Registrierung von höheren Diensten
D2.1.3	Blaupausen für die organisatorische Struktur einer neuen Community
D2.1.5	Evaluation und Dokumentation existierender Architekturkonzepte
D2.1.6	Leitfaden zur Unterstützung des Software-Management-Prozesses
D2.1.7	Strategien für den Aufbau und Betrieb von Testumgebungen
D2.2.1	Erster Bericht zu Aktivitäten von WissGrid-Fachberater-Teams
D2.2.2	Aufbaukonzepte für neue Community-Grids, Aktivitäten und Ergebnisse der Fachberater teams
D2.3.2	Betrachtungen zur akademischen Sicherheitsinfrastruktur im D-Grid
AP3	
D3.1	LZA-Architektur
D3.2	Workshop: Architektur zur Langzeitarchivierung in D-Grid
D3.4.1	Erste WissGrid-Spezifikation: Langzeitarchivierungsdienste
D3.4.2	WissGrid-Spezifikation: Langzeitarchivierungsdienste

D3.4.3	LZA-Dienste: Charakterisierung, Validierung, Metadatenextraktion, Formatmodule für JHOVE2: NetCDF (Software, Dokumentation) und FITS (Software, Dokumentation) Unterstützung bei : JHOVE2 als GridJob (Skripte, Dokumentation)
D3.5.1	Erste WissGrid-Spezifikation: Grid-Repository
D3.5.2	WissGrid-Spezifikation: Grid-Repository
D3.5.3	Grid-Repository: Federico (Software und Dokumentation)