



Report über die ausgewählten Dienste (R 3.2.6)

Version 04.05.2016

Cluster 3

Verantwortlicher Partner DAASI, GWDG

DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: Final

Verfügbarkeit: Öffentlich

Autoren:

Heiko Hütter

Carsten Thiel

Tibor Kálmán

Ulrich Schwardmann

Peter Gietz

Xi Kong

Revisionsverlauf:

Datum	Autor	Kommentare
13.06.2015	Heiko Hütter	Erste Strukturierung und Roadmap zur Kommentierung
15.06.2015	Carsten Thiel	Kommentierung zu vorhandenen Diensten
15.10.2015	Xi Kong	Erfassung von Nutzungszahlen
23.10.2015	Heiko Hütter Tibor Kálmán Ulrich Schwardmann	Neue Struktur und Einarbeitung von Nutzungszahlen
05.01.2016	Heiko Hütter	Fertigstellung von Version 1
18.01.2016	Peter Gietz	Endredaktion Version 1
4.5.2016	Xi Kong	Die Erklärung von Anhang 1 ergänzt. Kommentare von Carsten Thiel eingearbeitet (zum Abschnitt 2 und die Zahlen der Tabelle) Kommentare von Benedikt von St. Vieth eingearbeitet (Anzahl der VM von JSC ergänzt)

Inhaltsverzeichnis:

1. EINLEITUNG	4
2. NICHT NACHHALTIG BEREITGESTELLTE DIENSTE	4
3. WEITERES VORGEHEN	5
4. AKTUELLE DIENSTÜBERSICHT	6
4.1. Bereitgestellte virtuelle Maschinen (Hosting)	7
4.2. Bereitgestellte Basis-, Generische- und Operative-Dienste	9

1. Einleitung

Dieses Dokument sollte eine Analyse des Status der in DARIAH-DE verfügbaren Dienste liefern, sodass vom DARIAH-DE Exekutiv-Komitee eine fundierte Entscheidung über den Weiterbetrieb jedes in DARIAH-DE verfügbaren Services nach der zweiten Förderphase erfolgen kann. Als Report von Cluster 3 standen dabei insbesondere technische Eigenschaften, betriebliche Abläufe und Nutzungszahlen im Vordergrund. Insbesondere wurde versucht für die Services folgende Informationen zu ermitteln:

- Bisherige Nutzungszahlen sollten, soweit verfügbar, Auskunft über die bisherige Nutzung des jeweiligen Services liefern. Zusätzlich sollte eine Prognose für das zukünftige Nutzungsverhalten gegeben werden
- Eine Kostenschätzung nach aktuell vorliegenden Zahlen und Berechnungsmodellen. Diese sollte sowohl für die aktuell vorliegenden Nutzungszahlen gegeben, als auch für die prognostizierten Bedarfe ermittelt werden.

Der Status des Dienstes soll beschreiben, wie nahe der Dienststatus an einen Produktivbetrieb heranreicht, der insbesondere auch geplante Abrechnungsmöglichkeiten erlaubt und in die Prozesse von DARIAH-DE integriert ist.

Eine Beurteilung aus fachwissenschaftlicher Perspektive kann von Cluster 3 nicht geleistet werden und muss im Bedarfsfall durch andere Cluster, die im Exekutivkomitee vertreten sind, erfolgen.

Bei dem Versuch eben diese Daten zu ermitteln, hat sich allerdings herausgestellt, dass für eine Realisierung dieses Vorhabens noch wesentliche Arbeit sowohl fachwissenschaftlicher als auch technischer Art geleistet werden muss.

- Fachwissenschaftlich muss geklärt werden über welche Impact-Faktoren die Entscheidung für einen Weiterbetrieb eines Dienstes unterstützt werden kann
- Technisch muss bei einigen Diensten eine Grundlage für eine Nutzungserhebung erst noch geschaffen werden, bei anderen Diensten muss die Erfassungsmethode noch transparenter gemacht werden.

Es wird empfohlen, diese Arbeiten aus Synergiegründen parallel zu Arbeiten an einem Accounting-System durchzuführen, wie sie für ein Folgeprojekt andiskutiert wurden.

2. Nicht nachhaltig bereitgestellte Dienste

Nicht nachhaltig bereitgestellt Dienste sind Dienste, deren Daten und/oder Funktionen entweder in anderen, durch DARIAH-DE betriebenen, Diensten verfügbar sind oder migriert werden können, oder die aus anderen Gründen nicht weiter betrieben

werden. Die AG Service Life Cycle (SLC) sieht im aktuellen Entwurf des SLC vor, dass sowohl die Kriterien für als auch der Prozess zur Außerbetriebnahme eines Dienstes bereits im Handover-State definiert werden. So soll noch vor Produktivgang eines solchen Services sichergestellt werden, dass es Strategien zur Nachnutzung der generierten Daten oder Funktionen eines Services gibt, aber auch sinnvolle Kriterien existieren, an Hand derer entschieden werden kann, ob die Außerbetriebnahme eines Dienstes beispielsweise aus Ressourcenknappheit erfolgen kann.

Aus ganzheitlicher Sicht wurde das Service-Portfolio von DARIAH-DE analysiert und folgende Dienste identifiziert, die erstmal nicht in DARIAH-DE betrieben werden sollen:

- **Sonar/SonarQube** ist ein Software-Qualitätsmanagement-Tool, das in DARIAH evaluiert wurde, geringe Nutzungszahlen aufweist. Es wird in anderem Kontext/Projekt eingebunden.
- **SVN** als Standalone-Lösung für die Verwaltung von Code ist vollständig im Software-Management-Tool **Chili Project** enthalten, das mittlerweile Teil des DARIAH-DE Portfolios ist. Daher kann der reine SVN-Service aus dem zukünftigen Portfolio entnommen werden.
- **Jira** verursacht hohe Lizenzkosten und die Nicht-Verlängerung der JIRA-Lizenz war im 2014 nach DARIAH-I entschieden. Die genutzten Funktionalitäten werden vollständig durch das lizenzkostenfreie **Chili Project** abgedeckt. Während DARIAH-DE II wurde eine Migration hin zu Chili vollzogen.

3. Weiteres Vorgehen

Unter dem Gesichtspunkt zunehmend knapperer Ressourcen muss DARIAH-DE qualifizierte Entscheidungen sowohl zum Betrieb bzw. zur Aufnahme von Services als auch zur Außerbetriebnahme bereits laufender Services zu Gunsten neuer, benötigter Services treffen. Einen wesentlichen Beitrag zu möglichen Verfahren, die eine Entscheidungsfindung unterstützen können, liefert Report R 1.3.3 „Finale Version Erfolgskriterien“ aus Cluster 1. Die Herausforderung, teilweise völlig unterschiedliche Services (im Sinne von Methodik, Fachrichtung usw.) nachvollziehbar vergleichen zu können, ist damit allerdings noch nicht überwunden. Eine solche Entscheidung ist zum aktuellen Diskussionsstand vom Exekutivkomitee zu treffen, das an Cluster 2 und 3 allerdings noch entsprechende Anforderungen im Sinne von Informationsbedarfen und insbesondere deren Metriken artikulieren muss, wie sich im Folgenden zeigt.

Aus operativer und technischer Sicht ergeben sich zum Einen verschiedene Möglichkeiten der Granularität der Erfassung von Metainformationen zu einem Service:

1. Eine Dienstgüte wird beispielsweise lediglich an Hand der Umgebung beurteilt (Entwicklungsumgebung vs. Produktivumgebung). Eine solche Erhebung ist Cluster 2 und 3 jeder Zeit möglich, führt mutmaßlich allerdings zu falschen Schlussfolgerungen, da die aktuelle Betriebsumgebung keine hinreichende

Aussage zur Betriebstauglichkeit, Reifegrad oder gar zu Bedarfen eines Services liefert.

2. Cluster 2 und 3 ist ebenfalls in der Lage oberflächliche Nutzungszahlen für die Services zu erheben¹. Diese sind allerdings zum jetzigen Zeitpunkt ebenfalls in Frage zu stellen, da die Daten extrem verunreinigt sein könnten.²
3. Aus Perspektive von Cluster 3 ergibt somit lediglich eine Erfassung von Nutzungszahlen Sinn, die auf speziell für den jeweiligen Service und seinen Einsatzzweck zugeschnittenen Metriken basiert. Die Definition dieser Metriken kann allerdings nur von fachwissenschaftlichen Experten erfolgen.

Es ist zu erwarten, dass für eine servicespezifische Erfassung von Nutzungszahlen für die meisten Services noch Entwicklungsaufwand besteht, um diese qualifiziert zu erfassen. Die Anforderungen werden allerdings äquivalent zu den Anforderungen erwartet, die sich durch die Einführung der Accounting-Infrastruktur ergeben, weswegen die Umsetzung optimaler Weise zeitgleich erfolgt.

Nach der erfolgreichen Umsetzung ist empfehlenswert, dass das Exekutivkomitee in regelmäßigem Turnus das DARIAH-DE Serviceportfolio an Hand der definierten und erfassten Metriken evaluiert. Aus Perspektive der Autoren sollte hierbei für die Entscheidungsfindung mindestens folgende Dimensionen berücksichtigt werden:

- Fachwissenschaftlicher Nutzen (Impactfaktoren)
- Nutzungszahlen
- Technische Pflege und Betrieb ist weiterhin gewährleistet bzw. auf dem aktuellen Stand
- Kompatibilität mit anderen relevanten DARIAH-DE Infrastruktur-Komponenten weiterhin vorhanden

4. Anhang 1: Aktuelle Dienstübersicht

Die folgenden Tabellen enthalten Daten, die bei den Serviceprovidern erhoben worden sind. Cluster 3 hält einerseits die erfassten Nutzungszahlen als noch nicht hinreichend vergleichbar und die Erhebungsmethode als noch nicht ausreichend „transparent“, sodass sie zunächst als unzureichend für die Unterstützung einer fundierten Entscheidung erachtet werden. Andererseits existieren aus technischer Sicht keine ausreichenden Werkzeuge, die die Nutzungszahlen aller in DARIAH-DE zu betreibenden Dienste in ihrer jetzigen Form leicht anheben könnten. Als ein erster Ansatz, seien die Ergebnisse dennoch an dieser Stelle dokumentiert.

¹ Beispielsweise durch Auslesen der Serverlogs, des genutzten Traffics, etc.

² Öffentlich zugängliche Dienste könnten deutlich größere Zugriffszahlen aufweisen, da sie von Bots und Suchmaschinen aufgerufen würden oder Backup-Prozesse könnten ebenfalls bei entsprechender Konfiguration falsche Nutzungszahlen implizieren, um nur zwei Beispiele zu nennen.

4.1. Bereitgestellte virtuelle Maschinen (Hosting)

Services Technische Infrastruktur und Unterstützungen		Nutzungszahlen*			
		GWDG	RZG	JSC	gesamt
Virtuelle Maschinen	A.1				
Virtuelle Maschine Small Bronze mit Betriebssystem	A.1.1.1				
Virtuelle Maschine Small Silber mit Betriebssystem	A.1.1.2				
Virtuelle Maschine Small mit OS und Backup, Bronze	A.1.1.3				
Virtuelle Maschine Small mit OS und Backup, Silber	A.1.1.4				
Virtuelle Maschine Small mit Backup, OS-Wartung, Bronze	A.1.1.5		2		2
Virtuelle Maschine Small mit Backup, OS-Wartung, Silber	A.1.1.6			5	5
Virtuelle Maschine Small mit OS-Wartung, Bronze	A.1.1.7	2	3	3	8
Virtuelle Maschine Small mit OS-Wartung, Silber	A.1.1.8				
Virtuelle Maschine Big Bronze mit Betriebssystem	A.1.2.1				
Virtuelle Maschine Big Silber mit Betriebssystem	A.1.2.2				
Virtuelle Maschine Big mit OS und Backup, Bronze	A.1.2.3				
Virtuelle Maschine Big mit OS und Backup, Silber	A.1.2.4	1			1

Virtuelle Maschine Big mit Backup, OS-Wartung, Bronze	A.1.2.5	17	1		18
Virtuelle Maschine Big mit Backup, OS-Wartung, Silber	A.1.2.6	3			3
Virtuelle Maschine Big mit OS-Wartung, Bronze	A.1.2.7	5			5
Virtuelle Maschine Big mit OS-Wartung, Silber	A.1.2.8	2			2
Beratung und Unterstützung	A.2				
Konfiguration und Installation	A.3				
Wartung	A.4				
Implementation	A.5				
Hosting von Komplett-Services	A.6				
Bereitstellung VM mit Backup, OS-Wartung, Silber	A.6	2		1	3
Administration des Service	A.6				
Summe (per Jahr)		32	6	9	47
<i>* Anzahl der virtuellen Maschinen die im Zeitpunkt der Erfassung im Betrieb waren</i>					

4.2. Bereitgestellte Basis-, Generische- und Operative-Dienste

		Einzelnutzung*
Services Basisdienste		
PID Service	B.1	
mit < 50.000 PID	B.1	
mit < 1.000.000 PID	B.1	**
Storage	B.2	
Fileserver/Massenspeicher	B.2.1	
Langfristige Datensicherung, HSM	B.2.2	
DARIAH-Repositoryum (Data Curation)	B.2.3	**
DARIAH Storage Service (Bitstream Preservation)	B.2.4	**
Cloud Share (DFN-Cloud, 10GB)	B.2.5	
Datenbanken	B.3	
Mailinglisten mit Mailman	B.4	**
Technical Survey Provisioning	B.5	

Confluence	B.6	710
EtherPad	B.7	2982
JIRA Issue Tracking	B.8	**
Chili Projektverwaltung	B.9	68
Version Control	B.10	
Jenkins	B.11	30
Sonar - Code Quality	B.12	3
Services Generische Dienste		
GeoBrowser	C.1	**
Collection Registry	C.2	**
Schema Registry	C.3	**
Generische Suche	C.4	**
Services Operative IT-Dienste		
Authentication and Authorization Infrastructure	D.1	2982
Monitoring	D.2	40
Help Desk	D.3	240

Bemerkungen:

- (*) Nutzungszahlen wurden auf Basis von verschiedenen Metriken ermittelt.
- (**) Nutzungszahlen konnten mit den bestehenden Accounting-Tools nicht erfasst werden.