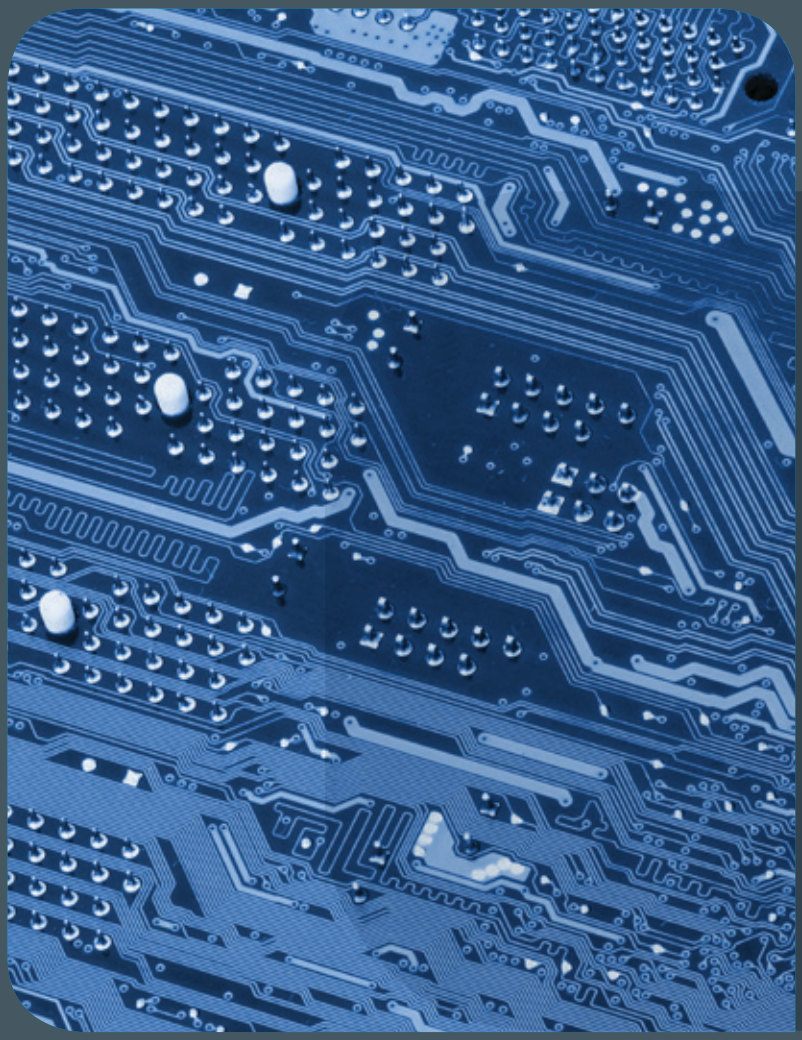
GWDG NACHRICHTEN 09|13

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen



GWDG
NACHRICHTEN

09|13

Inhalt

Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
36. Jahrgang
Ausgabe 9/2013

Erscheinungsweise:
monatlich

www.gwdg.de/gwdg-nr

Auflage:
500

Fotos:
© jamdesign - Fotolia.com (1)
© Spectral-Design - Fotolia.com (7)
© pizuttipics - Fotolia.com (14)
© Rainer Grothuis - Fotolia.com (23)
© contrastwerkstatt - Fotolia.com (24, 25)
© MPIbpc-Medienservice (3)
GWDG (2, 26)

Herausgeber:

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
Am Faßberg 11
37077 Göttingen
Tel.: 0551 201-1510
Fax: 0551 201-2150

Redaktion:
Dr. Thomas Otto
E-Mail: thomas.otto@gwdg.de

Herstellung:
Maria Geraci
E-Mail: maria.geraci@gwdg.de

Druck:
GWDG / AG H
E-Mail: printservice@gwdg.de

Prof. Dr. Ramin Yahyapour
ramin.yahyapour@gwdg.de
0551 201-1545

Liebe Kunden und Freunde der GWDG,

schon in der letzten Ausgabe war der Abhörskandal der Geheimdienste ein Thema meines Editorials. Es stellt sich jenseits der geäußerten Kritik die Frage, was man aktiv tun kann. Dabei gibt es einige Maßnahmen, die sich recht leicht etablieren lassen.

Die meisten E-Mail-Server und E-Mail-Programme erlauben eine Verschlüsselung der Kommunikationskanäle. Die GWDG betreibt mit ihrem Exchange-2010-Dienst mittlerweile 50.000 Postfächer für die Max-Planck-Gesellschaft und die Universität Göttingen. Da mittlerweile alle gebräuchlichen E-Mail-Server und E-Mail-Programme Transportverschlüsselung über TLS/SSL unterstützen, wird die GWDG in den kommenden Monaten diese Verschlüsselung für die Kommunikation innerhalb der MPG und der Universität statt als empfohlene Option als obligatorischen Standard etablieren.

Es bietet sich für unsere Kunden zusätzlich auch eine Ende-zu-Ende-Verschlüsselung der E-Mails an. Die GWDG bietet in Kooperation mit dem DFN-Verein einen Zertifikatsdienst an, über den Nutzer einfach ein Personenzertifikat erhalten können, das sich leicht in viele E-Mail-Programme für eine S/MIME-Verschlüsselung und Signatur installieren lässt.

In den vorliegenden und kommenden GWDG-Nachrichten werden wir diese Maßnahmen und Optionen näher vorstellen. All dies kann kein umfassender Schutz gegen Überwachung sein. Dennoch sollten alle vertretbaren, technischen Maßnahmen ergriffen werden, um die Kommunikation zu schützen. Die GWDG wird sich deshalb weiter aktiv dafür einsetzen, die IT-Sicherheit für ihre Kunden voranzubringen.

Ramin Yahyapour

GWDG – IT in der Wissenschaft

E-Mail-Verschlüsselung mit X.509-Zertifikaten – Teil 1: Beantragung und Sicherung von Zertifikaten

Text und Kontakt:
Thorsten Hindermann
thorsten.hindermann@gwdg.de
0551 201-1837

Aus aktuellem Anlass der National Security Agency (NSA) Prism-Affäre, bei der E-Mails massenhaft gespeichert und ihre Inhalte möglicherweise durchsucht wurden, soll in einer mehrteiligen Artikelserie, die in den kommenden Ausgaben der GWDG-Nachrichten fortgesetzt wird, gezeigt werden, wie sensible Informationen in E-Mails vor dem Zugriff Dritter geschützt werden können. Es kann natürlich mit dem Verfahren der E-Mail-Verschlüsselung nicht verhindert werden, dass diese aus den Datenströmen im Internet abgezweigt und gespeichert werden können. Aber es wird potenziellen Stellen im In- und Ausland wesentlich erschwert oder unmöglich gemacht, die Inhalte der E-Mails zu manipulieren oder diese gar zu lesen und nach bestimmten Begriffen durchzumustern.

BEGRIFFSERKLÄRUNGEN

Die zwei Hauptbegiffe, die im Zusammenhang mit dem Umgang von E-Mail-Verschlüsselung fallen, sind **X.509-Zertifikate** und **Public Key Infrastructure**, im Weiteren kurz **PKI** genannt.

Die PKI ist ein hierarchisch organisierter Aufbau von Zertifikatsautoritäten, engl. **Certification Authority** (im Weiteren kurz **CA** genannt), beginnend mit einer Wurzel, über Zwischenstationen hin zur ausstellenden Autorität für Zertifikate. Diese Kette der Autoritäten bildet die Grundlage einer PKI.

Die Zertifikate wiederum sind eine digitale Repräsentation von Benutzern, Diensten, Netzwerkgeräten oder Computern, die durch eine CA ausgestellt wurden. Diese Zertifikate sind zusammen mit jeweils einem privaten (private key) und einem öffentlichen (public key) Schlüssel miteinander verbunden.

Technisch betrachtet ist das Zertifikat eine digital signierte Ansammlung von Informationen, u. a. Informationen über den Benutzer, Dienst, Netzwerkgerät oder Computer, die ausstellende CA, die verwendeten Signier-/Verschlüsselungsverfahren, Informationen über die Abruf-URLs von Sperrlisten für gesperrte Zertifikate usw.

X.509 wiederum ist ein ITU-T-Standard (Internationale Fernmeldunion) für eine PKI zum Er-/Ausstellen digitaler Zertifikate.

ZERTIFIKAT BEANTRAGEN

Um nun ein Zertifikat zu beantragen, ist es als erstes wichtig zu wissen, welche ausstellende Registrierungsautorität, engl. **Registration Authority** (im Weiteren kurz **RA** genannt), für Antragsteller zuständig ist. Für die Kunden der Max-Planck-Gesellschaft und der Universität Göttingen gibt es jeweils einen Link, den es sich lohnt in der Lesezeichenliste aufzunehmen.

Für die MPG-CA ist das *https://ca.mpg.de/request* oder *https://ca.mpg.de/ras*.

Und für die Universität Göttingen-CA ist das *https://ca.uni-goettingen.de/request* oder *https://ca.uni-goettingen.de/ras*.

In jahrelanger Praxis hat es sich bewährt, das/die Zertifikat(e) mit dem Mozilla Firefox zu beantragen und zu verwalten. Ein weiterer Vorteil des Firefox-Webbrowser ist, dass dieser auf allen drei gängigen Plattformen Windows, Linux und Mac OS X zur Verfügung steht.

E-mail encryption using X.509 certificates

Due to recent events at the National Security Agency (NSA) Prism affair, in which the NSA stored mass of e-mails and searched their content for important keywords, a multipart series of articles, that will be continued in future issues of GWDG News, will show how sensitive information in e-mails can be protected from third party access. Of course with the method of the e-mail encryption it can not be prevented that e-mails can be diverted and stored from the data streams in the Internet. But it is made potential sites at homeland and abroad very difficult or impossible to manipulate the contents of the e-mails or even read this and to screen for certain terms.

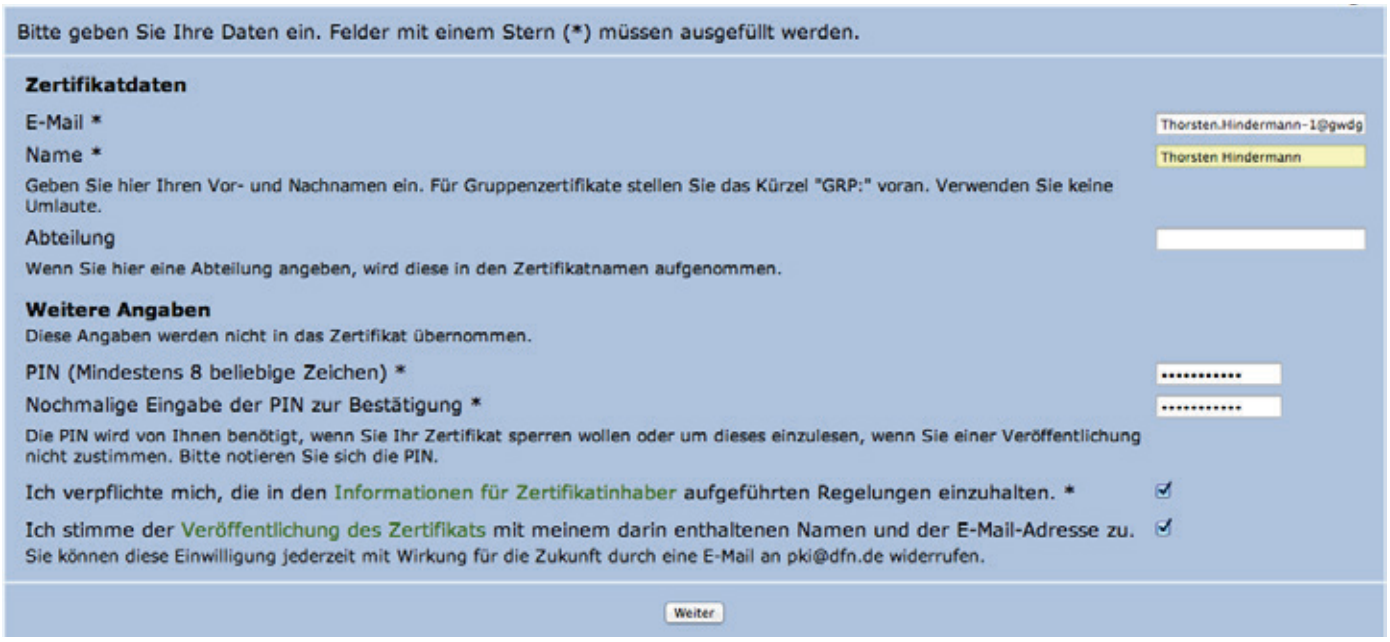


Abb. 1

Der eigentliche Antrag wird durch Klick auf die Schaltfläche „Nutzerzertifikat" in der Menüleiste des Registerreiters „Zertifikate" dann bei der ausgewählten RA mittels des Webbrowsers Firefox gestellt (siehe Abb. 1). Bei diesem Webformular ist es wichtig, dass die mit * gekennzeichneten Felder ausgefüllt werden. Eine Abteilung kann wahlweise angegeben werden. Nun muss noch eine PIN eingegeben werden. Diese wird oftmals beim Import des Zertifikats in den Firefox gebraucht und wenn der Anwender selbst sein Zertifkat sperren möchte. Auch hier wird dann die PIN abgefragt, bevor das Zertifikat dann gesperrt wird. Bitte diese Angabe gut merken! Da in einer PKI der öffentliche Schlüssel (engl. public key) ohne Bedenken weitergegeben werden kann, kann der Haken bei „Veröffentlichung des Zertifikats" ohne Bedenken gesetzt werden. Diese Möglichkeit kann sich sogar als vorteilhaft erweisen, wie später noch beschrieben wird. Die „Informationen für den Zertifikatinhaber" müssen auf alle Fälle durch Setzen des Hakens anerkannt werden. Jetzt auf „Weiter" klicken.

In der Übersichtsseite über die eingegebenen Angaben können diese noch einmal auf ihre Richtigkeit geprüft werden und mit einem Klick auf „Ändern" korrigiert werden. Andernfalls nun auf „Bestätigen" klicken.

Wurde auf „Bestätigen" geklickt, wird im Firefox der private Schlüssel generiert und im Firefox-Zertifikatspeicher abgelegt. Dieser Zertifikatspeicher ist unabhängig vom verwendeten Betriebssystem. Weiterhin wird der Zertifikatantrag, engl. certificate signing request (im Weiteren kurz CSR), in der ausgewählten RA hochgeladen.

Es wird eine Bestätigungsseite angezeigt. Mit einem Klick auf „Zertifikatantrag anzeigen" wird der generierte Antrag im PDF-Format entweder gleich angezeigt oder heruntergeladen und kann dann mit einem PDF-Anzeigeprogramm angezeigt und ausgedruckt werden. Diese Handhabung hängt vom verwendeten Betriebssystem und/oder installierten PDF-Anzeigeprogramm ab. Den ausgedruckten Antrag muss der Zertifikatnehmer eigenhändig unterschreiben.

Mit diesem Formular muss er dann zum RA-Operator der ausgewählten RA gehen. Dort wird die persönliche Identifizierung vorgenommen, d. h. mittels des Personalausweises des Zertifikatnehmers vergleicht und überprüft der RA-Operator die Angaben auf dem Zertifikatantrag mit dem Ausweis. Wenn alles in Ordnung ist, wird der RA-Operator das Zertifikat dann zeitnah ausstellen. Per Bestätigungsmail an den Zertifikatnehmer wird dieser über die Ausstellung des Zertifikats unterrichtet. Korrekterweise muss hier von der Signierung des öffentlichen Schlüssels, des hochgeladenen CSR, durch die entsprechende CA gesprochen werden.

In der Bestätigungsmail kopiert der Zertifikatnehmer den zweiten URL – das ist wichtig(!) – aus der E-Mail und kopiert diesen in die Adresszeile des Firefox. Ist auf dem System des Zertifikatnehmers der Firefox der Standardbrowser, genügt ein Klick auf diesen Link.

Nun werden im Firefox der private und signierte öffentliche Schlüssel zusammengeführt und beide ergeben zusammen das Zertifikat (siehe Abb. 2). Ist dieser Vorgang erfolgreich abgeschlossen, wird folgender Hinweis präsentiert. (siehe Abb. 3). Im

Abb. 2

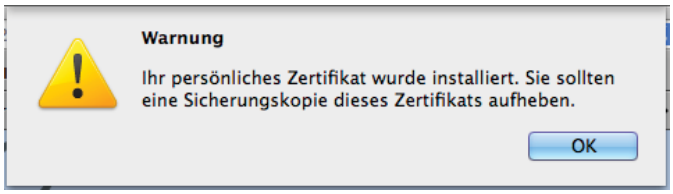


Abb. 3

folgenden Kapitel wird dieser wichtige Teil beschrieben.

Ein detaillierte Anleitung zur Beantragung eines Zertifikats ist unter dem URL *http://www.gwdg.de/index.php?id=zertifikat_beantragen* zu finden.

Anmerkung: Neben dem Firefox kann auch mit dem Internet Explorer (im Weiteren kurz IE genannt) auf grafische Weise ein Zertifikatantrag erzeugt werden. Manchmal kann es aber mit dem IE vorkommen, dass oftmals, abhängig von der eingesetzten Windows-Version, ein wichtiger betriebssystemseitiger Bestandteil noch nicht installiert ist, so dass es zu Fehlermeldungen kommen kann und die Beantragung scheitert. Hier müssen dann oftmals die VortOrt-Administratoren erst noch das fehlende Programmteil installieren, bevor die Beantragung gelingt. Es ist auch die Erzeugung eines CSR mittels des Kommandozeilenprogramms Open-SSL möglich, allerdings werden hier dann schon erweiterte Kenntnisse mit Zertifikaten und der Umgang mit der Kommandozeile vorausgesetzt. Aus diesem Grund hat sich die Verwendung des Firefox im Laufe der Jahre als am praktikabelsten herausgestellt.

SICHERUNG VON ZERTIFIKATEN

Eine der wichtigsten Handlungen ist es, eine Sicherheitskopie des gerade erstellten Zertifikats anzufertigen. Auch hier ist der Firefox Webbrowser dem Zertifikatnehmer behilflich.

Dazu muss unter „Extras“ der Einstellungen-Dialog im Firefox geöffnet werden. Hier das Zahnrad-Symbol mit der Unterschrift „Erweitert“ anklicken und auf der mehrfach geteilten Schaltfläche darunter auf „Zertifikate“ klicken. Nun die Schaltfläche „Zertifikate anzeigen“ anklicken (siehe Abb. 4).

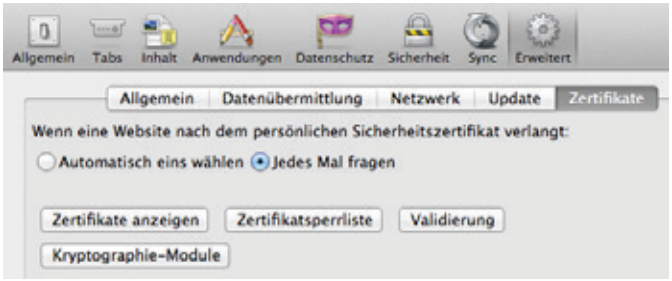


Abb. 4

In dem nun erscheinenden Dialog auf der mehrfach geteilten Schaltfläche/Registerreiter auf „Ihre Zertifikate“ klicken und das gerade zusammengeführte Zertifikat anklicken. Auf die Schaltfläche „Sichern...“ klicken (siehe Abb. 5). Es wird nach einem Kennwort gefragt, mit der die Container-Datei im PKCS#12-Format verschlüsselt wird. Der Grund dafür ist, dass diese Datei sowohl den privaten als auch den öffentlichen Schlüssel enthält, also das gesamte Zertifkat. Gerade wegen des privaten Schlüssels ist es wichtig, dass diese Datei entsprechend gesichert ist.

Im entsprechenden Speicherdialog muss ein Datenträger/Verzeichnis angegeben werden, wo die Datei mit der Dateiendung *.P12* gespeichert werden soll. Es empfiehlt sich ein externer Datenträger. Praktischer Hintergrund: Wenn der Rechner, auf dem das Zertifikat mal beantragt wurde, ausgetauscht wird, die Festplatte formatiert wird oder defekt ist, ist das Zertifikat unwiderruflich verloren. Dann ist ein Entschlüsseln von E-Mails, die mit diesem Zertifikat verschlüsselt worden sind, für immer unmöglich!

Anmerkung: Die Sicherung hat auch noch einen anderen, praktischen Aspekt, der nicht unterschätzt werden sollte. Im Laufe der Tätigkeit sammeln sich mit der Zeit einige Zertifikate an. Wenn nun mit einem oder mehreren Zertifikaten E-Mails verschlüsselt worden sind, können diese alten E-Mails nur mit dem dann aufbewahrten Zertifikat wieder entschlüsselt werden, selbst wenn zu diesem Zeitpunkt das Zertifikat sein Ablaufdatum überschritten hat. Deshalb ist die Sicherung und Aufbewahrung ein wichtiger Schritt. D. h. bei einem Rechnerwechsel müssen dann am besten

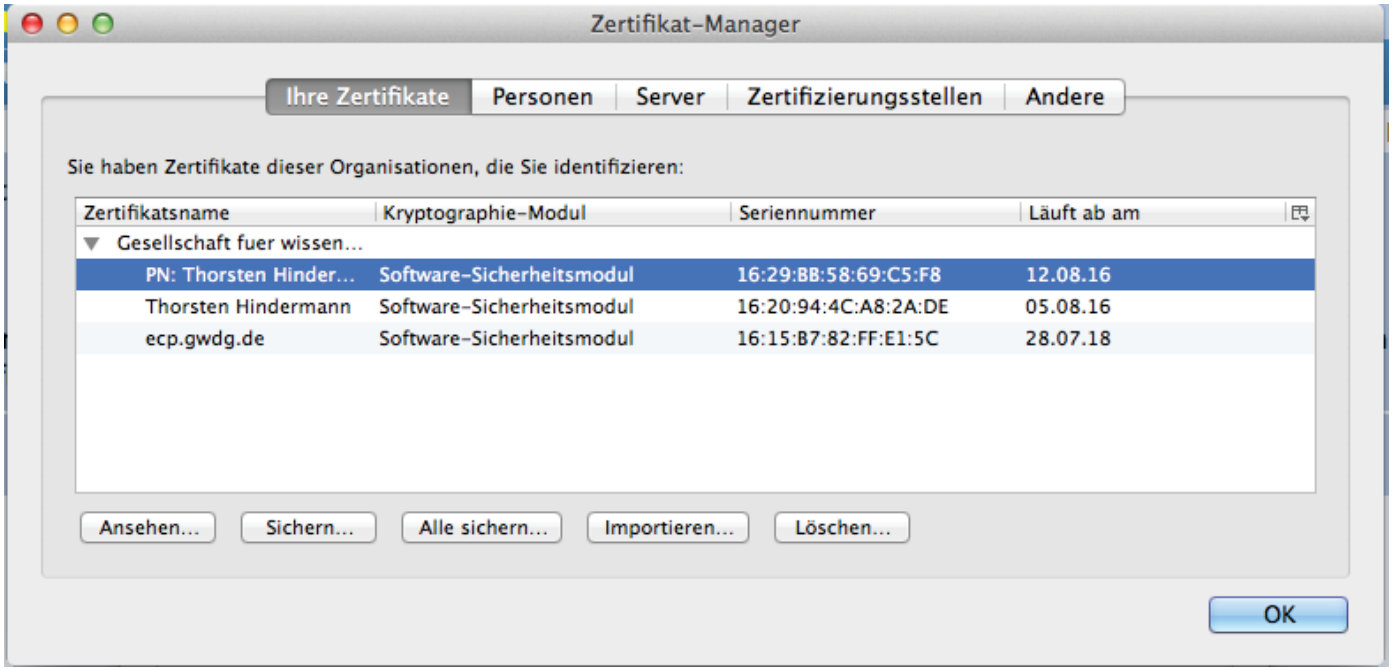


Abb. 5

alle alten und das aktuelle Zertifikat in die entsprechenden Zertifikatspeicher importiert werden. Dieser Vorgang wird im Teil 2 in den nächsten GWDG-Nachrichten näher beschrieben.

Hinweis: Ein persönliches Zertifikat, um E-Mail-Nachrichten zu signieren/verschlüsseln, hat eine Laufzeit von drei Jahren.

AUSBLICK

Nachdem in diesem Artikel die Beantragung und Sicherung von Zertifikaten zur E-Mail-Verschlüsselung erläutert wurden, sollen in den nächsten Teilen folgende Themen detailliert behandelt werden:

- Installation und Verteilung von Zertifikaten (GWDG-Nachrichten 10/2013)
- Verschlüsselung bei Outlook-Mailanwendungen (GWDG-Nachrichten 11/2013)
- Verschlüsselung bei Thunderbird, Notes 9 und Apple-Mail-anwendungen (GWDG-Nachrichten 12/2013)

Infobox

Wichtige URLs

Zertifikate beantragen

MPG: *https://ca.mpg.de/request oder https://ca.mpg.de/ras*
Universität Göttingen: *https://ca.uni-gottingen.de/request* oder *https://ca.uni-goettingen.de/ras*

Informationen

Allgemein: *http://www.gwdg.de/pki*
Public-Key-Infrastruktur: *http://www.gwdg.de/index.php?id=pki*
Detailinfomationen: *http://wiki.gwdg.de/index.php/Kategorie:PKI*
PKI-FAQ: *http://www.gwdg.de/index.php?id=faq#c2374*

Artikel in den GWDG-Nachrichten zum Thema „Zertifikate“

Ein zweiteiliger Artikel zur Einführung in die Welt der X.509-Zertifikate in den Ausgaben 9/2011 und 10/2011

Ein Artikel, wie Zertifikate für die VMware-Infrastruktur-Dienste erstellt werden, in der Ausgabe 5/2013

Kontakt

Bei weiteren Fragen zu diesem Thema schreiben Sie bitte eine entsprechende E-Mail an *support@gwdg.de*.

Identity Management bei der GWDG – die technische Lösung

Text und Kontakt:
Andreas Ißleiber
andreas.issleiber@gwdg.de
0551 201-1815

Der Artikel beschreibt im Detail, wie technische Probleme und Prozesse im Bereich des Identity Management bei der GWDG gelöst werden. Darüber hinaus zeigen wir das Zusammenspiel und die Abhängigkeiten zwischen den einzelnen IdM-Komponenten und wie diese eingesetzt werden können.

EINLEITUNG

In den GWDG-Nachrichten 8/2013 haben wir die Vorteile der Anbindung lokaler Benutzerverwaltungen der Institute an das zentrale Identity Management (IdM) der GWDG beschrieben. Darüber hinaus stellten wir unseren neuen Dienst „IdM as a Service" vor, der insbesondere auch für die Max-Planck-Institute eine erhebliche Vereinfachung der Benutzerverwaltung erlaubt und zusätzlich den raschen Zugang zu zentralen Diensten der GWDG ermöglicht. Auch zukünftige Dienste können damit zeitnah und automatisch an den Endanwender gebracht werden.

Dieser Artikel beschreibt nun die technischen Abläufe im Detail und stellt die Umsetzung der Prozesse aus dem Bereich der Benutzerverwaltung dar. Überdies werden Zusammenhänge zwischen den einzelnen IdM-Bausteinen näher erläutert.

KOMPONENTEN DES IDM

Das Identity Management der GWDG umfasst viele Komponenten, welche ganz unterschiedliche Bereiche abdecken und als Gesamtwerk für die Umsetzung zahlreicher Prozesse aus der Benutzerverwaltung der Universität Göttingen sowie Max-Planck-Instituten verantwortlich sind.

Das Identity Management bei der GWDG kann in fünf große Bereiche unterteilt werden, welche in Abbildung 1 dargestellt sind.

METADIRECTORY

Als zentrale Instanz fungiert das MetaDirectory (eDirectory von Novell), welches alle Identitäten aller angebundenen Verzeichnisse beinhaltet. Hierbei besteht das MetaDirectory aus zwei virtualisierten Servern, die, redundant aufgebaut, das zentrale Verzeichnis (eDirectory) zur Verfügung stellen. *Idm1* ist der Hauptserver, der auch als Master-Replica bezeichnet wird und sich mit dem Server *idm2*, der eine Read/Write-Replica hält, permanent synchronisiert. Durch diese Replikation zwischen *idm1* (Master) sowie *idm2* (Read/Write) erreichen wir eine stabile Redundanz und eine sehr hohe Verfügbarkeit des Verzeichnisses. Diese beiden Server stellen somit die Basis des MetaDirectory dar.

Anbindung der Verzeichnisse

Alle externen Verzeichnisse wie (Windows AD, LDAP, Datenbanken u. v. m.) sind über Remote-Loader, welche als Java application auf den externen Systemen laufen, mit dem MetaDirectory der GWDG verbunden (vgl. GWDG Nachrichten 8/2013). Die Kommunikation zwischen Remote-Loader und Me-taDirectory läuft prinzipiell gegen den Server idm1. Bei Ausfall von *idm1* wäre eine Kommunikation gegen den redundanten *idm2* jederzeit möglich.

Treiber

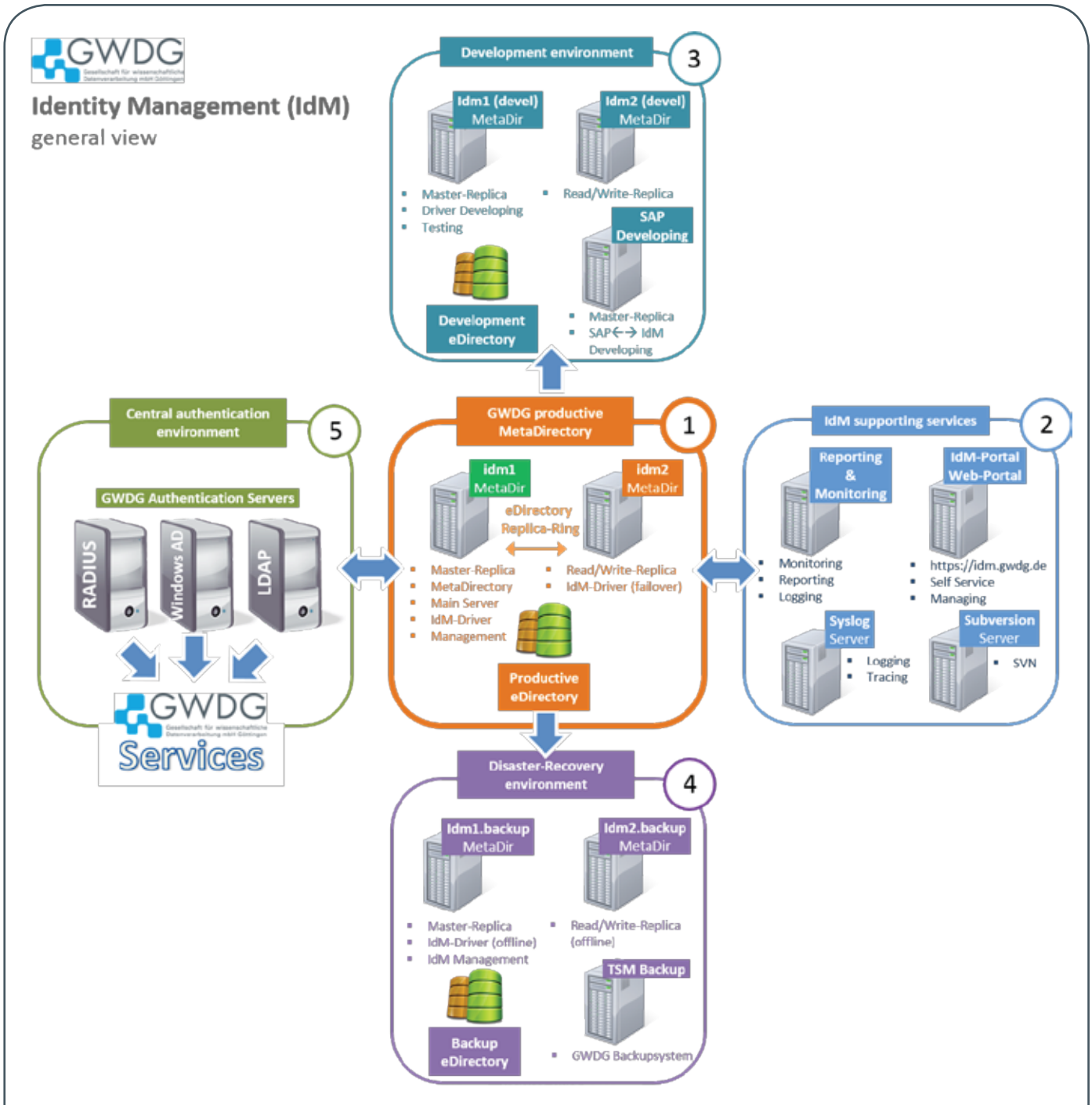
Zu jedem extern angebundenen Verzeichnis läuft ein entsprechender Treiber im MetaDirectory, der die Logik der Anbindung darstellt und damit auch alle Policies für die Verzeichnisanbindung beinhaltet. Alle Treiber laufen im Produktivbetrieb auf dem Server idm1. Auf idm2 sind ebenfalls alle Treiber vor-handen, jedoch sind diese gestoppt und müssten nur bei Störungen oder Ausfall von idm1 gestartet werden (Failover).

IDM-UNTERSTÜTZENDE SYSTEME

In direkter Umgebung zum MetaDirectory laufen weitere Server, die das MetaDirectory unterstützen sowie den Zugriff der Benutzer auf das IdM ermöglichen.

GWDG identity management – the technical solution

This article describes in detail how to solve technical problems and procedures in the field of identity management. Beyond that, we also show the correlations and dependencies between each single IdM components. How it works and how to use it.



1_Wesentliche Bausteine des Identity Management bei der GWDG

IdM-Portal

Insbesondere das IdM-Portal (*https://idm.gwdg.de*) ist die entscheidende zentrale Schnittstelle zwischen Benutzer/Administratoren und dem MetaDirectory. Über diese Webseite können Benutzer grundlegende Änderungen im Bereich des SelfService vornehmen. Vor allen Dingen aber ist es ein Portal für die Administration der an das MetaDirectory angebundenen Verzeichnisse. Viele Administratoren der einzelnen Institute finden in diesem Portal eine dem Institut angepasste Arbeitsumgebung vor, mit der alle wesentlichen administrativen Aufgaben erledigt werden können (Anlage, Löschen, Änderungen von Accounts u. v .m.; siehe hierzu die GWDG-Nachrichten 3/2013).

Monitoring, Reporting, Logging

Ein weiterer Server dient dem Monitoring und Reporting der Ereignisse aus dem MetaDirectory. Hier sind sämtliche Ereignisse, mit einem Timestamp versehen, abgelegt. Das ist insbesondere zur Nachverfolgung von Ereignissen entscheidend sowie auch für die Entwicklung und Test von Treibern und Policies sehr hilfreich. Der Bereich Report erzeugt darüber hinaus Statistiken, wie viele Änderungen oder Benutzeranlagen in welchen Bereichen durchgeführt wurden. Das Reporting wird bei der GWDG im September 2013 in Betrieb genommen und dient eher zur anonymen, nominellen Erfassung von Ereignissen, um auch eine Aussage über die Auslastung des Gesamtsystems treffen zu können und basierend auf den Ergebnissen im Vorfeld schon Maßnahmen für etwaige

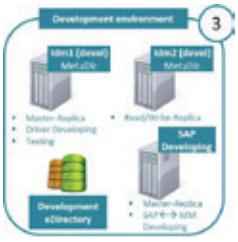
Erweiterungen des IdM zu ergreifen.

Das Monitoring dient zur unmittelbaren Sicherstellung der Verfügbarkeit des MetaDirectory. Hier werden die Treiber überwacht und bei etwaigen Störungen vordefinierte Aktionen ausgeführt. Das kann vom schlichten „restart" eines Treibers bis hin zur E-Mail-Alarmierung und zum Starten von externen Scripts gehen.

Subversion

Ein SVN-Server dient zur Versionierung und Verfolgung von Änderungen bei der Treiberentwicklung, ganz so wie es auch in anderen Softwareentwicklungen eingesetzt wird. SVN ist hier ein unverzichtbares Werkzeug, da in der Regel mehrere Personen an der Treiberentwicklung beteiligt sind und Änderungen an Treibern auch Auswirkungen auf mehrere Bereiche gleichzeitig haben können.

DIE ENTWICKLUNGSUMGEBUNG



Die Entwicklung von Treibern im MetaDirectory unterscheidet sich nicht von anderer Softwareentwicklung. Letztlich wird bei der Treiberentwicklung der Treiber in einer „höheren Logik" mithilfe eines Programms (Designer) entwickelt. Der Treiber selbst liegt dann im XML-Format vor (DirXML). Grundsätzlich entwickeln wir Treiber niemals an dem produktiven MetaDirectory.

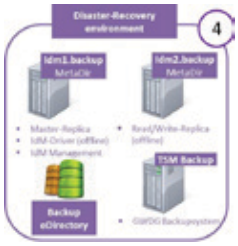
Treiberentwicklung

Eigens zu diesem Zweck haben wir eine Entwicklungsumgebung aufgebaut, die eine Kopie des produktiven MetaDirectory darstellt, bei dem aber keine relevanten Benutzerdaten übertragen werden. In dieser Test- und Entwicklungsumgebung können Treiber für die Institute in einer realistischen Umgebung entwickelt werden, ohne dabei Produktivdaten zu gefährden. Ein fertiggestellter Treiber durchläuft hierbei mehrere Tests, bis dieser anschließend vom Entwicklungs- in das Produktivsystem übernommen werden kann. Auch neue Policies oder Änderungen in größerem Umfang (Bsp.: Ändern eines Attributs bei 25.000 Benutzern) werden zunächst in der Entwicklungsumgebung vorgenommen, bevor diese in der Produktivumgebung ausgeführt werden.

SAP-Treiberentwicklung

Da am zentralen MetaDirectory der GWDG auch das SAP-System der Universität angebunden ist, haben wir eine eigene, spezialisierte Entwicklungsumgebung für den SAP-Bereich aufgebaut, da dieser sich in vielen Bereichen von der Treiberentwicklung von Verzeichnissen wie Windows AD oder LDAP deutlich unterscheidet.

DISASTER RECOVERY UND BACKUP



Trotz redundant ausgelegtem MetaDirectory haben wir zusätzlich noch eine exakte Kopie der gesamten produktiven Umgebung in der Servervirtualisierung erstellt. Diese wird täglich um 2:00 Uhr automatisch synchronisiert. Sollte der sehr unrealistische Fall eintreten, dass beide MetaDirectory-Server nicht verfügbar sind oder auch das eDirectory in beiden Systemen irreversibel zerstört ist, so können wir innerhalb von zwei Stunden die gesamte MetaDirectory-Umgebung wiederherstellen. Generell würde sich aber ein Totalausfall des MetaDirectory nicht unmittelbar auf die angebundenen Verzeichnisse auswirken. Die am MetaDirectory angebundenen Verzeichnisse bleiben autonom und Benutzeranmeldungen sind auch bei Ausfall der MetaDirectory-Server weiterhin möglich. Eine direkte Anmeldung an das MetaDirectory ist, abgesehen vom IdM-Portal, genau aus diesem Grund auch nicht vorgesehen. Anmeldungen der Benutzer erfolgen immer an den am MetaDirectory angebundenen Verzeichnissen. Allerdings wären Änderungen und Neuanlagen bei einem Totalausfall des MetaDirectory bis zur Wiederherstellung nicht mehr möglich.

Backup

Zusätzlich wird einmal täglich ein Backup zu den GWDG-TSM-Backupsystemem durchgeführt. Insbesondere das MetaDir-Verzeichnis (eDirectory) wird hierbei gesichert. Darüber hinaus wird auch untereinander eine Sicherungskopie des eDirectory verschlüsselt auf allen IdM-Servern abgelegt und mit bis zu 20 Tagen historisiert. Hierdurch haben wir die Möglichkeit, auf den Inhalt und Zustand des gesamten MetaDirectory zu jedem Zeitpunkt der vergangenen 20 Tage zurückzugreifen.

ZENTRALE AUTHENTIFIZIERUNGSDIENSTE

Die GWDG verfügt über mehrere Verzeichnisdienste, welche unter anderem auch für die Authentifizierung der Benutzer zuständig sind. Das sind primär der Windows-AD-Cluster, der LDAP-Cluster sowie die RADIUS-Server. Benutzer melden sich für die Nutzung zentraler Dienste der GWDG immer an einer der genannten Authentifizierungsumgebungen an. Eine Anmeldung an das MetaDirectory selbst findet hierbei nicht statt. Das MetaDirectory stellt das führende Verzeichnis für alle angebundenen Authentifizierungsserver dar, weil die Authentifizierungsserver direkt vom MetaDirectory versorgt werden.

Authentifizierungsserver am MetaDirectory

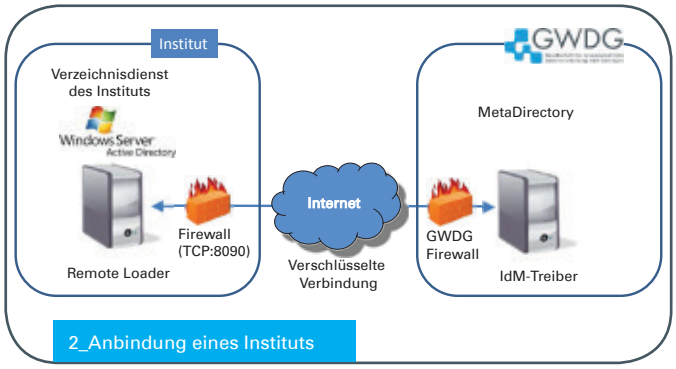
Alle Authentifizierungsserver sind unmittelbar am MetaDirectory angebunden, sodass sichergestellt ist, dass alle Benutzer und deren Attribute in allen Systemen nach den Vorgaben der Treiber synchronisiert sind. Die Harmonisierung unterschiedlicher Verzeichnisdienste und Authentifizierungssysteme ist ohnehin eine der entscheidenden Aufgaben eines MetaDirectory.

Zentrale Dienste der GWDG

Über die Authentifizierungsserver sind zahlreiche Dienste der GWDG verfügbar. Insbesondere Exchange, eduroam, Sharepoint, virtuelle Server, Cloudservices u. v. m. sind unmittelbar durch die Synchronisation der Authentifizierungsserver über das MetaDirectory für die Benutzer erreichbar. Institute, deren lokales Verzeichnis am MetaDirectory der GWDG angebunden ist, profitieren damit unmittelbar von der Bereitstellung neuer Dienste bei der GWDG.

METADIRECTORY IM DETAIL

Anbindung externer Verzeichnisse an das MetaDirectory



Im Folgenden werden die Anbindung eines Instituts an das MetaDirectory vorgestellt, sowie die Abarbeitung eines Prozesses exemplarisch demonstriert. Das Beispiel geht von einem Windows AD als lokalem Verzeichnisdienst im Institut aus. Die Verbindung zum Verzeichnisdienst des Instituts erfolgt über einen Remote Loader, der verschlüsselt die Daten in beide Richtungen überträgt. Alle Änderungen am Institutsverzeichnis (Windows) werden vom Remote Loader erkannt und anhand definierter Policies des dazugehörigen IdM-Treibers entsprechend übertragen.

Beispiel: Anlage eines Benutzers

Das folgende Beispiel (siehe Abbildung 3) stellt den Ablauf bei Anlage eines Benutzers im Verzeichnis des Instituts (Windows AD) vereinfacht dar.

1 Ein Benutzer mit den relevanten Attributen wird im Windows AD des Instituts durch den Institutsadministrator angelegt.

2 Der auf dem Windows-AD-Server installierte Remote Loader erkennt das Ereignis „Anlage eines Benutzers“ und überträgt die Daten (Attribute) sowie den Typ des Ereignisses (add) an das MetaDirectory in Form eines XML-Dokuments.

3 Der entsprechende Treiber für die Anbindung des Windows AD empfängt die Daten. Hier werden nacheinander verschiedene Bereiche im Treiber durchlaufen:

a) Mapping: Anschließend erfolgt ein Attributsmapping, in dem der Name der Attribute entsprechend angepasst werden kann. Das Mapping erfolgt zwischen der Attributsbezeichnung innerhalb des MetaDirectory sowie der Attributsbezeichnung im angeschlossenen Windows AD. Beispiel: Im Windows AD lautet der Username *samAccountName* im Vergleich dazu im IdM *UniqueID*. Es ist somit ein Mapping zwischen *samAccountName* und *UniqueID* definiert, das das gleiche Attribut in beiden Verzeichnissen unter einem unterschiedlichem Namen verfügbar macht.

b) Filter: Zunächst werden die Daten durch einen Filter geschickt, der definiert, welche Attribute für die Übertragung relevant sind bzw. auf welche Attribute der Treiber „reagieren“ soll.

c) Event-Type: In einem weiteren Schritt wird der Ereignis-Typ ausgewertet. Hier wird überprüft, ob das Ereignis ein „add“, „modify“, „delete“oder „sync“ ist. Abhängig vom Ereignis-Typ werden dann unterschiedliche Bereiche des Treibers durchlaufen, in denen jeweils Policies definieren, wie mit den Attributen zu verfahren ist. In unserem Beispiel handelt es sich um den Ereignis-Typ=add, da ein Benutzer unter Windows AD neu hinzugefügt wurde.

4 In Schritt 4 erfolgt die Überprüfung, ob der Benutzer bereits im MetaDirectory existiert. Ist das der Fall, so wird natürlich kein neuer Benutzer hinzugefügt, sondern es werden lediglich die Inhalte aller übertragenen Attribute mit den bereits im MetaDirectory gespeicherten Attributen des Benutzers verglichen und basierend auf den Einstellungen im Treiber entsprechend angepasst bzw. übernommen (Schritt 5b). Hierbei ist zusätzlich bei jedem Attribut eine Priorität festgelegt, die besagt, welcher Attributinhalt bei Ungleichheit zwischen angebundenem Windows AD und MetaDirectory Vorrang hat und damit übernommen wird.

5a Existiert der Benutzer im MetaDirectory noch nicht, so wird dieser basierend auf den Regeln des Treibers mit den entsprechenden Attributen angelegt. In der Praxis werden aufgrund der Neuanlage im MetaDirectory dann andere Treiber aktiv und fügen den Benutzer in anderen Zielverzeichnissen hinzu. Dieses können z. B. die zentralen Verzeichnisse der GWDG (Windows AD sowie LDAP) sein.

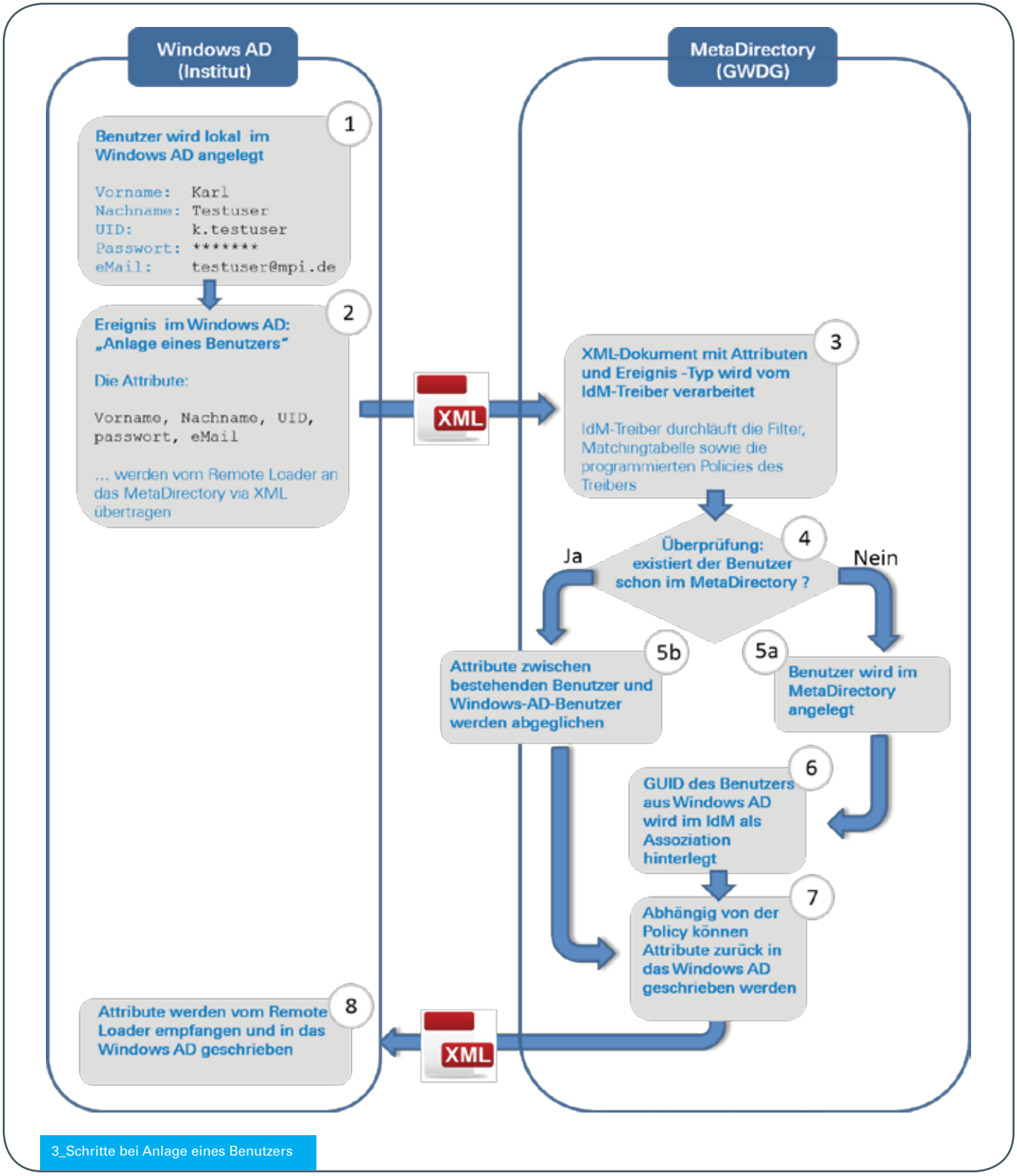
6 Wurde der Benutzer im MetaDirectory neu erzeugt, so wird anschließend durch den Treiber eine dauerhafte Verbindung zwischen dem im MetaDirectory angelegten Benutzer und dem Benutzer im Windows AD hergestellt. Hierbei wird die *GUID* des Windows-AD-Benutzers, welche im Windows AD eindeutig ist, als Attribut (Assoziation) am Benutzerobjekt im MetaDirectory abgespeichert. Dadurch ist bei etwaigen weiteren Attributsänderungen die Assoziation der Benutzer zwischen Windows AD und MetaDiretory immer gewährleistet und ein aufwendiges Suchen in den Verzeichnissen entfällt dadurch. Überdies kann auch die Position des Benutzers im Windows AD beliebig verändert werden, ohne dass dabei im MetaDirectory der korrespondierende Benutzer neu verknüpft oder gar neu angelegt werden muss.

7 Die Treiber im MetaDirectory sind bidirektional ausgelegt. Es können also in gleicher Weise auch Daten vom MetaDirectory zum angebundenen Verzeichnis übertragen werden. Abhängig von der programmierten Treiberlogik, können dabei Attribute in das Windows-Verzeichnis zurückgeschrieben werden.

8 Etwaige neue Attribute oder auch Änderungen bestehender Attribute werden vom MetaDirectory zum Windows AD über den Remote Loader entsprechend in das lokale Verzeichnis geschrieben.

Das Ergebnis

Nach der Anlage eines Benutzers im Windows AD und der daraus folgenden Anlage eines Benutzers im MetaDirectory, sehen die Attribute in den beiden Verzeichnissen wie in Abbildung 4 dargestellt aus.

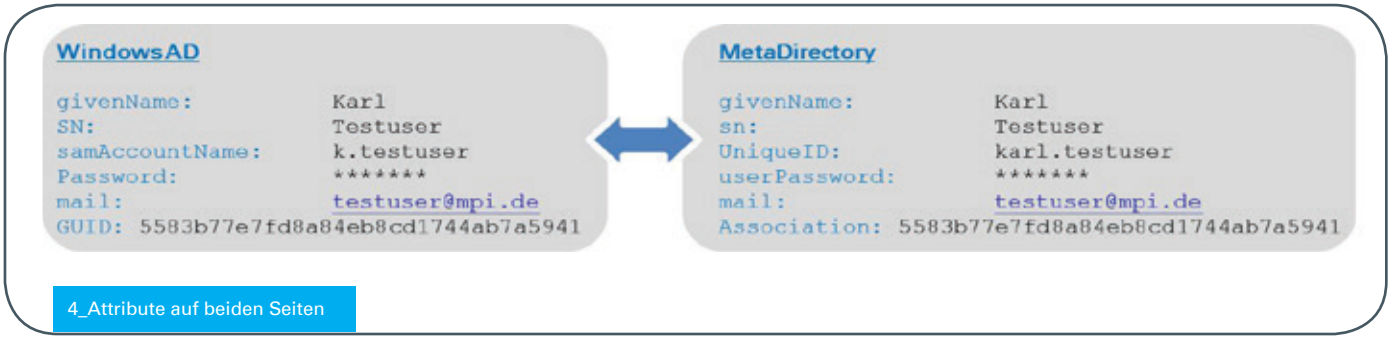


3_Schritte bei Anlage eines Benutzers

Das Beispiel ist im Vergleich zur Realität stark vereinfacht. Natürlich existieren in der Praxis viele weitere Attribute, die für die jeweiligen Verzeichnisse von Relevanz sind. Insbesondere im MetaDirectory wird noch eine Vielzahl von internen Attributen am Benutzerobjekt mitgeführt, die für die Verarbeitung im MetaDirectory entscheidend sind. Überdies ist im Windows AD das Passwort eines Benutzers natürlich nicht durch ein einfaches Attribut auszulesen. Diese Dinge wurden aufgrund der besseren Darstellbarkeit weggelassen.

Gruppenzugehörigkeiten

Natürlich sind die Treiber im MetaDirectory nicht nur auf Benutzerobjekte beschränkt. Es können auch Gruppen, Gruppenzugehörigkeiten und Rollen sowie ganze Strukturen (O,OU) zwischen angebundenem Verzeichnis und MetaDirectory synchronisiert werden. Das ist oft auch gängige Praxis, wenn ein externes Institut an das MetaDirectory angebunden wird.



4_Attribute auf beiden Seiten

WEITERE FÄHIGKEITEN DES METADIRECTORY

Abbildung von Workflows

Über das MetaDirectory können ganze Workflows abgebildet werden. Hierfür sind mehrere unterschiedliche Treiber zuständig, mit denen sich auch Genemigungsprozesse bei z. B. Anlage eines Benutzers abbilden lassen.

Workorder

Im MetaDirectory können auch Ereignisse zu vorgegebenen Zeitpunkten ausgeführt werden. Dafür sind sog. Workorder-Treiber zuständig. Diese werden zu den definierten Zeitpunkten aktiv und triggern dann weitere Prozesse. Das wird in der Praxis z. B. bei Löschvorgängen von Benutzern angewandt. Soll ein Benutzer gelöscht werden, so wird dieser zunächst nur als gelöscht markiert und eine „Workorder" sorgt dann zukünftig für das tatsächliche Löschen des Benutzers zu einem späteren Zeitpunkt. In gleicher Weise werden auch „expirationdate" für Kurzzeitaccounts realisiert.

Treiber für viele Anwendungsfälle

Im MetaDirectory der GWDG sind nicht nur Treiber für Windows AD und LDAP verfügbar. Das IdM der GWDG bietet Treiber für eine Vielzahl von Verzeichnissen, Datenbanken und Softwareschnittstellen bis hin zu spezialisierten Schnittstellen für Telefonzentralen. Natürlich stellt ein Treiber für eine standardisierte Anbindung nur das Grundgerüst dar und muss in jedem Fall durch Programmierung der Logiken in Form von Policies ergänzt werden. Hierbei erfolgt die Programmierung in einer eigenen MetaSprache, die speziell auf die Bedürfnisse eines MetaDirectory abgestimmt ist.

Erweiterbarkeit der Treiber durch andere Programmiersprachen

Sollte ein Treiber für einen speziellen Anwendungsfall nicht verfügbar sein oder eine Anbindung nicht standardisiert erfolgen können, so besteht in jedem Fall die Möglichkeit, durch eigene Erweiterungen die bestehenden Treiber zu ergänzen. Das kann bis zur Erstellung von eigenen Treibern gehen. Viele Funktionen lassen sich auch durch extern programmierte Funktionen in beliebigen Sprachen ergänzen (Java, Java-Script, ECMA) sowie auch andere Programmiersprachen (Perl, PHP, Python).

FAZIT

Das hier vorgestellte IdM der GWDG mit seinen wesentlichen Bausteinen kann nur in einem kleinen Umfang dargestellt werden. Die Integration sowie die Fähigkeiten des MetaDirectory der GWDG sind deutlich größer und würden den Rahmen dieses Artikels sicherlich sprengen. Es sollte vielmehr die prinzipielle Funktionalität und das Zusammenspiel der unterschiedlichen Komponenten beispielhaft dargestellt werden. Der Betrieb eines MetaDirectory ist in Umgebungen wie bei der GWDG und den damit angebundenen Instituten von Universität und Max-Planck-Gesellschaft ein unverzichtbares Werkzeug. Es erleichtert den Umgang mit Accountdaten in solch komplexen und heterogenen Strukturen und garantiert nachvollziehbar die Abbildung von Prozessen aus dem Bereich der Benutzerverwaltung. Wichtig bei Einsatz eines MetaDirectory ist die Tatsache, dass es bei einem etwaigen Totalausfall das MetaDirectory nicht zu einem weiteren Ausfall wesentlicher Dienste kommt. Die Autonomie der angebundenen Verzeichnisse ist entscheidend und eine Voraussetzung für den sicheren Betrieb eines MetaDirectory.

Für weitergehende Fragen erreichen Sie uns per E-Mail unter *idm@gwdg.de* oder *support@gwdg.de*

Workshop zu den Softwarelösungen Turnitin und iThenticate

Text und Kontakt:
Regina Bost
regina.bost@gwdg.de
0551 201-1831

Anke Bruns
anke.bruns@gwdg.de
0551 201-1519

Roland Groh
roland.groh@gwdg.de
0551 201-1838

Simon Heider
simon.heider@gwdg.de
0551 201-1840

Am 11.09.2013 findet von 14:00 bis 17:00 Uhr bei der GWDG ein Workshop zu den Softwarelösungen Turnitin und iThenticate zur Plagiatsprävention statt. Der Workshop, der von einem Vertreter der Herstellerfirma iParadigms abgehalten wird, dient zur Vorbereitung des Tests der Produkte. Er soll zur Klärung noch offener Fragen dienen, ist aber keineswegs Voraussetzung für die Teilnahme an der Testphase. Neben den Testteilnehmern steht der Workshop auch allen anderen Interessierten offen, die sich über diese Softwareprodukte informieren möchten.

Wie in den GWDG-Nachrichten 8/2013 angekündigt, stellt die GWDG von Mitte September 2013 bis Mitte November 2013 im Rahmen einer Testphase allen Interessierten kostenlos die Softwareprodukte Turnitin und iThenticate zur Verfügung.

Turnitin und iThenticate sind zwei Softwarelösungen des Herstellers iParadigms (*http://iparadigms.com*), die es ermöglichen, Texte mit einem Pool anderer wissenschaftlicher Texte zu vergleichen. Dies erleichtert die Erkennung von Plagiaten und ermöglicht so die Überprüfung der Einhaltung von wissenschaftlichen Standards.

Während Turnitin für die Originalitätsprüfung studentischer Arbeiten im Rahmen von Lehrveranstaltungen entwickelt wurde, eignet sich iThenticate besonders für die Beurteilung von Texten für Forschungseinrichtungen, wissenschaftliche Verlage oder Graduiertenkollegs.

Turnitin und iThenticate werden über eine Webschnittstelle bedient, so dass für die Nutzung außer einem gängigen Webbrowser keine weitere Softwareinstallation benötigt wird. Turnitin entspricht dem Learning Tools Interoperability (LTI) Standard und ist in verschiedene Learning-Management-Systeme integrierbar. Schnittstellen zu weiteren Systemen können mithilfe eines SDK für Java und PHP entwickelt werden.

Bei Interesse können Sie sich bis zum 9. September 2013 bei der unten angegebenen Kontaktadresse für die Testphase (Betreff: „Testphase“) und/oder den Workshop (Betreff: „Workshop“) anmelden. Aus organisatorischen Gründen bitten wir auch diejenigen, die bereits Interesse an der Testphase bekundet haben, sich für den Workshop mit einer kurzen E-Mail extra anzumelden. Danke! Nach Ablauf der Testphase möchten wir Sie um Ihre Einschätzung des getesteten Produkts bitten.

Bitte melden Sie sich auch, falls Sie grundsätzlich an der Plagiatserkennungssoftware interessiert sind, der Zeitraum für eigene Tests jedoch unpassend ist. So kann der Bedarf an einer derartigen Lösung besser abgeschätzt werden. Sollte die Testphase gut angenommen werden, ist geplant, Turnitin und iThenticate als Dienste der GWDG unseren Kunden zur Nutzung anzubieten.

Kontaktadresse: *test-plagiatserkennung@gwdg.de*

Workshop on Turnitin and iThenticate

Preceding the start of the testing period of plagiarism prevention software announced in GWDG News 8/2013, a workshop will be held at GWDG on September, 11, 2013 from 14 to 17 h with a representative of the software firm iParadigms. This workshop is not a formal requirement for test users but it offers the opportunity to clarify open questions around the software and the upcoming test. Of course, you are also invited if you are interested in the software but do not plan to participate in the test. If you are interested in the workshop and/or the test, please contact us via email at *test-plagiatserkennung@gwdg.de* with the subject „Workshop“ resp. „Test“ by September, 9. May we ask those who already registered for the test to let us know if you'll participate in the workshop as well? Thank you! We would like to ask your opinion on the software after the testing period.

Zentraler Fileservice mit dem HNAS-System

Text und Kontakt:
Stefan Teusch
stefan.teusch@gwdg.de
0551 201-1866

Dr. Eckard Handke
eckhard.handke@gwdg.de
0551 201-1548

Der von der GWDG betriebene HNAS-3090-Cluster ist ein gemeinschaftlich beschaffter, zentraler High-Performance-Fileserver der Projektpartner MPI für biophysikalische Chemie, studIT der Georg-August-Universität Göttingen und GWDG. Er ermöglicht eine nach Organisationen getrennte Datenhaltung und synchrone Datenspiegelungen zwischen entfernten Standorten. Getrennte Speicherung von Meta- und Nutzdaten und Festplattenpooling erhöhen die I/O-Leistung. Die Möglichkeit externer Datenauslagerung sichert die netto zur Verfügung stehende Kapazität.

EINLEITUNG

Im Jahr 2010 wurde durch das MPI für biophysikalische Chemie, die studIT der Georg-August-Universität Göttingen und die GWDG ein **Network Attached Storage (NAS)-System** als gemeinschaftliche zentrale Fileservice-Lösung beschafft. Mit Hilfe dieses Systems werden Homeverzeichnisse, Gruppen- und Abteilungslaufwerke der Projektpartner sowie Gruppen- und Institutslaufwerke der GWDG-Kunden bereitgestellt.

Ursprünglich wurde das System bei der Firma *BlueArc* beschafft, die mittlerweile von der Firma *Hitachi Data Systems (HDS)* akquiriert wurde. Die HDS hat wesentliche Komponenten des Systems ausgetauscht und zusätzliche Funktionen implementiert, durch die Stabilität, Redundanz und Leistungsfähigkeit des Systems nochmals gesteigert werden konnten. Den erfolgreichen Abschluss der damit verbundenen Umbauarbeiten möchten wir zum Anlass nehmen, das technische Konzept und die Funktionsweise des Massenspeichersystems **„Hitachi NAS" (HNAS)** im Detail vorzustellen.

NETWORK ATTACHED STORAGE

Im klassischen Massenspeichergeschäft lassen sich im Wesentlichen zwei Kategorien unterscheiden. **Raidsysteme** realisieren blockbasierten Speicher, der dem einbindenden Rechner als lokale Festplatte erscheint. Nach Formatierung mit einem Filesystem können auf diesem sogenannten Blockdevice Dateien abgelegt werden. Sollen diese Daten über die lokale Verwendung hinausgehend auch anderen über ein Netzwerk angebundenen, entfernten Rechnern (Klienten) verfügbar gemacht werden, sind entsprechende Fileservices zu installieren und zu betreiben.

NAS-Systeme arbeiten bereits auf dieser höheren, dateiorientierten Ebene. Die integrierten Fileservices sind für diese Aufgaben optimiert und bieten weitergehende Funktionalitäten an. Der Datenzugriff der Klienten erfolgt im Allgemeinen über Cifs/SMB2 oder NFS. Erstere bezeichnen die klassischen Protokolle der Windows-Umgebungen; NFS ist der Standard in der UNIX-Welt.

DER KONZEPTIONELLE AUFBAU

Das HNAS-System ist zur Erhöhung der Verfügbarkeit an zwei geographisch getrennten Standorten (GWDG und SUB) verteilt modular aufgebaut. Konkret besteht es im Frontend aus zwei NAS-Köpfen **HNAS-3090** im Clusterverbund. Besonderheiten dieser HNAS-Köpfe bestehen in der Implementierung eines system-internen, objektorientierten Filesystems, der Implementierung der Netzwerkstacks, der Fileservice-Protokolle und vieler weiterer Funktionen in sogenannten **Field-programmable Gate Arrays (FPGA)**. Der Fileservice ist somit fast vollständig in Hardware realisiert und nicht als wesentlich langsamere Softwarelösung implementiert.

Auf dem Köpfen werden sog. **Enterprise Virtual Server (EVS)** als logisch getrennte, eigenständige virtuelle Fileserver konfiguriert. Diese bieten den jeweiligen Projektpartnern und den Instituten der Universität Göttingen teilweise auch protokollübergreifenden Fileservice an. Die einzelnen EVS können Mitglied

Central file service with HNAS-3090

The HNAS-3090 cluster – a project of the MPI for Biophysical Chemistry, studIT of the Georg-August-Universität Göttingen and the GWDG – is a shared high-performance file server. It allows separated data management and synchronous data mirroring between remote locations. Separation of meta data and user data as well as disk pooling increase the I/O performance. The possibility to migrate data to external archives guarantees sufficient capacity.

unterschiedlicher Mircosoft-Active-Directory-Umgebungen sein. Insgesamt werden sowohl Datenzugriffe und -sichtbarkeiten als auch der Netzwerkverkehr der einzelnen Organisationen strikt voneinander getrennt.

Aufgrund des Clusterverbunds lassen sich einzelne EVS zwischen den Köpfen beider Standorte migrieren. Dies dient hauptsächlich der Dienstverfügbarkeit auch im Fehlerfall eines Kopfes, indem der HNAS-Kopf des jeweils anderen Standortes die EVS des ausgefallenen automatisch übernimmt. Durch EVS-Migrationen lässt sich darüber hinaus die Lastverteilung optimieren, und sie stellt im Wartungsfall einzelner Komponenten ein unverzichtbares Managementwerkzeug dar.

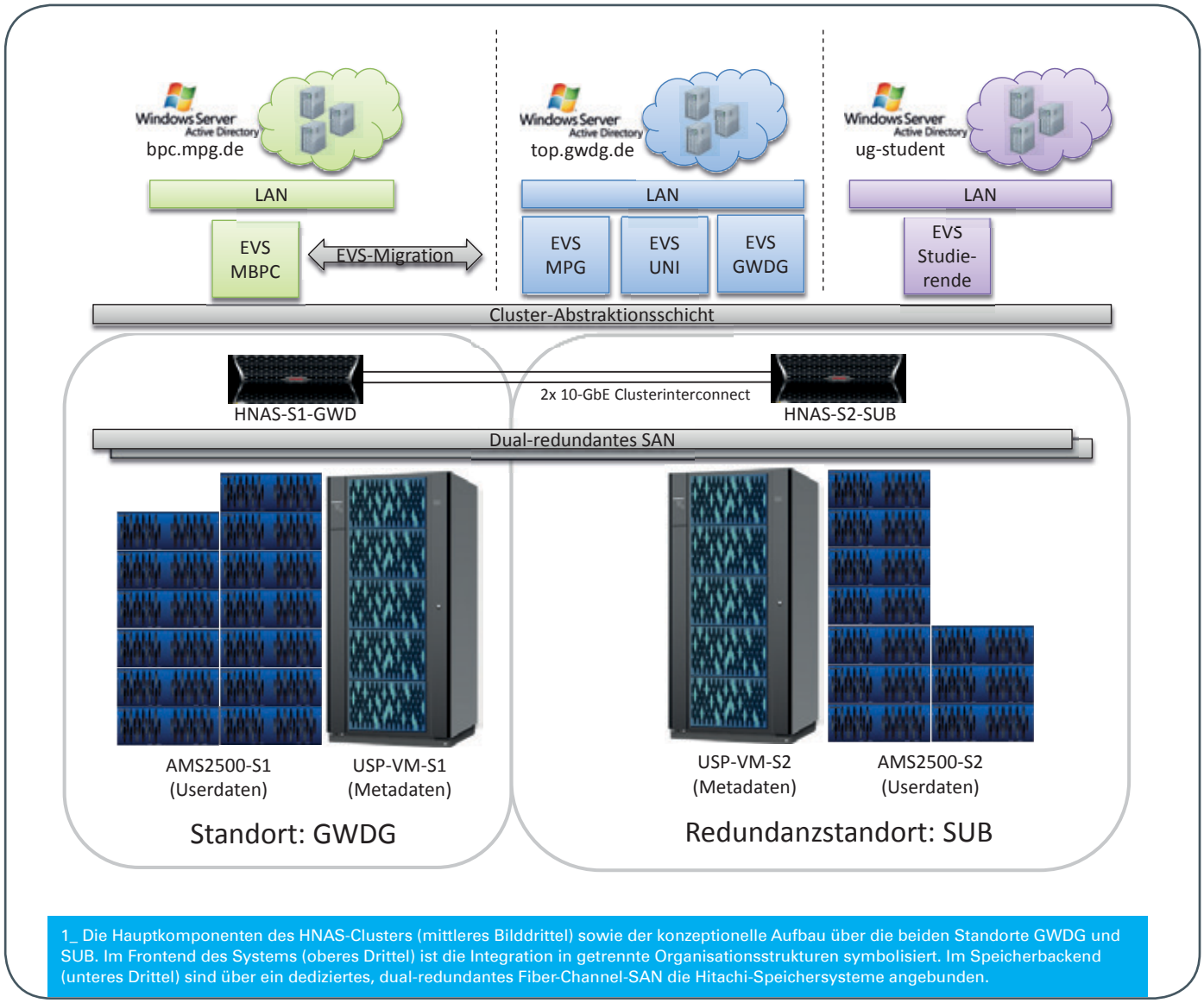
Im sog. Speicherbackend greifen die NAS-Köpfe auf Hitachi Speichersysteme mit insgesamt über einem halben PetaByte Kapazität zu.

Da bei Cifs/SMB2-Zugriffen der Windows-Welt ca. 70 % der Zugriffe aus lesender Operation der Metadaten (Dateizugriffsrechte, -eigentümer, Zeitstempel etc.) bestehen, werden die Metadaten von den eigentlichen Dateiinhalten getrennt und dediziert auf leistungsstarke Fiber-Channel-Festplatten gespeichert. Die eigentlichen Dateiinhalte, die Userdaten, werden auf kapazitiven SATA-Festplatten abgelegt. Durch diese getrennte Speicherung wird der Fileservice maßgeblich beschleunigt.

DAS SPEICHERBACKEND

Die Zahl der von einer Festplatte leistbaren Ein-/Ausgabeoperationen pro Sekunde ist konstruktionsbedingt relativ gering. Durch parallele, verteilte Datenspeicherung auf mehrere Festplatten kann jedoch die Leistung bei einem normalen Dateizugriff massiv erhöht werden. Daher werden mehrere Festplatten, genauer mehrere Raid-6-Verbünde, in einem Speicherpool zusammengefasst, in dem anschließend die Filesysteme angelegt werden.

Pro Standort existiert jeweils ein primärer Speicherpool mit 113 TeraByte Kapazität. Am Standort GWDG werden die Daten des MPIs für biophysikalische Chemie gespeichert, am Standort SUB die der Studierenden, die Institutslaufwerke der Universität Göttingen und einige GWDG-interne Daten. Zur Gewährleistung der Verfügbarkeit auch bei Ausfall eines Standortes werden beide Speicherpools synchron an den jeweils anderen Standort kopiert. Im Fehlerfall einzelner Speicherkomponenten bis hin zum Wegfall eines Standorts, z. B. aufgrund eines Stromausfalls, kann dadurch der Betrieb ohne zeitlichen Datenverlust automatisch mit der Datenkopie des zweiten Standorts fortgeführt werden.



1_ Die Hauptkomponenten des HNAS-Clusters (mittleres Bilddrittel) sowie der konzeptionelle Aufbau über die beiden Standorte GWDG und SUB. Im Frontend des Systems (oberes Drittel) ist die Integration in getrennte Organisationsstrukturen symbolisiert. Im Speicherbackend (unteres Drittel) sind über ein dediziertes, dual-redundantes Fiber-Channel-SAN die Hitachi-Speichersysteme angebunden.

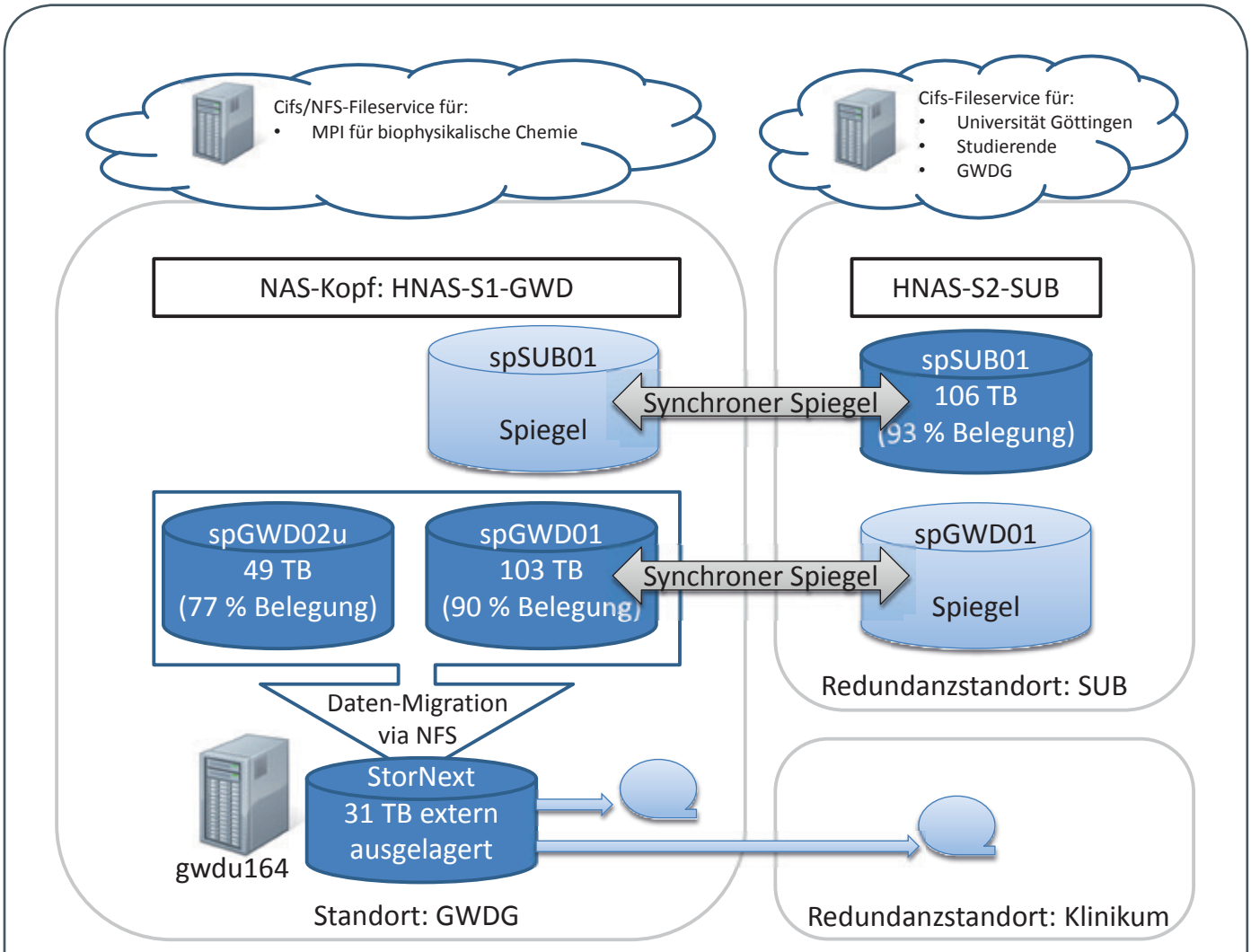
DATENMIGRATION

Eine weitere Besonderheit des HNAS-Systems ist die automatische Datenmigration, bei der eine transparente Verlagerung einzelner Dateien zwischen Speicherbereichen unterschiedlicher Güteklassen erfolgt. Anhand diverser Kriterien (z. B. Pfad, Dateigröße, Änderungs- und Zugriffszeiten) können z. B. Dateien von einem leistungsstarken, teuren in einen leistungsärmeren, kostengünstigen Speicherbereich verschoben werden. Im Original-Filesystem bleibt ein sog. „Stub" der ursprünglichen Datei bestehen, in dem der neue Ablageort der Datei referenziert wird. Greift ein Klient auf ein migriertes File zu, wird der Stub durch das Speichersystem interpretiert und die gewünschte Datei aus dem entsprechenden Speicherbereich bereitgestellt.

Die Datenmigration kann Dateien sowohl innerhalb des Systems als auch auf am Speicherbackend angeschlossene externe Systeme verschieben. Auf diese Weise werden beispielsweise selten genutzte Daten des MPIs für biophysikalische Chemie aus den gespiegelten Speicherbereichen herausgelöst und via NFS im zentralen UNIX-Filecluster der GWDG abgelegt. Das dort verwendete „StorNext"-Filesystem von Quantum ist ein Cluster-Filesystem, das sich über mehrerer Rechnerknoten erstreckt und eine sog. HSM-Funktionalität bietet, bei der Daten für den Nutzer transparent in zwei Kopien an zwei Standorten auf Magnetbänder verschoben werden.

Der Einsatz solcher Techniken dient vor allem der Kostenreduktion. Gerade bei der Speicherung unstrukturierter Daten wie PDFs, Office-Dokumente, Bilder etc. zeigt sich, dass lediglich mit einem sehr geringen Anteil der Daten täglich gearbeitet wird. Der Hauptanteil der Daten wird faktisch nicht oder äußerst selten verwendet. Diese Charakteristik besteht seit Anbeginn der Datenhaltung und führt gerade in Zeiten allgemeiner Datenexplosion häufiger zu Differenzen zwischen Speicherbetreibern und -nutzern. Im Hinblick auf nicht oder nur sehr selten verwendete Daten sind die Investitionen in leistungsstarke Online-Speichersysteme im Allgemeinen nicht zu rechtfertigen. Nicht mehr verwendete Daten, z. B. Dateien abgeschlossener Projekte etc., sind in Offline-Speichern wesentlich kostengünstiger und langfristig auch sicherer aufgehoben als auf Online-Speichern.

Die Datenmigration des HNAS-Systems kann obigen Prozess der Archivierung nicht ersetzen. Indem sie die Kapazitätsanforderung des synchron gespiegelten Speicherbereichs aber langsamer



2_Die Filesysteme sind in drei Speicherpools untergebracht: Je Standort ein primärer, synchron gespiegelter Pool mit 113 TB und ein nicht-gespiegelter Pool mit 63 TB am Standort GWDG. Selten verwendete Daten am Standort GWDG werden mittels Datenmigration in das zentrale UNIX-Cluster-Filesystem „StorNext" der GWDG ausgelagert. Diese hält zusätzlich zwei Kopien auf Magnetband an verteilten Standorten.

anwachsen lässt, führt sie jedoch zu deutlichen Einsparungen beim genutzten Online-Speicher. Der Einsatz solcher Datenmigrationen ermöglicht darüber hinaus die Realisierung noch weitergehender Konzepte beispielsweise im Bereich des Backups: Die unterschiedlichen Speicherebenen könnten anhand ihrer jeweiligen Anforderungen mit unterschiedlichen Snapshot- und/oder Backupzyklen versehen werden, um so die speziell beim Backup entstehende Belastung des Systems deutlich zu reduzieren.

WEITERE FUNKTIONALITÄTEN

Die internen Filesysteme können mit steigender Belegung vergrößert werden, so dass keine ungenutzten Kapazitäten längerfristig vorgehalten werden müssen. Im Bedarfsfall lassen sich Kapazitätserweiterungen sogar im laufenden Betrieb durchführen. Die maximale Größe eines Filesystems kann je nach Konfiguration bis zu 256 TeraByte betragen. Der den einzelnen Klienten zugewiesene Speicherplatz lässt sich durch frei konfigurierbare Obergrenzen (Quotas) flexibel verwalten.

Snapshots ermöglichen ein regelmäßiges Einfrieren des Filesystemzustands, womit eine zeitliche Versionierung der Daten aufgebaut werden kann, so dass ältere Versionen, versehentlich gelöschte oder anderweitig korrumpierte Dateien ohne Probleme wiederhergestellt werden können.

FAZIT

Das HNAS-System ist ein sehr flexibles Speichersystem, das mit umfangreichen Funktionen nahezu alle Erfordernisse des Einsatzes in einer wissenschaftlichen, heterogenen und vor allem nach Organisationen getrennten Umgebung abdeckt. Der bestehende Aufbau nutzt bei Weitem noch nicht alle bestehenden Möglichkeiten aus.

Die Verwaltung und der Betrieb eines gemeinsamen Fileservices für mehrere Organisationen konnten im Vergleich zum Betrieb mehrerer dedizierter Fileserver in jeder Organisation wesentlich vereinfacht bzw. verbessert werden. Dies drückt sich sowohl in gesteigerter Effizienz als auch in erhöhter Stabilität aus. Insbesondere bei sehr großen Filezahlen spielt das HNAS-System seine Stärken aus. Derzeit werden über 340 Mio. Dateien zentral verwaltet.

Ganz unabhängig von allen technischen Merkmalen des Systems zeigt das Projekt des gemeinschaftlich beschafften zentralen Fileservers, dass Mehrwerte und Synergien für die beteiligten Projektpartner mit verhältnismäßig geringem Aufwand erkannt und geschaffen werden können. Dies betrifft nicht nur den finanziellen Aspekt, sondern vor allem die gewonnene Arbeitszeitersparnis bei der Administration eines zentralen Systems im Vergleich zu mehreren ggfs. sogar unterschiedlichen dedizierten Systemen.

Kurz & knapp

Kontingentzuweisung für das vierte Quartal 2013

Die nächste Zuweisung von Institutskontingenten für die Inanspruchnahme von Leistungen der GWDG erfolgt am Dienstag, dem 1. Oktober 2013. Die Höhe der Kontingente wird den Instituten per Brief oder per E-Mail mitgeteilt. Die Bemessung der Institutskontingente erfolgte nach den Vorläufigen Richtlinien des Beirats der GWDG und den Ergänzungen der Beiratskommission für die Verteilung von IT-Leistung entsprechend dem Verbrauch im Zeitraum vom 01.03.2013 bis 31.08.2013. Nicht verbrauchte Kontingente werden zu 50 % in das nächste Quartal übertragen. Negative Verbrauchswerte werden zu 100 % mit dem neuen Institutskontingent verrechnet.

Jeder Benutzer kann den aktuellen Stand des Institutskontingents durch die Eingabe des Kommandos *kontingent* auf einer Workstation des UNIX-Clusters oder im WWW unter *http://www.gwdg.de/index.php?id=1678* abfragen. Dort besteht auch die Möglichkeit, Informationen über den Stand des separaten Druckkontingents abzurufen.

Falls in Ausnahmefällen das Institutskontingent nicht ausreichen sollte, können begründete Anträge über *http://www.gwdg.de/index.php?id=799* gestellt werden. Solche Anträge sollen bis zum **18.11.2013** eingereicht werden.

Glässer

Zusatztermin für den SharePoint-Kurs

Aufgrund größerer Nachfrage bieten wir einen weiteren Termin für den Kurs „Die SharePoint-Umgebung der GWDG“ an. Er findet am **29.10.2013 von 9:30 – 15:30 Uhr** im Kursraum der GWDG statt. Ausführliche Informationen zu diesem Kurs finden Sie unter *http://www.gwdg.de/index.php?id=1403*.

Otto

Betriebsausflug der GWDG am 12.09.2013

Am Donnerstag, dem 12.09.2013, findet der diesjährige Betriebsausflug der GWDG statt. Das Rechenzentrum bleibt an diesem Tag zwar zu den üblichen Zeiten geöffnet, es wird aber nur eine Minimalbesetzung an Personal anwesend sein. Wir bitten alle Benutzer und Besucher der GWDG, sich hierauf einzustellen.

Otto

Datenmanagement bei der GWDG – einheitliche Prozesse und integrierte Softwarelösungen für die Forschung

Text und Kontakt:
Oliver Schmitt
oliver.schmitt@gwdg.de
0551 201-2176

In der Ausgabe 1/2013 der GWDG-Nachrichten wurde das Datenmanagement der GWDG konzeptionell und inhaltlich beleuchtet, um die Herausforderungen im Wissenschaftsbetrieb zur Sicherstellung der Zitierbarkeit, Nachnutzung und Langzeitarchivierung von Forschungsdaten zu benennen und in die Datenmanagement-Dienste der GWDG einzuführen. Dieser Artikel ergänzt nun die konzeptionellen und technischen Realisierungsmöglichkeiten des Datenmanagements und zeigt, wie Aspekte der Ablage, Verwaltung, Langzeitarchivierung, Zugriffssteuerung und Referenzierbarkeit mit einem einheitlichen Ansatz in Forschungsprojekten mit Hilfe der Dienste der GWDG gelöst werden.

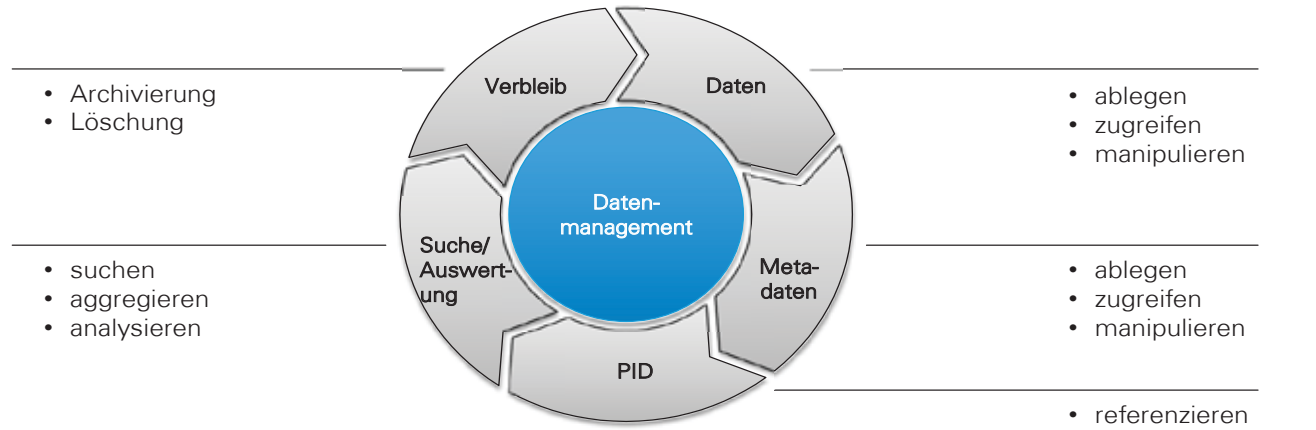
Der Umgang mit Forschungsdaten stellt die Wissenschaft unter verschiedenen Gesichtspunkten vor große Herausforderungen. Auf der einen Seite fordert die DFG seit langer Zeit, wissenschaftliche Primärdaten für mindestens zehn Jahre aufzubewahren [1]. Auf der anderen Seite ergeben sich durch eine sinnvolle Vernetzung und Nachnutzung neue Möglichkeiten in der Forschung. Dies ist für die unterschiedlichen Wissenschaftsdisziplinen mit ihren teilweise sehr großen Datenmengen keine triviale Aufgabe. Daher ist es sehr wichtig, das Datenmanagement bereits in der Antragsphase von Forschungsvorhaben zu berücksichtigen. Die GWDG bietet als verlässlicher Partner verschiedene Möglichkeiten, das Datenmanagement konzeptionell, technisch und operativ durchzuführen und zu unterstützen.

KONZEPTIONELLE UMSETZUNG DES DATENMANAGEMENTS

Datenmanagement in der Forschung ist stark mit der Arbeit der Forscher und Wissenschaftler verzahnt. Diesem Grundsatz zu folgen, bedeutet, dass bei der Realisierung von digitalen Forschungsinfrastrukturen, wie virtuellen Forschungsumgebungen, kollaborativen Portalen oder fachgebietsspezifischer Individualsoftware, das Datenmanagement und seine Prozesse in die Prozesse der Wissenschaftler einzubeziehen. Datenmanagement ist kein Selbstzweck, sondern ein generischer Teil der computergestützten Forschungsarbeit und daher aus kaum einer Fachdisziplin wegzudenken. Datenmanagement muss hierbei den kompletten Zyklus aus Datenentstehung in Experimenten, Literaturarbeit und Beobachtungen, Datenaggregation und Analyse durch

Data management at the GWDG – uniform processes and integrated software solutions for research

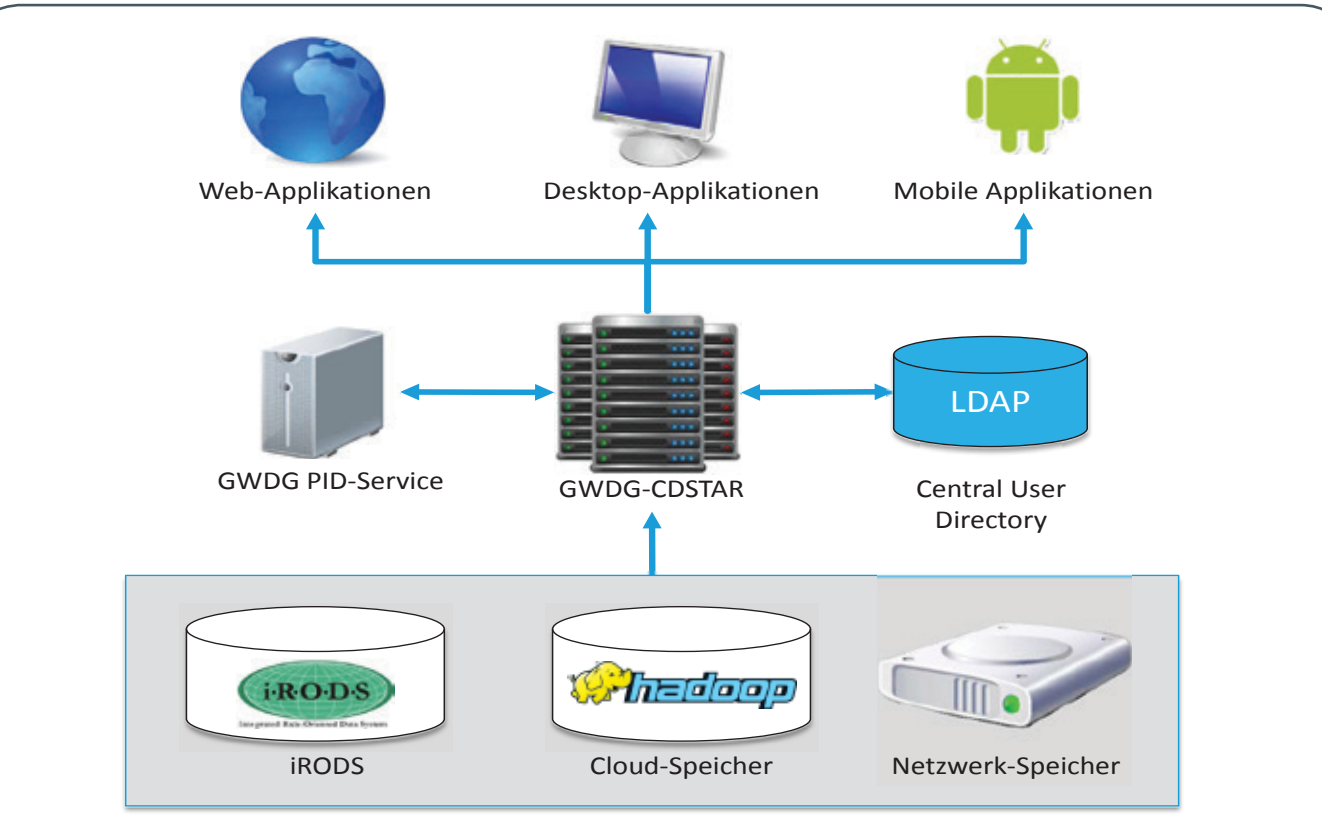
Number 1/2013 of the GWDG News points out the aspects and challenges of data management in the scientific usage for citing, long-term archiving and reusing of research data. This article follows the topic of research data management and shows the holistic concept of the GWDG for managing research data and gives in insight into the technical solutions offered by the GWDG. The importance of data management is to understand it as an accompanying process where software and research activities are tied together interfering the daily work of the researchers, while not impeding research activities. Also the importance for running data management infrastructure sustainable is a key for successful data management and should be considered in project design phase.



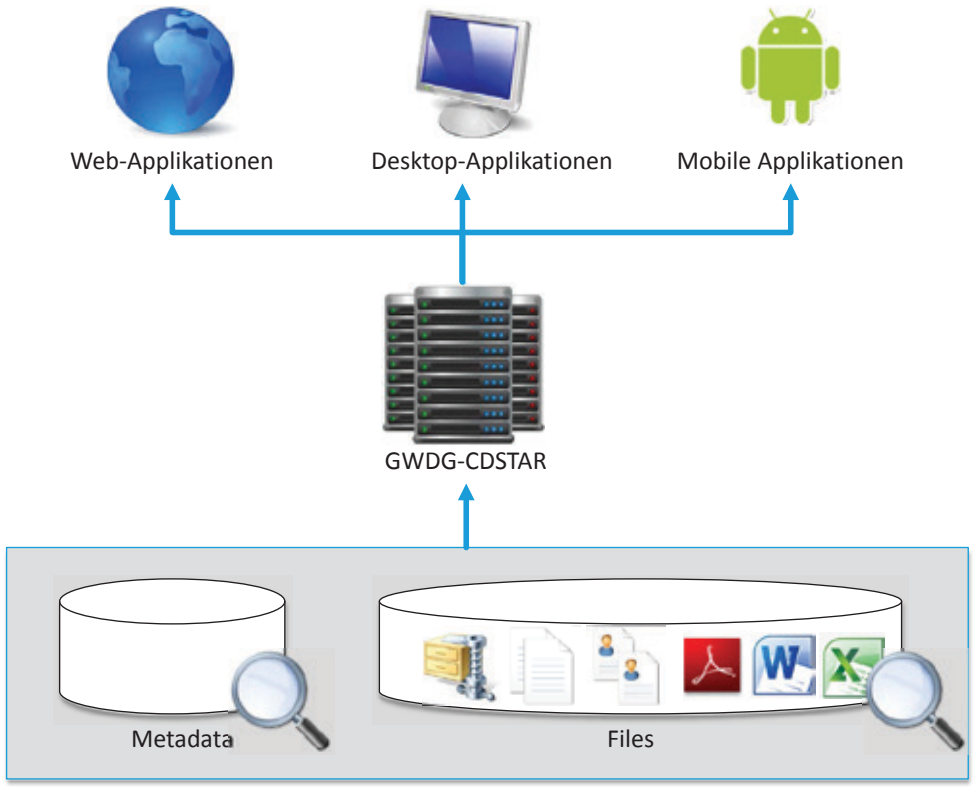
1_Forschungsdatenmanagement ist ein einheitlicher Prozess, der die fachwissenschaftlichen Arbeitsabläufe begleitet.

Algorithmen und Auswertungstools, Ablage der Daten in Repositorien sowie Überführung in Langzeitarchivierungslösungen erfassen. Die formale Erfassung der relevanten wissenschaftlichen Arbeitsschritte erlaubt es, das Datenmanagement individuell für das jeweilige Forschungsvorhaben anzupassen und damit den Anforderungen der Projektträger zur Sicherung der guten wissenschaftlichen Praxis Rechnung zu tragen und die Förderbarkeit von Forschungsvorhaben auf nationaler und internationaler Ebene sicherzustellen.

Zur Umsetzung des Datenmanagements bietet die GWDG als Projektpartner die Möglichkeit, wissenschaftliche Arbeitsabläufe in Form von Prozessdiagrammen und Workflow-Charts in gängigen Prozessmodellierungssprachen wie BPMN (Business Process Model and Notation) [2] zu explizieren und Maßnahmen zum Forschungsdatenmanagement in die wissenschaftlichen Arbeitsabläufe zu integrieren. Ziel ist es dabei, die Arbeit der Wissenschaftler durch das Forschungsdatenmanagement zu unterstützen und die dafür nötigen Aufwände in der täglichen wissenschaftlichen Arbeit gering zu halten. Zusätzlich sind die Anforderungen an das Datenmanagement in Datenmanagementplänen zu sammeln, die eine globale Sicht auf die Handhabung von Forschungsdaten und die Datenströme im Forschungprojekt haben. Die GWDG unterstützt und berät hierzu bei der Ausarbeitung sog. Datenmanagement Policies, die dann z. B. Teil der Geschäftsordnung in



2_Datenmanagementsoftware steuert als Middleware zentral den Datenzugriff und vereinfacht die Komplexität.



3_Eine einheitliche Middleware im Datenmanagement schlägt Brücken zwischen fachwissenschaftlichen Anwendungen.

Sonderforschungsbereichen (SFB) werden können. In Ausgabe 8/2013 der GWDG-Nachrichten wurden bereits GWDG-Projektbeteiligungen aus den Natur- und Geisteswissenschaften genannt, bei denen Datenmanagement in die Prozesse der Wissenschaftler integriert wird und flankierende Maßnahmen wie Datenmanagementpläne und Datenmanagement-Policies sich in Vorbereitung befinden oder bereits erfolgreich verabschiedet worden sind.

TECHNISCHE UMSETZUNG DES DATENMANAGEMENTS

Da die Digitalisierung von Forschungsarbeit im Jahr 2013 sehr weit fortgeschritten ist und immer mehr wissenschaftliche Disziplinen verstärkt darin investieren, stellt die Integration von Datenmanagement in die Anwendungslandschaft des Forschungsprojekts zunächst eine architektonische Herausforderung in der Softwareentwicklung dar. Je nach angestrebter Unterstützung bei den wissenschaftlichen Arbeitsprozessen kann eine Datenmanagementlösung als ein Baustein neben den wissenschaftlichen Fachanwendungen, z. B. ausschließlich für Archivierungszwecke genutzt werden oder vollintegriert als Datendrehscheibe im Forschungsprojekt verwendet werden. Die Datenmanagement-Software übernimmt hierbei die Abstrahierung zur eigentlichen Speicherlösung (z. B. Solid State Disks, Festplatten oder Tape), die Verwaltung der Daten und die automatische Registrierung der Daten bei einem Persistent-Identifier-Service zur dauerhaften Referenzierung der Daten. Hierbei übernimmt die Datenmanagement-Software die Rolle einer integrativen Komponente als sog. Middleware zwischen Datenzugriff, Recherche und fachspezifischer Software (siehe Abbildung 2).

Der Einsatz der Datenmanagement-Middleware bedeutet die Verortung grundlegender Funktionalitäten bei der Integration von Datenmanagementkomponenten. Zum einem müssen Datenspeicher angebunden werden, die die Anforderungen an Datennutzbarkeit und Archivierung erfüllen. Dies sind in der Regel hierarchische Speichersysteme (HSM), die in einer Kombination aus Festplatten und Tape funktionieren. Zusätzlich müssen existierende und neuzuschaffende wissenschaftliche Fachanwendungen wie virtuelle Forschungsumgebungen, Desktop-Anwendungen oder gar mobile Applikationen, die auf Tablet und Smartphone ablaufen, integriert werden. In Einzelfälle werden auch Geräte zur Beobachtung und Analyse von Forschungsexperimenten wie Hochgeschwindigkeitskameras oder OCR-Softwarekomponenten mit angebunden.

TECHNISCHE UMSETZUNG MIT GWDG CDSTAR ALS INTEGRIERTE SOFTWARELÖSUNG

Die GWDG bietet als zentrale Komponente einer Datenmanagementlösung **GWDG CDSTAR** (Common Data Storage Architecture) als einen integrierten Object Storage an, der den Anforderungen an das Forschungsdatenmanagement im wissenschaftlichen Kontext Rechnung trägt. GWDG CDSTAR folgt hierbei den Prinzipien etablierter Object-Storage-Dienste wie Amazon S3 [2] oder Microsoft Azure Blob Storage [3] durch der Bereitstellung einer offenen, HTTP-basierten REST(Representational state transfer)-Schnittstelle [4] [5] zur leichten Integration von Anwendungen. GWDG CDSTAR bietet die Möglichkeit zur Speicherung von Forschungsdaten in Objekten, die mit Metadaten versehen werden können. Alle Datensätze werden von GWDG CDSTAR automatisch mit einem Persistent Identifier versehen, der von dem GWDG-betriebenen EPIC-Dienst [7] bereitgestellt wird. Zusätzlich umfasst GWDG CDSTAR eine Suchmaschine, mit der über die Metadaten und Forschungsdaten gesucht werden kann. Hierzu können Suchanfragen über den Volltext und die einzelnen Metadatenfelder

gestellt werden. GWDG CDSTAR unterstützt über 40 Dateiformate u. a. PDF, Microsoft-Office-Dateien, XML-Dateien, ZIP-Archive und Metadaten aus Video- sowie Audiodateien [6]. Der Suchindex wird automatisch im Hintergrund bei Dateizugriffen aktualisiert. Da sowohl der Datenzugriff auf Forschungsdaten und die dazugehörigen Metadatensätze als auch das Absetzen von Suchanfragen eine einheitliche REST-Schnittstelle nutzt, können existierende und neue wissenschaftliche Fachanwendungen leicht an alle GWDG-CDSTAR-Funktionalitäten angebunden werden. Die reichhaltige Unterstützung des REST-Paradigmas führt zu schnellen Ergebnissen bei der Nutzung von GWDG CDSTAR in allen gängigen Programmiersprachen wie Java, C/C++, Python, Ruby und C#.

GWDG CDSTAR nutzt ein rollenbasiertes Modell für Datenzugriff, um selektive Zugriffe feingranular zu steuern und Datenzugriffsrichtlinien in besonderen Einsatzszenarien in der Forschung durchzusetzen. Die Anbindung an LDAP-Berechtigungsdatenbanken kann erfolgen und eine Shibboleth-basierte [9] Single-Sign-On-Anbindung ist im Herbst 2013 verfügbar. Hiermit lässt sich GWDG CDSTAR auch als Archivierungsanwendung für bestehende Forschungsinfrastrukturen nutzen, die bereits über ein Rollenmodell verfügen und sich für eine Langzeitarchivierungslösung für die im Projekt gesammelten Forschungsdaten interessieren.

DATENMANAGEMENT ERFORDERT NACHHALTIGE BETRIEBSKONZEPTE

Datenmanagement als fortwährender Prozess begleitet alle Projektphasen einer digitalen Forschungsinfrastruktur von den ersten Anwendungsfallbeschreibungen bis zur Fertigstellung der Software. Ist diese Grundlage verstanden, so muss auch über einen nachhaltigen Betrieb der Datenmanagementlösung nachgedacht werden. Wie im ersten Datenmanagementartikel der GWDG-Nachrichten in Ausgabe 1/2013 beschrieben, muss eine lange Verfügbarkeit der Daten gewährleistet werden. Dies umfasst beispielsweise den Betrieb der Server über einen Zeitraum, der länger als das Forschungsprojekt ist, oder die Überführung der Daten in ein Datenarchiv zur Offline-Lagerung mit anschließender Aufbewahrung der Softwareinfrastruktur. Um dies zu realisieren, sind zwei primäre Dinge nötig. Zum einen ist die Erstellung eines Betriebskonzepts nötig, das beschreibt, wie die Datenmanagementlösung und die dazugehörigen Anwendungen im Regelbetrieb betrieben und gewartet werden. Zum anderen umfasst es die Abstellung von Projektressourcen für die Umsetzung der Maßnahmen im Betriebskonzept. Für ein erfolgreiches Datenmanagement in einem zeitlich begrenzten Forschungsvorhaben ist es daher nötig, Ressourcen für die Einlagerung der Daten im Langzeitarchiv und die fachgerechte Aufbewahrung der digitalen Infrastruktur am Ende der Projektlaufzeit bereitzustellen und möglichst dauerhafte Verantwortlichkeiten in den teilnehmenden Forschungseinrichtungen zu benennen. Die Archivierung der Daten und die fachgerechte Aufbewahrung stellt sicher, dass eine Nachnutzung der Daten möglich ist und dass die digitale Forschungsinfrastruktur sich in einem definierten, d. h. überwachten und gewarteten Zustand befindet. Die GWDG berät und unterstützt gerne im Dialog mit den Fachwissenschaftlern bei der Ausarbeitung von Betriebskonzepten und analysiert bestehenden Konzepte und Lösungen auf Nachhaltigkeit in Bezug auf Datenmanagement und Nachnutzbarkeit von Forschungsdaten.

FAZIT

Datenmanagement für Forschungsdaten ist ein Prozess, der die gesamten fachwissenschaftlichen Aktivitäten begleitet und dessen Umsetzung in Software Teil einer digitalen Forschungsinfrastruktur ist. Die massiv zugenommene Digitalisierung der Forschungsaktivitäten in den letzten Jahren benötigt daher leistungsfähige Datenmanagementlösungen. Die GWDG bietet mit CDSTAR eine integrierte Datenmanagementlösung als Middleware an, die für ein breites Spektrum an Forschungsdatenmanagement in vielen Fachdisziplinen genutzt werden kann und die Integration vielfältiger Systeme über einheitliche REST-Schnittstellen erlaubt. Die Expertise der GWDG in zahlreichen Forschungsprojekten hilft dabei, Datenmanagement im Rahmen der Forschungsprojekte zu verankern und Maßnahmen zum Betrieb und der Langzeitarchivierung erfolgreich umzusetzen, um eine Nachnutzung der Daten zu ermöglichen und Förderungsrichtlinien der Projektträger gerecht zu werden.

REFERENZEN

[1] Deutsche Forschungsgemeinschaft. (1998) Vorschläge zur Sicherung gute Wissenschaftlicher Praxis. [Online]. *http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf*

[2] Object Management Group. (2013, August) BPMN Specifications. [Online]. *http://www.omg.org/spec/BPMN/index.htm*

[3] Amazon Web Services, Inc. (2013, June) Amazon S3, Cloud Computing Storage for Files, Images, Videos. [Online]. *http://aws.amazon.com/s3/*

[4] Microsoft Corporation. (2013, June) Blob Service REST API. [Online]. *http://msdn.microsoft.com/en-us/library/windowsazure/dd135733.aspx*

[5] Fielding, Roy T. and Taylor, Richard N., „Principled design of the modern Web architecture," ACM Trans. on Internet Technol., pp. 115-150, May 2002.

[6] Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., Berners-Lee, T., Fielding, R. (1999, April) Hypertext Transfer Protocol – HTTP/1.1. [Online]. *http://www.w3.org/Protocols/rfc2616/rfc2616.html*

[7] European Persistent Identifier Consortium. (2013, June) EPIC – European Persistent Identifier Consortium. [Online]. *http://pid-consortium.eu/*

[8] The Apache Software Foundation. (2013, June) Apache Tika – Supported Document Formats. [Online]. *http://tika.apache.org/0.10/formats.html*

[9] Shibboleth Consortium. (2013, June) Shibboleth. [Online]. *http://shibboleth.net/*

Die GWDG sucht für innovative Projekte aus den Bereichen Datenmanagement und virtuelle Forschungsumgebungen zum nächstmöglichen Zeitpunkt zwei

Wissenschaftliche Mitarbeiterinnen/ Mitarbeiter

Die Stellen sind zur Teilzeit geeignet. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgeltgruppe E 13 TVöD vorgesehen. Die Stellen sind auf zwei Jahre befristet.

In mehreren Projekten werden bei der GWDG Dienste entwickelt, mit denen Datenmanagement und virtuelle Forschungsumgebungen für Forscher unterschiedlicher Fachdisziplinen nutzbar gemacht werden. An den Projekten sind neben der GWDG weitere Partner aus Industrie und Wissenschaft beteiligt.

Wir bieten Ihnen:

- Ein Umfeld für Forschung und Entwicklung zu aktuellen IT-Trendthemen
- Ein spannendes Aufgabengebiet an der Schnittstelle zwischen Industrie und Forschung
- Einen direkten praktischen Bezug Ihrer Forschungsarbeit mit direkter Anbindung an ein Rechenzentrum
- Die Möglichkeit zur Promotion an der Georg-August-Universität Göttingen

Zu Ihren Aufgaben gehört:

- Erarbeitung von Datenmanagement- und Langzeitarchivierungskonzepten
- Entwurf und Entwicklung von Datenmanagementmethoden und Implementierung der zugehörigen Prozesse
- Konzeption und Umsetzung von virtuellen Forschungsumgebungen
- Präsentation Ihrer Ergebnisse auf nationalen und internationalen Konferenzen sowie bei den Projektpartnern

Wir erwarten von Ihnen:

- Ein abgeschlossenes Studium der Informatik oder eines verwandten technischen oder naturwissenschaftlichen Faches
- Erfahrung in einer für web-basierte Anwendungen geeigneten Programmiersprache
- Engagement und Leistungsbereitschaft
- Gute Kommunikationsfähigkeiten
- Gute Englischkenntnisse in Wort und Schrift
- Flexibilität, Eigeninitiative und die Fähigkeit zu selbständigem Arbeiten sind ebenso unerlässlich wie ein hohes Maß an sozialer Kompetenz und der Bereitschaft zur Integration in ein hochmotiviertes Team
- Kenntnisse in mindestens einem der folgenden Gebiete:
  - Datenmanagement
  - Langzeitarchivierung
  - Fortgeschrittene Kenntnisse in Java und Java-Enterprise-Technologien
  - Fortgeschrittene Kenntnisse in einer Skriptsprache (Python, PHP, Perl, Ruby o. ä.)
  - Administration von Linux-Servern

Die GWDG will den Anteil von Frauen in den Bereichen erhöhen, in denen sie unterrepräsentiert sind. Frauen werden deshalb ausdrücklich aufgefordert, sich zu bewerben. Außerdem sind Bewerbungen Schwerbehinderter ausdrücklich erwünscht. Interessiert? Dann bewerben Sie sich bitte online bis zum **30.09.2013** unter *https://s-lotus.gwdg.de/gwdgdb/age/bewerbungen_ag_e_2013_09.nsf/bewerbung*

Fragen zur ausgeschriebenen Stelle beantworten Ihnen:

Herr Dr. Ulrich Schwardmann
Tel.: 0551 201-1542
E-Mail: ulrich.schwardmann@gwdg.de oder

Herr Dr. Philipp Wieder
Tel.: 0551 201-1576
E-Mail: philipp.wieder@gwdg.de

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Stellenangebot

Die GWDG sucht zum 1. November 2013 zur Unterstützung der Arbeitsgruppe „Nutzerservice und Betriebsdienste“ eine/einen

Technische Angestellte/Technischen Angestellten

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgeltgruppe E 10 vorgesehen. Die Stelle ist zunächst auf zwei Jahre befristet, eine spätere Verlängerung oder Entfristung ist je nach Mittelfreigabe nicht ausgeschlossen.

Aufgabenbereich

- Mitarbeit beim Ausbau und Betrieb einer Microsoft-SharePoint-2013-Umgebung
- Mitarbeit bei der Integration einer umfangreichen externen Microsoft-Sharepoint-2010-Umgebung
- Mitarbeit bei Beratung und Problemlösung im Bereich SharePoint aufgrund von Kundenanfragen

Anforderungen

- Fachhochschul- oder Bachelor-Abschluss vorzugsweise in Informatik oder verwandten Fächern
- Gute Kenntnisse im Bereich Microsoft SharePoint ab Version 2010 für Support und Customizing
- Schnelle Lernfähigkeit sowie gute Kommunikations- und Teamfähigkeit
- Wünschenswert sind eingehende Kenntnisse von Windows-Betriebssystemen sowie im Bereich Microsoft Active Directory
- Sprachkenntnisse in Wort und Schrift in Deutsch und Englisch

Die GWDG will den Anteil von Frauen in den Bereichen erhöhen, in denen sie unterrepräsentiert sind. Frauen werden deshalb ausdrücklich aufgefordert, sich zu bewerben. Außerdem sind Bewerbungen Schwerbehinderter ausdrücklich erwünscht. Bewerbungen bitte bis zum **09.09.2013** über das Online-Formular unter *https://s-lotus.gwdg.de/gwdgdb/agh/bewerbungen_ag_h_2013_07.nsf/bewerbung*

Stellenangebot

Fragen zur ausgeschriebenen Stelle beantworten Ihnen:

Frau Katrin Hast
Tel.: 0551 201-1808
E-Mail: katrin.hast@gwdg.de oder

Herr Dr. Konrad Heuer
Tel.: 0551 201-1540
E-Mail: konrad.heuer@gwdg.de

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Die GWDG sucht zum 1. November 2013 zur Unterstützung der Arbeitsgruppe „Nutzerservice und Betriebsdienste“ eine/einen

Technische Angestellte/Technischen Angestellten

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist in Entgeltgruppe E 10 vorgesehen.

Aufgabenbereich

- Mitarbeit im allgemeinen Kundensupport der GWDG im Rahmen eines Funktionszeitmodells
- Mitarbeit beim Betrieb einer Microsoft-Sharepoint-2013-Umgebung
- Mitarbeit bei Beratung und Problemlösung im Bereich Sharepoint aufgrund von Kundenanfragen
- Mitwirkung bei Beschaffungen im Rahmen der Aufgaben der Arbeitsgruppe

Anforderungen

- Fachhochschul- oder Bachelor-Abschluss vorzugsweise in Informatik oder verwandten Fächern
- Gute Kenntnisse von Windows-Betriebssystemen
- Schnelle Lernfähigkeit sowie gute Kommunikations- und Teamfähigkeit
- Vorteilhaft sind zusätzliche Kenntnisse von UNIX- oder Linux-Betriebssystemen
- Sprachkenntnisse in Wort und Schrift in Deutsch und Englisch

Die GWDG will den Anteil von Frauen in den Bereichen erhöhen, in denen sie unterrepräsentiert sind. Frauen werden deshalb ausdrücklich aufgefordert, sich zu bewerben. Außerdem sind Bewerbungen Schwerbehinderter ausdrücklich erwünscht. Bewerbungen bitte bis zum **30.09.2013** über das Online-Formular unter *https://s-lotus.gwdg.de/gwdgdb/agh/bewerbungen_ag_h_2013_08.nsf/bewerbung*

Stellenangebot

Fragen zur ausgeschriebenen Stelle beantworten Ihnen:

Herr Thomas Körmer
Tel.: 0551 201-1555
E-Mail: thomas.koermer@gwdg.de oder

Herr Dr. Konrad Heuer
Tel.: 0551 201-1540
E-Mail: konrad.heuer@gwdg.de

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

INFORMATIONEN:
support@gwdg.de
0551 201-1523

September bis Dezember 2013

Kurse

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
GRUNDLAGEN DER BILDBEARBEITUNG MIT PHOTOSHOP	Töpfer	04.09. – 05.09.2013 9:30 – 16:00 Uhr	28.08.2013	8
INDESIGN – GRUNDLAGEN	Töpfer	10.09. – 11.09.2013 9:30 – 16:00 Uhr	03.09.2013	8
GRUNDKURS UNIX/LINUX MIT ÜBUNGEN	Hattenbach	17.09. – 19.09.2013 9:15 – 12:00 und 13:30 – 16:00 Uhr	10.09.2013	12
USING THE GWDG SCIENTIFIC COMPUTE CLUSTER	Dr. Boehme, Ehlers	23.09.2013 9:00 – 12:00 Uhr	16.09.2013	2
PARALLELRECHNERPROGRAMMIERUNG MIT MPI	Dr. Haan	23.09.2013 14:00 – 17:00 Uhr 24.09. – 25.09.2013 9:15 – 17:00 Uhr	16.09.2013	10
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	26.09.2013 9:15 – 12:00 und 13:00 – 16:00 Uhr	19.09.2013	4
PHOTOSHOP FÜR FORTGESCHRITTENE	Töpfer	01.10. – 02.10.2013 9:30 – 16:00 Uhr	24.09.2013	8
CLIENT-MANAGEMENT MIT BARAMUNDI	Becker, Körmer, Quentin, Rosenfeld	17.10.2013 9:00 – 12:30 und 13:30 – 15:30 Uhr	10.10.2013	4
INDESIGN – AUFBAUKURS	Töpfer	22.10. – 23.10.2013 9:30 – 16:00 Uhr	15.10.2013	8
DIE SHAREPOINT-UMGEBUNG DER GWDG	Buck	29.10.2013 9:00 – 12:30 und 13:30 – 15:30 Uhr	22.10.2013	4
UNIX FÜR FORTGESCHRITTENE	Dr. Sippel	04.11. – 06.11.2013 9:15 – 12:00 und 13:15 – 15:30 Uhr	28.10.2013	12

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	13.11. – 14.11.2013 9:00 – 12:00 und 13:00 – 15:30 Uhr	06.11.2013	8
EINFÜHRUNG IN DAS IP-ADRESSMANAGEMENTSYSTEM DER GWDG FÜR NETZWERKBEAUFTRAGTE	Dr. Beck	28.11.2013 10:00 – 12:00 Uhr	21.11.2013	2
UNIX/LINUX-ARBEITSPLATZRECHNER – INSTALLATION UND ADMINISTRATION	Gerdes, Dr. Heuer, Körmer, Dr. Sippel	02.12. – 03.12.2013 9:15 – 12:00 und 13:30 – 16:00 Uhr	25.11.2013	8
UNIX/LINUX-SERVER – GRUNDLAGEN DER ADMINISTRATION	Gerdes, Dr. Heuer, Körmer, Dr. Sippel	04.12. – 05.12.2013 9:15 – 12:00 und 13:30 – 16:00 Uhr	27.11.2013	8
UNIX/LINUX – SYSTEMSICHERHEIT FÜR ADMINISTRATOREN	Gerdes, Dr. Heuer, Körmer, Dr. Sippel	06.12.2013 9:15 – 12:00 und 13:30 – 15:00 Uhr	29.11.2013	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	11.12. – 12.12.2013 9:00 – 12:00 und 13:00 – 15:30 Uhr	04.12.2013	8

Teilnehmerkreis
Das Kursangebot der GWDG richtet sich an alle Mitarbeiterinnen und Mitarbeiter aus den Instituten der Universität Göttingen und der Max-Planck-Gesellschaft sowie aus einigen anderen wissenschaftlichen Einrichtungen.

Anmeldung
Anmeldungen können schriftlich per Brief oder per Fax unter der Nummer 0551 201-2150 an die GWDG, Postfach 2841, 37018 Göttingen oder per E-Mail an die Adresse *support@ gwdg.de* erfolgen. Für die schriftliche Anmeldung steht unter *http://www.gwdg.de/antragsformulare* ein Formular zur Verfügung. Telefonische Anmeldungen können leider nicht angenommen werden.

Kosten bzw. Gebühren
Unsere Kurse werden wie die meisten anderen Leistungen der GWDG in Arbeitseinheiten (AE) vom jeweiligen Institutskontingent abgerechnet. Für die Institute der Universität Göttingen und der Max-Planck-Gesellschaft erfolgt keine Abrechnung in EUR.

Absage
Sie können bis zu acht Tagen vor Kursbeginn per E-Mail an *support@gwdg.de* oder telefonisch unter 0551 201-1523 absagen. Bei späteren Absagen werden allerdings die für die Kurse berechneten AE vom jeweiligen Institutskontingent abgebucht.

Kursorte
Alle Kurse finden im Kursraum oder Vortragsraum der GWDG statt. Die Wegbeschreibung zur GWDG sowie der Lageplan sind unter *http://www.gwdg.de/lageplan* zu finden.

Kurstermine
Die genauen Kurstermine und -zeiten sowie aktuelle kurzfristige Informationen zu den Kursen, insbesondere zu freien Plätzen, sind unter *http://www.gwdg.de/kurse* zu finden.


GWDG
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen