

GWDG NACHRICHTEN 08-09|17

Exchange 2016

Windows Server Update
Service

TSM-Skriptsammlung

PID/DOI im Batch-Upload
von BExIS++

International Summer
School on Data Science

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG





GWDG NACHRICHTEN

08-09|17 Inhalt

4 Upgrade des E-Mail-Systems auf
Exchange 2016 6 Kurz & knapp 7 Windows
Server Update Service 9 TSM-Skriptsammlung
15 PID/DOI im Batch-Upload von BExIS++
20 First International Data Science Summer
School in Göttingen 22 Stellenangebot
23 Personalia 26 Kurse

Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
40. Jahrgang
Ausgabe 8-9/2017

Erscheinungsweise:
monatlich

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:

© cybrain - Fotolia.com (1)
© Mathias Rosenthal - Fotolia.com (7)
© momius - Fotolia.com (8)
© vege - Fotolia.com (13)
© fotogestoeber - Fotolia.com (14)
© Contrastwerkstatt - Fotolia.com (22)
© Sashkin - Fotolia.com (25)
© Jan Vetter (20, 21)
© MPLbpc-Medienservice (3, 23, 24)
© 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:

Franziska Schimek
E-Mail: franziska.schimek@gwdg.de

Druck:

Kreationszeit GmbH, Rosdorf



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

Liebe Kunden und Freunde der GWWDG,

vom 10. bis 21. Juli 2017 fand in Göttingen die erste internationale Summer School zum Thema „Data Science“ statt. Diese wurde vom Institut für Informatik in Kooperation mit der Göttingen eResearch Alliance und Göttingen International unter Beteiligung der GWWDG organisiert. Die Veranstaltung fand sehr großen Zuspruch, was sich insbesondere an über 400 internationalen Bewerbungen für einen der 35 Teilnehmerplätze erkennen ließ. Das Thema trifft offensichtlich den Nerv der Zeit. Das umfangreiche Programm lieferte ein breites Spektrum an Vorträgen, das von den Methoden aus Mathematik, Statistik und Informatik bis zu Fragen zu Tools und zur Infrastruktur reichte. Ein besonderer Fokus lag auf der Darstellung von disziplinspezifischen Anwendungen.

Ich habe mich insbesondere gefreut, dass wir im Rahmen der HeKKSaGOn-Kooperation mehrere Vortragende aus Japan gewinnen konnten. Seit 2010 gibt es mit HeKKSaGOn ein japanisch-deutsches Universitätsnetzwerk, in dem die Universitäten Heidelberg, Kyoto, Karlsruhe/KIT, Tohoku/Sendai, Göttingen und Osaka zusammenarbeiten. Seit 2016 gibt es in diesem Verbund eine Arbeitsgruppe, die sich mit Data Science in Methodeneinsatz, Lehre und Infrastruktur beschäftigt.

Ich wünsche Ihnen viel Spaß mit dieser Ausgabe der GWWDG-Nachrichten.

Ramin Yahyapour

GWWDG – IT in der Wissenschaft

Upgrade des E-Mail-Systems auf Exchange 2016

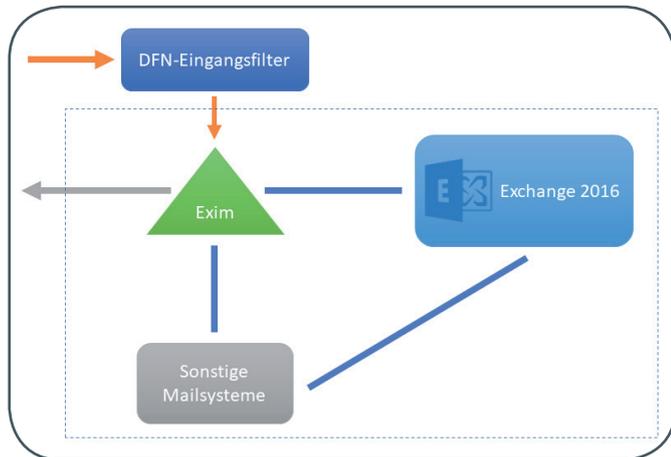
Text und Kontakt:

Jan-Nikolas Fahrenholz
jan-nikolas.fahrenholz@gwdg.de
0551 201-1536

Die GWDG betreibt bereits seit einigen Jahren den E-Mail und Groupware-Dienst Microsoft Exchange. Um die Stabilität und Ausfallsicherheit dieses aktuell für über 75.000 Nutzer bereitgestellten zentralen Basisdienstes weiter zu verbessern und den bisher schon großen Funktionsumfang noch zu erweitern, führt die GWDG zurzeit in mehreren Schritten ein umfangreiches Upgrade von Microsoft Exchange 2010 auf die Version Exchange 2016 durch.

AUFBAU DER NEUEN E-MAIL-UMGEBUNG MIT EXCHANGE 2016

Im Unterschied zu Exchange 2010 wurde die Anzahl der Serverrollen aufgrund der mittlerweile viel günstigeren CPU-Leistung auf zwei reduziert: „Postfach“ und „Edge Transport“. Letztere findet in der E-Mail-Umgebung aber keine Verwendung, da sie durch den UNIX-Mailer *mailer.gwdg.de* bzw. *mailer.mpg.de* (realisiert mit Exim) bereits vorhanden und fester Bestandteil der E-Mail-Umgebung ist (siehe Abb. 1).



1_Die E-Mail-Umgebung der GWDG

Die Postfach-Serverrolle in Exchange 2016 enthält alle Serverkomponenten der bisherigen Exchange-2010-Clientzugriffs-, -Postfachdatenbank- und -Hub-Transport-Serverrolle:

- Die **Clientzugriffsdienste** stellen Authentifizierung, eingeschränkte Umleitung und Proxydienste zur Verfügung und sie bieten alle üblichen Clientzugriffsprotokolle: HTTP, POP, IMAP und SMTP.
- Zu den **Postfachdatenbankdiensten** gehören alle herkömmlichen Serverkomponenten der Exchange-2010-Postfach-Serverrolle: Backend-Clientzugriffsprotokolle, Transportdienst, Postfachdatenbanken und Unified

Messaging. Der Postfachserver verarbeitet alle Vorgänge für die aktiven Postfächer auf diesem Server.

- Die **Hub-Transport-Serverrolle** ist für das Weiterleiten von E-Mails zuständig, nicht nur zwischen dem Internet und der Exchange-Infrastruktur, sondern auch zwischen den einzelnen Exchange-Servern. E-Mails werden immer mit der Hub-Transport-Serverrolle weitergeleitet, auch wenn sich Quell- und Zielpostfach auf demselben Server oder in derselben Postfachdatenbank befinden.

Dies bietet in Exchange 2016 gegenüber Exchange 2010 u. a. folgende Vorteile:

- **Flexibilität beim Versionsupgrade:** Es gibt keine starren Upgrade-Anforderungen mehr. Postfachserver können in Bezug auf andere Postfachserver unabhängig und in beliebiger Reihenfolge aktualisiert werden. Auf

E-Mail Service Upgrade to Exchange 2016

Communication via e-mail is one of the most important aspects of communication inside an organization and with external partners. To ensure a more stable mailing environment, the GWDG upgrades the e-mail service to Microsoft Exchange 2016.

Therefore, we set up a new server farm of thirteen servers divided on three sites GWDG, Fernmeldezentrale (FMZ) and Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) in order to provide a redundant environment.

The upgrade to Exchange 2016 includes various new and improved features like a new Outlook Web App (OWA) interface, and MAPI over HTTP as the standard protocol, which improves connection stability for Outlook clients.

At the moment the frontend is already migrated to Exchange 2016. The next step will be the migration of Exchange 2010 mailboxes to the new environment, which will be coordinated with the institutes' administrators.



2_Die Verteilung der Exchange-2016-Server auf die Standorte



3_Der Anmeldebildschirm der Exchange 2016 Outlook Web App

Wunsch können somit Institute mit besonderen Anforderungen gesondert von der restlichen Umgebung gewartet werden.

- **Sitzungsindifferenz:** Bei Exchange 2010 war für mehrere Protokolle Sitzungsaffinität zur Clientzugriffs-Serverrolle erforderlich. In Exchange 2016 befinden sich die Clientzugriffs- und Postfachkomponenten auf demselben Postfachserver. Dadurch kann für eingehende Clientverbindungen auf Postfachservern ein Loadbalancing mithilfe von Methoden wie „Geringste Anzahl von Verbindungen“ oder „Round-Robin“ erfolgen, wodurch es zu deutlich schnelleren Antwortzeiten kommt.

Um die Exchange-2016-Umgebung redundant zu gestalten, wurden insgesamt 13 Server vom Typ Dell PowerEdge R730xd beschafft und an den Standorten „Am Faßberg“ (GWDG), „Fernmeldezentrale“ (FMZ) und „Niedersächsische Staats- und Universitätsbibliothek Göttingen“ (SUB) aufgebaut (siehe Abb. 2). Ausgestattet sind diese mit jeweils 28 physikalischen CPU-Kernen, 128 GB RAM und NL-SAS-HDDs – die Umgebung liegt somit leicht über der von Microsoft empfohlenen Preferred Architecture für Exchange 2016. Durch die Verteilung auf drei Standorte gibt es von jeder Datenbank eine aktive Kopie an einem und zwei passive Kopien an den anderen Standorten. Außerdem wird ein Backup aller Datenbanken für den Katastrophenfall betrieben.

NEUE FUNKTIONEN BEI EXCHANGE 2016

Zu den wichtigsten Neuerungen für die Nutzer gehört die Verbesserung der **Outlook Web App (OWA)**, welche, ergänzend zu Desktop-Computern und Laptops, für Tablets und Smartphones optimiert wurde. Dazu gehören u. a.:

- **E-Mail-Verbesserungen** einschließlich einer neuen einzeiligen Ansicht des Posteingangs mit optimiertem Lesebereich, Archivierung, Emojis und der Möglichkeit, Postfachaktionen wie das Löschen einer E-Mail oder Verschieben einer E-Mail rückgängig zu machen.
- Der **Kalender** hat ein neues Aussehen erhalten und verfügt nun über neue Features, einschließlich E-Mail-Erinnerungen für Kalenderereignisse, die Möglichkeit zum Vorschlagen einer anderen Uhrzeit bei Besprechungseinladungen, eine verbesserte Suche sowie Geburtstagskalender.
- Vorschläge bei der **Suche** und Verfeinern der Suche zur

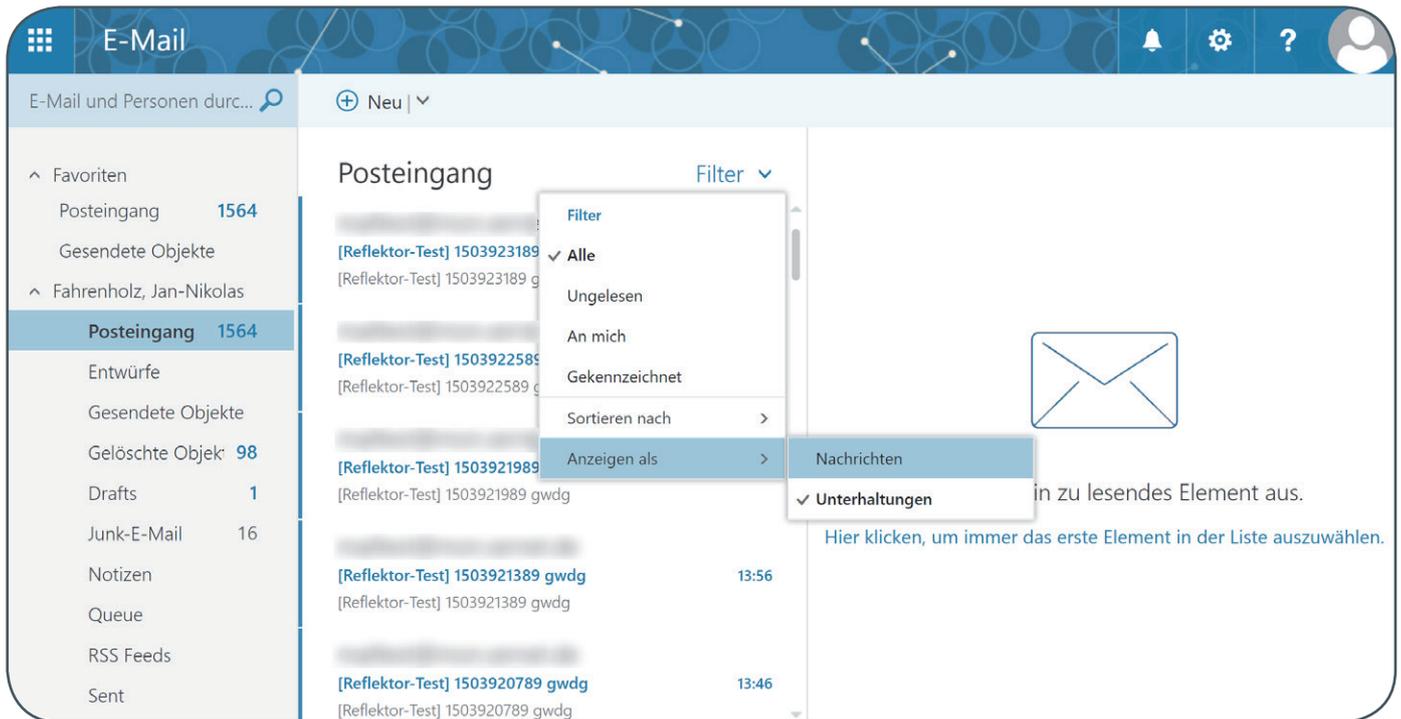
Verbesserung der Suchfunktionalität, sodass Benutzer die gewünschten Informationen noch schneller finden können. In Suchvorschlägen wird versucht vorwegzunehmen, was der Benutzer sucht, und entsprechende Ergebnisse werden geliefert. Sucheinschränkungen helfen dem Benutzer durch kontextbezogene Filter, die gesuchten Informationen leichter zu finden. Filter können Datumsbereiche, verwandte Absender usw. enthalten.

- **Pins und Kennzeichnungen** ermöglichen Benutzern, wichtige E-Mails dauerhaft oben im Posteingang anzuzeigen (Pins) und andere E-Mails zur Nachverfolgung (Kennzeichnungen) zu markieren. Die Pins sind jetzt auf Ordner bezogen und eignen sich daher optimal für Benutzer, die ihre E-Mails in Ordnern organisieren. Sie können markierte Elemente nun mithilfe von Posteingangsfiltern (siehe Abb. 4) oder mit dem neuen Aufgabenmodul, auf das sie über das App-Startfeld zugreifen können, schnell wiederfinden.
- Die **Linkvorschau** ermöglicht Benutzern, einen Link in E-Mails einzufügen. Outlook im Web generiert dann automatisch eine Vorschau, um Empfängern einen kurzen Einblick in den Inhalt des Linkziels zu geben. Dies funktioniert auch mit Videolinks.

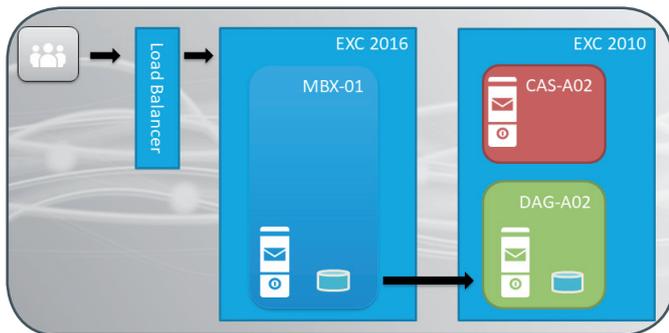
Zusätzlich zu den Änderungen an der OWA ist **MAPI über HTTP** nun das Standardprotokoll, das Outlook für die Kommunikation mit Exchange verwendet. MAPI über HTTP verbessert die Zuverlässigkeit und Stabilität der Outlook- und Exchange-Verbindungen durch Verschieben der Transportschicht auf das HTTP-Modell nach Industriestandard. Dies ermöglicht bessere Sichtbarkeit von Transportfehlern sowie verbesserte Wiederherstellbarkeit. Die weitere Funktionalität umfasst die Unterstützung für eine explizite Funktion zum Anhalten und Fortsetzen. Dies ermöglicht unterstützten Clients das Ändern des Netzwerks oder das Fortsetzen aus dem Ruhezustand heraus unter Beibehaltung desselben Serverkontextes.

Während MAPI über HTTP jetzt das Standardprotokoll für die Kommunikation zwischen Outlook und Exchange ist, verwenden Clients, die dieses Protokoll nicht unterstützen, Outlook Anywhere (RPC über HTTP).

Weiterhin ist es mit **Cumulative Update (CU) 6** endlich möglich, gelöschte Elemente nicht im Posteingang, sondern im ursprünglichen Ordner wiederherzustellen.



4_Der Posteingang mit den Filtereinstellungen. Hier kann nach ungelesenen E-Mails gefiltert und die Konversationsansicht (de-)aktiviert werden.



5_Clientenanfragen werden von den Exchange-2016-Servern entgegen-
genommen. Befindet sich das Postfach noch auf Exchange 2010,
kommuniziert Exchange 2016 direkt mit dem Exchange-2010-Postfach-
server.

MIGRATIONSABLAUF

Als erste Tests im Produktivsystem wurden bereits im Februar 2017 alle Mitarbeiter- und Funktionspostfächer der GWDG erfolgreich auf die Exchange-2016-Postfachdatenbanken migriert. Danach wurden, leider aufgrund unerwarteter Probleme, in

mehreren Anläufen und mit der Unterstützung von Microsoft letztendlich auch erfolgreich *email.gwdg.de* und somit die Clientzugriffsdienste auf die Exchange-2016-Server migriert.

Auch die Postfächer einiger Institute wurden nach vorheriger Abstimmung zu Testzwecken der Institute bereits auf die neuen Datenbanken kopiert.

Im nächsten Schritt werden, in Absprache mit allen betroffenen Instituten, alle Postfächer von Exchange 2010 auf 2016 migriert. Die Netzwerkbeauftragten der beteiligten Einrichtungen werden dazu frühzeitig gesondert informiert.

REFERENZEN

- Microsoft TechNet – Neuerungen in Exchange 2016: [https://technet.microsoft.com/de-de/library/jj150540\(v=exchg.160\).aspx](https://technet.microsoft.com/de-de/library/jj150540(v=exchg.160).aspx)
- The Microsoft Exchange Team Blog – The Exchange 2016 Preferred Architecture: <https://blogs.technet.microsoft.com/exchange/2015/10/12/the-exchange-2016-preferred-architecture/>

Kurz & knapp

Erratum zu den GWDG-Nachrichten 7/2017

Im Artikel „E-Mail-Signierdienst“ in den letzten GWDG-Nachrichten 7/2017 hat sich leider ein Fehler eingeschlichen.

Auf der Seite 12 wird mehrfach irrtümlicherweise das Feld „RCTP TO“ erwähnt. Korrekt muss das Feld aber „MAIL FROM“ lauten.

Otto



Windows Server Update Service

Text und Kontakt

Michael Reimann
michael.reimann@gwdg.de
0551 201-1826

Der nach wie vor wichtigste Schutz gegen Angriffe durch Schadsoftware jeglicher Art ist, das Betriebssystem und die Anwendungen stets auf dem aktuellen Stand zu halten. Im Falle von Microsoft Windows und den Office-Anwendungen geschieht das über den Update Service des Herstellers.

Bestes Beispiel für die Notwendigkeit der laufenden Aktualisierung von Betriebssystem und Anwendungen ist der sich im Mai 2017 verbreitende Verschlüsselungstrojaner *Wannacry*, der vorzugsweise Windows-7-Systeme befiel, die den zwei Monate zuvor freigegebenen Patch MS17-010 (Sicherheitsupdate für den SMB-Server) nicht eingefahren hatten. Dadurch entstand eine gefährliche Sicherheitslücke, die diesem Schädling den Zugang zum System ermöglichte, um so dann die Dateien zu verschlüsseln.

Dieser seinerzeit medienwirksame Vorfall bestätigte einmal mehr die dringende Empfehlung, die eigenen Windows-Systeme und Office-Anwendungen sicherheitstechnisch zeitnah auf dem aktuellen Stand zu halten. Das geschieht für gewöhnlich über den **Microsoft Update Service**, über den jeden zweiten Dienstag im Monat (dem sog. **Patchday**) die entsprechenden angefallenen Sicherheitskorrekturen verteilt werden. Nun kennt jeder das Phänomen, dass diese Updates immer zur Unzeit auf den Rechner wollen, obwohl man doch eigentlich gerade Dringenderes zu erledigen hat und dabei ungern gestört sein möchte. Also werden die Updates dann oft hinausgezögert, wenn nicht gar ganz ausgesetzt, was – wie in dem oben erwähnten Fall – gefährlich werden kann.

In den (zumeist über ein *Active Directory*) betreuten Institutsnetzen bemühen sich daher die Administratoren, die Verteilung der Patches so zu steuern, dass die Anwender auf ihren Systemen davon möglichst wenig mitbekommen. Das gelingt u. a. durch die Zwischenschaltung eines speziellen Servers: den **Windows Server Update Service (WSUS)**. Dieser lädt die Updates vom Microsoft Update Server herunter und bietet den Administratoren die Möglichkeit der gezielten Einflussnahme auf die

Weiterverteilung, so dass beispielsweise bestimmte Rechnergruppen eine besondere Patchzusammenstellung erhalten können oder sich der Zeitpunkt der Aktualisierung genauestens steuern lässt.

Solch einen WSUS-Server betreibt auch die GWDG für ihre Kunden: *sus.gwdg.de*. Befinden sich beispielsweise in Instituten an Standorten außerhalb von Göttingen sehr viele zu betreuende PCs, lassen sich dort weitere WSUS-Server installieren, die als Downstream-Server nun selbst wiederum die Updates von dem zentralen WSUS *sus.gwdg.de* gemäß den dort eingerichteten Zielgruppen laden und weiterverteilen und so zur Lastverteilung beitragen. (So wird auch das Active-Directory-Netzwerk der GWDG von einem eigenen WSUS gespeist, der ebenfalls am zentralen *sus.gwdg.de* hängt.) Gerade bei größeren Servicepacks oder den zweimal im Jahr anfallenden Funktionsupdates für Windows 10 müssen große Datenmengen bewegt werden, die ein einzelner Server kaum zu stemmen vermag. Nicht umsonst setzt Microsoft hier große Serverfarmen ein und verteilt die Funktionsupdates zudem auch noch zeitlich in Wellen.

Der zentrale WSUS der GWDG verteilt derart umfangreiche Updates daher nicht automatisch, sondern reicht sie erst in

Windows Server Update Service

The first protection against attacks by malicious software is to keep the operating system and applications up-to-date. In the case of Microsoft Windows and the Office applications, this will be done by the Microsoft update service.

Abgabe mit den Administratoren in den Instituten nach entsprechenden Vorbereitungen weiter.

Neben dem Aufbau von derartigen WSUS-Kaskaden aus Up- und Downstream-Servern könnten natürlich auch Einzelrechner von *sus.gwdg.de* ihre Updates beziehen, anstatt sie von Microsoft direkt zu laden. Offenbar wird das auch immer noch gerne genutzt, denn es laden ziemlich viele gerade auch heimische PCs ihre regelmäßigen Updates von *sus.gwdg.de* – vermutlich aus historischen Gründen. Das bringt jedoch keinerlei Vorteile, bei Windows 10 sogar eher Nachteile mit sich, denn es werden – wie ja bereits erwähnt – speicherintensive Updates gar nicht automatisch weitergereicht; ein heimisches Windows-10-System bekäme die wichtigen Funktionsupdates nicht angeboten und liefe somit nach 18 Monaten unweigerlich aus dem Support. Hier ist also der Microsoft Update Service unbedingt vorzuziehen.

Um seine Funktionalität zu verbessern, hat die GWDG *sus.gwdg.de* nun noch einmal verstärkt, seine Komponenten modernisiert und großzügiger ausgelegt, damit er die an ihm

hängenden Systeme besser und zügiger bedienen kann. Er unterliegt einer ständigen Beobachtung, so dass er unverzüglich aufgestockt wird, sollte er wieder einmal an seine Leistungsgrenzen kommen. Doch diese Maßnahmen helfen leider wenig bei den sich alle halbe Jahre regelmäßig wiederholenden Problemen mit den aktuellen Funktionsupdates für Windows 10. Vermutlich liegt die Ursache hier wohl eher in dem unzureichenden Zusammenspiel von WSUS und Windows 10. Um das **Windows-as-a-Service**-Konzept besser abbilden zu können, hat Microsoft inzwischen auch schon eine alternative Lösung am Start: **Windows Update for Business**. Die GWDG untersucht diesen neuen Service bereits darauf hin, inwieweit er die Verteilung innerhalb eines Active-Directory-Umfeldes für Windows-10-Systeme verbessern kann.

Weitergehende Informationen zum WSUS der GWDG finden sich direkt auf *sus.gwdg.de*, und natürlich können Anfragen und Anregungen auch immer gerne an die Service-Hotline (*support@gwdg.de*) gerichtet werden. ■



FTP-Server

Eine ergiebige Fundgrube!

Ihre Anforderung

Sie möchten auf das weltweite OpenSource-Softwareangebot zentral und schnell zugreifen. Sie benötigen Handbücher oder Programmbeschreibungen oder Listings aus Computerzeitschriften. Sie wollen Updates Ihrer Linux- oder FreeBSD-Installation schnell durchführen.

Unser Angebot

Die GWDG betreibt seit 1992 einen der weltweit bekanntesten FTP-Server, seit sieben Jahren mit leistungsfähigen Ressourcen für schnellen Service.

Ihre Vorteile

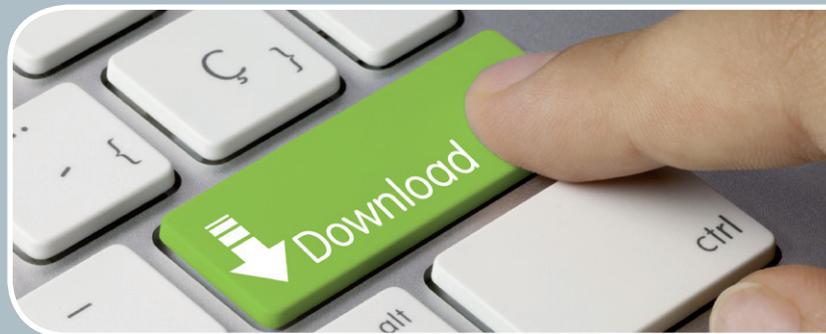
- > Großer Datenbestand (50 TByte), weltweit verfügbar
- > Besonders gute Anbindung im GÖNET

- > Aktuelle Software inkl. Updates der gebräuchlichsten Linux-Distributionen
- > Unter pub befindet sich eine aktuell gehaltene locatedb für schnelles Durchsuchen des Bestandes.
- > Alle gängigen Protokolle (http, https, ftp und rsync) werden unterstützt.

Interessiert?

Wenn Sie unseren FTP-Server nutzen möchten, werfen Sie bitte einen Blick auf die u. g. Webseite. Jeder Nutzer kann den FTP-Dienst nutzen. Die Nutzer im GÖNET erreichen in der Regel durch die lokale Anbindung besseren Durchsatz als externe Nutzer.

>> www.gwdg.de/ftp-server



TSM-Skriptsammlung

Text und Kontakt:

Björn Nachtwey
bjoern.nachtwey@gwdg.de
0551 201-2181

Bei der GWDG sind im Rahmen der ISP-Backup-Administration zahlreiche Skripte zur Vereinfachung der Administration entstanden. Diese werden sukzessive auf dem GWDG-GitLab veröffentlicht, um das Wissen teilen zu können. Bestandteil dieser Skriptsammlung ist auch die Neuimplementierung des Parallel-Backup-Tools „dsmci“, über das in den GWDG-Nachrichten 11/2016 und beim 33. DV-Treffen der Max-Planck-Institute 2016 berichtet wurde.

Gerade in größeren Umgebungen mit mehreren ähnlichen oder gleichartigen Systemen neigt der Administrator dazu, sich die tägliche Arbeit mit Skripten zu vereinfachen. Dies ist bei der TSM- bzw. ISP-Umgebung der GWDG nicht anders. Zwar gibt es zahlreiche, teilweise kommerzielle Hilfsprogramme, jedoch sind diese meist nicht ganz das, was ein Administrator sich wünscht.

ERGÄNZENDE MONITORING-SKRIPTE

Die GWDG setzt zwar den TSMManager (<http://tsmmanager.com>) außer zur Administration auch für das Monitoring der TSM-Umgebung ein, jedoch sind gerade die Übersichten zu allen Servern mit Informationen überladen und damit wenig hilfreich (siehe Abb. 1).

Wünschenswerter wären also Zusammenstellungen von einzelnen Informationen aller Server zur Kontrolle „auf einen Blick“. Hierzu wurden bei der GWDG einfache Skripte erstellt, die dies tun und das Ergebnis als einfache Webseite im Intranet bereitstellen (siehe Abb. 2).

Im Laufe der Zeit haben sich so einige Skripte angesammelt (siehe Abb. 3), die nun auch sukzessive im GitLab (<https://gitlab.gwdg.de>) online gestellt werden sollen.

SKRIPTE ALS ERSATZ FÜR ISP-FUNKTIONEN

Gerade in TSM-Umgebungen mit mehreren Servern kommt man relativ schnell an den Punkt, „Housekeeping“-Aufgaben über alle Server laufen zu lassen. Mit dem Command Routing und der TSM-eigenen Skriptsprache bietet IBM hierfür auch Hilfsmittel an, jedoch lässt die Steuerung, also die Ausführung in Abhängigkeit von Ereignissen oder Abfrageergebnissen, sehr zu wünschen übrig. Außerdem laufen verschiedene Prozesse jeweils nur lokal und vollkommen losgelöst von den übrigen Serverinstanzen. Dieser letzte Punkt kann zu Problemen führen, wenn gemeinsame Ressourcen genutzt oder die Nutzung optimiert werden soll. Mit den TSM-eigenen Bordmitteln ist es häufig nur möglich, bestimmte Prozesse zu unterschiedlichen Zeiten laufen bzw. starten zu lassen und dann zu hoffen, dass es zu keiner Konkurrenz kommt.

Zwei Beispiele, die besonders gravierend sein können, nutzen die gemeinsamen Laufwerke:

- Backup der internen Datenbank auf Band
- Nachverdichten der nur noch teilweise gefüllten Bänder (Reclamation-Prozess)

In beiden Fällen geht der jeweilige TSM-Server davon aus, ausreichend freie Laufwerke zu finden, um diese Aufgaben bewältigen zu können.

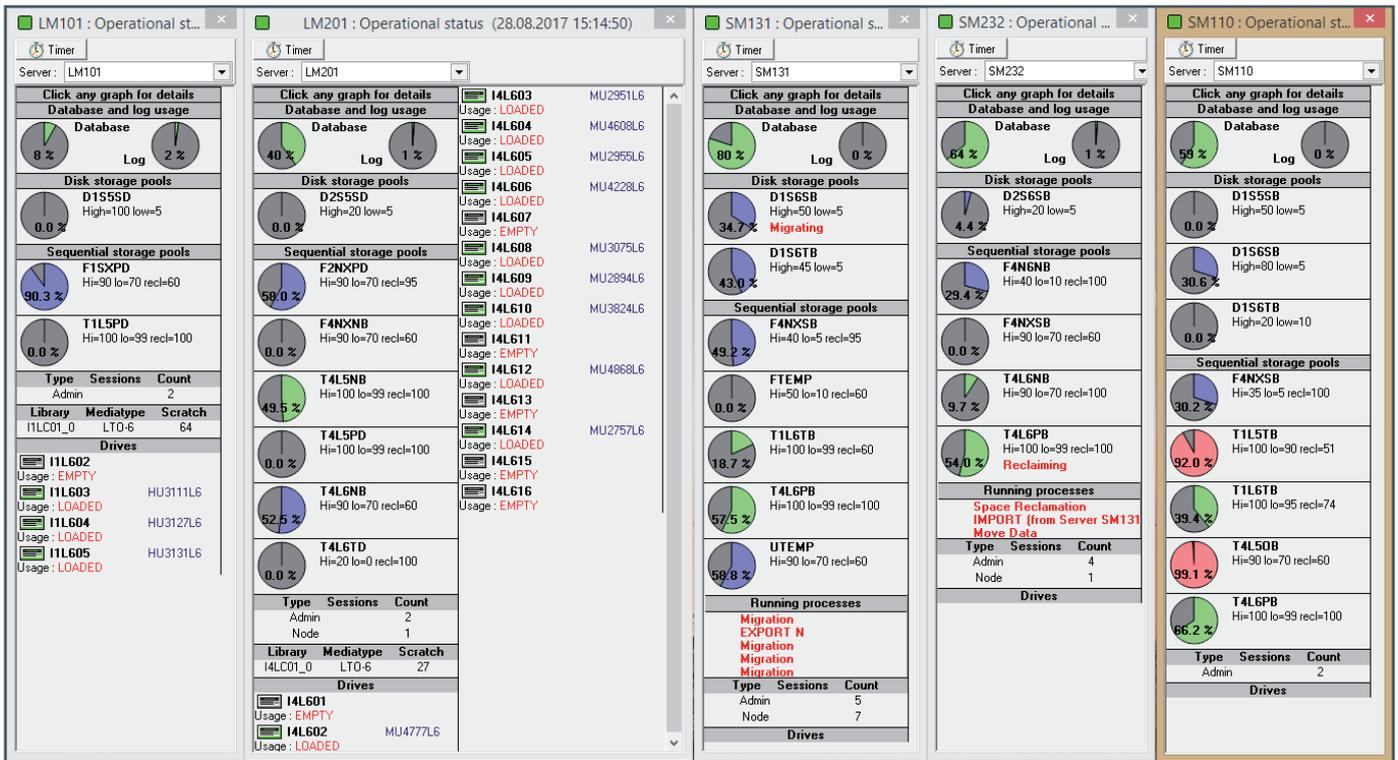
Im Bandroboter der GWDG ist dies aber nicht der Fall: Aktuell nutzen 14 Serverinstanzen lediglich 16 Laufwerke, was auch ziemlich gut funktioniert. Lediglich die Reclamation-Prozesse führten in der Vergangenheit immer wieder zu Problemen, so dass hierfür ein eigenes Skript geschrieben wurde.

In dem Skript finden dich folgende Überlegungen wieder:

- Das Zusammenkopieren soll nur laufen, wenn mindestens drei Laufwerke frei sind – so bleibt mindestens eines für andere Aufgaben verfügbar.
- Es sollen neben vollen Bändern (*STATUS='FULL'*) auch Bänder berücksichtigt werden, die zwar *STATUS='FILLING'* sind, aber eine gewisse Zeit (im Beispiel der GWDG: 15 Tage) nicht mehr genutzt wurden. So können z. B. durch das Neuorganisieren von Collocation Groups und das Löschen von Knoten pro Gruppe auch mehrere angefangene Bänder zugeordnet sein, die sich im normalen Betrieb aufgrund geringer oder abnehmenden Datenmengen nicht mehr füllen würden.
- Über die (einheitliche) Device-Klasse(n) lassen sich Tapevolumes identifizieren, so dass die anschließenden

TSM Scripts

IT administrators like to automate tasks by writing scripts. They often have their own opinion on how to do the monitoring of “their” servers, therefore they also develop scripts. At the GWDG it’s just the same: running the TSM environment many scripts have been written. Sometimes build-in functions do not work in the “expected” way, e.g. doing the tape reclamation by every TSM server indepently from the others needs lots of drives. A more sophisticated approach will first look for empty drives and then start as many reclamation jobs as drives are available. Unfortunately the TSM scripting language does not support such complexity, such a script must be run on an external host. At last, the “dsmci” script has been recorded in PERL running as well on Linux as on Windows. Actually some testing is done, discovering a problem with case sensitivity. To share this knowledge we start to put this scripts into the GWDG GitLab, starting with some monitoring scripts.



1_Statusinformationen desTSMManagers (hier nur ein Auszug: 5 von 15 Instanzen)

ISP Server State -- Database: Mon Aug 28 15:05:01 2017 CEST

Servername	current size (GB)	utilized (GB)	free (%)	Date of last ReOrg	Date of last Backup
tsm110	984.29	585.55	40.5	2017-08-04	2017-08-28
tsm112	499.75	320.84	35.8	2017-08-28	2017-08-28
tsm113	999.74	641.49	35.8	2017-08-23	2017-08-28
tsm130	492.15	334.16	32.1	2017-08-27	2017-08-28
tsm131	1574.90	1262.00	19.9	2017-06-21	2017-08-28
tsm180	999.99	511.59	48.8	2017-08-27	2017-08-28
tsm182	999.50	484.04	51.6	2017-01-19	2017-08-27
tsm230	492.15	241.62	50.9	2017-08-28	2017-08-28
tsm231	1476.44	987.66	33.1	2016-08-19	2017-08-27
tsm232	999.50	640.46	35.9	2016-09-17	2017-08-27
tsm233	492.15	382.44	22.3	2017-08-28	2017-08-27
tsm234	499.75	83.59	83.3	2017-08-28	2017-08-27
lm101	39.36	3.31	91.6	2017-08-26	2017-08-27
lm201	39.36	15.84	59.8	2017-08-27	2017-08-27

2_Status der Datenbanken derTSM-Server

- Befehle nicht auf DISK oder FILE arbeiten.
- Statt der Reclamation findet ein `move data <volume>` statt. Daher müssen die ggfs. unterschiedlichen Namen der Storagepools nicht berücksichtigt werden, allerdings können „Reclamation Storagepools“ so nur durch die dedizierte Angabe genutzt werden und erfordern eine entsprechende individuelle Anpassung; das Verschieben der Daten innerhalb einer Storagepools ist jedoch ohne Kenntnis des Namens möglich.
- Die Angabe eines Grenzwertes „Reclamation Threshold“ auf den einzelnen Storagepools der Server macht wenig Sinn, da diese sonst parallel eine echte Reclamation machen.
- Der Grenzwert für das Verschieben von Daten ist also im Skript anzugeben. Mit Blick auf die „Filling Tapes“ ist die tatsächliche Datenmenge (Kapazität x Füllgrad) ggfs. sinnvoller als nur der Füllgrad.

- Die Ergebnisse sind – mindestens pro Server – nach der Größe zu sortieren, so dass Daten von nur wenig gefüllten Bändern zuerst verschoben werden. Es ist davon auszugehen, dass bei relativ vollen Bändern noch weitere Daten im Rahmen des „Expiration“-Prozesses herausfallen.
- Für das Verschieben der Daten ist eine Administratorerkennung mit Passwort im Skript zu hinterlegen. Diese sollte daher nicht die „SYSTEM“-Klasse haben, sondern lediglich „STORAGE“, vorzugsweise beschränkt auf die Tapeools.

Der Ablauf des Skriptes sieht dann wie folgt aus:

1. Abfrage der „freien“ Laufwerke auf dem Libman:

```
SELECT COUNT(*) FROM drives WHERE -
drive_state='EMPTY'
```

TSM-Status @ GWDG

Allgemeine Legende:

SCHWARZER TEXT : keine besondere Bedeutung
 GRÜNER TEXT : Alles OK
 ORANGER TEXT : Wert im Auge behalten!
 ROTER TEXT : TSM-Admins fragen, ob es bekannt ist

Drives and Paths

ISP Library Manager D R I V E S and P A T H S State: Mon Aug 28 14:14:34 2017 CEST

Server	# Drives online	# Drives offline	# Drives n/n	# Drives online [%]	# Paths online	# Paths offline	# Paths online [%]
lm101	4	0	0	100.00	53	0	100.00
lm201	16	0	0	100.00	215	0	100.00

[Host-Status::Softwarestatus \(new Servers only\)](#)

[Server-Status::Database](#)

[Server-Status::Storagepools \(new Tab\)](#)

Volumes

- [Volumes readonly, unav. with errors -- Overview](#)
- [Volumes readonly with usage -- List and MOVE DATA commands](#)
- [f6k-S1: Volumes filling with usage -- List and MOVE DATA commands](#)

3_Übersichtsseite zu den Skripten

- Falls weniger als drei Laufwerke *EMPTY* sind: Abbrechen, sonst Anzahl der Prozesse bestimmen: Anzahl *N* der MOVE-Prozesse entspricht dem abgerundeten Ergebnis der durch 2 geteilten Anzahl freier Laufwerke.
- Schleife über alle ISP-Server, dort Abfrage der zu verschiebenden Volumes:

```
SELECT volume_name, -
CAST(FLOAT((est_capacity_mb/1024) * pct_utilized / 100) -
AS DEC(8,2)) AS movable_gb -
FROM volumes WHERE -
(devclass_name='I4L6' AND -
(DAYS(current_date) - DAYS(last_write_date))>$dayslimit) -
ANS movable_gb < $sizelimit) -
ORDER BY movable_gb
```

Die Parameter *\$dayslimit* und *\$sizelimit* sollten als Variable im Kopf des Skriptes definiert sein, um einfach geändert werden zu können.

- Start von *N* MOVE-Prozessen

Ein vollständiges Skript wird nach der notwendigen Überarbeitung im GitLab online gestellt.

NEUIMPLEMENTIERUNG „DSMCIS.PL“

Im Rahmen des Backup-Workshops im März diesen Jahres gaben die Kollegen von der MPCDF dankenswerterweise den sowohl naheliegenden wie letztendlich entscheidenden Hinweis, wie das Skript zur Parallelisierung des Backups sinnvoll und einfach auf Windows-Betriebssystemen eingesetzt werden kann: Statt einer Implementierung in der PowerShell möge doch einfach mal PERL ausprobiert werden.

Diesem Hinweis folgend steht man zwar vor dem „Problem“,

dass PERL nicht Bestandteil von Windows ist (ggfs. bringt hier das Linux-Subsystem für Windows eine einfache und praktikable Lösung) und man sich für eine der zahlreichen PERL-Varianten entscheiden muss. Dafür löst PERL aber zahlreiche Probleme, die in der ursprünglichen Fassung auf Basis der BASH nur mit einigem Aufwand zu lösen waren. Insbesondere das Starten von Sub-Threads und das Warten auf das Beenden derselben sowie das Kontrollieren, dass nur eine begrenzte Anzahl Threads tatsächlich parallel läuft, ist in PERL sehr einfach (siehe Exkurs in diesem Text).

Das Umcodieren des BASH-Skriptes nach PERL dauerte nur wenige Tage. Nach dem Lösen einiger kleinerer Probleme (in der Regel durch unterschiedliche Codierungen begründet) stand eine PERL-Version von *dsmcis* sowohl für Windows als auch für Linux zum Testen bereit. Mit wenig zusätzlichem Aufwand ließ sich auch noch ein rudimentäres Profiling der Läufe ergänzen: Im Rahmen der Auswertung wird eine Datei *dsmcis.prof* erzeugt, die für jedes Unterverzeichnis die Laufzeit in Sekunden angibt. Auf Basis dieser Laufzeit werden im nächsten Lauf die Verzeichnisse mit der längsten Laufzeit beginnend gestartet. Neue Verzeichnisse werden mit einer Laufzeit von 1 Mrd. Sekunden (= ~ 31 Jahre) am Anfang eingereicht, die „Kurzläufer“ machen den Abschluss.

Der Quelltext steht derzeit noch unter dem alten URL <https://gitlab.gwdg.de/bnacthw/dsmci/blob/master/perl/dsmcis.pl> im GitLab, wird aber zeitnah in die Sammlung <https://gitlab.gwdg.de/bnacthw/TSM-Scripts> eingereicht.

Für die nähere Zukunft ist geplant, die Variante „paralleles Durchsuchen aller Verzeichnisse“ wieder aufzunehmen. Hierfür muss aber noch eine Routine zum Löschen der Backupdaten von bereits gelöschten Verzeichnissen aufgenommen werden: Der Schalter *-subdir=no* verhindert, dass Unterverzeichnisse durchsucht werden; also werden auf dem Knoten gelöschte Verzeichnisse auch nur selbst gelöscht, nicht aber die darunter liegenden weiteren Verzeichnisse.

Das neue *dsmcis.pl*-Skript wird derzeit getestet, für die Sicherungen der StorNext-Filesysteme sind die Ergebnisse sehr gut. (Es verhält sich wie das alte BASH-Skript, ist durch das Profiling aber schneller). Auf den Windows-StorNext-Servern wird es in Verbindung mit „StrawberryPERL“ [<http://strawberryperl.com>; genutzte Version 5.20. (64Bit)] getestet. Hierbei kommt es vereinzelt zu Abbrüchen des Clients. Der Fehler wird derzeit im IBM Development analysiert. Vermutlich liegt er in der Case-Sensitivität begründet. Die Option *CASESENSITIVEAware YES* scheint Abhilfe zu schaffen.

Alle Interessierten sind herzlich eingeladen, auch diese Skripte zu nutzen, und wir freuen uns natürlich über jede Rückmeldung.

EXKURS: VERGLEICH PARALLELE THREADS IN SHELLS UND IN PERL

In „klassischen“ Shell-Umgebungen wird ein Subthread durch den Aufruf eines Kommandos gestartet und das Skript wartet dieses ab, also beispielsweise in der BASH:

```
#..
# starte Applikation X
aplx
```

Ergänzt man das *&*, läuft das Kommando im Hintergrund und blockiert das eigentliche Skript nicht:

```
#..
# starte Applikation X im Hintergrund
aplx &
```

Auch PERL sieht den Start von Subkommandos vor und bietet sogar zwei Varianten:

- „Wartend“

```
#..
<Rückgabewert> = system (<Kommando ggf. mit Optionen>);
```

- „Nicht-wartend“

```
#..
<Rückgabewert> = exec (<Kommando ggf. mit Optionen>);
```

Beiden nicht-wartenden Ansätzen (BASH und PERL) ist gemeinsam, dass das startende Skript die Kontrolle über den Subthread an das Betriebssystem abgegeben hat und somit auch keine Rückmeldung beim Beenden des Subthreads bekommt. Soll nun die Anzahl der parallelen Subthreads begrenzt werden, müssen diese vergleichsweise aufwändig über das Betriebssystem ermittelt und gezählt werden. Die hierzu notwendigen Ansätze sind für Linux/UNIX und Windows vollkommen unterschiedlich.

Threads zählen mit der BASH

Der Aufwand wird am Beispiel der BASH unter Linux aufgezeigt:

```
nthreads=$(ps axo ppid,cmd | grep $ppid | grep "$server | grep -v grep | wc -l)
```

Zunächst wird über das *ps*-Kommando eine Liste aller laufenden Prozesse abgefragt, über die *o ppid,cmd*-Option wird die Ausgabe auf die *ppid* (parent process id) und den Aufruf (*cmd*) reduziert, anschließend wird nach der ID des aufrufenden Skriptes gefiltert (*grep \$ppid*). Da bei einer kurzen *ppid* auch Prozesse geliefert werden, in deren *ppid* die gesuchte enthalten ist (z. B. ist „34“ ja auch in „234“ oder „110347“ zu finden), wird zusätzlich nach der korrekten Angabe der TSM-Serverinstanz gesucht (*grep „\$server*). Da auch die jeweiligen *grep*-Prozesse Treffer liefern, müssen diese herausgefiltert werden (*grep -v grep*). Das verbleibende Ergebnis wird gezählt (*wc -l*) und gibt somit die Anzahl der laufenden Subprozesse an. (Das aufrufende Skript hat die gesuchte *\$ppid* als *\$pid* (process id) und fällt daher raus).

Threads zählen mit der PowerShell

Für Windows könnte analog der BASH über die PowerShell gearbeitet werden. Entsprechende Ansätze wurden aber durch den Autor verworfen, nachdem er bereits erfolglos reichlich Zeit aufgewandt hatte – letztendlich ein Vielfaches im Vergleich zur PERL-Rekodierung. Daher nur eine kleine Idee hierzu: Das Zählen der Subthreads sollte über das *get-Process*-cmdlet möglich sein. Über entsprechende Optionen kann analog der BASH auf die Abhängigkeit vom Elternprozess gefiltert werden.

Threads in PERL

Grundsätzlich ließe sich in PERL eine Routine analog der BASH oder der PowerShell nachbilden, im Zweifel durch den Aufruf der entsprechenden System-Kommandos. Dies führt aber zu dem Aufwand, entweder zwei PERL-Skripte für Linux und Windows zu schreiben oder entsprechend zahlreiche *if*-Verzweigungen für die beiden Betriebssysteme vorzusehen. PERL bietet jedoch ein eigenes Thread-Modul und damit eine wesentlich elegantere Methode:

Über den *fork()*-Befehl erzeugt der PERL-Interpreter einen zweiten Thread, der an just dieser Stelle im Skript startet. Dieser Thread bearbeitet wie das originale Skript alle nachfolgenden Anweisungen. Es ist daher sinnvoll, die unterschiedlichen Aufgaben über eine *if*-Anweisung zu verzweigen. Hierfür bietet sich die Abfrage des Rückgabewertes der *fork()*-Routine an: Ist der Wert nicht definiert, konnte kein Thread erzeugt werden, ist der Wert „wahr“, gibt es einen neuen Thread – und es ist die Elternroutine, in der dieses *if* ausgeführt wurde. Im Kindthread ist der Wert ebenfalls definiert, aber „falsch“. Die Abfrage könnte also wie folgt aussehen:

```
my $cpid = fork();
if (! defined $cpid)
{
    # forking failed!
    exit 1; # Abbruch des Skriptes
}
if ($cpid)
{
    # parent process
    # ... was der Elternprozess auch immer tun soll
}
else
{
    # child process
    # ... was der Kindprozess auch immer tun soll
}
```

Auch das Einsammeln/Abwarten der gestarteten Threads gestaltet sich wesentlich einfacher: Über die *wait()*-Funktion wird auf das Beenden eines Kindthreads gewartet; alle können also mit einer einfachen Schleife abgewartet werden:

```
while (wait() != -1) ;
```

AUSBLICK

Wie angesprochen, werden im Laufe der Zeit immer mehr Skripte überarbeitet und online im GitLab verfügbar sein. Dies umfasst auch solche Skripte, mit denen Aufgaben erledigt werden,

die über die ISP-eigene Skriptsprache nicht möglich sind, beispielsweise das automatische Löschen von Admin-Accounts (wenn man die Option *-USERID=NONE* beim Anlegen eines Knotens vergessen hat) oder das automatische Zuordnen von Knoten zu Collocation Groups. Im Rahmen der „Kurz und knapp“-Mitteilungen werden wir hierüber in kommenden Ausgaben der GWGD-Nachrichten berichten.

Das parallele Backup über das *dsmci.pl*-Skript wird weiter optimiert und die angesprochenen Funktionen ergänzt werden. Vielleicht kann schon zum diesjährigen 34. DV-Treffen der Max-Planck-Institute, das vom 17. - 19. Oktober in Göttingen stattfinden wird, von wesentlichen Fortschritten berichtet werden. ■



Bewerbungs-/Registrierungssysteme

Wir unterstützen Sie in Ihrer Organisationsarbeit!

Ihre Anforderung

Sie möchten ein Bewerbungs- oder allgemeines Registrierungsverfahren durchführen, z. B. für offene Stellen oder Tagungsmanagement. Bewerber sollen sich online bewerben und automatisiert per E-Mail benachrichtigt werden können. Gutachter sollen über das WWW auf die Bewerbungen bzw. Registrierungen zugreifen und Bewertungen online einstellen können.

Unser Angebot

Wir erstellen Ihnen nach Ihren Wünschen eine Lotus-Notes-Datenbank, die allen Kandidaten oder Registranten über einen Webbrowser offen steht. Die eingereichten Dokumente können aber nur von ausgewählten Gutachtern über das WWW und von speziellen Bearbeitern Ihres Instituts eingesehen, bearbeitet oder bewertet werden. Die Eingänge werden nach Ihren Kriterien sortiert und dargestellt. Weitere Workflows sind individuell gestaltbar.

Ihre Vorteile

- > Leistungsfähiges ausfallsicheres System zum Aufnehmen von Bewerbungen oder Registrierungen über das WWW
- > Datenschutzgerechte Speicherung und Verarbeitung der Daten
- > Die Verteilung der Unterlagen auf Papier ist überflüssig, da der Zugriff der Gutachter oder sonstigen Bearbeiter über das WWW erfolgt.
- > Jeder Workflow ist an Ihre Situation anpassbar.
- > Kandidaten können automatisiert per E-Mail benachrichtigt werden (z. B. Absagen).

Interessiert?

Der Service wie auch die individuelle Beratung können über support@gwdg.de angefordert werden. Nähere Informationen zum Workflow Management mit der Lotus-Software von IBM sind auf der u. g. Webseite zu finden.

>> www.gwdg.de/lotus



Mailinglisten

MAILVERSAND LEICHT GEMACHT!

Ihre Anforderung

Sie möchten per E-Mail zu oder mit einer Gruppe ausgewählter Empfänger kommunizieren, auch außerhalb Ihres Instituts. Sie möchten selbstständig eine Mailingliste verwalten, z. B. Empfänger hinzufügen oder entfernen. Bei Bedarf sollen sich auch einzelne Personen in diese Mailingliste einschreiben dürfen.

Unser Angebot

Wir bieten Ihnen einen Listserver, der zuverlässig dafür sorgt, dass Ihre E-Mails an alle in die Mailingliste eingetragenen Mitglieder versendet werden. Die E-Mails werden automatisch archiviert. Das Archiv kann von allen Mitgliedern der Liste nach Schlagwörtern durchsucht werden. Die Anzahl Ihrer Mailinglisten ist unbegrenzt

Ihre Vorteile

- > Leistungsfähiges ausfallsicheres System zum Versenden von vielen E-Mails
- > Sie senden Ihre E-Mail lediglich an eine Mailinglisten-Adresse, die Verteilung an die Mitglieder der Mailingliste übernimmt der Listserver.

- > Listenmitglieder können an diese E-Mail-Adresse antworten. Eine Moderationsfunktionalität ist verfügbar, mit der Sie die Verteilung einer E-Mail genehmigen können.
- > Voller administrativer Zugriff auf die Einstellungen der Mailingliste und der Listenmitglieder
- > Obsolete E-Mail-Adressen werden vom System erkannt und automatisch entfernt.
- > Wenn Ihre E-Mail-Domäne bei uns gehostet wird, können Sie auch die Adresse der Mailingliste über diese Domäne einrichten lassen.

Interessiert?

Für die Einrichtung einer Mailingliste gibt es zwei Möglichkeiten: Zum einen als registrierter Benutzer der GWDG im Selfservice über das Kundenportal der GWDG und zum anderen, indem Sie bitte eine entsprechende E-Mail an support@gwdg.de senden, die die Wunsch-E-Mail-Adresse der Liste sowie die E-Mail-Adresse der Person, die die Liste bei Ihnen administrieren soll, enthalten sollte. Die administrativen Aufgaben sind leicht zu erlernen.

PID/DOI im Batch-Upload von BExIS++

Text und Kontakt:

Thorsten Hindermann
thorsten.hindermann@gwdg.de
0551 201-1837

In einem Artikel der GWDG-Nachrichten 8-9/2016 wurde der Batch-Upload für das DFG-Projekt „BExIS++“, an dem die GWDG beteiligt ist, bereits vorgestellt und die Zusammenarbeit einer weit verteilten Gruppe von Programmierern in dem Projekt erläutert. In diesem Artikel wird das Merkmal PID/DOI des BExIS++-Batch-Uploads näher beschrieben.

WAS BISHER GESCHAH

Wie gerade erwähnt, war das BExIS++-Projekt (<http://bexis2.uni-jena.de/>) mit der Vorstellung der Programmierung des Batch-Uploads sowie der Zusammenarbeit einer weit verteilten Gruppe von Programmierern schon in einer Ausgabe der GWDG-Nachrichten vertreten.

Weiterhin hat auch ein dreitägiger Besuch bei der BExIS++-Kerngruppe in Jena stattgefunden, bei dem viele technische Fragen geklärt werden konnten. Nebenbei gab es immer mal lose Videokonferenztermine und E-Mail-Kontakte, in deren Verlauf kleine Abstimmungen stattgefunden haben.

WAS IST EIN PID/DOI EIGENTLICH?

Ein persistenter Identifizierer (engl. Persistent IDentifier, kurz PID) oder digitaler Objektbezeichner (engl. Digital Object Identifier, kurz DOI) ist ein eindeutiger und dauerhafter digitaler Identifikator für physische, digitale oder abstrakte Objekte. Er wird vor allem für Online-Artikel wissenschaftlicher Fachzeitschriften (DOI) oder Publikationen von Forschungsartikeln (PID) verwendet. Das PID/DOI-System baut auf einem Handle-System auf und ist, vereinfacht ausgedrückt, mit ISBN und ISSN vergleichbar.

PIDs/DOIs haben den Zweck, mit ihnen dauerhaft auf digitale Objekte verweisen zu können. Dies wird erreicht, indem jedem Objekt ein einzigartiger und dauerhafter Name (Identifikator) zugewiesen wird. Anhand dessen kann aus einer zentralen Datenbank der URL abgefragt werden, unter dem das Objekt derzeit zugänglich ist. Es können auch mehrere URLs hinterlegt werden; es werden dann beim Zugriff auf das Objekt zunächst alle URLs aufgelistet.

PIDs/DOIs lösen also die Schwierigkeit von toten Links, indem Objekte nach der Änderung ihrer URL trotzdem noch unter ihrer PID/DOI zugänglich bleiben; es muss bloß der Eintrag in der zentralen Datenbank aktualisiert werden. In der Praxis erfolgt aber oftmals auch diese Aktualisierung nicht und so kann es gleichfalls zu toten PIDs/DOIs kommen. Ein aktuelles Beispiel für einen toten DOI ist 10.10.1038/nphys1170.

Die zentrale Datenbank, in der die Identifikatoren und URLs gespeichert werden, wird von der International DOI Foundation (<http://www.doi.org/>) betrieben und bei den PIDs ist das das ePIC Consortium (<http://www.pidconsortium.eu/>).

BATCH-UPLOAD BEI BEXIS++ - AKTUELLER STAND DER ENTWICKLUNG

Neben der Möglichkeit, mehrere Dateien von einem Client-Computer aus in BExIS++ hochzuladen oder zuvor schon hochgeladene, auf dem Server vorhandene Dateien zu verarbeiten, ist ein weiteres wichtiges Merkmal im Formular des Batch-Uploads vorhanden: die Angabe von PID/DOI-Referenzen zur weiteren Verarbeitung.

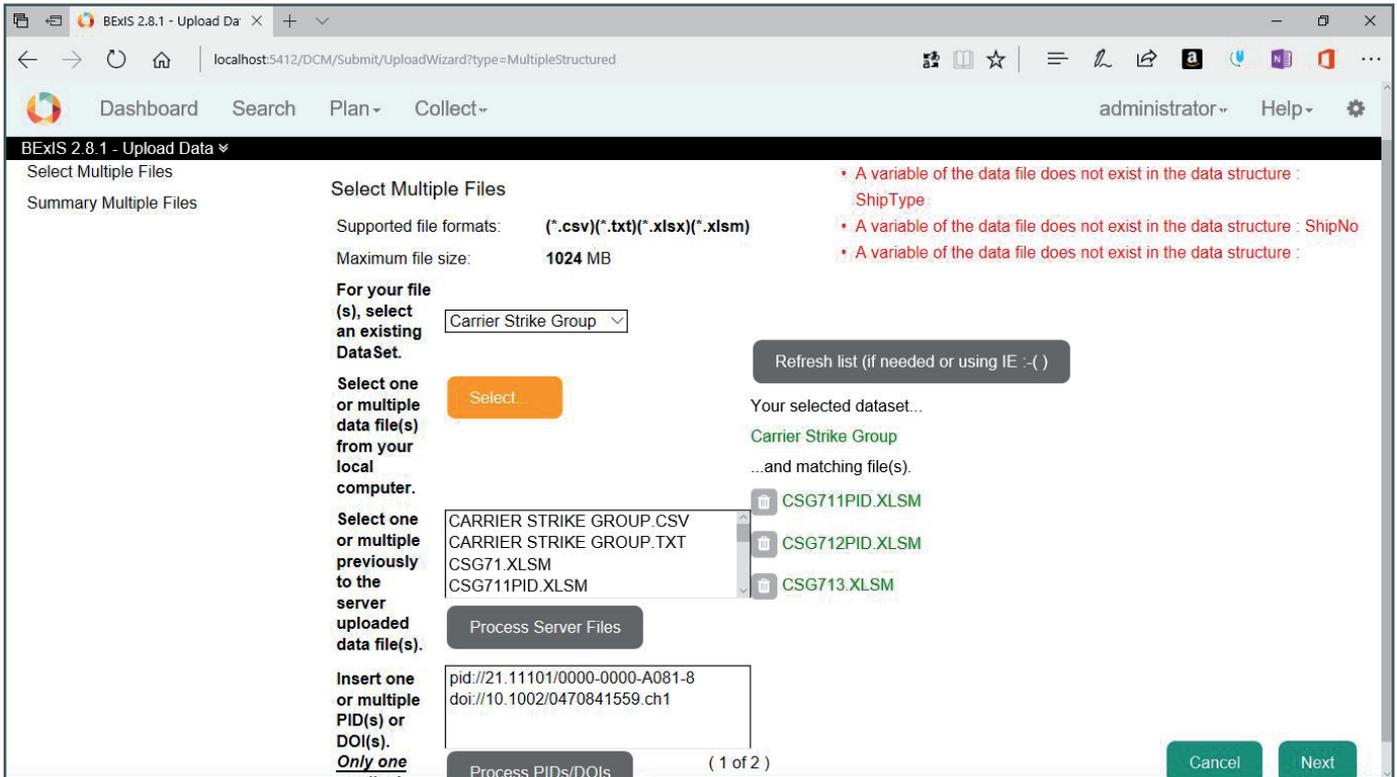
Weiterhin ist der gesamte Batch-Upload-Prozess nun so vereinfacht, dass nur noch ein Schritt für die Angabe und Verarbeitung der Dateien notwendig ist und im zweiten Schritt Informationen über die verarbeiteten Dateien ausgegeben werden. Das Hauptformular, Schritt 1, ist in Abb. 1 zu sehen.

Hier ein kurzer Überblick über das Formular und seine Funktionsweise: Das Layout ist dreispaltig. In der linken Spalte sind die zwei erwähnten Arbeitsschritte zu sehen. In der rechten Spalte gibt es oben einen Bereich für Meldungen, die während der Prüfung der Dateien auftreten können. Der untere Bereich enthält die Liste der ausgewählten Dateien, die die Prüfung erfolgreich bestanden haben.

Die Spalte in der Mitte enthält die entscheidenden Elemente. Oben die herunterklappbare Auswahlliste mit den zur Verfügung stehenden Datensätzen. Nach der Auswahl eines Datensatzes werden alle damit verknüpften Dateien in den drei folgenden Bereichen geprüft. Als nächstes folgt die Möglichkeit (Select-Button), Dateien vom Client-Computer aus hochzuladen. Darunter die Ansicht und Auswahlmöglichkeit der Dateien, die zuvor auf den Server hochgeladen worden sind. Ist ein Datensatz ausgewählt worden, werden in dieser Liste dann nur noch die Dateien angezeigt, die mit diesem Datensatz verknüpft sind. Danach folgt der

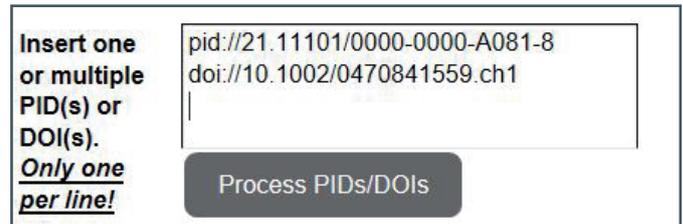
PID/DOI in the Batch Upload of BExIS++

In an article of the GWDG News 8-9/2016, the batch upload for the DFG project „BExIS ++“, in which the GWDG is involved, has already been presented and the cooperation of a widely distributed group of programmers in the project was explained. This article describes the feature PID/DOI of the BExIS++ batch upload.



1_Hauptformular im Batch-Upload-Prozess

Eingabebereich für die PID- oder DOI-Referenzen. Alle angegebenen Dateien werden gegen den ausgewählten Datensatz geprüft. Wird diese Prüfung erfolgreich bestanden, werden diese Dateien in die Liste im unteren Bereich der rechten Spalte eingefügt. Nach einem Klick auf „Next“ beginnt dann die eigentliche Verarbeitung der Dateien, d. h. die Speicherung der in den Dateien enthaltenen Datenzeilen in der BEXIS++-Datenbank.



2_PID/DOI-Eingabebereich

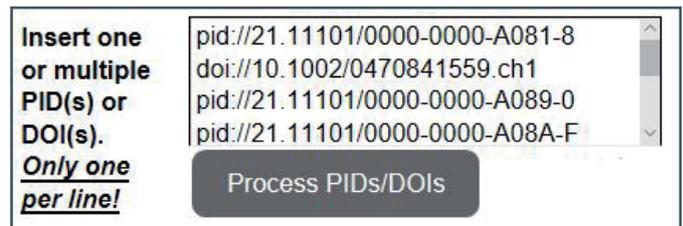
PID/DOI IM BATCH-UPLOAD VON BEXIS++ IM DETAIL

In diesem Abschnitt werden wir uns näher mit der Realisierung der Angabe und des Hochladens von PID/DOI-Referenzen in BEXIS++ befassen.

Technisch basiert die BEXIS++-Webanwendung auf ASP.NET MVC5 und wird mit C# programmiert. Im Folgenden werden im Text hier und da einige C#-Quellzeilen abgebildet sein, die wesentliche Teile der PID/DOI-Verarbeitung aufzeigen. Auch werden einige XML-Zeilen aus der Konfigurationsdatei der BEXIS++-Webanwendung (*web.config*) zu sehen sein.

Technisch ist der Teil der Webseite eine Mischung aus ASP.NET MVC5, JavaScript und der klassischen formularbasierten Webseiten-Programmierung. Konzentrieren wir uns nun auf die PID/DOI-Verarbeitung. Die Angabe von PIDs/DOIs findet im unteren Bereich der mittleren Spalte des Webformulars (siehe Abb. 1) statt. Dieser Bereich ist in Abb. 2 vergrößert dargestellt. In diesem Eingabebereich kann eine große Zahl von PID/DOI-Angaben eingefügt werden. Der Texteingabebereich erstellt automatisch einen Rollbalken an der rechten Seite, wenn mehr Angaben gemacht werden, als aktuell angezeigt werden können (siehe Abb. 3).

Um zu unterscheiden, welcher Auflösungs-Service von PID/DOI zu URL genutzt werden soll, wird dem PID/DOI, ähnlich einem URL wie z. B. *http://www.gwdg.de*, *pid://* bzw. *doi://* vorangestellt.



3_PID/DOI-Eingabebereich mit Rollbalken

Im korrespondierenden Abschnitt in der Konfigurationsdatei *web.config* ist ersichtlich, welcher Auflösungs-Service gefragt werden soll. Würde sich zukünftig der URL des Service ändern, kann ein Systemverwalter diesen Eintrag ändern, ohne dass die Programmierung der Webanwendung geändert werden müsste.

In der Quellzeilen-Auflistung 1 ist ersichtlich, dass die Werte (*value=*), der Schlüssel (*key=*), *HandleServiceURL* und/oder *DOI-ServiceURL* entsprechend angepasst werden können. Neben diesen beiden Schlüssel/Wert-Paaren gibt es noch andere Schlüssel/Wert-Paare, die für den Batch-Upload wichtig sind und von Systemverwaltern geändert werden können, ohne dass die Webanwendung geändert werden müsste.

Nach dem Klick auf „Process PIDs/DOIs“ passiert eine ganze Menge hinter den Kulissen. Als erstes schauen wir uns den Quellzeilen-Ausschnitt der Webseite an, der die Eingabeliste und Schaltfläche für die PIDs/DOIs repräsentiert (siehe Quellzeilen-Auflistung 2).

```

<!-- Begin block of Key/Value pairs for batch upload function programmed by GWDG -->
<!-- Supported file extensions for structured files-->
<add key="StructFExt" value=".csv,.txt,.xlsx,.xslm" />
<!-- Supported file extensions for unstructured files-->
<add key="UnStructFExt"
value=".avi,.bmp,.csv,.doc,.docx,.gif,.jpg,.mp3,.mp4,.pdf,.png,.tif,.txt,.xls, ,.xsd,.zip" />
<!-- Support order files value: true or false -->
<add key="OrderFiles" value="true" />
<!-- Order files extension -->
<add key="OrderFilesExt" value=".orf" />
<!-- Handle server URL -->
<add key="HandleServiceURL" value="http://hdl.handle.net/api/handles/" />
<!-- DOI server URL -->
<add key="DOIServiceURL" value="http://dx.doi.org/api/handles/" />
<!-- End block of Key/Value pairs for batch upload function programmed by GWDG -->

```

Quellzeilen-Auflistung 1: web.config

```

<tr>
<td><b>Insert one or multiple PID(s) or DOI(s).<br /><i><u>Only one per line!</u></i></b></td>
<td style="vertical-align: top">
    @using (Html.BeginForm("ProcessPiddois", "SubmitSelectMultipleFiles", new {area = "DCM"},
        FormMethod.Post, new {id = "processPiddoisForm", enctype = "multipart/form-data"}))
    {
        @(Html.TextAreaFor(m => m.SelectMultipleFilesFromPidsDois, new {title = "Enter one or
multiple PID(s)/DOI(s). ATTENTION: Please prefixes with doi:// or pid://!", @class = "bx-list-multi"}))
        <button type="submit" class="bx-button">Process PIDs/DOIs</button>
    }
</td>
</tr>

```

Quellzeilen-Auflistung 2: SelectMultipleFiles.cshtml

```
foreach (var pid in reqFrmSplit) { ... }
```

Quellzeilen-Auflistung 3: ProcessPiddois

```

// If there are blank lines in the text field break the actual iteration
if (string.IsNullOrEmpty(pid) || string.IsNullOrWhiteSpace(pid))
    break;

```

Quellzeilen-Auflistung 4: ProcessPiddois

Die Tabellenzeile (<tr>...</tr>) enthält zwei Spalten (<td>...</td>). Die linke Spalte, in der Auflistung die erste <td>-Zeile, enthält die kurze Erklärung, was hier erwartet wird. Die zweite Spalte enthält den Eingabebereich @(<Html.TextAreaFor(...) und die darunter befindliche Schaltfläche <button>...</button> zum Absenden (type=submit) sowie das das ganze Konstrukt einkleidende Formular @using(Html.BeginForm(...) { ... }. Die eigentliche Verarbeitungs-Routine, die aufgerufen wird, wenn die Schaltfläche geklickt wurde, ist angegeben mit ProcessPiddois in der Quellzeilen-Auflistung 2.

Wir verlassen nun die Ansicht (engl. View) und wenden uns der Verarbeitung in der Steuereinheit (engl. Controller) zu. Somit hätten wir auch zwei der Buchstaben der dreibuchstabigen Abkürzung (engl. Three Letter Acronym, oder ganz kurz TLA) MVC – steht für Model/View/Controller – geklärt.

In der Steuereinheit wird die Methode ProcessPiddois angesprungen und hier findet die weitere Verarbeitung der Angaben in dem Textbereich statt. Hier wird nun aber nicht die ganze Methode abgebildet, sondern nur relevante Teile der Quellzeilen, um bei

der großen Zahl von Quellzeilen nicht den Überblick zu verlieren.

Da mehrere PIDs/DOIs in dem Textbereich angegeben werden können, müssen all diese Textzeilen abgearbeitet werden. Nach etwas Vorbereitung wird der Durchlauf aller Textzeilen in der Programmiersprache C# mit einer „Für jedes Element“-Anweisung eingekleidet (siehe Quellzeilen-Auflistung 3).

Sollte sich eine Leerzeile eingeschlichen haben, wird diese nach Prüfung und Erkennung übersprungen (siehe Quellzeilen-Auflistung 4).

In den folgenden Quellzeilen (siehe Quellzeilen-Auflistung 5) wird anhand von pid:// oder doi:// entschieden, welcher Auflösungs-Service gefragt werden soll. Der entsprechende Auflösungs-Service wird nach Erkennung gezielt durch den Aufruf von UploadWizardHelper.GetWebConfKeyValue("HandleServiceURL") oder UploadWizardHelper.GetWebConfKeyValue("DOIServiceURL") aus der web.config abgefragt (siehe dazu die Quellzeilen-Auflistung 1).

Die eigentliche Auflösung des PID/DOI zu einer realen URL geschieht mit dem in Quellzeilen-Auflistung 6 dargestellten

```

var pidSplit = Regex.Split(pid, "://"); //split the actual pid into protocol and value
// Check the array-element 0 for the protocol part: PID or DOI
if (pidSplit[(int) PidProtVal.Protocol].ToLower() == _pidProtStrings[(int)
PidProt.Pid])
{
    uriPidsDoIs = UploadWizardHelper.GetWebConfKeyValue("HandleServiceURL") +
        pidSplit[(int) PidProtVal.Value];
}
else
{
    uriPidsDoIs = UploadWizardHelper.GetWebConfKeyValue("DOIServiceURL") +
        pidSplit[(int)PidProtVal.Value];
}

```

Quellzeilen-Auflistung 5: ProcessPiddois

```

var resolvedUri = await ResolveUriFromHandleAsync(uriPidsDoIs);

```

Quellzeilen-Auflistung 6: ProcessPiddois

```

var aBool = await GetUriAsync(resolvedUri, filePath, resolvedUri.Split('/').Last());

```

Quellzeilen-Auflistung 7: ProcessPiddois

Konstrukt.

Der so ermittelte URL enthält eine Excel-Datei mit entsprechenden Datenzeilen, die zur weiteren Verarbeitung auf den Server heruntergeladen werden muss. Diese Aufgabe erledigt der in Quellzeilen-Auflistung 7 dargestellte Aufruf.

Um den Ablauf noch einmal kurz zu veranschaulichen, was die abstrakt wirkenden C#-Quellzeilen bewirken: *pid://21.11101/0000-0000-A089-0* wird zu *http://hdl.handle.net/api/handles/21.11101/0000-0000-A089-0* und die URL wird aufgelöst zu *http://wwwuser.gwdg.de/~thinder/bpp/CSG7110Pid.xlsm*.

Gehen wir zum Abschluss noch kurz auf die zwei zuletzt genannten Aufrufe in den Quellzeilen-Auflistungen 6 und 7 ein.

Der Kern der Routine zum Herunterladen der ermittelten Excel-Datei aus Quellzeilen-Auflistung 7 ist in Quellzeilen-Auflistung 8 zu sehen, die komplette Routine zur Ermittlung des eigentlichen URL aus Quellzeilen-Auflistung 6 in Quellzeilen-Auflistung 9.

Somit sind die wichtigsten Komponenten der zur Ermittlung und zum Herunterladen der durch PIDs oder DOIs angegebenen Excel-Dateien in diesem Artikel kurz vorgestellt worden.

AUSBLICK

Inzwischen wird an zwei weiteren zentralen Routinen gearbeitet: Zum einen an der Prüf-Routine, die auf gültige Datendateien des ausgewählten Datensatz (engl. Dataset) prüft und nur diese zur Verarbeitung zulässt, zum anderen an der Verarbeitungs-Routine, die die Datensätze der Datendateien verarbeitet und endgültig in der Datenbank speichert.

Ziel ist es, eine Machbarkeits-Version (engl. Proof of Concept) auf der nächsten BEXIS++-Benutzerkonferenz vorzustellen.

FORTSETZUNG FOLGT

In einer der kommenden Ausgaben der GWDG-Nachrichten im ersten Halbjahr 2018 wird ein weiterer Artikel erscheinen, um über den Fortschritt sowie Entwicklungserfahrungen der Batch-Upload-Programmierung zu berichten. Schwerpunktmäßig werden die beiden im Ausblick genannten Routinen den Inhalt dieses zukünftigen Artikels bestimmen. ●

```

/// <summary>
/// Download a file from a given PID or DOI URI and store this file to the given
/// path on the Server with the given filename
/// </summary>
/// <param name="piddoiUri">Resolved PID or DOI URI</param>
/// <param name="filePath">Path where to store the downloaded file</param>
/// <param name="fileName">Name to store the downloaded file</param>
/// <returns>True if the the file is downloaded and stored. Otherwise false.</returns>
private async Task<bool> GetUriAsync(string piddoiUri, string filePath, string fileName)
{
    // Create a new WebClient instance.
    using (WebClient webClient = new WebClient())
    {
        try
        {
            // Download the Web resource and save it into the current file system folder.
            await webClient.DownloadFileTaskAsync(piddoiUri, filePath + "\\ " + fileName);
            return true;
        }
        catch (WebException e)
        {
            return false;
        }
        ...
    }
    ...
}

```

Quellzeilen-Auflistung 8: GetUriAsync

```

/// <summary>
/// Resolve URI from Handle service
/// </summary>
/// <param name="piddoiUri">URI of the PID or DOI to resolve</param>
/// <returns>The resolved DOI/PID URI string</returns>
private async Task<string> ResolveUriFromHandleAsync(string piddoiUri)
{
    string resolvedUri;
    byte[] resultBytes;

    HttpClient webClient = new HttpClient();
    webClient.DefaultRequestHeaders.Accept.Add(new
    MediaTypeWithQualityHeaderValue("application/json"));

    try
    {
        resultBytes = await webClient.GetByteArrayAsync(piddoiUri);

        using (MemoryStream memoryStream = new MemoryStream(resultBytes))
        {
            DataContractJsonSerializer dataContractJsonSerializer = new
            DataContractJsonSerializer(typeof(SelectMultipleFilesPIDDOIModel.Rootobject));
            SelectMultipleFilesPIDDOIModel.Rootobject rootobject =
            dataContractJsonSerializer.ReadObject(memoryStream) as SelectMultipleFilesPIDDOIModel.Rootobject;
            memoryStream.Dispose();
            resolvedUri = rootobject.values[0].data.value.ToString();
        }

        return resolvedUri;
    }
    catch (HttpRequestException e)
    {
        return e.Message + " ==> " + piddoiUri;
    }
}

```

Quellzeilen-Auflistung 9: ResolveUriFromHandleAsync



First International Data Science Summer School in Göttingen

Text and Contact:

Lena Steilen
lena.steilen@gwdg.de
0551 201-2153

For the first time an International Data Science Summer School was hosted in Göttingen. It has successfully welcomed 35 students from a multi-disciplinary background for an intensive two-week course from July 10th to July 21st 2017. The Summer School was part of the newly established Short Term Programs Göttingen of the Göttingen University and organized by the Institute of Computer Science in cooperation with the Göttingen eResearch Alliance and Göttingen International. The GWDG was actively involved in the implementation of the Summer School.

PROGRAM

The participants faced a varied program to learn about data science. The topics ranged from the data life-cycle, over how to manage research data correctly and data analytics methods to ethical and legal aspects of data driven research. The program was based on methods and data use in several disciplines.

Beside the lectures hands on sessions were part of the program. This gave the participants the opportunity to practice the learned methods and see what data science means for different disciplines. Additionally the Data Science Research Bazaar gave researchers from Göttingen University the opportunity to show their practical work with research data to the international participants, and discuss their approaches. Further information about the program can be found on the Summer School web page under <https://www.uni-goettingen.de/de/555483.html>.

LECTURES

The lectures of the first week were focused on more general topics and gave an overview about data management and

commonly used data science methods. The second week focused on data analytics and practical examples.

The organizing committee brought lecturers from diverse disciplines together. Lecturers from computer science were accompanied by researchers from economics, medicine, earth sciences and geography, biology, arts and humanities. The lecturers came

Erste internationale Data Science Summer School in Göttingen

Zum ersten Mal fand eine internationale Data Science Summer School in Göttingen statt. Sie konnte 35 Studierende mit multidisziplinärem Hintergrund für diesen intensiven zweiwöchigen Kurs vom 10. bis 21. Juli 2017 willkommen heißen. Die Summer School war Teil des Short Term Programs Göttingen, welches seit einiger Zeit an der Universität Göttingen existiert, und wurde vom Institut für Informatik in Kooperation mit der Göttingen eResearch Alliance und Göttingen International organisiert. Die GWDG war aktiv an der Durchführung der Summer School beteiligt.



not only from the corresponding faculties but also from various institutions at the Göttingen Campus for example the Göttingen State and University Library, the GCDH, the UMG, the DZNE Göttingen, the MPI for Solar System Research and the MPI for Biophysical Chemistry. Very special thanks go to the scientists of the HeKKSa-GOn network from Japan who joined this event.

KEYNOTES

Highlights of the course were the keynote and the evening lecture. To start things off, Prof. Dr. Hardi Peter (MPI for Solar System Research) gave an enlightening and well-received keynote on “The role of data acquisition and processing in solar astronomy”, where he presented some tools used by scientists to explore data and model the solar atmosphere and its effects on the earth. Prof. Dr. Gerhard Wörner (Geoscience Centre at the Faculty of Geoscience and Geography) talked in his evening lecture about “What to do with 100.000 trace element analysis in volcanic rocks from the Central Andes? Clues to the origin of magmas in time and space”.

CONCLUSION

The Data Science Summer School made clear how important knowledge of data science topics is for young researchers. Raising awareness of the challenges in dealing with research data is in this context as important as teaching methods and the use of tools – although the use of tools was highly requested by the participants. ■



Stellenangebot

Die **GWDG** sucht ab sofort für ihr Netzwerk-Team einen

Senior IT-Netzwerkarchitekten (m/w)

Entgelt und Sozialleistungen richten sich nach dem Tarifvertrag für den öffentlichen Dienst (TVöD). Je nach Qualifikation ist eine Bezahlung bis nach Entgeltgruppe 14 TVöD möglich. Die Stelle ist nicht zur Teilzeit geeignet.

Die GWDG betreibt in ihrem über mehrere Redundanz-Standorte verteilten Rechenzentrum neben üblichen IT-Basisdiensten wie Storage und E-Mail eigene Cloud-Anwendungen sowie Hosting-, Infrastructure-as-a-Service- und HPC-Dienste. Hierbei werden hochverfügbare Routing- und Switching-Infrastrukturen sowie Firewall-Lösungen eingesetzt. Mit der geplanten Inbetriebnahme des Hochleistungsrechners HLRN-IV in 2018 und eines Rechenzentrum-Neubaus in 2020 wird dieses Netz weiter ausgebaut.

Im Auftrag der Georg-August-Universität Göttingen betreibt die GWDG das Datenübertragungsnetz der Universität Göttingen (GÖNET), das die Institute des Göttinger Campus untereinander und mit dem Deutschen Forschungsnetz X-WiN verbindet sowie das campus-weite WLAN eduroam. Sie berät und unterstützt zudem die Institute der Max-Planck-Gesellschaft bezüglich Datenübertragungsnetze.

Ihre Aufgaben

- Planung und Umsetzung von komplexen Netzwerk-Infrastrukturen für die GWDG und die Institute der Universität Göttingen und Max-Planck-Gesellschaft sowie Beratung zu und Begutachtung von Netzwerk-Projekten in der Max-Planck-Gesellschaft
- Konzeption und Weiterentwicklung der Datenübertragungsnetze der GWDG und der Universität, u. a. für die Integration des HLRN-IV sowie den RZ-Neubau
- Koordination und Anleitung des Netzwerk-Teams
 - › bei der Installation, Konfiguration, Überwachung und Störungsbeseitigung der Netzwerk-Komponenten, insbesondere von Routern, Firewalls, VPN-Gateways, Switches, WLAN-APs und -Controllern, u. a. der Marken Cisco, Checkpoint und HPE / Aruba,
 - › beim Betrieb von IPAM-, DNS- und DHCP-Servern, u. a. der Marke BlueCat Networks
 - › unter Vorgabe von Strategien und bei Analyse komplexer Problemsituationen.

Unsere Anforderungen

- Abgeschlossenes technisches Studium mit

- IT-Orientierung oder eine abgeschlossene Berufsausbildung mit adäquater beruflicher Erfahrung
- Mindestens fünf Jahre Berufserfahrung im Bereich Netzwerk-Administration
- Sehr gute Kenntnisse und Erfahrung bezüglich des Aufbaus und der Funktionsweise von großen, heterogenen, IP-basierten Datenübertragungsnetzen im LAN-Bereich
- Erfahrung mit Techniken wie BGP, OSPF, NAT, IPsec, STP, HSRP, vPC, FabricPath, WLAN, IEEE 802.1X, IPv6 und DNSSEC
- Sehr gute Kenntnisse in der Administration von Cisco Nexus, Catalyst und ASA oder vergleichbaren Netzwerk-Komponenten
- Erfahrung im Projektmanagement
- Fähigkeit, selbstständig und systematisch zu arbeiten, analytisches Denken sowie eine rasche Auffassungsgabe
- Überdurchschnittliche Kommunikations- und Teamfähigkeit
- Gute Deutsch- und Englischkenntnisse in Wort und Schrift

Ihre Chance

- Festanstellung
- Interessante und sehr abwechslungsreiche Aufgabenstellung
- Weitreichende Gestaltungsmöglichkeiten
- Flache Hierarchien
- Flexible Arbeitszeiten mit Zeitkonto
- Attraktive soziale Rahmenbedingungen nach dem TVöD (39-Stunden-Woche, 30 Tage Urlaub, jährliche Sonderzahlung)
- Attraktive Universitätsstadt zwischen Harz und Weser

Die GWDG will den Anteil von Frauen in den Bereichen erhöhen, in denen sie unterrepräsentiert sind. Frauen werden deshalb ausdrücklich aufgefordert, sich zu bewerben. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht.

Bitte reichen Sie Ihre Bewerbung mit allen wichtigen Unterlagen **bis zum 15. September 2017** über das Online-Formular unter <https://s-lotus.gwdg.de/gwdgdb/agi/20170809.nsf/bewerbung> ein.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

Herr Stefan Teusch

Tel.: 0551 201-1866

E-Mail: stefan.teusch@gwdg.de

ABSCHIED VON DR. SONG YANG

Herr Dr. Song Yang hat die GWGD nach zweijähriger Tätigkeit als Senior Researcher in der Arbeitsgruppe „eScience“ (AG E) am 31. Juli 2017 verlassen, um seine Forschungen in seinem Heimatland China fortzusetzen. Während seines Forschungsaufenthaltes bei der GWGD war Herr Dr. Yang im Projekt „CleanSky – Network for Cloud Computing Eco-System – ITN (Initial Training Network)“ tätig. Bei diesem Trainingsnetzwerk handelt es sich um ein innovatives Ausbildungsnetz der Europäischen Union, gefördert durch das Marie Skłodowska-Curie actions – Research Fellowship Programme. Dr. Yangs Forschungsthema lautete „Langzeitspeicherung von datenintensiven Anwendungen in der Cloud am Beispiel von Gen-Sequenzierungs-Daten“. Zur Vermeidung von Datenverlusten, zum Beispiel infolge von Serverausfällen aufgrund von Hardware-Defekten oder Software-Problemen, entwickelte er ein Datenspeicherungs-Framework, das überprüft, wie zuverlässig die Daten gespeichert werden. Für das häufig auftretende Problem der Variabilität der Datenzugriffszeit bei Cloud-Storage-Services entwarf er eine latenzempfindliche Datenplatzierungsstrategie. Er arbeitete ferner an Lösungsansätzen zur Ressourcenskalerung bei datenintensiven Anwendungen. Zur Integration der zuvor genannten Verfahren erstellte er ein Framework. Über seine Forschungen bei der GWGD veröffentlichte er eine Reihe von Publikationen, darunter in IEEE RNDM 2016, IFIP / IEEE IM 2017, IEEE Transactions on Parallel and Distributed Systems und Elsevier Journal of Computer Networks. Wir danken Herrn Dr. Yang für die Zusammenarbeit und wünschen ihm für seinen beruflichen und privaten Lebensweg weiterhin alles Gute und viel Erfolg.



Wieder



NEUE MITARBEITERIN JANINE NOWACK

Seit dem 1. Juli 2017 unterstützt Frau Janine Nowack die Arbeitsgruppe „Nutzer-service und Betriebsdienste“ (AG H). Schwerpunkte ihrer Tätigkeit werden der Benutzersupport und verschiedene administrative Aufgaben im engeren Arbeitsumfeld sein. Frau Nowack ist ausgebildete Fachinformatikerin für Anwendungsentwicklung mit Kenntnissen in C und C# sowie in Betriebssystemen für mobile Geräte. Im Rahmen einer längeren ehrenamtlichen Tätigkeit hat sie zudem Schülerinnen und Schüler mit IT-Grundlagen vertraut gemacht. Frau Nowack ist per E-Mail unter janine.nowack@gwdg.de und telefonisch unter 0551 201-1543 erreichbar.

Heuer

NEUE MITARBEITERIN RENATE SCHOLZ

Seit dem 1. Juli 2017 wird die Verwaltung durch eine neue Mitarbeiterin unterstützt: Frau Renate Scholz übernimmt vorübergehend Aufgaben im Sekretariat der GWGD. Sie ist ausgebildete Industriekauffrau mit mehrjähriger Berufserfahrung als Finanzbuchhalterin. Ihr Aufgabenschwerpunkt wird im Bereich des Sekretariats und in vorbereitenden Tätigkeiten für die Verwaltung sowie in der Drittmittelverwaltung und Buchhaltung liegen. Frau Scholz ist telefonisch unter 0551 201-1803 und per E-Mail unter renate.scholz@gwdg.de zu erreichen.



Suren

NEUER MITARBEITER MARCEL SCHNEIDER NEUE MITARBEITERIN MAIKE DAWIDOWSKI

Das Support-Team der GWDC hat Verstärkung durch zwei neue studentische Hilfskräfte bekommen: Herr Marcel Schneider ist dort seit dem 15. Mai 2017 tätig, Frau Maike Dawidowski seit dem 1. August 2017. Schwerpunkte ihrer Arbeitszeiten werden die Abendstunden und Wochenenden sein, um dann Anrufe oder Anfragen per Ticket-System von ratsuchenden Nutzern der GWDC entgegenzunehmen und zu bearbeiten. Beide studieren an der Georg-August-Universität Göttingen im Bachelor-Studiengang Angewandte Informatik. Wir freuen uns über die Verstärkung im Support-Bereich und wünschen ihnen eine erfolgreiche Tätigkeit bei der GWDC.

Helmvoigt



NEUER AUSZUBILDENDER THARMILAN RAJAKUMAR

Am 1. August 2017 hat Herr Tharmilan Rajakumar seine Ausbildung zum Fachinformatiker (IHK) in der Fachrichtung Systemintegration bei der GWDC begonnen. Während seines Studiums der Wirtschaftsinformatik konnte Herr Rajakumar bereits Kenntnisse in der Informatik erwerben und möchte jetzt dieses Wissen mit einer mehr praxisorientierten Ausbildung zum Fachinformatiker für Systemintegration weiter ausbauen. Herr Rajakumar ist per E-Mail unter tharmilan.rajakumar@gwdg.de und telefonisch unter 0551 201-1533 zu erreichen..

Herbold

NEUE STIPENDIATIN SHIRIN DABBAGHI VARNOSFADERANI

Seit dem 1. August 2017 ist Frau Shirin Dabbaghi Varnosfaderani als Stipendiatin in der Arbeitsgruppe „eScience“ (AG E) tätig. Frau Dabbaghi Varnosfaderani bringt einen Bachelor-Abschluss in „Hardware Engineering“ der Islamic Azad University of Najafabad, Iran (2011) und einen Master-Abschluss in „Distributed Systems“ (2014) der Universiti Putra Malaysia mit. Sie hat sich auf das Arbeitsfeld „Security in Cloud Computing“ spezialisiert und arbeitet bei der GWDC an ihrer Dissertation „A combined approach to ensure data security in cloud computing“. Frau Dabbaghi Varnosfaderani ist per E-Mail unter shirin.dabbaghi-varnosfaderani@gwdg.de und telefonisch unter 0551 201-2198 erreichbar.

Wieder





Using the Parallel Processing Power of the GWDG Scientific Compute Cluster

Upcoming Introductory and Parallel Programming Courses

GWDG operates a scientific compute cluster with currently 17,048 cores and a total compute power of 369 Teraflops ($3.69 \cdot 10^{14}$ floating point operations per second), which can be used by all scientists of the institutes of GWDG's supporting organisations, University of Göttingen and Max Planck Society.

In order to facilitate the access to and the efficient use of these computing resources, GWDG offers introductory and parallel programming courses, held at GWDG's site 'Am Faßberg'.

The next courses in 2017 are

> October 16th, 9:30 am - 4:00 pm

Using the GWDG Scientific Compute Clusters – An Introduction

This course explains all steps for accessing GWDG's clusters, to compile and install software, and to work with the batch system for the execution of application jobs. The course is intended for new or inexperienced users of the clusters.

> October 17th - 18th, 9:15 am - 5:00 pm

Parallel Programming with MPI (Including MPI for Python)

This course introduces the message passing interface (MPI) for programming parallel applications in FORTRAN, C, and in Python. All con-

cepts will be illustrated with hands on exercises. Examples of parallel applications will be presented and analysed.

> November 7th, 9:15 am - 5:00 pm

Programming with CUDA – An Introduction

Graphic processors (GPUs) are increasingly used as computational accelerators for highly parallel applications. This course introduces hardware and parallelization concepts for GPUs and the CUDA programming environment for C and Fortran, including the language elements for controlling the processor parallelism and for accessing the various levels of memory.

These three courses are repeated regularly. Other courses on parallel computing, dealing with more specialized topics can be arranged on demand. The possible subjects include parallel programming for shared memory systems and using extensions of C or Fortran with high level parallel constructs.

More Information about the courses held regularly or on demand at www.gwdg.de/scientific-computing-courses.

Information for registering for the courses at www.gwdg.de/courses.

If you have any further questions please contact support@gwdg.de.



INFORMATIONEN:
support@gwdg.de
0551 201-1523

September bis
Dezember 2017

Kurse

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ANWENDER	Buck	07.09.2017 9:00 – 12:30 Uhr	31.08.2017	2
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ADMINISTRATOREN	Buck	07.09.2017 13:30 – 17:00 Uhr	31.08.2017	2
INDESIGN – GRUNDLAGEN	Töpfer	12.09. – 13.09.2017 9:30 – 16:00 Uhr	05.09.2017	8
NETZLAUFWERKE FÜR DIE GE- MEINSAME DATENABLAG	Quentin	21.09.2017 9:00 – 12:30 und 13:30 – 15:30 Uhr	14.09.2017	4
FUNDAMENTALS OF IMAGE PROCESSING WITH PHOTO- SHOP	Töpfer	26.09. – 27.09.2017 9:30 – 16:00 Uhr	19.09.2017	8
USING THE GWDC SCIENTIFIC COMPUTE CLUSTER – AN IN- TRODUCTION	Dr. Boehme Ehlers	16.10.2017 9:30 – 16:00 Uhr	09.10.2017	4
PARALLELRECHNERPRO- GRAMMIERUNG MIT MPI	Prof. Haan	17.10. – 18.10.2017 9:15 – 17:00 Uhr	10.10.2017	8
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ANWENDER	Buck	19.10.2017 9:00 – 12:30 Uhr	12.10.2017	2
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ADMINISTRATOREN	Buck	19.10.2017 13:30 – 17:00 Uhr	12.10.2017	2
PHOTOSHOP FÜR FORTGE- SCHRITTENE	Töpfer	24.10. – 25.10.2017 9:30 – 16:00 Uhr	17.10.2017	8
ADMINISTRATION VON PCS IM ACTIVE DIRECTORY DER GWDC	Quentin	02.11.2017 9:00 – 12:30 und 13:30 – 15:30 Uhr	26.10.2017	4

KURS	VORTRAGENDE/R	TERMIN	ANMELDEN BIS	AE
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	07.11. – 08.11.2017 9:00 – 12:00 und 13:00 – 15:30 Uhr	31.10.2017	8
PROGRAMMING WITH CUDA – AN INTRODUCTION	Prof. Haan	07.11.2017 9:15 – 17:00 Uhr	31.10.2017	4
INDESIGN – AUFBAUKURS	Töpfer	14.11. – 15.11.2017 9:30 – 16:00 Uhr	07.11.2017	8
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ANWENDER	Buck	16.11.2017 9:00 – 12:30 Uhr	09.11.2017	2
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ADMINISTRATOREN	Buck	16.11.2017 13:30 – 17:00 Uhr	09.11.2017	2
QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R	Cordes	21.11. – 22.11.2017 9:00 – 12:00 und 13:00 – 15:30 Uhr	14.11.2017	8
EINFÜHRUNG IN DAS IP-ADRESSMANAGEMENTSYSTEM DER GWDG FÜR NETZWERKBEAUFTRAGTE	Dr. Beck	28.11.2017 10:00 – 12:00 Uhr 13:30-15:30 Uhr	21.11.2017	2
NETZLAUFWERKE FÜR DIE GEMEINSAME DATENABLAGE	Quentin	30.11.2017 9:00 – 12:30 und 13:30 – 15:30 Uhr	23.11.2017	4
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	07.12.2017 9:15 – 12:00 und 13:00 – 16:00 Uhr	30.11.2017	4
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	12.12. – 13.12.2017 9:00 – 12:00 und 13:00 – 15:30 Uhr	05.12.2017	8
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ANWENDER	Buck	14.12.2017 9:00 – 12:30 Uhr	07.12.2017	2
EINFÜHRUNG IN SHAREPOINT 2013 FÜR ADMINISTRATOREN	Buck	14.12.2017 13:30 – 17:00 Uhr	07.12.2017	2

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 <https://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 Institutskontin-

gent 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 <https://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 <https://www.gwdg.de/kursprogramm> zu finden.



Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen