GWDG-Bericht Nr. 77

Andreas Oberreuter, Stefan Vollmar,
Alexander Weiße (Hrsg.)

27. DV-Treffen der Max-Planck-Institute

14. - 16. September 2010
in Göttingen

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Andreas Oberreuter, Stefan Vollmar,
Alexander Weiße (Hrsg.)

27. DV-Treffen der Max-Planck-Institute

14. - 16. September 2010
in Göttingen

Andreas Oberreuter, Stefan Vollmar,
Alexander Weiße (Hrsg.)

27. DV-Treffen der Max-Planck-Institute

14. - 16. September 2010
in Göttingen

GWDG-Bericht Nr. 77
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
Am Faßberg 11
D-37077 Göttingen
Telefon: 0551 201-1510
Telefax: 0551 201-2150
E-Mail: gwdg@gwdg.de

Satz: Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Druck: Goltze Druck, Göttingen

ISSN 0176-2516

Inhalt

Vorwort

Seit vielen Jahren ist das DV-Treffen der Max-Planck-Institute eine wichtige Plattform für den Austauch über aktuelle Trends der Informationstechnologie, neue Ideen und Lösungen technischer Probleme, die die IT-Mitarbeiter verschiedener Institute gleichermaßen betreffen. Das 27. DV-Treffen, zu dem sich vom 14. bis 16. September 2010 circa 180 IT-Mitarbeiter und IT-Interessierte am bewährten Tagungsort auf dem Göttinger Faßberg trafen, stand deshalb unter dem Motto „Geteiltes Wissen ist mehr als halbes Wissen“.

Schwerpunkte des vielfältigen Programms waren neben typischen Fragen aus den Bereichen Nutzer/Rechner-Administration oder IT-Sicherheit vor allem die Themen Langzeitarchivierung, IPv6 und Content-Management. Darüber hinaus nutzten zentrale Partner, wie das RZG und der DFN-Verein, die Veranstaltung, um ihre vielfältigen Dienstleistungen vorzustellen. Der zeitliche Ablauf des Treffens wurde 2010 leicht abgewandelt. Nachdem sich zuerst die IT-Verantwortlichen zu ihrem 2. Treffen zusammengefunden hatten, startete das DV-Treffen am ersten Tag mit Vorträgen und der Wahl des IT-Sprecherkreises. Die beliebten Workshops füllten diesmal den Nachmittag des zweiten Tages aus. Um mehr Raum für persönliche Kontakte und Diskussionen zu bieten, wurden die Pausen verlängert, was allseits sehr begrüßt wurde. Die Präsenz von externen Firmen und damit verbundene Marketingaktivitäten wurden wieder reduziert.

Eine weitere Neuerung gab es auch zum Abschluss des Treffens: zwei Preisverleihungen. Als Würdigung für Verdienste um die IT-Community der

MPG wurde erstmals der „IT Community Award“ vergeben, dessen Idee wir in diesem Band erläutern. Außerdem wurde ein Preis für den besten Tagungsbeitrag vergeben, über den die Teilnehmer „online“ abstimmen konnten.

Details zum Programm, die Folien der meisten Beiträge und weitere Informationen zum Treffen finden sich auf der Webseite http://27dvt.mpim-bonn.mpg.de/. Neben diesem modernen – und flüchtigeren – Medium wollten wir jedoch nicht auf eine traditionelle Form der Publikation verzichten und haben eine Reihe von Beiträgen in diesem Band zusammengestellt.

Sowohl die Herstellung dieses Bandes als auch das Treffen in Göttingen wären kaum so reibungslos verlaufen, wenn wir nicht die unermüdliche und großzügige Unterstützung durch die GWDG und das Team von Thomas Otto erfahren hätten. Herzlichen Dank!

Bonn und Köln, 04.08.2011

Andreas Oberreuter, Stefan Vollmar, Alexander Weiße

DNSSEC und seine Auswirkungen auf DNS-Dienste in der MPG

Holger Beck

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

1. DNS – damals und heute

Das Domain Name System DNS wurde 1983 in den RFCs 881-883 eingeführt. Damals bestand das Internet gerade einmal aus 562 Host genannten angeschlossenen Netzteilnehmern (Geräten). Das Internet war „nur“ das Netz eines Forschungsverbunds mit Schwerpunkt in den USA.

Damals war das Netz dramatisch gewachsen. Gegenüber dem Start mit fünf Hosts hatte sich die Größe des Netzes verhundertfacht. DNS war die Lösung eines drängenden Problems: Bis dahin wurde die Liste der Netzteilnehmer zentralisiert in einer Datei (`hosts.txt`) vom Network Information Center (NIC) des Internet gepflegt und an die Teilnehmer verteilt. Das Wachstum machte dieses Vorgehen zunehmend impraktikabler.

DNS ersetzte dieses zentralisierte System durch ein Netz dezentralisierter Auskunftsstellen (DNS-Server). Die wesentliche Voraussetzung, durch die eine Dezentralisierung erfolgen konnte, war die Einführung des heute so selbstverständlichen, hierarchisch organisierten Namensraums aus Domänen und Subdomänen. Für jede Domäne oder Subdomäne wurden ein oder

(meist) mehreren DNS-Servern die Zuständigkeit delegiert. Jeder Administrator eines Netzes konnte seinen eigenen Teil des Namensraums erhalten und selbst pflegen.

Heute sind mehr als 650 Millionen Hosts am Internet angeschlossen. Aus einem reinen Forschungsnetz ist ein Netz für jedermann geworden. Im Netz werden nicht nur Forschungsdaten ausgetauscht. Das Internet ist ein Netz für tausenderlei Anwendungen geworden. Für viele Menschen ist es so gut wie unentbehrlich geworden.

So sehr das Netz auch mutierte, das DNS ist seit 1983 im Wesentlichen gleich geblieben. Es ist im Umfang mit gewachsen – an den Prinzipien des DNS hat sich nichts geändert.

2. Integritätsprobleme

DNS ist heute eine zum Funktionieren des Netzes unverzichtbare Komponente geworden. Deutlich erkennbar wurde das beim DNS-Ausfall am 12. Mai 2010. Teile des deutschen Internet-Teiles waren durch Fehler in den Daten der DNS-Server für die de-Domänen praktisch nicht mehr vorhanden. Die Netzwerkinfrastruktur funktionierte unverändert, alle eigentlichen Dienste waren in Betrieb – aber weil für Teile des Netzes die Umsetzung von DNS-Namen zu IP-Adressen nicht mehr erfolgte, war dieser Teil des Netzes praktisch nicht mehr vorhanden.

So dramatisch der Ausfall im Mai 2010 war, so zeigt dieser doch nur besonders deutlich die Bedeutung des DNS und die Fehlbarkeit auch doppelter und dreifacher Absicherungen gegen menschliche Fehler.

Die täglichen Probleme des DNS liegen an anderer Stelle.

Die Bedeutungssteigerung des Internet durch Wachstum in Quantität und Funktionalität machte es interessant für Kriminalität und Sabotage. Ein Teil der Angriffstechniken basiert auf dem Missbrauch oder der Sabotage der DNS-Dienste. DNS wurde für ein relativ kleines und abgeschottetes Forschungsnetz entwickelt. Kriminelle Aktivitäten standen beim Design nicht im Fokus. So waren Sicherheitsfunktionen, die insbesondere die Richtigkeit der erteilten Auskünfte garantieren, nicht vorgesehen.

In den vergangenen Jahren wurden immer wieder einmal Sicherheitslücken in DNS-Software gefunden. Durch Programmierfehler bedingte Sicherheitsupdates sind in jeglichen Softwarekomponenten immer wieder einmal fällig, solange Menschen Software programmieren und dabei nicht fehlerfrei arbeiten. Die 2008 von Dan Kaminsky veröffentlichten Probleme, die eine Manipulation von DNS-Caches mit falschen Daten (Cache-Poisoning) erlaubten,

haben dagegen großen Aufwand verursacht, weil hier nicht ein Programmier-, sondern ein Designfehler ausgenutzt wurde, um gefälschte Antworten zu verteilen. Dieses Designproblem konnte zwar kurzfristig gelöst werden, zeigt aber doch, dass Fälschungssicherheit nicht systematisch gegeben ist.

3. Die Lösung: DNSSEC

Das Problem ist lange bekannt und seit Jahren wird an einer Lösung gearbeitet. Schon im Januar 1997 wurde RFC 2065 verabschiedet, der die DNS Security Extensions (kurz DNSSEC) definierte. Im Laufe der Jahre wurden Verbesserungen, Ergänzungen und Erweiterung an DNSSEC vorgenommen (RFCs 4033-4065, März 2005; RFC 4470, April 2006; RFC 5155, Februar 2008).

Das Konzept von DNSSEC basiert auf dem Einsatz asymmetrischer Kryptographie. Public Keys für DNS-Zonen (DNSKEY-Records) werden über DNS-Server veröffentlicht. Die in der Hierarchie jeweils übergeordnete Zone (Parent-Zone) veröffentlich mit den Delegationseinträgen eine Signatur der Keys der untergeordneten (Child-)Zone in DS-Records. Jede Zone signiert durch zusätzliche RRSIG-Records die eigenen DNS-Einträge (genauer alle Resource Recordsets, d.h. alle Einträge mit gleichem Namen und Typ werden gemeinsam signiert). Über die kryptographische Signatur ist dann überprüfbar, ob die erhaltene Auskunft, vom autorisierten DNS-Server kommt und unverfälscht ist.

Zusätzlich wird auch sichergestellt, dass Auskünfte über die Nichtexistenz von Einträgen überprüfbar sind. Dazu werden sogenannte NSEC- (ältere Variante) bzw. NSEC3-Records (neuere Variante) verwendet.

4. Verfügbarkeit von DNSSEC

Die Einführung von DNSSEC im Internet ist lange nicht vorangekommen, aber 2010 gab es wesentliche Fortschritte. Insbesondere die Spitze der DNS-Hierarchie (die Root-Zone) wurde in 2010 signiert. Auch `.org` und einige ccTLD (z. B. `.se`) sind bereits signiert. Für die `de`-Zone wurde in 2010 ein Testbed bereitgestellt und betrieben.

Neben der Eintragung von Schlüsseln und Signaturen in den DNS-Servern wird auch die Unterstützung der Verfahren in DNS-Server- und DNS-Clienten-Software benötigt. Auch die Unterstützung von DNSSEC in Softwarekomponenten ist 2010 vielfältig vorhanden.

Die Validierung von DNS-Antworten, kann vom DNS-Clienten auf dem Endgerät oder ersatzweise vom lokalen DNS-Server für die Endgeräte erfol-

gen, wobei dann der als lokaler DNS-Resolver eingesetzte DNS-Server aus Sicht des Endgerätes vertrauenswürdig sein muss. Die Validierung durch den DNS-Server unterstützt der weit verbreitete Bind-Server in aktuellen Versionen. Die DNS-Server von Microsoft validieren DNSSEC ab Windows 2008 R2, unterstützen dabei aber nur NSEC-Records und keine NSEC3-Records.

Auf Ebene der DNS-Clienten unterstützen Linux, MacOS und BSD mit zusätzlichen Libraries die direkte Validierung durch den DNS-Clienten. Windows ab Version 7/2008 R2 ist ebenfalls DSNSEC-aware und kann somit DNS-Einträge über DNSSEC validieren.

5. Umsetzung und Folgen

Zunächst benötigt DNSSEC kleinere Erweiterungen des **DNS-Protokolls**, insbesondere ein paar Flags zur Nutzung von DNSSEC im DNS-Header. Probleme können hier mit DSL-Routern als DNS-Proxies oder DNS-Router auftreten, wenn diese DNSSEC nicht korrekt implementieren.

Wesentlich für die Nutzung von DNSSEC ist die Etablierung von **Vertrauensbeziehungen**. Denn selbst wenn innerhalb der DNS-Hierarchie das Vertrauen zum Schlüssel untergeordneter (Child-)Zonen durch die Signatur in der übergeordneten (Parent-)Zone abgeleitet werden kann, müssen irgendwo durch einen oder mehrere Trust Anchor eine Wurzel oder mehrere Wurzeln des Vertrauens definiert werden. Mit Signierung der Root-Zone ist die Voraussetzung geschaffen, dass mittelfristig nur die Signatur dieser Zone als Trust Anchor benötigt wird (sobald der Weg zu allen signierten Subzonen lückenlos signiert ist).

Die **Signierung von Resource Recordsets** fordert einen gestiegenen Rechenaufwand bei der die Signaturen erstellenden Instanz – also in vielen Fällen bei dem Master-DNS-Server der Zone. Der Aufwand entsteht dabei nicht nur einmalig beim Anlegen der Records, sondern auch beim aus Sicherheitsgründen notwendigen Schlüsselwechsel (Key Rollover) und dem damit notwendigen Resignieren alle Recordssets mit dem neuen Schlüssel.

Die **Speicherung** zusätzlicher Records für DNSSEC bedingt größere DNS-Dateien in DNS-Servern, was aber nur für Registries oder sehr große Zonen relevant sein dürfte.

Im Zusammenhang mit **Schlüsselgenerierung und Signierung** sind zu beachten:

- Kommunikation mit **Parent** wegen Signierung

  Hier ist die Entwicklung von Prozeduren notwendig, die sicherstellen, dass dem Parent Public Keys auf sicherem Weg mitgeteilt werden können, damit dieser auch verlässliche Signaturen erstellen und bereitstellen kann.

- **Sichere Speicherung** der privaten Schlüssel

  Hier ist zu klären, wie ein privater Schlüssel, der zur Signatur von Records ja immer wieder benötigt wird, sicher gespeichert werden kann. Systeme, auf denen private Schlüssel liegen, müssen dabei besonders geschützt werden. Hier kann eigentlich nicht mehr die DNS-Server-Funktionalität einer kleineren Einrichtung nebenbei von einem Server erledigt werden, der wegen anderer Hauptfunktionen allgemein zugreifbar sein muss.

- **Signierung** der Records nach Änderungen von Datensätzen

  Die Signierung muss organisiert werden, damit jeder neue Datensatz immer sofort auch signiert wird.

Besonders zu beachten bei Schlüsselgenerierung und Signierung ist die Notwendigkeit der Erneuerung von Schlüsseln (Key Rollover). Eine periodische Änderung der Schlüssel und damit der Signaturen erfolgt zum Schutz gegen Angriffe auf Schlüssel. Dabei muss dann auch eine erneute Signatur durch den Parent erfolgen. Als Kompromiss zwischen sichereren (längeren) Schlüsseln und damit verbunden höherem Rechenaufwand bei Signierung und Validierung und möglichst langer Verwendbarkeit von Schlüsseln erfolgt eine Unterteilung in

- **Key Signing Keys (KSK)**, die kryptographisch stärker und länger gültig sind und vom Parent signiert werden müssen (mit allen damit verbundenen Kommunikations- und Organisationproblemen) und
- **Zone Signing Keys (ZSK)**, die kürzer und weniger sicher sind, daher schneller gewechselt werden müssen, aber mit eigenem KSK signiert werden. Damit ergibt sich zwar ein häufiger wiederkehrender Arbeitsaufwand, aber kein Kommunikationsproblem mit dem Parent, sowie durch den kürzeren Schlüssel ein geringerer Rechenaufwand bei Signierung und Validierung.

Ein weiterer Nebeneffekt der DNSSEC-Einführung ist, dass **DNS-Antwortpakete größer** werden. Bisher waren die für DNS-Abfragen verwendeten UDP-Datagramme maximal 512 Byte groß. DNSSEC-Antworten enthalten mehr Informationen, so dass größere Datagramme benötigt werden. DNS-

Datagramme dürfen daher bis zu **4.096 Byte** groß sein. Bei typischen MTUs von 1.500 Byte können durch **Fragmentierungen** Probleme entstehen. Besonders problematisch kann die traditionelle 512-Byte-Grenze in **Firewalls** sein. Diese häufig verwendete Regel muss aus allen Firewalls entfernt werden, damit DNSSEC funktioniert. Falls die größeren UDP-Pakete im Netz nicht durchkommen, bleibt sonst nur das Ausweichen auf **Abfragen über TCP**, was im Standard vorgesehen ist, aber möglicherweise in Firewalls auch blockiert wird (weil bisher nur Zonen-Transfers über den TCP-DNS-Port liefen, die man ja meist verhindern wollte).

Dynamische DNS-Einträge stellen ebenfalls ein Problem dar. Bei Windows-DNS-Servern gibt es keine DNSSEC-Unterstützung für DDNS. Hier können DNSSEC und DDNS nur nebeneinander leben, wenn eine Trennung in Zonen mit DDNS und ohne DNSSEC und solche ohne DDNS und mit DNSSEC vorgenommen wird. BIND-Server bieten eine DDNS-Unterstützung, aber eine Signierung von dynamischen Einträgen kann nur erfolgen, wenn der private Schlüssel auf Servern vorhanden ist. Das ist zum Schutz des privaten Schlüssels auch nicht wirklich erwünscht. Daher ist auch hier ein Verzicht auf Signierung dynamischer Einträge zu überlegen (durch Trennung in Zonen wie oben erwähnt).

6. Beispielzonendaten

Nachstehend sind einige Beispiele für die zusätzlichen DNS-Records bei Verwendung von DNSSEC aufgeführt

6.1 SOA-Record

```
example. 3600 IN SOA ns1.example. bugs.x.w.example. 1 3600 300 (
              3600000 3600 )

         RRSIG SOA 7 1 3600 20150420235959 20051021000000 (
               40430 example.
               Hu25UIyNPmvPIVBrldN+9Mlp9Zql39qaUd8i
               q4ZLlYWfUUbbAS41pG+68z81q1xhkYAcEyHd
               VI2LmKusbZsT0Q== )
```

6.2 NS-Records

```
NS      ns1.example
NS      ns2.example.

RRSIG   NS 7 1 3600 20150420235959 20051021000000 (
        40430 example.
        PVOgtMK1HHeSTau+HwDWC8Ts+6C8qtqd4pQJ
        qOtdEVgg+MA+ai4fWDEhu3qHJyLcQ9tbD2vv
        CnMXjtz6SyObxA== )
```

6.3 MX-Record

```
MX      1 xx.example.

RRSIG   MX 7 1 3600 20150420235959 20051021000000 (
        40430 example.
        GgQ1A9xs47k42VPvpL/a1BWUz/6XsnHkjotw
        9So8MQtZtl2wJBsnOQsaoHrRCrRbyriEl/GZ
        n9Mto/Kx+wBo+w== )
```

6.4 DNSKEY-Records

```
DNSKEY  256 3 7 AwEAAaetidLzsKWUt4swWR8yu0wPHPiUi8LU (
        sAD0QPWU+wzt89epO6tHzkMBVDkC7qphQO2h
        TY4hHn9npWFRw5BYubE= )
DNSKEY  257 3 7 AwEAAcUlFV1vhmqx6NSOUOq2R/dsR7Xm3upJ (
        j7IommWSpJABVfW8Q0rOvXdM6kzt+TAu92L9
        AbsUdblMFin8CVF3n4s= )

RRSIG   DNSKEY 7 1 3600 20150420235959 (
        20051021000000 12708 example.
        AuU4juU9RaxescSmStrQks3Gh9FblGBlVU31
        uzMZ/U/FpsUb8aC6QZS+sTsJXnLnz7flGOsm
        MGQZf3bH+QsCtg== )
```

7. DNSSEC in der MPG

7.1 Notwendigkeit?

Die Frage nach der Notwendigkeit der Einführung von DNSSEC in der MPG ist in zwei Teile zu gliedern:

- Ist es notwendig, eigene DNS-Zonen der MPG zu signieren?
- Ist es notwendig, DNS-Abfragen durch DNS-Clienten oder DNS-Server der MPG validieren zu lassen?

Bezüglich der Signatur eigener Zonen ist eine dringende Notwendigkeit, wie sie z. B. bei DNS-Zonen von Banken zu sehen ist, sicherlich im wissenschaftlichen Umfeld nicht in gleichem Umfang zu sehen, da die Auswirkungen bei Fälschungen in der Regel gering sind. Dennoch sollte eine wissenschaftlich führende und innovative Einrichtung wie die MPG auch hier durch Einsatz moderner Technologie in angemessenem Umfang in Erscheinung treten.

Die Notwendigkeit der Validierung von DNS-Anfragen auf Rechnern bzw. DNS-Servern der MPG hängt von den Risiken ab, die mit der Nutzung von gefälschten DNS-Antworten verbunden wären. In der Regel sind diese Risiken nicht als besonders hoch einzuschätzen. Dennoch bedeutet die Nutzung

von DNSSEC einen Sicherheitsgewinn, auf den nicht verzichtet werden sollte, soweit die Einführung von DNSSEC nicht zu erheblichen Betriebsproblemen oder Kosten führt (was nicht zu erwarten ist).

7.2 Umsetzungsprobleme

Probleme bzw. Aufwände bei der Einführung von DNSSEC können entstehen bei

- **Anpassung von Firewallregeln:** Hier muss, wie oben erwähnt, sichergestellt werden, dass eine Begrenzung von Paketgrößen auf 512 Byte bei DNS-Paketen aus den Konfigurationen entfernt wird und DNS-Abfragen auch über TCP erlaubt werden. Der Aufwand dürfte hier seitens der Institute gering sein, da die Anzahl der Firewalls doch gering ist.
- **Erstellung und Erneuerung von Signaturen bei Delegation der Institutszonen:** Die Signatur der einen Zone `mpg.de` stellt keinen erheblichen Aufwand dar (und würde von der GWDG übernommen). Bei der Delegation von Zonen unterhalb `mpg.de` an DNS-Server der Institute entsteht pro Zone ein Kommunikationsaufwand bei der Erstsignatur und jeder Erneuerung der KSKs zwischen Institut und GWDG. Der Arbeitsaufwand selbst dürfte für die Institute bei einer meist begrenzten Anzahl von Zonen und KSK-Gültigkeiten von mehreren Jahren eher gering sein. Größer könnte das organisatorische Problem sein, wenn nicht rechtzeitig an die Erneuerung der KSKs gedacht wird.
- **Signierte Domänen außerhalb `mpg.de`:** Hier ist zusätzliche Kommunikation zwischen GWDG und ISP bzw. Institut und ISP bei Signierung und KSK-Erneuerung nötig, wenn auch diese Zonen signiert werden sollen. Im Prinzip gelten hier dieselben Abwägungen wie bei der Delegation von mpg.de-Subzonen, nur sind hier andere Kommunikationspartner involviert.
- **Signierung von Zonendaten:** Prinzipiell sind Tools zur Signierung von Zonen vorhanden. Für eine Automatisierung und die Ergebniskontrolle bei Automatisierung gibt es Lösungen. Diese Lösungen sind aber meist noch nicht so verbreitet und ausgetestet, wie das wünschenswert wäre.
- **Designfragen:** Die Strukturierung von DNS-Diensten ist im Forschungsumfeld häufig noch sehr einfach gehalten und entspricht teilweise dem Stand der frühen 90er Jahre, also des Zeitpunktes, zu dem das Internet sich in der deutschen Forschungslandschaft verbreitet hat. Die Einführung von DNSSEC wirft auch die Designfragen auf. Eine Unterscheidung in externe und interne DNS-Server, die Rollentrennung zwischen

autoritativen Servern und Resolvern oder die Einführung eines Hidden Master (der dann samt aller privaten Schlüssel von außen unerreichbar sein kann) ist prinzipiell zu empfehlen. Eine DNSSEC-Einführung sollte zum Anlass genommen werden, dass Design zu prüfen (auch wenn das keine zwingende Notwendigkeit ist).

7.3 Perspektive: Nutzung des IPAM-Systems der GWDG

Das von der GWDG betriebene IP-Adressmanagementsystem der Firma Bluecat bietet bei der Einführung von DNSSEC Vorteile. Prozeduren für Key-Rollover und Signierung der Records sind im System vorhanden. Eine Automatisierung dieser Vorgänge kann daher unproblematisch erfolgen.

Soweit die Master-Server von delegierten Subzonen von `mpg.de` im IPAM-System betrieben werden, erfolgt auch der Schlüsselaustausch zwischen `mpg.de` und den Subzonen automatisch. Dazu können von den Max-Planck-Instituten eigene IPAM-Appliances (von Bluecat) beschafft und integriert werden, oder der Master-Server wird auf eine vorhandene IPAM-Appliance der GWDG verlegt und die bisherigen DNS-Server der Institute werden als Slave-Server angebunden. Damit übernimmt das IPAM-System der GWDG der GWDG die DNSSEC-spezifischen Aufgaben bei der Bereitstellung signierter Zonen.

Ansätze und Randbedingungen für eine MPG-weite wissenschaftliche Cloud-Lösung

Ist Cloud Computing produktiv nutzbar oder braucht es dafür noch Jahre? Der Versuch einer Ortsbestimmung.

Bertram Smolny, Samy Elshamy

Max-Planck-Institut für Biogeochemie, Jena

1. Prolog

Mobiles (in Zukunft ubiquitäres) Computing erfordert ständige Verfügbarkeit von Daten und Diensten. Dem kann mit neuen Technologien Rechnung getragen werden: Cloud Computing soll Rechnerleistung, Services und Anwendungen flexibel per Internet liefern. Liegen aber nicht noch Welten zwischen Theorie und Praxis? Soll doch das, was mit „Grid Computing“ begann, dann „Utility Computing“ hieß, zukünftig flexibel lieferbar aus der „Wolke“ kommen.

2. Was ist eine Wolke?

Die Informationsverarbeitung durchläuft derzeit wieder einen jener Paradigmenwechsel, an dessen vorderster Front jetzt Technologien wie Cloud Computing stehen. Dabei sollen (die inzwischen virtualisierten) Plattformen,

Infrastrukturen und Softwareleistungen als Services für Kunden bereit gestellt werden. Das ist so neu nicht, denn schon seit Jahren macht Grid Computing von sich reden, ohne richtig außerhalb der Wissenschaft aus den Startlöchern gekommen zu sein.

Es mag unterschiedliche Wertungen dieser Technik geben, unbestritten ist jedoch, dass es signifikante Vorteile für Organisationen gibt, die Cloud Computing anwenden: Skalierbarkeit, Flexibilität und Kostensenkung.

Trotzdem muss sich jede Organisation, die Services aus der Cloud nutzt, über deren Risiken im Klaren sein und angemessene Risikobehandlungsstrategien umsetzen. Diese Risiken können eine Vielzahl von Themen betreffen, u. a. neben dem Problem des Zugriffs über nichtredundate Netzwerke, Datensicherheit oder auch Risiken des Lieferanten. Hier sei nur kurz erwähnt, dass im Zuge der sogenannten „cablegate"-Affäre einer der Marktführer der Storage Cloud seinen Kunden Dienste auf politischen Druck hin verweigerte.

2.1 Die Beratersicht[1]

„Cloud Computing kommt auf jeden Fall", verkündet Russ Daniels, CTO Cloud Services Strategy bei Hewlett Packard. „Das sehen wir am steigenden Interesse an Virtualisierung und Software-as-a-Service (SaaS)." Gleichzeitig räumt er ein: „Bis zum Durchbruch braucht Cloud Computing noch Jahre." Da sei noch Einiges zu lernen.

Zum Beispiel, eine einheitliche Definition zu finden. Daniels gibt zu, dass der Hype um die Wolke derzeit ebenso groß ist wie die Verwirrung. Zum Teil werde alter Wein in neuen Schläuchen verkauft.

Sein Verständnis: Cloud Computing basiere auf der Nutzung von Virtualisierung. Ziel sei es, Anwendungen und Services unabhängig von einer bestimmten Infrastruktur nutzen zu können. Auslastung und Ressourcen müssten sich flexibel an den wechselnden Anforderungen des Business ausrichten lassen.

2.2 Definition

Das National Institute for Standards and Technologies (NIST) definiert Cloud Computing wie folgt:

1. http://www.cio.de/knowledgecenter/storage/871063/

„Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.“[2]

Schlüsselcharakteristiken könnten sein:

- bedarfsorientierter „Self Service“, wie Speicherzuordnung von Endbenutzern
- „elastische“ Dienste, die nutzungsorientiert skalieren
- „pay per Use“-Dienste
- Netzzugänge für mobile Endgeräte
- Ressourcenpooling für Organisationen oder Endkunden

2.3 Ein Blick ins Buch (und zwei ins Leben)

Nicolas Carr, zweifellos einer der provokantesten und prophetischsten Denker gegenwärtiger IT-Trends, setzt in seinem Buch „The Big Switch“ auf eine Manifestierung der bestehenden Oligopole: Die Informationsindustrie liefert bald sämtliche ITK-Dienste wie „Strom aus der Steckdose“. Nach dem Vorbild der Energiegiganten braucht man nur noch eine Handvoll von diesen und man ist aller Sorgen ledig. Wo das hinführt, kann man am deutschen Energiemarkt ablesen – Resultat sind Ressourcenverschwendung und Preisdiktat.

Richtig dagegen ist: Das Internet hat einen dezentralen Ansatz. Hier kann, im Gegensatz zu den großen Energieerzeugern, jeder einspeisen. Das hat jetzt auch die ITK-Industrie entdeckt.

IBM und Cisco z. B. werben mit Smart- und Grid-Technologie, die auf einem dezentralen Ansatz fußen. Peer-2-Peer-Netze mit intelligenter Steuerung werden als ein Markt gesehen und intensiv beworben. Hier passen auch Beispiele wie die Kooperation von Volkswagen und LichtBlick mit einem Fokus auf die Einführung von Netzstrukturen, die sich dezentraler Architekturen bedienen.

2. http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Volkswagen und LichtBlick haben sich zusammengetan, um das innovative „Zuhause-Kraftwerk“ realisieren zu können. Volkswagen bringt sein Expertenwissen im Bau von PKW-Serienmotoren sowie die Fähigkeit zur Produktion großer Stückzahlen in die Kooperation ein. LichtBlick ist Experte im Sektor erneuerbare Energien und verfügt über langjährige Erfahrungen im Vertrieb von Strom und Gas. Zusammen wollen sie ein intelligentes Grid von sogenannten „Zuhause-Kraftwerken“ entwickeln, dessen Vorbild ein Peer-to-Peer-Netzwerk aus der IT ist.

2.4 Grenzen des Wachstums

Die vier größten IT-Firmen der Welt, Microsoft, Google, Apple und Facebook, können auf ihren Kerngeschäft nicht mehr so weiter wachsen.

Das ist nur noch im Kerngeschäft der Konkurrenz möglich, denn die Nutzerzahlen steigen nicht so stark wie erhofft – also wildern sie in den Revieren der anderen und versuchen, mit anderen Ansätzen Vorteile am Markt zu erlangen. Waren früher Steigerungsraten mit „Enterprise“-Lösungen zu erzielen, hat man jetzt die „Konsumerisierung“ entdeckt, nach der Losung „Kleinvieh macht auch Mist“.[3]

Der Nutzer in der Rolle des „kreativen Individualisten“ will sich nicht mit „Monstern“ wie Lotus Notes oder Sharepoint einlassen – so einfach wie die Technologie des altbewährten USB-Sticks muss für ihn ITK des 21. Jahrhunderts vulgo Cloud Computing funktionieren, allerdings ohne dessen Nachteile.

2.5 Konsumerisierung

Um Daten auf mehreren Computern zur Verfügung zu haben, nutzte man bisher so etwas wie USB-Sticks – aber das Risiko, diese zu vergessen, zu verlieren oder Datenverlust durch technischen Ausfall zu erleiden, möchte man nicht mehr eingehen.

Eine „Online-Festplatte“ als Ersatz wäre gut; die müsste im Wesentlichen zwei Bedingungen erfüllen: Die Daten müssen vor neugierigen Augen so wie auf dem USB-Stick z. B. durch Verschlüsselung geschützt werden. Dennoch sollte man überall und jederzeit Zugriff auf seine Daten erhalten können. Für die iPhone- oder iPad-Welt bieten viele Online-Dienste mehr als der USB-Stick, da diesen Geräten der USB-Anschluss fehlt. Ein weiterer Grund für die „Cloud“-Lösung.

3. http://www.cio.de/schwerpunkt/k/Konsumerisierung.html

Man unterscheidet dabei zwischen Public und Private Cloud, wobei die Hardcore-Cloud-Anbieter bei der Definition einer Private Cloud nicht mitziehen: Sie sind der Meinung, nur eine Public Cloud ist eine richtige Cloud, weil sie per Definition immer durch einen Dienstleister angeboten wird. Im Folgenden soll das Augenmerk hier nur auf relevante Technologien zum Thema „Storage Cloud“ gelegt werden, obwohl man ja bereits die Definitionsphase von „Everything as a Service“ als Zusammenfassung von „Software as a Service“ (SaaS), „Platform as a Service“ (PaaS), „Infrastructure as a Service“ (IaaS) und „Humans as a Service“ (HuaaS) hinter sich gelassen hat.[4]

3. Verschiedene Wolken

3.1 Public Cloud

Amazon und Google sind vermutlich die beiden Namen, die einem spontan einfallen, wenn man an die Public Cloud denkt. Ob Rechenleistung, Datenspeicher oder Software, beide Anbieter liefern dem Kunden, was er braucht. Die Verwaltung obliegt dem jeweiligen Anbieter, so dass der Kunde keinen Mehraufwand zu fürchten hat. Mit E-Mail, Office-Programmen, die paralleles Arbeiten ermöglichen, und Groupware bietet Google auf dem SaaS-Sektor vielschichtige Möglichkeiten für den Endnutzer. Generell dreht es sich bei dem Public-Cloud-Konzept darum, dass der Endnutzer sich nicht um physische Grundlagen wie Hardware, Stellplatz sowie Administration zu kümmern hat, sondern flexibel in der Skalierung ist und in dem Umfang Ressourcen mietet, wie es zum jeweiligen Zeitpunkt nötig ist. Hier liegt mit Sicherheit ein auf den ersten Blick sehr verlockender Punkt für z. B. kleinere Firmen mit überschaubarem Budget. Auf den zweiten Blick muss man jedoch die Kehrseite der Public Cloud sehen und die Aufgaben rund um Datenschutz, Sicherheit und Administrierung in fremde Hände geben.

3.2 Private Cloud

Die Idee der Private Cloud setzt generell auf der Public Cloud auf. Ziel ist es jedoch, die Wolke zu privatisieren bzw. selbst zu konstruieren. Angefangen bei dem physischen Part der Wolke bis hin zum visuellen Layer kann die Installation nach Wunsch gestaltet werden. Dadurch bleiben wichtige Aspekte wie Datenschutz, Wartung und Administrierung direkt in den Händen des Betreibers. Im Kontrast dazu steht die Public Cloud, bei der ledig-

4. http://de.wikipedia.org/wiki/Everything_as_a_Service

lich Ressourcen gemietet werden und keinerlei Kontrolle über Daten, Verfügbarkeit und Sicherheit vorhanden ist.

3.3 Hybrid Cloud

Die hybride Wolke steht für parallele Nutzung von Public und Private Cloud. Organisationen benutzen für kritische, sicherheitsrelevante Aufgaben und Storage ihre Private Cloud und können bei Bedarf weniger sensible Prozesse auf die Public Cloud auslagern. Es wird gezielt die bequeme Skalierung der Public Cloud und die Integrität der Private Cloud ausgenutzt. Ob man dabei aber immer alle Datensicherheitsprobleme im Auge hat, steht auf einem anderen Blatt.

4. Technologien

4.1 Infrastruktur

Untersuchung von Servern und Clientkomponenten: Virtualisierung und Automatisierung: Nun ist ja die Wolke keine neue Erfindung. Es gibt seit vielen Jahren schon bewährte Lösungen, um Dateien auszutauschen und verfügbar machen zu können – aber „früher war alles besser“ ist Unsinn. („Es gab früher Dinge, die waren gut, und die wären es auch heute noch, wenn man die Finger davon gelassen hätte.“[5])

Einfachste und bewährte Lösungen sind seit Jahren z. B.:

- FTP
- WEBDAV
- RSYNC
- AFS
- sshFS

Nicht zu vergessen Dienste wie GIT und Subversion, die das Rückgrat der Code-Entwicklung darstellen. So vielfältig wie der Dschungel der Handy-Tarife, sind auch die Angebote bei Diensten, die Storage Cloud betreffen.

Beispiele für Clienten, die untersucht worden sind: ADrive, Humyo, DropBox und Wuala. Beispiele für Server sind iFolder, Teamdrive, Atmos sowie partiell Eucalyptus.

5. J. Malmsheimer: http://www.youtube.com/watch?v=mRggWhM70Wc

Was wir nicht direkt beobachtet haben, sind Serverdienste wie Azure, Amazon S3 oder „Nischenanbieter“ wie SpiderOak oder SparkleShare. Auf Letzterem ruhen derzeit viele Hoffnungen aus der Open-Source-Community. Die Tests zeigen dabei aber nur eine Momentaufnahme, da sich die Technologien mit dem Mooreschen Gesetz ändern.

4.2 Private-Lösungen

Es wurden folgende Komponenten untersucht:

4.2.1 iFolder

iFolder ist der von Novell entwickelte Kandidat für Windows, Linux und Mac OS X. Ob 32 oder 64 Bit, beides wird unterstützt. Voraussetzung für Windows ist .NET 2.0 + 3.0 und für Mac OS X bzw. Linux die entsprechende Mono-Version. Dies trifft für den Client zu, der Server ist lediglich für Linux (OpenSUSE 11.1, Ubuntu nur mit viel Aufwand) verfügbar.

iFolder ist komplett frei zugänglich und ermöglicht lokale Kontrolle über Benutzerkonten. Im Test fiel uns ein Bug in der Passwortänderung auf. Ein einmal gesetztes Passwort ließ sich nicht mehr ändern!

Ebenso hat der Client Probleme mit dem simultanen bearbeiten von Files und unterstützt kein Versioning – eigentlich ein K.O.-Kriterium.

4.2.2 TeamDrive

TeamDrive, entwickelt von einer deutschen Firma in Hamburg ist auch für die drei oben genannten Betriebssysteme verfügbar, Client wie Server; an einer Version für das iPhone wird gearbeitet.

Als kleines Feature wird der Client für Portable Apps und für USB-Installationen angeboten. Konten müssen jedoch zentral bei TeamDrive angelegt werden, nur dann ist eine Benutzung möglich. TeamDrive legt alle Daten AES-verschlüsselt ab und verkündet, keine Nachschlüssel anzufertigen. Der „digitale Nomade“ ist also von seinen Daten befreit, wenn er den Schlüssel verbummelt hat.

Ein privater Server kann nur eingebunden werden, sofern man die „TeamDrive Personal Lizenz“ für 29,90 EUR erwirbt. Rabatte von 30 % sind angekündigt worden.

Gemeinsamer Vorteil der beiden Applikationen (iFolder wie TeamDrive) ist, dass beide automatisch die Synchronisation durchführen.

Ein Zugriff auf Files via Webinterface ist auch von einem „Fremdrechner“ („Publish-URL“-Funktion in der Professional-Lizenz) aus möglich. Es ist aber nicht möglich, seine Dateien via Webinterface selbst zu managen; das steht wohl noch auf der ToDo-Liste. In der Enterprise-Version verspricht der Anbieter, auch Dateien (keine Folder) via kryptischer URL an Nicht-TeamDrive-Mitglieder ausgeben zu können. Wie die Dateien hier entschlüsselt werden, ist uns nicht so ganz klar.

Das Produkt macht einen ausgereiften Eindruck, auch wenn es unter OS X ab und zu hakt. Nach Dropbox die gefühlt zweitbeste Lösung im Test.

4.2.3 unison

Unison und Capivara sind die einzigen hier getesteten Werkzeuge mit Zugriff auf den Sourcecode. Beide werden privat entwickelt und sind für sämtliche Plattformen verfügbar. Capivara setzt auf Java auf: Man benötigt lediglich die JRE. Unison stellt für die unterschiedlichen Systeme jeweils Clienten zur Verfügung.

Bei beiden basieren die Algorithmen unter anderem auf SSH. Die Oberfläche von Unison ist simpel gehalten und auf das Nötigste reduziert, Capivara hingegen bietet einen an den MidnightCommander angelehnten Dateibrowser mit diversen Funktionen.

Die Backupserver werden bei beiden mit persönlichem Login direkt angesprochen.

Vorteil hier ist, dass man nicht extra Ordner deklarieren muss, die man synchronisiert haben möchte, sondern direkt auswählen kann, was, wie, wann und wo synchronisiert werden soll. Der daraus entstehende Nachteil ist, dass man den Prozess von Hand anschieben muss. Ob man sich da nicht lieber für die Straßenbahn entscheidet?

Unison ist plattformübergreifend verfügbar (Windows, Linux, UNIX, Mac OS X etc.) und kann Änderungen im Dateibestand auf beiden Seiten der Replikas verarbeiten.

Konflikte (gleiche Datei auf beiden Seiten geändert) werden angezeigt, können aber nur manuell gelöst werden.

Durch Nutzung des rsync-Algorithmus, mit dem nur geänderte Blöcke von Dateien übertragen werden müssen, ist eine Bandbreitenoptimierung möglich, sowohl getunnelt als auch via Socket.

Unison ist fehlertolerant, um auch bei Programmabbrüchen oder Netzwerkfehlern einen konsistenten Zustand der Replikas zu bewahren. Ein weiterer

Vorteil ist, dass das Programm im Userlevel arbeitet und so keine Root-/ Administratorenrechte benötigt.

4.2.4 capivara

Wer oft die Inhalte von Verzeichnissen auf den gleichen Stand bringen muss und grafische Programme bevorzugt, könnte bei Capivara, einem SourceForge-Projekt richtig sein. Es muss nicht immer rsync sein. Das betriebssystemunabhängige Java-Programm synchronisiert über eine grafische Oberfläche die Inhalte zweier Verzeichnisse.

Capivara kann mit lokalen Verzeichnissen umgehen und beherrscht auch die Netzwerkprotokolle FTP und SFTP. In einer Ansicht à la MidnightCommander (siehe auch Atmos) wählt man die Verzeichnisse aus und startet bei Bedarf zunächst den Duplikat-Finder, der doppelte Dateien aufspürt und wahlweise löscht oder verschiebt. Zur Sicherheit bietet der Dubletten-Detektiv einen Vorschaumodus, in dem er die gefundenen Dateien auflistet.

Beim Synchronisieren vergleicht das Tool die Dateien nicht nur anhand ihres Zeitstempels, sondern bietet optional auch einen Vergleich anhand der Dateigröße sowie der SHA-1-Prüfsumme. Die Dateimenge lässt sich mit Hilfe von Filtern eingrenzen. Ähnlich wie bei rsync kann man beim Synchronisieren auswählen, ob Capivara die Verzeichnisse identisch befüllen soll oder ob nur neue Dateien den Weg ins Zielverzeichnis finden sollen. Auch beim Synchronisieren gibt es einen Vorschau-Button, über den man die geplanten Aktionen des Tools vorab begutachten kann.

4.2.5 Atmos (EMC)

Atmos ist eine sogenannte Service-Linie von EMC2. Der Storage-Marktführer bietet zum Einem Lösungen an, die sich an die Storage-as-a-Service-Betreiber, also an Dienstleister, richten. Hier kommen vor allem klassische Storage-Lösungen zum Einsatz, aber auch die Lösungen, die das Unternehmen neuerdings im Rahmen der Virtual-Computing-Environment-(VCE)-Koalition zusammen mit Cisco und seiner Tochter VMware geschaffen hat. Alle Produkte zielen auf flexible IT-Infrastrukturen mit geringen Verwaltungs- und Energiekosten ab.

Zum Anderen offeriert EMC2 die Servicelinie „Atmos“, die sowohl für SaaS-Betreiber als auch für Anwenderunternehmen gedacht ist. Die Provider, die Storage-Services an Kunden weiterverkaufen wollen, können mit der Atmos-Lösung Speicherdienste und -anwendungen entwickeln und ausrollen.

Für Geschäftskunden betreibt EMC2 hingegen selbst einen entsprechenden Dienst mit der Typenbezeichnung „Atmos onLine“. Hier stellt der Anbieter sowohl Speicher- als auch Rechenleistung online bereit, die Kunden nach Bedarf nutzen können. „Atmos onLine” richtet sich an Kunden, die große Datenmengen bewegen und verwalten wollen, die zudem hohe Anforderungen an Service-Levels, Informationssicherheit und Zugriffsschutz stellen.

EMC2 versichert, dass die Nutzer stets die Kontrolle über Daten behalten und selbst entscheiden, welche Informationen wo gespeichert werden. Allerdings konnte EMC auf Nachfrage keine Standorte in Deutschland nennen.

Zur Bedienung gibt es ein Firefox Add-On – aber keine richtigen Klienten. Das Atmos-Fox genannte Plug-In für den Firefox (und nur für den) bietet Look & Feel eines Norton Commanders der 90er Jahre. Glücklich, wer damit zufrieden ist.

4.2.6 Eucalyptus

Beim „Elastik Utility Computing Architecture for Linking Your Programs To Useful Systems“ (Eucalyptus) handelt es sich um eine Open-Source-Software-Infrastruktur zum Aufbau von skalierbaren Cloud-Computing-Umgebungen in Rechenclustern.

Es wurde als ein Forschungsprojekt am Institut für Computer Science an der University of California Santa Barbara entwickelt und wird mittlerweile von der Eucalyptus Systems Inc. vermarktet, wird aber weiterhin als Open-Source-Projekt gepflegt und weiterentwickelt.

Eukalyptus ist kompatibel mit den Schnittstellen zu Amazon EC2 und S3 (SOAP und REST). Es unterstützt virtuelle Maschinen, die auf einem Xen Hypervisor oder einer KVM ausgeführt werden und hat Administrationstools für die System- und Benutzerverwaltung. Populäre Anwendung von Eukalyptus ist die Ubuntu Enterprise Cloud UEC. Dabei handelt es sich um eine Open-Source-Initiative von Ubuntu, um auf eine einfachere Art und Weise skalierbare Cloud-Infrastrukturen auf Basis von Eucalyptus bereitzustellen und diese zu konfigurieren.

Mit der Ubuntu Enterprise Cloud können Public Clouds erstellt werden, welche Amazon’s EC2 infrastructure nutzen. Mit UEC sollen aber auch Private Clouds entwickelt werden, die auf der eigenen Infrastruktur im eigenen Rechenzentrum hinter der eigenen Firewall gehostet werden.

4.3 Public-Lösungen

Es wurden folgende Komponenten untersucht:

4.3.1 ADrive

Die „ADrive-Online-Harddisk“ gibt es seit 2007. Damit ist ADrive eines der ältesten Angebote. Mit 50 GByte bietet das Unternehmen nach eigenen Angaben das größte kostenlos erhältliche Speichervolumen an, Windows SkyDrive (hier nicht getestet) hat 25 GByte im Portfolio.

Die Funktionen der kostenlosen Version von ADrive sind etwas eingeschränkt. Will man mehr Funktion als nur Webzugriff (etwa webDAV), muss man auf den kostenpflichtigen Dienst umsteigen. Also nur Werbegerassel?

Nach der Anmeldung und Bestätigung per E-Mail kann man über den Browser sofort auf den virtuellen Speicher zugreifen. Für die kostenpflichtigen „Signature“- und „Premium“-Produkte gibt es eine Software (für Windows, Mac und Linux) zum Herunterladen, aber keinen Client für mobile Endgeräte.

Wie die Testergebnisse hier aussahen, ist nicht zu sagen. Die kostenlose Produktlinien kann man nur 14 Tage probieren, sofern man seine Kreditkartennummer hinterlegt hat.

Die Geschwindigkeit beim kostenfreien Angebot hat uns davon abgehalten.

4.3.2 humyo.com

humyo.com: Humyo, mittlerweile eine Trendmicro-Tochter bietet Online-Archivierung und -Backup für bis zu 10 GByte (kostenlose Version) bzw. 100 GByte (Premium-Version) persönlicher Daten auf einem Webserver.

Bei Humyo fehlen die Download-Links für Surfer ohne Account, was wohl der Kundengewinnung dienen soll, für den auf Kooperation ausgelegten Wissenschaftsbetrieb aber eher unhandlich ist.

Dabei erfolgt der Zugriff auf alle gespeicherten Daten via Webbrowser, und laut Hersteller besteht ein maximales Level an Sicherheit im Kontext der Verfügbarkeit durch Nutzung redundanter (gespiegelter) Server.

Der Browser nutzt ein Java-Applet; etwas mehr als Funktionalität als Drag & Drop beim Upload sucht man jedoch vergebens.

4.3.3 Dropbox

Einfacher Online-Backup-Dienst zum Sichern wichtiger Daten auf einem Webserver. Der Nutzer schiebt dabei nach einer Anmeldung die zu sichernden Ordner in den Dropbox-Ordner auf der Festplatte, der alle darin gespeicherten Dateien und Ordner automatisch per SSL in die Wolke synchronisiert. Dropbox bietet Download-Links nur in den Ordnern „Public“ und „Photos“. Die übrigen Daten kann man nur mit anderen Dropbox-Usern teilen, die dann aber auch gleich Schreibrechte bekommen.

Besonders hervorzuheben ist die Möglichkeit, Dateien über mehrere Rechner zu synchronisieren. In der Online-Benutzeroberfläche gibt es dafür eigens eine Funktion „Add a computer“, mit deren Hilfe man weitere Rechner auf die auf Dropbox gesicherten Daten zugreifen lassen kann.

Dropbox nutzt einen Algorithmus zur Datenduplizierung, bei dem der Client jede Datei mittels einer Hash-Summe prüft. Gibt es diese schon einmal in den unendlichen Weiten (auch anderer) Dropbox-Nutzer, gelingt der Upload auch großer Dateien innerhalb von Sekunden, da er nur in der Datenbank verlinkt werden muss. Sind die Dateien hingegen eineindeutig, kann ein Upload von größeren Dateien auch schon mal mehrere Stunden in Anspruch nehmen. Dabei erscheint dann einen Meldung auf dem Desktop, man möge sich derweil einen Schokoriegel gönnen.[6]

Dropbox gibt an, als Backend Amazon S3 zu benutzen und dies mittels REST anzusprechen. Dabei würden alle Daten beim Storage-Cloud-Anbieter verschlüsselt abgelegt. Um diese Schlüsselverwaltung kümmert sich Dropbox allein – ein Nachteil, der in der Wissenschafts-Community als ein Vorteil angesehen wird: Man muss sich nicht mit Zertifikaten herumquälen. Der Nachteil, dass Dropbox damit Zugang zu allen gespeicherten Daten hat, wird geflissentlich übergangen.

Man kann seinen Online-Speicher bei Dropbox erweitern, indem man entweder auf ein kostenpflichtiges Angebot umsteigt (50 GByte oder 100 GByte) oder durch Einladung an Freunde bis zu 10 GByte „geschenkt“ bekommt. Ein Gruppentarif ist mittlerweile ebenso erhältlich wie eine Lösung für Portable Apps. Weiterhin hervorzuheben ist die Möglichkeit, gelöschte Dateien und auch ganze Verzeichnisse per Webinterface einfach wiederherzustellen. Bei der „Pro“-Version ist das für 30 Tage in die Vergangenheit möglich, in der kostenpflichtigen gibt es eine zeitlich unbeschränkte Wiederherstellungsfunktion.

6. http://farm6.static.flickr.com/5300/5425620367c36c90f81c.jpg

4.3.4 Strato HiDrive

Strato HiDrive: In der Eigenwerbung als „geniale Online-Festplatte“ vermarktet, können Strato-Nutzer anderen Daten zur Verfügung stellen, mittlerweile bis zu einer Größe von 2.500 GByte Online-Speicherplatz für Firmen, die mit zehn Admin-Accounts 60 Nutzer bedienen können.

In der Regel bekommt man einen Link mit einem Schlüssel oder einer Session-ID, den man dem Partner per Mail zustellt. So kann HiDrive nur für einzelne Dateien einen Browser-Link erzeugen, für den man dann festlegt, wie oft und wie lange er genutzt werden darf.

Die Protokolle sind dabei vielfältig: Via SMB/CIFS gibt es eine Einbindung als Festplatte, auch über VPN mittels http/https: kann man „mit jedem gängigen Internet-Browser“ auf seine gespeicherten Daten zugreifen. Auch rsync und FTP sind machbar. Nachteil bei Strato: Man „kauft die Katze im Sack“ – ein Test vorab ist kostenfrei nicht möglich.

4.3.5 wua.la

WUALA ist ein SPIN-OFF der ETH Zürich und jetzt von Lacie vermarkteter Dienst: Die in Java geschriebene Client-Software läuft unter Windows, Linux sowie Mac OS X und lässt sich über einen Button auf der Wuala-Website mittels Java-Webstart aktivieren, ohne extra einen Client installiert haben zu müssen. Um Dateien online zu speichern, schiebt man sie einfach in den Ordner „Meine Dateien“ des Wuala-Clients. Einzelne Unterordner lassen sich für alle Wuala-Nutzer oder selbst angelegte Benutzergruppen freigeben. Der Client stellt den Online-Speicher auch als lokale Netzwerkfreigabe bereit, so dass Anwendungen Dateien daraus direkt öffnen, bearbeiten und speichern können. Unter Windows erstellt die Software die Freigaben ohne weitere Eingriffe. Ein Ubuntu benötigte jedoch etwas Handarbeit, die eine README im Linux-Paket beschreibt. Im Online-Speicher abgelegte Dateien lassen sich kommentieren und zu einer Favoritenliste hinzufügen. Eine Suchfunktion, eine Kategorisierung nach Dateityp, Benutzernamen oder dem Zeitpunkt der Veröffentlichung sowie selbstgewählte Tags erleichtern die Orientierung.

Der Online-Dienst Wuala spendiert 1 GByte Speicherplatz, den man gegen Geld oder im Tausch gegen eigenen lokalen Plattenplatz erweitern kann („Pro“-User).

Bei den kostenlosen Accounts mit 1 GByte Speicher blendet der Client Werbung ein. „Pro”-User erhalten gegen Geld oder im Tausch gegen lokalen Plattenplatz mehr Online-Speicher. Den lokalen Speicher bindet Wuala über

Peer-to-Peer-Technologie in seinen Speicherdienst ein, wenn der Rechner wenigstens vier Stunden pro Tag aus dem Internet erreichbar ist. Dabei werden die Daten clientseitig AES-verschlüsselt [10].

5. Fazit

5.1 Best of Breed

Die hier betrachtete „Storage Cloud“ wird durch Produkte wie Dropbox oder TeamDrive real erlebbar.

Einfach zu benutzen, für viele Plattformen (auch mobile) verfügbar, hat sich Dropbox zum Liebling der digitalen Wissenschafts-Nomaden aufgeschwungen. Diese Usability wird durch keinen anderen Klienten erreicht und ist wohl auch der Hauptgrund für die pandemische Ausbreitung.

Verfügbar für iPhone & iPad, Blackberry und Android (hier auch PGP-Verschlüsselung möglich) ist dieser Dienst mittlerweile das Produkt der Wahl; andere versprechen immer nur, Dropbox kann.

Durch die Einbindung ins Filesystem kann auf jedem benutzten Rechner eine Indexierung stattfinden, so dass man nicht ewig nach Daten suchen muss. Auch ein Anwendungsszenario für ein TrueCrypt-Image ist mittlerweile vorhanden und erprobt. Hier muss man allerdings auf die Indizierung verzichten.

Durch eine geschickte Versionsverwaltung werden Konflikte bei verschiedenen Dateiarten (die auch folderbasiert sind wie z. B. Apples Pages) korrekt behandelt: Die alte Datei wird als „deleted“ markiert und kann zur Not im Webinterface der Dropbox wiederhergestellt werden. Ein weiteres Alleinstellungsmerkmal ist die Möglichkeit, nur für bestimmte Ordner eine Synchronisation auszuwählen, hiermit schützt man seine mobilen Geräte vor einem Filesystem-Überlauf.

5.2 Datenschutz

Anwender, die Datenschutzgesetze zu beachten haben, und dazu zählen wir die digitalen Wissenschafts-Nomaden der MPG, sind aufgefordert, zu überdenken, ob sie Dienste wie Dropbox weiterhin sorglos nutzen können.

Bei Amazon ist in den AGB zu lesen, dass man die Wahl hätte, ob sich Daten auf deren Online-Festplatte innerhalb der EU aufhalten sollen. Im Zweifelsfall wären wir da nicht so sicher (siehe Patriot-Act).

Wenn es also um die private Storage Cloud geht, bleiben nur noch Lösungen wie ein institutseigenes „TeamDrive“ oder ein Projekt unter Eucalyptus.

Literatur

[1] Tom DeMarco: „Wien wartet auf Dich!“, Hanser Fachbuch, 1999

[2] LIFE 2:
http://www.studie-life.de

[3] Mark Hinkle: „Die 11 besten Open Source Cloud Computing Projekte 2009“:
http://tinyurl.com/ypwc34

[4] wikipedia: „Next Generation Network”:
http://de.wikipedia.org/wiki/NextG enerationN etwork

[5] Cloud Computing in der öffentlichen Verwaltung:
http://tinyurl.com/5uk4y5j

[6] http://www.allthingsdistributed.com/

[7] EMCQR Atmos Conceptual Overview:
http://tinyurl.com/48dyyh8

[8] http://open.eucalyptus.com/

[9] Dr. Fabius Klemm:
http://25dvt.bgc-jena.mpg.de/Z/Abstracts/Klemm.html

[10] Dr. Fabius Klemm: „P2P Storage in the Cloud“:
http://tinyurl.com/6yns8vj

[11] Hansjörg Küster: „Das ist Ökologie“. Beck, 2005

[12] http://de.wikipedia.org/wiki/USA_PATRIOT _Act

Beweissicherheit und Archivierung von Forschungsdaten in der MPG

Jan Potthoff

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Paul C. Johannes

Universität Kassel

1. Einleitung

In der Max-Planck-Gesellschaft muss jeder Mitarbeiter bei Neuanstellung mit seiner Unterschrift bestätigen, dass er die Regeln zur Sicherung guter wissenschaftlicher Praxis zur Kenntnis genommen hat. Die Regelung basiert auf den Empfehlungen der Deutschen Forschungsgemeinschaft vom Januar 1998 [DFG]. Für den Nachweis erbrachter Leistungen wird darin eine Sicherung der Forschungsprimärdaten für einen Zeitraum von mindestens zehn Jahren gefordert. Zu archivieren sind sowohl die Primärdaten, die für eine Veröffentlichung genutzt wurden, als auch die Protokollierung der Arbeitsschritte in einer nachvollziehbaren Form.

In der Praxis werden die im Forschungsprozess durchgeführten Arbeitsschritte und deren Ergebnisse in einem Laborbuch (auch Laborjournal, Pro-

tokoll-, Notiz- oder Tagebuch genannt) dokumentiert. Unabhängig von der Bezeichnung bilden Laborbücher nach Ebel, Bliefert und Greulich (2006, S. 16) die Keimzelle der naturwissenschaftlichen Literatur, auf welchen Zwischenberichte, Berichte und Publikationen aller Art aufbauen.

Durch elektronische Laborgeräte und automatisierte Laborprozesse entstehen im Forschungsprozess in zunehmendem Maße digitale Daten, z. B. produzieren Versuche am Windkanal am Max-Planck-Institut für Dynamik und Selbstorganisation in Göttingen digitale Bilddaten im Terabyte-Bereich. Daher ist eine Dokumentation mithilfe von herkömmlichen Laborbüchern in Papierform nicht länger realisierbar.

Die Daten, z. B. die erzeugten Bilder in einer Versuchsreihe im Windkanal, werden nicht in das papiergebundene Laborbuch übernommen, sondern existieren daneben in digitaler Form. Somit entstehen hybride Laborbücher, d. h. eine Dokumentation, die teilweise auf Papier und teilweise am Computer durchgeführt wird. Diese Hybriden führen zu Problemen sowohl bei der Gewährleistung der Vollständigkeit der Daten als auch hinsichtlich der langfristigen Interpretierbarkeit, Authentizität und Integrität.

Um weiterhin eine langfristige Interpretierbarkeit und die Sicherstellung der Datenauthentizität und -integrität, wie sie in den Regeln der guten wissenschaftlichen Praxis gefordert werden, zu gewährleisten, bedarf es für eine rein elektronische oder hybride Dokumentationsweise neuer Konzepte. In diesem Zusammenhang erarbeiten die Universität Kassel, die Physikalisch Technische Bundesanstalt in Braunschweig und die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen im Rahmen des DFG-geförderten Verbundprojektes „Beweissicheres elektronisches Laborbuch“ (BeLab), eine entsprechendes Konzept, mit dem eine beweiswerterhaltende Langzeitarchivierung von digitalen Forschungsprimärdaten und dazugehörigen Metadaten erreicht werden kann und zu dem die Vollständigkeit der im Forschungsprozess entstandenen Daten erhalten bleibt.

2. Ablauf eines Forschungsprozesses und dessen Dokumentation

Zur Analyse der Forschungspraxis wurden im Rahmen des BeLab-Projekts u. a. Gespräche mit Wissenschaftlern und IT-Mitarbeitern aus Max-Planck-Instituten durchgeführt. Es zeigte sich, bezogen auf den Wissenschaftler, eine individuelle Dokumentationsweise des wissenschaftlichen Forschungsprozesses. Dies ist nicht nur in den unterschiedlichen Forschungsrichtungen zu begründen, sondern auch in der Organisationsweise des Forschers selbst.

Grob lässt sich der Forschungsprozess in fünf Phasen einteilen:

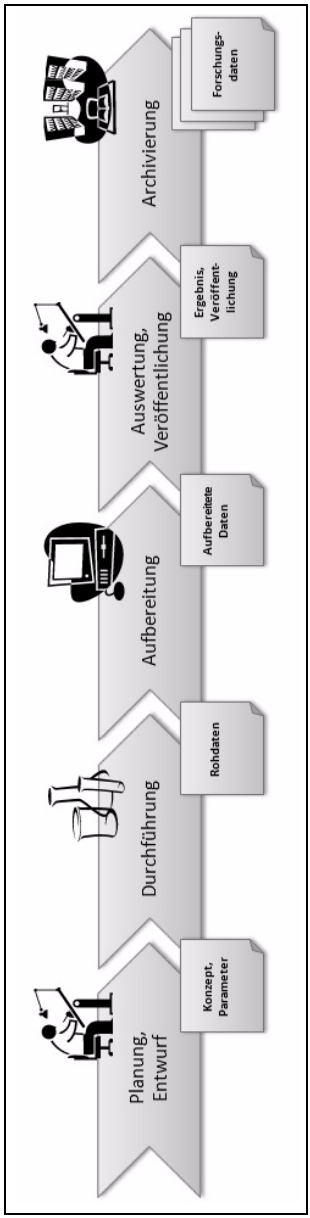

Abb. 1: Phasen des Forschungsprozesses

In der Planungsphase werden eine Literaturrecherche durchgeführt, das Experiment definiert und Parameter festgelegt. Die anschließende Durchführung, Aufbereitung und Auswertung des Versuchs ist abhängig von der jeweiligen Forschungsrichtung. Im Allgemeinen werden Messgeräte verwendet, deren Datenoutput (zum größten Teil in digitaler Form) in der Datenmenge stark variiert. In den durchgeführten Gesprächen wurde der Umfang der sogenannten Rohdaten eines Experiments von einigen Megabyte bis einigen Gigabyte genannt. Teilweise bereiten Messgeräte die Rohdaten eigenständig auf, bevor sie ausgegeben werden, meist erfolgt jedoch die Aufbereitung in einem gesonderten Prozess durch den Forscher. Anschließend werden die so vorverarbeiteten Daten ausgewertet. Die am Windkanal des Max-Planck-Instituts für Dynamik und Selbstorganisation aufgezeichneten Bilder werden beispielsweise zur Analyse von Turbulenzen durch High Performance Computer mittels Simulationsberechnungen aufbereitet. Nach einer Veröffentlichung oder der Dokumentation der Ergebnisse erfolgt abschließend die Archivierung. Damit stehen Daten und Ergebnisse für eine spätere Nachnutzung oder zur Darlegung von Forschungsergebnissen weiterhin zur Verfügung.

In den durchgeführten Gesprächen mit Wissenschaftlern der Max-Planck-Institute konnte festgestellt werden, dass das klassische Laborbuch in Papierform weiterhin Anwendung findet. Als großer Vorteil gegenüber der elektronischen Erfassung wurde die flexible Dokumentationsweise, wie das schnelle Schreiben von Anmerkungen, genannt. Sowohl bedingt durch den digitalen Output als auch durch den Einsatz von Applikationen, die zur Weiterverarbeitung genutzt werden, werden im Forschungsprozess zusätzlich digitale Daten erzeugt. Das Laborbuch in Papierform kann daher nicht ausschließlich verwendet werden. In diesem Fall ist es üblich, die Daten in einer selbstdefinierten Verzeichnisstruktur abzulegen und mit einem entsprechenden Verweis in das Laborbuch einzutragen. Als problematisch wurde in diesem Zusammenhang die selbstdefinierte Vorgehensweise genannt, die dazu führt, dass nach dem Ausscheiden von Mitarbeitern Daten verloren gehen können, da Daten auf nur für den Mitarbeiter zugänglichen Bereichen gesichert wurden.

Im Max-Planck-Institut für biophysikalische Chemie in Göttingen wurde im Gespräch mit Wissenschaftlern eine rein elektronische Dokumentationsweise vorgestellt. Dabei handelte es sich um eine für den chemischen Bereich ausgelegte Datenbank, die u. a. das Suchen von chemischen Strukturen erlaubt. Zur freieren Dokumentation dienen Ausdrucke, auf denen Anmerkungen schnell vermerkt werden können. Anschließend werden evtl. aufgenommene Anmerkungen in das System übertragen. Der Mehrwert der

rein elektronischen Dokumentation wurde insbesondere durch das schnelle Auffinden von bereits durchgeführten Experimenten und deren Ergebnissen sichtbar.

In anderen Abteilungen des Max-Planck-Instituts für biophysikalische Chemie erfolgte eine elektronische Dokumentation auf der Grundlage unterschiedlichster Systeme, wie beispielsweise Microsoft-Office-Anwendungen oder Wiki-Systemen. Generell wurde der Wunsch nach einer einheitlichen elektronischen Dokumentationsform deutlich.

Ein Vorteil der rein elektronischen Dokumentation, z. B. mithilfe eines elektronischen Laborbuchs, ist die zentrale Datenhaltung. Daten, Metadaten und die Dokumentation können zentral gepflegt und für andere Forscher freigegeben werden. Des Weiteren kann durch die Anpassung der Dokumentationsweise an den Forschungsprozess eine Effizienzsteigerung erzielt werden.

3. Archivierung der Forschungsdaten

Die Archivierung und Bereitstellung von Forschungsdaten dient nicht nur zur Nachnutzung wissenschaftlicher Ergebnisse, sondern auch zur Darlegung von Forschungsergebnissen. Dies kann jedoch nicht immer gewährleistet werden. So wurde neben der Form der Speicherung der Daten in selbstdefinierten Verzeichnisstrukturen auf lokalen Rechnern auch von Wechseldatenträgern berichtet, die während Experimenten, die außerhalb des Instituts durchgeführt wurden, Verwendung fanden.

Im Falle der vorgestellten rein elektronischen Dokumentation am Max-Planck-Institut für biophysikalische Chemie wurden die Daten neben der elektronischen Archivierung/Sicherung zusätzlich in Papierform archiviert, um so einen höheren Beweiswert zu erzielen.

3.1 Interpretierbarkeit der Daten

Im Rahmen der Langzeitarchivierung wird von einer möglichst langlebigen Interpretierbarkeit der Daten gesprochen. Trotz der Ungewissheit, welche Datenformate langfristig interpretierbar bleiben, existieren nach Ludwig (2010) Kriterien, nach denen Formate hinsichtlich der Eignung für eine Langzeitarchivierung bewertet werden können. Frei verfügbare Formatspezifikationen, das Einhalten von Standards und die Angabe von Metadaten im Dateiformat fördern beispielweise eine langfristige Interpretierbarkeit. Die Komplexität eines Datenformats und evtl. verwendeter Schutzmechanismen, wie Kopierschutz oder Verschlüsselung, erschweren dagegen die Interpretationsfähigkeit.

Wie auch in den Gesprächen mit Wissenschaftlern der Max-Planck-Institute festgestellt wurde, liegt die Verantwortung einer möglichst langlebigen Interpretierbarkeit der Daten auf Seiten des Datenerzeugers. So wurde teilweise die verwendete Programmversion mit den Daten abgelegt oder die Versionsnummer als Metadatum erfasst. Des Weiteren wurde über den Einsatz von MD5-Summen zur Überprüfung der Datenintegrität berichtet. Die Verantwortung über die Verwendung der Checksummen lag auch in diesem Fall beim Wissenschaftler.

3.2 Strukturierung durch Metadaten

Metadaten werden nicht nur als Interpretationshilfe des Datenformats verwendet, sondern dienen zusätzlich zum Auffinden im Archiv abgelegter Daten. Welche Metadaten für eine spätere Suche relevant sind, ist anwendungsbezogen. In jedem Fall liegt die Verantwortung für die Angabe der Metadaten zum Auffinden gewünschter Daten beim Wissenschaftler. Zusätzlich können automatisch erfassbare Metadaten gesichert werden, um die Eingabe der Metadaten zu vereinfachen. Elektronische Laborbücher bieten z. B. die Möglichkeit, Metadaten wie Person, Datum, letzte Änderung etc. automatisch zu erfassen. Nach Borghoff (2003) sollte dabei beachtet werden, dass das Metadatenformat auf Standards, wie z. B. Dublin Core, METS oder MARC 21, basiert, um einen Metadatenaustausch gewährleisten zu können .

Im Falle der rein elektronischen Dokumentation am Max-Planck-Institut für biophysikalische Chemie wurden durch die Software individuell zusammenstellbare Metadaten zentral aufgenommen und gepflegt. In anderen Fällen wurden Metadaten in separaten Dateien dokumentiert und im Verzeichnis der Daten abgelegt. Eine einheitliche Vorgehensweise zur Eingabe und Pflege von Metadaten wurde nicht festgestellt.

4. Beweiswerterhaltung im Forschungsprozess

Neben der Datenintegrität, die im genannten Fall durch MD5-Summe realisiert wurde, ist die Datenauthentizität aus Sicht der Beweiswerterhaltung von großer Bedeutung. Gerade im digitalen Umfeld sind Manipulationen leicht durchzuführen und ohne besondere Maßnahmen nicht nachzuvollziehen. Gleiches gilt für die Verwendung von Metadaten, um die Authentizität eines Dokuments nachzuweisen. Allgemein kann festgehalten werden, dass durch den Nachweis eines lückenlosen und vollständig geführten Laborbuchs die Richtigkeit von Forschungsdaten und -ergebnissen untermauert werden kann.

Auf das Laborbuch kann es beispielsweise in universitären Disziplinarverfahren, beim Patentrechtsstreit (so z. B. nach Entscheidungen des LG Düsseldorf (30.04.2002) und des Europäischen Patentamts (20.08.1997)), im Zulassungsbereich, beim Urheberrechtsstreit und, nach Tiedemann (2010), bei Verfahren um die Vergabe von Fördergeldern als Beweismittel ankommen.

Im Zivilprozessrecht gelten nach § 371 Zivilprozessordnung (ZPO) elektronische Dokumente nur als Objekte des Augenscheins und unterliegen der freien Beweiswürdigung durch das Gericht. Im Beweisrecht ist jedoch der Urkundsbeweis nach §§ 415 ff. ZPO das stärkste Beweismittel. Die Privaturkunde erbringt den Beweis dafür, dass die darin enthaltenen Erklärungen durch den Aussteller abgegeben wurden. Allein aufgrund dieser Beweiswertungsregel ist es sehr vorteilhaft, in einem Gerichtsverfahren auf die Urkunde als Beweismittel zurückgreifen zu können.

Während herkömmliche Laborbücher ohne weiteres als Urkunden gewertet werden können, sind für elektronische Laborbücher besondere Maßnahmen zu treffen.

5. Lösungsmöglichkeiten im BeLab-Projekt

Um die Datenauthentizität und -integrität von elektronischen Daten mit der Archivierung überprüfbar zu halten, können fortgeschrittene oder qualifizierte elektronische Signaturen nach dem Signaturgesetz (SigG) verwendet werden.

Sind elektronische Dokumente mit einer qualifizierten elektronischen Signatur versehen, sind sie nach § 371a ZPO als Urkunde zu werten. Nur eine qualifizierte Signatur ersetzt nach den Vorstellungen des Gesetzgebers die eigenhändige Unterschrift. Daher können qualifizierte elektronische Signaturen zum Erhalt der Beweiskraft in Verbindung mit einer elektronischen Dokumentationsweise verwendet werden.

Soll primär die Integrität erhobener Rohdaten nachweisbar sein, kann möglicherweise auf die qualifizierte elektronische Signatur verzichtet und durch eine meist kostengünstigere, fortgeschrittene Signatur ersetzt werden. Des Weiteren besitzen Messgeräte teilweise bereits Signaturkomponenten, mit denen das Signieren der Rohdaten im Gerät selbst durchgeführt werden kann und somit die Integrität über den gesamten wissenschaftlichen Forschungsprozess (von der Datenerhebung bis zur Archivierung) überprüfbar ist.

Neben den (qualifizierten oder fortgeschrittenen) Signaturen können elektronische Zeitstempel zur Steigerung des Beweiswertes eingesetzt werden. Sie

ermöglichen eine authentische und unverfälschbare Verknüpfung von Daten mit einer Zeitaussage. Damit kann ein unveränderter Zustand eines elektronischen Dokumentes von einem bestimmten Zeitpunkt an nachgewiesen werden. Zeitaussagen sind im Forschungsprozess an mehreren Stellen von Bedeutung, z. B. beim Experiment selbst, beim Authentifizierungsprozess oder bei Änderungen an den Daten.

Aufbauend auf den im Vorfeld beschriebenen etablierten (kryptographischen) Verfahren wird im Rahmen des BeLab-Projekts ein Konzept für eine beweissichere Langzeitarchivierung von Forschungsprimärdaten entwikkelt. Die Implementierung eines Prototypen (im folgenden BeLab-System genannt) auf der Basis des entworfenen Konzepts wird sowohl aus technischer als auch juristischer Sicht beurteilt.

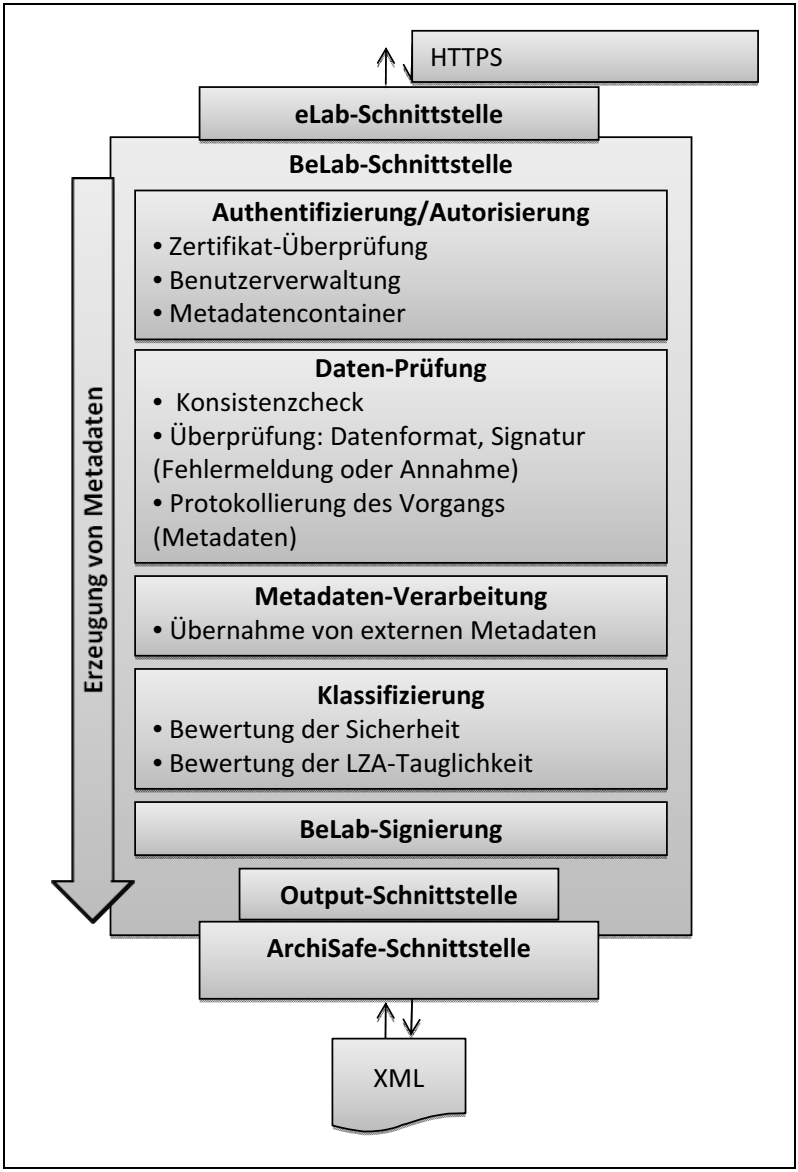

Abb. 2: Arbeitsweise des BeLab-Systems

Die an das BeLab-System übergebenen Daten werden in mehreren Schritten verarbeitet (s. Abb. 2). Eine im Vorfeld verwendete Signatur, beispielsweise

durch ein signierendes Messgerät, wird im ersten Schritt auf Gültigkeit überprüft. Da im Forschungsprozess in der Regel mehrere zu einem Experiment gehörige Dateien entstehen, können Datensammlungen an das BeLab-System übergeben werden, die durch das System auf Konsistenz geprüft werden. Des Weiteren wird eine Aussage über das verwendete Datenformat der übermittelten Daten bzgl. der Eignung zur Langzeitarchivierung getroffen. Basierend auf der technischen Richtlinie des Bundesamtes für Sicherheit in der Informationstechnik [BSI] TR-03125 erfolgt abschließend die Archivierung der Forschungsdaten. Durch das ArchiSafe-Modul in Kombination mit dem zur BSI TR-03125 gehörigem ArchiSig- und Krypto-Modul wird eine langfristige Archivierung inklusive der eingesetzten Signaturen sichergestellt. BSI TR-03125, ArchiSafe und ArchiSig übernehmen die Spezifikationen der RFC 4810 „Long-Term Archive Service Requirements“ und RFC 4998 „Evidence Record Syntax“ der Internet Engineering Task Force [IETF] und basieren damit auf international anerkannten Standards.

6. Fazit und Ausblick

Wie in den Gesprächen mit Wissenschaftlern und IT-Mitarbeitern aus Max-Planck-Instituten bestätigt wurde, liegen Forschungs- und Messdaten in zunehmendem Maße digital vor. Eine reine Dokumentation mithilfe eines papiergebundenen Laborbuchs ist daher nur noch selten möglich. Daher wird zum einen eine (teilweise) elektronische Dokumentation und in jedem Fall eine digitale Archivierung notwendig, zum anderen können Vorgehensweisen, die zur Überprüfung der Authentizität und Integrität in einem Laborbuch in Papierform etabliert waren, nicht konsequent eingesetzt werden.

Da die Dokumentationsweise des Forschers und die Anforderungen an ein elektronisches Laborbuch im starken Maße vom Forschungsbereich und dem Wissenschaftler selbst abhängig sind, ist ein generischer Ansatz zur Gewährleistung einer beweissicheren Langzeitarchivierung von Forschungsprimärdaten sinnvoll. Des Weiteren muss der Forschungsprozess in allen Stufen betrachtet werden, um einen möglichst hohen Beweiswert zu erzielen.

So wird im Rahmen des BeLab-Projekts ein Konzept, basierend auf fortgeschrittenen und qualifizierten Signaturen und Zeitstempeln, entworfen, das die ganzheitliche Beweiswerterhaltung (vom Messgerät bis zur Archivierung) gewährleisten soll. Damit soll es zukünftig möglich sein, das elektronische Laborbuch äquivalent zu herkömmlichen Laborbüchern als Beweismittel verwenden zu können.

7. Referenzen

Borghoff, U. M., Rödig, P., Scheffczyk, J. & Schmitz, L. (2003). Langzeitarchivierung – Methoden zur Erhaltung digitaler Dokumente. Heidelberg : dpunkt.verlag.

Bundesamtes für Sicherheit in der Informationstechnik [BSI]. (2009). Technische Richtlinie 03125 – Vertrauenswürdige elektronische Langzeitspeicherung, zu finden unter https://www.bsi.bund.de/ContentBSI/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html - Stand vom 23.12.2010.

Deutsche Forschungsgemeinschaft [DFG]. (1998). Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ – Vorschläge zur Sicherung guter wissenschaftlicher Praxis, zu finden unter http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf - Stand vom 12.01.2011.

Ebel, H. F., Bliefert, C., Greulich, W. (2006). Schreiben und Publizieren in den Naturwissenschaften (5. Aufl.). Weinheim : Wiley-VCH.

Europäisches Patentamt [EPA]. Entscheidung vom 20.08.1997, Az. T0886/1993.

Internet Engineering Task Force [IETF] (2007). RFC 4810. Long-Term Archive Service Requirements. Wallace, C., Pordesch, U. & Brandner. R., zu finden unter http://datatracker.ietf.org/doc/rfc4810/ - Stand vom 23.12.2010.

Internet Engineering Task Force [IETF] (2008). RFC 4998. Evidence Record Syntax. Gondrom, T., Brandner. R. & Pordesch, U., zu finden unter http://datatracker.ietf.org/doc/rfc4998/ - Stand vom 23.12.2010.

LG Düsseldorf. Urteil vom 30.04.2002, Az. 4aO95/01.

Ludwig, J. (2010). Formate – Auswahlkriterien. In Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S. & Jehn, M. (Hrsg.), Nestor Handbuch – Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Version 2.3) (Kap.7.3), zu finden unter http://nestor.sub.uni-goettingen.de/handbuch/ - Stand vom 23.12.2010.

Signaturgesetz (SigG) vom 16. Mai 2001 (BGBl. I S. 876), das zuletzt durch Artikel 4 des Gesetzes vom 17. Juli 2009 (BGBl. I S. 2091) geändert worden ist.

Tiedemann, P. (2010), Entzug des Doktorgrades bei wissenschaftlicher Unlauterkeit, Zeitschrift für Rechtspolitik, 2, 53-55.

Zivilprozessordnung (ZPO) in der Fassung der Bekanntmachung vom 5. Dezember 2005 (BGBl. I S. 3202; 2006 I S. 431; 2007 I S. 1781), die zuletzt durch Artikel 3 des Gesetzes vom 24. September 2009 (BGBl. I S. 3145) geändert worden ist.

Langzeitarchivierung am MPI für biophysikalische Chemie – ein simples Verfahren

Petra Küster

Max-Planck-Institut für biophysikalische Chemie, Göttingen

Zusammenfassung

Die Regeln zur Sicherung guter wissenschaftlicher Praxis verlangen die sichere Archivierung von Daten, die Grundlage von Publikationen sind, für mindestens zehn Jahre. Um dies erfüllen zu können, wurde 2001 am MPI für biophysikalische Chemie ein simples Verfahren entwickelt. Die Arbeitsgruppen kopieren die Dateien auf einen dedizierten Server. Über eine Webschnittstelle geben sie Metadaten ein und stoßen die Archivierung an. Die Dateien werden in TAR-Dateien von maximal 2 GByte gepackt und zusammen mit der TAR-Log-Datei und einer Datei, die die Metadaten und MD5-Checksums der Dateien enthält, ins TSM-Archiv auf zwei Sätzen von Bändern geschrieben, die an unterschiedlichen Standorten stehen.

1. Motivation

Am 24.11.2000 beschloss der Senat der MPG die „Regeln zur Sicherung guter wissenschaftlicher Praxis“, die folgende Anforderung enthält:

„Primärdaten als Grundlagen für Veröffentlichungen müssen auf haltbaren und gesicherten Trägern in den Instituten oder Forschungseinrichtungen, wo sie entstanden sind, für mindestens zehn Jahre aufbewahrt werden, sofern dies möglich ist. Für berechtigte Interessenten muss der Zugang zu den Daten gewährleistet sein.“

Der EDV-Gruppe des Instituts war klar, dass man die Erfüllung dieser Regel nicht den einzelnen Arbeitsgruppen überlassen kann, sondern eine zentrale Lösung anbieten muss.

Da viele Forschungsgruppen nach einigen Jahren das Institut wieder verlassen, ist nur über eine zentrale Lösung sichergestellt, dass die Primärdaten für zehn Jahre im Institut bleiben und zugänglich sind.

Diese Auffassung wird auch vom Senat der MPG vertreten. In einer aktualisierten Version vom März 2009[1] findet man folgende Ergänzung:

„Es muss entweder vom Institut oder zentral sichergestellt werden, dass Daten zumindest für diesen Zeitraum lesbar verfügbar bleiben.“

Welche Möglichkeiten hatten die Arbeitsgruppen bisher, ihre Daten über einen längeren Zeitraum zu sichern?

1. Externe Speichermedien

Häufig wurden die Daten auf externe Speichermedien kopiert und in den Schrank gestellt. Als Speichermedien wurden CDs, DVDs oder externe Festplatten verwendet. Zum einen ist fraglich, ob diese Medien nach zehn Jahren noch lesbar sind, zum anderen ist die Datenintegrität nicht sichergestellt, Medien können ausgetauscht oder Dateien manipuliert werden. Außerdem kann man kaum sicherstellen, dass diese Medien nach mehreren Jahren auch auffindbar sind.

2. Archivserver der GWDG

Die GWDG betreibt einen Archivserver auf Basis von UniTree, auf den alle Mitarbeiter Dateien ablegen können. Der Archivserver kann aber mit vielen kleinen Dateien nicht arbeiten, die Arbeitsgruppen müssen also angehalten werden, die Verzeichnisse zusammenzufassen. Ob dies geschieht und ob die Dateien in zehn Jahren noch entpackt werden können, ist dabei nicht sichergestellt. Der Benutzer kann seine Dateien im Archiv löschen, und es gibt

1. http://www.mpg.de/229457/Regeln_guter_wiss_Praxis__Volltext-Dokument_.pdf

wieder das Problem der Auffindbarkeit, unter welcher Nutzerkennung wo die Primärdaten zu einer Publikation abgespeichert sind.

Neben der notwendigen Archivierung von Primärdaten von Publikationen gab es im Institut auch den Wunsch, Daten von Benutzern zu archivieren, die das Institut verlassen haben.

2. Anforderungen

Es wurden folgende Anforderungen an ein System zur Langzeitarchivierung gestellt:

1. Speichermedien müssen nach zehn Jahren garantiert lesbar sein.

   Dazu muss es möglich sein, im Hintergrund einen Medienwechsel durchführen zu können.

2. Datenintegrität muss sichergestellt sein.

   Es darf nicht möglich sein, Daten zu manipulieren. Ein Benutzer darf nicht die Möglichkeit haben, seine archivierten Daten zu löschen.

3. Die Eingabe von Metadaten muss möglich sein.

   Über Metadaten wird das Auffinden von Dateien erleichtert.

4. Das Archivierungssystem muss einfach zu bedienen sein, damit es von den Arbeitsgruppen angenommen wird.

Was solch ein Archivierungssystem nicht leisten kann, ist sicherzustellen, dass die Daten in Formaten abgespeichert werden, die auch in zehn Jahren noch lesbar sind. Dies liegt in der Verantwortung des Nutzers.

Für den Benutzer sollte folgender Workflow realisiert werden:

1. Er stellt die Daten in einem vorgegebenen Bereich zusammen.
2. Über ein Web-Interface gibt er die Metadaten ein.
3. Mit dem Abschicken des Formulars wird der Archivierungsprozess gestartet.
4. Der Benutzer bekommt Rückmeldung per E-Mail, wenn die Archivierung erfolgreich abgeschlossen ist.

3. Realisierung

3.1 Speichermedien

Wir haben uns entschieden, für die Speicherung der Daten Tivoli TSM im Archiv-Modus zu nutzen mit einer Policy, dass die Dateien elf Jahre aufgehoben und zwei Kopien angefertigt werden. Eine Tape-Library steht im Maschinenraum der GWDG am Faßberg, die andere im Rechenzentrum der Universitätsmedizin Göttingen 2 - 3 km entfernt.

Ein Grund, Tivoli TSM zu verwenden, war, dass die Technologie vorhanden ist und genügend Erfahrungen vorliegen, um sicherzustellen, dass Dateien auch in zehn Jahren noch restauriert werden können. Ein Wechsel auf neue Speichermedien ist im Backend möglich.

3.2 Zusammenstellen der Daten

Es gibt einen zentralen Rechner `archive.mpibpc.mpg.de` mit Verzeichnissen für die verschiedenen Gruppen. Der Datentransfer erfolgt mit SMB oder NFS (früher wurde auch NETATALK und FTP angeboten). Durch ACLs ist sichergestellt, dass die Arbeitsgruppen nur in ihrem eigenen Bereich lesen und schreiben. Der Zugriff ist nur für einzelne Personen der Arbeitsgruppe, den Arbeitsgruppenleiter und/oder den Abteilungsadministrator freigeschaltet.

Es wird empfohlen, die Daten einer Publikation in ein Unterverzeichnis zu kopieren, das einer bestimmten Namenskonvention genügt. Somit sind die wichtigsten Informationen schon in dem Verzeichnisnamen enthalten. Außerdem wird eine Datei `readme.txt` gefordert, aber nicht überprüft, ob sie sinnvolle Inhalte enthält.

3.3 Web-Frontend

Mit dem Web-Frontend werden einige Metadaten eingegeben und die Archivierung angestoßen. Es gibt zwei unterschiedliche Eingabemasken: eine für die Archivierung von Publikationsdaten, eine für die Archivierung von Daten eines Benutzers.

Abb. 1: Frontend, Auswahlmenü

Abb. 2: Archivierung von Primärdaten von Publikationen

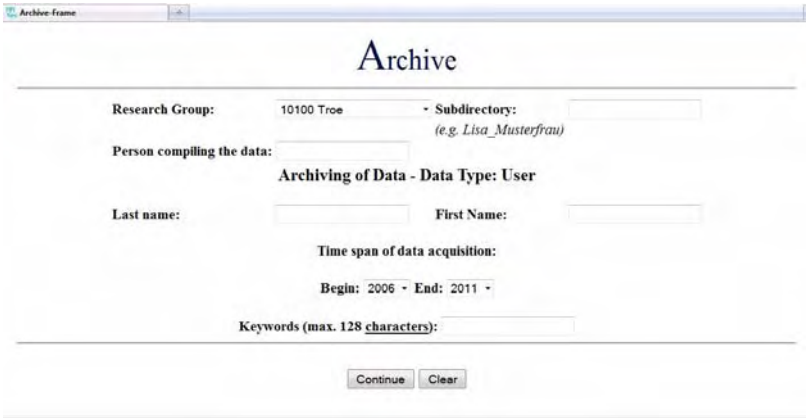

Abb. 3: Archivierung von Daten ausgeschiedener Benutzer

Bevor der Benutzer das Formular aufruft, muss er sich anmelden. Aus dem Login-Namen wird die E-Mail-Adresse des Benutzers ermittelt.

3.4 Archivierungsprozess

1. Nach Betätigung des Submit-Buttons wird ein Job-Auftrag geschrieben. Ein cron-Job schaut nach Aufträgen und startet den Prozess zur Archivierung. So wird sichergestellt, dass nicht mehrere Archivierungsprozesse gleichzeitig laufen.
2. Aus den Eingaben im Frontend wird eine Text-Datei geschrieben.
3. Die abgelegten Dateien werden mit TAR zusammengefasst, mit einer maximalen Größe der TAR-Dateien von 2 GByte.
4. Der TAR-Log wird gezippt. Größe und MD5-Checksum von jeder TAR-Datei werden an die Text-Datei mit den Metadaten angehängt.
5. TAR-Dateien, TAR-Log, Info-Datei und `readme.txt` werden in einem Standardverzeichnis zur Archivierung abgelegt. Die Dateien werden folgendermaßen benannt:

TAR-Files	`abt_verzeichnisname_timestamp.tar#`
TAR-Log	`abt_verzeichnisname_timestamp.tar.log`
Info-Datei	`abt_verzeichnisname_timestamp.txt`
Readme.txt	`abt_verzeichnisname_timestamp.readme`

   Hierbei ist „`abt`“ das Kürzel der Abteilung.

6. Die Archivierung wird mit dem Befehl `dsmc` von Tivoli TSM gestartet.
7. Der Archivierer bekommt zwei E-Mails. Zuerst wird eine E-Mail mit dem gezippten TAR-Log versandt, anschließend eine weitere E-Mail, wenn die Archivierung erfolgreich abgeschlossen wurde.

Der `dsmc`-Befehl wird mit dem Parameter „`-deletefiles`“ gestartet, sodass die Dateien nach erfolgreicher Archivierung automatisch gelöscht werden. Die Originaldateien werden dann nach einer Woche gelöscht.

3.5 Suche

Zurzeit wird die Suche in den Archivierungsvorgängen über die Suche in den Dateinamen realisiert. Abhängig von den Anmeldeinformationen werden nur die Vorgänge der zugehörigen Arbeitsgruppe angezeigt.

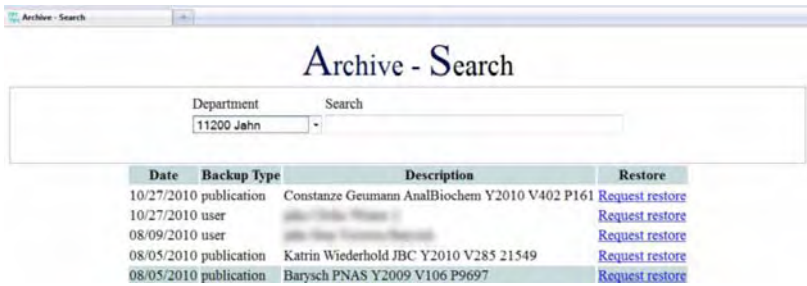

Abb. 4: Screenshot der Suchmaske

Geplant ist der Aufbau einer einfachen MySQL-Datenbank, die die eingegebenen Metadaten und die `readme.txt` enthält, um so eine differenziertere Suche zu ermöglichen. Außerdem soll für einen ausgewählten Archivierungsvorgang die TAR-Log-Datei angezeigt werden.

Das Skript zur Wiederherstellung der Daten kann nur vom Systemadministrator angestoßen werden.

4. Erfahrungen und Probleme

Das im Institut eingeführte Archivierungssystem hat den Vorteil, dass es vom Benutzer einfach zu bedienen ist und mit TSM ein System benutzt wird, mit dem es langjährige positive Erfahrungen gibt.

Ein Problem sind die TAR-Dateien. Archivierungen mit bis zu 400 GByte konnten in einer Session durchgeführt werden, bei der dann 200 TAR-Dateien erzeugt wurden. Der Restaurierungsprozess dauert dann entspre-

chend lang. Für die Archivierung von Primärdaten im TByte-Bereich ist diese Aufteilung in TAR-Dateien nicht mehr sinnvoll.

Große Datenvolumen fallen nur in einigen Arbeitsgruppen an. Diese legen die Primärdaten auf dem Archivserver der GWDG unter einer speziellen Userkennung ab. Wir empfehlen diesen Arbeitsgruppen aber trotzdem, zusätzlich das Archivierungssystem des Instituts zu nutzen, um die Metadaten und die Information, wo die Primärdaten der Publikation zu finden sind, einzugeben.

Das auf dem Server vorhandene Filesystem ist für TSM von großer Bedeutung. Bei einer Änderung des Frontend-Servers muss immer sichergestellt werden, dass die früher archivierten Dateien auch restauriert werden können. Ansonsten gibt es nur die Möglichkeit, skriptbasiert die Dateien in das ursprüngliche Filesystem auszulesen, dann auf das neue Filesystem zu kopieren und anschließend wieder ins Archiv zu schreiben.

Persönliche Mitarbeiterseiten mit Emacs Org-Mode

Stefan Vollmar, Cornelia Weigelt, Michael Sué, Andreas Hüsgen, Roman Krais, Ingo Alt, Timm Wetzel, Alexander Schuster

Max-Planck-Institut für neurologische Forschung, Köln

1. Einleitung

Es gab eine Reihe von Anfragen unserer Nutzer nach „persönlichen Mitarbeiterseiten“ und wir haben eine Möglichkeit gesucht, folgende Punkte zu berücksichtigen: (1) dem Institut dürfen keine rechtlichen Nachteile entstehen (Haftung der Nutzer für eigene Inhalte), (2) möglichst viel Gestaltungsfreiheit: die Mitarbeiter sollen Inhalte innerhalb gewisser Vorgaben selbst gestalten können, (3) Wahrung der „Corporate Identity“ / Integration in den Webauftritt des Instituts, (4) Kompatibilität mit der „Mission“ des Instituts, (5) einfache und unbürokratische Umsetzbarkeit.

2. Methoden

Der zentrale Teil der Umsetzung beruht auf der Nutzung von Plain-Text-Dateien im Org-Mode-Format [1] (s. Abb. 1): Mitarbeiter entwerfen eine Seite, die weitgehend einem formalen Lebenslauf entsprechen soll. Die

Arbeitsgruppenleiterin oder der Arbeitsgruppenleiter hat eine Kontrollfunktion und leitet den Text dann als E-Mail an die Redaktion der Homepage weiter, wenn es keine inhaltlichen Einwände gibt. Die Redaktion kann aus der Text-Datei im Org-Mode-Format automatisch statische HTML-Seiten generieren [2]. Es gibt eine ausführliche Anleitung [3] und eine Reihe von Beispielen, die sich für eigene Zwecke leicht anpassen lassen. Auch wenn es einen institutsinternen Preview-Service gibt und „Power-User“ die HTML-Seiten selber erzeugen können [4], ist unsere Erfahrung, dass fast alle Nutzer ohne diese Hilfsmittel auskommen und nur eine einzige automatische Wandlung in das HTML-Format erforderlich ist.

```
#+setupfile: cv-setup-hcard.org
#+language: de

#+author: Titel Vorname Mittelname Nachname
#+title: Titel Vorname Mittelname Nachname

{{{mhead-hcard(
Titel Vorname Mittelname Nachname,
Vorname,
Mittelname,
Nachname,
Titel,
vorname-nachname.jpg,
vorname-nachname.html,
Physiker\,<br/> Software-Entwickler,
vorname.nachname@nf.mpg.de,
cornet,
+49 (0)221-4726-123,
+49 (0)221-4726-298
)}}}

* Expertise
- Medizinische Bildverarbeitung
- C, C++, Lisp, SQL
- Positronen-Emissions-Tomographie

* Werdegang
- 2008 :: Lorem ipsum dolor sit amet [fn:lorem], consectetur adipisi
          sed do eiusmod tempor incididunt ut labore et dolore magna
          aliqua.
- 2003 :: Ut enim ad minim veniam, quis nostrud exercitation ullamco
          laboris nisi ut aliquip ex ea commodo consequat. Duis aute
          irure dolorin reprehenderit in voluptate velit esse cillum
          dolore eu fugiat nulla pariatur [fn:ipsum].
- 2000 :: Excepteur sint occaecat cupidatat non proident, sunt in
          culpa qui officia deserunt mollit anim id est laborum.

# Sie müssen natürlich keine Fußnoten benutzen - es ist aber angenehr

[fn:lorem] Ut enim ad minim veniam.

[fn:ipsum] Lorem ipsum dolor sit amet, consectetur adipisicing elit,
sed do eiusmod tempor incididunt ut labore et dolore magna
aliqua.
```

Abb. 1: Die HTML-Seite (Abb. 2) wurde aus diesem Text im Org-Mode-Format generiert. Es handelt sich dabei um Plain Text, d. h. man kann einen beliebigen Texteditor einsetzen. In diesem Beispiel handelt es sich um einen „Screenshot“ von Emacs (Org-Mode), daher sind bestimmte Teile farblich hervorgehoben. Das Homepage-Team bekommt keine HTML-Seiten, sondern Org-Mode-Code (s. o.) und generiert daraus statische HTML-Seiten.

Max-Planck-Institut für neurologische Forschung
mit Klaus-Joachim-Zülch-Laboratorien der
Max-Planck-Gesellschaft und der
Medizinischen Fakultät der Universität zu Köln
Cologne, Germany http://www.nf.mpg.de

Titel Vorname Mittelname Nachname

Titel Vorname Mittelname Nachname
Physiker,
Software-Entwickler

E-Mail: vorname.nachname@nf.mpg.de
http://www.nf.mpg.de/cornet
Tel.: +49 (0)221-4726-123
Fax: +49 (0)221-4726-298

MPI für neurologische Forschung
Gleueler Str. 50, 50931 Köln, Germany

Expertise

- Medizinische Bildverarbeitung
- C, C++, Lisp, SQL
- Positronen-Emissions-Tomographie

Werdegang

2008
Lorem ipsum dolor sit amet [1], consectetur adipisicing elit, sed do eiusmod temp
dolore magna aliqua.

[...]

2000
Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt

Fußnoten:

[1] Ut enim ad minim veniam.

[2] Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor in
magna aliqua.

Abb. 2: Beispielseite für einen Mitarbeiter, die aus dem Org-Text aus Abb. 1 generiert wurde. Der „Impressum“-Link ganz unten zeigt auf eine Seite mit dem Inhalt von Abb. 3.

„Die persönlichen Mitarbeiterseiten des Max-Planck-Instituts für neurologische Forschung werden eigenverantwortlich von den jeweiligen Mitarbeiterinnen und Mitarbeitern gepflegt. Es wurden Regeln vorgegeben, die für die Gestaltung eigener Inhalte gelten. Haftungsansprüche, die aus Verstößen gegen diese Regeln erwachsen, gehen zu Lasten des entsprechenden Mitarbeiters“.

Abb. 3: Text des Impressum-Links

Optional können die Mitarbeiter auch ein Template wählen, welches die persönlichen Daten im hCard-Format [5] generiert: Dabei werden Meta-Informationen mit angeboten, die die automatische Interpretation der Daten vereinfachen, so dass z. B. Vorname, Nachname oder Stadt automatisch als solche interpretiert werden (und nicht, wie üblich, von Suchmaschinen „geraten“ werden). Natürlich ist die Kehrseite einer verbesserten „Maschinenlesbarkeit“ der so veröffentlichten Daten auch ein erhöhtes Missbrauchspotenzial – wir weisen die Nutzer darauf hin und sie selber können entscheiden, ob sie ein erhöhtes Missbrauchsrisiko tragen wollen.

Auf den Anleitungsseiten [3] ist die vollständige Umsetzung in Org-Mode dokumentiert, insbesondere sei auf die Nutzung von sog. Makros hingewiesen: wie man im oberen Teil von Abb. 1 sehen kann, werden über das `mhead-hcard()`-Makro die für das hCard-Format benötigten Meta-Informationen übermittelt. Das Makro dient hier dazu, Platzhalter in einem HTML-Block zu ersetzen, der an dieser Stelle automatisch in die finale HTML-Datei eingefügt wird. Dieser Mechanismus erlaubt es also prinzipiell, die bereits vorhandenen Möglichkeiten von Org-Mode zur Generierung von HTML-Content um beliebig komplexe HTML-Strukturen zu erweitern. Außerdem unterstützt Org-Mode die übliche Einbindung von Cascading-Stylesheets (CSS), so dass in unserem Beispiel die Mitarbeiterseiten die gleichen Stylesheets nutzen können wie Teile der Homepage. So erreichen wir trotz der individuellen Gestaltungsmöglichkeiten die gewünschte Homogenität aller Institutsseiten.

3. Ausblick

Der Inhalt dieses Beitrags wurde inzwischen auf der FOSDEM 2011-Konferenz als (kleines) Beispielprojekt für die Nutzung von Emacs Org-Mode im Forschungsumfeld präsentiert (Session B. Guerry, GNU DevRoom) [6].

Neben der beschriebenen Anwendung von Emacs Org-Mode für persönliche Mitarbeiterseiten sehen wir eine Reihe anderer Möglichkeiten, dieses Werkzeug sinnvoll im Rahmen wissenschaftlicher Arbeit einzusetzen, etwa zur

Erzeugung statischer HTML-Seiten [7] oder zur Dokumentation von Software [8].

Allerdings bleiben auch diese Anwendungen noch weit hinter dem Spektrum zurück, welches von Org-Mode abgedeckt wird: Für Wissenschaftler interessant ist z. B. auch die Nutzung als Outliner, als integrierter Kalender, „elektronischer Zettelkasten“ oder allgemeines Meta-Format zur Generierung von HTML-Seiten, LaTeX-Publikationen, DocBook-Format oder Wiki-Markup. Noch gar nicht abzusehen sind die neuen Möglichkeiten, die sich für Reproducible Research [9] durch die Org-Babel-Bibliothek ergeben (Dokumentation und Ausführung von Code-Blocks) und die sich bereits in diesem Übersichtsvortrag [10] ankündigt.

4. Danksagung

Wir danken Frau Heidi Schuster (Gruppe Prof. Gerling) für die Beratung bei den rechtlichen Fragen sowie den Org-Mode-Maintainern Carsten Dominik und Bastien Guerry, die zusammen mit anderen Mitgliedern der Org-Mode Mailing List <emacs-orgmode@gnu.org> zum Gelingen des Projekts beigetragen haben.

5. Fußnoten

[1] Org-Mode: http://orgmode.org

[2] Beispiele für Mitarbeiterseiten, die aus Org-Files generiert wurden: http://www.nf.mpg.de/mitarbeiter (deutsch: http://www.nf.mpg.de/cv (englisch)

[3] Anleitung zur Erstellung persönlicher Mitarbeiterseiten: http://www.nf.mpg.de/cv-howto/cv-de.html

[4] Power-User können sich unter Linux / Mac OS X / Windows Emacs mit Org-Mode installieren, um selber HTML-Content aus Org-Files zu erzeugen (und alle anderen Vorteile von Org-Mode zu nutzen): http://www.nf.mpg.de/cv-howto/cv-emacs.html

[5] Das „microformat“ hCard: http://en.wikipedia.org/wiki/HCard

[6] FOSDEM 2011, „Free and Open Source Software Developers’ European Meeting“: http://www.fosdem.org/2011, Brüssel, Februar 2011

[7] KinderUni 2009/2010: Programmierkurs für Kinder mit besonderem Bezug auf die Visualisierung von Daten unserer Tomographen; die HTML-Seiten zum Kurs (CD-ROM) wurden vollständig mit Org-Mode geplant (Outliner) und generiert: http://www.nf.mpg.de/kinderuni

[8] VHIST, ein OpenSource-Projekt zur Dokumentation von Workflows, besonders im Bereich Multi-Modality Imaging. Auf der Homepage des Projekts befindet sich ein Link zur Online-Dokumentation der Referenz-Implementierung von VHIST, welche vollständig mit Org-Mode erstellt wurde: http://www.nf.mpg.de/vhist

[9] Reproducible Research: „An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.“ D. Donoho: http://reproducibleresearch.net

[10] Gastvortrag Carsten Dominik, „Emacs Org-mode: Organizing a Scientist's Life and Work“: http://www.nf.mpg.de/orgmode/guest-talk-dominik.html

Warum IT-Sicherheit am MPI (nicht?) funktionieren kann

Sind mobiles Arbeiten und IT-Sicherheit im Zeitalter „ergebnisorienterter Arbeitsumgebung“ miteinander vereinbar?

Rainer Kleinrensing

Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig

Bertram Smolny

Max-Planck-Institut für Biogeochemie, Jena

IT-Sicherheitsschulungen der MPG, München, Göttingen und Hamburg, Dezember 2010

»Der Glaube, dass die eigene Sicht der Wirklichkeit
die Wirklichkeit schlechthin bedeutet,
ist eine gefährliche Wahnidee.«

Paul Watzlawick

Zusammenfassung

IT-Sicherheit ist kein statisches Gebilde und keine digitale Größe. Sie besteht aus sehr vielen Teilaspekten, die im Zeitalter des Web 2.0 oftmals auf dem Altar der „Usability“ geopfert werden.

Vollständige IT-Sicherheit kann nur durch die totale Kontrolle des modernen „Wissenschafts-Nomaden“ hergestellt werden; empfohlene operative Maßnahmen kollidieren oft mit dem chaotischen Verhalten des „kreativen Individualisten“. Es gibt nicht „Das IT-Sicherheitskonzept der MPG“; die Diversität der Anforderungen der Institute ist viel zu groß, als dass man sie mit einem einheitlichen Konzeptpapier umfassend behandeln kann.

1. Einleitung

Die zunehmende Nutzung des Internet ermöglicht globale Prozesse in der Wissenschaft und erleichtert die Zusammenarbeit und den Austausch von Erkenntnissen. Arbeit und insbesondere wissenschaftliches Arbeit wird immer aber stärker geprägt durch eine „ergebnisorienterte Arbeitsumgebung“ – wissenschaftliche Arbeit ist kein Ort, zu dem man geht, sondern etwas, was man tut, völlig unabhängig vom „wo“ und „wann“. Dem folgen aber viele Vorstellungen der Verwaltungen nicht, die meinen, der „kreative Individualist“ kommt morgens ins Institut, schließt sein Büro auf, setzt sich an den Schreibtisch und schaltet seinen Computer ein.

Daraus folgen auch Forderungen nach flexiblen Arbeitszeitmodellen mit Formen wie Home-Office, alternierender Telearbeit und anderen mobilen Arbeitsformen.

Der Grund dafür ist genauso einfach wie auch komplex, denn wie andere Forschungsorganisationen muss sich die Max-Planck-Gesellschaft einerseits immer mehr als attraktiver Arbeitgeber positionieren. Andererseits erfordern diese durch die sogenannte „Konsumerisierung der Enterprise-IT“ entstandenen Effekte einen völlig neuen Denkansatz in Bezug auf Verfügbarkeit, Datensicherheit und -integrität.

In der Meisterung von mobilem (in Zukunft ubiquitärem) Computing[1] sehen wir in den nächsten Jahren die größten Herausforderungen für die ITK der MPG.

1. http://de.wikipedia.org/wiki/Ubiquitous_Computing

2. Thesen

Die Forderungen nach flexiblen Arbeitszeitmodellen, Home-Office, alternierender Telearbeit und mobilem Arbeiten werden immer häufiger an die Geschäftsführungen der Institute herangetragen.

Der Grund dafür ist genauso einfach wie auch komplex, denn die Institute müssen sich einerseits immer mehr als attraktive Arbeitgeber positionieren, aber andererseits wollen sich die Geschäftsführungen nicht mit den Problemen der IT-Sicherheit, die daraus erwachsen, herumschlagen. IT-Sicherheit wird dabei häufig als „Wissenschaftsverhinderung“ verbrämt.

Die daraus resultierende Frage ist:
(Wie) Lassen sich IT-Sicherheitsanforderungen mit den Anforderungen an den IT-Service, die die Flexibilisierung der Arbeit der Wissenschaft mit sich bringt, vereinbaren?

Das Thema ist an sich polemisch angelegt. Wenn man sich der Materie aus Sicht einen IT-Verantwortlichen/IT-Leiters eines Max-Planck-Instituts nähert, wird man feststellen: Der IT-Leiter kann Verantwortung kaum an einen Sicherheitsverantwortlichen delegieren.

Die IT-Abteilung versteht sich als Dienstleister der wissenschaftlichen Nutzer. Das macht seine Existenzberechtigung aus. Die Nutzer wiederum sehen nur die möglichen Dienstleistungen, nicht die (dahinter liegenden) Datenschutzprobleme. Sie zu sensibilisieren, fällt nicht leicht, wird doch oft nur die Lösung in den Mittelpunkt gestellt, nicht die Diskussion von Problemen (von IT-Sicherheit). Zitat: „Die IT hat lange genug bestimmt, was zu machen ist. Jetzt muss die Forschung wieder mal sagen, was sie wirklich braucht.“

Des Weiteren ist die IT-Abteilung bestrebt, gute Arbeit zu leisten statt sich in „unproduktiven“ Rechtsinterpretationen zu verlieren, weil oftmals der polemische Satz gilt: „Zwei Juristen äußern drei Meinungen“[2]. Die Antworten seitens der Verwaltung sind manchmal eher problem- als lösungsorientiert!

Das hohe Tempo der Änderungen in der Arbeitswelt schlägt direkt in der IT-Abteilung durch, weil die Informationstechnologie alle Arbeitsbereiche durchdrungen hat. Hier wird auch nur noch nach „ergebnisorientierter

2. Die bayrische Justizministerin Beate Merk: „Da Daten anders als Autos oder Handys keine Sachen sind, kann man sie nicht stehlen. Und wo es keine gestohlene Ware gibt, da gibt es auch keine Hehlerei.“ http://www.netzpolitik.org/2010/daten-kann-man-nicht-stehlen/

Arbeitsumgebung“ (Results only Work Environment (RoWE)) geurteilt, die wiederum nur geringen Spielraum für IT-Sicherheit lässt:

1. Es zählen nur Ergebnisse – sonst nichts!
2. Arbeit ist kein Ort, zu dem man geht, sondern etwas, was man tut, gleichgültig wo und wann.
3. Es gibt keine Arbeitszeitregelungen. Die Mentalität der 9-bis-17-Uhr-Jobs ist überwunden.
4. Keine Besprechung ist verpflichtend. Die Wissenschaftler entscheiden selbst, ob sie wichtig oder Zeitverschwendung sind.
5. Es steht niemandem zu, darüber zu urteilen, wie die Wissenschaftler ihre Zeit verbringen.
6. Niemand spricht darüber, wie viele Stunden er oder sie arbeitet.
7. Jeder Tag fühlt sich an, als sei er Samstag.

Kann IT-Sicherheit unter diesen Randbedingungen funktionieren und, wenn ja, wie?

Die Komplexität und Dynamik nimmt so stark zu, dass wir sie als Individuen kaum mehr beherrschen können. So titelt die „Süddeutsche Zeitung“ vom 20.11.2010: „Wie die Technik uns beherrscht“. Auch wird die Datendichte weiter zunehmen. Die Frage, wer auf die Daten berechtigt zugreifen darf, wird ausgeblendet. Beispiel ist die „Wolke Google“.

Sind die Daten erst einmal erhoben, dann werden auch die Verwertungsmodelle dafür entwickelt (siehe hier auch „Illegale Steuer-CD darf für Ermittlungen genutzt werden“)[3].

Statt sich mit Datenschutz und der Sicherheit auseinander zu setzen, wird oftmals Bequemlichkeit als Effizienz verbrämt; sporadisches Micromanagement statt permanentem Arbeiten nach Datenschutzrichtlinien.

Auch wird der Umgang mit Daten autoritärer: Die Geschäftsmodelle im sogenannten Web 2.0 basieren auf den Inhalten, die nur ein geringer Teil der Nutzerschaft einbringt (Nielsen-Regel: 90-9-1-Regel zur Nutzung von Communities: Verteilung 90 % konsumieren, 9 % fügen Inhalt hinzu, aber nur 1 % bewerten diesen[4].

3. http://www.nzz.ch/nachrichten/wirtschaft/aktuell/illegale_steuer-cd_darf_fuer_ermittlungen_genutzt_werden_1.8517531.html

3. Begriffsdefinitionen: Normung der Informationssicherheit

„Als Informationssicherheit bezeichnet man Eigenschaften von informationsverarbeitenden und -lagernden Systemen, welche die Vertraulichkeit, Verfügbarkeit und Integrität sicherstellen.

Informationssicherheit dient dem Schutz vor Gefahren bzw. Bedrohungen, der Vermeidung von Schäden und der Minimierung von Risiken.

In der Praxis orientiert sich die Informationssicherheit heute unter anderem an der ISO/IEC-Standard-Reihe 2700x, aber auch zunehmend an ISO/IEC 15408.

Der deutsche Anteil an dieser Normungsarbeit wird vom DIN-NIA-01-27 IT-Sicherheitsverfahren betreut.

Die Informationssicherheit umfasst neben der Sicherheit der IT-Systeme und der darin gespeicherten Daten auch die Sicherheit von nicht elektronisch verarbeiteten Informationen.“[5]

In dieser Sichtweise ist die Informationssicherheit eine ökonomische Größe: Es liegt Informationssicherheit vor, wenn über einen bereits bekannten Weg kein Angriff auf das System mehr möglich ist.

Oftmals wird aber auch Sicherheit als indirekt proportional zur Zeit des zurückliegenden Ereignisses betrachtet.

Unter dem Dach der Informationssicherheit gibt es einen Teilaspekt IT-Sicherheit, der wiederum Untermengen ausbildet: Computersicherheit, Netzwerksicherheit, Datensicherheit, Datensicherheit als Synonym für Backup.

Nicht vollständig geklärt ist in der Max-Planck-Gesellschaft, was der IT-Sicherheits-beauftragte zu diesen Szenerien exakt beizutragen hat. Was muss der tun? Was unterlassen? Es gibt unserem Kenntnisstand nach keine Funktionsbeschreibung im Personalhandbuch.

IT-Sicherheit ist aber entgegen landläufiger Meinung keine binäre, es ist eine analoge Größe.

Außerdem täuscht hier die Wahrnehmung: Da der Tag meistens mit anderen Dingen ausgefüllt ist, wird Sicherheit nur manchmal wahrgenommen und ist „gefühlt“ indirekt proportional zur Zeit des letzten Sicherheitsvorfalls.

---

4. http://de.wikipedia.org/wiki/Jakob_Nielsen_(Webdesignexperte)
5. http://de.wikipedia.org/wiki/Informationssicherheit

Dabei ist IT-Sicherheit ein fortwährender Prozess, der eine kontinuierliche Beachtung erfordert.

4. Fundamentaler Gegensatz

Die Durchdringung aller Arbeits- (und Lebens-)bereiche mit Informations-Technologie ist fast vollständig.

Dieser Fakt führt zu einem enormen Anwachsen der technisch unterstützten Funktionalität, die gleichzeitig auch die Komplexität (und mit ihr die Fehleranfälligkeit) steigen lässt.

Mit der Präsenz der Informationstechnologie steigt in gleichem Maße auch die Abhängigkeit von ihr. Aber die Zufriedenheit bei Kunden (hier Entscheider) und Nutzern steigt nicht.

Deren Haltung wird eher kritischer, die Akzeptanz der Arbeit der IT-Abteilung schlechter.

Sogar die „Alltagsmedien“ haben das mittlerweile erkannt: Die Technik beherrscht uns.

Der Druck, der damit seitens der „kreativen Individualisten“ aufgebaut wird, wird immer stärker, die Entropie dahinter immer größer.

Aber die Zufriedenheit der Entscheider, hier wissenschaftliche Kunden, wächst nicht mit.

Es ist eine Trendumkehr zu beobachten: Firmen, die früher für den Enterprise-Markt Produkte entwickelten, wenden sich heute dem Consumer-Markt mit mehreren 100 Mio. Konsumenten zu.

Kann diesen Anforderungen eine IT-Abteilung bestehend aus drei bis sieben Mitarbeitern Rechnung tragen?

5. Die ITK-Abteilung

Einen wichtigen Beitrag zur Umsetzung der IT-Sicherheitsrichtlinien leistet ohne Zweifel auch die ITK-Abteilung des jeweiligen Instituts. Diese ist je nach Anforderungen, Geschichte und Signifikanz verschieden aufgestellt. Eine Standardisierung dieser IT über die gesamte MPG wäre nicht durchsetzbar und würde die Kreativität erdrücken, weil oftmals und gerade von den Instituten die maßgeblichen Innovationen im IT-Bereich kommen.

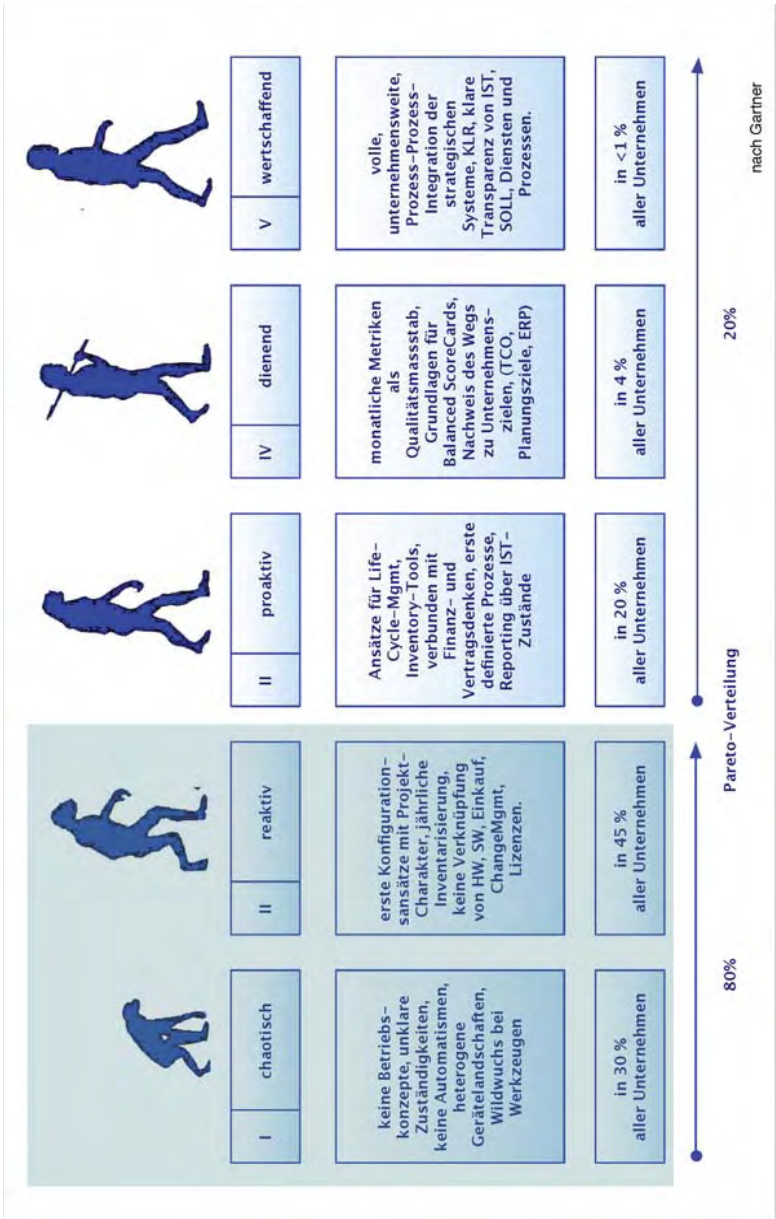

Abb. 1: Die ITK-Abteilung nach Gardner

Erläuterung nach Gardner (s. Abb. 1):

1. chaotisch: keine Betriebskonzepte
2. reaktiv: Projektcharakter
3. proaktiv: Ansatz zu Life-cycle-Management
4. dienend: monatliche Metrik
5. wertschaffend: Prozess-Integration.

Dies gilt für ITK in Industrie, Handel und Banken – über Wissenschaft ist hier keine Aussage gemacht. Man kann aber davon ausgehen, dass viele der IT-Abteilungen noch nicht vollständig in der Prozess-Integration [5] angekommen sind.

In bestimmten Werbeschriften kommt auch schon keine IT-Abteilung mehr vor: Der „digitale Wissenschafts-Nomade“ braucht sie nicht mehr?!

Auch ist die Arbeit „in der Wolke“ losgelöst von den (oftmals schwerfälligen) Verwaltungsprozessen, die im Institut ablaufen. Wie soll da die IT-Sicherheit des Instituts integriert werden?

Diese Werbebotschaften signalisieren aber auch der Geschäftsführung, dass die IT-Abteilung eventuell ineffizient arbeitet, weil eben doch Zeit für die Einrichtung eines Mail-Kontos oder einer Konfigurationsänderung des Servers gebraucht wird, alleine, um diese auch sauber zu dokumentieren. Ob das im Privaten wirklich so viel besser läuft, ist völlig unbewiesen.

6. Konsumerisierung der Enterprise-IT

Einen unserer Meinung nach nicht zu vernachlässigenden Beitrag zum Thema IT-(Un-)Sicherheit spielt auch die Vermischung von Privat- und Geschäftsinteressen, auch unter dem Schlagwort „Konsumerisierung“ bekannt.

Generation iPhone und iPad sorgen dafür, dass die Grenzen zwischen Arbeit und Freizeit verschwimmen – ein Phänomen, das nicht einzudämmen ist, da es auch und verstärkt über die Leistungsträger „eingeschleppt“ wird.

„Genau das führt bei IT-Abteilungen mitunter zu Kopfschmerzen“[6], weil diese Geräte fortgeschrittene Sicherheitsstandards unterlaufen, viele Mitar-

6. http://www.cio.de/_misc/article/printoverview/index.cfm?pid=587&pk=2276164&op=lst

beiter aber mittlerweile zu Hause eine (ihrer Meinung nach) bessere IT-Ausstattung vorfinden als im Institut.

„Das Vermengen von privaten und geschäftlichen Daten ist nicht unproblematisch, weil es beispielsweise möglich macht, vertrauliche Informationen vom einen in den anderen Bereich zu kopieren“, gibt Andrew Jaquith zu bedenken.[7]

Interessant ist es, zu beobachten, dass auch dies wieder ein neues Geschäftsfeld erzeugt: Aufsichtsräte großer Unternehmen, die vom Standpunkt der IT gesehen „extern“ sind, wollen die zu besprechenden Dokumente auf ihren iPads und schlecht gewarteten Notebooks sehen und bearbeiten. Um hier einen unkontrollierten Abfluss der sensiblen Informationen zu verhindern, gibt es bereits Software, die solcherlei Informationen nur auf einer Portal-Seite sichtbar macht, aber verhindert, dass sie auf das Endgerät heruntergeladen werden. Das Unternehmen wird so vor der IT seiner Aufsichtsräte geschützt. [10]

7. Phänomen Apple & Co.

Da kein Wachstum im Enterprise-Bereich mehr möglich schien (hier waren Anbieter wie Microsoft, IBM oder Oracle schon zu weit enteilt), wurde der Fokus auf den „kreativen Individualisten“, der dann sein Gadget auch im Institut verwenden will, gelegt. Dieses Phänomen ist in der Wissenschaft konkret zu beobachten.

Zum Anderen erwarten Anwender, die mit ihren eigenen Geräten ins Büro kommen, dass sie sich ihr persönliches Portfolio an Applikationen und Services zusammenstellen und damit arbeiten können. Gleichzeitig wollen sie institutsfremde Anwendungen, wie etwa gMail, Facebook oder Skype, nutzen.

Können sie das nicht, fühlen sie sich von ihrer wissenschaftlichen Community ausgegrenzt. Man kann also derartige Anwendungen schwer verbieten; unserer Meinung nach bringt eine Sensibilisierung für das Gefahrenpotenzial viel mehr als ein Verdammen in Bausch und Bogen. Was kann man tun?

Ein erster Ansatz könnte eine Datenverschlüsselung auf mobilen Datenträgern mit allen damit in Verbindung stehenden Konsequenzen und Problemen sein.

7. http://www.cio.de/_misc/article/printoverview/index.cfm?pid=560&pk=2243417&op=lst

Das wiederum führt zu Themen wie PKI, Admin-Rechte auf Notebooks, Verschlüsselungsverbote bei Auslandsreisen etc.

Ein anderer Ansatz ist die Arbeit mit virtuellen Desktops per Webzugriff. In diesem Fall liegen alle Daten sicher im Institut und die Mitarbeiter können von jedem beliebigen internetfähigen Computer arbeiten; die Akzeptanz dieses Dienstes vorausgesetzt.

Dies aber setzt voraus, dass eine Risikoabschätzung vorliegt, um daraus Sicherheitspolicies zu definieren.

Die Umsetzung dieser Sicherheitspolicies führt dann zu verschiedenen Profilen: PC im Internetcafé im Ausland, privater Home-Arbeitsplatz oder Firmen-Notebook.[8]

Weitere Sofortmassnahmen können u. a. sein, die Mitarbeiter ständig zu qualifizieren, um ihnen auch so den Freiraum für eigene Projekte zu geben. Diese Projekte wiederum sollten sich auch nur an einem wichtigen Parameter orientieren: Offene Schnittstellen. Wenn jetzt noch die Stellenanzahl der IT-Mitarbeiter adäquat wächst, ist man auf gutem Wege.

8. Beispiele aus dem Alltag

8.1 Literaturzugang

Die Verfügbarkeit von elektronischer Literatur, insbesondere elektronischer Zeitschriften (e-Journals) ist heutzutage unverzichtbar, um den aktuellen Stand der Forschung innerhalb eines bestimmten Themengebietes einzusehen. Da die Wissenschaftler keine persönlichen Logins bei den Verlagen haben und anonymisierte Login-Verfahren wie Shibboleth[9] noch nicht flächendeckend zur Verfügung stehen, besteht ein gewisses Missbrauchspotenzial. Dies muss gegen die Freiheit der Forschung abgewogen werden.

8.2 Skype

Das Internet-Telefonie-Programm hat sich zu einem Quasi-Standard entwickelt. Da der Quellcode nicht vorliegt und dank starker Verschlüsselung auch nicht geklärt werden kann, ob etwa Daten des Benutzers nach außen abflie-

8. Hier muss auf eine „Risikoabschätzung mobile Rechner im internen Netzwerk“ verwiesen werden, ohne die der Betrieb sicherer ITK-Strukturen mit mobilen Wissenschafts-Nomaden nicht möglich ist.

9. http://shibboleth.internet2.edu/

ßen (oder auch illegale Inhalte dank des Peer-to-Peer-Prinzips das eigene Netz passieren) ist es oft nicht gestattet, dieses Programm einzusetzen. Allerdings ist dann evtl. auch die wissenschaftliche Zusammenarbeit gefährdet, da z. B. Gruppen in den USA diese Bedenken nicht teilen. Eine Lösung könnte sein, Skype nur in dedizierten Netzen zuzulassen. [9]

8.3 Personal Cloud Storage

Dienste wie Wuala, Teamdrive und vor allem Dropbox erfreuen sich zunehmender Beliebtheit. Neben rechtlichen Fragen (dürfen institutseigene Daten überhaupt auf Nicht-MPG-Servern gelagert werden) stellen sich auch Fragen nach der Vertraulichkeit, etwa bei Daten, die zu einem Patent führen sollen.

8.4 Zertifikate

Das Ausrollen einer PKI gehört zu den eher komplizierten Aufgaben der IT; nicht deshalb, weil es geheimnisvoll wäre, sondern wegen des organisatorischen Aufwandes, der zum sicheren Betrieb nötig ist. Obwohl in der MPG eine PKI vorhanden ist, die dank dezentraler Strukturen sogar sehr schnell reagieren kann, gibt es immer wieder Stellen, an denen man selbstsignierte Zertifikate vorfindet, damit die Einrichtung des Dienstes noch schneller vonstatten geht. Wenn diese Situation dann noch Eingang in Dokumentation findet („beim Aufruf der Webseite müssen Sie die Warnung vor unsicheren Zertifikaten zweimal ignorieren“) kann von Sicherheitsbewusstsein beim Anwender keine Rede mehr sein. [8]

Allgemein ist festzustellen, dass es ein zweidimensionales Spannungsfeld von Sicherheit versus Arbeitsqualität auf der einen Achse und von Freiheit versus Standardisierung auf der anderen Achse gibt. Der eigene Dienst muss hier verortet werden.

8.5 Externer Audit – Netzwerkprüfer

Seit einiger Zeit gibt es Werkzeuge, die das eigene Netz von außen auf Schwachstellen prüfen können (DFN-Warnmeldungen). Eigentlich eine sehr sinnvolle Sache, leider ist die Akzeptanz innerhalb der MPG nicht sehr hoch. [6]

8.6 Mobiltelefon

Wie stelle ich sicher, dass mein Telefon sicher ist? Leider ist auch im Zeitalter von iPhone & Co. immer noch „Turnschuh-Administration“ zur Installation der neuesten Sicherheitsupdates bei diesen nicht managebaren Geräten nötig. Wie man dann konkret die Disziplin der Nutzer einfordert, starke

Passwörter bzw. PIN-Codes zu verwenden, steht noch auf einem anderen Blatt.

8.7 Backup

Wenn einmal eine Datensicherung angefertigt ist, soll sie vom Standpunkt der IT-Abteilung aus möglichst lange erhalten bleiben bzw. konserviert werden. Aber auch hier gibt es Löschfristen, deren Nichteinhaltung juristische Konsequenzen nach sich ziehen kann.

8.8 Archiv

„Langzeitarchivierung verletzt Urheberrecht“ – diese These war auf dem letzten DV-Treffen in Göttingen zu hören. Hier ist die Gesetzgebung noch sehr dynamisch, aber umso schneller läuft man in juristische Fallstricke hinein. [7]

8.9 USB-Stick

An jedem MPI finden mindestens wöchentlich Seminarvorträge von bedeutenden Wissenschaftlern statt. Meist bringen sie ihre Vorträge auf einem USB-Stick mit, der eine Minute vor Beginn des Vortrages in den Instituts-Vortragsrechner gesteckt werden muss. Wir verlassen uns darauf, dass der MPG-Virenscanner zuverlässig erkennt, welche Malware auf dem Stick ist. Dass der Vortrag vorab per Mail verschickt wird, ist in einer Forschungslandschaft nicht durchsetzbar.

9. Schlussfolgerungen

Auch zukünftig wird das Internet die wissenschaftliche Entwicklung maßgeblich mitprägen. Das Recht auf informelle Selbstbestimmung, Daten- und Urheberschutz stehen dabei im Mittelpunkt. Dazu müssen technische und organisatorische Rahmenbedingungen angepasst werden.

Entgegen landläufiger Meinungen haben sich aber die IT-Abteilungen der Max-Planck-Institute längst von der Rolle des „Druckerpapiereinleger“ emanzipiert – sie sind zum „Enabler“ der Wissenschaft geworden.[10]

Ohne die intelligente Organisation der ITK-Struktur ist heute Spitzenforschung nicht mehr möglich.

10. https://init.mpg.de

a) IT-Sicherheit ist ein Gebot der Stunde

   Aber wie lässt sich IT-Sicherheit messen und damit vergleichbar machen?

b) IT-Sicherheit verursacht Kosten

   Wer ist bereit, für die erforderlichen Ressourcen zu bezahlen?

c) Es gibt nicht „Das IT-Sicherheitskonzept der MPG“.

   Die Diversität der Anforderungen der Institute ist viel zu groß, als dass man sie mit einem einheitlichen Konzeptpapier umfassend behandeln kann, aber es gibt Maßnahmen, die die IT-Sicherheit signifikant erhöhen können: von Risikoabschätzung über Mitarbeiterschulung bis hin zu Planungshandbüchern.

d) In einer sich ständig ändernden Arbeitswelt muss den Anforderungen des „digitalen Wissenschafts-Nomaden“ mehr Rechnung getragen werden.

   Wissenschaftsarbeit ist kein Ort, zu dem man geht, sondern etwas, was man tut, gleichgültig wo und wann.

e) Schlussfolgerungen nach „Cable-Gate“:

   1. Alle Informationen, die digital vorliegen, erblicken irgendwann das Licht der Welt.
   2. Man muss sich der Abhängigkeit von fremden Regierungsentscheidungen bzgl. Technologie bewusst sein.
   3. Rezipienten können nicht mehr so tun, als hätten sie von nichts gewusst.

Also: IT-Sicherheit kann im MPI nur **unter gewissen Umständen** funktionieren.

Diese Umstände müssen gemeinsam von den IT-Mitarbeitern, den Wissenschaftlern, aber auch den Juristen definiert werden.

Literatur

[1] Ilija Trojanow; Juli Zeh: Angriff auf die Freiheit – Sicherheitswahn, Überwachungsstaat und der Abbau bürgerlicher Rechte. Hanser, München 2009.

[2] Rainer W. Gerling, Heidi Schuster: Der transparente Bürger: http://www.mpg.de/print/1050554

[3] Thomas Baumann, Dieter Ruder, Bertram Smolny (Hrsg.): 25. DV-Treffen der Max-Planck-Institute – 18. - 20. November 2008 in Göttingen. GWDG-Bericht Nr. 75: http://www.gwdg.de/fileadmin/inhaltsbilder/Pdf/Publikationen/GWDG-Berichte/gwdg-bericht-75.pdf

[4] Thomas Pelkmann: „Widerstand gegen das iPad sinnlos“: http://www.cio.de/_misc/article/printoverview/index.cfm?pid=562&pk=2229450&op=lst

[5] Christiane Pütter: Die fünf Typen des mobilen Mitarbeiters: http://www.cio.de/_misc/article/printoverview/index.cfm?pid=160&pk=2235793&op=lst

[6] Automatische Warnmeldungen des DFN: https://www.cert.dfn.de/autowarn/

[7] Harald Müller: Langzeitarchivierung verletzt Urheberrecht. Beitrag im Plenum des 27. DV-Treffens der MPG vom 14. - 16.09.2010 in Göttingen.

[8] PKI des DFN-Vereins: http://pki.pca.dfn.de

[9] Skype-Policy der University of Santa Barbara: http://www.oit.ucsb.edu/connect/skype.asp

[10] Andrea König: „Hilti-CIO verkündet Ende der IT-Diktatur“: http://www.cio.de/2264665

Der „IT Community Award“

Die Organisatoren des DV-Treffens 2010

Max-Planck-Institut für Biologie des Alterns, Köln

Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern, Bonn

Max-Planck-Institut für Gesellschaftsforschung, Köln

Max-Planck-Institut für neurologische Forschung, Köln

Max-Planck-Institut für Mathematik, Bonn

Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln

Max-Planck-Institut für Radioastronomie, Bonn

Der „IT Community Award“ wurde zum DV-Treffen 2010 zum ersten Mal verliehen. Mit dem Preis soll die Anerkennung der IT-Community der MPG zum Ausdruck gebracht werden, dass der Beitrag des Preisträgers (Person/ Gruppe) vielen geholfen hat, ihre Arbeit besser und erfolgreicher zum Ziel zu bringen (jeweils bezogen auf den Zeitraum zwischen den DV-Treffen). Das kann z. B. durch konstruktive Beiträge in Mailinglisten, Foren und in Peer-to-Peer-Gruppen einzelner Einrichtungen, durch Einbringen von Zeit und Energie in die Ausrichtung von Workshops oder die Beschäftigung mit

zentralen Kernthemen im Sinne des gemeinsamen Vorwärtskommens auch außerhalb des lokalen Arbeitsfeldes geschehen sein.

Kandidaten für den Preis können bis zu drei Wochen vor dem DV-Treffen durch die Community der MPG benannt werden, wobei Mitglieder des jeweils aktuellen Organisationsteams nicht aufgestellt werden dürfen. Eine kurze schriftliche Begründung ist erforderlich und aus den Kandidaten wird vom jeweiligen Organisationsteam die Preisträgerin, der Preisträger oder vielleicht auch eine Gruppe bestimmt.

Es folgt der Originaltext der Urkunde, welche zusammen mit einer bedruckten Tasse, dem diesjährigen Preis, übergeben wurde. Die Tasse ist ein Unikat; wir bedanken uns bei http://xkcd.com für die Erlaubnis, ein (besonders IT-typisches) Motiv der Comic-Serie für die Gestaltung benutzen zu dürfen.

Urkunde 2010 (Originaltext)

Der IT-Community Preis würdigt einen wesentlichen Beitrag zur Förderung des Community-Gedankens in der IT der MPG.

Während die meisten Max-Planck-Institute schon aufgrund ihrer vielfältigen und spezifischen Forschungsrichtungen nur wenige bis gar keine Schnittmengen in der täglichen Kooperation aufweisen, ist dies für die zugehörigen IT-Einrichtungen der einzelnen Institute zum Glück anders.

Aus diesem Grunde haben die jährlichen DV-Treffen, die themenbezogenen Workshops und institutsübergreifenden Kooperationen einen sehr großen Stellenwert in der Arbeit der IT-Mitarbeiter/innen.

Konstruktive Beiträge in Mailinglisten, Foren und in Peer-to-Peer-Gruppen einzelner Einrichtungen, das Einbringen von Zeit und Energie in die Ausrichtung von Workshops, die Beschäftigung mit zentralen Kernthemen im Sinne des gemeinsamen Vorwärtskommens auch außerhalb des lokalen Arbeitsaufwandes sind für alle Beteiligten ein großer Gewinn.

Mit diesem Preis soll daher die Anerkennung der IT-Community zum Ausdruck gebracht werden, dass der Beitrag vielen geholfen hat, ihre Arbeit besser und erfolgreicher zum Ziel zu bringen.

Im Besonderen:

Der diesjährige MPG IT-Community Preis, der zum ersten Mal verliehen wird, soll daher die Mühen und Leiden, aber auch den Einsatz und die ersten Erfolge beim Aufbau einer neuen IT-Community Plattform in der MPG zum

besseren Austausch von Ideen, von Best-Practise-Lösungen und Diskussionen, vor allem aber einem neuen Level an Kooperationen, belohnen.

Viele Mitstreiter und Initiatoren haben daran im Vorfeld mitgewirkt, und müßten hier genannt werden, aber letztlich ist es dem Einsatz des IT-Verantwortlichen und seinen Mitarbeitern am Max-Planck-Institut für Informatik zu verdanken, dass diese Kollaborationsplattform bereitgestellt und mit ersten Strukturen und Inhalten gefüllt worden ist, so dass alle IT-Kollegen/innen in der MPG daran gewinnbringend partizipieren können.

Somit verleihen wir (in diesem Jahr das Organisationsteam der Köln-Bonner Max-Planck-Institute) den IT-Community Preis der Max-Planck-Gesellschaft an:

Herrn Jörg Herrmann und seine IT-Mitarbeiter am MPI in Saarbrücken.

Das Organisationsteam möchte es aber nicht unterlassen, alle IT-Kollegen/innen in der MPG zu ermuntern, diesen Invest an Zeit und Wissen nicht zu ignorieren, sondern vielmehr durch eigene Beiträge für die IT-Community auszubauen und so eine Grundlage für eine gemeinsame Wissensbasis zu schaffen.

MAX-PLANCK-GESELLSCHAFT

IT Community Award 2010

Der IT-Community Preis würdigt einen wesentlichen Beitrag zur Förderung des Community-Gedankens in der IT der MPG.

Während die meisten Max-Planck-Institute schon aufgrund ihrer vielfältigen und spezifischen Forschungsrichtungen nur wenige bis gar keine Schnittmengen in der täglichen Kooperation aufweisen, ist dies für die zugehörigen IT-Einrichtungen der einzelnen Institute zum Glück anders.

Aus diesem Grunde haben die jährlichen DV-Treffen, die themenbezogenen Workshops und institutsübergreifenden Kooperationen einen sehr großen Stellenwert in der Arbeit der IT-Mitarbeiter/innen.

Konstruktive Beiträge in Mailinglisten, Foren und in Peer-to-Peer-Gruppen einzelner Einrichtungen, das Einbringen von Zeit und Energie in die Ausrichtung von Workshops, die Beschäftigung mit zentralen Kernthemen im Sinne des gemeinsamen Vorwärtskommens auch außerhalb des lokalen Arbeitsaufwandes sind für alle Beteiligten ein großer Gewinn.

Mit diesem Preis soll daher die Anerkennung der IT-Community zum Ausdruck gebracht werden, dass der Beitrag vielen geholfen hat, ihre Arbeit besser und erfolgreicher zum Ziel zu bringen.

Im Besonderen:

Der diesjährige MPG IT-Community Preis, der zum ersten Mal verliehen wird, soll daher die Mühen und Leiden, aber auch den Einsatz und die ersten Erfolge beim Aufbau einer neuen IT-Community Plattform in der MPG zum besseren Austausch von Ideen, von Best-Practise-Lösungen und Diskussionen, vor allem aber einem neuen Level an Kooperationen, belohnen.

Viele Mitstreiter und Initiatoren haben daran im Vorfeld mitgewirkt, und müßten hier genannt werden, aber letztlich ist es dem Einsatz des IT-Verantwortlichen und seinen Mitarbeitern am Max-Planck-Institut für Informatik zu verdanken, dass diese Kollaborationsplattform bereitgestellt und mit ersten Strukturen und Inhalten gefüllt worden ist, so dass alle IT-Kollegen/innen in der MPG daran gewinnbringend partizipieren können.

Somit verleihen wir (in diesem Jahr das Organisationsteam der Köln-Bonner Max-Planck-Institute) den IT-Community Preis der Max-Planck-Gesellschaft an:

Herrn Jörg Herrmann und seine IT-Mitarbeiter am MPI in Saarbrücken.

Das Organisationsteam möchte es aber nicht unterlassen, alle IT-Kollegen/innen in der MPG zu ermuntern, diesen Invest an Zeit und Wissen nicht zu ignorieren, sondern vielmehr durch eigene Beiträge für die IT-Community auszubauen und so eine Grundlage für eine gemeinsame Wissensbasis zu schaffen.

Originalurkunde 2010

In der Reihe GWDG-Berichte sind zuletzt erschienen:

Nähere Informationen finden Sie im Internet unter
`http://www.gwdg.de/gwdg-berichte`

Nr. 40 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1994
1995

Nr. 41 *Brinkmeier, Fritz* (Hrsg.):
Rechner, Netze, Spezialisten. Vom Maschinenzentrum zum Kompetenzzentrum - Vorträge des Kolloquiums zum 25jährigen Bestehen der GWDG
1996

Nr. 42 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1995
1996

Nr. 43 *Wall, Dieter* (Hrsg.):
Kostenrechnung im wissenschaftlichen Rechenzentrum - Das Göttinger Modell
1996

Nr. 44 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1996
1997

Nr. 45 *Koke, Hartmut und Engelbert Ziegler* (Hrsg.):
13. DV-Treffen der Max-Planck-Institute - 21.-22. November 1996 in Göttingen
1997

Nr. 46 **Jahresberichte 1994 bis 1996**
1997

Nr. 47 *Heuer, Konrad, Eberhard Mönkeberg und Ulrich Schwardmann*:
Server-Betrieb mit Standard-PC-Hardware unter freien UNIX-Betriebssystemen
1998

Nr. 48 *Haan, Oswald* (Hrsg.):
Göttinger Informatik Kolloquium - Vorträge aus den Jahren 1996/97
1998

Nr. 49 *Koke, Hartmut und Engelbert Ziegler* (Hrsg.):
IT-Infrastruktur im wissenschaftlichen Umfeld - 14. DV-Treffen der Max-Planck-Institute, 20. - 21. November 1997 in Göttingen
1998

Nr. 50 *Gerling, Rainer W.* (Hrsg.):
Datenschutz und neue Medien - Datenschutzschulung am 25./26. Mai 1998
1998

Nr. 51 *Plesser, Theo und Peter Wittenburg* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1997
1998

Nr. 52 *Heinzel, Stefan und Theo Plesser* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1998
1999

Nr. 53 *Kaspar, Friedbert und Hans-Ulrich Zimmermann* (Hrsg.):
Internet- und Intranet-Technologien in der wissenschaftlichen Datenverarbeitung - 15. DV-Treffen der Max-Planck-Institute, 18. - 20. November 1998 in Göttingen
1999

Nr. 54 *Plesser, Theo und Helmut Hayd* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 1999
2000

Nr. 55 *Kaspar, Friedbert und Hans-Ulrich Zimmermann* (Hrsg.):
Neue Technologien zur Nutzung von Netzdiensten - 16. DV-Treffen der Max-Planck-Institute, 17. - 19. November 1999 in Göttingen
2000

Nr. 56 *Plesser, Theo und Helmut Hayd* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2000
2001

Nr. 57 *Hayd, Helmut und Rainer Kleinrensing* (Hrsg.):
17. und 18. DV-Treffen der Max-Planck-Institute
22. - 24. November 2000 in Göttingen
21. - 23. November 2001 in Göttingen
2002

Nr. 58 *Plesser, Theo und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2001
2003

Nr. 59 *Suchodoletz, Dirk von*:
Effizienter Betrieb großer Rechnerpools - Implementierung am Beispiel des Studierendennetzes an der Universität Göttingen
2003

Nr. 60 *Haan, Oswald* (Hrsg.):
Erfahrungen mit den IBM-Parallelrechnersystemen RS/6000 SP und pSeries690
2003

Nr. 61 *Rieger, Sebastian:*
Streaming-Media und Multicasting in drahtlosen Netzwerken - Untersuchung von Realisierungs- und Anwendungsmöglichkeiten
2003

Nr. 62 *Kremer, Kurt und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2002
2003

Nr. 63 *Kremer, Kurt und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2003
2004

Nr. 64 *Koke, Hartmut* (Hrsg.):
GÖ* – Integriertes Informationsmanagement im heterogenen eScience-Umfeld: GÖ*-Vorantrag für die DFG-Förderinitiative „Leistungszentren für Forschungsinformation“
2004

Nr. 65 *Koke, Hartmut* (Hrsg.):
GÖ* – Integriertes Informationsmanagement im heterogenen eScience-Umfeld: GÖ*-Hauptantrag für die DFG-Förderinitiative „Leistungszentren für Forschungsinformation“
2004

Nr. 66 *Bussmann, Dietmar und Andreas Oberreuter* (Hrsg.):
19. und 20. DV-Treffen der Max-Planck-Institute
20. - 22. November 2002 in Göttingen
19. - 21. November 2003 in Göttingen
2004

Nr. 67 *Gartmann, Christoph und Jochen Jähnke* (Hrsg.):
21. DV-Treffen der Max-Planck-Institute
17. - 19. November 2004 in Göttingen
2005

Nr. 68 *Kremer, Kurt und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2004
2005

Nr. 69 *Kremer, Kurt und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2005
2006

Nr. 70 *Gartmann, Christoph und Jochen Jähnke* (Hrsg.):
22. DV-Treffen der Max-Planck-Institute
16. - 18. November 2005 in Göttingen
2006

Nr. 71 *Hermann, Klaus und Jörg Kantel* (Hrsg.):
23. DV-Treffen der Max-Planck-Institute
15. - 17. November 2006 in Berlin
2007

Nr. 72 *Kremer, Kurt und Volker Macho* (Hrsg.):
Forschung und wissenschaftliches Rechnen - Beiträge zum Heinz-Billing-Preis 2006
2007

Nr. 73 *Baumann, Thomas, Dieter Ruder und Bertram Smolny* (Hrsg.):
24. DV-Treffen der Max-Planck-Institute
6. - 8. November 2007 in Jena
2008

Nr. 74 *Schwardmann, Ulrich* (Hrsg.):
Grid-Technologie in Göttingen - Beiträge zum Grid-Ressourcen-Zentrum GoeGrid
2009

Nr. 75 *Baumann, Thomas, Dieter Ruder und Bertram Smolny* (Hrsg.):
25. DV-Treffen der Max-Planck-Institute
18. - 20. November 2008 in Göttingen
2009

Nr. 76 *Assmann, Wolfgang, Christa Hausmann-Jamin und Frank Malisius* (Hrsg.):
26. DV-Treffen der Max-Planck-Institute
22. - 24. September 2009 in Berlin
2010

Nr. 77 *Oberreuter, Andreas, Vollmar, Stefan und Alexander Weiße* (Hrsg.):
27. DV-Treffen der Max-Planck-Institute
14. - 16. November 2010 in Göttingen
2011