

# Von Gigaflops zu Teraflops: VIER JAHRE SUPERCOMPUTING MIT „SCHRÖDINGER“

## Was bisher geschah

Zahlreiche Wissenschaftszweige – Theoretische Chemie, Festkörperphysik, Mathematik, Meteorologie u.a. – haben einen sehr hohen Bedarf an Rechenleistung. Um diesen Bedarf abzudecken, wurde im Jahr 2001 vom Zentralen Informatikdienst gemeinsam mit der Fakultät für Naturwissenschaften und Mathematik ein Supercomputing-Projekt initiiert. Die Laufzeit des Projekts sollte vier Jahre betragen. Weil ein Supercomputer in vier Jahren hoffnungslos veraltet, wurden von Anfang an Budgetmittel für zwei Ausbaustufen vorgesehen.

In mehreren *Comment*-Artikeln wurde über den bisherigen Verlauf des Projekts berichtet: Nach einer öffentlichen Ausschreibung erfolgte im Sommer 2001 der Zuschlag an den Bestbieter, die auf Cluster-Lösungen spezialisierte Wiener Firma *init.at* ([www.init.at](http://www.init.at)). Das von *init.at* konzipierte System war ein Cluster aus 192 Rechenknoten und mehreren Servern. Im Herbst 2001 wurden die ersten Komponenten geliefert, Ende 2001 nahm der neue Supercomputer – der den Namen „Schrödinger I“ erhielt – den Produktionsbetrieb auf. Planmäßig erfolgte die zweite Ausbaustufe Anfang 2003, wobei die Rechenleistung deutlich gesteigert werden konnte: Im Juni 2003 erreichte „Schrödinger II“ Platz 190 in der Liste der 500 schnellsten Supercomputer der Welt ([www.top500.org](http://www.top500.org)).

Während in Schrödinger I Athlon-Prozessoren von AMD eingesetzt wurden, verwendete Schrödinger II Pentium-Prozessoren von Intel. Bereits Anfang 2004 wurde mit Überlegungen begonnen, wie die dritte Ausbaustufe aussehen könnte. Intels Erzrivale AMD hatte mit dem Opteron einen leistungsfähigen 64-Bit-Prozessor auf den Markt gebracht. Um diesen Prozessor zu testen und Erfahrungen mit der 64-Bit-Architektur zu sammeln, wurde der Cluster im Mai 2004 um vier Knoten mit je zwei Opteron-Prozessoren 246 (2,0 GHz) erweitert.

Nach der ursprünglichen Planung hätte die dritte Ausbaustufe im Herbst 2004 erfolgen sollen. Es stellte sich jedoch bald heraus, dass zu diesem Zeitpunkt mit den zur Verfügung stehenden Budgetmitteln nur ein mäßig attraktiver Ausbau – etwa 40% mehr Durchsatz – möglich gewesen wäre. Daher wurde beschlossen, auf die „Frühjahrskollektionen“ der Hersteller zu warten und den Ausbau ins Jahr 2005 zu ver-



Schrödinger III, Außenansicht

schieben. Auch zu diesem Zeitpunkt war es schwierig, im Rahmen des Budgets eine Lösung zu finden: Verschiedene Varianten wie z.B. ein „Schrödinger 2 1/2“ (wobei nur die Hälfte der Knoten ausgebaut und mit einem schnellen Netzwerk von Infiniband verknüpft werden sollte) wurden diskutiert und wieder verworfen. Durch zusätzliche Budgetmittel, die teils vom Zentralen Informatikdienst und teils vom Institut für Theoretische Chemie<sup>1)</sup> zur Verfügung gestellt wurden, konnte schließlich doch noch ein Ausbau des gesamten Clusters erreicht werden.

Die ersten Knoten wurden bereits im Mai 2005 geliefert, wurden aber noch lange und gründlich getestet, sodass der eigentliche Umbau erst vom 19. Juli bis 2. August stattfand. Nach einem zweiwöchigen Probetrieb wurde der Cluster für abschließende Tests am 17. August noch einmal abgeschaltet; am 24. August 2005 begann schließlich der Produktionsbetrieb von „Schrödinger III“.

## Schrödinger III

Die bisherige Architektur des Clusters (192 Rechenknoten, vernetzt mit Gigabit-Ethernet) wurde praktisch unverändert

1) Das Institut für Theoretische Chemie finanziert 16 Knoten, die nur diesem Institut zur Verfügung stehen.



Schrödinger III, Innenansicht

beibehalten. Es wurden jedoch fast alle Komponenten ausgetauscht:

- **Prozessor:** Die Opteron-Prozessoren haben im Test nicht überzeugt, weshalb weiterhin Intel-Prozessoren eingesetzt werden. Der Intel Pentium 4 processor 840 gehört einer neuen Generation an, die unter dem Codenamen „Prescott“ bekannt ist und gegenüber dem Vorgänger („Northwood“) entscheidend verbessert wurde. Die Taktfrequenz beträgt 3,2 GHz (bisher 2,53 GHz), und auch sonst ist er in allen Komponenten leistungsfähiger: Der Prozessor hat 2 MB (statt bisher 512 KB) Level 2-Cache (schneller Zwischenspeicher) und einen schnelleren *Front Side Bus* (800 MHz statt bisher 533 MHz), über den die Daten zum Motherboard und zum Hauptspeicher transferiert werden. Der Pentium 840 ist ein 64-Bit-Prozessor: Der Hauptvorteil dieser EM64T-Technologie ist die Möglichkeit, sehr großen Speicher zu adressieren. Für die Rechenknoten ist dies zwar unerheblich, aber auch die Performance mancher Applikationen kann von der 64-Bit-Architektur profitieren.
- **Hauptspeicher:** Die Hauptspeicher-Ausstattung aller Knoten wurde verdoppelt (2 GB statt bisher 1 GB).
- **Plattenplatz:** Auf den beiden Fileservern standen bisher knapp 1,2 TB (Terabyte) für permanente Daten sowie ein weiteres Terabyte für temporäre Daten (*Scratch Space*) zur Verfügung. Der *Scratch Space* auf den File-

servern wurde aufgelassen und der dadurch gewonnene Plattenplatz für permanente Daten bereitgestellt. Für temporäre Daten wurde ein eigener Scratch-Server mit 1,6 TB Plattenkapazität angeschafft. Auch die einzelnen Knoten haben nun eine größere lokale Platte (80 statt bisher 30 GB).

- **Netzwerk:** Schon bei Schrödinger II waren alle Knoten untereinander mit Gigabit-Ethernet verbunden. Ein Engpass war jedoch der Switch (Cisco 4500): Zwar konnte damit jeder einzelne Knoten zu jedem beliebigen anderen eine Verbindung mit der vollen Bandbreite von 1 Gbit/s aufbauen; wenn viele Verbindungen gleichzeitig aktiv waren, stand jeder einzelnen aber nur mehr ein Bruchteil der Bandbreite zur Verfügung. Herzstück von Schrödinger III ist ein wesentlich leistungsfähigerer – und teurerer – Cisco 6500, mit dem auch dann noch fast die theoretische Bandbreite erreicht werden kann, wenn alle Knoten simultan miteinander kommunizieren.
- **Software:** Für Schrödinger III wurden auch einige neue Software-Produkte angeschafft: Der Fortran-Compiler von PathScale erzeugt Code, der für 64-Bit-Architektur optimiert ist, und mit *gridMathematica* ist die Parallelversion der bekannten Software *Mathematica* verfügbar.

Nachdem die Gehäuse der neuen Knoten schmaler sind als die bisherigen, finden auf einem Regal fünf statt vier Knoten Platz. Dadurch wurde die Möglichkeit geschaffen, auf der gleichen Stellfläche 48 Knoten mehr unterzubringen. Diese Möglichkeit blieb nicht lange ungenutzt: Noch während der Testphase im Juni kam seitens des Instituts für Mathematik das Angebot, weitere 48 Knoten zu finanzieren, welche gleichzeitig mit den übrigen Knoten geliefert und eingebaut wurden. Auch die oben erwähnten vier Knoten mit Opteron-Prozessoren bleiben weiterhin Bestandteil des Clusters.

Gegenüber Schrödinger II wurde die Rechenleistung mehr als verdoppelt: Beim LINPACK-Benchmark erreicht Schrödinger III eine Rechenleistung von 1,14 Teraflop/s<sup>2)</sup> (unter Verwendung aller 240 Knoten, inklusive der für das Institut für Theoretische Chemie bzw. Institut für Mathematik reservierten). Schrödinger I brachte es auf 204,5 Gigaflop/s, Schrödinger II erreichte 504,3 Gigaflop/s.

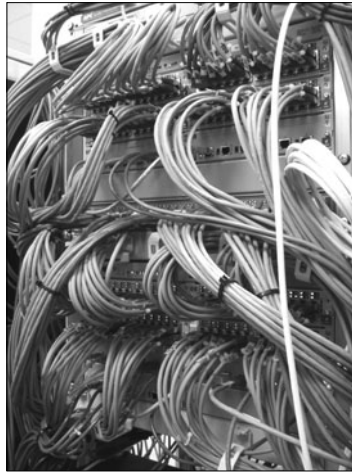
## Rückblick und Ausblick

Das Projekt „Schrödinger“ war in vielen Bereichen außerordentlich erfolgreich:

- Die wichtigsten Erfolge sind selbstverständlich die wissenschaftlichen. Zahlreiche Projekte aus unterschiedlichen

2) Flop/s steht für *Floating point operations per second*, Rechenoperationen pro Sekunde. Ein Teraflop/s entspricht einer Billion ( $10^{12}$ ) Operationen. Beim LINPACK-Benchmark, auf dem die Liste der 500 schnellsten Supercomputer beruht, wird die Rechengeschwindigkeit beim Lösen eines linearen Gleichungssystems gemessen.

Gebieten der Naturwissenschaften, Mathematik und Informatik konnten mit Hilfe von „Schrödinger“ durchgeführt werden, viele Publikationen sind in renommierten Fachzeitschriften erschienen. Ein Überblick über die wissenschaftlichen Ergebnisse ist unter [www.univie.ac.at/qccd/hpsc/hpsc\\_main.htm](http://www.univie.ac.at/qccd/hpsc/hpsc_main.htm) zu finden.



**Verkabelung an den zentralen Switches von Schrödinger III**

- Der „österreichische Supercomputer“ fand beachtliches Interesse in den Medien (siehe [www.univie.ac.at/ZID/schroedinger-doku/](http://www.univie.ac.at/ZID/schroedinger-doku/)). Während über die Uni Wien sonst häufig etwas zu „überfüllten Hörsälen“ und „Chaos bei der Prüfungsanmeldung“ in den Zeitungen zu lesen ist, hat die durchwegs positive Berichterstattung über den Cluster zum Ansehen der Universität Wien in der Öffentlichkeit beigetragen.
- *Outsourcing* ist im IT-Bereich weit verbreitet, aber keineswegs immer erfolgreich. Aufgrund der geringen Personalressourcen, die der ZID in dieses Projekt investieren konnte, war weitgehendes Outsourcing unerlässlich: Fast alle laufenden Administrations- und Wartungsarbeiten sowie der Benutzer-Support wurden von der Firma [init.at](http://init.at) durchgeführt. Die Zusammenarbeit zwischen der Universität Wien und [init.at](http://init.at) hat reibungslos funktioniert und damit viel zum Gelingen des Projekts beigetragen.
- Im Supercomputing ist Know-how ebenso wichtig wie Rechenleistung. Durch den Betrieb wurde viel Know-how gewonnen und Erfahrungen gesammelt, die auch anderen Projekten zugute kommen: An etlichen Instituten (z.B. Experimentalphysik) wurden kleinere Cluster errichtet, die in vielen Punkten Schrödinger ähnlich sind.

Obwohl das Supercomputing-Projekt offiziell beendet ist, wird der Support-Vertrag mit der Firma [init.at](http://init.at) bis auf weiteres verlängert. Dadurch ist sichergestellt, dass Schrödinger III weiterbetrieben werden kann, solange es technisch sinnvoll ist. Aufgrund des raschen Fortschritts in der Computertechnik ist mit einer Lebenszeit von etwa drei Jahren zu rechnen.

Ob und in welcher Form es ein Nachfolgeprojekt geben wird, ist noch ungewiss: Eine Finanzierung nach dem bisherigen Modell ist aufgrund des neuen Organisationsplans der

3) 1,14 Teraflop/s hätten im Juni 1997 Platz 1, im Juni 2000 Platz 5 und im Juni 2002 Platz 22 ergeben, aber für einen Platz in der derzeit gültigen Liste vom Juni 2005 nicht mehr gereicht. Durch Tuning und Optimierung wäre wahrscheinlich einer der letzten Plätze erreichbar gewesen; um den Betrieb nicht zu lange zu unterbrechen, wurde jedoch darauf verzichtet.

4) Schrödinger III verwendet die *Sun Grid Engine* als Batchsystem, das die Aufgaben („Batchjobs“) auf die einzelnen Knoten verteilt. Für die Verwaltung der Warteschlange („Batchqueue“) und eine möglichst gerechte Verteilung sorgt der *Mauji Scheduler*.

Universität Wien bedeutend schwieriger, da nun anstelle der Nawi-Fakultät deren fünf Nachfolge-Fakultäten (Physik, Chemie, Mathematik, Lebenswissenschaften und Geowissenschaften) beteiligt sind. Es muss allen Verantwortlichen klar sein, dass Supercomputing in vielen Wissenschaftszweigen ein unerlässliches Werkzeug ist und dass dafür beträchtliche Investitionen erforderlich sind. Schon die Tatsache, dass Schrödinger III nicht in der nächsten Liste der 500 schnellsten Supercomputer (November 2005) aufscheinen wird<sup>3)</sup>, illustriert, dass bedeutend mehr Anstrengungen erforderlich sind, um international nicht den Anschluss zu verpassen.

Es wäre auch denkbar, dass in Zukunft kein zentraler Supercomputer mehr betrieben wird und Rechenleistung nur mehr in Form von Clustern oder sonstigen Rechnersystemen erbracht wird, die von den einzelnen Instituten und Arbeitsgruppen betreut werden.

Für eine zentrale Lösung spricht:

- Die Kosten für die Infrastruktur (Stellfläche, Strom, Klimatisierung) werden oft unterschätzt. Ein Rechenzentrum kann diese Infrastruktur meistens günstiger zur Verfügung stellen und hat auch viel Erfahrung im Betrieb von großen Rechnersystemen (Datensicherung usw.).
- Der Betreuungsaufwand für ein einzelnes großes System ist in Summe geringer.
- Für große Rechenaufgaben können temporär sämtliche Ressourcen eines zentralen Systems zur Verfügung gestellt werden. Damit können komplexere Probleme gelöst werden, als es mit kleineren Rechnern möglich ist.

Auch eine dezentrale Lösung hat Vorteile:

- Verschiedene Arbeitsgruppen haben unterschiedliche Anforderungen, daher kann ein Rechnersystem, das nur von einer Gruppe genutzt wird, leichter an individuelle Bedürfnisse angepasst werden.
- Ein zentrales System braucht eine aufwendige Ressourcenverwaltung (z.B. ein Batchsystem mit einem ausgefeilten Scheduling-Algorithmus<sup>4)</sup>), um die Ressourcen gerecht zu verteilen. Dezentrale Systeme, die nur von wenigen Mitgliedern einer Arbeitsgruppe genutzt werden, kommen hier mit wesentlich einfacheren Mitteln aus.

Welche Lösungen auch gewählt werden: Beträchtlicher finanzieller und personeller Aufwand ist in jedem Fall erforderlich. Die attraktivste Lösung ist wohl eine Kombination aus einem zentralen Supercomputer und kleineren dezentralen Systemen.

Peter Marksteiner ■

# ALTERNATE DATA STREAMS: „TROJANISCHE PFERDE“ VON MICROSOFT?

Woher weiß Ihr Rechner, welches Programm er starten soll, wenn Sie eine bestimmte Datei aktivieren? Richtig: Eine Funktion des Betriebssystems sorgt für die korrekte Zuordnung. Das marktbeherrschende Betriebssystem MS-Windows verwendet dafür (wie sein Vorgänger MS-DOS) so genannte Assoziationen zwischen dem jeweiligen Anwendungsprogramm und dem Dateityp, der über die Erweiterung des Dateinamens definiert ist. Beispielsweise bedeutet `xy.exe`, dass es sich um eine ausführbare Datei handelt, `xy.doc` signalisiert ein MS-Word-Dokument usw. Auf einem Macintosh hingegen wird diese Information unabhängig vom Dateinamen gespeichert: Unter MacOS besteht jede Datei aus zwei parallelen Datenströmen – der erste Strom beschreibt die so genannten Ressource-Informationen (dazu gehört u.a. der Typ der Datei), der zweite enthält die eigentlichen Daten.

Diese grundlegend unterschiedliche Behandlung der Ressource-Informationen erschwerte lange Zeit den direkten Datenaustausch zwischen Windows- und Macintosh-Rechnern. Mit der Einführung des Dateisystems NTFS 5 präsentierte Microsoft dann eine Lösung für das Problem: Dank *Alternate Data Streams*, kurz ADS, konnten endlich auch Apple-BenutzerInnen auf so genannte *Windows Shares* (an andere Rechner exportierte Verzeichnisse oder virtuelle Festplatten) zugreifen. Jede Datei, jedes Verzeichnis, aber auch jede Verknüpfung besteht im – unter Windows heute einzig sinnvollen – Dateisystem NTFS aus einem Hauptdatenstrom und optional mehreren alternativen Datenströmen.

Betrachten wir beispielsweise die Datei `hugo.txt` (Abb. 1). Unter Windows wird der Hauptdatenstrom mit `hugo.txt` angesprochen; weitere alternative Datenströme der Datei sind, durch einen Doppelpunkt getrennt, diesem Namen angefügt. Der erste alternative Datenstrom in unserem Beispiel ist damit `hugo.txt:strom1`, der zweite `hugo.txt:strom2` usw. Dabei kann `strom1` oder `strom2` jeder beliebige Name sein, er darf nur keinen Doppelpunkt enthalten. Verbirgt sich in einem alternativen Datenstrom ein ausführbares Programm (z.B. `hugo.txt:calc.exe`), könnte es aus der Sicht von Windows jederzeit gestartet werden.

## Warum „Trojaner“?

Microsoft hat die Existenz von Alternate Data Streams zwar in den technischen Unterlagen zum Betriebssystem Windows dokumentiert, es jedoch verabsäumt, ein Programm zur Verfügung zu stellen, das solche „Anhängsel“ auflistet. Weder mit dem Windows-Explorer noch mit dem Kommando `dir` lassen sich alternative Datenströme anzeigen. Auch die Anzeige der Dateigröße richtet sich unter Windows nur nach dem Hauptdatenstrom – ADS werden nicht

mitgerechnet. Selbst die „Quota“, mit der einzelnen BenutzerInnen eine Obergrenze an belegbarem Plattenplatz zugewiesen werden kann, ignoriert alle alternativen Datenströme. Für den Windows-Benutzer bedeutet dies, dass er ADS mit den standardmäßig verfügbaren Werkzeugen nie zu sehen bekommt.

Viele Inhalte von alternativen Datenströme sind durchaus sinnvoll und legitim. Beispielsweise speichert Windows XP Miniaturbilder (*Thumbnails*) von Grafiken in ADS, und der Virens Scanner von Kaspersky Lab reduziert seinen Scan-Aufwand, indem er an jede von ihm durchsuchte virenfreie Datei einen Datenstrom mit einer Kennung anfügt. Dennoch können sich – z.B. nach einer Hacker- oder Wurmattacke – hinter jeder harmlosen, vom Benutzer selbst erstellten Datei auch durchaus gefährliche Programme verbergen.

Erschwerend kommt hinzu, dass die bloße Existenz von Alternate Data Streams der überwiegenden Mehrzahl der BenutzerInnen (und auch vielen AdministratorInnen) nicht bekannt ist. Hinsichtlich der Anwendungen, die mit ADS umgehen können, herrscht in Windows ein heilloses Durcheinander: Es gibt viele Programme, die mit dem zusätzlichen Doppelpunkt im Dateinamen nichts anfangen, und sogar einige Virens Scanner, die alternative Datenströme nicht durchsuchen.

## Wissen ist der beste Schutz

Schon die Tatsache, dass Microsoft relativ wenig Informationen über Alternate Data Streams zur Verfügung stellt, macht sie gefährlich. Wie die bekannteren Risikofaktoren am PC kann auch die potentielle ADS-Bedrohung nur durch richtiges Verhalten reduziert werden; ein geübter Blick im Erkennen von Unregelmäßigkeiten ist dabei hilfreich.

Nehmen wir dazu ein einfaches Beispiel: Angenommen, der Datei `calc.exe` (das ist der Windows-Taschenrechner) wurde als alternativer Datenstrom die Datei `cmd.exe` (die

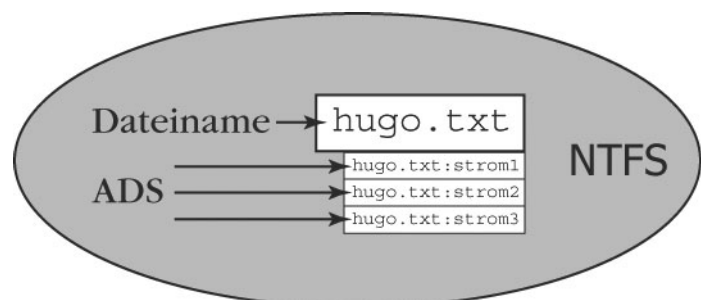


Abb. 1: Alternate Data Streams verstecken sich im Dateisystem NTFS hinter dem eigentlichen Dateinamen.

Windows-Eingabeaufforderung, auch bekannt als *Ausführen* oder „DOS-Box“) angefügt. Sowohl `calc.exe` als auch `calc.exe:cmd.exe` sind ausführbare Programme. Betrachtet man die Datei `calc.exe` im Windows-Explorer, so bleibt ihr alternativer Datenstrom unsichtbar. Auch unter den Eigenschaften der Datei, die durch Klick mit der rechten Maustaste aufgerufen werden, ist nichts darüber in Erfahrung zu bringen. Beim Starten bzw. Verwenden des Taschenrechners bemerkt man die zusätzlichen Daten ebenfalls nicht. Wird jedoch der alternative Datenstrom – also `cmd.exe`, die DOS-Box – ausgeführt (z.B. mittels `start`-Befehl aus einer anderen DOS-Box), ist oft erkennbar, dass es sich um ein via ADS aufgerufenes Programm handelt.

Windows selbst bietet zwei Möglichkeiten, den alternativen Datenstrom zu enttarnen. Beide erreicht man über den Windows Task-Manager, welcher durch Drücken der Tastenkombination `<Strg>+<Alt>+<Del>` aktiv wird. Auf den Registerkarten *Anwendungen* und *Prozesse* ist hier die tatsächliche Herkunft der aktiven DOS-Box (`calc.exe:cmd.exe`) ersichtlich. Der Haken: Ein Doppelpunkt im Dateinamen ist zwar ein sicherer Hinweis auf ADS; fehlt der Doppelpunkt, könnte das Programm aber trotzdem über ADS gestartet worden sein, da sowohl der Anwendungs- als auch der Prozessname mit geeigneter Programmierung änderbar sind.

Um Alternate Data Streams einigermaßen zuverlässig zu finden zu können, benötigt man deshalb ein entsprechendes Zusatzprogramm – z.B. den *StreamExplorer* ([www.rekenwonder.com/streamexplorer.htm](http://www.rekenwonder.com/streamexplorer.htm)), den *ADS Locator* von Spybot ([www.safer-networking.org/de/paragraphs/tools\\_ads.html](http://www.safer-networking.org/de/paragraphs/tools_ads.html)) oder das mittels Befehlszeile gesteuerte Programm *lads* (steht für *list ADS*, [www.heysoft.de/ep-lads.htm](http://www.heysoft.de/ep-lads.htm)). Alle angeführten Programme sind Freeware.

Der StreamExplorer zeigt beispielsweise zu jeder Datei die Anzahl ihrer alternativen Datenströme sowie deren Name, Dateigröße und Inhalt (siehe Abb. 2). Aus der Abbildung ist ersichtlich, dass jede Datei in NTFS mindestens zwei Datenströme hat: Im Datenstrom `<default>` (EDV-Jargon für „standardmäßig“) ist der eigentliche Inhalt der Datei zu finden; der Datenstrom `<no name>` enthält gewisse Dateiattribute. An die Datei `test1.dat` wurde zusätzlich der alternative Datenstrom `nochEinStrom` angefügt, der mit 33 Bytes größer ist als der Hauptdatenstrom (21 Bytes).

## Was hilft gegen ADS?

Manche Windows-Standardfunktionen sind ADS-tauglich, andere nicht. Beim Kopieren von Dateien (sowohl mit der Kopierfunktion als auch mit dem Befehl `copy`) werden alle alternativen Datenströme mitkopiert und bleiben erhalten. Auch Notepad, der Standard-Editor von Windows, kann mit ADS umgehen – allerdings nur auf Umwegen: Wenn Sie in Notepad unter *Datei – Öffnen* einen ADS-Namen eingeben, erhalten Sie eine Fehlermeldung, dass der Dateiname nicht gültig ist. Öffnen Sie die Datei aber über die DOS-Box (z.B. mit dem Befehl `notepad test1.dat:nochEinStrom`), so wird der alternative Datenstrom im Editor angezeigt.

Der Befehl `type` in der DOS-Box, der den Inhalt einer Datei am Bildschirm ausgibt, ist hingegen nicht ADS-fähig – er spricht stets nur den Hauptdatenstrom an. Daher kann diese Funktion zum Entfernen von ADS verwendet werden: Lenkt man die Ausgabe statt auf den Bildschirm in eine neue Datei um, wird nur der Hauptdatenstrom übernommen. Ausgehend vom Beispiel in Abb. 2 muss man also in der DOS-Box den Befehl `type test1.dat > test2.dat` eingeben, dann `test1.dat` löschen und `test2.dat` wieder in `test1.dat` umbenennen, um die zusätzlich angehängten Datenströme zu entfernen.

Auch wenn der alternative Datenstrom an einem Verzeichnisnamen hängt, lässt er sich im Allgemeinen (d.h. wenn man der Eigentümer des Verzeichnisses ist) unkompliziert entfernen: Zunächst legt man einen neuen Ordner an und gibt diesem die gleichen Berechtigungen, die auch das „ADS-befallene“ Verzeichnis hat. Dann kopiert man alle Dateien und Unterverzeichnisse des Original-Verzeichnisses in das neue. Anschließend wird das alte Verzeichnis gelöscht und das neue wieder entsprechend umbenannt.

**Achtung:** Bitte versuchen Sie das *nicht* mit dem Verzeichnis `C:\WINDOWS\` und dessen Unterverzeichnissen! Dies würde unent-

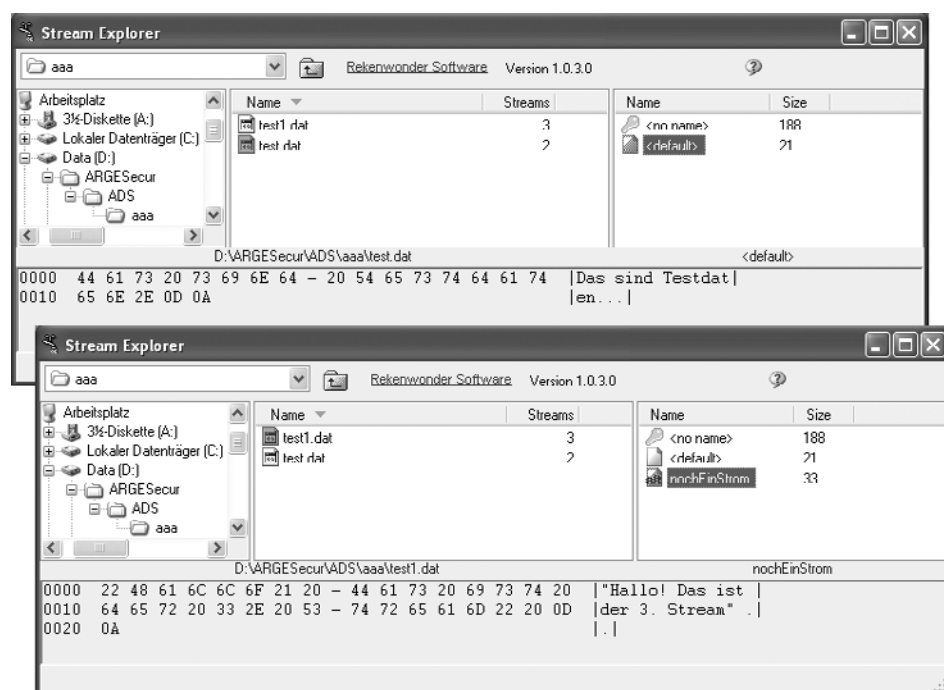


Abb. 2: StreamExplorer-Ansichten von Dateien mit zwei (oben) bzw. mit drei (unten) Datenströmen

behrliche Systemprogramme in Mitleidenschaft ziehen – das Betriebssystem wird instabil bzw. kann unter Umständen irreparabel beschädigt werden. Auch im obersten Verzeichnis einer Festplatte (z.B. D:\) ist dieses Verfahren nicht anwendbar.

## Viren und ADS

Obwohl bereits Viren aufgetreten sind, die sich in alternativen Datenströmen verstecken, werden ADS noch immer nicht von allen Virenschaltern etc. berücksichtigt. Manche – z.B. die Ad-Aware SE Personal Software 1.06, die unter [www.lavasoft.de](http://www.lavasoft.de) als Freeware erhältlich ist – beherrschen das Suchen von digitalem Ungeziefer in Alternate Data Streams hervorragend, viele andere finden Viren in ADS zumindest beim direkten Zugriff auf die Virus-Daten (d.h. wenn diese aktiviert oder auch nur angezeigt werden). Eine unrühmliche Ausnahme ist Norton Internet Security 2004.

Der McAfee VirusScan, der an der Uni Wien als Gratissoftware angeboten wird (siehe [www.univie.ac.at/ZID/gratissoftware/](http://www.univie.ac.at/ZID/gratissoftware/)), ist voll ADS-tauglich. Er entdeckt Viren in ADS beim Zugriff auf die Datei selbst dann, wenn nur der Hauptdatenstrom angesprochen wurde, und wird auch beim Durchsuchen der Festplatten fündig. Wesentlich dafür ist jedoch, dass Sie den Virenschalter so einstellen, dass er immer *alle* Dateien<sup>1)</sup> untersucht: Da Windows beim Erkennen der Funktion einer Datei auf Assoziationen setzt, die jedoch relativ einfach verändert werden können, sind ausführbare Programme (z.B. Viren) nicht zwingend an der Dateieindung .exe erkennbar. Den Virenschalter nur nach .exe-Dateien suchen zu lassen, greift daher zu kurz.

Mit Vorsicht zu genießen ist die Koppicillin-CD, die von AdministratorInnen gern verwendet wird, um Windows im abgeschalteten Zustand nach Viren und Trojanern zu durchsuchen. Dies ist insbesondere hilfreich, wenn man befürchtet, dass sich auf dem System ein aktives Rootkit befindet, ein Hacker-Werkzeug, das sich Windows gegenüber unsichtbar macht.<sup>2)</sup> Aufgrund der Konstruktion der Knoppicillin-CD findet jedoch keiner der dort verfügbaren Scanner Viren in alternativen Datenströmen. Damit bleiben Rootkits, die in ADS versteckt sind, mit diesem Werkzeug unauffindbar.

## Vom Wurm befreit ...

... ist nicht unbedingt geheilt! Wer kennt schon alle Detailfunktionen des Wurms, der soeben vom System entfernt wurde? Welche Programme wurden mit Hilfe des mittlerweile

1) Wählen Sie dazu (Pfad je nach Aufruf): *VirusScan Konsole – Bei-Zugriff-Scanner* (bzw. *Alle stationären Datenträger*) – *Eigenschaften* – (Alle Vorgänge –) *Entdeckung* – *Alle Dateien*

2) siehe *Comment 04/1*, Seite 11 bzw. [www.univie.ac.at/comment/04-1/041\\_10.html#rootkits](http://www.univie.ac.at/comment/04-1/041_10.html#rootkits)

3) siehe *Comment 04/1*, Seite 20 bzw. [www.univie.ac.at/comment/04-1/041\\_20.html](http://www.univie.ac.at/comment/04-1/041_20.html)

eliminierten Trojaners in das System geschmuggelt? Diese Fragen und mehr sollten alle diejenigen beschäftigen, die sich eben aus der Umklammerung eines digitalen Ungezieters befreit haben. Dank Rootkits, ADS und anderen Eigenheiten ist es unter Windows immer möglich, dass ein „Erbstück“ auf einem kompromittierten System zurückbleibt. Ein einmal gehacktes oder auch nur von einem digitalen Schädling befallenes System ist nie mehr wirklich sicher. Ein in Alternate Data Streams verborgenes, aktives Rootkit lässt sich nur mit großem Aufwand enttarnen – auch ADS-taugliche Scanner wie McAfee sind den Tricks eines Rootkits hilflos ausgeliefert, sobald dieses erst einmal läuft.

Deshalb ist Vorbeugen stets besser als Heilen und Sicherheit von Anfang an<sup>3)</sup> die beste Lösung: Ein immer aktueller, immer aktiver, ADS-tauglicher und gut konfigurierter Virenschalter, eine ebenfalls von Beginn an aktive Personal Firewall (die in Windows XP integrierte Firewall genügt meist), regelmäßige ADS-Routinekontrollen sowie ein generell defensives Verhalten sparen langfristig viel Zeit und Mühe.

Aron Vrtala ■

## Neue Standardsoftware

([www.univie.ac.at/ZID/standardsoftware/](http://www.univie.ac.at/ZID/standardsoftware/))

### Neue Produkte (Stand: 15. 9. 2005)

- Adobe Acrobat 7 Prof. für Win. und Mac
- Adobe After Effects Prof. 6.5 für Win. und Mac
- Adobe Audition 1.5 für Win.
- Adobe Encore DVD 1.5 für Win.
- Adobe GoLive CS2 8.0 für Win. und Mac
- Adobe Illustrator CS2 12.0 für Win. und Mac
- Adobe InDesign CS2 4.0 für Win. und Mac
- Adobe Photoshop CS2 9.0 für Win. und Mac
- Adobe Photoshop Elements 3.0 für Win. und Mac
- Adobe Premiere Prof. 1.5 für Win.
- Apple iWork '05 1.0 für Mac
- Apple MacOS X 10.4
- Apple QuickTime Pro 7 für Win. und Mac
- Corel Designer Technical Suite 12 für Win.
- InfoZoom Prof. 3.80 für Win.
- MS Digital Image Suite 2006 für Win.
- MS Windows 2003 Server Standard x64 Edition
- MS Windows XP Prof. x64 Edition
- **SPSS 13.0** für Win. (Achtung: Wegen eines neuen Vertrags und der geänderten Lizenzierung ist dieses nicht als Gratis-Update zu SPSS 12 verfügbar, sondern muss bei Bedarf neu bestellt werden!)
- VMware Workstation 5 für Win.

### Updates (Stand: 15. 9. 2005)

- SAS 9.1.3 TS1M3 für Win. (bisher 8.2)

Peter Wienerroither

# DIGITALE DOKUMENTE AUSTAUSCHEN – ABER RICHTIG!

Das Lesen bereits gedruckter Texte erfordert – sofern eine geeignete Schriftart und -größe gewählt wurde – keine technischen Hilfsmittel. Zum Lesen bzw. Verarbeiten eines Word-Dokuments, einer JPEG-Grafik oder einer PDF-Datei (um nur drei Vertreter aus dem gewaltigen Zoo von Text-, Grafik- und Druckbeschreibungsformaten zu nennen) wird jedoch immer das Hilfsmittel Computer inklusive geeigneter Software benötigt. Zudem kämpft die Verarbeitung digitaler Daten mit dem noch erheblicheren Problem der Kurzlebigkeit: Neue Programmversionen mit zusätzlichen Funktionen bewirken Änderungen in den Dateiformaten. So lassen sich z.B. die Features von Acrobat 7 (= PDF-Format Version 1.6) mit einem PDF-Reader der Version 6 oder älter nicht darstellen. Ähnliches gilt für alle Softwareprodukte.

Solange Sie nur Ihre eigenen Dokumente auf Ihrem eigenen Rechner bearbeiten, gibt es meist keine Probleme: Eine ältere Datei ist üblicherweise mit einer neueren Version des entsprechenden Programms lesbar. Umgekehrt funktioniert das leider nicht – je nach Software ist die Datei nicht mehr zu öffnen oder die neuen Features werden lediglich nicht dargestellt. Durch unseren hohen Grad der Vernetzung ist das Austauschen von elektronischen Daten aber mittlerweile tägliches Geschäft. Dazu müssen die Dateien von unterschiedlichen Betriebssystemen und verschiedenen Programmversionen kompatibel verarbeitet werden können. Der Haken dabei: Die Zahl der verfügbaren Dateiformate und ihrer Versionen ist heute größer als die Anzahl der Sprachen und Dialekte der Welt.

## Dateien tauschen – aber wie?

### Freiwillige Selbstbeschränkung

Wenn Sie z.B. einem Journal eine Publikation in digitaler Form übermitteln wollen, erhalten Sie üblicherweise in den Informationen für AutorInnen des Verlags Hinweise darüber, welche Dateiformate zulässig sind. So wie Sie von einem Verlag nicht erwarten können, dass er für Sie alle denkbaren Formate handhabt, können Sie auch vom Empfänger eines elektronischen Dokuments nicht erwarten, dass er den aktuellsten Reader für ein bestimmtes Dateiformat installiert hat bzw. zu dessen Installation bereit ist (vielleicht ist er nicht einmal in der Lage, ihn zu installieren). Der Austausch digitaler Dokumente erfordert daher immer das Einhalten gewisser Mindeststandards – damit beispielsweise Acrobat 5 eine PDF-Datei lesen kann, darf sie eben keine *Dernier Cri-* Funktionen der Version 7 beinhalten.

Zusätzlich müssen alle Elemente der Datei, die nicht auf allen Systemen als verfügbar vorausgesetzt werden können, im Dokument mittransportiert werden („eingebettet sein“).

Das betrifft in erster Linie Grafiken und Schriftarten (engl. *Fonts*): Standardmäßig werden diese nicht eingebettet, um die Dateigröße möglichst gering zu halten und damit die Verarbeitungsgeschwindigkeit zu erhöhen. Insbesondere wenn Dateien an Druckereien bzw. Drucksysteme übermittelt werden (siehe dazu auch Artikel *Onlinedruck: Die Hausdruckerei per Mausclick* auf Seite 29), müssen die verwendeten Fonts und Grafiken aber inkludiert sein, damit die Dokumente in der gewünschten Form ausgedruckt werden können.

### Welche Formate sind sinnvoll?

Derzeit dominiert Microsoft die PC-Welt. Bei **MS-Office**-Formaten (Word, PowerPoint usw.) kann man daher in der Regel davon ausgehen, dass die EmpfängerInnen die Datei verarbeiten können. Daneben hat sich auch das *Portable Document Format (PDF)* der Firma Adobe durchgesetzt, das eigens für den sicheren und verlässlichen Austausch von elektronischen Dokumenten entwickelt wurde. Im Druckwesen ist außerdem das Format **PostScript** seit vielen Jahren Standard. Selbstverständlich existieren auch zahlreiche Dateiformate, die bei bestimmten Benutzergruppen sehr beliebt sind und innerhalb dieser Gruppen daher meistens unkompliziert verschickt werden können – z.B. die Textverarbeitungen TeX und LaTeX unter NaturwissenschaftlerInnen. Im Folgenden werden die für einen problemlosen Datenaustausch erforderlichen Einstellungen jedoch nur anhand der drei oben genannten, am weitesten verbreiteten Formate beschrieben.

Als Beispiel für die prinzipielle Vorgangsweise bei den Programmen von Microsofts Office-Suite (die alle vergleichbar aufgebaut sind) dient hier MS-Word 2003 in der Windows-Variante.

## MS-Office

Ein Tipp zu Beginn: Office-Dokumente, die für die Weitergabe bestimmt sind, sollten nach ihrer Fertigstellung nochmals unter einem anderen Namen gesichert werden. Verschicken Sie stets die „neue“ Datei! Sie vermeiden damit Probleme, die durch die Schnellspeicherung und die *Rückgängig*-Funktion entstehen können – unter Umständen hat der Empfänger sonst die (meistens unerwünschte) Möglichkeit, alle Änderungen nachzuverfolgen, die Sie an dem Dokument vorgenommen haben.

### Einbetten von Fonts

Um eine Datei mit eingebetteten Schriftarten abzuspeichern, wählen Sie in MS-Word das Menü **Extras – Optionen**

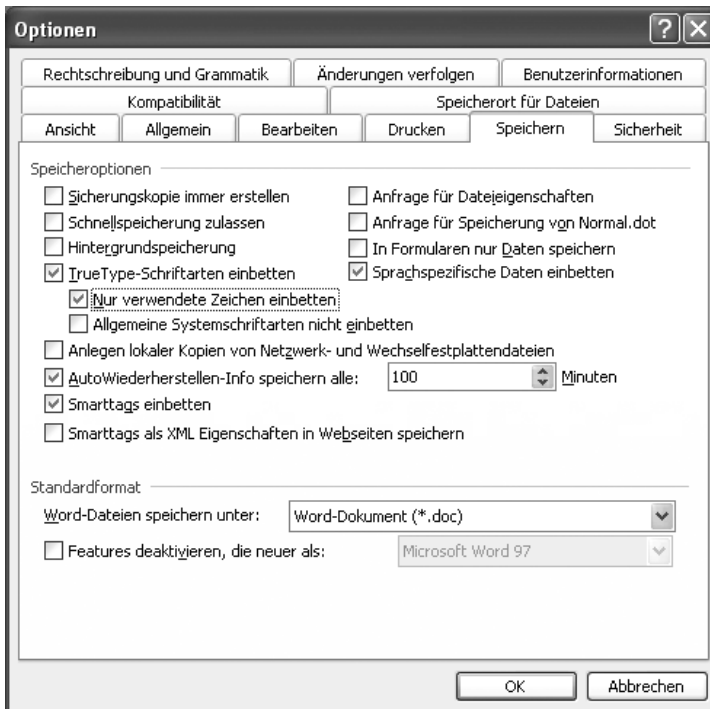


Abb. 1: Speichern mit eingebetteten Schriftarten (MS-Word)

und dann die Registerkarte **Speichern** (siehe Abb. 1). Dort aktivieren Sie **TrueType-Schriftarten einbetten** sowie (zum Zweck der Plattenplatzeinsparung) die Unteroption **Nur verwendete Zeichen einbetten**. Vorsicht: Diese bewirkt, dass die EmpfängerInnen der Datei keine wesentlichen Änderungen am Dokument vornehmen können, weil nicht alle Schriftzeichen enthalten sind. Soll die Datei für andere Personen bearbeitbar bleiben, dürfen Sie in Word diese Option nicht anklicken (bei MS-PowerPoint müssen Sie in diesem Fall die Option **Alle Zeichen einbetten** aktivieren). Abschließend überprüfen Sie noch die Option **Smarttags einbetten**, die aktiviert sein sollte, damit alle Funktionen des Dokuments verfügbar sind. Wenn Sie sichergehen wollen bzw. wenn die EmpfängerInnen ältere Windows-Versionen verwenden, können Sie in MS-Word zusätzlich von der Option **Features deaktivieren, die neuer als:** Gebrauch machen (in MS-PowerPoint würden Sie dann unter *PowerPoint-Dateien speichern als* beispielsweise *PowerPoint 95-Präsentation* auswählen).

## Verknüpfungen

Ein weiterer Stolperstein beim Austausch von Dateien sind Verknüpfungen: In MS-Office können Teile des Dokuments – z.B. Grafiken – über das Menü **Einfügen** und den gewünschten Unterpunkt mit einer anderen Datei verknüpft werden. Dazu muss im Dialogfenster des entsprechenden Unterpunkts die Option **Verknüpfen** aktiviert sein. Diese bewirkt, dass alle Änderungen der externen Datei automatisch im aufgerufenen Office-Dokument sichtbar werden. Die Grafiken (oder andere Objekte) sind jedoch nur als Verknüpfung in das Dokument integriert, d.h. es wird

nicht der Inhalt, sondern lediglich die Lokation der externen Datei gespeichert. Wenn Sie das Office-Dokument verschicken, so ist dieses zwar klein, aber nicht vollständig. Die EmpfängerInnen, die auf die extern gespeicherten Bilddaten nicht zugreifen können, sehen die Datei nur bruchstückhaft: Für die fehlenden Objekte ist zwar ein Rahmen in der jeweils richtigen Größe vorhanden, dieser enthält jedoch nur ein kleines rotes X.

Verknüpfte Daten lassen sich folgendermaßen in einer Datei mitspeichern: Wenn ein Dokument externe Verknüpfungen hat bzw. hatte, ist im Menü **Bearbeiten** die Funktion **Verknüpfungen** verfügbar (andernfalls ist sie grau dargestellt und kann nicht ausgewählt werden). Im Fenster **Verknüpfungen** selektieren Sie alle vorhandenen Quelldateien, indem Sie die <Shift>-Taste (für Großbuchstaben) gedrückt halten und die erste und die letzte Datei der Liste anklicken (siehe Abb. 2). Danach aktivieren Sie die Option **Grafik in Dokument speichern**. Nun klicken Sie auf **Jetzt aktualisieren**, um die aktuellen Bildinhalte nochmals von der Festplatte zu laden, und anschließend auf **OK**. Dann speichern Sie die Datei – sinnvollerweise unter einem neuen Namen. Ein Vergleich der ursprünglichen mit der neuen Dateigröße muss einen Zuwachs zeigen. Beachten Sie, dass nun Änderungen an den Quelldateien nicht mehr automatisch im Office-Dokument nachgezogen werden!

## Adobe Acrobat

Das *Portable Document Format* (PDF) der Firma Adobe ist aufgrund seiner technischen Möglichkeiten und Fortschritte mittlerweile nicht mehr so portabel, wie man es aufgrund seines Namens erwarten würde. Deswegen hat Adobe eine ISO-Norm namens PDF/X vorgelegt, die größtmögliche Kompatibilität gewährleistet. Mit dem an der Universität Wien als Standardsoftware verfügbaren Programmpaket Adobe Acrobat Professional (siehe [www.univie.ac.at/ZID/standardsoftware/](http://www.univie.ac.at/ZID/standardsoftware/)) können Sie PDF/X-Dateien erzeugen – entweder aus anderen PDF-Dateien oder durch Konversion, z.B. aus einem Word-Dokument. In allen Fällen

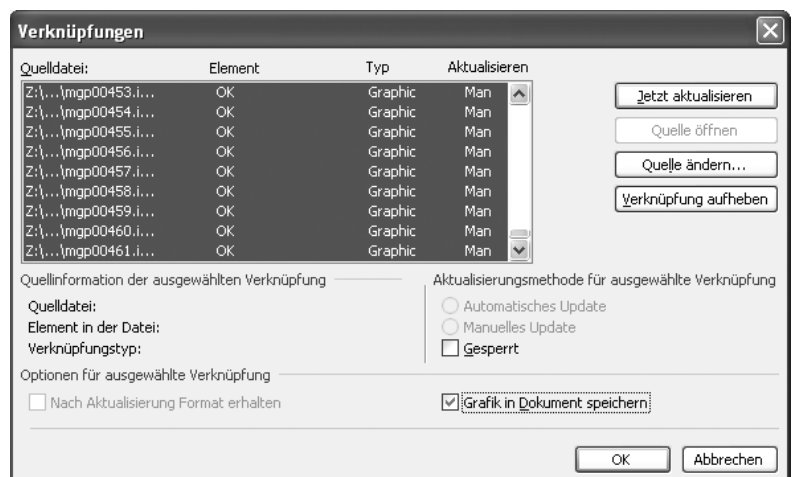


Abb. 2: Verknüpfte Dateien im Dokument speichern (MS-Word)



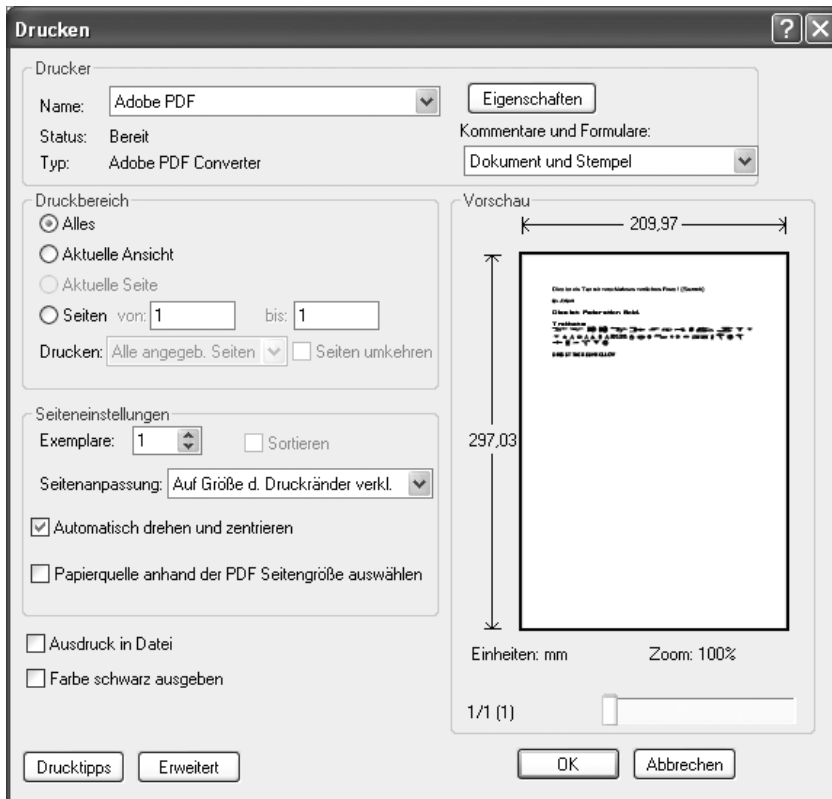


Abb. 3: Auswählen des Adobe PDF-Druckertreibers im Dialogfenster *Drucken*

führt der sicherste Weg über den Druckertreiber *Adobe PDF* und den Speicherstandard PDF/X-3:2002. Wählen Sie also in MS-Office oder in Adobe Acrobat **Datei – Drucken** und selektieren Sie aus der Liste *Name* den Drucker **Adobe PDF** (siehe Abb. 3). Anschließend klicken Sie rechts daneben auf die Schaltfläche **Eigenschaften**.

Es erscheint das Dialogfenster *Eigenschaften von Dokument „Adobe PDF“*. In der Registerkarte **Layout** können Hoch- oder Querformat ausgewählt werden. Es ist hier auch möglich, die Anzahl der Seiten pro Blatt festzulegen, was insbesondere beim Ausdruck von PowerPoint-Folien sehr nützlich sein kann. Für die Definition des PDF/X-Formats wird jedoch die Registerkarte **Adobe PDF-Einstellungen** benötigt (siehe Abb. 4).

Wählen Sie in dieser Registerkarte aus der Liste *Standardeinstellungen* die Option **PDF/X-3:2002**. Unter *Adobe PDF-Ausgabeordner* ist **Eingabeaufforderung für PDF-Dateiname** am günstigsten. Danach müssen Sie noch das Kontrollkästchen **Schriften nicht an Adobe PDF senden** deaktivieren (kein Häkchen) und auf **OK** klicken. Fertig. Sie werden nach einem Dateinamen gefragt und können das Dokument abspeichern. Wenn Sie anschließend den Menüpunkt **Datei – Dokument-eigenschaften** aufrufen, sehen Sie, dass auf der Registerkarte *Beschreibung* die ältere PDF-Version 1.3 (Acrobat 4.x) angegeben ist. Auf der Registerkarte *Schriften* sind alle Schriftarten als *Eingebettete Unter-*

gruppe deklariert. Das bedeutet, dass das Dokument nun auch die entsprechenden Schriftzeichen enthält und daher von jedem Acrobat Reader der Version 4 oder höher darstellbar ist.

Da der *Adobe PDF*-Druckertreiber auch die Grafiken neu berechnet, entsteht unter Umständen ein Verlust an Bildqualität, was zwar in den meisten Fällen kein Problem darstellt, bei größeren Darstellungen aber möglicherweise unerwünscht ist. Sie können diese Funktion deaktivieren, indem Sie neben der Liste *Standardeinstellungen* auf die Schaltfläche **Bearbeiten** klicken (siehe Abb. 4). Es erscheint ein neues Fenster, in dem Sie das Symbol *Bilder* auswählen. Schalten Sie dann die Neuberechnungen für Farb-, Graustufen- und Schwarzweißbilder und die zugehörigen Komprimierungen aus.

## PostScript

Bei PostScript (Dateityp: .ps) handelt es sich um eine technisch sehr ausgereifte Seitenbeschreibungssprache, die ausschließlich für den professionellen Druck entwickelt wurde. Das Nachbearbeiten solcher Dateien ist nur eingeschränkt möglich – z.B. mit CorelDraw. PostScript gibt es in drei Versionen, die „Level“ genannt werden: Level 1 ist veraltet; Level 2 wird derzeit von Level 3 abgelöst. Für einen möglichst kompatiblen Datenaustausch ist daher PostScript Level 2 zu empfehlen.

Ausgehend von Ihrem PDF-Druck von vorhin sind Sie nur einen Katzensprung von einem PostScript-Level 2-Doku-

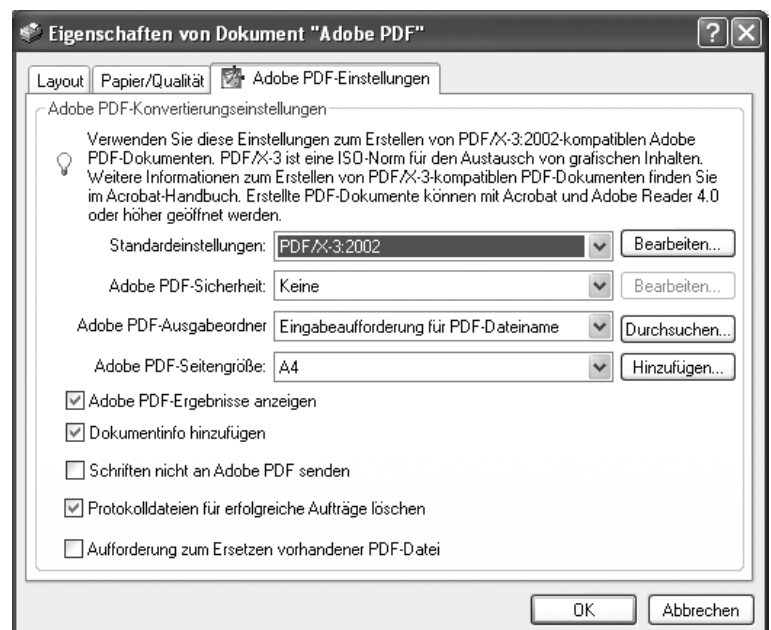


Abb. 4: Registerkarte *Adobe PDF-Einstellungen*

ment entfernt: Aktivieren Sie im Fenster *Eigenschaften von Dokument „Adobe PDF“* (aufzurufen über die Schaltfläche *Eigenschaften* rechts oben im Dialogfenster *Drucken*, vgl. Abb. 3) die Registerkarte **Layout** und klicken Sie rechts unten auf die Schaltfläche **Erweitert**. Im daraufhin erscheinenden Fenster *Erweiterte Optionen für Adobe PDF Converter* (siehe Abb. 5) öffnen Sie **Dokumentoptionen** und dann **PostScript-Optionen**. Bei *PostScript-Sprachebene* ist standardmäßig 3 eingestellt. Ändern Sie diesen Wert auf 2, indem Sie auf die Ziffer klicken; danach klicken Sie auf **OK**.

Abschließend muss der *Adobe PDF*-Druckertreiber noch angewiesen werden, eine PostScript-Datei zu erzeugen. Dies geschieht durch Aktivieren des Kontrollkästchens **Ausdruck in Datei** im Dialogfenster *Drucken* (siehe Abb. 3). Der Druckertreiber generiert dann eine PostScript-Datei, der Sie einen beliebigen Dateinamen mit der Erweiterung *.ps* zuweisen können.

Aron Vrtala ■

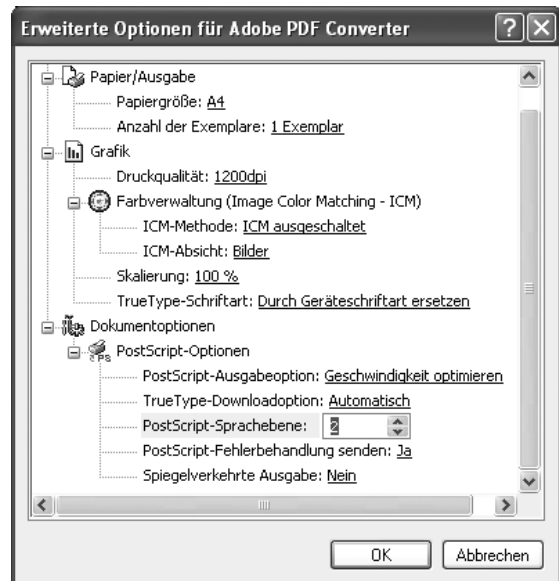


Abb. 5: *Erweiterte Optionen für Adobe PDF Converter* (Einstellen von PostScript-Level 2)

## ONLINEDRUCK: DIE HAUSDRECKEREI PER MAUSKLIICK

Seit Gutenberg ist viel geschehen. Wir drucken heute auf Laser- oder InkJet-Druckern, lokal oder über das Netzwerk. Fast auf jedem Schreibtisch steht ein Drucker – aber keine

Druckerei. An der Uni Wien ist das nun anders. Mit dem *Onlinedruck*, einem neuen Service der Hausdruckerei der Universität Wien, steht per Mausclick ein hochqualitatives

Druckereizentrum zur Verfügung, das von allen Uni-MitarbeiterInnen mit Mailbox-UserID (siehe [www.univie.ac.at/ZID/mailbox/](http://www.univie.ac.at/ZID/mailbox/)) genutzt werden kann.

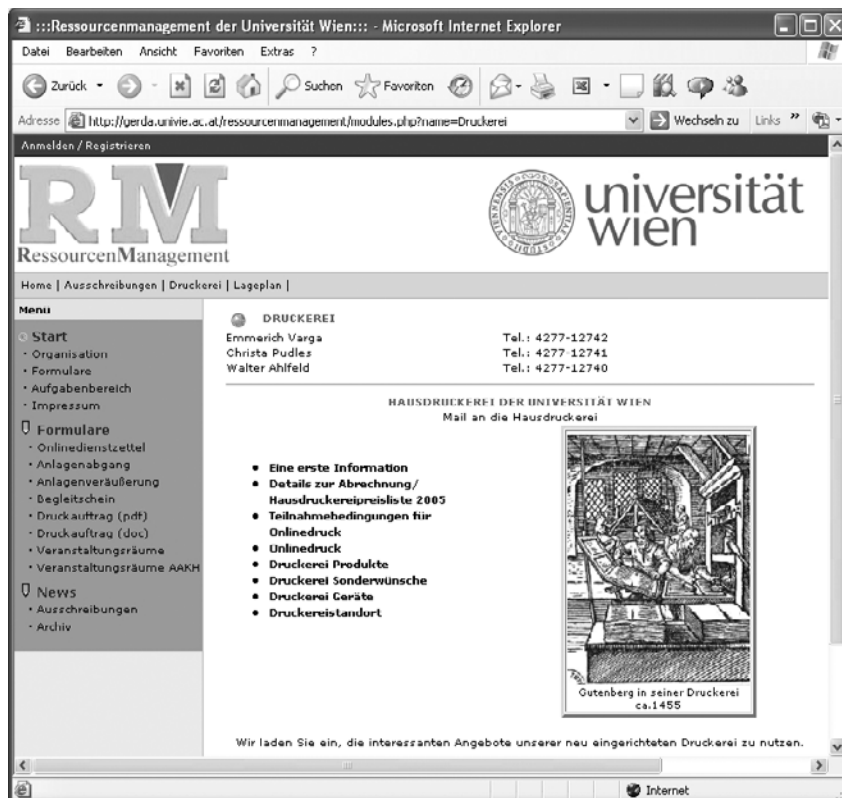


Abb. 1: Homepage der Hausdruckerei

Worin liegen die Vorteile? Ein typischer PC-Drucker ist relativ langsam. Hunderte Skripten auszudrucken benötigt viel Zeit und bringt manche Hardware an die Grenzen ihrer mechanischen Belastbarkeit. Zudem ist der lokale Drucker meist verhältnismäßig teuer im Preis pro Blatt; Herstellerfirmen erwirtschaften ihre Kosten für die Geräte mitunter erst über den Verkauf von Tintenpatronen oder Tonerkassetten. *Last but not least*: Der Instituts-Drucker oder -Kopierer hat meist nur wenige Optionen für die Endfertigung Ihres Druckwerks – die Hausdruckerei kann u.a. Falten, Falzen, Rillen, Lochen, Heften, Broschüren, Mappen binden, Spiralisieren, Leinenbinden und Bücher mit Heißkleberbindung herstellen. Aber kein Vorteil ist ohne Nachteil: Im Gegensatz zum privaten Drucker wickelt die Hausdruckerei die Druckaufträge nicht sofort und nicht zu jeder Tageszeit, sondern nach Maßgabe der zeitlichen Möglichkeiten ab.