

Datenaustausch – Formate und Schnittstellen

H. Möbus, Groß Düben

Die Hoffnung auf das papierlose Büro hat sich in den letzten zwei Jahrzehnten nicht erfüllt. In heutigen Büros wird mehr denn je Papier bedruckt. Aber gleichzeitig ist das Informationsvolumen, das weltweit auf maschinenlesbaren Datenträgern gespeichert und über diese ausgetauscht wird, um mehrere Größenordnungen gewachsen. Der Anwender steht daher immer öfter vor der Aufgabe, als Dateien vorliegende Informationen zu übernehmen bzw. Informationen als Dateien weiter zu geben. Über die in diesem Zusammenhang wichtigen Dateiformate informiert Sie der nachfolgende Beitrag.

1 Begriffe und technische Voraussetzungen

Die Begriffe Format und Schnittstelle werden häufig synonym genutzt, was zu Missverständnissen führt. Besser ist es, den Begriff Format im Sinne von Dateiformat zu benutzen und den Begriff Schnittstelle ausschließlich als die Fähigkeit eines Programms zu verstehen, ein bestimmtes Dateiformat lesen und/oder schreiben zu können. Damit ist der Begriff Format an das Vorhandensein einer Datei¹⁾ und der Begriff Schnittstelle an das Vorhandensein eines Programms geknüpft. Bild 1 veranschaulicht den erläuterten Sachverhalt. Um Daten austauschen zu können, bedarf es darüber hinaus technischer Voraussetzungen, unabhängig von den Formaten (Tafel 1). Wenn zwischen den Rechnern A und B Daten ausgetauscht werden sollen, müssen diese Rechner entweder über ein Netzwerk miteinander verbunden oder in der Lage sein, Datenträger (z. B. Diskette, CD-ROM etc.) zu beschreiben und wieder einzulesen. Während in der Vergangenheit der Datenaustausch vorzugsweise per Datenträger (Bild 2a) erfolgte, werden dazu jetzt mehr und mehr die durch das Internet (Bild 2b) bereitgestellten Dienste (E-Mail und Filetransfer) genutzt. Der Datenaustausch ist aber in jedem Fall mit dem Transfer einer oder mehrerer Dateien (Filetransfer) verbunden. Client-Server-Anwendungen (Bild 3) bedienen sich einer speziellen Form des Datenaustausches und unterscheiden sich vor allem durch die Tatsache, dass beide Kommunikationspartner zeitgleich aktiv sein müssen und dass der Datenaustausch nicht in je-dem Fall mit der Übertragung kompletter Dateien verbunden sein muss.

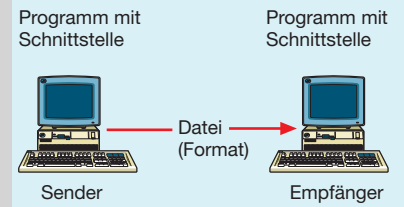
2 Standardaustauschformate

2.1 Formate zum Austausch alphanumerischer Informationen

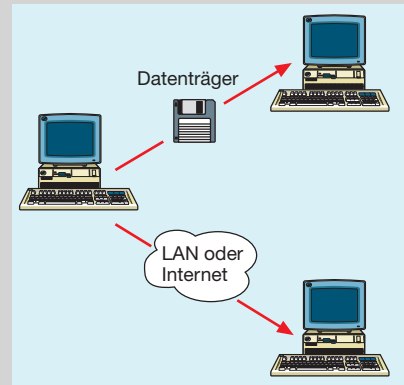
Zur Abspeicherung alphanumerischer Informationen, also in erster Linie Texte, sind von den Entwicklern der Textverarbeitungsprogramme eine Vielzahl unterschiedlicher Formate entwickelt worden. Davon haben sich nur einige wenige als Austauschformate wirklich durchgesetzt.

ASCII und ANSI. Die Unterscheidung von ASCII und ANSI ist zwar gebräuchlich, aber fachlich nicht korrekt. ASCII ist eine aus der Fernschreibtechnik übernommene Codierung für Buchstaben, Ziffern, Sonder- und Steuerzeichen. Die Codierung erfolgt in einem Byte, also mit 8 Bit. Da man anfangs aber ein Bit als Prüfbit verwendet hat, standen nur 7 Bit zur eigentlichen Codierung zur Verfügung. Die damit codierbaren 128 Zeichen werden als einfacher ASCII-Zeichensatz oder auch als Standard-ASCII-Code bezeichnet. Später wurde das achte Bit nicht mehr als Prüfbit, sondern ebenfalls zur Zeichencodierung genutzt. Die dadurch zusätzlich zur Verfügung stehenden 128 Zeichen werden zur Darstellung länderspezifischer Sonderzeichen (z. B. die deutschen Umlaute) genutzt. Da aber auch mit dem so erweiterter ASCII-Code nichtlateinische Alphabete nur mit Einschränkungen darstellbar waren, wurde durch die amerikanische Standardisierungsbehörde (ANSI - American National Standard Institut) ein universeller Code (UNICODE) entwickelt, bei dem deutlich mehr als 8 Bit zur Codierung eines Zeichens verwendet werden. Die ersten 128 Zeichen

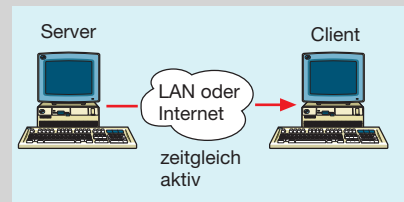
¹⁾ Unter einer Datei versteht man eine bestimmte Menge zusammengehörender Daten, die unter einem Namen (Dateinamen) eindeutig identifiziert werden können. Für die Größe des Dateinamens und die dafür nutzbaren Zeichen gibt es je nach verwendetem Betriebssystem Einschränkungen. In einem Zusatz zum Dateinamen werden Informationen zum Format und zur Verwendung der Datei abgespeichert.



1 Format und Schnittstelle



2 Datenaustausch
a) mittels Datenträger
b) über das Netz



3 Client-Server-Prinzip

Tafel 1 Zusammenstellung ausgewählter Standardformate

Nur-Text-Formate	
ASCII	American Standard Code for Information Interchange
UCS	Universal Multiple-Octet Coded Character Set (UNICODE)
Text-Dokumente	
DOC	Word Dokument
RTF	Rich Text Format
Bitmap-Formate	
BMP	Windows-Bitmap
TIF	Tagged Image Format
GIF	Graphics Interchange Format
Vektorformate	
DXF	Drawing Exchange Format
SVG	Scaleable Vector Format
Spezielle Formate	
PDF	Portable Document Format
HTML	Hypertext Markup Language
XML	Extensible Markup Language

dieses Codes entsprechen dem ASCII-Standard. Dieser Code ist auch unter dem Namen UCS bekannt, wird auch als ANSI-Code bezeichnet und z. B. unter Windows NT verwendet. Die so abgespeicherten Texte enthalten neben den Ziffern, Buch-

Autor

Dr.-Ing. Horst Möbus ist freiberuflich als EDV-Berater, Fachjournalist und Honorar-dozent tätig

Datanorm-Dateien

- DATANORM.RAB
- Rabattgruppen...
 - DATANORM.WRG
- Warengruppen
 - DATANORM.XXX
- Artikeltexte und Preise
 - DATPREIS.XXX
- Preisänderungen
 - DATASETS.XXX
- Stücklisten
- XXX Δ 001...999

4 Datanorm-Dateien

staben, Sonder- und Steuerzeichen keine weiteren Informationen und werden im allgemeinen als Textdateien (.txt) gekennzeichnet. Diese Textdateien enthalten keinerlei auf das einzelne Zeichen bezogene Formatierungsinformationen (z. B. Farbe, fett/kursiv etc.) und in diese Dateien können auch keine Bilder eingebunden werden.

DOC oder RTF. Zum Austausch von Textdokumenten werden vorzugsweise DOC- oder RTF-Dateien genutzt. Z. B. bieten Hersteller Leistungsverzeichnisse ihrer Produkte in diesen Formaten an. Die Verwendung des von MS-Word genutzten DOC-Formates bereitet wegen der Tatsache, dass die von den verschiedenen Word-Versionen erzeugten DOC-Dateien nicht immer zueinander kompatibel sind, häufig Schwierigkeiten. Das RTF-Format ist diesbezüglich wesentlich günstiger und wird von nahezu allen Windows-Textverarbeitungen unterstützt.

2.2 Formate zum Austausch grafischer Informationen

Zur Abspeicherung grafischer Informationen (Skizzen, Zeichnungen und Bilder) gibt es grundsätzlich zwei Möglichkeiten:

Bitmapgrafik-Formate. Bei diesen Formaten werden Orts- und Farbangaben für jeden Bildpunkt (Pixel) abgespeichert, daher wird auch von Pixelgrafiken gesprochen. Die rechentechnische Realisierung von Bitmapgrafiken ist relativ einfach. Daher werden diese vorzugsweise bei einfachen Grafikprogrammen genutzt. Pixelformate haben folgende Nachteile:

- Der zur Abspeicherung dieser Grafiken nötige Speicherplatz steigt mit der Anzahl der Bildpunkte.
- Pixelgrafiken können nur in relativ engen Grenzen vergrößert oder verkleinert werden.

Um Pixelgrafiken auszutauschen, werden diese in der Regel mit Komprimierungsprogrammen (Packer) verkleinert.

Vektorgrafik-Formate. Im Unterschied zu den Pixelgrafiken werden bei den Vektorgrafiken die zur Erstellung der Grafik notwendigen mathematischen Zusammenhänge abgespeichert. Eine Linie wird hier durch die Koordinaten des Anfangs- und Endpunktes und ein Kreis durch die Mittelpunktkoordinaten und den Radius beschrieben. Darüber hinaus müssen noch Informationen zur Farbe und zur Strichstärke abgespeichert werden. Der Komprimierung von in Vektorgrafik-Formaten vorliegenden Dateien sind relativ enge Grenzen gesetzt.

2.4 Spezielle Formate

Neben Standardformaten für alphanumerische und grafische Daten haben sich eine Reihe weiterer Standardformate etabliert:

HTML-Format. Das Web basiert auf dem HTML-Format. An sich ist das HTML-Format ein typisches Nur-Text-Format, aus dem heraus eine Vielzahl anderer Dateien aufgerufen und angezeigt werden können. Programme zur Anzeige von HTML-Dateien werden als Browser bezeichnet. HTML ist ein typisches „Austauschformat“ auf der Basis des Client-Server-Modells. Neben der Anwendung im Web findet das HTML-Format zunehmend bei Online-Hilfen Verwendung. Daher kann es manchmal auch nötig werden, einen Browser auf einem Rechner zu installieren, der keinen Zugang zum Internet hat.

PDF-Format. Dieses Format wird häufig für die Speicherung von Dokumenten genutzt, die aus Texten und Grafiken bestehen. Gegenüber der Speicherung als HTML- oder Textdokument hat dieses Format den Vorteil, dass der Inhalt wesentlich besser gegen Manipulationen und Kopien von Teilen geschützt ist. Zur Anzeige und zum Ausdruck wird ein spezielles Programm, der Acrobat Reader, genutzt. Das PDF-Format wird daher zunehmend auch innerhalb von Web-Präsentationen und zur Weitergabe von Unterlagen (z. B. per E-Mail) genutzt, bei denen Veränderungen weitestgehend ausgeschlossen werden sollen.

TIF-Format. Dieses Bitmap-Format wird wegen der weiten Verbreitung und der guten Komprimierbarkeit zunehmend zur Weitergabe von gescannten Unterlagen genutzt. Ein im TIF-Format vorliegendes Dokument ist gegenüber Manipulationen ähnlich gut geschützt wie ein Dokument im PDF-Format.

2.5 Komprimierung

Um Speicherplatz und/oder Übertragungszeit zu sparen, werden auszutauschende

Links zum Thema Datenformate

www.dateiformate.de
www.buergerle.de
www.gaeb.de
www.gaeb2000.de
www.datevstadt.de
www.beuth.de
www.zveh.de

Dateien häufig komprimiert. Je nach Dateiformat wird damit eine Reduzierung des Volumens um bis zu einem Faktor 10 erreicht. Derzeit wird für komprimierte Dateien vorzugsweise das ZIP-Format verwendet. Die Mehrzahl der gebräuchlichen Komprimierungsprogramme unterstützen dieses Format. Der Austausch komprimierter Dateien setzt aber bei Sender und Empfänger ein solches Programm voraus. Wenn dies nicht der Fall ist, kann die komprimierte Datei auch als selbstentpackende EXE-File gespeichert und übertragen werden.

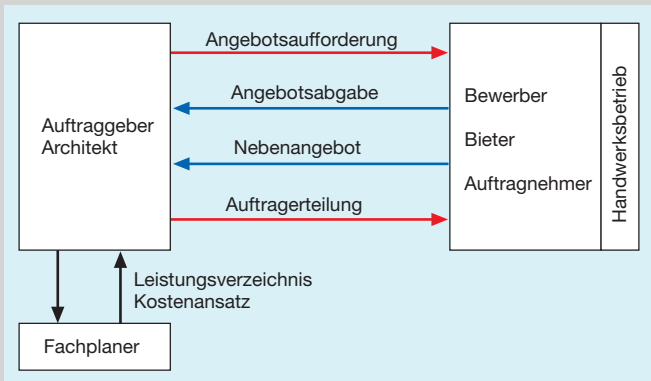
3 Branchenspezifische Austauschformate

3.1 Artikel und Leistungen

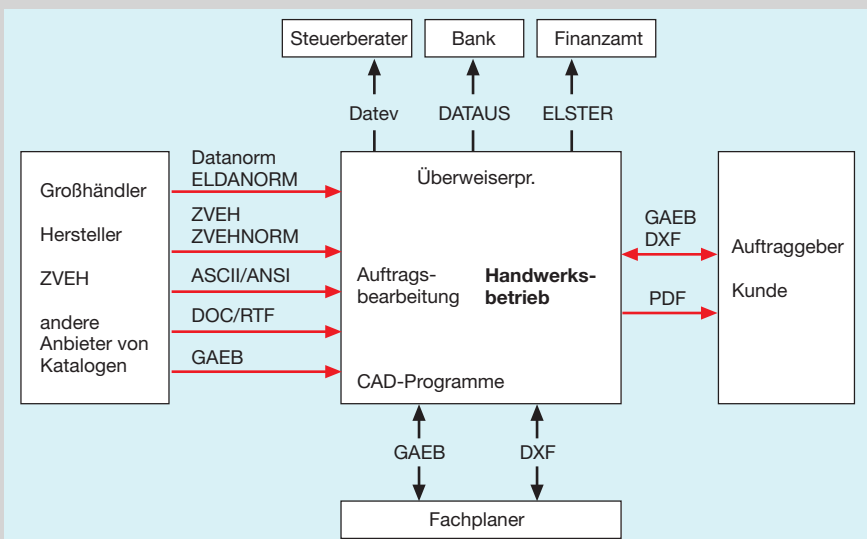
Informationen zu Artikeln und Leistungen sind für den Elektrohandwerker und den Planer unverzichtbar. Diese Informationen werden durch den Großhandel bzw. durch Anbieter von herstellernerneutralen Artikel- und Leistungskatalogen in verschiedenen Formaten bereitgestellt.

Datanorm. In diesem Format werden durch den Großhandel Artikelinformationen bereitgestellt. Die Entwicklung des Datanorm-Formates wurde durch eine Arbeitsgruppe unter Federführung des Ing.-Büros Bürgerle (www.buergerle.de) realisiert. Für das Datanorm-Format wurden im Laufe der Jahre verschiedene Versionen genormt. Obwohl bereits seit 01.01.02 die Version 5 zur Anwendung empfohlen wird, überwiegt nach wie vor der Einsatz der Version 4. Datanorm wird als Oberbegriff für einen Satz von Dateien (Bild 4) verwendet.

ZVEH. Seit mehr als einem Jahrzehnt ist die vom ZVEH entwickelte „Kalkulationshilfe für die elektro- und informationstechnischen Handwerke“ (KfE) ein wertvolles Hilfsmittel zur Erstellung von Angeboten und Rechnungen. Die als Buchwerk vorliegenden Informationen können aber auch in Branchenlösungen übernommen werden. Mit der ZVEH-Schnittstelle wurde eine für die Programmierung außerordentlich einfache Schnittstelle definiert, die entscheidend zur weiten Verbreitung der KfE beigetragen hat.



⑤ GAEB – Datenaustausch bei Planung und Durchführung



⑥ Möglichkeiten des Datenaustausches

ZVEHNORM/ELDANORM. Die ZVEH-Schnittstelle hat einen entscheidenden Mangel. Zu den einzelnen Leistungspositionen der KfE können keine Stücklisten mit den Artikeldaten des Großhändlers hinterlegt werden, von dem der Handwerker beliefert wird. Zur Lösung dieses Problems wurden die Formate ZVEHNORM und ELDANORM entwickelt. Unter der Voraussetzung, dass die KfE im ZVEHNORM-Format in das Branchenprogramm eingelesen wird und der Großhändler den Artikelstamm im ELDANORM-Format bereitstellt, kann jeder Handwerker mit seinen individuellen Preisen kalkulieren. Die dadurch im Handwerk gegebenen Möglichkeiten zur Erleichterung der Arbeit können jedoch nicht immer genutzt werden, da durch den Großhandel das ELDANORM-Format (noch) nicht durchgängig unterstützt wird.

3.2 Baudurchführung

Der Gemeinsame Ausschuss Elektronik im Bauwesen (GAEB) hat die GAEB-Formate mit dem Ziel definiert, die während des Baugeschehens zwischen den beteiligten Partner auszutauschenden Informationen zu klassifizieren. Gegenwärtig erfolgt die Ablösung der GAEB 90/91 durch die GAEB 2000.

Bei den GAEB-Formaten erfolgt eine Unterscheidung nach Datenaustauschphasen (Bild ⑤).

3.3 Buchführungsdaten

Beim Schreiben von Rechnungen entstehen Informationen, die bei der Buchführung im Steuerbüro wieder benötigt werden. Um diese nicht erneut eingeben zu müssen, können diese Daten aus der Auftragsbearbeitung im Datev-Format exportiert und durch die Finanzbuchhaltung in diesem Format wieder eingelesen werden. Namensgeber für dieses Format ist die in Nürnberg ansässige Datev eG, ein EDV-Dienstleister für Steuerberater und Wirtschaftsprüfer.

3.4 Überweisungen und Lastschriften

Das Ausfüllen von Überweisungen und Lastschriften ist eine zeitaufwändige und fehlerintensive Tätigkeit und die Erfassung dieser Belege verursacht bei den Geldinstituten wiederum erhebliche Aufwendungen. Zur Vereinfachung dieser Datenübergabe wurde zum Datenaustausch das DATAUS-Format (auch DTA-Format genannt) entwickelt. Überweisungen und Lastschriften werden mit einem Programm erfasst und auf einer Diskette gespeichert. Diese Diskette wird dann mit einer unterschriebenen Liste der auf der Diskette ge-

speicherten Überweisungen und Lastschriften an die Bank gegeben. Dieses Verfahren bringt erhebliche Arbeitserleichterungen und birgt keinerlei Sicherheitsrisiken.

3.5 Steuererklärung

Seit Kurzem haben auch die Finanzämter die Vorteile des elektronischen Datenaustausches entdeckt. Mit dem ELSTER-Verfahren (ELEktronische STEuerERklärung) können Steuererklärungen und andere Unterlagen dem Finanzamt zugestellt werden. Da aber nach wie vor diese Unterlagen auch in ausgedruckter Form (wegen der Unterschrift) übersandt werden müssen, hält sich der praktische Nutzen dieser Möglichkeit in Grenzen.

4 Aktuelle Situation und Tendenzen

Bedingt durch die anhaltend schwierige gesamtwirtschaftliche Lage und die krisenhafte Entwicklung in weiten Teilen des IT-Bereiches setzen sich neue Schnittstellen wie ELDANORM/ZVEHNORM, Datennorm V 5.0 und GAEB 2000 derzeit nur zögerlich durch. Das betrifft sowohl die Anbieter- als auch die Anwenderseite. Die Tatsache, dass mit XML eine neue Auszeichnungssprache für Daten (und damit auch ein neues Datenformat!) zur Verfügung steht, die alle Voraussetzungen mitbringt, sich als universelles Format zu etablieren, unterstützt diese Zurückhaltung. XML selbst befindet sich noch in einer Entwicklungs- und Standardisierungsphase. Eines der ersten Programme, das als Standardformat mit XML-Dateien arbeitet, ist OpenOffice 1.0/StarOffice 6.0. Wegen des vergleichsweise großen Umfangs von XML-Dateien werden die mit diesem Office-Paket erzeugten XML-Dokumente vor der Speicherung komprimiert.

5 Fazit

Für den Elektrowerkbetrieb gibt es vielfältige Möglichkeiten (Bild ⑥) Informationen auf Datenträgern oder unter Nutzung des Internets auszutauschen. Diese Möglichkeiten helfen, Arbeitszeit einzusparen und Kosten zu senken. Das Wissen um Datenformate und Schnittstellen ist vor allem bei der Auswahl geeigneter Software unverzichtbar. Die Programmierung gerade der branchenspezifischen Schnittstellen ist zum größten Teil mit einem erheblichen Aufwand verbunden. Diese Schnittstellen werden daher von den Softwarefirmen in der Regel als kostenpflichtige Zusatzmodule angeboten, deren Preis oft recht nahe an dem des Basismoduls angesiedelt ist. Hier rechtzeitig zu prüfen, welche Schnittstelle wirklich benötigt wird und auf welche ggf. verzichtet werden kann, ist unerlässlich ■