Ing. Mag. Horst Greifeneder
Allgemein beeideter und gerichtlich zertifizierter Sachverständiger, Fachgebiet Informationstechnik;
Computer-Forensik-Spezialist

Forensische Metadatenanalyse bei Office Word-Dokumenten (2007-2013)

Das Programm Microsoft Office Word ist der *De-facto*-Standard für Textverarbeitung im Privatleben und Geschäftsalltag. So ist es nicht weiter verwunderlich, dass Word-Dokumente als Beweismittel in zivil- oder strafrechtlichen Rechtssachen immer wieder eine wichtige Rolle spielen.

Beim vorliegenden Artikel handelt es sich um den zweiten Teil der Artikelserie zur forensischen Metadatenanalyse bei Office Word-Dokumenten. Im ersten Teil wurden jene des Dateityps Word 97-2003 behandelt.[1]

Im vorliegenden Artikel folgen Dokumente des Dateiformats Word 2007-2013.

1. Forensische Fragestellungen

Im Zusammenhang mit der Auswertung von sichergestellten Word-Dokumenten stehen in der Regel nachfolgende Fragestellungen im Mittelpunkt des richterlichen Interesses:

- Wer hat das Dokument ursprünglich erstellt?
- Wer hat das Dokument zuletzt bearbeitet?
- Wann wurde das Dokument erstmals erstellt?
- Wann wurde das Dokument zuletzt bearbeitet?
- Wurde das Dokument ausgedruckt?
- Wie häufig wurde das Dokument bearbeitet?

Eine wertvolle Grundlage zur Beantwortung der angeführten richterlichen Sachfragen bilden dabei die im Zusammenhang mit dem Dokument gespeicherten Metadaten.

2. DOCX-Dateiformat für Word 2007-2013

Das Standardformat für Microsoft Word ab der Version 2007 ist das auf den Office Open XML-Standard beruhende Dateiformat DOCX.[2]

Das Dateiformat DOCX ist ein Containerformat bestehend aus einem ZIP-Archiv mit XML und binärem Code. DOCX-Datei können ohne Modifikation des Dateiinhalts am einfachsten durch Dekomprimieren der Datei mit einem Komprimierungsprogramm wie 7zip analysiert werden.[3]

Office Open XML-Dokumente werden in Packages gespeichert, die den Open Packaging Conventions entsprechen. Ein Package ist eine ZIP-Datei, die alle Bestandteile (Parts und Items) eines Dokuments enthält.[4]

2.1. Parts und Items

Parts sind die einzelnen Bestandteile (Bausteine) des Inhalts des Dokuments (Text, Grafiken, Bilder etc), während Items beschreibende Metadaten sind, die festlegen, wie die einzelnen Bestandteile des Dokuments zusammengestellt und dargestellt werden sollen.[5]

Ein minimales DOCX-Dokument enthält im Wurzelverzeichnis der ZIP-Datei eine XML-Datei namens [Content_Types].xml sowie drei Verzeichnisse _rels, docProps und ein Verzeichnis word mit den eigentlichen Dokumentendaten.

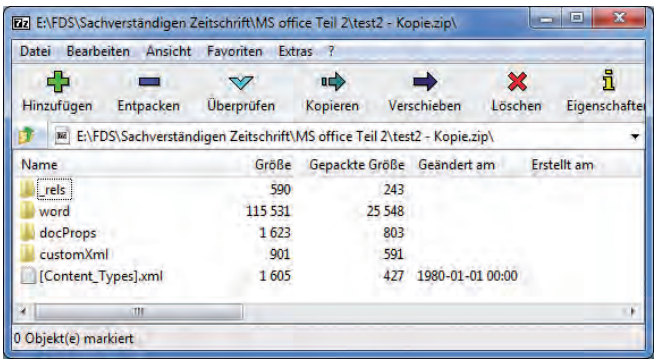

Abbildung 1: Inhalt einer dekomprimierten DOCX-Datei

Abbildung 1 zeigt den Inhalt einer dekomprimierten DOCX-Datei mit dem Namen test.docx. Die DOCX-Datei beinhaltet eine Verzeichnisstruktur mit den Verzeichnissen docProps, word und _rels sowie eine XML-Datei mit dem Namen [Content_Types].xml.

2.2. _rels

Die Datei _rels im Verzeichnis _rels beinhaltet Informationen über die Struktur des Word-Dokuments und zu den Speicherorten der Metadaten des Dokuments sowie des XML-Dokuments mit dem eigentlichen Inhalt der Word-Datei.

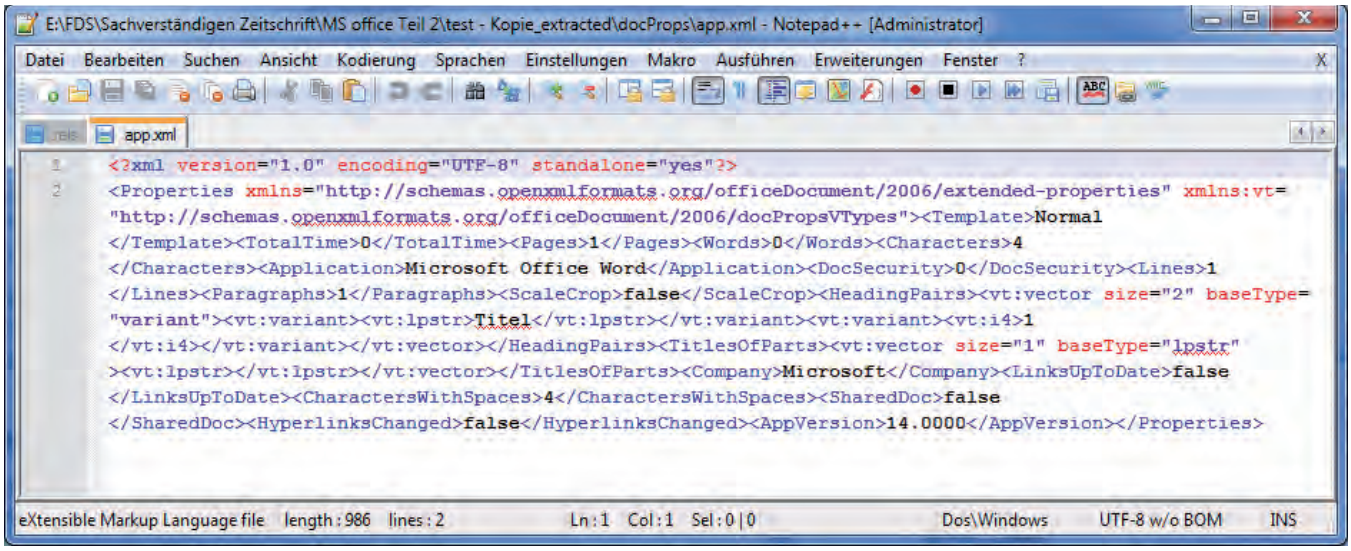

Abbildung 2: Die Datei app.xml enthält applikationsspezifische Metadaten des Word-Dokuments

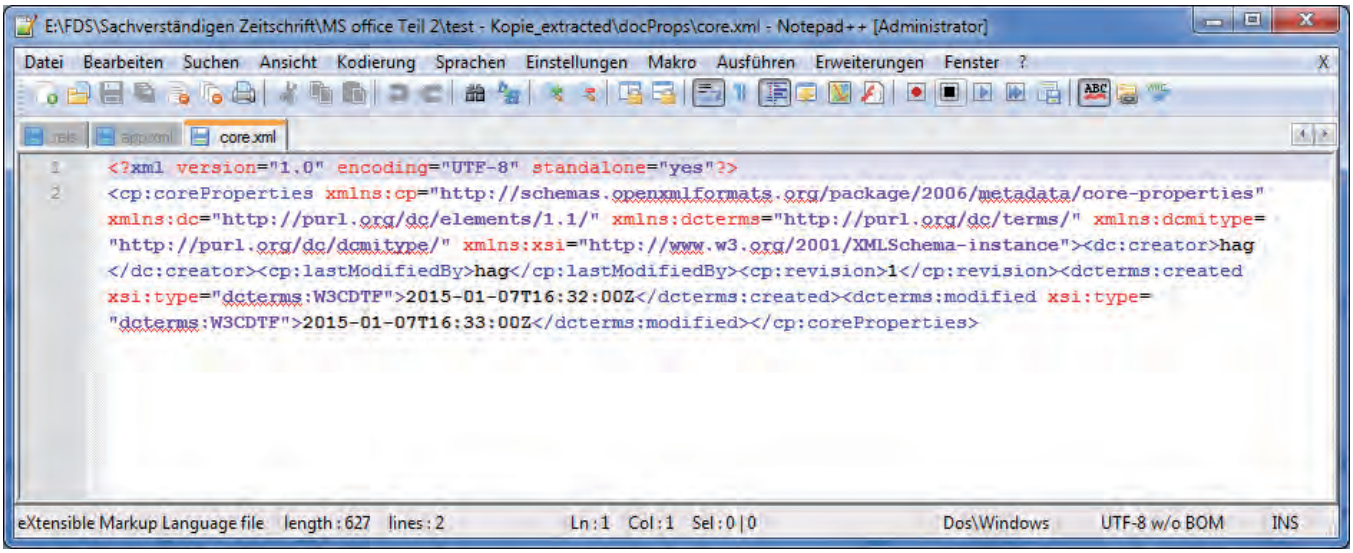

Abbildung 3: Die Datei core.xml enthält dokumentenspezifische Metadaten des Word-Dokuments

2.3. docProps

Im Verzeichnis docProps werden in zwei oder mehreren XML-Dateien die Metadaten des Word-Dokuments nach dem Dublin-Core-Standard (ISO 15836:2003) der „Dublin Core Metadata Initiative“ (DCMI) gespeichert.[6]

2.3.1. app.xml

Die Datei app.xml enthält programmspezifische Metadaten des Dokuments, unter anderem den Programmnamen (zB Microsoft Office Word), die Programmversion (zB 14.0000 für Word 2010), statistische Informationen wie die Anzahl der Seiten, Wörter und Buchstaben im Dokument sowie Daten zu Sicherheitseinstellungen des Dokuments (DocSecurity).

2.3.2. Die Datei core.xml

Die Datei core.xml enthält dokumentenspezifische Metadaten des Dokuments, unter anderem den Autor, den letzten Bearbeiter, die Anzahl von Überarbeitungen sowie Datums- und Zeitangaben im W3CDT-Format (zB 2015-01-07T16:32:00Z) zum Zeitpunkt der Erstellung, letzten Modifikation und des letzten Ausdrucks des Dokuments.

Bei einem reproduzierbaren Versuch konnte der Autor feststellen, dass bei einer entpackten DOCX-Datei die metadatenspezifische Inhalte der Dateien app.xml und core.xml im Klartext problemlos manipuliert werden können.

Die manipulierten Daten werden in die Eigenschaften des Dokuments übernommen. Der Zeitpunkt der Modifikation ist in den dokumentenspezifischen Zeitstempeln nicht er-

sichtlich. Allerdings ändern sich die Zeitangaben für den Zeitpunkt des letzten (lesenden) Zugriffs auf die DOCX-Datei bzw der letzten Modifikation bei den dateispezifischen Zeitstempeln des Dateisystems.

2.4. Das Verzeichnis word

Im Verzeichnis word befindet sich eine Reihe von Verzeichnissen und XML-Dateien mit den Daten zum Dokumenteninhalt.

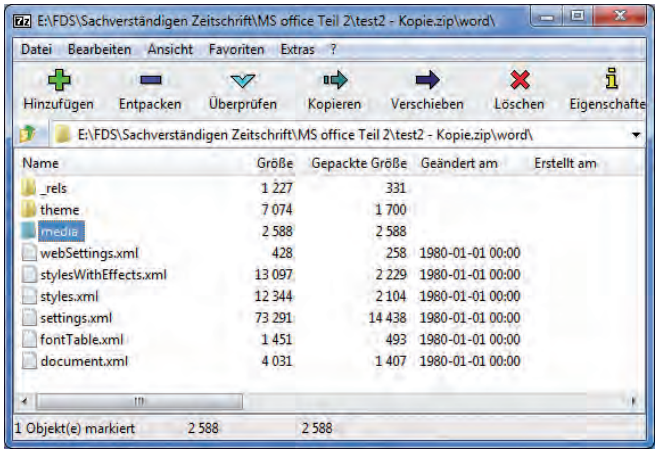

Abbildung 4: Inhalte des Verzeichnisses word

Das Verzeichnis theme enthält Angaben zum gewählten Design des Dokuments. In das Dokument integrierte Bilder werden im Standardverzeichnis /word /media gespeichert. Die Originalbezeichnung der Bilder wird durch die Bezeichnung imageN (zB image1.jpg) ersetzt.

Die Datei document.xml beinhaltet schlussendlich den Hauptteil des eigentlichen Dokumenteninhalts sowie Referenzen auf im Dokument enthaltene Medien (zB Bilder mit den Originaldateinamen).

2.5. Das Verzeichnis customXml

Das Verzeichnis customXml enthält in der Regel keine für die forensische Metadatenanalyse relevanten Informationen.

2.6. [Content_´Types].xml

Diese Datei enthält eine Beschreibung des Inhaltes der ZIP-Datei und ist aus forensischer Sicht wenig ergiebig.

3. Ausgewählte Metadaten

Metadaten sind Daten über Daten, das heißt, Metadaten beschreiben bestimmte Eigenschaften eines Dokuments. Bei einem DOCX-Dokument beinhalten Metadaten unter anderem Informationen über den Autor, den Zeitpunkt der Erstellung und letzten Bearbeitung des Dokuments.

Immer wenn ein Benutzer ein Dokument speichert, werden – neben dem eigentlichen Dateiinhalt – die Metadaten des Word-Dokuments nach dem Dublin-Core-Standard in den Dateien app.xml und core.xml im Klartext gespeichert.

3.1. creator

Das Metadatum „creator“ wird in der Datei core.xml gespeichert und beschreibt den Autor des Dokuments mit dessen Benutzernamen. Der Wert wird vom Programm beim erneuten Speichern der Datei standardmäßig nicht geändert und stellt damit einen Hinweis auf den ursprünglichen Ersteller des Dokuments dar. Aus forensischer Sicht zu beachten ist, dass beim Speichern des Dokuments mit dem Befehl „Speichern unter“ der Name des Autors unverändert bleibt, obwohl es sich beim neuen Dokument aus inhaltlicher Sicht um ein ganz neu erstelltes Dokument handelt.

3.2. Zuletzt geändert von

Das Metadatum „LastModifiedBy“ wird in der Datei core.xml gespeichert und beschreibt den Namen des letzten Benutzers, der Änderungen am Dateiinhalt vorgenommen hat.

In diesem Zusammenhang ist aus forensischer Sicht zu beachten, dass beim Speichern der Datei der Wert jedes Mal aktualisiert und der bisherige Eintrag überschrieben wird.

3.3. Revision

Das Metadatum „Revision“ wird in der Datei core.xml gespeichert und beschreibt die Anzahl von Überarbeitungen des Dokumenteninhalts.

Bei einem Versuch des Autors konnte festgestellt werden, dass – abweichend vom Verhalten des Feldwertes REVNUM bei DOC-Dateien – der Wert des Feldes nur bei gespeicherten Änderungen am Dokumenteninhalt erhöht wurde.

3.4. Datums- und Uhrzeitinformationen

Unabhängig von den Zeitstempeln des Dateisystems speichert Microsoft Word nachfolgende Zeitangaben im W3CDTF-Datums-/Zeitformat (UTC) in der core.xml-Datei:[7]

- Created – Erstellzeitpunkt des Dokuments,
- LastModified – Zeitpunkt der letzten Speicherung des Dokuments und
- LastPrinted – Zeitpunkt des letzten Ausdrucks des Dokuments.

Die Zeitstempel können über Microsoft-eigene Feldfunktionen ausgelesen werden und stehen auch bei wiederhergestellten Dateien zur Verfügung, bei denen Zeitstempelinformationen aus dem Dateisystem meist nicht mehr nutzbar sind.

3.5. Programmversion

Welche Word-Version zur Erstellung des Dokuments verwendet wurde, lässt sich aus den in app.xml gespeicherten Metadaten „Application“ und „AppVersion“ entnehmen. Ein Wert von 14.0000 für AppVersion bezieht sich zB auf Microsoft Office Word 2010.[8]

4. Auswertung von Metadaten

Neben den bereits im ersten Teil zur forensischen Metadatenanalyse bei Office Word-Dokumenten beschriebenen Verfahren zur Auswertung (wie die programmeigene Anzeige von Eigenschaften des Word-Dokuments oder Extraktion der Metadaten per Spezialprogramm)[9] ermöglicht das DOCX-Dateiformat eine direkte Klartext-Analyse der Metadaten in den entpackten Dateien core.xml und app.xml.

5. Fazit

Metadaten in Word-Dokumenten mit dem DOCX-Dateiformat stellen eine wichtige Informationsquelle zur Beantwortung von forensischen Fragestellungen im Zusammenhang mit der Erstellung und Bearbeitung von Word-Dokumenten dar.

Im Vergleich zu den Metadaten in älteren Versionen von Word (DOC-Dateien) ist jedoch zu berücksichtigen, dass die Metadaten einer DOCX-Datei durch den Benutzer in den Dateien core.xml und app.xml direkt im Klartext modifiziert werden können. Der Zeitpunkt derartiger Modifikationen spiegelt sich nur in den entsprechenden Zeitstempeln des Dateisystems wider.

Anmerkungen:

1 *Greifeneder*, Forensische Metadatenanalyse bei Office Word-Dokumenten, SV 2014/2, 93.

2 Seite „Office Open XML“, URL: http://de.wikipedia.org/wiki/Office_Open_XML.

3 Seite „Word Document (DOCX)“, URL: http://www.forensicswiki.org/wiki/Word_Document_(DOCX).

4 Das ZIP-Dateiformat ist ein Format für komprimierte Dateien, das einerseits den Platzbedarf bei der Archivierung reduziert und andererseits als Containerdatei fungiert, in der mehrere zusammengehörige Dateien oder auch ganze Verzeichnisbäume zusammengefasst werden können (Quelle:Wikipedia).

5 Seite „Office Open XML“, URL: http://de.wikipedia.org/wiki/Office_Open_XML.

6 Seite „Office Open XML“, URL: http://de.wikipedia.org/wiki/Office_Open_XML.

7 Aus einer Zeitangabe in UTC ergibt sich die entsprechende in Österreich geltende mitteleuropäische Zeit (MEZ), indem man eine Stunde, und die im Sommer geltende mitteleuropäische Sommerzeit (MESZ), indem man zwei Stunden addiert.

8 Seite „Microsoft Word“, URL: http://de.wikipedia.org/wiki/Microsoft_Word.

9 *Greifeneder*, SV 2014/2, 96 f.

Korrespondenz:
Ing. Mag. Horst Greifeneder
Schenkelbachweg 32, A-4600 Wels
Tel.: 07242 / 77715
Fax: 07242 / 77716
E-Mail: office@fds.at

Wichtig für alle im Jahr 2010 zertifizierten und im Jahr 2005 sowie im Jahr 2010 rezertifizierten Sachverständigen Rezertifizierung 2015

Wir machen darauf aufmerksam, dass alle Sachverständigen, die während des Jahres 2005 auf weitere 10 Jahre bzw während des Jahres 2010 auf weitere 5 Jahre eingetragen wurden sowie all jene, die im Jahr 2010 erstmalig allgemein beeidet und gerichtlich zertifiziert wurden, bis längstens Ende September 2015 den Antrag auf Verlängerung der Eintragung an die Präsidentin oder den Präsidenten des Landesgerichts, bei dem sie seinerzeit den Antrag auf Eintragung gestellt haben, zu richten haben.

Im Antrag sind die gerichtlichen Verfahren, in denen Sie seit Ihrer Eintragung, bei häufiger Heranziehung in einem maßgeblichen Zeitraum unmittelbar vor der Antragstellung, also etwa im letzten Jahr vor der Antragstellung, tätig geworden sind, mit Aktenzeichen und Gericht anzuführen. Der Rezertifizierungsantrag hat auch einen Hinweis auf die absolvierten Fortbildungsaktivitäten zu enthalten. Legen Sie daher auch – soweit vorhanden – dem Antrag einen Ausdruck des Bildungs-Passes bei.

Die Präsidentin oder der Präsident kann weitere Ermittlungen anstellen und ein Gutachten der Kommission nach § 4a SDG oder eine Äußerung eines qualifizierten Mitglieds dieser Kommission einholen.

Es wird empfohlen, den Antrag auf Rezertifizierung nicht erst gegen Ende der dafür offenstehenden Frist, sondern möglichst bald zu stellen, um eine gleichmäßige Auslastung der mit der Rezertifizierung befassten Stellen zu erreichen.