

# JPEG2000 - Neue Dateiformate in der Archivierung

## JPEG2000 - New file format in archiving

Thomas Zellmann  
LuraTech GmbH  
Kantstr. 21, D-10623 Berlin  
Tel : +49-30-39 40 50 – 0 Fax : +49-30-39 40 50 – 99  
Email : t.zellmann@luratech.com Internet : www.luratech.com

### Zusammenfassung :

Die LuraTech GmbH zeigt die neuen JPEG2000-Dateiformate mit den Teilen :  
JPEG2000/Part1 für die effiziente Kompression von BILDERN und  
JPEG2000/Part6 für die effiziente Kompression von DOKUMENTEN  
Diese Formate eignen sich auch für historische Objekte bei Archiven, Bibliotheken und Museen.

### Abstract :

LuraTech GmbH presents the new JPEG2000 file formats with its parts :  
JPEG2000/Part1 for efficient IMAGE compression and  
JPEG2000/Part6 for efficient compression of digitised DOCUMENTS.  
Those formats are also quite interesting in cultural heritage projects of archives, libraries and museums.

### JPEG2000 -

#### Neue Dateiformate in der Archivierung

Immer mehr Organisationen nutzen elektronische Archive, um ihre Dokumente und Bilder langfristig aufzubewahren. Die Vorteile – im Vergleich zum papierbasierten Archiv – liegen auf der Hand: Schnellerer Zugriff und damit einhergehend die schnellere Beantwortung von Rückfragen, geringerer Platzbedarf sowie eine sichere Aufbewahrung von wichtigen Dokumenten sind nur einige Gründe. Doch mit der Anzahl der Dokumente und Bilder steigt auch der benötigte Speicherplatz. Zwar sind die Kosten für Speichermedien in den vergangenen Jahren gesunken, doch nicht der Administrationsaufwand. Darüber hinaus sind die Ladezeiten für größere Dokumente länger. Der folgende Beitrag stellt die derzeit aktuellen Dateiformate vor, die für die Speicherung von Dokumenten und Bildern geeignet sind. Dabei wird auch auf die Vor- und Nachteile dieser Formate eingegangen.

### JPEG2000 -

#### New file format in archiving

More and more organisations are using electronic archives to store their documents and images over the long term. The advantages – compared to paper-based archives – are obvious: faster, company-wide access and thus faster responses to inquiries, low space requirements and secure storage of critical documents, to name only a few. However, the increasing volume of documents and images requires increased electronic storage capacity as well. The costs for storage media have fallen over the past few years, but the administrative expenses have not. The loading times for larger documents are also longer. The following article presents the latest file formats suited for storing documents and images. The advantages and disadvantages of the particular formats are also discussed.

### **Zurzeit genutzte Bildformate**

Schaut man sich in Organisationen um, so stellt man fest, dass insbesondere die Bildformate JPEG, TIFF, BMP oder GIF genutzt werden – mit ihren bekannten Stärken und Schwächen: Das JPEG-Format hat zwar den Vorteil, dass das Dateivolumen relativ gering ist, dafür die Bildqualität um so schlechter. Beim TIFF- und auch beim BMP-Format ist es genau anders herum: Das Datenvolumen ist groß, dafür die Bildqualität sehr gut. Auch im GIF-Format ist das Datenvolumen relativ groß und die Bildqualität aber nur mittelmäßig.

### **Neuer Bildstandard JPEG2000 – der Nachfolger von JPEG**

Als neuer Bildstandard etabliert sich zunehmend JPEG2000 – der offizielle Nachfolger von JPEG. Auch dieses Format ist von der ISO-Gruppe standardisiert und in zahlreichen Anwendungen, wie beispielsweise Quicktime von Apple, Adobe Photoshop oder CorelDRAW 11, bereits fester Bestandteil. In diesem Format abgespeicherte Bilder überzeugen mit einer hohen Qualität und sind dennoch bedeutend kleiner – im Vergleich zu TIFF LZW oder JPEG um den Faktor zwei bis drei. Damit nehmen sie weniger Speicherplatz in Anspruch, der Datentransfer ist schneller und die Download-Zeiten reduzieren sich. Darüber hinaus können JPEG2000 Dateien umfangreiche Metadaten im standardkonformen XML-Format aufnehmen. Somit lassen sich ausgewählte Metadaten einer Bilddatenbank an den Anwender übermitteln. Zusätzlich können die Bilder schnell in kleinere Auflösungen konvertiert werden. Das Vorhalten von unterschiedlichen Versionen eines Bildes, z.B. in geringerer Auflösung für das Internet und in einer höheren Auflösung für den Ausdruck, erübrigt sich.

Im Bereich der Archivierung kann JPEG2000 LOSSLESS speichern und somit als digitales Original (englisch : digital master) für die Langzeitarchivierung eingesetzt werden.

### **Image formats in current use**

If you look around in various companies, you'll notice that the image formats JPEG, TIFF, BMP or GIF are used particularly frequently – with all their familiar strengths and weaknesses. The JPEG format has the advantage that the file size is relatively small, but the image quality is also poorer. With the TIFF and BMP formats, the reverse is true: the file size is large, but the image quality is very good. With the GIF format, the file size is also relatively large, but the image quality is mediocre.

### **New image standard JPEG2000 – the successor to JPEG**

JPEG2000 is rapidly establishing itself as the new image standard – the official successor to JPEG. This format has also been standardized by the ISO group and is already an integral feature in numerous applications, such as QuickTime by Apple, Adobe Photoshop or CorelDRAW 11. Images saved in this format are of a high quality, yet significantly smaller by a factor of two or three compared to TIFF, LZW or JPEG files. This means that they require less storage space, data transfer is faster and download times are reduced. Furthermore, JPEG2000 files can record comprehensive metadata in a format compliant with XML standards. Consequently, selected metadata can be conveyed to the user from an image database. Moreover, the images can quickly be converted to lower resolutions. Saving different versions of images, i.e. in lower resolution for the Internet and in a higher resolution for printouts, is no longer necessary.

For archiving, JPEG2000 can save images in a lossless manner and can thus be used as a digital master for long-term archiving.

## **Archivierung von zusammenhängenden Dokumenten**

Das neue JPEG2000-Format ist in sechs Teile untergliedert, weitere sind in Bearbeitung. Teil eins als Basis beinhaltet die Syntax des JPEG2000 Datenstroms und alle notwendigen Schritte zum Kodieren und Dekodieren von JPEG 2000 Bildern. Für die Verarbeitung von sogenannten „compound documents“, also farbigen Dokumenten, die sowohl bitonale Elemente (Text, technische Zeichnungen) als auch Bilder enthalten, wurde JPEG2000/Part 6 entwickelt und von der ISO-Gruppe verabschiedet. Es definiert das JPM Dateiformat für die Dokumentenbildverarbeitung und nutzt dafür das Mixed Raster Content (MRC) Modell.

## **Die Herausforderung der Dokumentenkompression meistern**

Das Mixed Raster Content-Modell trägt der Situation Rechnung, dass es kein einheitliches optimales Kompressionsverfahren für sämtliche Komponenten eines Dokumentes gibt. Wenn man die Dokumentengröße über ein Bildkompressionsverfahren reduziert, wird der Text schnell schlecht lesbar. Bei Nutzung des Fax-G4-Formates verschwindet die Farbe. Die Lösung besteht darin, Text- und Bildanteile in einzelne Ebenen zu segmentieren und anschließend separat mit den jeweils bestmöglichen Algorithmen zu komprimieren. Bitonale Elemente wie Text und technische Zeichnungen werden verlustfrei im Fax Group 4- oder JBIG2-Format gespeichert, während die Kompression des Vorder- und Hintergrunds im JPEG2000-Standard erfolgt. So lässt sich eine exzellente Schrift- bzw. Bildqualität bei extrem kleinen Dateigrößen erreichen. Farbige Dokumente, die im TIFF-Format eine Größe von 23 MB haben, schmelzen zu 75 KB zusammen. Gleichzeitig können enthaltene Texte durch OCR-Programme besser erkannt und ausgelesen werden, da sie vom Hintergrund freigestellt sind. Dies trifft insbesondere zu, wenn die Schrift weiß und der Hintergrund dunkel ist.

## **Archiving of documents that belong together**

The new JPEG2000 format is divided into six parts, and others are in the works. Part one, as the base, includes the syntax of the JPEG2000 data stream and all necessary steps for encoding and decoding JPEG 2000 images. JPEG2000/Part 6 was developed and approved by the ISO group for processing so-called "compound documents", i.e. color documents that contain bitonal elements (such as text or technical drawings) as well as images. It defines the JPM file format for the document image processing and uses the mixed raster content (MRC) model for this purpose.

## **Mastering the challenge of document compression**

The mixed raster content model takes into account that there is no single optimal compression process for all components of a document. When the document size is reduced using an image compression process, the text rapidly becomes illegible. When the fax G4 format is used, the color disappears. The solution is to create separate layers for text and image components and then to compress them individually using the most appropriate algorithms. Bitonal elements such as text and technical drawings are saved without loss in the Fax Group 4 or JBIG2 format, while the foreground and background are compressed using the JPEG2000 standard. This results in excellent text and image quality and files that are extremely small in size. Color documents which are 23 MB in TIFF format shrink to a mere 75 KB. At the same time, text in the document can be recognized and read better by OCR programs, since it is free of the background. This is especially true when the text is white and the background dark.

## **LuraTech GmbH**

Einer der Pioniere bei der Entwicklung von Bild- und Dokumentenkompressionsverfahren ist die LuraTech GmbH aus Berlin (Deutschland). Zehnjährige Forschungserfahrung im Bereich Wavelet-basierter Datenkompression mit angewandter Engineering-Kompetenz sowie eine enge Zusammenarbeit mit Universitäten und Forschungseinrichtungen sind für LuraTech die Basis zur Entwicklung innovativer Kompressionssoftware. Der Entwicklungsleiter der LuraTech leitet die deutsche Delegation des JPEG Komitee bei der International Organization for Standardization (ISO). Zusätzlich ist das Unternehmen als Co-Autor von JPEG2000/Part6 tätig und hat den größten Teil der JPEG2000/Part4 Testströme erstellt.

LuraTech hat mit LuraWave und LuraDocument zwei Hochleistungs-Kompressionsverfahren für digitale Bilder und Dokumente entwickelt, die insbesondere in der Medizin, geographischen Informations- und digitalen Archivsystemen sowie allgemein in Imaging-Lösungen eingesetzt werden. Das neueste Produkt ist LuraDocument JPM, die weltweit erste Implementierung des neuen ISO-Standards JPEG2000/Part6 (ISO 15444-6), dem internationalen Standard für das sogenannte Multi-Layer Verfahren. Als einziges auf offenen Standards zur Multi-Layer Technologie basierendes Produkt ist LuraDocument JPM führend im Bereich der Kompression farbiger Dokumente. Mit der Multi-Layer Technologie werden farbige Dokumente in Schichten aufgeteilt und auf ca. ein Prozent ihrer ursprünglichen Größe reduziert, ohne sichtbaren Qualitätsverlust.

### **Fazit**

Durch Nutzung von ISO-standardkonformen Formaten können so selbst farbige Dokumente bei geringem Speicherbedarf revisionssicher archiviert werden.

## **LuraTech GmbH**

LuraTech GmbH, based in Berlin, Germany, is one of the pioneers in the development of image and document compression. A decade of research experience in the area of wavelet-based data compression together with applied engineering competence as well as close cooperation with universities and research institutes form the basis for LuraTech's development of innovative compression software. The development manager at LuraTech heads the German delegation of the JPEG committee at the International Organization for Standardization (ISO). Furthermore, the company is a co-author of JPEG2000/Part 6 and created most of the JPEG2000/Part 4 test streams.

With LuraWave and LuraDocument, LuraTech has developed two high-performance compression processes for digital images and documents which are used particularly in medicine, geographical information and digital archiving systems as well as in imaging solutions in general. The latest product is LuraDocument JPM, the first implementation worldwide of the new ISO standard JPEG2000/Part 6 (ISO 15444-6), the international standard for the so-called multi-layer process. As the only product based on open standards of multi-layer technology, LuraDocument JPM is a leader in the area of color document compression. With multi-layer technology, color documents are separated into layers and reduced to approximately one percent of their original size without any visible loss of quality.

### **Conclusion**

The use of formats that comply with ISO standards facilitates the archiving of color documents with minimal storage requirements in a form suitable for auditing.