

Uwe M. Borghoff  
Peter Rödiger  
Jan Scheffczyk  
Lothar Schmitz

# Long-Term Preservation of Digital Documents

Principles and Practices

## **Erratum**

Due to an error that occurred during the production process, the Table 9.2 on p. 186 is partially cut off. Please find attached the correct version of p. 186.

This project aimed to obtain the following goals:

- Identification of high-risk file attributes
- Assessment of the requirement to convert files from the TIFF 5.0 format into the current version 6.0
- Investigation of the status of the upcoming TIFF 7.0 revision
- Assessment of the risks that might occur if a TIFF version would be skipped, i.e., TIFF 5.0 files were directly converted to TIFF 7.0 and
- Assessment of the risks and possible data losses if RDO were converted into an open format, here *Cornell Digital Library* (CDL)

*risk categories and examples*

The project team carried out extensive investigations and analyzes on the data formats TIFF and RDO, but also on other data formats and projects that had provided reasons for a conversion from TIFF 4.0 to TIFF 5.0. The result of this analysis was, among others, a collection of examples of risk categories, a selection of which we exemplary show in Table 9.2.

*assessment of TIFF*

Although it was easy to obtain information about both old and current versions of TIFF, a direct contact with the developer team of TIFF 7.0 at Adobe could not be established. The advantages of TIFF are platform independence, support for a wide range of applications, and storage of metadata. The flexibility of TIFF causes, however, some disadvantages, e.g., nonstandard tags and the option to choose the bit order either to be *MSB* (*most significant bit*, i.e., the left most bit is most significant) or *LSB* (*least significant bit*, i.e., reversed order). After a thorough evaluation of TIFF 5.0 and TIFF 6.0, and various tests concerning the migration quality, the project team saw no risk to skip this migration. Instead, TIFF was placed on a watch list.

**Table 9.2.** Examples of risk categories

Risk Category	Examples
Content Fixity	The file format uses a new compression method that changes the order of bits
Security	Format migration may impact digital watermarks, digital stamps, or other cryptographic techniques for the “stability” of the content
Context and Integrity	Links to other files, e.g., metadata or variants are modified by migration
References	File extensions change due to updates in the file format. This also affects the URL of the file
Costs	The costs for long-term preservation are generally unpredictable because every migration cycle require different steps depending on the kind of the migration, e.g., routine work or introduction of a new paradigm
Staff	The staff lacks required technical knowledge
Functionality	Certain properties of a new file format, can influence the creation of variants, e.g., for print
Legal	Copyright issues can restrict the use of derived versions, e.g., versions having a better resolution