

# The Storage of Macros and OLE Objects

---

*DiaLOGIKa/makz 22 September 2008*

## **Contents**

Introduction.....	2
The Compound File Binary Format.....	2
OLE Objects in the Compound File Binary Format .....	2
OLE Objects in the OpenXML File Format .....	3
Macros in the Compound File Binary Format .....	4
Macros in the OpenXML File Format.....	4
Extending the Structured Storage Reader.....	5

## Introduction

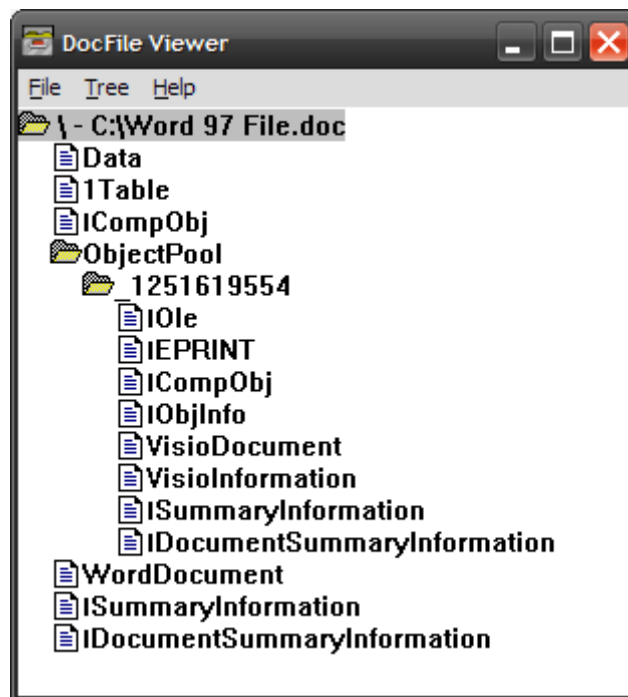
This document gives an overview on the implementation of macros and OLE objects in Microsoft's binary Office file format. It explains how they are stored in a binary Word file and why it doesn't take a specification to convert them to the new OpenXml Format.

The following descriptions and algorithms are based on *[MS-CFB]: Compound File Binary File Format* and *MS-DOC: Word Binary File Format (.doc) Structure Specification*.

## The Compound File Binary Format

The *Compound File Binary Format* was created by Microsoft to store structured binary data in a single file. The format is very similar to the well known file system FAT. It consists of storages (folders) and streams (files). These storages and streams are accessed via an index table in the header of such a file.

This file format is used as the base of many high level Microsoft file formats and has also been adopted by other companies. Moreover, it is the base of the binary Office file formats including macros and OLE objects.



The storages and streams of a Word 97 file.

## OLE Objects in the Compound File Binary Format

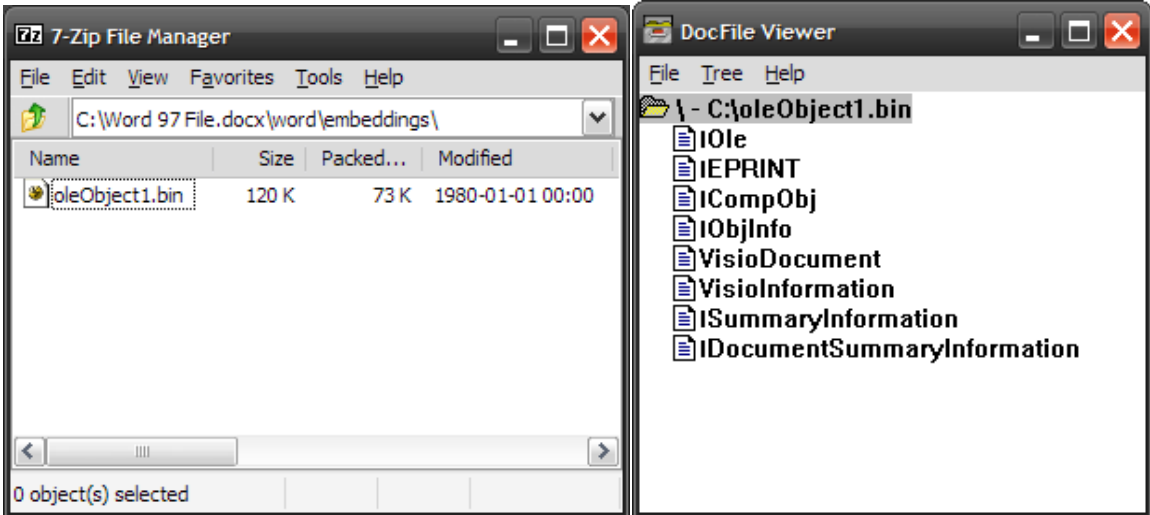
Each of the three binary Office file formats has a special storage where OLE objects are stored in. This storage is called "ObjectPool". Each OLE object of a .doc file is placed in the "ObjectPool" in a sub-storage with a randomly generated name (see screenshot above).

Each OLE object consists of several streams. It depends on the format of the object which streams are stored in the storage. The picture above shows an embedded Visio Document.

### OLE Objects in the OpenXML File Format

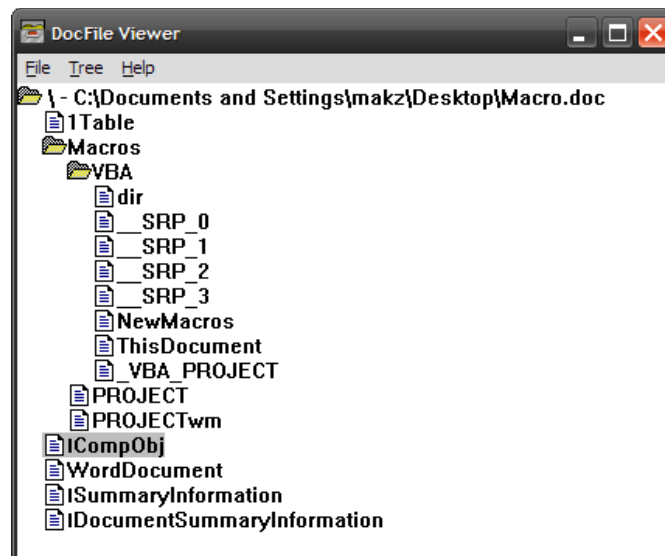
Embedded objects are stored in an OpenXML file inside the zip archive in a folder called “embeddings”. The files usually have the extension .bin and are Compound Files.

The following example demonstrates the same Word 97 file with the embedded Visio Document in OpenXml. The second picture shows the internal CFB structure of the OLE object.



## Macros in the Compound File Binary Format

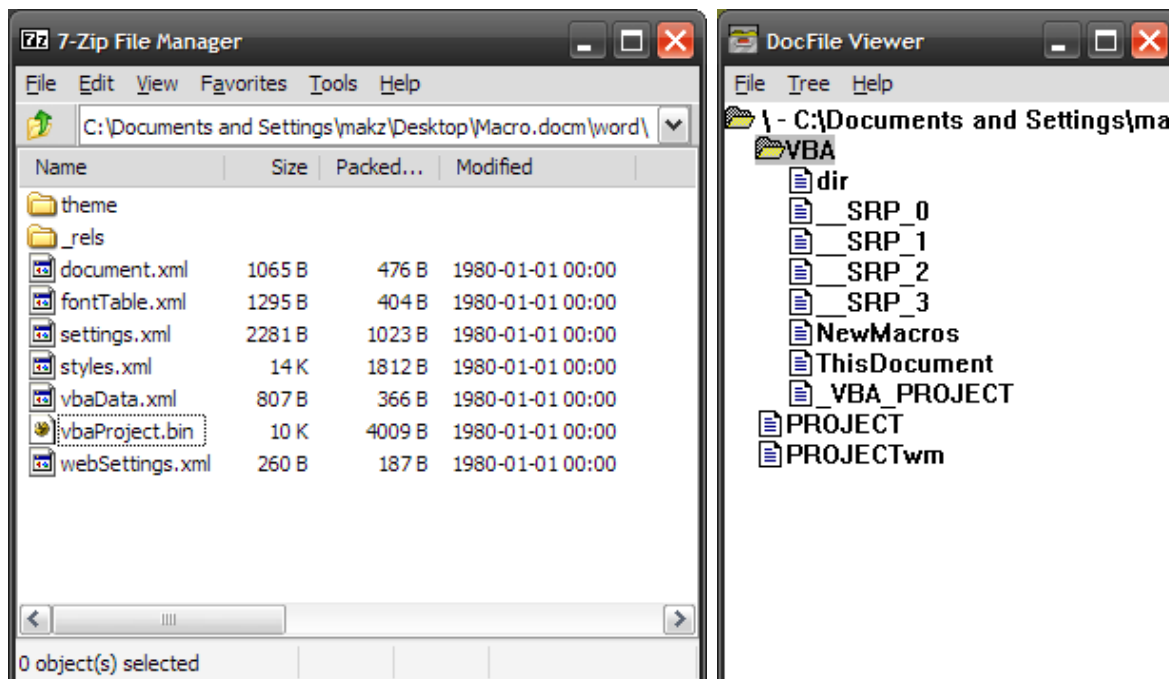
The binary Word files has one storage for all Macros in the document. This storage is called “Macros” and contains the binary macro data. Additionally meta information about the macros in the document is stored in the table stream of the binary Word file:



A Binary Word File with several macros.

## Macros in the OpenXML File Format

There is no special folder for macro data in OpenXML. The meta information is stored in a file called “vbaData.xml”, the macro data is stored as a Compound File in “vbaProject.bin”:



## **Extending the Structured Storage Reader**

At the current state of the project (Phase I – Milestone 2) the “StructuredStorageReader” library is only able to read Compound File Binaries. In order to implement the conversion of OLE objects and macros, the library must be extended to provide a functionality to write streams to a Compound File Binary.