Chapter 9. Object and Cross-Reference Streams

This chapter provides information about the implementation of object stream and cross-reference stream support in qpdf.

9.1. Object Streams

Object streams can contain any regular object except the following:

- stream objects
- objects with generation > 0
- the encryption dictionary
- objects containing the /Length of another stream

In addition, Adobe reader (at least as of version 8.0.0) appears to not be able to handle having the document catalog appear in an object stream if the file is encrypted, though this is not specifically disallowed by the specification.

There are additional restrictions for linearized files. See Section 9.3, “Implications for Linearized Files”, page 41for details.

The PDF specification refers to objects in object streams as “compressed objects” regardless of whether the object stream is compressed.

The generation number of every object in an object stream must be zero. It is possible to delete and replace an object in an object stream with a regular object.

The object stream dictionary has the following keys:

- `/N`: number of objects
- `/First`: byte offset of first object
- `/Extends`: indirect reference to stream that this extends

Stream collections are formed with `/Extends`. They must form a directed acyclic graph. These can be used for semantic information and are not meaningful to the PDF document's syntactic structure. Although qpdf preserves stream collections, it never generates them and doesn't make use of this information in any way.

The specification recommends limiting the number of objects in object stream for efficiency in reading and decoding. Acrobat 6 uses no more than 100 objects per object stream for linearized files and no more 200 objects per stream for non-linearized files. ***QPDFWriter***, in object stream generation mode, never puts more than 100 objects in an object stream.

Object stream contents consists of *N* pairs of integers, each of which is the object number and the byte offset of the object relative to the first object in the stream, followed by the objects themselves, concatenated.

9.2. Cross-Reference Streams

For non-hybrid files, the value following `startxref` is the byte offset to the xref stream rather than the word `xref`.