# Apple Pages Format Profile

*For the identified Class A and B formats, develop brief profiles of each format that minimally includes for the format:*

## Full name (taken from the specification if applicable) and common aliases

Apple Pages, Apple iWork Pages

## Brief description

*Include the key things we should know about it (e.g. this is a compound format which always includes 3 files)*

Apple Pages is part of the iWork office suite and is a competitor to Microsoft Office Word. It is only officially supported on iOS and OS X. Since the first version was released in 2005 (iWork ’05), the increasing sophistication of the software has been reflected by the increasing complexity of the format. In particular, on the release of the iWork ’09 suite, the underlying file formats were radically changed from OS X Bundle directories to a ZIP-based format (similar to DOCX or ODF). However, only the iWork ’09 suite is capable of reading the older format and migrating to the newer format, because later versions can only read the ZIP-based form.

Note that since the Mac App Store was launched, Pages has been distributed independently, and no longer packaged together under the iWork name. During this transition, some ‘advanced’ features were dropped including RTF import/export for Pages (Wikipedia). Furthermore, the native file format has been changed from being XML based into a binary form (see below).

**Structure of the obsolete iWork format (iWork ’05 to iWork ’08)**

* Container: OS X Bundle
  + In this case, the Bundle is a folder with a “.pages” extension, containing a “Contents” sub-folder that holds the package information file (“PkgInfo”).
  + Note that the entire Bundle directory may be zipped for export/email and given the file extension “.pages.zip”
* Content:
  + Compressed XML under a proprietary schema
  + Optionally, a preview/thumbnail image.

**Example (generated using Pages 3.0.3)**

|  |  |
| --- | --- |
| **Path** | **Description** |
| lorem-ipsum.pages | Pages Document |
| lorem-ipsum.pages/Contents/PkgInfo | OS X Bundle package information file |
| lorem-ipsum.pages/index.xml.gz | XML document content |
| lorem-ipsum.pages/QuickLook/Thumbnail.jpg | Thumbnail for desktop previews. |

You can download the example file from [here](https://github.com/anjackson/format-corpus/tree/master/corpora/variations/variations/application/x-iwork-pages-sffpages/08-3.0.3).

**Structure of the deprecated iWork ’09 Pages format**

* Container: ZIP
* Content:
  + Compressed XML under proprietary schema (as used by iWork ’08)
  + Optionally, a preview/thumbnail image.

**Structure of the current Pages format**

* Container: ZIP
* Content:
  + Proprietary binary format – a compressed variant of Protobuf (iWork 2013, 2014).
  + Optionally, a preview/thumbnail image.

## Key adopters of the format (e.g. large repositories or academic libraries, domains)

OS X is popular with academic researchers and iWork is well known, but it is not currently known whether this has led to significant usage of Pages within that community.

References in key preservation policies / format guidance for repositories:

* Archaeology Data Service: Not listed (Archaeology Data Service).
* Library of Congress: Not listed (Library of Congress).
* National Archives and Records Administration: Not listed (NARA).
* Archivematica: Not listed (Archivematica, 2014).
* Library and Archives Canada: Not listed (Library and Archives Canada).
* The National Archives of Australia: Not listed. (National Archives of Australia, 2011).

## Applicable MIME media types

application/x-iwork-pages-sffpages, application/vnd.apple.pages

## Applicable file extensions

.pages .pages.zip

## The organization/individual/company that originally developed it

Apple

## The organization/individual/company that currently maintains it

Apple

## Availability and location of specifications (direct URLs if available)

None

## Brief information about patent/license issues

Proprietary. No known licencing scheme.

## Key related links (Websites describing it, documentation, etc.)

* <http://en.wikipedia.org/wiki/IWork>
* <http://fileformats.archiveteam.org/wiki/IWork>
* <https://wiki.documentfoundation.org/DLP/Libraries/libetonyek>
* Sample [Pages](https://github.com/anjackson/format-corpus/tree/master/corpora/variations/variations/application/x-iwork-pages-sffpages) files.

## Risk summary

* **Forced Obsolescence**
  + The oldest format implementation is obsolete (no longer support by Apple)
  + The second generation of the format is now deprecated and may be at risk (if Apple repeats their prior behaviour and drops format import support for the iWork ’09 format when the next major release of the software is made)
* **Proprietary Format**
  + All generations are proprietary, with single vendor support, closed source, closed specification, and almost no viable open source implementations at this time
  + LibreOffice support for iWork is very limited at present (see [here](https://wiki.documentfoundation.org/DLP/Libraries/libetonyek))
* **Encryption**
  + Since iWork ’09, Pages documents can be password-protected
* **Missing font information**
  + Where not easily substituted, non-embedded fonts could lead to loss of critical information
* **External References**
  + Externally referenced content may be difficult to identify, collect and preserve
* **Embedded Resources**
  + Resources can be embedded within the main bytestream, each of which may present new preservation issues.

## Mitigation of key risks

* **Mitigating forced obsolescence risk**
  + Re-save in the latest iWork format, and defer potential issues around migration until a later date, by which time the open source tools may have improved sufficiently
* **Mitigating proprietary format risk**
  + Export to alternative formats. Pages supports export to:
    - PDF, Word (DOC or DOCX), ePub, Plain Text.
  + Access via emulation:
    - Providing legitimate copies of OS X and Pages can be found.
    - Further complicated by the fact that current versions of Pages are distributed via the AppStore and so older versions are not easily available.
* **Mitigating encryption, missing fonts, external references and embedded resources risks**
  + Examine incoming material to check for the presence of these dependencies

# References

Archaeology Data Service. (n.d.). *Archaeology Data Service / Digital Antiquity Guides to Good Practice*. Retrieved from http://guides.archaeologydataservice.ac.uk/g2gp/TextDocs\_2

Archivematica. (2014). *Format Policies*. Retrieved from https://www.archivematica.org/wiki/Format\_policies

*iWork 2013*. (2014). Retrieved 2014, from File Formats Wiki: http://fileformats.archiveteam.org/wiki/IWork#iWork\_2013

Library and Archives Canada. (n.d.). *File Format Guidelines for Preservation and Long-term*. Retrieved from http://www.collectionscanada.gc.ca/obj/012018/f2/012018-2200-e.pdf

Library of Congress. (n.d.). *Recommended Format Specifications. Textual Works and Musical Compositions*. Retrieved from http://www.loc.gov/preservation/resources/rfs/textmus.html

NARA. (n.d.). *NARA 2014-04: Appendix A, Revised Format Guidance for the Transfer of Permanent Electronic Records – Tables of File Formats*. Retrieved from http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#textualdata

Wikipedia. (n.d.). *Pages (word processor) - Version history*. Retrieved from http://en.wikipedia.org/wiki/Pages\_(word\_processor)\_-\_Version\_history