**Test ARCHive Format Risk Report**

**Overview**

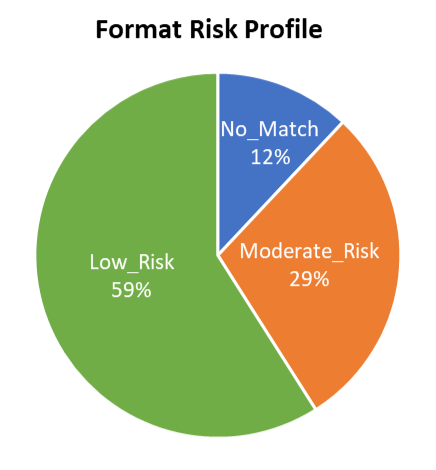
This report summarizes the test group holdings in ARCHive as of 2023-11-03.

* 3 collections
* 15 AIPs
* 17 versions of 9 formats (each format has 1-4 versions)

For the rest of the report, “format” refers to the specific version of each format.

**Risk Profile**

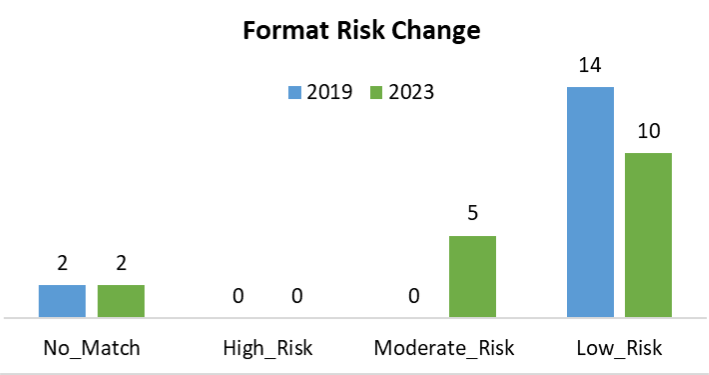
The percentage of formats at each NARA risk level for the entire group.



The formats that do not have a NARA risk level are Comma-Separated Values (CSV) with no PUID, Office Open XML Document, and Office Open XML Workbook. Not matching NARA does not necessarily mean they are high risk, but simply that no files in that format have been donated to NARA.

**Risk Change**

The percentage of formats at each NARA risk level from the 2019 and 2023 analyses.

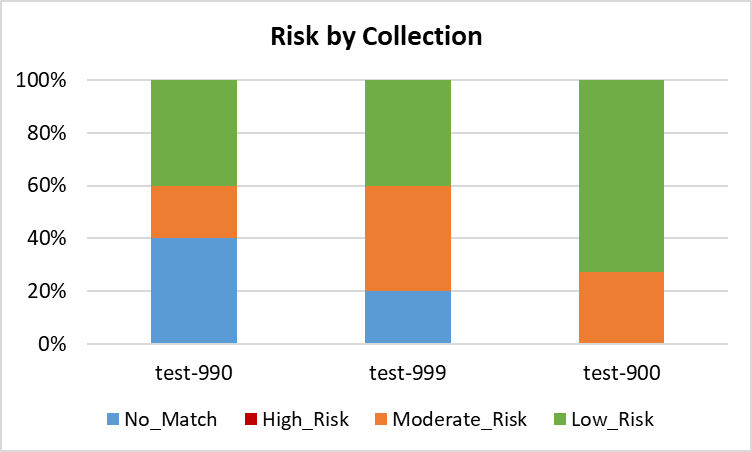


There is one new format since the 2019 analysis: Office Open XML Document 2007 onwards.

Macromedia Flash Data (versions 5-7) and PDF/A version 1a all increased to moderate risk.

**Risk By Collection**

The percentage of formats at each NARA risk level in each collection. The goal is to visualize how many collections have a high percentage of risk vs. low percentage of risk. The risk report spreadsheet (“Collection\_Risk\_Levels”) has the risk data for individual collections.



**Formats per AIP**

The purpose of this analysis is to visualize the complexity of AIPs in terms of the number of formats. Having more complex AIPs could make preservation actions like format migration more difficult, and you may wish to narrow the boundaries of future AIPs. Having more formats than you expect in AIPs may also mean that your format identification tool listed multiple possible formats for the same file and needs to be adjusted.

There are 15 AIPs. Of these, 10 (67%) contain a single format. There are 2 AIPs (13%) with 5 or more formats, with 10 formats being the highest.

In this histogram below, other than the first bar, the first number in the label is not included in the bar. So, the bar [1,2] has the number of AIPs with 1-2 formats but the bar (2,3] has the number of AIPs with 3 formats.

