Image Type Usage

The set of images present in the repository allows for the building of a document in its original look and feel.

➤ *Page thumbnails* can represent the general look of an entire issue or section, or can be used to graphically represent search results.

➤ *Page snapshot images* are used to simulate the original document look and feel in the browser reading application

➤ *Primitive images* are used to simulate the original view of a page component such as article or article elements.

3.4.2 PDF Documents

The original documents formats such as MS Word, PostScript or TIFF, are converted and stored in the Olive repository as unified PDF files. This is order to create a set of homogenic files which my be needed for download and PDF delivery purposes

3.4.3 XML and XSL

Aside from the entity XML files that represent the content itself, the repository contains XML files that serve as auxiliary data. These describe the Repository's structure and provide additional sources of information used by the client applications:

➤ Documents XML (Table Of Contents)

➤ Section XML

➤ Authors dictionary

➤ Predefined collections description

➤ Keywords

3.5 Storage size considerations

The Repository size may vary, and depends on a number of factors. These include page size, number of pages, and usage of graphics, color and page layouts. The numbers below are averages provided as a rough guideline to the necessary storage calculation. It is important to note that repository size is flexible and can be optimized by the ability to adjust the image resolution levels and compression methods.

3.5.1 Born Digital source materials

The follow are the page size approximations of example document processed by Olive software:

Format	A3 (Newspaper Size)	A4 (Standard Size)
PDF	100-200K	20-50K