Preserving the past and present for the future:

Digital Preservation Programme at the National Library of New Zealand

Hōtaka Matapopore Matihiko | Te Puna Mātauranga o Aotearoa
Digital Preservation at the National Library

“At the National Library our job and our passion is to look after the memory of our nation. Increasingly that memory is encapsulated in a digital world… Our challenge is preserving these precious items so that future generations can access these memories.”

- Bill Macnaught, National Librarian

Like physical items, digital items need care and attention to make sure they will be preserved and remain accessible in the future.

The National Digital Heritage Archive (NDHA) ensures that the National Library of New Zealand can meet its mandate to build, preserve and provide access to the Library’s digital collections.

Digital items collected or digitised by the Library are held in the library’s digital preservation system, also referred to as the National Digital Heritage Archive.
The NDHA Collection

This year, 2018, marks 10 years since the NDHA was first launched in 2008. To date, the archive holds approximately 21 million files.

The following diagram shows what the NDHA contains (as of 31 October 2018).
The following graph shows the growth in the number of intellectual entities (IEs) and files contained in the repository.

Since 2010, we have seen a marked increase in the number of files used to make up an object (intellectual entities), reflecting an increase in complexity.

An intellectual entity consists of all the files that relate to a single item. An intellectual entity can be a single file (e.g. a scan of a painting) or it can be many files that make up a whole item (e.g. all the scanned pages of a book).
Growing Repository

The size of the digital preservation repository has been growing since its inception in 2008.

![NDHA Repository Size (Terabytes) 2008-2018](chart.png)
Content Format: A Key Challenge

One of the key challenges for digital preservation is their file formats which can impede access.

Digital content in the repository is in 210 different file formats. The pictures below show the most common formats in the NDHA.

Formats with more than one million files:

- **TIFF**: 10.6m Files
- **XML**: 6m Files
- **JP2**: 2.5m Files

File formats with more than one hundred thousand files:

- **PDF**: 566k Files
- **TXT**: 367k Files
- **W**: 225k Files
- **LOG**: 159k Files
- **JPG**: 133k Files
- **ARC**: 110k Files
- **WARC**: 110k Files
The pie chart below shows the proportion of the various types of content in the NDHA, as of 31 October 2018:

- **Periodicals**: 39.99%
- **Unpublished**: 17.52%
- **Digitised Image**: 17.73%
- **Monographs**: 7.14%
- **Newspaper**: 0.29%
- **Magazine**: 0.01%
- **HTML Serial**: 0.22%
- **Digitised Sound**: 0.46%
- **Digitised Video**: 0.01%
- **Digitised Books**: 0.01%
- **Audio**: 0.38%
- **Web Harvest**: 1.57%
- **Video**: 0.03%
- **Archives NZ Content**: 14.63%
Preserving the NZ Web

The National Library collects New Zealand web content via two different initiatives within its web archiving programme.

The first initiative selects specific sites for regular harvesting, cataloguing, digital preservation and access.

The second initiative, known as the ‘Whole of Domain’ (WOD) harvest, bulk collects the New Zealand web presence, typically once a year. These harvests collect all .nz websites that are publicly accessible. These domain harvests total over approximately 1.6 billion files and 120 terabytes of data.


For more information, visit programme website: https://digitalpreservation.natlib.govt.nz/

Contact email: ndha.admin@dia.govt.nz