INTRODUCING THE FORMAT LIBRARY HELPER

Jay Gattuso

National Library of New Zealand

New Zealand

Jay.gattuso@dia.govt.nz

Abstract – This poster describes the Format Library helper tool built by the National Library of New Zealand to aid file format research and analysis.

Keywords – format identification, PRONOM, format library, knowledge base

Conference Topics – Building the Capacity & Capability.

I. INTRODUCTION

This poster describes the 'Format Library Helper" tool that was built by the National Library of New Zealand.

The Poster will show the user interface, and be annotated to describe each section, and what its purpose is.

A. Summarizing the tool

The poster will describe:

- the links/crosswalk we've made between common format ID spaces, PRONOM, LOC, NARA and Wikidata.

- The search tool for exploring format identifiers via various facets (keyword, MIME, ext, puid)

- The documentation we collect per format including specification and technical writeups

- How the tool easily links a user to PRONOM, WIKIDATA, LOL, NARA information sources

- The file/collection counting we do, including the augmented ubiquity count that includes sources of files as well as counts of files

- The TestSet files we collect per format to support format research

- The incorporated early version format risk recording system that covers technical documentation, access to high quality digital preservation format notes (LOC and NARA) and render pathway.

- The sharable JSON file that records all our data per format

- A summary of the inputs required to populate / update the tool.

- A summary of the automated monthly reporting we're driving from the tool data

- The github location of the codebase for standing up a new instance for any user/organisation + instructions

- An invitation to use / comment / contribute towards progressing the tool

II. UNIQUE CONTRIBUTION TO DIGITAL PRESERVATION

There is no similar tool in the digital preservation toolkit (based on extensive experience in the format space). As well as the feature set of tool itself, the exposed format mappings are new, and the underlying risk model whilst immature and ripe for comment, is also new.