QUESTING FOR INTEROPERABILITY WITH NEWSPAPER METADATA

Dr Emily Bell
Loughborough University
e.bell@lboro.ac.uk

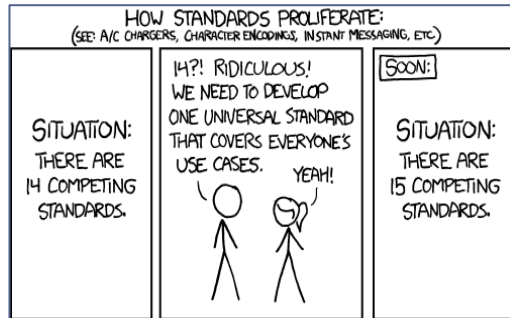## The problem:

9 Nine Collections

Including the National Libraries of Mexico, Finland, the Netherlands, Australia, New Zealand, and the British Library

2 Two Formats

1. XML
2. JSON

Every one structured slightly differently…

Including METS/ALTO, METS/PREMIS, METS/MODS, METS/MIX, MPEG-21/DIDL, JISC, Dublin Core, Bespoke, API

The solution:

Using…
- Ontologies
- Digitisation Guidelines
- Websites
- Academic Literature
- Grey Literature
- Staff Interviews
- Rumours
- Retro-Engineering

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!

SOON:

SITUATION: THERE ARE 15 COMPETING STANDARDS.

xkcd.com/927/

The Oceanic Exchanges Atlas of Nineteenth Century Newspapers

18 Copyright

18.0 Language Variants

Urheberrecht ©; Copyright ©; tekijänoikeus; Derechos de autor.

18.1 Technical Definition

Specifies a copyright holder.

18.2 Usage Notes

Copyright in the nineteenth century was a controversial topic. The 1842 Copyright Act, which focused on copyright for books, did not mention newspapers directly; as such, the reprinting of articles from other newspapers continued unabated. This was also complicated by the 1842 Act binding copyright to the author, when the majority of newspapers published articles anonymously. An 1881 judgement classed newspapers as books for copyright purposes, but there was still need for further rulings to specify which **article categories** were protected. In America, there was no copyright protection for foreign authors until 1891. This meant that the reprinting of articles remained an integral feature of nineteenth-century American journalism.

All newspaper materials in Delpher have been cleared for third-party copyright claims and released into the public domain, while Trove has done so with most of its collection; any data from in-copyright works are clearly identified on the web interface and in API results. Although many of the newspapers would now be in the public domain in the UK anyway, the British Library retains copyright for the digital files because of the skill, judgement and labour used in the creation of the digital image.

18.3 Examples:

- "The extent of **copyright** in news material itself was still at issue in the 1890s, when there were important decisions on the protection of press agency information." [DNCJ, GL, 143]
- "The 1842 Copyright Act protected **copyright** in "books," defined in a broad manner. However, section 18 contained special rules applying to the following genres: encyclopaedias, reviews, magazines, periodical works, or works published in a series of parts." [Cooper, 662]
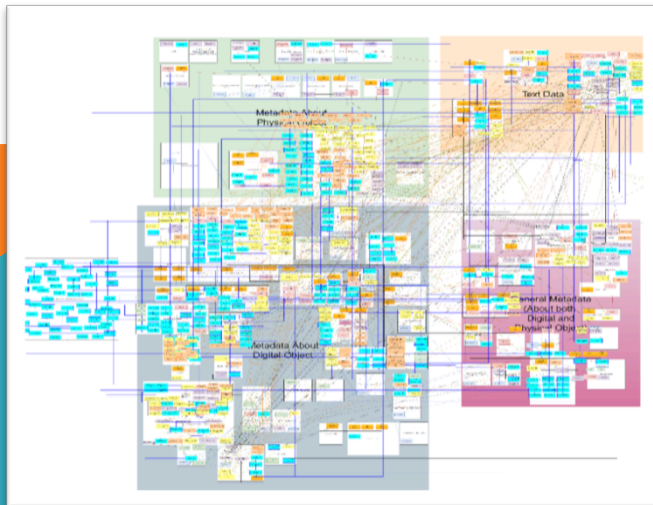
18.4 Metafield Notes

This field appears in the adapted Dublin Core file structures, and always as a string.

18.5 Individual Collection Notes

DEMP: There are different fields for the copyright holder for the **newspaper title** (didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ dc:rights) and the digital record (DEMP::didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\dcx:recordRights).

18.6 Instantiations

B1GI::issue\metadataInfo\sourceLibrary\copyrightStatement	DUB::SLI	Copyright © The British Library Board
B1GL::issue\cp	DUB::SLI	Copyright © The British Library Board
B2GL::issue\cp	DUB::SLI	Copyright © The British Library Board
B2GI::issue\metadataInfo\sourceLibrary\copyrightStatement	DUB::SLI	Copyright © The British Library Board
DEMP::didl:DIDL\didl:Item\didl:Component\didl:Resource\srw_dc:dcx\ dc:rights	DUB::SLI	AD NieuwsMedia B.V.

The solution: