

Labels / Types / Categories within the PAGE XML format (2013 version)

Page Types

- Front cover
- Back cover
- Title
- Table of contents
- Index
- Content
- Blank
- Other

Levels of Text Objects

- Regions (blocks)
 - Text Lines
 - Words
 - Glyphs (characters)

Region Types / Subtypes / Labels

- Text Region
 - Paragraph
 - Heading
 - Caption
 - Header
 - Footer
 - Page number
 - Drop capital
 - Credit
 - Floating
 - Signature mark
 - Catch word
 - Marginalia
 - Footnote

- Footnote continued
 - Endnote
 - Table of contents entry
 - Other
- Image
- Line Drawing
- Graphic
 - Logo
 - Letterhead
 - Decoration
 - Frame
 - Handwritten-annotation
 - Stamp
 - Signature
 - Barcode
 - Paper grow
 - Punch hole
 - Other
- Table
- Chart
 - Bar
 - Line
 - Pie
 - Scatter
 - Surface
 - Other
- Separator
- Maths
- Chemical Formula
- Musical Notation
- Advert
- Noise
- Unknown / Other