Visiolink's XML Import Requirements

General XML requirements – just as a reminder

All XML files must be well-formed, and if DTDs or Schemes are referenced from them, valid with respect to these DTDs or Schemes.

File formats and file naming

XML files: You may deliver one xml file per article, or one file per page, or one per page and one per article, or even one file containing the whole publication, or whichever output your productional system generates. Chances are high that Visiolink can read your xml feed.

Zip files: You even may submit XML feeds as zip files

Filenames must contain **publication date**, e.g. in YYYYMMDD form. If it is zip files you deliver, the naming of the contained xml files is irrelevant.

Structure

Visiolink's generic xml import can handle many different xml structures, as long as these few conditions are met:

1. It must be possible to identify articles, i.e. information about items as text spans, text blocks, images and where on which page they are placed, will not do alone.

2. For each identifiable article, it must be possible to identify the following **metadata** and **contents**:

Metadata:

Section number or section indicator: Which section the article is part of, given the newspaper is organised in sections.

Page number: Which page the article is placed on; page numbering may be absolute or section relative.

ID: a unique article identifier

Geometry: Where on the page, and other following pages, the article is placed and how big an area it covers. Article geometry can be defined as a set of boxes or contour paths.

- **Boxes** can be defined as either
  [x position left, y position top, width, height], or
  [x position left, y position top, x position right, y position bottom],

and may contain [page number] with multi-page articles.

- **Contour paths** can be defined as
  $[x_1, y_1, x_2, y_2, x_3, y_3, ..., x_n, y_n, x_1, y_1]$,

where each $[x_a, y_a]$ describes one point, and for each a: $x_a = x_{a+1}$ or $y_a = y_{a+1}$, i.e. each point must be connected to the next point through an either horizontal or vertical line.

Coordinates are expected to be specified in one of the following units: pts, mm, dmmm, percentage of page width or height. During import all coordinate values will be translated to pts.

Article category (optional): A marker that might be shared between several articles with the same topic, such as "Sports", "Local news" etc.

Contents - all of which are optional:

Supertitle: The supertitle of an article.

Title: The main title of an article.

Subtitle: The subtitle of an article.

Blurb: Introductory section, a.k.a. ingress of an article.

Byline: Author information, preferably sub structured into:

- **Name:** The author's name, i.e. first name(s) and family name.
- **Role:** Additional relevant information bout the author.
- **Email:** The author's e-mail address.

Body: The main body text of an article, preferably sub structured in paragraphs and paragraph headings.

Fact boxes: All kinds of an article's satellite elements, such as background information, tables, that are either realised as floating objects or accumulated after the body.

Top boxes: Satellite elements that are supposed to be presented on top of the article, e.g. with book reviews, in which case a top box would contain book title, author, ISBN, and five star score.

Quotes: Quotes with or without quote source. Note: This is currently not supported in clients.

Images: Images linked to an article. For each image, the following subelements must be identifiable:

- **File reference:** The filename of the image file, perhaps together with a path. Make sure the file itself is submitted as well, and retrievable by either a static path or the path specified in the file reference.
- **Caption (optional):** The caption that is supposed to appear together with the image.
- **Photographer (optional):** Name of the photographer in charge of the image.
- **Copyright (optional):** Name of the copyright holder for the image.
- **ID (optional):** An image identifier, only important when needed as sort key.
- **Width (optional):** The width on the page that the image covers, only important when the image area on the page is needed as sort key.
- **Height (optional):** The height on the page that the image covers, only important when the image area on the page is needed as sort key.