Puzzled Pint
Puzzle Archive Format

2016-12-04

Intent / In-Scope

- An intermediate data layer to take what we have (a collection of PDFs and freeform CMS text) and normalize it into structured data.

- Not a final, permanent, format.

- Data that can be easily parsed, massaged, and then loaded into a final format to seed the new website's database.

- Minimal and flexible.

- XML. Easy to hand-edit. Easy to start with a template. Easy to copy/paste from website into editor.

- Data format validation. test.sh wrappers calls to xmllint to look for bad XML, divergence from the DTD, or missing linked files.
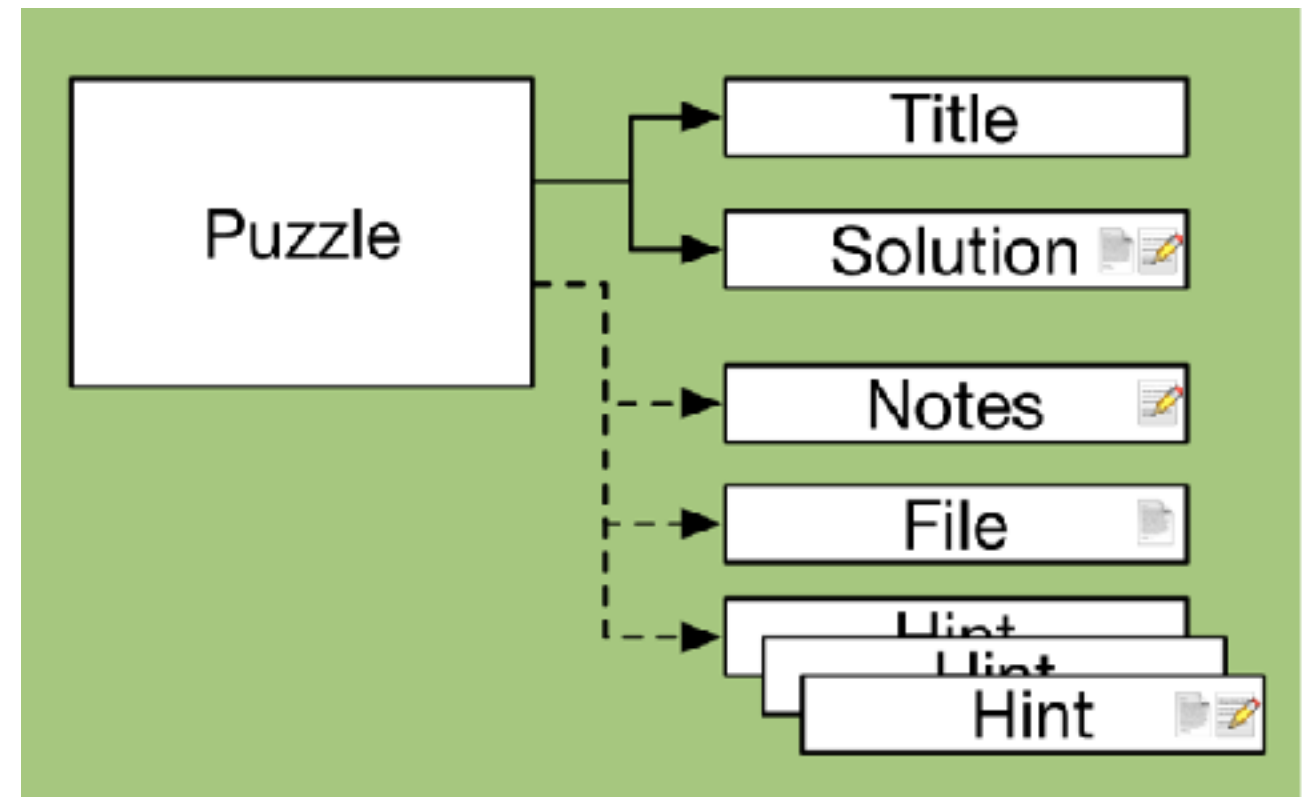
Out-of-Scope

- Wasting time arguing over the data structure. I had to start with something, then revised it a couple of times as I found non-conformate content. It's fairly solid now. Unless we run across more content that doesn't conform to the existing data structure, consider the DTD final for this task. Save the arguing for the final website's data structure.

Structure

- This is a "file based" database.

- Each month is in a YYYY/MM directory.

- Each month has a month.xml describing the month and its puzzles. See also: sample_month.xml and month.dtd

- There are also the various support files — puzzle PDFs, answer sheets, solutions, etc.

- Upon import, I was anal-retentive and renamed the puzzle PDFs to {number}-{title}-{puzzle | solution}.pdf. That let me better see what files mapped to what puzzle. Some of the original names were pretty silly and embarrassing ("mypuzzle.pdf" and "final, version 3" and such). We don't have to stick to this naming.
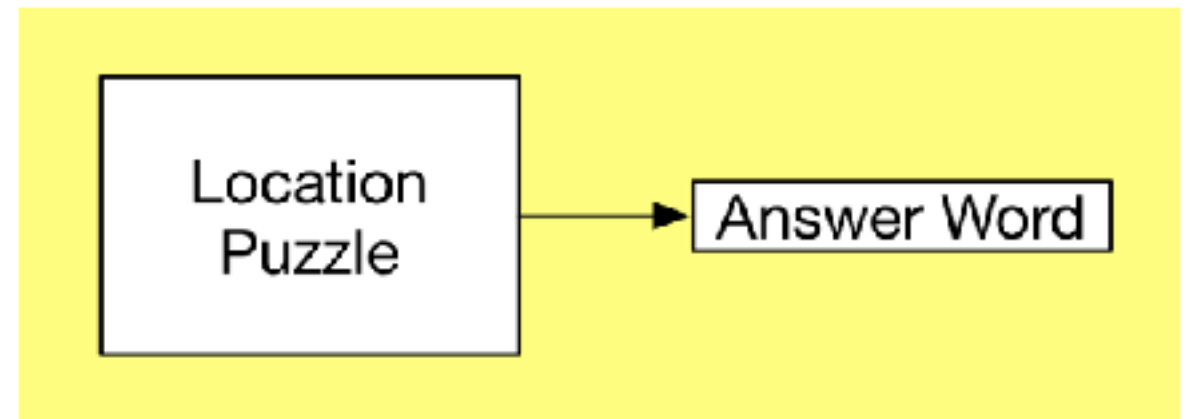
The Puzzle

- The smallest "interesting" unit of data is the puzzle.

- The puzzle has, at minimum:

  - A title.

  - Content: freeform text and/or a linked [typically PDF] file.

  - A solution: freeform text and/or a linked [typically PDF] file.

- It can also have any number of hints (including zero). Any of these can also be freeform text and/or a linked file. In theory, the website could render these as progressive clues.
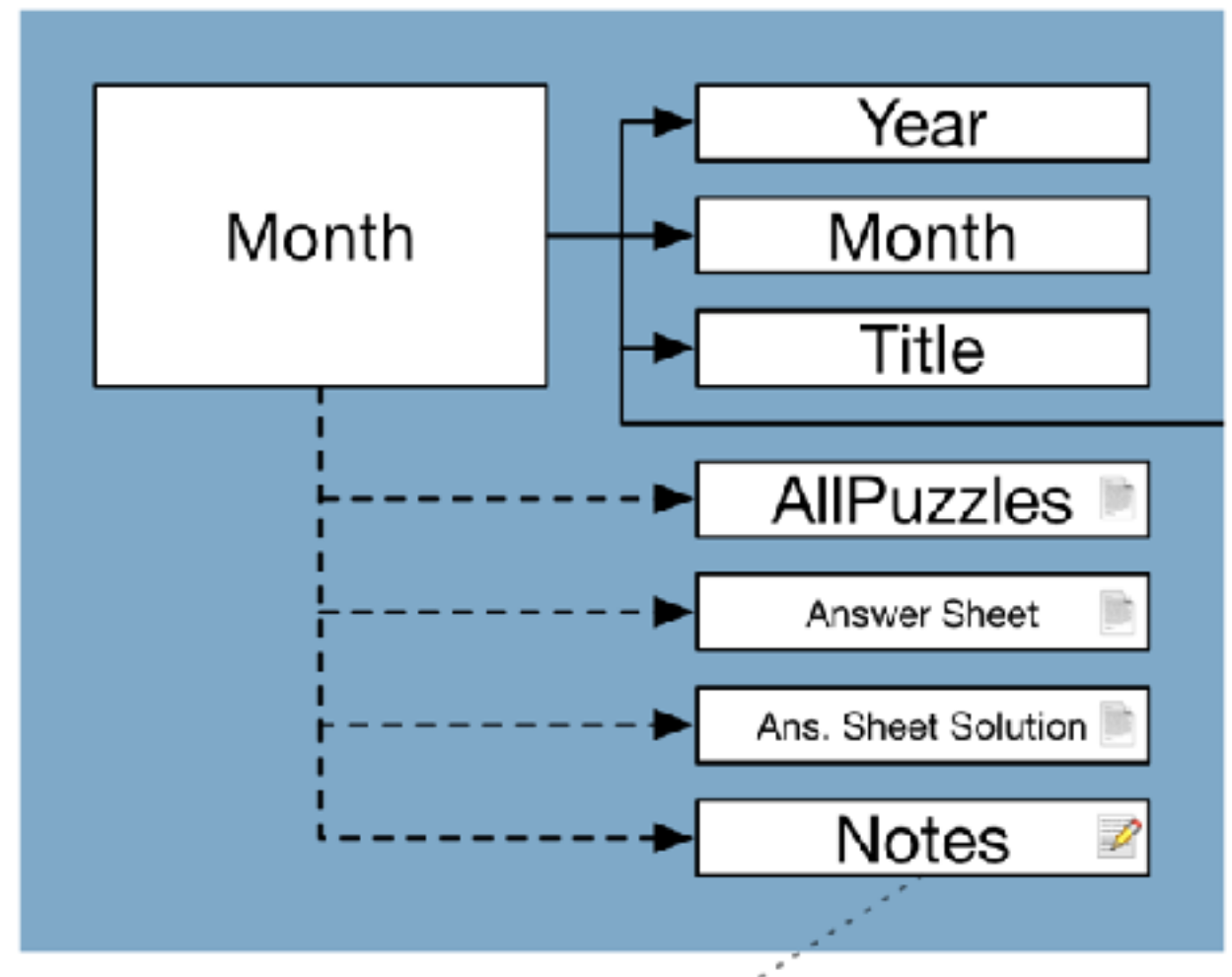
Location Puzzle

- The Location Puzzle builds on the normal puzzle by including an official answer word.
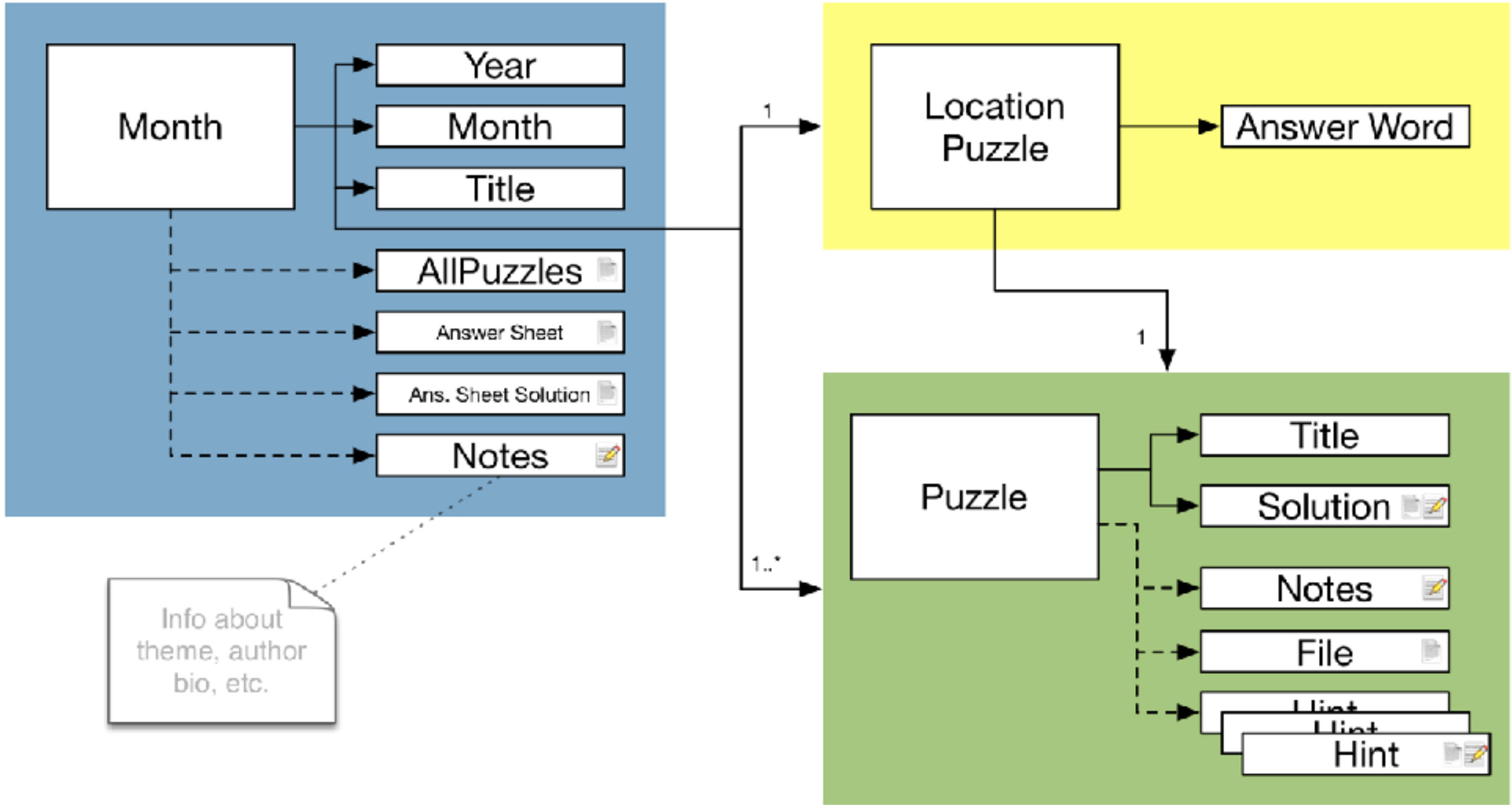
Month

The Month is the largest unit of data in this format. It includes:

- Metadata — year, month, title.

- Freeform text, typically used to introduce the theme, include an author bio, and so on.

- A location puzzle.

- One or more puzzles.

- Optionally, an allpuzzles.pdf.

- Optionally, an answer sheet.

- Optionally, an answer sheet with solutions filled in.

Month
- Year
- Month
- Title
- AllPuzzles 📄
- Answer Sheet 📄
- Ans. Sheet Solution 📄
- Notes 📝

Info about theme, author bio, etc.

Location Puzzle
- Answer Word

Puzzle
- Title
- Solution 📄📝
- Notes 📝
- File 📄
- Hint
- Hint
- Hint 📄📝

1

1..*

1

1

Legend

📄 = Attachment (PDF)

📝 = Freeform Text (Markdown)

Validation

Run the top level test.sh script and this will spider into each year/month folder and:

- Validate the XML against the DTD using xmllint.

- Validate that any href attributes point to real files.

- Warn of any href attributes that point to external domains.

Assorted Notes

- All freeform text in the XML is Markdown. We don't have to stick to this — for example, we can render it as HTML upon database import — but it's a start. They should be considered Github-flavored markdown. They occasionally make use of links, bold, italics, tables, and occasional code-blocks.

- Linked files are typically PDF, but could technically be anything. We have a few media files here and there (MP3, MP4).

- The hint file ./2015/06/00-location-hint1.html hotlinks images on snout.org.

- December 2011 is missing some solutions.