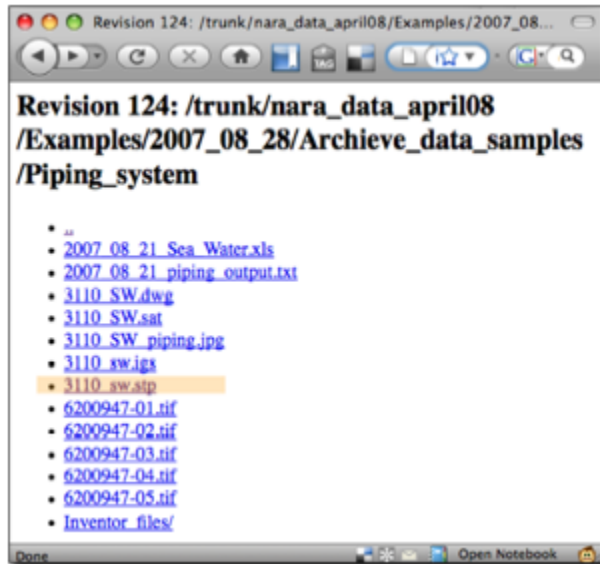


Basic Issues

Following on from some challenge problems set by Robert Chadduck and Mark Conrad (ERA), the WVU conducted a set of brainstorming sessions where they took stock of their progress to date and reflected over best future directions.

A repeated theme in those sessions was “finding the associations”. To explain this concept, consider the following fragment of ERA data.



This figure shows the contents of one directory containing technical specification data of the kind stored at ERA. The two important features of this directory are:

1. The relative infrequency of STEP files. This directory has only one: “3100_sw.stp”) while all the others are the products of other CAD tools (e.g. “3110_sw.igs”), graphics files (the .tif files), amongst other things.
2. The presence of some technical support documents. This directory has a “piping_ouput.txt” file and a “Sea_Water.xls” file .

(Note that there is nothing special about this directory- we can point to many other ERA-supplied data directories that have the same two features.)

In terms of our research direction, the above two features are tremendously significant:

- *Our current focus is on STEP documents.* Such a focus is inappropriate if, as shown above, much of ERA’s technical information is stored in a non-STEP format.
- *Our current focus ignores the technical support documents.* Such a focus is inappropriate if, as shown above, it is common for design documents are routinely augmented with such support documents.

The presence of the technical support documents in the archive is both interesting and significant:

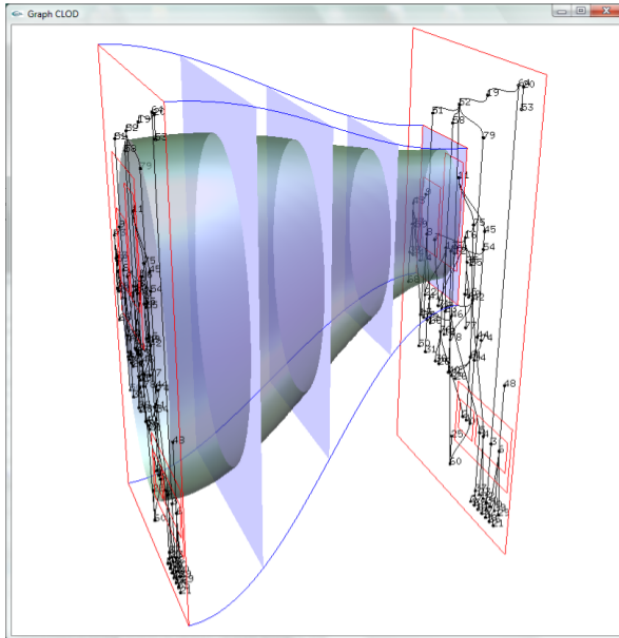
- Based on our own experiments with designing artifacts at WVU, we note that while designing an artifact, we often generate intermediary documents, intended to answer one particular question. For example, we might write a spreadsheet with the specific goal of determining the best trade-off between the wall thickness of a pipe, its cost, and its strength.
- These intermediary documents explain *why* a particular design was selected.
- However, *all too often these intermediaries are thrown away.*

If our goal is the use and reuse of technical documents, perhaps decades after they were first created, then we need to better support the capture and use of the supporting documents. The core technical problem of this goal is the discovery of *associations* between technical documents.

Operationally, this goal effects our work as follows:

1. *We need to thoroughly document the use of intermediary technical support documents*, as well as how they often are not part of a finished design stored in an archive. This is the current task of Victor Mucino and Jim Mooney.
2. *When data is being searched within the archive:* (possibly, decades after it was first entered into an archive). This is the task of Tim Menzies. Association discovery during data use can be augmented with information retrieval methods that combine (a) knowledge of the user and (b) smart indexes over the archive contents.
3. *When data is being entered into an archive:* We need to develop tools that increase the probability that we will capture these intermediaries. This is the current task of Tim McGraw. Association discovery during data entry can be augmented by adaptive graphical browsers that study who views what in the archive, and in what order (so two

concepts are associated more if they are commonly viewed together).



4. Also, we may need to look beyond STEP as the core representation of the technical specifications. This is the task of Jim Mooney.