An Automated Method to Store, Search, and Retrieve Wellbore Data

Case Struyk*
KC Petrophysics, Calgary, AB
kc.petrophysics@gmail.com

Summary

The oil industry currently uses manual and non-standardized methods to store, search, and retrieve wellbore data. The purpose of this paper is to show that this process can be automated, using techniques similar to those of a public library, by defining a metadata element set for wellbore data. The proposed method can be applied immediately to "structured data", such as LAS files, because software can extract information automatically from the body of the file to populate the metadata element set. However, application of this method to "unstructured data", such as Adobe "pdf" files, cannot be automated unless their "summary information tags" contain a populated metadata element set. The process of adding a populated metadata element set to a summary information tag is best done at the time the unstructured files are being created. This will require the wireline and coring industry to make some changes to their operation.

This paper explains data storing concepts using metadata, defines a metadata element set for wellbore data, and explains its application to structured and unstructured data files. The focus of is on wellbore data relating to logs and core. Logs, and possibly core, will benefit most from an industry standard because they are frequently accessed and/or exchanged between various parties. The method was successfully tested on a pilot project consisting of 20,000 LAS files.

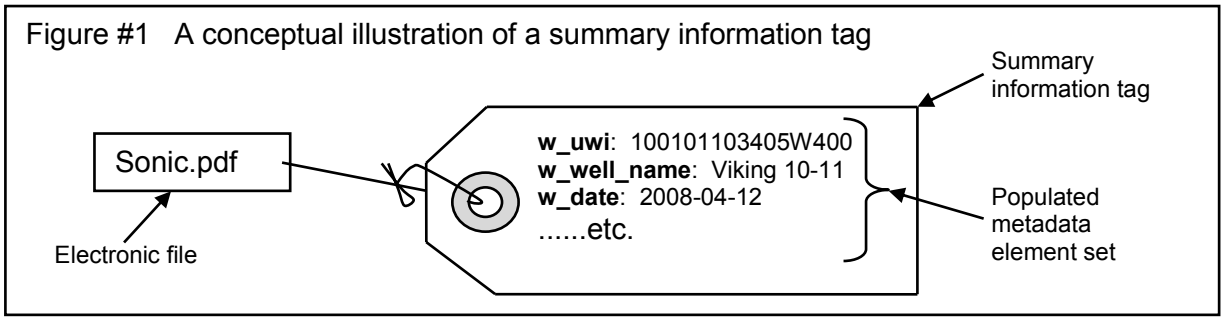
Introduction

The concept proposed for organizing wellbore data uses techniques similar to that of a public library. A public library collects basic information about each book using key words such as "title", "author", "publisher", etc., transfers this information to a library reference card, and then organizes the reference cards so that they can be searched to retrieve the book.

In the computer world, electronic files are analogous to books. Electronic files generally reserve an area within their file structure for "summary information" where metadata can be stored. This space will be referred to as the "summary information tag". The summary information tag is located under "file properties", and is rarely used by the oil industry.

Present day nomenclature refers to key words (e.g., title, author, subject, date, etc.) as "metadata elements", and a group of key words for a specific object (e.g., library book) as a "metadata element set". The reference card becomes a "populated metadata element set", that can be stored in the summary information tag (Figure #1), or programmatically extracted from structured files. The entire collection of library reference cards (i.e., populated metadata element sets) is today stored in a reference database.

Figure #1 A conceptual illustration of a summary information tag



Method

The Metadata Element Set for Wellbore Data

The metadata element set proposed for wellbore data is shown in Figure #2. The first three metadata elements are of an organizational nature while the remaining ones are descriptive. The three elements of an organizational nature are further illustrated in Figure #3.

Figure #2: Metadata Element Set for Wellbore Data

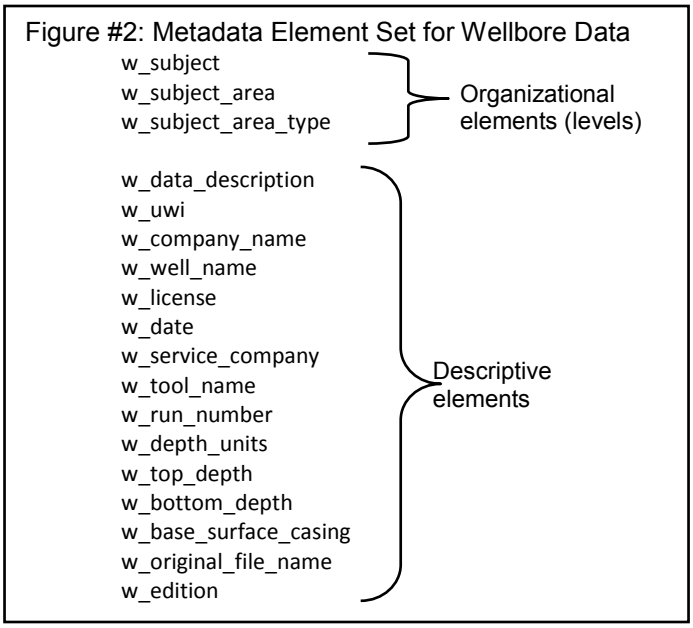
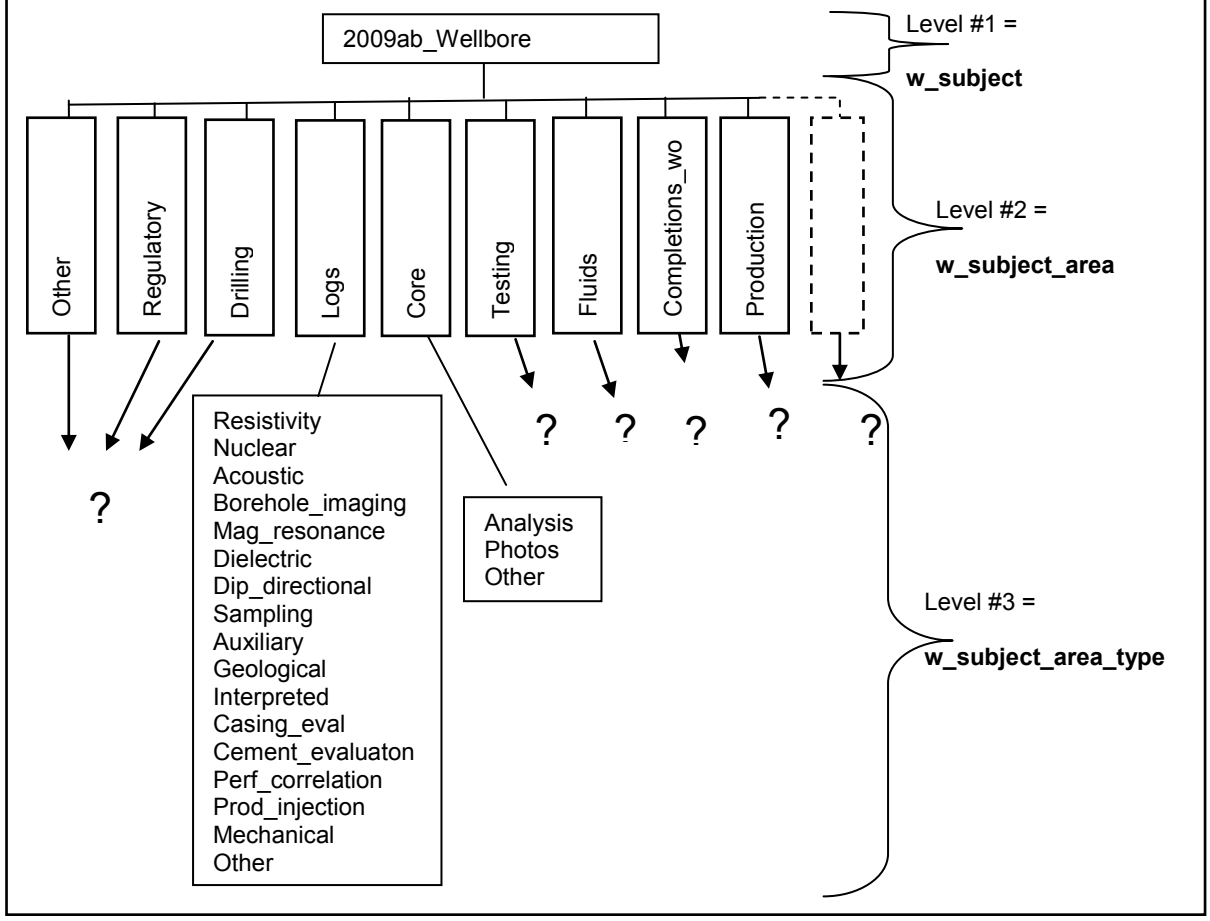


Figure #3: The three organizational levels for wellbore data (the schema)



Principal Categories for Wellbore Data

As illustrated in Figure #3, the metadata element "**w_subject_area**" is subdivided into a number of categories. These categories are intended to represent the principal wellbore related activities common to resource exploration/exploitation:

1) Application for regulatory permissions
2) Drilling the well
3) Gathering log information
4) Gathering core information
5) Gathering well tests information
6) Gathering fluid analysis information
7) Completion, work over, and abandoning the well
8) Recording production data.

Additional categories may be added as needed (e.g., AFE, accounting, etc.). Note that the above list includes four categories for information gathering (items #3 through #6). The categories that can benefit most from an industry standard are logs, and possibly core (items #3 and #4), as these are frequently accessed and/or exchanged between various parties. These two categories have been subdivided in the metadata element "**w_subject_area_type**" (Figure #3). The remaining categories for the metadata element "**w_subject_area**" have not yet been subdivided, because it is unclear if they require an industry standard.

Populating a Metadata Element Set

A populated metadata element set provides a summary of the contents of the electronic file. The means by which this information is obtained depends on whether the file is considered "structured" or "unstructured".

Files are considered "structured" if software can extract information automatically from the body of the file to populate the metadata element set. LAS files are considered "structured".

Files are considered "unstructured" if software is unable to extract this information from the body of the file. An example of an "unstructured" file is an Adobe Acrobat "pdf" file. The storing and searching of these files cannot be automated unless their "summary information tags" contain a populated metadata element set (Figure #1). The process of adding a populated metadata element set to a summary information tag is best done at the time the unstructured files are being created.

Testing the Concept

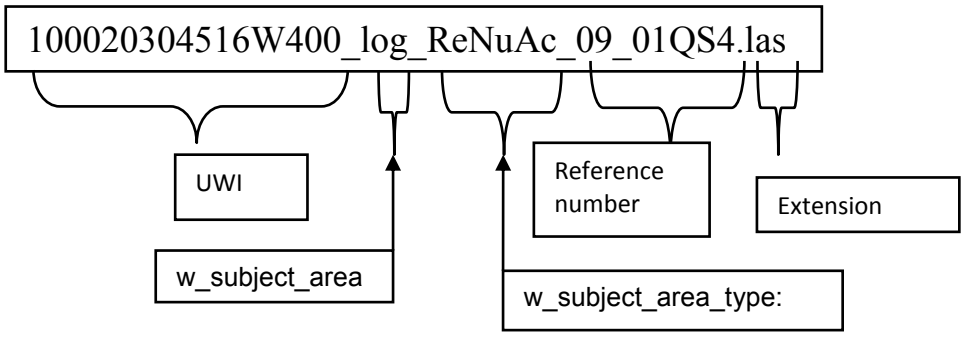
A pilot project consisting of 20,000 LAS files was used to test the concept. The test results were reviewed by a geological group, and some important shortcomings were identified. The model was then revised and the test rerun. The new test successfully automated the storage and retrieval of log data, and was also able to identify wells that failed to comply with well data submission regulations.

File Naming Convention Used in Pilot Project

Populated metadata element sets and their associated file names are collected into a reference database that can be used to retrieve the names of desired files. File names reveal important information about their contents (e.g., sonic.pdf, TVDMAIN.LAS, 01-35_HI_RES.LAS, etc.), but these names are not always unique. File names must become unique, however, when they are collected into a common directory in order to prevent overwriting issues. The files therefore require renamed during the archiving process. To avoid a loss of information, the original file name is stored in both the metadata element set, and within the LAS file. The file naming convention used in the pilot project is shown in Figure #4. The design of the file name is as follows:

1) The file name begins with the 16 character UWI, allowing files to be sorted by UWI.
2) A three character abbreviation for the “w_subject_area” follws.
3) A series of two character abbreviations for the “subject_area_type” is added next.
4) A reference number follows. It consists of “yy_aaaaa”, where “yy” refers to the year and “aaaaa” refers to five alphanumeric characters. This creates a numbering system with a base of 36, rather than the more common base 10 for decimal or 16 for hexadecimal. This reference number will handle about 60,000,000 files annually. The underscore within the reference number may be replaced by a short version of the company name.
5) The file name finishes with an extension indicating file type.

Figure #4: File naming convention used in pilot project



Conclusions

A method was proposed for automating the storing, searching, and retrieving of wellbore data. It was successfully tested on a pilot project consisting of 20,000 LAS files.

The concept can be applied immediately to structured data, such as LAS files. However, application of this method to unstructured data, such as Adobe “pdf” files, requires that the metadata information be made available in the summary information tag that is found in most electronic files. The process of adding metadata information to unstructured files will require the wireline and coring industry to make some changes to their operation. These changes are more likely to occur if there is general agreement on the metadata element set for wellbore data.

All parties will benefit by automating the storage and retrieval of log and core data. Regulatory bodies will be able to design an electronic submission process, improve their ability to monitor compliance to well data submission requirements, and provide industry better access to non-confidential data. Automation will allow oil, gas, and service companies to reduce their shipping and handling costs, while increasing their efficiency in data tracking, storage, and retrieval.

The Canadian Well Logging Society (CWLS) has recently set up a Metadata Committee to explore the possibility of designing a standard metadata element set for log and core data.

Acknowledgements

Special thanks to the many people who helped this article through its various stages of development.