



Maximizing the Value of Unstructured Sub-surface Data

The Resurgence of the Librarian in Digital
Form as a Knowledge and Learning Source

Presenter: Scott Tidemann, Vice President Petrosys Middle East & Asia Pacific
Email: scott@petrosys.com.au

Paper authors: Dirk Adams and Volker Hirsinger, Petrosys USA, Houston, Texas.
Robert Winsloe, Digital Earth Inc., London, UK

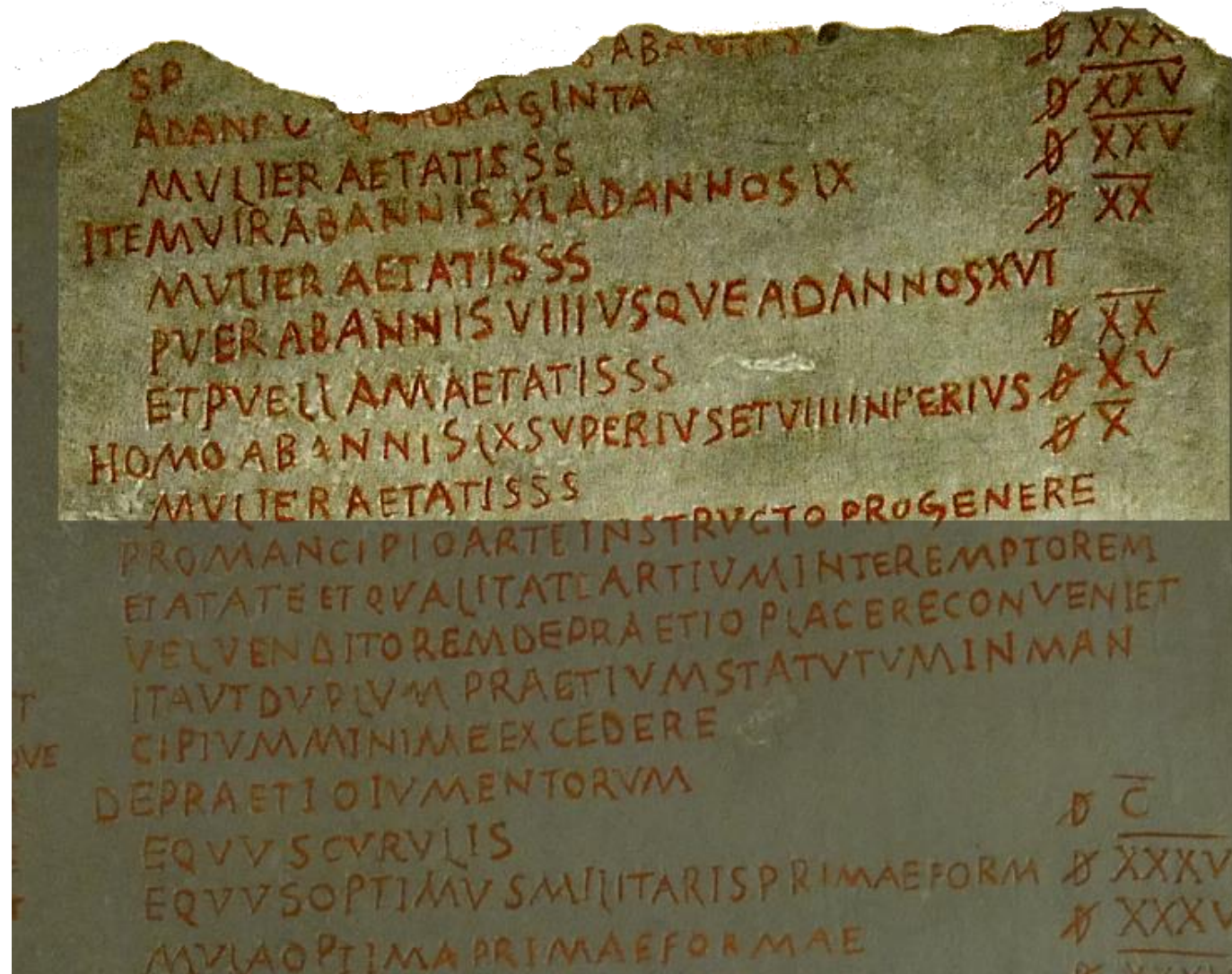
A Brief History of the Written Word

Egyptian
hieroglyphics



A Brief History of the Written Word

Roman
pricelist



A Brief History of the Written Word

18th. Century
Vellum



A Brief History of the Written Word

20th. Century Paper





A Brief History of the Written Word

Recent High Tech Printing

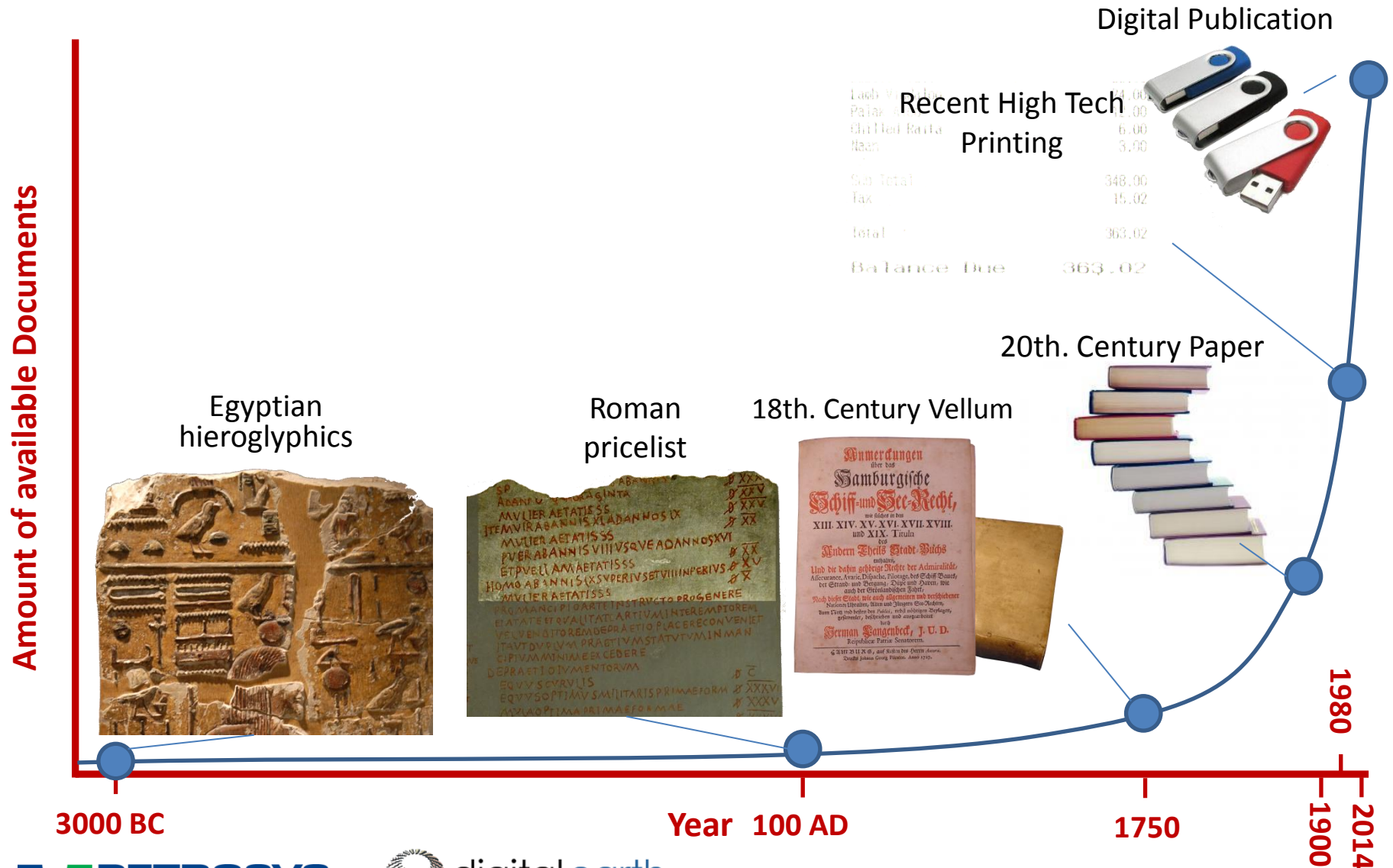
Lamb Vindaloo	24.00
Palak Aloo	12.00
Chilled Raita	6.00
Naan	3.00
Sub Total	348.00
Tax	15.02
Total	363.02
Balance Due	363.02

A Brief History of the Written Word

Digital Publication



History of the Written Word and Volume of Written Data



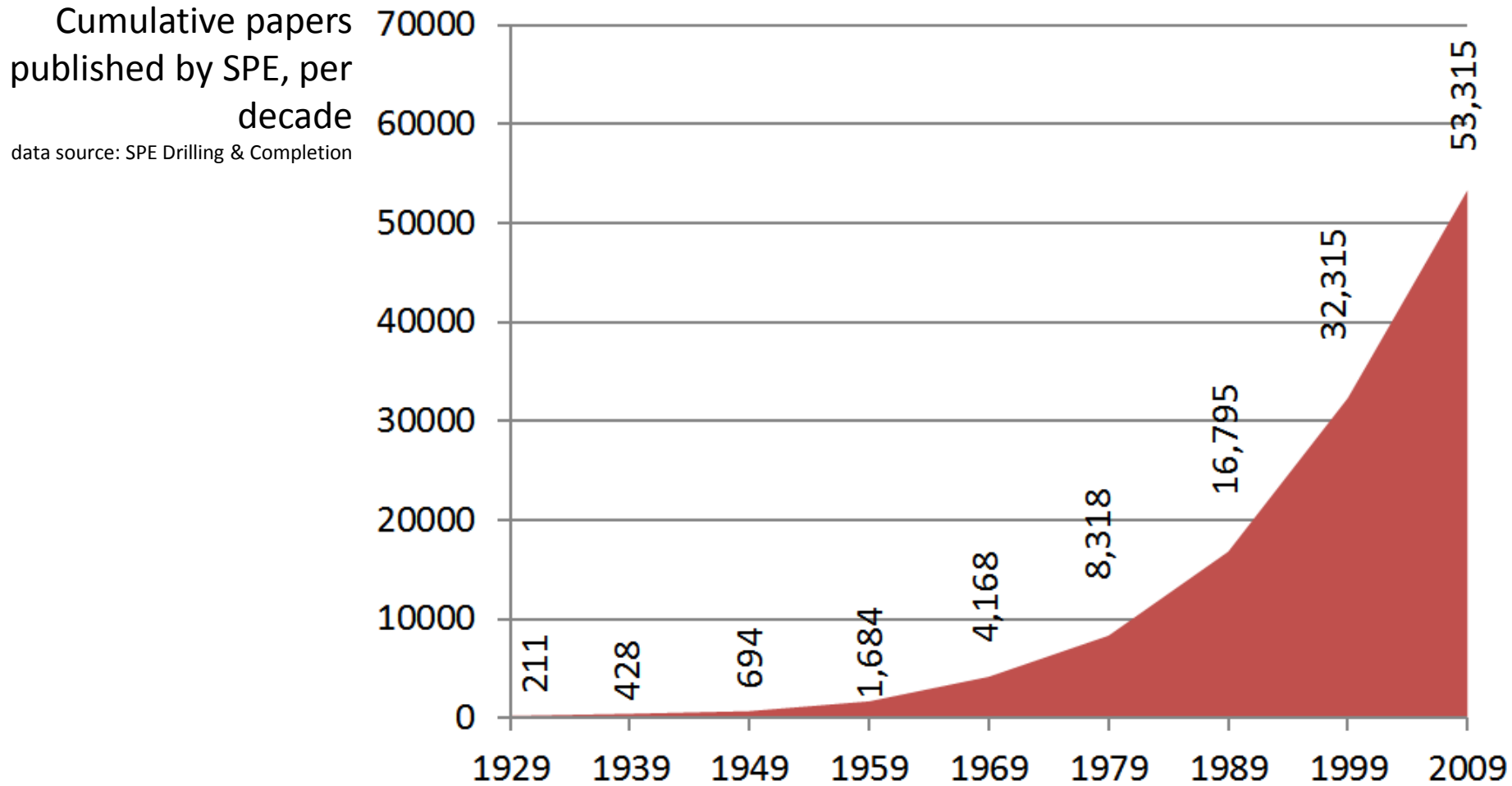
History of the Written Word and Volume of Written Data

As data managers, we have a responsibility to add value to the past, present and future investment of the petroleum E&P industry in the creation of knowledge, and to ensure that information remains readily accessible, highly visible, and secure in the long term.

Relevance Per Recorded Word



The Scale of Unstructured Text in E&P



The Scale of Unstructured Text in E&P


There is a huge volume of unstructured text containing E&P knowledge:

- Tabular reports – DST summaries, dumps from data stores
- Operational reports – well completions, reservoir performance, ...
- Field studies
- Scientific papers
- Text content in graphic documents – side labels, well log headers, ...

TABLE 2
ROCKEVAL PYROLYSIS DATA

TOB 8667	GEOCHEM SAMPLE	DEPTH/IDENTITY	TOC (%)	S1 (mg/g)	S2 (mg/g)	S3 (mg/g)	PRODUCTION INDEX	HYDROGEN INDEX	OXYGEN INDEX	TMAX (°C)
----------	----------------	----------------	---------	-----------	-----------	-----------	------------------	----------------	--------------	-----------

WELL COMPLETION LOG
ESSO NAUTILUS A-1

DEPT. NAT. RES & ENV

 PE600324

CONCESSION: PEP. 49 Hematite **STATE:** VICTORIA **BASIN:** OTWAY

LOCATION: Lat. 38° 58' 41" S Long. 142° 32' 46" E

ELEVATION: G.L. M.S.L. K.B. 93' **WATER DEPTH:** 327'

SPUDDED: April 13, 1968 **COMPLETED:** May 5, 1968 **DRILLED BY:** OCEAN DRILLING & EXPLORATION CO.

CLASSIFICATION: STRATIGRAPHIC TEST **STATUS:** PLUGGED & ABANDONED

TOTAL DEPTH: 6,597 ft **PLUGGED BACK T.D.:** 460 ft

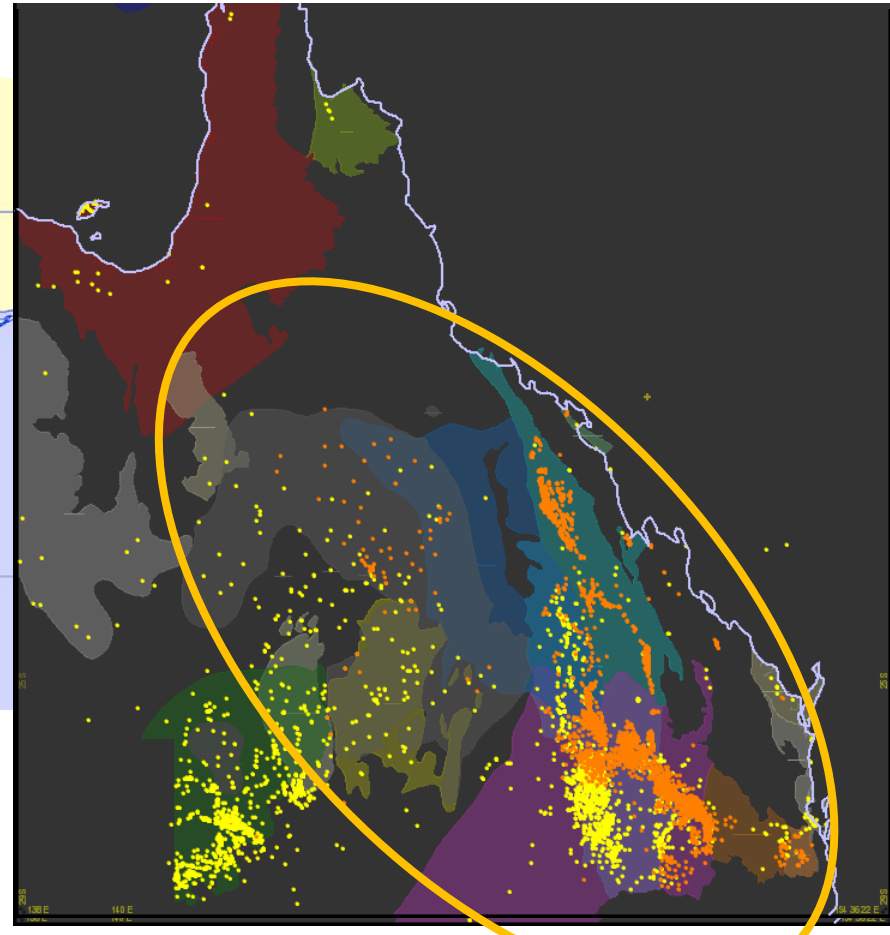
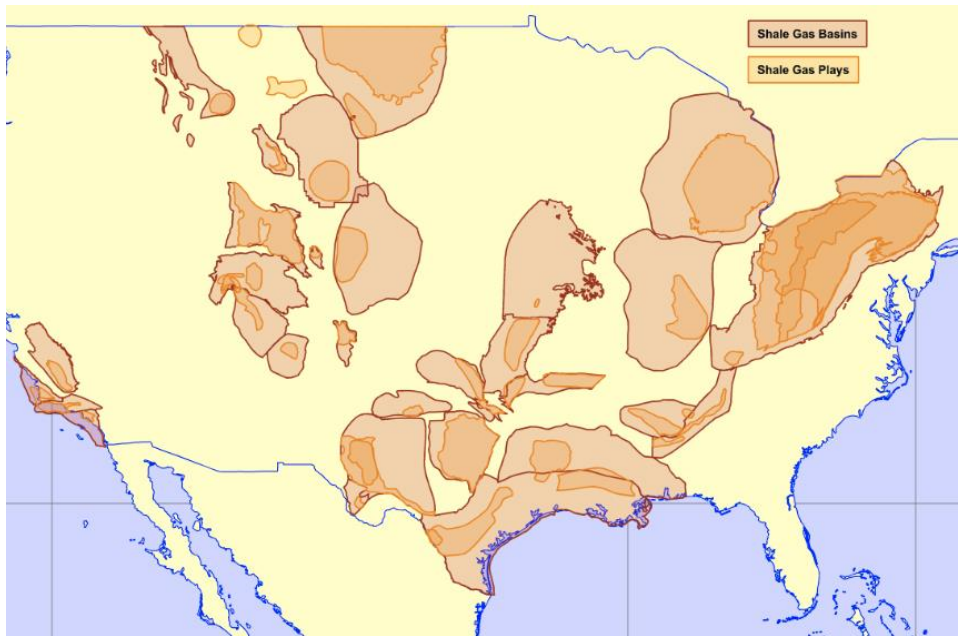
CASING AND TUBING				PLUGS			PERFORATIONS		
SIZE	SET AT	SX. CMT.	FORMATION	FROM	TO	SX. CMT.	FROM	TO	No. FT.
30"	539'	500		2,350'	1,850'	265'			
20"	950'	1200		650'	460'	148'			
13 3/8"	2137'	1950							

The Value of Unstructured Text to E&P

- Structured data management only addresses a tiny fraction of the wide range of knowledge acquired in petroleum E&P
- The investment in structured data management is biased towards the data that is relevant to current E&P processes
- ‘Uncommercial’ unstructured text has traditionally received second class treatment –
 - Geological environments that don’t promise traditional plays
 - Frontier areas with high geotechnical or economic risk

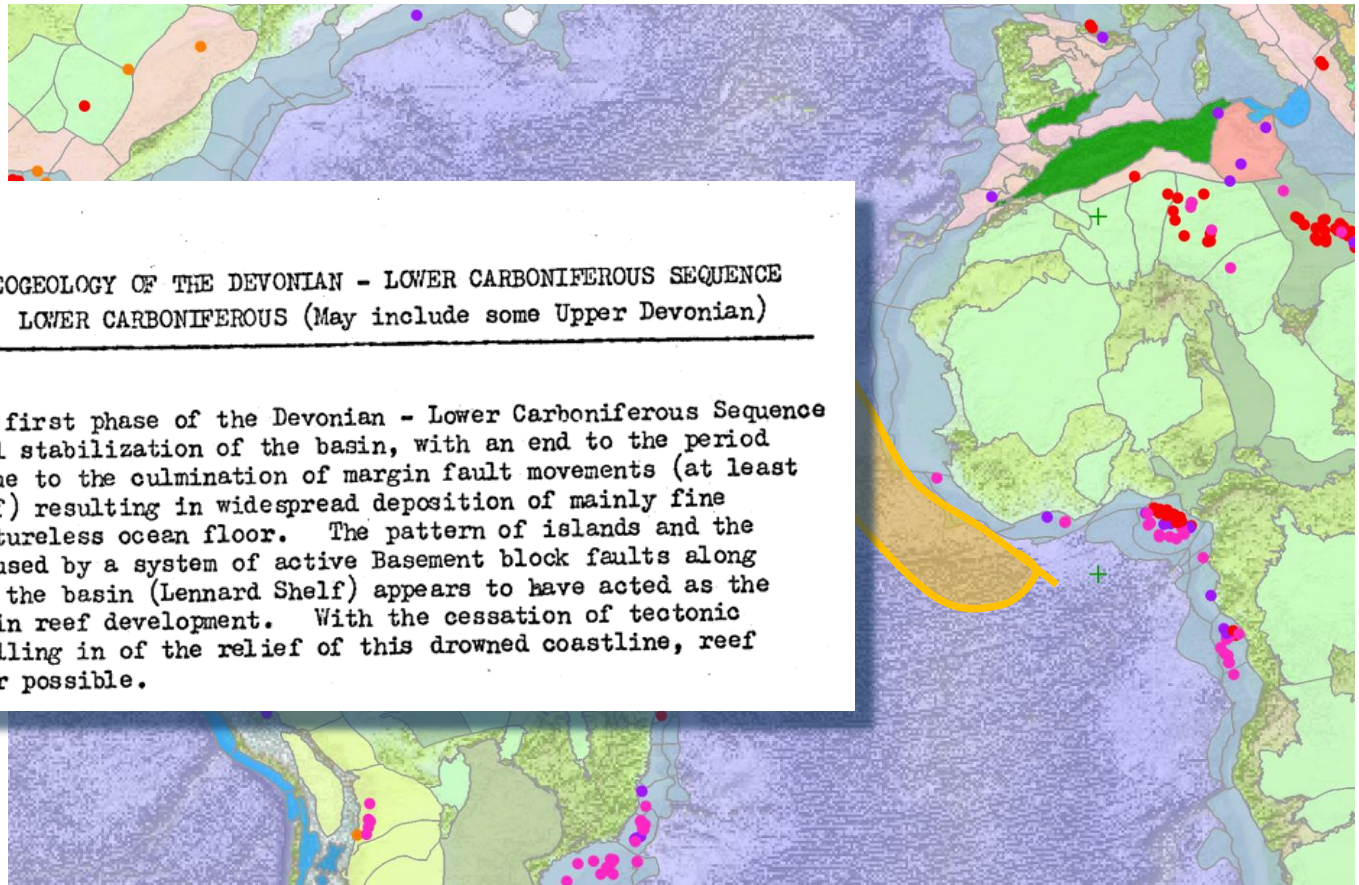
The Value of Unstructured Text to E&P

- Changes in E&P technology and energy economics encourage us to take a second look at previously rejected areas
- For example, development of oil & gas from coals and shales means that we need to revisit geotechnical studies from the 50's through the 70's.



The Value of Unstructured Text to E&P

- Combinations of better offshore engineering and evolving geopolitical circumstances revitalise our interest in frontier plays
- Historic reports take on new significance – discovering them becomes a challenge



Finding Knowledge In Unstructured Data

The mass of unstructured data accessible for E&P is a classical 'Big Data' issue. How do we help people find things in this?

- Library Classification Schemes
 - the Dewey Decimal (Library) System
 - More recent domain specific taxonomies
- Keyword / Subject Indexes and Thesauruses
 - Library of Congress catalogue
 - Australian AESIS Index
 - Anarchistic tags
- Natural Language Search Mechanisms
- Linking to Data Objects in a Structured Database

Classification Schemes – Library Coding Systems

Library classification systems played a key role in helping people find things in unstructured data collections, from their inception in the early 1800's until their gradual replacement with more elaborate electronic indexes at the present day -

- Dewey Decimal Classification
 - Based on

- Library of Congress

- ISBN number

Holmes Principles of physical geology /

by Holmes, Arthur, 1890-1965., Holmes, Doris L., 1899-

New York : Wiley, c1978.

Subjects ● Physical geology.

Call #: 203 H73p 1978

ISBN: 0470263369

Imprint: 730 p. : ill., maps ; 24 cm.

Edition: 3d ed. / rev. by Doris L. Holmes. --

Notes: Previous ed. published in 1965 under title: Principles of physical geology.
"A Halsted Press book."
Includes bibliographies and index.

OCLC No.: ocm03705998

Library Indexes and Taxonomies

- Most library indexing systems are based on classification schemes – taxonomies.
- The problem with defining a taxonomy is building classifications for future knowledge that has not yet been conceived.
 - For example, the Dewey Decimal Library Classification ...

Dewey Decimal Classification

The Ten Main Classes

The Hundred Divisions

The Thousand Sections

000 C
100 P
200 R
300 S
400 L
500 S
600 T
700 A
800 L
900 H

000 Comput
010 Bibliogr
020 Library
030 Encyclo
040 [Unassi
050 Magazi
060 Associa
070 News m
080 Quotati
090 Manusc

100 Philoso
110 Metaph
120 Epistem
130 Parapsy
140 Philoso
150 Psychol
160 Logic
170 Ethics
180 Ancient
190 Modern

200 Religio
210 Philoso

000 Computer science, information & general works
001 Knowledge
002 The book
003 Systems
004 Data processing & computer science
005 Computer programming, programs & data
006 Special computer methods
007 [Unassigned]
008 [Unassigned]
009 [Unassigned]

010 Bibliography
011 Bibliographies

012
013
014
015
016
017
018
019

020 Library & information sciences
021 Library relationships
022 Administration of physical plant
023 Personnel management

050 General serial publications
051 Serials in American English
052 Serials in English
053 Serials in other Germanic languages
054 Serials in French, Occitan & Catalan
055 In Italian, Romanian & related languages
056 Serials in Spanish & Portuguese
057 Serials in Slavic languages
058 Serials in Scandinavian languages
059 Serials in other languages

060 General organizations & museum science
061 Organizations in North America

... was developed in the 19th century essentially by one man and was built on a top-down approach to classify all human knowledge, which makes it difficult to adapt to changing fields of knowledge. (Wikipedia)

070 News media, journalism & publishing
071 Newspapers in North America
072 Newspapers in British Isles; in England
073 Newspapers in central Europe; in Germany

Library Indexes and Taxonomies

- Most library indexing systems are based on classification schemes – taxonomies.
- The problem with defining a taxonomy is building classifications for future knowledge that has not yet been conceived.
 - For example, the Dewey Decimal Library Classification ... *was developed in the 19th century essentially by one man and was built on a top-down approach to classify all human knowledge, which makes it difficult to adapt to changing fields of knowledge.* (Wikipedia)
- Present day document indexing systems still rely largely on using taxonomies.
 - Taxonomies remain mostly time dependent. We must not lose sight of potential losses of knowledge value as we apply new taxonomies to old reports with a severe mismatch in context.

Keyword Indexes, Tags and Thesauruses

- Thesauruses and subject indexes became popular tools for librarians in the second half of the 20th century –
 - USA:
 - ... Petroleum Abstracts at Tulsa University
 - ... Library of Congress Subject Catalogue
 - Australia:
 - ... AGI Glossary of Geology
 - ... EJC Thesaurus of Engineering Terms
 - ... AMF Australian Earth Sciences Information System (AESIS)¹
- In a document collection, a subject index might be the ‘alphabetical index’ that you can search for keywords or phrases
- The creation of a thesaurus for a given discipline is a significant investment in professional time, and represents a knowledge management for that discipline
- In present day unstructured data management, the use of ‘metadata tags’ to unstructured data objects is becoming popular
 - Unfortunately, few people appreciate the value of applying a rigorous subject index analysis before they allow creation of tags

LIBRARY OF CONGRESS CLASSIFICATION OUTLINE

CLASS Q - SCIENCE

Q - 390
Q1-295
Q300-390
Q350-390

Subclass QA

QA1-939
QA1-43
QA47-59
QA61-69
QA75-76.95
QA76.75-76.765
QA76.75-76.765
QA150-272.5
QA273-280
QA299.6-433
QA440-699
QA801-939

General
Cybernetics

machines
Computers. Computer science
computer software
Elementary mathematics. Algebra
Probabilities. Mathematical statistics
Analysis
Geometry. Trigonometry. Topology
Analytic mechanics

ANARCHISTIC TAGS!

Geographic Tags and Indexes

- Most E&P data deserves to have a geographic reference.
- Geotechs and IT people like to use numeric spatial indexes. The geographic 'context' of a knowledge element is not always tied to specific coordinates –
 - Regions are often defined by context, eg. Political vs. climate vs. tectonic
 - Basin identities and outlines are interpreted
 - Field boundaries may be split by commercial and political considerations
- Our industry has invested in the creation of a range of geographical dictionaries and data sets, which are mostly proprietary, to enable consistent geographic tagging of sometimes interpretive data:
 - IHS geographic dictionary
 - Fugro
 - Government agencies, eg. MMS (USA), ANH (Colombia), GA (Australia), ...
- Geographic tags and indexes remain a high value addition to the cataloguing of unstructured E&P data.

Natural Language Search Mechanisms

- Natural language search – ‘like Google’ – is hugely successful in helping people find things to get through their daily lives.
- Effective natural language searches are very context dependent. The success that we attribute to Google is because it is taking into account a huge range of factors that reflect a constantly changing ‘knowledge’ rather than the current ‘truth’:
 - Popular search queries and their most frequently accepted results
 - Location and other personal information about the questioner
 - Paid search preferences
 - Search criteria based on keyword and other indexes
- Natural language search may find a lot of data that is relevant, however...
 - It may also display a lot of noise
 - It will not guarantee that the searcher will see the most important data
 - Present technologies are anglo-centric, a lot of E&P literature is not in English.

Subject Indexes, Tags and Language

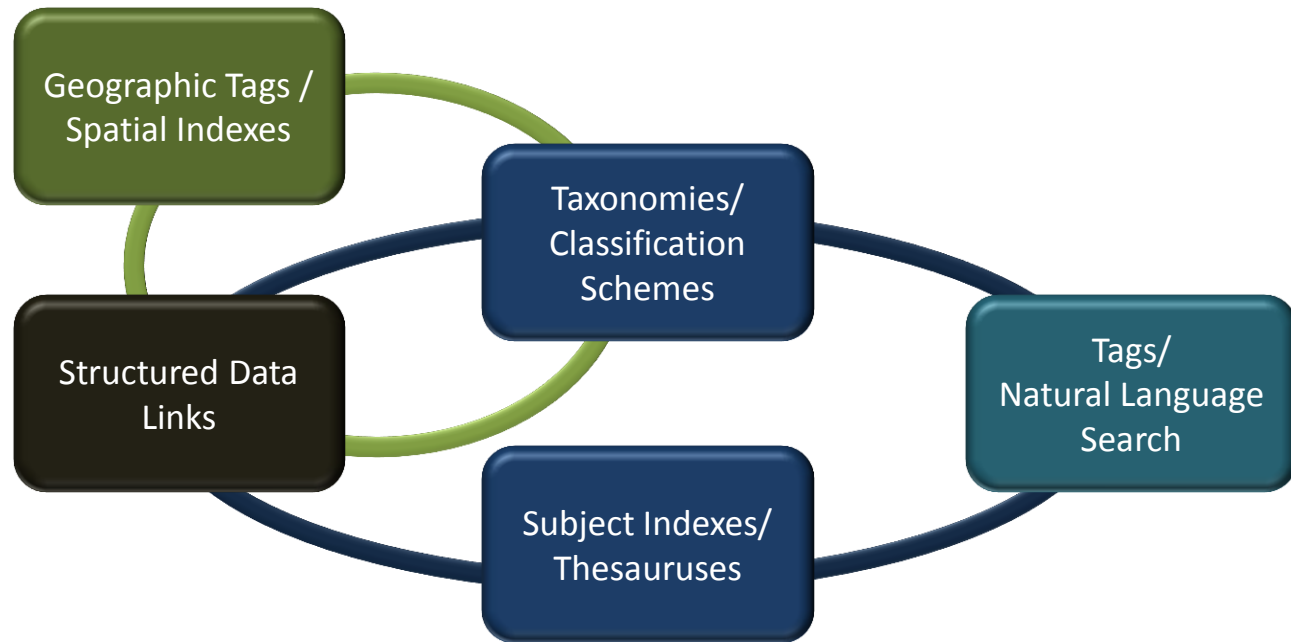
- Subject indexes, search algorithms and taxonomies have evolved with a strong bias to the English language.
- Practical application in today's industry, particularly in Latin America and Asia requires a multi-lingual approach:
 - English – TOP WELL COMPLETION REPORT
 - Spanish
 - Mexico – CIMA CARPETA DE POZOS
 - Venezuela – TOPE CARPETA DE POZOS
 - Argentina – TOPE LEGAJO
 - Chinese, Malaysian, Vietnamese, Indonesian ...

Linking Unstructured to Structured Data

- Increasingly, we have structured master data for at least the key components of our operations –
 - Well master
 - Seismic survey master
 - Production entities – fields / blocks
- The PPDM Records Management (RM) module is an effective industry standard way of storing such links in a database.
- Tools for creating the links are the software industry's challenge –
 - For new data, good data governance can promote reliable linking
 - For old data, creation of links is largely dependent on manual processing using the other tools, from natural language searches through library classifications.

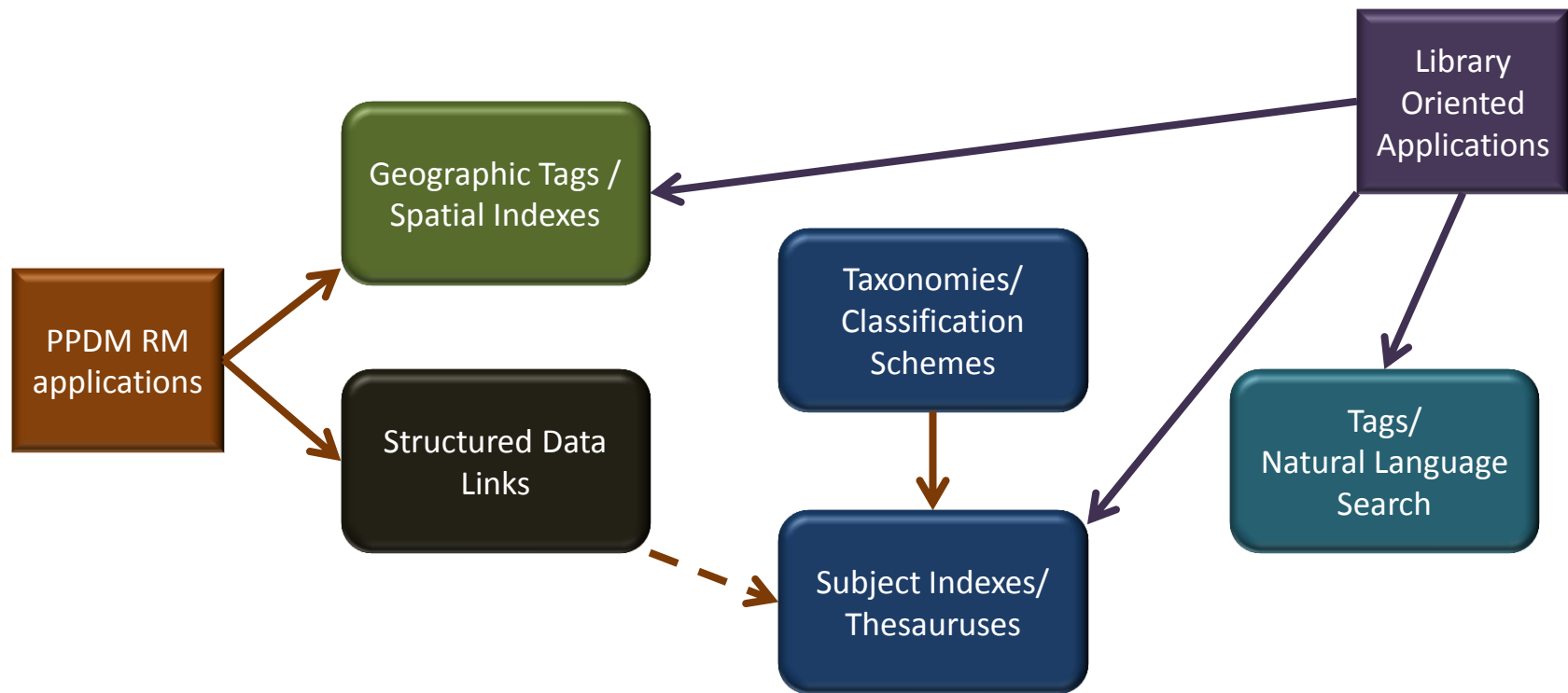
Finding Knowledge In Unstructured Data

Ideally, our unstructured data management needs to leverage all available technologies.

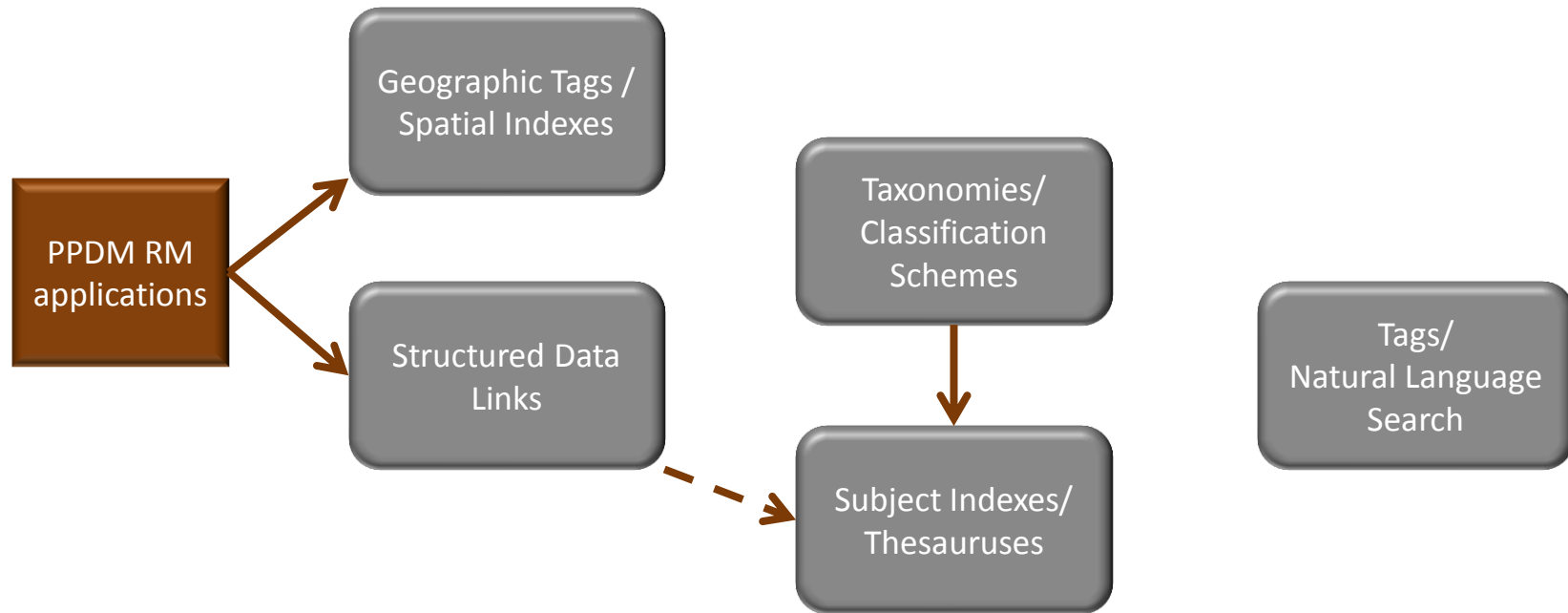


Finding Knowledge In Unstructured Data

Different applications do this in different ways, but at the end of the day they all rely on taxonomies in one way or another.



PPDM RM application example – dbMap/web



dbMap/web – A PPDM/RM approach

The dbMap/web unstructured data management system follows the PPDM RM model, where taxonomies are applied within the constraint of predefined tables such as R_ITEM_CATEGORY

The screenshot displays the dbMap/web interface with a 'Reference table lookup (PPDM Reference : R_ITEM_CATEGORY)' dialog box open. The background interface includes a top navigation bar with seven tabs: 1 Item media, 2 Storage, 3 Well Assets, 5 Media content, 6 Originator, and 7 Description. The 'General' section of the form is visible, showing fields for mediaType (PAPER), Location ref (1000735), Circulation allowed (checked), Physical Item Id (1000735), Physical Item Status (AVAILABLE), and Label. The 'Item Description' and 'Taxonomy' sections are also partially visible. The dialog box contains a table with the following data:

	Type	Id	Name
1			
2	Well	BIOSTRAT	Biostratigraphy
3	Well	BOREHOLE SEISMIC	Borehole Seismic
4	Well	CUTTINGS	Cuttings
5	Well	FIS	Fluid Inclusion Stratigraphy
6	Well	FLUID ANALYSIS	Fluid Analysis
7	Well	GENERAL	General
8	Well	GEOCHRONOLOGY	Geochronology
9	Well	LWD_MWD	Log/Measure While Drilling
10	Well	MUDLOG	Mudlogging
11	Well	ORGANIC GEOCHEM	Organic Geochemistry
12	Well	PETROGRAPHY	Petrography

The dialog box also includes a 'Cancel' button and a 'View 1 - 25 of 25' indicator.

dbMap/web – A PPDM/RM approach

Each level of taxonomy in this part of the system has to be allocated it's own unique table in the PPDM schema – category vs. sub-category, for example:

The screenshot shows a web browser window titled "dbMap/Web - YDSH - Mozilla Firefox". The address bar shows "Go to a Web Site". The main content area has a navigation bar with six tabs: "1 Item media", "2 Storage", "3 Well Assets", "5 Media content", and "6 On". The "1 Item media" tab is selected and highlighted in orange. Below the navigation bar, there is a "General" section with a table of fields:

General			
mediaType	PAPER	Location ref	1000735
Physical Item Id	1000735	Created on	
Physical Item Status	AVAILABLE	Catalogue date	11-MAR-2013
Label		Catalogue name	SUSANH
		Circulation allowed	
		Circulation out	
		Original	
		Preferred	
		Sale allowed	
		Certified copy	

Below the "General" section is an "Item Description" section with a text area for "Item Description". Below that is a "Taxonomy" section with two dropdown menus: "Category" (set to "Fluid Analysis") and "Sub category". At the bottom right, there is a link "Color Index - Data Entry" and three buttons: "Previous", "Next", and "Finish".

dbMap/web – A PPDM/RM approach

Addition of arbitrary additional levels of subject indices are possible using the 'Descriptions' on an item ... what would the ideal relationship be between the intent of 'Descriptions' and the keyword indices of library thesauruses?

The screenshot shows the dbMap/web application interface in a Mozilla Firefox browser window. The interface has a top navigation bar with seven tabs: 1 Item media, 2 Storage, 3 Well Assets, 5 Media content, 6 Originator, and 7 Description. The '7 Description' tab is currently selected. Below the navigation bar, there is a section titled 'Add appropriate Descriptions:' with input fields for 'Name', 'Description', and 'Description date' (set to 11-MAR-2013). Below this is a 'Descriptions list:' table with columns 'Descriptions' and 'Type'. A 'Reference table lookup (PPDM Reference : R_RM_DESC_TYPE)' dialog box is open, displaying a list of description types: 1, 2 BASE DEPTH, 3 COMMENT, 4 ENCLOSURE, 5 GENERAL, 6 MAIN CLASS, 7 OTHER, 8 PRODUCT TYPE, and 9 TOP DEPTH. The dialog box has a 'Cancel' button at the bottom right. The browser's address bar shows the URL: psad19:7016/webdbmap_rm_demo/getPane2?block=INDEXING_WORKFLOW&display=wizard&mode=add&id=-1#.

Description type
1
2 BASE DEPTH
3 COMMENT
4 ENCLOSURE
5 GENERAL
6 MAIN CLASS
7 OTHER
8 PRODUCT TYPE
9 TOP DEPTH

dbMap/web – A PPDM/RM approach

Below links from Core Analysis, Wells & Fields/Pools (master objects) help the user find the details for a wide range of structured & unstructured analysis.

The screenshot displays the dbMap/web application interface. The top navigation bar includes links for Wells, RM, Analysis, Culture, Fields/Pools, Facilities, Legal, and Reference Tables. A search bar is located below the navigation bar. The main content area shows an analysis report for a wellbore, with a table of analysis results. The table has columns for Analysis ID, Analysis source, Sample date, Analysis date, Analysis purpose, Analysis quality, and Study type. The table contains three rows of data. To the right of the table, there is a sidebar with metadata and links. The sidebar includes sections for Analysis Report/Group, Metadata, Dates, Sample location, and Depths. At the bottom of the page, there are links to other objects, including Wells (via Origin), Links (Anl Component), and RM Items.

	Analysis ID	Analysis source	Sample date	Analysis date	Analysis purpose	Analysis quality	Study type
1	✓ V...	CARIN			Analysis of ... core and core plug sam	GOOD	PETROLOGY
2	✓ X...	XRAY Min Serv Ltd			XRD of whole rock and clay fraction for various	GOOD	GEOCHEMISTRY_XRD
3	✓ XRF_ABC...	...			XRF of 7 selected samples, including Major &	GOOD-FAIR	GEOCHEMISTRY_XRF

Analysis Report/Group Metadata

Analysis ID: ... Analysis source: CARIN

Analysis purpose: Analysis of ... core and core plug samples

Analysis quality: GOOD Study type: PETROLOGY

Dates: Analysis- and Report- related milestones

Sampling	Analysis	Reported	Received	Start	End

Sample location: WELLBORE Stratigraphic name set

Depths: Top + Base + TVD + Stratigraphy

Type	Depth	OUOM	Strat Unit ID
Top	1,992.22 metre		
Base	2,311 metre		
TVD			

Created by PETROSYS on 30-JUL-2014. Last updated by PETROSYS on 30-JUL-2014.

Edit New Save as... Delete

View 1 - 3 of 3

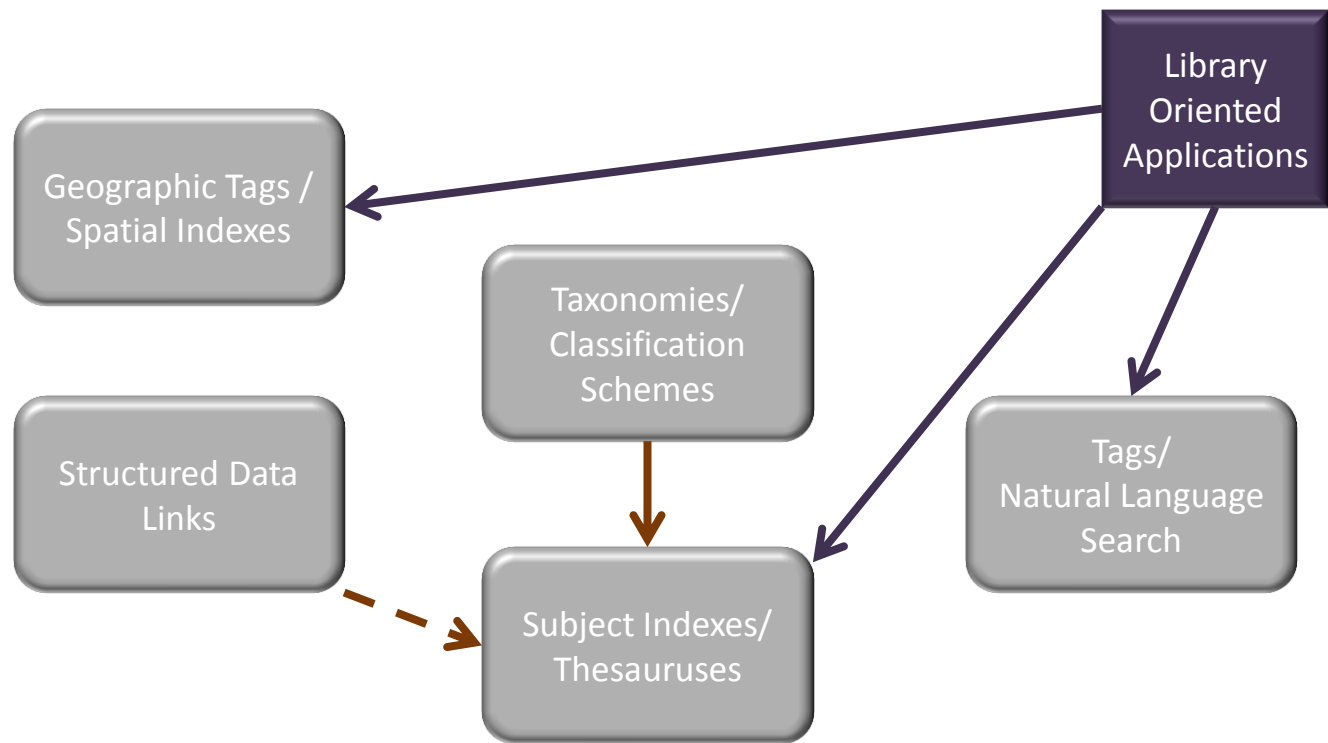
Analysis report data: results, companies, aliases, remarks, and individual sample results

Analysis detail (incl. XRD)	XRD Spreadsheet (Clays)	XRD Spreadsheet (Whole Rock)	Elemental analysis	Elemental detail (incl. XRF)	Isotopic analysis
Oil analysis	Gas analysis	Water analysis	Analysis company	Analysis report alias	Analysis remark
Analysis step/part	Analysis samples				

Links to other objects

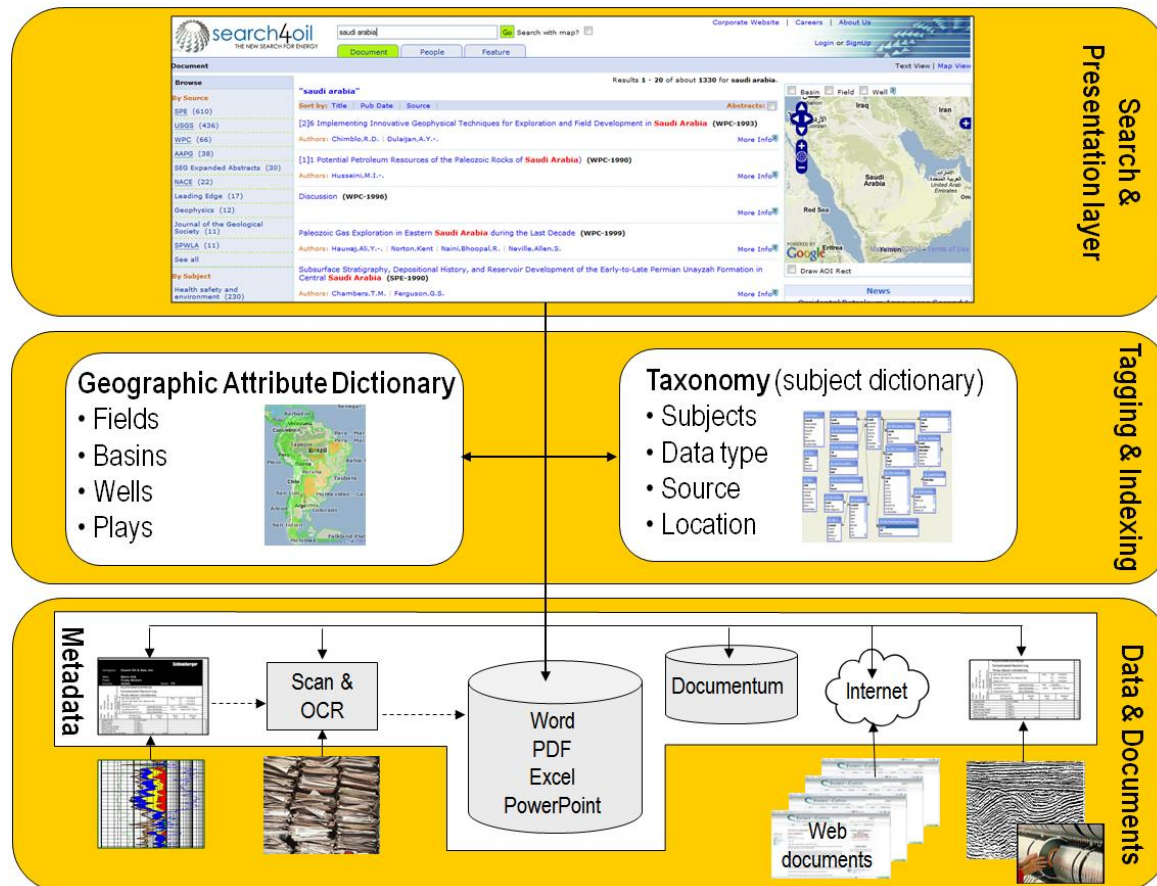
Wells (via Origin)	Links (Anl Component) 1	RM Items 1

Digital Library Approach – Search4Oil



Search4Oil – A more library centric approach

A composite approach to indexing of unstructured information is used by the Search4Oil web application -



Geographic Tags in Search4Oil

Publicly available geographic tagging has been used in Search4Oil. More sophisticated geographic tags, such as IHS, are not available to the public as they are proprietary.

Publicly available geographic tags –

Political Boundaries

... to municipality level

Basin Outlines ...

Field Outlines for 18 thousand fields

Well Locations

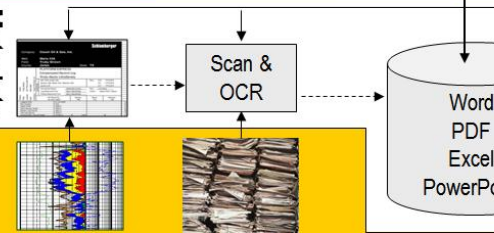


Geographic Attribute Dictionary

- Fields
- Basins
- Wells
- Plays



Metadata



Geographic and Free Text Searches

☒ Search with map?[Careers](#) | [About Us](#)[Corporate Website](#)[Login or SignUp](#)[Document](#)[People](#)[Feature](#)

Document

[Text View](#) | [Map View](#)

Browse

By Source

[SPE](#) (45)[AAPG](#) (10)[Sedimentary Geology](#) (2)[USGS](#) (2)[Applied Geochemistry](#) (1)[Chemie der Erde -
Geochemistry -
Interdisciplinary Journal for
Chemical Problems of the
Geosciences](#) (1)[Journal of Geochemical
Exploration](#) (1)[Leading Edge](#) (1)Results **1 - 20** of about **65** for **hydrocarbon systems** in zoom:4.

"hydrocarbon systems"

Sort by: [Title](#) | [Pub Date](#) | [Source](#) |**Abstracts:** ☐

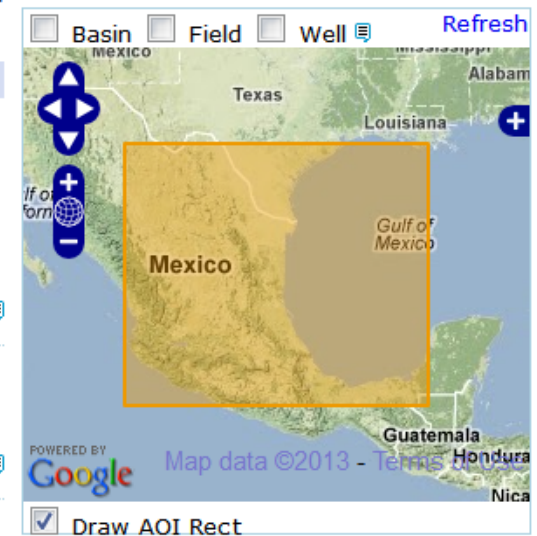
Chapter 5. Assessment of undiscovered conventional oil and gas resources-Lower Cretaceous Travis Peak and Hosston formations, Jurassic Smackover interior salt basins total petroleum **system**, in the East Texas basin and Louisiana-Mississippi salt basins (**USGS-2005**)

Authors: [Dyman,T.S.](#) | [Condon,Steven](#)[More Info](#)

Temperature Dependence of L1/L2/V Behavior in CO2/**Hydrocarbon Systems** (**SPE-1988**)

Authors: [Orr,F.M.](#) | [Silva,M.K.](#) | [Larson,L.L.](#) | [Taylor,M.A.](#)[More Info](#)

Geochemical and isotopic approach to maturity/source/mixing



E&P Specific Keyword Indexes

Journal of Geochemical Exploration (1)

Palaeogeography, Palaeoclimatology, Palaeoecology (1)

By Subject

Hydrocarbon Systems and Basin Analysis (11)

Health safety and environment (8)

Geochemistry (6)

Sedimentology and Stratigraphy (4)

Paleontology (3)

Numerical simulation (2)

Tectonics (2)

Pipelines (1)

Reservoir characterization (1)

Reservoir engineering (1)

By Basin

Kerr (23)

Midland - Permian (23)

Orogrande (23)

Pedregosa - Chihuahua (23)

Authors: Gurgey,Kadir | Clayton,Chris | Emiroglu,Nasim | Philip,Faust, Jr.

[More Info](#)

The sequence stratigraphy, sedimentology, and economic importance of evaporite-carbonate transitions: a review **(Sedimentary Geology-2001)**

[More Info](#)

Fracturing Microseismic Response in Turbidite Reservoirs in Tajin Field **(SPE-2003)**

Authors: Kaiser, Penelope | Rodriguez, Juan, M. | Gachuz, Heron | Swaan, Abraham

[More Info](#)

Petroleum **Systems** of the Permian Basin, USA **(AAPG-2003)**

Authors: Claxton, Brenda, L. | Burgess, Jack, D. | Hill, Ronald | Tobey, Mark, H.

[More Info](#)

Polycyclic aromatic **hydrocarbons** as evidence of **hydrocarbon** migration in marine and lagoon sediments of a recent rift zone (Skjalfandi and xarfjurr), Iceland **(Chemie der Erde - Geochemistry - Interdisciplinary Journal for Chemical Problems of the Geosciences-2006)**

Authors: Geptner, Alfred, R. | Chernyansky, Sergey, S. | Pikovskii, Yurii, I. | Richter, Bjarni

[More Info](#)

Case Studies on New Methods of Technology Transfer and Development **(SPE-2000)**

Authors: Kiker, Robert, D.

[More Info](#)

Phase Behavior, Fluid Properties, and Displacement Characteristics of Permian Basin Reservoir Fluid/CO2 **Systems** **(SPE-1992)**

Authors: Creek, J.L. | Sheffield, J.M.

[More Info](#)

Permeability and Relative Permeability Measurements at Reservoir Conditions for CO2-Water **Systems** in Ultra Low Permeability Confining Caprocks **(SPE-2006)**

Native Language Search Scores

Document

[External Link to the paper](#)



Integrated Geostatistical Reservoir Characterization of Turbidite Sandstone Deposits in Chicontepec Basin, Gulf of Mexico

Related Features

Hydrocarbon Systems and Basin Analysis Petrophysics **Tampico** Mexico
Reservoir characterization

What is this?

Authors:

[Yamamoto, Hiroshi](#), [Sandria, Francisco](#) [Murguia](#), [Abbaszadeh, Maghsood](#),
[Yazawa, Nintoku](#), [Guerrero, Fernando](#) [Rodriguez Garza](#), [Shimamoto, T.](#),
[Takano, Osamu](#), [Zamora, David, H.](#)

Serial:

SPE Annual Technical Conference and Exhibition, 5-8 October
2003, Denver, Colorado

Basin:

[Tampico](#)

Countries:

[Mexico](#)

Subjects:

Reservoir characterization(45%), Hydrocarbon Systems and Basin
Analysis(36%), Petrophysics(34%)

Summary

- There is a huge wealth of information stored as unstructured text
- The discovery of new play types and exploitation technologies has given new value to previously unheeded reports and data sets
- Data managers need to provide the best ways of navigating through increasingly noisy unstructured data collections ... and to help clients reflect on the importance of improving the signal to noise ratio of their text content
- A combination of structured master data links, peer reviewed taxonomies and subject index thesauruses, E&P aware geographic tags, and appropriately applied natural language search is likely to provide the best outcome.

Presenter: Scott Tidemann, Vice President Petrosys Middle East & Asia Pacific

Email: scott@petrosys.com.au