

Implementation of Machine Learning Systems to Enhance the Value of the CDA North Sea Data Set*

Philip Neri¹

Search and Discovery Article #42132 (2017)**

Posted September 18, 2017

*Adapted from oral presentation given at 2017 AAPG Geoscience Technology Workshop, Big Data & Deep Learning in the Oil Industry: Basic Applications, Houston, Texas, May 22, 2017

**Datapages © 2017 Serial rights given by author. For all other rights contact author directly.

¹AgileDD (pneri@me.com)

Abstract

The CDA maintains a collection of well and seismic data submitted by the UKCS operators since the early days of the North Sea Exploration and Production in the 1960's. The collection of CDA well data has been made available to operators and authorities as a database of 11,500 well headers and as a set of 450,000 documents under various formats such as .pdf, .xls, .doc, .tiff, .jpg, .las, .dlis.

This collection of data is similar in its organization and content with legacy datasets that can be found in any industry: around 20% of the information is available in a structured form such as a relational database, and 80% in a semi-structured or unstructured form, typically grouped in folders containing various documents formatted as described above. Since most of the software and data management tools used in E&P can only access the structured information and in some cases some half-structured formats, it transpires that E&P decisions are based on a small part of the available stored information.

The low benchmark of 20% of available data is due to several factors, primarily the cost of indexing (classifying the documents per topic) and cataloguing the documents (extracting metadata from the document) which is currently a work-intensive process. But the cost is not the only limitation. The fixed nature of most of the subsurface data-models makes it almost impossible to catalog information which was not planned to be extracted in the initial stage of the data model design.

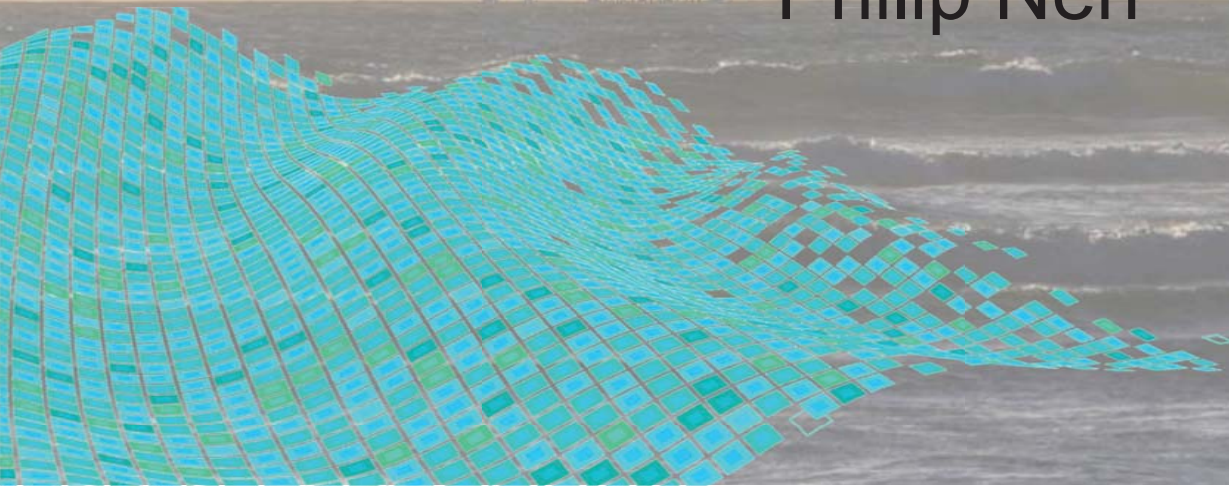
In 2016, the CDA launched a challenge to find new ways to extract value from its unstructured data assets. This article explores the application of newly developed Machine Learning Systems (MLS) to automate part of the indexing and cataloguing. MLS demonstrated a reduced time (and therefore cost) of access to information but also enriched the extracted information by qualifying its extraction confidence and source, and identifying replicates. They make it possible to perform data analysis of larger datasets in term of volume and variety.

The performance of Machine Learning Systems when applied to subsurface data management will be discussed, the limitation criteria listed, and some future possibilities to overcome the current limitations will be overviewed.

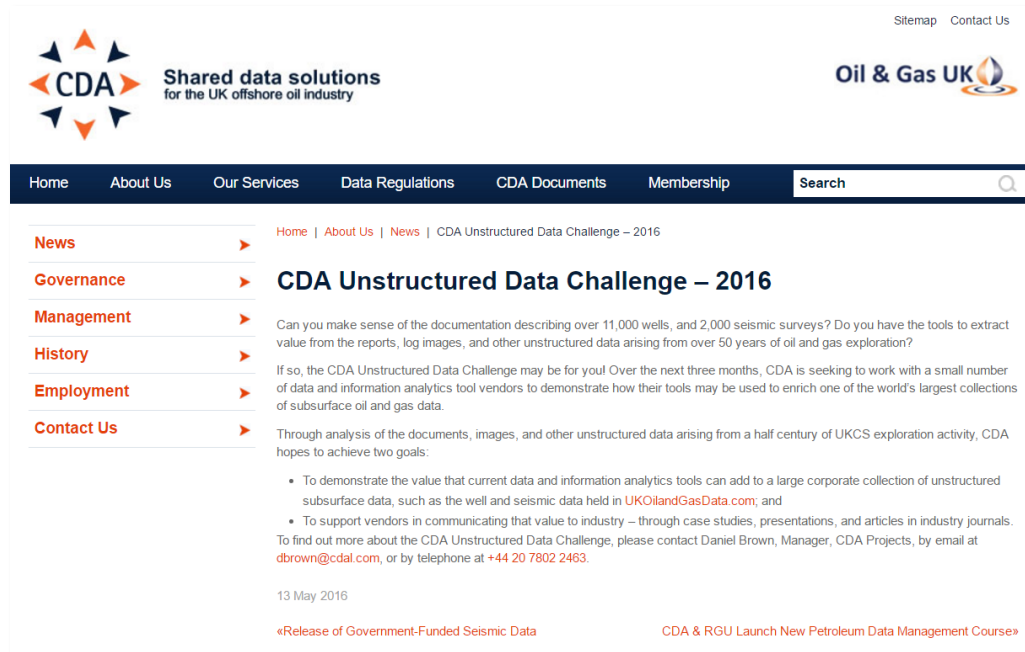


Implementation of **machine learning** systems to
enhance the value of the CDA North Sea data set

Philip Neri



The CDA unstructured data challenge



Presenter's notes: In summer 2016 AgileDD, together with 7 other technology providers, was selected to participate in the first CDA unstructured data challenge. The idea was to illustrate how new technology could “enrich one of the world largest collections of subsurface O&G data.

This presentation will show you what we have achieved in a short period of 4 weeks in August 2016.

The CDA unstructured data challenge

The screenshot shows the CDA website with the following elements:

- Header:** CDA logo (a star-like shape with 'CDA' in the center) and the text 'Shared data solutions for the UK offshore oil industry'. To the right is the 'Oil & Gas UK' logo. Further right are links for 'Sitemap' and 'Contact Us'.
- Navigation Bar:** A dark grey bar with links: 'Home', 'About Us', 'Our Services', 'Data Regulations', 'CDA Documents', 'Membership', and a search box labeled 'Search'.
- Main Content Area:**
 - A list of bullet points:
 - To demonstrate the value that current data and information analytics tools can add to a large corporate collection of unstructured subsurface data, such as the well and seismic data held in UKOilandGasData.com; and
 - To support vendors in communicating that value to industry – through case studies, presentations, and articles in industry journals.
 - A sidebar on the left with two links: 'Employment' and 'Contact Us', each with a right-pointing arrow.
 - Main text below the bullet points:

of data and information analytics tool vendors to demonstrate how their tools may be used to enrich one of the world's largest collections of subsurface oil and gas data.

Through analysis of the documents, images, and other unstructured data arising from a half century of UKCS exploration activity, CDA hopes to achieve two goals:

 - To demonstrate the value that current data and information analytics tools can add to a large corporate collection of unstructured subsurface data, such as the well and seismic data held in UKOilandGasData.com; and
 - To support vendors in communicating that value to industry – through case studies, presentations, and articles in industry journals.

To find out more about the CDA Unstructured Data Challenge, please contact Daniel Brown, Manager, CDA Projects, by email at dbrown@cda.com, or by telephone at +44 20 7802 2463.
 - Footer area with the date '13 May 2016' and two links: '«Release of Government-Funded Seismic Data' and 'CDA & RGU Launch New Petroleum Data Management Course»'.

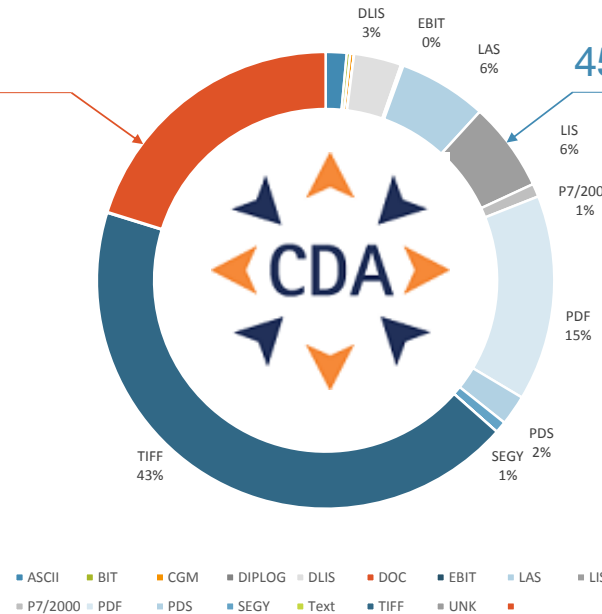
Presenter's notes: In summer 2016 AgileDD, together with 7 other technology providers, was selected to participate in the first CDA unstructured data challenge. The idea was to illustrate how new technology could “enrich one of the world largest collections of subsurface O&G data.

This presentation will show you what we have achieved in a short period of 4 weeks in August 2016.

The CDA well data set

2 relational tables

- Well header
- CS8 file index



450,000 unstructured files

- End of well reports
- Wirelines, MWD
- Composites
- CCA and SCAL
- Geochemistry
- Fluid analysis
- Biostratigraphy
- Lithostratigraphy
- Cementation report
- Checkshots and VSP
- ...etc...

Presenter's notes: CDA is a significant and vitally important component of the private/public partnership that constitutes the NDR (National Data Repository) for the UKCS

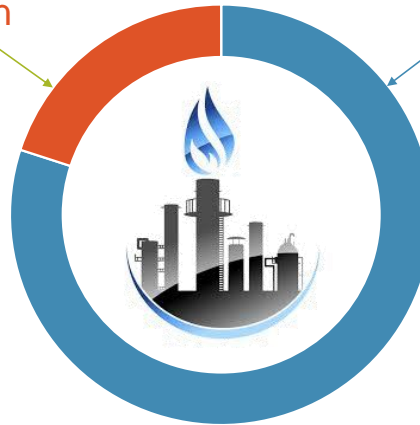
The CDA has collected information about 11,500+ offshore wells drilled over the last 60 years

This information is under the form of
450,000 unstructured files

The 80/20 ratio

20% structured information

- Easy to access
- Easy to query
- Easy to QC
- Easy to model and analyze



80% unstructured information

- Costly to index
- Costly to extract metadata
- Difficult to access and query
- Cannot feed analytical tools

Presenter's notes: The CDA data set is not so different from many others we can find in our industry and in some other industries. According to various sources (Merryl Lynch, EMC, Oracle ...) 80 % of the information is available in an unstructured format (PDF, TIFF ...)

This type of information is difficult to use, it cannot be used directly by the modeling of BI tools. Extracting metadata to create indexes and populate DBs is extremely costly and need SMEs that today are increasingly scarce in our industry.

This translates into the fact that only 20% of the information is available within structured database.

The big problem is that since this small part is easy to access and query, it tends to be the unique source of information to base decision

Your decision is at risk

20% structured information

- Easy to access
- Easy to query
- Easy to QC
- Easy to analyze



80% unstructured information

- Costly to index
- Costly to extract metadata
- Difficult to access and query
- Cannot feed analytical tools

Presenter's notes: And that makes your decision a very risky one.

What if we reverse the ratio ?

80% structured information

- Easy to access
- Easy to query
- Easy to QC
- Easy to analyze
- Easy to source
- With a known confidence level



20% unstructured information

- Automatic cataloging and indexing

Make your decision more reliable based on more structured data

Presenter's notes: Our ambition is to reverse the ratio. Using an automated process based on a machine learning system, we extract more information from your documents at a fraction of the cost and time, and this makes your decisions more reliable.

The advantages of automated cataloging

**SAVE
MONEY**

Avoid populating
databases manually

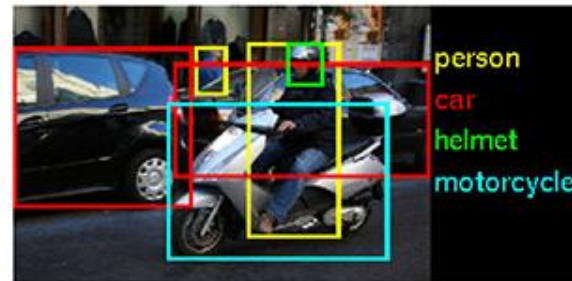
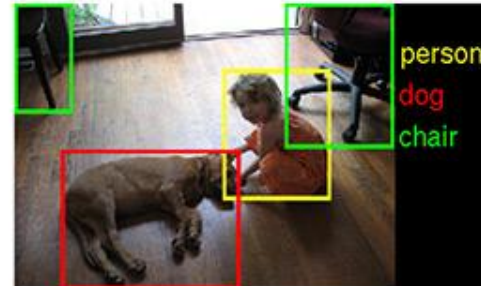
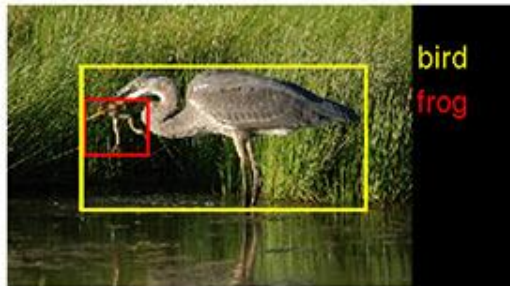
GO FASTER

From data to decision

**DE-
RISK**

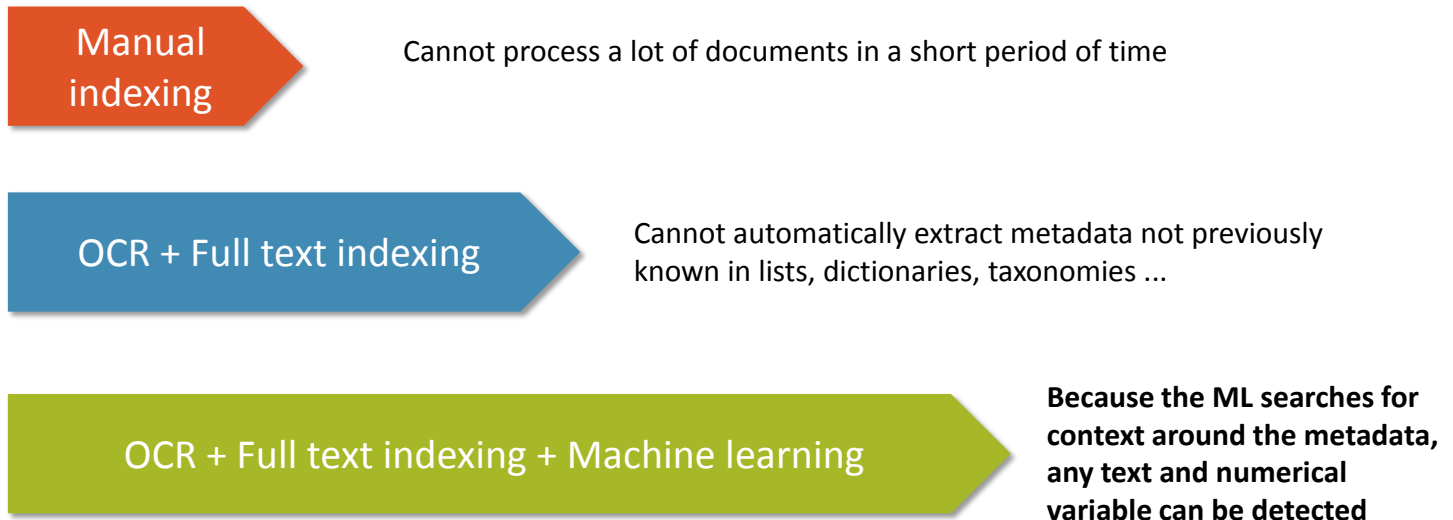
Using more verified
information

Why use a machine learning system?



Presenter's notes: Supervised Machine Learning excels at recognizing a pattern in an unstructured context

ML exceeds manual and full text indexing



10

Presenter's notes: The capacity to detect the pattern around a target metadata item allows us to make a more efficient cataloging compared to a full text indexing. ML detects a numerical variable such as a coordinate, a depth, a temperature using the context of the value which is not possible even with the best Full text indexing

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

The screenshot displays a software interface for document management and analysis. On the left, a tree view shows a collection of documents, including '208/11-1' with a confidence score of 100. The middle pane shows the details for '208/11-1', including document type, category, and country. The right pane shows a 'PETROPHYSICAL ANALYSIS' report for the same well, detailing various parameters such as depth, temperature, and lithology.

Document Details (208/11-1):

- Well Name: 208/11-1
- Document Category: REPORT IMAGE/GEOL_PP/PHYS
- Document Title: PETROPHYSICAL ANALYSIS
- Country-Region: UKCS
- Document Language: English
- Original Document Resolution: 300
- Original image size: 300
- Number of pages: 30

PETROPHYSICAL ANALYSIS 208/11-1

DEPTH SCALE: 1:500 MD

PROJECT: UK_FO_WOS COUNTRY: UKCS WELL CLASSIFICATION: Exploration
LOCATION: West Shetland LICENCE: P1454 COMPLETION STATUS: Plugged & Abandoned

INTEREST: DONG E&P (UK) Ltd OPERATOR: DONG Energy E&P (UK) Ltd ANALYSED BY: NCGL APPROVED BY: XRAFR

MUD TYPE: Water based mud (WBM) DATUM FOR ELEVATION: MSL SPUD DATE: 23/9/2012
TOTAL DEPTH: 4295 MEASUREMENT REF: EP LOOKING COMPLETE: 15/1/2013
BOTTOM HOLE TEMP: 130DEG C ELEVATION MEAS. REF: 36.00 DATE PLOTTED: 30 May 2013
BOTTOM HOLE F.M.: Surface Fin. GROUND LEVEL ELEV.: -1106.6

REMARKS: Shale volume from Density Neutron logs. Porosity from logs by neutron corrected density porosity with matrix density at 2.65 g/cc. Water saturation from modified formation with logs at 2 and RT from neutron which is slightly optimistic compared to RT from LWD. Formation water salinity is very accurate. 22,000 ppm NaCl has been used. (20,000 ppm NaCl based on 4000 Ohm-cm @ 25°C). Temperature from bottom hole logs (Cragghead 208/11-1) 25°C @ 21.5m MD. No more logs can be seen in logs - it is estimated due to the neutron saturation. Colours in red of page (4-7). Possible correlation through water. Value on table is correlation? The BC is not expected to be one continuous accumulation. It is more likely to be isolated or dispersed in the water. Core might say result in the analysis that look like fine scale variations.

LITHOLOGY SYMBOLS

Water	Shale	Coal
Clay Bound Water	Sand	Carbonate
Clay - Carbonate - Oil - Shale	Clash	Volcanics
Clay	Contest	

OPERATIONAL SYMBOLS

Core Blockout	Core Loss	Shallow Core Blockout	Shallow Core Loss
Lost	Lost	Lost	Lost
Lost	Lost	Lost	Lost
Lost	Lost	Lost	Lost

Well Header Metadata Table:

Correlation	Depth/Depth	Porosity Logs	Resistivity	Sonic	Mudlog	Annotations	Porosity	Saturation	CPI

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its "container" such as a report, log, digital document ... and its "contents" such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

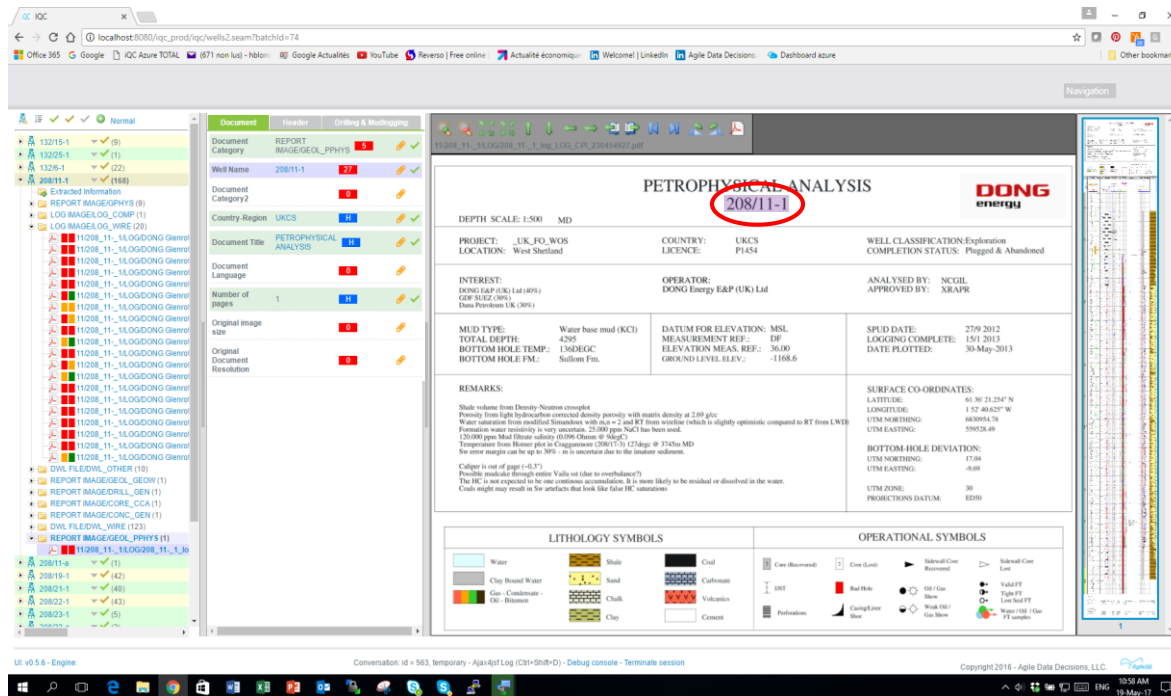
The screenshot displays a software interface for document management and analysis. On the left, a file explorer shows a hierarchy of documents, including '208/11-1' and 'LOG/208_11-1_log_LOG_CPI_236454927.pdf'. The central panel shows the document's metadata, including the title 'PETROPHYSICAL ANALYSIS 208/11-1' and the company 'DONG Energy'. The right-hand panel displays a detailed petrophysical analysis report, including fields for project, location, well classification, and various technical parameters. A legend at the bottom shows different symbols for well types and data series.

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata



Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

The screenshot displays a software interface for well data analysis. The top section shows a list of documents on the left and a detailed view of a specific well on the right. The well name is 208/11-1, highlighted with a red circle. The document category is REPORT IMAGEGEOL_PHYS. The well header metadata includes: Well Name: 208/11-1, Document Category: REPORT IMAGEGEOL_PHYS, Country/Region: UKCS, Document Title: PETROPHYSICAL ANALYSIS, PROJECT: _UK_PO_WOS, COUNTRY: UKCS, WELL CLASSIFICATION: Exploration, COMPLETION STATUS: Plugged & Abandoned, ANALYSED BY: NCGIL, APPROVED BY: XRAPR, SPUD DATE: 27/9/2012, LOGGING COMPLETE: 15/1/2013, DATE PLOTTED: 30-May-2013. The petrophysical analysis section includes: TOTAL DEPTH: 4295, BOTTOM HOLE TEMP.: 134DEGC, MEASUREMENT REF.: DF, BOTTOM HOLE FM.: Sullivan Fin., ELEVATION MEAS. REF.: 36.00, GROUND LEVEL ELEV.: -1168.6. The remarks section contains text about shale volume from Density-Neutron correction and water saturation. The bottom section shows LITHOLOGY SYMBOLS and OPERATIONAL SYMBOLS. The interface also includes a navigation pane on the right and a status bar at the bottom.

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

The screenshot displays a software interface for well header metadata extraction and petrophysical analysis. The interface is divided into several panels:

- Left Panel:** A tree view showing a hierarchy of documents. The selected document is "11/208_11_1_LOG/DONG Glenrothes Final MWD_LWD/175in_Sect".
- Header Panel:** A table of extracted metadata. The table has columns for the field name, value, and a confidence factor (0-100).

Field	Value	Confidence
Latitude Degrees	61°	75
Latitude Minutes	36'	81
Latitude Seconds	21.254"	96
Latitude Indicator	n	11
Longitude Degrees	1°	100
Longitude Minutes	52'	86
Longitude Seconds	40.625"	86
Longitude Indicator	w	79
Surface Latitude	61°36'21.254" n	11
Surface Longitude	01°52'40.625" w	11
- PETROPHYSICAL ANALYSIS Panel:** A detailed report for well "11/208_11_1" dated "208/11-1". It includes project information, operator details, and various measurements.

PETROPHYSICAL ANALYSIS 208/11-1			
PROJECT: UK_PO_WOS LOCATION: West Shetland		COUNTRY: UKCS LICENCE: P1454	WELL CLASSIFICATION: Exploration COMPLETION STATUS: Plugged & Abandoned
INTEREST: DONG E&P (UK) Ltd (40%) GDF BUREZ (20%) Dong Petroleum UK (30%)		OPERATOR: DONG Energy E&P (UK) Ltd ANALYSED BY: NCGIL APPROVED BY: XRAPR	
MUD TYPE: Water base mud (KCI)	TOTAL DEPTH: 4295 BOTTOM HOLE TEMP.: 130DEGC BOTTOM HOLE FM.: Sulfon Fm.	DATUM FOR ELEVATION: MSL MEASUREMENT REF.: DF ELEVATION MEAS. REF.: 36.00 GROUND LEVEL ELEV.: -1168.6	SPUD DATE: 27/9/2012 LOGGING COMPLETE: 15/1/2013 DATE PLOTTED: 30-May-2013
REMARKS: Shale volume from Density-Neutron crossplot Porosity from light hydrocarbon corrected density porosity with matrix density at 2.65 g/cc Water saturation from modified Simandoux with $m = 1$ and K_{f0} from wireline which is slightly optimistic compared to RT from LWD Formation water resistivity is very uncertain. 25,000 ppm NaCl has been used. 120,000 ppm Mud filtrate salinity (0.06% (mass) @ 30°C) Temperature from Hunter plot in Cragganmore (208/17-1) 125deg @ 3745m MD By error margin can be up to 30% - so is inaccurate due to the immature wellbore. Caliper is out of page (-0.3") Possible miscalculation through entire Vals or due to overbalance? The HC is not expected to be one continuous accumulation. It is more likely to be residual or dissolved in the water. Coals might be present in low sandstones that look like false HC saturation		SURFACE CO-ORDINATES: LATITUDE: 61° 36' 21.254" N LONGITUDE: 01° 52' 40.625" W UTM NORTHING: 6630054.75 UTM EASTING: 1724 BOTTOM-HOLE DEVIATION: UTM NORTHING: 1724 UTM EASTING: -0.69 UTM ZONE: 30 PROJECTIONS DATUM: EDNS	
- LITHOLOGY SYMBOLS Panel:** A table of symbols for different lithologies.

LITHOLOGY SYMBOLS	
Water	Shale
Clay bound Water	Sand
Gas - Condensate - Oil - Bitumen	Chalk
	Clay
	Coal
	Carbonate
	Volcanics
	Consolid
- OPERATIONAL SYMBOLS Panel:** A table of symbols for different operational parameters.

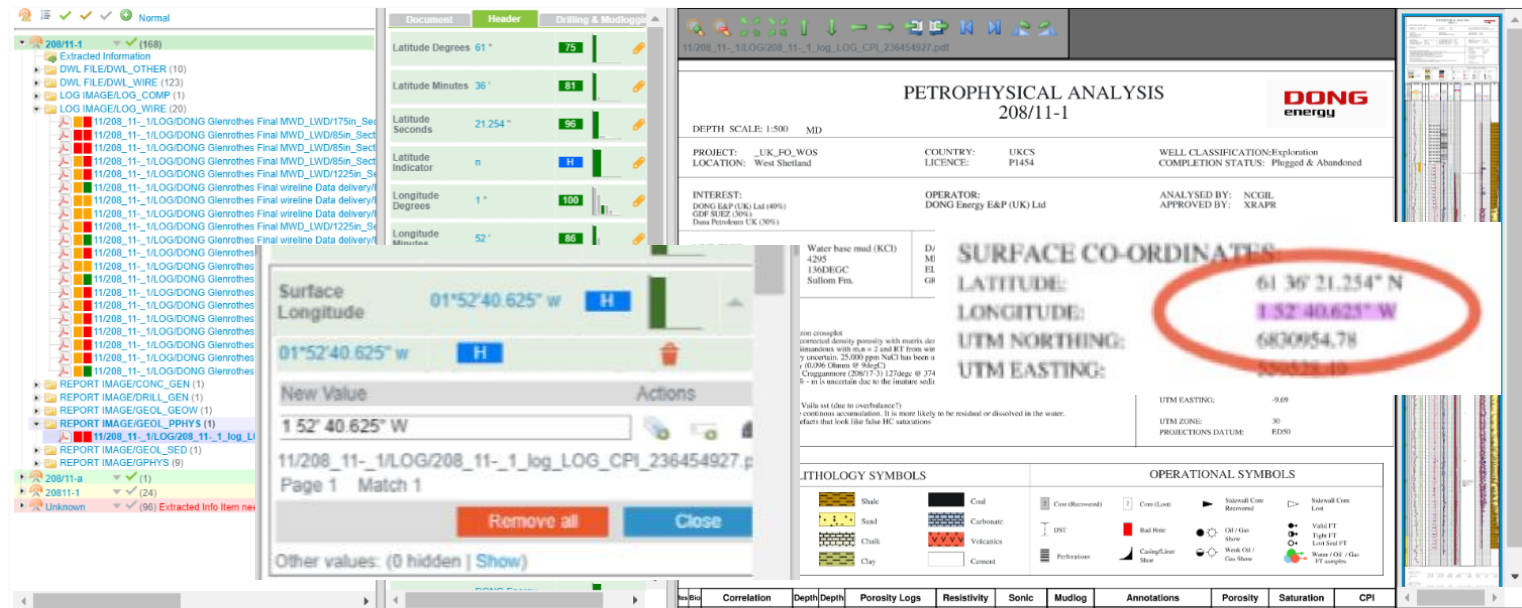
OPERATIONAL SYMBOLS	
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)
Core (Recomm)	Core (Load)

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its "container" such as a report, log, digital document ... and its "contents" such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata



Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

The screenshot displays a software interface for well data management. On the left, a file explorer shows a directory structure for well 208/11-1. The central panel features a 'Well Result' dialog box with a 'plug and abandon' status and a confidence score of 45. Below this, a 'New Value' field is set to 'Plugged & Abandoned'. The right panel shows a 'PETROPHYSICAL ANALYSIS' report for well 208/11-1, which includes fields for project, location, country, operator, and various technical specifications. A red circle highlights the 'COMPLETION STATUS' field, which is set to 'Plugged & Abandoned'.

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

The screenshot displays a software interface for well data management. On the left, a file explorer shows a hierarchy of files, including '208/11-1' and 'LOG IMAGE/LOG_WIRE'. The central document viewer shows a 'WELL CLASSIFICATION: Exploration' and 'COMPLETION STATUS: Plugged & Abandoned', with the latter circled in red. Below this, a 'Well Result' section shows 'plug and abandon 45'. The right-hand panel contains a table of metadata, including 'DATUM FOR ELEVATION: MSL', 'SPUD DATE: 27/9/2012', and 'LOGGING COMPLETE: 15/1/2013'. It also includes a 'GY SYMBOLS' section with a legend for various well types and a 'Porosity Logs' section with a legend for various log types.

Well Classification	Completion Status
Exploration	Plugged & Abandoned

Well Result	Value
plug and abandon	45

Well Header Metadata	Value
DATUM FOR ELEVATION	MSL
MEASUREMENT REF.	DF
ELEVATION MEAS. REF.	36.00
GROUND LEVEL ELEV.	-1168.6
SPUD DATE	27/9/2012
LOGGING COMPLETE	15/1/2013
DATE PLOTTED	30-May-2013

GY SYMBOLS	OPERATIONAL SYMBOLS
Shale	Core (Left)
Sand	Core (Right)
Chalk	Core (Left)
Clay	Core (Right)

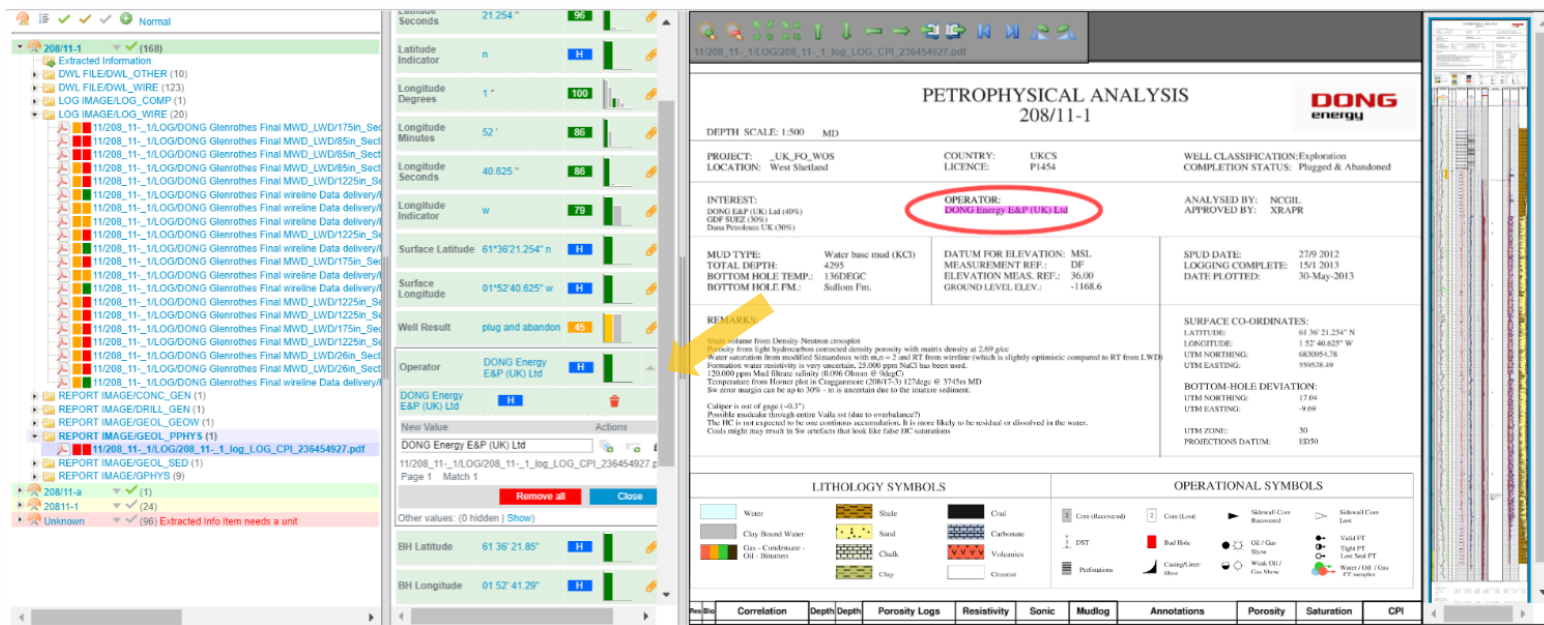
Porosity Logs	Resistivity	Sonic	Mudlog	Annotations	Porosity	Saturation	CPI

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata



Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

Example of indexing using the CS8 taxonomy and extraction of some well header metadata

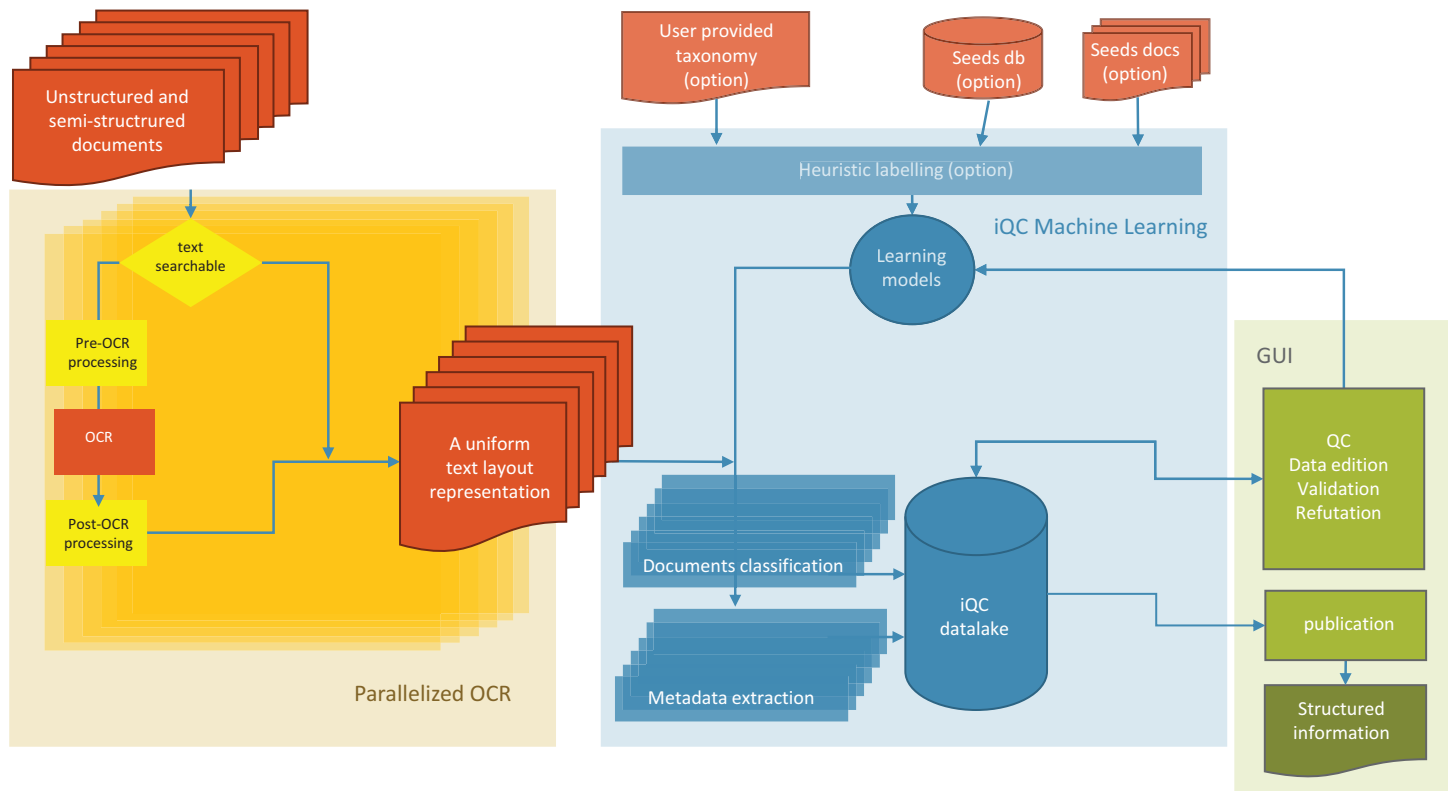
The screenshot displays a software interface for well header metadata extraction. On the left, a file explorer shows a directory structure with files like 'DWL_FILE/DWL_OTHER (10)', 'DWL_FILE/DWL_WIRE (123)', and 'LOG IMAGE/LOG_COMP (1)'. The central panel shows a table of extracted metadata for well '208/11-1'. The table includes fields like 'Latitude Indicator', 'Longitude Degrees', 'Longitude Minutes', 'Longitude Seconds', and 'Longitude Indicator'. A red oval highlights the 'OPERATOR: DONG Energy E&P(UK) Ltd' text in the central panel. The right-hand panel displays various metadata fields, including 'SPUD DATE', 'LOGGING COMPLETE', 'DATE PLOTTED', 'SURFACE CO-ORDINATES', 'BOTTOM-HOLE DEVIATION', and 'PROJECTIONS DATUM'. At the bottom, there are tabs for 'Correlation', 'Depth/Depth', 'Porosity Logs', 'Resistivity', 'Sonic', 'Mudlog', 'Annotations', 'Porosity', 'Saturation', and 'CPI'.

Presenter's notes:

These few snapshots illustrate the capacity of our tools to:

- Detect the document category using the CS8 taxonomy used in the UK to describe each subsurface document. The CS8 taxonomy defines a document according to its “container” such as a report, log, digital document ... and its “contents” such as mud-logging, petrophysics, seismic, engineering ... Does the same automatically after training.
- Detect some well header metadata, text or numerical, using their context
- Associate a confidence factor to each extraction
- Display the variability of a particular metadata value for the same well

How does it work?

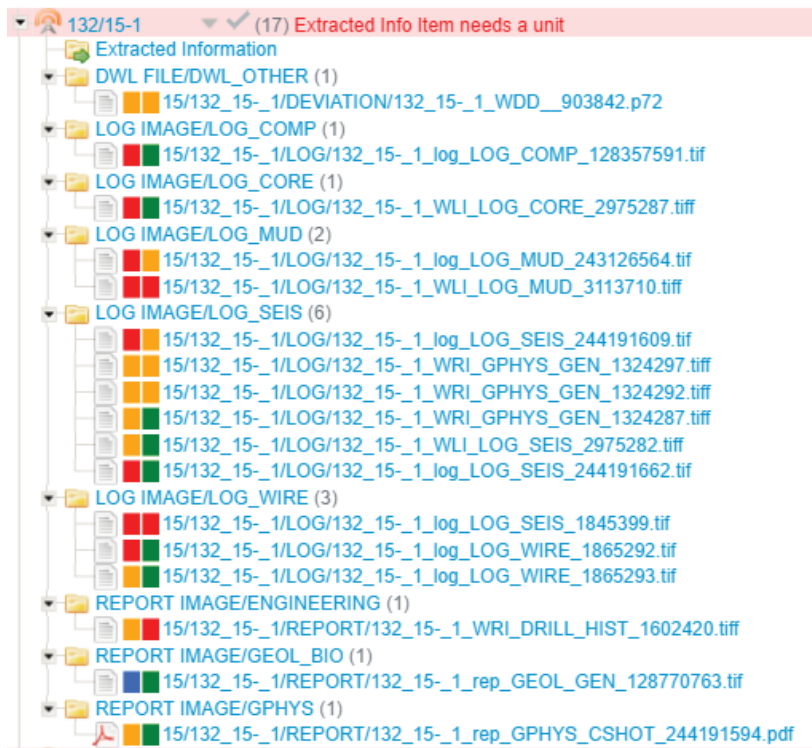


Document category



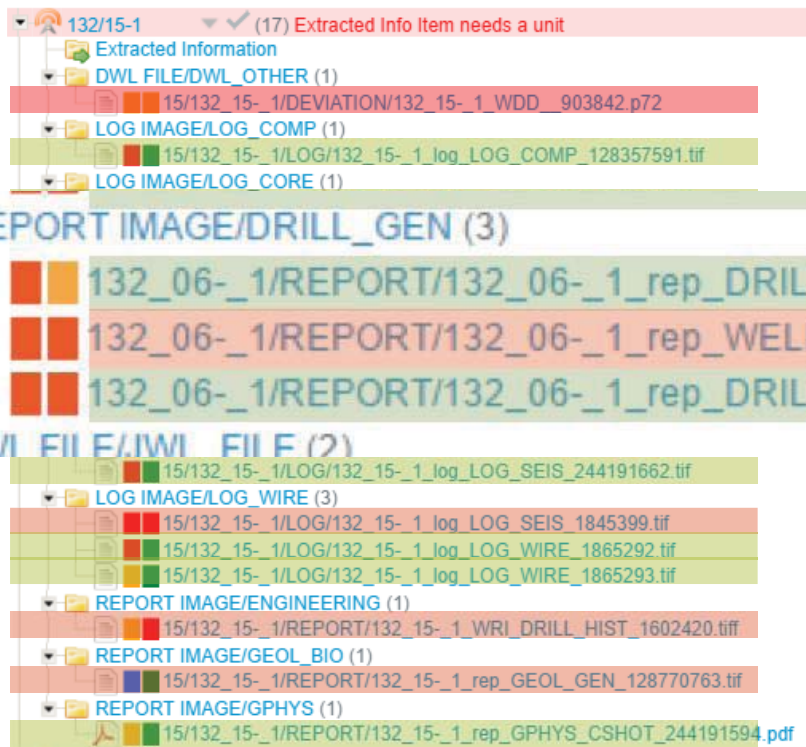
- In 2 days, iQC has been trained to use the CS8 taxonomy, using 2000 “seed” examples

Document category



- This initial training was sufficient to have an 80% match with the manual indexing done over many years on the 450,000 documents.
- The ratio moved up to 90% with a seed of 5,000 documents
- The 10% of remaining discrepancies relate to “manual errors” or rare and ambiguous taxonomy classes.

Document category

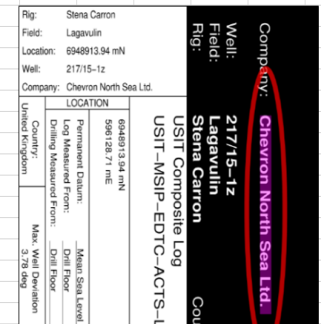
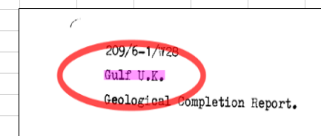
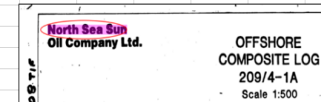


- This initial training was sufficient to have an 80% match with the manual indexing done over many years on

relate to “manual errors” or rare and ambiguous taxonomy classes.

QC of the CDA tables according to the source documents

Wellbore	Attribute	Value	Confid CDA wellbore	Original operator		
209/12-1	Operator	Marathon Oil U.K.. Ltd.	100 209/12- 1	MARATHON OIL (UK) LTD		
209/3-1a	Operator	MOBIL NORTH SEA LTD.	100 209/03- 1a	MOBIL NORTH SEA LIMITED		
209/4-1a	Operator	North Sea Sun Oil Company Ltd	100 209/04- 1a	AGIP (U.K.) LIMITED	OperatorU69	
209/6-1	Operator	Gulf U.K.	100 209/06- 1	CHEVRON U.K. LIMITED	OperatorU75	
213/23-1	Operator	MOBIL NORTH SEA LTD	100 213/23- 1	MOBIL NORTH SEA LIMITED		
213/26-1z	Operator	chevron	76 213/26- 1z	CHEVRON		
217/15-1z	Operator	Chevron North Sea Ltd.	100 217/15- 1z		0 OperatorU82	
220/26-1	Operator	B.P. PETROLEUM DEVELOPMENT	100 220/26- 1	BP EXPLORATION OPERATING COMPANY LIMITED		
220/26-1st	Operator	B.P. PETROLEUM DEVELOPMENT	100 220/26- 1st	#N/A		
220/26-2	Operator	BP Petroleum Development Ltd	100 220/26- 2	BP EXPLORATION OPERATING COMPANY LIMITED		
74/1-1	Operator	Gas Council (Exploration) Ltd	100 74/01- 1	BG INTERNATIONAL LTD		
83/24-1	Operator	DEPARTMENT OF ENERGY	100 83/24- 1	HMG DEPARTMENT OF TRADE AND INDUSTRY		
86/17-1	Operator	Murphy Petroleum Limited	100 86/17- 1	MURPHY PETROLEUM LIMITED		
86/18-1	Operator	The British National Oil Corporation	100 86/18- 1	BP EXPLORATION OPERATING COMPANY LIMITED		
87/12-1a	Operator	BP Petroleum Development Ltd	100 87/12- 1a	BP EXPLORATION OPERATING COMPANY LIMITED		
87/14-1	Operator	Department of Energy, UK	100 87/14- 1	HMG DEPARTMENT OF TRADE AND INDUSTRY		
87/16-1	Operator	Department of Energy, U.K.	100 87/16- 1	HMG DEPARTMENT OF TRADE AND INDUSTRY		
88/2-1	Operator	Department of Energy, U.K	100 88/02- 1	HMG DEPARTMENT OF TRADE AND INDUSTRY		
8/15- 1	Operator	UNOCAL U.K. LIMITED	100 8/15- 1	UNOCAL NORTH SEA EXPLORATION LIMITED		
8/15-2a	Operator	Total Exploration and Production UK Ltd.	100 8/15- 2a	TOTAL		
8/20-1	Operator	Agip (UK) Ltd	100 8/20- 1	AGIP (U.K.) LIMITED		
8/25a-1	Operator	nautical petroleum	100 8/25a- 1	NAUTICAL		
8/30-1	Operator	Sovereign Oil & Gas Plc	100 8/30- 1	SANDS EXPLORATION LIMITED	OperatorU108	
8/4-1	Operator	Chevron UK Ltd	100 8/04- 1	CHEVRON U.K. LIMITED		



Presenter's notes: The automatic well header metadata extraction authorize to QC the CDA DB according to the source documents. This example shows the discrepancies on a text metadata (the well operator)

QC of the CDA tables according to the source documents

IQC wellbore	title	IQC val	confiden		IQC value	CDA Water		Delta		
			unit	ceScore		Depth (m)	Depth (m)			
214/17-1	Water Depth	3832 ft	80	214/17-1	1167.99	1167.99	0.00			
214/19-1	Water Depth	1119 m	100	214/19-1	1119.00	1000.05	118.95	Water Depth IM3		
214/4-1	Water Depth	5317 ft	60	214/04-1	1620.62	1620.62	0.00			
214/9-1	Water Depth	5089 ft	60	214/09-1	1551.13	1556	4.87	Water Depth IM23		
1/4-1	Water Depth	506 ft	100	1/04-1	154.23	154.23	0.00			
1/4-2	Water Depth	479 ft	100	1/04-2	146.00	146	0.00			
18/3-1	Water Depth	263 ft	100	18/03-1	80.16	80.1624	0.00			
18/5-2	Water Depth	92.6 ft	100	18/05-2	28.22	119	90.78	Water Depth IM32		
26/12-1	Water Depth	144 ft	100	26/12-1	43.89	43.89	0.00			
26/14-1	Water Depth	225 ft	100	26/14-1	68.58	68.58	0.00			
26/4-1	Water Depth	219 ft	100	26/04-1	66.75	67.06	0.31			
26/7-1	Water Depth	64.9 m	100	26/07-1	64.90	64.9	0.00			
26/8-1	Water Depth	186 ft	100	26/08-1	56.69	56.69	0.00			
27/10-1	Water Depth	262 ft	100	27/10-1	79.86	79.86	0.00			
27/3-1	Water Depth	246 ft	100	27/03-1	74.98	74.98	0.00			
4/26-1A	Water Depth	351 ft	100	4/26-1A	106.98	132.89	25.91	Water Depth IM51		
4/26-2	Water Depth	342 ft	100	4/26-2	104.24	104.24	0.00			
7/16-1	Water Depth	417 ft	100	7/16-1	127.10	127.1	0.00			
102/28-2	Water Depth	315 N	100	102/28-2	96.01	96.01	0.00			
103/1-1	Water Depth	316 ft	100	103/01-1	96.32	96.32	0.00			
103/1-2	Water Depth	314 ft	100	103/01-2	95.71	no value	no value			
103/18-1	Water Depth	232 N	100	103/18-1	70.71	70.71	0.00			
103/18-2	Water Depth	314 ft	26	103/01a-2	95.71	98.45	2.74			
103/2-1	Water Depth	375 ft	100	103/02-1	114.30	114.3	0.00			
103/21-1	Water Depth	276 ft	100	103/21-1	84.12	84.12	0.00			
106/18-1	Water Depth	299 ft	100	106/18-1	91.14	91.14	0.00			
106/20-1	Water Depth	79.5 m	100	106/20-1	79.50	79.5	0.00			
106/24-1	Water Depth	252 ft	100	106/24-1	76.81	76.81	0.00			
106/24a-2B	Water Depth	241 ft	100	106/24a-2B	73.46	73.46	0.00			
106/28-1	Water Depth	359 ft	100	106/28-1	109.42	109.42	0.00			
107/1-1	Water Depth	237 ft	100	107/01-1	72.24	72.24	0.00			
107/16-1	Water Depth	140 ft	100	107/16-1	42.67	42.67	0.00			
107/21-1	Water Depth	176 ft	100	107/21-1	53.64	53.64	0.00			
93/2-1	Water Depth	97.5 m	100	93/02-1	97.50	97.54	0.04			
93/2-3	Water Depth	98.45 m	100	93/02-3	98.45	98.15	0.30			
93/6-1	Water Depth	336 ft	100	93/06-1	102.41	104.24	1.83			
99/12-1	Water Depth	53 ft	100	99/12-1	16.15	16.46	0.31			

214/19-1 SHELL EXPRO

WELL INFORMATION

WELL 214/19-1
WELL TYPE EXPLORATION
COMPANY SHELL EXPRO
BLOCK UK 214/19
LOCATION FAROE-SHETLAND BASIN
CO-ORDINATES
Latitude 61° 24' 45.78" N
Longitude 02° 15' 23.5" W
PERMANENT DATUM Mean Sea Level
DEPTH MEASURED FROM Below Rotary Table
ROTARY TABLE ELEVATION 76 m
WATER DEPTH 1119 m
TOTAL DEPTH 4587 m MDBRT

Mobil North Sea Ltd.
214/9-1
West of Britain
Jack Bates
U.K.

CBT-GR-CCL
21-OCT-2000
SCALE 1:200

COMPANY : Apache North Sea Ltd.
WELL NUMBER : Golden Arrow 19/5-2
TYPE : Exploration
RIG OPERATOR : Petrolia Drilling Ltd.
RIG NAME : Petrolia
SURFACE LOCATION : 57° 53' 4.632" N

Presenter's notes: Our capacity to extract numerical variables allowed us to QC the CDA DB according to the sources docs. It showed surprising "human errors" on some very easy to detect metadata. The main value is not only to alert on discrepancies but also to show immediately the documented source of information we have used to alert us, as well as our confidence in the automatic detection.

QC of the CDA tables according to the source documents

IQC wellbore	title	IQC val	unit	confiden	CDA wellname	IQC value	CDA Water	Delta	
214/17-1	Water Depth	3832 ft	80	214/17-1	1167.99	1167.99	0.00		
214/19-1	Water Depth	1119 m	100	214/19-1	1119.00	1000.05	118.95	Water Depth!M3	
214/4-1	Water Depth	5317 ft	60	214/04-1	1620.62	1620.62	0.00		
26/12-1	Water Depth	144 ft	100	26/12-1	43.89	43.89	0.00		
26/14-1	Water Depth	225 ft	100	26/14-1	68.58	68.58	0.00		
26/4-1	Water Depth	219 ft	100	26/04-1	66.75	67.06	0.31		
26/7-1	Water Depth	64.9 m	100	26/07-1	64.90	64.9	0.00		
26/8-1	Water Depth	186 ft	100	26/08-1	56.69	56.69	0.00		
27/10-1	Water Depth	262 ft	100	27/10-1	79.86	79.86	0.00		
27/3-1	Water Depth	246 ft	100	27/03-1	74.98	74.98	0.00		
4/26-1A	Water Depth	351 ft	100	4/26-1A	106.98	132.89	25.91	Water Depth!M51	
4/26-2	Water Depth	342 ft	100	4/26-2	104.24	104.24	0.00		
7/16-1	Water Depth	417 ft	100	7/16-1	127.10	127.1	0.00		
102/28-2	Water Depth	315 \N	100	102/28-2	96.01	96.01	0.00		
103/1-1	Water Depth	316 ft	100	103/01-1	96.32	96.32	0.00		
103/1-2	Water Depth	314 ft	100	103/01-2	95.71	no value	no value		
103/18-1	Water Depth	232 \N	100	103/18-1	70.71	70.71	0.00		
103/1a-2	Water Depth	314 ft	26	103/01a-2	95.71	98.45	2.74		
103/2-1	Water Depth	375 ft	100	103/02-1	114.30	114.3	0.00		
103/21-1	Water Depth	276 ft	100	103/21-1	84.12	84.12	0.00		
106/18-1	Water Depth	299 ft	100	106/18-1	91.14	91.14	0.00		
106/20-1	Water Depth	79.5 m	100	106/20-1	79.50	79.5	0.00		
106/24-1	Water Depth	252 ft	100	106/24-1	76.81	76.81	0.00		
106/24a-2B	Water Depth	241 ft	100	106/24a-2B	73.46	73.46	0.00		
106/28-1	Water Depth	359 ft	100	106/28-1	109.42	109.42	0.00		
107/1-1	Water Depth	237 ft	100	107/01-1	72.24	72.24	0.00		
107/16-1	Water Depth	140 ft	100	107/16-1	42.67	42.67	0.00		
107/21-1	Water Depth	176 ft	100	107/21-1	53.64	53.64	0.00		
93/2-1	Water Depth	97.5 m	100	93/02-1	97.50	97.54	0.04		
93/2-3	Water Depth	98.45 m	100	93/02-3	98.45	98.15	0.30		
93/6-1	Water Depth	336 ft	100	93/06-1	102.41	104.24	1.83		
99/12-1	Water Depth	53 ft	100	99/12-1	16.15	16.46	0.31		

214/19-1 SHELL EXPRO

UK 214/19

LOCATION FAROE-SHETLAND BASIN

CO-ORDINATES Latitude 61° 24' 45.78" N Longitude 02° 15' 23.5" W

PERMANENT DATUM Mean Sea Level

DEPTH MEASURED FROM Below Rotary Table

ROTARY TABLE ELEVATION 26 m

WATER DEPTH 1119 m

TOTAL DEPTH 4867 m MDBRT

Mobil North Sea

214/9-1

West of Britain

Jack Bates

U.K.

larger

COMPANY

WELL NUMBER

TYPE Exploration

RIG OPERATOR : Petrolia Drilling Ltd.

RIG NAME : Petrolia

SURFACE LOCATION : 57° 53' 4.632" N

Presenter's notes: Our capacity to extract numerical variables allowed us to QC the CDA DB according to the sources docs. It showed surprising "human errors" on some very easy to detect metadata. The main value is not only to alert on discrepancies but also to show immediately the documented source of information we have used to alert us, as well as our confidence in the automatic detection.

Lessons learnt from the CDA challenge

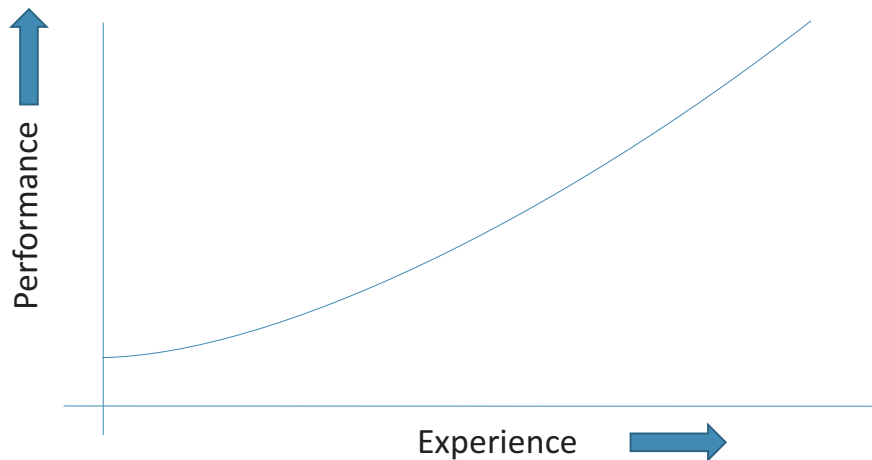
- Machine learning detects metadata in unstructured documents that other methods cannot detect
- It supports the QC of a structured database using unstructured sources
- It makes it possible to easily extend the contents of the database “on demand”
- For us, the CDA challenge was also an opportunity to enrich our learning model and make it more stable

Presenter's notes: Our capacity to extract numerical variables allowed us to QC the CDA DB according the sources documents. It showed surprising “human errors” on some very easy to detect metadata. The main value is not only to alert on discrepancies but also to show immediately the documented source of information we have used to alert as well as our confidence in the automatic detection.

The (machine) learns with us!

Overview [\[edit\]](#)

Tom M. Mitchell provided a widely quoted, more formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."^[9] This definition is notable for its defining machine learning in fundamentally operational rather than cognitive terms, thus following Alan Turing's proposal in his paper "Computing Machinery and Intelligence" that the question "Can machines think?" be replaced with the question "Can machines do what we (as thinking entities) can do?"^[10]



The (machine) learns with fresh data!

Looking ahead

- We are looking for documents to crunch
- We are engaging with early adopters
 - To perform pilot projects and feasibility studies system relative to their needs and objectives
 - To evaluate the performance of our machine learning
- We work in the cloud (Microsoft Azure®)
 - Our Learning models improve continuously
 - New users benefit from all accrued learning
 - An alternative configuration is to install our system locally



Most promising IT & web
startup 2017



philip.neri@agiledd.com

www.agiledd.com

