EAGE
EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

EAGE Conference on Reservoir Geoscience

3-5 DECEMBER 2018 • KUALA LUMPUR, MALAYSIA

WWW.EAGE.ORG

Elastic Docs As An Automated Information Retrieval Platform For Unstructured Reservoir Data Utilizing a Sequence of Smart Machine Learning Algorithms within a Hybrid Cloud Container

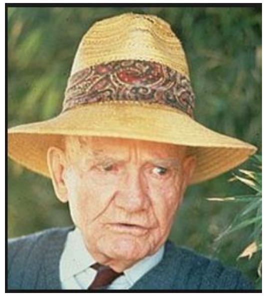
N.M. Hernandez, P.J Lucanas, I. Panganiban, C. Mamador, C.Yu *Iraya Energies*

3-5 DECEMBER 2018 • KUALA LUMPUR, MALAYSIA

Outline of Presentation

- Problem Statement
  - *Dealing with Unstructured Reservoir Data*
- Methodology
  - *Machine Learning*
  - *Database & Infrastructure*
- Results
  - *Supervised and Unsupervised*
  - *NLP & Reservoir Image Processing*
  - *Metrics*
  - *API Demonstration*
- Conclusion

EAGE



“Where oil is first found is in the minds of men”

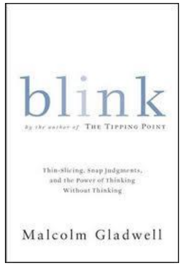
-Wallace E. Pratt

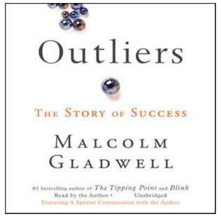
Pioneer Petroleum Geologist (1959)

“Good intuition requires years of practice”

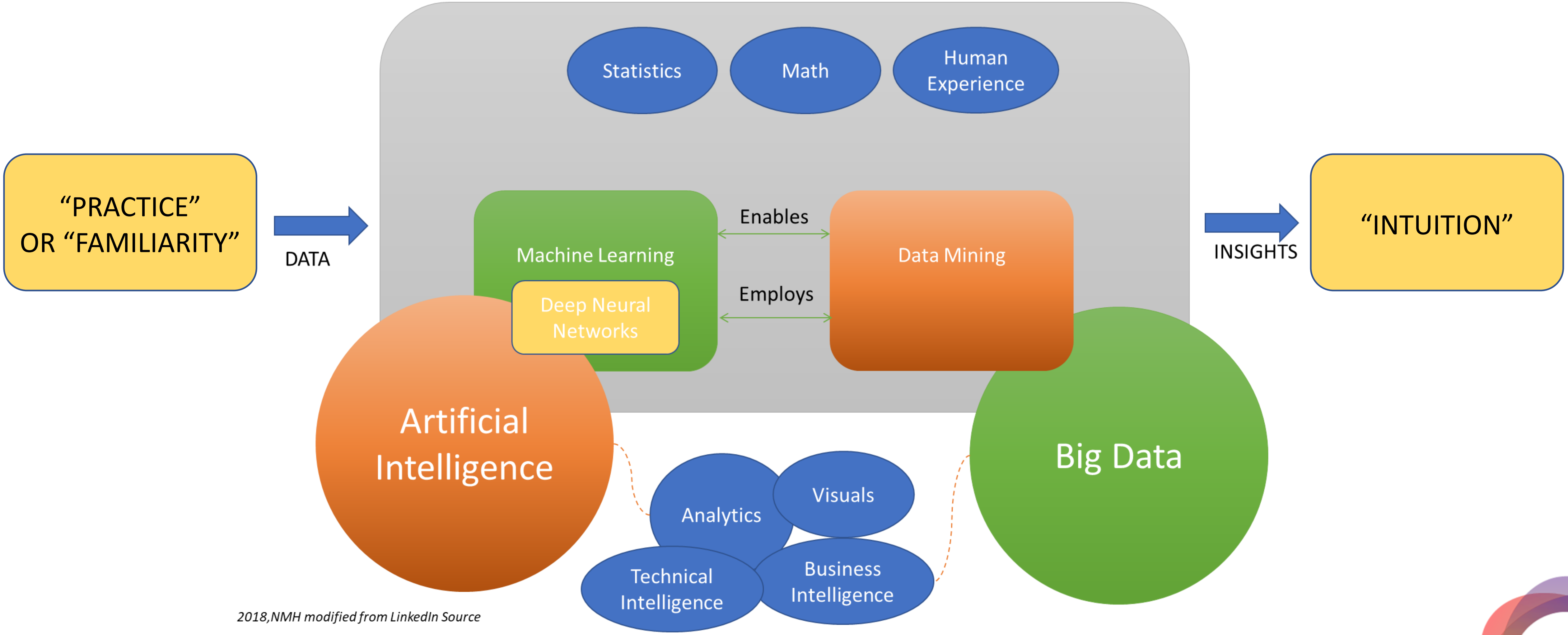
-Malcolm Gladwell

Author, Blink





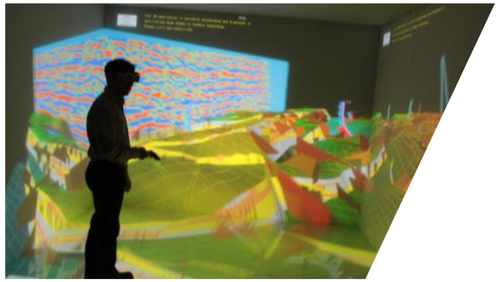
Symbiosis between ML and Data Mining

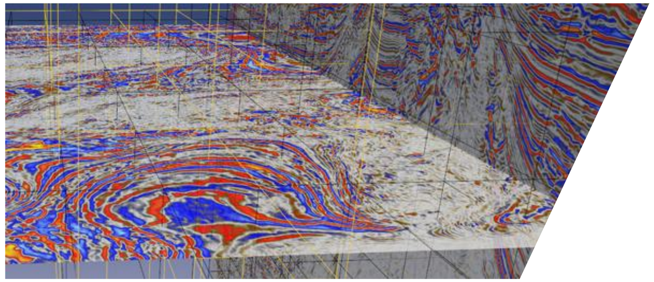


2018,NMH modified from LinkedIn Source

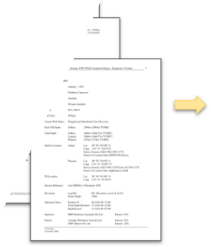
Problem Statement: ElasticDocs

Highly advanced platforms for well and seismic data





Unstructured documents are inaccessible



	WELL_NAME	BASIN	PERMIT	OPERATOR	LATITUDE	LO
Abalone1_FWR	Abalone -1	Browse Basin	AC/P29	Operations Summary	[illegible]	
Abalone_1_Final_Well_Report_Govt	Abalone—In	Browse	AC/P 29	J APEX AC Ltd\n	[illegible]	[illegible]
Acme 1 FWR	Acme 1	Carnarvon Basin	WA - 205-P\n	Wayne Talbot, Kunal Mishra	[illegible]	[illegible]
Amulet_1_CH1_Well_Completion_Report_Basic	Amulet-1		WA-8-L	Tap Oil Limited	[illegible]	[illegible]
Amulet_2_Well_Completion_Report_Basic_W21078A1	Amulet-2		WA-8-L	Kufpec Australia Pty Ltd	[illegible]	[illegible]
Antiope_1_ST1_Interpretive_Well_Completion_Report_main_text	Antiope — USTI	Northern Carnarvon		Antiope-USTI	[illegible]	
Appendix_08_Backpacker_1_BH_Mudlog_FWR	Backpacker-1	Darryl Meakins		Newfield Australia (Cartier) Pty. Ltd.	[illegible]	[illegible]
Argus_1_Well_Completion_Report_Basic_Data_Vol_Main_Text	Argus—1	Ocean General	AC/P30, Argus-1	BHP Petroleum	[illegible]	
BALNAVES_DEEP-3_SDL_FWR	Balnaves Deep-3			Apache Energy Limited (AEL)	[illegible]	[illegible]
Backpacker1_FWR	Backpacker-1	Darryl Meakins		Newfield Australia (Cartier) Pty. Ltd.	[illegible]	[illegible]
Balnaves Deep-2_SDL_FWR	Balnaves Deep-2			Apache Energy Limited (AEL)	[illegible]	[illegible]
Banambu Deep-1 Final Well Completion Report	Banambu Deep-1	Analysis Banambu		Banambu Deep-1 Tectonic Elements and Location ...		
Final_Well_Report_MEO_Australia_Artemis_1_FINAL	Artemis-1		WA-360-P	MEO Australia Ltd\n	[illegible]	[illegible]
Well Completion Report FINAL	Audacious—5		AC/RL6	OMV Australia Pty Ltd.	[illegible]	

- Reservoir Properties
- Geological Interpretations

Technical Memory is Lost

EAGE

Problem Statement

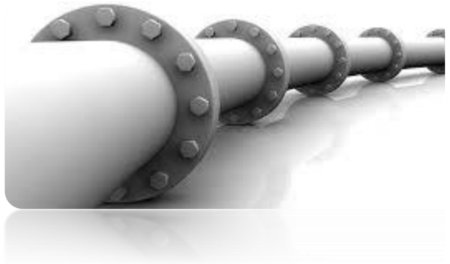
Using the latest advances in Machine Learning:

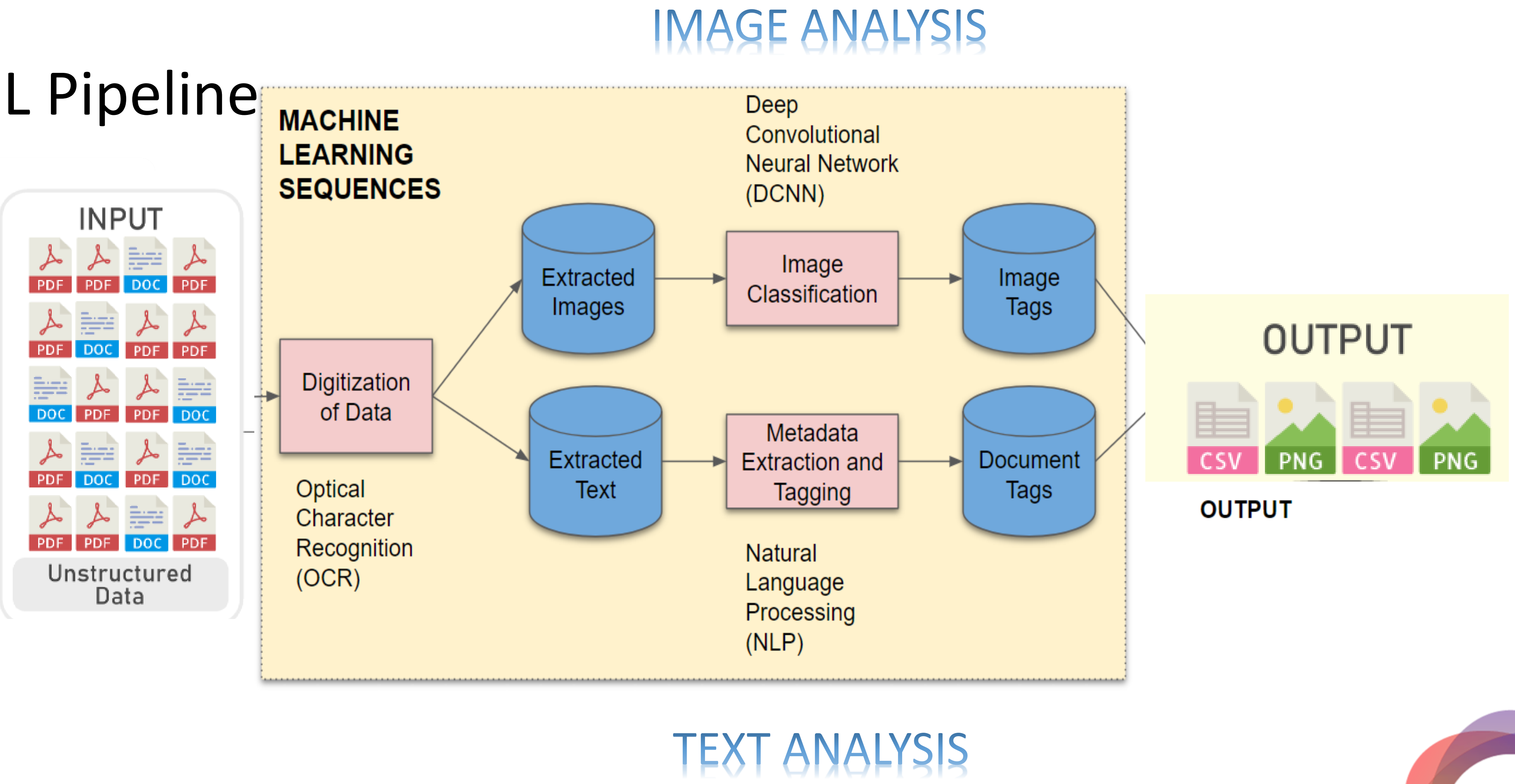
- How to gain reservoir experience leveraging from existing data
- How to maintain or recover corporate memory

EAGE

Methodology

ElasticDocs ML Pipeline





EAGE

Methodology

IT Infrastructure

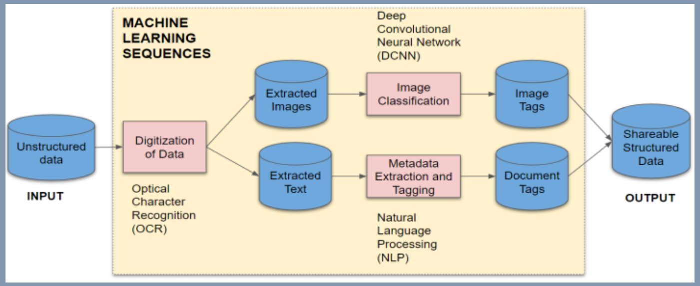
GPU
CPU

Cloud-Based
Or Hybrid
Infrastructure

Database
SQL
NoSQL

Microservices

ML Core



User-Facing API

Web Layer

GIS Layers

CURATED, OPEN SOURCE LIBRARIES *Python, Tensorflow, Leaflet, Elasticsearch,etc*

Natural Language Processing

Named Entity Recognition

Geology Identification

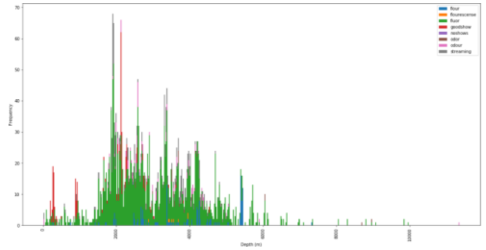
("the glauconitic claystone sample is barren of foraminifera. no definitive environment interpretation is possible.",{'entities':[(4,25,'GEOL')]}),

("massive claystone interbedded with silty claystone and thin argillaceous siltstone",{'entities':[(8, 17, 'GEOL'), (35, 50, 'GEOL'), (60, 82, 'GEOL')]}),

("chevron australia pty ltd acme 1",{'entities':[]}), *(example of no geological entity to be detected)*

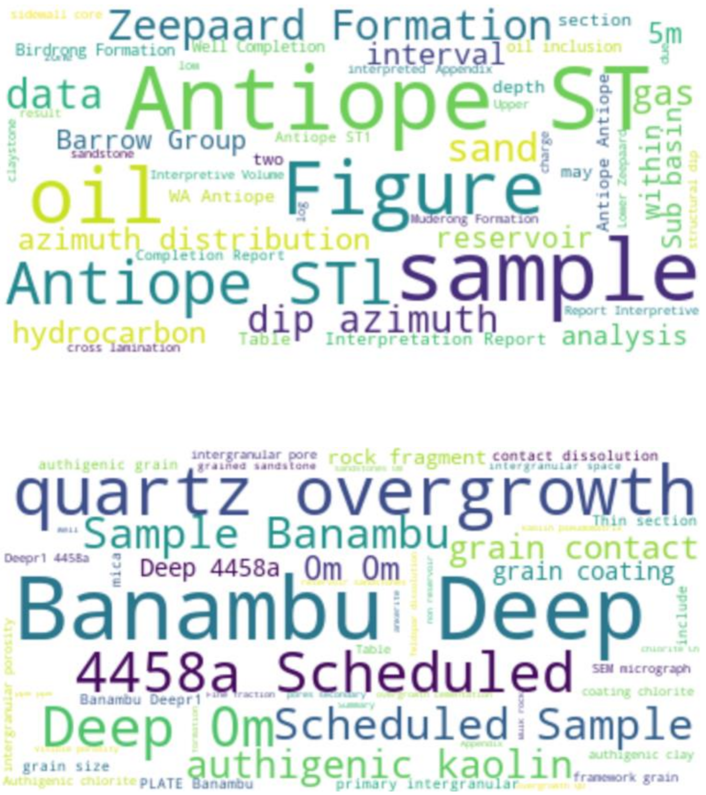
Hydrocarbon show Identification

("STAIN N~D FLUORESCENCE 3699-3709m This interval constitutes the upper part of the ",{'entities':[(23, 27, 'DEPTH'), (28, 32, 'DEPTH')]})



Natural Language Processing

Word Clouds



Unsupervised Clustering with Topic Models

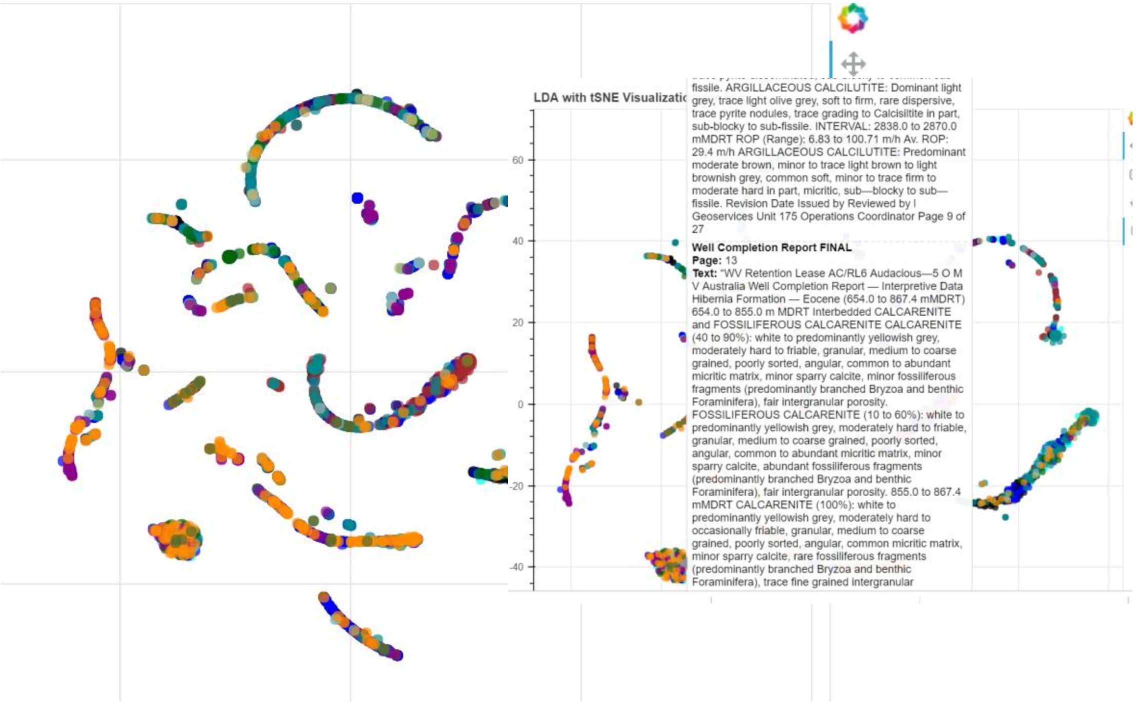


Image Recognition

Multi-Format Images

	Precision	Recall	F1-score
Map	0.83	0.96	0.89
Seismic	1.00	0.95	0.97
Core	0.89	0.98	0.94
SEM	0.95	0.93	0.94
Others	0.91	0.73	0.81

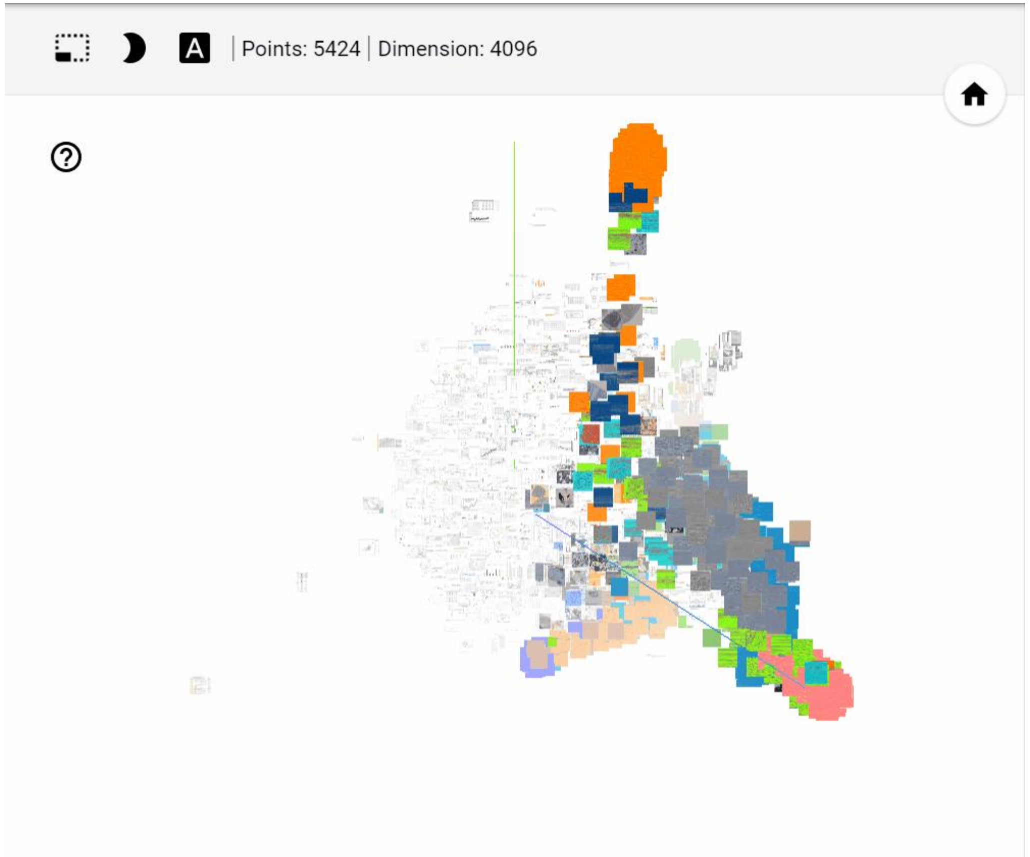
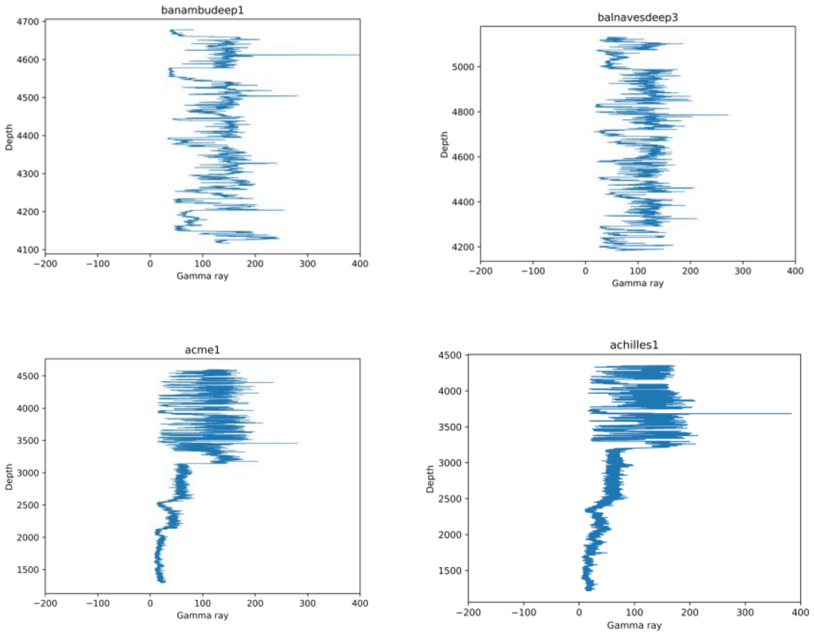


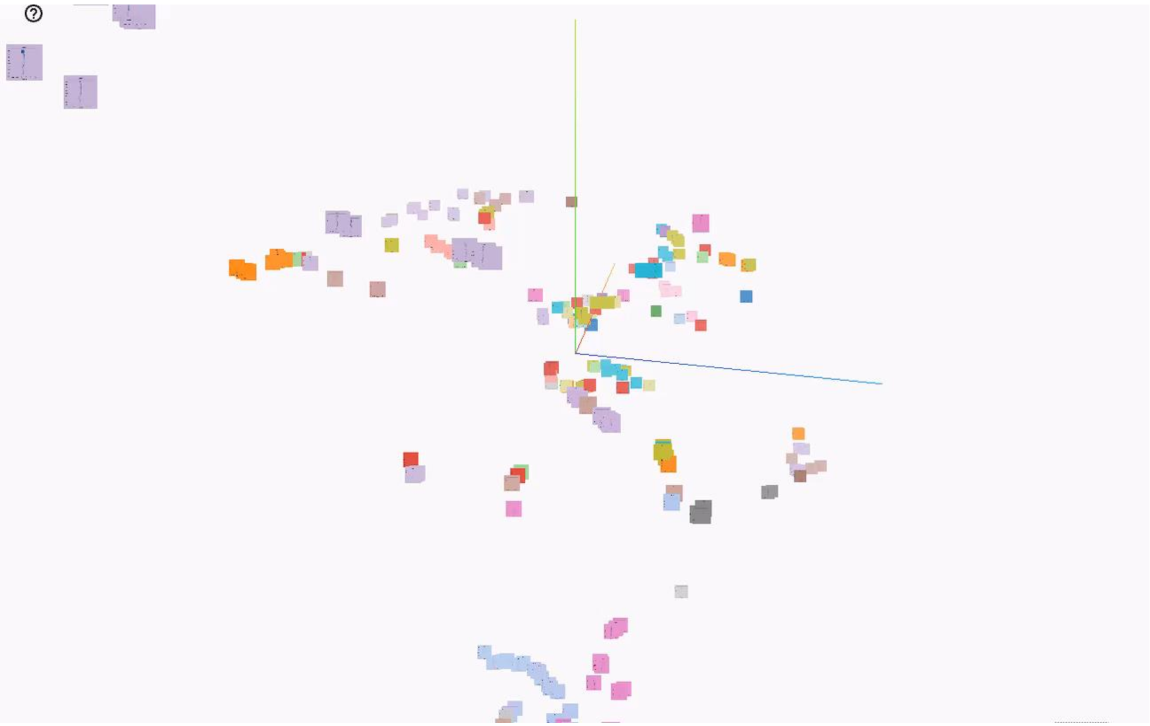
Image Recognition

Well Twinning

Twin Set A

Twin Set B





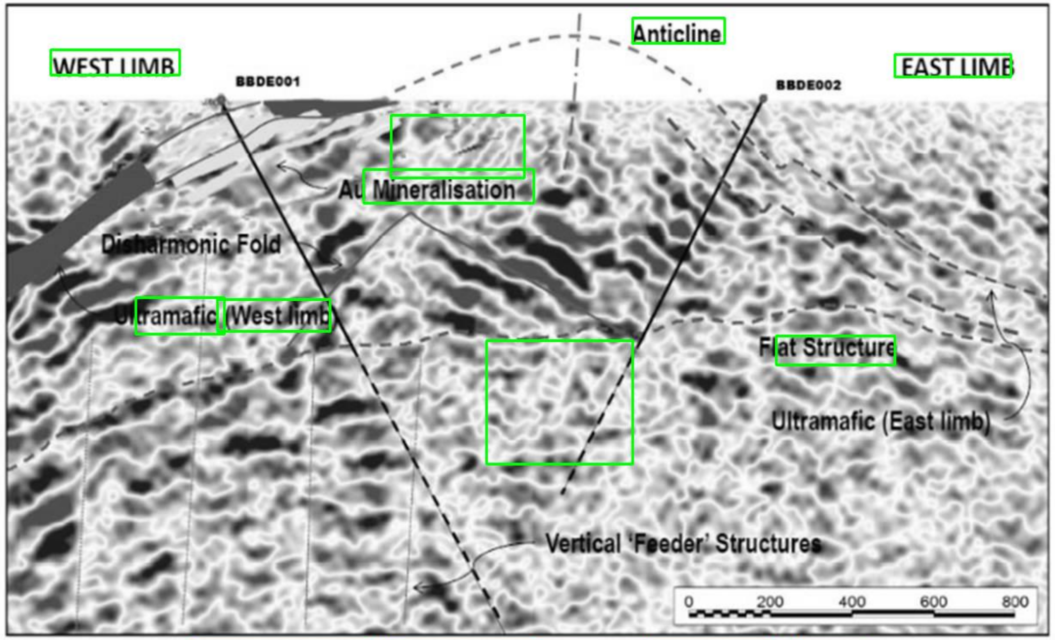
Applications in:

- ❑ Analog Search
- ❑ Anomaly detection
- ❑ Quality control

Scene Detection

Train Station Signage

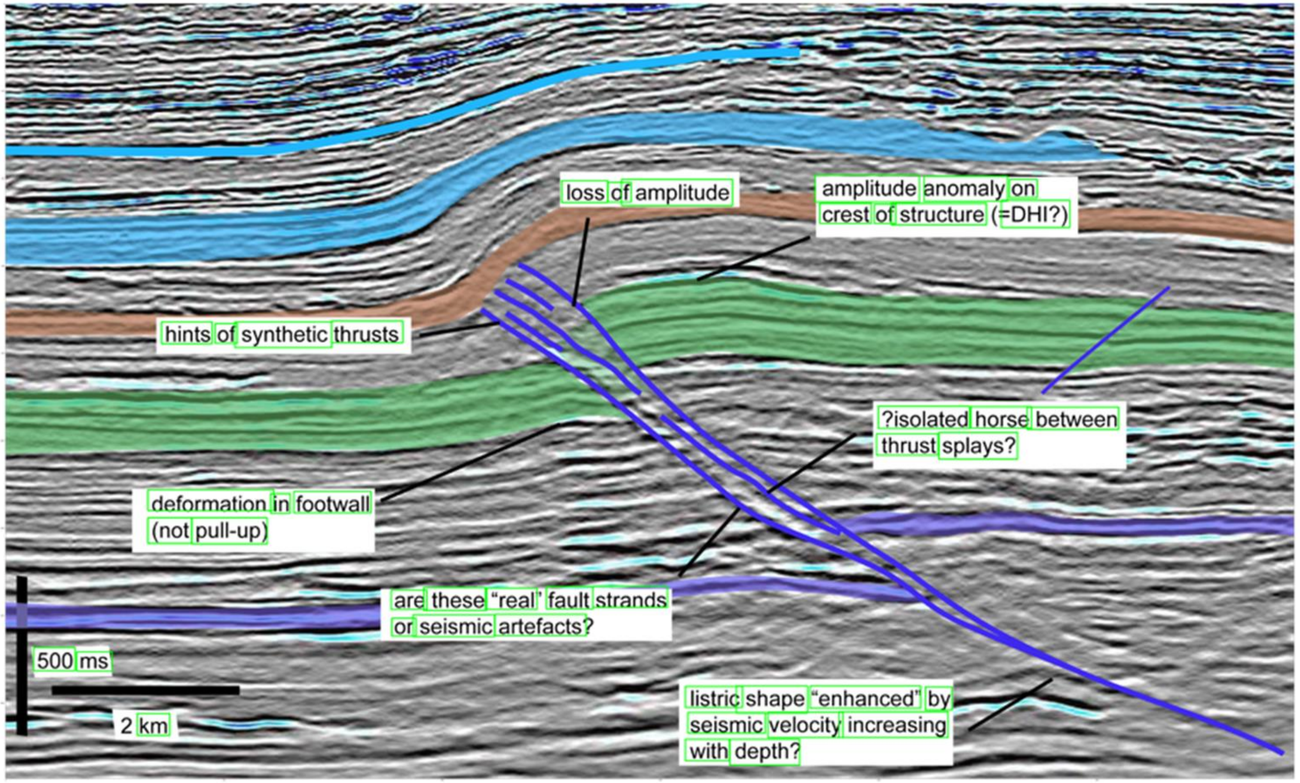
Seismic Signage



Cross-section on Seismic Line Showing Planned Drill Holes

EAGE

Scene Detection

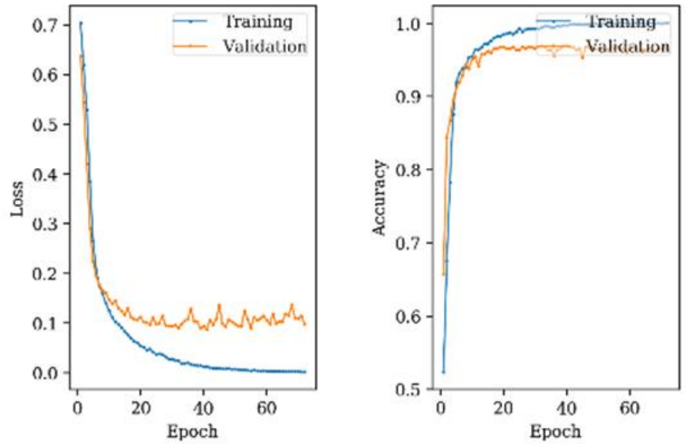


thrust
on
or
listric
'with
ms
are
crest
500
'increasing
"anhanced'
of
<trands
loss
seismic
seismic
f amplitude
between
artefacts'
these
hints
thrusts
anomaly

km
(not
footwall
velocity
"real
'horse
by
fault
splays?
deformation
?isolated
structure
> shape
'amplitude
depth?
pull-up)
of
in
of
'synthetic
=DHI?)

Metrics

Speed



Training Loss

Accuracy

- Precision: proportion of positive identification is correct

$$Precision = \frac{T.P.}{T.P. + F.P.}$$

- Recall: proportion of actual positives is correct

$$Recall = \frac{T.P.}{T.P. + F.N.}$$

- F1 score: harmonic mean of precision and recall

$$F1\ score = \frac{2(precision * recall)}{precision + recall}$$

		Predicted	
		0	1
Actual	0	T.N.	F.P.
	1	F.N.	T.P.

Precision and Recall

EAGE

Metrics

ML Application	Task	Speed
OCR	Text Extraction only, excluding Image Classification	150,000 pages 13 hrs
	Text Extraction and Image Extraction	4,542 pages 6.31 GB 25 Final Well Reports 10 hrs
NLP	Lithology / Geology Indicator Frequency Analysis (i.e Carbonates, Sandstone, etc)	4 hours
	Well Cataloguing	1,500+ input las files 66, 515 curves identified 5,681 top log curves (cali, gr, neu, por..) 2 hrs 33.66 min
DCNN	Imag Classification	2,598 tagged images input 16% Tables 6% Figures 19% Map 24% Charts 33% Noise 20-30 mins during training <s after training Currently includes 8 classes: thin section SEM, seismic, stratigraphic chart, cores, map and general classes such as chart, figure, table

Tying it all together with an API

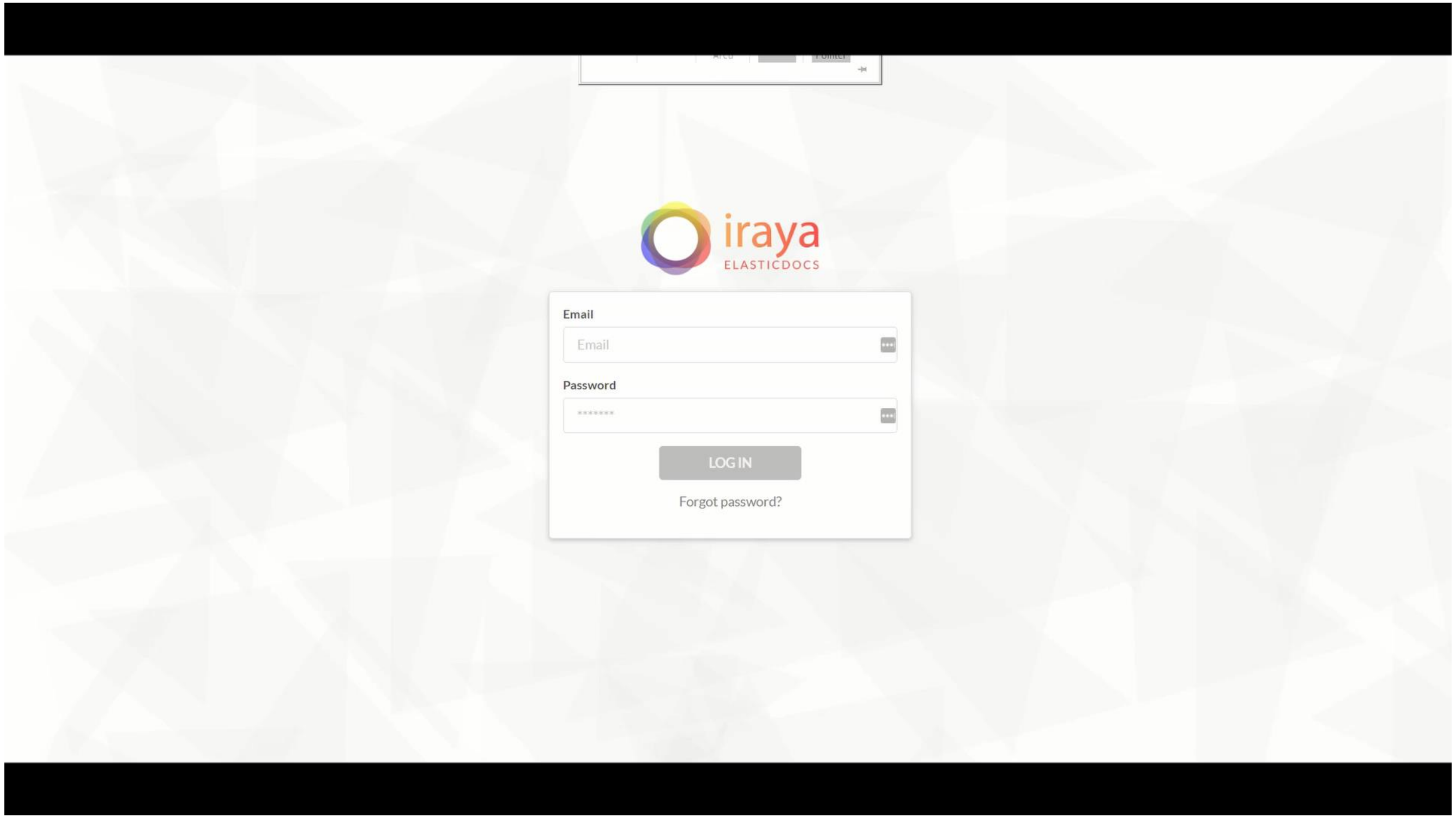
ElasticDocs

- ElasticSearch
- Geolocation
- Metadata extraction
- AutoImage Recognition

Supports geoscientists'

- Knowledge
- Intuition
- Experience

through accessible, verifiable big data



Conclusion

Dealing with huge amount of unstructured reservoir dataset is made more effective in ElasticDocs by:

- Curation and thorough investigation of appropriate machine learning algorithms
- Creating both structured and non-structured database to host and properly standardize reservoir data as input to machine learning algorithm
- Apply appropriate compute infrastructure, leveraging on availability of compute resources, either with on-prem or cloud
- Design a user-friendly API that all geoscientists can access and analyze their own data

There is huge potential in application of machine learning, and we are barely scratching the surface

- complex networks
- Increasing granularity in object identification