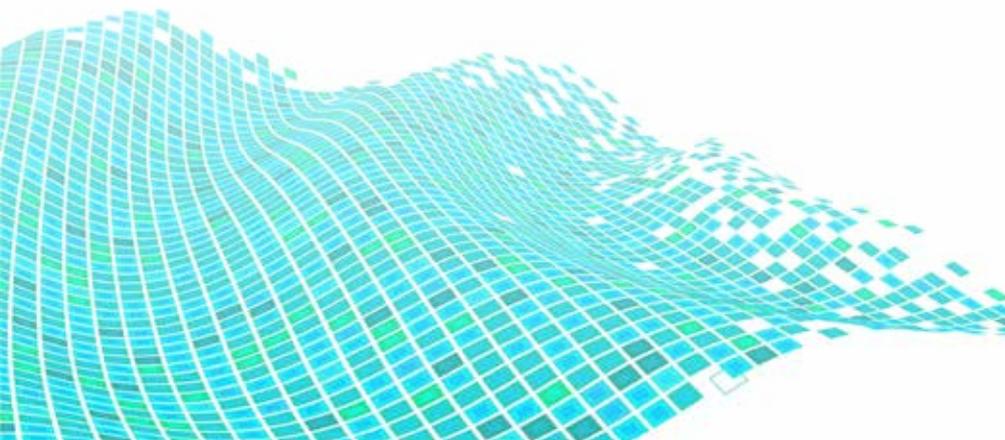


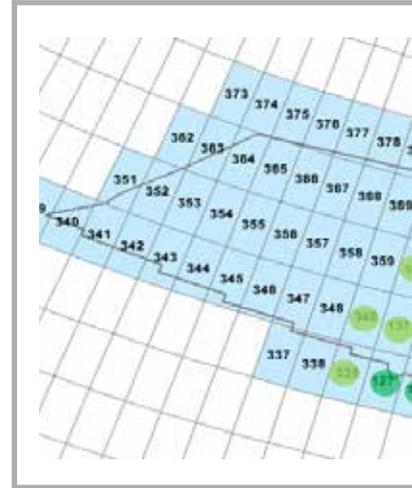


Structured information delivers reliable decisions

“80% of digital data is stored in unstructured files.
It is not classified and it cannot be searched efficiently.”



www.agileDD.com



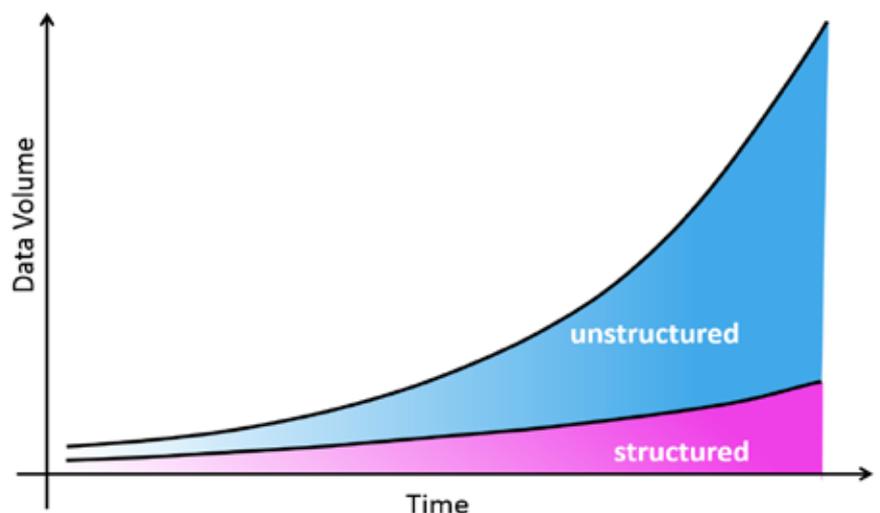
A breakthrough for data retrieval

For oil & gas professionals needing to put to work information stored in free-format documents, AgileDD's iQC solution automates the process to analyze and extract information from the electronic files and deliver an indexed and catalogued repository that can be used for searches and for analytics.

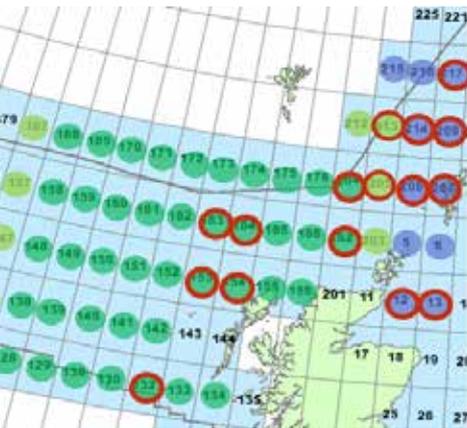
Contrary to earlier solutions, AgileDD's iQC keeps a live link to every source document, allowing users to verify that each data item is correctly transposed from the original.

Scale of the problem

Studies* indicate that between 70% and 85% of all data is unstructured. Information in these MS Office, PDF, JPEG or TIFF files is difficult to search effectively and cannot be used for analytics. These dormant data assets comprise business-critical data and metadata that will support better decision-making across many areas of activity in the company.



* Merrill Lynch: 80% of data is unstructured
GATE project / Sheffield University: 85% unstructured



Hundreds of thousands of unstructured documents related to borehole data were identified, and the relevant data items were extracted and fed into a structured data repository based on the clients's own taxonomy



The system has flagged low probabilities and some variability for the reading of the water depth variable on a set of scanned documents. The operator can immediately see these documents, with a highlight on the location from which the data was extracted, and validate or correct the reading. The logic of the Machine Learning System registers any such validations or refutations, which contributes to the improvement of the results for future analysis.

A user-driven interface

iQC offers a workflow-driven process that takes the user through the steps needed to:

- set up the data
- verify the integrity of the existing structured database
- train the Machine Learning System to index the unstructured documents and to extract targeted information
- verify and adjust the results to improve the MSL's performance

Results the way you want them

MSL results can be used to populate a project or master database, a data lake or to QC an existing database against the contents of unstructured documents

No orphan information

Every data item that is extracted and stored into the structured environment carries a probability and a variability parameter that reflects the degree of confidence that the data is the right one and that it has been correctly interpreted. If the probability is low, or there is abnormal variability for a parameter, the system can instantly bring up the original unstructured documents to verify the reading and correct it if needed. Each correction is registered by the MSL and contributes to the learning set.

Versatile deployment

iQC is designed to operate equally well on a local server or in a SaaS cloud-based environment. Designed to be used by subject matter experts who understand the data, the Machine Learning System does not require any particular knowledge of its inner logic, and is therefore easy to configure and update.

Data efficiency for a North Sea operator

Our customer had thousands of wells to run through analytics to gain a better grasp of the regional trends for formation characteristics. Each well had on average 100 unstructured documents associated to it: reports from the drilling and logging crews, progress reports, core and fluids analysis, interpretation reports. Their existing workflow involved the manual screening of the documents and the manual input of the relevant information into a hierarchical folder/file system. It could take up to 6 hours to handle one well.

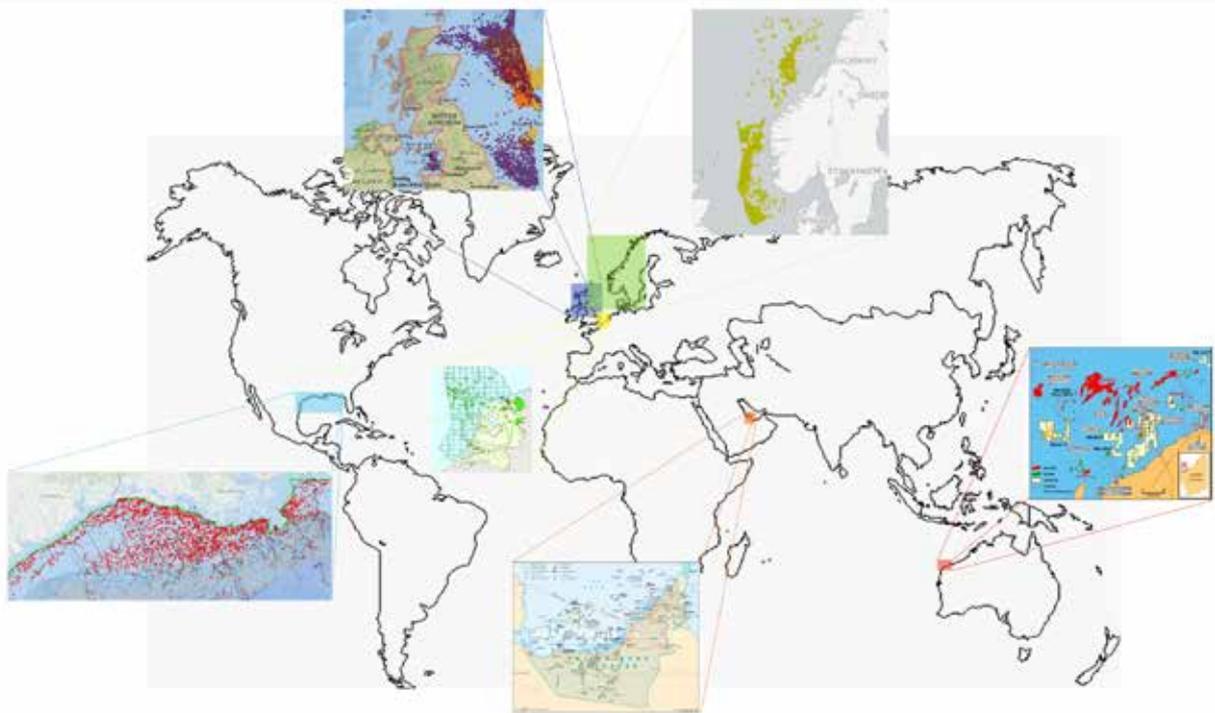
AgileDD selected a subset of the wells and their associated documents, and set the iQC Machine Learning System (MLS) to train itself on locating and validating the required information in each document. The process is iterative, with a domain specialist reviewing each pass and providing additional criteria for the MLS to improve its success score.

The training took about 2 days, and once finished iQC processed all the well data and

documents in a few hours. The result was a structured database where the logs had been enhanced with a rich set of metadata. Every retrieved data item has an associated probability attribute as well as a flag indicating variability in the readings. This informs the user as to the reliability of the information. The database is easily searchable using any combination of criteria. At any time the user can point back at the original document if he or she wishes to verify an entry.

The benefits to the customer were threefold:

- Important and highly valuable well information hitherto buried in unsearchable documents was now at the fingertips of the geoscientists and engineers
- The whole process had taken less time than would have traditionally been used to process a tenth of the data volume
- The new database was also cleared of any duplicate wells or replicated documents, improving the reliability of the analytics.



Areas where AgileDD has applied Machine Learning