WHITE PAPER

Creating value from National Data Repositories

Introduction

As the name suggests, National Data Repositories or NDRs are typically set up by government agencies and regulators as part of a long-term strategy to protect and optimize the value of a nation's natural resources. By collecting, standardizing and making data available quickly and efficiently they reduce barriers blocking investor entry, eliminate competitive positions based on data holdings and thereby maximize inward investment on exploration, development and production operations.

To achieve this, NDRs must create value for key data generators and users – the operating companies and the service providers. In many respects, the value of NDRs to operating companies, especially in territories that have implemented public data release policies, can be argued to be the data volumes they enable them to access. In a study commissioned by Common Data Access, "The business value case for data management – a study" (Hawtin and Lecore, 2011), it is stated that, since oil company value is directly related to an understanding of the subsurface, data contributes 25-33% of the value. Of course, unless the data can be efficiently accessed and integrated into a company's systems and processes and be considered of recognized quality, that value is diminished.

CGG & NDRs

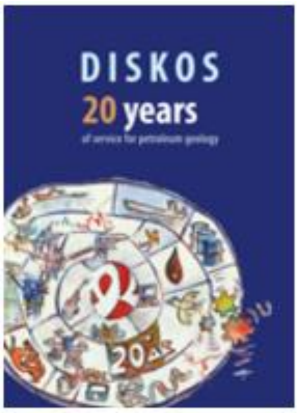
CGG Data Management Services has implemented and operated NDRs around the world since 2008. In 2013 CGG was awarded the contract to operate Diskos, Norway's National Data Repository. Widely regarded as the first true NDR and referenced as an example of international best practice, Diskos serves over 60 different companies and over 300 regular users. It celebrated its 20th anniversary of operation in 2015. A core team of CGG experts deliver the services together with staff from our project partners, EVRY and Kadme. The initial phase of the contract included migration of a proprietary data model format into the PPDM data model, the development of new functionalities, the design and implementation of the IT environment and the transfer of nearly 1 PB of data. Today the data volume exceeds 5 PB. Each month an average of 150 TB of new data is loaded. The 300+ Diskos users place over 1000 data orders every month, averaging 65 TB of downloads and distribution.

In this white paper we discuss current NDR best practice and share our experience on how best to initiate, implement, operate and create value from NDRs. We reveal how NDRs have developed to provide greater automation and new capabilities and consider how they must further evolve to meet the future needs of the industry.

Efficient access to trusted data comes at a cost. Although there are different business models for NDRs with many countries choosing to pay for their NDR from central funds, it is the operating companies that ultimately directly or indirectly pay for the service. In addition to paying for the NDR, the operating companies and their service suppliers also carry the cost burden of data submission. Additional copies of data, preparation to required standards, and administration efforts all contribute to this cost.

If the proposed range of 25-33% of company value derived directly from data is accepted, the outcome of any cost/benefit analysis performed on a planned NDR implementation or ongoing operation should be a foregone conclusion. This does not mean however that there is no need for a focus on efficiency. Opportunities for NDR operations to realize and deliver further efficiencies are required and need to be pursued. Modern NDR solutions should both reduce cost and save time for their stakeholders, whether that relates to the cost to the regulator to operate the environment or minimizing the burden on operating companies conducting their data submissions, for example. Additionally, NDRs must evolve to be more than a repository, delivering direct access to the data source. In the future, they should provide a feedstock of data for new and emerging analytics and machine learning capabilities whilst at the same time preserving information security.

“We have big expectations for the new partnership agreement with CGG. The goal is to build on 20 years of success and bring Diskos to the next level,” says Diskos manager Eric Toogood, and points out several focus areas for the upcoming years.”



***Diskos goes next level*, an article in Diskos 20 years of service for petroleum geology:**
http://www.npd.no/en/Publications/Presentations/DISKOS/Diskos-goes-next-level/

Evolution of NDRs

In many ways, the progression of NDRs parallels changes we have all experienced in consumer banking. Until comparatively recently, depositing or withdrawing money from a bank involved a visit to a “branch” and the performance of manual handling tasks involving physical items. In a relatively short time period this conventional way of doing business has been replaced by online banking with instructions and transactions carried out digitally.

NDRs are not so different. Submissions or "deposits" of data involved preparing file sets on media and shipping them to a physical location or "branch". Whilst downloads or "withdrawals" of data could be planned and defined online, typically the actual delivery of data required the physical shipment of media.

Such physical transfers of data take time, cost money and are inherently inefficient. When the operation of Diskos was tendered in 2013 the technical and commercial requirements indicated that an evolutionary jump was necessary, marking a transition to a new, highly automated, efficient and lower-cost solution. CGG delivered this, providing an NDR solution that has greater automation and new capabilities. This is what has enabled the significant monthly uploads of data and the rapid, accurate self-service fulfilment of over 99% of the approximately 1000 data orders made each month. With Diskos celebrating 20 years of service, CGG is proud to have enabled this next evolutionary phase.

Automation

When deciding on the services to automate, those accessed most frequently with the greatest urgency should take priority. Exactly what services does an NDR typically offer? Considering different geographies and NDRs the services can be grouped into the following broad categories:

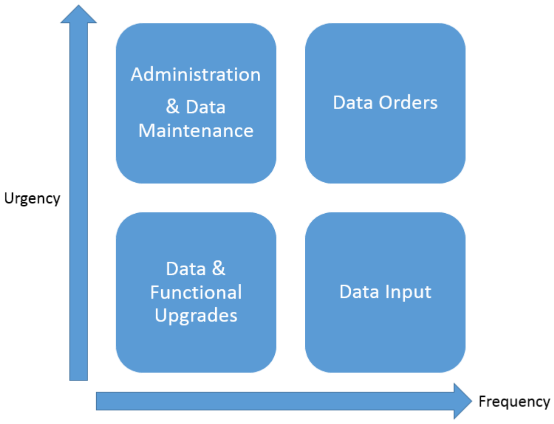
Data Input: The submission of original or reprocessed data or reporting to satisfy monthly or annual requirements, including validation and QC against defined standards.

Data Orders: Requests for data transfer to authorized parties.

Data Maintenance: Updating meta-data such as entitlement settings that control data access. Correcting, replacing and amending records.

Data & Functional Upgrades: Clean-up and data extraction projects and changes to functionality.

Administration: Information security routines, process improvement and risk management, quality monitoring, user support and updates to standards and guidelines.



Of these tasks, in terms of service demand, *data inputs* are typically the most frequent activity, closely followed by *data orders*. However, although Service Level Agreements (SLAs) may govern the rate at which data is to be loaded there is a significant business driver for fulfilling data orders quickly. For this reason, we will look at the automation of data orders first.

Data Orders

The ability for a user to search, browse and view, both textually and on a map, the data they are entitled to is a pre-requisite. Invariably, this functionality is delivered over a web interface. Automation beyond this point occurs in two ways:

1) Delivery of the data order
2) Access to the data via an Application Programming Interface (API)

A data order actually consists of a user-defined set of data that needs to be located and copied. Since the user defines the order on the web interface it consists of meta-data describing the data. To fulfil the order it is necessary to use the meta-data to locate, extract, copy and send the data to the user. Where the user has defined part of a seismic volume, the SEG-Y data (pre- or post-stack) for example, there is also a need to 'cut' the data to the spatial limits defined in the order.

CGG's Akon system automates this entire process with an API between the web interface and the NDR 'back-end' managing the process of locating, cutting and copying the data. Since the web interface only allows users to see and order data to which they are entitled, access rights are also honored. To complete the order, the extracted data volume or order is either placed onto an FTP/SFTP site or copied to media and dispatched, according to client instructions.

In addition to the conventional approach of browsing and ordering data via a web interface, CGG's system can be 'exposed' to user applications via an API. This is fully secured and honors data entitlement. The API enables users to integrate details of the data they are entitled to into their own desktop applications. For example, if they have built a GIS application that displays their corporate data availability, they can add a layer that shows data available from the NDR, providing a complete picture of data coverage. It is also possible to construct and apply regular data downloads for well data. This can be useful when an operating company has a particular group of wells of interest and wishes to maintain local file storage of all relevant data for those wells. Whilst it is easy to create an order for all of this data, as new data is being loaded all the time, it quickly becomes necessary to update and submit new data orders to gain access to newly loaded data. Using the API, this update process can be fully automated, ensuring that the latest and most complete data coverage is always available to the company geoscientists.

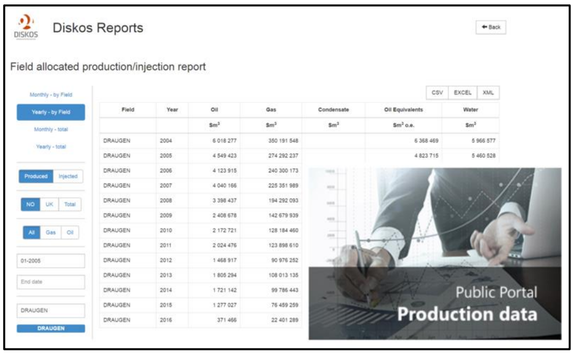
Data Input

The ease of performing loading or data input to an NDR can be highly variable. In a perfect world, where there are clear submission guidelines, each contractor and company involved in data acquisition and supply uses industry-standard file formats and they are consistently and accurately applied, loading would be a straightforward process. However, the gathering and supply of the data that must eventually be submitted to an NDR involves a multitude of stakeholders acting over a long period of time. Consider also that requirements and values that should be populated in header records do sometimes change over time and this can therefore lead to genuine data inconsistencies.

This means that, even in mature provinces with well-established NDRs, data preparation for loading remains a significant task. There are, however, various aspects of this process that can be automated.

For simpler data types that do not include associated files, such as consolidated production data reporting, it is possible to use industry data exchange protocols. For the Diskos operation in Norway the Norwegian Petroleum Directorate participated in *an NDR community initiative* to create a derivative of the Production Reporting Markup Language (PRODML) from Energistics. This was termed Monthly Production Reporting Mark-up Language (MPRML). Using this as the target format, operating companies extract the relevant data from their internal production reporting systems using scripts. Each operating company creates a single XML file using the MPRML format that is automatically submitted. The contents of the file are validated before being published to the NDR. The submission and publication of monthly reporting data are almost entirely automatic for both the supplier and recipient of the data.

Certain data types such as well and seismic are more complex, consisting of many different files and file types. Full automation of the loading of this type of data is possible, but is not yet desirable as it would inevitably lead to incomplete or inconsistent data.



Diskos Reports

Field allocated production/injection report

Field	Year	Oil	Gas	Condensate	Oil Equivalents	Water
		Sm^3	Sm^3	Sm^3	Sm^3 o.e.	Sm^3
DRAUGEN	2004	6 018 277	350 191 548		6 368 469	5 966 577
DRAUGEN	2005	4 549 423	274 292 237		4 823 715	5 460 528
DRAUGEN	2006	4 123 915	240 300 173			
DRAUGEN	2007	4 040 166	225 351 989			
DRAUGEN	2008	3 398 437	194 292 093			
DRAUGEN	2009	2 408 678	142 679 939			
DRAUGEN	2010	2 172 721	128 184 460			
DRAUGEN	2011	2 024 476	123 898 610			
DRAUGEN	2012	1 468 917	90 976 252			
DRAUGEN	2013	1 805 294	108 013 135			
DRAUGEN	2014	1 721 142	99 786 443			
DRAUGEN	2015	1 277 027	76 459 259			
DRAUGEN	2016	371 466	22 401 289			

Production data gathered using MPRML as seen on the Production Data Public Portal:

http://www.diskos.no/public-portal

Collaboration Works

MPRML is itself the result of collaborative work coordinated by Energistics. This global, not-for-profit industry consortium serves as an independent body to facilitate and manage open data, information and process standards for the upstream oil and gas industry.

The concept of a standardized production data reporting structure that would enable regulators to meet government objectives was conceived at the National Data Repository conference held in Rio de Janeiro, Brazil in 2010. A work group was formed and six different countries contributed alongside several service companies. The results of the work were presented at NDR11, held in Kuala Lumpur, Malaysia, and the MPRML, a derivative of PRODML, was completed shortly afterwards. MPRML is proof that collaboration works.

However, aspects of the quality control process have been automated. For example, industry-standard file types that include header records, such as LAS and SEG-Y, are automatically validated. This takes the form of rule-based checks for each attribute in the header. As an illustration, the spud date of a well can be validated to be consistent with the data held in an external master repository for wells – such as the regulator's well licensing database. It can also be validated as being in the correct format according to the submission standards and to be present if it is a mandatory field.

Integrate to Validate

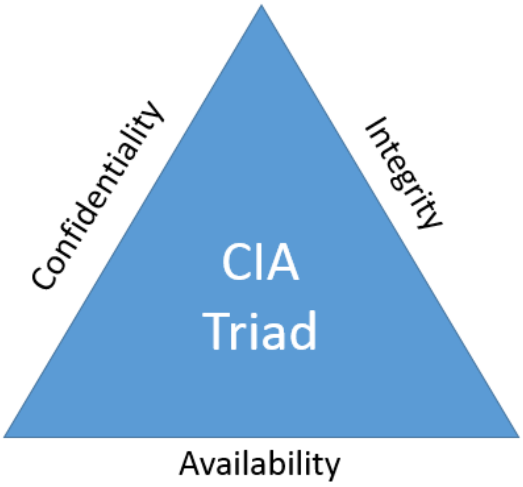
Integrating the NDR with other associated repositories, such as the regulatory body's licensing or permitting systems, enables validation and consistency of data. It can also provide a 'warning' or planning aid to highlight data that should be submitted. In turn, this enables control over completeness, helping to ensure that all required data are gathered and reported in a timely manner.

By building an extensive library of validation rules it is possible to fully automate the QC and verification of these files, such that a file that passes validation can then be loaded to the system ready for entitlements to be set. Further efficiency is gained by exposing these validation tools to the operating companies and reporting back to them on the rules that have not been met. In this way, the tools are used to validate the files at the point they are received from their contractors. This enables the user to be aware of any specific failure and have them corrected at source.

New Capabilities

Automation is not the only enhancement to have been delivered to the Diskos NDR solution by CGG. In addition to extended capabilities and new functionalities, such as completeness control, CGG was the first NDR provider required to deliver a solution compliant with the ISO 27001 information security standard.

Information security covers not only confidentiality of data but also its availability and integrity.



The information security 'triangle'

This has far-reaching consequences for the design and operation of an NDR, especially when accessibility and ease of use must be balanced with security requirements. Information security is similar to data management in that it relies equally on people, process and technology.

Technologically, security requirements dictate and inform the hardware, communications and application design. For example, the data itself must be completely separate from the user front-end; 'data segregation' must be evident as a control. The user therefore accesses and uses only meta-data. Any access (ordering) of that data creates a request that is made on the data servers via an API.

Processes must be designed, documented, followed and updated to ensure the integrity and confidentiality of the data. Setting entitlement to data is a good example. No matter how secure and well designed the solution is, if an entitlement is set incorrectly someone could gain access to data to which they have no rights. To control and prevent this risk, it is necessary to design a safe process. This needs to be holistic in that it considers how the application will guide and assist the user, for example by asking the operator to confirm the selection they request before applying it, and how fail-safes and checks can be applied. In the case of entitlement setting, it is such a critical process to complete correctly that CGG determined that, to adequately mitigate the risk, one system operator should input the data, while a second must check their work before the change can be applied; 'segregation of duties' must also be evident as a control.

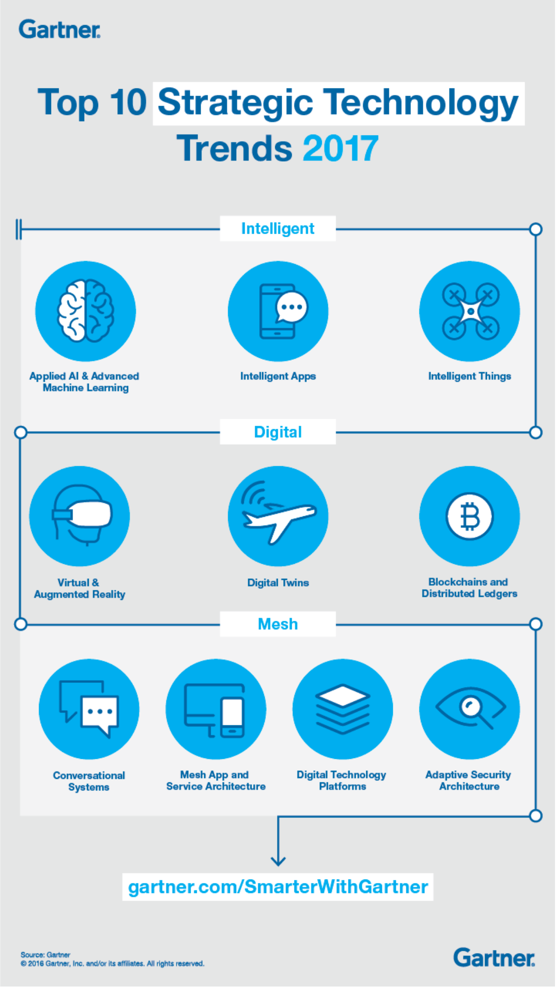
People are also at the center of security, since it is almost always a human action that creates an information security incident. To meet the criteria for ISO 27001 all staff need to be capable and competent in their role. They need to be trained to understand the implications of their actions and to understand information security requirements. Above all, they must be encouraged to be open and honest about any incidents and weaknesses they experience, to report them and to be engaged in investigations and improvement actions.

Data Management Competence

Irrespective of automation, people remain at the center of an NDR operation. The accuracy, integrity and security of the data rely on their competence. How do you measure oil & gas data management competence and staff progression? The answer can be found at www.cdacompetency.com. This data management competency system (available free of charge) lists different aspects of the profession and describes five different levels of competence for each. Competency maps can be created for an individual or for a role. This is a very useful tool for planning career development or for ensuring that staff have the right competencies for the job. Alternatively, staff may wish to become a certified petroleum data manager. Find out how under the Certification tab at www.ppdm.org

Ready for the Future

As an industry we are experiencing an unprecedented time of change driven by a combination of low oil prices and emerging technologies. The changes are exciting as new technologies and techniques deliver new possibilities.



Gartner

Top 10 Strategic Technology Trends 2017

Intelligent

Applied AI & Advanced Machine Learning

Intelligent Apps

Intelligent Things

Digital

Virtual & Augmented Reality

Digital Twins

Blockchains and Distributed Ledgers

Mesh

Conversational Systems

Mesh App and Service Architecture

Digital Technology Platforms

Adaptive Security Architecture

gartner.com/SmarterWithGartner

Source: Gartner
© 2016 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Gartner's Top 10 Strategic Technology Trends for 2017: http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/

As data managers of national or corporate solutions and environments, we are responsible for evaluating and understanding the changes and predicting and preparing for the next data management challenges.

NDRs often represent a very significant collection of data gathered over a long period of time. Emerging technologies hold the promise of extracting and organising a greater amount and variety of such data.

Analytics & Artificial Intelligence

Predictive analytics and artificial intelligence (AI) are already being used in various ways in the oil & gas industry. Whilst experimental R&D projects are underway in the exploration arena, both techniques are actively being used in projects for producing assets to predict and optimize facilities maintenance and production.

Each of these projects requires 'base-data' to be supplied as complete, accurate and up-to-date information, in a timely and cost-effective manner, and in an acceptable format. This is the challenge faced by data managers. To meet it, the data stored in national and corporate data repositories therefore needs to meet these requirements whilst also remaining accessible and transferable in open-standard formats. CGG's NDR solution meets this challenge via robust processes and automation to ensure accuracy and completeness. Accessibility and transferability of data is ensured by making use of the open and public PPDM data model from the Professional Petroleum Data Management Association. Files are also stored in standard formats, such as LAS and SEG standards, with consistent nomenclature and use of headers.

Increasingly, access to data needs to be faster and more efficient. As already described, CGG is making use of an Energistics exchange standard for the submission of monthly production data. In the future, NDR solutions will need to be capable of exchanging a greater variety of subsurface data in exchange standards such as WITSML and RESQML.

Machine Learning

Machine learning aims to replicate how we as humans think and learn. Traditional computer programming is more deterministic, meaning that the inputs and desired outputs have to be known. Non-deterministic programming enables different outcomes depending on 'choice points', but the programmer must be aware of all such choices at the outset. For many routine tasks in data management, variation of the input data, in terms of documents and formats, is so great that it has proved impossible to create programmes that can reliably complete the task. Machine learning offers a chance to automate what are currently manual tasks, such as locating and transferring core analysis data within well files into a corporate or national data repository.

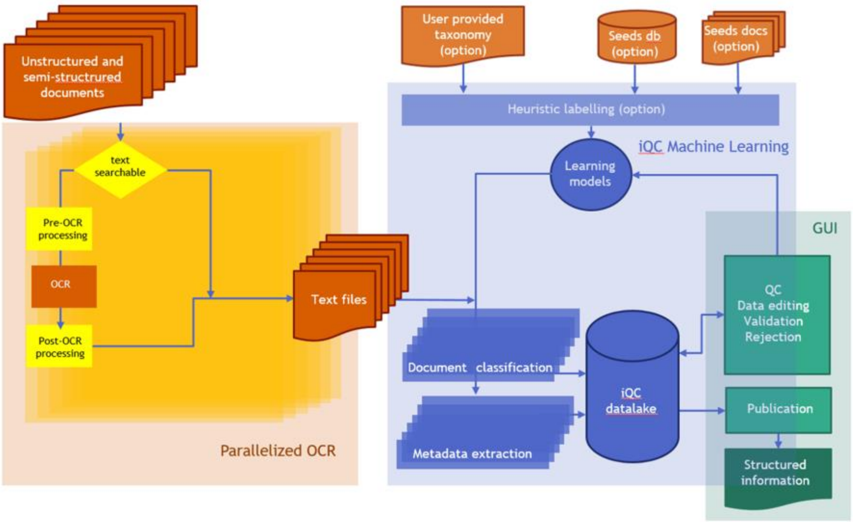
Most national data repositories have been successful in gathering critical well information. However, much of the data particularly related to wells is presented in unstructured document collections or semi-structured files. In these formats, data such as tops, pressures and bottom hole temperatures need significant 'preparation' work before use as the 'base-data' for analytics.

Machine learning promises to transform the speed and cost of accessing this data by changing it from a manual task to one that is largely machine-driven (Blinston and Blondelle, 2017).

The data manager's challenge is to assess the potential of these emerging technologies, consider the data present within the repositories they manage and build the business case for unlocking the full range of data.

Pre-stack Interpretation

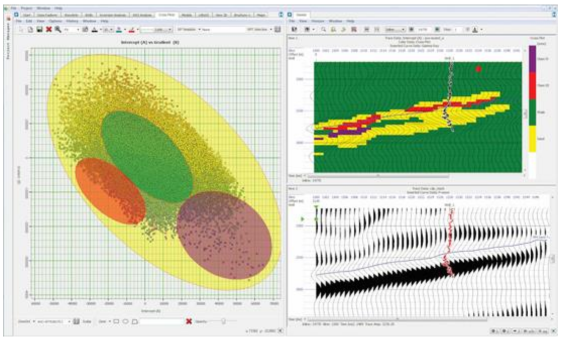
Until recently, most seismic interpretation and analysis was carried out on post-stack seismic volumes. Pre-stack volumes were often an intermediary step produced and QCed during processing to enable the



A typical machine learning workflow (from Blinston and Blondelle, 2017).

production of relevant and useful post-stack volumes.

As seismic acquisition technologies, compute power and our understanding of the seismic response have progressed, it has become more common to conduct interpretation and analysis on pre-stack volumes, for example to analyze rock properties using Amplitude Versus Offset (AVO). This has an impact on the data manager as pre-stack volumes are orders of magnitude larger than the equivalent post-stack volumes. With advances in storage technologies there are many viable Hierarchical Storage Management (HSM) and private/public "cloud" storage options. Storage of larger volumes is not therefore the issue. However, fulfilling multiple data order requests for data volumes into the 10 TBs or even 100 TBs currently requires physical shipment of media.



Pre-stack data conditioning, attribute computation and analysis using CGG's HampsonRussell AVO module. http://www.cgg.com/en/What-We-Do/GeoSoftware/AVO/AVO

Where high-bandwidth communication networks are available, CGG overcomes this problem by enabling replication of 'extracted' volumes direct to the user's local file storage. However, even this is inefficient as the large data volumes are being stored twice. In proprietary applications this is solved by giving direct access to the data and to applications used to visualize, analyze and interpret it. The NDR of the future requires a similar solution, perhaps co-locating interpretation and analytics applications with the data. In this vision of the future the user will not place an order for data retrieval. Instead, that 'order' will request access to the data and the chosen application.

Conclusions

The business case for NDRs is as strong, if not stronger, than it ever was. This does not, however, guarantee acceptance and adoption by all stakeholders. To be truly successful, these solutions must become an intrinsic part of the wider community's data management solution and to do this they must deliver their services more efficiently and effectively than could be achieved by one party alone.

This is CGG's driver to stay at the forefront of NDR evolution and automation. But we are not content to stop there. We see an exciting transformation taking place in our industry and we are preparing our NDR solutions for this bright future.

Glossary of terms

API (Application Programming Interface): Functionality within a computer application that can be 'exposed' to other applications to enable data and commands to be passed "machine to machine" from one application to another.

HSM (Hierarchical Storage Management): Hardware that enables seamless storage of files and data across different types of disk and/or media such as tape. Commonly combines disk and tapes moved between storage slots and tape drives by a robotic system.

LAS (Log ASCII Standard): Standard file format used to store well log information.

MPRML (Monthly Production Reporting Markup Language): XML exchange standard developed as a derivative of PRODML used to exchange monthly production figures.

PPDM (The Professional Petroleum Data Management association): Frequently used to refer to the open, public data model developed and maintained by the association.

PRODML: XML exchange standard developed and maintained by the Energistics Industry Consortium covering production data.

RESQML: XML exchange standard developed and maintained by the Energistics Industry Consortium covering data relating to the reservoir.

SEG-Y: File standard developed and maintained by the Society of Exploration Geophysicists designed to enable the transfer of seismic data.

SLA (Service Level Agreement): A specified rate, time or other measure for a given service type, commonly contained within a contract.

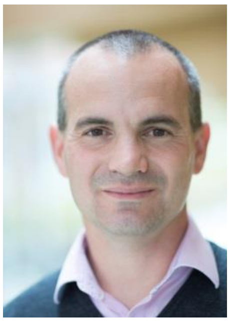
WITSML: XML exchange standard developed and maintained by the Energistics Industry Consortium covering well, drilling and log data.

References

Blinston, K. and Blondelle, H., 2017, Machine Learning systems open up access to large volumes of valuable information lying dormant in unstructured documents. The Leading Edge, **36**, no.3, 257-261.

Hawtin, S. and Lecore, D., 2011, The business value case for data management – A study: Common Data Access Limited, http://cdal.com/wp-content/uploads/2015/09/Data-Management-Value-Study-Final-Report.pdf

About the authors



Kerry Blinston is the Global Commercial Director for CGG Data Management Services. He has worked for CGG for nine years and been actively engaged within national and corporate data repository projects over that time, working in countries such as Sri Lanka, Azerbaijan and Ethiopia. Currently he is the contract holder for CGG's operation of Norway's NDR, Diskos. He holds a BSc in Geological Sciences from Leeds University and an MSc in Petroleum Geology from the University of Aberdeen.

Karen Blohm is currently the Diskos Programme Manager for CGG Data Management Services. Karen fulfils an operational leadership role with strategic and high-level data management engagements in National/Governmental Authorities and oil & gas companies. She has had technical roles in subsurface interpretation and petroleum geoscience followed by technical, consultative and operational leadership roles in data management. Data management engagements have provided opportunities around the globe, including in Norway and Australia, and have focused on scoping and delivery of major data management and service delivery improvement programs. Karen holds a BSc. in Geology from the University of Hull, UK.

About Data Management Services

Data Management Services (www.cgg.com/dms) are part of CGG GeoConsulting and offer a full range of services and solutions to answer E&P data challenges, from physical storage and transcription to the implementation of corporate and national data repositories. They were awarded the Diskos NDR contract by the Norwegian Petroleum Directorate in 2013.

About CGG

CGG (www.cgg.com) is a fully integrated Geoscience company providing leading geological, geophysical and reservoir capabilities to its broad base of customers primarily from the global oil and gas industry. Through its three complementary businesses of Equipment, Acquisition and Geology, Geophysics & Reservoir (GGR), CGG brings value across all aspects of natural resource exploration and exploitation.

CGG employs around 5,600 people around the world, all with a Passion for Geoscience and working together to deliver the best solutions to its customers.

CGG is listed on the Euronext Paris SA (ISIN: 0013181864) and the New York Stock Exchange (in the form of American Depositary Shares. NYSE: CGG).