EnergyIQ White Paper

Data Objects as a Foundation for E&P Data Management

March, 2016

Contents

1. Introduction

More than ever, E&P companies are relying upon information to gain insight and make the critical business decisions that drive operational effectiveness and gain a competitive advantage. They need to know that they are making decisions based upon the best information available, and they require full visibility to the quality and history of that information.

To be able to achieve this level of information certainty, organizations are increasingly viewing data management as a core competence. Effective E&P data management involves the integration of multiple data sources across the Well Lifecycle, and then making sure that the business has access to the best version of this data when and where it is needed.

There are many aspects to effective data management, with one of the key foundations being provided through the practical implementation of Data Objects. Data Objects group attributes that define a recognized entity such as a Wellbore into a single unit, and establish the rules and standards required to manage that entity. Having a standard set of attributes and rules defined for a Well Origin, Wellbore and Wellbore Completion, for example, greatly facilitates the validation, management, exchange, and visualization of information across the well lifecycle. Data Objects are a practical extension to accepted industry standards such as the PPDM 'What is a Well?'.

From an implementation perspective, Data Objects provide a level of abstraction between the underlying storage and transport mechanisms that enable well lifecycle integration and long term sustainability. Handling small, self-contained packages of information, provides support for event-based data exchange to keep applications synchronized across the Well Lifecycle in real-time, and to ensure that the business has access to the data when and where they need it.

This white paper discusses the role of Data Objects as a key component in establishing a platform for effective E&P data management. The paper includes a section on the definition and benefits of Data Objects together with a section that discusses one example of the practical implementation of Data Objects.

Note that frequent reference is made to the Well Hierarchy and the associated components within this paper. This important concept is not the focus of this paper; instead, the reader is referenced to the following: PNEC 2015: White Paper – The Well Hierarchy as a Foundation for MDM: A Case Study.

2. What is a Data Object?

This section describes the components that make up a Data Object and how it can be applied to support the Well Hierarchy as defined through the PPDM 'What is a Well?' initiative.

2.1. Data Object Definition

A Data Object is defined as a collection of data attributes combined with the information required to manage the object to support business workflows; see **Figure 2.1**. In some quarters, a distinction is made between a Data Object and a Business Object where a Data Object is primarily comprised of just the attribute definitions, while a Business Object includes all of the rules for managing the object. For the purposes of this discussion, we will refer to the combination as a Data Object.

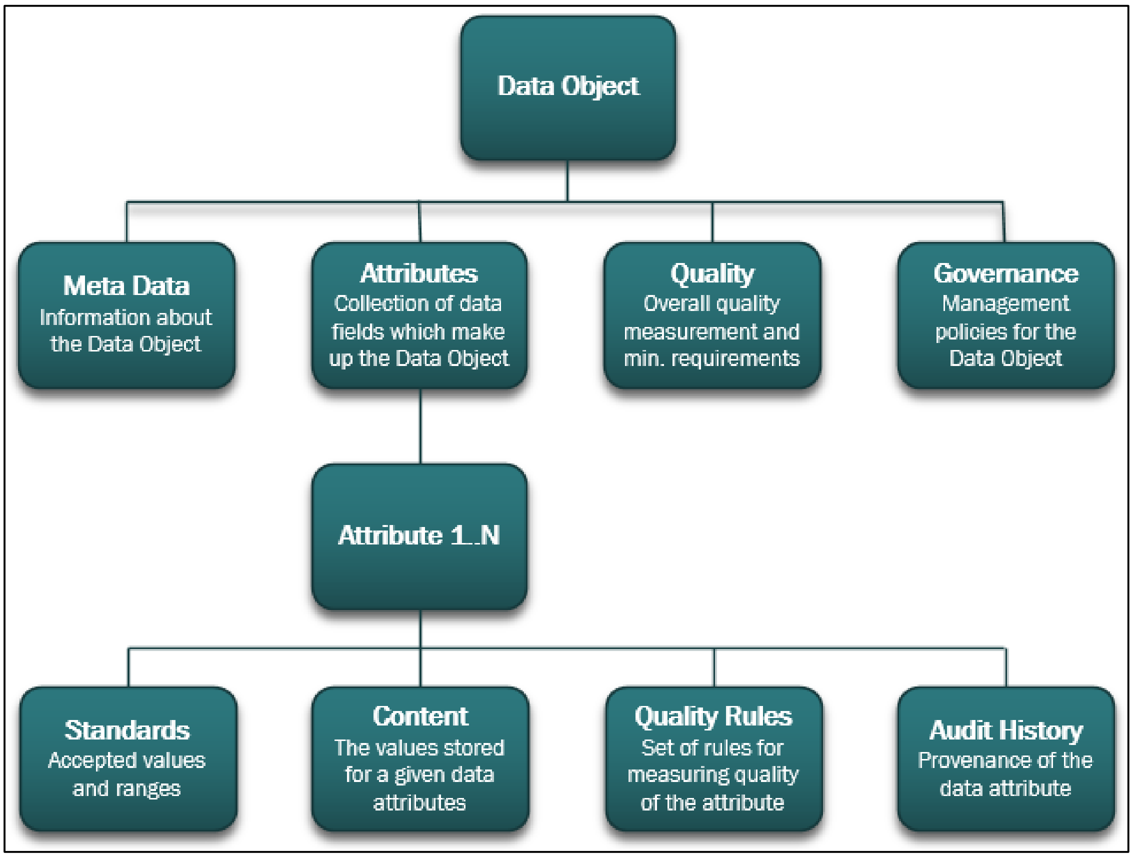

Figure 2.1: Data Object Definition

Meta Data: This contains information primarily about the content of the Data Object including when it was created, what version of the standard it represents or extensions to the standard definition.

Attributes: These tend to get the most attention especially as it relates to those objects that form part of the Well Hierarchy such as Well Origin, Wellbore, and Wellbore Completion.

The attributes have a defined data type that is implementation agnostic, and a meaningful description of the intent of the content. An example of the definition of a subset of attributes within the Wellbore Data Object is illustrated in **Figure 2.2**.

***WellboreDO** - path of drilled footage from the Well Origin (top/start) to a terminating point (bottom/end)*		
Attribute	**Type**	**Description**
Uid	string	Unique identifier of this data object (system-generated)
UidParent	string	Unique identifier of the associated (parent) Well or Wellbore object
Source	string	Source (individual, company or government agency) of this data object
GovernmentID	string	Identifier assigned to the Wellbore by a regulatory agency (e.g., API)
WellboreName	string	Name of this Wellbore
BHSpatialLocation	SpatialLocationDO	Bottom Hole spatial location for the Wellbore (includes latitude, longitude and CRS)
BHLegalLocation	LegalLocationDO	Bottom Hole legal location for the Wellbore (e.g., Congressional, Texas, Offshore, etc)
LegalSurveyType	string	Survey type of the legal description for this Wellbore (e.g., CONGRESS, TEXAS, OFFSHORE)
CurrentClassification	ReferenceDO	Current classification of the Wellbore
CurrentStatus	ReferenceDO	Status of this Wellbore (e.g., drilling, completing, producing)
FluidType	ReferenceDO	Type of fluid associated with this Wellbore (e.g., Gas, Oil, Condensate)
ProfileType	string	Shape or profile of this Wellbore (e.g., vertical, horizontal, directional)
WellboreType	string	Type of Wellbore (e.g., sidetrack, bypass, pilot)
DepthDatum	ReferenceDO	The point from which depths are measured (e.g., KB, RT, DF, CF)
DepthDatumElev	float	Elevation associated with the DepthDatum value
CasingFlangeElev	float	Elevation of the casing flange
DerrickFloorElev	float	Elevation of the derrick floor
KellyBushingElev	float	Elevation of the kelly bushing (generally from sea level)

Figure 2.2: Sample Wellbore Attribute Definitions

Quality: From a data quality perspective, we need to be able to assess the quality of Data Objects as a whole derived from rules applied to the individual attributes. This quality measurement at the Data Object level will fall into a number of different categories (facets) that could include, but are not limited to accuracy, timeliness, completeness and standards. If Data Objects are to be exchanged between applications and data stores that are independent of one another, then standards will need to be defined to establish a common interpretation of the quality results.

Governance: This component includes the business rules that are used to manage the Object. Examples of these could include the rules for creating a new entity, as well as updating or deleting an existing entity.

The following definitions apply to the individual attributes within the Data Object:

Standards: In many cases, an attribute value will be part of a well-defined subset or fall within an acceptable range. The list of standard values or accepted ranges should be provided within the Data Object definition. Note that this may not be passed as part of the data object exchange, but used for reference purposes in the definition or implemented as part of the Quality Rules.

Content: The actual content of the Data Object is delivered with the Data Object and must conform to the defined data types and standards definitions. For all Data Objects there will be a minimum set of attributes required to be populated for completeness. These attributes will be part of the definition.

Quality Rules: The data quality rules that were used to validate the data must be passed along with the Data Object for the results to be meaningful. As noted previously, if standard sets of quality rules can be established then it may not be necessary to pass the individual rules.

Audit History: The history of changes to the contents of the Data Object should be available to the calling application. It is important to the consumer of the data to understand where it came from and how it has been managed.

2.2. Attribute Definition

There is always a temptation when defining which attributes belong with each Data Object to allow the contents to be driven by the practical implementation. Nowhere is this more evident than in the desire to include links to other Data Objects within the existing object. For example, within the Wellbore DO, should a link to the parent be included along with links to children? Should the child links only be to those objects included in the Well Hierarchy or should it also include objects outside of the Well Hierarchy such as logs, directional surveys or tests? This is further complicated by the fact that it is incorrect to assume that the Well Hierarchy is fixed in nature. For example, a Wellbore Completion would typically be the child of a Wellbore but there are occasions where it must be tied to a Well Origin due to a lack of available information. As drilling technology continues to evolve at a rapid rate, it is likely that additional components will be required to extend the Well Hierarchy such as the Wellbore Segment.

It is the opinion of the author that links to other Data Objects, while useful from an implementation perspective, add too much rigidity to the definition, and should be minimized at the expense of requests for additional information from the Data Object producer. It is valid to include the Well Hierarchy parent of the object, and possibly direct children within the Data Object. Everything else should be retrieved via a call back to the producer of the Data Object in question.

2.3. Data Object Flexibility

The challenge with defining standard Data Objects (or any format) is to not include everything so the result becomes complicated and unwieldy. The intent should be to include all of the attributes that would typically be associated with a Data Object and then provide a well-structured mechanism for extensions that enable both producing and consuming applications to include data that is outside the standard definition.

3. The Benefits of Data Objects

This section discusses some of the benefits of Data Objects within an E&P data management solution.

3.1. Support for Standards

The 'What is a Well?' standard has gained a great deal of acceptance in the industry over recent years, as it provides a baseline for data integration across the enterprise. The full benefit of this standard begins to be realized, however, when it can be practically implemented within a data management solution. EnergyIQ's Well Hierarchy provides the platform for this implementation by aggregating multiple data sources to derive the key Well components such as Well Origin, Wellbore, and Wellbore Completion, for example.

An obvious evolution of the Well Hierarchy is to establish which attributes belong at which level of the Well Hierarchy. It is considered that this should become an extension to the 'What is a Well?' standards definition. Having a clearly defined set of attributes for each level of the Well Hierarchy will greatly facilitate the management, exchange, and interpretation of data.

These attributes would be created within the respective Data Objects and, having an industry accepted set, would greatly facilitate future data exchange. Consequently, EnergyIQ is participating in a consortium of application and data vendors to create an initial set of Data Objects that can then be adopted and further developed by one of the standards organizations.

3.2. Well Lifecycle Data Exchange

The primary benefit of Data Objects is considered to be in the area of data exchange. Creating a packet of information containing a well-defined collection of attributes with supporting information, allows complying applications to rapidly process the data with confidence based upon the standards definitions, quality results and audit history. Data Objects can be delivered as part of a batch process or individually as part of a real-time data exchange platform.

E&P data management involves large volumes of information, and these volumes are rapidly increasing, driven by technological advances. The ability to transfer large numbers of smaller packets of information, based upon modern technology, delivers greater flexibility, reliability and performance than traditional methods. We see this type of approach successfully implemented in the financial industry, as an example, along with social networking environments.

In addition, the Data Object exchange approach provides a foundation for real-time data exchange as part of an event management solution. In this case, a central hub or processor is notified of new data or data changes, and determines what application or process needs to be notified of that change. It can then request the Data Object or Objects related to that event and pass the objects to the consuming applications. In this way, data can be kept synchronized across the Well Lifecycle in real-time, with full visibility to the extent and quality of the data transfers. This is the ultimate goal of a Well Lifecycle based data management solution where users have access to the best data available--all the time.

3.3. Validation and Governance

The design of Data Objects allows for the validation of the content with reference to the other attributes within that same Data Object so that the information can be validated to a known level of quality. It may be that the consuming application performs additional validation against the target data store or some other data store, but the consumer can be confident that quality rules have been applied within the context of the Data Object itself.

The overall quality measurement of the Data Object along with the individual results, should be delivered to the consuming application to build trust in the data. It is important that the individual rules used to measure quality are delivered with the Data Object or at least referenced in an unambiguous way.

One key aspect of the quality process is to deliver the contents of the Data Object in the required coordinate reference system, units of measure, and master reference values. The source values of these components should be available to the consuming application, if required.

From a governance perspective, it is important to know the history or provenance of data, and not just the current quality. This should answer questions such as where did the data originate from, what changes were made to the data, who made those changes, and why were the changes made. Consuming applications may handle data differently based upon this information, and it is another significant component towards building trust. Additionally, providing business rules for the management of the Data Object establishes consistency between both consuming and producing applications.

3.4. Abstraction

Using Data Objects for the movement of data establishes an important layer of abstraction between the consuming applications and the physical model of the producer. This minimizes interpretation errors by hiding the complexity of the model (particularly relevant in the case of PPDM,) and protects the consuming applications from changes to the underlying model over time.

4. Data Object Implementation

It is important that the Data Object definition is not reliant on any specific technology. A Data Object should be capable of being created and transferred using any appropriate technical platform, as this will encourage adoption and protection in the long term, as new and better solutions for data movement become available.

This section discusses one practical implementation of Data Objects and a transfer mechanism developed by EnergyIQ through the ActiveExchange platform.

4.1. Creating Data Objects

Data Object *definitions* should be independent from any particular technology platform, but utilization of Data Objects within a data exchange solution requires the *implementation* of a particular technology and data format. In order to achieve wider adoption across both applications and organizations, Data Object definitions should be supported by a variety of language-specific bindings. The EnergyIQ ActiveExchange platform, discussed in more detail in **4.2**, leverages a Java language binding, which supports the ability to pack and unpack data objects as both JSON and XML. Other applications may require a C# language binding or perhaps a JavaScript language binding. In the future, there may be a need to support binary formats such as AVRO. These options are all implementation details, and remain separate from the definition of the Data Objects.

4.2. ActiveExchange

An E&P company's data management ecosystem continuously generates Well-related events, such as "Well Spudded", "AFE Created" and "New Wellbore Added". These events are typically associated with a particular Data Object. The ActiveExchange solution is an 'event management platform' that relies heavily on having a common understanding of Data Objects in order for different types of events to be processed, and acted upon. For example, the creation of a new Well in a Planning System such as enersight™ may spawn an event. This event is captured and, based on the company-specific data flow, the associated Well Data Object is retrieved from enersight and delivered to a Well Master solution, such as EnergyIQ's Trusted Data Manager™ (TDM). In another example, a user may express interest in new Wellbore Data Objects that fall within a particular area of interest that matches up with the extents of their interpretation project. When a new Wellbore is added to the Well Master system, if it matches the user's criteria, then an event is generated which results in the retrieval and subsequent routing of that Data Object to the user's specified project.

In all cases, the retrieval, routing and delivery of data is done via the same Data Object definition, and in this case, also the same Data Object technology and format implementation. What is important to understand is that the normalization of Data Object definitions is what allows solutions, such as ActiveExchange, to easily support a variety of consuming and producing applications, as well as employ a number of different types of Data Object services in order to perform transformation, validation, and calculation routines on Data Objects as they are routed throughout the platform. A depiction of how this type of orchestration occurs is shown below in **Figure 4.2**.

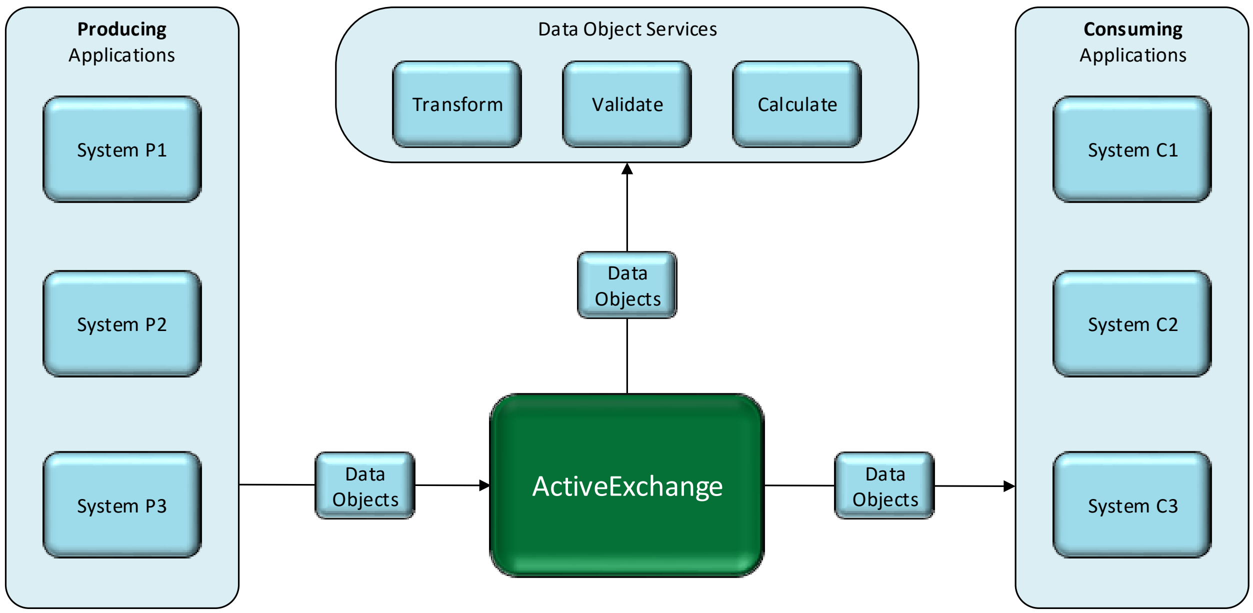

Figure 4.2: Producing, Routing, and Consumption of Data Objects

4.3. Producing and Consuming Agents

An ecosystem of applications that all adhere to the same definition of Data Objects possess inherent advantages specific to data exchange. The normalization of the definition allows for implementation of specific APIs to be developed and adopted by applications that want to produce or consume Data Objects. This creates an opportunity for applications to communicate directly with one another without the need for different translation packages. For the past several decades, this type of communication has been relegated to bulky, inefficient sets of data that require multiple steps to export, manipulate, and trial and error attempts to import. Now imagine a scenario where applications across different vendors all support the same definition of Data Objects, and advertise implementation details (e.g. supported formats) via standards-based APIs. This is not a brand new concept; indeed, this type of scenario has been envisioned before. It would be naïve to expect that all application vendors will come together and adopt the same Data Object definition. But what if a few do? It would be a start, and the normalization of data exchange could begin and grow from there. We should ask ourselves why the oil and gas industry has been unable to make even minimal progress in this area, compared to other verticals such as the financial and medical industries.

Increased adoption of a standard definition for Data Objects also allows for the creation of solutions such as ActiveExchange. Although point-to-point communication between applications serves some use cases, there are many more complicated use cases that demand an event management and data flow engine that can orchestrate the proper sequence of events, all while taking advantage of a standard Data Object definition across all producing and consuming applications.

Also to be considered is that because application vendors will not all choose to support the same Data Object definitions, solutions such as ActiveExchange can play a role in bridging the gap. There are opportunities to create solutions that implement a façade around many different vendor- and customer-specific applications and databases that still achieves the "normalization of data objects", at least for any processes that interact with that integration solution.

4.4. Quality and Transformation Services

The normalization of data object definitions also allows for the creation of reusable quality and transformation services within an organization's data management ecosystem. These services support the ability to assess the quality of Data Objects while in transit to the indicated destination, and also perform calculations on the data in order to prepare it for the target.

By attaching quality metrics to each Data Object, consumers can make more informed decisions when deciding whether or not to pull in new or updated data. With standard Data Object definitions, it is possible for organizations to compile a multitude of different data quality rules, and wrap those inside of a data quality service that can be run on demand or within the context of a data exchange solution, such that data quality metadata and metrics are always attached to each Data Object before it reaches its target. This allows consumers to each enforce their own rules regarding the type of data that is allowed within the underlying application or database. Furthermore, this also allows for higher level logic to be enforced within the data exchange solution such that Data Objects that fall below a certain level of quality are never offered up to any consumers. This gives an organization a large amount of control, and flexibility, over the quality of data that is allowed to be distributed throughout the enterprise.

The ability to leverage data transformation services is also crucial to any data exchange solution. These services, like data quality services, should function as a "black box" that allow for the ability to pass in data objects and receive back the same Data Object, but with specific content modified. Examples include the ability to datum-shift surface locations, perform unit of measure calculations, rotate directional survey azimuths, and cross reference various types of reference data. These are all examples of *content transformations*, where the structure of the Data Objects does not change. The centralization of this logic makes it much easier to maintain and also achieve more uniformity specific to how these types of transformations and calculations take place.

5. Summary

This White Paper has introduced the concept of Data Objects as one of the key building blocks of effective E&P data management. In the context of this paper, a Data Object is defined as a collection of data attributes combined with the information required to manage the object to support business workflows.

It is considered that the definition of a standard set of attributes and data quality rules corresponding to the different levels of the Well Hierarchy is a logical extension to the PPDM 'What is a Well?' initiative. Gaining industry acceptance of these definitions will greatly facilitate data understanding, exchange, management and quality measurement across the entire Well Lifecycle. It is not a stretch to say that adoption of Data Object definition standards would have a profound impact on the way that data is managed, and exchanged, within the E&P industry, leading to better business decisions and workflow processes.

From an implementation perspective, there are robust open source technologies available today that make the event-based exchange of Data Objects across the Well Lifecycle a viable proposition. This paper has presented one such implementation in the form of EnergyIQ's ActiveExchange application.