

Experiencias de evaluación de calidad de datos

Por **Rubén Velazco** y **Carla Olmo**

Este trabajo fue seleccionado del VII Congreso de Exploración y Desarrollo de Hidrocarburos

Para una correcta gestión del conocimiento, la empresa debe establecer previamente un correcto sistema de gestión de la información: controlar su almacenamiento y su flujo y en general, la correcta gestión del dato. Este sistema debe proveer un conjunto de políticas y lineamientos que definan la manera en que la información va a ser utilizada y administrada dentro de la organización

Frases como “¿dónde encuentro la última versión de la simulación nodal?” o “el volumen en el control de pozo ¿será correcto?” o “los datos de tiempo real no concuerdan con los resumidos en la base de datos para el cierre de producción” o “¿de dónde tomo el último registro de presión de fondo de un pozo?” son indicadores de baja confiabilidad del dato, inexistencia de políticas y procedimientos para su tratamiento y una cultura organizacional que no contempla perfiles específicos con responsabilidad sobre la administración de la información.

Si bien conceptos como “calidad del dato” e “integración de información” se conocen e intentan aplicar desde hace años, sólo recientemente las compañías están reconociendo el valor que aporta a las decisiones de negocio contar con información accesible, confiable y a tiempo, que responde a las necesidades de los procesos operativos y de gestión.

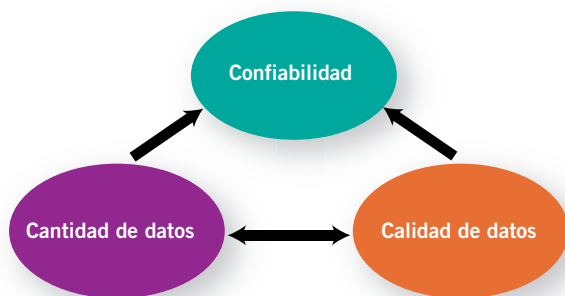
Esto se debe, en gran medida, a que durante estos años, tanto dentro como fuera de la industria, las operadoras y las empresas consultoras han ido formando un cuerpo de conocimiento específico en lo relativo a gestión de la información.

En este documento, revisaremos algunas de las mejores prácticas existentes en la materia, a la luz de una experiencia particular de aplicación de una metodología de evaluación del nivel de calidad de los datos asociados a los procesos de ingeniería de producción, en tres de los activos de PESA en la Argentina.

Primeramente, en el apartado “Contexto”, describiremos cómo surge la necesidad de este ejercicio para luego, en el apartado “Metodología aplicada” presentar las técnicas y métodos utilizados y luego enunciaremos, en el apartado “Sistema de gestión de datos”, las conclusiones y recomendaciones surgidas de la aplicación de dicha metodología, las cuales guiarán los próximos pasos de PESA en temas de gestión de datos y calidad.

De aquí en más, haremos mención a conceptos que quisiéramos especificar desde el inicio:

- En el contexto de este documento, la “confiabilidad de la información” se mide por la cantidad y calidad de los datos que la componen.
- Para estudiar el nivel de completitud en cantidad de datos de un determinado conjunto de información, se debe estudiar el nivel de completamiento de los campos correspondientes de la(s) base(s) de datos que soporta(n) el proceso de trabajo bajo estudio.
- La calidad de datos, en cambio, se determina según su grado de conformidad con requisitos previamente



establecidos: formatos, obligatoriedad, periodicidad, nomenclaturas, ubicación en la base de datos, etc. Esta definición de calidad se alinea con la de Philip Crosby [1], autor de importantes prácticas de gestión de la calidad.

Contexto

Una reciente investigación realizada por el CDI Institute [2] muestra que en un alto porcentaje de las empresas (31%), la iniciativa de gestión de datos surge en el área de Tecnología Informática (TI). Sin embargo, creemos que para que logre los objetivos debe contar con el compromiso del nivel gerencial del negocio y la designación de gestores de datos, esto es, profesionales formados en las disciplinas específicas, en nuestro caso, petróleo y gas, y en la problemática particular de administración de flujos de información.

La experiencia de PESA se basó en esta creencia: motivado por la percepción de un bajo nivel de calidad de la información asociada a los procesos de trabajo de ingeniería de producción, la línea gerencial de negocio fue quien solicitó a un equipo de soporte técnico la revisión del estado de situación de los datos y posterior propuesta de mejora.

Falta y/o duplicidad de datos, inconsistencia entre bases de datos origen y centralizadoras de información, desconocimiento de la información existente para consultas, datos catastrados con distintos criterios a través de la compañía y del tiempo, falta de información confiable para la elaboración de indicadores de desenvolvimiento; fueron algunos de los problemas detectados por la gerencia que motivaron la solicitud del estudio.

Metodología aplicada

La siguiente figura grafica el proceso repetido en las tres áreas de PESA.

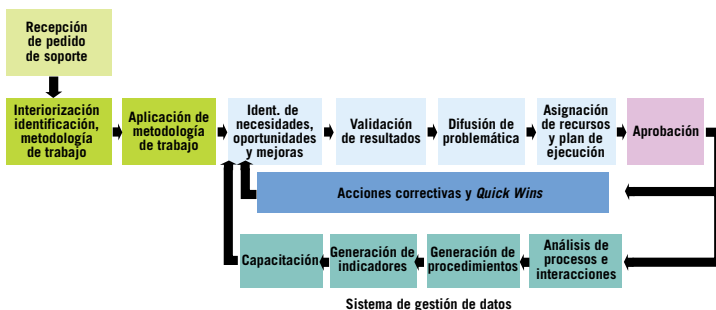


Figura 1. Secuencia de pasos de la metodología aplicada en PESA

El ciclo de actividades desde la “recepción del pedido” hasta la ejecución de “acciones correctivas y Quick Wins”, corresponde al ejercicio de evaluación de la calidad de la información y presenta un ciclo típico de mejora continua (Planificar- Hacer-Controlar-Actuar, conforme a la definición de Shewhart, modificada por Deming [3]).

El segundo ciclo (Implementación del “sistema de gestión de datos”), presenta la conclusión de largo plazo del trabajo realizado en PESA: los lineamientos generales

para un sistema de gestión de datos que abarca no sólo el primer ciclo (puntual de análisis de calidad), sino también las actividades de largo plazo que apuntan a la optimización del uso de la información, aplicando teorías y técnicas de análisis de procesos, mapeo de flujos de datos, estandarización de procedimientos y roles asociados a la administración de información, generación de indicadores para seguimiento y control de desvíos, capacitación y difusión a la organización.

Proceso de trabajo

El proceso se inicia en el área de Operaciones con un pedido de soporte por parte de gerencia. Las actividades del proceso de trabajo se agruparon en las siguientes etapas, que serán descritas a continuación:

- Interiorización.
- Relevamiento de información.
- Clasificación de causas raíces.
- Agrupamiento y priorización.
- Validación.

Atendiendo a las mejores prácticas que indican que este tipo de trabajos deben ser ejecutados con alta participación de los interesados directos y coordinado por recursos con conocimiento de teoría y técnicas de gestión de datos; se designó como coordinador de los trabajos al Analista de extracción, referente de datos de la disciplina de Ingeniería de producción.

Este analista cubría los requerimientos del rol de gestor de datos (que a posteriori de estos ejercicios se formalizaría como parte de la implementación del sistema de gestión de datos): contar con el conocimiento específico de los procesos de negocio y de la teoría y las técnicas de gestión de datos.

Interiorización

La primera etapa del trabajo consiste en generar un

espacio para la interiorización en la problemática, cuyas principales acciones apuntan a la comunicación con los clientes con el objeto de entender el tipo y magnitud de la cuestión, para poder adaptar la metodología de trabajo al caso presentado.

Las mejores prácticas sugieren que todo análisis con foco en la optimización se inicie por el estudio detallado de los procesos abarcados con los responsables de su ejecución, mapeando los flujos de información asociados tal cual se encuentran en la actualidad y documentando cómo deberían fluir en el caso óptimo.

En el caso de PESA, esta actividad se nutrió de información de actividades anteriores y del conocimiento de detalle que tenía el equipo de evaluación de los procesos operativos. Como resultado, se contó con una lista inicial de problemas a profundizar, que guiaron el trabajo posterior de relevamiento de información.

Relevamiento de información

Con foco en ahondar en los temas identificados y avanzar en la determinación de sus causas, se aplicaron en paralelo las siguientes técnicas:

- **Entrevista con la gerencia:** con esta herramienta se pudo profundizar en la percepción de la gerencia sobre las evidencias de problemas y se pudo determinar el alcance (procesos, flujos de datos y aplicaciones) del trabajo a realizar.
- **Muestreos e investigación en bases de datos:** para la aplicación de esta técnica, se seleccionó un conjunto de datos mínimos indicado por los referentes de los procesos y se realizó un análisis estadístico del estado de calidad y cantidad de dicho grupo de datos. Mediante actividades de estudio de tendencias, comparación contra valores deseados o estándares, extrapolación y graficación, se pudo demostrar el nivel de calidad y completamiento. Estos resultados permitieron orientar hacia las posibles causas de los

problemas identificados, soportar las preguntas de las entrevistas e identificar a qué actores dentro del proceso realizárselas.

- **Observación in situ:** actividad orientada principalmente a obtener información relacionada a la forma en que se utilizan las aplicaciones (*software*), modo de interacción de los profesionales y técnicos, horarios de trabajo, relación con personal externo, herramientas que se utilizan en forma rutinaria y no rutinaria, carga real de tareas que cumplen las personas abocadas al control de flujos de datos, recursos tecnológicos disponibles, nivel de experiencia y capacitación de los usuarios, etc. Surgen durante esta actividad nuevas preguntas a incorporar al cuestionario de las entrevistas.
- **Análisis de aplicaciones:** esta técnica fue desarrollada para detectar posibles problemas de *performance*, evaluar complejidad de acceso y dificultades de generación de salidas de datos.
- **Entrevistas personales a usuarios:** en su gran mayoría fueron informales ya que el perfil de los entrevistados así lo requería para un óptimo resultado. Para la actividad, se tomó como guía el cuestionario realizado en función de las conclusiones de las técnicas desarrolladas anteriormente. Esta actividad permitió detectar y analizar evidencias de problemas (desvíos),

identificar causas y avanzar en la determinación de posibles soluciones.

Clasificación de causas raíces

Contando con la lista de problemas existentes, se les asoció las causas identificadas, clasificándolas en tres grupos:

1. **Cognitivas o de conocimiento:** causas relativas a falta de capacitación o capacitación insuficiente (en el uso de los *softwares*, en la administración de los *softwares*, en los procedimientos, etc.) de los recursos, desconocimiento del rol de los recursos en los procesos, desconocimiento de las capacidades de administración de información de los *softwares* instalados, etcétera.
2. **Aplicativas o relacionadas con la aplicación:** causas relativas al desenvolvimiento de los *softwares* instalados: lentitud, errores, configuraciones complejas; necesidad de trabajos manuales intermedios, carencia de un *software* que soporte una porción de los procesos, etcétera.
3. **Operativas, enfocados a los desvíos en los procedimientos:** falta de formalización de procesos y procedimientos operativos, de aseguramiento de la calidad de la información y de administración de los *softwares* instalados; falta de formalización y estanda-

| Problemas | Clasificación de las causas | | | Corto plazo | | | Largo plazo |
|--|---|---|---|-------------|-------------|---|---|
| | Cognitivas | Aplicativas | Operativas | Im-pacto | Apli-cación | Prop. acción | Propuesta |
| Faltan datos en la base de DFW y algunos de los existentes no cumplen con los parámetros definidos en los procedimientos (Ej.: Causas de <i>Pulling</i> no coincidentes con los <i>Picklist</i> establecidos). | <p>La capacitación brindada para la carga de datos al personal de <i>Pulling</i> y Construcción y Perforación de Pozos no fue efectiva. En parte, debido a que esta no se realizó según los procedimientos desarrollados por Ing. de Producción.</p> <p>No están explicitadas ni informadas las funciones de los "propietarios de datos" en cuanto a su papel en garantizar la confiabilidad de la información.</p> <p>Como forma de proceder normal (erróneamente), los reportes de bombas se generan cada vez que estas sufren cambios en una intervención, de otra manera no se crean dichos reportes.</p> <p>No existe un plan de capacitación/actualización para usuarios de carga de DFW (<i>Pulling</i>, copo, WL, <i>Coiled Tubing</i>, otros).</p> | <p>Ausencia de campos en la aplicación para la carga de datos requeridos, por ejemplo:</p> <ul style="list-style-type: none"> • datos de tránsito de fluidos, • movimientos de válvulas en pozos con inyección selectiva; • datos de bombas PCP (<i>Dims</i> no soporta este sistema de extracción, se utiliza el reporte correspondiente a BM). | <p>No existe un procedimiento que establezca las acciones que deben seguir los propietarios de datos para garantizar la confiabilidad de la información.</p> <p>La administración de los datos de WO y Perf. se realiza en forma independiente de los datos de <i>Pulling</i> y WL, por lo que existen diferencias de criterio en la carga de los campos de uso común para los distintos eventos. (Ej.: sartas de producción, WO no carga reportes de bombas, etcétera).</p> <p>El soporte de DFW está enfocado en la revisión de la información generada solo en los eventos de Perforación y WO.</p> <p>No existe dentro de este grupo una atención específica para los eventos de <i>Pulling</i> y WL, por lo cual no hay casi interacción con los equipos de Ing. de Prod. que son los que establecen los criterios de calidad.</p> <p>No existe un procedimiento para el alta y modificación de eventos de <i>Pulling</i> y WL.</p> <p>Se creó una base paralela en Access para el seguimiento de los costos de <i>Pulling</i>, los campos registrados en esta son soportados por DFW.</p> <p>No existen procedimientos de carga para las operaciones de <i>wireline</i> y <i>coiled tubing</i>. Esto hace que los operadores carguen en la base de datos de acuerdo a su propio criterio.</p> | A | DFW | <p>Difundir desde los niveles superiores, la necesidad de la organización de que la información sea registrada con calidad en las bases corporativas.</p> <p>Realizar una revisión en conjunto entre los referentes de Construcción y Perforación de Pozos, Ing. de Producción y Reservorios de los procesos de Perforación, WO, <i>Pulling</i>, WL, <i>Coiled Tubing</i> y otras operaciones especiales de Ing. de Producción, actualizando los procedimientos que regirán la carga de datos.</p> <p>Realizar capacitación a usuarios de carga de <i>Pulling</i> y Construcción y Perforación de Pozos según los procedimientos actualizados de Ing. de Producción. Hacer hincapié en los datos que cada sector necesita del otro.</p> <p>Difundir las responsabilidades actualmente establecidas para los propietarios de datos.</p> <p>Gestionar la unificación de los criterios utilizados por el soporte de DFW en la administración que realiza actualmente para la información de Perf y WO, para adecuarlos a <i>Pulling</i> y WL.</p> <p>Solicitar la elaboración y difusión de un procedimiento único para el alta y modificación de eventos.</p> | <p>Establecer un sistema eficiente para la gestión de la información, de manera de asegurar sistemas confiables en el tiempo.</p> <p>Este debería contemplar los procedimientos, las herramientas, el soporte, mapas de datos, figura de gestores de datos, etcétera.</p> |

Figura 2. Matriz resumen del trabajo

| | | | | |
|-----------------|----------|--|---|--|
| Impacto | M | <p>Realizar los cambios necesarios en SIGEDIP para soportar la dosificación por bacheo que baña la instalación.</p> <p>Evaluar si la solución de carga de Análisis de oil y gas es posible en Tow.</p> <p>Realizar capacitación para administradores cubriendo temas de configuración de sistemas colectores, ZLA, TDML.</p> <p>Definir y difundir entre los responsables de las transmisiones y administración de la base de datos el procedimiento de transmisiones de datos entre las distintas bases de datos.</p> | <p>Evaluar el desarrollo de funcionalidades en SIGEDIP para la carga de los análisis de oil y gas que no son cubiertos totalmente por TOW.</p> | <p>Realizar una evaluación y mejora en los procedimientos de registro y control de información en Tow.</p> |
| | B | <p>Evaluar y definir si los datos de armadura pueden llevarse en DFW (reporte WellHead).</p> <p>Realizar capacitación para usuarios de consultas, Data Analyzer (DFW), Reportes Tow, QControl. Distribuir manuales de aplicación.</p> | <p>Desarrollar una funcionalidad para registrar los datos generados en la PRM.</p> <p>Realizar capacitación para la carga apuntando a reforzar la carga de datos que hoy faltan en Tow.</p> | <p>Mejorar las comunicaciones de los yacimientos.</p> |
| | B | | M | A |
| Esfuerzo | | | | |

Figura 3. Matriz de análisis de impacto vs. esfuerzo de acciones correctivas

rización de criterios de carga, etcétera.

Los mencionados grupos causantes de problemas son normalmente los que impactan en forma negativa en la cantidad o calidad de los datos. La figura 2 muestra el formato de la matriz utilizada para listar los diferentes pro-

blemas detectados, las causas clasificadas en los tres grupos mencionados, la categorización según impacto, y las acciones correctivas (“paliativos”) de corto y largo plazo.

Agrupamiento y priorización

Para cada problema, el equipo de evaluación enumeró acciones correctivas, las cuales buscan presentar una solución inmediata o de mejora en el corto plazo. Teniendo

en cuenta un escenario de corto plazo, se definió, para cada acción correctiva, el nivel de aporte/impacto sobre la confianza de la información y se le asignó una estimación del esfuerzo necesario para implementarla. El aumento en la confiabilidad se refiere a la cantidad/calidad de la información de los procesos internos y externos (definidos a continuación). Los niveles de aporte/impacto así como el esfuerzo se clasificaron como alto, medio y bajo.

Procesos externos: asociados con información requerida por socios, organismos del Estado, y regulación legal (Ley SOX).

Procesos internos: asociados con procesos de análisis con alto impacto en CAPEX/OPEX.

Esta matriz muestra el resultado de parte del ejercicio de clasificación de acciones correctivas.

Por otro lado y teniendo en cuenta un escenario de largo plazo, el análisis de la situación en su conjunto, permitió establecer un diagnóstico del sistema y presentar una propuesta de acción a largo plazo que apunte a la solución definitiva o de raíz del problema. Esto resultó necesario para darle sustentabilidad en el tiempo al trabajo realizado.

Validación

En tareas conjuntas con los referentes de los procesos relevados, se procedió a la revisión, validación de los resultados obtenidos y difusión del documento generado, esto último a los efectos de que los recursos involucrados en las acciones a desarrollar entiendan la problemática y estén predispuestos a participar en la implementación de acciones de corto y largo plazo.

Las acciones fueron volcadas en un plan de ejecución, el que involucraba recursos de las operaciones, de los centralizados y área de servicio de tecnología informática (TI).

Aprobado el plan de ejecución, se comenzó con las tareas descritas en el plan de tareas identificadas. Este plan se estableció de modo tal que las acciones correctivas del corto plazo se ejecutaran prioritariamente según el grado de impacto en la operatoria. En forma paralela, se comenzaron

a ejecutar tareas relacionadas a la implementación de un sistema de gestión de datos, el cual se definió como fundamental para lograr una máxima confiabilidad de los datos.

Sistema de gestión de datos: conclusiones y recomendaciones

El análisis de los hallazgos de aplicar la metodología expuesta nos permitió enunciar las siguientes conclusiones:

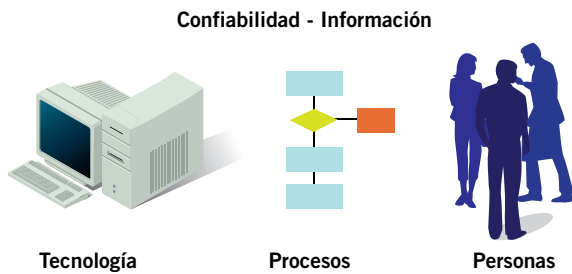


Figura 4. Pilares de la confiabilidad de la información

- Los problemas detectados corresponden, en su gran mayoría, a ineficiencias en los procesos de flujos de información, causados por razones cognitivas, aplicativos y operativas. Estas ineficiencias tienen un impacto negativo en la confiabilidad de la información de la compañía.
- En general, los problemas son comunes a todos los sitios donde se ejecutan los mismos procesos, por lo cual los hallazgos encontrados y las soluciones planteadas pueden y deben ser replicados.
- A los efectos de cubrir la necesidad de respuestas a las evidencias de problemas en forma inmediata a las operaciones, se deben plantear acciones de corto plazo. Estas acciones deben priorizarse por el nivel de impacto (en el corto plazo), entendiendo por tal el impacto en el aumento de confiabilidad de los datos asociados a procesos internos y externos. En paralelo, se debe trabajar en profundidad sobre las causas de estos y la implementación de soluciones sustentables.
- Se evidencia la necesidad de establecer e implementar sistema de gestión de la información (*data governance*) que provea el conjunto de políticas y lineamientos que definen la manera en que la información debe ser utilizada y administrada dentro del sector/compañía y la vinculación de dichas políticas y lineamientos con los procesos, roles y herramientas que soportan la aplicación.

El foco de este sistema sería que existan los flujos de datos requeridos por los procesos de negocio y que existan los procesos de administración de datos que aseguren que los datos sean confiables, eficientes, accesibles, seguros. Entre los beneficios esperados podríamos mencionar:

- Disponer de la información necesaria en el momento oportuno, con el nivel de calidad y granularidad requerido por los tomadores de decisiones.
- Alcanzar una organización que garantice y facilite la comunicación y coordinación entre sus miembros.
- Integrar y administrar la información interna y externa como un todo para facilitar la toma de decisiones más eficientes.
- Establecer y adecuar las herramientas que viabilizan la consulta de datos.
- Cumplimiento de reglamentaciones internas y externas.
- Reducir los costos de la administración del *software* instalado.

El objetivo por alcanzar debe apuntar a lograr y mantener una mejora sustancial en la confiabilidad de la in-

formación requerida por los procesos y que sea sostenible en el tiempo.

Las mejores prácticas del MDM Institute [4] muestran que la existencia de un proceso de gobierno de los datos que presente los lineamientos de la política de gestión de los datos y vincule a los procesos, roles y herramientas que soportan la aplicación de dicha política, es vital para alcanzar el objetivo de contar con información confiable, adecuada, completa, en tiempo y que resulte consistente.

Los pasos iniciados por PESA para la instauración de este proceso pueden enumerarse de la siguiente manera:

- Concientización de los niveles de dirección con relación al valor de la calidad, la conformidad y la seguridad de los datos.
- Establecimiento y documentación de normas, reglas, procedimientos y tecnologías adecuadas que permitan garantizar la disponibilidad, accesibilidad, calidad, coherencia, trazabilidad y seguridad de todos los datos producidos y utilizados en la empresa.
- Designación de un área centralizada cuya responsabilidad sea coordinar el proceso de implementación y guiar metodológicamente a la organización, proporcionando la teoría y las técnicas necesarias en cada paso.
- Designación de gestores de datos por disciplina de negocio, con localización cercana a los procesos específicos.
- Planificación del trabajo de los gestores de datos con casos específicos de ingeniería inversa sobre flujos de información asociados a procesos críticos o estratégicos para la organización. Se procura que esta forma de trabajo permita apuntar a solucionar necesidades reales de la organización y en paralelo ir formalizando los procesos de gestión de la información, roles y funciones asociadas.

Por último, se aspira a que, en forma iterativa, la documentación de una metodología de trabajo de los gestores de datos surgida de la aplicación práctica de su función, pueda irse generando con el tiempo.

- Elaboración de un detallado plan de capacitación de los gestores de datos en teoría y técnicas específicas de la problemática de administración de información.

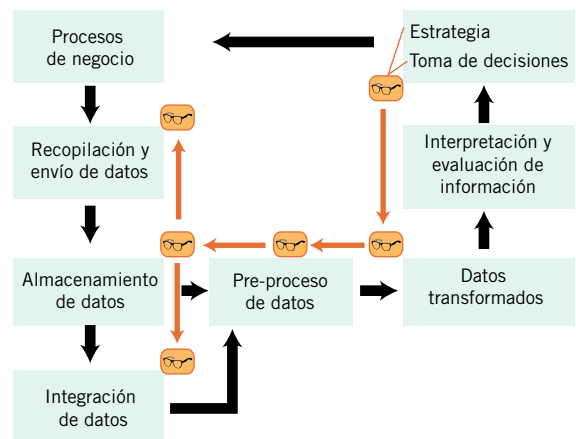


Figura 5. Esquemática del proceso de ingeniería inversa en un contexto de datos

- Elaboración de un plan de capacitación en temas relacionados a gestión de datos para los profesionales específicos de las disciplinas de negocio. La realidad de organizaciones con un alto grado de madurez en gestión de la información, muestra que los gestores de datos son profesionales con amplia trayectoria en su disciplina que evolucionan a responsables por los flujos de información de la disciplina en cuestión.
- Comunicación y divulgación (por medio de reuniones con actores clave) de la teoría de base, los planes y forma de trabajo, procurando la generación de conciencia y alineamiento a los objetivos por alcanzar.
- Involucramiento temprano de TI para lograr su concientización sobre el impacto estratégico de estas iniciativas y lograr su alineamiento y colaboración en la fase de implementación de mejoras. ■

producción de E&P, actualmente se desempeña como Gestor de Datos Especialista de Petrobras Argentina S.A.; desde 2009 forma parte de la Gerencia de Ingeniería de Producción de dicha empresa.

Carla Olmo es licenciada en Sistemas e Información de la Universidad Católica Argentina con un posgrado en Ingeniería de Petróleo del ITBA (2010), trabajó del 2006 al 2008 en la primera etapa de implementación de los procesos de Gestión de Datos en Petrobras Argentina S.A. Actualmente es Project Manager de la Gerencia de Reservas y Reservorios de la misma compañía.

Bibliografía

- [1] Philip Crosby (1979): *Quality is Free*. New York: McGraw-Hill.
- [2] The CDI Institute, "Corporate Data Governance Best Practices – 2006-7 Scorecards for Data Governance in the Global 5000", abril 2006.
- [3] Shewhart & Deming (1999): *Manual de la ASQ*, páginas 13–14, American Society for Quality.
- [4] The Master Data Management (MDM) Institute, <http://www.tcdii.com/whatIsMDM.html>

Los autores agradecen el aporte al armado del presente documento y la revisión de Gabriel Cuello, Yacomo Lattarulo, Gumersindo Novillo y Graciela Funes; y a Petrobras la autorización para publicarlo.

Rubén Velazco es Analista en Sistemas con más de quince años de experiencia en gestión, análisis de datos e implementación de mejoras tecnológicas en los procesos de