

PNEC Data and Information Integration 2003 Houston, May 2003

Introduction

The 7th edition of Phil Crouse's Data Integration Conference was quite well attended considering the prevailing world situation – with 140 pre-registered¹. Since last year there have been some significant changes in the data management scene which require some background to follow. The first 1997 PNEC conference was sponsored by the Geoshare User Group and had a whole session on Geoshare applications and was followed by the Geoshare AGM. PNEC also attracted the Canadian Public Petroleum Data Model PPDM as exhibitor and sponsor – PPDM also piggy-backed its Houston spring member meeting onto PNEC. The association of PNEC, PPDM and Geoshare established a pattern which was continued until last year when Geoshare was kind of disbanded and handed over to the Petrotechnical Open Standards Corporation (POSC) for 'retirement'. This has proved a significant development in that both POSC and PPDM now cohabit at PNEC – and appear to be working towards sharing some XML-based developments. This is quite an achievement for Phil Crouse's entrepreneurship – he has succeeded in bringing the two warring standards bodies together where years of arm twisting by major oil companies has conspicuously failed.

Elsewhere, the PNEC conference offered a fair cross section of where data management is at today – with a fair crop of XML-based initiatives, more on searching and spidering of multiple data sources and contributions from the e-business crowd. Industry thinking on 'who does what' seems to have retreated some from the vigorous outsourcing advocacy of prior years. Today, independents don't see any political incorrectness in developing fairly substantial solutions in-house. Majors may 'buy not build' but this is likely to involve considerable remodeling of off the shelf products.

Highlights

- [Kerr McGee's Desktop well log delivery](#).
- [Migrating upstream computing to Linux](#) (Anadarko)
- [XML Modules](#) (POSC)
- [Well log nomenclature](#) (panel discussion)
- [Data viewers for wide area networks](#) (INT)

¹ The demographics are interesting: 14 from Schlumberger/GeoQuest against 2 from Halliburton/Landmark; 12 from Shell against 2 from ExxonMobil and 1 from BP.

Contents

INTRODUCTION	1
CONTENTS	2
DATA SOURCE RECONCILIATION WITH FUZZY LOGIC – JO KOSTECKA, MARATHON	3
DESKTOP WELL LOG DELIVERY – PAUL HAINES, KERR MCGEE	3
TAKING UPSTREAM COMPUTING TO LINUX – WILL MORSE, ANADARKO	3
REDUCING COSTS BY NETWORK MONITORING – RICHARD CROUNSE, IT MONITOR	4
XML MODULES – JOHN BOBBITT, POSC	4
PANEL DISCUSSION I.....	4
INTEGRATING TECHNOLOGY AND PROCESS – TRUDY CURTIS, PPDm ASSOCIATION	4
WITSML STANDARDS FOR WELLSITE INFORMATION TRANSFER, ALAN DONIGER, POSC	5
PRODUCTION DATA REPORTING – AZHAR SINDHU, RCG INFORMATION TECHNOLOGY.....	5
PROJECT VERACRUZ - DATA AND INFORMATION MANAGEMENT – RAY FLORES, PEMEX.....	5
STATOIL’S E&P RESULTS DATABASE – NAJIB ABUSALBI, SCHLUMBERGER	5
PANEL DISCUSSION II NOMENCLATURE.....	6
SUPPLY CHAIN WISDOM – PETER FLANAGAN, OILDEX.....	6
KM AND WEB TECHNOLOGIES – JEFF PFERD, PETRIS.....	6
WEB TECHNOLOGIES FOR INFORMATION ACCESS – UGUR ALGAN, LANDMARK.....	7
ONLINE ACCESS TO WELL LOG DATA – RICHARD HERRMANN, A2D TECHNOLOGIES, GENE RHODES, LANDMARK	7
DATA VIEWERS FOR WIDE AREA NETWORKS – JIM VELASCO, INT.....	7
INTEGRATING DISPARATE DATA REPOSITORIES – STEVE HAWTIN, SCHLUMBERGER	7
‘LET THE ENGINEERS ANALYZE THE DATA’, TIM ELSER, PIONEER RESOURCES CANADA	8
IM – THE <i>CULTURAL</i> CHALLENGE – FLEMMING ROLLE, PARAS	8
DYNAMIC DATA MANAGEMENT – CARL HUCSALL AND RAY CLINE, SAIC.....	8
28 SPATIAL SEARCH TECHNOLOGY – BRUCE SANDERSON, GEODYNAMIC SOLUTIONS.....	8

Data source reconciliation with fuzzy logic – Jo Kostecka, Marathon

Marathon has previously reported on its efforts to unify a multiplicity of heterogeneous databases – notably in its ‘Portal of Portals’ paper at last year’s PNEC conference. Uniting data from different sources implies consistent naming conventions – which in the case of publicly available data, and data from acquisitions, is not always the case. 12 and 14 digit API’s and other well naming conventions require matching to avoid duplicates or missing data. Kostecka has developed a [Perl](#) program for semi-automated well number matching. The program uses what Kostek describes as ‘fuzzy logic’ to locate partial matches based on API number, location, total depth etc. The program assigns weights to different variables. A two pass process brought the number of suspect matches down to manageable proportions.

Desktop Well Log Delivery – Paul Haines, Kerr McGee

Kerr McGee’s IM strategy can be summarized as ‘buy, apply and modify’ – application software is bought from established vendors, but KMG finances software development and attempts to influence future directions. Haines described KMG’s IM function as responding to internal client requirements – in a relatively short time frame – “there are no more ‘Project Mercuries’, with a 3-4 year time horizon”. Users of KMG’s in-house developed well log databases ‘don’t want LIS – they want LAS’. Haines recommends making sure ‘you receive good documentation from data vendors’. KMG uses [OilWare](#)’s Well Log Indexing System (WLIS). This ‘crawls’ corporate file systems for LAS, LIS and DLIS log files which are converted to XML. [InnerLogix](#)’ ILX Cataloguer is also being deployed to crawl and index raster logs recognizing TIF, PDF, CGM etc. formats.

Q&A

Is the WLIS database POSC compliant?

No it is PPDM-based

Is the UWI internal to KMG?

Yes- but it is a composite of IHS Energy’s P2000 and Probe UWIDs.

Is Datalogix software deployed?

Yes but not yet extensively tested

What sort of savings have been achieved?

60-70% time savings for users, 30-40% savings on log costs.

Taking upstream computing to Linux – Will Morse, Anadarko

Morse stressed that the subject of his talk is about bringing Linux-based desktop computing to the upstream – Linux’s role in seismic processing is considered a given. The motivation for moving interpretation software to Linux is ‘be there or be square’². Morse failed to explore this logic further – but pointed out that developers ‘now prefer Linux’ – although the costs of a migration to Linux ‘may not be as low as you think’. Morse noted that most all E&P software is 32 bit – 64 bit software ‘is coming’. The tough part of using Linux in high end graphics is the poor support for graphics cards. The most important point to remember in specifying Linux software is that it should be **Linux Standard Base** (LSB) compliant. Vendors should be encouraged to supply LSB-compliant applications. More from www.linuxbase.org. Particular attention should be made to testing NFS, NIS, networked license managers, remote Oracle etc. as vanilla Linux does not support these out of the box. Certification should ensure qualification in these esoteric Linux extensions. In view of all this, Linux deployers should ‘manage expectations for cost savings’.

Q&A

It would appear that the graphics card issue is a major obstacle to Linux take-up!

Yes – just something we’ll have to live with though!

What software are you running on Linux?

Anadarko is running Landmark and Paradigm apps on Linux today.

² For those who are not familiar with this 1960’s slang – the implied reason is that Linux is an unstoppable trend.

Reducing costs by network monitoring – [Richard Crouse](#), IT Monitor

IT downtime was proving a major problem on Pemex's Cayo Arcas platform. Crouse described a number of commercial tools which provided monitoring of infrastructure performance in real time. [HP OpenView](#) provides basic monitoring, but 'can be improved on'. Trouble shooting routers and switches can spot faults developing. Network services can be monitored with Open Source [Software Process Dashboard](#), with [Cisco NetFlow](#), Osi Software's [PI System](#) and [Real Time Performance Management](#). Activity can also be captured as a 'net bot' image – and stored in an Oracle blob – like a security video of network activity – ready for post-intrusion analysis. PI System claims 10,000 installations – in business for 20 years – also offers '[Wired City](#)' hosted remote network monitoring service.

XML Modules – [John Bobbitt](#), [POSC](#)

Bobbitt started out with a comment on previous discussions on units of measure in GIS and related fields. OpenGIS and ISO use the XML quantity data type. Bobbitt commented "pretty good solutions already exist for these issues – or you can just do what you want to!" Bobbitt's presentation of XML modules presented work done by OASIS/UNCEFACT – X12 – XML. These different groups use the same concept. An XML 'Module' is an XML business object such as a well header. Different XML modules can be assembled to create compound documents 'on demand'. To cater for specific (and limited) instances where local requirements mandate refinement or restriction in use the 'Profile' concept is introduced. Thus the North Sea Profile contains rules and guidelines for N. Sea use (DTI Wellname etc.) whereas the Gulf of Mexico Profile uses the API number. Note that profile creation must follow certain rules for legitimacy³; geographical variants are OK – but company-specific profiles are proscribed. More from www.posc.org/ComponentDocs/main.html. Foreign modules can be included – such as a PPDM Business Associate module within a POSC well header module. PPDM are said to be profiling the POSC well header ML module which may be accessed as an XML registry - <usesWell xlink:href='http://www.x.gov/well_12345'>. More explanation from [John Bobbitt](#).

Panel Discussion I – IT at large

Spam – regular spam is a pain – but easily dealt with. Harder is internal 'spam' – internal memos to be read and responded to in a political situation.

Understanding IT and the business – problems arise because managers don't understand IT, and developers don't understand the business. Another dimension is witnessed by the 'silo-orientation' of the vendors – contrasted with the 'interoperability – oriented' users. One speaker teaches an IT class for geoscientists – but this has lukewarm support from management. According to a McKinsey study – best Fortune 500 companies are those whose board members 'take an interest in IT'! A suggestion for an oil and gas data management degree course – taken very seriously – but a proposal for a degree in systems admin was pooh-poohed by one faculty because it was not considered professional/academic.

Integrating Technology and Process – [Trudy Curtis](#), PPDM Association

Trudy Curtis – CIO and acting CEO following the sudden departure of Scott Beaugrand from the PPDM Association, believes that technology is 'an opportunity for gain – or chaos'. Data volumes are exploding, and information proliferating (and duplicating). To sort out the chaos, Curtis offers the three 'P's – Policies, Practices and Procedures – on what to keep and how. Legislative pressure (Sarbanes-Oxley in US, PIPEDA in Canada) is forcing corporations to 'know their data' and to adopt a holistic approach to management.

Q&A

It's OK to talk of managing and storing – but what about throwing stuff away – look at Enron and the Tobacco industry? There has to be a retention limit for G&G data – even if it's 50 years!

³ One aspect of POSC's best practices is whether element and attribute names should be in UPPER CASE, lower case – or CamelCase (new terminology for us!).

There has been a failure to apply records management to G&G. It is OK to throw stuff away – but what's there needs organizing – following regulations, cataloging and best practices.

[WITSML standards for wellsite information transfer – Alan Doniger, POSC⁴.](#)

The substance of WITSML has been extensively covered in previous Technology Watch reports. Doniger announced 'significant expansion of WITSML to cover completions and cost data. At the same time POSC is trying to keep WITSML 'aligned with existing and future standards from POSC and other organizations.

Q&A

By moving into cost data there will there be overlap with the API standards?

We anticipate collaboration in this field.

What purpose is WITSML actually fit for?

Data capture in field – and movement to the office. But the schema may encounter valid use elsewhere, later in the data lifecycle for post processing. WITSML has also been mooted for regulatory reporting.

[Production data reporting – Azhar Sindhu, RCG Information Technology](#)

Sindhu described [RCG IT's](#) approach to solving a US operators production reporting requirements. Integration can be achieved with a variety of technologies – API-oriented, server-based, messaging-based, enterprise application integration (EAI) or web services. The project involved integration of well data, land information, operations, applications such as TietoEnator's [Energy Components](#), Landmark's [DIMS](#) and [DSS](#), MRO Software's [Maximo](#) and portal projects. RCG's analysis is based on its productized [Business and Technology Roadmap](#). RCG IT clients include Apache Corp. BP, ExxonMobil, Texaco and Unocal.

[Project Veracruz - Data and Information Management – Ray Flores, Pemex](#)

A walk-through the development of Pemex's E&P technical database for the Veracruz region. Pemex is a big Finder shop. Finder integrates data from Pemex in-house applications along with GeoFrame, OpenWorks, Eclipse and OilField Manager. A data integration 'bus' is planned to extend integration across SAP and Merak.

[Statoil's E&P Results Database – Najib Abusalbi, Schlumberger](#)

Abusalbi presented 'one aspect' of the work performed by Schlumberger Information Solutions for Statoil⁵ – the interpretation results database. This is not (just) a database, neither is it a 'silver bullet' – technology needs to be coupled with the business process. Along with the usual problems of data volumes, Statoil have experienced the high cost of disparate data stores, of the lack of context stored with data, data quality issues and generally the weak integration offered by commercial and internal solutions. On the positive side, continuing efforts in data management have increased the time spent interpreting and documenting data from 20% in 1996 to around 50% in 2001. Abusalbi slipped in a plug for SIS' data streams – which are now presented with an orthogonal 'workflow' component – data management 'streams' are decoupled from usage 'workflows' – "quality and context management underpins the solution". The Statoil project set out to 'implement an IM solution to manage corporate data (specifically interpretation results of validated quality and context that supports decision making, and effective re-use of data throughout the asset's lifecycle.⁶'. Statoil interprets in OpenWorks and GeoFrame and results are captured back to the InfoTrack results database. 'Results' – interpretations, grids etc. are stored in a vendor-neutral format. Interestingly, each component database (OpenWorks, GeoFrame etc.) is accessed with its dev kit API – no mention was made of OpenSpirit.

Q&A

Norway benefits from a very structured data environment – with an abundance of clean data – what of other environments?

⁴ POSC is now the Petrotechnical Open Standards Corp.

⁵ Presented last year at PNEC by Kjetil Tonstad.

⁶ From the Schlumberger – Statoil letter of intent.

Again, there is no silver bullet – it is easier to achieve these results in Norway.

No mention of PetroBank?

Norwegian operators use PetroBank as the source of seismic data. Statoil is connected to PetroBank. The current presentation concerns a complementary strategy for corporate information.

End user uptake?

The project has been structured to ensure user participation.

Panel Discussion II Nomenclature

This Shell-chaired discussion was centered on nomenclature – in particular **well naming systems**. IHS Energy now offers a service to US States to check that API well numbers have not been issued before. Some speakers asked if this should be extended worldwide – but this was met with some reticence – ‘if it ain’t broke, don’t fix it!’ An AAPG committee looked into this, and today, Petroconsultants (an IHS Energy unit) offers a well-number generation service to operators – there is a problem that now many are aware of this service. ConocoPhillips does make use of the Petroconsultants service, but the problem is that this generates a ‘private’ in-house well number. The true benefit will come when the Petroconsultants numbers are broadcast to other companies. Other complicating well numbering issues were raised – like complex side-tracks, multiple completions. Shell Expro migrated its numbering system from the Shell internal convention (based on slots) to the DTI system (based on drilling order). So a well like Brent A 9/19 became 9/3 – this was not highly appreciated by users. Interest was expressed in kicking-off a POSC well numbering project.

Well log mnemonics – POSC announced that as part of its Practical Well Log Standards (PWLS) work, it had been granted the right to use and migrate the Schlumberger classification including curve mnemonics. Both Baker Atlas’ Recall and Schlumberger’s LogDB manage the mnemonics tables.

Well status classification standard – Work is in progress with the Minerals Management Service on well class reference standards. John Bobbitt explained that there existed an orthogonal classification of purpose (water, oil, injector etc.) and result (dry hole, P&A, suspended etc.). These twin identities are often confused for plotting purposes – where this issue is of symbology, not classification. The UK DTI has suggested standard names for purpose and status (drilling, POOH, fishing etc.) – a classification which has expanded into a detailed hierarchy of activities. Richard Herrmann (A2D) suggested using a subset of the SPE’s 200 graphics for plotting.

IT/IM at large – Borland described Shell’s ‘struggle’ between corporate, ‘global solutions’ IT and G&G users asking – what/who is IM/DM? What career paths are on offer for IT professionals? Who should they report to, exploration management or CIO? ExxonMobil has separated out global IT (email, office automation etc) but has kept engineering and G&G IT in a separate company with its own CEO, career paths – which belongs to upstream exploration. Interestingly, ExxonMobil recognizes the key role played by data management professionals “we’re still around – have not been downsized – ExxonMobil wouldn’t even offer a voluntary redundancy package to data managers!”

Supply Chain Wisdom – Peter Flanagan, [Oildex](#)

Oildex was founded on the observation that a \$16 invoice could cost as much as \$30 to process.

Using digital invoice this cost has been brought down to \$5 while the payment cycle time has been reduced from 60/90 days to 14 days.

KM and web technologies – Jeff Pferd, [Petris](#)

KM used to be called training! Petris uses web technologies to capture information on user patterns – workflow monitoring, search engine logging, application launching etc. in a ‘passive knowledge capture.’ This leverages ‘meta data mining’ – eg which departments are using what information etc. Examples were shown of user access to the AAPG’s set of publications which is available through Petris. Around 7000 documents have been downloaded in 3 years – various statistics on most accessed papers etc. presented.

Web technologies for information access – Ugur Algan, Landmark

(Paper not presented – from handout). Algan described Landmark's development of e Technical Workspace Portal for AGIP ENI (this was the subject of a paper in First Break, Jan 2003). The portal set out to offer geoscientists and engineers the ability to respond to business needs such as multi-database query, project preparation and asset disposal. Landmark's Team WorkSpace was customized to ENI's requirements. The portal was hosted internally by ENI in Milan. Queries are executable across PGS Tigress, IHS Energy's Iris21, C&C Reservoirs Analogues, in-house ENI data sources and Microsoft Office documents. More from [Ugur Algan](#).

Online Access to Well Log Data – Richard Herrmann, A2D Technologies, Gene Rhodes, Landmark

According to a study by McKinsey, "55% of industry costs are associated with information exchange – such exchange currently experiences many 'friction points'". The Landmark/A2D presentation went over similar ground as covered in our report from the 2003 Salt Lake City AAPG. Herrmann's presentation vaunted the merits of web services – and the Microsoft .NET environment which has been used to offer cross platform access from Landmark's interpretation environments to A2D's hosted well log libraries. .NET calls return well header, spatial extent, log curve, authentication and purchase order. Delivery is in a choice of TIF, SmartRaster, LAS or XML formatted logs. GIS publishing supports wells, XML, Shapefiles and ArcIMS. On the Landmark side, OpenWorks makes Java calls through, SOAP, RMI, XML and https. Herrmann commented on the large number of clicks and screens that were required to place an order from OpenWorks saying "Web Services should really be about computer to computer interaction – we are still some way away from the day when all data pops up into a 'meta-project' itself!"

Data viewers for wide area networks⁷ – Jim Velasco, INT

INT's products allow centrally managed data to be viewed over wide area networks, the Internet etc. The latest INT product line exemplified by the Web LogViewer, offers thin client access to data in multiple data stores. By 'thin client', INT means a J2SE 1.4 runtime, a web browser or WebStart browser. A 'lightweight' API (compared with INT's traditional offerings) allows for quick integration into enterprise solutions. The LogViewer provides area curve filling, lithology display, curve attribute editing, image integration and pan/zoom.

Q&A

What is the target market – developers for embedded deployment or shrink-wrapped for end users?

This product is 90-99% a shrink-wrap solution – as far as you can go in shrink-wrapping this kind of functionality. The tools still require configuration, but are designed to be customized by the IT department. Some J2EE knowledge is required – but further shrink wrapping would require drilling down into corporate security models which is undesirable.

Integrating disparate data repositories – Steve Hawtin, Schlumberger

Hawtin kicked off his talk with a very impressive bit of juggling, before outlining Schlumberger's 'streams'-based integration strategy for disparate data repositories. The strategy involves tailored workflow analysis and data streams configured and deployed as required. Data cleanup is done during data use – with back population to a consolidated single reference repository [*à la* Statoil].

Q&A

Data clean up should not be performed by end-users. "Users need to use data – cleanup should be done upstream by data managers."

What happened to Open Spirit? – this was predicated on the deployment of multiple repositories as being preferred over a single repository.

There has been some down-playing of the role to be played by Open Spirit, in some cases the cost of maintaining distributed data sets has proved too high – in reality "the devil is in the detail".

⁷ This paper was also presented at the 2003 San Francisco JavaOne conference.

[‘Let the engineers analyze the data’ – Tim Elser⁸, Pioneer Resources Canada](#)

Pioneer had its main IT components in place deploying best of breed applications, but using complex, manual data entry. However, there were infrequent data updates, and only annual reporting – a drive towards higher frequency reporting and greater granularity meant that manual data entry was no longer an option. The installed base was made up of Hyperion’s SBASE OLAP. Third party data mining tools lacked functionality – so Aclaro was called in. PetroLook and PetroShare provided required data mining functionality. PetroLook and PetroShare are both developed in Microsoft’s C#.NET⁹. The end result has ‘piqued the imagination of Pioneer’s data users – who no longer feel that disparate data stores are an obstacle to data access.

[IM – The *cultural* challenge – Flemming Rolle, Paras](#)

Rolle stated that most (all?) data management initiatives fall short of their initial objectives – if they do not just fall flat! Such failures can be explained by Roll’s formula viz:

Change=Discomfort+Vision+Capability+(a first step)+Energy and Enthusiasm

Rolle discussed various disconnects and other causes of discomfort to conclude that by aligning the vision, building a team, acknowledging work done and allowing time, project success was assured¹⁰. Oh we nearly forgot – ‘communicate, communicate, communicate!’.

[Dynamic Data Management – Carl Hucsall and Ray Cline, SAIC](#)

SAIC has been working on a project which sets out to ‘transform the digital oilfield of the future’. The aim is to link SCADA data feeds to the field office. SAIC recognizes four levels of operation value (and complexity!) from remote monitoring, through exception management and diagnostics to dynamic reservoir optimization. A tiered IT architecture is proposed. A resource layer provides base level IT services scheduling, transactions, triggers and exceptions. An intermediate services layer manages business logic and analytical services such as trends, curve fitting and historical analysis. Finally a user interaction layer provides modeling, simulation and ‘discovery’. These are delivered in a “web services situation – not as a behemoth of a database”. Real time information benefits production, sales trading and supply chain. Production is the first field where web services will contribute – the technology is here today.

[28 Spatial Search Technology – Bruce Sanderson, Geodynamic Solutions](#)

Search technologies start with vanilla Google-type full text searches – these are fast, free and easy, but often suffer from a large number of search results. Google on an internal intranet can help constrain searching – but these are expensive to deploy. Other options include document management systems such as OpenText’s LiveLink – this is ‘a great tool’ but may not always integrate the upstream workflow. But indexing and document check in/out is an unwelcome overhead. Enterprise search systems can be very powerful. These crawl file systems – looking in documents, web pages and databases – again these are expensive to deploy. Geographical Information Systems (GIS) can dovetail with enterprise search technology. Deployment requires a search ‘hub’ coupling GIS to text search. Best of all worlds would be something like [Autonomy](#)¹¹ plus a GIS.

⁸ Presented by Chris Faig, [Aclaro](#) Software.

⁹ We had a chat with Faig about how Microsoft’s programming environment has evolved with the introduction of .NET. The latest version of Visual Basic with its object oriented manipulation of Microsoft’s complex document object models has proved a steep learning curve – beyond many VB 3-6 hacker’s skills. This has effectively shut-out several of Aclaro’s hackers and contract workers – to the satisfaction of Aclaro’s VB .NET professionals!

¹⁰ It is interesting to turn some of these entreaties on their head to see how they sound. In other words if a consultant advises you to “align the vision” – what would you think of the opposite advice – “don’t align the vision”? If such an opposite is obviously wrong – then the original advice may be tending toward the platitude!

¹¹ It is our understanding that BP – a flagship Autonomy customer – found that the costs of this expensive and much-hyped tool exceeded the benefits.