

Session #1 – Data Mining / Threat Assessment

Session #1

Data Mining / Threat Assessment

Session Chair: Chad Zamarin

Facilitators: Mick Collins (GE), Jerry Rau (SUG), Sheila Wilson (PODS)

Government/Industry Pipeline R&D Forum - New Orleans, Louisiana, February 7-8, 2007

Session #1 – Data Mining / Threat Assessment

Attendance Breakdown

Approximate total attendance	30 persons
Federal Regulators	2 persons
State Regulators	0 persons
International Regulators	0 persons
Pipeline Industry	22 persons
Standard Organizations	2 persons
Researchers	4 persons
Academics	0 persons

50 gaps grouped into 9 categories with 3 receiving consensus endorsement

Session #1 – Data Mining / Threat Assessment

Top 3 Identified R&D Gaps

Gap #1 – Interchange standards and tools between operators and regulators (Technology and Consensus Standard)

Gap #2 – Advanced Risk Management Tools and Technologies (Technology, Consensus Standard and General Knowledge)

Gap #3 – Best Practices for Enterprise Integration (Consensus Standard and General Knowledge)

Session #1 – Data Mining / Threat Assessment

Interchange standards and tools between operators and regulators (Gap #1)

New or Improved Technology / New or Revised Consensus Standards

Target:

Oil & gas operators and regulators

Operating Environments:

Internal Operators, Internal Regulators, Web

Goals:

Standardize Reporting systems (incidents, annual reports, FERC, etc.)

Develop integration between operator data systems and regulator systems

Collaborative management of standards and system development

Shared tools for visualization, performance measurement and reporting (Across industry, within organization)

Reduce duplication of effort (less time gathering data, more time analyzing)

Road Blocks/Barriers

Political / Internal

Agreeing to appropriate levels of reporting details

Timeline

24 months

Government/Industry Pipeline R&D Forum - New Orleans, Louisiana, February 7-8, 2007

Session #1 – Data Mining / Threat Assessment

Advanced Risk Management Tools and Technologies (Gap #2)

Creation and Dissemination of General Knowledge / New or Revised Consensus Standards

Target:

Oil & Gas Operators

Operating Environments:

Internal Operators

Goals:

Holistic approach to risk mgmt / threat assessment (quantified industry data)

Evolution of algorithm standards

Predictive modeling systems

Systemic analysis of trends (“Google Search” your assets)

Implementation of quantitative and predictive models

Road Blocks/Barriers

Reaching consensus

Sensitivity and availability of information

Where to manage industry shared data

Timeline

24 months

Government/Industry Pipeline R&D Forum - New Orleans, Louisiana, February 7-8, 2007

Session #1 – Data Mining / Threat Assessment

Best Practices for Enterprise Integration (Gap #3)

Creation and Dissemination of General Knowledge / New or Revised Consensus Standards

Target:

Oil & Gas Operators

Operating Environments:

Internal Operators

Goals:

Develop a consensus roadmap for evolution of integrated data systems (multi-phased approach)
Address multiple reference systems and database standards
Reduce the integration pain
Methodologies, tools, standards

Road Blocks/Barriers

Reaching consensus

Timeline

24 months

Government/Industry Pipeline R&D Forum - New Orleans, Louisiana, February 7-8, 2007

Session #1 – Data Mining / Threat Assessment

Other Take-Aways

- Data Mining / Threat Assessment is dependent on (and limited by) the fundamental data management foundations and integration abilities.
- Historically, data management development lagged behind new technology introduction. Recognition that the data management component must be considered in other areas of R & D.
- Data quantity and complexity is growing and growing and growing.
- Consensus that in many areas...
 - Data already exists that is not fully leveraged
 - Technology already exists (maybe not in our industry) to bring data together and get more out of the existing data
 - Therefore effective technology transfer, high ROI and quick wins should be attainable

Additional Identified Gaps

1. Data Security Standards and Protocols
2. Communications / Bandwidth
3. GPS data accuracy, transformations, limitations (DOC, etc.)
4. Publicly available data sets / accuracy (Imagery, Landowner, Basemap, DEM)
5. Legacy data Issues
6. Many More...

“Spend more time analyzing and responding and less time gathering, correcting and integrating”

Appendix – Additional Gaps

New / Improved Technology (1)

High Level	Gap	
New / Improved Technology	Data Model for Transmission	Bill
New / Improved Technology	Data Model for Distribution	Bill
New / Improved Technology	Spatial Accuracy Coordinate Transform / Translation. Software standards aimed at leg GIS.	Bill
New / Improved Technology	Temporal Connectivity Persistence, Parent - Child (Historical data, maintain the data set with current pipe) Parent to child, in either direction, and look at how the pipeline has changed) Time dependent changes.	Bill
New / Improved Technology	Tools to decentralize data integration field data collection of that gets all the way into the corporate database. 300 people going via 2 GIS guys, and get the data out.	Gary
New / Improved Technology	Integration of traditional paper oriented data. OCR	Gary
New / Improved Technology	Data Quality - Tools to maintain / correct the data, quality. ETL (Extract, Transform, Load)	Craig
New / Improved Technology	Association of Critical data with applications that use this data (GIS, Risk, Asset Mgmt, SCADA)	Chris
New / Improved Technology	Grade & wall thickness from ILI runs	David
New / Improved Technology	GPS Signal Scrambling, accuracy of signal	David
New / Improved Technology	Communication speed & protocols to remote areas. Data to the field. Point of source data collections. Data in quick. On demand, linking to fields poits, CP, Uploads data from field.	David
New / Improved Technology	Ground penetrating Radar, grab the Z	David
New / Improved Technology	Collection and comparison of ILI data to understand repeatability of data on re-inspection with different inspection tools.	Bill Mohr
New / Improved Technology	Search in depth of "failures" to tell the deeper stories of why and how (show the love for internal investigations)	Bill Mohr
New / Improved Technology	Internal data on facilities to be improved including radiography and location of small pipe within facilities	Bill Mohr
New / Improved Technology	Blink corporate automation for aircraft photography for changes in ROW (Bill Mohr). Aerial Photography change recognition (David)	Bill Mohr

New / Improved Technology (2)

New / Improved Technology	Integration of databases - legacy system issues as well as changing definition of data elements	Piyali
New / Improved Technology	Spend more time analyzing data than gathering & correcting data	Sheila
New / Improved Technology	One call Rates - Ability to mine data to look for trends, predictive models.	Chad
New / Improved Technology	Standard and protocols for comparing field inspection results with ILI data for multiple threats. Feedback into the ILI.	Chad
New / Improved Technology	Legacy Data - Global involvement in leading this industry. Same faces. Tools to collect new data or convert old data. 90% of the time is used getting the data less important.	Craig
New / Improved Technology	Data Management Best Practices, How can we come up with strategies that can enhance company specific data challenges?	Craig

Reporting / Sharing

High Level	Gap	
Reporting / Sharing	Ability to Automate the reporting... and the receipt of the report by PHMSA. How can we make this an automated transaction. That hits all reg'd enviroments. Intercompany reproting..points that touch PHMSA, FERC and others. Fed - State; Better communication of data between different government agencies; federal - federal - state - federal	Ed
Reporting / Sharing	Understanding of core technology capabilities, available public data. What public data, accuracy can be leveraged by the pipeline company.	Ed
Reporting / Sharing	Reporting tools to pull data into meaningful formats, dashboard concepts maybe? Have data presented in a meanifull format. Leadership reporting. Standard output for reporting, mining. (PPTS, DOT, Internal Operator)	Gary
Reporting / Sharing	Fuels - Relationship with material for transportation of new fuels	Emily
Reporting / Sharing	Normalizing leakage history (Distribution) for trend analysis	Ed Newton

Reporting / Sharing (1)

High Level	Gap	
Standards / Guidelines / Practices	Before data is integrated "completely" data can be aggregated. Process to get them to the dream of a enterprise solution. Tools that cut across data sets.in the interium period. Model independent. Road map	Greg Smith
Standards / Guidelines / Practices	How to "remove" a threat even with indications that the threat should be considered. Joker in the data set. One indication that offsets a result. Need to look at other data sets that would remove this data, outlier	Sean
Standards / Guidelines / Practices	Management of Data Change:(Chris); Management of change/workflow tracking (Gary) , HCA Changes (eg, need to justify changes, answer, why these changes have occurred.	Chris
Standards / Guidelines / Practices	Operators are not clear what constitutes Critical Data sets & how they are managed. What is critical v necessary v enhancement.	Chris
Standards / Guidelines / Practices	Consistency / Clarity in data definition. Common use of terms, across the board use of term.	Piyali
Standards / Guidelines / Practices	Better communication of data with & between government agencies & vendors. Leverage data across more of a customer/vendor etc base. Grab more data, how accurate is it.	Sheila
Standards / Guidelines / Practices	The challenge of mapping legacy mechanisms of asset locations (linear referencing like engineering stationing) lat/long or odometer (LI) is real. Until the field embraces a different reference mechanism, stationing will be required. Each company must decide to "phase out" the use of engineering stationing eg. before a common GPS Focused mechanism can be effective. Change Management at its best. A common reference. Therefore agree that this value is ref'd xxxxx not xxxxxx	Greg Smith

Reporting / Sharing (2)

Standards / Guidelines / Practices	Risk/Threat Assessment algorithm methodologies for multiple levels of analysis (ranking, predicting, modeling). Use of available corrosion data to penetrate statistical distributions - better integrity model - risk assessment (Re thinking of what we to do with the data collected, inspection data). Holistic defect database, trends, stat's. Predictive	Mariano
Standards / Guidelines / Practices	Security in data transferring - encryption. What level do we need to set for the various data sets we want to share, as well as gather.	Mariano
Standards / Guidelines / Practices	Data Integration Standards, For business decisions "data integration" seems to be the key concept. Data integration, is making decisions from from multiple data sets that you can't make from just one (Brian). How do we ensure / improve confidence that data integration that is needed can be adapted (Sean). Integration of databases/Standard (Mariano) - End users need to understand what accuracy is needed....	Brain Sitterly
Standards / Guidelines / Practices	Threat specific Data Requirements (required, recommended, etc.). Critical data needed.....	Chad Zamarin
Standards / Guidelines / Practices	Data format standards or requirements (For all types of operators, Transmission, etc...) Look to resolve the gap between Trans, Liquid, Gs...	Chris
Standards / Guidelines / Practices	Institutional barriers to unified data within enterprise. This may be data education. Owners, trust the data, owners a fine with sharing. Aim to go to a mass sharing etc.	Chris