Lessons Learned: Creating the Digital Twin Concurrently for Pipeline Projects

Keith Winning

Pipeline Data Solutions

Esri European Petroleum GIS Conference

London, October 30-31, 2018

Lessons Learned: Creating the Digital Twin Concurrently for Pipeline Projects

Keith Winning – Managing Director, Pipeline Data Solutions

Abstract: This paper reviews the lessons learnt in the early adoption of pipeline data modelling for a major pipeline project. It addresses the challenges and enumerates the benefits of undertaking the data modelling concurrent with the design and construction phases of the project. Potential solutions to some of the challenges are also presented, using a combination of revised process, procedures and the adoption of emerging technology.

Keywords Pipeline Data Modelling • PIM • Data Interoperability • PODS

Keith Winning (✉)

Managing Director, Pipeline Data Solutions

27, Old Gloucester Street, London WC1N 3AX

e-mail: kwinning@PipelineDataSolutions.com

Preface

This paper was presented at the 2018 Esri European User Conference in London. The presentation slides are attached at the end of this paper.

Glossary

APDM	ArcGIS Pipeline Data Model
GIS	Geographical Information System
PIM	Pipeline Integrity Management
PODS	Pipeline Open Data Standard
QA	Quality Assurance
QC	Quality Control

1. Introduction

Central to effective pipeline integrity management (PIM) is the need to manage, analyse and disseminate pipeline geospatial data (Perich, Van Oostendorp, Puente, & Strike, 2003). To meet this requirement, pipeline operators use geographical information systems (GIS) and pipeline specific data models, such as the ArcGIS Pipeline Data Model (APDM) and the Pipeline Open Data Standard (PODS).

Early implementations of these pipeline data models sought to provide an inclusive data structure to meet all pipeline requirements. This inevitably led to complex models with high levels of redundancy. Despite the comprehensive nature of the models, pipeline operators still needed to add additional corporate specific data. Our data models are now evolving to reflect this, focusing on providing the core structure and functionality around which the pipeline operator can build their own specific data model.

With the growing use of pipeline data models to manage existing assets, the challenge now is to model the pipeline through the design and construction phases of the project. Leveraging the value of the pipeline data model in the earlier project phases offsets the cost of building the model for the operational phase of the project. The challenges of building a pipeline data model concurrent with the construction of the pipeline are varied and numerous and form the basis of this paper.

1.1 Entropic Data

The concept of entropic data was introduced in the paper *'The Challenge of Managing Entropic Data: A Pipeline Case* Study', presented at the ESRI European Petroleum User Group conference in 2017 (Winning, 2017).

Entropy, as a thermodynamic concept is defined as:

> *"A thermodynamic quantity representing the unavailability of a system's thermal energy for conversion into mechanical work, often interpreted as the degree of disorder or randomness in the system."*
>
> *(Oxford Dictionaries, 2010)*

and in a more general sense may be defined as the

> *"Lack of order or predictability; gradual decline into disorder."(Ibid)*

The second law of thermodynamics states that entropy always increases. Although the increase mentioned in the second law is time, for geospatial data, Tobler's first law implies that entropy increases with distance (Miller, 2004).

The term entropic data is being used, not to describe data that contains errors, but geospatial data that exhibits a randomness in the error distribution, with errors tending to increase with both time and distance from the source data **(Slide 12)**.

> ***Entropic data has unpredictable error distribution which increases with both time and distance from the source of the data.***

An example of Entropic Data are the weld numbers. These are initially generated by hand on the pipes (Figure 1) during construction and are then subsequently transcribed by hand. Finally, they are computerised and entered into the pipeline tracking system **(Slide 13)**.

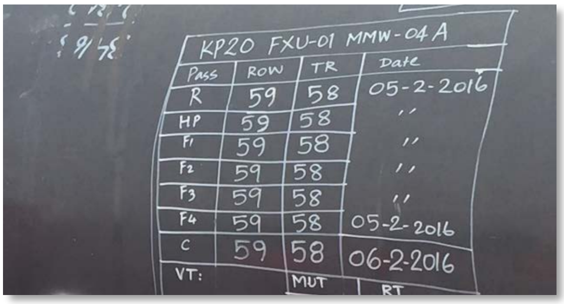


Figure 1 – Weld Data

2. Project Scope

The case study presented in this paper involved the construction of a 480km 48" diameter gas pipeline. This required approximately 40,000 lengths of pipe, 350 induction bends, 50,000 welds and numerous valves and fittings. These were stored in 11 models with a combined record count in excess of 1.5 million **(Slide 4)**.

3. Approaches

There are a number of approaches to pipeline data modelling, based on the project phase when the modelling commences. Historically, pipeline data modelling has grown out of the as-built process, aiming to provide better quality asset data for the Pipeline Integrity Management (PIM). As such, it was usually part of the post construction activities.

3.1 Post construction

With the late adoption of the data modelling, the ability to realise value against the modelling cost is reduced. The impact of this is two-fold. This approach means that the data modelling is only available to support this final project phase, increasing the cost to value ratio. Crucially however, it prevents the verification of the construction to the design intent. This can be a benefit to construction management, as it can help to manage change and its associated cost **(Slide 5)**.

3.2 Concurrent with construction

With this approach the quality of the data model is increased as errors can be resolved at source with the installation contractor, if required. Also, it can assist in the construction management through the use of geospatial progress reporting of the individual construction activities **(Slide 6)**.

3.3 Concurrent with design

The early adoption of the data modelling in the design phase offers the greatest return on the data modelling investment. In addition, it significantly improves the quality and the completeness of the data model. It enables the verification of the construction to the design intent, thereby ensuring the installation is compliant. Examples of this could be:

- The identification of the installation not meeting the specification break requirements for the installation of low temperature pipe above ground.
- The extents of the installation of the heavy wall pipe, either to code or design requirements.
- Confirmation of the minimum depth of cover requirement along the pipeline.
- Installation of mechanical protection at crossing locations.

While these issues are a function of the quality management of the installation contractor, they are easier to identify when looking at the data holistically. This is a major benefit of the early adoption of the data modelling **(Slide 7)**.

A summary of the characteristics and benefits of the modelling by project phase are tabulated in Table 1.

Pipeline Data Modelling Characteristics by Project Phase	Post Construction	Construction	Design
Data			
Completeness	×	×	✓
Quality	×	✓	✓
Applicable Project Phases			
Asset Management	✓	✓	✓
Construction Management	×	✓	✓
Construction Verification	×	×	✓
Record of Design	×	×	✓

Table 1 – Modelling Benefits by Project Phase

4. Challenges

This approach presents a number of challenges, both internal and external. On major projects, the duration from the start of detail design through construction and commissioning can be significant.

This presents issues in being able to adopt schema changes during the project data modelling process. One method, is to freeze the schema at the start of the process. While this addresses the issue, it presents problems for operators with large numbers of assets from optimising and improving the schema, particularly when they have multiple concurrent projects.

The other main internal challenges relate to the management of the data modelling. While creating the model in parallel with construction leads to improved quality and completeness, the issue of re-cycled data submissions needs to be addressed.

This data re-submission requires additional re-work and leads to resource loading issues, potentially impacting on both cost and schedule.

There are a number of key external challenges with modelling during the construction phase. There is the inherent uncertainty of data quality and phased delivery. Although an appropriate level of quality should be assumed, procedures and processes need to be in place to mitigate and deal with data quality and schedule issues.

The main data quality issues were:

- Out of sequence data submission, leading to stationing problems.
- Re-cycling of data submissions from the contractor and the associated re-work involved.
- Record level dataset change management.
- Incentivising data providers to submit data to the agreed quality and schedule **(Slide 8)**.

5. Lessons Learned

The lessons learned from this project fell into five categories.

5.1 Data

> *"All animals are equal, but some animals are more equal than others."*
>
> *(Orwell, 2018)*

With limited time and resources, it is important to differentiate between primary and secondary data within the model. By doing this it is possible to focus the checking and validation on the key components of the system. In order to do this, a real understanding of the use of the model during the operation of the asset is essential. This approach would enable resource constrained projects were able to prioritise the primary data quality reviews.

However, this is neither project or record specific. Therefore, consideration should be given by schema providers to assist in the identification and differentiation of primary and secondary tables.

Data will always contain errors and omissions, so this should be planned for and procedures and processes put in place before the start of the data modelling. The treatment of these errors should be driven by their priority as primary or secondary datasets.

Datasets were found to exhibit entropic properties (discussed in Section 1.1). This can be useful, as the level of entropy for a given dataset can be calculated and used to select the best primary keys within the datasets to minimise the number of errors when performing joins.

Finally, the use of data templates for collecting contractor supplied data proved beneficial in simplifying the data gathering requirements but did lead to issues of data interoperability induced errors. A simple solution to this is to add an alpha character to a padded numerical field. This ensures that there was no translation error on transfer between systems as the value is now seen as a text field **(Slide 11)**.

5.2 Procedures and Processes

5.2.1 Data Acceptance and Review

One of the key procedures to establish is that of data acceptance and review. These should be based on, and acknowledge the concept of primary and secondary data within the schema. It is recommended that a 'reject and contractor fix' approach be adopted initially for primary data. For secondary data, a more pragmatic method is to adopt an 'accept and fix' approach. This also applies where primary errors are not being fixed by the contractor or data supplier.

Another significant challenge is tracking changes within the contractor supplied datasets for different revisions. While procedures can be put in place, there must be an awareness that any system is fallible. To this extent, automated processes should be developed to identify record level changes between different revisions of datasets. This is further complicated as the formats of the datasets (if supplied as spreadsheets or delimitated files) can change.

Central to the integrity of the model, is the requirement to record at record level any changes or fixes to the supplied data **(Slide 15)**.

5.2.2 Tagging

Given the importance of tagging within the data model for performing joins, any issues relating to missing, duplicate or incorrectly formatted tags causes a significant issue. At present (within the PODS schema) only one alternative asset tag is in the schema and there is no reference association.

It is recommended that external tags (weld IDs, Pipe Numbers, etc) are stored in a CrossRef table against the document for which the tag is associated and linked to an internally created TagID which guarantees uniqueness across the entire data model. This would also support multiple tags for each item **(Slide 16)**.

5.2.3 Automated Processes

To improve quality and reduce cost, automated processes that can be adapted to meet project specific criteria need to be developed. These should address the key areas of:

- Validating, importing and storing of vendor data prior to the availability of the geospatial component.
- Providing temporary tables for the storing of data duplicated across multiple tables, thus ensuring consistency.
- The creation of internal TagIDs with alpha prefixes based on the feature class.
- Database checking for referential integrity, domain values and out of bound values.

- Web-based progress reporting of construction activities by type, based on the contractors daily/weekly reports.
- Reporting model completeness and build progress and visualisation of the data model using web-based tools and automated production of drawings directly from the data model **(Slide 17)**.

5.3 System

5.3.1 Schema

The PODS schema contains significant duplication of information across multiple tables, leading to an additional overhead to ensure consistency. Use of additional CrossRef tables would reduce this and reduce the model build effort. In particular consideration needs to be given to the handling of tags, cut pipes and coating records.

The addition of separate event tables for handling interface changes, specification breaks (ASME B31.3/ASME B31.3) and design data (temperature, pressure, etc) should also be considered. The current approach of the information residing on each individual pipe presents an unnecessary overhead and leads to potential errors.

The preferred solution would be the revision of the schema so that this could be leveraged across projects. Where this is not possible, additional tables could be created, conforming to the schema rules on a project basis to provide a single source for this duplicated data **(Slide 19)**.

5.3.2 Technology

Many of the errors fond in the data could have been reduced at source had there been a better use of technology. Products such as ESRI Survey 123 and Collector are easy to configure and would greatly assist in this. This approach would potentially reduce errors, improve the data transmission rate and ensure that the data was more structured.

However, in spite of the compelling evidence to support this approach it requires additional costs for equipment, software and training in addition to a change in the current working practices common in the pipeline construction world. This will surely improve with time **(Slide 20)**.

5.4 Interfaces

As with any activity, the identification and management of interfaces is critical for delivery of the project. Areas that should be considered are the documentation of roles and responsibilities with identified methods for escalating problems to management. The primary data should be prioritised and if necessary be subject to increased levels of supervision and checking.

A key issue with respect to data delivery is the value that the contractor puts on it. It is recommended that consideration is given during preparation of the construction contract to the linking of data delivery and acceptance by company to defined key activities. Linking data delivery and acceptance to hydrotesting of a section would ensure that the contractor built this into his schedule. The contract could also provide some form of incentive to promote the data delivery **(Slide 22)**.

5.5 Resources

Key to the successful modelling of any system is a deep understanding of both the real-world asset, through all project phases and its digital twin. It also requires both knowledge and experience in a wide range of systems and skills covering data schemas, software, web applications, data management and programming.

This is best achieved by using an integrated team of specialist pipeline engineers, GIS analysts and engineering designers with significant knowledge of the complementary disciplines to provide a holistic solution **(Slide 24)**.

This level of resourcing requirement also presents a challenge. Not only is it difficult to find individuals with the right combination of experience and knowledge that the team requires, but retention can be a risk to the successful outcome of the project.

This risk could be mitigated through the use of specialist pipeline data providers. This approach could also deliver cost and schedule benefits at project phase boundaries. With a single data modelling provider throughout the project life-cycle, there would be no changes to the procedures, processes or team at these interfaces **(Slide 25)**.

6. Conclusion

In conclusion, the key lessons may be summarised as **(Slide 26)**:

- Managing and incentivising data providers.
- Understanding and using data entropy to your advantage.
- Acknowledging the data hierarchy and directing focus accordingly.
- Accepting and planning for data that will contain gaps and errors.
- The use of integrated teams with expertise in pipeline engineering and data modelling.
- The use of robust procedures supported by automated processes.

The lessons learned covered almost every aspect of the project. This is not to say that the project was not successful, on the contrary it was very successful, but possibly its true success will be measured in the improvement that it will bring to the projects to follow.

References

Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers, 94*(2), 284-289. doi:10.1111/j.1467-8306.2004.09402005.x

Orwell, G. (2018). *Animal Farm: Special Edition*: CreateSpace Independent Publishing Platform.

Oxford Dictionaries. (2010). *Oxford Dictionary of English* (A. Stevenson Ed.). Oxford: Oxford University Press.

Perich, W., Van Oostendorp, D. L., Puente, P. P., & Strike, N. D. (2003). Integrated data approach to pipeline integrity management. *Pipeline & gas journal, 230*(10), 28.

Winning, H. K. (2017). *The Challenge of Managing Entropic Data: A Pipeline Case Study*. Paper presented at the ESRI European Petroleum User Group Conference, London, 1st - 2nd November. http://proceedings.esri.com/library/userconf/epug17/papers/epug_14.pdf

Keith Winning is the Managing Director of Pipeline Data Solutions and has almost 30 years' experience in the field of pipeline design and data management. During his career he has been involved in all aspects of the pipeline project life-cycle for major projects in the UK, Europe, the Former Soviet Union, Africa, the Middle East and the Far East.

He has Master's degrees in both Mechanical Engineering and Geographical Information Science and a PhD in the *'Application and Development of Advanced Engineering Geographical Information Systems for Pipeline Design.'*

He is a Chartered Engineer, Environmentalist and Geographer, and Fellow of the Institution of Mechanical Engineers, the Institution of Engineering Designers and the Royal Geographical Society. He has authored and reviewed a number of technical papers as well as presenting at major international conferences.

His unique range of experience, academic and professional qualifications make him one of the leading exponents in the field of pipeline data management and enable him to provide technical leadership to an integrated tripartite team of engineers, engineering designers and GIS analysts.

Pipeline Data Solutions specialise in managing complex data to deliver quality solutions to the pipeline industry. Utilising both our integrated proprietary systems and some of the leading exponents in the field of pipeline data management, we are able to offer an unparalleled service combining both quality and value.

We offer a complete range of pipeline geospatial data services for the oil and gas industries, including:

- Capture & Cleansing
- Creation
- Validation
- Management & Maintenance

We can support all phases of the project life cycle, from feasibility and FEED through detail design and construction to operation.

We can support all pipeline database schemas, including PODS, APDM, UPDM and client specific schemas.

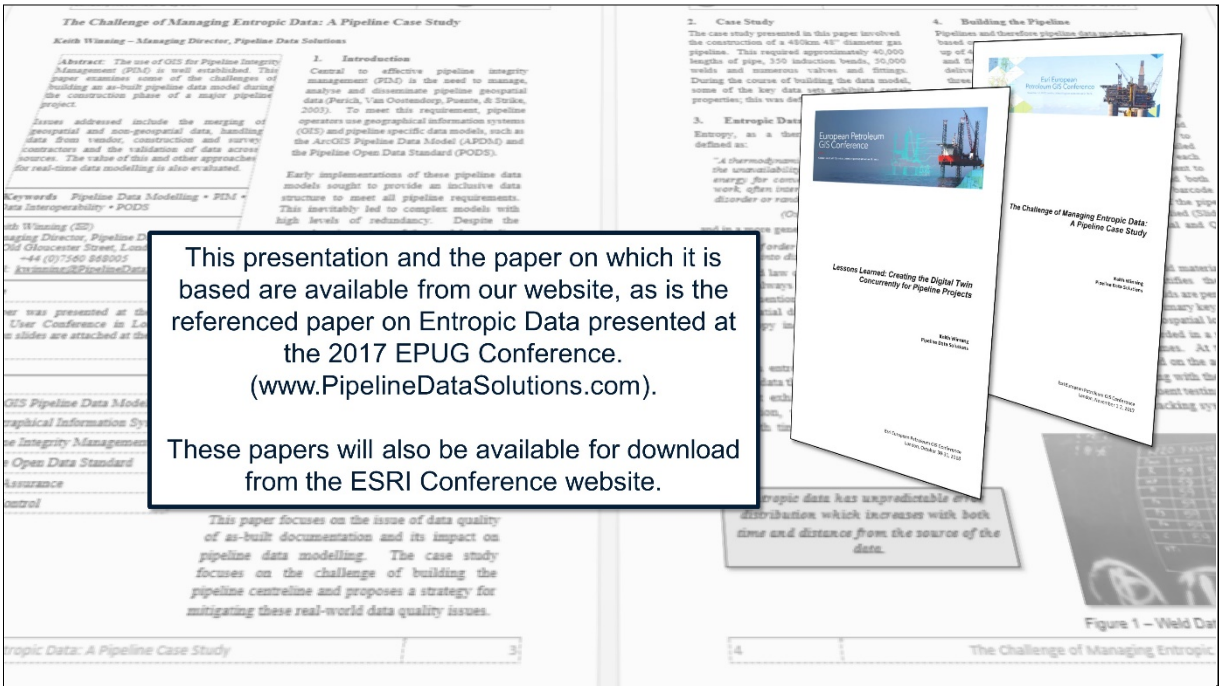
Centralising the data management processes across the entire project life cycle improves both data quality and integrity. It also reduces costs and enables clients to leverage greater value from their data management investment.

For more information on our services please visit our website at:

www.PipelineDataSolutions.com

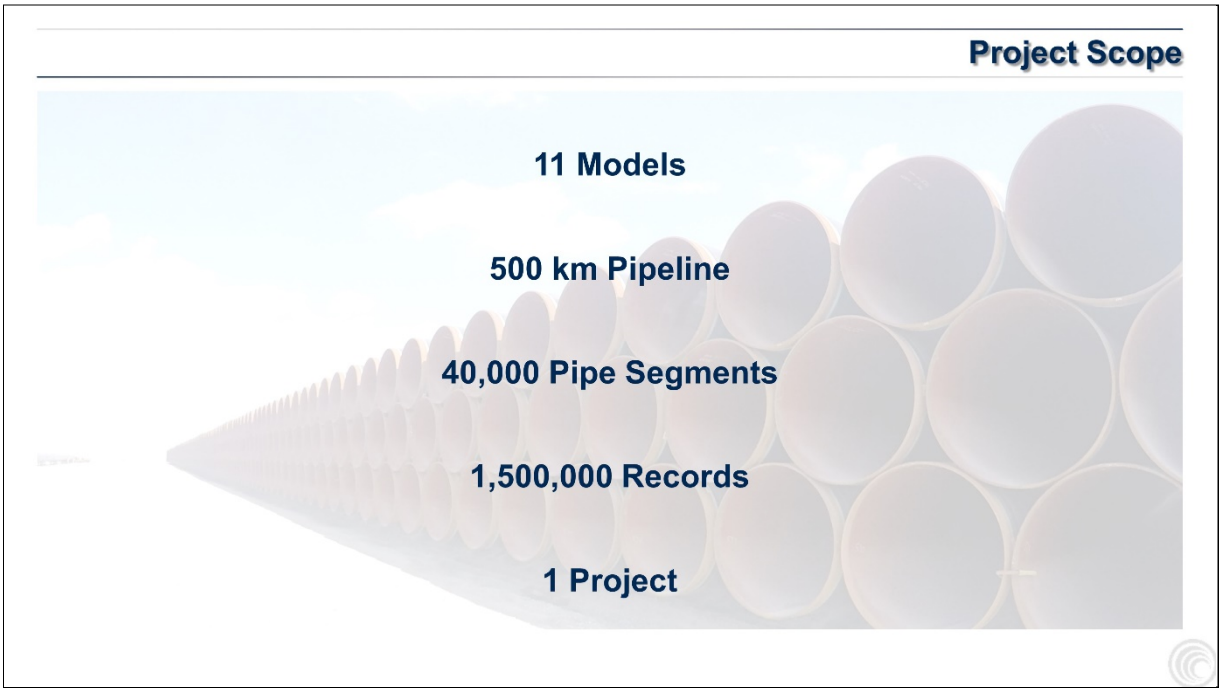
Appendix – Presentation Slides

Slide 1

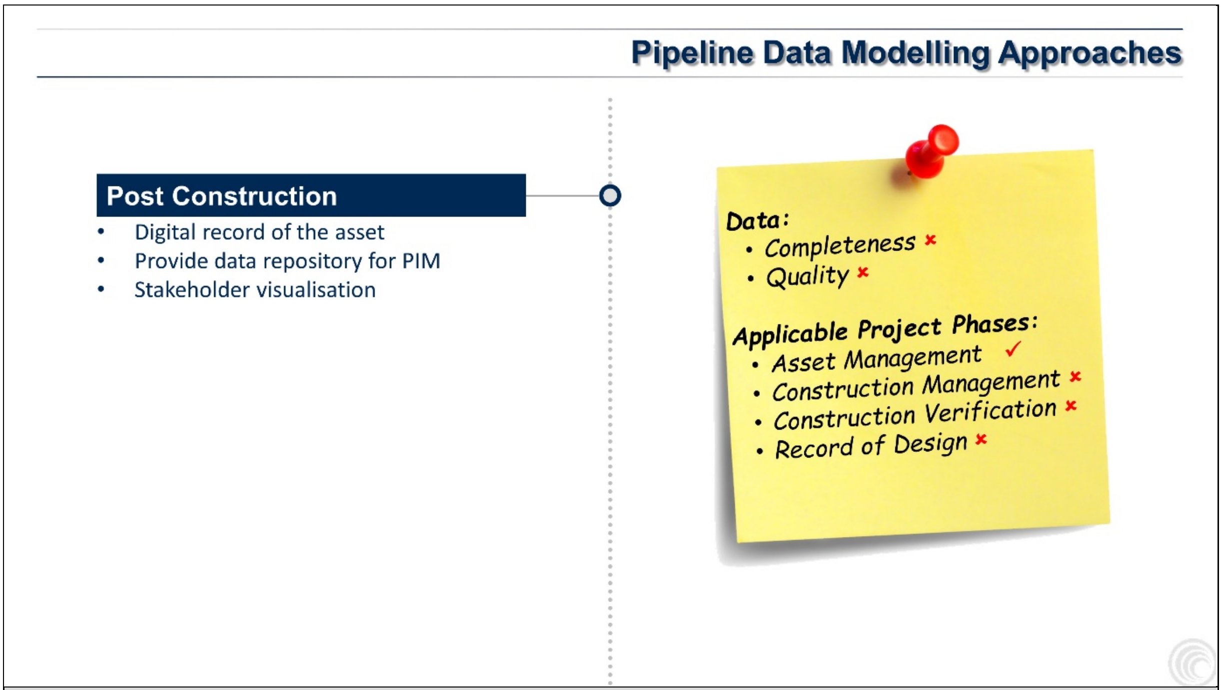


Slide 2

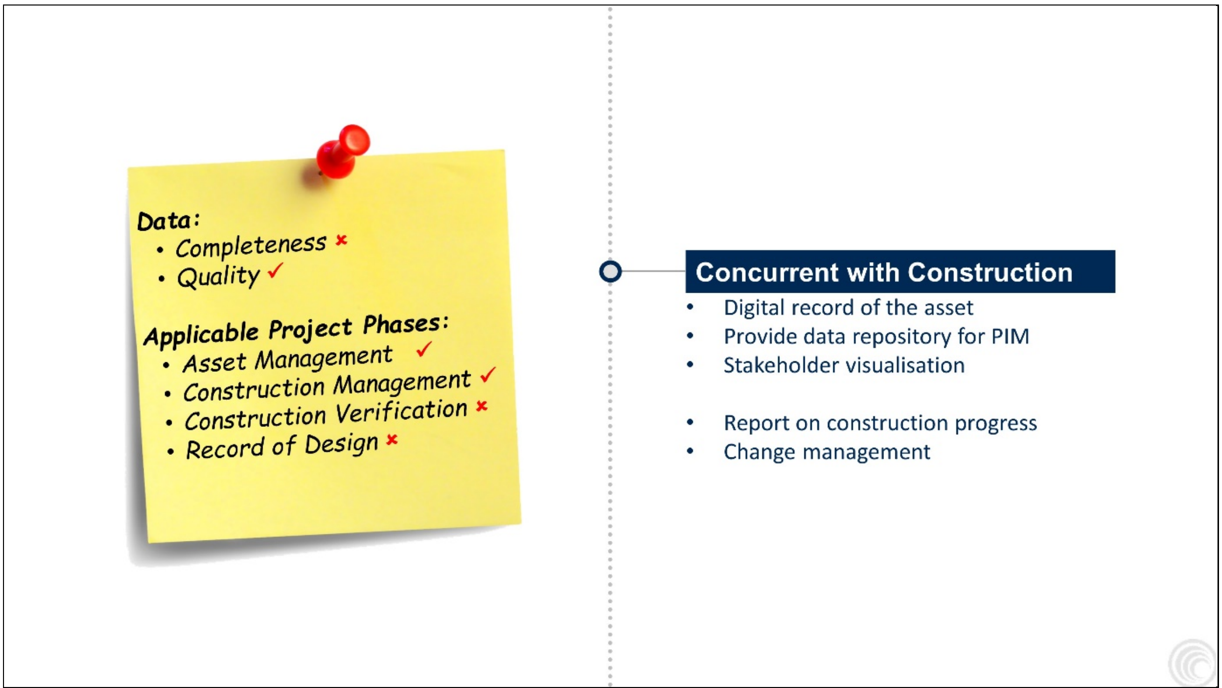
Slide 3



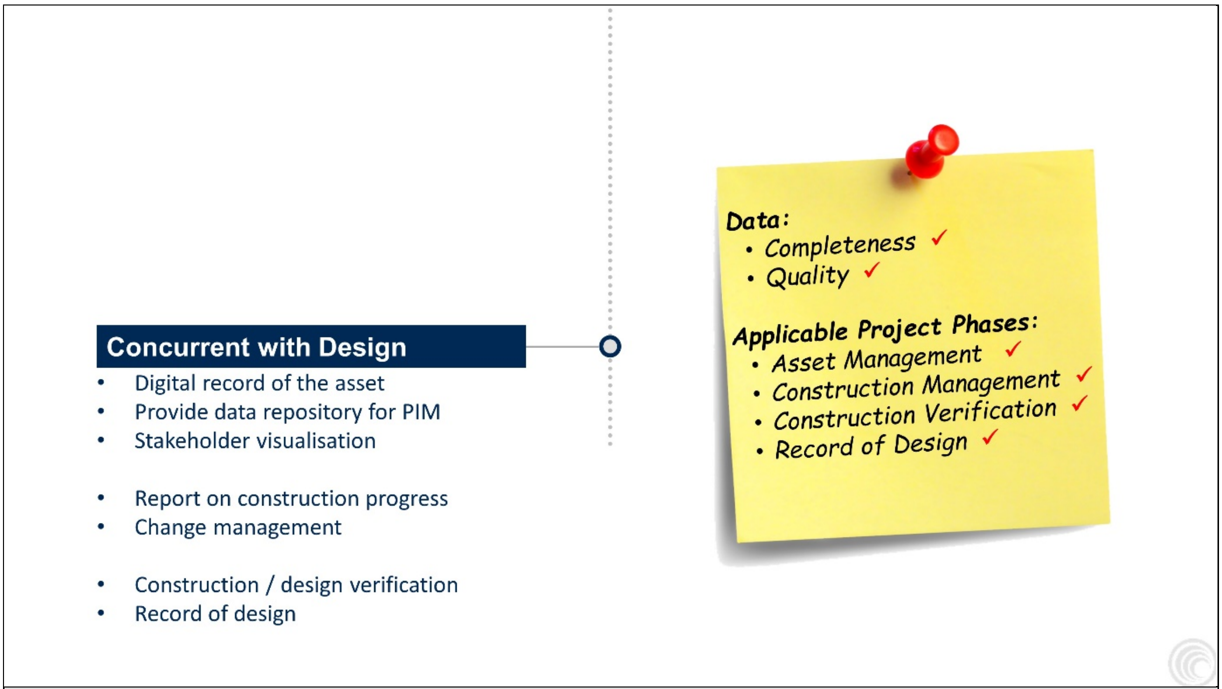
Slide 4



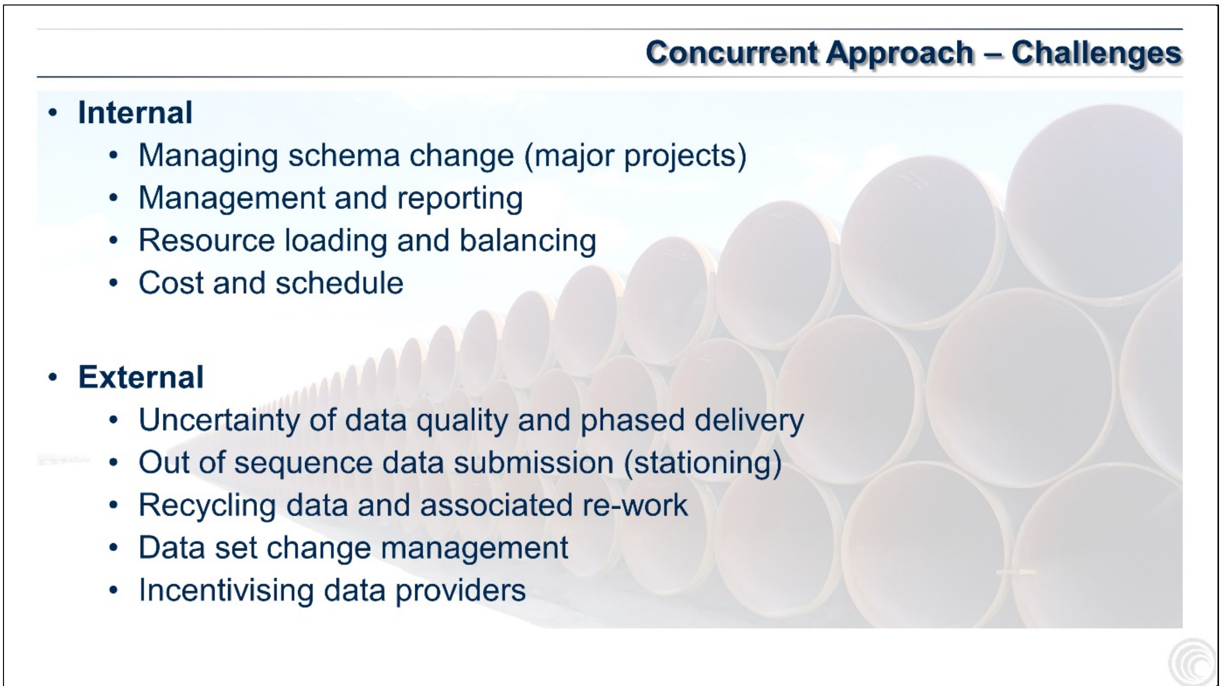
Slide 5



Slide 6



Slide 7



Slide 8

Slide 9

Slide 10

Data – General

- Differentiate between primary and secondary data
- Prioritise primary data
- Plan for errors and omissions
- Identify potential interoperability induced errors
- Identify and measure entropic data

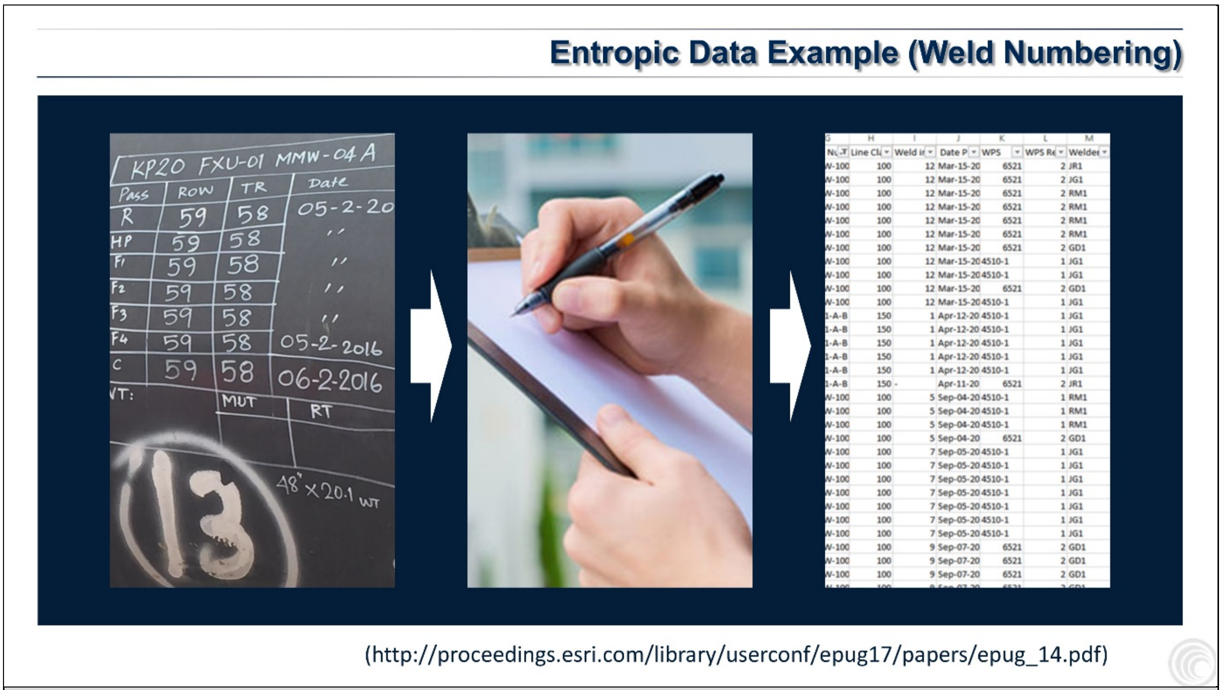
Slide 11

Data – Entropic Data

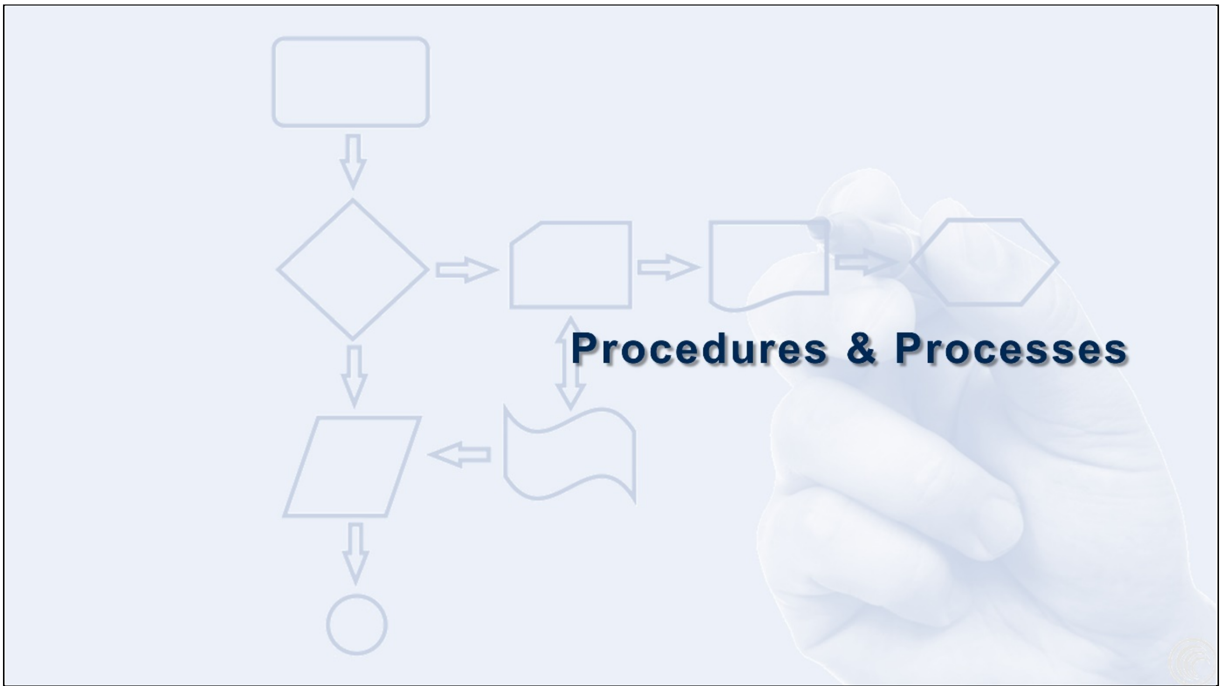
Entropic data has unpredictable error distribution which increases with both time and distance from the source of the data

- Data entropy cannot be removed but it can be managed
- Data entropy can be measured
- Processing effort is inversely proportional to data entropy
- Selection of primary keys based on data entropy

Slide 12



Slide 13



Slide 14

Procedures – Data Acceptance and Review

- **Define data acceptance procedures based on data classification**
- **Adopt 'reject and contractor fix' approach for primary data**
- **Adopt 'accept and fix' approach for secondary data**
- **Define change tracking procedures**
- **All fixes to be recorded in record level metadata**

Slide 15

Procedures – Tagging

- **External Tagging**
  - Manage multiple tags for the same item
  - Store tag ID against source document (additional CrossRef Table)
- **Internal (Data Model) Tagging**
  - Develop and document internal (data model) tagging procedure
  - Fixed length with alpha characters to improve interoperability
  - Guaranteed unique across disparate data sets
  - Use for all joins

Slide 16

Processes – Develop Automated Solutions

- **Vendor Data**
  - Validating, importing and storing
- **Duplicated Data**
  - Storing in a single location for cross population
- **Tagging**
  - Internal data model Tag IDs
- **Checking**
  - Referential integrity, domains and out of bound values
- **Reporting and Visualisation**
  - Progress by activity (modelling, RoW, stringing, welding, etc)
  - Web interface and drawings

Slide 17

Slide 18

Systems – Schema

- **Additional CrossRef Tables**
  - Support one-to-many relationships (tags, cut pipes and coating material)
- **Separate (Additional) Event Tables**
  - Interfaces (construction contractor)
  - Specification breaks (ASME B31.3 / ASME B31.8)
  - Design data (temperature, pressure, design factor, etc)
- **Alternative Approach**
  - Single source temporary tables

Slide 19

Systems – Technology

Use of Technology to Collect Field Data (ESRI Survey 123 and Collector)

- **Benefits**
  - Reduced errors
  - Improved data transmission rate
  - Structured data
- **Challenges**
  - Equipment / software costs
  - Training
  - Change in existing working practices

Slide 20

Slide 21

Interfaces

- Identify and Manage Interfaces
- Document Roles and Responsibilities
- Prioritise Primary Data
- Link Construction Data Delivery and Acceptance to Activity
- Incentivise Data Delivery in the Contract

Slide 22

Slide 23

Resources – Requirement

- **Significant Knowledge and Experience**
  - Pipeline engineering (design and construction)
  - Pipeline data modelling (databases and schemas)
  - Engineering, GIS and CAD software
  - Additional skill sets (web apps, data management and programming)
- **Integrated Team of Pipeline Engineers and GIS Analysts**
  - Deep understanding of both the data and the model
  - Knowledge of complementary disciplines to manage interfaces
  - Provide holistic solution

Slide 24

Resources – Challenges and Solutions

- **During Model Build**
  - Change of key personnel due to project duration
- **Project Phase Boundaries**
  - Change in procedures, processes, systems and personnel
- **Specialist Pipeline Data Management Services**
  - Mitigate loss of key personnel
  - Can provide continuity across project phase boundaries

Slide 25

Key Findings …

- Managing and incentivising data providers
- Understanding and using data entropy to your advantage
- Acknowledging the data hierarchy and directing focus accordingly
- Accepting and planning for data that will contain gaps and errors
- The use of integrated teams with expertise in pipeline engineering and data modelling
- The use of robust procedures supported by automated processes

Slide 26


PDS
MANAGING COMPLEX DATA | DELIVERING QUALITY SOLUTIONS
www.PipelineDataSolutions.com | info@PipelineDataSolutions.com | 27 Old Gloucester Street London WC1N 3AX

Slide 27