The Challenge of Managing Entropic Data: A Pipeline Case Study

Keith Winning
Pipeline Data Solutions

Esri European Petroleum GIS Conference
London, November 1-2, 2017

The Challenge of Managing Entropic Data: A Pipeline Case Study

Keith Winning – Managing Director, Pipeline Data Solutions

Abstract: The use of GIS for Pipeline Integrity Management (PIM) is well established. This paper examines some of the challenges of building an as-built pipeline data model during the construction phase of a major pipeline project.

Issues addressed include the merging of geospatial and non-geospatial data, handling data from vendor, construction and survey contractors and the validation of data across sources. The value of this and other approaches for real-time data modelling is also evaluated.

Keywords Pipeline Data Modelling • PIM • Data Interoperability • PODS

Keith Winning (✉)
Managing Director, Pipeline Data Solutions
27, Old Gloucester Street, London WC1N 3AX
Tel: +44 (0)7560 868005
e-mail: kwinning@PipelineDataSolutions.com

Preface

This paper was presented at the 2017 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.

This paper focuses on the issue of data quality of as-built documentation and its impact on pipeline data modelling. The case study focuses on the challenge of building the pipeline centreline and proposes a strategy for mitigating these real-world data quality issues.

2. Case Study

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. During the course of building the data model, some of the key data sets exhibited certain properties; this was defined as entropic data.

3. Entropic Data

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 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.

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

4. Building the Pipeline

Pipelines and therefore pipeline data models are based on a centreline. This centreline is made up of 4 main components: pipes, bends, valves and fittings; all of these are long-lead items, delivered to site prior to construction. There are three main data components to the pipeline centreline.

4.1 Vendor Data

These components are manufactured to the required design parameters in pipe mills, and fabrication facilities. They are subject to stringent QA and QC processes with detailed documentation being provided for each component. This requires each component to be tagged. Pipes are typically tagged both inside and outside with stencils and a barcode placed on the outside near the end of the pipe were coating will not be factory applied (Slide 6). This data provides the material and QC properties of the components.

4.2 Weld Data

In addition to providing the weld material and QC properties, it also identifies the join between components. The welds are performed in the field and provide the primary key for both the welding data and the geospatial location of the welds. These are recorded in a number of places and at different times. At the time of welding the weld is noted on the adjacent pipe **(Figure 1 / Slide 8)** along with the QC data for the weld and its subsequent testing. This is then recorded in the pipe tracking system.

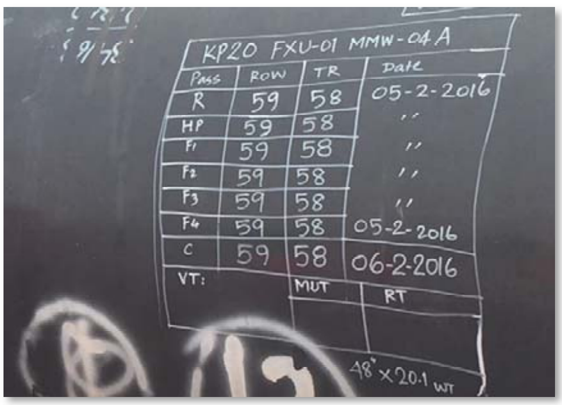


Figure 1 – Weld Data

4.3 Survey Data

Finally, once the pipeline is installed in the trench the geospatial location of the welds is recorded by the post installation survey crew. It should be noted that at the time the post installation survey is carried out, the trench is normally backfilled to the crown of the pipe for safety reasons. This final dataset provides the geospatial component to the pipeline centreline.

5. The Challenge

The weld and survey data is submitted on a regular basis but inherently there are a number of issues.

- The datasets do not have common boundaries due to the temporal differences in the collection of the two datasets. In addition, they will have gaps were tie-ins are still required.
- They are bi-directional, insofar as the data is not necessarily collected in station or chainage order, to reduce the data collection effort.
- The three datasets do not have common primary keys and there are frequent errors in the primary key fields **(Slides 12 & 13)**, particularly where these are collected during or after construction (weld and survey data).

6. The Solution

The key to meeting this challenge is to recognise that the data is entropic and to select primary keys based on the level of data integrity.

6.1 Correcting the Weld Data

The weld data contains references to the two elements that have been joined by the weld, with the weld being the primary key. The component identifiers (pipe, bend, valve or fitting) can be compared to the inherently more accurate vendor data. Where no match is found, the component identifier is corrected by inspection. While it is possible to identify potential candidates for the component identifier which is incorrect using fuzzy match criteria code, it ultimately requires an element of manual intervention to resolve these issues **(Figure 2 / Slide 15)**.

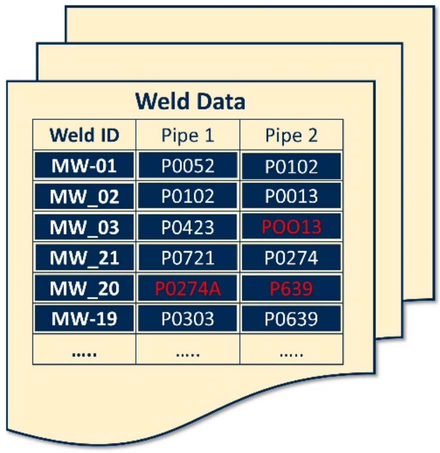


Figure 2 – Correcting Weld Data

6.1.1 *Create Pipe Data from Weld Data*

Having corrected any pipe/component identifier errors in the weld data, the pipe data is created with the two pipe/component identifiers becoming the primary keys for this dataset **(Figure 3 / Slide 16)**.

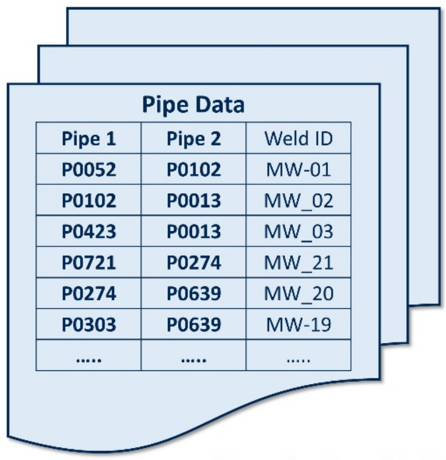


Figure 3 – Creating Pipe Data

6.1.2 *Identify Start & End of Strings*

Using this dataset, another dataset is created for those pipe/component identifiers that only appear once. These elements only have one join and therefore occur at the start or end of a pipe string **(Figure 4 / Slide 17)**.

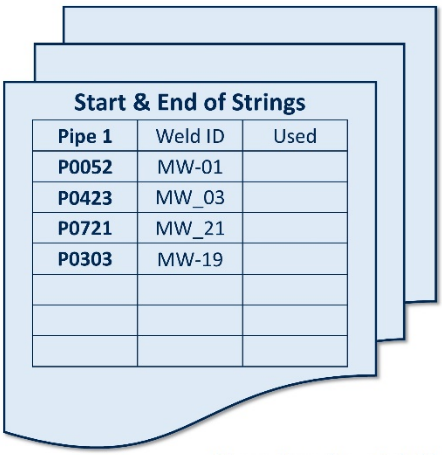


Figure 4 – Identifying Start & End of Strings

6.1.3 *Building String Data*

The pipe data is comprised of a number of individual pipe strings, with a variable distance between pipe strings. These gaps represent tie-in sections or breaks in the mainline welding. Using the dataset with the start and end of strings, the first record is used as the start of the first string **(Figure 5 / Slide 18)**.

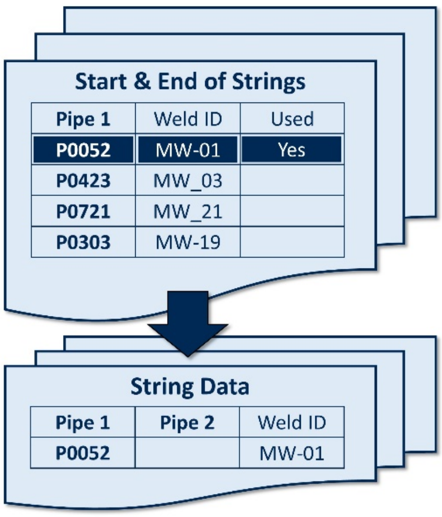


Figure 5 – Build Pipe String (Start of String)

Using the pipe data, the pipe string is built using the pipe/component identifiers. As part of this process, the pipe/component identifier is placed into either the Pipe 1 or Pipe 2 field ensuring that the direction of the pipe string is now honoured. At this point, the data is no longer bi-directional. This process continues until there are no further matches **(Figure 6 / Slide 19)**.

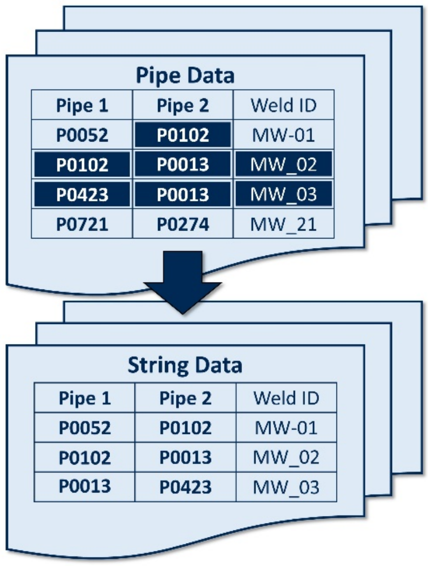


Figure 6 – Build Pipe String

The final pipe/component identifier is confirmed using the dataset containing the start and end of strings and the final pipe/component identifier of the string flagged as used **(Figure 7 / Slide 20)**.

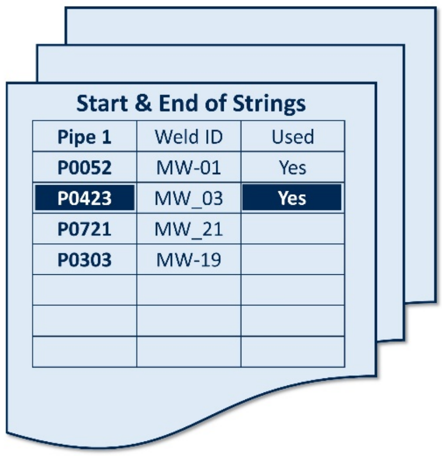


Figure 7 – Build String (End)

6.2 Correcting the Survey Data

The survey data is corrected using a similar process to the weld data. First, the pipe/component identifiers are checked against the vendor data and any errors are identified and corrected. It should be noted, that unlike the weld data, the survey data only contains a single pipe/component identifier and this is not consistently the positive or negative component. The corrected survey data is then sorted into the same order as the pipe string data based on the pipe/component identifiers.

At this point the pipe string data is joined using the pipe/component identifiers. Using the coordinates of the first and last pipe/component identifier of each string, against the design route for the pipeline, it is possible to correct the pipe string direction if required, transforming the scalar data to vector data **(Figure 8 / Slide 20)**.

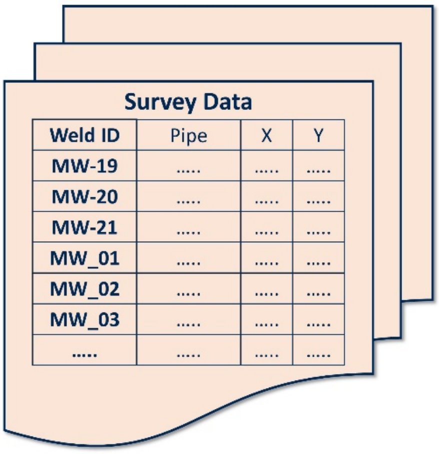


Figure 8 – Survey Data

7. Conclusions

The key concepts of managing entropic data can be summarised as:

- Data entropy cannot be removed but it can be managed.
- Data entropy can be quantified.
- Data processing effort is inversely proportional to data entropy.
- Primary keys should be selected based on the level of data entropy.

The recognition of the notion of entropic data has enabled the development of a solution for those performing pipeline data modelling during the construction phase of the pipeline project. In its broader context, this approach could also offer insights outside of pipeline data modelling where entropic data is encountered.

Arguably the use of dedicated data collection tools such as Collector and Survey123 could eradicate many of the errors seen on this project and should be seen as best practice and the way forward. This approach would require a significant change to working practices, investment in technology (hardware and software) and training. However, until pipeline data modelling becomes a standard requirement for pipeline operators during both the design and construction phases of pipeline projects, there is unlikely to be the appetite by pipeline construction contractors to embrace this technology.

In the absence of a better technical solution, GIS analysts with responsibility for pipeline data modelling will need to seek innovative solutions to improve the process of handling and managing a disparate range of field collected data.

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

Oxford Dictionaries. (2010). *Oxford Dictionary of English*. 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.

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 also 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:

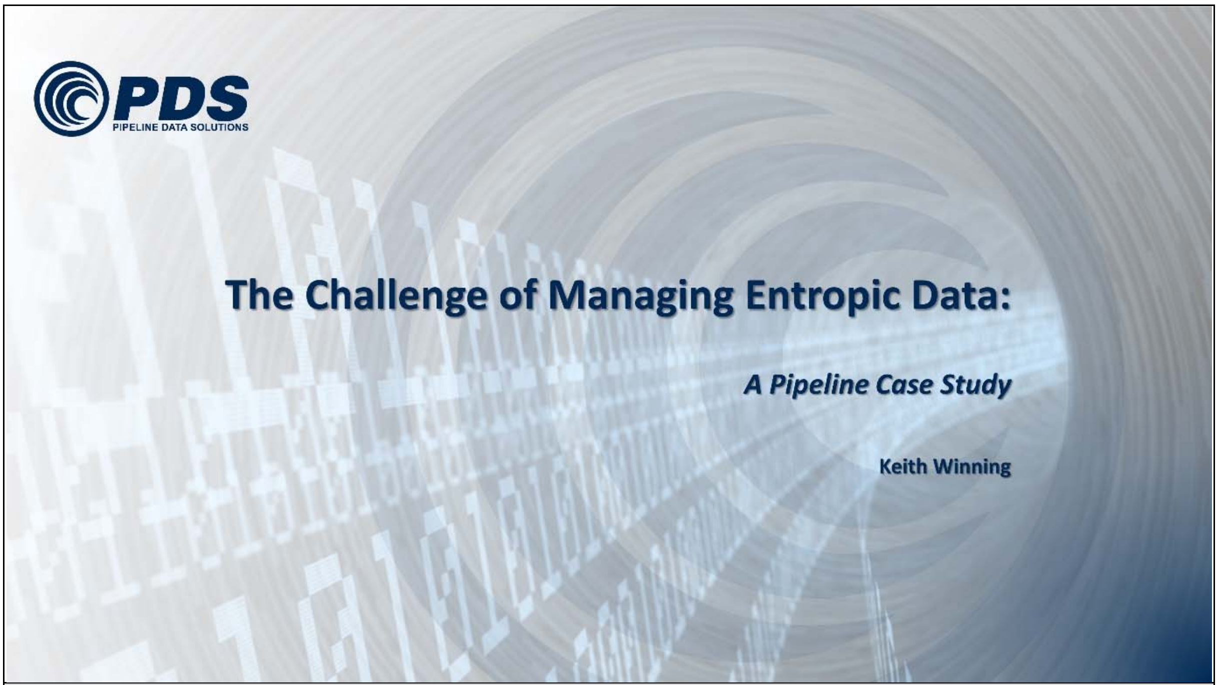
- Capture & Cleansing
- Creation
- Validation
- Management & Maintenance

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

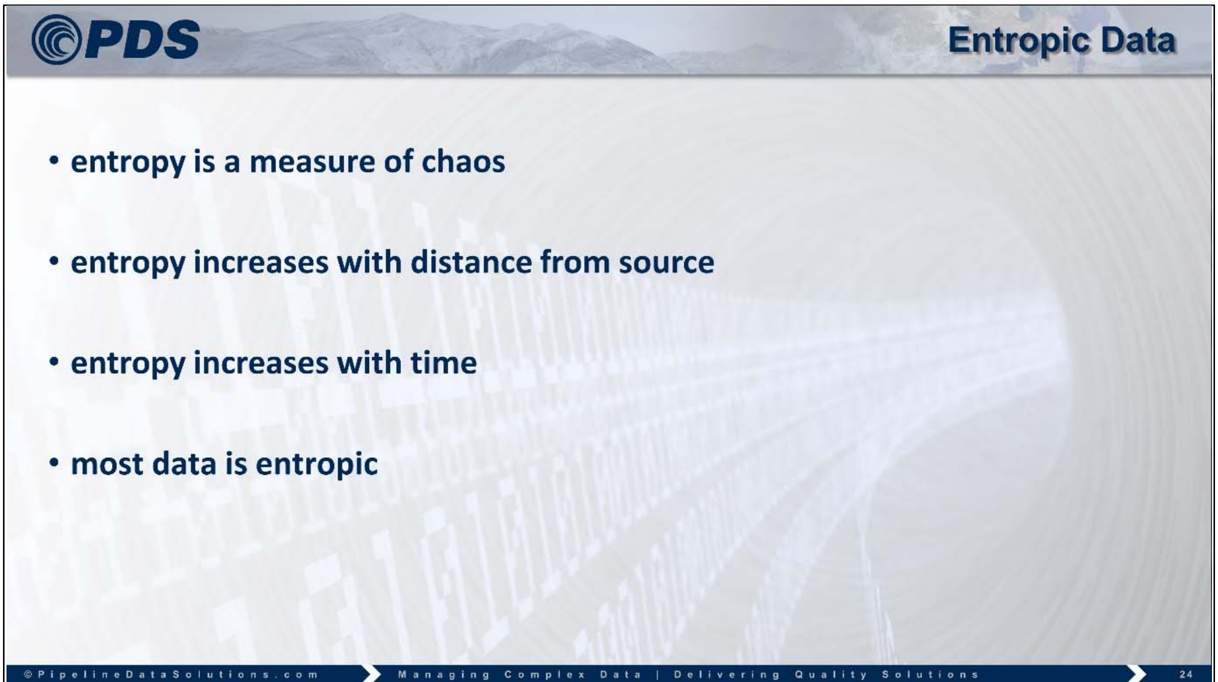
We can support all pipeline database schemas, including PODS, APDM, UPDM and client specific schemas.

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.

Appendix – Presentation Slides



Slide 1

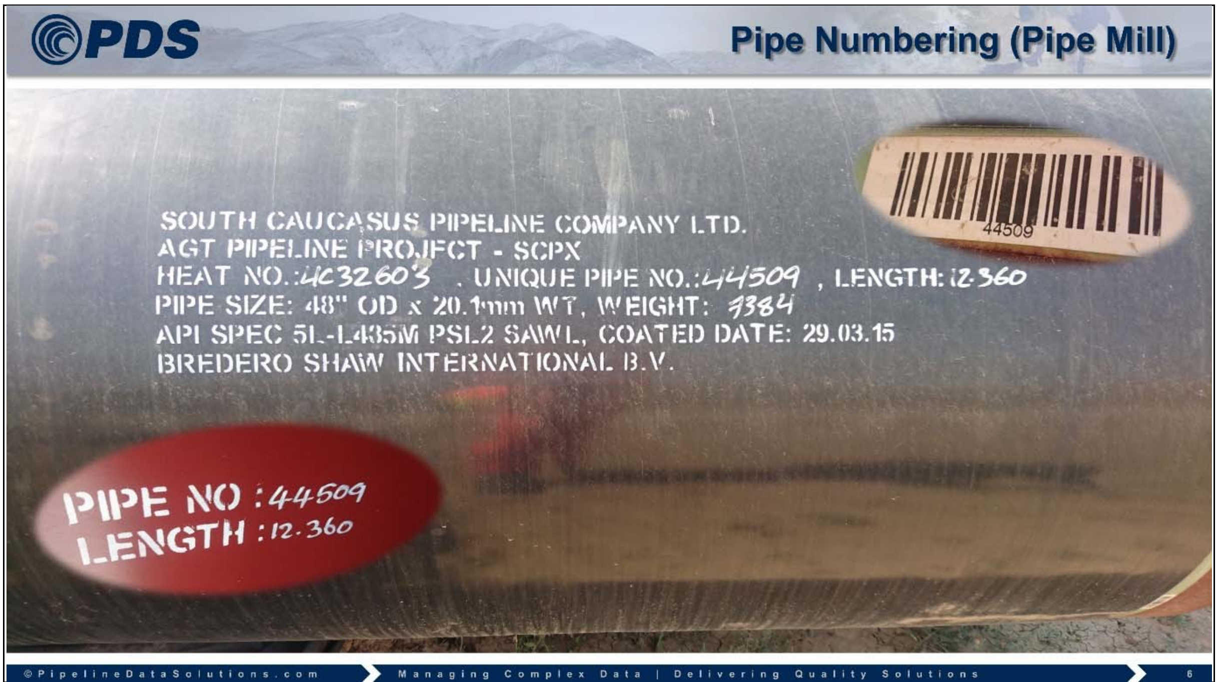


Slide 2

Slide 3

Slide 4

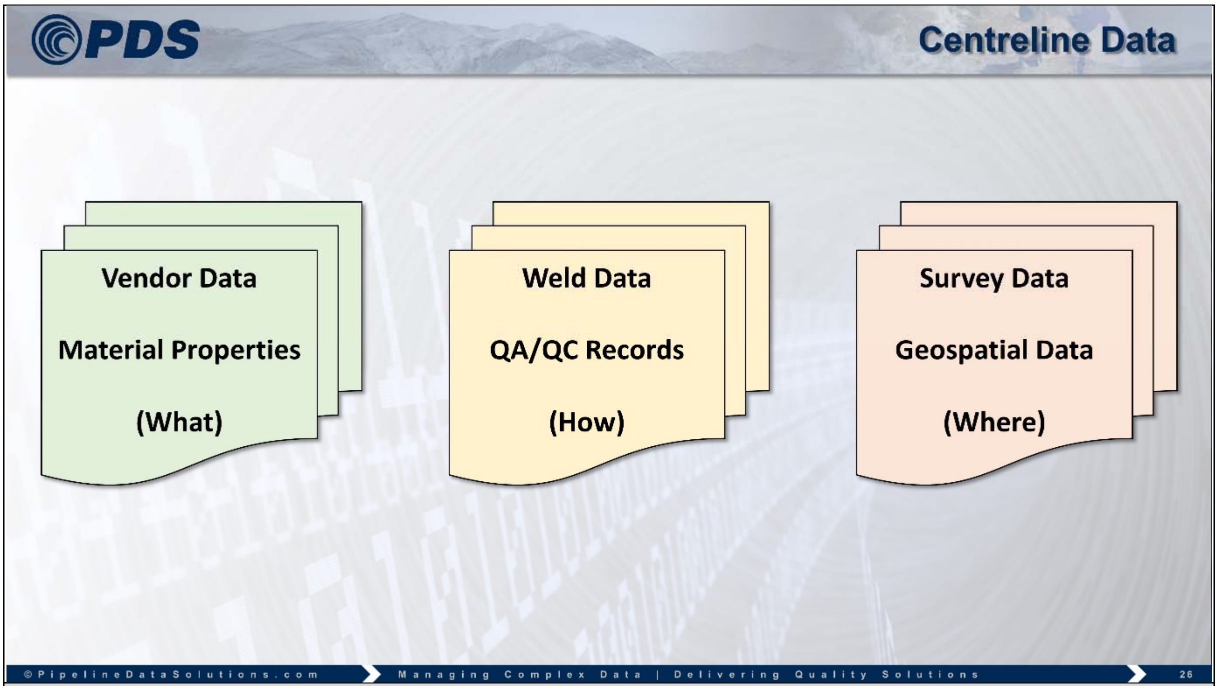
Slide 5



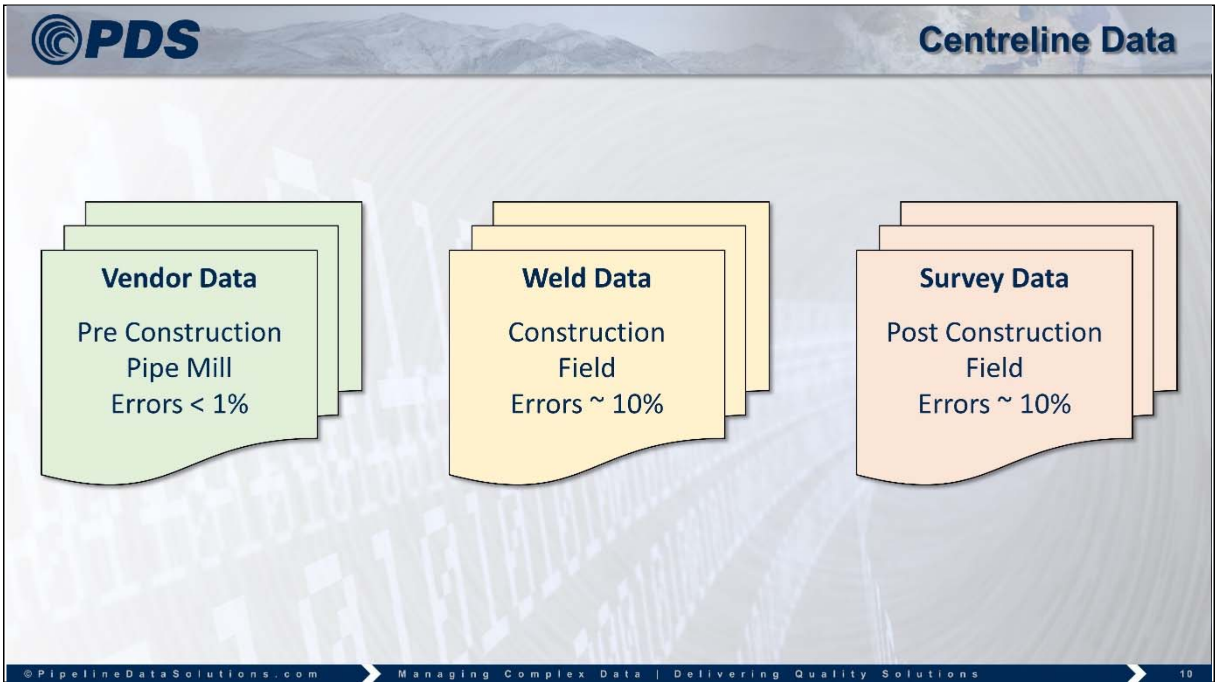
Slide 6

Slide 7

Slide 8



Slide 9



Slide 10

Some of the Challenges

- Transforming bi-directional data with gaps
- Datasets with different primary keys
- Errors in the data including the primary keys

Slide 11

Weld & Survey Data Primary Key Errors

Weld Data

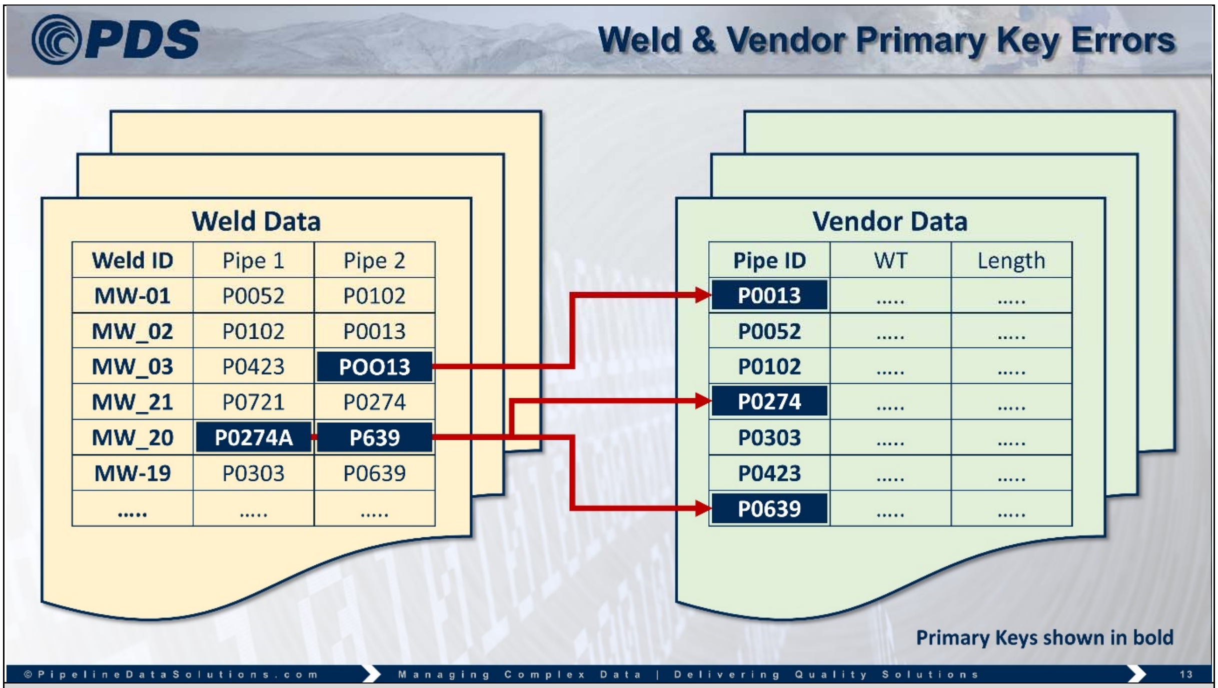
Weld ID	Pipe 1	Pipe 2
MW-01	~~P0052~~	~~P0102~~
MW_02	P0102	P0013
MW_03	P0423	POO13
MW_21	~~P0721~~	~~P0274~~
MW_20	~~P0274A~~	~~P639~~
MW-19	P0303	P0639
.....	.....	.....

Survey Data

Weld ID	Pipe	X	Y
MW-19	.....	.....	.....
MW-20	.....	.....	.....
MW-21	.....	.....	.....
MW_01	.....	.....	.....
MW_02	.....	.....	.....
MW_03	.....	.....	.....
.....	.....	.....	.....

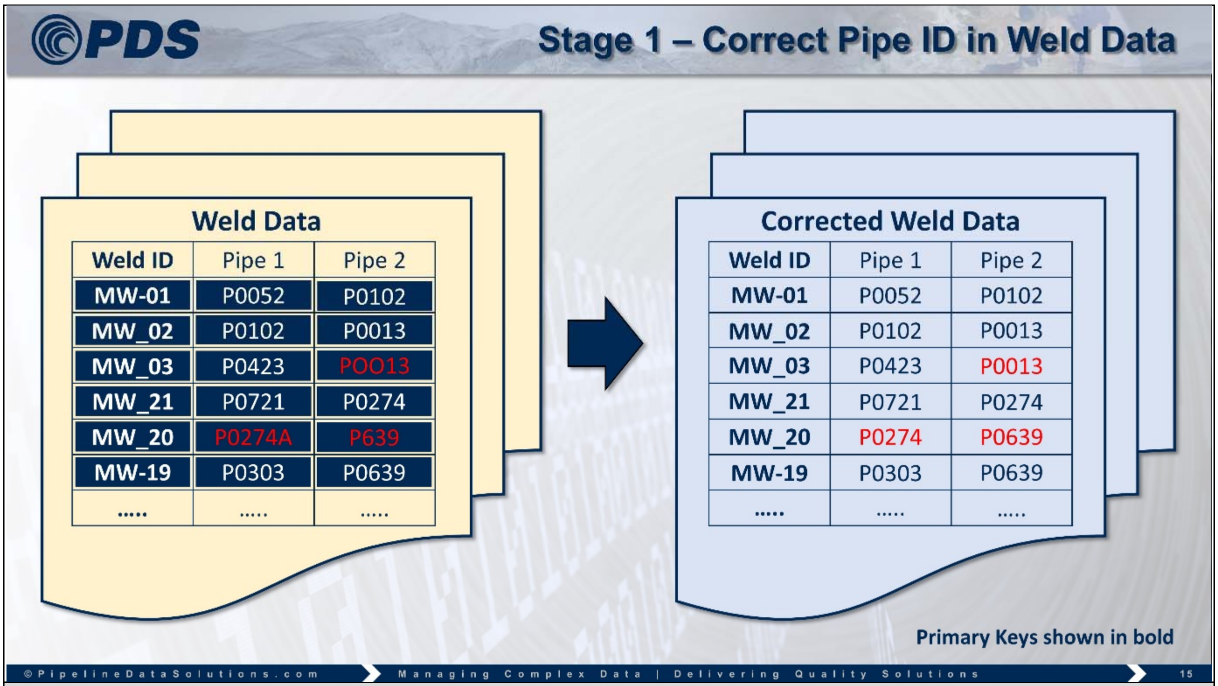
Primary Keys shown in bold

Slide 12

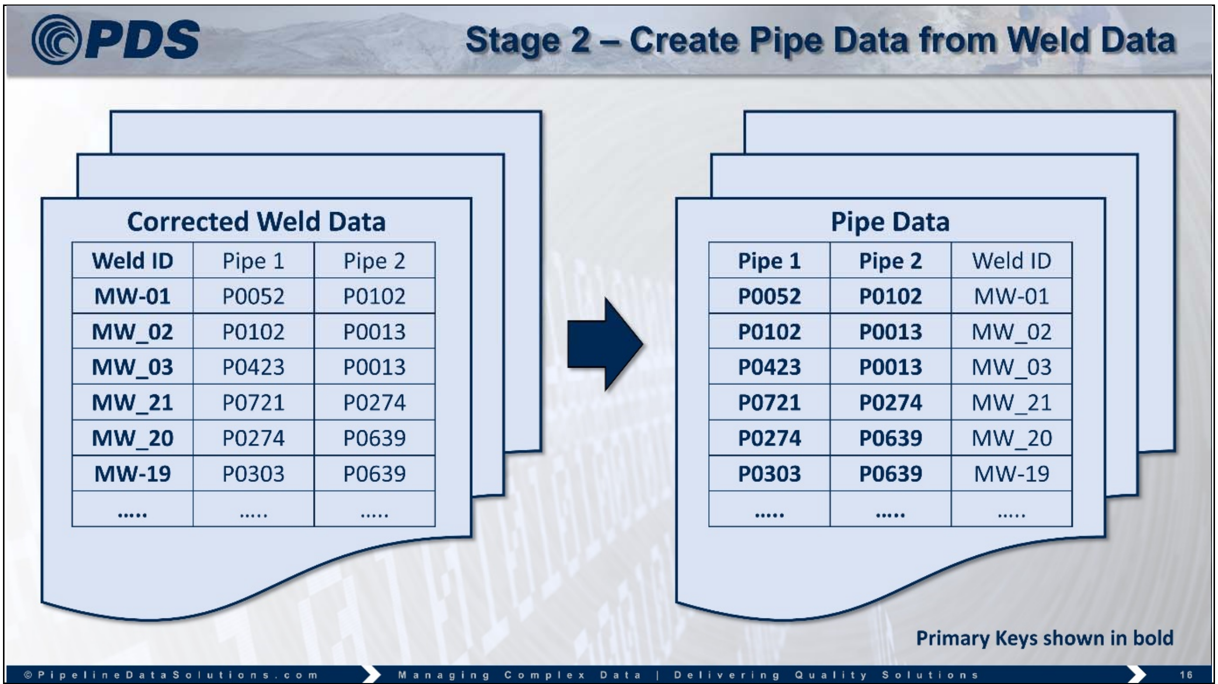


Slide 13

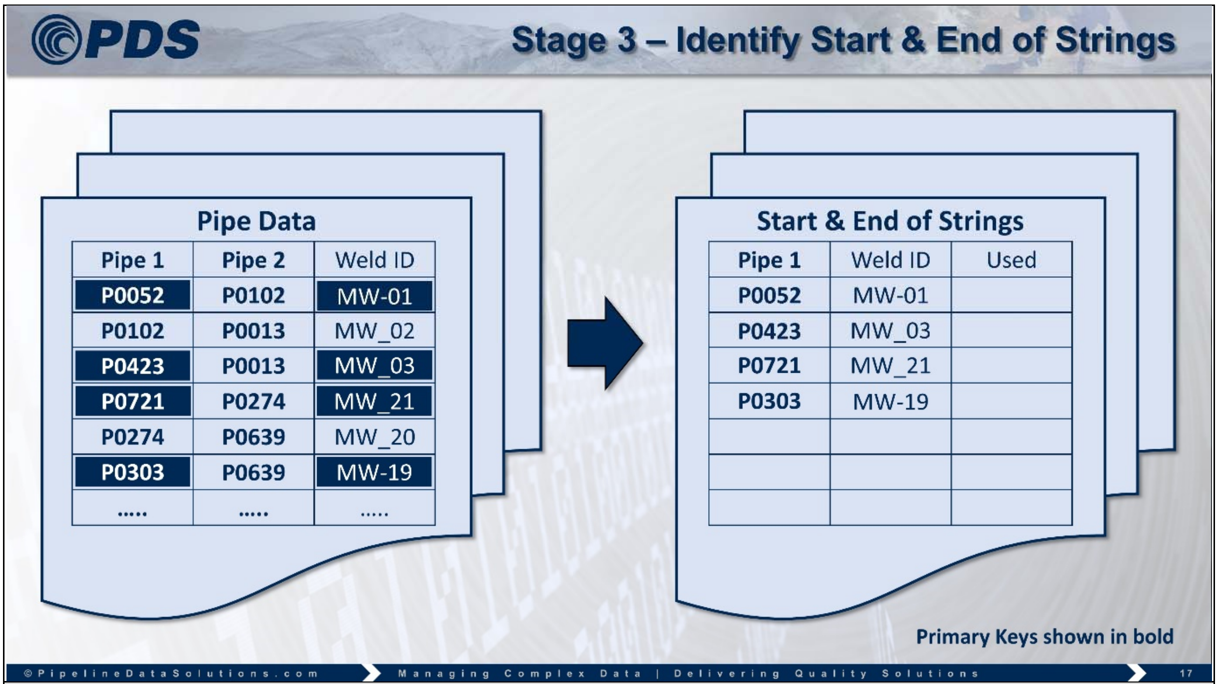
Slide 14



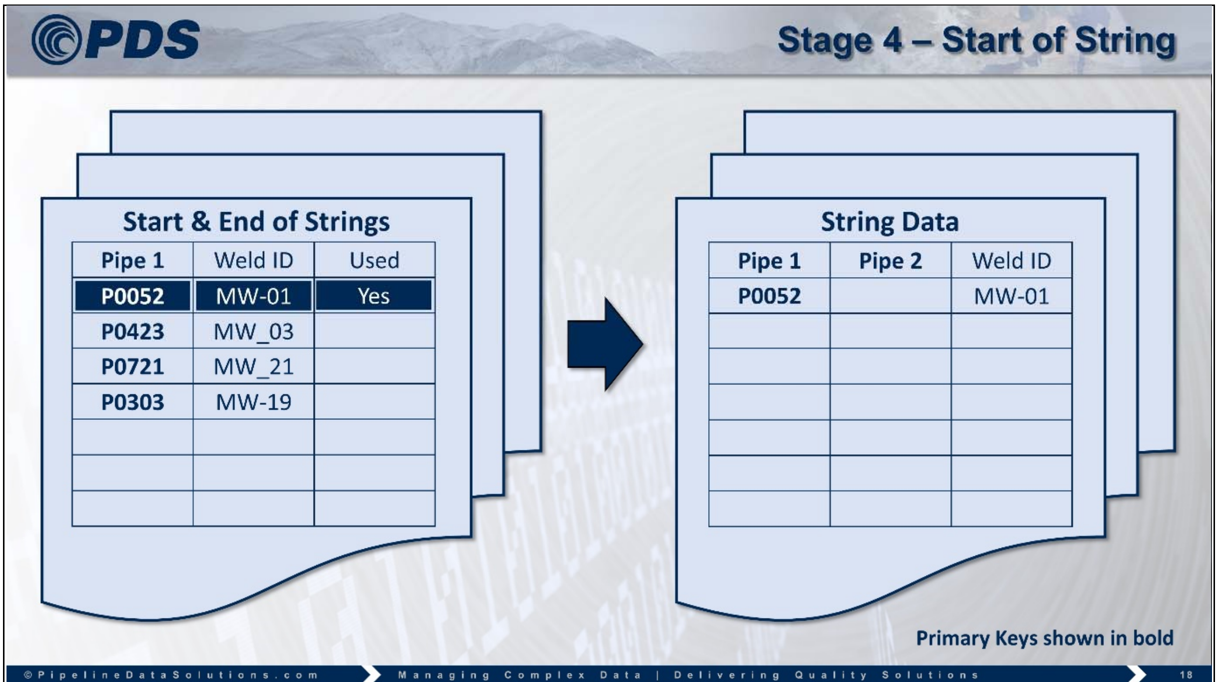
Slide 15



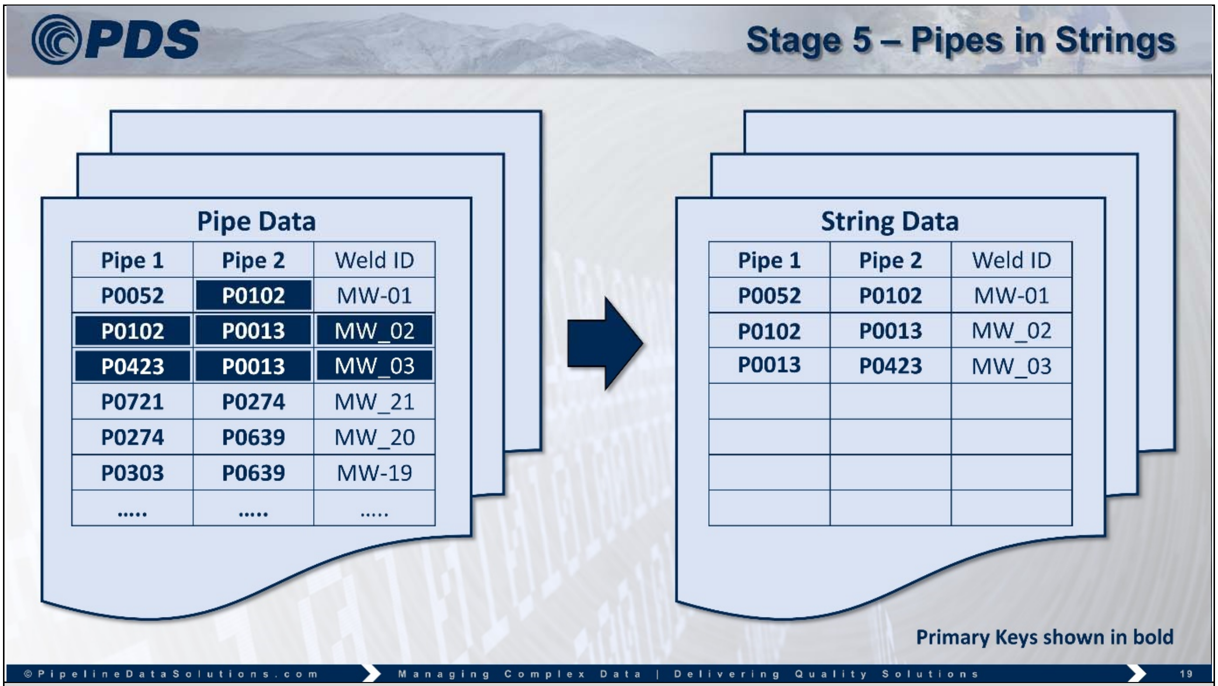
Slide 16



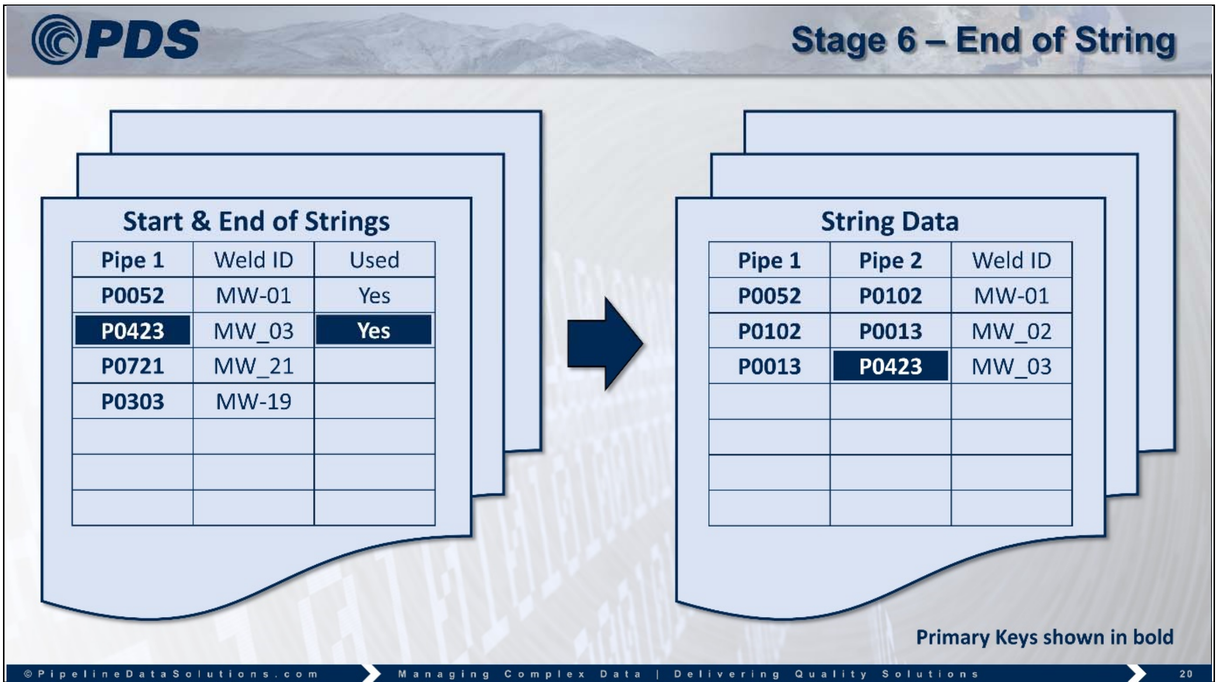
Slide 17



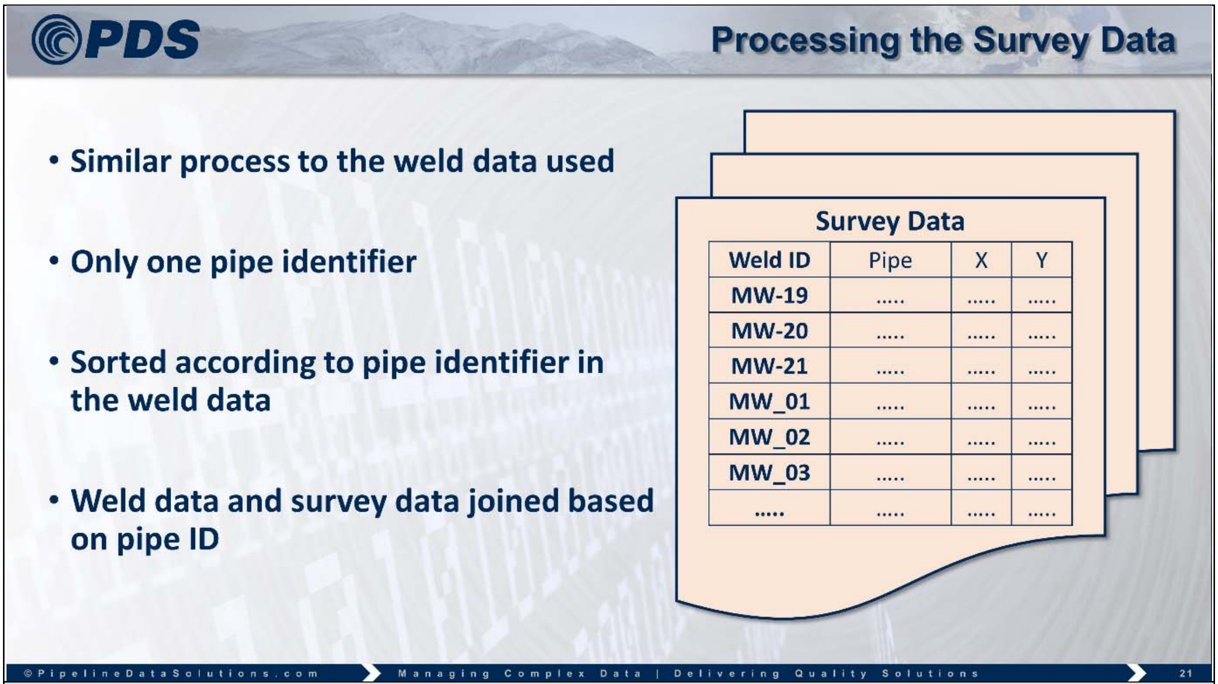
Slide 18



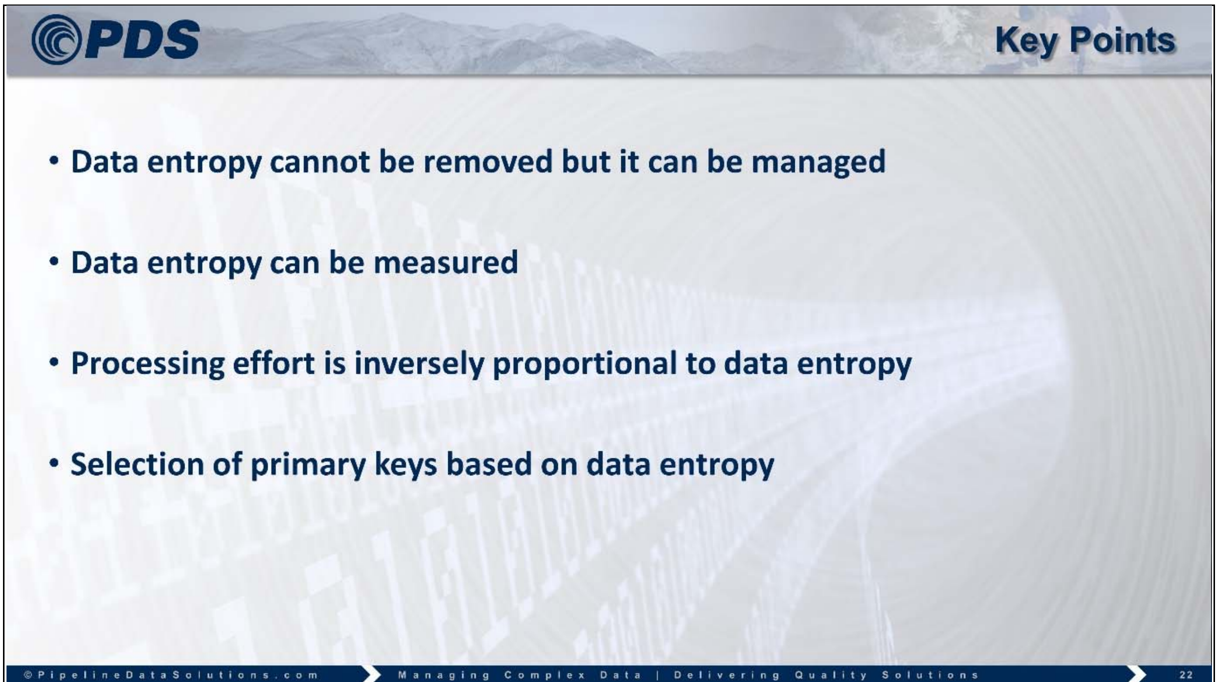
Slide 19



Slide 20



Slide 21



Slide 22


PDS
PIPELINE DATA SOLUTIONS
Managing Complex Data | Delivering Quality Solutions
www.PipelineDataSolutions.com | info@PipelineDataSolutions.com | 27, Old Gloucester Street, London WC1N 3AX

Slide 23