

# Streamlining data validation, transformation and loading:

## Reducing cost and time for data handover and migration of legacy data

By A Park, Intergraph

*Performing a high-quality data handover is critical to ensure the safe, reliable and effective operation of a process plant. If the quality of information handed over to operations is questionable, there is a constant need to physically verify the true physical state of the plant. This article discusses ways to ensure a smooth handover of data from project to operations.*

**M**anaging the handover of data and documentation from CAPEX (Capital Expenditure) projects to operations is a formidable and labour-intensive task. The volumes of data handed over are enormous, with a typical \$1 billion CAPEX plant having 200 000 tags, 100 000 documents and up to 20 million characteristics and relationships. Data are submitted from multiple sources (contractors, suppliers, authorities, etc) and are often delivered incomplete and with errors.

The challenge of validating and correcting errors can be compounded if data are delivered very late in the project life cycle, with staff demobilised, budgets exhausted and management focused on how to finish the project as quickly as possible. The handover from a \$1 billion CAPEX plant can typically cost \$10-15 M and take up to one year to successfully validate and load operations systems.

Performing a high-quality data handover is critical to ensure the safe, reliable and effective operation of a process plant. If the quality of information handed over to operations is questionable, there is a constant need to physically verify the true physical state of the plant. This can drive up the cost of plant modifications by 30%.

### **Smooth handover of data**

A best-practice approach to ensure a smooth handover of data from project to operations includes the following prerequisites:

- A well-defined information handover specification incorporated into all contracts.
- An organisation responsible for follow-up of handover requirements.
- A process of continual incremental handover of data.
- A mechanism to receive and verify data received and loaded into target systems.

Intergraph has extensive experience in assisting their customers with data handover by providing information specifications, assistance in quality control and loading of data. The new tool, SmartPlant Enterprise for Owner Operators – VTL (Validation, Transformation and

CAD - Computer Aided Design  
 CAPEX - Capital Expenditure  
 CSV - Comma Separated Value  
 DLL - Dynamic Link Library  
 EPC - Engineering, Procurement, and Construction  
 OCR - Optical Character Recognition  
 PDF - Portable Document Format  
 PL/SQL - Procedural Language/Structured Query Language  
 VTL - Validation, Transformation and Loading  
 XSLT - Extensible Stylesheet Language Transformations

**A**bbreviations

Loading), combines many years of experience in data quality control and loading with requirements identified by key customers. A major oil and gas customer has collaborated closely with Intergraph on defining both requirements and testing the solution, ensuring that the product is closely aligned with industry needs.

**Handover from CAPEX projects**

VTL is a comprehensive solution that manages data acquisition from multiple sources. Data are held in a staging area, subject to rigorous quality control before being extracted for loading into target systems, including Intergraph’s SmartPlant Enterprise suite and third-party applications, as suggested in *Figure 1*.

The solution enables owner operators and project management contractors to verify the quality of incoming data prior to loading into target project or operations systems. It empowers EPCs (Engineering, Procurement and Construction) to check the quality of information deliverables before sending them to customers. The extensive tracking process of all data submissions, checks performed and results will offer complete auditable traceability of the handover process.

**Data validation during maintenance and operations and legacy data migration**

VTL can play a key role in efforts to improve and maintain data quality during the operations phase of a plant. Data may be extracted from operations and maintenance systems or by scanning and use of OCR (Optical Character Recognition) or parsing of PDF (Portable Document Format) and CAD files and loaded into VTL to validate data quality and perform corrections as part of a data quality improvement exercise before reloading back into operations systems.

Data from turnaround projects and OPEX (Operating Expenditure) modifications can be very extensive. Information can be validated in VTL to ensure correctness and completion prior to approval for loading into operations systems.

VTL can also play a key role in data migration from legacy systems. Data can be loaded from existing applications for verification before loading into new target systems.

**Data import**

Incoming data submissions from one or more sources are imported into the VTL staging area. The customer provides mapping and

transformation tools. Any mapping tool capable of generating XSLT (Extensible Stylesheet Language Transformations), such as MapForce from Altova, can map incoming data or export data to be loaded into target systems.

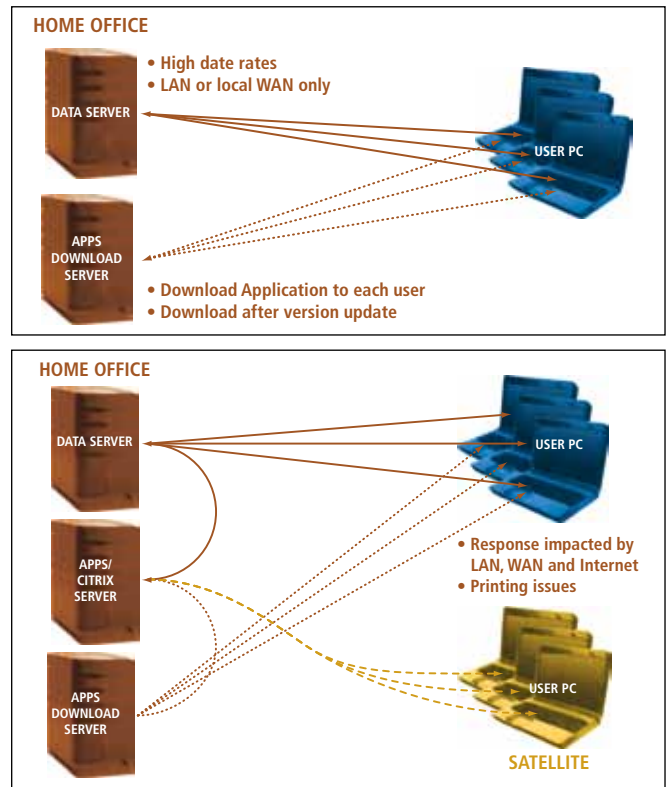


Figure 1: Validation, transformation and loading of data is critical.

The import module reads incoming files and transforms these using the XSLT generated from a mapping tool. The incoming file format may be provided in accordance with the supplier’s standard, the owner’s handover specification or an international standard such as ISO 15926 [1]. The import module’s advanced features prompt users for input or can calculate values based on incoming data if needed. Users can check incoming CSV files for compliance with RFC 4180 [2] standard to validate CSV file structure. Invalid characters can also be identified for exclusion.