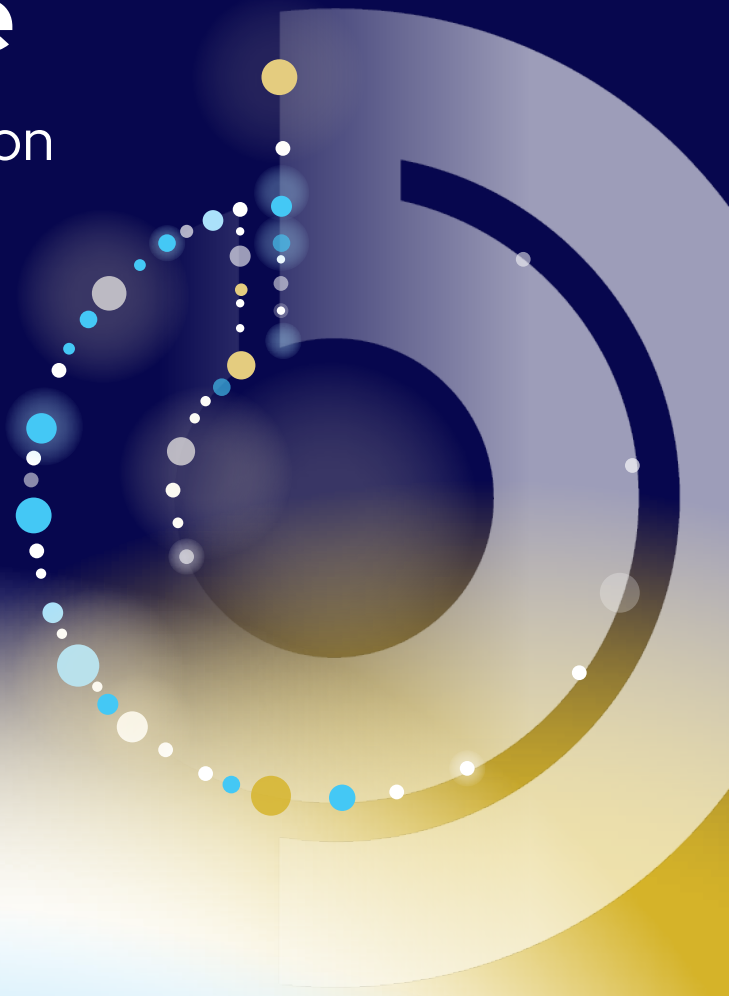
OvationData

Ultimate Guide

Data Cleaning and Ingestion



January 2026

ovationdata.com

Contents

Who this Guide is for

This guide is for data managers, digital transformation leads, and geoscience teams in upstream oil and gas who are responsible for managing, cleaning, or migrating subsurface and operational data.

It assumes you're familiar with industry data types but not necessarily with the details of an Extract, Transform, and Load (ETL) implementation.

Inside, you'll find:

- A practical framework for approaching data cleaning projects
- A detailed seven-step workflow covering the main data types you'll encounter
- Guidance on evaluating data management partners

The workflow and recommendations draw on Ovation Data's research using industry-standard data models and public benchmark datasets. If you want technical depth on these foundations, the appendices provide additional information.

Data cleaning isn't flashy work, but it's what makes everything else possible - the analytics, the insights, the decisions that save time, cut costs, and spark discoveries. This guide is designed to help you get it right.

Why Data Cleaning Matters

Oil and gas has always been a data-intensive industry. High-stakes, capital-intensive projects demand rigorous information management, and the sector was working with large-scale datasets long before “Big Data” became a common phrase.

Today, that data exists in countless formats, from seismic surveys and well logs to production records and sensor readings, on media ranging from legacy tapes to modern cloud storage. It’s often scattered across departments and systems, with some records spanning decades.

At the same time, advances in predictive analytics, data science, and machine learning offer real opportunities to optimize operations, control costs, and make better decisions. But none of these technologies can deliver results if the underlying data is incomplete, inconsistent, or inaccessible. If you intend to use AI or predictive analytics to accelerate decision-making, clean data isn’t just helpful; it’s essential. Many new tools simply won’t work without data that’s properly structured and organized.

That’s where data cleaning comes in.

Operational efficiency

The cleaning and preparation of data, a process known as Extract, Transform, and Load (ETL), can account for up to 50 per cent of a data scientist’s time.

An ETL workflow does three things:

Extracts

Extracts data from various sources (databases, spreadsheets, APIs, legacy formats).

Transforms

Transforms it by cleaning, filtering, and structuring it into a usable format.

Loads

Loads it into a database for analysis and reporting.

Get this process right, and your data becomes a genuine asset. Get it wrong, and your teams spend their time wrestling with files instead of finding insights.

Improved decision-making

Effective data management doesn’t just save time, it improves the quality of decisions across the organization. When data is clean, consistent, and accessible, teams can trust what they’re seeing and act on it with confidence.

The recommendations in this guide draw on Ovation Data’s research into ETL challenges in upstream oil and gas. That research used industry-standard data models as reference points, particularly the Open Subsurface Data Universe (OSDU) and the Professional Petroleum Data Management (PPDM) model.

These frameworks represent current best practice for managing oil and gas data, and understanding them can help shape an effective ETL approach.

For a detailed overview of OSDU and PPDM data models, see Appendix A.

Cost management

Poor data management has direct cost implications - from duplicated effort and delayed projects to compliance failures and missed opportunities. Conversely, investing in proper data cleaning can deliver measurable savings.

Geographic Information System (GIS) tools play an important role in the ETL process, helping teams integrate, analyze, and visualize spatial data. Open-source options like QGIS now offer functionality comparable to commercial alternatives at substantially lower cost.

For long-term projects, like Carbon Capture and Storage (CCS) monitoring, which can span decades, integrating data management with analytics from the outset is essential. Ovation's research into CCS data challenges identified several key requirements, including affordable petabyte-scale storage, automated metadata extraction, and shared access through web-based GIS portals.

For full details on Ovation's CCS research findings, see Appendix B.

How this Guide is Structured

The rest of this guide is organized to take you from high-level principles through to practical implementation:

- **Before you Begin - Auditing and Managing Your Data**
  Provides a simple three-phase framework for thinking about data cleaning projects - to audit, manage, and realize the benefits.
- **Four Data Types**
  Introduces the four main data types you'll encounter in upstream oil and gas operations, and why each presents distinct challenges.
- **A Seven-Step Workflow for Data Cleaning and Ingestion**
  Walks through a detailed seven-step workflow for cleaning and ingesting these data types, based on Ovation's research and practical experience.
- **Choosing the Right Partner**
  Offers guidance on evaluating data management partners, including the questions to ask and the pitfalls to avoid.

If you're looking for technical depth on industry data models and Ovation's research methodology that informs this document, you'll find that in the appendices.

Before you Begin – Auditing and Managing Your Data

Data cleaning techniques vary depending on how your organization stores and manages its data. Ovation's approach follows three phases, which provide a useful framework for whatever your specific circumstances.

Step 1

Audit your data

Start by understanding what you have. This means:

- Performing **a system health check** to identify where data sits across departments (and where silos exist).
- Running a **data quality assessment** to spot inconsistencies, gaps, and errors.
- **Searching for duplicates** and removing repetitive records.
- **Validating data** before any migration.

For large databases, automation is essential - manual auditing simply isn't feasible at scale.

Step 2

Manage your data

Once you know what you have, put systems in place to manage it properly:

- **Standardize formats** so data can be integrated and compared.
- Use **a single platform** for access wherever possible.
- Ensure your system can **handle all data types**, including legacy formats.
- **Archive data** that isn't currently in active use.
- **Manage risk** through secure storage and disaster recovery capabilities.

Step 3

Realize the benefits

Done well, data cleaning delivers measurable results. You gain clarity about what data you have and where to find it. Teams spend less time searching and more time analyzing. Business decisions are based on reliable, consistent information rather than guesswork.

There are organizational benefits too. Breaking down silos means knowledge flows more freely across departments. Costs fall as downtime, bottlenecks, and duplication of effort are reduced.

Clean, well-managed data is the foundation for what comes next - advanced analytics, machine learning, and genuine digital transformation. These benefits compound over time. Organizations that invest in data cleaning often find that it unlocks value they didn't know was there.

Four Data Types

Upstream oil and gas operations (exploration, drilling, well completion, and production) generate a wide range of data. This guide focuses on the four that matter most for data cleaning and ingestion - GIS data, seismic data, well data, and production data.

GIS data

Geographic Information System (GIS) data helps operators understand and manage their assets, operations, and risks. It includes two components: geospatial data and attribute metadata.

Geospatial data describes the location and geometry of features on the earth's surface or subsurface, including licensed block boundaries, wellhead and well bottom locations, pipelines, and facilities. Attribute metadata captures the characteristics of these features, things like ownership, depth, timestamps, and other relevant properties.

Seismic data

Seismic data from geophysical surveys is crucial to upstream operations. It provides insight into subsurface structures, enabling teams to locate potential hydrocarbon reservoirs. The standard industry format is SEG-Y, with the current revision being SEG-Y 2.1 (with a clarification on Extended Textual Headers released in May 2025).

Industry data models such as the Open Subsurface Data Universe (OSDU) provide comprehensive support for seismic data, including acquisition surveys, processed data, and 2D or 3D interpreted data. OSDU's master data types (such as Seismic Acquisition Survey and Seismic Processing Project) reference specific elements, including play type (geological setting), hydrocarbon prospect, field name, and basin name, enabling seismic datasets to be placed in their relevant business context.

For more details on the OSDU data models, see Appendix A.

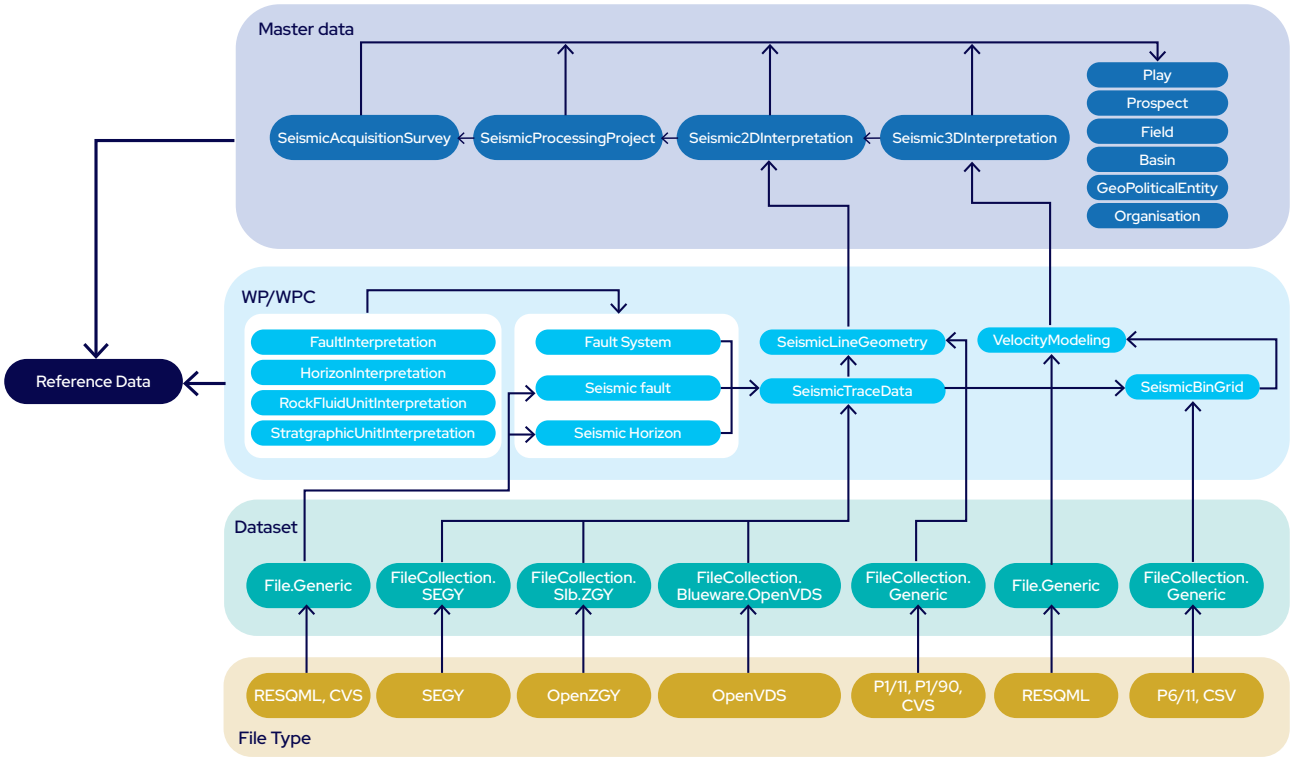


Figure 1: OSDU Seismic Model (not all WP/WPC and File Types are shown)

Well data

Well data captures the physical and chemical properties of rocks and fluids within a well. This data is typically collected via sensors deployed in a wellbore or attached to a drill string, measuring properties such as resistivity, porosity, and pressure. It forms the foundation for assessing reservoir quality and optimizing production.

The OSDU platform provides comprehensive support for well data, including well logs, well trajectories, and wellbore data. Its master data types ('well' and 'wellbore') reference essential elements such as play type, hydrocarbon prospect, field name, and basin name, allowing well data to be mapped within the appropriate exploration or production context.

For more details on the OSDU data models, see Appendix A.

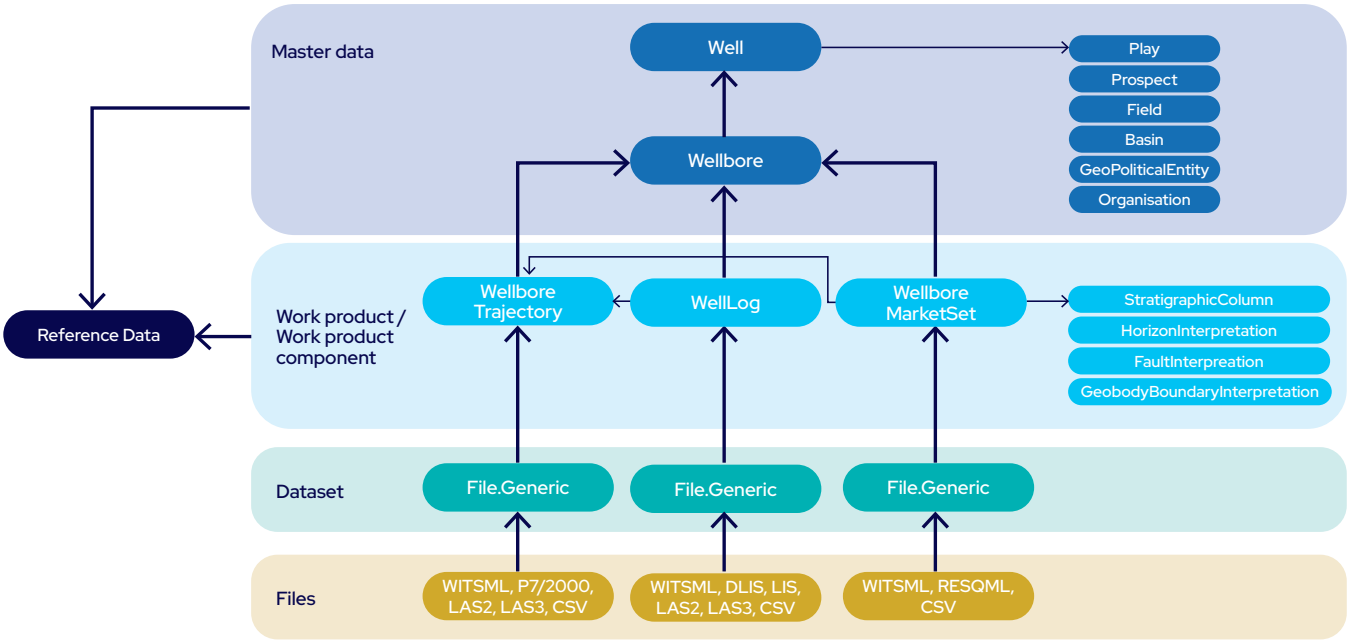


Figure 2: OSDU Well Model (not all WP/WPC and File Types are shown)

Production data

Production data is collected from sensors on surface or downhole equipment to monitor the performance of wells and facilities. It includes measurements such as oil and gas rates, pressure drop, and pump speed, and is used to optimize recovery and ensure operational safety.

Other data types (geological, geophysical, geochemical, petrophysical, reservoir simulation, and environmental data) are also important to upstream operations but fall outside the scope of this guide. Together, these data types enable a comprehensive understanding and management of reservoirs.

A Seven-Step Workflow for Data Cleaning and Ingestion

The previous sections explained what data cleaning involves and the data types you'll encounter. This section outlines a seven-step workflow, developed through Ovation's benchmark research, which shows what a robust ETL process looks like in practice.

Most organizations will need specialist support to implement a workflow like this, but understanding the steps will help you evaluate partners and ask the right questions.

For details on the public benchmark datasets and research methodology underpinning this workflow, see Appendix B.

1 Identify culture data and work with the Coordinate Reference System (CRS)

Culture data is the foundation of any project. It includes essential information on licensed block boundaries, roads, pipelines, power lines, and facility locations. This data (typically in formats such as ESRI Shapefile, GeoPackage, or GeoJSON) establishes the Area of Interest (AOI) for subsequent data cleaning. It's usually sourced from government agencies or professional survey-certified companies.

Throughout the ETL process, it's crucial to maintain the integrity of the attribute information attached to culture data.

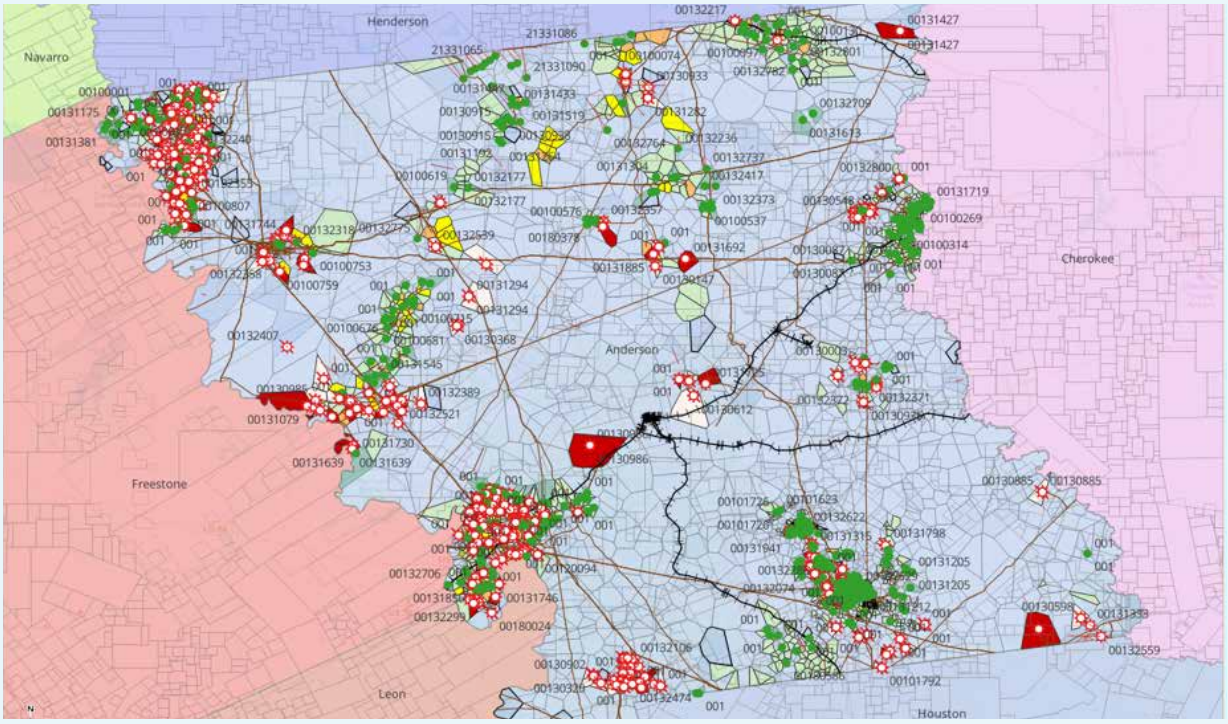
Location and projection

Location information is at the heart of culture data. Survey data is usually presented as Cartesian coordinates (X, Y), with the Coordinate Reference System (CRS) determining how those coordinates are projected onto maps. Most web-based GIS applications use the Universal Transverse Mercator (UTM) system with WGS 84 as the datum. During ETL, CRS conversion is often required to project culture data to the correct location.

For example, the Teapot Dome benchmark dataset (provided in ESRI ShapeFile format) uses CRS EPSG:32056 - NAD 27/Wyoming East Central. ETL tools must be able to load this data correctly and preserve both spatial and attribute integrity.

Quality control

Ovation uses QGIS, an open-source GIS application, to complement the ETL process. It allows for visualization, comparative analysis, and quality control of ETL outputs, helping ensure that culture data and CRS are accurately represented throughout.



Source: Wells and culture data (Texas public data)

2 Deploy a seismic metadata crawler

Seismic data (2D and 3D) plays a central role in geophysical surveying and often covers extensive regions. The industry-standard format for seismic trace data is SEG-Y.

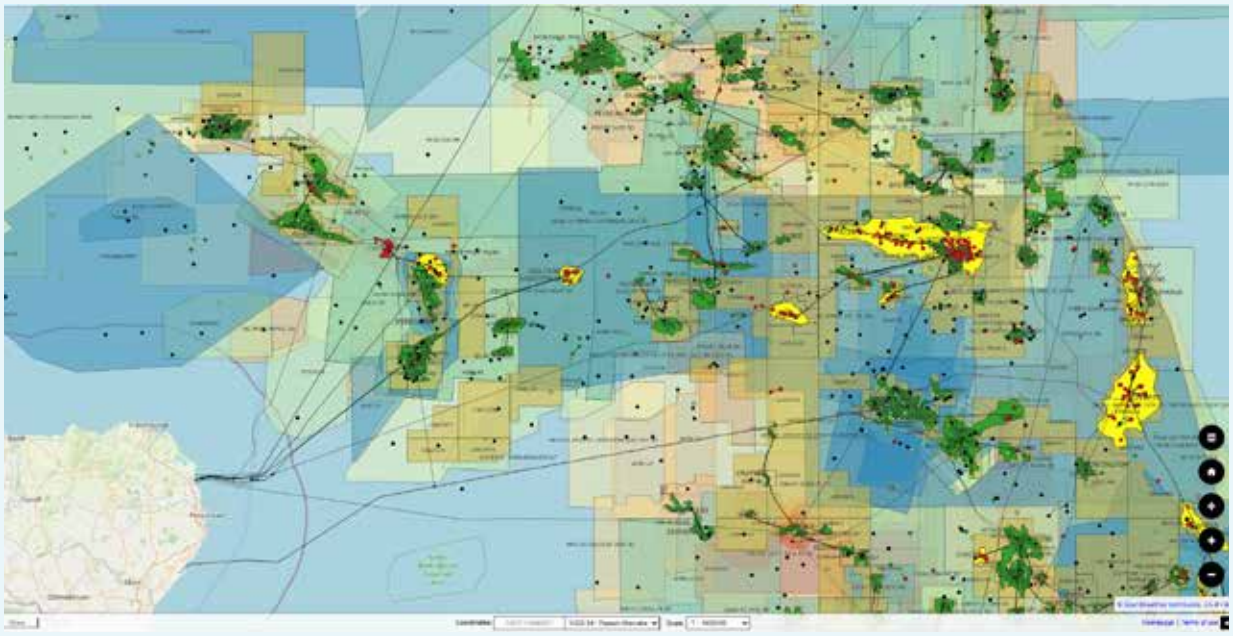
From a data management perspective, ETL tools need to scan hundreds or even thousands of seismic files in a single batch run. This makes a seismic metadata crawler a core requirement for any ETL solution.

What a SEG-Y crawler should do

A capable SEG-Y crawler scans the header information of SEG-Y files and extracts metadata, particularly the location information of seismic traces (the geometry of the seismic survey). This geometry can then be overlaid on a GIS map containing culture data, providing an essential quality control check. The successful extraction ratio is a key performance indicator for any SEG-Y crawler.

Other metadata, such as EBCDIC header or binary header information, must also be extracted and made searchable in the database.

When navigation data is absent from the header or stored in a separate file, georeferencing or digitizing is required to position seismic trace data accurately.



Source: Seismic surveys and culture data (UKNDR public data)

3 Confirm the Unique Well Identifier and locate wellheads and well bottoms

The Unique Well Identifier (UWI) is fundamental to well data management. In the United States, the API well number serves this role (a unique, permanent, numeric identifier). In other countries, a unique well name within a project area may be used instead, though this carries a risk of duplication across projects. Ovation strongly recommends that any data cleaning project should begin by confirming the UWI.

Well geometry

The geometry of a well (including wellhead location, Kelly Bushing (KB), well trajectory from deviation surveys, and the associated CRS) is critical to accurate data management.

Differences in CRS and survey precision are common, particularly when wells and surveys span multiple decades. This can make well borehole geometry complex. At a minimum, location attributes for a well borehole should include UWI, wellhead and well bottom locations, and CRS. Additional attributes typically include KB and Total Depth (TD), measured either by Measured Depth (MD) or True Vertical Depth (TVD).

Other attributes (such as well symbol and spud time) also hold operational significance. A wellbore is a valuable asset, and this information is typically carefully maintained in a database. ETL tools should be able to extract these attributes accurately and map them to their respective databases.

A note on data formats

In the Teapot Dome benchmark dataset, wellhead information is stored in an Excel spreadsheet. A modified SQL script was used to load the data into SQL Server. This illustrates how input formats can vary even when data is described as following a standard.

4 Deploy a well log crawler

Well log data is a rich source of subsurface information, crucial for understanding what lies beneath. It comes in various formats - image files requiring digitization, unstructured formats like CSV tables that need manual header input, or binary formats such as DLIS (pioneered by Schlumberger). For this benchmark workflow, Ovation has focused on Log ASCII Standard (LAS) format files containing well header information.

What a LAS crawler should do

A LAS crawler should be able to process thousands of LAS files in a batch run, extracting and organizing header information into a SQL database. Key requirements include:

- **Fuzzy UWI identification.** The crawler should automatically identify and assign LAS files to one or more UWIs, drawing from file paths, file names, or LAS headers.

- **Hierarchical virtual categories.** It should enable the creation of a Curve Alias table for each subcategory, helping manage log curves that measure the same physical property but have different names due to variations in service vendors or tool types.
- **Advanced query capabilities.** Users should be able to query wells or log curves using criteria such as map area, well attributes, virtual categories, or depth range.
- **Golden Curve Sets.** With appropriate approvals, the crawler should be able to create an optimum set of curves, known as Golden Curve Sets.

In the Teapot Dome benchmark dataset, 1,210 LAS files were processed.

5 Batch load well tops, zones, and facies

Geological information (well tops, geological zones, and lithofacies) plays an integral role in well data analysis. These data points form the foundation for regional geological studies and feed into reservoir property evaluation through petrophysical interpretation.

Managing complexity and interpretation

One of the challenges in managing this data is its inherent hierarchical structure, combined with the subjective nature of geological interpretation. Multiple interpretations of the same data often coexist. For example, the formation top of the Jurassic period may be subdivided into upper, middle, and lower Jurassic tops, each interpreted differently by different geologists. This hierarchy and variation extend to geological zone definitions and lithofacies classifications.

Batch loading is a fundamental ETL requirement, but it also demands a flexible hierarchical categorization system to organize information efficiently. The concept of Golden Sets (approved, authoritative versions of interpreted data) is crucial for managing multiple interpretations and presenting a consistent dataset to end users.

In the Teapot Dome benchmark dataset, formation tops information was prepared in an Excel spreadsheet, adding complexity to the loading process.

6 Incorporate well production and engineering data

Production from a well is the primary business concern for operations teams. Monitoring and managing production data (real-time pressure, temperature, and oil, gas, and water production rates) is essential. Comprehensive engineering data is also relevant, including well completion information, perforation records, and treatment reports.

ETL requirements for production data

Batch loading production data is a fundamental ETL requirement. The ability to automate real-time

data refresh significantly enhances both user experience and operational efficiency.

Given the dynamic nature of production data, access through a web portal with adjustable report templates is highly valuable. Users should be able to create tailored outputs, for example, a cumulative oil production curve for a specific geological zone, or a bubble map showing production breakdown on a GIS map.

7 Incorporate documents and unstructured data

Documents and unstructured data play an important role in upstream oil and gas operations. The workflow developed in Steps 1 through 6 provides the foundation for integrating this data into the system.

Linking documents to structured data

The key principle is to associate documents and unstructured data with existing objects in the database. For example, drilling reports or core photographs can be linked to a well's Unique Well Identifier (UWI) through keyword tagging. This association enables valuable analysis, such as visualizing data intensity on a map, identifying which wells have core images, and spotting gaps in documentation.

The specifics of how documents and unstructured data are handled will depend on each organization's needs and systems. The workflow can be adapted and extended as requirements evolve.

Choosing the Right Partner

A seven-step workflow is one thing on paper. Implementing it with the right tools, expertise, and quality control is another matter. Most organizations will need specialist support, whether that means augmenting an in-house team or working with an external provider.

Not all partners are equal. These five questions will help you evaluate potential providers and avoid common pitfalls.

Question 1: Will you be able to use the data once it's digitized?

Some providers digitize data and identify metadata, but store it in proprietary formats that are difficult to access without purchasing additional services. Before committing to a partner, ask how the data will be stored and whether there are any restrictions on retrieval.

At Ovation, we don't use proprietary wrappers to store data (unless a client specifically requests encryption), and we never charge egress fees. If you're currently locked into a restrictive arrangement with another provider, we can help extract your data so you can actually use it.

Question 2: Do you actually need the "bells and whistles"?

Some providers emphasize features that sound impressive but deliver limited practical value. For example, a web portal with data viewers is only useful if the underlying data is high-quality and properly structured. Before being swayed by feature lists, ask whether the functionality addresses your actual needs, and ask to see it working with real data.

Question 3: Will a "one size fits all" approach meet your needs?

Different organizations have different requirements, and a generic solution may not address yours. Even if you're focused on a single service right now (for example, transcription or migration), consider the bigger picture. Will this partner be able to support you as your needs evolve?

Ovation offers customized solutions across the full data lifecycle, from transcription and storage through to ongoing management and access.

Question 4: Does the provider have experience across industries and geographies?

Data management challenges vary by sector and region. A provider with broad experience is more likely to have encountered, and solved problems similar to yours. If your operations span multiple countries, ask whether the provider can support cross-border projects and whether they have experience with local regulatory requirements. For example, Ovation has delivered data service projects in more than 50 countries, and our teams regularly mobilize to client locations to perform work on site.

Question 5: Can the provider handle the full range of media formats?

Some providers can generate a copy of your data, but in a format that's no longer readable by modern systems, which defeats the purpose. Ask whether the provider can transcribe from your specific legacy formats to current industry standards, and what quality control processes they use.

Ovation's processing facilities include an inventory of more than 1,500 devices, supporting over 60 years of technology and more than 275 different media types. These include large-volume seismic formats. We transcribe from historical tape media to current standards, with real-time quality control throughout.

For a complete list of supported media types, see Appendix C.

Conclusion

Data cleaning isn't the most glamorous part of digital transformation, but it's what makes everything else possible.

The oil and gas industry sits on decades of valuable data, but much of it is locked in legacy formats, scattered across systems, or buried in archives. Without proper cleaning and ingestion, the data can't deliver the insights that organizations need to operate efficiently, control costs, and make better decisions.

The ETL process is complex. It demands specialist tools, deep domain knowledge, and rigorous quality control. But the investment pays off. Clean, well-structured data becomes a genuine asset, ready to support analytics and decision-making that drive competitive advantage.

This guide draws on Ovation's research, benchmark studies, and more than 45 years of experience working with data across the energy sector. We hope it provides a helpful framework, whether you're building internal capability, evaluating external partners, or simply trying to understand what good data management looks like.

If you'd like to discuss how Ovation can support your data cleaning and management needs, we'd welcome the conversation.

About Ovation Data

Much of the data held by oil and gas companies is irreplaceable. Seismic surveys, well logs, and production records represent decades of investment, and once they are lost or corrupted, they can often be unrecoverable. Ovation Data exists to preserve and protect that data, keeping it secure, accessible, and ready to deliver the insights that move your organization forward.

For more than 45 years, we've helped energy companies store, manage, and unlock the value in their data. We meet clients wherever they are on their data journey, working alongside in-house teams or leading the way, adapting our expertise to their specific needs.

Our services span the full data lifecycle, from transcription and recovery through to ongoing storage, stewardship, and access. Whatever the format (tapes, paper, film, or digital files) our specialists recover and prepare data so it's usable in your environment.

Data transcription and recovery

We transcribe more than 500,000 tapes and cartridges every year, converting data from legacy media to current industry standards. Our processing facilities house more than 1,500 devices supporting over 60 years of technology and more than 275 media types, including large-volume seismic formats.

We specialize in seismic and well tape transcription and offer AI-driven well log digitizing through our strategic partners. Where data has been compromised or is stored on failing media, our recovery specialists use custom equipment and techniques developed through decades of experience.

For a complete list of supported media types, see Appendix C.

Seismic data services

Our geophysical data services help operators get more from their seismic archives. We offer navigation merge (matching navigation data with seismic traces), data conversion across all standard formats, demultiplexing, subsetting, and duplication. Our quality control processes are rigorous, and our teams include experienced seismic geophysicists who understand both historical field practices and current standards.

Well data services

We convert well data across all formats including LIS, LAS, DLIS, BIT, and LISTIF. Our scanning and vectorizing services transform paper logs, maps, and seismic sections into usable digital formats, adding genuine analytical value to legacy archives.

Storage and access

We provide secure, scalable storage solutions, from traditional infrastructure to cloud-based systems, with disaster recovery capabilities built in. Your data stays always secure, always accessible, and always ready to run, with no hidden fees, no complex file types, and no lock-in.

Global reach

With data service projects delivered in more than 50 countries, our teams are experienced in cross-border work and regularly mobilize to client locations. Whatever the format, wherever the location, we can help.

References and further information

1. The National Data Repository (NDR) encompasses 2D/3D seismic data, offshore wells, and research documents.
2. The Bureau of Ocean Energy Management (BOEM) offers valuable insights with its 2D/3D seismic data and offshore wells. BOEM hosts major U.S. Carbon Capture Storage (CCS) projects.
3. The Railroad Commission of Texas (RRC) has a wealth of data on horizontal wells, production data, and well documents.

Appendices

Appendix A: Industry Data Models (OSDU and PPDM)

This appendix provides additional detail on two industry-standard data models referenced in this guide - the Open Subsurface Data Universe (OSDU) and the Professional Petroleum Data Management (PPDM) model. Together, OSDU and PPDM represent the current best practice for managing oil and gas data. Understanding their structure can help inform an effective ETL approach.

Open Subsurface Data Universe (OSDU)

OSDU is an open-source data platform developed to address one of the energy industry's persistent challenges - siloed data. By standardizing how data is managed and supporting diverse data types, OSDU enables organizations to integrate workflows, accelerate deployment, and improve decision-making.

The platform supports most data types used in the energy industry. Its master data types provide a consistent way to reference key business elements (including play type, hydrocarbon prospect, field name, and basin name) across different datasets. This allows seismic, well, and other data to be placed in their appropriate exploration or production context.

OSDU and seismic data

OSDU provides comprehensive support for seismic data models, including data collected through acquisition surveys, processed seismic data, and 2D or 3D interpreted data. Key master data types include SeismicAcquisitionSurvey, SeismicProcessingProject, Seismic2DInterpretation, and Seismic3DInterpretation.

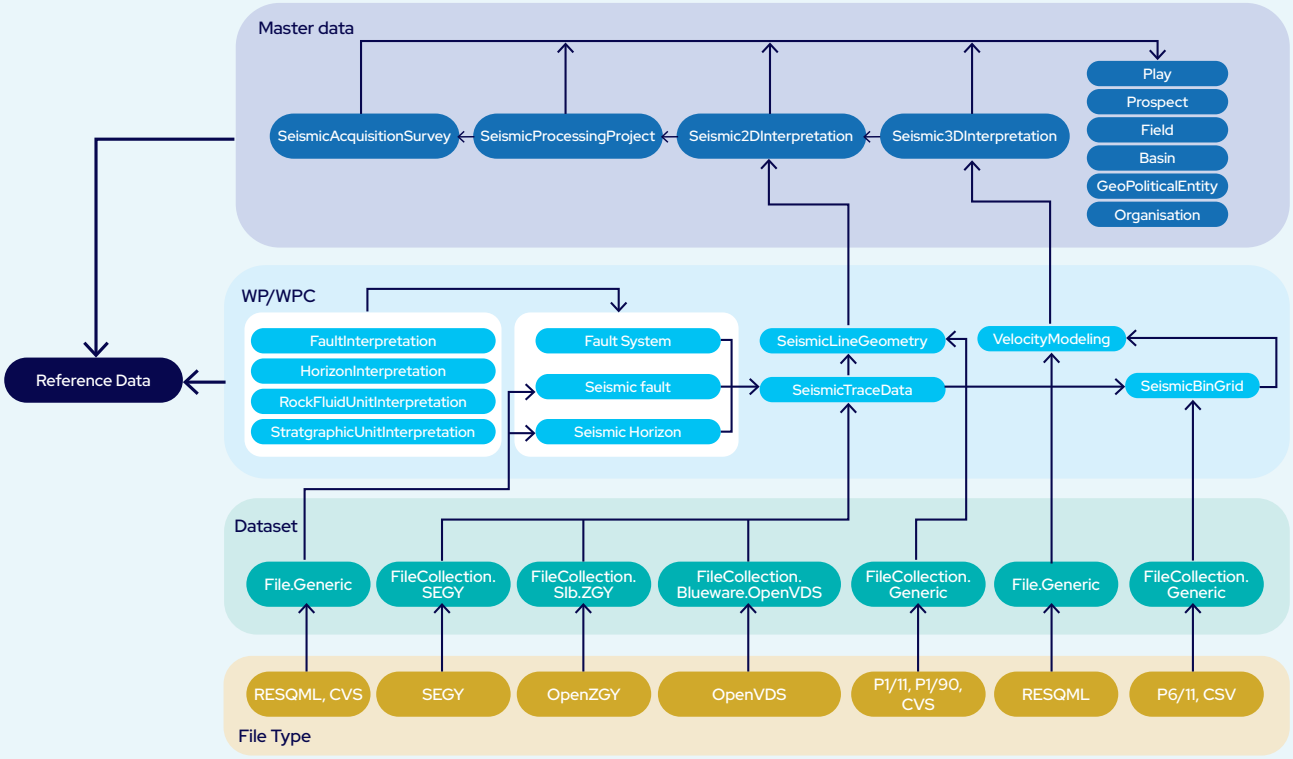


Figure A1: OSDU Seismic Model (not all WP/WPC and File Types are shown)

OSDU and well data

OSDU also provides comprehensive support for well data, including well logs, well trajectories, and wellbore data. Its master data types ('well' and 'wellbore') reference essential elements such as play type, hydrocarbon prospect, field name, and basin name — allowing well data to be mapped within the appropriate exploration or production context.

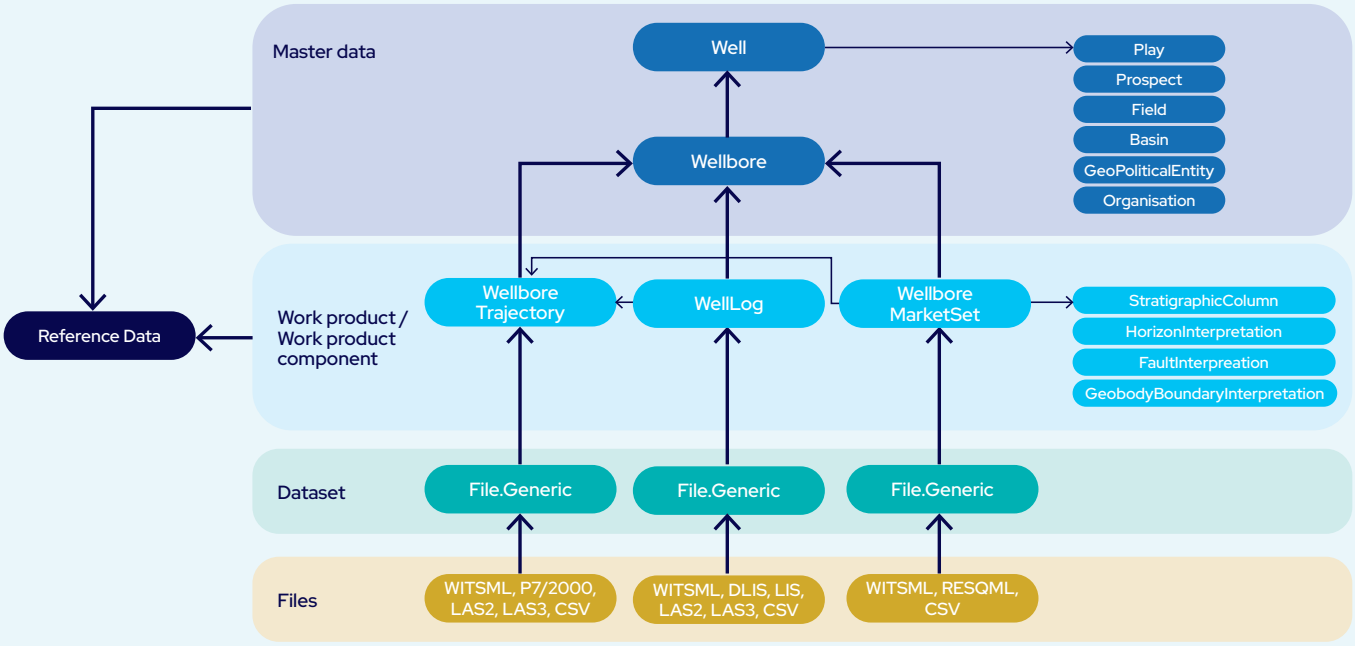


Figure A2: OSDU Well Model (not all WP/WPC and File Types are shown)

Professional Petroleum Data Management (PPDM)

The PPDM model is a well-established framework for managing petroleum data, developed and maintained by the PPDM Association. It provides a standardized data model and best practices for organizing upstream oil and gas information, and has been widely adopted across the industry.

Appendix B: Public Benchmark Datasets and Research Methodology

The seven-step workflow presented in this guide was developed and validated using public benchmark datasets. This appendix explains Ovation's research methodology and the data sources used.

Why public data?

Client datasets cannot be used for benchmarking without explicit permission, and showcasing client data publicly raises legal and ethical challenges. Public datasets provide an unrestricted foundation for workflow research and development, enabling Ovation to demonstrate methodologies openly.

Public data sources

Multiple government and institutional sources provide oil and gas datasets suitable for benchmarking:

1. **Texas Railroad Commission (RRC):** Culture data, well logs (mostly image format), documents, and production data. Seismic data is largely absent.
2. **Bureau of Ocean Energy Management (BOEM):** Comprehensive culture data, 2D/3D seismic data, well logs, documents, and production data. BOEM is integrating its database with the US Geological Survey (USGS) for US offshore seismic data.
3. **Department of Energy (DOE):** Project-based public data releases, including the Teapot Dome dataset.
4. **UK National Data Repository (NDR):** Comprehensive offshore petroleum information including culture data, 2D/3D seismic data, well logs, documents, and production data.
5. **Australia's Geophysical Archive Data Delivery System:** 2D/3D seismic data, well logs, and documents.

Downloading large datasets from government portals can be challenging due to connection limitations and data integrity issues. Ovation uses specialized web crawling tools to harvest public datasets efficiently.

The Teapot Dome dataset

Ovation selected the Teapot Dome dataset, released by the Department of Energy, as the primary benchmark for the workflow presented in this guide. This dataset is particularly suitable because of its comprehensive range of data types:

- GIS culture data
- Geological maps
- 2D/3D seismic data
- Thousands of well log files (1,210 LAS files)
- Core photographs
- Geological tops
- Production data

The completeness and variety of this dataset made it an excellent test case for validating the data cleaning and ingestion workflow across multiple data types and formats.

Appendix C: Supported Media Types

Ovation's processing facilities support an extensive range of media types spanning more than 60 years of technology. The list below covers the major formats we can read, recover, and transcribe. If you have media not listed here, please contact us - our inventory of more than 800 devices includes rare and customized equipment for unusual formats.

Reel, cartridge, and cassette — ½-inch

- 7-track reel tape — 1200 ft., 2400 ft. (200bpi, 256bpi, 512bpi, 800bpi) (read-only)
- 9-track reel tape — 600 ft., 1200 ft., 2400 ft., 3600 ft. (800bpi, 1600bpi, 3200bpi, 6250bpi)
- 14-track reel tape H.D.D.R. — 1200 ft., 2400 ft. (800bpi) (read-only)
- 3480 tape cartridge — SL, XL, XXL — 180MB, 250MB, 280MB
- 3490 and 3490E tape cartridge — SL, XL — 650MB, 850MB
- 3590 and 3590E Magstar MP tape cartridge — 10GB, 20GB, 40GB
- 3590H Magstar MP tape cartridge — 60GB
- 3592 J1A (GEN 1 Jaguar) tape cartridge — 60GB, 300GB
- 3592 (GEN 2 TS1120) — 100GB, 500GB, 700GB
- 3592JB (GEN 3 TS1130) — 128GB, 640GB, 1TB
- 3592JC (GEN 4 TS1140) — 1.6TB, 4TB
- 3592JD (GEN 5 TS1150) — 7TB, 10TB
- 9490EE cassette tape
- T9940A/B cassette tape

Cartridge and cassette — ¼-inch, 4mm, 8mm, 19mm

- 8mm Exabyte cassette tape — 8200, 8500, 8700
- 8mm Mammoth cassette tape — 8900, Mammoth LT
- 8mm Mammoth-2 cassette tape — M2 (20GB, 40GB, 60GB)
- 8mm AIT cassette — AIT-1, AIT-2, AIT-3 (100GB), AIT-3Ex (150GB), AIT-4 (200GB), AIT-5 (400GB)
- 8mm AIT-Turbo cassette — AIT-1 Turbo (40GB), AIT-E Turbo (20GB)
- 8mm VXA cassette — VXA-1 (33GB), VXA-2 (80GB), VXA-320 (160GB)
- 4mm DAT cassette tape — DDS1, DDS2, DDS3, DDS4, DDS5 (36GB)
- 19mm Sony D1 cassette tape — medium, large

Removable disc, disk, and NAS

- USB and FireWire (1394) portable drives (FAT32, NTFS, NFS)
- Storage area network (SAN) and network attached storage (NAS) drives
- Redundant Array of Inexpensive Disks (RAID)
- CD-R(OM), CD-RW — 650MB, 700MB
- DD-R(OM), DD-RW (double-density CD) — 1.3GB
- PD650 rewritable optical disk cartridge — 650MB
- DVD-R, DVD+R, DVD-RW, DVD+RW — 3.95GB, 4.7GB
- DVD+R DL (double layer) — 8.5GB
- DVD-RAM — 2.6GB, 5.2GB, 4.7GB, 9.4GB
- HD DVD — 15GB
- Blu-ray Disc — 25GB, 50GB
- UDO disk cartridge (ultra density optical) — 30GB
- PDD (Blu-ray), PDD23, BD-ROM, BD-R — 23GB
- 3½-inch and 5¼-inch floppy disk — SS, DS, LD, HD
- Removable memory — sticks, flash, cards, drives

- USB pen drives, flash drives, memory drives — JumpDrive, MicroVault, JetFlash, DataTraveler
- CompactFlash (CF) — Types I and II, MicroDrive
- Memory Stick (MS), Memory Stick PRO
- MultiMediaCard (MMC)
- Secure Digital Card (SD)
- SmartMedia (SM/SSFDC)
- xD Picture Card
- PCMCIA Memory Card — Types I and II
- PC hard disk card — Types I and II

Other media types

- 1-inch 21-track reel tape — 1200 ft., 2400 ft. (356bpi, 712bpi) (read-only)
- 35mm ICI-1012 terabyte reel optical tape
- Analog tape — TECHNO AM and FM (read-only)
- NTCP tape cartridge — 20GB
- Redwood SD-3 tape cartridge — small, medium, large
- Metrum VHS RSP-2150 (VLDS) cassette tape
- DLTtape cartridge — 2000, 4000, 7000, 8000, DLT1
- DLT VS Cartridge Tape — DLT1/VS80 (40GB), VS160 (80GB), VS1, DLT-V4 (160GB)
- DLT-S4 (3rd Generation SuperDLT) — 800GB
- Super DLTtape — SDLT1 (220/320), SDLT2 (600) — 110GB, 160GB, 300GB
- TK tape cartridge — TK30, TK50, TK70, TK85, TK87, TK88, TK89 (CompacTape I, II, III, IIIXT, IV)
- Sony DTF cassette tape — GW240S, GW730L
- Sony DTF2 cassette tape — GW2-60GS, GW2-200GL (200GB)
- Sony SAIT1 and SAIT2 tape cartridge — 500GB, 800GB
- LTO/IBM 3580 Ultrium tape cartridge — LTO1 (100GB), LTO2 (200GB), LTO3, LTO4 (TS2240) — 400GB, 800GB
- 19mm Ampex DD2 DST cassette tape — small, medium, large
- ¼-inch ADR and ADR2 — 15GB, 25GB, 30GB, 60GB, 120GB
- ¼-inch 3570 Magstar MP B-format, C-format, C-XL format tape cartridge
- ¼-inch Mini QIC cartridge (including Irwin, Colorado, Travan, DC2xxx, MC3xxx)
- ¼-inch Mini QIC cartridge (including Iomega Ditto Max, Ditto Easy, Ditto)
- ¼-inch Standard QIC cartridge (including DC6xxx, DC9xxx, SLR2–SLR7)
- ¼-inch Standard QIC SLR32 (13GB, MLR1), SLR40, SLR50 (MLR3), SLR60, SLR75, SLR100, SLR140
- 3½-inch high-capacity floppy disk — LS-120, LS-240 (SuperDisk), HiFD (200MB)
- 3½-inch floptical disk — 21MB
- 3½-inch, 5¼-inch, 12-inch, and 14-inch M.O., WORM, optical disk cartridge (various including IBM, Plasmon, Sony, HP, Panasonic)
- 12-inch optical disk cartridge — Philips LMS/Plasmon LMS 4000, 6000, 8000
- 5¼-inch Bernoulli Disk — 44MB, 90MB, 105MB, 150MB, 230MB
- Iomega Jaz Disk — 1GB, 2GB
- Iomega PocketZip (Clik!) Disk — 40MB
- Iomega Zip Disk — 100MB, 250MB, 750MB
- SyQuest disk cartridge — EZFlyer (135MB, 230MB), SyJet (1.5GB), SparQ (1GB)
- 3½-inch SyQuest disk cartridge — SQ310 (105MB), SQ327 (270MB)
- 5¼-inch SyQuest disk cartridge — SQ400 (44MB), SQ800 (88MB), SQ1100 (105MB), SQ2000 (200MB)

Documents and other formats

- Paper documents and vellum film
- Drawings on paper and vellum film
- Punched cards, microfiche, film, paper tape (input only)


OvationData