Lineage Metadata as a Critical Component of Data Trustworthiness for Subsurface and Analytics Applications*

Philip Neri[1]

Search and Discovery Article #42408 (2019)**
Posted August 5, 2019

*Adapted from oral presentation given at 2019 AAPG Annual Convention and Exhibition, San Antonio, Texas, May 19-22, 2019

**Datapages © 2019 Serial rights given by author. For all other rights contact author directly. DOI:10.1306/42408Neri2019

[1] Energistics Inc., Houston, TX (philip.neri@energistics.org)

Abstract

Knowledge regarding the origin of data being used for modeling or analytical purposes is essential to establish trust in the results. Source encompasses not only the device(s) used to create the data, but also the history of subsequent operations performed on the data. Trust is not a quantity or a parameter. For data originating outside the controlled data environment of a company or institution, trust in the usability of the data is a decision based on varying criteria. The completeness of the lineage metadata will play a significant role in allowing the data receiver to establish whether the data can be trusted “as is” or whether a verification process is required prior to usage. To address this issue, companies providing data to customers, partners or other entities would seek guidance on what lineage information requirements, or in the absence of a formal request provide the data with accompanying metadata as they saw fit. Such case-by-case procedures are onerous and, in most cases, do not entirely satisfy the requirements of the recipient. A number of factors over the past 20 years have exacerbated this issue. Foremost is the increasing complexity of upstream datasets and their exponential growth in volume. The trend towards shorter oilfield development project cycles puts further pressure on staff. Finally, the attrition in subject matter experts (SME) able to assess data validity limits the resources that can be deployed. Modern technologies such as machine learning can provide valuable efficiencies to overcome the lack of lineage metadata. However, this would still require some level of supervision and outcome verification, and the level of detail regarding e.g. exact processing history (which software package(s), which version, what parameters, identity of users, dates, etc..) would be limited. The more rational and less ambiguous approach is to make sure that all the necessary information is attached to the data. Starting in 2010 the industry came together to define a standard for metadata, including lineage, data assurance and integrity components, and this was published in 2016 as the Energy Industry Profile (EIP) of ISO 19115-1:2014. It is an ISO Conformance Level 1 profile of the published international standard. The biggest challenge lays ahead: convincing all industry players that the investment in implementing the metadata standards is an effort that will deliver a step-change in data trustworthiness while freeing up valuable SME resources.

ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

Lineage Metadata as a Critical Component of Data Trustworthiness for Subsurface and Analytics Applications

Philip Neri

Energistics

Agenda

- Digital collaboration requires connectivity
- Moving data while retaining knowledge
- The concept of data lineage
- Applications in geoscience projects
- Applications in analytics
- Industry standards
  - RESQML

ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

Digital Collaboration Requires Connectivity

Three Criteria for a Digital System

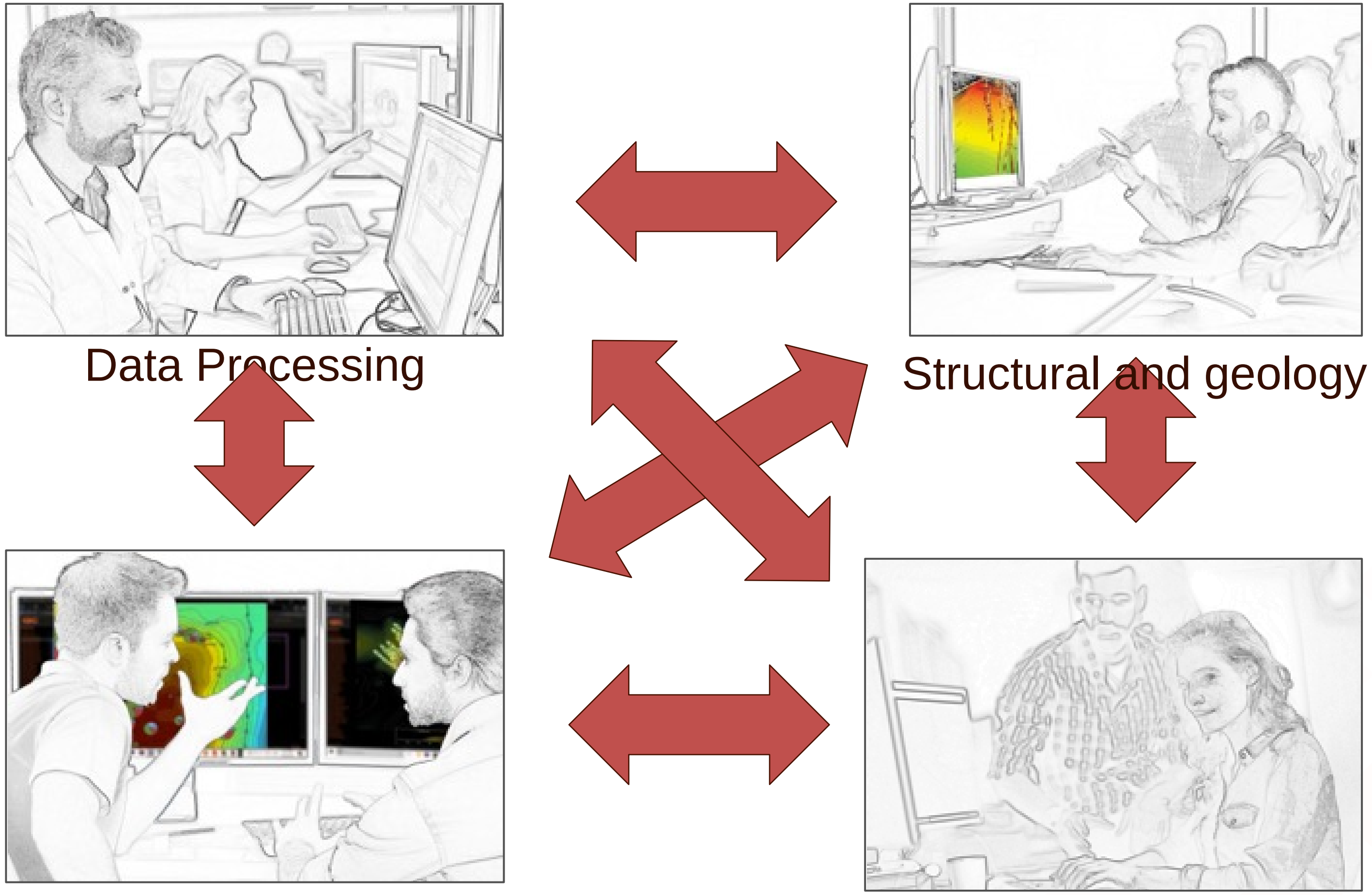
» DATA: Accessible – Verified – Informed – Secure

» ANALYTICS: Value – Insight – All-Inclusive – Robust

» CONNECTED: No Silos – Shared – Completeness - Integrity

Presenter's notes: Started in 1990 as POSC…rebranded in 2006 as Energistics Added to standards development mission…promote standards adoption Volunteer led, facilitation is key

Subsurface Modeling is increasingly complex



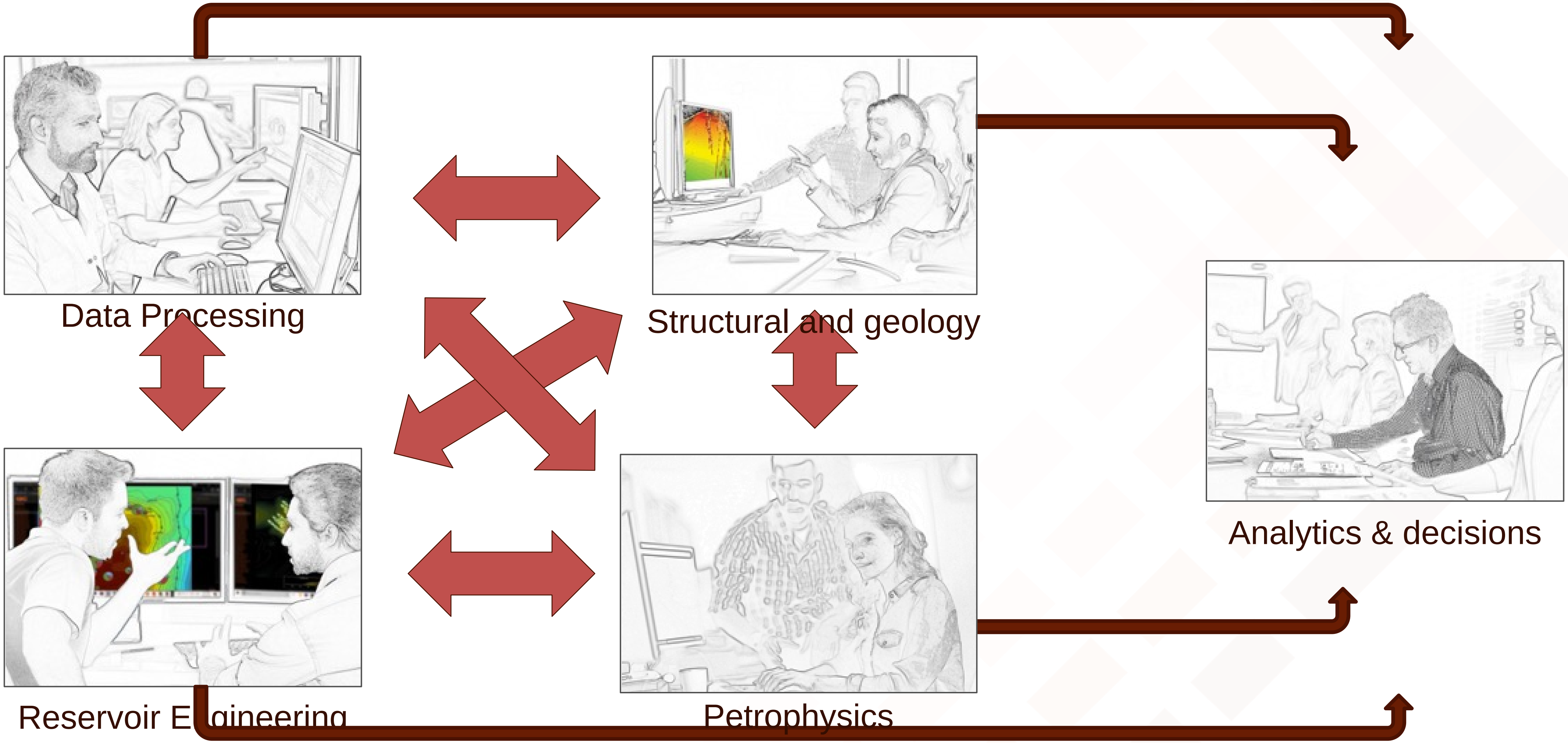
Data Processing

Structural and geology

Reservoir Engineering

Petrophysics

And needs to run faster



Data Processing

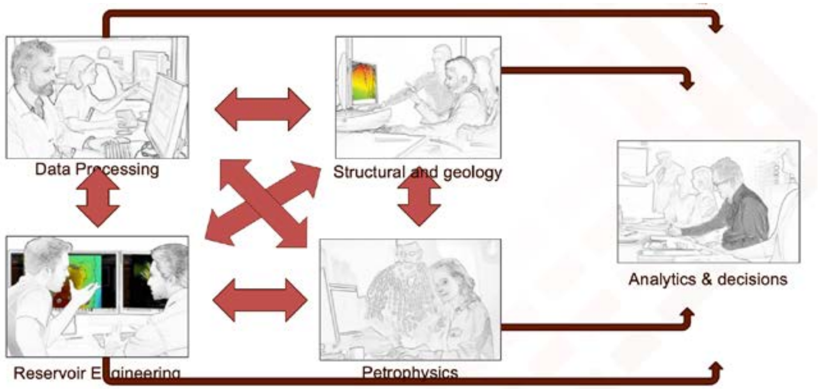
Structural and geology

Reservoir Engineering

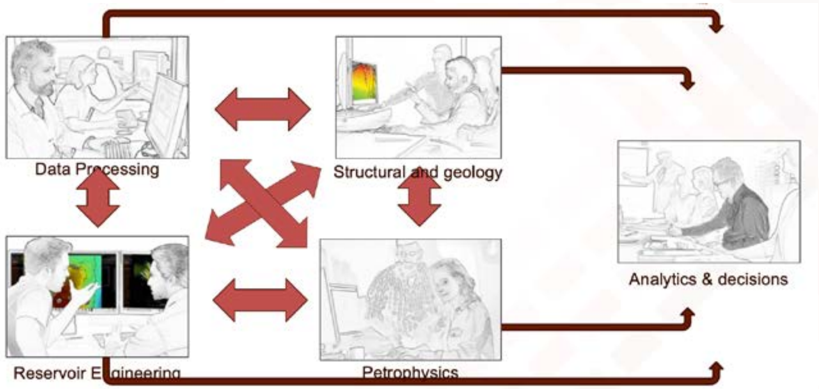
Petrophysics

Analytics & decisions

Data ecosystems are increasingly connected

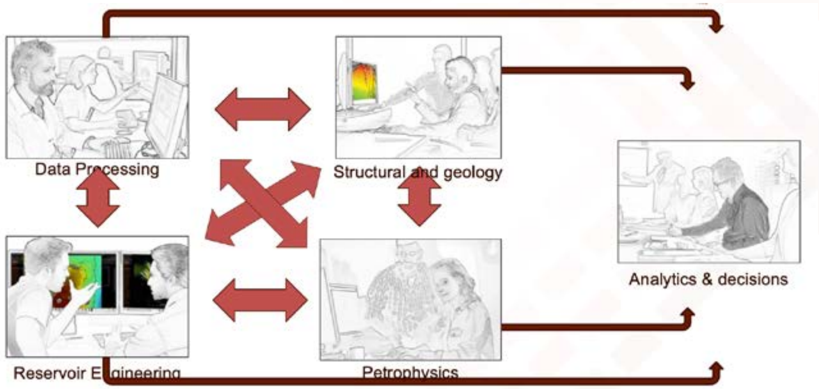


Operator



Partner

Data Brokers



Regulator

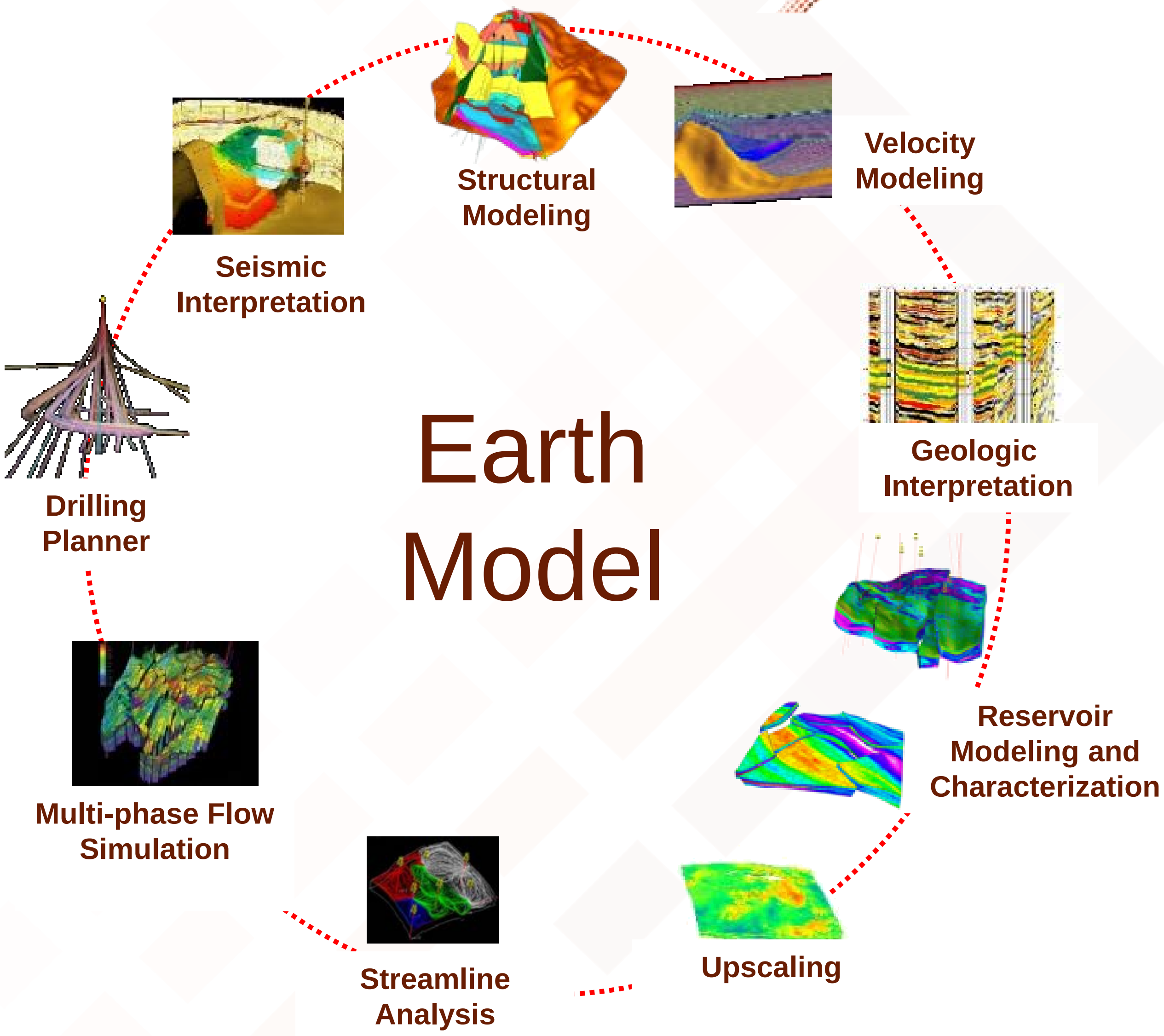
ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

Lineage: Moving Data While Retaining Knowledge

Modern workflows are more complex and durable

- An asset team will invoke many disciplines from new data to final model(s)
- Data is sourced from different suppliers
- Assets are re-visited for monitoring and EOR purposes over long periods (10+ years?)



Earth Model

Seismic Interpretation

Structural Modeling

Velocity Modeling

Geologic Interpretation

Reservoir Modeling and Characterization

Upscaling

Streamline Analysis

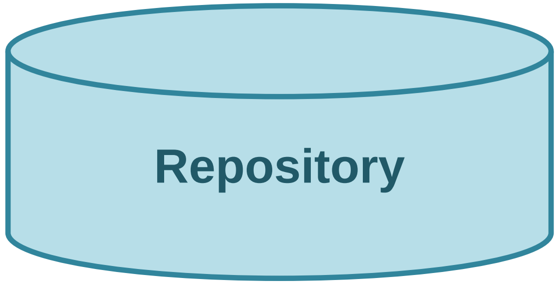
Multi-phase Flow Simulation

Drilling Planner

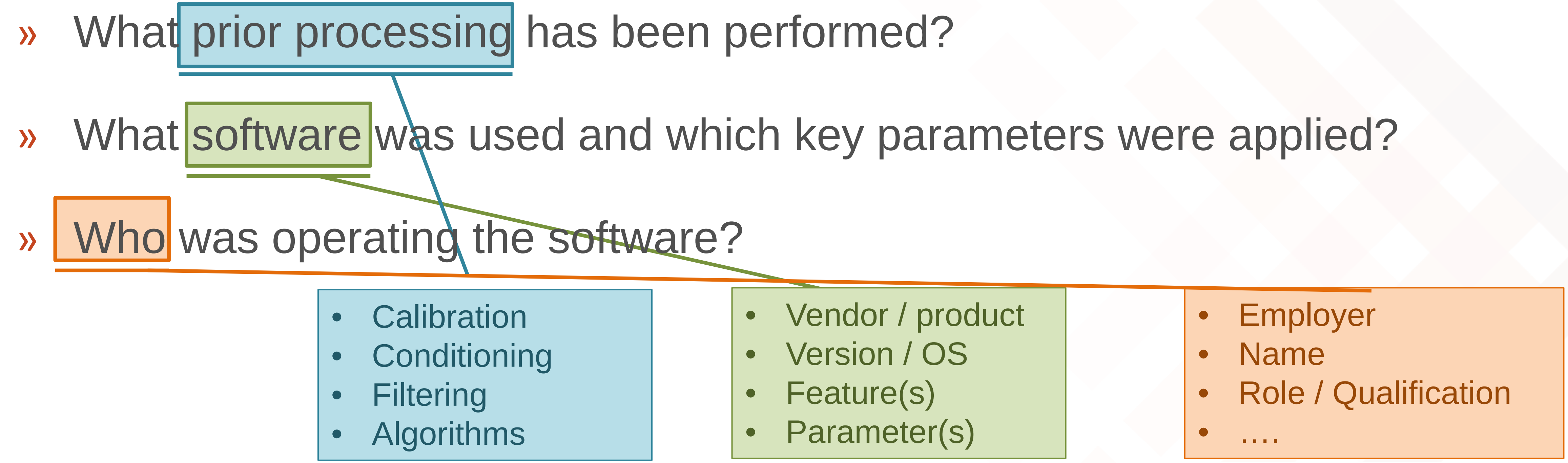
Presenter's notes: Sharing earth model data between products
Using both XML and HDF5
SIG working on V2.0 now (French influence)

Is data ready for use and can it be trusted?

- What prior processing has been performed?
- What software was used and which key parameters were applied?
- Who was operating the software?



Is data ready for use and can it be trusted?



- » What prior processing has been performed?
- » What software was used and which key parameters were applied?
- » Who was operating the software?

- Calibration
- Conditioning
- Filtering
- Algorithms

- Vendor / product
- Version / OS
- Feature(s)
- Parameter(s)

- Employer
- Name
- Role / Qualification
- ….

Lineage is not always simple

- Capturing objective data transformation steps
  - Data creation (date,…) e.g. markers
- Documenting other activities
  - Partial deletes
  - Manual / interpretative edits
- Observations
  - Lithologies, classifications, …

Is data ready for use and can it be trusted? [2]

- Users need to know to make informed decisions
  - Reprocess data that does not meet required benchmarks
  - Set aside data that does not meet benchmarks
- Lack of information results in a higher workload
  - Qualitative verification based on experience with similar data
  - Quantitative verification using available (metadata) information
  - Workload can include searching for information in unstructured documents

Considerations for Analytics

- Data must not be co-mingled if it is of differing quality
- AI / LM front-end to perform triage at data ingestion
  - Apply criteria related to available and credible metadata
  - Reject or group into sets with different weightings
  - Unsupervised data searches into unstructured documents has risks
- AI / ML judging data quality without prior criteria and/or training?
  - Fast-evolving field of investigation
  - Litmus test: would you make a decision based on it?

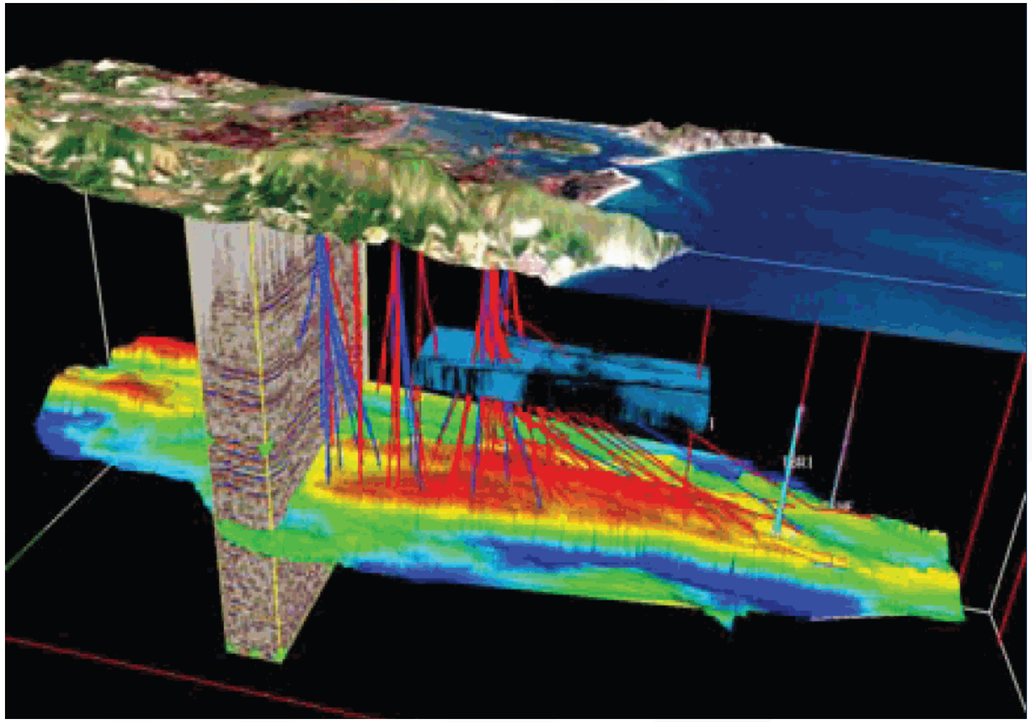
ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

The RESQML Standard

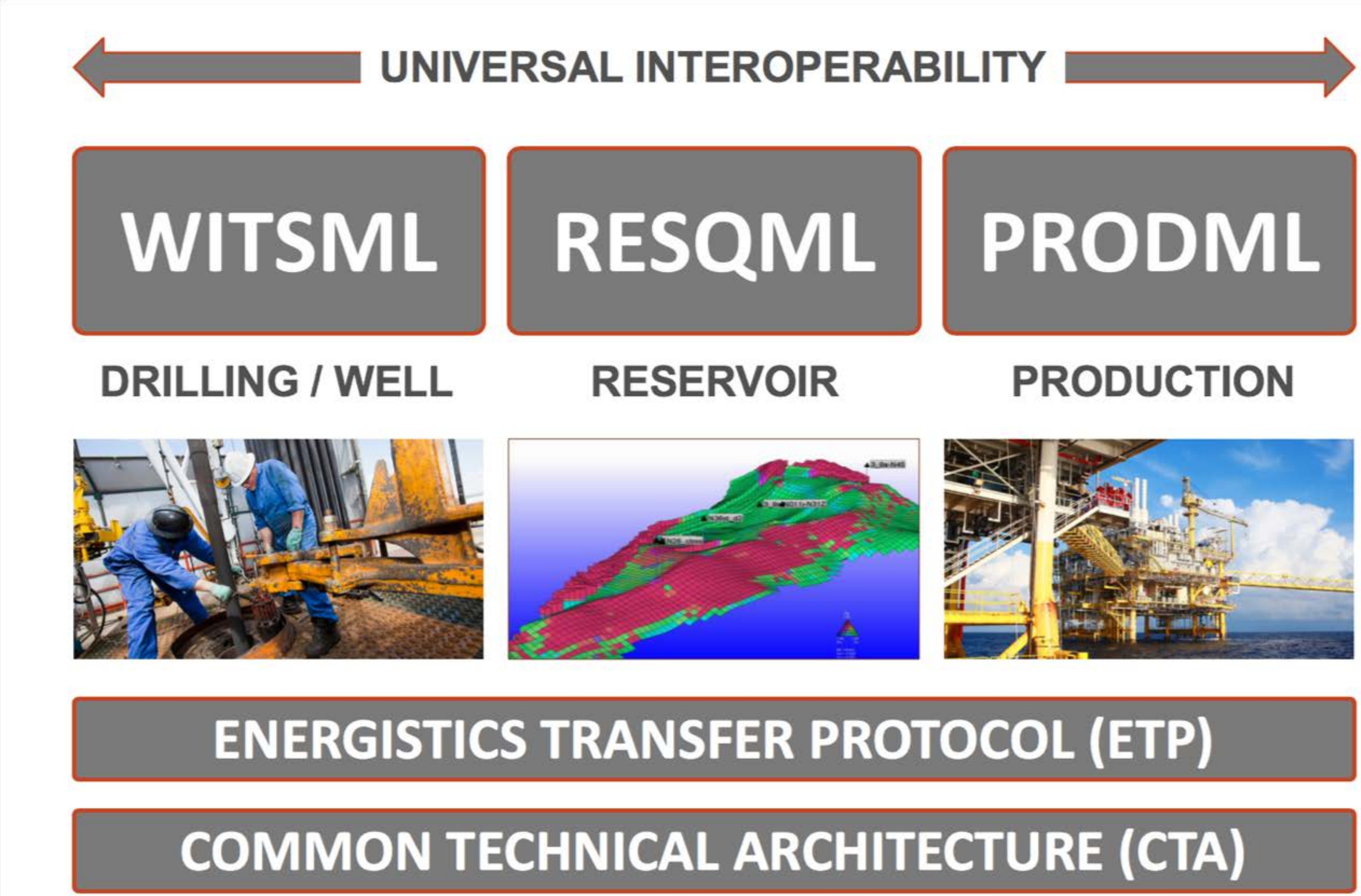
RESQML v2.0.1

- » RESQML is not an acronym
- » RESQML moves earth models
  - Each part individually
  - As a package (HDF5)
- » In a vendor-neutral way
- » Using modern technology



"Using RESQML, the same project type and size took less than an hour to transfer and check, knowing that the standard handles comprehensively all reference and unit information."

Energistics' Spectrum of Standards



UNIVERSAL INTEROPERABILITY

WITSML	RESQML	PRODML
DRILLING / WELL	RESERVOIR	PRODUCTION

ENERGISTICS TRANSFER PROTOCOL (ETP)

COMMON TECHNICAL ARCHITECTURE (CTA)

Presenter's notes: Our flagship family of standards…RESQML, WITSML, PRODML
Collect data once, use in all three verticals
Seamless integration of subsurface data

ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

Conclusions

Conclusions

- Data is “orphaned” if it is not referenced
  - References
  - Ownership
  - Lineage
- Modern multi-disciplinary workflows and long life-cycles
  - Cannot assume that people-related knowledge is available
  - Knowledge must be able to move with the data to different ecosystems
- Standards are detailed and require investment
  - Time, resources and people
  - It is a long-term, strategic investment

ENERGISTICS

ADOPT > ADVANCE > ACCELERATE

THANK YOU

QUESTIONS?

www.energistics.org