CCUS: 4014747

End-to-End Workflow for Managing Large Volume Data from CCUS

Jess Kozman*[1], Stuart Lowery[1], James Lamb[1], 1. Katalyst Data Management

This paper was prepared for presentation at the Carbon Capture, Utilization, and Storage conference held in Houston, TX, 11-13 March.

Abstract

Challenges for subsurface carbon storage work include the large volume of data delivered from Measurement, Monitoring, and Verification (MMV) protocols defined by government regulatory agencies. Volumes can exceed one petabyte (1000 terabytes) in a year of operation from wells fully instrumented using distributed fiber optic sensors (DFOS). Processing and curation of data volumes will be critical for operators defining future sustainable carbon management projects. Petroleum geoscientists are familiar with storage technologies for delivering large volumes of digital data that is Findable, Accessible, Interoperable, and Reusable (FAIR). We have evaluated the adoption and adaptation of industry accepted optimum practices for the high sampling density data sets being delivered by time-lapse monitoring programs in subsurface reservoirs, before, during, and after CO2 injection. Many carbon storage projects are at planning stages where well engineers more familiar with smaller data sets require data management plans in well budgets, and petabyte scale data management is new territory for their discipline and function. A fully integrated CCUS project requires effective custody handover of large time-series digital data streams from the design phase to provide compliance with regulatory reporting guidelines. This can measurably contribute to a positive return on investment across long-term asset lifecycles. Increased visibility, transparency, and democratization of large volume data can drive innovation for ingestion, enrichment, and consumption of MMV data for sustainable carbon storage projects. The deployment of new field-validated, intelligently tiered data management workflows for large data volumes will benefit practicing geoscientists who work with DFOS data. This study applies new and additive data modeling concepts developed by industry consortia on open source and technology agnostic data platforms, based on newly available industry standards. The combination of experience with tiered storage, a skill pool of data subject matter expertise, and embedded data workflows accelerates successful delivery of data to support critical business decisions for CCUS exploration, planning, development, and operations. This work shows that development times can be reduced and extensibility to new data types can be accelerated by application of novel data management methodologies.

Introduction

The energy and resource industry is inherently cyclical in activity levels, and those activities drive data management methodologies through acquisition and delivery of expanding volumes of geotechnical data (Mohammadpoor and Torabi, 2020). Early cycles for subsurface data were marked by rapid increases in recording channels in the early 1980's, widespread use of 3D and 4D surveying from the 1990's to the 2000's, and recent upturns in visibility of "Big Data", data science, and data analytics. An additional resurgence in attention to digital data sets has been driven by media hype around a combination of emerging digital technologies comprising Generative algorithms, Large language models, Artificial intelligence, and Machine learning (Reinert, 2023), which data managers are referring to by the nested acronym GLAM. As more resource industry operators emphasize low carbon projects and initiatives, we are also seeing a rapid expansion of digital data entering data management platforms and ecosystems from DFOS systems, many of which are being deployed to support continuous monitoring of CCUS reservoirs (Jacobs, 2022). Large volumes of distributed fiber optic sensor data and the potential for regulatory requirements to retain raw data sets for decadal timeframes are now ushering in a new set of dimensions vital to Big Data management. Professional data management practitioners in the resource sector are now being tasked with developing managed workflows for Distributed Acoustic Sensor (DAS) data as it starts to be delivered to enterprise-scale data platforms (Gabriel and Wheelwright, 2023). The large volumes of data involved are pushing the boundaries of existing workflows and inspiring new strategies pulling from multiple areas of data management. Here we review the process and culmination of assembling a collection of industry accepted optimum practices and lessons learned, into a deployable end-to-end solution for managing data from CCUS projects.

Theory and/or Methods

Embedded data workflows for large-volume continuous reservoir monitoring data sets from CCUS projects should build on existing data management methodologies. For this project we evaluated existing solutions that have demonstrated and published business and financial benefits. We compiled information about business tested workflows for petabyte scale (Barrow, 1997) resource industry digital subsurface data currently under management on optimized tiered storage. The available datasets comprise more than 100 petabytes* of digital data from global resource operators (oil and gas and mining companies), service and acquisition providers, data processing facilities, national oil companies, and government regulators. From this analysis we identified data consumption dimensions that define data management requirements for continuous monitoring digital data, including pervasiveness, proliferation, propagation, and persistence. A data workflow methodology for approaching the unique challenges of fiber optic data is to adopt and adapt existing industry accepted optimum practices that have been proven to provide a positive return on investment with other, similarly large geotechnical data sets. In this context, the business objectives of data management are to provide data which is Findable, Accessible, Interoperable, and Reusable, or FAIR (Wilkinson et al, 2016), and independent software vendors will continue to provide subsurface data management solutions that enable FAIR technology, methodology, best practices and optimized work flows. As seen in Figure 1, this can be accomplished by applying existing data management best practices already adopted by the resource sector and other data intensive, capital intensive, and highly regulated industries. (Sadowski, 2019). Since CCUS operators are expected to

* Both the National Institute of Standards and Technology (NIST) and the Institute of Electrical and Electronics Engineers, Inc. (IEEE) differentiate between storage volumes calculated using decimal and binary arithmetic. In this paper we adopt the convention of 1 petabyte = 1000 terabytes using decimal prefixes. The 1998 International System of Units (SI) prefixes for binary multiples approved by the International Electrotechnical Commission (IEC), including the term "pebibytes", while more accurately describing binary storage, have not been widely adopted by storage providers or users and would be confusing to readers. See the NIST Reference on Constants, Units, and Uncertainty for a full explanation: https://physics.nist.gov/cuu/Units/binary.html

exhibit relationships that scale between organizational size and data complexity (Kobayashi et al, 2021), that analysis can be used to select the most appropriate level of data management capability maturity and deploy a solution that will likewise scale with the growth of incoming CCUS continuous monitoring data sets.

	Pervasiveness	Proliferation	Propagation	Persistence
	Data is used across multiple functions and disciplines	*Multiple copies are duplicated rapidly within a system*	*Data spreads widely across multiple systems*	*Data continues to be used for a prolonged period*
• Metadata • Modelling & Design • Reference & Master	NOT FINDABLE			
• Architecture • Governance • Security	FINDABLE	NOT ACCESSIBLE		
• Quality • Integration • Content	FINDABLE	ACCESSIBLE	NOT INTEROPERABLE	
• Warehousing & BI • Security • Storage & Operations	FINDABLE	ACCESSIBLE	INTEROPERABLE	NOT RE-USABLE
	FINDABLE	ACCESSIBLE	INTEROPERABLE	RE-USABLE

Figure 1. Relationship between dimensions of large volume data management, FAIR implementation principles, and DMBOK2 knowledge areas. If any intersection of a dimension (top horizontal row) and set of knowledge areas (left vertical column) is left unmanaged (or willfully mismanaged), one of the FAIR implementation principles will be violated (red text). The conclusion is that CCUS data can only be sustainably FAIR if all the knowledge areas are applied across all the dimensions.

The behavior of an organization at a data technology frontier (Al-Obaidan and Scully,1991) is controlled by a ratio between the outward acting expansion of data democratization that supports data driven decision making for CCUS projects, effectively balanced against the inward acting pull of access control, confidentiality, security and governance for the supporting data (Kozman, 2023). The dimensions that control and describe expansion of large volume digital data from subsurface operations have been previously described (Kozman, 2003) and updated as new data types enter into more common usage. The dimensions most frequently noted in management of DFOS data from CCUS projects are

- Pervasiveness: The usage of data products across multiple technical functions and disciplines within an organization, which can result in multiple copies needing to be managed over the lifecycle of a project, and the need for multiple data custodians and data custody handoff points in the data lifecycle. This can be measured by the number of unique users accessing a particular data product, and the number of unique reporting structures represented by those users. Unmanaged pervasive data can lead to data finding decision latency, and this latency can be reduced by agreeing to standardized and consistent metadata fields in a catalogued index, appropriate levels of design for conceptual, logical, and business data models, and managed systems of record for reference and master data.
- Proliferation: The speed at which these multiple copies can be generated on multiple data applications or systems in an enterprise data ecosystem. This can be measured by analyzing the output from industry standard de-duplication algorithms operating on network attached or object level storage systems, and correlating with the growth and expense of maintaining data storage systems. Unmanaged proliferating data can lead to data access decision latency. This latency can be reduced with well documented and future-proofed technology agnostic enterprise data

architectures, a combination of top-down and bottom-up data governance frameworks, and data classifications schemes that balance access and enforceable cybersecurity standards based on discoverable data confidentiality, privacy, and protection.

- Propagation: The rate at which the same data product is re-formatted or re-ingested into multiple applications or internal fit-for-purpose databases. This can include non-unique versions or unmanaged personal data stores. This property can be measured by the number of times a single data entry is moved between applications in a captured data workflow. Unmanaged propagating data can contribute to data interoperability decision latency, and is addressed through quality and confidence metadata flags that allow users to determine if data is fit-for-purpose in an enterprise master data system, technologies that support sharing of data between applications without duplication, and content management frameworks that allow multiple catalog items to link to optimized bulk storage of binary large objects on tiered storage.
- Persistence: The longevity of data products on data storage and archiving systems and platforms, including multiple iterations of upgrades to media and technology. This is easily measured by categorizing the earliest creation dates of data types that can be accessed and extrapolating to a useful life-span or regulatory requirement for retention. Unmanaged persistent data manifests in extended periods of data reusability decision latency, and can be reduced through the use of robust warehousing and business intelligence strategies, security standards that recognize and enforce retention and discovery standards, and operational storage processes that anticipate the value of legacy data in the later stages of a project lifecycle.

Industry standard principles for balancing these requirements is provided by the knowledge areas and environmental factors described in the Data Management Association (DAMA) Body of Knowledge v. 2 or DMBOK (Earley and Henderson, 2017). Figure 1 shows the relationship between FAIR implementation principles, DMBOK2 knowledge areas, and dimensions of rapidly growing CCUS continuous monitoring data. A properly scaled solution will allow a CCUS operator to "future proof" their data platform for multiple dimensions of data ingestion, enrichment, and consumption. The ability to deliver geotechnical data against a FAIR data implementation profile (Arofan and Simon, 2022) has been consistently ranked as a way to remove barriers to investment (Mejía and Aliakbari, 2023) and levels of data management capability maturity continue to be correlated with financial return on investment (Gimenez and Kozman, 2004, Kozman, 2014) and value creation from increasing volumes of digital data (Tordo, Tracy and Arfaa, 2011). In the following section we describe the development of an end to end embedded digital data workflow that builds on these existing industry data management standards and methodologies and delivers FAIR data for continuous carbon operations monitoring.

Results

Distributed fiber optic instrumentation for CCUS projects provides data at higher orders of magnitude for both spatial and temporal sampling density, across multiple physical properties in the subsurface. As operators move from CCUS projects utilizing depleted oil and gas fields into deep saline reservoirs to access larger volumes of available pore space, continuous active or passive monitoring at multiple kilometer-depth wells will be required. With the ability to sample continuously at sub-meter lengths over injection periods spanning decades, these projects will become petabyte-scale data producers fairly quickly (He and Paulsson, 2021). Field validation studies with resource sector organizations have shown that an embedded data workflow methodology (DWM) is effective in managing and delivering geotechnical data from boreholes to support new solutions and maximize efficiency (Li et al, 2023). Previous studies focused on communicating large volumes of data consumed from automated systems, and used across multiple teams, functions, disciplines, and business units, requiring collaboration between geoscience and data science subject matter experts. As in the DWM case study, our development of an end-to-end data workflow for CCUS monitoring data has evolved over multiple years from a proof of concept to continuous improvement (Figure 2). The development included important concepts from other areas of professional grade data management, including data custody handoff (Kozman, 2016), and the assignment of self-identified data governance roles and responsibilities (Li et at, 2023). Data custody handoff points are based on a similar concept built into exiting data standards for duration-based and continuous production monitoring with real-time digital data streams (Henson, 2021). The result of the applications of these optimum accepted practices is a sustainable end-to-end workflow that can be deployed on existing data ecosystems and scaled to accommodate exponentially increasing data volumes as fiber optic sensing technologies advance and are adopted by different sectors of the energy industry.

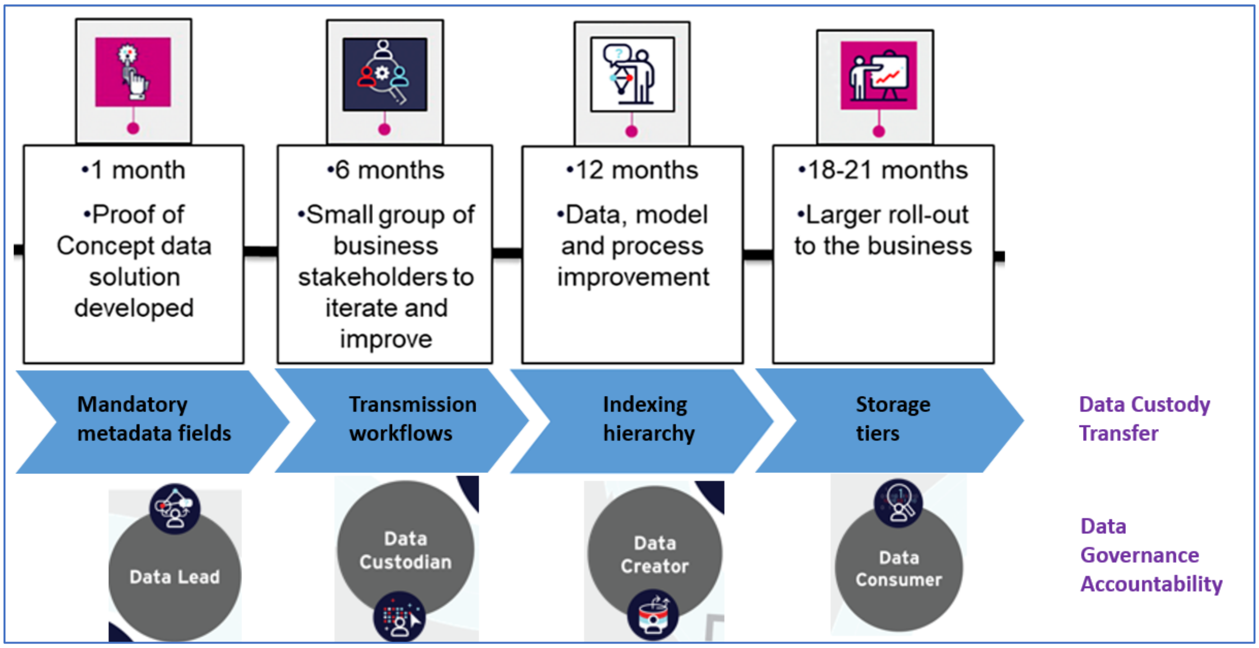

Figure 2. Development cycle of an embedded Data Workflow Methodology (DWM) for continuous reservoir monitoring DAS data (After Li et al, 2023, used by permission). The workflow shows a timeline for deployment, the transfer points for custody of data workflow elements, and the self-identified data governance roles that participated and validated the workflow at the different stages.

The DWM recognizes that in order to use and trust a data pipeline, multiple data roles need to work in a whole of organization approach with a single common goal for deriving value from continuous improvement. The program identifies 4 stages in a digital data workflow and 6 roles that recognize subject matter expertise. Highlights from data workflow mapping initiatives include increased autonomy, control and feedback, value, and trust in the data to support impact on the business bottom line.

Fully instrumented monitoring wells on existing projects are already generating data volumes requiring up to a petabyte of storage and compute capacity at the well head (Isaenkov et al, 2021). At these volumes, current data upload speeds, even across broadband wireless, are impractical for transmitting raw DAS measurements in real time to a central server for decision making. In fact, in some pilot projects, raw data is collected on RAID arrays of solid state drives at the wellsite and subsequently transported by vehicle to a data center processing and interpretation. With DAS arrays now able to generate multiple terabytes of raw acquisition data in a single day (He and Paulsson, 2021), developing a system to make insights available to end users from a remote wellsite location to a central office would seem to overwhelm most real-time wireless broadband systems. Scalable real-time processing systems can deliver up to 500 terabytes (TB) of fiber optic data in a week to support processing, interpretation and real-time monitoring. Data streaming solutions will need to be scalable to handle raw data on the order of 150 terabytes per day for multiple years of automated operations monitoring through web-based front end solutions (Bakka, 2022). To meet these requirements, we examined the current capabilities of open source industry standards based on vendor agnostic data ecosystems and platforms (Gabriel and Wheelwright, 2023).

We have developed a unique and scalable end-to-end workflow by adopting and adapting existing and proposed work product components (WPC's) as jointly developed by industry contributors to the OSDU® Data Platform and Open Footprint™ forum. Members of these groups include over 20 operators, national oil companies and regulators, and close to 200 technology and service providers to the resource industry. The workflow is designed to facilitate the transmission and storage of raw and preprocessed data, access to processed data sets, and interoperability and re-use of interpreted data volumes. It begins with collection of raw data and application of pre-processing algorithms including decimation, deconvolution and stacking. In most cases this reduces the data volumes to the point that they can be transmitted over an LTE broadband wireless link (see example in Figure 3). A recent Southeast Asia case study also demonstrated the feasibility of using high speed satellite links from remote locations to ship data volumes after pre-processing.

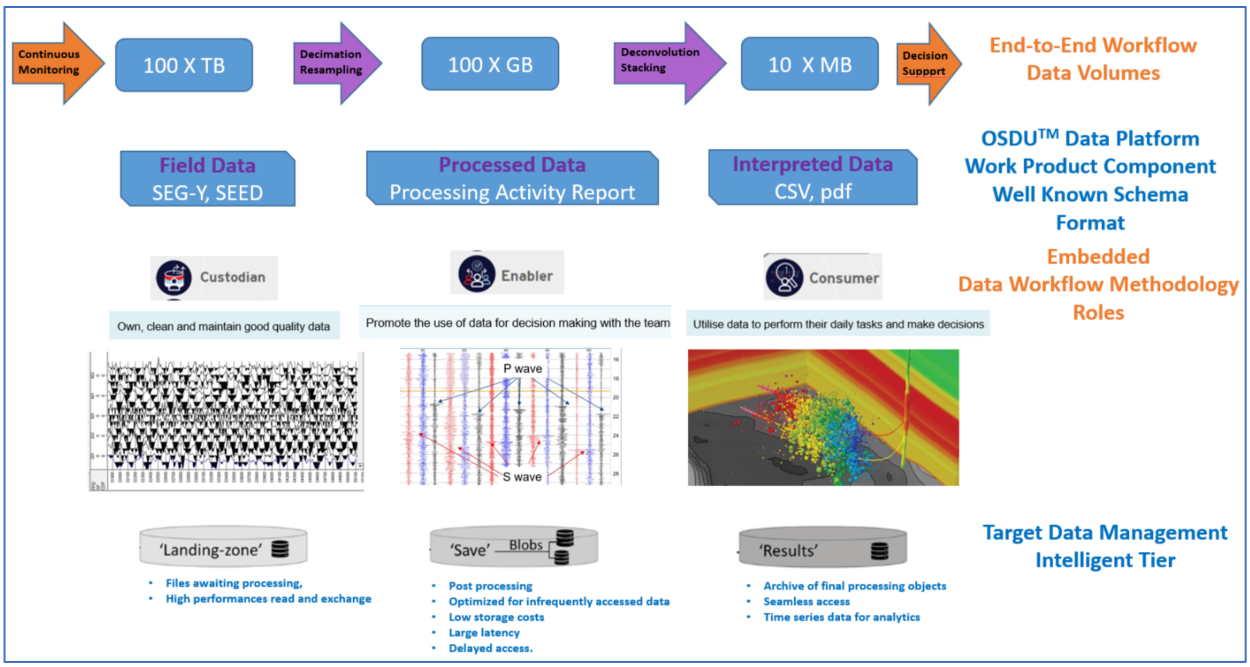

Figure 3. The end to end workflow as developed for continuous monitoring of CCUS digital data, showing the reduction in data volume and use of intelligent tiers of optimized storage (after Azzola et al, 2022).

Acquired data can be routed by automated workflows to a "landing zone" in the nearest storage provider's point of presence, where it can be read and exchanged by high performance computing resources for subsequent processing steps. After further processing the raw acquired data can be moved to a high-latency, low cost level of intelligent tiered storage. Data awaiting pre-processing is managed by a data custodian, who has accountability for data quality and completeness in the embedded workflow, while processed data is managed by a data enabler, who is accountable for the use of data within a team to support data decisions. The version of the processed data in a storage tier that is optimized for infrequent usage will meet the requirements of government regulators, who are in some cases mandating that operators store data related to their carbon storage projects for up to 25 years (BP, 2021), or a period of time equivalent to the duration of injection (Dean and O'Brien, 2024), without overwhelming immediate access storage tiers. A mandatory set of minimum metadata attributes is defined, agreed by the data custodian, and collected and stored with the indexed data to facilitate search and filter capabilities through a web-enabled and map-based user portal.

Final results and interpreted data can be maintained on an intelligent storage tier with seamless access, where an industry standard Application Programming Interface (API) can make them available to analytics, machine learning, or artificial intelligence programs. The interpreted data can be presented as a point cloud of monitored subsurface measurements imaged as points in space (Staněk, 2022). Accountability for data being available for organizational decision support is assigned to a data consumer at this point in the workflow. Recent harmonization work within the data standards forums and special interest working groups is focused on how continuous monitoring data from the subsurface can be successfully linked to surface-based carbon counting workflows for validation of mass or volume balanced calculations of carbon stored and sequestered in geologic reservoirs (Bates et al, 2022). This will allow opportunities for the efficacy of carbon sequestration in a target reservoir to be verified through mass and volume balance calculations integrated with surface facility carbon counting. The proper audit trails to provide the transparency and trust required for this full accounting will only be accessible if the basic tenets of FAIR data management and delivery are adhered to at all the steps and stage gates in a fully embedded data workflow methodology.

Discussion

The promise of DFOS is to enable continuous real time monitoring of multiple physical properties in the subsurface (including temperature, strain, and both active and passive acoustic response), in three dimensions over time, with either permanently installed or fit-for-purpose deployment of cables. Multiple options for deployments are currently available and tested, including on tubing, behind casing (Carbon Capture Journal, 2020), or suspended in a fluid filled borehole (Pevzner et al, 2023). Unused or underutilized “dark fiber” in existing installations such as a subsea umbilical or telecommunications network can also be activated to be used as a sensor (Zhan et al, 2020).

Fiber optic deployments and a new generation of fiber optical sensing have the advantage of making acquisition of longer-term decadal monitoring projects more affordable, and this requires a parallel set of workflows to make end-to-end data delivery workflows also viable and affordable. With the ability to augment or replace more conventional surface seismic acquisition using nodes, simultaneous downhole acoustic recording will drive increasing volumes of raw, processed and interpreted data onto enterprise data management platforms. These volumes will be further increased by volumes of data reprocessed or digitized during exploration and development of the target carbon storage reservoir. Fiber optic sensing also lowers environmental impact, exposure to HSE risk in the field, mobilization costs, and technical challenges of repeatability of receiver equipment and locations (Isaenkov, 2023). Distributed acoustic sensing arrays also record larger lateral imaging distances than geophone deployments (Miller et al, 2016). But all these advantages come with the added burden of Big Data volumes that strain current embedded data workflow methodologies.

Operator experience has been that in dealing with the drilling and completions or well planning discipline and functions within a typical resource operator, recognition of the large data volumes involved with continuous reservoir monitoring technologies can create a barrier to adoption. Presenting an industry standard end-to-end data workflow that addresses data storage at the wellsite, additional power and computing requirements on the rig floor, connectivity and bandwidth issues, and storage and archiving of raw and processed data, can assist with breaking down those barriers and facilitate a discussion around the return on investment from deploying fiber optic sensors during well construction versus repeated 4D seismic surveys over the life of a carbon capture and storage project (Warner et al, 2020). In addition to initial government funding such as that incentivized by the United States Inflation Reduction Act, operators are now beginning to acquire and budget funding to develop CCUS projects at much larger scale. The ongoing viability of these projects requires long-term profitability including sustainable data management and storage strategies.

Geophysicists who are already familiar with data management systems for large volumes of geophysical data will also find efficiencies from the embedded data workflow methodology described. In part of our evaluation, we made geophysical sensing data available on a web-based map-indexed platform with standardized metadata fields available for search and filter. User communities surveyed reported that on average, time to find and access the data after it was loaded to a standards-based data management platform was reduced from weeks to minutes compared with previous fragmented storage limited by geographic or functional siloes. Applying average reported fully loaded costs for geotechnical end users and reported number of data operations per quarter demonstrates that the investments in embedded data workflows had a positive return on investment in as little as 5 months.

Reduction in cycle time also applied to workflow development work. In extending the data footprint of an existing data management platform to support new data types in support of low carbon energy projects, it was found that more than half of the new data entities were well represented by schemas and components already in a subsurface open source code base, and less than 30 required customized development of an industry standard REST API endpoint. This represented a data interoperability latency reduction of 43% of resource allocation, which translated into reduced cost for the client.

Conclusions

The use of an embedded end-to-end digital data workflow for CCUS data streams has the potential to make the data more findable, accessible, interoperable, and reusable for end users in support of critical business decisions during a carbon injection and storage project lifecycle. Leveraging existing industry standards and accepted optimum practices for data management means that business value can be recognized more quickly and less time is spent on development of custom workflows or data handling operations. As more operators move to low-carbon projects and as regulations for measurement, monitoring and verification (MMV) of carbon storage reach a higher level of data management maturity, we expect these type of workflows to be more widely adopted and standardized by the resource industry.

Acknowledgements

The author wishes to acknowledge the consulting and operations teams at the Houston and Perth Katalyst Data Management offices and data centers for their contributions and assistance in preparing this material.

References

Al-Obaidan, A. H., and Scully. G. W. 1991. "Efficiency differences between private and State-Owned enterprises in the international petroleum industry." *Applied Economics*, Volume 23 237–246.

Azzola1, J., Thiemann, K., and Gaucher, E. 2022. Towards Real-Time Seismic Monitoring of a Geothermal Plant Using Distributed Acoustic Sensing. https://doi.org/10.5194/egusphere-2022-1417

BP Exploration Operating Company Limited. 2021. Measurement Monitoring and Verification (MMV) Plan for Endurance, NS051-SS-REP-000-00018. Department of Business, Energy and Industrial Strategy. https://assets.publishing.service.gov.uk/media/629527708fa8f5038dcd288c/NS051-SS-REP-000-00018-MMV_Plan.pdf UK Crown Copyright.

Bates, M., Cosby, T. and Perrons, R.K. 2022. Trust but Verify: How the Open Footprint Forum Will Help the Oil and Gas Industry Reduce Fugitive Emissions. *J Pet Technol,* Data Science and Digital Engineering, May 27, 2022. ISSN: 1944-978X (Online) https://jpt.spe.org/trust-but-verify-how-the-open-footprint-forum-will-help-the-oil-and-gas-industry-reduce-fugitive-emissions

Bakka, H.S., Dümmong, S., Haavik, K.E., et al. 2022. Paper presented at the SPE Norway Subsurface Conference, Bergen, Norway, April 2022. Real-Time Fiber Optics: A Door Opener for the Wellbore Environment. Paper Number: SPE-209540-MS, https://doi.org/10.2118/209540-MS

Barrow, B. 1997. A Lesson in Megabytes. *IEEE Standards Bearer*, January 1997, (5). https://www.thierry-lequeu.fr/data/PELS/Comm/Publications/Newsletter/9704/STORY18.HTML

Correa, J. 2020. Fibre-optic based DAS technology for Long-term Seismic Monitoring of Carbon Capture and Storage Projects. U.S. Department of Energy, Office of Scientific and Technical Information, Presented at *EAGE Workshop on Fibre Optic Sensing for Energy Applications in Asia Pacific.*

Dean, M. and O'Brien, S. 2024. Best Practices for Risk-Based Measurement, Monitoring, and Verification in CCUS Projects. The Way Head, Society of Petroleum Engineers, January 2, 2024. Risk Management. https://jpt.spe.org/twa/best-practices-for-risk-based-measurement-monitoring-and-verification-in-ccus-projects

Earley, S., and Henderson, D., Sebastian-Coleman, L (Eds.) 2017. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DM BOK). Bradley Beach, NJ: Technics Publications, LLC.

Gabriel, C. and Wheelwright, J. 2023. Collaboration in Carbon Counting: OSDU® CCUS and Open Footprint™. Presented at: *The Open Group Summit, Delivering a Secure and Sustainable Tomorrow*, Houston, USA, 30-Oct to 2-Nov 2023. https://events.opengroup.org/website/59050/home/#agenda

Gimenez. L. and Kozman, J. 2004. Maturity Models for E&P Data and Information Management Organizations. https://doi.org/10.2118/88666-MS.

Gregory, A., and Hodson, S. 2022. WorldFAIR Project (D2.1) 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' (1.0). Zenodo. https://doi.org/10.5281/zenodo.7378109

He, R. and Paulsson, B.N.P. 2021. Development of Big-Data Edge-Computing Analytic Framework for Distributed Fiber Optic Sensors". United States Department of Energy, Office of Scientific and Technical Information. OSTI Identifier 1901727, Report Number DE-SC0020467, DOE Contract Number SC0020467. https://www.osti.gov/biblio/1901727

Henson, A. 2021. Four steps to extract the full potential of a novel data platform. *World Oil*, November 2021. https://www.worldoil.com/magazine/2021/november-2021/features/four-steps-to-extract-the-full-potential-of-a-novel-data-platform/

Isaenkov, R., Pevzner, R., Glubokovskikh, S. et al. 2021. An Automated System for Continuous Monitoring of CO2 Geosequestration Using Multi-Well Offset VSP With Permanent Seismic Sources and Receivers: Stage 3 of the CO2CRC Otway Project. Lawrence Berkeley National Laboratory, Publications Preprint. https://doi.org/10.1016/j.ijggc.2021.103317

Isaenkov, R. et al. 2023. Long-term continuous DAS VSP monitoring for CCS: advanced repeatability analysis. *Curtin University Reservoir Geophysics Consortium,* 2023 Annual Meeting. https://ceg.curtin.edu.au/curtin-reservoir-geophysics-consortium/consortium-resources/crgc-agm-publications/2023-crgc-annual-meeting/

Jacobs, T. 2023. Silixa Chief Expects CCS Boom To Create 'Huge Opportunity' for Fiber Optics. *J Pet Technol*, Carbon capture and storage, September 1, 2022. https://jpt.spe.org/silixa-chief-expects-ccs-boom-to-create-huge-opportunity-for-fiber-optics

Kobayashi Y, Takayasu H, Havlin S, Takayasu M. 2021. Robust Characterization of Multidimensional Scaling Relations between Size Measures for Business Firms. *Entropy* (Basel). Jan 29;23(2):168. doi: https://doi.org/10.3390/e23020168. PMID: 33573072; PMCID: PMC7910913.

Kozman, 2003. Managing the three Ps of seismic data: proliferation, pervasiveness, and persistence. ASEG Extended Abstracts 2003(2):1. https://doi.org/10.1071/ASEG2003ab090

Kozman, J. 2014. Optimising Technology Adoption for E&P Strategies. *Proc. Indon. Petrol. Assoc., 38th Ann. Conv., 2014.* https://www.ipa.or.id/en/publications/optimising-technology-adoption-for-e-p-strategies

Kozman, J.B. 2016. Data Custody Transfer Points. Presented at the Professional Petroleum Data Management Perth Data Management Symposium. https://dl.ppdm.org/dl/1760

Kozman, J. 2023. Balancing Data Democratization and Entitlements - Technology vs Human Behavior. Presented at the Society for Professional Data Management, End-Year Event 2023, *Energising Data Management – Disruptive or Newly Native?* https://www.societypdm.org/events/spdm-end-year-event

Li K, Griffin MA, Barker T, et al. 2023. Embedding data science innovations in organizations: a new workflow approach. *Data-Centric Engineering*. 2023;4:e26. https://doi.org/10.1017/dce.223.22

Mejía, J. and Aliakbari, E. 2023. Annual Survey of Mining Companies, 2022, Published on May 4, 2023, the Fraser Institute, Calgary, Alberta, Canada. https://www.fraserinstitute.org/studies/annual-survey-of-mining-companies-2022

Miller, D.E. et al. 2016. Simultaneous Acquisition of Distributed Acoustic Sensing VSP with Multi-mode and Single-mode Fiber-optic Cables and 3C-Geophones at the Aquistore CO2 Storage Site. *The Recorder,* official publication of the Canadian Society of Exploration Geophysics. **41** (06). https://csegrecorder.com/articles/view/simultaneous-acquisition-of-distributed-acoustic-sensing-vsp-with-multi-mod

Minimising the environmental impact of CO2 storage monitoring with fibre-optics. 2021. *Carbon Capture Journal*, July - Aug 2020. https://silixa.com/wp-content/uploads/Minimising_the_environmental_impact_Silixa.pdf

Mohammadpoor, M. and Torabi, F. 2020. Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*, **6,** (4): 321-328. https://doi.org/10.1016/j.petlm.2018.11.001.

Reinert, A. 20923, How To Attain "The Prize" with Big Data Management. *J Pet Technol,* Data Science and Digital Engineering, March 29, 2023 by PPDM. https://jpt.spe.org/how-to-attain-the-prize-with-big-data-management

Sadowski, J. 2019. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). https://doi.org/10.1177/2053951718820549

Staněk, F., Jin, G., and Simmons, J. 2022. Fracture Imaging Using DAS-Recorded Microseismic Events. *Front. Earth Sci.*, 19 July 2022, Sec. Solid Earth Geophysics, Volume 10 – 2022. https://doi.org/10.3389/feart.2022.907749

Tordo, S., Tracy, B.S., and Arfaa, N. 2011. National oil companies and value creation. World Bank working paper ; no. 218, eISBN 978-0-8213-8832-7, HD9560.5.T647 2011, 338.2'7282—dc23, ISSN: 1726-5878, https://doi.org/10.1596/978-0-8213-8831-0

Warner, T., Rampton, D., Morgan, D. et al. 2020. Using Geophysical Technologies Deployed in Inexpensive Monitoring Wells to Monitor the Evolution of a CO2 Plume: Potential Benefits and R&D Needs. National Energy Technology Laboratory, Pittsburgh, PA, United States, October 1, 2020. https://www.osti.gov/biblio/1677509

Zhan, G., van Gestel, J.P., and Johnston, R. 2020. DAS data recorded by a subsea umbilical cable at Atlantis field. *SEG Technical Program Expanded Abstracts 2020. https://doi.org/10.1190/segam2020-3427669.1*