PROCEEDINGS, 44th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2019
SGP-TR-214

Getting Data Out of the Ground: Modern Challenges Facing EGS Collab, the DOE Geothermal Data Repository, and the Geothermal Industry

Jon Weers, Jay Huggins

National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401-3305

Keywords: EGS, Collab, data, collaboration, challenge, geothermal, information, management, security, AWS, cloud, future, NREL, DOE, OpenEI, SURF, GDR, transfer, storage, big, mine, real-time, decision, making, utilization, ES Net, Globus, Snowball, sensors

ABSTRACT

The amount of data being generated by the geothermal industry is increasing exponentially. Advanced tools, higher resolution sensors, and digitally integrated equipment are providing valuable insight into geothermal exploration and operational efficiencies, but the volumes of data they generate and the unique properties of Enhanced Geothermal System (EGS) projects are creating challenges of their own. This paper explores those challenges and the steps the U.S. Department of Energy (DOE) is taking to address them through comprehensive data management strategies, improvements to DOE's Geothermal Data Repository (GDR), and the development of a secure data collaboration tool for DOE's Frontier Observatory for Research in Geothermal Energy (FORGE) and EGS Collab projects.

1. GEOTHERMAL DATA CHALLENGES

Data-driven innovation is propelling the petroleum industry to new heights. The collection and analysis of big data from seismic arrays is giving companies like Shell higher resolution and more accurate pictures of target reservoirs (Marr 2015) and helping others gain efficiency and operational insight (Venables ,2018). Advances in instrumentation including smaller, more affordable sensors, Distributed Temperature Sensing (DTS) and Distributed Stress Sensing (DSS) fiber optic cables, and networks of interconnected devices are making the collection of downhole data easier and more prolific. For example, the analysis of Rayleigh backscattering of Ultra Violet (UV) light in DTS cables and DSS cables is increasing their resolution by several orders of magnitude, allowing, in some cases, the measurement of temperature and stress at 1cm intervals (Loranger et al. 2015). Leveraging these advances to gain valuable insight into exploration, operational efficiencies, and EGS stimulation projects, the geothermal industry is generating more data than ever before (Weers and Anderson 2016). Timely access to these data can inform exploration, minimize risk during drilling, enable teams to make smarter decisions, and increase the efficiency of geothermal operations. However, the volume of data being generated can present its own series of challenges. Data that are too large or too complex to store, transport, or utilize via conventional means (i.e. "big data") can be difficult to process effectively. Additionally, data from geothermal exploration and drilling projects can be more difficult to acquire than their petroleum counterparts.

1.1 Acquisition

The typical lithology of geothermal resources can make it difficult to acquire data in the first place. This is especially prevalent in the development of deep geothermal and EGS reservoirs where ideal conditions often require drilling to deeper target depths, in harder rock where harsh vibrations can rattle equipment, and in hotter temperatures (above 250ºC) that exceed the tolerances for which many instruments have been designed. The geothermal industry is in need of reliable downhole instruments that can survive the harsh conditions of geothermal drilling (Pennewitz et al, 2012).

Also difficult to acquire are existing data from previous projects that have been lost to a desk drawer, stored in a company vault, bound by proprietary agreement, or are otherwise unavailable to the public. Lack of access to these data presents a serious challenge to the geothermal industry as a whole. Unable to learn from others' mistakes or build upon their successes, potential users of such data are at risk of duplicating efforts or missing opportunities for collaboration and innovation. The avoidance of this issue is one of the primary drivers behind the GDR, which houses data from all DOE Geothermal Technologies Office (GTO) funded projects and research. While some of the data on the GDR are subject to a moratorium for a predetermined period of time to protect proprietary interests, all data submitted to the GDR is eventually made available to the public. More importantly, the metadata for these data are all available to the public immediately, including a general description of the data and a primary contact. Not only does this inform the geothermal community that these data exist, it also allows anyone interested in these data to contact the data owners and attempt to obtain an advanced copy or even team up to collaborate on ongoing, related work.

1.2 Storage

The storage of large datasets presents several challenges to the geothermal industry; the most obvious being hard drive storage capacity. With recent advancements in instrumentation including higher sample rates and increased resolution, it's not uncommon for a geothermal stimulation experiment to generate more than 1TB of data per day. Simply storing these data as they are generated can be problematic.

Less common but equally challenging, certain file systems can be subject to a maximum number of files and folders, called *inode* limits. In many cases, these limits are configurable (e.g. maximum number of folders allowed in an *ext4* filesystem) but require the configuration to be made at the time the filesystem is created. Running into an *inode* limit in the field can lead to unanticipated downtime or worse, unnoticed data loss.

At scale, the storage of big data starts to present unusual maintenance problems. For example, one petabyte of data, if stored on an array of 10TB hard drives, can require a substantial amount of maintenance. One petabyte (1PB) would occupy roughly 108 10TB drives operating at 95% storage efficiency (reserving 5% for index tables and other filesystem components). While some hard drive manufacturers are better than others, the average failure rate of hard drives over 1TB is 1.27% (Klein 2018), indicating that at least one drive, or 10TB of data could be lost each year. When operating at these scales, even the most reliable of storage solutions can require additional redundancy to prevent data loss.

To avoid these issues, the GDR currently stores data on Amazon's cloud-based Elastic File System (EFS), a redundant file-system that is not subject to *inode* limits and is theoretically infinitely expandable. In reality, the filesystem is limited by the total storage capacity of the Amazon cloud, but in practice, the drives storing GDR data automatically scale in proportion to the data stored within, allowing for continual support without risk of reaching capacity or the need to swap out drives to accommodate additional data. This service costs more than standard cloud storage but is competitive when compared with the cost of maintaining, expanding, and troubleshooting traditional file systems.

Effective cloud storage requires high-speed connections to the cloud, which may not be possible in the field. This can require creative storage solutions both out in the field and back at the lab, or wherever data analyses, visualization or modeling activities are taking place. For collaborative projects or larger organizations, these could be distributed among multiple partners creating a need not only for multiple storage solutions, but also for an efficient transport mechanism capable of moving large amounts of data.

1.3 Transfer

Moving large amounts of data presents its own set of challenges, especially in the field. High speed connections, like Google Fiber or ESnet (a high-speed alternative to the internet for DOE research institutions) are wonderful when they are available, but are still limited by the number of people with access to those connections. Every transfer involves at least a sender and a receiver, and unless both parties have an uninterrupted high speed connection, there is a risk that the transfer will have to cross through a slower part of the internet. Compounding the issue, the firewalls adopted by many institutions to prevent cyber-attacks also prevent the transfer of large volumes of data. Some firewalls have set limits on the size of files and will terminate any connection attempting to move more data than allowed. Others inspect the packets of transmitted information for nefarious payloads, a process that while prudent, can slow down the transfer rate of complex data significantly.

Peer-to-peer transfer protocols employed by tools like Globus can circumvent some of these issues and help increase transfer speeds by sending data over multiple ports simultaneously. However, the very nature of peer-to-peer tools often limits the data transfer to a pair of individuals who must coordinate the data transfer carefully to ensure the appropriate software is installed and to initiate the transfer at both ends. These types of solutions do not efficiently support group collaboration and can make it difficult to disseminate a large dataset to a team of collaborators.

For projects generating more than 1TB of data each day, the risk of slow transfer speeds or the need to transfer to multiple locations can create a bottleneck of information. A system generating 1TB of data per day generates 728.18MB of data each second. With typical high speed internet service topping out at around 100Mbits per second (or about 12.5MB/sec) per second, regular transfer methods won't be able to move data as fast as it is being generated. Google Fiber, whose upper limit is around 1,000Mbits per second (125MB/s) would take 5.8 days to transfer the data generated in a single day, assuming a perfect connection and gigabit speeds on both ends.

A requirement to push data to multiple locations exacerbates this problem considerably. In some cases, physically transporting drives can still be the most efficient means of moving data. This is the logic behind Amazon's Snowball, a service in which, for a fee, Amazon will send an enormous hard drive shielded in a protective shipping case to allow for data to be copied over and shipped to a facility for loading onto the cloud directly. Solutions like these are still subject to the time it takes to copy data from one drive to another.

1.4 Utilization

Finally, the utilization of big data can be challenging. Assuming the data are accessible, many geothermal projects generate tens of thousands of files over the lifetime of the project. Daily logs, core photos, and streaming data from sensors, typically split into daily or even hourly files, can amass to an unwieldly number of data resources. Organizing these data into meaningful and accessible structures can be a formidable barrier to access.

The utility of big data can be improved through strategic file formats, like the Hierarchical Data Format (HDF), which is designed to store and organize large amounts of data for the purposes of supporting massive data queries, parallel computation, and high performance computing. However, data from instruments is rarely available in these formats and must often be translated before it can be effectively used.

Centralized databases built around accessibility (i.e. "data lakes") can be utilized to provide teams of people with access to data that would otherwise be prohibitive to transfer or access using conventional means. Where datasets are simply too large to download efficiently or too complex to utilize without a high performance computer, users of cloud-based data lakes can send queries to the data and have answers returned via API, or park a cluster of cloud-based compute nodes near the data to perform analysis (Figure 1).

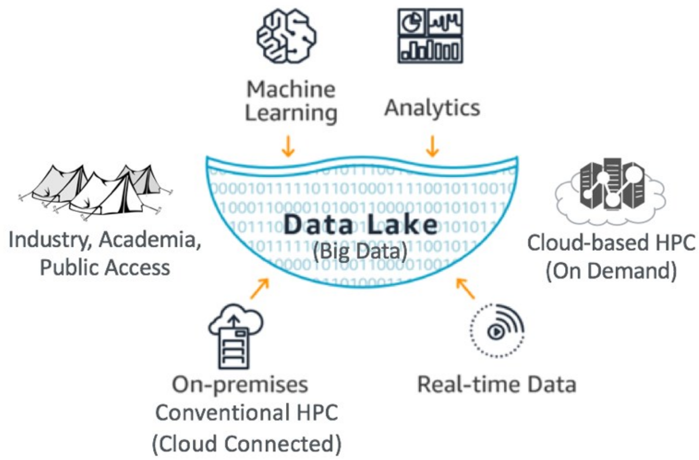

Figure 1: A data lake, or central data repository, accessible by people and tools through APIs, represented by orange arrows.

Especially advantageous for large data sets, data lakes reduce duplication of storage by providing a single access point. Without transfer costs, collaborators can spend more time using the data and generating results. However, the data lake must be established in a way that supports the many needs of the potential users. Developing the appropriate Application Programming Interfaces (APIs) can be time consuming. Unless established in advance and fed by a live data stream, data lakes are typically not suitable for real time or near-real time decision making.

The utility of a dataset can be increased through the use of standardized naming conventions and project-specific organizational strategies, which can help users find things more quickly, but are no guarantee that critical data will be found. A data management system with a robust, faceted search can assist with data discovery and organization by allowing data owners to tag key datasets with specific terms and provide data seekers with a faceted search or other means of finding them.

The EGS Collab project faced many of these challenges and more, due in part to its location nearly a mile underground.

2. UNIQUE CHALLENGES FOR EGS COLLAB

Initiated by the DOE GTO to demonstrate the potential for EGS and to function as a stepping stone from lab experiments to FORGE, the EGS Collab project is undertaking a series of stimulation experiments at the intermediate scale (~10 meter) to improve understanding of stimulation and response in an appropriate rock mass and to validate thermal-hydrological-mechanical-chemical (THMC) models. In an effort to seek a test bed congruent with the lithology of potential EGS sites without incurring substantial drilling costs, the EGS Collab experiments are being performed approximately 1 mile underground at the Sanford Underground Research Facility (SURF) (Figure *X*), in Lead, South Dakota (Heise, 2015).

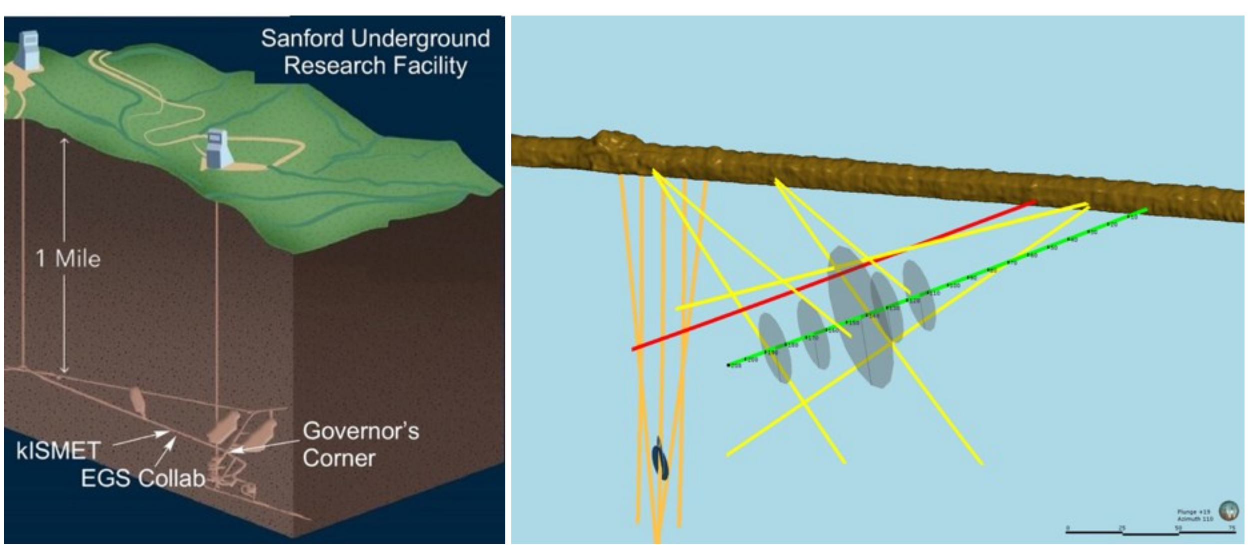

Figure 2: (left) Diagram of the Sanford Underground Research Facility (SURF), illustrating access shafts and the 4850 level, home to Experiment 1 of the EGS Collab project. (right) Depiction of boreholes and fracture planes for Experiment 1. The West Drift of the 4850 level appears as the thick brown line. The gray disks represent fracture planes, and the colored lines

represent wells used in the experiment. The production well is shown in red, the stimulation (injection) well is shown in green, and the monitoring wells are shown in yellow.

Formerly a gold mine, the SURF provides access to a deep crystalline rock mass via a network of existing, underground tunnels that allow for the cost-effective drilling of boreholes and monitoring wells. Infrastructure within the tunnels include ventilation, water, and power, which allow EGS Collab team members to monitor experiments onsite. While the temperature (max ~35ºC) and stress levels are not ideal for geothermal projects, the location within SURF allows for testing under realistic in-situ stress conditions without the need for costly deep drilling, and provides the EGS Collab team with the ability to perform detailed characterization and monitoring of the experiments directly from the drift (Kneafsey et al, 2019).

However, the location is not without its challenges. Limited space within the drift requires careful management of sensing equipment, including the rationing of power and bandwidth (Figure 3). Digital communication to the outside world must be passed through a single, shared internet connection, a standard Cat-5 ethernet cable limited to around 100Mbits per second.

Figure 3: A monitoring station located on the 4850 level of the SURF near Experiment 1.

The EGS Collab project is employing a variety of monitoring systems, including passing seismic monitoring and distributed fiber optic sensors to monitor temperature (DTS), seismicity via acoustic sensing (DAS), and strain (DSS) (Daley et al., 2013), each of which are recording in high resolution, and combined are generating nearly 2TB of data per day (Kneafsey et al, 2019). Even a high speed internet connection over a fiber optic cable would be insufficient to transport data being generated at these rates.

2.1 Getting Data Out of the Mine

Monitoring stations, like the one depicted in Figure 3, allow team members within the drift to keep an eye on critical data during experiments. However, space in the drift is too limited to accommodate all of the team members necessary to effectively manage these experiments, so some data must be transmitted out of the mine. Limited by a shared internet connection with a maximum bandwidth of 100Mbits per second, the EGS Collab project has to be strategic about selecting which data are broadcast out. Priority is given to data needed to understand how the experiment is unfolding and to support decision-making in near-real time during critical stages of the experiment. Many of the data necessary to make these decisions aren't inherently useful in their raw form (e.g. raw DTS data) and must be translated or visualized in order to convey information effectively. These data are pre-processed on computers located in the mine before being transmitted to remote team members.

2.2 Edge Processing

Preprocessing on premises before connecting to the internet is considered "edge processing" and is becoming more prevalent in data-rich industries. This is nothing new for the automotive industry, especially with the advances being made in autonomous vehicles and newer cars with intelligent collision avoidance systems. As automobiles become more laden with sensors they are generating more and more data. A Tesla Model S, for example, generates more than 4TB of data every day; information used by the vehicle in real time for adaptive cruise control, collision avoidance, and energy efficiency (Lambert, 2016). Only a small fraction of the data generated by each car is recorded. The rest is processed in real time and discarded.

The oil and gas industry is leveraging Artificial Intelligence (AI) to shift the processing of large amounts of data from the cloud to the edge. Where volumes of data would be shipped to a central repository and processed by high performance computers, field-tested AI algorithms are now being applied to data streams in the field allowing for faster processing of data into usable information and empowering employees in the field to make critical, data-driven decisions in near-real time. Furthermore, many oil and gas companies are applying machine learning algorithms to instruments in the field to automate the decision-making process altogether. This level of edge processing changes the role of equipment in the field from providing data to determining outcomes (Venables, 2018). This level of automation requires a significant amount of trust in the developed algorithms, which must be field tested rigorously and compared against the original source datasets before widespread deployment.

Automation in the EGS Collab project was limited to the development of data plots to visualize the raw data streams being generated and transmit them to remote team members. Screens like the one depicted in Figure 3 were broadcast to the team using traditional web-based screen sharing software, which was easily accommodated by the limited bandwidth in the mine, and recorded externally for later review.

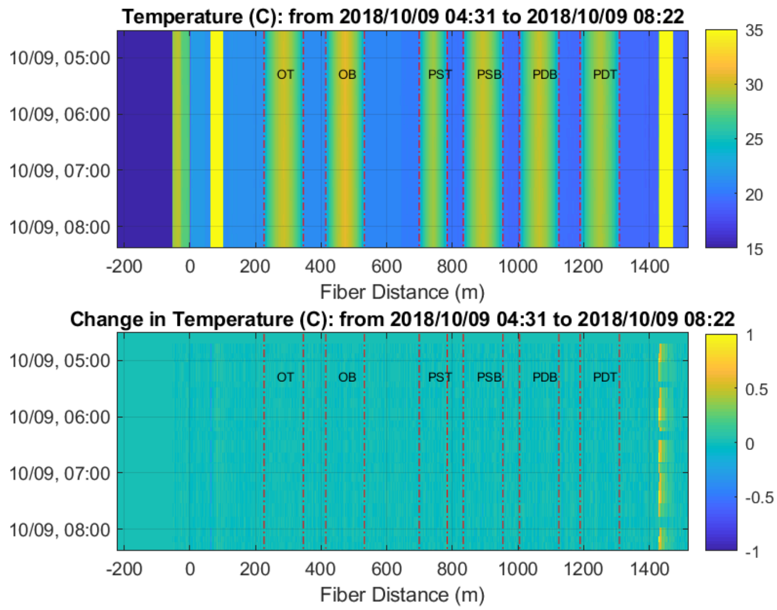

Figure 4: Plot of DTS data across multiple wells over time.

Plots like the one shown in Figure 4 above were generated in the mine and uploaded at regular intervals to a collaborative data management system (DMS) on OpenEI (Open Energy Information, https://openei.org), where they are programmatically sorted and indexed so the entire EGS Collab team can access them securely.

By processing the data in this way, critical information can be conveyed to the team in a near-real time, efficient manner. Images, after all, can contain a thousand data points or more. However, visualizing data as images introduces certain limitations. Visualizations are excellent for quickly identifying trends but cumbersome for determining the precision of a single data point. Unlike raw data, images are not machine readable, making them more difficult to import into additional analyses or reuse in another format. Lastly, images are a terminal format. Once data have been visualized in image form, further data compression or aggregation is severely limited.

Edge processing can be a valuable tool for managing large amounts of data in the field, but picking the right information to convey can be tricky, and selective storage of the raw data may be necessary to validate the algorithms processing the data to ensure that they are generating actionable information.

3. FROM DATA TO INFORMATION

In their most basic form, data are recorded observations. Transforming data into usable information requires, at the very least, the attribution of any context necessary to understand the data. In the data science world, this context often takes the form of metadata. Axes and labels present in the visualizations generated in Figure 4, combined with a general background knowledge of the project, provide the project team with the context needed to interpret the visualization and extract meaningful information. For those not familiar with the project, additional context may be needed to interpret the data correctly. This need drives the metadata requirements for data submitted to the GDR and is included in the secure DMS used by the EGS Collab project (Figure 5).

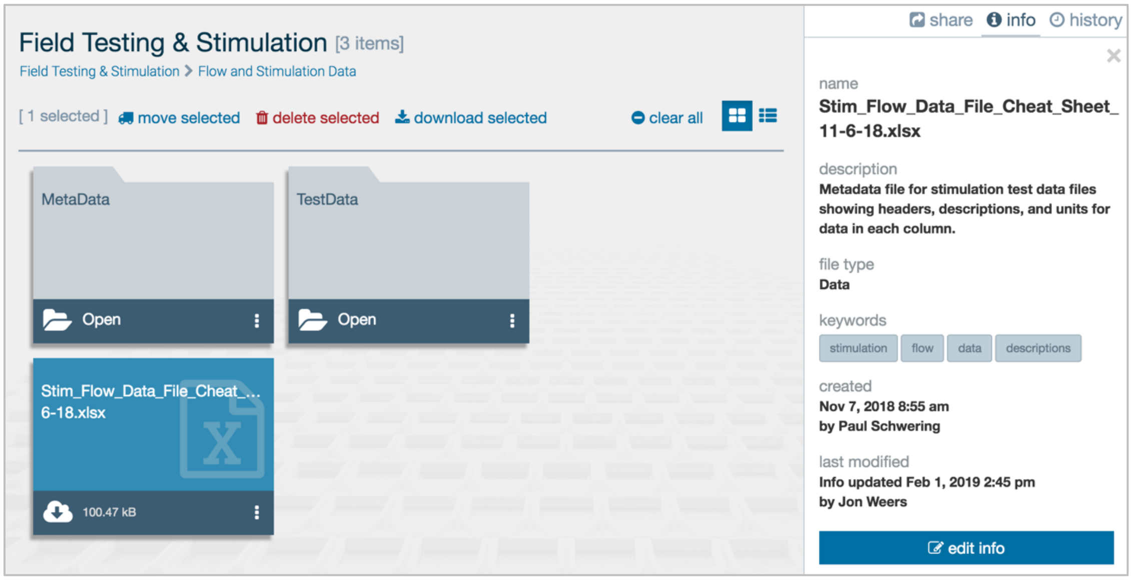

Figure 5: Screenshot of the EGS Collab DMS showing metadata for the selected item in the "info" panel (right) (OpenEI, 2019).

An optional information panel allows EGS Collab researchers to attribute additional notes, metadata, or other information to uploaded data files. When ready for public release, any number of individual files or folders can be selected within the DMS and sent to the GDR as a data submission. During this process, any metadata attributed in the DMS will automatically populate the appropriate metadata fields in the corresponding GDR submission, which will open in a new browser window allowing the researcher to complete the submission at their convenience using the existing GDR interface.

3.1 Information Dissemination

The connection developed between the EGS Collab DMS and the GDR will utilize the existing GDR submission workflow to publish EGS Collab data to the greater scientific community. Once publicly available, data from the EGS Collab project will be disseminated to all of the existing GDR data partners, including Data.gov, the Office of Science and Technical Information's (OSTI) DOE Data Explorer, and Thompson Reuters' Data Citation Index (Weers et al, 2018) (Figure 6).

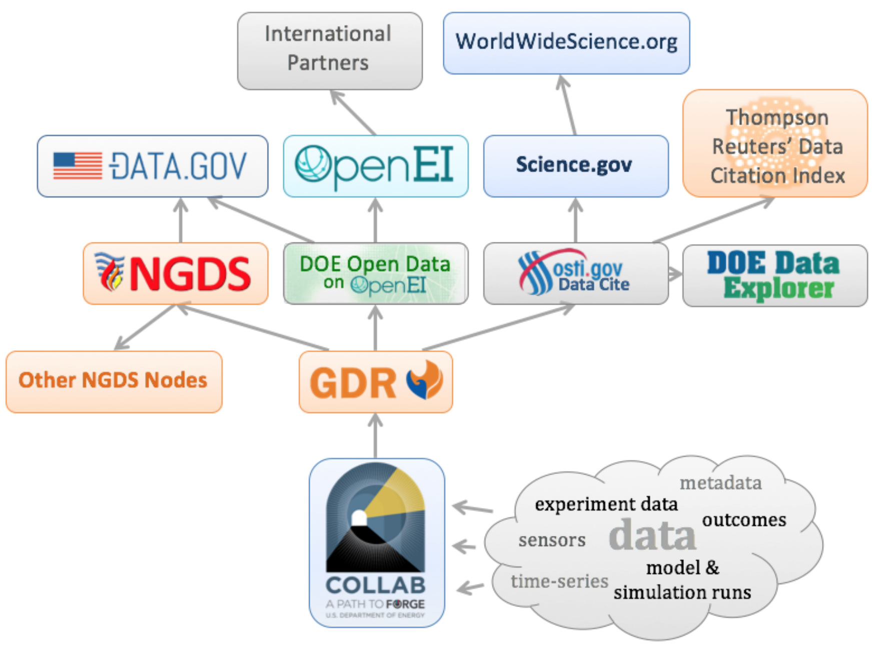

Figure 6: Propagation of EGS Collab data through the network of GDR data partners.

4. FUTURE OF GEOTHERMAL DATA TECHNOLOGY

The amount of geothermal data being collected through research and development activities will likely continue to grow at substantial rates, eventually surpassing the storage capacity of conventional desktop and laptop computers. The data generated by the EGS Collab project alone is already exceeding the storage capacity of most computers. Barring considerable advances in storage technology, the processing and analysis of geothermal research data will likely transition to data lakes and other cloud-based data platforms such as the collaborative DMS on OpenEI.

As artificial intelligence matures, the processing of data essential to near-real time decision making could shift from the cloud toward the edge; where intelligent instruments might selectively report summaries of salient information rather than simply stream raw data. To bridge this gap, however, trust must first be established. Gaining confidence in the application of machine learning for real time decisions making will likely require the storage and post-mortem analysis of raw data to review decisions being made autonomously and evaluate the effectiveness of algorithms operating in the field. The collection, storage, and dissemination any raw data collected now will be essential to the development of future automation. Ultimately, the challenge facing both human and machine will be knowing which data to keep and which not to.

Preserving large amounts of data for future research can be costly. Advances in storage technology are already driving the cost of conventional storage down and allowing more data to be stored for cheaper. Additionally, concepts such as "cold storage" represent new business models in the cloud storage market and can provide cheaper alternatives to conventional long term storage. Legacy architectures for data storage typically presume the data must be immediately available. In some use cases, it may be acceptable to wait for a period of time between the request to access the data and the delivery of those data. In these cases cold storage becomes a viable option and can significantly reduce the cost of storage by several orders of magnitude. Most cloud storage providers offer variants of cold storage with latency ranging from seconds to days, but all offer high reliability and continue to innovate new storage.

As an alternative to full downloads, transfers of selected data from data lakes can reduce transfer time and cost. As data lake utilization options evolve, cold storage and data lakes could work together to make it possible to flexibly process data at scales that are currently unfeasible.

To support research and innovation in the geothermal community, all data generated from the EGS Collab project will eventually be made available to the public via the GDR. The EGS Collab team is working closely with the OpenEI and GDR teams to develop data management solutions that not only meet the needs of the EGS Collab project today, but which accommodate the needs of future geothermal data research innovation.

ACKNOWLEDGEMENTS

This research was supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Geothermal Technologies Office (GTO) under Award Number DE-AC52-07NA27344 with LLNL, Award Number DE-AC05-76RL01830 with PNNL, Award Number DE-AC02-05CH11231 with LBNL, and Contract Number DE-AC36-08-GO28308 with the National Renewable Energy Laboratory, as part of the EGS Collab project. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. Some of the research supporting this work took place at the Sanford Underground Research Facility (SURF) in Lead, South Dakota. The assistance of the SURF and its personnel in providing physical access and general logistical and technical support is acknowledged.

REFERENCES

Burwell, S., VanRoekel, S., Park, T. and Mancini, D.: Memorandum For The Heads of Executive Departments and Agencies, M-13-13 "Open Data Policy – Managing Information as an Asset." Director Executive Office of the President, Office of Management and Budget (2013).

Daley, T., Freifeld, B., Ajo-Franklin, J., Dou, S., Pevzner, R., Shulakova, V., Kashikar, S., Miller, D., Goetz, J., Henninges, J., and Lueth, S.: Field testing of fiber-optic distributed acoustic sensing (DAS) for subsurface seismic monitoring, The Leading Edge, 32(6), 699-706, doi:10.1190/tle32060699.1.

Dobson, P., Kneafsey, T.J., Blankenship, D., Valladao, C., Morris, J., Knox, H., Schwering, P., White, M., Doe, T., Roggenthen, W., Mattson, E., Podgorney, R., Johnson, T., Ajo-Franklin, J., and E.C. Team (2017): An Introduction to the EGS Collab Project. *GRC Transactions Vol 41*, 41st Geothermal Resources Council Annual Meeting, Salk Lake City, UT (2017).

GDR: "DOE Geothermal Data Repository." OpenEI: Open Energy Information. National Renewable Energy Laboratory, 15 Jan. 2018. Web. https://gdr.openei.org.

Heise, J. (2015), The Sanford Underground Research Facility at Homestak, *Journal of Physics: Conference Series*, *606*(1), 26

Hilbert, M., and López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025), 60 –65. doi:10.1126/science.1200970.

IBM: "What is big data?" IBM: Bringing big data to the enterprise. IBM. 12 Feb. 2012 Web. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

Klein, A.: "Hard Drive Stats for Q3 2018: Less is More." Blog. Backblaze. 16 Oct 2018. Web. https://www.backblaze.com/blog/2018-hard-drive-failure-rates/.

Kneafsey, T.J., Blankenship, D., Hunter, K., Johnson, T., Ajo-Franklin, J., Schwering, P., Dobson, P., Morris, J., White, Fu, P., Podgorney, R., Roggenthen, W., Doe, T., Mattson, E., Ghassemi, A., Valladao, C., and E. C. Team(2019): EGS Collab Project: Status and Progress, *Proceedings, 44th Workshop o Geothermal Reservoir Engineering*, Stanford University, Stanford, CA (2019).

Knox, H., P. Fu, J. Morris, Y. Guglielmi, V. Vermeul, J. Ajo-Franklin, C. Strickland, T. Johnson, P. Cook, C. Herrick, M. Lee, and E.C. Team (2017), Fracture and Flow Designs for the Collab/Sigma-V Project in *GRC Transactions, Vol. 41,* 41st Geothermal Resources Council Annual Meeting, Salk Lake City, UT (2017).

Lambert, F.: "Tesla has now 1.3 billion mile of Autopilot data going into its new self-driving program." Electrek. Electrek, Co, 13 Nov. 2016. Web. https://electrek.co/2016/11/13/tesla-autopilot-billion-miles-data-self-driving-program/.

Loranger, Sébastien & Gagné, Mathieu & Lambin Iezzi, Victor & Kashyap, Raman. (2015). Rayleigh scatter based order of magnitude increase in distributed temperature and strain sensing by simple UV exposure of optical fibre. Sci. Rep.. 5. 10.1038/srep11177.

Marr, Bernard.: "Big Data In Big Oil: How Shell Uses Analytics To Drive Business Success." Forbes Tech. Forbes, Inc. 26 May 2015. Web. http://www.forbes.com/sites/bernardmarr/2015/05/26/big-data-in-big-oil-how-shell-uses-analytics-to-drive-business-success.

OpenEI: "Collab DMS: Secure Data Management System." OpenEI: Open Energy Information. National Renewable Energy Laboratory, 1 Feb. 2019. Web. https://collab.openei.org/.

Pennewitz, E., Schilling, M., Kruspe, T., Jung, S., and Ruehs, A.: "Active cooling of downhole instrumentation for drilling in deep geothermal reservoirs," *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Graz, 2012, pp. 600-603. doi: 10.1109/I2MTC.2012.6229454

Weers, J. and Anderson, A.: The DOE Geothermal Data Repository and the Future of Geothermal Data, *Proceedings, 41st Workshop o Geothermal Reservoir Engineering*, Stanford University, Stanford, CA (2016).

Weers, J., Anderson, A., and Taverna, N.: The Geothermal Data Repository: Five Years of Open Geothermal Data, Benefits to the Community, *GRC Transactions Vol 41*, 41st Geothermal Resources Council Annual Meeting, Salk Lake City, UT (2017).

Venables, M.: "Maturation Of AI Offers Compelling Prospects To Oil And Gas Companies." Forbes Energy. Forbes, Inc. 20 Dec. 2018. Web. https://www.forbes.com/sites/markvenables/2018/12/20/maturation-of-ai-offers-compelling-prospects-to-oil-and-gas-companies/.