SAND2021-7964C

GRC Transactions, Vol. 45, 2021

Efforts and Innovations to Promote Data Sharing and Data Accessibility in the EGS Collab Experiments

Pengcheng Fu[1], Jon Weers[2], Mark White[3], Hunter Knox[3], Paul Schwering[4], Joseph Morris[1], Doug Blankenship[4], Tim Kneafsey[5], and EGS Collab Team[a]

1. **Lawrence Livermore National Laboratory, Livermore, California, USA**
2. **National Renewable Energy Laboratory, Golden, Colorado, USA**
3. **Pacific Northwest National Laboratory, Richland, Washington, USA**
4. **Sandia National Laboratories, Albuquerque, New Mexico, USA**
5. **Lawrence Berkeley National Laboratory, Berkeley, California, USA**

fu4@llnl.gov

Keywords:

EGS, Collab, fracturing, circulation, experiment, model validation

ABSTRACT

Large-scale scientific research programs such as the EGS Collab experiments and the FORGE program play extremely important roles in advancing geothermal technologies. Such efforts involve a large number of researchers from multiple institutions, last multiple years, and generate large, complex datasets. The value of the research efforts is only realized when the datasets are used by a large community of researchers in the decades to come. A challenge is that due to the

[a] Members of the EGS Collab team include: J. Ajo-Franklin, T. Baumgartner, K. Beckers, D. Blankenship, A. Bonneville, L. Boyd, S. Brown, J.A. Burghardt, C. Chai, A. Chakravarty, T. Chen, Y. Chen, B. Chi, K. Condon, P.J. Cook, D. Crandall, P.F. Dobson, T. Doe, C.A. Doughty, D. Elsworth, J. Feldman, Z. Feng, A. Foris, L.P. Frash, Z. Frone, P. Fu, K. Gao, A. Ghassemi, Y. Guglielmi, B. Haimson, A. Hawkins, J. Heise, Chet Hopp, M. Horn, R.N. Horne, J. Horner, M. Hu, H. Huang, L. Huang, K.J. Im, M. Ingraham, E. Jafarov, R.S. Jayne, T.C. Johnson, S.E. Johnson, B. Johnston, S. Karra, K. Kim, D.K. King, T. Kneafsey, H. Knox, J. Knox, D. Kumar, K. Kutun, M. Lee, D. Li, J. Li, K. Li, Z. Li, M. Maceira, P. Mackey, N. Makedonska, C.J. Marone, E. Mattson, M.W. McClure, J. McLennan, T. McLing, C. Medler, R.J. Mellors, E. Metcalfe, J. Miskimins, J. Moore, C.E. Morency, J.P. Morris, T. Myers, S. Nakagawa, G. Neupane, G. Newman, A. Nieto, T. Paronish, R. Pawar, P. Petrov, B. Pietzyk, R. Podgorney, Y. Polsky, J. Pope, S. Porse, J.C. Primo, C. Reimers, B.Q. Roberts, M. Robertson, V. Rodriguez-Tribaldos, W. Roggenthen, J. Rutqvist, D. Rynders, M. Schoenball, P. Schwering, V. Sesetty, C.S. Sherman, A. Singh, M.M. Smith, H. Sone, E.L. Sonnenthal, F.A. Soom, D.P. Sprinkle, S. Sprinkle, C.E. Strickland, J. Su, D. Templeton, J.N. Thomle, C. Ulrich, N. Uzunlar, A. Vachaparampil, C.A. Valladao, W. Vandermeer, G. Vandine, D. Vardiman, V.R. Vermeul, J.L. Wagoner, H.F. Wang, J. Weers, N. Welch, J. White, M.D. White, P. Winterfeld, T. Wood, S. Workman, H. Wu, Y.S. Wu, E.C. Yildirim, Y. Zhang, Y.Q. Zhang, Q. Zhou, M.D. Zoback

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

complexity of the data, it could require a user to devote a serious effort into understanding the data before the data can be effectively utilized. In the EGS Collab experiments, the team has devoted remarkable efforts and developed many innovative solutions to make the data more accessible to broader team members and future users. This paper documents the experience gained and lessons learned by the EGS Collab team in disseminating the data in the most informative and inspiring forms to maximize the value of this precious dataset. "Accessibility" in the title does not only mean making the raw data available for download. Particularly, we want to emphasize the importance of organizing, annotating, and presenting the data in ways to make it easy to digest by prospective consumers of the data.

1. Introduction

The EGS (enhanced geothermal system) Collab project, sponsored by the United States Department of Energy (DOE), Geothermal Technologies Office (GTO), focuses on intermediate-scale (~10-20 m) EGS reservoir creation processes and related model validation at crystalline rock sites (Kneafsey et al., 2021). The Experiment 1 and Experiment 2 testbeds are located near the 4850 ft and 4100 ft level drifts, respectively, at SURF (Sanford Underground Research Facility), formerly the Homestake Gold Mine, in South Dakota, USA. An important objective of the experiments is to validate computer codes that simulate and predict the hydraulic stimulations of fractures in EGS reservoirs. This is achieved by (1) extensive modeling and prediction of hydraulic stimulations in the design phase of the experiments, (2) thorough monitoring of the hydraulic stimulations to delineate hydraulic fracture propagation and natural fracture stimulations, and (3) using the observation data to validate and further improve numerical models.

A large quantity of data has been generated during Collab Experiment 1. Considerable experience has been gained and important lessons have been learned in disseminating the data internally within the Collab team as well as to the broader EGS community. EGS Collab has generated many types of data from the characterization of the site/testbed, the experiments, and various analyses. The quantity of the data is massive and the inter-relationship between the data is very complex. Simply making the data "available" by sharing the raw form and sharing individual data types separably does not address the accessibility issue. Digesting the raw, disconnected data requires a considerable amount time and deep data processing expertise. We had to make numerous innovations to share and present the data in a more digestible or useable form. In this paper, we use our experience to emphasize the importance of organizing, annotating, and presenting many streams of data in an easily understood manner for making informed decisions by the experiment team as well as for future use by prospective data consumers.

2. Types of Data from EGS Collab

We broadly divide Collab data into the following categories:

- *Background data*: Data collected before the onset of the stimulation and circulation tests, mainly to characterize the testbed.
- *Experiment data*: Data generated by or during the stimulation and circulation experiments.
- *Simple data*: Data that in their "raw" forms can be directly consumed by most end users.
- *Processed data*: Data that requires processing before they can be consumed by most end users. In this context, processing refers to simple translations or conversions to normalize

formats, synchronize timesteps, or provide additional contextual information necessary to understand the data. The processing should be largely objective and does not involve inversion or substantial subjective interpretation.

- *Interpreted data*: Data resulting from analyses, simulations, modeling efforts, and expert interpretations. Inversions are often involved, and the analyses often take place in multiple iterations to allow continuous improvements of the results.

Table 1. Examples of various types of data are listed in the table below.

	Background data	Experiment data
Simple data	Well trajectories; rock properties from literature; wellbore temperature survey	Stimulation and flow data (flow rate, pressure, temperature) as the associated notes; daily experiment plan and logs
Processed data	Well logs; fracture picks	DTS measurements; seismic sensor time histories
Interpreted data	Rock velocity field; discrete fracture network (DFN) connectivity; inverted electrical conductivity field	Microseismic catalog; CASSM interpretation; inverted electrical conductivity evolution; interpreted water jetting pattern;

Each of the following sections presents one aspect of our effort to facilitate and maximize data dissimilation.

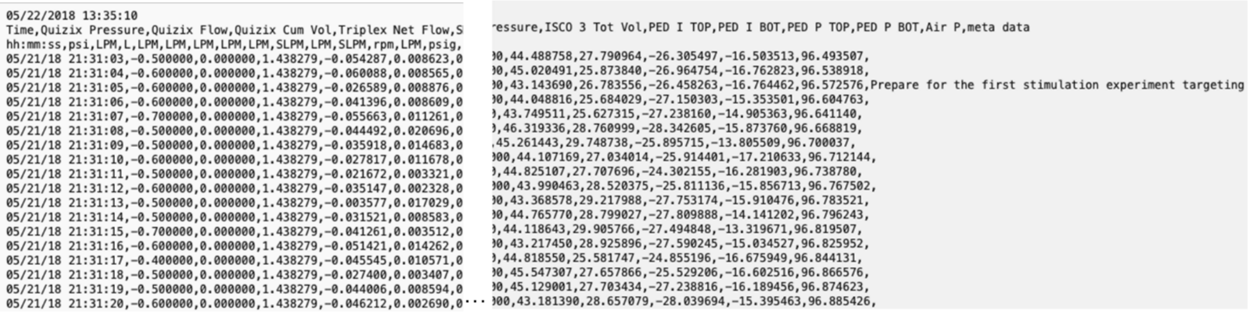
2. Documentation and Record Keeping

The Collab team has diligently and thoroughly documented all activities, incidents, and observations that occurred in the field or in the concurrent tele-meetings supporting the field crew while field tests are being carried out. A few practices that the Collab team found valuable are summarized below. These practices are very much in line with standard experiment documentation procedures.

- **Details in daily shift reports.** Common contents of daily shift reports include field work personnel, plans for the day, observations and anomalies, simple direct measurements and sensor calibration results, scanned notes, as well as work photos. See Section 5 for the dissimilation of daily shift reports.
- **Notes embedded in time-series data.** The flow-and-stimulation system deployed in the Testbed 1 recorded more than 90 channels of time-series data from various sensors. Along with the instrument-captured data channels, the operator of the system could input and embed short notes in real time and stream/store these notes in the same comma-separated value (CSV) file along with other data. Figure 1 shows the beginning segment of a data

file. The notes can be seen in the last column of the data informally named as “meta data”. Often trivial-looking incidented happen while experiment was in progress; having a record of those incidents and observations with real timestamps helped correlate and interpret measurements observed in other time-series data streams with related events.

- **Managing and tracking discussions.** Once the Collab team realized the challenge of tracking technical discussions occurring over emails, the team adopted the Teamwork (teamwork.com) platform to house all discussions. This proved to be highly effective and served as a form of permanent record.



```
05/22/2018 13:35:10
Time,Quizix Pressure,Quizix Flow,Quizix Cum Vol,Triplex Net Flow,S
hh:mm:ss,psi,LPM,L,LPM,LPM,LPM,LPM,LPM,SLPM,LPM,SLPM,rpm,LPM,psig,
05/21/18 21:31:03,-0.500000,0.000000,1.438279,-0.054287,0.008623,0
05/21/18 21:31:04,-0.600000,0.000000,1.438279,-0.060088,0.008565,0
05/21/18 21:31:05,-0.600000,0.000000,1.438279,-0.026589,0.008876,0
05/21/18 21:31:06,-0.600000,0.000000,1.438279,-0.041396,0.008609,0
05/21/18 21:31:07,-0.700000,0.000000,1.438279,-0.055663,0.011261,0
05/21/18 21:31:08,-0.500000,0.000000,1.438279,-0.044492,0.020696,0
05/21/18 21:31:09,-0.500000,0.000000,1.438279,-0.035918,0.014683,0
05/21/18 21:31:10,-0.600000,0.000000,1.438279,-0.027817,0.011678,0
05/21/18 21:31:11,-0.500000,0.000000,1.438279,-0.021672,0.003321,0
05/21/18 21:31:12,-0.600000,0.000000,1.438279,-0.035147,0.002328,0
05/21/18 21:31:13,-0.500000,0.000000,1.438279,-0.003577,0.017029,0
05/21/18 21:31:14,-0.500000,0.000000,1.438279,-0.031521,0.008583,0
05/21/18 21:31:15,-0.700000,0.000000,1.438279,-0.041261,0.003512,0
05/21/18 21:31:16,-0.600000,0.000000,1.438279,-0.051421,0.014262,0
05/21/18 21:31:17,-0.400000,0.000000,1.438279,-0.045545,0.010571,0
05/21/18 21:31:18,-0.500000,0.000000,1.438279,-0.027400,0.003407,0
05/21/18 21:31:19,-0.500000,0.000000,1.438279,-0.044006,0.008594,0
05/21/18 21:31:20,-0.600000,0.000000,1.438279,-0.046212,0.002690,0···
```

```
ressure,ISCO 3 Tot Vol,PED I TOP,PED I BOT,PED P TOP,PED P BOT,Air P,meta data

00,44.488758,27.790964,-26.305497,-16.503513,96.493507,
00,45.020491,25.873840,-26.964754,-16.762823,96.538918,
00,43.143690,26.783556,-26.458263,-16.764462,96.572576,Prepare for the first stimulation experiment targeting
00,44.048816,25.684029,-27.150303,-15.353501,96.604763,
0,43.749511,25.627315,-27.238160,-14.905363,96.641140,
0,46.319336,28.760999,-28.342605,-15.873760,96.668819,
,45.261443,29.748738,-25.895715,-13.805509,96.700037,
000,44.107169,27.034014,-25.914401,-17.210633,96.712144,
0,44.825107,27.707696,-24.302155,-16.281903,96.738780,
000,43.990463,28.520375,-25.811136,-15.856713,96.767502,
00,43.368578,29.217988,-27.753174,-15.910476,96.783521,
00,44.765770,28.799027,-27.809888,-14.141202,96.796243,
0,44.118643,29.905766,-27.494848,-13.319671,96.819507,
00,43.217450,28.925896,-27.590245,-15.034527,96.825952,
0,44.818550,25.581747,-24.855196,-16.675949,96.844131,
00,45.547307,27.657866,-25.529206,-16.602516,96.866576,
00,45.129001,27.703434,-27.238816,-16.189456,96.874623,
000,43.181390,28.657079,-28.039694,-15.395463,96.885426,
```

Figure 1: A truncated snapshot of one of the stimulation-andflow time history files (CSV format). The notes (in-file metadata) can be seen in the last column.

3 Prompt Data and Result Sharing

An organized and self-explanatory directory structure, established on OpenEI (https://foundry.openei.org), was used to share, upload, and store data as soon as the data became available (theonly exception being large-size raw seismic data). In addition, stimulation and flow data were streamed to OpenEI in real time for timely analyses and decision making by a team observing the tests online. Eventually, data files are directly submitted to GDR from Data Foundry for public dissemination (Weers and Huggins, 2019). Data from the OpenEI database are made available to all project personnel, and other interested parties.

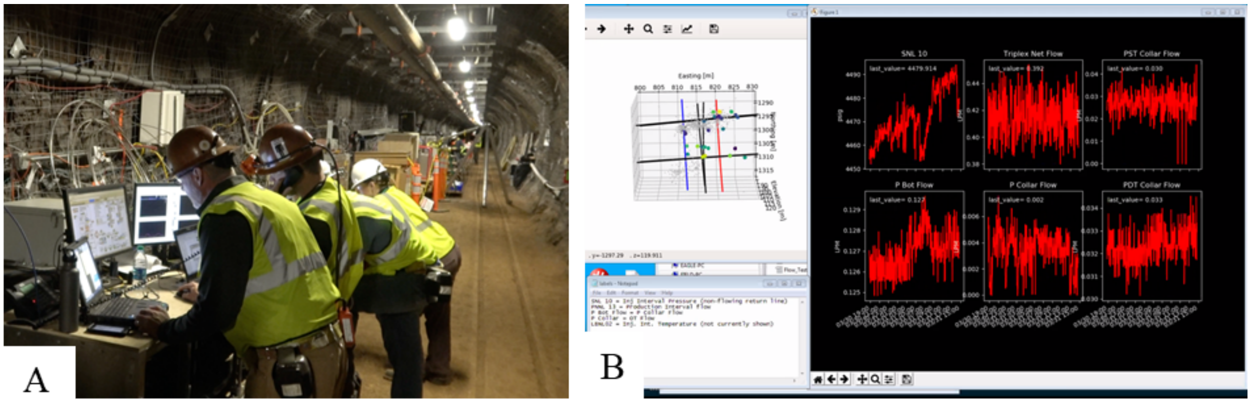


Figure 2: Real-time timeseries sharing through an online video conference system. A) Field operation team at the SURF 4850 level. B) Shared screen from the data acquisition computer, with the six dark panels showing time histories in a 24-hour time window. The upper left corner of shared screen shows microseismic events, resolved in near-real time.

Collab experiments were planned and carried out iteratively. On each field-work day, the crew (Figure 2A) started the work according to a predetermined plan, but the plan was usually adjusted based on observations. Very often, the daily plan itself had a decision tree built in with many "if" conditions. Project managers and other experts remotely monitor the experiments and they are involved in the adjustments of the plans. It is therefore critical to share the data as early as possible to allow informed decisions to steer the experiments to achieve the optimal results. During each day of testing, six channels (out of 90+) that were most relevant to the objective of the day were displayed on the screen of the data acquisition computer and shared with all remote participants (Figure 2B). A near-real time (within one to two minutes) edge-computing microseismic inversion capability played a crucial role in real-time decision-making during experiments as well as for planning the subsequent experiments.

4 Synchronization of Multi-source Data

When different measurements are synchronized spatially and/or temporally, they may reveal processes that are obscured in individual measurements, or corroborate each other to result in more definitive conclusions. Figure 3 illustrates synchronized injection data, the microseismic events, and DTS data in two wells for a hydraulic test conducted in the Fall 2018. This plot shows that injection at higher rates (0.8 to 5 L/minas noted by green bars "a" through "l" in the upper panel; baseline rate was 0.4 L/min) created MEQs and helps to pinpoint where the created/stimulated fractures intersected the wells in the DTS plots.

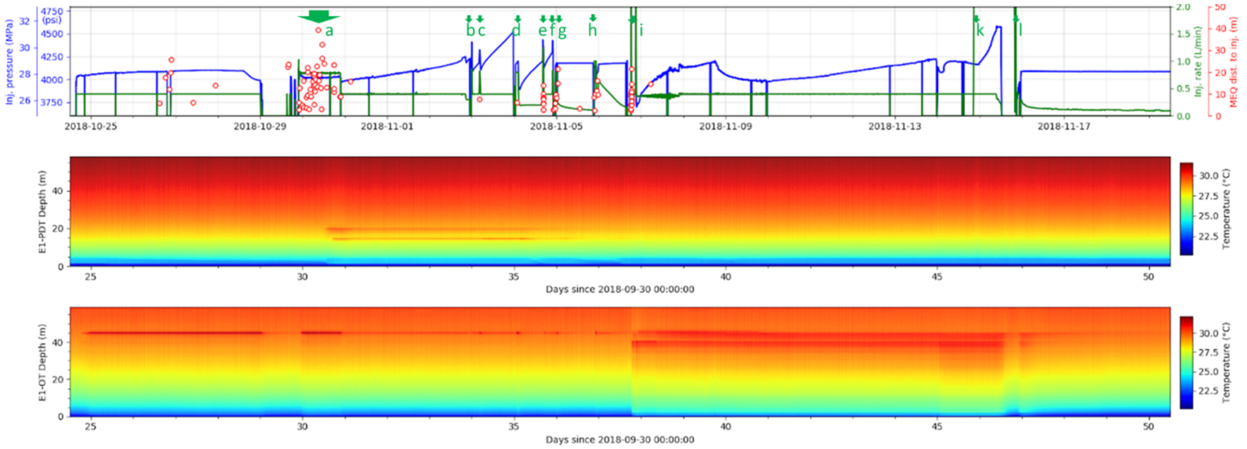


Figure 3: Synchronized flow data (blue line is pressure curve, green line is flow rate curve, and green arrowhead with markers "a" through "l" are periods of higher flow (=>0.8 L/min injection rates in the upper panel), MEQs (red circles in the upper panel), and DTS data from two wells (middle and lower panels). Each vertical line of DTS data show temperatures from the drift wall (lowest part of the panel – blue) to the deepest (top, red). Repeated measurements allow for the apparent continuous plot over time.

Another example, illustrated in Figure 4, shows DTS data measured along the six monitoring holes during the first five months of the long-term circulation test in 2019. In addition to the thin warmer bands that show intersections between flowing fractures and wellbores, wider bands of cooling signals gradually emerged in the E1-PST and E1-PSB (indicated by markers "4" and "5", respectively). Such wider cooler temperature anomalies also emerged along E1-OT and E1-OB. In a complementary figure (Figure 5), the change in temperature from baseline temperature in the monitoring wells is shown at their actual spatial locations. Although the cooling segment in E1-PST seems to coincide with some logged fractures, there are no flowing fractures near the cooling segment in E1-PSB. In addition, the overall cooling in these four affected wells was correlated with the first 50 m of the E1-I which was used for chilled-water circulation. Because of this synchronization of multiple data streams, we hypothesized that the cooling in these four wells was induced by thermal conduction from E1-I. Subsequent modeling work validated that the radial conductive cooling did happen, and suggested that the first 50 m of E1-I withdrew about 10 times more heat out of the testbed than did the circulation of chilled water through the main fractures. This hypothesis has been validated by a high-fidelity model (Wu et al., 2021). Because this hypothesis was not apparent, the true physics would have been obscured if it were not for the thought-provoking visualization that accurately presents the spatial relationships between the measured signals.

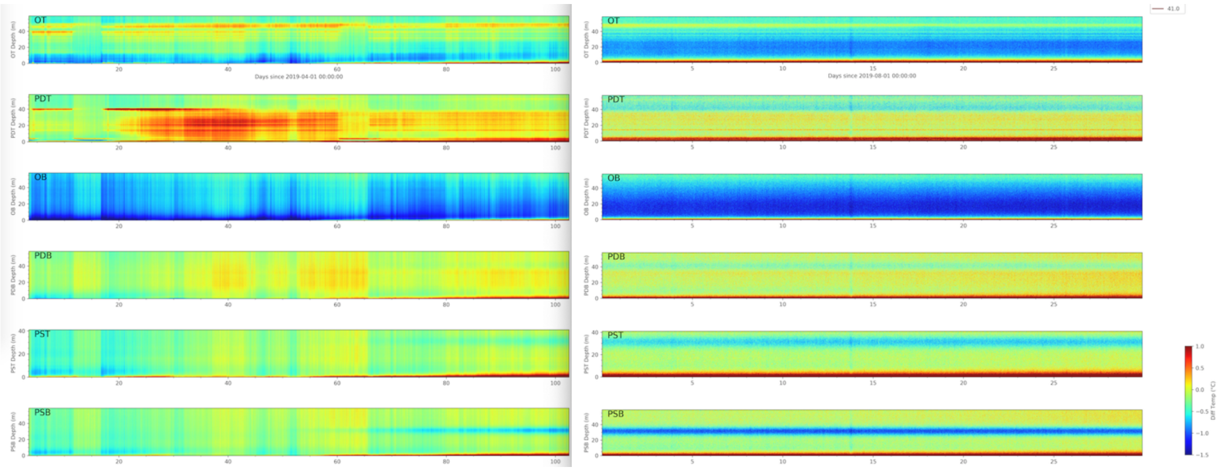


Figure 4 DTS measurements in terms of temperature change in six wells. The left panels show results in April to June, and the right panels show results in August 2019.

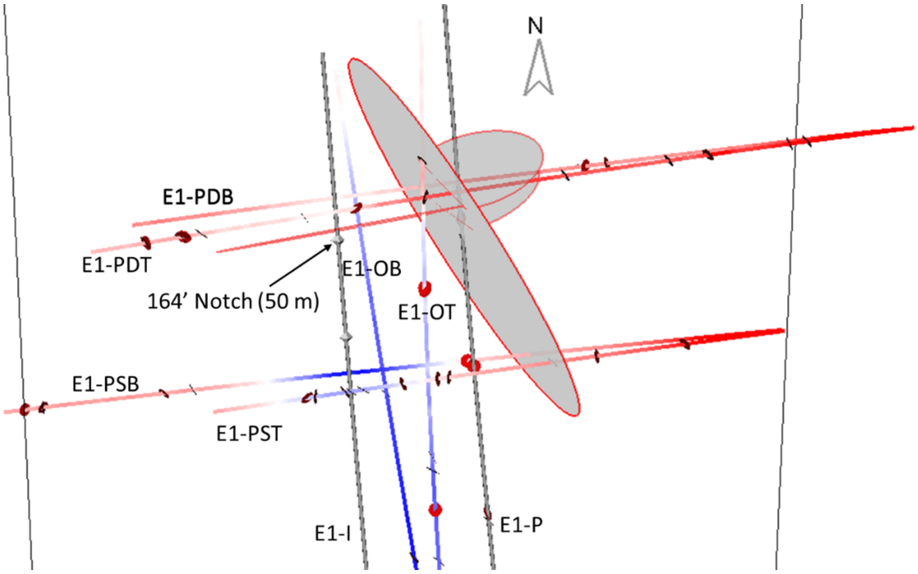


Figure 5: Spatially synchronized data: trajectories of the eight wells, temperature change (blue meaning cooler and red meaning hotter than the background) along the six observation wells as of September 2019, logged open fractures along wellbores shown as small disks, and prominent hydraulic fractures stimulated in earlier experiments (larger disks). The chilled-water circulation in E1-I was limited to 164’ Notch, and the deeper part of the hole was isolated from chilling.

5. Data Accessibility

A successful experience of EGS Collab is that we try to make the data available as early as possible and available to as many people as possible. Different participants of the experiments, including modeler, experimentalists, geophysicists, hydrologists, senior scientists, graduate students, etc., bring different perspectives to the interpretation. Many important discoveries in Collab were not made by the "direct owners" of the data, namely those who collected and/or processed the data, but by "spectators". This was only possible in a "borderless" data sharing environment.

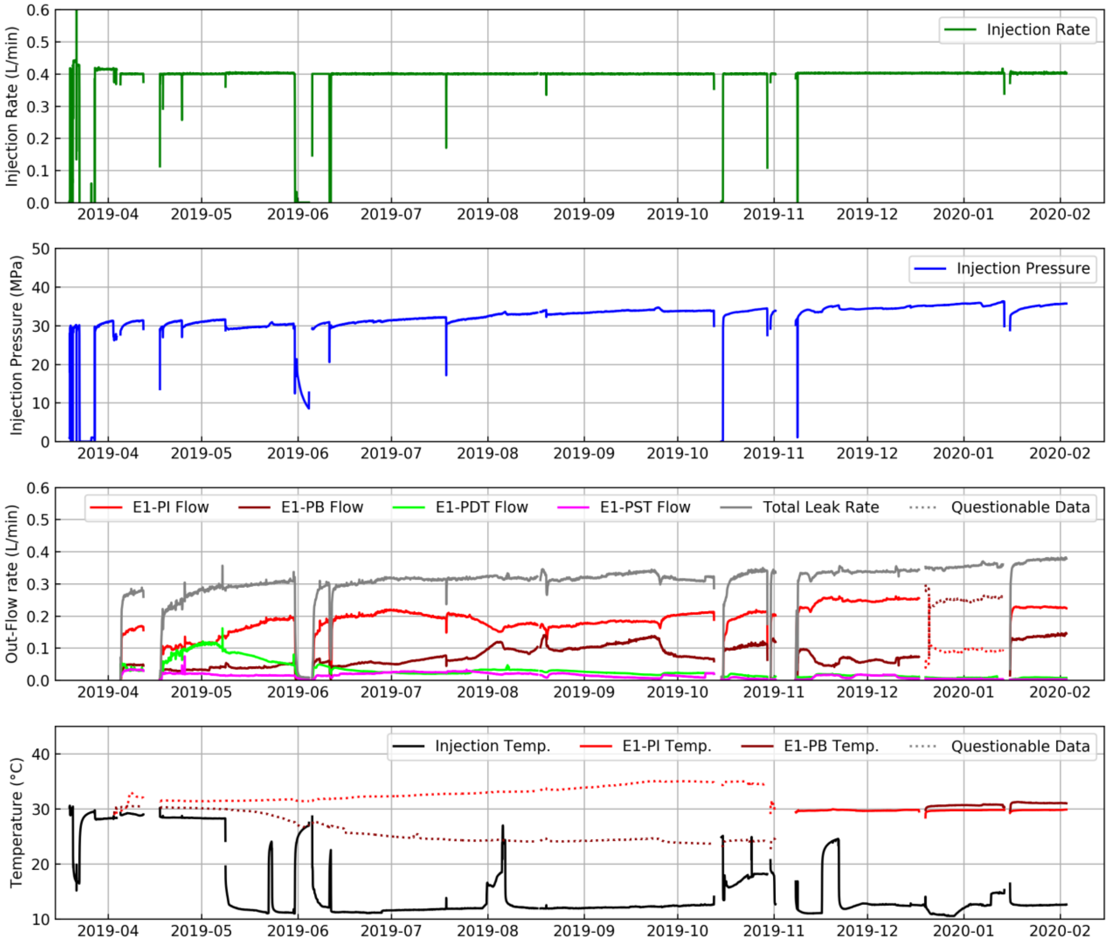


Fig. 6 A plot combining key measurements during the long-term water circulation test performed at the EGS Collab Experiment 1 testbed. Injection rate and pressure are shown in the first and second panels respectively. Outflows were mainly observed at E1-P, E1-PDT and E1-PST as shown in the third row. The dotted line segments denote questionable temperature and outflow measurements as discussed in the text. The total outflow rate is also plotted. The fourth panel shows the injection and production temperatures. Note that systematic and continuous measurements outflows started in early April 2019 as shown in the third panel.

To make "borderless" data sharing possible, using an open data sharing platform like OpenEI is necessary but not sufficient. We encouraged all participants to use free and open-source data formats. The "Fat Crayon" toolkit (https://github.com/joepmorris/FatCrayonToolkit) developed by Joe Morris for the design of the testbeds was written in Python and the outputs are in the open-source VTK format. We also converted DTS data from Matlab to the Python Pickle format and to a more accessible CSV format for participants who do not use Matlab or Python. Additionally, we devoted a significant amount of effort to visualizing the complex dataset in an intuitively accessible way, as exemplified by some of the cases discussed in this paper, such as Figure 3 and Figure 6.

A recent effort to make Collab data and learnings accessible to a wide audience was to create a Collab wiki site. Significant benefits of a wiki site, compared with more traditional documentation method, include:

- Centralized: All information in a variety of formats can be housed in or at least linked to wiki pages.
- Open, live, and lasting: Anyone can access the information, even long after the closure of the project.
- Democratic: Anyone with proper knowledge of the work, the data, and the results, including future users of the data, can update the wiki pages.
- Logical: All the information can be cross-linked, including internal links among the wiki pages and external links such as to the GDR submissions.

The EGS Collab Wiki pages (https://openei.org/wiki/EGS_Collab_Project_Overview) are hosted by openei.org/wiki and are being continuously created and updated. In addition to "topical" pages, we have also used wiki to synchronize wellbore logs and core photos. All the Experiment 1 shift reports have also been converted to indexed and searchable wiki pages (https://openei.org/wiki/EGS_Collab_Shift_Reports).

6. Multi-level Data Reduction of Massive Time Histories

The climax of EGS Collab Experiment 1 was the long-term water circulation test that was conducted from early 2019 to early 2020. The primary objective of the flow test was to provide data to validate computer models concerned with flow processes and heat exchange processes in EGS. The circulation test generated a large amount of data. The flow system collected data from ~100 channels at an output rate of 10 s (sometimes at 1 s), generating approximately 3 million rows of data, stored in more than 100 CSV files with a total size of 3 GB. In addition to the large data volume, another challenge for using the data is that the correspondence between physical quantities and data collection channels is very complicated for several reasons: (1) Two pumps were used in an alternating fashion: a Quizix pump and triplex pump; (2) Some channels were added in the middle of the test; (3) Occasionally, mistakes in connecting the sensors were made after system repair or reconfiguration; (4) Some sensors were later found to be faulty.

To facilitate the use of this valuable dataset byusers with different levels of sophistication in computing tools and different levels of familiarity to the test, we have created several products. All products and the Python scripts used to generate them are available at http://gdr.openei.org/submissions/1301.

Raw channel plots

Data streams from selected channels (not all channels recorded meaningful data) are plotted in a multi-page PDF file. Each page includes one month of data as shown in Figure 7. The nominal meanings of these channels were annotated on the plots while the meanings were not always accurate due to the reasons mentioned above. The Python script used to generate these plots is included in the GDR release. Note that the channels are groups based on similarities, namely, flow rates are grouped together and temperatures are grouped together. The plots are visually distracting due to the lack of further processing.

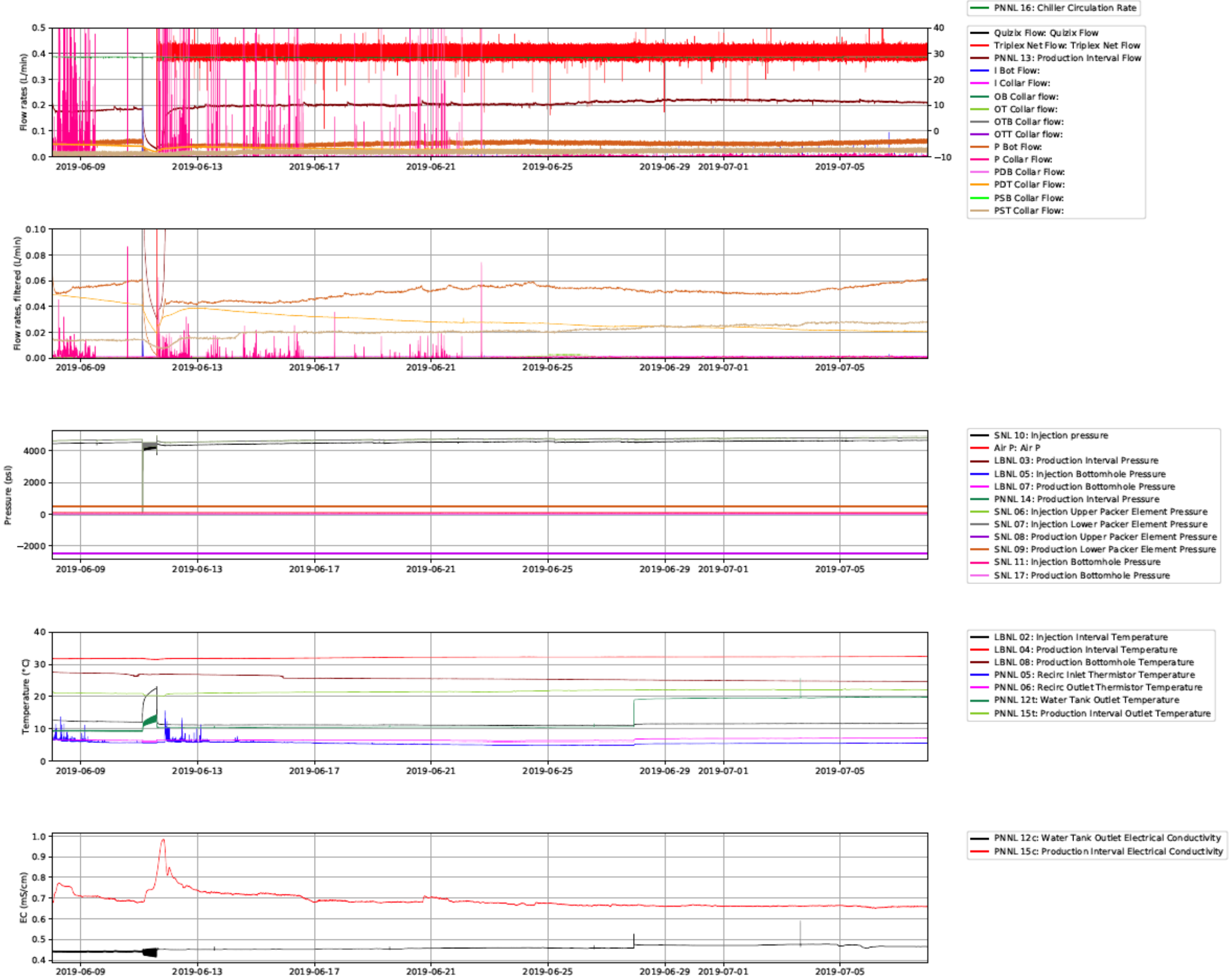


Figure 7 One page in the raw channel plot connections covering one month of data streams.

Single-quantity streams

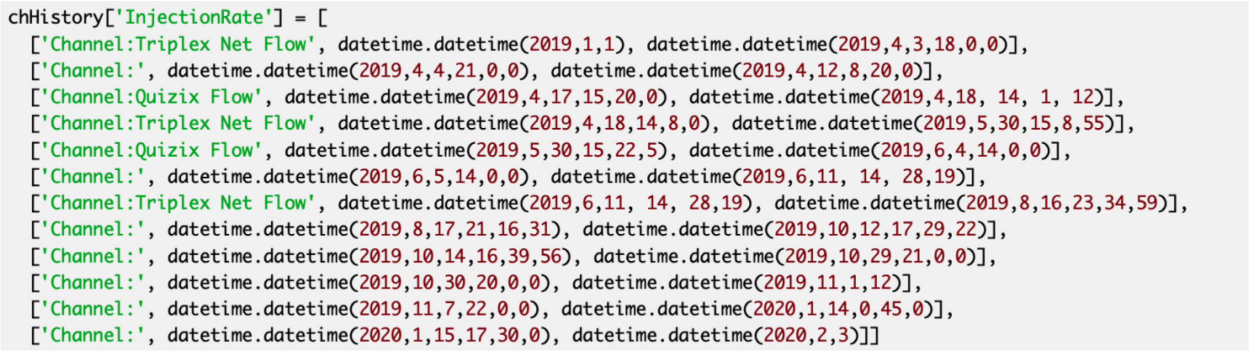
Nine quantities were selected to generate single-quantity stream files. These are the most relevant quantities for EGS modeling.

1. Injection Pressure
2. Injection Rate
3. Flow rate out of E1-P Interval
4. Flow rate out of E1-P Bottom
5. Flow rate out of E1-PDT
6. Flow rate out of E1-PST
7. Injection Interval Temp

8. Temperature at E1-P Interval
9. Temperature at E1-P Bottom

As noted before, each quantity could have been measured by different channels at different stages of the long-term circulation test. The data stream for each quantity is organized into a single CSV file. The user would not need to worry about or even understand which sensors were used to collect the data. All the embedded notes (stored in the "metadata" column of the original CSV files) in the original CSV files are included in every stream file, regardless of whether they are relevant to the steam itself or not. The name of the original files and the channel names are also included as notes so that interested users could track where the data were from. Because the PI and PB temperature thermistors had issues, we store the readings before the thermistor replacement and after the replacement into separate files, named as "_Questionable" and "_Upgraded", respectively

The Python script used to generate these steam files is included in the GDR release. The channel switching and skipping of meaningless segments (from diagnostic testing or interruptions) are explicitly "documented" in the script as shown in Figure 8.



```
chHistory['InjectionRate'] = [
  ['Channel:Triplex Net Flow', datetime.datetime(2019,1,1), datetime.datetime(2019,4,3,18,0,0)],
  ['Channel:', datetime.datetime(2019,4,4,21,0,0), datetime.datetime(2019,4,12,8,20,0)],
  ['Channel:Quizix Flow', datetime.datetime(2019,4,17,15,20,0), datetime.datetime(2019,4,18, 14, 1, 12)],
  ['Channel:Triplex Net Flow', datetime.datetime(2019,4,18,14,8,0), datetime.datetime(2019,5,30,15,8,55)],
  ['Channel:Quizix Flow', datetime.datetime(2019,5,30,15,22,5), datetime.datetime(2019,6,4,14,0,0)],
  ['Channel:', datetime.datetime(2019,6,5,14,0,0), datetime.datetime(2019,6,11, 14, 28,19)],
  ['Channel:Triplex Net Flow', datetime.datetime(2019,6,11, 14, 28,19), datetime.datetime(2019,8,16,23,34,59)],
  ['Channel:', datetime.datetime(2019,8,17,21,16,31), datetime.datetime(2019,10,12,17,29,22)],
  ['Channel:', datetime.datetime(2019,10,14,16,39,56), datetime.datetime(2019,10,29,21,0,0)],
  ['Channel:', datetime.datetime(2019,10,30,20,0,0), datetime.datetime(2019,11,1,12)],
  ['Channel:', datetime.datetime(2019,11,7,22,0,0), datetime.datetime(2020,1,14,0,45,0)],
  ['Channel:', datetime.datetime(2020,1,15,17,30,0), datetime.datetime(2020,2,3)]]
```

Figure 8 A snippet of the Python script that "stitches" data streams together based on physical meaning. The script serves as a full disclosure of processing of the raw data.

Hour-interval streams

The single-quantity streams, in 10 s or 1 s intervals, still resulted in large files, with up to more than two million lines of record each. The data have been processed to generate hour-interval data streams for easy handling by prospective users. To resample the data, each stream is first divided into separate, continuous segments. Within each segment, the first data point and the last data point are preserved. The data in-between are resampled by averaging values within hour-length intervals. After reduction, each stream only includes several thousands of lines. The script to resample the streams is also included in the GDR release.

Scripts as documentation

In the data processing mentioned above, all scripts were released along with the processed data. This reflects the philosophy of "scripts as documentation". The two main benefits of this approach are: (1) The prospective users would know what has been done to the data. Some subjective judgements were exercised and ad-hoc treatments applied in processing the data. The scripts serve as permanent records of these treatments. (2) The scripts serve as examples and instructions on how to use the raw data.

6. Closing Remarks

As a part of the EGS Collab effort, mainly through a collaboration among the participating national laboratories, we innovated in many ways to share data and knowledge from our ongoing experiment effort with a broad audience. The philosophy underlying this effort is that a complex dataset's value is only realized if a diverse community can access, digest, and utilize it.

Acknowledgement

This research was performed in support of the EGS Collab project; taking place in part at the Sanford Underground Research Facility in Lead, South Dakota. The assistance of the Sanford Underground Research Facility and its personnel in providing physical access and general logistical and technical support is acknowledged. Support from the EGS Collab team is gratefully acknowledged. This work was supported by U.S. Department of Energy, Geothermal Technologies Office, and performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This paper is LLNL report LLNL-CONF-823860. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

REFERENCES

Kneafsey, T. et al. "The EGS Collab Project: Status and Accomplishments." *Geothermal Rising Conference*, 2021.

Wu, H., Fu, P., Frone, Z., White, M.D., Ajo-Franklin, J.B., Morris, J.P., et al. "Modeling heat transport processes in enhanced geothermal systems: Validation at EGS Collab Experiment 1." *Geothermics* (2021), submitted.

Weers, J. and Huggins, J. "Getting Data Out of the Ground: Modern Challenges Facing EGS Collab, the DOE Geothermal Data Repository, and the Geothermal Industry", (2019) *44th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, California.