The Station Information System (SIS): A Centralized Repository for Populating, Managing, and Distributing Metadata of the Advanced National Seismic System Stations

by Ellen Yu, Prabha Acharya, Justin Jaramillo, Sue Kientz, Valerie Thomas, and Egill Hauksson

ABSTRACT

Creating, maintaining, and archiving accurate station metadata is critical for successful seismic network operations, data discovery, and research. The Station Information System (SIS) is a centralized repository of seismic station equipment inventory, instrument response, and site information of stations operated by regional seismic networks (RSNs) of the Advanced National Seismic System (ANSS; Filson and Arabasz, 2017). It has a web-based user interface that enables the creation and manipulation of the corresponding metadata. The system can track the installation, maintenance, and removal of equipment from a site, which often results in the creation of new metadata epochs. SIS also computes the overall response, including gain, of a data channel by combining the responses of the underlying hardware components. SIS distributes this information in standard formats such as Federation of Digital Seismic Networks StationXML and dataless Standard for the Exchange of Earthquake Data. SIS can also be used to manage inventory of field equipment such as power, telemetry, or Global Positioning System antenna, as well as links to other site-related repositories external to SIS to give the network operator the most complete view of a site and the overall network. This article summarizes the main features in SIS. We present its basic infrastructure, holdings, workflow, and how RSNs retrieve data from it. We also explain the reasoning to pursue one centralized repository and why it supports the goals of SIS and the ANSS. We demonstrate that by providing the ANSS network operator with a comprehensive site view, SIS enables the production of high-quality metadata, a necessary prerequisite for producing high-quality seismic data.

INTRODUCTION

We define station metadata as the information needed to convert digital time series recorded at a station to the observed earth units. In seismic networks, this is most commonly, but not limited to, the ground motion. Accurate station metadata are essential to researchers and other consumers of geophysical data. The quality of all seismic network products, such as the earthquake catalog, ShakeMaps, or moment tensor solutions, depends on the availability of accurate site locations and channel response. Higher quality metadata such as derived instrument-specific calibration information can help researchers model earth structure with more accuracy than using nominal instrumental response data (e.g., Davis and Berger, 2012; Ringler et al., 2012). Accurate station metadata also facilitate faster and easier discovery of information (Giles, 2011).

The seismological research community has made considerable progress with metadata standards with the Federation of Digital Seismic Networks' (FDSN) adoption of the Standard for the Exchange of Earthquake Data (SEED) format in 1988 (Incorporated Research Institutions for Seismology [IRIS] SEED manual v2.4, Ahern et al., 2009, with an abbreviated description found in Ringler and Evans, 2015). This has allowed for an easy exchange of data among various data centers as well as development of software for data discovery and data analysis. This data format has also enabled the creation of metadata with utilities such as the IRIS Nominal Response Library (NRL; Templeton, 2017) and Portable Data Collection Center (Casey, 2012).

Accurate station metadata are also critical to the network operator, the provider of the data. For instance, instrument response and geographic location are two of the critical factors considered by network operators to guide their decisions on what equipment to purchase (Bockholt, 2017). However, there are many other pieces of station information needed by a network's personnel to implement its scientific and public safety goals that do not contribute to instrument response and therefore are not typically considered station metadata. These pieces of information are also not well described in the SEED format. While a SEED volume might describe the data accurately, it does not completely describe the installation of the seismic equipment that produced the data. The SEED format was not designed to provide a complete record of the equipment
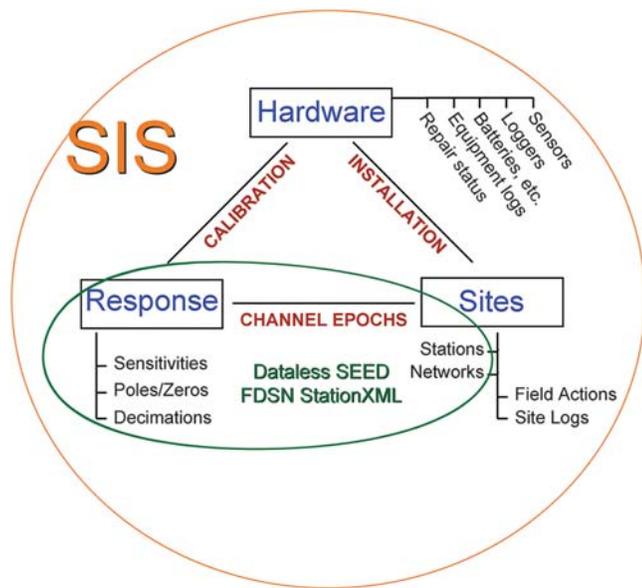
▲ **Figure 1.** While dataless Standard for Exchange of Earthquake Data (SEED) and Federation of Digital Seismic Networks (FDSN) StationXML describe seismic data, they only contain the subset of information describing a seismic station contained in the green oval. Network operators who maintain seismic stations must know additional information contained in the orange oval such as the specific equipment at a site, how they are connected, their configuration, firmware version, and more. SIS, Station Information System.

repairs, upgrades, and site installations that is needed by network operators to effectively maintain or improve the network. For example, a common installation type involves digitization of a sensor signal at a different site from where it is recorded. A channel response-only approach does not accurately reflect the location of all equipment in this type of deployment, especially if there are preamplifier elements in the response chain. The SEED format cannot easily describe equipment types such as those needed for power and communication of seismic data to processing applications, which are also vital to the operation of station.

The FDSN StationXML format allows greater description of equipment deployed at a site. However, it still lacks information important to the network operator, such as equipment component information (e.g., components on a triaxial sensor) or equipment attributes (e.g., firmware and voltage) to name a few. FDSN StationXML also does not track inventory, so a sensor not installed at a station cannot be described in this format.

Without a systematic and comprehensive approach to integrating the information about the field equipment deployment and the resulting metadata, efforts to create and maintain accurate station metadata could compete against rather than complement other important tasks such as improving the network. In these situations, much of the operator's valuable time is spent fixing data entry errors, struggling with shell scripting

or Structured Query Language (SQL), rather than mission-focused activities that would improve data quality or completeness, such as planning, deployment, data quality control, or equipment repair. Such competition for resources results in lower data quality as well as possibly erroneous and delayed metadata.

Station Information System (SIS) provides a valuable bridge between field equipment and the end user's metadata (Fig. 1). It does this with a centralized relational database and a user interface (UI) that provides a comprehensive view of site equipment and metadata, and it is integrated with the workflow of operations, making it a closed-looped system. As a result, the time and effort to update metadata after a change in the field is greatly decreased.

SIS was originally developed by the Southern California Earthquake Data Center (SCEDC) at the California Institute of Technology (Caltech) to manage the metadata needs of the Caltech/U.S. Geological Survey (USGS) Southern California Seismic Network. SIS has been under development during the past four years to meet the needs of multiple regional seismic networks (RSNs) with the goal of becoming the repository for USGS/Advanced National Seismic System (ANSS) funded networks. A technical group comprised representatives from various ANSS RSNs oversees its development. The development staff is located at Caltech.

OVERVIEW OF STATION INFORMATION SYSTEM

We show an overview of SIS infrastructure in Figure 2, which consists of an UI, relational database, and webservers. The UI is written in a Python-based web framework. The UI accesses an Oracle database to store or retrieve information. Through web-based forms (no client needed), authenticated users access a repository shared by all SIS users. The user can enter and edit inventory, instrument response, field actions, and site information, all of which are stored in the underlying database. Users can also make updates or corrections to historical metadata. When the user is finished making changes to a site in SIS, site information can be published to a public file server in standard formats such as dataless SEED and FDSN StationXML. These files are then available for download to consumers of station metadata such as seismic networks, data centers, and researchers. The published files also provide a consistent image of the station response at any given time. This allows SIS users to do multiple updates to metadata and quality control without being disruptive to the processes querying the metadata. It also means that critical real-time processes need not rely on SIS being available.
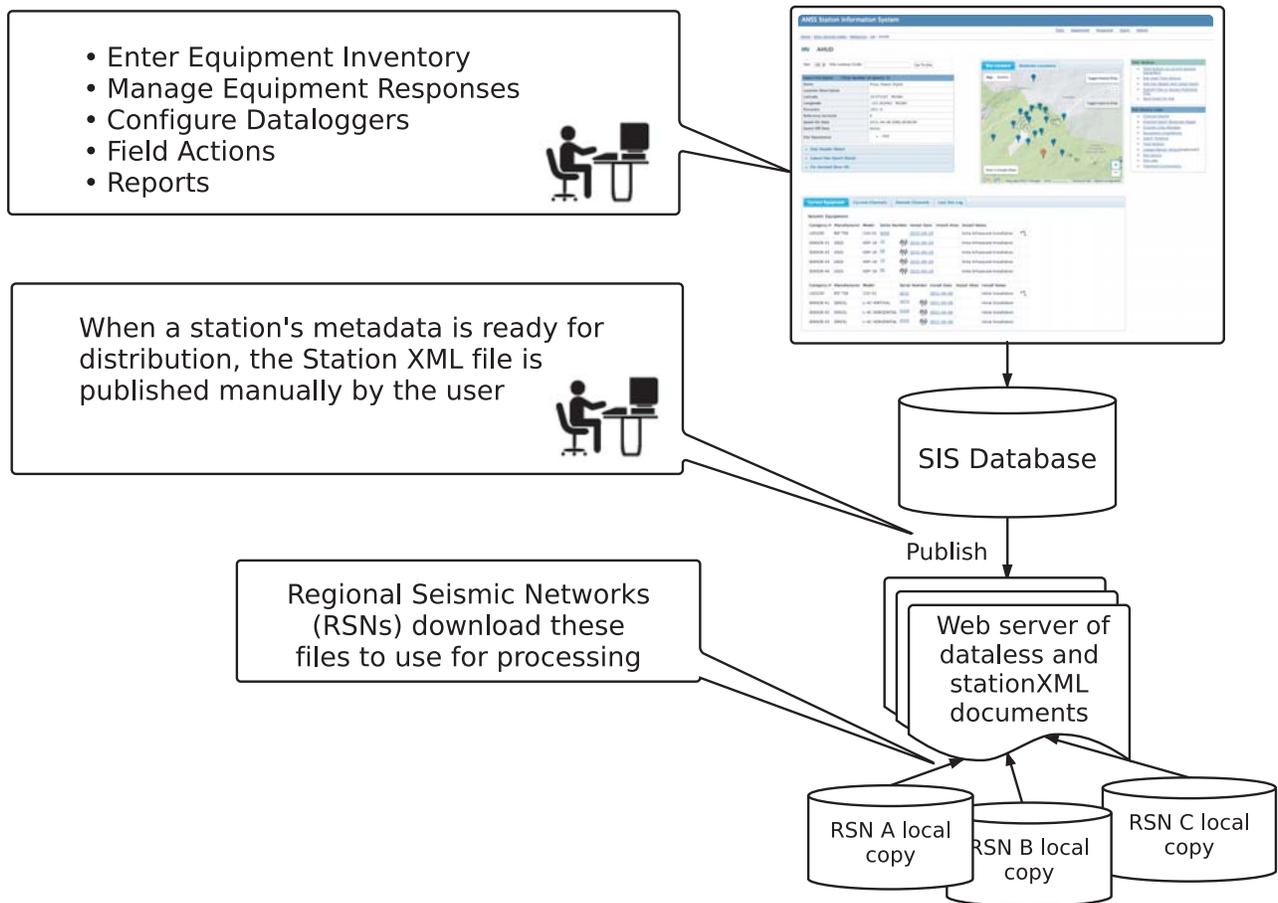
To ensure proper documentation of SIS, we use a wiki that hosts user documentation, design documents, the trouble ticketing system, and also minutes of meetings of the Technical Implementation Committee.

SIS INPUT FORMATS

In addition to entering data through the UI, SIS allows users to batch load station metadata in a format called ExtStationXML.

Station Information System (SIS) Overview

• Enter Equipment Inventory
• Manage Equipment Responses
• Configure Dataloggers
• Field Actions
• Reports

When a station's metadata is ready for distribution, the Station XML file is published manually by the user

SIS Database

Publish

Regional Seismic Networks (RSNs) download these files to use for processing

Web server of dataless and stationXML documents

RSN A local copy

RSN B local copy

RSN C local copy

▲ **Figure 2.** Overview of SIS and its main components. Advanced National Seismic System (ANSS) regional seismic networks (RSNs) use the metadata contained in the published files from SIS. This ensures that users are getting a consistent image of a site, and they do not depend on the availability of SIS.

This extension of FDSN StationXML also describes hardware components and how they are connected to their respective response information in an added element called SubResponse. This format can also describe equipment installed at a site that does not contribute to seismic data channel response—for example, telemetry and power information. If site information is only available in FDSN StationXML, a converter developed by Philip Crotwell at South Carolina Seismic Network can be used to convert FDSN StationXML to ExtStationXML.

SIS OUTPUT FORMATS

SIS writes out metadata for sites in its repository in three primary formats: FDSN StationXML, dataless SEED, and ExtStationXML. They are organized by FDSN network code and available for download at the SIS public file server. SIS uses the IRIS FDSN StationXML-SEED Converter to generate the dataless SEED from the FDSN StationXML file. SIS can also produce .kml files and reports in .csv format.

COMMUNITY REPOSITORY OF SEISMIC EQUIPMENT AND INSTRUMENT RESPONSE

SIS is one of the most comprehensive repositories of modern seismic instrumentation (equipment models and responses) in the United States because the repository is maintained by the SIS user community and by automated updates from the IRIS NRL (Fig. 3). To date, SIS contains information on 220 sensor models and 120 logger models compiled from the IRIS NRL and networks using SIS. Because this hardware and response information is in a centralized repository, SIS allows networks to share and standardize seismic equipment model names and responses. Once a new equipment model is entered into SIS by any network, other users can use it as well. This allows easy transfer of information among SIS users, and shortens the time an analyst needs to research and enter a new equipment model into the system (Fig. 4).

The use of a centralized database also makes it easy for SIS to track equipment that is used by multiple RSNs. This is a
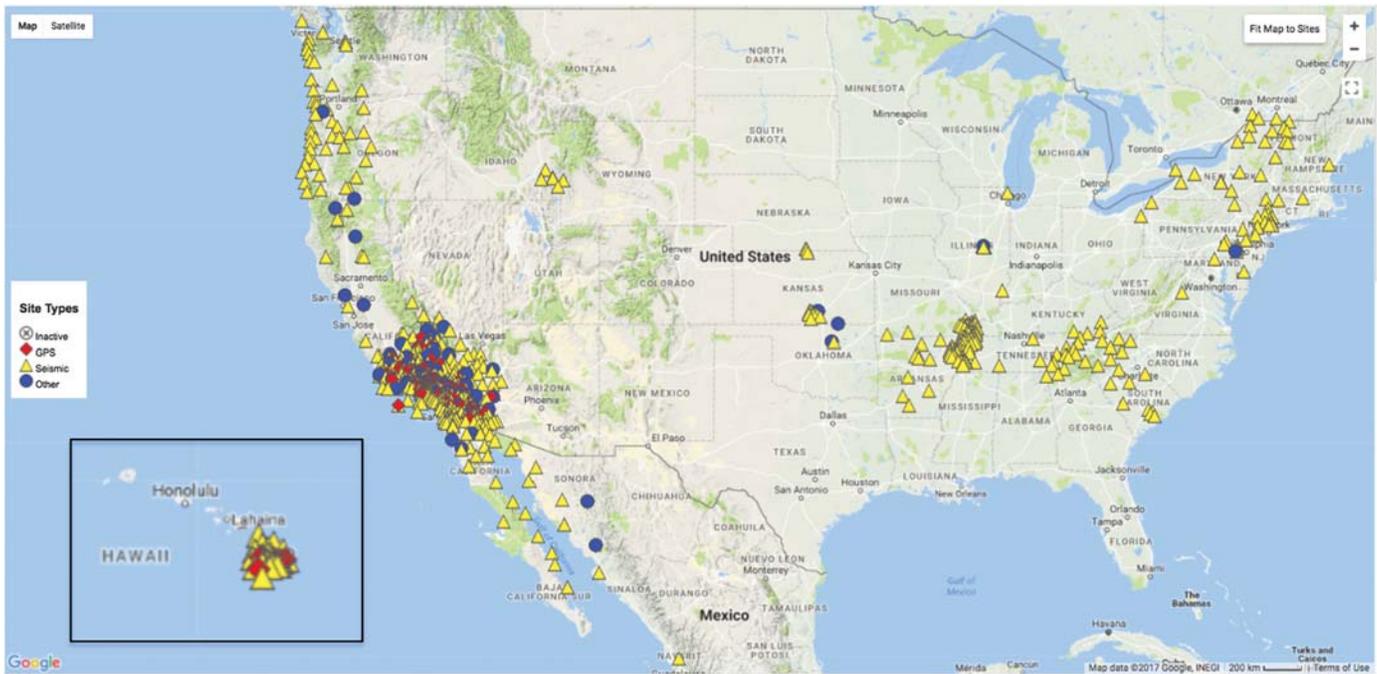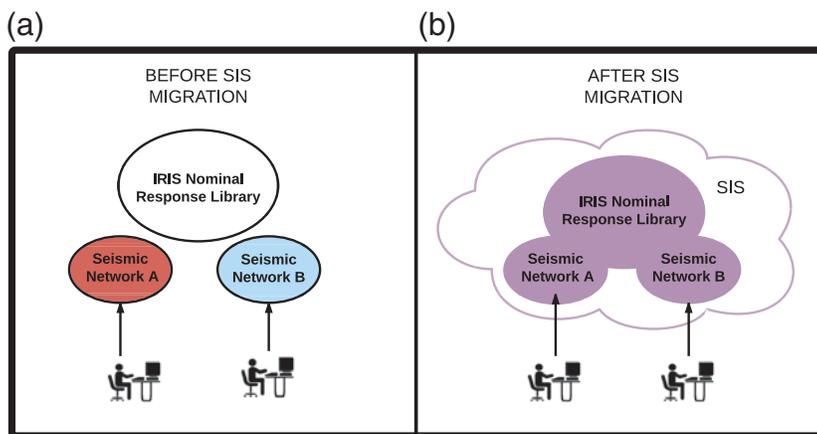
▲ **Figure 3.** ANSS networks and their stations that use SIS for handling metadata. This list is not final and expected to grow to the full complement of U.S. Geological Survey (USGS)/ANSS funded networks.

common case with the use of the ANSS Depot, a seismic equipment depot used by several ANSS networks. The ANSS Depot may receive a logger for repair from one RSN, distribute a different instrument back as a replacement, and send the logger to a different RSN after it has been repaired.

▲ **Figure 4.** (a) Prior to using SIS, users from network A would only have access to equipment models and responses that had been entered into network A's repository. Similarly, users from network B would only have access to what had been entered in network B's repository. Both users can retrieve data from the Incorporated Research Institutions for Seismology (IRIS) Nominal Response Library (NRL), but this would be a manual process. (b) In SIS, users can access all entries. If users from network B enter a new equipment model and/or response, network A users can begin using the new entry as well. Likewise, new entries into the IRIS NRL are automatically made available to SIS users.

SIS also reduces the effort that networks need to describe their instrument response. Like equipment models in SIS, once a response is entered into the SIS repository, it is available for all networks to review and use.

Users can associate response such as poles and zeros with seismic instruments. Users can create their own responses, use existing ones in SIS, or responses from the IRIS NRL using a graphical tool developed by Philip Crotwell. This feature is useful not only for field technicians who can save time by quickly using a nominal response from the IRIS NRL, but also for analysts who wish to further calibrate their responses and in turn produce more accurate ground-motion measurements (Davis and Berger, 2012; Ringler *et al.*, 2012), or for networks wishing to describe instruments deployed using earlier generations of hardware that might not be currently described in the IRIS NRL.

ACCURATELY DESCRIBING A WIDE VARIETY OF EQUIPMENT IN SIS

One major design goal of SIS is to track all equipment that directly contribute to seismic channel response as well as all field equipment deployed at a seismic station. To accomplish this, SIS has a generalized and flexible data model that allows users to create and store new types of equipment (Fig. 5). Users also have
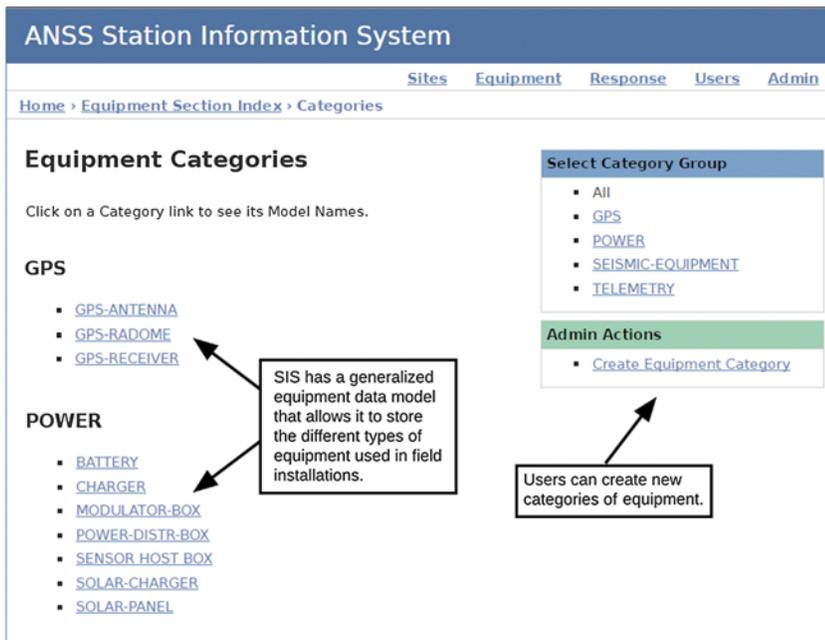
▲ **Figure 5.** SIS has a flexible data model to store many types of seismic instrumentation and associated power systems, communications, and other equipment to fully describe the seismic station.
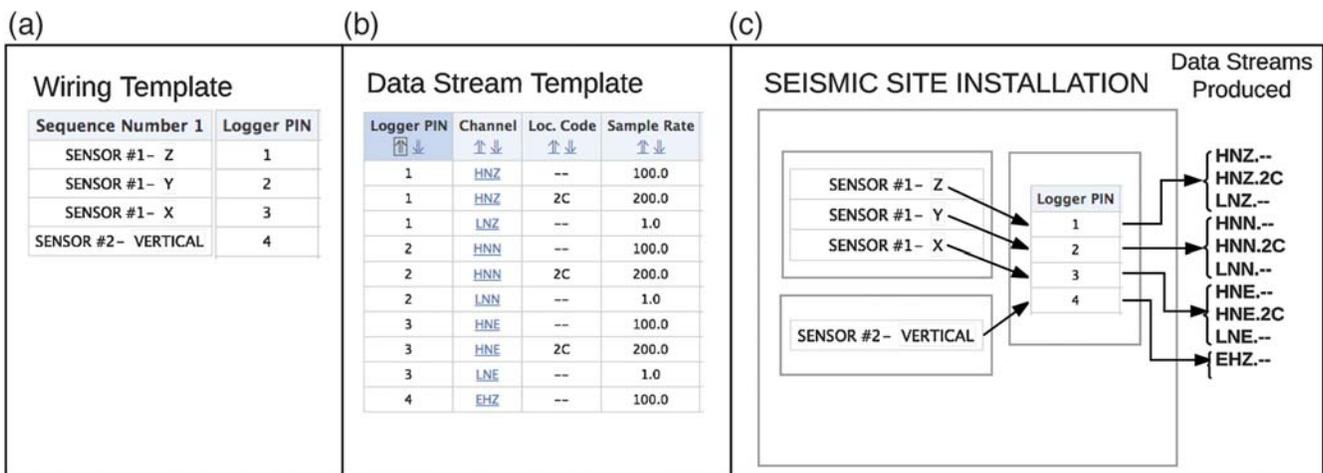
DESCRIBING A VARIETY OF SEISMIC INSTALLATIONS AND SEISMIC DATA CHANNELS IN SIS

Two of the greatest challenges in maintaining station metadata are the wide variety of sensor and preamplifier elements that can be connected to a data logger, as well as the wide variety of data streams that can emerge from the logger. SIS manages these complexities with data stream and wiring templates. In wiring templates, users specify any combination of sensors and/or preamplifier elements that will connect to a logger. In data stream templates, users list the SEED channel names of the data streams and their decimation cascades used by the data logger to digitize the signal to a final sample rate. Then, users can assign different combinations of these templates to specific data loggers, and in turn describe which d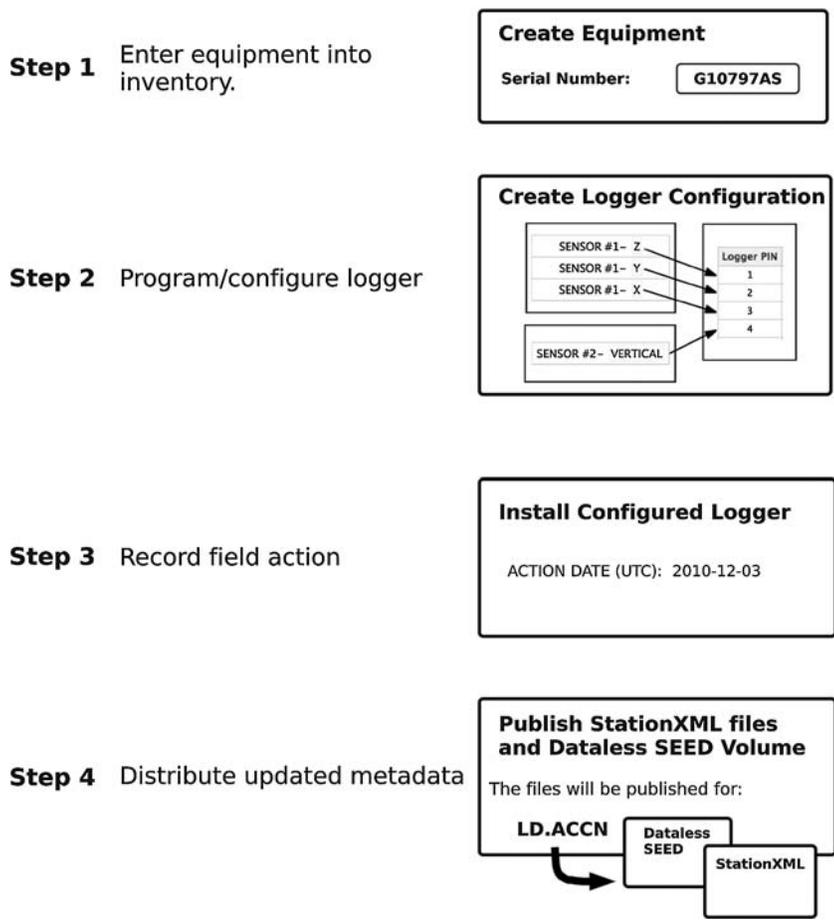ata streams will be produced by a logger. Users can reuse these combinations for common installation configurations (Fig. 6). For example, for temporary deployments for an aftershock sequence, SIS users can design these templates and assign them to digitizers allocated for this deployment ahead of time. This would allow the field crew to focus on field installation rather than metadata creation in such time critical situations.

SIS is not only able to handle single-site installations (digitizer and sensor at the same site), it can also describe multisite

the flexibility of storing important information that varies between equipment models in the form of key-value pairs. These key-value pairs can be time based (epochs) or held constant. For example, users can track firmware revision with loggers over time and voltage of batteries.

▲ **Figure 6.** With wiring and data stream templates, users are able to describe a variety of seismic installations in SIS. The user can create these templates using forms in the SIS user interface. For example, (a) the wiring template, combined with (b) the data stream template will describe the seismic installation and (c) the data streams produced. The first sensor (SENSOR #1) has three components measuring ground acceleration which is digitized by the logger at three sample rates (200, 100, and 1 samples/s). The second sensor (SENSOR #2) is a short-period instrument with one component measuring ground velocity and digitized by the logger at a single sample rate (100 samples/s).

Step 1 Enter equipment into inventory.

Create Equipment

Serial Number: G10797AS

Step 2 Program/configure logger

Create Logger Configuration

	Logger PIN
SENSOR #1– Z	1
SENSOR #1– Y	2
SENSOR #1– X	3
	4
SENSOR #2– VERTICAL	

Step 3 Record field action

Install Configured Logger

ACTION DATE (UTC): 2010-12-03

Step 4 Distribute updated metadata

Publish StationXML files and Dataless SEED Volume

The files will be published for:

LD.ACCN   Dataless SEED   StationXML

▲ **Figure 7.** Data entry in SIS is designed to follow the workflow of network operations. While the steps described above must occur in this order, they do not have to happen at the same time, nor do they need to be done by the same person. When a field action is recorded, SIS automatically translates this information into the creation, and/or the end of channel epochs with any appropriate changes to response and overall gain. As a result, the only action needed to produce dataless SEED after entering the field action is for the user to publish files for the site.

installations, where the logger is installed at a different physical site location than the sensors. In systems where only channel response is tracked, the site location of instruments is missing and may be erroneously inferred to be that of the FDSN station code.

USING SIS TO CREATE DATALESS SEED AND FDSN STATIONXML

Prior to SIS development, the process of translating a field action into updated dataless SEED files frequently required gathering information from several people and manually preparing the data, which made it time consuming and error prone. In SIS, the data entry steps required (entering equipment and response into inventory, configuration of the logger for what channels it will produce, and recording the field action in SIS) mirror the steps needed to prepare equipment for deployment and then installing it in the field. Then, the

creation of metadata is not a separate and distinct process but rather a result of the work already recorded in SIS. As a result, SIS users maintain station metadata by entering field actions.

When SIS users enter a field action such as swapping a sensor and decide to distribute the change, a new dataless SEED file is automatically created. SIS uses the wiring and data stream template information associated with the logger to determine if new data streams must be created, offlined, or changed. SIS also uses the response information associated with the equipment to calculate the sensitivity (Fig. 7).
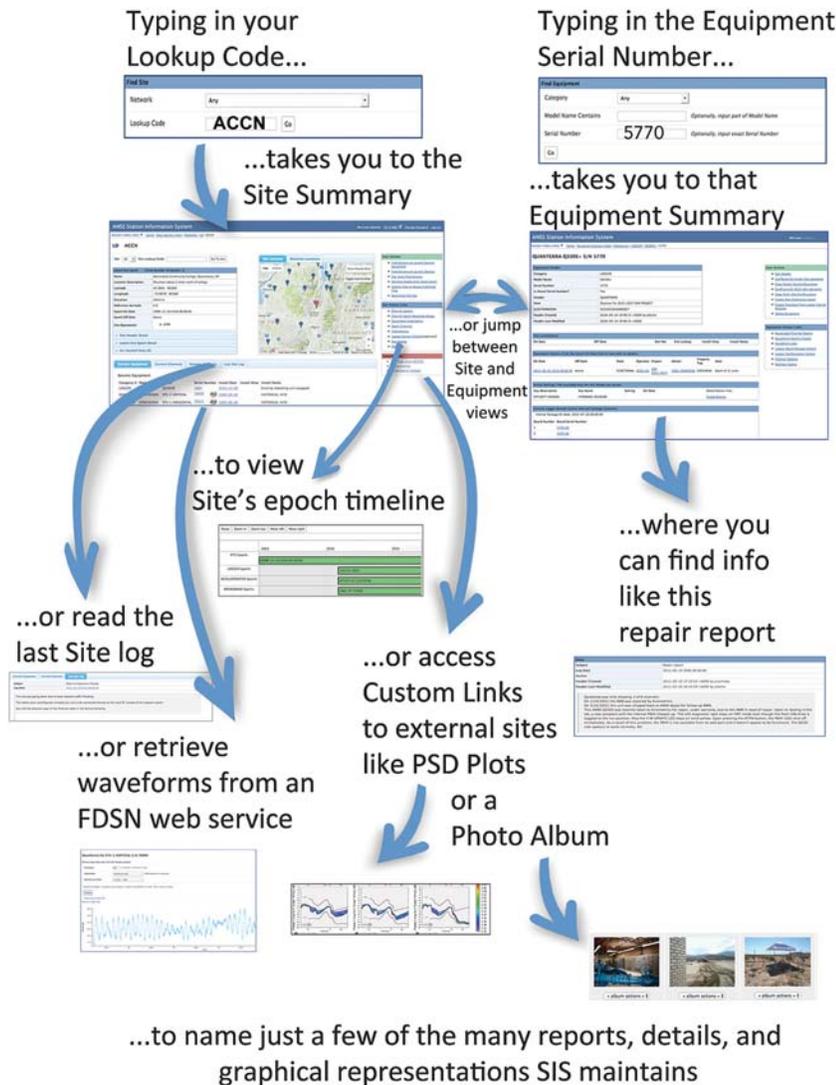
USING SIS FOR QUALITY CONTROL AND PLANNING

When data are stored and maintained in SIS, they become a powerful analysis tool for field technicians, analysts, and management. Network equipment and deployment is multifaceted and contains several many-to-many relationships. SIS provides a holistic view of network sites, their equipment, and the responses of equipment.

Users can easily navigate to the selected equipment information from the site summary page, and they can also access the sites that equipment has been installed at from the Equipment Summary page. Users can further drill down into details, view field equipment history, such as its ownership, repair status and logs, or equipment inventory, and site information such as site-visit logs, current and past deployed equipment, and what seismic data channels were recorded (Fig. 8). SIS also provides a level of data validation to prevent certain data entry errors. For example, users are not able to put in incorrectly formatted dates or install equipment if it is currently deployed elsewhere.

Although SIS uses a centralized database, it also allows users to create custom links that can go to URLs outside of SIS for further information. These links can appear on the site or equipment pages and can be generalized to appear on any site or equipment page according to user specified rules. This is useful for information for which access needs to be further restricted such as an external landowner database that has its own authentication, or to leverage existing efforts such as photo archives or webservices. Then, SIS users can easily access all relevant repositories regarding a seismic station and its equipment, even if that information is not stored in SIS. Furthermore, because SIS maps installed equipment to channel response, it can be integrated with FDSN webservices that use parameters such network, station, channel, and location code. For example, SIS users can view waveforms retrieved from the FDSN waveform webservice from the site summary

Typing in your Lookup Code...

...takes you to the Site Summary

...or jump between Site and Equipment views

...to view Site's epoch timeline

...or read the last Site log

...or retrieve waveforms from an FDSN web service

...or access Custom Links to external sites like PSD Plots or a Photo Album

Typing in the Equipment Serial Number...

...takes you to that Equipment Summary

...where you can find info like this repair report

...to name just a few of the many reports, details, and graphical representations SIS maintains

▲ **Figure 8.** Users can easily navigate through SIS and browse through site logs, channel responses, as well as equipment installation epochs. This graphic shows how a user might navigate through SIS. At the top, they would look up a station or an equipment serial number. (Middle) Then, this would allow them to look at various information associated with the lookup. (Bottom) Finally, they could look at various data or other station info.

page for a specific sensor with a couple of mouse clicks instead of the user having to learn the webservice application programming interface (API) and type in URL. This type of feedback will allow for better data quality control (Fig. 9).

DISCUSSION: FUTURE DIRECTIONS

There remain a number of areas of SIS development that would benefit network operators. One is to support describing telemetry connections between the equipment entered in SIS and the central processing site. Such work would help users to diagnose outages, data latency, and to further improve data completeness and quality. Further development in integrating SIS with outside repositories would capitalize on the growing use and standardization of webservices by the research community. More work is also needed to track and monitor the planning aspect of field work. Tasks such as scouting prospective sites, getting permits, selecting type of telemetry, and more happen before a site begins recording data. By capturing such information in SIS, network operators would be able to have a record of logistics associated with getting the site installed. Finally, the question of how tools such as the SIS UI and the resulting repository could be made available to a wider community beyond ANSS merits further discussion.

CONCLUSIONS

Seismic network operators must invest significant time, resources, and personnel in acquiring, deploying, and maintaining equipment. SIS enables the network operator and its data users to fully reap the value of these efforts, by tracking equipment inventory, deployment, and instrument response. The SIS UI saves valuable time for network operators by allowing them to easily manage and publish metadata without requiring operators to learn programming, database administration, or the intricacies of response cascades. Besides being a repository of an individual network's station metadata, SIS is also an archive of community knowledge on seismic instrumentation and field equipment. As an ANSS community resource, it provides standards in naming and channel response representation, which will not only benefit the network operators that produce seismic metadata and waveforms, but also the seismic research community at large.

DATA AND RESOURCES

The data used in this article are from Advanced National Seismic System (ANSS) Depot (https ://earthquake.usgs.gov/monitoring/anss/docs/ANSS_Equipment _Policy.pdf, last accessed November 2017), ExtStationXML format (https://anss-sis.scsn.org/xml/ext-stationxml/current/docs/ sis_extension.html, last accessed November 2017), Federation of Digital Seismic Networks (FDSN) StationXML format (https://www.fdsn.org/xml/station/, last accessed November 2017), Incorporated Research Institutions for Seismology (IRIS) Nominal Response Library (https://ds.iris.edu/NRL/, last accessed November 2017), IRIS Nominal Response Library Picker (https://github.com/crotwell/nrlBrowser, last accessed November 2017), IRIS FDSN StationXML-SEED Converter (https:// seiscode.iris.washington.edu/projects/stationxml-converter, last accessed November 2017), Standard for Exchange of Earthquake Data (SEED) format (https://www.fdsn.org/seed_manual/
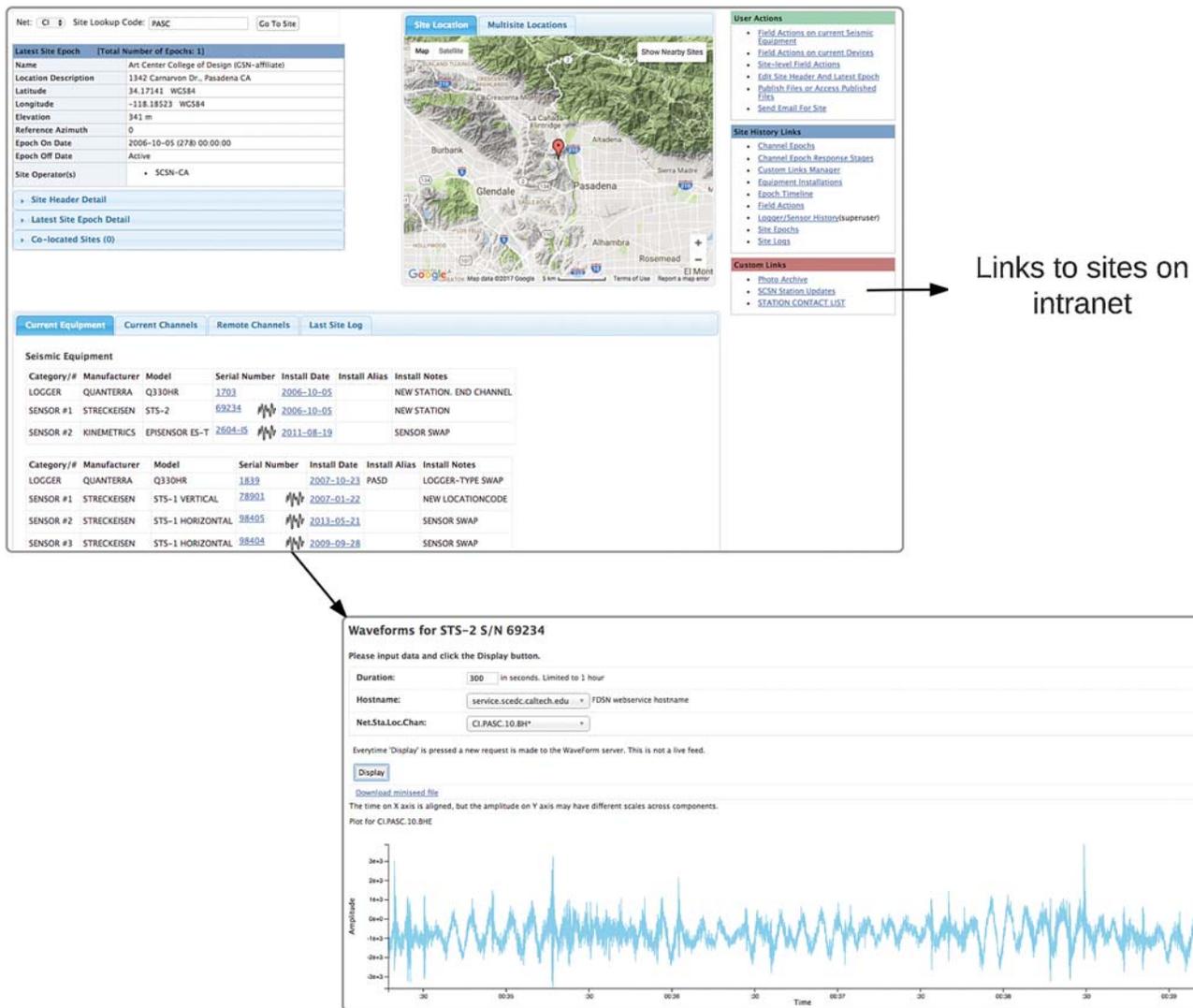
▲ **Figure 9.** SIS users can also store links on SIS pages which go to external repositories, such as a photo archive or landowner database. SIS also allows users to retrieve and display waveforms from FDSN webservices. In the figure, the user can easily see waveforms from a specific broadband sensor from a site where there are four (three single-component STS-1 sensors and one three-component STS-2 sensor) sensors. This is because SIS has already stored which equipment has produced which FDSN channel codes. Such connections would be difficult to determine with just the SEED channel codes. The data from the vertical-component STS-2 sensor at CI network station PASC (Art Center College of Design, California) are shown.

ACKNOWLEDGMENTS

REFERENCES

Ahern, T., R. Casey, D. Barnes, T. Knight, and C. Trabant (2009). *SEED Reference Manual, version 2.4*, available at http://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf (last accessed October 2017).

Bockholt, B. (2017). Instrument response and self-noise analysis of dataloggers at the Idaho National Laboratory, *Seismol. Res. Lett.* **88,** no. 1, 104–112.

Casey, R. (2012). Portable data collection center v3.9, *Data Services Newsletter* Vol. **14,** no. 3, Winter 2012, Incorporated Research Institutions for Seismology, https://ds.iris.edu/ds/newsletter/vol14/no3/ (last accessed October 2017).

Davis, P., and J. Berger (2012). Initial impact of global seismographic network quality initiative on metadata accuracy, *Seismol. Res. Lett.* **83,** no. 4, 697–703.

Filson, J., and W. Arabasz (2017). Origins of a National Seismic System in the United States, *Seismol. Res. Lett.* **88,** no. 1, 132–143.

Giles, J. (2011). Geoscience metadata—No pain, no gain, *Geol. Soc. Am. Spec. Pap.* **482,** 29–33.

Ringler, A. T., and J. R. Evans (2015). A quick SEED tutorial, *Seismol. Res. Lett.* **86,** no. 6, 1717–1725.

Ringler, A. T., C. R. Hutt, R. Aster, H. Bolton, L. S. Gee, and T. Storm (2012). Short note—Estimating pole-zero errors in GSN-IRIS/USGS network calibration metadata, *Bull. Seismol. Soc. Am.* **102,** no. 2, 836–841, doi: 10.1785/0120110195.

Templeton, M. (2017). *IRIS Library of Nominal Responses for Seismic Instruments*, Incorporated Research Institutions for Seismology, Dataset, doi: 10.17611/S7159Q.

Ellen Yu
Prabha Acharya
Justin Jaramillo
Sue Kientz
Egill Hauksson
California Institute of Technology
Seismological Laboratory
1200 E. California Boulevard, MS 252-21
Pasadena, California 91125-2100 U.S.A.
eyu@gps.caltech.edu

Valerie Thomas
U.S. Geological Survey
525 South Wilson Avenue
Pasadena, California 91106-3212 U.S.A.

Published Online 22 November 2017