Distributed Acoustic Sensing (DAS) Metadata Model

Rob Mellors[1], Kathleen M Hodgkinson[2], Voon Hui Lai[3],
and the DAS Research Coordination Network
Data Management Working Group

[1] University of California, San Diego
[2]Sandia National Laboratory
[3] Australian National University

Contact: rmellors@ucsd.edu

August 15, 2022

Members of the DAS RCN Metadata working group:
Jonathan Ajo-Franklin; Kent Anderson; Sandra Barajas; Paul Bodin; Jerry Carter; Luigia Cristiano; Ben Evans; Ge Jin; Meghan Miller; Helle Pedersen; Javier Quinteros; Nigel Rees; Diane Rivet; Jonathan Schaeffer; Nicole Taverna; Chad Trabant; Arantza Ugalde; Antonio Villaseñor; Jon Weers; Ray Willemann; Lesley Wyborn

Other contributors:
Nate Lindsey; Zack Spica; Andreas Wuestefeld

Summary

The goal of this whitepaper is to suggest a starting point for a common DAS metadata standard for archival purposes regardless of the data format ultimately used and to guide data collection at experiments. The intent is that this metadata data standard should be independent of the specific implementation and the emphasis is on content. We provide a suggested template for content and then evaluate using three scenarios based on existing DAS datasets, although the scenarios are nominal and intended for evaluation purposes only.

The metadata is divided into five major blocks: overview, cable and fiber, interrogator, acquisition, and channel. The information in each block consists of either text, integer, or floating points numbers and each item is specified as either required or recommended. Cable and fiber, interrogator, acquisition, and channel blocks may be repeated as necessary. The three datasets are DAS data collection from the SAFOD borehole, a surface deployment along a levee in Louisiana (NHERI) and a surface dataset from the Brady's geothermal area (POROTOMO). We conclude with suggestions for further work and next steps.

Table of Contents

Introduction

Distributed Acoustic Sensing (DAS) refers to a type of vibration sensing that uses an optical fiber as a continuous sensor (*Hartog*, 2000). In a typical implementation, laser pulses are sent into the fiber at one end and the reflected back-scattered light from along the fiber is measured over time to estimate the dynamic strain induced in the fiber by transient elastic waves, both scalar and vector (Figure 1).

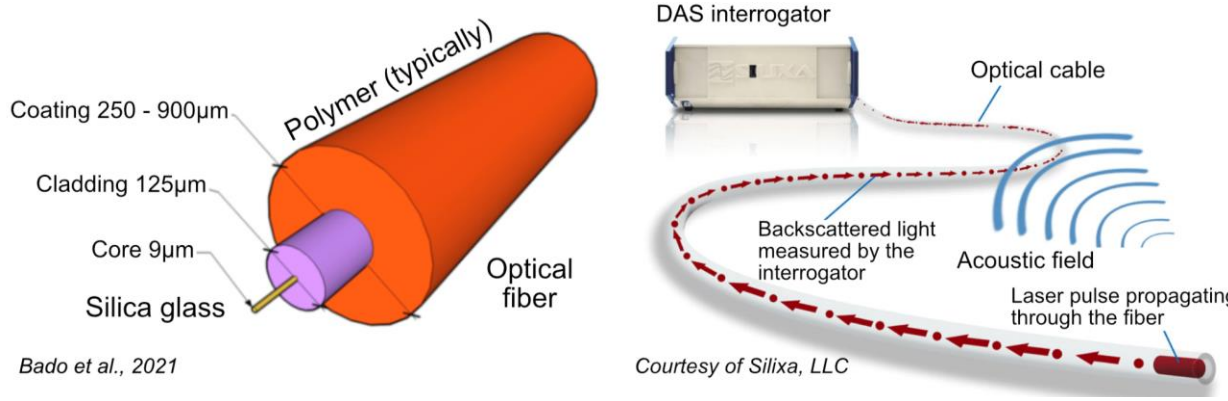

Figure 1. (Left) Schematic of an optical fiber. (Right) Diagram of an interrogator measuring an acoustic field.

As the wavefield is measured across discrete intervals (channels) along the fiber, each DAS acts as an array of sensors simultaneously (Figure 2). Each measurement represents the average strain (or strain-rate) along that length of fiber (the gauge length).

These sensors differ from traditional seismic sensors in several ways:

- They measure a directional component of strain, a tensor quantity, over a length rather than at a specific point.
- Many channels are collected simultaneously (up to 10,000's)
- Acquisition parameters and response are different from acquisition systems using inertial sensors.
- The installation environment can vary significantly along the length of the cable resulting in variations in sensitivity across the array.
- Data rates (1-10 mb/s) and resulting volumes (up to a Tb per day) are large.

As a result, standard seismic metadata (e.g., SEED) and file formats (e.g., SEGY) are a poor fit for DAS data. Standard formats cannot accommodate all the acquisition parameters needed for proper characterization and the large volumes overwhelm commonly used formats and archival systems. The goal of this white paper is to explore the requirements for DAS metadata and provide a set of recommendations and suggested content that will facilitate the re-use of these data sets in future research.

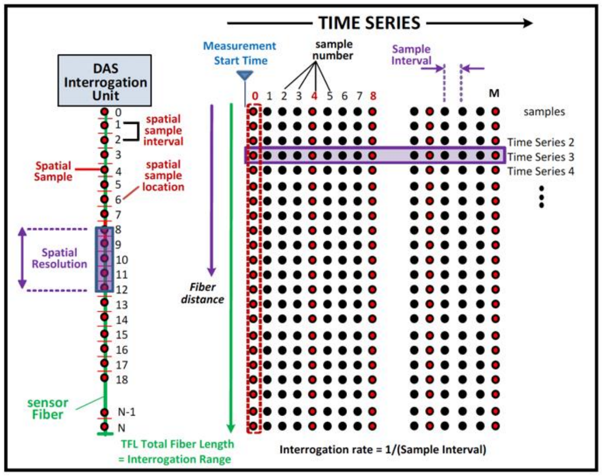

Figure 2. Diagram of data collection from a DAS fiber. From SEAFOM (2018). The time series recorded at each spatial sample form a data channel.

DAS data has been collected on a large scale for geophysical purposes for approximately ten years. Many of these datasets have been collected during hydrocarbon exploration and production (e.g., *Karrenbach et al*., 2019) and this has driven much of the development of geophysical DAS systems and analysis. In recent years, the collection of research grade datasets for DAS has increased. The applications span a range of uses and include subsurface monitoring, geotechnical assessment, ocean floor, and ice flow. New uses, such as earthquake early warning, are being explored and to fully maximize the potential of these datasets they should be made available to the research community in an open data repository.

In general, these research datasets fall into two main categories: 1) custom fiber deployments (*Wang et al*., 2018) and 2) data collection from existing but unused ('dark fiber') telecommunications fiber (*Martin*, 2017). Custom fiber deployments may be optimized for high quality data with engineered fiber and cables. Dark fiber relies on existing standard telecommunication fiber and may yield poorer quality data than custom deployments but require much less effort and cost to collect data.

Each data collection consists of two main components: the fiber (and associated cable) which serves as a sensing element and the electronics ("interrogator") located at one end which collects the data. The cable is typically fixed in place, either in a borehole, trench, conduit, or on the seafloor. The interrogator can be switched easily.

The cables and fibers vary in design depending on purpose and logistical constraints (*Soga and Luo*, 2018). The signal travels through an optical fiber, which is composed of silica glass coated with polymer, polyimide, or metal. The design of the optical fiber controls the type of propagation as well as other attributes such as polarization. As DAS relies on back-scattered signals, special engineered fiber may vary the scattering properties to enhance sensitivity. Standard telecom fiber is usually optimized for transparency. The fiber (or fibers) are enclosed in a cable. The cable construction depends on the requirements. Cables for borehole deployments may include a steel jacket as well as gel (or powder) for strain relief to avoid fiber breakage during installation. Surface and near-surface deployments may use Kevlar for rodent protection. Ocean floor cables typically include multiple fibers, steel jacket, and a core of copper fiber for power transmittal. One key item are connectors and splices, which may cause attenuation of the signal. Installation settings will also vary and could range from grouted cables in borehole, shallow burial for surface installations, or conduit for dark fiber. These variations in design and installation variations will change the coupling and response of the fiber to seismic waves. Attenuation along the fiber may be measured using an OTDR (optical time-domain reflectometry) but the frequency and amplitude of the fiber/cable itself is difficult to quantify or model exactly.

Interrogators and acquisition also vary. Most interrogators are commercial, and the exact details of the design and algorithm may be proprietary. The fundamental design may differ and hence the measured signal (strain or strain-rate) will be different, similar to the difference between traditional velocity or acceleration seismic sensors. While all interrogators use a laser pulse (or pulses), the repetition rate, length, and shape of the pulse can be varied for specific purposes, which will change the measurement length (gauge length) or the expected range.

As a result of the varied possibilities for both cable, interrogator, and acquisition, the associated metadata required for analysis of DAS data is, in general, more complex than for seismic time series data collected from traditional sensors. In addition, research grade seismic sensors have been carefully calibrated to allow estimation of the basic ground motions units to within 5 or 10 % in many cases (*Ringler et al*., 2012). This capability does not exist yet for DAS, Nominal conversion factors based on laser wavelength, gauge length, and refraction index may be available, and benchtop strain measurements (*SEAFOM*, 2018) are possible, considerable uncertainties exist in coupling and understanding of cable response to seismic waves in realistic settings. Most field calibrations rely on co-located seismic sensors (*Lindsey et al*., 2020) but the comparison between a point sensor measuring velocity or acceleration and the linear DAS strain measurement is not completely straightforward.

The level of maturity metadata associated with DAS data collects varies. One organization, SEAFOM, coordinates “industry recommended” practices and procedures for collection and calibration of DAS data but this work has not, as far as we know, been extended to standardized metadata. Common data formats for the data have been either industry proprietary, SEGY, or a Hierarchical Data Format (HDF). A current trend appears to be towards HDF, as SEGY requires extensive customization and is in general inadequate to hold all parameters. Cloud-optimized formats such as Zarr (https://zarr.readthedocs.io/en/stable/), designed with object storage in mind, offer a promising path for DAS data management as it enables researchers to utilize highly scalable on-demand compute resources available in the cloud.

The goal of this whitepaper is to suggest a starting point for a common metadata standard for archival purposes regardless of the data format ultimately used and to guide data collection at experiments. The intent is that this metadata data standard should be independent of the specific implementation and the emphasis is on content. We provide a suggested outline for content and then evaluate using several case studies. We conclude with suggestions for further work and next steps.

Requirements for metadata

The metadata requirements are:

- Accommodate most use cases (data collection scenarios)
- Permit cloud-based processing
- Allow pre-processing
- Accessible, i.e., the data should be contained in an open, non-proprietary format.
- Facilitate discoverability, i.e., the metadata and data should be easy to find and interpret by both humans and computers.
- Reusable, i.e., the metadata should be sufficient to allow the data to be re-analyzed in future research.

Metadata outline

We propose to divide the metadata into distinct groups.

- Overview metadata
    - Gives a high-level overview of the DAS deployment. This metadata group is intended to describe the geographic location, or region, of the installation and dates of the project. This metadata group is intended to facilitate discovery based on spatio-temporal searches.
- Cable and fiber metadata
    - Describes the cable environment and the fibers used within the cable. More than one cable may be used over the course of an experiment and there may be several fibers within a cable. The intention is to uniquely identify the fiber used to make the measurements.
- Interrogator metadata
    - Contains information about the interrogators used. More than one interrogator may be operating during an experiment. Each interrogator is described in one interrogator metadata block. The metadata consists of a unique identifier, the interrogator model, and the unit of measure.
- Acquisition metadata
    - Contains information on data collection, sample rate, pulse length, gauge length, channel spacing, and signal-processing steps.
- Channel metadata

- Describes each individual channel along the fiber, e.g., position, pulse direction, and location method. One possibility here is to have a pointer to a file(s) with a known format (e.g. kml, csv) rather than repeat all information.

In addition, we distinguish between required and recommended metadata. Required metadata is information that is considered essential to enable re-use of the data. This information makes the data self-describing and no further information required from the provider to work with the data. Recommended metadata is information that could be useful in interpreting the measured signal but is not essential. For example, gauge length is required metadata while the geographic coordinates of tap tests used to provide gauge positions is recommended.

Metadata Relationships

The various metadata categories are nested as shown below (Figure 3). In the simplest of examples there is one fiber within one cable connected to one interrogator. The suggested metadata model allows for much more complex installations with multiple fibers, cables, and interrogators

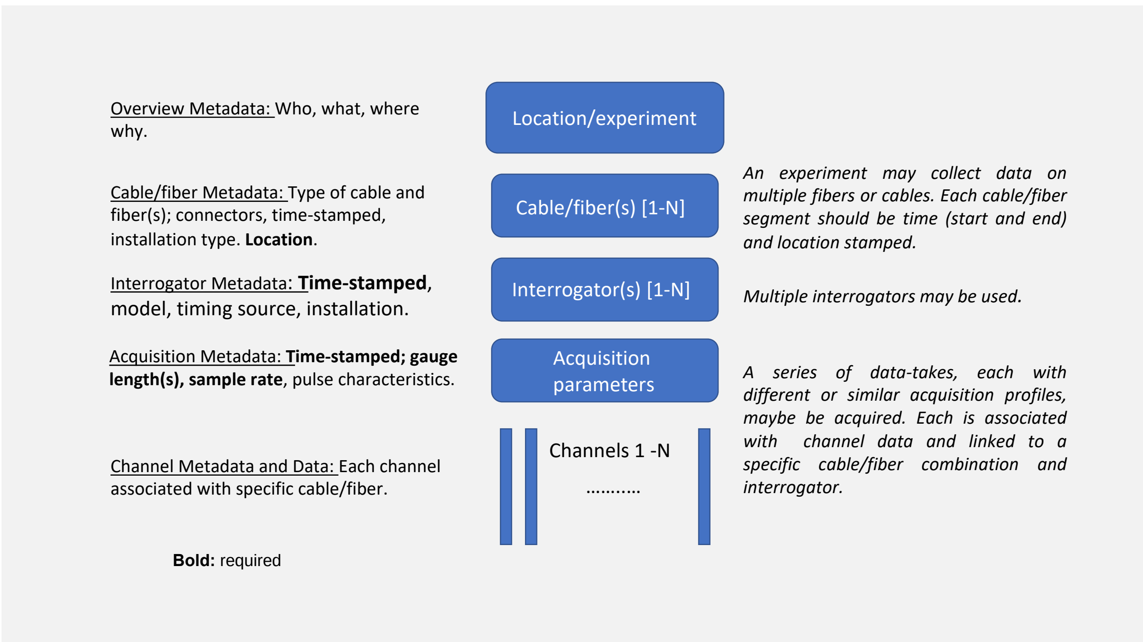

Figure 3. Schematic of metadata.

Metadata tables

Overview Metadata

The following are proposed as metadata in the overview category:

Overview Metadata	**Required?**	**Metadata Type**	**Description**
Location Description	Yes	Text	Description of geographic location e.g., Parkfield, California, USA
Deployment Type	Yes	Text	Permanent or Temporary
Network Name	Yes	Text	A network name should be managed by an organization accepted by the international community to avoid duplicating names
Data collection site name	Yes	Text	Name of data collection site e.g., SAFOD Borehole
Number of Interrogators	Yes	Integer	Number of interrogators used to collect data over the course of data collection
Principle Investigator(s)	Yes	Text	Point of Contact(s)
Start Date	Yes	ISO 8601	Start date of data collection, yyyy-mm-dd (UTC)
End Date	Yes	ISO 8601	End date of data collection, yyyy-mm-dd (UTC)
Purpose of data collection	No	Text	Brief explanation (1 or 2 sentences).
Data Collection	No	Text	Continuous or segmented
Digital Object Identifier	Yes	DOI link	An identifier that uniquely identifies the data, this identifier may only become available following discussion with a data repository.
Product ID	No	Integer	If this is a derivative product of the original data, it must have a Product ID.
Funding Agency	Yes	Text	Name(s) of agency that funded the data collection e.g. NASA, NSF, DOE

Project Number	No	Text	Funding project number e.g., EAR-9794799 . Should be supplied if a number has been assigned by funding agency(s).

Cable and Fiber Metadata

The following are proposed as metadata in the Cable category, more than one cable may be used over the course of an experiment and there may be several fibers within a cable. The intention is to uniquely identify the fiber used to make the measurements. One metadata block is used to describe each cable. The fiber ID should uniquely identify the fiber used. The identifier is user defined.

Cable and Fiber Metadata	Required?	Metadata Type	Units	Description
Fiber ID	Yes	Alphanumeric identifier		A term that identifies the fiber used within an experiment, e.g., alpha-numeric code defined by the researcher, default = 1
Start time	No	Date-Time	ISO 8601	Installation time of the cable
End time	No	Date-Time	ISO 8601	Removal of the cable
Cable characteristics	No	Text		example: tightly buffered, armored, gel-filled
Cable installation environment	No	Text		Examples: Conduit, trench, outside borehole casing, wireline
Cable model	No	Alphanumeric identifier		Cable manufacturer model
Cable diameter	No	float	meters	
Cable coordinates	No	float		List of cable coordinates. List should include a descriptor that indicates how positions were determined and a version number.

Connector coordinates	No	float	meters	List of coordinates of connectors (if any) along a cable. List should include descriptors for how positions were determined.
Fiber mode	No	Text		Example: single, multimode
Fiber refraction index	No		float	
Signal loss along fiber	No	float	dB/km	
Fiber geometry	No	Text		Linear or helical
Winding angle	No	float	degrees	With respect to cable axis
Fiber start location	No	float		
Fiber end location	No	float		if different from cable
Fiber length	No	float	meters	if different from cable
Comments	No	Text		

Interrogator Metadata

The following are proposed as metadata in the Interrogator category:

Interrogator Metadata	**Required?**	**Metadata Type**	**Description**
Interrogator ID	Yes	Alphanumeric	The Interrogator ID uniquely identifies the interrogator used within the network. User defined. Default = 1
Manufacturer	Yes	Text	Interrogator manufacturer e.g., Silixa
Model	Yes	Text	A model number or name that uniquely identifies the interrogator (e.g., iDAS, ODH4)
Interrogator unit	Yes	Text	Interrogators natural unit of measure: e.g.,

of measure			strain, strain-rate, velocity

Acquisition Metadata

The following are proposed as metadata in the Acquisition category:

Acquisition Metadata	Required?	Metadata Type	Units	Description
Acquisition ID	Yes	Alphanumeric identifier	Integer	User defined, one identifier per acquisition period, default = 1
Start time	Yes	date-Time	ISO 8601	UTC
End time	Yes	date-Time	ISO 8601	UTC
Acquisition sample rate	Yes	float	Hz (1/s)	The rate at which the interrogator provides output data
Pulse repetition rate	No	float	Hz (1/s)	Number of pulses per second
Interrogator rate	No	float	Hz (1/s)	Rate at which the Interrogator Unit interrogates the fiber (pulse repetition rate)
Pulse width	No	float	seconds	The width of the pulse, or burst of light, sent down the fiber
Gauge length	Yes	float	meter	The length along the fiber between a pair of pulses, determined at experiment setup
Number of channels	Yes	integer		total number of channels in archived data set
Channel spacing	Yes	float	meter	Spacing (m) between channels

Data sample rate	Yes	integer	Hz (1/s)	Sample rate of archived data set (this will equal the acquisition sample rate if there is no decimation)
Data unit of measure	Yes	text	strain / strain rate / velocity	Unit of measure of archived data set (e.g., strain, strain rate, velocity). This may be the same unit as the Interrogator Unit of Measure if the data are raw.
Decimation	Yes	Integer		If no decimation applied the default = 0
Time filtering	Yes	Text	Filter Type/frequency band	Filtering process applied to time series Example: Bandpass 1-100 HZ

Channel Metadata

The following are proposed as metadata in the Channel category:

Channel Metadata	**Required?**	**Metadata Type**	**Units**	**Description**
Name of associated file	No	Text		
File format of associated file	Yes, if associated file provided	Text		Example: KML, CSV (with headers), etc.
Channel ID	Yes	Integer		Uniquely identify channel
Geographic reference frame	Yes	Text		Coordinate system: e.g., WGS84, UTM
Distance along fiber	Yes	float	meters	Distance along fiber between the interrogator and channel.
Y-coordinate	Yes	float	Latitude/meters	Y position of channel within the defined

				reference frame
Y-coordinate error	No	float	meters	Uncertainty in Y location
X-Coordinate	Yes	float	Longitude/meters	X position of channel within the defined reference frame
X-Coordinate error	No	float	meters	Uncertainty in X location
Elevation [m]	No	float	meters	Height above Sea level in meters
Elevation error	No	float	meters	Uncertainty in elevation
Depth	Yes	float	meters	Depth (positive for depth below surface)
Depth error	No	float	meters	Uncertainty in elevation, depth below surface is positive??
Strike	No	float	degrees	Strike of channel, degrees clockwise of east positive
Strike error	No	float		
Dip	No	float	degrees	
Dip error	No	float		
Location method	Yes if X and Y coordinates are provided	Text		Example: Tap tests, GPS, survey
Direction of laser pulse	No	float	degrees	Increasing trace number

Use case scenarios

As a starting point, we describe several used case scenarios using recent datasets as examples:

- Borehole data collection on legacy fiber (SAFOD)
- Segmented active source recording of horizontal DAS array from the NHERI Levee Imaging Experiment
- Continuous recording of horizontal DAS array from the Porotomo Project

SAFOD – borehole installation

The SAFOD DAS experiment used an existing fiber installed in the San Andreas Fault Observatory at Depth (SAFOD) borehole.

Overview Metadata

Overview Metadata	Required?	Description
Location description	Yes	Parkfield
Deployment type	Yes	Permanent
Network name	Yes	DAS
Data collection site name	Yes	San Andreas Fault Observatory at Depth (SAFOD) borehole
Number of interrogators	Yes	1
Principal investigator	Yes	
Start Date	Yes	2017-06-01
End Date	Yes	2017-07-31
Purpose of data collection	No	Measure local earthquakes
Data collection	No	Segment
Digital Object Identifier	Yes	An identifier that uniquely identifies the data, this identifier may only become available following discussion with a data repository.
Product ID	Yes	If this is a derivative product of the original data, it must have a Product ID.
Funding Agency	No	Name(s) of agency that funded the data collection e.g. NASA, NSF, DOE

Project Number	No	

Cable and Fiber Metadata

Cable Metadata	Required?	Description
Fiber ID	Yes	1
Start time	Yes	2017152:00:00:00.00
End time	Yes	2017212:00:00:00.00
Cable characteristics	No	
Cable installation environment	No	Borehole
Cable model	No	-
Cable diameter	No	-
Cable coordinates	No	-
Connector coordinates	No	-
Fiber mode	No	-
Fiber refraction index	No	-
Signal loss along fiber	No	-
Fiber geometry	No	Linear
Winding angle	No	0
Fiber start location	No	0
Fiber end location	No	800
Fiber length	No	800
Comments		

Interrogator Metadata

Interrogator Metadata	Required?	Description
Interrogator ID	Yes	1
Manufacturer	Yes	Optasense
Model	Yes	ODH3.1
Unit of Measure	Yes	Counts

Acquisition Metadata

Acquisition Metadata	Required?	Units	Description
Acquisition ID	Yes	Integer	1
Start time	Yes	Time	UTC
End time	Yes	Time	UTC
Acquisition sample rate	Yes	Hz (1/s)	2500
Pulse repetition rate	No	Hz (1/s)	-
Interrogator rate	No	Hz (1/s)	
Pulse width	No	s	-
Gauge length	Yes	meter	10
Number of channels	Yes	integer	800
Channel spacing	Yes	meter	1
Data sample rate	Yes	Hz (1/s)	250
Data unit of measure	Yes	text	counts
Decimation	No	integer	
Time filtering	Yes	text	

Channel Metadata

Channel Metadata	Required?	Description
Channel ID	Yes	1
Geographic reference frame	Yes	
Distance along fiber	Yes	
Y-coordinate	Yes	Latitude (WGS84) / UTM coordinate
Y-coordinate error	No	
X-Coordinate	Yes	Longitude (WGS84) / UTM coordinate
X-Coordinate error	No	
Elevation [m]	No	
Elevation error	No	
Depth [m]	Yes	
Depth error	No	
Strike	No	
Strike error	No	
Dip	No	90
Dip error	No	
Location method	Yes if X and Y coordinates are provided	
Direction of laser pulse	No	Coordinates (azimuth)

NHERI Levee Imaging - Horizontal DAS data

The NHERI project deployed a fiber along a levee in Louisiana and conducted an active source seismic experiment.

Figure: Seismic source generated using vibroseis truck at various shot positions (round markers) recorded by a segment of horizontal DAS array buried in shallow trench (blue line) at Black Hawk Levee, Louisiana.

Overview metadata

Overview Metadata	Required	Description
Location description	Yes	Section of Mississippi River levee in Black Hawk, Louisiana
Deployment type	Yes	Permanent
Network name	Yes	
Data collection site name	Yes	Black Hawk Levee
Number of interrogators	Yes	1
Principal investigator	Yes	Brady Cox (Utah State University)
Start date	Yes	2021-10-22
End date	Yes	2021-10-22
Purpose of data collection	No	1. To demonstrate field DAS imaging on a section of Mississippi River levee for a workshop

		2. To assess the extent of internal erosion at Mississippi River levee and evaluate the use of DAS for such applications
Data collection	No	Segmented
Digital Object Identifier	Yes	https://doi.org/10.17603/ds2-c96x-pg70
Product ID	No	
Funding agency	Yes	Natural Hazards Engineering Research Infrastructure
Project number	No	2037900

Cable and Fiber Metadata

Cable and Fiber Metadata	**Required?**	**Description**
Fiber ID	Yes	1
Start time	No	-
End time	No	-
Cable characteristics	No	Outside plant (OSP) loose tube cable with corrugated steel armor and PFM™ gel-filled buffer
Cable installation Environment	No	Outdoor, buried in a shallow trench
Cable model	No	Loose Tube Single Jacket Single Armor Series 12
Cable diameter	No	
Connector coordinates	No	-
Fiber mode	No	single mode
Fiber refraction index	No	-
Signal loss along fiber	No	-
Fiber geometry	No	linear, order starts from south to north
Winding angle	No	-

Start Location	No	31.232360, -91.653479
End Location	No	31.233023, -91.654428
Fiber length	No	240 m
Comments	No	-

Interrogator Metadata

Interrogator Metadata	**Required?**	**Description**
Interrogator ID	Yes	1
Manufacturer	Yes	OptaSense
Model	Yes	ODH4+
Unit of measure	Yes	Radians relative to optical wavelength of 1550.12 nm

Acquisition Metadata

Acquisition Metadata	**Required?**	**Units**	**Description**
Acquisition ID	Yes	Integer	9
Start time	Yes	Time	10/22/2021 20:03:51
End time	Yes	Time	10/22/2021 20:04:09 (+18s)
Acquisition sample rate	Yes	Hz (1/s)	20 kHz
Pulse repetition rate	No	Hz (1/s)	-
Interrogator rate	No	Hz (1/s)	-
Pulse width	No	s	-
Gauge length	Yes	meter	2.04 m
Number of channels	Yes	integer	235
Channel spacing	Yes	meter	1.02 m
Decimation	No	Integer	20

Data sample rate	Yes	Hz (1/s)	1 kHz
Decimation	Yes	Integer	0
Time Filtering	Yes	Text	High-pass filtered above 3 Hz

Channel Metadata

Channel Metadata	Required?	Description
Channel ID	Yes	Trace number 692
Geographic reference frame	Yes	WGS84
Y-coordinate	Yes	31.232360
Y-coordinate error	No	-
X-Coordinate	Yes	-91.653479
X-Coordinate error	No	-
Elevation [m]	No	-
Elevation error	No	-
Depth [m]	Yes	-
Depth error	No	-
Strike	No	-
Strike error	No	-
Dip	No	-
Dip error	No	-
Location method	Yes if X and Y coordinates are provided	Google earth kmz file
Direction of laser pulse	No	South to north (increasing trace number)

Porotomo Project - Horizontal DAS array

The POROTOMO project collected active and passive seismic data from an 8 km custom fiber installation at Bradys geothermal field in Nevada. One section was deployed in a borehole.

Overview Metadata

Overview Metadata	Required?	Description
Location description	Yes	Brady's Geothermal Field
Deployment type	Yes	Permanent
Network name	Yes	DASH
DAS collection site name	Yes	Brady Geothermal area
Number of interrogators	Yes	4 (2 DAS units and 2 DTS units)
Principal investigators	Yes	Kurt Fiegl (University of Wisconsin)
Start date	Yes	3/8/16, paused, and then started again on 3/11/2016
End date	Yes	3/26/2016
Purpose of data collection	No	To assess an integrated technology for characterizing and monitoring changes in an enhanced geothermal system (EGS) reservoir in three dimensions with a spatial resolution better than 50 meters
Data collection	No	Continuous
Digital Object Identifier	Yes	https://doi.org/10.15121/1778858
Product ID	No	
Funding agency	Yes	USDOE Office of Energy Efficiency and Renewable Energy (EERE), Renewable Power Office. Geothermal Technologies Program (EE-4G)
Project number	No	DOE contract number EE0006760

Cable and Fiber Metadata

Cable and Fiber Metadata	Required?	Description

Fiber ID	Yes	1
Start time	No	-
End time	No	-
Cable characteristics	No	Tight-buffered
Cable installation environment	No	outdoor, trenched
Cable model	No	-
Cable diameter	No	-
Cable coordinates	No	-
Fiber mode	No	-
Fiber refraction index	No	-
Signal loss along fiber	No	-
Fiber geometry	No	linear
Winding angle	No	-
Fiber start location	No	-
Fiber end location	No	-
Fiber length	No	8400 m
Comments	No	-

Interrogator Metadata

Interrogator Metadata	**Required?**	**Description**
Interrogator ID	Yes	1
Manufacturer	Yes	Silixa
Model	Yes	iDAS
Unit of measure	Yes	Raw units (radians of optical phase change per time sample)

Acquisition Metadata

Acquisition Metadata	Required?	Units	Description
Acquisition ID	Yes	Integer	2
Start time	Yes	Time	2016-03-11
End time	Yes	Time	2016-03-26
Acquisition sample rate	Yes	Hz (1/s)	-
Pulse repetition rate	No	Hz (1/s)	-
Interrogator rate	No	Hz (1/s)	-
Pulse width	No	s	-
Gauge length	Yes	meter	10 m
Number of channels	Yes	integer	8721
Channel spacing	Yes	meter	1.021 m
Data sample rate	Yes	Hz (1/s)	1000 Hz
Decimation	Yes	Integer	-
Time filtering	Yes	Text	-

Channel Metadata

Channel Metadata	Required?	Description
Channel ID	Yes	431
Geographic reference frame	Yes	UTM Zone 11
Y-coordinate	Yes	327806.874
Y-coordinate error	No	-
X-Coordinate	Yes	4407447.965
X-Coordinate error	No	-
Elevation [m]	No	-

Elevation error	No	-
Depth [m]	Yes	-
Depth error	No	-
Strike	No	-
Strike error	No	-
Dip	No	-
Dip error	No	-
Location method	Yes if X and Y coordinates are provided	Tap test
Direction of laser pulse	No	-

Possible Future Extensions of the Format

Comments here on compression and edge computing at interrogators or data products (introduce basic idea of levels, but not necessary to specify products here- leave that to workshop in late 2022 early 2023)

Acknowledgements

Thanks to all who provided input. 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.

References

Becker, M. W., & Coleman, T. I. (2019). Distributed Acoustic Sensing of Strain at Earth Tide Frequencies. *Sensors, 19*(9), 1975. https://www.mdpi.com/1424-8220/19/9/1975

Hartog, A. H. (2000). Distributed Fiber-Optic Sensors: Principles and Applications. In K. T. V. Grattan & B. T. Meggitt (Eds.), *Optical Fiber Sensor Technology: Advanced Applications — Bragg Gratings and Distributed Sensors* (pp. 241-301). Boston, MA: Springer US.

Karrenbach, M., Cole, S., Ridge, A., Boone, K., Kahn, D., Rich, J., et al. (2019). Fiber-optic distributed acoustic sensing of microseismicity, strain and temperature during hydraulic fracturing. *GEOPHYSICS, 84*(1), D11-D23. https://library.seg.org/doi/abs/10.1190/geo2017-0396.1

Lindsey, N. J., Rademacher, H., & Ajo-Franklin, J. B. (2020). On the Broadband Instrument Response of Fiber-Optic DAS Arrays. *Journal of Geophysical Research: Solid Earth, 125*(2), e2019JB018145. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JB018145

Martin, E., (2018), Passive imaging and characterization of the subsurface with distributed acoustic sensing, Ph.D Thesis, Stanford University.

Ringler, A. T., Hutt, C. R., Aster, R., Bolton, H., Gee, L. S., & Storm, T. (2012). Estimating Pole–Zero Errors in GSN-IRIS/USGS Network Calibration Metadata. *Bulletin of the Seismological Society of America, 102*(2), 836-841. https://doi.org/10.1785/0120110195

SEAFOM. (2018). *DAS Parameter Definitions and Tests*. Retrieved from https://seafom.com/published-documents/

Soga, K, and L. Luo (2018) Distributed fiber optics sensors for civil engineering infrastructure sensing, Journal of Structural Integrity and Maintenance, 3:1, 1-21, DOI: 10.1080/24705314.2018.1426138

Wang, H. F., Zeng, X., Miller, D. E., Fratta, D., Feigl, K. L., Thurber, C. H., & Mellors, R. J. (2018). Ground motion response to an ML 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. *Geophysical Journal International, 213*(3), 2020-2036. https://doi.org/10.1093/gji/ggy102

Appendix A. Glossary

Cable. Contains the fiber(s) as well as a protective jacket.

Channel. Each time series that corresponds to the signal for a specific length of fiber (gauge). Distance between channels is often less than the gauge length and therefore signals from two adjacent channels are not independent.

Channel spacing. Spatial distance measured along the fiber between channels. Differs from the gauge length.

Phase coherent. Interrogator that returns a phase that is linear with applied strain.

Phase incoherent. Interrogator that returns a phase that is non-linear with applied strain.

Engineered fiber. Fiber that has been custom fabricated to enhance signal-to-noise for DAS. Variety of methods are possible including enhanced Rayleigh back-scatter, either by doping or inducing artifacts, or by periodic variations in back-scatter. Often increases attenuation.

Fiber. Optical glass 'wire' that includes core, cladding and coating. Exact geometry and design may vary depending on desired properties (e.g., single-mode, polarization maintaining, or photonic crystals).

Fiber length or distance. The maximum distance from the interrogator to the end of the terminated end of the fiber.

Gauge length. The length over which the signal is averaged to improve signal-to-noise. May be set as an acquisition parameter or may be generated in post-processing with some interrogators.

Helical fiber. Fiber wound in a spiral and intended to enhance azimuthal response.

Interrogator. The box containing the electronics that produces the laser pulse(s) and records the returning back-scatter. Design varies with vendor and details are usually proprietary.

Interrogation Rate (or pulse rate): The rate at which the Interrogator Unit interrogates the fiber sensor.

Pulse. Burst of laser light into the fiber. Pulse may be carved into various shapes.

Pulse width. The length of the pulse, equivalent to the velocity of the pulse in the fiber multiplied by the length of time the light is pulsed.

Tap Test. A test to identify the physical location of buried fiber. This is often done by tapping the ground with a heavy weight and then identifying the specific channel at that location. Other techniques, such as using the noise from a moving vehicle with known position, are possible.

Strain. Extension per unit length, in this context the change in gauge length divided by the original gauge length, unitless.

Vibration isolation plate. Device useful for some interrogators to reduce contamination of recorded signal along fiber caused by vibration of the interrogator itself.

Appendix B. Useful links

https://seafom.com/
https://www.energistics.org
https://www.energistics.org/portfolio/prodml-data-standards/
http://docs.energistics.org/EO_Resources/Worked_Example_DAS_v1.0.pptx

https://seismic-data.org

Appendix C. Comments

A summary of comments on the poster presentation based on this white paper at 2022 Seismological Society of America meeting, Seattle, WA:

- Important parameters for using a DAS dataset:
  - Method of photonic estimation (dual-pulse, single-pulse, chirp, local oscillator) or at least require the category of photonic estimation (quantitative, non-quantitative)
  - SEAFOM standard of noise level (e.g., noise level estimate in rad/rt-hz at 1, 5,10,50 km)
- Less important, but still relevant:
  - provenance of location estimation (why, how, when, who)
  - Dark fiber vs direct install
  - Fiber owner
  - Fiber operator
  - OTDR for array (w provenance as fibers do change)

EGU comments

(related) Ways to transfer large amounts of data?
Add trace start time to channel metadata? [but would this also imply sample rate and number of samples?

Other comments

Timing could be other than GPS (e.g., NTP or PTP).
Could add a timing metric for segments where timing lock was missing.
Ownership of the cable, location of the first repeater, depth of water (marine or lake)
Use a pointer to a file containing locations rather than repeat (implemented in July 25 version)
Add metadata version number
Add “user-defined” space