

Initial steps towards a DAS metadata model

Rob Mellors¹, Kathleen M Hodgkinson², Voon Hui Lai³, and the DAS RCN Data Management Working Group⁴

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

The number of Distributed Acoustic Sensing (DAS) research datasets is increasing rapidly and cover a range of applications, from ice movement to subsurface monitoring to earthquake hazard. As the number of available interrogators increases, it is expected that the need for archiving of the datasets will also increase. Long term archiving faces two main challenges at present: the need for large amounts (100's TB) of storage and a standardized metadata format. As part of the DAS Research Coordination Network (RCN), a DAS data management working group is constructing a metadata model for DAS data. The objective is to develop a common metadata standard for archival purposes and to guide data collection at experiments. The metadata requirements include: 1) accommodation of most use cases (data collection scenarios); 2) permitting of cloud-based processing; 3) allowing of pre-processing; and 4) reduction of the burden of data transport. We propose to divide the metadata into distinct groups consisting of overview (basic information about the experiment), acquisition (type of interrogator and acquisition parameters), and cable and fiber descriptions. A subset will be denoted as "required" metadata, and standard metadata principles, such as findability, accessibility, interoperability, reusability (FAIR), and machine-readability, will be adhered to. The purpose of this presentation is to inform potential users of these efforts, encourage adoption of the proposed standard, and invite community input.

Metadata Requirements

- 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.

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. It also describes the provenance, or origin, of the data set.

Cable and fiber metadata

Describes 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.

Overview Metadata: Who, what, where why.

Location/experiment

Cable/fiber(s) [1...N]

Cable/fiber(s) [1...N]

Interrogator(s) [1...N]

Interrogator(s) [1...N]

Acquisition Metadata: Time-stamped; gauge length(s), sample rate, pulse characteristics.

Acquisition parameters

Channel Metadata and Data: Time and location stamped. Each channel associated with specific cable/fiber.

Channels 1 - N

Figure 1. Schematic of metadata

This research was funded by the U.S. Department of Energy. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Objective: What are the minimum technical specifications required for DAS data? (and how might they be organized?)

Overview

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 that is accepted by the international community to avoid duplicate DAS project names
Data collection site name	Yes	Text	Name of data collection site e.g., SAFOB Borehole
Principle Investigator(s)	Yes	Text	Point of Contact
Start Date	Yes	ISO 8601	Start date of data collection (UTC)
End Date	Yes	ISO 8601	End date of data collection, (UTC)
Purpose of data collection	No	Text	Brief explanation (1 or 2 sentences).
Data Collection	No	Text	Continuous or segmented
Number of Interrogators	Yes	Integer	Number of interrogators used to collect data over the course of data collection
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., USGS, NSF
Project Number	No	Text	Funding project number e.g., EAR-9794799. Should be supplied if a number has been assigned by funding agency(s).

Channel

Channel Metadata	Required	Metadata Type	Units	Description
Channel ID	Yes	Integer		Uniquely identify channel
Geographic Reference Frame	Yes	Text		Coordinate system: e.g., WGS84, UTM
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	degrees	Uncertainty in strike
Dip	No	float	degrees	Orientation of channel wrt horizontal
Dip Error	No	float	degrees	Uncertainty in dip
Location Method	Yes if X and Y are provided	Text		Example: Tap tests, GPS, survey
Direction of Laser Pulse	No	float	degrees	Azimuth

Cable and fiber

Cable and Fiber Metadata	Required	Metadata Type	Units	Description
Fiber ID	Yes	Alphanumeric		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		Cable manufacturer model
Cable diameter	No	float	meters	
Connector Coordinates	No	float	meters	List of coordinates of connectors (if any) along a cable.
Fiber Mode	No	Text		Example: single, multimode
Fiber refraction index	No	float		A measure of the speed of light in the material
Signal loss along fiber	No	float	dB/km	Example: 15 dB/km
Fiber Geometry	No	Text		Linear or helical
Winding Angle	No	float	degrees	With respect to cable axis
Fiber Start Location	Yes	float		
Fiber End Location	Yes	float		
Fiber Length	No	float	meters	
Comments	No	Text		

Interrogator

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)
Unit of Measure	Yes	Text	Radians relative to optical wavelength of 1550.12 nm

Acquisition

Acquisition Metadata	Required	Metadata Type	Units	Description
Acquisition ID	Yes	Alphanumeric	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	
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		Default = 0 if no decimation
Time Filtering	Yes	Text	Filter Type/ frequency band	Filtering process applied to time series Example: Bandpass 1-100 Hz

Example Test Case

NHERI Levee - Horizontal DAS data collected using active source

We use a segmented active source recording of horizontal DAS array from the NHERI Levee Imaging Experiment as a test case for the proposed schema.

Figure 2. 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. Two km of fiber optic cable had been installed in both surface and borehole arrays as a means to image the levee at high spatial resolution and monitor seepage. (Cox, B., 2021, Distributed Acoustic Sensing (DAS) data from the NHERI@UTexas Thumper Shaker Truck and 240m of Fiber Along the Blackhawk Levee)



Cable and fiber

Cable and Fiber Metadata	Required	Description
Fiber ID	Yes	1
Start time	No	-
End time	No	-
Deployment type	Yes	Permanent
Network name	Yes	e.g., MRL (Mississippi River Levee)
Data collection site name	Yes	Black Hawk Levee
Number of interrogators	Yes	1
Principle Investigator(s)	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

Interrogator

Interrogator Metadata	Required	Description

<tbl_r