SnT 2025 8 SEPTEMBER ONLINE DAY 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

A Federation of Online Legacy Data in Seismology (FOLDS)

Tim Ahern[1], Lorraine Hwang[2], Joseph Batllo[3], Javier Quinteros[4]
[1]EarthScope, emeritus [2]UC Davis, [3]ICGC, [4]GFZ

INTRODUCTION AND MAIN RESULTS

FOLDS is a system that will enable seismological data centers with legacy data collections to create, curate and distribute **standardized** metadata and image files from analog seismic recordings. The system will promote data reuse following FAIR principles, leveraging a new suite of International Federation of Digital Seismograph Networks (FDSN) web services similar in capability to the current FDSN services supporting digital timeseries. These services will allow access to metadata and digital representations of analog recordings by Internet through a federation of worldwide FOLDS data centers. This is a distributed approach to generate metadata as well as distributing digital images and metadata. Data can be replicated in multiple centers if needed. The FDSN has approved 19 Required, 25 Recommended, and 12 Optional elements of metadata. FAIRness considerations were identified for Required elements. Several geographically distributed legacy data centers have agreed to participate as beta testers of the system. • Proposal due in December 2025. • If funded, development likely begins in 2026.

SnT 2025 8 SEPTEMBER ONLINE DAY 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

A Federation of Online Legacy Data in Seismology (FOLDS)

Tim Ahern, Lorraine Hwang, Josep Batlló, Javier Quinteros

P2.4-145

FOLDS Metadata Fields

Metadata that had been collected by several legacy data preservation efforts were identified. There were significant differences in the various contributor's metadata. The resulting 39 metadata elements were reviewed by participants attending a workshop in Albuquerque in September 2019.

Next, surveys of the worldwide community were conducted to determine if the proposed elements should be REQUIRED, RECOMMENDED, or be OPTIONAL. Additional fields could be recommended. An FDSN Technical Review Committee spent months refining the results and included some fields suggested in the surveys. This resulted in a final list of 56 proposed fields.

The community process resulted in 19 fields **REQUIRED**, 25 fields **RECOMMENDED**, and 12 fields **OPTIONAL**. The general classes of metadata fell into 7 Element Types with 56 elements

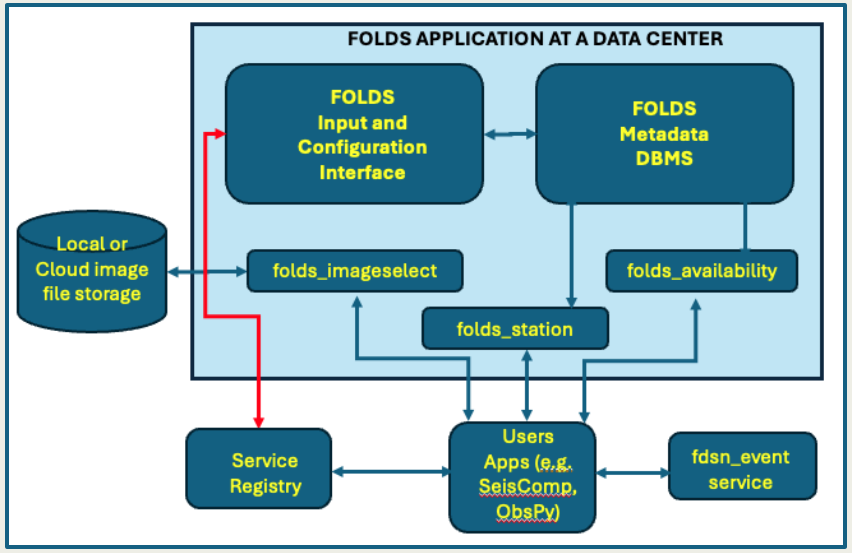
Element Type	NUMBER OF ELEMENTS	Required	Recommended	Optional
TIME	3	2	1	0
STATION/CHANNEL	12	5	3	4
SENSOR	8	6	0	2
RECORDING SYSTEM	4	1	2	1
TESEO	6	0	6	0
IMAGE FILE	20	5	10	5
ADDITIONAL	3	0	3	0
TOTAL	56	19	25	12

FOLDS metadata has been accepted as an FDSN standard and available for use now.

FOLDS Application Design

The FOLDS application will run at each legacy data centre and will have the following components:

- Entry of site-specific configuration parameters
- Metadata entry by GUI or by XML ingest
- A DBMS to store site specific FOLDS metadata
- 3 web services to expose the data centre's metadata
  - folds-availability: a catalogue of holdings
  - folds-station: selected metadata in XML format
  - folds-imageselect: the actual image files
- Image files stored on local disk or in cloud storage
- An external service registry will show services available from the various FOLDS centres.



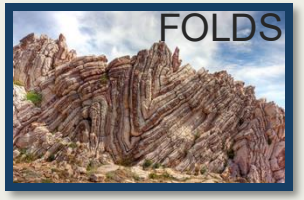
FOLDS FAIRness and Access

FOLDS was built to ensure that the system supports FAIR access to information. For all metadata fields that are REQUIRED, FAIRness is assessed as shown on the next slide containing the FOLDS metadata fields.

Effort was taken to manage the amount of work that will be required at a centre using the FOLDS system. This was done by only REQUIRING entry of 19 fields per image. Most REQUIRED fields can be auto-entered based on the fields in the previous image. REQUIRED fields can always be used for search and when present so can RECOMMENDED fields.

FOLDS Status

- A proposal is now in preparation with anticipated submission to the US National Science Foundation in December 2025
- Proposal funding is anticipated in mid-late 2026
- Initial Beta test available mid to late 2027
- There are ~15 identified data centers that have agreed to participate in the FOLDS beta test
  - These data centres are from Asia, Europe, North America, & South America



SnT 2025 CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

A Federation of Online Legacy Data in Seismology (FOLDS)

Tim Ahern, Lorraine Hwang, Josep Batlló, Javier Quinteros

P2.4-145

FOLDS Metadata Fields

Metadata	Metadata Element	Metadata Element Description		F	A	I	R	Req	Rec	Opt
Time of Data	3							19	25	12
	Start Time	the time of the first sample in the image	YYYYMMDDTHH:MM:SS.FFFF format	Y	Y	Y	Y	Y		
	End Time	the time of the last sample in the image	YYYYMMDDTHH:MM:SS.FFFF format	Y	Y	Y	Y	Y		
	Time Correction	any time correction applied to the data	YYYYMMDDTHH:MM:SS.FFFF format						Y	
Channel Details	12									
	Latitude	latitude using WGS84 datum	SEED format convention	Y	Y	Y	Y	Y		
	Longitude	longitude using WGS84 datum	SEED format convention	Y	Y	Y	Y	Y		
	Elevation	Elevation above (+) or below (-) sea level	real in meters						Y	
	Depth of sensor below ground surface	depth below ground surface at specified	real in meters							Y
	Network Name	network to which the station belongs (e.g. WWSSN, GSN, EREBUS)	text						Y	
	FDSN Network Code	FDSN network code- Earliest FDSN Code in use for the station (use SS if not associated with a network)	text						Y	
	Site Name	site name (e.g. Albuquerque, New Mexico, USA)	text	Y	N	N	N	Y		
	IR Station Code	station's code in the International Registry	text	Y	Y	Y	N	Y		
	Channel/component	channel code as in SEED format	text as in SEED Manual Appendix A	Y	Y	Y	Y	Y		
	Open Date	date when station was opened	YYYYMMDDTHH:MM:SS.FFFF format							Y
	Close Date	if closed, Date when station was closed. Leave empty if still operating or not known	YYYYMMDDTHH:MM:SS.FFFF format							Y
	FDSN Time Series Identifier	Proposed new FDSN Time series identifier	see https://docs.fdsn.org/_/downloads/source-identifiers/en/v1.0/pdf/							Y
Sensor	8									
	Type of sensor	type of sensing instrument (e.g. Streckheisen STS-2, Ewing, Beniof)	text	Y	Y	Y	N	Y		
	Sensor serial number	manufacturer's serial number of	text							Y
	Galvo Free period	the free period of the instrument	real in seconds	N	N	Y	Y	Y		
	Galvo Damping constant	the instrument's damping constant	real dimensionless	N	N	Y	Y	Y		
	Horizontal 1 dip/azimuth	the dip/azimuth of the first horizontal	SEED convention	Y	Y	Y	Y	Y		
	Horizontal 2 dip/azimuth	the dip/azimuth of the second horizontal	SEED convention	Y	Y	Y	Y	Y		
	Vertical dip/azimuth	the dip/azimuth of the vertical channel	SEED convention	Y	Y	Y	Y	Y		
	Nature of instrument	Mechanical (e.g., Wiechert) or electromagnetic (e.g., Golitsyn).	text							Y
Recording System	4	Teseo. Some of these can be calculated if you know the paper size. See p. 18-19 of manual								
	Type of recording system	type of recording system (e.g. Teledyne helicorder)	text	Y	N	N	N	Y		
	Recording system serial number	manufacturers serial number if known	text							Y
	Scale/gain/amplification	scale or gain factor (scaler)	real dimensionless						Y	
	Period of scale/gain	period at which the gain is valid	real in seconds						Y	

Metadata	Metadata Element	Metadata Element Description	see TESEO spec https://github.com/INGV/teseo2/blob/master/Documentation/Teseo2UserManual/Teseo2UserManual.md	F	A	I	R	Req	Rec	Opt
Parameters for drum recorders	6									
	Paper speed	paper speed (linear velocity of paper)	real in meters/s						Y	
	R	length of the writing arm, from its rotating axis to the tip of the needle	real in meters/s						Y	
	r	radius of the drive cylinder bearing the smoked paper	real in meters						Y	
	a	distance from the rotating arm axis to the driving cylinder axis	real in meters						Y	
	b	shift of the arm axis to the base line on the smoked paper	real in meters						Y	
	d	length of 1 minute on paper	real in meters						Y	
Image file details	20									
	DOI of scanned Image	Enter the DOI if one has been assigned	DOI format	Y	Y	Y	Y		Y	
	Date of Scanning	the data the image was scanned	YYYYMMDDTHH:MM:SS.FFFF format							Y
	Resolution	the resolution of the scanned image	pixels per meter	Y	Y	Y	Y	Y		
	Vertical pixels	the number of pixels in the vertical dimension	number of pixels						Y	
	Horizontal pixels	the number of pixels in the horizontal	number of pixels						Y	
	Image format	image file type	e.g. heic,jpeg, jpeg-2000, openEXR, pdf, png, tiff	Y	Y	Y	Y	Y		
	Image size	the total size of the image in bytes	integer						Y	
	Analog image length	length of the original document	real in meters						Y	
	Analog image width	width of the original document	real in meters						Y	
	Color depth	the color depth of the scanner if applicable	integer						Y	
	Phase Markings present	Indicate if phase notations were placed in the image	text (Y or N)							Y
	Associated Bulletin	Earthquake phases are present on image. Phases were reported to a bulletin or otherwise publlished.	text						Y	
	Occlusions	Indicate true if tears or other flaws obscure trace data	text (True or false)							Y
	Earthquake signal	present	text (True or false)							Y
	Timemark Format	Positive real to indicate length of vertically offset timemarks, negative real to indicate length of gapped timemarks, null to indicate no timemarks	real in m							Y
	Polarity of recording	Ground motion up = up on paper or down on paper	Text (up or down)						Y	
	Original recording type	Photographic paper, drum recordings (smoke, hot stylus, ink)	text	Y	Y	N	Y	Y		
	Location of original record	Country, state or province, city, institution, room of original analog document when scanned	text	N	N	N	N	Y		
	Vectorized_trace	The trace in this image has been vectorized.	text (Y or N)	Y	N	N	N	Y		
	Contact information of owner	Contact information for the original owner of the data	text						Y	
Additional	3									
	Notes and Comments	Optional Notes and/or comments	text						Y	
	Source of information	Information about source of metadata entered - e.g., lat/lon adopted from a published source, or response assumed based on X information or publication.	text						Y	
	Date of metadata creation	Date and time when the metadata was created or last updated	YYYYMMDDTHH:MM:SS.FFFF format						Y