

# FEDERATION OF ONLINE LEGACY DATA IN SEISMOLOGY

## FOLDS

A SYSTEM TO ACQUIRE, MANAGE, AND DISTRIBUTE METADATA & DIGITAL VERSIONS OF LEGACY SEISMIC DATA

---

TIM AHERN, IRIS (NOW EARTHSCOPE) EMERITUS

LORRAINE HWANG, UC DAVIS (CIG)



## 2

# OUTCOME OF 2021 FDSN VIRTUAL MEETING

---

- WG 2 formed a Proposal Review Team after the 2021 IASPEI meeting as part of the FDSN Framework Process
- The Review Team consisted of
  - Central Asia, Inna Sokolova, Institute of Geophysical Research, KNDC (Now GSRAS, Obninsk)
  - North America Bob Engdahl, U. of Colorado, Boulder
  - Europe, Josep Batllo, Institut Cartogràfic i Geològic de Catalunya

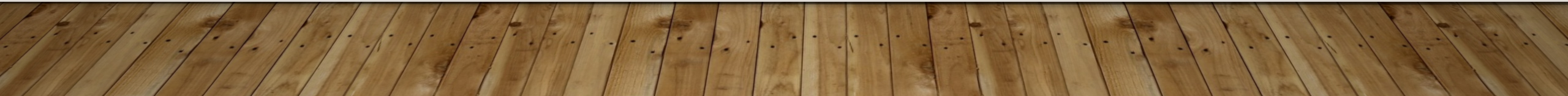


# 3

## RESULTS OF THE PROPOSAL REVIEW

---

- The Review Team Identified the key elements of metadata that should be collected in Legacy Data preservation projects drawn from community surveys that had been collected previously
- An analysis of the metadata elements resulted in two outcomes
  - Identification of metadata elements that should be Required, Recommended, Optional, or Not Included
  - Metadata elements deemed required that make the data FAIR
- The Review team reviewed 7 categories of metadata that were in the community surveys
  - Data Timing Metadata
  - Sensor Metadata
  - TESEO Metadata for drum recorders
  - ADDITIONAL Metadata
  - Station Channel Metadata
  - Recording System Metadata
  - Image File Metadata



# TIME OF DATA METADATA FIELDS

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A	Comments
Time of Data	3										
	Start Time	the time of the first sample in the image	Y	Y	Y	Y	Y				
	End Time	the time of the last sample in the image	Y	Y	Y	Y	Y				
	Time Correction	any time correction applied to the data						Y		Y	

Note that the FAIRness of **REQUIRED** metadata is assigned as well

Findable  
 Accessible  
 Interoperable  
 Reusable

**Findable**  
**Accessible**  
**Interoperable**  
**Reproducible**  
**Required**  
**Recommended**  
**Optional**  
**Omitted**

**Required** – parameters must be available and can be *used for searching*. These fields are NOT OPTIONAL. Metadata not containing these elements will not be included in the federated system

**Recommended** – metadata should be available and *when available used in searching*. Strictly speaking Recommended fields are NOT REQUIRED but ENCOURAGED

**Optional** – totally optional if information is available. Will *not be used for search* but will be returned to user when available

**N/A** – fields that have been suggested but the process advises not to include them in the system



# STATION CHANNEL INFORMATION

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A	Comments
Station Channel Details	12										
	Latitude	latitude using WGS84 datum	Y	Y	Y	Y	Y				Also IR station data
	Longitude	longitude using WGS84 datum	Y	Y	Y	Y	Y				Also IR station data
	Elevation	Elevation above (+) or below (-) sea level						Y			Also IR station data
	Depth of sensor below ground surface	depth below ground surface at specified longitude							Y		
	Network Name	network to which the station belongs (e.g. WWSSN, GSN, EREBUS)						Y			
	FDSN Network Code	FDSN network code- Earliest FDSN Code in use for the station (use SS if not associated with a network)						Y			
	Site Name	site name (e.g. Albuquerque, New Mexico, USA )	Y	N	N	N	Y				Also IR station data
	IR Station Code	station's code in the International Registry (ISC)	Y	Y	Y	N	Y				
	Channel/component	channel code as in SEED format	Y	Y	Y	Y	Y				
	Open Date	date when station was opened							Y		
	Close Date	if closed, Date when station was closed. Leave empty if still operating or not known							Y		
	FDSN Time Series Identifier	Proposed new FDSN Time series identifier							Y		Maybe too complex for old seismograms

# SENSOR METADATA

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A	Comments
Sensor	8										
	Type of sensor	type of sensing instrument (e.g. Streckheisen STS-2, Ewing, Beniof)	Y	Y	Y	N	Y				
	Sensor serial number	manufacturer's serial number of seismometer if							Y		
	Galvo Free period	the free period of the instrument	N	N	Y	Y	Y			Y	It may be not available
	Galvo Damping constant	the instrument's damping constant	N	N	Y	Y	Y			Y	It may be not available
	Horizontal 1 dip/azimuth	the dip/azimuth of the first horizontal	Y	Y	Y	Y	Y				Default 0
	Horizontal 2 dip/azimuth	the dip/azimuth of the second horizontal	Y	Y	Y	Y	Y				Default 0
	Vertical dip/azimuth	the dip/azimuth of the vertical channel	Y	Y	Y	Y	Y				
	Nature of instrument	Mechanical (e.g., Wiechert) or electromagnetic (e.g., Golitsyn).							Y		



# RECORDING SYSTEM METADATA

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A
Recording System	4	Many of these are parameters used in 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)	Y	N	N	N	Y			
	Recording system serial number	manufacturers serial number if known							Y	
	Scale/gain/amplification	scale or gain factor (scaler)						Y		Y
	Period of scale/gain	period at which the gain is valid						Y		Y

# TESEO PARAMETERS

---

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A
TESEO Parameters for drum recorders	6									
	Paper speed	paper speed (linear velocity of paper)						Y		
	R	length of the writing arm, from its rotating axis to the tip of the needle						Y		
	r	radius of the drive cylinder bearing the smoked paper						Y		
	a	distance from the rotating arm axis to the driving cylinder axis						Y		
	b	shift of the arm axis to the base line on the smoked paper						Y		
	d	length of 1 minute on paper						Y		

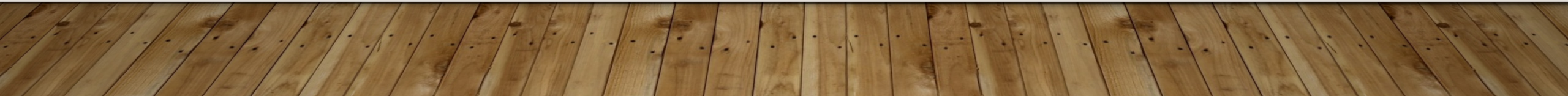


# IMAGE FILE METADATA

Metadata	Metadata Element	Metadata Element Description	F	A	I	R	Req	Rec	Opt	N/A
Image file details	20									
	DOI of scanned Image	Enter the DOI if one has been assigned	Y	Y	Y	Y		Y		
	Date of Scanning	the data the image was scanned							Y	
	Resolution	the resolution of the scanned image	Y	Y	Y	Y	Y			
	Vertical pixels	the number of pixels in the vertical dimension						Y		
	Horizontal pixels	the number of pixels in the horizontal dimension						Y		
	Image format	image file type	Y	Y	Y	Y	Y			
	Image size	the total size of the image in bytes						Y		
	Analog image length	length of the original document						Y		
	Analog image width	width of the original document						Y		
	Color depth	the color depth of the scanner if applicable						Y		
	Phase Markings present	Indicate if phase notations were placed in the image							Y	
	Associated Bulletin	Earthquake phases are present on image. Phases were reported to a bulletin or otherwise published.						Y		
	Occlusions	Indicate true if tears or other flaws obscure trace data							Y	
	Earthquake signal	Indicate true if an earthquake signal is present							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							Y	
	Polarity of recording	Ground motion up = up on paper or down on paper						Y		
	Original recording type	Photographic paper, drum recordings (smoke, hot stylus, ink)	Y	Y	N	Y	Y			
	Location of original record	Country, state or province, city, institution, room of original analog document when scanned	N	N	N	N	Y			

# 10 AN APPLICATION IS NEEDED TO TO CAPTURE 51 ELEMENTS

- 18 REQUIRED elements present legacy data centers with a manageable task
  - Many of the elements can either be generated automatically (e.g., date of metadata creation)
    - If processing in a station sorted order, most fields do not change from record to record
  - Such considerations reduce the level of effort considerably
- 21 RECOMMENDED elements, ~9 of which would require manual input
- 12 OPTIONAL elements, ~5 of which would require manual input
- **Only ~14 metadata fields would require manual entry for a single image**
  - **In the future, some of these could be determined via algorithmic processes (AI)**





# II FEDERATION OF ONLINE LEGACY DATA IN SEISMOLOGY

---

- **FOLDS** will be an application to
  - Capture/generate metadata at legacy data centers
  - Curate the metadata as needed
  - Expose the metadata and digital artifacts by
    - query metadata database using Web Services
    - Providing unique identifiers to digital artifacts (e.g., scanned images)
    - Distribute metadata and image files via web services
- Legacy Data Centers continue to manage and distribute metadata and artifacts from their local data center as part of a federated system
- This effort does not include the actual digitizing of the original data such as paper records but will accommodate earlier digitizing efforts

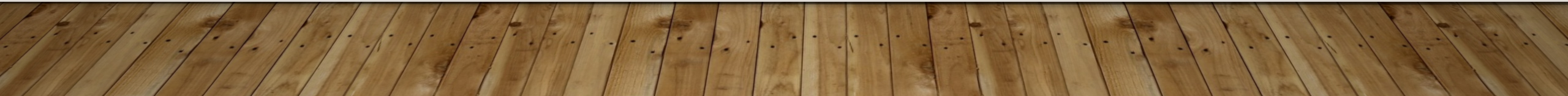
# 12 INTELLIGENT WORKFLOWS

---

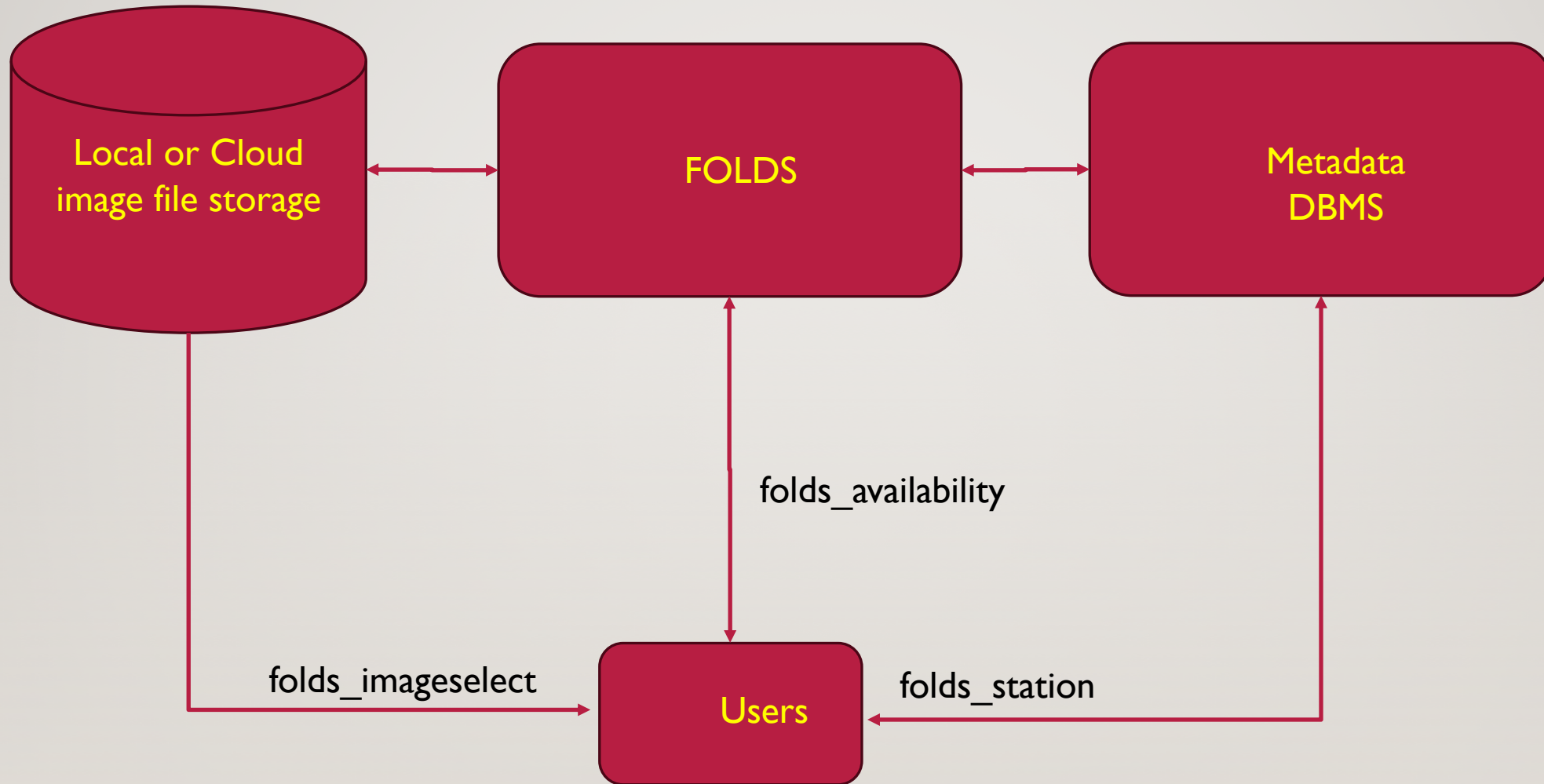
- For each image, auto-entry of all metadata that do not change between consecutive images will automatically fill fields for next image.
  - Network Codes, Channel IDs, sensors used, station coordinates, site names,
  - Ensure fields fall within acceptable limits before committing to FOLDS DBMS
- Dealing with either station sorted images or time/event sorted images allows adaption to how digital images are managed locally
- **Enforce** entry of **Required** metadata fields and make it **easy** to include Recommended and Optional metadata fields



- Development of APIs based on Web Services
  - **folds\_availability**: availability of a digital artifact matching a user's request based on Required and populated Recommended query parameters. (similar to **fdsn\_availability** service)
    - Optional information can be returned if user requests this but optional fields would not be queryable
  - **folds\_station**: providing metadata elements meeting a user's query (similar to **fdsn\_station** service)
  - **folds\_imageselect**: capable of recovering
    - a thumbnail of the image or
    - the actual digital image files, meeting search criteria (similar to **fdsn\_dataselect**)
  - Make use of **fdsn\_event** at existing Federated Data Centers

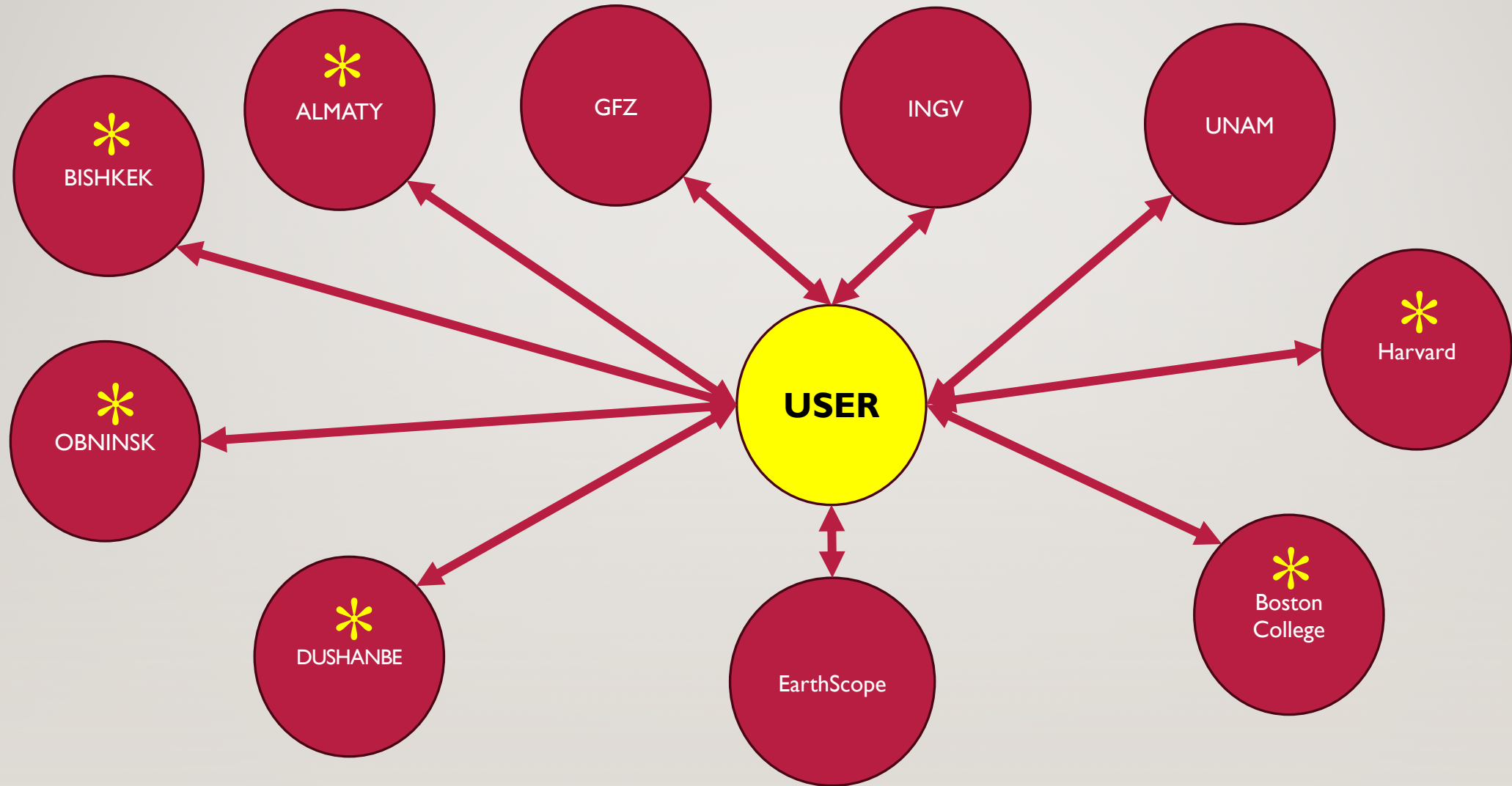


# FOLDS WILL BE A FEDERATED AND DISTRIBUTED SYSTEM





# A FEDERATED SYSTEM FOR LEGACY DATA



# NEXT STEPS

---

- Ask WG II to form a group of 3-5 FDSN members to:
  - Help develop/review the FOLDS specification
  - Review progress periodically
  - Participate in periodic project status updates
  - Act as initial beta testers
- Seek funding
  - Initial steps have already been taken
- Select Contract Team
- Seek Partners
  - Identify potential testers at multiple Legacy Data Centers as shown in previous slide \*
- Promote the system and provide training when it becomes available
- Optimistic Timeframe
  - Funding 6 months
  - Beta Application 12-18 months
  - Working system 2025 IASPEI meeting



- The Technical Review Community devoted a great deal of time to this effort and did an excellent and thorough job.
  - Thanks to Inna, Josep, and Bob
- Thanks to Javier for shepherding this through the Framework process in WG2
- Thank you for your attention

