Review: FDSN Framework Proposal for a Metadata Standard for Legacy Seismic Data

Review Team: Bob Engdahl (University of Colorado, Boulder), Josep Batllo (Institut Cartogràfic i Geològic de Catalunya), Inna Sokolova (Institute of Geophysical Research, KNDC)

A proposal has been submitted to WG2 of FDSN by Tim Ahern that attempts to standardize the metadata needed to support an internationally coordinated collection system of digital legacy seismic data elements generated from analog recordings. The product of this effort will be the development of an FDSN standard set of metadata elements and associated documentation. These standards will make easier the work of those at collection centers creating the metadata and at data centers that manage metadata and corresponding legacy data. However, to manage the level of effort required for data collection centers to participate, enough metadata is required to make legacy data Findable, Accessible, Interoperable and Reusable (FAIR).

A proposal Review Team has been assembled to recommend a final list of Required, Recommended, and Optional metadata elements. For this evaluation the Team needs to consider a comprehensive international survey of metadata categories and elements, keeping in mind the FAIR requirements and the ramifications of having too many or too few parameters. The following documents have been made available to the Team to assist in the proposal evaluation.

- FDSN Framework Proposal for a Metadata Standard for Legacy Seismic Data
- Workshop report on Securing Legacy Seismic Data to Enable Future Discoveries
- Final Pre-Workshop Legacy Data metadata list
- New Elements Legacy Data metadata list
- FDSN WGII Legacy power point presentation

There follows the report of the proposal Review Team.

The two metadata lists in spread sheet format provided to us were combined into a single list for review. This list was reconfigured by replacing the right side of the list, after the column for Metadata Description, with labeled columns to be used for the review of each element. These include four columns (F, A, I, R) for each of the

FAIR attributes, three columns to indicate whether an element was Required (Req), Recommended (Rec), or Optional (Opt), one column to indicate if an element is Not Always Available (N/A), and a final column for Comments. Each member of the Team was responsible for independently assigning a Yes (Y) or No (N) for each element in these columns where applicable. The three lists were compared, differences resolved, and a final list adopted as shown in Table 1.

General overview

First, the Review Team welcomes the proposal as submitted and believes that it is an important step in facilitating the use of analog seismograms. When the last of those people involved in the daily recording and processing of analog seismograms anywhere go into retirement, it is important to introduce methods that allow these legacy seismograms to be managed by present day information systems.

Second, the Review Team considered all the proposed elements as valuable to identify/characterize legacy seismograms and none were discarded, but it would be necessary for the Team to classify the elements into different levels of importance. Finally, some duplications among elements were noticed and some elements could be obtained from others. We'll discuss these issues in the following paragraphs.

Use of the metadata

The proposed metadata should be applied to legacy seismograms, most likely seismograms for which a scanned image exists (i.e., raster images made up of pixels). But it is not necessary for the traces they contain to be converted from a raster image into a vector graphics format (vectorized). The usual situation is that they will not be. The proposed metadata set can be mainly used for two different tasks: to locate a seismogram and/or to have enough information to vectorize the traces/waveforms and get them into mseed or similar formats. For the first task the most common procedure is that one searches for a date/time and place. In this case the searched record can be characterized by a time window, a site, and the place where it can be found now. For this search it is not necessary to have the records scanned but inventoried following the defined rules. In this case, just a handful of elements are needed. For the second task it is necessary to get the records (the scanned images) and a set of parameters characterizing the transfer function of the recording seismograph and physical characteristics (size, record speed, etc.) of the record itself. In this case, the number of necessary parameters/elements increases.

The Review Team analyzed/discussed the different elements keeping in mind the final objective is to get enough information to vectorize traces.

How to define FAIR elements

It was a decision of the Review Team to analyze the FAIR characteristics only for elements we identified as required. We found that this was not always easy and, in some cases, maybe not even be the best approach. For example, Latitude (using WGS84 datum – Table 1, line 7) is clearly a FAIR element. However, "Galvo Free period" (Table 1, line 22) may or may not be a FAIR element, because it depends on the availability of other factors like Galvo damping, etc., to properly recover the seismograph transfer function. More details are given in the Conclusions section.

Block discussions

Time of data

This block is composed of three elements. "*Start time*" and "*End time*" are fundamental when searching for seismograms in a time window. Thus, we propose that these elements are all "req". Some remarks are important for its practical use. The proposed format in the FDSN Framework Proposal asks for a precision of up to a tenth of second. This does not seem realistic. The first and last minute marks appearing on an analog seismogram are usually annotated on it (but not always). However, determination of the first and last recorded second implies, in general, specific calculations. These factors should be considered when defining the exact properties and format of the start and end time. We suggest a flexibility for these elements allowing different accuracies, from just the day (it may be useful for old seismograms) to a tenth of a second if desired. It is worth noting that in the case of most analog seismograms from USSR stations, the time course (time increment) was from right to left, but there were also some stations with recordings where the time course was from the left to right. There are also seismograms characterized by local time/UTC. For example, Italian seismograms recorded before WWII were almost all recorded on local time. Even the bulletins were published in local time. A final point is that there are stations (not many) for which seismograms do not show GMT time. This probably should be noted somewhere. By default, this parameter should be "00", but for the stations operated with local or regional time it will be, for example, GMT+3. As rule this information is not shown on historic seismograms. So, without knowing this factor one cannot correctly determine the

time on the seismogram. Of course, the number of such stations is small, but their seismograms can be very important.

The "*time correction*" is another important factor. Its proper definition requires two values: the time correction itself and the moment it was observed/calculated (hour and minute of the day). In some places it was not calculated every day. The proposed format for it is (nn.nnn) seconds. At some observatories the time correction may be several minutes, and the format of the defined element should allow that to be noted. The last point is that the time correction is not always annotated on the seismogram (for old seismograms never) and this information may have been lost for different reasons. This is the main reason why we suggest this element to be assigned "Rec" instead of required. We also suggest that in this case (and similar others) a "not always available" (N/A) indication may be introduced as a response. This tells the user that the element may not have been available when the metadata were created and, in this case, they cannot use the seismogram for absolute time determination.

Station Channel Details
Twelve elements are included in this section. The Review Team considered just five of them as "Req". "*Latitude*" and "*Longitude*" identify the place the seismogram was recorded and allows spatial searches to be made.

"*Site name* and *IR*" - The International Registry (IR) of station codes maintained by the ISC is a second way to identify the recording site. Let's remember that Latitude and Longitude are parameters associated with the station code in the IR. Thus, there is some duplicity. But we recognize it is easier for the user to directly have latitude and longitude than to search for them in the IR. Also, it helps to satisfy the FAIR principles. Moreover, the IR contains a significative number of mismatches (among them, stations that changed coordinates and this is not reflected in the IR, as well as coordinates not compliant with the WGS84 datum). It may happen that a collection of analog seismograms exists for a station never registered at the IR. In this case we suggest that the first step, before introducing the metadata of the seismograms into the system, would be to register a code for the station in the IR.

We propose the "*site name*" as "Req" element and, somehow, it seems redundant. Even this element may not be necessary for FAIR compliance. Curators of the seismograms know this, and it is easy to introduce the site name in the system. Moreover, it could be helpful when trying to locate complementary material (bulletins, publications, station books, etc.) for the seismogram. Let's remember

that the present station code system was introduced in 1963, when ISS reorganized as ISC, and the present stations codes are not found in materials prior to that time (e.g., French stations are identified with a two-letter code in bulletins of the fifties, and for Leopolis/Lemberg/Lvov/Lviv sites the last two codes are "true" and should be written in Cyrillic characters and referred to the present station code as LVV).

"*Channel/Component*" may be properly managed using the present SEED format. Sometimes it may be difficult to enter the characteristics of old records with those in use today (HH, HN, etc.). Some remarks in a "User's Manual" may help. But there is a problem to consider. Analog seismograms may contain the record of more than one component. How should we proceed in this case? An easy way is to introduce a different set of metadata for each channel where all sets point to the same image. If the decision is made that an image should have a unique set of metadata, then the SEED format should be modified. This discussion extends to elements on lines 24, 25 and 26.

Sensor
Eight elements are included in this category and the Review Team recommends considering six as required.

The "*Type of sensor*", as now introduced (text format) helps the user in the same sense as "site name".

We recommended as required the "*Galvo Free Period*" (line 22 in Table I) and "*Galvo Damping constant*" (line 22) just to indicate the need for parameters and elements to properly assess the transfer function of the instrument. In this section there are elements for the galvo; but nothing to characterize the transductor (maybe at least free period and damping?). The element on line 78 of the table is clearer than lines 22 and 23. Certainly, element 78 will be more complex than 22 and 23, but a similar element exists in SEED format.

Elements at lines 24, 25 and 26, about channel orientation are like those existing in SEED and do not need comment. But we point to the comments about several components in one seismogram presented with the channel/component element (line 15).

Recording System
In this case we just recommend the "*recording system*" element as "Req".
We have some doubts on the "*Scale/gain/amplification*" and "*Period of Scale/gain*". Do they refer to the recording system (helicorder in many cases) or do

they refer to the whole system? In the last case they should be referred together with the system transfer function. In the first case its value may not be available in many cases (again a N/A entry is suggested).

TESEO Parameters for drum recorders
None of the presented elements was selected as "Req". Of course, their availability helps ease the task of waveform correction. But in many cases, they are not easy to recover as the original instruments do not exist anymore. Also, personal experience shows that, after some iterations, it is not difficult to guess them.

Image file details
This block contains many elements. These elements are related to the characteristics of the raster image (we think this fact is not stated as such but looking at the proposed elements it is clear), but also with those of the original physical seismogram and its contents. We selected five of them as required. The first two are the "*resolution*" of the image and the "*image file format*". When "*resolution*" (more precisely, the dots per inch -dpi- density) is known it is possible to obtain the dimensions of the original image and any other desired dimension (length of the minute, wave amplitudes, etc.). "*Image format*" is also important and considered as "req". It may be the case, depending on the characteristics of the designed system, that the search engine returns to the user the name of the file with the proper file extension identifying the image format. If this is the case, this element will be "implicitly" managed by the metadata system, and it may be skipped.

Two more "req" elements are the "*original recording type*" and the "*location of the original record*". It will be good for the user to know about the physical nature of the record (smoked, photographic, ink, thermal, etc.) but, as we point in the comment's column, it is even more important to know about the registration type (displacement, velocity, acceleration or any other). This last point is important in case there are any doubts about applying an instrument transfer function to a vectorized trace. The FDSN should evaluate the possibility of redefining or unfolding this element. "*Location of original record*" should help the user when serious doubts or supplementary details on the original arise. This element is related to the "*contact information of the owner*".

A last required element is the "*vectorized trace*". We included it because it is good to know if this work has already been done and because (as it will not be a common case) it is not making extra work for the people in charge of introducing metadata into the system if it is defined as "No" by default.

Additional

This group contains (excepting lines 62, 63 and 64) elements suggested by users participating in the survey. None has been selected as required. Some of them can be considered duplications of previous elements or very similar to them. This fact is pointed out in the comment's column of Table I and, in these cases, they have not been evaluated.

Nevertheless, element in line 78, "*sensor*" deserves a detailed comment. As pointed out when dealing with elements in lines 22 and 23 (Galvo free period and Galvo damping constant), several elements in the proposed list are devoted, at least in part, to define the characteristics of the transfer function of the seismograph that recorded the original image. Also, the element in line 80 can be included among them. But it seems that the present set of elements does not offer a clear way to introduce and properly define the seismograph transfer function. The Review Team has been discussing this item as we think it is one of the key points to make the proposed system useful (see the last to entries of the Appendix with scripts of discussions). The FDSN WGII should devote time to a careful analysis of the problem and get a definitive solution.

Conclusions, recommendations, and final remarks

The Review Team concludes and accepts that the system of metadata applied to analog seismograms described in this proposal is an important FDSN WG asset.

The Review Team has evaluated each element in the spread sheets provided and summarized its findings in this report and in Table 1 (separate attachment). The elements were classified as follows: 15 Required (Req); 27 Recommended (Rec), 5 of which were considered not always available (N/A); and 16 Optional (Opt). Classification of the FAIR attributes was determined only for Required elements. Arguments by the Team for the classifications of each element can be found throughout the report.

The final number of "required" elements selected is driven by two divergent positions. From the point of view of curators (those in charge of the original seismograms and, most likely, of their raster images) and those in charge of introducing the metadata in the system, it is clear a reduced set of elements will simplify the task. From the point of view of users, they would like to have the most complete information about the seismograms they want to use. Users would prefer

to get it directly from the metadata parameters and from characteristics of the seismograms they can otherwise obtain from a reduced set of elements (unless available supporting software automates these tasks, as in the case of SEED format).

It may seem that number of required/recommended elements is quite high. But it is necessary to consider that, when introducing metadata sets in the system, the most common case will be to document/introduce a set of seismograms from the same seismic station, even the same instrument. In such cases many elements will be "fixed" for tens/hundreds of seismograms and only time, component, and maybe some other elements will change from one to the next.

Another item to consider is that in the period from the end of WWII to the IGY (International Geophysical Year – 1957/58) the procedures and instruments at seismic observatories were standardized almost everywhere. Thus, seismograms from the fifties onwards share many common elements and are not difficult to fit into standard procedures (elements in the proposed system). Going backwards, diversity increases, and, in the early years of seismic observation, it is possible to find a counter example for any proposed standard. It is also true that these situations involve a reduced number of seismograms, but they may contain information on relevant earthquakes and other events.

On the one hand, it has not been difficult to evaluate/prioritize the elements. On the other hand, the Review Team notes that the present definition of some of the elements may be considered blurred (e. g., the time correction or the general definition of the transfer function). More work should be devoted to a proper definition and use of the metadata elements. Specifically, identification of the type of sensor (mechanical, electromagnetic, strainmeter, etc.) and how the transfer function of the seismograph is represented in the metadata set is not clear at this time and needs further elaboration. Most likely, this last case will lead to some rearrangement of the proposed elements (there will be new elements and others will disappear).

Evaluation of the FAIR characteristics of the elements posed some difficulties. Nevertheless, as a conclusion, if the objective is locating seismograms, a minimum set of elements composed by *Start time, End time, Latitude, Longitude, and IR Station code* should be compliant with the FAIR principles. If the objective is to obtain digital waveforms from the analog seismograms, the *transfer function* of the recording seismograph (introduced as a unique element or as several ones) and the

resolution of the image (or equivalent parameters) should be also a minimum compliant set.

An optimum for the proposed system is that all images supported have a DOI or similar digital identifier. At present, many institutions do not have easy access to DOI codes. Maybe the FDSN may improve this access as it does with the citation of seismic networks. Certainly, this case involves millions of DOI's and maybe it is not possible to manage it from the FDSN. In any case, this is a topic to evaluate and search for available alternatives.

As previously pointed out the definition of some elements leaves doubts. More work should be devoted to their precise definition. It is true that it is not always possible to define elements with a few written words (right now) in a table. For this reason, a reference manual, with clear description of all elements and different possible cases ought to be written as a necessary complement to the proposed metadata system. Seismograms dating from the fifties of the XX Century onwards fit easily in the proposed categories/elements, but earlier seismograms are diverse and in many instances it may be difficult to fit their characteristics into the defined elements. For this reason, the reference manual should give detailed descriptions of different cases and we suggest illustrating it with examples including images of real seismograms.

There also were many considerations discussed in the report which may be of interest to the FDSN WG. Also useful may be the Appendix which documents some of the lengthier discussions by members of the Team on related issues.

Table I (separate attachment)
A table used by the Review Team to evaluate each Metadata element as described in the body of the report.

APPENDIX

Bob Engdahl: One concern I have is that too many of the elements are Required or Recommended in the FDSN WG survey because it is easy to see how they might be helpful and when one takes a survey the answers don't impact the work that would result when a data center must produce the metadata. I think the Team needs to carefully make sure that each Required element is necessary to be FAIR. To really be helpful I think the number of elements in the Required and Recommended categories should be reduced. Another thought is that if there are any elements that can be automatically determined then those should be identified as such and be noted. But it would not fall on a data center to produce those manually but rather the focus could be on finding a way to fund a system that can do the automated tasks so that the process can minimize the impact on a center.

Josep Batllo: Comments about required/recommended/optional elements:
At first sight, I think all elements presented/submitted in the proposal by Tim are worth to consider as elements that allow to better identify a seismogram and its record characteristics. Thus, I think the whole submitted list can be accepted (unless duplication among some one of them is evident) in a first round. All of them will be, at least, optional in the new defined metadata. Another way to see this… We accept the whole list with a "general comment"; but we should give good specific arguments to reject any one of them.

About required/recommended elements:
This is the key point! Too many or too few? Let's look at it in another way. How do I search for seismograms/waveforms on the IRIS WaveServers? Normally I search for an earthquake/a time window and/or I search for a region. The result are waveforms from different stations2. Thus, if I'm looking for available analog seismograms, I'll look for start/end time and "region" can be changed by "Station codes in the IRS". Waveform components will be "available seismograms for each station"

Thus (again), a search for four elements: "Start time/ End time /Station code/Components" will return all available seismograms. May we say that these four elements define in a "unique way" an analog seismogram? Maybe yes. All other elements may be recommended/optional (now, please, put all your arguments against this proposal).

There are, of course, many "nuisances". To say "Station code" is equivalent (to me) to say lat, lon, altitude. Thus, we find three redundant elements. Of course, it is

good for a user to find these elements attached directly to the image. In this way he/she has not to go to search the IRS for them. But, in another step, introduction of the latter elements can be automatized if there is a connection between the proposed metadata system and the IRS.

Another "turn of the screw" on this point. Potential users of the new system will not look directly for waveform as it is done with present WaveServers. In a first step, they will search the "availability" of scanned images (not waveforms!!) and to download them. In some cases, it would not be images to download; but the system will inform you about the existence of a seismogram and where to ask for it. The four previously proposed elements define quite well an analog seismogram (or its scanned image).

And more nuisances:
If you ask for start/end time (some hours) and code of a WWSSN it will return you 6 seismograms. When you ask for the same for contemporary Russian (or Kazakhstan -Inna, this is for you!!) station with beautiful Kirnos instruments you will get back "three times the same seismogram?" (As the three components were recorded in the same sheet of paper). But, most likely, this is to go to depth for us as evaluators, this will be a topic for those writing the application (and to consign in manuals). In fact, the last one is the solution given by INGV SISMOS (to have three different sets of elements pointing to the same image, one for each component).

Let's point again that the users will receive "images", not waveforms and I think this makes these issues acceptable. Thus, there is a previous step before to use seismograms and its waveforms. On a first step, the unavailability of instrument transfer function and all complementary details (record speed, etc.) is not playing against the "location" of the seismogram. Of course, if the received set of metadata gives you all needed inputs to, once the raw seismogram vectorized, to pass from raw waveform to "ground motion" analog seismograms will be much more used than now. This is the reason because it may be necessary to ask for more "required" elements.

Another point to think about. Those working with old seismograms know (I know) that when looking for records of an earthquake we do not get so much records as it can be for a search with IRIS WS for an earthquake occurred just yesterday (unless you look for Chile 1960/Alaska 1964, etc.). In fact, the number of seismograms available (and still existing) for earthquakes occurred prior to the IGY

(International Geophysical Year) is much reduced than after that point. Thus, the search for needed information to obtain the “ground motion” may not be too big.

Inna Sokolova: I recently worked with colleagues from Central Asia to collect and check metadata on stations whose seismograms are stored in the archives of Kyrgyzstan, Kazakhstan, and Tajikistan. We have collected data for over 400 stations. Of course, many of them are not registered with the ISC. And those stations that are registered with the ISC, many have inaccuracies. In addition, IR registration has a significant disadvantage: there is no information about the time period over which the information presented is correct. Many stations move to another location over time, but there is no way to reflect this information in IR. So, it is not possible to register all stations all over the world, but of course, we should should be tends to do it. Secondly, it is necessary to revise the IR registration standards because stations description is not comprehensive at present time.

Josep Batllo: I suppose that would be true of the elevation as well? But could there also be a case where we have no IR station code?" Yes, of course!! (Some old stations with interesting seismograms never received a code). In this case I think the procedure should be to get an IRS code for the site (they are for free) and "proceed as usual". Yes, also for the "elevation"

Inna Sokolova: We have already discussed that the database of stations metadata IR (ISC) requires some revision. First, there are some discrepancies in coordinates and altitude, the stations operation period is not shown, its moving to another place, network and other. It is not convenient to search information about station in IR. I like working with the IRIS DMC data base describing the digital stations SeismiQuery (http://ds.iris.edu/SeismiQuery/station.htm). Each station there is well described. Can we suggest developing a new format for IR? And recommend that National and Regional seismological Centers revise the information of IR stations.

Inna Sokolova: To facilitate the work with historical seismograms, I have the following suggestions: In the USSR there was limited number of manufactured seismometers and galvanometers. The most of them are described in the book by (Aranovich, Z.I. et al. Main types of seismic instruments, Equipment, and methodology of seismic observations in USSR. - M.: Science, 1974). Most of the stations were equipped with standard instruments having standard parameters. Of course, there were calibrations, but as I remember their minor deviations from standard parameters. I think the same was for other countries too. I think it would

be better to make a brief reference book of analogue seismic equipment (for all known equipment in the world) with tables of standard parameters in unified form. Probably, such reference book already exists, but I have not met. The availability of such reference book will facilitate much the work with historical seismograms.

Josep Batllo: About a “reference book of analogue seismic equipment” I do not remember it exists. It is true that it is possible to find information on the web.
As you say, for soviet stations:
https://www.rand.org/content/dam/rand/pubs/reports/2009/R1204.pdf
https://apps.dtic.mil/sti/citations/ADA016246
For WWSSN, I think the following two publications give enough information
http://ds.iris.edu/seismo-archives/info/stations/WWSSN1964.pdf
https://pubs.er.usgs.gov/publication/ofr20141218
Myself I did something for Spain (in Spanish)
http://www.ign.es/web/resources/sismologia/publicaciones//Catalogosismografos.pdf (and I tried to give transfer functions in poles and zeroes)
Other useful information can be found at “SEISMOARCHIVES”
http://ds.iris.edu/seismo-archives/info/
And many other scattered documents, e.g., for the Austro-Hungarian Empire:
https://www.ig.cas.cz/wp-content/uploads/2018/03/kozak_history-of-seismol-czech-ii.pdf(See Table 2) -in the way, if you look at figure 1, station 4 corresponds to Leopolis/Lvov/Lviv (the last two should be written in Cyrillic), a city now in the news for awful reasons and one of the earliest seismic stations in the region. Also, attached, I’m sending you an unknown document (I remember how surprised Jim Dewey was when I show it to him) I found it years ago at the Library of Ebro Observatory. As you may see, it contains a good description of characteristics of the American stations (North and South) around 1950. Certainly, after the second world war (or maybe because of the International Geophysical Year) there was a clear unification of the seismographs. This is not the case for the period 1890-1950 (or, in another words, there was much more variety of instruments and "versions" of instruments).

Josep Batllo: Certainly, all this should be put together (in fact, this is one of my “main objectives” as responsible of the ESC WG on legacy seismograms, against this objective I’ve the problems of “not to have so much time” and “not to know so much about websites”). On another side, to prepare a draft of a list of available information on the web will not be difficult. Going into the objectives of the ESC WG, in the midterm there is the project to write a new chapter, appendix or whatever, of the NMSOP dealing with the specific problems of analog records. We

are the last people that have been working regularly with analog records. Young people do not know them, and we should keep the "know-how" about how to deal with them (e. g. just a couple of weeks ago a young researcher -around 30- told me that she found a big problem because the time read on a ROM seismogram of 1947 does not correspond to that consigned in the bulletin…). All this goes, of course, much further away that our present objective; but it is also part of the problems to solve if analog records should be used. A last comment on the topic. Inna, as you mentioned Aranovich, similar "original" information should be available in Russian? It will be good to have the references of original information worldwide (also in Japanese, or whatever). If we know the sources, we may search them and, if it is necessary, in a second step, to translate them to other languages. More on the same. It will be good to have these old references scanned and available on the net (at the RAS web, the Kazakhstan seismological survey web- or whatever – also ISC or IRIS may be a good places).

Bob Engdahl: My feeling is the time correction should be required because we really need accurate phase arrival times to have any chance of getting better locations for historical earthquakes when those data are limited. However, I am willing to listen to any strong arguments by Josep why it should be only recommended.

Inna Sokolova: The "time correction element" is, in general available directly on the records of the WWSSN stations and, if I remember properly, in the contemporary stations of Soviet Union (Inna, please, confirm). But it is not so common in secondary stations and older records. These records may contain valuable and useful information. If "Time correction" is a required element, they will be discarded, and I think they should fit in the system. There are two solutions. One is that element "time correction" would be just "recommended". A second one is that a codification "not available" N/A can be introduced in the system. To me the second option is better because it gives indication to the user that this parameter is "unknow/not available". This N/A may be useful for other cases, as when no minutes are available in the elements "start time"

Josep Batllo: The first point is that, if a N/A entry (or similar) is possible, I'll be happy to classify this element as "required". If not…. I've been looking for seismograms available at SISMOS website and others I've access, and it confirms what I was thinking previously. The time correction written in the seismograms becomes more generalized on the fifties. In early times (let's say up to the middle of XX century) the time correction was, of course, introduced when picking phases; but it was not written in the seismograms. My search has been "aleatory";

but I can give some approximate results Stations for which I didn't find time correction written on the seismograms prior to 1950 approx.: LPB, PAS, SVE, TUO (Starts in 1940), ROM, SCO (starts in 1939), BUC (starts 1960), LIS (starts 1967), And stations without "time correction" written in the seismograms at any moment: ATH, BER, FBR, GTT, HEL, ISK, TRS, MNH, EBR. Certainly, I know observatories (GTT, FBR, EBR) that keep the information in station books, etc.; but in others has been lost. In the case of EBR (A place I know well), the time correction was not checked every day. Thus, introduce the time correction means to check different documents and, even, to perform some calculations. I prefer to have the "real available data" and do "approximations" by myself. My main conclusion is that, if we force the "time correction" as required, we'll loss many important early seismograms (mainly because the amount of work to identify a seismogram is growing). Certainly, it is important to know this datum when available, this is the reason because I prefer "recommended" instead of "required" I do not think the main use of legacy seismograms is to pick up times. A good time correction of a phase picking needs two "time corrections": those obtained before and after the time of interest. In fact, the time correction element will need to entries: the time correction itself and the moment it has been calculated (it was written this way in the WWSSN seismograms). Maybe we should point this in our report. I've been reading again old picks (for regional seismograms, Batlló et al. 2008, BSSA) and the conclusion is that I cannot do better than the original ones stated in the bulletins (they are properly made and, considering the accuracy of the different parameters involved, it is not possible to precise better than 1 sec.). Instead, it is true that it is possible (and recommended) to use them to correct misidentified phases (before 1940 many PPP published readings correspond to PKiKP). To me, the main use of the scanned legacy seismograms is to obtain vectorized waveforms for present use. Maybe, the general conclusion and recommendation for the FDSN WG is that definition of this element (required or recommended) should be studied (more) carefully.

Inna Sokolova: I looked at the SEED Reference Manual V2.4. Appendix A contains information about Channel Naming. These naming standards are designed for digital equipment. The first letter specifies the general sampling rate and the response band of the instrument.

Band code Band type Sample rate (Hz) Corner period (sec)
F ... ≥ 1000 to < 5000 ≥ 10 sec
G ... ≥ 1000 to < 5000 < 10 sec
D ... ≥ 250 to < 1000 < 10 sec
C ... ≥ 250 to < 1000 ≥ 10 sec

E Extremely Short Period ≥ 80 to < 250 < 10 sec
S Short Period ≥ 10 to < 80 < 10 sec
H High Broad Band ≥ 80 to < 250 ≥ 10 sec
B Broad Band ≥ 10 to < 80 ≥ 10 sec
M Mid Period > 1 to < 10
L Long Period ≈ 1
V Very Long Period ≈ 0.1
U Ultra Long Period ≈ 0.01
R Extremely Long Period ≥ 0.0001 to < 0.001
P On the order of 0.1 to 1 day1 ≥ 0.00001 to < 0.0001
T On the order of 1 to 10 days1 ≥ 0.000001 to < 0.00001
Q Greater than 10 days1 < 0.000001
An Administrative Instrument Channel variable NA
O Opaque Instrument Channel variable NA

This standard does not describe analog instruments. If so, it is necessary to add a description. Maybe I don't know something and there is already a description of the channels for analog instruments?

Tim Ahern: The current effort is more about capturing the analogue data. IF/when some of the analog get digitized is the time that the sample rate comes into things. I am assuming that such digitizing would determine the sample rate as well and the data would be transformed into miniSEED with the appropriate channel naming at that time. We try to capture the type of instrument and therefore the response characteristics would also help in determining appropriate sample rates as well as some other items such as the resolution of the image capture.

Josep Batllo: Now I come back to transfer functions. There are several elements related to them in the elements list we are reviewing: 22, 23, 27, 31, 32, 78, 80. In fact, as it is written right now in Tim's proposal, the transfer function for an instrument is not properly solved. Elements 22 and 23 ask for free period and damping of galvanometer; but free period and damping of sensor is not asked for anywhere (just indirectly in 78). In fact, we have two main ways to introduce the characteristics of the transfer function: We may use the "old one", directly related to contemporary information (bulletins, reports, etc.). For mechanical instruments Free period, sensor damping, and static magnification are needed. If we look for completion in the future, also dry friction should be introduced (it is not used just because if one introduces this term in the motion equation it has not an analytical solution… but it has numerical solution!!) For electromagnetic instruments we need free period and damping of sensor and galvanometer, as well as coupling.

Also, magnification and the frequency at which it is measured. Another way is to introduce the transfer function in “modern” notation, this is, in poles and zeroes. It is more general (it allows any instrument type), and its management is already solved in present mseed(?) format (also, young people know how to manage). In this last case, some tools (small programs/scrips) should be given to people in charge to introduce the metadata in the system allowing a quick transformation from “old constants” to poles and zeroes. It is not difficult. In the catalogue of Spanish seismographs, I applied some easy rules (in this case I assumed “coupling = 0”) (Batllo and Bormann, SRL, 2000). Most likely it is not our work to solve these problems now, but to point them to the FDSN WG for a better result.

Inna Sokolova: I agree with Josep that representation of responses in the form of zeros and poles and application of Dataless-seed format and XML is convenient for use. I agree that we need script for PAZ using the instrument parameters. It is also possible to create a library with calculated PAZ for standard analog equipment and creation of a script that generates responses in Dataless-seed format the same as PDCC software (distributed by IRIS DMC).