

Template for suggesting changes to miniseed specification

Commenting on document version #	M=modification, N= add new section, D=Delete existing section	Modification
Topic	Topic	General Compression and Opaque data encoding
Type of Action (M=modification, N= add	Type of Action (M=modification, N= add	Modification
Current Wording from document	Applies to M or D	<p>Define encodings for a general compression method for fundamental data types. Also define an encoding for opaque data.</p> <p>Suggestion: Encoding 50: 32-bit integers,general compressor Encoding 51: 32-bit IEEE floats,general compressor Encoding 52: 64-bit IEEE floats,general compressor Encoding 100: Opaque data</p>
New wording	Applies to M or N	<p>Suggestion: Encoding 50: generic tagged compressor Encoding 100: Opaque data</p>
Rationale		<p>There are numerous LZ-based and other generic lossless compression algorithms. This is a field of continuous development. "Standards" can have a half-life of a few years.</p> <p>The suggestion is to define a single new "generic" encoding type, and to use the payload, for example with an introductory ascii tag or other means,, to define the specific encoding method to follow. This is extensible, and flexible. This approach is already discussed in the VAR field for opaque headers. For example, "ALGORITHM=BROTLI2027,TEXT~", followed by whatever.</p> <p>While LZ-based algorithms like 7-zip or Brotli are great with text (e.g. station xml), generic dictionary based compression is not particularly effective for continuous 1-D time series data (like seismic, acoustic, barometric...etc). A time series format based on generic compression methods consumes significantly (ordes of magnitude) more resources (memory, cycles) than the efficient methods in use now with little benefit in data volume reduction.</p>
Comments: Author, organizatino, and email		Edelvays Spassov, Kinemetrics, ens@kmi.com
Date of Comment		5/18/16