SAND2014-16766R

SANDIA REPORT

SAND200X-XXXX
Unlimited Release
July 2014

IDC Reengineering Phase 2 Glossary

Version 0.1 (Initial Draft)

Christopher J. Young, James Mark Harris

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND200X-XXXX
Unlimited Release
July 2014

IDC Reengineering Phase 2 Glossary

Christopher J. Young, James Mark Harris
Next Generation Monitoring Systems Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MS0401

Abstract

This document contains the glossary of terms used for the IDC Reengineering Phase 2 project. This version was created for Iteration I1.

REVISIONS

Version	Date	Author/Team	Revision Description	Authorized by
0.1	7/31/2014	SNL IDC Reengineering Team	Initial Release for I1	M. Harris

TABLE OF CONTENTS

Revisions 4

1. Terms 9
   - 1.1. Agglomerative Hierarchical Clustering 9
   - 1.2. Alphanumeric Data 9
   - 1.3. Alphanumeric List 9
   - 1.4. ALT (Alternate Subsystem) 9
   - 1.5. Amplitude Correction Factor 9
   - 1.6. Analysis Time Interval 9
   - 1.7. Area of Interest 9
   - 1.8. Array 10
   - 1.9. Array Coherence 10
   - 1.10. Aseismic Region 10
   - 1.11. Association 10
   - 1.12. Attenuation 10
   - 1.13. Automatic Processing 10
   - 1.14. Azimuth 11
   - 1.15. Backup 11
   - 1.16. Basemodel 11
   - 1.17. Beam 11
   - 1.18. Beam Steering 11
   - 1.19. Beam, Coherent 11
   - 1.20. Beam, Continuous 11
   - 1.21. Beam, Detection 11
   - 1.22. Beam, fk 12
   - 1.23. Beam, Incoherent 12
   - 1.24. Beam, Origin 12
   - 1.25. Beamforming (Beaming) 12
   - 1.26. Beamforming (Beaming), Coherent 12
   - 1.27. Beamforming (Beaming), Incoherent 13
   - 1.28. CD-1.0 Format and Protocols 13
   - 1.29. CD-1.1 Format and Protocols 13
   - 1.30. Cepstral Domain Measurements 13
   - 1.31. Channel 13
   - 1.32. Channel, Derived 14
   - 1.33. Channel, Raw 14
   - 1.34. Correlation Coefficient 14
   - 1.35. CSS3.0 Database Schema 14
   - 1.36. Data Acquisition 14
   - 1.37. Data Acquisition Partition 14
   - 1.38. Data Authentication 15
   - 1.39. Data Processing Partition 15
   - 1.40. Data Provider 15
   - 1.41. Data Quality Errors 15

1.42. Data Source 15
1.43. Database Backup 15
1.44. Database Restore 15
1.45. Defining/Non-Defining....15
1.46. DOE 16
1.47. Earth Model....16
1.48. Earth Model, 1D 16
1.49. Earth Model, 2D 16
1.50. Earth Model, 3D 16
1.51. Empirically Derived Corrections....16
1.52. Event 17
1.53. Event Bulletin....17
1.54. Event Change History....17
1.55. Event Epicenter 17
1.56. Event Hypocenter 17
1.57. Event Hypothesis....17
1.58. Event Hypothesis Location Solution 17
1.59. Event Hypothesis Quality Metric 18
1.60. Event Location 18
1.61. Event Magnitude, Network 18
1.62. Event Magnitude, Station....18
1.63. Event Quality Metric 18
1.64. Event Report....18
1.65. Event Set 18
1.66. Expected Signal Detection 18
1.67. F-statistic 19
1.68. False Event 19
1.69. Filter Cascade 19
1.70. Filter, Waveform 19
1.71. Frequency Domain Measurements 19
1.72. Frequency-Wavenumber (fk) Measurements....20
1.73. Frequency-Wavenumber (fk) Power Spectrum....20
1.74. Frequency-Wavenumber (fk) Processing....20
1.75. Geographic Data 20
1.76. Geographic Information System (GIS)....20
1.77. Geometric Spreading 20
1.78. Geospatial Data 20
1.79. GSE 2.1/IMS 1.0 Formats and Protocols....21
1.80. Historical Data....21
1.81. Hydroacoustic Arrival Group (HAG) 21
1.82. Hydroacoustic Time Domain Measurements....21
1.83. IDC (International Data Center)....21
1.84. IMS (International Monitoring System) 22
1.85. Interactive Processing 22
1.86. Late Signal Association 22
1.87. Late-Arriving Waveform 22

1.88. Machine Readable Earthquake Data Report 22
1.89. Magnitude Estimation 22
1.90. Magnitude Type 23
1.91. Maximum Likelihood Magnitude Estimation (MLE) 23
1.92. Mini-SEED Format 23
1.93. Missed Event 23
1.94. Mission Transfer 23
1.95. Monitoring Mission Performance 23
1.96. Near Real Time 23
1.97. Network 24
1.98. Network Magnitude Detection Threshold 24
1.99. Network Processing 24
1.100. Noise 24
1.101. Operational Processing Time Period 24
1.102. Operational Subsystem 24
1.103. Operations Log 24
1.104. OPS (Operational Subsystem) 25
1.105. Origin 25
1.106. Phase 25
1.107. Phase Grouping 25
1.108. Pipeline Processing 25
1.109. Pipeline Transfer 25
1.110. PKI (Public Key Infrastructure) 25
1.111. PKI Credentials 25
1.112. Playback 25
1.113. Polarization Features 26
1.114. Power Detector 26
1.115. Preliminary Phase Label 26
1.116. Primary 26
1.117. Processing Sequence 26
1.118. Processing Stage 26
1.119. Processing Unit 27
1.120. Q (Quality Factor) 27
1.121. QC Mask 27
1.122. QuakeML 27
1.123. Quality Control (QC) 27
1.124. Reference Event 27
1.125. Reference Event Database 27
1.126. Regression Testing 28
1.127. Rejected Event Hypothesis 28
1.128. Relative Event Magnitude Estimation 28
1.129. Relative Event Relocation 28
1.130. Remote Operations Center (ROC) 28
1.131. Rotation 28
1.132. SEED Format 29
1.133. Sensor Orientation 29

1.134. Signal 29
1.135. Signal Association 29
1.136. Signal Characterization 29
1.137. Signal Detection....29
1.138. Signal Detection Feature 29
1.139. Signal Detection Feature Measurement 29
1.140. Signal Detection Feature Prediction 30
1.141. Signal Detector Threshold....30
1.142. Signal Enhancement....30
1.143. Single-Station Event Hypothesis 30
1.144. Site 30
1.145. Slowness....30
1.146. STA/LTA 30
1.147. State-of-Health (SOH) Station Data 30
1.148. Station 31
1.149. Station Data 31
1.150. Station Magnitude....31
1.151. Station Magnitude Detection Threshold 31
1.152. Station Processing....31
1.153. Subspace Detector....31
1.154. System Backup 32
1.155. System Configuration 32
1.156. System Message....32
1.157. System Performance 32
1.158. System Restore 32
1.159. System Storage Format 32
1.160. Test Data Set....32
1.161. Test Suite....33
1.162. Three Component Station....33
1.163. Time Domain Measurements 33
1.164. Training Data Set....33
1.165. Training Subsystem 33
1.166. Unit Testing....33
1.167. Virtual Event Hypothesis 33
1.168. Virtual Origin 33
1.169. Waveform....33
1.170. Waveform Correlation Event Processing....34
1.171. Waveform Format, Native....34
1.172. Waveform Format, Storage 34
1.173. Waveform Processing 34
1.174. Waveform QC Mask 34
1.175. Waveform Quality Control....34
1.176. Waveform, Derived....34
1.177. Waveform, Raw 35

1. TERMS

1.1. Agglomerative Hierarchical Clustering

A method to find clusters of similar individuals within a population. The population is the set of *waveforms* recorded at a single station for a set of *events* and the measure of similarity is *correlation coefficient*. Agglomerative hierarchical clustering iteratively clusters waveforms into a single tree-like structure called a dendrogram. For more information on the agglomerative hierarchical clustering algorithm, see the book Cluster Analysis, by Everitt et al.

1.2. Alphanumeric Data

Data that is displayed to an analyst as numbers or letters, e.g. the latitude and longitude of an *event hypothesis*. This is in contrast to waveform data, which is displayed as a line on a 2D plot.

1.3. Alphanumeric List

Alphanumeric data presented in a multi-column table. A list of *events*, a list of *signal detections* for an event, or a list of *stations* are all examples.

1.4. ALT (Alternate Subsystem)

A duplicate of the *Operational Subsystem* (*OPS*) software and hardware at a geographically separate location. The ALT is continuously available to assume mission responsibilities.

1.5. Amplitude Correction Factor

A phase and frequency specific multiplicative factor which represents the decrease in the amplitude of a wavefront due to *attenuation* as it travels from source to receiver. Amplitude correction factors are applied in conjunction with corrections due to *geometric spreading*.

1.6. Analysis Time Interval

The time interval bounding the data reviewed by an analyst. That data includes *waveforms*, *event hypotheses* and their associated *signal detections*, and unassociated *s*ignal detections.

1.7. Area of Interest

A geographic region deemed to be important for the operational monitoring mission.

1.8. Array

A group of *stations* (sometimes referred to as "elements") deployed with the intent of processing waveform data with specialized techniques exploiting signal coherence between the stations. This processing improves SNR and also provides estimates of the direction (*azimuth* and *slowness*/velocity) of signals. Arrays are used for seismic, infrasonic, and hydroacoustic monitoring.

1.9. Array Coherence

A measure of the similarity of *waveforms* during a given time interval between the elements of an *array*. High array coherence indicates that a signal is propagating across the array.

1.10. Aseismic Region

A region of the Earth where no seismic events have ever been recorded. Many seismic monitoring techniques (e.g. *magnitude estimation*, estimating *station magnitude detection threshold, waveform correlation event processing*) leverage results from past events, so aseismic regions present additional challenges.

1.11. Association

See *signal association*.

1.12. Attenuation

The decrease in amplitude of a signal due to loss of energy as the signal propagates away from its source. Attenuation consists of two components: intrinsic attenuation which is the phenomenon in which kinetic energy is converted to heat by anelastic processes or internal friction, and scattering attenuation, which is caused by energy reflecting off of small scale material heterogeneities. In addition to attenuation, *geometric spreading* also acts to modify the amplitude of seismic signals.

1.13. Automatic Processing

Data processing that is performed by the system without analyst interaction. This type of processing is initiated by a configurable trigger such as availability of data or the completion of a previous processing step. Includes *station processing*, *network processing*, post-analysis processing, post-evaluation processing, and late data processing.

1.14. Azimuth

The angle in degrees measured clockwise from geographic North of a signal arriving at a *station*. Azimuth and *slowness* completely describe the vector direction of arrival for a signal at a station.

1.15. Backup

The Subsystem acting as the backup location for the System. The pipeline and analyst processing do not actively run on the Backup. May be either *OPS* Subsystem or *ALT* Subsystem. The *Primary* always synchronizes data to the Backup. During a *pipeline transfer*, the Primary and Backup roles switch.

1.16. Basemodel

An *Earth model* used to compute base-level predictions (e.g. seismic travel time) that can be made more accurate by applying *empirically derived corrections* developed specifically for that basemodel. An Earth model is only referred to as a basemodel when predictions are further refined using empirically derived corrections.

1.17. Beam

The product of *beamforming*; a single derived channel (see *channel, derived*) representing the sum of the raw channels (see *channel, raw*) for all the elements of an *array*.

1.18. Beam Steering

See *beamforming*.

1.19. Beam, Coherent

The derived channel (see *channel, derived*) resulting from coherent beamforming (see *beamforming*, *coherent*).

1.20. Beam, Continuous

A *beam* channel formed continuously in time, i.e., without any specified start or end time.

1.21. Beam, Detection

The *beam* on which a *signal detection* occurs. When detecting signals for data from an *array*, a set of beams is formed spanning the regions of monitoring interest. Detectors are run on all of these beams to look for signals. If a signal is detected, the beam on which the detection occurred is known as the detection beam. If there are detections on more than one beam, then the best of these will be selected (e.g. best SNR) as the detection beam.

1.22. Beam, fk

A *beam* steered to point to the maximum *fk power spectrum* direction (*azimuth* and *slowness*) in 2D fk space (see *frequency wavenumber processing*) corresponding to a *signal detection* on an *array*. The fk beam should show the best possible signal for a signal detection and hence is the beam that is automatically shown to an analyst reviewing an *event hypothesis*.

1.23. Beam, Incoherent

The derived channel (see *channel, derived*) resulting from incoherent beamforming (see *beamforming*, *incoherent*).

1.24. Beam, Origin

A *beam* steered to the location of an *event hypothesis*. A particular *phase* (typically first arriving P) must be assumed for each origin beam. When an event hypothesis is displayed for an analyst to review, *fk beams* are shown for detecting *stations*, and origin beams are shown for non-detecting stations. Assuming the *Event hypothesis* is well-located, the origin beams should allow the analyst to manually add any *signal detections* that were missed.

1.25. Beamforming (Beaming)

Beamforming (also known as beaming) is a multichannel signal processing technique taking advantage of the direction-dependent arrival of a signal across an the elements of an *array*. Beamforming sums the waveform data from the elements of an array to produce a single derived channel (see *channel, derived*). The intent is to boost SNR.

1.26. Beamforming (Beaming), Coherent

A method for increasing the SNR of signals arriving at an *array* from a particular *azimuth* and *slowness*. Coherent beamforming (also known as *beam steering*) time shifts the *waveforms* from an array's elements before summing under the assumption that a plane wave is arriving from that direction. The shifting is done for each element by subtracting the time delay relative to the array beampoint (a reference location for the array) that would be expected for a plane wave arriving from that azimuth and slowness. If there is a signal arriving from the specified azimuth and slowness, an SNR gain occurs when summing the time delayed waveforms due to the simultaneous constructive interference of coherent directional signals and destructive interference of incoherent background noise. In theory, a gain of SNR equal to the square root of the number of elements can be achieved, e.g. a factor of 3 for a 9 element array. In practice, the realized gain is usually less.

1.27. Beamforming (Beaming), Incoherent

The same as *coherent beamforming* except that the *waveforms* are rectified (absolute values) before summing.

1.28. CD-1.0 Format and Protocols

A standard used for transmitting continuous waveform data from *stations* of the International Monitoring System (see *IMS*) to the International Data Center (see *IDC*) and for transmitting these data from the IDC to National Data Centers (NDCs). CD-1.0 uses a TCP/IP program-to-program socket communication to send binary formatted waveform data (frames). The CD-1.0 standard has been superseded by the CD-1.1 standard (see *CD-1.1 Format and Protocols*). The format and protocols are described in the document Formats and Protocols for Continuous Data CD-1.0 from the CTBTO found at ctbto.org/fileadmin/user_upload/procurement/2011/Attachment_3_to_the_TOR.pdf.

1.29. CD-1.1 Format and Protocols

A standard used for transmitting continuous waveform and state-of-health (see *SOH*) data from *stations* of the International Monitoring System (see *IMS*) to the International Data Center (see *IDC*) and for transmitting these data from the IDC to National Data Centers (NDCs). This protocol is designed to support multicasting of continuous data, which entails more complex data flow topologies. The format and protocols are described in the document Formats and Protocols for Continuous Data CD-1.1 from the CTBTO found at ctbto.org/fileadmin/user_upload/procurement/2011/Attachment_4_to_the_TOR.pdf.

1.30. Cepstral Domain Measurements

Measurements made in the cepstral domain for characterizing a signal of interest, in particular hydroacoustic signals. The cepstral domain highlights periodicities in the spectrum (frequency domain). A cepstrum of a *waveform* is the Fourier transform of the waveform's power spectrum. The independent variable of the cepstrum is called quefrency, expressed in units of time but representing the period of harmonic features of the waveform.

1.31. Channel

A generic term for either a raw channel (see *channel, raw*) or a derived channel (see *channel, derived*). A time interval of data from a channel is a *waveform*.

1.32. Channel, Derived

A source for time series data created by processing one or more raw channels (see *channel, raw*). Examples of common types of processing to form derived channels are filtering (see *filter, waveform*), beaming (see *beamforming*), and *rotation*. Derived channels are generally created to enhance the SNR of signals.

1.33. Channel, Raw

A source for unprocessed time series data from a seismic, hydroacoustic, or infrasonic sensor. E.g. the output from a short period, vertical component seismometer.

1.34. Correlation Coefficient

A measure of similarity between two *waveforms*. The absolute value of the correlation coefficient ranges from 0 (no similarity) to 1 (a perfect match). Formally, the correlation coefficient for two entities is the covariance divided by the square root of the product of the variances. For waveforms, this is the cross-correlation of the two waveforms divided by the square root of the product of the auto-correlations.

1.35. CSS3.0 Database Schema

A schema that was developed by the former Center for Seismic Studies (CSS) in ~1990 to facilitate storage and retrieval of seismic data used for seismic monitoring of test ban treaties. It consists of 21 relations (database tables) that describe parametric data associated with seismic *events*, including *station* information. The CSS3.0 Format is described in the document Center for Seismic Studies Database Schema Version 3.0, June 1990, SAIC-90/1235.

1.36. Data Acquisition

Receiving data from a *data provider* and storing it for subsequent access by the System. Data can be acquired in a variety of formats including *CD-1.0*, *CD-1.1*, *IMS-1.0*, *SEED*, and *miniSEED*. Data acquisition can be unclassified or classified, depending on the data provider.

1.37. Data Acquisition Partition

The portion of the System where data is acquired (see *data acquisition*) and re-distributed.

1.38. Data Authentication

The corroboration that data have not been altered in an unauthorized manner by using such mechanisms as digital signatures, *PKI* (public key infrastructure), etc.

1.39. Data Processing Partition

The portion of the System where data are processed, analyzed, evaluated, and where data processing products are distributed and archived.

1.40. Data Provider

See *data source.*

1.41. Data Quality Errors

Errors in *waveforms* that can lead to problems with processing and analysis. Data quality errors include data gaps, amplitude spikes, repeated amplitude values, linear trends and invalid gain. Prior to signal processing, data is analyzed by *quality control* (QC) software to identify and record any errors (see *QC Mask*).

1.42. Data Source

A source of waveform and/or *event* data for the System. Data sources can be *stations* (e.g., *IMS* stations) or external data centers (e.g., *IRIS-DMC*).

1.43. Database Backup

The copying or archiving of the contents and state of database at a particular point in time. The purpose of making a database backup is to provide the information needed for a *database restore*, should that be necessary.

1.44. Database Restore

The restoration of a database's previous state. To restore a database, a particular *database backup* must be selected.

1.45. Defining/Non-Defining

Any observation that contributes to the determination of an *event* attribute is considered to be "defining" for that attribute. The detection of an event, the location of an event (see *event location*), the magnitude of an event see *event magnitude, network)*, and the *source type* assigned to an event are all determined by specific types of *defining* observations (travel time, *azimuth*, *slowness*, amplitude) from one or more *phases* recorded by one or more *stations*. If an observation is linked to an event (associated) but does not contribute to the calculation of an event attribute, then it is considered to be non-defining for that attribute.

1.46. DOE

The United States Department of Energy

1.47. Earth Model

A representation of one or more physical properties of the Earth, generally used for determining or validating some aspect of an *event hypothesis* (e.g. signal travel times, signal azimuths, signal *slownesses*).

1.48. Earth Model, 1D

An *Earth model* in which the physical properties represented vary only with one dimension. Generally this dimension is depth for seismic and hydroacoustic models and elevation for infrasonic models. 1D depth/altitude varying models are effective because gradients of physical properties are generally much stronger in the vertical direction due to gravitational effects on the solid Earth, oceans, and atmosphere.

1.49. Earth Model, 2D

An *Earth model* in which the physical properties represented vary with two dimensions. Generally these dimensions are latitude and longitude. There are not many examples of physical properties that vary only with latitude and longitude and not with depth, so 2D models are less common than 1D or 3D models. Examples of types of 2D models are surface wave group velocities (for predicting dispersion) and body or surface wave attenuation (for predicting amplitude).

1.50. Earth Model, 3D

An *Earth model* in which the physical properties represented vary with three dimensions. Generally these dimensions are latitude, longitude, and depth. 3D Earth models represent the highest level of fidelity of static (non-time varying) Earth models.

1.51. Empirically Derived Corrections

Corrections that are applied to predictions (e.g. seismic travel time) calculated using *basemodels* to improve fidelity of the predictions. Empirically derived corrections are based on measurements from ground-truth events and are developed specifically for a particular basemodel. Empirically derived corrections are only used by entities that have a need for accuracy beyond what can be provided by an Earth model.

1.52. Event

The occurrence of some source of energy within the Earth's body, oceans, or atmosphere that can be detected by seismic, hydroacoustic, and/or infrasonic sensors. For the same event, many different *event hypotheses* may be created at different *processing stages*. One of these event hypotheses must be designated as preferred.

1.53. Event Bulletin

A list of *event hypothesis location solutions*, with or without the associated *signal detections.* The primary product of most seismic monitoring agencies (e.g. IDC, NEIC) is an event bulletin. Event bulletins can be constrained by region, by time, by magnitude, etc.

1.54. Event Change History

A complete record of the evolution of the *event hypotheses* corresponding to a single event, from the initial detection of the *event* through the final version approved for release.

1.55. Event Epicenter

An *event's* 2D geographic location as described by latitude and longitude, i.e. the position on a map. This is in contrast to an *event hypocenter*, which refers to latitude, longitude, and depth.

1.56. Event Hypocenter

An *event's* 3D spatial location as described by latitude, longitude, and depth. This is in contrast to an *event epicenter*, which refers only to the latitude and longitude, i.e. the position on a map.

1.57. Event Hypothesis

A set of *signal detections* and calculated event parameters that are associated with an *event*. An event hypothesis has one or more *event hypothesis location solutions*, one of which must be designated as preferred.

1.58. Event Hypothesis Location Solution

An estimate of an *event location* tied to a particular *event hypothesis*. Another term for an event hypothesis location solution is an *origin*.

1.59. Event Hypothesis Quality Metric

See *event quality metric*.

1.60. Event Location

The combination of an event's spatial location (see *event hypocenter*), temporal location, spatial location uncertainty, and temporal location uncertainty.

1.61. Event Magnitude, Network

An estimate of the size of a seismic *event* determined by combining the set of available station event magnitudes (see *event magnitude*, *station*). Separate network event magnitudes can be calculated for each available station *magnitude type*, e.g. mb, Ms.

1.62. Event Magnitude, Station

An estimate of the size of a seismic *event* determined by processing the waveform data recorded by one *station*. Separate station event magnitudes can be calculated for different *magnitude types*, e.g. mb, Ms.

1.63. Event Quality Metric

A quality metric computed as a number in the closed interval [0.0, 1.0] (low to high) for each *event hypothesis* formed on the System. This metric indicates the quality of the event hypothesis as a function of the event hypothesis' associated *signal detections* and related measurements, location solution, *station* state-of-health, and *network* state-of-health information. The System computes a new event quality metric whenever any parameter used for calculating the metric is updated and stored.

1.64. Event Report

An automatically or manually prepared document providing detailed information about a particular *event*. Could include text, tables, figures, maps.

1.65. Event Set

The set of one or more *event hypotheses* that an analyst selects for analysis.

1.66. Expected Signal Detection

A *signal detection* that is anticipated to exist but has not been detected by the System. Expected signal detections are typically acquired from empirical knowledge. An Expected Signal Detection is associated with a *station*, a *phase* label and an *event hypothesis*, and may include an expected waveform and/or a set of *signal detection feature measurements*. It may also contain information about parameters that were used in the past to detect the signal, such as filter settings, fk parameters, etc. An expected signal detection cannot become an observed signal detection directly, but can guide the processes that search waveform data for observed signal detections.

1.67. F-statistic

The power on a *beam* divided by the average computed over all the array elements of each element's residual power. The F-statistic can be used as the basis for a signal detector for data from an *array* or to characterize the coherence of the *signal detection* on a particular beam. Because each point in a *frequency-wavenumber (fk) power spectrum* corresponds to an *azimuth* and *slowness* pair, calculating an F-statistic for each point in the spectrum can help identify the peaks. See *"An automatic event detector at the Tonto Forest Seismic Observatory", Blandford, 1974.*

1.68. False Event

An *event* detected by the System that does not actually correspond to a real source of energy, i.e. an artifact of the data processing. Adjustments can be made to the system to reduce the number of false events, but usually only with a corresponding increase in the number of *missed events* built by the system.

1.69. Filter Cascade

The application of two or more waveform filters (see *filter, waveform*) in series. Complex filters can be designed to meet a set of specific needs by applying a series of simple filters, each of which is designed to meet a more basic need. Once the sequence of filters is known, the successive filter operations can be replaced with a single filter operation equivalent to the filter cascade.

1.70. Filter, Waveform

An algorithm that operates on a *waveform* to produce a second waveform with a limited portion of the original frequency content. Typical waveform filters are low-pass, high-pass, or band-pass. Waveform filters are usually applied to enhance signals relative to noise.

1.71. Frequency Domain Measurements

Measurements made in the frequency domain for characterizing a signal of interest, including: total energy, total power, average power, power spectral density, and signal cepstrum. A segment of waveform data (a time-series) is transformed to the frequency domain using a Fourier transform.

1.72. Frequency-Wavenumber (fk) Measurements

Measurements made from the fk transformed data (see *frequency-wavenumber processing*) of an *array*: *azimuth*, azimuth uncertainty, *slowness*, slowness uncertainty, and *array coherence*.

1.73. Frequency-Wavenumber (fk) Power Spectrum

The representation in the frequency-wavenumber domain of coherent *signal* power in the *waveform* data from an *array* for a specified time interval. Calculation of fk power spectrum requires three Fourier transforms to convert time-sampling to frequency (designated by f), and longitude and latitude sampling to x and y wavenumbers (designated by k). Typically fk power spectra are represented as 2D plots (not 3D), by collapsing the frequency information to a single value for each x and y *slowness* by averaging values across the range of frequencies.

1.74. Frequency-Wavenumber (fk) Processing

A signal processing technique that can be applied to a short interval of *waveform* data from an *array* to determine if a *signal* is present and estimate the signal's *azimuth* and *slowness*. Waveform data are first converted to a *frequency-wavenumber power spectrum* and then further processing is done in the fk domain.

1.75. Geographic Data

See *geospatial data*.

1.76. Geographic Information System (GIS)

A software application designed to capture, store, manipulate, analyze, manage, and present *geospatial data*.

1.77. Geometric Spreading

The decrease in signal amplitude as a wavefront expands away from its source that accounts for the increasing wavefront size.

1.78. Geospatial Data

Geospatial data is information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the Earth, typically represented by points, lines, polygons and/or complex geographic features, and may contain information attached to a location. Geospatial data is often accessed, manipulated or analyzed through *Geographic Information Systems*.

1.79. GSE 2.1/IMS 1.0 Formats and Protocols

Formats and protocols developed for the Group of Scientific Experts Third Technical Test (GSETT-3) for communication of IMS monitoring data over the internet. The formats and protocols are described in the document Provisional GSE2.1 Message Formats and Protocols from GSETT-3 found at seismo.ethz.ch/prod/autodrm/provisional_GSE2.1.pdf and also in the document Formats and Protocols for Messages IMS1.0 from the International Data Center found at isc.as.uk/standards/isf/download/ims1_0.pdf.

1.80. Historical Data

Waveform and alphanumeric data referenced to times prior to the *operational processing time period*.

1.81. Hydroacoustic Arrival Group (HAG)

A group of *signal detections* on different hydrophones from a single hydroacoustic station that have been determined to come from the same event and that can be used to determine azimuth back to the event. HAGs can only be formed for multiple element (array) stations, and hence apply to H stations but not T stations.

1.82. Hydroacoustic Time Domain Measurements

Measurements made directly on *waveforms* from hydroacoustic sensors, including: signal termination time, signal summation time, signal time spread, signal skewness, signal kurtosis, peak energy in a defined time period, intensity average time in a defined time period, peak energy in a defined time period, and crossing counts (the number of times a waveform crosses a threshold in a defined time period).

1.83. IDC (International Data Center)

The International Data Center (IDC) is the data processing component of the monitoring system associated with the proposed Comprehensive Nuclear-Test-Ban Treaty (CTBT). The IDC processes the data from a global sensor *network* known as the *IMS*. National Data Centers (NDCs) receive IMS station data and bulletins from the IDC.

1.84. IMS (International Monitoring System)

The International Monitoring System (IMS) is a world-wide *network* of seismic, hydroacoustic, infrasound, and radionuclide *stations* deployed as part of the CTBT verification regime.

1.85. Interactive Processing

Analyst directed data processing. Interactive processing can be performed to review and refine existing *event hypotheses* (automatic or analyst-built) or to build new event hypotheses missed by prior processing.

1.86. Late Signal Association

A network *signal association* process triggered when a *signal detection* is made on a *waveform* after normal network signal association has occurred. Late signal association includes associating the new signal detections to previously formed *event hypotheses* as well as using them to form new event hypotheses.

1.87. Late-Arriving Waveform

Waveform data acquired by the System after *pipeline processing* has been initiated.

1.88. Machine Readable Earthquake Data Report

The USGS electronic bulletin format which contains hypocenter parameters, macro-seismic observations and phase and amplitude data. A detailed description can be found at the web site hazards.cr.usgs.gov/weekly/mchedr.txt

1.89. Magnitude Estimation

The process whereby the magnitude (size) of an *event* is estimated based on the observed waveform characteristics for a specified seismic *phase* at one or more *stations*. Magnitude is calculated for each station observing the event (see *event magnitude, station*) and these results are combined to come up with a network magnitude (see *event magnitude, network*). The magnitude calculation requires that the location of the event (see *event location*) be known, so location must be estimated before magnitude. A magnitude estimation calculation formula must account for the decreases in signal amplitude between source and receiver due to geometric spreading and anelastic attenuation. These factors are determined empirically based on the observed amplitudes for a set of events of well-known sizes.

1.90. Magnitude Type

A particular magnitude estimation method based on a specified *phase*, frequency band, and instrument.

1.91. Maximum Likelihood Magnitude Estimation (MLE)

A method of estimating the magnitude of an *event* using information from both detecting and non-detecting *stations*. For the latter, an amplitude measurement is made at the theoretical arrival time of the *phase* used for the type of magnitude being calculated (see *magnitude type*), and the assumption is that the amplitude for that phase from the event must be less than or equal to the amplitude measured at the theoretical arrival time.

1.92. Mini-SEED Format

A format used to identify *SEED* data records without any control header information, also known as Data Only SEED. The Mini-SEED format is described in Appendix G of the SEED Reference Manual from the Incorporated Research Institutions for Seismology (IRIS) found at fdsn.org/seed_manual/SEEDManual_V2.4.pdf.

1.93. Missed Event

An *event* that was known to have occurred but that was not detected by the System (automatic processing and/or analyst review). Adjustments can be made to the system to reduce the number of missed events, but usually only with a corresponding increase in the number of *false events* built by the system.

1.94. Mission Transfer

The process of transferring the execution of System operations from the *Primary* to the *Backup*.

1.95. Monitoring Mission Performance

A measure of how well the System is performing the monitoring mission. This is based on the System's metrics and statistics supporting sensor state-of-health, geophysical network capability, algorithmic performance, and detected *event* statistics.

1.96. Near Real Time

An implication that the delay between a triggering action and response of the system to that action is not significant. Depending on the type and amount of processing required to produce a result, this delay could range from seconds to 30 minutes.

1.97. Network

A group of *stations* used for monitoring. For example, the *IMS* network is the group of stations used by the *IDC* to monitor the CTBT. Spacing between stations in a network is much larger than in an *array*, and array processing techniques are generally not applied across a network.

1.98. Network Magnitude Detection Threshold

The geographically varying magnitude below which *events* monitored by a specified *network* can no longer be detected with a specified level of confidence (generally 90% or 95%). This threshold can either be determined empirically (if a sufficient number of events spanning a range of sizes have occurred in a region) or by modeling.

1.99. Network Processing

Integrated processing of data from two or more *stations* within a *network*. If the network processing requires processed station data, then *station processing* must be completed for all required stations before network processing can begin. Types of network processing include: *signal association*, *event location* refinement, and *magnitude estimation*.

1.100. Noise

Portions of a *waveform* not containing *event* information, i.e. without apparent signals.

1.101. Operational Processing Time Period

The time during which analysts may refine and save *events* without special procedures (currently 45 days).

1.102. Operational Subsystem

The portion of the System residing at the principal location that supports the operational mission. This subsystem can serve as *Primary* or *Backup*. To insure there is no impact on mission capability, development and testing are supported with separate subsystems (DEV and SUS).

1.103. Operations Log

A log for capturing operational activities, e.g., significant *events*, *station* outages, shift changes.

1.104. OPS (Operational Subsystem)

See *Operational Subsystem*.

1.105. Origin

See *event hypothesis location solution.*

1.106. Phase

An indication of the path and type of a signal originating from an *event* traveling through the body of the Earth, the oceans, or the atmosphere. For example the seismic P phase refers to a compressional wave refracting within the mantle of the Earth, while the seismic ScS phase refers to a shear wave reflecting off the outer core boundary.

1.107. Phase Grouping

The process of grouping all of the *signal detections* from one *station* that are assumed to come from the same *event*. Each signal represents a different *phase*. The system bases grouping on signal detection measurements (e.g. relative timing, *azimuth*, *slowness*). Various operations can be made on signal detections after placing them in groups, including phase assignment and creation of a *single station event hypothesis*.

1.108. Pipeline Processing

The sequence of real-time automatic data processing by the System that begins after acquisition of raw waveform data and results in a set of *event hypotheses* with associated *signal detections*.

1.109. Pipeline Transfer

The transfer of *pipeline processing* from the *Primary* to the *Backup*.

1.110. PKI (Public Key Infrastructure)

Infrastructure used for the secure exchange of information.

1.111. PKI Credentials

Digital keys used for authenticating messages exchanged between systems.

1.112. Playback

The insertion of a previously captured waveform data acquisition feed in place of the real time raw waveform feed.

1.113. Polarization Features

Features derived from the analysis of three component data (see *three component station*) that characterize how a signal has been partitioned across the components. Polarization features can be used to identify *phase* type and for *association* with an *event hypothesis*. Polarization features include: *azimuth*, azimuth uncertainty, *slowness*, slowness uncertainty, rectilinearity, planarity, horizontal to vertical power ratio, and short and long axis incidence angles.

1.114. Power Detector

A signal detector that triggers on changes in *waveform* amplitude, which is proportional to power. The most well-known of these is the *STA/LTA* detector.

1.115. Preliminary Phase Label

A seismic *phase* label determined using only *station* signal detection feature measurements, i.e. prior to network *signal association*. The possible preliminary phase labels are: compressional (P), shear (S), teleseismic (Tx), regional (Rx), and noise (N).

1.116. Primary

The system role defining the location of the active pipeline and analyst processing. May be either *OPS* or *ALT*. The Primary is always syncing data to the *Backup*. During a *pipeline transfer*, the *Primary* and *Backup* roles switch.

1.117. Processing Sequence

An ordered grouping of *processing units* or other nested *processing sequences* connected by logic elements that specify sequencing, branching, concurrency, and entry and exit criteria. Processing sequences may specify control parameters such as data buffering and *data source*, and must adhere to an interface standard for invocation, status return, data access, logging, messaging, etc.

1.118. Processing Stage

A named group of data processing and analysis functions, used to track status of increments of work performed on time intervals and *events* through the System. The flow of data through the System, from *data acquisition*, through *automated processing* and multiple reviews, to reporting of an *event*, is defined as a series of *processing stages* (e.g. Pipeline, Analyst 1, etc.). A processing stage may define automatic sequences (see *processing sequences*), interactive-only activities, or interactive and automatic sequences. A stage description includes: a list of functions that are performed, entry criteria (time, event, or data availability triggers), and exit criteria (completion of processing, recognition of an important event, or declaration by an Analyst).

1.119. Processing Unit

A basic processing action with defined inputs and outputs. Processing units may specify control parameters such as data buffering and *data source*, and must adhere to an interface standard for invocation, status return, data access, logging, messaging, etc.

1.120. Q (Quality Factor)

A parameter used to characterize intrinsic *attenuation* for signals propagating through a particular type of material. Q is inversely proportional to intrinsic attenuation. While strictly speaking Q does not include scattering attenuation, tomographic Q models generally include the effects of both scattering and intrinsic attenuation.

1.121. QC Mask

See *waveform QC mask*.

1.122. QuakeML

A flexible, extensible and modular XML representation of seismological event data.

1.123. Quality Control (QC)

See *waveform quality control*.

1.124. Reference Event

An *event* recognized by an analyst as containing unique or important characteristics that may help in the analysis of future events that are related. For example, a nuclear test could be designated as a reference event for any subsequently detected nearby events thought to be tests.

1.125. Reference Event Database

Data for a collection of *reference events*, including both parameters and *waveforms*.

1.126. Regression Testing

Testing of a modified software component that compares current results against past results prior to the modification (for the same inputs) to find unexpected changes (regressions).

1.127. Rejected Event Hypothesis

An *event hypothesis* determined to be invalid by either the System or an Analyst. The history of rejected event hypotheses, including *signal detection* associations, are available on the System and rejected event hypotheses can be reopened by Analysts. All signal detections are unassociated from an event hypothesis when it is rejected, making those signal detections available to form other event hypotheses.

1.128. Relative Event Magnitude Estimation

The process of estimating magnitudes (see *magnitude estimation*) for *events* in a localized source region by scaling relative to a designated master event with a well-established magnitude.

1.129. Relative Event Relocation

The process of estimating the location of an event hypothesis (see *event location*) in a localized source region by comparing observations (usually travel times) relative to observations from other nearby reference event hypotheses. Relative event locations calculated in this manner can be more accurate than locations produced by independently relocating the event hypothesis. The process can simultaneously update the locations of all the event hypotheses or just the particular event hypothesis that is being analyzed, if the reference event locations are fixed.

1.130. Remote Operations Center (ROC)

A facility used for short-term data analysis and processing control. The workstations in the ROC allow for remote presentation of displays from either *OPS* or *ALT*.

1.131. Rotation

A coordinate system transform that rotates raw channel (see *channel, raw*) data from a *three component station* to align the data's axes parallel and perpendicular to a specific *azimuth* and *slowness* (i.e. ray path). Rotation produces derived channels corresponding to an arriving signal's radial and transverse ground motion. The purpose of rotation is to enhance the SNR of signals of interest.

1.132. SEED Format

Standard for the Exchange of Earthquake Data (SEED) Format. An international standard format for the exchange of digital seismological data. It was designed for use by the earthquake research community, primarily for the exchange of unprocessed waveform data between institutions. The SEED Format is described in the document SEED Reference Manual from the Incorporated Research Institutions for Seismology (IRIS) found at fdsn.org/seed_manual/SEEDManual_V2.4.pdf

1.133. Sensor Orientation

The orientation of the axis of ground motion recorded by a sensor relative to vertical (inclination) and to geographic North (declination).

1.134. Signal

A portion of a *waveform* containing information from an *event*.

1.135. Signal Association

The process of linking (associating) a set of *signal detections* from a *network* of *stations* to an *event hypothesis*, either existing or new. Association is based on consistency of observed and predicted signal detection feature measurements (e.g. arrival time, *azimuth*, *slowness*). Signal association can be done automatically by the system (see *pipeline processing*), or manually by an analyst.

1.136. Signal Characterization

The process of measuring *signal detection* features for the purposes of determining the *phase* of a signal detection, and for determining whether or not a signal detection is consistent with an *event hypothesis* (see *signal association*).

1.137. Signal Detection

A specific interval on a *waveform* marking the arrival of a *signal* of interest. Other portions of the waveform are *noise*.

1.138. Signal Detection Feature

A feature associated with a *signal detection*, e.g. arrival time, back *azimuth*, horizontal *slowness*, amplitude, frequency content, etc.

1.139. Signal Detection Feature Measurement

A measurement of a *signal detection feature*, including measurement uncertainty.

1.140. Signal Detection Feature Prediction

A prediction of a *signal detection feature*, including prediction uncertainty.

1.141. Signal Detector Threshold

The threshold that determines when a *signal detection* will be declared for a given data *channel* for a given signal detection algorithm.

1.142. Signal Enhancement

Signal processing techniques including filtering (see *filter, waveform*), *beamforming*, and three component waveform data *rotation* used to enhance the *signal* content and reduce the *noise* content of waveform data.

1.143. Single-Station Event Hypothesis

An *event hypothesis* with associated *signal detections* coming from only a single *station*.

1.144. Site

See *station*.

1.145. Slowness

A measure of the inverse apparent velocity of a wave moving across the surface of the Earth at a *station*. The inverse of slowness is the apparent velocity of such a wave. Slowness is often used in *phase* identification and is sometimes used for determining *event location*.

1.146. STA/LTA

A signal detection method that is based on a sudden change in signal power resulting from the onset of a *signal*. STA/LTA is calculated as a ratio of short-term average (STA) signal power (or energy) to a long-term average (LTA) power (or energy). The STA time window typically precedes the LTA time window. When there is no signal present, hence only *noise* in both windows, the STA/LTA ratio is ~1. When a signal is present in the STA time window and noise is present in the LTA time window, the STA/LTA ratio is > 1. A detection is declared when the STA/LTA ratio exceeds a specified threshold.

1.147. State-of-Health (SOH) Station Data

Information describing the status of *stations* that is part of the CD1.1 Format, e.g., status of sensor *channels*, instrumentation, security, timing, communication, power.

1.148. Station

An installation where monitoring sensors are installed. Multiple sensors can be installed at the same station (see *channel*). An *array* is a group of stations, but is sometimes referred to as a station.

1.149. Station Data

Data sent to the System by contributing *stations*. This includes both the ground-motion data (see *waveform*) as well as *state-of-health (SOH) station data*.

1.150. Station Magnitude

See *event magnitude, station*.

1.151. Station Magnitude Detection Threshold

The geographically-varying *event magnitude* below which a *station* is not expected to detect *events*. This can be empirically estimated as the magnitude at which a station's *signal detections* no longer follow the Gutenberg & Richter frequency-magnitude relationship predicting that for every decrease in size of 1 magnitude unit the number of events detected should increase by a factor of 10. The highest magnitude at which a station's detected events deviate from this prediction indicates the threshold below which the station does not detect events presumed to have occurred.

1.152. Station Processing

Processing of data from a *station* that is independent of processing of data from any other stations. Types of station processing include: *waveform quality control*, waveform filtering (see *filter, waveform*), *beamforming*, *signal detection*, *signal characterization,* and *phase grouping*.

1.153. Subspace Detector

A special class of waveform correlation signal detector (see *waveform correlation event processing*) in which the incoming data is compared against a suite of similar template *waveforms* (i.e. a subspace) rather than a single template waveform. The comparison is to a linear combination of the separate waveforms; the weights for each are determined dynamically as part of the detection processing to achieve the best composite match. Using a subspace detector instead of a single template can make the correlation more robust to differences in source mechanism or source location.

1.154. System Backup

The archiving of data, files, software, etc., required to restore the System to a previous state.

1.155. System Configuration

The complete set of system parameters that define the operation of the system software. Examples include sensor thresholds, filters (see *filter, waveform*), the particular version of an earth model in use, and *processing sequences*. Each instance of a system configuration is saved so the state of all parameters at any time can be recalled.

1.156. System Message

A message generated on the System and distributed to system users. System messaging is controlled by setting various parameters including level (failure, warning, notification), target (intended recipient), and triggering criteria.

1.157. System Performance

The overall performance of the system as measured by CPU usage, disk space usage, system up-time, network usage, and status of software processes.

1.158. System Restore

The restoration of the System to a previous state (see *system backup*).

1.159. System Storage Format

A collection of database table descriptions, waveform file formats, and parameter file formats that are used to store the data used by the System and are documented in the Database Design Document (DBDD).

1.160. Test Data Set

A group of artifacts that comprise the required components to define a single test. This group contains a test name (required), a unique identifier (required), the system components to be exercised in the test (required), a set of system configuration settings (optional), a pre-existing captured waveform data feed capable of being replayed on an independent subsystem (optional), alphanumeric data source (optional), a test time duration (optional), a sequence of component specific commands/event processing operations to be performed in a specific order including where in the sequence a waveform data feed is to be injected (optional), a set of calibrated test results (optional), and a test results evaluator (optional).

1.161. Test Suite

A group of *test data sets*.

1.162. Three Component Station

A seismic *station* with separate instruments measuring ground motion in three perpendicular directions: up-down, north-south, and east-west. These directions are often referred to as Z, N, E. Also referred to as a 3C station.

1.163. Time Domain Measurements

Measurements made directly on time-series data, i.e., *waveforms*. Examples of general time domain measurements include onset time, amplitude, and period.

1.164. Training Data Set

Raw waveforms (see *waveform, raw*) and *alphanumeric data* (processing results) for a specific time period that provide a good selection of known *events* suitable for analyst training.

1.165. Training Subsystem

A subset of the event analysis portion of the System that is used to interactively teach analysts how to perform the mission.

1.166. Unit Testing

A method by which individual units of source code (sets of one or more computer program modules together with associated control data, usage procedures, and operating procedures) are tested to determine if they are fit for use. (*definition from Wikipedia*).

1.167. Virtual Event Hypothesis

A trial *event hypothesis* created for analyzing waveform and alphanumeric data in an attempt to discover evidence supporting the existence of an actual *event*.

1.168. Virtual Origin

See *virtual event hypothesis*.

1.169. Waveform

A generic term for either a raw waveform (see *waveform*, *raw*) or a derived waveform (see *waveform*, *derived*).

1.170. Waveform Correlation Event Processing

A technique used to find *events* by matching current *waveforms* to waveforms of known historical events. Waveform similarity is determined using the *correlation coefficient*. When a match is found, there is high probability that a new event has occurred that is of the same *source type* and near the same location (see *event location*) as the historical event.

1.171. Waveform Format, Native

The format a *waveform* is in when it is acquired by the System.

1.172. Waveform Format, Storage

The persistent waveform storage format used by the System, as described in the Database Design Document (DBDD).

1.173. Waveform Processing

Any operation on waveform data involving signal detection (see *signal detection algorithms*), polarization measurements (see *polarization features*), time domain measurements, frequency domain measurements, *beamforming*, magnitude measurements (see *magnitude estimation*), filtering (see *filter, waveform*), waveform correlation (see *waveform correlation event processing*), and ambient noise calculations.

1.174. Waveform QC Mask

The tag applied to a segment of waveform data with a QC (see *quality control*) problem. Each QC mask includes a start and stop time and a description of the type of problem. Subsequent *waveform processing* algorithms may use this information to mask (i.e. ignore) these segments.

1.175. Waveform Quality Control

The processing of waveform data to identify problems related to *data acquisition* and/or transfer (e.g. dropouts, spikes, etc.). In particular waveform quality control is focused on identifying problems that can lead to false *signal detections* and/or to missed true signal detections.

1.176. Waveform, Derived

Output from a derived channel (see *channel*, *derived*) during a particular time interval. E.g. a bandpass filtered version of a raw waveform (see *waveform, raw*); a *beam* created by summing multiple raw waveforms from the elements of an *array*.

1.177. Waveform, Raw

Output from a raw channel (see *channel, raw*) during a particular time interval.

Sandia National Laboratories