Specifications of EMSC event ID service

Version	2.0
Status	Final
Authors	Matthieu Landès (EMSC)
Dissemination level	Public
Related project	EPOS, Grant agreement n°676564, WP 8.5
Keywords	EPOS, UNID, Event ID, Services

Document history

Date	Version	Author	Object
30/03/2017	0.1	Matthieu Landès	First draft
26/07/2024	2.0	Matthieu Landès	Update of the service

Acknowledgment

This EventID Service is made within the EPOS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676564.

Summary

Changes from version 0.1

Correct the option name “Include_info” in the documentation to match the one used by the service. And concerning the parameter “preferred only”, in the service parameter, we have renamed the previous “prefered_only’ to “preferred_only”. Associated references in the documentation are also updated.

Moreover, the backend no longer uses openresty in lua, and we have switch to the python library fastapi.

Context of this service

The intent is to provide a web service that links the event identifiers of different institutes. We have identified two main approaches. The first is to link the DOI/PID of events and the second approach is to link existing event identifiers created by seismological institutes.

DOI/PID offers the advantages of consistency, persistency of data and are external hosted. However very few institutes have DOI/PID for seismic events and the integration of these identifiers may require time to be implemented.

In this service, we pursue the second option via an automatic solution that only includes institutes with FDSN-event web services. This solution is faster to implement, does not rely on hosted data or archiving, is real time and is always updated.

For small seismic events, origin parameters and magnitude estimations may have large uncertainties and the event association between different institutes may become problematic. One solution that was used for the EMSC bulletin was to relocate events with arrivals. In the case of a web service, this computation has to be done on the server side and may be archived to facilitate future requests; but this creates the problem of identifier updates…

To avoid these kinds of difficulties and to keep this service simple and “easy to verify”, we choose to use a cost function approach and to consider only events with magnitudes higher than XX (which may be 4.5 or 5.0). From experience, this threshold is enough to simplify the event association step. Furthermore, to avoid problematic event associations with false positive results, we chose to have very conservative criteria. The drawback is that this association operation may miss candidate events.

Mapping seismic event identifiers is a difficult topic and the choice of linking existing event identifiers accessible by FDSN-event web services allows us to give a pragmatic and feasible solution, on a time scale compatible with the “implementation phase” of the EPOS project. However, the specifications of the service may need to be re-evaluated with future evolutions and future practices, in particular with the diffusion of DOI/PID.

I. Aim of the document

The aim of this document is to describe the specifications of the web service developed in order to manipulate event identifiers of several contributors and to link them with the UNID identifier used at the EMSC. The main characteristic of this service is to allow a dynamic mapping of existing event identifier without any local data storage.

The first section describes how the event identifier mapping is performed. The second presents what contributors are included in the service. The last section describes the specifications of the web service.

Glossary

EventID is used to define the identifier of a seismic event. Generally, each seismological institute that locates events has their own identification system.

Catalog (out and source): refers to the institutes "hosting" the identifier. We have for instance EMSC, USGS, INGV or ISC.

Origin parameters are a set of parameters describing the location, the depth of the source mechanism of a seismic event.

The **source event** refers to the initial event given by the user to the service.

Event association consists in searching and identifying same events in different catalogs. One earthquake may be detected by different networks and each network will locate this event with distinct origin parameters. And this leads to "distinct" events in different catalogs.

Definition of the UNID identifier

The unified identifier (UNID) allows to link together all the data related to a same event. The UNID was created at the EMSC within the NERIES project and is described in the document "UNID_deliverable_v5.pdf".

II. How the service works?

This service is designed to provide a dynamic mapping of existing event identifier (eventID). The service aims to associate the identifier of one catalog to another for a single seismic event.

This service takes into account event catalogs that provide a FDSN-event web service. For each request, the service will ask each web service to collect possible earthquakes and try to make the association to identify the same event and so to find the corresponding eventID.

This dynamic approach has the main advantage to always provide updated eventID and do not rely on hosted data (within EMSC).

1. Use case

Simply, the user gives as inputs one eventID, the associated catalog and the output catalog. The latest is the catalog that the user seeks to get the ID.

For instance, I know the EMSC ID of the Manipur seismic event, M6.7 January the 3th, 2016. This ID is 480233. I would like to know the USGS id of this event. In this case, the eventID is 480233, the source catalog is EMSC and the output catalog is USGS.

2. Workflow of the eventID service

This eventID web service is based on 2 requests on FDSN-event web services and on an event association to find the ID correspondence. The workflow can be synthetized by the Figure 1.

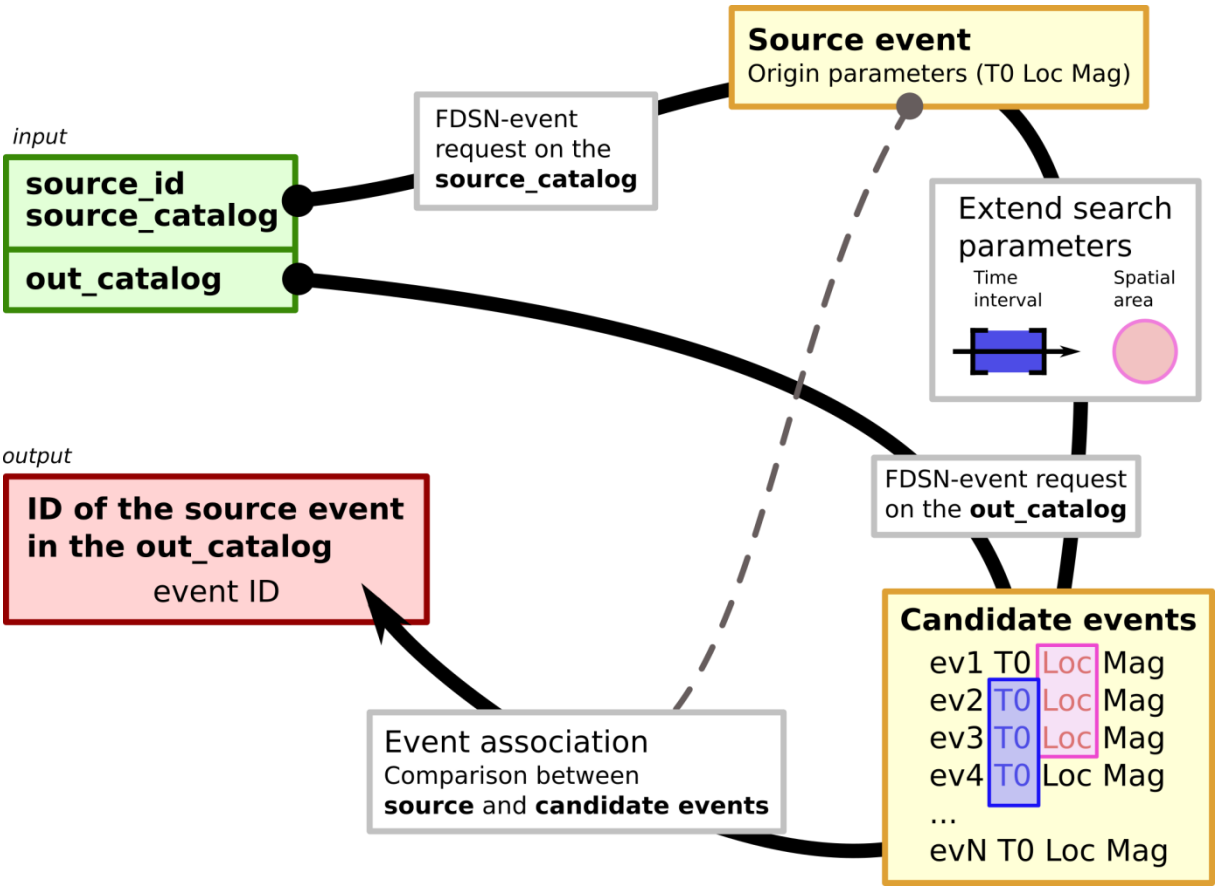

Figure 1 : Global overview of the eventID system.

The input parameters define the source event.

1. Get the origin parameters of this source event via the FDSN-Event of the source catalog. We consider the origin time, the geographic location and the magnitude.

2. Search in the out_catalog the event that is “probably” the same as the source event. From the origin parameter previously collected, we extend the time interval (parameter collect_dtime, fixed by default at 5 minutes) and the spatial area (parameter collect_dloc, fixed as a disk with a radius of one degree) and we harvest events in this time and spatial domain with FDSN-event of the out_catalog. This set of events regroups candidate events.

3. To identify the source event among the candidate events, we perform an association step to find the most probable event using a “cost function” described in the annex V.

Note that all parameters used for the harvesting and association steps can be customized through options in the web service.

III. Descriptions of event ID data

All eventID data distributed by this service are collected through different FDSN-event web service. This is the criteria that determines which event identifiers can be mapped. For now, we have incorporated 6 event identifiers and this can be easily extended.

	Web service URL	Basic web service options
unid	http://www.seismicportal.eu/fdsnws/event/1/	format=text
isc	http://www.isc.ac.uk/fdsnws/event/1	Format=xml
emsc	http://www.seismicportal.eu/fdsnws/event/1/	catalog=EMSC-RTS&format=text
usgs	https://earthquake.usgs.gov/fdsnws/event/1/	format=csv
ingv	http://webservices.rm.ingv.it/fdsnws/event/1/	format=text
gfz	https://geofon.gfz-potsdam.de/fdsnws/event/1	Format=text

Notes:

- Both EMSC and UNID event identifiers are created at the EMSC and refer to the same earthquake origin.
- When it’s available, we choose to prefer the text format for requesting the web services to minimize the cost of QuakeML parsing.
- The speed of the service depends mainly on the performance of the other web-services.

Parameters describing ID information

Event ID information are simple and are limited to 4 main parameters and other 7 additional (see below). For a given request, the information returned by the service are link to the best solutions

with his “id”, his catalog, the misfit of the association and the url to find more parameters that describes the associated seismic event.

With the web service option “include_info”, additional parameters are given. They represent convenient information to eventually check the association procedure: longitude, latitude, origin time and magnitude of the event, and the differences in time, in distance and in magnitude between the origin parameters of the event in the source and the out catalog.

Main parameters	
id	Identifier of the event
catalog	Catalog of the id
misfit	Resulting misfit of the association
url	Web url to access event information (of the catalog)

Additional parameters	
eq_lon	Longitude of the event
eq_lat	Latitude of the event
eq_time	Origin time of the event
eq_mag	Magnitude of the event
delta_time, delta_loc, delta_mag	Difference between the origin parameters of the event in the source and in the out catalog. The differences are in second for the time, in kilometer for location and in “magnitude”.

IV. Specifications of the EventID service

This service is a part of the EPOS Thematic Core Service and aims to provide a dynamic mapping of existing event identifiers from catalogs providing a FDSN-event web service.

User requests are based on **ID parameters**.

- The origin event identifier is defined by the *source_id* and the *source_catalog*. And the desired output identifier is defined by the *out_catalog*.

The user may choose to add other options.

- **Harvesting options** allow the user to customize the preselection of events in the *out_catalog* by modifying the spatial radius *collect_dloc* and the time interval *collect_dtime* centered at the reference event.

- **Association options** allow the user to modify parameters $\sigma_t, \sigma_m, \sigma_d$ used during the event association (see Annex V). They correspond respectively to parameters *misfit_dtime*, *misfit_dmag*, *misfit_dloc*.
- Requests may contain **output options**. When "preferred_only" set to true, only the event with the minimum misfit function is returned. This is the default choice. Otherwise, all events selected at the harvesting step are return. This can be useful to test the association function and to optimize misfit and collect parameters. The Boolean parameter "include_info" activate the additional parameters in the output.
  The format of the output is defined by the *format* parameter. By default, the output of the web service is in json format. The other choice is the text format, which is a csv-like format.

Table 1: Description of parameters used in the web service

	Parameter	abbreviation	Min	Max	Type	Units	default
Id parameters							
	source_id				string		
	source_catalog				string		
	out_catalog				string		
Harvesting options							
	collect_dloc		0	180	float	degree	1.5
	collect_dtime		0		float	second	60.0
Association options							
	misfit_dloc		0		float	km	105.0
	misfit_dtime		0		float	second	13.0
	misfit_dmag		0		float		0.8
Output options							
	preferred_only				boolean		true
	Include_info				boolean		false
	format		Json or text		string		

V. Annex: Event association algorithm

For a given event Ev_r and a list of events LEV, the problematic of the association is to find the event among LEV with a minimum misfit function. This misfit function is simply a distance function with Ev_r in time, magnitude and location.

For an element I in LEV, we define the misfit function as $\text{misfit}_i = {}^1/_3\left(\frac{\Delta\text{Time}_i}{\sigma_t} + \frac{\Delta\text{Mag}_i}{\sigma_m} + \frac{\Delta\text{Loc}_i}{\sigma_d}\right)$

with $\begin{cases} \Delta\text{Time}_i = |\text{T}_i - \text{T}_r| \\ \Delta\text{Mag}_i = |\text{Mag}_i - \text{Mag}_r|. \\ \Delta\text{Loc}_i = \text{Dist}(\text{Loc}_i, \text{Loc}_r) \end{cases}$

T, Mag and Loc are respectively the origin time, the geographic location and the magnitude; $\sigma_t, \sigma_m, \sigma_d$ are normalization factors.

Moreover, we define another function that estimate how many parameters are in the range of $\sigma_t, \sigma_m, \sigma_d$increased by 20%.

$$m2_i = (\Delta\text{Time}_i < 1.2\ \sigma_t) + (\Delta\text{Mag}_i < 1.2\ \sigma_m) + (\Delta\text{Loc}_i < 1.2\ \sigma_d)$$

In this expression, we suppose that true is converted to 1 and false to 0.

Finely, the association algorithm is the following:

- Among LEV, we select events such that $m2 = 3\ or\ (m2 = 2\ and\ \text{misfit} < 1)$ (1)
- and the associated event is that one with the minimum misfit estimation.

Note that the filtering (1) may reject all events and in that case, there will be no association.