REPORT

Development of a software to deliver siteXML site characterization metadata

funded by

the ORFEUS Strong-Motion Support Grant

AUTHOR: KIRIAKI KONSTANTINIDOU (KIRIAKI@ITSAK.GR)

INSTITUTE OF ENGINEERING SEISMOLOGY AND EARTQUAKE ENGINEERING

EARTHQUAKE PLANNING AND PROTECTION ORGANIZATION (GREECE)

VERSION: 0.1 (12-12-2025)

Table of Contents

1 Scope of the project

The scope of the project as described in the Agreement signed between ORFEUS and … is to develop a software to exchange site characterization metadata using the SiteXML schema proposed by the SERA project. More specifically the software should:

1. Be developed using free and open-source tools.
2. Be freely and openly available on a public software repository (e.g., Github, Gitlab).
3. Be released with an open-source license (e.g., GPL or similar).
4. Be accompanied by dedicated public documentation.
5. Be accompanied by a demonstration dataset for integration in ORFEUS coordinated platforms (e.g., ESM, RRSM, StationBook).

2 Background Information

During the SERA project, Work Package 7 was dedicated to proposing a roadmap for the exchange of site characterization metadata among network operators.

Evaluation of available site characterization metadata at permanent strong motion stations in Europe within SERA Task 7.3 clearly outlined that a large proportion of site characterization information is currently available only upon request to specific scientists or through scientific papers, which critically hampers promoting scientific research and engineering use of European seismological data. The European Geotechnical Database (EGD, http://egd-epos.civil.auth.gr/), ESM (https://esm.mi.ingv.it) or ORFEUS station book (https://www.orfeus-eu.org/opencms/stationbook/) are European services that could gather in a systematic way and with harmonized format, the site characterization information available at European seismological networks.

Thanks to international questionnaires and an international workshop, Task 7.2 has defined a list of 7 indicators considered as mandatory for a reliable site characterization and has proposed a quality metrics strategy aiming at evaluating in a quantitative way the overall quality for a site characterization analysis. The indicators that were identified are:

1. Fundamental resonance frequency, f0
2. S-wave velocity profile, Vs(z)
3. Time-averaged S-wave velocity in the upper 30 m (VS30)
4. Surface geology
5. Depth of bedrock (Engineering)
6. Depth of bedrock (Seismological)
7. Building-Code Site class (soil class)

SERA Task 7.1 focused on the description of the site condition metadata included in the European Geotechnical Database (EGD), a Europe-wide geotechnical, geological and site conditions inventory. EGD aims to collect and harmonize data from national/regional databases, for both permanently instrumented sites and non-instrumented well-studied sites.

Finally, SERA Task 7.4 proposed a strategy to disseminate site characterization information between network operators and European seismological services based on the already existing standard FDSN Station Web Service (StationXML). A site characterization XML Schema was proposed (SiteXML) that includes both site characterization information required by EGD, and site characterization most advocated indicators proposed by SERA Task 7.2. It was also proposed to use the already existing standard FDSN Station Web Service (StationXML) to disseminate the site characterization information: the URL of SiteXML would be introduced in the <ExternalReference> element in the StationXML structure.

3 SiteXML - The Site Characterization Metadata Schema

The SiteXML schema includes selected parameters for a site characterization. In general, it consists of three main parts:

- Site ownership, organizational and personal contact details
- Location information, geological, geotechnical and geophysical parameters (like geology and soil classification) that define what type of ground the site is built on.
- Geophysical parameters (like resonance frequency and velocity profiles) that define how the ground responds dynamically.

3.1 Schema update

The SiteXML schema proposed by the SERA Project (version 1.2), was evaluated during the current project. In this evaluation phase, some weaknesses were identified. Namely:

1. Missing associations between major objects like analysis, site description and velocity profiles.
2. Use of nested complex-types, which increases complexity and reduces readability.
3. Many of the nested complex types were repetitive (e.g., site indicator reference).
4. Missing resource IDs for major objects.
5. Missing schema-imposed restrictions especially for enumerated types.
6. VelocityProfile definition allowed for ambiguous content.

Following a round of technical discussions, it was decided to update the schema using the following guidelines:

1. The SiteXML document should be uniquely identified using a resource ID
2. The Schema should support more than one Analysis element, for site operators that have contacted many surveys on a site.
3. The Schema should use unique resource identifiers to associate site description with analysis and velocity profiles.
4. The Schema should include elements for preferred analysis and preferred velocity profile.
5. The Schema should impose value restrictions.

Additionally, the new schema was refactored to define and use complex and simple types for all element/object definitions, following the paradigm of StationXML and QuakeML. This way:

- There is a clean separation of type and structure.

- The schema has improved readability and modularity.
- Type derivation and substitution is possible.
- The schema is better future proofed; extensions will be easier to implement.

The actual structure and contents of the resulting XML file were affected only slightly.

The resulting schema (version 1.3) is provided for reference at the end of this document (Appendix I).

3.2 Unique Resource Identifiers

The schema includes several resource identifiers that are assigned to elements as attributes. They are used to uniquely identify the respective elements and to provide associations among them (for example associate a Velocity Profile with a specific analysis). Each site operator is responsible to provide meaningful unique identifiers, that follow the typical QuakeML Resource ID convention:

<authority>:<domain>/<resource-type>/<local-id>

For example:

- Resource identifier for a site → quakeml:isterre.fr/site/001
- Resource identifier for a site owner → quakeml:isterre.fr/siteOwner/001

Below is the description of each component:

ID Component	Description	Example
Authority	The prefix indicates that the identifier conforms to the QuakeML Resource ID convention	quakeml
Domain	This is the authority string, typically the fully qualified domain name (FQDN) of the organization responsible for issuing the ID. It ensures global uniqueness — no two authorities will issue the same local paths.	isterre.fr, itsak.gr
Resource type	This identifies the type of resource being described. This allows tools and databases to classify the nature of the resource.	site, analysis, person
Local Identifier	This is the local ID assigned by the authority (isterre.fr). It uniquely identifies the specific resource within the authority's domain. Usually numeric or alphanumeric.	001, 002, etc.

Table 1. Description of resource identifier components.

Below is a list of the Resource IDs used in SiteXML:

- **Site ID:** The unique identifier of the SiteXML document.
- **Site Description ID:** A unique identifier for the site description information
- **Analysis ID:** A unique identifier for each analysis included in the document.
- **Velocity Profile ID:** A unique identifier for each velocity profile included in the document.
- **Site Owner ID:** A unique identifier for the site owner

- **Person ID:** A unique identifier for a contact person for the site
- **Institution ID:** A unique identifier for the institution a contact person is affiliated with.

3.3 Site Indicators

All site indicators proposed by the SERA Project are included in SiteXML. For each site indicator, site operators may include the following information:

- A value that depends on the specific site indicators. It may be a real value/uncertainty pair, a fixed string value or free text.
- A reference (title, doi, url, author, date, etc). Literature or web reference providing detailed information on the methodology adopted for determining the provided value.
- A quality index.

Some site indicators may have additional properties, like the method used for determining the provided value. In Table 2 below, you can find a summary of the supported site indicators and the additional properties that are associated with them.

Property	Type	Part of
geologicalUnit	Free text	Site Description
h800	Double value/uncertainty pair	Site Description
bedrockDepth	Double value/uncertainty pair	Site Description
siteClassEC8	Enumerated value	Site Description
resonanceFrequency	Double value/uncertainty pair	Analysis
resonanceFrequencyMethod	Enumerated value	Analysis
velocityS30	Double value/uncertainty pair	Analysis
velocityS30Method	Enumerated value	Analysis
velocityS30ManualIndex	Enumerated value	Analysis
velocityS30MethodCompIndex	Enumerated value	Analysis
velocityProfile	Structure	Analysis

Table 2. Site indicators included in SiteXML.

3.4 Quality indexes

The quality index for each site indicator should be calculated using the guidelines that were proposed by SERA Deliverable 7.2. One factor that affects the quality index in all cases is the existence of a reference for the provided value (a published study). Another factor is the method used to determine the provided values (where applicable).

4 Description of the SiteXML Schema

4.1 Schema Declaration

The schema declaration defines the schema namespace as http://www.orfeus-eu.org/xml/site/1 and the schema version as "**1.3**".

```
<xs:schema                          xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns="http://www.orfeus-eu.org/xml/site/1"
targetNamespace="http://www.orfeus-eu.org/xml/site/1" version="1.3">
```

4.2 Schema Root Element

The entry element of the SiteXML schema is **`SERA_quakeml`**, defined as a complex structure (**`RootType`**) that contains all site characterization information.

```
<xs:element name="SERA_quakeml" type="RootType"/>
```

`RootType` defines the overall structure of a SiteXML document. It contains mandatory metadata such as the document creation time, information about the site owner, a detailed site description, and optional analysis objects. It also includes identifiers that uniquely reference the document and specify the schema version.

The content model consists of:

- **creationTime** *(required)* — The timestamp of when the SiteXML document was generated, expressed in ISO 8601 format.
- **externalReference** *(optional, repeatable)* — A list of external URIs pointing to additional resources containing site characterization information.
- **siteOwner** *(required)* — Information about the site owner, including personal, institutional or organizational details.
- **siteDescription** *(required)* — A single object describing the site itself, including location, geology information, and other metadata.
- **analysis** *(optional, repeatable)* — One or more analysis entries providing site condition indicators or processing results.

It also includes the following required attributes:

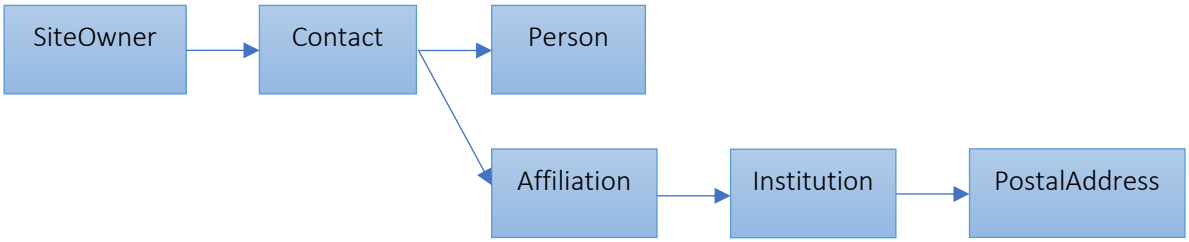
- **publicID** *(required)* — A globally unique identifier (ResourceIdentifier) for the SiteXML document.
- **schemaVersion** *(required)* — The version number of the SiteXML schema used to generate the document.

4.3 Site Owner

Each SiteXML document includes information on the site owner. This information is mandatory.

Note: The site owner structure follows closely the QuakeML 2.0 Draft.

This part of the schema includes several complex types that are hierarchically associated to form the SiteOwner element. Here is a basic diagram of this structure:



Summary of the relationships:

- SiteOwnerType → contains ContactType
- ContactType → contains PersonType + AffiliationType
- AffiliationType → contains InstitutionType
- InstitutionType → contains PostalAddressType
- PostalAddressType → contains CountryType

4.3.1 SiteOwnerType

Represents the owner of a site. It includes the following elements and attributes:

Property	Type	Required	Description
publicID	ResourceIdentifier	optional	Unique resource ID for the site owner
codeName	Free text	required	Short code name of the site owner (e.g., ISTERRE).
fullName	Free text	required	Full name of the site owner (e.g., Institut des Sciences de la Terre)
contact	ContactType	required	Contact person for the site

Table 3. SiteOwner type

4.3.2 ContactType

Provides contact information for a person, including personal info and institution affiliation. It includes the following elements:

Property	Type	Required	Description
person	PersonType	required	Personal contact information
affiliation	AffiliationType	optional	Affiliation of the person with an institution

Table 4. Contact type

4.3.3 PersonType

Personal contact details of a person. It includes the following elements and attributes:

Property	Type	Required	Description
publicID	ResourceIdentifier	optional	Unique resource ID for person
firstname	Free text	required	First name of the person
lastname	Free text	required	Last name of the person
mbox	Free text	required	Email address

homepage	URI	optional	Personal web page

Table 5. Person type

4.3.4 AffiliationType

Describes the person's institutional affiliation. It includes the following elements:

Property	Type	Required	Description
institution	InstitutionType	required	Institution details
department	Free text	optional	Department name (e.g., Engineering Seismology Department)
function	Free text	optional	Role/function within the institution (e.g., Senior researcher)

Table 6. Affiliation type

4.3.5 InstitutionType

Information about an institution. It includes the following elements and attributes:

Property	Type	Required	Description
publicID	ResourceIdentifier	optional	Unique resource ID for institution
name	Free text	required	Name of the institution
mbox	Free text	optional	Email address
phone	Free text	optional	Phone number
homepage	URI	optional	Website
postalAddress	PostalAddressType	optional	Physical address

Table 7. Institution type

4.3.6 PostalAddressType

Physical postal address. It includes the following elements:

Property	Type	Description
streetAddress	Free text	Street address (e.g., Universite Grenoble Alpes CS 40700)
locality	Free text	City (e.g., Grenoble)
postalCode	Free text	Postal code (e.g., 38058)
country	CountryType	Country information

Table 8. PostalAddress type

4.4 Site Description

This element includes site description metadata for a site. It includes location information and several geotechnical and geophysical parameters like **soil, morphology, topography, and geological classifications**.

Note: Site Description is modeled after QuakeML Draft 2.0 class

Table 9 below, summarizes the metadata included in site description:

Propery	Description
publicID	Unique resource identifier for site description parameters
latitude	Latitude of the site
longitude	Longitude of the site
altitude	Altitude of the site
minDistanceFromStation	Minimum distance between the permanent seismological station and the site characterization measurement.
maxDistanceFromStation	Maximum distance between the permanent seismological station and the site characterization measurement.
station	Station name (in case of a station characterization)
morphology	Qualitative description of the shape of the earth's surface
topography	Quantitative description of the shape of the earth's surface
geologicalUnit	Surface geology
h800	Engineering bedrock depth
bedrockDepth	Seismological bedrock depth
siteClassEC8	Building-Code Site class (soil class)
Preferred site analysis ID	Resource ID of the preferred analysis
Preferred velocity profile ID	Resource ID of the preferred velocity profile

Table 9. Site description parameters

4.4.1 Site Topography

Site operators may include two different values for site topography, according to two different schemas (For details on the schemas used see **SERA Deliverable D7.1 – Appendix I**)

Schema A is the topography description schema of the Italian Code and may take the following values:

- **T1 :** Flat surface, isolated slopes and cliffs with average slop angle less than or equal to 15°
- **T2 :** Slopes with average slope angle more than 15°
- **T3 :** Ridges with crest width significantly less than the base width and average slope angle between 15° and 30°

- **T4 :** Ridges with crest width significantly less than the base width and average slope angle more than 30°

Schema B is the one proposed by Burjanek et al. (2014). It may take the following values:

- Valley
- Lower slope
- Flat
- Middle slope
- Upper slope
- Ridge

4.4.2 Site Morphology

Site morphology is the qualitative description of the shape of the earth's surface. This also takes fixed values:

- Plain
- Valley-Basin
- Slope
- Ridge

4.4.3 Site Class EC8

Ground type according to Eurocode 8 (EC8 § 3.1.2, Table 3.1), based on the velocity s30 value and the geotechnical description.

- A
- B
- C
- D
- E
- S1
- S2
- Undefined

4.5 Analysis

The Analysis element includes information on geophysical parameters, that define how the ground responds dynamically.

Table 9 below, summarizes the metadata that may be provided in each analysis:

Propery	Description
publicID	Unique resource identifier for analysis parameters
siteDescriptionID	The Site Description this analysis object refers to
creationTime	Date that this analysis was generated or published
resonanceFrequency	Resonance frequency of the soil column at each location/site
velocityS30	Average shear-wave velocity between 0 and 30 meters depth.

velocityProfile	One or more velocity profiles associated with this analysis.
velocityProfileCount	Number of Velocity profile(s) that are available for this site.
sptLogsCount	Number of SPT log profile(s) that are available for this site
cptLogsCount	Number of CPT log profile(s) that are available for this site.
boreholeLogsCount	Number of borehole log profile(s) that are available for this site.

4.5.1 Fundamental resonance frequency

Regarding the resonance frequency value, the site operator may provide one or more methods that were used to determine the provided value (**resonanceFrequencyMethod**). The quality index for this site indicator is calculated based, among other things, on the method or methods used. This property takes specific values:

- HVSR EARTHQUAKE RECORDS
- HVSR NOISE
- SSR EARTHQUAKE RECORDS
- SSR NOISE
- INFERRED

4.5.2 Share-wave velocity in the upper 30 m (Vs30)

Regarding the shear-wave velocity (Vs30) value, the site operator may provide one or more methods that were used to determine the provided value (**resonanceFrequencyMethod**). The quality index for this site indicator is calculated based, among other things, on the method or methods used. This property takes specific values:

- Geology
- Topographic Slope
- SPT
- CPT
- Laboratory
- S-REFR
- S-REFL
- SASW
- MASW
- SWI
- SPAC/F-K
- ReMi
- Crosshole
- Downhole
- Uphole
- P-S Log
- Seismic Cone
- DH Strong Motion Arrays

Also, the site operator may provide two additional quality indexes, that affect the calculation of the final quality index for the indicator:

- velocityS30MethodCombIndex: Carries the information on whether a combination of two methods or more has been applied to estimate the Vs30 value. It is used for the estimation of the Vs30 quality index. Allowed values:
  - 1.0: if only one method has been used to estimate the Vs30 value
  - 1.2: if a combination of two or more methods has been applied to estimate the Vs30 value
- velocityS30ManualIndex: Overall qualitative factor on the knowledge of the maximum depth of Vs measurements, which is most commonly related to the depth the EC8 engineering bedrock (Vs≥800 m/s). The reasoning for introducing this index and description of its values is provided in SERA Deliverable 7.1, Appendix III. It takes the values:
  - 0.2
  - 0.4
  - 0.8
  - 1.0

4.6 Velocity Profiles

The site operator may provide one or more velocities profiles for a specific site, with the following restrictions:

- Each velocity profile belongs to an analysis.
- A reference to the preferred velocity profile is provided in the site description element.

Each velocity profile consists of many ***VelocityProfileData*** elements, one for each velocity profile layer. For each layer, the site operator can provide the following values:

- velocityP
- velocityS
- density
- layerTopDepth
- layerBottomDepth

All values are value/uncertainty pairs.

4.7 Utility Types

The SiteXML schema includes several complex and simple Types that are used all over the document. These include:

- ExternalReferenceType: This type contains a URI and a description for external data that users may want to reference in SiteXML. It is currently used only by the Root Element.
- ReferenceType: This type is used to provide reference information for site indicators. Includes two properties: ***LiteratureSource*** and ***FileResource***.
- LiteratureSourceType: This type is used to provide information on a published resource.

- **FileResourceType**: This type is used to provide a link (URI) for the literature source.
- **RealQuantityType**: Type used for all value/uncertainty pairs.
- **ResourceIdentifierType**: Type used by all resource identifiers.
- **CounterType**: Simple type for integer values > 0 (e.g. counters).

5 SiteXML ObsPy Module

The main scope of this project was to develop a software, to handle SiteXML files in order to facilitate the site characterization data exchange. The main requirements that should be met were:

1. To be developed using free and open-source tools.
2. To be freely and openly available on a public software repository (e.g., Github, Gitlab).
3. To be released with an open-source license (e.g., GPL or similar).

After a round of technical discussions, it was decided to build this software as an ObsPy module, a seismological Python library, that is widely used by the seismological community. This way all three requirements would be met, and additionally the distribution of the software with ObsPy, would enhance the adoption of the software by site operators.

Further, due to with QuakeML and some conflicts with structures already implemented in ObsPy for FDSN StationXML, it was decided to be built as a separate I/O module.

5.1 The structures

The implementation of the classes follows closely the structure of the SiteXML schema. There is one parent class (SERAsite) that includes all site characterization information for a site (ownership, site description, one or more analysis, etc). The rest of the classes implement the rest of the schema objects. In the table below you can find a summary of the classes, arguments they take and a brief description for each one.

Class Name	Arguments	Description
ValueWithUncertainty	value, uncertainty	(no description in source)
LiteratureSource	title, first_author, secondary_authors, year, booktitle, language, doi	(no description in source)
SiteIndicator	name, value, methods, quality_index, literature_source	(no description in source)
SiteCharacterizationParameters	publicID, analysis	(no description in source)

Class Name	Arguments	Description
SERASite	site_owner, site_description, publicID, analysis, overall_quality_index, created, external_references	*This is the parent class for the siteXML object tree.*

5.2 The functionality

The software supports the following main functions:

- **Validate** a SiteXML file against the schema.
- **Read** from SiteXML file and store metadata in data structures.
- **Write** a SiteXML file from metadata in the data structures.
- **Import** SiteXML metadata from either CSV or Excel files and produce the resulting SiteXML files.

Additionally, it implements a function to associate the SiteXML file for a station with the respective StationXML by adding a reference, as suggest by SERA Task 7.4.

5.2.1 Validate SiteXML

The validate functionality is implemented by the function **`validate_sitexml()`**. The function takes one argument, the file name or file-like object to read from and returns True if the validation is successful, or False otherwise. The SiteXML file is validated against the schema version that is designated in the header of the XML file.

Example:

```
if validate_sitexml("SiteXMLfile.xml"):
      continue_processing()
else:
      print("Not a valid SiteXML file")
```

5.2.2 Read SiteXML

The read functionality is implemented by the function **`read_sitexml()`**. The function takes one argument, the file name or file-like object to read from and returns a new **`SERASite`** object that contains all the site characterization information. Example:

```
sera_site = read_sitexml("SiteXMLfile.xml")
```

Main constrain: At least site owner and site description metadata should be present in the XML file to create the SERASite object.

5.2.3 Write SiteXML

The write functionality is implemented by the function **`write_sitexml()`**. The function takes three arguments:

- The **`SERASite`** object that contains the site characterization information.

- The file name or file-like object to write to.
- A flag, to optionally validate the produced XML file against the schema (default is True)

Example:

```
write_sitexml(sera_site, ”SiteXMLfile.xml“, validate=True)
```

Main constrain: The input **SERASite** object should contain at least the site owner and site description metadata.

5.2.4 Import from CSV/Excel

The import functionality is implemented by the function **csv_to_sera_site()**. The function takes four main arguments and returns a dictionary of **SERASite** objects. The unique resource identifier of the site is used as the dictionary key.

- **site_owner_csv:** An input CSV file to read site ownership information. This file should contain only one line with the site ownership information.
- **site_description_csv**: An input CSV file to read site description information. This file should contain one line per site.
- **analysis_csv**: An input CSV file to read analysis metadata. This file should contain one line per analysis (could be many lines per site)
- **velocity_profiles_csv**: An input CSV file or a directory that contains any number of CSV files with velocity profiles. The site operator can choose between using a single file for all velocities profiles or separate files per site.

Main constrains:

- The input csv/excel files should use specific column names.
- At least site owner and site description metadata should be provided for the import function to proceed.
- The use of resource identifiers for all main structures (site ID, site description ID, analysis ID and velocity profile ID) is required for the software to make the correct associations.

Sample csv/excel files are provided in Appendix II of this report.

6 Documentation

Part of this project is to provide comprehensive documentation on SiteXML and usage of the software.

Regarding the first one, it was decided to follow the StationXML paradigm and include all documentation regarding the schema inside the schema file using the <annotation> and <documentation> tags. The <documentation> “source” attribute has been used to add external references where needed. The documentation is then produced in HTML format using an open-source XSLT stylesheet (xs3p, https://github.com/neeraj9/xs3p) that was adapted for the needs of this project.

The documentation will be uploaded in a location decided by the ORFEUS Strong Motion Committee, so it can be accessible by all users of the software.

Regarding the software, the ObsPy guidelines were followed to provide proper documentation in the classes and functions using **docstrings.** This documentation includes:

- A description for each function or class.
- A description and type of all input arguments.
- A description and type of the returning object of a function.
- Usage examples

This way, the documentation of the SiteXML module, will be integrated with the rest of the ObsPy documentation and it will be available at https://docs.obspy.org.

7 Dissemination of SiteXML

It was decided that each site operator, should put an effort in using the SiteXML schema and this software to produce site characterization metadata in a homogenized way. Since the schema is compatible in a large degree with the QuakeML concept model and with the European Geocharacterization Database (EGD), the produced XML files could then be imported in various site characterization metadata databases (ESM, StationBook, RRSM, EGD).

The update of the SiteXML files with new information, should be caried on by the network operators. The creation date of the SiteXML, designates whether it has been updated or not.

Appendix II – SiteXML Reference

8 RootType

Element	Type	Occurrence	Description
creationTime	xs:dateTime	1	Generation timestamp of the document, in ISO 8601 format.
externalReference	ExternalReferenceType	0..∞	URI references to external site-related resources.
siteOwner	SiteOwnerType	1	Metadata describing the site owner (person/institution).
siteDescription	SiteDescriptionType	1	Narrative and structured description of the site.
analysis	AnalysisType	0..∞	Optional analysis entries documenting site condition indicators or results.
Attribute	Type	Use	Description
publicID	ResourceIdentifier	required	Unique identifier of the SiteXML document.

Element	Type	Occurrence	Description
schemaVersion	xs:string	required	Version identifier of the SiteXML schema.