Open Geospatial Consortium

Publication Date: 2014-02-25

Approval Date: 2013-06-22

Posted Date: 2013-02-28

Reference URL for this document: http://www.opengis.net/doc/BP/obs-via-sos/1.0

Reference number of this OGC® project document: OGC 13-015

Version: 1.0

Category: OGC® Best Practices

Editor: EO2HEAVEN Consortium

OGC Best Practice for Sensor Web Enablement: Provision of Observations through an OGC Sensor Observation Service (SOS)

Warning

This document defines an OGC Best Practices on a particular technology or approach related to an OGC standard. This document is not an OGC Standard and may not be referred to as an OGC Standard. It is subject to change without notice. However, this document is an official position of the OGC membership on this particular technology topic.

Document type:	OGC® Best Practice
Document subtype:	
Document stage:	Approved for publication as a BP
Document language:	English

Preamble to the "OGC Best Practices for Sensor Web Enablement: Provision of Observations through an OGC Sensor Observation Service (SOS)"

An important aspect in order to improve the applicability of the Sensor Web technology to application domains such as environment and health is to facilitate the practical use of the OGC Sensor Web Enablement (SWE) framework. This document reflects experiences and recommendations when using OGC SWE concepts in the research project EO2HEAVEN (Earth Observation and Environmental Modelling for the Mitigation of Health Risks), co-funded by the European Commission as part of the 7th Framework Programme (FP7) Environmental theme. EO2HEAVEN contributes to a better understanding of the complex relationships between environmental changes and their impact on human health.

This document focuses on one basic issue: the provision of observations in an OGC SOS.
This includes the definition of a lightweight profile of the OGC Sensor Observation Service (SOS), an analysis of and contribution to the specification of the SOS 2.0 as well as an approach how the data used within Earth observation (EO) applications can be integrated more easily into SOS instances. It focuses on sensor data (observations) originating mainly from in-situ or mobile sensors providing measurements that may be mapped to a tabular format.

The intended audience of this document includes system architects, information modelers and system developers engaged in designing sensor service networks and related applications taking into account relevant OGC standards.

For more details on EO2HEAVEN see the project Web site at http://www.eo2heaven.org . The essence of this document relies upon the contents of the EO2HEAVEN public deliverable D4.8 "Specification of Advanced SWE Concepts". For further EO2HEAVEN specifications see http://www.eo2heaven.org/category/documents-categories/public-deliverables.

The lightweight OGC SOS profile was developed in close cooperation between the FP7 projects EO2HEAVEN and UncertWeb (see http://www.uncertweb.org).

SEVENTH FRAMEWORK PROGRAMME

THEME 6

ENVIRONMENT (INCLUDING CLIMATE CHANGE)

EO2HEAVEN

Earth Observation and ENVironmental modelling for the mitigation of HEAlth risks

244100 - EO2HEAVEN CP-IP

OGC Best Practices for Sensor Web Enablement: Provision of Observations through an OGC Sensor Observation Service (SOS)"

March 2012

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 244100

Document Control Page

Title	OGC Best Practices for Sensor Web Enablement: Provision of Observations through an OGC Sensor Observation Service (SOS)
Creator	Ed.: Thomas Usländer (Fraunhofer IOSB)
Description	This document comprises experiences and recommendations when using Sensor Web Enablement (SWE) concepts. This document focuses on one basic issue: the provision of observations in an OGC SOS. This includes the definition of a lightweight OGC SOS profile (OGC 11-169r1), an analysis of and contribution to the specification of the Sensor Observation Service (SOS) 2.0 as well as an approach how the data used within Earth observation (EO) applications can be integrated more easily into SOS instances. These recommendations result from the work performed in 2010-2013 as part of the research project **EO2HEAVEN** (**E**arth **O**bservation and **Env**ironmental Modelling for the Mitigation of **Hea**lth Risks), co-funded by the European Commission as part of the 7th Framework Programme (FP7) Environmental theme. EO2HEAVEN contributes to a better understanding of the complex relationships between environmental changes and their impact on human health. See http://www.eo2heaven.org/ . The lightweight OGC SOS profile has been developed in close cooperation between the FP7 projects EO2HEAVEN and UncertWeb (see http://www.uncertweb.org/).
Publisher	EO2HEAVEN Consortium
Contributors	Bröring, Arne (52°North) Hollmann, Carsten (52°North) Jirka, Simon (52°North) Jürrens, Eike Hinderk (52°North) Kunz, Siegbert (Fraunhofer IOSB) Osmanov, Artur (52°North) Schnürer, Raimund (52°North) Simonis, Ingo (OGCE) Stasch, Christoph (IfGI, University of Münster, on behalf of the UncertWeb project) Usländer, Thomas (Fraunhofer IOSB) Watson, Kym (Fraunhofer IOSB)
Type	Text
Format	Microsoft Word 2003
Language	en-GB
Creation date	08/01/2013
Version number	1.0
Version date	27/02/2013
Last modified by	Thomas Usländer (Fraunhofer IOSB)

Revision history

Version	Date	Modified by	Comments
0.1	08/01/2013	Usländer, Thomas (Fraunhofer IOSB)	Initial version.
0.2	01/02/2013	Usländer, Thomas (Fraunhofer IOSB)	Draft based upon input from partners
0.3	06/02/2013	Usländer, Thomas (Fraunhofer IOSB)	Consolidated Draft
0.4	07/02/2013	Jirka, Simon (52°North)	Proofreading and minor revision.
0.5	11/02/2013	Usländer, Thomas (Fraunhofer IOSB)	Consolidated Draft
1.0	27/02/2013	Usländer, Thomas (Fraunhofer IOSB) Jirka, Simon (52°North) Stasch, Christoph (IfGI, University of Münster)	Update of Annex A: OGC SOS lightweight profile (OGC 11-169r1) Overall consolidation

Table of Contents

1 Architectural Context 7
2 Provision of Observations 8
2.1 Sensor Observation Service (SOS) Version 2.0 8
2.1.1 Introduction 8
2.1.2 Challenges and Questions Addressed by the SOS 2.0 Standard 8
2.1.3 Summary of the SOS 2.0 Specification 9
2.1.4 Summary and Conclusion 10
2.2 Lightweight SOS Profile 11
2.2.1 Introduction 11
2.2.2 Challenges and Questions 11
2.2.3 Results 11
2.2.4 Conclusion 15
2.3 Import of Sensor Data Archives into the SOS 16
2.3.1 Introduction 16
2.3.2 Challenges and Questions 16
2.3.3 Results 16
2.3.4 Summary and Conclusion 32
3 References 33
4 Abbreviations and acronyms 35
5 Annex A: SOS Lightweight Profile 36

Index of figures

Figure 2-1: Choosing a CSV file on the local file system containing the observation data 19
Figure 0-2: Choosing a CSV file on a FTP server 20
Figure 0-3: Selecting settings for parsing the CSV file 21
Figure 0-4: Describing the content of the columns (date and time) 22
Figure 0-5: Describing the content of the columns (measured value)... 23
Figure 0-6: Creating associations between measured values and features of interest.................. 24
Figure 0-7: Adding missing metadata (missing time zone) 25
Figure 0-8: Adding missing position information (EPSG code and height) 25
Figure 0-9: Automatically generating feature of interest identifiers 26
Figure 0-10: Manually adding information about the unit of measurement 27
Figure 0-11: Creating sensor identifiers 28
Figure 0-12: Setting the positions of the features of interest through a map view 29
Figure 0-13: Setting the positions of the features of interest through text fields 30
Figure 0-14: Setting the URL of a Sensor Observation Service.. 31
Figure 0-15: Import finished 32

Index of Tables

Table 1 Supported Observation Types of the Lightweight SOS Profile 13
Table 2 Different steps of the application 18

1 Architectural Context

The OGC Sensor Web Enablement (SWE) architecture (OGC, 2008c) defines a conceptual approach to build a Spatial Data Infrastructure (SDI) for sensor data. This architecture provides the conceptual foundation and standard framework of the work presented in this document.

In addition to the OGC SWE architecture, the present recommendations and tools rely upon the architectural work that has been carried out in a series of European research projects and submitted to OGC as best-practices work follows:

- Reference Model of the ORCHESTRA Architecture (RM-OA) (OGC, 2007c), accepted as OGC best-practices paper in 2007, that instantiated and tailored the OGC Reference Model for the domain of environmental risk management following the ISO Reference Model for Open Distributed Processing (RM-ODP) (ISO/IEC 10746-1:1998), and
- the Sensor Service Architecture (SensorSA) (OGC, 2009d), accepted as OGC discussion paper in 2009, that extended the RM-OA towards a sensor-based environment following the OGC SWE architecture.

The EO2HEAVEN architectural work has continued this architectural specification line in its multi-part specification of a Spatial Information Infrastructure (SII) (EO2HEAVEN, 2011a) integrating both in-situ and space-borne EO sensors. The present document is an excerpt of the EO2HEAVEN SII focusing on the advanced SWE concepts (EO2HEAVEN, 2011b) for the provision of observation in SOS instances.

2 Provision of Observations

This paper provides recommendations on the provision of observation data in a consistent way. It covers the following three main aspects:

1. In section 2.1 the OGC Sensor Observation Service (SOS) 2.0 (OGC, 2012a) standard is introduced as a web service interface offering interoperable access to sensor data. The emphasis in this section is put on the specific advancements of the SOS 2.0 compared to version 1.0.
2. In order to facilitate the application of the SOS 2.0 standard, section 2.2 recommends a lightweight SOS profile for stationary in-situ sensors that focuses on the SOS core elements.
3. Finally, section 2.3 introduces an approach how data owners can easily publish their data on SOS servers through a graphical import tool.

2.1 Sensor Observation Service (SOS) Version 2.0

2.1.1 Introduction

Standardized access to sensor observations and sensor metadata is provided by the OGC SOS. Observations offered by the SOS can be collected by physical or virtual sensors (e.g. simulations or results of model calculations). The service acts as a mediator between a client and the sensor data archive or a physical sensor system. The heterogeneous communication protocols and data formats of the associated sensors are hidden by the standardized interface of the SOS. Sensor data requested by a client are returned as observations following the structure of the Observations and Measurements (O&M) model (ISO 19156:2011). The interface of the SOS supports access to heterogeneous sensor types, stationary as well as mobile sensors which gather their data in-situ or remotely.

2.1.2 Challenges and Questions Addressed by the SOS 2.0 Standard

This paper recommends the use of the SOS 2.0 (OGC, 2012a) as it is an evolutionary improvement of the SOS 1.0 standard. Based on the experiences gained with the SOS 1.0 specification (OGC, 2007a) the SOS 2.0 standard addresses several challenges. These are summarized as follows and described in more detail in the following sub-section:

- Complexity reduction
  - Division into core, extensions, and requirements (classes)
  - Introduction of a key-value-pair (KVP) binding
  - Definition of O&M as default and mandatory response format for observations
- Capabilities redesign
  - One sensor per observation offering to uniquely describe relationships between sensors, observed properties, and features
  - New section: InsertionCapabilities for providing all metadata necessary for inserting new sensors and observations
  - Features of Interest are not listed in Contents anymore to reduce the length of the GetCapabilities responses
- Result handling redesign
  - New operations: InsertResult and InsertResultTemplate
  - New operation: GetResultTemplate
- Advanced feature retrieval

2.1.3 Summary of the SOS 2.0 Specification

The development of the final version of the SOS 2.0 specification was completed in 2012 and the final standard document was published in the same year (OGC, 2012a).

The main advancements from SOS 1.0 to SOS 2.0 are described in the following.

Complexity Reduction

A central goal of the development of SOS 2.0 has been the reduction of complexity of the standard to facilitate its application in software development. In this respect and by aiming at a better readability of the specification document, SOS 2.0 divides its operations and functionalities into a specification core and its extensions. The core comprises the mandatory operations for retrieval of the service metadata and its content (GetCapabilities), for querying observations (GetObservation), and for accessing sensor descriptions (DescribeSensor). The transactional extension contains operations for inserting new sensor descriptions and sensor observations. The result handling extension specifies operations for insertion and retrieval of pure observation results without observation metadata to increase performance and scalability. The enhanced extension amends the SOS functionality by providing optional operations, e.g. to enable the retrieval of observed features. By following OGC's new modular specification model (OGC 2009a) the core and extensions of the SOS 2.0 specification consist of requirement classes containing requirements, which convey criteria to be fulfilled if compliance with the standard is claimed. This new design of the specification is more precise and facilitates conformance testing of compliant software components.

In order to facilitate the usage of the SOS, version 2.0 of the standard adds an http GET binding for selected operations. The operation parameters are passed to the service as key-value pairs in the URL of the service endpoint. This binding is reduced in complexity, but also in functionality, compared to the usual XML-based SOAP binding and simplifies the usage of the SOS. In the future, also a RESTful SOS binding would be desirable as mentioned in section 2.2.

The SOS 1.0 specification did not define any response format as mandatorily supported. To restrict this generality and eventually improve interoperability the SOS 2.0 standard defines O&M as its mandatory and default response format for sensor data. Other response formats may still be supported by SOS servers. However, a formally accepted extension of the standard has to define how the service behaves in detail when responding in that format. The SOS 2.0 interface still does not define a default and mandatory response format for the retrieval of sensor metadata. However, SensorML (OGC, 2012b) is mentioned as the recommended format. The Transducer Markup Language (TML), named in the SOS 1.0 specification document as a potential response format, is not mentioned in the specification anymore. This decreases the options and increases interoperability among SOS 2.0 servers.

Capabilities redesign

A step towards simplifying the standard and streamlining the different SWE services is the introduction of the SWE Common specification (OGC 2008b) upon which the SOS 2.0 specification is based. A main part of that specification is the description of a common, SWE-wide model for the part of the self-describing metadata document of a service, also known as capabilities document, which describes the content of a SWE service. For this contents section of the capabilities document, abstract types are defined and reused by the SOS 2.0.

Note: Other SWE services such as the Sensor Planning Service (SPS) 2.0 (OGC, 2009c) rely upon the SWE Common specification as well.

The contents offered by a service are grouped into so-called offerings. In case of the SOS it is an observation offering. The SOS 1.0 specification has already used this concept. However, a redesign of the offering type restricts it now to aggregating only the observations gathered by exactly one instead of multiple sensor systems. Formerly, it has been up to the SOS provider to group observations to offerings. This could have been done by different criteria: spatially, thematically (e.g., per sensor or observed property), or temporally. The simple conceptual change of limiting the offering to one sensor system eases the set-up and the access to an SOS server, since grouping of observations to offerings is not ambiguous anymore.

An important concept within the SWE framework defined by the O&M model is the feature of interest. A feature of interest is the computational representation of a real-world entity modeled with a certain set of properties. This could be for example the feature “Gulf of Mexico”. Also, a sampling point “P_42” within the Gulf of Mexico, where a measurement was taken by a certain (maybe mobile) sensor system, is a feature of interest. Both could have properties such as water depth, salinity or geometry. In case of the SOS 1.0 standard all features of interest of the aggregated observations needed to be listed for each observation offering. This is helpful to provide clients a list of features for which observations can be requested. However, this listing of all features has been identified as a problem for mobile sensor systems (e.g., a boat taking measurements in the Gulf of Mexico), which create many sampling features (e.g., sampling points) during operation. Those sampling features could accumulate to huge numbers and could increase the capabilities document up to an unusable size. Hence, SOS 2.0 servers do not list sampling features of interest in their capabilities document anymore. Instead, only those features of interest shall be listed to which the sampling features belong to. In the example above, this would be “Gulf of Mexico” and not “P_42”. Listing those sampled features of interest supports the discovery of observations.

Result handling redesign

The SOS 1.0 standard has already supported the retrieval of the pure observation results for a specified timestamp without the complete set of associated observation metadata. The purpose of this functionality is to allow clients to repeatedly obtain sensor data without receiving responses that largely contain the same data except for a new timestamp and result value. This is in particular useful in scenarios with restricted bandwidth or for SOS clients with restricted processing power. The SOS 2.0 specification redesigns and simplifies the GetResult operation involved in this retrieval of pure observations results. In particular, the response from the SOS server containing the results is defined in a more precise way. An additional operation is introduced (GetResultTemplate) which returns an exact description of the structure and the encoding of the results by making use of the SWE Service Common 2.0 data model.

A functionality added by the SOS 2.0 standard is the insertion of pure observation results by means of new operations (InsertResultTemplate and InsertResult). These allow client components to insert sensor results into an SOS without the need to repeatedly transmit the entire set of observation metadata. Similar to the result retrieval functionality, this functionality is useful if the communication bandwidth of the client, in this case the sensor data producer, is limited.

Also, the capabilities model of the SOS 2.0 has been improved to better support the insertion of new data to SOS servers. A new section of the capabilities document (called insertion capabilities section) now states the observation type, result type, feature types, and encodings supported by a SOS server for insertion of results or observations.

Advanced feature retrieval

However, clients still need to be able to retrieve a list of sampling features from the SOS. The knowledge about existing sampling features is, for example, necessary for the construction of queries for sensor observations. For the retrieval of features of interest a separate operation, called GetFeatureOfInterest, has already been defined in the SOS 1.0 specification. In version 2.0 of the SOS standard this operation is extended in its parameterization. Not only a specific feature, or features for a certain spatial filter can be requested, also features which are observed by specified sensors or which carry specified observed properties can be requested. And further, SOS 2.0 servers allow the clients to request all stored features of interest.

2.1.4 Summary and Conclusion

SOS 2.0 is a valuable evolutionary advancement of the SOS 1.0 standard. Its improvements are expected to be very useful for implementing efficient mechanisms for sensor data access as well as sensor data publication. Special advantages of SOS 2.0 are the more compact approach of the capabilities as well as the unique definition how the offering concept has to be interpreted. This way, in combination with the lightweight SOS profile (section 2.2) interoperability between different SOS servers and clients is increased. Furthermore, the advanced feature handling capabilities of SOS 2.0 make it possible to build advanced client applications that allow filtering of observations based on spatial criteria.

2.2 Lightweight SOS Profile

2.2.1 Introduction

The Sensor Observation Service (SOS), in its version 1 (OGC, 2007a) and in its version 2 (OGC, 2012) as assessed in section 2.1, plays a central role in the OGC SWE architecture as it defines an interface for accessing sensor data and metadata.

However, the SOS specification and the most important additional standards it relies on, are relatively complex to implement and use. These encompass the Observations and Measurements (O&M) specification for conceptualizing (ISO 19156:2011) and encoding measurement data (OGC, 2011a) as well as the Sensor Model Language (SensorML) (OGC, 2007b) for encoding metadata,

The reason for this is the extremely broad range of applications that these standards are supposed to support. On the one hand, this makes it possible to flexibly apply SWE to many different applications, on the other hand, the effort for understanding, implementing and applying the SWE standards increases considerably.

This section introduces an approach for a lightweight profile of the SOS by focusing on a set of commonly used elements of these standards and thus reducing the complexity of the specification.

At a later stage this work may serve as a basis for developing domain-specific SWE profiles.

2.2.2 Challenges and Questions

This profile was designed in a way that is, on the one hand, efficient and easy to implement, and, on the other hand, standard compliant. Especially the following challenges were considered during the creation of the profile:

- Reducing the number of operations: certain operations of the SOS standard were designed for very specific needs; these operations were left out of this profile
- Reducing the complexity of the SOS operations (e.g. less complex filters for requesting sensor data)
- Limiting the supported data formats to those used in practice (i.e. no support of the Transducer Markup Language (TML) within the SOS)

In summary, the main aim that guided the development of this profile was

- to support those use cases which are regularly occurring in practice, and
- to leave out those with very specific requirements that go beyond the mainstream of SWE use cases.

In the next section, it is described how a lightweight profile of the SOS may look like. The profile described in this best practices paper is based on the 2.0 versions of the different SWE specifications. A more detailed description of this profile is attached as Annex A to this document and can also be found in (OGC, 2011b).

2.2.3 Results

The OGC SOS defines standardized interfaces for accessing sensor data and metadata. Thus, the SOS may be considered as the core element of the SWE architecture. It provides a means for integrating sensor data into Spatial Data Infrastructures like it is already possible for other kinds of geospatial data such as maps (through the OGC Web Map Service (WMS, OGC 2006), raster data (through the OGC Web Coverage Service (WCS, OGC 2008a) and geospatial objects (through the OGC Web Feature Service (WFS)).

The following subsections explain which parts of the SOS need to be considered for a minimum profile that reduces its complexity and increases the efficiency.

2.2.3.1 SOS Operations

This lightweight SOS profile comprises those operations, which are mandatory within the SOS standard (GetCapabilities, DescribeSensor and GetObservation) so that they shall be implemented by every standard compliant SOS server. Furthermore, the GetFeatureOfInterest operation was included as it allows accessing the geometries of sensor stations.

The GetCapabilities operation is common for every OGC web service specification. It is important to support this operation as it provides the metadata necessary for clients to generate valid SOS requests. It contains especially the following information:

- Supported operations of the SOS instance
- Identifiers of all sensors encapsulated by the SOS instance
- Identifiers of all observed phenomena
- Offerings
  Note: As described in section 2.1.3, the SOS 2.0 standard requires a unique relationships between exactly one sensor and its observed phenomena.

The DescribeSensor operation allows the client to retrieve metadata about the sensors encapsulated by an SOS instance. It is a mandatory element of the SOS specification. Thus, it is included in the minimum profile so that servers must offer this operation. However, clients need not necessarily make use of the DescribeSensor operation. The request of the DescribeSensor operation is very simple as the only parameter consists of the identifier of the sensor for which a metadata document is requested. The response is a SensorML document (a simple structure for such a SensorML document is defined below in section 2.2.3.3).

The GetObservation operation is an essential part of the SOS specification as it is the operation for requesting sensor data. Thus, it is obvious that this operation is needed for both servers and clients. However, in order to reduce the complexity of the GetObservation operation several restrictions have been defined that limit the query parameters. In detail the following parameters are included in this lightweight SOS profile:

- Offering: the identifier of the offering that contains the requested sensor data
- Temporal filter for specifying the time for which sensor data is requested; however the temporal filters are limited within the lightweight SOS profile to the following types: during a time period and equal to a time instant
- Procedure: the identifiers of the sensors from which data is requested
- Observed property: the identifier of the phenomena for which data is requested
- Feature of interest: the geometry object for which data is requested; within the lightweight SOS profile this is limited to identifiers of the geometric objects for which sensor data is requested
- Spatial filter: Bounding box of an area for which data is requested

The GetFeatureOfInterest operation allows retrieving the geometries of the stations at which measurements were performed. As the access to the geometries of measurements is a requirement for creating map visualizations of the sensors and their data, the GetFeatureOfInterest operation has been included for servers in this lightweight SOS profile. However, not every client might rely on this operation (e.g. clients that only show diagrams of time series data). In order to reduce the complexity of this operation, its request parameters are restricted within this profile. Clients may request geometries using only one of these parameters:

- featureOfInterest: the identifier of the requested feature of interest
- spatialFilter: bounding box describing an area for which features are requested
- observedProperty: identifiers of the properties that shall be observed at the requested features.

- procedure: identifiers of the procedures that shall observe the requested features

In case of this profile, the response of the GetFeatureOfInterest operation shall always contain one or more SamplingPoints.

2.2.3.2 Observations and Measurements

The Observations and Measurements standard offers an encoding for data observed by sensors. However, as O&M has been designed to be extremely versatile and comprehensive it has also a certain complexity. Thus, this lightweight profile has been defined containing only those elements of O&M, which are essential for most common use cases. Within the lightweight SOS profile, the following observation types are included:

Observation Type	Result Type	Description	Example
Measurement	gml:MeasureType	scalar numerical value with unit of measurement	<om:result uom="Cel">36 </om:result>
CountObservation	xs:integer	count of an observed property	<om:result>12</om:result>
TruthObservation	xs:boolean	truth value (often existence) of an observed property	<om:result>true </om:result>
CategoryObservation	gml:ReferenceType	value from a controlled vocabulary	<om:result codeSpace= http://www.uncertweb.org/ tempCodeSpace/tempCategories.xml >warm </om:result>
TextObservation	xs:string	any kind of textual description of an observed property	<om:result>some text </om:result>
om:ComplexObservation with swe:DataArray	swe:DataArray	compact representation of multiple observation values (e.g. time series); this observation types needs to be further restricted and enhanced by guidance in future revisions of the profile.	-

Table 1 Supported Observation Types of the Lightweight SOS Profile

Note: Table 1 is taken from the OGC Discussion Paper (OGC, 2011b) and has been slightly enhanced (addition of SWE Common Observations)

In order to minimize the complexity of O&M documents, the responses of SOS servers need only to contain a limited set of O&M elements. These are:

- gml:identifier (mandatory): this element shall be provided for uniquely identifying or referring to a specific observation.
- om:phenomenonTime (mandatory): the time instant or time period for which the observation contains sensor data; within the lightweight profile this is restricted to the data types TimeInstant and TimePeriod.
- om:resultTime (mandatory): the time when the result became available.
- om:procedure (mandatory): the identifier of the sensor that has generated the observation
- om:observedProperty (mandatory): the identifier of the phenomenon that was observed
- om:featureOfInterest (mandatory): an identifier of the geometric feature (e.g. sensor station) to which the observation is associated; within the lightweight profile this is limited to sampling

points

- om:result (mandatory): the observed value; the type of the result is restricted to the types shown in Table 1.

A typical sensor observation encoded using an om:Measurement looks as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<!--=============================================================
This observation depicts an example
-point as sampling location
-time instant as phenomenon time
-result is a measure value (double with units of measure information)
==============================================================-->
<om:OM_Measurement gml:id="obsTest1" xmlns:om="http://www.opengis.net/om/2.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:sams="http://www.opengis.net/samplingSpatial/2.0"
    xmlns:sf="http://www.opengis.net/sampling/2.0"
    xmlns:xlink="http://www.w3.org/1999/xlink"
    xmlns:gml="http://www.opengis.net/gml/3.2"
    xsi:schemaLocation="http://www.opengis.net/om/2.0 ../observation.xsd">
   <!--=============================================================
   global identifier of the observation
   ==============================================================-->
   <gml:identifier codeSpace="http://www.myWebSite.org">obsTest1</gml:identifier>
   <!--=============================================================
   phenomenon time represents the time when the observation was taken (so when the
    sensor has interacted with reality
   ==============================================================-->
   <om:phenomenonTime>
      <gml:TimeInstant gml:id="ot1t">
         <gml:timePosition>2005-01-11T16:22:25.00</gml:timePosition>
      </gml:TimeInstant>
   </om:phenomenonTime>
   <!--=============================================================
   result time represents the time when the result value was produced; might differ for
    example in case of a soil sample; sample has been taken during phenomenonTime and
    then evaluate later in a laboratory during resultTime
   ==============================================================-->
   <om:resultTime xlink:href="#ot1t"/>
   <!--=============================================================
   procedure has produced the observation's result (usually a sensor)
   ==============================================================-->
   <om:procedure xlink:href="http://www.example.org/register/process/scales34.xml"/>
   <!--=============================================================
   observedProperty usually contains a link to a vocabulary containing the definition of
    the property which has been observed; in case of environmental modelling this is
    usually a physical phenomenon (e.g. temperature, air pollutant concentration, etc.)
   ==============================================================-->
   <om:observedProperty xlink:href="urn:ogc:def:phenomenon:OGC:temperature"/>
   <!--=============================================================
   feature of interest carries the sampling geometry of the observation
   ==============================================================-->
   <om:featureOfInterest>
      <sams:SF_SpatialSamplingFeature gml:id="SamplingPoint1">
         <sf:type xlink:href="http://www.opengis.net/def/samplingFeatureType/OGC-
                    OM/2.0/SF_SamplingPoint"/>
         <sf:sampledFeature xsi:nil="true"/>
         <sams:shape>
            <gml:Point gml:id="UOMlocation">
               <gml:pos srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
                  52.87 7.78</gml:pos>
            </gml:Point>
```

```xml
        </sams:shape>
      </sams:SF_SpatialSamplingFeature>
    </om:featureOfInterest>
    <!--============================================================
    observation result is an numerical value with information about the unit of
     measurements
    ==============================================================-->
    <om:result uom="Cel">36</om:result>
</om:OM_Measurement>
```

2.2.3.3 Sensor Model Language

The Sensor Model Language (SensorML) is used within the SOS for encoding the sensor metadata documents that are returned in case of DescribeSensor requests.

This lightweight profile defines a minimum set of metadata that shall be provided in a SensorML document. Complex elements of SensorML that are intended for very specific applications are not considered within this lightweight profile but may be added if necessary.

The following metadata items are considered mandatory for a SensorML document:

- gml:description: A short textual description of the sensor or sensor system
- gml:identifier: A unique identifier of the sensor system
- sml:keywords: Terms which help to describe the sensor system and serve for discovery purposes (e.g. observed phenomena, name of the sensor location)
- sml:identification: A short and a long name referring to the sensor system
- sml:classification: At least a classifier identifying the sensor type
- sml:contacts: Contact information about the operator of the sensor
- sml:featuresOfInterest: abstraction of the real world entity, the feature of interest, which is observed by the sensor system. In case of this profile, the feature of interest is a station and modelled as a SamplingPoint.
- The outputs of the sensors (identifiers of the observed phenomena and if necessary the units of measurement)

A further restriction is that every sensor shall be modelled as a SensorML system. Other types are not considered for this profile.

2.2.4 Conclusion

In summary this section has provided a pragmatic approach how the complexity of the SOS, O&M and SensorML standards can be reduced by defining according minimum profiles. This makes it easier to apply SWE concepts more quickly and with less overhead for familiarizing with the specifications. Although these minimum profiles reduce the flexibility of these standards, we expect that most common applications can be covered by these profiles.

A future version of this lightweight profile may include a RESTful SOS binding in order to provide an additional simplified means for accessing SOS servers.

2.3 Import of Sensor Data Archives into the SOS

2.3.1 Introduction

Within various projects and initiatives many existing (sensor) datasets are used which cover time periods in the past. However, these datasets are currently only available in form of existing files (e.g. comma separated values) or local databases. In order to integrate these datasets into systems which make use of distributed SDI components, it is necessary to provide the existing sensor data though standardized web-based interfaces, i.e. the OGC Sensor Observation Service (SOS) (OGC, 2007a and OGC, 2012). The following section investigates how the import of these sensor data archives into SOS instances may be facilitated to turn into a task manageable for non-IT experts.

The current implementation of the SOS Import tool was developed in the course of the EO2HEAVEN project and is available online at: https://wiki.52north.org/bin/view/SensorWeb/SosImporter

2.3.2 Challenges and Questions

Currently, a significant amount of knowledge is required when setting up an SOS instance filled by existing sensor data. Usually the users have three options for publishing their sensor data:

- Usage of the transactional SOS operations: In this case the user has to convert the data into the Observations and Measurements (O&M) data model (ISO 19156:2011 and OGC, 2011a) such that the data can be inserted into the SOS instance through standardized operations. However, a domain expert should not be obliged to spend the effort to understand the O&M model before being able to map the existing sensor data to the O&M model.
- Direct insertion of sensor data into SOS databases: In this case the existing data is transmitted into the database of the SOS, e.g. by means of SQL statements in case of a relational database. However, this approach is neither suitable for non-IT-experts. Users should not be obliged to spend time to understand the data model of a proprietary SOS database, familiarize with SQL and create according SQL statements.

  Note: The data model of the database "behind" the SOS interface is not standardized.
- Connection of existing databases with SOS instances: In this case a SOS implementation would be customized in order to use a specific database as source for observation data. But as this approach would require significant programming work, it is not suited to non-IT-experts as well.

The following section describes an approach that facilitates the import of sensor data into SOS instances. The envisaged solution is a wizard tool that guides users through the import process of sensor data. The users are able to describe the structure of the existing data through a graphical user interface such that the wizard is able to automatically create a mapping of the existing datasets into the SOS/O&M/SensorML data models, respectively.

After specifying the structure of the datasets, the wizard pushes the sensor data into the SOS instance in a final step.

2.3.3 Results

The current version relies upon comma-separated values (CSV) files as the input for the application. CSV files are a common approach to represent tabular data in a textual format. Outside the Sensor Web, they are often used to exchange sensor data. The structure of CSV files has never been standardized, though there are some guidelines (Shafranovich, 2005):

- Columns are separated with a consistent character and rows are separated with newlines.
- Traditionally, a comma delimits columns.

- In countries where decimal numbers are already separated with commas, other characters, like the semicolon or colon, are taken.
- Fields that contain a special character (e.g. a column separator or a newline) are enclosed by two quote symbols.
- Often, double quotes are used here. If a field includes a quote symbol, it is escaped by placing another quote symbol in front (e.g. “”best“” for “best”).
- Sometimes, comment characters (e.g. #) may appear which escape a whole line.

It is assumed that the content of the CSV file, measurements and sensor metadata, is stored column-wisely. For example, it could look like this:

01.01.2011; Sensor1; 1,3

01.01.2011; Sensor2; 2,5

02.01.2011; Sensor1; 1,7

02.01.2011; Sensor2; 2,2

…

At least one column of measurements should be included in the file. The results of measurements can be numeric, positive integers (counts), Booleans (true or false) or textual. For numeric values (e.g. 2.3), thousands and decimal marks have to be specified in addition. This is necessary since regional differences exist for these separators. For example, in Germany a decimal comma is used, whereas the United Kingdom and the United States use a decimal point.

In addition to the measurement column, there can be one or more date and time column in arbitrary format. Date and time could be split over several columns (e.g. date in one, time in another one). Moreover, date and time need not be complete to be compliant to OGC SWE standards. For instance, there are only times and no dates, or the time zone is missing. Analogously, this applies to sensor locations, too, in case they were provided within a CSV file. The format in which sensor positions are described can be arbitrary, so patterns for converting them into a standardized positional scheme have to be specified. Position data could be available partitioned in several columns (e.g. latitude in one, longitude in the other one) and, again, need not be complete. For example, only latitude and longitude are given and neither height, nor a spatial reference system. Also, the units of latitude, longitude and height might be missing.

It is assumed that within one column the same format is used (e.g. comma as a decimal separator or throughout the same date and time pattern). Other values than these (e.g. NULL-values or headings) will be ignored. Further columns in the CSV file might include units of measure, observed properties, sensors or features of interest. It is not very likely that URIs can be found here, so these columns will be interpreted as names. In the scope of the work described in this document, only stationary (not mobile) and in-situ (not remote) sensors are taken into account.

The application makes use of the wizard design pattern which guides the user through different steps. These and their purposes are briefly characterized in the table below. Screenshots of all steps are shown below in Figure 2-1 to Figure 2-15.

Step	Function
Step 1	Choose a CSV file from the file system to publish in a SOS instance or specify a CSV file on a FTP server that shall be (automatically/regularly) fetched by the SOS importer.
Step 2	Provide a preview of the CSV file and select settings for parsing (e.g. which character is used for separating columns)
Step 3	Display the CSV file in tabular format and assign metadata to each column (e.g. indicate that the second column consists of measured values). Offer customizable settings for parsing (e.g. for date/time patterns)
Step 4	Solve ambiguities: In case of more than one date/time, feature of interest, observed property, unit of measurement, sensor identifier or position has been identified in step 3, select the correct associations to the according measured value columns (e.g. state that date/time in column 1 belongs to the measured values in column 3 and date/time in column 2 belongs to the measured values in column 4). When there is exactly one appearance of a certain type, automatically assign this type to all measured values
Step 5	Check available metadata for completeness and ask the user to add information in case something is missing (e.g. time zone, EPSG-code for positions)
Step 6	When there is no metadata of a particular type present in the CSV file (e.g. sensor id), let the user provide this information (e.g. name and URI of the observed property or sensor, missing position data)
Step 7	Enter the URL of a Sensor Observation Service where measurements and sensor metadata in the CSV file shall be uploaded to
Step 8	Assemble all information from previous steps and convert the CSV file into XML files according to OGC's Observations and Measurements and SensorML specifications. Register sensors and insert observations at the given Sensor Observation Service using the Transactional Profile of the SOS specification. Show the progress and provide a report of errors and success

Table 2 Different steps of the application

Note that steps 4 to 6 can sometimes be skipped in case of CSV files of a particular structure. Also, their functionality is split up in the application according to the type of metadata they represent (a - date/time, b - features of interest, observed properties, units of measurement, sensors, c - positions).

In the first step, the user selects the file (in this version a CSV file), which contains the observation data set that shall be published. This can either be a CSV file on the local file system (Figure 2-1) or a file on an FTP server (Figure 2-2). In case the user chooses a file on a FTP server, the SOS Importer can be configured to automatically check in regular intervals if the file has been updated. Furthermore, it is possible to provide regular expressions for the file name so that the SOS Importer can also handle those cases in which new CSV files are created for certain time periods (e.g. new CSV files created every day that contain the creation date in their file name).

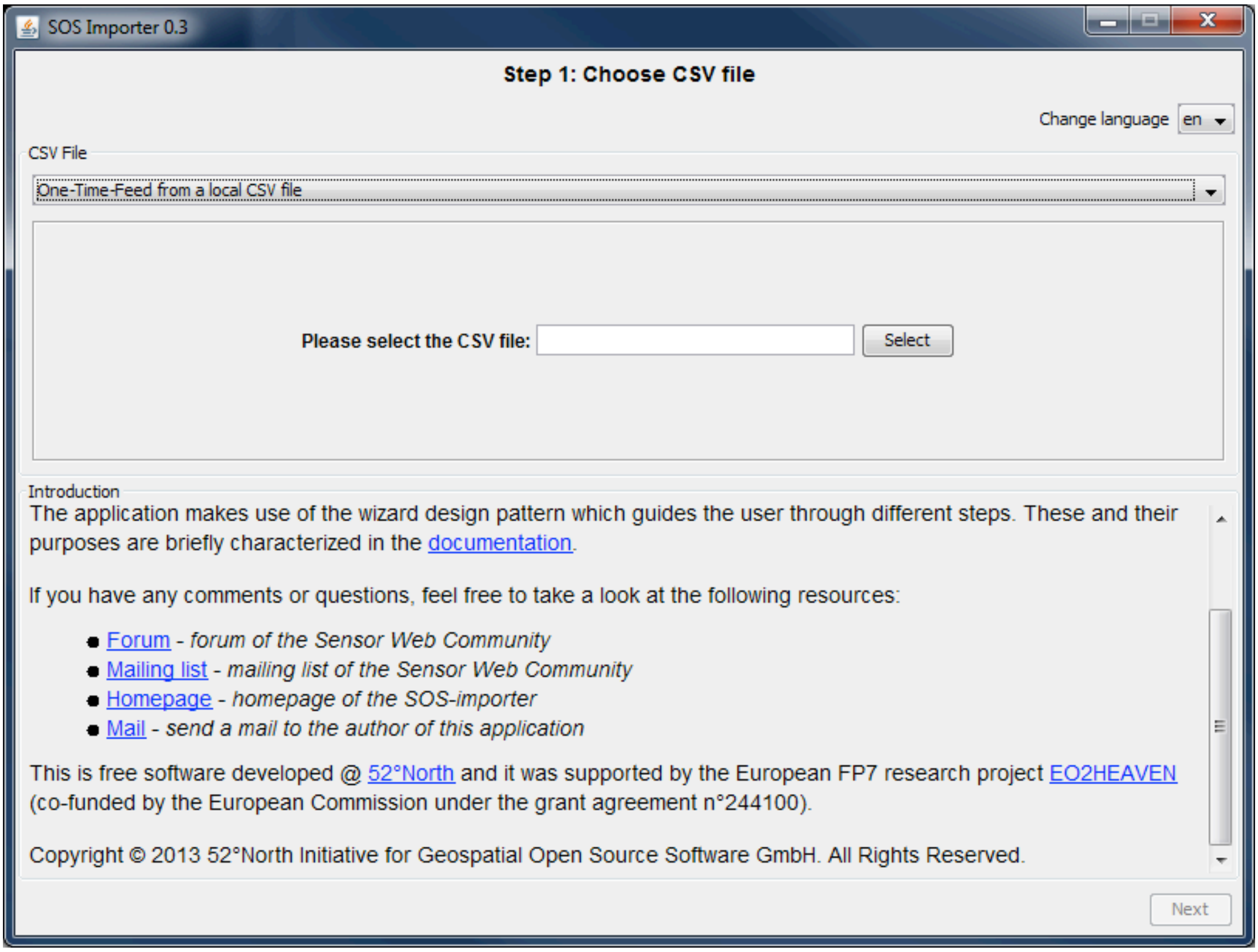

Figure 2-1: Choosing a CSV file on the local file system containing the observation data

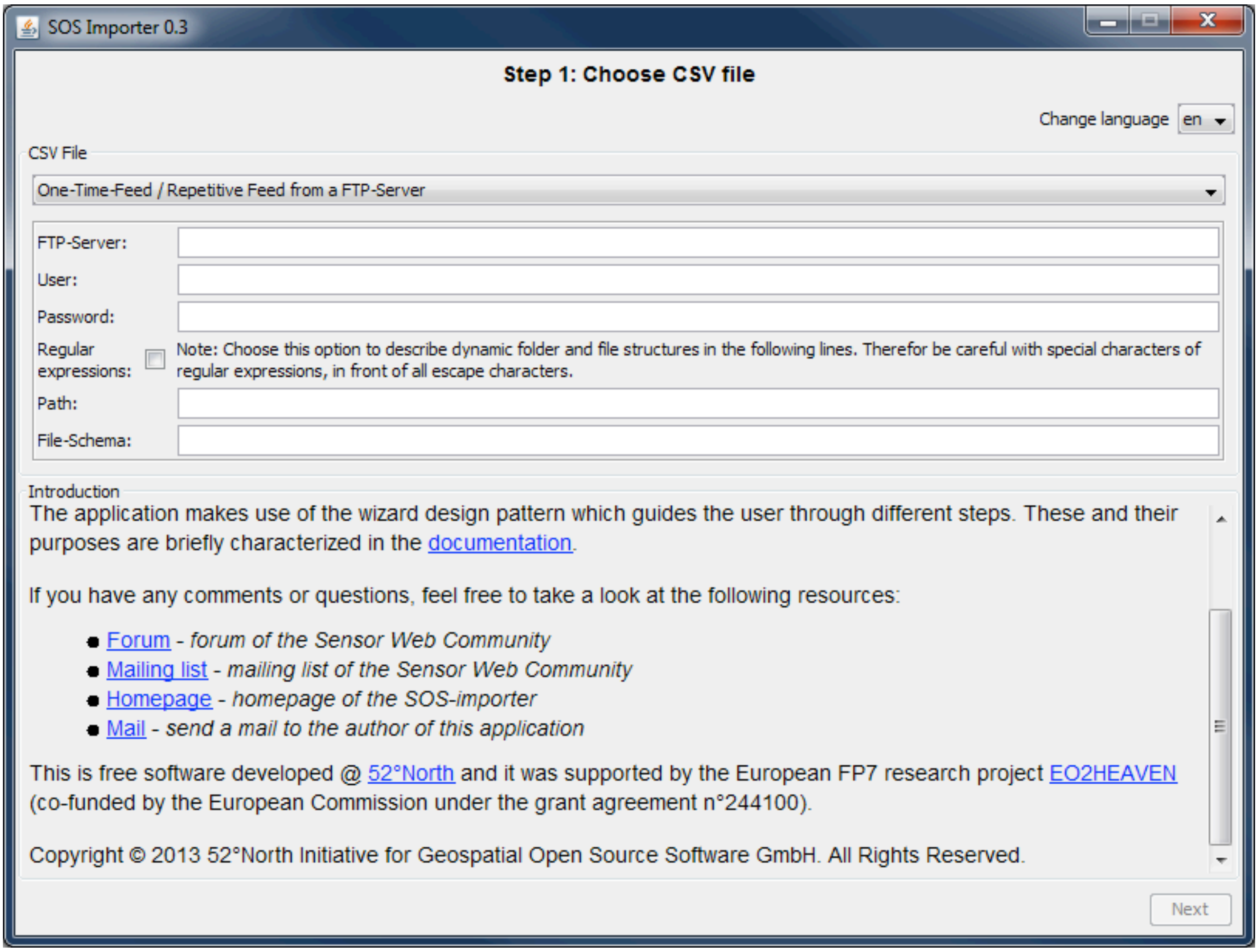

Figure 2-2: Choosing a CSV file on a FTP server

After this, the general settings for parsing the file need to be set (column separator, comment indicator, text qualifier and decimal separator). Furthermore the number of lines at the beginning of the file that shall be ignored can be specified (e.g. for ignoring headers in CSV files) (Figure 2-3).

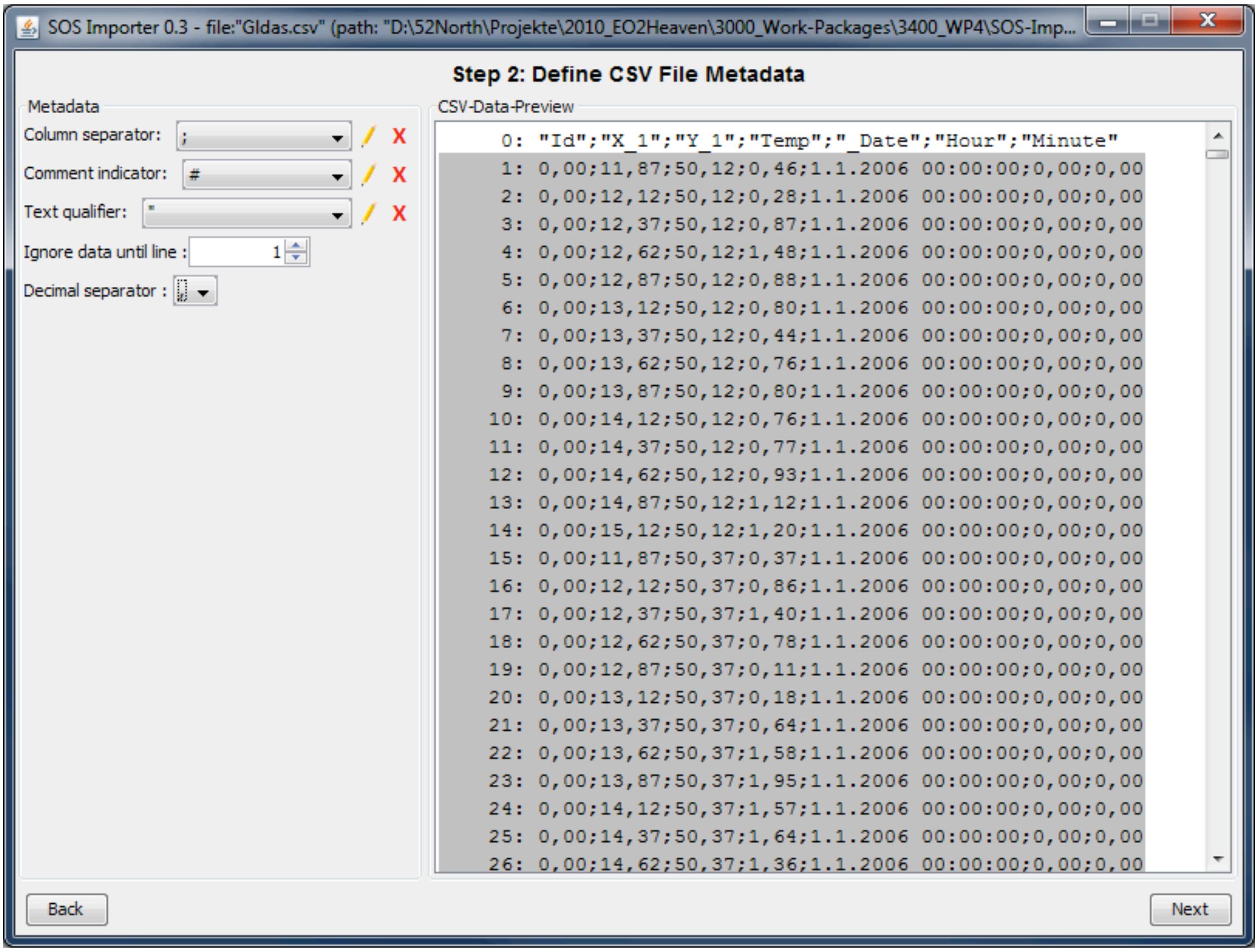

Figure 2-3: Selecting settings for parsing the CSV file

In the next steps the content of the different columns needs to be defined. Depending on the type of data contained in a column further information needs to be set by the user. For example:

- If a column contains data and time, the according format (e.g. dd/MM/yy HH:mm) needs to be described (Figure 2-4)
- If a column contains observed values the data type (e.g. numeric, Boolean, category) need to be set and if applicable also thousands and decimal separators (Figure 2-5)

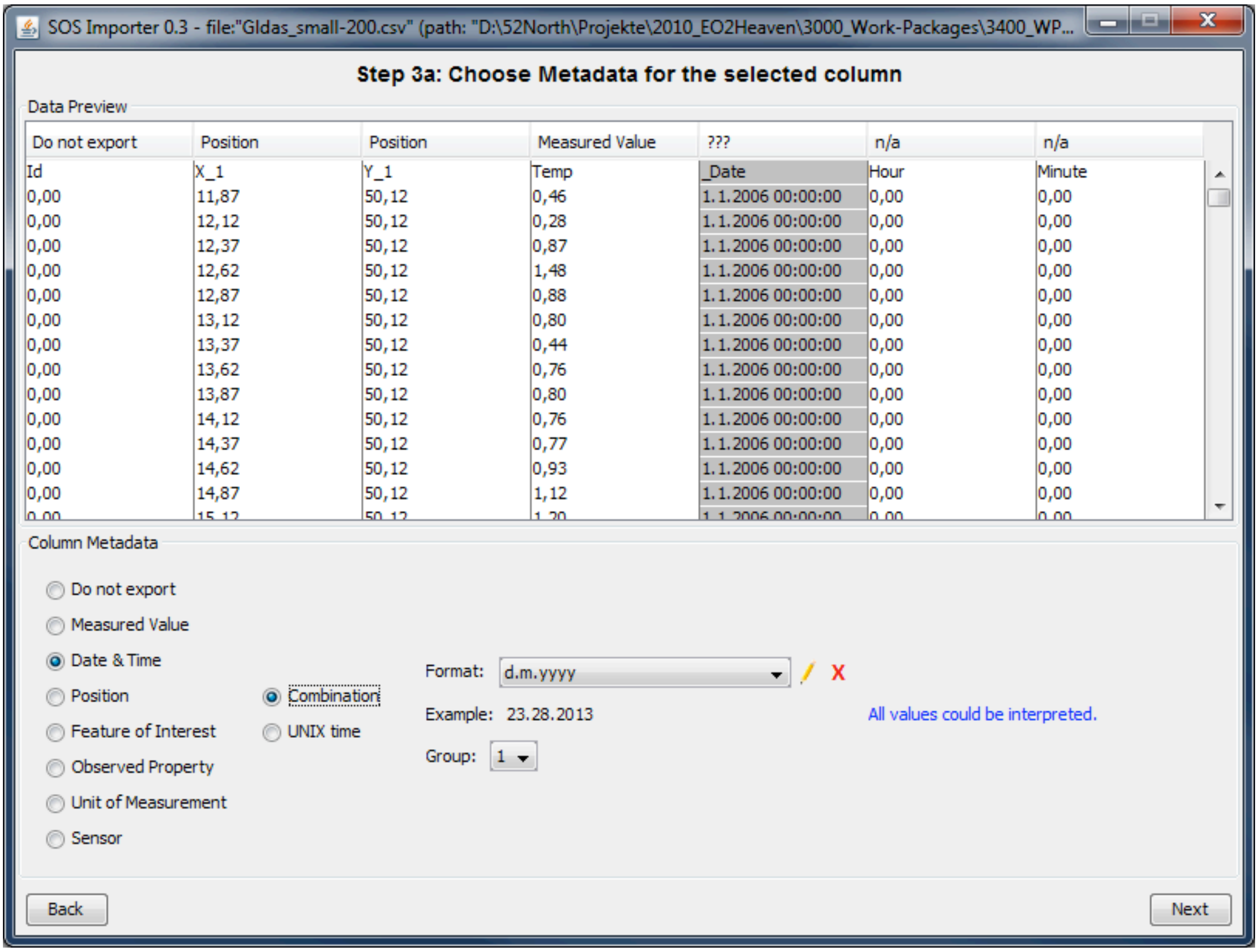

Figure 2-4: Describing the content of the columns (date and time)

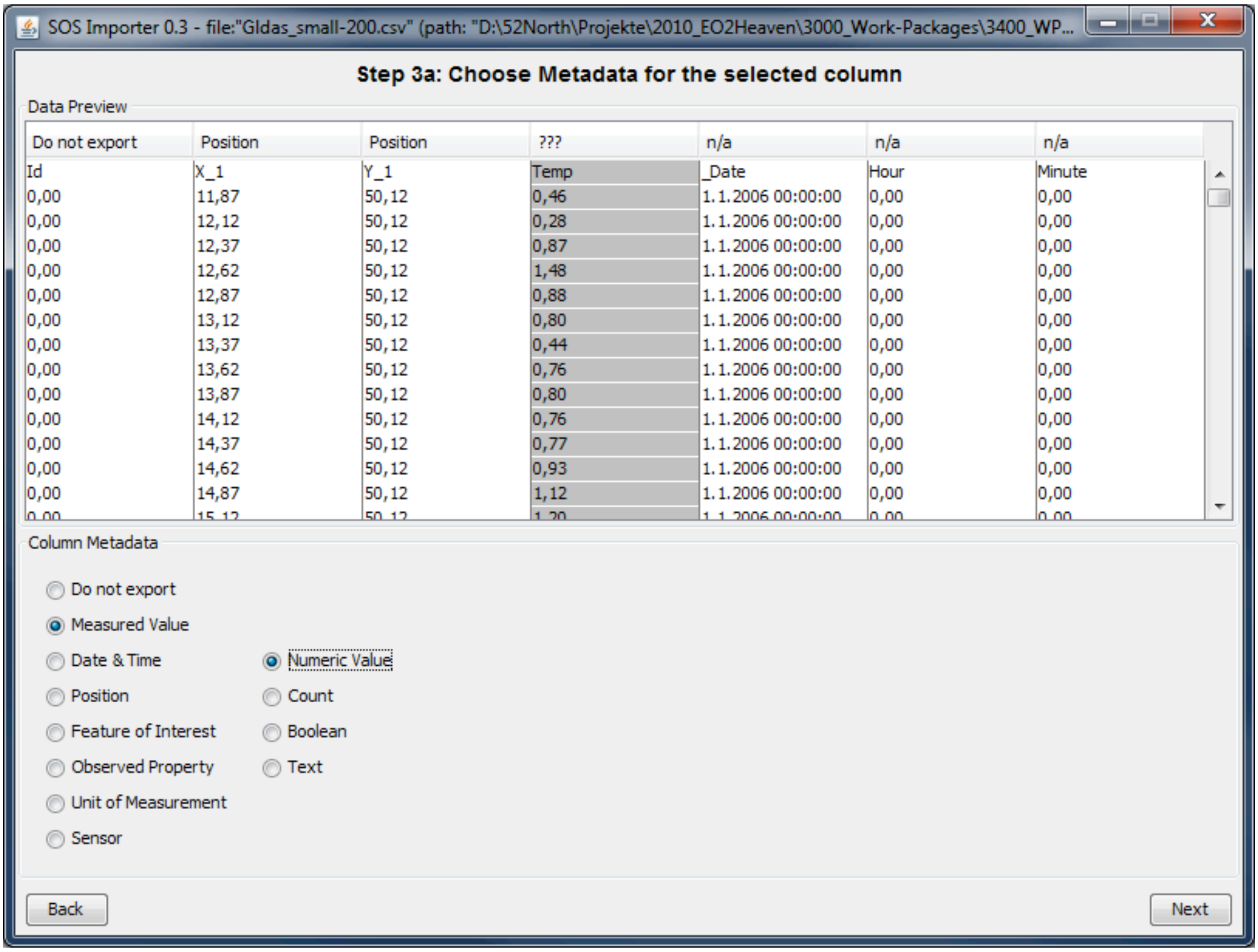

Figure 2-5: Describing the content of the columns (measured value)

If multiple columns with the same content type are present in a CSV file, according associations must be created, for example

- If there are multiple feature of interest columns it must be defined which feature of interest column belongs to which measured value column (Figure 2-6)
- If there are multiple timestamp columns it must be defined which time stamp column belongs to which measured value column
- If there are multiple unit of measurement columns it must be defined which unit of measurement column belongs to which measured value column

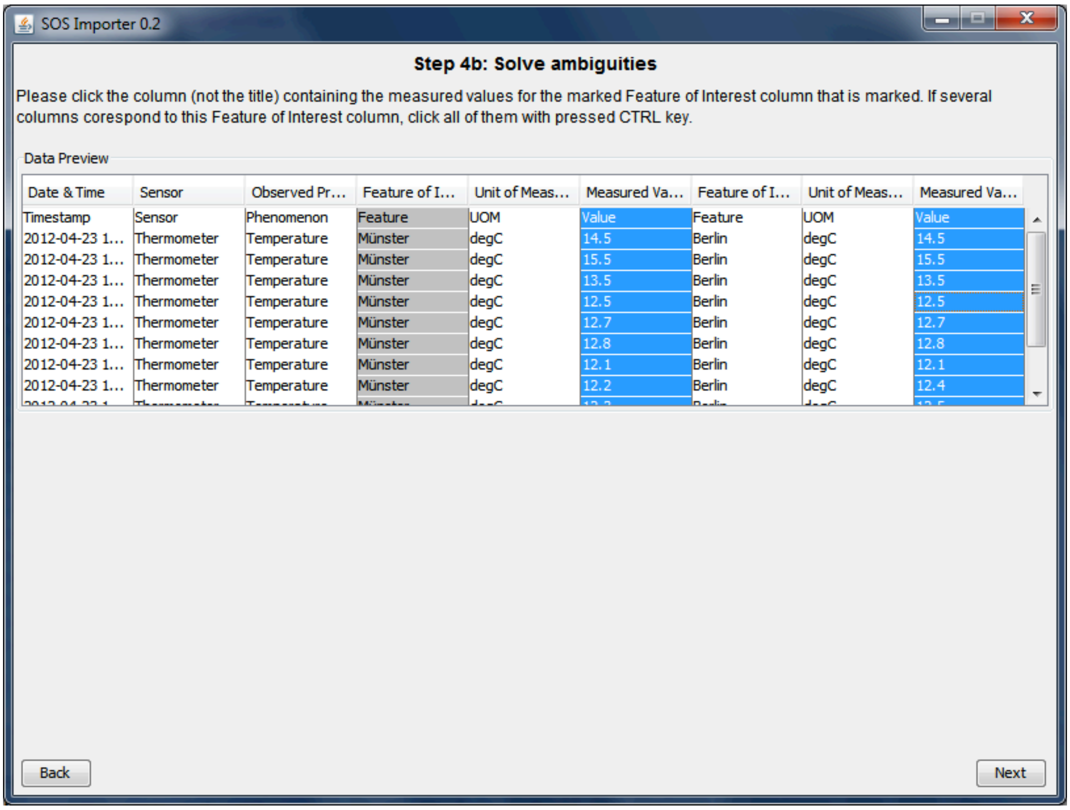

Figure 2-6: Creating associations between measured values and features of interest

If there is missing metadata for a time stamp column (e.g. time zone) the user has to provide this information in the next step (Figure 2-7).

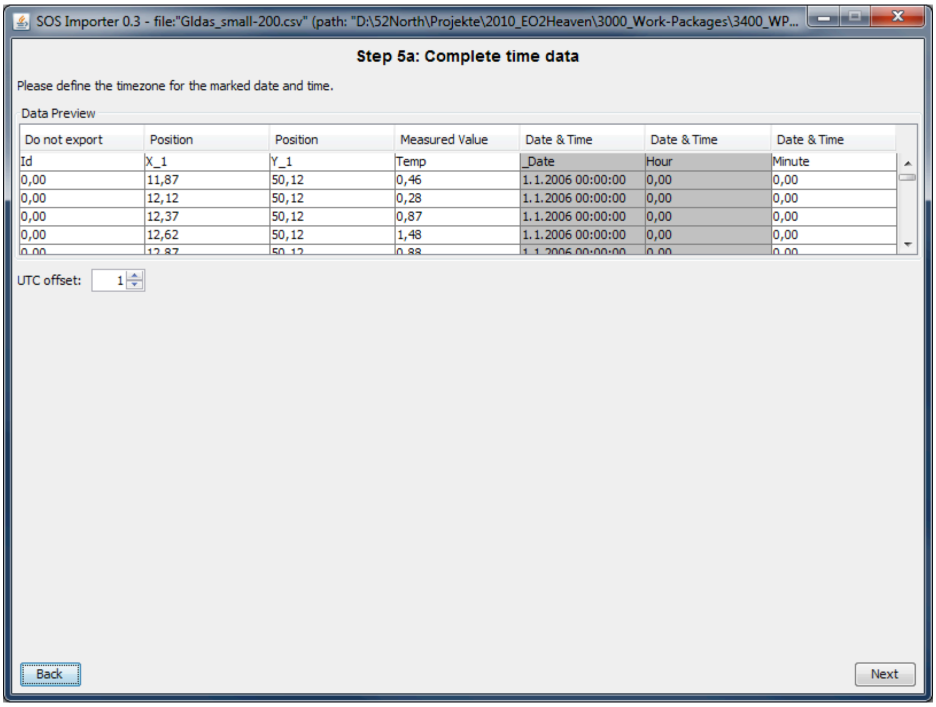

Figure 2-7: Adding missing metadata (missing time zone)

In case certain metadata elements are missing within the positions of sensor observations, the next step (Figure 2-8) can be used for inserting the according information (e.g. height, EPSG code of the reference system).

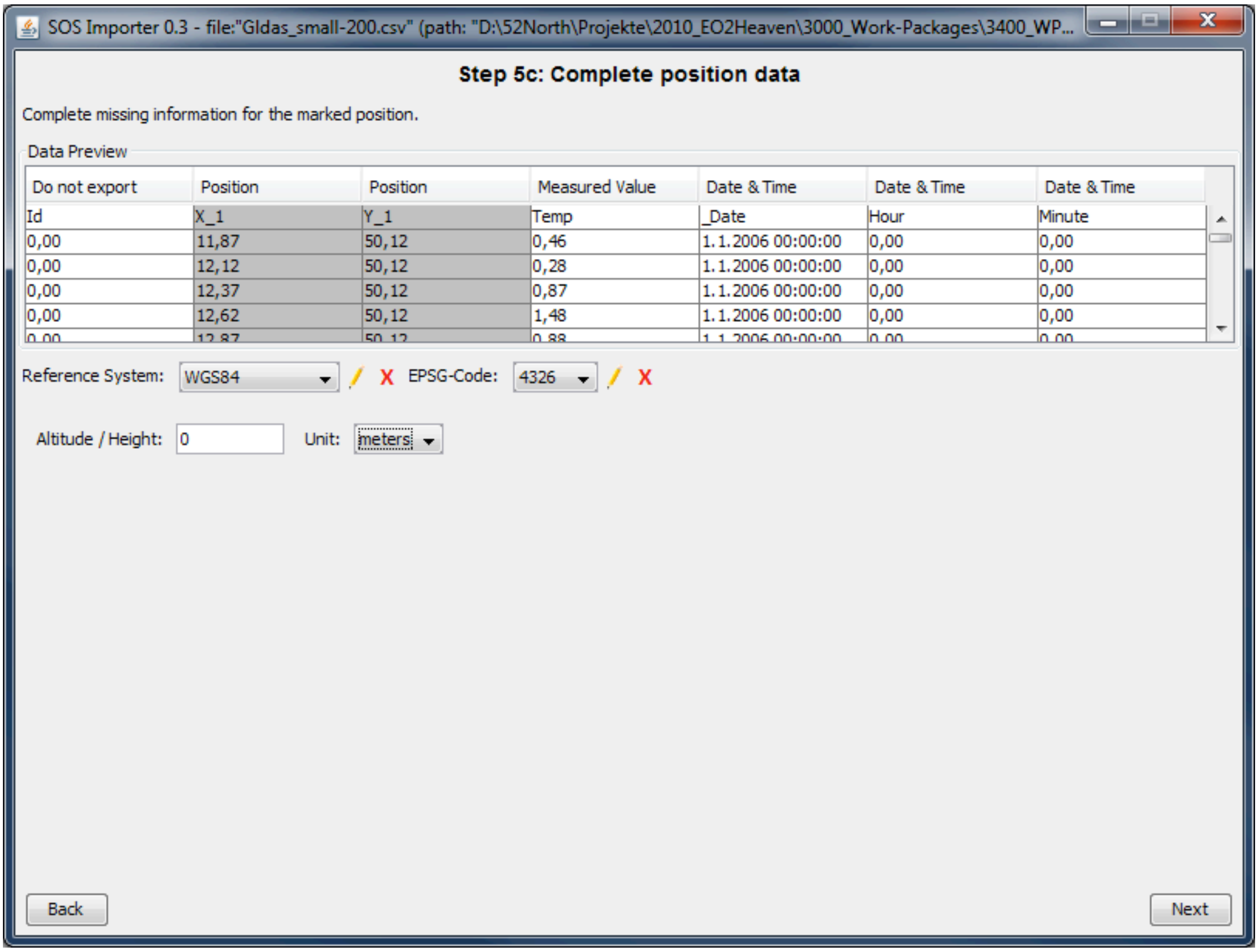

Figure 2-8: Adding missing position information (EPSG code and height)

In some cases, certain metadata elements (e.g. names of measurement stations, units of measurement, names of sensors, positions) may be missing completely. However, as they are required by the SWE standards, the import tool provides a means for manually adding or even automatically generating such metadata elements.

If no identifiers for the geographic features the measurements are associated to are provided, the import tool offers the automatic generation of feature of interest identifiers. In the example below the identifiers are generated from the coordinates of the measurements (Figure 2-9).

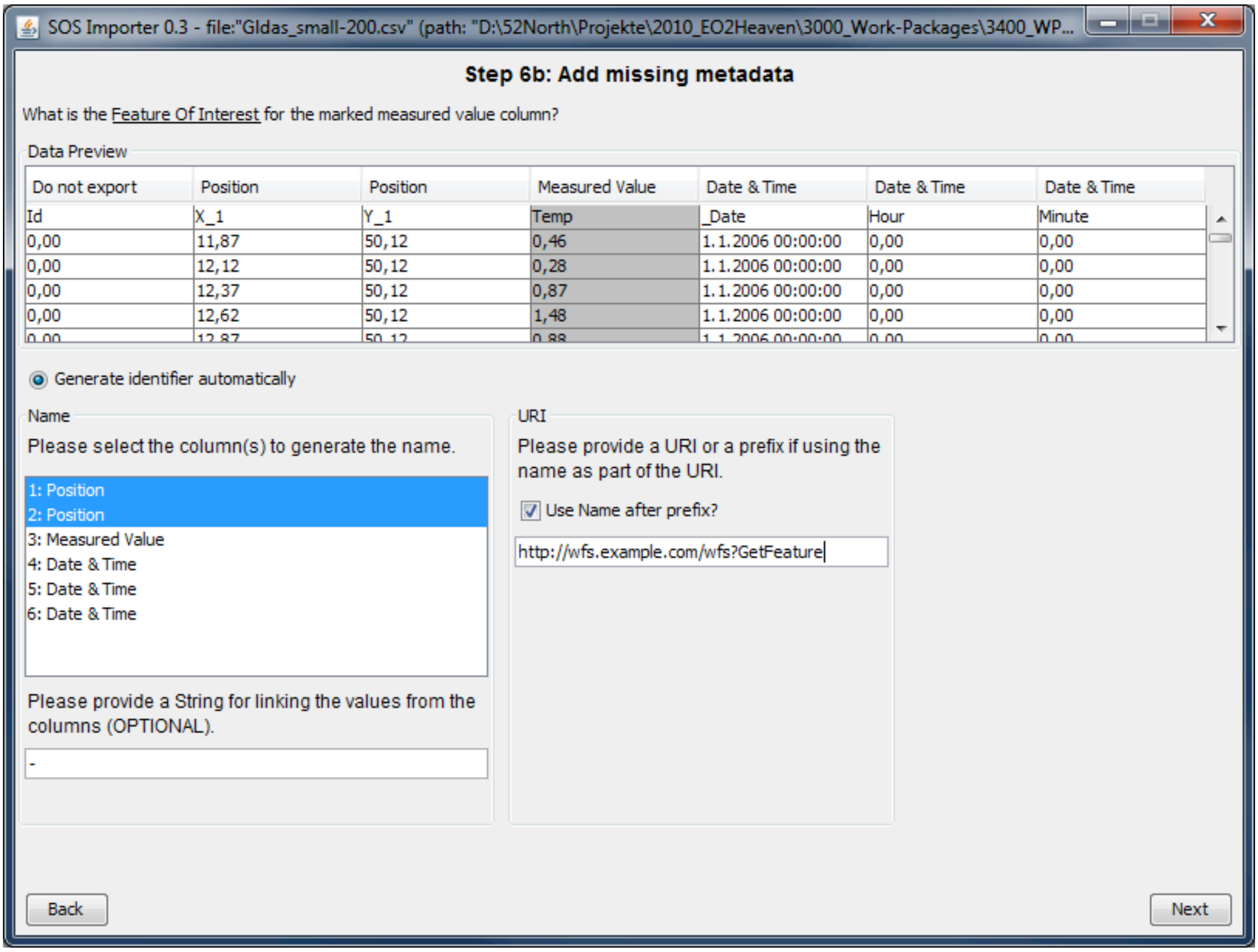

Figure 2-9: Automatically generating feature of interest identifiers

In case no unit of measurement is provided for the observed values, it can be manually added as a metadata element (Figure 2-10).

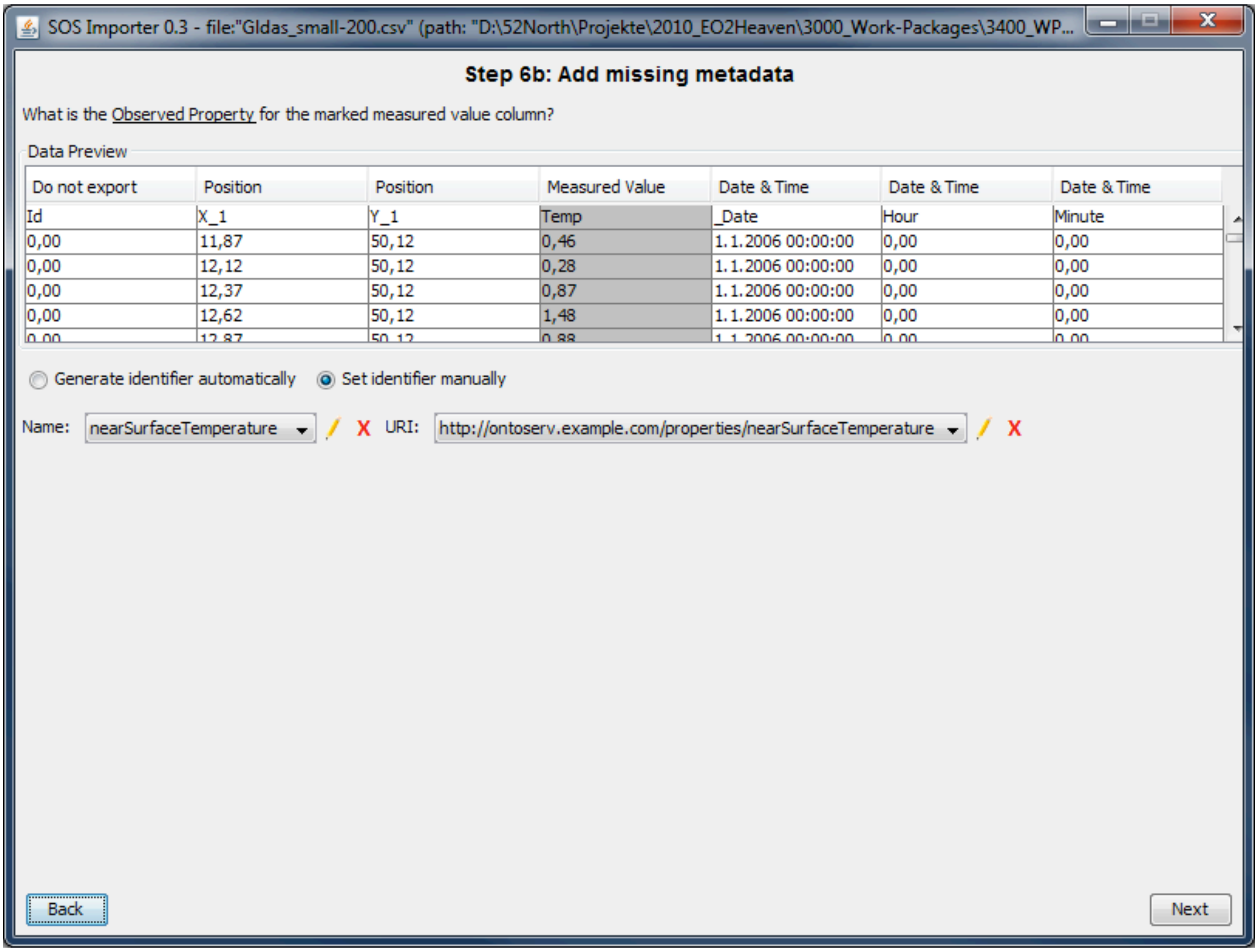

Figure 2-10: Manually adding information about the unit of measurement

If the CSV file does not contain the identifiers of the sensors which have generated the observation data, these identifiers can be added through the import tool (Figure 2-11). This can be achieved either automatically or manually.

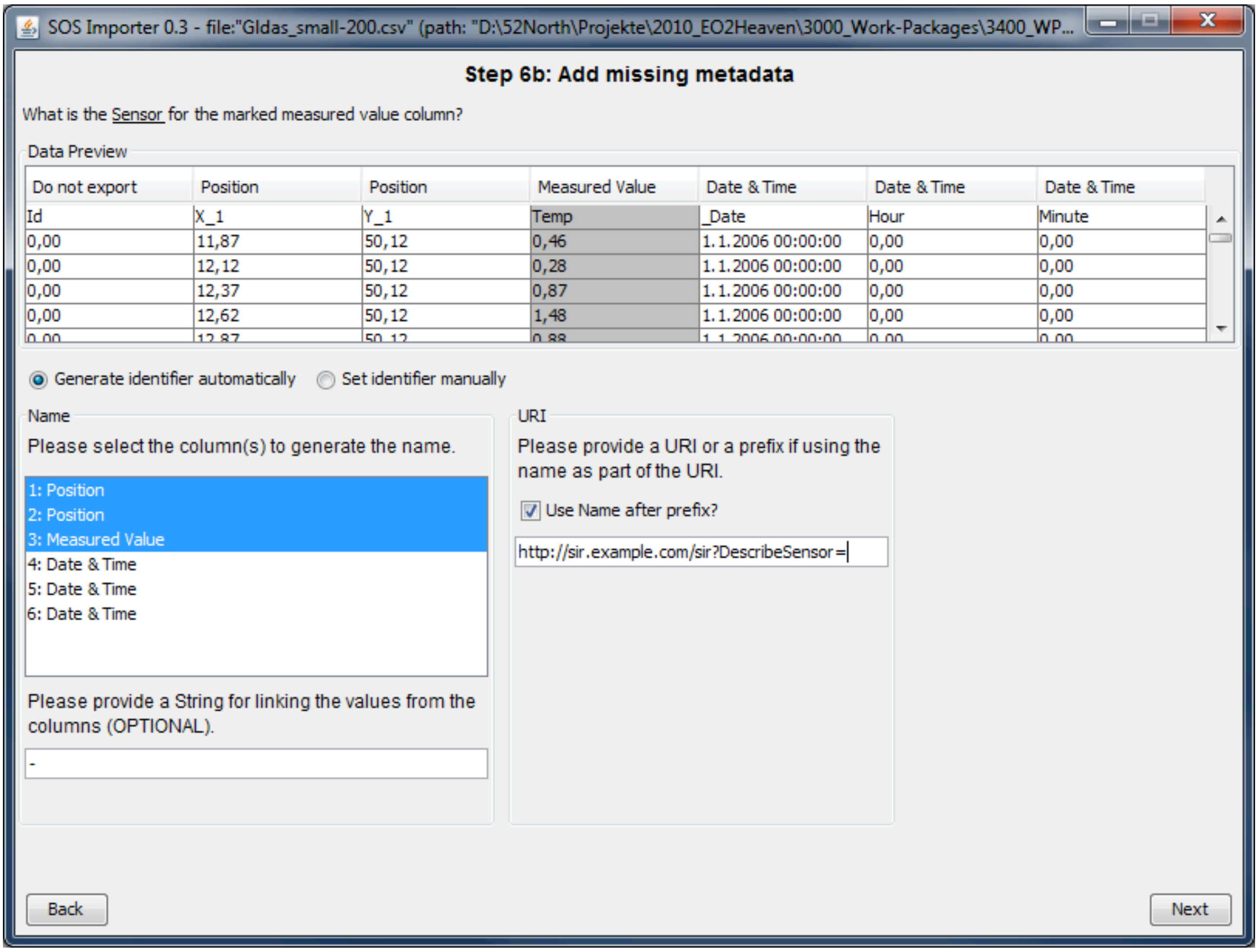

Figure 2-11: Creating sensor identifiers

In case the CSV file does not provide any spatial reference for the measured values, the menu below offers the possibility to enter positions manually. The user has the option to enter the coordinates through a map view (Figure 2-12) or through text fields (Figure 2-13).

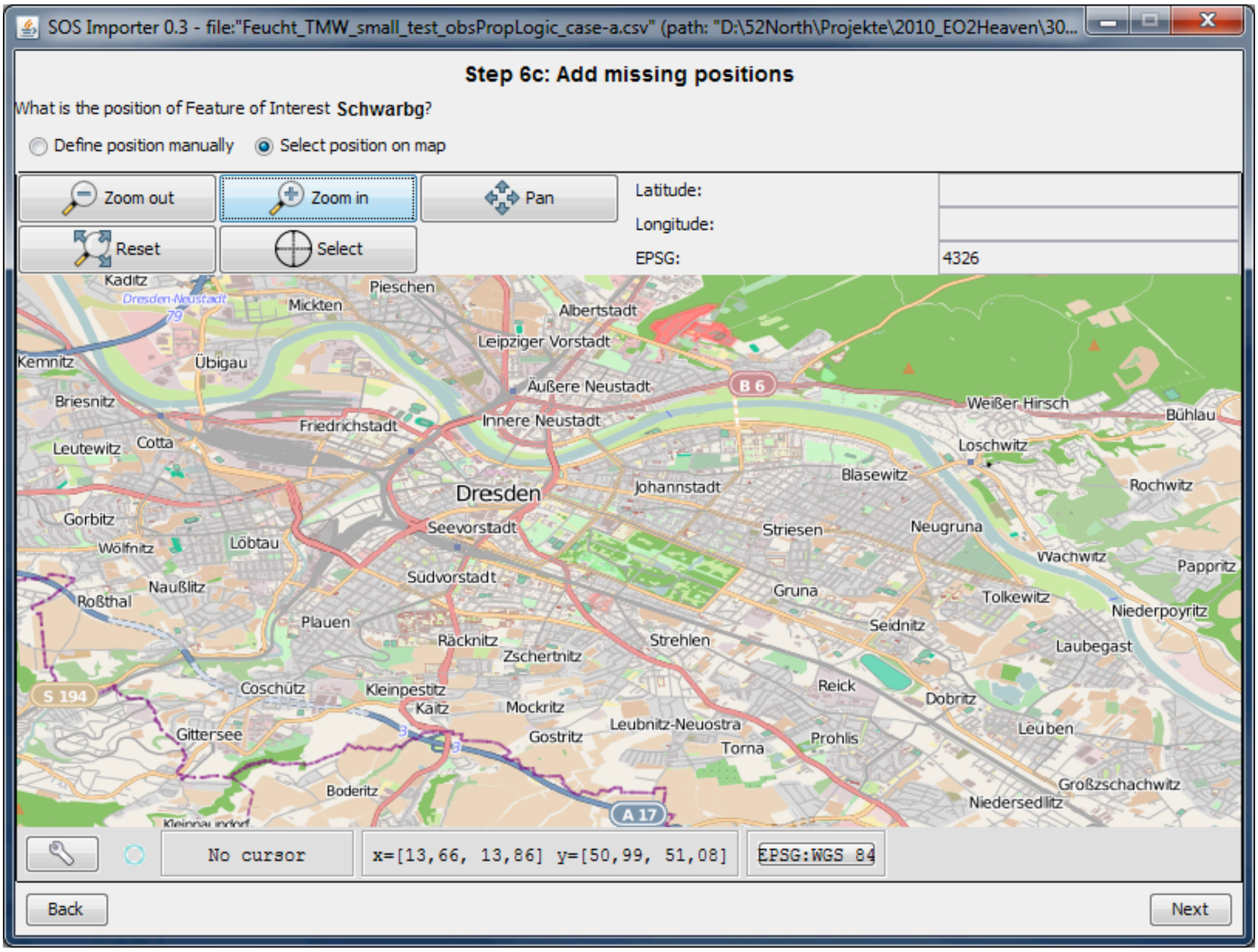

Figure 2-12: Setting the positions of the features of interest through a map view

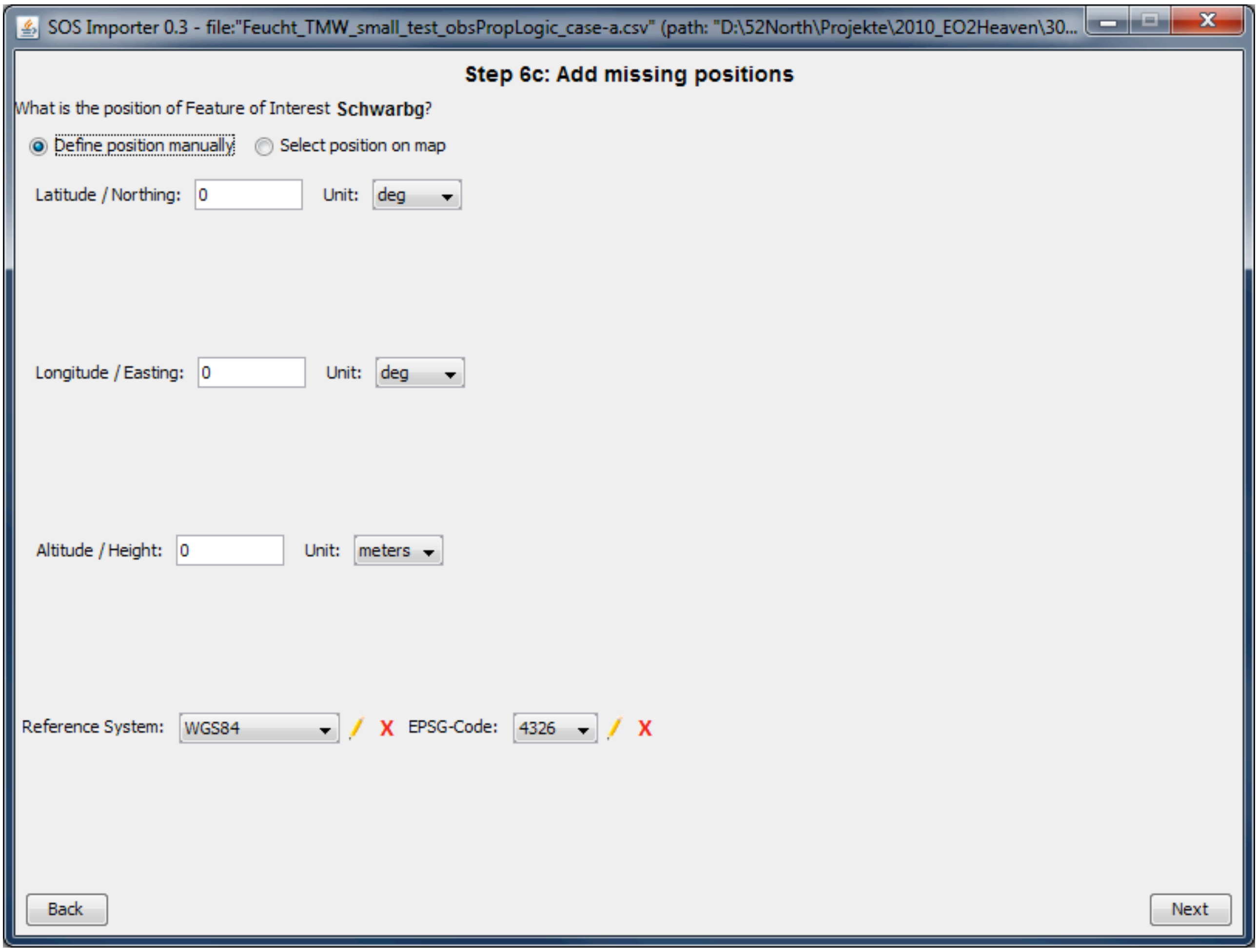

Figure 2-13: Setting the positions of the features of interest through text fields

After all necessary data and metadata elements have been entered, the SOS instance to which the sensor observations shall be published must be determined (Figure 2-14). In addition the user is able to store the configuration of the import in an external file. Later on this can be used for parameterising an automatically running import tool that regularly checks for new available data and subsequently pushes these data sets into a SOS server.

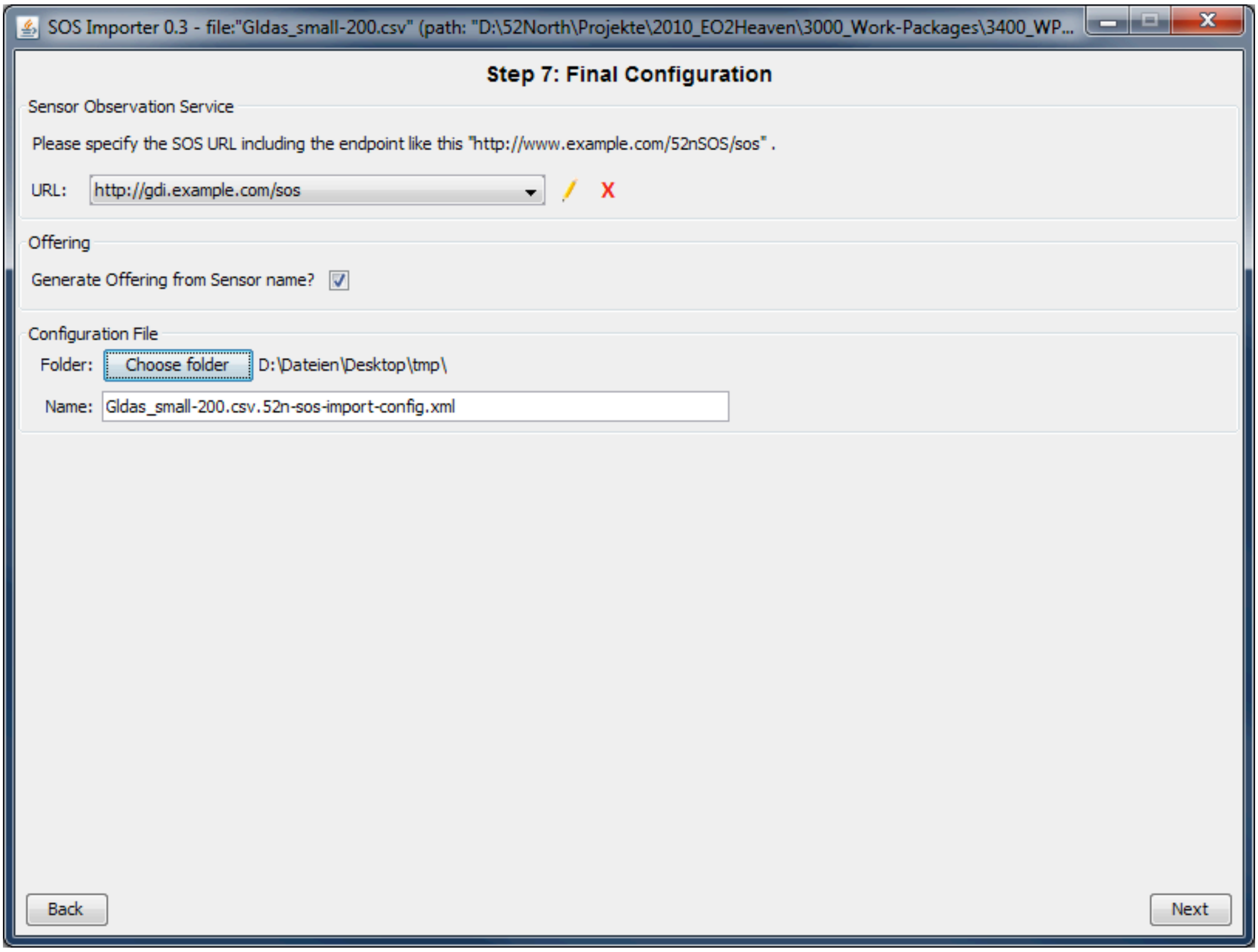

Figure 2-14: Setting the URL of a Sensor Observation Service

Finally, the user receives feedback, if the insertion of the measurement data into the SOS was successful (Figure 2-15).

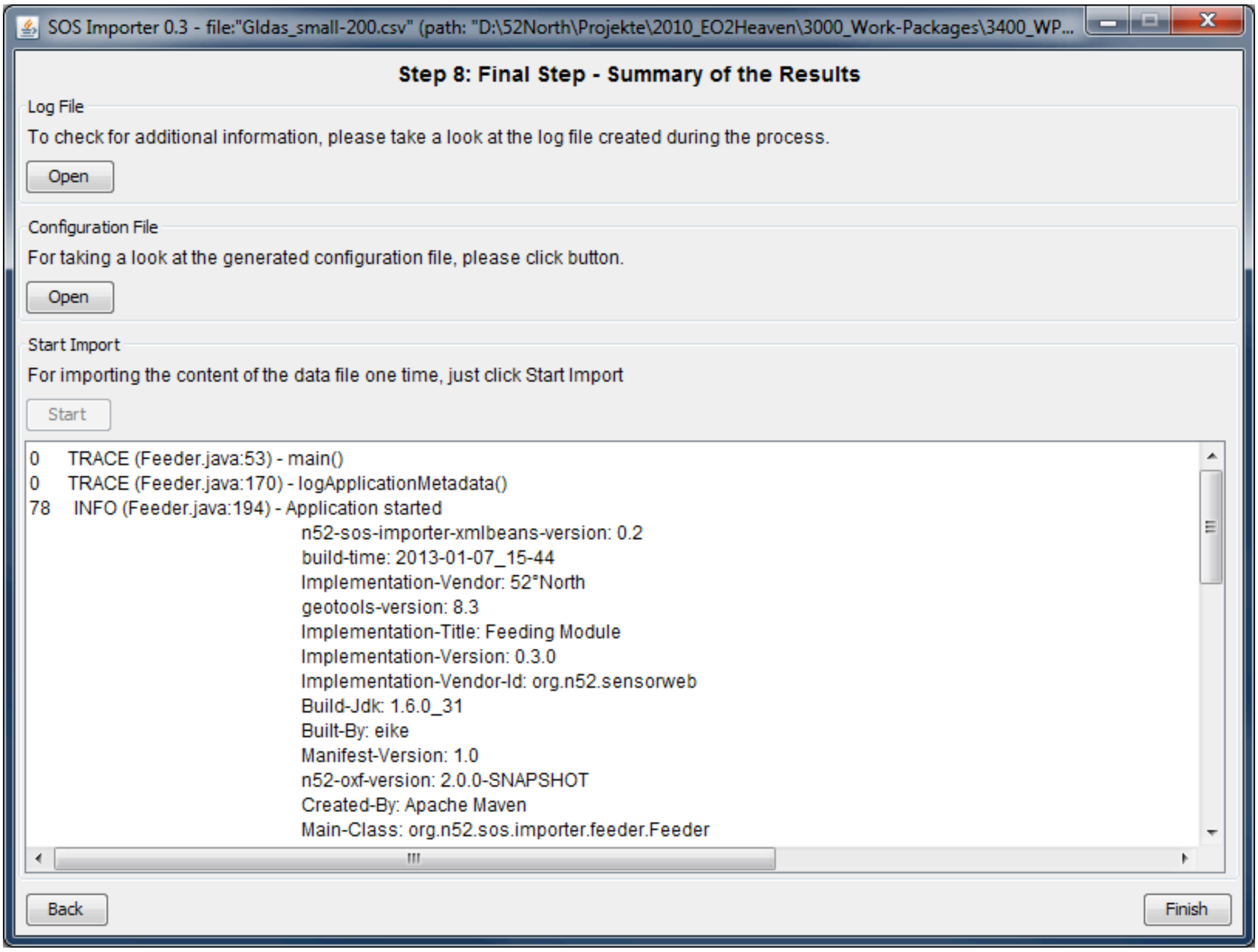

Figure 2-15: Import finished

2.3.4 Summary and Conclusion

The presented tool is a significant step forward to facilitate the practical application of the SOS. By offering an easy-to-use approach for publishing sensor data through a SOS, the acceptance of the SOS can be significantly increased. Furthermore it is a valuable tool for end users to integrate their sensor data into a SWE-compliant system.

The current version demonstrates that the general concept of an import tool for legacy data sets into the SOS is feasible. It may already be used for publishing existing data sets with limited background knowledge about the SWE standards.

However, further work will be necessary, especially covering the following aspects:

- Better hiding of underlying SWE concepts: the current version of the import tool still confronts the user with several SWE-specific terms (e.g. feature of interest); instead the next version should be further optimized for avoiding the use of SWE-specific terms and replace them by more natural language explanations
- More flexible support of data structures: the current version was developed on the basis of a set of sample files that were used within the EO2HEAVEN project. For the future a more flexible support of CSV files (i.e. different ways of how data is organized) is desirable. In addition, further data formats (e.g. Excel files) should be supported, as well.

3 References

EO2HEAVEN, 2011a. D4.7 Specification of the Spatial Information Infrastructure (SII) Implementation Architecture (second issue). Fraunhofer IOSB (Ed.), Deliverable D4.7 of the EO2HEAVEN project

http://www.eo2heaven.org/sites/default/files/D4.7%20Specification%20of%20the%20SII%20Implementation%20Architecture.pdf

EO2HEAVEN, 2011b. Specification of Advanced SWE Concepts (second issue). 52°North (Ed.), Deliverable D4.8 of the EO2HEAVEN project

http://www.eo2heaven.org/sites/default/files/D4.8_Specification_of_the_Advanced_SWE_Concepts_Issue_2.pdf

ISO/IEC 10746-1:1998: Information technology – Open Distributed Processing – Reference model.

ISO 19156:2011 - Geographic information – Observations and measurements

OGC, 2006. OpenGIS Web Map Server Implementation Specification. Version 1.3.0. OpenGIS Implementation Specification (Status: final), OGC 06-042.

OGC, 2007a. *Sensor Observation Service 1.0*. OGC 06-009r6

http://portal.opengeospatial.org/files/?artifact_id=26667

OGC, 2007b. *Sensor Model Language (SensorML) Implementation Specification*. OGC 07-000.

http://portal.opengeospatial.org/files/?artifact_id=21273

OGC, 2007c. Reference Model of the ORCHESTRA Architecture. Version 2.1. OGC Best Practices Document 07-097.

OGC, 2008a. *Web Coverage Service (WCS) Implementation Standard*. OGC 07-067r5.

http://portal.opengeospatial.org/files/?artifact_id=27297

OGC, 2008b. Robin, A. (2009). OGC SWE Common Data Model Encoding Standard. Version 2.0.0. OGC Encoding Standard 08-094r1.

OGC, 2008c. OGC Sensor Web Enablement (SWE) Architecture. OGC Best Practices Document 06-021r4

http://portal.opengeospatial.org/files/?artifact_id=29405

OGC, 2009a, *OGC Policy Standard - The Specification Model - A Standard for Modular specifications*. OGC 08-131r3.

https://portal.opengeospatial.org/files/?artifact_id=34762

OGC, 2009c. *OGC Sensor Planning Service Implementation Standard*. Version 2.0. OpenGIS Implementation Standard. OGC 09-000

OGC, 2009d. *Specification of the Sensor Service Architecture, Version 3.0 (Rev. 3.1).* OGC Discussion Paper 09-132r1. Deliverable D2.3.4 of the European Integrated Project SANY, FP6-IST-033564, 2009.

http://portal.opengeospatial.org/files/?artifact_id=35888

OGC, 2011a. *OGC Implementation Specification: Observations and Measurements (O&M) - XML Implementation 2.0*. OGC Document OGC 10-025r1.

http://portal.opengeospatial.org/files/?artifact_id=41510

OGC, 2011b, Lightweight SOS Profile for Stationary In-Situ Sensors. OGC Discussion Paper 11-169r1.

https://portal.opengeospatial.org/files/?artifact_id=52803&version=1

OGC, 2012a. *Sensor Observation Service 2.0*. OGC 12-006.

https://portal.opengeospatial.org/files/?artifact_id=47599

OGC, 2012b: OGC Sensor Model Language. Version 2.0 Implementation Specification (draft). OGC 12-000.

Shafranovich, Y., 2005. *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. The Internet Society.

http://tools.ietf.org/html/rfc4180

4 Abbreviations and acronyms

CSV	Comma Separated Values
EO	Earth Observation
EO2HEAVEN	Earth Observation and ENVironmental modelling for the mitigation of HEAlth risks
EPSG	European Petroleum Survey Group
FOI	Feature of Interest
FP6/7	6th/7th Framework Programme
GEOSS	Global Earth Observation System of Systems
GML	Geography Markup Language
INSPIRE	Infrastructure for Spatial Information in the European Community
IS	International Standard
ISO	International Standardization Organisation
IST	Information Society Technology
KVP	key-value-pair binding
OGC	Open Geospatial Consortium
ORCHESTRA	Open Architecture and Spatial Data Infrastructure for Risk Management
O&M	Observations and Measurement
REST	Representational State Transfer
RM-OA	Reference Model of the ORCHESTRA Architecture
RM-ODP	Reference Model for Open Distributed Processing
SANY	Sensors Anywhere (FP6 project)
SDI	Spatial Data Infrastructure
SensorML	Sensor Model Language
SensorSA	Sensor Service Architecture
SII	Spatial Information Infrastructure
SOS	Sensor Observation Service
SPS	Sensor Planning Service
SQL	Structured Query Language
SWE	Sensor Web Enablement
TML	Transducer Markup Language
UncertWeb	Uncertainty Enabled Model Web
URL	Uniform Resource Locator
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
XML	Extensible Markup Language

5 Annex A: SOS Lightweight Profile

The OGC Public Discussion Paper 11-169r1 (OGC, 2011b) “OGC Lightweight SOS Profile for Stationary In-Situ Sensors” shall be considered an integral part of this best-practices document.