

SUSHI

Information on usage events must be transferred to a central database. The SURE Statistics project has formulated as a requirement that the transfer of data must take place in a reliable manner. In the infrastructure to be built in this project, the log aggregator will bear the primary responsibility for obtaining the statistical data from individual repositories. Once every 24 hour, it will send a request to each repository for the usage events that have occurred on that particular day. Secondly, the log aggregator is also responsible for ensuring that automated requests by internet robots are filtered as much as possible.

Communication between log aggregator and repository must take place as is detailed in figure 1.

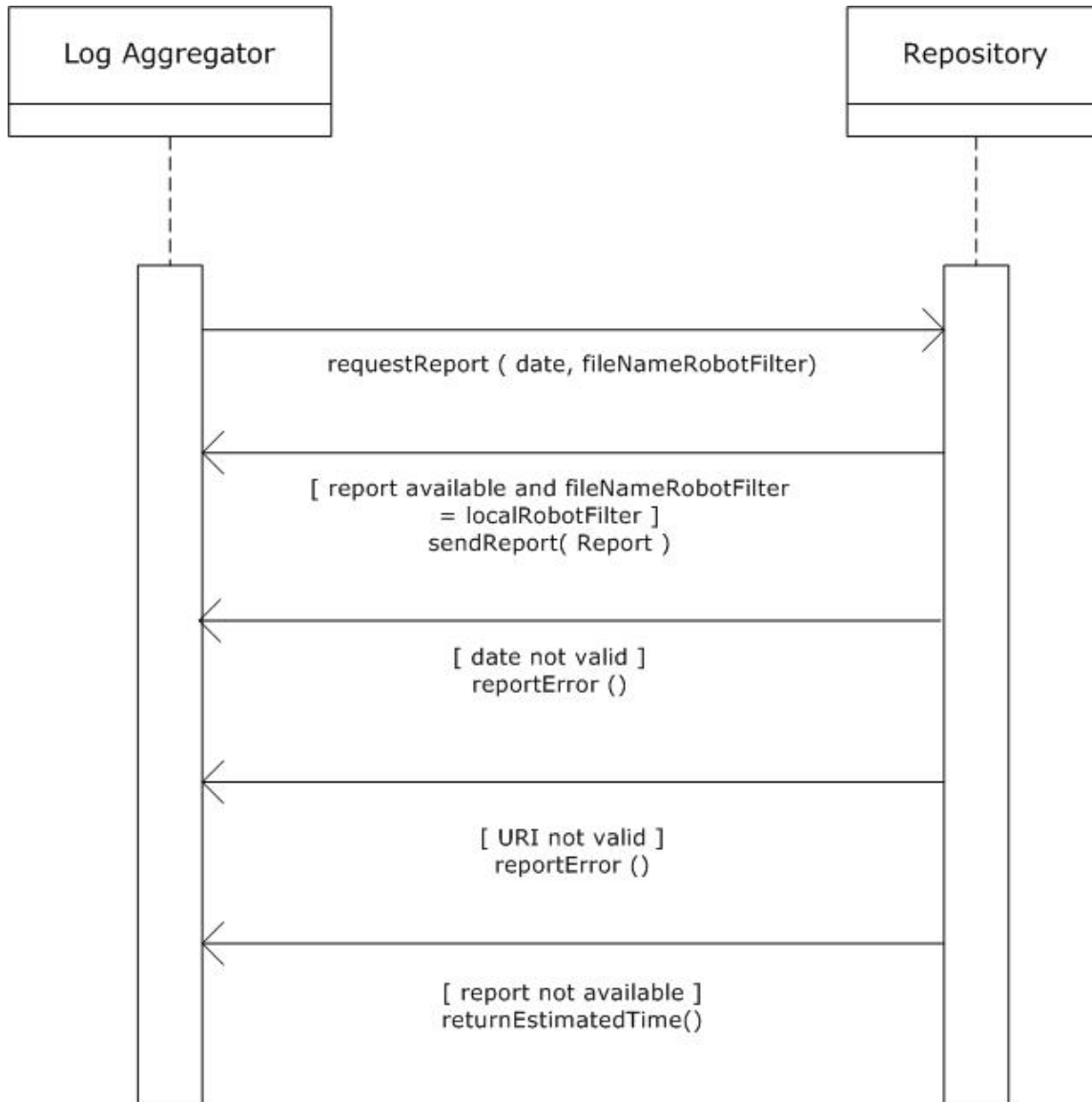


Figure 1.

The interaction commences when the log aggregator sends a request for a report about the daily usage of a certain repository. Two parameters must be sent as part of this request: (1) the date of the report and (2) the file name of the most recent robot filter. The filename that is mentioned in this request will be compared to the local filename. Four possible responses can be returned by the repository.

- If the filename that is mentioned in the request exactly matches the filename that is maintained locally, and if a report for the requested data is indeed available, this report will be returned immediately.
- In this protocol, only daily reports will be allowed. This was decided mainly to restrict the size of the data traffic between the servers. If a request is sent for a period that exceeds one day, an error message will be sent indicating that the date parameter is incorrect.
- If the URI of the robot filter file, for some reason, cannot be resolved to an actual file, an error message will be sent about this.
- If the parameters are correct, but if the report is not yet available, a message will be sent which provides an estimation of the time of arrival.

This protocol involves a conversation in two directions. To implement the transfer of data, the protocol that was developed by the *Standardised Usage Statistics Harvesting Initiative* (SUSHI) shall be used. SUSHI^[1]http://www.surffoundation.nl/#_ftn1 was developed by NISO (National Information Standards Organization) in cooperation with COUNTER. SUSHI enables parties to harvest usage statistics. The protocol works with only two types of messages: requests and responses. The protocol was originally developed for the exchange of COUNTER reports, but fortunately, other types of reports can also be used as the payload of the transfer. The standard does stipulate, however, that the requirements for report naming are adhered to. SUSHI is based on SOAP. The services that can be offered by the Web Service are described in a WSDL document.^[2]http://www.surffoundation.nl/#_ftn2

In a number of other projects, OAI-PMH is used to synchronise the central database with local databases. In SURE, SUSHI was favoured over OAI-PMH because the latter technique only allows for a one-way-traffic. It was found that SUSHI allowed for a more reliable transfer of data.

In SUSHI version 1.0., the following information must be sent along with the request:

- § Requestor ID
- § Name of requestor
- § E-mail of requestor
- § CustomerReference ID (may be identical to the Requestor ID)
- § Name of the report that is requested
- § Version number of the report
- § Start and end date of the report

This request will activate a special tool that can inspect the server logging and that can return the requested data. These data are transferred as OpenURL Context Object log entries, as part of a SUSHI response.

The response must repeat all the information from the request, and provide the requested report as XML payload

The usage data are subsequently stored in a central database. External parties can obtain information about the contents of this central database through specially developed web services. The log harvester must ultimately expose these data in the form of COUNTER-compliant reports.

Listing 2 is an example of a SUSHI request, sent from the log aggregator to a repository.

```

<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://schemas.xmlsoap.org/soap/envelope/ http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ReportRequest
      xmlns:ctr="http://www.niso.org/schemas/sushi/counter"
      xsi:schemaLocation="http://www.niso.org/schemas/sushi/counter
http://www.niso.org/schemas/sushi/counter_sushi3_0.xsd"
      xmlns="http://www.niso.org/schemas/sushi"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
      <Requestor>
        <ID>www.logagggregator.nl</ID>
        <Name>Log Aggregator</Name>
        <Email>logagggregator@surf.nl</Email>
      </Requestor>
      <CustomerReference>
        <ID>www.leiden.edu</ID>
        <Name>LeidenUniversity</Name>
      </CustomerReference>
      <ReportDefinition Release="urn:robots-v1.xml" Name="Daily Report v1">
        <Filters>
          <UsageDateRange>
            <Begin>2009-12-21</Begin>
            <End>2009-12-22</End>
          </UsageDateRange>
        </Filters>
      </ReportDefinition>
    </ReportRequest>
  </soap:Body>
</soap:Envelope>

```

Listing 2.

Note that the intent of the SUSHI request above is to see all the usage events that have occurred on 21 December 2009. The SUSHI schema was originally developed for the exchange of COUNTER-compliant reports. In the documentation of the SUSHI XML schema, it is explained that COUNTER usage is only reported at the month level. In SURE, only daily reports can be provided. Therefore, it will be assumed that the implied time on the date that is mentioned is 0:00. The request in the example that is given thus involves all the usage events that have occurred in between 2009-12-21T00:00:00 and 2009-12-22T00:00:00.

As explained previously, the repository can respond in four different ways. If the parameters of the request are valid, and if the requested report is available, the OpenURL ContextObjects will be sent immediately. The Open URL Context Objects will be wrapped into element <Report>, as can be seen in listing 3.

```

<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://schemas.xmlsoap.org/soap/envelope/ http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ReportResponse
      xmlns:ctr="http://www.niso.org/schemas/sushi/counter"
      xsi:schemaLocation="http://www.niso.org/schemas/sushi/counter
        http://www.niso.org/schemas/sushi/counter_sushi3_0.xsd"
      xmlns="http://www.niso.org/schemas/sushi"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
      <Requestor>
        <ID>www.logagggregator.nl</ID>
        <Name>Log Aggregator</Name>
        <Email>logagggregator@surf.nl</Email>
      </Requestor>
      <CustomerReference>
        <ID>www.leiden.edu</ID>
        <Name>LeidenUniversity</Name>
      </CustomerReference>
      <ReportDefinition Release="urn:DRv1" Name="Daily Report v1">
        <Filters>
          <UsageDateRange>
            <Begin>2009-12-22</Begin>
            <End>2009-12-23</End>
          </UsageDateRange>
        </Filters>
      </ReportDefinition>
      <Report>
        <ctx:context-objects
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xmlns:dcterms="http://dublincore.org/documents/2008/01/14/dcmi-terms"
          xmlns:ctx="info:ofi/fmt:xml:xd:ctx">
          <ctx:context-object timestamp="2009-11-
            09T05:56:18+01:00">
            ...
          </ctx:context-object>
        </ctx:context-objects>
      </Report>
    </ReportResponse>
  </soap:Body>
</soap:Envelope>

```

Listing 3.

If the begin date and the end date in the request of the log aggregator form a period that exceeds one day, an error message must be sent. In the SUSHI schema, such messages may be sent in an <Exception> element. Three types of errors can be distinguished. Each error type is given its own number. An human-readable error message is provided under <Message>.

```

<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://schemas.xmlsoap.org/soap/envelope/ http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ReportResponse xmlns:ctr="http://www.niso.org/schemas/sushi/counter"
      xsi:schemaLocation="http://www.niso.org/schemas/sushi/counter http://www.niso.org/schemas/sushi/counter_sushi3_0.xsd"
      xmlns="http://www.niso.org/schemas/sushi"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
      <Requestor>
        <ID>www.logaggregator.nl</ID>
        <Name>Log Aggregator</Name>
        <Email>logaggregator@surf.nl</Email>
      </Requestor>
      <CustomerReference>
        <ID>www.leiden.edu</ID>
        <Name>LeidenUniversity</Name>
      </CustomerReference>
      <ReportDefinition Release="urn:DRv1" Name="Daily Report v1">
        <Filters>
          <UsageDateRange>
            <Begin>2009-12-22</Begin>
            <End>2009-12-23</End>
          </UsageDateRange>
        </Filters>
      </ReportDefinition>
      <Exception>
        <Number>1</Number>
        <Message>The range of dates that was provided is not
          valid. Only daily reports are
          available.</Message>
      </Exception>
    </ReportResponse>
  </soap:Body>
</soap:Envelope>

```

Listing 4.

A second type of error may be caused by the fact that the file that is mentioned in the request can not be accessed. In this situation, the response will look as follows:

```

<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://schemas.xmlsoap.org/soap/envelope/
    http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ReportResponse
      xmlns:ctr="http://www.niso.org/schemas/sushi/counter"
      xsi:schemaLocation="http://www.niso.org/schemas/sushi/counter
        http://www.niso.org/schemas/sushi/counter_sushi3_0.xsd"
      xmlns="http://www.niso.org/schemas/sushi"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
      <Requestor>
        <ID>www.logaggregator.nl</ID>
        <Name>Log Aggregator</Name>
        <Email>logaggregator@surf.nl</Email>
      </Requestor>
      <CustomerReference>
        <ID>www.leiden.edu</ID>
        <Name>LeidenUniversity</Name>
      </CustomerReference>
      <ReportDefinition Release="urn:DRv1" Name="Daily Report v1">
        <Filters>
          <UsageDateRange>
            <Begin>2009-12-22</Begin>
            <End>2009-12-23</End>
          </UsageDateRange>
        </Filters>
      </ReportDefinition>
      <Exception>
        <Number>2</Number>
        <Message>The file describing the internet robots is not
          accessible.</Message>
      </Exception>
    </ReportResponse>
  </soap:Body>
</soap:Envelope>

```

Listing 5.

When the repository is in the course of producing the requested report, a response will be sent that is very similar to listing 6. The estimated time of completion will be provided in the <Data> element. According to the documentation of the SUSHI XML schema, this element may be used for any other optional data.

```

<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://schemas.xmlsoap.org/soap/envelope/
    http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ReportResponse
      xmlns:ctr="http://www.niso.org/schemas/sushi/counter"
      xsi:schemaLocation="http://www.niso.org/schemas/sushi/counter
        http://www.niso.org/schemas/sushi/counter_sushi3_0.xsd"
      xmlns="http://www.niso.org/schemas/sushi"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
      <Requestor>
        <ID>www.logaggregator.nl</ID>
        <Name>Log Aggregator</Name>
        <Email>logaggregator@surf.nl</Email>
      </Requestor>
      <CustomerReference>
        <ID>www.leiden.edu</ID>
        <Name>LeidenUniversity</Name>
      </CustomerReference>
      <ReportDefinition Release="urn:DRv1" Name="Daily Report v1">
        <Filters>
          <UsageDateRange>
            <Begin>2009-12-22</Begin>
            <End>2009-12-23</End>
          </UsageDateRange>
        </Filters>
      </ReportDefinition>
      <Exception>
        <Number>3</Number>
        <Message>The report is not yet available. The estimated
          time of completion is
          provided under "Data".</Message>
        <Data>2010-01-08T12:13:00+01:00</Data>
      </Exception>
    </ReportResponse>
  </soap:Body>
</soap:Envelope>

```

Listing 6.

Error numbers and the corresponding Error messages are also provided in the table below.

Error number	Error message
1	The range of dates that was provided is not valid. Only daily reports are available.
2	The file describing the internet robots is not accessible
3	The report is not yet available. The estimated time of completion is provided under "Data"

----[1]http://www.surffoundation.nl/#_ftnref1] <http://www.niso.org/schemas/sushi/>
 [2]http://www.surffoundation.nl/#_ftnref2] It can be accessed at http://www.niso.org/schemas/sushi/counter_sushi2_5.wsdl.

Application profiles

SURFshare use of Usage Statistics Exchange
 DRIVER Annex Use of Usage Statistics Exchange