



**Specification:
Data Format and Exchange
for OA Statistics
Version 0.5**

Specification: Data Format and Exchange for OA Statistics

Hans-Werner Hilse <hilse@sub.uni-goettingen.de>

Daniel Metje <metje@sub.uni-goettingen.de>

Goettingen State and University Library

Date	Version	Comments
29.07.2009	0.1	First Draft
10.08.2009	0.2	Policy edited and connections to other projects outlined
27.08.2009	0.3	Compatibility aspects renewed
30.09.2009	0.4	Translation into English language
05.11.2009	0.5	Addition of requester classification, minor XML schema clean-up

Table of contents

1 Preface.....	4
1.1 Terminology.....	4
1.2 Data Format.....	4
1.3 Data Exchange.....	5
1.4 Related Projects and Initiatives.....	8
1.5 Compatibility Aspects.....	9
2 Data Provider: Infrastructure.....	9
3 Data Format: Example with Comments.....	10
4 Data Format: External Policies.....	13
5 General Comments.....	13
6 Annex.....	15

1 Preface

1.1 Terminology

The model described in this documents references the following entities:

- a **user** downloads (and possible reads) a certain
- **document** from a
- **service/server** of a certain kind,
 - optionally mediated by a **referring document**,
 - optionally mediated by a **referencing context** and a **link resolver**
- which in turn generates **usage data**,
- which is stored in a certain **data format**
 - optionally **aggregated** for multiple services
- and offered by a **data provider** for
- harvesting by a central **service provider**

1.2 Data Format

The data format is based on XML which allows a formal verification of input and output data. It readily provides a lot of existent programming libraries for reading and writing data for a wide range of programming environments and languages.

Therefore, OpenURL ContextObject encoded in XML have been selected as the data format for the encoding of usage data. This choice provides existent XML schema documents for the verification out of the box.¹

The data which will be aggregated in the context of the OA Statistics project is more complex than the minimal set of data that the OpenURL ContextObject standard requires, but fully complies with those minimal requirements.

The OpenURL ContextObject format makes it possible to extend the format by adding custom data structures while staying compatible with the basic format. Thus the choice of OpenURL ContextObject is a compromise that was also made for compatibility with other projects with similar aims (e. g. in the Netherlands, UK, USA).

Moreover, the OpenURL ContextObject format offers an aggregation feature that allows for early combination of data records in a container for aggregated data storage:

```
<context-objects>
  <context-object> ... </context-object>
  <context-object> ... </context-object>
  <context-object> ... </context-object>
</context-objects>
```

¹ see ANSI/NISO Z39.88-2004, Part 3: The XML ContextObject Format, S. 47ff

1.3 Data Exchange

The data exchange between data provider and service provider is based on the widely established OAI Protocol for Metadata Harvesting (OAI-PMH)².

OAI-PMH was originally designed for the exchange of document metadata. Thus, this standard is mainly adapted to a specific way of handling a certain kind of metadata, as usage data does not meet the general requirements of typical formats used.

In principle, the protocol specifies a data synchronization mechanism which supports a reliable implementation of one-way data synchronization. OA Statistics exploits this synchronization functionalities.

The document-centric approach of OAI-PMH results in the following central problems:

- *Requirement for metadata record identifiers*³
While being irrelevant for the later use of the aggregation, nevertheless data providers must issue identifiers for data records to formally comply with OAI-PMH. These identifiers must be valid URIs.
- *Datestamp for records*⁴
OAI-PMH requires datestamps for all records of provided data. In the context of the OA Statistics project, this information has to be kept separately from the datestamp of the usage event itself:
 - Datestamp within the usage data contained within the metadata part of the OAI record, i. e. within the ContextObject's data: This is the time at which the actual usage event took place.⁵
 - Datestamp within the OAI-PMH record header: This is the time the ContextObject or the ContextObjects container has been stored in the database which feeds the OAI-PMH interface.
- *Mandated metadata in Dublin Core format*
This requirement may be lifted in the context of usage data since currently there is no direct use for this format itself. Nevertheless it is strongly recommended to implement it anyway to comply with the requirements for a standards compatible OAI-PMH interface. It seems advisable that the offered data is at least a rudimentary DC data set (identifier & description) which should logically describe the data offered and linked to by a certain identifier (see above regarding the identifier).
Example⁶:

```
<record>
  <header>
    ... (compare notes about the record header)
  </header>
```

² <http://www.openarchives.org/OAI/openarchivesprotocol.html>, referred to as OAI-PMH

³ OAI-PMH, 2.4

⁴ OAI-PMH, 2.7.1, also see below regarding OAI datestamps

⁵ Also see notes in the example data set given later on

⁶ Warning: the XML excerpts given in this document for example purposes do not necessarily contain all details regarding XML namespaces and XML schema. Nevertheless this omitted information is to be included in actual implementations and must not be considered optional!

```

<metadata>
  <dc xmlns="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <identifier>ID2</identifier>
    <description>
      Usage Event Data for Server ... from ... until ...
    </description>
  </dc>
</metadata>
</record>

```

Also, the choice of identifiers imposes problems: According to the OAI-PMH specification, the identifier within the DC metadata set must link to the described document. When understood as being metadata, the data contained in one <context-object> or in a <context-objects> aggregation is best described as being metadata of the usage events in a given time frame. Those usage events, however, regularly do not have their own identifiers yet. So in order to comply with DC requirements, too, identifiers have to be generated for those usage events as well (ID2 in the excerpt above). However, by now there seems to be no immediate use case for such identifiers. Therefore, in the context of the OA Statistics project, offering DC metadata is not required.

- *Usage of Sets*⁷
 OAI-PMH optionally allows for structuring the offered data in „sets“ to support selective harvesting of the data. Currently, the OA Statistics project does not exploit this possibility. Future refinements of this specification may use this feature, e. g. for selecting usage data for certain services.
 Provenance information is already included in the ContextObjects.
- *Datestamps, Granularity*⁸
 (also compare the notes about datestamps in the OAI-PMH record header versus datestamps within the ContextObjects)
 The OAI-PMH specification allows for either exact to the second or exact to the day granularity for record header datestamps. The data providers may chose one of these possibilities. The service provider implemented by the OA Statistics project will rely on overlapping harvesting, i. e. the most recent datestamp of the harvested data is used as the „from“ parameter for the next OAI-PMH query. Thus, the data provider will provide some records that have been harvested before. Duplicate records are matched by their identifiers (those in the OAI-PMH record header) and are silently tossed if their datestamp is not renewed (see notes below on deletion tracking).
 OA Statistics strongly recommends to implements exact to the second datestamps to keep the data redundancy as low as possible.
- *Deletion tracking*⁹
 The OAI-PMH provides functionalities for the tracking of deletion of records. Compared to the classic use case of OAI-PMH (metadata of documents) the use case of OA Statistics (usage data) falls in a category of data which is not subject of long-

⁷ see OAI-PMH, 2.7.2

⁸ see OAI-PMH, 2.7.1

⁹ OAI-PMH, 2.5.1

term storage. Thus, the tracking of deletion events does not seem critical since the data tracking deletions would summarize to a significant amount of data.

However, the service provider will accept information about deleted records and will eventually delete the referenced information in its own data store. This way it is possible for data providers to do corrections (e. g. in case of technical problems) on wrongly issued data.

It is important to note that old data which rotates out of the data offered by the data provider due to its age for storage reasons will not to be marked as deleted. This kind of data is nevertheless valid usage data, but not visible anymore.

The information about whether a data provider uses deletion tracking has to be provided in the response to the „identify“ OAI-PMH query within the <deletedRecords> field. In the context of the OA Statistics project, the only options are „transient“ (when a data provider applies or reserves the possibility for marking deleted records) or „no“.

The possible cases are:

- Incorrect data, that has already been offered by the data provider, shall be corrected. There are two possibilities:
 - Re-issuing of a corrected set of data carrying the same identifier in the OAI-PMH record header as the set of data to be corrected, with an updated OAI-PMH record header datestamp, or
 - When the correction is a full deletion of the incorrect issued data, the OAI-PMH record has to be reissued without a ContextObject payload, with specified „<deleted>“ flag and updated datestamp in the OAI-PMH record header.
- Records that fall out of the time frame for which the data provider offers data: These records are silently neglected, i. e. not offered via the OAI-PMH interface anymore, without using the deletion tracking features of OAI-PMH.

- *Metadata formats*¹⁰

All data providers have to provide support for <context-object> documents or <context-objects> aggregations, respectively.

This choice also has to be announced in the response to the „listMetadataFormats“ query¹¹ by the data provider. While a specific „metadataPrefix“ is not required, the information about „metadataNamespace“ and „schema“ is fixed for implementations:

```
<metadataFormat>
  <metadataPrefix>ctxo</metadataPrefix>
  <schema>http://www.openurl.info/registry/docs/xsd/info:ofi/▶
    ▶fmt:xml:xsd:ctx</schema>
  <metadataNamespace>info:ofi/fmt:xml:xsd:ctx</metadataNamespace>
</metadataFormat>
```

- *Inclusion of ContextObjects in OAI-PMH records*

Corresponding to the definition of XML encoded ContextObjects as data format of the data exchanged via the OAI-PMH, the embedding is to be done conforming to the OAI-PMH:

¹⁰ OAI-PMH, 3.4

¹¹ OAI-PMH, 4.4

```

<record>
  <header>
    <identifier>urn:uuid:fd23522e-c447-4801-9be4-c93c60a2d550▶
      ▶</identifier>
    <timestamp>2009-06-02T14:10:02Z</timestamp>
  </header>
  <metadata>
    <context-objects xmlns="info:ofi/fmt:xml:xsd:ctx">
      <context-object> ... </context-object>
      <context-object> ... </context-object>
    </context-objects>
  </metadata>
</record>

```

In the aforementioned example, the OAI-PMH record is identified by a UUID (in form of a URI).¹²

When offering single <context-object> documents rather than an aggregation using <context-objects> containers like above, a conformal OAI-PMH record may look like the following:

```

<record>
  <header>
    <identifier>urn:uuid:fd23522e-c447-4801-9be4-c93c60a2d550▶
      ▶</identifier>
    <timestamp>2009-06-02T14:10:02Z</timestamp>
  </header>
  <metadata>
    <context-object xmlns="info:ofi/fmt:xml:xsd:ctx"▶
      ▶timestamp="2009-06-01T19:20:57Z">
      ...
    </context-object>
  </metadata>
</record>

```

1.4 Related Projects and Initiatives

The design of this specification relies heavily on existing literature¹³ and is intended to be compatible to similar initiatives.

The OA Statistics data model is therefore closely modelled along the description of the usage of ContextObjects proposed by Bollen et al. 2006 and is only slightly extended to meet the needs of the OA Statistics project.

Moreover, the exchange and aggregation of usage data is more or less in line of Scenario B of the Final Report¹⁴ of the „Publisher and Institutional Repository Usage Statistics“ (PIRUS) project.

¹² see RFC 4122

¹³ Bollen, J. / Van de Sompel, H.: "An Architecture for the Aggregation and Analysis of Scholarly Usage Data". In: Proceedings of the Joint Conference on Digital Libraries, 2006

¹⁴ http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus_finalreport.pdf

1.5 Compatibility Aspects

The data format accepted by the OA Statistics service provider allows some variations on the side of the data provider. However, strategically it may make sense for data providers to not use these freedoms to ensure compatibility to other, more restrictive service providers.

In this regard, the following aspects have to be considered:

- Usage of <context-objects> containers
Other projects may not allow aggregations of single <context-object> documents within a <context-objects> container, although fully specified by the default ContextObject XML schema. Therefore, OAI-PMH records consisting of a single <context-object> document are perfectly valid for OA Statistics, as well.
- The ContextObjects Schema allows the specification of an identifier for every Context Object. As defined by the XML schema, one may use the „identifier“ attribute of the <context-object> element. This specification might be used for compatibility reasons, but is not mandatory for OA Statistics and is not analysed or used otherwise by the service provider:

```
<context-object
  xmlns="info:ofi/fmt:xml:xsd:ctx"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="info:ofi/fmt:xml:xsd:ctx ▶
    ▶http://www.openurl.info/registry/docs/xsd/info:ofi/▶
    ▶fmt:xml:xsd:ctx"
  identifier="urn:uuid:fd23522e-c447-4801-9be4-c93c60a2d550"
  timestamp="2009-06-11T16:05:08Z">
  ...
</context-object>
```

2 Data Provider: Infrastructure

The implementation of an abstract data provider basically operates along the following abstract steps:

1. Storage log data of a service as XML encoded Context Object:
 - either by creating a logger service that generates XML encoded Context Objects directly, or
 - by converting log data into this format.
2. Now, the data may be aggregated by storing the usage data of multiple usage events in a common <context-objects> container. This would reduce the amount of single data entities on the storage level.
3. Another option is the institutional aggregation of usage data, e. g. the usage data of different services (on possibly different servers) may be aggregated and then provided for access by a single central OAI-PMH data provider.
4. Provision of the data via an OAI-PMH data provider interface.

- The OAI-PMH data may also be compressed. This can be supported on the HTTP level by allowing for a compressing Content-Transfer-Encoding by the harvesting service provider, such that the data provider can compress the data stream.¹⁵

3 Data Format: Example with Comments

```
<context-objects
  xmlns="info:ofi/fmt:xml:xsd:ctx"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="info:ofi/fmt:xml:xsd:ctx http://www.openurl.info/▶
  ▶registry/docs/xsd/info:ofi/fmt:xml:xsd:ctx">
```

This starts the container for multiple ContextObjects. Both, the information on the XML namespaces and the schema information is mandatory.

```
<context-object timestamp="2009-06-11T16:05:08Z">
```

This starts an actual data set of a single usage event. It can be repeated within the aggregation (but remember to keep the amount within some bounds, see notes at the end of this document).

If the optional aggregation feature based on <context-objects> containers is not implemented, the XML namespace and schema details must be given as attributes to the <context-object> document itself instead.

The attribute „timestamp“ specifies the time of the usage event which is described by the data set and time has to be registered exact to the second and has to be implemented based on the ISO8601 norm.

The optional attributes „identifier“ and „version“ allowed by the ContextObject specification are not used by the OA Statistics service provider but may be additionally specified, e. g. for compatibility reasons.

```
<administration>
  <oa-statistics xmlns="http://dini.de/namespace/oas-info">
    <status_code>200</status_code>
    <size>234500</size>
    <document_size>234567</document_size>
    <format>application/pdf</format>
    <service>http://webdoc.sub.gwdg.de/</service>
  </oa-statistics>
</administration>
```

The <administration> container may be extended by specific data formats. OA Statistics uses this possibility to provide more information about the context of the usage event and data on the usage event itself:

- The XML namespace of the <oa-statistics> container has to be given as „http://dini.de/namespace/oas-info“.
- <status_code> specifies the HTTP status code for the HTTP query underlying the usage event,
- <size> specifies the transmitted amount of data

¹⁵ See RFC 2616 (HTTP 1.1), Sections 3.5, 14.3, 14.11

- <document_size> specifies the full size of the document of which a part is being transmitted
- <format> specifies the MIME type of the data transmitted
- <service> is an URI identifier of the service that has handled the usage event, e. g. the digital repository

As link resolver log data will not provide all information mentioned above. Therefore several elements have to be left aside, for example the <service> element. Compare the remarks on the <resolver> element of the ContextObject below.

```
<referent>
  <identifier>http://webdoc.sub.gwdg.de/diss/2000/kirchner/ii-▶
    ▶2.pdf</identifier>
  <identifier>urn:nbn:de:7-webdoc-12345</identifier>
  <identifier>oai:elib.uni-stuttgart.de-opus:100</identifier>
</referent>
```

The <referent> container describes the resource that the usage event was targeted at. The <identifier> element therefore contains an identifier of that resource and is repeated for every known identifier of the resource. This should also include identifiers of more abstract resources which the accessed resource is a part of, e. g. a journal identifier. In order to facilitate compatibility with the German OA Network project¹⁶, the OAI-PMH identifier of the resource's metadata (as issued by the repository's metadata OAI-PMH data provider) shall also be included. The latter will support the usage statistics to the documents covered by the OA Network Platform. All identifiers must be given in URI format.

```
<referring-entity>
  <identifier>http://goedoc.uni-goettingen.de:8080/▶
    ▶goescholar/handle/goescholar/2088</identifier>
  <identifier>urn:isbn:3-7643-6612-5</identifier>
</referring-entity>
```

If available, a HTTP referrer has to be included in the ContextObject's <referring-entity> element. This indicates the entity which was directing to the used resource at the time of the usage event (if it was not forged). As a minimal requirement, this would be the URL provided by the HTTP referrer string. Additionally all known other identifiers for that resource may also be specified.

```
<requester>
  <metadata-by-val>
    <format>http://dini.de/namespace/oas-requesterinfo▶
      ▶</format>
  <metadata>
    <requesterinfo xmlns="http://dini.de/namespace/oas-▶
      ▶requesterinfo">
      <hashed-ip>b505e629c508bdcfbf2a774df596123dd001▶
        ▶cee172dae5519660b6014056f53a</hashed-ip>
      <hashed-c>d001cee172dae5519660b6014056f5346d05▶
        ▶e629c508bdcfbf2a774df596123d</hashed-c>
      <hostname>uni-saarland.de</hostname>
      <classification>institutional</classification>
      <hashed-session>660b14056f5346d0</hashed-session>
```

¹⁶ <http://www.dini.de/projekte/oa-netzwerk/>

```

        <user-agent>mozilla/5.0 (windows; u; windows nt
        ▶5.1; de; rv:1.8.1.1) gecko/20061204▶
        ▶</user-agent>
      </requesterinfo>
    </metadata>
  </metadata-by-val>
</requester>

```

The `<requester>` element holds information about the agent that generated the usage event, basically identifying the user that triggered the event. This includes the IP address (in a hashed form for privacy reasons), Class-C network address (also hashed), host name (reduced to only first and second level domain name, also for privacy reasons), a classification of the agent, a session ID and User Agent string.

To integrate this detailed data set, the `ContextObject` has to be extended by a custom data scheme. Accordingly, the pre-defined extension facility of the `ContextObject` format is utilized to define a OA Statistics specific data scheme:

- `<format>` and XML namespace of the `<requesterinfo>` container must be set to „`http://dini.de/namespace/oas-requesterinfo`“.
- `<hashed-ip>` is the IP address corresponding to the usage event, made anonymous by using a salted hash function on the address string.
- `<hashed-c>` is the class-C network part of the IP address, made anonymous by the same means.
- `<hostname>` the client's host name, truncated to the second-level domain part, if available (some hosts may not have a hostname referring to their IP address)
- `<classification>` is a first classification of the agent that generated the usage event. If no classification is possible, it must be omitted. The following classifications can be given in an enumerated list, i.e. it must consist of one or more of the following string values (if multiple classifications are made, they must be given separated by white-space):
 - “internal”: classification for technical, system-internal accesses. Examples would be automated availability and consistency checks, *cron* jobs, keep-alive queries etc.
 - “administrative”: classification for accesses that are being made due to human decision but are for administrative reasons only. Examples would be manual quality assurance, manual check for failures, test runs etc.
 - “institutional”: classifies accesses that are made from within the institution running the service in question, regardless whether they are for administrative reasons.
- `<hashed-session>` optionally (if available/in use) specifies a session ID (if it is itself a hash) or a hash of a session ID of the complete usage session of a given user if available
- `<user-agent>` specifies the full HTTP user agent string if available.

```

<service-type>
  <metadata-by-val>

```

```

    <format>info:ofi/fmt:xml:xsd:sch_svc</format>
    <metadata>
      <abstract xmlns="info:ofi/fmt:xml:xsd:sch_svc">▶
        ▶yes</abstract>
    </metadata>
  </metadata-by-val>
</service-type>

```

The `<service-type>` element classifies the used resource. This is based on metadata in the format specified by the „info:ofi/fmt:xml:xsd:sch_svc“ scheme¹⁷. This catalogue of classifications may be extended in a later stage of the OA Statistics project.

The method of expressing this classification is prescribed by the ContextObject XML schema.

```

<resolver>
  <identifier>http://sfx.gbv.de:9004/sfx_sub/</identifier>
</resolver>
<referrer>
  <identifier>info:sid/dlib.org:dlib</identifier>
</referrer>

```

This additional information may be optionally specified and is only sensible for link resolver usage data (as opposed to web servers or repositories). The `<resolver>` element specifies the URL of the OpenURL resolver itself. The `<referrer>` element specifies the identifier of the context from within the user triggered the usage of the target resource which is given via the `<referent>` element, and which was itself referenced by the `<referring-entity>` element (see above).

```
</context-object>
```

This line closes a single ContextObject document, which marks the end of an OAI-PMH record if the optional aggregation within `<context-objects>` containers is not applied.

Otherwise, following the rules outlined above several ContextObjects may follow:

```

<context-object> ... </context-object>
<context-object> ... </context-object>
<context-object> ... </context-object>
...
</context-objects>

```

This finally closes the `<context-objects>` container and completes the OAI-PMH record if the aggregation feature is being used.

4 Data Format: External Policies

Information about the single elements of a ContextObject is described in the standardized schema describing XML encoded ContextObjects.¹⁸

This includes information about the appropriate order of elements, their binding character (mandatory or optional) and repeatability.

¹⁷ see http://www.openurl.info/registry/docs/xsd/info:ofi/fmt:xml:xsd:sch_svc

¹⁸ <http://www.openurl.info/registry/docs/xsd/info:ofi/fmt:xml:xsd:ctx>

A proper XML Schema for the encoding of OA Statistics specific information is provided in the annex of this specification.

5 General Comments

The responses to OAI queries should not exceed a size of about 50 MB in order to guarantee a reliable data transfer.

Similarly, the amount of aggregated records provided by the <context-objects> containers should be reduced accordingly.

This also has consequences for responses to „listRecords“ OAI-PMH queries, which are used for harvesting the data. A single response typically consists of many OAI-PMH records and thus has to be limited to a sensible number of records. The „resumptionToken“ feature of the OAI-PMH protocol allows for segmenting full data set into sequentially harvested parts.

Finally, to restrict stored and transmitted data to a manageable size, it is recommended to avoid unnecessary white space characters in the ContextObjects' structures.

6 Annex

oas-info.xsl:

```
<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://dini.de/namespace/oas-info"
  elementFormDefault="qualified"
  attributeFormDefault="unqualified">
  <xs:annotation>
    <xs:documentation>
      XML Schema defining the XML metadata format for usage event
      metadata, specifically for purposes within a larger client/server
      architecture for harvesting usage information. This specific
      metadata is only a subset of the whole metadata that describes a
      usage event. The main data structure is an OpenURL ContextObject.
    </xs:documentation>
    <xs:appinfo xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:dcterms="http://purl.org/dc/terms/">
      <dc:title>XML Format oas-info</dc:title>
      <dc:creator>Hans-Werner Hilse</dc:creator>
      <dc:creator>OA-Statistik Project (Germany)</dc:creator>
      <dc:description>
        This XML Schema defines a format to express usage event
        specific metadata as an XML document.
      </dc:description>
      <dc:identifier>http://dini.de/namespace/oas-info</dc:identifier>
      <dcterms:created>2009-10-12</dcterms:created>
    </xs:appinfo>
  </xs:annotation>
  <xs:element name="oa-statistics">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="status_code">
          <xs:annotation>
            <xs:documentation>
              specifies the HTTP status code for the HTTP query
              underlying the usage event
            </xs:documentation>
          </xs:annotation>
          <xs:simpleType>
            <xs:restriction base="xs:integer">
              <xs:pattern value="[0-9][0-9][0-9]"/>
            </xs:restriction>
          </xs:simpleType>
        </xs:element>
        <xs:element name="size" type="xs:nonNegativeInteger">
          <xs:annotation>
            <xs:documentation>
              specifies the transmitted amount of data
            </xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="document_size" type="xs:nonNegativeInteger">
          <xs:annotation>
            <xs:documentation>
              specifies the full size of the document of which a
              part is being transmitted
            </xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```

        </xs:documentation>
      </xs:annotation>
    </xs:element>
  <xs:element name="format">
    <xs:annotation>
      <xs:documentation>
        specifies the MIME type of the data transmitted
      </xs:documentation>
    </xs:annotation>
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:pattern value="\c+/\c+"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="service" type="xs:anyURI">
    <xs:annotation>
      <xs:documentation>
        an URI identifier of the service that handled the
        usage event, e.g. the document server
      </xs:documentation>
    </xs:annotation>
  </xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>

```

oas-requesterinfo.xsl:

```

<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://dini.de/namespace/oas-requesterinfo"
  elementFormDefault="qualified"
  attributeFormDefault="unqualified">
  <xs:annotation>
    <xs:documentation>
      XML Schema defining metadata for "Requester" metadata within the
      context of an OpenURL ContextObject. It was designed to express
      more information on a user of electronic resources but with
      careful consideration
      of the user's privacy.
    </xs:documentation>
    <xs:appinfo xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:dcterms="http://purl.org/dc/terms/">
      <dc:title>XML Format oas-requesterinfo</dc:title>
      <dc:creator>Hans-Werner Hilse</dc:creator>
      <dc:creator>OA-Statistik Project (Germany)</dc:creator>
      <dc:description>
        XML Schema defining metadata for "Requester" metadata within
        the context of an OpenURL ContextObject. It was designed to
        express more information on a user of electronic resources but
        with careful consideration of the user's privacy.
      </dc:description>
      <dc:identifier>http://dini.de/namespace/oas-requesterinfo
      </dc:identifier>
      <dcterms:created>2009-10-12</dcterms:created>
    </xs:appinfo>
  </xs:annotation>

```

```

<xs:element name="requesterinfo">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="hashed-ip" type="xs:hexBinary">
        <xs:annotation>
          <xs:documentation>
            the IP address from which the usage event originates,
            made anonymous by using a salted hash function on the
            address string
          </xs:documentation>
        </xs:annotation>
      </xs:element>
      <xs:element name="hashed-c" type="xs:hexBinary">
        <xs:annotation>
          <xs:documentation>
            the class-C network part of the IP address, made
            anonymous by the same means as for the IP address
          </xs:documentation>
        </xs:annotation>
      </xs:element>
      <xs:element name="hostname" type="xs:string" minOccurs="0">
        <xs:annotation>
          <xs:documentation>
            the client's host name, chopped to the second-level
            domain part. This element is to be omitted if there
            is no hostname for the client.
          </xs:documentation>
        </xs:annotation>
      </xs:element>
      <xs:element name="classification" minOccurs="0">
        <xs:annotation>
          <xs:documentation>
            further information about the class of the user which
            can only be determined locally at the service
          </xs:documentation>
        </xs:annotation>
        <xs:simpleType>
          <xs:list>
            <xs:simpleType>
              <xs:restriction base="xs:string">
                <xs:enumeration value="internal">
                  <xs:annotation>
                    <xs:documentation>
                      Usage events that happen just due to
                      internal system reasons, e.g.
                      automated integrity checks, high
                      availability checks etc.
                    </xs:documentation>
                  </xs:annotation>
                </xs:enumeration>
                <xs:enumeration value="administrative">
                  <xs:annotation>
                    <xs:documentation>
                      Usage events that happen due to
                      administrative decisions, e.g. for
                      quality assurance.
                    </xs:documentation>
                  </xs:annotation>
                </xs:enumeration>
              </xs:restriction>
            </xs:simpleType>
          </xs:list>
        </xs:simpleType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>

```

```

        <xs:enumeration value="institutional">
          <xs:annotation>
            <xs:documentation>
              Usage events triggered from within the
              institution running the service for
              which usage events are collected.
            </xs:documentation>
          </xs:annotation>
        </xs:enumeration>
      </xs:restriction>
    </xs:simpleType>
  </xs:list>
</xs:simpleType>
</xs:element>
<xs:element name="hashed-session" type="xs:hexBinary"
  minOccurs="0">
  <xs:annotation>
    <xs:documentation>
      optionally (if available/in use) specifies a session
      id or hash of it for the full usage session of the
      given user
    </xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="user-agent" type="xs:string" minOccurs="0">
  <xs:annotation>
    <xs:documentation>
      specifies the full HTTP user agent string if given
    </xs:documentation>
  </xs:annotation>
</xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>

```