



## MILESTONE 05

**Project Acronym:** OpenUp!  
**Grant Agreement No:** 270890  
**Project Title:** Opening up the Natural History Heritage for Europeana

---

### MS05 – Sample data service mock-up: data integrity service

**Revision:** Final

---

**Authors:**

**Anton Güntsch (BGBM) and the BGBM Biodiversity Informatics Team**

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	



## **REVISION HISTORY AND STATEMENT OF ORIGINALITY**

### **Revision History**

<b>Revision</b>	<b>Date</b>	<b>Author</b>	<b>Organisation</b>	<b>Description</b>
1	2011-03-28	Anton Güntsch	BGBM	1 <sup>st</sup> compilation of integrity service requirements and a diagram showing the basic information flows on the TMG scratchpad-site.
2	2011-06-20	Anton Güntsch & BDI Team	BGBM	Full specification including an xml schema for annotations and the URL of the service mockup.
2a	2011-06-21	Project coordination	BGBM	Minor edits

### **Statement of Originality:**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## Distribution

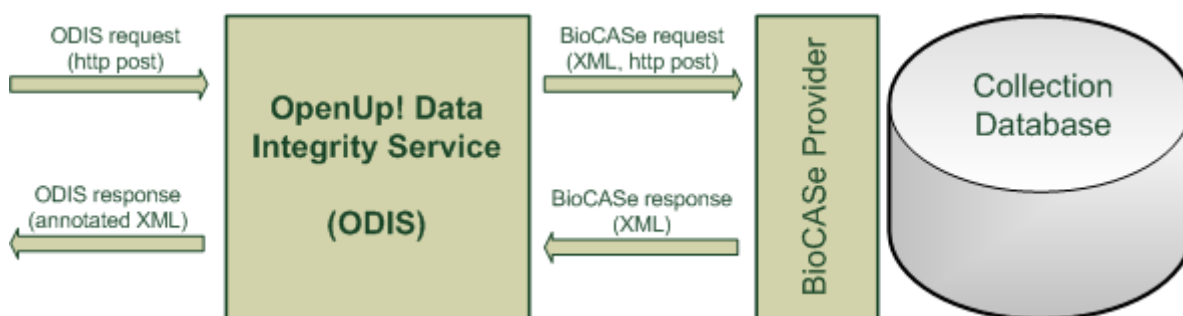
Recipient	Date	Version	Accepted YES/NO
TMG	2011-02-28	1	YES
Work Package Leader	2011-06-20	2	YES
Project Coordinator	2011-06-21	2a	YES

## MS05 – Sample data service mock-up: data integrity service

**Please note:** The following description has been copied from the OpenUp! Scratchpad site of the Technology Management Group (TMG). The deliverable itself is the integrity service mock-up, which is publicly available at <http://ww2.biocase.org/odis/queryform>.

### Overview

The OpenUp! Data Integrity Service (ODIS) is a rule-based system for checking the syntactical correctness of ABCD data accessible via a given BioCASE provider service installation. It can be used as a standalone service and will also be part of the Collections Data Quality Toolkit together with services for botanical and zoological names.





## ODIS requests

An ODIS request can be submitted by using HTTP Post with the following three key/value pairs:

Key	Value	Description
Provider URL	URL	The access point of the BioCASE provider installation to be analysed
Sync	true/false	Decides whether the client expects a synchronous response for the given request. If Sync=true is submitted the service will return a full ABCD document containing ABCD unit-records containing annotations encoded as XML-comments. If Sync=false, the service will return a URL from which annotated ABCD-document can be downloaded.
BioCASeRequest	BioCASE-Protocol 1.3 search request	The BioCASeRequest value is used to restrict quality control to a reasonable number of records, for example by considering a specificity genus only. As the BioCASE protocol request format has already the necessary mechanisms to filter records in a SQL-like XML-encoded syntax, we are using this very protocol syntax here as well. For a detailed description of the BioCASe protocol see <a href="http://www.biocase.org/products/protocols/">http://www.biocase.org/products/protocols/</a> .

## ODIS responses

The ODIS response will be an annotated ABCD-document, which will be delivered directly or as a downloadable file depending on the value of the Sync-argument in the ODIS request. If sync=false has been requested the client receives the URL of the ABCD document to be downloaded. In this case, trying to fetch the result-document before it is completely available for download will produce an HTTP server error (temporarily not available).

For storing annotations generated by the integrity checker service we use XML comments (<!-- comment -->) with the assumption that an annotation is always directly following the opening tag of the element it belongs to. The following example shows an annotation of a scientific name belonging to a determination in an ABCD 2.06 document:

```
<?xml version='1.0' encoding='UTF-8'?>
<DataSets xmlns='http://www.tdwg.org/schemas/abcd/2.06'>
<DataSet>
  <TechnicalContacts></TechnicalContacts>
  <ContentContacts>
    <ContentContact>
      <Name>John Smith</Name>
      <Email>j.smith@NaturalHistoryCollection.org</Email>
    </ContentContact>
  </ContentContacts>
  <Metadata>
    <Description>
      <Representation language='en'>
```



```
<Title>herbarium collection</Title>
</Representation>
</Description>
<RevisionData>
  <DateModified>2001-03-01T00:00:00</DateModified>
</RevisionData>
</Metadata>
<Units>
  <Unit>
    <SourceInstitutionID>BEBOP</SourceInstitutionID>
    <SourceID>HerbCol</SourceID>
    <UnitID>1136</UnitID>
    <Identifications>
      <Identification>
        <Result>
          <TaxonIdentified>
            <ScientificName>
              <FullScientificNameString>Calendula arvensis (Vail.) L.</FullScientificNameString>
            </ScientificName>
          </TaxonIdentified>
        </Result>
      </Identification>
      <Identification>
        <Result>
          <TaxonIdentified>
            <ScientificName>
              <FullScientificNameString>
                <!--
              <Annotations>
                <Annotation>
                  <Context>OpenUp</Context>
                  <ISODateTime>2011-03-29T12:24Z</ISODateTime>
                  <MethodOrAgent>ODIS V0.7</MethodOrAgent>
                  <Type>Warning</Type>
                  <Message>Scientific name string seems to be malformed</Message>
                  <Suggest>Calendula incana Willd.</Suggest>
                </Annotation>
              </Annotations>
            </FullScientificNameString>
          </ScientificName>
        </TaxonIdentified>
      </Result>
    </Identification>
  </Identifications>
</Unit>
</Units>
</DataSet>
</DataSets>
```



The precise syntax of an annotation is defined with the schema [annotation.xsd](#). Each annotation consists of the following 6 elements:

Element name	Description
Context	A short but meaningful string indicating the context of an annotation. For annotations in the context of OpenUp! this will simply be "OpenUp". The context can be used for example for searching for annotations that are relevant for a specific purpose.
ISODateTime	ISO8601 encoded date and time.
MethodOrAgent	An indication of who or what created an annotation. This might be a software application, a service, or a person. In the context of OpenUp, this will be something like "ODIS V1.7".
Type	Type of annotation such as "comment" or "warning" we should try to define a controlled vocabulary here.
Message	A free-text representation of the annotation (for human consumption).
Suggest	In some cases the service might be able to suggest a correction of the annotated element.

If the client chooses asynchronous access (`sync=false`), the server/service will return the HTTP status code 303 ("see other") and the HTTP header "Location" with the new URI for download. The body will also contain this URI as an html-encoded link. If a client tries to access the response file before it has been put up for download, the server will return the http status code 503 ("service unavailable") and the http header will contain a message "retry-after number", with number being the number of seconds to wait before the response document can be expected. For a detailed list of HTTP status codes used by OpenUp! validation services please refer to [A common API for name data quality services](http://open-up.eu/content/common-api-name-data-quality-services#http_status_codes) ([http://open-up.eu/content/common-api-name-data-quality-services#http\\_status\\_codes](http://open-up.eu/content/common-api-name-data-quality-services#http_status_codes)).

## **ODIS Mock-Up**

URL: <http://ww2.biocase.org/odis/queryform>

The ODIS Mock-Up demonstrates the integrity service interface without implementing it to a full extent. However, for a limited set of example queries it is fully compliant to its specification so that it can be included in other implementations already. The ODIS Mock-Up access point is INSERT URL here. Please note, that the production-ready service will be installed at a different location.

The service "simulates" a simple rule-set for

- checking the correctness of ISO-3166 country codes
- checking the syntactic correctness of URLs

It is also capable of sending http status codes (e.g. 303 - "see other" for asynchronous responses).