

# TAPIR Workshop February 2007 (12 to 15) GBIF, Copenhagen

## 1. Summary

TAPIR, the TDWG Access Protocol for Information Retrieval, is being actively developed. There is a draft specification<sup>1</sup>, two fully functional data provider software packages (PyWrapper<sup>2</sup> and TapirLink<sup>3</sup>) and a TAPIR Network Builders Guide will soon be available.

The aims of this workshop were to advance the deployment of the TAPIR protocol and to build the community of TAPIR expertise. The workshop started with a training session and then a hands-on development sessions in separate working groups.

The list of participants forms the Appendix. For more information please contact Renato De Giovanni: [renato \[at\] cria . org . br](mailto:renato[at]cria.org.br). For more information about the meeting, see <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirWorkshop2007>

## 2. Goals

- Explain the main features and concepts of TAPIR to new developers.
- Amend the existing TAPIR documentation where needed.
- Develop (at least partially) a new TAPIR client software including libraries and specific data harvesters to be used by biodiversity data networks.
- Discuss, implement and document LSID resolution service in existing TAPIR provider software.
- Create a generic query tool to interact with TAPIR providers.
- Create new output models and query templates to be used by TAPIR networks.
- Plan and implement interoperability unit tests for TAPIR providers.
- Develop proposals related to further advance TAPIR.

## 3. Outcomes

- The training session included the following presentations and demonstrations:
  - History and overview of TAPIR (Renato De Giovanni).
  - Main concepts about TAPIR: conceptual schemas, output models, query templates, concept name servers (Markus Döring).

---

<sup>1</sup> <http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/>

<sup>2</sup> <http://www.pywrapper.org/>

<sup>3</sup> [http://sourceforge.net/project/showfiles.php?group\\_id=38190&package\\_id=217873](http://sourceforge.net/project/showfiles.php?group_id=38190&package_id=217873)

- Demonstrations (Javier de la Torre).
- Guidelines for building TAPIR networks (Renato De Giovanni).
- All participants were invited to talk about their interests and plans related to TAPIR. This information has been summarized at the end of the wiki page of the workshop<sup>4</sup> and was used to update the original list of proposed themes for discussion and initial prototyping. Work continued in separate groups.
- All discussions about a TAPIR client were focused on the “harvesting” functionality. A general design was initiated<sup>5</sup> and development will probably follow as some of the participants demonstrated interest to do it as part of their own projects.
- A new theme for discussion was related to Concept Name Servers. It was agreed that an enhanced service will soon be needed for issues such as identifying what query templates and output models are available for a specific set of concepts and determining equivalent concepts across different conceptual schemas<sup>6</sup>. One of the participants started prototyping a PHP implementation for a Concept Name Server.
- A test suite for TAPIR provider implementations was discussed<sup>7</sup>
- A simple Java implementation capable of validating responses to specific requests was discussed. One of the participants demonstrated interest in continuing the implementation afterwards.
- During the meeting, a number of XML Schema validation issues were detected with the new DarwinCore schema. These problems were addressed and an “official” DarwinCore response structure for TAPIR was created<sup>8</sup>, and two corresponding output models for ABCD and DarwinCore mappings produced. The existing RSS2 and KML output models for ABCD 2.06 were normalized (the response structure definition was separated from the output model and just referenced by it) and two corresponding output models were created for DarwinCore mappings.
- GBIF started development of a TAPIR indexer which will be available in the new GBIF portal. The indexer will be able to interact with TAPIR providers that have mapped their local databases using the new DarwinCore (including the geospatial and curatorial extensions) or ABCD 1.20/2.05/2.06 ‘standards’.
- There was a lot of discussion about how to use TAPIR with the existing TDWG ontology, in particular about how to derive “TAPIR-usable” concepts from the ontology. In cases where properties from generic classes in the ontology need to be referenced in a more specific context (like Person.Name in the context of a collector name) it was suggested to create concepts that also consider relationships from the ontology (like /CollectionEvent/HasCollector/Person/Name indicating collector name). There was initial prototyping to produce such concepts from the ontology, map them in a TAPIR instance and then generate RDF output from the TAPIR provider. Related discussions can be found in the TDWG Architecture Group wiki<sup>9</sup>.
- Some tests were initiated about LSID resolution with TAPIR and work will continue by one of the participants.

---

<sup>4</sup> <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirWorkshop2007>

<sup>5</sup> <http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirClientDesign>

<sup>6</sup> <http://wiki.tdwg.org/twiki/bin/view/TAPIR/ConceptNameServer>

<sup>7</sup> <http://wiki.tdwg.org/twiki/bin/view/TAPIR/ProviderServiceTests>

<sup>8</sup> [http://rs.tdwg.org/dwc/tdwg\\_dw\\_record\\_tapir.xsd](http://rs.tdwg.org/dwc/tdwg_dw_record_tapir.xsd)

<sup>9</sup> <http://wiki.tdwg.org/twiki/bin/view/TAG/LsidVocs>

- Both existing TAPIR provider software packages were adjusted during the workshop. One of the participants also discovered how run a TAPIR provider in IIS with “clean URLs” (when the resource code is part of the REQUEST\_URI and not an HTTP query parameter).
- The existing TAPIR mailing list will be used to discuss any developments related to TAPIR.
- The following changes in the protocol were suggested during the workshop and then incorporated in the specification and in the XML Schema:
  - The "envelope" parameter on search operations should be turned on by default.
  - Include a note about the possibility of restricting the allowable domains related to the location of style-sheets that are specified through the "xslt" parameter. When the style-sheet comes from an unknown location it can be ignored with a corresponding warning being raised in the diagnostics section.
  - Add specific XML Schema capability for xsd:include.
  - Include note about the fact that a provider is not forced to guarantee the entire validity of search responses according to the XML Schema defined in the response structure, except to the extent of its own declared XML Schema capabilities.
  - Include a recommendation for providers to raise warnings instead of errors when an unsupported XML Schema construct is found in the response structure.
  - Include a new search parameter called "omit-ns" instructing providers to omit or not, anynamespace declarations and the corresponding prefixes in search responses when the envelope is turned off.
  - Include a new attribute "required" for concepts in the output model mapping (defaults to false). When “required” is set to true, providers should raise an error if the concept is unmapped, or the corresponding value is null.
  - Node paths in output model mapping are not considering namespaces. When the response structure needs to make use of different namespaces, all namespaces need to be declared in the output model element, and the nodes referenced in the xpaths must include the associated namespace prefix.

## 4. Conclusions

All participants had the opportunity to get more information about TAPIR and to address any uncertainties. A list of suggestions to improve the protocol and to clarify the specification was proposed. There are good perspectives that many developments that were discussed and prototyped during the workshop will be further developed by the participants as part of their own plans and projects. It was strongly suggested to make use of any available collaborative environment so that other people can participate. The meeting was considered successful and there is great expectation about new TAPIR software and the first TAPIR networks.

# Appendix

## List of participants

- Donald Hobern, Global Biodiversity Information Facility, dhobern [at] gbif . org
- Giorgos Ksouris, Global Biodiversity Information Facility, ksouris [at] gbif . org
- Greg Whitbread, Australian Dept. of Environment and Heritage, ghw [at] anbg . gov. au
- Javier de la Torre, Imaste-IPS, jatorre [at] imaste-ips . com
- John Wieczorek, Museum of Vertebrate Zoology, Berkeley, tuco [at] berkeley . edu
- Jörg Holetschek, BGBM (Botanischer Garten und Botanisches Museum Berlin-Dahlem), j . holetschek [at] bgbm . org
- Jose Miguel Cuadra, INBio (Instituto Nacional de Biodiversidad), josecuadra [at] gmail . com
- Kevin Richards, Landcare Research, RichardsK [at] landcareresearch . co . nz
- Markus Döring, BGBM, m . doering [at] bgbm . org
- Milko Skofic, Biodiversity International, m . skofic [at] cgiar . org
- Patrick Leary, Marine Biological Laboratory, pleary [at] mbl . edu
- Renato De Giovanni, CRIA (Centro de Referência em Informação Ambiental), renato [at] cria . org . br
- Roger Hyam, TDWG (Biodiversity Information Standards), roger [at] tdwg.org
- Wouter Addink, ETI Bioinformatics, wouter [at] eti . uva . nl
- Youjun Guo, Yale University, youjun . guo [at] yale . edu