Deliverable 1.4 (D1.4)

Summary report of operational EU BON services and data provision for the European taxonomic backbone

M48

Project acronym:	EU BON
Project name:	EU BON: Building the European Biodiversity Observation Network
Call:	ENV.2012.6.2-2
Grant agreement:	308454
Project duration:	01/12/2012 – 31/05/2017 (54 months)
Co-ordinator:	MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany

Delivery date from Annex I:	M48 (November 2016)
Actual delivery date:	M48 (November 2016)
Lead beneficiary:	FUB-BGBM
Authors:	Andreas Kohlbecker (BGBM), Anton Güntsch (BGBM), Andreas Müller (BGBM), Eckhard von Raab-Straube (BGBM), Florian Wetzel (MfN), Caroline Ring (MfN), Matúš Kempa (IBSAS), Pavel Stoev (PENSOFT), Aaike De Wever (Naturalis), Guido Sautter (PLAZI), Donat Agosti (PLAZI)

1. Executive Summary

1.1. Introduction

The goal of the deliverable is to provide a technical summary of web services providing a harmonised and unified access layer to taxonomic information resources defined in Appendix 3 of the INSPIRE Directive. We describe the basic web service architecture, the unified access protocol, the harmonisation of contributing databases, integration into the EU BON portal, as well as its documentation.

1.2. Progress towards objectives

The overall objective of the EU BON taxonomic backbone is to improve and harmonize European and national level taxonomic backbone data for European plant and animal species. The main sources for the European taxonomic backbone are the databases Fauna Europaea, Euro+Med Plantbase and the European Register of Marine Species (ERMS) that focus on different taxonomic groups. To this end, we focused on the following activities:

Database harmonisation: migration of the Fauna Europaea database to the EDIT Platform for Cybertaxonomy (http://cybertaxonomy.eu/). Implementation of a new data portal and services. Integration into the merging process of the Pan-European Species directories Infrastructure. Development of a fully functional remote-editing environment for the EDIT Platform for Cybertaxonomy. User workshops.

Database updates: Continuous updates of the databases contributing to the EU BON taxonomic backbone based on the gap analysis provided by task 1.3.

Service layer: Implementation of a web service layer providing unified access to relevant taxonomic data sources, in particular to the sources defined in Appendix 3 of the INSPIRE Directive. Integration of this service layer into the EU BON Portal. Documentation.

Service documentation and registration: Thorough and standardised documentation of services as well as registration in a community service registry for Metadata annotation and discovery.

Service testing: Set up of a framework for automated service monitoring and analysis. Performance improvements.

1.3. Achievements and current status

The migration of Fauna Europaea to the EDIT Platform for Cybertaxonomy has been completed. The database of Fauna Europaea (FaEu, http://www.fauna-eu.org) assembles all living, multi-cellular, land and freshwater animals in Europe based on the current state of knowledge. In the most recent version of FaEu, 145,000 metazoan terrestrial and freshwater species and subspecies are listed, in total the database covers about 235,708 taxa. On a European scale, it is one of the most relevant data sources in the field of taxonomy for terrestrial animal species. Fauna Europaea is the standard reference for political decisions in biodiversity questions by its provision to PESI and the "Infrastructure for Spatial Information in the European Community" (INSPIRE). Database entries are fully expert based, which is a unique feature of the service.

All data have been incrementally imported into a Fauna Europaea instance of a Common Data Model (CDM) compliant database using the EDIT Platform import layer. This activity contributed to the long-term sustainability of the portal, which was hosted on an outdated server infrastructure before. The EDIT platform portal software has been modified to meet the needs of the European Faunistic community. The new Fauna Europaea Portal is now available at http://fauna-eu.org/ and hosted by the Museum für Naturkunde, Berlin, Germany.

With both Euro+Med Plantbase and Fauna Europaea being migrated to the EDIT Platform a major step towards harmonisation of data prepared and provided by the European taxonomic communities has been taken. This affects not only the technical merging of checklist data into the data warehouse

of the Pan-European Species directories Inventory providing the priority resource mentioned in the INSPIRE directive. Also, it now allows to share service and user documentations, to harmonise server infrastructures, and to jointly organise and hold user tutorial workshops. In parallel to the technical works on the backbone databases, the participating taxonomists were (and are) continuously updating the taxonomic backbone. Updates are cyclically published via the PESI workflow for merging and publishing the major European checklist databases Euro+Med Plantbase, Fauna Europaea, European Register of Marine Species (ERMS), and Index Fungorum.

Based on the improved structure and content of the European checklists infrastructure, a new Unified Taxonomic Information Service has been implemented, tested, and published (UTIS, http://cybertaxonomy.eu/eu-bon/utis/). The service implements a unified protocol for simultaneous access to the following (INSPIRE-conform) taxonomic information systems:

- Pan-European Species directories Infrastructure (EU-Nomen)

- EUNIS

- Natura2000

In addition, the service mediates access to the Catalogue of Life (CoL) and the World Register of Marine Species (WORMS). To speed up queries requiring data from less capable source databases (e.g. EUNIS), a caching mechanism has been integrated using a NoSQL database, which is particularly fast responding on graph-like (taxonomic) data structures.

A service documentation for developers is consistently integrated into the service endpoints and queries can be manually composed and tested using pre-cooked forms.

The new service layer has been implemented into the EU BON portal and provides query expansion capabilities for taxon-level requests. In addition, the services have been registered in the BiodiversityCatalogue (www.biodiversitycatalogue.org/services/79) to increase their visibility and advertise their use outside the context of EU BON.

For automated testing of services, a standardised Apache JMeter (https://jmeter.apache.org/) based platform has been set up which sends and evaluates service requests to UTIS. This platform was also used to test and significantly improve the PESI services providing the major contribution of taxonomic content to UTIS.

1.4. Future developments

On the basis of the working EU BON infrastructure for taxonomic data, we will continue to update the underlying databases. For example, the completion of Euro+Med plantbase (http://ww2.bgbm.org/EuroPlusMed) is expected for the beginning of 2017. Also for Fauna Europaea updates of the existing database, i.e. the taxonomic content, are planned in 2017. A suitable tool for this updating process is available and was already tested at the MfN (the so-called Taxonomic Editor). Furthermore, the BGBM together with VLIZ will continue to prepare, perform, and publish merges of the PESI backbone feeding into UTIS.

To facilitate UTIS-integration into R-based workflows and analyses, a connector module has been prototyped and will be integrated into analyses services provided by the EU BON Portal (www.biodiversity.eubon.eu).

The BGBM will also continue to host UTIS and support its integration into the emerging landscape of biodiversity informatics services and applications. It is planned, for example, to integrate UTIS into the "Terminology Services" of the German Federation for Biological Data (GFBio, http://www.gfbio.org/) as a basis for annotating research data during the archiving process. Furthermore, the integration of UTIS as a taxonomic query expansion mechanism for BioCASE specimen portals is currently being assessed.

Table of content

2. Service architecture

The unified taxonomic backbone for EU BON is a mechanism for integrating and harmonising relevant taxonomic resources including the Pan-European Species directories Infrastructure (PESI, http://www.eu-nomen.eu, [de Jong et al. 2015]), Catalogue of Life (CoL), EUNIS, Natura2000, and the World Register of Marine Species (WoRMS), some of which are explicitly requested in the INSPIRE directive:

"*EU-Nomen is the preferred reference list to be used. If a taxon is listed in EU-Nomen, this reference must be used as first choice. If it is not listed in EU-Nomen, the second choice is EUNIS, if not in EUNIS, Natura2000 can be used.*"[Appendix 3 Recommendation 7 in INSPIRE – Data Specification on Species Distribution, Draft Technical Guidelines: p.34-35].

The chosen architecture for the EU BON taxonomic backbone reflects the directive by mediating access to the listed resources and returning results following the defined priorities.

2.1. Unified Taxonomic Information Service (UTIS)

The EU BON taxonomic backbone has been implemented as a generic engine for federated searches for taxonomic information (UTIS, Unified Taxonomic Information Service). The EU-BON taxonomic backbone is an instance of UTIS which is publicly available at http://cybertaxonomy.eu/eu-bon/utis. The UTIS software is based on the Data Refinement Workflow (Mathew et al. 2014) library developed in the context of the EU 7th Framework project BioVeL (http://www.biovel.eu/, [Hardisty et al. 2016]). The source code of this library is available from GitHub (https://github.com/BioVeL/drf-workflow.git). By building the EU-BON taxonomic backbone on the foundations of the Data Refinement Workflow library EU BON can benefit from optimisations and implementations made by BioVeL and vice versa.

Two versions of the UTIS have been released so far. Each of these releases is installed on a server at the BGBM and is available at a version specific URI:

- Version 1.0: http://cybertaxonomy.eu/eu-bon/utis/1.0/
- Version 1.2: http://cybertaxonomy.eu/eu-bon/utis/1.2/

The main URI http://cybertaxonomy.eu/eu-bon/utis redirects to the latest version.

When accessed with a web browser the service delivers a web page which links to the human-readable documentation of the web service API (**Fig. 1**).

Figure 1: UTIS service documentation web page.

2.2. EU BON taxonomic backbone as federated search engine

The EU BON taxonomic backbone has been developed as federated search engine that connects an extensible set of checklists. Requests sent to the taxonomic backbone are simultaneously distributed to the connected data providers. Result sets returned from the individual checklists are then mapped into a unified data model. Finally, the results are combined and presented to the client, who sent the request (**Fig. 2**).
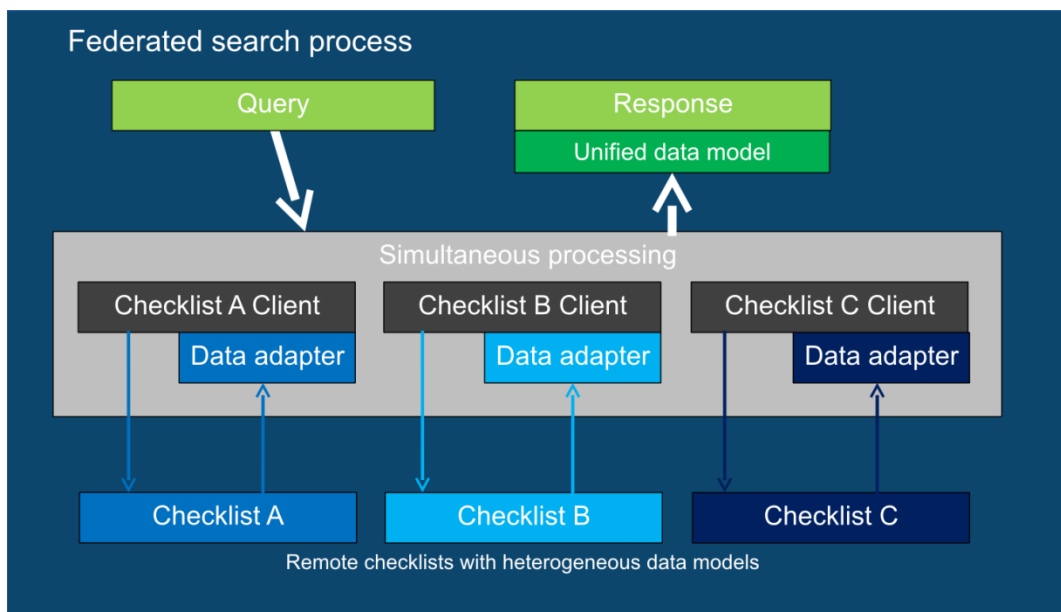

Figure 2: Basic UTIS information flow.

2.3. Assuring conformity with the INSPIRE directive

In order to conform with INSPIRE Appendix 3, Recommendation 7 a defined order of priorities is crucial for this architecture. The taxonomic backbone therefore allows to configure these priorities and returns responses accordingly. The results of the checklist with the highest priority are returned first, followed by the results of checklist with the next lower priority, and so on (**Fig. 3**).

2.4. Software components

All source code is publically available as open source code from the BGBM git repository following the Mozilla Public License Scheme Version 1.1 (https://www.mozilla.org/en-US/MPL/1.1/). The UTIS software consists of three major parts:

- **utis-core**: Core library providing the federated search backend, checklist adapters, data conversion and harmonization functionality. (https://dev.e-taxonomy.eu/gitweb/utis-core.git)

- **utis-webapp**: Web application module, which allows running utis-core as RESTful web service. (https://dev.e-taxonomy.eu/gitweb/utis-webapp.git)

- **utis-client**: This library provides the full utis functionality to client applications like the EU BON Portal. Client applications can use this library to consistently access the UTIS web service in an easy to use and reliable way. The whole client library is generated automatically on base of the machine readable documentation provided by swagger (see below). (https://dev.e-taxonomy.eu/gitweb/utis-client.git)

The web application component utis-webapp integrates the Swagger (http://swagger.io/) framework which provides powerful mechanisms to automaticaly generate technical representations of the RESTful web service API. This way it can be guaranteed that the documentations of the web service API are always up to date and consistently presented. The machine readable documentation produced by Swagger is being used in other subsequent processes of the application building process:

1. The utis-client library, is directly generated from the swagger representations.

2. The HTML front-end mainly consists of an interactive service documentation. The entire documentation is generically built based on the Swagger API representations. Please refer to chapter 4.1 for more details on the documentation.
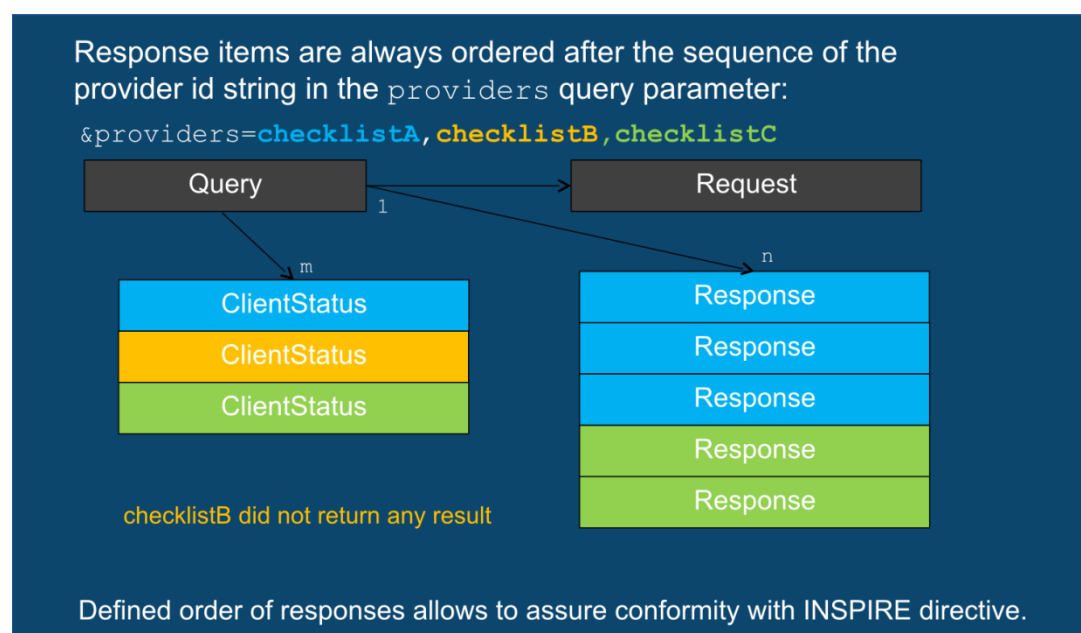

Figure 3: Configuration of priorities.

Query responses always contain the full taxonomic information as returned by each individual checklist. Each result-set entry can therefore always be traced back to its source.

2.5. Connecting checklist providers

For each of the checklist providers, UTIS has an adaptor that connects to the specific web service in order to send requests, retrieve the responses and to harmonize them into the unified data model. An exception to this principle is the EUNIS information system:

EUNIS is exposed through a SPARQL web service maintained by the EEA (http://www.eea.europa.eu/), which turned out to be too slow for our use case. UTIS needs to wait for the responses of all providers which are requested simultaneously during a federated search. One slow service involved into the overall request can thus slow down the whole process. In order to circumvent this bottleneck it has been decided to cache all relevant EUNIS data locally in an NoSQL database. This also correlates with the recommendation of the EEA that users get the RDF directly from the source for large queries.

The EUNIS service is periodically queried for changes. In case updated data are detected by UTIS, the relevant data are fetched as an archive containing an RDF/XML database dump. This dump is imported through a Tinkerpop Blueprints (https://github.com/tinkerpop/blueprints/wiki) pipeline into a Neo4j graph store (https://neo4j.com/). The benefit of using a graph store is that these systems are providing an excellent performance of operations that need to traverse graphs. Furthermore, Neo4j has a freetext index together with an extremely fast search engine. By using this offline strategy, UTIS can now process EUNIS queries in a few milliseconds.

2.6. REST service endpoints

UTIS provides a RESTful web-service layer with four endpoints:

- /capabilities : List the available checklist providers. The information for each provider includes, documentation url, copyright information, version and supported search modes.

- /search : Execute a federated search. The search endpoint supports the different search modes:

  o scientificNameExact

  o scientificNameLike (begins with)

  o vernacularNameExact

  o vernacularNameLike (contains)

  o findByIdentifier

  Each of the modes can be combined with a set of options:

  - addSynonymy: Adds the synonymy of the accepted taxon the response.

  - addParentTaxon: Adds the parent taxon of the accepted taxon to the response.

  - dedup: Allows to deduplicate the results by making use of a deduplication strategy. The deduplication is done by comparing specific properties of the taxa in the result set.

- /classification/{taxonId}/children: List all taxonomic children of the taxon specified by the given identifier ({taxonId}).

- /classification/{taxonId}/parent: List the higher taxa of the taxon specified by the given identifier ({taxonId}) in the order as they occur in the branch of the classification tree.

All response data are returned as JSON (http://json.org/).

Fig. 4 shows an overview of the tnr (Taxon Name Resolution)-schema, which is the return type of search and classification operations. The response is contained in a query object together with a tnrRequest object. Repeating the request together with the response is advantageous when further processing the data in asynchronous pipelines like in workflow environments.
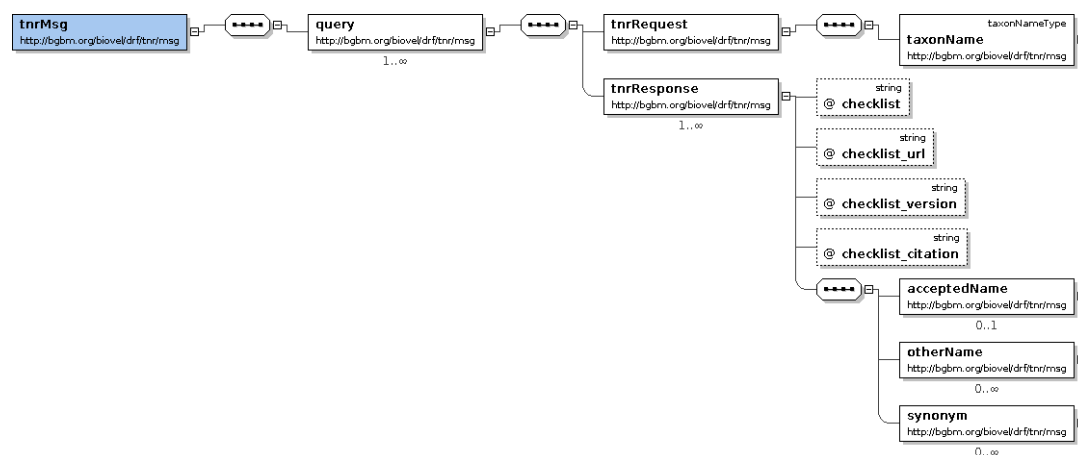

Figure 4: Taxon Name Resolution Schema.

For more details on the RESTful API please refer to online documentation at http://cybertaxonomy.eu/eu-bon/utis/1.2/doc.html#!/utis-controller.

3. Harmonisation of underlying taxonomic databases and data provision

3.1. Database harmonisation

An important aspect of the EU BON taxonomic backbone is the harmonisation of contributing database systems. To this end, the Fauna Europaea information system [de Jong et al. 2014] has been migrated to the EDIT Platform for Cybertaxonomy, which provides already the technical infrastructure for Euro+Med Plantbase.

The Platform for Cybertaxonomy (http://www.cybertaxonomy.org, [Berendsohn 2010]), developed as part of the EU NoE EDIT (European Distributed Institute of Taxonomy), is an open-source software framework covering the breadth of the taxonomic workflow, from fieldwork to publication [Ciardelli et al. 2009]. It provides coupled tools for full, customized access to taxonomic data, editing and management, and collaborative team work. At the core of the platform is the Common Data Model (CDM), offering a comprehensive information model covering all relevant data domains: names and classifications, descriptive data (morphological and molecular), media, geographic information, literature, specimens, persons, and external resources (http://dev.e-taxonomy.eu/trac/wiki/CommonDataModel). The model adheres to community standards, including the Biodiversity Data Ontology developed by the Biodiversity Information Standards organization TDWG (http://www.tdwg.org). Apart from its role as a software suite supporting the taxonomic workflow, the platform is a powerful information broker for a broad range of taxonomic data providing solid and open interfaces including a Java programmer's library and a CDM Rest Service Layer.

- The Fauna Europaea migration included the following steps:

- Data import from the existing (Oracle) Fauna Europaea database using the EDIT Platform Import Layer with additional methods for dealing with specific Fauna Europaea data structures.

- Set up of a CDM-service instance at the Museum für Naturkunde (MfN) hosting the Fauna Europaea information system.

- Implementing a new Fauna Europaea Portal (http://www.fauna-eu.org/) based on the existing Drupal-based EDIT Platform Portal software. The new web portal was launched at the end of 2015.

- Updating the webpage content and creating a new graphical surface of Fauna Europaea by the graphic designers at the MfN (**Fig. 5**).

- Organising a series of EDIT Platform tutorial workshops Fauna Europaea data managers and involved taxonomists at the Museum für Naturkunde.
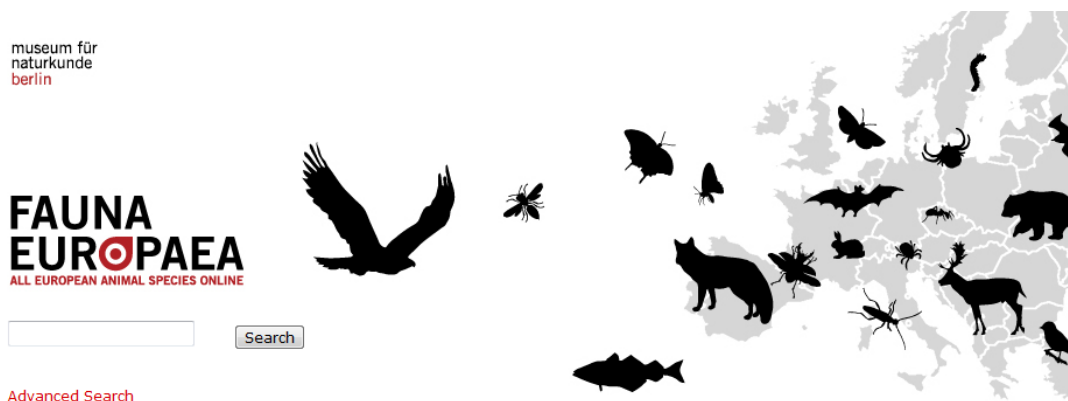

Figure 5: Search interface of the new web portal of Fauna Europaea at http://fauna-eu.org/

Apart from the technical harmonisation of Euro+Med Plantbase and Fauna Europaea, the merging pipeline for taxonomic checklists contributing to the Pan-European Species directories Infrastructures was further optimised. Two merges of Euro+Med, Fauna Europaea, European Register of Marine Species (ERMS) and Index Fungorum were performed in the course of EU BON and integrated into PESI-services and the EU BON taxonomic backbone. In this context, PESI services were tested using JMeter and performance problems were reported to the hosting institution (VLIZ) and resolved. BGBM and VLIZ will continue to manage the PESI merging processes. The next merge is planned for the beginning of 2017.

An important improvement was the development of a fully functional remote-editing solution for the EDIT Platform based on the existing Eclipse-based editor software. Editorial contributions to Fauna Europaea and Euro+Med plantbase used to be hindered by technical and usability problems of the Taxonomic Editor (Taxeditor) component of the EDIT platform. Contributing taxonomists needed to have a direct database connection which was often difficult to establish due to the local network topology and for security reasons. This problem was circumvented by providing a centralised (Ulteo based) terminal server solution on a BGBM server. The solution, however, was never satisfying due to performance and usability problems.

In order to fully solve these problems, the Taxeditor was transformed into a real remote application. The Taxeditor can now connect to the CDM-service instances directly via a HTTP-Invoker remoting service (http://docs.spring.io/spring/docs/current/spring-framework-reference/html/remoting.html#remoting-httpinvoker), which completely obsoletes the need for a direct data base connection. This change in the Taxeditor architecture is a major improvement and allows taxonomists to contribute to the EU BON taxonomic backbone in a direct and hazzle-free manner.

3.2. Data provision

In parallel to the technical implementation of the EU BON taxonomic backbone and streamlining of the merging workflow, the content of contributing taxonomic source database has been continuously and significantly improved:

For the European taxonomic database Fauna Europaea, there was also work done on the content side, however, the most important task was to secure the long-term availability of the content by transferring the system onto a new technological platform. On the content side, an update of the Fauna Europaea CDM database was conducted after optimizations and improvements in the EDIT Platform. This update included improved handling and data quality of misapplied names.

For the updates of the data, an important part was to add vernacular names ('common names') to the database of Fauna Europaea. Over 36,000 vernacular names for over 9,900 species were added so this information serves as a new feature for Fauna Europaea. This means that species can be searched also by using their common name in the advanced search interface (in case a database entry for that species exist). When showing the results for a taxon, also the vernacular names are listed as an additional source of information. This is particularly important, as often data on species exist that are connected to their common name - in order to use such information for scientific research, a clear link to an unambiguous scientific name is an important prerequisite. The vernacular names were provided by the Flanders Marine Institute in Oostende in Belgium. There common names are provided in up to 46 different languages, from Western and Northern European languages, e.g. English, Spanish and Swedish to Ukrainian, Russian and Hebrew.

Compared to the first CDM versions of Fauna Europaea, further improvements have been made in order to show the complete content of the database. In addition to names, synonyms and distributional data, also literature references and taxonomic specialists are now displayed to the datasets (see also **Appendix A**).

Further work has been conducted to update also another part of the content of FaEu, which is information on its network of collaborators, which are in this case a network of taxonomic

coordinators and experts. The expert community has been contacted and updated to retired and new specialists and this information was included in the FaEu expert database. A similar improvement has been applied to the list of national focal points, which were updated list the regional responsibilities. In addition to the existing network, further experts were contacted for the provision of dataset updates. A tentative solution for database updates was installed until thanks to the implementation of a taxonomic editor, database entries are now being updated according to higher taxon groups. In the future, the existing system of national focal points, taxonomic group coordinators and specialists will be revised to become appropriate for a broader and flexible taxonomic community with changing responsibilities.

The content of Euro+Med Plantbase, the information resource for vascular plants in the European-Mediterranean area, was considerably updated and enlarged during the last years. To start this process, a gap analysis was done to identify the most important tasks that were needed to reach full taxonomic and geographic coverage (corresponding to ca. 220 plant families) by the end of 2016. As a result of that gap analysis, the expert network of Euro+MedPlantbase was partly activated again and partly widened.

The newly acquired taxonomic expertise for many critical plants group enabled us to complete the database as planned. Taxonomic experts for their respective groups delivered completely new treatments for large and important plant groups such as Ranunculaceae, Chenopodiaceae, Ferns, Conifers and Polygonaceae. Geographical additions included data entry for the Caucasus region from relevant literature. New literature, including many floristic works, national and regional checklists both available online and published in print, have been continuously screened for relevant new taxonomic information and new distribution data.

3.3. Integration with publishing platforms

Pensoft together with a research team led by Dr Jose Fernandez-Triana from the University of Guelph, Canada, has developed an innovative workflow for integration of DNA barcoding data from Barcode of Life Data Systems (BOLD) into Biodiversity Data Journal (BDJ) and its publication in the form of a scholarly article. A recently published article used specimen records downloaded from BOLD in tabular format to be imported into a structured text developed within the Pensoft Writing Tool (PWT). These data were used to study the distribution of ten Nearctic species of braconid wasps from the Microgastrinae subfamily. Currently almost 4 million sequences (over 3.4 million of them DNA barcodes) are stored in BOLD, including coverage for more than 143K animal species, 53K plant species, and 16K fungi and other species, and this impressive storage of information is continuing to grow every day. This pilot study has explored how the unique amount of data stored on the BOLD platform can be utilised for new research purposes. Data stored on BOLD were used to uncover a significant number of new records of locality, provinces, territories and states.

> **Resource:** Fernandez-Triana, J., Penev, L., Ratnasingham, S., Smith, M., Sones, J., Telfer, A., deWaard, J., Hebert, P. (2014) Streamlining the use of BOLD specimen data to record species distributions: a case study with ten Nearctic species of Microgastrinae (Hymenoptera: Braconidae). Biodiversity Data Journal 2: e4153. doi: 10.3897/BDJ.2.e4153

Pensoft launched a new series in Biodiversity Data Journal called 'Contributions on Fauna Europaea' which embraces thematic data-papers structured in a common pattern extracted from a large database. This novel publication model will assemble in a single-issue 56 data-papers on different taxonomic groups covered by the Fauna Europaea project in the period 2000-2014 and a range of accompanying papers highlighting various aspects of this project (gap-analysis, design, taxonomic assessments, etc.). This is the first collection of data-papers of this scale. It will formalise and effectively publish the results of nearly 500 contributors building the largest European animal (taxonomic) database. The new publication model provides a reliable mechanism for citation and bibliographic indexing of large and uniformly structured databases. The publication of Fauna Europaea data papers brings a number of benefits for science, for example it stimulates experts to hand-over descriptive details on their groups, triggers new ways of community networking and participation, motivates experts to update

their data, supports a better documentation of their achievements, including issues like 'micro-publications', and increase an ownership feeling with the associated effort.

Resource:
http://bdj.pensoft.net/browse_user_collection_documents.php?collection_id=9&journal_id=1

Furthermore, Pensoft, together with Naturalis and Plazi published in BDJ an editorial describing a mechanism for correction of erroneous datasets published via the GBIF IPT through publication of data paper corrigendum. A recent paper on central European spiders included a number of taxonomic errors (Čandek et al. 2013). In a corrigendum published in this issue (Čandek et al. 2015), the authors duly correct the record. Data from the original publication have already been harvested by online resources including GBIF. To guarantee that the data is corrected not only in the scientific literature but also in GBIF, the Darwin Core Archive (DwC-A) file (which is the vehicle for distributing content to a collection of online resources; GBIF 2010, Wieczorek et al. 2012) has been updated and submitted to GBIF. The supplier (Pensoft) needs to trigger a re-indexing through the API (Application Programming Interface, a set of protocols that, in this context, is used to share data between software applications) and the content will be added to the indexing queue. Normally it takes few hours to be indexed (Markus Döring, GBIF senior software developer, pers. comm.). However, the original DwC-A file remains available for users to download from the journal web site. The original and corrected data files are clearly labeled as such and visible alongside the original publication. A link landing at the corrigendum will be added to the original publication metadata to facilitate its discoverability. In addition, the XML data file from the original article has been retained on the servers of Plazi, but the XML tags have been amended to render them no longer exposed for harvest. A modified XML document combining the original data with all corrections specified in the corrigendum (i.e., a single corrected document) has been made available as a supplementary document linked to the corrigendum, and will be uploaded to Plazi upon publication of the corrigendum. This will present the corrected data in XML form, permitting the export of treatments and data to various aggregators (Penev et al. 2012).

Recource: Miller, J., Georgiev, T., Stoev, P., Sautter, G. and Penev, L. (2015) Corrected data re-harvested: curating literature in the era of networked biodiversity informatics. Biodiversity Data Journal 3: e4552. doi: 10.3897/BDJ.3.e4552

4. Service documentation and registration

4.1. Service documentation

A consistent service documentation is generated fully automatically from the Swagger component that is integrated into the web service (see chapter 2.5) and provides both the necessary technical information for programmers using UTIS services and forms for conveniently constructing and testing queries (**Fig. 6**). In particular, the service documentation provides for each endpoint the complete list of UTIS query parameters and their description, a specification of data types and controlled vocabularies as well as the structure and content of response messages and associated HTTP Status Codes and their interpretation.

Figure 6: Form-based interface for creating and testing UTIS-service.

4.2. Registration of the PESI web services

The PESI web services have been contributed to the CSR and are registered with the following resource details:

- Resource Id: urn: geoss:csr:resource:urn:uuid:be5e6bea-e3f6-8adc-9f28-3536058e4bbd

- Resource Name: PESINameService

Resource URL:
http://geossregistries.info/geosspub/resource_details_ns.jsp?compId=urn:geoss:csr:resource:urn:uuid:be5e6bea-e3f6-8adc-9f28-3536058e4bbd

4.3. Registration of the taxonomic backbone services

The EU BON Description of Work (DoW) demands that the taxonomic backbone is registered at the EU BON Portal (EBP). This, however, can only be done after the release of the EBP which is planned for project month 51 (MS242, Updated registry and metadata catalogue, service release). In the meantime, the services have been registered directly at the GEOSS Component and Service Registry (CSR) (http://geossregistries.info). Additionally, we have registered the services in the BiodiversityCatalogue (https://www.biodiversitycatalogue.org/) to open up and advertise the services to the wider biodiversity informatics community:

- BiodiversityCatalogue: https://www.biodiversitycatalogue.org/services/79/

- GEOSS Component and Service Registry (CSR):

    o Resource Id: urn:geoss:csr:resource:urn:uuid:45fc8fb3-d81b-99ee-d0a1-7845101ba42d

    o Resource Name: EU-BON taxonomic backbone

    o ResourceURL: http://geossregistries.info/geosspub/resource_details_ns.jsp?compId=urn:geoss:csr:resource:urn:uuid:45fc8fb3-d81b-99ee-d0a1-7845101ba42d

5. Integration into the EU BON Framework

The taxonomic backbone is currently integrated into the EU BON framework in following contexts.

5.1. EU BON UTIS Taxonomic Search Widget

The *Data Search* application (http://api.eurogeoss-broker.eu/eu-bon-portal) of the EU BON Portal (EBP) enables users to search for occurrence datasets matching specific taxa. The taxonomic backbone is integrated into this process to help filtering and correcting taxon names. At the beginning of the search process, users can ask the taxonomic backbone to assist the selection of accepted taxon names as main criteria for the search process (**Fig. 7**, step 1). The EBP in turn presents the user the *EU BON UTIS Taxonomic Search Widget*. This widget offers an easy to use interface to the taxonomic backbone and the available search modes. Finally, users choose the desired name (step 2) and execute the search for data sets (step 3).
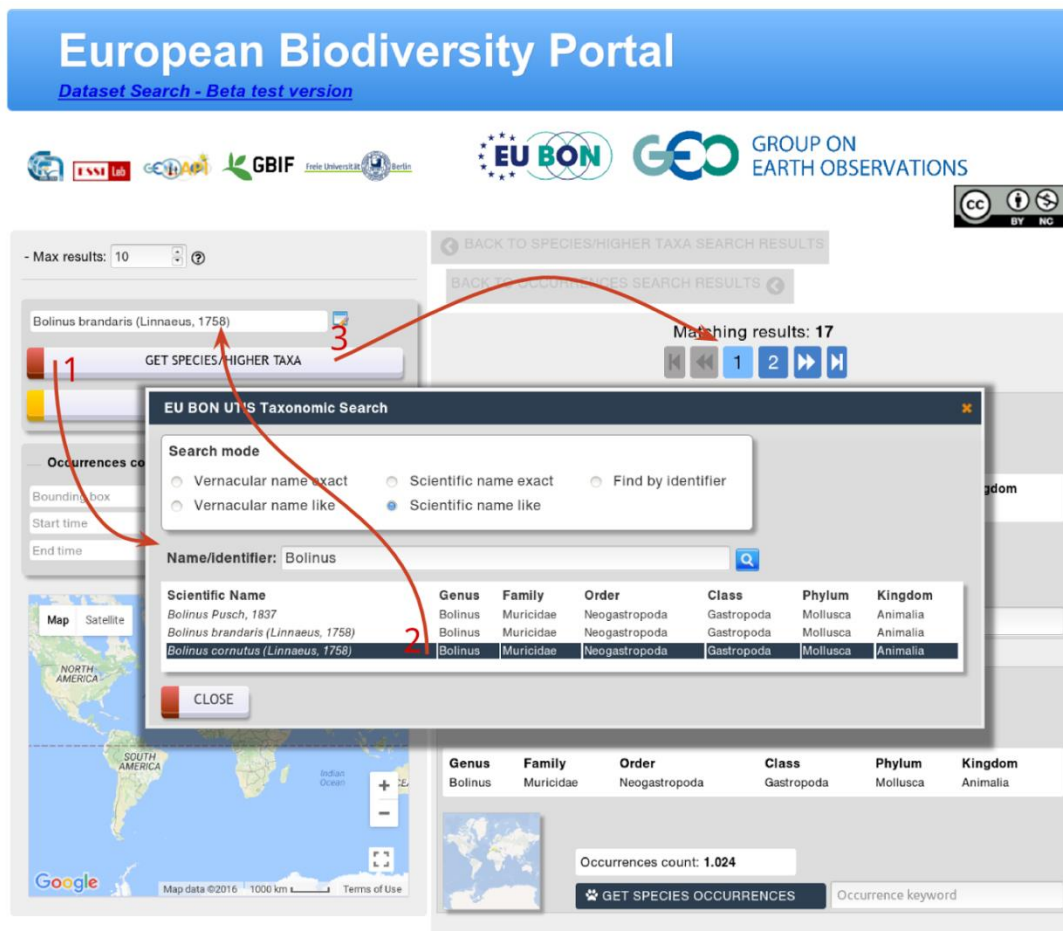

Figure 7: EU BON UTIS Taxonomic Search Widget in the EU BON Portal Data Search application

5.2. Integration into EU BON eLabs R-Studio installation via the taxize

The federated taxonomic search as provided by the taxonomic backbone is made available to scientists running analyses in the EU BON eLabs R-Studio installation via the taxize (https://github.com/ropensci/taxize) R-package. Taxize is an extension to the R-Studio, which is maintained by Scott Chamberlain (Technical lead of rOpenSci). The integration of the UTIS API into taxize was realised in a collaboration of the BGBM with Scott Camberlain.

In the EU Bon eLab, scientists can now use taxonomic data in workflows by using taxize. Data from the taxonomic backbone provided through taxize can be integrated with data from other sources and domains. Finally, R-packages provided by WP3&4 can be used to analyse the data and return results.

Taxize is also bundled in the "mirroreum" Docker image through all the ROpenSci-packages (https://ropensci.org/) bundled in there. The "mirroreum" Docker image has been provided by Markus Skyttner from NRM as component for raquamaps. Raquamaps (https://raquamaps.github.io/) is an R package providing some core functionality of the AquaMaps (http://www.aquamaps.org/) modelling, with the intent to speed it up and make it easier to automate and also provide it in an open source format, following these best practice guidelines: http://r-pkgs.had.co.nz/

6. Service testing

All tests have been implemented on the basis of the Apache JMeter test framework (https://jmeter.apache.org/). The complete test plans are available from the BGBM git repository (http://dev.e-taxonomy.eu/gitweb/jmeter-tests.git). This repository also contains additional tests for EDIT Platform driven CoL name catalogue services which have been developed in the context of the BioVel project and the Data Refinement Workflow.

6.1. Performance and stress testing of the PESI web services

The PESI web services maintained and hosted by VLIZ are the priority source of taxonomic information for EU BON. Therefore, robustness of the PESI endpoints and their capability to handle frequent requests from the EU BON backbone are of particular importance. The taxonomic backbone sends requests to the PESI web service making use of four different SOAP actions:

- search for taxa by an exact scientific name string (SOAP action: getPESIRecords, like=false)

- search for taxa which start with a scientific name string, also covers search with wildcard characters (SOAP action: getPESIRecords, like=true)

- fuzzy search for taxa by a scientific name (SOAP action: matchTaxon)

- search for taxa by an vernacular name (SOAP action: getPESIRecordsByVernacular)

Two assertions have been tested for each of the responses

- HTTP status header is 200

- Response body contains the string <item xsi:type="tns:PESIRecord"> (makes sure that records are returned)

Two fixed lists of names have been used to generate the test requests, one for scientific names and another for vernacular names whereas the tests have been run with the top 500 entries of the respective list.

- http://dev.e-taxonomy.eu/gitweb/jmeter-tests.git/blob/HEAD:/pesi/PESI-scientific-names.txt

- http://dev.e-taxonomy.eu/gitweb/jmeter-tests.git/blob/HEAD:/pesi/PESI-vernacular-names.txt

Three test scenarios have been run so far with this setup. In each scenario 500 different SOAP requests for each of the above-named actions have been sent to the service. During the first scenario, all requests were run sequentially. The other scenarios made use of multiple threads (5 and 20) to send parallel requests to the service. **Tables 6.1.1 – 6.1.3** shown in **Appendix B** document the results from the first test run.

The results of this first run clearly showed that the PESI SOAP web service is scaling well under the chosen test conditions. The error rate shown in the summary report tables stayed constant; the increased load by parallel requests is not causing any response drop outs or errors. The average response time also stays more or less constant in all three test scenarios and is about 2.2 seconds. The response time will of course directly affect the responsiveness of EU BON taxonomic backbone, since the backbone cannot respond faster than its primary source of data. These findings haven been reported to the Flanders Marine Institute (VLIZ) hosting the PESI web service layer together with recommendations for improving PESI services. Based on this, VLIZ was able to improve their service performance significantly. Another run of the JMeter performance tests after the bug fixes and performance improvements showed that the PESI web services are now about six times faster than before. **Figures 6.1.1 - 6.1.3** and **Tables 6.1.4 - 6.1.6** shown in **Appendix B** document these results in more detail.

6.2. Performance and stress testing of WoRMS

The World Register of Marine Species - WoRMS (http://www.marinespecies.org/) offers a

SOAP/WSDL based web service to access its data by different operations. We tested three relevant actions:

- exact matching of scientific name - getAphiaRecords

- fuzzy matching of scientific name - matchAphiaRecordsByName

- exact matching of vernacular name - getAphiaRecordsByVernacular

- fuzzy matching of vernacular name - not applicable

Two assertions have been tested for each of the responses

- HTTP status header is 200

- Response body contains the string '<item xsi:type="tns:AphiaRecord">' (makes sure that records are returned)

Table 6.2.1: Summary report - 1000 runs in single thread. Average response time for single request was 1.6 sec.

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getAphiaRecordsByVernacular	1000	184	88	1322	143,04	27,70%	5,2/sec	12,92	2560,5
getAphiaRecords [SOAP]	1000	208	80	2369	180,59	11,20%	4,6/sec	9,61	2123,6
matchAphiaRecordsByName	1000	4442	3669	19118	667,26	0,00%	13,5/min	0,54	2442,1
TOTAL	3000	1611	80	19118	2042,59	12,97%	40,5/min	1,56	2375,4

Table 6.2.2: Summary report - 1000 runs in 5 parallel threads. Average response time for single request was 1.9 sec.

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getAphiaRecords...	1004	266	76	2018	173,54	27,59%	11,7/sec	29,27	2560,7
getAphiaRecords ...	1004	269	77	3698	265,68	10,46%	11,4/sec	23,68	2134,2
matchAphiaRecor...	1004	5276	1068	36228	2443,51	0,70%	56,3/min	2,22	2427,4
TOTAL	3012	1937	76	36228	2756,67	12,92%	2,8/sec	6,52	2374,1

Table 6.2.3: Summary report - 1000 runs in 20 parallel threads. Average response time for single request was 5.8 sec.

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getAphiaRecords [SOAP]	1019	264	119	1416	145,82	10,01%	10,3/sec	21,68	2152,9
getAphiaRecordsByVernacular	1019	268	122	1238	139,93	27,48%	10,3/sec	25,68	2557,5
matchAphiaRecordsByName	1019	16913	3941	56591	13219,94	0,00%	1,2/sec	2,74	2438,4
TOTAL	3057	5815	119	56591	10947,73	12,50%	3,5/sec	8,04	2382,9

All tests were run using 1000 names randomly picked from WoRMS database (**tables 6.2.1 – 6.2.3**). There is a separate test file with 100 vernacular names. Even though these names were directly provided by WoRMS, we can see error percentage being more than 0. These erroneous records are those categorised as not marine. Although used web service functions provide parameter to include non-marine taxa into search, and we used it, this functionality seems not to work.

The graphs (**Figure 6.2.1a - 6.2.3b**) in **Appendix C** show the response times over time. Because the fuzzy matching took much longer to complete, it would not be scaled well on a single graph with exact matching, therefore we use separate graphs.

6.3. Continuous reliability and performance testing of the EU BON taxonomic backbone

The UTIS installation acting as taxonomic backbone for EU BON is continuously monitored and tested. This does not only assure the reliability of this important EU BON framework component, but also tests the availability of the connected underlying checklist providers. The necessary JMeter test plans have been developed by IBSAS and finally installed into the BGBM continuous integrations system at http://int.e-taxonomy.eu/jenkins/view/All/job/utis-jmeter-tests/

The test suite is run once per night. A detailed test plan result summary and performance trend analyses are generated automatically after each test run.:

- Performance Trend:
  http://int.e-taxonomy.eu/jenkins/view/All/job/utis-jmeter-tests/performance/

- Performance trend per test case:
  http://int.e-taxonomy.eu/jenkins/view/All/job/utis-jmeter-tests/performance/testsuiteReport/?performanceReportPosition=EUBON-UTIS_TestPlan-T1.jtl

- Test plan summary:
  http://int.e-taxonomy.eu/jenkins/view/All/job/utis-jmeter-tests/EUBON-UTIS_TestPlan-T1/

In case of errors an alerting email will be send to the persons in charge at the BGBM.

7. Future developments

The EU BON Unified Taxonomic Information Service provides a consistent and stable service-oriented software platform for creating and processing simultaneous queries to distributed heterogeneous taxonomic information resources. In the context of EU BON, UTIS integrates PESI, EUNIS, Natura2000 as well as the Catalogue of Life (CoL) and WORMS and acts as a INSPIRE-conform backbone for the EU BON Portal services. The BGBM will continue to host the UTIS services and to maintain and develop the UTIS software. Priority will be given to the identification and integration of additional taxonomic information resources.

For example, TreatmentBank (Plazi) is providing access to over 170,000 taxa and their taxonomic treatments and illustrations. The service is continuous extended through ongoing text and data mining of recent taxonomic publications and entire journal runs. The former is covering an estimated 25% of all the new species described annually, totalling with re-descriptions 43,000 names. The uniqueness of TreatmentBank is that additional to the names and the metadata, taxonomic treatments and the referenced illustrations are provided, including several million of facts about the taxa. A prototype for integrating taxa from TreatmentBank into UTIS is presently implemented. Taxon names together with their publication metadata are imported into the UTIS graph store in order to make these data available to the EU BON taxonomic backbone. Plazi provides a customized RDF-XML export which focuses on data which are relevant for EU BON. New publication entries in the TreatmentBank are recognized by UTIS and are imported in the underlying store in order to keep the taxonomic backbone up do date.

Being entirely based on freely available open source software, UTIS can easily be set up and configured for different purposes. It is planned to advertise the use of UTIS in different project contexts and to attract the attention of biodiversity informatics developers requiring a system for searching distributed taxonomic resources. An integration of UTIS-Services in the Terminology Services of the German federation for Biological Data (GfBio, http://www.gfbio.org/) is already planned.

8. References

- Berendsohn, W. G. 2010: Devising the EDIT Platform for Cybertaxonomy. Pp. 1-6 in: Nimis P.L. & Vignes-Lebbe R. (ed.): Tools for identifying Biodiversity: Progress and Problems. ISBN 978-88-8303-295-0.

- Čandek K, Gregorič M, Kostanjšek R, Frick H, Kropf C, Kuntner M (2013) Targeting a portion of central European spider diversity for permanent preservation. Biodiversity Data Journal 1: e980. doi: 10.3897/BDJ.1.e980

- Čandek K, Gregorič M, Kostanjšek R, Frick H, Kropf C, Kuntner M (2015) Corrigendum: Targeting a portion of central European spider diversity for permanent preservation. Biodiversity Data Journal 3: e4301. https://doi.org/10.3897/BDJ.3.e4301

- Ciardelli, P., Kelbert, P., Kohlbecker, A., Hoffmann, N., Güntsch, A., Berendsohn, W. G.: The EDIT Platform for Cybertaxonomy and the Taxonomic Workflow: Selected Components. In: Fischer, S., Maehle, E., Reischuk, R. (Hrsg.): INFORMATIK 2009, Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9. - 2.10. in Lübeck. Lecture Notes in Informatics (LNI) 154 - S. 28;625-638

- de Jong, Y., Verbeek, M., Michelsen, V., de Place Bjørn, P., Los, W., Steeman, F., Bailly, N., Basire, C., Chylarecki, P., Stloukal, E., Hagedorn, G., Wetzel, F., Glöckler, F. Kroupa, A., Korb, G., Hoffmann, A., Häuser, C., Kohlbecker, A., Müller, A., Güntsch, A., Stoev, P., Penev, L.: Fauna Europaea – all European animal species on the web. Biodiversity Data Journal 2: e4034 (17 Sep 2014), doi: 10.3897/BDJ.2.e4034

- de Jong, Y., Kouwenberg, J., Boumans, L., Hussey, C., Hyam, R., Nicolson, N., Kirk, P., Paton, A., Michel, E., Guiry, M., Boegh, P., Pedersen, H., Enghoff, H., Raab-Straube, E. von, Güntsch, A., Geoffroy, M., Müller, A., Kohlbecker, A., Berendsohn, W., Appeltans, W., Arvanitidis, C., Vanhoorne, B., Declerck, J., Vandepitte, L., Hernandez, F., Nash, R., Costello, M., Ouvrard, D., Bezard-Falgas, P., Bourgoin, T., Wetzel, F., Glöckler, F., Korb, G., Ring, C., Hagedorn, G., Häuser, C., Aktaç, N., Asan, A., Ardelean, A., Borges, P., Dhora, D., Khachatryan, H., Malicky, M., Ibrahimov, S., Tuzikov, A., De Wever, A., Moncheva, S., Spassov, N., Chobot, K., Popov, A., Boršić, I., Sfenthourakis, S., Kõljalg, U., Uotila, P., Olivier, G., Dauvin, J., Tarkhnishvili, D., Chaladze, G., Tuerkay, M., Legakis, A., Peregovits, L., Gudmundsson, G., Ólafsson, E., Lysaght, L., Galil, B., Raimondo, F., Domina, G., Stoch, F., Minelli, A., Spungis, V., Budrys, E., Olenin, S., Turpel, A., Walisch, T., Krpach, V., Gambin, M., Ungureanu, L., Karaman, G., Kleukers, R., Stur, E., Aagaard, K., Valland, N., Moen, T., Bogdanowicz, W., Tykarski, P., Węsławski, J., Kędra, M., M. de Frias Martins, A., Abreu, A., Silva, R., Medvedev, S., Ryss, A., Šimić, S., Marhold, K., Stloukal, E., Tome, D., Ramos, M., Valdés, B., Pina, F., Kullander, S., Telenius, A., Gonseth, Y., Tschudin, P., Sergeyeva, O., Vladymyrov, V., Rizun, V., Raper, C., Lear, D., Stoev, P., Penev, L., Rubio, A., Backeljau, T., Saarenmaa, H., Ulenberg, S. (2015): PESI - a taxonomic backbone for Europe. Biodiversity Data Journal 3: e5848. doi: 10.3897/BDJ.3.e5848.

- Fernandez-Triana, J., Penev, L., Ratnasingham, S., Smith, M., Sones, J., Telfer, A., deWaard, J., Hebert, P. (2014) Streamlining the use of BOLD specimen data to record species distributions: a case study with ten Nearctic species of Microgastrinae (Hymenoptera: Braconidae). Biodiversity Data Journal 2: e4153. doi: 10.3897/BDJ.2.e4153

- Hardisty, A.R., Bacall, F., Beard, N., Balcázar-Vargas, M.P., Balech, B., Barcza, Z., Bourlat, S.J., De Giovanni, R., de Jong, Y., De Leo, F., Dobor, L., Donvito, G., Fellows, D., Guerra, A.F., Ferreira, N., Fetyukova, Y., Fosso, B., Giddy, J., Goble, C., Güntsch, A., Haines, R., Ernst, V.H., Hettling, H., Hidy, D., Horváth, F., Ittzés, D., Ittzés, P., Jones, A., Kottmann, R., Kulawik, R., Leidenberger, S., Lyytikäinen-Saarenmaa, P., Mathew, C., Morrison, N., Nenadic, A., de la Hidalga, A.N., Obst, M., Oostermeijer, G., Paymal, E., Pesole, G., Pinto, S., Poigné, A., Fernandez, F.Q., Santamaria, M., Saarenmaa, H., Sipos, G., Sylla, K.H., Tähtinen, M., Vicario, S., Vos, R.A., Williams, A.R., Yilmaz, P. 2016. BioVeL: a virtual

laboratory for data analysis and modelling in biodiversity science and ecology. BMC Ecol. 2016 Oct 20;16(1):49. DOI:10.1186/s12898-016-0103-y

- INSPIRE – Data Specification on Species Distribution, Draft Technical Guidelines. http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SD_ v3.0rc3.pdf

- Mathew, C., Güntsch, A., Obst, M., Vicario, S., Haines, R., Williams, A. R., de Jong, Y. & Goble C.: A semi-automated workflow for biodiversity data retrieval, cleaning, and quality control Biodiversity. Data Journal 2: e4221 (11 Dec 2014), doi: 10.3897/BDJ.2.e4221

- Miller, J., Georgiev, T., Stoev, P., Sautter, G. and Penev, L. (2015) Corrected data re-harvested: curating literature in the era of networked biodiversity informatics. Biodiversity Data Journal 3: e4552. doi: 10.3897/BDJ.3.e4552

- Penev L, Catapano T, Agosti D, et al. Implementation of TaxPub, an NLM DTD extension for domain-specific markup in taxonomy, from the experience of a biodiversity publisher. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2012 [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2012.

- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. doi:10.1371/journal.pone.0029715

Appendix A

Example for content provided by the Fauna Europaea web platform, exemplified by the Red Fox (*Vulpes vulpes*)

> Animalia > Eumetazoa > Chordata > Vertebrata > Gnathostomata > Mammalia
> Eutheria > Carnivora > Caniformia > Canidae > Vulpes > V. vulpes

Home >

Vulpes vulpes (Linnaeus, 1758)

Back to search result

Experts   Synonymy   Citation   Distribution   Common Name   Bibliography

Experts

Expert	Role
Dr. Igor Zagorodniuk	associate specialist
Prof. Dr. Wieslaw Bogdanowicz	taxonomic specialist
Prof. Dr. Wieslaw Bogdanowicz	group coordinator

Synonymy

Vulpes vulpes (Linnaeus, 1758)
  ≡ *Canis vulpes* Linnaeus, 1758

Distribution

Distribution	export

Region	Status
Albania	present
Austria	present
Belarus	present
Belgium	present
Bosnia and Herzegovina	present
Britain I.	present
Bulgaria	present
Central European Russia	present
Corsica	present
Croatia	present

Showing 1 to 10 of 74 entries   ◄ Previous   Next ►

Common Name

Albanian: dhelpa, dhelpra, skilja; **Bokmål, Norwegian; Norwegian Bokmål**: rev, rødrev; **Danish**: Ræv; **Dutch, Middle (ca.1050-1350)**: Vos; **English (North America)**: Red Fox; **German**: Fuchs, Rotfuchs; **Greek, Modern (1453-)**: Αλεπού; **Hebrew**: שועל מצוי; **Lithuanian**: rudoji lapė; **Norwegian Nynorsk; Nynorsk, Norwegian**: raudrev, rev; **Romanian**: Vulpea; **Russian**: лисица обыкновенная, лисица рыжая; **Slovenian**: lisica; **Swedish**: rödräv; **Turkish**: Kizil tilki, Kızıl tilki, Tilki; **Ukrainian**: Лис звичайний; **Welsh**: cadno

Bibliography

Mitchell-Jones A. J., Amori G., Bogdanowicz W., Krystufek B., Reijnders P.J.H., Spitzenberger F., Stubbe M., Thissen J.B.M., Vohralik V. & J. Zima - The atlas of European Mammals in Academic Press, London, 484 pp. 1999

Appendix B

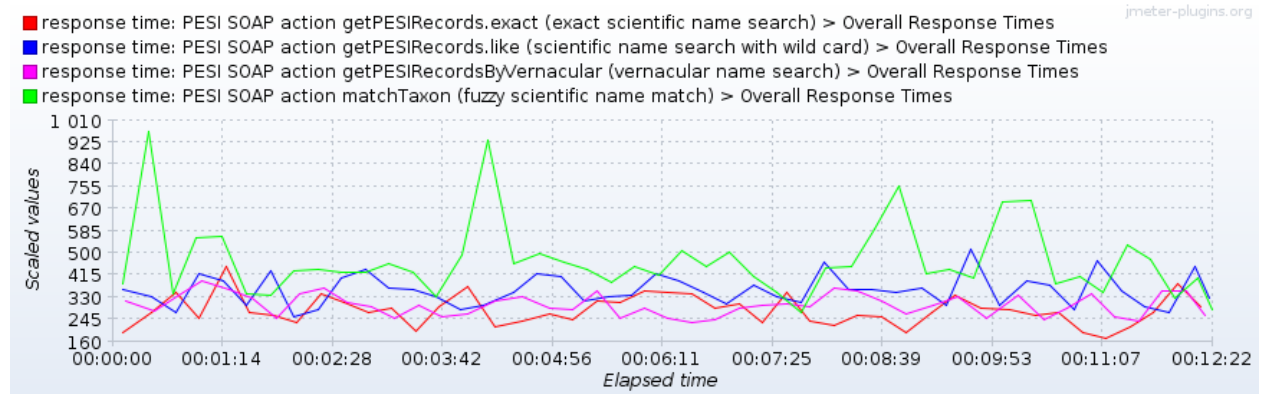Figures and Tables from Chapter 6.1: Service testing results

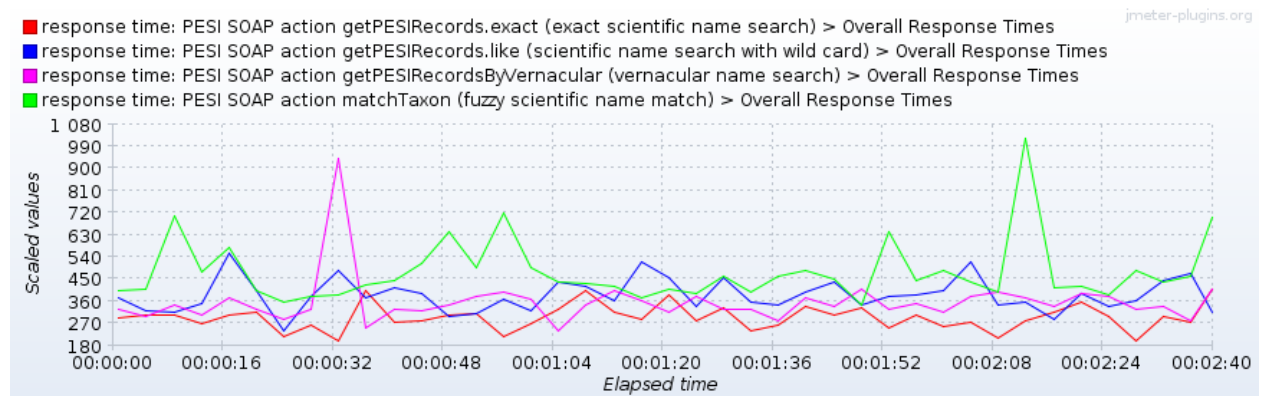**Figure 6.1.1:** Response time graph 1 thread 500 requests

Table 6.1.4: Summary report 1 thread 500 requests

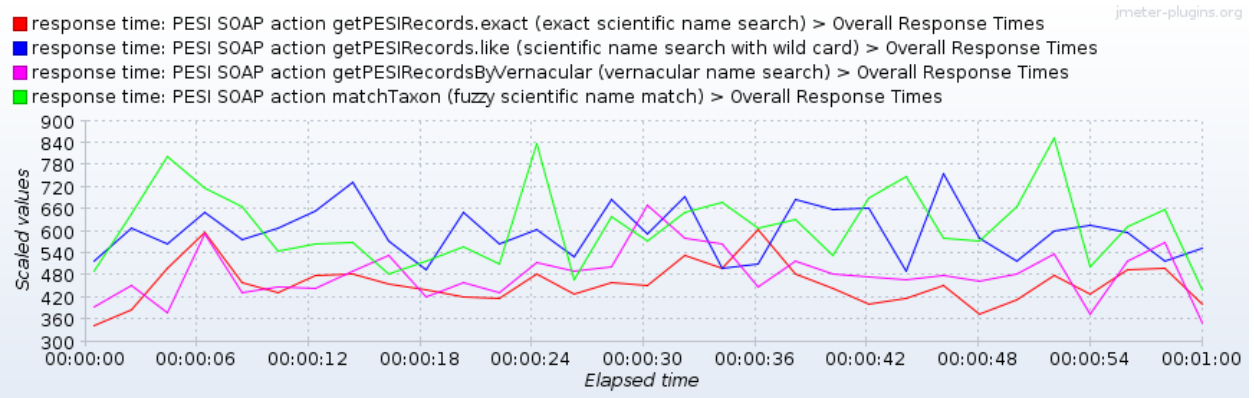| Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getPESIRecords.exact[SOAP]	500	276	120	1713	161.20	2.80%	40.4/min	1.30	1976.8
getPESIRecords.like [SOAP]	500	347	129	2118	197.43	0.00%	40.4/min	10.07	15310.1
matchTaxon [SOAP]	500	462	200	4940	406.00	3.80%	40.4/min	1.27	1934.4
getPESIRecordsByVernacular [SOAP]	500	301	120	828	140.73	0.80%	40.4/min	1.63	2471.5
TOTAL	2000	347	120	4940	259.75	1.85%	2.7/sec	14.25	5423.2

Figure 6.1.5: Response time graph 5 threads 500 requests

Table 6.1.5: Summary report 5 threads 500 requests

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getPESIRecords.exact[SOAP]	504	291	133	930	156.30	2.78%	3.2/sec	6.10	1978.0
getPESIRecords.like [SOAP]	504	385	150	1577	233.84	0.00%	3.2/sec	46.91	15227.1
matchTaxon [SOAP]	504	475	177	5617	418.30	3.77%	3.2/sec	5.97	1935.9
getPESIRecordsByVernacular [SOAP]	504	353	123	3346	219.57	0.79%	3.2/sec	7.62	2471.2
TOTAL	2016	376	123	5617	282.80	1.84%	12.6/sec	66.22	5403.1

Figure 6.1.3: Response time graph 20 threads 500 requests

Table 6.1.6: Summary report 20 threads 500 requests

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	KB/sec	Avg. Bytes
getPESIRecords.exact[SOAP]	519	459	163	3953	242.05	2.70%	8.6/sec	16.78	1987.3
getPESIRecords.like [SOAP]	519	603	176	3377	330.63	0.00%	8.6/sec	128.23	15257.2
matchTaxon [SOAP]	519	620	226	5860	453.63	3.66%	8.6/sec	16.34	1946.0
getPESIRecordsByVernacular [SOAP]	519	486	143	1819	200.02	0.77%	8.6/sec	20.87	2472.5
TOTAL	2076	542	143	5860	329.24	1.78%	33.7/sec	178.40	5415.8

Appendix C

Figures from Chapter 6.2: Performance and stress testing of WoRMS
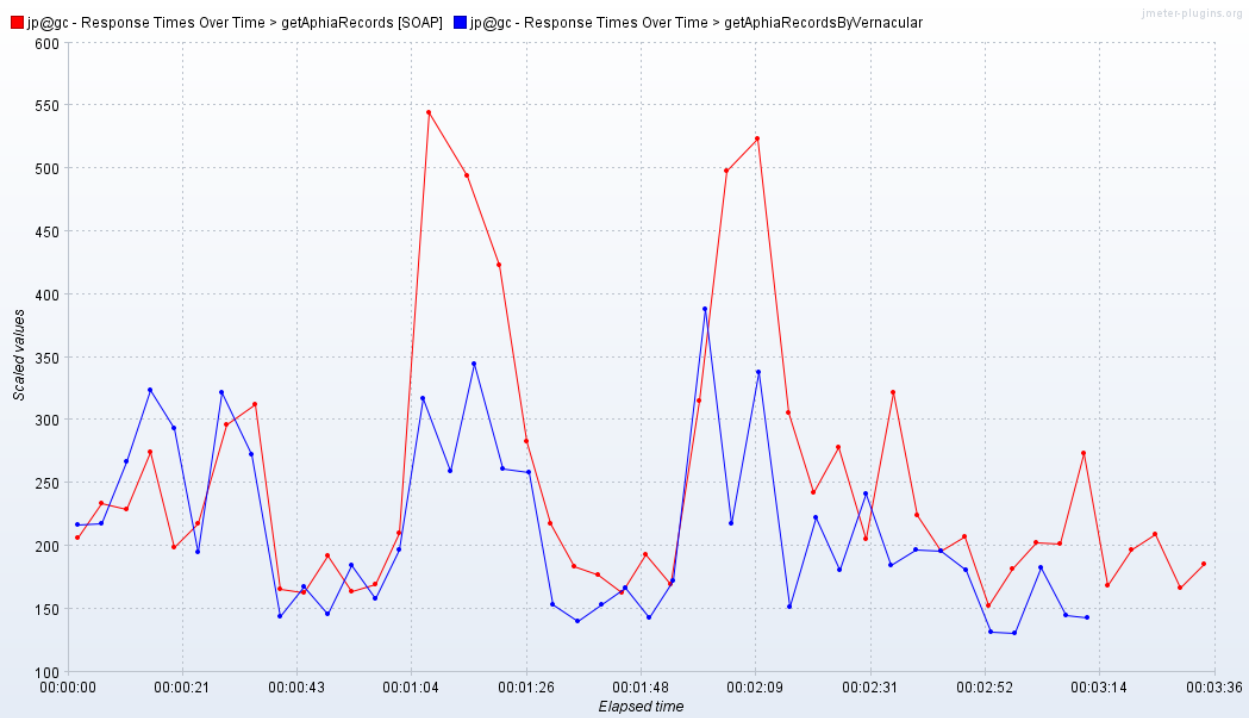

Figure 6.2.1a: Response time graph - scientific and vernacular exact match, 1 thread, 1000 requests
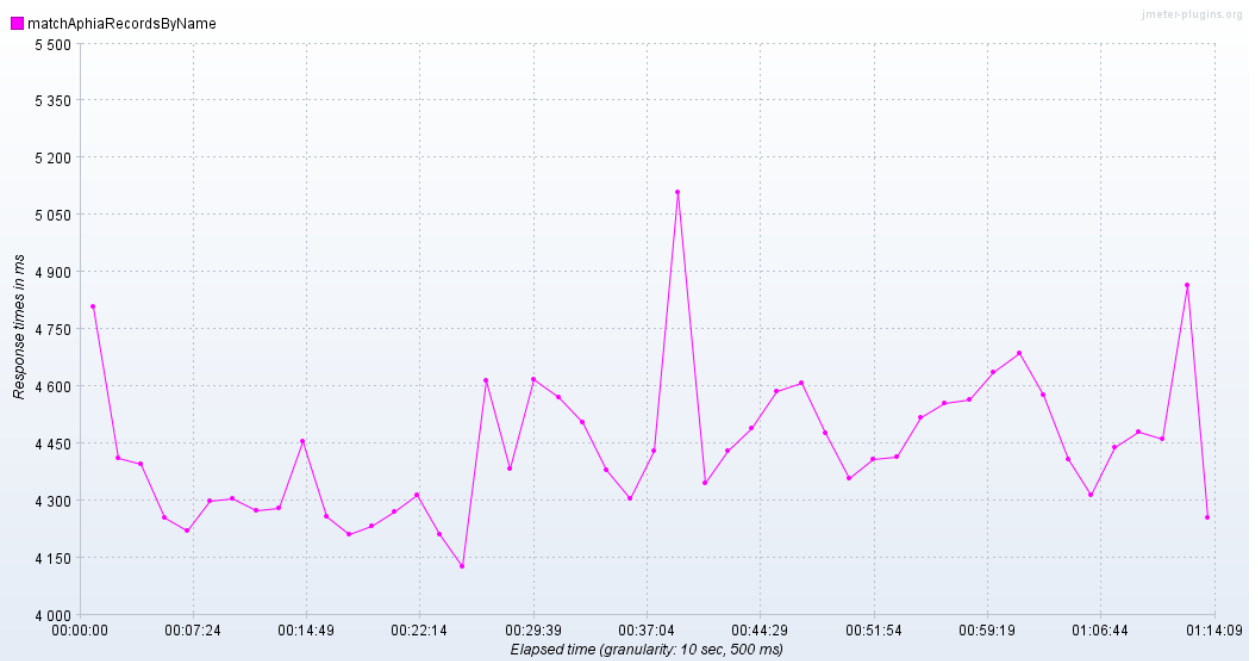

Figure. 6.2.1b: Response time graph - scientific fuzzy match, 1 thread, 1000 requests
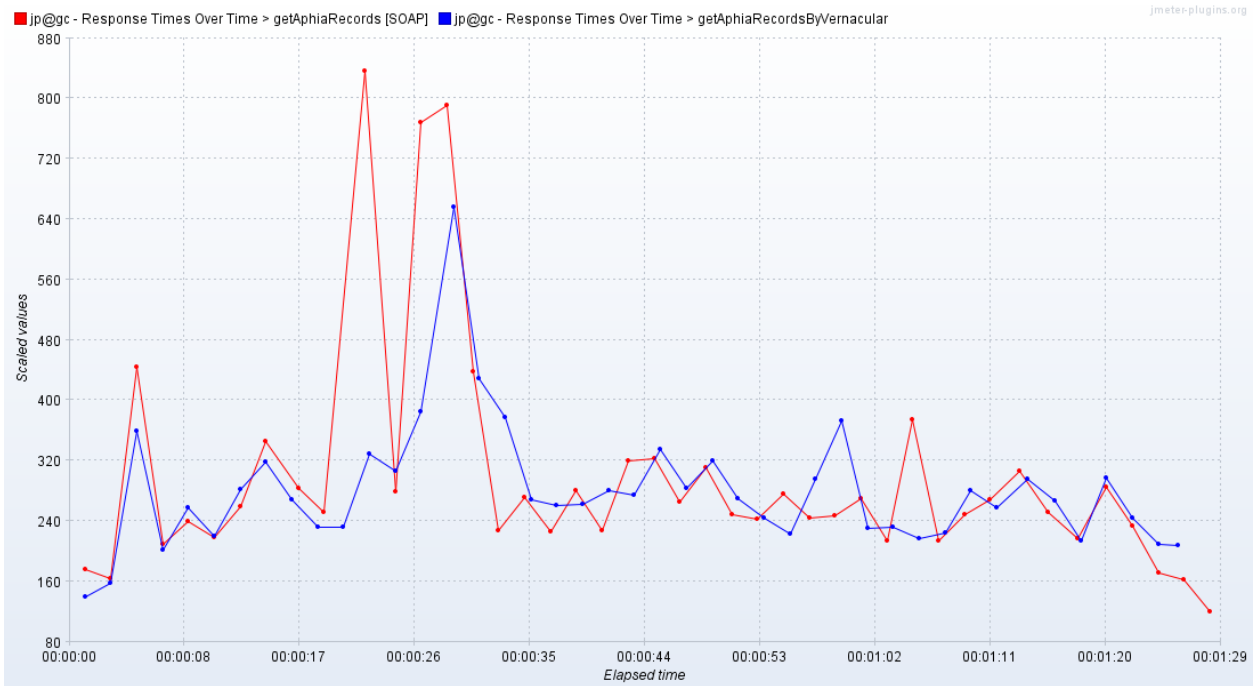
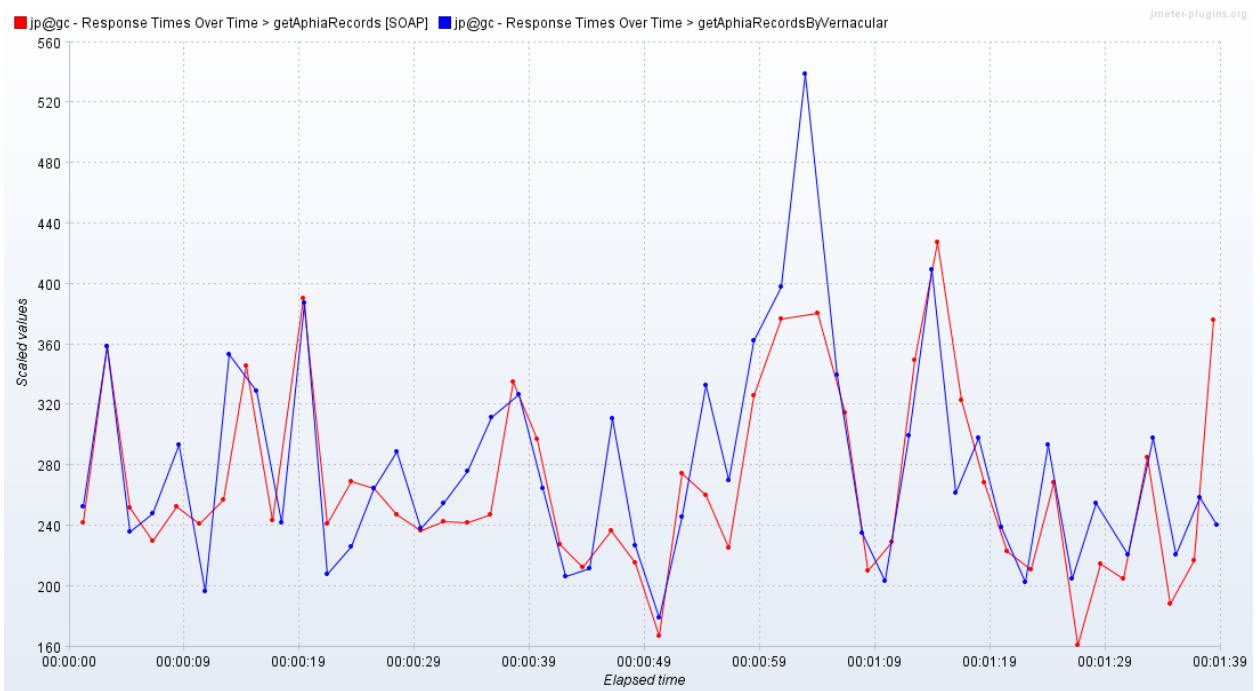**Figure 6.2.2a:** Response time graph - scientific fuzzy match, 5 threads, 1000 requests

Figure 6.2.3a: Response time graph - scientific and vernacular exact match, 20 threads, 1000 requests
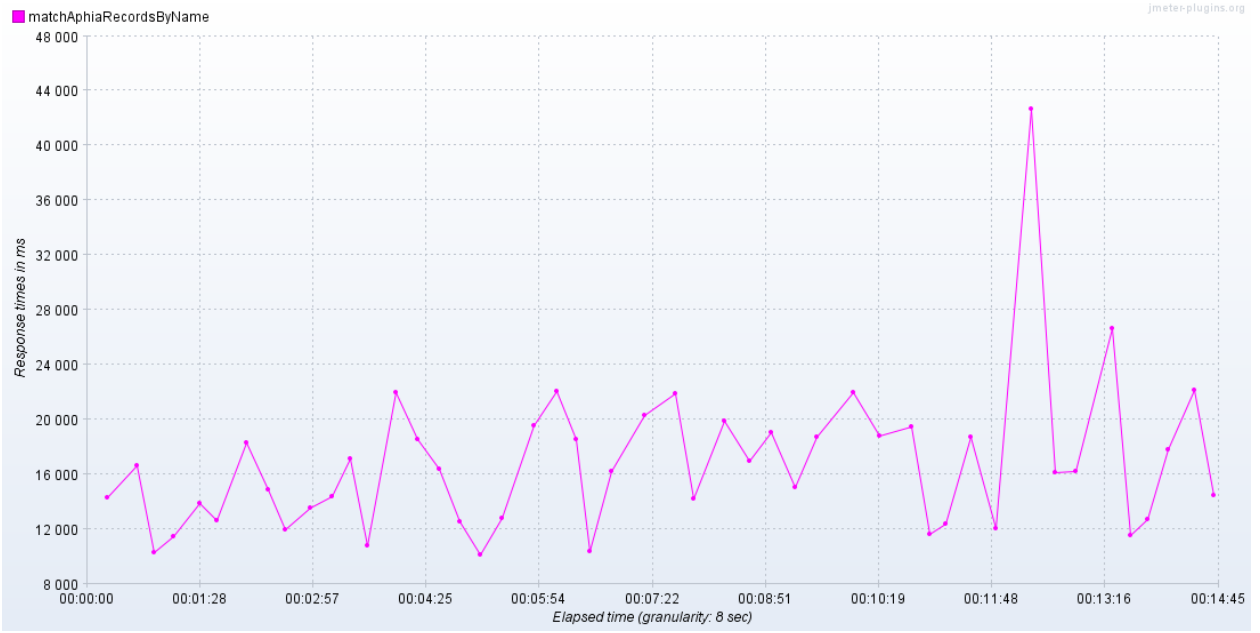
Figure 6.2.3b: Response time graph - scientific fuzzy match, 20 threads, 1000 requests