eXist Solutions
Magdalena Turska
magdalena@existsolutions.com

LOD API for Lexicon of Greek Personal Names - Final report

February 2nd 2019

OVERVIEW

I.Sicily is an EpiDoc corpus of the inscriptions of Sicily (http://sicily.classics.ox.ac.uk). One of the major priorities of the project once EpiDoc transcriptions are ready is getting names and people properly annotated using citable external authority URIs. Lexicon of Greek Personal Names (http://www.lgpn.ox.ac.uk) is an obvious choice for ancient personography and its spin-off linguistic project (LGPN-Ling) could provide corresponding onomastic resource. Important prerequisite for using LGPN and LGPN-Ling in that way by I.Sicily but also other projects is exposing Linked Open Data API that allows for linking with LGPN resources via stable URIs and querying LGPN databases.

LGPN

The Lexicon of Greek Personal Names was established in 1972 and the use of computer technology to facilitate the data collation process and the eventual publication was an obvious choice, even as far back as 1975. Since then several generations of computers have come and gone but until 2016 the LGPN records have been created in a workflow established around two decades earlier, based on manually composed text files which were later imported into relational database Ingres (cf http://www.lgpn.ox.ac.uk/online/computerization/index.html).

In 2010 the first attempt to convert the database into an XML resource compliant with the Text Encoding Initiative (TEI) guidelines has been made for the purposes of the Database search project which exposed the majority of the information gathered in the relational database to the public (http://clas-lgpn2.classics.ox.ac.uk). Resulting XML documents were stored in eXist-db, a native XML database which also served as data publication and querying platform.

Since 2016 LGPN underwent another major change, which moves also the editing and ultimately print publishing workflows completely into TEI/XML domain. The forthcoming volume VI of LGPN covering Palestine, Syria and Trans-Euphratic Regions is the the first one edited in the new graphic editing environment and will be typeset with the XML to LaTeX transformation based on the records in XML database eXist-db. Concurrently, the LGPN-Ling, a new accompanying linguistic resource has been added, which employed the same principles of using eXist-db and TEI from its very beginning and uses similar form-based editing environment to curate the data.

Nevertheless, at the time of writing, the only public interface is only serving content for the voluems I to V and in a form which doesn't facilitate its use as Linked Open Data resource. Currently LGPN offers existing persistent URIs in the format 1) **http://www.lgpn.ox.ac.uk/name/{NAME}** and 2) **http://www.lgpn.ox.ac.uk/id/{ID}**
Main problem for using LGPN as external authority is it dereferences to usual HTML web pages with content that is meant for human interaction, not machine processing. Furthermore, type 1 URIs dereference to a list of person records,

not an onomastic resource. For backwards compatibility, these older URIs need to be preserved but a new scheme designed in parallel to expose parts of the data for the purpose of querying and referencing LGPN and LGPN-Ling resources from external systems.

LGPN API Project

Development of the 1.0.0 version of the LGPN API (Application Programming Interface) to facilitate external referencing and use of LGPN resources has been the subject of LGPN LOD project funded by Oxford John Fell grant awarded to I.Sicily project and supervised by Jonathan Prag. Implementation was carried out by eXist Solutions Gmbh with Magdalena Turska, lead developer of LGPN project, acting as a project manager and lead developer also for LGPN API.

Tasks required for launching the basic API service have been initially estimated as follows:

- Scoping: 8
- Implementation: 32
- Documentation: 10
- Contingency & project management: 10

adding up to a total of 60 hours of development work.

In hindsight, the scoping part has been hugely underestimated, not taking into consideration difficulties of scheduling and requesting feedback from potential future users and time required for communication. To a lesser extent also implementation has been estimated with certain dose of optimism, assuming more limited requirements than were eventually implemented. Moreover, in the course of development work we have discovered a bug in the eXist database affecting negatively the API implementation and had to reconsider some decisions made for LGPN project itself.

After completion of the LGPN API project is handed over to LGPN for further development and maintenance.

Goals

Main goals for the initial version of the API were:

- Design and implement stable URI system for person records from LGPN and name records from LGPN-Ling databases.
- Implement an API for dereferencing the URIs into JSON that could be consumed by external applications.
- Design and implement an API for querying LGPN and LGPN-Ling in order to find matching records following basic criteria (eg. name, period or place) useful for the purposes of annotating references to people and names in I.Sicily texts with canonical identifiers.
- Launch the LOD services.
- Publish the sources and documentation in an open repository.

Implementation

All LGPN resources are stored as data or application packages within eXist database instance hosted on the server of Faculty of Classics. Server maintenance remains the responsibility of Faculty of Classics IT department, while eXist Solutions in practice maintains the eXist-db instance during the active development period.

Service address

During the implementation phase the API service is accessible at
http://clas-lgpn4.classics.ox.ac.uk:8080/exist/apps/lgpn-api

LGPN Packages

Core LGPN resources are grouped in 4 packages:

- LGPN - application modules for the editing environment for LGPN, allowing to browse, search and edit records related to persons
- LGPN-data - data collection where person records and other related resources (eg place or bibliographical records) are stored
- LGPN-ling - application modules for the editing environment for LGPN-Ling, allowing to browse, search and edit name records
- LGPN-ling-data - data collection where name records and other resources relevant for linguistic part (eg meanings of morphemes) are stored

LGPN API project adds one more package to this list, likewise running on the same eXist-db instance and subject to the same backup and maintenance regime

- LGPN-api (linked open data) - application exposing the API for linking and machine querying of LGPN resources

Repository

All LGPN-API code is hosted in Git repository available at https://gitlab.existsolutions.com/lgpn/lod with an issue tracker used for discussion during the implementation stage and reporting bugs and feature requests in the future.

Initial work on the API design has been documented via Swagger tool and is available at https://app.swaggerhub.com/apis/lgpn7/lgpn-lod/1.0.0 as well as in the gitlab repository.

LGPN API project can be installed on any eXist-db instance running version 4.5.0 or higher but requires all other lgpn packages.

API Overview

Main concepts in LGPN database are people (living in broadly understood ancient Greek world) who are usually known by name (or names), related to their contemporaries by family or social relationships, attached to physical places, and for whom there may exist a body of references in scholarly literature. Elaborate information on names is stored in the LGPN-Ling and only referenced by LGPN part. These concepts are all represented as TEI-conformant XML documents and elements with unique identifiers, stored in eXist database and interlinked according to mechanisms allowed by TEI Guidelines, which employ these identifiers.

From Linked Open Data perspective, LGPN API exposes these two types of resources: **persons** and **names**.

Person and name reference URIs

Persons and names stored in LGPN or LGPN-Ling databases
have their unique identifiers and can be annotated in external editorial projects using stable, citable external authority URIs. For this purpose two API

endpoints are established:

- /lgpn-api/names/{ID} eg /lgpn-api/names/Arteas
- /lgpn-api/persons/{ID} eg /lgpn-api/persons/V6-10001

Both expose details of a single name or person record.

Person and name queries

Another important use case for the API is enabling machine queries, which return a list of matching name or person records. For this purpose two further API endpoints are provided:

- /lgpn-api/persons?{QUERY}
- /lgpn-api/names?{QUERY}

These always expose a list of records (even if the list is empty or contains just single entry). Examples of queries are:

- List all persons named Αβδαλλας: /lgpn-api/persons?name=Αβδαλλας
- List all persons with names start with Αβδ: /lgpn-api/persons?name=Αβδ*
- List all persons named Αβδαλλας from Bostra (by place-id): /lgpn-api/persons?name=Αβδαλλας&place-id=P96234789-1fb5-4246-ba6b-a1f948a1bac6
- List all persons named Αβδαλλας from Volume 6 : /lgpn-api/persons?name=Αβδαλλας&volume=V6
- List all persons named Αβδαλλας from Wad* (by place name): /lgpn-api/persons?name=Αβδαλλας&place=Wad*
- List all persons named Αβδαλλας from Volume 6, dated before 475: /lgpn-api/persons?name=Αβδαλλας&volume=V6&end-date=475

Query parameters

v.1.0.0 exposes following person query parameters:

- name
- place
- place-id
- reference
- volume
- start-date
- end-date

Queries are case and diacritics insensitive. Wildcards are allowed in non-leading position.

Formats

Format of the response is dependent on the request headers and could be HTML, XML or JSON for the v.1.0.0 of the API.

Example curl request for JSON output:

curl --header "Accept: application/json" http://localhost:8080/exist/restxq/lgpn-api/persons\?name=Αβδαλλας

Documentation

Current project documentation including swagger.yaml file conforming to the OpenAPI specification is part of the Git repository available at https://gitlab.existsolutions.com/lgpn/lod