VGI Discovery with Linked Data

General introduction

The Energic COST action aims to demonstrate the potential of Volunteered Geographic Information such that data, generated by a wide range of participants ranging from authoritative bodies across scientists to individual citizens, can be used to provide information relevant to scientific, societal and policy in a European context. The objective of activities within the Energic Datathon is to allow anyone to participate in our activities, and demonstrate the potential of transformations of data to knowledge. To lower the barrier to entry for participants each Datathon task provides:

- a description of the underlying motivation for the task

- sets out an initial set of questions that might be explored through the data

- gives access to some prepared data and suggests potential additional sources, and

- suggests potential tools and methods which might be used in the task.

However, these guidelines are only intended to give a starting point to the activity, and we encourage you to be as creative as possible. Entries to the datathon will be judged by a panel of Energic members, and the best will be invited to present their results at the Energic closing meeting in London.

Specific introduction including overarching objectives

This challenge deals with the publication and discovery of VGI resources in the COST ENERGIC project repository. In an on-going effort within ENERGIC to model the domain of VGI, we have started to create a structured set of relevant concepts and relationships between them in an ontology. The combination of these concepts, such as 'Contributor', 'Information type' and 'Data quality', can be used to create descriptions of volunteered data, VGI systems (such as OSM), VGI analysis methods and detailed VGI publication references. In its turn, with these descriptions, we will be able to find relevant VGI entities.

The ontology can be found here:
http://webprotege.stanford.edu/#Edit:projectId=c09a73fb-6089-40d2-9224-678f9bfd63ce
As the concepts in the ontology are related to other domains, the descriptions (which will be represented by Linked Data) need to refer to other domain ontologies as well. The VGI entities described in this way are supposed to land in the *ENERGIC repository*, which is meant to be utilized by anyone who wants to use VGI.

Available datasets and additional potential data sources

ENERGIC VGI Ontology
GeoNames
OSM RDF dataset on http://linkedgeodata.org/ ;
http://downloads.linkedgeodata.org/releases/2015-11-02/
DBPedia

Authoritative data

Linked Data sources at https://datahub.io/

Questions to be asked of the data

Concerning the publication and discovery of VGI resources in the ENERGIC repository we distinguish the following sub challenges 1A (basic) and 1B and 2 (advanced):

1. Expanding and testing ENERGIC ontology
a. Create as many as possible relevant VGI system descriptions and detailed VGI publication references based on the ENERGIC ontology and publish these descriptions online (e.g., at the ENERGIC website or repository).
b. Extra challenge (not compulsory): As evidence of the usefulness of this, create and execute a few relevant (Geo)SPARQL queries.

2. Align the ENERGIC ontology with other ontologies, such as Geonames, OSM and DBPedia, and create a triple store that facilitates VGI system descriptions and detailed VGI publication references. As evidence of the usefulness of this, execute a few relevant (Geo)SPARQL queries.

Possible methods and tools

Ad 1a. You can create the VGI system descriptions and detailed VGI publication references with yEd graph editor, see the document below:
https://www.dropbox.com/s/c75hbcsau0fu66a/VGI%20descriptions%20as%20graphs.pdf?dl=0

Ad 1b. (Geo)SPARQL queries can be created with a Linked Data / Ontology explorer (e.g., SPEX) and/or SPARQL query tool, e.g., provided within a triplestore tool such as Parliament.

Ad 2. Alignment can be done within a triplestore, such as Parliament. (Geo)SPARQL queries can be created with a Linked Data / Ontology explorer (e.g., SPEX) and/or SPARQL query tool, e.g., provided within the triplestore tool.

More details on data sources, methods, tools, and tips & tricks can be found here:
https://docs.google.com/document/d/1C3B5d7SXjXDQuS6SvBbMCZv6lyZPKE7ywCHSiYbYJcQ/edit ?usp=sharing

Reporting your results

The expected outputs include:
Ad 1a. RDF data from yEd, describing VGI entities. This RDF is supposed to be used in a triplestore for querying after the Datathon as part of the ENERGIC repository.

Ad 1b and 2: (Geo)SPARQL queries and results (within SPARQL editor, such as Parliament or SPEX). You should prepare a report of your results which explains briefly:

▪ The data and methods you used (and provides links to these such that your work can be reproduced)

▪ Interprets your results, concentrating on what you learnt through the datathon and linking to the questions set out above

▪ Emphasises challenges in carrying out the datathon

- Illustrates the originality and novelty of your approach

- References any external sources you used to help you complete the task

- A 2 minute video pitch presenting your report

Your report should be prepared as a self-contained set of HTML pages which can be accessed by the judges and uploaded to the Energic website after the challenge. All content on the website should be licensed CC-BY-SA (where you use data sources covered by other licenses you should provide tools and access to these and make clear any limitations in their use).

Judging criteria

A panel of Energic members will judge the quality of entries to the Datathon and select the best examples for presentation at the final Energic meeting in London. The following criteria will be used in judging entries:

- Overall quality of the entry to the datathon

- Originality and novelty of the approach taken

- Quality of the description of the data and tools used, especially with respect to reproducibility

- Soundness of the approach taken

- Potential scientific, societal and policy impacts of the results

- Quality and engagement in the video pitch

Information for organisers

Target group:
Basic (objective 1a): BSc, MSc and PhD students

Advanced (objectives 1b and 2): Linked Data professionals

The Energic Datathon is open to anyone. However, it will be most fun, and probably also most productive for small groups (typically 3-4 people). The tasks have been designed such that they can be carried out by groups with different levels of skills, ranging from basic spatial analysis using standard GIS to creation of more complex workflows using programming skills. We estimate that typical time investment for a Datathon task should be of the order of 12 hours - however, it is of course up to participants how much or how little time you invest. The only hard rule is our deadline for submissions of **31.07.2016**.

There is no need to register for the Datathon, just submit your report to ross.purves@geo.uzh.ch by the deadline. However, we'd like to know that you're taking part, so feel free to drop us a mail telling us who you are, how many of you are participating in which challenges, and whether or not others are welcome to join you. Please Tweet about the event using the HashTags #Energic and #Datathon. Some useful information about running datathons events can be found at:

- https://hackathon.guide/

- http://guide.mlh.io/

Contact information

For this particular datathon:

r.l.g.lemmens@utwente.nl
Gilles.Falquet@unige.ch

For the Energic Datathon challenge:

ross.purves@geo.uzh.ch
f.o.ostermann@utwente.nl
r.l.g.lemmens@utwente.nl