



POLITÉCNICA



Volunteering assistance to online geocoding services through a distributed knowledge solution

José Pablo Gómez-Barrón Sierra, Miguel A. Manso Callejo, Ramón Alcarria Garrido

June 9th, 2015
Workshop at 18th AGILE Conference on
Geographic Information Science
Lisbon, Portugal

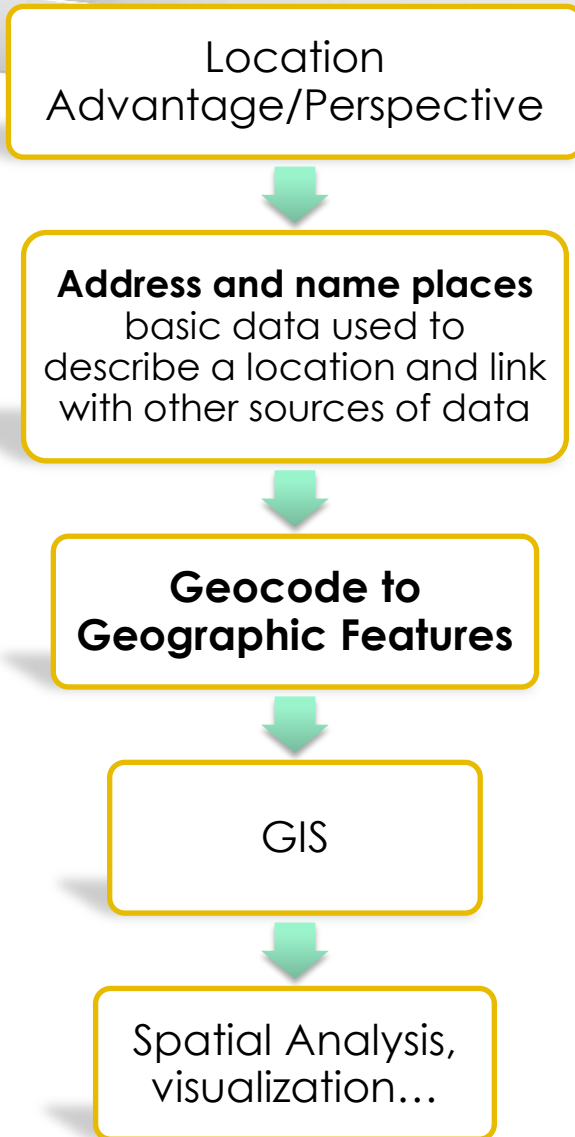
**RICH-VGI: enRICHment of volunteered geographic information (VGI):
Techniques, practices and current state of knowledge**



Roadmap

- Introduction
- Platform Development
- Platform-User Interaction
- Data Management
- Demo time!
- Conclusions and Future Work

Introduction



- Geocoding process assigns a geographic coordinate pair to a particular place by comparing its descriptive location elements with those in a reference database [1, 2, 3].
 - Searches in reference data
 - Assigns scores to potential candidates
 - Filters based on minimum match score
 - Delivers the best match
- Main components: input, output, processing algorithm, reference database [2].

Online Geocoding Services



- **Digital mapping services** and **open-data mapping platforms** are constantly **improving their web service technologies and APIs** (Application Programming Interface) to tackle geocoding complexity and to make it transparent to end users.
- **Immediate output**
- **High match rates**
- **Basic user knowledge**
- **Low or no cost**

Online Geocoding Services



- **Output quality values**, like the calculation method used or entity type that was obtained.
- **Guidelines** to:
 - Understand the output,
 - **Data quality assessment**,
 - Complementing data quality common metrics: completeness, positional accuracy, repeatability [3] and similarity [4].

Online Geocoding Services

NOT TO



- Service providers are responsible for:
 - maintaining the reference matching data,
 - improving predefined algorithms
- **User cannot customize the geocoder settings or rules** to manipulate the response according to their needs or specific input data.
- **End users must analyse the quality of the geocoded results** for each service to choose the best option to their applications [4] and data characteristics.

Online Geocoding Services



- Sometimes, **low quality results**, mainly **with ambiguous input**.

- With **ambiguous or unstructured** data addresses **inputs**, **variety of online geocoding services** response can be an advantage.

SUPER



The Crowd-Geocoding Platform

MORE SUPER



- Proposes to combine and analyse different geocoders outputs as options for incomplete or imprecise data and obtain a better geocoded result.
 - Is based on crowdsourcing geospatial data [5, 6] and Volunteered Geographic Information [7] approaches.
 - Relies on open cognitive abilities and local knowledge of collaborators.
 - Facilitates user online assistance to analyse quality and geographic precision of geocoding results,
 - Help to identify and saves user selected best candidate or manually geocoded address.
 - Enables distributed users online participation on geocoding tasks.
 - Using the platform database results, comparative evaluations can be made.

Platform Development

What we want?

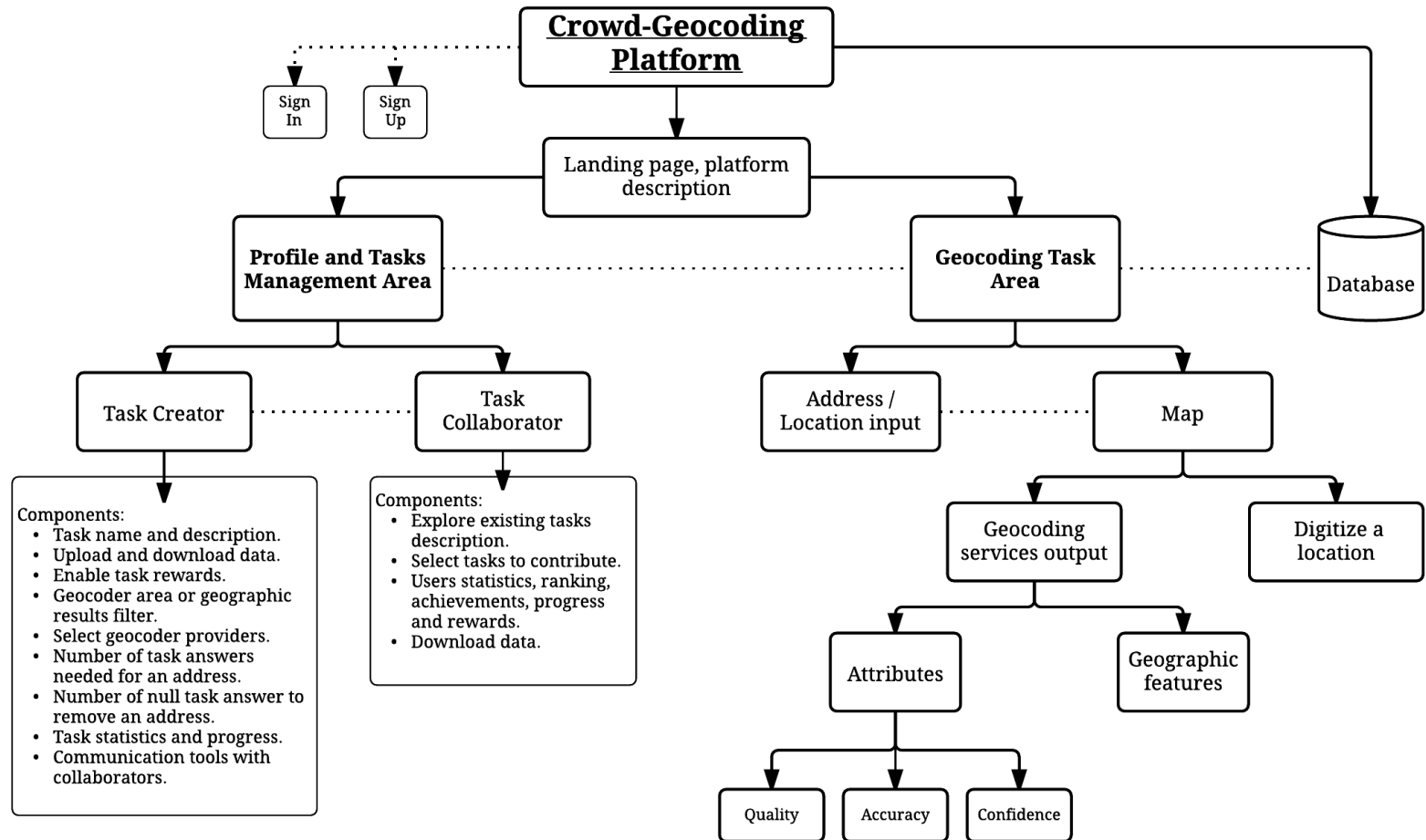
- Guarantee the quality of a geographic final layer, requiring human interaction to check and analyze the results.

What will we need?

- A web mapping client to browse and explore geodata.
- Make comparisons of output locations and quality attributes to choose best option.
- Save user-selected service, and each geocoder coordinates and quality info.
- Have a data model to further quality evaluation analysis.

Platform Development

What are the main components?

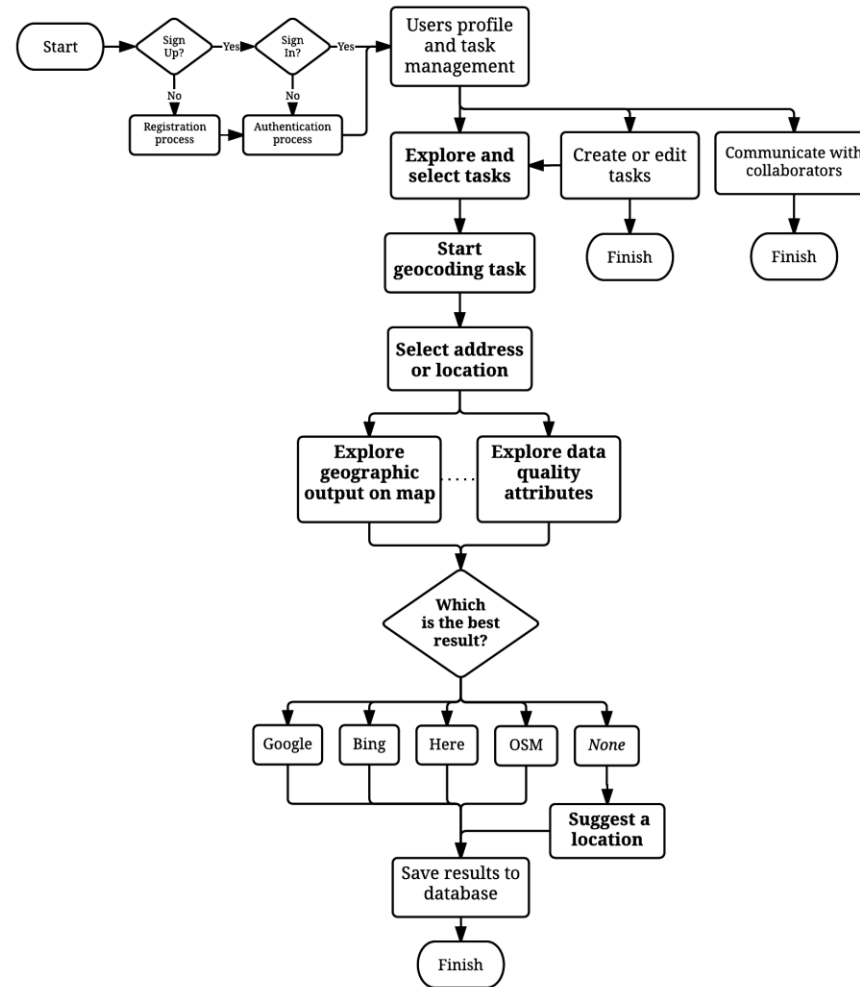


Platform Development

What are the used technologies?

Technologie		Server-side	Client-side
Python/Django	Clean design and organized application structure	x	
	Python models and defined functions to process POST and AJAX		
	Template language to easy pass output values and context data		
	Security integration: authentication, registration and account management		
	Django models: gives automatically generated database-access API		
Python Geocoder API	Python wrapper client: supports most popular geocoding web services	x	
	Converts different responses into a consistent and unified JSON response		
	Google, Bing, Here & OSM (Nominatim) providers are enabled in this platform		
PostgreSQL	Django models fields and data behaviour are related to a single database	x	
CartoDB	Geospatial database to manage the geographic data	x	
	Python client API for CartoDB SQL to insert geographic results	x	
	SQL JavaScript API to select random addresses based on user ID		x
JavaScript/Jquery	Easier web development		x
Bootstrap	Responsive web development		x
Leaflet.js / Mapbox.js	Interactive web maps.		x

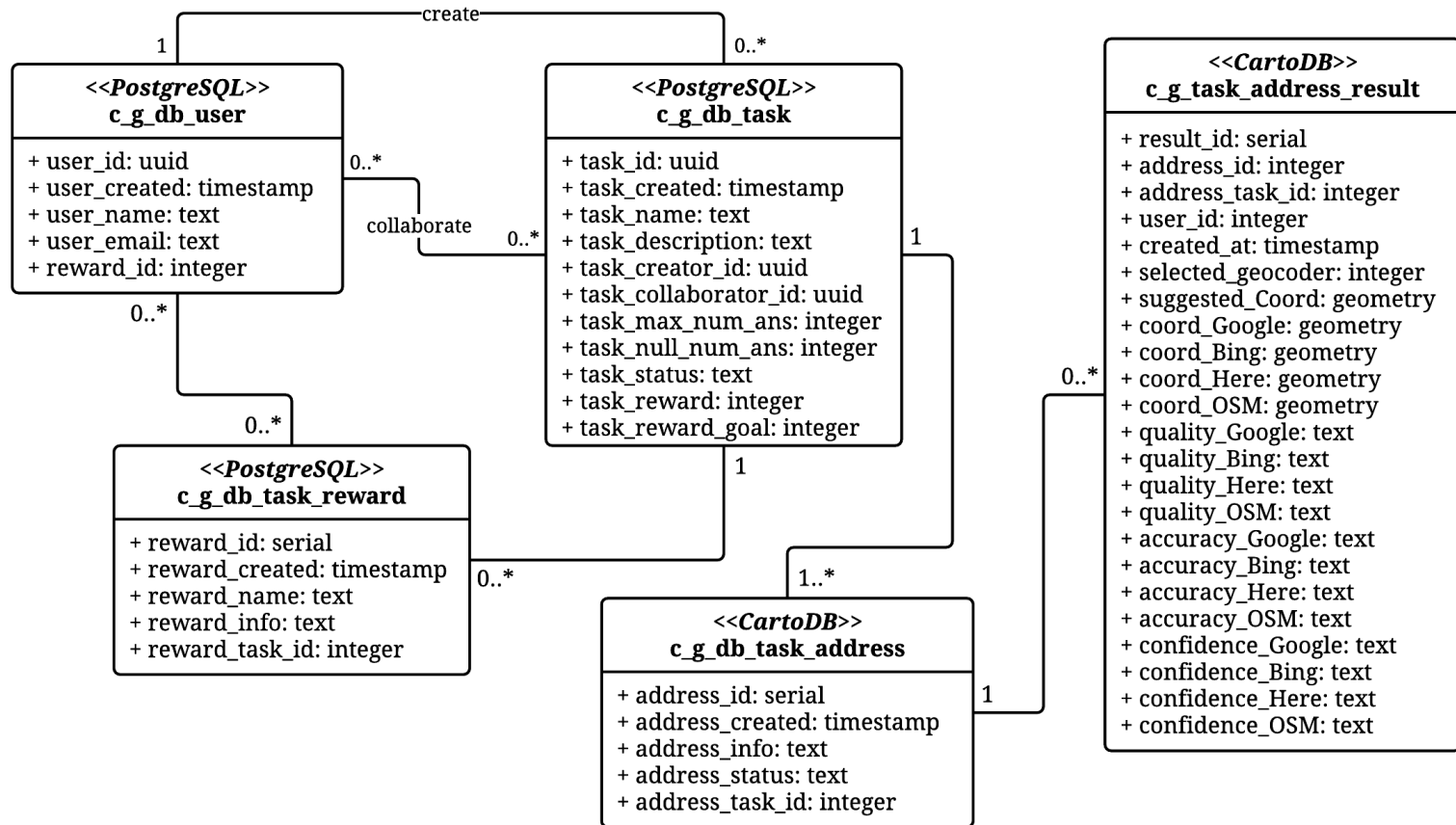
Platform-User Interaction



Data Management

- Geocoded location + 3 parameters stored in database:
- **Accuracy:** method used to calculate the location or result type.
- **Quality:** output match level or granularity of the match, location entity type.
- **Confidence:** from OpenCage API calculation method, use data response from the each API to create a confidence range between 0 and 10 based on the distance in kilometres between the South West and North East corners of each resulted associated bounding box.

Data Management



It's demo the time!

Crowd-Geocoding

Q?

My Profile

User-

Crowd-Geocoding:

Online assistance to analyse quality and geographic precision of online geocoding services

I like to CrowdGeocode!

[Register to help someone!](#)

Based on Crowdsourcing geospatial data and Volunteered Geographic Information approaches

Conclusions

- Crowdsourced collaborative approach to deal with an actual problem in the use of online geocoding services.
- Facilitates user interaction to control and evaluate the accuracy of geocoded outputs:
 - relying in the amount of collaborators reviewing results
 - Combining diverse reference sources to increase data availability.
- Task results database to generate descriptive statistics, comparative evaluation, data quality assessment using common geocoder metrics like completeness, positional accuracy against base line data and similarity between services.

Future Work

- Possibility for the user to modify the input address text to reduce the ambiguity of the entry to the geocoding process and iterate the process with the accumulative text editions.
- Enable in the task creation, to indicate the geographic feature that corresponds to the input address to geocode (e.g swimming pools), hence the user can identify the best accurate geocoder related to the searched physical object in the base map.



POLITÉCNICA



AGILE

THANKS FOR YOUR ATTENTION



Any questions please, send me a email

José Pablo Gómez Barrón Sierra

E: jp.gomez@alumnos.upm.es, jpablo.gomez@gmail.com

t : @jpablogomez

T: +34 651024537

References

- [1] H. A. Karimi, M. Durcik, and W. Rasdorf, "Evaluation of uncertainties associated with geocoding techniques," *Comput. Civ. Infrastruct. Eng.*, vol. 19, no. 3, pp. 170–185, 2004.
- [2] D. W. Goldberg, J. P. Wilson, and C. a. Knoblock, "From Text to Geographic Coordinates: The Current State of Geocoding," *URISA J.*, vol. 19, pp. 33–46, 2007.
- [3] P. A. Zandbergen, "A comparison of address point, parcel and street geocoding techniques," *Comput. Environ. Urban Syst.*, vol. 32, no. 3, pp. 214–232, 2008.
- [4] D. Roongpiboonsopit and H. a. Karimi, "Comparative evaluation and analysis of online geocoding services," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. April 2015, pp. 1081–1100, 2010.
- [5] R. Hudson-Smith, A., Batty, M., Crooks, A., Milton, "Mapping for the masses: accessing web 2.0 through crowdsourcing.," *Soc. Sci. Comput. Rev.*, vol. 27 (4), pp. 524–538, 2009.
- [6] C. Heipke, "Crowdsourcing geospatial data," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 6, pp. 550–557, Nov. 2010.
- [7] M. F. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov. 2007.