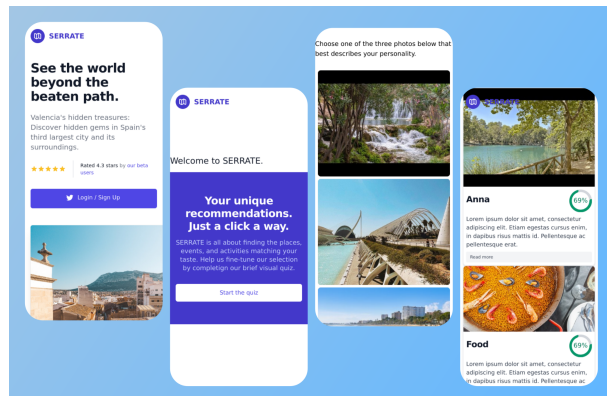


Technical Specification - SERRATE

1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

SERRATE helps DMOs to effectively direct the flow of tourists from hotspots to less known rural areas by:

- 1) integrating **cross-domain data sources** like real-time sharing of posts on social media around the region (e.g. Valencia) with info about the specific points of interests, search keywords; in order to
- 2) **extract sentiment, keywords, topics** using AI (e.g. NLP, image recognition) technologies from these posts,
- 3) matching them with a curated and extendable **storytelling knowledge base** about activities (outdoor / nature, culture, gastronomy, actual local events, e.g. festivals)



Visitors/tourists login with Twitter and take a short quiz to select what best fits their preferences. Accordingly relevant trip recommendations and activities are presented based on an index that shows how sustainable is a visit to the suggested places at the current moment.

2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

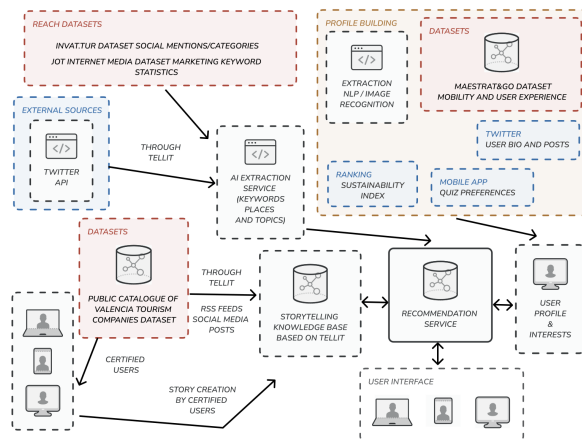
For point 1) of the project's technical scope above, we used the dataset from Turisme Comunitat Valenciana – INVAT.TUR that provides mentions in social networks for different points of interest, per month, for different product categories as a filter for the type and amount of posts that SERRATE will consider relevant and b) a curated list of hashtags and user profiles on Twitter from the dataset of public catalogues of Valencia tourism companies.

For point 2), SERRATE uses existing IN2 technologies for AI processing and enrichment and take advantage of the Maestrat&Go dataset that provides mobility and user experience, including userID, date, point of interest, geolocation, objectives (check-ins done in that point of interest), and ratings to build a more accurate user profile. Because users are registering in the mobile app through Twitter, the user profile is also additionally enhanced by the visitors Twitter bio and posts. Furthermore to capture the "mood of the day" we provide a short photo-based quiz and ask the user to select the 3 most interesting photos.

For point 3) we have customised our own storytelling platform Tellit (<https://tellitapp.com/about>) to serve as the story knowledge base around regions of interest. Stories in Tellit are able to be created on the platform from certified touristic users.

SERRATE envisions the whole process described above to be fully automated at the end, but it will start with an editorial workflow which will allow a manual human intervention at any step for social media managers of the destinations to control the user/tourist experience.

The diagram on the right shows the implemented system architecture for the SERRATE backend.



3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in?

The solution thrives from working with huge datasets. This is because the underlying AI/machine learning algorithms used for extracting features, topics, and recommendations generally increase in accuracy as more data is available. Scalability is achieved through on demand parallelisation (horizontal scaling). The computing intensive components are packaged in individual (docker) containers which can be deployed on-demand (on different VMs) and run in parallel in order to cope with high loads. Intermediate results from the analysis components are stored in a shared database that supports sharding. For ensuring flexibility of the solution, use our own cloud infrastructure as a service orchestrator and as the main entry point (e.g. serving API requests, providing search results, storing metadata) to the user-interface and frontend. Doing so allows us to separate the processing and feature extraction layer and scale it based on demand either internally or in the public cloud (e.g. dedicated AWS EC2 instances, offloading computing intensive tasks to VMs with GPU processors that are provisioned and stopped as needed). Using monitoring tools for AWS (e.g. CloudWatch) and our own infrastructure (e.g. DataDog and InfluxDB) we can scale any time to master any given workload. Many territories are facing the same issue related to managing tourism flows. While flexibility has been built in from the start, due to its nature on how different customer datasets are used, the SERRATE solution can not yet be an off-the-shelf offer. A certain customisation and adaptation is needed for each new DMO that wants to use it. This means also that it is better able to answer the specific needs of a customer. For instance DMOs that do not yet have an app can use the SERRATE tourist app as their own, while others that already have an official app can simply integrate SERRATE into their existing marketing and communication mix. SERRATE has now reached at TRL 5. Next, a pilot with a DMO will be needed in order to validate the complete solution. We expect that based on user feedback the UI will be further improved before a commercial product can be launched.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR).

IN2 is well aware of the data security and legal compliance requirements of working with data, especially with regards to GDPR. Thus from the start we defined a data management workflow which considers aspects of secure data storing, processing and sharing as well as ethical and legal compliance. Serrate will use data not only from open public sources (e.g. Wikidata) but also from partner organisations (e.g. Play&Go), clients (e.g. Invattur) and open platforms like Twitter and blogs. Whenever possible we will anonymize personal data and present aggregated results. In SERRATE we are mainly building a database of places information and visitor preference information, matching a type of place to a type of user. The information about places are completely agnostic and don't require storing any personally identifiable information. For registered users we require storing personally identifiable information (e.g. email address, name, preferences). We plan to store this only for users who provide an explicit and informed consent accepting the privacy policy and terms of the SERRATE service that will make clear for which purposes, why, what and when personally identifiable information is used, retained and deleted. Of course users will retain all their rights to access, amend, correct or delete their personal information at any time. Moreover, indeed the use of minimal personal information is necessary to facilitate the basic functionality of SERRATE, giving personalised recommendations (rather than one-size-fits all recommendations). To increase security, we operate a firewall for each server that allows only specific ports to be open (e.g. port 22 for ssh, 80 for http and 443 for https) and only specific IPs to access other services. Additionally we only allow connections using the SSH key-based authentication mechanism. During data transfer only secure protocols are used (e.g. https). After processing and extraction of metadata and topics, raw data from social media are permanently deleted and only the extracted aggregate knowledge is retained.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfils the challenge/Theme Challenges and data provider requirements

We used agile methodologies during the technical development work (SCRUM meetings, 2-week sprints). The work was carried out iteratively, aiming at producing a first MVP early, and continuously improving it based on feedback from ITI. Regular meetings with the Coach has helped validate progress achieved along the way. There were some minor technical issues encountered related to data access but these have been resolved in time. Moreover, we received access to some additional datasets after the first MVP released which helped us further improve the solution. One challenge we faced was the calculation of a sustainability score of a given POI, as there are currently no standard ways to do this, but rather different indicators from literature that could be used.