

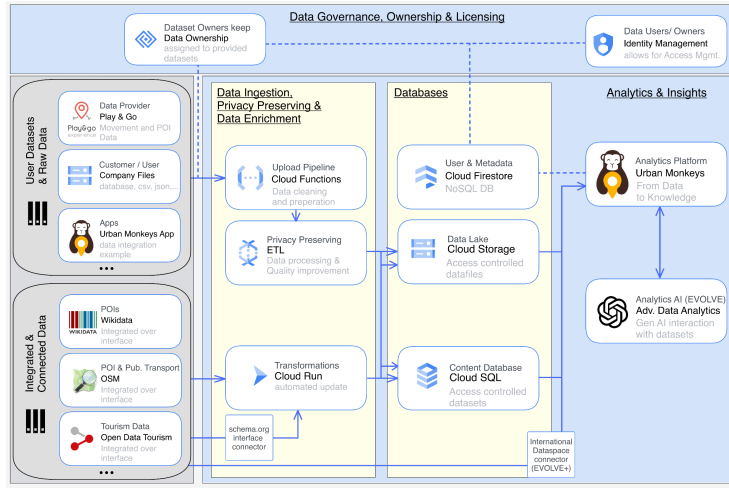
Technical Specification Double-side Page: Urban Monkeys

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.

One key challenge of the tourism industry is that it consists of many distributed and disconnected stakeholders. Employees are often experts in their domain, but lack data science expertise, especially in small/medium sized companies. Therefore, we have developed our solution to enable easy data analysis and data combination, and thus allow high-quality data analytics results for tourism and beyond.

Furthermore, we are enabling data providers to share their data and thus benefit from collaboration and shared knowledge within the industry. We can apply our solution in a scalable way to the current challenge and add value to overarching data-science challenges, not only in the tourism industry. For the MVP we focus on combining datasets and enabling problem solving by high-quality visual analytics. Through this approach, our users can easily identify the behaviour of tourists, understand local infrastructure and Places of Interest.

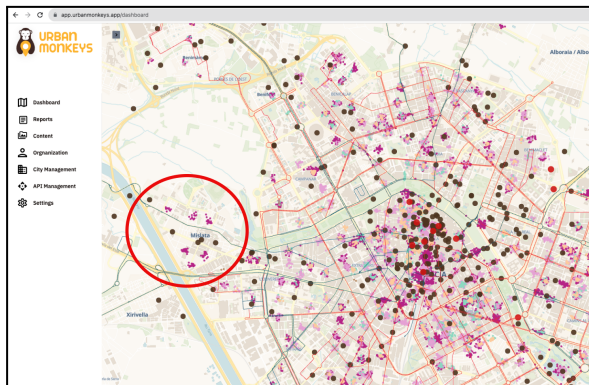
Our application empowers users to gain easy understanding of their data and identify new opportunities.



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.

We build our backend infrastructure on Google Cloud to ensure integration and security benefits for the overall ecosystem. From Google Cloud we utilise a broad array of features, such as Identity, SQL & NO-SQL Databases, Storage, Server & Serverless capabilities. Our backend code is written in Python, while our frontend is created based on the React framework. The main data sources integrated today are OpenStreetMap, Wikidata, AtlasObscura and OpenData: Tourism. For integration purposes, we have created several interfaces and automated pipelines, which allow regular updates of our data stack. Mainly the schema.org interface for tourism data enables easy integration for all data providers using the same schema. Additionally, we are utilising several open-source projects to enable automatic labelling, as well as our own ML-models to enhance available data and increase data quality. In total, we have prepared several million data points and can scale even further. This allows plug & play capabilities for many datasets. The key findings with regards to the REACH challenge and the data provided were:

- In '23, users were either using "AR" or "selfie" at each sight, whereas in other years usage was mixed
- High correlation among years, with the exception of individual POIs for specific years (e.g. '18 in Mislata)
- Individual event triggers as "selfies" have a high versatility between the years and sights, which could point to high-performing events
- Conversion after check-in depends significantly on the place (e.g. Maria Ros-Manuel Iranzo has high conversion (1.0), whereas Convent de Jerusalem-Matemàtic Marzal converts badly (0.33) in 2023)



However, it's also very clear that a significant portion of results are remaining to be uncovered within the data, which leads us to further enhance our application to include a new AI-analytics-based feature, as described in the next section.



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? Discuss how your solution could be integrated within multi-stakeholder data value chains.

The main focus of our approach was to provide a scalable solution for multiple stakeholders, as we have seen the importance of sharing data across the ecosystem, especially in the tourism industry. Already today, our solution is available globally and ready to scale for thousands of users. Besides the tourism industry we see that many other industries can benefit from our offering as well, as visualising, analysing and sharing data has never been that easy.

However, we see the current implementation as a first step towards a greater vision. On our roadmap we have two core functionalities planned, which will improve the value for our users drastically. Within the EVOLVE phase of REACH we will additionally implement a connector to international data spaces to easily integrate and extend the data value-chain. With this, our application will also become an onboarding ramp for smaller companies to share their data with a broader audience, while still keeping the sovereignty of their data.

While testing our application for the challenge and exploring additional value propositions we made an exciting observation with regards to generative AI for data science. When combining datasets with leading Gen.-AI-Models, users can conveniently interact with their data and derive meaningful results with low entry barriers. This way, we can democratise data science across all kinds of organisations and employees, by enabling anyone in a company to analyse, understand and share data in seconds, and thus make decisions based on easily understood data. Today, there is no technical integration publicly available, however we expect this to change over the next few weeks. With our data sharing and combination approach we are in a unique position to provide value in that domain as the combination of datasets is especially vital to derive meaningful results. The current TRL level is 7.

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). Explain how you tackle data governance and sovereignty aspects in your solution.

We are guaranteeing a high security standard and strict data governance and sovereignty. To accomplish this, we have implemented several security and compliance approaches: General authentication is implemented with the Google Cloud Identity Platform, where users can sign up onto our service using a classic email / password approach. From this authentication, the user receives a temporary token, which is required to interact with our platform, e.g. for data access. Users can even use our application without uploading the data to our server at all: In case there are regulatory restrictions, data analysis and rendering can be carried out on the user device directly. When deciding to upload data to our servers, the user can share their files with partners or other users conveniently. In the file's metadata, file owners are defined who can decide how the file is shared (or deleted). Thus, the file owner can always decide to revoke access. In a later iteration we will include organisation management, where company structures can be re-created and files can easily be shared and managed within team/supervisor structures. All data is encrypted in transit and in rest. Within our upload process we can also implement checks to ensure privacy preserving data sharing (planned for EVOLVE phase).

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 fulfills the challenge/Theme Challenges and data provider requirements.

We have been following a generic sprint based agile development approach with weekly code and progress reviews, coordinated with Jira. Over the project, we first conducted an additional architecture validation while starting already on data engineering and design. We've encountered a scalability challenge with our Explore implementation, as the solution had a very specific focus on the tourism industry only. After we have conducted several interviews with industry experts from tourism and beyond, we have identified a greater opportunity to provide an industry agnostic solution, which aligns with the overall vision and scope of REACH. This led us to switch our technology stack to better scalable frameworks with more relevant data science packages. Our preferred technology framework was next.js, which we had to abandon due to incompatibility with critical packages. We then settled for React, due to their maturity and integration capabilities. This solution is now behaving as desired and we've built a solid application based on this. With regards to backend, the designed architecture mainly remained, with the exception that we're using a combination of a separate user database instead of using the Identity Access Management due to authorization limitations (and cost implications).

With regards to data quality and data challenges, we had several iterations on data preparation. Our own solution, however, proved immensely valuable for troubleshooting and testing, as we could simply detect abnormal behaviour when visualising our data.

