****

# 4 Summary of R&D Projects

The company has developed a unique solution for both customers and brands by providing them with unadulterated information regarding consumer needs and patterns. This highly unique platform enables brands to use consumer authenticated and provided, browser, mobile, and productivity data alongside large social media feeds, to analyse and identify trends in customer needs and desires and incorporate those into their products and services.

Significant research and development activities have been, and are continuing to be, undertaken. These activities continue to show solid research based results, which is evident in the continued development of the products.

For users, they have built a user dashboard and insights panel to help consumers understand patterns in their online footprint through metrics and graphs.

For brands they are in the process of developing products that can perform thorough cross-examination of consumer behaviours and provide important analytics related to their trends from data collected through various social media platforms, devices and technologies.

The company are also using various Application-Programming-Interfaces (APIs) and building browser extensions in order to further expand their data collection sources. Their core analytics engine can make any user data anonymous while extracting insights from it that are utilised by brands for creating desired products and developing effective marketing strategies.

# 5 Baseline Technology

The company conducted a thorough analysis of the competitive marketplace and studied many different offerings that were available for both the customers and the brands.

Existing competitors relied primarily on indirect methods for collection of large amounts of consumer data through focus groups, survey, website cookies and web scraping. This comprise an inefficient method for analysing consumer behaviour since only a small number of consumers can be contacted / interviewed in order to gain insights for developing products and strategies for larger audiences. Their target market required large volumes of online consumer data through social media that is live, natural and unaltered; however, the existing baseline technologies were unable to meet such demands.

# 6 Scientific / Technological Uncertainties

There are a number of scientific/technological uncertainties within these projects where knowledge of whether or not the project/activity was technologically feasible, or how to achieve it in practice, was not readily known by a relevant competent professional.

The main uncertainties are detailed below:

1. If the analytics system can be designed to handle the large volume of data (both scale and variety of data sources).
2. Whether the system will be able to handle mass volume queries for different types of data.
3. Will it be possible to process large volumes of data in a timely manner in a simple interface for the company’s services to be made available for small and large businesses?
4. Also, can server power be used in a more efficient way in order to only use what is necessary thereby reducing costs incurred on third party cloud computing servers?
5. Would we will be able to break down URLs to identify search habits, search topics?
6. Would it be possible to understand customer habits across apps, web browsing, and locations?
7. Whether it would be possible to identify duplicates and detect fraud documents in the database.
8. Whether the platform can be scaled by global partnerships with existing players in a seamless way.
9. Whether it would be possible to create a secure iFrame into the company’s site to allow for partners to use functionality for testing different methods of creating a seamless integration for partnerships globally. Would it be possible to create this type of access to large number of partners?
10. Whether the system would be able to produce meaningful and accurate management information in a timely manner.
11. How to minimise the amount of data stored thereby reducing costs of server and maintenance? If it is possible to grab summary level data from APIs to minimise the number of calls and information pulled and stored will it be possible to integrate with large number of bespoke systems to transfer and obtain the data?
12. Would it be possible to integrate an additional messaging queue system that will allow the company to overcome the limitations of the IBM Softlayer messaging queue?
13. Whether the system would be compliant with EU data regulations?
14. Will it be possible to create a personal identity for the digital world which will enable users to port their data as an asset class?
15. Would it be possible to capture user behaviour on mobile devices, especially iOS and Android, while not impacting battery life?

# 7 How the Uncertainties Were Overcome

The scientific/technological uncertainties outlined above required substantial research & development work and were addressed as follows:

1. The company is continually developing and adapting their data structure with a noSQL database (MongoDB). They’re working to scale our analytics with Python as its proven to be the most scalable. Analytics development is in progress as they continue to scale their SaaS offering.
2. Significant and ongoing research & development has been undertaken with regards to establishing and maintaining reliable communication between the various APIs that they are integrating with as they are changed by the owners (especially Facebook and Twitter), and they’re adapting the data schema to make it more efficient for queries as they are already looking at over 100k data points on average for each user.
3. The company is still researching how best to make the data and analytics available real-time. Amongst their advancements over the past year, are moving from being able to do mostly manually triggered runs of the insights for users to now doing it consistently on systematic scheduled processes every 14 days with error handling schedulers to account for exceptions and process them in an automated fashion.

The company is now performing development and research to move it from every 14 days to be within a week and real-time thereafter. The complexities mainly come from the various APIs being differently structured with bespoke error handling and rate limiting criteria. So the key activity has been to optimise not only the way they call to the APIs but then how they process the data into their systems and then optimally run the insights processes. They are conducting research & development work with a goal to create the data structure to generate and enhance complex analytics using a graphical drag & drop interface.

1. The company has set up their servers in the EU to comply with the data laws accordingly. They have undertaken significant research in configuring the servers to optimally support our production and development environments so as to minimize the impact of ongoing scheduler runs to fetch data and process insights. They’ve also separated the key processes to add redundancy and controls for their production sites as the traffic has increased consistently over the year. They are continuing to invest significantly in research into this area.
2. The company has devised the algorithm to parse URLs to analyse search keywords, phrases along with e-commerce search habits for UK shopping sites.
3. They conducted research & development work resulting in the creation of a data schema to capture data for each user in a coherent way so that they can cross-reference behavioural data regardless of source. This is especially challenging as the data continues to grow and they are finding new ways to keep the metadata for analysis as they archive old data.
4. The company has started building algorithms to identify duplication and detecting incorrect data across data sources starting with demographic information. They are also cross-referencing user details to ensure accuracy. Expanding the type of data they can review systematically with research.
5. They are developing a widget to simplify onboarding with panel partners in the UK, US and Canada this year. The widget configurations are still being researched as they are spanning across numerous devices and screen sizes.
6. They have implemented their secure iFrame via the widget. There are complications around the iFrame sizing across devices and ensure the user interactions are done in a secure environment across multiple domains. They will be going live with panel partners in 2016.
7. The company has now implemented controls and management reporting in place to ensure the system is up and running with minimal downtime. Addressing these issues took considerable research & development time and resources as they are dealing with multiple servers and databases.
8. They have optimised their querying process for Facebook APIs as that resulted in the largest calls for info pulled and the most rate limiting error. This has reduced Facebook data and error rate by 95%. Research is now in progress for Twitter API.
9. The company has assessed multiple options and has now implemented and integrated RabbitMQ to handle messaging queues across all schedulers. They are now able to monitor processing live through a dashboard.
10. Their systems are compliant with all EU regulations currently and they are monitoring the progress on EU GDPR implementation guidelines. At the moment, they are compliant with any published guidelines, including not storing data for Russian users.
11. The company has setup the data schema around each user and are now researching the portability of the data across to other systems. The research is currently in progress.
12. After considerable research the company has developed the iOS and Android apps, that required much testing and enhancements to make it effective.

On Android they have developed their own algorithm to optimise capturing location, app usage and connectivity data so that there’s visibly no impact to the battery life. The next phase is to enhance behaviour capture to include music and health habits.

On iOS they have enhanced their methods for location and usage tracking further to save 25% of usage and processing power required. This R&D work is ongoing.