HRA Project: Creating a Dashboard for Usage of the HRA Portal

Jeet Patel, Aatman Vasoya, Srihari Kamath, Harsh Lotia, Harsh Chauhan

Abstract—This study addresses a significant gap in understanding user interaction with the Human Reference Atlas (HRA) Portal, a critical resource for researchers and educators in the biomedical field. We developed a comprehensive dashboard utilizing Google Analytics Data to visualize user engagement and behavior on the portal. Through a meticulous analysis of session durations, and user activity, we derived insights into the portal's reach and effectiveness in retaining users. Our findings indicate a substantial number of unique visitors, coupled with an in-depth exploration of specific sections of the portal by returning users. The dashboard reveals key interaction patterns, including a high bounce rate juxtaposed with meaningful engagement from users who choose to remain on the portal. These metrics underscore potential areas for optimizing content and enhancing user experience. By comparing two domains within the HRA Portal, we identified discrepancies in user engagement, informing tailored improvements. The paper highlights the significance of visual analytics in the strategic development of educational web resources and the pivotal role of user-centric data in shaping content strategy. The HRA dashboard can be viewed at PowerBi Dashboard.

Index Terms— HRA, Visual Analytics, Dashboard, User Engagement, Google Analytics, PowerBi.

Introduction

In an age where data is everywhere, the efficacy of scientific web portals largely depends on their ability to present information in a user-friendly and engaging manner. The Human Reference Atlas (HRA) Portal serves as a critical interface for researchers and educators, offering comprehensive data and resources on human health. The Human Reference Atlas (HRA) is an ambitious initiative designed to create a comprehensive, high-resolution, three-dimensional atlas of all cells in the healthy human body. Such an atlas is not only a pivotal reference for biomedical research but also a cornerstone for understanding human biology at the cellular level. The HRA Portal serves as the digital interface where this information is curated, stored, and made accessible to researchers, educators, and the broader scientific community. It includes various features such as 2D/3D Reference Objects, Anatomical Structure, Cell Types, plus Biomarker Tables, and Organ Mapping Antibody Panels (OMAPs).

Despite the crucial role that the HRA Portal plays, there has been limited research into how users interact with the platform. Insights into user engagement are essential for improving the portal's design, enhancing its educational value, and ensuring that the HRA remains a pertinent and utilized scientific resource.

This study aims to address this gap by developing a comprehensive usage dashboard for the HRA Portal. The dashboard's goal is to filter complex user interaction data into actionable insights, addressing specific questions about user behavior, patterns of page visits, and geographic distribution of the portal's audience.
By accurately mapping user interactions and preferences, the project aims to inform and drive strategic improvements to the HRA Portal. Such enhancements will not only augment the user experience but also ensure the portal's alignment with its educational mission and scientific objectives. The insights gleaned will be invaluable for tailoring content, optimizing features, and directing outreach and training efforts such as the Visible Human Massive Open Online Course (VHMOOC) and other educational events.

1 Literature Review

This study is situated within a broader context where analytics have been harnessed to distill complex data into strategic insights across various domains. While the specific application of Google Analytics data for the HRA Portal is novel, the concept of leveraging such data for in-depth analysis and visualization has precedents in the field. For instance, Murphy's exploration [1] of Ohio State University's library data visualization with Tableau demonstrates the power of integrating diverse data sources. Similarly, the work of Hagen et al. [2] showcases the potential of combining Google Analytics and Trends for policy analysis, underscoring the versatility of these tools. The framework by Kumar et al. [3] illustrates how analytics can be adapted to various data environments, including cloud-based platforms like AWS. Furthermore, the approach by Hanamanthrao and Thejaswini [4] to visualize clickstream data for an e-learning portal echoes the need for real-time insights in an educational context. Each of these examples informs and enriches the current study's approach to understanding user engagement through data visualization, providing a foundation upon which the HRA Portal's dashboard is developed.

2 Methodology

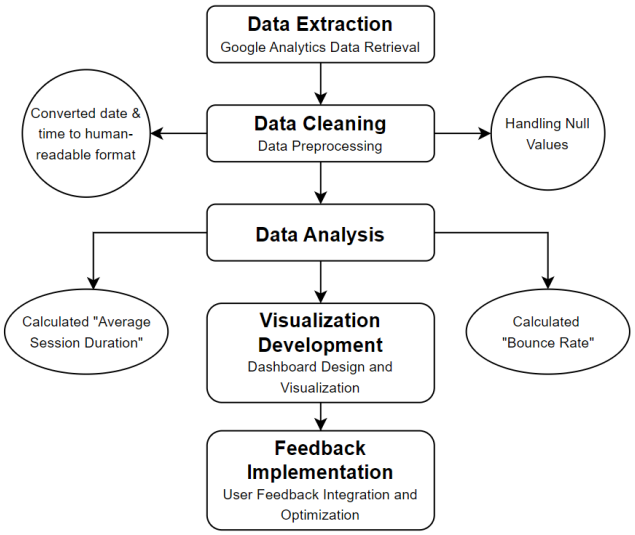

Fig. 1. Workflow Design

2.1 Synopsis of the Data

The dataset for our project is procured from Google Analytics, stored within a CSV file that records user interactions on the HRA portal. This data explains the patterns of visitor engagement with the site, encompassing nuances such as visit duration and the specific actions undertaken. By meticulously tracking these interactions, we can obtain invaluable insights into user behavior, an important component in enhancing the overall user experience.
The dataset contains 192,271 entries and is structured with the following attributes:

- **user_pseudo_id**: A numerical identifier for users, represented as a float.
- **hostname**: The hostname of the website visited, as a string.

- **page_location**: The specific URL visited on the website, as a string.
- **event_date**: The date of the event, formatted as a string.
- **user_first_touch_timestamp**: The timestamp of the user's first interaction, as a string (includes date) in UNIX format. There are 188,804 non-null entries in this column.
- **event_timestamp**: The timestamp of the event, as a string (includes date). In UNIX format
- **event_name**: The name of the event (e.g., first_visit, session_start, page_view), as a string.
- **event_category**: The category of the event (e.g., mousemove), as a string. There are 165,773 non-null entries in this column.
- **event_label**: Additional details about the event, as a string. There are 155,728 non-null entries in this column.

Aligning our analysis with the client's requirements, we have strategically focused on two critical domains: humanatlas.io and apps.humanatlas.io. The humanatlas.io is the web home page for the HRA portal, it gives access to data, tools, use of HRA and a general overview of the website. The app.humanatla.io allows access to Exploration User Interface(EUI) , Registration User Interface (RUI), Improve Cell Type Annotations application and Predict 3D Spatial Origin of Tissue Samples application. Data was collected from January 2 through February 26, 2024, however data does not exist for humanatlas.io for the entire month of January, this must be kept in mind while looking at the visualization. By concentrating our efforts on these domains, we can tailor our analysis to the areas most frequented by users, ensuring that our findings are not only pertinent but also actionable for implementing informed enhancements.

2.2 Data Cleaning

In preparation for analysis, meticulous data cleaning procedures were undertaken to ensure the integrity and interpretability of the dataset. Initial scrutiny revealed the presence of missing values in crucial fields such as user_first_touch_timestamp, event_category, and event_label. Recognizing the importance of maintaining data comprehensiveness, a decision was made to retain rows with missing values. Subsequently, UNIX timestamp formats were transformed into human-readable date formats, facilitating temporal analysis and enhancing data interpretability. Addressing the complexity of link data within the page_location field, encoded URLs underwent a transformation to readable formats into description, detail, and location. These rigorous data cleaning efforts were instrumental in guaranteeing the accuracy, reliability, and usability of the dataset for subsequent analysis.

2.3 Data Analysis

We conducted a comprehensive examination of key metrics to gauge user engagement with the HRA Portal. This involved calculating the number of visits and page views, fundamental indicators of website traffic and content consumption. To ensure the accuracy and reliability of our findings, we implemented a rigorous verification process, cross-referencing the calculated metrics with the original dataset.

Additionally, we delved deeper into user behavior by calculating key metrics crucial for understanding user engagement with the HRA Portal. One such metric, average session duration, was computed by aggregating session durations and grouping them by hostname. This allowed us to distinguish patterns in user behavior across different sections of the portal. The resulting average session duration, presented in both seconds and HH:MM:SS format, provides valuable insights into how long users typically interact with the portal during each visit. For instance, on the "apps.humanatlas.io" hostname, the average session duration was calculated to be 1539.72 seconds, or approximately 25 minutes and 39 seconds, while on "humanatlas.io" it was 392.72 seconds, or about 6 minutes and 32 seconds. Understanding average session duration aids in optimizing content to enhance user engagement and retention.

We also calculated the bounce rate, a pivotal metric indicating the proportion of sessions where users navigate away from the portal after viewing only one page. This calculation involved identifying sessions with zero duration (indicating immediate exits) and comparing them to the total number of session starts. The resulting bounce rate offers insights into the effectiveness of the portal's landing pages and overall user experience. For instance, on the "apps.humanatlas.io" hostname, the calculated bounce rate was 0.676, similarly, on "humanatlas.io," the bounce rate was calculated to be 0.954. These values provide valuable insights into user behavior and highlight areas of potential improvement in content relevance and navigation. By identifying areas with high bounce rates, we can pinpoint potential issues and implement strategies to improve user engagement, ultimately fostering greater user satisfaction and retention.

Furthermore, in our analysis, we also calculated the average pages per session visited by users on the HRA Portal. This metric provides valuable insights into user engagement and navigation behavior, indicating the average number of pages users interact with during each session. For instance, on the "apps.humanatlas.io" hostname, the average pages per session was calculated to be approximately 1.50, while on "humanatlas.io," it was approximately 1.82. These values highlight differences in user interaction patterns across different sections of the portal.

3 Results and Discussion.

Data cleaning, pre processing and analysis was performed using Python. Excel files generated from the analysis were: cleaned_data, processed_HRA and sorted_data (includes avg session per duration and bounce rate) were used to make the dashboard using PowerBi. The visuals created are discussed below.

By using the event_date and event_timestamp fields, we performed a time series analysis to uncover trends and patterns in user activity over time. This analysis helped identify peak activity periods, revealing insights into when users are most likely to visit the website and engage with content. Visualization was accomplished through area graphs, showing user activity trends over the observed period.

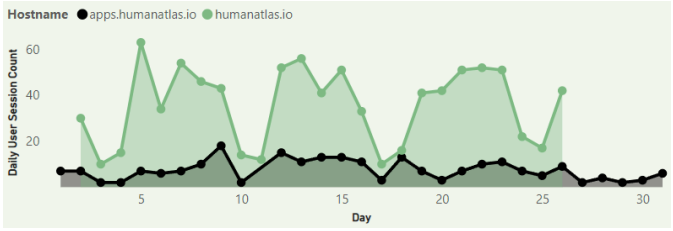

Fig. 2. *Daily User interaction analysis by event type*

A bar chart details the most trafficked pages on the portal, providing clear insight into user preferences and content effectiveness. This visualization serves as a direct indicator of what content resonates with the user base, and it can inform content creation and optimization efforts to better align with user interests and needs.

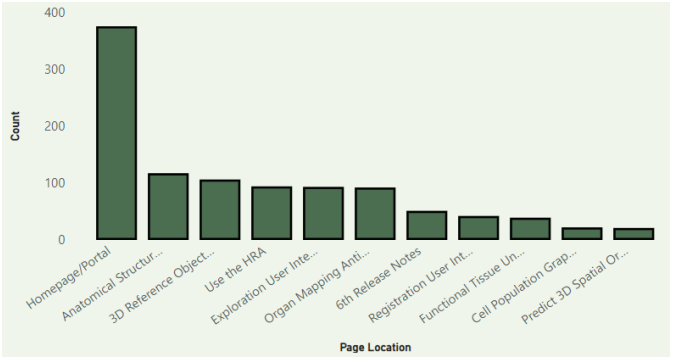

Fig. 2. *Most Frequently visited page*

Next in the visualization of the HRA Portal's user base, here we employed a pie chart to differentiate between unique and repeating users, revealing that 85.4% of users were visiting for the first time, while 14.6% were returnees. The larger proportion of unique visitors highlights the portal's success in attracting a new audience, possibly reflecting its outreach efficacy or the value of its content to newcomers. Meanwhile, the notable portion of repeating users points to a core of engaged users who find the portal worth revisiting. These insights are important in shaping the portal's user experience and content strategy, ensuring it appeals to both new and returning

visitors, and in gauging the effectiveness of any modifications to the platform.

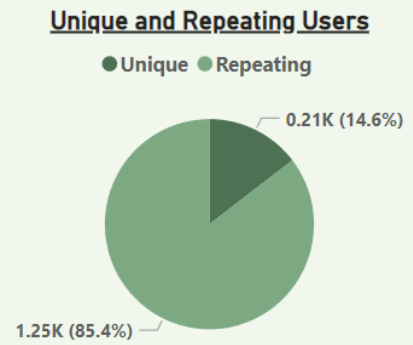

Fig. 3. *Unique and Repeating Users*

This part of the dashboard is particularly insightful, breaking down user interactions by 'event category', 'page description', and 'page location' and quantifying the occurrences of these events. Understanding the user journey through this lens allows us to identify which actions are most common and which pathways through the portal are preferred by users. This can reveal the intuitiveness of the site layout and highlight potential areas for improvement in navigation and user experience design.

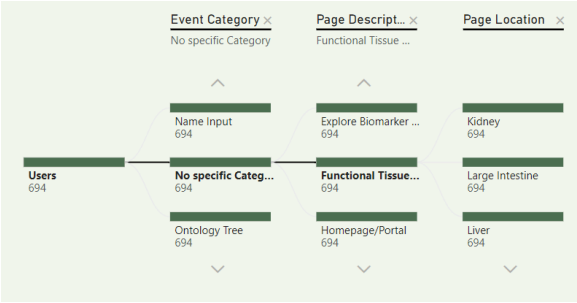

Fig. 4. User Journey Analysis

In evaluating the HRA Portal's user engagement, we tracked key performance indicators, revealing a bounce rate of 0.95 and an average session duration of 25 minutes and 39 seconds across app.humanatlas.io domains. For humanatlas.io bounce rate 0.95 and avg session duration of 6 minutes and 32 seconds. The recorded 2,589 page views and 1,479 visits offer a quantitative measure of the portal's traffic, suggesting a relatively high interest and interaction level.

The bounce rate, which appears high, coupled with the substantial average session duration, indicates that while many users leave after viewing just one page, those who engage tend to do so meaningfully. These metrics provide essential insights into the portal's user engagement patterns, with the potential to inform strategies for content optimization and user retention.

hostname	Bounce rate	Avg session duration
app.humanatlas.io	0.68	00:25:39
humanatlas.io	0.95	00:06:32

Table 1: Comparison of User Engagement Metrics

4 Conclusion

The in-depth analysis of user engagement with the HRA Portal unveils valuable insights pivotal for its optimization. Metrics such as average session duration, bounce rate, and average pages per session offer a comprehensive understanding of user behavior patterns, facilitating strategic decisions aimed at enhancing user satisfaction and retention. Supported by meticulous data cleaning procedures and robust analytical methodologies, these insights ensure the reliability and relevance of the findings. By aligning the analysis with client requirements and focusing on critical portal domains, actionable insights are derived, guiding future improvements to the portal's usability and educational value.

In conclusion, the development of a usage dashboard represents a significant milestone in advancing the HRA Portal's functionality and aligning it with scientific objectives. The insights gained from this study serve as a roadmap for strategic enhancements, ensuring that the portal remains a vital resource for researchers, educators, and the broader scientific community. Through continuous optimization efforts driven by data-driven insights, the HRA Portal can evolve to meet the evolving needs of its users, maintaining its relevance and impact in the field of biomedical research and education.

5 Acknowledgments

We extend our deepest gratitude to Libby Maier for allowing us to work on HRA's google analytics data. We also thank professors Andreas Bueckle and Michale Ginda for their invaluable guidance and insights throughout this project, which have been instrumental in enhancing our understanding of the HRA data and visualizations.

6 References

[1] Sarah Anne Murphy murphy.465@osu.edu (2013) Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making, Journal of Web Librarianship, 7:4, 465-476, DOI: 10.1080/19322909.2013.825148

[2] Loni Hagen, Thomas E. Keller, Xiaoyi Yerden, Luis Felipe Luna-Reyes, Open data visualisations and analytics as tools for policy-making, Government Information Quarterly, Volume 36, Issue 4, 2019, 101387, ISSN 0740-624X, https://doi.org/10.1016/j.giq.2019.06.004

[3] K. Kumar, J. Bose and S. K. Soni, "A Generic Visualization Framework based on a Data-Driven Approach for the Analytics data," 2017 14th IEEE India Council International Conference (INDICON), Roorkee, India, 2017, pp. 1-6, doi: 10.1109/INDICON.2017.8487236.

[4] R. Hanamanthrao and S. Thejaswini, "Real-time clickstream data analytics and visualisation," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 2139-2144, doi: 10.1109/RTEICT.2017.8256978.