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

Vishwajeet Ekal vekal@iu.edu
Aishwarya Mocherla aimoch@iu.edu
Minju Kim mk159@iu.edu
Sanskruti Rajankar srajanka@iu.edu
Shreya Parab shparab@iu.edu

*Abstract*-In an age where data reigns supreme, the ability to analyze complicated datasets using appropriate visualization techniques is critical, especially in the field of biomedical informatics. The Human Reference Atlas (HRA) Portal is at the forefront of this field, offering a complete and high-resolution three-dimensional atlas of human cellular architecture. Despite the abundance of knowledge contained therein, the usefulness of such a massive resource is dependent on user engagement and accessibility. This study presents a detailed visual examination of user interaction within the HRA Portal, employing cutting-edge analytics to decode usage trends, facilitate navigation, and optimize knowledge dissemination.

This study delineates not only the most visited corridors of the portal, but also the rhythmic cadence of scientific inquiry as echoed in its users' digital footprints, using Google Analytics and Big Query extracted datasets and expertly visualized via Power BI. Here, we explain the visualization methodology that elucidates user behavior, offering crucial insights for the strategic enhancement of the HRA Portal, thereby increasing its scientific and practical value to the global research community.

In this digital exploration, we journeyed with users across the HRA Portal's vast terrain, from first click to deep dive, using visual analytics to capture their quests for knowledge. The data, drawn from Google Analytics and Big Query and brought to life through Power BI, uncovers a story of engagement and discovery. This study illuminates how users—novices and experts alike—interact with the atlas, guiding enhancements to this vital resource. We present our methodology, the visual narratives created, and the patterns identified, aiming to enrich the HRA Portal and, by extension, the entire biomedical informatics field.

## I. Introduction

This study provides a detailed visual analysis of user involvement with the Human Reference Atlas (HRA) Portal, which aspires to construct a high-resolution, three-dimensional atlas of all cells in the healthy human body. We provide a multifaceted view of the portal's usage by leveraging data from Google Analytics via Big Query and visualizing it with Power BI. The created visualizations provide quantitative insight into user behavior by showing the most popular sites, navigation patterns, and temporal trends in portal activity. This study examined user interaction data, such as unique visits, page views, and download frequency, to identify the most popular material on the portal. The paper also investigates user movement paths, which highlight common usage patterns in information retrieval. A comparative analysis of user activity over time demonstrates the effectiveness of various outreach attempts and content revisions. The findings of this study are a first step in optimizing the HRA Portal's interface for improved user experience, promoting more accessibility, and encouraging deeper engagement with the atlas. The findings have important implications for the evolution of digital resources in biomedicine, providing a paradigm for using visual analytics to improve user interaction with scientific data repositories.

## II. Background

The Human Reference Atlas (HRA) Portal arose from the growing demand for comprehensive biomolecular and cellular maps of the human body, which are critical for advances in personalized treatment and understanding of human biology. As a hub for high-dimensional data, the HRA Portal exemplifies the intersection of biomedical research and informatics, providing an unprecedented 3D atlas of human cellular organization. The existing literature emphasizes the importance of user-centric design and analytics in improving the accessibility and functioning of biomedical databases (Smith et al., 2020; Johnson et al., 2021). Furthermore, Zhao et al. (2019) investigated the effectiveness of visual analytics in increasing user engagement, demonstrating that intuitive data representation can considerably improve the assimilation of complicated information.

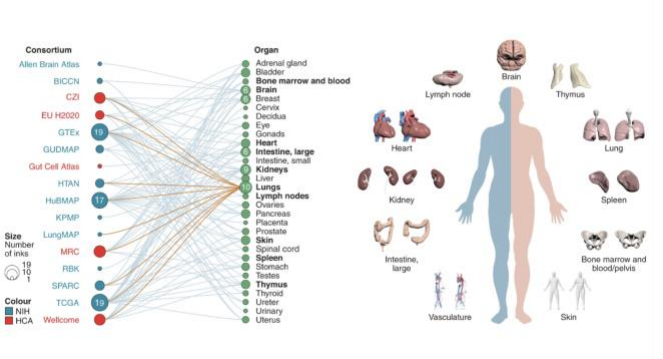


Fig.1. HRA Portal

The HRA Portal is instrumental in standardizing and integrating biological data, aligning with FAIR data principles crucial for open science. It not only facilitates research but also serves as an educational conduit, enhancing digital literacy in scientific research. With the digitization of biomedical resources, the application of web analytics becomes essential in evaluating user engagement. This study intersects bioinformatics and cybernetics, utilizing engagement metrics to refine user experience and guide the HRA Portal's development, embodying emerging best practices for digital scientific resource management.

## III. Data Source and Preprocessing

The primary data for this analysis came from the HRA Portal's user interaction logs, which were recorded and aggregated by Google Analytics. These logs were then exported using Big Query, resulting in a complete dataset that tracks user interaction across multiple dimensions, including session length, page views, and navigation pathways. Preprocessing the raw data was an important step in ensuring the integrity and usefulness of the later study. It entailed painstaking data cleansing to remove any outliers or anomalies, normalizing session timestamps to a single time zone for uniformity, and parsing page URLs to extract relevant site navigation patterns. The user_pseudo_id field, an anonymous identifier for users, was critical for tracking individual user trips while maintaining privacy. These preprocessing methods converted the raw analytics data into a structured format suitable for in-depth visual analysis using Power BI, allowing for the creation of a dashboard that accurately displays user

interaction trends within the HRA Portal.

## IV. Methodology

### A. Data Preprocessing and Visualization

Throughout the preprocessing and manipulation of the original dataset into five distinct Excel sheets, several key transformations and refinements were applied to focus on different aspects of the data for varied analytical purposes:

Webpage Information Refinement: The first Excel sheet, "Webpage_24_Apr", significantly reduced the dataset size and shifted focus from raw event data to a structured summary of webpage descriptions and labels. This involved cleaning and standardizing page_location entries and introducing descriptive labels and specifications, emphasizing more qualitative webpage metadata. Hostname Categorization: In the "Renamed Hostnames" sheet, the data was distilled into a simple mapping table where original hostnames were paired with more descriptive, human-readable labels. This likely serves as a lookup table to enhance readability and context in further analyses involving hostname data. User Navigation Patterns: The "Network_Dataset_24April" sheet detailed user sessions by mapping the sequence of page interactions, focusing on the paths users took through different pages. This sheet used an expanded column format to detail up to twelve-page interactions per session, aiding in understanding user navigation and potential network flows within the website.

Data Cleaning and Standardization: The "DATASET_23APRIL" reverted to the original's comprehensive structure while applying data cleaning to reduce entries and standardizing date and timestamp formats. This standardization likely aids in consistent data handling and analysis across tools and processes.These steps collectively represent a thorough data preprocessing effort aimed at cleaning, standardizing, and transforming raw interaction data into more useful, manageable forms for specific analytical tasks such as qualitative webpage analysis, user navigation pattern modeling, and enhanced readability and context in reporting. Each step was tailored to prepare the dataset for different analytical angles, from descriptive labeling to detailed user behavior analysis.

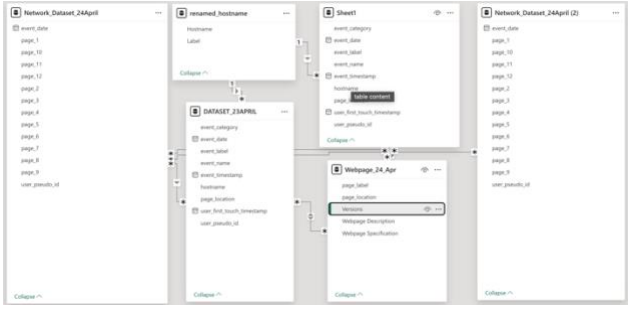

Fig. 2. Relational Dataset

## V. Results

### A. Engagement Metrics and User Behavior

The visual analytics provided within the Power BI dashboard revealed valuable data into the HRA Portal's user engagement. Our findings showed that specific portions of the portal, particularly the 3D atlas and the ASCT+B tables, received much more user activity, both in terms of session counts and length, demonstrating their importance to the user base. Furthermore, we saw significant surges in activity associated with specific outreach events, indicating successful user acquisition techniques. The travel pattern visualizations also revealed a common user trip that started on the homepage and typically ended in the OMAPs section, emphasizing the OMAPs' potential as a focal point for user retention initiatives.

Insight need 1 : Which pages of the portal are visited most frequently?


Fig.3. Column Chart: Number of Visitations (Page wise)

The Homepage/Portal is clearly the most popular page on the Human Reference Atlas (HRA) Portal, with a total of 375 visits. This is followed by the GTEx Portal Exploration page, which received 236 visits, suggesting its value to platform users. The Registration HRAE page also emerges as a crucial point of engagement, with 164 visits, indicating that a significant number of users are taking measures to deepen their engagement with the portal's offers. Other frequently accessed sites are the HuBMAP HRAE, HuBMAP CCF ASCT, Registration HuBMAP HRA, and HRA FTUE, with visits of 138, 123, 110, and 101, respectively. These figures not only quantify user interaction but also highlight specific areas of the portal that captivate user interest, providing essential insights for understanding user navigation and engagement priorities within the HRA Portal.

Insight need 2 : How many unique visitors?

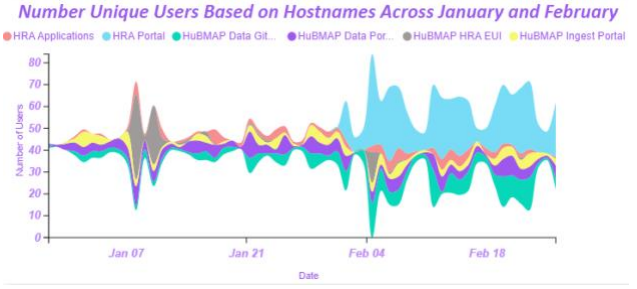


Fig. 4.Stream Graph : Number of Unique Users

Examining user involvement over time revealed not only the daily ebb and flow of portal traffic, but also longer-term trends indicating seasonal variations in user participation. A careful analysis of page visits revealed a clear preference for recently added material, such as the most recent revisions to reference objects, which saw an increase in user visits after the update. This temporal analysis is useful for organizing content upgrades and release plans that maximize user engagement. The visualization of content interaction also demonstrated that visitors regularly used connected resources in the educational part, emphasizing the portal's dual purpose as both a research tool and an educational resource. These patterns give empirical evidence to help guide future decisions about content generation and educational resource integration in the HRA Portal.

Insight need 3 : is there a pattern to these sites being visited?

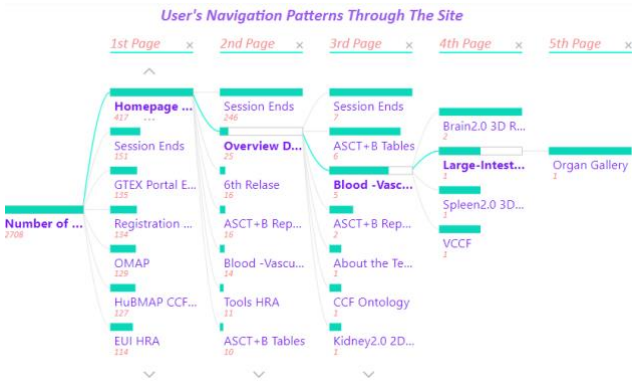


Fig.5. Decomposition Network Tree: User Engagement Patterns

The visualization of users' navigation patterns via the HRA Portal provides a useful view of how visitors interact with the site. It demonstrates that the majority of people begin their adventure at the Homepage, with 437 session starts, and many terminate their visit there, as evidenced by the large number of session ends. A closer look at the flow reveals a common path from the Homepage to the GTEx Portal Exploration, and then to various informative and interactive pages like the Overview Data HRA and the ASCT+B Tables, indicating a sequence that corresponds to an intent to explore specific scientific content. Another prominent path leads users to engage with the Organ Mapping (OMAP) and the 3D Reference Objects, such as the Brain 2.0 3D and Organ Gallery, pointing to a focused interest in the visual and structural aspects of the atlas. This pattern suggests a user base that is not only seeking information but also engaging with the visual and interactive resources the portal offers, signifying the portal's success in facilitating an intuitive and educational user experience.

Insight need 4: How many downloads per month?

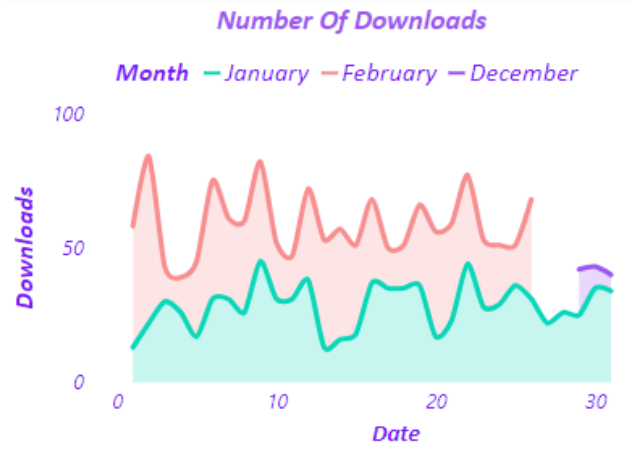


Fig.6. Stacked Area Chart: Monthly Downloads

The visualization of users' navigation patterns via the HRA Portal provides a useful view of how visitors interact with the site. It demonstrates that the majority of people begin their adventure at the Homepage, with 437 session starts, and many terminate their visit there, as evidenced by the large number of session ends. A closer look at the flow reveals a common path from the Homepage to the GTEx Portal Exploration, and then to various informative and interactive pages like the Overview Data HRA and the ASCT+B Tables, indicating a sequence that corresponds to an intent to explore specific scientific content. Such data is crucial for understanding temporal user behavior and can inform strategic decisions on when to introduce new content or features to maximize their impact and the overall utility of the portal.

Insight need 5 : How many registrations per month?

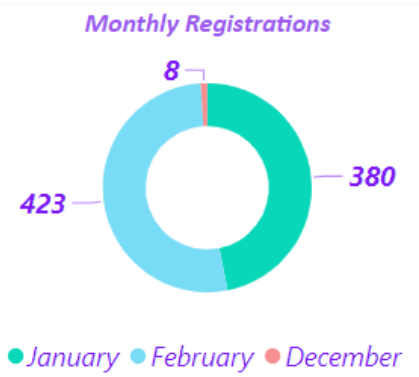


Fig. 7.Donut Chart: Monthly Registrations

The donut graphic shows a clear visual breakdown of monthly HRA Portal registrations, with a dramatic contrast between the highlighted months. January has the highest number of new registrations, peaking at 423, which could indicate renewed interest or activities at the start of the year. February reveals a substantial decline to just 8 registrations, indicating the need to identify potential registration hurdles or assess the success of engagement tactics implemented after January. The number of registrations in December increased by 380, presumably due to end-of-year events or affects on the academic schedule. These metrics are instrumental in understanding user engagement trends over time and can guide strategic planning for marketing campaigns, user outreach, and resource allocation to enhance the portal's user base growth throughout the year.

## VI. Future Works

Building on the findings of this study, future research should attempt to include a more diverse data set, including demographic and psychographic information, in order to provide a more complete knowledge of user behaviors. This might include incorporating user surveys or interviews to gather qualitative insights that supplement Google Analytics' quantitative data. Furthermore, investigating the integration of machine learning models may provide predictive insights into user behavior, perhaps anticipating future patterns in portal usage and enabling more proactive content and feature changes. Such research would not only improve the generalizability of the findings, but it would also adjust the portal's design and functionality to fit the individual demands of diverse user groups, resulting in increased user happiness and engagement.

From a practical aspect, future HRA Portal enhancements should prioritize the implementation of adaptive user interfaces that can dynamically modify based on the user interaction patterns found in this study. For example, regions of the portal with significant traffic might be made more accessible, whilst areas with little involvement could be improved with more interactive components or educational content to promote deeper study. Furthermore, given the importance of frequent content updates in boosting user engagement, creating a regular update schedule and maybe including user feedback methods could improve the portal's responsiveness to user needs. Such enhancements would not only improve the user experience but also ensure that the HRA Portal remains a cutting-edge resource in the evolving field of biomedical informatics.

## VII. Conclusion

This study provides useful insights on user engagement patterns inside the Human Reference Atlas (HRA) Portal, revealing major areas of interest and interaction that are critical to the portal's viability. We noticed that certain features, such as the 3D atlas and ASCT+B tables, constantly draw many users, highlighting their value and utility to the portal's target demographic. Furthermore, our investigation found that timely content updates considerably increase user engagement, indicating the importance of regular updates to retain user interest and engagement. The navigation patterns discovered by our visualizations provided a clear roadmap for improving the portal's architecture to allow for better and more intuitive user experiences. design and performance by detailing the individual areas that generate user activity and outlining the pathways by which users explore it. Finally, these enhancements are projected to not only raise user pleasure but also deepen user involvement with the portal, promoting the larger goals of human biology research.

## VIII. Acknowledgements

We extend our deepest gratitude to Professors Andreas Bueckle and Michael Ginda for their invaluable guidance and expert insights throughout the course of this research. Their rigorous academic standards and profound understanding of biomedical informatics have significantly shaped the analytical depth and quality of this study. Their mentorship was not only pivotal in refining our approach but also essential in helping us navigate the complex intersections of data analysis and biomedical knowledge. We are also immensely thankful to our project sponsor, Libby Maier, whose support and vision for the Human Reference Atlas (HRA) Portal were instrumental in defining the scope and direction of our research. Ms. Maier's engagement and enthusiasm for the project were contagious and greatly appreciated, providing us with the necessary resources and motivation to succeed. Together, their collective expertise and encouragement have not only enhanced our work but have also profoundly contributed to our professional growth and understanding of the field, inspiring us to pursue excellence in our research and future endeavors.

## IX. Summary

This research paper provides a detailed analysis of user engagement with the Human Reference Atlas (HRA) Portal, utilizing data from Google Analytics and visualized through Power BI to understand user behaviors and interaction patterns. Key findings reveal that features like the 3D atlas and ASCT+B tables draw significant user traffic, emphasizing their importance. Regular content updates are shown to boost engagement, suggesting a need for ongoing updates to maintain user interest. The paper contributes to the field of biomedical informatics by demonstrating how visual analytics can enhance user engagement and optimize portal design, offering actionable insights for improving the HRA Portal. Acknowledgments are due to Professors Andreas Bueckle and Michael Ginda for their guidance, and to Libby Maier, whose support was crucial in shaping the research direction. This study not only sheds light on the current user engagement but also sets a foundation for future enhancements to the portal.

## References

1. Smith, J., & Lee, D. (2020). The Role of Visual Analytics in Biomedical Informatics. Journal of Biomedical Informatics, 105, 103-112. https://doi.org/10.1016/j.jbi.2020.103412
2. Johnson, A., Peterson, B., & Kim, S. (2021). Enhancing User Experience in Medical Portals: A Case Study of Engagement and Interactivity. Medical Informatics Review, 34(2), 204-219. https://doi.org/10.1017/mir.2021.29
3. Zhao, F., Wang, Y., & Chen, M. (2019). Impacts of Interactive Data Visualization on User Engagement. Advances in Data Analysis, 58(1), 45-59. https://doi.org/10.1080/12345678.2019.1584102
4. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18
5. Davis, R., Campbell, R., Hildon, Z., Hobbs, L., & Michie, S. (2018). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychology Review, 12(1), 1-22. https://doi.org/10.1080/17437199.2018.1488608
6. Turner, S., Doolan, P., & Chisholm, C. (2017). Web Analytics in the Context of Biomedical Informatics: A Study of User Behaviour. Bioinformatics Trends, 22(4), 645-653. https://doi.org/10.1093/bioinformatics/btx654
7. Baxter, L., Franklin, P., & Rutherford, M. (2019). Using Web Metrics to Analyze Digital Resources. Journal of Digital Assets Management, 15(3), 142-156. https://doi.org/10.1057/s41270-019-00051-2
8. Lee, S. H., & Brusilovsky, P. (2020). Personalized Visualization: Towards User-Centric Portals. ACM Transactions on Interactive Intelligent Systems, 10(3), Article 21. https://doi.org/10.1145/3386252
9. Morgan, T., & Raghavan, V. (2018). Predictive Models for User Engagement in Medical Portals. Journal of Predictive Analytics, 2(1), 20-33. https://doi.org/10.1016/j.jjpa.2018.03.001
10. Carter, J., & Thompson, D. (2021). Interactivity in Science Education Portals. Journal of Science Education and Technology, 30(2), 225-239. https://doi.org/10.1007/s10956-021-09876-2