SmartHelm: Design and Implementation of Open Source based Behavioral Data Repository

Harish Moturu, Johannes Schering, Jorge Marx Gómez
Very Large Business Applications (VLBA),
University of Oldenburg,
Oldenburg, Germany.
emails: {harish.moturu, johannes.schering, jorge.marx.gomez}@uni-oldenburg.de

***Abstract*—Data is the driving force in the current world of digitalization for efficient decision making. Data availability plays a key role in almost every research field for improving data understanding and knowledge acquisition. There are many ways in which data can be made available to end users. In some cases, the user may be required to pay for using the data: similarly, there are some other data resources available as open source. In the case of sensitive data such as business related, personal and bio-physical data, there are certain challenges in making these kind of data publicly available. This paper addresses exclusively how an open data portal enables users to freely access bio-physical data that is developed as part of the SmartHelm research project. The open data portal *SmartHelm Behavioral Data Repository* constitutes of two key components: (1) a data management platform that organizes and structures the raw data into a well defined structural form and (2) a RestAPI server that has the functionality of providing the data in the desired format to the end user. The first prototypical version of the data portal constitutes of data obtained from experiments conducted in the project. These data sources are ElectroEncephaloGraphy (EEG), eye tracking and geographical location/ Global Positioning System (GPS) data. The main advantage of this data repository is that it facilitates the users to access all kinds of data without any special limitations such as user account, subscriptions etc. This bio-physical and mobility data will not only help other researchers working in this field, but also it will help municipalities to improve the bicycling infrastructure at the certain junctions in the city where the cargo bike logistic rider experiences more stress.**

***Index Terms*—bio-physical-data, EEG, Eye Tracking, data repository, behavioural data, bicycle data, mobility data, open data.**

I. INTRODUCTION

New bio-physical data may have a positive influence on different aspects of life. One topic that is often discussed is livability and sustainability of future cities [1] [2] . Knowledge about distraction factors in mobility applications could contribute to increase traffic safety and to adjust infrastructures to the demands of real users. A central problem here is that related data sets are often not available or at least not openly available on the Internet. In addition, the combination of behavioral data and mobility data is hard to find. One reason is that the behavioral data that is discussed in this contribution (EEG, Eye Tracking) is often gathered as part of scientific works under laboratory conditions. Another factor is that available mobility related data is often not openly available, but remains inside organizations because of data privacy or security concerns. To find a solution to this problem, the University of Oldenburg has developed a new data repository related to behavioral data. This includes an open data component to enable a complete open accessibility to the data sets, a visualization component to display the results in the form of geographic maps and Key Performance Indicators (KPIs) to make different data sets comparable. In this contribution, we focus on the aspect of the data portal. The development and the working steps of the implementation of the data management system (back-end) and the website (front-end) will be described in detail. The development and calculation of KPIs will be discussed in further publications. This contribution consists of the following individual sections. Section II is a survey of existing literature in the research field we are discussing in this contribution. It provides an overview about existing platforms related to behavioral data to identify the research gap. As we will see, existing solutions are not totally open, which means that the user has to register to get access to the data sources. Registration and reading user instructions takes time and could decrease user satisfaction. Existing website solutions do not seem to be very intuitive. It takes time to learn about functionalities and opportunities of the portals. In addition, mobility data in combination with behavioural data is not available so far. Mobility related distraction data could be a useful contribution to support practical and scientific applications. Section III gives an overview about the implementation of the behavioural data management system of the SmartHelm [23] research project. This includes the working steps of data integration, data processing, data visualization and publication. Section IV describes the functions and services of the behavioral data repository in detail. Section V are summarizing the preliminary results of the data management and data analysis in the SmartHelm project. The state of the art which is presented in the literature review section is compared with the functionalities of the behavioral data repository in Section VI. Section VII summarizes the results of the article and gives an outlook to the future development.

II. RELATED WORK

This next section describes the results of the literature survey carried out on the current state of the art of relevant platforms that provide human behavioural data. [3] lists out

 ISBN: 978-1-61208-962-1

different online based platforms for sharing and analyzing behavioural data. It describes in total 20 online platforms including data sharing, data visualization, offline or online analysis, and community cooperation. Among them, a few data repositories are short-listed in our work depending upon the type of data availability and relevance. Here, the factors considered for short-listing are if the type of data in the portals is bio-physical, sensory, mobility data and also the mode of accessibility of the data to the general public.

The ElectroEncephaloGraphy (EEG) / Event-Related potentials (ERP) based portal EEGbase [25] described in [4] - [6] is built as an open source web based data platform to store both raw and meta bio-physical data. This portal combines Semantic Web architectures such as Resource Description Framework (RDF) and Ontology Web Language (OWL) with common data structures. The developing methodology is based on the standalone relational data models. Although the main aim of the EEGbase portal is to provide open-source EEG/ERP data, users need to register and to create an account in order to access the data sources. This drawback is considered in our current SmartHelm Behavioural data repository, where end-users may directly access the data sources without any limitations. Another data platform referenced in our work is Neurobot [25] which is discussed in [7]. Neurobot is a web-based platform to publish neuroscience data for research and is developed in focus with Study data managers, Researchers, and Platform administrators. This platform does not only store and share the data, but also allows other researchers to upload data-sets and to contribute for further research. The key features of Neurobot [7] are data dictionary, user management, saved search, data query tool and API access. These additional features enhance the organization and increase the searching ability of the platform. Similar to EEGbase, the platform is not easily accessible because it requires an user account. Contributions [8] - [10] describe about the language independent object model Neuroscience Electrophysiology Objects (Neo). The model Neo [26] provides common representation of data acquired from electroencephalographic, intracellular or extracellular experimental recordings, or those that were generated from simulations. Unlike the platforms described in the previous papers, Neo is a standlone package developed as an open source software in Python programming language. The main functionality of Neo is to read and write different formats of electrophysiological data into a common representation, which would enhance interoperability and facilitate data-sharing. Although Neo is not exclusively a web or software based platform for data sharing, it indirectly acts as an open source component for building electrophysiological data sharing platforms. Another innovative EEG data sharing platform is IEEG-Portal [27] . It is a cloud based collaborative research platform developed for sharing, analyzing and visualizing data. [11] - [14] illustrate the methodology and technology behind the IEEG portal. One of the most significant features of IEEG portal is that it stores numerous large sized EEG data-sets based upon google web toolkit and Amazon S3 service. Due to the cloud feature, this

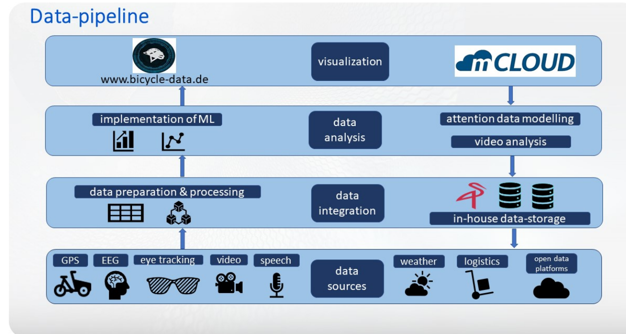

Fig. 1. Four layers architecture representing the data flow and their corresponding significance

platform facilitates the storage of data in terms of Terabytes. A disadvantage is mainly that the data privacy issue is not well explained. As the data is stored in the cloud, additional security and privacy measures need to be followed in order not to have any kind of data breach. Another drawback is that users need to have an individual account to access the data. Finally, the last paper in the literature review section is [15]. [15] it introduces an ontology based application Ontology of Heart Electrophysiology (OHE) portal for study of heart related biomedical activities. This ontology based tool provides a web based application for analyzing and to visualizing using the data acquired from heart electrophysiology data. This tool is an advancement in the bio-medicine sector as this type of ontology portals usually are not publicly available.

III. Methodology

The principle methodology to develop the SmartHelm open source based data portal is illustrated in this section. It explains the related data pipeline or data flow from data collection and processing to visualizing the results in various forms. Fig 1 shows the architecture behind the developed platform. In this section, each layer is explained in a separate paragraph.

In Fig 1, at the bottom is the source layer, where all relevant data sources are gathered and stored in the raw form in a project-owned storage server at the University of Oldenburg. The bio-physical data of the participants is critical from both ethical and legal perspectives and needs to be handled carefully. Due to data privacy, this database is not supposed to be stored in an external cloud server. Therefore, in the source layer, an open-source based simple storage solution (S3) client MinIO is deployed on the storage server to store the raw data on S3-based object storage. This MinIO client serves as a Data Lake to store the raw data. In principle, core data contains EEG sensors, eye-tracking, speech input commands, Global Positioning Systems (GPS) signals, and distraction marker data streams acquired from a group of subjects. Additional external mobility data sources for the data analysis are also integrated in the system depending upon the requirement. These include weather data [28], open street map and open bicycle data [29], historical accident data from the city of Oldenburg which details the various types of accidents happening at different crossings, road-sections and narrow bicycle lines in Oldenburg. This open bicycle data is part of an open data portal which contains a comprehensive set of bicycle data such as bicycle counting, parking and near accident data. The data warehouse and the related website were developed as part of a student project Bicycle Data. On top of the source layer is the second layer known as an integration layer, which is further described in the following section.

The Integration-layer serves the purpose of aggregating all the heterogeneous data sources and storing the data in a data management system. The process of aggregation of the data sources is implemented in a number of steps, as described in the following. The primary step is pre-processing the raw data sources of the source layer. Individual data sources are pre-processed to view type, format, variation and important features lying among the different data sources. The output from the pre-processing step is selecting the detailed information individually from all the available data sources. The next step is designing the schema for the data management system. For conception of the schema, a detailed state-of-the-art study was carried out. Among these are the methods for data warehousing design [16], recent trends in the conceptual schema design [17], and Extraction Transformation Loading (ETL) methodologies & [18] and [19] are considered as reference design architectures. Therefore, as a benchmark from the reference methodologies, in this second step, a schema design is developed by considering the time as the central dimensional parameter in the data warehouse. Based on the central design parameter described in the previous step, the third step transformation script is developed according to the reference ETL methodologies [18] [19]. The transformation step constitutes each form of the heterogeneous data according to the schema structure. The pre-processed data sources are transformed depending on the steps in a data pipeline. That means that each and every heterogeneous data source is extracted from the Data Lake and then transformed separately to fit into the schema. After the transformation is completed, each data source is uploaded to fit the individual tables in the schema. Therefore, the last step in the integration layer is to automate the above three steps in the loop to populate the data warehouse with all the available new data sources and to store the data that would be acquired in future experiments (in real traffic situation).

The data analysis layer can also be called the knowledge layer. From the variety of data stored in the data warehouse, different types of analysis can be performed such as attention modelling with EEG data, detection of Area of Interest (AOI) of the rider using the eye-tracking data and speech recognition system modelling. This layer serves the purpose of analysing the structured data stored in the Data Warehouse depending upon the goals of the research. Among these, some of the important methods are application of Machine Learning algorithms, implementation of Cross Industry Standard Process for Data Mining (CRISP-DM) [20], predictive analysis, and query based data analysis techniques. In the current paper, two types of analysis are illustrated. The first is the theoretical evaluation made on the feedback obtained after conducting interviews with experts and delivery professionals in cargo bike logistics and the second is analysing the data obtained from the experiments described above. As soon as attention related data from real traffic situations is available in a later stage of the SmartHelm project, the corresponding analysis with new Global Positioning Systems (GPS) routes will be carried out and published as a part of further publications.

The top most layer in the data flow is called the visualization layer. The results from our first visualizations are presented in Section 5. In general, this stage in the data flow and the graphical representation supports the understanding of the results and information obtained from the data analysis. The visualization layer, illustrated in Fig 2, consists of two functionalities a) publishing the raw data on an open data platform (i.e Data as a Service DaaS [21] [22]) and b) presentation of the results in various graphical forms on a dashboard. The implementation includes the development of a portal to download the raw data of the experiments and a dashboard for the visualization of the analysis results as part of the Bicycle Data website that was described earlier. To connect the data from the data warehouse to the respective visualization platform a Rest API service is developed.

IV. Functional Components of the Data Repository

The next section describes the implementation and functional components developed based on the methodology described in the previous section. It focuses on the individual aspects of the portal i.e what kind of data will be made available on the portal, then how the portal design looks like and finally explaining in detail what are the functionalities it provides to the users.

A. Overview of the available data in the Repository

ElectroEncephaloGraphy – data: As mentioned in the previous section, the repository mainly consists of bio-physical data. Among these, the first data type is ElectroEncephaloGraphy (EEG) data which is acquired from a set of participants as part of scientific studies. In principle, raw EEG data consists of brain signal values obtained from 7 electrodes in the form of impedance. The frequency of the impedance values is 250 Hz and the data is filtered in terms of the participant, data and session acquired. Beside impedance values from 7 electrodes, other parameters such as timestamps, session and location of experiment and participant related information are included in the raw data.

 ISBN: 978-1-61208-962-1

Eyetracking – data : The second type of the bio-physical data as part of the repository is eye tracking data. It consists of eye gaze values obtained from an Augmented Reality Glass (Microsoft HoloLens 2). In total, there are six eye gaze values calculated in x, y and z directions from the origin and the current position respectively. These eye-gaze values measure the 3 dimensional distance from the object to the participant viewing position, which further facilitates in estimation of the point of vision of the participant during the course of the experiment. In comparison with the EEG data, the frequency of data recording for eye-tracking data is less (around 30 Hz).

Location – data : During the experiment, the Global Positioning Systems (GPS) coordinates of the participant are logged while riding a bike using an open-source application known as OwnTracks [30]. The OwnTracks application logs the location information with various position based measurements namely altitude, latitude, longitude coordinates, speed of the rider, radius around the region, and including relevant metadata information. Here, the measurement frequency of Global Positioning Systems (GPS) data is significantly lower than the above two data types, with a rough estimate is 5Hz frequency. This location information along with other behavioral data is also made available on the open data platform for the end-user to download. In principle, location data enables the scope of mobility point of view on the data, which means behavioural data can be integrated and analyzed on the basis of location data.

Additional – data – sources : In addition to the above mentioned data sources, some another potentially relevant data sources are also prepared to be integrated in the data management platform. These are primarily mobility data such traffic volume data collected from various streets in the city of Oldenburg, which consists of counting information on the number of different types of transport vehicles crossing a particular traffic junction on an hourly basis. In addition to that, other types of behavioral data such as speech events, user interaction events and attention hot-spots data will also be integrated in the database.

B. Smarthelm API for data integration

In the methodology, integration layer describes the building of the data warehouse through integrating all the available data sources. The main functional component developed to access the data in the data warehouse is the API which is required to connect the data warehouse and the visualization component where the end-users have access to download and view the results. The API is developed based on the RestAPI technology using an open source Python web framework known as FastAPI. In contrast to other API framework technologies, one of the main advantages of FastAPI is the generation of automatic API docs depending upon the design of the download platform. The SmartHelm API doc shown in Fig 2 illustrates different kinds of accessible data parts represented as end-points. Here, each end-point indicates one data element which can be accessed to pull the specific data queried by the end-user. The SmartHelm API doc Interface(Fig 2) in general provides the information about all kinds of data elements that are available for integration. The key functional utility of SmartHelm-API underlies in linking the API for other external projects and can be used as an open source service to integrate the behavioural data repository with the back-end. For example, if an external project has the requirement of accessing only the EEG data, the API facilitates a gateway to pull the specific EEG data. This end-points format can be obtained from the API doc. Fig 2 clearly shows a sample end-point on how to obtain the EEG data for a time interval.

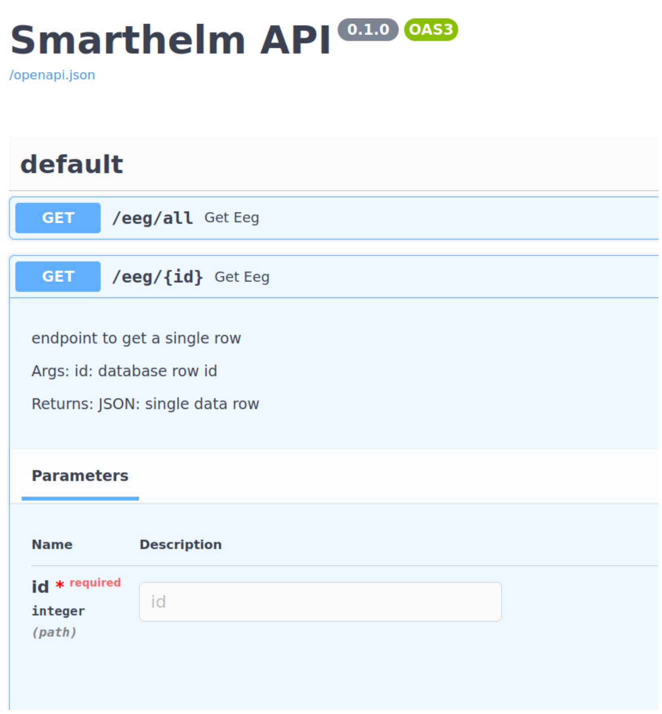

Fig. 2. An example from SmartAPI doc showing some of the available endpoints provided by the API where each endpoint serves the purpose of integrating the platform directly from the back-end

C. SmartHelm Data platform User Interface

The key concept behind the SmartHelm Data Platform User Interface (UI) is providing intuitive user experience throughout the portal. Fig 3 and Fig 4 illustrate a quick insight into the data portal. The design is made using a simplistic approach with a wider range focus and more intuitive for the general public. Along with the research community, other stakeholders such as politicians, municipality officials, city traffic planners, and the general public can easily understand what kinds of data are openly available to access and their functionality. In addition, the other driving force for developing the open source data platform is to enable data availability in the form

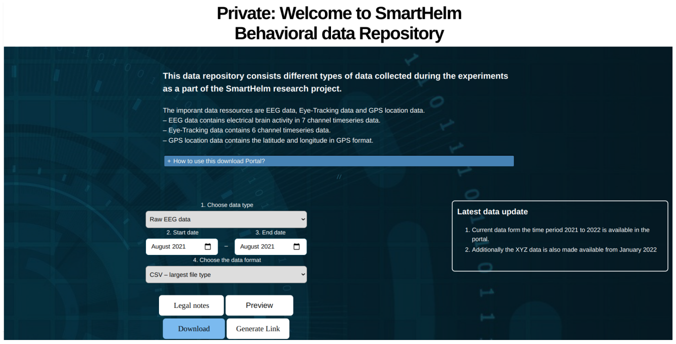

Fig. 3. This full size image from the front-end of SmartHelm Behavioural data Repository presents a complete overview of the portal which shows the description text and below is the data selection menu along with various functionality buttons.

Fig. 4. The Fig 4 displays functionality of various buttons below the selection menu such as generate link on the top and preview of the selected data to download.

of Data as a Service (DaaS). The UI design in Fig 3 reflects the sole purpose of the platform and the steps taken to ensure the goal of providing Data as a Service to the end-user. In the following, the content and usage of the UI are described in detail. The header menu explains briefly to the end-user which different types of data sources are currently available for download. After the header menu, a quick 'how-to' steps menu is integrated to make it more intuitive for the end-user to understand the various steps involved from the selection of the data until downloading the data in the desired data. Below 'how-to' steps menu is the data downloading section. In this section, different types of drop-down boxes are available for the user different categories such as type of data, selection of date interval, type of the data format to be downloaded and so on.

V. Challenges in the development

The next section describes the different findings during the development of the open source platform which constitutes a blend of data collected from sensors and mobility systems.

The previous sections described technical difficulties for the implementation. Various challenges regarding data privacy and security problems have to be addressed before the technical implementation of the above represented high-level system architecture can take place. In this section, all the necessary aspects related to ethical and legal issues are described, which are taken into consideration before modelling the system. Ethical and legal issues play an vital role in the SmartHelm project. Permanent surveillance of the participants during experiments is not the intention of the project and needs to be avoided to increase the acceptability of the riders in the future. There is no kind of real-time data transfer during the experiments which guarantees that misuse by the intermediate user of the gathered bio-physical data (EEG, Eye-Tracking, etc.) can be ruled out. In the methodology section, we explained how the raw data is stored in the grass root level using the S3-storage(Data-Lake) technique, but before that step the data transfer from data collection to landing into our Data Lake is also a challenging step. Here, certain measures are evident for a secure data transfer between the two components. To safeguard this critical problem, dedicated user policies are created on the Data Lake server to facilitate the data collection supervisors to upload the data directly on the S3-storage-server, which bridges the data transfer gap without any external channel. Here, the policies are confined in such a way that the users who are uploading the data into the server have minimal access to the server, which means they do not have rights to delete or edit the existing data, but only upload new data sets. The next challenging factor specifically about the data is time synchronization between different data streams. As explained, the data is heterogeneous data and obtained from different sensors and mobile applications. There does not exist by default by an synchronization in time-steps for all kinds of data-streams. Maintaining an uniformity is a crucial factor for further processing.

VI. Advantages in comparison with other portals

First and foremost, the main advantage of SmartHelm behavioural data repository is the unique characteristics in comparison with the other bio-physical data platforms. From the Section 2 through literature study, it is evident that compatibility, end-user serviceability and degree of simplicity of the older bio-physical data platforms are considerably behind in some key aspects than the SmartHelm data portal. In this context, primarily comes user accessibility to the platform in SmartHelm data repository. All external visitors do not essentially require any kind of user-agreement before downloading the data as well as a separate user account is not needed to be registered on the portal. These is one among the two key entities which differentiate the SmartHelm download portal from other human bio-physical data portals. In addition, a few other components which enhance the portals' wider scope of application are as follows. First comes the degree of intuitiveness for the user, which is explained in the previous section. Unlike other portals, the required information for the end-user is served without any heterogeneity. The next factor would be the user-acceptance, which means, to increase the user acceptance, the portal is designed in such a way that the end user has the complete control to be served by themselves by interacting with the portal. For example, the advantage of choosing the specific time interval in order to access the data represents the kind of options that increase the probability of user-acceptance. To achieve this, depending upon on the requirement, a user survey will also be conducted and the feedback will be considered to enhance the usability of the portal. The third additional factor would be the preview of the data before download. Although it is not a standalone functionality or a modern plugin, this kind of functionality is not found in other portals and it facilitates the user to obtain an overview of the data.

VII. Conclusion

The availability of mobility data integrated with behavioural data is a rare combination and very few research works acquire this kind of data. In addition, accessing such kind of sensible data through an open source medium would be more versatile. The SmartHelm behavioral data repository introduced in this work scientifically illustrated an open source platform consisting of an unique combination of mobility and bio-physical data. The problem statement i.e., the research gap described in the introduction section is that there exist few platforms which provide bio-physical data to the public, but still there are certain demerits which are clearly explained and the alternative solution is provided with the data portal developed in this work. The behavioural and mobility data integrated will be made accessible to the general public through the SmartHelm data portal as well as the mCLOUD / Mobilithek portal, which is provided by the German Ministry of Transport and Digital Infrastructure (BMDV). Moreover, some other benefits include the fact that city infrastructure planning of municipalities can be supported by these data sets during the improvement of road infrastructures or bike lanes. In the context to the cargo bike riders, data-driven decision-making probably increases work efficiency. Therefore, as an end-note to brief the further steps in the project, a visualization service is well planned in order to present the data analysis results in various graphical, visual forms such as heat-maps, digital city map, etc., which fulfills the main goal of the project to assist cargo bike riders by reducing the mental and visual stress in everyday parcel delivery. From a wide range perspective, this open accessible data to the public and publishing the results obtained from the overall project will, in return, encourage other major cities and municipalities in Germany to switch cargo bike logistics, which would become one of the major contributions to green logistics.

Acknowledgments

SmartHelm is funded by the German Federal Ministry of Transport and Digital Infrastructure (BMDV) as part of the mFUND program (project number 19F2105B) with a funding amount of around 1.48 Mio. Euro. As part of mFUND, the BMDV supports research & development projects in the field of data based and digital mobility innovations. Part of the project funding is the promotion of networking between the stakeholders in politics, business, administration and research as well as the publication of open data on the mCLOUD portal (which will be superseded by Mobilithek in 2022/23).

References

[1] E. Taniguchi, ”Concepts of city logistics for sustainable and liveable cities.” Procedia-Social and Behavioral Sciences no. 151, 2014, pp 310-317

[2] E. Taniguchi, & R. G. Tompson (eds.), ”City Logistics 3: Towards Sustainable and Liveable Cities.” John Wiley & Sons, 2018.

[3] Y. Chen, W. Zhong-yi, Y. Gang, and H. Lan, ”An overview of online based platforms for sharing and analyzing electrophysiology data from big data perspective.” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7, no. 4 2017: e1206.

[4] P. Ježek, and M. Roman, ”EEG/ERP Portal–Semantic Web Extension: Generating Ontology from Object Oriented Model.” In 2010 Second WRI Global Congress on Intelligent Systems, vol. 3, pp. 392-395. IEEE, 2010.

[5] P. Ježek, and M Roman. ”Database of EEG/ERP experiments.” In HEALTHINF 2010-Proceedings of the Third International Conference on Health Informatics, pp. 222-227, 2010.

[6] P Brüha, and M Roman. ”Portal for research in electrophysiology—Data integration with neuroscience information framework.” In 2012 5th International Conference on BioMedical Engineering and Informatics, pp. 1099-1103. IEEE, 2012.

[7] R. Badia et al., ”INCF Program on Standards for data sharing: new perspectives on workflows and data management for the analysis of electrophysiological data.” In Techn. Report https://www.incf. org/about-us/history/incf-scientific-workshops, 2015.

[8] S. Garcia et al., ”Neo: an object model for handling electrophysiology data in multiple formats.” Frontiers in neuroinformatics 8 2014: 10.

[9] J. Grewe, T. Wachtler, and J. Benda, ”A bottom-up approach to data annotation in neurophysiology.” Frontiers in Neuroinformatics, 5 2011: 16.

[10] A. Sobolev et al., ”Integrated platform and API for electrophysiological data.” Frontiers in neuroinformatics 8 2014: p.32.

[11] JB. Wagenaar et al., ”A multimodal platform for cloud-based collaborative research.” In 2013 6th international IEEE/EMBS conference on neural engineering (NER), pp. 1386-1389. IEEE, 2013.

[12] L G. Kini, A. Davis Kathryn, and B. Wagenaar Joost, ”Data integration: Combined imaging and electrophysiology data in the cloud.” Neuroimage, 124 2016: pp. 1175-1181.

[13] M R. Bower et al., ”Metadata and annotations for multi-scale electrophysiological data.” In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2811-2814. IEEE, 2009.

[14] B H.Brinkmann et al., ”Multiscale electrophysiology format: an open-source electrophysiology format using data compression, encryption, and cyclic redundancy check.” In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 7083-7086. IEEE, 2009.

[15] B. Gonçalves, Z. Veruska, G. Giancarlo, and G G.P.F. José, ”An ontology-based application in heart electrophysiology: Representation, reasoning and visualization on the web.” In Proceedings of the 2009 ACM symposium on Applied Computing, pp. 816-820. 2009.

[16] A. Bauer, and G. Holger. Data-Warehouse-System: Architecture, Development, Application, dpunkt. verlag, 2013.

[17] P. Vassiliadis, A. Simitsis, L. Liu, and M. Özsu, Encyclopedia of Database Systems. New York: Springer, https://doi.org/10.1007/978-1-4899-7993-3 2009, pp.1095-1101.

[18] P. Digra, P. Abrol, and P. Lehana, Design and Development Of Distributed Image Acquisition and Recording System for Network Based Appli-cations. In International Journal of Scientific and Technical Advancements 4, no. 3, 2018, pp.1-8.

[19] A. Dearmer, Complete Guide to Database Schema Design https://www.xplenty.com/blog/complete-guide-to-database-schema-design-guide/ Published:16th of July 2021. Last Access: 29th of November 2021.

[20] R Wirth, and H. Jochen, "CRISP-DM: Towards a standard process model for data mining." In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, vol. 1, pp. 29-40. 2000.

[21] Z. Zheng, Z. Jieming, and R. Lyu. Michael. "Service-generated big data and big data-as-a-service: an overview." In 2013 IEEE international congress on Big Data, pp. 403-410. IEEE, 2013.

[22] X Wang, T. Yang Laurence, L Huazhong, and M. Jamal Deen. "A big data-as-a-service framework: State-of-the-art and perspectives." IEEE Transactions on Big Data 4, no. 3 2017: pp. 325-340.

[23] Research project SmartHelm official website, funded by German Federal Ministry for Digital and Transportation; 2022; accessed on 15.March. 2022; website url = https://smart-helm.com/

[24] EEGbase, RRID:nif-0000-08190;year 2022; accessed on 12.March.2022; URL = http://eegdatabase.kiv.zcu.cz/home-page;jsessionid=10r3hyplemadg?0

[25] International Neuroinformatics Coordinating Facility (INCF); 2022; accessed on February 2022; URL = https://www.incf.org/form/neurobot-user-account

[26] Neo, 2010-2022, accessed February 2022; URL = https://neo.readthedocs.io/en/stable/io.html

[27] IEEG portal, 2022; accessed on February 2022; URL = https://www.ieeg.org/

[28] Visual Crossing Corporation, Visual crossing weather API; year 2022; accessed on March 2022; URl= https://www.visualcrossing.com/weather-api

[29] Project Group Bicyle Data, VLBA, University of Oldeburg; year 2021; accessed on February 2022; URL= https://www.bicycle-data.de/bicycles-data/

[30] Owntracks MQTT broker application, year 2022, accessed on Febraury 2022, URL= https://owntracks.org/booklet/tech/json/