Enabling Participatory Flood Monitoring Through Cloud Services

Diego Pajarito Grajales
University of Glasgow
diego.pajaritograjales@glasgow.ac.uk

Lívia Castro Degrossi
Fundação Getulio Vargas – São Paulo
livia.degrossi@fgv.br

Daniel D. R. Barros
Federal University of ABC
daniel.barros@ufabc.edu.br

Mohammed Rizwan Khan
University of Heidelberg rizwan@uni-heidelberg.de

Fernanda Lima E. Silva
Fundação Getulio Vargas – São Paulo
fernanda.silva@fgv.br

Maria Alexandra Cunha
Fundação Getulio Vargas – São Paulo
alexandra.cunha@fgv.br

Rachel Trajber
National Center of Monitoring & Alerts of Natural Disasters CEMADEN
rachel.trajber@cemaden.gov.br

João Porto de Albuquerque
University of Glasgow
Joao.Porto@glasgow.ac.uk

ABSTRACT

Flooding events are more impactful due to climate change, while traditional top-down approaches to flood management give way to new initiatives that consider citizens and communities as active strategic actors. Researchers and practitioners have started to place communities in the centre of creation processes or invite them to co-design digital platforms. However, many citizen science projects re-use well-known technological components without reflecting about how the technology is able to effectively support citizen participation in data generation, including the provision of flexible data storage and exchange. This paper describes a novel digital platform design which adopts cloud services to integrate official and citizen-generated data about urban flooding. It summarises the results of a participatory design process of a digital platform to collect, store and exchange flood-related data, which includes components such as data lakes, Application Programming Interfaces (APIs), and web and mobile interfaces. This work in progress paper presents insights and lessons learned from using cloud services to enable citizen participation and engage communities with flood monitoring.

Keywords

Flood monitoring, communities, citizen-generated data, cloud services, citizen participation.

1. INTRODUCTION

Among the multiple consequences of climate change, flooding events now show higher frequency, intensity and overall impact. Communities worldwide increasingly face flooding, but vulnerable communities, people living in informal settlements, are mostly affected by socio-economic impact and human losses (Dottori et al., 2018; Williams et al., 2018). Official authorities have traditionally carried out flood management, monitoring and data production. However, the advances in information and communication technology have increased the involvement of citizens in flood management, turning them into strategic actors capable of working to improve monitoring and response schemas (e.g., joining local governance schemas, building adaptation capacity and enhancing flood awareness). As a result, communities are increasingly placed in the centre of design processes and invited more often to co-design and co-produce information systems and high-resolution data.

In addition to community engagement in flood management, data and user interaction technologies have evolved significantly in the last two decades. Also, citizen participation is encouraged and technically possible through technologies based on cloud services that simplify data storage and exchange across multiple formats and structures. This work in progress attempts to provide insights and lessons learned from the exploration of using cloud technologies in citizen science and processing of citizen-generated data and official data sources for flood monitoring and data generation.

In the pursuit of this goal, the team developed software components to ingest, store and publish flood-related data produced by citizens and authorities. These components are needed to manage multiple data formats and schemas, so the proposed platform builds upon cloud services to manage, clean, transform and integrate data. The architectural design considered the efforts to engage citizens and practitioners with information technology as a tool to reduce the impact of flooding in Sao Paulo and Acre in Brazil carried out as part of the international research project "Waterproofing Data: Engaging Stakeholders in Sustainable Flood Risk Governance for Urban Resilience". Additionally, the architecture included user interfaces (e.g., the mobile application “Dados a prova d’água” and a web portal).

This paper contributes to increasing the knowledge about the integration of cloud-based technologies in citizen science, particularly by reviewing the benefits of cloud services to integrate data from multiple sources and reduce the time needed for development and deployment of data-driven platforms. The results describe the architecture to process and integrate citizen-generated with the official flood-related data (e.g., areas susceptible to flooding and historical rainfall records). The relevance of this work stands in the reduced time to develop and deploy information platforms for citizen science projects and the simplified management of the infrastructure involved. Reduced times and simplified processes can help communities replicate the approach and positively impact citizen participation and citizen-driven data production.

The paper is structured as follows: Section 2 depicts the related work on citizen-generated data production and current data services available from cloud providers. In Section 3, the participatory design process applied in the project is described which has driven the architectural design. Section 4 discusses how cloud services can support the requirements while implementing data flows to connect the endpoints (i.e., data sources and user interfaces). Finally, the paper describes insights and lessons learned from the results obtained, which were achieved with the help of citizens from São Paulo and Acre in Brazil, whilst also discusses the contributions to the fields of information technologies and citizen science and highlights directions for future work.

2. RELATED WORK

Our project stands between two complementary research fields. First is the engagement of communities and citizens in scientific duties, usually named citizen science or citizen-generated data production. Second is the adoption of cloud computing for data integration, especially the architectural design based on cloud services to enable and optimise data pipelines. Although both fields have been present in the academic literature in the last decade, we combine them to strengthen citizen science. For example, researchers find novelty in establishing a cloud-based architecture to complement official flood-related data with citizen-generated data.

2.1. Citizen Science and Citizen-Generated Data Production

Citizen science, defined as the participation of people from outside professional organisations in data collection and analysis, is an important and growing trend in scientific practice (Bonney et al., 2009; Phillips et al., 2021) and disaster risk management (Marchezini et al., 2018). This participatory form of carrying out scientific activities has great potential to advance knowledge, generate a shared understanding of hazardous phenomena, provide early warning of hazards, and help communities at risk to take actions to improve their resilience during and after disasters (Hicks et al., 2019).

Such participatory science has also been used to provide valuable data in flood risk management. For instance, citizen science is being used to fill the data gaps and increase understanding and assessment of flood hazards at the local level, particularly in areas with scarce measurements of past events (Sy et al., 2020). In fact, flood is a privileged theme in citizen science initiatives globally (Chari et al., 2019; Marchezini et al., 2018).

2.2. Cloud Services for Data Management, Storage and Exchange

In the last decade, cloud computing or cloud services have called scientists' attention when working in data management, data analytics, and computer sciences (Kobusińska et al., 2018; Wu et al., 2020; Yang et al., 2017). Among other research fields, data integration and automated development and operation DevOps have expanded to fields within academia (Munappy et al., 2020). Moreover, current architectures for information technologies and data-driven applications might take cloud services for granted. However, strategies such as citizen-generated data production have not fully embraced the cloud (Sturm et al., 2018).

Existing work around data pipelines (Roman et al., 2021), data integration (Dong and Rekatsinas, 2018), automated development and operation, known as DevOps, and automated data production and operation, known as DataOps (Munappy et al., 2020), provide insights into the benefits for optimising data management in community-led strategies and integrating heterogeneous data sources. For example, automated data integration worked for structured data sources but streamlining it to semi-structured and non-structured data sources matches the different scenarios in which communities and citizens struggle to make sense of data (Phillips et al., 2021; Wu et al., 2020). These potential benefits make cloud services appealing to official authorities regarding data management, storage and exchange since they can operate decoupled from components but still make integration possible (Chari et al., 2019).

From the related work reviewed, the authors see an opportunity to build upon the existing knowledge in citizen science and information technologies. Considering citizen-generated data a new high-resolution source capable of complementing official flood-related records is a novel approach worth analysing (Sy et al., 2020). Researchers have multiple opportunities to develop this work further based on the existing progress going from the technical development and the different component arrangements (Liu et al., 2011; O'Grady et al., 2016) to the validation methodologies needed to ensure data quality and integrity (Balázs et al., 2021; Kelling et al., 2015). Our work in progress innovates in designing a hybrid platform for data pipelines that integrate citizen generated data into an existing analysis platform of flood-related data. These contributions emerged from this work in progress, and, to the best of the team's knowledge, they have not yet presented in the literature.

3. DATA AND METHODS

This section describes conceptual and methodological building blocks used to integrate citizen participation and citizen science with flood monitoring. This approach aligns with a general aim of integrating citizens and citizen generated data in social processes and decision making (Conrad and Hilchey, 2011; Eleta et al., 2019). We, therefore, present in detail the participatory design carried out to define user needs and how these were translated into data flows. Later, we present the technology and cloud services considered as well as the proposed architecture to enable participatory flood monitoring.

Considering that the case study of flood monitoring is well-grounded in data management practices and a good testbed to combine data-driven and dialogical research methods. The platform's underlying concept was "data flow" due to the idea of moving data between stakeholders and improving decision making (Matheus et al., 2020). To define the most convenient arrangement of components (i.e., the architecture), the research team engaged stakeholders from the community (e.g., teachers, students, and all those living in the neighbourhood) to integrate their perspectives of the flooding events, record their perceptions through a platform and complement authoritative data sets.

3.1. Participatory Design Process

The participatory design process, particularly the one leading to designing the mobile app "Dados a prova d'água" (Waterproofing Data), was inspired by the critical pedagogy proposed by Paulo Freire in which educators and learners – in our case, researchers and citizens – collectively discuss and problematise citizens' reality (Freire, 2000). This approach has been commonly used to promote learning and develop critical awareness during the educational/research process, but in our project, we adopted the principles of critical pedagogy for citizen science activities (Porto de Albuquerque and Albino de Almeida, 2020). Following this, the platform design is the result of an underlying development of community engagement activities in the study area. These activities allowed researchers to understand residents' worldviews, how they perceive, understand the research topic (i.e., flood events) and their demands.

After these engagement activities, the team conducted a focus group with the residents to elicit the requirements of the mobile app. The focus group aimed at bringing together scientists, developers and community residents to discuss the existence (or lack) of flood-related data from the study area and the way the app could enhance data production and circulation as a pathway for transformations to improve resilience and sustainability (Porto de Albuquerque et al., 2021). After some time, the participants were able to reflect about their reality and develop a critical consciousness about their present situation regarding the lack of flood data. Later, groups of participants discussed problem-solving alternatives and suggested different data and functionalities they would like to have in a mobile app. Researchers then asked (provocative) questions to instigate the participants. Finally, the participants presented the outcomes of their discussions. As a result, the focus group activity based on the residents' reality contributed to the increasing of participants' critical awareness of flood events, engendered interest in the mobile application.

3.2. Data Flows and Integration

Data flows are considered one of the platform's building blocks (Rawat and Narain, 2019) since they allow data to move from the communities to the authorities and complement the existing top-down information flow (Matheus et al., 2020). The research team worked with the communities, schools and practitioners to identify data sources to describe flooding in two cities of Brazil. The communities produced flood-related data and complement authoritative sources through six different data flows. The first two data flows emerged from a strategy to produce self-made rainfall gauges with plastic bottles or pluviometers (named pluvipets) and deploy them across five country regions (i.e., the states of Santa Catarina, Acre, São Paulo, Mato Grosso and Pernambuco). Trained school teachers and students reported daily the rainfall measures registered with their self-made pluviometers. The next two data flows allow citizens to share perceptions of rain and flooding events with a precise location, description and photos. The last two data flows capture past flood memories and perceptions from citizens.

These citizen-oriented data flows, listed in Table 1, would provide a different view of flooding and complement the existing monitoring networks while engaging communities with flood monitoring. Additionally, three complementary flows integrate authoritative or traditional flood data: Official rainfall monitoring through the official rainfall gauge network, hydrogeological delimitation of susceptible areas based on physical variables and geological studies and, future development will integrate weather forecasts.

Table 1. Identified data flows

Source	Flow	Description
Citizen Generated	Pluviometer registration	A citizen reports a self-made pluviometer. It includes the location and description.
	Rainfall Measure	A citizen reports daily rainfall measures from a self-made pluviometer. All recorded values are available to view and update.
	Rain Event	Any citizen can report a rain event. It includes location, description and sometimes images.
	Flooding Event	Any citizen can report a flooding event. It includes location, the transit condition, description and sometimes images.
	Flood Memories	Any citizen can report a flood memory described in their own way. It can be textual information about the event and contain images or videos of the event.
	Flood Perception Map	Any citizen can report their perception of an area during flooding classified into Overflow, Elements at Risk and Overall Areas
Authoritative	Rainfall records	Authorities running monitoring networks provide periodic rainfall measures. Citizen generated data will complement official records.
	Flooding areas	A hydrogeological study provides areas susceptible to flooding and landslides. The areas provide a context for citizen reports and official rainfall records.
	Weather forecast	***To be developed***. Official and private providers generate short-term weather forecasts to notify citizens.

The team also carried out a four-month testing process to assess the functionality of the app and allow them to provide feedback on its usability. The Centre for Studies in Public Administration and Government (CEAPG/FGV) and CEMADEN Education Program at the Brazilian National Early Warning and Monitoring Centre for Natural Disasters/Ministry of Science, Technology and Innovation organised the participants. Partners run the activity under the name "WPD++ Waterproofing data citizen science: pollination of a mobile app in communities". Testing activities involved 21 'pollinators' (i.e., trained champions in charge of recruiting and monitoring the activities) and 200 participants from schools and civil defences in municipalities in five country regions.

3.3. Software Components

The team evaluated the related work in the field of citizen science, the main cloud service providers (e.g., Microsoft Azure, Amazon Web Services, Google Cloud) as well as consolidated web architectures to identify alternatives to building up a data platform able to support data flows. Based on the evaluation, the team started to group components to support functionalities and interactions presented in Figure 1. Firstly, clients' support for interaction required ingestion and query of flood-related data. To connect clients (e.g., mobile and web applications) and official endpoints (i.e., other official providers), the platform's back end is needed to ensure citizens and authorities can store their data properly and access it after integration. Also, all data assets must be stored safely and made available despite their format, structure or source. To optimise storage, the team combined a PostgreSQL relational database for quick execution of predefined queries and non-SQL storage to handle ingestion of heterogeneous data and records.

Data ingestion relies on a server-less pool of functions (i.e., MS Azure functions) that handle requests, data processing pipelines and dynamic loads. These functions store ingested JSON files into a data lake (i.e., MS data lake storage) and then process them to feed the PostgreSQL relational metadata-oriented model. Data query relies on a Node.js web API that executes dynamic queries in PostgreSQL to prepare small data sets for mobile and web clients. The team combined a server-based Node.js instance with a scalable azure functions pool as shown in Figure 1 to achieve the tight times for development. On the one hand, the combination optimised the capabilities of the development team to quickly deploy a Query API to interact with client requests and, on the other hand, deploy an interface to ingest data directly to the data lake.

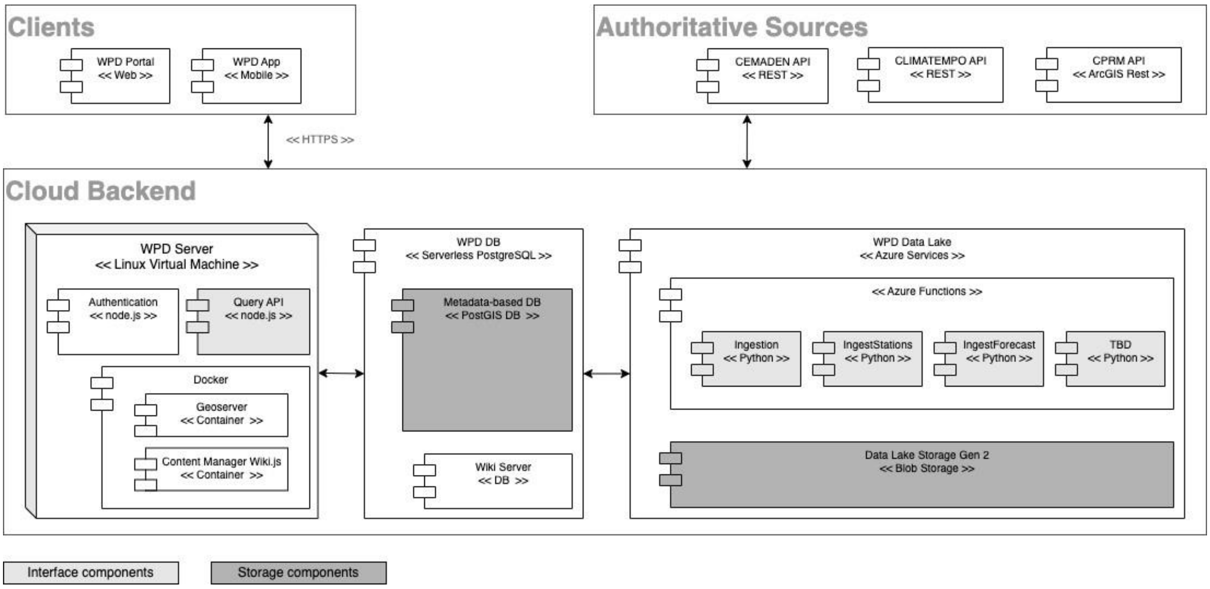

Figure 1 Proposed platform architecture

To integrate citizen participation and flood monitoring and, in general, integrate citizen-generated data decision making, the approach considered the participatory design to select components and cloud services aiming at enabling participatory flood monitoring and supporting the identified functionalities. Using the underlying concept of "data flow" researchers worked on designing a data platform to engage stakeholders with data production (e.g., by recording rainfall measures, reporting flooding events, memories, and perceptions) to enhance official data and provide a comprehensive vision together with vulnerable communities.

4. RESULTS

The results report on the technical process of implementing pipelines using cloud services (i.e., Microsoft Azure's python-based ingestion functions), data lake storage and a metadata-oriented database to ingest and process data. Additionally, results describe how client interfaces connect the pipeline endpoints and access processed data through dynamic SQL queries. There is a special emphasis on the role of the mobile app and the collaborative design used to define functionalities and user needs. Finally, this section compiles the work in progress results on how cloud technologies, especially the proposed hybrid architecture (i.e., cloud and serverbased), help integrate citizen participation and flood monitoring.

4.1. Data Pipelines

Data pipelines are considered as implemented data flows within a platform integrating interfaces or APIs dedicated to data ingestion, query and publication, together with other authentication functionalities. Table 2 depicts the pipeline's definition, endpoints, and interfaces. In this context, the team implemented the pipelines through Microsoft Azure functions that receive data in JSON formats and store it in the data lake. Likewise, these functions trigger validation and processing flows to prepare data for mobile and web interfaces. Design choices were about a cloud provider that supported integration, authentication and transformation services and allowed the team to focus on optimising data flows.

Data pipelines served for two main tasks as explained in Table 2: data integration and data access. For data integration, pipelines ingest based on user requests (i.e., when a mobile app user reports a new event) or periodic updates for official sources (i.e., regular updates for rainfall stations and on-demand for susceptibility areas). After, the pipeline quickly checked format and saved into the data lake, integration tasks later extracted key features to dispose data through the API. Data access operates via on-demand requests and notification updates while the clients call functions that execute parametric queries and start a subscription (i.e., initialising web sockets) for insert and update events. One complementary API controls authentication for the mobile and web interfaces.

Endpoint	Pipelines	Accessed via
Azure functions	Handle requests and JSON files send through the mobile app. The function saves all the files in the data lake and triggers data validation for data flows: *pluviometer registration*, *rainfall measure*, *flood event*, *rain event*.	Mobile / Web
Azure Pipelines / Azure functions	Handle automated requests and ingestion of official rainfall measures. They capture data from stations in two states in Brazil so far.	Data Lake
	Ingestion of susceptibility to flooding areas was not automated. Ingestion of weather forecast not developed yet.	
Node JS API	A Query API handles requests for citizen-generated and official data flows: *registered pluviometers*, *rainfall measures*, *flood events*, *rain events*, *susceptibility to flooding areas*.	Mobile / Web
	A second API allows the web portal to save flood memories and query those memories and the flood perception map	Web
Auth API	Handles user authentication across the platform. It does not include authentication for azure functions	Mobile / Web

4.2. Mobile Application

"Dados à Prova d'Água" is a mobile app serving as an interface of the Waterproofing Data project. It aims to enable the generation and circulation of flood-related data between communities and government agencies. The app was co-designed with communities from two cities in Brazil affected by floods and materialised a subset of the requirements identified. The app allows citizens to register their gauges and record periodic measures. Considering that most citizens do not have access to this kind of instrument or know how to use it; thus, they can use the app to provide the intensity of rainfall events instead.

The app also serves to make flood events visible to local authorities and other citizens. Therefore, citizens can provide data regarding flooded areas in the app. Another type of data that is often lacking is the water level in the riverbed and now citizens can provide this data through the app. The above-mentioned considerations were the foundations of the data flows, such as rainfall measure, rain and flood events and others listed in Table 1. Additionally, most participants did not know where and how to access official flood data; thus, they suggested including two official data sources: data from authoritative rainfall gauges and data regarding flood prone areas. By having access to both data sources, citizens would know which areas they should avoid in the case of a heavy rainfall event and look for alternative routes to go, for instance, from work to home.

Finally, users can combine the various sources and use the app as a "situation room". Moreover, citizens can access data based on user location and pre-defined parameters due to interface connecting the app and the query API. The final design comes out from the co-designed process and the later validation with the communities.
Figure 2 shows how the visual components present data and allow data to report through the mobile app interface.

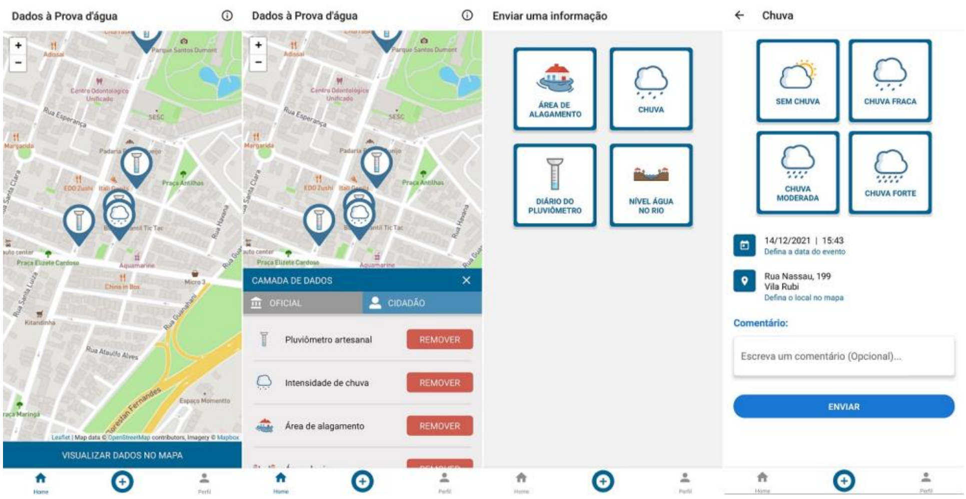

Figure 2 Mobile application user interface

4.3. Web Portal

The team developed a web portal as a second user interface to offer a complementary view of the flood-related data generated in the project as seen in Figure 3. With a single view of different data types, the web portal provides researchers and authorities with a better understanding of data and the gaps in knowledge. In addition, sharing citizen-generated experiences and authoritative data is relevant to understanding the role of data integration and community engagement within the project. However, the team faced additional difficulties in visualising different ranges, categories or points in time for the data flows listed in Table 1 within a geographic context (i.e., every data set became a geospatial data layer). The classification of different data sets based on their source (i.e., Official or Citizen-generated) controlled through a map-based application allows officers to explore and understand the integrated data in tandem with their crisis management tasks.

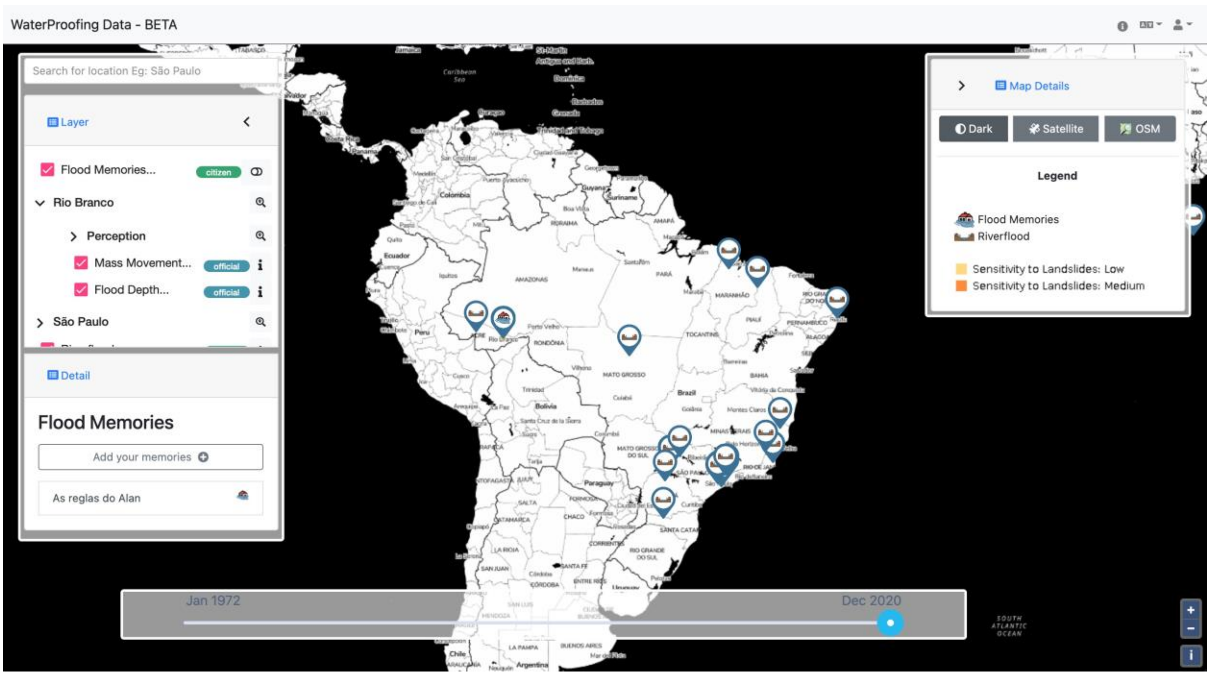

Figure 3 Web portal user Interface

5. DISCUSSION

To develop the project, the research team adopted an iterative agile development process that considered tasks for identifying needs and functional requirements together with the community members, data flow modelling, and selecting suitable components to develop and adapt. The process served to start a dialogue between citizens and stakeholders involved in flood management. This piece of research stands as an initial step to improve flood monitoring through cloud services while the proposed architecture helps to enable flood data production. Additionally, the two interfaces offered alternatives for visualising and analysing flood-related data. These are still early outcomes and a first step towards achieving improvements in flood monitoring. The adoption of cloud services made the case study rich in experiences to enable flood data collection and visualisation. Unlike traditional citizen science initiatives, this project aimed to incentivise citizens' participation through data generation in two moments. The first one – analysed here – is on the co-development of an interface and establishing general design principles whereas the second one is the data collection process. The team continuously reviewed the design principles emerging from the joint work of communities and researchers. This interactionresulted into an interface design and a set of requirements for data integration. The data flows listed in Table 1 were implemented through "data pipelines", whereas the lake storage component played a role in data ingestion, storage, and transformation. Technical advisors and researchers agreed on the data lake's potential to drive new research around ingesting citizen-generated and improving formatting or data cleaning tasks. Initial benefits achieved were related to the time saved in ingesting and structuring data. Another interesting idea to develop further is the combination of the data lake with a metadata-oriented database to manage storage, cleaning and processing tasks, simplifying data queries.

The team also found practical benefits from adopting the hybrid architecture that can be grouped into coding and combining multiple languages and frameworks. First, the time to develop the platform was shorter when ingesting multiple data formats despite their complex structure. Second, the platform is flexible by design as it integrates various programming languages and frameworks. It is possible to streamline the integration in the future of visual and codeless languages. This flexibility is common in different cloud providers (i.e., MS Azure) and from the citizen science perspective, it can lead to development teams working for components based on their skills.

The described benefits rise considerations for sustainability, maintenance and migration. However, balancing the pros and cons of novel and agile development approaches is also an interesting aspect investigated within this knowledge area. Projects adopting hybrid architectures have faced comparable results as the works of Wu et al. (2020), Liu et al. (2011) and Kelling et al. (2015). Since the final solution adopted a hybrid approach that combines server and server-less components, the team needed stronger coordination and more accurate documentation. In addition, server-based components took advantage of developers' previous experience, whereas cloud-based components forced team members to innovate and face a learning curve for novel approaches. The platform also considered non-functional requirements derived from the communities' environments. For example, designing lighter user interfaces and releasing them from handling business logic was especially important for mobile users in deprived communities that might have low-end or old devices with technical specifications unable to run demanding apps. In addition, for data science projects, it is crucial to know the devices available prior to designing the architecture. This understanding can directly come from community engagement, their involvement with citizen-generated data production and the periodic assessments with members of the user communities.

The discussion around non-functional requirements also covered time and privacy constraints. Since the team had a tight schedule for development and launch, they had to review the requirements, coordinate back-end and front-end development, and quickly agree on contracts ruling the API services. Additionally, the need to comply with the recently adopted general personal data protection law (known as LGDP from the acronym in Portuguese) impacted data treatment and storage within the Brazilian borders. Therefore, the international research team decided to allocate and isolate cloud services accordingly, bypassing the restrictions of on-premise deployments.

Considering project stakeholders, especially authorities and technicians currently working on flood management, the team analysed the benefits and challenges the proposed platform in day-to-day work. The most evident advantage comes from the option to visually identify the gaps between official flood data and citizen generated data. The combined information can serve to compare the water levels registered by residents in the street (i.e., represented by blue circles) in different locations during stronger flooding and the polygons of areas susceptible to flooding (i.e., represented by blue polygons) as shown in Figure 4. These comparisons and the user interface most suitable for the different audiences are still topics to cover in future research.

Figure 4 Example of citizen-generated and official data comparison

This research has served to identify challenges of integrating cloud-based and hybrid architectures and frame future research, partner's requirements and other platform needs. The two final considerations were about operational costs and the technical skills demanded. These two are access barriers faced by community-led projects and relevant issues to address in future research.

6. CONCLUSION AND FUTURE WORK

The project set an overarching goal of building up a platform to store and exchange flood-related citizen-generated and authoritative data. By integrating the communities as part of a dialogical process, citizens and researchers agreed on jointly defining the requirements of a mobile app that would serve to interact with the data platform and complement official data. The paper presents the progress made in designing a platform to integrate citizen-generated data and flood monitoring, especially highlighting the results from the participatory design to define a user interface and select components to enable participatory flood monitoring. Using the underlying concept of "data flow", researchers worked on designing a data platform to engage vulnerable communities and traditional stakeholders with data production and circulation.

The innovative approach of engaging citizens in designing data platforms provided insights into participation strategies for future citizen science initiatives. Although most of the outcomes have been positive, the approach implied considerable learning and resource availability (i.e., access to cloud platforms, basic understanding of cloud services and software development skills). Therefore, involving communities in citizen science projects should come with the support needed to identify requirements and components as part of the data platform development. Future works might consider preparing tools, documents or boot camps for communities and researchers interested in engaging communities with data collection or flood monitoring.

ACKNOWLEDGMENTS

The authors would like to acknowledge the substantial contributions of the researchers of the Waterproofing Data project and residents of our partner communities at M'Boi Mirim in the city of São Paulo and Rio Branco (Acre). The authors also gratefully acknowledge the financial support by the Belmont Forum and NORFACE Joint Research Programme on Transformations to Sustainability, which is co-funded by DLR/BMBF, ESRC, FAPESP, and the European Commission through Horizon 2020, as well as funding from the United Kingdom Research and Innovation Global Challenges Research Fund and the support from Alan Turing Institute.

REFERENCES

Balázs, B., Mooney, P., Nováková, E., Bastin, L., Jokar Arsanjani, J., 2021. Data Quality in Citizen Science, in: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K. (Eds.), The Science of Citizen Science. Springer International Publishing, Cham, pp. 139–157. https://doi.org/10.1007/978-3-030-58278-4_8

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience 59, 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Chari, R., Petrun Sayers, E.L., Amiri, S., Leinhos, M., Kotzias, V., Madrigano, J., Thomas, E.V., Carbone, E.G., Uscher-Pines, L., 2019. Enhancing community preparedness: an inventory and analysis of disaster citizen science activities. BMC Public Health 19, 1356. https://doi.org/10.1186/s12889-019-7689-x

Conrad, C.C., Hilchey, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ Monit Assess 176, 273–291. https://doi.org/10.1007/s10661-0101582-5

Dong, X.L., Rekatsinas, T., 2018. Data Integration and Machine Learning: A Natural Synergy, in: Proceedings of the 2018 International Conference on Management of Data. Presented at the SIGMOD/PODS '18: International Conference on Management of Data, ACM, Houston TX USA, pp. 1645–1650. https://doi.org/10.1145/3183713.3197387

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R.A., Feyen, L., 2018. Increased human and economic losses from river flooding with anthropogenic warming. Nature Clim Change 8, 781–786. https://doi.org/10.1038/s41558-018-0257-z

Eleta, I., Galdon Clavell, G., Righi, V., Balestrini, M., 2019. The Promise of Participation and Decision-Making Power in Citizen Science. Citizen Science: Theory and Practice 4, 8. https://doi.org/10.5334/cstp.171 Freire, P., 2000. Pedagogy of the oppressed, 30th anniversary ed. ed. Continuum, New York.

Hicks, A., Barclay, J., Chilvers, J., Armijos, M.T., Oven, K., Simmons, P., Haklay, M., 2019. Global Mapping of Citizen Science Projects for Disaster Risk Reduction. Front. Earth Sci. 7, 226. https://doi.org/10.3389/feart.2019.00226

Kelling, S., Fink, D., La Sorte, F.A., Johnston, A., Bruns, N.E., Hochachka, W.M., 2015. Taking a 'Big Data' approach to data quality in a citizen science project. Ambio 44, 601–611. https://doi.org/10.1007/s13280015-0710-4

Kobusińska, A., Leung, C., Hsu, C.-H., S., R., Chang, V., 2018. Emerging trends, issues and challenges in Internet of Things, Big Data and cloud computing. Future Generation Computer Systems 87, 416–419. https://doi.org/10.1016/j.future.2018.05.021

Liu, Y., Piyawongwisal, P., Handa, S., Yu, L., Xu, Y., Samuel, A., 2011. Going Beyond Citizen Data Collection with Mapster: A Mobile+Cloud Real-Time Citizen Science Experiment, in: 2011 IEEE Seventh International Conference on E-Science Workshops. Presented at the 2011 IEEE Seventh International Conference on e-Science Workshops (eScienceW), IEEE, Stockholm, Sweden, pp. 1–6. https://doi.org/10.1109/eScienceW.2011.23

Marchezini, V., Horita, F.E.A., Matsuo, P.M., Trajber, R., Trejo-Rangel, M.A., Olivato, D., 2018. A Review of Studies on Participatory Early Warning Systems (P-EWS): Pathways to Support Citizen Science Initiatives. Front. Earth Sci. 6, 184. https://doi.org/10.3389/feart.2018.00184

Matheus, R., Janssen, M., Maheshwari, D., 2020. Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly 37, 101284. https://doi.org/10.1016/j.giq.2018.01.006

Munappy, A.R., Mattos, D.I., Bosch, J., Olsson, H.H., Dakkak, A., 2020. From Ad-Hoc Data Analytics to DataOps, in: Proceedings of the International Conference on Software and System Processes. Presented at the ICSSP '20: International Conference on Software and System Processes, ACM, Seoul Republic of Korea, pp. 165–174. https://doi.org/10.1145/3379177.3388909

O'Grady, M.J., Muldoon, C., Carr, D., Wan, J., Kroon, B., O'Hare, G.M.P., 2016. Intelligent Sensing for Citizen Science: Challenges and Future Directions. Mobile Netw Appl 21, 375–385. https://doi.org/10.1007/s11036-016-0682-z

Phillips, T.B., Parker, A., Bowser, A., Haklay, M., 2021. Publicly Generated Data: The Role of Citizen Science for Knowledge Production, Action, and Public Engagement, in: Ferreira, C.C., Klütsch, C.F.C. (Eds.), Closing the Knowledge-Implementation Gap in Conservation Science, Wildlife Research Monographs. Springer International Publishing, Cham, pp. 83–107. https://doi.org/10.1007/978-3-030-81085-6_4

Porto de Albuquerque, J., Albino de Almeida, A., 2020. Modes of engagement: Reframing "sensing" and data generation in citizen science for empowering relationships, in: Toxic Truths. Manchester University Press. https://doi.org/10.7765/9781526137005.00028

Porto de Albuquerque, J., Anderson, L., Calvillo, N., Coaffee, J., Cunha, M.A., Degrossi, L.C., Dolif, G., Horita, F., Klonner, C., Lima-Silva, F., Marchezini, V., Martins, M.H. da M., Pajarito-Grajales, D., Pitidis, V., Rudorff, C., Tkacz, N., Traijber, R., Zipf, A., 2021. The role of data in transformations to sustainability: a critical research agenda. Current Opinion in Environmental Sustainability 49, 153–163. https://doi.org/10.1016/j.cosust.2021.06.009

Rawat, S., Narain, A., 2019. Managing Flow, in: Understanding Azure Data Factory. Apress, Berkeley, CA, pp. 265–309. https://doi.org/10.1007/978-1-4842-4122-6_6

Roman, D., Nikolov, N., Soylu, A., Elvesaeter, B., Song, H., Prodan, R., Kimovski, D., Marrella, A., Leotta, F., Matskin, M., Ledakis, G., Theodosiou, K., Simonet-Boulogne, A., Perales, F., Kharlamov, E., Ulisses, A., Solberg, A., Ceccarelli, R., 2021. Big Data Pipelines on the Computing Continuum: Ecosystem and Use Cases Overview, in: 2021 IEEE Symposium on Computers and Communications (ISCC). Presented at the 2021 IEEE Symposium on Computers and Communications (ISCC), IEEE, Athens, Greece, pp. 1–4. https://doi.org/10.1109/ISCC53001.2021.9631410

Sturm, U., Schade, S., Ceccaroni, L., Gold, M., Kyba, C., Claramunt, B., Haklay, M., Kasperowski, D., Albert, A., Piera, J., Brier, J., Kullenberg, C., Luna, S., 2018. Defining principles for mobile apps and platforms development in citizen science. RIO 4, e23394. https://doi.org/10.3897/rio.4.e23394

Sy, B., Frischknecht, C., Dao, H., Consuegra, D., Giuliani, G., 2020. Reconstituting past flood events: the contribution of citizen science. Hydrol. Earth Syst. Sci. 24, 61–74. https://doi.org/10.5194/hess-24-612020

Williams, D., Máñez Costa, M., Celliers, L., Sutherland, C., 2018. Informal Settlements and Flooding: Identifying Strengths and Weaknesses in Local Governance for Water Management. Water 10, 871. https://doi.org/10.3390/w10070871

Wu, B., Tian, F., Zhang, M., Zeng, H., Zeng, Y., 2020. Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. Geography and Sustainability 1, 25–32. https://doi.org/10.1016/j.geosus.2020.03.006

Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017. Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth 10, 13–53. https://doi.org/10.1080/17538947.2016.1239771