

GEO COLLECTOR BOT: A TELEGRAM-BASED OPEN TOOLKIT TO SUPPORT FIELD DATA COLLECTION

D. Oxoli¹* E. Pessina¹, M. A. Brovelli¹

¹ Dept. of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Anzani 42, 22100 Como, Italy - daniele.oxoli@polimi.it, edoardopessina.priv@gmail.com, maria.brovelli@polimi.it

Commission IV, WG IV/4

KEY WORDS: Field Data Collection, Survey, Telegram Bot, FOSS4G, Volunteered Geographic Information

ABSTRACT:

Georeferenced field data collection has become a popular practice allowing everyone to contribute to mapping objects or reporting events. The spread of mobile devices - capable of recording and sharing location coordinates, media and features while on the go - is primarily accountable for such diffusion. Accordingly, a number of mobile apps and software frameworks have been developed and released to perform field data collection. These frameworks allow to customize and dispatch collection forms as well as to manage contributors and records through web interfaces or database management systems. From the contributors' perspective, specific mobile client apps need to be installed to access selectively the collection forms and contribute to the data collection on the field using their mobile devices. This operation might inhibit the sporadic contribution of occasional users who may not be willing to install additional software. To overcome this limitation, this work presents the Geo Collector Bot, an alternative software toolkit to empower field data collection projects avoiding the development and/or the installation of a specific mobile app on contributors' devices. The Geo Collector Bot is a configurable Telegram-based chatbot enabling to dispatch of data collection forms that can be activated and filled in through Telegram chats. The ultimate goal of the presented work is to provide an alternative free and open-source software framework suitable for general-purpose field data collection applications. Development patterns and system architecture are described in detail alongside future improvements and outlooks for the Geo Collector Bot project.

1. INTRODUCTION

The collection of georeferenced information in the field has become an established and popular practice allowing professionals, volunteers and citizens to contribute to mapping objects or reporting events. Field data collection is essential to a variety of domains (See et al., 2016) including many scientific and humanistic disciplines, humanitarian and rescue operations, location reviews and professional engineering surveys, to mention a few.

The spread of mobile devices that can record location coordinates, media and features while on the go (and share them through the web) is primarily accountable for such diffusion (McLester and Piel, 2021). As a result, several mobile apps and software frameworks (both proprietary as well as free and open-source) have been developed and released to perform data collection in the field. Most of these frameworks allow developers or data collection promoters to customize collection forms according to the characteristics of each collection task and manage both users and records through web dashboards or database management systems (Brovelli et al., 2015). From the user perspective, mobile client apps are available to access selectively the collection forms and contribute to the data collection in the field using mainly smartphones or tablets.

Nevertheless, limitations connected to the final user experience are common in many of these apps (Steinberg et al., 2019). On one hand, mobile client apps are not always available or optimized for all mobile OS, therefore preventing their use in the field by a significant number of potential contributors (Ovando-Leon et al., 2022). On the other hand, each of these frame-

works requires the installation of a specific mobile app on the user's device. This operation may not represent a significant obstacle to the contribution to data collection projects by very active or committed users. However, it might inhibit the contribution of occasional users who may not be willing to install additional software on their device for sporadic mapping of objects or event reporting (Pau et al., 2016).

Given the above, this work presents the Geo Collector Bot, an alternative free and open-source software toolkit to empower field data collection projects avoiding the development and/or the installation of a specific mobile app on contributors' devices. This by leveraging as a mobile client the popular Telegram messaging service (Telegram Messenger Inc., 2022), currently counting 550 million monthly active users worldwide (Statista, 2022).

The Geo Collector Bot has been developed in the framework of the INSUBRI.PARKS project (Insubriparks, 2022), funded by the Interreg Co-operation Programme 2014-2020. This project aims at increasing the tourism attractiveness of the Insubria region (between Northern Italy and Southern Switzerland) through the provision of infrastructure as well as integrated marketing and management strategies for the Insubria natural parks. The Bot represents a component of the virtual infrastructure supporting the project (Oxoli et al., 2019, Oxoli et al., 2020). It was originally designed to allow both park visitors and managers to easily collect and share geolocated records on parks' status and feedback on points of interest. However, the ultimate goal of the presented work is to provide an open and general-purpose data collection software framework suitable for various applications.

The remaining of the paper is as follows. Section 2 present

* Corresponding author

a comparative analysis with existing free and open-source application for field data collection. Features and system architecture of the Geo Collector Bot are described in Section 3 while ongoing and future developments are outlined in Section 4. Finally, Section 5 includes overviews and discussion of the achieved results as well as future directions and expectations for the presented work.

2. APPLICATIONS BENCHMARKING

Focusing on general-purpose geospatial data collection software frameworks, some of the most popular free and open-source solutions are the Open Data Kit (ODK) (Open Data Kit Development Team, 2022), the KoboToolbox (Kobo Inc., 2022) and Epicollect (Centre for Genomic Pathogen Surveillance, 2022). Proprietary or pay-per-use solutions, developed by major software firms, are also available on the market but they were not considered in this work.

The above listed free and open-source software frameworks provide client and server components and both web and mobile apps to support the full development of field data collection projects (Brovelli et al., 2015).

From the developer (or data collection promoter) perspective, the adoption of such frameworks is facilitated by the availability of open Application Programming Interfaces (APIs) and web-based dashboards to generate, deploy and manage collection forms, users and records (see Figure 1). Such systems can be either self-hosted or deployed on cloud infrastructure according to the developers' needs (Steinberg et al., 2019).

Despite differences in data collection workflow and development skills required to set up such services, a critical concern for the adoption and diffusion of these applications is connected mainly to their client components and, consequently, to the users' experience. The reason is two-fold. First, mobile client apps are not always cross-platform, therefore preventing their use in the field by a significant number of potential contributors (Ovando-Leon et al., 2022). This is the case of the ODK on iOS devices, for example. Last, there is the necessity of installing a specific mobile app on the user's device to access data collection forms on each of the software frameworks mentioned above. While frequent users might not be discouraged by those limitations, the contribution of occasional users might be inhibited due to the unwillingness of installing additional software on their device for sporadic reporting (Pau et al., 2016).

Other relevant examples of free and open-source software for field data collection are QField (OPENGIS.ch, 2022) and Geopaparazzi/SMASH (HydroloGIS S.r.l., 2022). These platforms implement a more technical approach to field data collection and they are designed to support also professional surveys. Indeed, complex spatial items can be mapped and manipulated within these platforms and both raster and vector standard geospatial formats can be handled and used as data sources in the data collection project setup. Accordingly, interactions with Geographic Information Systems (GIS) software, such as QGIS, are generally required for both customizations of data collection projects as well as the analysis of collected data (Montagnetti and Guarino, 2021). This might not make the adoption of QField and Geopaparazzi/SMASH always preferable for both not-specialized users and data collection promoters of simple citizen-science projects (Lotfian et al., 2020), where the

geospatial dimension of collected data is often limited to tuples of point coordinates.

A final relevant characteristic, common to all the considered platforms, is the enabled possibility for users to collect data in the field without the need for an internet connection (Gichohi, 2016). Therefore, records can be temporarily stored in the mobile device memory and then synchronized with the project databases (Olyazadeh et al., 2017).

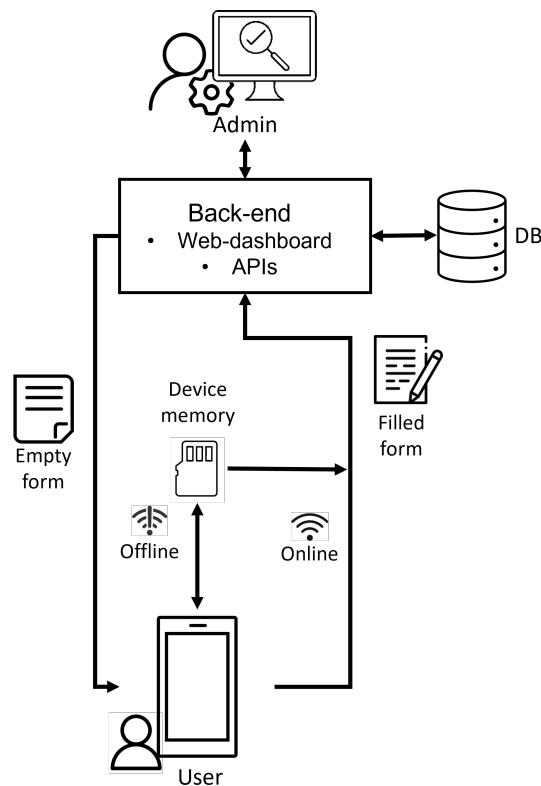


Figure 1. Schematic of common data collection software frameworks functionalities.

The Telegram-based open-toolkit - Geo Collector Bot - presented in this paper is conceived to mitigate some of the drawbacks in the user experience typical for the considered field data collection platforms and software frameworks. Details on development patterns, system architecture and functionalities of the Geo Collector Bot are described in the following section.

3. THE GEO COLLECTOR BOT

In this section, the current version of the Geo Collector Bot is presented alongside its system architecture and functionalities.

The Geo Collector Bot is a configurable Telegram-based chatbot enabling the dispatching of data collection forms that can be activated and filled through Telegram chats. It is released under MIT License and source code, documentation and a demo are available on GitHub (<https://github.com/opengeolab/geocollectorbot>).

The Geo Collector Bot works as a standard Telegram Bot. To collect the data, the Bot asks a series of questions to the user and persists the answers to a database. Available question

types comprise most of the traditional survey options including multiple-choice check-boxes, open text, user's location coordinates, and media file (foto or video) uploads (See Figure 2). Options for making questions mandatory or optional are also provided.

The Geo Collector Bot currently consists of a back-end application written in Typescript and running on Node.js. Some configuration is required to deploy the system. This can be performed by manual editing of set-up files for deployment environment variables definition, database and Telegram API connections. Both survey forms content and well as questions flow can be fully customised (according to the data collection project requirements) by editing a JSON configuration file that is also included in the Geo Collector Bot code-base. Documentation on configuration is available with the code repository.

As a supporting mobile client, the Telegram app is exploited. Telegram APIs mediate the bidirectional communication between the user and the system. These allow to dispatch the collection form to the user's device and redirect collected data to a database. The current version of Geo Collector Bot supports only PostgreSQL-PostGIS as data storage systems.

Local deployments of the system are facilitated by the provision of a Docker container (<https://hub.docker.com/r/geolabpolimi/geo-collector-bot>). A schematics of the Geo Collector Bot system architecture is included in Figure 3.

4. ONGOING AND FUTURE DEVELOPMENTS

It is worth mentioning that the development work for the Geo Collector Bot - at the time of writing - is still ongoing. New components under development are outlined as follows.

Ongoing development activities are focused mainly on the generalization of system configuration patterns, and the improvement of both users and developer documentation (<https://geocollectorbot-doc.readthedocs.io>, see Figure 4), and the support to other databases than PostgreSQL.

The current version of the Geo Collector Bot does not provide any dedicated back-end supporting modules for system set-up, data collection tasks management, and records analysis. These features are instead available for the reference data collection platforms and software frameworks considered in Section 2. Accordingly, future activities will mainly focus on the development of a Web-dashboard to facilitate the adoption of the Geo Collector Bot also by data collection promoters (or project administrators) without hard programming skills (see Figure 3).

The Web dashboard is conceived to ease the creation and publication of data collection projects by supplying survey forms creation and customization functionalities through graphical user interfaces. Data exploration and management modules will be also embedded in the dashboard to allow visualisation, analysis and sharing operations on the collected data. The Web dashboard will be key to aligning the Geo Collector Bot functionalities to the ones of the reference data collection platforms.

Finally, tools for feedback interaction between the project administrator and data collection contributors (users) will be included in the Geo Collector Bot (see Figure 3) coupled with dedicated Web dashboard modules for operating it. The Telegram APIs generate a unique chat ID for each bot chat (i.e.

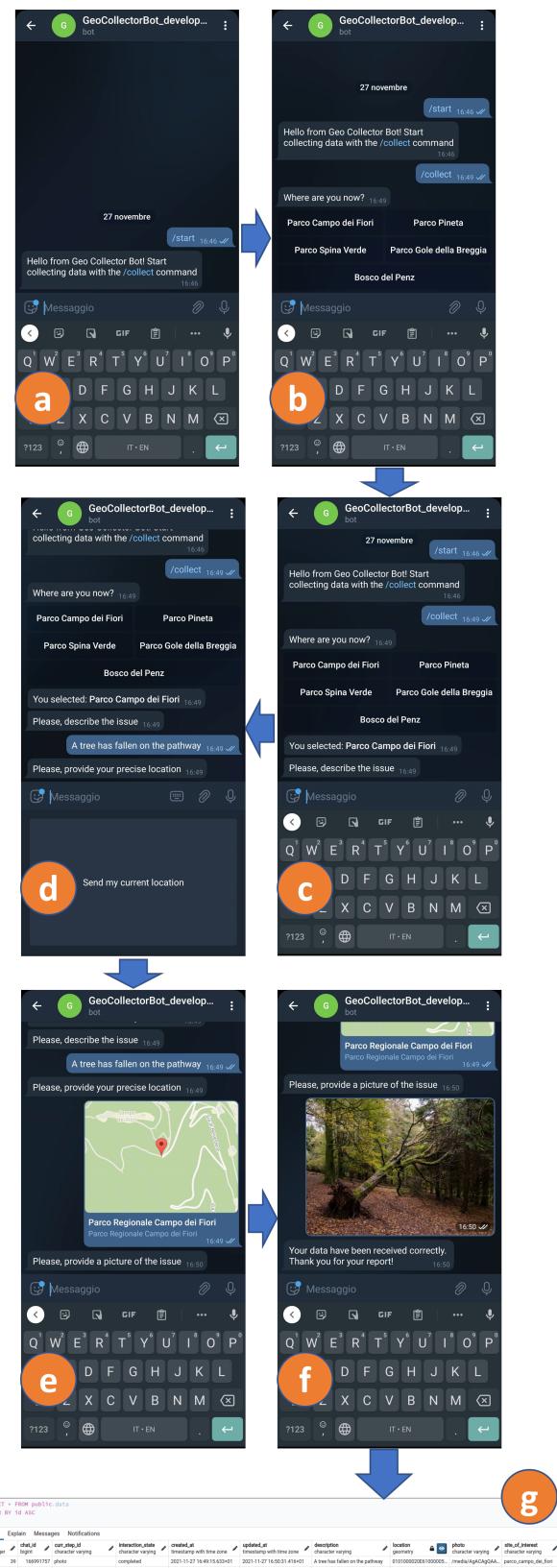


Figure 2. Example of Geo Collector Bot questions flow including (a) form activation, (b) multiple-choice check-boxes, (c) open text, (d) user's location coordinates, and (e) media uploads. Once the form is filled (f), collected data are forwarded and stored in the application database (g).

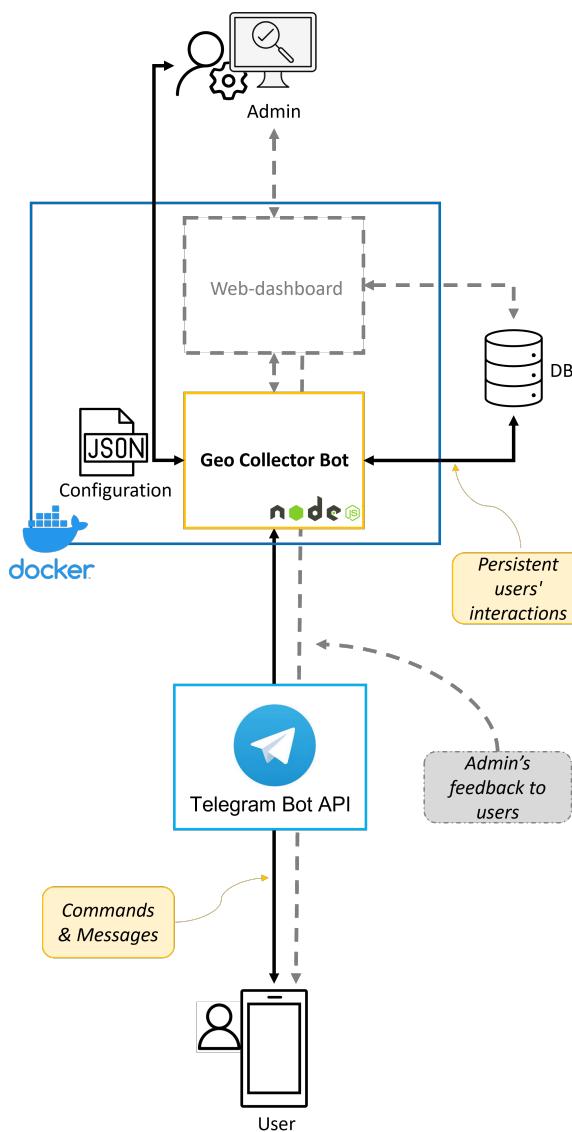


Figure 3. The Geo Collector Bot system architecture.
 Grey-dashed components and interactions are currently under development.

each connection between a specific user and the Geo Collector Bot) and transfer it to the system. By storing also this ID in the application database, the Geo Collector Bot will be enabled to communicate selectively with users. This pattern will allow the project administrator to provide feedback to data collection contributors and deliver it directly through the bot chat. The feedback functionality is expected to be particularly useful in those cases when the Geo Collector Bot would be used by the user to issue reports on events that might require the intervention of the project administrator to be solved.

A descriptive example of the above-outlined usage would be the case of a natural park visitor (user) who reports via the Geo Collector Bot to the park patrol (project administrator) that a tree has fallen on a path (see Figure 2). In this case, once the tree will be removed, the patrol could communicate to the visitor that the issue is solved. Not surprisingly, the feedback tool was requested by the INSUBRI.PARKS project partners that aim at using the Geo Collector Bot for operative monitoring of natural park areas.

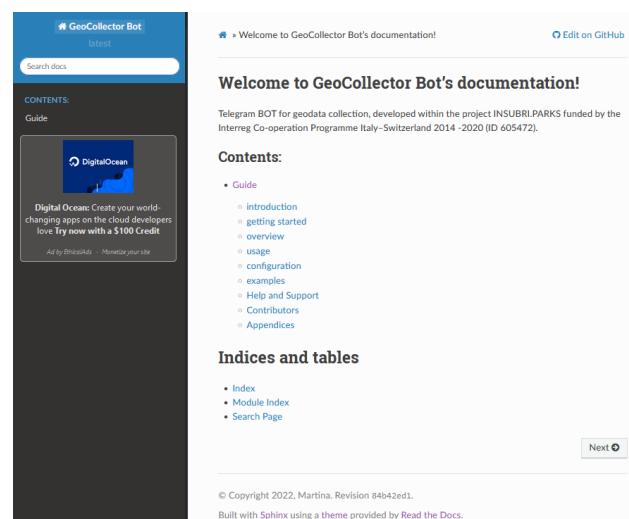


Figure 4. Preview of the Geo Collector Bot online documentation.

On the one hand, the feedback tool will empower the Geo Collector Bot with additional functionalities that are not typical of the existing data collection platforms by widening its possible application range. On the other hand, the possibility of establishing a feedback loop with data collection contributors is expected to contribute to user engagement in the data collection projects.

5. CONCLUSIONS AND OUTLOOKS

In this paper, the preliminary development of the Geo Collector Bot - a Telegram-based free and open-source software toolkit for field data collection - is described. The main goal of the presented system is to provide an alternative tool to existing platforms for which data collection contributors are not requested of installing additional software on their devices. This is achieved using the Telegram app as the mobile client.

Given the size of the Telegram app user base and thanks to its availability for different mobile OS, the typical issues in user experience, described in Section 2, can be overcome. Potentially, every Telegram user - 550 million monthly active users as of April 2022 (Statista, 2022) - could contribute to data collection projects based on the Geo Collector Bot thus enabling a large number of users to contribute, even sporadically, to data collection projects.

From the developer's perspective, the use of an open and cross-platform external mobile client, such as the Telegram app, allows for maintaining a single code base for the Geo Collector Bot, which results in a significant reduction of the development effort.

Nevertheless, the Telegram app also brings some limitations connected to mobile client customisation and operations. First, collection form design is restricted by the available Telegram widgets that can be directly embedded in the chats. The last and more critical limitation is the impossibility for the Geo Collector Bot of managing user devices' memory thus implicitly preventing offline data collection via standard Telegram chats. Despite this latter drawback, the use of Telegram APIs enables selective bi-directional communication between bots and chats

so that personalized messages can be automatically forwarded to a specific group of users or, eventually, to a specific user. The Geo Collector Bot will exploit this functionality to allow providing feedback to users, therefore enriching interactions and contributors' engagement in the data collection projects.

The current version of the Geo Collector Bot lacks of some components and functionalities, typical of the reference open data collection platforms and software frameworks available on the market. In particular, back-end supporting modules for system set-up, data collection tasks management, and records analysis are missing. To that end, the development of a Web-dashboard is planned and it will be included in the stack of the Geo Collector Bot Docker container as an auxiliary component.

The Web dashboard will allow to configure the Geo Collector Bot system, create and customize data collection projects, dispatch survey forms, manage users, and collect and analyse data through simplified interfaces. The addition of these components is key to lining up the Geo Collector Bot to the reference data collection platforms in terms of functionalities and usability.

Additional ongoing improvements to Geo Collector Bot regard the possibility of plugging other spatial database systems than PostgreSQL-PostGIS as well as the improvement of both user and developer software documentation.

Finally, a first extensive test phase for the Geo Collector Bot is planned. This will mainly involve partners of the INSUBRI.PARKS project with a focus on natural park managers and rangers. The system will be tested against current partners' practices of managing visitors' reports via direct communication channels (i.e. phone calls, SMS, WhatsApp messages, etc.) to assess real benefits connected to intervention time reduction and robust archiving of reported information. Thanks to the geographical position of the INSUBRI.PARKS project's natural parks, which extend alongside periurban areas, the unavailability of offline data collection functionalities for the Geo Collector Bot is expected to not significantly affect data collection.

The foreseen development and testing activities are expected to be concluded within the fourth quarter of the Year 2022, which corresponds to the INSUBRI.PARKS project end date.

Despite the specific application in the context of the INSUBRI.PARKS project, the ultimate goal of this work is to rise interest and grow a community around the Geo Collector Bot so that the continuation and sustainability of the project can be ensured and, in turn, contribute to both the spread and diversification of the available free and open-source geospatial software solutions.

ACKNOWLEDGEMENTS

This research has been funded by the project "INSUBRI.PARKS" within the Interreg V-A Italia-Svizzera Co-operation Programme 2014 - 2020 (ID 605472).

REFERENCES

Brovelli, M. A., Minghini, M., Zamboni, G., 2015. Public Participation GIS: a FOSS architecture enabling field-data collection. *International Journal of Digital Earth*, 8(5), 345–363.

Centre for Genomic Pathogen Surveillance, 2022. Epicollect5. <https://five.epicollect.net>. Accessed 23.05.2022.

Gichohi, B. W., 2016. Web surveys for offline rural communities. *Statistical Journal of the IAOS*, 32(4), 627–630.

HydroloGIS S.r.l., 2022. Geopaparazzi, SMASH and the survey server. <https://www.geopaparazzi.org>. Accessed 23.05.2022.

Insubriparks, 2022. The insubriparks project. <https://insubriparksturismo.eu>. Accessed 23.05.2022.

Kobo Inc., 2022. KoBoToolbox. <https://www.kobotoolbox.org>. Accessed 23.05.2022.

Lotfian, M., Ingensand, J., Brovelli, M. A., 2020. A framework for classifying participant motivation that considers the typology of citizen science projects. *ISPRS International Journal of Geo-Information*, 9(12), 704.

McLester, E., Piel, A. K., 2021. Mobile data collection apps. *Conservation Technology*, 177.

Montagnetti, R., Guarino, G., 2021. From Qgis to Qfield and Vice Versa: How the New Android Application Is Facilitating the Work of the Archaeologist in the Field. *Environmental Sciences Proceedings*, 10(1). <https://www.mdpi.com/2673-4931/10/1/6>.

Olyazadeh, R., Sudmeier-Rieux, K., Jaboyedoff, M., Derron, M.-H., Devkota, S., 2017. An offline–online Web-GIS Android application for fast data acquisition of landslide hazard and risk. *Natural Hazards and Earth System Sciences*, 17(4), 549–561.

Open Data Kit Development Team, 2022. Open Data Kit. <https://opendatakit.org>. Accessed 23.05.2022.

OPENGIS.ch, 2022. QField. <https://qfield.org>. Accessed 23.05.2022.

Ovando-Leon, G., Veas-Castillo, L., Gil-Costa, V., Marin, M., 2022. Bot-Based Emergency Software Applications for Natural Disaster Situations. *Future Internet*, 14(3), 81.

Oxoli, D., Cannata, M., Terza, V., Brovelli, M., 2019. NATURAL HERITAGE MANAGEMENT AND PROMOTION THROUGH FREE AND OPEN SOURCE SOFTWARE: A PRELIMINARY SYSTEM DESIGN FOR THE INSUBRI-PARKS PROJECT. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 179–183.

Oxoli, D., Terza, V., Cannata, M., Brovelli, M., Verde, P. R. S., 2020. An Open IT Infrastructure for Green Tourism Management and Promotion: The INSUBRI.PARKS Project. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, B4–2020.

Pau, E., Spano, D., Demontis, R., 2016. Gischatbot: una guida per la creazione di dati gis nelle chat di telegram. *Atti ASITA 2016*, 529–536.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M. et al., 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5(5), 55.

Statista, 2022. Number of monthly active Telegram users worldwide from March 2014 to April 2022.
<https://www.statista.com/statistics/234038/telegram-messenger-mau-users>. Accessed 23.05.2022.

Steinberg, M., Schindler, S., Klan, F., 2019. Software solutions for form-based, mobile data collection—A comparative evaluation. *BTW 2019—Workshopband*.

Telegram Messenger Inc., 2022. Telegram. <https://telegram.org>. Accessed 23.05.2022.