BIG DATA IN BIOGEOGRAPHY: FROM THE MUSEUM COLLECTION TO CITIZEN SCIENCE

2016 Roundtable Report
Italian Society of Biogeography
September 1st, 2016, Milan (Italy)

Coordinator: Valerio SBORDONI

Discussant: Paolo AUDISIO, Stefano DE FELICI, Leonardo LATELLA, Stefano MARTELLOS, Alberto ZILLI

The roundtable meeting was held in the framework of the joint Conference of UZI (Italian Zoological Union) SIB (Italian Society of Biogeography) and SITE (Italian Society of Ecology). The event brought together participants from universities, museums, and civil society.

In the meeting's opening statement, the President of SIB, Valerio Sbordoni, underlined the necessity to deeply understand how citizen science practices could be effective for data collection in biogeography, and for producing big data repositories. The creation of massive volumes of data does require the joint effort of many observers, which are normally no more available in the universities and/or in other research centers. Furthermore, modelling tools do require far more data than in the past. Hence, the involvement of volunteer citizens in the framework of citizen science activities can support researchers and decision makers. Combined with new data analysis tools, big data collected by citizens are a disruptive innovation, which is reconfiguring the idea on how research activities are conducted. This is even perceived by some authors as a revolution in Science, giving rise to new forms of empiricism that declare 'the end of theory', the creation of data-driven rather than knowledge-driven science.

While the production of big data is already common practice in some domains, such as remote sensing, weather prediction, and financial markets, their application in the field of biodiversity is still fragmentary. Yet, they represent an essential base to portray temporal and spatial distribution of organisms, model their past and present occurrence, and predict their future distribution under different scenarios.

Several countries have invested considerable efforts and resources for building national biodiversity repositories of primary biodiversity data, while others are still at the beginning of this process, in spite of recommendations expressed by the UN Convention on Biological Diversity (CBD).

International aggregators of data, like GBIF, Geobon, NCBI, TREEbase, and Tree of Life are examples of international efforts for collecting massive amounts of data. However, severe gaps still limit the use of these data at global scale. These gaps are primarily taxonomic knowledge (Linnean shortfall), and geographic coverage (Wallacean shortfall). The participation of volunteer citizens in the framework of citizen science programs can be extremely useful in filling these gaps.

The first discussant, Stefano De Felici (LifeWatch-ITA, University of Rome Tor Vergata), highlighted the importance of data aggregators (e.g. GBIF, OBIS and iNaturalist), and the relevant number of records reported for Italy, even if Italy officially do not participate in these networks. The discussant highlighted the increasing contributes from Citizen Science based sources, so the focus of the discussion involved the quality and reliability of data. Spatial and taxonomic accuracy of data aggregated in GBIF, as well as in other portals, has been disapproved. The discussant emphasized the importance of the taxonomists' role, especially in data quality control and validation, opposed to the decreasing number of taxonomists in the world.

The second discussant, Leonardo Latella (Natural History Museum of Verona), pointed out the role of the museums, and of their collections, as well as the importance of the involvement of museums in data collection processes. In Italy, ca. 20 millions specimens are preserved in ca. 1800 natural history collections, according to the first estimates, still partial, obtained in a national project aimed at mapping the collections of Italian natural history museums (Collmap). The National Biodiversity Network of the Italian Ministry of Environment aggregates data of several collections from the Natural History Museums of Verona and Florence only, hence leaving out a relevant portion of those data, because of funding shortage.

Leonardo Latella stressed the importance of the taxonomists' role, as essential for data validation. For this reason, during the last years, in the Museums of Verona, Florence and Ferrara, schools of taxonomy were developed (with one course on mammals, two on coleopterans and two on plants). The discussant described the positive response of the citizens to these activities, which can be considered as a very important first step towards a new generation of taxonomists. In fact, having huge amounts of data could be of no use for research in absence of taxonomists expertise for interpreting them. These initiatives are important also because there is, unfortunately, a steady decrease of less research projects, PhD positions, or degree thesis, which focus on taxonomy.

The following discussant, Stefano Martellos (University of Trieste) described the citizen science activities developed in the framework of the project CSMON-LIFE (LIFE13 ENV/IT/842), highlighting the interest demonstrated by citizens in terms of participation. Furthermore, the discussant underlined that the use of modern technologies, such as smartphones and other mobile devices for collecting data, do not restrict participation to young adults. CSMON-LIFE aims at developing a virtuous circle involving citizens, research institution, and decision makers has been created, in order to involve all the relevant stakeholders in the development of novel and more effective environmental policies. In fact, the public involvement is a fully recognized phenomenon and many European projects cannot be approved if the public dimension isn't considered an integral part of the project. However, the discussant raiser several issues related to the involvement of volunteers in data gathering activities:

1) it is difficult to create engagement, as well as retaining the volunteers after the first "enthusiastic" phase of involvement. This require a continuous process of feedback between volunteers and researchers and decision makers. Furthermore, it is necessary to raise awareness by using social media, as well as the more traditional mainstream media.

2) The quality of data, already addressed by Stefano De Felici, is the source of major criticisms on citizen science activities by many academics. How it is possible to verify the reliability of data in a citizen science project? Forcing volunteers to take images of the observed organisms can help the validation process, but has also the effect of lowering the amount of data provided by citizens (taking images is a more complex task than simply providing a record, and takes more time). However, on too many projects, the observer is the only responsible for data quality, and this lowers the interest of researchers for those data. On the contrary, it is of pristine importance to encourage the use of citizen science data by the scientific community, which should accept that data collected within a citizen science project, if going through a rigorous validation process, can be of good quality, with a rejection ratio of observations which is generally of ca. 10%.

Martellos concluded with a consideration of the possibility to raise the level of public engagement. In Italy, almost all the projects belong to the lower category of citizen science approach, where the citizen is "used "as a sensor to monitor a phenomenon. However, the actual potential of citizen science is higher, and volunteers can raise the ladder of engagement, becoming also interpreters of the data, and planners of new monitoring activities. However, bigger is the difficulty of engagement required, lower the number of participants. Hence, the major part of volunteers will be engaged in collecting data only, while more complex activities will be reserved to the more skilled and trained.

Finally, museums are starting to be involved in citizen science projects and data collection. Museums have the potential to be effective aggregating centers, where new ideas can be proposed, and where it could be possible to raise expertise and engagement level of volunteers.

Paolo Audisio (Sapienza University of Rome), shared his experience in a large project on the European fauna, which was funded with six million EUR. The project aims at developing a nomenclatural barebone for the entire European fauna. In four years, more than 135.000 species have been classified. The project portal has more than 5.000 visitors per day from all around the world. However, after this brilliant start, the project is temporarily on hold because of shortage of funding. The database is now managed by few organizations, and most of the work is managed by volunteers. At European scale, it seems to be evident that no perspective for the long term sustainability of the system does exist. The lack of structural funds is a major issue since they are fundamental for sustaining this (and several other) infrastructures which could be the heart of the biodiversity safeguard strategies at European level.

The last discussant, Alberto Zilli (London Natural History Museum), raised again the attention on the relevance of taxonomy, and the necessity to train students and volunteers, making them also capable to validate citizen science data. The discussant also comeback to the issues generated by the lack of funding. Lack of taxonomists and shortage of funding seem to be synchronized, and clearly affect each other. Hence, the question is: how to make taxonomy attractive again, and above all, how to demonstrate and communicate that the taxonomy is essential to biodiversity conservation? In the agrifood industry, there are several examples of actions taken on many species to minimize the copious taxonomic mistakes, which arise from the lack of taxonomic expertise.

During the final discussion, Martellos underlined the important role of decision makers, and the importance of public activities for raising awareness on the role of taxonomy. Citizen science activities, which provide training and knowledge to volunteers, can also have a relevant role in provide a better understanding of the role of biodiversity, and of taxonomy, to citizens. This opinion was supported by Valerio Sbordoni, who reported how to researchers the most relevant stakeholders are no longer the decision makers, but the increasing number of the citizens involved in citizen science and lifelong learning activities. Zilli, in turn, proposed to focus the attention on the relationship between citizen and food, in order to highlight wrong directions taken; he reported the heavy damages caused by poor biological control strategies, and how these waste costly resources. The discussion moved to the role of citizen science, and the issues related to the lack of taxonomists for data validation activities, and to the problems related to the taxonomic instability. Taxonomy, in fact, seems to evolve like a language, and its instability makes it intrinsically difficult to communicate, also among experts. Hence, the need for finding new ways to control the lack of taxonomists was highlighted, and the potential positive role of new ICT tools for supporting taxonomists' activities have been discussed.

The roundtable ended with a final discussion on citizen science and the involvement of the schools. Children engagement since early age is believed to be essential to raise environmental awareness. However, the age of participants is often inversely related to the scientific value of citizen science projects. Often, working with very young children leads to collection a lot of unreliable, or wrong, data. On the other side, working with high school students could really open new pathways. In citizen science, it is fundamental to find a good balance between educative aims, essential especially in projects that work with children, but relevant also when dealing with adults, and scientific aims, which can provide to citizen science the wide acceptance among scientists it was proven to deserve.

(testo di Flavia Bartoccioni)