

Masó, J. & Pons, X. (2025). Can we ask the same level of “FAIRNESS” to citizen generated data / ¿Podemos exigir el mismo nivel “FAIR” a los datos generados por la ciudadanía? *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), 35, I-VI. <https://dx.doi.org/10.21138/GF.922>

CAN WE ASK THE SAME LEVEL OF “FAIRNESS” TO CITIZEN GENERATED DATA

¿PODEMOS EXIGIR EL MISMO NIVEL “FAIR” A LOS DATOS GENERADOS POR LA CIUDADANÍA?

^{1a}JOAN MASÓ , ^{2,1b}XAVIER PONS 

^{1,2}Grumets Research Group

¹CREAF. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

²Departament de Geografia, Edifici B. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

^ajoan.maso@ieec.org ^bxavier.pons@uab.cat

Under the term *Citizen Science* we refer to practices that involve members of the general public—who are often not professional scientists—in contributing as volunteers to various stages of the scientific process. However, we can extend the concept into community-based activities that empower ordinary people to participate in the monitoring and management of environmental issues sometimes referred to as *citizen observatories*. In both cases, a common activity is the collection, sharing, and interpretation of data. Participatory activities existed before the advent of digital technologies, but they have been significantly enhanced by mobile devices, GNSS, sensors, apps, or web platforms, which have increased not only the number of observations but also their quality. As in professional science, data is merely a means to achieve an objective. In citizen science, the objective may be to demonstrate a scientific hypothesis (such as identifying galaxies in telescope images (Raddick *et al.* 2009), to raise awareness of a societal problem (such as air pollution (Wróblewski *et al.* 2021) or to gather sufficient evidence to persuade policymakers that something needs to be changed (such as reducing noise levels in city centers (SoundCollect, <https://scienceforchange.eu/en/news/news-mapping-the-sound-of-our-cities-with-citizen-science/>)). Despite the efforts invested in collecting citizen-generated data, many projects often overlook the value of their datasets beyond their initial objectives, and fail to make sufficient efforts to make the data openly accessible, and this limits the potential for unforeseen applications to emerge.

Environmental scientists have adopted data-sharing and FAIR principles (Findable, Accessible, Interoperable, and Reusable) to make data easily available to others, as part of a collective effort to adapt to and mitigate the effects of climate change and the pressure of the human activities on the environment. Despite being one of the most popular areas for citizen science, many Citizen Science initiatives forget to make their data FAIR. Examining experiences in European Union-funded projects (such as CitiObs (Castell *et al.* 2025), COS4Cloud (Soacha *et al.* 2020), ILIAD (Parkinson *et al.* 2024) and more4nature (MoRe4Nature, 2024–2027), it can be seen that this is not yet a common practice. An exception can be found in the case of “species occurrence” data, one of the biodiversity variables that is well organized. This is largely due to the maturity of well-established web platforms such as iNaturalist (Di Cecco *et al.* 2021).

Most citizen science projects focus their efforts on the challenging task of stimulating participation and maintain volunteer engagement in project activities. As a result, they often lack the resources or dedicated personnel to implement the necessary improvements to make their data FAIR, or they may simply overlook the importance of doing so. In a study involving more than 500 citizen science projects in the marine domain, only two initiatives demonstrated full compliance with all four

aspects of the FAIR principles (Wehn 2024), based on strict interpretation of go-FAIR sub-criteria (Henning *et al.* 2019). According to this interpretation; the “F” (findable) is achieved by through comprehensive metadata and the use of persistent identifiers, the “I” (interoperable) is demonstrated by the adoption of standardized vocabularies and agreed semantics; and the “R” (reusable) is ensured by applying a clear and open license (Thuermer *et al.* 2023).

A deeper analysis could help to identify where the main barriers are. In the same marine study, less than half of the projects offered their data as open access. While the reasons for this can only be speculated, it seems to be due to the lack of resources, or to a perceived lack of return on investment. Among the projects that did provide open access, only about half shared their data using common standards and protocols. This is likely due to the absence of adequate technical tools to support proper data access mechanisms. Therefore, greater efforts are needed to disseminate existing tools and ensure they are as user-friendly and cost-efficient as possible.

However, the most important barrier to achieving FAIR data is the lack of semantic interoperability. To make data useful beyond its original intended application, it is essential to ensure that the variables collected are explicitly and meaningfully described. This involves associating the recorded fields with accepted vocabularies and providing data in formats that transport these definitions as part of the metadata. Using a linked data approach, each definition is expressed as a URL that points directly to the web page of the concept. For example, if a citizen science project measures particulate matter with an average diameter of 2.5 micrometers (commonly tagged as PM25), the measurements can be linked to the definition provided by <https://www.eea.europa.eu/help/glossary/eea-glossary/pm2.5>. There are different ways to associate such definitions. In the AD4GD project (Masó *et al.* 2025), a semantic enrichment tool created by the Poznan Supercomputing and Networking Center has been tested, which transforms data into a version of JavaScript Object Notation adequate for Linked Data applications (JSON-LD) format, semantically tagging every aspect of the dataset with URLs from vocabularies. Within the Group on Earth Observations (GEO), another tool has been proposed to transform data into NetCDF format u the vocabularies standardized in the CF conventions (Eaton *et al.* 2003) for semantic tagging. However, both alternatives require converting data into formats that may be unfamiliar to end users. Thus, while improving interoperability, they may inadvertently reduce accessibility. To address this, a simpler approach has been proposed: semantically tagging CSV files by adding a companion JSON file known as CSV for the Web (CSVW) (Trojahn *et al.* 2022).

Another frequently observed barrier to FAIR data in the marine study is the lack of a clear and easy way to find the data license. When no license is provided, users downloading the data cannot determine whether it is reusable under specific conditions and are often forced to act on the side of caution and disregard the dataset altogether. In most cases, Citizen Science projects have no particular reason to restrict data usage and can easily adopt one of the Creative Commons licenses, such as the CC-0 (public domain dedication) or the CC-BY (attribution required).

In summary, with a bit of extra work it is possible to distribute data with an open license while adhering to the FAIR principles. Doing so facilitates the reuse of valuable datasets beyond their original purpose. Many citizen science projects have a limited geographical scope, which restricts the interest of their data specific local contexts. To overcome this limitation and enable their integration into global studies, data aggregation platforms pay a crucial role. These platforms are digital infrastructures that spatially aggregate data from different projects or activities, providing semantically consistent data and with clearly defined licenses. Such platforms are critical for managing large volumes of data, whether generated by citizens or professional scientists, providing support to the coordination among projects and initiatives, ensuring data quality, and making datasets accessible for both research and public use (Lush *et al.* 2024). Another example of data aggregation platform is the infrastructure being developed by the CitiObs project, which aggregates air quality data from initiatives that have adopted the Sensor Things API (Liang S. *et al.* 2021) standard to ensure data accessibility and reusability. If a Citizen Science project prioritizes open and FAIR data access, there should be no major difficulty to being integrated into these initiatives (such as the Global Biodiversity Information Facility, GBIF), thereby increasing both visibility and impact.

This issue features a study on the integration of transport and traffic data within the network in Madrid, aimed at improving access to shopping centers, and another in Buenos Aires, focused on

identifying centralities within a particularly complex urban fabric. It also includes research on bioclimatic classification in Venezuela. Furthermore, readers will find how investigations based on multitemporal satellite remote sensing enable detailed mapping of land use and land cover in Uruguay, the differentiation of numerous sugarcane varieties in Costa Rica, and the identification of musaceae monocultures in that same country, in this case using SAR imagery. Finally, the issue presents a contribution to studies on creek health in Puerto Rico.

Bajo el término *Ciencia Ciudadana* se hace referencia a prácticas que implican a miembros del público en general —que a menudo no son científicos profesionales— en la contribución voluntaria a diversas etapas del proceso científico. Sin embargo, el concepto puede extenderse a actividades comunitarias que empoderan a personas no profesionales para participar en la monitorización y gestión de cuestiones medioambientales, a veces denominadas *observatorios ciudadanos*. En ambos casos, una actividad común es la recopilación, el intercambio y la interpretación de datos. Las actividades participativas ya existían antes del auge de las tecnologías digitales, pero se han visto notablemente potenciadas por dispositivos móviles, GNSS, sensores, aplicaciones o plataformas web, lo que ha incrementado no solamente el número de observaciones, sino también su calidad. Al igual que en la ciencia profesional, los datos son meramente un medio para alcanzar un objetivo. En la ciencia ciudadana, dicho objetivo puede ser demostrar una hipótesis científica (como identificar galaxias en imágenes de telescopios (Raddick *et al.* 2009), visibilizar un problema social (como la contaminación del aire (Wróblewski *et al.* 2021) o reunir evidencia suficiente para convencer a los responsables políticos de que es necesario un cambio (como reducir los niveles de ruido en los centros urbanos (SoundCollect, <https://scienceforchange.eu/en/news/news-mapping-the-sound-of-our-cities-with-citizen-science/>)). A pesar de los esfuerzos invertidos en la recopilación de datos generados por la ciudadanía, muchos proyectos suelen pasar por alto el valor de sus conjuntos de datos más allá de sus objetivos iniciales y no realizan esfuerzos suficientes para hacer que los datos sean accesibles de forma abierta, lo que limita el potencial para aplicaciones no previstas inicialmente.

Los científicos ambientales han adoptado principios de compartición de datos y FAIR (Localizable, Accesible, Interoperable y Reutilizable, por sus siglas en inglés) para facilitar el acceso a los datos como parte de un esfuerzo colectivo para adaptarse y mitigar los efectos del cambio climático y la presión de las actividades humanas sobre el medio ambiente. A pesar de ser uno de los ámbitos más populares de la ciencia ciudadana, muchas iniciativas olvidan aplicar los principios FAIR a sus datos. A partir de experiencias en proyectos financiados por la Unión Europea (como CitiObs (Castell *et al.* 2025), COS4Cloud (Soacha *et al.* 2020), ILIAD (Parkinson *et al.* 2024) y more4nature (MoRe4Nature, 2024–2027), se observa que esta práctica aún no está generalizada. Una excepción se encuentra en los datos de "presencia de especies", una de las variables de biodiversidad que está bien organizada, en gran parte gracias a la madurez de plataformas web consolidadas como iNaturalist (Di Cecco *et al.* 2021).

La mayoría de los proyectos de ciencia ciudadana concentran sus esfuerzos en la difícil tarea de fomentar la participación y mantener el compromiso de los voluntarios en las actividades del proyecto. Como resultado, a menudo carecen de recursos o personal especializado para implementar las mejoras necesarias que permitan que sus datos sean FAIR, o simplemente no reconocen la importancia de hacerlo. En un estudio que analizó más de 500 proyectos de ciencia ciudadana en el ámbito marino, solo dos iniciativas demostraron cumplimiento total con los cuatro principios FAIR (Wehn 2024), según una interpretación estricta de los subcriterios de go-FAIR (Henning *et al.* 2019). De acuerdo con esta interpretación, la "F" (Localizable) se logra mediante metadatos exhaustivos y el uso de identificadores persistentes; la "I" (Interoperable) se demuestra mediante la adopción de vocabularios estandarizados y semánticas acordadas; y la "R" (Reutilizable) se garantiza mediante la aplicación de una licencia clara y abierta (Thuermer *et al.* 2023).

Un análisis más profundo podría ayudar a identificar dónde se encuentran las principales barreras. En el mismo estudio marino, menos de la mitad de los proyectos ofrecían sus datos en acceso abierto. Aunque sobre las razones solamente se puede especular, parece deberse a la falta de recursos o a la percepción de una escasa rentabilidad. Entre los proyectos que sí proporcionaban acceso abierto, sólo aproximadamente la mitad compartía los datos utilizando estándares y protocolos comunes. Esto probablemente se deba a la ausencia de herramientas técnicas adecuadas para ofrecer mecanismos apropiados de acceso a los datos. Por tanto, es necesario realizar mayores esfuerzos para difundir las herramientas existentes y garantizar que sean lo más fáciles de usar y rentables posible.

Sin embargo, la barrera más importante para lograr datos FAIR es la falta de interoperabilidad semántica. Para que los datos sean útiles más allá de su aplicación prevista originalmente, es esencial que las variables recopiladas estén descritas de forma explícita y comprensible. Esto implica asociar los campos registrados con vocabularios aceptados y proporcionar los datos en formatos que incluyan estas definiciones como parte de los metadatos. Utilizando un enfoque de datos enlazados, cada definición se expresa como una URL que apunta directamente a la página web del concepto. Por ejemplo, si un proyecto de ciencia ciudadana mide materia particulada con un diámetro medio de 2.5 micrómetros (comúnmente etiquetada como PM2.5), las mediciones pueden vincularse a la definición proporcionada por <https://www.eea.europa.eu/help/glossary/eea-glossary/pm2.5>. Existen diferentes formas de asociar estas definiciones. En el proyecto AD4GD (Masó *et al.* 2025), se ha probado una herramienta de enriquecimiento semántico creada por el Poznań Supercomputing and Networking Center, que transforma los datos al formato JSON-LD (JavaScript Object Notation for Linked Data), etiquetando semánticamente cada aspecto del conjunto de datos con URL de vocabularios. En el marco del Group on Earth Observations (GEO), se ha propuesto otra herramienta para transformar los datos al formato NetCDF utilizando los vocabularios estandarizados en las convenciones CF para el etiquetado semántico (Eaton *et al.* 2003). Sin embargo, ambas alternativas requieren convertir los datos a formatos que pueden resultar desconocidos para los usuarios finales. Con ello, al mejorar la interoperabilidad, se podría reducir inadvertidamente la accesibilidad. Para abordar este problema, se ha propuesto un enfoque más sencillo: etiquetar semánticamente archivos CSV mediante un archivo JSON complementario conocido como CSV para la Web (CSVW) (Trojahn *et al.* 2022).

Otra barrera observada con frecuencia en el estudio marino es la falta de una forma clara y sencilla de identificar la licencia de los datos. Cuando no se proporciona una licencia, los usuarios que descargan los datos no pueden determinar si estos son reutilizables bajo ciertas condiciones, y a menudo se ven obligados a actuar con cautela y descartar el conjunto de datos. En la mayoría de los casos, los proyectos de ciencia ciudadana no tienen motivos para restringir el uso de los datos y pueden adoptar fácilmente una de las licencias Creative Commons, como CC-0 (dedicada al dominio público) o CC-BY (requiere atribución).

En resumen, con un pequeño esfuerzo adicional es posible distribuir datos con una licencia abierta cumpliendo los principios FAIR. Esto facilita la reutilización de conjuntos de datos valiosos más allá de su propósito original. Muchos proyectos de ciencia ciudadana tienen un alcance geográfico limitado, lo que restringe el interés de sus datos a contextos locales específicos. Para superar esta limitación y permitir su integración en estudios globales, las plataformas de agregación de datos desempeñan un papel crucial. Estas plataformas son infraestructuras digitales que reúnen espacialmente datos de diferentes proyectos o actividades, proporcionando datos semánticamente consistentes y con licencias claramente definidas. Tales plataformas son fundamentales para gestionar grandes volúmenes de datos, ya sean generados por ciudadanos o científicos profesionales, brindando apoyo a la coordinación entre proyectos e iniciativas, garantizando la calidad de los datos y haciendo que los conjuntos de datos sean accesibles tanto para la investigación como para el uso público (Lush 2024). Otro ejemplo de plataforma de agregación de datos es la infraestructura que está desarrollando el proyecto CitiObs, que agrega datos de calidad del aire de iniciativas que han adoptado el estándar Sensor Things API (Liang S. *et al.* 2021) para garantizar la accesibilidad y reutilización de los datos. Si un proyecto de Ciencia Ciudadana prioriza el acceso abierto y FAIR a los datos, no debería haber mayores dificultades para integrarse en estas iniciativas (como la Global Biodiversity Information Facility, GBIF), aumentando así su visibilidad e impacto.

En este número contamos con un estudio sobre la integración de datos de transporte y tráfico en la red en Madrid, orientado al acceso a centros comerciales, y otro en Buenos Aires, enfocado en la

identificación de centralidades dentro de una trama urbana especialmente compleja. También se incluyen trabajos de clasificación bioclimática en Venezuela. Asimismo, podremos leer cómo las investigaciones basadas en teledetección satelitaria multitemporal permiten realizar cartografía detallada de usos y cubiertas en Uruguay, diferenciar numerosas variedades de caña de azúcar en Costa Rica, o identificar monocultivos de musáceas en ese mismo país, en este caso mediante imágenes SAR. Finalmente, se presenta una aportación a los estudios de salubridad en arroyos de Puerto Rico.

References

- Castell, N., Hassani, A., Wehn, U., Masó, J., Tavares, J., Chtzigoulas, A. (2025). Enhancing Citizen Observatories for healthy, sustainable, resilient and inclusive cities (No. EGU25-15040). *Copernicus Meetings*. <https://doi.org/10.5194/egusphere-egu25-15040>
- Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., Hurlbert, A. H. (2021). Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, 71(11), 1179-1188. 10.1093/biosci/biab093. <https://academic.oup.com/bioscience/article/71/11/1179/6357804?login=false>
- Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., ..., Lee, D. (2003). NetCDF Climate and Forecast (CF) metadata conventions. <https://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html>.
- Henning, P. C., Ribeiro, C. J. S., Santos, L. S., dos Santos, P. X. (2019). GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. *Em Questão*, 25(2), 13. DOI: 10.19132/1808-5245252.389-412. <https://seer.ufrgs.br/EmQuestao/article/view/84753>
- Liang, S., Khalafbeigi, T., van Der Schaaf, H., Miles, B., Schleidt, K., Grellet, S., ..., Alzona, M. (2021). OGC SensorThings API part 1: Sensing version 1.1. <https://docs.ogc.org/is/18-088/18-088.html>
- Lush, V., Bastin, L., Otsu, K., Masó, J. (2024). *Assessing FAIRness of citizen science data in the context of the Green Deal Data Space*. *International Journal of Digital Earth*, 17(1). <https://doi.org/10.1080/17538947.2024.2344587>
- Masó, J., Brobia, A., Palma, R., EliceGUI, I. (2025). Making Geospatial Data Assets Discoverable and Accessible in the Green Deal Data Space. *AGILE: GIScience Series*, 6, 38. <https://doi.org/10.5194/agile-giss-6-38-2025>
- MoRe4Nature. (2024–2027). *Empowering citizens in collaborative environmental compliance assurance via MOnitoring, REporting and action*. <https://www.more4nature.eu/>
- Parkinson, S., Ceccaroni, L., Edelist, D., Robertson, E., Horincar, R., Laudy, C., ..., Woods, S. M. (2024). The Iliad digital twins of the ocean: opportunities for citizen science. *ARPHA Proceedings*, 6, 61-65. DOI: 10.3897/ap.e126643. <https://ap.pensoft.net/article/126643/>
- Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., ..., Vandenberg, J. (2009). Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925*. <https://doi.org/10.48550/arXiv.0909.2925>.
- Soacha Godoy, K. A., Piera, J., Rodríguez Arias, M. À. (2020). Cos4Cloud-Integrating Citizen Science in the European Open Science Cloud: Challenges and opportunities for developing a new generation of Citizen Science Observatories. *ECSA Conference* (2020). <http://hdl.handle.net/10261/240817>
- Thuermer, G., González Guardia, E., Reeves, N., Corcho, O., Simperl, E. (2023). Data management documentation in citizen science projects: bringing formalisation and transparency together. *Citizen Science: Theory and Practice*, 8, 25. <https://doi.org/10.5334/cstp.538>

- Trojahn, C., Kamel, M., Annane, A., Aussenac-Gilles, N., Nguyen, B. L. (2022, September). A FAIR core semantic metadata model for FAIR multidimensional tabular datasets. *In 23rd International Conference on Knowledge Engineering and Knowledge Management (EKAW 2022)*, pp.174 - 181, 10.1007/978-3-031-17105-5_13ff. hal-03872685. <https://hal.science/hal-03872685v1/document>
- Wehn, U., Bilbao Erezkano, A., Somerwill, L., Linders, T., Masó, J., Parkinson, S., Semasingha, C., Woods, S. (2024). Past and present marine citizen science around the globe: A cumulative inventory of initiatives and data produced. *Ambio*. <https://doi.org/10.1007/s13280-024-02119-z>
- Wróblewski, M., Suchomska, J., Tamborska, K. (2021). Citizens or consumers? Air quality sensor users and their involvement in sensor. community. results from qualitative case study. *Sustainability*, 13(20), 11406. <https://doi.org/10.3390/su132011406>