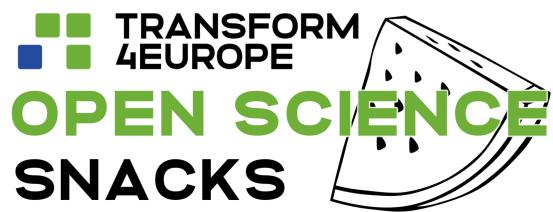


T4EU Open Science Snack 04: Open Data

Dear Open Science Snackers,

Welcome to our fourth Open Science Snack, this time on Open Data. In this edition, we will give an overview of the questions *what is, why and how to* Open Data and introduce another important related concept: the FAIR principles.



What is Open Data?

According to one prominent [definition](#), the “Open” in Open Data means that “... **anyone can freely access, use, modify, and share for any purpose.**”

What does this definition mean in practice? It means that open data must be available to interested parties, preferably as a download in a common, machine-readable format (e.g., [CSV](#)). Costs for accessing the data should not exceed a reasonable reproduction fee, if any. Anyone should be able to use, reuse and redistribute the data for any purpose, including combining it with other datasets - always with attribution to the author(s). And: “Anyone” includes commercial use, at least according to the above referenced [definition](#). If you want to know why commercial use is beneficial - for instance because it allows Wikipedia and bloggers to include this content - have a look [at this booklet](#) that explains why a non-commercial clause often won’t serve your needs. In any case, you should think carefully about who should have access to your data.

As for the type of data we are talking about, “the key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals.”¹



By [David Mellor](#) | Original: Ainali | Vector: AntiCompositeNumber - <https://osf.io/tvyxz/>, [CC BY 4.0](#), <https://commons.wikimedia.org/w/index.php?curid=64318704>

Why Open Data?

There are several reasons why data should be open. First, it is an integral part of the scientific endeavor that other researchers can check and reproduce published results. **Reproducibility** is a hallmark of transparency that helps to ensure trust in scientific findings. Therefore, all data necessary to reproduce published findings should be made available. Several analyses show that data is otherwise predominantly not accessible - even when authors assure the contrary in published papers^{2,3}. Second, an oft-repeated demand is that **publicly funded research should**

be publicly available. Third, your **data may be relevant** to other researchers who might use it instead of conducting a similar study again. This saves a lot of resources. Finally, there are also **selfish reasons** for making your data open, as you are more likely to spot errors and generally make the data more useful to others, including your future self, if you prepare your data thoroughly for publication in a repository. As data in a repository is usually given a DOI, this allows others to cite your data, thereby increasing your visibility.

Apart from these arguments that convinced many researchers to openly share their data, there may also be **formal obligations** to make your data open, for example as funders have certain requirements. Note that these requirements may come with an additional benefit: Many funders allow you to apply for funding to cover the costs associated with data management and sharing.

Despite all these reasons to openly share research data, it is also true that there may sometimes be good reasons not to do so. For example, privacy reasons, intellectual property rights, or other legal reasons may preclude sharing some or even all of the data. Data protection always trumps openness. Thus, open scientists and funders such as Horizon Europe do not unreservedly argue for Open Data, but rather for making data “[as open as possible, as closed as necessary](#)”. And most importantly: FAIR.

What are the FAIR principles?

The [FAIR principles](#) are guidelines to improve the **Findability, Accessibility, Interoperability, and Reuse** of data and other digital objects that were first published in 2016.⁴ The goal of the FAIR principles is to support the discoverability of data to simplify and increase knowledge discovery.

An important rationale for the development of the FAIR principles was that “humans increasingly rely on computational support to deal with data”.⁵ Machine-actionability is therefore a focus of the FAIR principles.

It is important to stress that FAIR data are not necessarily open. It is very well possible to work FAIRly without making your data publicly available.

If you want to learn more about the FAIR principles, the [FAIRy tale](#) is an entertaining story to start with. There are a lot more resources on the internet such as a [FAIR teaching and training handbook](#) for higher education institutions or a [discipline-agnostic tool](#) to help you assess your knowledge of the FAIR principles and that can also be used by trainers to assess the knowledge of course participants.



How to Open Data?

If you would like to open a dataset, make sure you choose a common and machine-readable format and decide for a trusted repository for your data. There is a lot of help on the Internet about choosing a repository, for example, from [OpenAIRE](#) or the [Registry of Research Data Repositories](#). A good repository will then help you a lot with the next steps:

1. Providing rich and machine-readable metadata (data about your data such as format, author, or source). Even if you decide not to (fully) open your data, at least your metadata should always be open.

2. Giving your data a persistent identifier, such as a [DOI](#), to make it uniquely identifiable and reliably accessible.
3. [Choosing](#) an appropriate [license](#) for your data, so others know if and how they can access, use and share your data.

A common way to address all these issues and more right from the start of a research project is through a [data management plan \(DMP\)](#). A DMP describes how the data will be handled during and after a research project, usually covering aspects such as data description and collection, documentation, storage, sharing and long-term preservation, and responsibilities. DMPs are now mandatory for projects funded by the European Union ([Horizon Europe](#), ERC). There are several online tools to help you write a DMP such as [argos](#) from OpenAIRE or the [Data Steward Wizard](#).

If you want to learn more about Open Data, you can take a look at the [official portal for European Data](#) and their [e-learning programme](#), which consists of 16 modules covering topics such as the legal side of Open Data or how to measure the impact of Open Data.

In the next Snack, we will have a look at preregistrations: a way of registering your research and analysis plan before you collect any data. We will discuss what preregistrations are and how they can improve the quality of your research.

Best regards,
The Science4All Initiative
[\[Transform4Europe research & innovation project T4ERI\]](#)

Upcoming Snacks

- Preregistration 101
- The value of replications

Past Snacks

- What is Open Science and what is in it for you?
- Myths about Open Science
- Open Access

You can find all the Snacks also on our website: <https://transform4europe.eu/t4eri/science4all/>

References

- (1) <https://opendatahandbook.org/guide/en/what-is-open-data/>
- (2) Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, 150, 33–41. <https://doi.org/10.1016/j.jclinepi.2022.05.019>
- (3) Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>
- (4) Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- (5) <https://www.go-fair.org/fair-principles/>
- (6) <https://www.openaire.eu/how-to-make-your-data-fair>

The Open Science Snacks by Science4All Initiative are licensed under [CC-BY 4.0](#).

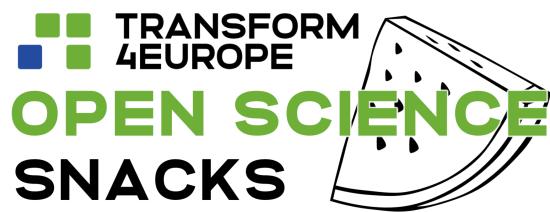


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101035805

T4EU Open Science Snack 04: Datos Abiertos (Open Data)

Queridos Snackers de Ciencia Abierta,

Bienvenidos a nuestro cuarto Snack de Ciencia Abierta, esta vez sobre datos abiertos. En esta edición, ofreceremos una visión general de las preguntas: ¿qué son los datos abiertos?, ¿por qué son importantes?, y ¿cómo se utilizan? Además, introduciremos otro concepto clave relacionado: los principios FAIR.



¿Qué son los datos abiertos?

Según una destacada [definición de The Open Definition](#), el "abierto" en datos abiertos significa que "... cualquiera puede acceder libremente a ellos, utilizarlos, modificarlos y compartirlos con cualquier fin" ("... anyone can freely access, use, modify, and share for any purpose.").

¿Qué significa esta definición en la práctica? Significa que los datos abiertos deben estar a disposición de las partes interesadas, preferiblemente como descarga en un formato común legible por máquina (por ejemplo, [CSV](#)). Los costes de acceso a los datos no deben superar una tasa de reproducción razonable, en su caso. Cualquiera debe poder utilizar, reutilizar y redistribuir los datos para cualquier fin, incluida su combinación con otros conjuntos de datos, siempre con atribución al autor o autores. El término "Cualquiera" incluye el uso comercial, al menos según la [definición](#) mencionada. En caso de que quiera conocer por qué el uso comercial es beneficioso -por ejemplo, porque permite a Wikipedia y a los blogueros incluir este contenido- revise [este folleto](#) que explica por qué una cláusula no comercial a menudo no sirve para sus necesidades. En cualquier caso, se recomienda pensar detenidamente quién debe tener acceso a sus datos. En cuanto al tipo de datos de que hablamos, "el punto clave es que, al abrir los datos, la atención se centra en los datos no personales, es decir, los que no contienen información sobre individuos concretos".¹



By [David Mellor](#) | Original: Ainali | Vector: AntiCompositeNumber - <https://osf.io/tvyxz/>, [CC BY 4.0](#), <https://commons.wikimedia.org/w/index.php?curid=64318704>

¿Por qué datos abiertos?

Hay varias razones por las que los datos deben ser abiertos. En primer lugar, es parte integrante del quehacer científico que otros investigadores puedan comprobar y reproducir los resultados

publicados. La **reproducibilidad** es un sello distintivo de la transparencia que ayuda a garantizar la confianza en los hallazgos científicos. Por lo tanto, todos los datos necesarios para reproducir los resultados publicados deben estar disponibles. Varios análisis muestran que, por lo general, no se puede acceder a los datos, incluso cuando los autores aseguran lo contrario en los artículos publicados^{2,3}. En segundo lugar, una exigencia que se repite a menudo es que la investigación **financiada con fondos públicos debe estar a disposición del público**. En tercer lugar, sus **datos pueden ser relevantes** para otros investigadores, que podrían utilizarlos en lugar de volver a realizar un estudio similar. Esto ahorra muchos recursos. Por último, también hay **razones egoísticas** para hacer que sus datos sean abiertos, ya que es más probable que detecte errores y, en general, que los datos sean más útiles para otros, incluido su futuro yo, si prepara sus datos minuciosamente para publicarlos en un repositorio. Como los datos de un repositorio suelen tener un DOI, esto permite que otros autores citen sus datos, aumentando así su visibilidad.

A parte de estos argumentos, que han convencido a muchos investigadores para compartir abiertamente sus datos, también puede haber **obligaciones formales** para hacer que sus datos sean abiertos, por ejemplo, porque las convocatorias o los organismos que financian las iniciativas lo requieran. Tenga en cuenta que estos requisitos pueden conllevar un beneficio adicional: muchas convocatorias le permiten solicitar financiación para cubrir los costes asociados a la gestión y puesta en común de los datos.

A pesar de todas estas razones para compartir abiertamente los datos de investigación, también es cierto que a veces puede haber buenas razones para no hacerlo. Por ejemplo, razones de privacidad, derechos de propiedad intelectual u otras razones legales pueden impedir compartir algunos o incluso todos los datos. La protección de los datos siempre prevalece sobre la apertura. Por ello, los científicos defensores del acceso abierto y los programas como Horizonte Europa no abogan incondicionalmente por los Datos Abiertos, sino por hacer que los datos sean "tan abiertos como sea posible, tan cerrados como sea necesario" ("[as open as possible, as closed as necessary](#)"). Y lo más importante: que sean FAIR.

¿Qué son los principios FAIR?

Los [principios FAIR](#) son directrices para mejorar la encontrabilidad, accesibilidad, interoperabilidad y reutilización de datos y otros objetos digitales que se publicaron por primera vez en 2016.⁴ El objetivo de los principios FAIR es facilitar la localización de los datos para simplificar y aumentar el descubrimiento de conocimiento.



Una razón importante para el desarrollo de los principios FAIR fue que "los seres humanos dependen cada vez más de la ayuda informática para tratar los datos".⁵ Por lo tanto, los principios FAIR se centran en la capacidad de acción de las máquinas.

Es importante subrayar que los datos FAIR no son necesariamente abiertos. Es muy posible trabajar de forma FAIR sin poner los datos a disposición del público.

Si desea aprender más sobre los principios FAIR, el [cuento FAIRy](#) es una historia entretenida para empezar. Hay muchos más recursos en Internet, como un [manual de enseñanza y formación FAIR](#) para instituciones de educación superior o [una herramienta independiente de la](#)

[disciplina](#) que le ayudará a evaluar sus conocimientos sobre los principios FAIR y que también pueden utilizar los formadores para evaluar los conocimientos de los participantes en los cursos.

¿Cómo abrir los datos?

Si desea abrir un conjunto de datos, asegúrese de elegir un formato común y legible y decídase por un repositorio de confianza para sus datos. Hay mucha ayuda en Internet sobre cómo elegir un repositorio, por ejemplo, en [OpenAIRE](#) o en el [Registro de Repositorios de Datos de Investigación](#). Un buen repositorio le ayudará mucho en los siguientes pasos:

1. Proporcionar metadatos ricos y legibles por máquina (datos sobre sus datos como formato, autor o fuente). Aunque decida no abrir (totalmente) sus datos, al menos sus metadatos deberían estar siempre abiertos.
2. Asignar a sus datos un identificador persistente, como un [DOI](#), para que sean identificables de manera única y accesibles de forma fiable.
3. [Elegir una licencia](#) adecuada para sus datos, de modo que otros sepan si pueden acceder a ellos, utilizarlos y compartirlos, y cómo hacerlo.

Una forma habitual de abordar todas estas cuestiones y más desde el principio de un proyecto de investigación es mediante un plan de gestión de datos ([data management plan, DMP](#)). En él se describe cómo se gestionan los datos durante y después del proyecto de investigación, y suele abarcar aspectos como la descripción y recopilación de datos, la documentación, el almacenamiento, la puesta en común y la conservación a largo plazo, así como las responsabilidades. Los DMP son ahora obligatorios para los proyectos financiados por la Unión Europea ([Horizon Europe](#), ERC). Existen varias herramientas en línea que le ayudarán a redactar un DMP, como [argos](#) de OpenAIRE o [Data Steward Wizard](#).

Si desea aprender más sobre Datos Abiertos, le recomendamos visitar el [portal oficial de Datos Europeos](#) y su [programa de e-learning](#), que consta de 16 módulos que cubren temas como el aspecto legal de los Datos Abiertos o cómo medir el impacto de los Datos Abiertos.

En el próximo Snack, abordaremos los prerregistros: una forma de registrar su plan de investigación y análisis antes de recopilar ningún dato. Discutiremos qué son los prerregistros y cómo pueden mejorar la calidad de su investigación.

Saludos,
La iniciativa Science4All
[\[Transform4Europe research & innovation project T4ERI\]](#)

Próximos Snacks:

- Preregistration 101
- El valor de las replicaciones

También puedes encontrar todos los aperitivos en nuestro sitio web: <https://transform4europe.eu/t4eri/science4all/>

Snacks pasados:

- ¿Qué es la Ciencia Abierta (Open Science) y en qué le beneficia?
- Mitos sobre la ciencia abierta
- Acceso abierto

Fuentes utilizadas

- (1) <https://opendatahandbook.org/guide/en/what-is-open-data/>
- (2) Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, 150, 33–41. <https://doi.org/10.1016/j.jclinepi.2022.05.019>
- (3) Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>
- (4) Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- (5) <https://www.go-fair.org/fair-principles/>
- (6) <https://www.openaire.eu/how-to-make-your-data-fair>

Los Open Science Snacks de la Iniciativa Science4All están licenciados bajo [CC-BY 4.0](#).



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101035805