

# Data Policy

Héctor G. Ceballos ([ceballos@tec.mx](mailto:ceballos@tec.mx))

Oct 26, 2021

This document describes the policies that will be applied to the data shared by Tec departments through the [IFE Living Lab & Data Hub](#).

## DATA PRIVACY

1. Personal information goes through an un-identification process including (but not limited to) sensible data removal and masquerading of identifiers (e.g. Matricula, Nomina).
2. Textual information such as comments of students are encoded through numerical representations that enables making inference without revealing actual information (e.g. [TF-IDF matrixes](#)).
3. The anonimacy of students and professors is protected by removing data that could be used to infer their identity.

## DATA ACCESS

4. Any researcher from Tecnológico de Monterrey or another institution (national or foreign) can request access to data published by the IFE Data Hub.
5. Access to data will be granted only after the applicant's credentials as a consolidated researcher and a justification of his/her data request, has been verified.
6. Applicants will get access only to the data they requested using a spreadsheet format indicating the required data.
7. Applicants must sign a [Terms of Use](#) letter where they commit to use the data only for the purpose they specify on it.

## DATA PROVENANCE

8. Data provenance is acknowledged through data papers made in co-authorship with data owners.
9. Each data collection is identified through a Digital Object Identifier (DOI), which is internationally accepted.

## DATA USAGE

10. Users must inform the IFE Data Hub about the usage and products (papers) they produce with the given data in order to get access to more data.
11. Data usage is tracked through DOIs and data papers, which must be referenced in any scientific paper based on the provided data.

12. Every semester, the IFE Data Hub will provide statistics on the usage of the data provided by the respective data owner. These statistics include the number of views of the dataset and the number of data requests during the period.

## **DATA PROTECTION**

13. Data is granted under a limited version of the [Open Data Commons Attribution License \(ODC-By\)](#), where users are free to produce works from the database (create) and to modify, transform and build upon the database (adapt), whenever they attribute any public use of the database, or works produced from the database (attribute). Users are not allowed to share the data or adaptations they make of it.
14. Data requests will be denied to researchers that do not comply with informing the usage they made of the data that was previously given to them.

## **FAIR PRINCIPLES**

15. FINDABLE. Data collections will be annotated with metadata for describing its content, and will be assigned a Digital Object Identifier (DOI) for referencing it. The content of data collections will also be documented through a data dictionary and a data teaser (brief example of data contained in the data collection).
16. ACCESSIBLE. Data collections will be published in a [Dataverse platform](#) that enables the protocols for data discovery in international indexes, as well as the authentication and authorization procedures for getting access to data as specified above.
17. INTEROPERABLE. Data will be shared in open formats such as Comma Separated Values (CSV). The description of every variable is specified in a data dictionary using an institution-independent vocabulary.
18. REUSABLE. Data is provided with a [Open Data Commons Attribution License \(ODC-By\)](#). Data provenance is indicated through metadata and acknowledged in data papers that describe each data collection. Data is encoded using international standards (e.g. the [ISO/IEC 5218](#) is used for encoding human sexes).