IRTA — Institute of Agrifood Research and Technology

This data curation checklist is based on data curation best practices from DCN CURATED Steps[1].
This document is a practical guide for data curators to ensure that all datasets submitted to multidisciplinary repositories are consistent to be shared and reproduced.

Step	Question	Items	Check
AUTHORIZATION	Who is in charge of the data?	Identification. Corresponding author behind the dataset is identified [Name, Affiliation and Identifier]	☐
		Contact. Corresponding author behind the dataset contact is provided [Email]	☐
		Right holder of the data is identified. [Specify]	☐
	Is the submitter the creator of the dataset?	Document the creator [Name and Affiliation]	☐
		Is the submitter authorized?	☐
	Has data been reused for the creation of the dataset?	Data origins are documented	☐
		Reuse of the original data is allowed	☐
	Is the dataset license specified?	CC BY license or another is specified	☐
DOCUMENTATION AND FILES	Does the dataset include apropiate documentation?	These can be README files, Codebooks, Data dictionaries....	☐
	Is the summary of data files complete?	NO undocumented. All provided files are detailed in the documentation	☐
		NO missing. All the files acording to the documentation have been provided	☐
		Relationship between files is detailed	☐
		Directory structure is specified, if applicable	☐
	Is the naming convention adequated?	Consistent	☐
		Logical	☐
		Descriptive	☐
		No special characters	☐
		Adequate versioning	☐
		Upper/lowercase letters do not contribute meaning differences	☐
METADATA	Dataset DOI	DOI is provided	☐
	Dataset metadata	Title	☐
		Authors [Name, affiliation and identifiers (ORCID/SCOPUSID..)]	☐
		Description [Summary describing the purpose, nature, provenence and scope of the dataset.] What, what for, where, why, how, who, and repository	☐
		Date [Submited date. Responsibility of the data curator]	☐
		Subject [Domain specific subject provided by the repository]	☐
		Keywords, 3 to 5 (Key terms that describe the dataset)	☐
		Keyword vocabulary [Associated vocabulary and URL, If not associated the data curator will do it]	☐
		Related publication. Handle,URL or DOI [That use the data from this dataset]	☐
		Related datasets	☐
		Kind of data [Drop-down list]	☐
	Content metadata	Domain-specific shemas used [if yes, specify]	☐
		README file template is followed and complete	☐
		Described variables [Description, coding-values, mesurament units, range, linkages]	☐
		Detailed information for transformations	☐
DATA FILES CONTENT	Tabular data	Clarly defined and consistent headers	☐
		Name of variables in, and only, the first row	☐
		Only one table per spreadshead file	☐
		Usability issues [Empty fields, Dummy data, default values, duplicates, privacy sensitive data...]	☐
		Clearly defined codes [missing data, not aplicable...]	☐
		No multivalue columns [Eg. Units can't be in the same column as values; use different columns]	☐
		NO embedded documentation in the data [place them in README file or another column]	☐
		Extract hidden documentation [Eg. Excel embedded notes]	☐
	File formats	Files in open, non-propietary, machine readable formats	☐
		If not open formats, is it possible to transform them to open formats? Do it.	☐
	Software	Documentation specifies which software and version are required to read the files	☐
		If propietary, documentation explains how to obtain it or open alternatives	☐
VALIDITY & QUALITY	Validity	Files open and code runs as expected	☐
		Files are not corrupt	☐
	Quality assurance	Quality control is clearly defined	☐
		Risks related to data acuracy are minimized	☐
OTHERS	Sensitive data	Sensitive data has been removed / anonymized	☐
		If NOT, data is shared as restricted or closed	☐
	Embargo	If applies, is it defined?	☐
	How long is data going to be preserved?	IRTA Dataverse repository provides 10 years of data preservation	☐
Requests	Requests	Describe concerns	☐
		Describe issues	☐
		Describe needed improvements	☐
		Other	☐

[1] Data Curation Network. (2018). Checklist of CURATED Steps Performed by the Data Curation Network, v.1. Data Curation Network. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/226891.