

CIRAD: CIRAD-TEMPLATE-ENG

1. Information about the project

DMP version and publication date

Exemple de réponse:

v2, published on 2018-02-25

Project acronym and complete name

Exemple de réponse:

STRADIV - System approach for the TRAnstition to bio-DIVersified agroecosystems,

BFF - Biomass For the Future

List of the main funders

Exemple de réponse:

Fondation Agropolis, ANR, Ademe, Interreg (Sudoe), etc.

Identifier of the call for proposals

Exemple de réponse:

AAP ACC 2017, AAP PIA2 [cid56862](#), etc.

Topic of the call

Exemple de réponse:

Climate change, Biotechnologies and Bioresources, etc.

Grant agreement number

Exemple de réponse:

ANR-11-BTBR-0006

Project start date

Exemple de réponse:

Recommandations:

ISO format: YYYY-MM-DD

Project end date

Recommandations:

ISO format: YYYY-MM-DD

Project coordinator (name of organization)

Exemple de réponse:

CIRAD

CIRAD project chief scientist

Exemple de réponse:

Surname first name, CIRAD, UR XXX, Montpellier, France, first.name.surname@cirad.fr

Recommandations:

Surname first name, organization, unit or laboratory, town, country, email
[+ d'informations sur l'affiliation](#)

Project goals

Exemple de réponse:

Define the conditions for ecological transition

Keywords

Exemple de réponse:

Biodiversity, Plant anatomy, Remote sensing, Data mining

Recommandations:

Opt for some keywords from Agrovoc ([Thésaurus Agrovoc FAO](#)), to be completed if necessary by specific keywords

Project partners (names of organizations)

Exemple de réponse:

INRA (France), ISRA (Senegal)

Recommandations:

Name of the organization (country).

Only mention partners having contributed to data collection, analysis or processing.

2. Dataset description

Dataset description

Exemple de réponse:

Data collected in the STRADIV design (60*9*15m) in Ivory (Madagascar) + 17 farmers' fields : ITK, plant material, trial design, inputs, biomass (rice, stylo), weeds, soil fauna (monoliths, pit fall trap), white grubs, estimated yield (rice, peanut, sorghum), rice yield components.

Recommandations:

Specify, among other things, the extent, scale, temporal coverage, possible application of a quality process, etc.

Nature of the dataset

Exemple de réponse:

- Experimental or observation data,
- Satellite images,
- Sorghum short read nucleotide sequences of 100 pb, digested by Apekl,
- Nuclear alignments on the sorghum reference genome,
- Small structural variations (SNP, INDEL),
- Surveys,
- Raw or secondary data, physical collections, models, softwares, etc.

Data collection country

Recommandations:

Indicate the country or countries where the data were collected.

Data collection period

Exemple de réponse:

- 2012-2013
- July-September 2015

Dataset production methods

Exemple de réponse:

- Statistical experimental design
- Vegetation index NDVI calculation
- Illumina HiSeq 2000 Sequencing System of the GenoToul (GeT) genomic platform
- Interviews in Dakar following a fixed pattern

Recommandations:

Methods and tools used to acquire and process the data

Data processing tools

Exemple de réponse:

- Excel, Stata, SPSS
- Qgis,
- Matlab v8 with Signal Processing toolbox,
- Tassel 5 GBS v2 Pipeline,
- R...

Recommandations:

Software package(s) used (specify the version if necessary), algorithm, etc.

Estimated volume of data

- Less than 50 MB
- Between 50 and 100 MB
- Between 100 MB and 1 GB
- Between 1 and 10 GB
- Between 10 and 50 GB
- More than 50 GB

Exemple de réponse:

- 100 MB
- 1 GB
- ~ 250 GB
- 5 TB

Recommandations:

Assessment of the expected data size (if known)

Existing publications related to this dataset

Exemple de réponse:

Dépigny S., Noupadja P., Tomekpe K., Lescot T., Bonneviot P., Bakry F., 2016. CARBAP K74: a plantain-like hybrid designed to promote sustainable plantain-based cropping systems, in Van den Bergh I., Risède J.M., Johnson V., 2018. X International Symposium on Banana : ISHS-ProMusa symposium, Agroecological approaches to promote innovative banana production systems, Montpellier, France, pp.63-70 DOI 10.17660/ActaHortic.2018.1196.7

Recommandations:

Specify the reference of the article, with the hyperlink, preferably DOI

Re-use of existing data

Exemple de réponse:

- Statistics from FAOSTAT on sorghum production, free use
- Data from Land Cover Map 2000, Center for Ecology and Hydrology, Forest Research has licence agreement

Recommandations:

If the dataset relies on the re-use of existing data, cite their origin.

If material protected by specific rights is used during the project, specify the access or re-use restrictions.

Are the existing data that you are going to re-use protected by a licence that limits their re-use? Explain, if needed, the terms of re-use and the means used to ensure compliance with those terms.

Will you be adding your new data to these existing datasets? Again, check that the licence allows it.

Intended future uses

Exemple de réponse:

- Multi-site comparison,
- Classification and matching methods for multi-source data,
- Meta-analyses,
- Knowledge extraction by text-mining methods.

Recommandations:

What re-uses of the dataset might be possible?

What might be the prospects for application or development?

In similar data already exist, what is the contribution of this dataset?

Data users

Exemple de réponse:

Scientific community, private sector, general public, policymakers, etc.

Recommandations:

Which audience might potentially be interested in the produced dataset?

3. Standards and metadata

Metadata description

Exemple de réponse:

- The metadata are collected by a scientist and organized according to the MIAPPE standard: general metadata, timing and location, biosource, environment, treatments, experimental design, sample collection, processing, management, observed variables.
- Metadata specific to SPOT images: spatial resolution, spectral field, temporal coverage.

Recommandations:

Which metadata describe this dataset? How were they produced (laboratory notebook, GPS, instrument type, manual input)? Who produced them?

Add a quality score if available.

Metadata standards

Exemple de réponse:

- MIAPPE standard in ISA-Tab format and Crop Ontology

- Survey with DDI standard
- Specific standard proposed by the consortium (see previous section)

Recommandations:

State the standard used, specify whether it is the recommended standard for the discipline. If there is no standard for the discipline, explain the preferred solution. For informations on discipline standards:

- <https://fairsharing.org>

- <http://www.dcc.ac.uk/resources/metadata-standards>

Data format

Exemple de réponse:

CSV, JPEG, TXT, FASTQ, BAM, VCF, etc.

Recommandations:

Opt to use open or widely used formats to facilitate sharing

<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>

File management

Exemple de réponse:

File tree

Each digitized field book is classified in a numbered folder whose name begins with the author's initials; e.g.: "XY_Field Book_1".

Naming rules

Each digitized page is named in the following way: <author's initials>_<book production year>_< ?? >_page_<page number>

Recommandations:

File management and organization, file naming rules, versioning management, etc.

<https://intranet-data.cirad.fr/gerer-ses-donnees/organisation-nommage-et-formats-des-fichiers>

4. Responsibilities, intellectual property rights

Person in charge of data collection

Exemple de réponse:

Surname first name, CIRAD, UR XXX, Montpellier, France, first name.surname@cirad.fr

Recommandations:

Surname first name, organization, unit or laboratory, town, country, email

More information about Cirad addresses:

<https://intranet-dist.cirad.fr/rediger-et-publier/affiliations>

Person in charge of data analysis

Exemple de réponse:

Surname first name, CIRAD, UR XXX, Montpellier, France, first name.surname@cirad.fr

Recommandations:

Surname first name, organization, unit or laboratory, town, country, email

Owner of intellectual property rights

Exemple de réponse:

CIRAD

See the "mode of execution" section in the consortium agreement under the aegis of the Green Biotechnologies GIS for the Biomass for the Future project.

Recommandations:

Who will hold the intellectual property rights on the data produced during the project?

Data ownership is specified in the consortium agreement, whose terms must be negotiated upstream of the project and the DMP: give the hyperlink to the consortium agreement if possible.

Ethical aspects

Exemple de réponse:

- Personal data (collected during surveys), anonymised / aggregated
- Data collected during animal testing validated by an ethics committee
- Research with partners outside Europe : recommendations of the H2020 manual have been respected
- Until the question is settled, omics data from genetic resources obtained in the context of the Nagoya protocol are subject to the same rules, particularly with regard to access and benefit sharing (ABS). It must be ensured, before the end of the project, that the partners have access to the digital resources they are entitled to according to the BFF consortium agreement, and know how to use the associated analysis tools (assess the need for documentation and training).

Recommandations:

Take into account, among other things, the elements presented in the guide accompanying H2020 projects: How to complete your ethics self-assessment (July 2016)

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

5. Data sharing, dissemination and re-use

During the project

Exemple de réponse:

- The dataset produced by the STRADIV 2 unit in Ivory will be deposited in the CIRAD Dataverse and shared with all the project partners.
- The BFF project brings together private and public partners and studies several plants. It will therefore be necessary to detail access to data for the different cases, referring to the consortium agreement.
- Concerning the sorghum study at CIRAD, the GBS data will be accessible to the scientists who are partners in the project.

Recommandations:

Data sharing and access arrangements for the project partners or other persons.

After the project

Exemple de réponse:

- The datasets produced by the STRADIV 2 unit will be published in open access, with an embargo period of 3 years for publication.
- The data will be publicly accessible either via NCBI (fastq) or the BFF portal (Bam and VCF).

Recommandations:

Specify which dataset will be shared.

Data dissemination and access arrangements: open or restricted access, embargo period, etc.

Type of licence

Exemple de réponse:

CC-BY, CC0, Open Knowledge Foundation Open database Licence, open licence, etc.

Recommandations:

Licence type used for dissemination.

For more information, see the CoopIST site (in French) : [Les principales licences de diffusion des jeux de données](#)

Data repository

Exemple de réponse:

CIRAD Dataverse, Huma-Num, Ortolang, NCBI SRA, GenBank, Pangaea, etc.

Recommandations:

Name of the repository where the data will be deposited, if it is identified

See the directory of research data repositories: <http://www.re3data.org>

Dataset identifier assigned by the repository

Exemple de réponse:

DOI

Recommandations:

Different types of digital identifier exist (DOI, Handle, POI).

CIRAD Dataverse assigns a DOI with the following format : [doi:10.18167/DVN1/LWT7BG](https://doi.org/10.18167/DVN1/LWT7BG)

Dissemination date

Exemple de réponse:

2021-01-01

Recommandations:

Date on which the dataset was made accessible in a repository or by a publication.

Date in ISO format: YYYY-MM-DD

Data reading

Exemple de réponse:

- For each dataset in Excel, a tab is used to describe the studied variables.
- JBrowse from the BFF portal for BAM and VCFfiles, otherwise Tabix or Integrated Genome Viewer (IGV).
- The documentation associated with the survey data includes the questionnaire and the pollster manual.

Recommandations:

Specify what documentation or software is needed to understand and access the data (codes, abbreviations, versions of the software for reading, explanatory documents, etc.)

Sensitive data

Exemple de réponse:

- Survey data that have not been anonymised
- SNPs that can be used as part of a marker kit in support of phenotyping.

Recommandations:

Existence of sensitive data justifying waiving of the dissemination principle: protected, personal, strategic data, or data from private partnerships, etc.

6. Archiving and preservation

Storage and backup during the project

Exemple de réponse:

- Personal computer of <name of person>, B2, CIRAD Lavalette, Montpellier
- External drive of the xxx team (yyy research unit).
- The content of the external drive is synchronised with a CIRAD server every Thursday.
- In the storage space of the HPC data centre of the AGAP bioinformatics platform

Recommandations:

Storage media, geographical location, intermediate volumetrics, policy for computer system security

After the project, procedures for long-term preservation

Exemple de réponse:

- Dataset deposited in Zenodo : DOI:10.167/toto21
- The data will be kept in the storage space with backup (gs7k) by HPC data centre of the AGAP bioinformatics platform, and in the NCBI repository
- The climatic data is not reproducible and will be stored in the CIRAD Dataverse.

Recommandations:

Which data have to be preserved over the long term ?

Archiving media (e.g.: local server, distant server, external drive, repositories), location

Recommended lifetime

Exemple de réponse:

BnF: 5 years

Recommandations:

Recommended duration (taking into account the existing legal and/or prescribed requirements)

Final volume of archived data

Exemple de réponse:

- 1 TB
- ~ 250 GB

Recommandations:

Once the data to be archived have been selected, specify their volume.

Data to be destroyed

Exemple de réponse:

- The name of the interviewed farmers will be deleted from the survey on cultural practices
- Removal of the intermediate data analysis files

Recommandations:

Specify whether some types of data will be destroyed (e.g. personal data, according to CNIL recommendations).

Name of person in charge of archiving and preservation

Exemple de réponse:

- Surname first name. organization. unit or laboratory. town. country

- *Surname, first name, organization, unit or laboratory, town, country*
- EMBL
 - BnF

Recommandations:

Name and affiliation of the person in charge of data archiving and preservation. Or the organization producing the repository, if it guarantees perennial archiving.

Associated costs

Exemple de réponse:

- BnF: 1 TB = € 2,545, supplementary archiving on a disk: € 315 exclusive of tax.
- € 300/TB/year.

Recommandations:

Monetary, material and human costs associated with data archiving and preservation, and arrangements made to cover them (in particular after the end of the project).

Costing tool and checklist :

www.data-archive.ac.uk/media/247429/costingtool.pdf