## Government of Canada: GC maDMP

### DMP General Info

Has someone granted approval to this project?

* Yes
* No - default

Provide the name(s) of the person(s) that gave approval.

Provide the position(s) of the person(s) that gave approval.

Provide the email(s) of the person(s) that gave approval.

Provide a description of the approval given.

Provide the status of the approval.

* Approved
* Rejected
* Submitted
* Unknown - default
* Not applicable

Date at which the approval was given.

What is the classification level of this DMP form?

* Unclassified
* Sensitive but unclassified
* Restricted
* Confidential
* Secret
* Top Secret
* Unknown - default

What language is the DMP? Use the ISO 693-3 code.

*Example Answer*:

fra

*Guidance*:

Visit https://en.wikipedia.org/wiki/List\_of\_ISO\_639-3\_codes for a list of codes.

### Project

Add the following information each of the research project(s) that you will be working on and for which the data and work are described in this DMP.

*Guidance*:

Add additional blocks if necessary (i.e more than one research project).

### People

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

*Guidance*:

Add additional blocks if more than one contributor is needed.

### Resources

Specify what technical instruments you are using to collect the data. Provide a name, description, type, and label.

*Guidance*:

Add additional blocks if more than one instrument was used.

To execute the DMP, is additional specialist expertise required?

* No - default
* Yes, trained support staff is available
* Yes, we will be training existing staff
* Yes, we will be hiring new people with additional expertise

What kind of training?

What kind of expertise?

Do you require hardware or software in addition to what is usually available in the institute?

* Yes
* No - default

What specific hardware/software do you need, and why?

Will you be storing physical samples?

* Yes
* No - default

How will the samples be identified?

* Samples will get an internal code
* Samples will receive a universally unique persistent identifier (PID)

Where will the samples be stored

Will you be acquiring data using measurement equipment?

* Yes
* No - default

Specify what data sets you will acquire using measurement equipment

Do you have any non-equipment data capture?

* Yes
* No - default

Will you be collecting questionnaires?

* Yes
* No - default

Will you be collecting case report forms?

* Yes
* No

Will you be collecting data from electronic patient records?

* Yes
* No - default

Please list all non-equipment datasets you will collect

Do you have all tools to couple the necessary data types?

* Yes
* No - default

### Computing environment

Is the data and/or code reliant on the computing environment?

* Yes
* No - default

Provide the following information about the dependencies

*Guidance*:

Copy this block for any additional dependencies

Is there a data integration tool that can handle and combine all the data types you are dealing with in your project?

* Yes
* No - default

Can all data be brought into the same format, e.g. RDF?

* Yes
* No - default

What software will you be using to collect all data?

* We will not use a data warehouse
* TranSMART
* Software not listed here
* We will develop our own (explain why)

Will you be using a shared working space to work with your data?

* Yes
* No - default

Are data that project members store themselves adequately backed up and traceable?

* Yes, protected against both equipment failure and human error
* No - default

Will this work space be run by dedicated specialists?

* Yes
* No - default

What kind of data will you have in your work space?

Do you need the work space to be close to the compute capacity?

* Yes
* No - default

Will you be working with your data in another form than the way it will be archived?

* Yes, archival will require a conversion step
* No, data format will be archived in the same way we work with it - default

How does the storage need change over time?

* Storage needs will be the same during the whole project
* Storage needs are large at the beginning and will be reduced later
* Storage needs are small at the beginning and will grow later
* Storage needs are largest in the middle of the project
* Explore

When will your raw data become available?

* Raw data will come in right at the start
* Raw data will come in during the project

How much of the raw data do you need to keep in the work space?

* Raw data will need to stay in the work space
* Raw data can be cleaned out or archived quickly
* Raw data do not form a major part of the storage needs
* Raw data will only be loaded in the work space on demand from the primary archives

Do your raw data need to be archived?

* Yes
* No, it is also stored elsewhere can be recovered easily
* No, it can be remeasured easily and more cheaply than archiving

Did you plan how much intermediate data you will get during analysis and how long each step needs to be kept?

* The volume of intermediate data will not be significant - default
* A large volume of intermediate data will be in the work space

Is it possible to store intermediate temporary data on a separate (scratch) area that is not backed up?

* We will use the main work space for temporary data
* We can offload intermediate results to a scratch area that is not backed up

Are you sure you will not need a backup of the data stored on the scratch file systems (any scratch you use)?

* Yes
* No

Are you using a filesystem with files and folders?

* Yes
* No - default

Will you use a folder for each sample/subject?

* Yes
* No - default

What is the naming convention for this folder?

Will you use a (sub)folder for each (repeated) analysis?

* Yes
* No - default

What are the naming conventions for the analysis folders?

Will you use a (sub)folder for each step in the analysis workflow?

* Yes
* No - default

What are the naming conventions for these folders?

What appointments have you made about the naming of files?

Will you be storing data in an "object store" system?

* Yes
* No - default

What are the naming conventions for objects in the store?

*Guidance*:

Make sure names are relatively short, and avoid spaces and special characters. You can use underscore characters, and consider using unique identifiers for the samples/experiments, workflow invocations, and processing steps. You can consider to add versioning using the date in YYYYMMDD format.

Will you use a relational database system to store project data?

* Yes
* No - default

How will you handle changes in the data?

* Data in the database will be read-only
* We will be allowing Create, Update and Delete
* Modifications will be made by Expiring the existing data and Adding updated data

Will you use a graph database for data in the project?

* Yes
* No

Will you be storing data in a triple store?

* Yes
* No

Workflow development status

* This has been arranged - default
* More guidance is desired

Will you be exploring parameters to the workflow, or run in bulk?

* We will be exploring parameters to the workflow
* We will be running in bulk
* A bit of both

What data will the workflow developers or implementers use?

* They do not need data
* They can use existing data from another project
* They can use data from our project

When will they have access?

* Data will be available at the start of the project, no waiting needed
* They can start with a subset as soon as our first data comes
* Workflow development will commence as soon as our data is available

How will data security be dealt with?

* Data security is not an issue
* Data will not be on systems owned by workflow developers
* Data will always be encrypted
* We have made other arrangements

List existing software components you will use in the analysis workflow

*Guidance*:

Your workflow may be available in components from different sources. Specify the different parts that you recognize and that you will each acquire in a different way

List new software components you will develop for the analysis workflow

### Funding

Add all the funding that are part of this project.

### Cost

Provides a list of costs related to data management.

### Legal or Ethical Issues

Do any ethical issues exist?

* Yes
* No - default

Describe why there may be ethical issues

Provide how to access the ethical issues report

Are there Indigenous considerations for this data?

* Yes
* No - default

Is there Indigenous community approval for the dataset?

* Yes
* No
* Unknown
* Not applicable

Describe the Indigenous community approval

Describe the control over the data

Describe the governance of the data

Describe the Indigenous considerations

Is there Indigenous control over the data?

* Yes
* No - default
* Unknown

Does Indigenous data governance exist for the data?

* Yes
* No - default
* Unknown
* Not applicable

What classification level has been established?

* Top secret
* Secret
* Confidential
* Restricted
* Sensitive but unclassified
* Unclassified
* Unknown

What data security and privacy measures have been implemented?

What is the protection level of the distribution?

* Protected A
* Protected B
* Protected C
* Not protected
* Unknown

Does the dataset contain personal data?

* Yes
* No
* Unknown

Is there a privacy impact assessment?

* Yes
* No - default

Has the assessment been approved?

* Yes
* No
* Unknown

Has the assessment been approved by [tba]

* Yes
* No
* Unknown

Has the assessment been conducted?

* Yes
* No
* Unknown
* Not applicable

What is the download link for the assessment?

Has the assessment been registered?

* Yes
* No
* Unknown

Are there additional security and privacy measures needed?

* Yes
* No - default

What is the title of the security or privacy need?

Describe the security or privacy need

Does the dataset contain sensitive data?

* Yes
* No
* Unknown

What is the DMP protection level?

* Protected A
* Protected B
* Protected C
* Not protected
* Unknown

Does the distribution contain intellectual property?

* Yes
* No - default

Describe the copyright of the IP

What is the extent of the copyright?

Who is the copyright holder?

What was the IP publication date?

Does the IP contain Indigenous traditional knowledge?

* Yes
* No

Describe the data use constraints

Is there a data sharing agreement?

* Yes
* No
* Unknown

Is there an Indigenous information sharing agreement?

* Yes, all of it
* Yes, part of it
* No - default
* Not applicable
* Unknown

Describe the Indigenous information sharing agreement

### Dataset Metadata

For each piece of metadata, provide the following information:

* Description
* Language of the metadata
* Is there a metadata standard ID?
  + What is the identifier?
  + What type is the ID?
* How will you be collecting and keeping your metadata?

### Dataset

What is the earliest start date of data collection?

What is the latest end date of data collection?

Describe the data collection period

What quality assurance was performed on the data?

Does the dataset have an ID?

* Yes
* No - default

What is the identifier?

What type of ID is used?

* Handle
* DOI
* ARK
* URL
* Other

Describe the dataset

What is the geographic coverage of the dataset?

What is the annual growth of the dataset in terabytes (TB)

Is this dataset related to industry?

* Yes
* No - default

What classification system is used?

What is the industry code?

What is the name of the industry?

On what date was the dataset issued?

What keywords describe the dataset?

*Guidance*:

Separated by commas

What language is the distribution?

Enter the preservation statement

Is there a subject?

* Yes
* No - default

What is the classification of the subject?

What is the indigenous knowledge classification?

What is the title of the dataset?

What type of dataset is it?

### Dataset Distributions

What is the access URL to the distribution?

Until when is the distribution available?

What is the size in bytes of the distribution?

How is access to the data controlled?

* Open
* Shared
* Closed

Is the distribution data aggregated?

* Yes
* No
* Unknown

Describe the distribution data aggregation

What is the data class of the distribution?

* Continuous - default
* Discrete (point data)
* Geospatial (GIS raster)
* Geospatial (GIS vector)
* Longitudinal (panel)
* Qualitative - nominal
* Qualitative - ordinal
* Quantitative - continuous
* Quantitative - discrete
* Time Series

What is the data completeness of the distribution?

* 95-100%
* 90-95%
* 85-90%
* 80-85%
* 75-80%
* 70-75%
* 65-70%
* 60-65%
* 55-60%
* 50-55%
* 45-50%
* 40-45%
* 35-40%
* 30-35%
* 25-30%
* 20-25%
* 15-20%
* 10-15%
* Less than 10%

What is the data criticality of the distribution?

* Business function continuity and improvement
* IM or IT modernization
* Intergovernmental agreement
* Minister / Deputy Minister Priorirty
* Mission Critical
* Policy compliance (departmental)
* Policy compliance (federal)
* Regulatory
* Risk mitigation
* Treasury board submission
* Vote netted revenue (VNR)
* Other program priority

Does the distribution have data integration?

* Yes
* No
* Unknown

Describe the data integration

What are the sources of the data integration?

When was the distribution data last updated?

What is the latest end date for the data?

What is the data priority of the distribution?

* Active data
* High throughput
* Long term storage
* Transitory

What is the size of the data?

For above, what unit is used for the data size

* Bytes - default
* KB - kilobyte
* MB - megabyte
* GB - gigabyte
* TB - terabyte
* PB - petabyte
* EB - exabyte
* ZB - zettabyte
* YB - yottabyte

What is the source of the data in the distribution?

* Field analyzer
* Laboratory instrument
* Modeling-simulation output
* Qualitative observations
* Quantitative measurements
* Satellite
* Text document
* Survey
* Wildlife tracking device

Describe the data source of the distribution

Is the data source external?

* Yes
* No - default
* Unknown

Describe the external data source

What is the start date of the data?

What is the maximum latency for data transfer in msec?

What is the minimum throughput for the data transfer in Mbps?

What type of data is it?

* Administrative
* Audit
* Calibration
* Chemical
* Citizen science
* Climate
* Emissions
* Human resources
* Inventory
* Medical
* Monitoring network
* NLP (natural language processing)
* QA-QC
* Operational
* Regulatory-legislative
* Repurposed data
* Pilot Study
* Research & development
* Spill
* Surveillance

How frequently is data updated?

Describe the distribution

What is the download URL for the distribution?

What is the format of the distribution?

* CSV
* Excel
* JSON

What geodetic datum was used?

* ED50 (older European)
* ETRS89 (European)
* GDA94 (Australian)
* JGD2011 (Japanese)
* KGD2002 (Korean)
* NAD27 (older North American)
* NAD83 (North American)
* OSGB36 (Ordnance Survey of Great Britain)
* Tokyo97 (Japanese)
* TWD97 (Taiwan)
* WGS 84 (World Geodetic System)

Does the distribution use indigenous data?

* Yes
* No - default
* Unknown

What is the Indian band number?

What Indigenous data identification method is used?

* Indian band
* Self-identification
* Not applicable
* Other

What is the linked data star rating?

* 5 stars
* 4 stars
* 3 stars
* 2 stars
* 1 star

What quality control level is used?

* Level 0 - raw
* Level 1
* Level 2
* Level 3
* Mixed
* Other
* Unknown

What is the shared drive address?

What is the status of the distribution?

* Finished
* Minor issues
* Major risk encountered
* On track
* Not started

Describe the distribution status

What is the title of the distribution?

Provide the following information for your physical data assests, if any are needed

* What is the city where the physical data asset is located?
* What is the building name where the physical data asset is located?
* What is the room number of the physical data asset?
* What is the full computer name of the physical data asset?
* Describe the physical data asset
* What type of physical data asset is used?

### Host

What is the availability of the system?

*Guidance*:

Answer in percentage, such as 99.5%

How frequently is the system backed up?

*Guidance*:

For example, daily, weekly, biweekly, etc.

What type of backup is performed?

What is the system certified with?

* coretrustseal
* din31644
* dini-zertifikat
* dsa
* iso16363
* iso16919
* trac
* wds
* Not certified

Describe the system

Where is the system located? Enter the 2-digit country code.

What PID system is used?

* ark
* arxiv
* bibcode
* doi
* ean13
* eissn
* handle
* igsn
* isbn
* issn
* istc
* lissn
* lsid
* pmid
* purl
* upc
* url
* urn
* Other
* No PID system

What type of storage is used?

Does the host support checksum?

* Yes
* No
* Unknown

Does the host support versioning?

* Yes
* No
* Unknown

What is the title of the host?

What is the URL of the host?

What is the minimum number of CPU cores needed?

What is the minimum CPU speed in Ghz needed?

What is the minimum type of machine?

* Desktop workstation
* Laptop
* Server
* Supercomputer

What is the minimum amount of RAM needed in GB?

### Disposition and Retention

Does this dataset have archival value?

* Yes, all of it
* Yes, part of it
* No - default

Describe the archival value

Are there any legal issues with the retention of this dataset?

* Yes
* No - default
* Unknown

Describe any possible legal issues

List any other remaining distributions of the data

Will it be necessary to destroy the dataset?

* Yes, all of it
* Yes, part of it
* No - default
* Unknown
* Not applicable

Describe why destruction is necessary

Will this dataset need to be used perpetually?

* Yes, all of it
* Yes, part of it
* No - default
* Unknown
* Not applicable

Describe why perpetual use is needed

What will trigger a review of the retention of the dataset?

* A fixed date
* Not applicable (transitory record)
* Not applicable (permanent retention)
* Other
* Unknown (based on Electronic records as documentary evidence s. 6.4.5.)

What date triggers a review of the retention?

Describe the retention review trigger for the data

Has a disposition action been planned?

* Yes
* No - default

Has the archival value of the dataset been considered?

* Yes
* No
* Unknown

Have legal issues regarding its disposition been considered?

* Yes
* No
* Unknown

Is the destruction required?

* Yes, all of it
* Yes, part of it
* No
* Not applicable
* Unknown

Is perpetual use required for the dataset?

* Yes
* No
* Unknown

What is the retention review trigger date?

Has a disposition action been completed?

* Yes
* No - default

What is the authorization number for the disposition action?

On what date was the disposition action completed?

Describe the disposition action

What type of disposition action was performed?

* Destroyed
* Alienated from government holdings
* Transferred to national archives
* Transferred to another government department or agency