

# Guidance for Review of Data Management Plans Submitted with Applications and Proposals for NIST Funding for Scientific Research

NIST G 5701.00

Issue Date: 12/10/2020

Effective Date: 11/29/2017

## PURPOSE

This document establishes guidance for review of data management plans (DMPs) submitted with applications and proposals for financial assistance and other types of federal funding for the conduct of scientific research, as well as for monitoring compliance with special award conditions.

## APPLICABILITY

This guidance is provided for Federal Program Officers and other NIST staff who review financial assistance proposals and other applications for NIST funding for the conduct of scientific research.

## REFERENCES

- [Department of Commerce Grants and Cooperative Agreements Manual](#)
- [NIST Policy 5700.00 Managing Public Access to Results of Federally Funded Research](#)
- [NIST Order 5701.00 Managing Public Access to Results of Federally Funded Research](#)
- [NIST Procedure 5702.01 Managing Public Access to External Research Funded by NIST](#)

## GUIDANCE

Notices of Funding Opportunities (NOFOs)<sup>1</sup> require that applicants for federal funding for the conduct of scientific research submit a data management plan (DMP) that includes, at a minimum:

1. a summary of activities that generate data
2. a summary of the types of data generated by the identified activities
3. a plan for storage and maintenance of the data generated by the identified activities, and

---

<sup>1</sup> <https://www.nist.gov/about-nist/funding-opportunities>

4. a plan describing whether and how data generated by the identified activities will be reviewed and made available to the public, both during and after the award's period of performance.

All applications for activities that will generate scientific data using NIST funding are required to adhere to a DMP or explain why data sharing and/or preservation are not within the scope of the project. A template for DMP generation and a sample DMP are provided in Appendices A and B, respectively. An applicant is not required to use the template as long as the DMP contains the required information.

All DMPs must comply with the requirements of NIST Policy 5700.00, *Managing Public Access to Results of Federally Funded Research* and NIST Order 5701.00 *Managing Public Access to Results of Federally Funded Research* or any successor Policy and Order.

The DMP must be evaluated against the rubric provided in Appendix C. Data is expected to be made publicly available unless such a release is not permitted (e.g., data contains personally identifiable information (PII) or business identifiable information (BII)). If data will not be made publicly available, justification must be provided.

After selection has been made, the FPO should work with the NIST Grants Office to have the applicant revise their DMP as needed.

## **DIRECTIVE OWNER**

602 – Special Programs Office

## **APPENDICIES**

- A. Template for DMP to Accompany Request for NIST Funding
- B. Example DMP
- C. Rubric for Evaluation of DMPs
- D. Revision History

## APPENDIX A

### TEMPLATE FOR DATA MANAGEMENT PLAN (DMP) TO ACCOMPANY REQUEST FOR NIST FUNDING

Your DMP should explain how you plan to manage and make available to the public the data generated in your proposed project. The plan must include sufficient detail for evaluation of its appropriateness in a review conducted by someone who is unfamiliar with your work.

Your DMP is required to contain four sections, which are listed below along with types of questions you should consider when developing your plan. These questions are intended to clarify the type of information expected.

#### **I. Summary of activities that generate RESEARCH data**

Provide a brief explanation of research activities that generate data. Research data is defined as the recorded factual materials commonly accepted in the scientific community as necessary to document and support research findings.

Consider these questions:

- What research are you doing?
- How will you collect, capture, or create the data? (include methods, instruments, software, and infrastructure if relevant.)
- Will you be using data from other data sources?

#### **II. Data types and classification**

Provide a summary of the types of data generated by the activities identified above.

Consider these questions:

- What data will you collect, capture, or create in the research?
- What data will be shared?
- What metadata and documentation will be provided with the data in order to make it reusable?
- What contextual details (metadata) are needed to make the data you capture or collect meaningful?
- How will you make these details available?

#### **III. Preservation**

Provide a plan for storage and maintenance of the data both during and after the award's period of performance. In some cases, NIST may be willing to make awardees' data available through NIST's data management infrastructure; see

[https://www.nist.gov/system/files/documents/2018/06/19/final\\_g\\_5702\\_01.pdf](https://www.nist.gov/system/files/documents/2018/06/19/final_g_5702_01.pdf). (Note that financial records, supporting documents, statistical records, and all other non-Federal entity records pertinent to an award must be retained for at least 3 years, consistent with 200.334 - <https://ecfr.federalregister.gov/current/title-2/subtitle-A/chapter-II/part-200/subpart-D/subject-group-ECFR0e38f1d21dd8c05/section-200.334>.)

Consider these questions:

- How will data be stored during the period of performance?
- How will data be stored and archived after the period of performance?
- How will the data be transitioned from short-term storage to long-term storage and archival?
- What is the format of the data files?
- Are the files in a non-proprietary, machine-readable, machine-actionable format?
- How much storage will be required?

#### **IV. PROVIDING PUBLIC ACCESS**

Provide your plan for making your data discoverable and available to the public unless such a release is not permitted. You are expected to take reasonable measures to safeguard protected personally identifiable information and other sensitive information consistent with applicable laws. If data will not be made publicly available, justification must be provided, and a waiver granted.

Consider these questions:

- How will the data be made publicly available? (If it won't be, why not?)
- Are there risks associated with disclosure of the data? (If so, how will they be addressed?)
- How will the data be reviewed prior to release?
- How will the data be made discoverable and accessible?
- Where will the data be published or deposited?
- When will the data be published or deposited?
- For how long will it be made available to the public?
- Will the data be provided in a non-proprietary, machine-readable, and machine-actionable format?
- How is the public permitted to re-use your data?

## APPENDIX B

### EXAMPLE DMP

#### I. Summary of activities that generate RESEARCH data

Different varieties of edible oils will be purchased in local grocery and natural foods stores. The types of oils tested will depend on local availability. We will measure vitamin E as individual tocopherols in these oils using liquid chromatography with fluorescence detection. Oils will be weighed (mass manually recorded in lab notebook), spiked with an internal standard (mass manually recorded), and saponified following the procedure in [J AOAC Int.](#) 1999 Mar-Apr; 82(2):288-96. Tocopherols will be extracted. Tocopherols and the internal standard will be separated and measured using a reversed-phase chromatographic column with fluorescence detection as described in the above paper. Software associated with the instrument will record the fluorescence signal as peaks of interest elute from the column, storing the chromatogram on the computer's hard drive. Peak areas of analytes and internal standards will be integrated and peak areas transferred to a spreadsheet, where they will be related mathematically to calibration response factors to calculate concentrations in the oil samples. Calibrant concentrations will be determined spectrophotometrically with corrections made for purity.

#### II. Data types and classification

We will collect the following: mass of oil sample, mass of internal standard, peak areas of chromatographic peaks in samples and calibrants, absorbance of calibration solutions, analyte purity in calibration solutions. These will be used to calculate tocopherol concentrations in the oil samples. The tocopherol concentrations (mean and standard deviation) in oil samples will be published. Individual values that were used for calculation of summary statistics will be provided upon request. Analytical methodology will be provided in the publication that accompanies the data. A readme.txt file will be prepared if there are any requests for data in order to explain the meaning of columns names and units for tables of tocopherol concentrations in edible oils. The paper itself will be available from the journal as well as through NIST's publications repository, accessible via [www.nist.gov/publications](http://www.nist.gov/publications).

#### III. Preservation

Raw data in the form of chromatograms and integrated peak areas will be stored on the hard drive of the instrument's computer. Integrated peak areas, sample masses, and derived data (i.e., concentration calculations) will also be stored in spreadsheets on Elwood, an OISM-managed file server that is backed up nightly. Derived data (i.e., integrated peak areas and concentration calculations in spreadsheets) will be stored on Elwood. A spreadsheet containing final concentrations and summary statistics will be saved as a .csv file and uploaded to Midas, which is also managed by OISM. Peak areas are exported by the data acquisition system to a file saved to Elwood. Data in that file is combined with manually entered numeric data and manipulated in spreadsheets that are also stored in Elwood. The spreadsheet ultimately uploaded from Elwood to Midas. Peak areas are exported by the data acquisition system to a file saved to Elwood. Data in that file is combined with manually entered numeric data and manipulated in spreadsheets that are also stored in Elwood. The spreadsheet ultimately uploaded from Elwood to Midas. Chromatograms in a proprietary format are stored on the hard drive of the instrument. Peak areas are stored in Excel files and backed up to Elwood nightly. Data made available to the public will be in the form of a .csv file accompanied by a readme.txt file. Less than 1 megabyte of storage required for final .csv file stored by Midas.

#### IV. PROVIDING PUBLIC ACCESS

The paper will indicate that data is available upon request. A record to indicate data availability will be created on Midas, which will make the data locatable through data.gov as well. Initially data will be made available to the public by request. If a request is made, a readme.txt file will be developed, and the dataset will be made publicly available via data.nist.gov. Oils will be described generically (e.g., 'sunflower oil' without identification of brand names or manufacturers. There are no PII, BII, deemed exports, or economic or security risks associated with disclosure of this data. Paper and data will be reviewed in accordance with NIST Order 1801.00, <https://inet.nist.gov/adlp/directives/review-fundamental-research-communications>. Data will be archived in perpetuity, consistent with federal records retention requirements for public data. The public will be able to access the data as long as we determine the data to be valid. Summary statistics will be made available in a paper in a peer-reviewed journal. Underlying data will be stored in Midas and can be made available to the public upon request. Data will be made available to the public through nist.data.gov if one request is received. Data will be deposited when the publication is submitted to a journal for consideration, expected by December of this year. Data will be archived and can be made available in perpetuity. See <https://www.nist.gov/topics/data/public-access-nist-research/copyright-fair-use-and-licensing-statements-srd-data-and> for licensing and fair use.

## APPENDIX B

### RUBRIC FOR EVALUATION OF DATA MANAGEMENT PLANS

#### I. SUMMARY OF ACTIVITIES THAT GENERATE DATA

Performance Criteria	Complete	Incomplete	Did not address
1. Describes the project	Clearly describes the project “We will measure vitamin E as individual tocopherols in edible oils using liquid chromatography with fluorescence detection.”	Some description of the project is included, but DMP is missing information or wouldn’t be well understood by someone outside of the project "Tocopherols in oils."	No information about the project is included
2. Describes how data will be collected, captured, or created	Clearly describes how data will be collected, captured, or created, including methods, instruments, software, or infrastructure where relevant	Missing some details regarding how some of the data will be produced; makes assumptions about reviewer knowledge of methods or practices “LC-fluorescence with an internal standard.”	Does not include information regarding how the data will be collected, captured or created

	<p>“Commercially available edible oils will be weighed (mass manually recorded in lab notebook), spiked with an internal standard (mass manually recorded), and saponified following the procedure in <u>J AOAC Int.</u> 1999 Mar-Apr; 82(2):288-96. Tocopherols will be extracted. Tocopherols and the internal standard will be separated and measured using a reversed-phase chromatographic column with fluorescence detection. Software associated with the instrument will record the fluorescence signal as peaks elute from the column, storing the chromatogram on the computer’s hard drive. Peak of analytes and internal standards will be integrated and peak areas transferred to a spreadsheet, where they will be related mathematically to the relative response factors for calibrants to calculate concentrations in the oil samples. Calibration concentrations will be determined spectrophotometrically with corrections made for purity.”</p>		
--	--	--	--



## II. DATA TYPES AND CLASSIFICATION

Performance Criteria	Complete	Incomplete	Did not address
3. Describes the data that will be collected, captured, or created	<p>Clearly describes the data that will be collected, captured, or created</p> <p>“We will collect the following: mass of oil sample, mass of internal standard, peak areas of chromatographic peaks in samples and calibrants, absorbance of calibration solutions, analyte purity in calibration solutions. These will be used to calculate tocopherol concentrations in the oil samples.”</p>	<p>Missing some details regarding data that will be collected, captured, or created; makes assumptions about reviewer knowledge</p> <p>"Tocopherol concentrations."</p>	Does not describe the data that will be collected, captured, or created
4. Describes what data will be shared	<p>Clearly describes the data that will be shared</p> <p>“The tocopherol concentrations (mean and standard deviation) in edible oil samples will be published. Individual values that were used for calculation of summary statistics will be provided upon request.”</p>	<p>Missing some details regarding what data will be shared</p> <p>“Concentrations.”</p>	Does not describe data that will be shared
5. Describes what metadata and documentation will be provided with the data in order to make it reusable	<p>Clearly describes the documentation and metadata that will be provided to make the data reusable</p> <p>“Analytical methodology will be provided in the publication that accompanies the data. A readme file will explain the meaning of columns names and units for tables of tocopherol concentrations in edible oils for requested data.”</p>	<p>Missing some details that would make the data reusable</p> <p>“In publication.”</p>	Does not include information regarding metadata or documentation

6. Describes other contextual details that are needed to make the data captured or collected meaningful	<p>Clearly describes any other contextual details needed to make the data meaningful</p> <p>“Different varieties of edible oils will be purchased in local grocery and natural foods stores. The types of oils tested will depend on local availability.”</p>	<p>Missing some details regarding contextual details that would make the data more meaningful</p> <p>“Tocopherol concentrations.”</p>	Does not include information regarding contextual details
7. Describes how the metadata, documentation, and other details will be provided	<p>Clearly describes how the metadata, documentation, and other details will be shared</p> <p>“A paper published in a peer-reviewed journal will be linked, via DOIs, to data, metadata, and a readme file made available through data.nist.gov. The paper itself will be available from the journal as well as through NIST’s publications repository, accessible via www.nist.gov/publications.”</p>	<p>Missing some details describing how the metadata, documentation, and other details will be shared</p> <p>“Will publish.”</p>	Does not include information regarding how the data will be collected, captured or created

### III. PRESERVATION

Performance Criteria	Complete	Incomplete	Did not address
----------------------	----------	------------	-----------------

8. Provides a short-term storage plan	<p>Clearly describes how the data will be stored in the short term</p> <p>“Raw data in the form of chromatograms and integrated peak areas will be stored on the hard drive of the instrument’s computer. Integrated peak areas, sample masses, and derived data (i.e., concentration calculations) will also be stored in spreadsheets on Elwood, an OISM-managed file server that is backed up nightly.”</p>	<p>Missing some details regarding short-term storage of the data</p> <p>"Computer"</p>	Does not describe short-term storage of the data
9. Provides a long-term storage/archival plan	<p>Clearly describes how the data will be stored/archived in the long term</p> <p>“Derived data (i.e., integrated peak areas and concentration calculations in spreadsheets) will be stored on Elwood. A spreadsheet containing final concentrations and summary statistics will be saved as a .csv file and uploaded to Midas, which is also managed by OISM.”</p>	<p>Missing some details regarding long-term storage of the data</p> <p>“CD”</p>	Does not describe long-term storage of the data
10. Provides a plan for transitioning from data acquisition to short-term storage to long-term storage and preservation of the data, including after the end of the award’s period of performance	<p>Clearly describes how data will be moved among acquisition and storage systems</p> <p>“Peak areas are exported by the data acquisition system to a file saved to Elwood. Data in that file is combined with manually entered numeric data and manipulated in spreadsheets that are also stored in Elwood. The spreadsheet ultimately uploaded from Elwood to Midas.”</p>	<p>Missing some details regarding the movement of data among systems</p> <p>“Drag and drop.”</p>	Does not include information regarding the movement of data among systems

11. Identifies the format of data that will be stored	<p>Clearly identifies format(s) of data that will be stored, including use of any domain-specific metadata standards, and identifies any complicating factors that would prevent publication of data in a non-proprietary machine-readable, machine-actionable format</p> <p>“Chromatograms in a proprietary format are stored on the hard drive of the instrument. Peak areas are stored in Excel files and backed up to an institutional drive nightly. Data made available to the public will be in the form of a .csv file accompanied by a readme.txt file.”</p>	<p>Only partially describes data formats that will be used for storing data and/or the rationale or complicating factors</p> <p>"System software and Excel"</p>	Does not describe data formats that will be used for storing data and does not explain rationale or complicating factors
12. Provides an estimate of the amount of data storage required	<p>Provides a quantitative estimate of the amount of data storage required</p> <p>“Less than 1 megabyte of storage required for final .csv file stored by Midas”</p>	<p>Missing a quantitative estimate of the amount of storage required</p> <p>“Spreadsheets.”</p>	Does not provide an estimate of storage requirements

#### IV. PROVIDING PUBLIC ACCESS

Performance Criteria	Complete	Incomplete	Did not address
----------------------	----------	------------	-----------------

13. Describes how the data will be made publicly available, including whether it will be released to the public or shared with restrictions	<p>Provides details about how the data will be made publicly available</p> <p>“The paper will indicate that data is available upon request. A record to indicate data availability will be created on Midas, which will make the data locatable through data.gov as well. Initially data will be made available to the public by request. If a request is made, a readme.txt file will be developed, and the dataset will be made publicly available via data.nist.gov.</p>	<p>Provides limited information about how the data will be made publicly available</p> <p>"By request."</p>	Does not describe how the data will be made publicly available
14. Describes any risks associated with disclosure of the data; describes security measures to prevent unauthorized access or other means by which risks will be mitigated, privacy protected, etc.	<p>Clearly describes risks associated with disclosure of the data</p> <p>“Oils will be described generically (e.g., ‘sunflower oil’ without identification of brand names or manufacturers. There are no PII, BII, deemed exports, or economic or security risks associated with disclosure of this data.”</p>	<p>Missing some details regarding risks associated with disclosure of the data</p> <p>“Not applicable.”</p>	Does not describe risks associated with disclosure of the data
15. Describes how the data will be reviewed before it is made available to the public	<p>Clearly describes how data will be reviewed before it is made available to the public</p> <p>“Paper and data will be reviewed in accordance with NIST Order 1801.00, <a href="https://inet.nist.gov/adlp/directives/review-fundamental-research-communications">https://inet.nist.gov/adlp/directives/review-fundamental-research-communications</a>.”</p>	<p>Missing some details describing how data will be reviewed before it is made available to the public</p> <p>“Peer review.”</p>	Does not describe how data will be reviewed before it is made available to the public

16. Describes how the data will be made discoverable and accessible if it is not being published	<p>Provides details about how the data will be made discoverable and accessible if it is not being published</p> <p>“Data will be archived in perpetuity, consistent with federal records retention requirements for public data. The public will be able to access the data as long as we determine the data to be valid.”</p>	<p>Provides vague information about how the data will be made discoverable and accessible if it is not being published</p> <p>“Contact PI.”</p>	Does not describe how data will be made discoverable and accessible if it is not being published
17. Describes where the data will be published or deposited	<p>Provides details describing where the data will be published or deposited</p> <p>“Summary statistics will be made available in a paper in a peer-reviewed journal. Underlying data will be stored in Midas and can be made available to the public upon request. Data will be made available to the public through nist.data.gov if one request is received.”</p>	<p>Provides indeterminate information about where the data will be published or deposited</p> <p>“Data will be deposited.”</p>	Does not describe where data will be published or deposited
18. Describes when the data will be published or deposited	<p>Clearly describes when the data will be published or deposited</p> <p>“Data will be deposited when the publication is submitted to a journal for consideration, expected by December of this year.”</p>	<p>Provides indeterminate information about when the data will be published or deposited</p> <p>“Data will be deposited in our repository after completion of the project.”</p>	Does not describe when the data will be published or deposited
19. Describes how long the data will be made available to the public.	<p>Clearly describes how long the data will be made available to the public</p> <p>“Data will be archived and can be made available in perpetuity.”</p>	<p>Provides indeterminate information about long data will be made available to the public</p> <p>"Data will be deposited in our repository after completion of the project.”</p>	Does not describe how long data will be made available to the public

20. Describes copyright, fair use, and licensing of the data.	<p>Clearly describes copyright, fair use, and licensing of the data.</p> <p>“See <a href="https://www.nist.gov/topics/data/public-access-nist-research/copyright-fair-use-and-licensing-statements-srd-data-and">https://www.nist.gov/topics/data/public-access-nist-research/copyright-fair-use-and-licensing-statements-srd-data-and</a>.”</p>	<p>Missing some details or assumes knowledge regarding copyright, fair use, and licensing of the data.</p> <p>“Federal data.”</p>	Does not describe risks associated with disclosure of the data
---	--	---	--

## APPENDIX C

### REVISION HISTORY

+	Date	Responsible Person	Description of Change
Initial	6/9/2017	Katherine Sharpless	Initial draft
Rev. 0.1	11/22/2017	Dan Cipra	Formatting updates only
Rev. 0.2	1/24/2020	Katherine Sharpless	Replaced narrative guidance with a DMP template, example language, and rubric.