Ref. Ares(2018)4839737 - 20/09/2018

Beyond Planck -- delivering state-of-the-art observations of the microwave sky from 30 to 70 GHz for the next decade - Final review DMP

1. Data summary

State the purpose of the data collection/generation

Explain the relation to the objectives of the project

Specify the types and formats of data generated/collected

Specify if existing data is being re-used (if any)

Specify the origin of the data

State the expected size of the data (if known)

Outline the data utility: to whom will it be useful

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Outline the identifiability of data and refer to standard identification mechanism. Do

you make use of persistent and unique identifiers such as Digital Object Identifiers?

Outline naming conventions used

Outline the approach towards search keyword

Outline the approach for clear versioning

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Specify how the data will be made available

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Specify where the data and associated metadata, documentation and code are deposited

Specify how access will be provided in case there are any restrictions

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Describe data quality assurance processes

Specify the length of time for which the data will remain re-usable

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Clearly identify responsibilities for data management in your project

Describe costs and potential value of long term preservation

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Beyond Planck -- delivering state-of-the-art observations of the microwave sky from 30 to 70 GHz for the next decade - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

Explain the relation to the objectives of the project

Specify the types and formats of data generated/collected

Specify if existing data is being re-used (if any)

Specify the origin of the data

State the expected size of the data (if known)

Outline the data utility: to whom will it be useful

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Outline the identifiability of data and refer to standard identification mechanism. Do

you make use of persistent and unique identifiers such as Digital Object Identifiers?

Outline naming conventions used

Outline the approach towards search keyword

Outline the approach for clear versioning

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Specify how the data will be made available

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Specify where the data and associated metadata, documentation and code are deposited

Specify how access will be provided in case there are any restrictions

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Describe data quality assurance processes

Specify the length of time for which the data will remain re-usable

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Clearly identify responsibilities for data management in your project

Describe costs and potential value of long term preservation

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Beyond Planck -- delivering state-of-the-art observations of the microwave sky from 30 to 70 GHz for the next decade - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

The data employed in this project were taken by the ESA funded Planck satellite between 2009 and 2013, and in particular by the Low-Frequency Instrument (LFI) aboard that satellite, observing at 30, 44 and 70 GHz. These data provide some of the strongest constraints on the early universe physics available to date, and represents a foundation for modern cosmology.

These data are publicly available from via the Planck Legacy Archive (*https://pla.esac.esa.int/*) in several forms, including unprocessed time-ordered data; calibrated time-ordered data; detector and frequency maps; and cosmological and astrophysical component maps. However, the currently available maps show hints of unresolved systematics, particularly in the form of calibration errors. The main goal of the current project is to derive new detector, frequency and component maps from the raw time-ordered data, and therefore release the full scientific potential of the Planck LFI observations. This will be done by integrating calibration, map making and component separation, such that what is learned from one stage can be fed into the following stages.

Both time-ordered data and sky maps are available in FITS-formatted data files, which is an effective standard in the community today; see https://heasarc.gsfc.nasa.gov/fitsio/ for details regarding this library. In addition, all sky maps are defined with the HEALPix pixelization; see https://healpix.jpl.nasa.gov/. This is by far the most widely used sky pixelization

scheme in cosmology today.

The total data volume of the raw time-ordered data is about 3TB, including both detector readouts and pointing. The number of derived sky maps is O(100), each of which requires 300MB, for a total of O(30 GB) volume in terms of final map products. In addition, we will provide new calibrated time-ordered data, which will require another 3 TB of long-term storage.

Once completed, these products will be fed back into the Planck Legacy Archive, after proper vetting by the Planck Science Team. These products will then be useful to the entire cosmological community for the foreseeable future.

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

As stated in the introduction, both the original data and final products will be available through the Planck Legacy Archive. This is a long-term platform supported by the European Space Agency that still undergoes significant development. In particular, it supports a wide range of search options and data exploration facilities, and our products will be naturally integrated into this environment.

We will adopt a similar naming convention for our final map products as the official Planck products, for instance taking the form:

BP_SkyMap_030_1024_R1.00_full.fits

for frequency sky maps, and

COM_BP_CompMap_CMB_2048_R1.00.fits

for component maps. This convention will ensure that external users are both familiar with the naming convention, but the BP tag will still allow direct recognition that these products are new.

Likewise, we will adopt the same convention regarding version numbering as Planck, marking each file with RX.YY, where X is major release number, and YY is minor release number. However, we expect there will be only one version.

Regarding meta-data, we will employ precisely the same information as the official Planck products, specifying all relevant information in the FITS headers.

2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

Both the raw input and final data products will be made publicly available through the Planck Legacy Archive interface. Intermediate processing products derived during the testing and burn-in phase of the project will be kept closed, simply because these are not interesting or useful to external users, and they will require large amount of disk spaces.

The methods and software tools required to access the data are provided through the CFITSIO and HEALPix libraries, which are both OpenAccess software.

2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

The data will be bully interoperable through the Planck Legacy Archive, and they will be fully operable with industry standard codes. All data will include standard meta-data that will allow processing with the same tools as the official Planck products.

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

The data will be provided with full access rights, fully analogous to the existing Planck observations. The data will be made available at the end of the project period, after careful vetting and quality control. There will be no additional embargo period beyond this.

The data will remain re-usable as long as the Planck Legacy Archive is available, which is anticipated to be indefinitely in some form or other.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

The only cost associated with making these data FAIR though the Planck Legacy Archive is associated with the manpower for actually uploading the data to the database. This work will be done in close collaboration with the PLA managers.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

The data will be maintained on back-up servers at the University of Oslo, as well as on the Planck Legacy Archive.

There are no sensitive data involved in this project.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

There are no significant legal or ethical aspects regarding the data used in this project.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Not applicable.