Request to Archive
With The National Climatic Data Center
For Total Solar Irradiance (TSI) and Solar Spectral Irradiance (SSI) FCDR
Provided by University of Colorado & NRL

2014-07-24

This information will be used by the data center to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Odele Coddington
University of Colorado, LASP
303 492 9318
odele.coddington@lasp.colorado.edu

2. Summarize the scope of data to be archived. Give an overview of the data collection as a whole and the types of outputs within the collection. Describe the data variables, including their measurement resolution and coverage.

The Total Solar Irradiance (TSI) and Solar Spectral Irradiance (SSI) FCDR provide daily values of the TSI and the concurrent (self consistent) SSI on multiple time scales, from days to centuries, in wavelength bands designed to be suitable for use in climate and atmosphere models. A composite observational record of daily TSI is constructed by cross calibrating extant observations from 1978 to the present. Historical time series are calculated using the NRL-developed models. Provided are daily and monthly TSI and SSI from 1882 to the present, and annual TSI and SSI since 1615. The data set covers, in aggregate, the period from the beginning of the acceptable historical index record to the last achievable processing period prior to the CDR delivery date. The data sets will be delivered to NCDC in netCDF4 file format following CF metadata conventions. The FCDR delivery will conform to CDRP requirements for source code, data and metadata, and utilize a data delivery mechanism to be agreed upon in a CDR Submission Agreement.
One additional file will contain three reference solar spectra: the solar spectral irradiance averaged over the past 30 years, and for representative minimum and maximum conditions.

3. Name the organization or group responsible for creating the dataset.

LASP (Odele Coddington, Peter Pilewskie, and Marty Snow) & Naval Research Laboratory (Judith Lean)

4. Describe the level to which the data are processed. For example, are these unprocessed raw observations, derived parameters, quality controlled or inter-calibrated data, etc.?

Derived, Level 3

5. List the input datasets and ancillary information used to produce the data.

- Mg II Index: SORCE Solstice, GOME (Univ. Bremen), GOME2 (Univ. Bremen)
- USAF White Light: Sunspot Regions Obtained from NOAA/NGDC-Bill Denig
- Fixed scaling coefficients to convert the input Mg II index to the scale of the NOAA Mg II index composite (Viereck et al., 2004)
- Fixed Beta Sun Angle (generic sun orientation) of one value per day throughout a generic year

6. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 1615

Ongoing as continuous updates to the data record

7. Edition or version number of the dataset:
Version 1, Revision 0

8. Describe its relationship to other archived datasets, such as earlier versions or related source data.
None

9. Approximate date when the dataset was or will be released to the public:
2015-01-01

10. Has the dataset undergone user evaluation and/or an independent review process?
The technical aspects of the composite construction and model formulation have been published in multiple peer-reviewed journal papers. The corresponding models, designated NRLTSI and NRLSSI, have been used as inputs for many climate and atmospheric model simulations, including IPCC AR4 and AR5 (e.g., Schmidt et al. 2011). The NRLTSI2 and NRLSSI2 models, which will be transitioned, are revised versions of the NRLTSI model, which is based on a composite record of TSI from 1978 to 2003 (Fröhlich and Lean 2004) and the NRLSSI model, which is based on UARS UV (120 to 400 nm) spectral irradiance observations from 1992 to 1994 (Lean et al. 1997). Theoretical estimates of wavelength-dependent sunspot and facular contrast functions are used in the region 400 100000 nm (Lean 2000, Lean et al. 2005). Lean (2000) describes the various indices available for use in irradiance variability models. Initially, the long-term irradiance changes were estimated from the variability of Sun-like stars (Lean, White and Skumanich 1995; Lean, Beer and Bradley 1995) then subsequently revised following flux transport model calculations of the evolution of magnetic flux with solar activity (Wang et al. 2005). Lean and Woods (2010) provide an overview of the NRL solar irradiance variability models, and Thuillier et al. (2013) compare NRLSSI with other spectral irradiance variability models. The NRLTSI2 and NRLSSI2 models are the next versions of these models, and incorporate new understanding of solar irradiance afforded by the SORCE/TIM and SORCE/SIM observations.

11. List the available documentation items that support reading, using and completely understanding the dataset. This list can include data format specifications, user guides, algorithm documentation, software documentation, design documents, papers, platform/instrument metadata, data/process flow diagrams and other informational sources. Provide URLs or identifiers as available.
1. Solar Irradiance CDR Implementation Plan
2. Solar Irradiance Data Flow Diagram
3. Maturity Matrix
4. Algorithm Theoretical Basis Document (In Draft)

12. Describe the type and scope of metadata (information about the data) included with the data files. Do the metadata comply with a standard?
netCDF-4 files will follow CF metadata conventions

13. Indicate the data file format(s).
1. netCDF-4

14. Are the data files compressed?
No

15. Detail how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

File naming convention: [data-short-name]_[data-version]_[extra-attribute(s)]_s[begin-date]_e[end-date]_c[file-date].[extension]

The TSI data will be 1 month files, with approximately 1-3 month updates. The SSI data will also be one month files, with the same update cadence.
There will also be historical datasets, using same naming convention, and organized in 1 month files.

The TSI daily and monthly values will be in 1-year and 10-year files, respectively, with approximately 1-3 month updates. The historical datasets (for dates prior to and including 2014), will use the same naming convention, and also be organized in 1-year and 10-year files,designated day and mon, respectively. In each case, V001 indicates Version 1 and Rev00 indicates revision 0.

Example: the initial netCDF files of daily TSI during 2000 and monthly TSI from 2000 to 2009 created on 31 Dec 2014 are designated:
TSI_V001_R00_day_20000101_20001231_31Dec14.nc
TSI_V001_R00_mon_200001_200912_31Dec14.nc

The TSI annual values will be in one file.
Example: the initial netCDF file of annual TSI created on 31 Dec 2014 is designated:
TSI_V001_R00_ann_1615_2014_31Dec14.nc

The SSI daily and monthly values will be in 1-year files, with approximately 1-3 month updates. The historical datasets (for dates prior to and including 2014), will use the same naming convention, and also be organized in 1-year files. Files of daily and monthly SSI will be designated day and mon, respectively. In each case, V001 indicates Version 1 and Rev00 indicates revision 0.

Example: the initial netCDF files of daily and monthly SSI during 2000 created on 31 Dec 2014 are designated:
SSI_V001_R00_day_20000101_20001231_31Dec14.nc
SSI_V001_R00_mon_20000101_20001231_31Dec14.nc
The SSI annual values will be in 10-year files, with approximately annual updates. The historical datasets (for dates prior to and including 2014), will use the same naming convention, and also be organized in 10-year files. Files of annual SSI will be designated ann.

Example: the initial netCDF file of annual SSI from 2000 to 2009 created on 31 Dec 2014 is designated:
SSI_V001_R00_ann_2000_2009_31Dec14.nc

16. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

The updates to regression coefficients, which are constant in a given model version, will be infrequent (every several years at most); the update will reflect new understanding of solar irradiance variability obtained from the observations as the record continues to length in time.

17. What is the total data volume to be submitted?

Enter the total data volume for static data to be submitted as a completed collection.

Total Data Volume: 8.6GB
Number of Data Files: 436

Enter the total data volume rate for a continuous data production.

Total Data Volume Rate: 5.25MB per Month
Data File Frequency: 2 per Month
Data Production Start: 2015-01-01

18. Explain how to access sample data files and/or file inventories for previewing.

Sample netCDF data needed for CF metadata review check

19. Identify the server that will connect to the ingest server at the data center for submitting the data.

Physical Location:	Laboratory for Atmospheric and Space Physics, Boulder, CO, 80303
System Name:	LISIRD/LaTiS
System Owner:	UCO/LASP > Laboratory for Atmospheric and Space Physics, University of Colorado
Additional Information:	We would initiate the sftp put from LASP.

20. What are the possible methods for submitting the data to the data center? Select all that apply.

1. FTP PUSH

HTTP service request

21. Identify the wanted type(s) of web access for distributing the data from the data center. Web access support depends on the resources available for the dataset.

1. Direct download links
2. Advanced web services (e.g., THREDDS Catalog Service)

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

No known constraints apply to the data.

23. Who are the expected users of the archived data?

Industry - renewable energy, water resources, hydrology, atmospheric chemistry and applications, photocell and solar energy applications
Science Communities - Global Climate Model (GCM) groups, atmospheric chemistry assimilation models (autochem), stratospheric and stratospheric/climate models, Community Radiative Transfer Models (MODTRAN)
Federal and State Government - NASA (Earth Science and Heliophysics Divisions), NOAA, NRL, NREL, NIST

24. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at the data center.

Dataset(s) are supported by the NOAA CDR Program

25. Are the data archived at another facility or are there plans to do so?

The data will be archived at the Laboratory for Atmospheric and Space Physics (LASP) and distributed through the LASP Interactive Solar Irradiance Data Center (LISIRD).

26. Is there an existing agreement or requirement driving this request to archive?

Per CDR Program. See Lean and Pilewskie SOWs.

27. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.
N/A

28. Have funds been allocated to archive the data at the data center?
NOAA CDR Program funds.

29. Time range when the data center would need to be ready to start archiving the data:
Between 2014-09-01 and 2014-09-26

30. Add any other pertinent information for this request.
The data sets cover these listed time periods:

1615 to present, annually
1882-01 to present, monthly
1882-01-01 to present, daily

Files of solar spectra to the end of 2014 will be provided as part of the initial transition. Additional files will be generated as continuous updates to the data record become available.

The file sizes are currently estimates based on the following formula:
A 1-year file containing daily values of SSI is 63 MB. A 1 year file containing monthly-averaged values of SSI is 2.5 MB. A 10 year file containing annually-averaged values of SSI is 2.2 MB. Yearly files of monthly-averaged SSI from 1978 to 2014 = 93 MB. 10-year files of annually-averaged SSI over the time period = 22 MB.
A 1-year file containing daily values of TSI is 44 KB. A 10-year file containing monthly-averaged values of TSI is 37 KB. Files of monthly-averaged TSI from 1882 to 2014 = 492 KB. There will also be yearly-averaged TSI from 1882-2014 requiring approximately 20 KB.
I approximated the data volume using the approximated data file sizes and a record length of 133 years, which covers the "historical dataset" from 1882 to 1978 and the dataset from 1978 to 2014.
There will be 1-3 month updates of this data beginning roughly 1 Jan 2015 (every 3 months at first, improving to monthly updates as processing matures).