**Xman and xcheck Operations Guide (PDS3 Version)**

*Produced by the NSSDCA*

*modified*

July 16, 2021

**OVERVIEW**

The xman program generates XML-formatted manifests of PDS3 Submission Information Packages (SIPs) to be delivered to NSSDCA. Xman generated manifests convey:

* Global SIP identifying and descriptive information
* Information about Transfer Objects within the SIP. (PDS3 SIPs will have only one Transfer Object containing the contents of a single PDS volume)
* Information about individual files with the Transfer Object, including checksums

See Appendix A for the XML schema of the manifest.

The xcheck program validates and extracts information from an xman-generated manifest file.

***XMAN* Installation**

The xman package is distributed in two forms: a setup.exe file for use on Microsoft platforms, and a source distribution kit for use on all other platforms.

On Microsoft platforms, the *xman* program is distributed in a standard *setup.exe* file which when run will offer the user a choice of the directory into which the *xman* program should be installed. After installation, the *xman* manifest generator program can be run from the XP command line shell, *cmd.exe*.

On all other platforms, the source code is extracted from the xman tar.gz file and built using the usual "./configure; make; make check" commands to collect system information, built the xman and xcheck programs, and run the regression tests. Building the programs will require a C compiler, its support libraries, and the Expat XML parsing library. On Apple OS-X platforms, these are supplied by the Xcode tools from the operating system installation disk; On Red Hat, Fedora, Centos, and other Red Hat derived operating systems, these can be installed using "yum groupinstall 'Development Tools'; yum install expat-devel"; On Solaris, these tools are available from the Sun Freeware collection at [www.sunfreeware.com](http://www.sunfreeware.com).

The xman software may be obtained from the following URL: <https://nssdc.gsfc.nasa.gov/pds/software/>

The manifest generation program, *xman.exe*, is a command-line driven program. To simplify the installation and operation of the program, all files in the *xman* distribution should be unpacked into a single directory, the program should be run from this directory, and the program should have permission to write the manifest and log file that will be generated into this directory. This imposes a requirement that the current working directory have enough space to hold the entire manifest, which will accrue at a rate of about a kilobyte per file being recorded.

***XMAN* CONFIGURATION**

Two configuration files are required for the operation of the *xman* program: *producer-id*.tsv, which identifies the producer running the xman program, and id-map.tsv, which maps producer and volume identifiers to a producer archive product identifier. These configuration files should be placed in the same directory from which the xman program will be run.

The *producer-id*.tsv files are node-specific and need only be installed once.

The id-map.tsv file is shared by the data nodes. This file is updated every time a new PDS Producer Archive Project is registered at NSSDCA. A data node must re-install the id-map.tsv when notified by NSSDCA that a Producer Archive Project submitted by that node has been registered. Additional information concerning registration of Producer Archive Project is in the document <https://nssdc.gsfc.nasa.gov/pds/PDS3_Submission_Process_v2.1.docx>.

The *producer-id*.tsv configuration files and the up-to-date id-map.tsv file available are accessible on-line at <https://nssdc.gsfc.nasa.gov/pds/configuration_files>.

***XMAN* OPERATION**

After installation, the *xman* manifest generator program can be run from the XP command line shell, *cmd.exe,* on Microsoft platforms, or from the bash, tcsh, or any other standard command shell on any other platform. From the shell prompt, change the current working directory to the directory containing the *xman* program and its support files. The *xman* program is then run from this directory and supplied with a single argument: the top level directory of the volume to be recorded in the manifest. Once started, the *xman* program will run to completion with no further interaction. While the run time depends on a wide range of factors, several hours of processing time should be allocated for each terabyte of data to be processed. See the *xman* and *xcheck* MAN pages for additional information.

**INPUTS**

The main input to the *xman* program is the path to the top-level directory of the volume to be recorded in the generated manifest file. The only other input normally supplied to the *xman* program is the *--comment=<message>* command line option. This option provides a way to record a small item of manually generated information about a data volume in the manifest. The message string supplied to this option must be a short string of printable UTF-8 characters.

**The Data Volume**: Because the manifest schema specifies a UTF-8 character set encoding, all directory names and file names must be valid UTF-8 character sequences. Due to limitations on the size of the magnetic tapes currently used for archival storage of the data, no data file may exceed 300 GB in size.

**Volume Description:** The top level of the volume for which a manifest is to be generated must contain a properly formatted ODL *VOLDESC.CAT* file. This file must contain an "OBJECT = VOLUME" block containing VOLUME\_ID parameter/value pair. Omission of this parameter will result in a fatal error condition.

**OUTPUTS**

If no errors are encountered during the execution of the *xman* program, the only output displayed will be a summary listing the number of files processed, their volume, and the time required. A typical program run might look like:

C:\Documents and Settings\jmkodis\xman-xp>xman.exe g:

xman: 462 files, 654153487 bytes in 417.671 seconds at 1.566 MB/sec

Along with this will be the primary output of the *xman* program, an XML-formatted manifest of the data volume and a log file documenting the execution of the *xman* program. Any warning or error messages will be recorded in this file, along with the version information for the *xman* program, the start and stop times of the manifest generation run, the total file count, byte count, run time, and data rate achieved during the program's execution. Unless specified otherwise, these files will be names Sip-manifest-<volume-id>.xml and Sip-manifest-<volume-id>.log.

The XML SIP manifest file will be as described in version 2.7 of the SIP manifest XML schema, included as Appendix A. Portions of a sample SIP manifest containing optional MD5 checksums is included as Appendix B. The *xman* program only records information on directories and regular data files. Any other file system objects encountered are silently ignored.

**Appendix A: SIP Manifest Schema**

**<?xml version="1.0" encoding="UTF-8"?>**

**<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">**

 **<xs:complexType name="DIRECTORY">**

 **<xs:sequence>**

 **<xs:element name="DIRECTORY\_NAME" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Full directory specification</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="MODIFICATION\_DATE\_TIME" type="xs:dateTime" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Directory modification date in ISO 8601 combined date and time format</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **</xs:sequence>**

 **</xs:complexType>**

 **<xs:complexType name="FILE">**

 **<xs:sequence>**

 **<xs:element name="FILE\_NAME" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Full path file specification</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="CHECKSUM">**

 **<xs:complexType>**

 **<xs:sequence>**

 **<xs:element name="METHOD" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Method used for generating CRC for object at this level, enumerated: CRC32, MD5,CRC value for object, no available it checksum method is "none"</xs:documentation>**

 **</xs:annotation>**

 **<xs:simpleType>**

 **<xs:restriction base="xs:string">**

 **<xs:whiteSpace value="preserve"/>**

 **<xs:enumeration value="CRC32"/>**

 **<xs:enumeration value="MD5"/>**

 **<xs:enumeration value="none"/>**

 **</xs:restriction>**

 **</xs:simpleType>**

 **</xs:element>**

 **<xs:element name="VALUE" type ="xs:string" minOccurs="1" maxOccurs="1" nillable="false"/>**

 **</xs:sequence>**

 **</xs:complexType>**

 **</xs:element>**

 **<xs:element name="SIZE">**

 **<xs:complexType>**

 **<xs:sequence>**

 **<xs:element name="UNIT" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Unit used for object size, enumerated: BYTE, KB, MB, GB, TB</xs:documentation>**

 **</xs:annotation>**

 **<xs:simpleType>**

 **<xs:restriction base="xs:string">**

 **<xs:whiteSpace value="preserve"/>**

 **<xs:enumeration value="BYTE"/>**

 **<xs:enumeration value="KB"/>**

 **<xs:enumeration value="MB"/>**

 **<xs:enumeration value="GB"/>**

 **<xs:enumeration value="TB"/>**

 **</xs:restriction>**

 **</xs:simpleType>**

 **</xs:element>**

 **<xs:element name="VALUE" type="xs:integer" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Object size in specified units</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **</xs:sequence>**

 **</xs:complexType>**

 **</xs:element>**

 **<xs:element name="MODIFICATION\_DATE\_TIME" type="xs:dateTime">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>File modification date and time in ISO 8601 combined date and time format</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **</xs:sequence>**

 **</xs:complexType>**

 **<xs:complexType name="SIP\_GLOBAL">**

 **<xs:sequence>**

 **<xs:element name="MANIFEST\_TYPE" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Identified NSSDC manifest type</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="PRODUCER\_ARCHIVE\_PROJECT\_ID" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Identifies a set of data from a single provider to be archived at NSSDC. Starting at 2009-01-01 the NSSDC collection ID is used as the PAPID.</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="PRODUCER\_SITE\_ID" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Identifies a provider site where the Submission Information Package was produced</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="SIP\_ID" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Unique identifier for the Submission Information Package (SIP)</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="PRODUCER\_COMMENT" type="xs:string" minOccurs="0" maxOccurs="1" nillable="true">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Optional producer-provided comment</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="CREATION\_DATE\_TIME" type="xs:dateTime" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Manifest generation date and time in ISO 8601 combined date and time format</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="ORIGINATING\_DATA\_DIRECTORY" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Absolute path spec or original data directory for sequence transfer objects</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **</xs:sequence>**

 **</xs:complexType>**

 **<xs:group name="DIRECTORY\_FILE">**

 **<xs:sequence>**

 **<xs:element name="DIRECTORY" minOccurs="0" maxOccurs="unbounded" nillable="true"/>**

 **<xs:element name="FILE" minOccurs="1" maxOccurs="unbounded" nillable="false"/>**

 **</xs:sequence>**

 **</xs:group>**

 **<xs:complexType name="TRANSFER\_OBJECT">**

 **<xs:sequence>**

 **<xs:element name="TRANSFER\_OBJECT\_ID" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Unique identifier consisting of SIPID + provider's name of the transfer object. Providers provide transfer objects in varying forms (e.g. volumes from PDS and data files from SPDF).</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="NUMBER\_OF\_FILES\_INCLUDED" type="xs:positiveInteger" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:annotation>**

 **<xs:appinfo/>**

 **<xs:documentation>Number of files in contained in this object.</xs:documentation>**

 **</xs:annotation>**

 **</xs:element>**

 **<xs:element name="TRANSFER\_OBJECT\_SIZE">**

 **<xs:complexType>**

 **<xs:sequence>**

 **<xs:element name="UNIT" minOccurs="1" maxOccurs="1" nillable="false">**

 **<xs:simpleType>**

 **<xs:restriction base="xs:string">**

 **<xs:whiteSpace value="preserve"/>**

 **<xs:enumeration value="BYTE"/>**

 **<xs:enumeration value="KB"/>**

 **<xs:enumeration value="MB"/>**

 **<xs:enumeration value="GB"/>**

 **<xs:enumeration value="TB"/>**

 **</xs:restriction>**

 **</xs:simpleType>**

 **</xs:element>**

 **<xs:element name="VALUE" type="xs:float" minOccurs="1" maxOccurs="1" nillable="false"/>**

 **</xs:sequence>**

 **</xs:complexType>**

 **</xs:element>**

 **<xs:group ref="DIRECTORY\_FILE" minOccurs="1" maxOccurs="unbounded"/>**

 **</xs:sequence>**

 **</xs:complexType>**

 **<xs:element name="SIP\_MANIFEST">**

 **<xs:complexType>**

 **<xs:sequence>**

 **<xs:element name="SIP\_GLOBAL" type="SIP\_GLOBAL" minOccurs="1" maxOccurs="1" nillable="false"/>**

 **<xs:element name="TRANSFER\_OBJECT" type="TRANSFER\_OBJECT" minOccurs="1" maxOccurs="1" nillable="false"/>**

 **</xs:sequence>**

 **</xs:complexType>**

**</xs:element>**

**</xs:schema>**

**Appendix B: Sample SIP Manifest**

**<?xml version='1.0' encoding='UTF-8'?>**

**<SIP\_MANIFEST>**

 **<SIP\_GLOBAL>**

 **<MANIFEST\_TYPE>pds</MANIFEST\_TYPE>**

 **<PRODUCER\_ARCHIVE\_PROJECT\_ID>PDSTEST:000001</PRODUCER\_ARCHIVE\_PROJECT\_ID>**

 **<PRODUCER\_SITE\_ID>PDSTEST</PRODUCER\_SITE\_ID>**

 **<SIP\_ID>PDSTEST:000001:1236107720:MGSC\_1046</SIP\_ID>**

 **<PRODUCER\_COMMENT></PRODUCER\_COMMENT>**

 **<CREATION\_DATE\_TIME>2009-03-03T19:15:20Z</CREATION\_DATE\_TIME>**

 **<ORIGINATING\_DATA\_DIRECTORY>/cygdrive/e</ORIGINATING\_DATA\_DIRECTORY>**

 **</SIP\_GLOBAL>**

 **<TRANSFER\_OBJECT>**

 **<TRANSFER\_OBJECT\_ID>PDSTEST:000001:1236107720:MGSC\_1046:1</TRANSFER\_OBJECT\_ID>**

 **<NUMBER\_OF\_FILES\_INCLUDED> 462</NUMBER\_OF\_FILES\_INCLUDED>**

 **<TRANSFER\_OBJECT\_SIZE>**

 **<UNIT>BYTE</UNIT>**

 **<VALUE> 654153487</VALUE>**

 **</TRANSFER\_OBJECT\_SIZE>**

 **<DIRECTORY>**

 **<DIRECTORY\_NAME>./</DIRECTORY\_NAME>**

 **<MODIFICATION\_DATE\_TIME>2000-06-01T00:52:34Z</MODIFICATION\_DATE\_TIME>**

 **</DIRECTORY>**

 **<DIRECTORY>**

 **<DIRECTORY\_NAME>./CATALOG/</DIRECTORY\_NAME>**

 **<MODIFICATION\_DATE\_TIME>2000-06-01T00:52:34Z</MODIFICATION\_DATE\_TIME>**

 **</DIRECTORY>**

 **<FILE>**

 **<FILE\_NAME>./CATALOG/CATINFO.TXT</FILE\_NAME>**

 **<CHECKSUM>**

 **<METHOD>CRC32</METHOD>**

 **<VALUE>fcf8c381</VALUE>**

 **</CHECKSUM>**

 **<CHECKSUM>**

 **<METHOD>MD5</METHOD>**

 **<VALUE>cf813f9201c8b591c0c5ee45aa5b3076</VALUE>**

 **</CHECKSUM>**

 **<SIZE>**

 **<UNIT>BYTE</UNIT>**

 **<VALUE>1309</VALUE>**

 **</SIZE>**

 **<MODIFICATION\_DATE\_TIME>2000-06-01T00:52:33Z</MODIFICATION\_DATE\_TIME>**

 **</FILE>**

 **...**

 **<FILE>**

 **<FILE\_NAME>./VOLDESC.CAT</FILE\_NAME>**

 **<CHECKSUM>**

 **<METHOD>CRC32</METHOD>**

 **<VALUE>89476380</VALUE>**

 **</CHECKSUM>**

 **<CHECKSUM>**

 **<METHOD>MD5</METHOD>**

 **<VALUE>14dc87facea5d3b0f62c4476c7785d60</VALUE>**

 **</CHECKSUM>**

 **<SIZE>**

 **<UNIT>BYTE</UNIT>**

 **<VALUE>1662</VALUE>**

 **</SIZE>**

 **<MODIFICATION\_DATE\_TIME>2000-06-01T00:52:34Z</MODIFICATION\_DATE\_TIME>**

 **</FILE>**

 **</TRANSFER\_OBJECT>**

**</SIP\_MANIFEST>**