OSKAR Binary File Format

Version history:

Revision	Date	Modification
1	2012-11-23	Creation.
2	2013-03-04	Fixed description of image data tag.
3	2013-04-18	Added telescope model path, channel bandwidth and time integration tags to visibility data group.
4	2013-11-29	Added image group tags for grid type and coordinate frame. Added sky group tag for rotation measure.
5	2014-07-16	[2.5.0] Added spline data and element data tag groups.

1 Introduction

This document describes the binary file format used by OSKAR applications. It is intended to be used for reference only, since there are library functions to read and write data files in this format.

2 Format Description

An OSKAR binary file contains a simple file header, and a sequence of data blocks, each with its own descriptive header tag. Every data block is independent, and each is stored as an array of values in the native byte-order of the system that wrote the file. (The byte order is recorded in the header.)

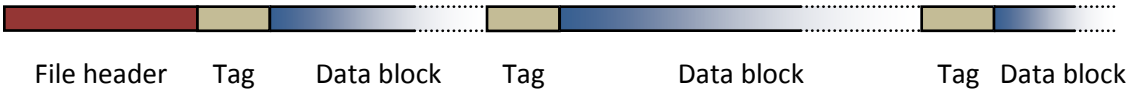

Figure 1: Overall structure of data in an OSKAR binary file.

Since the type and size of each data block is stored in the tag, any program reading such a file can simply skip over any data blocks which are not of interest. A single data block may consist of a single value, or an array of values; in the latter case, other data blocks within the file should contain the dimension sizes of multi-dimensional arrays.

When reading an OSKAR binary file, it is sensible to first construct a local tag index to help locate the correct data. All OSKAR library functions that read binary files will do this automatically if necessary.

2.1 File Header

The file header is exactly 64 bytes long, and is structured as follows:

Offset (bytes)	Length (bytes)	Description
0	9	The ASCII string "OSKARBIN", with trailing zero.
9	1	The OSKAR binary format version.
10	1	If data blocks are written as little endian, 0; else 1.
11	1	Size of void* in bytes.
12	1	Size of int in bytes.
13	1	Size of long int in bytes.
14	1	Size of float in bytes.
15	1	Size of double in bytes.
16	4	The OSKAR_VERSION as a little-endian, 4-byte integer.
20	44	(*Reserved; must be 0*).

Table 1: Description of the file header structure.

The OSKAR binary format version (at byte offset 9) is currently 1. This format is anticipated to remain stable, and will only change if the underlying header or tag structure is modified.

The OSKAR_VERSION (at byte offset 16) is given as a little-endian, 4-byte integer in hexadecimal format. It is a numeric value of the form 0xMMNNPP (MM = major, NN = minor, PP = patch) that specifies the version number of the OSKAR library that wrote the file. For example, in OSKAR version 2.1.3 this would be 0x020103.

2.2 Tag (Data Block Header)

Each data block that follows the file header (whether a single value or an array of values) is preceded by a tag, which is a structure exactly 20 bytes long. The data that follows the tag is uniquely identified by the combination of the group ID, tag ID and index. The two identifiers will take different forms depending on whether the tag is “standard” or “extended,” and this is specified by the tag flags. A user-specified index can be used to identify multiple versions of the same tag type within a file, and should be set to 0 if this is not required. The differences between the two tag types are detailed in Section 2.2.2 and Section 2.2.3.

Offset (bytes)	Length (bytes)	Description
0	4	The string "TAG" in ASCII format, with trailing zero.
4	1	Tag flags (see Section 2.2.1).
5	1	Data type of the data block (see Section 2.2.4).
6	1	The group ID, if not an extended tag; else the group name size in bytes.
7	1	The tag ID, if not an extended tag; else the tag name size in bytes.
8	4	User-specified index, as little-endian 4-byte integer.
12	8	Block size in bytes, as little-endian 8-byte integer. This is the total number of bytes until the next tag.

Table 2: Description of the tag structure.

2.2.1 Tag Flags

Bit	Description
0-6	(*Reserved; must be 0.*)
7	If true, this is an extended tag; if false, this is a standard tag.

Table 3: Description of the tag flags.

2.2.2 Standard Tags

A standard tag has both the group ID and tag ID specified as 8-bit integer values, which are found at byte offsets 6 and 7 in the tag header.

There can be a maximum of 256 different group types, and a maximum of 256 tags per group, so the total number of unique tag types supported for standard tags is 65,536. All data files written by current versions of OSKAR applications only use standard tag identifiers.

2.2.3 Extended Tags

If the tag is an extended tag, then the group name and tag name are specified as strings rather than 8-bit IDs: extended tags in an OSKAR binary file have the group name and then the tag name written as ASCII 8-bit character strings immediately after the main tag structure itself. Both strings must be less than 255 characters long, and both will have a null terminator. The length of the group ID string and tag ID string, including the null terminators, will be available at (respectively) byte offsets 6 and 7 in the tag header.

Note that the block size in the tag header is the total number of bytes until the next tag, including the extended tag names.

2.2.4 Data Types

The data type field in the tag header consists of a single byte, whose eight bits have the following meaning:

Bit	Meaning when set
0	Char type (1 byte), used also for string data.
1	Integer type (4 bytes).
2	Single-precision floating point type (4 bytes).
3	Double-precision floating point type (8 bytes).
4	(*Reserved; must be 0.*)
5	Complex flag: data consists of a pair of values that describe real and imaginary components. The real part is given first, then the imaginary part.
6	Matrix flag: data consists of four values that describe a 2x2 matrix. For a matrix written as $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the order of the values is *a, b, c, d*.
7	(*Reserved; must be 0.*)

Table 4: Data type flags.

Note that bits 5-6 are used to specify compound types with bits 2-3: so, for example, a double-precision complex matrix type would be represented as (binary) 01101000 = (decimal) 104.

3 Standard Tag Groups

This section lists the tag identifiers found in various OSKAR binary format files.

3.1 Standard Meta-Data Tags

Tags in this group have a group ID of 1.

Tag ID	Description
1	Date and time of file creation [string].
2	Version of OSKAR that created the file [string].
3	Username of user that created the file [string].
4	Current working directory for application that created the file [string].

3.2 Settings Tags

Tags in this group have a group ID of 3.

Tag ID	Description
1	Path to settings file [string].
2	Settings file contents [string].

3.3 Run Information Tags

Tags in this group have a group ID of 4.

Tag ID	Description
1	Run log [string].

3.4 Visibility Data Tags

Tags in this group have a group ID of 5.

Tag ID	Description
1	Number of channels [int].
2	Number of times [int].
3	Number of baselines [int].
4	Dimension order of visibility data [array, int(4)]. (See note 1, below.)
5	Data type of coordinate arrays [int]. (See Section 2.2.4.)
6	Data type of visibility data array [int]. (See Section 2.2.4.)
7	Start frequency, in Hz [double].
8	Frequency increment, in Hz [double].
9	Observation start time, as MJD(UTC) [double].
10	Time increment, in seconds [double].
11	Polarisation type [int]. (0 = none, 1 = linear)
12	Unit of baseline coordinates [int]. (1 = metres)
13	Baseline UU-coordinates [array; type given by tag ID 5].
14	Baseline VV-coordinates [array; type given by tag ID 5].
15	Baseline WW-coordinates [array; type given by tag ID 5].
16	Complex visibility amplitudes [array; type given by tag ID 6].
17	Phase centre Right Ascension [double].
18	Phase centre Declination [double].
19	Number of stations [int].
20	Unit of station coordinates [int]. (1 = metres)
21	Station X-coordinates in ECEF frame [array; type given by tag ID 5].
22	Station Y-coordinates in ECEF frame [array; type given by tag ID 5].
23	Station Z-coordinates in ECEF frame [array; type given by tag ID 5].
24	Channel bandwidth, in Hz [double].
25	Time integration per correlator dump, in seconds [double].
26	Path to telescope model directory [string].

Note 1: The visibility dimension order is given by tag ID 4, which is a four-element array containing the following integer IDs:

0. Channel
1. Time
2. Baseline
3. Polarisation

The order of the IDs gives the dimension order of the visibility data array, from slowest- to fastest-varying. Current versions of OSKAR write visibility data with polarisation varying fastest, then baseline, then time, then frequency channel.

Note 2: Baselines are formed by cross-correlating stations 0-1, 0-2, 0-3… 1-2, 1-3… etc. The autocorrelations are not computed, so for n stations, there will be $n(n-1)/2$ baselines.

3.5 Image Data Tags

Tags in this group have a group ID of 6.

Tag ID	Description
0	Image data [array; type given by tag ID 2].
1	Image type [int]. (See note 1, below.)
2	Data type of image data array [int]. (See Section 2.2.4.)
3	Dimension order of image data [array, int(5)]. (See note 2, below.)
4	Number of pixels, width [int].
5	Number of pixels, height [int].
6	Number of polarisation planes [int].
7	Number of time slices [int].
8	Number of frequency channel planes [int].
9	Image centre Right Ascension (degrees) [double].
10	Image centre Declination (degrees) [double].
11	Field of view in Right Ascension (degrees) [double].
12	Field of view in Declination (degrees) [double].
13	Observation start time, as MJD(UTC) [double].
14	Time increment, in seconds [double].
15	Start frequency, in Hz [double].
16	Frequency increment, in Hz [double].
17	Mean image value [double].
18	Image variance [double].
19	Minimum image value [double].
20	Maximum image value [double].
21	Image RMS [double].
3.5.1.1.1.1 **22**	Grid type [int].
3.5.1.1.1.2 **23**	Coordinate frame [int].
24	HEALPix nside [int].

Note 1: The image type is given as an integer code from the following list:

0. Undefined.
1. Full Stokes (in order: I, Q, U, V).
2. Stokes I.
3. Stokes Q.
4. Stokes U.
5. Stokes V.
6. All linear polarisations (in order: XX, XY, YX, YY).
7. Linear XX.
8. Linear YY.
9. Linear XY.
10. Linear YX.
50. Point-spread function (PSF).
100. Beam pattern, scalar.
101. Beam pattern, polarised.

Note 2: The image dimension order is given by tag ID 3, which is a five-element array containing the following integer IDs:

0. RA

1. Dec
2. Polarisation
3. Time
4. Channel

The order of the IDs gives the dimension order of the image data array, from slowest- to fastest-varying. Current versions of OSKAR write image data with RA varying fastest, then Dec, then polarisation, then time, then frequency channel.

Note 3: The image grid type (tag ID 22) is given as an integer code from the following list:

0. Undefined.
1. Rectilinear.
2. HEALPix Ring scheme.

Rectilinear grid type images are made using an orthographic (-SIN, in FITS nomenclature) tangent-plane projection.

Note 4: The image coordinate frame (tag ID 23) is given as an integer code from the following list:

0. Undefined.
1. Equatorial.
2. Horizontal.

3.6 Sky Model Data Tags

Tags in this group have a group ID of 7.

Tag ID	Description
1	Number of sources [int].
2	Data type of all arrays [int]. (See Section 2.2.4.)
3	Right Ascension values, in radians [array; type given by tag ID 2].
4	Declination values, in radians [array; type given by tag ID 2].
5	Stokes I values, in Jy [array; type given by tag ID 2].
6	Stokes Q values, in Jy [array; type given by tag ID 2].
7	Stokes U values, in Jy [array; type given by tag ID 2].
8	Stokes V values, in Jy [array; type given by tag ID 2].
9	Reference frequency values, in Hz [array; type given by tag ID 2].
10	Spectral index values [array; type given by tag ID 2].
11	FWHM (major axis), in radians [array; type given by tag ID 2].
12	FWHM (minor axis), in radians [array; type given by tag ID 2].
13	Position angle of major axis, in radians [array; type given by tag ID 2].
3.6.1.1.1.1 **14**	Rotation measure, in radians / m^2 [array; type given by tag ID 2].

3.7 Spline Data Tags

Tags in this group have a group ID of 9. Arrays will be present in both single and double precision.

Tag ID	Description

3.7.1.1.1.1 **1**	Number of knots in X or theta coordinate [int].
3.7.1.1.1.2 **2**	Number of knots in Y or phi coordinate [int].
3.7.1.1.1.3 **3**	Knot positions in X or theta [real array].
3.7.1.1.1.4 **4**	Knot positions in Y or phi [real array].
3.7.1.1.1.5 **5**	Spline coefficients [real array].
3.7.1.1.1.6 **6**	Smoothing factor [double].

3.8 Element Data Tags

Tags in this group have a group ID of 10.

Tag ID	Description
1	Surface type [int]. 1 = Ludwig-3

Element data files will contain a number of spline data tag groups, which are identified by an index.

For fitted coefficients in the Ludwig-3 system, the spline tags will have the following index values:

- H (real): 0
- H (imag): 1
- V (real): 2
- V (imag): 3