Data eXchange Unit Description

IWG4 WP1 (OSF) DXU Description

Document No.: VIS-TNO-4MOST-47110-9234-0001

Issue No.: 1.01

DocuShare URL: https://ds-web.aip.de/docushare/dsweb/Services/Document-5700

Prepared by

_____
Elmo Tempel

Approved by

_____
ODG System Engineer

Released by

_____
Operations Manager

Change Record

Issue	Date	Page, Section, Paragraph Affected	Reason, Remarks
0.01	2017-01-02	All	New document
0.02	2017-02-20	Front page	Updated version number, dates, etc.
0.03	2017-10-24	All	Updates from deprecated DMS - L2 ICD
0.04	2017-12-04	All	Updates from ODG comments
0.10	2017-12-19	All	Content added by IWG4
1.00	2018-03-13	Front page and header	Document released for FDR

List of Contributors

Name	First Name	Institute	Comments
Worley	Clare	IoA	Template document
Walcher	Jakob	AIP	Template document
Liske	Jochen	UHH	Template document
Krumpe	Mirko	AIP	Review of the IWG4 DXU structure
Tempel	Elmo	AIP	Principal writer
Louise	Howes	Lund	Minor comments
Sarah	Martell	AAO	Minor comments

Table of contents

List of Figures

No table of figures entries found.

List of Tables

1 Objectives

This document defines the Data eXchange Unit (DXU) for the IWG4 WP1 (object selection function) data products to be used by 4OR. IWG4 is responsible for calculating the geometric and object selection functions for the 4MOST surveys. This DXU describes the object selection functions that are calculated based on synthetic model spectra. The geometric and observed selection functions are described in [RD8].

Any changes to the DXU must be agreed upon by the IWG4 and DMS including the 4OR in advance of data product submission to allow all parties to update their software to accommodate any agreed changes.

2 Applicable Documents (AD)

The following applicable documents (AD) of the exact issue shown form a part of this document to the extent described herein. In the event of conflict between the documents referenced herein and the contents of this document, the contents of this document are the superseding requirement.

AD ID	Document Title	Document Number	Issue	Date
[AD1]	4MOST ICD – General Definitions	MST-ICD-PMO-02000-0002	1.a	2014-07-21
[AD2]	4MOST Acronym List	MST-LIS-PMO-30500-9350-0001	2.00	2015-03-06
[AD3]	VLT Software Programming Standards	VLT-PRO-ESO-10000-0228	2.00	11.02.2010
[AD4]	Software Engineering Development Standard	MST-STD-PMO-40200-9420-0001	1.00	06.04.2016

AD ID	Document Title	Document Number	Issue	Date
[AD5]	FITS Working Group, Commission 5: Documentation and Astronomical Data, International Astronomical Union. Definition of the Flexible Image Transport System (FITS),	http://fits.gsf .nasa.gov/fits_standard.html	3.0	07-2008
[AD6]	ESO Science Data Products Standard	GEN-SPE-ESO-33000-5335	5.0	11/01/2013

3   Reference Documents (RD)

The following reference documents (RD) contain useful information relevant to the subject of the present document.

RD ID	Document Title	Document Number	Issue	Date
[RD1]	Data Interface Control Document (ESO FITS Header standards)	GEN-SPE-ESO-19400-0794	5.0	8/07/2011
[RD2]	DMS DRPD	VIS-DER-4MOST-47110-1410-0002	1.0	2017-02-27
[RD3]	DMS DR Archive	VIS-DER-4MOST-47110-1440-0001	1.1	2016-11-15
[RD4]	Back-End ICD	VIS-ICD-4MOST-47110-9700-0002	1.0	2017-02-27
[RD5]	The UCD1+ controlled vocabulary Version 1.3	http://wiki.ivoa.net/internal/IVOA/PlanetaryUCD/WD-UCDlist-1.3-20170502.pdf	1.3	2017-05-02
[RD6]	IWG4 WP1 (OSF) description	MST-TNO-PSC-20304-9234-0002	1.0	01/03/2017
[RD7]	IWG4 WP2 (GSF) description	MST-TNO-PSC-20304-9234-0003	1.0	01/03/2017
[RD8]	IWG4 WP2 (GSF) DXU description		0.05	05/12/2017

4   Definitions

The L1 products are the extracted and reduced science-ready spectra upon which L2 (and L3) analyses will take place, as described in [RD2]. The L2 products are the deliverables from the L2 pipelines that are measured on the L1 products. The L1 and L2 products comprise the Phase 3 requirements for ESO.

The L3 products are added value products that are supplementary to the L2 products but are not included within the ESO Phase 3 product delivery.

4OR is the operational repository for 4MOST that will hold all 4MOST products ingested on a continuous basis as described in [RD3].

4PA is the public archive for 4MOST that will hold L1 and L2 products for release to the 4MOST community and the world on a regular basis as described in [RD3].

The data flow of the entire back end operations is described in [RD4].

The description of the selection functions (SFs) is described in [RD6] (object selection functions) and [RD7] (geometric selection functions). Object selection function (OSF) in IWG4 context means that these are calculated based on model spectra. The calculation of OSF does not require real data. Data products for the selection functions calculated using the real observed data is described in [RD8].

5 General Format

FITS format is the designated format for data transport for 4MOST. Standard notation and naming conventions for the files and content are provided at the start of the pertinent sections below.
There are four key standards that all DXUs must adhere to:

- The use of FITS must adhere to the ESO FITS standards outlined in [RD1],
- The data and metadata described here must meet the requirements for ingestion into the 4MOST archives as described in [RD3] section 4.3.2.2,
- The data and metadata must be VO compliant. Unified Content Descriptors[1] must be specified per measurable as listed in [RD5].
- For the purposes of delivery as ESO Phase 3 these products must also adhere to the ESO Science Data Product Standards as described in [AD6].

5.1 Data Package Delivery to 4OR

The data products will be delivered to the 4OR via secure ftp upload as described in [RD3] section 4.3.2.2.

6 FITS File and Structure

The data package is the output of the analysis by IWG4 on the L1 and L2 data products. The provenance of the input products (input product filename and version number) is retained in the data products by IWG4 in the Primary Header.

6.1 FITS Filename

The filename of each delivered product should have the format defined below where the naming fields are separated by an underscore:

FM_<Field1>_<Field2>_<Field3>.fits

The fields are defined in **Table 6-1**.

---

[1] http://www.ivoa.net/documents/REC/UCD/UCD-20050812.html

Table 6-1 Filename Fields		
Field	**Description**	**Example**
FM	Code-usable abbreviation for 4MOST	
1	Data Release abbreviation and number to 2.d.p	DR1.00
2	Source of FITS file	IWG4
3	Survey and sub-survey abbreviation (object selection function class abbreviation)	S6.1

The length of the FITS filename should not be excessive (≤30) and 3 fields should be sufficient to distinguish files for a particular data release. Further file specific information should be placed in the primary header (see Section 6.4).

Selection functions are calculated per sub-survey. Since SFs data products are different for each survey, there are separate file for each sub-survey. The structure of FITS file for each object class is described in Section 6.2.

Remark: it shall be coordinated with surveys what object selection functions are required. It might be that object selection functions are not required for all surveys or sub-surveys. It is assumed that each sub-survey contains only one object class. The content of OSF is defined by surveys.

The list of all object classes is defined in Table 6-2.

Table 6-2 List of surveys and sub-surveys		
Abbreviation	**Survey name**	**Sub-survey name**
S1.1	MW Halo LR Survey	k/m giants
S1.2	MW Halo LR Survey	Rare objects
S2.1	MW Halo HR Survey	Bright
S2.2	MW Halo HR Survey	Faint
S3.1	MW Bulge and Disc LR Survey	Extended Solar neighborhood
S3.2	MW Bulge and Disc LR Survey	Dynamical disc
S3.3	MW Bulge and Disc LR Survey	Faint dynamical disc
S3.4	MW Bulge and Disc LR Survey	Chemodynamical disc
S3.5	MW Bulge and Disc LR Survey	Inner Galaxy
S3.6	MW Bulge and Disc LR Survey	White dwarfs
S3.7	MW Bulge and Disc LR Survey	Compact X-ray emitting binaries
S4.1	MW Bulge and Disc HR Survey	Bulge
S4.2	MW Bulge and Disc HR Survey	Inner disc
S4.3	MW Bulge and Disc HR Survey	Outer disc
S4.4	MW Bulge and Disc HR Survey	Nearby disc
S5.1	Galaxy Cluster Survey	Bright cluster galaxies
S5.2	Galaxy Cluster Survey	Cluster members
S5.3	Galaxy Cluster Survey	High SNR cluster members
S6.1	AGN Survey	Main AGN survey

S6.2	AGN Survey	High-resolution AGN survey
S7.1	Galaxy Evolution Survey (WAVES)	WAVES-Wide
S7.2	Galaxy Evolution Survey (WAVES)	WAVES-Deep
S8.1	Cosmology Redshift Survey	WiFloZ: Bright galaxies (BG) sub-survey
S8.2	Cosmology Redshift Survey	BLESS-Red: LRG sub-survey
S8.3	Cosmology Redshift Survey	BLESS-Blue: ELG sub-survey
S8.4	Cosmology Redshift Survey	QURVE: Quasar sub-survey
S8.5	Cosmology Redshift Survey	QURVE-LyA: Lyman alpha forest sub-survey
S9.1	Magellanic Clouds Survey	Main sequence stars
S9.2	Magellanic Clouds Survey	RR Lyrae stars
S9.3	Magellanic Clouds Survey	Cepheids low-resolution
S9.4	Magellanic Clouds Survey	Red clumps stars
S9.5	Magellanic Clouds Survey	RGB stars
S9.6	Magellanic Clouds Survey	Supergiant stars
S9.7	Magellanic Clouds Survey	Cepheids high-resolution
S9.8	Magellanic Clouds Survey	O-rich AGB stars
S9.9	Magellanic Clouds Survey	C-rich AGB stars
S9.10	Magellanic Clouds Survey	Tip of RGB stars
S9.11	Magellanic Clouds Survey	Others
S10.1	Time-Domain Extragalactic Survey	TiDES-SH: supernovae hosts
S10.2	Time-Domain Extragalactic Survey	TiDES-LT: live transients
S10.3	Time-Domain Extragalactic Survey	TiDES-RM: AGN reverberation mapping

6.2   FITS Structure

The structure of the IWG4 data product is defined in the following tables. Each object class (listed in Table 6-2) has different FITS structure. The structure for each object class id defined in a separate table. The types of data delivered are multi-dimensional maps.

Remark: Below are only some examples of the tables. The content of possible tables (listed in Table 6-2) should be coordinated with surveys. The current examples are just placeholders.

Remark: Name of the main extension (e.g. GAL_MAIN, STARS_MAIN) should be given in Table 6.2. Do we need different extension name for each sub-survey?

Table 6-3 FITS Structure: S7.1		
EXTN #	**Extension Name**	**Description**
0	PHU	This is the primary header unit. There will be no data in this HDU. The header will have all the general information about the IWG4 OSF data products.
1	GAL_MAIN	This extension gives the joint probability as a function of SNR and all main galaxy parameters: magnitude, size, color, redshift, emission line strength.

2, …	…	Additional extensions are defined if necessary.

Table 6-4 FITS Structure: S3.1

EXTN #	Extension Name	Description
0	PHU	This is the primary header unit. There will be no data in this HDU. The header will have all the general information about the IWG4 OSF data products.
1	STARS_MAIN	This extension gives the joint probability as a function of SNR and all main stellar parameters: magnitudes in different bands. Bands are still TBD.
2, …	…	Additional extensions are defined if necessary.

Table 6-5 FITS Structure: S6.1

EXTN #	Extension Name	Description
0	PHU	This is the primary header unit. There will be no data in this HDU. The header will have all the general information about the IWG4 OSF data products.
1	AGN_MAIN	This extension gives the joint probability as a function of SNR and all main AGN parameters: magnitude, redshift, type of AGN.
2	AGN_TYPE	This extension defines the types of AGNs.
3, …	…	Additional extensions are defined if necessary.

6.3 Specific Usage Conventions for 4MOST DXUs

The following conventions are used across the 4MOST DXUs for consistency and standardisation when being ingested by the archives.

6.3.1 NULL Values

For NULL values the standard usage per data type are listed in Table 6-6.

Table 6-6 NULL Values per Data Type

Column Type	Data Type	NULL Value
AlphaNumeric (e.g. Object name, Flags)	String	' '
Measurements, Errors	Floating Point	NaN, INDEF, NULL
Limits, Number Counts	Integer (positive)	-1

6.3.2 Concatenation symbol

When multiple strings are concatenated within a string cell (e.g. multiple flags, names etc) the delimiter for use in this DXU is the comma (,). Currently, we do not see any need for delimiter, however, it might change in the future when DXU document is updated.

6.3.3 Coordinate Units and Precision

The coordinate columns used in this DXU are in units of degree decimals and specified in DOUBLE precision. The used coordinate system for RA and Dec is J2000 equatorial. Coordinates are given in GAIA reference system.

6.3.4 Boolean values

Use of any Booleans in this DXU adhere to the correct usage within FITS file of 'T' and 'F' with data format of L.

6.4 Primary Header

The primary header contains further information regarding the source and processing of the data products that is not encoded in the filename. Primary header has the same structure for all object classes.
Keywords added to the standard FITS primary header are listed in **Table 6-7**.

Table 6-7 Primary Header Keywords		
Keyword	**Content**	**Description**
FMNODE	IWG4	4MOST node that has produced this data product
RELEASE	DRX.xx	Data release of this data product
PVERS	XX.xx	IWG4 WP1 Pipeline Version
DXUDOC	<SOURCE DXU Document name>	Data eXchange Unit document in which this product is described.
CONTACT	<name of person>	Who to contact if any anomalies are found in this dataproduct.

6.5 FITS Extensions

Each of the Extensions listed in **Table 6-3** are described further in the following subsections.

6.5.1 Extension GAL_MAIN: main OSF for galaxies

This extension contains multi-dimensional data array (probability map). The header of the extension contains all the necessary information about the data product.

Table 6-8 Extension GAL_MAIN Header Keywords		
Keyword	**Content**	**Description**
EXTNAME	GAL_MAIN	Name of extension

The number of axes and the axis scales are defined in the following table. These are additional keywords in the extension header. The scale of axes is linear, where physical unit is given as $physical\_value = AZEROn + (i-1)*ADELTAn$, where 'i' is the axis index starting from 1.

The size of the data cube is given with mandatory keywords NAXISn. Remark: propose a better way if possible.

Table 6-9 Extension GAL_MAIN Header Keywords for axes				
#	ATYPEn	AUNITn	AZEROn	ADELTAn
1	SNR	-	<starting value>	<step size>
2	mag_x	mag	<starting value>	<step size>
3	color	mag	<starting value>	<step size>
4	size	arcsec	<starting value>	<step size>
5	redshift	-	<starting value>	<step size>
6	em_line_index	?	<starting value>	<step size>

In the table above, 'n' goes from 1 to the number of axes in a multi-dimensional array.
Remark: the parameters given in the table shall be coordinated with surveys. The magnitude system (band, fibre vs total) and exact meaning of other parameters are specified by surveys. Currently, this table is just a placeholder.

6.5.2 Extension STARS_MAIN: main OSF for stars

This extension contains multi-dimensional data array (probability map). The header of the extension contains all the necessary information about the data product.

Table 6-10 Extension STARS_MAIN Header Keywords		
Keyword	Content	Description
EXTNAME	STARS_MAIN	Name of extension

The number of axes and the axis scales are defined in the following table. These are additional keywords in the extension header. The scale of axes is linear, where physical unit is given as physical_value = AZEROn + (i-1)*ADELTAn, where 'i' is the axis index starting from 1. Remark: propose a better way if possible.

Table 6-11 Extension STARS_MAIN Header Keywords for axes				
#	ATYPEn	AUNITn	AZEROn	ADELTAn
1	SNR	-	<starting value>	<step size>
2	mag_x	mag	<starting value>	<step size>

In the table above, 'n' goes from 1 to the number of axes in a multi-dimensional array.
Comment: These will need the equivalent TTYPE Comment which describes the keyword, the TUCD and the other relevant TXXX. The archives will use these as tool tips, in-line descriptions etc.
Remark: There is alternative way how to define n-dimensional datacubes: define it as a one-dimensional array and then provide algorithm how to convert it to n-dimensional datacube. OpSys uses this. We should think which way we want to present the n-dimensional datacubes here. This will be decided later. *Decision is needed how to store multi-dimensional datacubes!*

6.5.3 Extension AGN_MAIN: main OSF for AGNs

This extension contains multi-dimensional data array (probability map). The header of the extension contains all the necessary information about the data product.

Table 6-12 Extension AGN_MAIN Header Keywords

Keyword	Content	Description
EXTNAME	AGN_MAIN	Name of extension

The number of axes and the axis scales are defined in the following table. These are additional keywords in the extension header. The scale of axes is linear, where physical unit is given as physical_value = AZEROn + (i-1)*ADELTAn, where 'i' is the axis index starting from 1. Remark: propose a better way if possible.

Table 6-13 Extension AGN_MAIN Header Keywords for axes

#	ATYPEn	AUNITn	AZEROn	ADELTAn
1	SNR	-	<starting value>	<step size>
2	mag_x	mag	<starting value>	<step size>
3	redshift	-	<starting value>	<step size>
4	AGN_Type	-	1	1

In the table above, 'n' goes from 1 to the number of axes in a multi-dimensional array.

The types of the AGNs are defined in the following table.

Table 6-14 Extension AGN_Type

Values	AGN Type description
1	??
2	??
3	??

7 Further Information on Provenance

7.1 Input data used

Describe what input files were used (e.g. 4MOST data products, external physical information, models etc) including version information.
Model spectra produced by surveys. 4FS_ETC to generate realistic spectra from model spectra. IWG7 and IWG8 pipelines to process the model spectra produced by surveys.

7.2 Algorithms and software

This section is currently TBD. The general idea of SFs is described in [RD6] and [RD7]. The exact algorithms and software are TBD.

7.3 Physical meaning of parameters

Please elaborate on the TTYPE COMMENT in Table 6 to provide a one paragraph description of each parameter that could be used to inform users of the worldwide data releases about the meaning of those parameters. Include what units are used and the meaning of any restrictions on data range.

7.4 Flag Definition Table

The DXUs of object selection functions does not contain any flags that need separate tables.

Appendix A Interface Control Drawings

Please insert any relevant drawings here.

Appendix B List of Acronyms

List of Acronyms	
4MOST	4-metre Multi-Object Spectroscopic Telescope
4OR	4MOST Operational Repository
4PA	4MOST Public Archive
AD	Applicable Document
DMS	Data Management System
DRS	Design Reference Surveys
DXU	Data eXchange Unit
FIBINFO	Fibre Information
FITS	Flexible Image Transport System
GSF	Geometric Selection Function
ICD	Interface Control Document
IWG	Infrastructure Working Group
L1-3	Level 1 to 3 Data Products
N/A	Not Applicable
OB	Observation Block
OpSys	Operational Systems
OS	Operations Scientist
OSF	Object Selection Function
PS	Project Scientist
QC	Quality Control
RD	Reference Document
TBD	To be defined
SF	Selection Function
SNR	Signal to Noise Ratio
VO	Virtual Observatory