

SKA metadata recommendations

WP3+ WP5 joint meeting
Bologna, Nov 12 th

1.5 h discussion

contacts: E. Liuzzo, M. Massardi, K. Rygl

Background

Metadata Definition

It is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier.

Types

From the National Information Standards Organization (NISO):

- ***Descriptive metadata:*** used for discovery and identification
- ***Structural metadata:*** how the components of an object are organized.
- ***Administrative metadata:*** information to help manage the source. (technical information, rights management metadata and preservation metadata)

Goals

- **What information should be in the metadata?**
→ content
- **How to compute/extract this information?**
→ methods
- **How to access to the metadata?**
→ Tools for quering

To consider...

- **SWGs requirements**
 - more complete as possible info from the beginning
- **General archive miner** point of view
 - not only radioastronomer!
- **SDP requests**
 - to be able to re-process at ESDC
- **lessons learn** from existing archives:
 - information clean and clear is needed

Standards in radioastronomy: Image metadata

Typically images are stored in **FITS files** (v 4 of 13 August 18)
https://fits.gsfc.nasa.gov/standard40/fits_standard40aa-le.pdf

Mandatory Keywords

- **Principal:** SIMPLE , BITPIX, NAXIS, NAXISn , END
- **Conforming Extensions:** XTENSION, PCOUNT , GCOUNT

Other Reserved Keywords

- **Keywords Describing the History or Physical Construction of the HDU:** DATE, ORIGIN, EXTEND, BLOCKED
- **Keywords Describing Observations:** DATE-OBS , DATExxxx , TELESCOP, INSTRUME, OBSERVER , OBJECT
- **Bibliographic Keywords:** AUTHOR, REFERENC
- **Commentary Keywords:** COMMENT , HISTORY , Keyword Field is Blank,
- **Array Keywords:** BSCALE , BZERO , BUNIT , BLANK , CTYPE_n , CRPIX_n , CRVAL_n, CDELT_n, CROTAN, DATAMAX, DATAMIN
- **Extension Keywords:** EXTNAME , EXTVER, EXTLEVEL, INHERIT

Additional Keywords: Requirements + Restrictions

Suggestions...

<https://docs.google.com/document/d/18f62tvZ1Xfmqv6VKzolwGYgOVAFDQxJHJYs68bqcQjY/edit?usp=sha>

1. *data acquisition and reduction kwds:* info from raw data

KRATARG, KDECTARG, KUVRANGE, KBAND, KMINPRLBL, KMAXPRBL, KMAXANGSC, KPADLIST, KNANT, KMINEL

2. *Image description kwds:* info from images

KBNDRES, KCHNRMS, KSPATRES, KSTOKES, KDYNRNGE, KBNDCTR, KBNDWID, KFLUXTOT, KCALIB, KTIMERES, KCONTSUB, KSELFAL, KDECONV

3. *others:* info from proposal, schedule, etc

KPROJID, KSB, KZSOURC, KSOFTVER, KFOV, KEXPTIME, KAVGT, KAVGF, KQAFLG ,KGRID

→ ***Add the kwds derived from the study of ObsCore for Italian Pulsar Search***

https://docs.google.com/document/d/11UKi8_XKsfznqnGjVdBSjjWhRW3vd8bCMEDJuqh1EQU/edit

→ ***For all kwds a clear definition is needed!!***

Suggestions

- catalogues specifying the source properties (position, flux density, size, etc.)
- surface brightness distribution of the sky as a function of position and frequency
- clean components table
- description of the point spread Function of the observations
- spectra line catalogue of at least the most brilliant sources in the field
- moments maps and masks used to make them
- signal-to-noise maps
- velocity field and velocity dispersion maps at various resolution
- Others?

Metadata content

We do not know what will be the SKA data format, however:

- will we start from FITS info? If yes, what will we recommend?
 - go through the v.4 standard FITS document?
- what else?
 - go through the WP5 SKA metadata recommendations?
 - descriptive, structural, administrative types?
- how to know more in details SWGs and SDPs requirements and non radioastronomer users?
 - surveys?
- uniformity of metadata content between archival images is required for full scientific exploitation of the archival
 - Implication on methods to obtain them (see after)

Methods to obtain the metadata

- How to extract (and ingest) the needed info?
 - Info extracted by ESDC tools or by PI or combinations?
 - Looking at the existing tools? What are they? Or give general possible design?
 - Guides where the procedures will be clearly described
 - validation problems (see later discussion)

Tools to query the metadata

- Which will we recommend?
 - an existing one or give a general design?
- If an existing one, why this?

Standards in astronomy

STANDARDS

- **AVM - Astronomy Visualization Metadata:** a standard defining discovery metadata for fully rendered astronomical imagery.
- **FITS - Flexible Image Transport System:** Used by the astronomy community to originally describe telescope images, but is now a family of standards to describe multi-dimensional data including spatial, spectral and temporal dimensions and the scaling and distortions that may be present.
- **International Virtual Observatory Alliance Technical Specifications:** A set of specifications, including metadata standards, that enable the integration of many astronomical archives into an international virtual observatory.
- **SDAC - Standard for Documentation of Astronomical Catalogues:** Used as an alternative to FITS for archiving astronomical data in a form that is more accessible to humans and standard Unix command-line tools.
- **SPASE Data Model:** An information model for describing the elements of the heliophysics data environment.

Standards in astronomy

EXTENSIONS


- **FITS World Coordinate System (WCS)**: An extension of FITS that enables data to be defined to specify physical, or world coordinates within each pixel in an image
- **IMPEX Data Model**: A simulation extension to the SPASE data model
- **Resource Metadata for the Virtual Observatory**: Defines metadata terms and concepts necessary for discovery and use of astronomical data collections and services.

Tools

AVM Adobe Metadata Panels

A set of metadata panels that can be added to Adobe Creative Suite 4 applications to allow **AVM**-compliant metadata to be entered directly into images.

AVM Web Tool

A web-based tool for assembling an **AVM**-compliant **XMP**  packet for insertion into an image file.

FITS Image Software Packages

Software packages that display or manipulate the relatively simple class of FITS data files that containing 2-dimensional images

GAVO DaCHS - Data Center Helper Suite

The software that underlies the German Astrophysical Virtual Observatory (GAVO), packaged so that it can be used to set up other Virtual Observatory-compliant data centres.

Saada

A tool for publishing astronomical data files as online databases suitable for integration into the international Virtual Observatory.

SDAC Tools

A set of four tools for working with **SDAC**-compliant archives: acut can be used to edit text files in a columnar fashion; trcol removes or transforms columns from a text file; anafit verifies that data files conform to their description; tofits converts **SDAC** tables to FITS tables.

SPASE Metadata Editor

A web-based editor for generating **SPASE** descriptions.

SPASE Tools

The **SPASE** website's list of tools for working with **SPASE** metadata and the **SPASE** framework.

And also INAF VO Tools:
VODANCE, TASMAN, IA2TAP

Standards in astronomy

Use Cases

Aus-VO - Australian Virtual Observatory [↗](#)

An initiative to provide a distributed, uniform interface to the data archives of Australia's major astronomical observatories, and to archives of astrophysical simulations, as part of the international Virtual Observatory.

CDS - Centre de Données astronomiques de Strasbourg [↗](#)

CDS (Centre de Données astronomiques de Strasbourg/Strasbourg astronomical Data Centre) is a member of the ICSU World Data System, specialising in astronomical data and related information. Its VizieR catalogue system uses SDAC to organise data and metadata holdings.

ESA - European Space Agency [↗](#)

ESA is an international organisation with 20 Member States that coordinates the financial and intellectual resources of its members to facilitate space related programmes and activities. Image data is released using FITS for many missions.

Euro-VO - European Virtual Observatory [↗](#)

A project to integrate and enable data analysis techniques over the astronomical archives of Europe, as part of the international Virtual Observatory.

HEASARC [↗](#)

The High Energy Astrophysics Science Archive Research Center (HEASARC) is the primary archive for NASA's (and other space agencies') missions dealing with electromagnetic radiation from extremely energetic phenomena ranging from black holes to the Big Bang.

IVOA - International Virtual Observatory Alliance [↗](#)

The International Virtual Observatory Alliance (IVOA) is an organisation that debates and agrees to technical standards that are needed to make the Virtual Observatory possible. FITS is one of those standards.

JAXA - Japan Aerospace Exploration Agency [↗](#)

Japan's independent administrative institution that handles research and development in space and aviation areas.

NSSDC SPASE Registry [↗](#)

The National Space Science Data Center's registry of SPASE-described space science mission data.

SPASE Inside [↗](#)

The SPASE website's list of systems that use SPASE compliant metadata to enable search services.

Spitzer Space Telescope Website [↗](#)

The image gallery for the Spitzer Space Telescope employs [AVM](#) within its content management system, and all the images have AVM tags embedded within them.

STEREO Science Center [↗](#)

Archive for STEREO telemetry, mission support data, and higher level instrument data and analysis software.

US Virtual Astronomical Observatory [↗](#)

An initiative to integrate astronomical archives and provide associated research capabilities in the United States as part of the international Virtual Observatory.

Virtual Solar Observatory [↗](#)

Extra slides

Suggestions...

@IT ARC

AKF: ALMA Keywords Filler project

E. Liuzzo et al. 2018

- Python code implemented in a CASA task
http://www.alma.inaf.it/index.php/ALMA_FITS_Keywords

```
-----> inp()
# FITSkwd :: FITS keywords
imName      =      ''      # Input image name
kwdlist     =      ''      # Select kwds: '' ==> all kwds will be
                        # calculated
outfile     =      ''      # Name of disk file to write output,
                        # ''==>to terminal
include     =      ''      # Include the calculated kwds in the
                        # FITS header, True or False. Default
                        # is ''==> False
```

- It returns a dictionary of kwds that could be included in metadata
- The FITS code is working on different types of data,
NOT ONLY ALMA e.g. VLA, VLBA, EVN

Suggestions

@IT ARC + TS

KAFE:

the Key-analysis Automated FITS-images Explorer

Burkutean et al. 2018

<https://arc.ia2.inaf.it/kafe.php>

Web interface providing with automatic processing:

- FITS kwds on the base of the AKF code

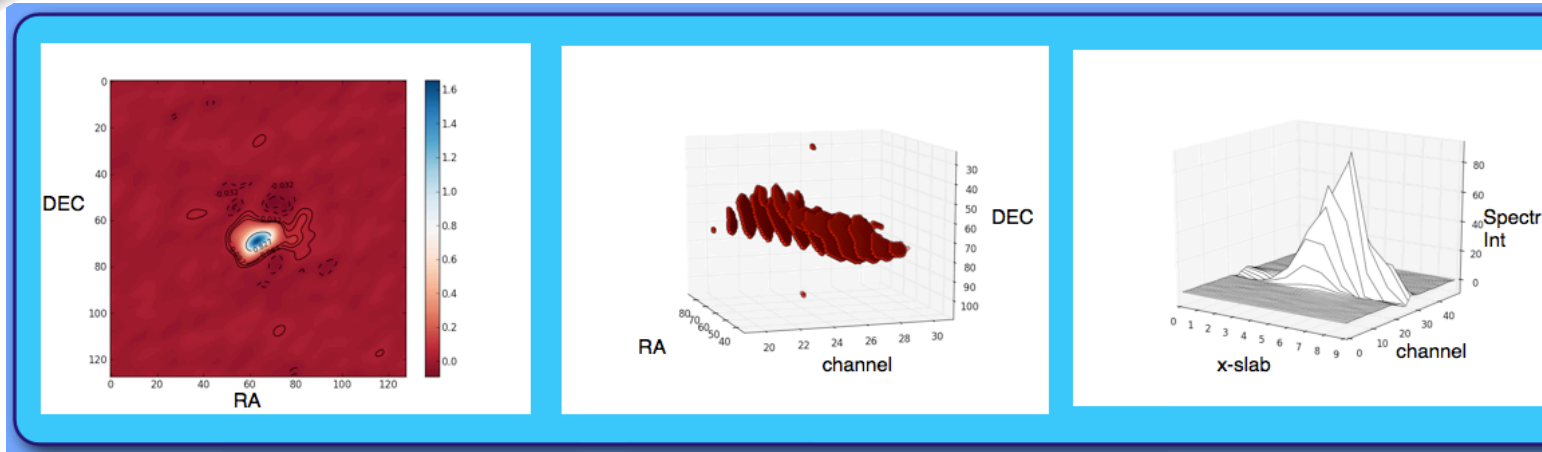
AND ALSO

- advanced image analysis plots
- catalogue cross-matching
- minimum input of user is required – the image computations and the required parameter settings are fully automated

Registration is needed



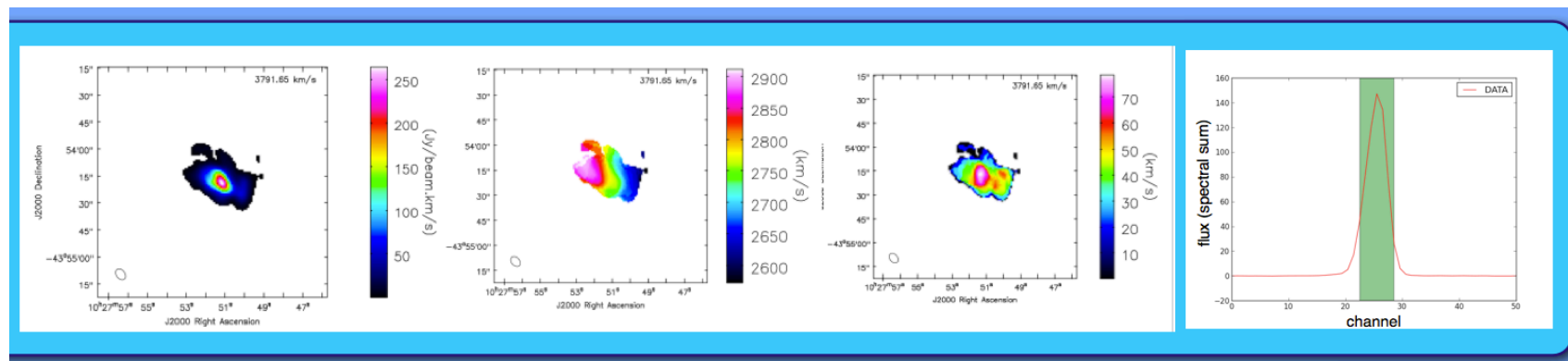
KAFE



Maximum channel map

3D view of the cube

3D spectral slabs



Moment 0,1,2 maps

line detection + spectral fit



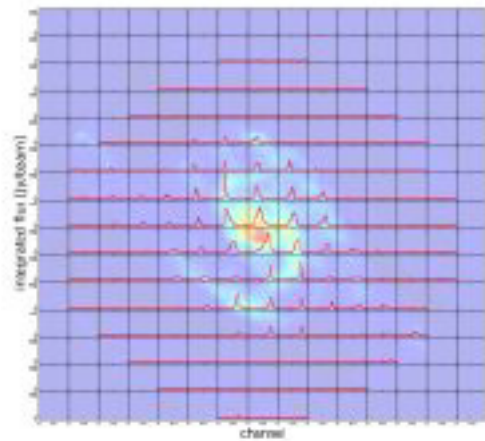
KAFE



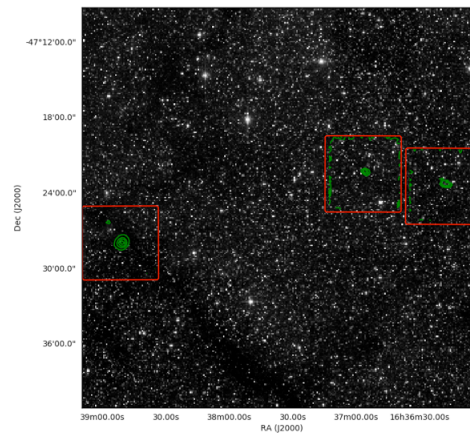
```
#####test.fits#####  
RA:          20.916 deg  
DEC:          33.256 deg  
SPATRES:      0.002 arcsec  
BNDCTR:       5022427500.0 Hz  
BNDRES:       64000000.0 Hz  
BNDWID:       64000000.0 Hz  
CHANRMS:      6.14e-05 Jy/beam  
DYNRANGE:     26.24  
FLUXTOT:      0.0023 Jy  
DATAMAX:      0.0016 Jy/beam  
DATAMIN:      -0.0001 Jy/beam  
STOKES:       ['I']  
#####test2.fits#####  
RA:          18.258 deg  
DEC:          33.2560645833 deg
```

Image Keywords

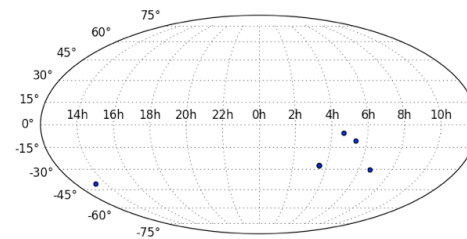
Spatial spectra



composite field plot



catalogue query





KAFE



```
#####test.fits#####  
RA:          20.916 deg  
DEC:         33.256 deg  
SPATRES:     0.002 arcsec  
BNDCTR:      5022427500.0 Hz
```

Please register your interest by writing to kafe@ira.inaf.it

Notes:

- Flexible for new features

Sp

ery

