This mail is not only to describe what we have discussed in Trieste (main session and side meetings) but also to give my personal views on a number of issues....

DISSEMINATION

A great part of what we have discussed was dealing with dissemination or how to make IVOA partners make use of characterization.

Concerns fall in different categories:

- a) How do I decide which values I have to put in which characterization DM attributes?
- b)Is characterization able to deal with my favorite data? Eg : 3D spectroscopy, Long slit spectra ,visbility data, polarimetry data, X-ray or gamma data, etc....
- c) When I know that, how do I publish that on the VO?
- d) How do my users can use these characterization data for different use cases ?

For a), it is possible to use Alberto's Micol tutorial document. I also wrote some slides for the last Euro-VO DCA workshop, based on ACS images examples. Igor wrote a couple of examples based on ASPID-SR data ore more recently Giraffe... These are all pragmatic "by example" approaches On the other side Fabien would like all the char stuff to be firmly based on theoretical grounds. I will discuss his proposal below in relationships to Level 4 stuff.

For b), we know from Igor's work that Char is perfectly able to manage 3D spectroscopy data. Discussions are going on with Catherine Boisson for gamma-ray data. Anita studied the "complex visibility" data case as well as the polarimetric data. I think she has shown how we can do it, but we have to build a full example. Igor has also shown that it is possible to give some descriptions of snapshots from simulations in char terms.

As for c), there is now a lot of technics which have been explored to format and serialize characterisation data and publish them.

For utype/based (VOTABLE formats) we can tight the attributes to SIA1 query responses (like in SaadaDB, or VOTABLE output of Aladin image server). SSA does it naturally by its characterization utypes. The hope is to get this in the same way for images and cubes in SIA2

DM mapper and MEX also allow to match utypes with outputs of the VO services.

For the XML version, CDS developped Camea to build and check interactivly valid characterization documents, and more recently a FITs2Char mapper, allowing to map FITS keywords and their combination into char files....

d) is probably the more critical part, because no VO application fully makes use of characterization.

In utype mode, Aladin only displays meta information with their utypes but makes no filtering on Field values according to utypes as it does for UCDs. TOPCAT can plot query response records against char fields, but utypes are not directly visible on the Axes.

SaadaQL and Igor's ASPID-SR interface as well as Alberto's implementation of ISO images log allow queries based on utypes constraints all using SQL oriented queries as will probably do extensions of ADQL in a near future. But none of these usages is fully standard at the moment.

Direct use of the xml structure has been implemented in ASPID-SR search engine, and is extensivly tested at CDS in relationship with Workflow developements and in experiments with E.Auden for Observation/VOEvent matching ... But we need a fully working science case along these lines to be fully understood by the community (this is what Alberto said since allready some time, and this is also a conclusion of the last char exercice in the EURO VO DCA tutorial in Garching).

USE CASES AND PRIORITIES

After that we discussed Use cases for Extension of characterization data model towards a

full observation model (including a characterization version 2, Provenance, DatasetID curation and Access reference) We had to prioriterize on that.

In Trieste (and before) we listed the following features/use cases as necessary to be added

- Transmission curves of 2D images
- PSF or beam profile for Spectra, Images, radio data
- Characterization,DataSetId and curation + whatever linkage in Xml
- complex data : resolution and sampling changes on sub-observations
- complex data : support estimated for different parts
- combining data: resolution for a coadded file (from individual detailled characterization)
- combining data: transmission curve from a coadded file...
- reduction stuff,
- various displays and views of a science dataset.
- Characterization of Theory data and linkage to SimDB data model

High Priority was set on

a) Transmission curves

b) fine resolution description including variability maps, PSF and mappings

c) explicit demand for Characterization, DatasetID, linkage

d) linkage to d ataset progenitors or derived products (cutouts, views, composition, compressed images).

A BIT OF ORGANISATION

How do we organize these features using

a)the concepts of characterization level 4, characterization of complex data and Provenance

b) ascendant compatibility with current IVOA char

Let's start from what has been described in the Char Dm document and interprate it:

Level 4

Level 4 is an apriori estimation of how an input signal will be affected by the observation process. Damping and smearing are the main phenomena Damping of an input signal in the parameter space is sensitivity and is the ground for definition of coverage. Smearing of the signal directly gives what we call resolution. It varies in width but also in shape with the position in parameter space.

Level 4 or fine level description of sampling maybe a little bit more cumbersome to define Because anything we use so far in charac is defined in terms of a given sampling. Generally the same sampling will be used for data and Damping or smearing information. But in which terms has the sampling information to be represented? If it is irregular we may have difficulties to describe it in its own terms. The exact form of the sampling function, if not trivial will have to be described in a continuous mode: in practice with a much higher sampling.

It is obvious how we can infer support, and lower levels from the level 4 for a simple dataset. But what about a complex one? In that case the inferred support may be made of different subparts, each of them with very different ranges of resolution or sampling. We may be able to partition our observation in "sub-observation" in that case. The really specific part for each subobservation will be a set of support+ resolution+sampling for each set in the partition.
This naturally introduces characterisation of complex data as global for level 1 and 2 but for each part of the partition we also need something specific going to level 3 and linking together sampling, resolution and coverage, that is a full IVOA char for each part !!!!

Characterisation of a dataset is not all what we want to know about an observation or

dataset because char is a static view of a dataset which is actually part of a process of information transformation through the observing path and data processing. So for a full knowledge of these data, linkage to previous stages of the data process will be necessary. The previous stages will be again described in term of characterization. In addition we have to describe the process leading from
the previous stages to the dataset : this is actually the provenance information.

Level 4 has been described above as the damping and smearing of the signal by the observing process. Actually this definition lack to show that this is very statistical in nature. This is basically what Fabien thought to be required for a correct definition of Char
So we can see char as the probabilty distribution of observing events for a given input signal.
Hereafter and contrary to Fabien, I will consider the mapping function as a deterministic process outside the scope of characterisation. Generally this mapping is well descfribed through WCS or WCS-like technics.

Basically I have input signals coming from the outside multiaxis world, and the observing process theoretically maps this at a given position in the internal grid (contrarily to fabien I do thing the observing device has internal axes). Each signal is made of a set of events (eg photons) and we can study the distribution of these events around the theroretical mapping position. This distribution shows both the damping (by its amplitude) and the smearing(by its shape and width) affecting the signal during the observing process. So it is both resolution and sensitivity or variablity map for all the axes together

If we want to adress the things in practice what we need is decomposition of the probability distribution along the axes, and into the properties.

We may write the distribution at a given position as the product of an amplitude function and a dimensionless "PSF". This will split sensitivity/coverage from Resolution.

We may also project the distribution on each of the output axis . And this will give us the distribution for each axis, asa function of all the physical axes. Some variable may vanish if axes are independant.

This statiscal view is a good introduction to the discussion of comapring IVOA char and characterization in SIMDB. Mireille and I discussed this with Gerard justr after Trieste.

RELATIONSHIPS with SimDB
A bit of datamodel theory (see Mireille for details) to begin

- At top we have Domain or Analysis data model: it is a very general model for concepts not intended to be implemented in practice.
- The "Logical model", derived from domain model is in UML. SimDB and IVOA char belong to this category
- The physical model is in practice an xml schema (char one eg)or a relational schema...

The discussion with Gerard showed that we could reconcile the characterization for Theoretical data and the one for observations at the Domain level.
At the Logical level we may be much closer tan we are today...

Because SimDB consider characterization for "properties" of objects in Object collections we have to consider how we can apply this "object collection" concepts in the case of observation.

Actually objects can match pixels and collection of pixels match our "datasets"...
Each of the properties match with our "axes".

SimDB doesn't have need for Resolution, sampling and error apparently (only Value) but it can be seen has a peculiar case (and actually I am not totally convinced).

SimDB can evoluate the CharacterisationTYpe content towards "reference value", "bounds","support", "distribution", much closer our four levels.
Reference value matches the "Location" of our IVOA char and distribution will be matching the level 4 "maps" or sensitivity

Generally the IVOA characterization data model is dealing with "a priori" distributions ("flat fields") while
the main interest of simDB is in "a posteriori" distributions. But Actual value distributions , and support can be
estimated for Observational data as well. So the flavor of the characterization (a priori or a posteriori) could be usefull as well...

Actually in SimDB there is a mediator class between ObjectCollections and Properties
This is called Valueassignment and contains values, resolution sampling and error. If we look at this as the starting point of our matching between the two datamodel Char classes, this drives to the conclusion that SimDB which is a more active process description than IVOA dealing with "stratic" datasets is probably "Property first" in the sense of IVOA char than Axis first. It is probably possible to better match both sides doing so.

All this shows that SimdB characterization and IVOA charcaterization for observations could be inferred from the same domain model char which remain to be defined in detail ....

TRANSMISSION curve use case

The spectral axis for an image : unsampled

We need to know the spectral shape of what can be considered as a spectral « pixel ». It is a peculiar case of spectral sampling "level 4" as discussed above. This description may be considered as constant for any position in space, time, polarisation, flux. So what we will need in that case is a single function.

It looks very much like a spectrum but with a few restrictions (NO Target, no Location in space, no location in time,etc ...) .... So a serialisation of transmission curve values using spectrum datamodel in eg VOTABLE will be rather well adapted.

We can hook to such a transmission curve from our level 4 structure but with a few additional information.

We may say that:

It is a simple function of lambda (no map of values, or matrix or map of matrices for the spatial axis)

It is a Dataset, a function or a set of moments (just choose),

we may give the URI in case of Dataset

We may give the Model/serialisation: here SpectruM/VOTABLE

we give (allready) additional documentation

what else ?

beam profile / psf use case / changing resolution:

Which part of the extended object is observed in the spectrum ?

Can we separate merged objects by deconvolution?

What we need is again a function giving the PSF. It can be constant, it can be varying on the spatial ,spectral or time axes... The shape may also be indifferent and the PSF WHM only given everywhere.

To describe this again we may use a functional or a matrix view. The WHM is a moment of

the functional view.

In the serialisation we have to say:

The variation along the other axes ("sampling" of the PSF variations)

If it is matrixrepresentation or

functional (nature, parameter)

moments (WHF + sometimes others)

Common things in the two use cases:

We give a variation domain on the other axes: may be constant as in transmission curve or single PSF across the field

We give the description type :map ("matrix") , functional (nature,parameters) or moments (including the case with one single moment = local sampling period or PSF / LSF WHM)

Provenance of complex data: Combination of observation

Beside Carac of the combined observation we need a new "Provenance" box containing Pointers towards the members Metadata (eg their own carac and provenance, etc....)

We describe the nature of the provenance : here combination...

We describe the algorithm : coaddition, drizzling, statistical fusion , etc ...

(we may give some parameters links to weight maps, etc ...)

Level 3 or 4 of the combined observation charac may sometimes be inferred from the level 3 or 4 of the individual data