



IUCr forums

Discussions on IUCr projects and activities

Discussion of namespace conventions for community-developed CIF dictionaries

Snapshot of recent discussion thread, 14 August 2013

[COMCIFS namespace and dictionary policy](#)

by [jamesrhester](#) » Tue Jan 15, 2013 2:34 am

Could I get some comments on the following statement of IUCr namespace and dictionary policy? Can we improve on this? I think we have a chance to act on the Madrid discussion of the status of different dictionaries from external providers in the context of namespaces as well. If all goes well at this stage I will present this to COMCIFS for further discussion.

Draft COMCIFS dataname and dictionary policy within the IUCr domain

COMCIFS must ensure the uniqueness of all extant datanames within the IUCr domain. The following policy is designed to maximise the chances that the status and meaning of any dataname encountered in the IUCr domain is unambiguous. A dataname is considered to be within the IUCr domain if the proposed `_audit.discipline` dataname has the value 'IUCr'.

1. *Datanames not explicitly approved by COMCIFS and appearing in CIF datafiles should either contain the string '[local]' or commence with a prefix handed out by COMCIFS*
2. COMCIFS makes no undertakings as to the uniqueness of datanames containing the string '[local]'.
3. In the register of approved prefixes, COMCIFS may provide certification that datanames with a given prefix will be unique. In order to obtain this certification, a prefix assignee should:
 - publish a publically-available dictionary defining all datanames with that prefix
 - have an organisational structure judged capable of enforcing dataname policy (a single person also suits this criterion)
4. Alternatively, if a prefix assignee provides to the IUCr a dataname dictionary and advises that the prefix is no longer in use, the IUCr will archive that dictionary and certify that the prefix is unique. If later workers wish to re-use such a 'closed' prefix, they must not define any items that appear in such archived dictionaries.
5. The IUCr cannot provide any guarantees as to the correctness or uniqueness of definitions in dictionaries published by third parties. COMCIFS may choose, on request, to bring such third party dictionaries into the IUCr domain, in which case datanames and details of definitions may change.

Last edited by [jamesrhester](#) on Wed Jul 17, 2013 2:18 am, edited 2 times in total.

[Re: COMCIFS namespace and dictionary policy](#)

by [jcbollinger](#) » Tue Jan 15, 2013 7:35 pm

jamesrhester wrote: **Draft COMCIFS dataname and dictionary policy for bodies within the IUCr domain**

I think it would be better and clearer to delete "bodies within" from the policy title.

COMCIFS must ensure the uniqueness of all extant datanames. The following policy is designed to maximise the chances that the status and meaning of any dataname encountered in the IUCr domain is unambiguous.

Those sentences seem inconsistent with respect to the scope of the policy or COMCIFS' responsibility. I suggest appending "in the IUCr domain" to the first.

1. *Datanames not explicitly approved by COMCIFS and appearing in CIF datafiles should either contain the string '[local]' or commence with a prefix handed out by COMCIFS*

This is probably the place to set policy on a question that persists in our other thread of discussion: to what, if any, extent may data names from different domains appear in the same data file? We are agreed, I think, that data names from different domains should not be mixed in the same save frame or directly in the same data block, but there are other, better-controlled possibilities that we might want to allow:

What about a save frame drawing from a different domain than its host data block?
 What about a save frame drawing from a different domain than a sibling save frame?
 What about different data blocks in the same file drawing from different domains?

I'm satisfied with the rest of the proposed policy.

John

Last edited by [Brian McMahon](#) on Mon Jan 21, 2013 1:36 pm, edited 1 time in total.

Reason: *Tidied formatting (quoted list)*

[Re: COMCIFS namespace and dictionary policy](#)

by [jamesrhester](#) » Tue Jan 29, 2013 6:12 am

I'm happy with John's suggested adjustments.

jcbollinger wrote: This is probably the place to set policy on a question that persists in our other thread of discussion: to what, if any, extent may data names from different domains appear in the same data file? We are agreed, I think, that data names from different domains should not be mixed in the same save frame or directly in the same data block, but there are other, better-controlled possibilities that we might want to allow:

What about a save frame drawing from a different domain than its host data block?
 What about a save frame drawing from a different domain than a sibling save frame?
 What about different data blocks in the same file drawing from different domains?

John

I guess that these are questions of scope: does the value of a dataname in a host block apply within a save frame? Can it be overridden in the save frame? Is it still overridden when that save frame is referenced from the enclosing host block? Do we apply any restrictions at the level of the standard or at the level of the dictionary?

Let's see: a save frame exists in a data block in order to be referenced from that data block - it serves no other purpose in a non-dictionary file. Therefore the save frame contents will be drawn into the host data block's data ontology and should therefore be at least anticipated by that ontology. Likewise and equivalently, a programmer writing a program to read in data from CIF files needs to know beforehand what to expect in the way of datanames and save frame contents (if referenced). As a final point for consideration, the usage of save frames and save-frame references envisioned in the demonstration DDLm dictionaries (see Spadaccini et. al. at <http://dx.doi.org/10.1021/ci300075z>) is explicitly controlled by the dictionaries through a 'ref-loop' category type. For example, experimental data described using datanames from the 'experiments' category and arising from multiple experiments can be bundled in separate save-frames and then references to those save frames can be looped over in the main datablock.

As a thought experiment - what if an IUCr-domain CIF-file wanted to refer to experimental results from a different domain in

this way, perhaps as a way of explaining the origin of a particular quantity used in further calculations? Firstly, such a use would need to be anticipated in the DDL dictionaries by using some sort of mechanism to indicate that the sub-category describing the save frame contents is equivalent to a category defined in some alien dictionary, as specified by a separate column (values from an enumerated list). That alien category would then be notionally grafted on to the parent category in the ontology tree. Given this description in the dictionary, programmers could then anticipate the likely contents of the column based on the dictionaries from that other domain. Perhaps more attractively for non-IUCr users, non-IUCr users could graft on crystallographic results in this way (e.g. astronomers referencing mineral structures detected in space). But this is all very speculative of course.

Note that in this example the initial decision as to whether or not to refer to foreign domains is with the dictionary writer, rather than the CIF file producer; therefore there is a chance to assess other domains for longevity of datanames etc, and the likelihood of encountering unexpected ad-hoc inclusion of other domains' datanames is reduced. Dictionary writers also have the option of importing chunks of other dictionaries directly into their own dictionaries. Therefore I think that it should in principle be possible for save frames to refer to different domains, but the wisdom of doing this is not something I think we can come to a conclusion on here.

To allow the above possibility, but reduce the amount of checking of the audit.domain tags in save frames, I suggest that the scope of the `_audit.domain` dataname (and indeed of any dataname) includes child frames *unless* it is overridden by that tag appearing in the child frame. The answer to your three questions would then be:

- (i) A save frame may draw from a different domain than its host data block. We note that this will only be useful if the relevant dictionary has explicitly allowed for this possibility
- (ii) Likewise sibling save frames are not constrained to come from the same dictionary
- (iii) Likewise data blocks may come from different domains.

[Re: COMCIFS namespace and dictionary policy](#)

□ by [jcbollinger](#) » Tue Jan 29, 2013 5:26 pm

jamesrhester wrote: To allow the above possibility, but reduce the amount of checking of the audit.domain tags in save frames, I suggest that the scope of the `_audit.domain` dataname (and indeed of any dataname) includes child frames *unless* it is overridden by that tag appearing in the child frame. The answer to your three questions would then be:

- (i) A save frame may draw from a different domain than its host data block. We note that this will only be useful if the relevant dictionary has explicitly allowed for this possibility
- (ii) Likewise sibling save frames are not constrained to come from the same dictionary
- (iii) Likewise data blocks may come from different domains.

That would suit me well.

John
