What is ISOcat?
ISOcat is a web-based implementation to store and make accessible concepts (a concept registry), more specifically data categories, that are relevant for the CLARIN infrastructure and for encoding linguistic phenomena. Basically it provides a persistent identifier for each data category, and a variety of properties of the data category. It allows one to uniquely refer to a data category (using a PID, e.g. this (link is external)) under abstraction from language-specific (e.g. English 'noun' v. French 'nom') and arbitrary differences in notations for data categories (e.g. 'noun' v. 'n'). This will make it possible for all kinds of resources and tools to 'interoperate', not only on the format level, but also on the level of content. A large list of data categories, mainly originating from the ISO TC 37/SC 4 project (which itself based its selection on earlier projects for best practices and standards such as EAGLES and ISLE) has been created. These data categories are currently considered as candidates for official inclusion in ISOcat and some have already been accepted (but all are already accessible for inspection, comments etc.). Of course, in some cases the same expression is used for two or more different concepts, sometimes dependent on a specific theoretical view on the matter. But ISOcat is open, one can add one's own concepts, and even organize a whole group of related concepts in a so-called 'profile'. Currently, ISOcat is basically a flat list of data categories, each with its properties. Data category specifications can be associated with a variety of data element names and with language-specific versions of definitions, names, value domains and other attributes. It is the intention to add, in a next stage, relations between concepts. This will allow one to specify many types of relations between concepts, e.g. that one concept is a hyponym of another one; that two concepts are not completely identical but very close; using such relations one can specify multiple hierarchical ontology's on these concepts, etc. etc.

Does it contain relations?
No - the data model was set up with the explicit intention to not include relations, since these in most cases are dependent on theories and practical intentions. Therefore a framework will be offered that allows users to easily manipulate and share relations according to their needs. From CLARIN we intend to offer at least one set of relations with a large coverage which users may want to use or manipulate.

Does it solve all semantic interoperability problems?
No - it is just a start to offer a reference, so that users creating new resources could use the registered categories and schemas describing legacy data can refer to them. But we found that not all tag sets which are in use for various purposes can easily be mapped on another one. It also will largely depend on the intended usage. For searching an imperfect mapping may result in less precision, but for a researcher this may not be a problem.

How should I work with ISOcat in my project?
For each concept that occurs in your resource or in the metadata of your resource, you should check whether a corresponding concept already exists in ISOcat. If this is not the case, you will have to add the concept in ISOcat. You will also have to make a formally represented mapping between the notations for concepts that occur in your resource or its metadata, and the PIDs of the corresponding ISOcat data categories. For example, if you use "zn" as the notation for the concept of 'noun', this mapping will have to include:

zn is-a (link is external). The way to express this relationship between a concept and an ISOcat data category depends on the resource format. In XML documents the DC Reference vocabulary (see here (link is external)) can be used like this:
<myResource xmlns:dcr="http://www.isocat.org/ns/dcr (link is external)">
<partOfSpeech dcr:datcat="http://www.isocat.org/datcat/DC-1345 (link is external)" dcr:valueDatcat="http://www.isocat.org/datcat/DC-1333 (link is external)">noun</partOfSpeech>
</myResource>
In this case the references are directly inserted in the instance document, but it's also possible and in many cases preferable to include them in the schema of the resource, e.g., in an XML Schema or Relax NG document. Also resource specific languages, e.g. ODD or TBX, have their own specific ways to declare the relationship between a data element and a data category.

Where can I obtain more information on ISOcat?
Concept Registry Short Guide (link is external) and ISOcat website (link is external). For technical questions, you might also send an e-mail to the clarin-nl helpdesk.

How do I adjust the font size in ISOcat?
This is possible in most web browsers by using the zoom function, in Firefox this is done through View > Zoom > Zoom in

How do I make data categories publicly available?
Select the data category, add this through the "+" icon to a Data Category Selection (DCS). Use "save the selected data categories" for the DCS and provide it with a name, after which the antenna-icon will be usable.

How do I contact the owner of a data category?
Use the ISOcat forum - see this site for details.

What should I enter for "origin" or the "source" (in a language section)?
In origin you can mention the inspiration source for the creation of the data category. If you do not know what to enter, please enter CLARIN. For the source of the language section, please enter CLARIN.

Should I use capitals for the name of the data category?
No, unless it is required by certain language rules (e.g. nouns in German), the name of a data category should not contain capitals.

How do I add a new data category for the use with CMDI?
Register at isocat.org
Send a mail with your name to cmdi@clarin.eu (link sends e-mail)
You will be added to the CLARIN group
My Workspace > Shared > CLARIN > MD > button “edit this data category selection”
My Workspace > button “create new data category”
When you save the new data category you'll be asked if you want to add it to the basket, which is the nickname of the are where the MD data category selection is edited, click yes and the new data category is added to that selection
Click on the icon for “save the selected data categories”

After inspection the new data categories can be moved to the Metadata thematic view (Finally, and optionally the data categories in the Metadata thematic view can be submitted to the Thematic Domain Group for official approval)

What is the granularity of the definitions?
There is much debate about this and there is no good universal answer yet. However, we need to start using the ISOcat registry to find out how the definitions can be improved, which categories are missing and which granularity should be chosen for metadata, morphology and semantic annotation to just mention a few examples.

Does ISOcat have cool URIs for the mimetypes?
It's unclear at the moment of ISOcat should replicate the contents of other registries. The IANA lists of mimetypes would be an example of such registry, the Dublin Core set of metadata element would be another. ISOcat could function as just a proxy, by not assigning it's own PIDs to categories from these registries, i.e., it just provides access no actual registration. However, it could only do that if these registries provide some own kind of persistent URIs. Dublin Core does in the form of PURL URLs, but currently we're not aware that IANA provides something similar, e.g., a cool URI, for the mimetypes.
Is there a standard reference document? Yes, there is a formal ISO document that counts as reference document. This can be bought here (link is external). Note that all relevant information from the standard should be processed in the (help pages (link is external)) or on the ISOcat 12620 page (link is external)). If information seems to be lacking there, this can be seen as a ISOcat feature request and will be treated as such.
What is a good place to start for people that are not introduced to ISOcat yet? Well, a good place would be the CLARIN-NL helpdesk portal (you found it already!) A short guide can be found here (link is external). A manual is featured on the ISOcat website (link is external). Especially useful is the ISOcat screencast (created by CLARIN) that covers pretty much all functionality of ISOcat. Also useful are the help pages (link is external), that offer a "Getting Started" page and contain many links to information sources.

What are the possibilities for the processing of large amounts of categories?
These possibilities are covered by the "DCIF import process". DCIF stands for the Data Category Interchange Format which is part of the ISO 12620:2009 standard. Its schema is available at the ISOcat 12620 web page (link is external). A basic requirement here is a valid DCIF document. Subsequently (or when greatly in trouble to get the document valid), one can contact the ISOcat system administration at isocat@mpi.nl (link sends e-mail). The system administration will then handle the actual import into ISOcat for you.

Why can't I find all elements in ISOcat when editing my profile in the Component Registry?
If you search for ISOcat categories in the Component Registry, a service of ISOcat is invoked that only searches through the ISOcat list for metadata categories. Categories that are private in ISOcat but also public categories which aren't a member of the metadata profile cannot be found in this manner. However, the component registry editor also allows you to paste in an ISOcat data category PID. This could be used for private testing of CMDI components and profiles. For public CMDI profiles all data categories should be public too and should member of the metadata profile.

What is the goal of ISOcat?
The goal of ISOcat is to provide a registry of, possibly standardized, data category specifications which can be shared among users and especially resources. By sharing data categories the semantics associated with them, as described in their specification, are shared. These shared semantics help to achieve a level of semantic interoperability between resources. For example, if both resource A and B refer to data category 1 these two resources may indicate that they contain information relevant to the users of both resources. This semantic overlap can be determined even when the resources are structurally very different.

Suppose one accepts a certain definition or provides a definition in your own specific domain and LINK to a more general one. What will happen to this link if the more general definition is changed in order to improve it? It can be the case that an improvement for a particular person is not an improvement for others. The same is true for descriptions.
Change of semantics is indeed a problem of data categories owned by users. For standardized data categories, which are owned by a Thematic Domain Group (TDG), a versioning mechanism is in place, which means that it's always possible to refer to a specific version which a specific definition. It is envisioned that the same versioning mechanism will become available for all public data categories, i.e., including the ones owned by users.

What strategy is chosen for communicating about changes to a linked element or about the addition of a new definition for an already defined term that is also present in another domains?
For each data category a change log is kept, it should be possible to use that to implement some form of a change tracker. There is now an ISOcat feature request (link is external) for such a change tracker.

Almost none of the current ISOcat data categories are yet part of an official standard. There are often multiple candidate data categories in ISOcat to use. How can we determine to which one we should map to? If mapped to one that will later not become a standard, the mapping would have to be redone
Indeed ISOcat currently (October 2010) doesn't contain any standardized data categories, as the implementation of the standardization workflow has only recently become available and TDGs still have to come up to speed. However, for the currently active TDGs the lists of active members is available. As these TDGs are preparing data category selections for standardization they are the owners of data categories likely to be standardized. Currently one has to keep these names in mind to select data categories which are likely to be standardized:
for the metadata profile data categories owned by Peter Wittenburg or Daan Broeder;
for the terminology profile data categories owned by Sue Ellen Wright; and
for the morphosyntax profile data categories owned by Gil Francopoulo.
However, there are also sets of data categories which can be seen as de-facto standards and who might be never enter the ISO standardization process. For example, the data categories, owned by the gold-user, associated with the GOLD ontology.