

# Do's and don'ts

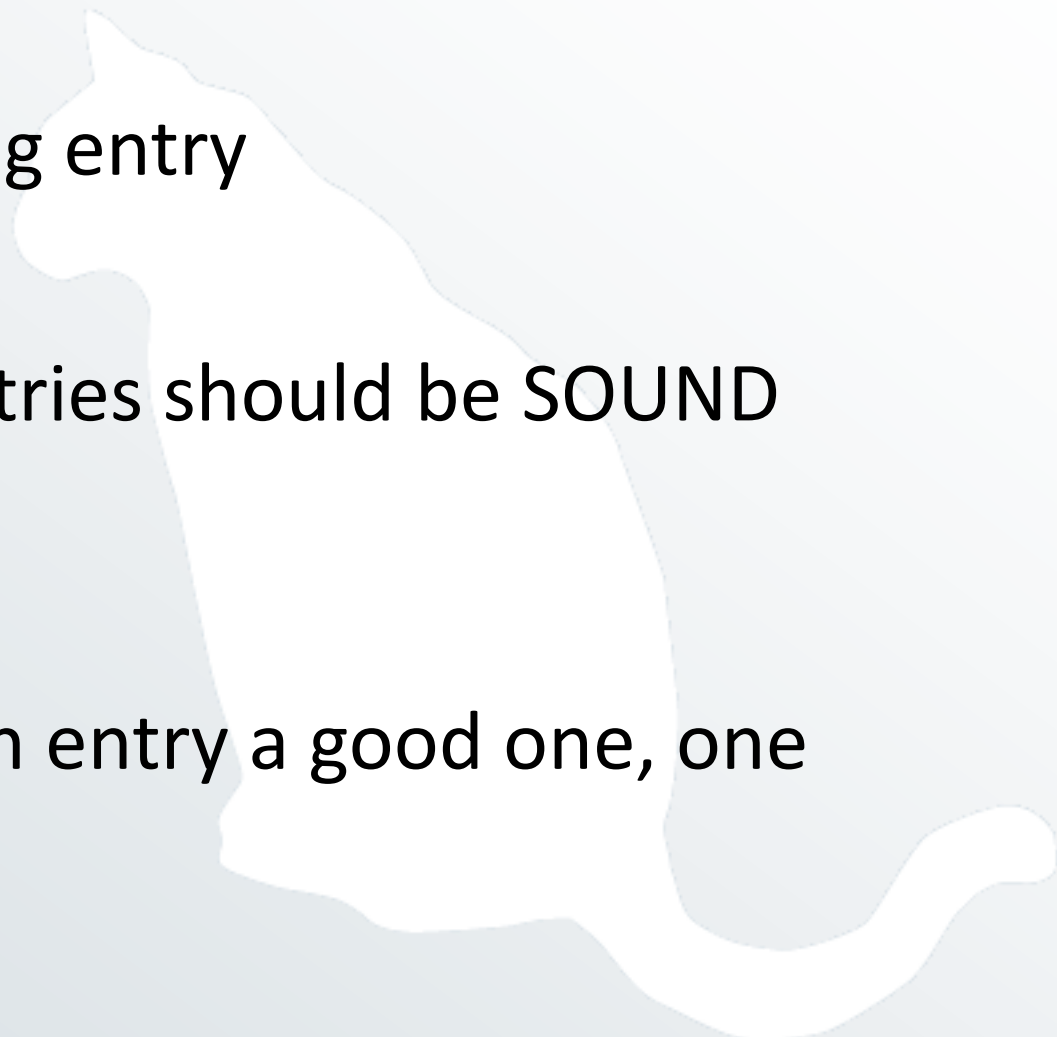


# Your work wrt ISOcat

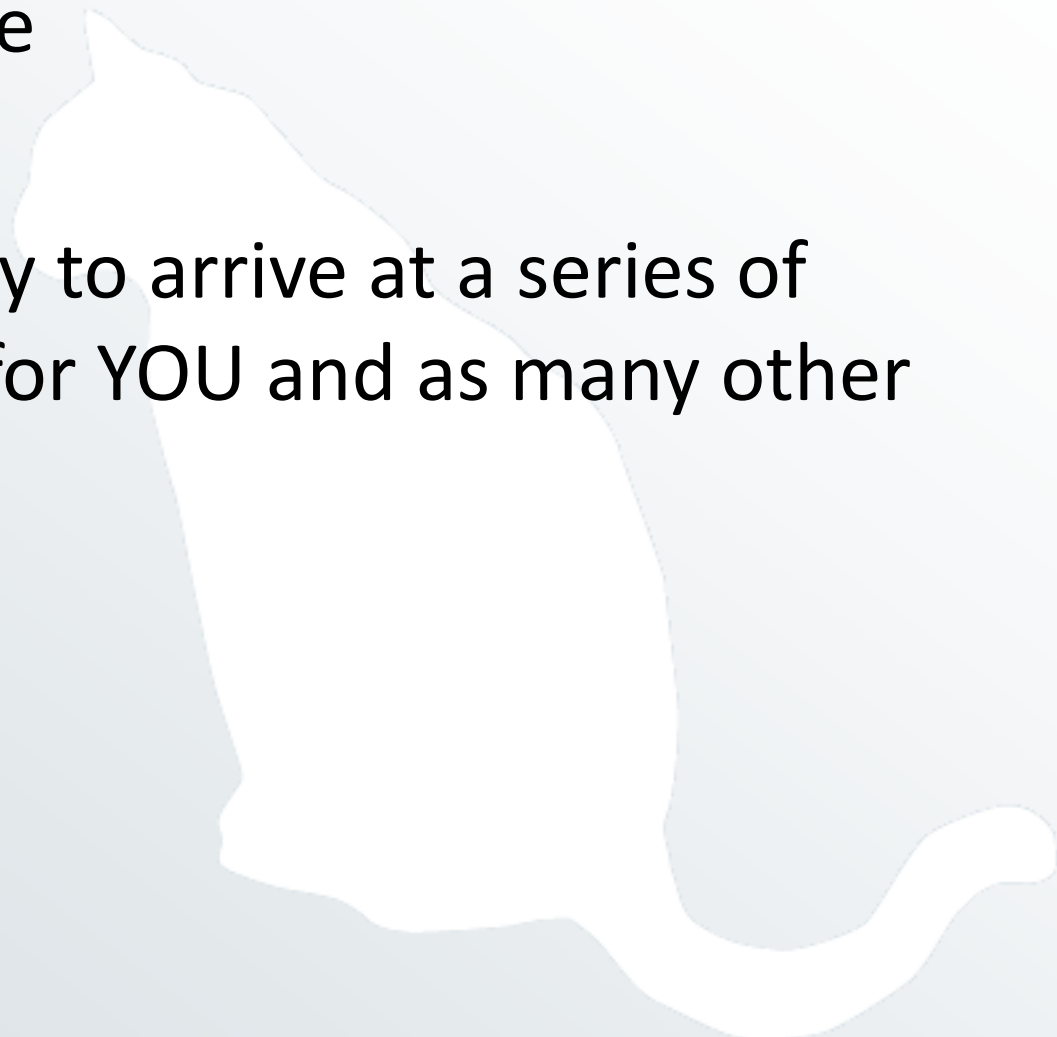
- Create an entry
- Link with an existing entry

In both cases: the entries should be SOUND ones

- But: what makes an entry a good one, one that you can use?



# Standards

- Hardly any available
  - We really should try to arrive at a series of sound DCs, useful for YOU and as many other people as possible
- 

# Matching DC

- What defines a matching DC?
  - It should ‘match’ with the way you use a specific notion in the annotation scheme, application, ... at hand
  - It should come with the same profile
  - It should handle the same phenomenon
    - **proper**  $\neq$  **proper noun**
    - **full stop**  $\neq$  **punctuation**

# What defines a sound DC?

- Meaningful definition

*Indefinite pronoun*

– Not: pronoun that is indefinite

Unless

both ‘pronoun’ and ‘indefinite’ are  
defined elsewhere AND

it is mentioned explicitly which ones are  
involved AND

these definitions are correct (for you)

# What defines a sound DC?

- Correct definition

*Personal pronoun*

– Not: pronoun referring to persons

As

*That cat has five kittens. SHE ...*

*This table was very expensive but I like IT very much*

[Note: \_in a particular tagset this definition may be correct! In general it is not.]

# What defines a sound DC?

- Reusable definition

*Personal pronoun*

Not: In CGN a personal pronoun ...

Not: In Dutch a personal pronoun ...

Not: A personal pronoun (*ik, ikke* and *ikzelf*) is characterized by ...

A definition should be as neutral (project, language) as possible, while still valid for your purposes!

# Flagged DCs

- Never link with 'deprecated' DCs !
- In other cases the flags show whether the DC specification is correct from a technical point of view.
- Note that only DCs with a green marking are qualified for standardization



# DC/DCS and profile

- Profiles are not added automatically, a DCS may contain elements with various profiles
- In case the profile you need is not yet available, contact Menzo

# What to include?

- In general: all linguistically meaningful notions mentioned in your schema, manual, definition
- Abbreviations (PST for /past tense/) are to be mentioned as Data Element Name

# Do's & don'ts

## Do's:

- Create a DCS for your scheme (name project, annotation scheme, ...)
- Provide clear definitions (short, to the point) for your scheme, application, ....
- Take care not to leave concepts used in your definitions undefined or vague
- Use appropriate vocabulary (per profile)
- Check 'adopted' DC's regularly till standardization !

# Do's (continued)

When creating a DC, fill out

- Justification: used in XYZ, part of tagset N

PLUS **why a new DC is needed !**

- Language section
  - Always English language section
  - Strong recommendation: sections for object language(s), for working language (like language in which manual is written)
  - Sections in the various languages should match (+/- be translations of each other)

# Do's (continued)

When creating a DC, fill out

- Example section
  - Note that \*negative\* examples may be very helpful!
- foreignWord
  - Dutch language section
  - example section: the, house, NOT: poster
  - explanation section: een woord als 'poster'
  - heeft Nederlandse diminutief: postertje, itt house (\*housje, \*houseje)

# Example sections

Suppose you want to illustrate a Dutch phenomenon:

- Ex.sec. in EN language section
  - Dutch ex with transl in English
- Ex.sec. in DE language section
  - Dutch ex with transl in German
- Ex.sec. in EN linguistic section
  - EN example
- Ex.sec. in DE linguistic section
  - DE example with translation in English

# Don'ts

- Confuse Language and Linguistic section
  - Latter contains language specific values for closed Confuse domains
- Be (too) language specific in definition
- Mention scheme in definition
- Use several definitions in one DC
- Use circular definitions
- Rely on authority
- Rely on standardized status
  - Definition should fit YOUR scheme, etc
- Make a DC private once it has been public

- Questions?

