

CATEGORISING DATA

Case Study on an issue resolved by the Connemara Programme

2016

Key Points

- A viable and flexible categorisation schema was essential for the operation of the Programme web interface on myconnemara.com
- The schema also had to be compatible with external schemas.
- A two level internal categorisation was adopted along with the use of NACE codes.
- All Category and sub category names have to be comprehensible in 81 languages
- A Third level of categorisation is being examined for greater granularity
- NACE coding is not used to service end user information queries.

The Issue

The categorisation scheme for the relational data posed an issue. There are distinct differences between the local and Institutional (authorities and Institutions) perceptions of the “correct” categories for artifacts and activities in all areas addressed by the Programme. Mediating these differences requires a flexible and cross comparable categorisation schema. The schema had to be comprehensible across 81 languages.

The Approach

The approach adopted involved a three level approach to the issue.

- 1) Identification of the levels to which Data was to be categorised
- 2) Identification of local and Institutional categorisation norms
- 3) Identification of a mediation schema that delivered cross comparability

The Result

A Two level schema of internal categorisation is presently in use. This provides a parent Main Category and sub categories as children of this parent. This was adopted as it provides a simple schema and for web presentation reduced the navigation overhead to end users. 51 Main categories and 459 sub categories are presently in use. Over 1400 Category / Sub category pairs have emerged from data capture and categorisation. A third level is presently being examined. This will be a Sub / Sub category layer.

The local schema involved identifying the “normal” term for a business, artefact or activity. These are then checked against non local schemas (National and Linguistic). Agreed ones are loaded onto the system and Categories and Sub Categories are then allocated to each item entered to the data set. These are what end users select to access information and are also used to create category and sub category interactive maps.

The NACE (Nomenclature Generale des Activites Economiques dans l’Union Europeenne (General Name for Economic Activities in the European Union) schema was adopted to provide external compatibility. The numeric not text definitions are used. The NACE information is for internal use and not displayed to the end user on the web.