The IPTC Taxonomy – Subject Codes and Subject Qualifiers

Document history				[Document URN: urn:iptc:workdoc:ncd:0619:1]
Revision	Issue Date	Pages	Author (revised by)	Remark
1	2006-06-15	2	John Minting	

To the IPTC Newscodes Working Party

The categorization of news items is intended to facilitate both the grouping of real-time items into packages and to help in searching an archive for items that relate to a specific subject, yet the requirements for each can be very different; the former not necessarily needing the same level of granularity as the latter.

Until the advent of "artificial intelligence" and computer aided categorization, this work was a manual task which resulted in the use of a small generic list of subjects that provided enough metadata to satisfy newspapers and the like, but now huge archives and very specialist information users demand the ability to find only items related to their very specific genre quickly.

In its wisdom, the IPTC designed a taxonomy in a hierarchical structure that allows users to choose the amount of granularity they need. The second level today providing approximately 400 terms that fit the needs of manual categorization, whilst the third level at approximately 900 terms is easily supported by mechanical means.

All those I have spoken to in the last few months, archive providers and categorization engine manufacturers agree that there is a need for more granularity in taxonomies to support the growing needs of "search and retrieval". For example the BBC has a 150,000 term taxonomy for their archives; Lexis Nexis uses a 3000 term, 6 level list for general news. They also agree that a single public taxonomy that provides a core for general news has many advantages, but that single taxonomy must be extensible to satisfy a changing world.

IPTC's Subject Codes and Qualifiers are being used in many different disciplines world-wide, sometimes as is, sometimes with propriety "add-ons" where either the IPTC list is missing general terms or where the user has some very specific needs.

To satisfy the many needs that have been suggested both at IPTC and elsewhere one conclusion is to base the future on the existing IPTC taxonomy, improve the Subject Matter to reflect better the needs of those who want to limit their categorization to a general base of 300/400 terms; to allow an almost limitless expansion of the third level (Subject Detail) for general news terms and provide the specifications for proprietary "add-ons" that can seamlessly plug into the core taxonomy.

Limiting and/or curtailing the extensibility of the IPTC Subject Codes will eventually lead to its demise and its position as an academic museum piece, as those who need to add value to their product go elsewhere.

=== END of document ===