"Steroid" Scandal Rocks Major League Libraries

By Daniel Cohen - Special Correspondent for Information Science and Levity

Washington, DC | December 14, 2007

More than half of the systems responsible for managing the nearly 17 million titles in the Library of Congress catalog have tested positive for prohibited performance enhancing content, according to an Inspector General's report expected to be published next month. In a public statement issued by the IG, he stated that "The use of taxonomies, controlled vocabularies, and other performance enhancing content often referred to under the catchall term 'metadata' has long been suspected by Library fans. For much too long, Library owners have simply turned a blind eye, choosing not to question how their high-priced key assets had been able to break one long-standing performance record after another. As long as they were seeing the performance, nobody wanted to question or acknowledge how it was being achieved."

The investigation was triggered by anonymous tips and overheard conversations between certain unnamed Library staff members alluding to "a card catalog on steroids." In a possibly-related development, recent published observations by certain professional Library experts that the renowned Library of Congress Subject Headings (LCSH) now resembled a "galaxy" have led some to wonder if the impressive bulk of not only the catalog, but of many finding aids, classification indexes, and other "players" in the Library were also being artificially enhanced.

According to an Integrated Library System Program office staffer who insisted upon anonymity, some of the obvious increased bulk is a predictable outcome of the adoption of new rules and policies intended to promote diversity in the languages and scripts supported by the Library. For example, according to the anonymous source, the adoption of the Unicode standard, "may have sounded quite benign to most casual users of Library catalogs, but it requires twice as many bytes as traditional data encoding standards used in earlier decades. "If the storage data census looks a lot bigger, that's because it *is* a lot bigger."

The expense of producing the high-quality metadata found in the Library of Congress systems leads many observers to believe that the problem may be limited to major league cataloging libraries. However, according to an Association of Research Libraries spokesperson, "For a long time we have seen evidence of metadata finding its way into the systems at the college library level, and even on the local public library level."

Expert analysis of the metadata found in these so-called minor league libraries suggests that much of it was probably produced at a major league institution like the Library of Congress. According to National Science Foundation metadata expert Mia Culpa. "You see amateur efforts at producing metadata by kids and wannabe catalogers. This stuff is known on the street by such names as 'folksonomy' and 'social tagging.' Some of it is even pretty powerful, but producing the real thing requires a level of expertise you won't find in many high school kids or web surfers. To get the kind of performance boost seen in the major leagues libraries, you need metadata made by people who really know what they are doing."

Many feel that the cost factor will limit future creation of such "pharmaceutical grade" metadata. According to Library Science experts, the discontinued support of "metada laboratories" at institutions like the Library of Congress will eventually cause the supply of high quality metadata to dry up. "When you no longer have places like the Library of Congress creating this powerful stuff, you will see the performance of library searching systems come back down to the level that nature intended," said one anti-metadata activist.

Some commentators feel that the crackdown on high-quality metadata labs is wrongheaded. "We have seen that this stuff works to enhance performance. Why not make the top quality stuff available to everyone? That's the only way to really level the playing field," says metadata advocate Harley Trion. "If we close down the labs creating high-quality metadata, you will see widespread adoption of street-quality metadata like social tagging and folksonomies, because that's all you will be able to get. I'd rather know that my kids were using metadata that is made in a clean lab with experts and quality assurance processes than have them experimenting with street-grade metadata of unknown origin and quality."