“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.”

Continued on A-49