Fixing Broken Records, Exploring the Value of Big Data

The 1964 Supreme Court case of *Jacobellis v. Ohio* produced a memorable quote relating to obscenity: "I know it when I see it."

Unfortunately, records and information management (RIM) challenges aren't so clear. While records management is well-defined in the hard-copy world, it's a different picture in the digital domain.

New Tools, More Documents, More Troubles

Trouble first surfaced in the early 1980s with the advent of word processing and computer-generated documents. As documents proliferated, enterprise content management solutions (ECMs) entered to save the day.

Then, widespread Internet adoption in the mid-1990s triggered a data surge that nearly drowned ECMs. Caught off-guard, vendors devised a "work-around" by using human judgment to compensate for lack of scalability: having end-users manually select and tag records, thus dramatically reducing the processing load. Problem solved … but not really.

Using Human Judgment as a Solution

The byproduct of this ECM work-around was that human capital was utilized, with all its attendant problems. Using humans for highly repetitive processes not only cost more, it also came with inherently human errors, including inconsistency, bias, insufficient training, and apathy.

By the late 1990s it became apparent that end-user classification of e-mail and files was headed for a brick wall. Why?

Causes for RIM's Perfect Storm

First, the volume of data spawned by e-mail alone had grown several orders of magnitude larger than anything the ECM solutions were architected to handle.

Second, document types had expanded to an ever-larger array, including not just e-mail and files, but also instant messages, social media, and others. Indeed, even the definition of a relevant record is now questioned, with the debut of "non-records," which fall outside current definitions of a business record but which are, nonetheless, important to business.

Third, the ecosystem began to raise the stakes in managing such digital content as regulatory agencies, such as the Securities and Exchange Commission, enacted strict retention requirements. Then, in 2006, the Federal Rules of Civil Procedure issued guidelines on electronic discovery, giving evidentiary weight to electronically stored information (ESI). This combined regulatory and legal assault underscored the need for digital recordkeeping.

Fourth, these problems prompted *ad hoc* solutions, which in turn created data silos. The rise of the silos has created intractable problems. Silos proliferate duplicates. Silos search inconsistently since each one has its own search engine. Silos have disjointed retention since each one has different retention capabilities. Combined, these issues result in loss of data control. Data silos now constitute the most challenging hurdle to managing digital content.

Seeking the Right Approach

Industry pundits are starting to opine on how best to proceed. Stalwarts, such as the National Archives and Records Administration, have begun questioning ECM work-arounds – namely, end-user classification – and a rising chorus of voices recommends "taking the end-user out of the records management business."

Even More Changes Lie Ahead

As textual analytics improves, the potential of intelligence to be gleaned from such data is attracting attention from top management, especially for extracting competitive advantage. Paradoxically, concerns over legal exposure still prompt many to purge "unnecessary" data. But there remains the nagging question of whether we are throwing the baby out with the bath water. The strategic value of such data is just beginning to be explored, the consequences of which could trigger another wave of changes in managing future content.

Guessing the right approach for this long-term problem will be critical to avoid very expensive re-directions along the way.

Perhaps the question isn't so much whether we'll know it when we see it, but whether we'll see it before it hits us in the face.