What Did We Learn About the Systems We Analyzed?
White Paper Written by Philip Bantin

In the first phase of the project, we reviewed several large systems in the financial and student records functional areas. We concluded that although data managers were very aware of the pertinent laws, regulations, and statements of best practices, the information systems were seriously flawed as recordkeeping systems. In particular, we noted that the systems did not: capture and preserve inviolate records; migrate records on a regular schedule; include complete and detailed documentation; and generate enough metadata, particularly contextual metadata, audit trail data, and management metadata on such issues as retention or preservation strategies.

We concluded that “we suspect that most of these deficiencies are not just confined to our test cases but will be present in the majority of the information systems on our campus. On the basis of discussions with numerous data stewards and managers all over campus, it is clear that disposition and use history data are not available in any form for most systems. And the problems of no defined mechanisms for migration (as opposed to simple back-ups) and of no systematic means of maintaining records inviolate are we fear endemic problems inherent in information systems.”

Experience reviewing IU systems in phase II of the project has confirmed our fears about the recordkeeping deficiencies of the information systems. I can say without hesitation that the primary data and information systems employed by IU and likely most institutions are poor recordkeeping systems. To illustrate this point, I would like to review the characteristics of the primary systems in use at IU, and follow this up with a discussion of why they are poor recordkeeping systems. The three prominent types of systems used at the institutional level at IU and in most organizations are On-Line Transaction Processing Systems, Decision Support or Data Warehouse Systems, and Electronic Document Management Systems.

On-Line Transaction Processing Systems Employing DBMS Software
The most basic business system and the heart of most organizations is the On-Line Transaction Processing system (OLTP). A transaction processing system is a computerized system that performs and records the daily routine transactions necessary to the conduct of business. The primary goal of these systems is to automate computing intensive business transactions, such as those undertaken in the financial and human resource functional areas. The emphases are on processing data (sorting, listing, updating, merging), on reducing clerical costs, and on outputting documents required to do business, such as bills, paychecks and orders. The guiding principles of these systems are to create data that is current, up-to-date, accurate, and consistent. To achieve these goals, these systems employ Database Management System (DBMS) software. One of the primary advantages of DBMS is its ability to limit and control redundant data in multiple systems. Instead of the same data field being repeated in different files, the information appears just once. Another advantage of DBMS is that it improves data
integrity. Updates are made only once, and all changes are made for that data element no matter where it appears.

**Decision Support Systems (DSS):**
Decision Support Systems are “an organized collection of people, procedures, databases and devices used to support problem-specific decision making.” (Ralph H. Sprague)
Unlike operational databases that are set up to handle transactions and that are kept current as of the last transaction or update, decision support systems are analytical, subject oriented systems that are designed to assist managers in strategic planning and in handling semi-structured or unstructured problems that cannot be easily programmed. DSS are decision focused systems usually aimed at top managers, with an emphasis on flexibility, adaptability, and quick response. Their primary objectives are to aggregate transactions as snapshots in time, and to produce flexible, on-demand reports for managers.

**Electronic Document Management Systems (EDMS):**
EDMS applications focus on the control and management of documents of all types - text, images, graphics, and complex/compound documents. The ultimate objective of the system is to support the day-to-day use of documents for the ongoing business needs of the organization. EDMS functionality typically includes: document capture, indexing of documents, storage management, version control, integration with desktop applications, and retrieval tools to access the documents. EDMS more closely resemble recordkeeping systems than any other type of information system. However, even EDMS lack some of the key requirements necessary to manage documents over their entire life cycle and to create reliable and authentic records that provide evidence of critical activities of the institution.

**Why are these Systems NOT good Recordkeeping Systems?**
The primary data and information systems employed by most institutions do not routinely and systematically fulfill five major requirements of a recordkeeping system.

**Data and Information Systems do not routinely capture records**
In OLTP and Decision Support Systems, business records are not routinely stored and captured as stable, finite, physical entities. OLTP systems are primarily concerned with providing an organizational resource to current data that will support on-going business needs. Decision Support Systems are largely focused on transforming data into reports and other information packages designed to assist decision-making. Neither system is designed to capture records or evidence of business transactions, and to manage those records for all long as required.

**Data and Information Systems do not systematically preserve immutable records**
Not only will the record not be captured at the time of creation, it may be impossible to recreate at some later date. Databases are dynamic, volatile systems, in a state of continual change. Data updates occur frequently, and with database management
software managing the system, these revisions are made in every file containing that revised data element. Consequently, once a virtual view of a document is erased from the screen, it may or may not be possible to recreate the record exactly as it originally appeared. Moreover, databases typically maintain only the current value for any given data element. Historical data, if kept at all, is usually incomplete or summarized. As a result, historical "snapshots" of a database do not routinely capture the data values needed to reconstruct a specific record. While EDMS do a better job of preserving and managing original documents and versions of records, these systems typically have less control over modification and deletion than would be permitted in a recordkeeping system.

** Data and Information Systems do not systematically preserve complete, fully documented records
All of the most popular information systems fail to meet the requirement for capturing and preserving fully documented digital objects. More specifically, two major problems exist: 1) Metadata that was never created, and 2) Metadata that exists but is not linked to or part of the record. System metadata as typically defined by systems designers and technologists is often not as complete as necessary to describe a record. Transaction logs maintained in typical information systems do contain some critical data on updates and revisions, but on the whole, these logs do not provide enough essential metadata describing and documenting the business transaction. Of particular concern are the paucity of metadata related to the context of creation and use, and metadata related to the management of the record over time. This latter category of metadata would include documentation on a preservation strategy and on how long to retain the record, including references to laws or best practices and a designation in years on the length of time the record must be retained. The availability of this contextual and management metadata could very well make the difference between a useful and a useless record and a record that is properly or improperly managed. In addition, some vital metadata may not have a logical association with the content data, i.e., there are no links between or pointers to and from metadata describing and defining the record and the data comprising the record content. In some cases, the metadata may not be a part of the automated system at all, but may exist only as a paper document totally disassociated from the records it is describing. Finally, even if links or pointers do exist, the logical relationship between the different parts of the record can become disconnected over time, or the various data elements comprising record content and the metadata may be preserved or discarded according to different timetables or schedules.

** Data and information systems do not provide for the long-term preservation of records
In most institutions the typical strategy for protecting data is to generate daily back-ups of all system data. While the back-up strategy is a necessary and useful activity, it must be regarded more as a disaster preparedness plan than as a preservation strategy. Unlike a true preservation strategy, the back-up solution does not deal with the root causes of the deterioration and obsolescence of digital records: hardware obsolescence, storage medium deterioration, and the biggest problem of all, software dependence or "the fact that digital documents are in general dependent on application software to make them accessible and meaningful." (Jeff Rothenberg) Overall, the back-up strategy does not
contribute to the overall preservation goals of ensuring readability and intelligibility in order to facilitate data exchange over time.

**Data and Information Systems do not provide for the systematic retention and disposition of records**
In many systems, the primary disposition strategy consists of copying data, and moving this data to near-line and eventually off-line storage status. In this strategy, no appraisal decisions are made on what to keep or destroy and why. Some EDMS may include retention controls, but the majority of these systems do not include the rigorous controls for disposing and retaining records that must be a required feature of any recordkeeping system.

To conclude, although the most commonly used data and information systems do contain some recordkeeping functionality, they are, on the whole, poor recordkeeping systems. This is not surprising since these systems were never designed to be recordkeeping systems. Systems do only what they’re designed to do, and the purpose, goals, and scope of OLTP systems, DSS, and EDMS are clearly not to manage records over their entire life cycle. To meet the requirements of recordkeeping, institutions will need either to modify existing data and information systems or design new recordkeeping environments that can be integrated with existing, legacy systems.