The World Bank Group Archives

ELECTRONIC RECORDS STRATEGY

FINAL REPORT

June 29, 2002

Peter Van Garderen
Artefactual Systems Inc.

artefactual
systems inc.

Table of Contents

Executive Summary

In light of the new Information Disclosure Policy, the WBG Archives has conducted a study of the Bank's capability to capture, preserve and make available its electronic records on an enterprise-wide basis. This study was carried out by consultant Peter Van Garderen of Artefactual Systems Inc. (Vancouver, Canada) between February 20 and April 5, 2002.[1]

Over the past two months the WBG Archives has begun to adopt the recommendations of this study into a workplan referred to as the Electronic Records Strategy. The goal of this strategy is to operationalize an electronic records management program at the Bank which will ensure the systematic, enterprise-wide management of electronic records according to international standards and best practices. The Strategy has been developed through a series of reports, presentations and meetings between Archives and ISG staff.

The Electronic Records Strategy consists of four primary initiatives:

1. A conceptual framework for the management of electronic records
2. An archival repository for electronic records
3. Standards for the management of electronic records
4. A systematic methodology for the retention scheduling of electronic records

This report summarizes the background, status and future plans for these initiatives. Additional material on these initiatives and reports on their progress will be made available on an Electronic Records Strategy homepage on the WBG Archives website.

What are Electronic Records?

Electronic records are business information (a.k.a. information assets, content, files, documents, data) in digital format that the bank has created and needs to preserve and make available because of their:

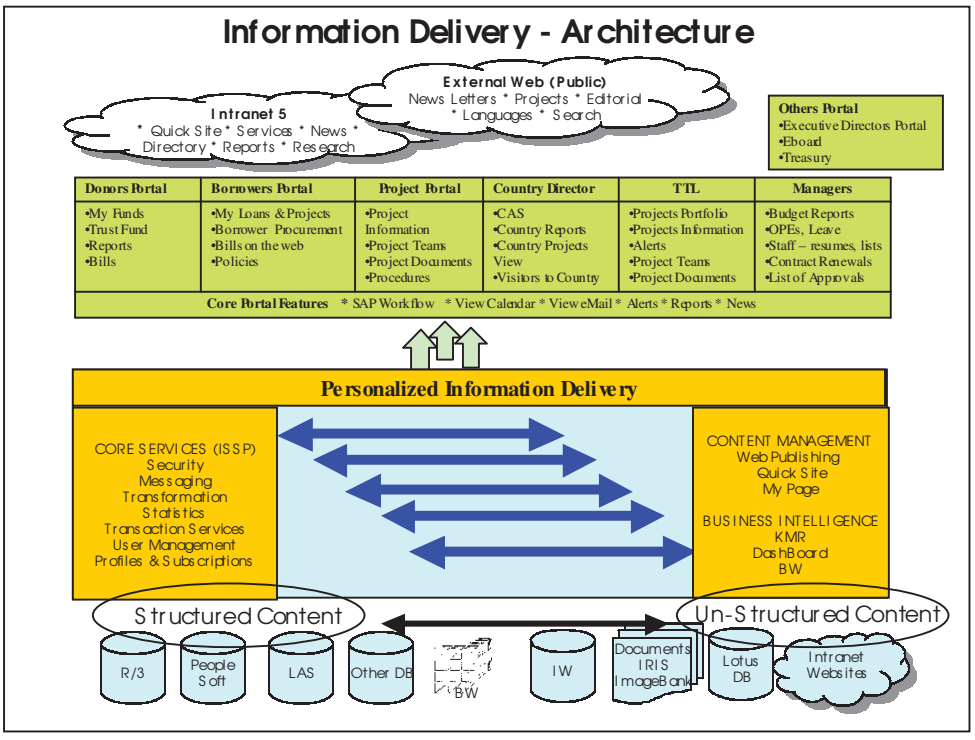
- business-use value
  - e.g. required for business processes, content re-versioning/re-purposing
- legal, regulatory requirements
- fiscal auditing requirements
  - e.g. invoices, statements
- research / historical value
  - e.g. knowledge products, economic histories
- accountability protection
  - e.g. disclosure - requests for information, evidence of actions & decisions

[1] See *Electronic Records Study: Interim Report* (April 5, 2002)

Electronic records are found in:

- Corporate information systems (e.g. SAP, Peoplesoft, LAS)
- Document management systems (e.g. IRIS 3 & 4, E-Docs)
- Email (Lotus Notes directories)
- Intranet & Websites
- In-house information systems (e.g. MS-Access databases)
- Legacy systems
- C:\Drives, Departmental Servers, Floppy Disks
- back-up tapes

Electronic Records are an important facet of the overal ISG Information Delivery Strategy. For the most part, they are the content that will be delivered via the planned information portals. In order to deliver this information we need to gain control over the repositories of structured and un-structured content in a consistent, enterprise-wide manner. This includes the 'capture' of electronic records by generating suitable metadata for their identification, retention and retrieval using the Bank's core metadata standard.[2]



[2] Developed and maintained by ISG Data Administration, see Bedford, Denise. "Appendix A: Core Metadata Data Dictionary" *World Bank Information Warehouse Content* (January 2001).

Findings: Electronic Records Study

The analysis of the electronic recordkeeping environment at the Bank revealed the following:

- as of February 2002, there was no integrated electronic records strategy to manage and standardize the enterprise-wide capture, access and preservation or destruction of the Bank's electronic records.

- existing standards related to electronic recordkeeping had not been promulgated or implemented

- electronic records issues were treated as an IT/systems problem rather than a recordkeeping (classification and retention) or cultural (education and training) problem.

- electronic records issues were dealt with on an ad-hoc basis, using approaches based on paper record management

- as with paper records, electronic records management was generally being addressed in the middle of the life-cycle, after the records had been already been created and used, instead of prior to their creation during systems requirements and design or as part of procedural requirements in the business process.

A widely accepted view amongst those dealing with electronic records management issues is that to assert any type of systematic, enterprise-wide control of electronic records they must be identified and scheduled at their point of creation otherwise it will too complicated or time-consuming to do it at a later time. This requires record capture requirements to be added to business systems that handle structured business processes and it requires employee awareness and willingness to ensure records capture in unstructured business processes (i.e. filing to IRIS). This represents the greatest challenge in gaining enterprise-wide control of electronic records. In comparison, the technological issues are relatively minor (generating and transferring metadata, rendering digital objects to storage formats, maintaining storage systems and migration strategies).

Role of the WBG Archives

Some may ask why the WBG Archives is concerned about management of digital business information across the enterprise as this is usually considered the domain of system administrators. However, the WBG Archives has been mandated the responsibility to manage all of the WBG's records (IBRD, IFC, MIGA), throughout their life-cycle. This includes records in digital format.

AMS 10.10 – 10.12: "The WBG Archives is responsible for developing any policies, standards, and procedures that may be needed:

to support the proper creation of WBG records …
...to support the proper management of WBG records …
...to support the proper retention and disposition of WBG records … "

The primary responsibility of system administrators is to keep their systems operational for business use. Without a centralized unit responsible for coordinating efforts and establishing best practices, electronic records management will continue in an ad hoc manner which means that the Bank runs the risk of mismanaging its electronic records if it treats them purely as system data. One unfortunate example of this is the recent controversy over Trust Fund records which forced the Bank to return millions of dollars to donors when it could not demonstrate proper recordkeeping. Therefore, the Bank's most valuable asset, **its reputation as a responsible and trustworthy organization**, will be at risk unless action is taken to implement an enterprise-wide electronic records strategy.

"Because of the Sept. 11 attacks, global enterprises have realized that they must assure customers, suppliers, and partners that the business is protected against internal security threats and external disaster – **such assurance has become an important part of an enterprise's overall brand value**"

- *Gartner Group Research Note (March 14, 2002)*

"In the aftermath of Enron's bankruptcy, customers, partners and suppliers will be more conscious of how an enterprise manages its vital business records. **Trustworthy partners will demonstrate that they have the systems and processes to protect those records**."

- *Gartner Group Research Note (March 14, 2002)*

The Electronic Records Strategy and the electronic records program which it seeks to launch will manage the bank's electronic records as a collective whole. This means centralized strategy, oversight, and resource allocation (e.g. enterprise storage strategy) for the full life-cycle management of electronic records from their creation through their identification, storage, retrieval, retention scheduling, destruction, and preservation. Establishing centralized control and strategies for the preservation of electronic records is especially important as this function includes more than merely storing digital objects. It includes storage media and storage location migration, file format migration, media refreshment, physical and networked access monitoring, creation of finding aids & search tools, ensuring redundant back-up, and maintaining migration strategies.

Conceptual Framework for the Management of Electronic Records

The conceptual framework provides an information architecture for the analysis and design products which will be created throughout the implementation of the Electronic Records Strategy. The framework is intended to establish a common vocabulary and vision amongst the stakeholders who will be analyzing and reconciling the various architectural layers and business viewpoints involved in the management of electronic records. It will also provide the structure and content for internal Bank standards on recordkeeping and provide tools and products for detailed analysis and design when implementing requirements in specific software systems. The framework consists of a glossary, concept diagrams, business & systems models (UML), business rules, requirement specifications, and requirement citations (standards, best practices).

The first version of these products are based on the *ISO 15489 Records Management* standard. As the Electronic Records Strategy progresses, these products will be expanded for further compliance with the DoD 5015.2 standard for recordkeeping software, the ISO/DIS Open Archival Information Systems standard and the research findings of the InterPARES Project.

The tools used to support the framework include a requirements database (MS-Access), diagramming software (MS-Visio) and thesauri software (Multi-Tes). The framework products will be posted on a project website on the WBG Archives homepage.

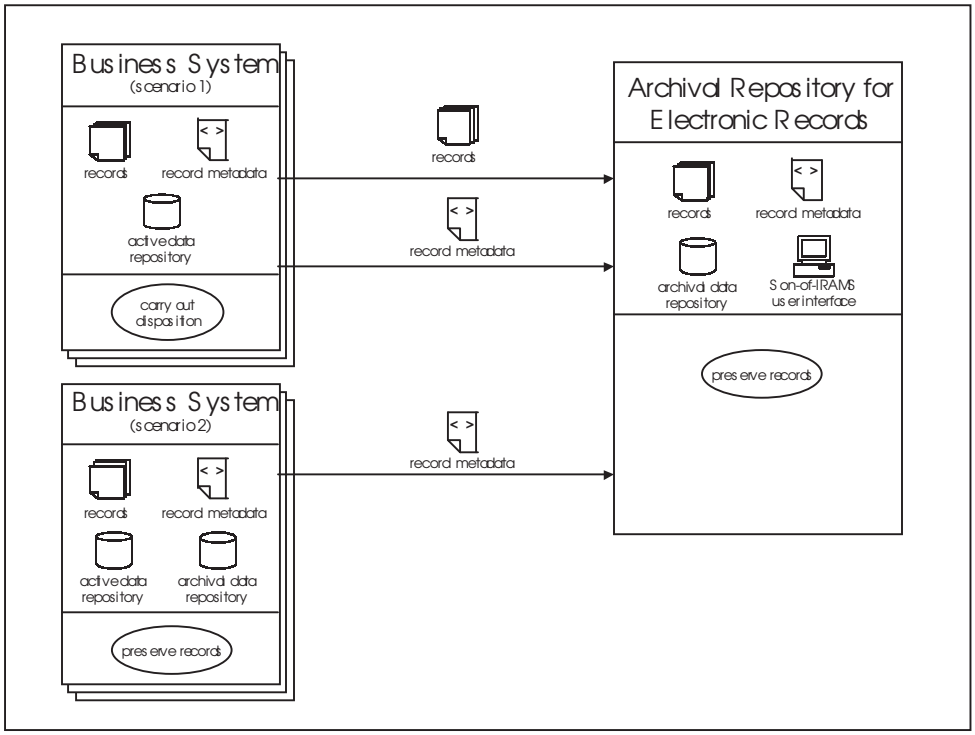
The primary concept established in the framework is that of a generic business system. A business system is defined as any combination of software, hardware, data and storage media that provides a business solution or supports a business process (i.e. procuring loans, processing invoices, creating documents, creating web-content, etc.). The term "business system" is used, therefore, in a generic way to refer to business information systems, content management systems, collaboration systems, messaging systems, e-commerce systems, etc.. This allows us to focus on the records creation and management attributes of the business system rather than the specific business functions it supports. In the course of providing business solutions, business systems create, store and manage business information. Electronic records are the subset of this business information which must be retained in an archival repository for its informational or evidential value.

Archival Repository for Electronic Records

Initially, the Archives assumed that all of the Bank's electronic records would eventually be accessioned into a centralized archival repository for electronic records. However, further analysis revealed that the large number and wide variety of electronic records in the Bank's many business systems would not make this feasible. The technical,

procedural, and organizational (i.e. political) complexity of identifying, scheduling, and transferring electronic records is significant and in the case of large, enterprise business systems (e.g. SAP, Peoplesoft) it is simply not efficient to transfer large amounts of electronic records on a regular basis. Especially since these systems have their own COTS *archiving* subsystems. The *archiving* systems provides the architecture and functionality to migrate business information out of the active data repository of the operational system to online, near-line or off-line data repositories for the purpose of improving the performance and storage capacity of the operational system. Therefore, the conceptual framework accounts for two possible scenarios.

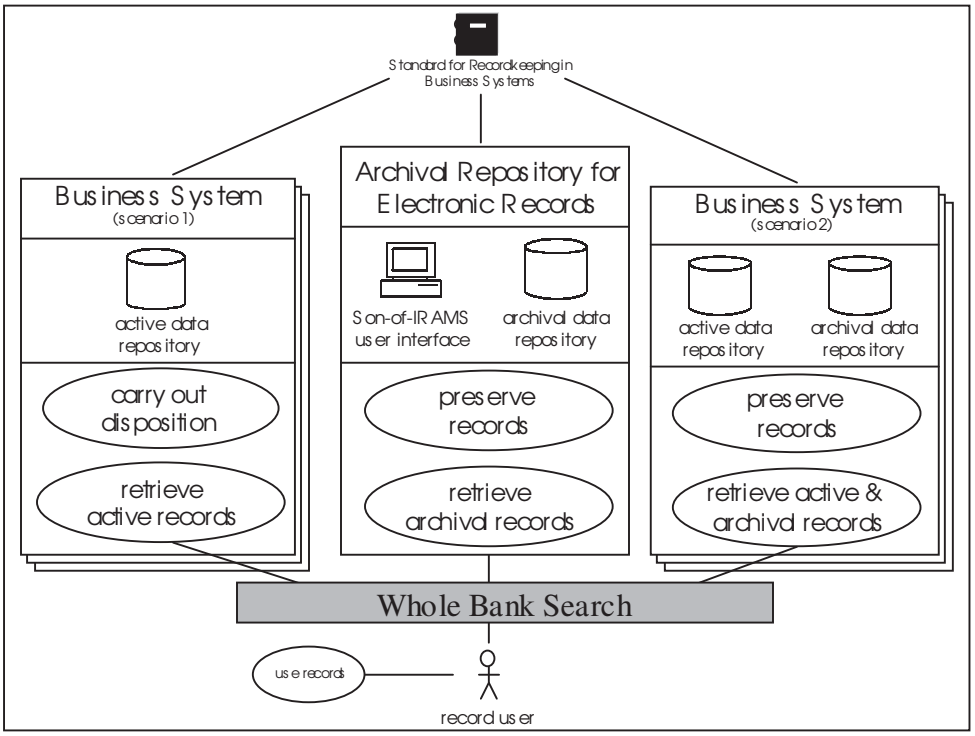
In the first scenario, business systems without archiving subsystems will transfer those electronic records which must be retained to the Archival Repository for Electronic Records along with their associated metadata. In the second scenario, business systems will use their own archiving subsystems to preserve electronic records while transferring their associated metadata to the central Archival Repository for Electronic Records so that intellectual control and access to the whole of the Bank's electronic records is centralized.



The user interface and application software used to manage the Archival Repository for Electronic Records will be the same software that the Archives uses to manage the physical repository of paper records. This system (IRAMS) will be replaced with a new COTS product ("Son-of-IRAMS") in the fiscal year 2003.

Standards for the Management of Electronic Records

The Electronic Records Strategy is closely allied with the Whole Bank Compendium & Search initiatives (led by Denise Bedford of ISG Data Administration) which seeks to provide single points of access to all of the Bank's content through web-based information portals. This initiative includes a Core Metadata Standard of 29 essential metadata attributes to which business system metadata will be mapped. The Core Metadata Standard includes recordkeeping attributes which will assist in the identification, retention scheduling, preservation and access of electronic records. The Whole Bank Search interfaces will also serve as the delivery mechanism for electronic records.

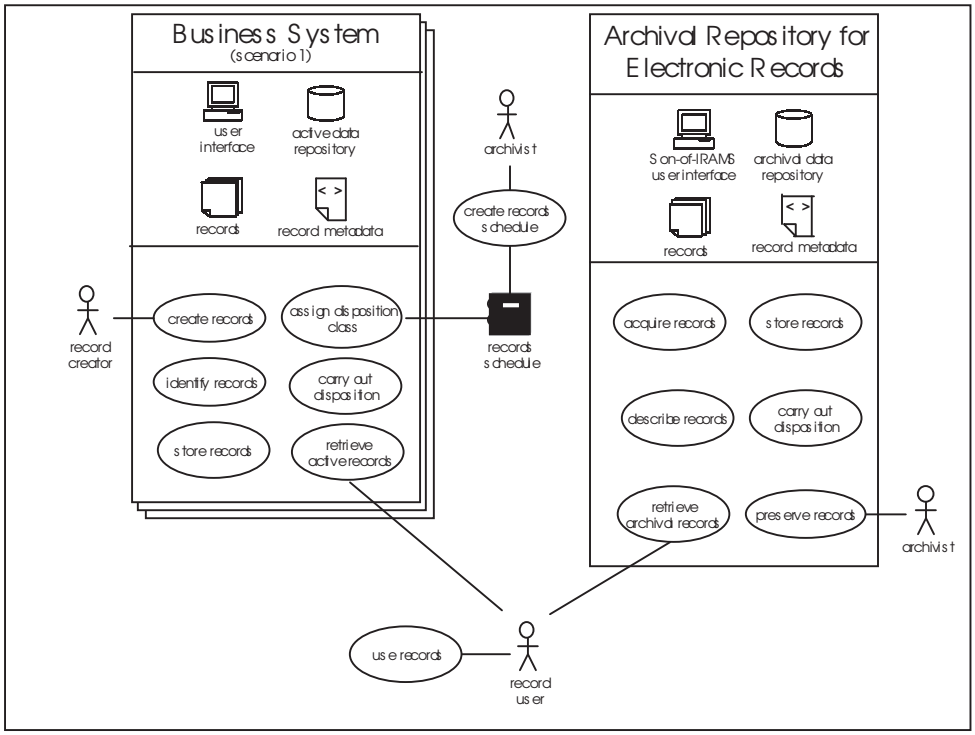


The WBG Archives will attempt to standardize the capture of the appropriate core metadata as well as the overall management of electronic records in business systems through the distribution, implementation and monitoring of the *Standard for Recordkeeping in Business Systems*. This standard is complemented by:

A. the *Standard for Electronic Recordkeeping* which is directed at Bank staff who will be using business systems to create and use electronic records
B. the *Standard for Unit Records Management* which is directed at business unit managers who will be responsible for implementing records management resources and responsibilities

The *Standard for Recordkeeping in Business Systems* contains functional requirements for specific recordkeeping processes including:

A. Record Identification	Assigning Core Metadata to electronic records to allow for their identification, retrieval and processing
B. Assigning Disposition Classes	Assigning an electronic record to a class of records with common retention periods and disposition actions
C. Carrying Out Disposition	The transfer or destruction of electronic records



The recordkeeping processes are represented by the oval shapes in the concept diagram above. Each of these oval shapes maps will decompose into more specific activity model diagrams which will specify in detail the entities, actors and processes that are involved. The key process in this domain is "Create Records Schedule" which is carried out by an archivist to identify disposition classes, assign retention periods and determine the final disposition of records (preservation or destruction) based on their business use requirements and fiscal, legal, and/or historical value.

A Systematic Methodology for the Retention Scheduling of Electronic Records

Even though the Bank's systems employ back-up procedures and storage costs are declining annually, we need to be selective about what we keep to avoid information overload. We need to distinguish between *useful* and *useless* information by **capturing** and **retaining** the right information and **destroying** the right information.

> "An information retention policy must consider who will be looking for information in the future, when that future might occur and the kinds of questions that might be asked…Failure to address these questions up front will result in **a mass of impenetrable information** that is more a liability than an asset."
>
> - Gartner Advisory (March 22, 2002)

The WBG Archives has long developed and maintained schedules for the retention and disposition of the Bank's paper records. One of the initial assumptions has been that these existing schedules would be mapped to electronic records. However, this has proven to be difficult in initial proof of concept tests. One of the main problems is that the Bank's record classification and record schedules are not integrated. Therefore, the criteria for classifying records for retrieval are not consistently the same as the criteria for identifying groups of records with the same retention and disposition requirements. This has not been a major issue in the paper environment as the physical aggregations of paper records are easier to group and analyze after their creation (e.g. they exist in filing cabinets or are transferred together in record boxes). However, electronic records are more difficult to identify, analyze and group together after they have been created due to "invisible" nature of the electronic domain, the layering of business systems (data, application logic, user interfaces), the integration and interdependency of business systems, and the lack of centralized classification and filing of electronic records created in unstructured business processes (e.g. word processing documents & email messages).

The generally accepted best practice for dealing with this issue is to introduce integrated records classification and retention tools based on business processes rather than the records' media, format, source business system, creator, custodian, etc.. The WBG Archives is currently conducting two parallel studies using the IRIS4 document management system in order to determine the most feasible way to standardize the creation of records schedules for electronic records in the Bank's various business systems. The first approach is testing the original hypothesis in which existing paper retention schedules are mapped to electronic records using the IRIS classification scheme and a disposition logic/rules engine. The second approach will attempt to re-classify the electronic records using a business process taxonomy which is directly tied to disposition classes.[3] This taxonomy will be developed in conjunction with ISG Data Administration which requires the same product to developed a business process view for access to Bank records in the Whole Bank Compendium & Search products. The findings of these pilot studies are due in October 2002 and the resulting analysis will lead to design recommendations and implementation specifications for IRIS4 and the Bank's other business systems.

[3] See Ian McAndrew (WBG Archives), "Electronic Records Strategy – Africa/Project Cycle Pilot Study" (June 2002).

Electronic Records Strategy Workplan

The WBG Archives has begun work on all of the initiatives described above. A high-level workplan is illustrated in the chart below. All of the related products and the latest information on Strategy developments will be posted on a project website which will be made available on the WBG Archives homepage. One final recommendation which will be key to the success of the Strategy and the tasks outlined below is the appointment of a full-time electronic records archivist to manage these interrelated projects and to lead the implementation of an operational electronic records management program based on the findings and products of the Electronic Records Strategy.