

Tools of the Trade – Putting Smart Technologies in the Hands of Information Professionals

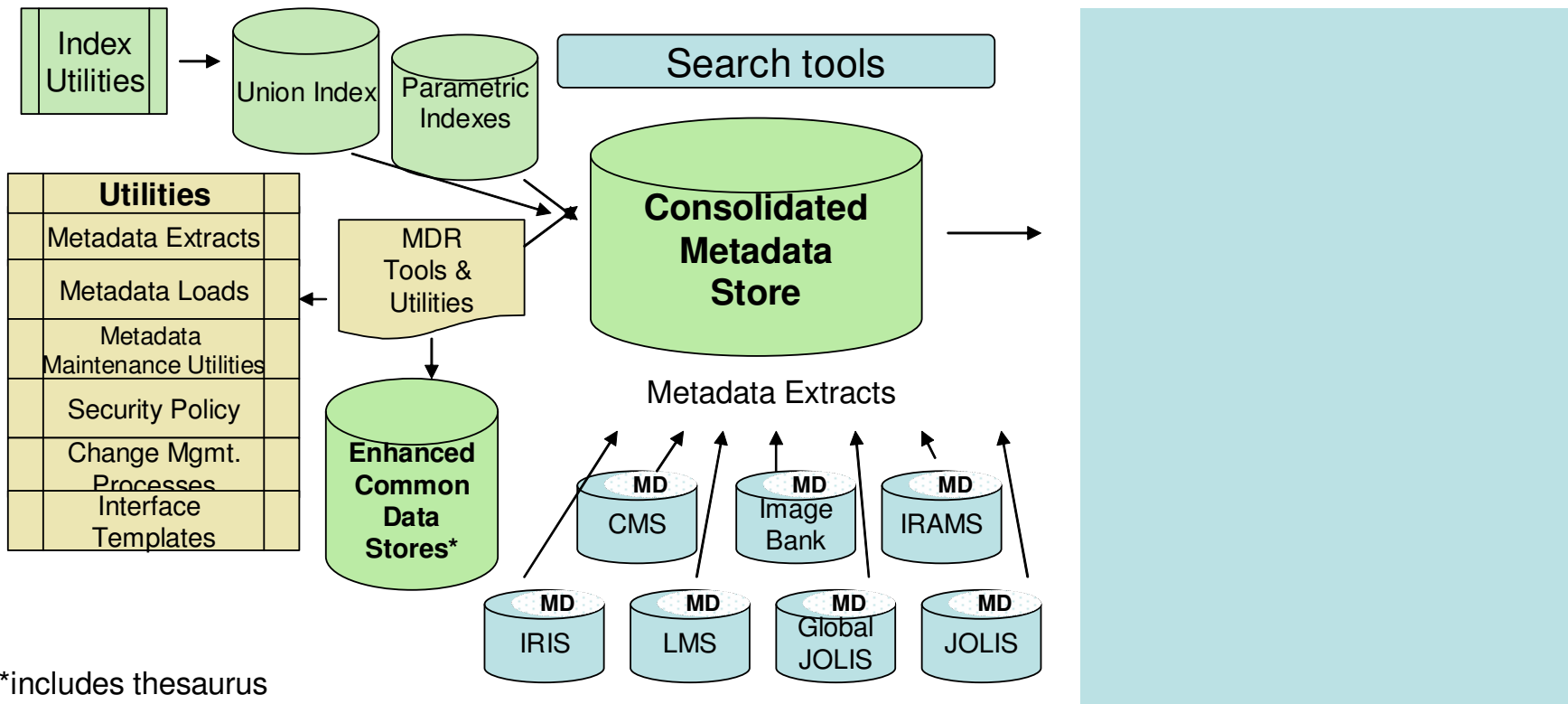
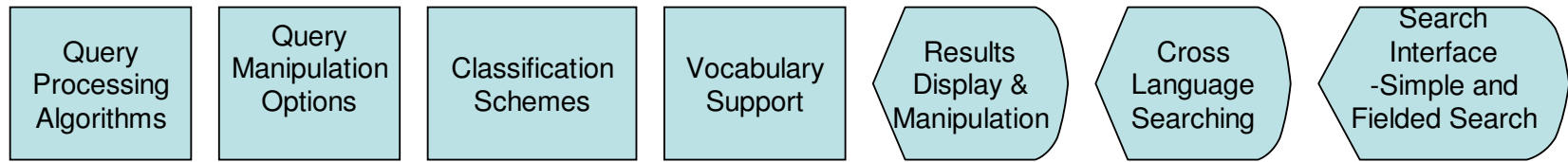
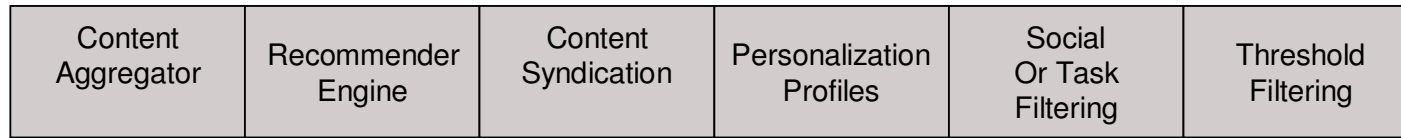
Denise Bedford, World Bank

Presentation to Managing Government Information
February 27, 2006

Vision of a Future Information Architecture

- Ten years ago we had a vision of what a future information architecture would look like
- Over the past ten years we have worked at realizing that vision. The vision involved:
 1. rationalizing and streamlining our information architecture in order to achieve economies of scale and scope in the delivery of information products and services
 2. Developing an information architecture which relied less on point-to-point integration and more on a middleware-based metadata and search architecture
 3. Implementing a consistent and enterprise wide approach to managing information
 4. Implementing a programmatic approach to capturing metadata and creating/maintaining our enterprise master data/reference stores
 5. Deploying an extensible and adaptable architecture to support sharing and transfer of knowledge, creation and collaboration to create new knowledge
 6. Developing the capability to support true cross-language access for our clients and staff

Information Architecture Blueprint



*includes thesaurus support and taxonomies

Teragram Metadata Capture

Rationalized & Streamlined Information Architecture

- Rationalizing the architecture depends on an agreement between the IT and the IM groups – this is not easy to achieve – programmatic approach involves...
 - Defining streamlined and rationalized approaches to content management, content publishing, and syndication
 - changing the way that development is done – from originally coding every new application to leveraging models, metadata and configurable components
 - Developing data models of all of the entities that make up our IA – content types, entity types
 - Moving to a Service Oriented Architecture from a 'tangled' information architecture
- This is not always easy because there are 'tradition-bound' IT folks just like there are 'tradition-bound' IM folks
- The 'tradition-bound' IT folks are the code writers – not the network and hardware people
- The IM folks are adapting more rapidly to new IA than are the IT folks

Middleware Architecture – Metadata & Search

- Business function taxonomy (using concept extraction, categorization tools)
- Topic taxonomy (concept extraction, categorization tools)
- Identify management architecture which respects security classification but supports customized provisioning of information
- Centralized approach to attribute management - approval, creation, provisioning
- Adaptable and dynamic reference sources which support staff and client participation
- TRUE Enterprise search functional requirements specification and delivery including a true advanced search capability, the ability to sort and narrow searching, synonym expansion and related term recommendations
- Using search to drive contextualization – moving from anonymous search to group-sensitive contexts to mass customization at the person level

Enterprise-Wide IM Approach

- We didn't have a common definition of what we meant by Information Management
- Information management has been supported in different units, following different processes and using different standards
- Developing an enterprise level IM program – in progress now
- Involved first addressing it on a virtual level before a physical level – the virtual level integration will make the physical level integration possible
- Enterprise information management starts with processes and policies before it can be effective at the workflow or practical level
- We need to bring together and synchronize processes and policies from multiple business systems

Programmatic Metadata and Master Data

- Began building a metadata repository to support integration and harmonization of information across systems vs. relying on labor-intensive point-to-point solutions
- Built enterprise level metadata profiles that could be used by any information unit or any end user (Records Management profile. Document management profile, library record profiles)
- Enabled our staff and clients to generate and manage their own metadata on demand while ensuring that it is compliant with quality standards set Bank-wide
- Use concept extraction tools to generate controlled vocabularies

Programmatic Metadata and Master Data

- Use clustering tools to identify first cut relationships for thesaurus and semantic relationships
- Used pattern matching tools to identify some key values – project ID, Loan #, Credit #, Trust Fund #, ISBN, ISSN, LCCN
- Creating 'subscription services' for individual systems to leverage metadata and reference sources
- "Re-indexed" and "re-cataloged" 1.3+ million historical and current records to provide better access in less than 30 days

Multilingual Architecture

- Designed an integrated and rationalized multilingual architecture to support clients and staff access to and use of information
 - Used semantic analysis tools to develop parallel language profiles to capture metadata and develop reference sources
 - Without the underlying semantic engines, none of what I've described would have been possible
 - Language detection software to ensure that the language values in the existing profiles were accurate – not always the case
 - Also involved developing data models to efficiently manage the equivalent relationships
 - Used concept extraction tools to extract parallel concepts for language experts to certify
 - Generate metadata in the language of the document – rather than trying to do suboptimal language translation on the fly

Realizing the Vision

- I do not want to downplay the challenges we have encountered and met, but we have had some exciting successes and we have learned a lot
- The successes have had a significant impacts on our working environment – impacts which we did not anticipate
- The changes we are experiencing are, I believe, a small steps along a path to a very different future – it is not just moving from manual or physical to digital
- The graphic depictions of the future architecture you see at conferences or presentations on the web don't capture some of these subtle environmental changes
- Today, I'd like to share with you some of our experiences with the goal of starting a conversation on the future impact to our profession and how we will handle it

Realizing the Vision

- In general, I would say we have been able to make greater progress than we expected because
 - we put smart tools in the hands of information professionals
 - We let information professionals guide the application of the tools to support good IM practice
- There have been expected and unexpected consequences
- I believe that both the expected and unexpected consequences are a direct result of the fact that we put the tools in the hands of the Information Professionals and did not simply leave them to the Information Technology professionals to install and deploy

Expected Impacts

- Information Access impacts
 - Increased precision of search
 - Better control over recall
 - Searching like we talk
 - Exact match searching – known item searching will work better
 - Metadata based searching now begins to resemble full-text searching but with all the advantages of structure & context, and a significant reduction in the amount of noise
- Productivity Improvements
 - Can now assign deep metadata to all kinds of content
 - Anyone can generate their own metadata and be assured that it is quality controlled
 - Remove the human review aspect from the metadata capture
 - Reduce unit times where human review is still used
- Information Quality impacts
 - All metadata carries the information architecture with it
 - Apply quality metrics at the metadata level to eliminate need to build 'fuzzy search architectures' – these rarely scale or improve in performance
 - Use the technologies to identify and fix problems with our data

Unexpected Consequences

- Unexpected consequences have been largely qualitative and environmental in nature
- Using technology in this way actually brought about significant changes in the working environment which were critical to moving forward
- Using the technologies in this way has had six types of impacts on our working environment
 - Changed the way we work
 - Changed our relationships with clients
 - Changed who skills needed to do the work
 - Changed how we manage information
 - Changed how we think about risks
 - Changed the scale and scope of our services and deliverables

Changed the Way We Work

- Changed...
 - How we index (now entirely machine-generated but also quality controlled)
 - Where we categorize (we didn't have resources to do it..)
 - When and how we catalog (it wasn't being done or it was very 'thin')
 - controlled vocabulary development – much richer and faster
 - Many institutional profiles into one institutional profile for all kinds of content
 - The way that we add new metadata attributes and provision attributes across profiles
 - Where and when we can do metadata quality audits and cleanup
- Who works...
 - Systems trained to think like indexers and catalogers
 - Staff with IM backgrounds
 - Staff without IM backgrounds
 - Clients who have begun to understand the value of the IM work we do and are now open to it because it is not burdensome

Changed the Way We Work

- Technologies have changed our process –
 - Whereas one person might have developed a single section in the thesaurus, we can now break up the work to resemble a 'mentofactoring' line
 - Some people do the training sets, some the concept extraction and vocabulary development, others do the testing
 - We can break up the work suitable to skill sets
- Tools have changed our relationships with our IT colleagues
 - There are traditional views in IT just as in the IM world
 - The new technologies, though, create an interest in IM among IT staff
 - This provides for common interests and knowledge sharing

Example of Change in How We Work

- Moving towards automated metadata generation means that catalogers shift their effort to reviewing the metadata generated and to more fully developing and maintaining subject headings/thesauri and classification schemes as part of a suite of categorization tools
- Level of effort shifts to training and developing the tools and away from original cataloging and metadata capture
- Continue to work closely with subject experts to define the controlled vocabularies and classification schemes
- It means that you have to have a metadata infrastructure that looks something like that ontology we just reviewed
- There is no silver bullet ontology tool out there that will do this work for you – your knowledge and skills are critical

Changed Our Relationships With Clients

- We now work more closely with our clients
- Our clients engage us with new opportunities for the tools, and we engage them earlier in the development process
- We can now realistically and practically involve them in feedback as we are creating
- This approach reduces the risk of failure – we adjust and refine as we develop
- This is actually an organizational culture change when IM, IT and clients participate in development on a regular basis
- We begin to develop common mental models and to see products and services from our client's perspective
- We are no longer an 'invisible' or 'indirect' service point – we are a visible service point
- Whereas previously the IT staff may have done all development, now their role is shifting to providing the hardware and infrastructure – more of the development work is being done by the IM staff

Changes The Skill Requirements

- Skills we need are now problem-solving and creativity, with a good grounding on principles
- Change from a single-loop learning to a double-loop learning environment
- More than just applying and interpreting the rules – making sure that the principles are honored
 - Bibliographic
 - Classification
 - Indexing
 - Retrieval
 - Records management
 - Authority control
- Really begins to resemble an IM environment, rather than fragmented units or departments performing targeted functions

Changes The Skill Requirements

- Everyone needs to have some comfort level with the technologies
- Everyone needs to have the ability to think outside the box – discovering the most effective way to use the tools
- Staff need to have a range of experiences but most important is that they need to be a 'user of information'
- Because we cannot hire these skills directly from graduate schools, we do a lot of one-on-one training and learning
- Becoming a knowledge sharing and learning environment
- This also means there is a lot more communication about what we're doing and how we're working – communication with middle and senior managers, with clients, with colleagues

Changes The Skill Requirements

- Because staff now perform multiple steps in a process, we have better backup and support across the team
- Through on the job training and knowledge transfer, we're achieving a consistent level of expertise
- Need to have a particular kind of technology, though, for the staff to work with – not just any technologies will do the job
- We need to teach technologies how to capture the metadata and master data that will be needed for this extreme semantic future
- We need to teach students how to maintain them and the new governance models

Effective Type and Use of Technologies

- You need to have tools which are “manageable” and “usable” by information professionals
- Many of the tools are so tightly integrated, you might generate rich metadata, but it will not make your information agile or mobile
- Statistical clustering engines do not get us to persistent meaning or contextualization.
- Clustering engines are great for thresholding or pattern tracings, but they will not generate the kind of metadata we need to realize this future
- We need semantic engines at the base of all our metadata efforts, and these engines need to be available in multiple languages -- semantics vary by language
- Magic black box approaches are neither meaningful nor sustainable -- you need to have access to the programs through a user-friendly interface so you can adapt them to your environment without having to have programming knowledge
- You need to have several different kinds of technologies to do what I’m going to describe today – not just one tool

Overview of Process & Tools

Activity	Approach	Tools
Create new metadata attribute	Human review & consultation with clients, data structures development, governance model definition	Oracle DBMS, in future Metadata Repository tools (ISO 11179); Oracle representation of data classes
Create new class in a class scheme	Human review & harmonization of existing information structures; tool based discovery of new structures through clustering & extraction	Teragram dynamic concept extraction using grammars, categorization, clustering; Oracle representation of data classes
Create new concepts for indexing	Create training sets working with experts, identify & integrate existing vocabularies	Teragram concept extraction, Oracle representation of values
Create new relationship	Human relationship creation, augmented by technological discovery	Teragram clustering engine, MultiTes Thesaurus Management System, Oracle copy of thesaurus relationships
Create new metadata standards and profile	Enterprise Profile Development with human review in some cases, no review in others; Metadata in the language of the document/content	Teragram enterprise profile leveraging concept extraction, categorization, and summarization

Changed How We Manage Information

- By manage information, I don't mean the typical information management life cycle functions. Rather, we find we now need to manage Information Management processes from a Knowledge Management perspective
- Experienced changes in four areas:
 - Governance
 - what we need to govern, how we govern and who governs
 - Best practices and lessons learned
 - New role for them, where we find them, who needs to know them
 - Change management and service requests
 - IA is now a dynamic and living system – change requests are constant - need to proactively manage change
 - Metrics and benchmarking
 - Our actions now have immediate consequences so we need to predict and assess them before we act

Governance

- What is governed has serious implications for access, use, efficiency and extensibility - Must be managed consistent with the level of impact
- Participation in governance is now key because using the technologies simply brings more people into the management process
- More people are participating in the process so a common and public set of criteria is needed
 - We can no longer retain the rules in our head – they need to be explicit and understandable
 - Governance groups must involve stakeholders not just the IM and IT staff
- More governance groups are needed – each attribute has a governance model & group
- Participation in governance becomes possible because the technologies free up our time to work with clients and peers
- Encourage stakeholder participation because we can make it easy for them
- Governance is essential to success because it is what ensure consistency in development and deployment across applications

Best Practices

- Always important to anchor in industry standards where they exist
- Any deviations from industry standards should be explained and understood by everyone
- Participation in standards work is key to keeping them relevant and practical
- Best practices and lessons learned are part of a continuous improvement environment
 - Manage IM with a KM process model (Mark McElroy)
 - Begins to look like a double-loop learning (Peter Senge) environment
 - Do not give greater weight to 'past practices' just because they exist
 - You have to have some context for judging what is 'best'
 - Risk-prefering behavior to share lessons learned
- Remain current on what others are doing and have learned
- Document and publish both BPs and LL's
- Continuous education and training now key because we can do more and imagine more having implemented the technologies as OUR tools
- Everyone now wants to understand how it all works – clients, stakeholders, IT staff, IM staff

Change Management & Service Requests

- All components of the IA are now alive and dynamic – JIT services, On Demand Services
- Change is not just possible it is probable so change must be managed
- IM Store Fronts - IM units are now client service points and service providers - no longer invisible backroom operations
- In the 1980's we expected catalogers to move to the front reference desk – now we need those catalogers and indexers back in technical operations but technical operations become direct service points for clients
- Clients request services directly - clients understand metadata and reference sources and they want their own – IM products are now services
- Service delivery rates, agreements and quality standards are now needed - expectations for Turnaround Time and Quality are now much higher
- All of these aspects seem to be propelling the demands for innovative use of the tools by IM professionals

Metrics & Benchmarking

- Because of the new Information Architecture, the impacts of change are immediate and more significant
- It is now possible to have an immediate impact – so we need to be able to accurately predict what the impact will be
- The IM community (stakeholders, IM professionals) understands more about cause/effect and will be more aware of when a change is made
- More cautious and complicated process for investigating and mitigating the impact of change is needed
- Impacts are not just quantitative – now they are qualitative – need metrics for both
- More opportunities to apply the measurement theory in practical settings (search metrics)
- We have been using benchmarking exercises – gauging the true 'pre-impact' state so we can measure the degree of change

Metrics & Benchmarking

- Don't underestimate the value of benchmarking - benchmarking exercises have been very valuable
- You don't know how much you don't know until you benchmark - there is not one benchmarking exercises that hasn't lead to at least one 'Ah-ha!'
- Benchmarking for a search impact brought about by richer metadata has helped us to identify problems with the underlying search architectures in several systems
- Benchmarking helps us to understand on a deeper level what our clients are experiencing
- Use case methodology is also very important for understanding client experiences
- 'Post-impact' metrics give us the management framework we need to achieve future improvements

Changed How We Think About Risks

- Our environment and our role in it is changing
- When we provided services and products which were fringe or peripheral, risks of change and innovation were low and less visible
- Now the products and services we provide have a higher impact, are critical to processes and are more visible
- Need to design for risk aversion – because we're not 'critical path' we need to make sure that our solution are sustainable – the tools, the people, and the standards
- Risk mitigation factors include:
 - Open architectures
 - Interchangeable skill sets
 - Adaptability of components rather than continuous redesign and renovation
 - Manage variations at the core not at the fringe
 - Think 'enterprise' as much as possible – enterprise profiles
 - Think 'master data' to manage quality
 - Manage once, provision often

Enterprise Profile Creation and Maintenance



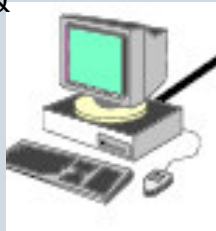
Update & Change Requests

UCM Service Requests

Data Governance Process for
Topics, Business Function,
Country, Region, Keywords,
People, Organizations, Project ID

e-CDS Reference Sources for
Country, Region, Topics
Business Function, Keywords,
Project ID, People, Organization

Enterprise Profile Development & Maintenance



TK240 Client



Teragram Team

Enterprise Metadata Profile

Concept Extraction Technology

- ✓Country
- ✓Organization Name
- ✓People Name
- ✓Series Name/Collection Title
- ✓Author/Creator
- ✓Title
- ✓Publisher
- ✓Standard Statistical Variable
- ✓Version/Edition

Categorization Technology

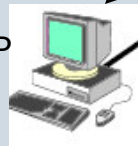
- ✓Topic Categorization
- ✓Business Function Categorization
- ✓Region Categorization
- ✓Sector Categorization
- ✓Theme Categorization

Rule-Based Capture

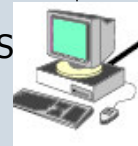
- ✓Project ID
- ✓Trust Fund #
- ✓Loan #
- ✓Credit #
- ✓Series #
- ✓Publication Date
- ✓Language

Summarization

ISP



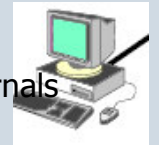
IRIS



ImageBank

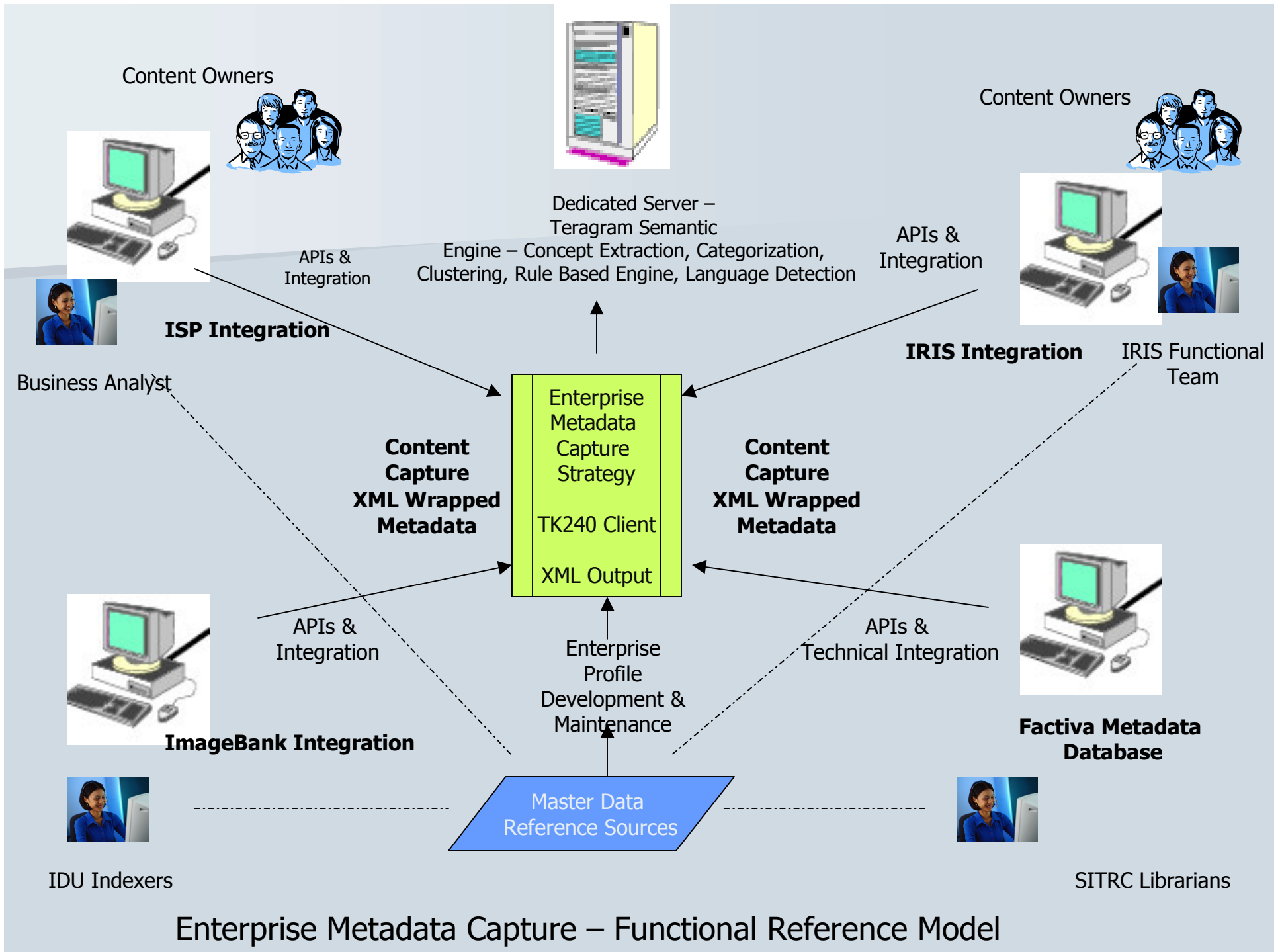


JOLIS
E-Journals



Factiva





Enterprise Metadata Capture – Functional Reference Model

Changed Scale & Scope of Services

- This is where there is a potential for explosive growth
- Let's first talk about the impact to **scale of services**
- It is now possible to extend our provision of support and services to more "I" than previously – it is now possible for us to cover magnitudes more "I" objects than previously
 - more content types
 - Greater numbers with existing staff resources
 - In multiple languages
 - Increased coverage in terms of richer metadata access points
 - JEL codes
 - Statistical indicators
 - People and Institutions References
- May be possible for us to provide services to other development organizations
- We can now provide deeper coverage of selected domains

Changed Scale & Scope of Services

- Now let's first talk about the impact to **scope of services**
- Office of Publisher and Country Office libraries have asked us to develop a single MARC profile for cataloging books and journal articles
- Operations Evaluation Department now uses our programmatic categorization metadata to identify projects for quality evaluation audits
- Business Function taxonomy is now used to data drive information from information repositories to web sites
- End users can self-generate metadata for their own resources
- Clients can self-generate metadata to support search and browse in CD products for clients with lower levels of connectivity
- Information management professionals can use the tools to generate deep profiles for individual clients – on their request and working closely with them

Extending Scope and Information Agility

- Most of our information use and access today is based on an anonymous access model
- Information is in a single place and people and systems go to that place to get it
- It is increasingly clear that anonymous access to information and the packaging of information for single use contexts is neither sufficient for users nor an efficient use of development/engineering resources
- We need to think in terms of contextualization and sensitization of information so that it can be used in any context where it pertains
- In the future, information will flow – information, not the systems in which it lives or was created, will be our focus
- Information needs to be agile and mobile – it needs to be sensitized to the contexts in which it might be used, to the interests of those who might use it, and to the applications that might consume it

Preparing for the Future

- We have only taken baby steps towards the the Meta-World but it is clear that any movement along this path will bring major changes to what we do and how we do it
- There are also significant implications for our profession and professional education
- We need to:
 - teach principles and applications not just rules – principles are adaptable and sustainable
 - teach more technology principles (and maybe in some cases to invent them) – this is missing from the IT professional education and is increasingly marginalized in education in the developing world
 - be more actively involved in standards development because this is where products come from
 - influence vendors and their products

Preparing for the Future

- We need to:
 - Develop a formal core IM curriculum focused more broadly than the library science curriculum or the records management curriculum
 - The IM curriculum needs to cover all of the IM processes
 - Define IM core competencies
 - Follow Canada's lead in implementing IMM processes and models
 - be open to teaching IM to everyone who will listen and wants to learn – it is no longer a sacred creed that we keep to ourselves
 - Create more support groups and sharing within and across "IM" professions and associations
 - Our professional organizations are good places for this but they represent silos within our professional practice – we need to bridge the silos

- Most importantly, we must be able to think outside the box and on our feet all the time – always referencing good professional principles but adapting it to current processes and shaping the technologies to use it

Thank You.

Questions & Discussions