

Euromap workshop
4 September 2002, London

transforming results
PWC CONSULTING

Knowledge Management
and the opportunities for
Language Technologies
such as Text Mining



Objectives this morning

- What is the business case for language technologies?
- How do they fit - the Knowledge Management context
- Examples of implementing semantically intelligent systems in the business environment – machine translation and taxonomies

Today, serious problems of information overload affect all knowledge-driven organisations. In the enterprise, employee time is wasted in inefficient search; and problems of navigation prevent adoption of knowledge management systems by users.

Existing search and knowledge management approaches have not solved these problems, and in many organisations information and knowledge sharing systems, such as B2E portals and intranets lie underused or unused.

As we move further into the knowledge economy era, we rely more on the effective use and exploitation of intellectual capital to gain competitive advantage. Increasingly organisations wish to place some basic structures behind the information that people need.

Structuring access to information not only prevents overload but eases the identification of information that may prove critical to the competitive position of an organisation in the marketplace.

Knowledge lifecycle



User accesses company information resources through Portal...

... that allows user to utilize hybrid searching and browsing tools...

...to find desired documents or resources

... that has undergone QA by Knowledge Managers throughout the organisation

... and have been created by staff that are actively recognized and evaluated on the basis of knowledge sharing

Information User



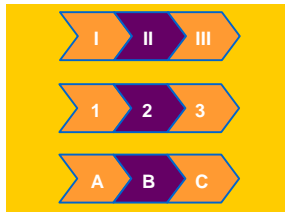
Personalized Intranet Portals



Controlled Metadata Vocabularies



Parallel Taxonomies



Thesaural Browsers



Corporate Knowledge Database



Required Document



Corporate Knowledge Management & Governance, Information Architecture & Technology



Content - Validation Protocols (QA)



Information Expiration Mechanism



Performance Reviews that includes assessment of Knowledge Sharing coupled with Change Management Programs



Embedded Authorial Recognition



Information Author



Access

Contribution

KM Model



KM Model

Methods:

- Workflow design for capture
- Action in Review; Peer Assist
- Tacit Knowledge SWAT Team
- Community of Practice

Tools:

- Collaboration tools (e.g. eRoom)
- Content Management (e.g. Vignette)
- Expert Finder (e.g. TacitMail)

Methods:

- Community of Practice
- Workflow design with explicit knowledge-creation
- Feedback on knowledge quality

Tools:

- Portal workflow incorporating knowledge
- Co-authoring tool (eg. eRoom)
- Topic visualization tools (e.g. TheBrain)



Methods:

- Knowledge Mapping
- Best practice synthesis
- Content approval, QA, legal workflow
- Taxonomy & Directory Services Workflows

Tools:

- Taxonomy (e.g. Semio)
- Meta-Data Tagging (e.g. Vignette)
- Security / Access (e.g. Netegrity)
- LDAP Directory

Methods:

- Data placement rules
- Repository design
- Refresh/Archive rules

Tools:

- Data Marts
- Content repositories, refresh/archive services
- Back-up / Recovery

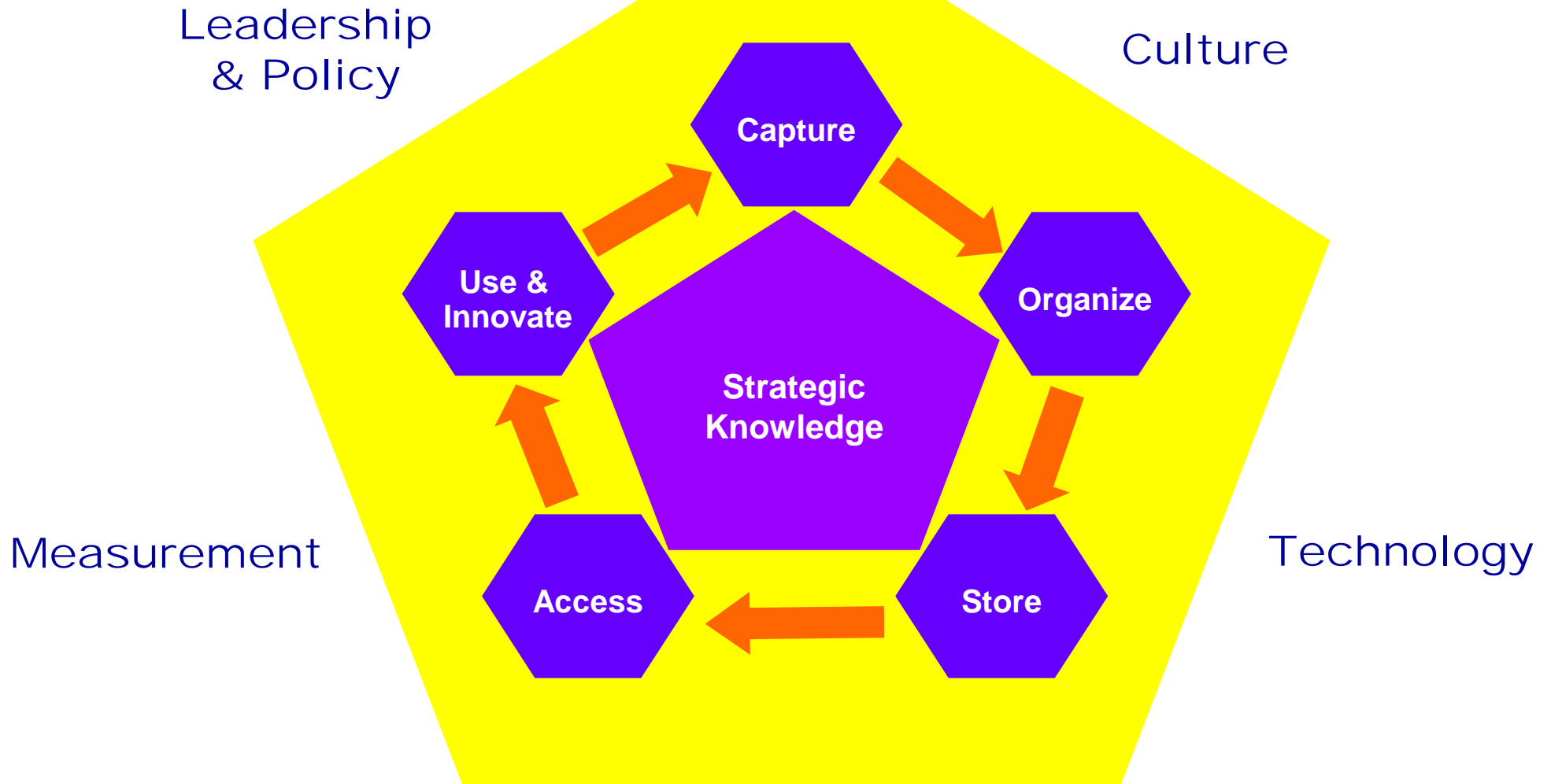
Methods:

- Workflow design with Access / Reuse
- Marketing of content
- Knowledge Helpdesk
- Community of Practice

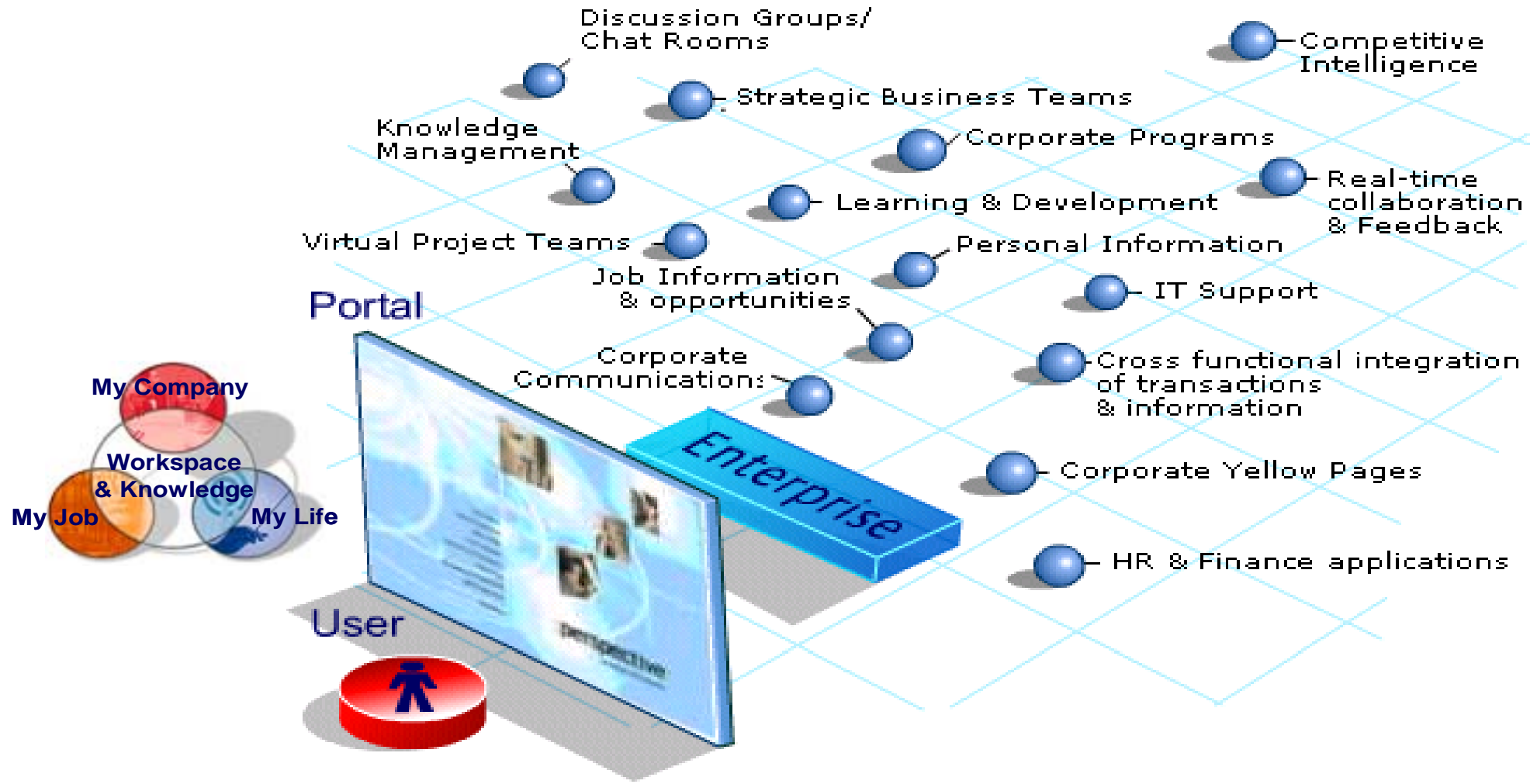
Tools:

- Portal Personalization / Push
- Search (e.g. Verity)
- Taxonomy (e.g. Semio)
- Autnotification (e.g. eRoom)

KM Holistic Enablers



Knowledge management can be embedded within a content-rich portal that incorporates management, integrated analytics, employee self-help and personal workspace functionalities



KM tools - example - taxonomies

[Library of Congress Classification](#)

www.meansbusiness.com

Bitpipe.com

Yahoo.com

Typically, what are taxonomies used for?

Core applications

- **Tagging content objects in the content development process** (either human or automated process)
- **Site navigation** - indexes and categories for browsing in websites
- **Information retrieval** - tying taxonomy to search engine
- **Personalisation** - targeting users with content objects

Other applications

- Providing a consistent source of terms for knowledge application development
- Providing multilingual terminology and mapping (eg English<> German)

What do taxonomies cost?

- Amazon.com - over 50 cataloguers on staff, though this is to a high degree automated (classification software). Need for expert “editorial” skills in team.
- Amazon reckon that the cost of tagging, maintaining and delivering is \$4 per page - hence need to focus!
- Microsoft’s internal portal has three full time taxonomists on payroll
- PwC has a team of six specialists in corporate taxonomy development

Where is the added value?

- Good information reduces cost and enhances productivity
- 20 minutes fruitless searching per day means at least 3,000 EUR per year
- Improved decision-making through better information access
- Reduction in time-to-market
- Improved staff satisfaction means better social conditions and less stress in workplace (inefficiency of traditional, text based search)

A concrete example - Microsoft

- Microsoft's internal portal invested heavily in taxonomies for navigation and search
 - over 78,000 staff worldwide
 - 2.2 million documents on intranet
 - 62% reduction in number of clicks
 - average of 16 seconds saved per task
 - 11% increase in task success rate
 - 42% very satisfied with search against 4% dissatisfied

KM tools - example – Machine Translation at PwC

- Provides automated translation on the PwC intranet in partnership with SYSTRAN
- Advises our consultants on translation best practice
- Provides and supports a large network of PwC professional translators
- Supports client engagement teams with guidance on e-content and translation in projects
- Supports the PwC brand

Business case for Machine Translation

- As a “knowledge transfer” tool in situations when it is not feasible to translate, because of cost and / or speed eg:
 - foreign language web pages, documents, cv’s
 - work engagement reports, progress reports, proposals
- As a method of understanding / gisting documents from knowledge repositories following search
- As a method of exploiting our expert translator knowledge and experience for re-use in dynamic translation environment

Why SYSTRAN?

- Best results for readability and accuracy following test phase with our own materials - we defined testing environment
- Flexible approach - ability to customise linguistics resources to a deep level (eg financial texts) and expertise in integration of multilingual functionalities eg processing and exchanges for KM
- Architecture / ASP set-up suits our global requirements and allows for customisation at different levels
- Good match with global consulting group - joint interest in building language and KM competency



Systranet.com

Login

Welcome to the PwC/Systran translation site. Please register by clicking on the Sign-up now link below. If you have accessed the site from the PwC network/KnowledgeCurve, your GUID (Global User ID) should appear in the first box. Do not alter this. Please complete the rest of the form. In future, you will access the translation site directly.

If you have accessed the site via the Internet, you need to provide your external PwC email address. After submitting your registration, you will receive a confirmation E-mail including a password that you will be able to change later. Make sure that your E-mail address is correct. When you access the site in future, you need to enter your password.

New User?	
Sign up now! <i>It's free.</i>	

[Why Sign Up?](#) [Forgot my Password](#) [Help](#)

Registered user	
GUID or E-mail	<input type="text" value="jsage002"/>
Password	<input type="password"/>
<input type="button" value="Continue"/> <input type="button" value="Reset"/>	

[Back to Top](#)



File Edit View Favorites Tools Help

Address <http://web.systranet.com/systran/pwc?email=jsage002&code=genericpw>
 Links [%QL10%](#) [%QL20%](#) [%QL30%](#) [Best of the Web](#) [EMAIL THIS](#) [Free Hotmail](#) [Internet Explorer News](#) [Internet Start](#) [KnowledgeCurve](#) [Microsoft](#)


Welcome Jonathan SAGE

[Help](#) | [Scope and Limitations](#) | [Modify my Profile](#)
Logout

Translate Now!

To translate your documents, follow the three steps below.

1 What do you want to translate?

 A Web page *(Enter the URL)* **A file** *(Upload your file)**Only TXT, RTF and HTML file formats are supported.* **Plain Text** *(Just enter or paste your text)*

2 Languages and options

Language pair

English to German

Dictionaries*Select the topic that is related to the subject of your document.*

General

Translation result layout

Separated

My Terms

[How to use it?](#)

Learning points in applying MT in real business contexts

- Educating users and managing their expectations critical - invest in communications and feedback
- Improving and customising linguistic resources (eg terminology) in specific subject areas
- Closing the gap between human translators and MT technology so technology is not perceived as a threat
- (then) - exploiting resources created by human translators in the MT environment
- Setting goals and measuring success - clearly positioning MT as a knowledge access tool alongside Search

Current projects

- Integration with xml standard in domain specific translation eg xbrl for internal and client purposes (eg Transaccount project)
- Provision of customer and domain specific linguistic resources within content management (eg corporate portals) for clients
- Language and e-learning - providing multilingual learning objects for global organisations (eg European Commission CODEX-IP project)
- Determining good practice in eContent Localisation (eg European Commission EEEL project)

Defining and measuring value - same issues as in KM

- **Metrics** - defining baseline / measurements at outset
 - need to show success at early stage of project - quick wins?
- **Measuring usability / user behaviour** - examining what people need
- **Managing expectations** - communications on an organisational policy towards improving information retrieval / goals of taxonomy or language project / expected outcomes