Enterprise Search, Faceted Navigation and Subject-centric Portals

About the presenter

- Started as a researcher at a university

  - Artificial Intelligence, Semantic Memory, Open Systems, Knowledge Representation, Justification-based reasoning, Knowledge conflicts, Knowledge evolution.

- But has been always fascinated by solving "real business problems"

- Shift to "enterprise computing"

  - programmer analyst, system architect, enterprise architect

- But never stopped researching ...

  - distributed systems, messaging, "active/composite" frontends, data constraints monitoring, Topic Maps, portals, faceted search, robotics

Agenda

- "Enterprise computing"

  - Traditional application landscape in an enterprise

  - Some recent enhancements: SOA, ROA, Search Engines, Portals

- One step further: Enterprise Knowledge Map

  - Creating a Knowledge Map

  - Subject-centric portals

  - Faceted Search and Navigation infrastructure

  - Issues

- Next stop: Subject-centric computing

Enterprise Computing

- Is tightly coupled with the internal organization and business processes

- Should support agile nature of contemporary business

Contemporary business requires:

- "Active" knowledge workers

  - Search for new information

  - Information assimilation

  - Evaluation

  - Re-ordering, synthesis

  - Identifying opportunities

  - Measuring risks

  - Defining and re-evaluating preferences

Traditional Applications

Interface/
front-end
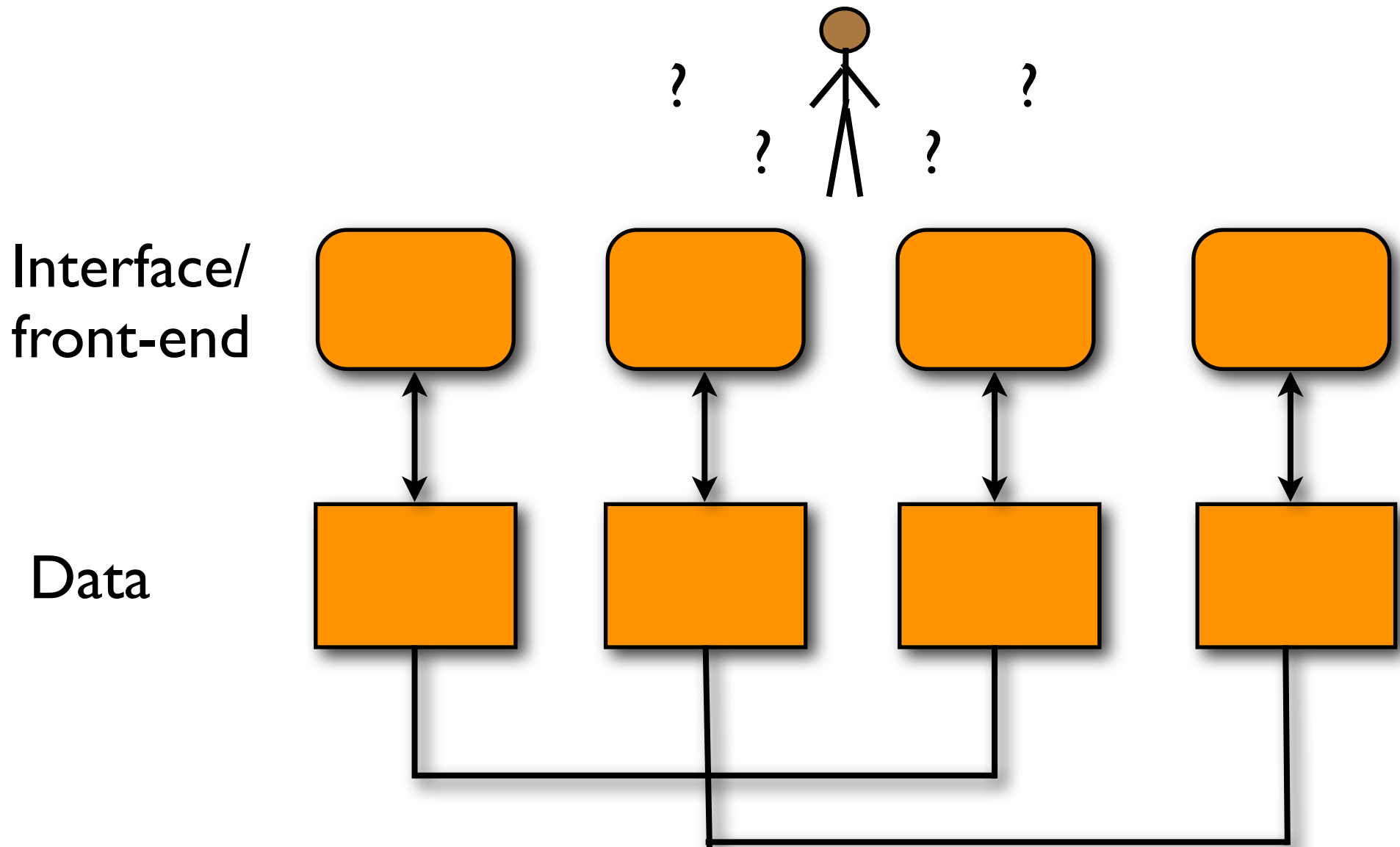
Data

Working with unstructured data
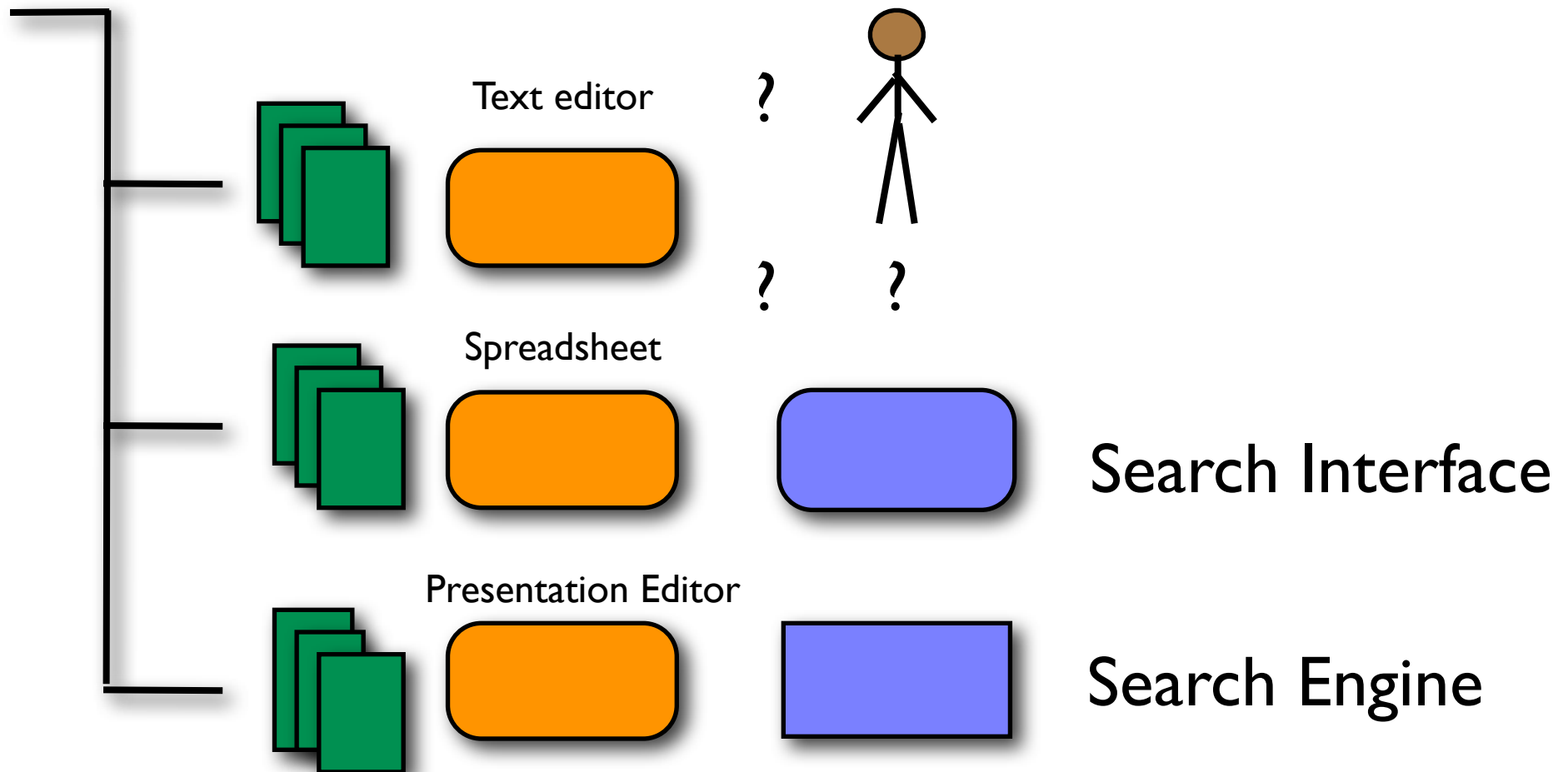
Summary of the traditional approach

- There is no holistic view on information resources in an enterprise

- Multiple applications are used for solving specific tasks

- It is not easy to relate information from different applications

- Structured information is not integrated with unstructured information

- In many cases we do not have enterprise-wide explicit representations of objects important for our business
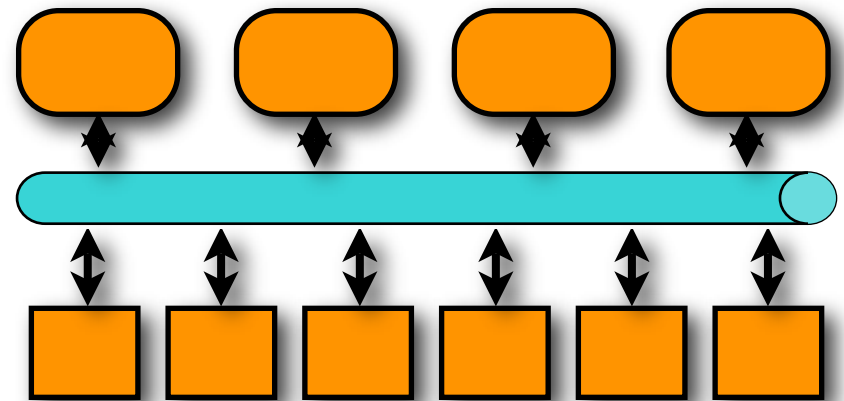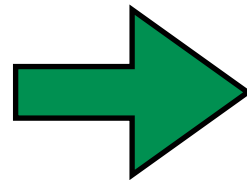
Recent enhancements

- Service Oriented Architecture (SOA)

- Resource Oriented Architecture (ROA & REST)

- Portals

- Enterprise Search

Recent development: SOA
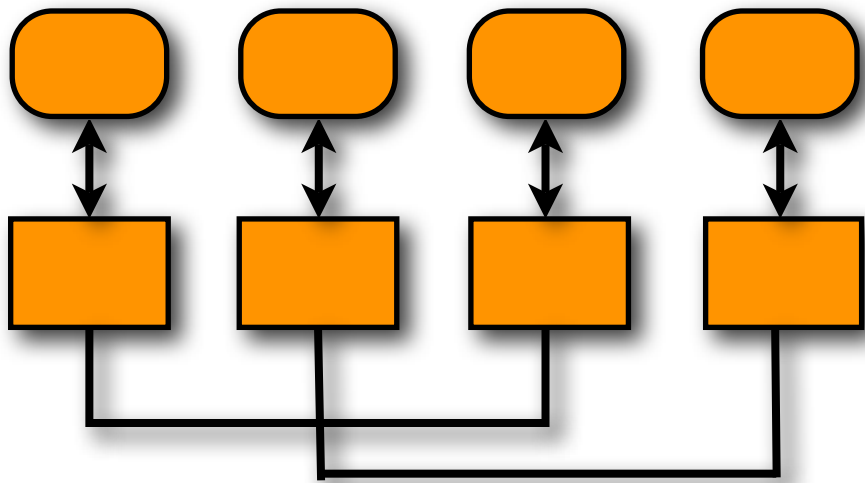
Application silos

Services, ESB,
Composite Applications

SOA

- It emphasizes the need for business architecture and understanding of business processes

- It helps to organize and optimize server side environment

- Modularity and mediation introduce some level of agility

- It is possible to achieve some level of re-usability

- Unified monitoring and governance

SOA?

- Does not help much for organizing and searching information

- Does not help much on the user interaction layer

  - May be a little bit more dynamic and user configurable interface

REST: some history

- 2000, Roy Fielding's Ph.D. dissertation

  - investigation of various architectural styles for building (networked) software

  - Chapter 5: defining REST

- REST - REpresentational State Transfer

- Clarification of the main principles of the Web

Key REST principles

- Give everything (important) a URI

- Serve one or more representations for "things" (using URI)

- Link "things" together (using URIs)

- Use uniform HTTP-based interface (POST, GET, PUT, DELETE)

  - kind of CRUD (Create, Read, Update, Delete)

- Communicate statelessly

Resource Oriented Architecture (ROA)

- Leonard Richardson

- Sam Ruby

- May 2007, First Edition

- Published by O'Reilly

- Coined the concept of the Resource Oriented Architecture

ROA?

- ROA shares the same understanding of a "Resource" as Semantic Web/RDF
  - no explicit distinction between "digital information resources" and "other things"

- Reality is ...
  - Subjects that are not information resources (people, companies, events, etc.) can have multiple representations in different "locations" on the Web
  - It is not just about a different form, representations can have different content
  - We need a robust mechanism to identify subjects and create 360° view based on merging information from multiple "locations"

Portals

- In many cases we just make windows smaller and present them on the same screen
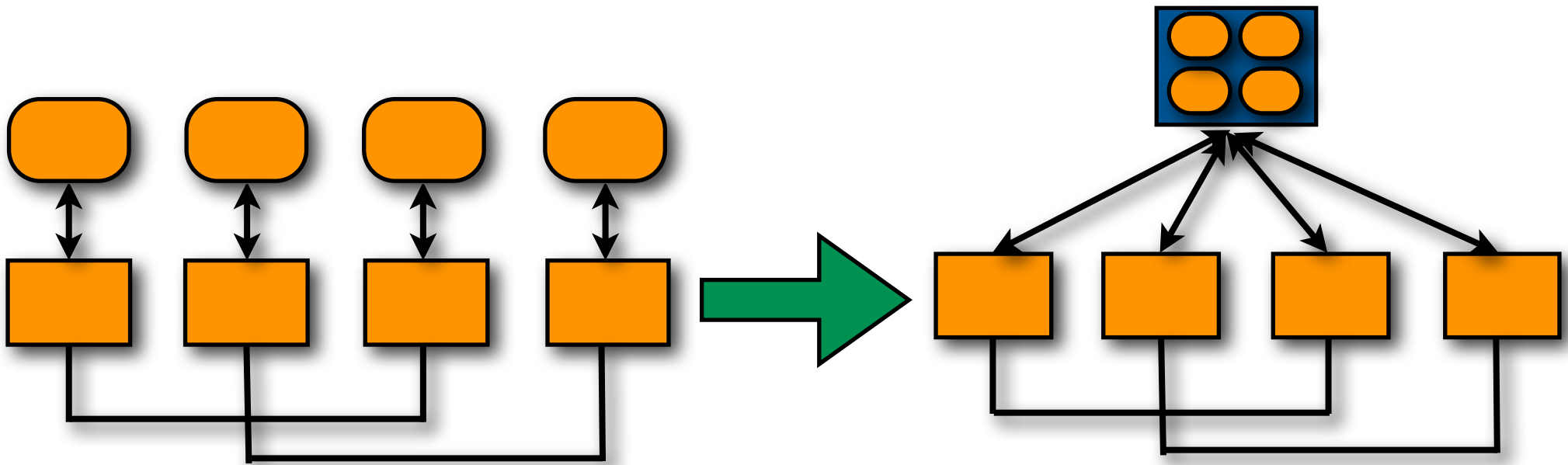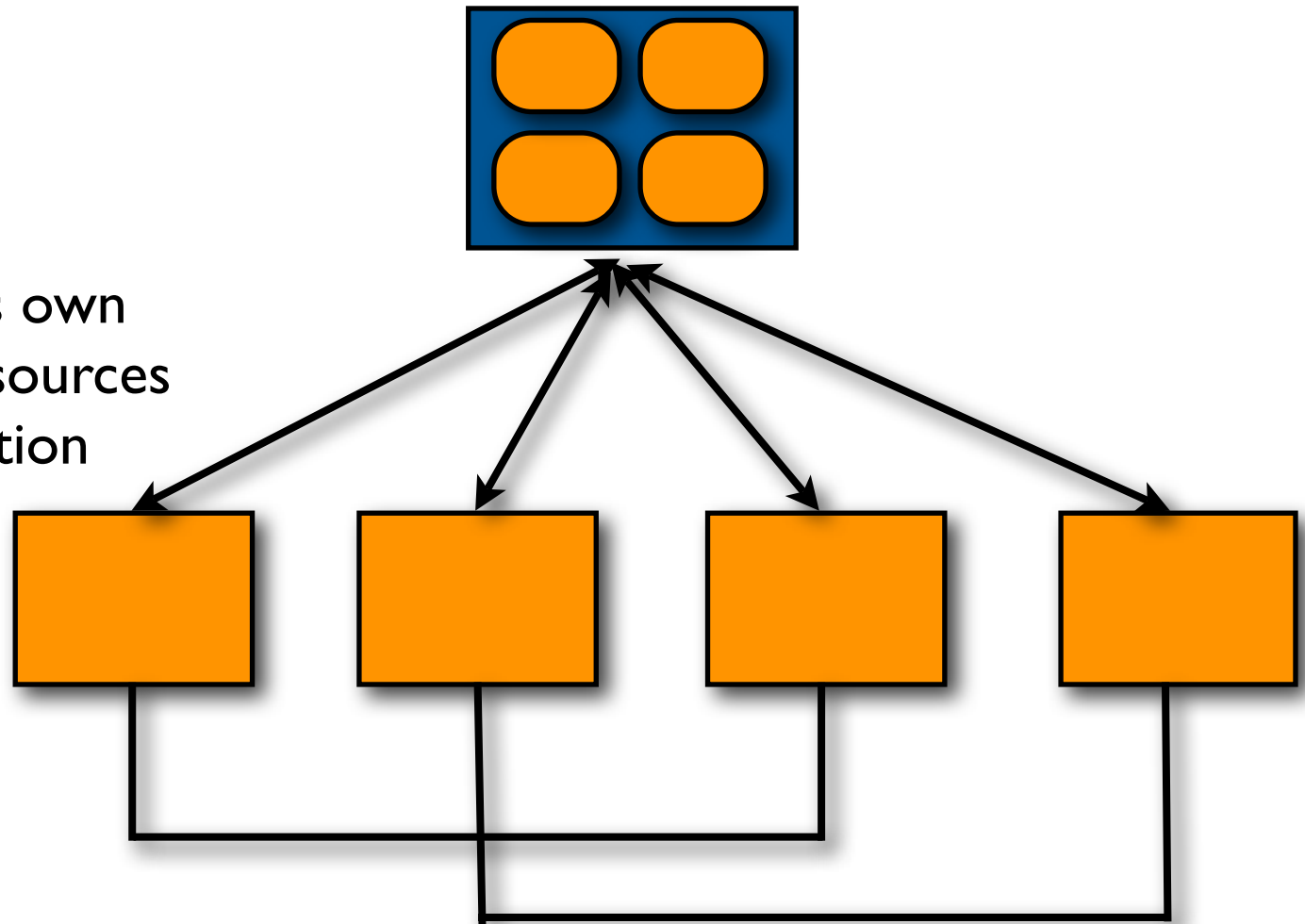
Portals?

- Traditional approach: integration "on the glass"

  - Direct connections to data sources

  - Each portlet defines own mapping from data sources to visual representation
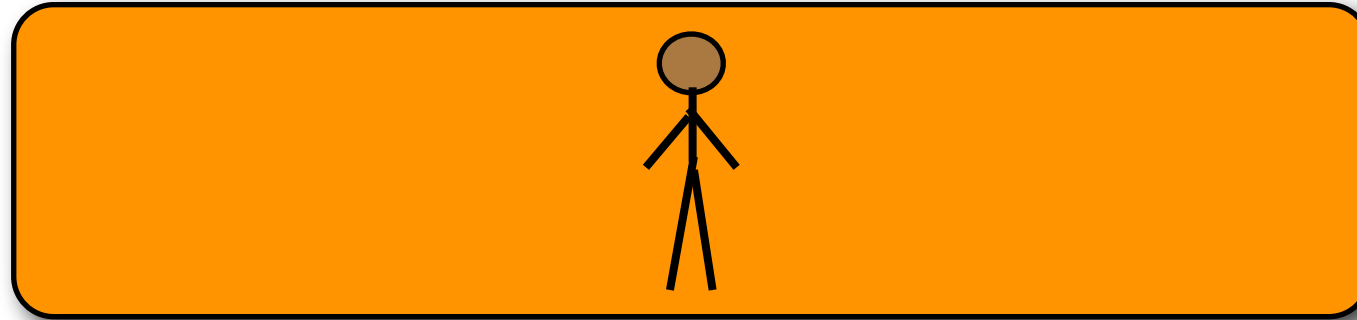
Enterprise Search

- Initially: full-text indexing of various documents

- Recently we see support for properties/metadata

- Extending searching from "documents" to various entities

- Some extras

  - Entity extraction, results clustering

Is it a right direction?

- SOA

- ROA

- Portals

- Enterprise search

- ...

- Something is missing !

Enterprise Architecture: adding knowledge layer

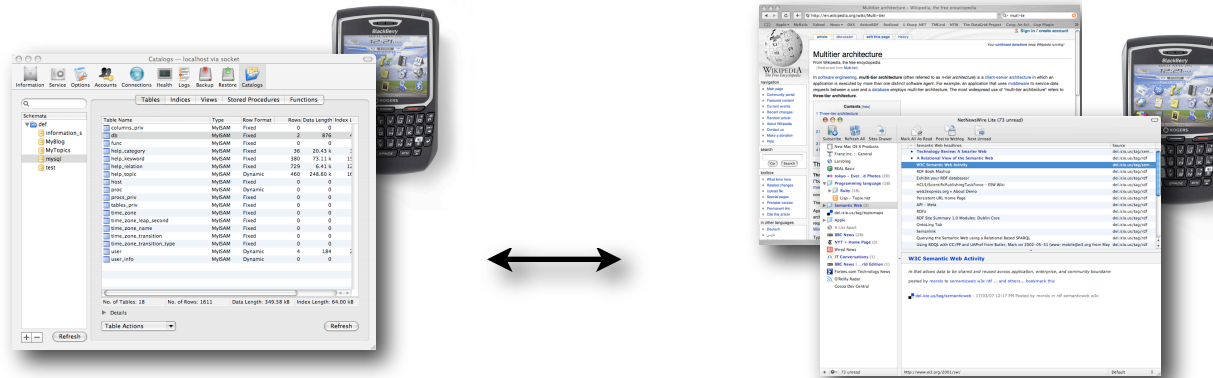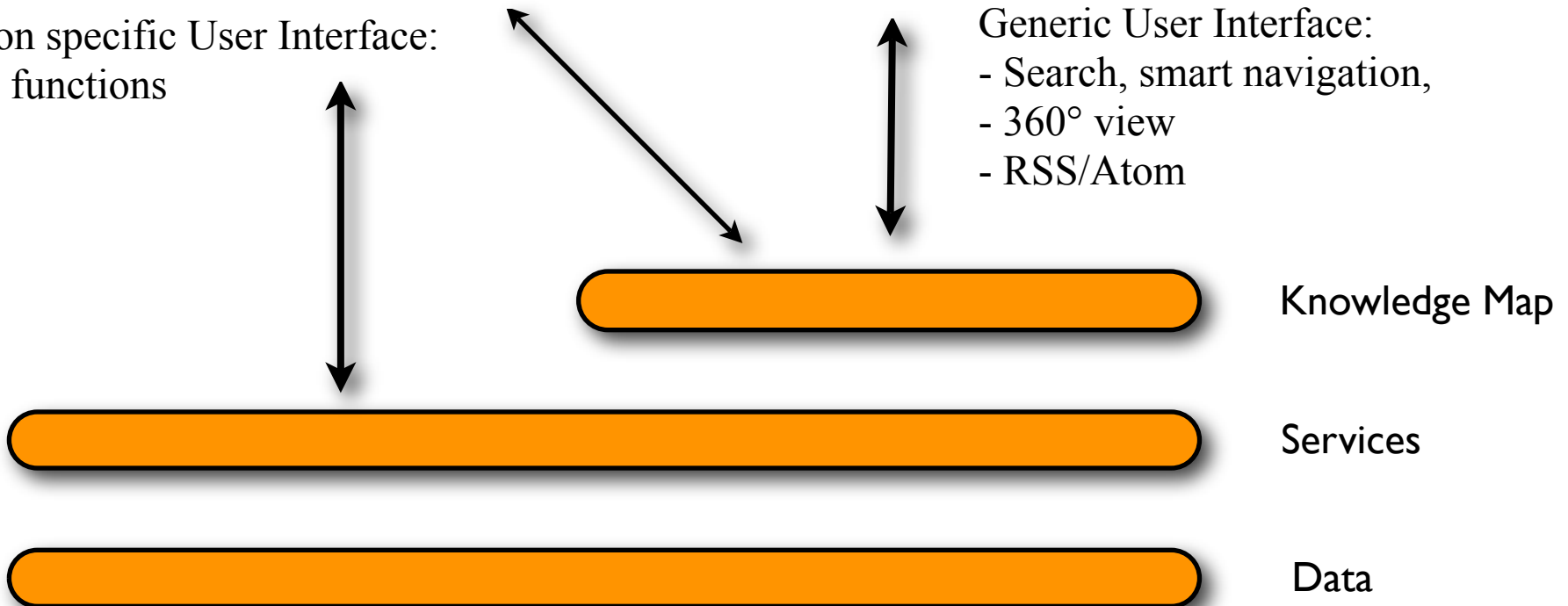Interaction Layer

Knowledge Map

Services

Data
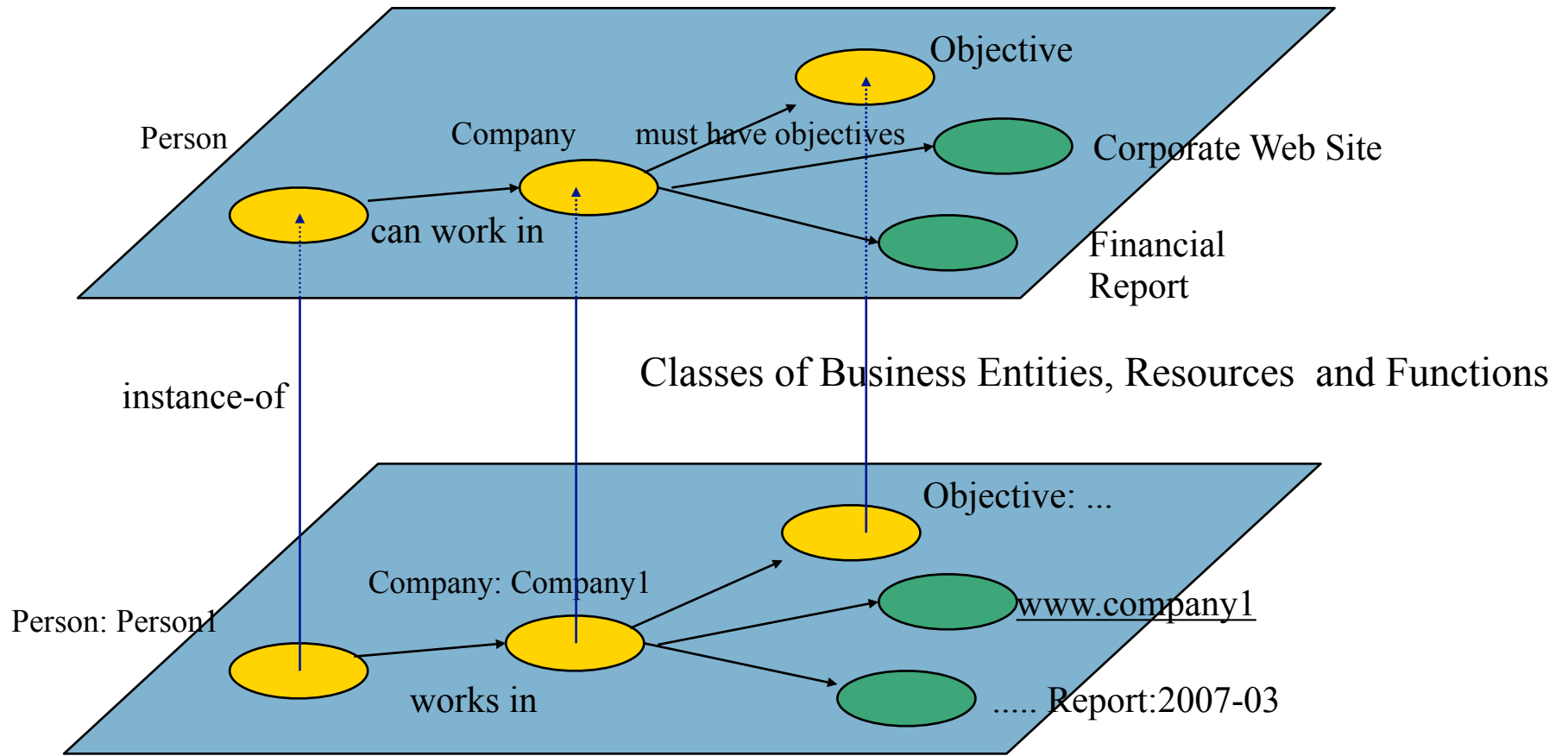
Task/Application specific vs. Generic User Interface

Application specific User Interface:
• Specific functions

Generic User Interface:
- Search, smart navigation,
- 360° view
- RSS/Atom

Knowledge Map

Services

Data

Knowledge Map

Person

Company   must have objectives   Objective

can work in   Corporate Web Site

Financial Report

Classes of Business Entities, Resources and Functions

instance-of

Person: Person1

Company: Company1   Objective: ...

works in   www.company1

..... Report:2007-03

Facts (summary)

How to create a knowledge map

- Identify

  - business functions

  - main entity types

  - main resource types

  - properties and associations for entities, resources and functions

- Map and integrate existing data sources

- Find and fill the gaps

- Use Topic Maps to implement Enterprise Knowledge Map!

How to integrate existing data into knowledge map

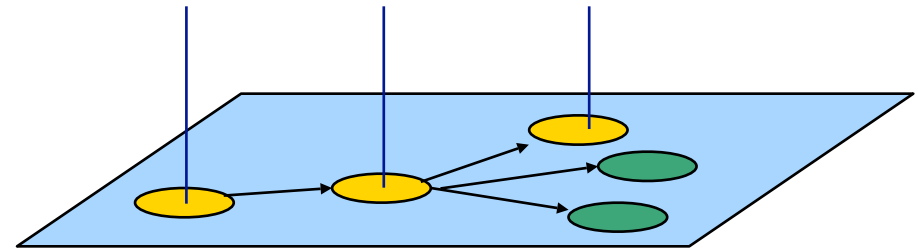Expose types of business entities and main relationships:
- Map existing models: UML, ER
- Use Topic Map ontology editor
- Re-use as much as possible existing definitions
- *Key attention: enterprise-wide identifiers*
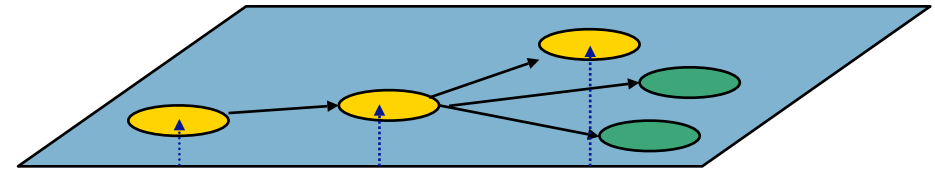
Types and Relationships (ontology)

Expose summary of existing data:
- Use mapping/exporting tools
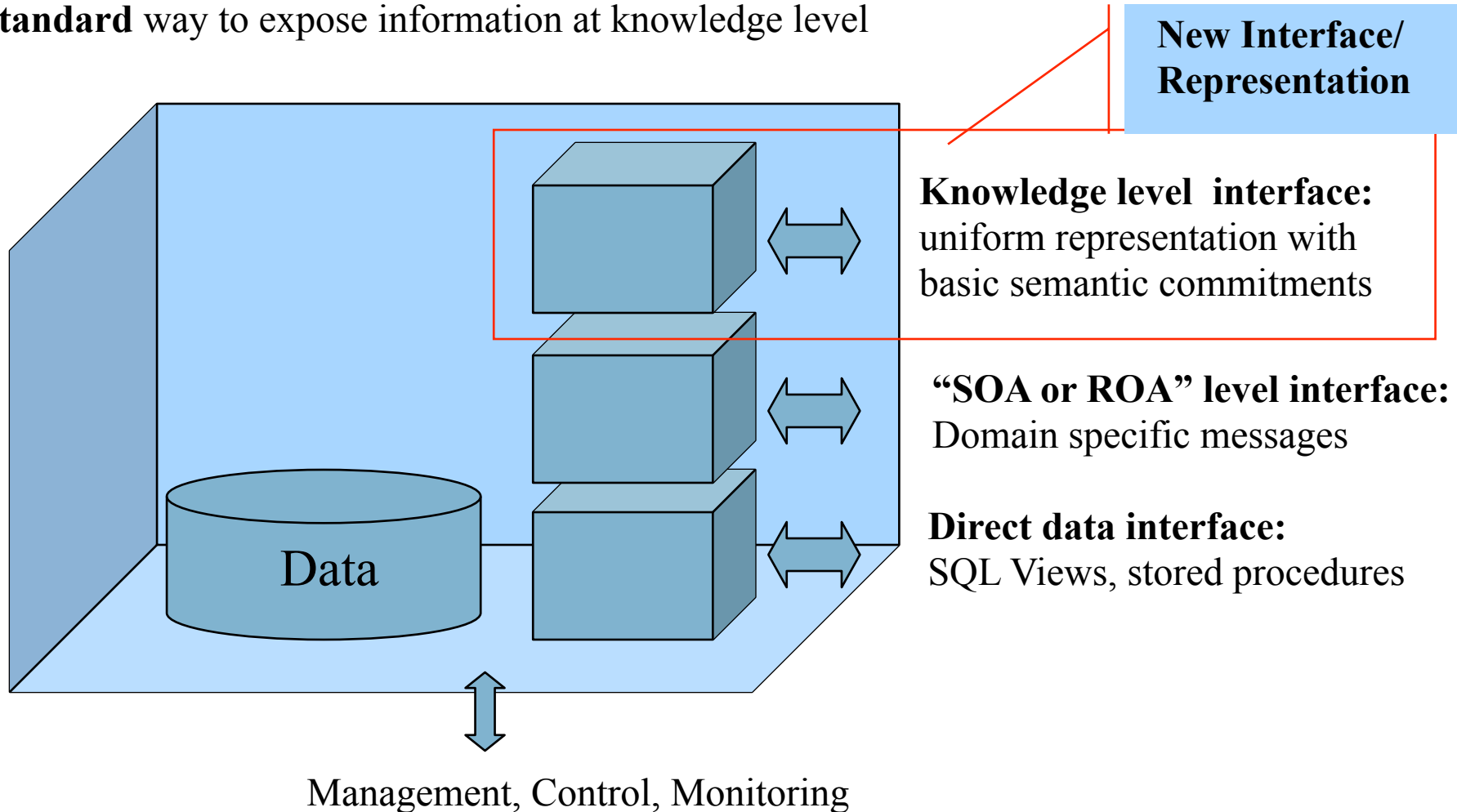- *Key attention: enterprise-wide identifiers*

Facts (summary)

Exporting summary of existing data

Application/Service:
- Add a new interface
- **Standard** way to expose information at knowledge level

New Interface/ Representation

Knowledge level interface: uniform representation with basic semantic commitments

"SOA or ROA" level interface: Domain specific messages

Direct data interface: SQL Views, stored procedures

Data

Management, Control, Monitoring

For Some services/applications

- **Use Topic Maps as a main representation mechanism**
  - Great way to handle directory-like master/reference data
  - Provide access to information at "business" level
  - Flexible in terms of defining new entities, properties and relationships (no need to change data structures)
  - Can represent various contexts/scopes including changes in time
  - Topic maps are often implemented on top of relational database (it helps to leverage investment in traditional databases)
  - It is like 4GL for information resources

Topic Maps as a "smart" storage

Topic Map Engine provides data as a web service:
creating, modifying, deleting entities, relationships and properties

Topic Map Engine

Data
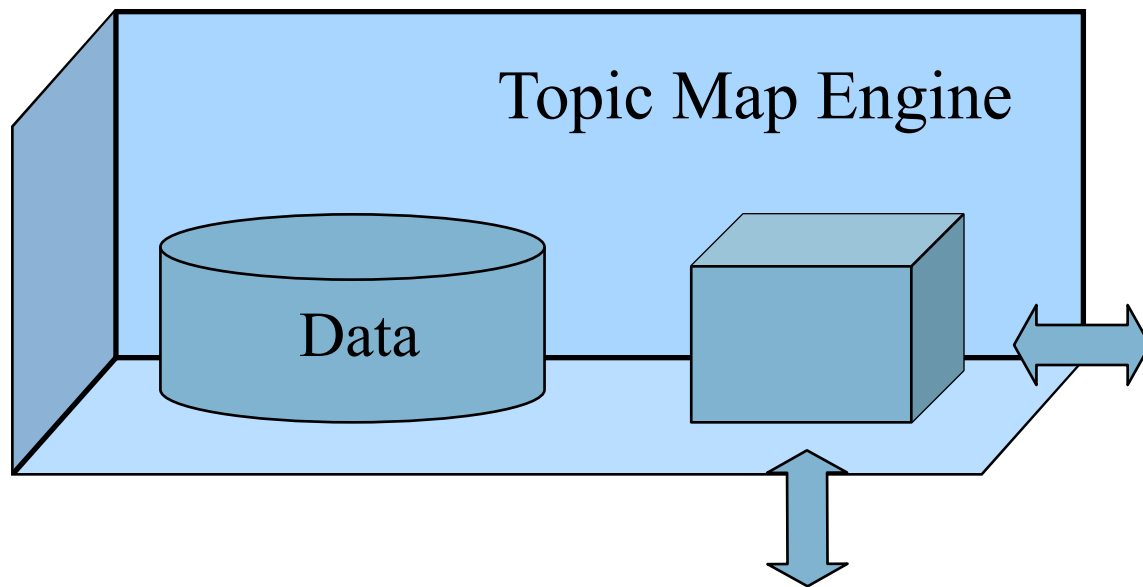
Knowledge/SOA/ROA level interface: uniform representation with basic semantic commitments

Management, Control, Monitoring

We have a knowledge map, what is next?

- Use knowledge map for "tagging" resources

- Create subject-centric portal/pages

- Introduce integrated faceted based search and navigation

- Define configurable subject-centric RSS/Atom feeds

"Tagging" resources

- Use knowledge map as a source for "fixed vocabularies"

  - support for inheritance (and querying)

  - synchronized with existing data sources

  - exists at the "conceptual level"

- Check if your ECM vendor/product supports ability to use "fixed vocabulary" from external sources for tagging resources

- Or create a resource map directly using topic maps

  - issues with moving, deleting resources

Subject-centric portal / pages

- Each main subject should have an explicit representation which is visualized through a "subject page"

  - Every person, department, company, project, business function ... has own subject page

- Subject page is a combination of "structural/reference" information and links to various resources

- Subject page renders integrated information about a subject (360° view), integration is already done at a knowledge map level

- Hint: it is like a combination of Wikipedia page + RSS + widgets

Using "off the shelf" Portals

- Composite interface is a powerful idea

  - Can be implemented with subject centricity in mind

- Configuration instead of coding

  - can save some time

- Personalization

- Role based interface

- Be careful with traditional "on the glass" and low level data integration patterns, promote integration at the "knowledge level"

Faceted Navigation and Search

- http://en.wikipedia.org/wiki/Faceted_classification

- Efficient way to implement "findability"

  - Integrated full-text, metadata search and browsing

  - Dynamic filtering: easy to add/delete conditions

  - Hints about possible directions for search/browsing

  - Multi-path access to information items

  - Universal: for resources, entities, activities, events

- Based on idea of "facets"

- Knowledge map is an excellent source for facets

Facets

- Facets are properties or attributes

- For example, documents can have:

  - Creator

  - Contributor

  - Publisher

    - Organization, Department

  - Document type

    - Presentation, Resolution, Newsletters …

  - Subject

    - Person, Company…

Facets

- Think how main subjects can be classified using faceted approach

- Example: categorization of people

  - http://en.wikipedia.org/wiki/Wikipedia:Categorization_of_people

- What is a correlation between a facet and subject properties/relationships?

  - facet can include some path from a main subject

- Each basic type can have a primary type-subtype hierarchy which can used as a facet for classification of other subjects

Example: Endeca Information Access Platform

- Disclaimer: Endeca related information is available on Endeca's public web site

- Commercial patented MDEX engine

  - http://endeca.com/technology/index.html

  - full text and faceted index

- 64-bit, clustering, scalability

- API for building guided navigation GUI

Faceted Search: example from the Web

Building index: traditional approach

- Direct access to data sources

Facets/Dimensions

Content acquisition using adapters

Data sources:
file system, web pages,
records in database

"Records" classified by various dimensions

Creating an index

Facets/Dimensions

Records classified
by various
dimensions
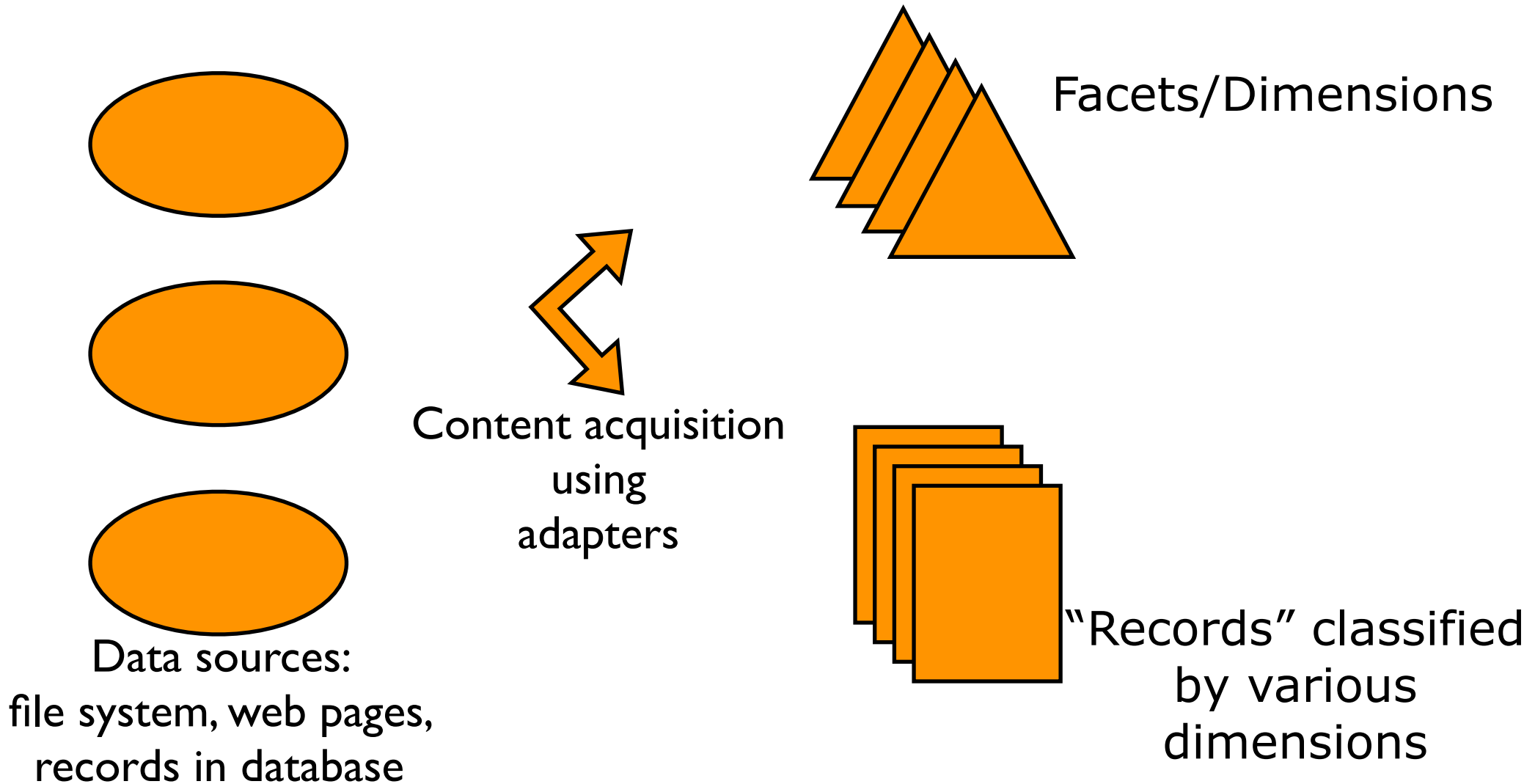
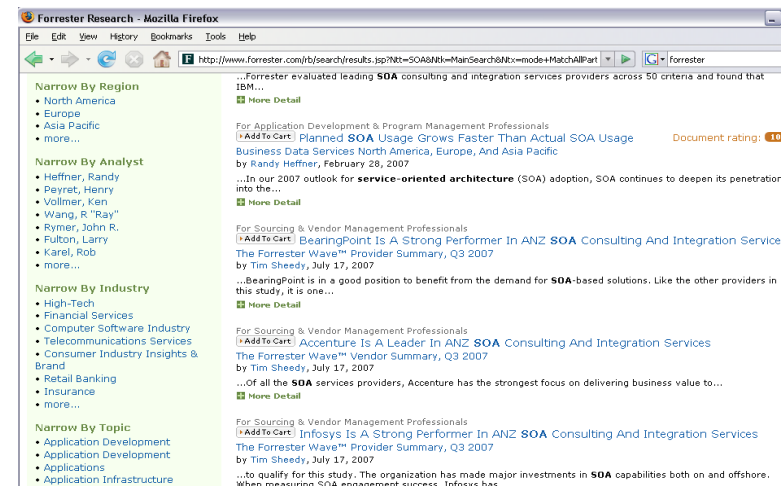Full-Text and Faceted
Index
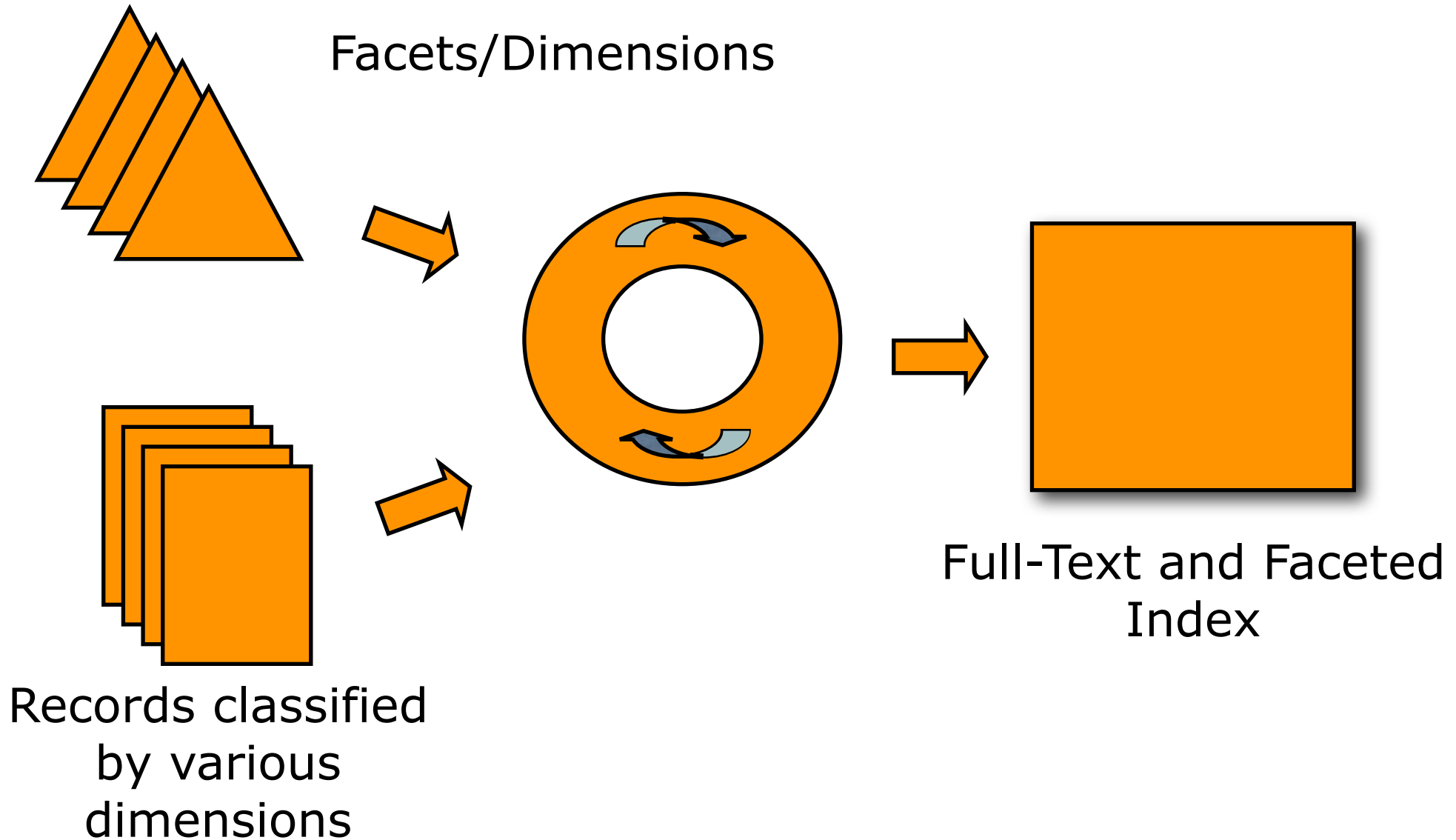
Faceted search and navigation

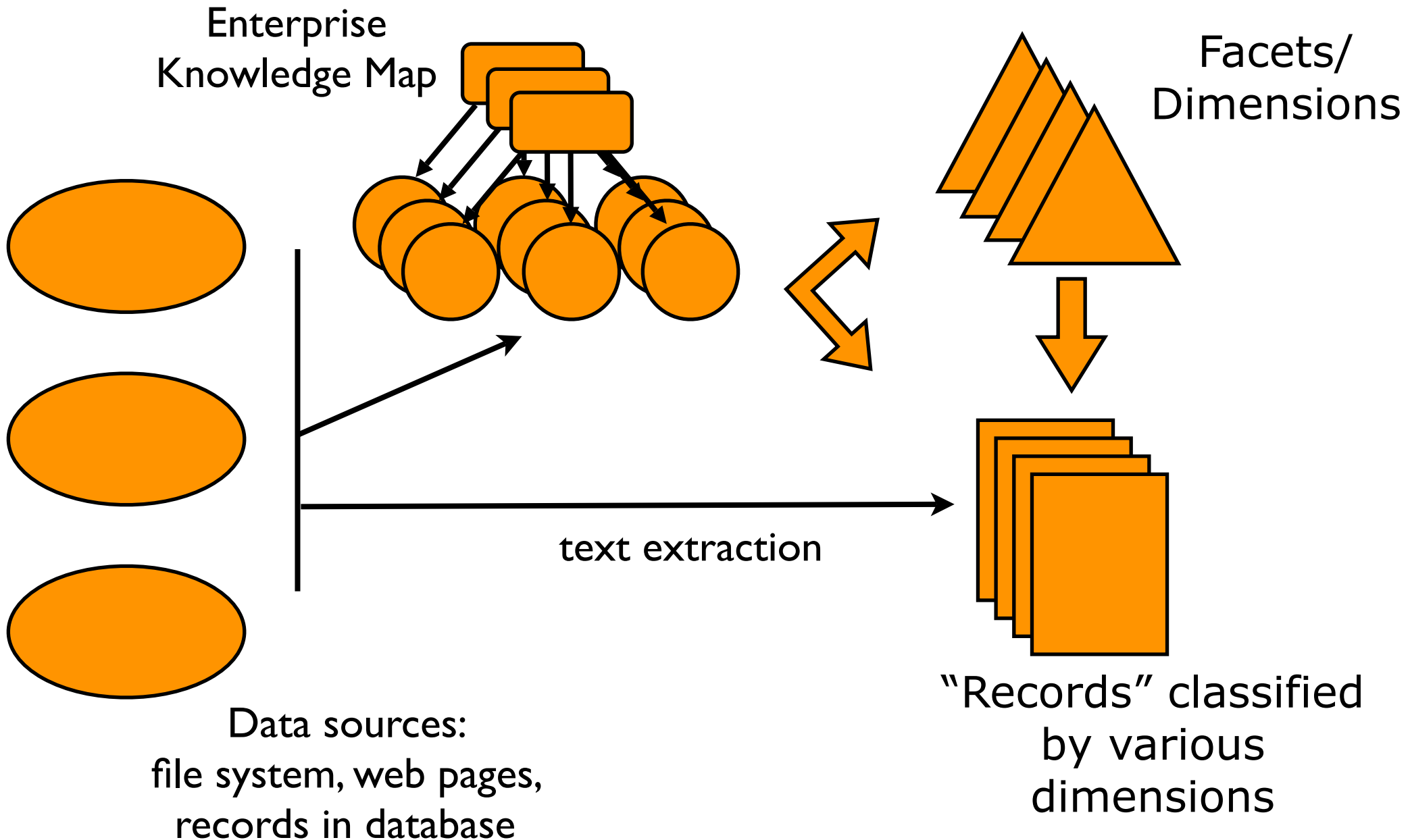*API, Web Services*

Full Text and
Faceted Index

Web server/Portal

Creating facets/dimensions and records: using Enterprise Knowledge Map

Enterprise Knowledge Map

Facets/ Dimensions

text extraction

"Records" classified by various dimensions

Data sources:
file system, web pages, records in database

It is not only about documents

We can search for people, for example:

Type: Person
First Name: John
Last Name: Smith
Works For Company: Example.com
Position: Researcher

- Index supports "records" with different properties

- Occurrences and associations can be used to define properties and classify topics of various types

- Commercial faceted navigation tools can be efficient and scalable

  - and have a nice integration with Enterprise Knowledge Map

Generating RSS/Atom feeds

- Leveraging existing RSS/Atom infrastructure (aggregators)

- Knowledge Map is an excellent source for RSS/Atom feeds

- Information integration is already done

- We can be subject centric or we can be source centric

- Integration with subject pages

 - easy to jump from information item related to a subject to a full subject page

Knowledge Map evolution

- Start with main object, resource and function types:

  - People, organizational units, main business functions

- Extend knowledge map, introduce new types, cover more areas

- Use logging for search and navigation, access to subject pages

- Analyze usage statistics

- Be proactive - modify knowledge map based on business objectives

Issues

- Information provenance

  - Topic Maps (as a standard) do not have a well established way to support provenance, although some tricks are well known and tried

- Representing time sensitive information

  - scope is a good start but without shortcuts modeling changes in time is not as easy as it could be

- Identity management

  - PSIs are important but in many cases other mechanisms have to be used: heuristics for identifying the same topics, workflows to adjust default identification decisions

- RDF/OWL compatibility

Next step: Subject-centric computing

http://psi.ontopedia.net/Subject-centric_computing
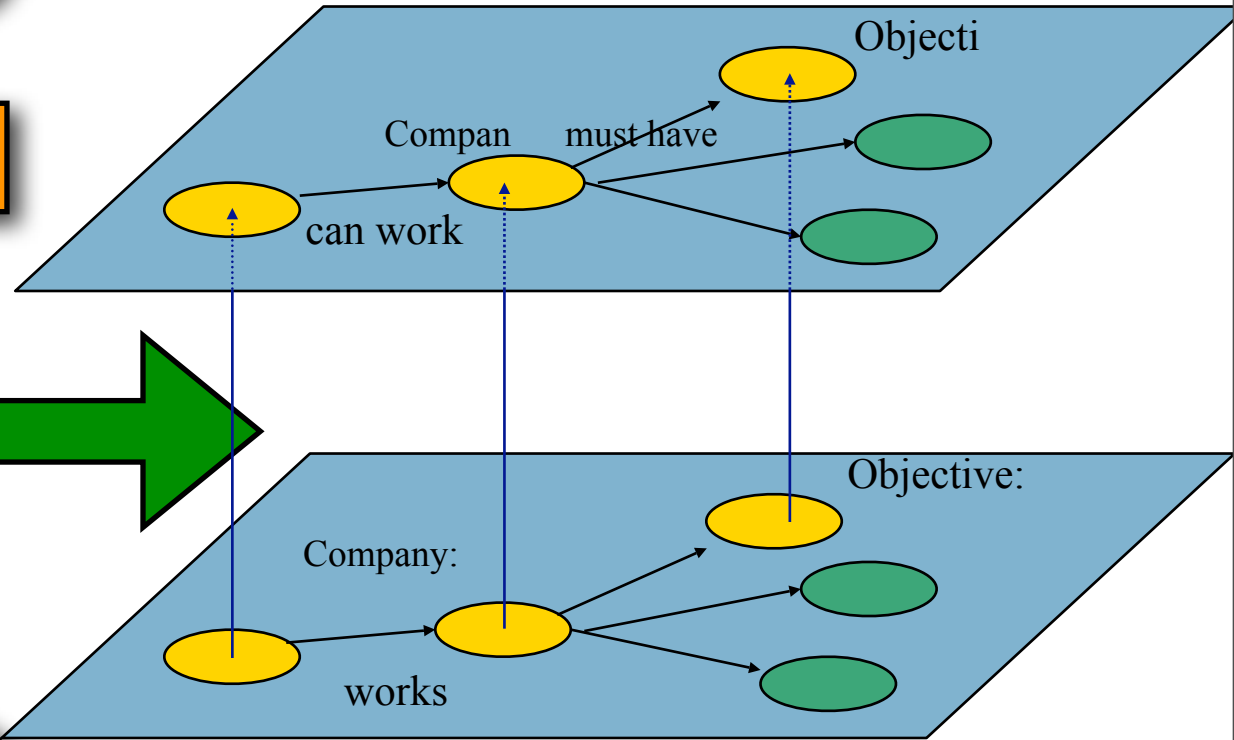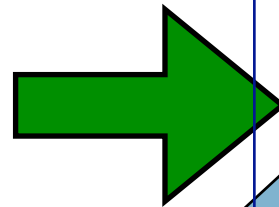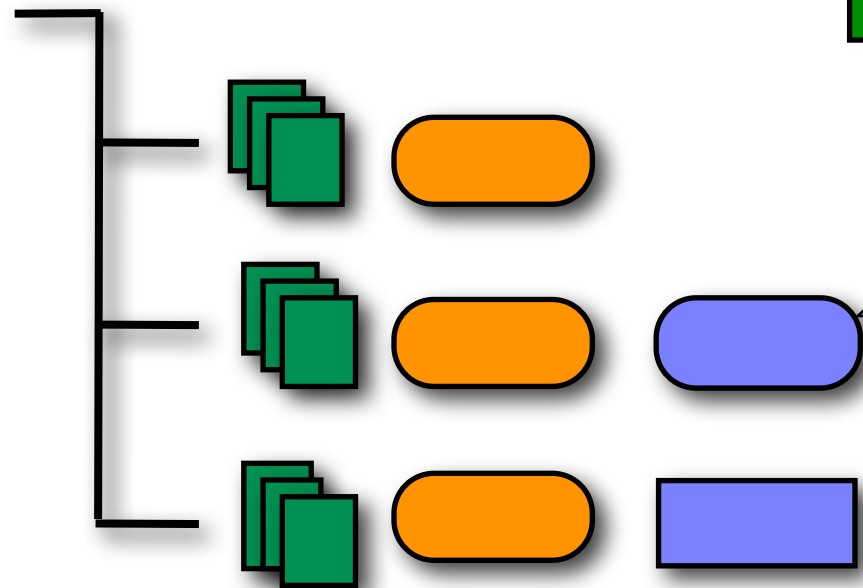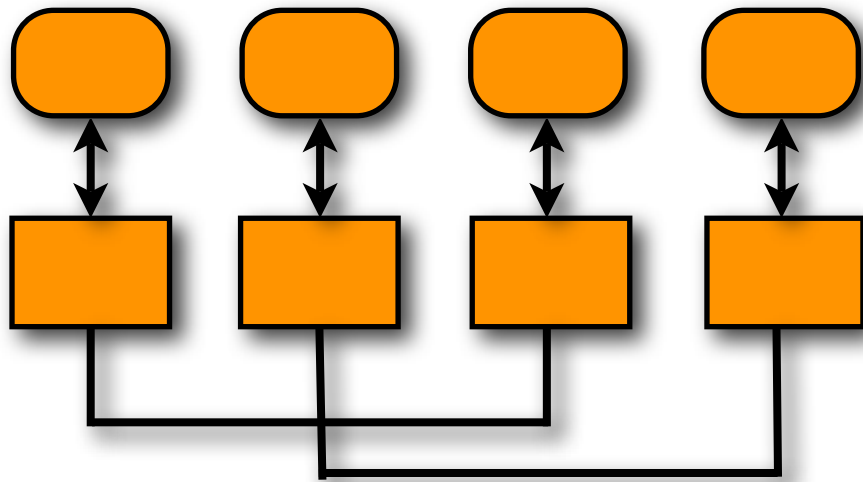
Summary

- Add a "knowledge map" to your Enterprise Architecture

    - Use Topic Maps to implement it

- Commercial "off the shelf " products can help with "enterprise features"

    - Portal

    - Faceted search and navigation platform

    - Topic Maps engine, data/topic map gateways

- SOA, ROA can help to implement "best of breed" solutions.

Questions