

SATURN Update



DAML PI Meeting

Dr. A. Joseph Rockmore

25 May 2004



SATURN: Needs and Challenges [1 of 2]

- SATURN = semantic access to time-ordered url's and related information
- Objective: easier and more accurate access to intelligence information
 - Unstructured intelligence (documents, in broadest sense)
 - Profiled and retrospective search, retrieval, and discovery
 - Based on time, geography, producing organization, **content**, etc.
 - Results displayed chronologically
- Federated search across document collections
- Incorporate promising advanced technologies
 - Especially Semantic Web



SATURN: Needs and Challenges [2 of 2]

- Profiled (“push”) search and retrieval
 - Users can establish any number of standing queries
 - Resulting metadata displayed chronologically (like Google News), in sections (per standing query), with pointer to document
 - Customizable display (number of results, sorting, etc.), updated as new documents become available
- Retrospective (“pull”) search and retrieval
 - On-the-fly queries using the same form as for establishing profiles
 - Resulting metadata displayed chronologically or by relevance, with pointer to document
- *Support making conclusions on metadata alone*
 - *Capability depends on the depth of metadata expressed and returned from query*

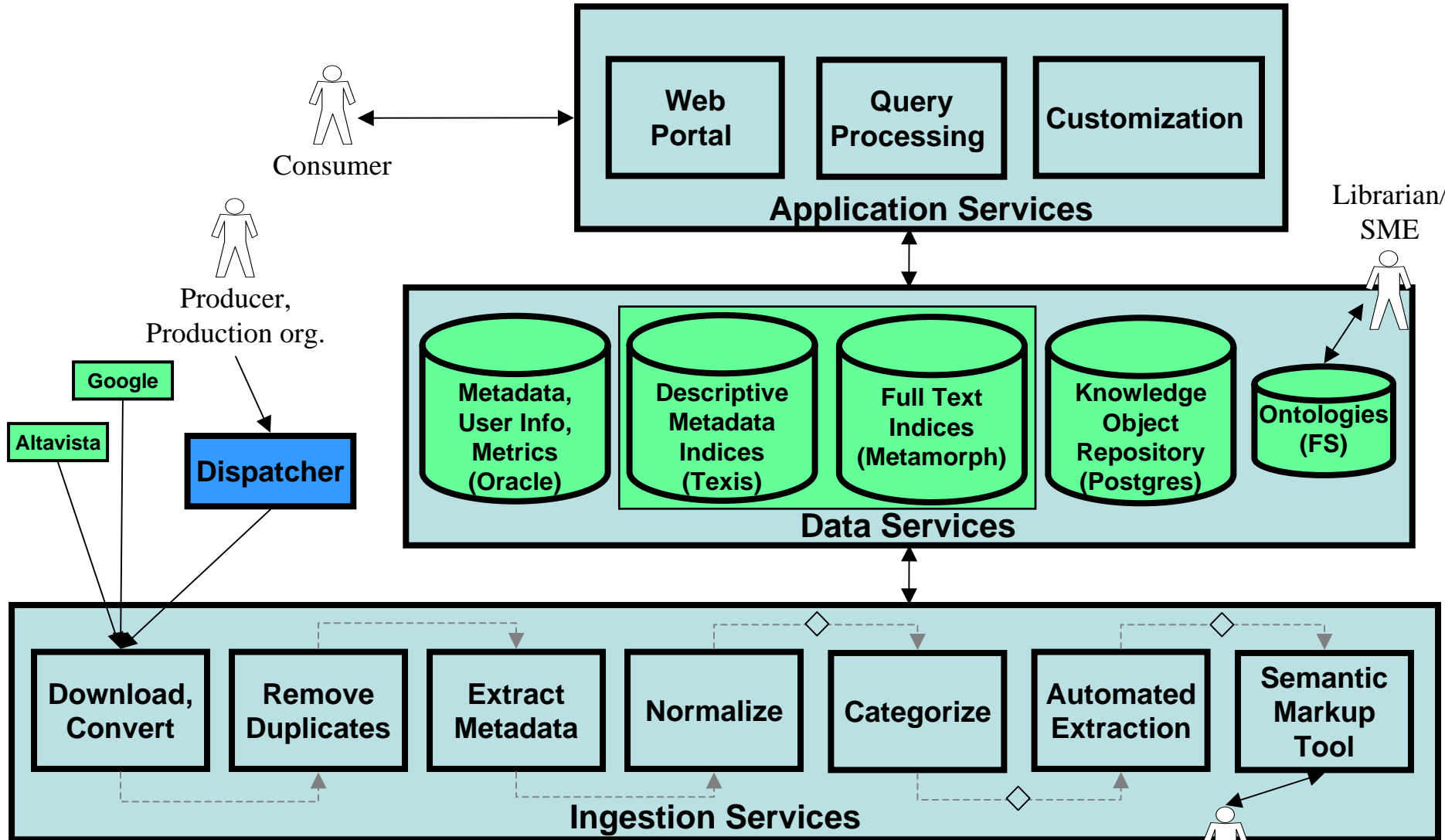


SATURN: Semantic Markup Approach

- If product has no metadata assigned
 - Derive estimates of time, organization, geographic reference, etc.
 - Use automated categorization tools to derive subject (according to common ontology)
 - For subset of all documents, use automated markup then semantic markup by SME to derive entities, relationships, etc.
- If product has metadata assigned by author or producing organization
 - Normalize to common syntax and semantics
 - Mappings may be difficult to produce and to apply
 - May have to add metadata not assigned as above
- Use assigned metadata to “train” automated tools

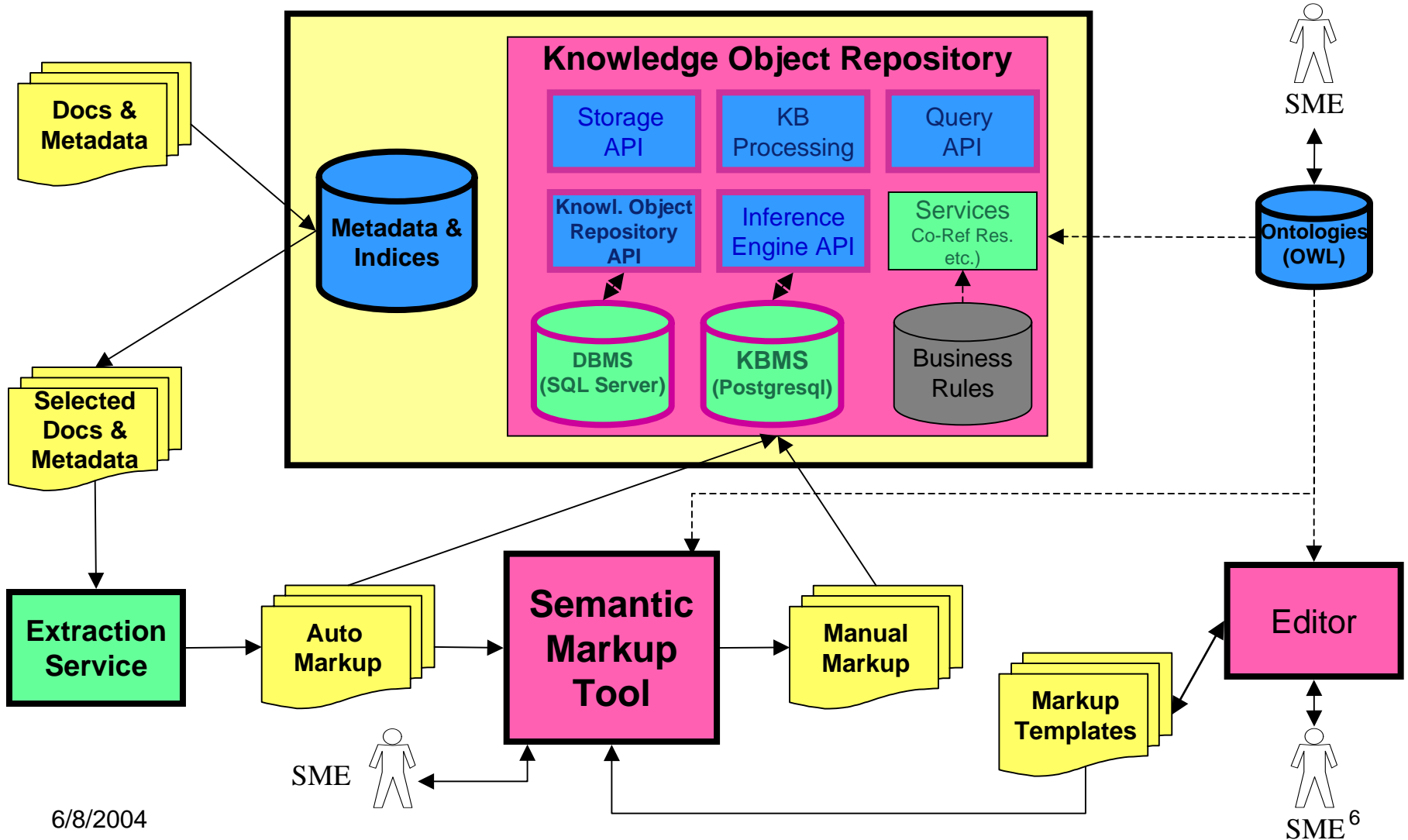


SATURN Functional Architecture

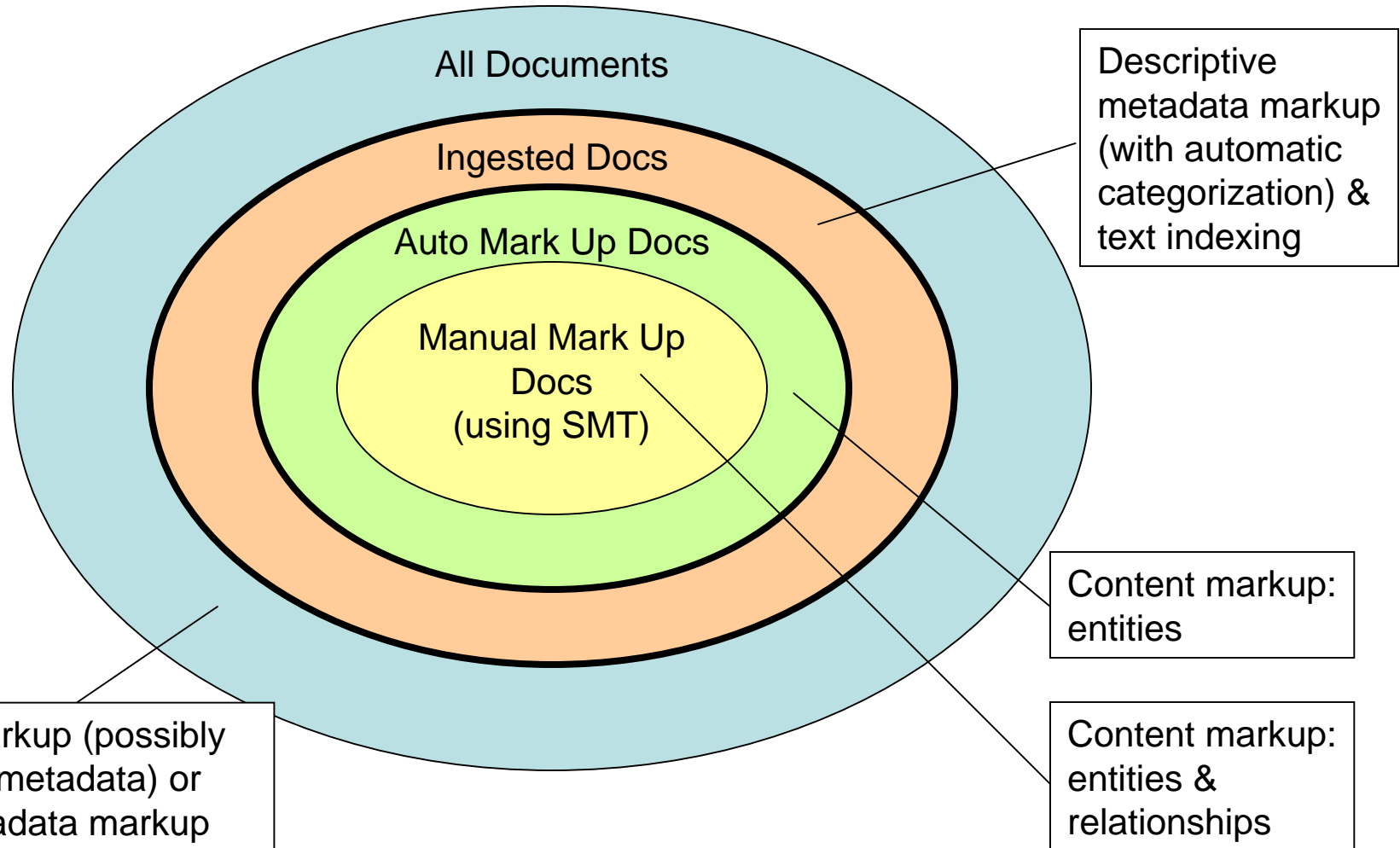




Semantic Markup Functional Architecture



Document Markup Landscape





SATURN API's

■ Portal

- TBD (possible ones include accessing an external knowledge/data source, calling a search tool or results ranker, etc.)

■ Document Ingest

- Arbitrary query including category

■ Data Services

- Accessible using JDBC (could implement as a web service)

■ Markup Tools & Knowledge Object Repository

- Knowledge object repository has an external API (OWL-based, queries in DQL/KBQL supported)
- Entity extractor API for plugging in different entity extractors
- Co-reference resolution service to use an external reasoner to match knowledge objects



Longer-Term Technical Challenges

- Knowledge Object Repository
 - Investigating various COTS/GOTS options
 - OWL support/inference making & scalability
- Co-reference Resolution
 - Investigating several approaches
- Integration of search with varying levels of metadata
 - “Virtual” documents (from knowledge objects)
 - Investigating applicable DAML Program tools
- Integration of Semantic Markup Tools into Early Adopters’ Production/Dissemination Approaches
 - Using .net for SMT (GUI) implementation to facilitate
 - Support for local storage/exploitation of OWL markup from SMT
- Support for ontology evolution once SMT & Knowledge Object Repository is deployed
 - DIONE work on ontology versioning by ISX & Lehigh Univ.



Example: Markup Tool

The screenshot displays the 'Semantic Markup Tool' interface. On the left, a form titled 'Iraqi debt' is shown with various fields for metadata. The 'Value' column contains dropdown menus for 'Iraqi debt', 'March 14th, 2002', 'President Bush', 'James Baker', and 'Iraq'. There are also buttons to 'Add Participant', 'Add Meeting Topic', 'Add Sponsor', and 'Add Facilitator'. The 'Meeting' tab is selected, and the breadcrumb path is 'Meeting > Iraqi debt > Donors' Conference > James Baker'. The main area on the right shows the 'Source Document' with a 'View' dropdown set to 'HTML'. The preview shows the following HTML output:

substantial reduction of iraq's foreign debt within the context of the Paris Club in 2004.

Prime Minister Berlusconi agreed with Baker over the opportuneness of an accord with the other creditor countries for reducing the burden caused by a dictatorial regime which has reduced iraq to poverty. The debt reduction will be decisive both for breathing new life into the economy of iraq and for the security of the country, as well as for building sovereign and democratic institutions. Baker expressed President Bush's great appreciation for the efforts of our country aimed at guaranteeing the iraqi people a future of freedom. Prime Minister Berlusconi noted that the human and financial resources made available by iraq, including through its participation at the Donors' Conference in October, are the expression of a commitment to the development of an iraq that is democratic, free, and prosperous, after decades of dictatorship and violations of human rights.

Template Text

A meeting called **iraqi debt** was held at **Iraq** on **March 14th, 2002** concerning **<Meeting Topic>**. It was organized by **James Baker** and **President Bush**. **James Baker**, and **<Participant>** were participants at the meeting.

Semantic Markup Editor



Summary

- In the real world, metadata ranges over a spectrum, from simple descriptive to deep content
 - Many documents will have no, bad, or simple metadata
 - It will be cost effective to add deep semantic markup to only a subset of all documents
- Integration of search and retrieval, metadata browsing, and analysis must be across documents with all types of metadata
- To the extent possible, support analysis and making conclusions on metadata alone
 - Use documents for detail and reference