servicefinder

ESTC2008
2ND ANNUAL EUROPEAN SEMANTIC TECHNOLOGY CONFERENCE

Realizing Service-Finder

Web Service Discovery at Web Scale

http://www.service-finder.eu

Lecturer:

Emanuele Della Valle

emanuele.dellavalle@cefriel.it

http://emanueledellavalle.org

Authors:

Emanuele Della Valle, Dario Cerizza, Irene Celino, Andrea Turati,
Holger Lausen, Nathalie Steinmetz, Michael Erdmann,
Wolfgang Schoch and Adam Funk

CEFRIEL FORGING INNOVATION

seekda!

The University Of Sheffield.

ontoprise know how to use Know-how

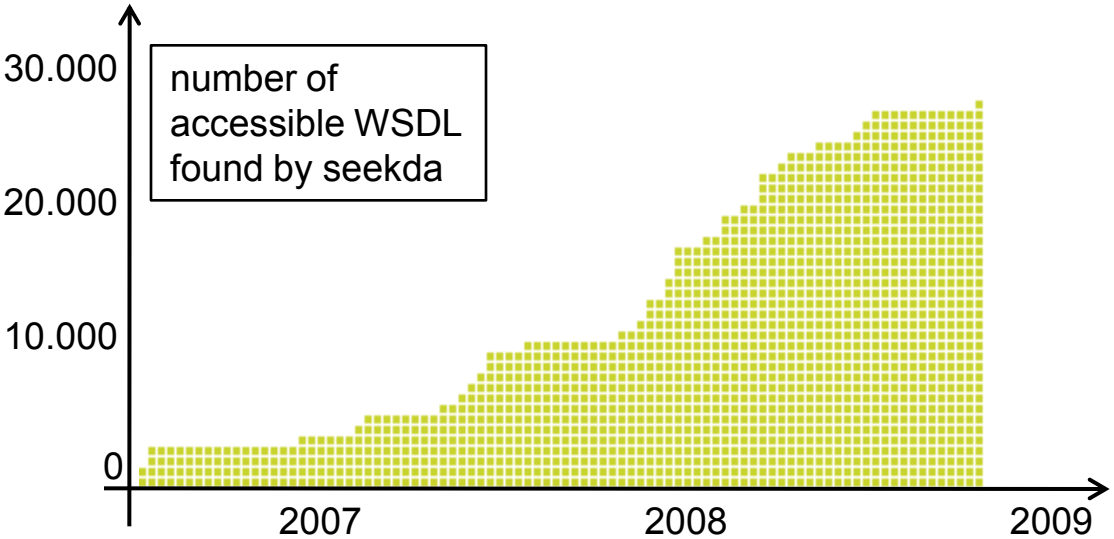
Agenda

- Introduction
  - SOA onto the Web
  - Drawbacks and pitfalls of public UDDI registries
  - Overcoming UDDI Limitation
- Service-Finder
  - Project idea
  - Key objectives
- Realizing Service Finder
  - Story boards
  - Requirements
  - Architecture and components
  - Work in progress
- Conclusions
  - Beyond state-of-the-art
  - Expected impact
  - Exploitation prospects

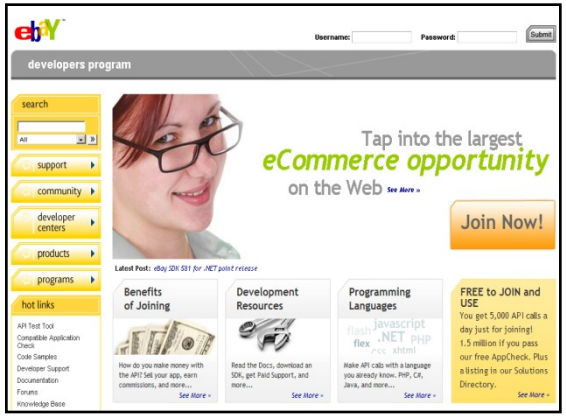
Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

SOA onto the Web

- Service Oriented Architectures (**SOA**s) along with Web Services technologies are widely seen as the **most promising fundament for** realizing service interchange in **business to business** settings.
- **However**, it is envisioned that **SOA**s and Web Services will increasingly **move** out of these settings and out **onto the Web**.



[source http://seekda.com/about/web_services (September, 2008)]



[http://developer.ebay.com/]

[http://aws.amazon.com/]

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

The rise and fall of public UDDI registries

- One of the **essential building blocks** for creating applications that utilize the vast quantities of services, which are available on the Web is making it easier to **discovery and select** the right services.
- **UDDI** was **initially proposed** as a component of Web Services usage process enabling registering and discovering services, **but** finally UDDI **did not reach** its expected potential.
- The **critical problem** in this new Web oriented environment is one of **scale** because services appear, disappear and change at a rate much higher than in business to business settings.

UDDI Business Registry Shutdown.

"*With the approval of UDDI v3.02 as an OASIS Standard in 2005, and the momentum UDDI has achieved in market adoption, IBM, Microsoft and SAP have evaluated the status of the UDDI Business Registry and determined that the goals for the project have been achieved. Given this, the UDDI Business Registry will be discontinued as of 12 January 2006.*"

[from "Registering for UDDI" 2005-12-17]

[see http://xml.coverpages.org/uddi.html]

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Pitfalls of public UDDI registries

1. UDDI is centered around programmatic access to the registry and **only** a few mostly **technically focused user interfaces are available**.
2. The **information** in public UDDI registry was often **outdated**. The **value** of the service in the public UDDI registry **is minimal if the service** itself **does not exist** anymore.
3. There are **no means for community feedback**. Practically there is **only** one possibility to provide feedback allowing the user to contact a provider by **email** listed in the service description.
4. A **WSDL definition** and a short description **is not sufficient** for a service consumer **to select a service**. To make decision about applicability of the service, **service consumer need** to become familiar with **pricing**, **terms and condition**, **s**ervice **l**evel **a**greements to name just a few.

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

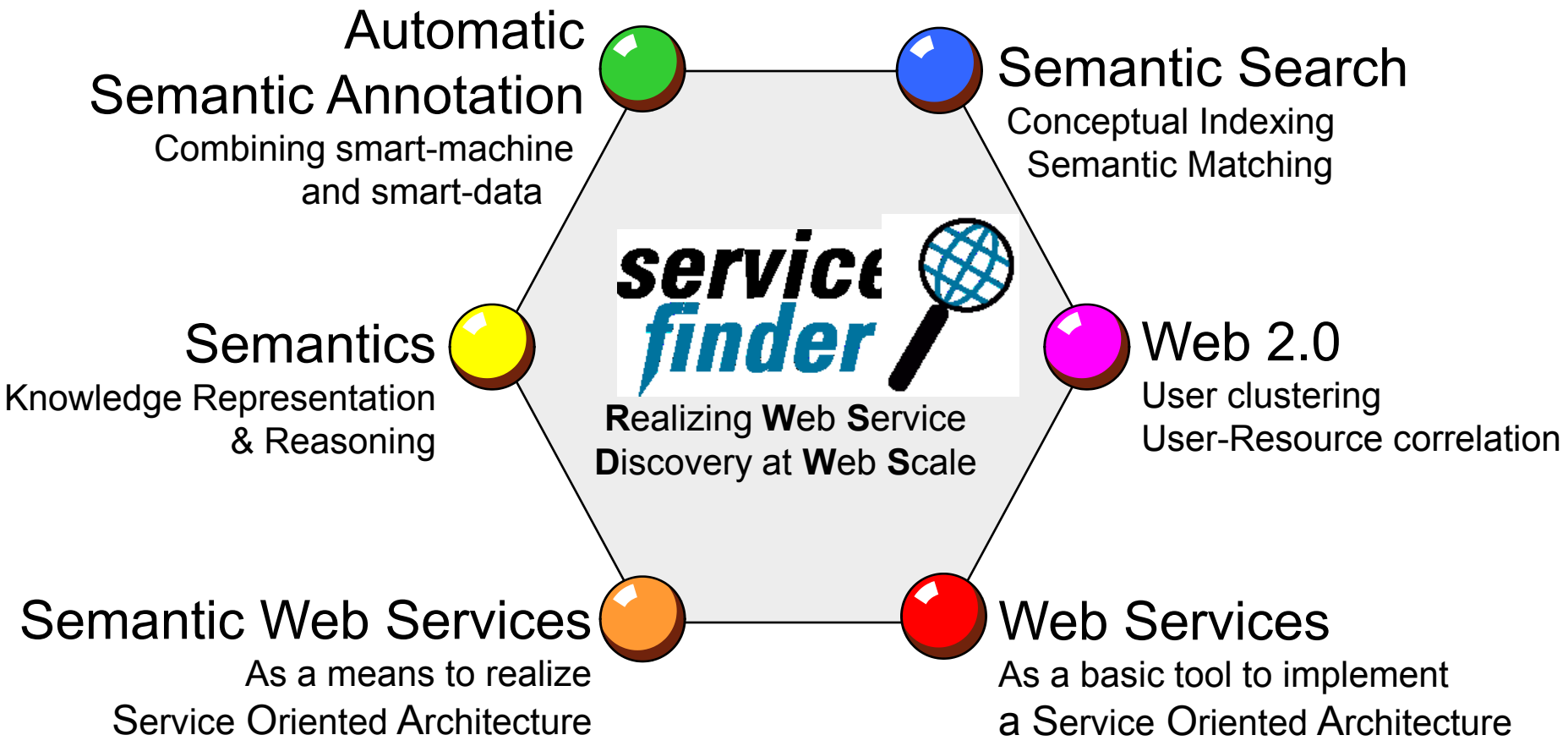
Overcoming UDDI limitation

1. **Easy to use GUI** - It is important that early adopters of Web Services technology, who learns about it for the first time, should be able to start exploring it with a few simply steps.
2. **Search Engine style** - Web is unpredictable and services can appear and disappear (the same as websites), but one can put up a mechanism (periodic crawling and availability check) allowing to eliminate these services which are not available any more.
3. **Architecture of participation** - Learn from Web 2.0 (e.g., wikis, blogs, etc.) in enabling community.
4. **More useful info** - Include all information required by a user to make decision about applicability of the service; e.g., pricing, terms and condition, service level agreements, etc.

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

servicefinder project idea

Service-Finder aims at developing a *platform for service discovery in which Web Services are embedded in a Web 2.0 environment*

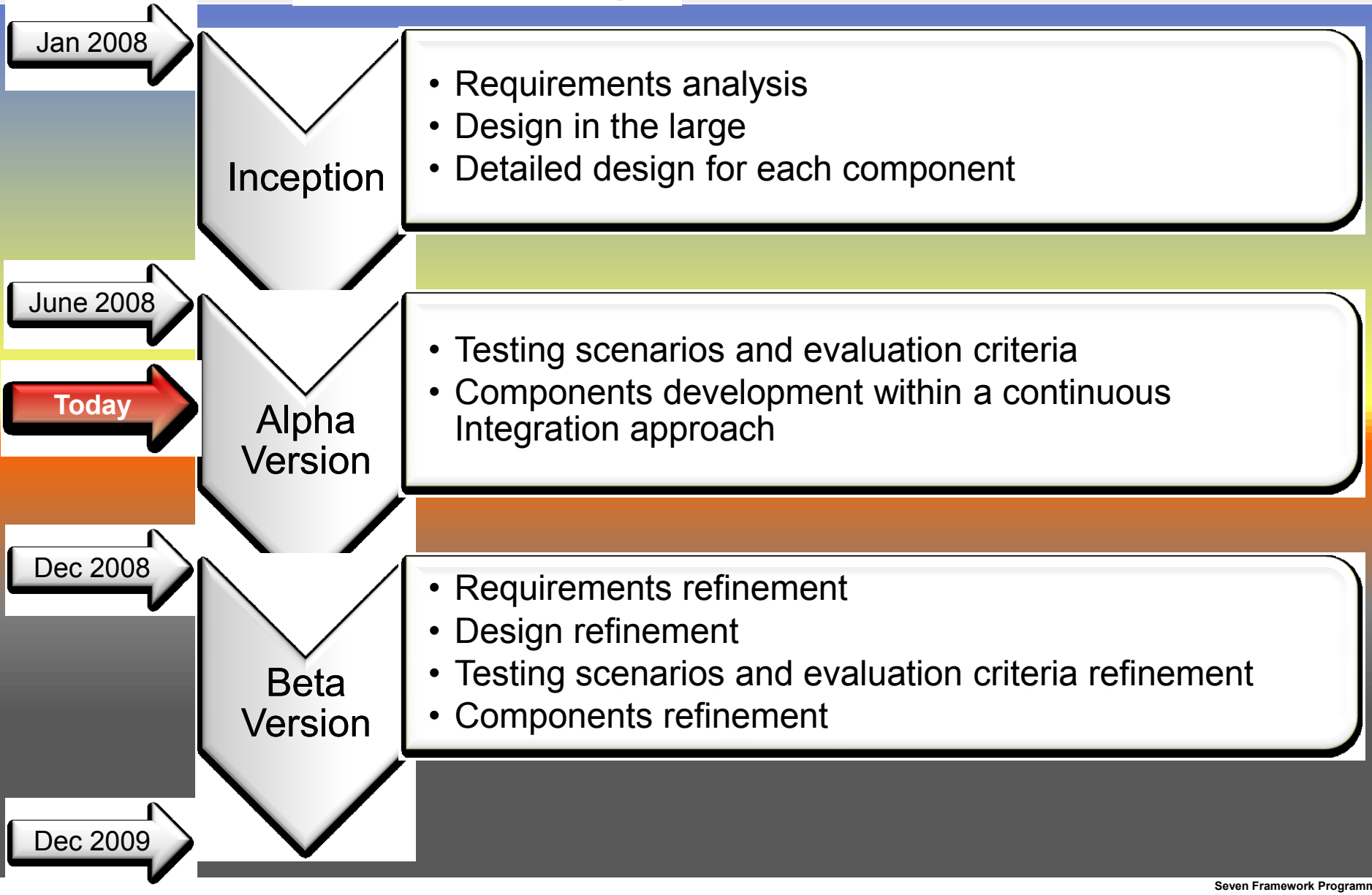


servicefinder key objectives

- Create a **Semantic Search Engine for Web Services**
  - Aggregates information from heterogeneous sources: WSDL, wikis, blogs and also users' feedbacks and behaviour
- Create a **Web Service Crawler** to identify Web Services and their relevant information
- **Automatically generate Semantic Service Descriptions** by analyzing heterogeneous sources
- Allow **efficient and effective search** of collected and generated data
- Provide a **Web 2.0 portal**
  - To support users in searching and browsing for Web Services
  - To give recommendations to users
  - To track user behaviour for improving accuracy of service search and user recommendations

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Realizing servicefinder



Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Use cases for servicefinder

- To gather requirements we imaged several use cases
  - **A system administrator** at a bank who is looking for an SMS Messaging **service that sends him an SMS** in any case failures with the on-line payment system of the bank.
  - **A business and technology consultant** working on a e-health project that needs to make it possible for general practitioners to **send and receive fax** directly from their patient record application using an on-line service.
  - **A web developer** that, after using a service listed on Service-Finder, decides to **edit the information on the portal** in order to improve it for other community users

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Requirements for servicefinder

- **We identified** within those previous use cases more than 60 **requirements** and we grouped similar requirements together into three main categories:
  - **Search related**: search for text, search for tag, search for concept, disambiguation, facet-browsing, ranking, sorting, comparing, etc.
  - **Web Service information related**:
    - **Services details**: interface, how can the service be used, its payment modalities, its terms and clauses, user-added information as ratings, comments and tags, measured values of service levels such as availability (uptime) or performance (response time) and the service level declared by the provider.
    - **Providers info**: name of the provider and its references, user-added information as ratings, comments and tags
  - **User Community related**: rating, commenting, tagging, editing, writing wiki entries, registration, recommendations

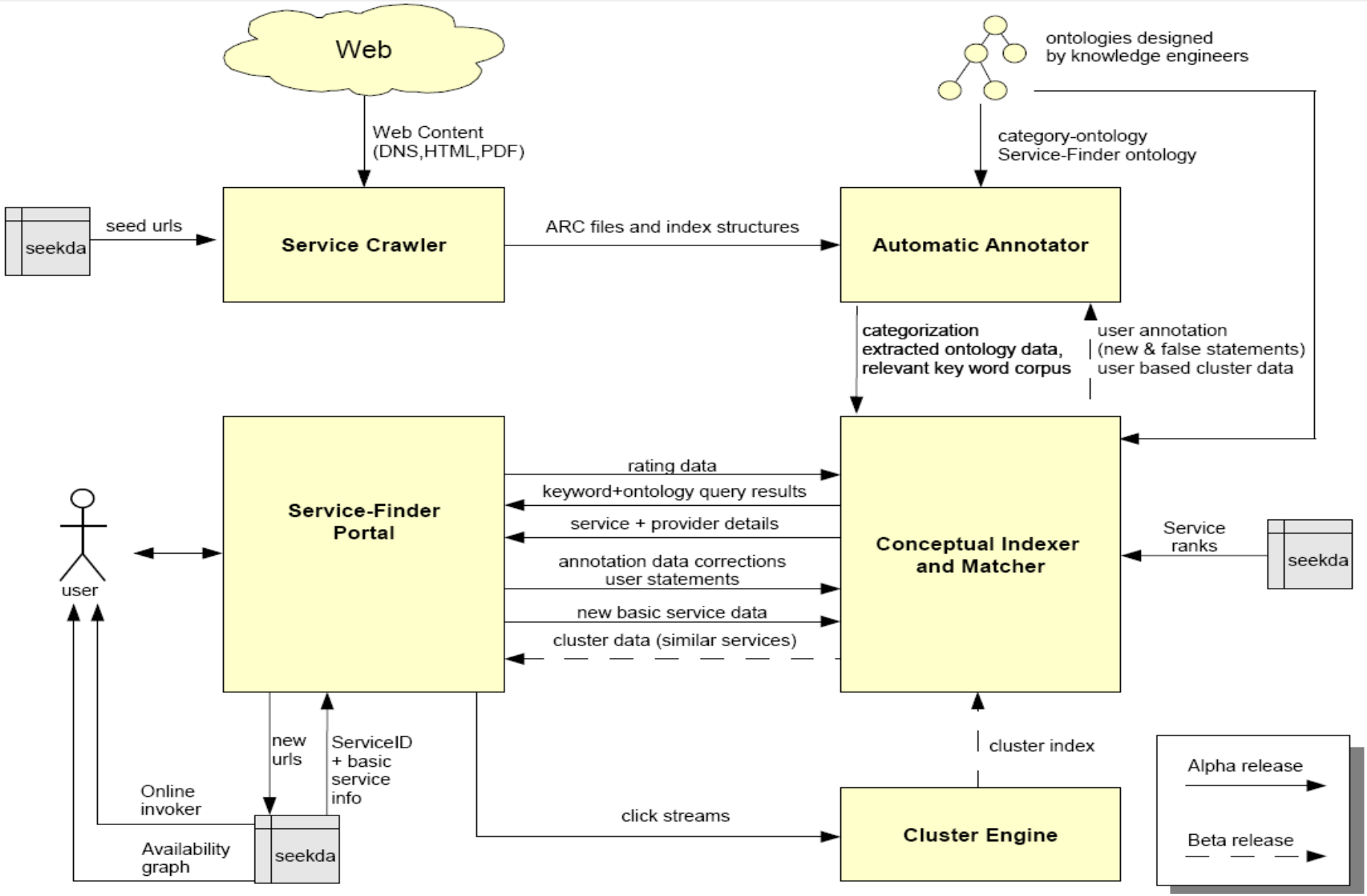
Provider-Related Requirements

- Any publicly available Web Service has somewhere on the Web a corresponding **interface description published** (e.g. using WSDL)
- **Addition Information** (e.g. service coverage, service availability, FAQs, price, etc.) about a service should be located **on the same domain** than the service description itself
- A **free service trial** should be **available**

- NOTE: we are, in this phase of the project, **focusing on WSDL** service descriptions, because they are both easier to detect on the Web and easier to analyze, as they have a standardized interface. Nevertheless not all publicly available services are described in WSDL, providers often use the **REST** paradigm or JSON. Thus, **in a second step** we will try to define methods to detect other service descriptions than WSDL on the Web.

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Architecture and components



Service Crawler by seekda!

- It **produces a snapshot** of the part of the Web that is relevant to Web Services, **including** both **service descriptions and related documents**
- It **proceeds with a first analysis** of the crawled data.
  - **Recognizing services**
    - either when Multiple WSDL files correspond to one service (e.g. multiple hosting of one service)
    - or when one WSDL description might contain more than one single service.
  - **Assigning a unique URL** to the service, e.g. http://seekda.com/providers/cdyne.com/IP2Geo
- It builds index for allowing random access to the snapshot to the Automatic Annotator
- It hangs over the pre-analyzed data set to the Automatic Annotator

Automatic Annotator by

- It unpacks and pre-processes the crawled data creating a **GATE**[1] serial datastore containing two corpora, wsdl and html.
- It **performs information extraction** on WSDL and HTML documents
  - extracting information from WSDL
  - **classifying HTML documents** in the data store as ContactDetails, Pricing, AutoGenerated, etc.
  - **extracting information** from those documents (such as company contact details and service descriptions).
- It hands over the Annotation Results for the Conceptual Indexer in RDF

[1] http://www.gate.ac.uk

Conceptual indexer and matcher by ontoprise

- The conceptual indexer and matcher is **the central data store for all information** that has to be used by multiple components within the Service-Finder architecture.
- It **stores the semantic annotations from** the **Automatic Annotator as well as** those provided by the **users** through the interface
- It also stores and indexes the textual description provided by the Automatic Annotator, as well as the textual comments provided by users.
- It provides **semantic querying capabilities** on top of the data stored that allow to do matchmaking between user request and service offers as well as retrieval capabilities.
- In order to allow the user to intuitively create queries it **allows combining a keyword search with an ontological query**.
- It is based on **OntoBroker**[1]

[1] http://www.ontoprise.de/de/en/home/products/ontobroker.html

Interface by CEFRIEL

- Service Finder Interface represents **the main entry point for a user who wants to search for services**. It provides the users with search functionalities to help them in finding the most appropriate services to fulfill their needs.
- In particular the user can:
  - **search** services **by keyword**, **tag** or **concept** in the categorization;
  - **sort** and **filter** query results by refining the query
  - **compare** and bookmark services for those services that offer this functionality, try out the service
  - register to the portal and **contribute in a Web 2.0 fashion** by tagging, rating, commenting and adding descriptions/properties to services
  - allow developers to invoke Service-Finder functionalities through an API access to service data
- It is based on lesson learned in implementing **Squiggle**[1] (CEFRIEL's semantic search engine) and SOIP[2] (CEFRIEL's semantic portal)

[1] http://swa.cefriel.it/Squiggle

[2] http://swa.cefriel.it/SOIP-F

Cluster Engine by

- This is **an experimental feature that aims at harnessing Wisdoms of the Crowds** as done in many Web 2.0 successful approaches (e.g. Amazon recommendations, Netflix movie clusters, Last.fm playlists, etc.)
- It will use the implicit and explicit feedback that users of the Service-Finder portal will leave when they interact with the portal in order to derive clusters of users and services.
- Intuitively, it does so by identifying from users' history, those users that behave similarly and, for each group of users, by identifying the services they usually interact with and group services used by users belonging to the same cluster.
- **It finds** (unlabeled**) clusters of users/services and it uses them to recommend services to users**.

Work in progress

- We are in the process of **finalizing the first internal** release of the alpha **prototype**.
- We will demonstrate such internal release to a group of expert during ISWC 2008 in Karlsruhe.
- We plan to **go live** with the alpha prototype by the end of **November 2008**.
- **Keep an eye on http://www.service-finder.eu !**
- **We are looking for testers and evaluators!!!**

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Key innovations of servicefinder

Research Activities	
Automatic Service Annotation	To **automatic create** Web Service **descriptions** by analyzing WSDL and related information • **coping with contradictions** • **using community process** to verify results
User and Service Clustering	To investigate and implement techniques for: • **clustering users** accordingly to their behaviours • **clustering services** accordingly to their usage by users belonging to the same clusters
Research and Engineering Activities	
Conceptual Indexing and Matching	To **apply semantic technologies** in the Web Service discovery domain To adopt them **to** the **new forms of input** descriptions: • **Automatic annotations, clusters, contexts**
Integration Activities	
Service-Finder Portal	To provide a **Web 2.0 portal** • **demonstrating** the developed technologies • **fostering communities participation**

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Beyond state of the art

Feature	State of the art	Improvement
Architecture for **lightweight semantic service discovery**	Approaches based on a registration process or an editorial team	Enables to **scale** service discovery with the upcoming increase of publicly available services
Largest and most accurate set of publicly available services	Specialized portals only containing subset of services	Focused crawler able to identify services and related information
Automatic metadata creation for Web Service	Innovative; under-researched	**Metadata generation from Web 2.0 data and services**
Integration of formal and informal (textual) knowledge	Indexed textual descriptions	**Hybrid match-making algorithm**
Automatic creation of both user and service clusters	Only general-purpose clustering techniques exist	**Specialize clustering algorithms** that jointly cluster users and services
Innovative interface that combines Web 2.0 features and service related features	Current Web 2.0 portals do not include semantic metadata.	Techniques that enable **handling of semantic metadata in Web 2.0 portals**

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Expected Impacts

- Service-Finder provides core **mechanisms to cope with changing environments**:
  - It uses **Web principles** such as openness and robustness;
  - It takes **explicit and implicit user interaction** for construction, improvement and validation of rich service description; and
  - It exploits **Semantic Web technologies** as means to organize internally the data on available services.
- It simplifies the service publishing process by removing the burden of any registration and **brings service discovery even to non-technical persons**.
  - **Publishers** increase their productivity, by being able to provide complex services without the need to register them explicitly.
  - **Creators** become able to design more communicative forms of content by integrating third party services.
  - **Organizations** can automate their processes by quickly finding adequate services.

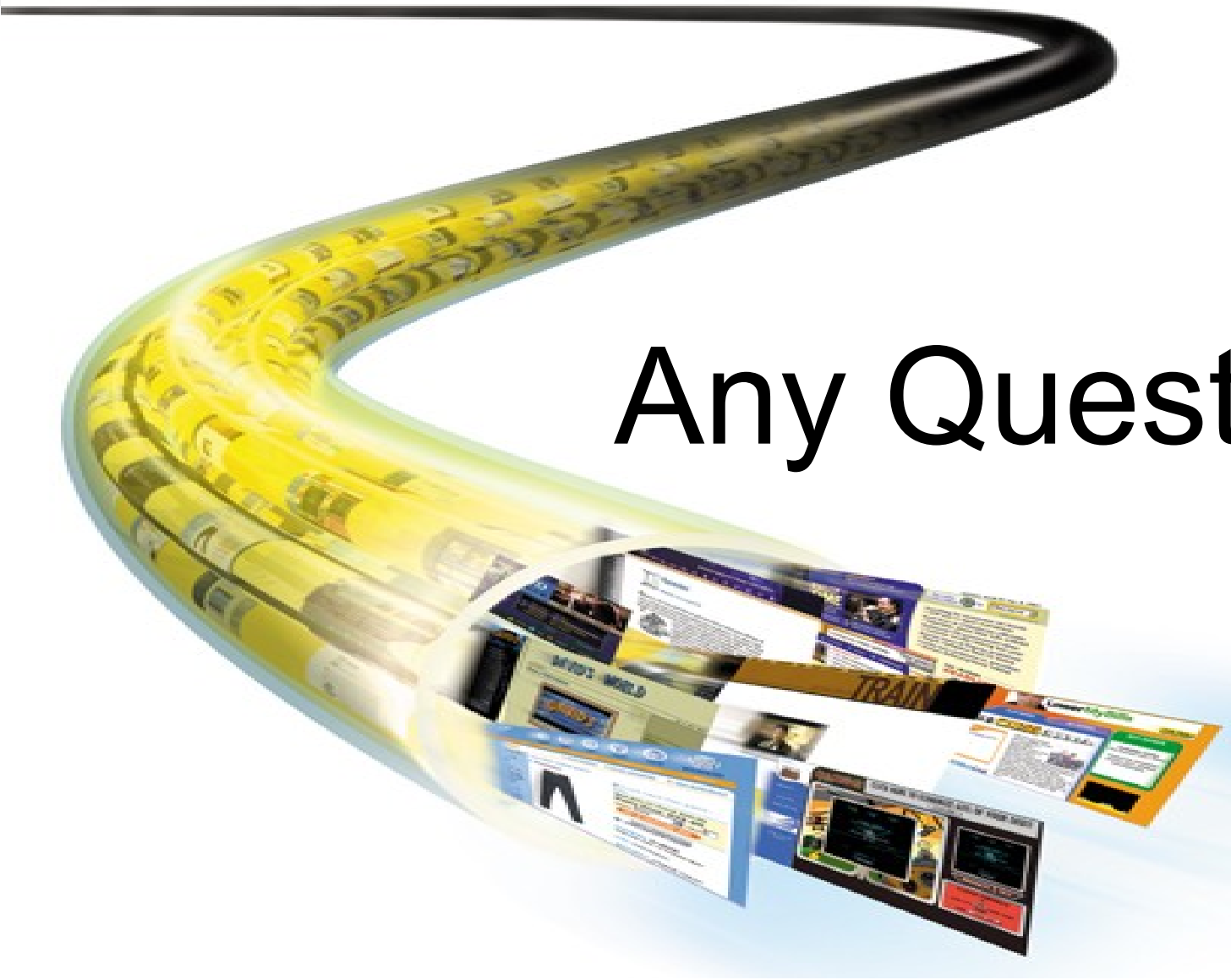
Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Exploitation Prospects

- **The results** of the Service-Finder project **have the potential to revolutionize this market and to outperform existing solutions**
- Using Service Finder for Public services
  - **Unique chance**
    - market for public services increases (xignite, cdyne, …)
  - **Missing Alternatives**
    - UDDI (has been shutdown in 2006)
    - Google (no reliable filter / no additional information)
    - Portals (rely on editorial process <=400 services)
- Service finder **can also be applied within organizations**
  - Number of Services increases in organizations
  - As within internet repositories in big companies can be quickly outdated
  - IT Manager like minimal invasive technology

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

Thank you for paying attention



Any Questions?

Seven Framework Programme
Information Society Technologies (IST)
Specific Targeted Research Project

servicefinder

ESTC2008
2ND ANNUAL EUROPEAN SEMANTIC TECHNOLOGY CONFERENCE

Realizing Service-Finder

Web Service Discovery at Web Scale

http://www.service-finder.eu

Lecturer:

Emanuele Della Valle

emanuele.dellavalle@cefriel.it

http://emanueledellavalle.org

Authors:

Emanuele Della Valle, Dario Cerizza, Irene Celino, Andrea Turati,
Holger Lausen, Nathalie Steinmetz, Michael Erdmann,
Wolfgang Schoch and Adam Funk

CEFRIEL FORGING INNOVATION

seekda!

The University Of Sheffield.

ontoprise know how to use Know-how