

SIRNET – Den Semantiska Webben

# Den Semantiska Webben vad som erbjuds, hur vi kan utnyttja det, och vilka effekter vi kan uppnå

Olle Olsson  
W3C & SICS  
olleo@w3.org olleo@sics.se

[V 0.2]

2004-10-25 - SIRNET



Svenska W3C kontoret



Page 1

## Vem?

**Olle Olsson**

SWEDISH  
INSTITUTE OF  
COMPUTER  
SCIENCE



Nationellt forskningsinstitut – stött av en industrigrupp och av staten –  
avancerad FoU inom IT: teknik och metodik



Svenska W3C kontoret

Internationell konsortium som tar fram generella webbstandarder

2004-10-25 - SIRNET



Svenska W3C kontoret



Page 2

# Vad och hur?

## *Frågor:*

- Vad som erbjuds?
- Hur vi kan utnyttja det?
- Vilka effekter det kan ge?

2004-10-25 - SIRNET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 3

# Bakgrund - trender

2004-10-25 - SIRNET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 4

# Semantiska webben

- Semantik
  - "Gammal skåpmat" ??
- Webben
  - "Hype" ??

2004-10-25 - SIRNET



Page 5

## Ett steg på vägen?

Från *form* ...

... till *innehåll* ...

... ... till *kunskap*

eller

Kejsarens nya kläder ?

2004-10-25 - SIRNET



Page 6

# Drivkrafter ...

- **Teknik**
- Teknologi mer kraftfull – processor, minne, ...
- **Nätverk** – global kommunikation
- **Allerstädes närvarande** – ubiquitous / ambient intelligence
- **Verksamhet**
- IT-ifieringen – ohejad frammarch
- **Snabbare processer** – JIT
- **Samverkan med andra** – mer och intimare
- **Dynamik i affärsrelationer** – korta samarbeten

→ Informationssamhället !

2004-10-25 - SIRINET

# Någon trend?

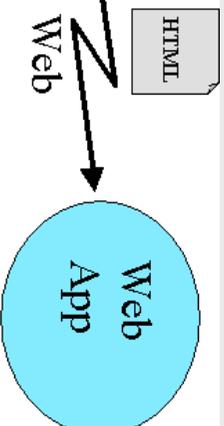
- **Ökad betoning på dynamisk anpassning** – vad som sägs ←→ hur det skall förstås
  - Domännamn
  - Java class-loader
  - URL: logiskt vs fysiskt
  - Web content negotiation
- **Ägande och kontroll**
  - distribuering och decentralisering
- **Globalisering** – språk, kultur, branch
  - Från representation till betydelse

2004-10-25 - SIRINET

# Webben – då & nu

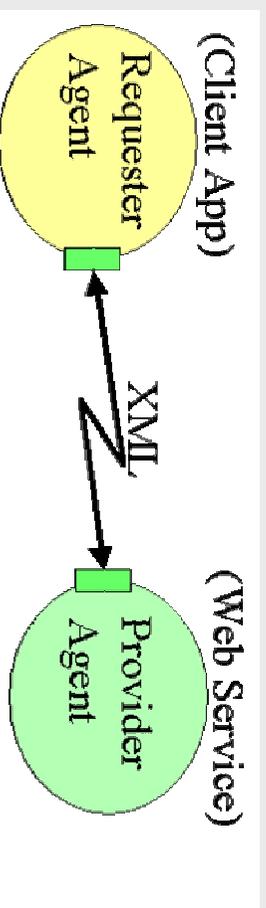


Human User



En separat värld  
icke  
verksamhetskritisk

Normal värld,  
verksamhets-  
kritisk



2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 9

## Hur kan man förstå detta?

- Behov:
  - Att få applikationer att kunna utnyttja data på ett mer informativt sätt
- Ansats:
  - Metadata beskriver data – standardformat, mekaniserbart
- Utmaning:
  - Fungera i verkligheten
- **Steg mot ökad automatik**
  - Bättre än idag

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 10

# Rima Patel, Java technology evangelist at Sun

“... Semantic Web is the next big thing--even beyond Web services. The Semantic Web is an extension of the current Web; information is given well-defined meaning, so that computers and people can cooperate more effectively. In the future, much more information will be immediately available on the Semantic Web than is possible today. Presently, it is in its early stages of development. Major building blocks of Semantic Web-- Ontologies and RDF (Resource Description Framework)- based languages (like OIL and DAML)--are currently being worked on. We are at least a few years away from seeing widespread deployment of the Semantic Web. But the vision is there.

**There seem to be two groups of people, each reacting differently to the Semantic Web: those who say it is too ambitious and too abstract to become a reality, and those, like myself, who believe in the virtues of this technology.**

We see its potential to change the way we present and consume information and do business on the Web--in fact, the very nature of the Web itself.”

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 11

## Stundens ingivelse Några specifika frågor

- Metadata vs. Data
  - För vem? I vems ögon?
- URI
  - Permanent vs. flyktigt
- Modell
  - Lagrad information vs begrepp

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 12

# ”sidebar” – W3C

2004-10-25 - SIRNET



Page 13

# Webbstandardisering – W3C

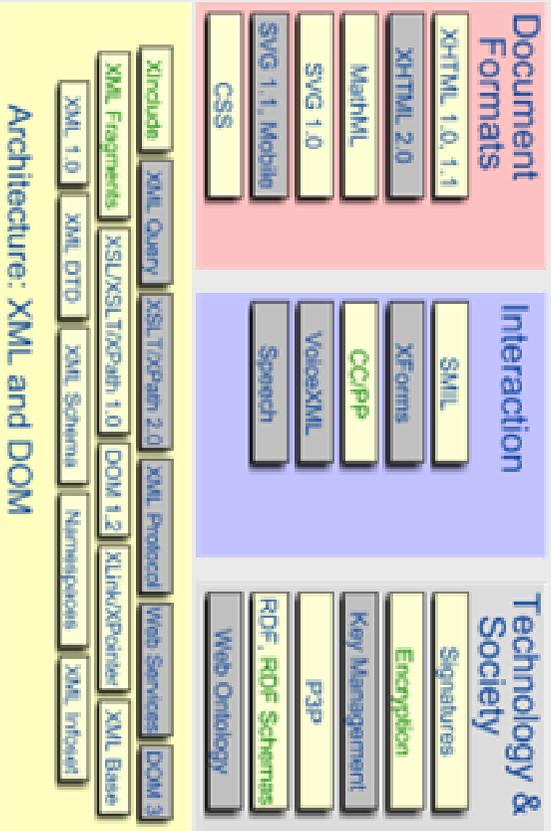
-  World Wide Web Consortium (1994- )
- Industrikonsortium – specificera webbstandarder
- Cirka 400 medlemmar – industri, samhälle, akademi
  - Global täckning
- Tar fram standarder (“W3C Recommendations”) i en konsensusbaserad process
  - HTML, HTTP, XML, .....

2004-10-25 - SIRNET



Page 14

# Webbstandarder



2004-10-25 - SIRNET

# Utveckling

1996	1997	1998	1999	2000	2001	2002	2003	2004
PNG	PICSRules	XML 1.0	CSS 1	ATAG 1.0	MathML 2.0	XML Signature	SVG 1.1	CC/PP
PICS 1.1		MathML 1.0	Namespaces	XML 1.0	Canonical XML	PP 1.0	SVG Mobile	DOM 3
CSS1		CSS 2	WebCGM	DOM2	XML Schemas	JML Canonicalization	XPointer	Infoset(2nd)
		SMIL 1.0	RDF (Old Version)	(X)HTML	Ruby	XPath Filter	SOAP 1.2	Infoset 1.1
		DOM1	WCAG 1.0		XLink 1.0	Decrypt Transform	XForms 1.0	XML 1.0 (3rd)
			Style Sheets PI		XML Base	XML Encryption	XML Events	XML 1.1
			MathML 1.01		SMIL 2.0	UAA 1.0	MathML 2.0	OWL
			XPath 1.0		SMIL Animation		PNG (2nd)	RDF
			XSL T 1.0		SVG 1.0			Voice Framework
					XSL 1.0			
					Infoset			
					WebCGM			

Date: 25-Sep-2004

2004-10-25 - SIRNET

# Webbteknologier

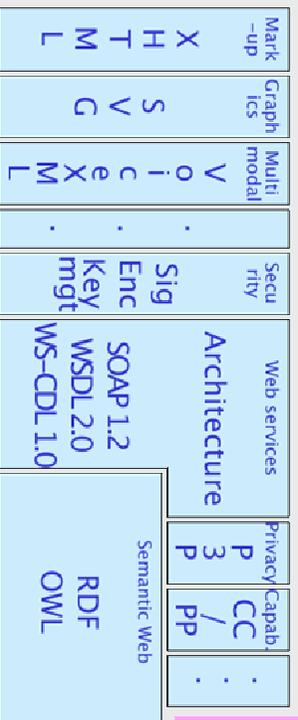
Initial Web



Tomorrow's Web

Higher-level Web standards and applications

Accessibility, Internationalization, Device Independence, QA



XML

(Namespaces, Base, Query, Schema, XSLT, XPath, Xlink, XForms)

HTTP/1.1

URI / IRI

HTML

HTTP

URL

2004-10-25 - SIRNET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 17

Vad som erbjuds  
... teknologier ...

2004-10-25 - SIRNET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret

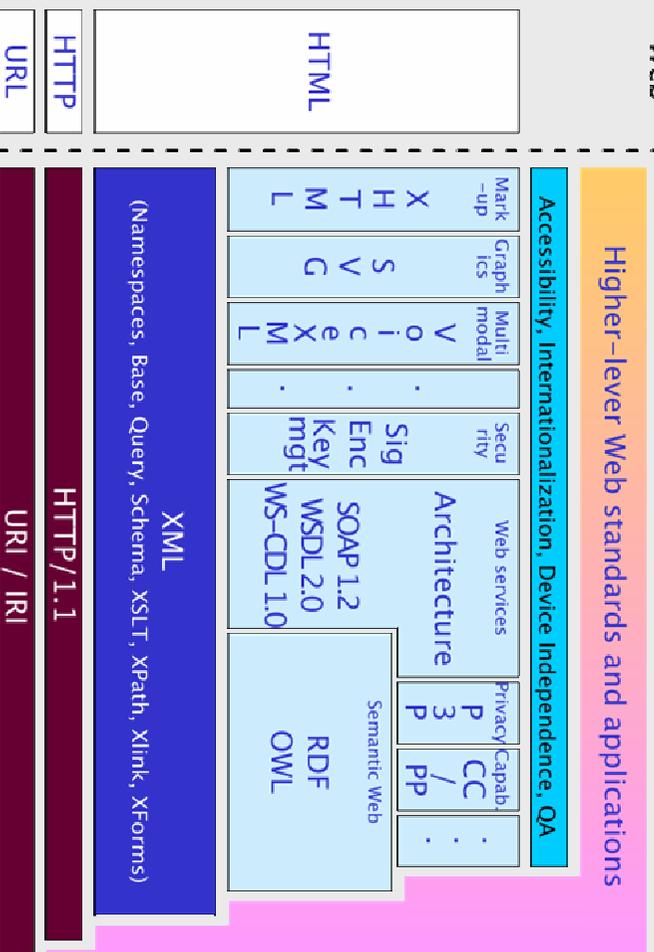


Page 18

# Teknologier

Initial  
Web

→ Tomorrow's Web



2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 19

## Den viktiga detaljen 😊

- Länken – URI / URL / URN
- Kan referera till vad som helst, var som helst, när som helst
- Omvärlden blir en del av oss
  - Vi blir en del av omvärlden
  - Vilket vi alltid varit
  - Utom då sett genom gamla tidens IT-glasögon

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 20

# Babels torn – igen?

- Webbtjänster kommunicerar strukturerad information
  - Representerad i viss begreppsbildning
  - Utnyttjas av producent och konsument
- Begreppsbildning bestämmer terminologi som används i meddelanden
  - Meddelanden uppbyggda av termer
- Webbtjänster populariseras
  - → proliferering av termuppsättningar ??
  - → Babels torn ??

2004-10-25 - SIRINET



Page 21

# Teknologi - RDF

- RDF: "Resource Description Framework"
- ... eller, varför inte, ....
  - "Relational Data Format"
  - "Reusable Data Format"
- "W3C Recommendation"
  - "webbstandard"
  - en komponent i den semantiska webben
- Språk för att uttrycka påståenden om saker:
  - om webb-resurser
  - om vad som helst

2004-10-25 - SIRINET



Page 22

# RDF - tripler

- Data uttrycks som tripler
  - [ subjekt, predikat, objekt ]
  - [ subjekt, relation, objekt ]
- Jämför: Entity / Relationship / Attribute
- Formatet: XML-syntax
- Subjekt, predikat, objekt kan vara URI:er
  - Globalt välbestämda referenser

2004-10-25 - SIRINET

# RDF - exempel

- *Informellt* exempel:
  - Subjekt: <http://www.example.org/foo.html>
  - Relation: <http://purl.org/dc/elements/1.1/creator>
  - Objekt: <http://example.org/staffids#12345>
- Betydelse:
  - "Webbsidan foo.html har skapats av Per Strand"

2004-10-25 - SIRINET

# Refererbarhet

- Vi behöver kunna tala om **saker**
  - Websidor, tjänster, telefoner, personer, ...
  - Konkreta - har existens, dvs är utpekbara
  - Kan identifieras m.h.a. en referens
    - "identifierare" – registreringsnummer, personnummer, ...
- Vi behöver kunna tala om **begrepp**
  - "skapad", "dokument", "person", ...
  - Abstrakta – mentala
  - Traditionellt identifierade med ord - språkberoende
- Identifiera med URI:er

2004-10-25 - SIRNET



Page 25

# Decentraliserad värld

- Behov att dela referenser över organisationsgränser
- Globalt unika identifierare
- Decentraliserat ansvar / ägande
- URI:er
  - Lokalt ägande
  - Lokal förvaltning
  - Global användbarhet

2004-10-25 - SIRNET



Page 26

**<http://purl.org/dc/elements/1.1/creator>**  
**<http://dublincore.org/2003/03/24/dces#creator>**

```
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/creator">
  <rdfls:label xml:lang="en-US">Creator</rdfls:label>
  <rdfls:comment xml:lang="en-US">
    An entity primarily responsible for making the content of the resource.
  </rdfls:comment>
  <dc:description xml:lang="en-US">
    Examples of a Creator include a person, an organisation, or a service.
    Typically, the name of a Creator should be used to indicate the entity.
  </dc:description>
  <rdfls:isDefinedBy rdf:resource="http://purl.org/dc/elements/1.1/">
  <dc:terms:issued>1999-07-02</dc:terms:issued>
  <dc:terms:modified>2002-10-04</dc:terms:modified>
  <dc:type rdf:resource="http://dublincore.org/usage/documents/principles/#element"/>
  <dc:terms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator-004"/>
</rdf:Property>
```

## Ontologier

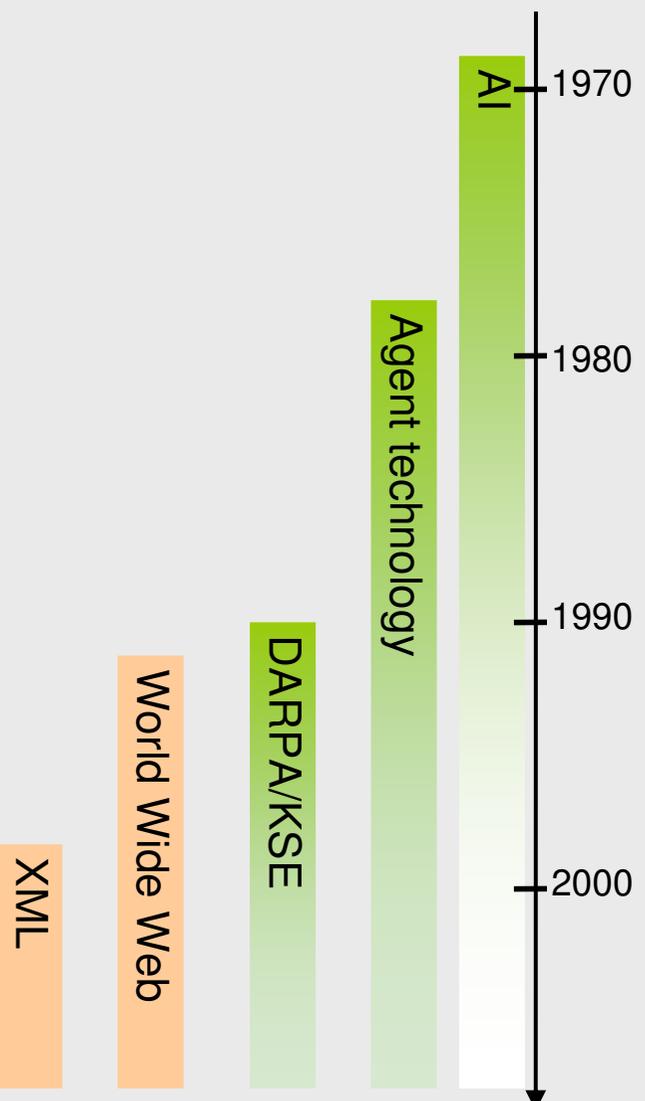
- Begrepp
  - Är identifierbara
  - Har betydelse
  - Kan rationellt användas på vissa sätt
    - Vad vi kan göra med strukturer som refererar till begreppet
    - Vilka slutsatser vi kan dra
- Ontologi = samling av definierade begrepp
  - taxonomi = vokabulär + struktur
  - Ontologi = taxonomi + begränsningsvillkor + regler + ...

# Teknologi - OWL

- OWL - Web Ontology Language
- Språk för representation av begreppsdefinitioner
  - Kombinerbarhet av decentraliserade begreppssamlingar
- Ger ett begreppsorienterat stöd för:
  - Finna, dela, kombinera information
- Metodik är nödvändigt komplement
- Behöver kuggas in i förvaltningsprocessen

2004-10-25 - SIRINET

# OWL - background/1



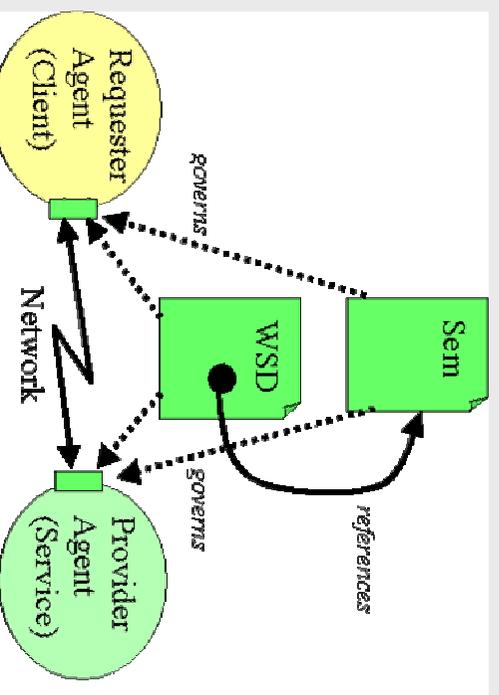
2004-10-25 - SIRINET

# OWL - language levels

- **OWL Lite** - support for classification hierarchies and simple constraints (e.g. simple cardinality constraints). Low complexity ==> high computational efficiency.
- **OWL DL** (Description Logic) - strong expressiveness with reasonable computational efficiency
- **OWL Full** - maximum expressiveness but no guarantees about computational efficiency

2004-10-25 - SIRINET

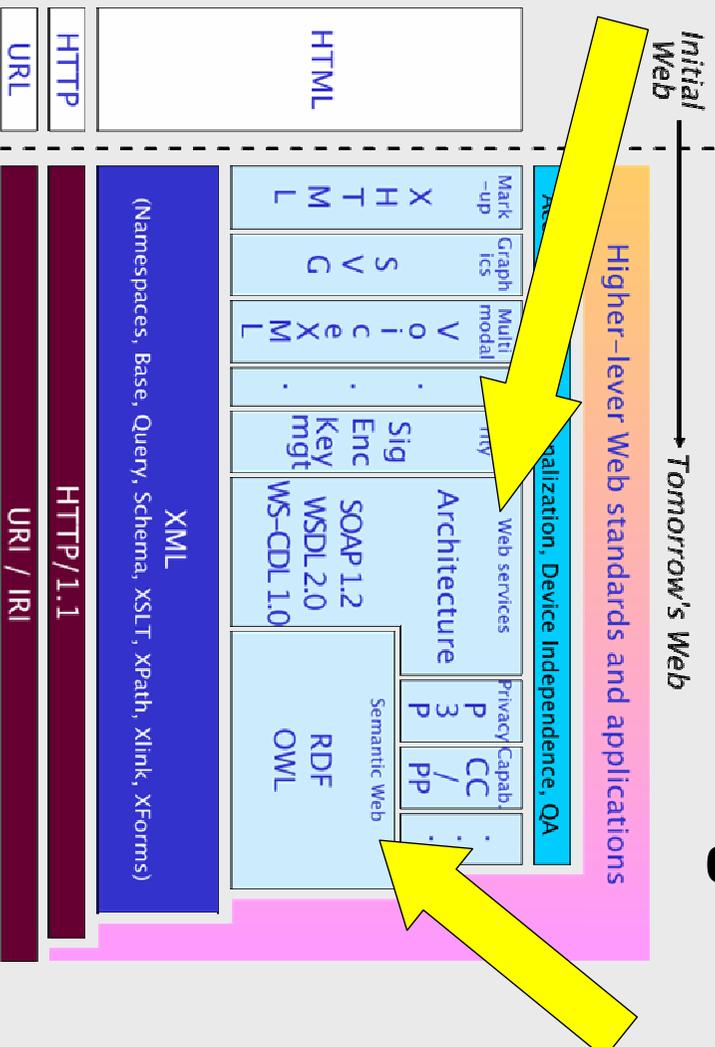
# Webbtjänster & semantik



- Vilket tjänsteinnehåll?
- Separat representation

2004-10-25 - SIRINET

# Semantiska teknologier



2004-10-25 - SIRNET

## Standarder?

2004-10-25 - SIRNET

# Standarder - mångfald

## "stacken av standarder":

- verksamhet
  - affärsprocessmodeller
  - datalager
  - distribuerad bearbetning
  - meddelandestruktur
  - katalogstrukturer
  - datakodning
  - protokoll för transport
  - nätverksprotokoll
  - fysisk konnektivitet
- infrastruktur



divergens



konvergens

# Standarder över tiden

- Kartan över standarder utvecklas över tiden
- För att kunna utnyttja standarder på bästa sätt bör man förstå vad som sker över tiden
- ... vore bra med en modell över standardslagfältet
  - som en ekosystem

# Leverantörer: varför?

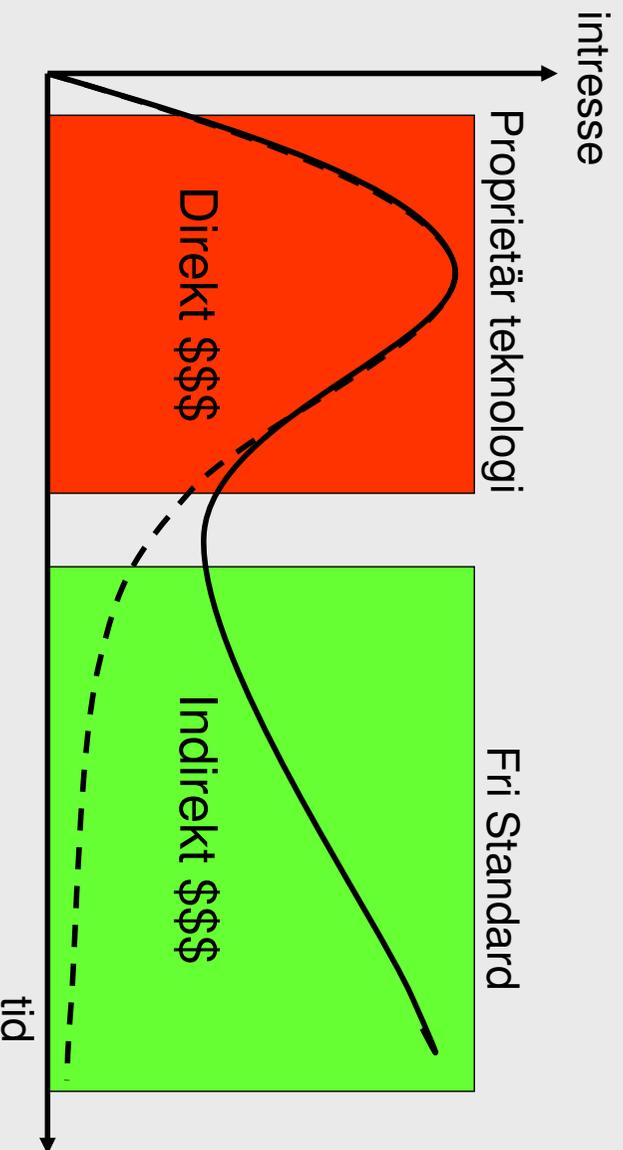
Drivkrafter för leverantörskategorin:

- Vidga kundbasen
- Tillhandahålla "pluggbara" teknologier
- Standardbaserade produkter utvidgas med "features"
- Stabilitet – investering i produktutbudet
- M.m.

*"Värva och binda kunder"*

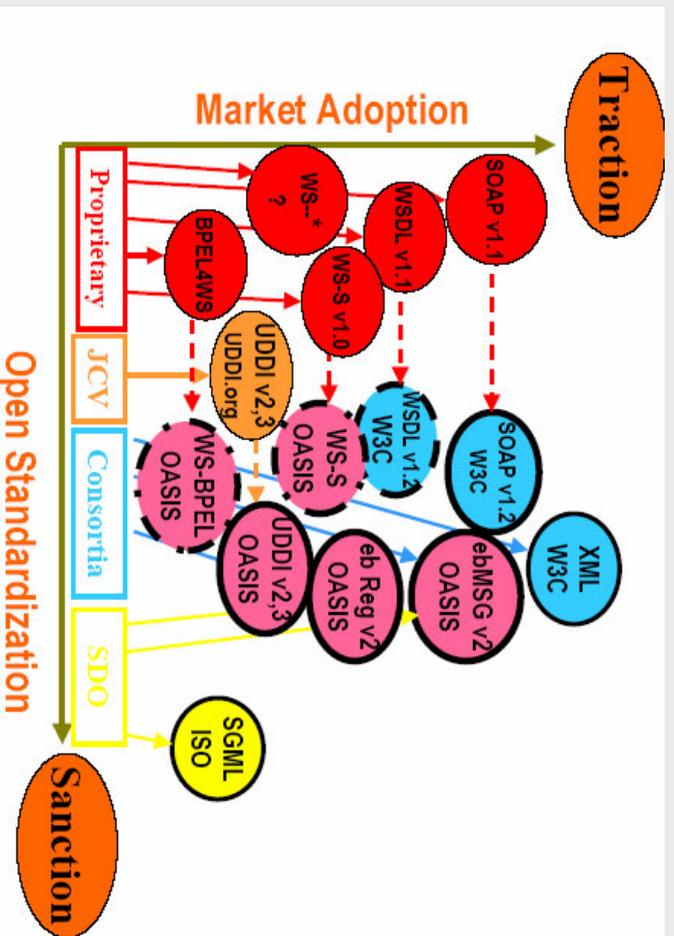
2004-10-25 - SIRINET

# Leverantör: lönsamhet?



2004-10-25 - SIRINET

# Standards livscykel



**Vem använder?  
Till vad?**



# Adobe

## XMP

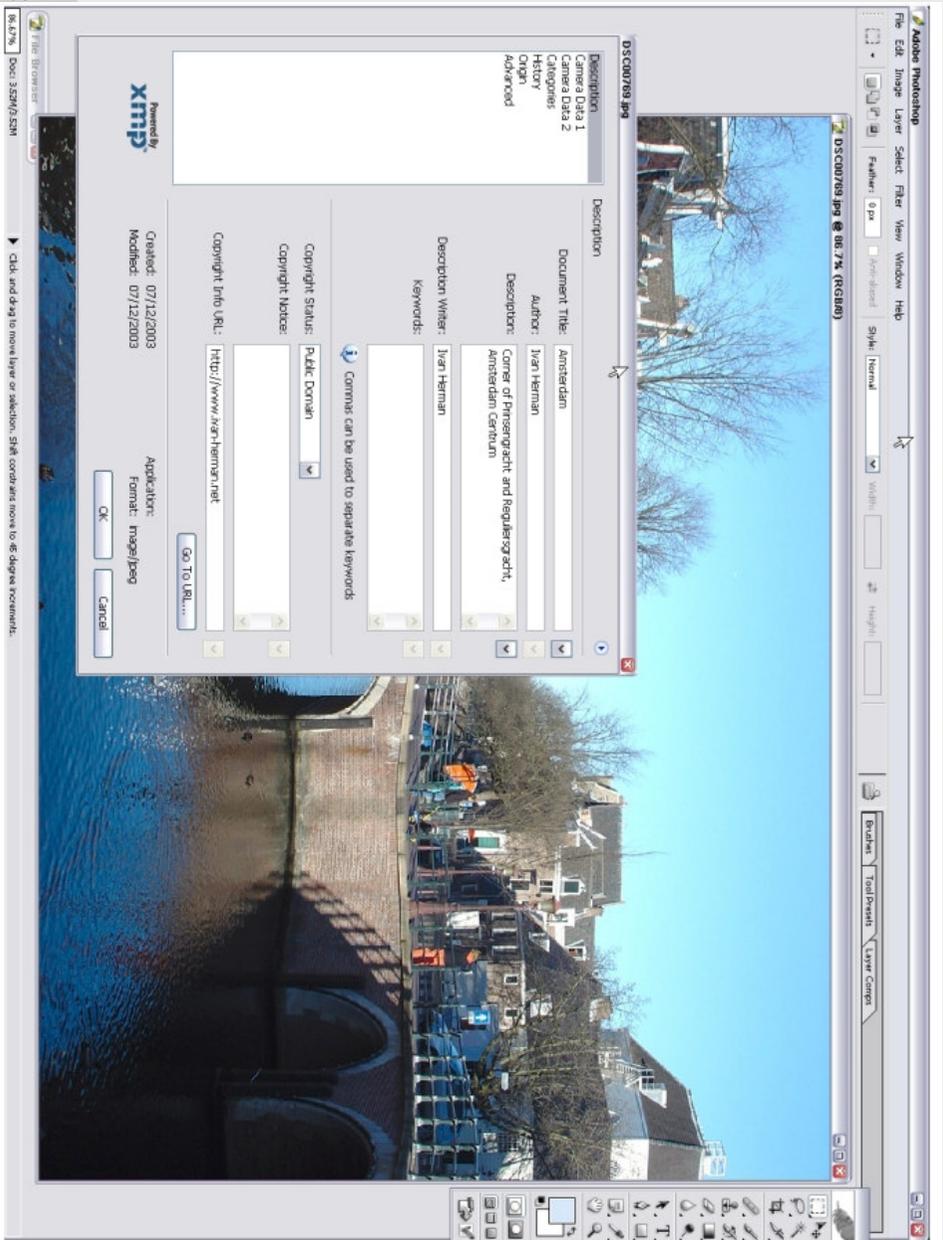
- extensible Metadata Platform (XMP), metadatatverktyg baserat på RDF/XML för att på effektivisera förvaltning av digitala resurser
- knyt RDF-baserad metadata till *alla* deras filformat
  - T.ex. Photoshop i Creative Suite
  - Många miljoner användare av RDF.. utan att veta om det
- Verktyetget allmänt tillgängligt
  - <http://www.adobe.com/products/xmp/main.html>

2004-10-25 - SIRNET

 **WORLD WIDE WEB**  
**C O N S O R T I U M**  
Svenska W3C kontoret



Page 41



2



# Sun

## Swordfish

- Användarstöd för för olika typer av produkter
- Frågor från användare processas av en intern RDF-motor som bl.a. känner till :
  - Sun's White Papers collection (<http://www.sun.com/servers/wp.html> )
  - Sun's System Handbooks collection ([http://sunsolve.sun.com/handbook\\_pub/](http://sunsolve.sun.com/handbook_pub/))

2004-10-25 - SIRNET



# Digital Harbor

## PIE - information integration platform

- "In today's enterprises, there are so many sources of data and so many applications that it is hard for users to 'connect the dots,' so they can make informed business decisions based on a holistic view rather than on one piece of the situation, **The W3C standards for semantics and ontologies represent the next-generation approach to this problem**. As the industry evolves, we look forward to being a key member of such an important industry consortium and working with leading companies to enable businesses to derive maximum value from their infrastructure and technology investments."

Austin Wells, Digital Harbor

2004-10-25 - SIRNET

# Dublin Core

## Dublin Core

- Vokabulärer för distribuerade digitala bibliotek
- En av de första metadatatavokabulärerna uttryckt i RDF
  - URI: <http://www.dublincore.org>
- Utvidgningar definierade (t.ex. PRISM) som stöder spårning av digitala rättigheter (digital rights)

2004-10-25 - SIRINET

Page 45



The Dublin Core Metadata Registry homepage features a blue header with the W3C logo and the text 'WORLD WIDE WEB CONSORTIUM Svenska W3C kontoret'. Below the header is a navigation menu with links: ABOUT THE INITIATIVE, DOCUMENTS, GROUPS, RESOURCES, DCMI NEWS, TOOLS AND SOFTWARE, MEETINGS AND PRESENTATIONS, and PROJECTS. The main content area has a blue background with the text 'Dublin Core Metadata Initiative' and 'Making it easier to find information.' Below this is a section titled 'The Dublin Core Metadata Registry' with a paragraph explaining its purpose: 'The Dublin Core Metadata Initiative's Metadata Registry is an application designed to enable users to explore the DCMI vocabulary in a way that simplifies the discovery and navigation of terms and their definitions, and that illustrates the relationship between terms. The goal of the Registry is to promote the discovery, reuse and extension of existing semantics, and to facilitate the creation of new vocabularies.'

### The Dublin Core Metadata Registry

The Dublin Core Metadata Initiative's Metadata Registry is an application designed to enable users to explore the DCMI vocabulary in a way that simplifies the discovery and navigation of terms and their definitions, and that illustrates the relationship between terms. The goal of the Registry is to promote the discovery, reuse and extension of existing semantics, and to facilitate the creation of new vocabularies.

#### Help

Please select from one of the following supported languages or click on the [Preferences](#) link above for additional options.

**Having trouble** displaying the international fonts? [Click here](#) for help.

\* DCEs-only translations

#### Preferences

Preferences	Search	Administration
<a href="#">العربية الشامية</a> [ar-SA]	<a href="#">Catalan</a> [ca-ES]	
<a href="#">Česky</a> [cs-CZ]	<a href="#">Cymraeg</a> [cy-GB] *	
<a href="#">Dansk</a> [da-DK] *	<a href="#">Deutsch</a> [de-DE]	
<a href="#">Ελληνικά</a> [el-GR]	<a href="#">English</a> [en-US]	
<a href="#">Español</a> [es-ES]	<a href="#">Suomeksi</a> [fi-FI]	
<a href="#">Français</a> [fr-FR]	<a href="#">Italiano</a> [it-IT]	
<a href="#">日本語</a> [ja-JP]	<a href="#">한국어</a> [ko-KR]	
<a href="#">සිංහල</a> [mr-IN]	<a href="#">Norsk</a> [no-NO] *	
<a href="#">Polski</a> [pl-PL]	<a href="#">Português</a> [pt-PT]	
<a href="#">Русский</a> [ru-RU]	<a href="#">Svenska</a> [sv-SE]	
<a href="#">ไทย</a> [th-TH]	<a href="#">Українська</a> [uk-UA]	
<a href="#">简体中文</a> [zh-CN]	<a href="#">繁體中文</a> [zh-TW]	

2004-10-25 - SIRINET

Page 46



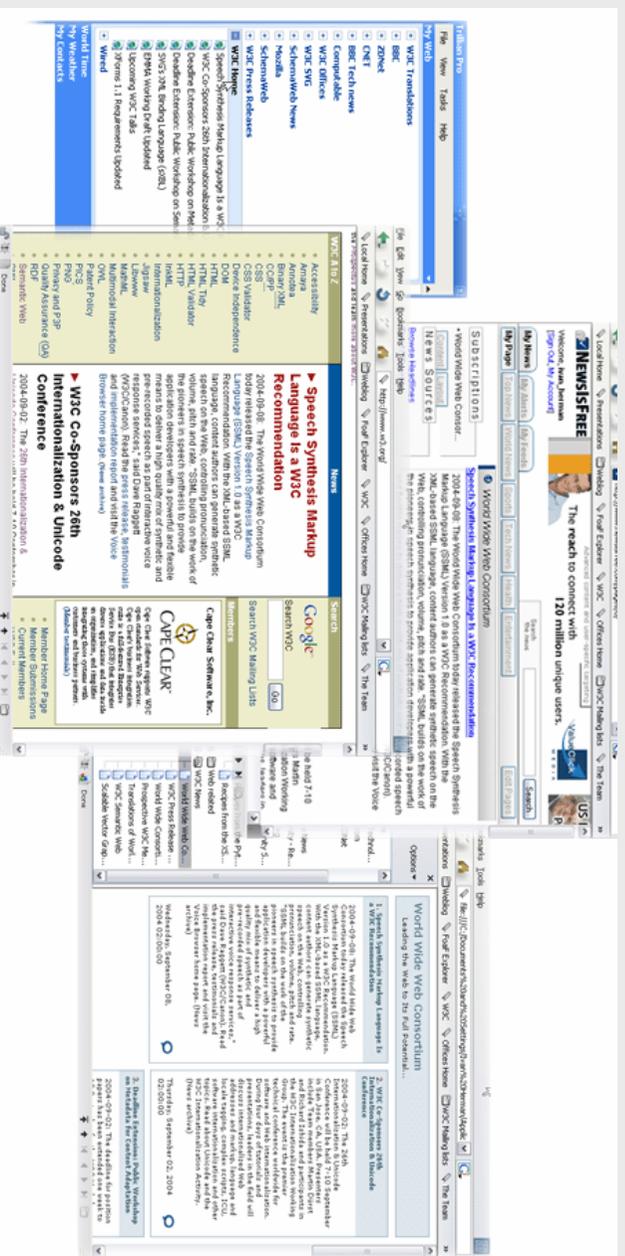
# RSS

## Web Content Syndication (RSS)

- Används för att beskriva viktigt (t.ex. föränderligt) innehåll på webbsidor
- Mycket aktivt område (RSS generationer, IETF...)
- Bred användning i webbloggvärlden
- example: W3C home page syndicated

2004-10-25 - SIRNET

Page 49



The screenshot shows a web browser window displaying the W3C website. The main content area features a 'SUBSCRIPTIONS' section with a heading 'W3C NEWSLETTERS' and a sub-heading 'W3C Co-Sponsors 26th Internationalization & Unicode Conference'. Below this, there are several RSS feed icons and titles, including 'W3C Newsletters', 'W3C Co-Sponsors 26th Internationalization & Unicode Conference', and 'W3C Co-Sponsors 26th Internationalization & Unicode Conference'. The browser's address bar shows the URL 'http://www.w3.org/'. The browser's status bar at the bottom indicates the page title 'W3C Home'.

2004-10-25 - SIRNET

Page 50

# Fler användningar av RDF

## Mozilla

- Intern data lagrad som RDF (t.ex. bokmärken)
  - Mozillas open directory project lagrar data i RDF
    - Använder RDF-filer större än 55MB...
- Creative Commons

- Ett ramverk för att uttrycka rättigheter för digitalt innehåll på webben
    - Juridiska begränsningar refereras från RDF, som finns i sidor
  - Det finns specialiserade webbläsare och plugins som förstår detta
  - Fler än 1,000,000 användare
    - ... som inte ser att de använder RDF...
- Brandsoft
- entreprise Web Management
  - Alla affärsmodeller lagras som RDF
  - Enkelt att sätta upp eget regelverk

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 51

# ”sök och dra slutsatser”

- Sökmotorer baserade på Semantic Web
  - Sökning baseras på komplexa villkor
- Ontologitjänster
  - “ge mig en webbtjänst som drar slutsatser åt mig”
  - (t.ex., om komplex metadata för ontologi)
    - Finn och hantera ekvivalenser
    - Dra logiska slutsatser
    - Kontrollera validitet av modeller
    - etc

2004-10-25 - SIRINET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 52

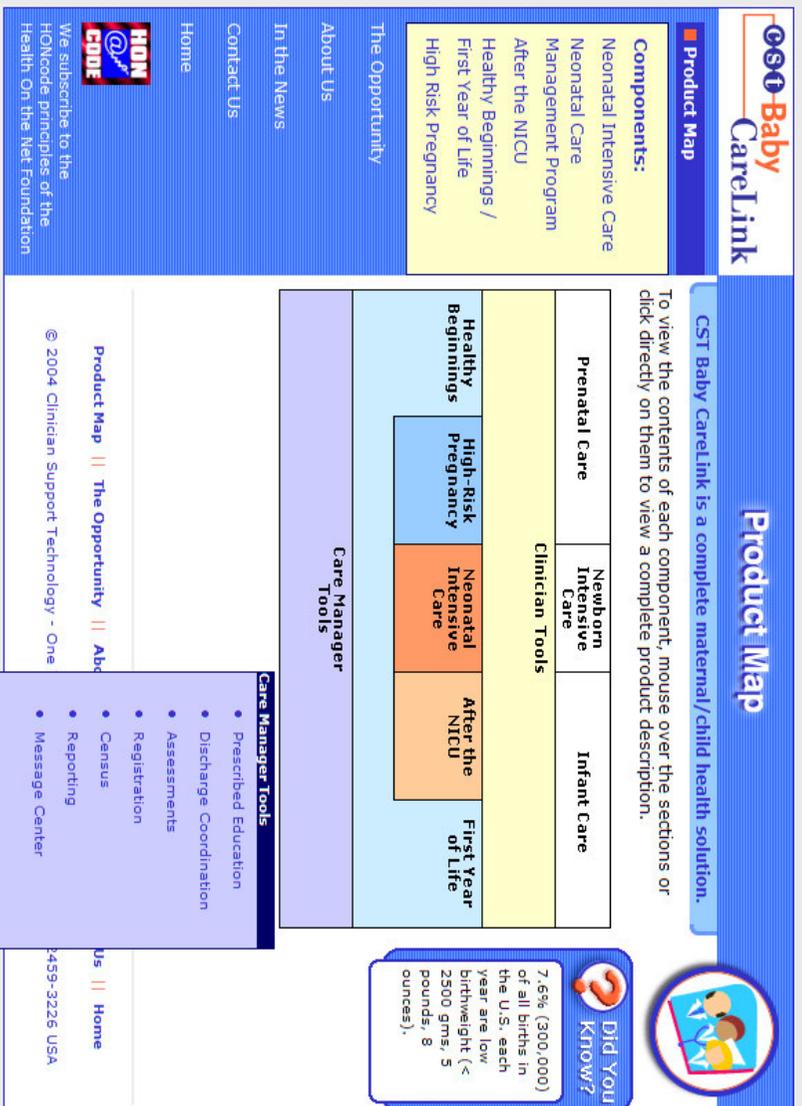
# Baby CareLink

## Baby CareLink

- Informationscentral om behandling för tidigt födda barn
- Erbjuder en OWL-baserad tjänst som en Web Service
  - Kombinerar skilda vokabulärer – medicin, försäkring, ...
  - Användare kan utvidga ontologierna
  - Komplexa frågor kan besvaras via tjänsten

2004-10-25 - SIRINET

Page 53



**CST Baby CareLink**

**Product Map**

CST Baby Carelink is a complete maternal/child health solution.  
To view the contents of each component, mouse over the sections or click directly on them to view a complete product description.

Prenatal Care	Newborn Intensive Care	Infant Care
Clinician Tools		
Healthy Beginnings	High-Risk Pregnancy	Neonatal Intensive Care
		After the NICU
		First Year of Life

Care Manager Tools

**Care Manager Tools**

- Prescribed Education
- Discharge Coordination
- Assessments
- Registration
- Census
- Reporting
- Message Center

**Did You Know?**  
7.6% (300,000) of all births in the U.S. each year are low birthweight (< 2500 gms; 5 pounds, 8 ounces).

**Product Map** || The Opportunity || About Us || Home

© 2004 Clinician Support Technology - One Health On the Net Foundation

2004-10-25 - SIRINET

Page 54



## HP Labs Semantic Web Research

“The HP Labs Semantic Web research group believes that the Semantic Web represents a huge potential technology disrupter, enabling new and more flexible approaches to data integration, web services, knowledge discovery etc. The group is committed to furthering the Semantic Web vision through:

- standards work via the W3C to support the growth of the Semantic Web;
- tools development to encourage the exploration and exploitation of the Semantic Web by developers;
- core research to help develop the field;
- applications research to demonstrate the value of Semantic Web;
- consultancy within HP to promote its use within the company.”

2004-10-25 - SIRNET



WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 55



## IBM

### *Semtag and seeker*

- Verktyg för automatstöd semantisk annotering

2004-10-25 - SIRNET



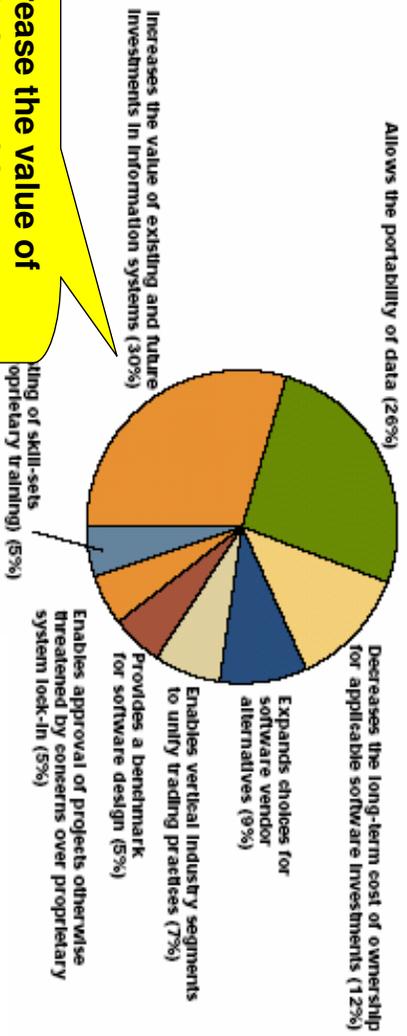
WORLD WIDE WEB  
CONSORTIUM  
Svenska W3C kontoret



Page 56

# Användarnas fördelar

Allow portability of data (26%)



Increase the value of existing and future investments in information systems (30%)

(c) 2003 Delphi Group

# Budskapet

- Möjligheter
  - Interoperabilitet mellan heterogena datakällor
  - Systematisk metod
  - Standardiserad teknologi
- Problem
  - Hur stor omfattning krävs för att få fördelar?
  - Metodik? Förvaltning?
  - Branchspecifika ontologier?

- “Webben” är en dominerande visionär drivkraft:
  - Standardisering
  - Distribuering
  - Decentralisering
- Semantiska teknologier finns – och blir allt viktigare
  - Nyckelteknologi för vissa affärsbehov
    - Internt
    - externt
- Starta nu!

# Frågor?

2004-10-25 - SIRNET



Page 61