Metadata tillämpningar i företag

"Ontologier i arbete - Språket som grund för framtida IT-utveckling"
31 augusti 2004, Electrum, Kista

Kerstin Forsberg
Senior Information Architect
Medical Informatics
AstraZeneca
Mailto:kerstin.l.forsberg@astrazeneca.com
Public homepage: http://www.viktoria.se/~kerstinf/

AstraZeneca

Metadata tillämpningar i företag

- Lite bakgrund och några tankeväckare
- Dublin Core
- EU metadata standard projekt
- ”Contexts”
- Topic Maps
- Avslutande funderingar

AstraZeneca

Våra forskningsområden

- Mage/tarm
- Hjärta/kärl
- Onkologi
- Andningsvägar
- Smärtlindring
- Centrala nervsystemet
- Infektion

AstraZeneca

Morgondagens läkemedel

Diprivan®
Losec®
Plendil®
Tenor
Pulmicort®
Xylocain®

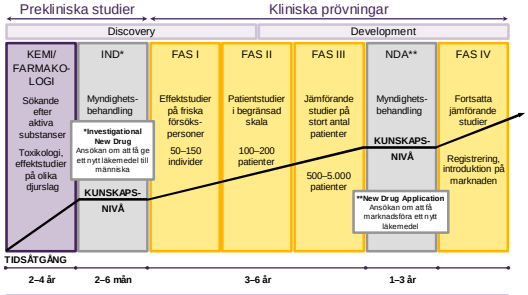
Exanta™
Iressa™
Crestor™
Faslodex™
Symbicort™
Nexium™

4

5

AstraZeneca

Forskningsprocessen

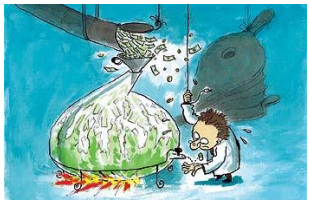


Cirka 10 år från idé till färdigt läkemedel

6

AstraZeneca

”Information Management” inom läkemedels industrin



“… the industry has not yet learned to make best use of the tools it already has, such as ways to share information across the various businesses”.

The Economist July 2002

7

AstraZeneca

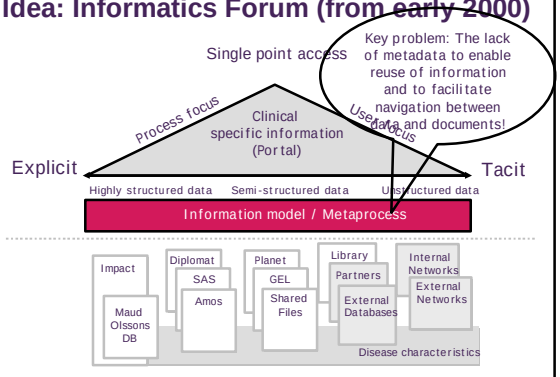
Medical Informatics: Vision

Enhance creativity, facilitate decision making and increase efficiency by enabling clinical researchers to exploit scientific information globally, and to enlarge personal networks.

8

AstraZeneca

Idea: Informatics Forum (from ~~early~~ 2000)



Några tankeväckare …

10

AstraZeneca

Några tankeväckare …

A semantic search for "Apolipoprotein A-I" will result in:

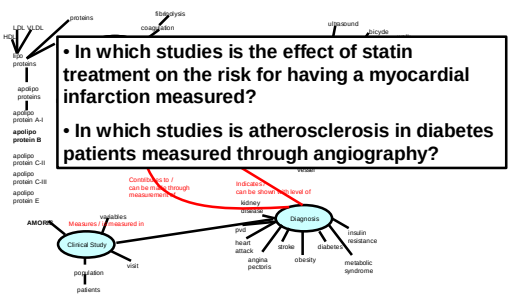
- A list of bibliographic data from Planet
- A list of clinical studies in which this is one type of captured lab data.
- Presentations material

For more information:

- A list to related communities is provided.
- One of the is the community set-up by professor Göran Walldius to share information about risk markers for death and myocardial infarctions related to abnormal lipids.

11

Några tankeväckare



12

AstraZeneca

Dublin Core, metadata element från 1995, ISO standard 2003

Content

- **Title:** A name given to the resource.
- **Subject:** The topic of the content of the resource
- **Description:** An account of the content of the resource.
- **Type:** The nature or genre of the content of the resource.
- **Source:** A Reference to a resource from which the present resource is derived.
- **Relation:** A reference to a related resource.
- **Coverage:** The extent or scope of the content of the resource

Intellectual property

- **Creator:** An entity primarily responsible for making the content of the resource.
- **Publisher:** An entity responsible for making the resource available.
- **Contributor:** An entity responsible for making contributions to the content of the resource
- **Rights:** Information about rights held in and over the resource

Instantiation

- **Date:** A date associated with an event in the life cycle of the resource.
- **Format:** The physical or digital manifestation of the resource
- **Identifier:** An unambiguous reference to the resource within a given context.
- **Language:** A language of the intellectual content of the resource.

13

AstraZeneca

Dublin Core metadata in corporate environments

Work Item 3
CEN, European Committee for Standardization, Workshop on Dublin Core Metadata

Objective: Identify and address corporate metadata needs to support modern business organizational functions like internal content reuse, federated search, & knowledge management.

14

AstraZeneca

Dublin Core metadata in corporate environments

- Vi ställer samman exempel på och erfarenheter av DC användning i företag:
  - Vilka specifika riktlinjer behövs för olika DC.element?
  - Hur DC elementen utvidgats?
  - Vilka verktyg som används för att skapa och underhålla metadata?
  - Vilka s.k. kontrollerade vokabulärer som används för metadata värden?

15

AstraZeneca

What specific guidelines are needed for the use of specific DC.elements?

- ***dc.subject*** for classifications of corporate products and services
- ***dc.publisher*** and ***dc.contributor*** for representing types of corporate roles
- ***dc.type*** for corporate content types
- ***dc.date*** for managing content lifecycle

16

AstraZeneca

Dublin Core, metadata element tillämpad på Volvo, slutet på 90-talet

- "These problems are a consequence of trying to describe information resources without taking into account the *context in which end users create and consume information*."

"Extensible use of RDF in a business context",
K. Forsberg and L. Dannstedt, presented at W3C conference in Amsterdam 2000

17

AstraZeneca

Metadata, olika perspektiv i ett företags sammanhang

- **Tekniska perspektivet**
  Detaljerar hur och var information är lagrad och dess accessmetod:
  T.ex. format, size, identifier, source
- **Producent perspektivet**
  Metadata som beskriver informations produktion och hanterande.
  "Micro-processen" dvs lifecykeln för den "container" som håller samman innehåller = document management
  T.ex. Author, Publish date, Version
- **Nuvarande sammanhang (administrative och vetenskapligt kontext)**
  Säkra att informationen relateras till ett sammanhang "här och nu" m'ha metadata värden från en "statisk" taxonomy
  T.ex. Project, Drug, Therapy Area
- **Evolution, nya sammanhang (administrative och vetenskapligt kontext)**
  Administrativa och vetenskapliga sammanhang förändras.
  Metadatan:s livscykel, "dynamisk" taxonomier.
  "Macro-processen"i verksamheten,
  T.ex. Compound becomes several Brands

18

AstraZeneca

”Context shapes content. The word context comes from Latin cum (with) and texere (to weave) and etymologically suggest a process of weaving together.”

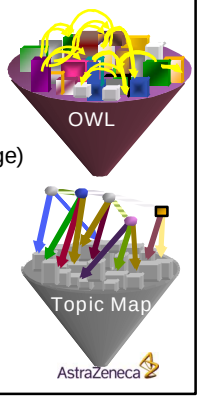
From John Seely Brown and Paul Duguid (2000), “The Social Life of Information”

19

AstraZeneca

Emerging IM standards

- Capturing contexts, encoding ontologies and enabling information navigation
  - OWL (Web Ontology Language)
  - Topic Maps



20

AstraZeneca

Topic Maps

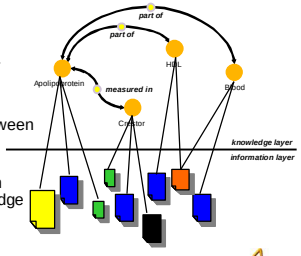
- ISO standard for organising, retrieving and navigating information resources
- Some of the key areas of application of Topic Maps are:
  - making information easier to find
  - connecting information that originates from multiple sources
  - capturing and sharing knowledge
  - integrating disparate applications
- Web Standard (XML Topic Maps 1.0, 2001)
  - XML version for use on the Web

21

AstraZeneca

Topic Maps
Summary of Core Topic Maps Concepts

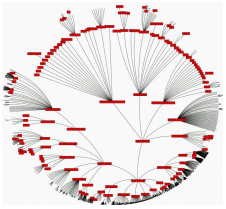
- A pool of information
  any type or format
- A knowledge layer
- Topics
  a set of knowledge topics for the domain in question
- Associations
  expressing relationships between knowledge topics
- Occurrences
  information that is relevant in some way to a given knowledge topic
- = The TAO of Topic Maps



AstraZeneca

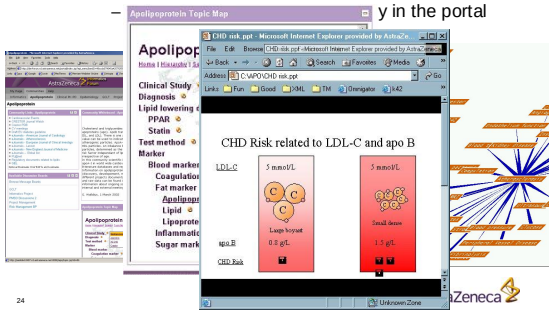
Topic Maps
Cases

- Autogeneration of Topic Maps from existing sources
  - PL@net thesaurus to publish as common taxonomies



Topic Maps
Cases

- Apolipoprotein research community
  - y in the portal



AstraZeneca

Några funderingar

- Identiteter!!!
  - Published Subject Indicators (PSIs)
    http://psi.astrazeneca.net/brands/#Crestor
    http://psi.astrazeneca.net/azglossary/#SubjectID
  - Life Science Identifier
    URN:LSID:ncbi.nlm.nih.gov:GenBank.accession:NT_001063:2
- Metadata Registries (ISO11179)
- Information Type Registry

25

AstraZeneca

Referenser

- Dublin Core
  - ISO/IEC 15836-2003,
    http://www.niso.org/international/SC4/n515.pdf
  - Dublin Core Community, http://dublincore.org
  - EU standard arbete baserat på Dublin Core, CEN Workshop Agreement on Guidance for the deployment of Dublin Core metadata
    http://www.cenorm.be/cenorm/businessdomains/businessdomains/isss/activity/wsmmi.asp
- Erfarenheter av metadata tillämpningar på Volvo
  - K. Forsberg and L. Dannstedt. "Extensible use of RDF in a business context." 33:1-6 Computer Networks Issues (June 2000) 347-364. Presented at the 9th International World Wide Web Conference, Amsterdam, Netherlands, (2000)
    http://www9.org/w9cdrom/323/323.html
- Topic Maps, en ontologi standard, utvärderad på AstraZeneca
  - "An evaluation of Topic Maps", Anna Carlstedt Mats Nordborg, Datalingvistik, Göteborgs Universitet
    http://www.ling.gu.se/~a8matsni/exjobb.htm

26

AstraZeneca