

Web Science: The Digital Heritage Case

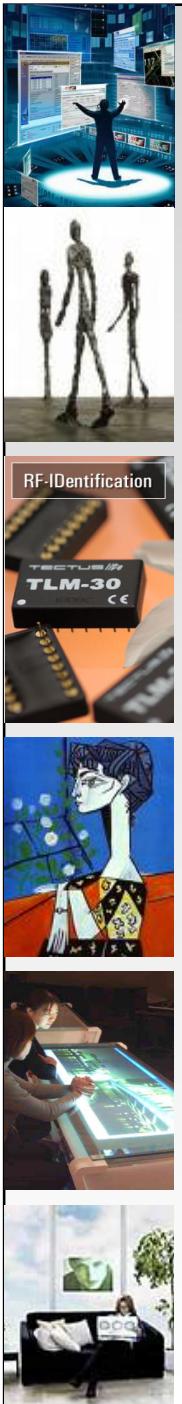
Guus Schreiber

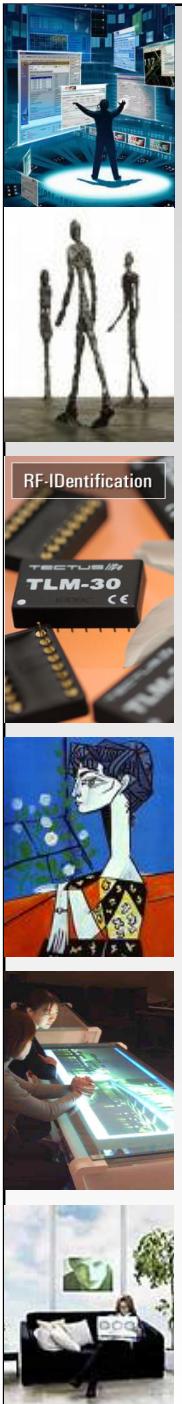
Informatics Dept., Web & Media

VU University Amsterdam

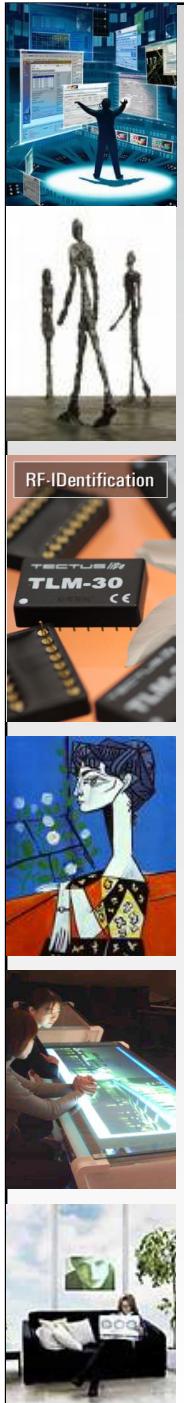
Acknowledgements

- This talk represents work of many people, in particular colleagues at the VU and of the Web Science Trust. Lora Aroyo, Nigel Shadbolt, Jacco van Ossenbruggen and Michiel Hildebrand provided slides



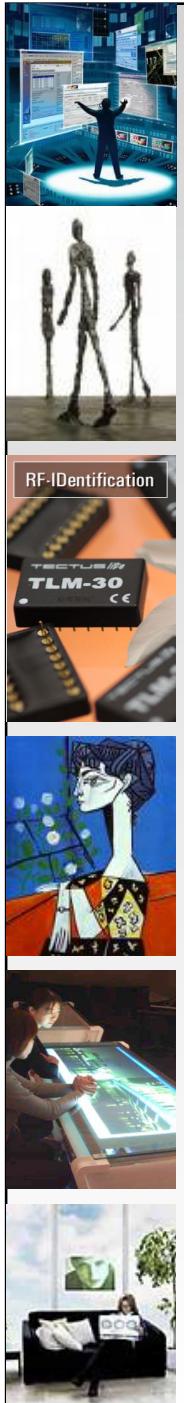


WHAT IS WEB SCIENCE?



21st century computer-science landscape

- Nowadays small (student) projects are achieving more than large multi-year multi-person projects 10 years ago
- Data storage and computing power are no longer an omnipresent problem
- Globalization of data and service availability
- Data and applications are moving to the Web

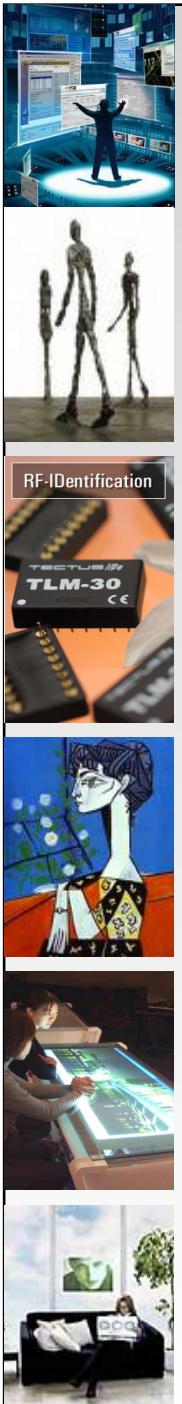
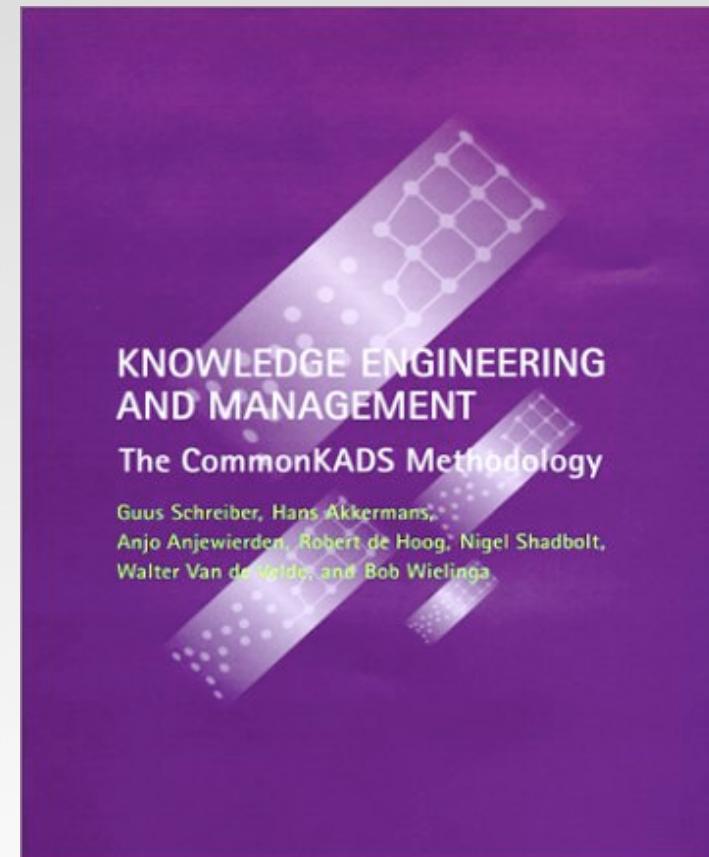


Web science

- Web science is not computer science transported to the Web
- Web science is **socially embedded**
- Broad scope of research issues: trust, reputation, security, governance, social networks, economic models
- “Shift from studying chips to studying clicks”

My journey knowledge engineering

- design patterns for problem solving
- methodology for knowledge systems
- models of domain knowledge
- ontology engineering





My journey access to digital heritage

E-Culture MultimediaN *cultural heritage search*

[Login](#) | [help](#) | [English](#) ▾

[search](#) [browse](#) [local view](#) [annotate](#)

This cultural search engine will give you access to artworks from several museum collections. Type a keyword, for example: Derain, calligraphy, or 1867.

Collections

Artchive.com (>3,000 objects)	Rijksmuseum.nl (>16,000 objects)	RMV.nl (> 10,000 objects)	KIT.nl (>78,000 objects)	Bibliopolis.nl (>1,600 objects)

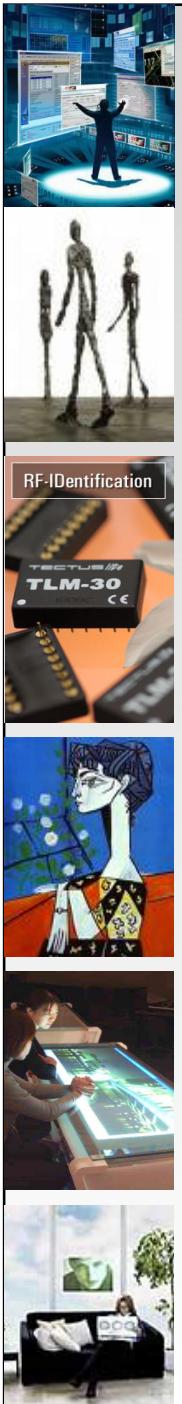
Vocabularies and thesauri

Getty AAT (>31.000)	Getty ULAN (>130.000)	Getty TGN (>890.000)	SVCN (Dutch ethnology, >11.000)	Princeton Wordnet (>115.000)

My journey Web standards



- OWL Web Ontology Language
- SKOS model for publishing vocabularies on the Web



A definition of Web science

“Web science is the study of

- (i) the **social behavior** in the Web at the personal, organizational and societal level,
- (ii) the Web **technology that enables** this behavior, and
- (iii) the **interactions** between technology and behavior”



Good introduction: Shneiderman, Comm. ACM, 2007

Web Science: A Provocative Invitation to Computer Science

Here's how it can awaken computer science to the interdisciplinary possibilities of the Web's socially embedded computing technology.



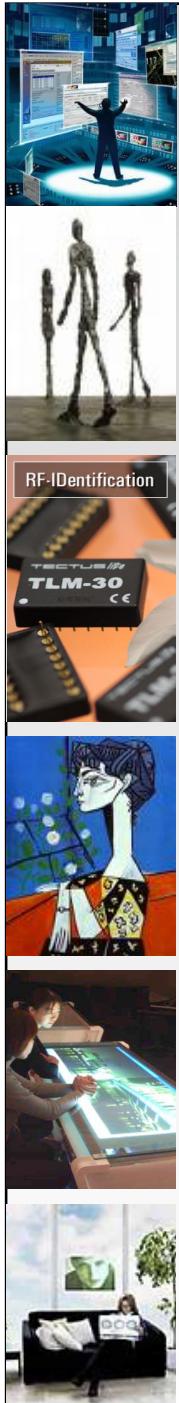
Much more than a provocative term, “Web science” signals a new way of thinking about computer science. Computer science researchers and professionals are familiar with the turbulence of technology innovation; Web science promoters challenge them to

But Web science advocates make clear that understanding Google’s technology and business success requires more than a discussion of Web crawling and distributed search algorithms. They insist something bigger and bolder is becoming important and will take an open mind for the traditional computer science community to absorb.

SOCIALLY EMBEDDED



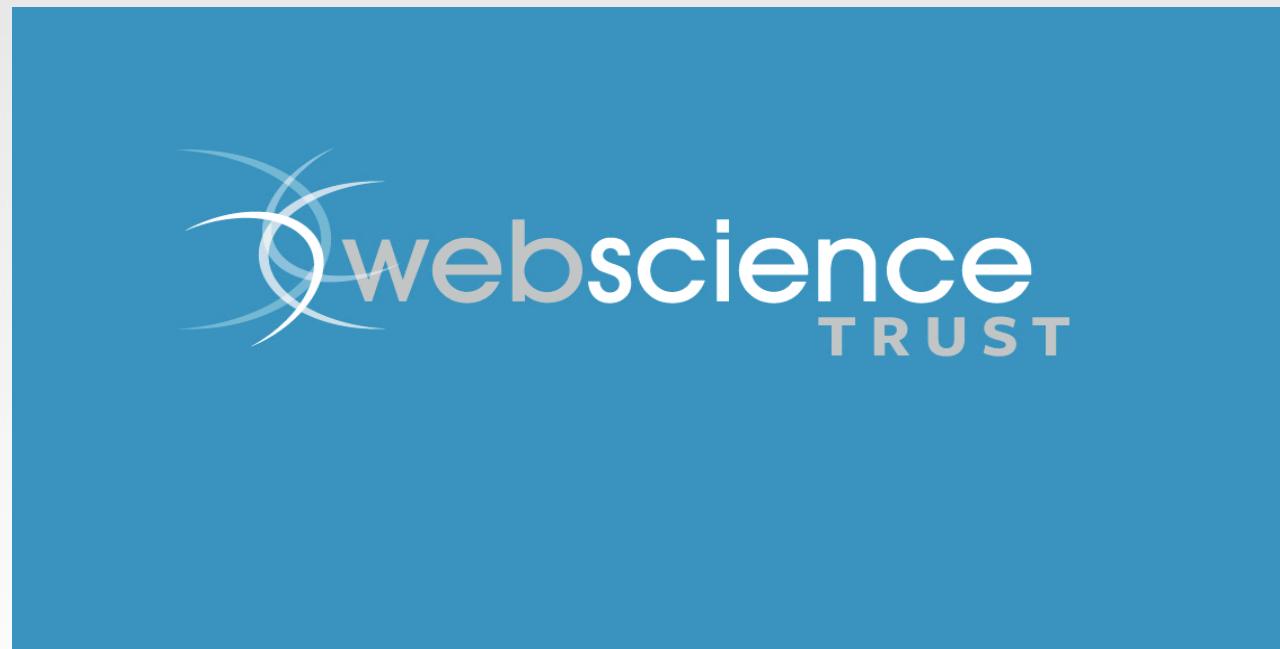
Computer Science	Web Science
Metrics	
Moore's Law	Page views
Order (n) algorithm analysis	Unique visitors/month
Gigabytes	Number of songs or videos
Topics	
Computer networks	Social networks
Packet switching	Voice over IP, music sharing
Information	Relationships
Programming languages	Wikis, blogs, tagging
Databases, operating systems, compilers	E-commerce, e-learning, e-government, medical informatics, financial analysis
3D graphics, rendering algorithms, computational geometry, object modeling	Creating and sharing video, animation, music, photos, maps
Focus	
Technology	Applications
Computers	Users
Supercomputers	Mobile devices
Proficient programmers	Universal usability



Web Science Trust

<http://webscience.org>

Network of academics promoting Web
Science





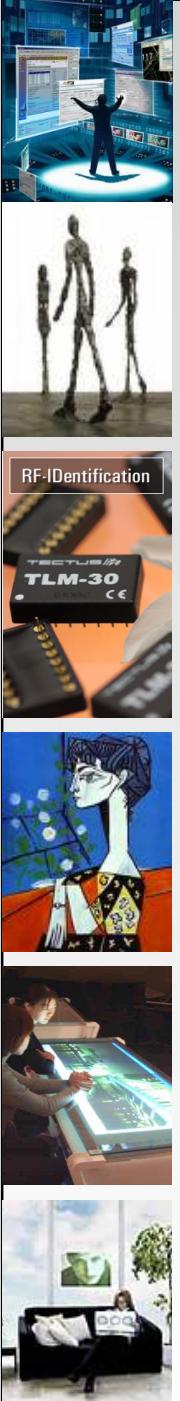
We need to study the Web as a phenomenon

- Is the Web changing faster than our ability to observe it?
- How to measure or instrument the Web?
- How to identify behaviors and patterns ?
- How to analyze the changing structure of the Web?

Sample research themes:

- Web dynamics
- Collective intelligence
- Privacy, trust and security
- Linked open data





Web science issues: social computing

- What makes online communities successful?
 - Role of moderation
 - Issues of trust and identity
 - Growth models
 - Role of type of discussion topic
 - Type of language used
- Driving role of technology



kassa



RF-IDentification



VRAAG & BEANTWOORD

U bevindt zich in het interactieve gedeelte van de Kassa-site. Dit is de plek waar u vragen kunt stellen, maar ook vragen binnen uw vakgebied kunt beantwoorden.

Klik op één van de onderstaande categorieën en subcategorieën om vragen te vinden die zijn gesteld of beantwoord.

U kunt hier ook zelf een vraag stellen, of een vraag zoeken via onze zoekmachine. Ook kunt u als deskundige iemand helpen en tegelijkertijd zelf uw problemen plaatsen. Om Vraag & Beantwoord leuk te houden hebben we een **aantal spelregels** opgesteld. Lees ze [HIER!](#)

Een overzicht van onze deskundigen (experts) vindt u [hier](#).

Zoek hier uitsluitend in Vraag & Beantwoord

alles

zoek

Zoek een vraag in een categorie en/of subcategorie

Auto & Vervoer

Kies subcategorie

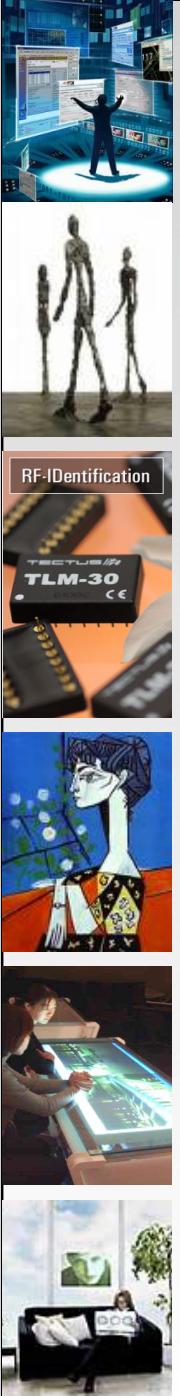
ok

[stel hier zelf een vraag](#)

KIES EEN CATEGORIE

Klik op een categorie en u krijgt een overzicht te zien van de subcategorieën die erbij horen en de laatste vragen die zijn gesteld en/of beantwoord.

- [➤ Auto & Vervoer \(2060 vragen\)](#)
- [➤ Computers & Internet \(10358 vragen\)](#)
- [➤ Elektronica & Telefonie \(4969 vragen\)](#)
- [➤ Geldzaken & Recht \(8614 vragen\)](#)
- [➤ Gezin & Gezondheid \(1672 vragen\)](#)
- [➤ Huis & Tuin \(4434 vragen\)](#)
- [➤ Levensmiddelen & Voeding \(515 vragen\)](#)
- [➤ Muziek & Hobby's \(583 vragen\)](#)
- [➤ Vakantie & Vrije tijd \(1090 vragen\)](#)



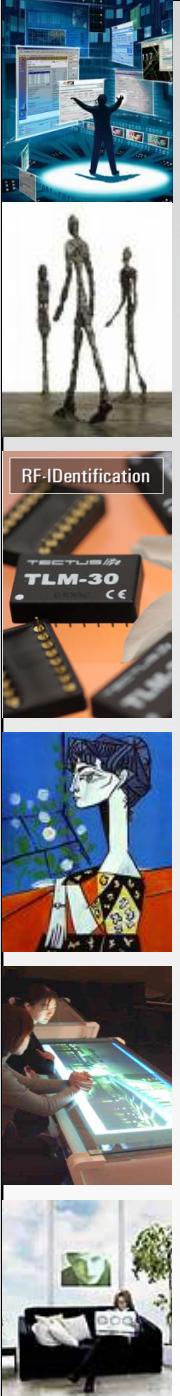
Web science issues: psychology and pedagogy

- Effect of chat/email on:
 - Mental development of children
 - Forming of relationships
 - Changes in cultural preferences
- Distance learning
- Distance psychotherapy

Web science issues: privacy

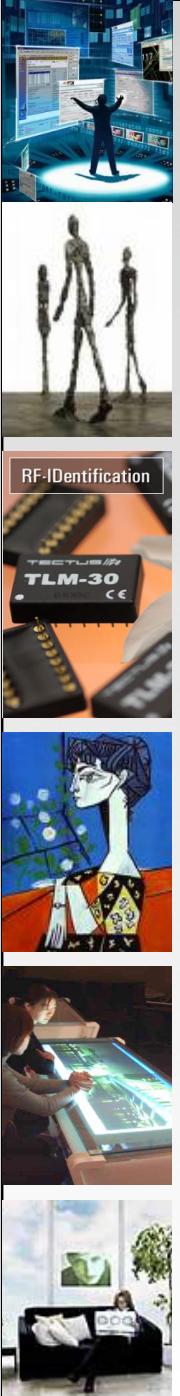
- Control over personal data
 - Video's and pictures of you
- Identity theft
- Who is allowed to store what?
- What do you accept as a user?





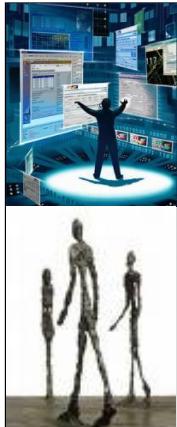
Web science issues: legal problems

- Lack of “location”
 - Which law to apply?
- Copyright
 - New types of licenses required
 - See for example Creative Commons
- Data aggregation: what can Google/
Facebook/... do with our data?



Web science issues: new economic models

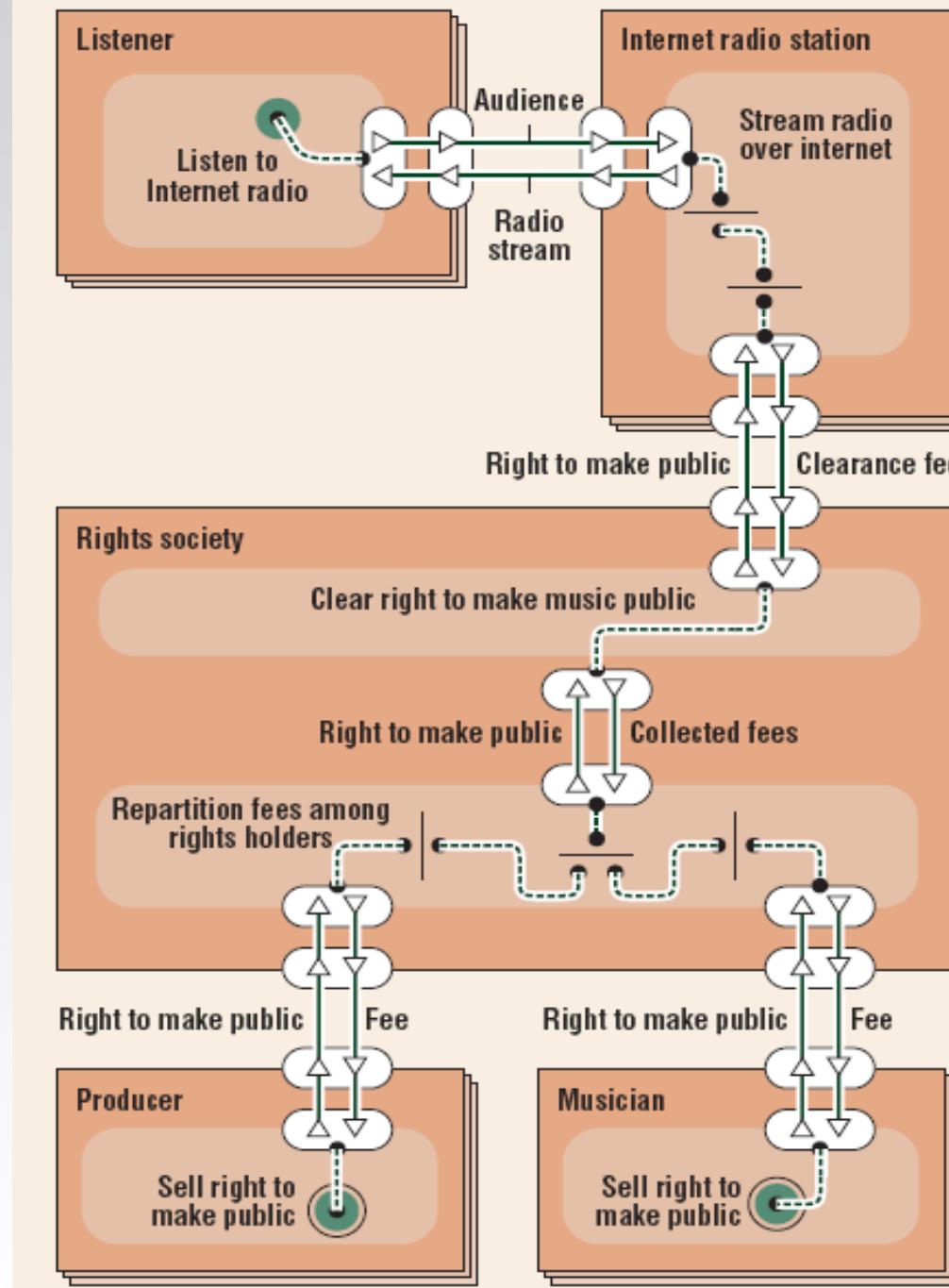
- Micro payments for mail messages?!
- Downloading material with copyright
 - What is the price of a song?
- Global personalized services replace shops

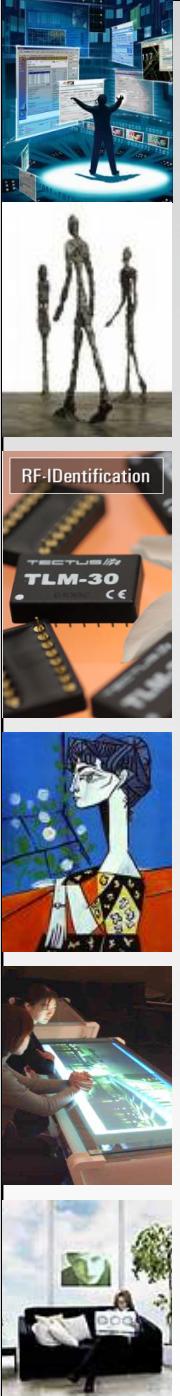


RF-IDentification



New economic models for distribution of fees for music rights





Web science issues: universal access

- Ideals: access for all!
- Limited Web access in particular countries
- Spreading of hate
- Quality of material gathered on a global scale
 - Wikipedia



Web for Social Development





Search: Advanced search

[Today on CNET](#) [Reviews](#) [News](#) [Downloads](#) [Tips & Tricks](#) [CNET TV](#) [Compare Prices](#) [Blogs](#)

[Business Tech](#) | [Cutting Edge](#) | [Green Tech](#) | [Access](#) | [Security](#) | [Media 2.0](#) | [Markets](#) | [Personal Tech](#) |
[News Blogs](#) | [Video](#) | [Extra](#) | [My News](#)

Brewster Kahle's modest mission: Archiving everything

By [Elinor Mills](#)

Staff Writer, CNET News.com

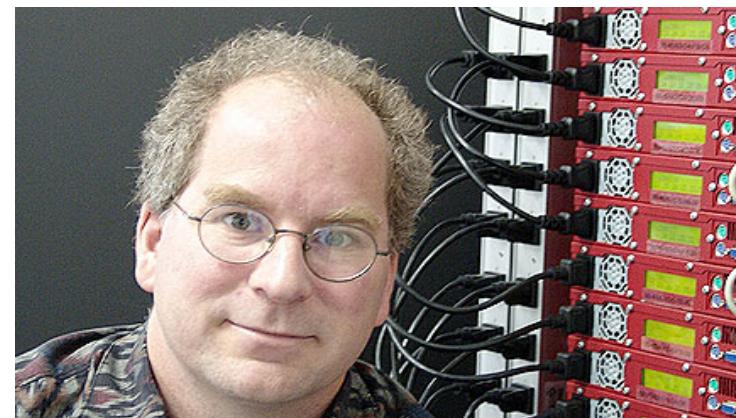
Published: June 23, 2006, 4:00 AM PDT

[TalkBack](#) [E-mail](#) [Print](#) [del.icio.us](#)

[Digg this](#)

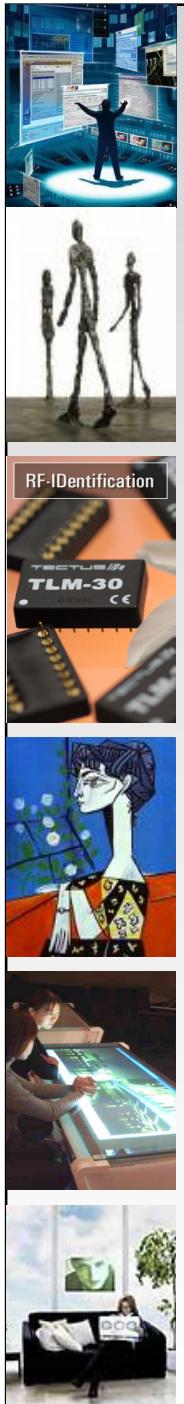
Brewster Kahle is on a mission. He wants the whole planet to have access to human knowledge. All human knowledge. And he's striving to make that possible--one byte at a time.

Ten years ago, Kahle founded the nonprofit Internet Archive, with the goal of preserving the hitherto ephemeral pleasures of the Net for posterity. But, unsatisfied with limiting himself



Featured gallery
[Photos: MiniCat car runs on compressed air](#)

<http://www.archive.com>



THE DIGITAL-HERITAGE CASE

X   <http://eculture.cs.vu.nl/europeana/session/search>   G Google

login | help | English ▾



This is a **research prototype of Europeana's semantic search engine**. Enter a search term, for example: **Egypt, Rembrandt, window**.

[Collections](#) [Thesauri](#)

Rijksmuseum
46,038 artworks

RKD
82,781 artworks

Louvre
11,327 artworks

[home](#) [disclaimer](#) [datacloud](#) [acknowledgments](#)

powered by **ClioPatria** 1.0 beta 2 (18/12/2008)
SWI Prolog 5.7.10-28-gaf37c11
eculture.cs.vu.nl

The Web: resources and links



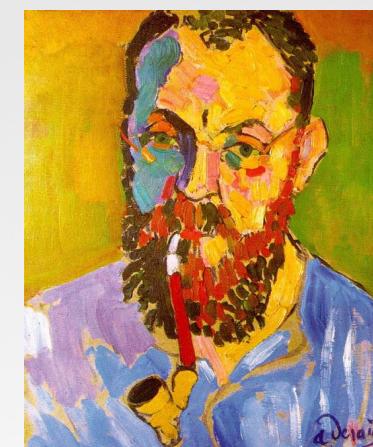
RF-IDentification



URL



Web link

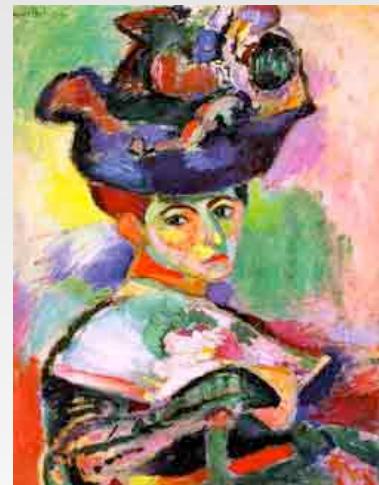


URL

The Semantic Web: typed resources and links



Painting
“Woman with hat”
SFMOMA

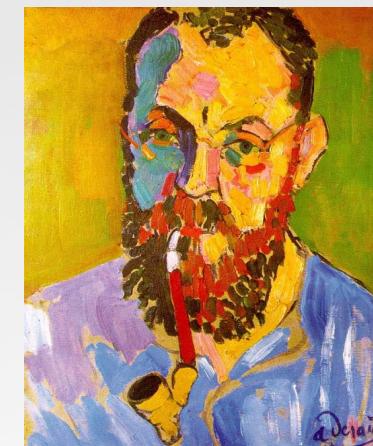


URL

Dublin Core
creator

Web link

ULAN
Henri Matisse



URL



http://www.getty.edu/vow/ULANFullDisplay?find=matisse&role=&nation=&prev_page=1&subjectid=5000173

Research

Research Home ▶ Conducting Research ▶ Union List of Artist Names ▶ Full Record Display

Union List of Artist Names® Online

Full Record Display

[New Search](#) [◀ Previous Page](#)

Click the  icon to view the record in a new window.

ID: 500017300

 **Matisse, Henri** (French)

Names

Matisse, Henri	artist (preferred)
Matisse, Henri	painter
Matisse, Henri	printmaker
Matisse, Henri	sculptor
Matisse, Henri	designer
Matisse, Henri	writer

Nationality: French

Roles:

artist	
painter	
printmaker	
sculptor	
designer	
writer	

Gender: male

Birth and Death Places:

Born: [Le Cateau-Cambrésis \(Nord, Nord-Pas-de-Calais, France\) \(inhabited place\)](#)

Died: [Nice \(Alpes-Maritimes, Provence-Alpes-Côte d'Azur, France\) \(inhabited place\)](#)



Research

[Research Home](#) ▶ [Conducting Research](#) ▶ [Thesaurus of Geographic Names](#) ▶ [Hierarchy Display](#)

Getty Thesaurus of Geographic Names® Online

Hierarchy Display

[New Search](#)

[◀ Previous Page](#)

[? Help](#)

[Vernacular Display](#) | [English Display](#)

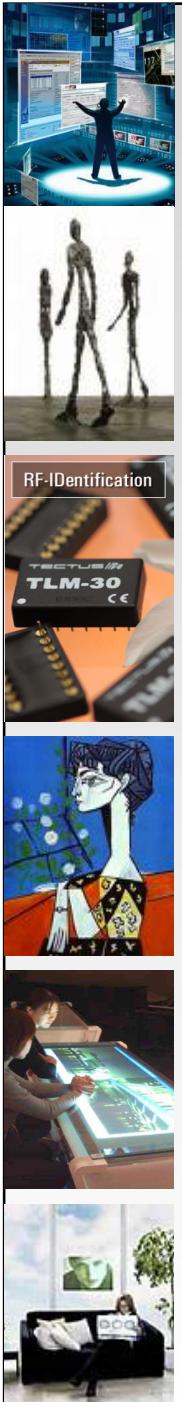
[View Selected Records](#)

[Clear All](#)

Click the  icon to view the hierarchy.

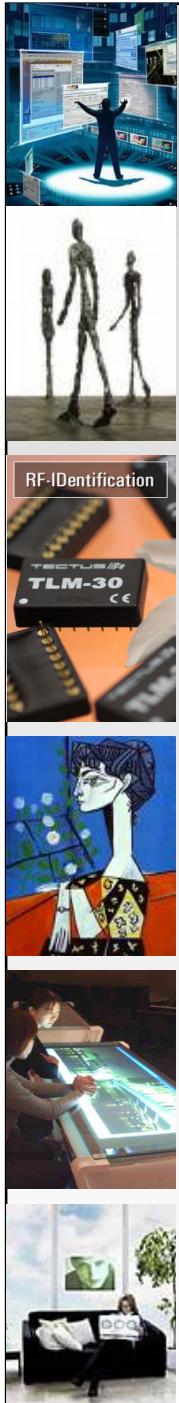
Check the boxes to view multiple records at once.

-  [Top of the TGN hierarchy \(hierarchy root\)](#)
-  [.... **World** \(facet\)](#)
-  [..... **Europe** \(continent\)](#)
-  [..... **Netherlands** \(nation\)](#)
-  [..... \[*view physical features* \]](#)
 - [..... **Aarkanal** \(canal\)](#)
 - [..... **Afsluitdijk** \(dam\)](#)
 - [..... **Alblasserwaard** \(general region\)](#)
 - [..... **Altena, Land van** \(general region\)](#)
 - [..... **Amstelland** \(general region\)](#)
 - [..... **Amsterdam Rijn Kanaal** \(canal\)](#)
 -  [..... **Aruba** \(dependent state\) \[N\]](#)
 - [..... **Bernisse Molen** \(mill center\)](#)
 - [..... **Biesbos** \(general region\)](#)
 - [..... **Brouwersdam** \(dam\)](#)
 - [..... **Calandkanaal** \(canal\)](#)
 - [..... **Delfland** \(general region\)](#)



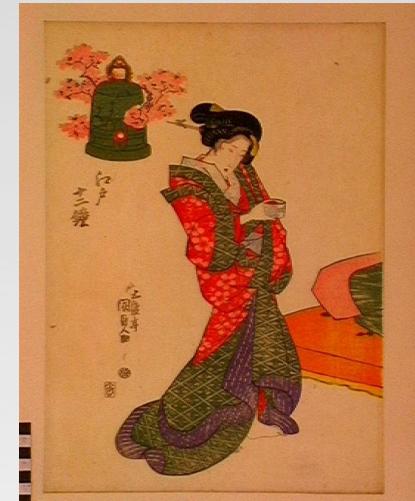
The myth of a unified vocabulary

- In large virtual collections there are always multiple vocabularies
 - In multiple languages
- Every vocabulary has its own perspective
 - You can't just merge them
- But you can use vocabularies jointly by defining a limited set of links
 - “Vocabulary alignment”
- It is surprising what you can do with just a few links



Example use of vocabulary alignment

“Tokugawa”



AAT style/period
Edo (Japanese period)
Tokugawa

*AAT is Getty's
Art & Architecture Thesaurus*

SVCN period
Edo

*SVCN is local in-house
ethnology thesaurus*



Semantic search: clustering and cluster -order principles

cultural heritage on the web :: **humans**

[Pictures](#) and metadata of artworks

Works created by (95)



Three Women
Picasso, Pablo



Glass of Absinthe
Picasso, Pablo



Accordionist
Picasso, Pablo



The Aficionado
Picasso, Pablo

Works by professionally related artist (31)



Fruit Dish, Ace of Clubs
Braque, Georges



Man with a Violin
Braque, Georges



Bottle, Newspaper, Pipe,
Braque, Georges



Still Life BACH
Braque, Georges

Works created by artists with style/period [Surrealist](#) also used by artist (6)



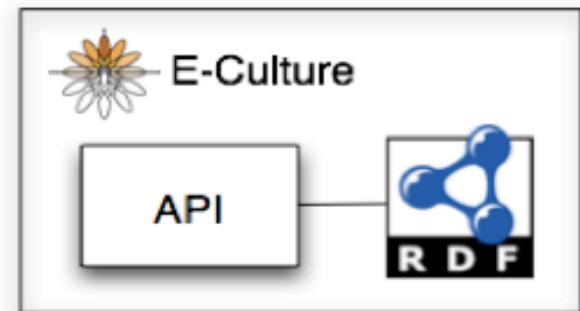
cultural heritage on the web :: machines

E-Culture Search API

- Search *HETEROGENOUS* data

No knowledge of schema required

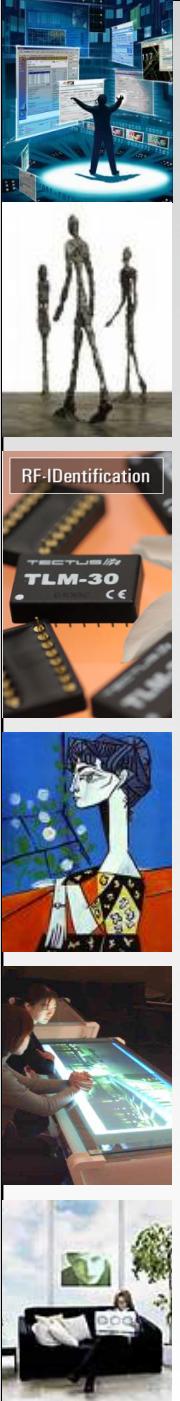
```
http://e-culture.multimedian.nl/demo/api  
/search?query=picasso
```



Result in JSON (Graph)

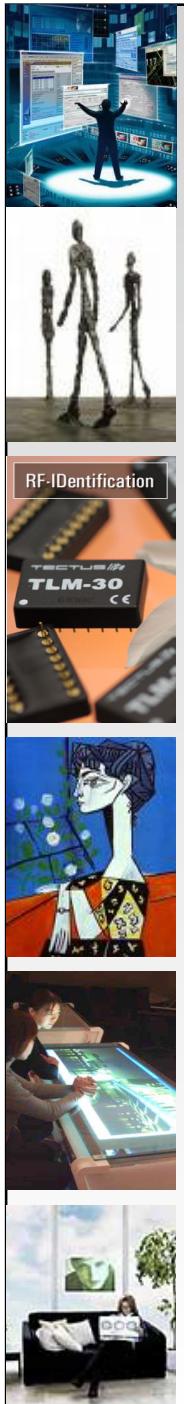
```
"items": [  
  {  
    "uri": "http://t-d-b.org?http://www.artchive.com/artchive/p/picasso/meninas.jpg",  
    "score": 0.794368,  
    "label": "Las Meninas (after Velazquez)",  
    "thumbnail": "/beta/thumbnail?uri=http%3a%2f%2fwww.artchive.com%2fartchive%2fp%2fpicasso%2fmeninas.jpg",  
    "sublabel": "Picasso, Pablo"  
  },  
  {  
    "uri": "http://t-d-b.org?http://www.artchive.com/artchive/p/picasso/self1.jpg",  
    "score": 0.794368,  
    "label": "Self-portrait",  
    "thumbnail": "/beta/thumbnail?uri=http%3a%2f%2fwww.artchive.com%2fartchive%2fp%2fpicasso%2fself1.jpg",  
    "sublabel": "Picasso, Pablo"  
  }]
```



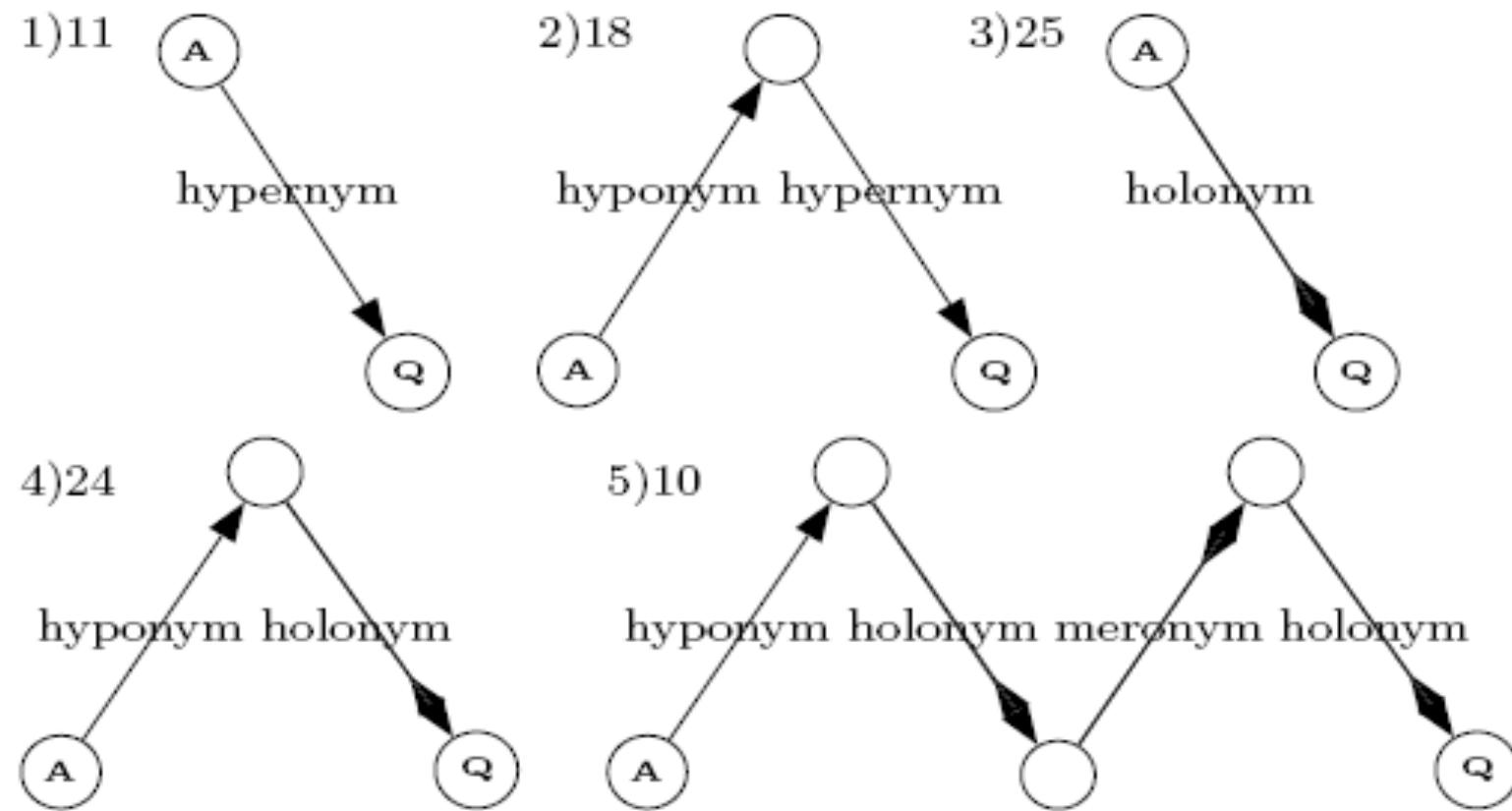


Web science issues: technical

- Information retrieval as graph search
 - more semantics => more paths
 - finding optimal graph patterns
- Vocabulary alignment
- Information extraction
 - recognizing people, locations, ...
 - identity resolution
- Multi-lingual resources

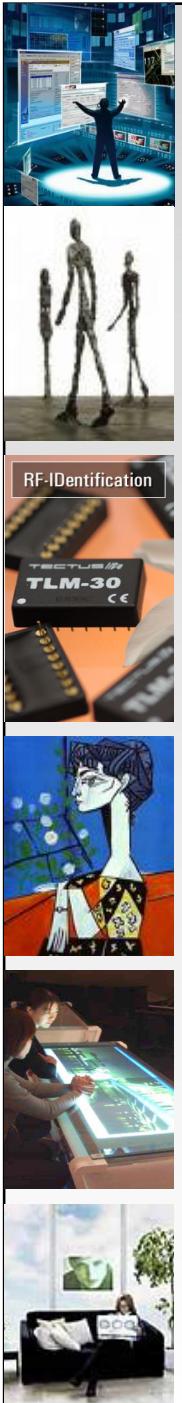


Search: WordNet patterns that increase recall without sacrificing precision



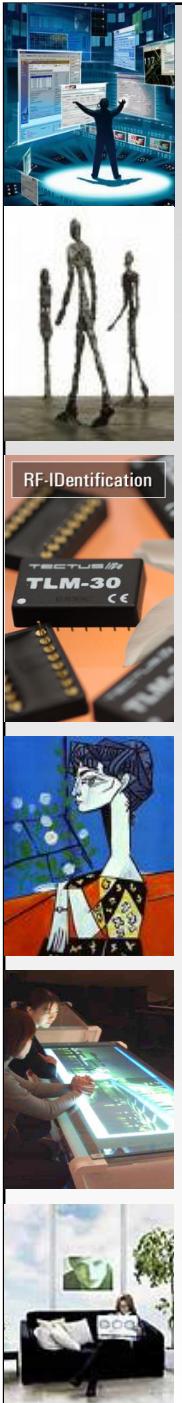
Web science issues: social

- Cultural heritage organizations find it difficult to “give away” their data
 - concerns for quality
- Re-orientation: Web is not derivative of physical presence; they should stand side-by-side
- Universal access: everyone should be able to enjoy the Rijksmuseum Amsterdam?!



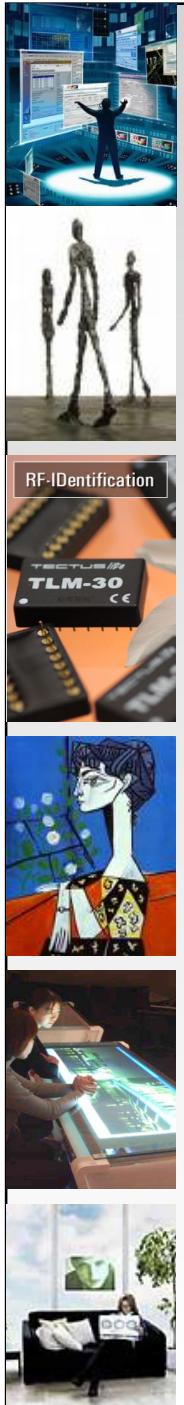
Web science issues: economic

- Primary access free
- Secondary services cost money
 - virtual museum shop can offer much larger collection
 - access to high-resolution images
 - tourist services on mobile devices



Web science issues: legal

- Europeana.eu faces enormous rights issues, in particular wrt recent works
- New licensing frameworks:
 - Creative Commons
 - Open Data Commons



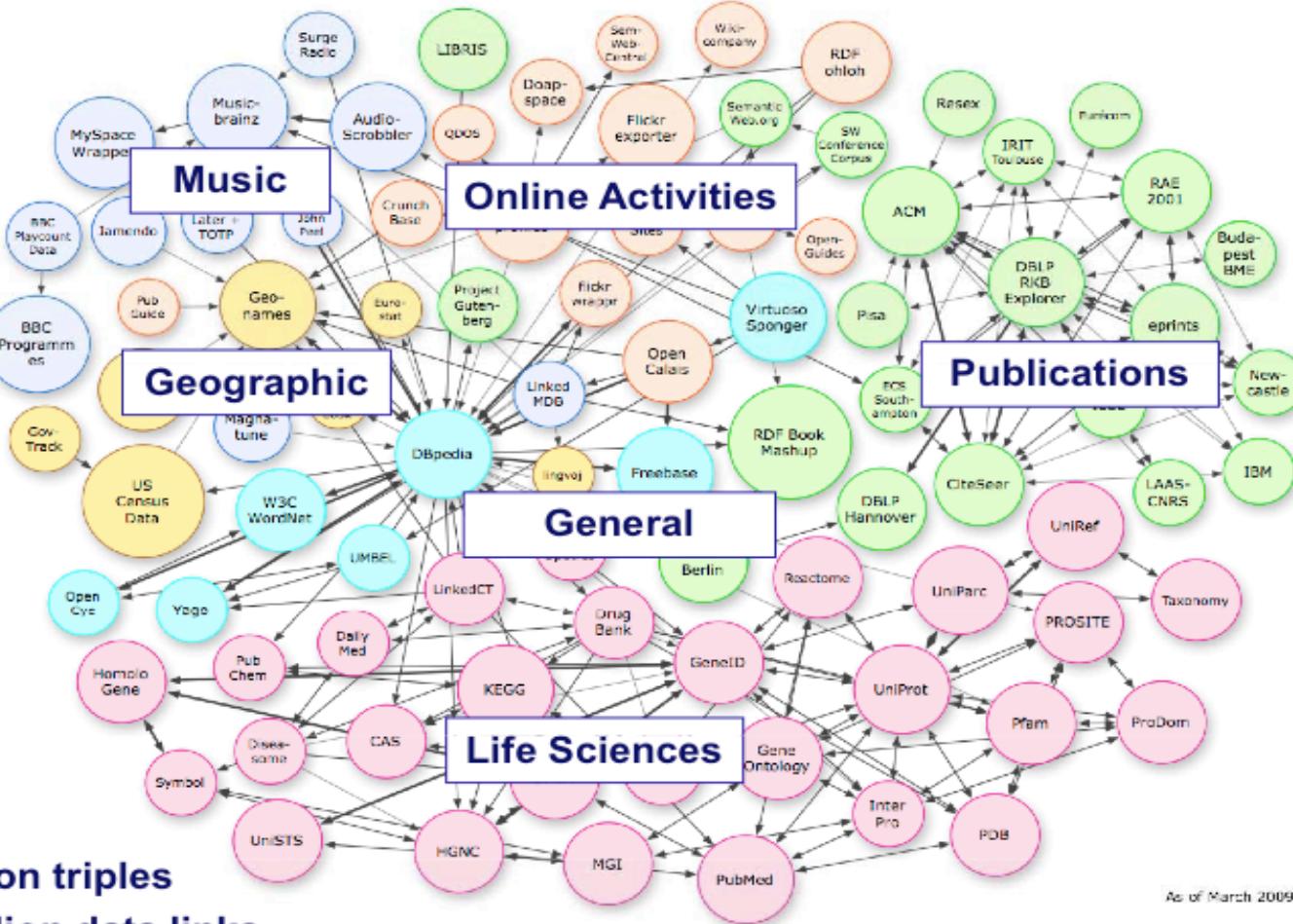
Availability of government data: <http://data.gov.uk>





Linked Open Data initiative

LOD Datasets on the Web: March 2009





Handling billions of statements

http://www.larkc.eu/

larKC

LARGE KNOWLEDGE COLLIDER

Home Overview People Deliverables Publications Presentations Community Early Adopters

Job offer

Subscribe to LarKC news

Enter search terms... SEARCH

Visit other LarKC sites

LarKC Weblog LarKC Wiki

News

2009-11-11 11:51 LarKC pharma expert to speak at ESTC09 panel

2009-11-06 12:52 WANTED: business expert for LarKC!

2009-08-14 12:24 Joint workshop of LarKC and SOA4All

Archives

LarKC: The Large Knowledge Collider

The aim of the EU FP 7 Large-Scale Integrating Project LarKC is to develop the Large Knowledge Collider (LarKC, for short, pronounced "lark"), a platform for massive distributed incomplete reasoning that will remove the scalability barriers of currently existing reasoning systems for the Semantic Web.

read more ↗



User-generated metadata

The execution of Johan van Oldenbarnevelt



RP-P-OB-77.320

Print with a scene

update cancel

Who Historical persons

person

Oldenbarnevelt, Johan van x

What Iconclass (en), WordNet (en), events (nl)
(mythological) concept, object or event

beheading x

Where Name of place or region

geographical place

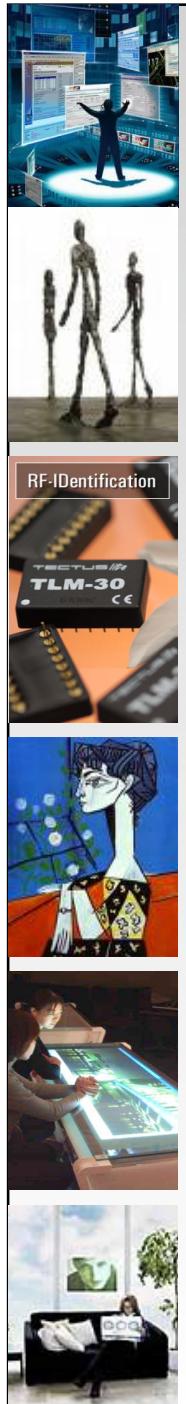
Den Haag x

When Date, year or period

enter date

done | cancel

Auto completion services



What Iconclass (en), WordNet (en), events (nl)

siege

Iconclass [view all 40 results](#)

- [45K21] **siege**
Society, Civilization, Culture
- [45K] **siege, position war**
Society, Civilization, Culture
- [94H] **last months of the siege and the fall of Troy**
Classical Mythology and Ancient History

WordNet [view all 6 results](#)

- siege** (beleaguered, besieging)
blockade
- siege of Orleans** (Orleans)
beleaguered
- Siege** Perilous
seat

siege, position war (more info)
subject on about 26 artworks



- Society, Civilization, Culture
- warfare; military affairs
 - siege, position war**
 - fortifications, military engineering
 - attack ~ siege
 - defence ~ siege
 - capture of city (after the siege)



Video tagging games

http://semanticweb.cs.vu.nl/prestoprime/www/bin-release/Waisda.html#

Waisda?

Select a video Write words describing what you see or hear Score points when your words match with other

Videos How to play?

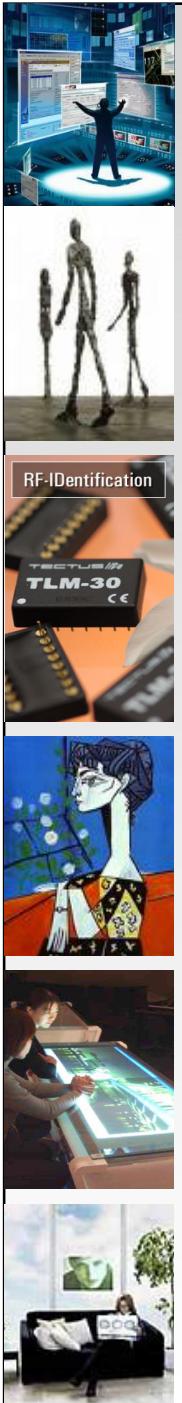
Boer zoekt vrouw afl. 1 Nr. of players: 11 Play!

Barend en van Dorp Nr. of players: 3 Play!

Boer zoekt vrouw afl. 2 Nr. of players: 5 Play!

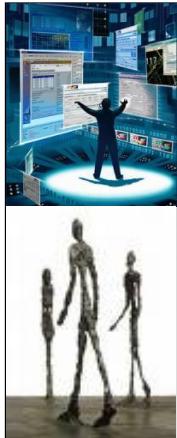
All Time Heroes Today's Champs Supertagger

Done



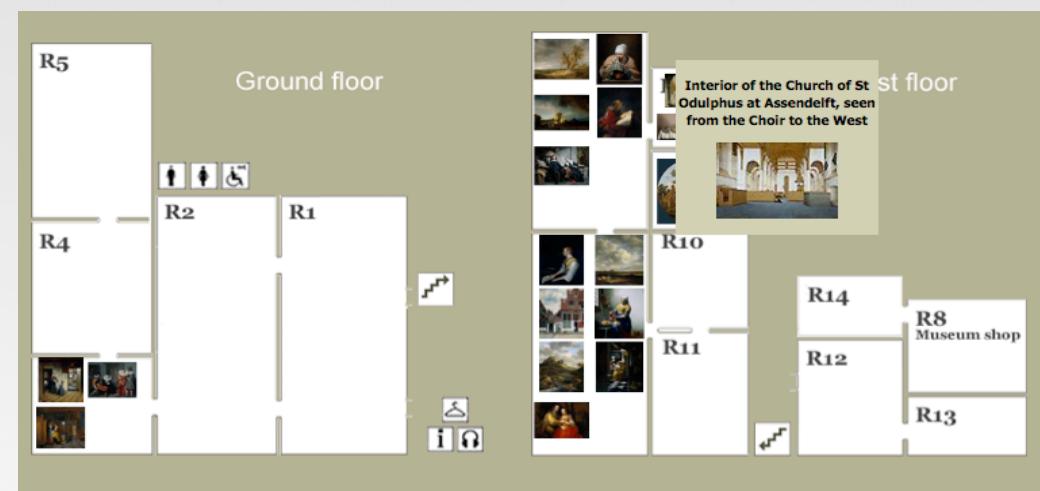
Web science issues

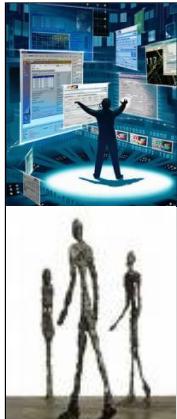
- CH organizations need outside help to annotate their objects
- Tension between in-house conservation -biased metadata and user view
 - who, what, where, when
- How can we derive trust levels?
- Auto completion is a difficult technical issue in case of many vocabularies



Personalized Rijksmuseum

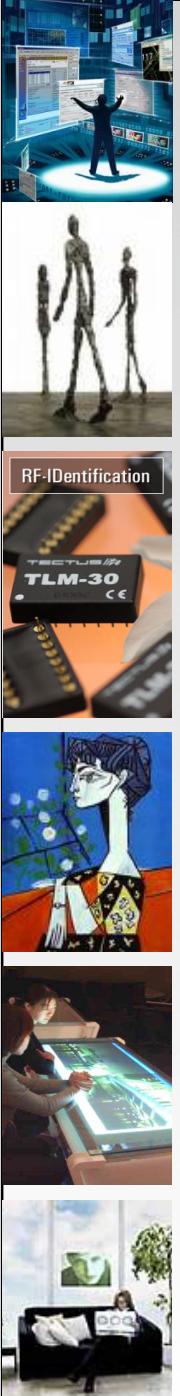
- *Interactive user modeling*
- *Recommendations of artworks and art topics*





Mobile museum tour



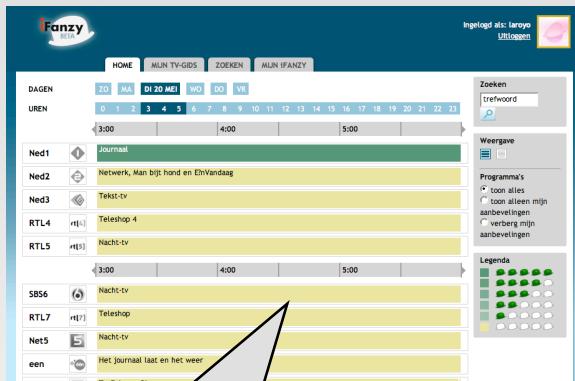


Web science issues

- Recommendation strategies
 - content-based vs. collaborative
- Combining your Web profiles for interests
 - OpenID
 - Google Social Graph API
<http://code.google.com/apis/socialgraph/>
- Tour connects virtual and physical world

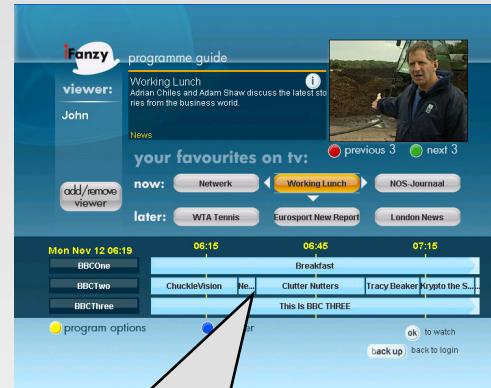


Online TV Guide



- Synchronization with STB
- Semantic Search

Settop Box EPG

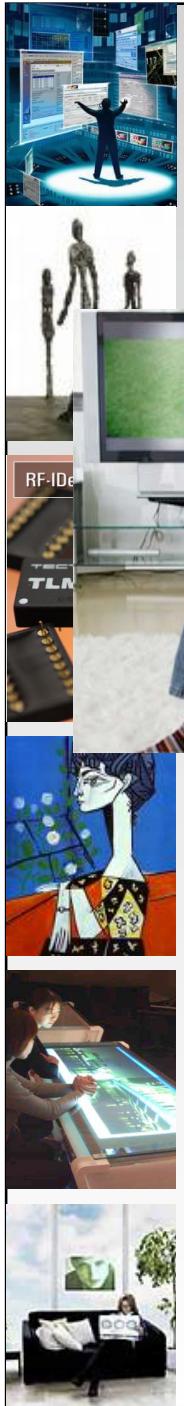


- My TV Night
- What's on for me
- Related Programs

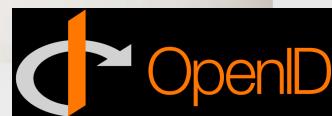
Mobile Identity



- ID Anywhere
- Notifications



Internet TV in Social Web





Web science issues

- TV and Web are becoming merged
 - decoupling of media content from device
- Connecting Web content with media content
 - use music preferences to suggest TV programs
- Trusted access to preferences of friends
- User profile standards: FOAF



User experience Lab



VU University
Amsterdam



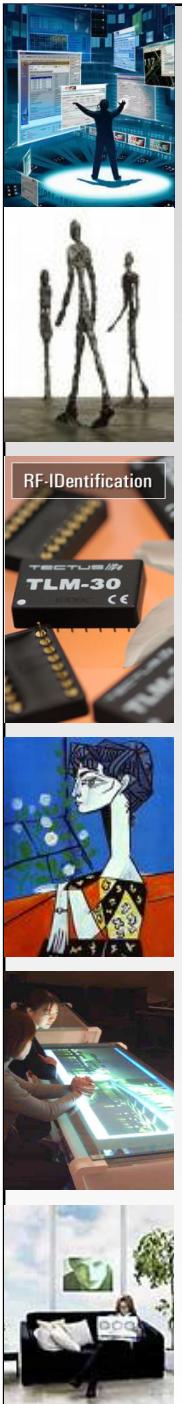
Department of Informatics 52

Rounding up



RF-IDentification





Some things I've learned

Principle 1: Be modest!
allow for multiple views and realities

Principle 2: Think large! cf.
Doug Lenat

Principle 3: Don't recreate but enrich and
align!

Principle 4: Beware of ontological
over-commitment!

Web science education

Typical curriculum topics

- Web technology
- Web communication
- Web society
- Web data

[http://wiki.websciencetrust.org/w/
Curriculum topics](http://wiki.websciencetrust.org/w/Curriculum_topics)



Web Science Trust

- What is Web Science?
- About The Trust
- News
- Research
- Events
- People
- Curriculum
- Friends and Supporters
- Contact Us

What is Web Science?

Nothing like the Web has ever happened in all of human history. The scale of its impact and the rate of its adoption are unparalleled. This is a great opportunity as well as an obligation. If we are to ensure the Web benefits the human race we must first do our best to understand it.

The Web is the largest human information construct in history. The Web is transforming society. In order to understand what the Web is, engineer its future and ensure its social benefit we need a new interdisciplinary field that we call **Web Science**.

Web Science Seminar

For a far more in depth introduction to Web Science, see this seminar from Professor Nigel Shadbolt [45 minutes]



<http://webscience.org/webscience.html>

Web Science

why this matters

- the Web matters
- an essential part of humanity
- An essential part of the current and future economy
- understanding the Web is a major challenge as big as any other global cause
- It is bound to reflect essential aspects of our own human capabilities
- It is certain to introduce new capabilities
- “...from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.”



Take home message

- Web science is much more than computer science for the Web
- Web science is strongly interdisciplinary
- Computer science is an important enabler for Web developments

