

# Information Standards on the Net Today and Tomorrow

Olle Olsson

Swedish W3C Office

Swedish Institute of Computer Science (SICS)

“Information Specialists” April 2014

© 2014 W3C



SWEDISH  
INSTITUTE OF  
COMPUTER  
SCIENCE

**SICS**

## Contents

- The information world & standards
- The web perspective – web standards
- Publishing and metadata
- Metadata formats
- Success, failure, surprise ... the future
- Summary

# The Information World and Standards

## Information world and standards

- Digital society: devices & information
  - Impact on everybody.
  - Some sectors more than others.
- Digital society == networked society
  - "If it is not on the net, it does not exist".
  - More stuff (things) possible to access.
  - Stuff needs to be usable.
    - Interoperability!
- Standards
  - Increased interoperability
    - more stuff can be used.
    - easier to make stuff usable

# Standards

- Standard:
  - **X** is some category of things.
  - Standard for **X** = specification of necessary characteristics for some “thing” **β** to be regarded as an **X**-thing.
  - A standard is a *document*.
- Types of standards:
  - *de jure*: published by officially recognized standards body;
  - consortia: published by a cooperation between major stakeholders in a business sector;
  - *de facto*: what just happens to have wide acceptance.

## The Web Perspective – Web Standards

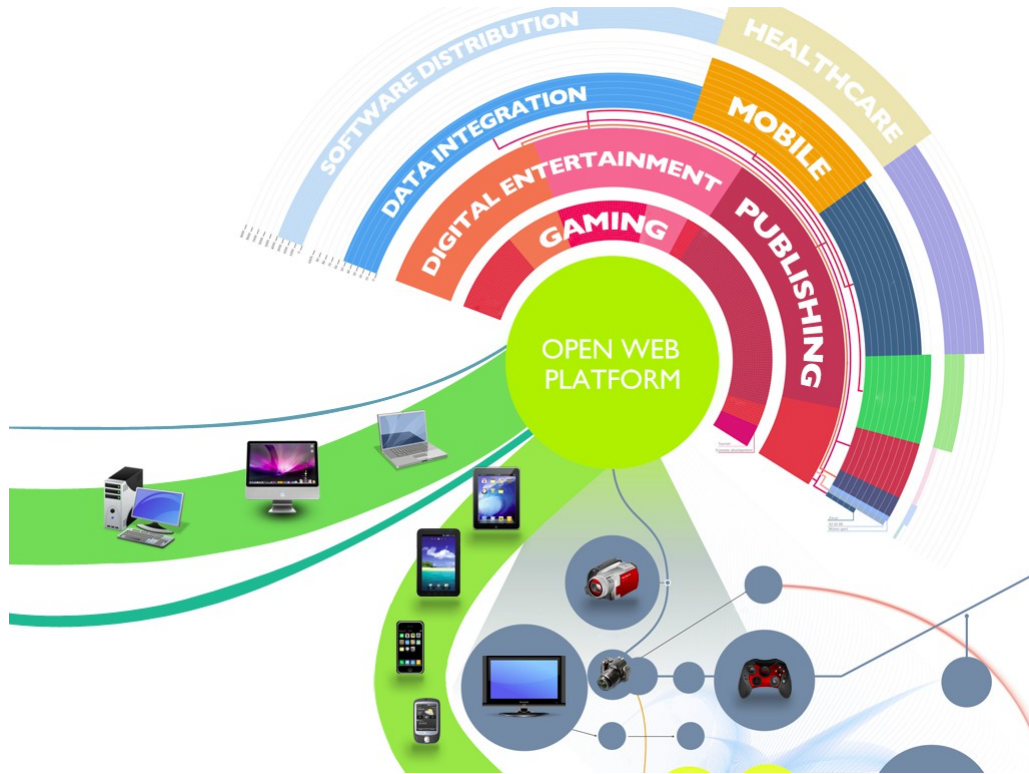
## Standardisation for the web – W3C

- **W3C<sup>®</sup> World Wide Web Consortium** (1994- )
- Industry consortium ... members of the consortium.
- Specify web technologies/standards.
- Contribute to good use of standards.
- Publish standards (“W3C Recommendations”)
  - HTML, HTTP, XML, CSS, RDF, .....
- The web is now 25 years old!
  - Tim Berner-Lee's proposal for a distributed hypertext system on top of the Internet (1989).
- W3C is now 20 years old.

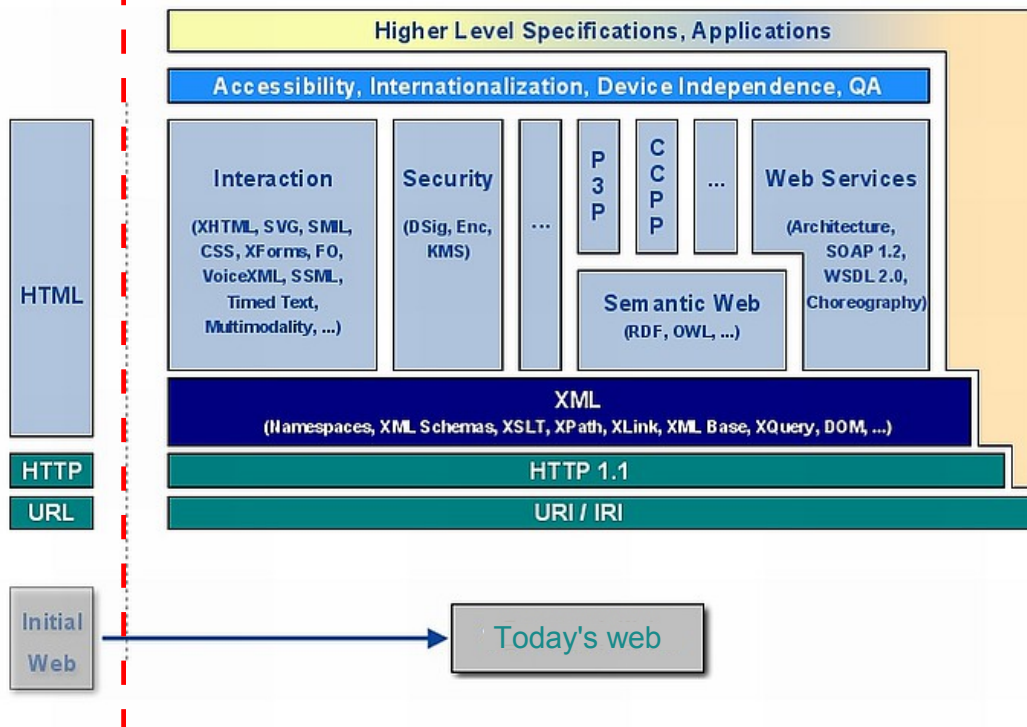
## Standards

- W3C produces “consortia standards”
  - core standards for the web.
  - some later become ISO standards
- Recognized as *the* standards defining organisation (SDO) for the web.
- Close cooperation with IETF, who is defining core Internet standards.
- Cooperation with many other SDOs, for specific industry sectors: automotive, mobile, TV&broadcasting, printing, ..., ***publishing***
- Unifying concept: *The Open Web Platform*

# Open Web Platform



# Open Web Platform



# Publishing

## Publishing today

25% of Americans own a tablet

"After 79 years in print, Newsweek goes digital only"

Google Ad Revenue Now More Than U.S. Print Publications Combined

E-Books See Triple Digit Growth As Paper Book Sales Dive

Pew: Survey Finds Rising Reliance on Libraries as a Gateway to the Web

## Publishing and the Web

- Publishing industry coming to the web
- eBooks using web technology
  - HTML5, HTTP, eBook readers on “web browser 2.0”, W3C EME (“Encrypted Media Extensions”), ...
- Device independence (larger markets)
- Distribution channels (incl. “app stores”)
  
- Contents published in standard web document formats
- Metadata using widely accepted formats for web metadata

## Information and metadata

- Metadata = information about information
  - What? Who? When? Where? ....
- On the web:
  - Well-structured metadata
    - Business: Amazon, ...
    - Services: Google Books, Google Scholar, ...
  - Semi-structured metadata:
    - Flickr, YouTube, ....
- What about old school publishing?

# Where was metadata?

Library of Congress Cataloging-in-Publication Data  
Narendra, Kumpati S.  
Stable adaptive systems / Kumpati S. Narendra and Anuradha M. Annaswamy.  
p. cm. — (Prentice Hall information and system sciences series)  
Includes index.  
ISBN 0-13-839994-8  
1. Adaptive control systems. I. Annaswamy, Anuradha M.  
II. Title. III. Series.  
TJ217.N37 1989 88-23841  
629.8'36—dc19 CIP

©1989 by Prentice-Hall, Inc.  
A Division of Simon & Schuster  
Englewood Cliffs, New Jersey 07632

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

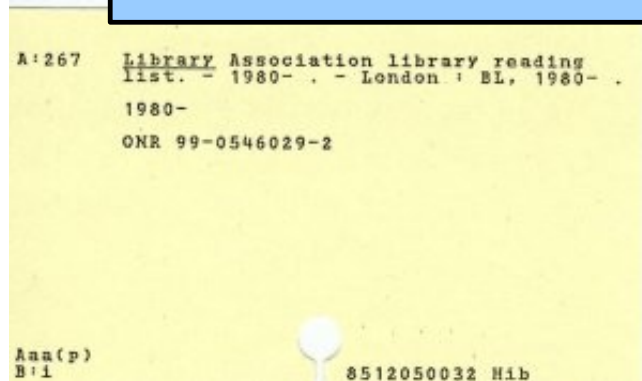
Printed in the United States of America  
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-839994-8

Prentice-Hall International (UK) Limited, London  
Prentice-Hall of Australia Pty. Limited, Sydney  
Prentice-Hall Canada Inc., Toronto  
Prentice-Hall Hispanoamericana, S.A., Mexico  
Prentice-Hall of India Private Limited, New Delhi  
Prentice-Hall of Japan, Inc., Tokyo  
Simon & Schuster Asia Pte. Ltd., Singapore  
Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro

Inside information object

A separate information object



# Where is Metadata? (1)

Inside information object

```
<html>
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Expressing Dublin Core in HTML/XHTML meta and link elements</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />

<meta name="DC.title" lang="en" content="Expressing Dublin Core in HTML/XHTML ..." />
<meta name="DC.creator" content="Andy Powell, UKOLN, University of Bath" />
<meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2003-11-01" />
<meta name="DC.identifier" scheme="DCTERMS.URI"
content="http://dublincore.org/documents/dcq-html/" />
<link rel="DCTERMS.replaces" hreflang="en"
href="http://dublincore.org/documents/2000/08/15/dcq-html/" />
<meta name="DCTERMS.abstract" content="This document describes how
qualified Dublin Core metadata can be encoded in HTML/XHTML &lt;meta&gt; elements" />
<meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
</head>
...
</html>
```



## Where is Metadata? (2)

Inside information object

Beräknad på  
en sammanlagd  
skatte- och  
avgiftssats av  
29-37 %

För beräkning av skatteavdrag för år 2014

# Skattetablell 29-37

Innehåll:

- Skatteavdrag
- Skatteavdrag s.k. stående
- Förhöjt skatteavdrag
- Skatteavdrag
- Skatteavdrag

Document Properties

Description Security Fonts Custom Advanced

Description

File: skatteverket-40335.pdf

Title: Skattetableller, löntagare (SKV 403)

Author:

Subject:

Keywords:

Created: 2013-12-13 10:32:11

Modified: 2013-12-13 18:12:27

Application: Adobe InDesign CS6 (Macintosh)

## Where is Metadata? (3)

A separate information object  
( Dbpedia )



### About: Moomin

[Sponge](#) [Permalink](#)

An Entity of Type : [yago:ComicStrip107003352](#), within Data Space : [dbpedia.org](#) associated with source [dataset\(s\)](#)


Type: [yago:ComicStrip107003352](#) Command: Start New Facet Go

*The Moomins are the central characters in a series of books and a comic strip by Swedish-Finn illustrator and writer Tove Jansson, originally published in Swedish by Schildts in Finland. They are a family of fairy tale characters, who are white and roundish, with large snouts, that make them resemble hippopotamuses.*

Attributes	Values
<a href="#">type</a>	<a href="#">Thing</a> <a href="#">book</a> <a href="#">work</a> <a href="#">written work</a> <a href="#">illustration</a>
<a href="#">author</a>	<a href="#">Tove Jansson</a>
<a href="#">country</a>	<a href="#">Finland</a>
<a href="#">illustrator</a>	<a href="#">Tove Jansson</a>
<a href="#">language</a>	<a href="#">Swedish language</a>
<a href="#">publisher</a>	<a href="#">Drawn and Quarterly</a> <a href="#">Farrar, Straus and Giroux</a> <a href="#">Macmillan Publishers (United States)</a> <a href="#">dbpedia:Schildts</a>

## Where is Metadata? (4)

A separate information object  
( OpenLibrary )



**Mumifar og havet**  
fortalt og illustreret af Tove Jansson ; oversat af Marie Svendsen.  
Published **1973** by [Carlsen](#) in [[Koebenhavn](#)].  
Written in [Danish](#).

**Edition Notes**  
Originaltitel: Pappan och havet.

**The Physical Object**

Pagination	206 s :
Number of pages	206

**ID Numbers**

Open Library	OL21678585M
ISBN 10	8756204906
Library Thing	<a href="#">37920</a>

**Read**  
No readable version available.

**Borrow**  
[Physical copy, local](#) WorldCat

**Buy**

- [Alibris](#)
- [Amazon](#)
- [AbeBooks](#)
- [Biblio.com](#)
- [Book Depository](#)
- [Powells](#)

**Lists**  
You *could* add **Mumifar og havet** to a [list](#) if you [log in](#).

**History** Created November 3, 2008 · [4 revisions](#) Download catalog record: [RDF](#) / [JSON](#) / [OPDS](#) | [Wikipedia citation](#)

## Metadata formats

# Metadata Formats



Mumifar og havet  
Tove Jansson

**Mumifar og havet**  
fortalt og illustreret af Tove Jansson ; oversat af Marie Svendsen.  
Published **1973** by [Carlsen](#) in [[Koebenhavn](#)].  
Written in [Danish](#).

**Edition Notes**  
Originaltitel: Pappan och havet.

**Read**  
No readable version available.

**Borrow**  
[Physical copy, local](#) WorldCat

**Buy**

Download catalog record: [RDF](#) / [JSON](#) / [OPDS](#) | [Wikipedia citation](#)

Number of pages 206

**ID Numbers**

Open Library [OL21678585M](#)

ISBN 10 [8756204906](#)

Library Thing [37920](#)

[Biblio.com](#)

[Book Depository](#)

[Powells](#)

**Lists**  
You could add **Mumifar og havet** to a [list](#) if you [log in](#).

Download catalog record: [RDF](#) / [JSON](#) / [OPDS](#) | [Wikipedia citation](#)

**History** Created November 3, 2008 · [4 revisions](#)

# Multimedia metadata formats (examples)

From: *Ontology for Media Resources 1.0*  
W3C Recommendation 09 February 2012

Format	Example
CableLabs 1.1	cl11:Writer_Display
DIG35	dig35:ipr_name/ipr_person@description
Dublin Core	dc:creator
EBUCore	ebucore:creator
EXIF 2.2	exif:Artist
ID3	id3:TCOM
IPTC	iptc:Creator
LOM 2.1	lom21:LifeCycle/Contribute/Entity
Media RSS	mrss:credit@role
MPEG-7	mpeg7:CreationInformation/Creation/Creator/Agent
OGG	ogg:track
QuickTime	qt:com.apple.quicktime.author
DMS-1	dms:Participant/Person
TTML	ttml:actor
TV-Anytime	tva:CredistsList/CredistItem
TXFeed	txf:author
XMP	xmp:CreatorTool
YouTube Data API Protocol	yt:author

## Europeana – European digital heritage

- A federation service – harmonizing metadata
- Using other services

external vocabulary

language variations

```
<edm:Agent rdf:about="http://dbpedia.org/resource/Raphael">
  <skos:prefLabel xml:lang="de">raffael</skos:prefLabel>
  <skos:prefLabel xml:lang="it">raffaello sanzio</skos:prefLabel>
  <skos:prefLabel xml:lang="no">raffaello santi</skos:prefLabel>
  <skos:prefLabel xml:lang="fi">rafael</skos:prefLabel>
  <skos:prefLabel xml:lang="pt">rafael sanzio</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">raphaël (peintre)</skos:prefLabel>
  <skos:prefLabel xml:lang="sv">rafael (konstnär)</skos:prefLabel>
  <skos:prefLabel xml:lang="en">raphael</skos:prefLabel>
  <skos:prefLabel xml:lang="ru">рафаэль санти</skos:prefLabel>
  <skos:prefLabel xml:lang="es">rafael (nombre)</skos:prefLabel>
  <skos:prefLabel xml:lang="es">rafael sanzio</skos:prefLabel>
  <skos:prefLabel xml:lang="nl">rafaël santi</skos:prefLabel>
</edm:Agent>
```

## General metadata standards – examples

- SKOS – Simple Knowledge Organization System
  - Representation of controlled vocabularies:
    - thesauri, classification schemes, taxonomies, subject-heading systems, ...
- RDF
  - General format – metadata models can be expressed:
    - Dublin Core, Creative Commons, ...
- PROV
  - Data model for provenance interchange on the Web
- SPARQL
  - Protocol for access to metadata (semantic information)

## Related activities @ W3C - examples

### ▪ Digital Publishing Interest Group

- With: the digital publishing ecosystem of electronic journals, magazines, news, or book publishing (authors, creators, publishers, news organizations, booksellers, accessibility and internationalization specialists, etc.)
- For: technical discussions, gathering use cases and requirements to align the existing formats and technologies (e.g., for electronic books) with those used by the Open Web Platform.

### ▪ Web and TV Interest Group

- With: the TV and moving images ecosystem
- For: review of existing work, as well as the relationship between services on the Web and TV services, and to identify requirements and potential solutions to ensure that the Web will function well with TV.

Success, failure, surprise ... the future

## Technology evolution just happens (not planned)

- The market selects among alternatives
  - XHTML: not enough uptake
  - HTML5: big success
- New markets => new needs
  - Mobile, tablets on the rise
  - PCs not the hot area
- Not enough value
  - Traditional metadata vs raw search
  - “write-only metadata” (no obvious usefulness)
- Industry preferences
  - Protected contents vs open contents (business models)

*Metadata adopted when consumer added value can be produced.*

## Summary

# Trends

- Standards proliferation has happened
  - Across sectors of society (industry sectors)
  - Within sectors
- The web is the common platform for access to information
- Web standards emerge as reusable common formats
  - General functionality / expressiveness
- Industry sectors move towards web standards
  - Web-aware devices and platforms invade “industry homelands”
  - Need to be interoperable
- Information is increasingly decentralised
  - Web paradigm provides a useful architectural solution.

---end---