PUBLICAÇÕES AMPLIADAS: UM NOVO MODELO DE PUBLICAÇÃO ACADÊMICA PARA O AMBIENTE DE E-SCIENCE

Luana Farias Sales
Luís Fernando Sayão
Rosali Fernandez de Souza

MOTIVAÇÃO!

- Atuação em um **instituto de pesquisa**
- Novas práticas do **fazer científico**
- Geração **intensiva de dados**
- Esvaziamento da **Biblioteca**
- Percepção das novas **necessidades dos usuários**
- Criação de um **novo serviço**? Um novo **modelo de biblioteca**:
  - Novo modelo de publicação
- Continuidade aos estudos de **Web Semântica**, especificamente quanto à questão das **relações conceituais**
- Expectativa em torno da investigação de novo **modelo de informação** para a área de **Ciências Nucleares**

CONTEXTUALIZANDO....
OS PARADIGMAS CIENTÍFICOS

1º PARADIGMA:

Ciência experimental ou empírica estuda a relação entre fenômenos por meio de experimentos

2º PARADIGMA:

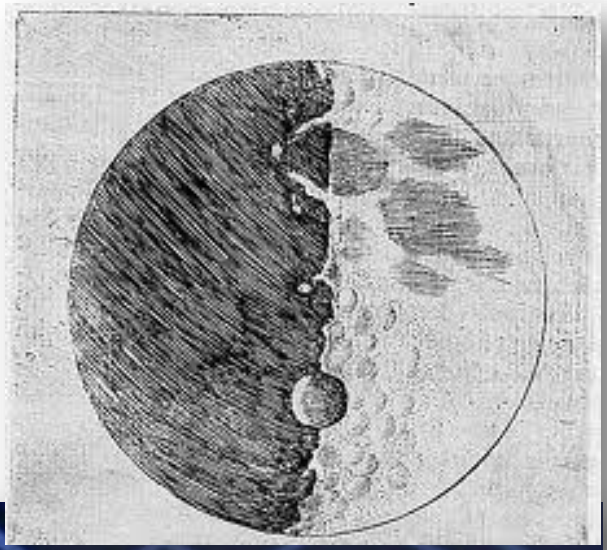
Ciência teórica ou **descritiva** formula modelos para descrição e explicação de modelos

3º PARADIGMA:

Ciência baseada em **simulação** uso de software e grande **geração de dados**

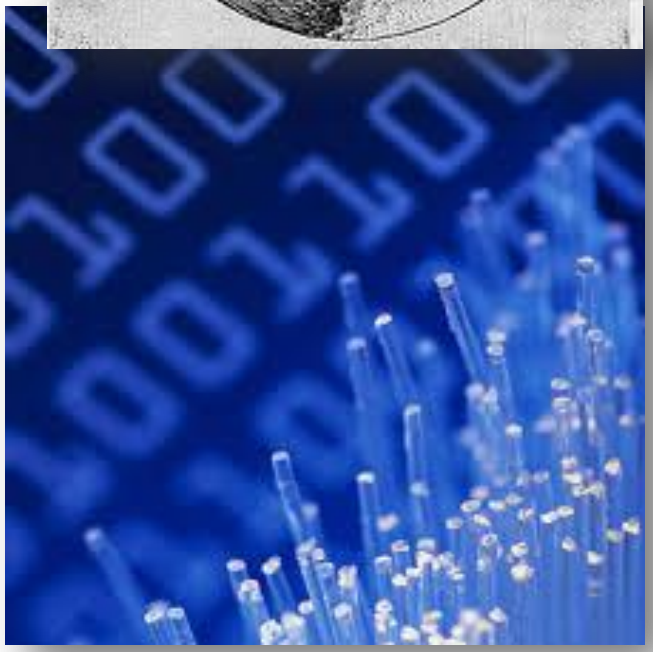
O QUARTO PARADIGMA CIENTÍFICO

E-SCIENCE



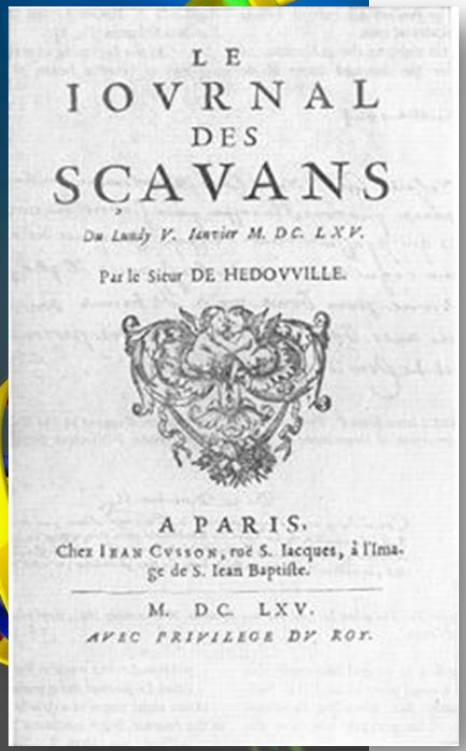
Pesquisa baseada no uso intensivo de dados de pesquisa, bem como no seu armazenamento, processamento, análise e compartilhamento.

E-science - “colaboração global em áreas chave da ciência e toda próxima geração de infraestrutura que vai possibilitar esta colaboração” (TAYLOR, 2001)



Ciência produzida a partir do uso, armazenamento, processamento, análise e compartilhamento de dados de pesquisa. (GRAY, 2005)

JUSTIFICATIVA



LE
IOVRNAL
DES
SÇAVANS
Du Lundy V. Ianvier M. DC. LXV.
Par le Sieur DE HEDOVVILLE.

A PARIS,
Chez IEAN CVSSON, ruë S. Iacques, à l'Image de S. Iean Baptiste.

M. DC. LXV.
AVEC PRIVILEGE DV ROY.

CONSIDERANDO:

Que a **publicação** é uma forte moeda para a ciência;

As alterações ocorridas na **produção científica** ao longo do tempo;

O aumento exponencial de **geração de dados**.

Concepção da pesquisa como um todo **integrado semanticamente**;

Necessidade de um modelo que reflita com fidedignidade a complexidade, a diversidade e a sofisticação da **atividade científica contemporânea**;

Necessidade de veículos de informação que proporcionem velocidade, interatividade, **compartilhamento e reuso** de informações;

Necessidade de poupar o tempo do usuário por meio da **recuperação integrada de informações** de naturezas distintas, mas que se relacionam semanticamente.

QUESTÕES

Como a atividade científica contemporânea pode ser disseminada, compartilhada por meio de um modelo que reflita com fidedignidade sua complexidade, diversidade e sofisticação?

É possível a criação de um modelo que reflita a complexidade dessa atividade científica, permitindo que ela seja recuperada e reutilizada para novas pesquisas, no todo ou em parte?

NOVOS MODELOS DE PUBLICAÇÕES

A publicação tradicional é limitada

Os dados curados podem adicionar valor aos resultados de pesquisa de publicações

PUBLICAÇÃO SOBREPOSTA **PUBLICAÇÃO COMPOSTA**

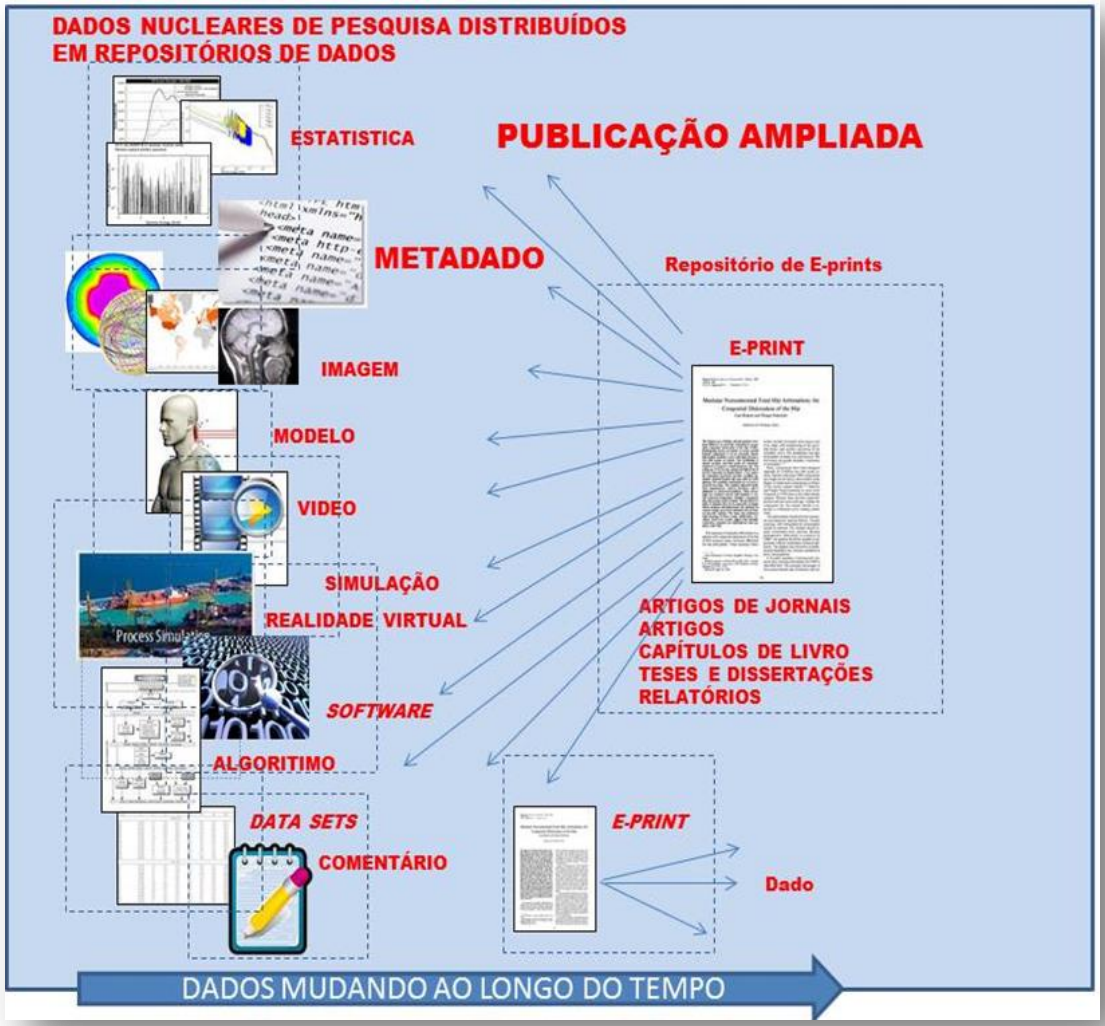
PUBLICAÇÃO AMPLIADA

PUBLICAÇÕES SEMÂNTICAS

PUBLICAÇÃO AMPLIADA: DEFINIÇÃO

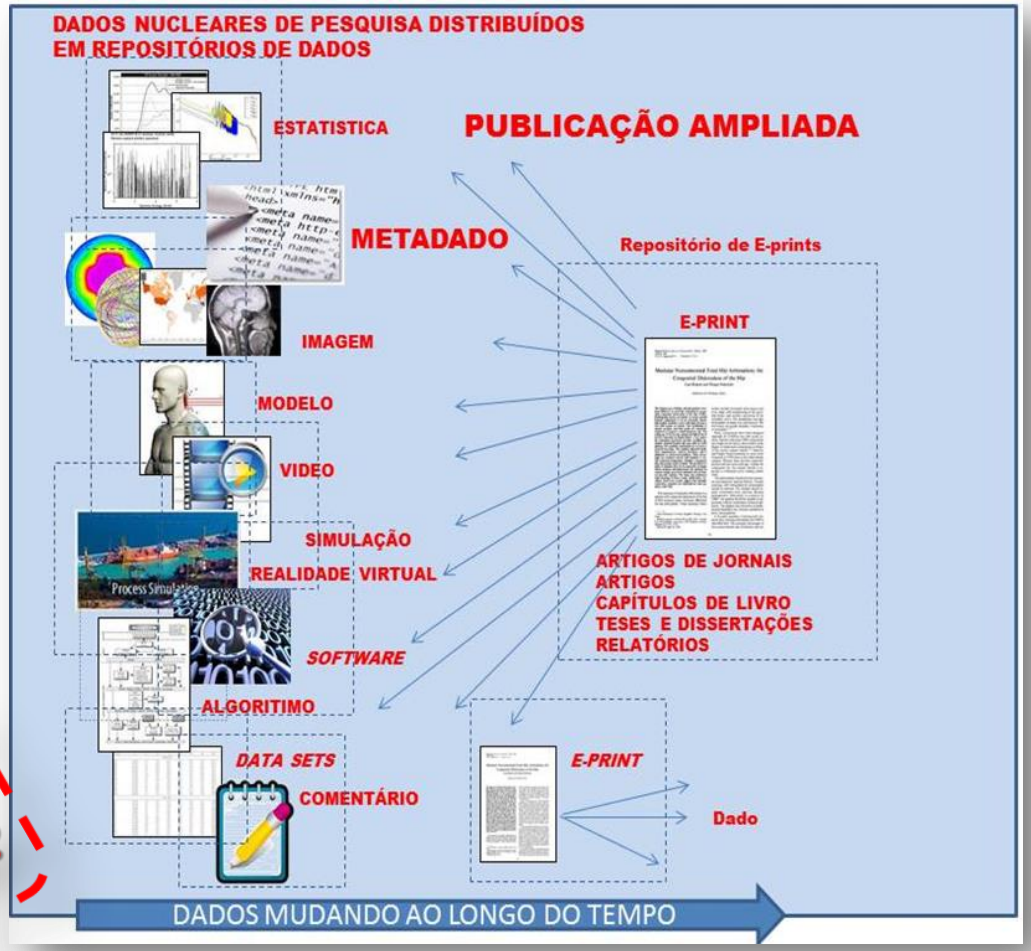
“

Objetos digitais compostos que combinam *e-Prints* com um ou mais recursos de dados, um ou mais registros de metadados, ou qualquer combinação deles (VERHAAR, 2008, p.11)



PUBLICAÇÃO AMPLIADA: COMPOSIÇÃO

- Objetos Digitais
  - E-prints
  - **DADOS DE PESQUISA**
  - Outros Recursos
- Metadados
  - Descritivos
  - Estruturais
  - Administrativos
- Relações
  - de inclusão
  - sequenciais
  - de versionamento
  - de linhagem
  - de manifestação
  - de citação



RELAÇÕES SEMÂNTICAS?

DADOS DE PESQUISA

“

FATOS, NÚMEROS, LETRAS, SÍMBOLOS QUE DESCREVEM UM OBJETO, UMA CONDIÇÃO, UMA SITUAÇÃO OU OUTRO FATOR NRC, 1999).

Não é possível exibir esta imagem no momento.

REGISTRO DE FATOS USADOS COMO FONTES PRIMÁRIAS NA INVESTIGAÇÃO CIENTÍFICA E QUE GERALMENTE SÃO ACEITOS NA COMUNIDADE CIENTÍFICA COMO NECESSÁRIOS PARA A VALIDAÇÃO DOS RESULTADOS DE PESQUISA (OCDE, 2007).

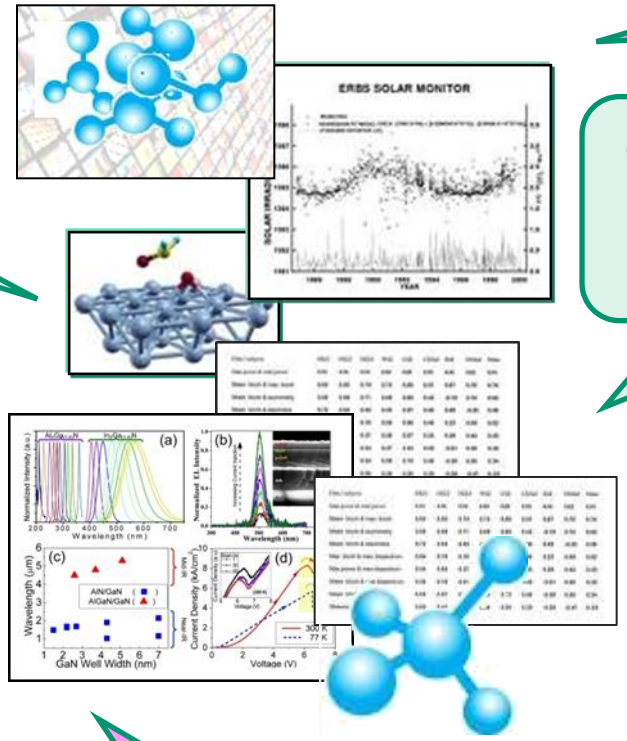
DADOS DE PESQUISA

CARACTERÍSTICA DE DIVISÃO	NFS	BORGMAN	HARVEY	LYON
Quanto à natureza	Número Imagem Software			
Quanto à origem	Observacionais Computacionais Experimentais	Registro		
Quanto ao nível de processamento	Intermediário Finais			
Quanto ao estágio de geração		Brutos Derivados		Crus Primários Derivados
Quanto à formação de coleções			De pesquisa De Comunidade De referência	
Quanto à mutabilidade				Canônicos Episódicos

O QUE É DADO DE PESQUISA ?

FORMULAS MATEMÁTICAS
expressas em MathXML ou algoritmos

ESPECIFICAÇÃO DE INSTRUMENTOS OU DE OUTROS *HARDWARE*



COLEÇÃO DE DADOS
resultados de experimentos, medidas, resultados de levantamentos

CERTIFICADOS DIGITAIS PARA INSTRUMENTOS CIENTÍFICOS

VISUALIZAÇÃO DE DADOS
gráficos, diagramas, tabelas ou modelos em 3D

COMENTÁRIOS E ANOTAÇÕES
de agentes que tenham consultado os objetos digitais

ESTRUTURAS QUIMICAS LÉGIVEL POR MÁQUINA

SOFTWARE
código fonte ou implementados como serviços web

DOCUMENTOS TEXTUAIS
que fazem parte de corpus criados para propósitos de pesquisa

MULTIMÍDIAS
imagens, vídeos e gravações em áudio

PUBLICAÇÃO AMPLIADA: FORMAÇÃO

PUBLICAÇÕES AMPLIADAS: REQUISITOS E RECOMENDAÇÕES

1. Deve ser possível **especificar as partes** componentes de uma publicação ampliada;
2. A publicação ampliada e seus componentes devem ser disponibilizados como recurso *web* **identificados por URI**, o mesmo vale para seus componentes;
3. Deve ser possível **agregar objetos digitais compostos** à publicação ampliada;
4. Deve ser possível **acompanhar as diferentes versões** das publicações ampliadas como um todo e das suas partes constituintes;
5. Deve ser possível **registrar as propriedades básicas** da publicação e dos outros recursos que estão adicionados a ela;
6. Deve ser possível **registrar a autoria** da publicação ampliada e dos outros recursos que a compõe;
7. Deve ser possível **assegurar a preservação de longo prazo** das publicações ampliadas;
8. Deve ser possível **registrar as relações** entre os recursos *web* que fazem parte da publicação ampliada;
9. As instituições que oferecem acesso a publicações ampliadas devem assegurar que elas **possam ser descobertas**;

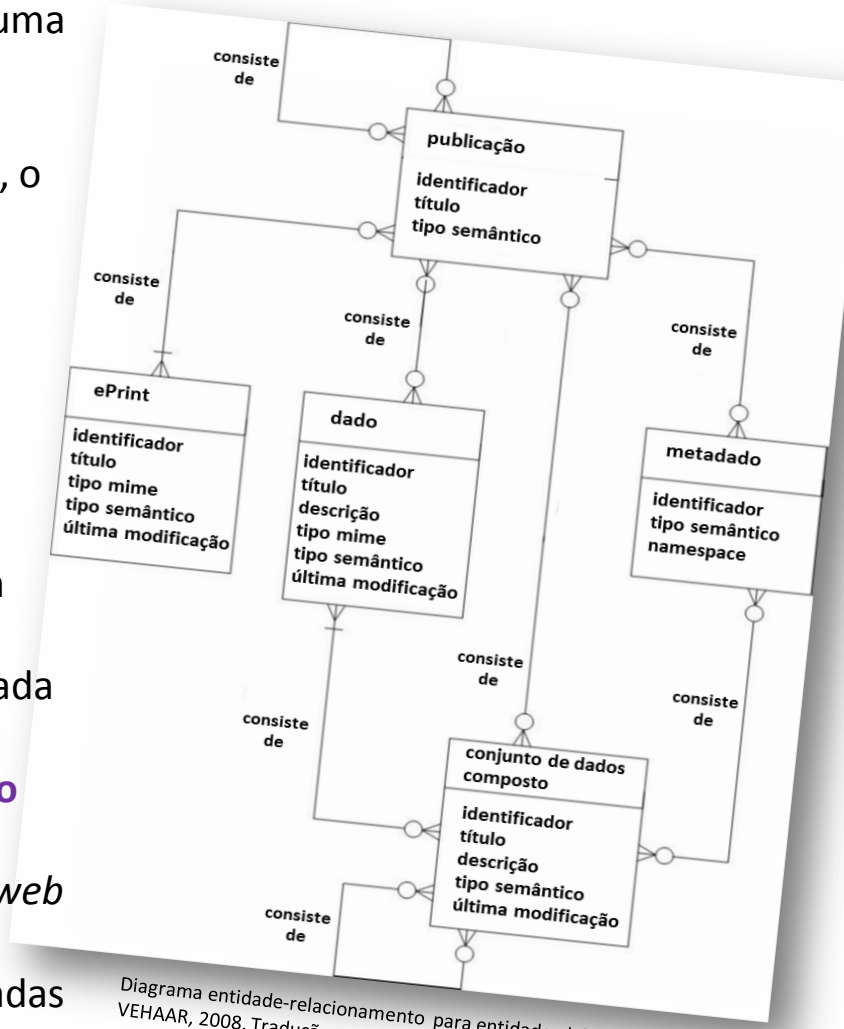


Diagrama entidade-relacionamento para entidades básicas e propriedades
VEHAAR, 2008. Tradução nossa

PUBLICAÇÕES AMPLIADAS: ESTRUTURA

MODELO ORE – apresenta um mapa de recursos permitindo a associação de uma identidade com agregações de recursos.

A NORMA OAI-ORE – visa padronizar a forma de descrever os elementos constituintes do objeto digital composto e os limites de uma agregação.

PUBLICAÇÕES AMPLIADAS: O MODELO *Object Reuse and Exchange* (ORE)

O modelo básico do OAI-ORE se resume em três entidades:

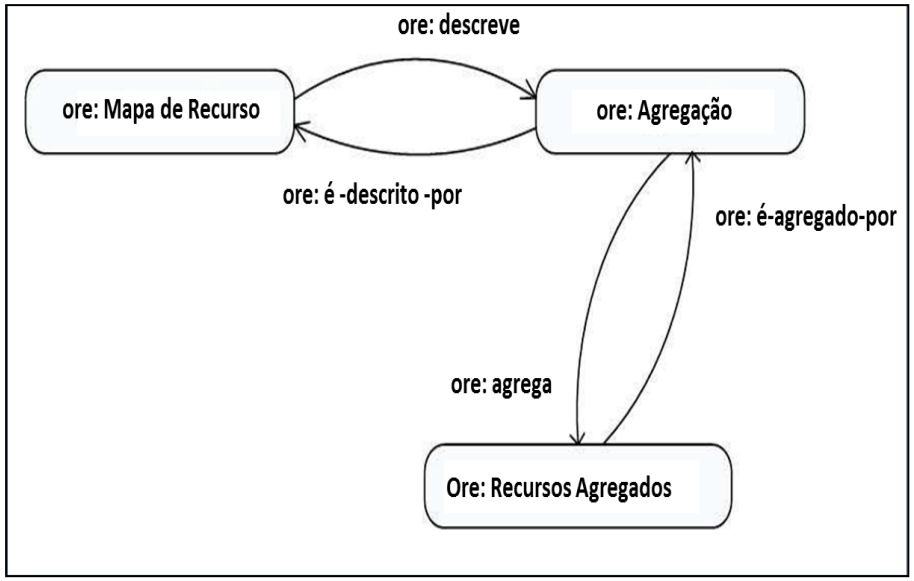
1. Agregação - coleção de recursos web

2. Recursos Agregados– recursos individuais que formam a agregação.

3. Mapa de Recursos - entidade que contém a descrição de uma "Agregação".

Essas entidades são ligadas por meio de propriedades que relacionam essas entidades:

descreve,
é-descrito-por,
agrega,
é-agregado-por
igual.



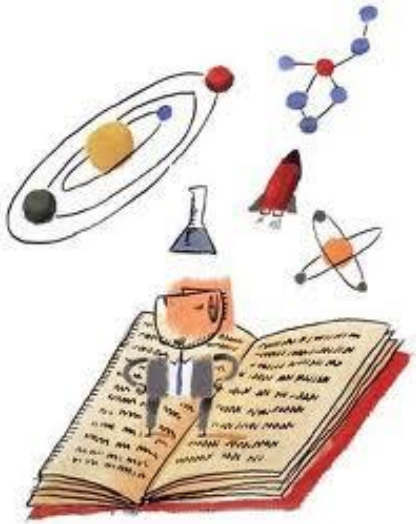
Modelo Básico do OAI-ORE.
SURF, 2009. Tradução nossa

PUBLICAÇÕES AMPLIADAS: A NORMA *Object Reuse and Exchange* (OAI-ORE)

Baseada no modelo ORE

Utiliza estrutura de grafos ao invés de uma estrutura em árvore.

Ponto de partida: cada unidade de um objeto complexo deve ser descrita separadamente com metadados que registrem sua estrutura, linhagem e identificação persistente.



O mapa de recursos torna explícita a associação de uma identidade a uma agregação de recursos e faz declarações sobre a sua estrutura.

Requere dos objetos digitais uma camada de interoperabilidade que vai além dos metadados puramente descritivos, considerando também na representação de cada elemento, os metadados de preservação e administrativos.

A OAI-ORE pode ser adotada para encapsular recursos distribuídos. É interessante observar que o foco da OAI-ORE está mais no recurso e menos nos repositórios.

O formato de serialização primário para mapear o recurso é o perfil da unidade atômica expressa em RDF (*Resource Description Framework*)

PUBLICAÇÕES AMPLIADAS: INICIATIVAS EXISTENTES

Content | Relations | Comments | Source

DRIVER 2 - Demonstrator

Parents: none

EP: Journalists in The Netherlands

An enhanced publication about journalists in The Netherlands. It is the result of a reserach to journalists and current social changes such as multiculturalism, the use of internet and the integration of news and entertainment. This enhanced publication consists of tha data from the survey and the book. The book itself is available as a complete file and a collection of separate chapters.

Author	M.J.P. Deuze
Persistent ID	urn:x-driver:378001
Publisher	Data Archiving and Networked Services
Created	2004-10-08
Modified	2008-10-17
Rights	OpenAccess

Journalists in The Netherlands

This book provides a profile of journalists in The Netherlands, looking at their professional role in contemporary multicultural, digital and commercial society. The analysis furthermore compares Dutch, Australian, German, British and American journalists, offering a comprehensive international overview of this cornerstond profession in todays' democracy.

Author	M.J.P. Deuze
Persistent ID	urn:x-driver:378004
Type	doctoralThesis
Publisher	University of Amsterdam
Created	2002-03-13
Modified	2002-03-13
Rights	RestrictedAccess

Open

Resources

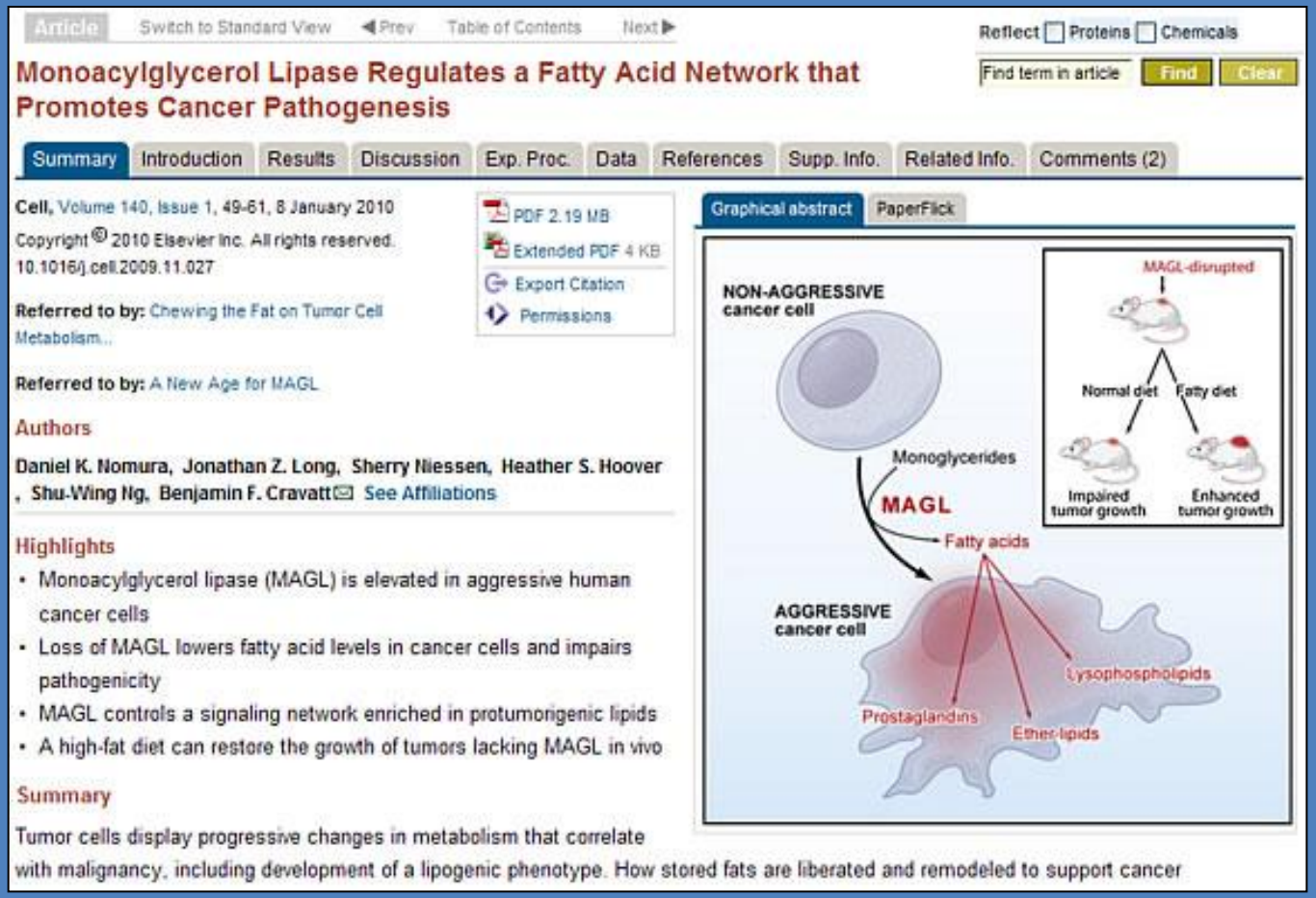
Surveydata 'Journalism in the Nether...

Author	M.J.P. Deuze
Persistent ID	urn:x-driver:378002
Type	dataset
Publisher	Data Archiving and Networked Services
Modified	2008-09-11
Rights	OpenAccess

Open

***driver2**.dans.knaw.nl/**demonstrator**/html/*

NEW Cell USER INTERFACE



The new **Cell** user interface. Outstanding features are the explicit layering and internal **integration of the information displayed**. In addition to the textual summary a **graphical abstract** is included, together with a **video** on which the authors introduce their findings. The tabs on top correspond with the **regular division of a scientific article**, while sections like **'Results'** and **'Data'** provide ample room for charts. **Figures** in the 'Data' section are **linked to the body of text so that each chart can be viewed in sync with its own context**. The **'Comments'** tab opens a blog-like web page for reactions

ABSTRACT GRÁFICO
VIDEO EM QUE OS AUTORES INTRODUZEM SEUS ACHADOS
RESULTADOS
DADOS – FIGURAS COM LINKS PARA O TEXTO
COMENTÁRIOS - BLOG

http://www.cell.com/abstract/S0092-8674%2809%2901439-1

NEW Cell USER INTERFACE

The new **Cell** user interface. Outstanding features are the explicit layering and internal **integration of the information displayed**. In addition to the textual summary a **graphical abstract** is included, together with a **video** in which the authors introduce their findings. The tabs on top correspond with the **regular division of a scientific article**, while sections like **'Results'** and **'Data'** provide ample room for charts. **Figures** in the 'Data' section are **linked to the body of text so that each chart can be viewed in sync with its own context**. The **'Comments'** tab opens a blog-like web page for reactions

ABSTRACT GRÁFICO
VIDEO EM QUE OS AUTORES INTRODUZEM SEUS ACHADOS
RESULTADOS
DADOS – FIGURAS COM LINKS PARA O TEXTO
COMENTÁRIOS - BLOG

http://www.cell.com/abstract/S0092-8674%2809%2901439-1

NEW Cell USER INTERFACE

Article | Switch to Standard View | Prev | Table of Contents | Next

Reflect ☐ Proteins ☐ Chemicals

Find term in article | Find | Clear

Monoacylglycerol Lipase Regulates a Fatty Acid Network that Promotes Cancer Pathogenesis

Summary | Introduction | Results | Discussion | Exp. Proc. | Data | References | Supp. Info. | Related Info. | Comments (2)

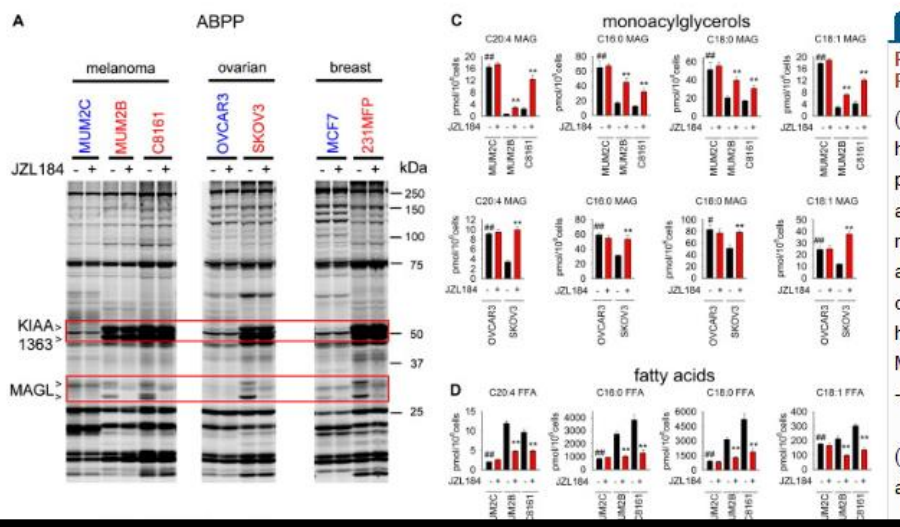
Cell, Volume 140, Issue 1, 49-61, 8 January 2010
Copyright © 2010 Elsevier Inc. All rights reserved.
10.1016/j.cell.2009.11.027

PDF 2.19 M | Extended | Export Cit

Graphical abstract | PaperFlick

Figures (13) | Excel (3) | Other (1)

Figure 1 | Figure 2 | Figure 3 | Figure 4 | Figure 5 | Figure 6 | Figure 7 | Figure S1 | Figure S2 | Figure S3 | Figure



Caption | Context in Article

Figure 1 MAGL Is Elevated in Aggressive Cancer Cells, where the Enzyme Regulates Monoacylglycerol and Free Fatty Acid Levels

(A) ABPP of serine hydrolase activities in nonaggressive (blue) and aggressive (red) human cancer cell lines. Serine hydrolase activities were labeled in whole-cell proteomes with the activity-based probe FP-rhodamine and detected by SDS-PAGE and in-gel fluorescence scanning (fluorescent gel shown in grayscale). Highlighted in red boxes are two enzymes, MAGL and KIAA1363 that are consistently elevated in aggressive versus nonaggressive cancer cells. Proteomes were also prepared from cancer cells pretreated with DMSO or the selective MAGL inhibitor JZL184 (1 μM, 4 hr) to confirm that the 33 and 35 kDa FP-rhodamine-labeled bands represented MAGL. See Table S1, Table S2, and Table S3 for a global analysis of FP-biotin-labeled serine hydrolase activities detected by ABPP in cancer cells.

(B) C20:4 MAG hydrolytic activity of cancer cells in the presence (red bars) or absence (black bars) of JZL184 (1 μM, 4 hr).

The new **Cell** user interface. Outstanding features are the explicit layering and internal **integration of the information displayed**. In addition to the textual summary a **graphical abstract** is ... with a **video** introduce ... on top **regular article, results'** sample ... in the **to the** each **chart with its comments'** web page

901439-1

ABST
VIDE
ACH
RESU
DADOS – FIGURAS COM LINKS PARA O TEXTO
COMENTÁRIOS - BLOG

PUBLICAÇÕES AMPLIADAS E WEB SEMÂNTICA

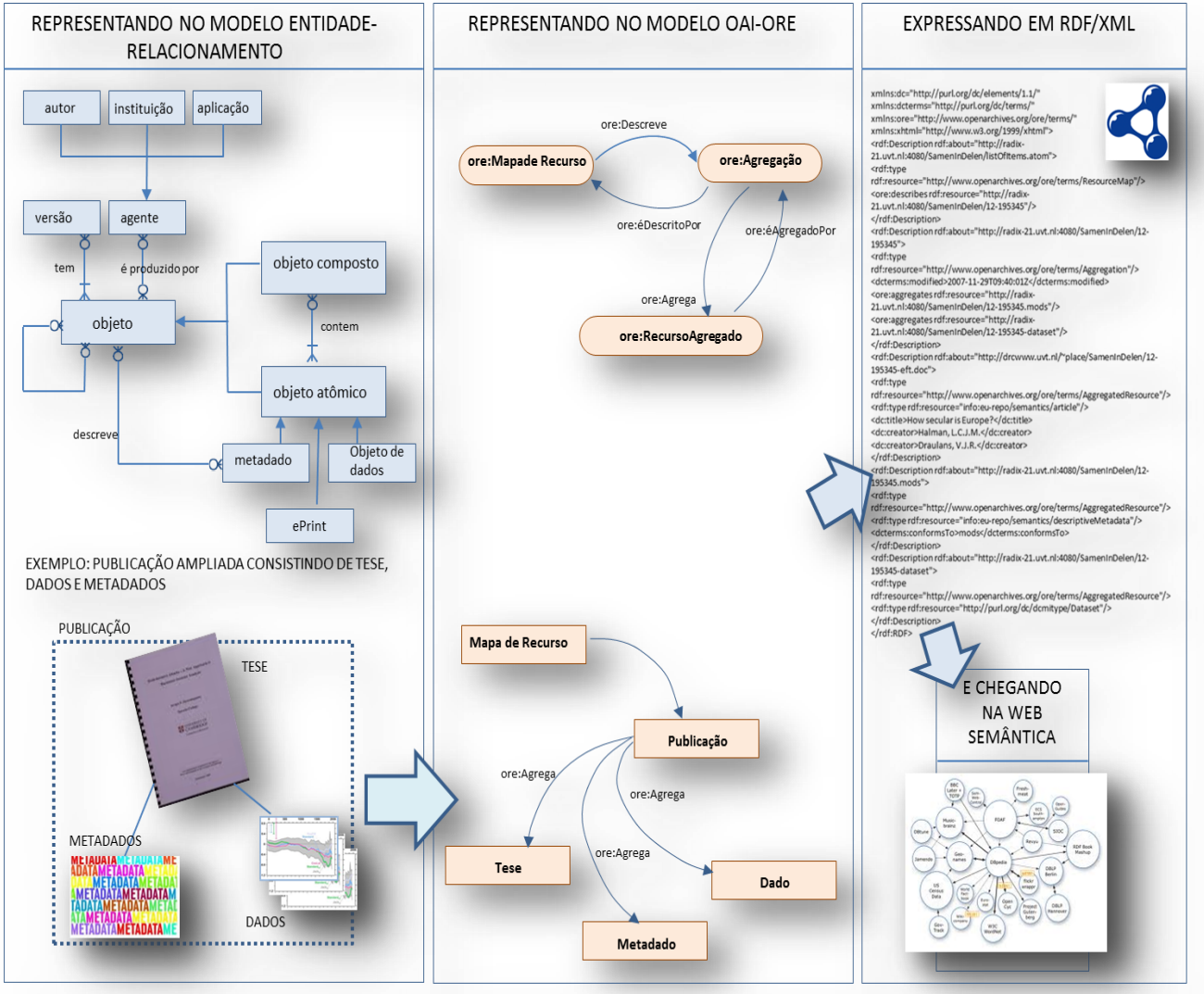
Publicações ampliadas consistem de dados e informações heterogêneas

A **Web semântica** é focada nas relações entre dados distribuídos

A **integração de dados** heterogêneos requer o uso de padrões.

Os **padrões do W3C** devem ser considerados na construção da publicação ampliada.

URI, XML, XML Schema e RDF, são exemplos de **recomendações do W3C que devem ser aplicadas à publicação ampliada**.



PUBLICAÇÕES AMPLIADAS: Vantagens

- A recuperação de e-*prints* e dados de forma integrada poupa o tempo gasto na busca por informações

- Os dados são contextualizados pela própria publicação, mantendo seu sentido original, mas podendo ser reutilizados para novas pesquisas e reinterpretados em outros contextos;

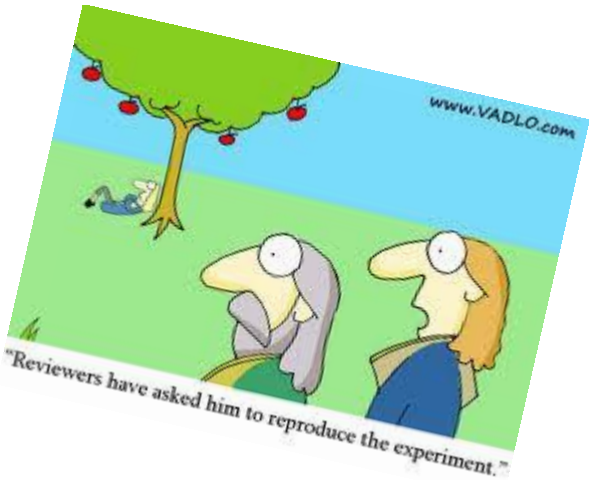
- Os dados passam a ser preservados para uso e reuso futuros, assim como as publicações tradicionais;

- Os dados podem ser citados isoladamente e como um todo, de acordo com a finalidade da pesquisa;

- Os dados permitem verificar a veracidade da pesquisa, apoiam os processos de replicação e dão mais confiabilidade aos seus resultados;

- As publicações ampliadas podem melhorar e tornar mais abertos e interativos os métodos de revisão por pares.

- A possibilidade de reuso dos dados financiados por recursos públicos abre possibilidades de encurtamento do ciclo de comunicação científica e de otimização dos recursos destinados à pesquisa.



PUBLICAÇÕES AMPLIADAS: DESAFIOS

- As **relações** entre coleções de dados e e-prints são muito **elementares** e precisam carregar mais pluralidade de significado;
- Necessidade de desenvolver **mecanismos de recuperação**, apresentação (renderização) e navegação dos dados e suas ligações, através de mecanismos visuais que usem **taxonomia ou outras ferramentas de visualização semântica**;
- **Confidencialidade x acesso aos dados**: a questão de propriedade intelectual das agregações e de suas partes ainda está aberta, além das questões éticas
- **Ausência de flexibilidade** ou extensibilidade, que permitam o **cientista escolher** interativamente um conjunto de recursos gerados a partir de uma experiência ou de investigação;
- Os sistemas de publicação existentes apoiam o **acesso livre** à publicações, mas não **aos dados**;
- Formatos de **metadados** para a **gestão da preservação** do significado e da estrutura de coleções de **dados de pesquisa** em ambientes de repositórios de dados;
- Interoperabilidade e integração com sistemas convencionais e sistemas avançados, como **web semântica** e **linked data**;
- Uso de ferramentas como **ontologias** que permitam não apenas a recuperação dos dados, mas o **processamento de conhecimento** de forma inteligente;
- Impactos no ciclo de comunicação científica, compartilhamento de recurso cooperativos, entre outros.

PUBLICAÇÕES AMPLIADAS: REFERÊNCIAS

AALBERSBERG, Ijsbrand Jan; DUNHAN, Judson; KOERS, Hylke. **Connecting scientific articles with research data**: new directions in online scholarly publishing. 2011. Disponível em: <http://www2.nict.go.jp/isd/ISDS-contents/wds-kyoto-2011.org/pdf/IS704.pdf >Acesso em: 19 maio 2013.

KIRCZ, Joost G. New practices for electronic publishing 1: Will the scientific paper keep its form? **Learned Publishing**, v. 14, n. 4, p. 265-272, 2001. Disponível em: <www.kra.nl/Webs**ite**/Artikelen/learned-publ1.pdf> Acesso em: 11 ago. 2013

KIRCZ, Joost G. New practices for electronic publishing 2: New forms of the scientific paper. **Learned Publishing**, v. 15, n. 1, p. 27-32, 2002. Disponível em: <www.kra.nl/Webs**ite**/Artikelen/Learnedpubl2.pdf > Acesso em: 11 ago. 2013

LAGOZE, Carl; VAN DE SOMPEL, Herbert. Compound Information Objects: The OAI-ORE Perspective. **Open Archives Initiative.** 2007. Disponível em: <http://www. openarchives. org/ore/documents/CompoundObjects-200705.html> Acesso em: 10 ago. 2013.

SAYAO, L. F.2012SAYAO, L. F. ; SALES, L. F.. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade** (UFPB. Online), v. 22, p. 1, 2012. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/12224/8586> Acesso em: 11 ago. 2013

SERINGHAUS, Michael R.; GERSTEIN, Mark B. Publishing perishing? Towards tomorrow's information architecture. **Bmc Bioinformatics**, v. 8, n. 1, p. 17, 2007. Disponível em: < www.biomedcentral.com/1471-2105/8/17 > Acesso em: 11 ago. 2013

VAN DEN HEUVEL, Charles; RAYWARD, W. Boyd. Facing interfaces: Paul Otlet's visualizations of data integration. **Journal of the American Society for Information Science and Technology**, v. 62, n. 12, p. 2313-2326, 2011. Disponível em: < http://onlinelibrary.wiley.com/doi/10.1002/asi.21607/abstract;jsessionid=B9BB9F523488602613B93F6D07543A08.d03t02?deniedAccessCustomisedMessage=&userIsAuthenticated=false > Acesso em: 11 ago. 2013

VERHAAR, Peter**. Report on object models and functionalities.** DRIVER II, 2007. Disponível em: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report_on_Object_Models_and_Functionalities.pdf?sequence=2>. Acesso em: 20 dez. 2011.

VERNOOY-GERRITSEN, Marjan (ed.). **Enhanced publications: linking publications and research data in digital repositories**. Amsterdam University Press, 2009. Driver and Surf studies. Disponível em: < dare.uva.nl/document/150723> Acesso em: 11 ago.

VERNOOY-GERRITSEN, Marjan (ed). **Emerging standards for enhanced publications and repository technology: survey on technology**. Amsterdam University Press, 2009a. Driver and Surf studies. Disponível em: <dare.uva.nl/aup/en/record/316870 > Acesso em: 11 ago. 2013

WOUTERSEN-WINDHOUWER, Saskia; BRANDSMA, Renze. **Enhanced publications**: state of the art. Part 1, p.19-91, 2009. Disponível em: <www.driver-repository.eu/**Enhanced**-**Publications**.html > Acesso em: 11 ago. 2013