Linked Data at Elsevier

DCMI Science & Metadata Community Workshop

2011-09-22

Who we are

Mike Lauruhn

Disruptive Technology Director, Elsevier Labs

- Part of Information Technology (Shared Services)
- Formerly: Librarian, Cataloger, Taxonomy & Metadata Consultant

Véronique Malaisé

Head of Taxonomy Center, Content Enrichment Center

- Electronic Production Department (Operations)
- Formerly: Post Doctoral researcher at the Free University Amsterdam, in Natural Language Processing aspects of Semantic Web projects

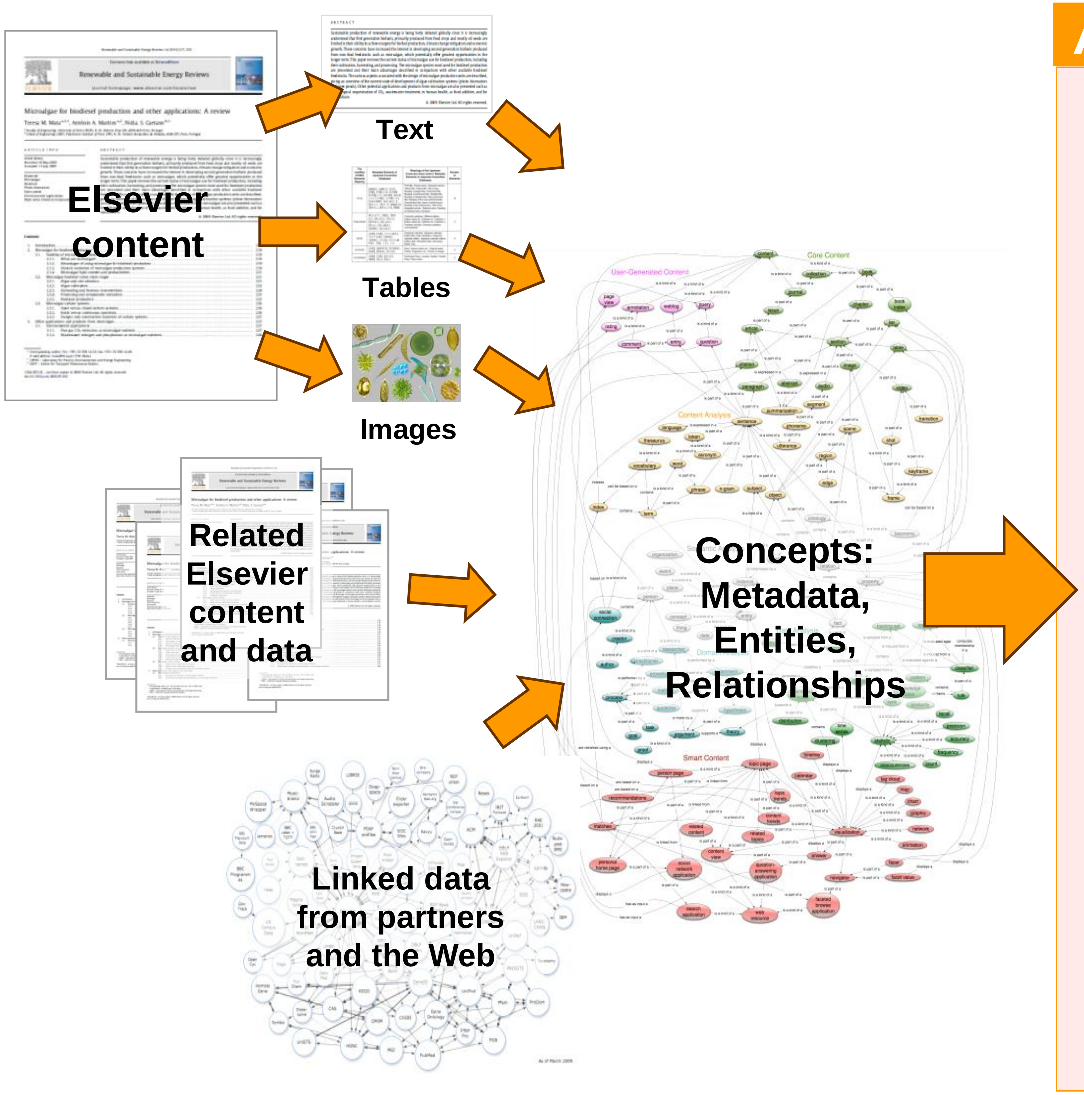
Today’s Session

Smart Content Overview

- Linked Data Repository
- Satellites & RDF
- Taxonomy

Some examples

Smart Content: Semantic Enhancements for Scientific Publishing


Elsevier content
Text
Tables
Images
Related Elsevier content and data
Linked data from partners and the Web
User-Generated Content
Core Content
Content Analysis
Smart Content
Concepts: Metadata, Entities, Relationships

Applied Smart Content

Better discovery

- Faceted search & browse
- Ontology-driven navigation
- Task-specific results
- Personalized/localized results
- Question answering

Better understanding

Countries - Fertility rate, total (births per woman)
Life expectancy at birth, total (years)
1960

- Tag clouds
- Heatmaps
- Streamgraphs
- Scatterplots
- Time series
- Animations

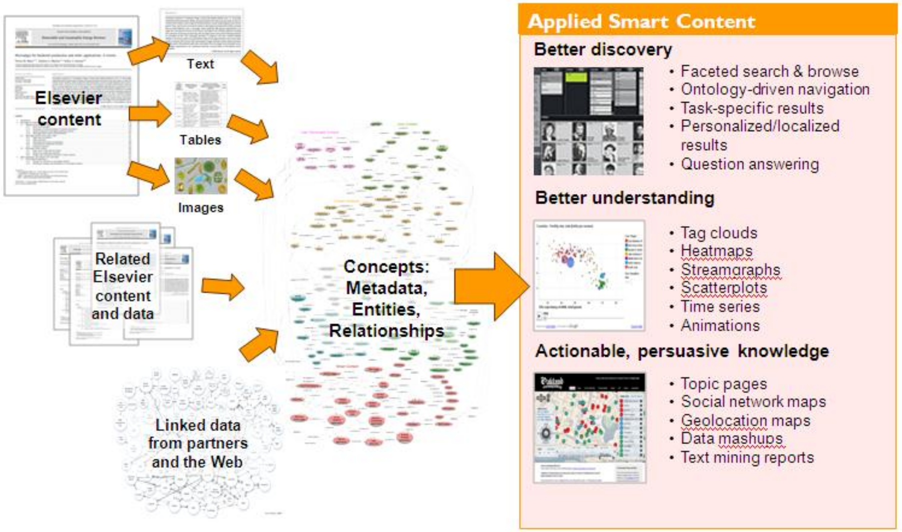
Actionable, persuasive knowledge

- Topic pages
- Social network maps
- Geolocation maps
- Data mashups
- Text mining reports

The Challenge

How to do semantic enhancement at scale for STM publishing?

- In harmony with our culture and legacy,
- Across the breadth of our content,
- Within an ecosystem of authors, institutions, libraries, repositories, publishers, content suppliers, and funding agencies.



Elsevier content

Text

Tables

Images

Related Elsevier content and data

Linked data from partners and the Web

Concepts: Metadata, Entities, Relationships

Applied Smart Content

Better discovery

- Faceted search & browse
- Ontology-driven navigation
- Task-specific results
- Personalized/localized results
- Question answering

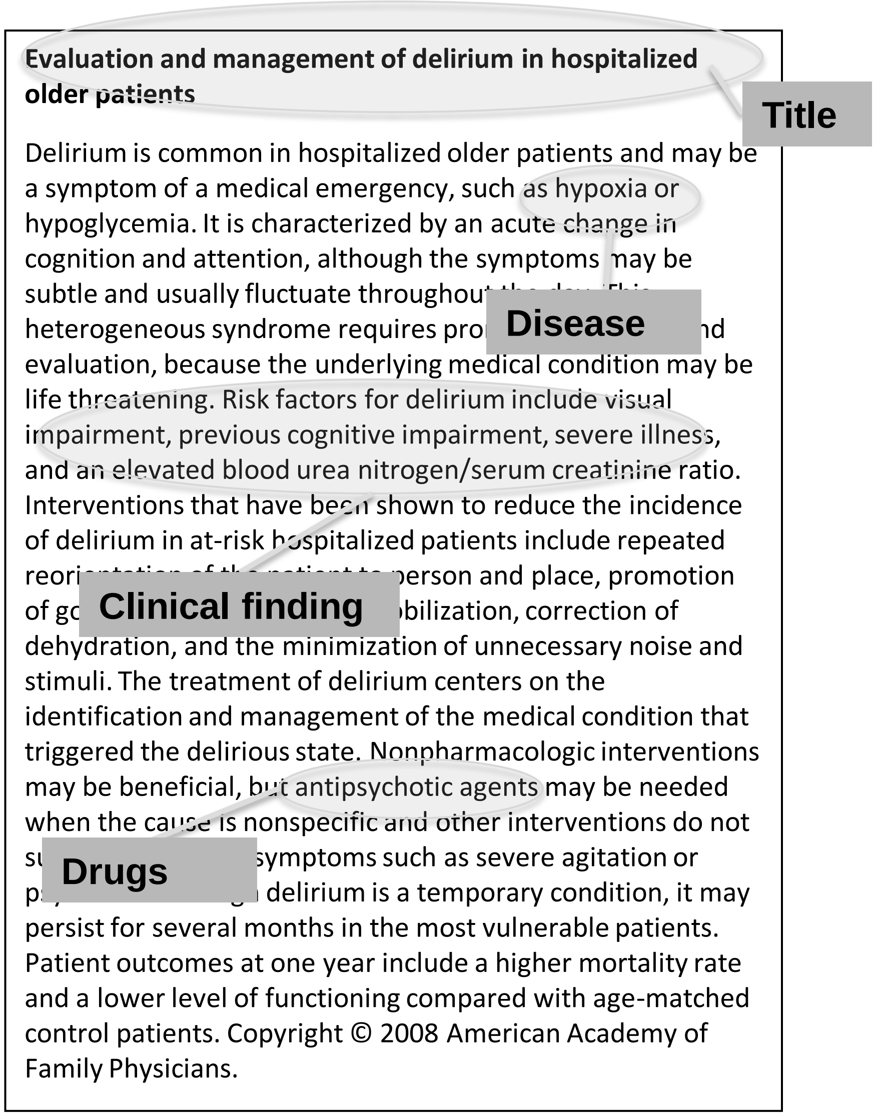
Better understanding

- Tag clouds
- Heatmaps
- Streamgraphs
- Scatterplots
- Time series
- Animations

Actionable, persuasive knowledge

- Topic pages
- Social network maps
- Geolocation maps
- Data mashups
- Text mining reports

Content Enrichment


Title
Disease
Clinical finding
Drugs

Evaluation and management of delirium in hospitalized older patients

Delirium is common in hospitalized older patients and may be a symptom of a medical emergency, such as hypoxia or hypoglycemia. It is characterized by an acute change in cognition and attention, although the symptoms may be subtle and usually fluctuate throughout the day. This heterogeneous syndrome requires pro[illegible]nd evaluation, because the underlying medical condition may be life threatening. Risk factors for delirium include visual impairment, previous cognitive impairment, severe illness, and an elevated blood urea nitrogen/serum creatinine ratio. Interventions that have been shown to reduce the incidence of delirium in at-risk hospitalized patients include repeated reorientation of the patient to person and place, promotion of go[illegible]bilization, correction of dehydration, and the minimization of unnecessary noise and stimuli. The treatment of delirium centers on the identification and management of the medical condition that triggered the delirious state. Nonpharmacologic interventions may be beneficial, but antipsychotic agents may be needed when the cause is nonspecific and other interventions do not su[illegible] symptoms such as severe agitation or p[illegible] delirium is a temporary condition, it may persist for several months in the most vulnerable patients. Patient outcomes at one year include a higher mortality rate and a lower level of functioning compared with age-matched control patients. Copyright © 2008 American Academy of Family Physicians.

- Concepts are identified in text, compared to Concepts and relations in a controlled vocabulary or semantic model, and stored as RDF in annotation files
- The storage mechanism for this information is the Linked Data Repository (LDR)

Linked Data Repository

Infrastructure that supports storing and linking of semantic annotation of content, supports apps that enable discovery and semantic search via an API.

- Used to store and structure data & relations derived from content.

- Interlinks data with other related sources of content (documents, sections of documents, data, multimedia).

- Provides service layer APIs for ease of interaction with both suppliers and internal processes.

Optimized for high-volume read and write access to RDF graphs.

Satellites

Linked Data compliant format that is used to capture, store and expose metadata objects.

Standards based, including DCMI, SKOS, and SWAN

Contains:

Metadata for the resource that the satellite is related to.

Provenance information of the metadata.

(Provenance & version based on Harvard SWAN standard)

Configurable for use case-specific information:

- Relevance and confidence numbers

  e.g. Medical integrity or tagging score

- Document fragment identification

EMMeT (Elsevier Merged Medical Taxonomy)

EMMeT Background

- Founded upon UMLS and utilizes standard vocabularies including, MeSH, SNOMED-CT, RxNorm, and ICD-9
- Structured around major classes (semantic types) including: diseases, procedures, drugs, symptoms, anatomy
- > 600,000 preferred terms
- ~ 2 million synonyms

Ontological Support

- Validate and update relationships against Elsevier sources
- Initial focus on disease, symptoms, and treatments

Internationalization

- Regionalization of EMMeT for key countries

EMMeT Components

Taxonomy relations:

- BT, NT
- Variants: synonyms, acronyms, abbreviations, other term types
- Scope notes & definitions

Ontological relations:

- Priority relations are declared between classes

Disease has Associated Drug
Disease has Treatment
Disease has Symptoms

SKOS-XL representation stored in the Linked Data Repository to support annotation formats for Elsevier content

EMMeT Components

Each Concept Includes	Example
Preferred Term	Coronary Artery Bypass Surgery
Class	Procedures
Variants (synonyms, jargon & vocabulary terms)	CABG; Bypass anastomosis for heart revascularization
Parent Relationships	Cardiovascular Surgery Procedures
Child Relationships	Bypass of Three Coronary Arteries
UMLS code	ICD9CM-2010, 36.1

EMMeT Components

Classes have defined relations:

- Disease has Associated Drug
- Disease has Treatment
- Disease has Symptom

Disease	isTreatmentProcedure	Procedure
Coronary Artery Bypass	treatmentProcedureFor	Acute Coronary Syndrome
Coronary Artery Bypass	treatmentProcedureFor	Acute Myocardial Infarction
Coronary Artery Bypass	treatmentProcedureFor	Angina Pectoris

EMMeT Example

```
<skosxl:literalForm xml:lang="en-US">Diabetes Mellitus</skosxl:literalForm>
<ebs:usageFlag rdf:resource="http://data.elsevier.com/EMMeT/Flags/MedicalName"/>
...
<skosxl:literalForm xml:lang="en-US">Diabetes</skosxl:literalForm>
<ebs:usageFlag  rdf:resource="http://data.elsevier.com/EMMeT/Flags/ConsumerFriendlyName"/>
...
<skos:notation rdf:datatype="http://data.elsevier.com/vocabulary/EMMeT">177824</skos:notation>
<skos:notation rdf:datatype="http://dbpedia.org/resource/UMLS">C0011849</skos:notation>
...

<rdf:type rdf:resource="http://data.elsevier.com/EMMeT/SemTypes/DiseaseOrSyndrome" />

<skos:broader rdf:ID="Relation-34359"
     rdf:resource="http://data.elsevier.com/vocabulary/EMMeT/Concept/48543"/>
<!-- BT: Disorders of endocrine system -->
...
<skos:narrower rdf:ID="Relation-9812"
     rdf:resource="http://data.elsevier.com/vocabulary/EMMeT/34565"/>
...
<emsem:hasSymptom rdf:ID="Relation-99999"
     rdf:resource="http://data.elsevier.com/vocabulary/EMMeT/Concept/53425"/>
...
```

Disorders of endocrine system

Abnormal metabolic state in diabetes mellitus

Abnormal Sense of Taste

Today’s Session

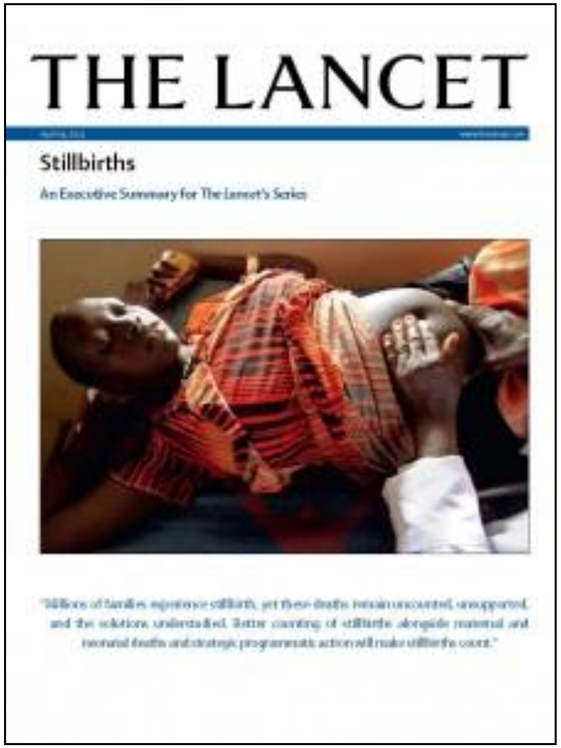
Smart Content Overview

- The Challenge
- Linked Data Repository
- Satellites & RDF
- Taxonomy

Some examples

Prototype for *The Lancet*

Special feature associated with *The Lancet* special issue: “Stillbirths” (Vol 377; Number 9774; April 14, 2011)



THE LANCET

Stillbirths

An Executive Summary for The Lancet's Series

“Millions of families experience stillbirth, yet these deaths remain uncounted, unsupported, and the solutions understudied. Better counting of stillbirths alongside maternal and neonatal deaths and strategic programmatic action will make stillbirths count.”

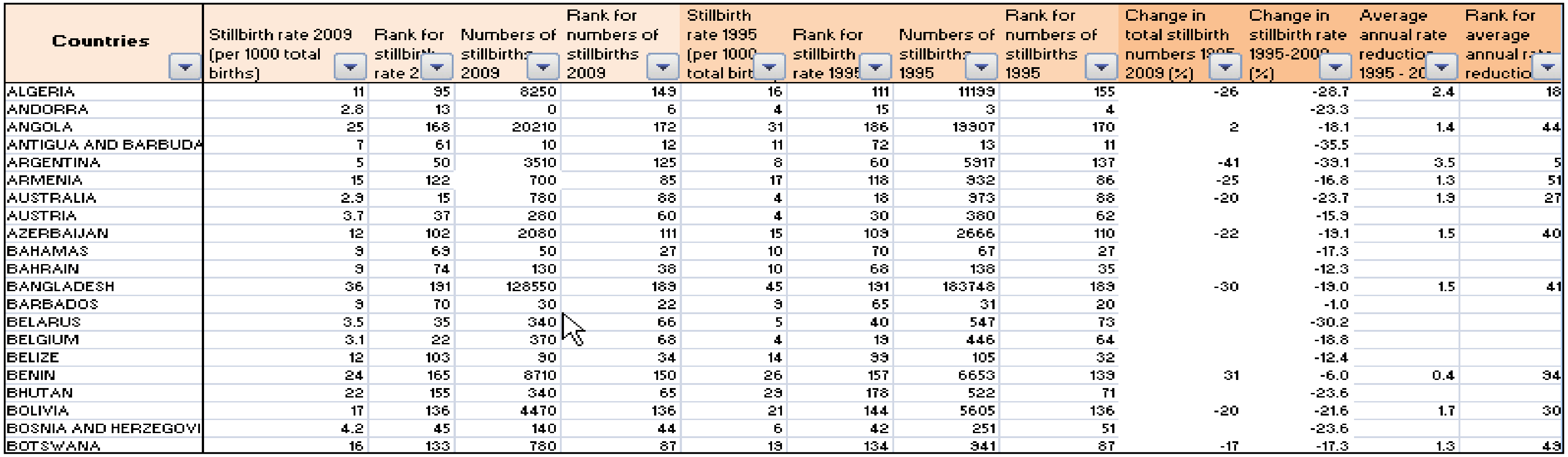
Creation LDR-enabled interactive application using:

- *The Lancet* content
- Datasets from *The Lancet* editorial research
- Datasets from The World Bank
- Subject (EMMeT) tagging from vendor
- Map

Prototype for *The Lancet*

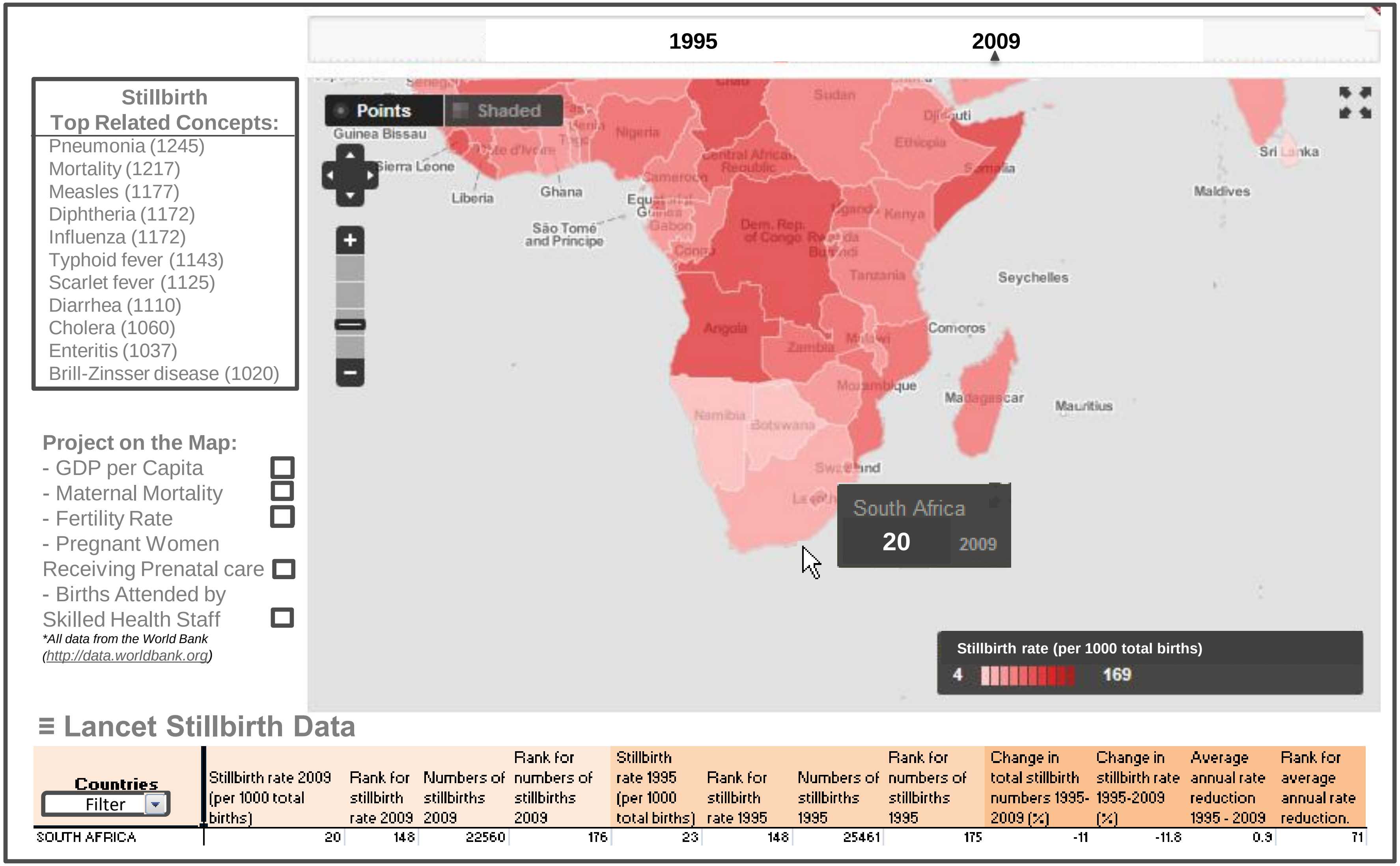
The Lancet and World Bank datasets loaded into the LDR as triples.

EMMeT concept tagging results loaded into LDR.



Countries	Stillbirth rate 2009 (per 1000 total births)	Rank for stillbirt rate 2	Numbers of stillbirth 2009	Rank for numbers of stillbirths 2009	Stillbirth rate 1995 (per 1000 total birt	Rank for stillbirth rate 1995	Numbers of stillbirth 1995	Rank for numbers of stillbirths 1995	Change in total stillbirth numbers 1 2009 (%)	Change in stillbirth rate 1995-200 (%)	Average annual rate reductio 1995 - 20	Rank for average annual r reductio
ALGERIA	11	95	8250	149	16	111	11199	155	-26	-28.7	2.4	18
ANDORRA	2.8	13	0	6	4	15	3	4		-23.3		
ANGOLA	25	168	20210	172	31	186	19907	170	2	-18.1	1.4	44
ANTIGUA AND BARBUDA	7	61	10	12	11	72	13	11		-35.5		
ARGENTINA	5	50	3510	125	8	60	5917	137	-41	-39.1	3.5	5
ARMENIA	15	122	700	85	17	118	832	86	-25	-16.8	1.3	51
AUSTRALIA	2.9	15	780	88	4	18	973	88	-20	-23.7	1.9	27
AUSTRIA	3.7	37	280	60	4	30	380	62		-15.9		
AZERBAIJAN	12	102	2080	111	15	109	2666	110	-22	-19.1	1.5	40
BAHAMAS	9	69	50	27	10	70	67	27		-17.3		
BAHRAIN	9	74	130	38	10	68	138	35		-12.3		
BANGLADESH	36	191	128550	189	45	191	183748	189	-30	-19.0	1.5	41
BARBADOS	9	70	30	22	9	65	31	20		-1.0		
BELARUS	3.5	35	340	66	5	40	547	73		-30.2		
BELGIUM	3.1	22	370	68	4	19	446	64		-18.8		
BELIZE	12	103	90	34	14	99	105	32		-12.4		
BENIN	24	165	8710	150	26	157	6653	139	31	-6.0	0.4	94
BHUTAN	22	155	340	65	29	178	522	71		-23.6		
BOLIVIA	17	136	4470	136	21	144	5605	136	-20	-21.6	1.7	30
BOSNIA AND HERZEGOVI	4.2	45	140	44	6	42	251	51		-23.6		
BOTSWANA	16	133	780	87	19	134	341	87	-17	-17.3	1.3	49

Lancet Prototype – Mock-up



Stillbirth
Top Related Concepts:

- Pneumonia (1245)
- Mortality (1217)
- Measles (1177)
- Diphtheria (1172)
- Influenza (1172)
- Typhoid fever (1143)
- Scarlet fever (1125)
- Diarrhea (1110)
- Cholera (1060)
- Enteritis (1037)
- Brill-Zinsser disease (1020)

Project on the Map:

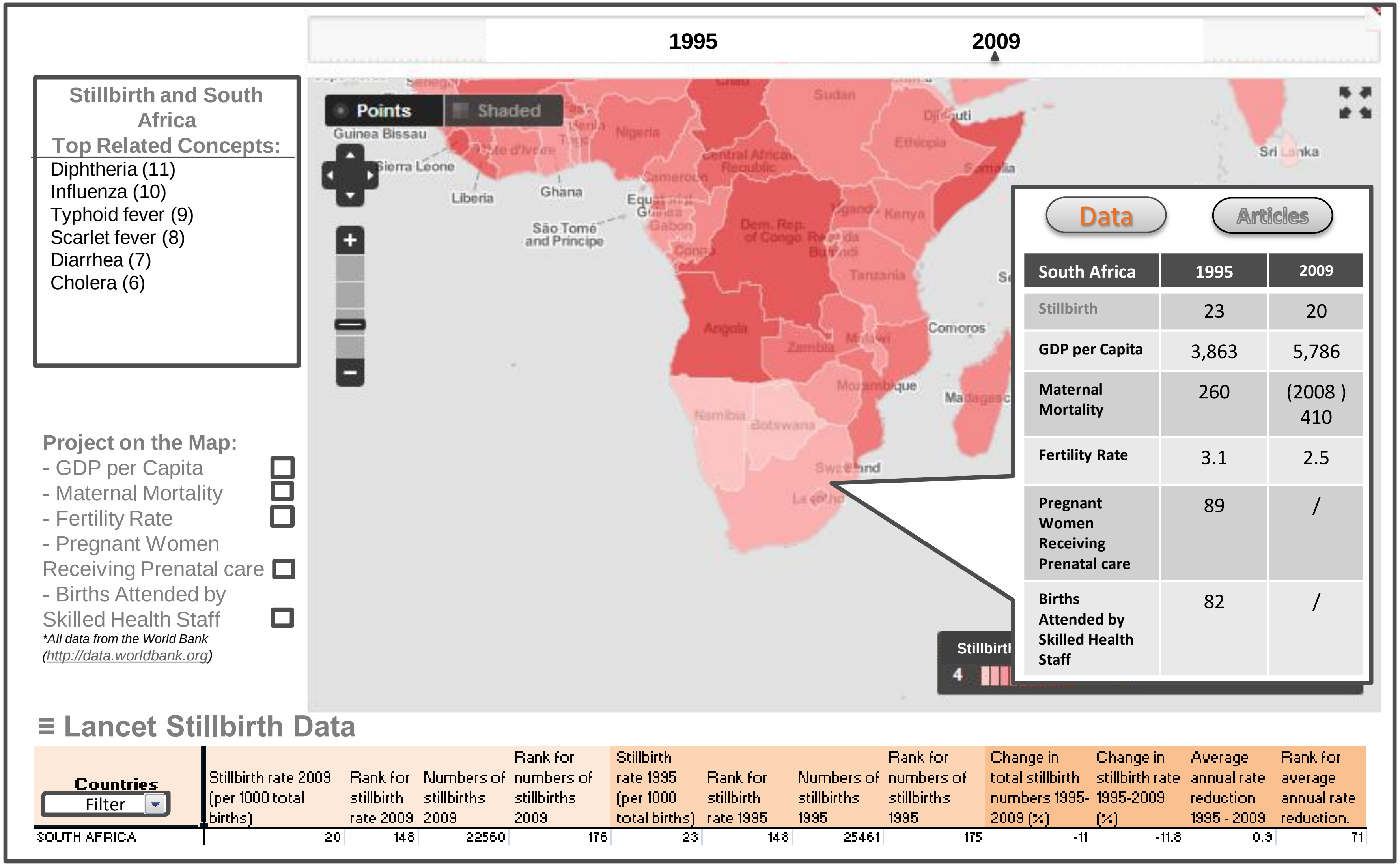
- GDP per Capita □
- Maternal Mortality □
- Fertility Rate □
- Pregnant Women Receiving Prenatal care □
- Births Attended by Skilled Health Staff □

**All data from the World Bank (http://data.worldbank.org)*

≡ Lancet Stillbirth Data

Countries Filter ▾	Stillbirth rate 2009 (per 1000 total births)	Rank for stillbirth rate 2009	Numbers of stillbirths 2009	Rank for numbers of stillbirths 2009	Stillbirth rate 1995 (per 1000 total births)	Rank for stillbirth rate 1995	Numbers of stillbirths 1995	Rank for numbers of stillbirths 1995	Change in total stillbirth numbers 1995-2009 (%)	Change in stillbirth rate 1995-2009 (%)	Average annual rate reduction 1995 - 2009	Rank for average annual rate reduction.
SOUTH AFRICA	20	148	22560	176	23	148	25461	175	-11	-11.8	0.9	71

Lancet Prototype – Mock-up



Stillbirth and South Africa
Top Related Concepts:

Diphtheria	(11)
Influenza	(10)
Typhoid fever	(9)
Scarlet fever	(8)
Diarrhea	(7)
Cholera	(6)

Project on the Map:

- GDP per Capita □
- Maternal Mortality □
- Fertility Rate □
- Pregnant Women Receiving Prenatal care □
- Births Attended by Skilled Health Staff □

*All data from the World Bank (http://data.worldbank.org)

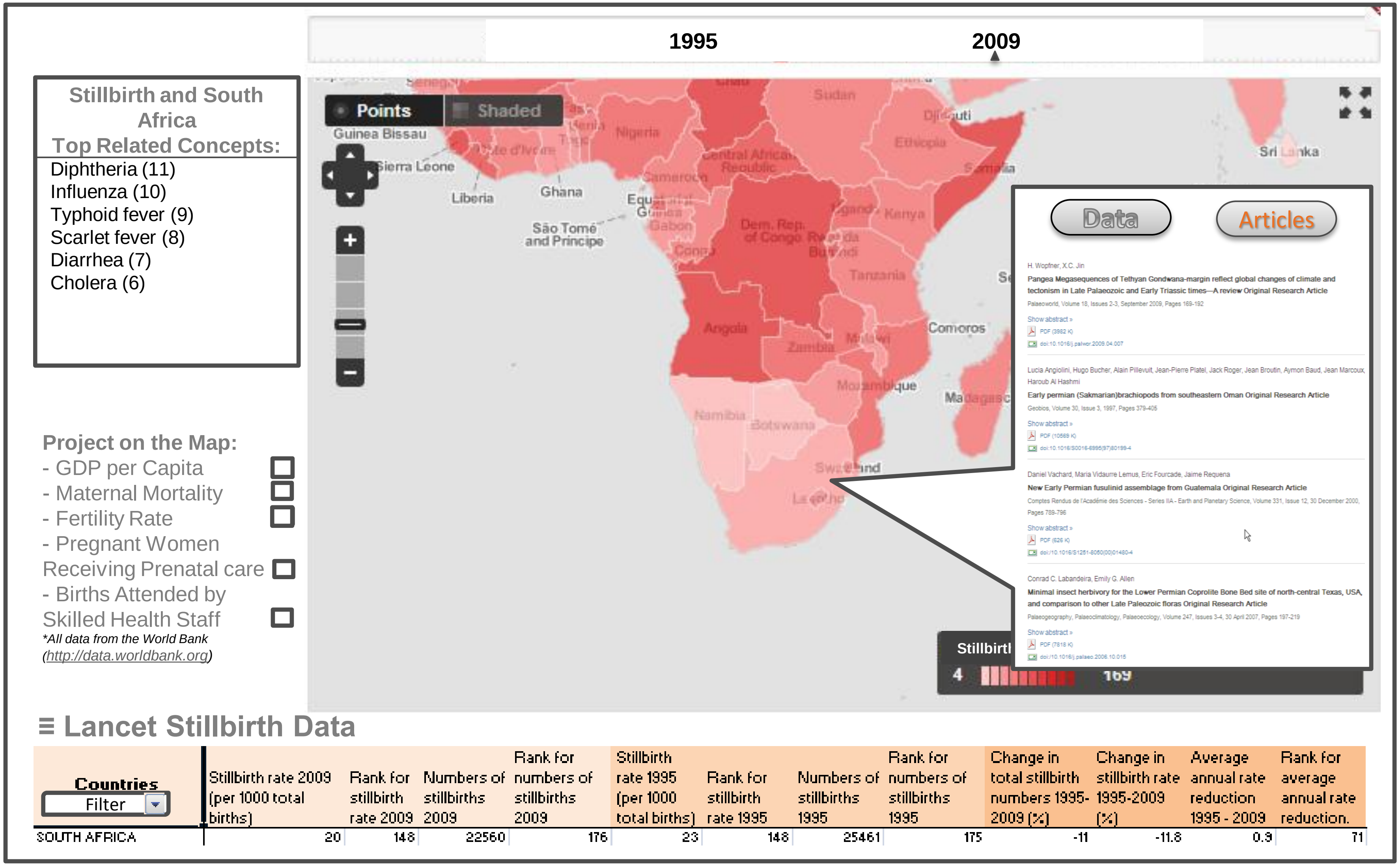
Data | Articles

South Africa	1995	2009
Stillbirth	23	20
GDP per Capita	3,863	5,786
Maternal Mortality	260	(2008) 410
Fertility Rate	3.1	2.5
Pregnant Women Receiving Prenatal care	89	/
Births Attended by Skilled Health Staff	82	/

≡ Lancet Stillbirth Data

Countries Filter	Stillbirth rate 2009 (per 1000 total births)	Rank for stillbirth rate 2009	Numbers of stillbirths 2009	Rank for numbers of stillbirths 2009	Stillbirth rate 1995 (per 1000 total births)	Rank for stillbirth rate 1995	Numbers of stillbirths 1995	Rank for numbers of stillbirths 1995	Change in total stillbirth numbers 1995-2009 (%)	Change in stillbirth rate 1995-2009 (%)	Average annual rate reduction 1995 - 2009	Rank for average annual rate reduction.
SOUTH AFRICA	20	148	22560	176	23	148	25461	175	-11	-11.8	0.9	71

Lancet Prototype – Mock-up



Stillbirth and South Africa

Top Related Concepts:

- Diphtheria (11)
- Influenza (10)
- Typhoid fever (9)
- Scarlet fever (8)
- Diarrhea (7)
- Cholera (6)

Project on the Map:

- GDP per Capita □
- Maternal Mortality □
- Fertility Rate □
- Pregnant Women Receiving Prenatal care □
- Births Attended by Skilled Health Staff □

**All data from the World Bank (http://data.worldbank.org)*

Data | Articles

H. Wopfner, X.C. Jin
Pangea Megasequences of Tethyan Gondwana-margin reflect global changes of climate and tectonism in Late Palaeozoic and Early Triassic times—A review Original Research Article
Palaeoworld, Volume 18, Issues 2-3, September 2009, Pages 169-192
Show abstract »
PDF (3982 K)
doi:10.1016/j.palwor.2009.04.007

Lucia Angiolini, Hugo Bucher, Alain Pillevuit, Jean-Pierre Platel, Jack Roger, Jean Broutin, Aymon Baud, Jean Marcoux, Haroub Al Hashmi
Early permian (Sakmarian)brachiopods from southeastern Oman Original Research Article
Geobios, Volume 30, Issue 3, 1997, Pages 379-405
Show abstract »
PDF (10569 K)
doi:10.1016/S0016-6995(97)80199-4

Daniel Vachard, Maria Vidaurre Lemus, Eric Fourcade, Jaime Requena
New Early Permian fusulinid assemblage from Guatemala Original Research Article
Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science, Volume 331, Issue 12, 30 December 2000, Pages 789-796
Show abstract »
PDF (626 K)
doi:/10.1016/S1251-8050(00)01480-4

Conrad C. Labandeira, Emily G. Allen
Minimal insect herbivory for the Lower Permian Coprolite Bone Bed site of north-central Texas, USA, and comparison to other Late Paleozoic floras Original Research Article
Palaeogeography, Palaeoclimatology, Palaeoecology, Volume 247, Issues 3-4, 30 April 2007, Pages 197-219
Show abstract »
PDF (7818 K)
doi:/10.1016/j.palaeo.2006.10.015

≡ Lancet Stillbirth Data

Countries Filter ▾	Stillbirth rate 2009 (per 1000 total births)	Rank for stillbirth rate 2009	Numbers of stillbirths 2009	Rank for numbers of stillbirths 2009	Stillbirth rate 1995 (per 1000 total births)	Rank for stillbirth rate 1995	Numbers of stillbirths 1995	Rank for numbers of stillbirths 1995	Change in total stillbirth numbers 1995-2009 (%)	Change in stillbirth rate 1995-2009 (%)	Average annual rate reduction 1995 - 2009	Rank for average annual rate reduction.
SOUTH AFRICA	20	148	22560	176	23	148	25461	175	-11	-11.8	0.9	71

‘Data to Semantics’ Research Project

Background: Proper implementation of clinical decision support systems (CDS) can:

- Reduce errors in medical care
- Bring research results faster to the front-line clinician
- Significantly improve patient outcome.

Overall budget:

5.2 M Euros over 4 years, paid by Dutch government (Ministry of Economic Affairs, Innovation and Agriculture)

Partnership:

Elsevier, Philips, Free University Amsterdam

‘Data to Semantics’ Research Project

Requirements:

Be able to answer complex questions

Aggregate data from multiple sources, combining complex patient specific data with information from external sources

Be semantically aware

Be continually updated with the latest validated research results.

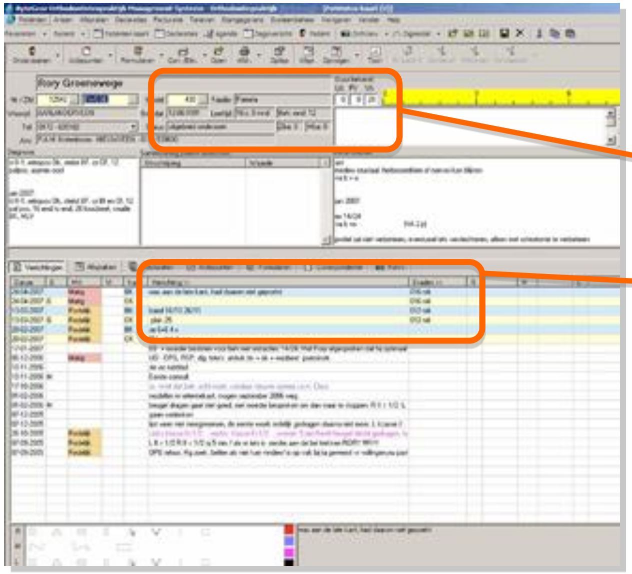
Components:

Flexible frameworks supporting the development of such applications

Integration of relevant, high quality content

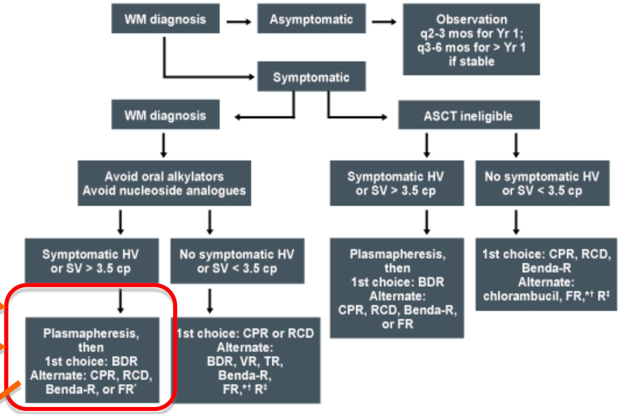
Tools enabling the extraction and aggregation of such content.

‘Data to Semantics’ Research Project



A. Philips’ Electronic Patient Records

Step 1: Patient data + diagnosis link to Guideline recommendation


WM diagnosis
Asymptomatic
Observation q2-3 mos for Yr 1; q3-6 mos for > Yr 1 if stable
Symptomatic
WM diagnosis
ASCT ineligible
Avoid oral alkylators Avoid nucleoside analogues
Symptomatic HV or SV > 3.5 cp
No symptomatic HV or SV < 3.5 cp
Symptomatic HV or SV > 3.5 cp
No symptomatic HV or SV < 3.5 cp
Plasmapheresis, then 1st choice: BDR Alternate: CPR, RCD, Benda-R, or FR*
1st choice: CPR or RCD Alternate: BDR, VR, TR, Benda-R, FR,*† R‡
Plasmapheresis, then 1st choice: BDR Alternate: CPR, RCD, Benda-R, or FR
1st choice: CPR, RCD, Benda-R Alternate: chlorambucil, FR,*† R‡

Benda-R, bendamustine, rituximab; BDR, bortezomib, dexamethasone, rituximab; cp, centipoise; CPR, cyclophosphamide, prednisone, rituximab; FR, fludarabine, rituximab; HV, hyperviscosity; R, rituximab; RCD, rituximab, cyclophosphamide, dexamethasone; VR, bortezomib, rituximab.

*Would avoid as monotherapy in patients with hyperviscosity and with FcγRIIIA-158 F/F polymorphism.
†Consider an attenuated schedule for fludarabine administration (see text). For rituximab based therapies, consider maintenance rituximab in responding patients (see text). Clinical trials should be considered for patients whenever possible.

B. Elsevier-published Clinical Guideline

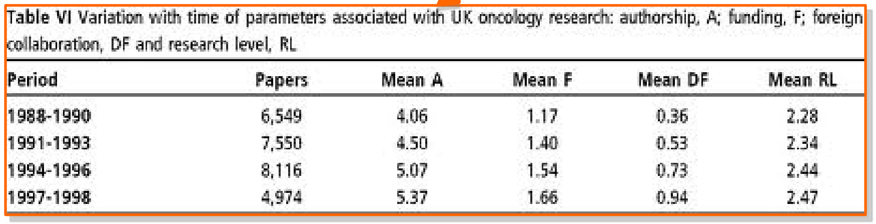


Table VI Variation with time of parameters associated with UK oncology research: authorship, A; funding, F; foreign collaboration, DF and research level, RL

Period	Papers	Mean A	Mean F	Mean DF	Mean RL
1988-1990	6,549	4.06	1.17	0.36	2.28
1991-1993	7,550	4.50	1.40	0.53	2.34
1994-1996	8,116	5.07	1.54	0.73	2.44
1997-1998	4,974	5.37	1.66	0.94	2.47

Step 2: Guideline recommendation links to research report/data

C. Elsevier (or other publisher’s) Research Report or Data

Contact Info

Mike Lauruhn	+1 619 206 6423	m.lauruhn@elsevier.com
Véronique Malaisé	+31 20 485 2254	v.malaise@elsevier.com
Alan Yagoda		a.yagoda@elsevier.com