ONLINE LEARNING AND LINKED DATA - LESSONS LEARNED AND BEST PRACTICES

Tutorial at the 23rd International World Wide Web Conference
WWW 2014
Co-organised by the EUCLID and LinkedUp projects
Linked Data as enabler for Open Education

Stefan Dietze
(L3S Research Center, DE, @stefandietze, http://purl.org/dietze)
Introduction

Research areas
- Web science, Information Retrieval, Semantic Web & Linked Data, data & knowledge integration (mapping, classification, interlinking)
- Application domains: education/TEL, Web archiving, ...

Some projects

(Linked) Web data & education
- Early activities in the field: tutorials, workshops, competitions (LILE – Linked Learning, LAK Data Challenge...)

See also: http://purl.org/dietze
Educational Data/Resources on the Web

(Open) Educational Resources

MOOCs

Distance Universities

Linked Open Data

Stefan Dietze

03/04/14

http://linkededucation.org

http://linkeduniversities.org
Linked Data: Pluto & the seven Dwarfs?

pluto dwarf

Planet Pluto is a dwarf planet located in the Kuiper Belt, beyond the orbit of Neptune. It is the most distant known dwarf planet in the Solar System.  The discovery of Pluto in 1930 was significant because it was the first object identified as a planet in over 50 years, and it expanded our understanding of the Solar System. Pluto is known for its distinctive features, including its large moon Charon and its seasonal changes due to its axial tilt. The dwarf planet status of Pluto was formally recognized by the International Astronomical Union in 2006.
“A little semantics goes a long way” (J. Hendler\(^1\))

**Semantic Web**
- Adding meaning through shared vocabularies and schemas (e.g., DBpedia)
- W3C standards RDF & SPARQL for data & knowledge representation and querying
- Persistent URIs to reference & interlink data on the Web

---

\(^1\) Hendler, J., The Dark Side of the Semantic Web, IEEE Intelligent Systems, Jan/Feb 2007
De-facto standard for sharing data on the Web

Vision: well connected graph of open Web data

350+ datasets and 50 billion triples in LOD Cloud alone

Other „incarnations“:
- Google Knowledge Graph
- Facebook Open Graph
- http://schema.org

„HTTP-accessibility“ (SPARQL, URI-dereferencing)
„Structure“ & „Semantics“ (=> shared/linked vocabularies)
  - „Interlinked“
  - „Persistent“
Actually, how “open” is open educational data?

(Open) Educational Resources

Distance Universities

Linked Open Data

MOOCs

Coursera

OpenLearn

OPEN COURSEWARE CONSORTIUM

OER

GLOBE

Ariane

Share

BBC

yovisto

dbpedia

Khan Academy

Open

openX

Udacity

World Wide Web

Athabasca University

Canada's Open University

http://linkededucation.org

http://linkeduniversities.org
Educational schemas/vocabularies – a case of „XKCD927“?

eg OER schemas, vocabularies, interfaces etc

How standards proliferate:

(See: A/C chargers, character encodings, instant messaging, etc)

SITUATION:
There are 14 competing standards.

14?! RIDICULOUS!
We need to develop ONE universal standard that covers everyone’s use cases.

YEAH!

SITUATION:
There are 15 competing standards.

https://xkcd.com/927/
Actually, how “open” is open educational data?
Just a short while ago...

Educational Data on the Web

- Open Educational Resource (OER) metadata & MOOC collections (e.g. OCW, OpenLearn, Merlot, Coursera)
- Competing Web interfaces (e.g. OAI-PMH, SOAP, REST)
- Competing metadata standards (e.g. IEEE LOM, ADL SCORM, DC...) & taxonomies & exchange formats (JSON, RDF, XML)
- Issues: heterogeneity & lack of interoperability
Actually, how “open” is open educational data?

Educational Data & Linked Data?

Educational Data on the Web

- **Open Educational Resource (OER) metadata & MOOC** collections (e.g. OCW, OpenLearn, Merlot, Coursera)
- Competing Web **interfaces** (e.g. OAI-PMH, SOAP, REST)
- Competing metadata **standards** (e.g. IEEE LOM, ADL SCORM, DC...) & **taxonomies & exchange formats** (JSON, RDF, XML)
- **Issues: heterogeneity** & lack of interoperability

Linked (Open) Data

- **Vision:** well connected **graph of open Web data**
- **W3C standards** (RDF, SPARQL) to expose data, URIs to interlink datasets
- Linking of schemas, vocabularies, data => interoperability

Pictures (c) by Paul Miller

"Stop hugging your data...”
Sir Tim Berners-Lee, 2009
A shorter while ago
Bringing OER into the LOD landscape

Automated data lifting & enrichment

- http://datahub.io/dataset/meducator
- Manually curated and automatically exported & enriched OER from variety of sources
- > 35000 triples so far
- > enrichments to DBpedia & Bioportal ontologies
- APIs used by several applications

Data tends to remain siloised.
Using LD in (medical) educational applications

Recommending educational resources across repositories

Using LD for integrating educational resources from a range of universities and OER repositories (e.g., Open Learn, Open Courseware, etc.).
“Structured” descriptions of OER?

Metadata of educational resource (virtual patient) lifted into RDF
Making sense of (medical) educational resources

HPV (disambiguation)
From Wikipedia, the free encyclopedia

HPV may refer to:
- Human papillomavirus, a virus which causes warts
- Human Powered Vehicle, a land, water, or air vehicle
- High Production Volume Chemicals
- Hypoxic pulmonary vasoconstriction, also known as shunt hypothesis
- Health Purchasing Victoria
- Hepatic portal vein

Human-powered transport
From Wikipedia, the free encyclopedia

Human-powered transport is the transport of person(s) and/or goods using human muscle power. An example is human-powered transport. Some forms of human-powered transport have existed since time immemorial in the form of walking, running, and swimming. Modern technology has enabled machines to enhance human-powered transport. Although motorization has increased speed and load capacity, many forms of human-powered transport remain popular for reasons of lesser cost, leisure, physical exercise and environmentalism. Human-powered transport is sometimes the only type available, especially in underdeveloped or inaccessible regions.
Adding semantics to OER descriptions

Entity recognition (NER) & disambiguation, via

Stefan Dietze

03/04/14

HPV (disambiguation)

From Wikipedia, the free encyclopedia

HPV may refer to:
- Human papillomavirus, a virus which causes warts
- Human Powered Vehicle, a land, water or air vehicle powered by human effort
- High Production Volume Chemicals
- Hypoxic pulmonary vasoconstriction, a physiological phenomenon
- Health Purchasing Victoria
- Hepatic portal vein

DBpedia

About: Human papillomavirus

An Entity of Type: disease, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org

Human papillomavirus (HPV) is a virus from the papillomavirus family that is capable of infecting humans. Like all papillomaviruses, HPVs establish productive infections only in keratinocytes of the skin or mucous membranes. While the majority of the known types of HPV cause no symptoms in most people, some types can cause warts (verrucae), while others can – in a minority of cases – lead to cancers of the cervix, vulva, vagina, penis, oropharynx and anus.

Property | Value
--- | ---
Virus del papil loma humana (VPH) es una expresión genérica que engloba a los virus que infectan la piel y mucosas en humanos y otros animales. Pueden originar diversas formas de cáncer, a menudo por vía sexual (en el caso de los tipos de virus que se toman en cuenta). La Generalitat de Catalunya estima un 80% de los casos de cáncer de cervizón y que probablemente es una de las infecciones más frecuentes. Más del 90% de infecciones de este tipo se curan de manera natural.
- Lípidos papa llomavirus (ve zlatota HPV a anglaiské Human papillomavirus)
One simple link – lots of knowledge/semantics

About: Human papillomavirus
An Entity of Type: disease, from Named Graph: http://dbpedia.org,
within Data Space: dbpedia.org

Human papillomavirus (HPV) is a virus from the papillomavirus family that is capable of infecting humans. Like all papillomaviruses, HPVs establish productive infections only in keratinocytes of the skin or mucous membranes. While the majority of the known types of HPV cause no symptoms in most people, some types can cause warts (verrucae), while others can – in a minority of cases – lead to cancers of the cervix, vulva, vagina, penis, oropharynx and anus.

Disease

ofType

Medicine

Life Sciences
Embedding educational repositories into a joint graph

- "HTTP-accessibility" (SPARQL, URI-dereferencing)
- "Structure" & "Semantics" (=> shared/linked vocabularies)
  - "Interlinked"
  - "Persistent"
Linked Data for Learning – How is it useful?

1. Linked Data for data sharing, enrichment and interpretation
   - RDF, SPARQL & shared vocabularies to improve interoperability of learning/user data
   - facilitate OER reuse vision across isolated platforms

2. Linked Data as body of knowledge for learning & education
   - Vast amount of publicly available resources and data
   - Dedicated OER and university data + „knowledge resources“ (from DBpedia to Slideshare)

   „HTTP-accessibility“ (SPARQL, URI dereferencing)
   „Structure“ & „Semantics“ (=> shared/linked vocabularies)
   „Interlinked“
   „Persistent“


Linked Data for Open and Distance Learning, Mathieu d’Aquin, report for the Common Wealth of Learning,
“LinkedUp” – Linking Web Data (for Education)

European project aimed at advancing take-up of open data and related technologies

http://linkedup-project.eu

“Success models”: data & applications

- LinkedUp Challenge to identify innovative tools & applications
- Evaluation methods and approaches

Data curation & profiling

- Collecting & exposing open data of educational relevance => LinkedUp Data Catalog
- Profiling and linking of Web Data => profiles of educational datasets

Technology transfer & community-building

- Disseminating knowledge & building communities (educators, computer scientists, data engineers)
- Gathering stakeholder feedback: use cases, and requirements

http://www.linkedup-challenge.org/

http://linkedup-project.eu/events

http://linkedup-challenge.org/#usecases
Who we are

LinkedUp Advisory Board

Sören Auer - University of Leipzig, Germany
Balaji Venkataaraman - Commonwealth of Learning, Canada
Dan Brickley - Google, United Kingdom
Philippe Cudre-Mauroux - EPFL, Switzerland

LinkedUp Network

Athabasca University, Canada's Open University

LinkedUp Consortium

Open Knowledge Foundation

UNIVERSITAT LEIPZIG

Stefan Dietze
Too many/diverse datasets, too little information

- Which datasets are **useful & trustworthy** for case XY (eg „learning about the solar system“)?
- Which **topics** (eg „Astronomy“) are covered?
- Accessability/reliability?
- Which datasets **describe/offer videos** (slides, publications, statistics etc)?

![Diagram showing possible datasets and tools](image-url)
Curating a catalog of data for education

- Which datasets are **useful & trustworthy** for case XY (e.g., “learning about the solar system“) ?
- Which **topics** (e.g., “Astronomy“) are covered?
- Accessability/reliability?
- Which datasets **describe/offer videos** (slides, publications, statistics etc)?

**Catalog of data (LinkedUp Catalog):** classification of datasets according to resource types, disciplines/topics, data quality, accessibility, etc.

**Infrastructure for distributed/federated querying**

Stefan Dietze
Building a registry for educational datasets

Dataset Metadata

LinkedUp Dataset Catalog

BBC Programme

Wonders of the Solar System

Brian Cox

Yovisto Video

Pluto & the Dwarf Planets

Entity disambiguation & linking [ESWC13]

AAISO

BIBO

FOAF

Schema mappings [WebSci13]

db:Astronomy

db:Astro. Objects

Topic profile extraction [WWW13, ESCW14]

Schema mappings

BBC

Yovisto

contains

DBpedia
LinkedUp Data Catalog

- RDF (VoID) dataset catalog: browse & query distributed datasets & mappings
- Live information about endpoint accessibility
- Federated queries using type mappings

Linked Education Cloud

Learning Analytics and Knowledge (LAK) Dataset
http://data.linkededucation.org/resource/lak-dataset

License: other-open
Responding: No
More info: http://www.solarresearch.org/resources/lak-dataset

Topics

Number of graphs: 20
Number of classes: 79
Number of properties: 471

Classes
- dcterms:subject
- dbpedia:uri
- rdf:type
- owl:toponame

Properties
- dc:subject
- dbpedia:uri
- rdf:type
- owl:toponame

Stefan Dietze
Dataset profiling: exploring LOD datasets/topics

- Automatic extraction of dataset “topics” [ESWC2014]
- Visualisation & exploration of dataset-topic graph (datasets, topics, relationships)
- Includes all (responsive) datasets of LOD Cloud
Type-specific exploration of dataset categories

- Type specific views on datasets/categories
  - “Document” (foaf:document)
  - “Person” (foaf:person)
  - “Course” (aaiso:course)

- Currently applied to datasets in LinkedUp Catalog only (as schema mappings already available here)
LinkedUp Challenge: using open data for learning

Open Data Competition to promote tools and applications that analyse / integrate (Linked) Web data

Organised by LinkedUp project over 2 years ("Veni", "Vidi", "Vici") with 40.000 EUR awards

Veni Competition - 22 submissions, 8 shortlisted for presentation at Open Knowledge Conference (17 September, Geneva Switzerland)
<table>
<thead>
<tr>
<th>Period</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>May – September 2013</td>
<td>Final events at OKCon 2013 (September 2013, Geneva)</td>
</tr>
<tr>
<td>October 2013 – May 2014</td>
<td>Final events at ESWC2014 (May, Crete)</td>
</tr>
<tr>
<td>May 2014 – October 2014</td>
<td>Final events at ISWC2014 (October, Riva del Garda, Italy)</td>
</tr>
</tbody>
</table>

- **Open Track only**
- **Open & focused track(s)**
- **Open track & focused tracks**
  - Submission details and calls to be released soon
Veni shortlist & winners

DataConf.

KnowNodes

ReCredible

YourHistory

Mismuseos

http://www.polimedia.nl/

PoliMedia – 1st price

http://www.globe-town.org/

GlobeTown - 2nd price

http://seek.cloud.gsic.tel.uva.es/weshare/

WeShare - 3rd price / people’s choice
Exploration of political debates and news coverage

Automatically generated links between transcripts debates, newspaper articles,

Generated data available as Linked Data (http://data.polimedia.nl)

Data sources: 1) newspapers in their original layout of the historical newspaper archive, and 2) radio bulletins of the Dutch National Press Agency (ANP)

- 9000+ debates (1945 – 1995)
- Over 3000 media links

Martijn Kleppe, Max Kemman, Henri Beunders (Erasmus Universiteit Rotterdam), Laura Hollink Damir Juric (Vrije Universiteit Amsterdam), Johan Oomen Jaap Blom (Nederlands Instituut voor Beeld en Geluid)
2nd Place: GlobeTown
Open data for sustainable development

- Exploring interaction between data about environment, economy and society (cause and effect in complex systems)

- Diverse open government data:
  - SWERA renewable energy data: [http://swera.unep.net/](http://swera.unep.net/),
  - Country names and codes: [http://opengeocode.org/](http://opengeocode.org/),

- Top three at Apps4Climate competition, held by the World Bank

Jack Townsend, Andrea Prieto-Vega, Richard Gomer, Will Fyson, Dom Hobson, Huw Fryer (University of Southampton, UK)
3rd Place/Peoples Choice: WeShare
Exploring, annotating, rating educational ICT tools

- Social & semantic annotation application for educational ICT tools
- Aids educators to find tools to support teaching at all educational levels
- Gathers and enriches data from existing registries and datasets
- Currently: approx. 7000 tool descriptions
- Crowdsourcing: educators, tutors etc can modify and enrich data

http://seek.cloud.gsic.tel.uva.es/weshare/

Adolfo Ruiz-Calleja, Guillermo Vega-Gorgojo, Juan I. Asensio-Pérez, Eduardo Gómez-Sánchez, Miguel L. Bote-Lorenzo, Carlos Alario-Hoyos
Universidad de Valladolid, Universidad Carlos III de Madrid, Valladolid, Spain
data.l3s.de – a DataHub for the L3S
Learning Analytics & Knowledge Data & Challenge
Facilitating Research on Learning Analytics and EDM

LAK Dataset & Challenge
Linked Data for the Learning Analytics & Educational Data Mining Community

About
The LAK Dataset makes publicly available machine-readable versions of research papers from the Learning Analytics and Educational Data Mining communities, where the main goal is to facilitate new research applications. This Website provides a home for the LAK Dataset as well as the associated LAK Data Challenges. This Website contains the latest information about the data itself as well as latest calls and updates about the LAK Data Challenges.

LAK Dataset (450 publications in RDF/R)
- ACM International Conference on Learning Analytics and Knowledge (LAK) (2011-13)
- International Conference on Educational Data Mining (2008-13)
- Journal of Educational Data Mining (2008-12)

LAK Data Challenge
- Analyse, explore correlate the LAK Dataset
- At ACM LAK 2014 (April 2014, Indianapolis)
Conclusions – “Lessons Learned”

- **Linked Data as enabler for educational resources sharing**
  - Interpretation of user & resource (meta)data
  - Understanding of learning contexts
  - Data disambiguation and integration
  - Cross-platform recommendations and analytics

- **Linked Data as body of knowledge for learning**
  - Dedicated educational/learning content
  - Repurposing of knowledge resources (social media etc)

- **Challenges**
  - Noise & diversity
  - Gap between technical community and end users /educators/learners
  - LinkedUp: technology transfer events, exploitation workshops etc
  - EUCLID
Outlook: some events & resources

- **LinkedUp Vici Workshop @ ESWC**
- **26 May, Crete, Greece**

**Workshop format**
The LinkedUp Vici Competition is open to anyone around the world who enjoys mashing up.

- **PROFILES workshop @ ESWC2014**
  - 26 May, Crete, Greece

- **LinkedUp Vici Challenge**
  - [http://www.linkedup-challenge.org](http://www.linkedup-challenge.org)

- **Vidi events at ESWC2014, Vici calls coming soon**

- **Learning Analytics & Knowledge Dataset**
  - [http://lak.linkededucation.org](http://lak.linkededucation.org)
  - + associated annual challenge

- **PROFILES workshop @ ESWC2014**
  - 26 May, Crete, Greece

- **26 May, Crete, Greece**

- **LiUD 2014 – Linking User Data**
  - [http://liud.linkededucation.org](http://liud.linkededucation.org)
  - **LiUD 2014**
  - Linking User Data 2014 at UMAP2014
  - Deadline: 1 April

- **Deadline: 1 April**

- **LinkedUp Challenge**
  - [http://www.linkedup-challenge.org](http://www.linkedup-challenge.org)

- **Vidi events at ESWC2014, Vici calls coming soon**

- **26 May, Crete, Greece**
Thank you!

WWW

See also (general)
- http://linkedup-project.eu
- http://linkedup-challenge.org
- http://linkededucation.org
- http://data.l3s.de
- http://linkeduniversities.org
- http://purl.org/dietze

See also (data)
- http://datahub.io/group/linked-education
- http://data.linkededucation.org
- http://data.linkededucation.org/linkedup/catalog/
- http://lak.linkededucation.org