Report on SCA Course

Title of the course
Linked data for science and innovation studies

Venue
VU University Amsterdam

Date
23-24 March 2017

Organizers
Peter van den Besselaar, Ali Khalili, Rinke Hoekstra, Klaas Andries de Graaf, Al Idrissou

Objectives
The aim of the course was that researchers firstly become aware of the potential benefits of Linked Open data – using (science and innovation studies) field specific examples, secondly to provide knowledge and skills to formulate the relevant questions related to data integration, and thirdly to introduce methods and tools for data linking, and for exploring, analyzing and visualizing linked data, as embedded in the Semantically Mapping Science (SMS) platform.

Main results
This course, organized as a mixture of lectures, hands-on and discussion sessions, helped participants in understanding Linked (Open) Data principles and how it can be used in science and innovation studies.

List of participants: **see ANNEX 1**

Programme: **see ANNEX 2**

Assessment: **see ANNEX 3**

Materials:

Linked Data Principles http://risis.eu/wp-content/uploads/2017/03/linked-data-principles.pdf

The Linked Data Lifecycle http://risis.eu/wp-content/uploads/2017/03/linked-data-lifecycle.pdf

SPARQL Query Language http://risis.eu/wp-content/uploads/2017/03/linked-data-sparql.pdf

Linked Data Publishing http://risis.eu/wp-content/uploads/2017/03/linked-data-publishing.pdf

ANNEX 1
LIST OF PARTICIPANTS

Name	Affiliation	Country	Email
Antonio García Romero	IE UNIVERSITY	Spain	agr22@faculty.ie.edu
Cristina Peñasco	CSIC	Spain	cristina.penasco@csic.es
Stephan Stahlschmidt	DZHW	Germany	stahlschmidt@dzhw.eu
Kiho Muroga	OECD	Japan	Kiho.MUROGA@oecd.org
Elsa Valente	Erasmus Uni	Netherlands	dealmeidavalente@ubib.eur.nl
Abdullah Gök	Manchester Uni	UK	abdullah.gok@manchester.ac.uk
Nuria BAUTISTA PUIG	Universidad Carlos III de Madrid	Spain	nbautist@bib.uc3m.es
Ulf Sandstrom	KTH	Sweden	forskningspolitik@gmail.com
Zohreh Zahedi	Leiden	Netherlands	z.zahedi.2@cwts.leidenuniv.nl
Wout	Leiden	Netherlands	w.s.lamers@cwts.leidenuniv.nl
Jakob Tesch	DZHW	Germany	tesch@dzhw.eu
Kostas Stathoulopoulos	Nesta	UK	konstantinos.stathoulopoulos@nesta.org.uk
Sisi. Li	Vrije Universiteit Brussel	Belgium	lisisi0212@gmail.com
Catalina Martinez	CSIC	Spain	Catalina.Martinez@csic.es
Isidro Aguillo	CSIC	Spain	Isidro.Aguillo@csic.es
Kathleen Gregory	DANS	Netherlands	kathleen.gregory@dans.knaw.nl
Wido van Peursen	Faculty of Theology, VU	Netherlands	w.t.van.peursen@vu.nl
Lucy Amez	Vrije Universiteit Brussel	Belgium	Lucy.Amez@vub.be
Xheneta Bilalli Shkodra	AIT	Austria	Xheneta.Bilalli-Shkodra@ait.ac.at
Ohid Yakub	University of Sussex	UK	o.yaqub@sussex.ac.uk

Maria Charitou	Expatriate Archive Center, The Hague.	Netherlands	charitou.ma@gmail.com
Patricia Laurens	UPEM	France	patricia.laurens@esiee.fr
Isabel Reis	FCT	Portugal	Isabel.Reis@fct.pt
Reinier de Valk	DANS	Netherlands	reinier.de.valk@dans.knaw.nl

ANNEX 2
PROGRAMME

Preparation

To make the course as effective as possible, we ask you to prepare by watching some material, by reflecting on it and by preparing for the first session.

Watch introductory videos on Linked Data

- Tim Berners-Lee on "The Next Web", http://www.ted.com/talks/tim_berners_lee_on_the_next_web
- Tim Berners-Lee at the GOV 2.0 expo ("bag of crisps"), https://youtu.be/ga1aSJXCFe0
- Google: Introducing the Knowledge Graph, https://www.youtube.com/watch?v=mmQl6VGvX-c
- Welcome to "MetaWeb", https://www.youtube.com/watch?v=TJfrNo3Z-DU

Reflection

- What, in your opinion, was the most interesting thing that came up? Write it down.
- Based on these videos, come up with two questions that you would like us to answer for you.

Preparation

- Write down your expectations: what research question(s) do you expect to be able to answer with Linked Data that you cannot answer, or is hard to answer, without it. Try to come up with two research questions.
- Choose a dataset, relevant for your questions, that you would like to see as Linked Data. Notes:
    - It is most convenient if the data is in tabular form, and is (somehow) related to the field of science, technology and innovation studies, but it doesn't have to be.
    - Make sure that the data is not sensitive (privacy, ethics), and that it is "open".
- Upload the dataset, along with a short description (one paragraph), to the DropBox folder that we have shared with you. This folder will be readable by all course participants.

Thursday 23 March
Location: main building of *Vrije Universiteit Amsterdam*, Room **HG 9A.16**

9:30   - 10:00   *Coffee*
10:00   - 10:30   *Introduction*
10:30   - 11:45   *Linked Data Principles*
      What is Linked Data, why do we need it, what does it look like and how does it work?
      Goal: by the end of this session you will understand the difference between URIs and URLs and can read Turtle files.
11:45   - 12:30   *Linked Data Hands-on*
    Build your own Linked Data, explore it, and link it to the Web of Data.
      Goal: By the end of this session you can write your own Turtle file, and use someone else's URIs to build Linked Data.
12:30   - 13:30   *Lunch*
13:30   - 14:00   *Linked Data & Research within RISIS* (Peter van den Besselaar)
14:00   - 14:30   *The Linked Data Lifecycle*
      What are the use-cases for linking data, and what research questions can you answer? What are the different phases in the lifecycle of Linked Data, what are the important considerations that need to be taken into account?
      Goal: by the end of this session you will understand the Linked Data lifecycle, and can apply it to your own research project.
14:30   - 14:45   *Coffee*
14:45 - 16:00   *Research Data Conversion Hands-on*
      Together with a fellow participant, select and (re)formulate a research question, decide on a plan for converting a relevant dataset to Linked Data, and perform the conversion.
      Goal: by the end of this session you will know what to take into account when converting tabular data to RDF, and are able to use OpenRefine to convert a dataset to RDF.
16:00   - 17:00   *Marketplace*
      This is where you showcase the end result of the conversion, and reflect on how you converted Linked Data. What are common pitfalls, what are best practices?
      Reflection and preparation for day 2
17:00   -   *Drinks*

Preparation for Day 2

To keep you busy over dinner, we ask you to prepare for the next day by watching some material, by reflecting on it and by doing some preparation.

Watch introductory video on Querying Linked Data

- SPARQL in 11 minutes
  https://www.youtube.com/watch?v=FvGndkpa4K0

Reflection

- Do you think the duration of the SPARQL movie is accidental?
- What, in your opinion was the most important take home message for the first day?
- Based on the first day, come up with two questions that you would like us to answer for you.

Preparation

- Revise the expectations and research questions you wrote down before the commencement of the course. What is different from before and why?
- What datasets brought by other participants (or discussed during the course) could be relevant for you, and what key entities should be linked?

Friday 24 March

Location: main building of *Vrije Universiteit Amsterdam*, Room **HG 11A.22**

9:00   - 9:15   *Coffee*

9:15   - 9:45   *Introduction & Reflection*

10:30   - 11:30   *SPARQL Query Language*

                 How can we query Linked Data, what is the syntax of SPARQL and how can I use it.

                 Goal: by the end of this session you will know the syntax of SPARQL and understand how you can query Linked Data on the Web.

11:30   - 11:45   *Coffee*

11:45   - 12:30   *SPARQL Hands-on*

      Query external datasets, query your own Linked Data, use queries to link it to the Web of Data.

                 Goal: By the end of this session you can write your own SPARQL queries, you can explain how a query produces results, and use someone else's SPARQL endpoint to answer questions.

12:30   - 13:30   *Lunch*

13:30   - 14:00   *Linksets and Lenses for Aligning Datasets* (Al Idrissou)

14:00   - 14:30   *Linked Data Publishing*

                 What are triple stores, and how do they work. How can we use a triple store to publish Linked Data on the web. How can we use mapping files to expose relational data on the web. How can we use (faceted) linked data browsers to explore our data.

                 Goal: by the end of this session you will understand how triple stores work, and can explain how Linked Data is published on the web.

14:30   - 14:45   *Coffee*

14:45 - 16:00   *Data Publication Hands-on*

                 Together with your fellow participant, revise your research question, make final adjustments to your Linked Data, and publish the Linked Data on the web..
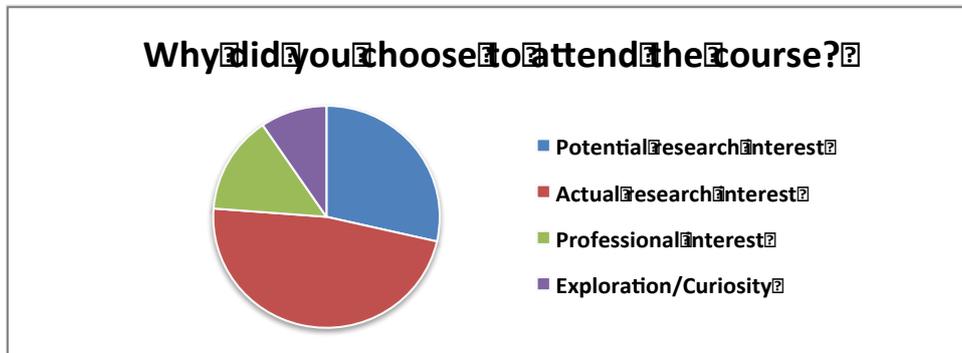
                 Goal: by the end of this session you can install a triple store, and publish your own Linked Data.

16:00   - 17:00   *Actual Research*

                 Now that we have datasets as Linked Data, can we formulate queries across our data that answer our research questions?

                 Goal: at the end of this session, you will give a short presentation of the design decisions of your data publication effort, and present "the" answer to your research question.

ANNEX 3
ASSESSMENT

Why did you choose to attend the course?

- ■ Potential research interest
- ■ Actual research interest
- ■ Professional interest
- ■ Exploration/Curiosity

RATINGS

Were the course objectives clearly defined?

no rather no rather yes yes

Were the contents of the course consistent with the course description?

no rather no rather yes yes

Has this course stimulated your interest?

no rather no rather yes yes

Was the course well organized?

no rather no rather yes yes

Was the course well structured?

no rather no rather yes yes

Have teaching materials facilitated learning?

no rather no rather yes yes

OVERALL SATISFACTION (scale 1-10)   8.5

BEST FEATURES

great overview on LD
nice set of tools for working with LD
step by step introduction into LD
knowledge about tools, methods & examples
Accessibility of teachers and language
The SPARQL lecture/exercise
the fact we used our own datasets
building our triples
very interesting subject
very nice lectures, very clear & to the point
the practical sessions
mix of lectures/presentations & assignments

ASPECTS TO IMPROVE

have more practice time
be a 3-day course with more exercises & hands-ons more days
a series of courses: intro, distance homeworks & applications
more time to learn this complex topic
more practical exercises needed as a follow up course
extend the duration, more specific explanation
give hands-out before the day for preview
would need more time to learn some concepts

nice presentations nice balance between theory and practice well organized very good explanation & material combination of theory/practice and hands-ons the hands-on parts	maybe one more day too much information for 2 days the duration of course should be longer