Online Media Provenance

State of the art

What did the others, techniques and methodologies used, evaluation methods, results

OpenSpending - aims to track every (public) government and corporate financial transaction across the world.

ProvVis - Experimental visualizations based on D3.js

ProvStore is a free service for storing, viewing and collaborating on provenance documents.

Important names in the field and research teams

An Overview of the PROV Family of Documents - The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web.

Paul Growth - research intelligent systems for the integration and use of diverse information with a particular emphasis on data provenance.

Articles, books and relevant links

Articles

- Provenance is becoming more important because it enables the link to see the source. The importance of provenance
- Data wrangling can introduce errors, data journalists should care about the validity of data. Provenance of data should include its primary source, but also all the

transformational steps performed by anyone.

How Spending Stories spots errors in public spending

- The provenance of a company's products have changed their practice to make their supply chain transparent because they worry about quality, safety, ethics, and environmental impact.
  The Transparent Supply Chain

Books

- Provenance: An Introduction to PROV - Synthesis Lectures on the Semantic Web: Theory and Technology, Luc Moreau and Paul Groth
- The Foundations for Provenance on the Web, Luc Moreau

Resources and tools available

Back-End

- NodeJS API using Express - one of the most powerful and versatile framework for building API in JS
- Sparql-client - a package for NodeJS which will make possible to run SPARQL queries on an ontology
- Lunr - another package for NodeJS which will help with articles comparison and see the similarities on different subjects
- For authentication we will provide an OAuth2 server which will provide the tokens used during the comunication

Front-End

- For the UI of our application we chose to work with the React framework for building a modern and performant web app

Data Model

- To be able to do SPARQL we will need to use an RDF database, having to choose between Apache Marmotta and GraphDB, both being NoSQL databases for storing triplets
- Also we will need another database for authentication, MongoDB being one of the best choices, being a NoSQL database too

Risk assessment

1.**Identify the hazards**

One of the hazards is that we can not process the date of the article, if the date of appearance if ambiguous.

2.**Decide who might be harmed and how**

One example is that advertisers because they will be confused about the segment of people who will have to launch certain campaigns.

3.**Evaluate the risks and decide on control measures**

Ex :risk of a Web Api is that we have no obvious security measures. And the solution is to introduce any further control measures in order to reduce the risk to the lowest possible level.

4.**Record your findings**

By recording the findings it shows that you have identified the hazards, decided who could be harmed and how, and also shows how you plan to eliminate the risks and hazards.

5.**Review your assessment and update as and when necessary**

You should never forget that few workplaces stay the same and as a result this risk assessment should be reviewed and updated when required.