Testing interfaces for navigating databases using metadata

# Geoff Ebbs s2980212 Masters Exegesis

Contents

[Geoff Ebbs s2980212 Masters Exegesis 1](#_Toc515373564)

[Abstract 3](#_Toc515373565)

[Prefacing material 3](#_Toc515373566)

[Author is the bomb 3](#_Toc515373567)

[Overview 3](#_Toc515373568)

[Research Question 4](#_Toc515373569)

[Significance 4](#_Toc515373570)

[Methodology 5](#_Toc515373571)

[Scope 7](#_Toc515373572)

[The interfaces 8](#_Toc515373573)

[Dirty prototype 8](#_Toc515373574)

[Second level prototype 8](#_Toc515373575)

[Timeline approaches 8](#_Toc515373576)

[Methodology adjustment as a result of initial testing. 9](#_Toc515373577)

[Currency charts 9](#_Toc515373578)

[Timeline libraries 9](#_Toc515373579)

[Chronoline.js 10](#_Toc515373580)

[HTML 5 video players. 10](#_Toc515373581)

[Imagehunt 11](#_Toc515373582)

[Storylines 11](#_Toc515373583)

[Presentation 11](#_Toc515373584)

[The backend 11](#_Toc515373585)

[Wordpress – source content 11](#_Toc515373586)

[HTML5, Javascript and css – Behind the interface 11](#_Toc515373587)

[Generation of metadata 12](#_Toc515373588)

[Metadex 12](#_Toc515373589)

[Presentation 12](#_Toc515373590)

[The Theory 12](#_Toc515373591)

[Findings 12](#_Toc515373592)

[Outcome 12](#_Toc515373593)

[Conclusion 12](#_Toc515373594)

[List of images 13](#_Toc515373595)

[List of figures 13](#_Toc515373596)

[List of abbreviations & terms 13](#_Toc515373597)

[Bibliography 13](#_Toc515373598)

# Abstract

In an increasingly complex information environment both consumers and researchers struggle to identify what is significant, meaningful and accurate. A number of projects exist to address this. Possibly the most prominent is the Semantic Web, established by web-founder Tim Berners Lee in 2000. The Semantic web sets out to provide a framework to allow comprehensive searching of the web using metadata.

Because metadata describes and defines content, it is more compact and better structured and, as a result, is much faster and more productive to search. Unfortunately, metadata is inconsistently generated, collected and stored which means that it is not readily usable on the wild web. The project to create a Semantic Web therefore, currently focuses on prepared databases such as dbpedia, a marked up version of Wikipedia, especially massaged for Semantic Web queries.

Partly because of its nascent status the challenge of how we might apprehend that metadata is yet to be seriously explored. We are used to the tabular layout of traditional computer systems and have rapidly become used to the stream of information we associate with our smartphone, but neither of those interfaces lends itself readily to the task of navigating complex information.

This project sets out to explore existing interfaces that can be used to navigate metadata with a view to determining how useful they might be in creating an interface for the Semantic web. It reduces the scope of the project by harvesting the native metadata available in current websites, specifically those published in WordPress.

The research on which this paper is based, explores the response of users to these interfaces and how they supported those users to make new inferences or draw new conclusions

# Prefacing material

## Author is the bomb

I am the bomb and I wrote everything in this paper.

## Overview

This project consists of

* a website that explores the narrative of a kayaker who paddled across three continents to get to COP21 in Paris using the navigation tools that are the subject of this paper
* a website that presents 15 years of environmental news using the same navigation tools
* a third website showing the tools in another environment.
* The tools that were used to build the site(s) (Github?)
* A presentation of the website(s) as a narrative at a gallery.
* Peripheral presentation materials to engage the viewer in using the interfaces (OUT OF SCOPE)
* The data collected on the responses to each interface
* This paper examining the effectiveness of these interfaces and the influence of modes of presentation on their perceived efficacy.

## Research Question

Identify those components of computer interfaces best suited to navigating databases using metadata to help build narrative and identify information in context.

## Significance

The Semantic Web project was initiated by Tim Berners Lee in 2000 (Berners-Lee and Fischetti 2001) to develop protocols and a framework to index the web built with metadata. This was in response to the explosion in the volume of content as a result of technology. Miniaturisation, connectivity, automation and artificial intelligence combine to generate visual, auditory and textual content in volumes completely beyond the capacity of humanity to absorb.

The challenge of finding verifiable, meaningful and relevant content in relationship to any particular topic increases with the variety and volume of the content itself. “As scientific instruments and computer simulations produce more and more data, the task of locating the essential information to gain insight becomes increasingly difficult.“ (Kesheng, Rishi et al. 2011) Metadata describes and defines content. Searching the metadata, then, is less onerous and more productive. Because we can sort, filter and organise content based on its metadata we can develop protocols for automating the sharing and delivery of content so that it is immediately usable and useful at the client. “Despite the unfamiliarity and power of the interface … 90% of the participants preferred the metadata approach overall, 97% said that it helped them learn more about the collection” (Mateas and Stern 2005)

An army of academics, developers and entrepreneurs are building the protocols, tools and content repositories to create the Semantic Web. “The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share and combine in formation more easily.” (Choudhury 2014)

This project sets out to contribute to that work in a very specific way. It seeks to test potential components of an interface designed specifically to navigate metadata.

Databases built before 1995 were designed to tabulate and report on information filtered by specific values of the content. The tabular nature of computer screens and applications such as spreadsheets complemented and supported the tabular view of business information systems commonly in use in most mainstream computer systems. While the visual interfaces of modern computing devices remain generally tabular in nature, we live in an increasingly visual world that requires more flexible responses. (Taylor and Benford 1997)

The development of text searching tools from the late eighties onwards shifted the form in which content is delivered from tables of content to a simple list of results. The notion of the stream is now firmly embedded in the public imagination as a sensible form of information delivery. “Compared with traditional databases, streaming databases pose new challenges for query processing due to the streaming nature of data which constantly changes over time.” (Liu and Ferhatosmanoğlu 2003) On the web, those results are generally web pages and documents posted to web servers granting them a URL. Early WAN-based text-searching returned simple lists of documents “text mining differs from search systems in that it can be incorporated into the organization’s business processes.” (Upshall 2014)

To interact meaningfully with the volume of information available to us, we must have a means of navigating it. That navigation system needs to respond intuitively to the character of the metadata which it reveals. “The use of automatic indexing tools, known also as text mining, or semantic content enrichment, has been for thirty years or more the subject of intense research and development,” (Upshall 2014)

This project provides examples of interfaces that consist of automatically generated navigation tools based on raw metadata that is native to the applications being examined.

Such an interface can potentially be used to navigate any metadata collection including the enormous set of data that is already converted to Semantic Web standards.

The prototype presented here can be attached to any WordPress site. This prototype is NOT YET installable as a WordPress plug in to simplify and this maximise control over the display device.

The project to implement it using WordPress methods is underway in GitHub and is expected to be implemented within months rather than years. That will allow for the expansion to networks of Wordpress sites and to the implementation of commercial features mining WooCommerce data. A separate project will port the interface to the dbpedia version of Wikipedia using SPARQL.

## Methodology

This project largely relies on a practice based methodology, as described in prototyping different interfaces in a consistent environment so they can be examined by usability testing and data analysis of log files. The approach conforms to a pragmatic research paradigm, setting out to explore solutions that work in the eyes of the subjects, rather than satisfying the requirements of a theoretical framework about how knowledge is formed. (Connell, d et al.)

The development process follows the double diamond of human centred design the process has been to initially use a literature survey to research the landscape of shared metadata analysis and visualisation and so refine the hypothesis and determine a specific focus for the practical WORK to test that hypothesis. It is thus iterative, using feedback to inform subsequent iterations. (Derrick 2018)

At a point half-way through the project, the literature survey has identified that navigating the semantic web itself is beyond the scope of this project. (Nowack 2009) Instead, the research will focus on navigating WordPress sites using the metadata natively available to them, or that metadata which can be automatically generated from them.

The methods used to implement this approach include:

A **literature review** to examine the various components of this overall project. The first of those components is the conceptual framework of the Semantic Web itself, largely a study of Berners Lee’s original document and some key works since then.

The second component is primarily made up the current tools being used to explore it. In addition to the **literature** identifying those tools, a **context review** of how those tools operate is a useful examination to inform the analysis of the prototypes tested in this research. The list of tools to be examined includes the query language SPARQL , commercial applications such as Pool Party (Koller 2009) and academic explorations such as YACY (Rudomilov and Jelínek 2011).

In addition to those well-established back-end tools it is important to examine what work has been done in providing front-end systems for navigating that data. The initial research indicates that these are primarily data visualisation approaches rather than navigation systems. (Rice 2018)

An overview of the nature of interfaces generally, the components of an interface that generally pertain to navigation rather than representation and any consideration or formal separation or metadata from the content that it describes are all useful areas of research. As these inform the theoretical underpinnings of the results, rather than the evaluation itself, that research will take place in parallel with the evaluation of user responses to the prototypes rather than informing the selection of the initial prototypes themselves.

There also needs to be a significant literature review and practical exploration of the wordpress tools available to develop the prototype interfaces to be used in the initial iterations of the evaluation.

The selection of the interfaces to form part of the initial prototype will emerge from a **context review** of existing applications that make use of metadata as a means of navigating narrative. Serious consideration has been given to **Whale Hunt** which is a project by New York based artist Harris who has developed an interface totally based on metadata to manipulate a library of images (Hennessy 2008). The web documentary **Prison Valley** has been selected as a classic example of web based interactive documentary (Nash 2012) that combines a traditional approach to online story telling with the use of metadata to access components of that story and to navigate through it on both initial and subsequent visits. The library of visual interfaces available at **D3 JS** is a powerful starting point for the rapid evaluation of a large number of existing interfaces to select a number of prototypes for consideration.

As well as these finished software projects it is proposed to **interview** three developers about projects which overlap in various ways to gain experience and understanding in the real world application of a theoretical framework such as described in real world software applications. The practitioners selected for that interview include Solid Software, Story Boxes and Ephox.

**Solid Software** is a Melbourne based developer (Goschnick 1998), well known to the author, who set out to build a multimedia story telling tool that allowed Australians to describe and share their multicultural heritage through an online application before the invention of the web, database publishing or the widespread adoption of the graphical user interface.

**Story Boxes** is a Brisbane film maker and corporate story teller, Griffith mentor to the author, that makes a range of visual content to promote corporate activity and generate narrative that is meaningful and exciting to the owners.

**Ephox** is a WordPress developer based in West End Brisbane that provides editorial interfaces used by WordPress, Blackboard and many other global content management systems responsible for a large portion of the content on the World Wide Web.

**Mindmapping** was used in the planning for this project to organise and map the components of the literature and context review and prioritise those areas most relevant to the final outcome.

The **prototyping** phase has commenced, taking the interfaces provided with Whale Hunt, the author’s Storylines and Swim projects and the Word Cloud from D3 JS to provide a consistent framework for sharing and demonstrating those interfaces such that the target audience can evaluate them naturally while their usage data is collected from those prototypes using log files and online response forms.

The **useability testing** and **data analysis** to determine what aspects of each interface are useful, study patterns emerging from the use of metadata as a primary navigation tool and explore the nature of the boundary between metadata and the content it describes will therefore be partly overt (log files) and partly overt (a mix of closed and open questions collecting user impressions)

Implicit in this iterative process is ongoing h**ypothesis testing** to examine whether the interfaces trialled during the project have provided the efficacy for which they were selected.

## Scope

This project provides examples of interfaces that consist of automatically generated navigation tools based on raw metadata that is native to the applications being examined.

Such an interface can potentially be used to navigate any metadata collection including the enormous set of data that is already converted to Semantic Web standards.

An increasing volume of data arrives on our communication devices in the form of a transient stream. The Semantic Web project is inherently limited to published data in the form of web pages. While this narrow scope needs to be addressed at some point it is well outside the scope of this project which will NOT examine the navigation or organisation of streamed content.

dbPedia represents the largest readily accessible content base of Semantic Web compliant material. It is accessible using the SPARQL query language and tools, such as YAYC, a distributed search engine that makes use of the SPARQL query language and Semantic Web prototocols. (Rudomilov and Jelínek 2011) There are also a range of corporate tools built on these protocols, such as Pool Party (Koller 2009). Initial research into using these tools and that framework as the back end for this project indicates that the development of the tools necessary to generate queries such that navigational interfaces of the sort implied by this research question cannot be easily generated. For that reason dbPedia will NOT be used as the knowledge base and WordPress has been selected as a simpler, more accessible platform for testing the concepts presented.

While the prototype presented here can be attached to any Wordpress site it is NOT YET installable as a WordPress plug in. Among other reasons, this allows the project to maximise control over the display device simply without navigating the WordPress development methods.

The project to implement it using WordPress methods is underway in GitHub and is expected to be implemented within months rather than years. That will allow for the expansion to networks of Wordpress sites and to the implementation of commercial features mining WooCommerce data. A separate project will port the interface to Wikipedia using SPARQL.

# The background - interfaces

## Timelines

### As a tool for visual story telling.

### Tips & tricks

1. Keep it short. We recommend not having more than 20 slides for a reader to click through.
2. Pick stories that have a strong chronological narrative. It does not work well for stories that need to jump around in the timeline.
3. Write each event as a part of a larger narrative.
4. Include events that build up to major occurrences — not just the major events.

# The practice - interfaces

These interfaces use a limited set of metadata that is readily available on most websites. The metadata natively stored on the site is complemented by additional data required by the some components of the interface and automatically generated when a session is initialised.

## Dirty prototype

The initial prototype simply inserted the Whale Hunt approach into an existing project prepared by the author for 7153QCA Web Technologies (Ebbs 2017 a) and then integrated the Storylines approach to representing links between WordPress posts based on their metadata developed and submitted for 7157 QCA Interactive Media Theories (Ebbs 2017 b). That approach appears to coincidentally be consistent with and complementary to an approach developed by Nguer and Spyratos in Portugal in 2008. (Nguer and Spyratos 2008) Initial feedback was informal and anecdotal and included feedback from the content producer and subject of the blog at lifejacket.ebono.com.au, Steve Posselt.

Even though I paddled every inch of the journey and took the vast majority of the photos, I was amazed how much the timeline view surprised me. There were about a third of the photos that I had forgotten taking and quite a few I used the back button to go back and have a second look.

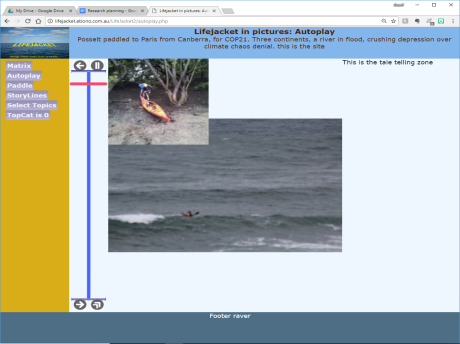
## Second level prototype

Having established that it is possible to present a range of components in one website, providing different views of the same data it became essential to invest some time in coordinating and streamlining those interfaces so that they operated in a consistent manner across interfaces. This section deals with the exploration of the approach to selecting images (or other content) based on the time at which it was published. In this project that component will be known as **the timeline as a selection tool**.

### https://lh6.googleusercontent.com/QVmsLOfMmv7IINMJw37yWILoXMPGITnLA7bnufFjruuU7qVwqtQtMhPpogQPz2oglZioa2htNDchF6Cw0ujzTMt9zD69nwl564WleROwFZJTbvJpj6eJAxXjS3xJmyi2Oboioh5uTimeline approaches

The second version of the dirty prototype provided a basic timeline that indicated what image or item was under the mouse. That item would then become active when clicked. This behaviour was adopted from The Whale Hunt.

Because it is the requirement of an information navigation system that we can filter by selecting ranges of data, the selection of a single item is only a small part of the puzzle.

The requirement to adjust the range of the timeline as well as the image, led to the addition of selectors at the top and bottom of the timeline which required much more complex analysis of the mouse position, the storage of states and so on.

To avoid the timeline becoming a major development project distracting from the overall project and it was decided to carry out a literature and context review into timeline libraries using Javascript.

This relatively simple component, thus became the test bed to define and refine the research approach.

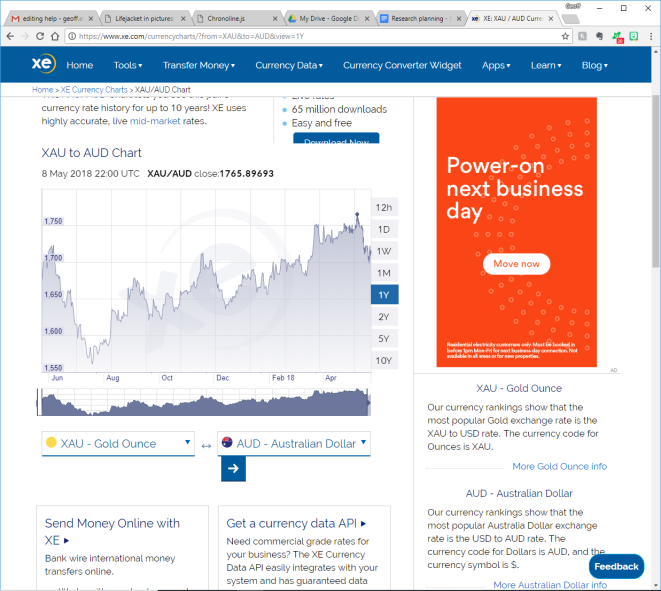
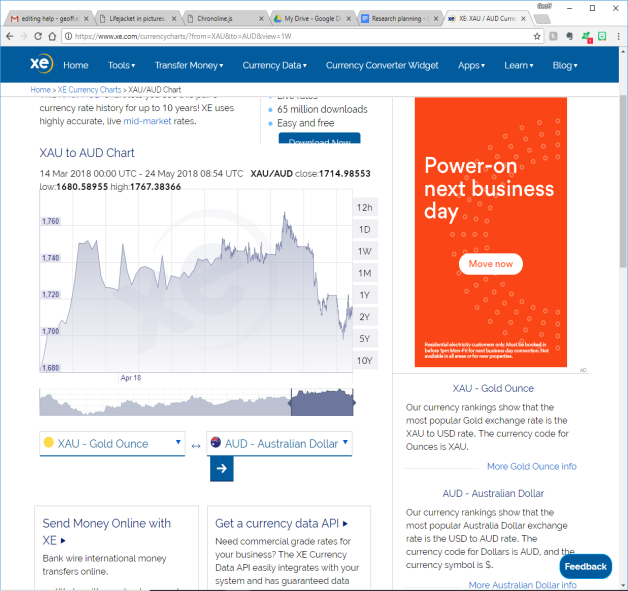
1. Prototyping defines a requirement.
2. A context review refines the requirement and suggests solutions
3. A literature review explores the available solutions and means of achieving them
4. Usability studies determines the efficacy of available solutions
5. The data analysis suggests an approach
6. Prototyping implements the results and the cycle starts again … or moves onto the next component.

### Methodology adjustment as a result of initial testing.

For practical reasons, this plan was altered so that usability studies can take place after implementation, allowing for testing of a number of aspects at once. The prototype as presented to the user therefore involves a number of different approaches and allows the user to compare thim. This has the benefit of ensuring a consistent test bed on which the testing of the different component takes place. Comparing the different methods of selecting timeframes in a currency charting application may produce completely different results than the same components being tested in a video control environment. Maintaining the prototype interface as the test bed also provides incidental feedback on the overall project that would otherwise require separate testing.

### Currency charts

The selection of a date within the available data is live (mouseover) and responds with useful information (date and closing price as subtitle - reflecting the dot at the peak in chart 1). The selection of the range is via external buttons. The range in chart 1 is one year. Note the capacity to arbitarily adjust the range using the slider at the bottom in chart 2.



### Timeline libraries

Rufus Pollock wrote a review of Javascript timeline libraries in 2012. His major finding was that they are too complex and not modular enough.

I want to start with a general comment. Timeline libraries consist of various components:

1. Data loading
2. Date parsing
3. Band (timeline) rendering
4. Showing render info on individual items

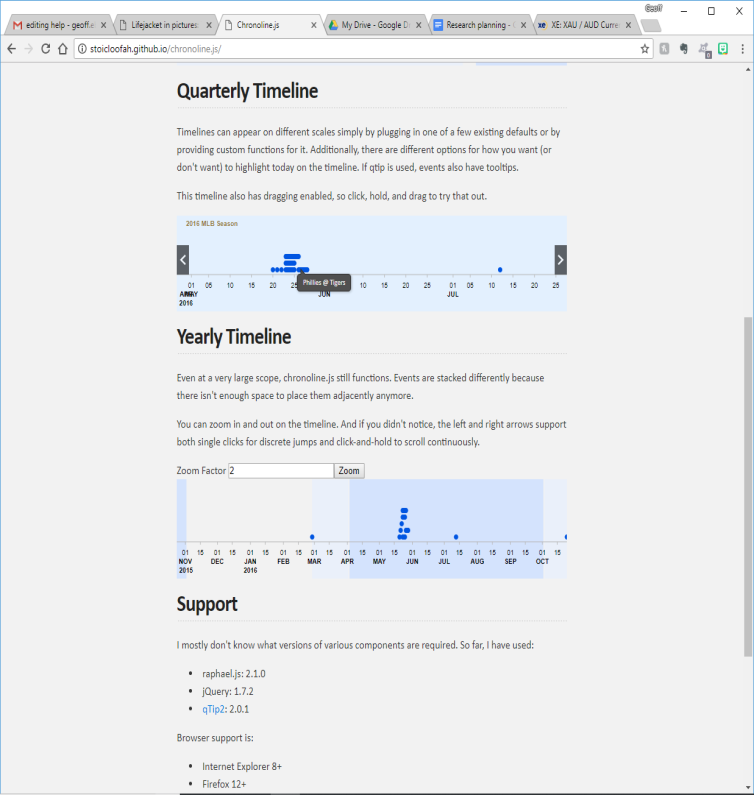
For me a timeline visualization library need only be the second of these but most that I’ve come across do more.

In fact a major issue in my opinion with most libraries is that they are under-componentized - they don’t separate cleanly into these different components and end up doing everything.

To take one example, the Verite timeline (in my view is one of the best libraries out there) has a whole bunch of its own custom date parsing built in inside an internal utility library which are hard to override or replace and also has a large chunk of code just for loading from google docs and other data sources. (You can of course somewhat solve this somewhat – as I do in Recline by parsing the dates directly and then submitting in a standardized form).

In my view, even if library authors do want to include these sorts of things, it would be good to do it in a way that allowed for a clean separation so that you could just use the parts you wanted (and/or over-ride parts more cleanly). (Pollock 2012)

### Chronoline.js

One of the lines reviewed by Pollock that appeared to have potential for this project is chronoline.js

Hovering over datapoints depicted on this timeline reveals useful data, indicating that making them into live links is trivial.

There appear to be no controls to adjust the scale (ie size of the range) though the line can be slid using the controls left and right that appear when you hover over the timeline.

Like most of the timelines reviewed by Pollock this is more concerned with data representation than navigation. The gap between the existing operation of the component and the requirement, combined with the observation that the components lack modularity determined that alternative approaches be sought.

### HTML 5 video players.

Given the relative simplicity of the basic requirement and the common appearance of HTML5 video it was decided to review the available code libraries.

Having used the Cloudfare library Clapprin the past and so having working examples to hand that was analysed initially. The code available from <https://cdnjs.com/libraries/clappr> includes a minimum library of 16,000 lines of code and a flexible library of 36,000 lines of code.

The bulk of the effort in building (and reviewing) the Javascript libraries that support HTML 5 video and audio goes into dealing with the actual media files. As the introduction to the review by OurCodeWorld.com says, “This [list] provides plugins that will deal with the harder tasks as compatibility, fallbacks for older explorers etc” <https://ourcodeworld.com/articles/read/148/top-7-best-html5-media-player-javascript-plugins>. All that is required for this project is a way to nominate the position in a list of files.

A second CloudFlare library, PLYR, was investigated. At only 8,000 lines of code it was comparitavely approachable but still quite dense. 1400 of those lines are the main function Plyr which is primarly concerned with configuration and loading elements of the interface. The actual controls presented to the end user and the passing of parameters to the other routines are buried up to four layers of functions deep.

An investigation of over an hour failed to identify whether the library allowed for multiple positions on the slider and whether a subset of the entire timeline could be selected.

It was decided to do a quick review of the libraries reviewed at OurCodeWorld startings with jPlayer. jPlayer was selected because it is completely open and customisable and has the reputation of being “developer friendly”. Again the vast majority of the code is about manipulating media files, buffering, loading drivers and codecs and managing the display of the content.

The scope of these players does not match the requirements of this project.

A final attempt was made to identify a class of libraries using more advanced search phrases.

javascript libraries to access image library using timeline returned some useful results.

Knightlab’s timeline.js describes itself as a reasonable fit. https://timeline.knightlab.com/

TimelineJS is an open-source tool that enables anyone to build visually rich, interactive timelines. Beginners can create a timeline using nothing more than a Google spreadsheet, [like the one we used for the Timeline above](https://docs.google.com/spreadsheets/u/1/d/1xuY4upIooEeszZ_lCmeNx24eSFWe0rHe9ZdqH2xqVNk/pubhtml). Experts can use their JSON skills to create custom installations, while keeping TimelineJS's core functionality.

Most usefully in terms of the literature review it identified the role of a timeline tool in allowing users to create narrative. (See relevant section above)

What is clear from this examination is that most published interfaces are sophisticated content delivery mechanisms in their own right and are not easily broken into components that are applicable to a generic test. Each of these interfaces can be analysed in its own environment and compared with other interfaces, but that analysis does not contribute to the development of a navigation system that consists of a combination of those components.

This represents a fork in the road of the research. One path is to persist with the development of a simple interface that includes many components and allows us to test how they might be combined, the other path is to test each interface in situ and then speculate on how they might be applied in conjunction.

Given the objective to produce a product that can be applied to WordPress sites it was decided to persist with the approach of developing simplistic versions of each interface. The aim of this stage is to throw as many intervaces onto the same data set as fast as possible. The hypothecated role of the timeline is to select a subset of images (content) based on their timestamp for use in other filtering components. The operation of the timeline interface itself is trivial in comparison to that back end work. A basic timeline constructed to provide that minimum functionality is the fastest way to move that project forward.

When the different selection components are in place, the first evaluation of the selection methods can begin.

The second interation of the framework will involve the ability to book mark content sets, for use as starting points to further exploration or as interactive but preselected, content. .Ideally, the selected content would be saved in a form usable by wordpress: a page, a content category, ….

The tool, thus forms the foundation for construction of narrative.

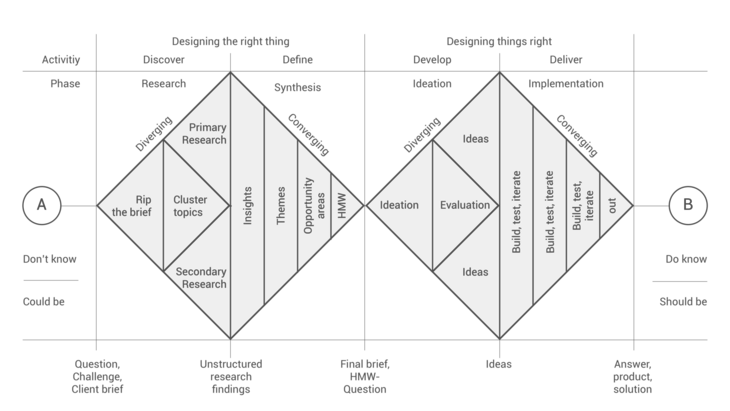
This stage of the project is throwing approaches at the ideation stage of Dan Nessler’s version of double diamond design approach, It is diverging as different approaches appear as alternatives. (Nessler, 2014)

At the end of July the divergence ceases and the total focus shifts to the evaluation of the interface at whatever stage it is. A range of implementations of many components may be visible as part of the testing.

The final prototyping stage is removing that duplication and dropping approaches that do not add sufficient value to the project.

Some effort will then be applied to using the tool to create a presentation of the tool’s capacity. It is likely that this paper will form the basis for the content in that presentation. Of course, it will be a WordPress site.

(Nessler 2014?) <https://www.hyperisland.com/community/news/how-to-apply-a-design-thinking-hcd-ux-or-any-creative-process-from-scratch>



This decision was reviewed in class on the last day of trimester one.

## Imagehunt

Whale Hunt (Hennessy 2008)

## Storylines

# Presentation

Death Row

Book

Chime

# The backend

### Wordpress – source content

It uses the content management system Wordpress as the example content repository and management system, because it is well-established, has a very large population of users and developers and, significantly, rigorous and open protocols for development. One of those rigorous and well-stablished protocols pertains to the storage of metadata.

Thus, Wordpress provides endless examples of typical websites with relatively consistent approaches to the organisation of metadata that provides a rich source of metadata on which to experiment.

Wordpress claims to host over one quarter of the world’s websites. Of course, a lot of information is stored in applications that are not included in that estimation and so its percentage of the total useful content is much smaller. For a prototyping project such as this, though, it is sufficiently complex and interesting.

### HTML5, Javascript and css – Behind the interface

The prototypes have been built as standalone web pages with HTML Canvas as the visual interface used to navigate the WordPress site. Again, these are the current, open standards ensuring the maximum capacity for integrating this approach in the broader development community. (Gailitx 2007) Ideally this would be tightly integrated with WordPress but the learning curve involved in achieving that may prove to be beyond the scope of this project.

Panels

## Generation of metadata

## Metadex

# Presentation

# The Theory

The displayed image is from the almost anonymous Bob “zazio815” at <https://smiy.wordpress.com/2011/01/10/the-common-layered-semantic-web-technology-stack/>. He credits the image as being derived from Benjamin Nowack’s interpretation of the Semantic stack. (Nowack 2009). Due to Bob’s preference for anonymity “It’s not about the who it’s about the what!” I have cited Nowack even though I prefer Bob’s image.

The Whale Hunt - (Hennessy 2008)

Prison Valley: arte.tv - (Nash 2012)

Corals of the World (Veron 2000)

# Findings

# Outcome

TaleSpin (a program that tells stories) (Meehan 1981)

Interactive Fantasy – populating virtual worlds (Laurel 1997)

Façade: A framework for interactive drama (Mateas and Stern 2005)

Skyrim: State of the art interactive game generation (Kuo, Hiler et al. 2017)

# Conclusion

Appendices, ethical approval, consent packages, interview survey questions, gantt chart/ timeline to completion

Bibliography / References

Design Research Question; What is your project about? What question does it seek to address?

Methods; How has (will) your project proceed? What methods will you use? Why have these methods been chosen? Describe how these methods will be (have been) done to a high standard.

Methodology; What is/are the overarching methodology(ies) that guides your research and organises your methods into a coherent process?

Exemplars and analysis; Provide an overview of some (hopefully the most important) related work done by others in your field, and provide a critical summary of it. How does your project fit into this landscape of existing/previous similar work? How is it informed by previous research and scholarship?

## List of images

## List of figures

## List of abbreviations & terms

Metadata

Metadex

µFormat

RDF

RFL

Semantic Web

SPARQL

URI

URL

# Bibliography

Berners-Lee, T. and M. Fischetti (2001). Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor, DIANE Publishing Company.

Choudhury, N. (2014). "World wide web and its journey from web 1.0 to web 4.0." International Journal of Computer Science and Information Technologies **5**(6): 8096-8100.

Connell, J. L., et al. Structured rapid prototyping :|ban evolutionary approach. Englewood Cliffs, N.J. |bYourdon Press,|cc1989.

Derrick, B. (2018). "Iterative Design Process." Retrieved 12/05/2018, 2018, from <http://www.ffkr.com/iterative-design-process-brian-derrick/>.

Goschnick, S. B. (1998). Design and development of Melbourne IT Creator/sup TM/-a system for authoring and management of online education. Technology of Object-Oriented Languages, 1998. TOOLS 28. Proceedings, IEEE.

Hennessy, K. (2008). "The Whale Hunt (Harris)." Museum Anthropology Review **2**(2): 134-135.

Kesheng, W., et al. (2011). "Finding regions of interest on toroidal meshes." Computational Science & Discovery **4**(1): 015003.

Fusion promises to provide clean and safe energy, and a considerable amount of research effort is under way to turn this aspiration into a reality. This work focuses on a building block for analyzing data produced from the simulation of microturbulence in magnetic confinement fusion devices: the task of efficiently extracting regions of interest. Like many other simulations where a large number of data are produced, the careful study of 'interesting' parts of the data is critical to gain understanding. In this paper, we present an efficient approach for finding these regions of interest. Our approach takes full advantage of the underlying mesh structure in magnetic coordinates to produce a compact representation of the mesh points inside the regions and an efficient connected component labeling algorithm for constructing regions from points. This approach scales linearly with the surface area of the regions of interest instead of the volume as shown with both computational complexity analysis and experimental measurements. Furthermore, this new approach is hundreds of times faster than a recently published method based on Cartesian coordinates.

Koller, A. (2009). SKOS Thesaurus management based on linked data with PoolParty. Third annual European semantic technology conference, Vienna, Citeseer.

Liu, X. and H. Ferhatosmanoğlu (2003). Efficient k-NN Search on Streaming Data Series, Berlin, Heidelberg, Springer Berlin Heidelberg.

Data streams are common in many recent applications, e.g. stock quotes, e-commerce data, system logs, network traffic management, etc. Compared with traditional databases, streaming databases pose new challenges for query processing due to the streaming nature of data which constantly changes over time. Index structures have been effectively employed in traditional databases to improve the query performance. Index building time is not of particular interest in static databases because it can easily be amortized with the performance gains in the query time. However, because of the dynamic nature, index building time in streaming databases should be negligibly small in order to be successfully used in continuous query processing. In this paper, we propose efficient index structures and algorithms for various models of k nearest neighbor (k-NN) queries on multiple data streams. We find scalar quantization as a natural choice for data streams and propose index structures, called VA-Stream and VA + -Stream, which are built by dynamically quantizing the incoming dimensions. VA + -Stream (and VA-Stream) can be used both as a dynamic summary of the database and as an index structure to facilitate efficient similarity query processing. The proposed techniques are update-efficient and dynamic adaptations of VA-file and VA + -file, and are shown to achieve the same structures as their static versions. They can be generalized to handle aged queries, which are often used in trend-related analysis. A performance evaluation on VA-Stream and VA + -Stream shows that the index building time is negligibly small while query time is significantly improved.

Mateas, M. and A. Stern (2005). Structuring Content in the Façade Interactive Drama Architecture. AIIDE.

Nash, K. (2012). "Modes of interactivity: analysing the webdoc." Media, Culture & Society **34**(2): 195-210.

Nguer, E. M. and N. Spyratos (2008). A user-friendly interface for evaluating preference queries over tabular data. Proceedings of the 26th annual ACM international conference on Design of communication. Lisbon, Portugal, ACM**:** 33-42.

Nowack, B. (2009). "The Semantic Web–Not a piece of cake."

Pollock, R. (2012, October 2013). "Javascript Timeline Libraries: A review." Retrieved 22 May 2018, 2018, from <http://okfnlabs.org/blog/2012/12/04/javascript-timeline-libaries-a-review.html>.

Rice, I. (2018). "Improved data visualisation through nonlinear dissimilarity modelling." Pattern Recognition **73**(Jan 2018): 12.

Rudomilov, I. and I. Jelínek (2011). Semantic P2P search engine. 2011 Federated Conference on Computer Science and Information Systems (FedCSIS).

Taylor, I. and S. Benford (1997). Exploration of the requirements in a visual interface to tabular data, Berlin, Heidelberg, Springer Berlin Heidelberg.

Upshall, M. (2014). "Text mining:Using search to provide solutions." Business Information Review **31**(2): 91-99.

Text mining is an umbrella term covering a wide range of software tools, including natural language analysis, use of statistical techniques, and machine learning, designed to extract entities (names of people and places), index terms, and relationships. This article outlines these tools, and describes how text mining is being applied in enterprises to solve user problems. A standard measure of accuracy enables text-mining tools to be compared. Text mining is often applied to unstructured content, but the most impressive results come from applying text-mining tools on structured text within a specific subject domain, using an iterative process to train the software.