

Abstract

Digital Narratives at the BBC is an experiment in short-form placements for academics within the BBC. Inspired by ideas of what a 'web adaptation' of a BBC property could be, it was determined that while the BBC had a world-leading implementation of a metadata backed publishing platform, it did not have the domain model to facilitate metadata based publishing of its programmes. The production information was there, but the content information wasn't.

Over the course of the placement this was addressed as a collaborative process between Paul Rissen, representing the BBC, Michael Jewell, a researcher with narrative modelling experience and myself, Toby Harris, a student on the Media and Arts Technology programme at Queen Mary, University of London.

An ontology was developed called 'BBC Stories' that aims to facilitate powerful journeys through BBC content. It presents a minimal, extensible core to meet the joint goals of simplistic application now while remaining compatible with more advanced use as adoption grows.

Acknowledgments



Rowena Goldman – Strategic Partnerships, BBC R&D

Paul Rissen – BBC Future Media and Technology

Michael Jewell – Goldsmiths College, University of London



Tassos Tombros – Queen Mary, University of London



Media and Arts Technology Doctoral Training Centre

Queen Mary, University of London

<http://www.mat.qmul.ac.uk>

Introduction	4
Paul Rissen’s Thesis	5
Applying the trend, in the here and now	7
Applying the trend, in a 3.0 Web	11
Transmedia and Crossmedia	16
At the BBC	17
SUDS	17
Archers Prototype Website	18
Mythology Engine	20
BBC Round-Up: the Pompeii Effect and a call to action	23
Research Problem	26
Literature Review: Towards modelling programmes	27
Semantic Web technologies and the BBC	27
Linked Data	27
Linked Data at the BBC	28
Linked Data conclusion	30
Modelling Content	31
Top-down, bottom-up, or both?	31
Media Information Spaces—A Semantic Challenge	32
Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology	32
Enter OntoMedia (stage-left)	33
Crowdsourcing the Narrative and Audience Interpretation	34

Reasoning with narrative	35
.....
Summary	35
.....
Ontology development	36
<hr/>	
OntoMedia 2	36
<hr/>	
OntoMedia2 – N3 annotation of Doctor Who episode	38
.....
BBC Stories	39
<hr/>	
BBC Stories – The ontology	43
.....
BBC Stories – N3 annotation of Doctor Who episode	45
.....
Applying ‘BBC Stories’	46
<hr/>	
Choice of media	46
<hr/>	
Application	46
<hr/>	
Creating transcriptions	46
.....
Using the data	46
.....
Visualising the data: Diagram	48
.....
Visualising the data: Interactive	49
.....
Conclusions	51
<hr/>	
References	52
<hr/>	

Introduction

Parts of BBC Future Media and Technology run a ‘10% time’ initiative. ‘The idea is to get lots of good ideas and quickly built prototypes rather than complete solutions. Then, if the idea is good, it will get picked up by our normal production process.’ (Ferne, 2008)

This project, Digital Narratives, is born out of Paul Rissen’s 10% time, and is an earnest contribution towards some of those ‘good ideas’ getting into the BBC’s production process. In the words of Jonathan Tweed, a BBC employee interested in taking on the ideas for his group’s web productions, ‘how can we get a powerful journey through our content now – and what investment do we need to inspire to get to the vision?’.

So what is this vision? Developed over postings on Paul’s ‘R4 is static’ blog, the ‘Fourth Medium’ post gets to the heart of it, talking about developing something that could be considered a true ‘web adaptation’ of a story – something that, just like a film or book adaptation, for example, can tell the same core story, but in a manner that best uses the opportunities and idiosyncrasies of that medium. (Rissen, 2009)

The potential impact of this direction is far greater than considering the nuances of one medium over another, however. It touches on just what the internet could be, and consequently just what an organisation like the BBC could become. If the the British Broadcasting Corporation was created not at the birth of broadcasting, but instead for a ‘Web 3.0’ era, what would it look like?

Paul Rissen's Thesis

What is the insight that can lead us towards a 'web adaptation'? Speaking at Future Everything 2010, Paul set out what I consider his thesis. He characterises the nature of effective drama as one of building up a network of information and the *jouissance* of making connections that suddenly give insight and draw disparate strands together. He then characterises the nature of the World Wide Web (hereon in, the Web) as intended to be one of a web of concepts. The argument is, then, that the World Wide Web could be an ideal medium for drama, its network nature being a native fit for drama's web of meaning.

The application of this thesis becomes particularly compelling when you then look at the development of the Web as a medium, and apply that to broadcast as a medium.

The Web as we have it today is largely a web of documents. A document is essentially a package of stuff fashioned into a single entity. Now if these bits of stuff were exposed as a whole, forming a linked world of data, then the role of the document would become that of a lens onto that data, allowing interpretation or journeys through it. This is the vision of the semantic web community, and is gaining acceptance as 'Web 3.0'. This view of the Web is also being taken up by the BBC itself in its role as a practitioner of web publishing, where it has found such an approach the only feasible way to ensure sufficient coverage of large, international sporting events where every team and individual's progress can be tracked.



What if you'd been involved in a horrific, decade-spanning war? What if you'd played a significant part in ending that war, but only in a way that meant both sides lost? What if you were the last of your kind? What if you'd been on the run?



What if you'd found a way to disguise yourself so completely, you forgot even yourself?



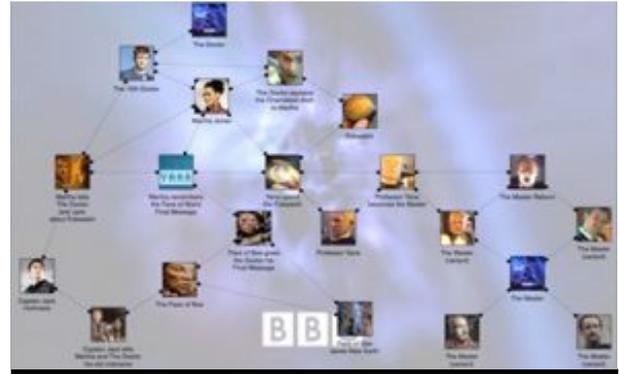
What if you'd made a friend, an equal? What if your new friend also had a fobwatch?



What if that new friend turned out to be something quite different? What if that friend, was, in fact, one of your deadliest enemies in disguise?

If, previously, we have said that the web as we have it largely consists of documents which package stuff together into a fashioned whole, well, you could say the same about a broadcast programme. In which case, can we apply this trend we observe in the Web to that of the programme? Based on his insight, Paul posits that we should. It will inform what we can do now, and where things can go.

The slides to the right of this section put the drama argument as presented by Paul at Future Everything. And this is where we leave Paul, and dive off ourselves.



When the threads of many narratives come together into a single dramatic moment such as this, as part of the audience, you put all the puzzle pieces together in your head, the hairs stand up on the back of your neck, and you punch the air in excitement. This is one of the secrets to any decent drama, whether it's science fiction, murder mystery, political thriller, or even soap. As you progress through the narrative, you learn key information about characters, about events, about places, and every new piece of information you learn, you attempt to fit into the network of knowledge about what you already know of this fictional world. You get involved. You treat it as if it is the real world, and you get drawn into the narrative. You talk to others about it, you share your experiences. This is true, irrespective of the media through which the narrative is delivered.

Applying the trend, in the here and now

If we are making an analogy between programmes and web pages, we can see that while they might both be ‘fashioned containers of stuff’, the web page already has network characteristics, unlike the programme. We might, however, be able to apply some of these network characteristics to programmes.

Within the page itself, the author may have created in-line links to other documents. The page as a whole may be tagged with a label such that it can be aggregated with other similarly labelled pages on the site. These links and the text of the page, accessible to computing services as well as end-users, can be statistically analysed in the context of the internet as a whole and ranked against search terms; search services such as those provided by Google seem almost god-like in their knowledge and reach, without even understanding what is represented by these documents – just processing patterns in the words and links.

So we might be interested in creating these in-line links. We might want to tag and so aggregate programmes together. We might want to make the contents of our programmes available to textual search.

What would this give us, and are there signs of this happening already?

In-line links give direct navigation. They are contextually specific to the content being consumed at that moment, and can link to into specific parts of documents as well as to the document as a whole. Enabling a focussed, direct journey through content, they are authored to be relevant to the matter at hand, sometimes even skipping the sections of the target document that aren’t relevant. This sounds very useful to our drama-as-network thesis. It would require the linking mechanism, and the ability to view the programme at a more granular level – to link to and from sections within that programme.

Tagging documents allows aggregation of related content. On absorbing one document, you can choose from a selection of others. This is evidently useful, but

does not sound so relevant to our drama-as-network thesis. It is, however, interesting in two particular ways.

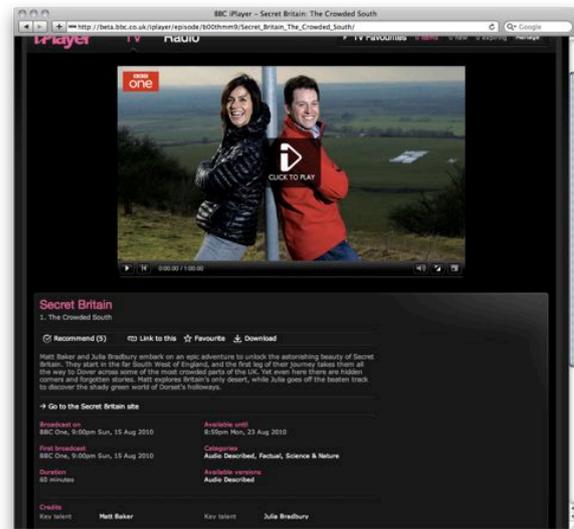
First, this is happening somewhat in Broadcast – and Broadcast programmes as delivered via the Web. Given the infrastructure of Broadcast, where a limited number of continuous streams of content – TV channels – can be delivered in parallel, these channels can be viewed as the Broadcast equivalent of such tagged aggregations. Without such constraints, such as when hosting programmes online with a view-on-demand model, there are examples of broadcasters who have followed the tagging pattern to some degree: BBC iPlayer has a category view, and programmes can be tagged with more than one category. The iPlayer currently in beta – v3 – includes production information taken from, and links to, a ‘programmes’ page. We will return to ‘/programmes’ in the following section, but for now we can note that there is increased linkage as per our tagging theme: linking by production information, and a list of related programmes.

Second, tagging is often an activity performed by the users, rather than the creators of the site. What is interesting in terms of drama, is that this means the categorisation is coming from how the product is perceived, rather than the intent in which it was made. This issue of intent vs. interpretation is one we shall return to.

Full text search on programmes would enable a whole range of functionality, some of which could be used for our drama-as-network thesis. With similar accuracy caveats as text-based web page search, programme listings, for instance, could be filtered by character or even onomatopoeia – bam! – and this information be used to locate the content within the programme. With such promise, it shouldn't be such a surprise that there is already such a project within the BBC. On a January 2010 posting to the BBC R&D Blog, Andrew McParland demonstrates work that successfully – or more pertinently, synchronously – combines broadcast delivery with internet delivery, and where subtitles are



BBC iPlayer v2, showing categories view



BBC iPlayer v3, showing linkage



BBC ‘programmes’, production and programme links

the demonstration application (Miller & McParland, 2010). With the subtitles now (privately) accessible to internet applications, the DataArt team built two visualisations that let you search for occurrences of words in the past seven days. While they were very excited by the results – finding clear trends around news events, for example, and discovering programmes they wouldn't otherwise watch, due to this extra insight into the topics covered – these benefits have not been made available to the public:

One thing I've come up against in the first few months of the project is that while departments across the BBC are eager to participate there are often obstacles to releasing the data. A good example of this is a visualisation I built in flash which used an API written by Andrew McParland from R&D to allow users to search for the instances of words mentioned on BBC TV channels. It threw up some really interesting results so I plugged the same API into a Microsoft Pivot to use the same data to create a new way to navigate through to iPlayer content. Due to concerns about the use of copyrighted material and issues about compliance these demonstrations were not able to be made accessible to the BBC's audiences on our website. (Littledale, 2010)

So here we have the nascent technical capability but corporate concerns holding back our potential for drama-as-network.

Therefore, in making an analogy between programmes and web pages to see what characteristics we could transfer and whether there are signs of this happening currently, we could conclude that:

- Programmes need to be viewed as containers of further content. We need to be able to get inside them and identify sections.
- There need to be linking mechanisms between these sections of content. There are no de-facto answers here on either the mechanism or how it is exposed to the user.

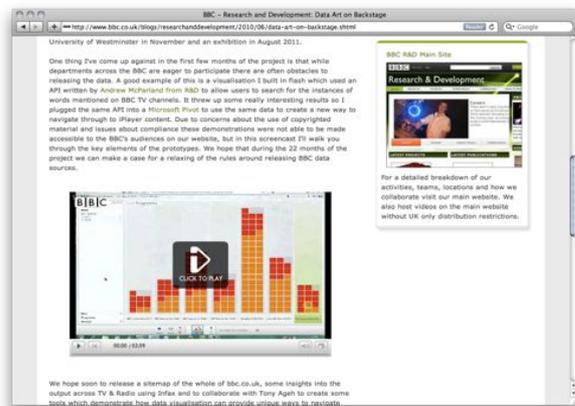
- The mediums of Broadcast and the Web are starting to blur, even through this narrow lens.

To follow the analogy through: currently, programmes tend to be viewed akin to discrete web pages, perhaps with a few tags at the bottom. We're interested in the in-line links that would be worked into the prose itself. As you watch a programme, it references events, characters are involved, interpretations are told... we want to be able to mark this up, to put this clip of that programme into a navigable world of such references, links and media. If a family drama such as Doctor Who has an episode set in the Battle of Britain, and the BBC has decades worth of output that touches on the Battle of Britain, that wealth of content should be available for that families' worth of follow-up interest.

Currently, the nearest equivalent we have is perhaps the experience of 'losing yourself' in a journey through wikipedia. Start searching for one thing, and find yourself on a journey you could not have predicted.



BBC R&D Blog, Subtitles + Internet = Advanced TV



BBC R&D Blog, Data Art post referencing subtitle data

Applying the trend, in a 3.0 Web

The term 'Web 3.0' is used here to denote the perceived future of the Web, typically centred on 'the semantic web' or 'the linked data web'. But these are abstract terms and largely a vision of the future, so let us start by looking at some counter, here-and-now, examples of what 'Web 3.0' is not.

According to a newspaper article that caught my eye, ['Johnny Carson is getting an upgrade for the YouTube era'](#). 3,300 hours of footage of an iconic American TV show that ran for 30 years and featured 22,000 guests ("Johnny Carson I Clip Licensing", 2010, Gold, 2010) has been digitised and a search tool created with access to full transcripts and associated metadata. This sounds very much like an example of the kind of thing we want, based on the previous section's analysis. However, on closer inspection, it proves not so. This new website functions as a device to allow a series of searches to be made from its 'licensing' home page. The issue is not that the actual video clip may be withheld at this stage for commercial purposes, but rather that there is no attempt to embody the network potential of journeying through this content. It's search, identify, go back and repeat. And while there is an archive of material whose content has been 'broken out' of the container of the original broadcast programme and made searchable, these clips are not addressable. There is no attempt to allow anybody on the internet to link into this site to reference a clip, or to allow contextualisation within the site linking out.

In summary, it is a bespoke application for a specific aim where the wealth of data is locked within whatever functionality that the specific application exposes. It would seem better for that data to be exposed in ways that could facilitate other functionality – that could take advantage of the medium. While the commercial parameters under which this project was undertaken cannot be known by this report, it is worth making the case that commercial parameters can be considered orthogonal to the availability of the data: if the



Los Angeles Times: Here's Johnny, digitized



JohnnyCarson.com: Our All New Clip Licensing Website - We've cataloged the entire Johnny Carson library, complete with transcripts. Search for clips, bits, guest stars, monologues and interviews.

monetisation comes from licensing the video clip, then facilitating more ways of these clips being found would be desirable.

Another example of what Web 3.0 is not, but perhaps the best present example of what a ‘Web adaptation’ could be of a linear programme, is something called ‘Prison Valley’.

‘Prison Valley’ is both a linear film – an hour long documentary structured as part road trip, part investigation, that has been shown on prime time French TV and at festivals around the world – and a website that presents the same world, but in a much expanded way.

The ‘web documentary’ – as its creators call – it comprises three main aspects:

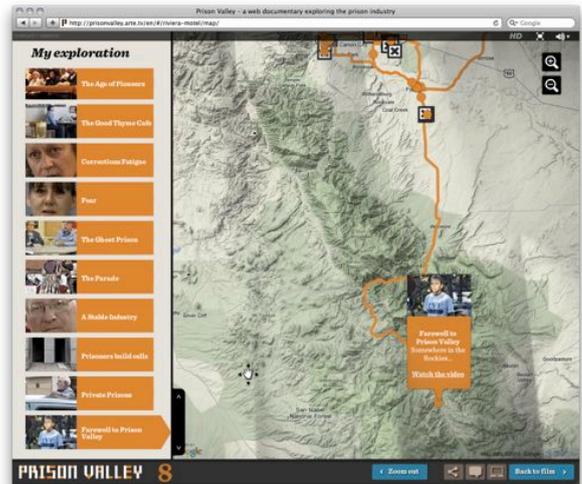
- Documentary film divided into chapters, navigable through a map plotting out the ‘road movie’ route.
- Journalistic evidence gathered through the investigation, such as more complete videos of people interviewed, photo galleries, and the odd souvenir.
- Community features such as forum discussions and live ‘webcast’ seminars with related experts and people featured in the documentary.

In comprising these aspects it realises the network effect, through telling its story with a web of connections that the user can explore, and also leverages the multimedia, social and realtime potential of the Web.

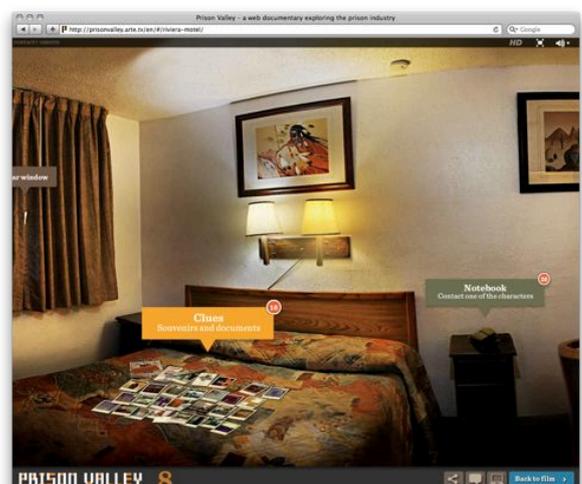
In my view, what is particularly compelling about ‘Prison Valley’ is the manner in which these elements have been combined. It takes advantage of the medium, but also walks the fine line between creating a feeling of sitting back and enjoying a good story, and allowing your interest to take detours or drop in and out.

The big ‘but’ is that ‘Prison Valley’ is a world unto itself. There is no way for its resources to become enmeshed in the wider Web.

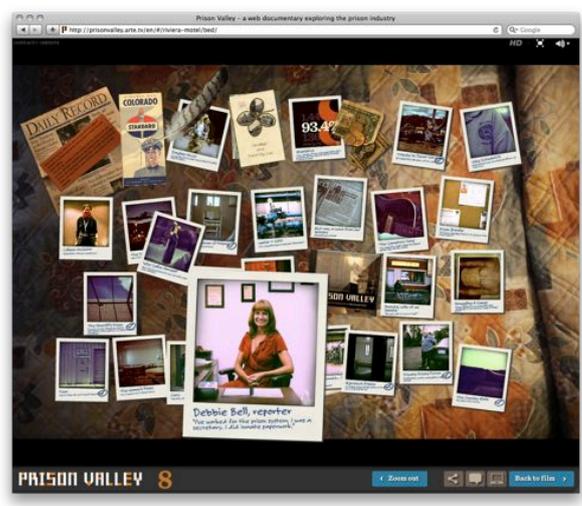
If the Web as we know it links the presentation layer – the documents that hold the information – then Web 3.0 is then about linking that information itself.



Prison Valley: A road movie, the journey as navigation



Prison Valley: Motel Room resources



Prison Valley: Explore the evidence further

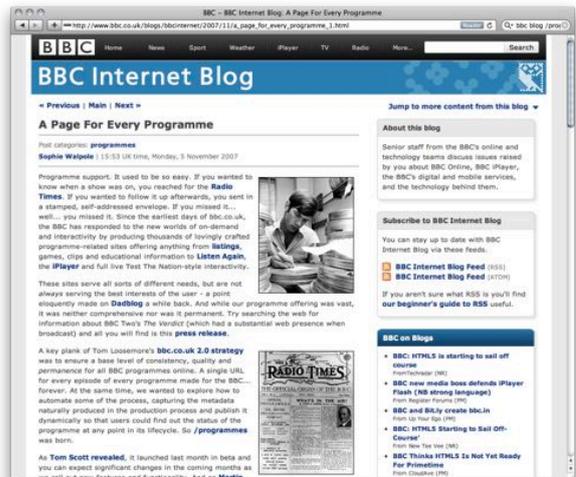
“Web 2.0 is a stovepipe system. It’s a set of stovepipes where each site has got its data and it’s not sharing it. What people are sometimes calling a Web 3.0 vision where you’ve got lots of different data out there on the Web and you’ve got lots of different applications, but they’re independent. A given application can use different data. An application can run on a desktop or in my browser, it’s my agent. It can access all the data, which I can use and everything’s much more seamless and much more powerful because you get this integration. The same application has access to data from all over the place.” – Tim Berners-Lee, inventor of the World Wide Web and Director of the World Wide Web Consortium. (Miller & Berners-Lee, 2008)

For our drama-as-network thesis, this sounds like an ideal environment. Information about the world is out there, in the world, we can add to it our pool of additions to serve our properties, and the drama experiences we want to build are the applications that sit on top of this integrated data.

Does the BBC somehow make the information about its programmes part of this vision of a shared, interoperable Web? The answer turns out to be yes:

A key plank of Tom Loosemore’s [bbc.co.uk 2.0](http://bbc.co.uk) strategy was to ensure a base level of consistency, quality and permanence for all BBC programmes online. A single URL for every episode of every programme made for the BBC... forever. [...] So /programmes was born. (Walpole, 2007)

BBC Programmes was launched in Summer 2007. Its goal is to provide a web identifier, with associated HTML pages and machine-readable feeds (RDF/XML, JSON and XML), for every programme the BBC broadcasts—allowing other teams within the BBC to incorporate those pages into new and existing programme support sites, TV Channel and Radio Station sites, and cross programme genre sites such as food, music and natural history. (Raimond et al., 2010)



BBC Internet Blog: A Page For Every Programme

The BBC W3C Case Study quoted above goes on in a manner that perfectly matches our aspirations for Drama:

BBC Music follows the same principles as BBC Programmes, and provides a web identifier for every artist the BBC has an interest in (featured in music programmes, in BBC events, etc.). BBC Music is underpinned by the Musicbrainz music database and Wikipedia, thereby linking out into the Web as well as improving links within the BBC site. BBC Music takes the approach that the Web itself is its content management system. Our editors directly contribute to Musicbrainz and Wikipedia, and BBC Music will show an aggregated view of this information, put in a BBC context.

So if, instead of every artist the BBC has an interest in, it was every 'thing' Drama were interested in, the wider Web would supply the context of what, say, a Viking was – also bringing the definition of Viking as 'a type of seafaring Scandinavian seen from the eighth to eleventh century A', and not one of the 70 alternatives currently listed on the Wikipedia disambiguation page – while the BBC website would host the characteristics of their Viking protagonist.

The idea of facilitating journeys across BBC content is also here, although currently it does not reach the programmes themselves.

The BBC Search Team is building on this newly-available linked data, by creating pan-BBC aggregations of content. These new pages are called Search+ – an indication that we are using this data to enhance the standard search experience. Each Search+ page shows the best content on its particular topic from around bbc.co.uk, and sometimes selected content from outside the BBC. The set of topics that make up the Search+ Pages essentially form a new BBC Controlled Vocabulary (CV) of concepts and entities. Each term in the BBC CV has an associated DBpedia resource, to enable us to use some of the metadata within DBpedia, and also enable links between our CV terms (and

associated Search+ pages) and content both inside and outside the BBC. We also intend to use sources other than just DBpedia (e.g. Musicbrainz and Geonames) to provide these 'Linked Open Data' associations in the near future.

So, it seems, this is the ideal environment to host our drama-as-network thesis, requiring us to expose our content in sufficient detail or granularity to facilitate the drama experience we desire.

Transmedia and Crossmedia

We have, so far, explored the idea of a 'Web adaptation' of a story, through exploring the insight that drama and networks could be native to each other. So what is going on outside of that insight? Quite a lot, as it turns out. The terms Crossmedia and Transmedia are both phenomenon largely associated with the rise of the Web, and are all about storytelling through different adaptations and mediums.

Crossmedia is very much the topic of this project, being the same story expressed across different media: what would the web one be like? Where Nicoletta Iacobacci says "It is generally the same program re-edited for different screens, fragmented content disseminated on different platforms, possibly incorporating extra content and channels to extend the viewers' experience", the prime motivation given is return-on-investment for the broadcaster. This project posits that the Web could allow that fragmented content not just to be a series of format-shifted cut-downs, but components of a story world that can be reassembled by the viewer in way that is a worthy drama experience in its own right.

Transmedia, however, is very different. While this project has aspirations to allow the viewer to interact with the story world, bringing it closer to transmedia's view of the viewer more as a player in a game, transmedia is about utilising many mediums in pursuit of a story experience. If crossmedia is one story in different mediums, transmedia is typically many stories in many mediums, with the user an active participant in exploring a story world bigger than each individual story. This gives fantastically interesting dramas that involve the Web – imagine a character you read about online phoning you up – and play out through things native to the Web such as social networking sites, but these are not Web adaptations. At time of writing, the state of the art is presented by Christy Dena in her PhD thesis 'Transmedia Practice: Theorising the Practice of Expressing a Fictional World across Distinct Media and Environments'

How do crossmedia and transmedia differ? Both are about content in a multiplatform distribution strategy. Both utilize the web as the main engaging space. Both relate to TV as one, maybe the most important, but just one of the media used to tell the story.

In a crossmedia environment, content is repurposed, diversified and spread across multiple devices to enhance, engage and reach as many users/viewers as possible. It is common to call crossmedia "content 360". It is generally the same program re-edited for different screens, fragmented content disseminated on different platforms, possibly incorporating extra content and channels to extend the viewers' experience. Brand here plays a key role and needs to be always identifiable. A typical form of crossmedia is when the plot of the story ends with a call-to-action, and drives the audience across different media. A good example is the BBC's Spooks, where, at the end of the TV episode, a cheerful announcement gives directions to a website.

In transmedia storytelling, content becomes invasive and permeates fully the audience's lifestyle. Stephen Erin Dinehart, who coined the term transmedia and created the VUP (viewer/user/player) relates this model to Richard Wagner and his concept of "total artwork" ("Gesamtkunstwerk") where the spectator becomes actor/player. A transmedia project develops storytelling across multiple forms of media in order to have different "entry points" in the story; entry-points with a unique and independent lifespan but with a definite role in the big narrative scheme.

Excerpt of 'From crossmedia to transmedia: thoughts on the future of entertainment', Nicoletta Iacobacci, Head of Interactive TV at the European Broadcasting Union. (Iacobacci, 2008)

At the BBC

Given that the BBC is the largest broadcasting organisation in the world (“BBC - About the BBC - Home”), and in its corporate structure has ‘Future Media and Technology’ as one of four content groups, you would expect to find traces of similar projects and thinking. And you would not be wrong.

SUDS

Prompted by an Eastenders website redesign in 2003, Celia Romaniuk – like Paul Rissen, an Information Architect – created a system intended to model soaps: ‘SUDS is Used to Describe Soaps’ (Romaniuk, 2003). From our perspective, amazingly prescient: it was ‘Eastenders in RDF’. Focussed on characters, it took the ‘friend of a friend’ (FOAF) ontology ‘devoted to linking people and information’ (Brickley & Miller), and extended it with the aim of modelling information sufficient to facilitate answering questions such as:

- › How many wives has Ian had?
- › Who have all the Queen Vic landlords been?
- › We hear about Phil's father a lot but has he ever appeared in the show?
- › Why did Frank leave?
- › Which characters have appeared in other TV shows?
- › Which actors have played Mark Fowler?
- › Does Steven know that Ian is not his real father?

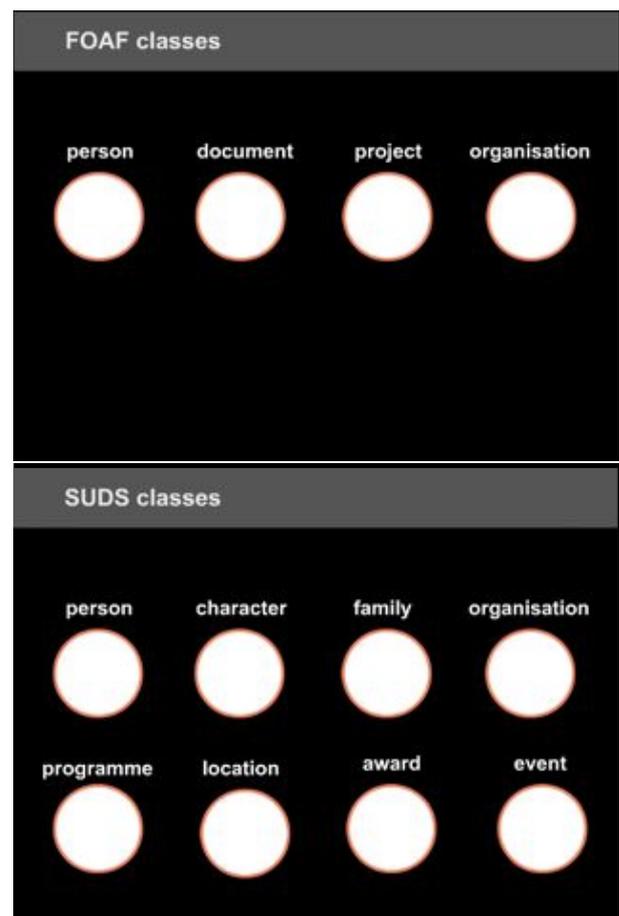
and to provide timelines for:

- › Characters
- › Families
- › Locations

as well as guides to events such as:

- › Past Valentine's Days
- › Weddings.

There is little trace of this project however. The general view of those I have asked in BBC FM&T is that the project while visionary was somewhat ahead of its time technologically. As an example, SPARQL, the RDF query language that facilitates those kinds of questions



SUDS Presentation: FOAF classes and the additional classes of SUDS

to be asked, only became a W3C recommendation in 2008 (“W3C Opens Data on the Web with SPARQL”). Perhaps more importantly, there was no wider movement within the BBC at that time working on integrating a linked-data publishing approach with their conventional web publishing systems.

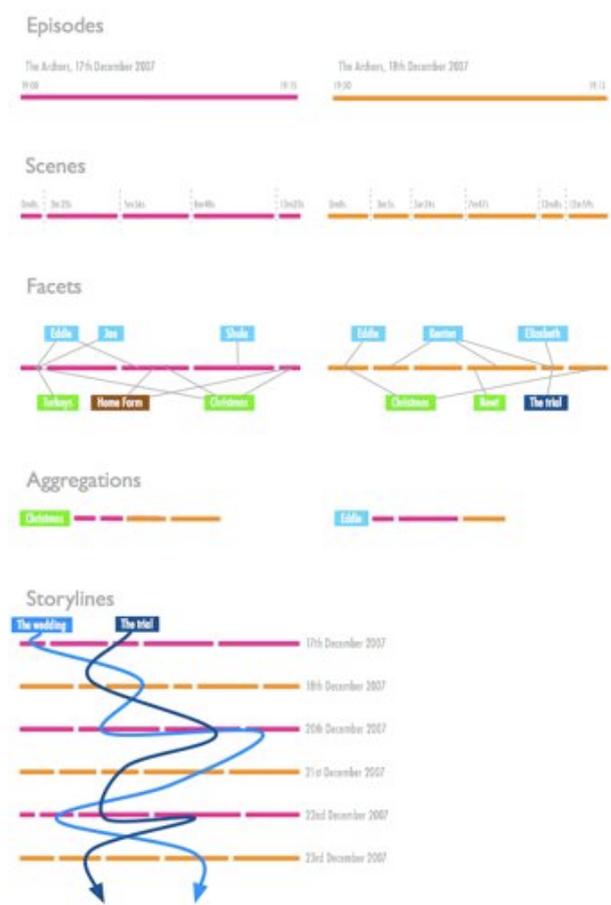
This is a shame, because in the SUDS presentation it is clear that the thinking was not just to provide some family trees for the website-as-was, but rather that viewing the EastEnders property exclusively as a series of TV episodes doesn’t serve what the users want: the stories, the things of interest, are ongoing, threaded through the seasons and episodes. And that the Web presents a real opportunity to represent the property in this way.

Archers Prototype Website

More recently, and again for a long-running soap, there was an experiment to enhance a BBC property’s website that allows viewing the world of the property through its characters, storylines, and other facets.

The ‘Archrs’ project had:

- a data-model representing
 - scenes
 - characters
 - character relationships
 - locations
 - storylines
 - tags
- a production tool that production staff could use to
 - segment podcast episodes into scenes
 - write title and description text for each scene
 - tag each scene
- a skeleton user-facing website
 - with an aggregation page for every
 - episode, scene, character
 - family, storyline, place, tag
 - podcast feed for every one of the above



‘Archrs’ Prototype RadioLabs Blog Post:

Episodes, Scenes, Facets, Aggregations, Storylines.

But most of all the project had conclusions we can learn from. Tristan Ferne, representing the Radio Labs team, writes:

Social history – The Archers often includes topical events and current affairs and generally represents the era in which we now live. So having an accessible, findable, addressable archive on the web going back 50 years could be a valuable social history resource.

Going smaller – Are scenes and characters the smallest units we can chop this drama into? Practically I think they are, though one could imagine going towards a finer grain of detail with marked-up conversations, characters' movements between locations, internal motivations...

Filling the 'static' data (e.g. characters, locations etc.) is basically done, some of it by some clever code that scrapes the Archers website. Annotating an episode took me about 10–15 minutes – but I'm not an expert and I wasn't annotating it that thoroughly. So starting from now it's not too bad and I think it's feasible. Annotating the entire Archers archive? That's another matter.

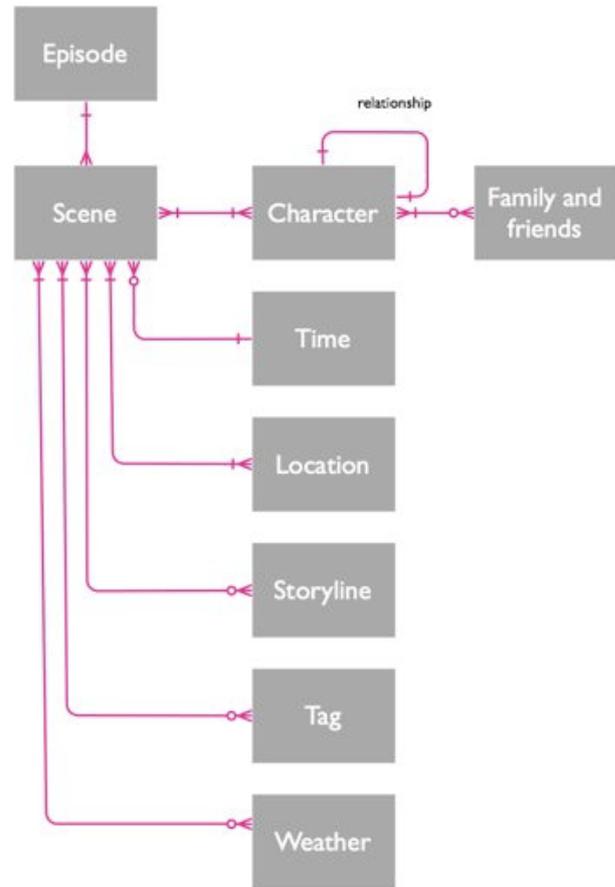
[Annotation:] Maybe it could be a community thing ?

Actually making the prototype live wouldn't be particularly hard but there's a lot of work around making a decent, usable site from this data.

It is also interesting from a transmedia point of view:

Thanks to the annotations we can create in the production interface and the superbly detailed episode synopses on the Archers website we should be able to build something that works as an audio-less, text-enhanced service.

*[Given] my initial thoughts about how the Archers is in some kind of pseudo-real-time. My favourite idea would be to deliver the scenes from the Archers to the audience *as they happen* – so at 11 in the morning, when over in Ambridge Kathy has gone over to Home Farm for coffee, your podcast application gets updated in real-time with that latest scene. Or maybe with some*



'Archrs' Prototype RadioLabs Blog Post: Data Model

The screenshot shows the 'Archers: 080219 Tuesday' production tool interface. At the top, there are navigation tabs for various data types: People, Maps, Locations, Episodes, Relationships, Relationship types, Scenes, Storylines, Social groups, Weather, and Tags. Below the tabs, the title 'Archers: 080219 Tuesday' is displayed, along with a warning: 'Do not refresh this page or use your browser back button! Cancel (all edits will be lost) Comments'. A timeline at the top shows a playhead at 00:12. Below the timeline, there are buttons for 'Jump to playhead', 'Start marker', 'End marker', 'Delete', and 'Done edit'. A table below the timeline lists scene segments with their start and end times:

Start Time	End Time	Description
01:44	02:22	
00:31	01:44	Adam cobby tells Brian that he's viewing a digester in Scotland next week and gives Brian no choice but to look after the stock while he's away.
02:22	05:07	

'Archrs' Prototype RadioLabs Blog Post: 'Episode Editor' production tool

clever VOIP shenanigans we could phone you up and let you listen in to the conversations in The Bull

(Ferne, 2008)

Mythology Engine

The Mythology Engine represents the coming together of the previous two prototype projects, with thinking about the nature of the web and drama as exemplified by the 'Fourth Medium' post mentioned in the Introduction text.

Our objectives for this project were to build something that would demonstrate how you could express stories in a form tailored for the web, to show how this would allow people to explore BBC dramas and unlock the archive, and to create a reusable framework that could apply to all dramas and stories. (Ferne, 2010)

Expressing stories in a form tailored for the web is the developing theme of this introduction. Aiming for something that can 'unlock the archive' and 'apply to all dramas and stories' is what truly differentiates the Mythology Engine from our previous examples, and we will return to this in the following section.

But first, what is it? Internal to the BBC, there is a functioning website with content sufficient to demonstrate what the basic user-experience of such a thing could be and a version of this website loaded with content sufficient to indicate it can be applied to different drama properties. Available to the public is a 1,300 word Research & Development blog post featuring a three minute demonstration video.

As a prototype, the Mythology Engine was both a proof of concept built for the experience of doing so, and a device to further the ideas embodied in it towards being realised in a live, production context. This project exists because of that work, enabling actors within the BBC and Queen Mary University of London to scope a project aimed at developing these ideas.

So what is the core idea? Tristan Ferne, writing in the Mythology Engine blog post:



'The Doctor Who Mythology Engine' homepage



Mythology Engine homepage for a different property

Most TV drama on the web is either deep and detailed fan-produced sites or visually rich but shallow sites from the broadcasters. We believe there is a middle way and it seems like there's a space for something here. Something that expresses the richness and depth of the stories that the BBC creates. Somewhere that will be the default place to find out about our stories and somewhere that people will link to and share with their friends. [...] One way of thinking about this that I've found helpful is to imagine the story existing in the writer's head before the scriptwriting and production creates the broadcast programme. The viewer then watches this, understands it and reconstructs the story. The Mythology Engine is designed to assist in this process; to let the audience explore complex plots or catch up on episodes they missed or stories they remember.

Complementary to this idea of completing the circle back from programme to the story that was in the writer's head is the notion that the Mythology Engine is not a vehicle for watching the full-length programmes, rather a way of structuring whatever elemental media there may be that contributes to that story.

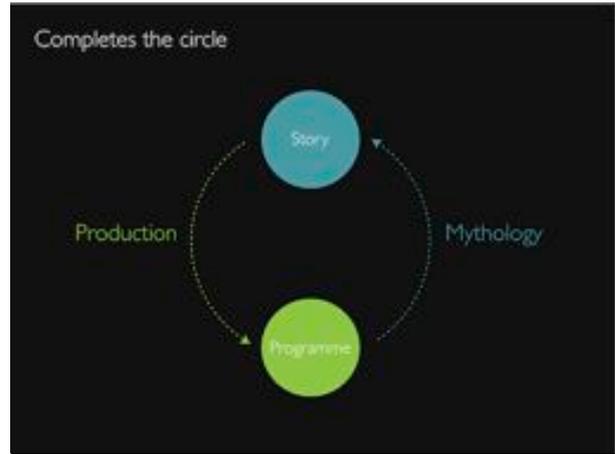
To get an idea of how this translated into the site, an excerpt of Theo Jone's narrated 'walk-through' demonstration of the site is featured on the following page.

Like the Archrs blog post, the Mythology Engine post concludes with a number of points for further consideration:

Investigating whether we could parse scripts, subtitles or video to automatically create the outline of the data for a story.

Using this framework to tell the stories behind the news and sport and to further explore the archive.

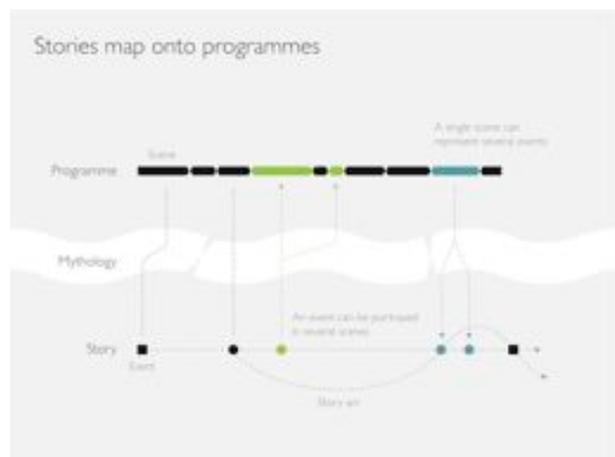
Looking at how user-generated content would be fitted into this framework. Is it something that sits on top? Or is it more fundamental than that and could we harness the fans to create the mythology for us?



Mythology Engine Blog Post: Completes the circle, from programme back to story



Mythology Engine Blog Post: 'Story' is made of several 'Events' which feature 'Characters', 'Things' and occur in 'Places'



Mythology Engine Blog Post: Stories map onto programmes; A single scene can represent several events, an event can be portrayed in several scenes, a programme might have its story and other story arcs running through it.

And should this model and framework make us think differently about how we write and produce stories?

Could we start to create narratives that are tailored for the web?

To these I would add the following questions related to statements made in the post:

Because there are always issues around the rights of distribution of programmes we designed it to work with and without short video clips, as these seemed relatively realistic to have. There is no long form video in the prototype for this reason, and also because it's not designed as a replacement for iPlayer. It should be complimentary to existing BBC sites.

Question: given the Web can host an array of media forms, could we build something that is transmedia by desire rather than necessity?

By designing the Mythology Engine to take advantage of architecture of the web with unique pages per concept and interconnecting links everywhere, we increase the findability and sharability of our content.

Question: could we also create good drama in doing so, if Drama has inherent network characteristics?

Stories are then collections of events, where an event is a specifically chosen, significant thing that happens in a story. This could be anything, but the important thing is that it is editorially chosen to tell the story.

Question: What constitutes an event is then highly interpretable: how do you define 'significant'? Moreover, that 'the important thing is editorial choice' is an assumption: perhaps with new mediums such as the Web come new possibilities, ie crowdsourcing, and can we even say one view is correct, be it the author's, some production staff, or a viewer?

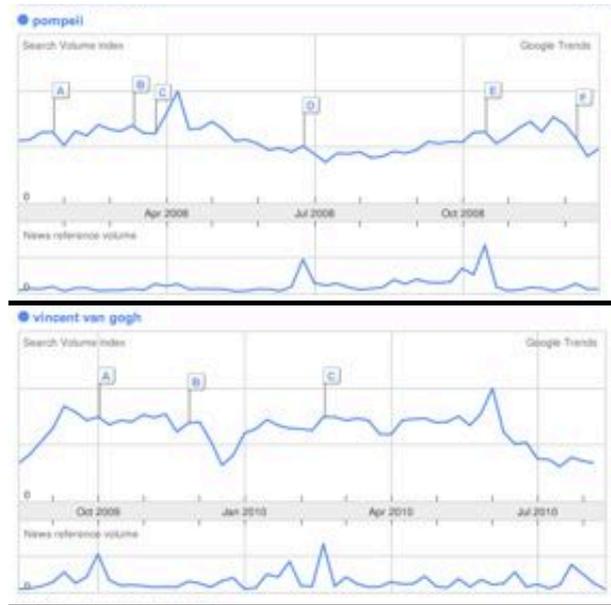
"Let's have a look at Genesis of the Daleks, a story from the fourth series with Tom Baker, first transmitted as six episodes in March 1975. This is a timeline showing the major events that make up the story. These events are editorially chosen highlights. In this case we've condensed the six episodes down to seven major events, focussing on the story of the Daleks. We didn't include any long-form video on this page, as we felt it was more about exploring than viewing. The synopsis gives a taste of the story, a fuller spoiler description is hidden at the bottom of the page. The featured characters, things and places, all link to other pages in the Mythology Engine. Let's explore one of those events in more detail. The Doctor is being interrogated by Davros. We've included some short-form video on this page, a couple of minutes seems about right. In this case there are two clips, as the event was originally a cliff-hanger between episodes. Ideally every event would have a video clip, but it seems to work fine without one. There's a snapshot of the timeline, showing your place in the story. And all the featured characters, with descriptions of their involvement. The Mythology Engine's richness comes from its interconnections. Here we have references to other stories. This one links off to the Daleks' invasion of Earth, mentioned in the clip above. If we clicked on this, we'd dive even further back into the archive. And this one works the other way round. A reference from a future story to here. In a more recent story, Sarah-Jane recalls this interrogation. We can click and arrive at that moment. So we've come from an archive story to a contemporary story. We skimmed over the characters before. Let's finish by clicking through and learning more about the evil master-mind, Davros. The simple visualisation at the top of this page is a rudimentary view on his relationships and encounters. It could do with some more work, but here we can see the Daleks, the Kaled bunker, and the fourth Doctor. Further down the page are his connections and relationships. We can see he's the creator of the Dalek race. Then we have all the places he's visited and the stories he's featured in."

Excerpt from Mythology Engine video narration.

BBC Round-Up: the Pompeii Effect and a call to action

 Reviewing SUDS, Archrs, and The Mythology Engine, it is clear that for some time people have been trying to express BBC properties on the Web in a way that is fundamentally different to how programmes are structured. There is a tacit or overt acknowledgement that the stories that drama properties tell are somewhat separate to the media the production process renders them into. While this is self-evident in a cross-media and transmedia world, the application to the Web seems both obvious and utterly opaque. It seems easy to look within a story and identify component parts, and these component parts seem a natural fit for being expressed on a network of information such as the Web. It seems obvious to ask for aggregations of those component parts in ways that are not by programme, but rather some form of facet: ie all scenes with character 'x'. But none of these prototypes have gone live, none have been filled with content, none have been experienced by a mass-user base, and nobody I have spoken to over the course of this project is talking confidently about just how such a thing might 'feel'. Just what might it be like, exploring something akin to an encyclopaedia of the story world? How much drama, in the dramatic sense of the word, might people want in that experience? What role might people use it for in relation to other adaptations of itself? Or other entertainment forms?

There is one concrete thing known about the here and now by BBC Future Media and Technology that can give us an insight as to what should be created and why: Doctor Who generates a lot of search traffic. A sharp rise peaking in the highest volume of searches conducted that year for 'Pompeii' correspond to the transmission of 'The fires of Pompeii'. Likewise for the search term 'Vincent Van Gogh' this year, corresponding to transmission of 'Vincent and the Doctor'. That is, people are following-up on topics viewed via broadcast by searching for information on the internet. And the BBC, most likely, has a wealth of programming on that topic already. The BBC could be facilitating exploratory



'The Pompeii Effect': Google Search Trends can be used to observe a dramatic rise in search volume for a factor of that week's Doctor Who episode. April 2008: "The fires of Pompeii". June 2010: "Vincent and the Doctor"

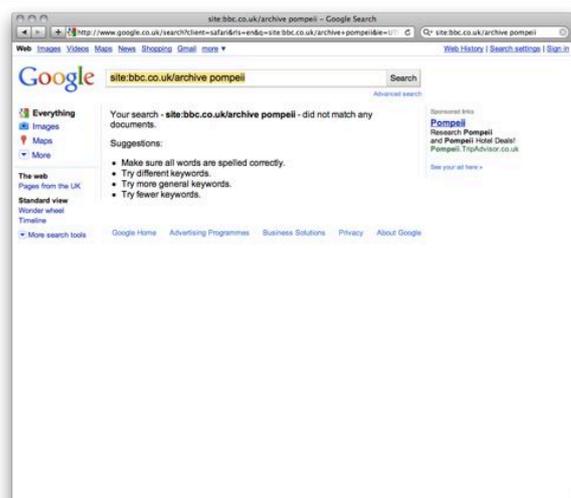
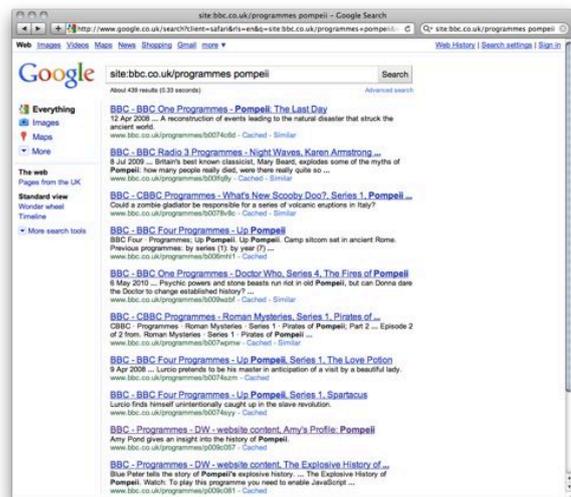
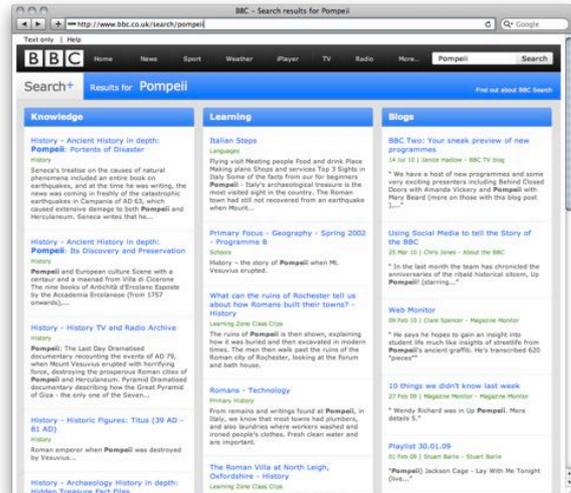
journeys not just through its own drama properties, but through its whole range of output.

Currently you can search within the BBC website, and for common terms you are returned a 'Search+' page with results categorised into 'Knowledge, Learning, Blogs and News'. This is a great resource but is limited in scope to website content: it does not turn up programme results currently. Thanks to the thinking behind 'programmes', this can be done by a standard internet search restricted to the domain 'bbc.co.uk/programmes'. So work is being done to realise the potential here, but to return to our theme there is neither a way of offering the user a contextual, click-through journey through programme content as discussed at the beginning of this chapter, or a way of curating journeys through disparate programmes and other BBC resources.

It should be noted that this vision is predicated on the BBC making its programming archive available to view on demand, which it currently isn't but the March 2010 BBC Trust Strategy Review clearly states this as an objective:

It should make the universal availability of its archive a key objective over the next ten years, creating an engine for new public value—connecting audiences with the best of everything the BBC has ever made. ("BBC Strategy Review March 2010")

How best to connect those audiences? The Mythology Engine points to a way of exposing programme content for exploration by users and a provides a way for editorial staff to tell stories through that content, but the project in scope is limited to Drama and its technical implementation is bespoke and its data 'siloe'd'. The 'programmes' domain points to a web-native infrastructure with the necessary completeness, permanence and modelling for cross-linking, but the modelling only encompasses production rather than content concepts.



Search results for 'pompeii' using: BBC.co.uk homepage search (top); a Google search restricted to bbc.co.uk/programmes (middle); and a Google search restricted to bbc.co.uk/archive.

All retrieved August 2010

This, then has to be the call to action: combine the best of the two worlds exemplified by ‘/programmes’ and ‘The Mythology Engine’, and in doing so transform forever what the BBC can be: not just an organisation broadcasting today’s timetable of content into the ether, but an organisation using that as the leading edge of a forever-building archive, interconnected for exploration, and editorialised as the fourth medium.

“Seeing the BBC as a swarming ball of amazing content is always how you should see it. You should always know that there's just tons of stuff. I know that when I go to Wikipedia it seems as if its got a page on everything, its got three million pages, one about everything, I can kick around it, I know its great. I go to the BBC and i search for 'x' and I get the one page that I was just on, but I know that its got a ton of content about that topic. Rather than read a Wikipedia page, lets watch a documentary. Lets see the big BBC through whatever lens you've chosen” - Henry Robbins

“At the moment we make programmes. And we make web pages. But i would like us to abstract the information from the things we make: all of the brilliant information and the connections that journalists, writers and production generally makes, that should be a big web, that should be the core of the BBC. On top of that people have plotted and extracted things in order to make programmes or websites or the app for teachers. For me its very much that moving us to the next level of just talking about information itself, relating the concepts to each other so that we build that big store of information and then we can start building the different useful views.” - Paul Rissen

*Quotes from conversation, 27th April 2010, between
Toby Harris - Author of this report
Henry Robbins - BBC Learning
Paul Rissen - BBC Future Media and Technology*

Research Problem

‘How can we get a powerful journey through our content now – and what investment do we need to inspire to get to the vision’. – Jonathan Tweed

The answer is the call-to-action in the previous section. For the powerful journey now, the ideas of the Mythology Engine need to be crossed with the practice of ‘programmes’. And for that to happen, programmes need to be modelled.

The research problem is then to create a way of modelling content that

- presents as universal, minimal core as possible
- is extensible as differing needs arise
- facilitates the BBC in its editorial role
- can be applied to fact and fiction
- is based on semantic web technologies as implemented at the BBC

And starting with Doctor Who probably isn’t such a bad idea: story arcs through decades of production, story lines mixing overt fancy with history, not to mention regeneration and time-travel to accommodate.

Literature Review: Towards modelling programmes

First we shall look at the state of semantic web technologies at the BBC to establish the technological framework for this project, and then we shall develop approaches to modelling content starting with criticism of the semantic web and its assumption of a shared world view.

Semantic Web technologies and the BBC

The Web publishing state of the art at the BBC was recently showcased in two BBC Internet blog posts. Entitled ‘The World Cup and a call to action around Linked Data’ and ‘BBC World Cup 2010 dynamic semantic publishing’, the latter especially sounds like the implementation of ‘semantic web technologies’ as per our research problem. But what is ‘linked data’?

Linked Data

Applications for the Web that are enhanced by using a data-set of another service are commonplace today. These are facilitated by a Web service publishing an API that gives programmatic access to that Web service’s information. Unfortunately, in practice these data-sets are rarely interoperable, and the APIs do not present a uniform interface to the data. The data is therefore ‘siloes’, with developers unable to implement applications against all the data available on the Web. (Bizer, 2009; Hausenblas, 2009).

In the seminal 2001 paper ‘The Semantic Web’, Tim Berners-Lee, James Hendler and Ora Lassila lay out a vision of expressing meaning and knowledge representation built on the same principles that made the Web so successful. The technologies outlined were designed to facilitate databasing and reasoning across the Web, and as such facilitate ‘applications implemented against all the data available on the Web’ as above, but moreover were presented as a route to the evolution of human knowledge itself.

In his 2003 rebuttal, Clay Shirky summarises his argument as “*The Semantic Web’s philosophical*

argument – the world should make more sense than it does – is hard to argue with. The Semantic Web, with its neat ontologies and its syllogistic logic, is a nice vision. However, like many visions that project future benefits but ignore present costs, it requires too much coordination and too much energy to effect in the real world, where deductive logic is less effective and shared worldview is harder to create than we often want to admit”. Shirky finishes by returning to the present-day API’s we started this section with: “*Much of the proposed value of the Semantic Web is coming, but it is not coming because of the Semantic Web. The amount of meta-data we generate is increasing dramatically, and it is being exposed for consumption by machines as well as, or instead of, people. But it is being designed a bit at a time, out of self-interest and without regard for global ontology. It is also being adopted piecemeal, and it will bring with it with all the incompatibilities and complexities that implies. There are significant disadvantages to this process relative to the shining vision of the Semantic Web, but the big advantage of this bottom-up design and adoption is that it is actually working now.”*

Hausenblas and Bizer argue that, as of 2007-9, it was starting to not work, and there were concrete things that were being done about it. To paraphrase, ‘Linked Data’ is Semantic Web technologies without the philosophy and a practical focus of getting data to developers now. As proof of the efficacy of this re-alignment, projects such as the W3C Linking Open Data community project and the DBpedia project – both of which Bizer initiated – were seeing take-up by developers and organisations, including the BBC.

It was Tim Berners-Lee who set out the defining document about Linked Data in 2006:

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. The URIs identify any kind of object

or concept. But whether for HTML or RDF, the same expectations apply to make the web grow:

- Use URIs as names for things
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
- Include links to other URIs. so that they can discover more things.

The core idea of Linked Data is to take HTTP URIs – universal, unique – and use them to represent anything. We have already seen in the introduction that the BBC has taken advantage of this – “/music provides a web identifier for every artist the BBC has an interest in (featured in music programmes, in BBC events, etc.)” – and that they are using the capabilities of Linked Data not just to aggregate BBC authored content with content available elsewhere on the Web, but as a mutually beneficial arrangement due to better linkage across the Web and editorial time spent on shared data sources.

Linked Data at the BBC

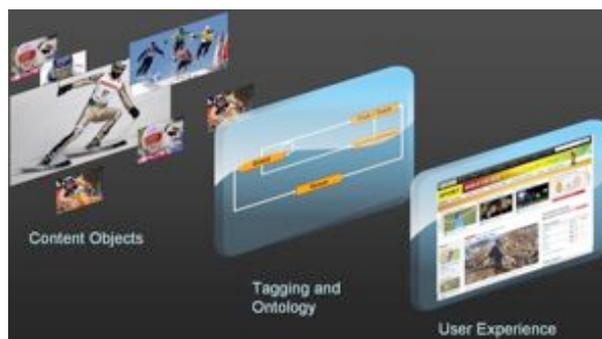
So what is the state of the art at the BBC? A July 2010 BBC Internet Blog post titled ‘The World Cup and a call to action around Linked Data’, hails a step change:

‘Though we have been using RDF and linked data on some other sites (such as BBC Programmes, BBC Wildlife finder, Winter Olympics) we believe this is the first large scale, mass media site to be using concept extraction, RDF and a Triple store to deliver content.’

The following is a synthesis of this article with another titled BBC World Cup 2010 dynamic semantic publishing. (Rayfield, 2010; Donovan, 2010)

Vision

- Easily and accurately aggregate content, find it and share it across many sources.
- From these simple relationships and building blocks dynamically build up rich sites and navigation on any platform.



BBC Internet Blog: The World Cup and a call to action around Linked Data. “We are not publishing pages, but publishing content as assets which are then organised by the metadata dynamically into pages, but could be re-organised into any format we want much more easily than we could before.” (Donovan, 2010)

- Use Linked Data to more accurately share this content and link out to other sites and content, a key goal for the BBC.
- Be at the cutting edge of development for the next phase of the Internet, Web 3.0.

Objectives

- Facilitate the publication of automated metadata-driven web pages that are light-touch, requiring minimal journalistic management, as they automatically aggregate and render links to relevant stories.
- Build on BBC Journalism
- Maintain accuracy through the metadata
- Implement at a scale that will deliver first for the 2010 World Cup and then for 2012 Olympics

The business need

- A page for every team, group and player in the 2010 World Cup would require more index pages than the existing BBC Sport site.
- Operating as per existing site with curation by an editor, automation rules and suchlike was judged unfeasible.
- Search technologies and previous methods for automation and metadata creation had proven to be sufficiently inaccurate: You don't want to get content

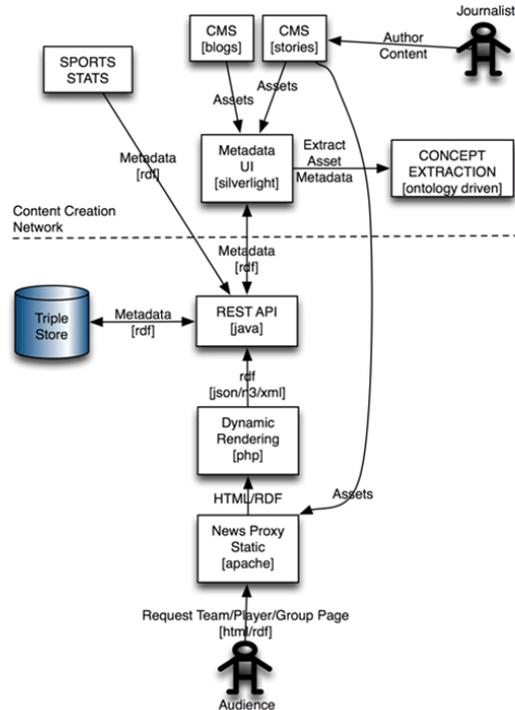
mixed up between different players with the same surname, for example.

Implementation – content in

- Journalist-published metadata is captured and transformed into an RDF representation.
- Automated XML sports stats feeds from various sources are delivered, processed, and transformed into an RDF representation. The transformation process maps feed supplier ids onto corresponding ontology concepts and thus aligns external provider data with the RDF ontology representation
- The journalists use a web tool, called 'Graffiti', for the selective association – or tagging – of concepts to content.
- In addition to the manual selective tagging process, journalist-authored content is automatically analysed against the World Cup ontology. A natural language and ontological determiner process automatically extracts World Cup concepts embedded within a textual representation of a story. The concepts are moderated and, again, selectively applied before publication. Moderated, automated concept analysis improves the depth, breadth and quality of metadata publishing.

Implementation – publishing approach

- The underlying publishing framework publishes metadata rather than content directly. The published metadata describes the world cup content at a fairly low-level of granularity, providing rich content relationships.
- Querying this published metadata enables dynamic page aggregations for teams, groups and players and semantic navigation.
- The description of the world cup is structured as per a ontological domain model: entity existence, groups and relationships between the things/concepts that describe the World Cup. The ontology also describes journalist-authored assets (stories, blogs, profiles, images, video and statistics) and enables them to be associated to concepts within the domain model.



BBC Internet Blog: BBC World Cup 2010 dynamic semantic publishing. "This diagram gives a high-level overview of the main architectural components of this domain-driven, dynamic rendering framework." (Rayfield, 2010)

Implementation – why RDF, triple store and SPARQL

- Facilitates agile modelling, whereas traditional relational schema modelling is less flexible and also increases query complexity.
- Capacity for inferred reasoning across the metadata based on ontological domain model. This inference capability makes both the journalist tagging and the triple store powered SPARQL queries simpler and quicker than a traditional SQL approach. Dynamic aggregations based on inferred statements increase the quality and breadth of content across the site.

Implementation – Technology 'Stack'

- All RDF metadata transactions are validated to ensure conformance to underlying ontologies and data consistency.
- Various flavours of RDF such as N3 or XML RDF can be ingested.

- The triple store is deployed multi-data centre in a resilient, clustered, performant and horizontally scalable fashion, allowing future expansion for additional ontologies and indeed linked open data (LOD) sets.
- The triple store is abstracted via an API designed as a generic façade onto the triplestore allowing RDF data to be re-purposed and re-used pan BBC. This service orchestrates SPARQL queries and ensures that results are dynamically cached with a low 'time-to-live'.
- The dynamic aggregation and publishing page-rendering layer is built using a Zend PHP and memcached stack. The PHP layer requests an RDF representation of a particular concept or concepts from the REST service layer based on the audience's URL request.
- The render layer will then dynamically aggregate several asset types. The resultant view and RDF is cached with a low TTL (1 minute) at the render layer for subsequent requests from the audience. The PHP layer dynamically renders views based on HTTP headers providing content negotiated HTML and/or RDF for each and every page.
- HTTP caching increases the scalability of the platform and also allows content delivery network caching (CDN) if demand requires.
- There are plans to exposing a public facing SPARQL endpoint. This will be a separate cluster/instance with a replicated snapshot exported from the live triplestore. We will be publishing our ontologies in tandem.

Validation

- This dynamic semantic publishing architecture has been serving millions of page requests a day throughout the World Cup with continually changing OWL reasoned semantic RDF data. The platform currently serves an average of a million SPARQL queries a day with a peak RDF transaction rate of 100s of player statistics per minute. Cache expiry at all layers within the framework is 1 minute proving a dynamic, rapidly changing domain and statistic-driven user experience.

Linked Data conclusion

.....

The previous exposition and synthesis of BBC documentation shows that from a technological point of view, the BBC is capable of implementing a semantic web technology backed platform to represent its programming as per the research problem. It also shows that there are the kind of resources outside of the BBC that would allow the degree of contextualisation an archive such as the BBC's would need: as the BBC's output is mostly about the world outside it, so the world outside it can provide that information through resources such as the linked-data version of Wikipedia, DBpedia. It also shows institutional desire to expose its data for use outside of the BBC, which specifically is very interesting in terms of engaging with fans of drama.

We also have a checklist:

- There are tools for manual, semantic-backed tagging of content
- There is a capability in automated tagging through natural language processing
- There is a belief in the long-term potential of representing BBC properties on the Web through separation of data/media and user experience as this will enable many applications built upon the same data. The applications can be multiform and ideas iterated through easily.
- The full BBC programme content can be represented on the web separately from the metadata, ie. the metadata can provide a sufficient experience in itself.
- We should use OWL with reasoning for our domain model
- We should generate RDF which will be hosted in a triple store
- We can use SPARQL queries
- We can build an API around our SPARQL queries
- We should anticipate having a separate SPARQL endpoint to expose this metadata as linked open data.

What remains unresolved are the more philosophical aspects Shirky discusses, and we will address this next in the context of modelling the content.

Modelling Content

In the video 'Web 3.0' by Kate Ray, Shirky reduces his many arguments about the viability of the Semantic Web to this:

And that gets very quickly to one of the deepest, you know, questions in all of Western philosophy, which is: Does the world make sense? Or do we make sense of the world? I don't think you can unambiguously describe the world. I don't think you can describe the world, or even large subdomains of the world, in a way that all observers or even most observers will agree with. (Ray, Shirky, 2010)

Thankfully, we are not aiming to describe the whole world, or put forward a vision that could. We are more limited in scope, and we have seen the technologies can facilitate real benefits. In fiction, we can also make claims about the world, for we can ask its creator directly. But the issue of agreement remains.

The question then becomes, does any one view trump another? Even if there was something inherent to that Doctor Who programme that a domain modelling group at the BBC can expose, would that representation have any relevance if it the programme is perceived differently amongst the audience it was made for?

There are two conclusions drawn by this project. First, that the notion of interpretation has to be built into the core of the modelling work to be done. Second, questions of authorial authority – of canon or correctness – are judged largely moot at best and counter-productive at worst: while the production organisation could be said to have a mandate to decide, any discrepancy between an imposed model and the perceived one will act to the detriment of the product. This line of thinking informs the next section.

Top-down, bottom-up, or both?

.....
On this matter of production of metadata, Faith Lawrence and M C Schraefel acknowledge it as a topic of some debate and go on to note "This division is most noticeable between those who believe in 'free tagging' and those who prefer the more formal construction of an ontology to define both the vocabulary of the domain and relationships of the concepts within it."

Their paper 'Freedom and Restraint Tags, Vocabularies and Ontologies' discusses how use of metadata has developed in many of the self-organising communities formed around amateur or fan fiction. They outline development from 'free-tagging', where users are free to add whatever text terms to the works of fiction, to 'folksonomy' where an emergent vocabulary develops. Observing "how a mature system may behave with the tags feeding the taxonomy which in turn informs the community as to what tags are expected and appropriate vocabulary", they argue that the system would benefit from application of some form of mapping between different vocabularies as synonyms are observed within, and between, these communities. They propose a system based on adding semantic relationship to the tags, linking them with external resources that can provide a shared statement. They also argue against Shirky's 2005 assertion regarding the superiority of free-tagging over structured ontologies by observing that their study group represents Shirky's ideal scenario for the assertion, and was judged in want of a structured ontology to link up the folksonomy terms and communicate their meaning across non-expert users as they emerge. (Lawrence & Schraefel, 2006)

While we can see that a hybrid of bottom-up and top-down approaches to creating metadata for fictional works is likely an optimum approach, this metadata has only been a characterisation for the fictional work as a whole, and we wish to model the work at a finer granularity, to open up the contents for contextualisation and so navigation. We shall address this next.

Media Information Spaces—A Semantic Challenge

.....
 An essay by Frank Nack in a promisingly titled 'Emergent Semantics' feature of IEEE Intelligent Systems provides eloquent insight into the domain. The part of the essay characterising a transition away from a society used to a hybrid system of traditional media and digital media and into "a knowledge space that facilitates new forms of creativity, knowledge exploration, and social relationships mediated through communication networks" reads as a manifesto for this project and has to be read in full.

"A basic aspect for such a space, which supports individuals but is still communal, is that information must be made accessible that is hidden in the unified structure of the single text, image, video, audio, or tactile unit. Thus, the goal is to create an environment in which media units and the relationships among them are understood as basic elements that can interrelate to produce new meanings.

To support this process of generating meaning, interpretation, and visualization, a system must know what is contained in the different media. For visual media, however, this poses a problem. Even though an image might provide a limited amount of visual information, it contains a wealth of meaning. This functionality is based on the two formal structures that can be assigned to every perceivable object in visuals: the signifier (which carries the meaning) and the signified (which is the concept or idea signified). The relation between the two elements is not a naming-process only, as the signified resembles not a thing but a concept. Secondly, the relation between the signifier and the signified is arbitrary. It is, in particular, the arbitrariness of the relationship between signifier and signified that enables the creation of higher-order sign systems and their diversity.

Thus, visual media requires more than characterizing its visual information on a perceptual level using objective measurements, such as those based on image or sound processing or pattern recognition. Creatively

reusing material for individual purposes, which usually opens up questions of aesthetics and subjective interpretation, has a strong influence on the descriptions and annotations of visual media data, either created during the data's production process or added later. Providing semantic, episodic, and technical representation structures that can change and grow over time is important. This also requires adaptable relations between the different type of structures. [...]

Yet, if we only had the information gathered during the production of media, including its reuse and modifications, we would still lack knowledge about the material's potential intrinsic meanings. Thus, it is important to make people aware that the notion of a completed work vanishes in such a system and leaves space for a creative and productive cycle, a living environment allowing all sorts of processes. These spaces are for investigation based on an interpreting, associative method rooted in a discourse-oriented collective interpretation of questions that, by following the branches of interdependencies, compare the most diverse theories." (Staab, Santini, Nack, Steels, & Maedche, 2002)

However it has no answers on how to achieve it. Or rather it suggests the Semantic Web and MPEG-7 are going to be the first realisations of this, without actually saying how. So what is MPEG-7 and does it have anything to do with the Semantic Web?

Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology

.....
 Jane Hunter's 2001 paper sets us on the right track: "The goal of [the MPEG-7] standard is to develop a rich set of standardized tools to enable both humans and machines to generate and understand audiovisual descriptions" and outlines her work translating the XML structures of MPEG-7 into a DAML+OIL Schema. The W3C incubator group that reviews Hunter's work in context of other MPEG-7 ontologies developed since have only published once in 2005 (Celma et al., 2007 – the page datestamp is

2007, but the content and url suggests the writing is from 2005), and it seems the state of the art is represented by the 2009 paper “A semantic-based framework for multimedia management and interoperability” (Tsinaraki & Christodoulakis, 2009). However, while MPEG-7 will very likely play a role in our modelling programmes, the more that is read about it the clearer it becomes that we’re missing the point somewhat: we can get very technical about how we model the media and stick the metadata on, but we are not addressing the core need of this project: to facilitate stories, to allow a powerful journey through content. First and foremost, we need a language to describe what we could call the story potential of the media.

Enter OntoMedia (stage-left)

OntoMedia is an ontology developed to allow the semantic annotation of the narrative components of media and the relationships between those components. It also “includes the annotation of monomedia (drawing on existing technologies such as the CIDOC Conceptual Reference Model, ABC, and the Functional Requirements for Bibliographic Records) and multimedia (drawing on MPEG and audiovisual media metadata standards)” which well encompasses the terrain of the previous section. (Lawrence et. al., 2006). The first statement sounds like it could meet “the story potential of the media” and the second statement neatly deals with the relationship to the actual media materials. As such, it deserves close inspection. The following is an extract from an early OntoMedia paper which characterises the domain well in addition to outlining the ontology.

The OntoMedia content model is divided into Events and Entities. We define an Entity as an object or concept. An Event describes an interaction between one or more Entities during which zero or more attributes of those entities are modified or a new entity is created. A number of Events subclasses are defined based on the most common types of events found in literature. The subclasses of the Entity construct fall into three different

types. Those related to objects both physical and abstract (ontomedia:Characters and ontomedia:Items), those related to spacial models (ontomedia:Space) and those relating to time (ontomedia:Timeline and ontomedia:Occurrence).

Entities of the ontomedia:character are defined as having a personality. For example a toaster would be a type of ontomedia:Physical Item, a sub- class of ontomedia:Item, whereas the talkie toaster from the television show Red Dwarf and the related books would be classed as an ontomedia:Character despite having the physical appearance of a toaster. The class ontomedia:Item is subclassed into physical and abstract items. An ontomedia:Occurrence is a specific instance of an event which occurs within a single ontomedia:Timeline. The reason for this is that any ontomedia:Inexpressible Content can have a ontomedia:Timeline associated with it. It is therefore likely that more than one time line instance will need to be defined for any given work. One event will there fore occur on multiple time lines but the relationship between the events on any given time line will not necessarily be the same as on any other time line. For example if we define one time line to describe the events which occur in the narrative and another to describe the events chronologically these two will differ whenever we encounter a ‘flashback’ or one of the characters mentions either a historical event or something from their past. This flexibility regarding time is especially important since stories may involve time travel and related paradoxes. While a 1:1 mapping exists between an instance of ontomedia:Timeline and an instance of ontomedia:Occurrence, a 1:Many relationship exists between any instance of an ontomedia:Event and the instances of ontomedia:Occurrence that contextualise it. This can even allow more than one occurrence of the same event to exist on the same time line for example if a character meets their future self on their personal time line this event will occur twice, once when then are the younger version of themselves and once when they are

the older version. Entities having their own time line allows us to see changes over time.

*Time travel was not the only occasion when characters have been known to meet themselves. The idea of multiple universes or realities is a staple in science fiction and fantasy. The `ontomedia:Context` class, a subclass of `ontomedia:Abstract Item`, was created to separate the many different versions of the same entity that may exist. It allows us to differentiate both between different representations of the same fictional character, different version of the same character and between real people, fictionalisations of real people and fictional characters. This is a particular issue when considering the contents of fiction, especially when those works have been reinterpreted across media, within the same work or after a period of time. Since these different interpretations can be physically distinctive, for example when a character is portrayed by different actors, or given different personality traits or back history it becomes necessary to recognise that while they may be supposed to be to same entity there are occasions when there differences are as important as their similarities. Examples of this can be seen in almost every movie adaptation of a book. For example the recent *Lord of the Rings* movie or the transformation of ‘*We Can Remember It For You Wholesale*’ to the movie ‘*Total Recall*’. In the first case the character of Faramir as portrayed in the movie by David Wenham was both physically and emotionally different to the character described in the book much to the disappointment of many fans. In the second the hero of the short story ‘*Douglas Quail*’ becomes ‘*Douglas Quaid*’. An even more extreme case can be seen in the new *Battlestar Galactica* mini-series in which two of the characters (*Starbuck* and *Boomer*) have changed gender since the original series aired. Interaction with users showed that they frequently distinguished between these different representations as well as creating their own. Since the separation was important to them it was also important to be able to model that distinction and allow*

the meta data to reflect it for the purposes of reducing ambiguity in search and retrieval.

*With the aim of modelling the contents of fiction, whatever media format it is presented in, we faced the potentiality of having to deal with anything that the human mind can come up with. The only way to deal with this lack of limitations is to plan for it. By making the ontology modular and expendable we leave the option open for those situations that arise which we had not envisioned. Beyond acknowledging that we are not going to be able to cover every situation, supporting extensibility allows us to reuse existing ontologies where they already exist. For example we extended the location ontology created for the *Signage* project [7] to provide a basic spacial model. We chose this ontology because it had a level of detail that matched what we envisioned for the rest of *OntoMedia*. This choice worked well in the examples that we created. However should the need have arisen for a different type of spacial representation then a different ontology could have been used instead.*

(Lawrence, 2005)

In short, *OntoMedia* has all the right credentials, it covers fact and fiction to the point of being time-travel compliant and more, and has been modelled with involvement from users. This seems the perfect fit: we shall start the implementation phase of this project with *OntoMedia*. But before we do, there is one thing outstanding and an area to demarcate off.

Crowdsourcing the Narrative and Audience Interpretation

While *OntoMedia* was modelled by a group working heavily with a mature community of fiction ‘users’, it still represents a top-down approach. While extensible and modular – properties facilitating change – *OntoMedia* does not however embody adaptability to some notion of emergent consensus from the end-users nor personalisation to each user’s own perception. If the story is in the eye of the beholder, and there are many eyes, *OntoMedia* cannot be the final answer here. On enquiry, it transpires two researchers – Lawrence and

Jewell – behind OntoMedia have proposed work in this direction, but have not had this accepted to date. The key passage in their proposal reads “to be of the most use an annotation system for describing narrative events should be capable of recording not only the multiple interpretations, but also where the ambiguity exists since the presence of the ambiguity itself may indicate a point of interest to researchers. While division of interpretation may highlight an ambiguous portion of the media, and thus a point for discussion of the narrative, it may also highlight the differences in user response. Where this information be recorded, it can offer valuable information about the way in which material is received and whether it is received in the same way by the different subsections of the audience.”

Reasoning with narrative

As a final stage to this review, the opportunity should be taken to draw a line between the objective we have developed – describing the story potential of the media – and thoughts of facilitating a powerful journey through content that go beyond that first and foremost objective. The discourse around topics such as reasoning with narrative is fascinating and the possibilities remarkable. But for the perspective of this report, the priority has to be a compelling realisation of some kind of rich-media wikipedia experience.

Summary

We’re going to build a language to describe the story potential of media. It will use semantic web technologies as practised at the BBC, so we’ll be representing the programmes in RDF, talking to it with SPARQL, and have all the mod-cons such as OWL reasoning. In practice this means the task is to facilitate semantic annotation of the narrative components of media and the relationships between those components.

Semantics are necessary, but to what rigour and reach is something to be tested: they are there to serve the purpose of linking up a story-world, not an end in their

own right. We also have to acknowledge the essential plurality of meaning given the audience is diverse and we are serving them: it is their perception that we need to work with.

The obvious next step is try to start modelling using the OntoMedia ontology, keeping in mind the ‘minimal, extensible’ requirements to allow institutional uptake, and the issue of interpretation.

Ontology development

This chapter summarises the development of what was to become the ‘BBC Stories’ ontology, a model designed for the BBC to use to unlock the stories that can be told through their programming on the medium of the Web.

‘BBC Stories’ is the joint effort of

- Paul Rissen. Information Architect at the BBC
- Michael Jewell, PhD. Co-author of OntoMedia
- Toby Harris, the author of this report

and took place in the context of a short form academic placement project at the BBC, April–August 2010. The development work has been guided by the evolving needs of the BBC during that time, with notable engagement and contributions from

- Yves Raimond
- Jonathan Tweed
- Andrew Dudfield
- Frances McNamara

OntoMedia 2

The result of the first major design cycle was OntoMedia2. There were two main aims

- Reform the OntoMedia core, with a view to an absolute minimum set of classes and rigour about what can be asserted
- Update OntoMedia data to be more interoperable with linked data principles by incorporating ontologies already in use at the BBC and that have become commonly used elsewhere since OntoMedia was released.

The core innovation of the transformation to OntoMedia2 is in its attempt to separate out semantic concerns from structural concerns. In OntoMedia, sub-classing is heavily used to add specific meaning to the core classes. For instance, as well as ‘Event’ there is ‘Travel Event’, which introduces its own properties of ‘from’ and ‘to’. It was viewed that this added complexity to the model, and such complexity would likely increase as more sub-classes were created by users unsatisfied with the

current definitions. OntoMedia2 provides a single ‘Event’ class, but provides a facility for arbitrary ‘facts’ about the world to be asserted by that event. So unlike in OntoMedia where programme annotations could develop an increasingly unique language of sub-classes, programme annotations made using OntoMedia2 would share a simple and consistent structure. The representation of whatever further detail is deemed needed is then made not as defining the event to have a specific meaning and providing a set of properties to detail that, but as set of self-contained facts. We are effectively creating a two tier representation, a higher-level one of coordination between events and their occurrences, that are tagged with what was involved, and a lower-level state-machine of precisely defined semantic statements.

The resultant annotations can be characterised as changing from ‘A travel event, involving character x, going from here to there’ to ‘This structural element called an event retracts fact character x is here and asserts fact character x is there’.

This is certainly less natural-language-like, but this is not a concern as fundamentally RDF is not natural language and regardless the actual form of the RDF will be abstracted from the users through tools developed for the annotation process. It also should mean less mental overhead when annotating as rather than having to have command of a broad number of concepts and their implementation, there is just the core concept of event, to which single statements of fact can be added, and any natural-language like grammar for these facts is universal and simple, essentially the subject-predicate-object of RDF. ‘The Doctor is in the Tardis’ -> “whoc:doctor om:locatedin whol:tardis”

Along similar lines, OntoMedia2 also attempts to separate interpretation from events. Conceived as having no interpretative value, the events are the bare references used for establishing the basic network structure, and with that established we can then bring in what the media provides for in its interpretation of that

event. A script might provide a structure of scenes and acts, or a DVD might provide chapters, and so in OntoMedia2 we can say a media region involves certain events, and from this patterns of narrative structure should emerge. The script also has a text representation of that event, and so just as we use occurrences to say this event has an occurrence on the timeline of the script, we can add that text representation to the occurrence. This should prove a useful (and semantically valid) feature in building story worlds that can exist without the long-form video they ultimately represent, and shows how clips of that programme can be linked to the OntoMedia2 data.

A push for having as rich as possible human-readable descriptions of all story world objects also formed a development for OntoMedia2, which as well as serving the user interfaces that the BBC will be using this data for, also frees us from attempting to capture all meaning in the text/media as RDF.

OntoMedia2 uses named graphs to group RDF statements, such that assertions of facts can be built up upon other assertions. This requires support for named graphs in the RDF tech stack being used.

OntoMedia2 – N3 annotation of Doctor Who episode

MEDIA

```

:DW_S4E1_ScriptDoc a foaf:Document.
#FIXME: foaf:primaryTopic po:pid "b009w049";
#FIXME: om:source <http://www.thewriterstale.com/pdfs/Doctor%20Who%204%20Ep.1%20-%20Shooting%20Script%20-%20Yellow%20-%2016.10.07.pdf>

:DW_S4E1_Script a om:MediaRegion;
om:content :DW_S4E1_ScriptDoc;
om:sub_region :scene01;
om:sub_region :scene02;
om:sub_region :scene03;
om:sub_region :scene04;
om:sub_region :scene05;
om:sub_region :scene06;
om:sub_region :scene07;
om:sub_region :scene08;
om:sub_region :scene09;
om:sub_region :scene10;
om:sub_region :scene11;
om:sub_region :scene12;
#FIXME [...]
om:sub_region :scene121;
om:sub_region :scene122;
om:sub_region :scene123;
om:sub_region :scene124;
om:sub_region :scene125;
om:sub_region :scene126;
om:sub_region :scene127;
om:sub_region :scene128.

#FIXME: dvd here as foaf:document

#FIXME: dvdvideo here with sub_regions as per dvd chapters. requires
finding out the timecode of the chapters etc.

#FIXME: dvd cast list linking actors to characters

#FIXME: something declaring equivalence of media (not just primaryTopic) of
dvd to broadcast episode

# TIMELINES

:day1 a om:timeline.
:night1 a om:timeline.
:day2 a om:timeline.
:night2 a om:timeline.

#FIXME: need to make night1 follow day1
#FIXME: need to place in whoniverse timeline, ie. 2008

# EVENTS FROM SCRIPT SCENES

# note - events and occurrences should auto-name themselves.
# ie. you could autogenerate
# donna_noble_isnt_in_noble_house_event
# script_occurrence01_on_day1

# note - i've left off all the "a type" triples 'cos so far they're all
inferred from the properties used and it makes it much more concise. This
may or may not be a mistake, but if creating this from a tool, no doubt
you'd be explicit.

:scene02 om:event :event01.

:donna_in_house { whoc:donna_noble om:location whol:noble_house. }

:occurrence01 timeline:timeline :day1;
  rdfs:comment "DONNA steps out of her front door. Smart, head held high;
she's on a mission."

:event01 om:retracts :donna_in_house;
  event:time :occurrence01.

:scene03 om:event :event02.

:doctor_in_tardis { whoc:doctor om:location whol:tardis. }

:occurrence02 timeline:timeline :day1.
  rdfs:comment "THE DOCTOR steps out of the TARDIS. Sets off. On a
mission."

:event02 om:retracts :doctor_in_tardis;
  event:time :occurrence02.

:scene06 om:event :event03.
:scene08 om:event :event03.

:donna_in_foyer { whoc:donna_noble om:location :adipose_industries_foyer. }

:occurrence03 timeline:timeline :day1;
  rdfs:comment "DONNA stops in the street, looks up... A TOWER BLOCK
looming above. Cool, sleek, stylish, the London HQ of Adipose Industries.
Deep breath, Donna heads towards it. DONNA walks through the revolving
doors.";
  om:follows :occurrence01.

:event03 om:asserts :donna_in_foyer;
  event:time :occurrence03.

```

```

:scene07 om:event :event04.
:scene09 om:event :event04.

:doctor_in_basement { whoc:doctor
om:location :adipose_industries_basement. }

:occurrence04 timeline:timeline :day1;
  rdfs:comment "THE DOCTOR stops in the street, looks up... THE TOWER
BLOCK looming above, Adipose Industries. But this is the opposite side to
Donna's, the back. Deep breath, the Doctor heads towards it. THE DOCTOR'S
down a flight of steps, sonicking a basement door - PRAC EXPLOSION on the
lock, and he slips inside.";
  om:follows :occurrence02.

:event04 om:asserts :doctor_in_basement;
  event:factor whoi:sonic_screwdriver;
  event:time :occurrence04.

:scene10 om:event :event05.

:donna_is_health_and_safety { :health_and_safety foaf:member
whoc:donna_noble. }
:guard_believes_donna_is_health_and_safety { :security_guard1
om:believes :donna_is_health_and_safety. }
:donna_in_cinema { whoc:donna_noble
om:location :adipose_industries_cinema. }

:occurrence05 timeline:timeline :day1;
  rdfs:comment "Posh foyer. DONNA shows her ID pass to the SECURITY
GUARD. DONNA Donna Noble, Health and Safety.";
  om:follows :occurrence03.

:event05 om:retracts :donna_in_foyer;
  om:asserts :guard_believes_donna_is_health_and_safety;
  om:asserts :donna_in_cinema;
  event:factor :id_pass;
  event:time :occurrence05.

:scene11 om:event :event06.

:doctor_is_health_and_safety { :health_and_safety foaf:member
whoc:doctor. :whoc:doctor foaf:name "John Smith". }
:guard_believes_doctor_is_health_and_safety { :security_guard2
om:believes :doctor_is_health_and_safety. }

:occurrence06 timeline:timeline :day1;
  rdfs:comment "'Backstage' corridor, all concrete and pipes. THE DOCTOR
passes a SECURITY GUARD, shows the psychic paper. THE DOCTOR John Smith,
Health and Safety.";
  om:follows :occurrence06.

:event06 om:asserts :guard_believes_doctor_is_health_and_safety;
  event:factor whoi:psychic_paper;
  event:time :occurrence06.

:scene13 om:event :event07;
  om:event :event08;
  om:event :event09;
  om:event :event10;
  om:event :event11;
  om:event :event12.

:event_presentation om:asserts { :adiposeindustries_staff
foaf:make :pill. };
  om:asserts :pill_reduces_fat.

:donna_is_a_journalist { :journalists foaf:member whoc:donna_noble. }
:journalists_at_presentation {
  :journalists om:location adipose_industries_cinema.
  :miss_foster om:location adipose_industries_cinema.
}

:occurrence07 timeline :day1;
  om:follows :occurrence05;
  rdfs:comment "Part of the Tower Block HQ, with Adipose Industries logos
on the walls. Slogan: The Fat Just Walks Away. The logo is on screen, as
MISS FOSTER steps forward, at the front. She's 40s, handsome, strong. She
addresses the audience, 40 PEOPLE or so, scattered about, taking notes -
they're JOURNALISTS; this is a Press Launch."

:event07 om:asserts :journalists_watch_miss_foster_presentation;
  event:time :occurrence07.

:journalists_believes_pill_reduces_fat {
  :journalists om:believes :pill_reduces_fat.
}

:occurrence08 timeline:timeline :day1;
  rdfs:comment "MISS FOSTER Adipose Industries. The twenty-first century
way to lose weight. No exercise, no diet, no pain. Just lifelong freedom,
from fat, the Holy Grail of the modern age. And here it is! Holds it up, an
ordinary red & white capsule. MISS FOSTER (CONT'D) You just take one
capsule. One capsule, once a day, for three weeks. And the fat, as they
say... ON SCREEN, GRAPHIC, the logo does a little spin, the jingle sings:
The Fat Just Walks Awwaaaaay!";
  om:follows :occurrence07.

:event08 om:asserts { :miss_foster
om:wants :journalists_believes_pill_reduces_fat};
  event:agent :miss_foster;
  event:factor :event_presentation; # Note what we did here -
presentation is an event, but we're not using it in the occurrence sense.
  event:time :occurrence08.

```

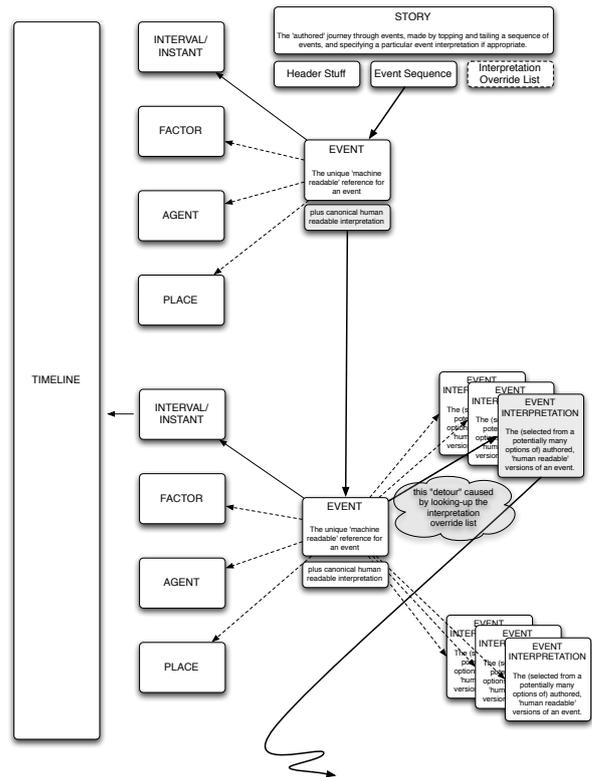
BBC Stories

The result of the second major design cycle was 'BBC Stories'. While a development from our interim OntoMedia2 ontology, it was sufficiently far removed from the official OntoMedia ontology in scope and operation that it would be a disservice to both to present it as a successor. Hence 'BBC Stories' was born.

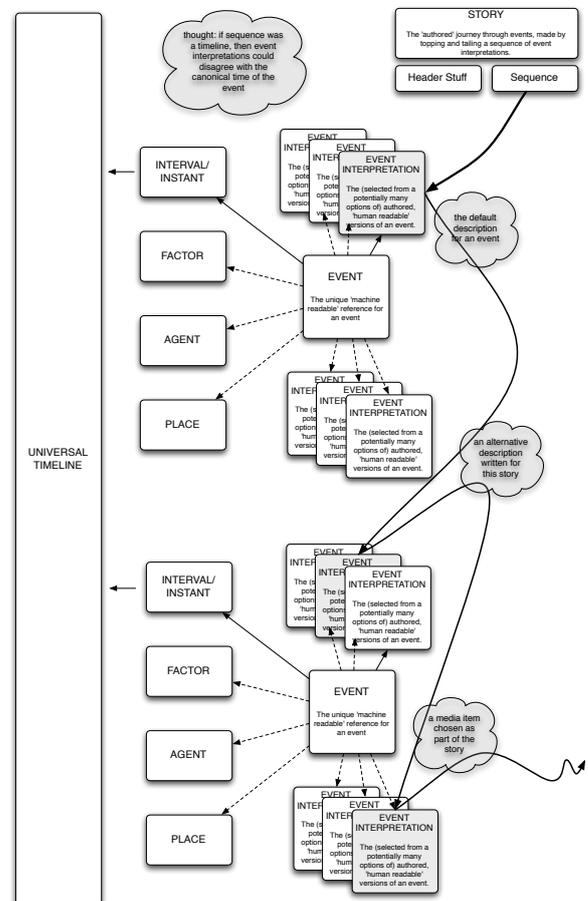
In OntoMedia core, there is no conceptualisation of a story. There are timelines, onto which occurrences of events are placed, and so a timeline can be considered as having the potential to tell a story. The timeline is actually a sequence of occurrences, so non-linear-in-time stories are possible. So a sequence of events can be reasoned, but you would have to know what other timelines were relevant to that story to find out things you may want to know about the events, such as their occurrence on the universal timeline, ie. when they actually happened. The point we are getting to is that

OntoMedia can create a rich tapestry of a story world, but from the point of view of the BBC, it misses the point completely: the primary aim is to editorialise a story world, so a core feature of the ontology has to be something that represents what the BBC might call a story, something that holds the kinds of things they would like to put on the webpage representing a story called "Doctor Who Victory of the Daleks", "The Battle of Britain", or "Winston Churchill". The need to provide such a simple, minimal core to the BBC became increasingly apparent for the BBC to be able to introduce this work. The impression gave was that, at some level, it had to appear as something akin to the next step on for tagging; there has to be the simple representation of what they want: we're going to define a story, its going to have these events in it, and these events link out into the shared world of people, places (and so on) that we are starting to build.

The other main innovation between OntoMedia2 and BBC Stories is that of an overt interpretation concept. Events no longer are structured through occurrences-with-added-interpretive-value on timelines, and on the



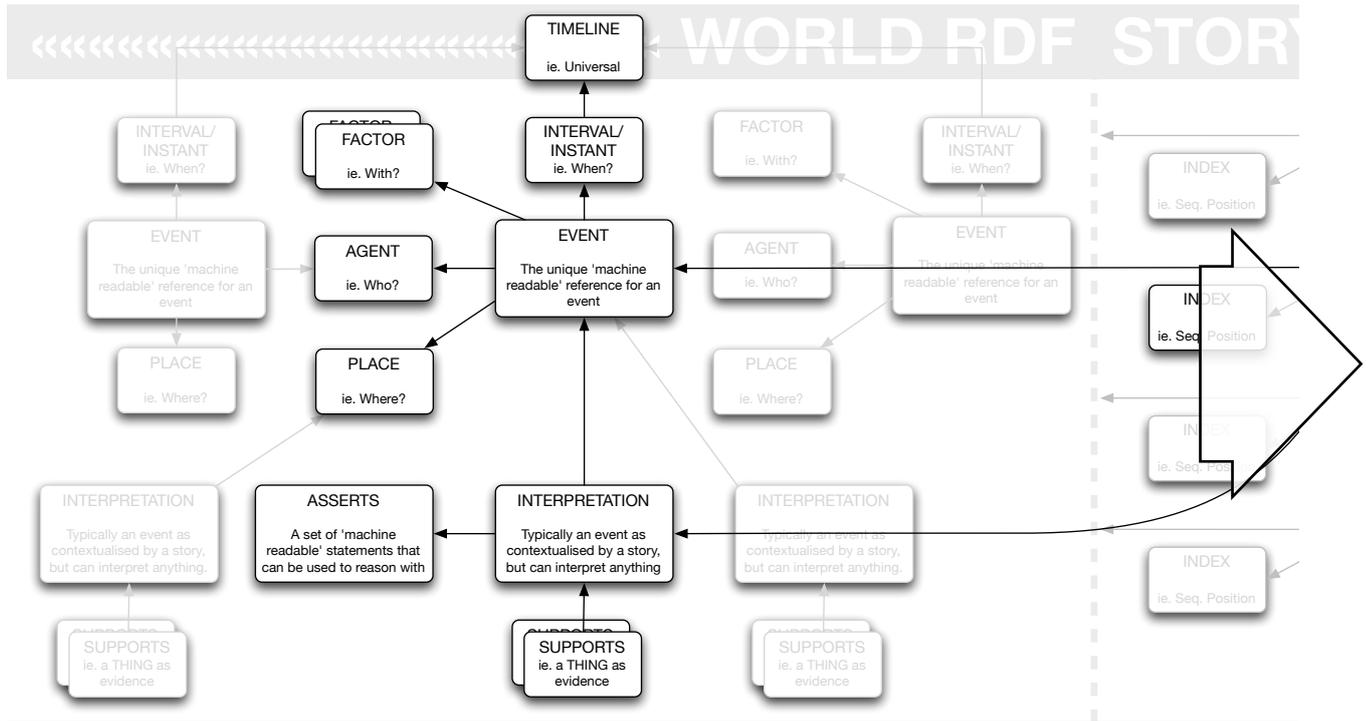
Debate at the time: is a 'story' a list of events...



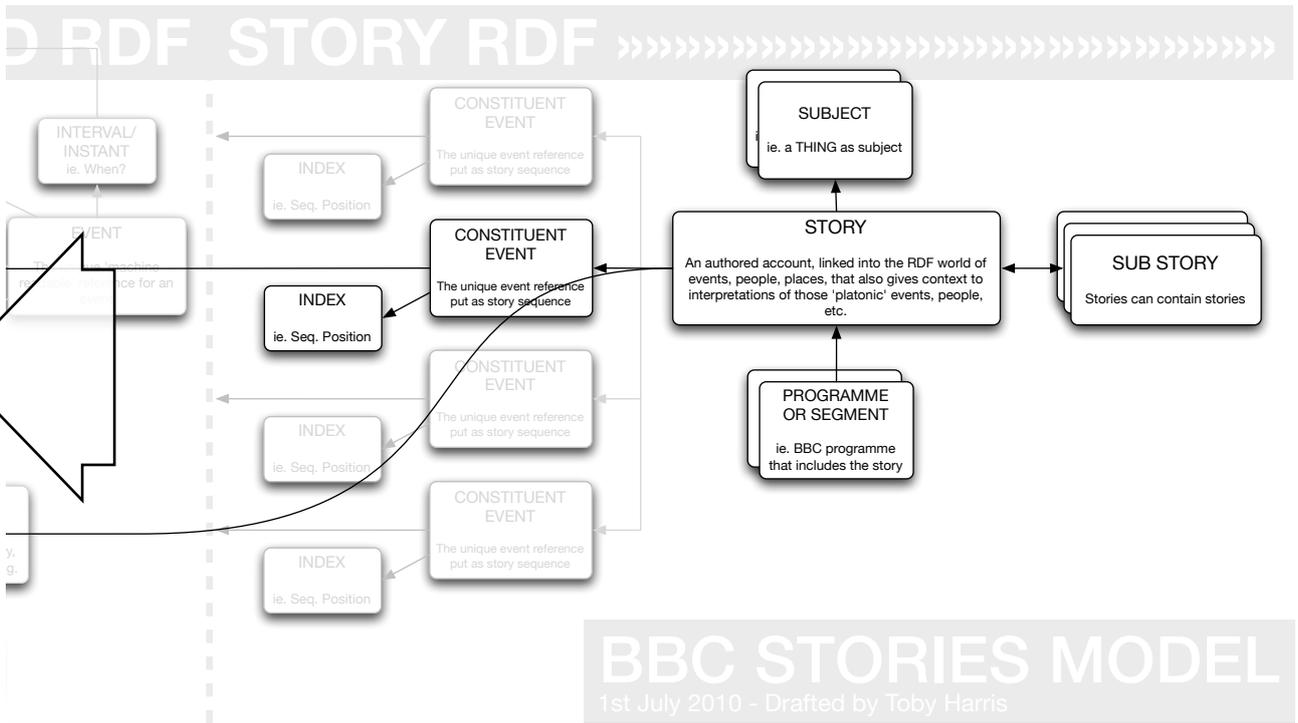
editorial side are editorially grouped into stories and on the 'world fact' side use the built-in time mechanism of the adopted Event ontology (already in use at the BBC, and originated at Queen Mary, University of London by Yves Raimond – <http://purl.org/NET/c4dm/event>). Anything that interprets this event, be it disputes a fact, or represents it in some media form, is now represented by an interpretation. In fact, interpretations can interpret anything, and are our unified way of separating out rich interpretive information from the simplistic network of events that are linked by people, places etc. in common. An interpretation can have things as supporting evidence, eg historians can cite the documents that informed the interpretative text written. Interpretations can have a direct source, at which point they are the media representation of whatever they are interpreting, eg this photograph interprets the event x. Interpretations, not events, now carry the assert and retract set of facts, so one set of interpretations could assert facts taken as canon, while any number of others could accommodate differing opinions and detail alternative constructions for the cause and event within that scene.

The ontology has been stable since late July, and while work continues, it should not effect any modelling performed with the current model. As a potential – and exciting – addition to the core, Michael Jewell working on being able to assign interpretations to people, and then perform queries from, and comparing, the perspectives of these people. Using the interpretation asserts/retracts facts mechanism, work continues on the absolute minimal way to express some kind of cause-and-effect through the story world – currently pruned to 'believes', 'wants', and 'facilitates' – and of finding ways to create further relations between programmes that haven't been considered due to the 'story world' focus of this project. For instance, an interpretation could assert that the set of this programme is really shaky, or it is a high-profile, big-budget documentary, and the viewer be more interested in following that kind of contextual relationship rather than one of shared topic.

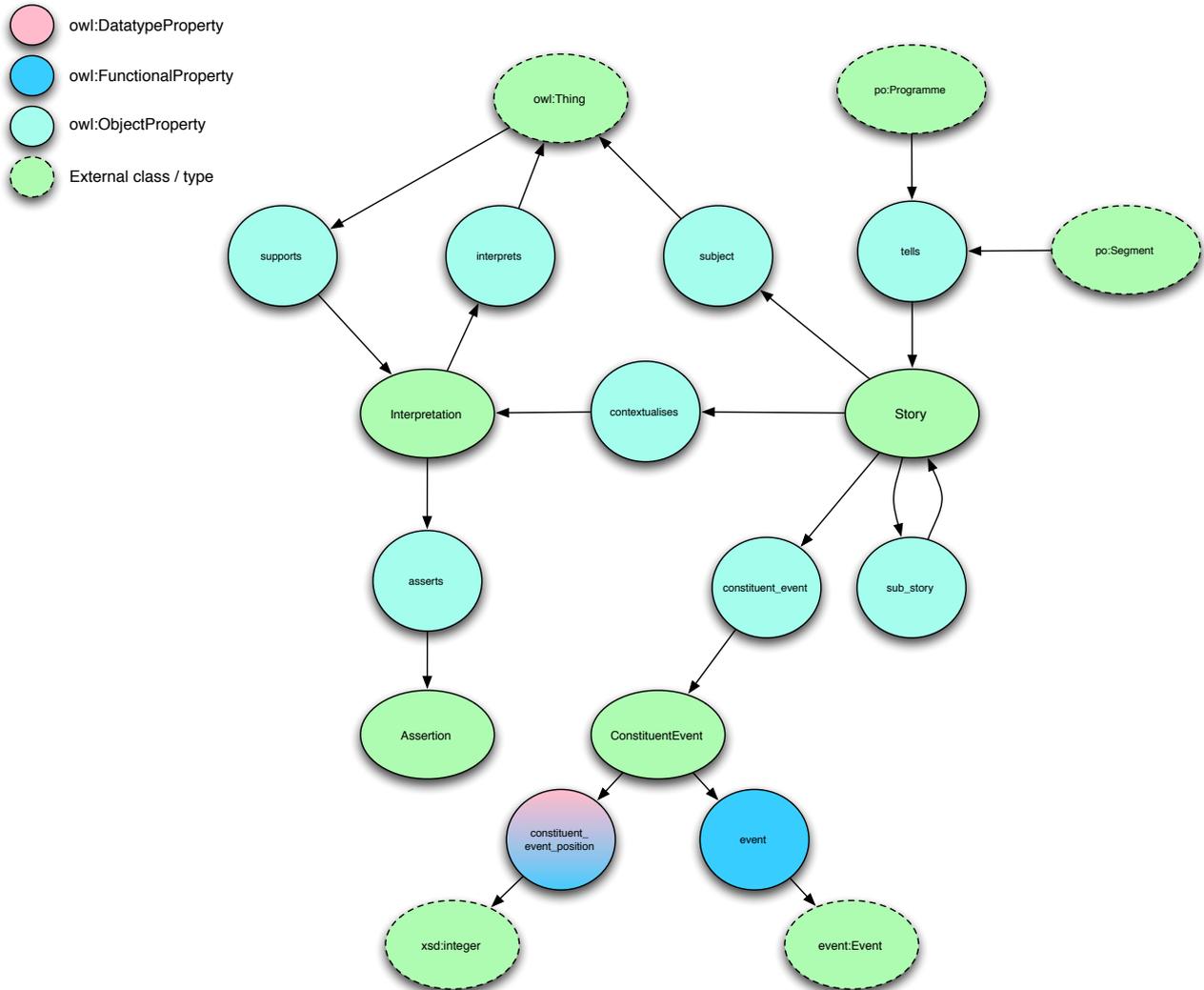
The ontology is easiest to understand when shown as a model, see following page, and the actual ontology is shown in diagram and OWL form on the following pages to that.



The world in RDF: events, people, places, factors, dates and times, and the interpretations, evidence and assertions that will build up around them...



...and the stories in RDF that editorialise that world, plotting journeys through it and contextualising the interpretations.



BBC Stories Ontology, as of July 2010. Diagram by Michael Jewell. Note Ordered List Ontology is not yet in use.

BBC Stories – The ontology

.....

This is the OWL (web ontology language) document for

BBC Stories, as of 4th August 2010. Future updates should be available at <http://contextus.net>

```
<?xml version="1.0"?
```

```
<!DOCTYPE owl [
  <!ENTITY xsd      "http://www.w3.org/2001/XMLSchema#">
  <!ENTITY rdf      "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!ENTITY rdfs     "http://www.w3.org/2000/01/rdf-schema#">
  <!ENTITY dc       "http://purl.org/dc/elements/1.1/">
  <!ENTITY dct      "http://purl.org/dc/terms/">
  <!ENTITY owl    "http://www.w3.org/2002/07/owl#">
  <!ENTITY base     "http://contextus.net/ontology/ontomedia#">
  <!ENTITY foaf     "http://xmlns.com/foaf/0.1/">
  <!ENTITY event    "http://purl.org/NET/c4dm/event.owl#">
  <!ENTITY timeline "http://purl.org/NET/c4dm/timeline.owl#">
  <!ENTITY geonames "http://www.geonames.org/ontology#">
]>
```

```
<rdf:RDF xmlns:rdf="&rdf;"
  xmlns:rdfs="&rdfs;"
  xmlns:dc="&dc;"
  xmlns:dct="&dct;"
  xmlns:owl="&owl;"
  xmlns:xsd="&xsd;"
  xmlns:foaf="&foaf;"
  xmlns:event="&event;"
  xmlns:timeline="&timeline;"
  xmlns:geonames="&geonames;"
  xml:base="&base;">
```

```
<owl:Ontology rdf:about="&base;">
  <rdfs:label>OntoMedia</rdfs:label>
  <dc:title xml:lang="en">OntoMedia</dc:title>
  <dct:created>2010-05-11</dct:created>
  <dct:modified>2010-05-18</dct:modified>
  <owl:versionInfo>0.02 2010/05/18 15:36:00 mjewell</owl:versionInfo>
  <dc:creator>Michael O. Jewell</dc:creator>
  <dc:creator>K Faith Lawrence</dc:creator>
  <dc:creator>Mischa M Tuffield</dc:creator>
  <dc:creator>Hugo Mills</dc:creator>
  <dc:creator>Paul Rissen</dc:creator>
  <dc:creator>Toby Harris</dc:creator>
</owl:Ontology>
```

```
<!-- ===== -->
<!-- Core: Timeline, Occurrence, Event, Assertion -->
<!-- ===== -->
```

```
<owl:Class rdf:ID="Timeline">
  <rdfs:subClassOf rdf:resource="&timeline;Timeline" />
</owl:Class>
```

```
<!-- Occurrence -->
```

```
<owl:Class rdf:ID="Occurrence">
  <rdfs:subClassOf rdf:resource="&timeline;Interval" />
  <rdfs:subClassOf rdf:resource="&timeline;AbstractInstant" />
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="precedes">
  <rdfs:label>precedes</rdfs:label>
  <rdfs:comment>This property defines the occurrence which immediately follows this occurrence</rdfs:comment>
  <rdfs:domain rdf:resource="&Occurrence" />
  <rdfs:range rdf:resource="&Occurrence" />
  <owl:inverseOf rdf:resource="&follows" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="follows">
  <rdfs:label>follows</rdfs:label>
  <rdfs:comment>This property defines the occurrence which comes immediately prior to this occurrence</rdfs:comment>
  <rdfs:domain rdf:resource="&Occurrence" />
  <rdfs:range rdf:resource="&Occurrence" />
  <owl:inverseOf rdf:resource="&precedes" />
</owl:ObjectProperty>
```

```
<!-- Event -->
```

```
<owl:Class rdf:ID="Event">
  <rdfs:subClassOf rdf:resource="&event;Event" />
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="asserts">
  <rdfs:label>suggests</rdfs:label>
  <rdfs:comment>A reference to an assertion made by this event.</rdfs:comment>
  <rdfs:domain rdf:resource="&event;Event"/>
  <rdfs:range rdf:resource="&Assertion" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="retracts">
  <rdfs:label>retracts</rdfs:label>
  <rdfs:comment>A reference to an assertion that this event retracts.</rdfs:comment>
  <rdfs:domain rdf:resource="&event;Event"/>
  <rdfs:range rdf:resource="&Assertion" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="causes">
```

```
<rdfs:label>causes</rdfs:label>
<rdfs:comment>An event that this event causes.</rdfs:comment>
<rdfs:domain rdf:resource="&event;Event"/>
<rdfs:range rdf:resource="&event;Event" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="requires">
  <rdfs:label>causes</rdfs:label>
  <rdfs:comment>An assertion required for this event to occur.</rdfs:comment>
  <rdfs:domain rdf:resource="&event;Event"/>
  <rdfs:range rdf:resource="&Assertion" />
</owl:ObjectProperty>
```

```
<!-- Assertion -->
```

```
<owl:Class rdf:ID="Assertion">
  <rdfs:label>Assertion</rdfs:label>
  <rdfs:comment>An assertion of some facts.</rdfs:comment>
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="facts">
  <rdfs:label>facts</rdfs:label>
  <rdfs:comment>An RDF graph containing facts in an assertion.</rdfs:comment>
  <rdfs:domain rdf:resource="&Assertion"/>
</owl:ObjectProperty>
```

```
<!-- ===== -->
<!-- Extensions to owl:Thing -->
<!-- ===== -->
```

```
<owl:ObjectProperty rdf:ID="represents">
  <rdfs:label>portrays</rdfs:label>
  <rdfs:comment>Something this individual represents.</rdfs:comment>
  <owl:inverseOf rdf:resource="&represented_by" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="represented_by">
  <rdfs:label>portrayed_by</rdfs:label>
  <rdfs:comment>Something represented by this individual.</rdfs:comment>
  <owl:inverseOf rdf:resource="&represents" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="location">
  <rdfs:label>location</rdfs:label>
  <rdfs:range rdf:resource="&geonames;Feature"/>
  <rdfs:comment>The location of an individual</rdfs:comment>
</owl:ObjectProperty>
```

```
<!-- ===== -->
<!-- Beings and Characters -->
<!-- ===== -->
```

```
<owl:ObjectProperty rdf:ID="has">
  <rdfs:label>has</rdfs:label>
  <rdfs:comment>A thing owned or possessed by this being.</rdfs:comment>
  <rdfs:domain rdf:resource="&foaf;Agent"/>
  <rdfs:range rdf:resource="&owl;Thing" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="believes">
  <rdfs:label>believes</rdfs:label>
  <rdfs:comment>A reference to a named graph of statements that this being believes.</rdfs:comment>
  <rdfs:domain rdf:resource="&foaf;Agent"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="wants">
  <rdfs:label>wants</rdfs:label>
  <rdfs:comment>A reference to an Event which the being wishes to occur.</rdfs:comment>
  <rdfs:domain rdf:resource="&foaf;Agent"/>
  <rdfs:range rdf:resource="&event;Event" />
</owl:ObjectProperty>
```

```
<!-- Being -->
```

```
<owl:Class rdf:ID="Being">
  <rdfs:subClassOf rdf:resource="&foaf;Agent" />
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="voices">
  <rdfs:label>voices</rdfs:label>
  <rdfs:comment>A character this being voices.</rdfs:comment>
  <rdfs:domain rdf:resource="&Being"/>
  <rdfs:range rdf:resource="&Character"/>
  <owl:inverseOf rdf:resource="&voiced_by" />
  <rdfs:subPropertyOf rdf:resource="&represents" />
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:ID="portrays">
  <rdfs:label>voices</rdfs:label>
  <rdfs:comment>A character this being portrays.</rdfs:comment>
  <rdfs:domain rdf:resource="&Being"/>
  <rdfs:range rdf:resource="&Character"/>
  <owl:inverseOf rdf:resource="&portrayed_by" />
  <rdfs:subPropertyOf rdf:resource="&represents" />
</owl:ObjectProperty>
```

```
<!-- Character -->
```

```
<owl:Class rdf:ID="Character">
  <rdfs:subClassOf rdf:resource="&Being" />
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="voiced_by">
  <rdfs:label>voices</rdfs:label>
  <rdfs:comment>A being that voices this character.</rdfs:comment>
  <rdfs:domain rdf:resource="&Character"/>
  <rdfs:range rdf:resource="&Being"/>
```

```

    <owl:inverseOf rdf:resource="#voices" />
    <rdfs:subPropertyOf rdf:resource="#portrayed_by" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="portrayed_by">
  <rdfs:label>voices</rdfs:label>
  <rdfs:comment>A being that depicts this character.</rdfs:comment>
  <rdfs:domain rdf:resource="#Character"/>
  <rdfs:range rdf:resource="#Being"/>
  <owl:inverseOf rdf:resource="#portrays" />
  <rdfs:subPropertyOf rdf:resource="#represents" />
</owl:ObjectProperty>

<!-- ===== -->
<!-- Space / Geography -->
<!-- ===== -->

<owl:Class rdf:ID="Space">
  <rdfs:subClassOf rdf:resource="#&geonames;Feature" />
</owl:Class>

<owl:ObjectProperty rdf:ID="owner">
  <rdfs:label>owner</rdfs:label>
  <rdfs:domain rdf:resource="#Space"/>
  <rdfs:range rdf:resource="#Being"/>
  <rdfs:comment>The owner of the space.</rdfs:comment>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="depicts">
  <rdfs:label>depicts</rdfs:label>
  <rdfs:domain rdf:resource="#Space"/>
  <rdfs:range rdf:resource="#Space"/>
  <rdfs:comment>A space depicted by this space (e.g. a filming
location).</rdfs:comment>
  <owl:inverseOf rdf:resource="#depicted_by" />
  <rdfs:subPropertyOf rdf:resource="#represents" />
</owl:ObjectProperty>

<!-- ===== -->
<!-- Items -->
<!-- ===== -->

<owl:Class rdf:ID="Item">
</owl:Class>

<owl:Class rdf:ID="PhysicalItem">
  <rdfs:subClassOf rdf:resource="#Item" />
</owl:Class>

<!-- ===== -->
<!-- Media -->
<!-- ===== -->

<!-- MediaRegion -->

<owl:Class rdf:ID="MediaRegion">
  <rdfs:label>Media Region</rdfs:label>
  <rdfs:comment>A piece of media.</rdfs:comment>
</owl:Class>

<owl:ObjectProperty rdf:ID="event">
  <rdfs:label>event</rdfs:label>
  <rdfs:comment>An event depicted by this region.</rdfs:comment>
  <rdfs:domain rdf:resource="#MediaRegion" />
  <rdfs:range rdf:resource="#Event" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="content">
  <rdfs:label>content</rdfs:label>
  <rdfs:comment>The content of this region.</rdfs:comment>
  <rdfs:domain rdf:resource="#MediaRegion" />
  <rdfs:range rdf:resource="#foaf;Document" />
</owl:ObjectProperty>

<!-- Collection -->

<owl:Class rdf:ID="Collection">
  <rdfs:comment>Represents a collection of things</rdfs:comment>
  <rdfs:label>Collection</rdfs:label>
</owl:Class>

<owl:ObjectProperty rdf:ID="number_of_parts">
  <rdfs:label>number_of_parts</rdfs:label>
  <rdfs:comment>This property specifies the number of parts contained by
the collection</rdfs:comment>
  <rdfs:domain rdf:resource="#base;Collection"/>
  <rdfs:range rdf:resource="#&rdfs;Literal"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="max_number_of_parts">
  <rdfs:label>max_number_of_parts</rdfs:label>
  <rdfs:comment>Specifies the maximum number of parts that can be
contained by the collection</rdfs:comment>
  <rdfs:domain rdf:resource="#Collection"/>
  <rdfs:range rdf:resource="#&rdfs;Literal"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="min_number_of_parts">
  <rdfs:label>min_number_of_parts</rdfs:label>
  <rdfs:comment>Specifies the minimum number of parts that can be
contained by the collection</rdfs:comment>
  <rdfs:domain rdf:resource="#Collection"/>
  <rdfs:range rdf:resource="#&rdfs;Literal"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="part_instance">
  <rdfs:label>part_instance</rdfs:label>
  <rdfs:comment>Specifies the template object which makes up the
components of this collection</rdfs:comment>
  <rdfs:domain rdf:resource="#Collection" />
  <rdfs:domain rdf:resource="#owl;Thing" />
</owl:ObjectProperty>

```

BBC Stories – N3 annotation of Doctor Who episode

.....

This shows a summary of how BBC Stories was looking when applied at the time the name was coined. There are some subtle differences between this and the final annotation in appendix 2.

```
### STORIES
# CANONICAL STORY
:canonicalStory a stories:story;
  stories:constituent_event [stories:constituent_event_position
"1"^^xsd:integer; stories:event event01.];
  stories:constituent_event [stories:constituent_event_position
"2"^^xsd:integer; stories:event event02.];
  stories:constituent_event [stories:constituent_event_position
"3"^^xsd:integer; stories:event event03.];
  stories:constituent_event [stories:constituent_event_position
"4"^^xsd:integer; stories:event event04.];
  # [etc....]
  stories:contextualises :programmeProduction;
  stories:contextualises :scriptScene01;
  stories:contextualises :scriptScene02;
  stories:contextualises :scriptScene03;
  stories:contextualises :scriptScene04;
  stories:contextualises :scriptScene05;
  # [etc....]
.

# MEDIA REPRESENTATIONS
:canonicalProgramme a programmes:programme;
  programmes:pid "b009w049";
  stories:tells :canonicalStory.

:cast {
  whoc:doctor portrayal:actor whop:david_tennant.
  whoc:donna_noble portrayal:actor whop:catherine_tate.
  # [etc.....]
}

:crew {
  # writer
  # [etc.....]
}

:programmeProduction a stories:interpretation;
  rdfs:label "Production info for the TV programme: writer, cast, etc.";
  stories:interprets canonicalProgramme;
  stories:asserts :cast;
  stories:asserts :crew;
  stories:source <http://www.imdb.com/title/tt115991/>.

:script a foaf:Document;
  foaf:primaryTopic :canonicalProgramme;
  foaf:page <http://www.thewriterstale.com/scr.html>;
  rdf:resource <http://www.thewriterstale.com/pdfs/Doctor%20Who%20Ep.
1%20-%20Shooting%20Script%20-%20Yellow%20-%2016.10.07.pdf>.
  stories:tells :canonicalStory;

### HIGH-LEVEL EVENTS
# Think DVD-Chapters

:adipose_industries_exposition a event:Event
  rdfs:label "The Doctor and Donna both investigate Adipose Industries";
  event:sub_event :building_entry;
  event:sub_event :presentation;
  event:sub_event :cubicles;
  event:sub_event :doctor_and_roger.

:parthenogenesis a event:Event
  rdfs:label "Donna visits Stacey, an Adipose Industries Customer.
Something dramatic happens to Stacey";
  event:sub_event :stacey_and_donna;
  event:sub_event :partho_activated;
  event:sub_event :stacey_partho;
  event:sub_event :donna_investigates.

[etc.....]
```

```
### PLOT-LEVEL EVENTS
# Think programme precis written out, and then make each sentence an
event.

# Scenes 2-12
:building_entry a event:Event;
  rdfs:label "Doctor and Donna independently enter the building";
  event:agent whoc:donna_noble;
  event:agent whoc:doctor;
  event:place whol:tardis;
  event:place whol:noble_house;
  event:place :adipose_industries_foyer;
  event:factor whoi:sonic_screwdriver;
  event:factor whoi:psychic_paper.

# Scene 13
:presentation a event:Event;
  rdfs:label "Journalists watch presentation at Adipose Industries";
  event:agent :journalists;
  event:agent :foster;
  event:agent whoc:donna_noble;
  event:agent whoc:doctor;
  event:place :adipose_industries_cinema;
  event:factor whoi:psychic_paper;
  event:factor :adipose_product.

# Scene 14
:cubicles a event:Event;
  rdfs:label "Doctor and Donna indepedently access Adipose Industries
customer information";
  event:agent :craig;
  event:agent :clair;
  event:agent :salespeople;
  event:agent whoc:donna_noble;
  event:agent whoc:doctor;
  event:place :adipose_industries_sales_cubicles;
  event:factor whoi:psychic_paper.

[etc.....]

### INTERPRETATIONS FROM SCRIPT
# Think a full marked-up transmogrification of programme's script into our
story world.
# Each interpretation is a scene, and breaks down the "this event has
agent, factor, location" with machine-reasonable precision.
# Note that there may be many script scenes for each 'plot level event',
or vice-versa, and that long scenes may require further breaking down to
model the plot development (ie. assertions and the script excerpt they
correspond to).

:scriptScene01 a stories:Interpretation;
  rdfs:label "Script - Scene 1";
  rdfs:comment "OMITTED";
  rdfs:resource <path/to/encodedscript/#scene1>;
  stories:source :script.

:donna_in_noble_house { whoc:donna_noble geonames:locatedIn
whol:noble_house. }

:scriptScene02 a stories:Interpretation;
  rdfs:label "Script - Scene 2";
  rdfs:comment "DONNA steps out of her front door. Smart, head held high;
she's on a mission.";
  rdfs:resource <path/to/encodedscript/#scene2>;
  stories:interprets :building_entry;
  stories:retracts :donna_in_noble_house;
  stories:source :script.

:doctor_in_tardis { whoc:doctor geonames:locatedIn whol:tardis. }

:scriptScene03 a stories:Interpretation;
  rdfs:label "Script - Scene 3";
  rdfs:comment "THE DOCTOR steps out of the TARDIS. Sets off. On a
mission.";
  rdfs:resource <path/to/encodedscript/#sceneX>;
  stories:interprets :building_entry;
  stories:retracts :doctor_in_tardis;
  stories:source :script.

:donna_noble_in_adipose_industries_foyer { whoc:donna_noble
geonames:locatedIn :adipose_industries_foyer. }
:donna_noble_is_health_and_safety { :health_and_safety foaf:member
whoc:donna_noble. }
:guard_believes_donna_noble_is_health_and_safety { :security_guard1
logic:believes :donna_is_health_and_safety. }
:id_pass_facilitates_guard_believes_donna_noble_is_health_and_safety
{ :id_pass
logic:facilitates :guard_believes_donna_noble_is_health_and_safety. }

# FIXME: Need to say something like...
#
guard_facilitates_donna_in_adipose_industries_because_member_of_adipose_sta
ff_permits_foafagent_in_adipose_industries
# ...expressed with the story logic ontology.
# staff *permits* entry? entry *requires* staff? latter more generic (idea
of dependencies) but not quite as natural english.

:scriptScene10 a stories:Interpretation;
  rdfs:label "Script - Scene 10";
  rdfs:comment "Posh foyer. DONNA shows her ID pass to the SECURITY GUARD.
DONNA: 'Donna Noble, Health and Safety.'";
  stories:interprets :building_entry;
  stories:asserts :donna_in_foyer;
  stories:asserts :donna_is_health_and_safety;
  stories:asserts :guard_believes_donna_is_health_and_safety;

stories:asserts
:id_pass_facilitates_guard_believes_donna_noble_is_health_and_safety;
stories:source :script.
```

Applying ‘BBC Stories’

Our ontology work is being applied in two arenas. There are live projects at the BBC that require a basic set of functionality, and a research strand exploring advanced functionality.

Choice of media

We had established Doctor Who would offer good test subjects to represent with our model. The Doctor Who production team also had the brief to develop the Doctor Who section of the BBC website to include more archive programming, and were keen to realise some of the benefits demonstrated by the Mythology Engine prototype.

An episode was therefore selected to act as our central test subject. While the public availability of any DW episode online could not be guaranteed, there was a season whose scripts were publicly available. We felt this would provide an interesting opportunity to work with multiple media representations of the same story, while also giving us more scope to deliver a powerful experience with textual descriptions alone. And so Doctor Who, Series 4 Episode 1: ‘Partners in Crime’ became the initial test subject for our work.

By July, with the BBC Stories ontology seemingly stable, the initial prospect of applying this work for Doctor Who had expanded into a pilot project based on the BBC Archive (good!).

A corpus of programmes were selected in a process at the BBC to represent the many domains of the BBC and have story or thematic potential taken as a whole. With this done, most of the implementation focus of ‘BBC Stories’ has been on facilitating this archive project. This has translated to a very high-level approach, annotating the programmes in a very broad way. While this could be characterised as a single, if important, step on from semantically backed tagging, it has brought interesting perspectives into our work, such as the opinion of archivists on the links they find interesting in

programming. Not to mention there is the chance of uptake and the pilot transforming into the live site!

This project, digital narratives, has kept the focus on the deeper, richer annotation possible with BBC Stories, with the aim of producing a complete transcription of the original DW episode as fully modelled as possible.

With these two in parallel and in sync, the core should be kept real, and the possible futures compatible.

Application

Creating transcriptions

We have been creating our transcriptions as plain text using N3 notation - “A readable language for data on the Web” (Berners-Lee, 1998). More precisely, we have been using TriG, which combines a subset of N3 with support for named graphs (Bizer & Cyganiak, 2004). Tools such as Protege were found more trouble than they were worth, providing you had a way to validate the document (ie. for grammatical mistakes, rather than what you’re actually asserting!).

As the ontology has developed, it has lent itself to different strategies for creating transcriptions. At the heart has always been heavy use of a text editor – N3 syntax highlighting highly advised – and in doing what can amount to significant data entry work ideas of how tools could ease the process come readily. With BBC Stories now largely stable, and a user-base looking likely in its use on the Archive project, developing such tools should be a first priority.

Using the data

For the development of ‘BBC Stories’ we have kept our experimental work separate from the deployed resources of the BBC, instead working with our own ‘sandbox’ environments.

To host and use the transcription data we have created, this project has modified a Mac OSX 10.6 installation with the following:

- ‘Raptor’ CLI utilities for verifying and manipulating between different forms of RDF serialisations. (“Raptor RDF Syntax Library”)
- ‘4Store’ triplestore and SPARQL endpoint for hosting and querying our BBC stories data. (“4store - Scalable RDF storage”)
- Mac OSX ‘Web Sharing’ featured enabled, turning on the built-in Apache web server software.
- Apache configuration to facilitate HTTP access to the triplestore from locally served webpages.

With the TriG text files imported into the triplestore, the data can be accessed using SPARQL queries using command line tools or via a HTTP SPARQL end-point.

For instance, to obtain all RDF statements with the chosen story as their subject, the query would be:

```
SELECT * WHERE
{
  <http://domain/story/thestory> ?p ?o .
}
```

This can be read as ‘select for me everything with “thestory” as subject, and return all the pairs of ?p predicate and ?o object.

A story in the BBC Stories ontology has an ordered list of events. Ordered lists conceptually combine the object to be listed with some form of ranking information (ie. before object x / after object y or absolute index), and then associate this information with the list object itself. Consider that as RDF is formed of statements made as subject, predicate, object triples, we cannot express an ordered list directly as we can only associate one single thing with another single thing in a single statement. We first need to create an object to represent the pair, associate this with the object to be listed and then associate it with the ranking information. Now we can associate all these pairing objects with the list object itself. When creating the data, this can be abstracted away somewhat is not a problem. When querying the raw RDF with SPARQL however, we will note that the above query doesn’t actually return us all the events of the

story. We are returned a ‘eventsList’ object, which if queried in turn will return a number of arbitrarily named objects holding the event object and a ranking object or literal.

SPARQL allows for more sophisticated searches, and we can create a single search to perform all the above for us and only return us what we want: all the first order properties of the story (ie everything that is story–predicate–object) and all the events in the event list with their position index.

```
SELECT ?p ?o ?index ?item WHERE
{
  <http://domain/story/thestory> ?p ?o .
  OPTIONAL
  {
    ?o <http://domain/ontology/bbcstories/slot> ?slot .
    ?slot <http://domain/ontology/list/index> ?index .
    ?slot <http://domain/ontology/bbcstories/item> ?
    item .
  }
}
```

This can be read as ‘select for me everything with “thestory” as subject, if any of those objects are then found to have objects associated by the concept “slot” then select them too, and if those slot objects have objects related by index or item concepts, select them too. Out of all of that, I’m return me the groupings of predicate, object, index and item.’

This is the tip of the iceberg as far as querying a full transcription of BBC Stories data is concerned, but is sufficient here to demonstrate the core functionality.

Visualising the data: Diagram

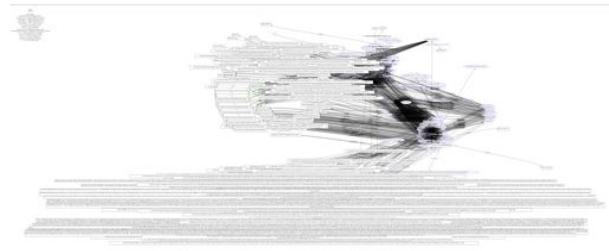
.....
 An attempt was made to visualise the full BBC Stories transcription of the test Doctor Who episode. It was hoped that given the strict structuring of the data, illuminating diagrams could be produced that could give some form of literal overview to the story structure, pacing and so on.

A workflow was found where RDF statements could be transformed into GraphViz ‘dot’ files using Raptor, which could then be imported into OmniGraffle either for automatic layout using various algorithms, or manual layout, or as envisioned, a combination of both.

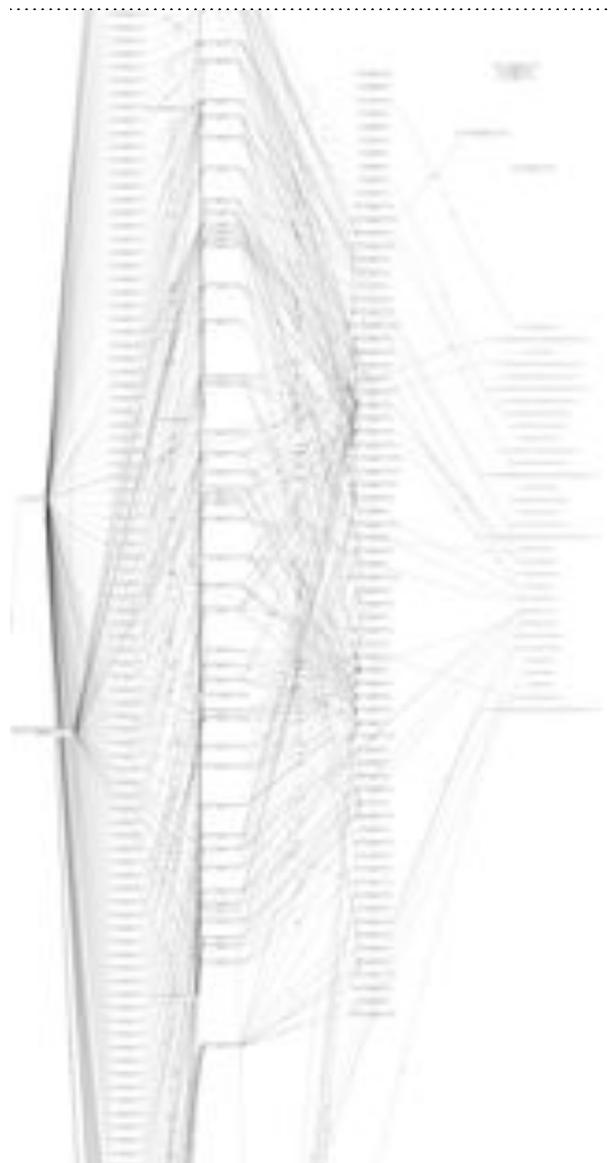
The result of this is shown to the right, which after many delays waiting for various layout options to compute, and a lot of time then trying to pick key nodes out and rearrange groups of nodes, was judged never going to work.

A dot file, on investigation, proved to be a simple format that states what nodes are connected to what else, and not a lot more. As one of the main issues with the initial diagram was that there was no control over what kind of objects or relationships were displayed, code was written to produce a DOT file with the desired characteristics. The key change was to take each of the main types of BBC Stories objects and roll certain properties into the labelling of the object, rather than have these as separate objects. This simplified the diagram considerably, and by adding rank information into the dot file, certain type properties could be inferred by their position on the diagram, so rather than all the event objects being linked to a ‘type:event’ object, they could simply line up in a row together. The amount of objects requiring display was still sufficiently large to pose problems to the layout algorithm, with the ranking information being largely – but not entirely – respected.

The more time that was spent on this, the more the impression grew that showing the completeness of the modelling and giving that illuminating overview were sufficiently contradictory the time would be better spent



DW S4E1 Visualisation using RDF->DOT file generator built-in to Raptor: after significant manual rearrangement the elements, its still a mess.



DW S4E1 Visualisation Excerpt using RDF->DOT file generator written specifically for BBC Stories data. Better, but prompting the question whether a flat 2D diagram is ever going to be as illuminating as hoped

on either producing some more selective interpretations diagrammed by hand – as all of the stunning ‘Information is Beautiful’ visualisations are, including the classic of the genre ‘Time travel in TV and Film’ (McCandless, 2009) – or moving on to the interactive visualisation intended for people to figuratively ‘walk around’ the story world and gain insight of the relationships from exploring within the visualisation, rather than as an overview.

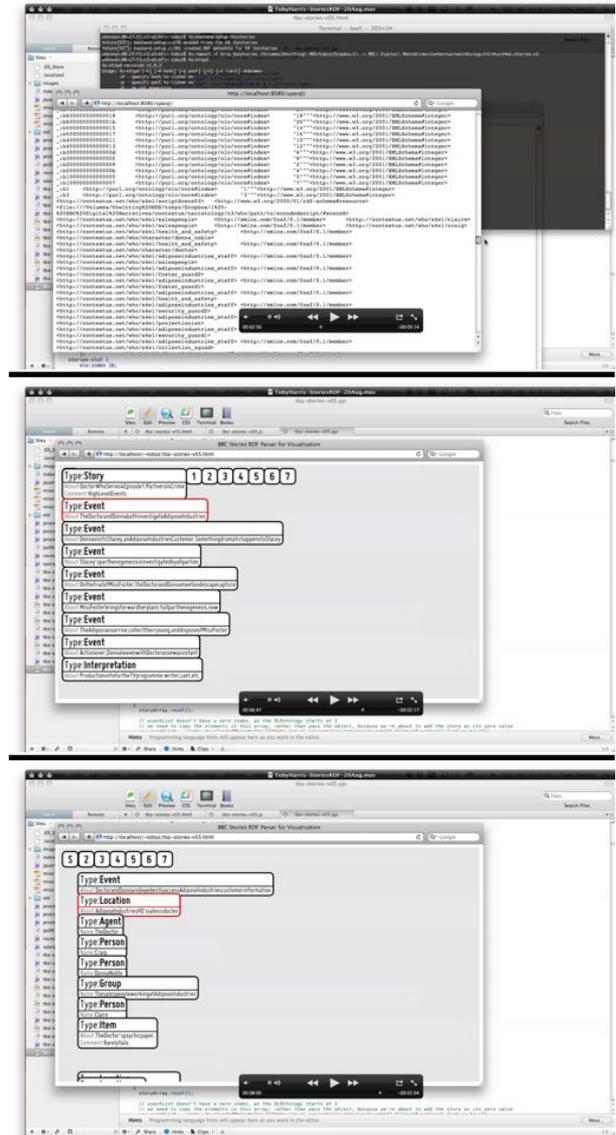
The code for the RDF->DOT converter functions is listed in appendix 6.

Visualising the data: Interactive

The foreseeable use of BBC Stories at the BBC would be to create webpages representing the objects of the story world – we could characterise this as a page for this character, and on it a number of links to other things relevant to this character. This is not all that could be done with the data however, and it was a goal of this project to explore the interactive, navigational possibilities of visualising this data. Rather than a series of webpages, what about a 3D world of dots and lines where you could experience that drama-as-network idea truly as a network, there full-screen floating in front of you. There wasn’t a way to characterise the experience of this in advance: that was the reason for doing it, to explore just what it could be.

This project has not been able to pursue that goal – developing the ontology and datasets has taken all but a sliver of the project’s run – but has created a proof-of-concept showing the technical foundations for how this could be achieved.

Cross-platform and open-source software was viewed as desirable by both those in the BBC and Queen Mary, and so the decision was made to use one of the processing.js, Processing, or openFrameworks. These all share the same approach to visualisation, but are implemented in different languages. The same codebase could then be ported between a HTML5 environment to a highly performant c++ compiled app without too much change in paradigm. processing.js was chosen as it



Top: Loading N3 into triplestore and test query.
Middle & Bottom: HTML5 interactive in-use.
Taken as stills from demo video “TobyHarris-StoriesRDF-20Aug”

represented a path to visualisation on the Web using technology native to the browser, thanks to the Canvas element introduced by HTML5.

A demonstration has been produced and the video walk-through should be viewed as part of this report. It is available online at <http://www.eecs.qmul.ac.uk/~tbz/digitalnarrativesbbc/TobyHarris-StoriesRDF-20Aug.mov>

Processing.js was still in beta at time of development, and was found to have significant performance issues when rendering text, which was central to the visualisation. The development generally of processing.js and browser support of the canvas element (and font support within it) appears in to be one of constant, fast-paced improvement, and so this issue is viewed as one of being an early adopter.

It is hoped the visualisation will be developed further for inclusion as part of the BBC’s data art project – the form of which will depend on what data does become available from any deployment of BBC Stories driven websites – or as an internal demonstration synchronising playback of the DW S4E1 programme with the visualisation of that story, such that you can navigate both by video timeline and by the story network.

The code for the HTML page, the API abstracting the SPARQL queries, and the processing.js sketch is listed in appendices 3, 4 and 5.

Conclusions

The introduction features a progression of interesting prototypes made by the BBC that have remained just that – prototypes. I would take the Bruno Latour inspired view that these projects have added layers of reality to a wider project, that they have affected the ever changing network of actors representing the stakeholders, the technologies, the vision. As an academic project in conjunction with the BBC, attempting to add another layer of reality on, playing a role as an actor within that network, these intangibles are the only possible actual outcome. But yet it is hard not feel that this project has to be judged by whether BBC Stories really gets to drive a website, whether that Archive pilot recently underway is the one that manages to take flight.

This project has made two broad contributions. Working as a group with Paul Rissen and Michael Jewell, an ontology for the BBC to use for modelling their content has been developed. We have debated its finer points and possibilities endlessly, and I have the chat logs to prove it. As such, my feeling is that the place for critical analysis of this is not from any of us, but from the people who take it on. This is not a cop-out: as discussed in the review section, for any claim of validity from the production or real alignment with end-users the ontology should somehow embody a constant feedback process with its users, seeing what balance can be made between a shared worldview and a world of interpretations. And really, the whole process of post-hoc creation of the story world data is quite ridiculous: surely it should be embedded in the production process, gradually taking form as writers, editors, production staff access and develop it as a shared resource? It would feed back into the production process (somebody is already making prop lists, right?), but more importantly, at that point you have the feedback process there in front of you.

The central notion of interpretations is the second broad contribution. The notion that we are trying to ‘wikipedia-ise’ BBC programming has been a common one, and

useful whenever explaining the project. But its actually exactly wrong for one very important reason: wikipedia attempts to present a shared world view on that topic, and despite this project being based around notions of ontology and modelling, this is exactly what we are trying not to do. Rather than present some form of lowest common denominator of agreement amongst the BBC content, we want the richness of all the BBC’s programming to be exposed, drawn together with a web of connections so people can immerse themselves in a world of content from which they can draw their own conclusions, and, perhaps yes, choose their own adventure.

References

- Berners-Lee, T. (1998). Notation3 (N3) A readable RDF syntax. Retrieved Thursday, 19 August 2010 from <http://www.w3.org/DesignIssues/Notation3>.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5):34-43.
- Berners-Lee, T. (2006). Linked Data - Design Issues. Retrieved Friday, 20 August 2010 from <http://www.w3.org/DesignIssues/LinkedData>.
- Bizer, C. & Cyganiak, R. (2004). The TriG Syntax. Retrieved Sunday, 22 August 2010 from <http://www4.wiwiss.fu-berlin.de/bizer/TriG/>.
- Bizer, C. (2009). The Emerging Web of Linked Data. *Intelligent Systems, IEEE*, 24(5), 87 – 92.
- Brickley, D. & Miller, L. FOAF Vocabulary Specification. Retrieved Wednesday, 18 August 2010 from <http://xmlns.com/foaf/spec/>.
- Donovan, J. O. (2010). BBC - BBC Internet Blog: The World Cup and a call to action around Linked Data. Retrieved Tuesday, 17 August 2010 from http://www.bbc.co.uk/blogs/bbcinternet/2010/07/the_world_cup_and_a_call_to_ac.html
- Ferne, T. (2008). BBC - Radio Labs: 10 percent time. Retrieved 16 August 2010 from http://www.bbc.co.uk/blogs/radiolabs/2008/01/ten_percent_time.shtml
- Ferne, T. (2008). BBC - Radio Labs: Archrs - an everyday story of web development. Retrieved Wednesday, 18 August 2010 from http://www.bbc.co.uk/blogs/radiolabs/2008/08/archrs_an_everyday_story_of_we.shtml
- Ferne, T. (2010). BBC - Research and Development: The Mythology Engine - representing stories on the web. Retrieved Wednesday, 18 August 2010 from <http://www.bbc.co.uk/blogs/researchanddevelopment/2010/03/the-mythology-engine-represent.shtml>
- Gold, M. (2010). Johnny Carson's 'Tonight Show' digitized for searchable database - latimes.com. Retrieved Monday, 16 August 2010 from <http://www.latimes.com/entertainment/news/la-et-0811-carson-20100811,0,7443493.story>.
- Hausenblas, M. (2009). Exploiting Linked Data to Build Web Applications. *Internet Computing, IEEE*, 13(4), 68 – 73.
- Hirst, T. (2009). The Dr Who Effect on Google Search Trends « OUseful.Info, the blog. Retrieved Tuesday, 17 August 2010 from <http://blog.ouseful.info/2009/03/11/the-dr-who-effect-on-google-search-trends/>

- Hunter, J. (2001). Adding Multimedia to the Semantic Web – Building an MPEG-7 ontology. *International Semantic Web Working Symposium*
- Iacobacci, N. (2008). Lunch over IP: From crossmedia to transmedia: thoughts on the future of entertainment. Retrieved Thursday, 19 August 2010 from <http://www.lunchoverip.com/2008/05/from-crossmedia.html>
- Lawrence, K. & Schraefel, M. (2006). Freedom and Restraint Tags, Vocabularies and Ontologies. *Information and Communication Technologies, 2006. ICTTA '06. 2nd, 1*, 1745 – 1750.
- Lawrence, K., Jewell, M., & Prugel-Bennett, A. (2006). Annotation of Heterogenous Media Using OntoMedia. *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*.
- Lawrence, K. (2005). OntoMedia-creating an ontology for marking up the contents of fiction and other media. *eprints.ecs.soton.ac.uk*.
- Littledale, A. (2010). BBC - Research and Development: Data Art on Backstage. Retrieved 16 August 2010 from <http://www.bbc.co.uk/blogs/researchanddevelopment/2010/06/data-art-on-backstage.shtml>
- McCandless, D. (2009). Time Travel In TV and Film. Retrieved Friday, 20 August 2010 from <http://www.informationisbeautiful.net/book/process/time-travel-in-tv-and-film/>
- Miller, A. & McParland, A. (2010). BBC - Research and Development: Subtitles + Internet = Advanced TV Research. Retrieved 16 August 2010 from <http://www.bbc.co.uk/blogs/researchanddevelopment/2010/02/subtitles-internet-advanced-tv.shtml>
- Miller, P., Berners-Lee, T. (2008). Nodalities » Blog Archive » Sir Tim Berners-Lee Talks about the Semantic Web. Retrieved Thursday, 19 August 2010 from http://blogs.talis.com/nodalities/2008/02/sir_tim_bernerslee_talks_about_1.php
- Ray, K & Shirky, C. (2010). Web 3.0 Transcript – Now with links! « kateray.net. Retrieved Friday, 20 August 2010 from <http://kateray.net/2010/05/17/transcript/>
- Rayfield, J. (2010). BBC - BBC Internet Blog: BBC World Cup 2010 dynamic semantic publishing. Retrieved Tuesday, 17 August 2010 from http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html
- Rissen, P. (2009). The Fourth Medium | R4isStatic.com. Retrieved 16 August 2010 from <http://www.r4isstatic.com/?p=68>
- Romaniuk, C. (2003). SUDS PowerPoint Presentation. Retrieved Wednesday, 18 August 2010 from <http://www.authorstream.com/Presentation/Quintino-47524-SUDS-Soap-operas-semantic-web-EastEnders-vs-American-soaps-as-Entertainment-ppt-powerpoint/>
- Shirky, C. (2005). Shirky: Ontology is Overrated – Categories, Links, and Tags. Retrieved Saturday, 21 August 2010 from

http://www.shirky.com/writings/ontology_overrated.html

Shirky, C. (2003). Shirky: The Semantic Web, Syllogism, and Worldview. Retrieved Friday, 20 August 2010 from http://www.shirky.com/writings/semantic_syllogism.html

Staab, S., Santini, S., Nack, F., Steels, L., & Maedche, A. (2002). Emergent semantics. *Intelligent Systems, IEEE*, 17(1), 78 – 86.

Walpole, S. (2007). BBC - BBC Internet Blog: A Page For Every Programme. Retrieved Tuesday, 17 August 2010 from http://www.bbc.co.uk/blogs/bbcinternet/2007/11/a_page_for_every_programme_1.html

Raimond, Y., Scott, T., Sinclair, P., Miller, L., Betts, S. & McNamara, F. (2010). Case Study: Use of Semantic Web Technologies on the BBC Web Sites. Retrieved Thursday, 19 August 2010 from <http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/>

4store - Scalable RDF storage. Retrieved Thursday, 19 August 2010 from <http://4store.org/>

BBC - About the BBC - Home. Retrieved Wednesday, 18 August 2010 from <http://www.bbc.co.uk/aboutthebbc/>

BBC Strategy Review March 2010. (2010). Electronic document, http://downloads.bbc.co.uk/aboutthebbc/reports/pdf/strategy_review.pdf, accessed Thursday, 19 August 2010.

DataArt on BBC Backstage - About. (2009). Retrieved Thursday, 19 August 2010 from http://backstage.bbc.co.uk/data_art/about.php

Johnny Carson I Clip Licensing. (2010). Retrieved Thursday, 19 August 2010 from <http://www.johnnycarson.com/clip-licensing.jsp>

Raptor RDF Syntax Library. Retrieved Sunday, 22 August 2010 from <http://librdf.org/raptor/>

W3C Opens Data on the Web with SPARQL. (2008). Retrieved Wednesday, 18 August 2010 from <http://www.w3.org/2007/12/sparql-pressrelease>