

The Programmes Ontology – making BBC programme metadata accessible to all

Tom Scott and Nicholas Humfrey

BBC (Future Media and Technology)

tom.scott@bbc.co.uk and nicholas.humfrey@bbc.co.uk

BBC Programmes online – a brief history

Historically the BBC website has focused on providing information and rich online experiences to support only the major BBC programmes but, with well over 1,000 programmes broadcast every day across radio and TV networks, web coverage to date has not been comprehensive. Nor has it been permanent. Those programmes that have had webpages have often not been represented by persistent URLs – often the same URL has been used to represent the latest episode.

In October 2007, the Audio and Music department, on behalf of the rest of the BBC launched a new service, [BBC Programmes](#), to address these issues. BBC Programmes aims to ensure that every programme the BBC broadcasts has a permanent, findable web presence.

BBC Programmes publishes programme metadata for all programmes on the eight BBC TV channels, ten national radio stations and the six stations covering Scotland, Northern Ireland and Wales. And the media for those programmes as found on iPlayer and Radio iPlayer; the BBC's on-demand service that allows UK-based users to view a selection of programmes broadcast on the BBC networks from the last seven days.

Our objective in developing BBC Programmes was to ensure that the data was made as accessible as possible. We didn't want to build another silo of data – instead we wanted to make a service that could be integrated into existing BBC online products as well as the wider web. This objective has significantly affected our design decisions and the resultant service. This paper explores these design decisions and the future direction of the service.

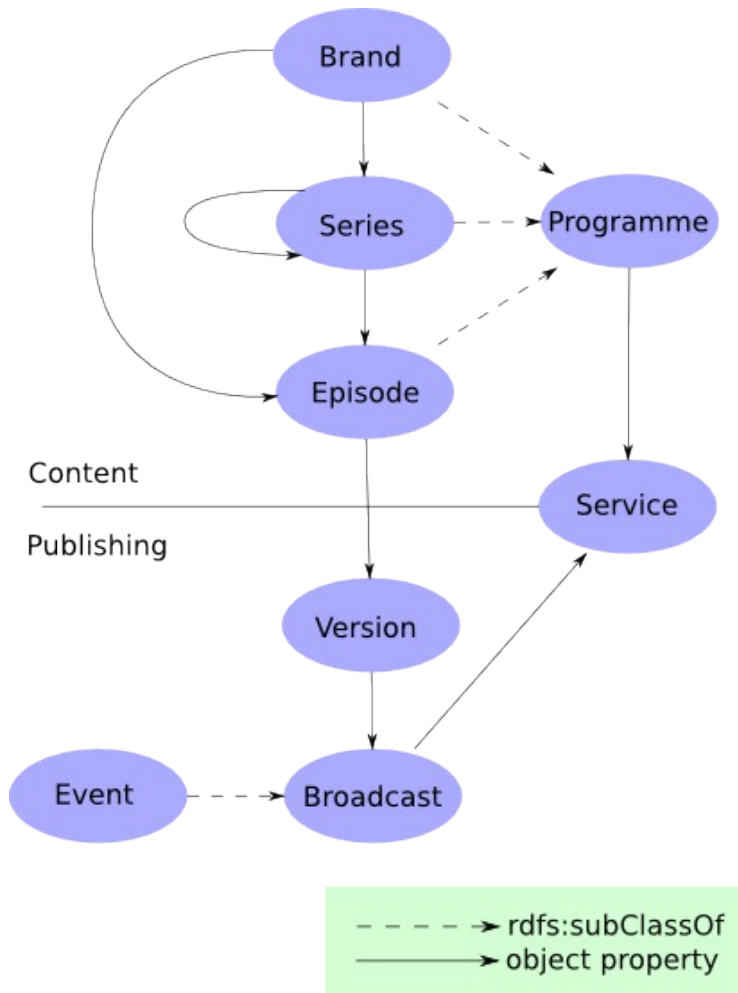
Linked data

To enable the sharing of programme metadata in a structured way, we are investigating the [linked data](#) approach, where resources on the web can be far more than just documents. They can identify anything, from a particular person to a particular programme. These resources have representations, which can be machine-processable (through the use of RDF, Microformats, RDFa, etc.), and these representations can hold links towards further web resources, allowing agents to jump from one dataset to another.

To facilitate this approach we have designed and implemented a service that exposes the data in a variety of representations via predictable, persistent URLs. We have also published, under a Creative Commons license, an ontology to describe how the different objects that make up a programme are modelled.

Programmes ontology

In order to provide direct access to the actual data backing BBC Programmes, we designed a Semantic Web ontology covering programmes data - the [Programmes Ontology](#). This ontology provides web identifiers for concepts such as brand, series, and episode. The ontology is divided into two main parts. First, it captures categorical information about programmes, and the relations between such categories. For example, it allows the description of a brand, a series constituting it, a sub-series and an episode in it. The second part of the ontology describes episodes' versions and their broadcast on a particular service.



We have designed the ontology so that it can describe any broadcasters' programmes, both live and on-demand. For example, the Southampton University student radio station, Surge, is using the Programmes Ontology to expose its schedule and programme information. Our hope is that the ontology will be used by broadcasters to interchange and interlink schedule information.

Data serialisation

When the service launched we ensured the standard desktop xHTML views were semantically marked up – this included the use of hCard and hCalendar microformats where appropriate. However, since then we have been working on a series of alternate representations, including: RSS, ATOM, iCal, XML, JSON, YAML and RDF.

We have recently launched schedule views as XML, JSON and YAML and episode, series and brand as RDF. In addition to these machine readable

representation we have also developed a mobile (xHTML MP) version of the site which is due to launch in early May.

Each representation is addressed by appending the name of the appropriate serialisation to the end of the URL. For example the URL for today's BBC 1 (London) schedule is:

```
http://www.bbc.co.uk/bbcone/programmes/schedules/london/today.json
```

In due course a combination of device detection and content negotiation will also be used to ensure the appropriate representation is delivered to the client device.

The BBC is licensing this data under a non-commercial license.

It is worth noting that our design approach to BBC Programmes has made the work of creating alternate serialisations relatively straightforward. Our approach was to start by identifying the individual resources and make these addressable; each public facing webpage is then made by transcluding a subset of those resources onto the page.

What this approach means is that we don't need to build an entirely separate and parallel mobile site; nor a mobile site with the same content and navigation as the desktop version. Instead we can pick and choose from a common set of resources but combine them in a different fashion to create a mobile site optimised for that platform. And the same logic applies when creating the RSS or RDF views.

URL design

Our primary focus when it came to designing the URLs for BBC Programmes was to ensure that they remain persistent – with each URL representing a single concept. What we've ended up with are two classes of page: aggregations and objects.

The aggregations are human readable, easily hackable and provide a list of objects. They include schedule views, aggregation by genre, format and a to z. For example, a list of music programmes on BBC 1:

```
bbc.co.uk{/ :service}/programmes/genres/:genre  
bbc.co.uk/bbcone/programmes/genres/music
```

Or Radio 4's FM schedule:

```
bbc.co.uk{/ :service}/programmes/schedules{/ :outlet}/{hack  
or date}  
bbc.co.uk/radio4/programmes/schedules/fm
```

The objects however follow a simpler pattern:

```
bbc.co.uk/programmes/:id
```

Where `:id` is an eight digit alphanumeric code. Our decision to use opaque URLs was driven by the need to provide persist identifiers. If we had included information about the programme brand (e.g. the Today Programme) or service (e.g. Radio 3) then there would have been a high risk that the URLs would have either changed or no longer reflected the ‘owning’ brand since programmes are often rebroadcast on different services. By stripping the URL back to a unique identifier we removed any future (or current) ownership issues.

Future plans – linking data

As discussed above we are also working on exposing music metadata in a similar fashion. Built around MusicBrainz (which is already part of the Linking Open Data project) to provide common GUIDs and basic metadata we are currently in the process of rebuilding the music website.

The relaunched music site will connect programmes and music e.g. artists played on with episode, series or programme brand; programme brands who feature this artist. These connections will be made available both explicitly on the relevant webpage but also via a SPARQL end point.

Using D2R Server, a Java application for mapping relational databases to RDF accessible through SPARQL, we have already internally trialled BBC Programmes as Linked Data. Around 5 million RDF triples are exposed this way. Multiple views are available for individual records, such as a brand, series or episode, and the SPARQL interface allows the data to be queried directly. By using D2R we avoid synchronization issues as the latest live data is always exposed.

Through the use of SPARQL, we can query the data using a variety of constraints that cannot be easily expressed through the Programmes web interface. We are also able to semantically connect to external data sources such as DBpedia to provide extra information that is not present in our dataset, such as date and place of birth of cast members.

Following programmes and music we will then address events (such as festivals), users (both end users and the programme's cast and crew) and topics (places, time periods and subjects). In each case we will give these resources permanent URIs and publish a series of representations.