

# How the BBC is approaching metadata quality measurement



Jeremy Tarling, Lead Data Governance Specialist  
BBC Design + Engineering

# How the BBC is approaching metadata quality measurement

- BBC content tagging 101
- Problems with the current approach
- Building better metadata - the Common Metadata project
- Introducing machine tagging
- Metadata quality measurement

# **BBC content tagging 101:** **Content <- about -> Thing**

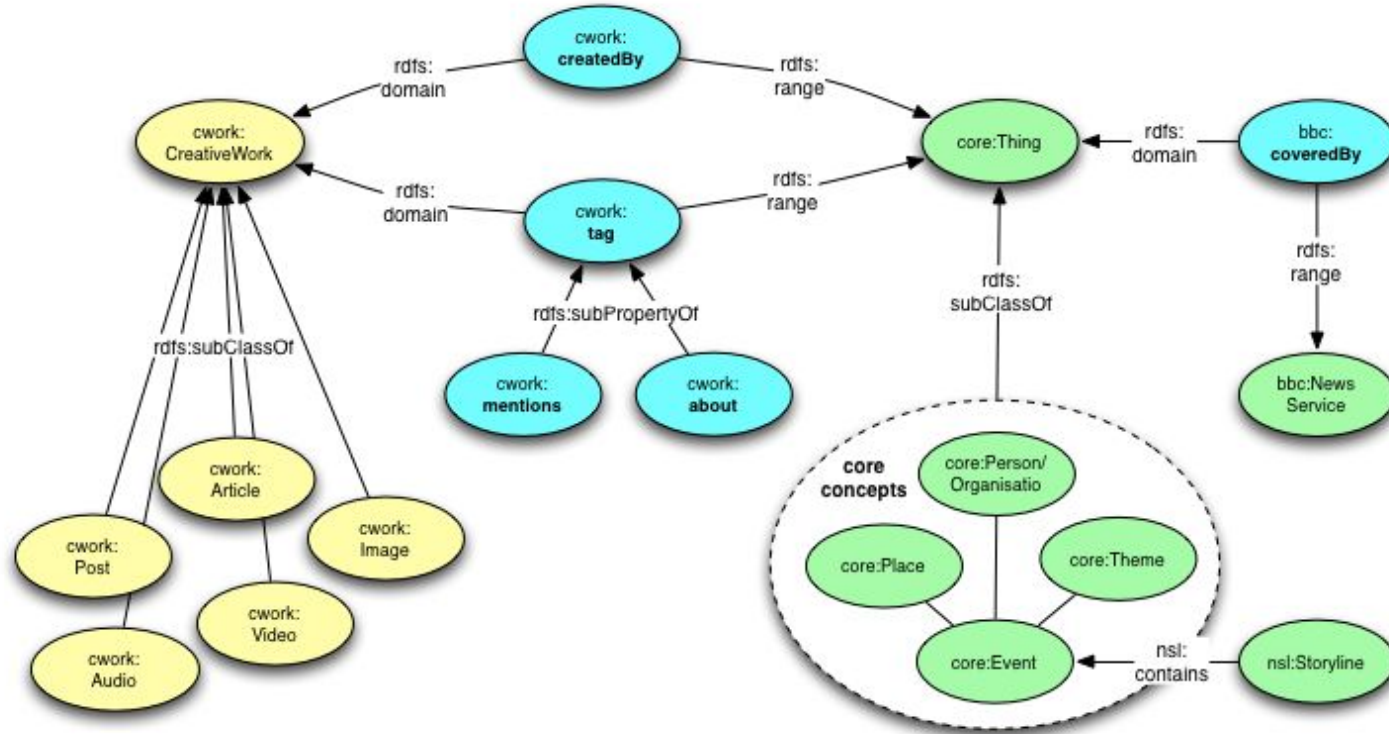
# BBC content tagging 101

## Journalism

### News, Sport

- Human tagging by subject
- > 2,000,000 tagged articles
- > 70,000 'Thing' tags
- > 5,000 topic pages

# BBC content tagging 101: journalism




# BBC content tagging 101: journalism

**BBC NEWS**

**Boris Johnson**


0:01  
**'Big gap' remains as Brexit deal talks to continue**



The future of Northern Ireland's border dominates, while Parliament reconvenes for the Queen's Speech.


[Read more >](#)

22:01 13 Oct  
**PM says 'significant' work to do on Brexit deal**



**BBC SPORT**

Manchester United



**Inter's Sanchez injured on Chile duty**  
15h

**World Cup winner Schweinsteiger retires**  
5d

**How bad could it get for Man Utd?**  
6d

**Man Utd displays unacceptable - De Gea**  
7d

**Newcastle win to deepen Man Utd worries**  
7d

James wondered 'is football for

**BBC NEWS**

**Wolverhampton**  
England, United Kingdom

Set as your local news?  Yes

+5 miles +10 miles  
+20 miles +30 miles

**LIVE Midlands Live: Breaking news and local stories**

**The pain that Wolverhampton stab victim's killers walk free**  
10h


**Four injured in 'mass brawl' at wedding**  
1h

**Cyrille Regis inducted into football's Hall of Fame**  
1d

**Pedestrian hit by 4x4 dies**  
1d

**BBC SPORT**

Athletics



**Kosgei shatters women's marathon record**  
10h

**Watch Kosgei smash Radcliffe's marathon world record**  
13h

**Farah eighth as Cherono wins Chicago Marathon**  
13h

**Kipchoge breaks two-hour marathon mark**  
1d

**Kipchoge expects others to repeat marathon feat**  
1d

**Reaction after Kipchoge runs**


# BBC content tagging 101: journalism

**BBC NEWS**

**Boris Johnson**

0:01

**Person**




The future of Northern Ireland's border dominates, while Parliament reconvenes for the Queen's Speech.

[Read more >](#)

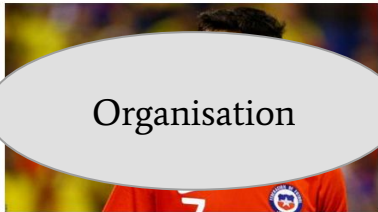
22:01 13 Oct

**PM says 'significant' work to do on Brexit deal**



**BBC SPORT**

Manchester United



**Organisation**

**Inter's Sanchez injured on Chile duty**

15h

**World Cup winner Schweinsteiger retires**

5d

**How bad could it get for Man Utd?**

6d

**Man Utd displays unacceptable - De Gea**

7d

**Newcastle win to deepen Man Utd worries**

7d

James wondered 'is football for

**BBC NEWS**

Wolverhampton

England

+20 miles

**Place**

**LIVE Midlands Live: Breaking news and local stories**

**The pain that Wolverhampton stab victim's killers walk free**

10h

**Four injured in 'mass brawl' at wedding**

1h

**Cyrille Regis inducted into football's Hall of Fame**


1d

**Pedestrian hit by 4x4 dies**

1d

**BBC SPORT**

Athletics



**Theme**

**Kosgei shatters women's marathon record**

10h

**Watch Kosgei smash Radcliffe's marathon world record**

13h

**Farah eighth as Cheroni wins Chicago Marathon**

13h

**Kipchoge breaks two-hour marathon mark**

1d

**Kipchoge expects others to repeat marathon feat**

1d

**Reaction after Kipchoge runs**

# BBC content tagging 101

## Journalism

News, Sport

- Human tagging by subject
- > 2,000,000 tagged articles
- > 10,000 topic pages

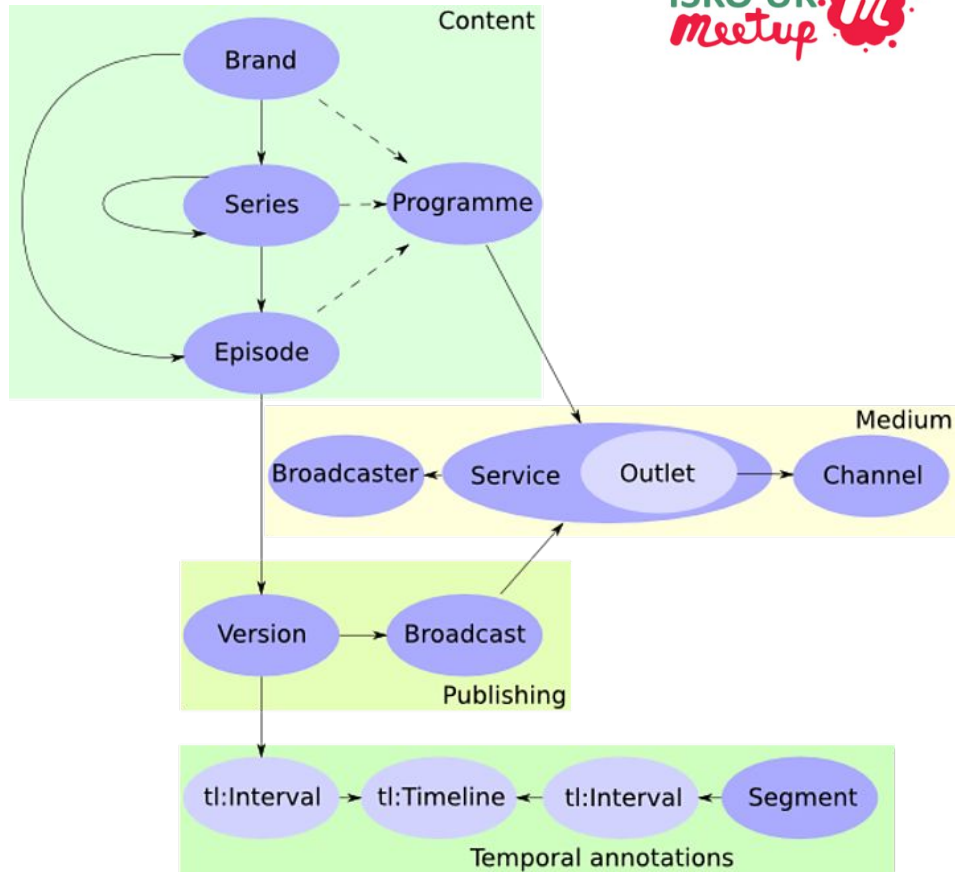
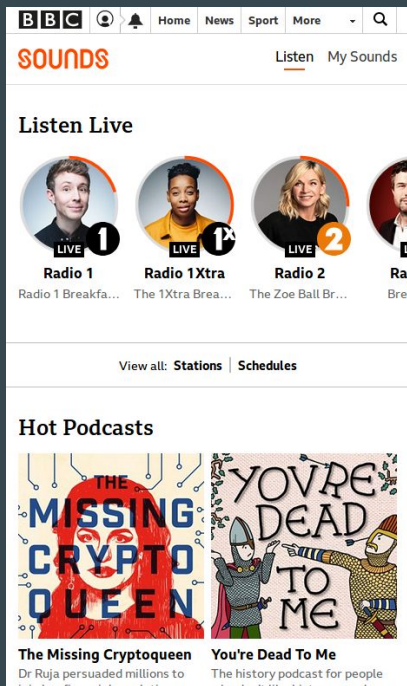
## Programmes

iPlayer, Sounds

- No tagging by subject (except R4)
- Human curation of content by genre and format



# BBC content tagging 101: iPlayer and Sounds



---> rdfs:subClassOf  
 —> object property

# BBC content tagging 101

## Journalism

News, Sport

- Human tagging by subject
- > 2,000,000 tagged articles
- > 10,000 topic pages

## Programmes

iPlayer, Sounds

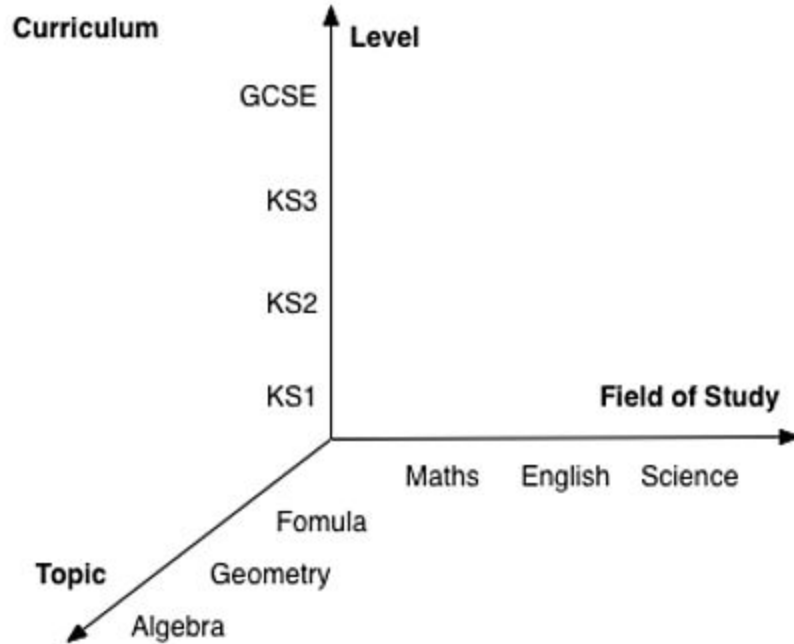
- No tagging by subject
- Human curation of content by genre and format

## the rest...

Bitesize, Recipes, etc

- Bespoke domain models
- Human tagging and curation

# BBC content tagging 101: Bitesize, Recipes, etc...



BBC CBBC CBBC iPlay

**Bitesize**

KS2

**Science**

Part of [Learn & revise](#)

**The living world >**

**Animals**  
4 guides

**What is a microorganism?**

BBC Home News Sport

**FOOD**

Discover recipes

Food occasions within the

**Diwali**  
27 Oct - 31 Oct  
Happy Diwali! Recipes and food ideas for the vibrant Hindu festival of lights.  
[See all recipes for Diwali \(82\)](#)

OCCASION

BBC Sign In News Sport

**Religions**

Religion can concern the of the univer as the creati agencies, usu ritual observ moral code g affairs.

**Featured religions and beliefs**

<p><b>Atheism</b></p> <p>Atheists are people who believe that god or gods are man-made constructs.</p>	<p><b>Baha'i</b></p> <p>One of the youngest of the world's major religions.</p>
<p><b>Candomblé</b></p> <p>A religion based on African beliefs, originating in Brazil.</p>	<p><b>Christianity</b></p> <p>The world's biggest faith, based on the teaching of Jesus Christ.</p>
<p><b>Islam</b></p> <p>Revealed in its final form by the Prophet Muhammad.</p>	<p><b>Jainism</b></p> <p>An ancient philosophy and ethical teaching that originated in India.</p>
<p><b>Judiasm</b></p> <p>Based around the Jewish people's covenant relationship with God.</p>	<p><b>Mormonism</b></p> <p>The Church of Jesus Christ of Latter-day Saints.</p>
<p><b>Rastafari</b></p> <p>A young religion founded in Jamaica in the 1930s.</p>	<p><b>Santeria</b></p> <p>Afro-Caribbean syncretic religion originating in Cuba.</p>

# **Problems with the current approach**

## **Process, accuracy, quality and scale**

# Poor tagging process

embedded in journalism workflows 😊

but...

70,000 tags to choose from 😞

no taxonomic hierarchy to filter on 😞

easy to mis-tag, no second pair of eyes 😞

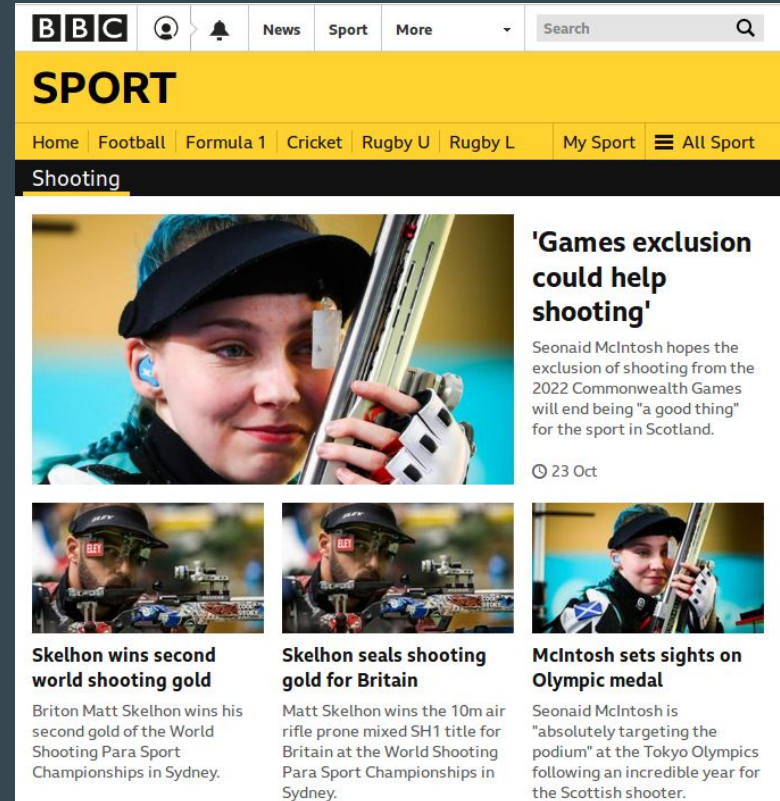
some tags are ambiguous 😞



# Poor tagging accuracy

It can be hard for a human to deal with tag ambiguity

For example, Sport have topic pages for all olympic disciplines, including shooting...



The screenshot shows the BBC Sport website interface. At the top, there is a navigation bar with the BBC logo, a search bar, and menu items for News, Sport, and More. Below this is a yellow header with the word 'SPORT' in large, bold letters. Underneath, there is a secondary navigation bar with links for Home, Football, Formula 1, Cricket, Rugby U, Rugby L, My Sport, and All Sport. The main content area is titled 'Shooting' and features a large article at the top with a photo of Seonaid McIntosh. Below this are three smaller article thumbnails, each with a photo and a headline.

**BBC** Search

News Sport More

**SPORT**

Home Football Formula 1 Cricket Rugby U Rugby L My Sport All Sport

**Shooting**

**'Games exclusion could help shooting'**

Seonaid McIntosh hopes the exclusion of shooting from the 2022 Commonwealth Games will end being "a good thing" for the sport in Scotland.

23 Oct

**Skelhon wins second world shooting gold**

Briton Matt Skelhon wins his second gold of the World Shooting Para Sport Championships in Sydney.

**Skelhon seals shooting gold for Britain**

Matt Skelhon wins the 10m air rifle prone mixed SH1 title for Britain at the World Shooting Para Sport Championships in Sydney.

**McIntosh sets sights on Olympic medal**

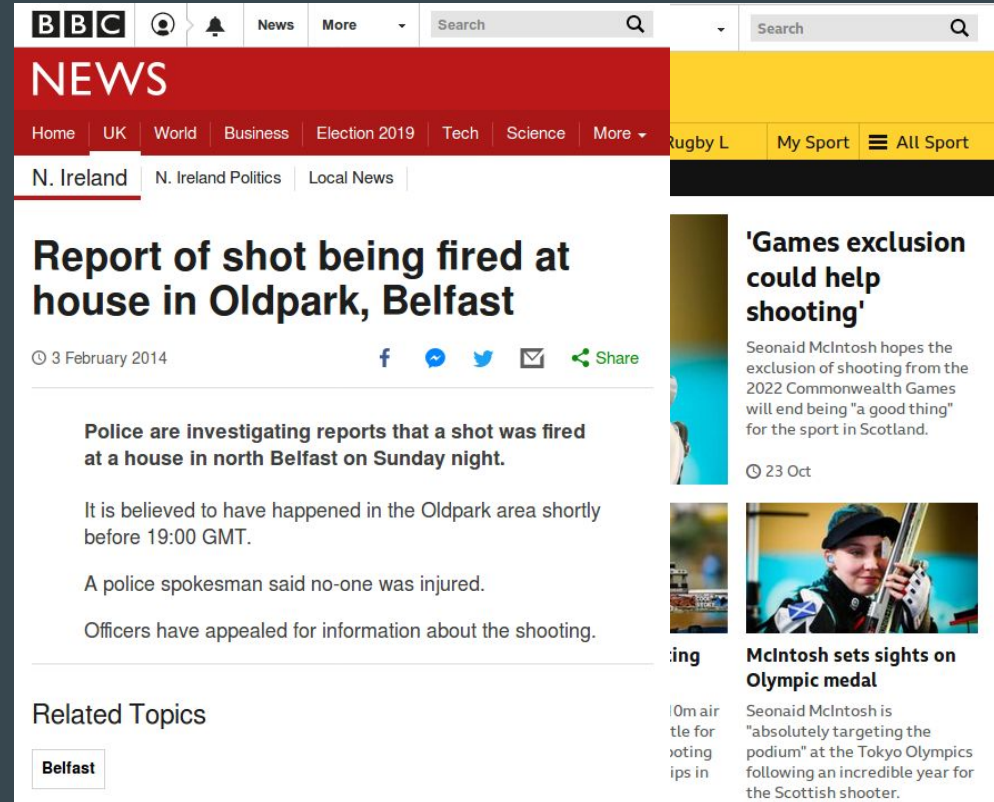
Seonaid McIntosh is "absolutely targeting the podium" at the Tokyo Olympics following an incredible year for the Scottish shooter.

# Poor tagging accuracy

It can be hard for a human to deal with tag ambiguity

For example, Sport have topic pages for all olympic disciplines, including shooting...

But shooting can sometimes happen outside of sport...



The screenshot shows a BBC News article titled "Report of shot being fired at house in Oldpark, Belfast" dated 3 February 2014. The article text states that police are investigating reports of a shot fired at a house in north Belfast on Sunday night, believed to have occurred in the Oldpark area shortly before 19:00 GMT. A police spokesman said no-one was injured, and officers have appealed for information about the shooting.

Below the main article, there is a "Related Topics" section with a tag for "Belfast".

On the right sidebar, there is a related article titled "'Games exclusion could help shooting'" dated 23 Oct. The article mentions Seonaid McIntosh, a Scottish shooter, who hopes the exclusion of shooting from the 2022 Commonwealth Games will end being "a good thing" for the sport in Scotland. Below this is another article snippet titled "McIntosh sets sights on Olympic medal" dated 10m air, mentioning her performance at the Tokyo Olympics.

# Poor tagging accuracy

## SPORT SHOOTING


Home | Football | Formula 1 | Cricket | Rugby U | Rugby L | Tennis | Golf | Athletics | Winter Olympics
More Sport ▾

Shooting ▾

Business | Election 2019 | Tech | Science | More ▾


Rugby L | My Sport | All Sport

Politics | Local News

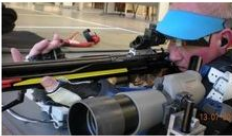


### Report of shot being fired at house

Police are investigating reports that a shot was fired at a house in north Belfast on Sunday night.



**Sykes waits for Glasgow call-up**  
Teenage shooter Larissa Sykes says an England call-up for the Commonwealth Games would cap a brilliant 12 months.




**Pub landlord targets eighth Games**  
A pub landlord from the Isle of Man is targeting his eighth successive Commonwealth Games in Glasgow.

### Headlines

- Shotgun star Hill eyes Glasgow 2014
- Hill wins Young Sports Personality
- Trio on Young Sports Personality list
- Shooters & triathlete join Scots team
- Bevis wins gold and silver at Euros
- Kerwood wins shotgun gold for GB
- Delays affect 2014 ticket emails
- 2014 ticket draw 'random and fair'
- Fans told about Glasgow 2014 tickets
- Allen puts Olympic pain behind her

### Olympics Shooting



**Shot being fired at Oldpark, Belfast**

Complete London 2012 archive


f
w
t
✉
Share

#### Things To Do

RUN BY THE BBC AND PARTNERS


[Find sport taster activities near you >](#)

### ing



**'Games exclusion could help shooting'**  
Seonaid McIntosh hopes the exclusion of shooting from the 2022 Commonwealth Games will end being "a good thing" for the sport in Scotland.

🕒 23 Oct

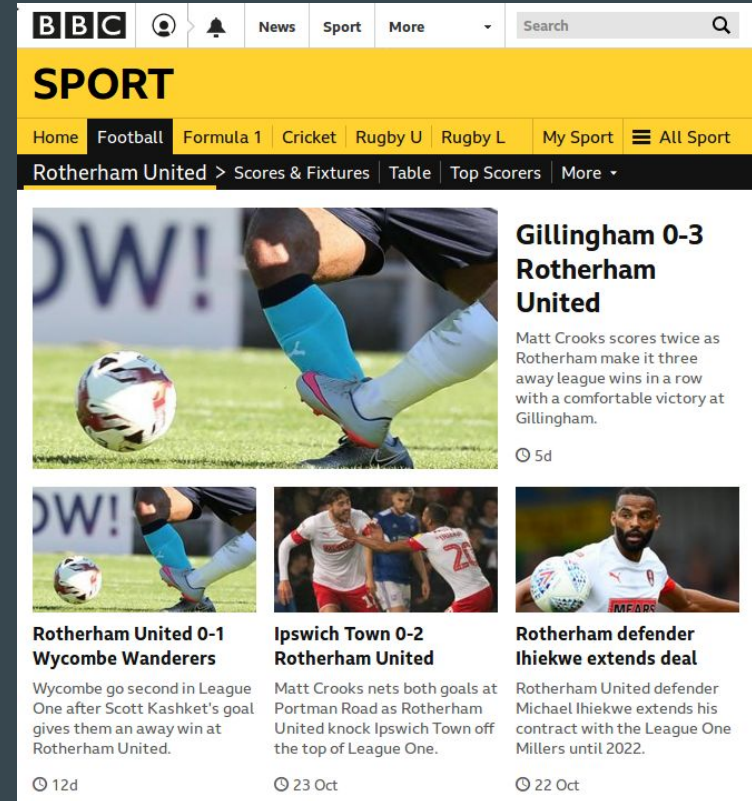


**McIntosh sets sights on Olympic medal**  
Seonaid McIntosh is "absolutely targeting the podium" at the Tokyo Olympics following an incredible year for the Scottish shooter.



# Poor tagging accuracy

Place names can be particularly problematic, especially when there are sports teams with the same name



**BBC** News Sport More Search

**SPORT**

Home **Football** Formula 1 Cricket Rugby U Rugby L My Sport All Sport

**Rotherham United** > Scores & Fixtures Table Top Scorers More

**Gillingham 0-3 Rotherham United**  
 Matt Crooks scores twice as Rotherham make it three away league wins in a row with a comfortable victory at Gillingham.  
 5d

**Rotherham United 0-1 Wycombe Wanderers**  
 Wycombe go second in League One after Scott Kashket's goal gives them an away win at Rotherham United.  
 12d

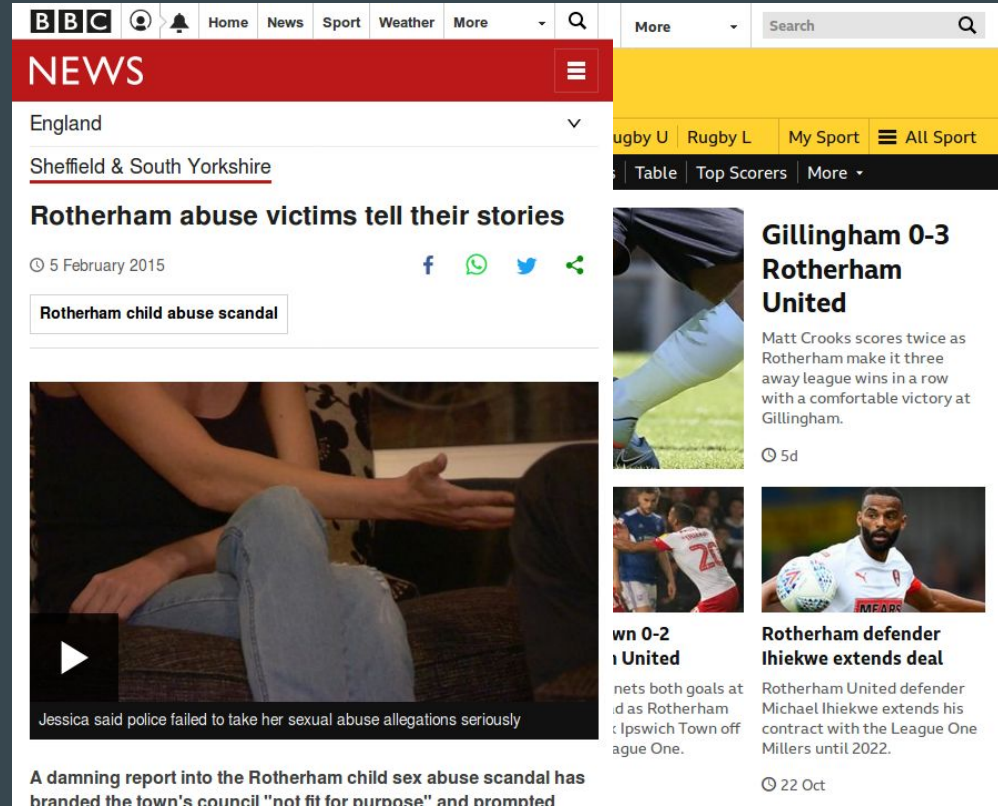
**Ipswich Town 0-2 Rotherham United**  
 Matt Crooks nets both goals at Portman Road as Rotherham United knock Ipswich Town off the top of League One.  
 23 Oct

**Rotherham defender Ihiekwe extends deal**  
 Rotherham United defender Michael Ihiekwe extends his contract with the League One Millers until 2022.  
 22 Oct

# Poor tagging accuracy

Place names can be particularly problematic, especially when there are sports teams with the same name

When a big News story is associated with a place it's all too easy to mistake the team for the location



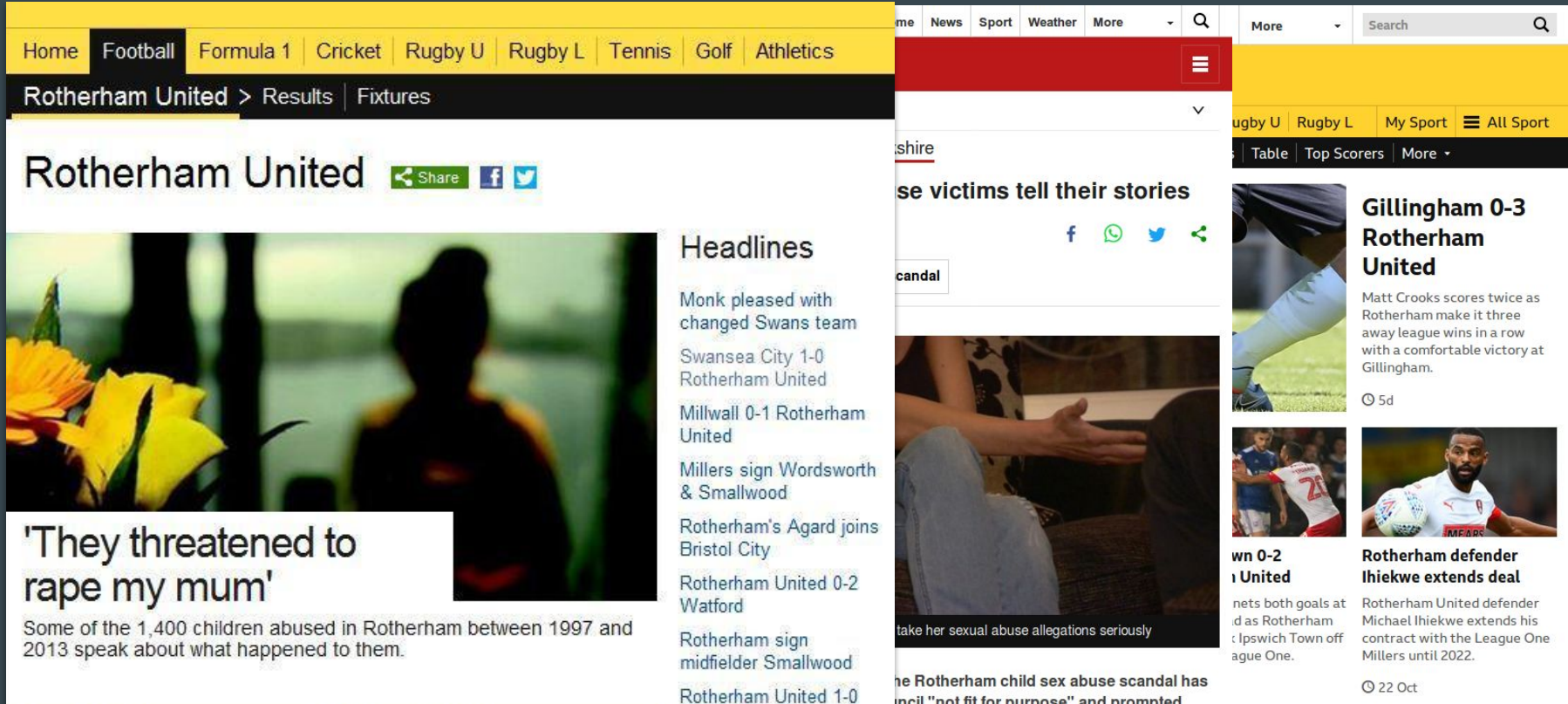
The screenshot shows a BBC News article titled "Rotherham abuse victims tell their stories" dated 5 February 2015. The article is tagged with "Rotherham child abuse scandal". A video player below the article shows a woman speaking, with a caption: "Jessica said police failed to take her sexual abuse allegations seriously".

The sidebar on the right features sports news:

- Gillingham 0-3 Rotherham United**: Matt Crooks scores twice as Rotherham make it three away league wins in a row with a comfortable victory at Gillingham. (5d)
- Wolves 0-2 Ipswich United**: Ipswich Town off to a winning start as they beat Wolves.
- Rotherham defender Ihiekwe extends deal**: Rotherham United defender Michael Ihiekwe extends his contract with the League One Millers until 2022. (22 Oct)

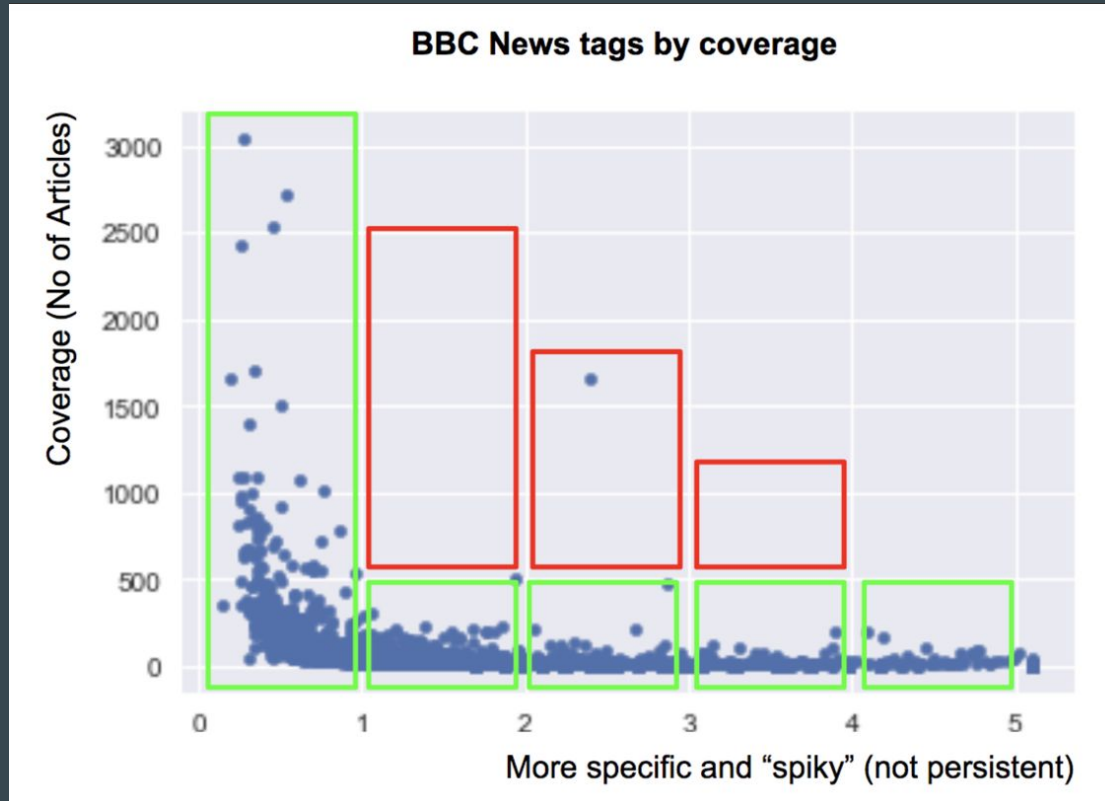
At the bottom of the page, a partial headline reads: "A damning report into the Rotherham child sex abuse scandal has branded the town's council 'not fit for purpose' and prompted..."

# Poor tagging accuracy



The screenshot shows a sports website with a yellow navigation bar containing links for Home, Football, Formula 1, Cricket, Rugby U, Rugby L, Tennis, Golf, and Athletics. Below the navigation bar, there are tabs for 'Rugby U', 'Rugby L', 'My Sport', and 'All Sport'. The main content area features a large headline: **'They threatened to rape my mum'**, with a sub-headline: **Some of the 1,400 children abused in Rotherham between 1997 and 2013 speak about what happened to them.** To the right of this headline is a 'Headlines' section with several links: **Monk pleased with changed Swans team**, **Swansea City 1-0 Rotherham United**, **Millwall 0-1 Rotherham United**, **Millers sign Wordsworth & Smallwood**, **Rotherham's Agard joins Bristol City**, **Rotherham United 0-2 Watford**, **Rotherham sign midfielder Smallwood**, and **Rotherham United 1-0**. On the right side of the page, there is a section for **Gillingham 0-3 Rotherham United**, with a sub-headline: **Matt Crooks scores twice as Rotherham make it three away league wins in a row with a comfortable victory at Gillingham.** Below this is a section for **Rotherham defender Ihiekwe extends deal**, with a sub-headline: **Rotherham United defender Michael Ihiekwe extends his contract with the League One Millers until 2022.** At the bottom of the page, there is a section for **The Rotherham child sex abuse scandal has**, with a sub-headline: **incil "not fit for purpose" and prompted**.

# Poor tagging quality



5% of our tags are broad and cover 85% of our content

95% of are tags are specific and cover 15% of our content

# Hardly any TV and radio tagging



Genre: 20010 (comedy)

Genre: 20010 (comedy)

according to our TV metadata these programmes are the same...

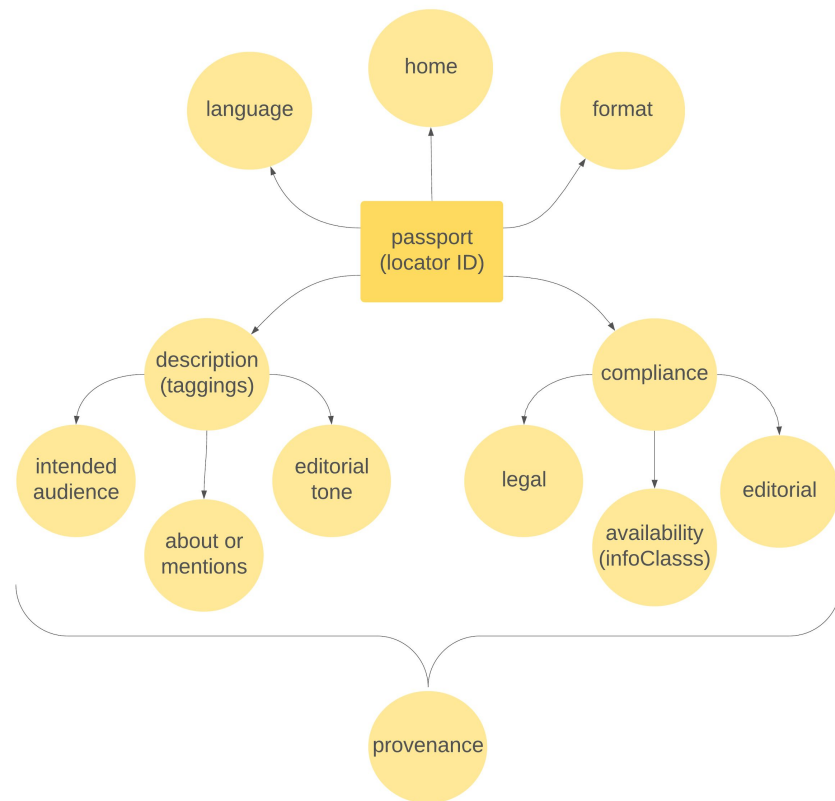
# Building better metadata: the Common Metadata project

# Building better metadata: defining common metadata

generalised schema for any BBC  
content object

new controlled vocabularies for  
editorial tone and intended audience

building in automation from the start  
by modelling *provenance*



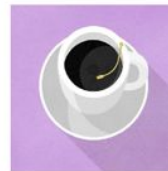
# Building better metadata: more than just 'about' tagging

There's a strong appetite in TV and Radio teams for better metadata, especially around editorial tone and audience behaviour

There's a willingness to experiment with automatic classification by these dimensions, but with human curation on top... for now



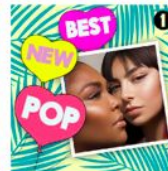
## Nostalgia



**Classical Fix**  
Elizabeth Day  
30 mins



**Radio 1's Workout Mix**  
Oliver Heldens, Krystal Klear and Disclosure  
28 mins



**Best New Pop**  
16 new pop songs you need to hear  
55 mins



**Alt. 00s from 6 Music**  
Heavy hitters and hidden gems from the 2000s  
62 mins

## Happy



**Radio 1 Playlist**  
Listen to the entire Radio 1 playlist  
137 mins



**Radio 1's Essential Mix**  
Andy C  
137 mins



**Pete Tong**  
The Month In Dance: May 2019  
120 mins



**Radio 1's Dance Anthems**  
Radio 1's Classic Dance Anthems with MistaJam  
60 mins

## Workout



**Behind the Album**  
Kate Tempest goes behind new album  
57 mins



**Jack Saunders' The Next Wave**  
Discover your new favourite band!  
56 mins



**Match of the Day Mix**  
FA Cup Final Mix  
92 mins



**Radio 1's Chill Mix**  
Jordan Rakei Chill Mix  
20 mins

**GiLL Sha**  
GiLL Sha  
30 m



# Building better metadata: editorial tone

working with commissioning experts  
and UX to define metadata to  
represent the editorial of a piece of  
content



preferredLabel	type(s)	disambiguationHint	URI
absurd	core:Thing core:EditorialTone tagging:TagConcept	nonsensical, ridiculous, silly	<a href="https://www.bbc.co.uk/things/161d7418-f265-4768-a348-b3cd4fcc8302#id">https://www.bbc.co.uk/things/161d7418-f265-4768-a348-b3cd4fcc8302#id</a>
camp		theatrical, exaggerated, ostentatious	<a href="https://www.bbc.co.uk/things/05ff0284-975e-49dc-b780-8dc9d695c816#id">https://www.bbc.co.uk/things/05ff0284-975e-49dc-b780-8dc9d695c816#id</a>
cerebral		intellectual, clever	<a href="https://www.bbc.co.uk/things/469989a7-ad22-48cb-b1f9-a4e1421b9952#id">https://www.bbc.co.uk/things/469989a7-ad22-48cb-b1f9-a4e1421b9952#id</a>
challenging		provocative, questioning, difficult	<a href="https://www.bbc.co.uk/things/90f03e7e-0f3b-4086-943e-7fcded85aa2a#id">https://www.bbc.co.uk/things/90f03e7e-0f3b-4086-943e-7fcded85aa2a#id</a>
complex		complicated, intricate	<a href="https://www.bbc.co.uk/things/c62d5f56-86da-4407-a609-060fa6cd022d#id">https://www.bbc.co.uk/things/c62d5f56-86da-4407-a609-060fa6cd022d#id</a>
complimentary		praising, approving, flattering	<a href="https://www.bbc.co.uk/things/8d273e3e-ee5b-40bc-8174-33260839e549#id">https://www.bbc.co.uk/things/8d273e3e-ee5b-40bc-8174-33260839e549#id</a>
critical		disapproving, fault- finding	<a href="https://www.bbc.co.uk/things/e8f0e6f3-b751-4148-967d-8568f17d6bb5#id">https://www.bbc.co.uk/things/e8f0e6f3-b751-4148-967d-8568f17d6bb5#id</a>
cynical		mistrusting, disbelieving, suspicious	<a href="https://www.bbc.co.uk/things/d0c8fd9a-dea8-4e6e-8381-35877f092812#id">https://www.bbc.co.uk/things/d0c8fd9a-dea8-4e6e-8381-35877f092812#id</a>
dark		sinister, bleak, morbid, mysterious	<a href="https://www.bbc.co.uk/things/119ea00e-55a4-4c97-a687-60ce9062e2d8#id">https://www.bbc.co.uk/things/119ea00e-55a4-4c97-a687-60ce9062e2d8#id</a>
deadpan		straight-faced, expressionless, impassive	<a href="https://www.bbc.co.uk/things/d47f1078-9498-4e0a-8cd3-d36de557f7bb#id">https://www.bbc.co.uk/things/d47f1078-9498-4e0a-8cd3-d36de557f7bb#id</a>

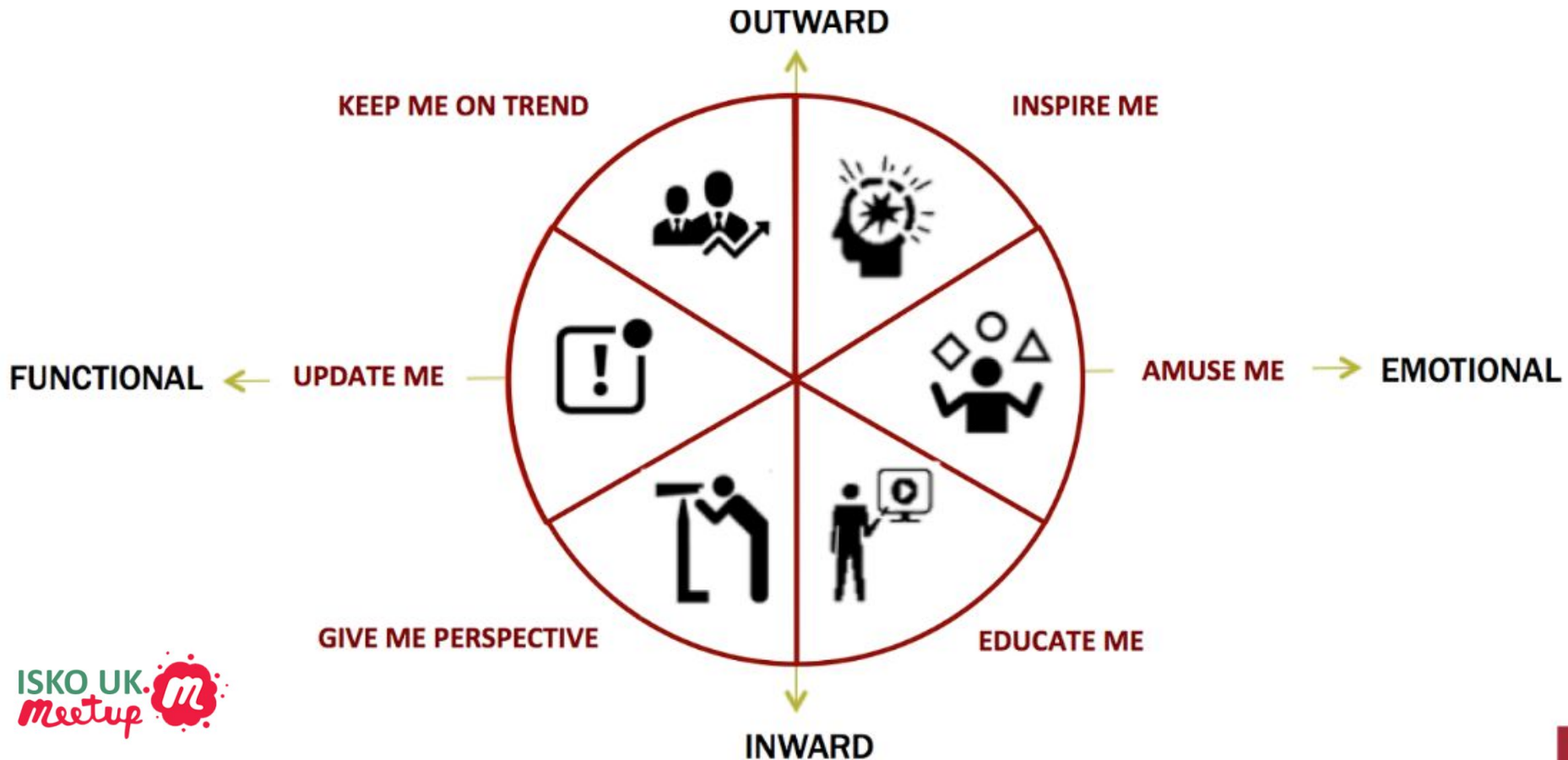
# Building better metadata: intended audience

working with UX and M&A experts to define metadata to represent the intended audience of a piece of content



preferredLabel	disambiguationHint	type(s)	URI
Male	Intended audience: gender	core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Gender	<a href="http://www.bbc.co.uk/things/15eb702e-c630-4612-b6c9-5c59da6c6634#id">http://www.bbc.co.uk/things/15eb702e-c630-4612-b6c9-5c59da6c6634#id</a>
Female		core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Gender	<a href="http://www.bbc.co.uk/things/3aa70400-7571-421c-97e6-01a8e762dfa1#id">http://www.bbc.co.uk/things/3aa70400-7571-421c-97e6-01a8e762dfa1#id</a>
Non-binary		core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Gender	<a href="http://www.bbc.co.uk/things/ed90b445-1f04-486a-9527-20b812aa203e#id">http://www.bbc.co.uk/things/ed90b445-1f04-486a-9527-20b812aa203e#id</a>
ABC1	Intended audience: situation	core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Situation	<a href="http://www.bbc.co.uk/things/8c91e0e6-1159-48b0-9dc9-73e5a2deda9e#id">http://www.bbc.co.uk/things/8c91e0e6-1159-48b0-9dc9-73e5a2deda9e#id</a>
C2DE		core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Situation	<a href="http://www.bbc.co.uk/things/2fc5c337-b0d7-404c-a852-fde8892c0b39#id">http://www.bbc.co.uk/things/2fc5c337-b0d7-404c-a852-fde8892c0b39#id</a>
Under-35	Intended audience: age	core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Age	<a href="http://www.bbc.co.uk/things/376c1218-21c8-496f-a66b-d025dc71dbce#id">http://www.bbc.co.uk/things/376c1218-21c8-496f-a66b-d025dc71dbce#id</a>
Over-35		core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Age	<a href="http://www.bbc.co.uk/things/e54e8a5a-9fde-41f4-accd-75177ce3c058#id">http://www.bbc.co.uk/things/e54e8a5a-9fde-41f4-accd-75177ce3c058#id</a>
Inform me	Intended audience: motivation	core:Thing <b>core:IntendedAudience</b> tagging:TagConcept audience:Motivation	<a href="http://www.bbc.co.uk/things/bf928ac3-b3bd-4d47-924e-cca1bdc29174#id">http://www.bbc.co.uk/things/bf928ac3-b3bd-4d47-924e-cca1bdc29174#id</a>
Educate me		core:Thing	<a href="http://www.bbc.co.uk/things/b1a660eb-">http://www.bbc.co.uk/things/b1a660eb-</a>

# Audience behaviours as metadata - “News needs”



# Building better tooling: Passport Control

Passport Control is the user interface  
to CRUD Passports

Useable as a standable tool or (better)  
embedded as an activity in an existing  
content production workflow



<https://passport-control.tools.bbc.co.uk/>

## Tags

Add tags to make your content discoverable. We recommend you add at least three tags.

### Add tags

### Review selected tags:

Tag types:

About **Cancer** ×  
Disease involving unregulate...

Editorial tone **Positive** ×  
Editorial tone: constructive, a...

Intended audience **Over 35s** ×  
Intended audience age range

# Improving metadata governance

Over the past decade we've evolved a vocabulary of 70,000 Things that can be used as tags

This makes it much harder for a human to tag a piece of content with (for example) Boris Johnson

Machine tagging can help, but really we need better structure and governance of our tags



BBC | jeremytarling | News | Sport | Weather

## THINGS BETA

Search results for **boris**

Only showing the first 20 results. F

**Boris**  
Japanese stoner rock band  
<http://www.bbc.co.uk/things/0500cb87-7ff4-4189-9ffd-1aeb90f47d90#id>

**Boris Yeltsin**  
Former Russian President  
<http://www.bbc.co.uk/things/235093c8-84a7-4a7e-b1b2-283724397d01#id>

**Boris Becker**  
German athlete  
<http://www.bbc.co.uk/things/5e56090a-ef99-4202-b116-72edacf6c277#id>

**Boris Dlugosch**  
Music Artist  
<http://www.bbc.co.uk/things/496783b9-f14f-4555-bdb9-a46a97f0abe9#id>

**Boris Brovtsyn**  
Music Artist  
<http://www.bbc.co.uk/things/83b31e67-e3b9-48c9-b58c-ab30e4707870#id>

**Boris Godunov**  
Opera by Modest Mussorgsky  
<http://www.bbc.co.uk/things/9a9a2fa6-2d93-4cde-99f9-64cd1388454f#id>

**Boris Christoff**

# Introducing machine tagging

## Starfruit, pace/valence and Vox

# First steps in automation: about and mentions

*Starfruit* - developed by BBC R&D

uses a standard machine learning approach to text classification (Multi-label Classification)

trained on a subset of the 70,000 tags that have been used at least 10 times to describe journalism content over the past few years

can be used to suggest tags to a journalist

# First steps in automation: Starfruit “aboutness”



Politics

## Queen's Speech: New laws on crime, health and the environment

1 hour ago 3166

**Brexit**



Queen outlines government's Brexit plan

Boris Johnson's government has set out "ambitious" policies on crime, health, the environment and Brexit in a Queen's Speech opposition parties have dismissed as a manifesto.

Plans for tougher sentences for violent offenders and legal targets for cutting plastic pollution are among the bills set out at Parliament's State Opening.

The BBC's Laura Kuenssberg said it was a "shopping list".

## Topics Results

Preferred Label	URI	Disambiguation Hint	Score	Types
Politics	<a href="http://www.bbc.co.uk/things/75612fa6-147c-4a43-97fa-fcf70d9cccd3#id">http://www.bbc.co.uk/things/75612fa6-147c-4a43-97fa-fcf70d9cccd3#id</a>	Governance of states	0.8580	tagging:TagConcept, core:Theme, core:Thing
Boris Johnson	<a href="http://www.bbc.co.uk/things/0b79dd85-35b5-40cd-b65e-cb152dd89fba#id">http://www.bbc.co.uk/things/0b79dd85-35b5-40cd-b65e-cb152dd89fba#id</a>	UK politician	0.7953	tagging:TagConcept, core:Person, core:Thing
Jeremy Corbyn	<a href="http://www.bbc.co.uk/things/71e97cc0-8bec-416b-82f0-a6bbb85d7f1f#id">http://www.bbc.co.uk/things/71e97cc0-8bec-416b-82f0-a6bbb85d7f1f#id</a>	UK politician	0.6565	tagging:TagConcept, core:Person, core:Thing
UK Parliament	<a href="http://www.bbc.co.uk/things/1d41fbc6-c505-4437-8139-7536c7cfc07e#id">http://www.bbc.co.uk/things/1d41fbc6-c505-4437-8139-7536c7cfc07e#id</a>	Legislative body of the UK	0.5871	tagging:TagConcept, core:Thing, core:Organisation
Unlawful Parliament suspension	<a href="http://www.bbc.co.uk/things/9f4b5957-ac41-4e5a-829d-36a173a4b1b9#id">http://www.bbc.co.uk/things/9f4b5957-ac41-4e5a-829d-36a173a4b1b9#id</a>	Suspension of Parliament in the run up to Brexit, 2019	0.5501	tagging:TagConcept, core:Thing, core:Event
Brexit	<a href="http://www.bbc.co.uk/things/f94293fd-ff83-400e-8bc0-89ec01930fce#id">http://www.bbc.co.uk/things/f94293fd-ff83-400e-8bc0-89ec01930fce#id</a>	UK departure from the EU	0.5337	core:Thing, tagging:TagConcept, core:Event



# First steps in automation: Starfruit “mentions”

BBC Home News More

## NEWS

Politics

### Queen's Speech: New laws on crime, health and the environment

1 hour ago 3166

Brexit



Queen outlines government's Brexit plan

Boris Johnson's government has set out "ambitious" policies on crime, health, the environment and Brexit in a Queen's Speech. Opposition parties have dismissed it as a manifesto.

Plans for tougher sentences for violent offenders and legal targets for cutting plastic pollution are among the bills set out at Parliament's State Opening.

The BBC's Laura Kuenssberg said it was a "shopping list".

Parliament	<a href="http://www.bbc.co.uk/things/02af43eb-7f41-4f06-9086-f52e6e46beb6#id">http://www.bbc.co.uk/things/02af43eb-7f41-4f06-9086-f52e6e46beb6#id</a>	Music Group	Parliament	0.3000	2	tagging:TagConcept, core:Thing
Country	<a href="http://www.bbc.co.uk/things/3a609713-8001-426f-a60b-a8cc4ee95e77#id">http://www.bbc.co.uk/things/3a609713-8001-426f-a60b-a8cc4ee95e77#id</a>	Distinct region	country	0.3000	1	tagging:TagConcept, core:Theme, core:Thing
Butter	<a href="http://www.bbc.co.uk/things/a44b7ad7-1211-492d-a014-a0d5a57f0c16#id">http://www.bbc.co.uk/things/a44b7ad7-1211-492d-a014-a0d5a57f0c16#id</a>	Dairy product	butter	0.3000	1	core:Theme, tagging:TagConcept, core:Thing
Bread	<a href="http://www.bbc.co.uk/things/04d84519-0db8-4b7c-805b-b1451a09398b#id">http://www.bbc.co.uk/things/04d84519-0db8-4b7c-805b-b1451a09398b#id</a>	Staple food	bread	0.2923	1	core:Thing, tagging:TagConcept, core:Theme
Shopping	<a href="http://www.bbc.co.uk/things/11ab5d9e-43db-492d-91a3-458bc84553fe#id">http://www.bbc.co.uk/things/11ab5d9e-43db-492d-91a3-458bc84553fe#id</a>	Buying goods	shopping	0.2890	1	core:Thing, core:Theme, tagging:TagConcept
Manifesto	<a href="http://www.bbc.co.uk/things/00a4152d-a172-4c79-bc51-f3fb1277cc9b#id">http://www.bbc.co.uk/things/00a4152d-a172-4c79-bc51-f3fb1277cc9b#id</a>	Published declaration of intentions	manifesto	0.2794	1	core:Theme, tagging:TagConcept, core:Thing
Healthcare	<a href="http://www.bbc.co.uk/things/27a675ec-1a68-4f13-b498-a77ee2bacbdc#id">http://www.bbc.co.uk/things/27a675ec-1a68-4f13-b498-a77ee2bacbdc#id</a>	Provision of medical care	health	0.2785	1	core:Theme, core:Thing, tagging:TagConcept
Exercise	<a href="http://www.bbc.co.uk/things/a73c80ca-eaa4-4ae6-9d55-b1600bb19eb4#id">http://www.bbc.co.uk/things/a73c80ca-eaa4-4ae6-9d55-b1600bb19eb4#id</a>	Physical activity	exercise	0.2710	1	core:Thing, tagging:TagConcept, core:Theme
Queen	<a href="http://www.bbc.co.uk/things/6d23f160-cbd6-4473-aa11-e3a1faa60cdf#id">http://www.bbc.co.uk/things/6d23f160-cbd6-4473-aa11-e3a1faa60cdf#id</a>	UK rock group	Queen	0.2578	2	tagging:TagConcept, core:Thing
Elderly people	<a href="http://www.bbc.co.uk/things/3892335e-43d8-489b-b0a1-9e2cf25ff486#id">http://www.bbc.co.uk/things/3892335e-43d8-489b-b0a1-9e2cf25ff486#id</a>	Senior citizens.,Stage of life	elderly	0.2517	1	core:Thing, tagging:TagConcept, core:Theme

“Our political editor said the PM was keen to focus on "bread and butter issues" like investment in schools and the NHS, or coming up with, at long last, a new way of funding care for the elderly.”

# First steps in automation: auto-tagging works for journalism

R&D's Starfruit is building confidence about machine classification in our editorial community

So far we are only using it for suggesting tags, but we are pushing for full auto-tagging

Content creators worry about automation of these editorial tasks, both for accuracy and for their job security

Last month, Sports Direct complained that its rival JD Sports' planned £90m takeover of Footasylum could reduce Mr Ashley's access to the top brands.

In the past, Adidas and Nike have preferred to work with JD Sports but Sports Direct has attempted to make inroads, appointing former Nike executive David Daly as chairman of the its board.

"Sports Direct has consistently aimed to provide the widest range of products at attractive prices and will continue to work constructively with all its suppliers to enhance its product offering for the benefit of consumers," the company said.

The BBC has asked Nike and Adidas for comment. In response to the Sunday Times story, Nike told the newspaper that it made adjustments to its sales channels from "time to time" to "optimise distribution".

---

## Related Topics

[Sports Direct](#)[Retailing](#)[Mike Ashley](#)[Companies](#)[Adidas](#)[Nike, Inc.](#)

---

# First steps in automation: editorial tone

BBC R&D are developing an automated editorial tone classifier for TV drama segments based on *valence* and *arousal*

“**Valence**, as used in psychology, especially in discussing emotions, means the intrinsic attractiveness/ "good"-ness (positive valence) or averseness/"bad"-ness (negative valence) of an event, object, or situation”

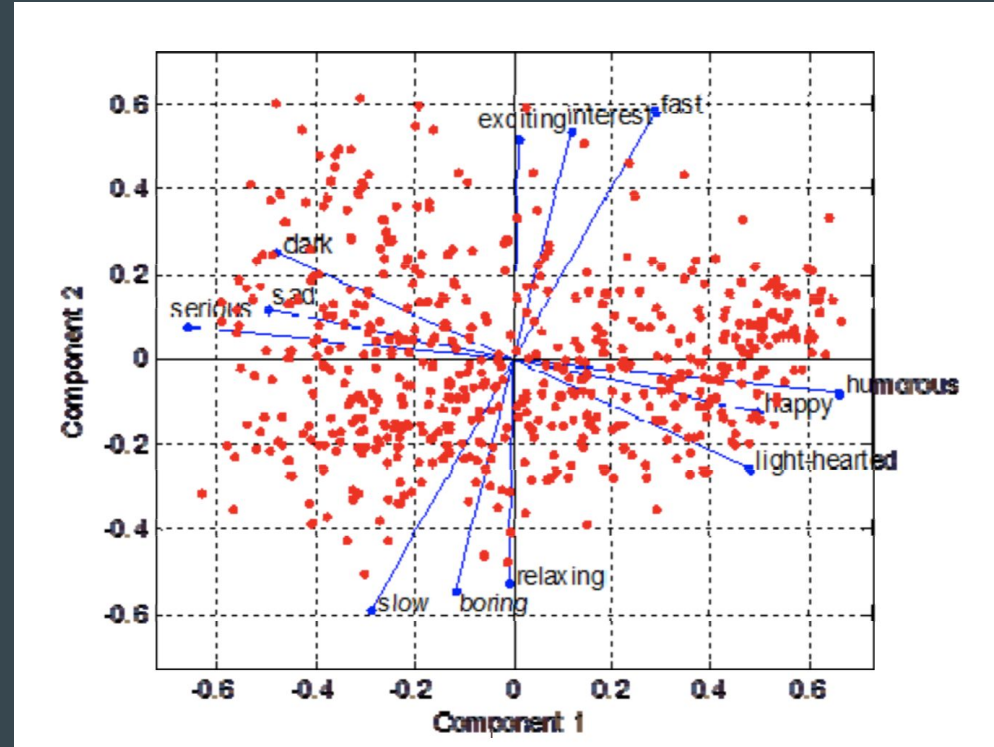
“**Arousal** involves activation of the ascending reticular activating system in the brain, which mediates wakefulness, the autonomic nervous system, and the endocrine system, leading to increased heart rate and blood pressure and a condition of sensory alertness, mobility, and readiness to respond.”

# First steps in automation: editorial tone

initial model built on human classification of comedy and drama clips

small controlled vocabulary representing the vectors - “slow/fast”, “exciting/boring”, “happy/sad”, etc.

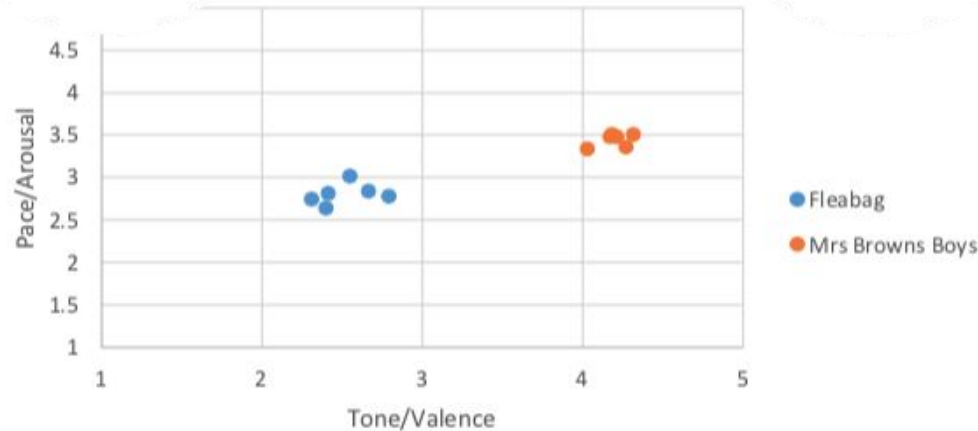
machine classification based on existing model then retained



# First steps in automation: editorial tone



Comedy



# Metadata quality measurement

## Completeness, Accuracy, Provenance

# Metadata quality measurement

Metadata quality measurement  
is well documented in library  
and information science

How might we develop a  
quality metric for the BBC's  
descriptive content metadata?

*completeness*

relevancy

logical consistency  
and coherence

conformance  
to expectation

*accuracy*

timeliness

accessibility

*provenance*

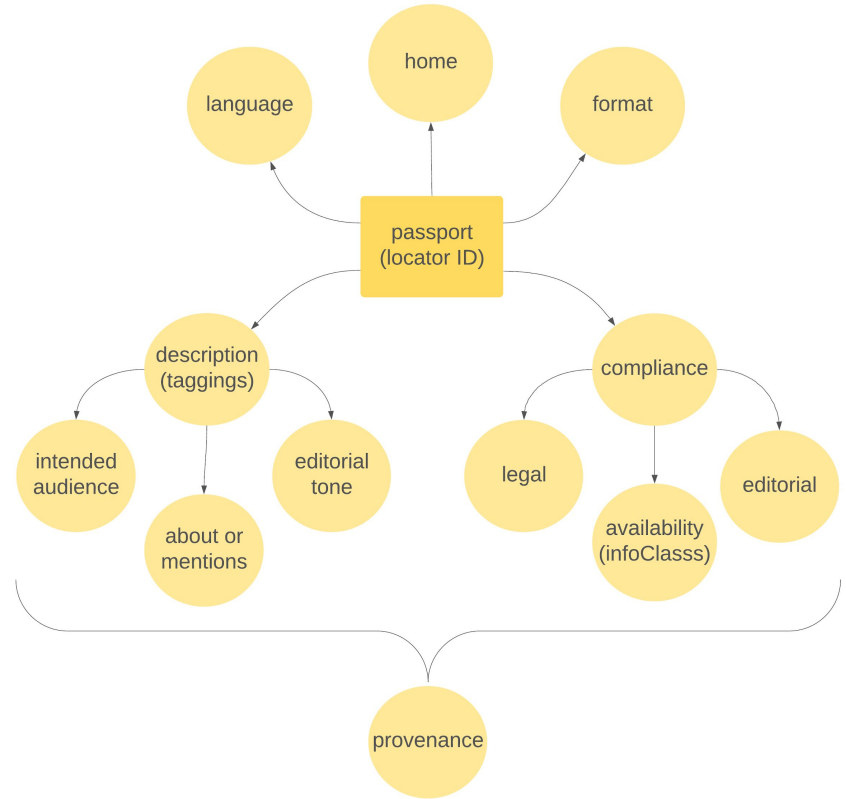
fitness for purpose

# Metadata quality measurement: Completeness

Given an agreed set of metadata properties, is there any metadata present for these fields?

Are all fields equal in importance?

Across all domains/products?





# Metadata quality measurement: Completeness

property	News articles	Sounds music mixes
about	90%	10%
editorial tone	10%	80%
Intended audience	20%	30%

# Metadata quality measurement: Provenance

Initially we are dealing with two provenances, human vs machine

But...

- Human tagging quality will be domain-specific and affected by experience and knowledge
- Machine tagging quality will be model-specific and only as good as the quality and timeliness of the training set

# Metadata quality measurement: Provenance

source	News articles	Sounds music mixes
Journalist	90%	40%
Music producer	50%	90%
Starfruit (model 94)	75%	25%

# Metadata quality measurement: Accuracy

Accuracy is calculated as a function of provenance and completeness at the individual property level

“User X with provenance score 25% for news articles says

<https://www.bbc.co.uk/news/uk-england-derbyshire-51611555> is

→ about → TripAdvisor, NatWest bank, Ilkeston”

“User Y with provenance score 75% for news articles says

<https://www.bbc.co.uk/news/uk-england-derbyshire-51611555> is

→ about → Ilkeston

→ tone → funny”

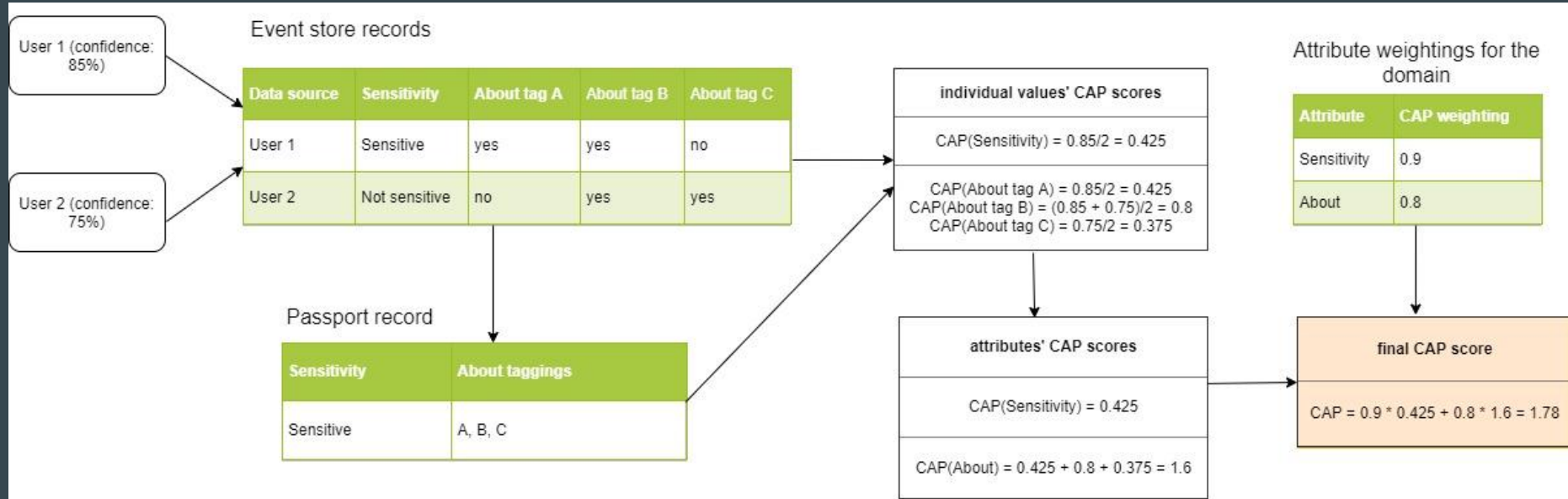
# Metadata quality measurement: CAP scores

We take **Curation**, **Accuracy** and **Provenance** together to generate a 'CAP score' for a given set of metadata assertions

$$\text{CAP} = (\text{threshold} + \text{SUM}(\text{relevant assertion confidence scores})) / (\text{number of relevant assertions} + 1)$$

(relevant assertions = assertions that support the value, e.g. assertions that would add the same About tag A to the metadata set)

# Metadata quality measurement: CAP scores



# Thanks for listening!

Links:

BBC Things (tags): <https://www.bbc.co.uk/things/>

BBC Ontologies: <https://www.bbc.co.uk/ontologies/>

Starfruit: <http://starfruit.virt.ch.bbc.co.uk/>

Vox: <http://vox.virt.ch.bbc.co.uk/>

[@jeremytarling](https://twitter.com/jeremytarling)

[jeremy.tarling@bbc.co.uk](mailto:jeremy.tarling@bbc.co.uk)