BBC Linked Data and the Semantic Publishing Benchmark

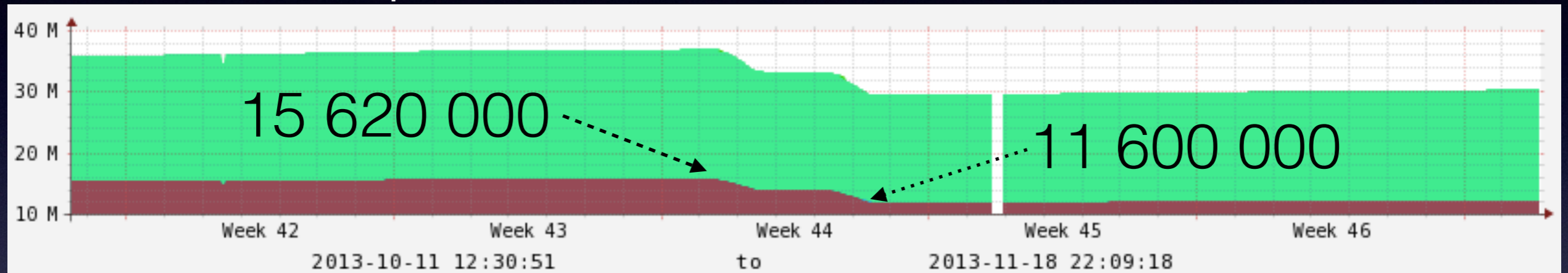Johan Hjerling

Linked Data Platform

BBC Linked Data and the Semantic Publishing Benchmark

- BBC Linked Data Platform metrics

- Linked Data Platform compared to the Semantic Publishing Benchmark

- Questions
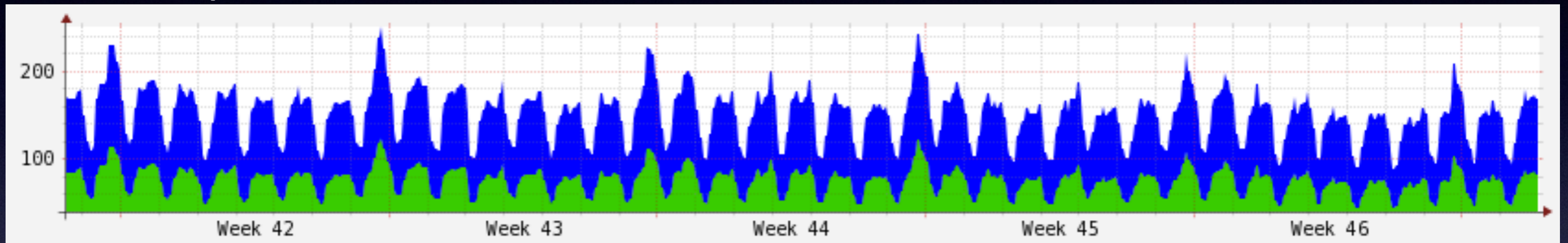
Metrics

Number of Triples

15 620 000

11 600 000

Current number of triples	12,100,035

Creative Works	47,964
Media Assets	2,391,996
Thing Graphs	21,191

Metrics

Reads per second

Average	around 150 per second
Spike	around 354 per second
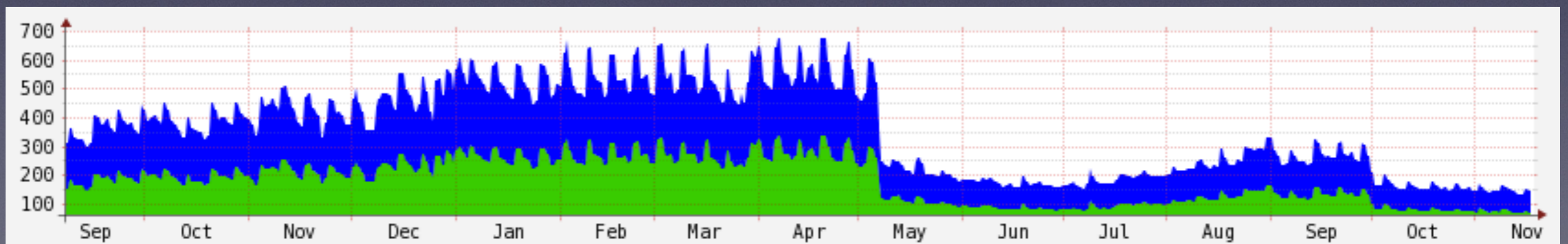
Metrics

Writes per second
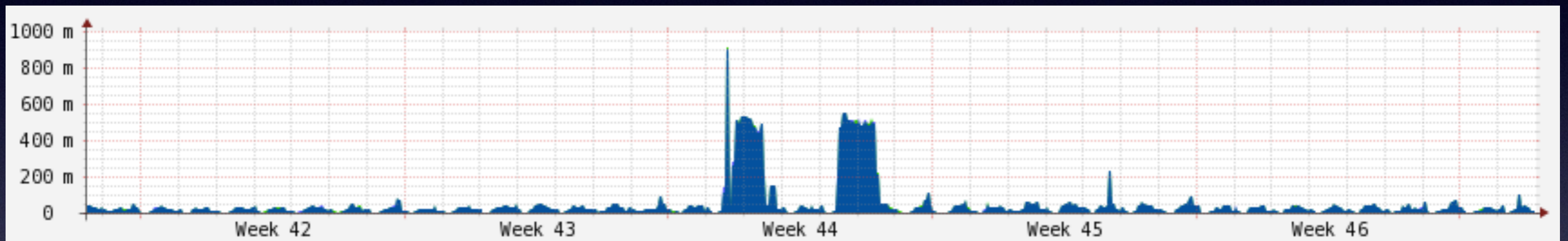
Average	around 0.047 per second	2.8 per minute
Spike	around 1.2 per second	72 per minute

Brief History

DSP
2010
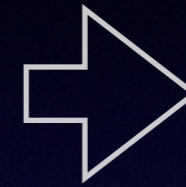
- First Dynamic Content
- 2010 World Cup
- Asset Ontology
- Currently drives the /sport pages and apps
- SPARQL 1.0
- Turtle, N3

Olympics
2012

- Libraries built as part of development
- 2nd iteration of DSP API

LDP
2012

- Built with the olympics libraries
- Instance data management
- Creative Work ontology
- SPARQL 1.1
- JSON-LD

Migration
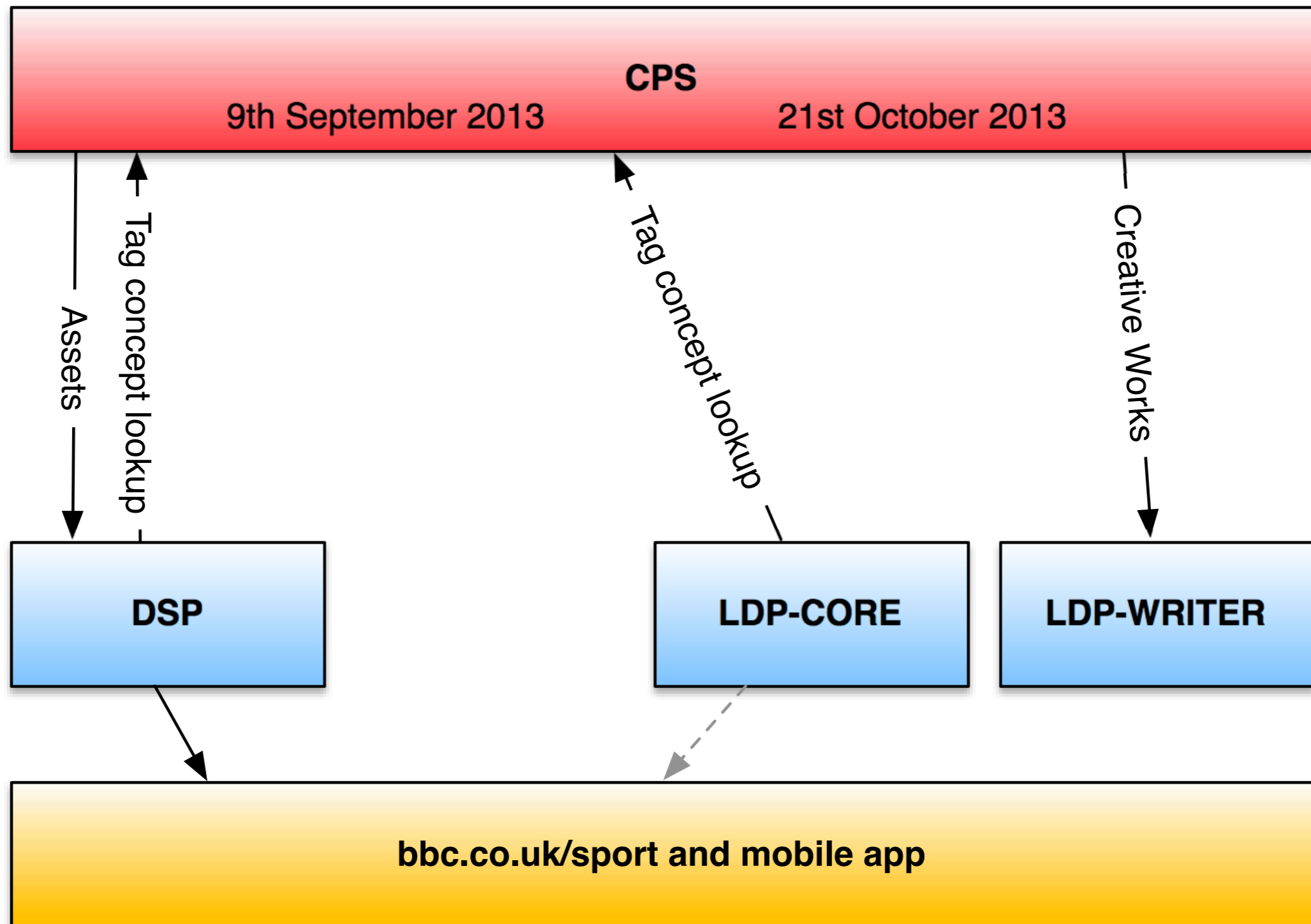
Instance data management

- Thing-Graphs

  - Singular units of data (*Person, Place, Event*)

  - Editorial data (*label, disambiguation hint*)

- Datasets

  - Grouped data (*Teams, Competitions*)

  - Relational data (*Team competes in*)

- Ontologies (diff)

Split Architecture

Websites & Apps

/sport /news /music

BBC Internal

CreativeWorks CreativeWorks CreativeWorks

REST Client

Linked Data Tagger

Linked Data Manager

CPS

Tag Concepts

Linked Data Core API

CreativeWork

CONSTRUCT ?cWork…

Instance data

Linked Data Writer API → Linked Data Active MQ → Triple Store
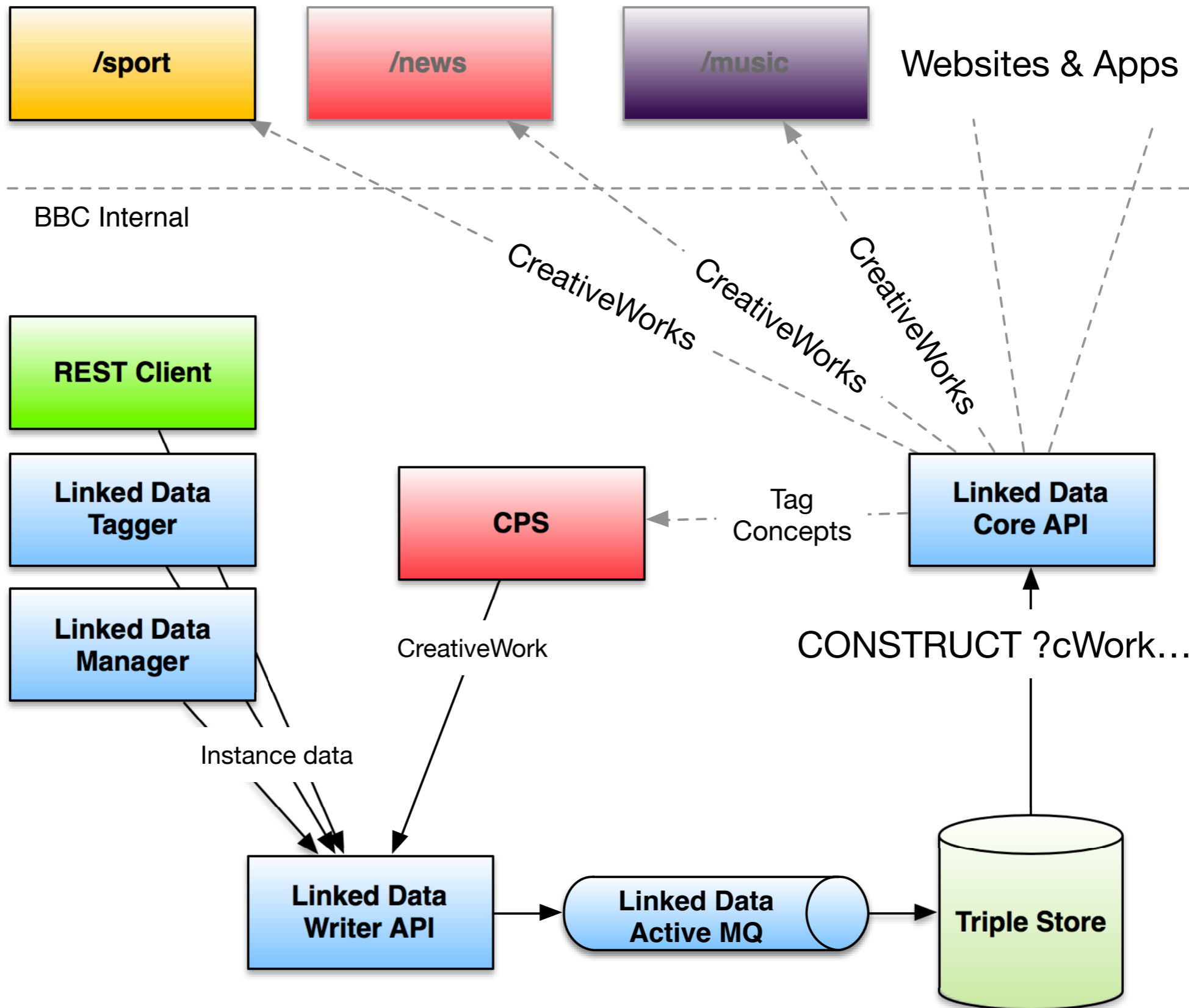
```
CONSTRUCT {
?creativeWork a cwork:CreativeWork ;
   a ?type ;
   cwork:title ?title ;
   cwork:shortTitle ?shortTitle ;
   cwork:about ?about ;
   cwork:mentions ?mentions ;
   cwork:dateCreated ?created ;
   cwork:dateModified ?modified ;
   cwork:description ?description ;
   cwork:category ?category ;
   cms:locator ?locator ;
   cwork:primaryFormat ?primaryFormat ;
   bbc:primaryContentOf ?webDocument .
  ?webDocument bbc:platform ?platform .
  ?webDocument bbc:webDocumentType ?webDocumentType .
  ?about a ?aboutType;
    rdfs:label ?aboutLabel ;
    core:shortLabel ?aboutShortLabel ;
    core:preferredLabel ?aboutPreferredLabel ;
    core:sameAs ?aboutSameAs ;
    core:primaryTopicOf ?aboutPrimaryTopicOf .
  ?aboutPrimaryTopicOf bbc:webDocumentCategory ?webDocumentCategory
  ?mentions a ?mentionsType ;
    rdfs:label ?mentionsLabel ;
    core:shortLabel ?mentionsShortLabel ;
    core:preferredLabel ?mentionsPreferredLabel ;
    core:sameAs ?mentionsSameAs .
  ?creativeWork cwork:thumbnail ?thumbnail .
  ?thumbnail a cwork:Thumbnail ;
    cwork:altText ?thumbnailAltText ;
    cwork:thumbnailType ?thumbnailType .
}
```

```
WHERE {
  {
  SELECT DISTINCT ?creativeWork
  WHERE {
    {{#about}}
      FILTER (?about = <{{about}}>) .
      ?creativeWork cwork:about ?about .
    {{/about}}

    {{#aboutTagType}}
      ?creativeWork cwork:about ?about .
      ?about a <{{aboutTagType}}> .
      {{#aboutTagPredicate}}
        {{#aboutTagObject}}
          ?about <{{aboutTagPredicate}}> <{{aboutTagObject}}> .
        {{/aboutTagObject}}
      {{/aboutTagPredicate}}
    {{/aboutTagType}}

    {{#aboutTagSearch}}
      ?creativeWork cwork:about ?about .
        ?about luc:ldpLuceneIndex "{{aboutTagSearch}}" .
    {{/aboutTagSearch}}

    {{#format}}
      {{#-first}}?creativeWork cwork:primaryFormat ?format .{{/-first}}
                {{#-first}}FILTER {{{/-first}}?format = <{{ . }}> {{^-last}} || {{/-last}}{{#-last}}} .{{/-last}}
    {{/format}}
    {{#mentions}}
      FILTER (?mentions = <{{mentions}}>) .
      ?creativeWork cwork:mentions ?mentions .
    {{/mentions}}
    {{#primaryContentOf}}
      ?creativeWork bbc:primaryContentOf ?webDocument .
      FILTER (?webDocument = <{{primaryContentOf}}>) .
    {{/primaryContentOf}}
    {{#locator}}
      FILTER (?locator = <{{locator}}>) .
      ?creativeWork cms:locator ?locator .
    {{/locator}}
    {{#audience}}
      OPTIONAL { ?creativeWork cwork:audience ?audience . }
      FILTER (?audience = <{{audience}}> || NOT EXISTS { ?creativeWork cwork:audience ?audience } ) .
    {{/audience}}
    {{#category}}
      FILTER (?category = <{{category}}>) .
      ?creativeWork cwork:category ?category .
    {{/category}}
    {{#within}}
      ?creativeWork cwork:tag ?location .
      ?location a geoname:Feature ;
        omgeo:within( {{within}} ) .
    {{/within}}
    {{#type}}
      FILTER (?type = <{{type}}>) .
      ?creativeWork a ?type .
    {{/type}}
    {{#since}}
      FILTER (?modified > "{{since}}"^^xsd:dateTime).
    {{/since}}
    {{#before}}
      FILTER (?modified < "{{before}}"^^xsd:dateTime).
    {{/before}}

    {{#orThings}}
      {{#-first}}
        OPTIONAL { ?creativeWork a ?type . }
        OPTIONAL { ?creativeWork cwork:category ?category . }
      {{/-first}}
      {{#-first}}FILTER {{{/-first}}{{#type}}?type = <{{type}}>{{/type}}{{#category}}{{#type}} && {{/type}}?category = <{{category}}>{{/category}}{{^-last}} || {{/-last}}{{#-last}}} .{{/-last}}
    {{/orThings}}

    ?creativeWork a cwork:CreativeWork ;
      cwork:dateModified ?modified .

    {{#excludeTypes}}
      {{#-first}}FILTER NOT EXISTS { {{/-first}}
        ?creativeWork a <{{ . }}>
      {{#-last}} } {{/-last}}
    {{/excludeTypes}}
  }
  ORDER BY DESC(?modified)
  {{#limit}}LIMIT {{limit}}{{/limit}}
  {{#offset}}OFFSET {{offset}}{{/offset}}
  }
  ?creativeWork a cwork:CreativeWork ;
    cwork:dateModified ?modified .
}
UNION {
  ?creativeWork a ?type .
  ?type rdfs:subClassOf cwork:CreativeWork .
}
UNION { ?creativeWork cwork:title ?title . }
UNION { ?creativeWork cwork:shortTitle ?shortTitle . }
UNION { ?creativeWork cwork:description ?description . }
UNION { ?creativeWork cms:locator ?locator . }
UNION { ?creativeWork cwork:category ?category . }
UNION {
  ?creativeWork cwork:about ?about .
  OPTIONAL { ?about rdfs:label ?aboutLabel . }
  OPTIONAL { ?about core:shortLabel ?aboutShortLabel . }
  OPTIONAL { ?about core:preferredLabel ?aboutPreferredLabel . }
  OPTIONAL { ?about core:sameAs ?aboutSameAs . }
  OPTIONAL {
    ?about core:primaryTopicOf ?aboutPrimaryTopicOf .
    OPTIONAL { ?aboutPrimaryTopicOf bbc:webDocumentCategory ?webDocumentCategory . }
  }
  OPTIONAL {
    ?about a ?aboutType .
    FILTER(
      ?aboutType != <http://www.w3.org/2002/07/owl#Thing>
      && ?aboutType != <http://www.w3.org/2000/01/rdf-schema#Resource>
      && ?aboutType != <http://purl.org/NET/c4dm/event.owl#Event>
      && ?aboutType != <http://xmlns.com/foaf/0.1/Organization>
      && ?aboutType != <http://xmlns.com/foaf/0.1/Agent>
      && ?aboutType != <http://www.geonames.org/ontology#Feature>
      && ?aboutType != <http://www.bbc.co.uk/ontologies/tagging/TagConcept>
      && ?aboutType != <http://www.bbc.co.uk/ontologies/tag/TagConcept>
      && ?aboutType != <http://www.bbc.co.uk/ontologies/domain/Thing>
      && ?aboutType != <http://www.bbc.co.uk/ontologies/cms/ManagedThing>
    )
  }
}
UNION {
  ?creativeWork cwork:mentions ?mentions .
  OPTIONAL { ?mentions rdfs:label ?mentionsLabel . }
  OPTIONAL { ?mentions core:shortLabel ?mentionsShortLabel . }
  OPTIONAL { ?mentions core:preferredLabel ?mentionsPreferredLabel . }
  OPTIONAL { ?mentions core:sameAs ?mentionsSameAs . }
  OPTIONAL {
    ?mentions a ?mentionsType .
    ?mentionsType rdfs:subClassOf core:Thing .
  }
}
UNION { ?creativeWork cwork:dateCreated ?created . }
UNION { ?creativeWork cwork:primaryFormat ?primaryFormat . }
UNION {
  ?webDocument bbc:primaryContent ?creativeWork .
  OPTIONAL { ?webDocument bbc:platform ?platform . }
  OPTIONAL { ?webDocument bbc:webDocumentType ?webDocumentType . }
}
UNION {
  ?creativeWork bbc:primaryContentOf ?webDocument .
  OPTIONAL { ?webDocument bbc:platform ?platform . }
  OPTIONAL { ?webDocument bbc:webDocumentType ?webDocumentType . }
}
UNION {
  ?creativeWork cwork:thumbnail ?thumbnail .
  OPTIONAL { ?thumbnail cwork:altText ?thumbnailAltText . }
  OPTIONAL { ?thumbnail cwork:thumbnailType ?thumbnailType . }
}
}
```

Show Query

Questions?

Johan Hjerling - BBC
johan.hjerling@bbc.co.uk