

# Data Standards & APIs



**Dr David Tarrant**

@davetaz

theODI.org



# Aim

Enable you to improve the use of data standards and APIs within the BBC and BBC services.

# Contents:

1. Data and data quality
2. Creating and using data standards
3. APIs and the web of data
4. Standards and APIs @ BBC

# The team



**Dr Jeni  
Tennison**  
CEO



**Sir Nigel  
Shadbolt**  
Chairman



**Sir Tim  
Berners-Lee**  
President

Founded in 2012, the Open Data Institute (ODI) is an international, independent and not-for-profit organisation based in London, UK.

# — Me

12+ years experience in Open Data



---

Dr David  
Tarrant  
Learning &  
Technology

Established first degree level module in Open Data  
at the University of Southampton

Part of the team that established the ODI

Have since helped transform governments and  
unlock over \$15m for startups

# Our mission

We work with companies and governments to build an open, trustworthy data ecosystem.

# Our vision

We want people,  
organisations and  
communities to  
use data to make  
better decisions  
and be protected  
from any harmful  
impacts.

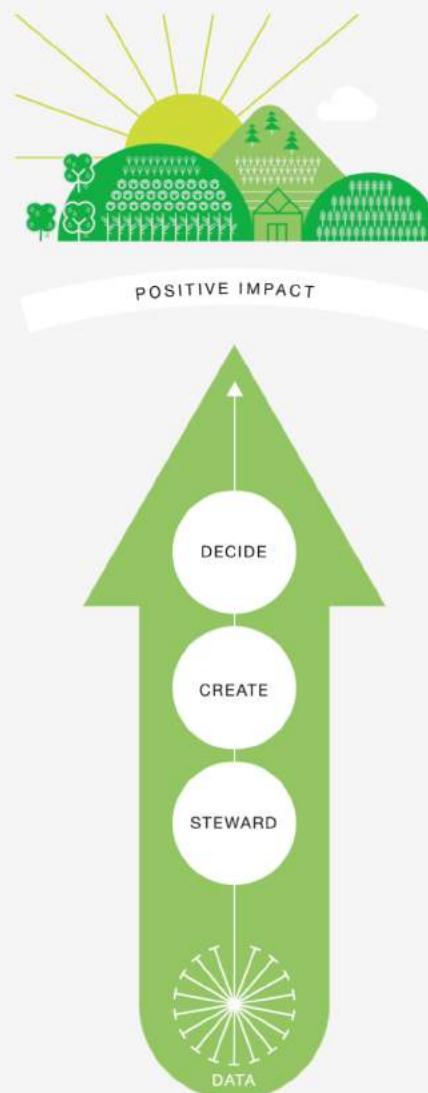
# Our theory of change

**How value is  
created from data**



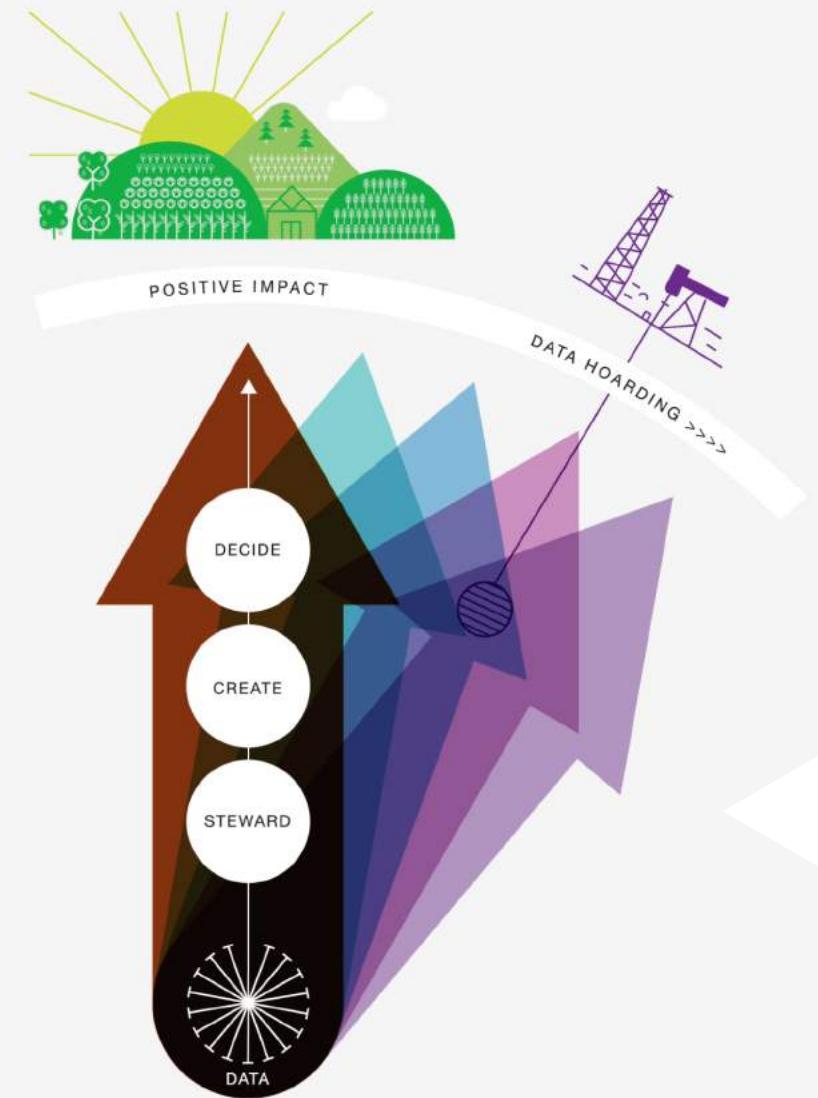
# Scenario:

What  
happens  
when we  
hoard data –  
the oil field



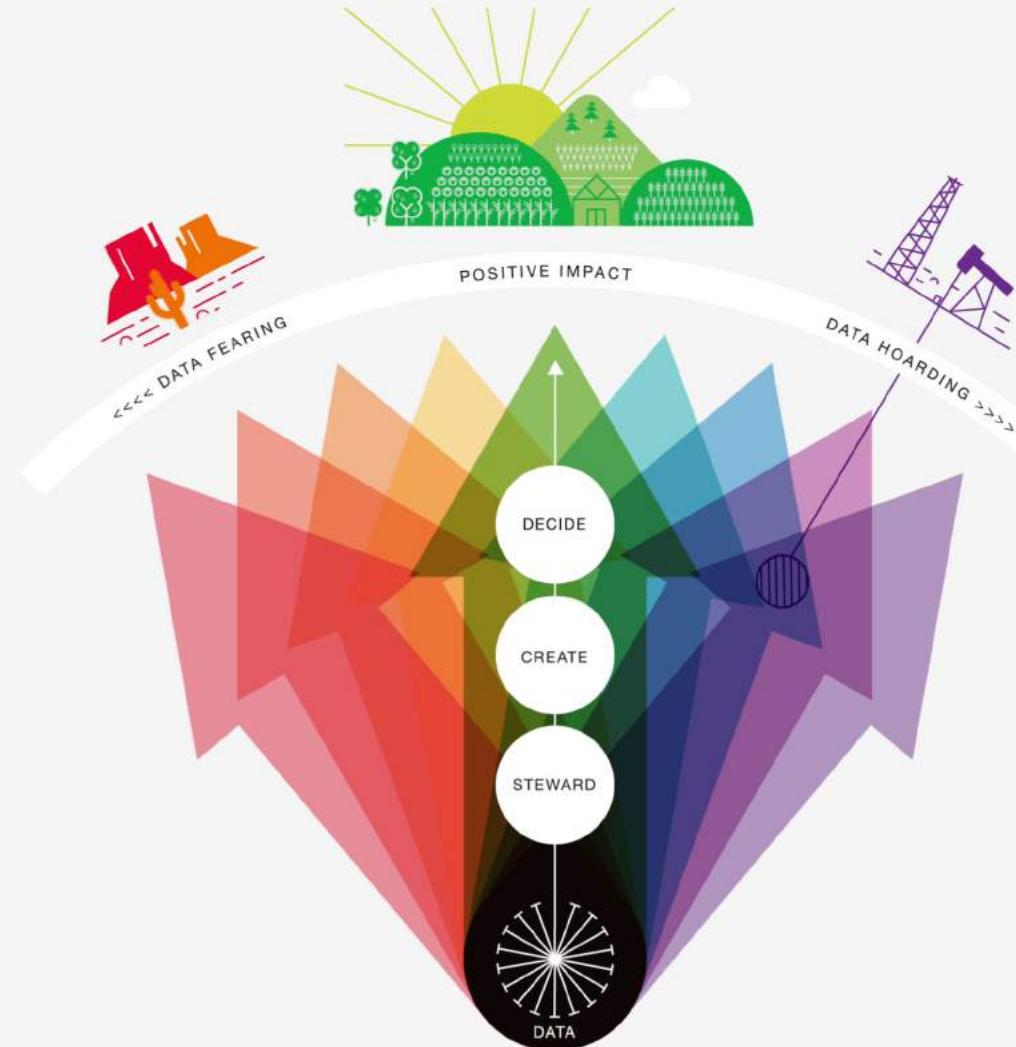
# Scenario:

What  
happens  
when we  
fear data –  
the wasteland



## Our theory of change

We are one of many organisations working towards a good balance between encouraging and restricting how data is collected and used.



---

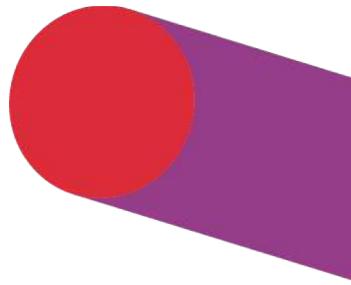
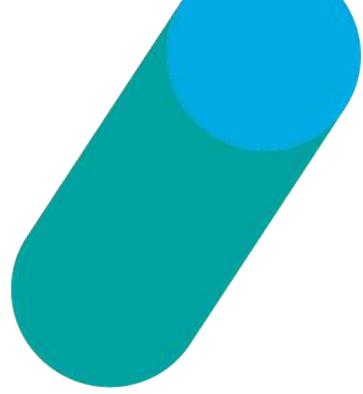
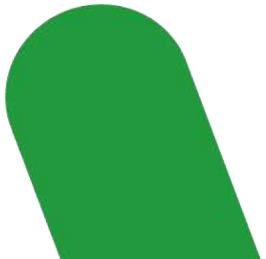
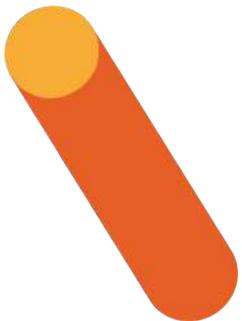
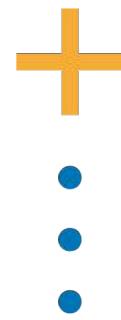
## Exercise (15 mins)

Download [http://bit.ly/tz\\_source](http://bit.ly/tz_source)

- 1) Identify as many quality problems as you can with this dataset.
- 2) Implement a new standard for this data

Did anyone search for other related datasets?

# What can we standardise?



# We can standardise...



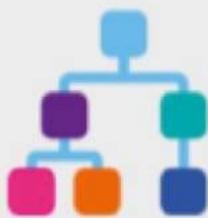
Models



Words



Identifiers



Taxonomies



File formats



Schemas



Data transfer



Units and measures



Data types

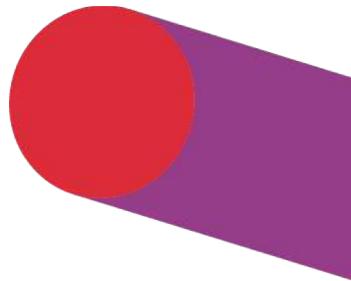
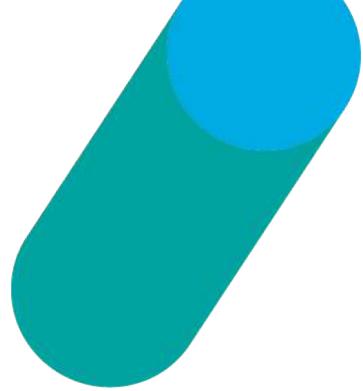
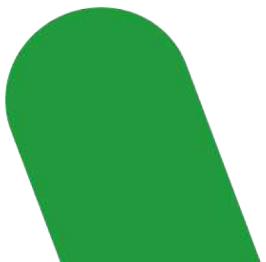
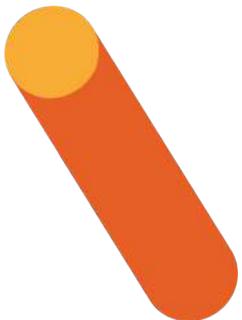
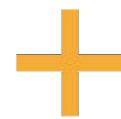


Code of practise



How we collect data

# What is a standard?



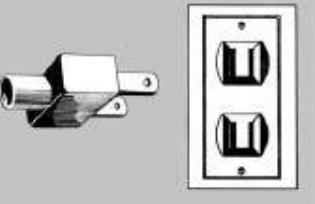
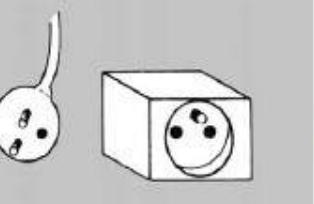
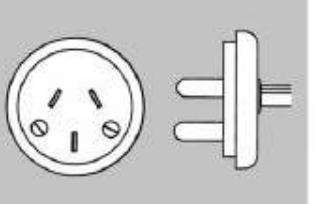
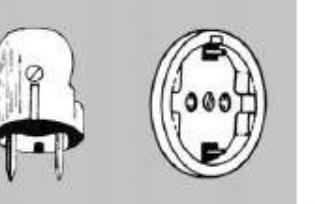
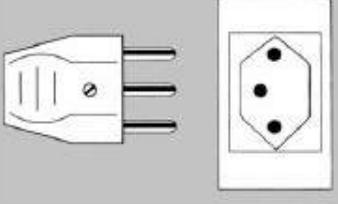
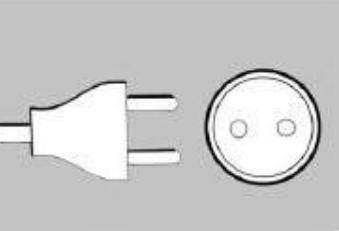
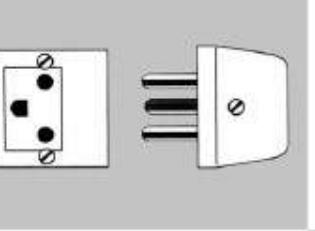
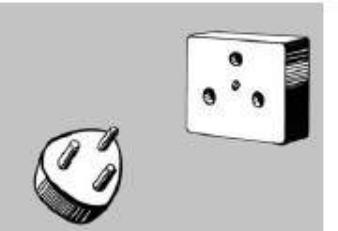
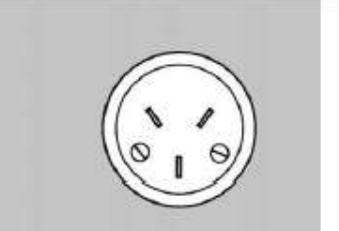
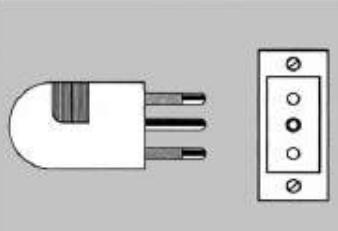
---

## Standard

A required or agreed level of quality or attainment  
– OED

Standards are documented, reusable agreements that affect us everyday - from addressing a letter, to changing a lightbulb, and looking up train times.

## Why are standards important?

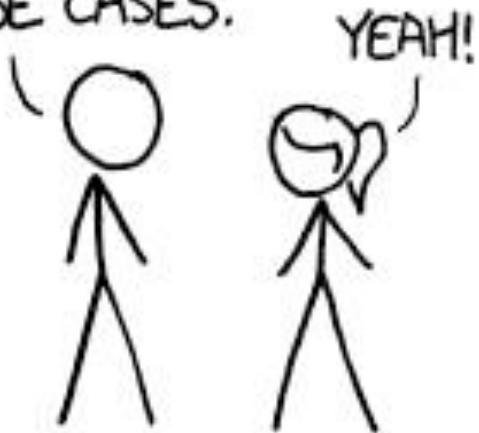
	Type <b>A</b> Flat blade attachment plug		Type <b>E</b> Round pin plug and receptacle with male grounding pin		Type <b>I</b> Oblique flat blade with ground
Type <b>B</b> Flat blades with round grounding pin		Type <b>F</b> “Schuko” plug and receptacle with side grounding contacts		Type <b>J</b> Round pins with ground	
	Type <b>C</b> Round pin attachment plug		Type <b>G</b> Rectangular blade plug		Type <b>K</b> Round pins with ground
Type <b>D</b> Round pins with ground		Type <b>H</b> Oblique flat blades with ground		Type <b>L</b> Round pins with ground	

## HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:  
THERE ARE  
14 COMPETING  
STANDARDS.

14?! RIDICULOUS!  
WE NEED TO DEVELOP  
ONE UNIVERSAL STANDARD  
THAT COVERS EVERYONE'S  
USE CASES.

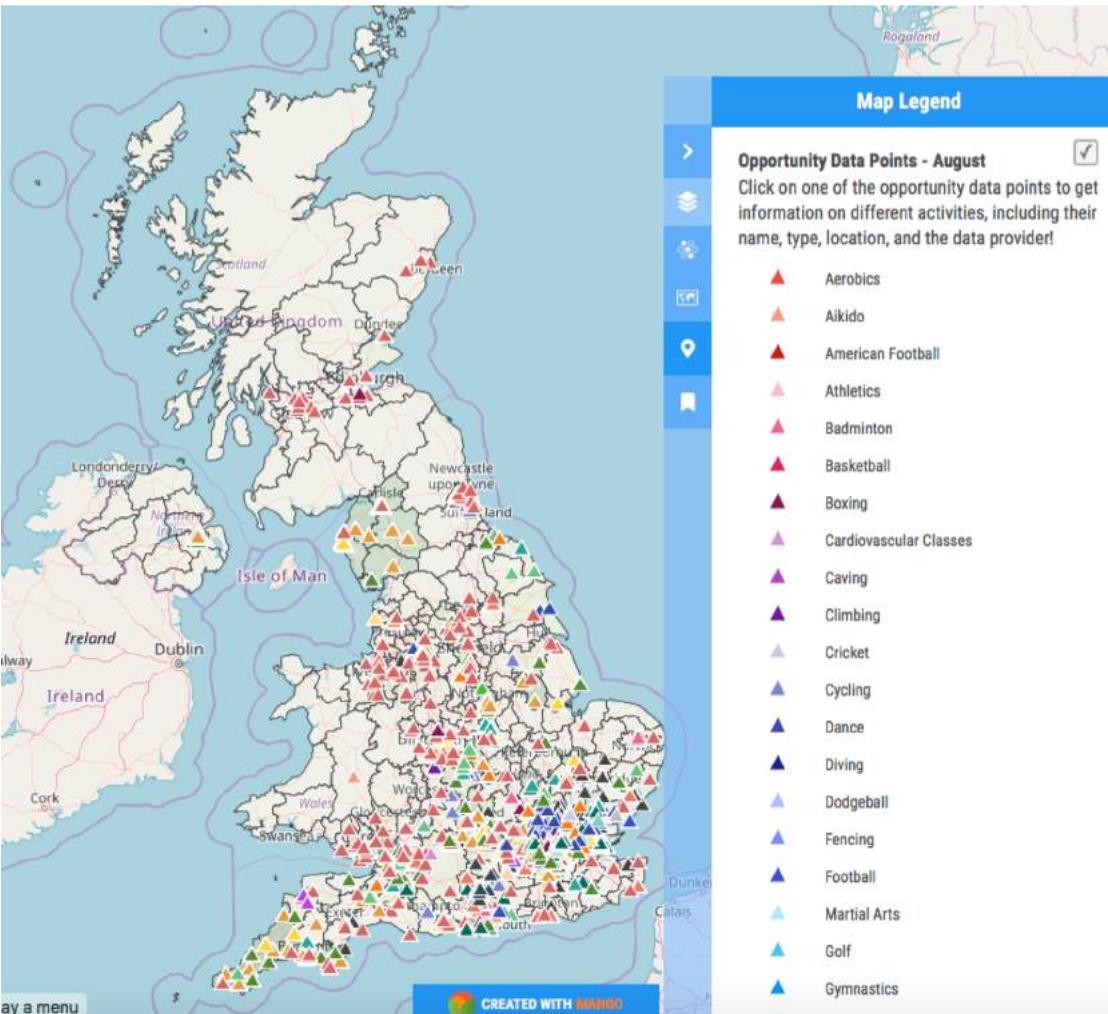


SOON:

SITUATION:  
THERE ARE  
15 COMPETING  
STANDARDS.

Fortunately, the charging one has been solved now that we've all standardized on mini-USB. Or is it micro-USB? S\*\*t!.

## Applications of Data Standards and APIs



Using data to help  
people get active



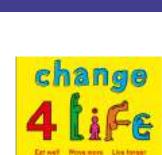
## Helping get people active



### Motivating



The Richmond Group of Charities



CHANGE ATTITUDES. FIND A CURE. JOIN US.

### Creating Opportunities



## User experience in different sectors



I want to go on holiday!



Loads of choice of how to book



Great search tools - easy to search through availability



Booked, paid and off I go!

vs



I want to be active!



Limited choice.  
Hard to find what I'm looking for.

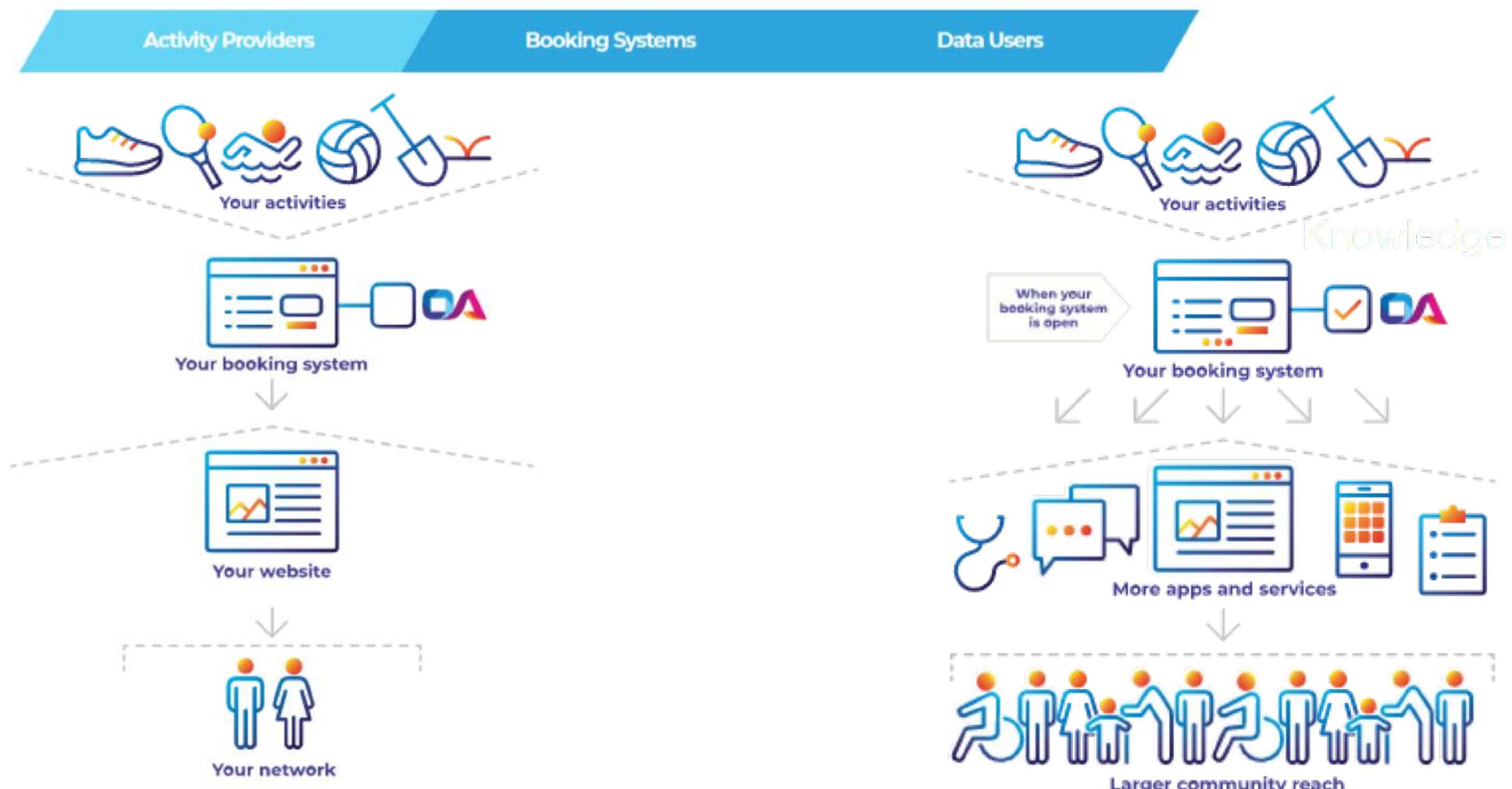


Nowhere to easily search what's available

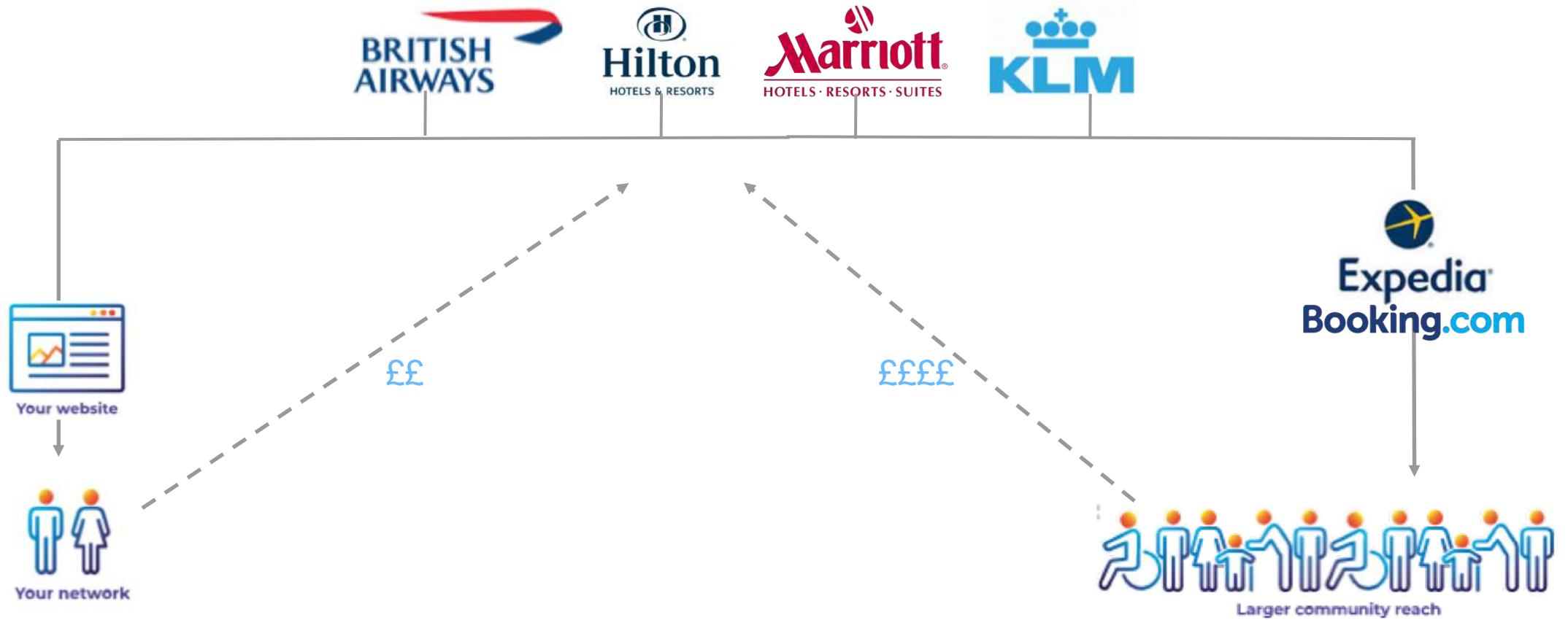


Give up.  
Remain inactive.

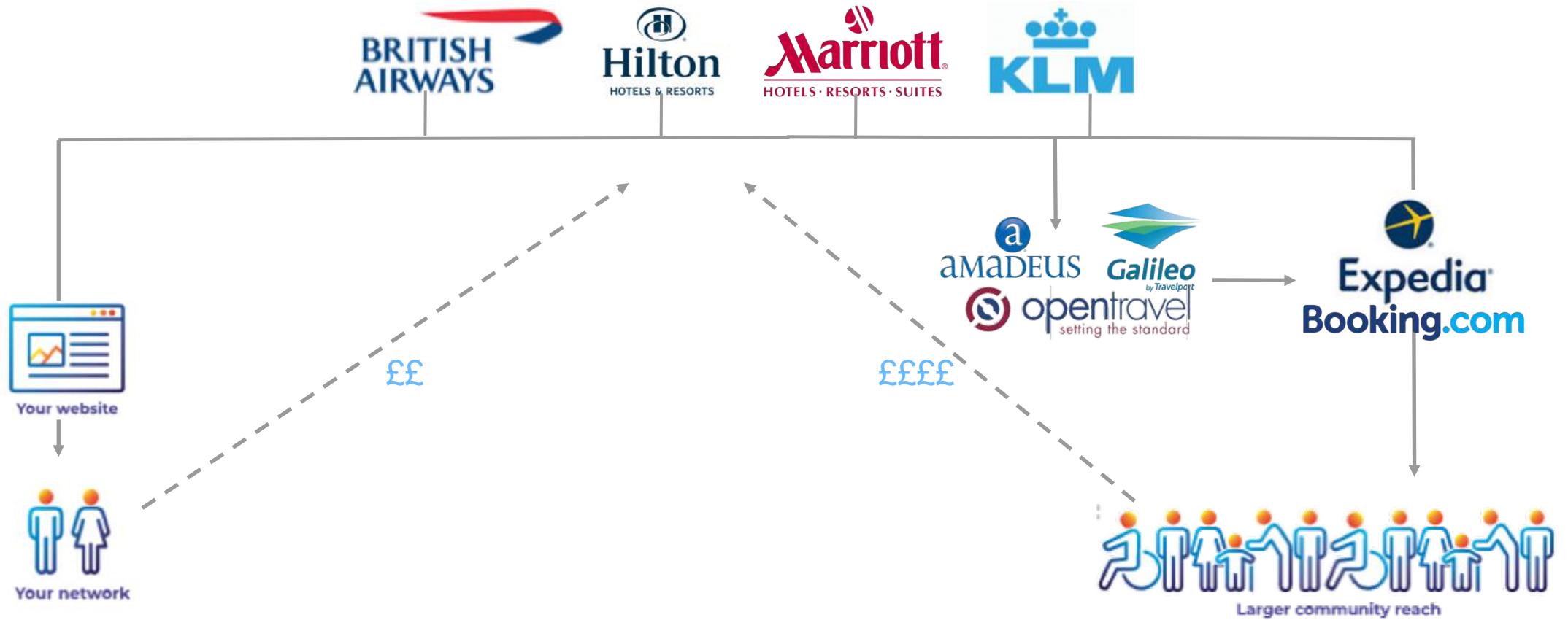
# OpenActive: How it works



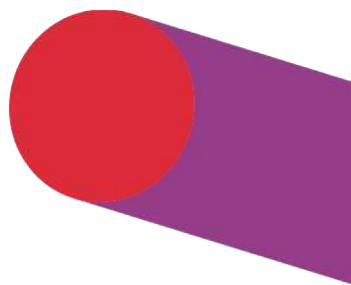
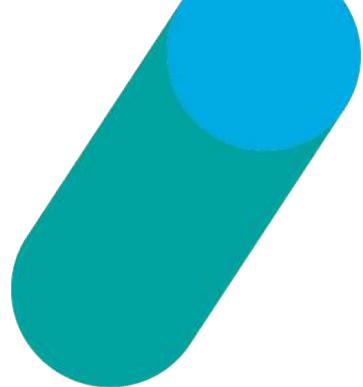
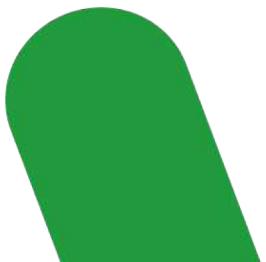
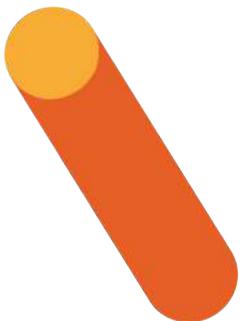
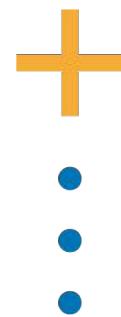
## Other ecosystems



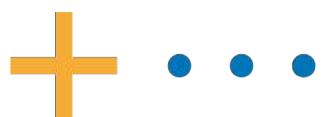
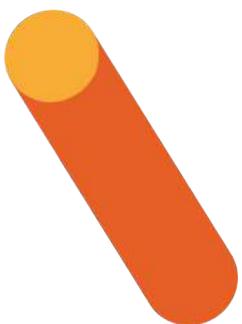
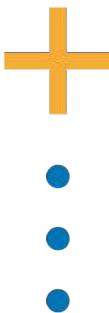
## Other ecosystems



# Creating a sporting opportunity...



Opportunity data is data  
about where and when  
sport and physical  
activities happen.





# Create

Create your own bit of  
**opportunity data?**

[bit.ly/openactive](http://bit.ly/openactive)

## Opportunity data in detail

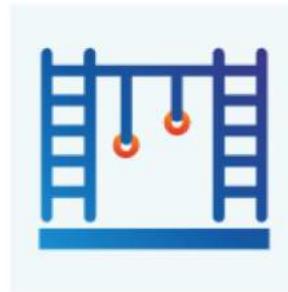
### Opportunity types



Events



Sessions



Facilities



Leagues and Ringers



Courses



Online Classes

### Properties of an opportunity



Time



Prices



Locations



Availability



Descriptions



Images



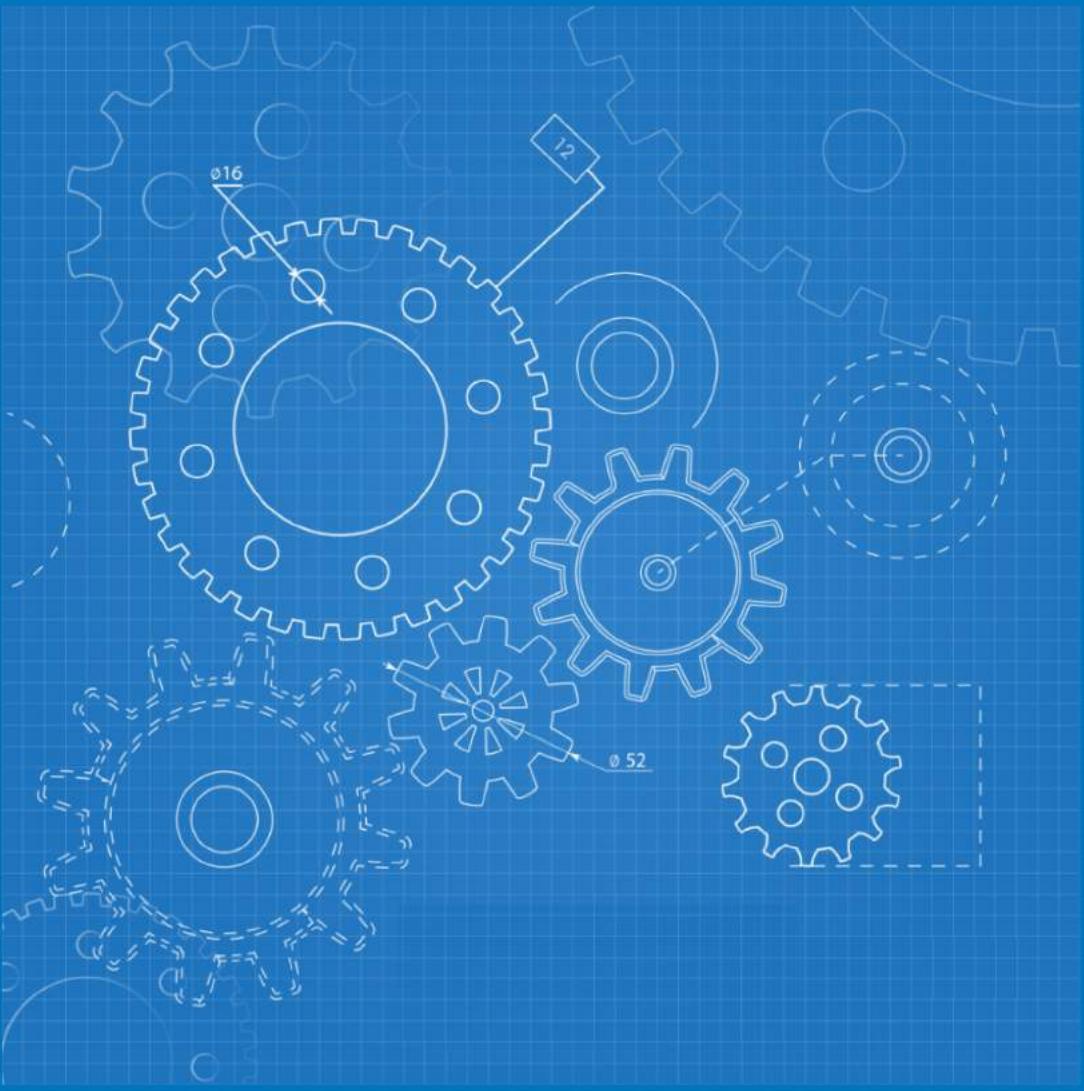
Gender  
Restriction



Age Range



Difficulty  
Level



Freepik [CC-BY]

# What is a schema?

A schema is a blueprint for data that defines a set of integrity constraints and rules relating to the structure and contents of a data resource.

The schema will define a number of key things:

- + Column/Key titles in the data
- + Value types
- + Value constraints

## Examples

Words - [PoliceUK crime categories](#)

Models – [BBC concepts](#)

Identifiers – [MusicBrainz](#)

Taxonomies – [Trade Tariff](#)

File formats and structures – e.g. CSV

Schema (format structure) – e.g. JSON

Data types – e.g. date

Units & measures – e.g. Celsius

Data transfer – e.g. API

Data collection & Code of practice – e.g.  
weather station quality

**1**

### **Respectful cooperation between standards organizations**

Whereby each respects the autonomy, integrity, processes and intellectual property rules of the others

**2**

### **Adherence to the fundamental parameters of standards development**

Including due process, broad consensus, transparency, balance and universal openness

**3**

### **Collective empowerment to strive to develop standards that are chosen & defined based on technical merit**

- As judged by the contributed expertise of an open and global experts community
- Provide global interoperability, scalability, stability, and resiliency
- Enable global competition
- Serve as building blocks for further innovation
- Contribute to the creation of benefit for humanity

**4**

### **Availability of standards specifications**

- Made globally accessible to all for implementation & deployment
- Defined procedures to develop specifications that can be implemented under fair terms
- Ensuring a broad affordability of the outcome of the standardization process (openness of input & output)

**5**

### **Voluntary adoption by the standards market**

The success of a standard is determined by the market.

Reference: [www.open-stand.org/principles](http://www.open-stand.org/principles)



---

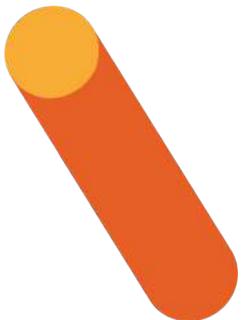
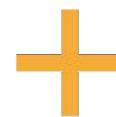
## Developing standards (if you have to)

- Collaboration, not ownership
- Don't expect it to happen fast
- Be happy with the outcome
- Keep it simple, make it extensible
- Keep it open
- Just because it isn't instantly successful doesn't mean it has failed

# Contents:

1. Data and data quality
2. Creating and using data standards
3. APIs and the web of data
4. Standards and APIs @ BBC

# What is an API?



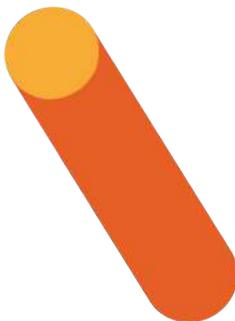
gg

*A promise by one system to  
consistently interact with  
another system in a  
standard way.*

jj

# What is the web's API?+

Describe an interaction  
and request...



THIS IS FOR EVERYONE



## **What Tim Invented**

Before the web we had lots of text contained in files.

These files were exchanged over FTP servers and people had to make requests using a console application.

Browsing wasn't possible. You looked at file names and downloaded what you thought interesting.

## **What Tim Invented**

The web enabled us to create rich text resources in the web.

Such resources could contain mark-up which gave structure to the text including:

- Lists
- of
- Things

It also enabled pictures to be added and most importantly links to other resources.

## Building web resources

Pick an artist on BBC Music

See if you can find:

- 1) Identifiers
- 2) Different content sources

Which identifiers and content  
are internal/external?

The screenshot shows the BBC Music website interface. At the top, there is a navigation bar with the word "MUSIC" in large letters, a search bar labeled "Search Music" with a magnifying glass icon, and dropdown menus for "Tracks", "Performances", "Playlists", "Artists", and "More". A "My Music" button is also present. Below the navigation, there is a profile section for Taylor Swift, featuring her photo, birth date (Born 13 December 1989), and a "Add to My Music" button. The main content area displays four categories: "Performances (10)", "Tracks (17)", "Events", and "News". Under "Performances", there is a link to "LATEST CLIP Taylor Swift". Under "Tracks", it says "LAST PLAYED ON BBC". Under "Events", there is a link to "Find events near you". Under "News", there is a link to "7 embarrassing prop malfunctions and onstage mishaps". The background of the main content area features a photograph of Taylor Swift singing into a microphone on stage. Below the main content, there is a section titled "Taylor Swift Biography (Wikipedia)" with a partial summary of her career.

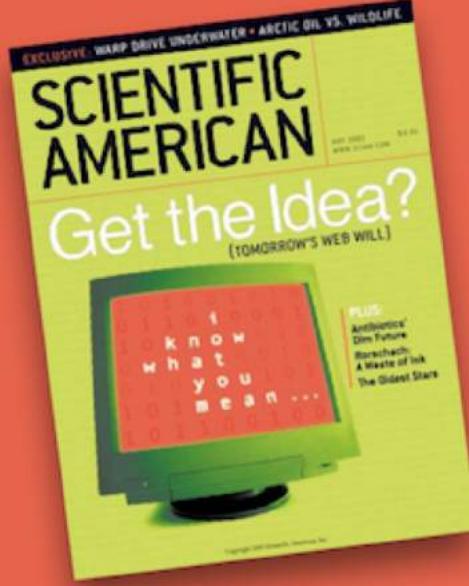
## The web as we know it





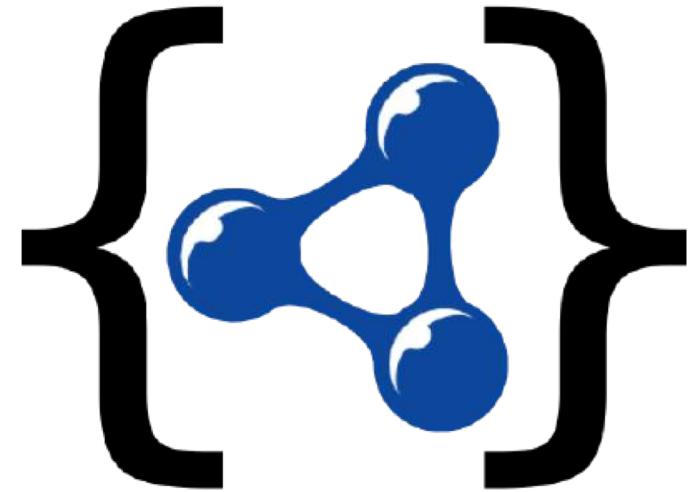
“ The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. ,,”

*Tim Berners-Lee 2001*

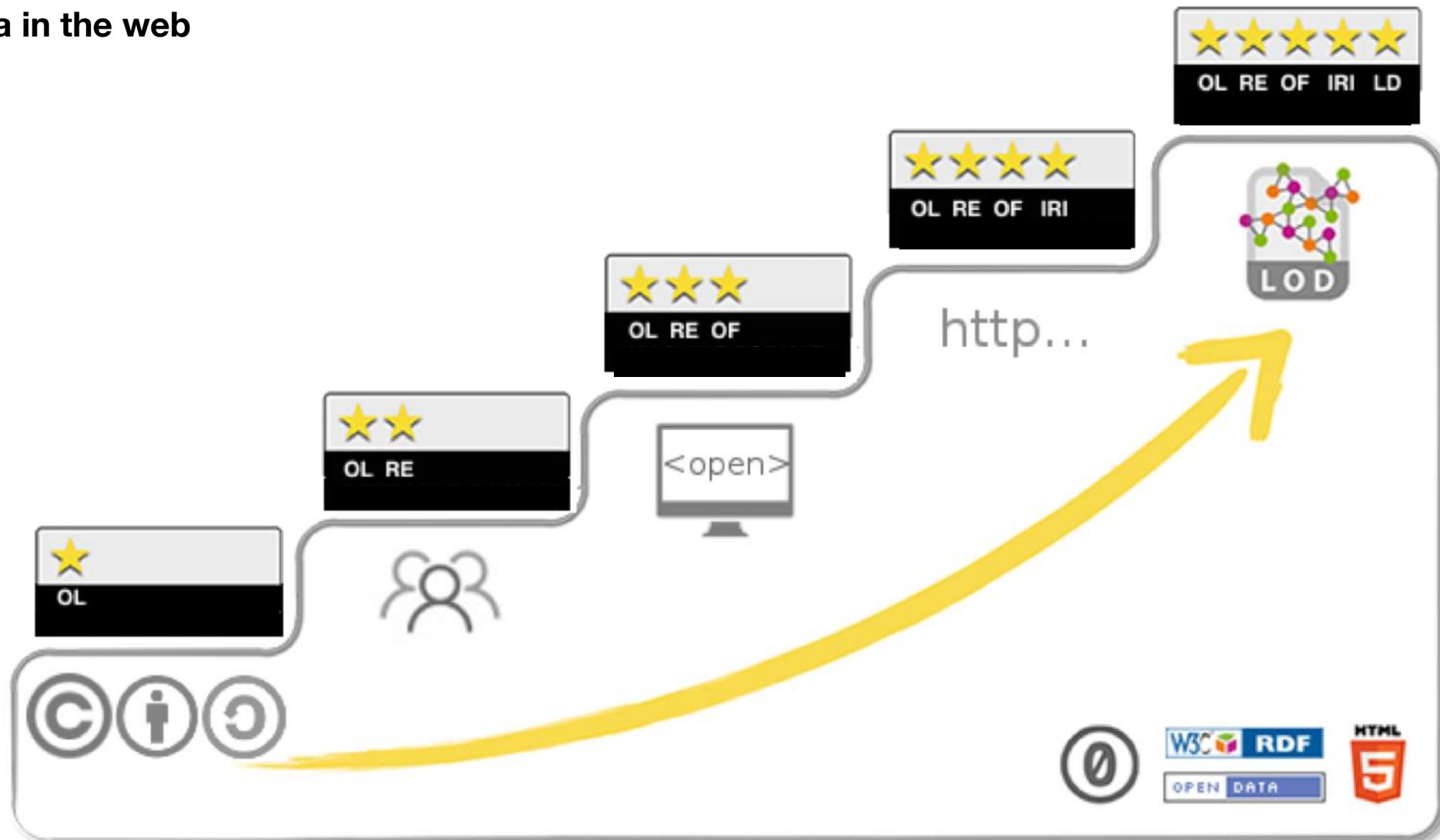


“ The real power of the ~~Semantic~~ Web will be realized when people create many programs that collect ~~Web~~ content from diverse sources, process the information and exchange the results with other programs. ,,”

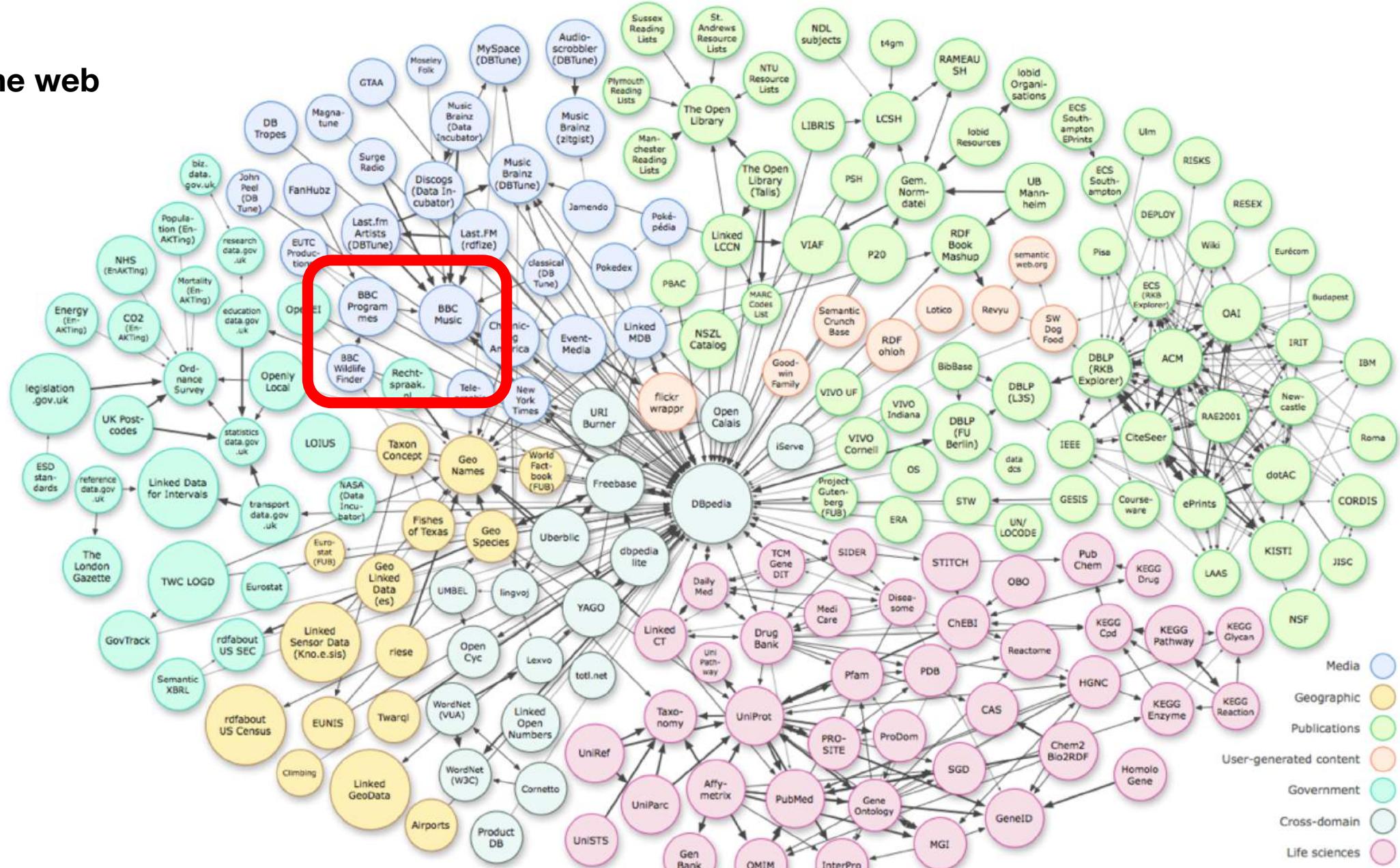
# Data



## Data in the web

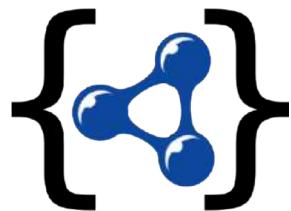


## Data IN the web





<http://bbc.co.uk/news>

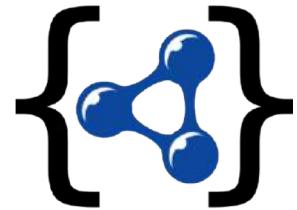


<http://feeds.bbci.co.uk/news/rss.xml>

How do we switch between document and data?



<https://www.trade-tariff.service.gov.uk/trade-tariff/sections/1>



?

How do we switch between document and data?

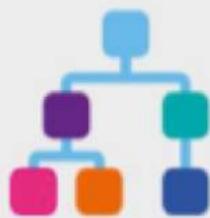
# We can standardise...



Models



Words



Taxonomies



Identifiers



File formats



Schemas



Data transfer



How we collect data



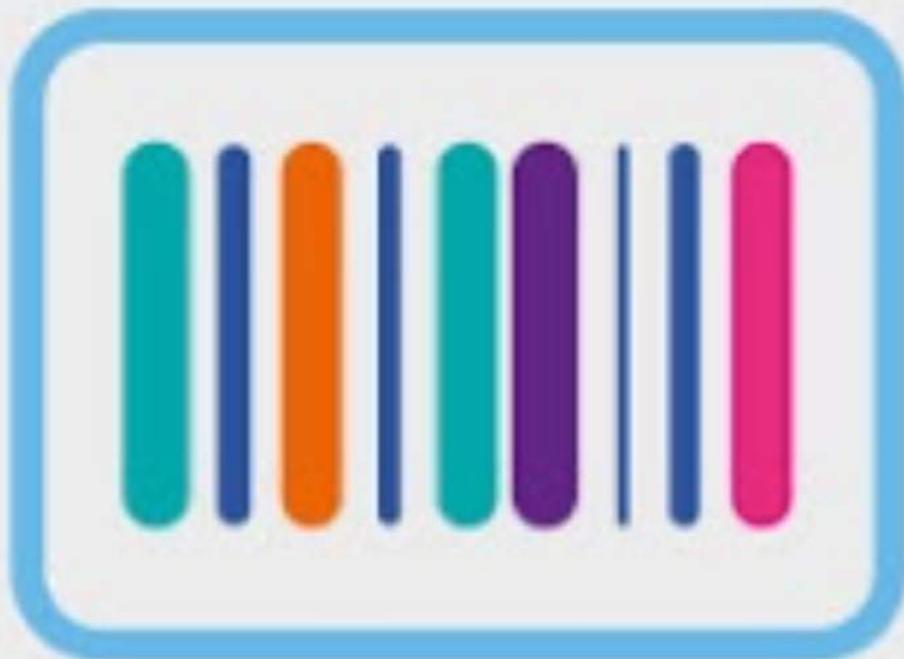
Units and measures



Data types



Code of practise



**Identifiers**

What is this the identifier for?

978-0471781172

What is this the identifier for?

1W75

What is this the identifier for?

<http://swrailway.com/train/1W75>

Authority

Database Identifier  
table?

## Dataset without identifiers

Payment Date	Expense Type Assets Under	Expense Area	Supplier Name	Transactions Number	Amount (Â£)
12-May-10	Construction (Oracle PA Control Account)	NON CASH & UNALLOCATED COSTS	GIBS LTD HAYS SPECIALIST	1002189	251,542.26
12-May-10	Corporate Strategy Consultancy	DG FINANCE	RECRUITMENT LTD	1163726	46.27
12-May-10	VAT (Input) (Oracle Sub-Ledger Control Account)	NON CASH & UNALLOCATED COSTS	HAYS SPECIALIST RECRUITMENT LTD	1163726	8,130.49
12-May-10	Fco Healthcare Scheme	DG CHANGE & DELIVERY	HEALIX INTERNATIONAL LTD	1163727	113,130.70
12-May-10	Programme Spend (Oracle Projects Control Account)	DG EUROPE AND GLOBALISATION	THE CARBON TRUST	1163753	34,500.00
12-May-10	P.A.Y.E Income Tax (Oracle PAY Control Account)	NON CASH & UNALLOCATED COSTS	HM REVENUE & CUSTOMS	4006772	4,000,783.16

## Dataset with identifiers

Company	Company ID	Amount (£)
3M UNITED KINGDOM PUBLIC LIMITED COMPANY Choose new match	<a href="https://opencorporates.com/companies/gb/01123045">https://opencorporates.com/companies/gb/01123045</a>	211,500.00
ACCOR SERVICES NORTH AMERICA, INC. Choose new match	<a href="https://opencorporates.com/companies/us_ny/1007862">https://opencorporates.com/companies/us_ny/1007862</a>	29,065.76
BRITISH COUNCIL(THE) Choose new match	<a href="https://opencorporates.com/companies/gb/RC000060">https://opencorporates.com/companies/gb/RC000060</a>	9,246,300.47
BULLET MARKETING LIMITED Choose new match	<a href="https://opencorporates.com/companies/gb/04842820">https://opencorporates.com/companies/gb/04842820</a>	31,222.10
THE CABINET OFFICE LTD Choose new match	<a href="https://opencorporates.com/companies/gb/05756325">https://opencorporates.com/companies/gb/05756325</a>	3,840,196.89
CAPGEMINI UK PLC Choose new match	<a href="https://opencorporates.com/companies/gb/00943935">https://opencorporates.com/companies/gb/00943935</a>	60,115.02
CHRISTIAN AID Choose new match	<a href="https://opencorporates.com/companies/gb/05171525">https://opencorporates.com/companies/gb/05171525</a>	32,657.79
CONCERTO CONSULTING LIMITED Choose new match	<a href="https://opencorporates.com/companies/gb/05040594">https://opencorporates.com/companies/gb/05040594</a>	32,432.45
CONSERVATION INTERNATIONAL CANADA	<a href="https://opencorporates.com/companies/can/0700015">https://opencorporates.com/companies/can/0700015</a>	100,255.16

What is this the identifier for?

[https://opencorporates.com/  
companies/gb/01123045](https://opencorporates.com/companies/gb/01123045)

What is this the identifier for?

<http://id.southampton.ac.uk/building/32>

Document or Data?

Wrong. It identifies the physical building, not the document or the data!

Building: <http://id.southampton.ac.uk/building/32>

Document: <http://data.southampton.ac.uk/building/32.html>

Data: <http://id.southampton.ac.uk/building/32.rdf>

How does a machine know to make the changes?

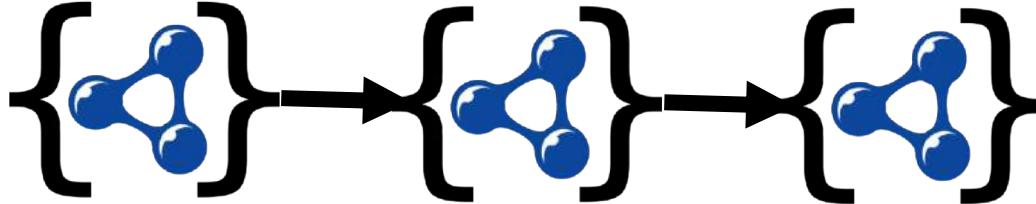
Physical



Documents



Data



## Linked data

# Amazing but challenging to publish and use

**EEE Building**

<http://id.southampton.ac.uk/building/32> ← This is the URI

Detail	Facilities	Services	Energy
Site: Highfield Campus			
Construction: 2006			
Architect: John McAslan & Partners			
Features: Building 32 is non-residential			

[View Disability Report for this Building](#)

**Occupants**

- Electronics & Computer Science
- Southampton Education School
- Agents, Interactions & Complexity
- Web & Internet Science
- Leadership School Improve &Effectiveness
- Lifelong & Work-Related Learning
- Mathematics & Science Education
- Social Justice & Inclusive Education
- Teaching Only Staff
- Deanery

 ©2010 Francois-Xavier Beckers (CC-BY)



[http://ton.ac.uk/building/32](#)

→ rooms:Building, [http://id.southampton.ac.uk/ns/UoSBuilding](#)  
→ "EEE Building"  
occupant → Electronics & Computer Science, Southampton Education School, Agents, Interactions & Complexity, Web & Internet Science, ...show 8 more...  
location → "32" [^http://id.southampton.ac.uk/ns/building-code-scheme](#)  
locations:within → Highfield Campus  
→ <http://www.soton.ac.uk/estates/ourestate/buildings/highfield/32.html>  
→ "50.9364157" [^xsd:float](#)  
→ "-1.395905" [^xsd:float](#)  
[southampton.ac.uk/ns/disabledGoPage](#) → <http://www.disabledgo.com/en/access-guide/building-32>  
locations:easting → "442544" [^xsd:integer](#)  
locations:northing → "115392" [^xsd:integer](#)  
Organization → University of Southampton  
[southampton.ac.uk/ns/ombielName](#) → "Bldg 32 (EEE)"  
feature → Building 32 is non-residential  
[southampton.ac.uk/ns/buildingDate](#) → "2006"  
[southampton.ac.uk/ns/buildingArchitect](#) → John McAslan & Partners  
spatial → "POLYGON((-1.3961073411331264 50.93683868764933, -1.3958347895092957 50.9368567227702, -1.3956958407975968 50.936065737417, -1.3959558923017397 50.93603859197583, -1.3961073411331264 50.93683868764933))"  
[southampton.ac.uk/ns/electricityTimeSeries](#) → "elec/b32/ekw"  
← is spatialrelations:within of ← 32 / 3077, 32 / 1015, Physical and Applied Science Faculty Deanery, Social and Human Sciences Faculty Deanery, ...show 54 more...  
← is foaf:depicts of ← <http://data.southampton.ac.uk/image-archive/buildings/raw/32.jpg>, <http://data.southampton.ac.uk/image-archive/buildings/1000/32.jpg>, <http://data.southampton.ac.uk/image-archive/buildings/800/32.jpg>, <http://data.southampton.ac.uk/image-archive/buildings/600/32.jpg>, ...show 5 more...  
← is event:place of ← RAeS Solent Branch Christmas Special Lecture - The Red Arrows

 a Building  
Google Infotex Ltd & Blazegraph, The GedInformation Group

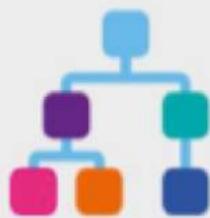
# We can standardise...



Models



Words



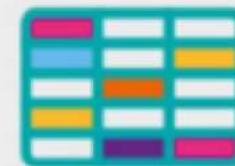
Taxonomies



Identifiers



File formats



Schemas



Data transfer



How we collect data



Units and measures



Data types



Code of practise

# REpresentational State Transfer (REST)

A way to interact with resources over the web.

GET <http://mydomain.org/themostat> (returns “16.2”)

PUT “18” -> <http://mydomain.org/thermostat>

## Data in the web



## Demo ([Tanzania dataset](#))

The Tanzania dataset from earlier is available via an API at

<http://api-demo.learndata.info/Tanzania>

You can get an individual facility using its ID:

<http://api-demo.learndata.info/Tanzania/ID/101>

Can you get a list of all the facilities in Mwanza?

How do you get the data and not the document?

## Representational (TANZANIA)

<http://api-demo.learndata.info/Tanzania/100.html>

<http://api-demo.learndata.info/Tanzania/100.csv>

<http://api-demo.learndata.info/Tanzania/100.json>

```
[  
  {  
    "ID": "100",  
    "Health facility name": "Bulima",  
    "Health facility type": "Dispensary",  
    "Health facility owner": "Not Specified",  
    "Region": "Simiyu",  
    "Star Rating": "1"  
  }  
]
```

## **Query APIs**

Query APIs can be used alongside REST APIs.

Are best used to find resources or data.

e.g. <http://api-demo.learndata.info/Tanzania/?Region=Mwanza>

or for multiple parameters:

<http://api-demo.learndata.info/Tanzania/?Region=Mwanza&Health%20facility%20type=Dispensary>

## Query API Representations

<http://api-demo.learndata.info/Tanzania/?Region=Mwanza>

Where do you put the .json?

Answer, you don't, use content negotiation!  
n.b. some APIs add &format=json (yuck!)

## Demo (BBC LightBulbs)

There are 10 light bulbs controllable at

<http://bit.ly/bbcbulbs>

Each one is identified by its number:

/bulbs/{bulbId}

You can get html, csv and json data about each bulb

You can use the API to turn them on and off:

/bulbs/{bulbId}/on & /bulbs/{bulbId}/off

---

## Real life example

<http://data.ordnancesurvey.co.uk/id/postcodeunit/EC2A4JE>

Try it in the different formats? (.json)

Can you click links?

Is it 5-star?

---

## APIs on the BBC website

How do you access data about Taylor Swift from the BBC website?

What other aspects of data standards and APIs can you spot in use here, what is missing?

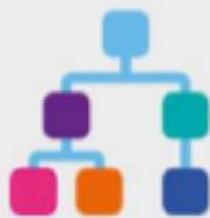
# We can standardise...



Models



Words



Taxonomies



Identifiers



File formats



Schemas



Data transfer



How we collect data



Units and measures



Data types



Code of practise



---

## What's next for the BBC?

The BBC has played a key role in the evolution of the web of data, but what is next?

# Any questions?

## Get in touch

If you would like to talk to us about collaborating, partnering, supporting our work, or anything else, we'd love you to get in touch.

[info@theodi.org](mailto:info@theodi.org)

+44 (0)20 3598 9395

@ODIHQ

