

The Web of Linked Data

A Tutorial

Nick Gibbins, Hugh Glaser, Christopher Gutteridge,
Wendy Hall, Ian Millard, Nigel Shadbolt
22 November 2010

The Plan

- Introduction
- How to do the “Five Stars”
- The Process at the University
- The Pragmatics at the University
- Group Breakout
- Coffee
- Feedback from Groups
- Discussion and Planning

Before You Start

- Strangely perhaps, we are following government
- Dataset Audit
- Governance

Where

- Some will be clear (legislation, etc.)
 - But not so much non-personal
- Service Database
- A spreadsheet on someone's PC
- A PDF on the web

What

- Low hanging fruit
- Or difficult but crucial set
- Facilitate linkage
 - “Pivot” sets
 - May seem rather mundane

What not

- Privacy
- Trust
 - Manage expectation of quality
- Identify “dangerous” datasets
 - Recognise what may cause trouble

When

- Static or Dynamic
- Release frequency

Governance

- Who decides what
- Who gives advice
- Empowering dataset owners
- URI schemas



Linked Data in a Nutshell

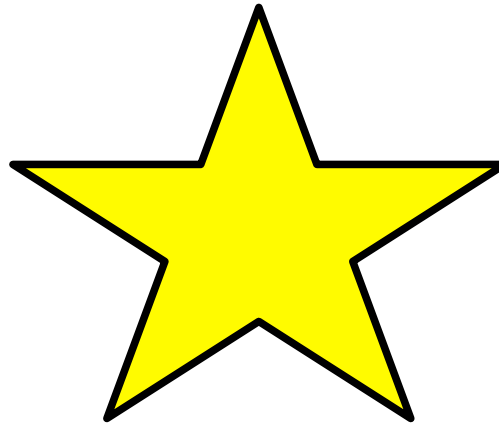
<http://www.flickr.com/photos/arielarielariel/322301228/>

The path to Open Linked Data

- Currently huge interest around the world
- Often social issues to be overcome
- May involve technologies which are new or different
- From “Raw Data Now” to semantic nirvana
- Often about making data available which already exists
- Incremental approach
- Five-star grading for your open data

One Star

Put your Data on the Web, with an Open License



One Star – What it means

- Simply put your data on the web
- Documents in any format
- License is very important
 - Public Domain Dedication and License
 - Open Data Commons
 - Open Government License

One Star – Benefits

- Huge power in making data available, well done!
- Users can
 - view, print
 - download documents for later use
 - start to do stuff with the data
- But data still locked in document

One Star – Requirements

- Web server
- Policy

Main Site Navigation

Home
A-Z Service Directory
Advice And Benefits
Business
Community And Living
Council And Democracy
Education And Learning
Environment And Planning
Health And Social Care
Housing
Jobs And Careers
Leisure And Culture
News And Publications
Transport And Streets
Frequently Asked Questions

Popular Links

Bus Services
Council Tax
Councillors
Jobs At The Council
Leisure Centres
Libraries
Parking
Planning Application Search
Recycling
Schools

Location: [Council and Democracy](#) » Transparency

Transparency

RBWM believes transparency is a key condition and driver for the delivery of council services. As a public funded organisation, the council has a duty to the residents that it serves to be transparent in its business operations and outcomes.

This data is freely re-usable under the same terms as data.gov.uk:
<http://data.gov.uk/terms-and-conditions>

■ [Annual Allowances paid to Individual Councillors](#)

Total amounts paid to individual Members for each financial year.

■ [Automatic Energy Reading](#)

On 25th June 2009 the Council agreed to implement a rolling programme of installing automatic meter reading (AMR) in the main civic buildings as a means of driving down energy consumption and reducing costs. The AMR will measure carbon emissions and the costs resulting from the electricity and gas consumption at the chosen sites.

■ [Corporate Strategy](#)

This Strategy provides a reference to ensure that residents, staff and partners are fully informed of the future aspirations and plans of the Royal Borough.

[Councillors' Allowances](#)

To ensure that they are not financially penalised for carrying out their Council duties, Councillors receive from the Council an allowance or allowances. The allowances paid to Councillors are contained within the Council's Members Allowances Scheme.

■ [Council Information and Data Transparency \(pdf\)](#)

This report describes how the council has reset its default position to transparent - so that all information (save three key areas) is available to the public.

■ [Council Properties Utilities Consumption](#)

Site providing consumption and cost information for Council properties. To log on to the site, enter the username: guest@guest.com. The password is guest

■ [Information Requests](#)

This page contains links to monthly information request summaries and performance statistics. With effect from January 2011 responses to requests will be accessible.

■ [Freedom of Information Act - RBWM Publication Scheme](#)

The Freedom of Information Act 2000 (FoIA) gives a general right of access to all types of recorded information held by public authorities, including the Royal Borough of Windsor & Maidenhead.

■ [Major Council Projects](#)

Details of the major projects that are currently being undertaken by the Borough Council

■ [Payment to Suppliers](#)

The attached reports list the [payments to suppliers](#) where the charge to a specific cost centre is greater than or equal to £500. Publication of these lists forms part of the Council's commitment to be open and transparent with its residents.

■ [Pay and Benefits of Staff Earning over £50,000 per annum](#)

Pay and benefits for staff earning over £50,000 are set by a group of Elected Members called the Employment Panel. Decisions are made by evaluating the content of each job using a set of nationally recognised criteria. Each job is then benchmarked against other national and local jobs before arriving at a specific salary, taking into account factors such as the competitiveness of the market for the role and the experience of the candidate.

■ [RecycleBank](#)

Recycling is easier and more rewarding for every local household following the council's ground-breaking RecycleBank scheme. The concept is simple - the more you recycle the more rewards you earn.

■ [Sources of RBWM Income](#)

The report details the major sources of actual income to the council. This data is represented graphically, with supporting data-sets of detailed information. Publication of this data forms part of the Council's

Related Links

[Access To Information](#)[Council & Democracy](#)[Information Requests 2010](#)[Information Requests 2009](#)[RecycleBank](#)

One Star – Best practise

- If you're publishing many resources, make them easy to find and access
- Include descriptions of what is contained
 - metadata, data about documents
- May wish to deploy a “data catalogue” or “portal”
 - social interaction, comments, forum

Datastore

Datasets A-Z

Categories

Keywords

Organisations

Search Datasets

Request Dataset

Popular Suggestions

Latest Suggestions

API Beta

Blog

Inspirational Uses

Useful Links

Digital Advisory Board

About

FAQs

Alcohol Related Hospital Admissions

[Home](#) › Alcohol Related Hospital AdmissionsJANUARY 2010 | [Health](#)Average:  Average: 4 (1 vote)

Numbers and rates of alcohol related hospital admissions. Hospital Stays for Alcohol Related Harm.

Measures are Numbers, Percentages and Percentage Change.

Warnings/Notes: Directly age standardised rates. Rates per 100,000 European Standard population.

Vital Signs Indicator VSC26: The Vital Signs framework provides a series of indicators from which PCTs may select a number that reflect the priorities for health in that area. [www.dh.gov.uk/en/Publicationsandstatistics/Publications/Publ...](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/Publications/Publ...)National Indicator Set NI39: The national indicators form part of a new performance framework for local authorities and local authority partnerships. www.communities.gov.uk/localgovernment/performanceframeworkp...

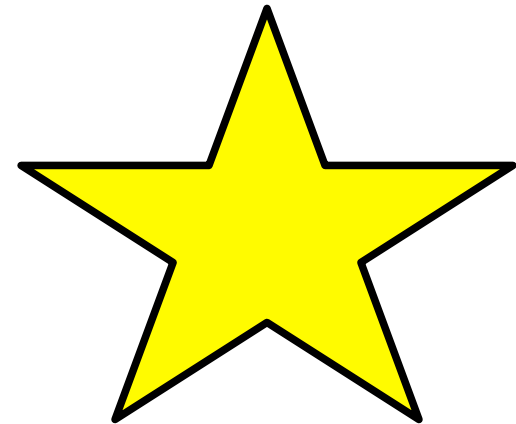
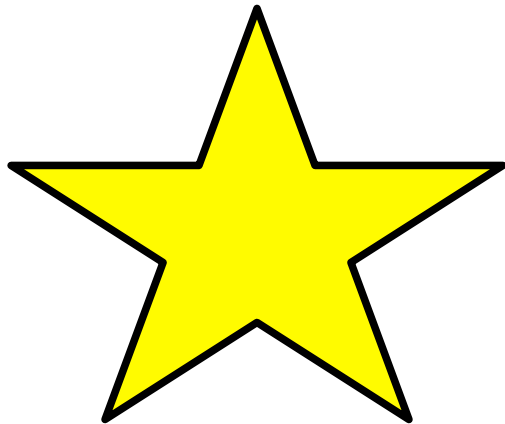
Date of Next Release Mar-10

Geography	Local Authority
Author Name	North West Public Health Observatory (NWPHO)
Update Frequency	Annual
Date range	2002/03 to 2007/08
Release Date	01/03/2009
Metadata update	2010-09-10
License Summary	Crown Copyright
URL	http://www.apho.org.uk/default.aspx?QN=HP_METADATA...
Download URL	http://www.nwph.net/alcohol/lape/download.htm
CSV URL	http://data.london.gov.uk/datafiles/health/alcohol...
XML URL	http://data.london.gov.uk/datafiles/health/alcohol...

TAGS | [admissions](#) | [alcohol](#) | [health](#) | [hospital](#)

Two Stars

Make it Available in a Machine Readable Format



Two Stars – What it means

- Publish your data in a structured format
 - lists, spreadsheet, database dump
- PDF, images, scanned documents are not acceptable!

Two Stars – Benefits

- Users now have access to the raw data
- They can
 - manipulate, edit, save data
 - search, annotate
 - export to different format(s)
- But may be constrained to proprietary software

Two Stars – Requirements

- Web server
- Policy

Two Stars – Best practise

- Ensure data is as regular and “clean” as possible
- Describe what columns and fields mean
- Consistent representation
- Don't mix numerical data with comments

Three Stars

Use an Open, Standard Format



Three Stars – What it means

- Publish your data in a structured format
 - that doesn't require proprietary software
- Excel spreadsheets are difficult to process
- CSV, JSON, XML

Three Stars – Benefits

- Users now have easy access to the raw data
- They can
 - freely access data in any way they like
 - open data in a variety of tools
 - write programs and scripts to process data
 - integrate into applications

Three Stars – Requirements

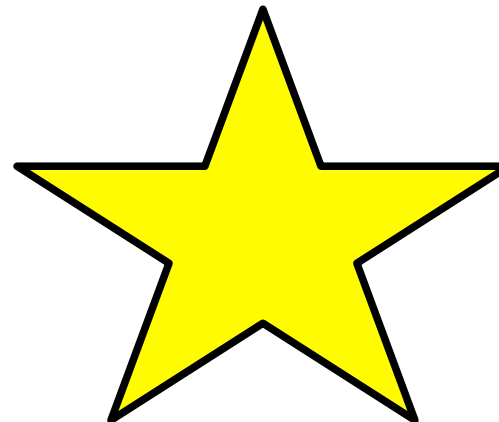
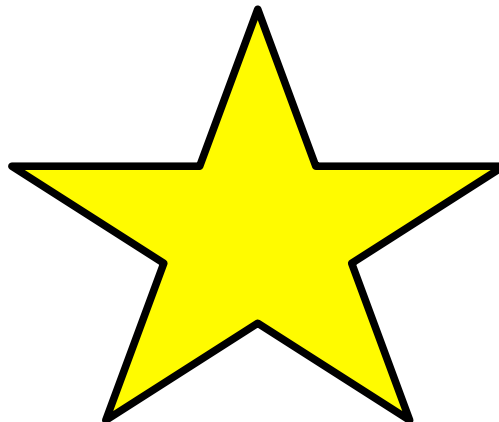
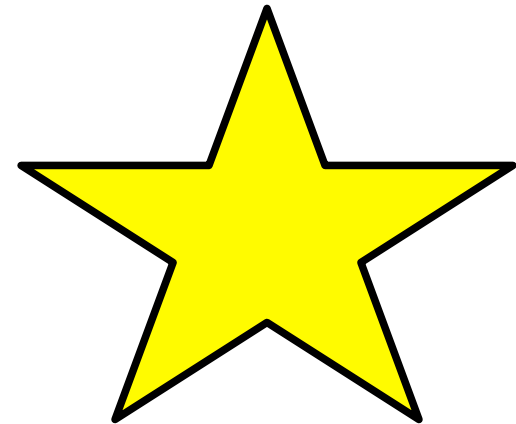
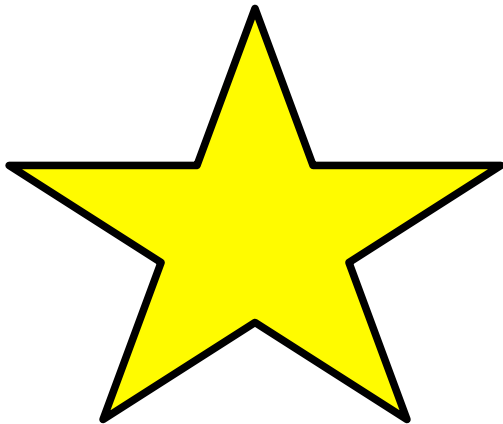
- Web server
- Policy
- May require plug-ins or effort to export/convert from proprietary tools into open format

Three Stars – Summary

- Can easily achieve 3* data
 - without expending large effort
- Few technical requirements
- Making data available opens up potential for reuse
 - often in ways not previously imagined
- Can still do more to join the “Web of Data”

Four Stars

Use an Open, Linked Data Format



Linked Data Principles

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information
4. Include links to other URIs, so that they can discover more things

1. Use URIs

- Uniform Resource Identifiers – standard means for naming things on the World Wide Web (and the Web of Linked Data)
 - <http://www.southampton.ac.uk/>
 - <mailto:nmg@ecs.soton.ac.uk>
 - <ftp://example.org/>
- The thing you type into the location bar of your browser

2. Use HTTP URIs

- Not all URIs are *resolvable* – can't look them up
- `mailto:nmg@ecs.soton.ac.uk` lets me send emails, but doesn't get me my mailbox
- HTTP is the standard protocol for the Web

Good URI Selection

- ‘Cool URIs don’t change’ – once you’ve chosen a URI convention for your organisation, it’s a pain to change it
- Getting this right is key to having your linked data used more widely

3. Provide Useful Information

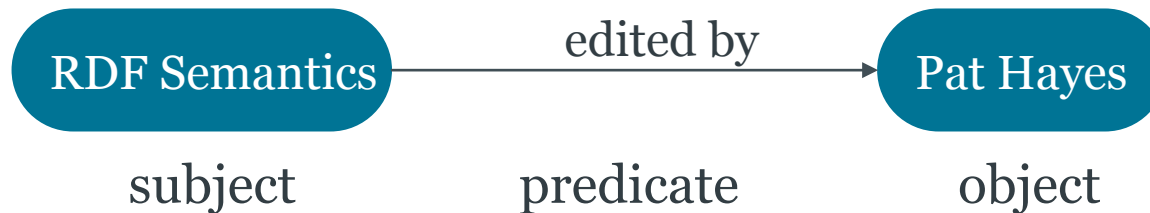
- Linked Data is about providing structured data on the Web
- RDF – the Resource Description Framework – is the ‘standard’ data format for structured Linked Data
- A W3C Recommendation
 - open standards process, royalty-free

Structure vs Syntax

- RDF is both a data model *and* a family of syntaxes
- Several RDF syntaxes exist:
 - RDF/XML
 - N₃
 - Turtle
 - Ntriples
 - TriX
 - RDFa
 - ...

The triple

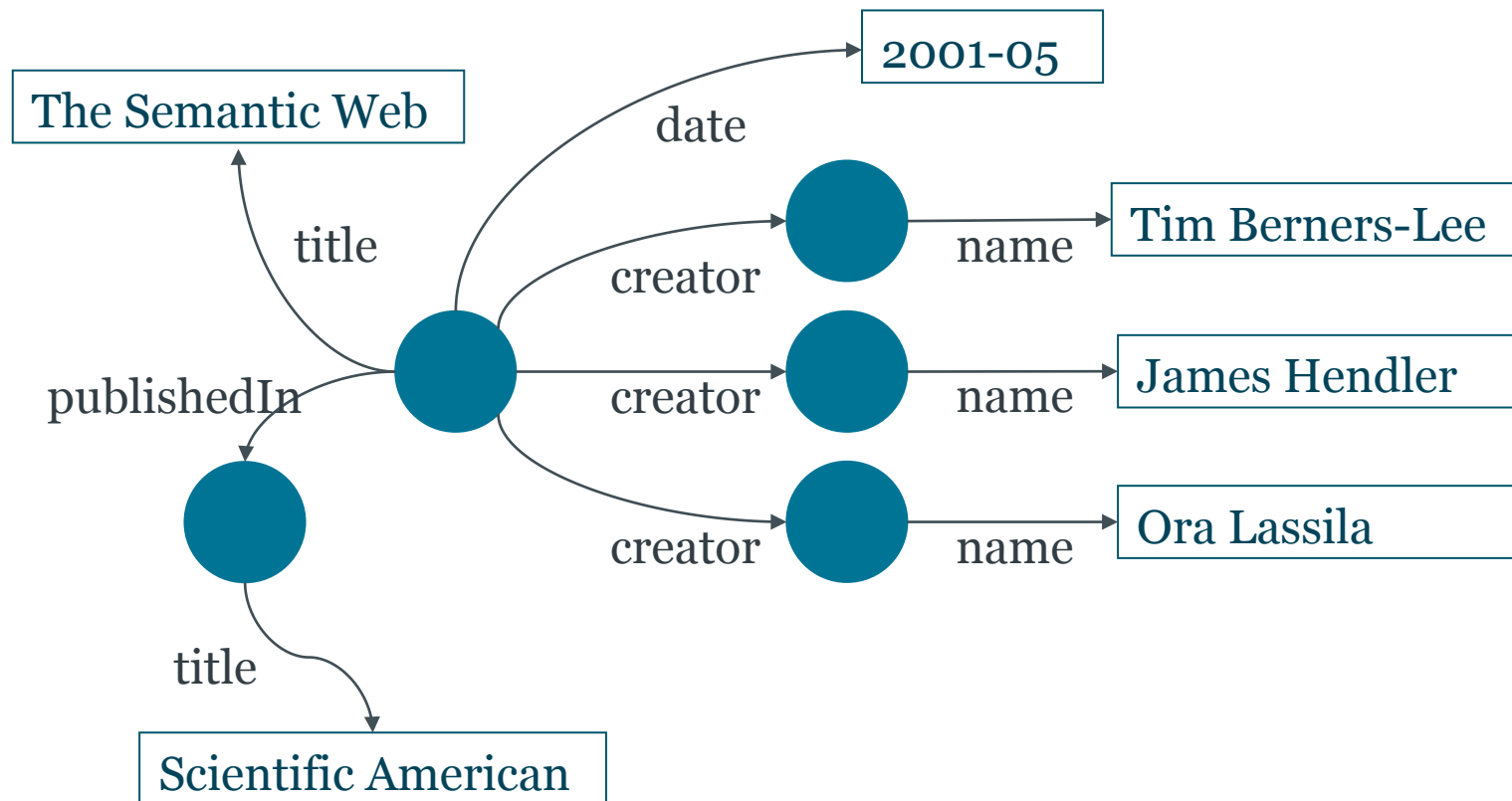
- Underlying model of triples used to describe the relations between entities in linked data
- (subject, predicate, object)
 - e.g. “RDF Semantics”, “edited by”, “Pat Hayes”



Example

- Take a citation:
 - Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. Scientific American, May 2001
- We can identify a number of distinct statements in this citation:
 - There is an article titled “The Semantic Web”
 - One of its authors is a person named “Tim Berners-Lee” (etc)
 - It appeared in a publication titled “Scientific American”
 - It was published in May 2001

Example

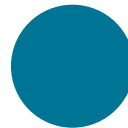


Example

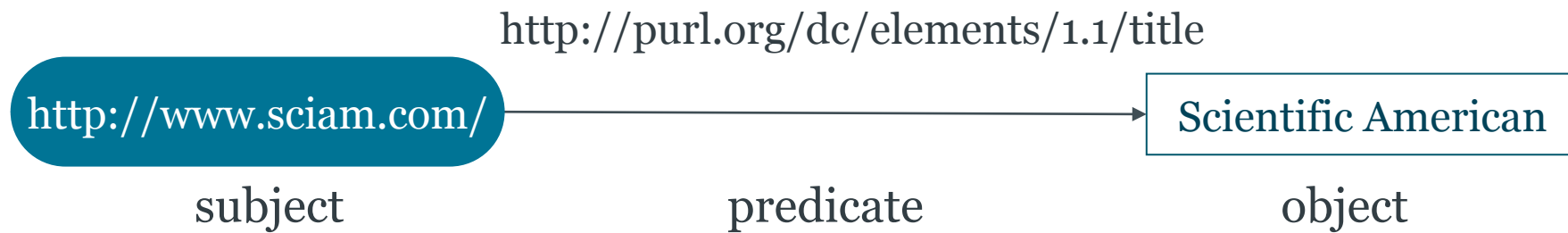
Literals have a value but no identity
(a string, a number, a date)

Scientific American

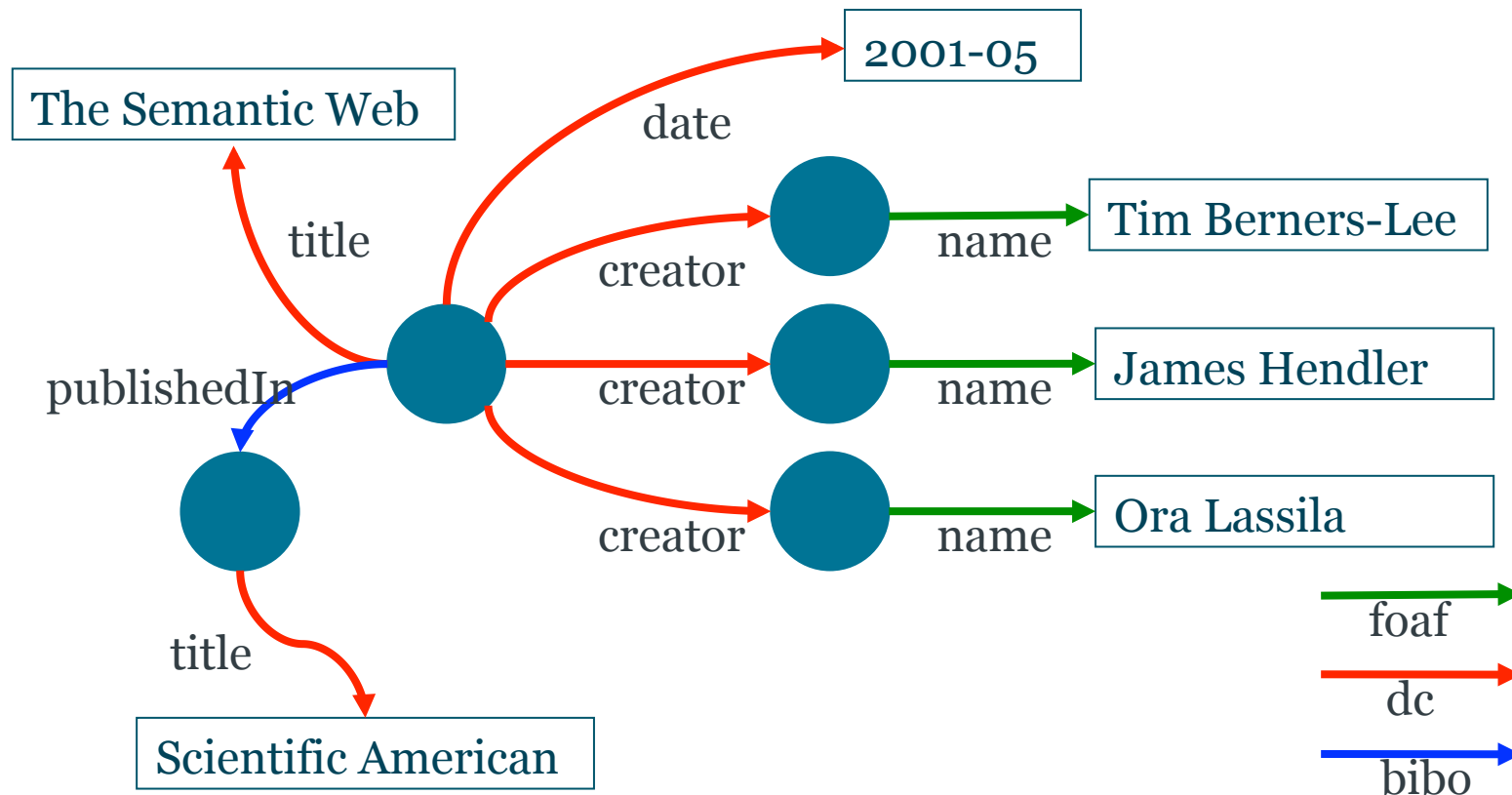
Resources represent objects with identity
(a web page, a person, a journal)



Example



Mixing Vocabularies



Defining the 'O' word

Ontology, *n.*

1. a. *Philos.* The science or study of being; that branch of metaphysics concerned with the nature or essence of being or existence.



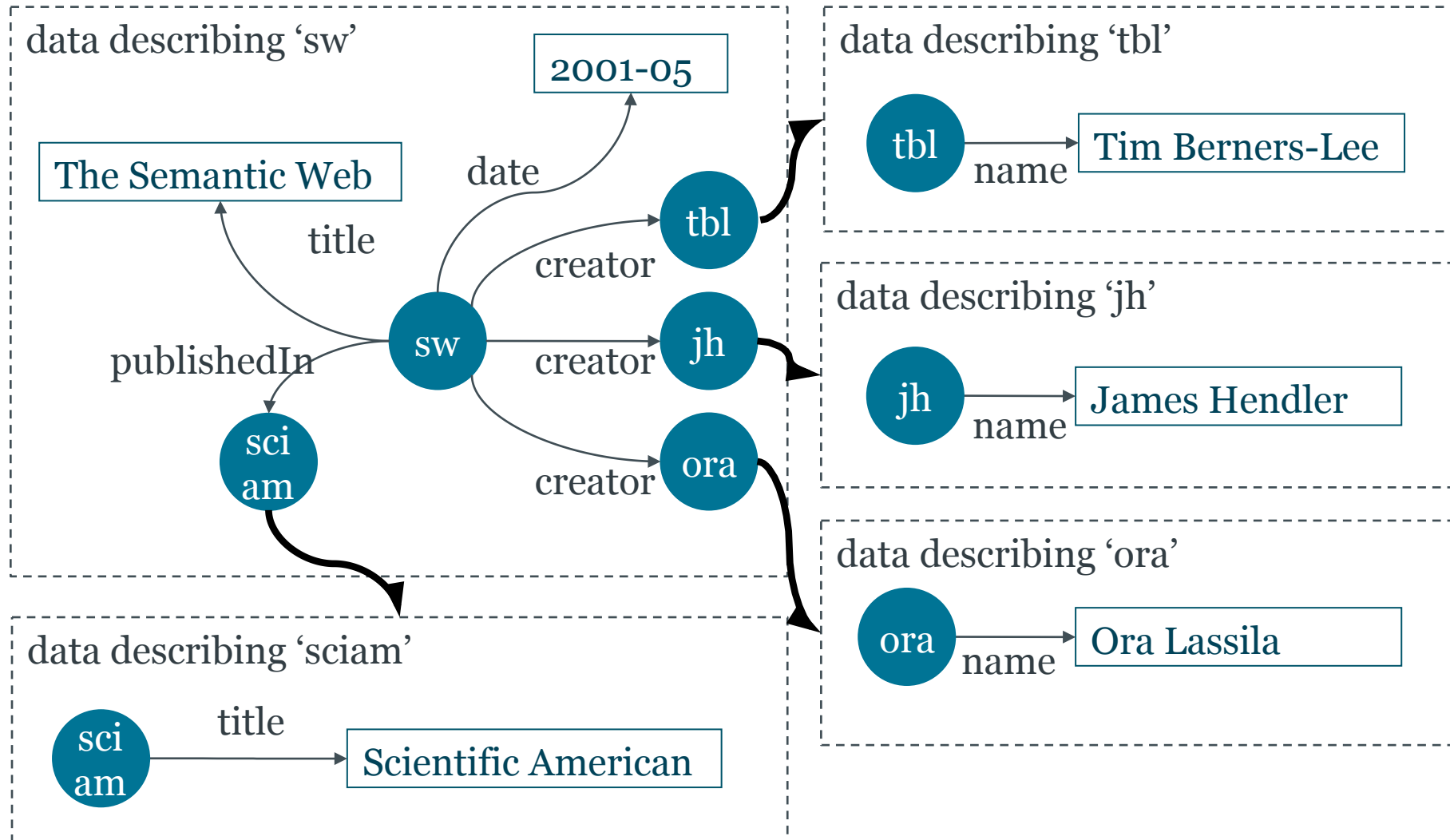
Defining the ‘O’ word

- A formal description of the objects, concepts, and relationships that are assumed to exist in some area of interest
- An engineered artifact
 - Facilitate communication
 - Promote sharing and reuse
 - Establish a joint terminology for a community of interest
- A little goes a long way

Linked Data Vocabularies

- Bibliographic metadata (Dublin Core, BIBO)
- News aggregation (RSS, ATOM)
- Cultural heritage (CIDOC)
- Knowledge Organisation (SKOS)
- Public Sector Information (IPSV)
- Social networking and online communities (FOAF, SIOC)
- Business information – products and services (GoodRelations, etc)
- ...

4. Link to other things



Five Stars

Link Your Data to Other People's Data



Five Stars – What it means

- The “Web of Data”
- Even within an organisation there will be different dataset publishing platforms
- So the Web of Data applies internally and externally

Links

- Achieve data fusion
 - Seeing how different datasets see resources
- What is the University view of Room 3008 in B32?
 - When was it cleaned?
 - Is there an outstanding maintenance ticket?
 - What AV is there, and is it working?
 - Am I scheduled to be taught in it at 11:00 today?
 - What energy does it consume?
 - Are there roadworks outside?
- Do the Council have plans?
- Are the Police concerned?

Linked Data Links

- Two sorts
 - “Foreign” URIs
 - sameAs

Foreign URIs

- Use someone else's URI in your dataset
 - Eg. Buildings & Estates may have URIs for rooms
 - So timetabling could use the same URIs
- Requires some coordination
 - Which can delay the earlier stars
 - But makes consumption somewhat easier

sameAs

- Actually often owl:sameAs
 - myURI owl:sameAs yourURI
 - ecs:learning-outcome-27 owl:sameAs ses:learning-outcome-46
- Usually directional from the source
 - ECS may want to say this, but SES may not, or may not have discovered this
- But can be Aggregated


In the Wild



<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.
This service helps you to find co-references
between different data sets.

Enter a known URI, or use Sindice to search first.

<sameAs> 

Equivalent URIs for <http://data.ordnancesurvey.co.uk/id/7000000000037254> –

- <sameAs> {
1. <http://dbpedia.org/resource/Portsmouth>
 2. http://dbpedia.org/resource/Portsmouth_UA
 3. <http://dbpedia.org/resource/UN/LOCODE:GBPME>
 4. http://dbpedia.org/resource/City_of_Portsmouth
 5. http://dbpedia.org/resource/Portsmouth,_Britain
 6. http://dbpedia.org/resource/Portsmouth,_England
 7. http://dbpedia.org/resource/Portsmouth_City_Council
 8. [http://dbpedia.org/resource/Portsmouth_\(district\)](http://dbpedia.org/resource/Portsmouth_(district))
 9. <http://www.okkam.org/ens/de2feafd9-ecbb-4e62-9bcb-290cc19a9194>
 10. <http://unlocode.rkbexplorer.com/id/GBPME>
 11. <http://unlocode.rkbexplorer.com/description/GBPME>
 12. <http://os.rkbexplorer.com/id/osr7000000000037254>
 13. <http://data.ordnancesurvey.co.uk/id/7000000000037254>
 14. <http://linkedgeodata.org/triptych/node/17721995#id>
 15. <http://mortality.psi.enacting.org/id/Portsmouth>
 16. <http://mpii.de/yago/resource/Portsmouth>
 17. <http://openlylocal.com/id/councils/353>
 18. <http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000003053f>
 19. <http://statistics.data.gov.uk/id/local-authority/00MR>
 20. <http://statistics.data.gov.uk/id/local-authority/24UK>
 21. <http://statistics.data.gov.uk/id/county?name=Portsmouth>
 22. <http://statistics.data.gov.uk/id/local-authority-district/00MR>
 23. <http://statistics.data.gov.uk/id/local-authority-district?name=Portsmouth>
 24. <http://statistics.data.gov.uk/id/local-authority?name=Portsmouth City Council>
 25. <http://statistics.data.gov.uk/id/local-authority-district?name=The ...Portsmouth>
 26. <http://sw.cyc.com/concept/Mx4ro602tsIDQdeU0OethnEzEA>
 27. <http://sw.opencyc.org/concept/Mx4ro602tsIDQdeU0OethnEzEA>
 28. <http://sw.opencyc.org/2008/06/10/concept/Mx4ro602tsIDQdeU0OethnEzEA>
 29. <http://sw.opencyc.org/2008/06/10/concept/en/CityOfPortsmouthEngland>
 30. <http://sw.opencyc.org/2009/04/07/concept/Mx4ro602tsIDQdeU0OethnEzEA>
 31. <http://sw.opencyc.org/2009/04/07/concept/en/CityOfPortsmouthEngland>
 32. <http://sws.geonames.org/2639996/>
 33. <http://sws.geonames.org/3333183/>
 34. <http://transport.data.gov.uk/id/local-authority/1745>
 35. <http://transport.data.gov.uk/id/local-authority-district/1745>
 36. <http://umbel.org/umbel/ne/wikipedia/Portsmouth>



<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.

This service helps you to find co-references

between different data sets.

Enter a known URI, or use Sindice to search first.

<sameAs>



Equivalent URIs for <http://data.ordnancesurvey.co.uk/id/7000000000017711> –

- <sameAs> {
1. <http://data.ordnancesurvey.co.uk/id/7000000000017711>
 2. <http://openlylocal.com/id/wards/4933>
 3. <http://statistics.data.gov.uk/id/cas-ward/00MSNC>
 4. <http://statistics.data.gov.uk/id/local-authority-ward/00MSNC>

[rdf+xml](#) · [n3](#) · [json](#) · [text](#)

Currently serving **39 221 073** URIs in **13 112 833** bundles!

[about](#) · [contact](#) · [get the widget](#)





Enter keyword(s)

Search

[home](#) [business](#) [leisure](#) [map shop](#) [explore](#) [magazine](#) [education](#) [about us](#) [data](#)

Portswood

A description of the resource identified by <http://data.ordnancesurvey.co.uk/id/7000000000017711>

<u>Preferred label</u>	Portswood
<u>Label</u>	Portswood
<u>Latitude</u>	50.928209
<u>Longitude</u>	-1.393214
<u>Type</u>	Unitary Authority Ward
<u>Gss code</u>	E05002465
<u>Area code</u>	UTW
<u>Census code</u>	00MSNC
<u>Unit i d</u>	17711
<u>In district</u>	The City of Southampton
<u>In european region</u>	South East
<u>Extent</u>	64470
<u>Easting</u>	442740.4
<u>Northing</u>	114482.4
<u>Touches</u>	Swaythling Bevois Bassett Shirley Bitterne Park
<u>Within</u>	The City of Southampton South East
<u>Point</u>	50.928209 -1.393214
<u>Notation</u>	00MSNC

openlylocal making local government more transparent

[ALL COUNCILS](#)
[HYPERLOCAL SITES DIRECTORY](#)
[OPEN DATA SCOREBOARD](#)
[COUNCIL SPENDING](#)
 Search by postcode

Southampton City Council: Portswood ward

Output Area	Student Communities A
Classification	
Size	280.731 hectares
Ons Snac Ref	00MSNC
Ons Gss Code	E05002465
Ordnance Survey Ref	7000000000017711
Fix My Street Reports	http://fixmystreet.com/reports/00MSNC

Councillors for this ward

- **Vincenzo Capozzoli** (Conservative)
- **Steve Sollitt** (Liberal Democrat)
- **Adrian Vinson** (Liberal Democrat)



About this Ward

Experimental

6,133 dwellings, 14,384 people live in this area, 35.6 years old on average

Religion





data.gov.uk

BETA

[Home](#) [Blog](#) [Data](#) [SPARQL](#) [Apps](#) [Ideas](#) [Ideas](#) [Forum](#) [Wiki](#) [Resources](#)
[About](#)

Portswood

A description of the resource identified by <http://statistics.data.gov.uk/id/electoral-ward/00MSNC>

Label	Portswood Portswood
Alternative label	Portswood
Type	Ward Unitary authority ward
Same as	7000000000017711
Contained by	SE280 427 921 00MS J
Country	921
District	00MS
Region	J
Parliamentary constituency	427
Local learning and skills council area	SE280
Notation	00MSNC 00MSNC

Foreign and sameAs

- Different approach for different URI types
- Actually, the 4* publisher can become 5* with no work
 - Third parties do these things
 - Or an enterprise can initiate the activity

I Favour sameAs

- Separation of concerns
- Can build RDF silos
- Links have different nature to the substantive data
 - Business Process
 - Provenance
 - Licence
- But you need a sameAs discovery and aggregation layer

data.soton.ac.uk: A Pragmatic Consideration

Authoritative Sources of Data

Timetabling

Banner

Buildings
& Estates

EdShare

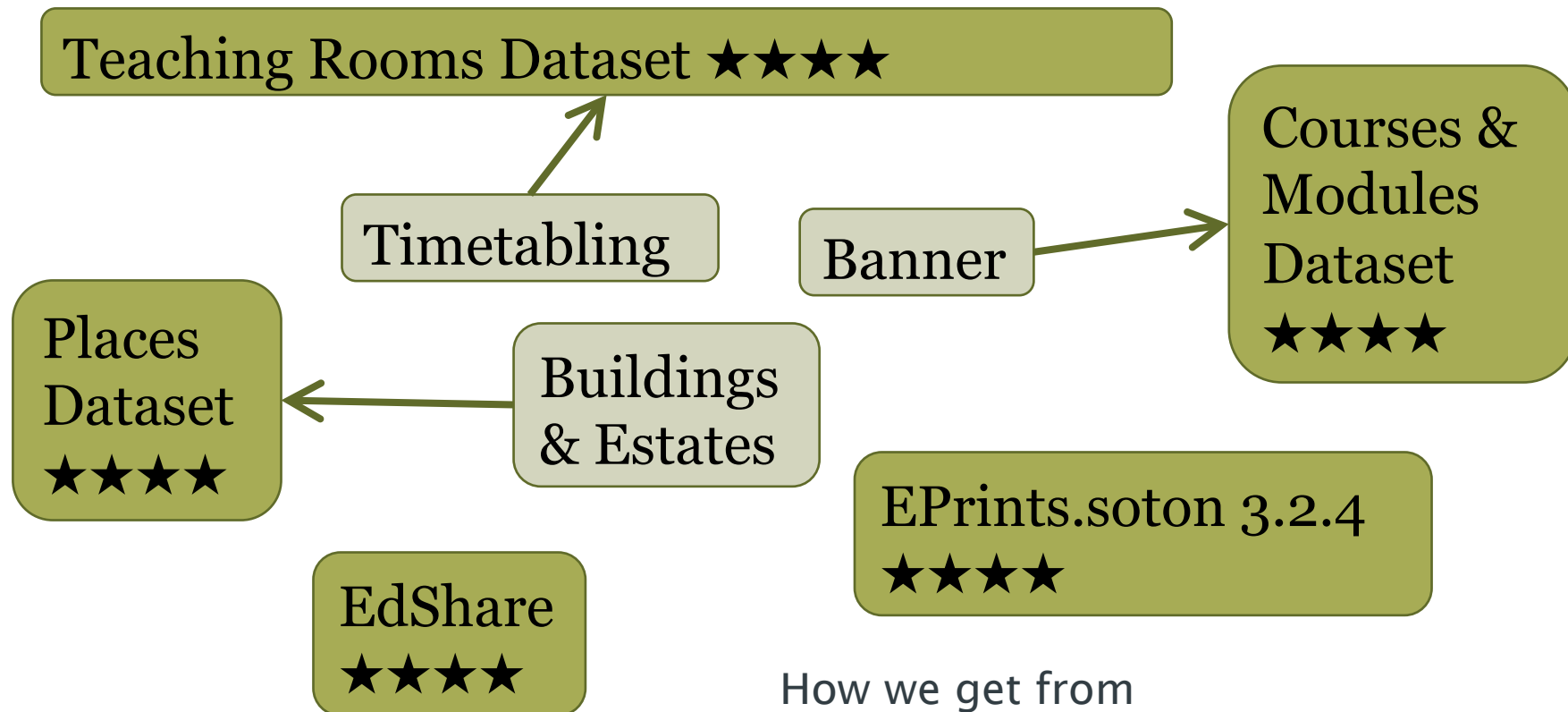
★★★★

EPrints.soton

★★★

+ many more!

Creating Open Datasets from Authoritative Sources



How we get from
a database to a public dataset will vary

One approach isn't likely to suit all cases.

Easy Links

Teaching Rooms Dataset ★★★★★

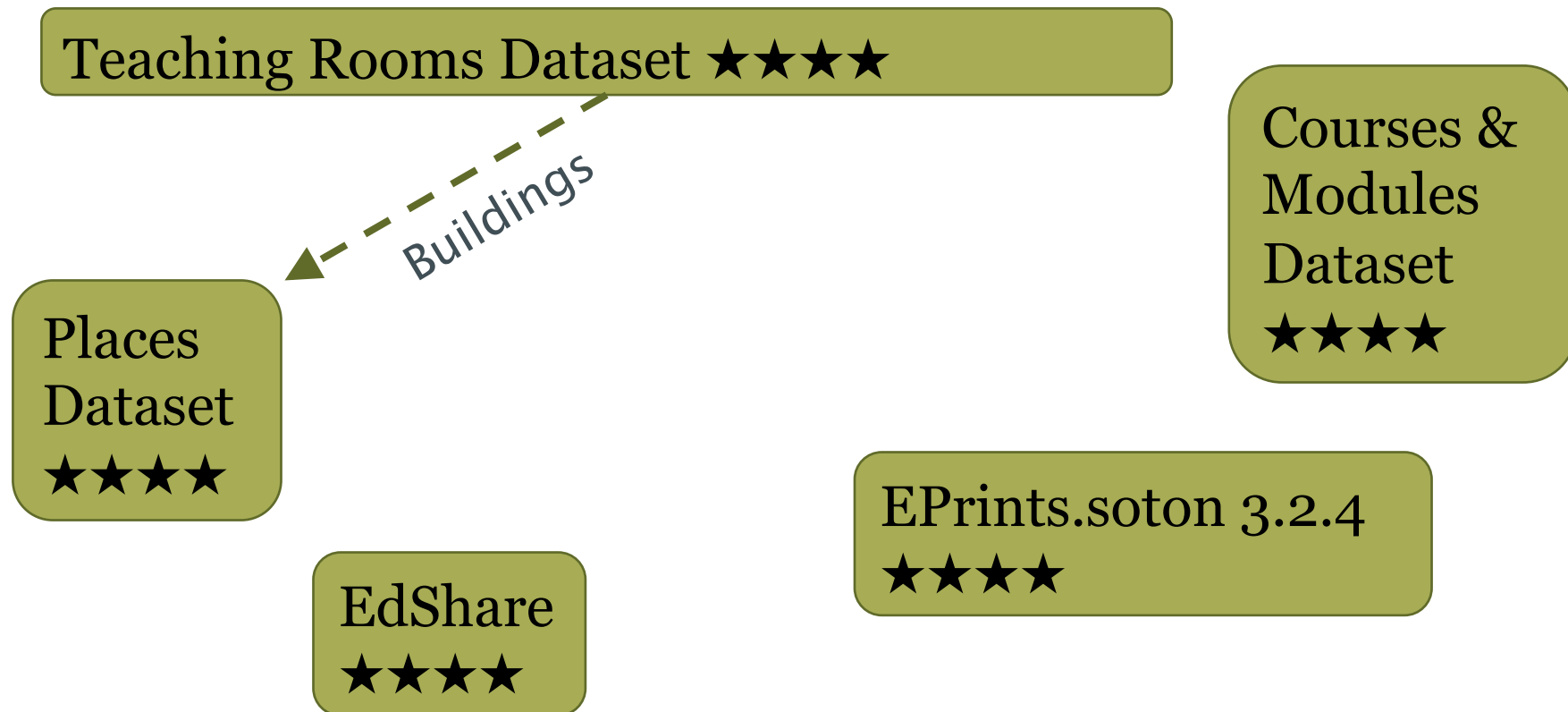
Places
Dataset
★★★★★

EdShare
★★★★★

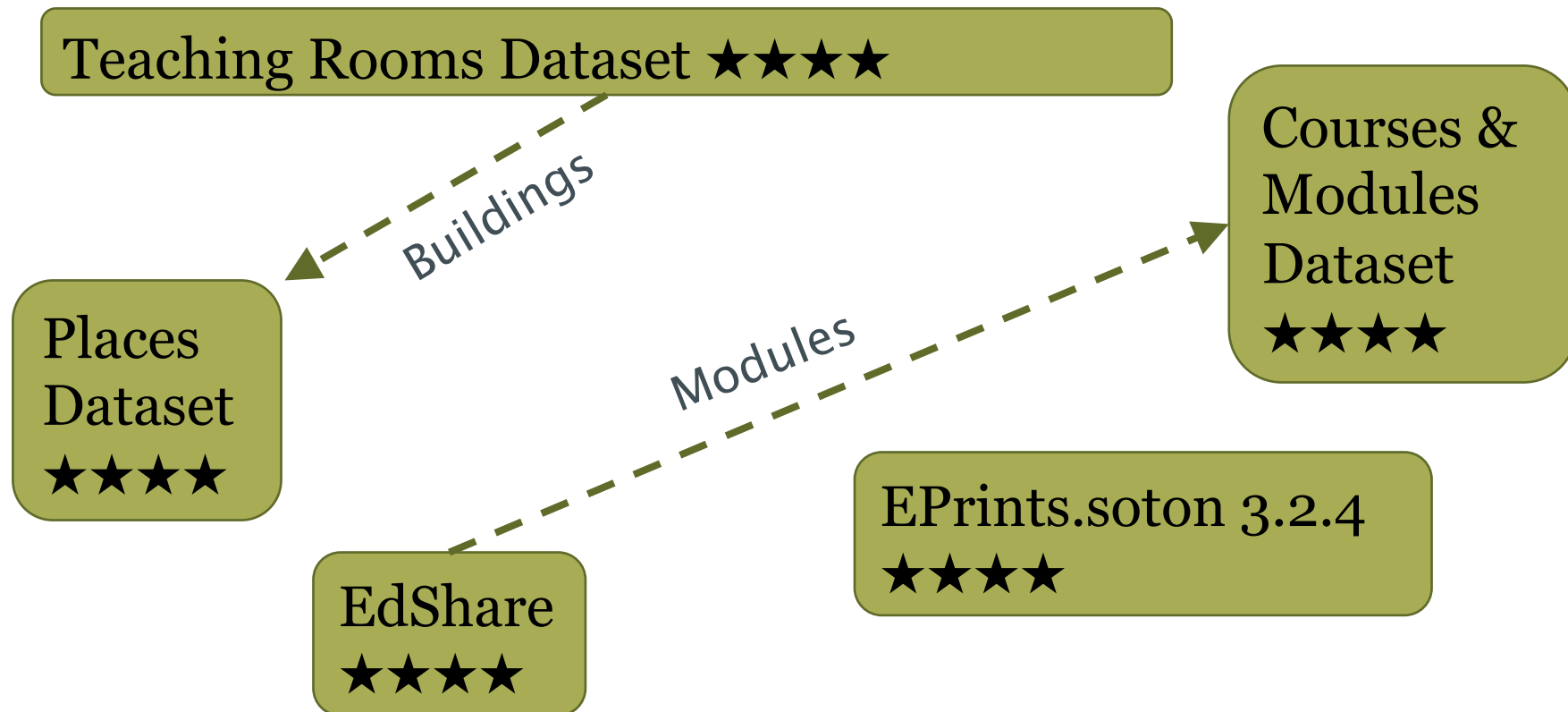
Courses &
Modules
Dataset
★★★★★

EPrints.soton 3.2.4
★★★★★

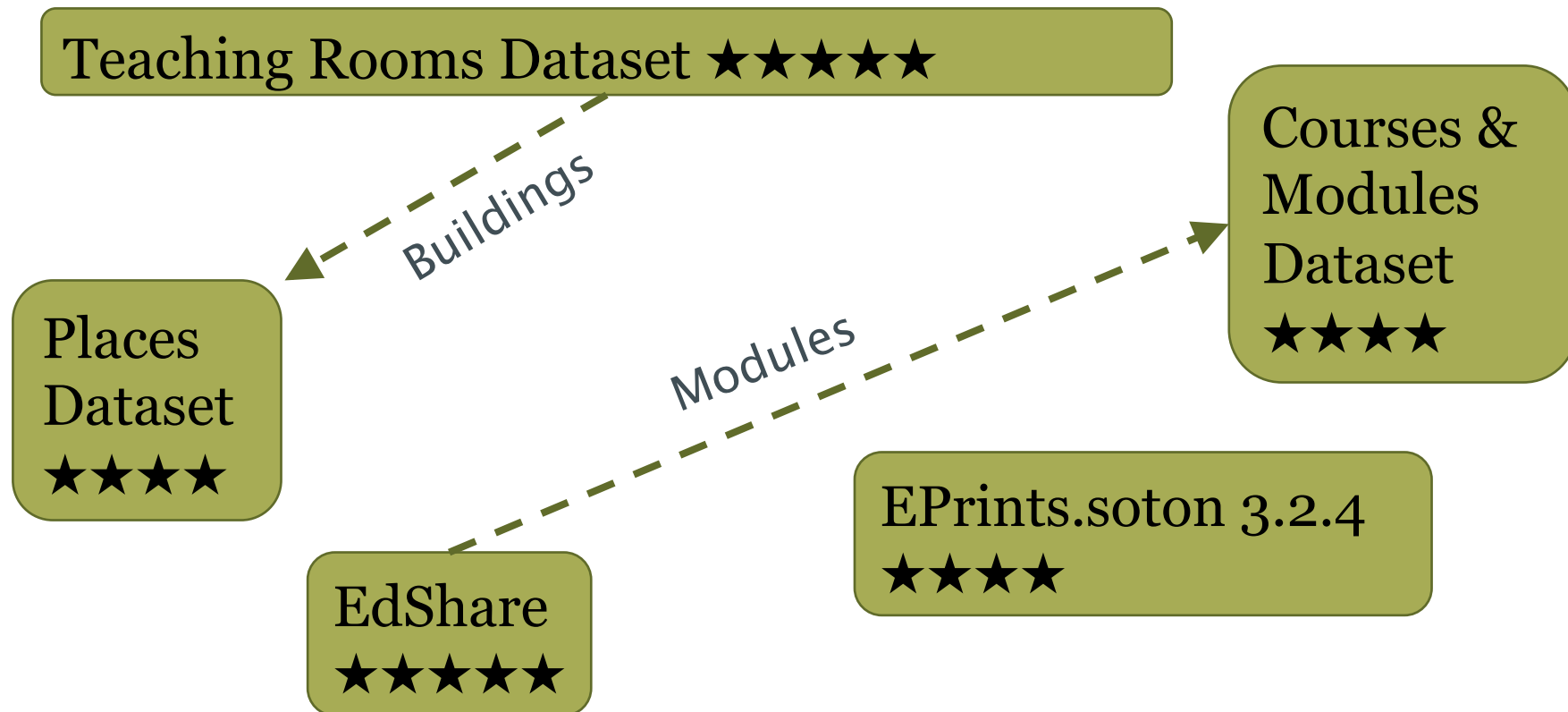
Easy Links



Easy Links



Easy Links: ★★★★★+★



Authoritative Open Datasets

EdShare

EPrints.soton

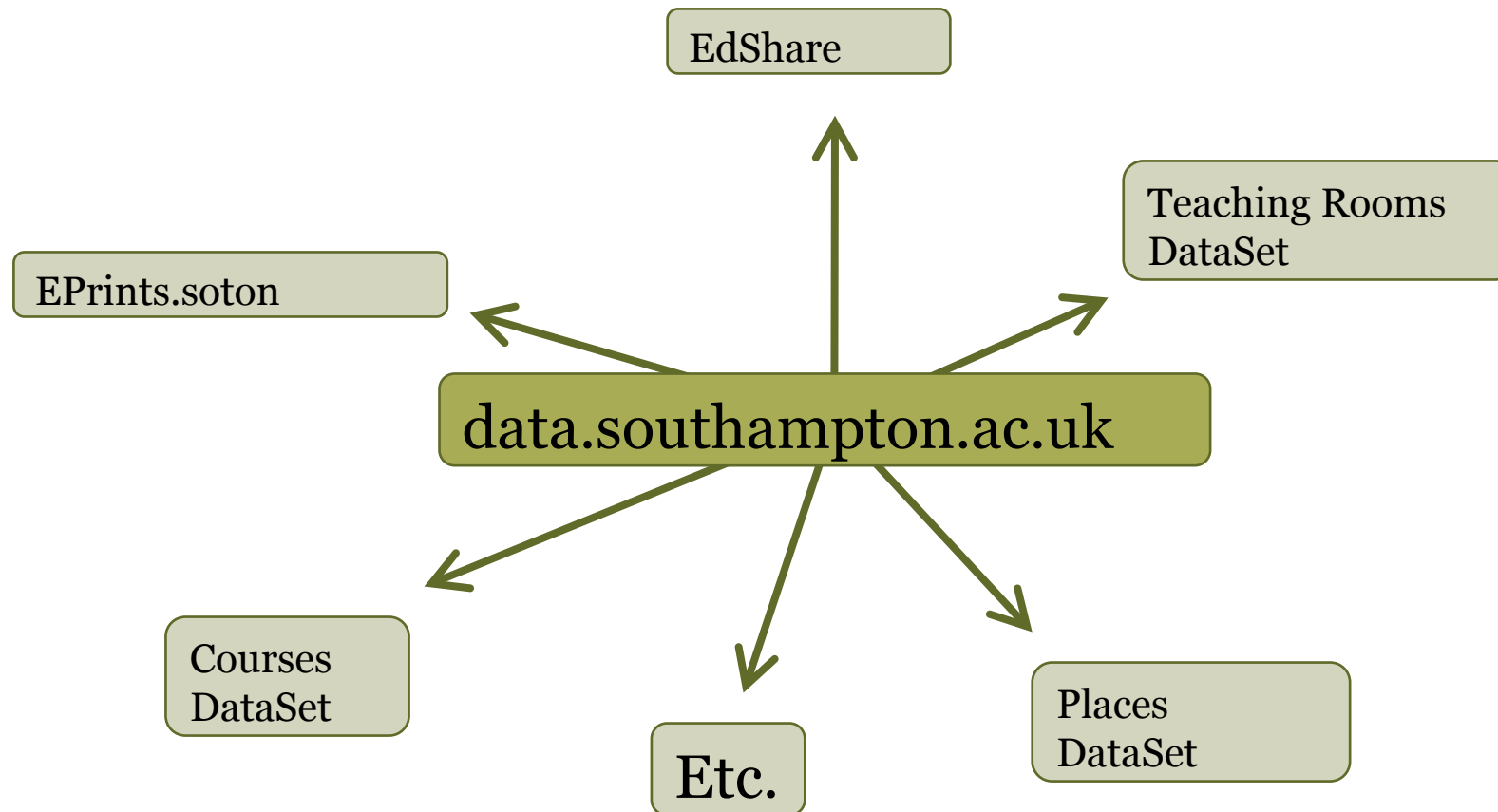
Teaching Rooms
DataSet

Courses
DataSet

Etc.

Places
DataSet

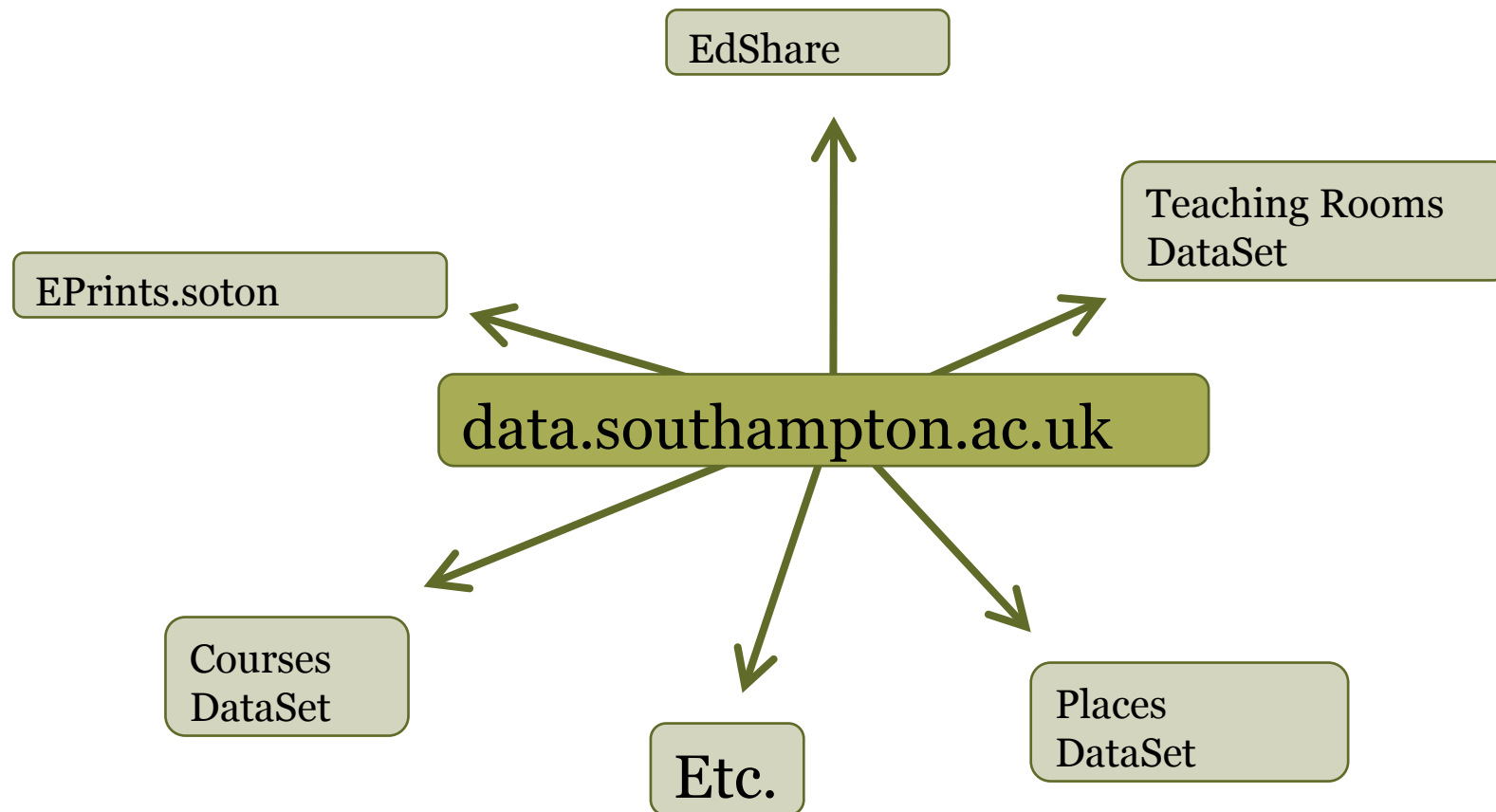
Catalog of Datasets



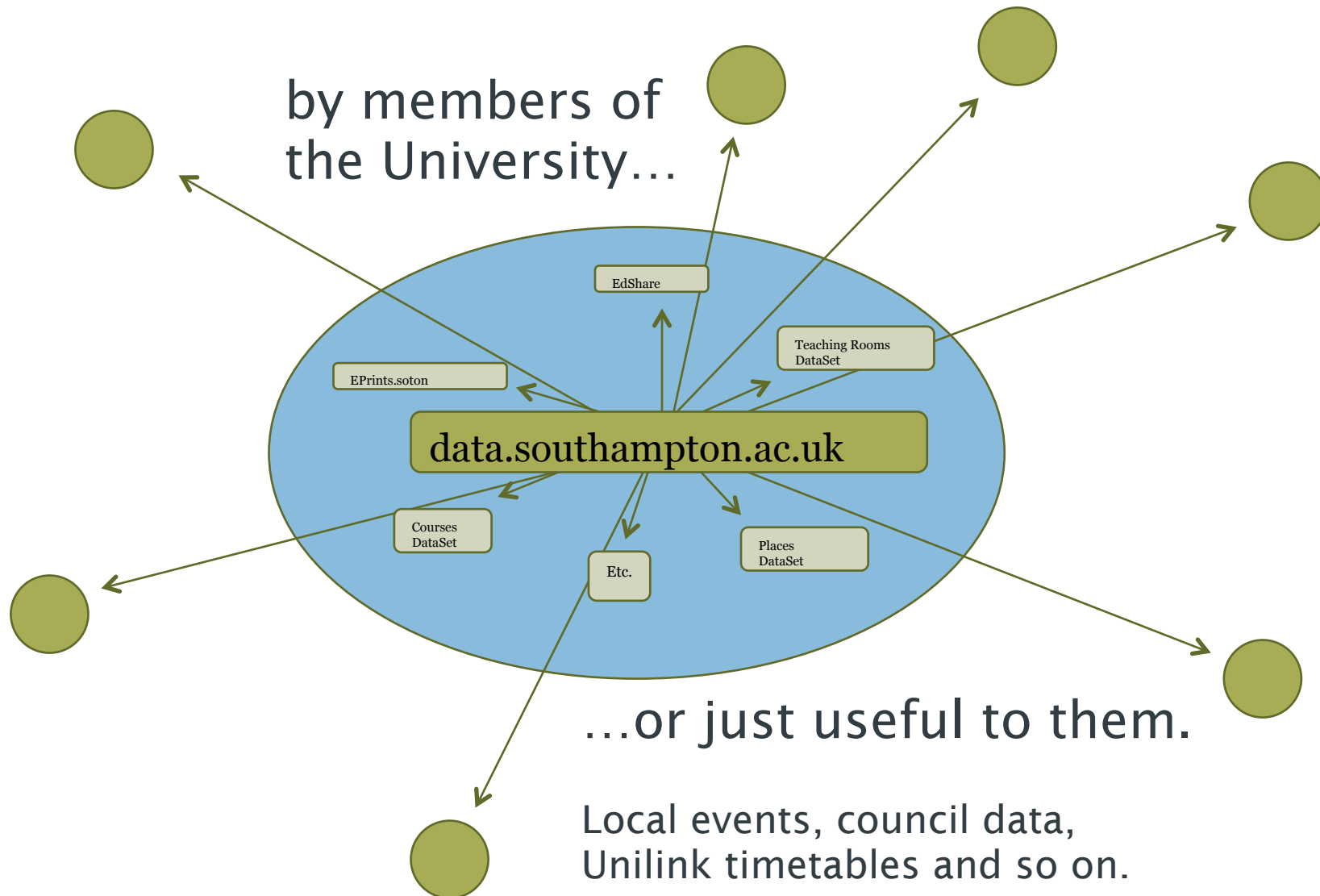
What to record?

- DataSet name & description
- License
- Update frequency & mechanism
- Corrections policy and instructions
- Dependencies
- List of known users
- What else?

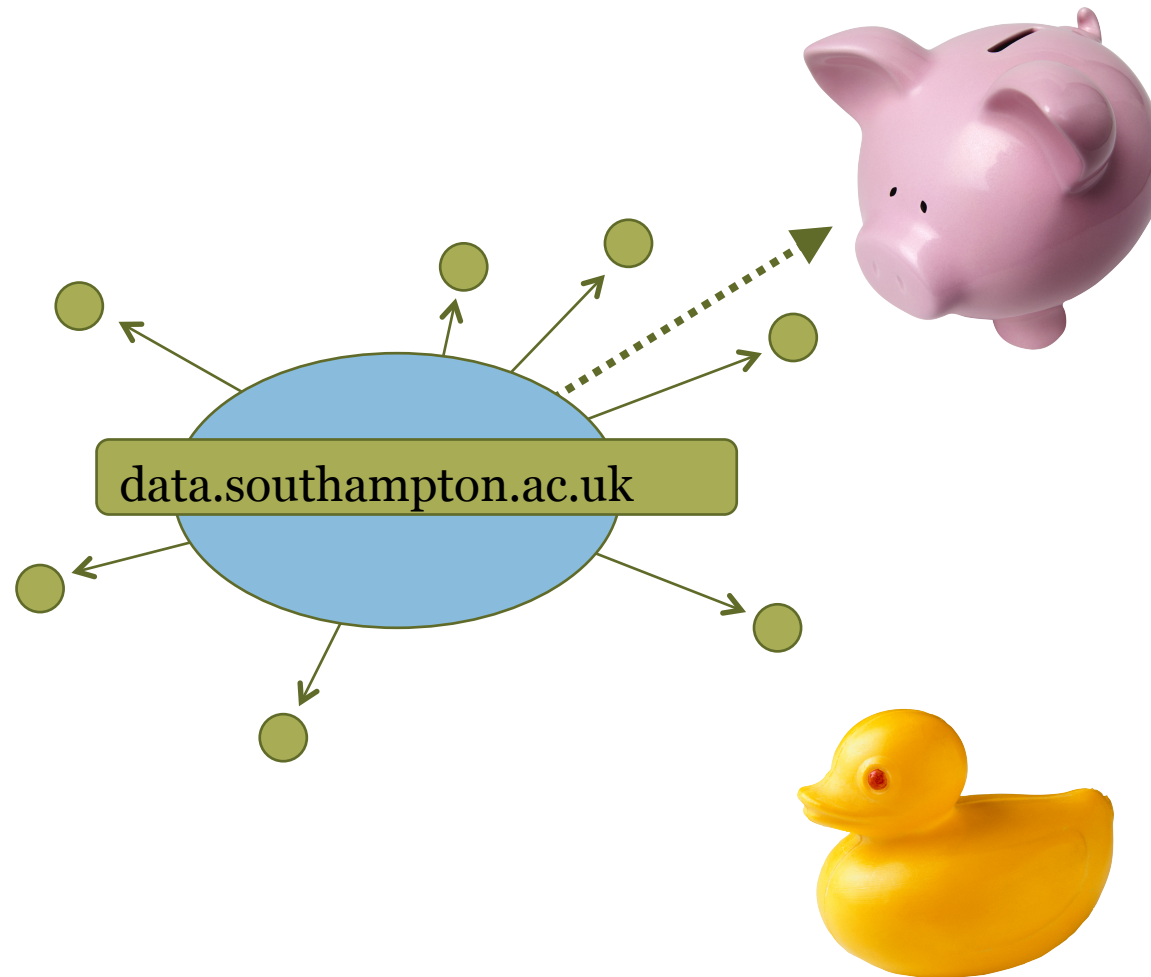
Catalog of Datasets



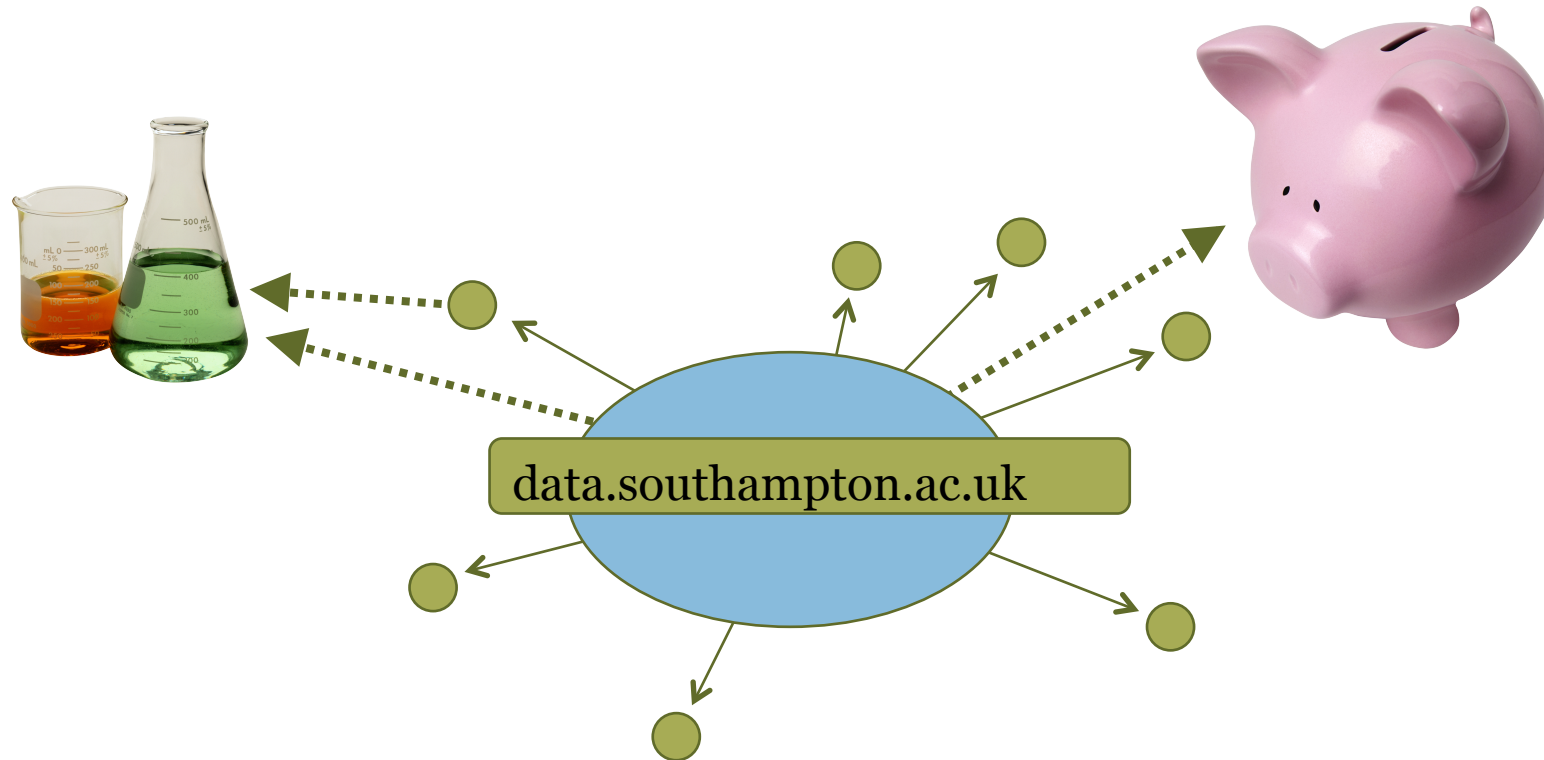
Non-Authoritative Datasets



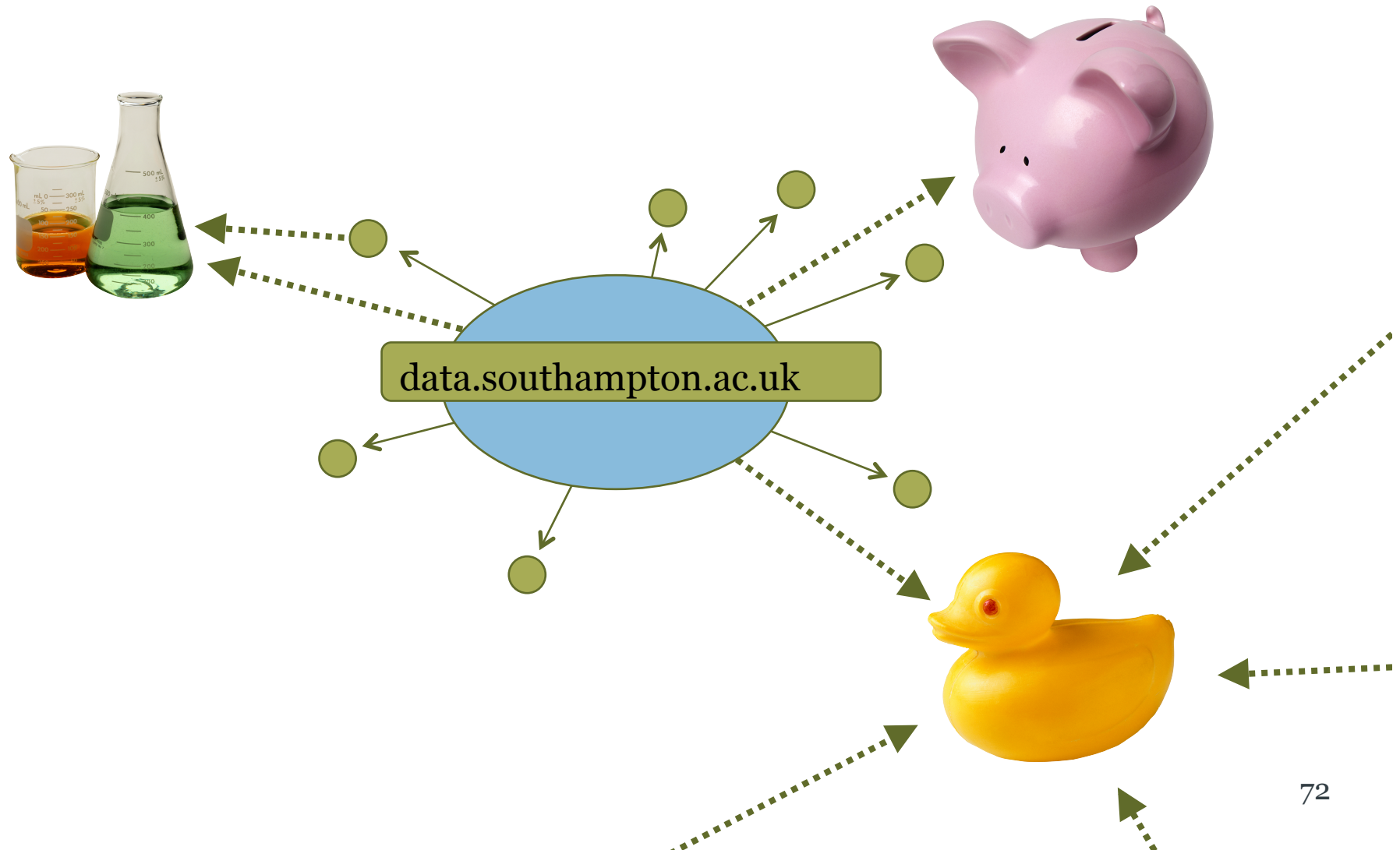
Catalog of known “Consumers”



Catalog of known “Consumers”



Catalog of known “Consumers”



Group Breakout

- What does it mean for you?
 - ★ Put your Data on the Web
 - ★★ Make it Available in a Machine Readable Format
 - ★★★ Use an Open, Standard, Format
 - ★★★★ Use an Open, Linked Data, Format
 - ★★★★★ Link Your Data to Other Peoples' Data
- Raw Data Now!
- Just Enough Ontology!
- Perfect is the Enemy of the Good!

Actions?

- data.southampton.ac.uk:
 - Web presence
 - Surface initial datasets
 - Community building
 - How to bring in the Southampton Research Environment?
 - Catalogue – bring up (gov) CKAN/ePrints?
 - southampton-partnership license
 - southampton.ac.uk license
 - Communication channel
 - Partnership channel

Further Reading

<http://www.flickr.com/photos/markhillary/337685031/>



Some Links

- <http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/>
- <http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>
- <http://www.w3.org/DesignIssues/LinkedData.html>
- <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>