Creating National Datasets

Christopher Gutteridge - @cgutteridge

- Previously; Lead Developer of EPrints.
- “Linked Open Data Architect” for University of Southampton.
- Benevolent dictator of data.ac.uk

Creating National Datasets

data.ac.uk

Creating ~~National~~ Datasets

data.ac.uk

Creating Useful Datasets

WHAT'S WRONG WITH THE STATUS QUO?

WHAT'S WRONG WITH THE STATUS QUO?

WHAT'S WRONG WITH THE STATUS QUO?

We can do better!

We can do better!

New technologies bring

- New efficiencies
- New possibilities

How it's normally done

- Agree XML schema
- Once per year send your XML to official body

- Official body provides either
  - Controlled access to services
  - Open access (web) tools to search & browse

Open Data

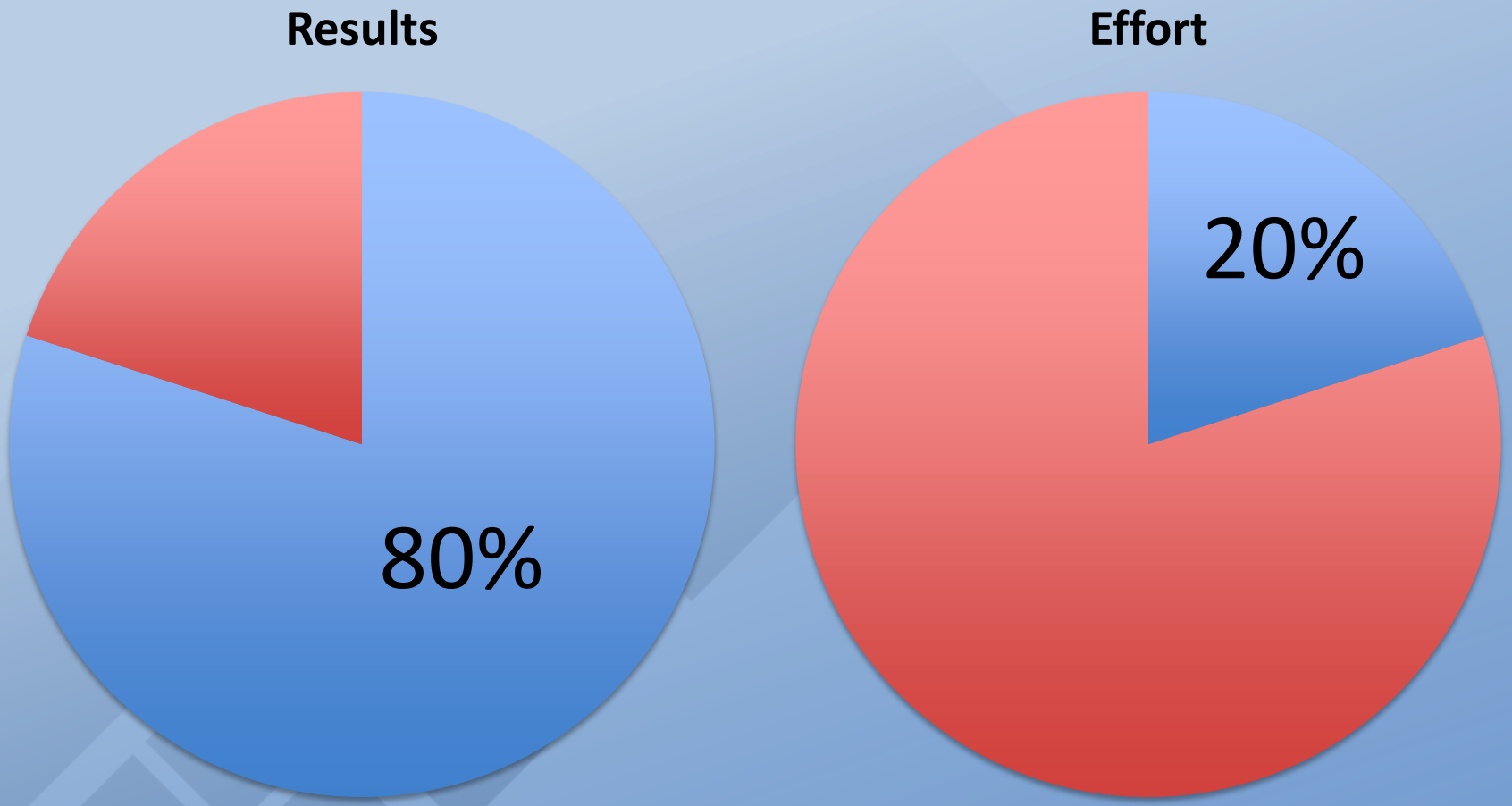
- Reusable Patterns
- Allows innovation
- Makes everybody's life easier
- Transparent
- Verifiable Provenance

Fruit Yoghurt (large)	0.85	0.75
Walkers Crisps	0.60	0.60
Kettles Crisps	0.70	0.70
Eat Natural Bar	1.10	1.10
Geobar	0.85	0.85
Croissant	1.05	0.90
Choc au Pain	0.90	0.90
Danish Pastry	1.25	1.05
Doughnut	0.90	0.75
Muffin	1.25	1.25
Cookie	1.00	1.00
Cup Cake	1.00	1.00
Traybake	1.00	1.00
Cake Slice	1.75	1.75
Cake Slice Premium	2.25	2.25
Honeybun Cake	1.50	1.50

	All	Student
	0.50	0.50
)	0.55	0.50
)	0.85	0.75
	0.60	0.60
	0.70	0.70
	1.10	1.10
	0.90	0.90
	1.25	1.05
	0.90	0.7
	1.00	1.00
	1.00	1.00
	1.75	1.75
n	2.2	2.25
	1.50	1.50
	1.00	1.00
	1.00	1.00
	1.00	1.00
	1.25	1.25
	0.75	0.75
s	0.70	0.70

≠

80/20 Rule



data.ac.uk

Case for OGL

- Open Government License for Public Sector
- Designed for UK public sector data.
- Protects against common concerns.

20%: Formats

- A little data is much better than no data
- RDF is easiest to extend and merge
- Tools can be created to map spreadsheets into RDF
- Common applications can export RDF
- Just don’t publish things that can’t be open data.

20%: Your end

- Publish what data you canon your organisation's website in a standard format (RDF or CSV)
- Create a simple machine-readable web-document stating
  - what license you're using
  - The URL of your Excel (or CSV, or RDF) document
  - What your university name & logo is.

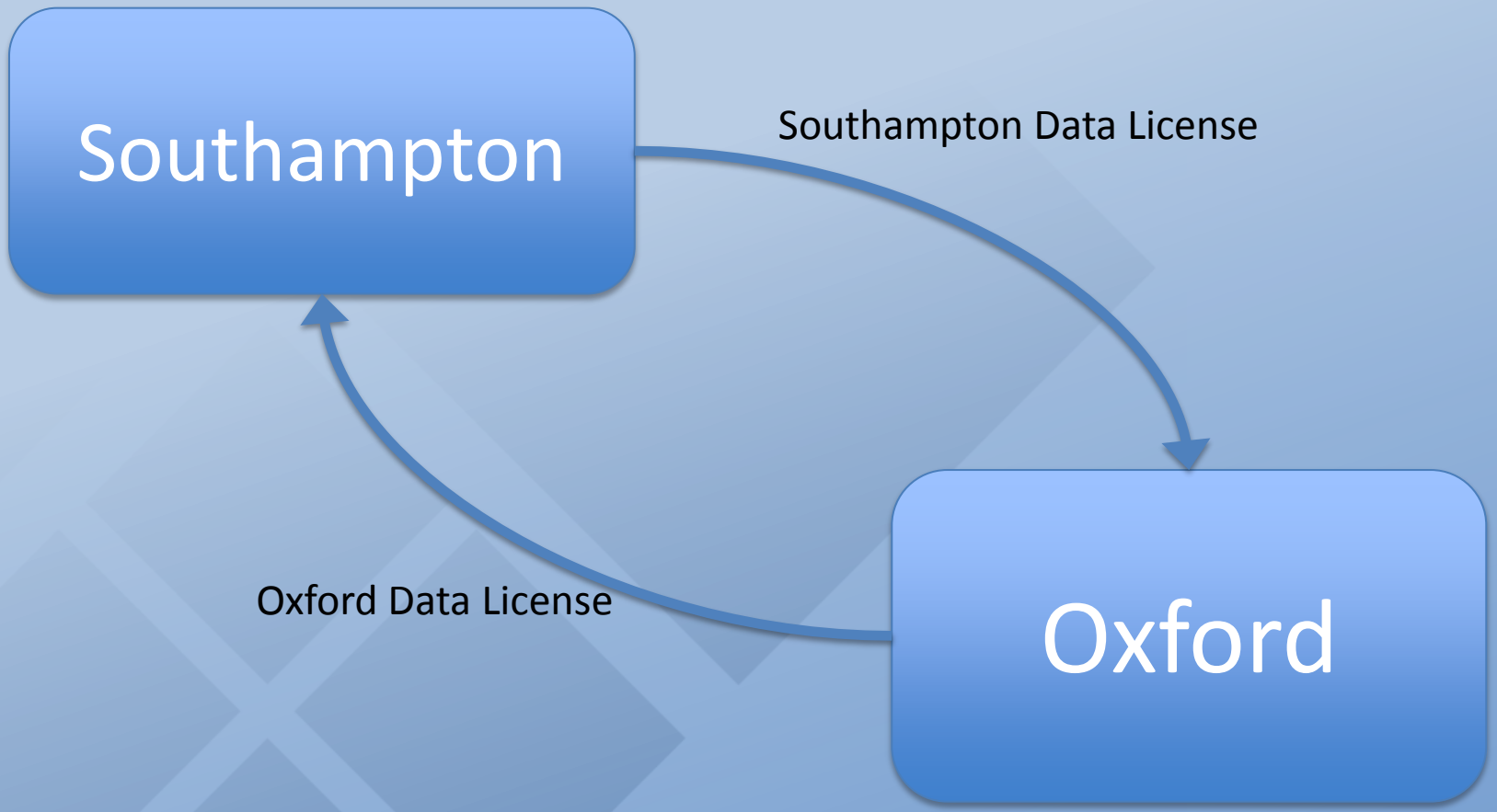
(we'll provide a template for these documents)

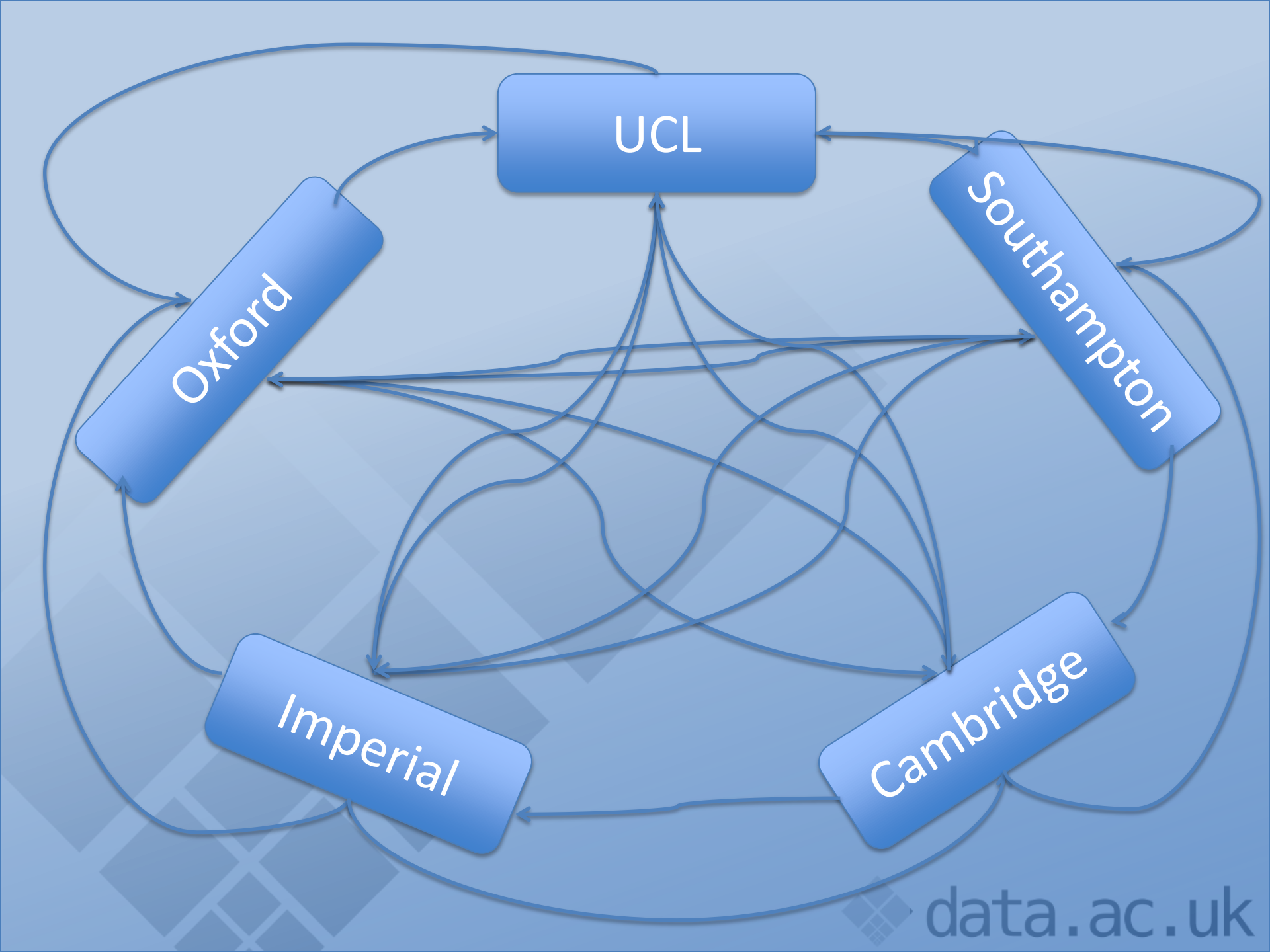
20%: KitCatalogue

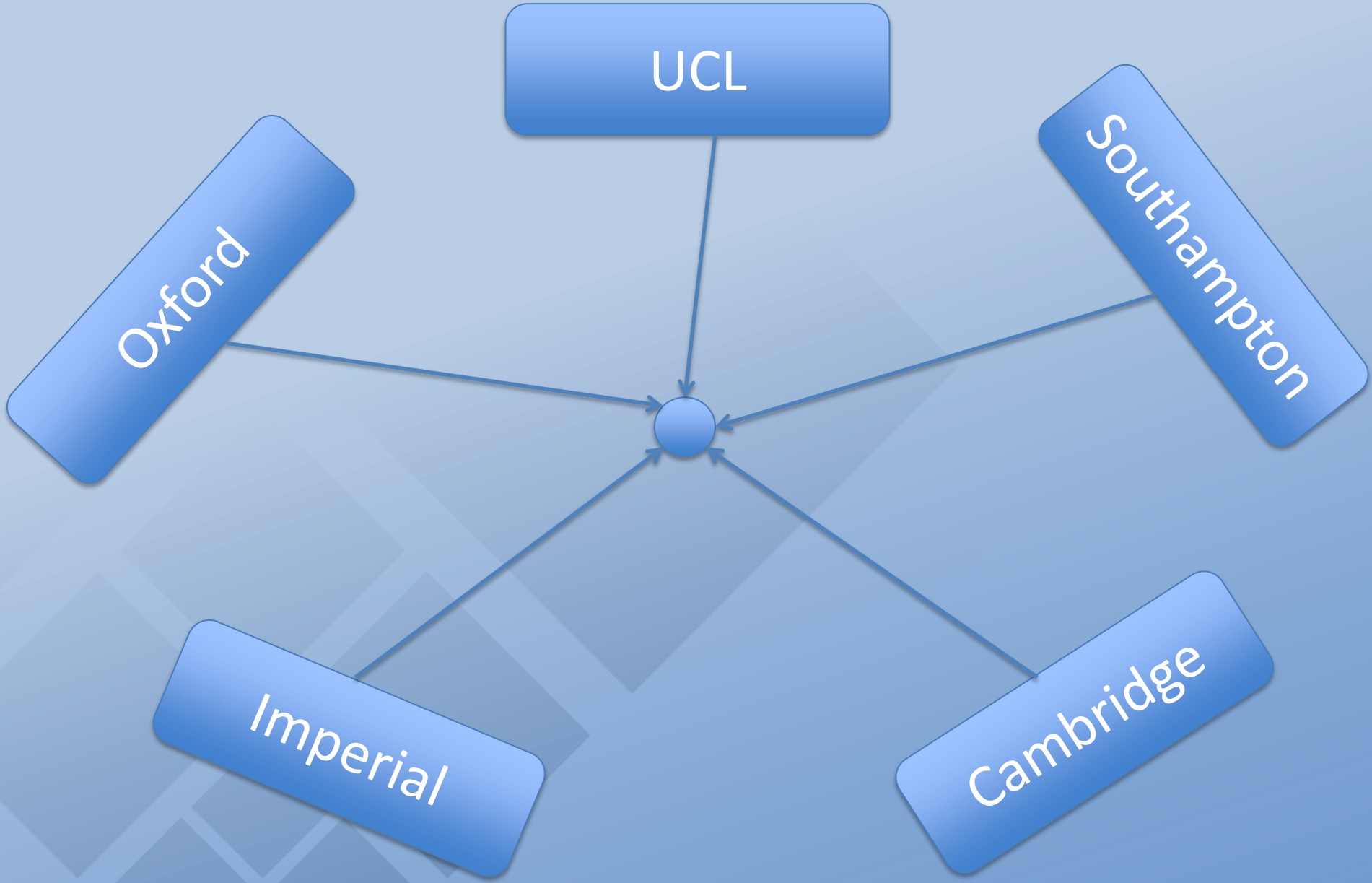
20%: Data.ac.uk

- Ask data.ac.uk to start harvesting it
- Once per day we
  - Download your data
  - Normalise it
  - Aggregate it with data from everyone else
  - Provide a search
  - Provide a download URL for the data as Excel, CSV and RDF

License Proliferation




UCL
Oxford
Southampton
Imperial
Cambridge


UCL
Oxford
Southampton
Imperial
Cambridge

DEMO:
www.events.southampton.ac.uk

DEMO: facilities.data.ac.uk

Science is International

- Funding isn't.

- data.ac.uk
  - scope is UK Academia
- Linked Open Data
  - scope is the World Wide Web

Useful aggregations for Equipment & Facilities data

- UK Academia
- South-coast public sector
- Universities in N.Ireland + R. of Ireland
- Hampshire Businesses

Other Useful Datasets

- Staff Expertise
- Vacancies
- Organisational Structure
- Buildings & Campuses
- Publications
- … research data

Support from Tools

- ABW (Agresso)
- KitCat
- …

data.ac.uk

Medium Future

- Auto discovery of datasets
- Semantic Web

/.well-known/void

http://www.ox.ac.uk/.well-known/void

data.ac.uk

Picture Credits

- Status Quo: Mtthshksm / Wikipedia Commons / GFDL
- Croissants: jnunoferreira /Flickr / CC-by 2.0
- Menu: data.soton.ac.uk / Open Government License