

Office for National Statistics

Accessing data from commercial sources

Kate Davies

Survey and Economic Indicators

Darran Tucker

Commercial Data Access Team

Producing statistics is changing

DATA COLLECTION

Generated

Surveys

Census

Sources

- The web
- Surveys

Challenges

- Effort vs. reward
- Legislation unclear in UK so larger companies cautious

Shared

Government

Sources

- Major Gov departments, DWP, DfE, DfT, HMRC
 - e.g. VAT register

Challenges

- Accessibility, legislative, cultural and technology barriers

Accessed

Industry

Potential sources

- Point of sales data, footfall
- User contributed data
- Social media
- Financial data flows

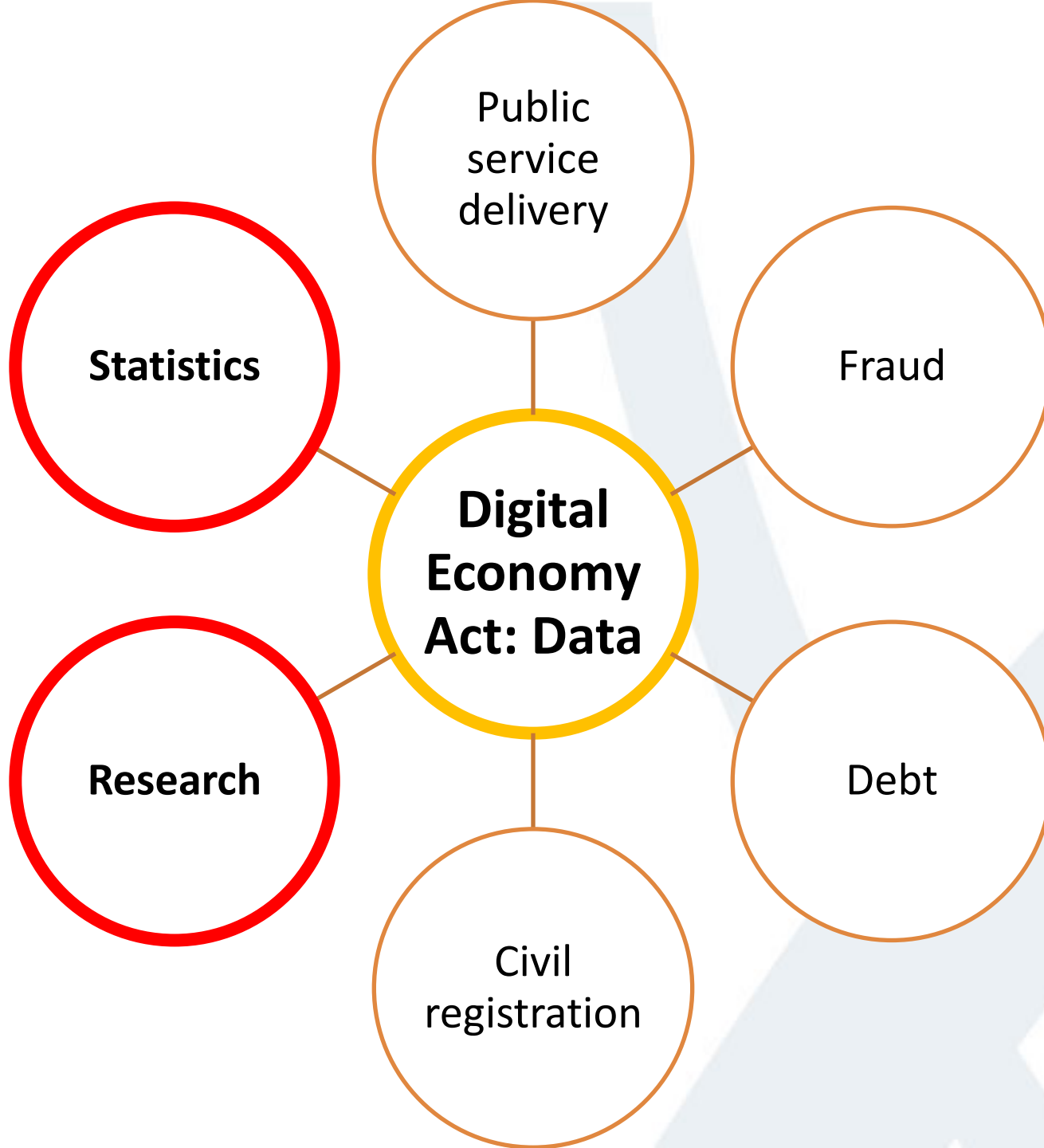
Challenges

- Knowing what is available
- Engaging and evaluating supply
- Reliability of supply
- Legislative...

Yesterday



Today and tomorrow...



Public
service
delivery

Fraud

**Digital
Economy
Act: Data**

Debt

Civil
registration

Research

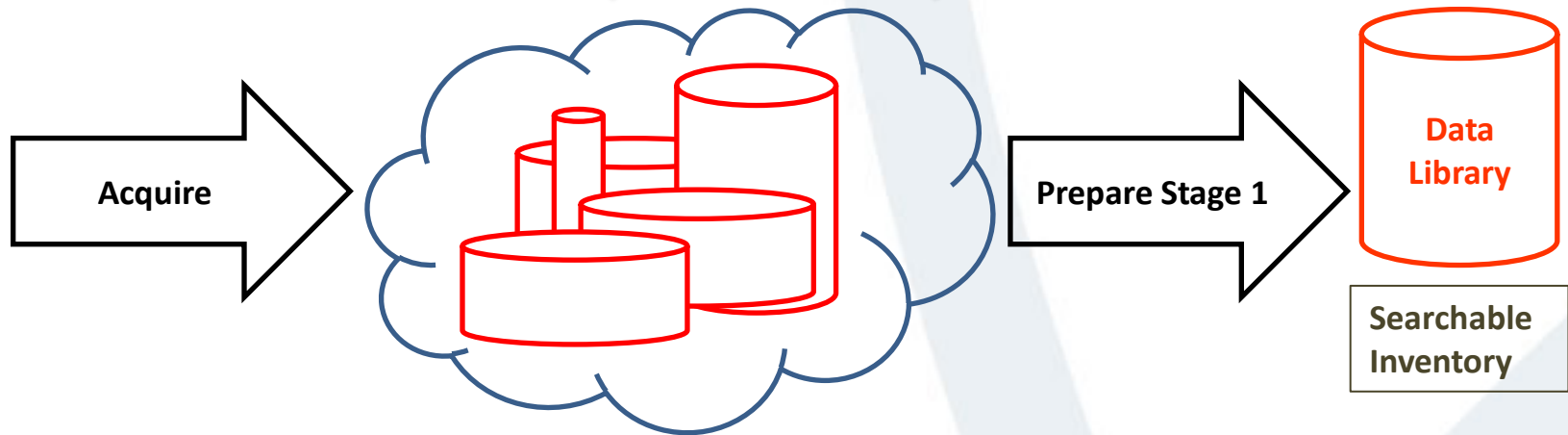
Statistics

Access non survey data

- A cross ONS service capability for all non survey data acquisition and preparation
 - Governmental: e.g. administrative
 - External: e.g. commercial, third sector
- Accelerate ONS' access to critical data infrastructure
 - linked data products
 - Wider data services based on user needs
 - Enabling users to maximise data utility, to better enable innovation and underpin better statistics

Acquire and prepare

Goal: Acquire once Use many time



Principles:

- Data is for Statistics and Research Purposes across ONS accessed via partnerships
- Investigate content and format of 'raw' data sample
- Minimise burden on data supplier (ingest data 'as-is' wherever possible)

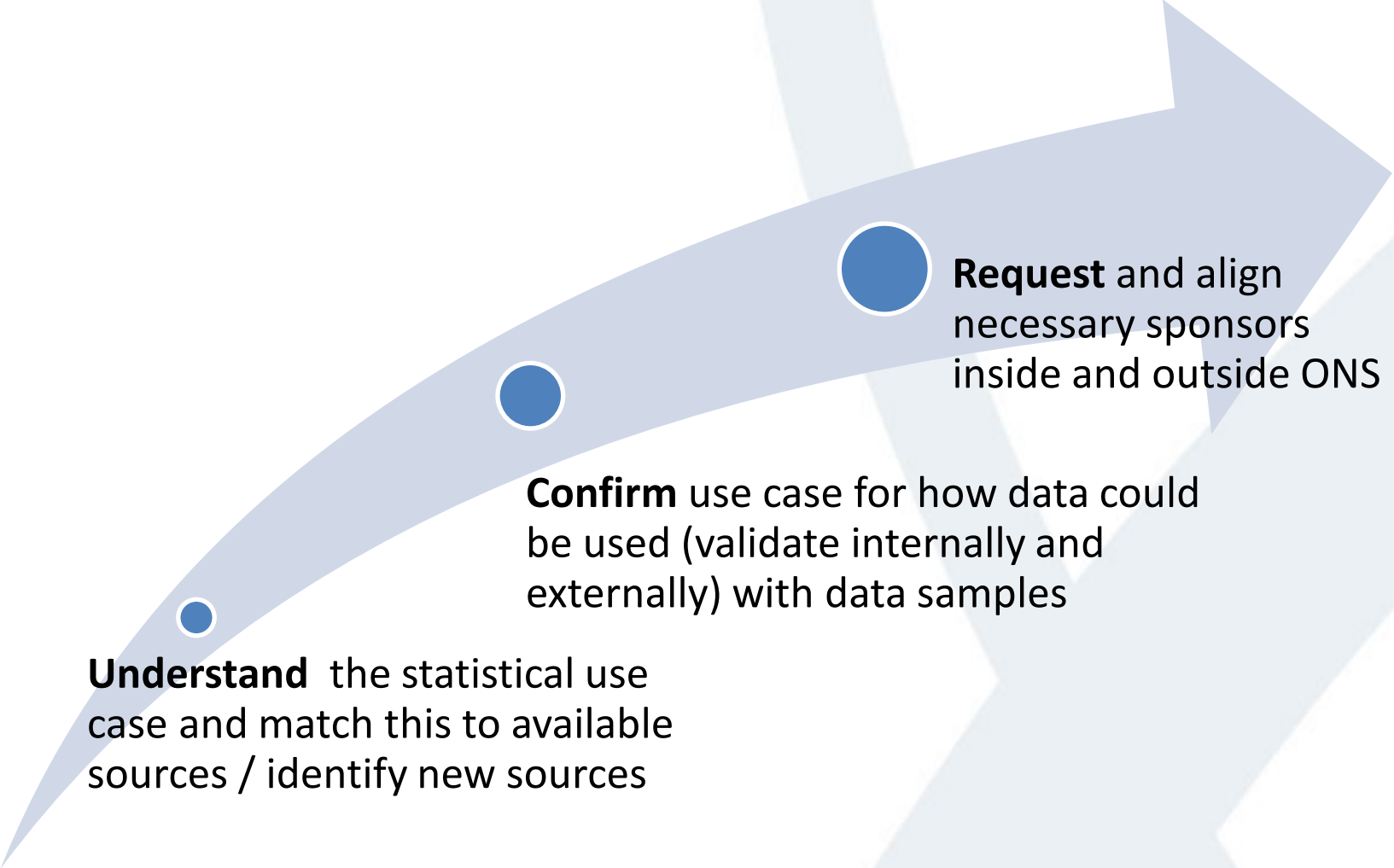
Processes:

- Identify what data are available from commercial entity
- Determine Content, Basic Quality, Security Classification
- Create Metadata

Outcome:

- An inventory of datasets, accessible inside ONS
- A single data repository

Acquire: data sharing made easy

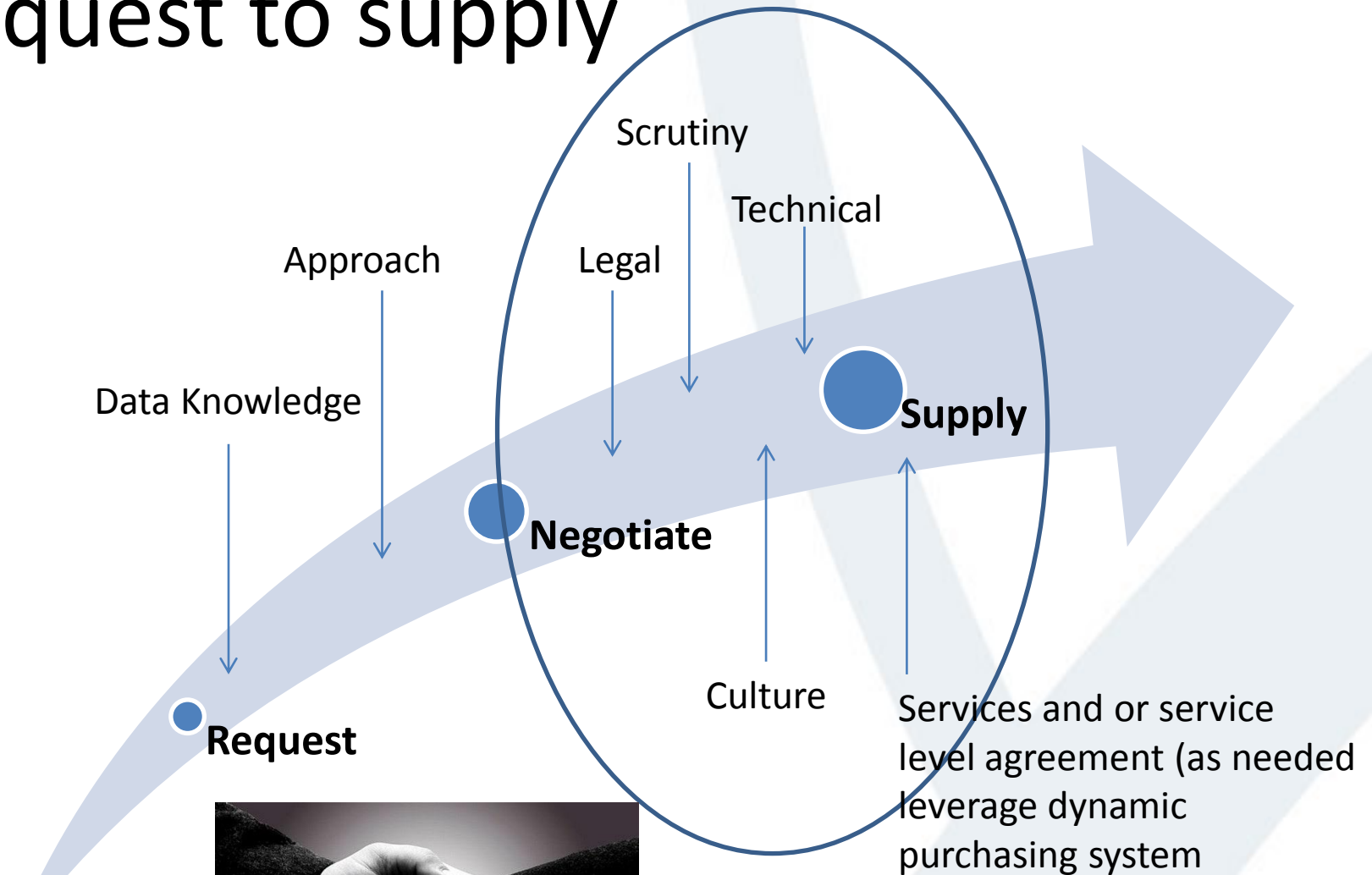


Understand the statistical use case and match this to available sources / identify new sources

Confirm use case for how data could be used (validate internally and externally) with data samples

Request and align necessary sponsors inside and outside ONS

Request to supply



Data Partnership Agreement

Monthly Commodity Index

- What did we agree?
 - Publish stats from MCI with the four category breakdown in £'000s
 - Replicate by taking data straight from sales systems
 - Use sales system to add more detail (next level down the COICOP) Need to map product codes to coicop (can this be done if ever changing code set is used)
 - Explore BRC categories and how we could best use this.
 - Internet split
 - Channel (app, web, mobile sales)
 - Geographical split

Proposal

- Together we refined the proposal with a number of retail organisations who then sought buy in from their management chains
- Advise all MCI respondents that we're doing it and they're welcome to join in.
- Increase MCI sample.
- After piloting. Advise all RSI respondents. (snow ball out).

Issues

- Collection method
 - Data feed / API / Spreadsheet / CSV but wont be survey....
- Definitions of categories.
 - Retailers have dynamic coding frames we will something more stable. Would require continual mapping to COICOP by each retailer or similar classification.
- Will highlight difference between accounting and sales systems.
- Commercial confidence
- ONS independence / no individual retailer gains commercial advantage, without others having the same opportunity.
- We would gather for statistical purpose only.
- Publish Vs closed pilot.

Working stages

- MCI £'000s data available (implemented August 2017)
- Retailers replicate from their sales data.
 - How to handle sales v accounts system differences?
- ONS produces analysis replica MCI analysis
- Retailers evaluate
- Define next step - what categories do we want as next granular level.
- Retailers provide further sales data based on these categories.
 - Do we want to try and do internet at the same time?
- ONS feedbacks analysis
- Stock take where next.

Stakeholder engagement

- Write to MCI respondents before publishing
- Take plan to BRC
- Retail event March 2018 when pilot has run for a few months.
- Plan wider RSI engagement beyond that point.