



Office for  
National Statistics





# UN GWG GP Committee



# Tech Partners



# Global Platform: How we got here



**UNSC Remit**  
World Data Forum  
January 2017



**Our Vision**  
A Global collaboration to harness the  
power of data for better lives  
[14/07/2017]

## WORK STREAMS

### WS1 Governance

Define the business operating model for the GP and make the business case for the GP taking into account the legal structure of the GP, legal and regulatory compliance, overall governance, structure of partner agreements, the customer network and the options for long-term sustainable financing



### WS2 Data Policy and Common Technology Framework

Draft and seek endorsement by the statistical commission of a data policy framework for governance and information management including ethics, privacy, confidentiality and security, which will shape the workings of the GP. Seek endorsement for a common technology infrastructure to enable its delivery



### WS3 Communications and Best Practice Sharing

Seek endorsement by the Statistical Commission of a framework for communications and best practice sharing for the GP. Overall progress towards delivering the GP will be widely communicated to encourage new partners and users to join this initiative.

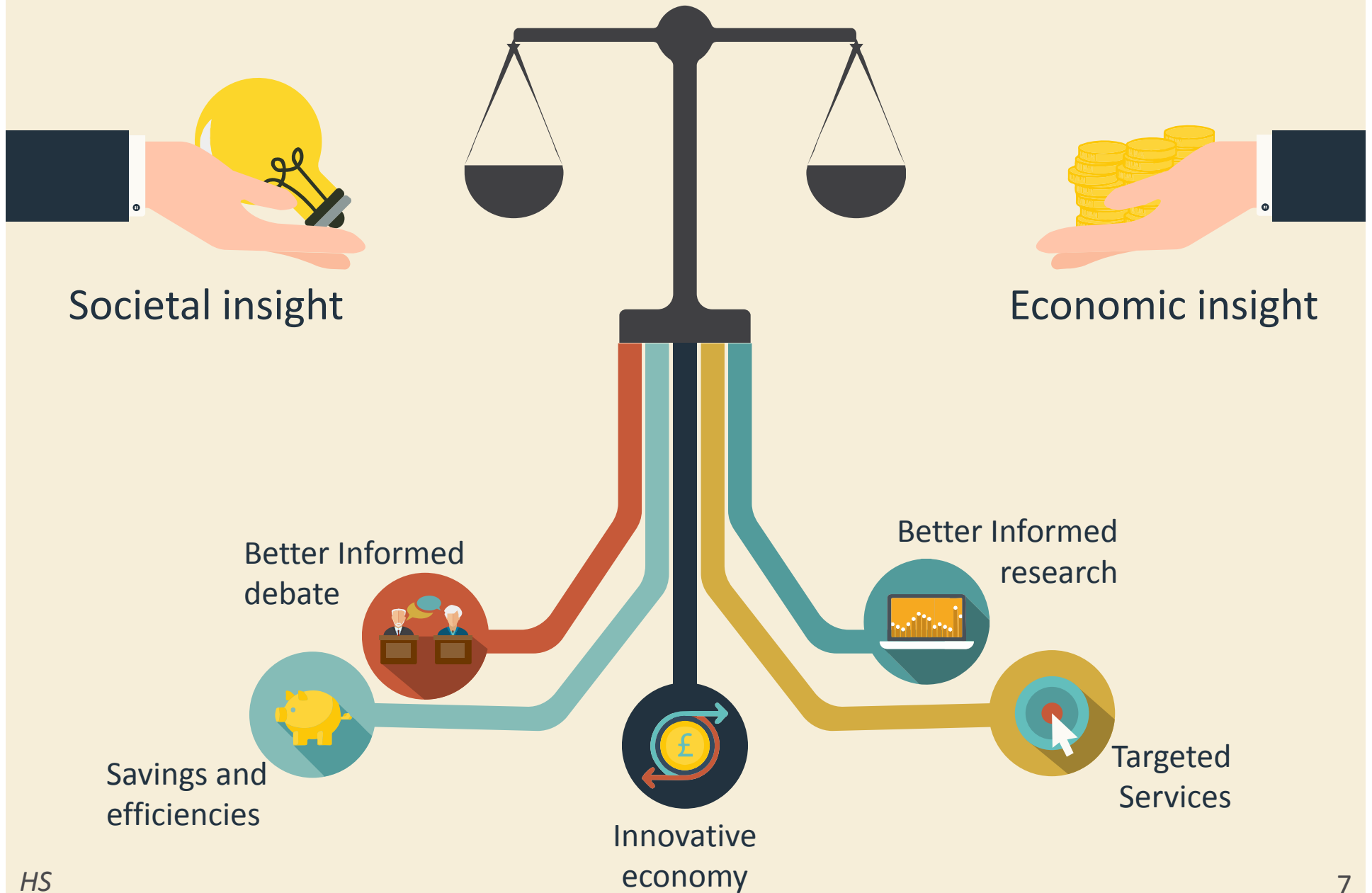


# Joining up with other data initiatives



# Statistical Modernisation @ONS

# Data for ONS, Government and Researchers

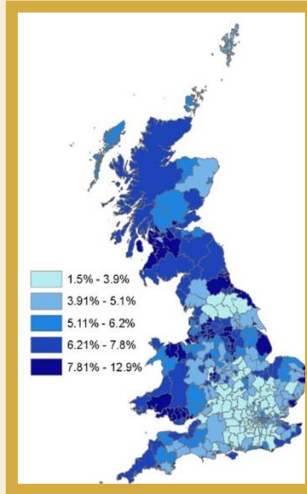




# Targeted Service Delivery

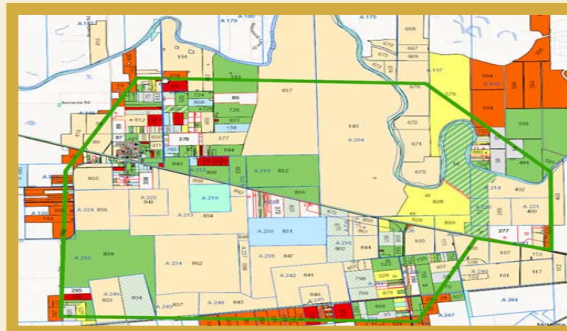
## Clustering

- Better policy decisions based demographics and geography



## Better Informed Public Debate: Migration

- Confused and mixed messaging from different sources
- Collaboration between ONS and OGDs
- Clearly present and explain all information



## Service integration: DfE/CLG Land Availability Tool

- Identify possible sites for new free schools

## Reduce reoffending rates

- Assess interventions



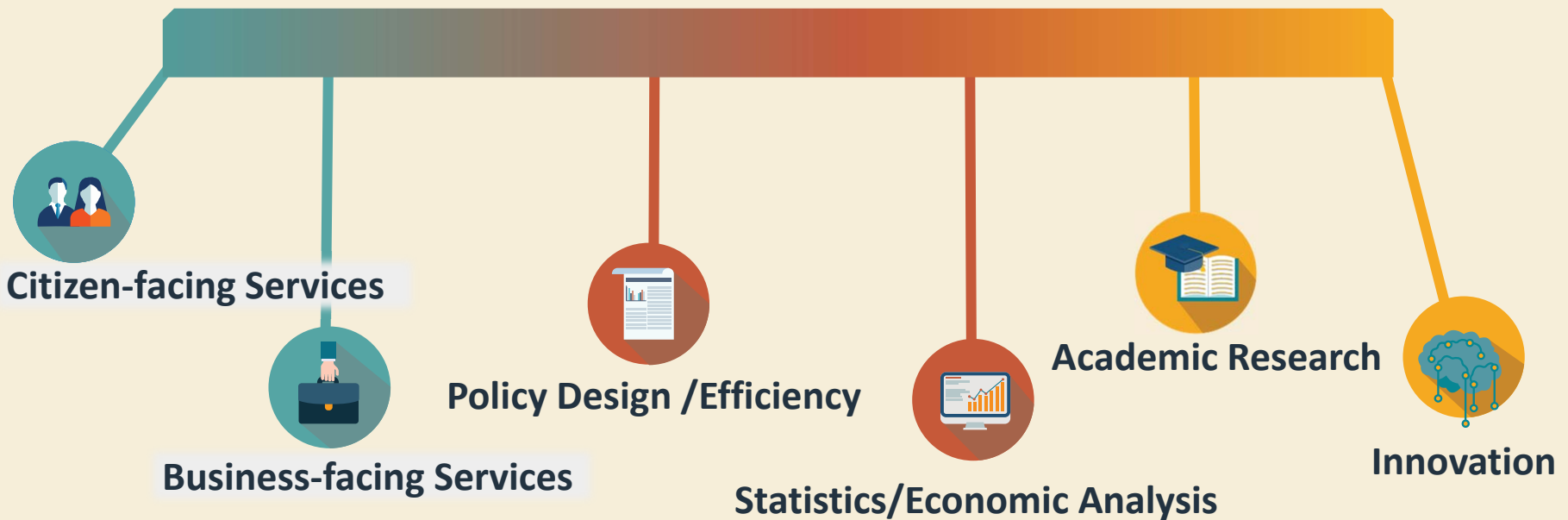
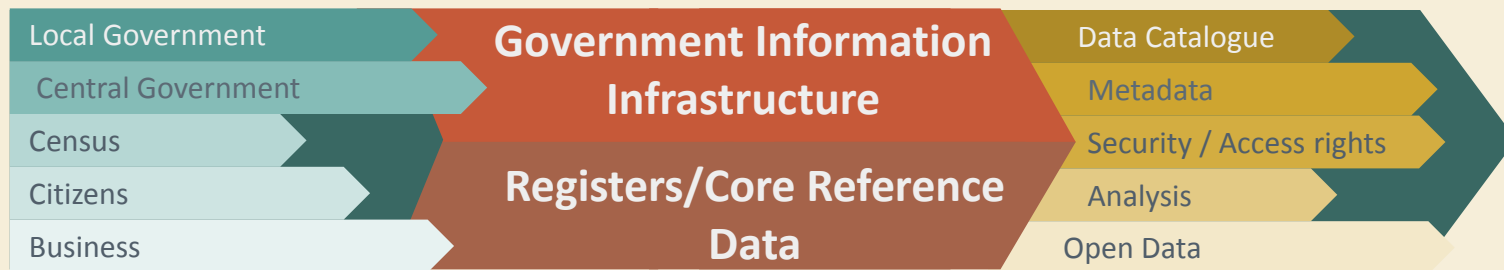
## Better policy decision : Flow of funds

- Closer monitoring of financial flows
- Reduce risk of another financial crisis
- Asset and liability position by sector
- One sectors liabilities are spread across economies





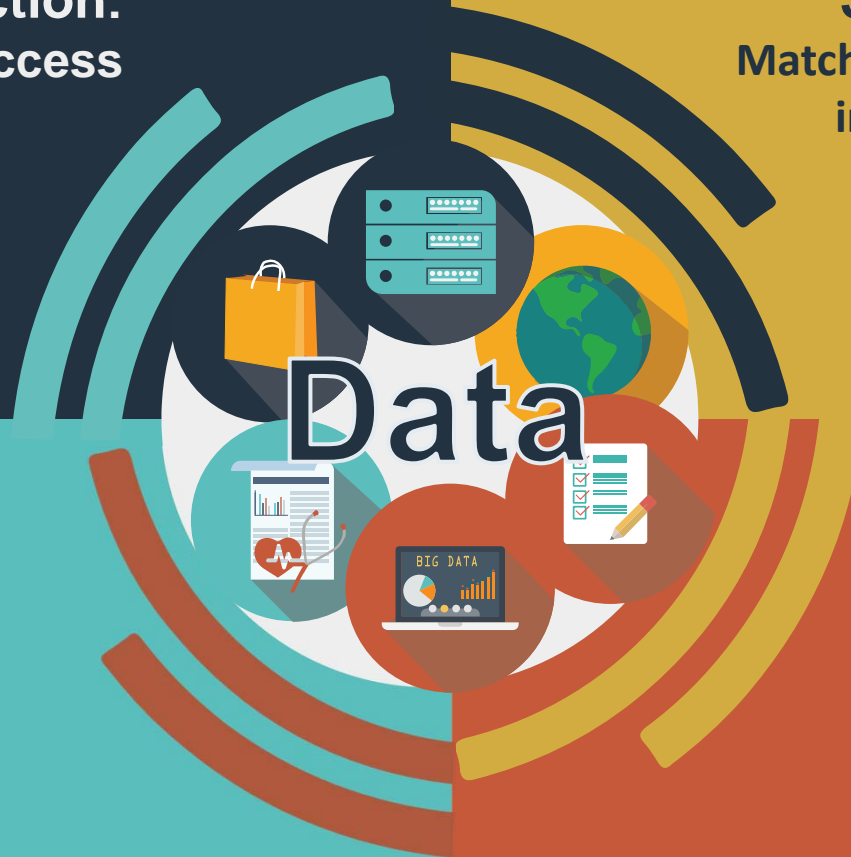
# Data Infrastructure: delivering a critical resource



# The challenge: multiple user lenses across data estate

**Statistical production:**  
Secure, in-house access  
ONS staff





**Statistical services:**  
Match, link, anonymise data:  
internal and 3<sup>rd</sup> party use



**Statistical research:**  
Working in partnerships, project-  
based access

**3<sup>rd</sup> party disclosure:**  
Accredited research and  
statistics

# Data use and access

IDENTIFIED   
 SAFE   
 UNRESTRICTED ACCESS   
 CONTROLLED ACCESS 

**DATA IN**

ACQUIRE



**STATUTORY**  
 SRSA (inc ISOs)  
 DEA  
 STA (business)  
 RSA (registration)  
 VAT/finance acts

Voluntary Surveys  
 Non-controlled admin data  
 Commercial partnerships  
 Open data

**NON-STATUTORY**



PREPARE

METADATA



**DATA LIBRARY**



STATISTICAL METHODS DEVELOPMENT

ANALYSE

STATISTICAL PRODUCTION

STATISTICAL RESEARCH

STATISTICAL METHODS ADVICE

3<sup>RD</sup> PARTY SERVICE

3<sup>RD</sup> PARTY DISCLOSURE

INFORM

STATISTICAL RELEASES

AD HOC OUTPUTS

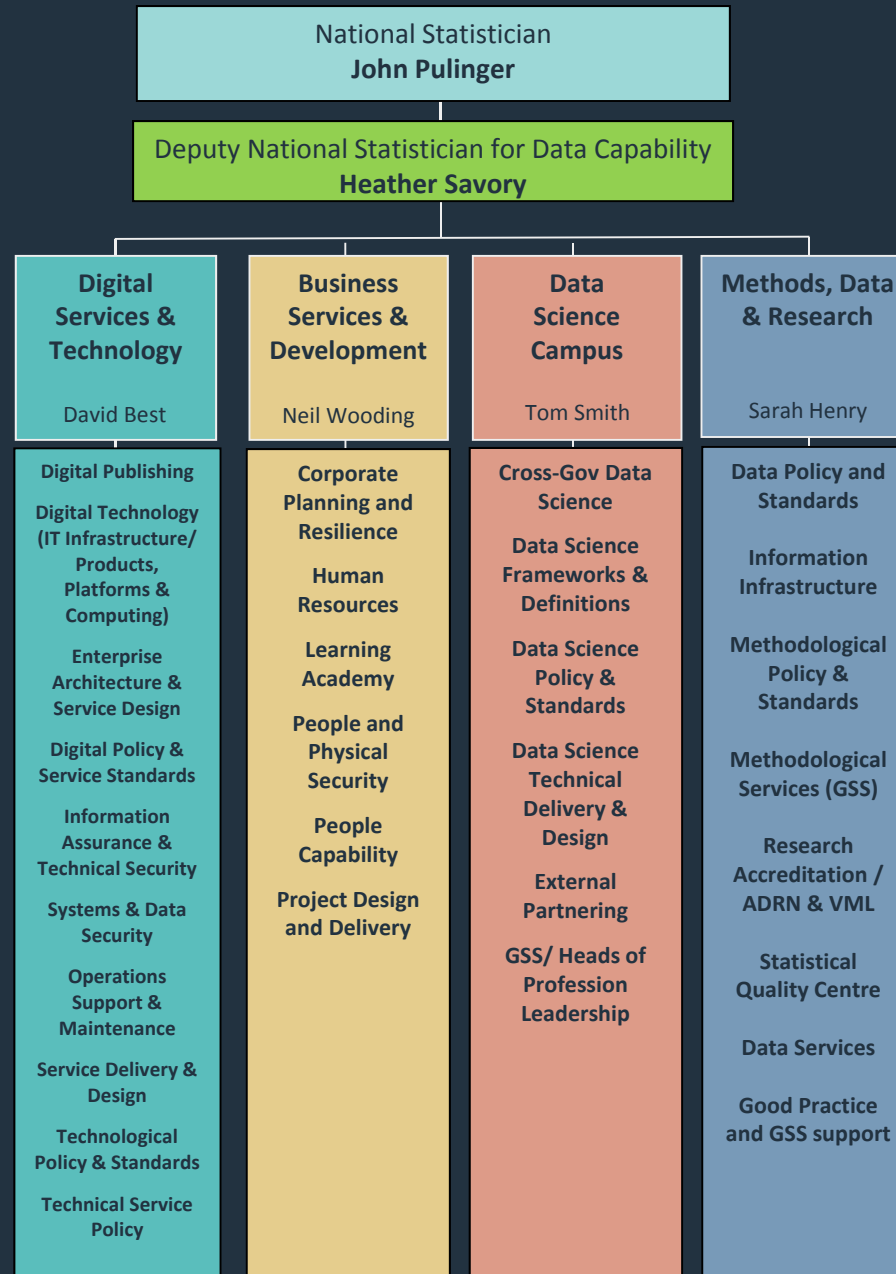
INFORM POLICY

DEVOLVED STATISTICS

ACCREDITED RESEARCH OUTPUT

**DATA OUT**

# Data Capability at ONS





# National Statistician's Data Ethics Advisory Committee (NSDEC)

## ETHICAL PRINCIPLES

### Public Good

The use of data has clear benefits for users and serves the public good.



### Confidentiality, data security, consent

The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.



### Methods and Quality

The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.



### Legal Compliance

Data used and methods employed are consistent with legal requirements such as the DPA, the Human Rights Act, the SRSA and the common law duty of confidence.



### Public views & engagement

The views of the public are considered in light of the data used and the perceived benefits of the research.



### Transparency

The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.



## Active Learning



## Experimentation

### Apprenticeship in Data Analytics

- 2 year programme
- Level 4 Diploma-Data Analytics

### MSc Data Analytics: Government

- September 2017
- Multiple academic partners
- Framework reviewed by GSS

### Government Data Accelerator

- Open to Public sector staff
- Run with GDS & GO-Science
- Network “hub” @ DSC

### Research Teams

- Launched September 2016
- 24 FTE
- 60 FTE projected March 2018

### Partnerships

- MoUs agreed including Universities (Cardiff), research institutes (Alan Turing), international statistical institutes (Stats Netherlands)
- Collaboration with national & devolved government, e.g. DEFRA, DCMS, DFID, Welsh Government

# ONS and National Addressing

Local Authorities



Royal Mail



Ordnance Survey

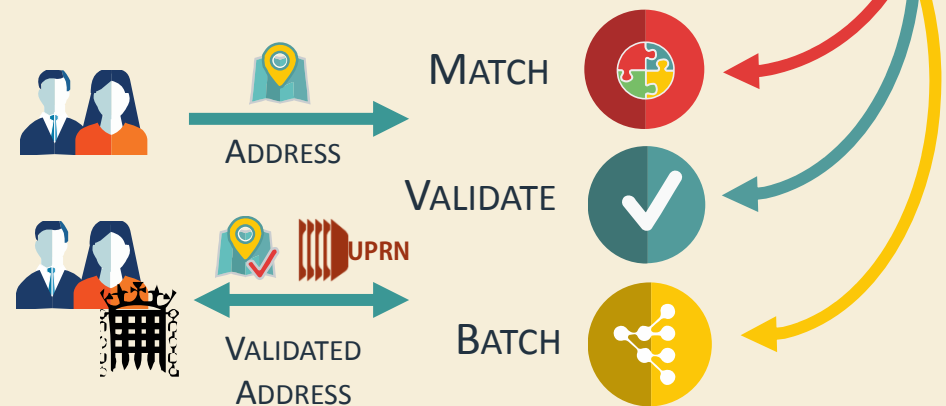


GeoPlace → UPRN

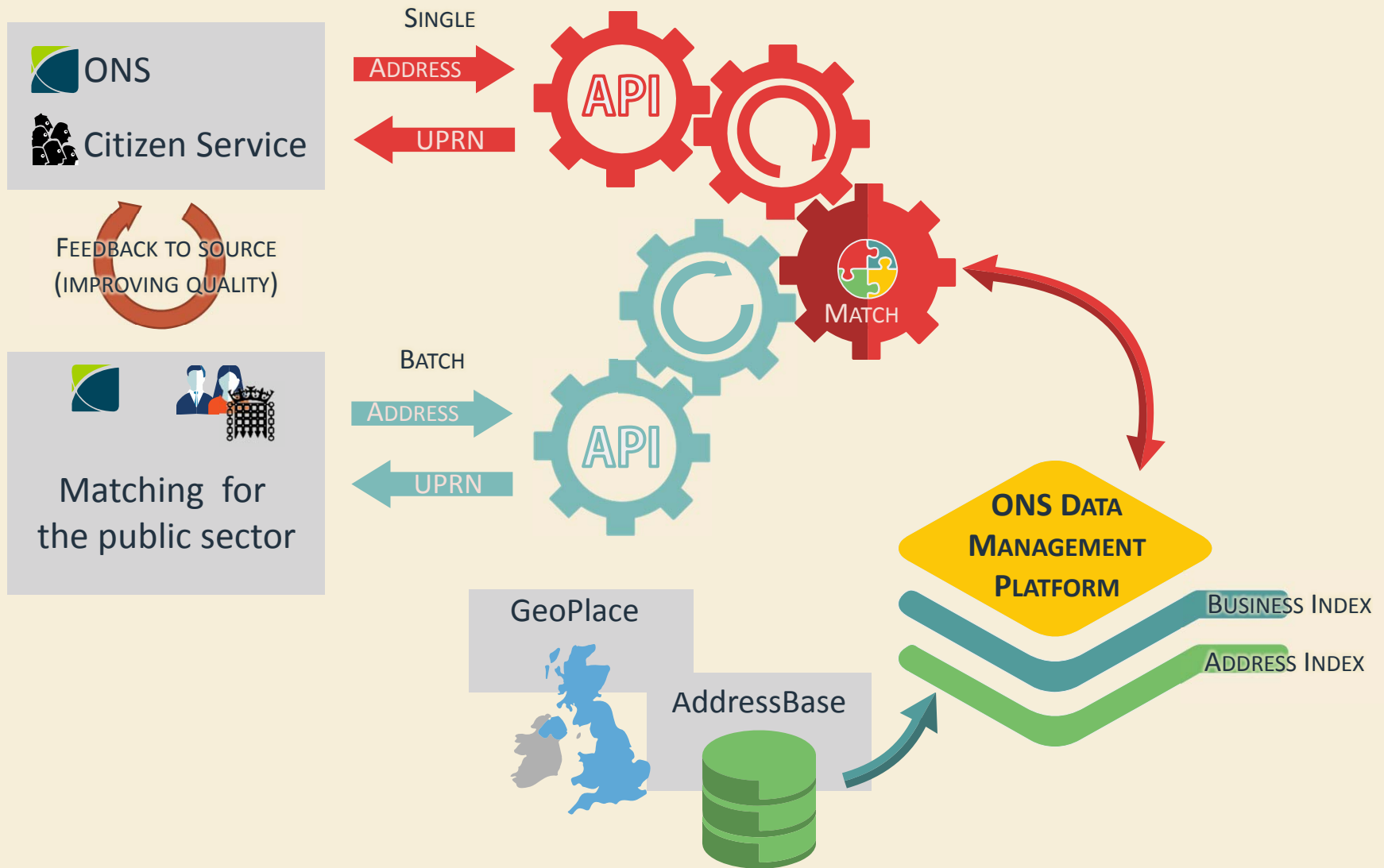


Addressing Service

Common secure services available to public sector  
(and potentially wider subject to licensing)  
common service > consistency & efficiency



# ONS Address Matching Service





# Business Index

## Input

### Current Data



Companies House



VAT



PAYE

### Future Data

- Self assessment
- Corporation Tax
- Charities and Commissions

HS

## Function

### Business Index

Matching & Linking algorithm: Combines records into linked data (legal units)

### Data Science



## Output

### Business Index Legal unit spine

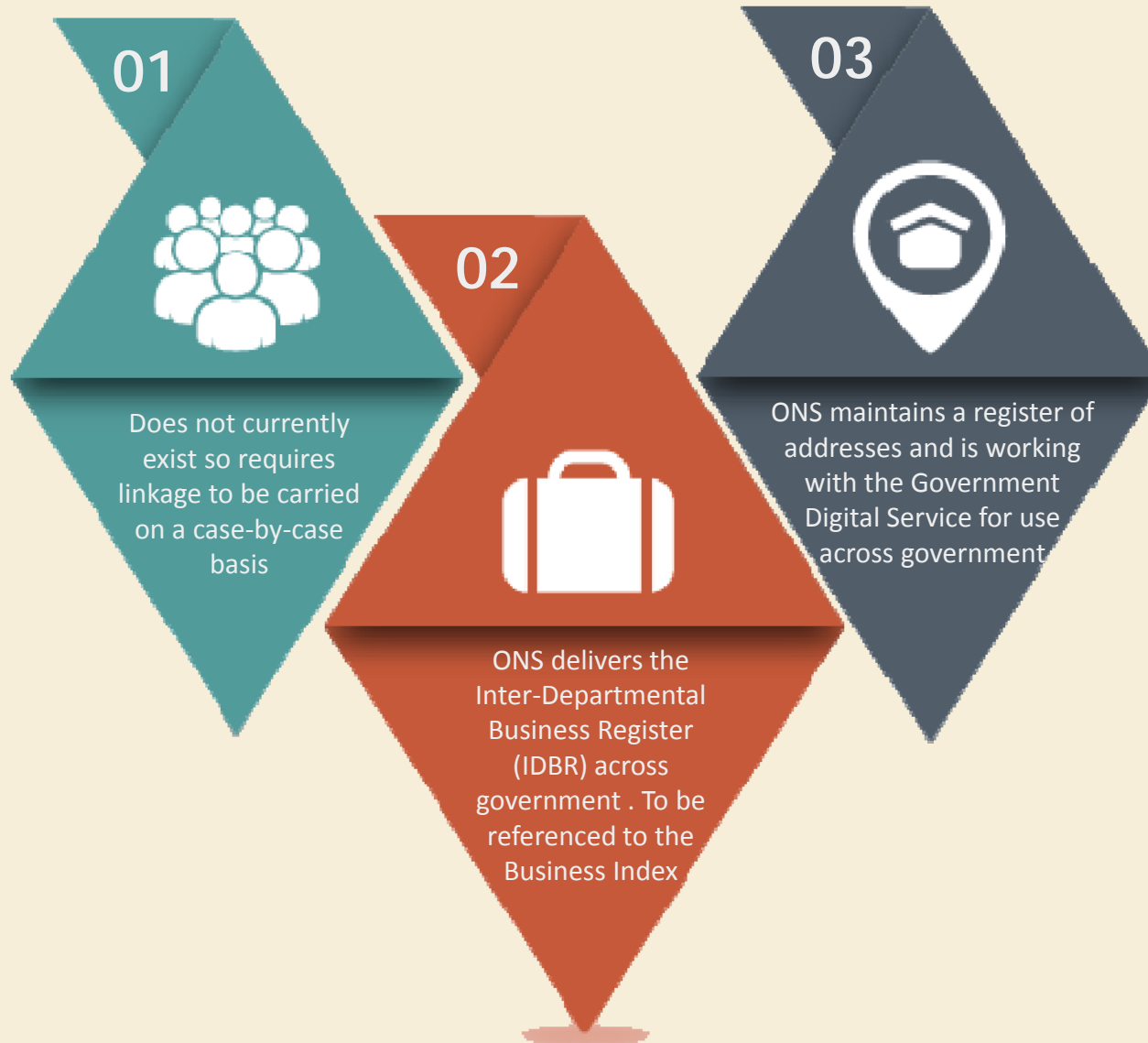


### Company details

#### Company details

- Companies House
- VAT
- PAYE

# Data Integration – 3 way index



# IDEAS – A New Data Model



# The Material Properties of a dataset

## CONTENT

What the data describes

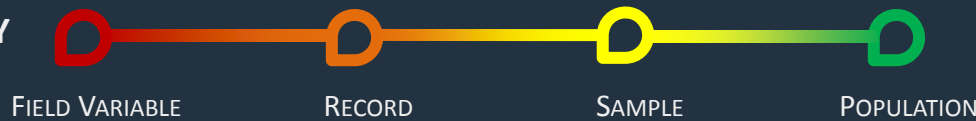
### SENSITIVITY



### IDENTIFIES



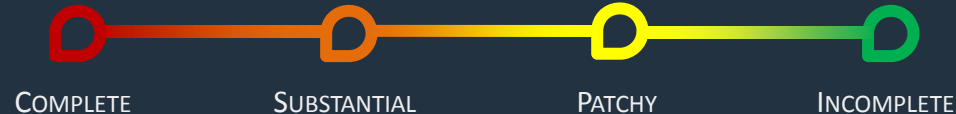
### GRANULARITY



### RECENCY



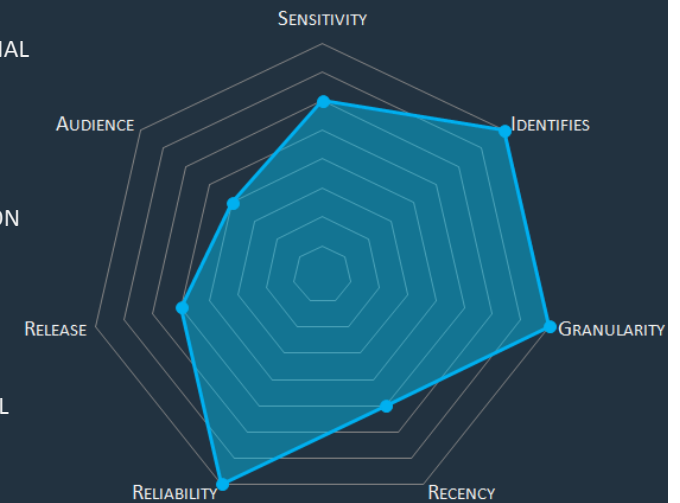
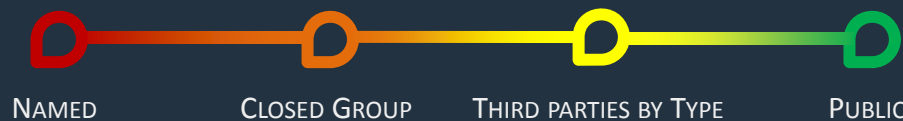
### RELIABILITY



### RELEASE



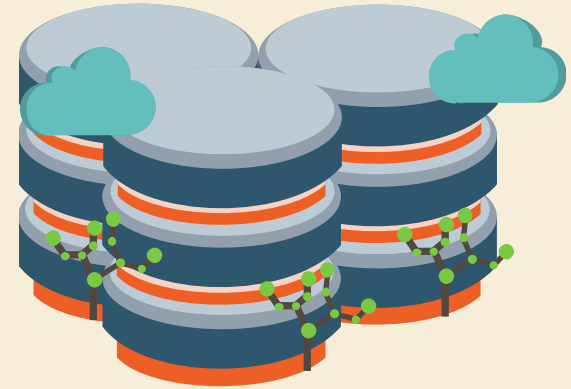
### AUDIENCE







User needs



Infrastructure

# WHY ARE WE HERE TODAY?



Collaboration



Data Access





Office for  
National Statistics

