Access without Boundaries

Richard Welpton,
UK Data Archive

Felix Ritchie,
UK Office for National Statistics

Introduction

- European Data Access
- Centralisation…or die?
- Framework for data access
- I didn't steal the title…..honest!

This work does not reflect the policy intentions of the Office for National Statistics or the UK Data Archive

*"The **research community** should enjoy **wider access** to **confidential data** used for the development, production and dissemination of European statistics, **for analysis** in the interest of **scientific progress** in Europe. **Access to confidential data by researchers** for scientific purposes **should** therefore **be improved** without compromising the high level of protection that confidential statistical data require"*

Article 26 of EC Reg 223/2009 of the European Parliament and of the Council of 11th March 2009.

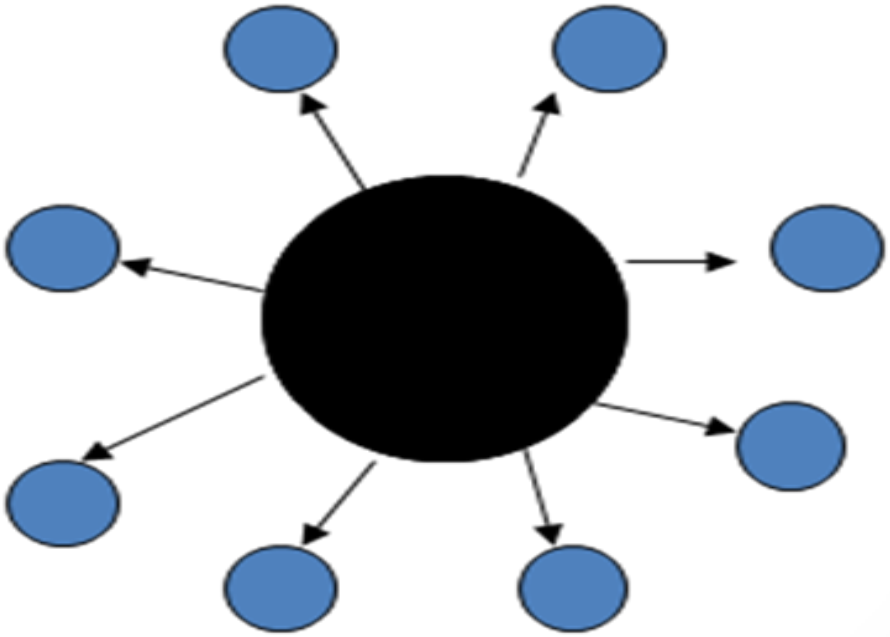
The Problem

- Analysis of data restricted to country where data are collected
- Risk Averse NSIs
- Researchers can only make comparative studies, but not statistically comparable
- Little or no harmonisation in data collection
- Small number of 'players' – NSIs, with no incentive to put Researchers first

The Solution: Centralise or die?

- One central 'node' for access across Europe
- Familiar to NSIs - safe
- Suits risk-averse NSIs
  - Highly controllable
  - Close supervision
  - Greater protection of data confidentiality

The ‘Fat’ Centre ‘System’

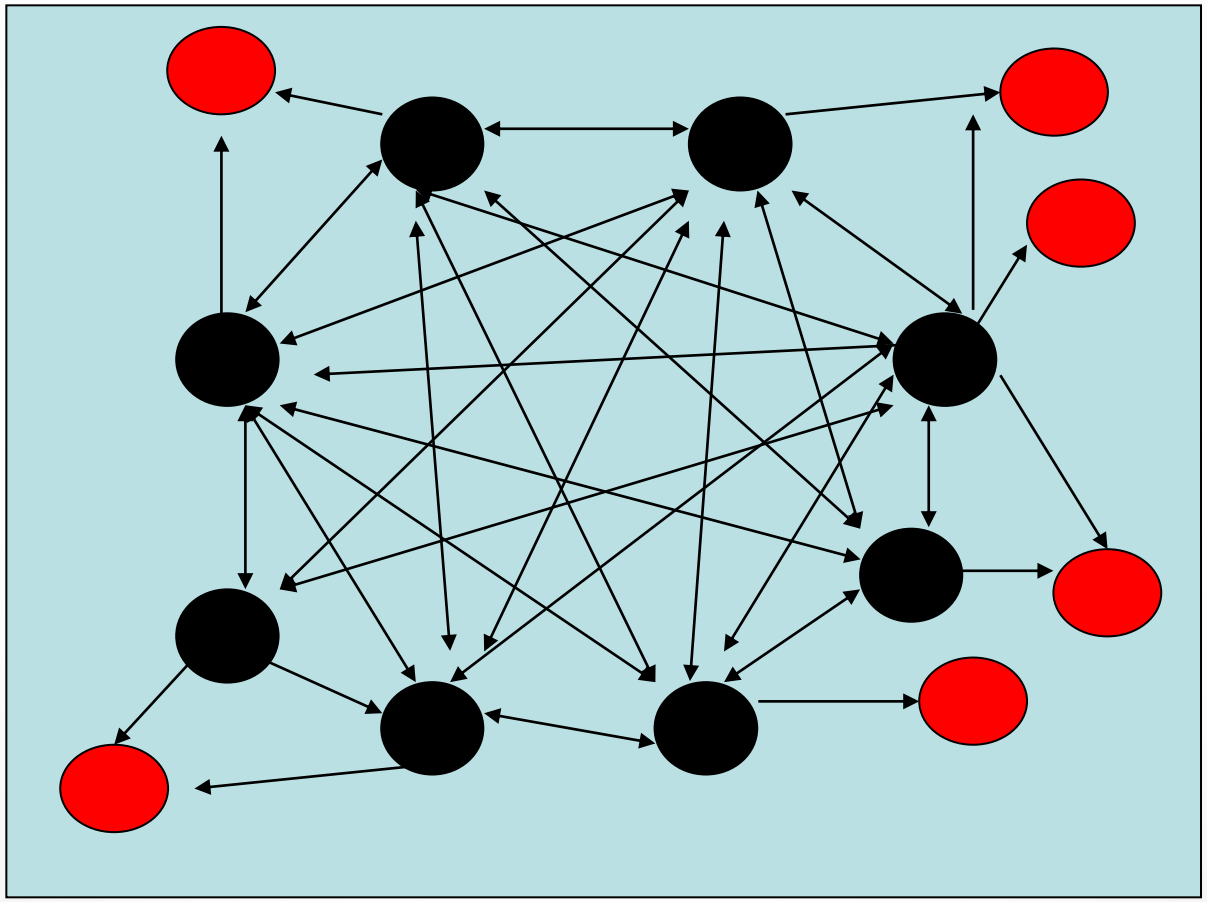


Systems vs Networks

- **System**
  - Specific solution to a specific problem
- **Network**
  - Not concerned about how things work, only how things work interactively

Systems vs Networks

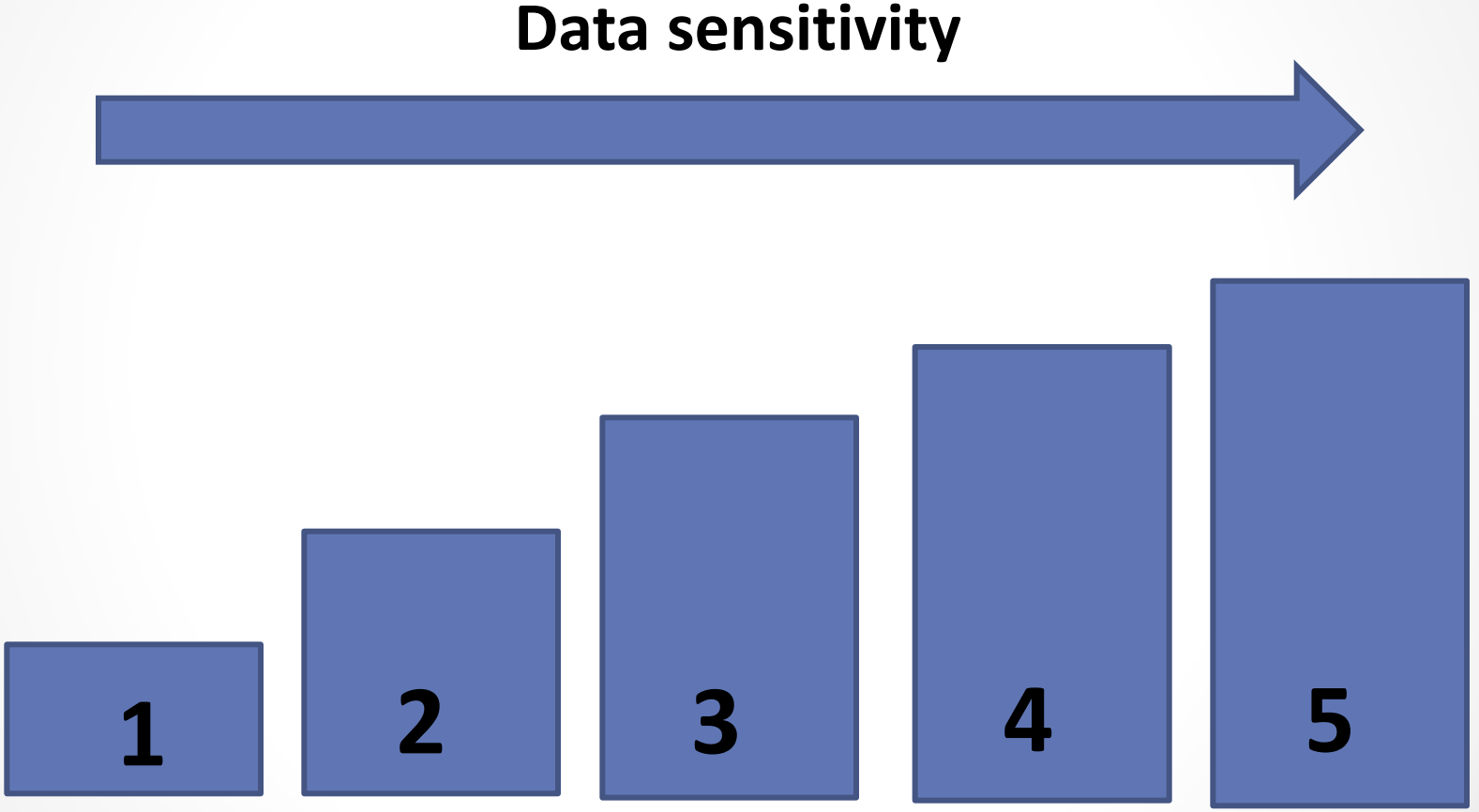
A Decentralised Network



Principles-based Network

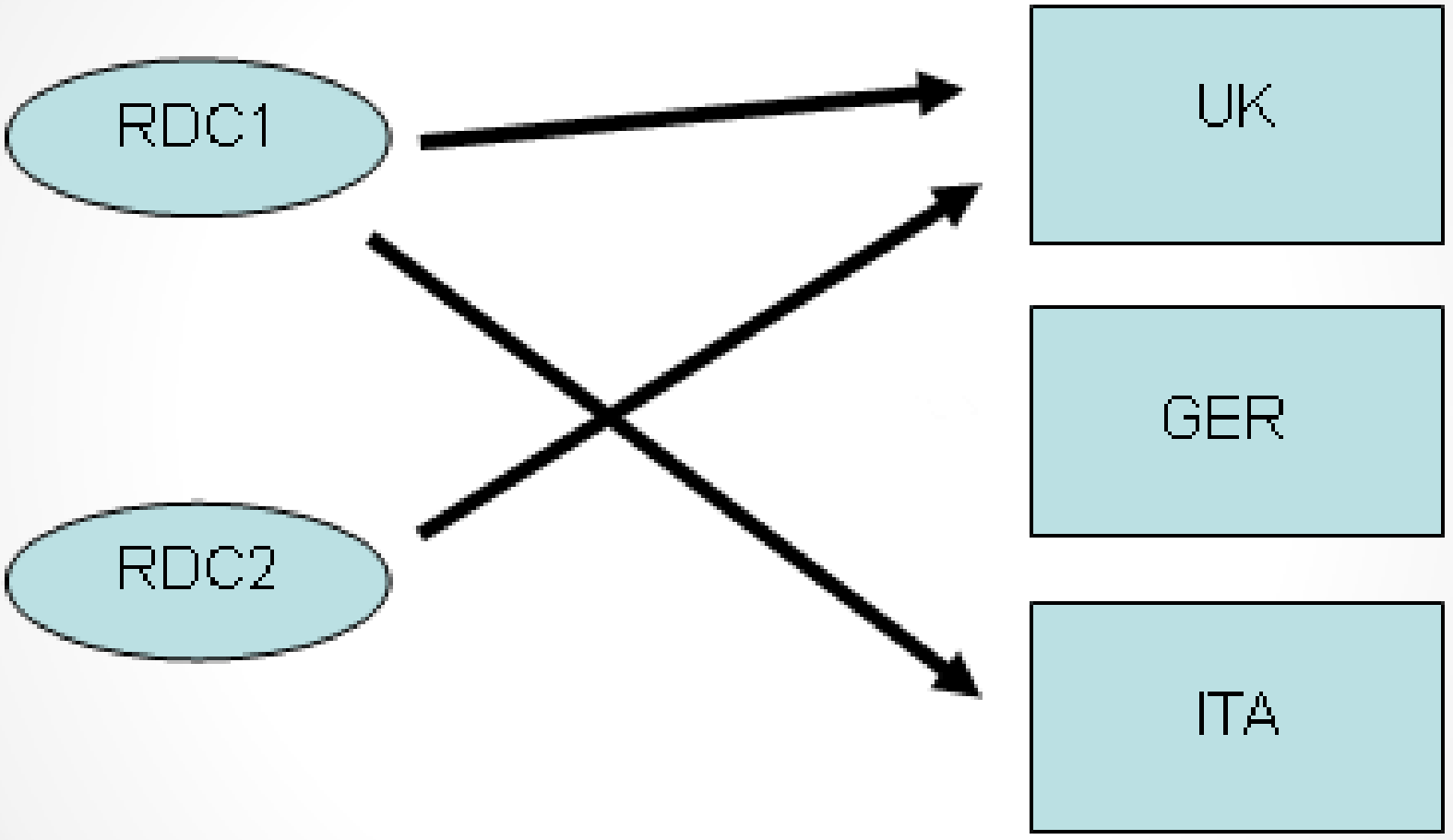
- **Common Analysis**
  - Ability to undertake comparative analysis
- **Common Access**
  - Any researcher from any country
- **Common Entry**
  - Any RDC can be established in any country to provide access to any other country's data
- **Common Standards**
  - RDCs choose their own protocols, but meet agreed standards

Common Standards (1)



Common Standards (2)

Level 1	• **Process applications** • **No data protection** • **No sanctions** • **No access restrictions** • **No output checks**
Level 5	• Full application screening • Compulsory interactive training • Criminal and civil sanctions • Secure technology that meets international standards • Full checking of outputs


RDC1
RDC2
UK
GER
ITA

Freedom or 'free-for-all'?

- Many users, many providers
- Providers choose a standard
- Markets by left by their own devices don't work
  - **Regulation by Eurostat**
  - **Regulation by NSIs**

Is there a free market for data?

A Race to the Top?

- **Better access**, more research, greater returns
- Data remains **safe** (some regulation required)
- **Innovation**
- Little evidence of a race to the bottom
- Researchers demand more sensitive/confidential data sources

Case Study: UK

- VML
- Secure Data Service
- HMRC Datalab
- Scottish Health Informatics Programme
- Welsh Assembly Government