Use of novel data sources for consumer prices
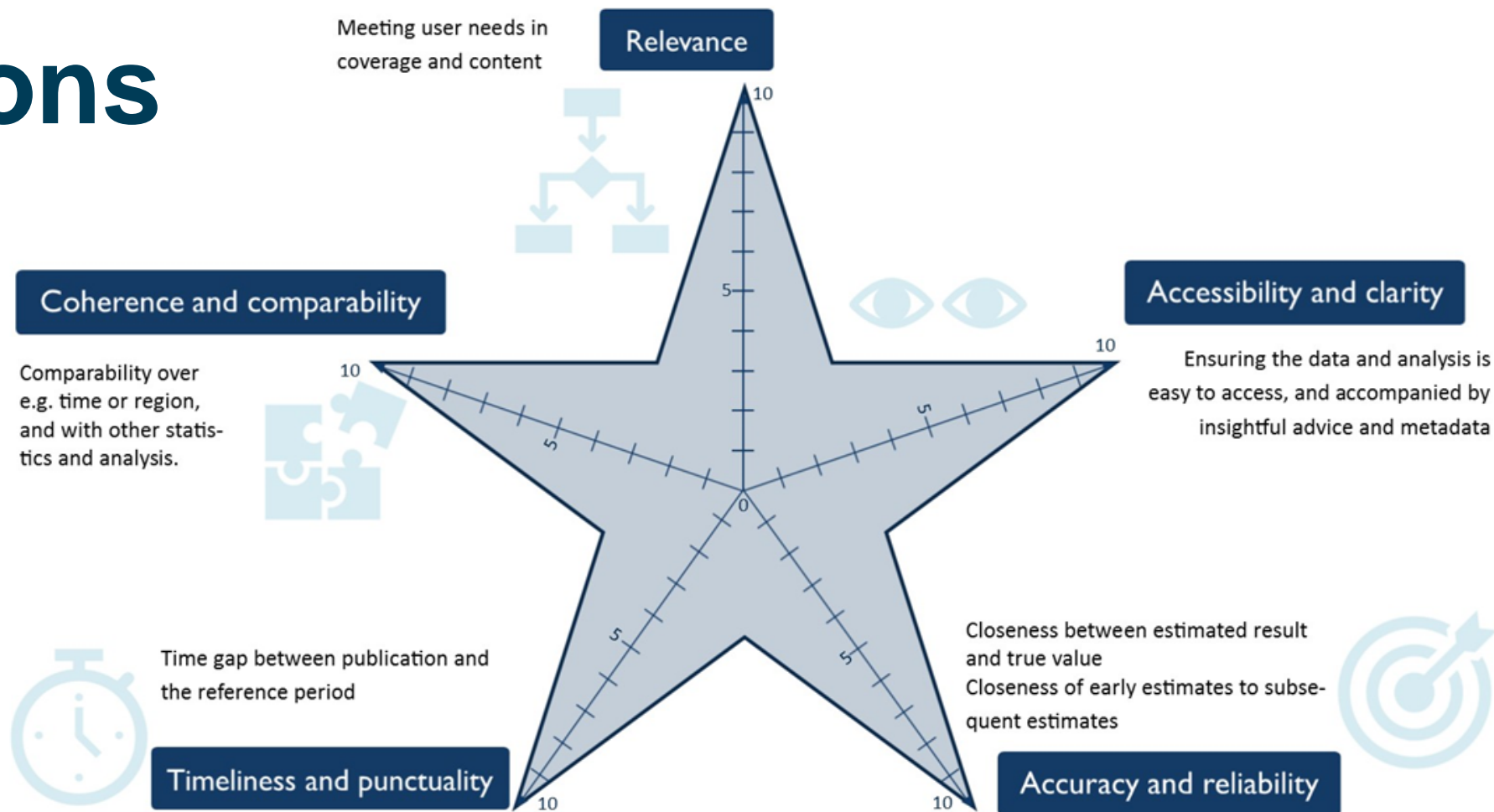
Catherine McGirr

Office for National Statistics, UK

Quality dimensions

Meeting user needs in coverage and content

Relevance

Accessibility and clarity

Ensuring the data and analysis is easy to access, and accompanied by insightful advice and metadata

Coherence and comparability

Comparability over e.g. time or region, and with other statistics and analysis.

Time gap between publication and the reference period

Timeliness and punctuality

Closeness between estimated result and true value
Closeness of early estimates to subsequent estimates

Accuracy and reliability

Office for **National Statistics**

What are alternative data sources?

- Scanner data

- Web scraped data

- Used in conjunction with data as currently collected

Why use alternative data sources?

- Improved coverage – for example regional

- High frequency of collection

- Reduced burden

- More granular information (eg expenditure)

Comparison of different data sources

Data dimension	Traditional	Scanner data	Web scraping
Data acquisition	Manual	Automated	Automated
Completeness/ scope	Sample items from sample of retailers	All transactions (census) from medium to large retailers	Bulk or sample from online retailers
Metadata	Item description + some attributes	Item description + (some) attributes Backseries may be available	Item description + attributes
Quantity data	None	Quantities sold	None
Timing	Single collection day	Daily	Daily/weekly

Ensuring quality of the data

- Initial data checks

- Standardising the data

- Monitoring the regular feeds

- Data sharing agreements

- Quality Assurance of Admin Data (QAAD)

Ensuring quality of the outputs

- Robust statistical systems

- Research into new methods

- Impact analysis

- Incorporate QA into regular monthly production

- Advisory panels – stakeholder and technical

- International best practice