

Session 2: Enabling Institutional Environment for data integration

Key findings of the Asia-Pacific Data Integration Capacity Assessment Survey

24 November 2020

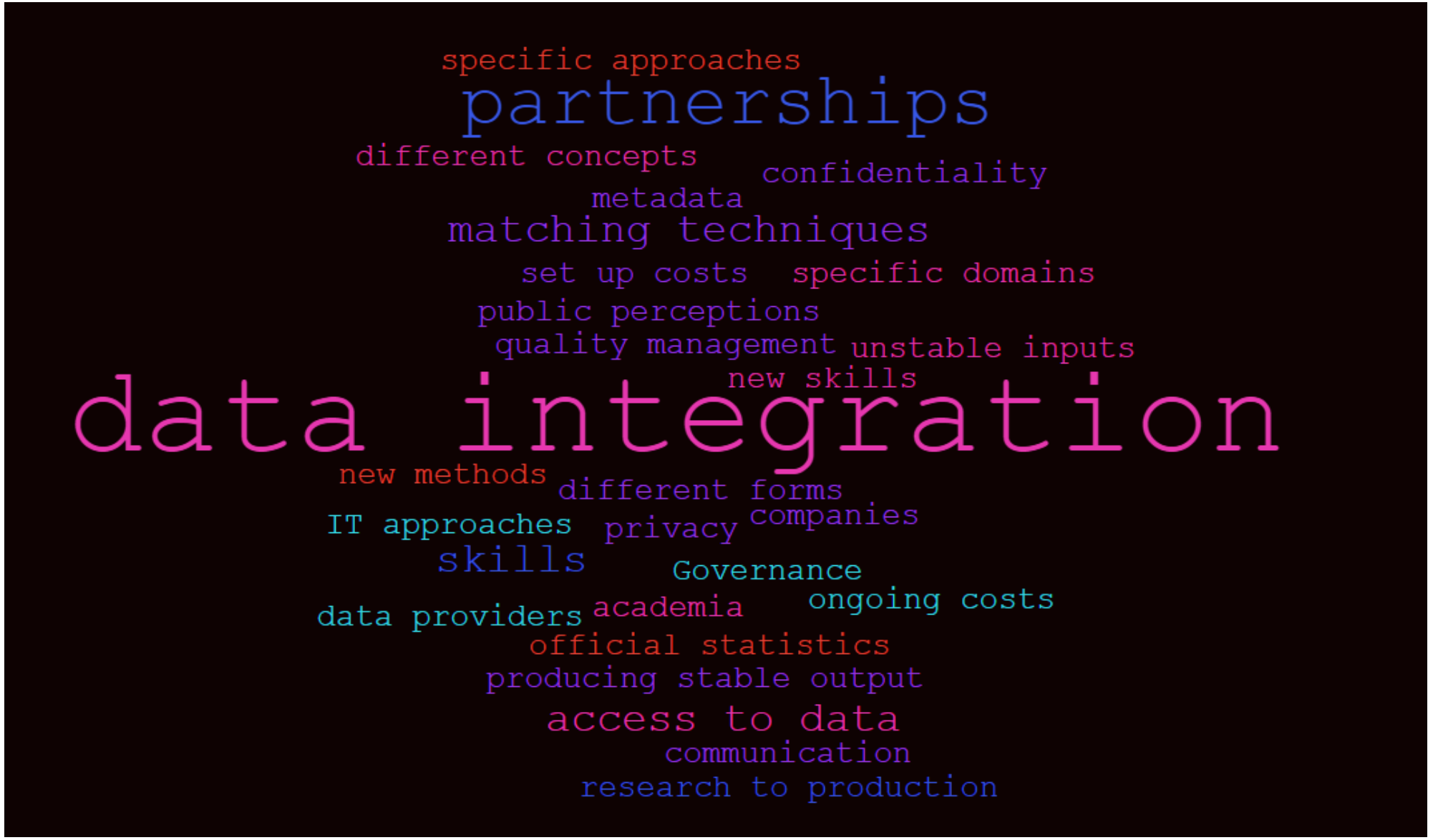
Ms. Jenine Borowik

Data Integration



Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information.

Challenges:


specific approaches
partnerships
different concepts
confidentiality
metadata
matching techniques
set up costs
specific domains
public perceptions
quality management
unstable inputs
new skills
data integration
new methods
different forms
companies
IT approaches
privacy
skills
Governance
ongoing costs
data providers
academia
official statistics
producing stable output
access to data
communication
research to production

Opportunities:



Key messages of Data Integration Capacity Assessment Survey

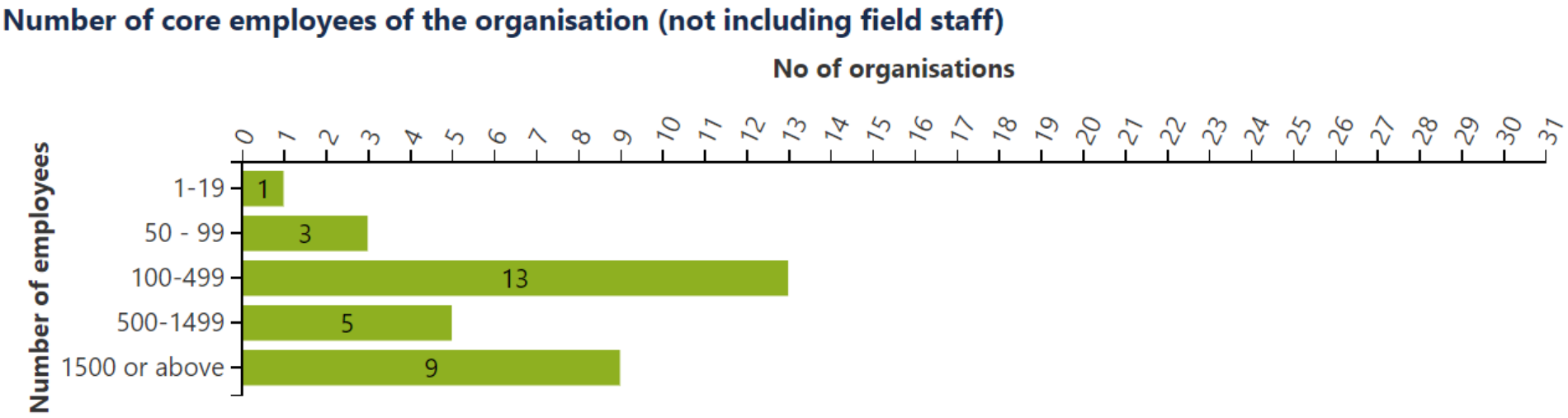


Aims of the survey

- Assess capacity for data integration in the region
- Adapt the "Guide to Data Integration for Official Statistics" to best meet the needs of countries in the region
- Plan and provide for future capacity building activities.

1- Responses (as at 23/11/2020)

- 31 responses - 28 NSOs, 3 others (2 Central banks, 1 Civil Registration)
- 10 centralised national statistical systems,
- 15 case studies submitted
- Good representation of organisation sizes



2- Skills

Of 31 organisations

- 26 using data integration to produce statistics
- 14 have access to required skills + 16 partly have access to required skills
- 22 have developed at least some training resources
- 28 are interested in further capacity building

Key message

A lot of expertise in region and interest in developing more

- Methods for linking (24)
- Data consideration and quality (24)
- Development of indicators for SDGs (24)
- ICT and tools (25)

3- Laws and Privacy

- 29 countries have laws which provide full or partial access to government data
- 23 have laws which provide full or partial access to private data
- Privacy issues are a barrier in many instances (21)

Key message

Access to data from private sources more difficult (not surprisingly)

4- Strategies for Data Integration

- 18 organisations have an organisation wide strategy for DI or are developing one
- 8 organisations are considering development of a strategy
- Approximately half of the organisations have specific data integration units which support or undertake DI activities

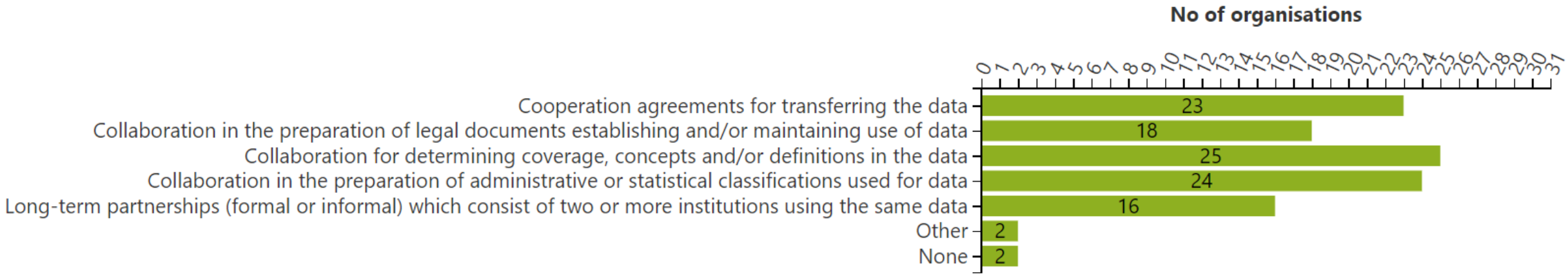
Key message

Opportunity to share strategies and design of DI units.

(Some countries already have)

5- Pre-Integration Practices

What pre-integration practices does your organisation use with data providers or other partners



Key messages

- Collaboration and agreements with other organisations are common
- Opportunity to share

6- Identifiers

- Unified personal identification (21) and unified business identification systems (18) are common
- Unified address identification systems are less common (13)
- Some countries have unified identity systems which do not completely cover the population (4 personal, 5 business, 5 addresses)
- There are restrictions on data custodians providing identifiers to the NSOs in 15 countries

Key message

use of unified identifiers is not available in all circumstances -> relevant to requirements for record linkage/statistical matching

7- Barriers

Most common barriers

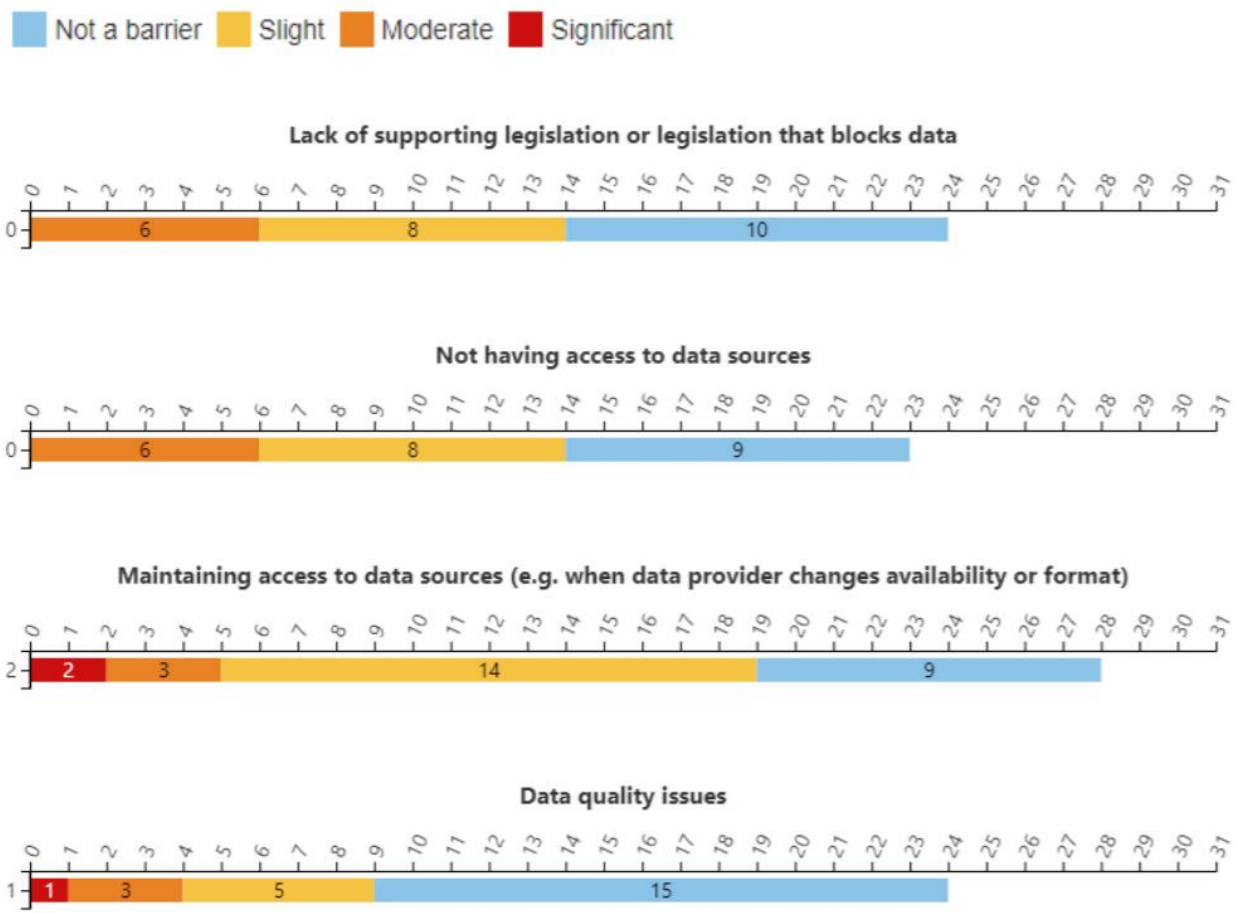
- Legislation
- Not having access to data
- Maintaining access to data
- Public acceptance and trust
- ICT issues

Key message

Benefit of sharing

Experiences to manage

Example of results (11 different barriers assessed)



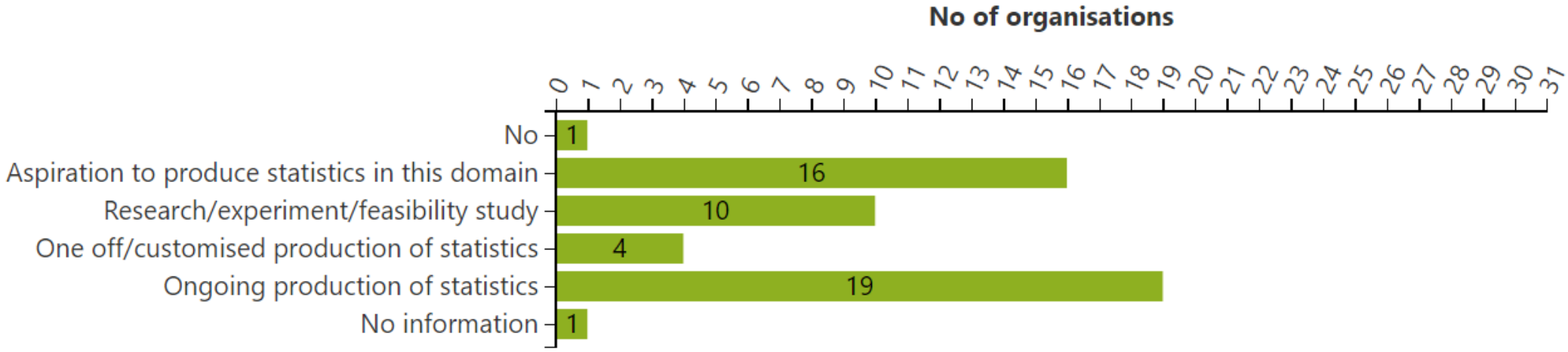
8- Statistical Domains

- Social, Economic, Environment, SDGs, others
- Aspiration to use more data integration in all statistical domains
- Aspiration to use DI for SDG indicators particularly strong
- Many research projects/feasibility studies underway

Key message - Benefit in sharing approaches for measuring SDGs

Does your organisation conduct, or plan to conduct, any data integration activity

To produce or improve SDG indicators



9- Tools

Wide variety of tools being used

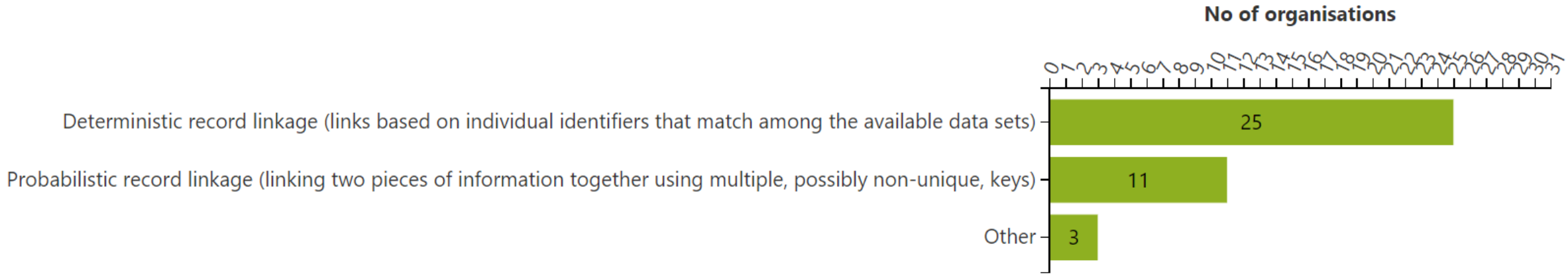
- R (15)
- SAS (10)
- SPSS (15)
- SDMX/API (10)
- SQL (and PHP) (21)
- Spreadsheets/EXCEL (23)
- MS Access
- Python
- STATA, Cspro
- GIS

Key messages –

Different tools for different types of integration? Further analysis required?

10- Methods

What methods does your organisation use?

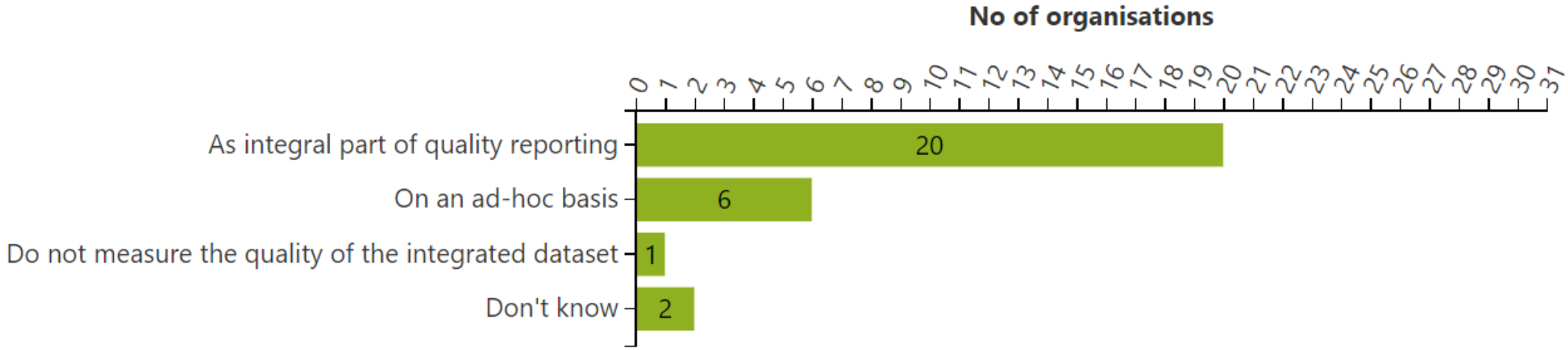


Key message/question?

Is the lower use of probabilistic record linkage restricting data integration activity?

11- Quality

How does your organisation measure the quality of statistical information produced as a result of data integration activities?



Key message

Opportunity to share quality frameworks and other approaches

Thank you



To view more detailed results

- Results available on DI-COP space (id required)
- If graph is ‘cramped’ or ‘not visible’ collapse the sidebar by clicking the symbol. (expand the sidebar by clicking again)

