UNSD/UNECE Survey and Project Templates

Ronald Jansen
United Nations Statistics Division
jansen1@un.org and BigData@un.org

International Conference on Big Data for Official Statistics
organized by UNSD and NBS China
28-30 October 2014, Beijing China

UNSD/UNECE Survey on Big Data - Objective

- To provide an overview of active Big Data projects for a more informed discussion and to further develop the programmes of the international working groups.
- We use the following definition of Big Data:
  - **Big data are data sources with a high volume, velocity and variety of data, which require new tools and methods to capture, curate, manage, and process them in an efficient way**

UNSD/UNECE Survey on Big Data - Overview

- Two different questionnaires:
  - Organization, Strategy and Governance
  - Big Data projects
- Conducted in September/October 2014
- Sent to the members of the UNECE Big Data group and of the UN GWG
- Received
  - 33 responses on Organization and Strategy
  - 57 responses on Projects (from 25 countries)

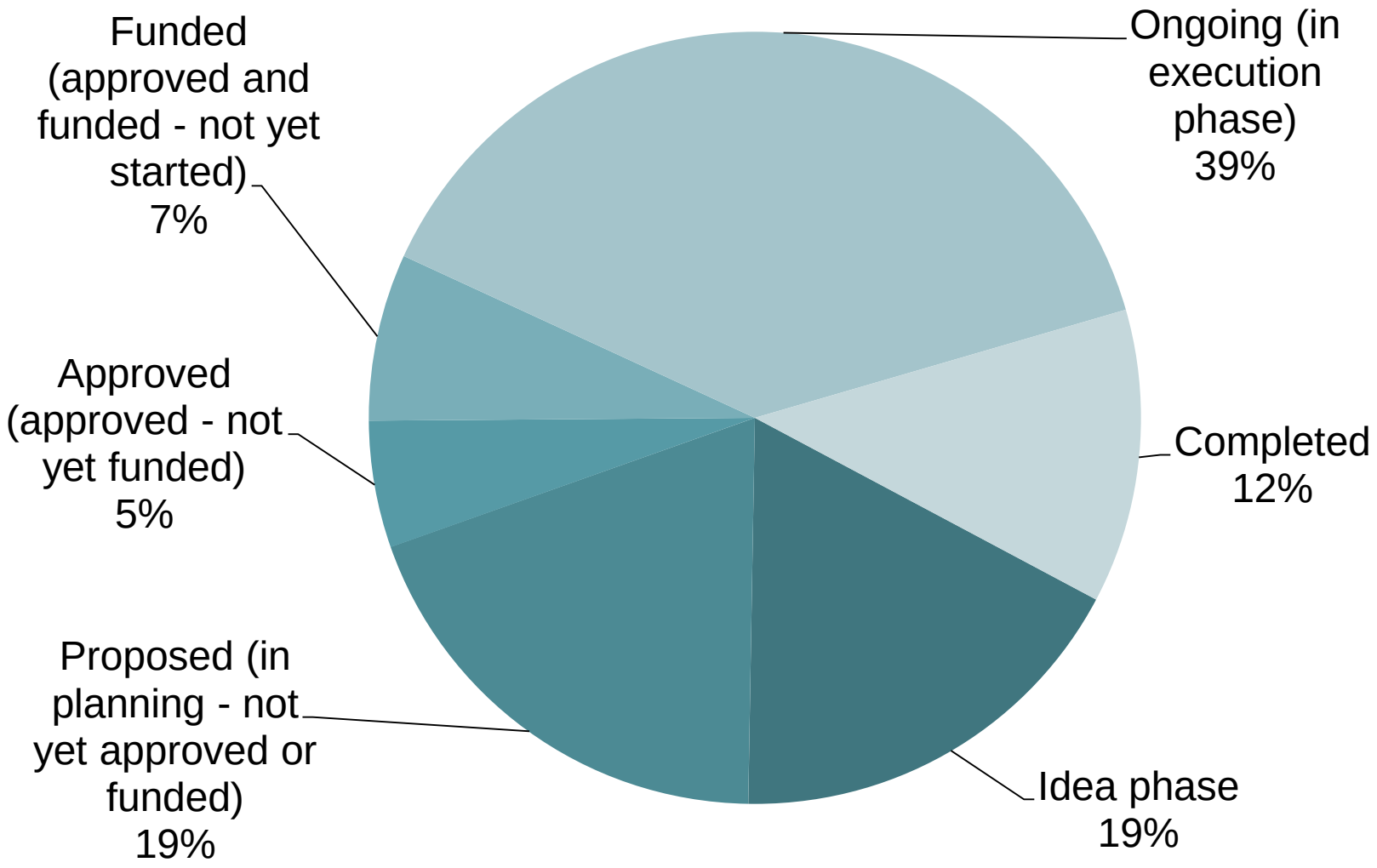
Summary of questionnaire on Big Data Organization and Strategy

- **Big Data strategy:** Only a few countries have developed a long-term vision for the usage of Big Data, some are formulating a Big Data strategy.
- **Quality -** The work of international organization is deemed important, in particular many countries mention the UNECE quality task team that aims to establish a quality framework
- **Privacy -** Crucial to protect privacy, even beyond what is defined by the rule of law, in order to protect public image
- **Skills -** So far NSOs have relied more on providing training to existing staff rather than hiring of a new type of staff ("data scientists").

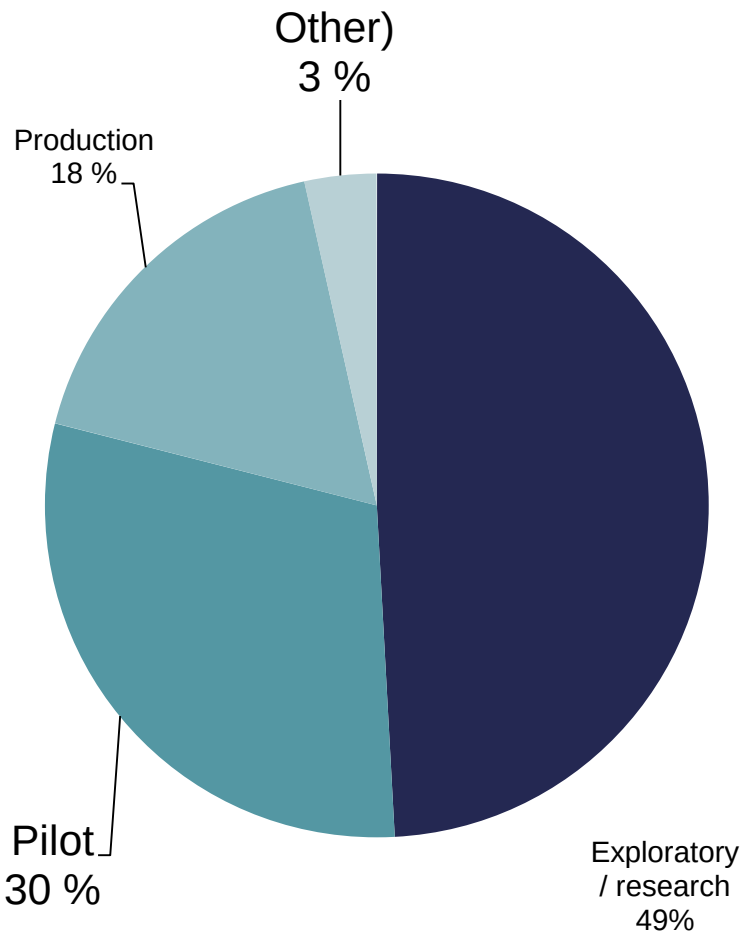
Project questionnaire – Main findings

- **57 projects** were submitted using the project questionnaire from 25 different countries/organizations.
- The countries/organization that submitted the largest number of projects where Eurostat, the Netherlands, Canada, Australia and Hungary and United Kingdom.

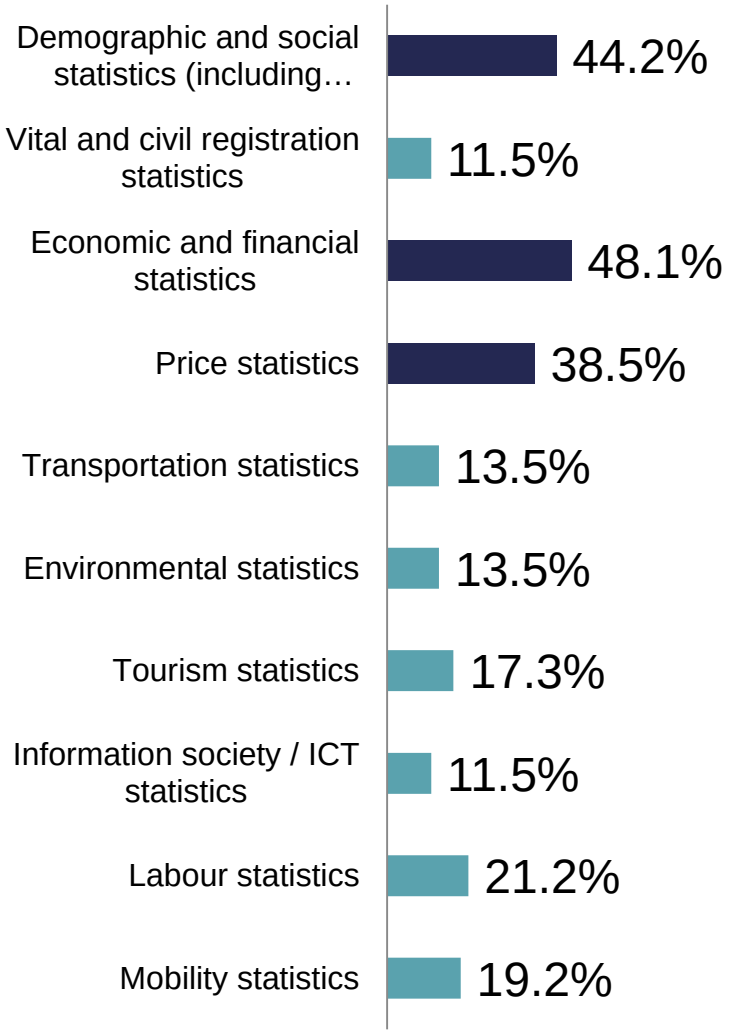
Project status



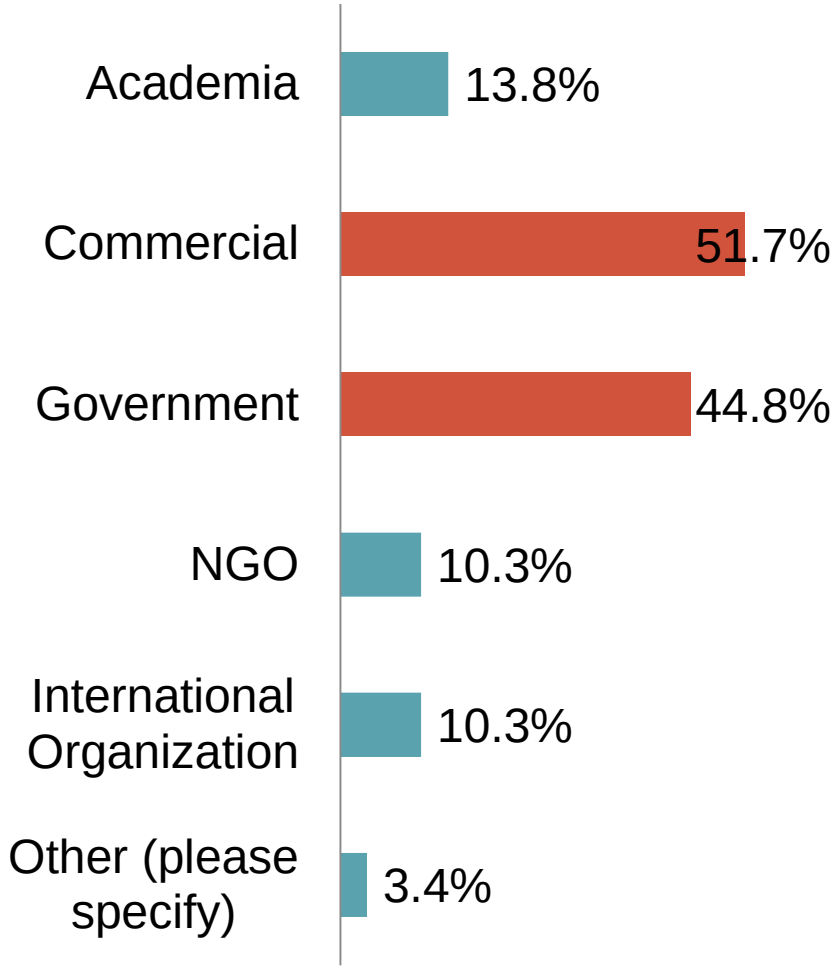
Would you qualify the project as:



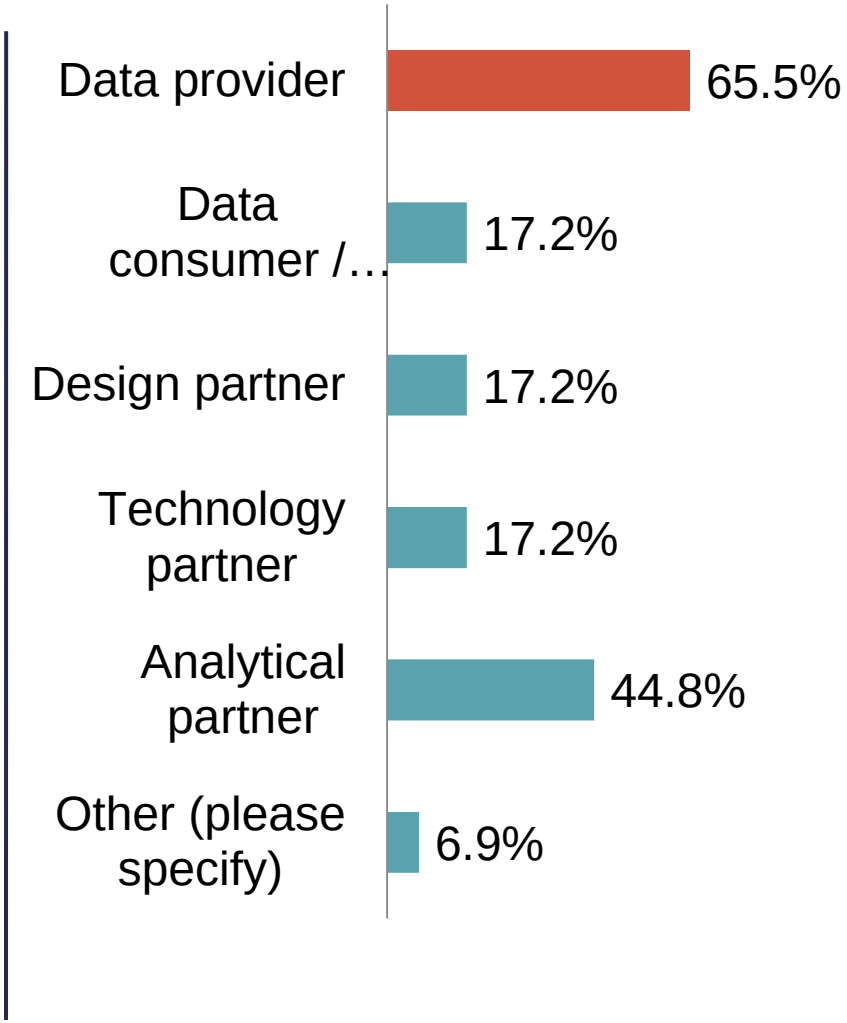
Potential areas of use for this project



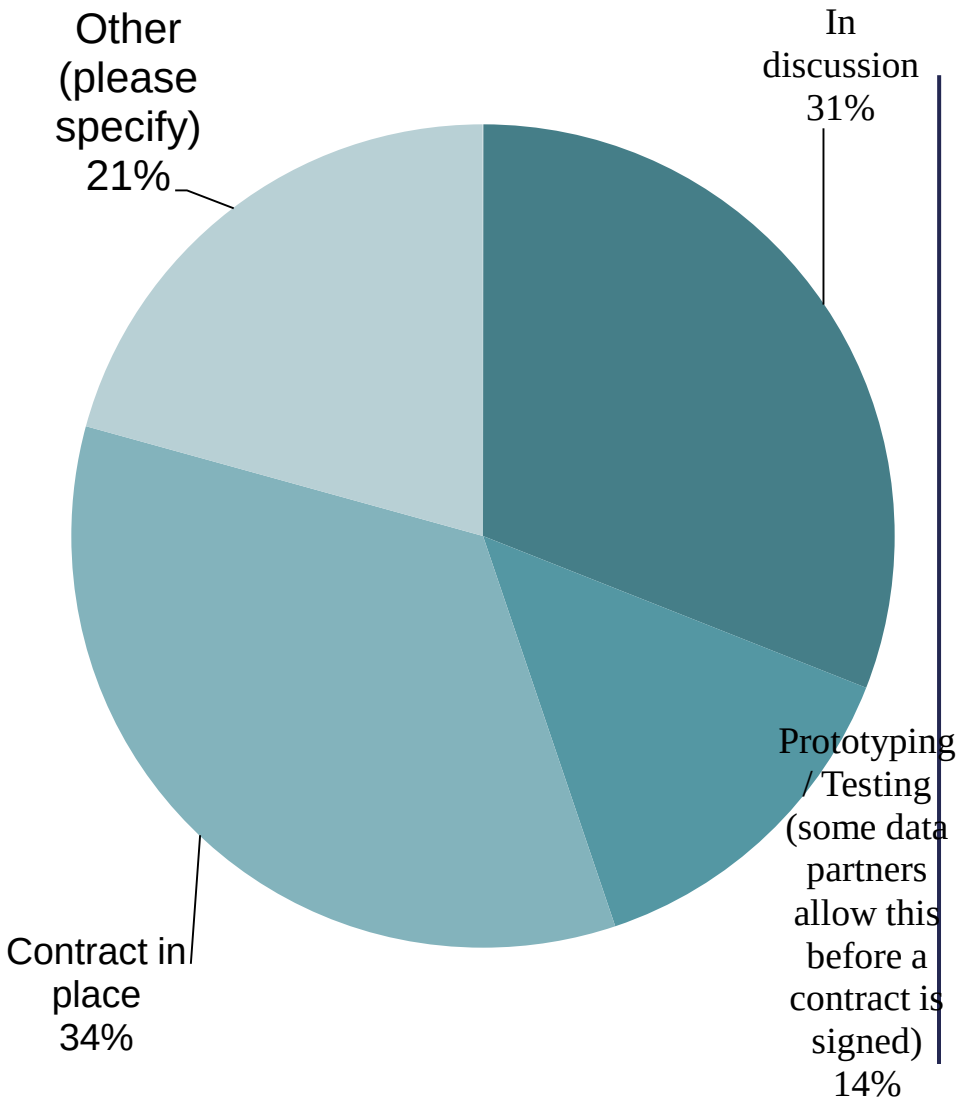
Type of partner organization



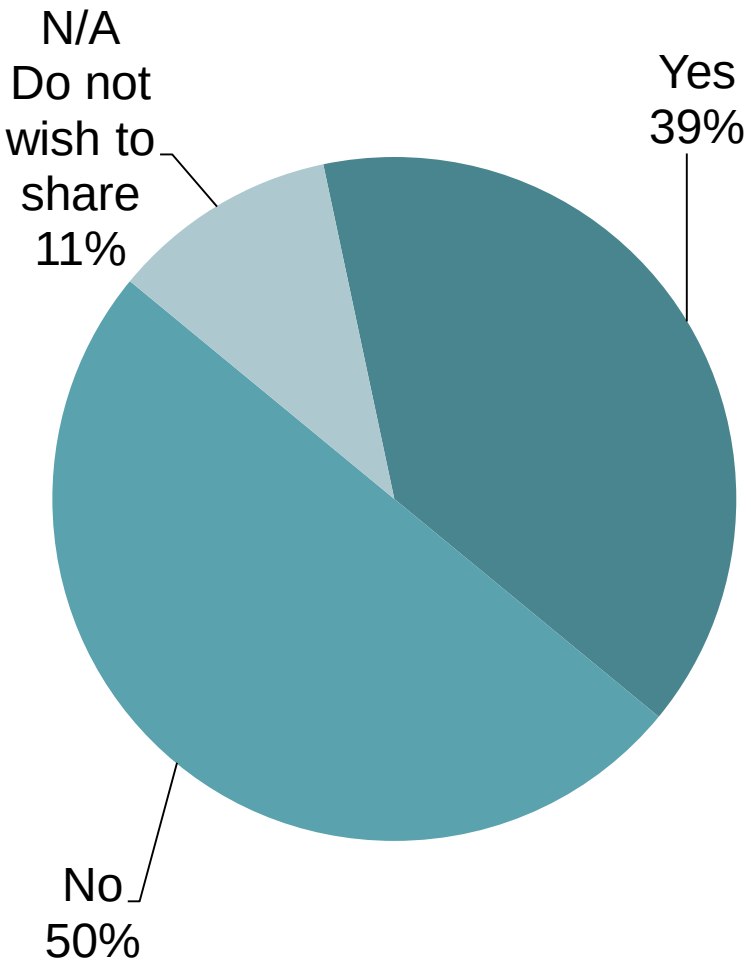
Type of partnership



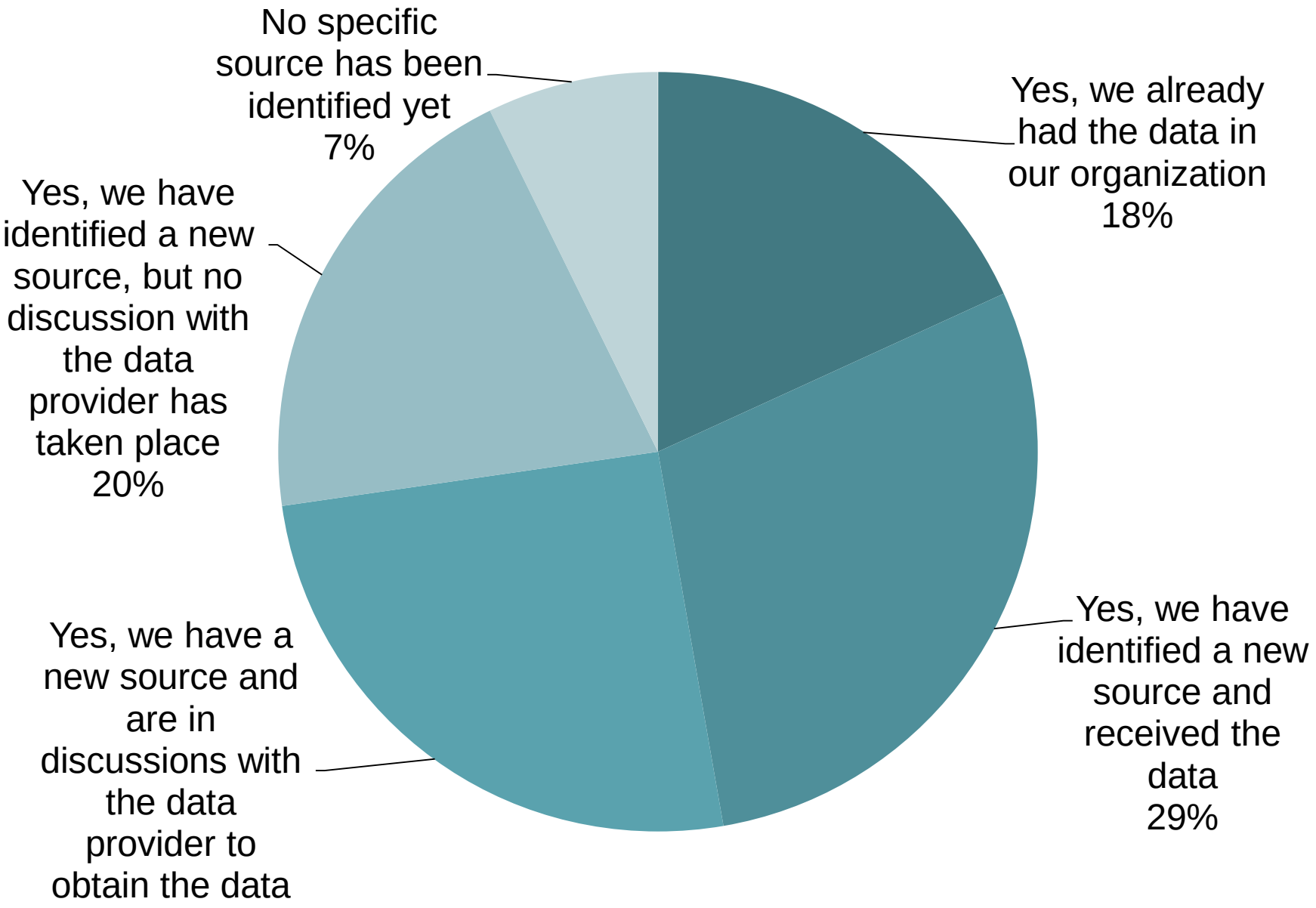
Current status of the partnership



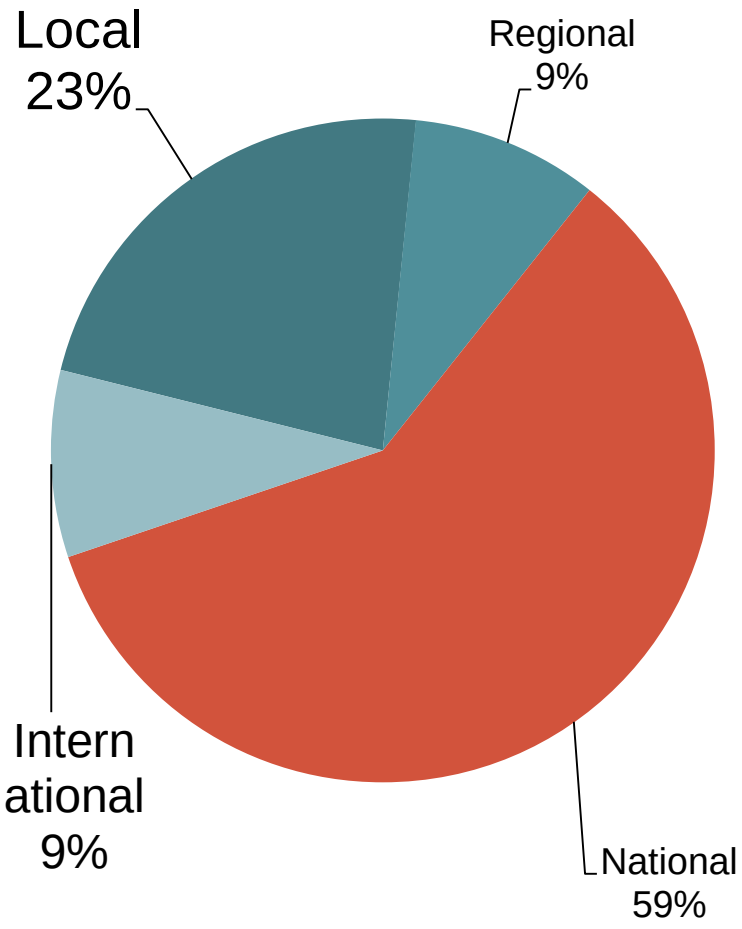
Are there any payments or financial arrangements with this partner?



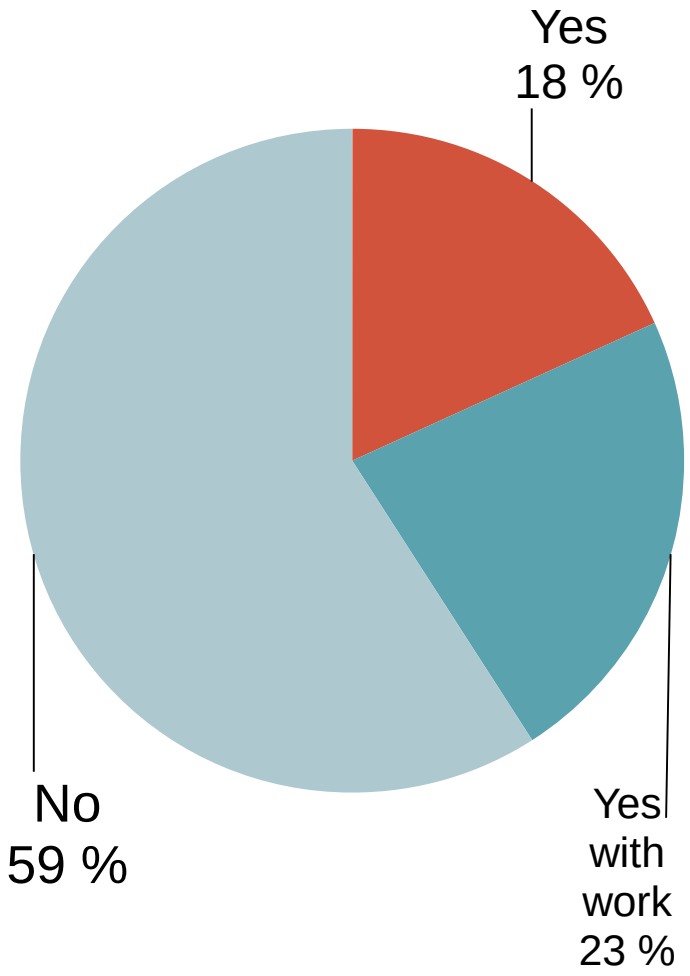
Do you have any data sources for this project?



What is the geographical scope of the data source?



Is this data source publicly available?



Technical tools

- Overwhelmingly survey participants chose to use internal hosting solutions rather than purchasing external hosting services.
  - Often due to privacy concerns
- Having to invest significant resources in a cluster before beginning a project represents a significant hurdle.
- Data analysis was mainly conducted using R and SAS.
- However, a large variety of tools were used in the projects, making it clear that no significant standard has emerged

Summary of Survey

- The various aspects can be tackled at different levels, i.e. national, regional and global
- Most respondents agree that the biggest challenge for most Big Data projects are the limited access to potential datasets.
  - Big data are, to a large extent, owned by the private sector and many of these players are global companies;
  - Thus, the global statistical community should use collective bargaining power to obtain the access to these data sets.
- Also, some respondents suggest that the international community should facilitate global partnerships and coordinate the development of quality and confidentiality frameworks

Thank You

Ronald Jansen
United Nations Statistics Division
jansen1@un.org and BigData@un.org

International Conference on Big Data for Official Statistics
organized by UNSD and NBS China
28-30 October 2014, Beijing China