

Statistical quality considerations that can be useful for the production of the innovation survey

Presentation to workshop on the review of innovation measurement in South Africa

9 November 2011

 Statistics
South Africa

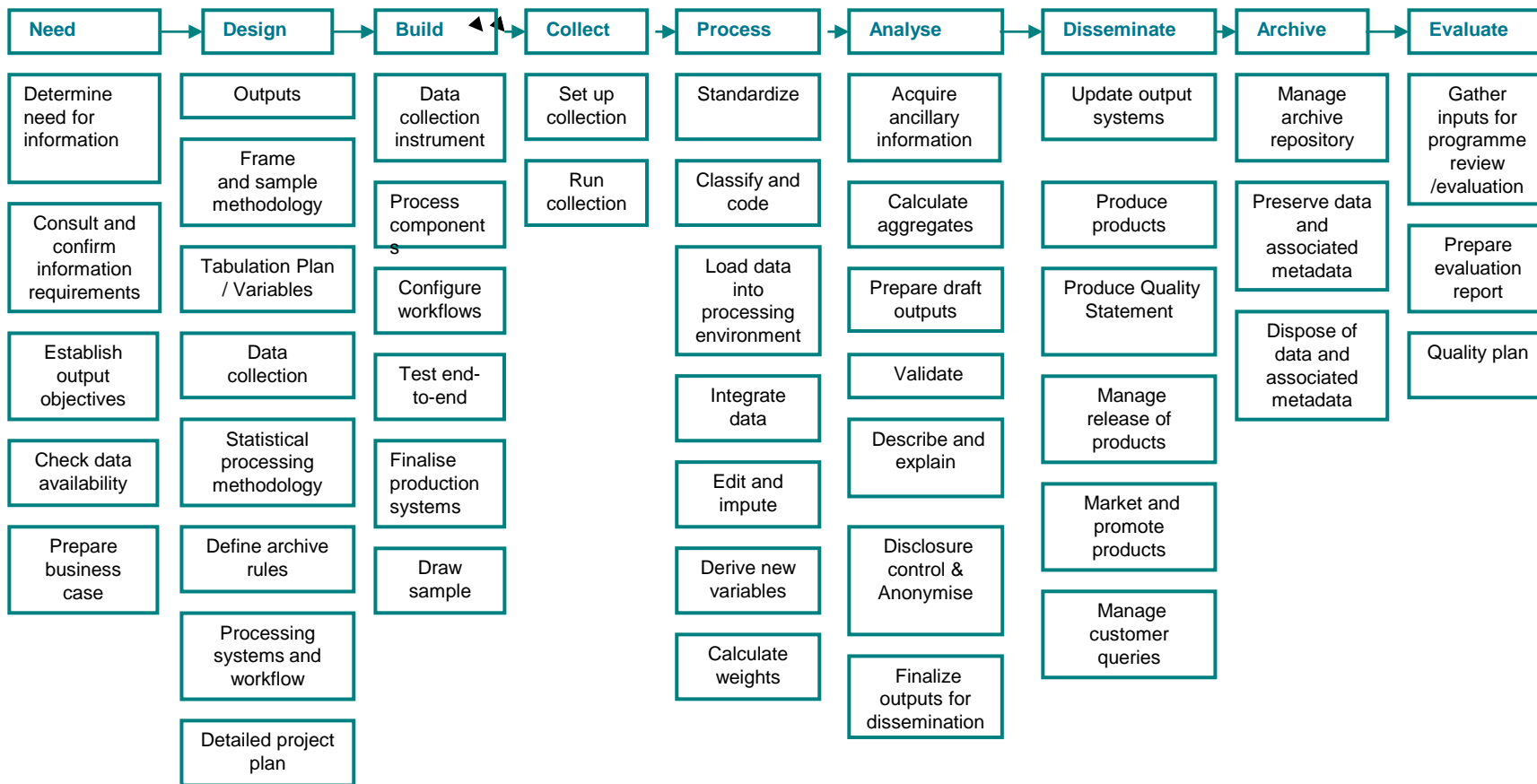


•••• your leading partner in quality statistics

Outline of the presentation

- The Statistical Value Chain
- Pre-requisite of Quality
- Accuracy
- Timeliness
- Comparability and Coherence

The Statistical Value Chain



Quality dimension: Pre-requisite of quality

1. A law/legal arrangement exists that explicitly mandates the production of statistics

Findings

- MoU signed in 2002 between the DST and CeSTII showing mandate for data production

Recommendations

- Ownership of the survey needs to be clarified
- MoU needs to be renewed

2. A set of policies must exist which covers all aspects of the statistical value chain (SVC)

Findings

- Oslo manual provides high level guidelines.
- Some aspects of the SVC are not covered by the manual

Recommendations

- Oslo manual is not prescriptive
- The provided guidelines and their associated processes can assist the DST and Cestii to develop policies covering all aspects of the SVC

3. A set of standards related to appropriate policies must exist.

Findings

- Oslo manual is a source document for conducting the survey
- Although it provides some standards others are lacking

Recommendations

- Specific standards associated to each policies need to be developed

4. There must be a law or other formal measures that inform respondents of their obligation to provide information, and any sanctions, which may apply if they fail to do so.

Findings

- No law or formal measures exists other than the covering letter by the Minister of Science and Technology that requests respondents to complete the survey instrument

Recommendations

- As this survey is not official statistics the Stats Act cannot be invoked
- However, the status of the R&D Survey can be used to solicit response
- Also, follow-up procedures for non-response should be systematic and documented

5. There must be a statistics unit or component responsible for compiling statistics

Findings

- A senior statistician was recruited recently.
- Lack of clerical/administrative staff for carrying out survey logistics

Recommendations

- The unit probably needs one or two survey statisticians as support staff
- Senior staff spend too much time on clerical work

6. A disaster recovery and business continuity plan must exist.

Findings

- No evidence of a business continuity plan

Recommendations

- DST needs to request a business continuity plan as part of the contract
- If no continuity plan exists, Cestii has to implement its own business continuity plan to safeguard the statistics produced (including the official statistics)

7. Computer software resources must be adequate in terms of:

- capturing systems;
- editing systems;
- coding systems;
- statistical software;
- up-to-date licences;
- virus protection; and
- appropriate access rights.

Findings

- There is no evidence that this standard has been addressed in the metadata provided

Recommendations

- A regular audit of the ICT hardware and software needs to be undertaken that pays specific attention to the items identified in the standard
- Ideally this can be conducted annually and form part of the IT audit conducted by the HSRC

8. Budgets must be adequate and timely.

Findings

- Cestii operates a combined budget for both surveys
- No evidence exists of how much of the budget goes directly to each survey
- HSRC funds partly some of the surveys

Recommendations

- Keeping separate budget for the surveys can ease the process of costing the surveys
- Assist in determining adequacy of the budget

9. Staff of a statistical programme must be employed in positions that are aligned with their skills profile.

Findings

- No evidence of this was found

Recommendations

- Cestii needs to keep records of job descriptions, qualifications and experience required along with a profile (i.e. skills, qualifications and experience) of incumbent

10. The agency must have a quality management system (QMS) in place.

Findings

- Evidence existence of some of quality management procedures
- The deficiency in the QMS is that there are no quality standards and policies that cover the SVC

Recommendations

- Develop a policy on quality
- Use the relevant SASQAF standards as part of the QMS
- Standards for the QMS will reflect compliance with processes in the QMS and if they are meeting the required standard

11. Staff members in production areas must have a data quality management requirement as part of their performance agreements or job descriptions.

Findings

- No evidence of this was found

Recommendations

- As a minimum requirement, managers of the survey must have the responsibility for data quality as part of their performance agreements

Quality dimension : Accuracy

1. Scientific sampling techniques must be used

Findings

- A probability sample used
- Frame: Stats SA business register
- Population subdivided into strata
 - 2 digit SIC (6 sectors)
 - by turnover size rather than by employment size
- Oslo manual prescribes employment size as a measure of size
- Sample size: 4000 enterprises later reduced to 2 836 (29% invalid)
- Realised sample size: 757 (27%)

Recommendations

- Need better sampling strategy
- This study cannot be used to make inference about the target population because it has high level of bias

2. Measures of sampling errors for key variables are calculated. Amongst others these are:

- standard error
- coefficient of variation (CV)
- confidence interval (CI)
- mean square error (MSE)
- design effect (DEFF)

Recommendations

- Measures of sampling errors must be calculated for the main variables
- They must be available for the other variables on request
- Work more closely with Stats SA since the Department has more established surveys which can be used to measure standard errors

3. Measures of non-sampling errors are calculated, viz.:

- Frame coverage errors
- Duplication in the frame/register used to conduct a survey
- The number of statistical units out of scope (i.e. number of ineligible units)
- Misclassification errors
- Measurement errors
- Processing errors
- Imputation rates
- Non-response errors

Findings

- Business register assessed for coverage before sampling is conducted
- Incorrect SIC codes still exist
- 18 months rule used
- Negative turnover
- Duplicate enterprises are flagged
- Snapshot of 78 785 enterprises used

Recommendations

- DST to negotiate frame sharing with Stats SA
- DST to draw sample
- Improvement of response rates needed urgently

Quality dimension : Timeliness

1. The preliminary results must be released according to the prescribed standard.

Findings	Recommendations
<ul style="list-style-type: none">•There is no standard for dissemination•No preliminary data is released	<ul style="list-style-type: none">•Adopt an international standard such as the GDDS and release accordingly•Consider releasing preliminary results while finalising the report

2. The final results must be released according to the prescribed standard.

Findings	Recommendations
<ul style="list-style-type: none">•There is no standard for when the results must be made available•Reference period: 2005-2007•Survey year: 2008•Release period: September 2011	<ul style="list-style-type: none">•Adopt an international standard such as the GDDS and release accordingly•Suggested standard for a biennial series: 12 months after the reference period (GDDS)



3. Project plan/schedule of key deadlines related to the statistical value chain must be compiled, viz.:

- data collection
- data processing

- data analysis
- dissemination

Findings

- There is a long delay between the reference period and the release of the data
- No evidence of each activity in the SVC has been taken into consideration e.g. dissemination

Recommendations

- Time lapse between planned activities and actual performance of the activities must be recorded
- Ensure that each activity in the SVC has been given sufficient time

4. The periodicity (e.g. monthly, quarterly, and annual) of release must conform to a data dissemination standard.

Findings

- Fixed periodicity
- Biennial: conforms to standard prescribed in Oslo manual

Recommendations

- Develop a data dissemination policy/standard that specify the adopted standard

Quality dimension : Comparability and Coherence

1. All data (including source data, related frame data, and related survey data) within the same series must use the same concepts and definitions, classifications, and methodology. Departures from common concepts and definitions, classifications, and methodology must be identified in the metadata and archived.

Findings

- Concepts and definitions, classifications and methodology are sourced from the Oslo manual

Recommendations

- Develop a data dictionary, classification standard and methodology document based on the Oslo manual

2. Statistics must be consistent over time

Findings

- Three previous innovation surveys, with different scope, methods, data producers
- No time series

Recommendations

- No conclusions should be drawn regarding the direction or scale of any real changes between the surveys
- Standardise survey methodology

3. The statistics must follow an expected trend established over time. Any inconsistencies in the key variables must be reconciled.

Findings

- No time series
- No established trend

4. Data across comparable series, or source data, are based on common frames, identifiers, concepts and definitions, and classifications, and departures from these are identified in the metadata

Findings

- Closest comparable series is R&D survey which is based on the frame of R&D performers

Recommendations

- Strengthen methodology for both surveys to achieve better comparability of the surveys

5. Statistics must be checked for consistency with a comparable dataset. Inconsistencies must be reconciled.

Findings

- No directly comparable results for the series exist
- Indirectly comparable series: R&D survey; Quarterly Employment Statistics; Annual Financial Statistics

Recommendations

- Figures related to expenditures on R&D and innovation can be compared to the R&D survey data
- Turnover and employment figures can be compared to Stats SA economic surveys
- Where results deviate from expected results data must be reviewed and updated if required

Thank you