National Workshop on Implementing Data Integration Approaches for Official Statistics

Kathmandu, Nepal

20-24 February 2023

DAY 5

Tell us about your experience using the toolbox yesterday

5350 4923

Communication of Integrated Statistics

Agenda

1. Introduction
2. Key Audiences
3. What is Different about Communicating Integrated Data
4. Examples
5. Final Words

1. Introduction

- To **communicate** means "to make something known, to impart and to transmit information". Importantly, communication is a two-way process, involving information and messages going back and forth between a sender and receiver.

- Communication goes beyond "dissemination", which can be a one-way activity to publish statistics online or in print. It involves talking with stakeholders and data users to capture and understand data needs, finding the most efficient way to meet their needs, explaining the methods and results and seeking (and getting) their feedback to continually improve statistical products.

1. Introduction

- Communication is vital for understanding the consequences of underlying data quality and errors in linkage, and for appropriately taking this into account when data users make use of integrated data.

  - Researchers should be proactive in evaluating quality of linkage: data users need to know what to ask for, in terms of information required to evaluate the quality of linkage

  - Data linkers (including trusted third parties) need to be willing to provide details of the DI processes used to create linked datasets

1. Introduction

• Communication is essential throughout the entire data cycle. Communication activities include the following:

Understanding needs and demands for data;

Establishing data-sharing agreements;

Explaining methods;

Presenting results;

Seeking from and giving feedback to suppliers and users, and so on.
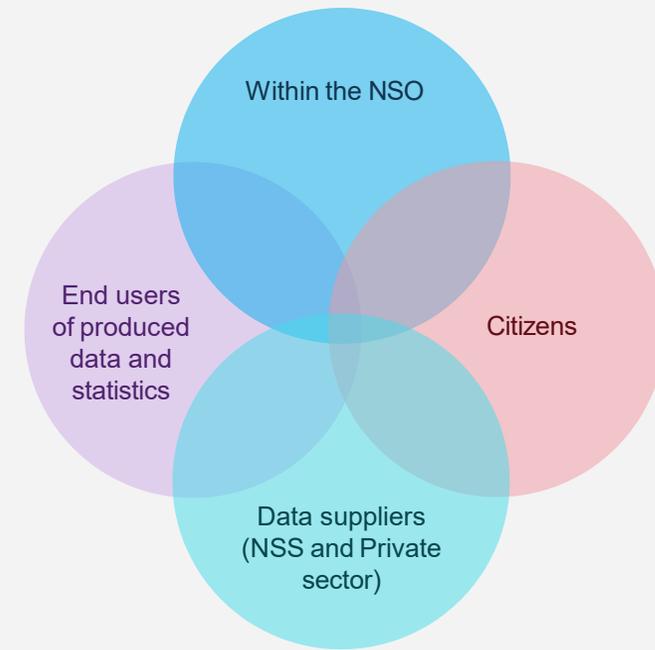
➢ Who are key audiences?

2. Key Audiences

- Producing data and statistics by integrating data from different sources is a relatively new practice to most of the NSSs.

- This highlights the importance of communication to keep all stakeholders informed and ensure integrated data reach their intended audiences.

- Regarding communicating integrated data, there are several target audiences to keep in mind.

If the **DI project** was **worth doing, then** the **communication of the integrated data** is **worth doing well.**

2. Key Audiences

Note: NSO, national statistical office; NSS, national statistical system.

- **Staff within the NSO** who need to know national and organizational policies for DI and what the DI is doing and planning in this matter.

- **Other data suppliers** across the NSS that the NSO is working with. They need information on (a) how to standardize their data, (b) what data-sharing agreements are in place, and (c) how they should format and transmit the data. They also need to know (d) how the data are being used by the NSO and how confidentiality is being protected.

- **End-users** of the integrated data. They require information on why and how the DI has been done, the strengths and limitations of final product, and issues to interpret, use and analyze the data effectively.

- **Citizens** and respondents may not use the integrated data products directly but have contributed their information and want to ensure the safe and ethical treatment of private information by the government.

10

➤ What is different about communicating integrated data?

3. What is Different about Communicating Integrated Data?

- Three main aspects make communicating integrated data slightly different from other sources of official statistics

  - DI relies on building relationships, which is dependent on good communication. Much time needs to be spent explaining the needs for and benefits of DI and the role that other agencies need to play in regularly providing the required data.

  - DI uses new, emerging, and sometimes complex methods that need to be explained. The methods used for DI require good communication skills and practices. Written communication is needed to document the approaches and provide guidelines for others that need to contribute to or to replicate integration methods. NSOs and NSSs must be able to explain the value proposition of integration so that key audiences understand why it is being done and the benefits that will be realized as a result. It is also necessary to communicate about data integration methods to non-technical users in layman terms verbally and in writing.

3. What is Different about Communicating Integrated Data?

- Statistics and data producers need to be credible. Mitigating risks to trust and credibility is essential. Citizens need to be informed that DI is being done and that their concerns about data use and confidentiality are being addressed. NSOs should communicate that statistical confidentiality is an overriding principle and always a priority in the production of official statistics, as per the [UN Fundamental Principles of Official Statistics Principle 6](#) (Confidentiality Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.)

  - According to a [paper about the future role of national statistical offices in Europe](#), NSOs should be leading discussions to clarify the difference between concepts of privacy, security and (statistical) confidentiality.

4. Examples

- The Data Integration Partnership for Australia, a three-year $130.8 million investment to maximize the use and value of the Government's data assets from 1 July 2017 to 30 June 2020, has developed a diagram that provides a good example of communicating with a broad audience about DI. (See also ABS website on DI)

4. Examples

- The brochure has a few features worth noting:
  - A 1, 2, 3 step diagram shows in a clear and simple way how data are sourced and transformed into datasets and analysis, and into meaningful information produced.
  - Clear and brief explanation of the DI partnership (main box with white background) and the "components"(Data infrastructure and integration; Data assets; Data analytical units; Better communication and engagement; Technical review and Advice) − easy to understand.
  - Clear explanation of the value proposition – the costs ($130.8 million investment) and the expected benefits (i.e., new insights into important and complex policy questions through DI and analysis)
  - Branding government logos, giving status and credibility to the product.

4. Examples

- Some ideas on communicating about data integration with different audiences were gathered in the [ESCAP regional workshops on implementing data integration in Asia and the Pacific](#), which are available on DI-CoP. These include suggestions on how to build relationships and mitigate risks to trust and credibility when communicating within the national statistical offices, with data suppliers, end users, citizens and respondents.

5. Technical Reports

- Any DI project should be accompanied by full documentation of the methods used for record linkage. It has two main uses
    - to allow a peer review of the methodology : A peer review is needed in order to provide assurance that the project has been carried out according to best practice
    - to provide a record of what has been done for the future : If methods are well documented, then capability can be greatly enhanced. Good documentation reduces the likelihood of making mistakes, and means that future work can build on past work, rather than rebuilding it from scratch

5. Technical Reports

- Suggested Contents:
    1. A stand-alone summary for the 'interested lay person'
    2. The reason for the integration
    3. A description of the input datasets
    4. A description of the matching methods used
    5. Details of the results achieved.
    6. A description of the final linked file

    And if appropriate

    7. Recommendations of things to do differently next time
    8. Other options that were thought of (or tried) and rejected

5.1. Why Reports Fail

1. There is no logical structure.

2. Ideas are not well thought out.

3. Work is disorganized.

4. Assumptions are made which cannot be justified by evidence.

5. There are too many grammatical and spelling mistakes.

6. Sentences and/or paragraphs are too long or too obscure.

7. It is obvious that ideas and sentences have been taken from other sources.

8. There is too much repetition.

9. There is too much irrelevant information.

10. Summary and conclusions are weak.

5.2. Presentations of Reports

- Another method of presenting a technical report is through an oral presentation
    - It is better to provide both a written report and an oral presentation
- If you want people to take notice of your results, you need to produce a good presentation.
- You can enhance your presentation with animation, artwork and diagrams which make it more interesting for your audience.
    - Know your audience

5.2. Presentations of Reports

- Effective oral communication is an important skill in research.
  - Some people have a natural talent, at giving effective public talks.
  - Not all may have such gifts, but **talents** and **competencies** in speaking can be **developed** (with **hard work** and **practice**).

- Most probably, not everyone in audience will read your report.
  - Your work and your research report will thus likely be judged by the quality of your presentation.
  - If you have a poor presentation, then your report will be poorly understood

- The primary purpose of your presentation is to educate, to provide information to the audience.
  - If your presentation is poorly expressed, your effort working on your report may be put to waste.
  - Too many details will likely not be remembered by the audience.

5.2. Presentations of Reports

- In report writing, there is a need to consider your readers and to plan your report.

- In considering your presentation, similarly, consider your audience and realistically plan your presentation.

- Remember the 9Ps:

    Prior Proper Preparation Prevents Poor Performance of the Person Putting on the Presentation.

5.2.1. Know Your Audience

- Identify areas of common interest
- Anticipate questions to be raised
- Do not underestimate nor overestimate
- Have an overall and didactic plan
  - type of talk and duration of talk
  - audience composition and expectations
  - **one shot at presentation!**

Composition :
general audience or specialists?
expected number of listeners?
a friendly or an interactive audience?

Expectations :
novel concepts or prior knowledge?
    basics have to be covered clearly, and early

5.2.2. Some Pointers

- **Short talks (15 minutes or less):** more difficult to prepare… essential points + lots of practice

- **Longer talks (say, 30 minutes):** must sustain people's attention

- **Rule of Three**: audience will likely remember only three things from your presentation.

  - *Vidi, Veni, Vici* (I came, I saw, I conquered) - Julius Caesar

  - **Stop, look and listen** - Public safety announcement

  - "Our priorities are Education, Education, Education" - Tony Blair

5.2.2. Some Pointers

- Create an outline of your talk.
- Be logical in your presentation.

Main point   Branches   Supporting points
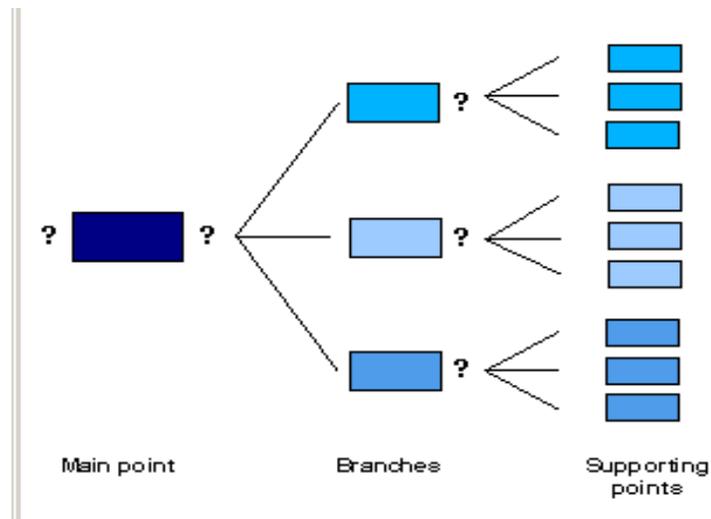
- Be realistic about how much material is appropriate

5.2.2. Some Pointers

- Try to have openings which will catch people's interest & attention

- Plan out transitions (links between successive elements of the talk)

- Conclusion of talk should summarize main concepts
    - Give a signal ("To summarize…")

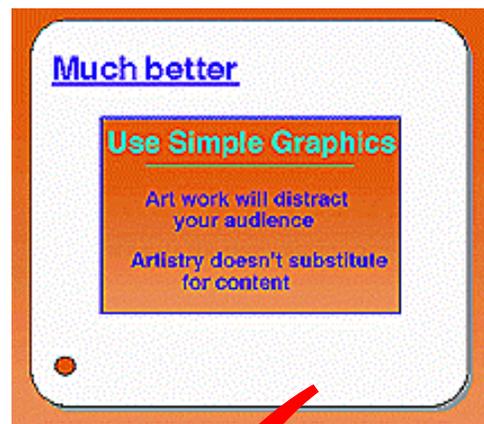- Communicate results effectively w/ visuals, e.g., graphs, pictures, tables, diagrams, photos.

    Confucius: "A picture = 1K words"

- Visuals must be VERY BIG!!! . . . but don't use big tables; .

5.2.2. Some Pointers

- **Be prudent** in the use of colors, fonts, artwork, animation and slide transitions!
(Too much art may be distracting).

Suggestion: Be artistic … but in moderation

- Use colors to emphasize some words.
- Run a spelling check on the words in your slides
- Don't use small fonts
- Don't overcrowd slides
- Don't change formats
- Don't use pale colors (such as yellow)
- Don't read off slides.
- Practice, practice, practice!

Final Words

- Increasingly, new data sources are becoming available to official statistics organizations. These new data sources can be used to provide new official statistics, address new or unmet data needs, lower response burden, overcome the effects of declining response rates, and address quality, coverage and bias issues in surveys to monitor national development plan and to examine success in meeting national and international reporting commitments under the SDGs. The focus on DI is increasing around the world.

Final Words

- The rich, informative datasets created by DI can potentially provide an invaluable source of information for research relating to health and society, allowing new insights into research questions that could not otherwise have been addressed with currently available datasets.

  - Although the value of administrative data linked to surveys and censuses in research studies is well accepted, the dynamic, error-prone or incomplete nature of administrative data can make linkage less than straightforward.

  - Methods for DI, particularly record linkage, have evolved over the years to accommodate imperfect data, but current techniques cannot eliminate linkage error entirely. With human involvement in the creation of these increasingly large and complex integrated data sources, recording errors will always be an issue and lead to uncertainty in linkage

Final Words

- Measurement error affects inference in the integrated data – it can lead to bias in estimation, which can be severe. The measurement error processes arising from probabilistic record linkage are complex and non-standard. A few causes of measurement error in data linking include:
  - units incorrectly linked, so that data from one unit is incorrectly associated with another unit (false positive links)
  - in many-to-one linking, statistics computed using only a few sub-units are used to measure characteristics of all sub-units
  - in many-to-one linking, characteristics of sub-units are inferred from features of major units (and vice versa).

Final Words

- This training provides a starting point for developing DI expertise om Nepal by reflecting on some of the issues faced and solutions found in DI projects, and by pilot testing a DI project. Combined with discussions and resources gathered and available through the DI-CoP of ESCAP, the 2020 DI-CAS, along with the regional workshops on implementation of DI in Asia and the Pacific, it is hoped that the NSO of Nepal, working with other data suppliers in government, can learn from this project for the future DI undertaking. To this end, all participants are invited to join DI-CoP,12 and learn more about DI, with the aim of accelerating this practice in Nepal.

Group Discussion:

- What are some of the data integration opportunities in Nepal?

- What are your suggestions for enhancing Nepal's capacity for data integration in terms of data access, coordination and collaboration among data holders, interoperability, etc.?

- After this training, what do you think the next steps should be?

- When you return to work, what is the one change you make to apply your learning from this training?

- Any other thoughts?

अन्त्य

END