

Developing an essential health data science skills curriculum for health research

1 Background and Rationale

Developing training and resources to build data science capability is critical. The development of new expertise in data science and the skills to overcome the well-recognised barriers that limit the collecting, sharing, analysis and use of high-quality data, are essential to addressing the current inequalities encountered by health research teams in low-resource settings.

There is a need to determine where the barriers and enablers to engagement in data science lie and so we will identify, by consensus and via an evidence-based approach, what are considered essential health data science skills. An essential data science curriculum can be built around those consensus themes. Resources, tools and training materials can be identified, developed or redeployed taking account of what already exists and what has proven effective. Mechanisms for dissemination of the resources, tools and training materials also need to be determined.

Definition of ‘essential health data science skills’: the minimum skills required for someone with limited experience in data science to be able to use data science approaches effectively for trustworthy and effective health research.

2 Aims and objectives

Aim

To identify the essential health data science skills needed to undertake effective and trustworthy health research using data science approaches.

Objectives:

1. to develop a new essential curriculum for health data science
2. to develop tools and resources to cover the priority gaps in health data science training

To note: this process aims to identify the *essential* health data science skills, i.e. the minimum skills required for someone with limited experience to be able to use data science approaches effectively for health research. We are not seeking at this stage to identify *all* or *advanced* health data science skills.

3 Process stages

All stages of the process will be carried out in collaboration with a cross-regional working group with representatives from three research institutions in Asia (icddr,b), Africa (Africa CDC) and Latin America (Fiocruz). Surveys will be piloted with partners before being delivered.

3.1 Setting the themes

- The landscaping and knowledge gap analysis (KGA) will be run through a web-based literature review, stakeholder interviews and a survey. A review of the results of the TGHN Essential Health Research skills Delphi to identify skills relevant to data science will also be undertaken at this stage; Working group discussions will take place to review the literature and to decide on the skills to be progressed into the crowd consensus review.

3.2 Crowd consensus review

Survey 1

- The survey will list the skills identified through the landscaping and knowledge gap analysis exercise and will be disseminated through a number of approaches; as an online survey, but also given out at regional data science or research capacity building workshops, conferences, training sessions where these take place.
- Respondents will be asked to consider whether each skill is 'essential' to a data science curriculum and rate them accordingly, using a 5-point Likert scale (Strongly agree, mostly agree, agree, mostly disagree and strongly disagree). Respondents will also be able to include any skills they feel are essential and which are not already included in the list. Essential will be defined as any skill reaching 75% consensus or above.

Survey 2

- The list of skills will be determined by the results of stage one and will include skills where no consensus was reached, and any new skills identified through survey 1. The survey will be disseminated as an online survey but also given out at regional data science or research capacity building workshops, conferences, training sessions where these take place.
- Respondents will be asked to consider whether each skill is 'essential' to a data science curriculum and rate them accordingly, using a 5-point Likert scale (Strongly agree, mostly agree, agree, mostly disagree and strongly disagree). Essential will be defined as any skill reaching 75% consensus or above.

3.3 Stakeholder review

- Results of the crowd consensus surveys will be reviewed and validated. The specific aims of the stakeholder review are to:
 - consider the suitability of the list of skills and groupings as an accurate reflection of the content;
 - evaluate the applicability of the proposed Essential Global Health Data Science Skills Curriculum findings to the global research community.
- The stakeholder review will conclude the process of developing an essential global health data science skills curriculum framework based on consensus findings.

3.4 Panel members/respondents

Led by working groups:

- Develop criteria to be met by panellists/respondents ensuring a diverse range of stakeholders and taking account of different geographical contexts and global health data science perspectives
- Define recruitment process for panellists/respondents (likely to be a combination of open and targeted recruitment with regional partners identifying key collaborators).

3.5 Implementation of the consensus

During this stage we will seek guidance from the community on how to disseminate the curriculum which could include sessions such as:

- Promotional webinar via TGHN Regional Hubs and the Global Health Data Science Hub and TGHN broader network. Investigators from Fiocruz, icddr,b and Africa CDC will lead the promotion and identification of key collaborators to be invited to participate in their region.

3.6 Ensuring equity

- Language considerations (ideally, we will deliver the surveys in 4 languages - English, French, Spanish and Portuguese - but the cost and time requirement will need to be taken into account;
- Led by working groups with global representation, identify broad but appropriate criteria for panellists/respondents;
- Transparency (we will publish the process and associated documents on TGHN Regional Hubs and Global Health Data Science Hub).

4 Survey platform

- Microsoft Forms for delivery of the Landscaping and Knowledge Gap Analysis and 2 crowd consensus surveys. Excel and 'R' (possibly NVivo) for quantitative and qualitative analysis.

5 Data management and analysis

5.1 GDPR compliance and consent

- Consider data storage and anonymisation
- Password protection
- Restricted access
- Consent to proceed included in each survey
- Statement on data storage and use in each survey
- Permission to collect name and email addresses included in each survey (not mandatory collection)

5.2 Analysis

- Analysis to be undertaken by HDR Global
- Results to be reviewed by the working groups and investigators ahead of stakeholder review webinars and/or meetings.

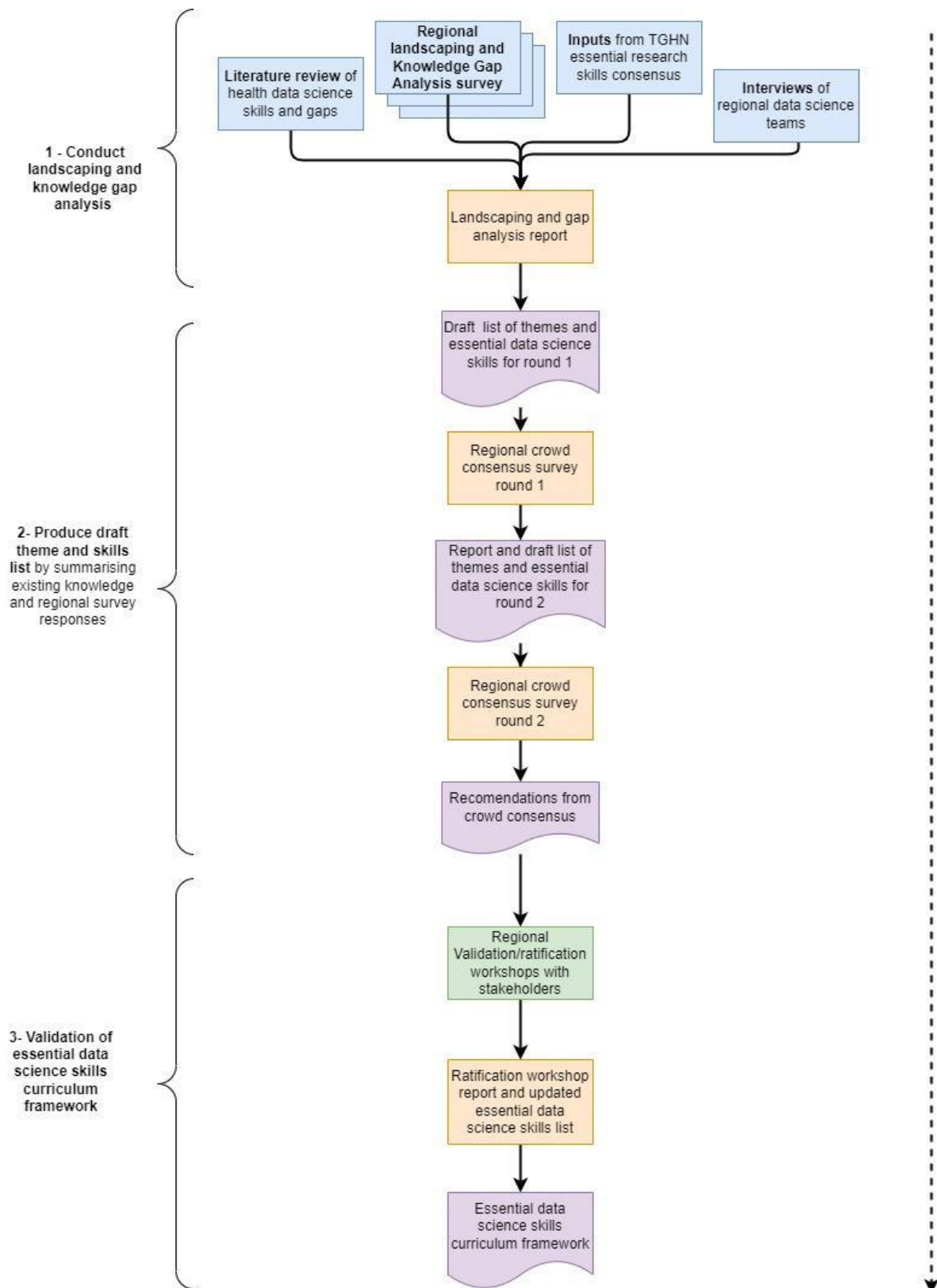
6 Results

- Shared via stakeholder review webinars;
- Published following implementation webinars, led by investigators in each region.

7 Outputs

- An essential data science curriculum *framework* based on consensus which meets the needs of the global health data science community in low- middle-income settings.

Appendix 1 - Crowd Consensus Process Flow



Appendix 2 - Methodology

Methodology background: Delphi process for 'crowd' consensus

The Delphi technique is a consensus building method that gathers experts to discuss issues (Developing a Consensus Led Essential Research Skills Training Curriculum, Arancha De La Horra Gozalo et al, 2022). The process goes through a series of cycles. In each cycle, a panel of experts is presented with a set of statements to rate. The Delphi process typically starts with a round of open-ended questions to form the basis of the initial pool of topics to rate. Controlled feedback is then given that shows how each individual's ratings compare with the whole group.

The Global Health Network has conducted several studies to reach a consensus using an adapted version of the Delphi approach (*Developing a Consensus Led Essential Research Skills Training Curriculum, Arancha De La Horra et al, 2022*; and *COVID-19 Research Priorities Identified by the Global Research Community, Trudie Lang et al, July 2020*). For example, one of these studies aimed to identify essential health research training skills gaps across Africa, Asia and Latin America in a three-stage approach:

- conducted a research training gap analysis reviewing responses of 7176 participants from 153 countries collected from 2017 to 2019, some of these were in-person workshops and some online. This generated a list of 98 research-training themes;
- a two-round Delphi study was carried out online to prioritise the outcomes of the gap analysis and find consensus on what constituted the minimum set of research skills;
- conducted a Stakeholder Review and Implementation workshops online which reviewed the global applicability of the findings and how they could be most effectively implemented.

The process resulted in an evidence-led research skills training curriculum formed by 108 themes mapped into 13 modules that any research group could use to implement and guide research training programmes. This work is now a formal TDR report.

The same Delphi approach was also used to assess the relevance of the WHO blueprint during COVID-19 (*COVID-19 Research Priorities Identified by the Global Research Community, Trudie Lang et al, 15 July 2020*) and more recently to determine whether a community of practice can support better pharmacovigilance across Africa.

These examples demonstrated that this approach of asking a large community of practice for their view, (collected online or in person) can achieve consensus. These data can then be combined to form the themes and also to serve as stages of review, as per the Delphi approach. This 'crowd' approach adds the important step to the standard Delphi of returning to the community of practice, the participants, in forming this consensus to take the findings back to them and ask them for their views on how to implement and take forward the findings.