

ISARIC Data tools Development Route

The ISARIC member assembly in Accra (September 2024) included three sessions dedicated to the ongoing development and use of ISARIC data tools within the scope of the Clinical Data/Epidemiology Platform. These sessions focused on presenting data innovations, gathering input from members, and prioritising future development efforts to ensure that the tools continue to address the member's needs. A brief description of the current tools is shown in Table 1.

During the **Data Innovation session**, we presented the current state of ISARIC data tools, with a particular focus on launching the first versions of ARC (Analysis and Research Compendium) and BRIDGE (BioResearch Integrated Data Tool GEnerator). This session highlighted the collaborative contributions from various networks, especially in the creation of new modules for ARC. Following participation and feedback from attendees, we received valuable suggestions aimed at fostering a more collaborative environment, seeking continuous improvement and making the tools more user-friendly.

In the **Analytical Pipelines Prioritisation session**, the focus was on aligning the development of reproducible analytical pipelines (RAPs) with the priorities of ISARIC members. We shared the findings from the 2023 ISARIC member assembly survey, highlighting how this survey guided the development of the tools introduced during the **Data Innovation session**. Additionally, we presented preliminary results from the 2024 survey conducted before the Member's Assembly, which emphasised the site resources, capabilities, and critical research questions that need to be rapidly addressed during outbreaks. The session facilitated a productive discussion on prioritising these research questions and their corresponding RAPs, ensuring that future development efforts align with the needs of the ISARIC members.

Finally, we conducted a **Workshop session** in which a step-by-step demonstration of how to use ISARIC tools was offered, with a particular focus on BRIDGE and ARC for CRF creation and data standardisation and collection. Participants were shown how to create CRFs using ARChetype templates or develop them from scratch. A CRF test was created and uploaded to REDCap, followed by real-time data collection. The session concluded with a demonstration of VERTEX (Visual Evidence & Research Tool for Exploration), where the data collected in REDCap was automatically visualised for descriptive analytics. This demonstrated the principle behind the RAPs and how the tools can be used from data collection to analysis, allowing researchers to gain valuable insights from their data efficiently.

In this report, we highlight and discuss the main results from the data prioritisation exercise and the feedback during the ISARIC Member's Assembly that will lead to the creation of a development route for the data analytics tools. A general summary status regarding the usage of tools:

- More than 30 sites have used BRIDGE (with ARC 1.0.0 as a source) to develop their CRFs for different diseases of interest (e.g., Oropouche in South America - Brazil, Dengue in Southeast Asia – India, mPox in Africa – DRC, and Nipah in Bangladesh).
- BRIDGE has been made publicly available with the four CRF templates (ARChetypes: COVID-19, Dengue, Mpox and H5Nx). Also,
- VERTEX has recently been used to analyse mPox data in DRC

Table 1 – Description of ISARIC data tools within the Clinical Epidemiological Platform

Clinical Epidemiology Platform (https://github.com/ISARICResearch)	An ecosystem of ISARIC tools designed to capture, process and analyse data in a standardised format for the purpose of efficient evidence generation. Includes (but is not limited to) ARC, BRIDGE, VERTEX.
Analysis and Research Compendium (ARC)	ARC is a comprehensive machine-readable document in CSV format, designed for use in Clinical Report Forms (CRFs) during disease outbreaks. It includes a library of questions covering demographics, comorbidities, symptoms, medications, and outcomes. Each question is based on a standardised schema, has specific definitions mapped to controlled terminologies, and has built-in quality control. ARC is openly accessible, with version control via GitHub ensuring document integrity and collaboration.
BioResearch Integrated Data Tool Generator (BRIDGE)	BRIDGE is a web-based application designed to operationalize ARC and edit any ISARIC CRF and tailor it to outbreaks in a particular context. By selecting and customising clinical questions and ensuring necessary data points for each, BRIDGE automates the creation of Case Report Forms (CRFs) for each disease and specific research context. It generates the data dictionary and XML needed to create a REDCap database for capturing data in the ARC structure. Additionally, it produces paper-like versions of the CRFs and completion guides.
Visual Evidence & Research Tool for Exploration (VERTEX)	VERTEX is a web-based application designed to visualise data through graphs and tables based on relevant research questions that need to be quickly answered during an outbreak. VERTEX uses reproducible analytical pipelines. Currently, we have pipelines for identifying the spectrum of clinical features in a disease and determining risk factors for patient outcomes.
Reproducible Analytical Pipeline (RAP)	A RAP corresponds to a set of instructions, activities and flows to read, prepare and analyse data, consisting of statistical, mathematical and/or machine learning modelling steps. RAP are currently implemented as open source codes in the ISARIC Clinical Epidemiological Platform, being customisable to different settings.

Data Prioritisation exercise

In the 2024 data prioritisation survey, responses were gathered from representatives of 37 sites and networks. The survey focused on two main aspects, the first being the level of resources and capabilities across several key areas, such as creating CRFs, conducting data analysis, and managing databases. The results revealed that levels of capabilities consistently exceed the level of resources across all tasks (Figure 1). This indicates that ISARIC members possess strong expertise and capacity and the future development of ISARIC data tools should provide the necessary resources to fully leverage the existing capabilities within the network. Here we present and discuss the main results from the survey.

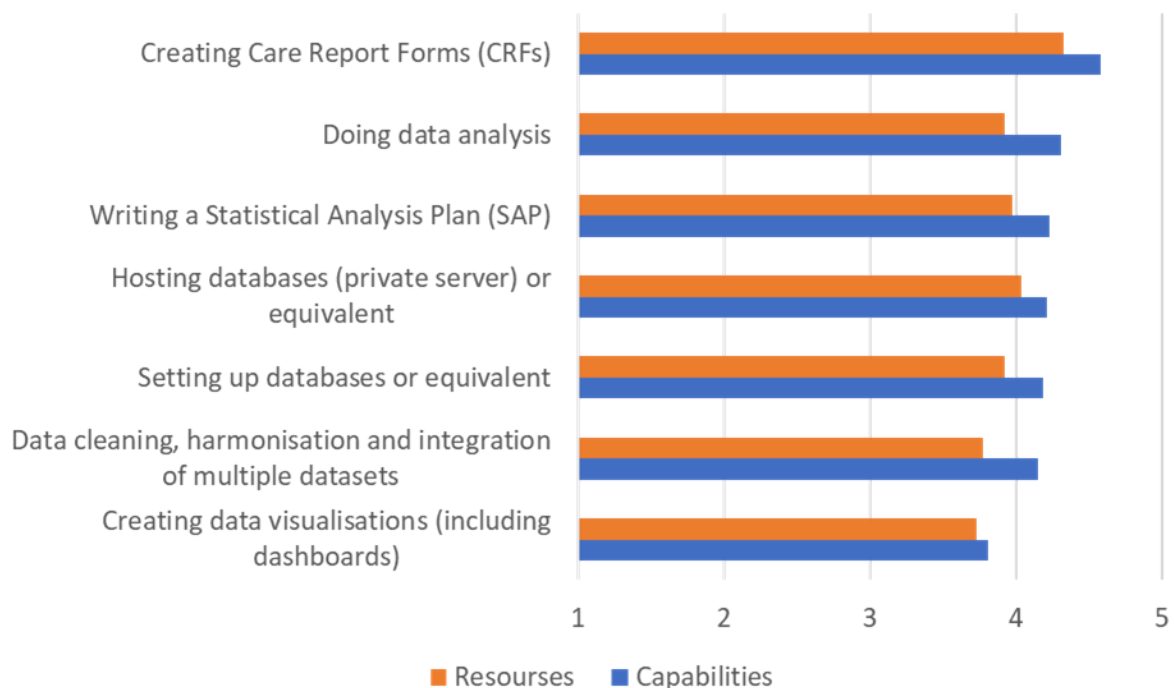


Figure 1 – Average scores regarding reported Resource and Capabilities levels from ISARIC partners (1 – Low and 5 - Strong). Responses were obtained from the ISARIC 2024 Data analytics tool survey. Participants were members or representatives of ISARIC networks or sites, and 37 sites responded the survey (As of Oct 08, 2024)

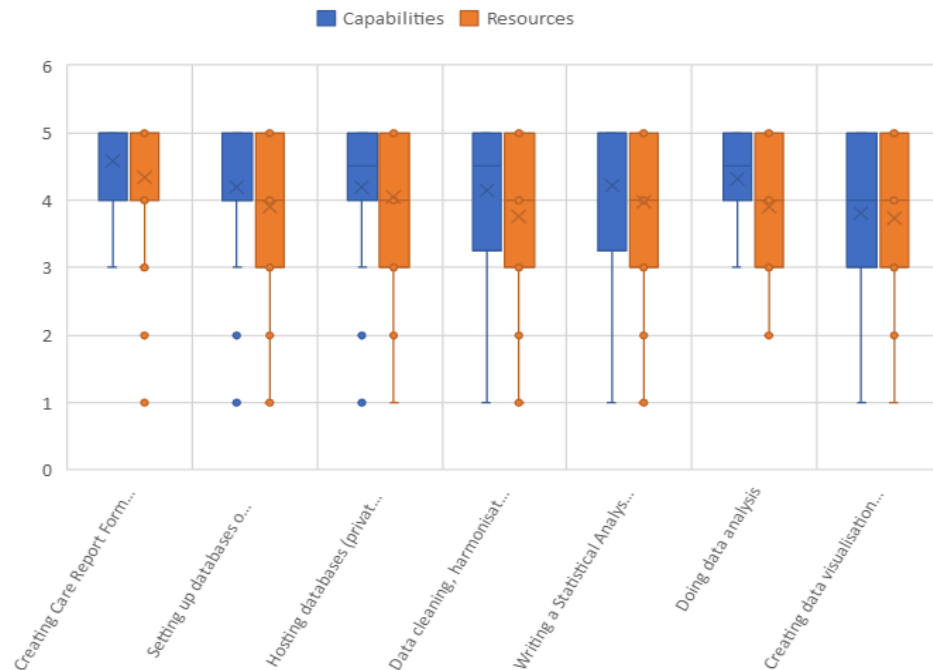


Figure 2 – Distribution of scores regarding reported Resource and Capabilities levels from ISARIC partners (1 – Low and 5 - Strong). Responses were obtained from the ISARIC 2024 Data analytics tool survey. Participants were members or representatives of ISARIC networks or sites, and 37 sites responded the survey (As of Oct 08, 2024)

With respect to the Resources and Capabilities (Figure 1) data curation and data visualisation were the lowest-rated items in the survey, indicating a need for improvement in these areas. This suggests that the development of reproducible analytical pipelines (RAPs) and their integration with VERTEX should be prioritised in future development efforts. Additionally, incorporating tools such as the auto-mapper and auto-coder will be essential to streamline the data harmonisation process, allowing the efficient integration of multiple datasets. Though results appear consistent regarding the distribution of scores for each item (Figure 2), a few sites reported low scores, especially with respect to Creating CRF and setting up and hosting databases. Those will be analysed individually.

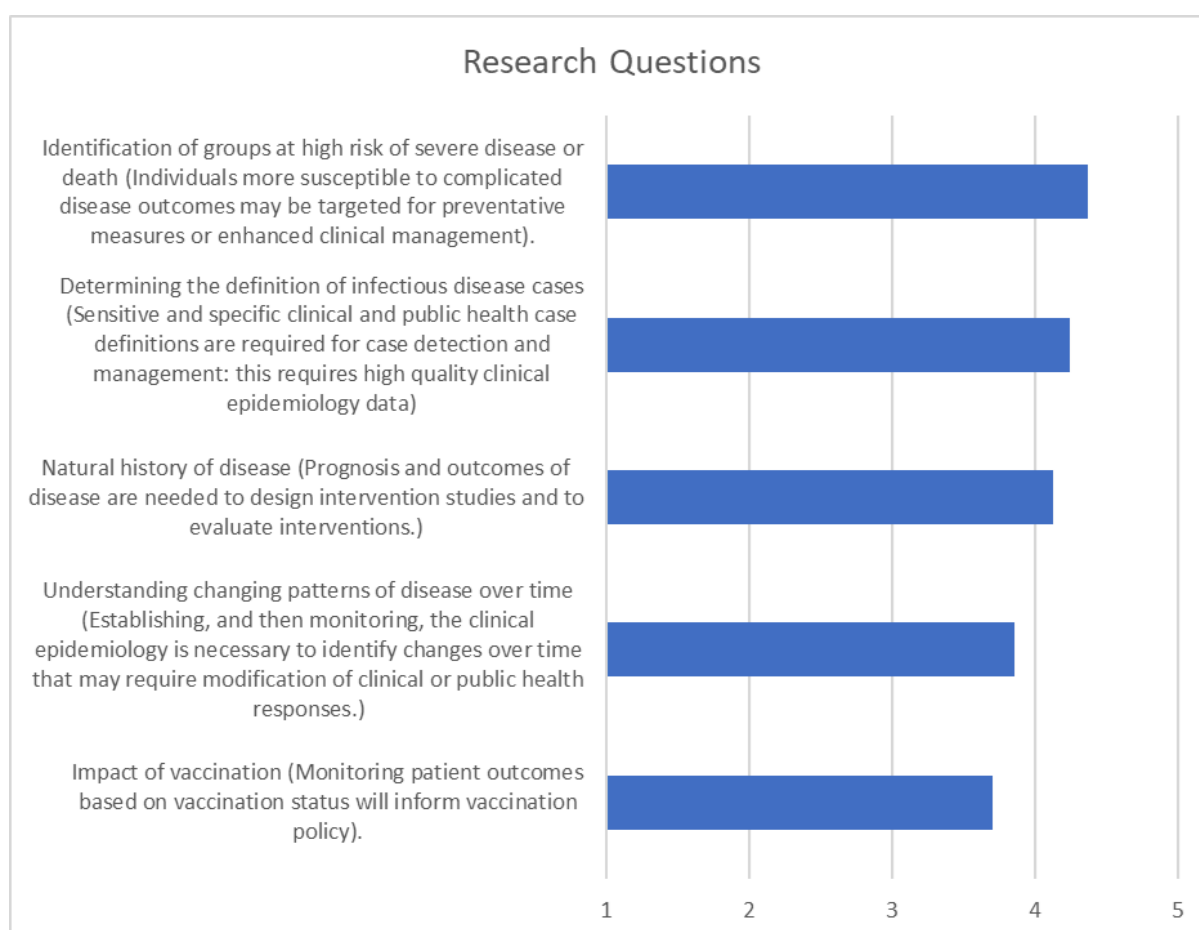


Figure 3 – Average scores regarding reported importance level of research questions from ISARIC partners (1 – Not important and 5 – More important). Responses were obtained from the ISARIC 2024 Data analytics tool survey. Participants were members or representatives of ISARIC networks or sites, and 37 sites responded the survey (As of Oct 08, 2024)

The second aspect of the survey focused on the prioritisation of research questions during an outbreak (Figure 3). The highest-rated questions highlight that the most critical answers at the onset of an outbreak revolve around quickly understanding the nature of the disease, identifying high-risk groups, determining case definitions, and understanding the natural history of the disease. These are essential for guiding early interventions and shaping the clinical and public health response.

Research questions regarding the later progression of the outbreak, such as changing disease patterns over time and the impact of vaccination, while important, are prioritised for later stages of development.

Development Route for ISARIC Data Tools

Based on input from the ISARIC member assembly sessions and survey responses, the following development route has been proposed for ISARIC data tools to address the needs of the members.

1) Development route for ARC (Analysis and Research Compendium)

- Definition of ARC Dimensions: Expand ARC's capability by defining new dimensions that fit different research areas. For example:
 - Specific modules: Pediatric, Maternal, Severe patients.
 - New data classes: Epidemiological, Clinical trials, socio demographics.
 - Syndromes: respiratory, neurological.

- ARC New Structure: Implement a new structural framework to accommodate these additional dimensions and enhance flexibility.
- Network Collaboration: Different networks will take ownership of developing specific dimensions within ARC, fostering a collaborative development environment.
- Multi-language Support: Introduce multi-language functionality to make ARC accessible to a broader global audience and enhance usability.
- Create an app to add questions to ARC (Define ARC update process)
 - Key ISARIC members/Partners: PENTA, VERDI, BACON

2) Development route for BRIDGE (BioResearch Integrated Data Tool GEnerator)

- Collaborative CRF Development: Introduce features for collaborative creation of CRFs, where users can create projects, share files, track changes and establish collaboration. Ensure user access management so that CRFs can be saved, edited, and accessed later. Data Governance
- Library of CRFs: Build a library where users can deposit and share CRFs (files), facilitating knowledge sharing and reuse across the network.
- Study Design: Develop an interface that helps users select study design elements and links them to applicable research questions.
- Integrate with ISARIC-VERTEX.
 - Key ISARIC members/Partners: CONTAGIO

3) Development route for Data Ingestion process (Auto mapper and Auto encoder)

- Transform data from national datasets (SINAN – Dengue in Brazil, MoH in Peru)

4) Development route for the Reproducible Analytical Pipelines (RAPs, including VERTEX)

- Code Contribution System: Create a system where anyone can suggest additions or edits to the code, encouraging open-source development and continuous improvement.
- Developer Integration: Allow developers to plug in new features and tools, ensuring the pipelines remain adaptive and scalable to evolving research needs.
- Collaboration and Communication: Encourage collaboration with other sites, allowing for technical development and exchange of feedback and ideas for RAPs and other tools.
- Develop RAPs for related to this research questions (in order of priority):
 - Identification of groups at high risk of severe disease or death (Individuals more susceptible to complicated disease outcomes may be targeted for preventative measures or enhanced clinical management).
 - Determining the definition of infectious disease cases (Sensitive and specific clinical and public health case definitions are required for case detection and management: this requires high quality clinical epidemiology data)
 - Natural history of disease (Prognosis and outcomes of disease are needed to design intervention studies and to evaluate interventions.)
 - Understanding changing patterns of disease over time (Establishing, and then monitoring, the clinical epidemiology is necessary to identify changes over time that may require modification of clinical or public health responses.)

- Impact of vaccination (Monitoring patient outcomes based on vaccination status will inform vaccination policy).
- Launch a generic SAP for the research question and related endpoints.
- Building software/analysis community
 - Key ISARIC members/Partners: Medical Research Council Gambia