

A New Standards-based Grammar for Linking Aggregate Datasets

Ritz, D¹, Jolliffe, B², Santas, X³, and Kariuki, J³¹ecGroup Inc.²HISP³Centers for Disease Control and Prevention

The theme of this session is the linking and cross-referencing of disparate aggregate datasets that need to be combined for purposes of reporting and/or analysis. The session leverages, as a global case study, the US Government's President's Emergency Plan for AIDS Relief (PEPFAR) programme. PEPFAR is a \$7 billion per year programme supporting the delivery of HIV-related services, medicines, and commodities in 58 low and middle-income countries (www.pepfar.gov). PEPFAR has an immense datastore of monitoring, evaluation and reporting (MER) indicators that have been collected from all its supported countries over the course of its 15 years of operations.

The goal of the session is to describe for attendees a newly-developed, standards-based grammar for describing interoperable aggregate data exchange and the message schemas needed to support it. The session facilitators are the primary authors of this new standard. Using the PEPFAR case study as a working example, the session explores how disparate HIV data elements and indicators from PEPFAR-supported countries are cross-referenced to each other and collected into a single central datastore to support analysis, management and reporting across the global programme. The specific HIV example will be elaborated upon to illustrate generalizable techniques that can be applied to linking aggregate datasets in other use cases (e.g. reporting to the annual WHO global health observatory, multiple provinces reporting to a federal health data institute, etc.).

The session will be facilitated by Xenophon Santas and James Kariuki of the US CDC, Bob Jolliffe of the University of Oslo's Health Information Systems Programme (HISP) and Derek Ritz of ecGroup Inc (a Canadian health informatics consultancy). All four facilitators are members of the Quality, Research and Public Health (QRPH) technical committee of the international digital health standards body, Integrating the Healthcare Enterprise (IHE; www.ihe.net). The session's content and examples will leverage the facilitators' first-hand experience working on HIV-related projects in low and middle-income countries (e.g. South Africa, Rwanda, Kenya, Malawi, Zimbabwe, Uganda, Sierre Leone, Vietnam, the Philippines and elsewhere).

It is intended that the session will be conducted using an interactive workshop style. Attendees who wish it will have an

opportunity to engage in participative (hands-on) learning. To get started, information will be provided about the standards-based grammar and how it works. Then, results from the facilitators' efforts leveraging this method to link multiple disparate HIV-related datasets will be presented. As a hands-on activity, attendees who have notebook computers will be able to connect to an open source software solution (www.dhis2.org) and "play in a sandbox" to try for themselves some of the techniques that have been described.

As learning objectives, it is expected that attendees will:

1. Be introduced to data linking use cases outside of their everyday experience
2. Learn about a new technique for expressing aggregate content schema that supports interoperable data exchange
3. Apply new skills in a hands-on, worked example.

