
A Health and Population Research Documentator for the Pentaho Data Integration Platform based on the GSBPM, GSIM and DDI

Chifundo Kanjala (London School of Hygiene and Tropical Medicine, UK)

Chifundo.Kanjala@LSHTM.ac.uk

Arofan Gregory (DDI, USA)

ilg21@yahoo.com

Jay Greenfield (DDI, USA)

jay_greenfield@hotmail.com

Abstract

DDI in partnership with the London School of Hygiene and Tropical Medicine has built a prototype documentator. It works with a data integration platform called Pentaho. Workflows on the platform are specified and execute at two levels — *what* and *how*. Each *what* decomposes into many *hows*. We map Pentaho *whats* to GSBPM subprocesses and specializations of these subprocesses that have been added to account for the domain (demographic and epidemiological surveillance). On the Pentaho platform, just like on many other data integration platforms, the *hows* are called transformations. Pentaho supports maybe twenty transformation types (Input, Output, Streaming, etc.) and across all the types maybe one hundred or more specific transformations (*hows*). The *hows* are the Pentaho service architecture. The documentator maps the *hows* for each *what* to DDI extended GSIM process steps and their services.

Keywords

[Keywords]