

Statistical Commission
Forty-fifth session
4 – 7 March 2014
Item 3(l) of the provisional agenda

Background document
Available in English only

Coordination of statistical activities in the United Nations system

Long-term vision for effective global data collection, data sharing and dissemination

Prepared by
National Institute of Statistics and Geography of Mexico

Long-term vision for effective global data collection, data sharing and dissemination

*Background document prepared by
National Institute of Statistics and Geography of Mexico*

Introduction

The purpose of this document is to define an appropriate and improved coordination mechanism for statistical data exchange within the United Nations system, at the international, regional and national levels. In this context, the UN Regional Commissions and the UN agencies play a strategic role in promoting the implementation of agreed standards within the countries and in supporting the integration and dissemination of statistical data across countries.

The proposed mechanism, based on a network of interconnected data hubs, envisions the progressive elimination of discrepancies in the data reported by countries through the implementation of statistical standards like SDMX.¹ Although the development of the proposed model is a long-term process, it can be implemented and adopted in stages in such a way that benefits can be achieved early.

Description of the Model

In this model, requesters and informants are connected by means of common hub, i.e., a mechanism on the Internet which acts as a transfer point capable of keeping records of each transaction, delivering notifications about each step in the exchange process, and also providing certifications to registered informants and requesters (Figure 1).

This model uses “push” and “pull” mechanisms in an alternating way. The requester initially uploads a request (push) and a notification to the informant is generated. The informant then downloads the request (pull), processes it, and uploads the response (push), after which a notification to the requester is generated. To complete this cycle, the requester downloads the response (pull).

¹ Statistical Data and Metadata Exchange, see www.sdmx.org,

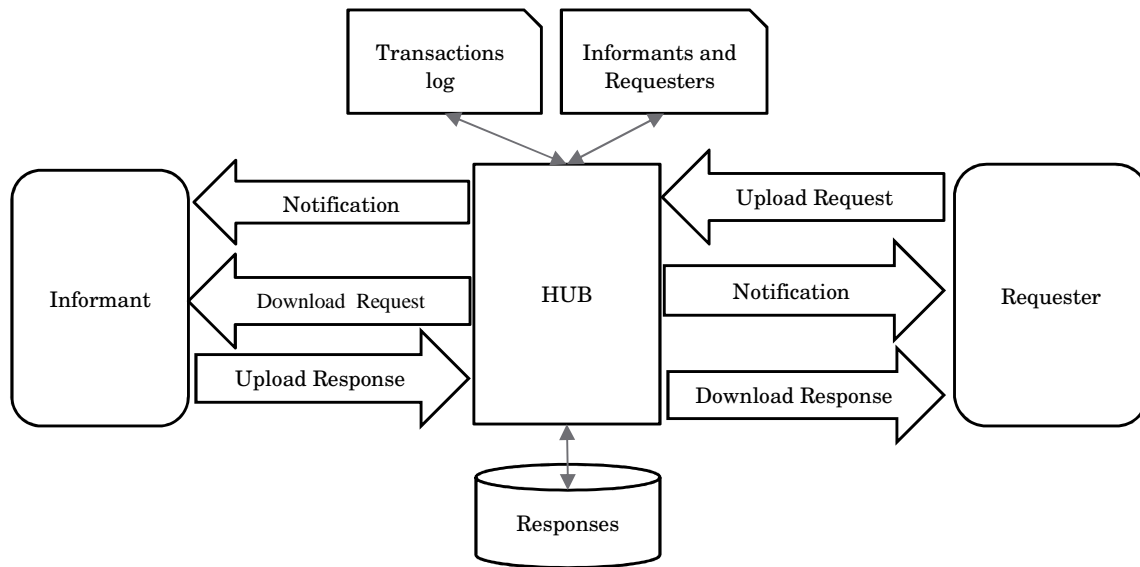


FIGURE 1; SIMPLIFIED VIEW OF THE MODEL

The hub architecture would allow the UN Entities to request and collect data-flows while informants are able to respond to a request by submitting a file or an URL providing access to an information source (the URL is expected to be in the future a call to a SDMX Web Service, as SDMX is gradually adopted by the countries).

Each of various hubs would manage an origin-destination matrix which would specify relationships between data flows and informants. In case a particular dataflow were missing in the matrix and the requester did not know who the informant is, a hub Manager would be able to determine which informant(s) could satisfy the request, redirecting it and updating the matrix.

The matrix, complemented with the log of the transactions managed by the hub and the authentication mechanisms, would be the basis for issuing electronic certificates to prove the origin of the information and the precise date when it was made available to the requester.

From a technical perspective, a hub would be a set of software modules based on a service-oriented architecture (SOA) with a common Application Programming Interface (API), allowing their implementation under different platforms and their interconnection in

different configurations. This common API would also allow requesters and informants to interact in different ways within the hub, such as through a web site or a direct connection to other systems. Thanks to this feature, the hub could represent a source of information for a global repository (like, e.g., UN Data), which by means of a software connector could extract the set of responses stored or accessed by the hub (Figure 2).

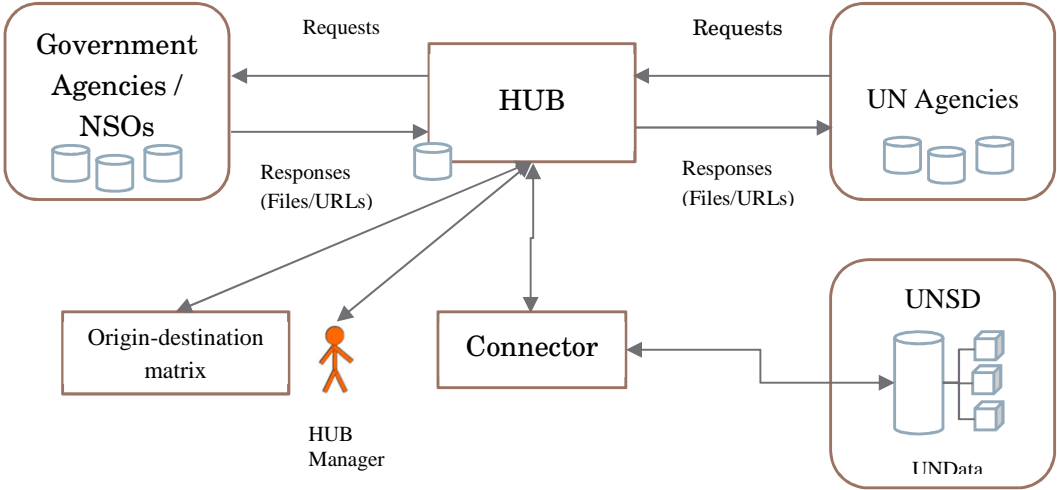


FIGURE 2; HUB MODEL

For a hub there is no difference between the interactions made with a human using a web interface and the interactions made with another software system accessing its API. Each hub is a piece of software that can be accessed by another hub.

The exchange mechanism can be extended in a very flexible way, because each hub can be connected to requesters, informants or to other hubs, as Lego pieces that can be arranged to conform different configurations. When connected to other hubs, they act as routers which can generate dynamic paths connecting requesters to informants with intermediate points of registry / control.

A feasible composition of hubs to support the coordination of information exchanges at national, international and global levels is represented in Figure 3. There is no limit on the number of levels that can be added on any side. For instance, countries can organize

functionalities will be complemented by a web form in which the informant can fill-in the information.

- Keep logs and copies of both requests and responses (files or URL's).
- Notify collectors when the responses to their requests are available for download.
- Determine a way to connect requesters and informants. A request can travel through several hubs until it reaches the appropriate informant (e.g., regional hubs, national hubs and government agencies' hubs). The response will go through the same chain of hubs back to the requester.
- Provide an Application Programming Interface (API) for external applications (like UN Data) to manage all the requests and responses.
- Support connection with SDMX-compliant artifacts, such as like web services and registries.

CASE STUDY

This section describes an instance of the request-response process, in order to illustrate the model proposed above. The process is described taking in account a possible evolution in the implementation of the mechanism, from an initial implementation stage to a more developed network of hubs.

In an initial implementation stage, a regional agency may make a data request through the hub by selecting an informant from a list of government agencies or NSOs registered in the origin-destination matrix. The hub would then send a notification of the request to the selected informant via email and internal inbox. Depending on the implementation, the request can be attached to the notification email or made available for download from the hub. Once it has processed the request, the informant responds through the hub by sending an URL or uploading a file. The hub then notifies the regional agency (the requester) that the response is available for download.

In a second implementation stage, the hub can be improved by including a hub manager, whose role is to help the regional agency identify an informant to answer its request. The

hub manager forwards the request to the appropriate informant, updates the origin-destination matrix, and monitors the request-response process.

The implementation of a hierarchy of hubs to improve the coordination of regional bodies in the UN statistical system and countries can be considered in a third stage. Hubs in each country would then keep logs of data exchange transactions for domestic purposes. A hub of each regional office would also maintain logs of the exchanges made by the countries of the region, while a global hub maintained by UNSD could have a global view of all the information exchanges made in the system.

In a fourth stage the global view can be connected to information systems like UN Data, feeding them with information flowing in the whole network. Then data from all the domains and countries can be queried by external users from a single system.
