Design of a Cyber Infrastructure for Accumulation and Distribution of Analytical Information Resource

Giorgi Ghlonti

N. Muskhelishvili Institute of Computational Mathematics

gg59ster@gmail.com

Project Scope

Attempts of reaching sustainable optimal position for subject areas are often faced with problems caused by presence of corrupted information resource, when information necessary for decision-making is not properly motivated and argued, sources of data unreliable, information obtained by the user - incomplete, inconsistent, irrelevant, etc. As a result, lack of clarity in cause-effect relationships, unawareness on the results of corrective actions is observed.

Making one of the major factors in social relations, analytical information becomes a social asset and turns into strategic resource of a society. Evidently, the problem of accumulation and utilization of this resource in industrial mode is actualized.

Industrial mode of information resource development implies possibility of its collection and completion on regular basis, its distribution among all stakeholders, with access to information limited only by the measure of its privacy and regulated (presumably) by economic and civil relations.

From technical point of view, development of a cyber infrastructure where accumulation of analytical information resource of a country, along with the measures providing appropriate quality of the resource, minimization of its cost and delivery time might be considered.

Existence of such a repository could be regarded as one of the conditions for sustainable development of a country, as its presence leads to:

- Improvement of interaction within different subject area;
- accumulation and sharing of knowledge in various spheres of activity;

- transparency of administration and control;
- realization of opportunities for decision making on the basis of modern management techniques;
- prevention (or even elimination) of conflicts.

Data collection according to coordinated models ensures integrity and consistency of information resource, comparability and compatibility of data, transparency of information space.

Among the requirements to the cyberspace, as a guarantee of quality of information resource provided, following may be mentioned:

- Openness, meaning ability to include new sources of information and new points and layers of data aggregation;
- Compatibility and comparability of data, arriving from different points of environment to data processing nodes and end user;
- Transparency as the possibility to trace sources and origins of any data aggregates;
- Security

Corporate analytical information resource is regarded as union of data clusters representing different subject areas (e.g. healthcare, education, industry, agriculture, etc). Solution involves, for each cluster, data acquisition, on some regular basis, from primary information providers of subject areas, its further processing and transformation into information aggregates, relevant to corresponding levels of subject area's organizational hierarchy, subsequent provision of information needs of different stakeholders, maintenance of long-term information archives.

Architectural decision should be flexible enough to be adaptable to the features of different communities, allowing them to organize their information space in suitable manner and to modify it easily in case of necessity.

The users of cyber infrastructure would be provided by different information services, both routine and ad-hoc. They would be equipped by the tools allowing specification of appropriate higher-order queries.

Management of cyberspace must be distributed among the levels of hierarchy.

A strategic level of management being responsible for planning information resource structure and content, forecasting evolution of the cyberspace, development and control of

quality standards for information resource, coordination and regulation of processes, associated with collection, transformation and distribution of data.

A local level of management - responsible for data processing.

Analysis of Requirements

The cyber infrastructure must provide the users with ability to cooperate at all stages of analytical information resource lifecycle, particularly:

- resource planning;
- resource gathering;
- resource accumulation;
- resource distribution.

Information Resource Planning

Information resource planning enables to ensure its completeness, integrity and consistency, relevance of information to the needs of particular consumers (located at different levels of subject areas organizational hierarchy).

The major tasks at the stage of information resource planning are the following:

- definition of the structure and content of information resource of the subject area;
- identification and classification of primary information providers;
- development of criterias for completeness, integrity and consistency of the resource;
- planning material, financial and other resources necessary for data collection and processing;
- determine access rights for different categories of users;
- schedule information collection process.

Definition of the Structure and the Content of Information Resource

By nature, analytical information resource is a hierarchical formation. Its structure and content is defined according to organizational model of particular subject area, accepted in a given community.

Thus it has to be allocated according to following dimensions:

- administrative hierarchy of subject area's organization model;
- set of subject-area specific descriptors;
- time dimension.

First dimension from mentioned above - administrative hierarchy of subject area's organization model, is a conformed dimension – it may be shared by several subject areas. For example, most subject areas related to social and economic relations in Georgia are organized according to administrative-territorial division of the country. Consequently, the above mentioned dimension of information resource would presumably contain following layers:

- country level (aggregated);
- regional level (aggregated);
- community level (aggregated);
- district level (aggregated);
- service provision level.

Fig. 1 denotes possible configuration for this hierarchy.

Along with this hierarchy there is a group of descriptors, subject area's information resource should as well be structured according to which. We've already mentioned it as the set of subject-area specific dimensions**.** They don't make a hierarchy, rather a set of independent dimensions describing activities in the subject area.

The last dimension to be mentioned is time dimension.

Facts to which dimensions are related are mostly panel data representing subject areas at appropriate level of detail. They provide either frequencies or absolute or relative indicators and are gathered in statistical tables.

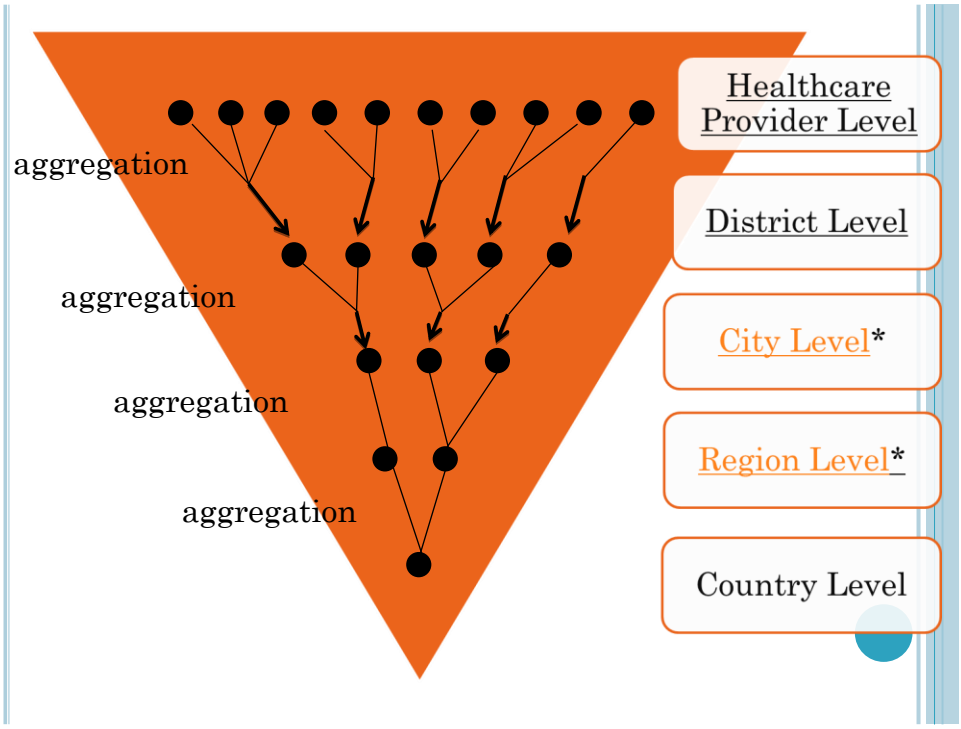

Fig. 1

The description of those tables, under the name of the Foundational Statistical Document (FSD) may serve as a source of metadata. Information Space (see below) of the Cyber infrastructure will be structured following FSD.

Identification and classification of Primary Information Providers

Primary information providers are granted by several subject-area specific characteristics, which are important from the point of view of content of information they provide. For healthcare management, among those characteristics may be denoted several:

- Type of services,
- Specialization,
- Type of HSP (hospital, pharmacy, clinic, health centers, etc) ,
- Ownership (private, state, etc.),
- Size of HSP (small, medium, large),
- etc.

These characteristics make the foundation of above mentioned set of subject-area specific descriptors;

Identification and classification of primary information providers is important, first of all, for determining the content of information resource, identification of homogeneous groups of information, provision of integrity of analytical information resource, comparability of different fragments of it, definition of criteria of classification.

Primary information provider is identified by cipher and FST

Scheduling Information Collection Process

Information collection is a multi stage process. It involves submission of data by primary information providers, check for completeness and consistency at given level of hierarchy, further transformation into information aggregates and indicators of higher levels of integration with necessary control at all stages.

This is the process of coordination of actions among different levels of organizational hierarchy, development of scenarios for standard and non-standard situations, schedule of compliance, finally leading to development of analytical information resource of the subject area for given period of time.

Data are accumulated in aggregation points. Each data aggregate receiving a unique identifier, representing its location in overall structure of analytical information resource.

Developing Criteria for Completeness, Integrity and Consistency of the Resource

Importance of this stage is obvious. It makes the basis quality control procedure of information resource. Use of IT technologies imposes additional requirements to formalization (semantic integrity constraints, etc)

Architectural Solution for Cyber Infrastructure

As it was mentioned above, designed Cyber Infrastructure should be responsible for providing computing and storage facilities for accepting data from primary information providers, its consecutive promotion up to the hierarchy of information resource structure and its transformation into forms relevant to various levels of subject area’s hierarchy and finally,

fulfillment of the needs of users in different information needs. Those requests include not only routine but also some special requests, provided (possibly) for additional fee.

The conceptual model of interaction with environment of the planned cyber infrastructure is depicted on Fig 2.

Here the major external entities interacting with the system are:

- Primary information providers submitting initial information;
- Global Managerial Unit;
- Consumers of the information resource accumulated in the cyber infrastructure.

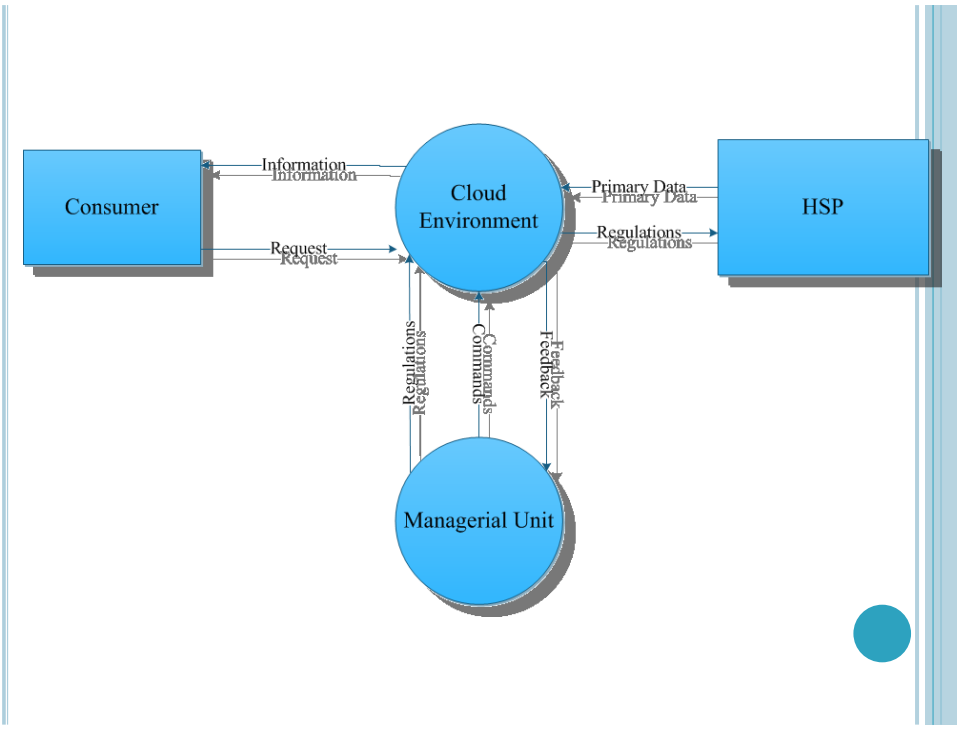

Fig. 2

The Global Managerial Unit is introduced for ensuring flexibility of the cyber infrastructure, its ability to maintain information resource of different subject areas.

Managerial Unit is responsible for conveying to the system information resource planning issues in form of commands thus determining the structure of its information space, quality assurance

of it, coordination and regulation of operations, associated with collection, processing and distribution of data.

The major information flows at this level are:

- Normative information and instructions sent by Managerial Unit and feedback from the cyber infrastructure;
- Basic data provided by the primary information providers;
- Request from consumers for different information services;

From architectural point of view the cyber infrastructure must contain at least following two layers:

- core service layer – encompassing components that present the core services of the cloud system;
- support layer – a middleware layer providing means for other layers to communicate and interact (e.g. message bus, database, transfer service, etc.);

Within core service layer following components are meant:

- Backend;
- Information Space;
- Frontend.

Backend is responsible for receiving information from primary information providers. Its existence is important because it serves as somewhat ETL point - a checking point against viruses and data transformation point necessary because apart from even in case of satisfying regulations the row data may be supplied in variety of formats (e.g. in form of Excel spreadsheets, xml files, etc.) and will becomes necessary to transform them in format supported by DBMS-s located in nodes of accumulation (see below) of cyber infrastructure.

Frontend is responsible for interaction with consumers of the information accumulated in the cyber infrastructure, providing them by common interface. From the frontend the cyber infrastructure receives also information from Global Managerial Unit.

Information space is a media for accumulation of information resource, supported with mechanism of data processing. It's a virtualized resource and must be treated both at logical and physical levels.

For being able to maintain several subject areas, Information Space must be split into several Sections, each dedicated to a particular subject area. We look at the structure of a Section as a hierarchy of points of information aggregation and further processing called *Nodes of Accumulation.* These nodes make up a hierarchy of suppliers and consumers of information and are arranged according to above mentioned structure of information resource

The lowest level of hierarchy in a given Section of Information Space receives initial information through the backend from the primary information suppliers.

Information thus got into the Section becomes subject to grouping, aggregation and additional processing gradually moving up and refilling the stages of Information Space.

In order to be able to perform its responsibilities, any node of accounting must be equipped by appropriate hardware and software (data mart, data warehouse, and accompanying software packages).

Software package at any given node of accumulation must be responsible for:

- Placement of data coming from subordinate nodes or input directly by the user;
- The check of integrity and consistency of data received by the node;
- Aggregation of data for obtaining the information relevant to given level of organizational hierarchy;
- Providing service for user's information needs at given level;
- Preparing the data for transmission to topping node.

DBMS schemas at the nodes of accumulation are build according to information, provided by FSD. Responsibility of support layer is to bring this information to the nodes of accumulation.

Conclusion

It may be claimed that implementation of proposed project will lead to creation of cyber-infrastructure supporting sustainable development of a country, through accumulation of knowledge and exchange of information among varieties of subject areas, increasing motivation for decision-making, administration, transparency of administration and optimization of management, prevention of conflicts and even ability to avoid them.

Analytical information will no longer be the privilege of a person or a social group, but rather a social asset, which created, maintained and shared by the entire community.