Implementing a Dashboard for Data Exchange on the REPOX Tool

Lizbeth Monica Cuba Samaniego
INESC-ID, Instituto Superior Técnico
Universidade de Lisboa
Portugal
Lizbeth.samaniego@ist.utl.pt

Abstract

The management of operational services where decisions depend on the perception of a lot of events can be supported by using interfaces with "dashboard" techniques. The usage scenarios of REPOX framework are examples of that, and therefore this thesis explores the hypothesis of a "dashboard" interface for that purpose. The REPOX platform supports the aggregation of bibliographic data for multiple channels (OAI-PMH, or even shared file Z39.50), its consolidation in local collections, and also its publication by multiple channels. In some real scenarios using REPOX, the number of these channels can reach the hundreds, with several million records. For that purpose we evaluated the various dimensions of information a manager of such a system requires, and developed and evaluated our proposals in line with this analysis.

Keywords
Dashboard; J2EE; Information visualization; Information presentation; Decision-making; REPOX.

1. Introduction

With continuous advances in technology, the rapid pace of today's business environment, and the need to create and manage data in increasing quantity, Managers are overwhelmed with reports and data produced from a multitude of Information Systems (IS). The problem is further exacerbated when the information in the reports is poorly presented and visualized, which is often distracting at the time of decision-making. This reality forces managers to consider implementing dashboards that could alleviate the problem of information overload.

Understanding the influence of information presentation and visualization on the effectiveness of decision-making is an important component of the user interface design in human-machine interaction. In this context, dashboards have been well received. The interest in dashboards is increasing because they are considered one of the most useful tools, for business intelligence.

Although dashboards have been adopted as a management tool, the scientific literature has failed to keep pace with the developments. While textbooks, e.g. (Few, 2006; Rasmussen et al., 2009) and articles in business mention e.g. (Miller and Cioffi, 2004; Kawamoto and Mathers, 2007) dashboards exist; only a few studies can be found in academic journals, providing little guidance for practitioners (Pauwels et al., 2009) and researchers (Yigitbasioglu and Velcu, 2012). More scientific work on dashboards is needed.
Motivated by these aspects, this thesis aims to study and implement a dashboard considering the available, open-source technology along with a substantial contribution from the limited academic point of view. In summary, the motivation for this work is to propose a dashboard interface for REPOX, an existing tool to manage processes for data transfer in digital libraries. The, also foundations for the design, implementation and evaluation of a dashboard for REPOX are also presented.

2. REPOX’s Dashboard

2.1 REPOX

Our case study for implementing a dashboard is REPOX. REPOX is an existing open-source tool to manage processes for data transfer in digital libraries. Figure 1 shows the functional areas involved in data transfer of digital libraries. The amount of data that REPOX handles is continuously increasing and consequently a practical way of reporting in real time is necessary. A general summary regarding REPOX is given in the next paragraph.

REPOX is an implementation of the concept of Metadata Repository (Nuno et al., 2007), which provides an open platform operating without commercial technology in its process of data preservation. This system manages collections of metadata, derived from various entities, and each interface is connected to a data source that is responsible for obtaining the records generated and checking them before their integration into REPOX.

2.2 Dashboards

Nowadays, dashboards is well-known; they are studied in IS, utilized in Business Intelligence (BI) and Business Process Management (BPM). However, a clear definition of a dashboard was not been given before (Few, 2006), indicating the lack of academic research on dashboards (Yigitbasioglu and Velcu, 2012). Although many different definitions exist, the most important ones are listed below.

“A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance” (Few, 2006).
2.3 Dashboard’s Purposes

To build a dashboard with the capability for the potential uses described above, the dashboard can be directed to four possible purposes, (Pauwels et al., 2009).

**Monitoring**

A dashboard helps to monitor performance by referring to the day to day evaluation of metrics that should result in corrective action. Monitoring could be considered as the dashboard’s most fundamental function.

**Consistency**

A dashboard enforces consistency between the alignment of measures and measurement procedures used across departments and business units.

Planning

A dashboard can be used for planning if its features include scenario analysis.

Communication

A dashboard communicates both the performance and the values of an organization to its stakeholders through the choice of the metrics.

2.4 Dashboards’ types

Currently, three major types of dashboards exist: operational, tactical and strategic (see Table 1). Each type emphasizes the top, middle and bottom layers of information and application to different degrees. The top layer graphically displays excepted conditions, the middle layer allows users to explore or “slice and dice” data from multiple dimensions and the bottom layer allows users to examine individual transactions and operational reports (Ikechukwu et al., 2012).

<table>
<thead>
<tr>
<th>Items/types</th>
<th>Operational</th>
<th>Tactical</th>
<th>Strategic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus</td>
<td>Monitor operations</td>
<td>Optimize process</td>
<td>Execute strategy</td>
</tr>
<tr>
<td>Emphasis</td>
<td>Monitoring</td>
<td>Analysis</td>
<td>Management</td>
</tr>
<tr>
<td>Users</td>
<td>Supervisors+</td>
<td>Managers+</td>
<td>Executives+</td>
</tr>
<tr>
<td>Scope</td>
<td>Operational</td>
<td>Departmental</td>
<td>Enterprise</td>
</tr>
<tr>
<td>Information</td>
<td>Detailed</td>
<td>Detailed/ Summary</td>
<td>Summary</td>
</tr>
<tr>
<td>Updates</td>
<td>Intra-day</td>
<td>Daily/Weekly</td>
<td>Monthly/Quarterly</td>
</tr>
<tr>
<td>&quot;Looks like a...&quot;</td>
<td>&quot;Dashboard&quot;</td>
<td>&quot;Business intelligence Portal&quot;</td>
<td>&quot;Scorecard&quot;</td>
</tr>
</tbody>
</table>

Because this thesis is about operational dashboard, in this section this type of dashboard will be defined.

**Operational Dashboard**

Dashboards for monitoring operations are often designed differently than those that support strategic decision-making. Operational dashboards are generally used by contract employees and their supervisors who directly interact with customers and services, or manage the creation and/or delivery of products. They are also used by managers or administrators to analyze data, monitor constantly changing activities, and respond quickly to events that might require attention. Consequently, an operational dashboard emphasizes monitoring over analysis and management. As a result, operational dashboards primarily deliver detailed information that is only slightly summarized, in real-time.
An operative user views information daily, for example to know the status of the information, information characteristics.

3. Dashboard Design for REPOX

In what follows, a set of steps for the implementation of a dashboard for REPOX are described.

Stage I: Dashboard Purpose

This step aids designers to identify user goals and understand the type of dashboard that is desired. Because this step limits the frontiers of the design, it is of paramount importance in the design phase. In this context, the following questions are necessary.

Who is my user?

In REPOX there are 3 different types of user: Administrative, Operator and Public. For example, the operator can be a library, which is responsible for the configuration of processes, check the status of implementation of processes and data release.

What value will the dashboard add?

A dashboard is currently unavailable for the open-source REPOX. The dashboard to be created will facilitate REPOX users’ understanding of data and assist their decision-making process through improved information presentation and visualization.

What type of dashboard am I creating?

An operational dashboard type is necessary and sufficient. This type of dashboard can for example shows the status of the settings, ingest status, etc.

Stage II: User requirements

In order to achieve users’ goals it is important to recognize the users’ requirements, since users requirements are considered to design the dashboard.

Because the dashboard is operational, operator requirement will be further considered.

For example, the user needs to know:

- Status of the records (Error, Cancel, ok)
- Amount of ingested records.
- Time spent in the ingestion.
- The most used protocols to ingest records.
- The status of the records, if they were deleted or updated.

Stage III: Selecting the Key Metrics

Ones the users’ goals and requirements are determined; the process of analyzing and selecting the PIs, KPIs, and general metrics must be carried out.

In order to monitor the data collection process, a set of KPIs and PIs, for Aggregator, Data provider, Data Set, and Records are detailed in the thesis. In this paper, metrics related to aggregators (defined below) are listed.

Aggregators (AGG)

An Aggregator is an entity that aggregates Data Sets from Data Providers, with the purpose of making them available also through the OAI-PMH protocol. The system administrator is allowed to register and manage Aggregators. The following metrics are usually requested by the operator.

- Total number of Aggregators, Data Providers, Data Sets and Records.
- Total Data Providers, Data Sets and Records for each Aggregator.
- Total monthly amount of Aggregators, Data Providers, Data Sets and Records.

Stage IV: Metadata management

The architecture of the data is very important. If the available data is collected and understood properly, it is possible to achieve good information for the optimization of the effort invested in
operations of the organization. The capability to handle data is also an important step in the design of a dashboard. To manipulate the available data by REPOX web services, it was necessary to implement transformation of XML files. The backups resulting from the transformation was stored in a DB.

Stage V: Solution technology selection

The capability to handle the DB along with the good choice of functionally features for ideal information presentation and visualization allows producing key metrics. However, the information presentation and visualization of the dashboard is contingent upon the selected solution technology. To this end, in this step, an open source technology solution that include dashboard, such as the OSBI platforms will be selected.

Technology for Dashboard

There exist two types of technologies for dashboards, the Commercial open source (COS) and the full open source (FOS), but several individual solutions implemented on the top of these technological solution. Because this thesis is all about OS technology, some aspects related to FOS is given in this paper.

FSO is completely free (no enterprise solution is available) and thus all the functionalities are available to the community for free. It is distributed under the GNU Lesser General Public License (GNU LGPL) (Golfarelli, 2009).

Due to the existence of several OS individual solution technologies (OSBI platforms can be also considered), such as SpagoBI, Pentaho, Opel, Vanilla, etc, a methodology to conduct the comparison which finally allowed to select the OSBI platform is described in what follows.

Method of conducting the comparison

First of all, a selection of a subset of the most promising platforms for the purposes of the thesis was selected. The steps followed for the election of the platform are:

- Investigate which platforms meet the minimum requirements for building a dashboard, such as including OLAP, etc.
- Evaluate and perform a comparison of the basic functionalities to create a dashboard by considering the interesting work (Golfarelli, 2009).
- Evaluate and finally select the OSBI platform.

For evaluating and selecting OSBI platforms, see Table 2. Gives a quantitative suggestion.

Stage VI: Design Features

Ones designers selected the key metrics to be included in the dashboards, it is time to think about how the information should be presented and visualized. In this sense the functional characteristics play an important role. A proper selection of functional characteristics, meaning the right choice of chart and table style (information presentation), allows efficient and attractive display of information (information visualization).

### Table 2. Final comparison of the platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>ETL</th>
<th>OLAP</th>
<th>Dashboard</th>
<th>Documentation/Usability</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpagoBI</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>4.50</td>
</tr>
<tr>
<td>Pentaho</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4.25</td>
</tr>
<tr>
<td>Opel</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3.25</td>
</tr>
<tr>
<td>Vanilla</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.00</td>
</tr>
</tbody>
</table>
With recent concepts of visual characteristics, the choice of colors and appropriate typographies to display information are not difficult. In short, in this step the main features of the dashboard (functional, form, structure, and visual feature) are design.

**Information presentation design**

The information can be presented in many different ways, for example, numbers, graphs, tables and bars.

Presentation format flexibility is the ability to view data in different ways (e.g. tables or graphs) point-and-click (like a drill down). This kind of features can be achieved in OSBI platforms having dashboard implement on it. Moreover, real-time notifications and alerts are also important for users.

The information utilized in monitoring REPOX has data obtained in real-Time, so for example it is possible to detect problems that may occur in the process of ingest. Concerning to alerts it may also implemented in a dashboard. However to do it you need threshold that should be discusses and defined in further improvements of the present dashboards. This kind of features also can be achieved in OSBI platforms.

Regarding drilling down capabilities, i.e. ability to display general to more detailed information, it should be stated that this feature can only be achieved in some OSBI platforms. This aspect should be also considered when an open source solution technology is persuaded. Some OSBI platforms need to be further implemented in this sense.

**Information visualization design**

Information visualization is an important component of the information presentation discussed above. In fact, the design of the information visualization helps to avoid distraction and comprehend given information displayed for example in a dashboard. Consequently, even if there is a fit (i.e. all the required information and features are available to the user), a poor visual design (e.g. excessive use of colours, low data-to-ink ratio, etc.) may confuse and distract users. Therefore, a visual feature helps to have an easy understanding of any given information.

In most of the cases it is necessary to have simple page or one glance view of the information provided in a dashboard, i.e. the idea is to concentrate on a single page all the facts and figures. Information should be visualized in a simple manner, and show what is important and highlight what deserves our attention "key factors". The key idea is to facilitate decision making for better results.

A good use of colours, i.e. the correct use of colour scheme, preattentive variables to make important information stand out, allows facilitating the understanding of the information. A row use of color scheme can produce distraction and cause consuming time labor for managers.

**Stage VII: Layout design**

Metrics (KPIs and PIs) are essential to build a dashboard for REPOX. This step schematically distributes KPIs (layout design) for the ultimate goal of this thesis, i.e. a dashboard for REPOX.

The layout of this dashboard is as follows (see Figure 2):

At the top (horizontally) the most important KPIs are given. When any interface is viewed by a user, the first glance is directed towards the top of the interface and for this reason; metrics of importance are fixed at the top of the dashboard. Obviously, the presentation of such information should be given along with simplicity and adequate read mode. This last aspect will be seen in stage IX. Metrics can be presented in the form of bars, columns, tables, pie charts, line graphs, etc.
The most important metrics of our case study were selected through a survey that was conducted to users of REPOX. Statistical analysis related to the evaluation can be viewed in Section 4.

At the middle part (horizontally), metrics with a good level of importance is normally found. The information contained in the metrics is red after the first ones, described in the previous paragraph.

At the bottom part (horizontally) metrics with a level of importance corresponding to the hierarchy shown in the previous two paragraphs are usually shown. Typically, users see this type of information at the end of an interface.

**Stage VIII: Deploy the dashboard**

A dashboard that displays daily information is typically used to monitor Week-to-Date, Month-to-Date, Quarter-to-Date, or Year-to-Date progressions. This type of data is frequently displayed in operational dashboards.

An operational dashboard helps managers understand their daily operations from various views and angles. This type of dashboard usually includes many charts and graphs. One major characteristic of an operational dashboard is that it combines data from multiple sources. In this dashboard general information of REPOX ingest process is displayed.

**Stage IX: Users training**

This section describes a fast and practical way to read information from the metrics of the dashboard for Operative users. This section describes details of how the information should be read. As it can be seen, at this level, dashboards were already designed based on the previous steps. In stage VII, the dashboards metrics were distributed hierarchically, containing three main parts. Based on this structure, user training should be implemented. Readers can found further information in the thesis.
4. EVALUATION

As stated above the first evaluation formats aim to select information presentation and visualization techniques which can be adopted for further uses in REPOX. The evaluation considers four different formats.

The first format of evaluation is to evaluate the effectiveness of the perception of the information in graphs for a generic metadata aggregator. In particular, it intends to evaluate the effectiveness of each technique, scheme color, gridlines and data-to-ink ratio. The three Options intend to represent, the total number by COUNTRY of Data Providers, Data Sets and Records ingested by a particular metadata aggregator.

The second format of evaluation is to evaluate the effectiveness of the perception of the information for a generic metadata aggregator. Specifically, it aims to evaluate the effectiveness the gridlines and fonts in the tables to present this type of information.

The third format of evaluation is to evaluate effectiveness of the information presentation (table vs graphs) for a generic metadata aggregator. Especially, it intends to evaluate the effectiveness of each technique, scheme color, gridlines, font’s contrast and data-to-ink ratio.

The final format of evaluation determines the effectiveness of the perception of the information for a generic metadata aggregator. Moreover, it intends to determine the effectiveness of each technique, scheme color, gridlines and data-to-ink ratio.

In each format of evaluation, three different Options are given. Each evaluation format (with different options) tries to evaluate different techniques of presentation of information, which are important when designing, for example, a dashboard. The options are described in what follows (see Table 3).

Option 1: This is the choice obtained from the present studies, i.e. good engineering practices on information presentation and visualization.

Option 2: This is commonly used in the literature. Its use is in debate.

Option 3: This is an alternative option also used in the literature.

Table 3. Average user satisfaction ratings for the 3 Options (first format of evaluation).

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty in perceiving the information</td>
<td>6,29</td>
<td>4,86</td>
<td>4,71</td>
</tr>
<tr>
<td>Amount of time needed to understand the information</td>
<td>6,29</td>
<td>5,57</td>
<td>5,00</td>
</tr>
<tr>
<td>Clarity of presentation of graphic elements</td>
<td>6,00</td>
<td>4,57</td>
<td>3,43</td>
</tr>
<tr>
<td>Simplicity of the presented information</td>
<td>6,00</td>
<td>4,29</td>
<td>4,43</td>
</tr>
<tr>
<td>The color is used so sparingly</td>
<td>4,43</td>
<td>5,29</td>
<td>5,00</td>
</tr>
<tr>
<td>Adequation for the color scheme</td>
<td>5,43</td>
<td>5,86</td>
<td>5,43</td>
</tr>
</tbody>
</table>
The formats of evaluation also have some questions that allow participants to rank the Options, Ranking from very low (1) to very high (7) accordingly and following the method by Likert (John, 2012). In this case, with six questions asked to obtain the preference of users, i.e. the option that helps to perceive information in short time and with simplicity and clarity. Also if the color is used so sparingly, and the good ad equation of color scheme.

A user satisfaction questionnaire was completed by the participants at the end of the session. To improve the methodological rigor, some statements were asked in a favorable way toward the prototypes tested and some were phrased in a negative manner. Responses were collected using a 7-point Likert scale with 1 = Very Low and 7 = Very High.

In order to improve readability, questions which required a lower response to reflect a positive satisfaction were flipped prior to analysis (e.g. if the user rated a question with 1, meaning the highest possible value, it was flipped to 7). A post-hoc analysis of the satisfaction ratings allowed the averages of the user satisfaction ratings and their standard deviation (σ) (see Figure 3).

![Figure 3. Average user satisfaction ratings, format of evaluation 1.](image)

In order to improve readability, questions which required a lower response to reflect a positive satisfaction were flipped prior to analysis (e.g. if the user rated a question with 1, meaning the highest possible value, it was flipped to 7). A post-hoc analysis of the satisfaction ratings allowed the averages of the user satisfaction ratings and their standard deviation (σ) (see Figure 3).

In order to improve readability, questions which required a lower response to reflect a positive satisfaction were flipped prior to analysis (e.g. if the user rated a question with 1, meaning the highest possible value, it was flipped to 7). A post-hoc analysis of the satisfaction ratings allowed the averages of the user satisfaction ratings and their standard deviation (σ) (see Figure 3).

The second part of the evaluation is similar to the first one, with the exception of the target and number of options for each format of evaluation. The questionnaire was also different.

Table 4. Comparison of the Indicators

<table>
<thead>
<tr>
<th>Indicadores</th>
<th>Description</th>
<th>Avrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicador 14</td>
<td>Mapping of Ingest Type, Data Sets, Records and Status</td>
<td>6,3</td>
</tr>
<tr>
<td>Indicador 1</td>
<td>Aggregators, Data Providers, Data Sets and Records</td>
<td>6,1</td>
</tr>
<tr>
<td>Indicador 6</td>
<td>Metadata Format, Data Sets and Records</td>
<td>6,0</td>
</tr>
<tr>
<td>Indicador 8</td>
<td>Status Ingest, Data Sets and Records</td>
<td>5,9</td>
</tr>
<tr>
<td>Indicador 11</td>
<td>Ingest type of Data Set and Records</td>
<td>5,8</td>
</tr>
<tr>
<td>Indicador 12</td>
<td>Data Sets and Records by Ingest Methods</td>
<td>5,6</td>
</tr>
<tr>
<td>Indicador 10</td>
<td>Data Sets and Records by Day</td>
<td>5,5</td>
</tr>
<tr>
<td>Indicador 2</td>
<td>Aggregators, Data Providers, Data Sets and Records by Months</td>
<td>5,4</td>
</tr>
<tr>
<td>Indicador 3</td>
<td>Data Provider types</td>
<td>5,3</td>
</tr>
<tr>
<td>Indicador 7</td>
<td>Type Collection, Data Sets and Records</td>
<td>5,2</td>
</tr>
<tr>
<td>Indicador 13</td>
<td>Data Sets, Records and Ingest Time by Month</td>
<td>5,1</td>
</tr>
<tr>
<td>Indicador 9</td>
<td>El porcentaje del uso del canal del ingest</td>
<td>4,8</td>
</tr>
<tr>
<td>Indicador 4</td>
<td>Data Providers, Data Sets and Records by Country</td>
<td>4,4</td>
</tr>
<tr>
<td>Indicador 5</td>
<td>Data Providers, Data Sets and Records by Aggregators</td>
<td>3,9</td>
</tr>
</tbody>
</table>
The second survey is to choose indicators with higher priority, see Table 4. The evaluation points were: relevance of the information shown, easiness of understanding the information, and confusion generated by the amount of information.

5. CONCLUSIONS

A novel operational dashboard for REPOX is presented in the context of this thesis. A solution methodology, based on OSBI Platforms and good information presentation and visualization along with ranked metrics is used to build the dashboard. Evaluation procedures support the paper output and following conclusion:

The SpagoBI OSBI platforms appear to be the best choice when an evaluation procedure is carried out. SpagoBI makes available a remarkable number of Good engineering practices on Information presentation and visualization should be considered to improve dashboards due to the limited academic research on the topic. The evaluation procedure of the dashboard corroborated this need.

A usability study, run with REPOX users, systematically investigated the various information presentation styles along with the degree of importance of the metrics to be used in a dashboard. The study participants were monitored while performing pre-defined task sets in seven different format of evaluation of metrics, and afterwards completed a satisfaction survey.

Study results revealed a significant support for good engineering practices on information visualization and presentation for dashboard, which supports the further research on the open topic. In addition, user satisfaction ratings corroborated those performance results, with the current present information presentation of and visualization receiving significantly higher ratings than the common ones.

Overall results show that, although being clearly a work in progress, the style how the information are presented and visualized are valid in helping to understand the metrics contained in the dashboard for REPOX.

6. REFERENCES


