Data Analytics and Decision Support in the context of ERP and beyond for giving CSIR a Competitive advantage – two case studies from HR (Human Resources) and MM (Materials Management) Modules.

Nalin Kumar Singh and Rajesh Kumar Singh Roushan
CSIR – National Metallurgical Laboratory, Jamshedpur

Abstract

The evolution of data is a natural corollary of the evolution of CSIR as an organization. Big data offers challenges as well as opportunities. ERP is an important milestone as a decision intervention towards transformation of CSIR. It is an integrated, fast and transparent enterprise level solution which overcomes unnecessary duplicity and the problem of silos. It has inbuilt data analytics which leads to decision support. The paper emphasizes upon the need for complementing ERP with other data bases, use of data mining and data analytics to produce Business Intelligence to provide decision support. In today's cut throat competitive environment, decisions should be based on facts and not on intuitions and beliefs. The paper gives many illustrations and details to processes of GPF and vendor management for developing analytics framework and show how it leads to vital decisions which make an impact.

Keywords

ERP, Data Mining, Data Analytics, Decision Support, Performance Mapping, HR Intervention, Organizational Effectiveness, Vendor Management and Performance System.

Introduction

Council of Scientific & Industrial Research is a premier R&D organization of India comprising of 38 Research Laboratories working in different spheres of scientific specialization. CSIR having completed 70 years has evolved from an organisation for substitution of imported technology to an important player in the global R&D scenario.

With the passage of time, the expectations from CSIR have increased manifold. In order to deliver, CSIR needs to reposition itself and accordingly re-engineer its delivery mechanism, communication at all levels both inside the organizational structure and the world outside. Needless to say, with the emergence of China and other new players in the business and technological arena, the need for excelling has become all the more prominent.

With the evolution of the organization in terms of both structural and cultural heterogeneity and growth, the way we look at the data has also evolved. The paper based organization of yester years was inefficient, slow and prone to errors. Electronic workplaces are not a new thing. Use of computers did this many decades ago. But any unbiased assessment makes it clear that it was only partially electronic and partially manual. Records were stored manually in Service Books and Registers. These were assessed and retrieved from various Sections/Divisions; notes were written at every stage for demanding and giving information which was amenable to errors, inadvertent and deliberate. Naturally, the whole process was time taking arbitrary and prone to error. Modern IT provided the facilities of speeding up

processes, increasing accuracy and data storage for recall. As it was the prevalent fashion, IT initially was used to reduce labour. Now it has been harnessed in innovative ways for competitive differentiation.

From file processing to database system and evolution of database system from hierarchical and network database to Relational Data Base Management System (RDBMS), data processing has evolved. Data are basically of two types – data that results within organizational boundaries and data resulting from business interactions which comprises of other business, suppliers, sellers, end-users, regulators of different kind, government, society at large etc. It makes an information network. The result is emergence of big data, a formal buzzword for data deluge which comes from various sources of digital communication e.g., images, streams, censors, multimedia, scientific and engineering databases, knowledge based information systems, geo-spatial and temporal databases coming from satellites and GIS systems etc. Data resulting from structured transactional data as well as unstructured data e.g., from social websites are equally important from the point of developing business intelligence and analytics. Negotiating with this big data (Volume, Variety & Velocity) is a challenge for all organizations whereas, latent in this challenge are the opportunities for building efficient organizations of tomorrow and drawing competitive advantage.

Methodology

The paper applies an exploratory survey of the concepts of ERP, data mining, data analytics and decision support system. Then, it moves on to a systematic study of One CSIR ERP system to decipher what kind of intelligence and analytics it offers and whether it can be utilized for decision support. It takes us to the stage where we are trying to see how efficiently the data from ERP modules offer data patterns which develop into Business Intelligence and how data mining and data analytics techniques can be added to offer better/advanced business analytics.

Two case studies of business analytics, one each from HR (Human Resources) and MM (Materials Management) modules are being attempted to show what kind of insights exploitation of data offers for decision making in CSIR.

One CSIR ERP

In continuation of this organizational evolution, CSIR has come up with its innovative project of ERP implementation having the following advantages.

1. Interoperability – To capture the data in a web based form and transfers the same to the enterprise application modules for processing and sharing such information in a common format.
2. Integrated planning and direction – An integrated enterprise develops consistent values, visions and goals and prescribes relevant programmes/functions to achieve such goals.
3. Integrated organizations – It keeps the internal organizational units and the external stakeholders to be integrated.
4. Integrated processes – Integration of business processes across varying functions.

5. Integrated Human Resources – An integrated enterprise application provides measureable parameters for achievement and facilitates most effective use of Human Resources across the organization.
6. Integrated Financial Resources – Policies, Practices and Procedures can be standardized to facilitate dynamic financial management.
7. Integrated Information Systems – Information is managed as an asset for the entire enterprise. Effective use of enterprise knowledge as a strategic resource for the enterprise needs.

To sum up, ERP helps CSIR in over coming duplicity of work and overcoming the problems of communication across the Sections/Divisions and helps in removing silos. It results into increased efficiency and better economy.

ERP has a modular architecture. CSIR ERP is profile based and form based and the data generated through transaction is available as information to the decision maker to make them capable to extract the critical business and operations information in a format for easy interpretation and decision making. (Source – CSIR Literature).

Once implemented ERP should generate business intelligence and predictive analysis. The execution of ERP provides the decision makers important insights into mapping efficiency of the employees and devising suitable performance based incentives. It also points out the skill gaps of specific individuals and groups and helps the HR in devising training and re-training packages as per the needs of the organization.

Implementation of ERP consisted of two related processes:–

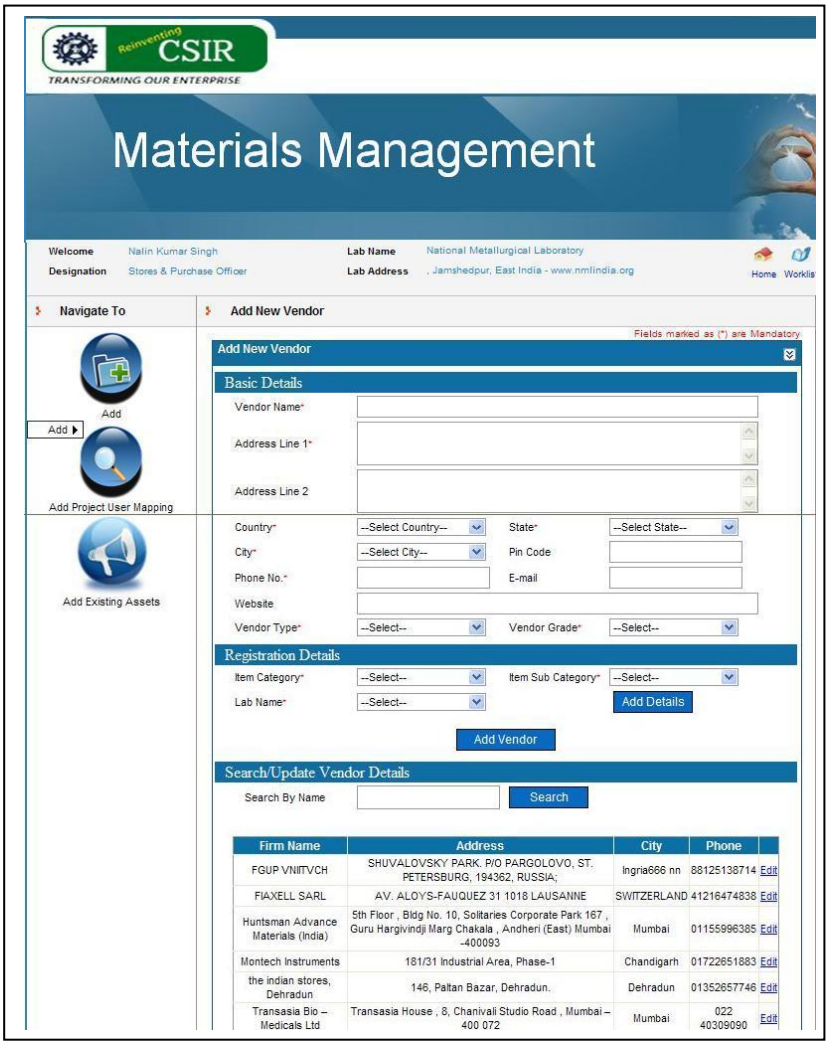
a. Putting existing business rule on an electronic process flow.
b. Re-engineer some of the business rules and work processes.

CSIR ERP MM module

In the vendor management system as configured in CSIR ERP System there is the provision to provide the vendor details as required at one place itself. The item category or the Item sub category, bank details for which the vendor is being registered is also shown. Based on the registration of vendors they would be provided a User ID and a Password. While floating a limited tender the names of firms are automatically displayed. We can select the required vendors and float tender enquiries within the shortest possible time. During the time of preparation of the comparative statement as well the data as provided for vendors is automatically displayed thus making the whole process very convenient. Similarly, while generating the reports in the cash purchase register or the daily Receipt register the data about vendors is automatically shown. In the e-tendering format the vendors can upload their quotations themselves. Similarly with the bank details of the vendors being available payment to the vendors can easily be transacted in e-mode.

The execution of the ERP in MM Module would help in reducing the inventory in all labs and would ensure that the organization can be run with Just in Time (JIT) inventory

It eliminates the need for creating safety and buffer stocks to meet exigencies and unpredictable situations. The ERP of two organizations if linked together can create value for the organization For example when all the labs are integrated through one ERP System then supplier margins would be easily decipherable by going into the statistics. For example if Dhanbad purchases an item at Rs 63 and Durgapur buys it at 53 and the same product and quality is purchased by Kolkata at Rs 43 then effective sourcing can easily be decided and value can be created for the organization.



Let us also look into how data analytics make use of association analysis to provide meaningful patterns. Below are some of the examples in which association analysis can make meaningful sense by co-relating multi dimensional data sources.

Suppose as the SPO I want to know which items are frequently purchased together (i.e within the same transaction). Suppose we want to find the association between the purchase of a computer and the purchase of Microsoft office together. Association analysis is better described with the help of confidence and support.

Suppose the data is interpreted as –

Buys (PL, ”Computer”) » buys (PL, “Microsoft office”) support = 2%, confidence = 60% where PL represents the project leader. A confidence or certainty of 60% means that if a project leader buys a computer then there is a 60% chance that the project leader will buy a Microsoft office software as well. A 2% support means that 2% of all transactions under analysis shows that computer and Microsoft Office has been purchased together. This association rule involves a single attribute that is (i.e, buys) that repeats. Thus it is called a single dimension association rule.

Suppose we have a relational database related to purchase again. That is age(PL, “30-40) ^ income (PL, “Rs 50000 to Rs 60000”) » buys(PL, “laptop”) (support =3% , confidence =70%). The data indicates that 3% of Project leaders between the age of 30 to 40 years and having an income of Rs.50000 to Rs 60000 have purchased a laptop. This association involving more than one attribute or predicate (i.e, age, income and buys) is an example of multidimensional association rule. Those data sets that do not satisfy the minimum support threshold and a minimum confidence threshold are discarded as uninteresting .

The Features in MM Module has elaborate features for Vendor registration which has made it possible to integrate the old vendor list with the newly enlisted vendor list giving rise to a more comprehensive data base. The data base for vendors is accessible to the entire stores& purchase section, to the users across divisions, to the laboratory as a whole and on an inter-laboratory basis for entire CSIR. Suitable data analytics if embedded in the vendor management module can provide the most efficient vendor in terms of delivery, quality, cost at the laboratory level and inter-laboratory level.

The organization can generate and track key performance indicators and in the case of CSIR the following information like –

i) Receivables Overdue (%)
ii) Order to ship cycle time(days)
Iii) Shipment to installation and commissioning time.
iv) Operational problems of the Equipment during warranty.
v) Most repaired equipment’s during a service contract.
vi) Equipment without service contracts but having greater frequency of repair and down time.
vii) In customer service rather than focus on first in and first out the business analytics should tell the most important user complaints needing prior attention depending on the exigency of the project.

Generally the requirement from different users within CSIR can be categorized into two heads – one of a customized nature and the other a more repetitive kind of demand. The customized requirements can be met from customized, tailored sources and the repetitive ones can be met by the most efficient vendors as per the list. The ERP MM Module should be able to map the users, the type of materials required by them and the database of vendors who can meet the demand successfully.

In CSIR we know that preventive maintenance and co-ordination of repair can ensure asset productivity and business analytics can help us in moving away from fixed maintenance charges to a form of predictive data which will be able to tell us about the machines that can run longer before servicing and which ones are showing signs that they should be serviced early.

By having consistent and reliable data about Order to Ship Cycle time we come to know the time that vendors from abroad would take to service an order or what time the Indian vendor would take to execute an Asset item or a consumable item.

Similarly by analyzing shipment to installation and commissioning time we would be able to know which vendors are the most efficient ones and the study can be analyzed for vendors across countries.

ERP would also help in minimizing the cost for maintaining inventory. The ideal concept would be to maintain a "Just in Time Inventory" to reduce the capital locked in idle inventory. Stocks can be held in fewer locations without increasing costs. Precise analytical models will be able to forecast demand which will help in focusing on inventory where it is required.

Broadly we can say that in Materials Management Module the data analytics and the applications required can be broadly categorized as :

i) Data which tells us about the spending patterns from projects and of individual projects leaders and other users within the projects.
ii) Analytics which tells about the performance of the supplier including complete supplier score cards, supplier price performance, delivery performance, product receipt quality, volume of business per supplier, payments due and overdue analysis.
iii) Analytics pertaining to the monitoring of procurement effectiveness - review of procurement process flow on a continuous basis so as to identify bottlenecks and take corrective actions. Thus it can be seen how with the use of information technology the supply chain and logistics functions can become more agile.

Current business or R&D environment is showing signs of stiff competition ahead and the pressure on supply chain to remain agile will continue on MM executives.

Decision making is fluid and dynamic part of any organization. Any decision making is contextual. Regulations change, policies change, markets change, competitors change and consumers change. Managing your decisions explicitly lets you respond to some of these changes but you also need to monitor your decisions for signs that hidden changes are undermining their effectiveness.

We can just say that in order to meet the uncertainty of the future, the ERP MM module would be a great tool in the evolution of CSIR as a dynamic R&D organization.

HR Module

HR module maintains a complete employee database.

i. Attendance.
ii. Payroll.
iii. Salary details.
iv. Performance evaluation.
v. Competency mapping.

All these information has been collated.

The ESS (Employee Self Service) module of HR makes the data available to the

USER→FUNCTIONARIES→APPROVING AUTHORITY→DECSION MAKERS AT CSIR

For each of these levels, a certain aspect or a certain part of the data is in view for their decision making. What we are trying to say is the data available to a certain individual in the chain is restricted and fixed. It can be used only for the purpose which is pre-defined. If someone wants to use this data for deriving patterns and develop intelligence, it is not available across the pyramid. The controls are highly centralized. Storing, access retrieval, visualization everything is controlled from the apex level. Naturally, if at all, it has any analytics imbedded in it, it is available only to decision makers at CSIR. Ideally, analytics should be available at every level in the hierarchy, sub-offices and head office for a flexible and decentralized decision making.

Having highlighted the huge benefits of an ERP system in CSIR, it is pertinent to critically appreciate that ERP covers only a part of the total activities of CSIR. It does not cover the whole gamut of activities. Any enterprise level system has to be based on uniformities or uniform patterns across the spectrum, the fact remains that besides these uniformities, there are activities which are unique, descriptive and analytical and fall outside the scope of ERP. Uniformity across the organization is desirable in some areas of activity but needless to say that uniqueness and heterogeneity among different units of the organization is desirable at times to promote healthy competition between different units of the same organization.

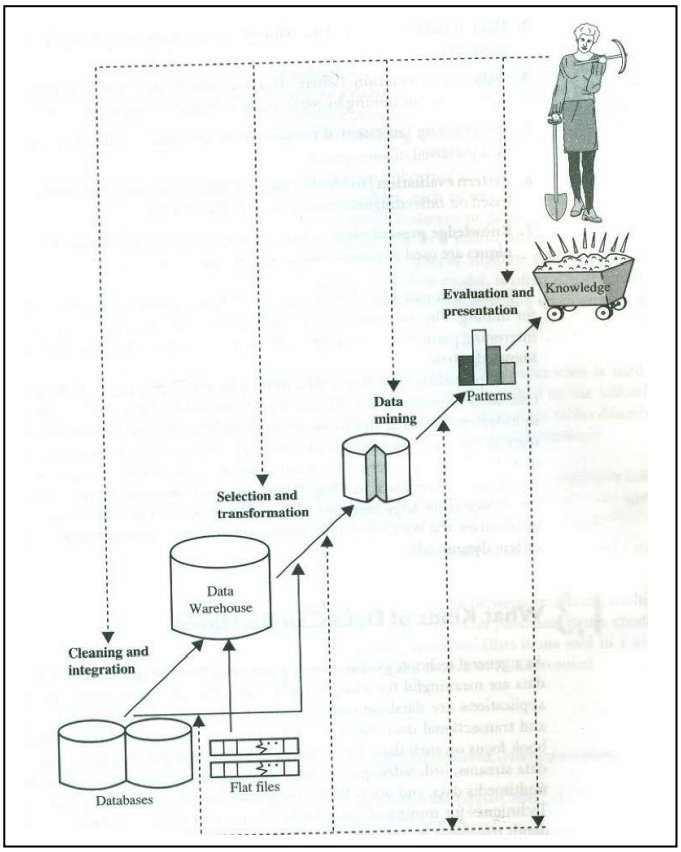
Even from the point of view of data, ERP makes use of and generates only a limited kind of data. To explain it further, ERP makes use of a database which is mostly a relational data base. The data ERP generates in turn is transactional data. This transactional data too, becomes a part of the relational data base. Again, the way this data is collected, classified, stored and retrieved creates many databases. These databases are structured data and their data models are known.

In addition to these we have data from www, data generated from the visits to our CSIR website and websites of all labs. Data from social websites and unstructured data from the various sources we have already mentioned. In a scientific organization like CSIR, the importance of none of these can be overlooked. For effective decision making, all these data type needs to be integrated and harnessed for providing useful analytics. In the absence of this, we may end into "a data rich but information poor situation" creating 'data tombs'.

This limitation of ERP can be solved by integration of data from other databases existing in CSIR and dynamic data from other sources. It can also solve the problem of data legacy from a previous system of data maintenance, either in physical or a digital format. This data may not be important from the point of workflow of the existing module but from the point of analytics, this may be integrated, mined and patterns can be deciphered for decision making.

Data Mining and Data Analytics

Invariably, in the lay man's language Data Mining and Data Analytics are used interchangeably. The two concepts have many things in common, yet these are different. Data Mining is an important step, an input to Data Analytics.



Data Analytics may require Data Mining; it may not require Data Mining at times. When data is present in clearly discernible Data Models, it requires some statistical analysis, some decision theory, a large input of behavioral sciences (OB and Physcology) and Mathematical Modeling. In the first case, it is a hypothesis driven analysis, in the second, it is a discovery driven analysis.

When data size is not too large, Data Mining may not be required for finding patterns which requires applying algorithms. At times patterns already visible may lead to decisions which require only a common sense. Patterns shown can lead to a good number of alternative decision paths. Although Data Analytics software's are available in the market and their literature tries to create the impression that all the problems related to data is a technological problem and only a technological solution to this problem is possible. It needs to be emphasized that that insight gained from data can be converted into good decisions only with a good subjective knowledge and it is where Managers matter.

Data Mining, instead of being one single concept or one method is a set of methods and according to the size and nature of data, the particular methods of mining varies. Data warehouses are very useful at times where databases are stored after cleaning and integration. Periodic loading and refreshing of data is done to keep it updated. These databases from data warehouses provide inherent support for OLAP (On Line Analytical Processing). OLAP uses background knowledge and allows the presentation of data for different levels of abstraction. It does so by different degrees of summarization. For example when data for different cities are available and we want to view it at the level of country, we have to roll up the data. In opposition to this when we have data available for the country and we want to see it for cities we have to drill down the data. Another example is, we have data available quarterly and we want to view it on a month's basis we have to drill down the data. Different SQLs can be applied for making sense of data from data warehouse. For

mining of these data from data warehouses, these steps of characterization, cleaning and integration are avoidable.

It is possible ERP generates some sequential data and a definite pattern can be seen. Now we may be interested to know the subsequent actions likely to take place based on this sequence. Suppose, we are interested in knowing whether such things happens in other organizations. In that case comparisons can be made between similar databases. If this data is available in a different format, data can be characterized, cleaned by removing noise and inconsistent data, integrated and mined.

Now look at another possibility, we are interested in seeing the pattern in data for a period of 20 years but the starting database of ERP is only 5 years old, the legacy data can be processed in the same way and it can be integrated before getting mined.

Suppose, from a transactional data base of ERP, we are interested in knowing some transactions using some dimensions with regard to one Scientist, Dr. Sanjay Kumar, ID E50G, the dimensions are – Patents, Papers, ECF generated, Technology transferred, Leave, Vigilance, Foreign visits.

Now, if we introduce one metrics of age i.e., 30 – 40 with regard to Principal Scientist, we can generate the same data for this age group using the same above mentioned dimensions. The results are available for use to the decision maker.

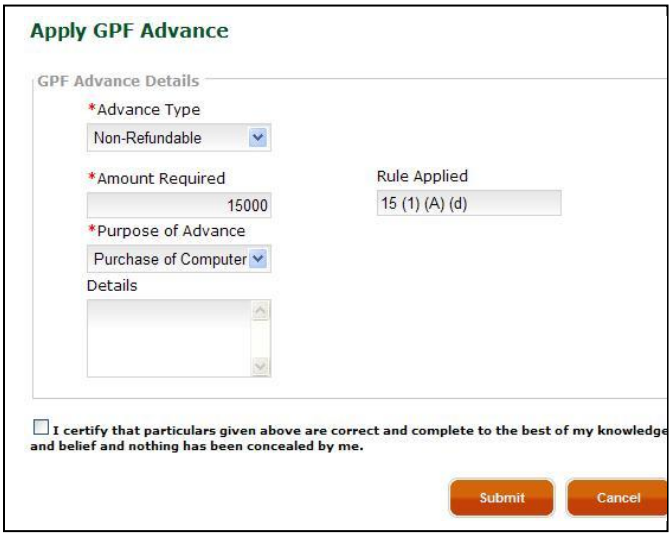
One more example of analytics in administration is developing business analytics for vigilance. A database of theft, financial misappropriation and fraud is collected for a 20 years period. We use the dimensions of Age, Designation, Pay Scale/Pay Band and Education of the accused, one interesting metrics of month of the incident/occurrence can be introduced. From the results generated, if we find that the maximum number of financial misappropriation where committed in the months of March and October before certain festivals, it provides a business intelligence for the decision maker who can use it for developing preventive mechanism for vigilance.

Analytics can be embedded to every stage of decision making. Suppose some persons visits NML website and spend 15 minutes time on technology developed and 10 minutes on patents. An Analytics Software can measure this for all the hits on the website.

Now, some person visits a Journal's archive to see papers on ore treatment or fly ash utilization. This electronic page of the Journal should provide a link to NML's website and the profiles of particular Scientist on different websites. This way the 'Referral Search Traffic' can increase manifold and a detailed geographical analysis of visitors both from 'Organic' and 'Referral Search Traffic' is possible. Some others write letters to NML Scientist and make queries regarding their technologies and patents. These three are three separate and distinct windows having no linkages. The result is communication gap and misuse of resources giving minimum output. Now if these three can be linked, alerts can be generated and data on past and probable customers can be shared across the spectrum.

Having discussed these possibilities in short, an effort has been made to develop a business analytics frame work using GPF data of CSIR laboratories.

General Provident Fund (GPF)



General Provident Fund is a mandatory deduction for all government employees joining before 01-01-2004. It is mandatory to deduct minimum 6% of the Basic pay of the employee and it can be maximum up to the basic pay of the employee except allowance. The quantum of deposits by the employees depends on how much surplus they have after meeting all necessary expenditures and fulfilling financial liabilities. Two kinds of advances are allowed from this fund, in case some employee wants to withdraw a part of this fund to meet some of their emergent financial needs–Refundable advance and Non Refundable advance. In more standard language, the first is known as Advance and the second as Withdrawal. Again there are reasons and objectives specified by GPF rules for which these advances/withdrawals can be made. We will try to see how GPF data for all CSIR laboratories and centers can give insight for making policies which go beyond GPF.

If the laboratories are named L1, L2...L41 we can tabulate deposits under GPF under columns of Gross Deposits, Average Deposits (i.e., gross deposits/employees subscribing to GPF).

1. A grade pay wise distribution of GPF deposits/deductions.
2. Grade pay wise distribution of Refundable advances.
3. Grade pay wise distribution of Non Refundable advances.

We can also create data bases for all laboratories on the reasons/objectives for which advances or withdrawals have been made e.g.,

a. Obligatory expenses.
b. Medical expenses.
c. Purchase of building house/Flats.

This data for all 40 laboratories can be seen along the dimensions of regions (North, South, East, West & Central) and City.

If one metrics of population is introduced, we come to know that

1. In cities with population above 50 lakhs, 60% of the withdrawals are for buying flats.

2. In cities below 50 lakhs population, withdrawals are more i.e., 40% compared to the average of big cities i.e., 15% for obligatory expenses.
3. In 9 Eastern laboratories, withdrawals and advances are more i.e., 30% for Medical expenditure against a national average of 8%.

Now, this data of GPF is giving us insights that

1. In bigger cities, employees are making withdrawals for creating assets whereas in smaller cities, employees are making withdrawals for social events like marriage and celebrations on child birth.
2. In Eastern laboratories, the fact that people are making withdrawals to meet medical exigencies describes how medical services are deficient and it results into an erosion of deposits and not into creation of permanent assets.
3. This report of data generated can be exploited in various ways and the knowledge which it generates can be used for making policies for creating health facilities in particular labs where it is needed. It is one small example of Data Analytics for decision support. If the data is voluminous, more unstructured, data mining tools can be applied before analytics is done.

Conclusion

The business intelligence within an ERP system has to be gradually evolved as per the decision support required by the organization. The gradual evolution of the ERP system would help in bringing efficiency within the R&D output of CSIR. From these simple illustrations of analysis, the organization can move to advanced predictive analysis. CSIR gradually can chart for itself a pioneering role in developing a unique research analytics in the form of Open Source Drug Discovery. Here, by opening an area of research to the world at large and by broadening and diversifying the collaborators' base at an extraordinary economy is a classic example of analytics. In a recent lecture "Indian S&T in the facebook era", Prof. Sameer Kumar Brahmachari speaks of leveraging the power of social websites like facebook in creating new platforms and connectivity for the education in the facebook era. Analytical ability and innovation-driven creativity is preferred to information-driven knowledge. It is possible by inculcating data analytics into the genes of the organization.

References :

1. Concepts in Enterprise Resource Planning by Ellen Monk and Bret Wagner [4th Edition, 2013, Course Technology, Cengage Learning 2013]
2. Data Mining Concept and Techniques by Jiawei Han, Micheline Kamber and Jian Pei [Published by Elsevier, 2011].
3. Harnessing the Growth Potential of Big Data by James Petter, Vice President and country manager EMC, UK & Ireland and Joe Peppard, Chair in Information Systems at Cranfield School of Management [].
4. Driving Operational Excellence with Business Analytics by James Taylor, CEO, IBM [2013 Decision Management Solutions].
5. Dynamics of Organizational Transformation and Executing Change by Dr. K. Jayakumar, Joint Secretary (Admin.), CSIR [1st Edition, 2011, CSIR]
6. Indian S&T in the Face book Era, Convocation address at Goa University, 27 January, 2012 by Prof. Samir Kumar Brahmachari, Director General, CSIR.

---

1. Store & Purchase Officer, CSIR-NML, spo@nmlindia.org
2. Section Officer (G), CSIR-NML, roushan@nmlindia.org