Managing the Content and Engaging the Nation in its Content Management: The WikiGyan Experience

Shalini R Urs

Keywords : Wiki, WikiGyan

1. Context

1.1 Information Friction

Most problems of our society - be it our failure to eliminate poverty, prevent terrorism, are essentially due to the 'information friction' phenomenon that exists. Information friction refers to the lack of systems that permit free-flowing information. Information friction happens when key facts, figures, data, records, content, best practices and process knowledge that are imperative for the successful functioning of a society, are wholly or mostly unavailable. While the information friction has been more or less eliminated to most of unstructured information on most topics, through the Wikipedia, and other information resources available on the Internet, the same is not true of 'data' relating to various kinds of societal data – whether it is the schools available in a particular area, a political candidate's crime records or assets or performance. Many organisations - Non Government Organisations (NGOs), Government agencies, academic researchers and others routinely collect enormous amounts of data pertaining to our society – socio economic, demographic, and governmental, health, and education. If only there were information systems and information management tools that were developed to harness these data, much of the existing information friction problems could be solved.

1.2 Content Management Technologies and Tools

A variety of Content Management tools are available today that can be very effectively deployed for managing the diverse kinds of data, mine them, analyse and provide business intelligence and insights. Such tools are being using by companies to build information systems that enable companies to use these data analytics and insights for business strategies and decision making. Text Mining, Data Mining and Business Intelligence are some of the key fields that are gaining grounds in these days of 'information' assets management in the business environments. Enterprise Content Management is a niche area that develops and deploys these tools for the purpose of enhancing the performance of an enterprise. In Politics and Governance as well, it is being used quite effectively as decision support systems. Policy makers need aggregated and analysed data for drafting government policies. One of the factors attributed to the success of US President Obama's campaign is his very strategic use of Information and Communication Technologies (ICTs) for managing data, for gaining insights and taking appropriate and immediate actions. The availability of plethora of open access content management tools makes developing content management tools and building information systems, fairly easy provided we have a well defined process model.

1.3. Information Sharing Environments

In the present world of Web 2.0 and Social Networking scenarios, Wikipedia and other such online resources and web services environments, information sharing for unstructured information seems to be very easy, rich and efficient. Anyone who has any information to share can do so easily. However, the key missing link in the current scenario is – we do not have a similar information sharing environment built for the structured data sets. Secondly, we also need to evolve effective process models that are scalable and cost effective.

2. Engaging the Nation

The Indian society can be viewed to consist of two segments – the developed and the emerging. The developed India (the privileged 10%) that includes among them the educated, professionals, technologists, bureaucrats, academicians, Government officials, and NGO workers who are literate, have access to these information sharing environments. While the other 90% constitutes the 'other' India, which may be referred to as the 'emerging India', which is not educated, and does not have access to these information sharing environments. Engaging the nation is the process of engaging the developed India in solving the problems of emerging India.

Over the past several decades, government organizations and Non-Governmental Organizations (NGOs) have addressed the plethora of India's physical, social, and economic challenges. They have been continuously implementing new programs to deliver better quality services to the citizens. They have an intimate understanding of the needs of the citizens, the objectives of their program and social issues that contribute to usability challenges.

Governments are keen on leveraging information and communication technologies to improve the tracking of service delivery, minimize corruption, and optimize service quality. Governments and bureaucrats demonstrate a strong understanding of the citizens, the required services, and technological options available to them. NGOs are keen on scaling their impact and realize that technology infusion is essential to achieving scale. As a result, these organizations have acquired a unique and detailed understanding of India's emerging class. Many of the NGOs perform cross-sector services, over time having grasped the powerful links between different problems. For example, one NGO might provide microfinance loans, while teaching computer classes, organizing cultural programs, and hosting health camps. This diversified approach is a diversion from the conventional business model, which encourages specialization.

This comprehensive understanding of emerging India's challenges naturally informs the NGOs' work, and today they have built a rich portfolio of effective development methodologies and solutions. Their professional nature, matched with a profound understanding of their constituents, makes NGOs one useful bridge between India's developed and emerging sectors. By strategically linking these two sectors, those in developed India can thoroughly understand emerging India's lives, aspirations, and consequential needs. This partnership will allow for the creation of solutions that will overcome the information friction that keeps the emerging class in relative poverty.

Paradoxically, the reason why NGOs have been able to impact India's emerging class is the same reason why they have been unable to actually *end* Indian

poverty; NGOs are so engrossed in emerging India's communities that they are subject to comparable levels of information friction. Therefore, solutions for emerging India will begin with breaking NGO's information barriers.

Academic institutions as part of their curriculum routinely incorporate and require class projects with the aim of effective learning of skills. But neither the reality is most of these projects remains just class projects and neither the student efforts are aggregated nor the students work in real life setting, thus defeating the purpose. Instead imagine a scenario where the entire student community is involved and engaged in working on solving nation's problems through their class projects, which are effectively part of and modules of a larger and broader well defined project.

Most companies and professionals working in them are engaged in some kind of corporate social responsibility related activities, which again tend to ad hoc, or disparate and/or not scalable. Imagine a scenario where in companies and/ or professionals working in the industry were to be engaged in solving nation's problems and building solutions towards larger projects, which are scalable and collaborative.

What is being proposed and conceptualised here is a vision of building information systems involving the different stakeholders – governments, NGOs, Academia and the industry through a process of well defined dialogue and engagement. A variant of the Open Source Development model – a Community Development Model which is directed under a broad framework of development model and a developer community. This 'engaging the nation' concept has been piloted through an experimental project called WikiGyan and the same is explicated in the following sections.

3. WikiGyan Project

3.1. What is WikiGyan?

It is a project to build an information system to empower every citizen with information and insight to build a civil society, an effective democracy and eGovernance. It is a system (of data and software tools and solutions) built by a student developer community at the International School of Information Management(ISiM), University of Mysore; mentored and monitored by a volunteer community of industry leaders (from Google, Yahoo and others), in response to and engagement with an NGO – the Deshapande Foundation. Its aim is develop a system and a process to solve the information problems of the nation. It is a live and living project – enabling the student community to learn, build and contribute towards this project.

WikiGyan aims to enable people/organizations to share data on the society (elections, health, education…) and gain insights through data analytics and intelligence based on the tools developed and integrated into the system. For simplicity sake we would like call this a combo of a YouTube for spreadsheets and Wikipedia for structured information. A system that lets people to upload spreadsheets, down load them, build databases, Data warehousing through the process of ETL (extract, transform and load) and visualise data in different ways.

Believing that Information friction (right person not having the right information at the right time) is the single most barriers towards awareness and right decision making for personal and societal advancement – be it voting, security, health, education, progress and development. WikiGyan will provide the data, information and insights to everyone – whoever needs and wants.

3.2. WikiGyan Process Model: Indians solving Indian information problems! It is a project by the people for the people – expanding and extending the open source movement to open systems movement.

The vast majority of the student community across the nation, continuously build, refine and fine-tune the system and enable NGOs, Governments, and researchers to upload their data. WikiGyan will empower the citizens to download the data, analyse, and visualise the data and gain insights leading to eGovernance, eHealth, and such systems. At the heart of this model is the idea of stakeholders coming together and being engaged in building information systems. The focus is especially on harnessing the power of students, and the voluntary involvement of industry and NGOs and others.

3.3. The Core Guiding Principles of WikiGyan model are –

♦ Open & Democratic System of development (driven by passion and built by capability and competence)

♦ Technical Excellence – to build a best of the breed information system

♦ Involvement and engagement – Compelled by passion and driven through capability and competence, people get involved

♦ Reach for the stars (High level framework) blended with low hanging fruits first approach – find it , fix it, fine tune it model

♦ Unleashing the students' potential and passion – bootstrapping through student project work and scaling greater heights of quality & standards through industry and NGOs /user groups engagement

3.4 The evolution of the process

The process model evolved for this idea (of engaging the nation in its service) involves and integrates two elements – education and engagement.

♦ **Education**

o Integrating the project work with their curriculum and academic programme

o Tying the different courses into a over all project thereby providing better understanding of how each course fits into the over all academic program and its goals of an academic program in information systems and management.

o We evolved a True blend of all the good learning models –

■ **Active Learning -** learners interacting with an environment; manipulating the objects and observing the effects of their manipulations; students explore and construct their own experiences.

■ **Authentic learning** – solving real world problems .Tasks connect the learners to the world outside them

■ **Constructive Learning:** Learners constructing their understanding and building their knowledge requiring that the learners find problems to be solved, explore multiple solutions to accepted problems, exemplify errors to clarify and refine knowledge

■ **Cooperative Learning:** Students work in groups with specific tasks and roles assigned to each one of them, shift roles and

responsibilities and communicate within and outside their team - working towards achieving targets, and reaching goals

♦ **Engagement :**

o Engaging the students to contribute – all the students of an institution (in this particular case ISiM) are involved

o Engaging the industry mentors – many thought leaders, project managers, product managers, team leaders from the information industry contributing to the project through their mentorship and participation

o Engaging the community – NGOs and other community to share their informational problems and providing insights

3.5 What is special and unique about this model?

♦ **Scalability** (hundreds of thousands of student community participating as developers, testers and QA)

♦ **Impact is transformational** –

o Transforming pedagogy - Bringing out the best in the students potential. Students Project work becomes more meaningful (through out the nation, hundreds of thousands of students carry out projects which to not see the light of the day nor do bring any value in real terms, where as in the WikiGyan model – every student work adds value to the over all system and enriches the student's learning and the system functionality and performance

o Transforming and achieving Academia– industry engagement. Visionaries, Product managers, program managers, coders and other professionals from the industry adding

to, monitoring, mentoring and ensuring quality and delivery of the product

o Ground Truthing the system by engaging the end users (NGOs, and others)

♦ **Societal** – a new model of nation building and Social Service – SST (Social Service through Technology)

♦ **Engagement** - Engages Industry , academia and society

♦ **Reality check** – problem definition by the community and reality checking by the community

4. ISiM –WikiGyan Case study

ISiM an innovative   institute offering a multidisciplinary masters program in information systems management took up the challenge of pioneering this new model of academic program and software product delivery for the societal information management.

How did we go about this?

4.1 Problem Statement

a. A real 'information friction' problem confronted by an NGO (Deshapande Foundation and its associates) was that different people, groups and organisations collect vast amounts of data in the form of spreadsheets during the course of their work and these data are not shared, integrated and made available to those who may need them, whenever they need, where ever they are.

b. They do not have the resources to develop and use tools and systems to share their datasets, deploy system integration tools and gain or provide information and insights.

c. End users cannot discover nuggets of information , patterns and insights that are hiding behind those datasets.

4.2 Project Design and Methodology

When the NGOs (led by Deshapande Foundation) approached some industry experts (Google, Yahoo and Infosys) with the above problem, during the initial briefings and meetings it became evident that this problem demands a new approach and model.

Over all Model and Approach

This particular model and process evolved over couple of weeks – with the industry volunteers and ISiM sitting down and conceptualising the model – an open source and open content and open system model; a students lead development model integrated with the curricular framework; integrating an inbuilt monitoring and mentoring system; living project following the agile and scrum approach

Specifics of the approach

♦ The ground rules and core guiding principles were laid out

♦ The ISiM curriculum was massaged and tweaked to integrate the project work into each and every course and over all mapping of project requirements with course outcome expectations was carried out

♦ Students buy in was ensured by building the academic credits and value in terms of learning outcomes in all aspects – competency building and technical skill sets ; product development and program management in a real setting ; contribution to society

♦ The overarching broad vision of building an Information systems (a YouTube and

Wikipedia combo for structured data) was broken down into four modules

♦ Chunking the project into four distinct modules –

   ■ Module 1 - upload and download module;
   ■ Module 2 - metadata and resource discovery ;
   ■ Module 3 - data integration ;
   ■ Module 4 - data visualisation

♦ Formation of project teams : Six teams were formed with one team each for the first two models and other two modules having 2 teams each

♦ Students were encouraged to select the modules and form their own teams based on their volition and interest (Empowerment and self determination)

♦ Students empowered to find out the problems and as well as the solutions. Go and try out – make mistakes fast and fix them along the way (enabling solution exploration and selection).

♦ Each team having project managers and product managers and roles defined and responsibilities assigned(honing Project management skills)

♦ The over all project specification outlined broadly by the industry mentors (interaction with industry).

♦ Detailed Product specifications developed by the students based on their understanding of the problem and solutions available.

♦ Weekly (or need based) Project review meetings, where in each of the teams presented

their work and achievements and challenges to the visiting industry mentors and based on the inputs and guidance, the road map for the future developed. Minutes of the meeting compiled and feedback tracking evolved.

♦ Tools and platforms selection : Students were encouraged to explore different software solutions, make calls regarding choice of software tools

♦ Communication and sharing – creating a Google group for the project; team meetings twice a week and sharing their thoughts and queries over the Google groups

♦ Reports and Documentation :

  • Weekly reports by each team and consolidated reports by Meta Project Manager.

  • Setting up the project site on Google sites and Source Forge net site to share software, documents and information

  • Each group documenting the technical specifications and progress and milestones accomplished

4.3 Technology Platform

Different technology options and tools study was carried out before finalising on XAMPP including PHP, Linux, MySQL, Apache, Media Wiki, Pentaho kettle, Weka and others such as Bugzilla and TRAC for tracking bugs and project management.

♦ **Key technologies and tools:**

  • Xampp for Linux 1.6: XAMPP is great because it takes only a few commands to get a fully integrated LAMP system up and running.

Installing the Components separately has typically been quite difficult and also a problem with having a number of separate open source components in its integration like Apache server, PHP, MySql.

  • Mediawiki 1.13.1(web based wiki software): MediaWiki offers a lot of features, including an optional file upload feature, a very comprehensive mark-up, very good internationalization support. MediaWiki is written in PHP and uses a MySQL database. Installation is incredibly simple. It is built to work in almost any Web - hosting environment where HTML can also be used.

  • Pentaho is a is Open Source application software for enterprise reporting, analysis, dashboard, data mining, workflow and ETL capabilities for Business Intelligence (BI) needs. It is easy to use and scalable and is being endorsed by the open source community. It has comprehensive capabilities include data integration, data mining and business intelligence and reporting. Kettle and WEKA are integrated parts of Pentaho

4.4 Over all Architecture and Technical Specification of the WikiGyan Project

WikiGyan consists of essentially four modules. The Module 1 is built on Media Wiki platform which enables people and organisations to create accounts and upload spreadsheets that they would like to share with others. A repository of spreadsheet is maintained by the system. The Module 2 essentially exploits the features and facilities of media wiki tools to build metadata and other resource discovery tools such as categories and tag clouds. Module 3

is a Pentaho Kettle based system that builds an aggregated database after the necessary data integration and transformations using a schema that is domain specific. Module 4 is a module that enables users to visualise data in different ways.

This module uses the data mining facility of WEKA which is an integrated part of Pentaho.

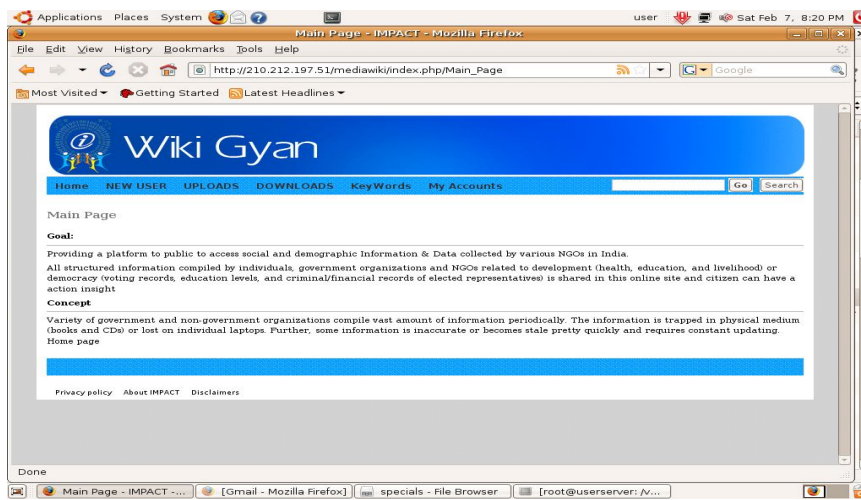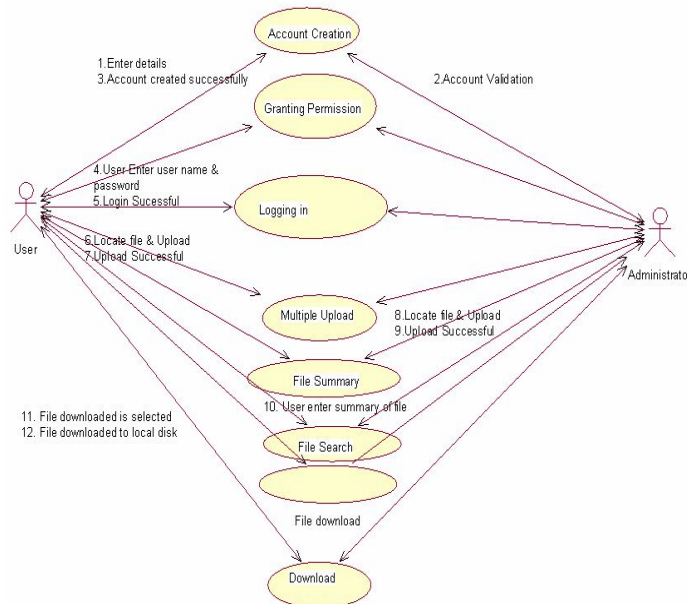The following figures * (1 to 4) provide an over view of the over all architecture and technical specification of the WikiGyan Project.

Figure 1: Screen shot of WikiGyan

Figure 2 : Use Case Diagram for uploading and authentication
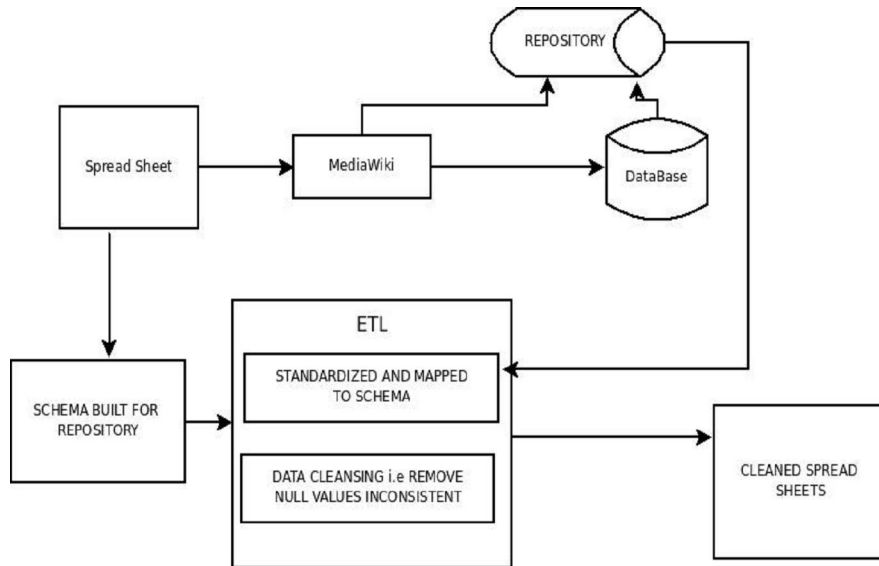
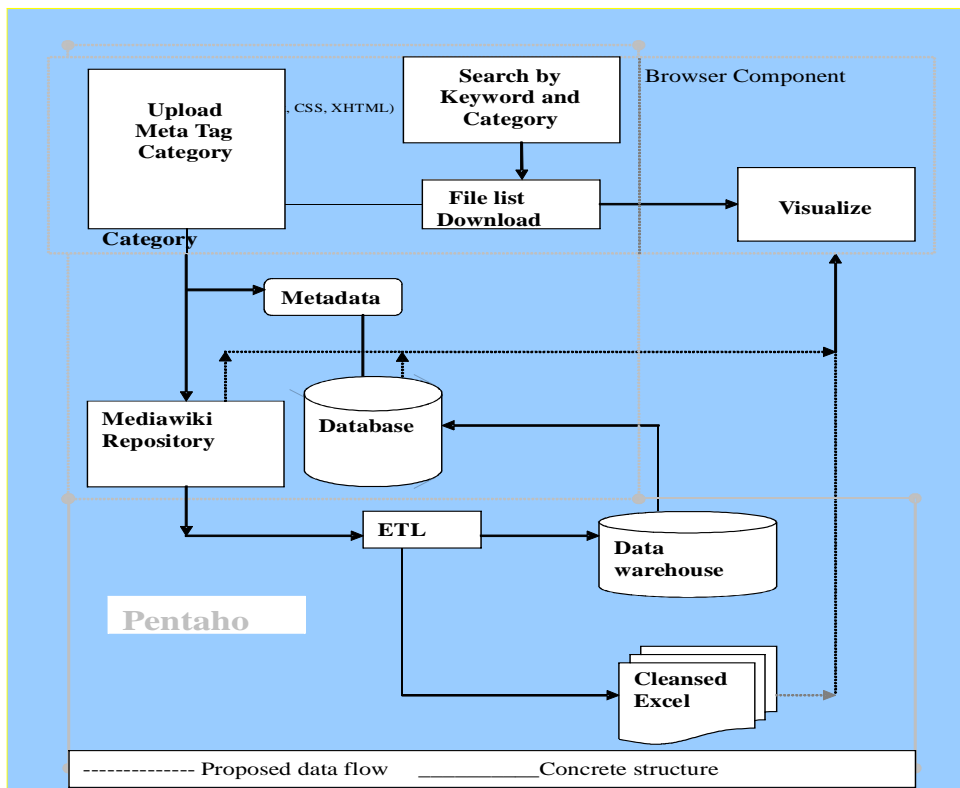**Figure 3 – Data Integration and ETL using Pentaho Kettle**

Figure 4 : Overall Technical Architecture of WikiGyan

5. Conclusion

The broad vision of WikiGyan is to initiate a movement to build information systems that would help solve the information friction scenarios obtaining in our country. In order to make it scalable and sustainable, a process model of open and community development model is also being experimented. A prototype solution has been built deploying Open Source tools such as XAMPP and Pentoho. This initiative is a model both for an open system product development process as well an innovative pedagogy model where in a combination of learning models such as active learning, authentic learning, constructive learning and cooperative learning are deployed through a living and live project. This initiative is also an exemplar for engagement with stakeholders – academia, industry and NGOs.

6. Acknowledgments

About Author

Ms. Shalini R Urs, Executive Director
International School of Information Management
University of Mysore.