Designing a website for TIFR: architecture and technology

Sourendu Gupta
(on behalf of the Web Development Team)

ASET Colloquium
TIFR
December 11, 2009

Web Development Team

Kausalya S.

S. Kyadari

A. Naik

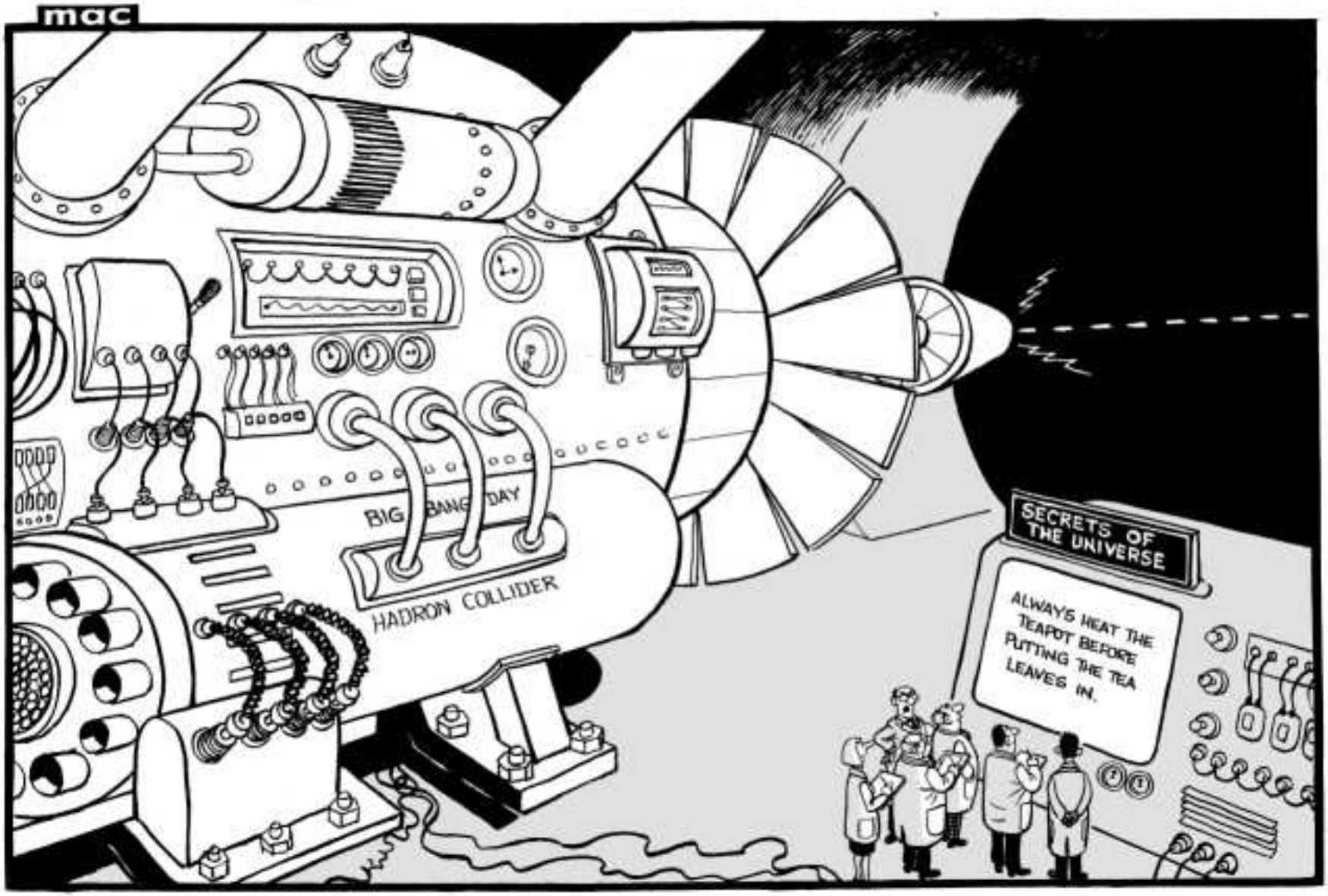
H. Raghavan

S. Raut

C. M. Vaidya

Outline

What is the purpose of a web site?



The first web site: circa 1994

- The whole web was amateur.
- The main browser was Mosaic. Lynx was still a major presence. Javascript was new.
- Yahoo! was hand made. Search engines were undependable.
- Ancient styling features included `<frame>` and `<blink>`.

TIFR web site had photos of some of the paintings in our collection. Main page had thumbnails, linking to high resolution images. Copyright issues rose after a few years.
TPY, Maths and TCS ran web servers and had their own sites. Individuals had home pages on one of these four servers.
Home pages were usually collections of links: bookmarks as a feature of browsers was yet to be invented.

The second web site: circa 2002

- Standards for corporate web sites emerged during the dot-com bubble
- HTML 2 and `<table>` was the standard graphics design paradigm
- `<body color>` and `<font color>` were the usual styling choices
- Javascript, PHP and MySQL were the main technology behind the scenes
- IE and Mozilla were the main browsers, Apache was the main server

TIFR web technical and graphic design was outsourced. Design based on MySQL data base manager to hold the page contents, PHP to extract and serve it up, Javascript and HTML 2 to control browser interaction. CC staff were trained in usage. User input was made through forms and a basic editor.

The second web site: circa 2002

- Standards for corporate web sites emerged during the dot-com bubble
- HTML 2 and `<table>` was the standard graphics design paradigm
- `<body color>` and `<font color>` were the usual styling choices
- Javascript, PHP and MySQL were the main technology behind the scenes
- IE and Mozilla were the main browsers, Apache was the main server

TIFR web technical and graphic design was outsourced. Design based on MySQL data base manager to hold the page contents, PHP to extract and serve it up, Javascript and HTML 2 to control browser interaction. CC staff were trained in usage. User input was made through forms and a basic editor.

User experience

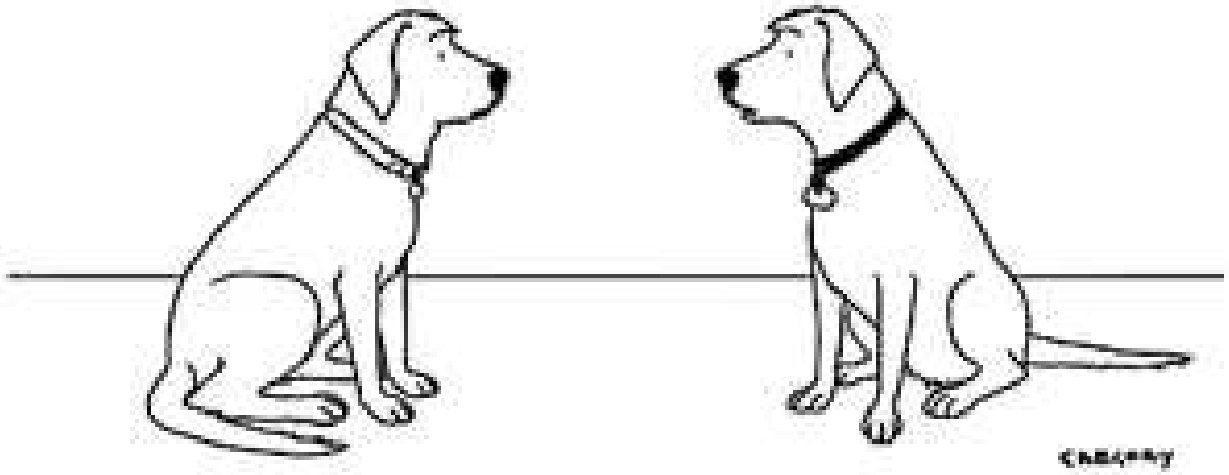
- Meeting of representatives of all departments, facilities and sections held in 2006 to get feedback on user experience.
- Many small features and changes requested.
- Staff unable to engineer these changes because of the complexity of the code and lack of documentation and training in code maintenance.
- Special needs of departments hard to incorporate.
- Work automation hard, since system resistant to re-engineering.

End of 2007, CC staff unable to port existing web site to new version of PHP. Meanwhile Datanet began to automate many administrative jobs. Decision made by web committee to re-design web site. Very useful design and technology inputs from committee members; especially G. Krishnamurthy, S. Saran, A. Thamizhavel.

Who is the web site for?

Web site has to be useful to TIFR users as well as the outside world. Automate everything that appears on notice boards: make the web the primary means of communication within TIFR.

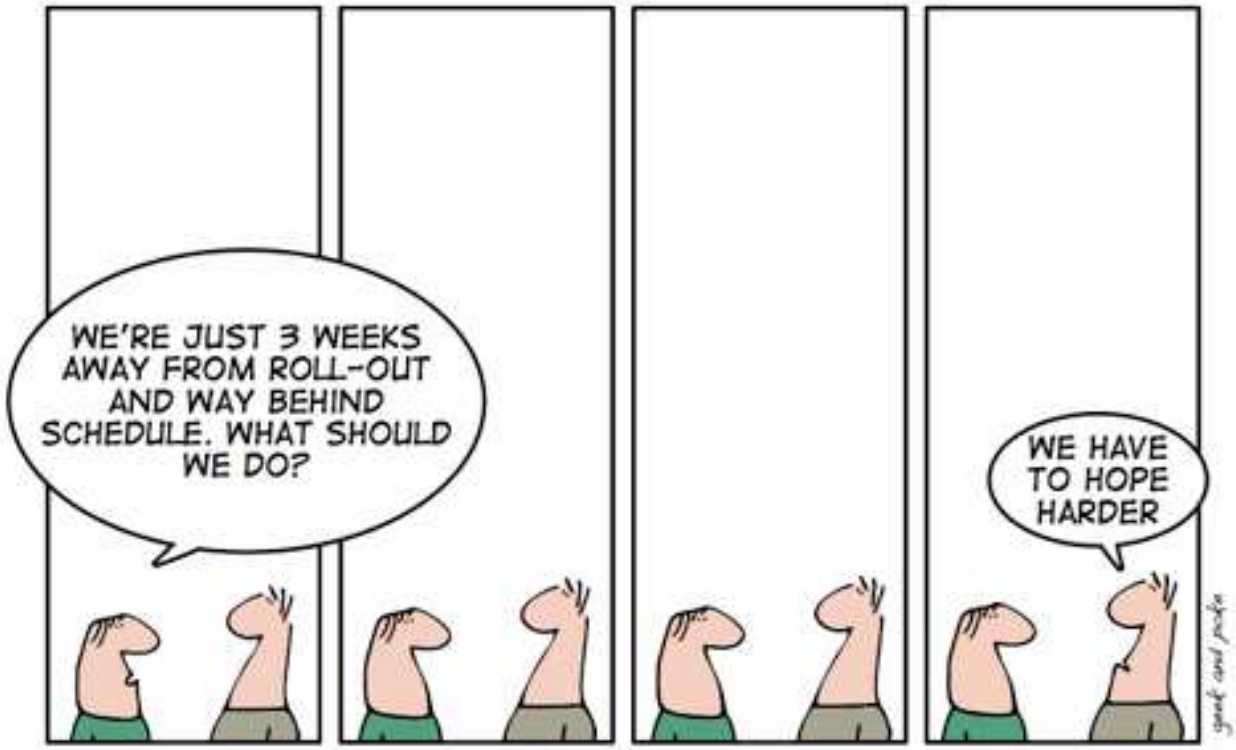
Content is king



"I had my own blog for a while, but I decided to go back to just pointless, incessant barking."

TIFR produces ideas; give the world a sense of that. Use original material; own copyright to everything we put out. Our main external target are students; tell them how they can come to us.

Build everything yourself



Design is an interactive process. Use standards and open source. Build everything else. Test the design: measure before you theorize. Track how people use the site.

The social web

All web activity is social engineering. Integrate the web into the work flow of TIFR. Look for niches where the web can simplify a job.

Outline

Design and implementation

First design principle

Uniformity of look, modularity of design.

Second design principle

Deliver content; deliver it fast and securely.

Third design principle

Automate jobs; make the web the easiest way to do things.

Design and implementation

First design principle

Uniformity of look, modularity of design.

Second design principle

Deliver content; deliver it fast and securely.

Third design principle

Automate jobs; make the web the easiest way to do things.

First technical decision

Learn new technology.

Second technical decision

Use only technology that is understood.

Visual design: form follows content

1. Each web page contains primary data, context and meta-data. Context and meta-data dictate the layout of the page.
2. A banner across the top of the page identifies it as part of the web site; this is the widest context.
3. A menu bar supplies the structure of the web site and the position of the page within the structure: this is the second level of context.
4. Usually there are quick links to a site map, a link for mail to the webmaster, *etc*. These are also part of the context.
5. Each page has a visible creation and last-modification date as well as an author and a copyright statement: this is metadata.
6. CSS is used for visual design (layout, colours, fonts) which integrates these elements. Changing visual design is simple with CSS.

Site design: core and hosted content

Core content

TIFR takes responsibility for content. Must have low volatility.

Content initially migrated from the old site. New content generated slowly. The most useful core content: PeopleFinder (formerly the phone book). Designed in house.

Site design: core and hosted content

Core content

TIFR takes responsibility for content. Must have low volatility.

Content initially migrated from the old site. New content generated slowly. The most useful core content: PeopleFinder (formerly the phone book). Designed in house.

Hosted content

User takes responsibility. May be volatile or nearly static.

Static includes personal home pages, departmental micro-sites, *etc*.

Volatile content most visible part of the TIFR web site: *e.g.*, notices, advertisements, replies to RTI queries, seminar and conference announcements, *etc*.

Safeguarding core content: the CMS

- CMS must be capable of version control, reversion and backup. Not as complicated as typical CMS used for blogs *etc*.
- Lightweight CMS can built from a collection of flat files using any scripting language. Chose `sh` because of familiarity.
  - Ascii text files in a unix directory hierarchy are input. Output is a website in plain HTML 4 styled with CSS.
  - Text files use subset of HTML, enforced by built in editor.
  - Metadata (author, copyright, *etc*) required.
  - Menu and site map generated automatically.
  - Backup and restoration performed by shell script.
- Collaborative work of Kausalya S., A. Naik and S. Kyadari. Further work to be done: modularization and configurability.

Safeguarding core content: the CMS

- CMS must be capable of version control, reversion and backup. Not as complicated as typical CMS used for blogs *etc*.
- Lightweight CMS can built from a collection of flat files using any scripting language. Chose `sh` because of familiarity.
  - Ascii text files in a unix directory hierarchy are input. Output is a website in plain HTML 4 styled with CSS.
  - Text files use subset of HTML, enforced by built in editor.
  - Metadata (author, copyright, *etc*) required.
  - Menu and site map generated automatically.
  - Backup and restoration performed by shell script.
- Collaborative work of Kausalya S., A. Naik and S. Kyadari. Further work to be done: modularization and configurability.

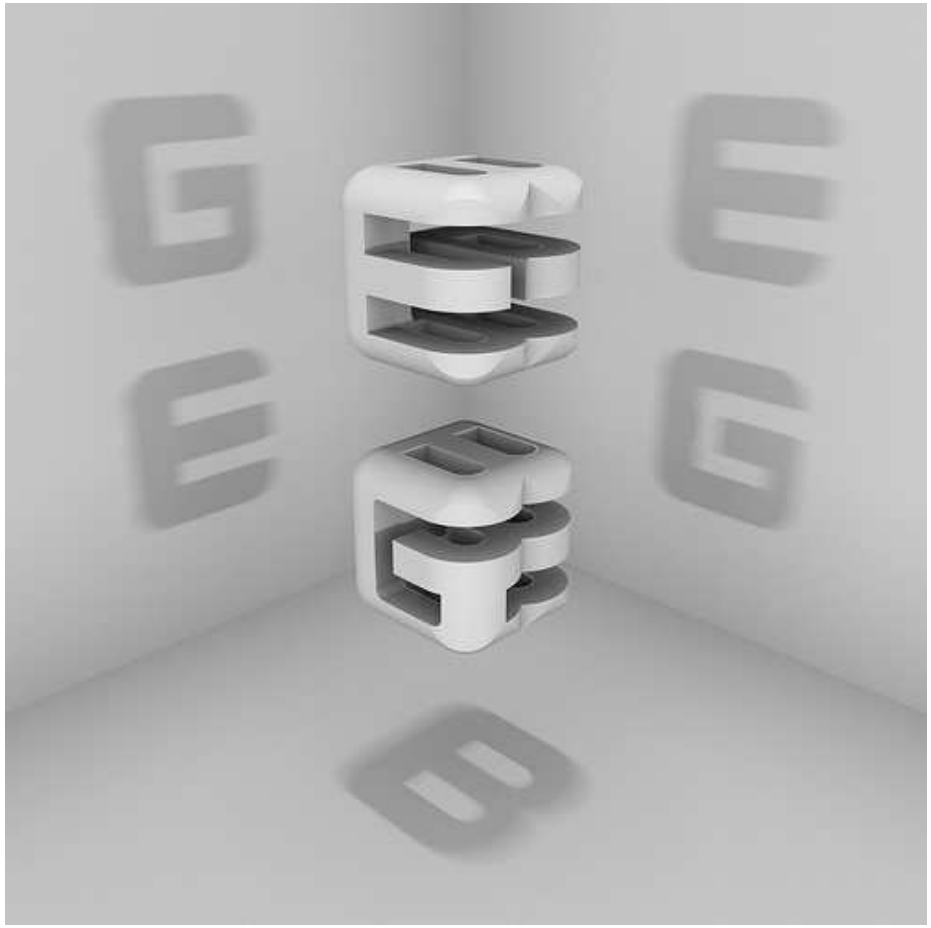
First Release

Initial version of current website released in December 2008. Designed to undergo continuous change.

PeopleFinder issues and development

- Old telephone directory was very useful but limited. Maintenance of records was hard: 50% of entries involved people who had left TIFR.
- No definitive database of people in TIFR. Datanet is close to complete but inaccessible to people outside of TIFR. Status of students is with university cell. Status of post-docs is with departments. Current home address, phone number, office room, not collected elsewhere. Setting up a definitive database is a major administrative effort.
- Technically a MySQL database with PHP front end. Data gathering and cross checking is major part of the effort.
- Data gathering help from G. Ajinkya, N. Goel, I. Murzello, N. Shirke, Bindu Jose. Maintenance by C. M. Vaidya.
- Database development by Kausalya S. Includes database integrity checks and ability to revert to previous versions.
- Standardizing operating procedures by Kausalya S., C. M. Vaidya and H. Raghavan.

PeopleFinder philosophy: Plato's cave



Records about a person lie in one place in one database. Single database, multiple views. Data validation becomes simple.

PeopleFinder views

- Record owner's view (YOU)
- View as telephone book: default.
- Generate default web page; user can over-ride.
- Create directories of Colaba and Mandala colony residents.
- Search by role in TIFR (*e.g.*, official mobile search).
- Search by research interest
- Administrator's view
- Visitor information?

PeopleFinder views

- Record owner's view (YOU)
- View as telephone book: default.
- Generate default web page; user can over-ride.
- Create directories of Colaba and Mandala colony residents.
- Search by role in TIFR (*e.g.*, official mobile search).
- Search by research interest
- Administrator's view
- Visitor information?

You can control your information

Each person is responsible for checking one's own entry: go to PeopleFinder, click on 2nd menu item ("Modify/Add entry") and search for your name. You get a complete view of all your data.

Electronic Notice Boards

- Tender notices and job advertisements have been put on the TIFR website since 2002. The Public Outreach Committee and the Subject Board for Physics have asked for the ability to put up volatile information at short notice. Awards to TIFR members and books published by them are also put up on the website by the PRO.
- The common idea behind these is the notion of an electronic notice board.
- Built by Kausalya S. and A. Naik. Further development by project students under their guidance. Software based on MySQL and PHP. Fully customizable, so a new notice board can be deployed very quickly.
- Now extracted into the main page of TIFR so that the titles of the notices are immediately visible. Reduces number of mails-to-all.
- Departments and sections are free to request notice boards by sending mail to `webmaster@tifr.res.in`. Example: canteen weekly menu is up on the notice board.

Calendar of Events

- Most important volatile information on the web till now is the list of formal scientific activity in the Institute: seminars, courses, conferences, public lectures, and meetings.
- The InDiCo (Integrated Digital Conferencing) tool allows management of complex conferences, workshops, and time-tables of meetings, to attach multimedia files to each event item and to store the resulting agendas in a multi-level hierarchical tree.
- Based on zope and python; built by a European consortium; available under GPL. Customized for TIFR by Kausalya S., A. Naik and a project student. Currently managed by S. Raut: eventmaster@tifr.res.in.
- Handles complete organization of a conference: announcement, registration, abstract and talk upload, printing of badges, proceedings. Can be used to send event reminders to a mailing list (customizable per event). Can be synchronized with desktop calendars. Can be used for managing room bookings.

Other uses of Indico

- SBP is announcing its courses on Indico. Complete course details and mailing lists for courses, day by day plan, lecture notes, problem sheets, *etc*, can be posted through Indico.

Other uses of Indico

- SBP is announcing its courses on Indico. Complete course details and mailing lists for courses, day by day plan, lecture notes, problem sheets, *etc*, can be posted through Indico.
- A recent meeting on high performance network for research is on Indico. This includes presentations on the National Knowledge Network and the transglobal link called GLORIAD-Taj. All talk pdf/ppt can be put on Indico.

Other uses of Indico

- SBP is announcing its courses on Indico. Complete course details and mailing lists for courses, day by day plan, lecture notes, problem sheets, *etc*, can be posted through Indico.
- A recent meeting on high performance network for research is on Indico. This includes presentations on the National Knowledge Network and the transglobal link called GLORIAD-Taj. All talk pdf/ppt can be put on Indico.
- Indico events can be made for either open viewing or restricted. It is possible to announce confidential meetings through Indico and let it handle details of reminders.

Other uses of Indico

- SBP is announcing its courses on Indico. Complete course details and mailing lists for courses, day by day plan, lecture notes, problem sheets, *etc*, can be posted through Indico.
- A recent meeting on high performance network for research is on Indico. This includes presentations on the National Knowledge Network and the transglobal link called GLORIAD-Taj. All talk pdf/ppt can be put on Indico.
- Indico events can be made for either open viewing or restricted. It is possible to announce confidential meetings through Indico and let it handle details of reminders.
- Facility booking can be made through Indico. Currently DBS uses a specially developed software for booking time on its confocal microscope. Indico can be adapted for such purposes.

Other uses of Indico

- SBP is announcing its courses on Indico. Complete course details and mailing lists for courses, day by day plan, lecture notes, problem sheets, *etc*, can be posted through Indico.
- A recent meeting on high performance network for research is on Indico. This includes presentations on the National Knowledge Network and the transglobal link called GLORIAD-Taj. All talk pdf/ppt can be put on Indico.
- Indico events can be made for either open viewing or restricted. It is possible to announce confidential meetings through Indico and let it handle details of reminders.
- Facility booking can be made through Indico. Currently DBS uses a specially developed software for booking time on its confocal microscope. Indico can be adapted for such purposes.
- Public lectures should be announced on the web: speaker and date as early as possible, title when available, poster/multimedia later.

Bandwidth and security

- Outgoing bandwidth is a small fraction of the incoming bandwidth. This is by design. We expect that TIFR users are a large fraction of all users of the website, so they will not show up in the external bandwidth. Also, many of our external users will be undergraduate students working over narrow links.
- Till now neglected segment is our peers in other institutions around the country. We need to put more material on the website directed at our peers. Let us know what you need
- Threats are of two kinds: denial of service and break in. Response to DoS attacks is not part of the website design. The barrier to break-in is physical security and password security to Indico and notice boards as well as accounts on the webserver. Recovery from break-ins is through regular backup schedules and CMS (original files of core site held on a separate server).

Outline

We are listening

USER1: The design of the new website is not good at al. Lot of us feel so.

USER2: Any good website will have the front page to give away the essence of the activities of the institute, this is at all not true with the one you have designed. Everyone that I talk to has the same feeling. I am sure TIFR deserves a much better web site than this.

USER3: This is terrible.

We are listening

USER1: `The design of the new website is not good at al. Lot of us feel so.`

USER2: `Any good website will have the front page to give away the essence of the activities of the institute, this is at all not true with the one you have designed. Everyone that I talk to has the same feeling. I am sure TIFR deserves a much better web site than this.`

USER3: `This is terrible.`

Hopeless despair is our only response to comments of this kind.

We need your help

USER: Search: It redirects to www.google.co.in. I am shocked. Even MIT uses the google search but it's done much more professionally.

DEVELOPER: ... you have offered to help out the web pages construction by building a custom search engine. Thanks for your offer; it is much appreciated.

USER: I would be happy to talk to you about the Custom Search Engine, although I am no expert myself. Unfortunately I am attending a conference at IISc, Bangalore and then would be taking a break. I would be there in January.

We need your help

USER: `Search: It redirects to www.google.co.in. I am shocked. Even MIT uses the google search but it's done much more professionally.`

DEVELOPER: `... you have offered to help out the web pages construction by building a custom search engine. Thanks for your offer; it is much appreciated.`

USER: `I would be happy to talk to you about the Custom Search Engine, although I am no expert myself. Unfortunately I am attending a conference at IISc, Bangalore and then would be taking a break. I would be there in January.`

Last mail from user on 10 December, 2008. Still waiting for next contact.

Debugging help was appreciated

USER: 1. under "Contact us" => "our Addresses" => email address of HEGRO, Pachmarhi "hegro2001@mantraonline.com" is outdated. Please use: hegro@tifr.res.in and hegro2001@sify.com
2. under "science" => "School of Natural sciences" => e. "Dept of High Energy physics"
3rd paragraph, 5th line: "These are studies at dedicated experimental facilities like GRAPES experiment at Udhagamandalam (Ooty) and PACT experiment at Pachmarhi"
==> please include the HAGAR experiment at Hanle (Ladakh).
"These are studies at dedicated experimental facilities like GRAPES experiment at Udhagamandalam (Ooty), PACT experiment at Pachmarhi and high altitude HAGAR experiment at Hanle (Ladakh)"

Debugging help was appreciated

USER: 1. under "Contact us" => "our Addresses" => email address of HEGRO, Pachmarhi "hegro2001@mantraonline.com" is outdated. Please use: hegro@tifr.res.in and hegro2001@sify.com
2. under "science" => "School of Natural sciences" => e. "Dept of High Energy physics"
3rd paragraph, 5th line: "These are studies at dedicated experimental facilities like GRAPES experiment at Udhagamandalam (Ooty) and PACT experiment at Pachmarhi"
==> please include the HAGAR experiment at Hanle (Ladakh).
"These are studies at dedicated experimental facilities like GRAPES experiment at Udhagamandalam (Ooty), PACT experiment at Pachmarhi and high altitude HAGAR experiment at Hanle (Ladakh)"

Appreciated the help and modified as requested.

Your suggestions influence design

USER: The site does not work when typing in the name of an institute member. However, it does give the correct details when the room number of the extension number is typed.

Please note: I would not like to have my residence phone number or any other residence details appearing on the web-pages.

Your suggestions influence design

USER: `The site does not work when typing in the name of an institute member. However, it does give the correct details when the room number of the extension number is typed.`

`Please note: I would not like to have my residence phone number or any other residence details appearing on the web-pages.`

We understood that many people had the same reservations. We accommodated this in the PeopleFinder. Now the residence address appears only on restricted pages, which cannot be viewed from outside the TIFR domain.

"What I need is a list of specific unknown problems we will encounter."