

# PREPARATION OF GEOSPATIAL DATA INFRASTRUCTURE IN GEOLOGICAL SURVEY OF INDIA: A REWARDING EXPERIENCE WITH BRGM OF FRANCE

G.S. Srivastava<sup>a\*</sup>, A.K. Malaviya<sup>b</sup>, Rajeev Srivastava<sup>b</sup> and A.K. Kulshrestha<sup>a</sup>

<sup>a</sup>Department of Geology, University of Lucknow, Lucknow- 226 007, India; gss\_geodata@yahoo.co.in

<sup>b</sup>Geodata Division, NR, Geological Survey of India, Lucknow- 226024, India; akmalaviya@yahoo.com

## Technical Commission IV, WG - IV/1

**KEY WORDS:** Oracle RDBMS, data domains, Geo-scientific database, bibliographic database, multi-criteria analysis, mineral targeting, softcopy conversion.

### ABSTRACT:

A large amount of earth science data, collected by Geological Survey of India since its inception in 1851, has been brought to electronic media for storage, quick access, retrieval and analysis in GIS environment, through collaboration with the BRGM of France, starting in 1996. In the initial stage, 10 data domains were identified out of which five, viz. Map50,000, Geochemical Exploration, Mineral Exploration, Drilling and Coal were developed and implemented for preparing geospatial database on Oracle RDBMS platform. Subsequently, data pertaining to the remaining domains viz., Geophysics, Photo-geology & Remote Sensing, Rock Analysis, Natural Hazards, Glaciology, and Environment were developed and implemented. Further domains of Map250,000 and Palaeontology are under process of development. Unpublished geological reports, numbering over 30,000, containing text and graphic data, were converted into soft copy and put in a separate bibliographic database, querriable through 17 indexed fields. Multivariate analysis pertaining to Tons Valley base metal prospect was taken up. The data included topographic contours, geochemical, geophysical, landuse/ geology and lineaments. Geochemical data were available for copper, lead, zinc, nickel, cobalt and silver. Of these, geochemical data for copper only were utilized as one layer as it was showing anomalous distribution pattern. Geophysical data utilized were Chargeability (induced polarisation), Self-potential and Resistivity. The weightages for different themes were given after discussions with the mineral exploration team, working in the area. Different map layers were prepared with SYNCOMPO and superimposed in GIS environment. The results were draped on the Digital Elevation Model of the block. Two significant anomalies were obtained; one coincided with the known zone of mineralization while the second anomaly was recommended for drilling, which indicated mineralised zone in depth. The study proved efficacy of the multivariate analysis for mineral targeting using data layers in GIS environment.

## 1. INTRODUCTION

In present era of information technology, storage, retrieval and dissemination of geological data through electronic media is prime requirement of an organisation like GSI, which has vast amount of geological data in terms of graphical and textual forms collected over a century and a half. Besides archival of such data for its long time preservation, the same could be processed using advanced software for deducing logical and useful information. The dissemination of information in different formats including hard copy outputs can be done speedily through computers.

### 2. GSI - BRGM COLLABORATIVE PROJECT

The ideas of using RDBMS for geological data were evolving in the mid nineties and a lot of data both in textural and graphic forms, were entered on dBase application developed on Disk Operating System (DOS). However the process of developing Database finally got crystallised through GSI-BRGM collaborative project on 'Setting-up of Geoscientific Data Centres in GSI, which acted as a catalyst (Srivastava *et al.*, 1998).

This project was implemented through seven sub-project viz., 1. Bibliographic Database, 2. Geo-scientific Database, 3. Digital map making facility, 4. Extension of Remote Sensing facilities, 5. Multiprocessing for mineral targeting, 6. Centre for study of ore deposit and metallogeny (CSOM) and 7. Computer \_\_\_\_\_

\* Corresponding author

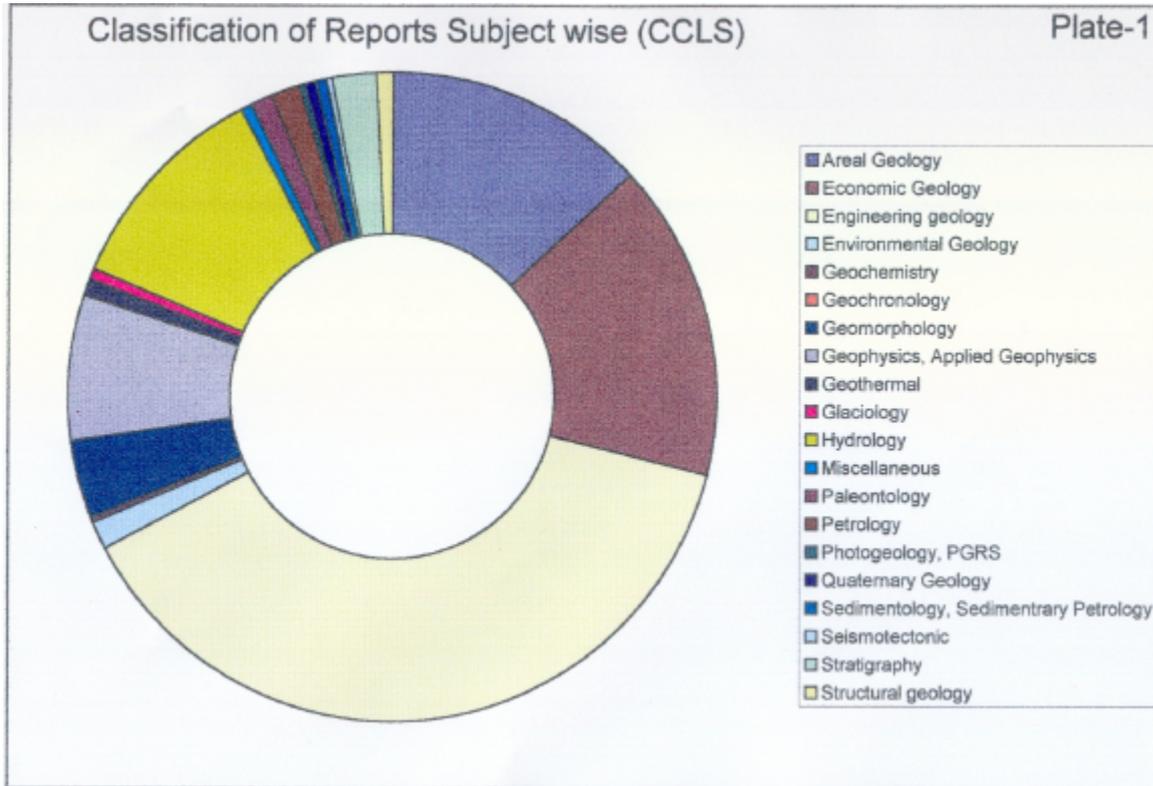
systems.

### 2.1 Bibliographic Database

The structure of bibliographic database was developed by GSI and BRGM Experts on an application TEXTO Software developed from a well known bibliographic software CHEMDATA. Now this has been transferred into RDBMS table formats. It was decided that all the published data of GSI will be entered in the database by CHQ while the regions will enter unpublished reports pertaining to their region. The Data from the Regions will be sent to CHQ for validation and incorporating the same in the National Geoscientific Data Centre (NGDC) database. The validated data from the Regions and that of CHQ will be sent to the Geoscientific Data Centres (GDCs) of the Regions so that similar database is available both at CHQ and Regional levels.

The bibliographic Database is incorporated under applications like PUBLIT (Published reports) UNREP (Unpublished reports) OREPORTS (Reports of other Regions) and NRCARDS (Designed for catalogue purposes in Northern Region), etc. For example, for the Pilot Project area of Northern Region, UNREP contained data from 521 reports, OREPORTS had data from 222 reports and NRCARDS contained data from all 743 reports. Pictorial representation of

the statistical analysis of a few fields viz., type of report (TODC), classification of report (CONF), Field Season (FSPR), date of publication (DPUB) and subject classification (CCLS) can also be made (Plate-1).



## 2.2 Geo-scientific Databases

Under this sub-project ten database domains (themes) were identified for the geological data of GSI. Out of these five domains, namely geological map on 1:50,000 scale, Geochemical Exploration, Mineral Exploration/Mining, Drilling and Coal (only for Coal Wing of GSI) were selected for designing of database structure. The remaining domains, viz., Geophysics, Photo-geology & Remote Sensing, Rock Analysis, Natural Hazards, Glaciology, and Environment were developed and implemented on the similar lines, subsequently after gaining experience from GSI-BRGM project. Further domains of Map250,000 and Palaeontology are under process of development. Other domains like Geothermics, Engineering Geology, etc. are planned for development and implementation in future.

After detailed discussions with the experts of each of the domains, lexicons and directories were prepared, validated and input in the system. Subsequently, transcripts from the reports pertaining to the Pilot Project Area were prepared and loaded in the system. As an example, the details of the work carried out in each domain for the Pilot Project area in Northern Region are given below:

### 2.2.1 Theme: Geological Mapping on 1:50,000

The work carried out in this theme can be divided into two subgroups:-

- a. Spatial data input was done in 'ARC/Info' using both Workstation and PC versions, and

- b. Thematic data input from the transcripts onto screen form, specially designed for this purpose, available on Oracle supported database.

The Methodology of the work carried out is as follows:

In order to create the spatial database, the geological maps were digitised manually by using PC/ArcInfo. Out of 17 themes (layers), 7 themes were present in the study area. Hence, the geometric data capture was made for these 7 layers only. Two additional layers of rivers and roads were also prepared to improve the output quality of the maps. The geometry data was then exported to UNIX ArcInfo on Sunsparc-20 workstation for on-line editing, building up of topology and creation and data encoding of Geometry\_id for linking with the Oracle database.

### Spatial Database

As per the database design, geometry capture to create spatial database, has a provision for the themes in the Oracle Database. These include 1. Lithological Units (LI)\*, 2. Boundaries (BO)\*, 3. Dykes (DY), 4. Zonal Structures (ZS), 5. Faults (FA)\*, 6. Regional Folds (FO)\*, 7. Shear Zones (SZ), 8. Trend Lines (TL)\*, 9. Non-oriented Point Structural Elements (NO), 10. Oriented Point Structural Elements (OS)\*, 11. Fossil Occurrences (OC)\*, 12. Mineralisation (MI), 13. Mines and Quarries (MQ), 14. Archaeological Sites (AS), 15. Other Point Elements (OP), 16. Palaeochannels (PA), 17. Linear Geomorphological Elements (GE), 18. Point Geomorphic Elements (GP), 19. Polygonal Geomorphological Elements (GG), 20. Point Secondary Surface Manifestation (SM), 21. Polygonal Secondary Surface Manifestation (SS), 22. Morphostratigraphic Unit (MU), 23. Morphostratigraphic

Surface (MS), 24. Morphostratigraphic Landform (ML), 25. Cultural Stratigraphic Surface (CS), 26. Landuse and Land Cover (LC), 27. Basin Related Drainage Morphometry (BD), 28. Point Related to Litholog (PS).

The themes marked with \* were available in the pilot project area and, therefore, tracings of targeted four sheets were prepared with these 7 themes available.

The working procedure of different systems is described below:

#### **Work on PC/ArcInfo**

1. Digitisation was done for all the lines including map boundary, lithological boundaries, faults and thrusts. Lithological units, which were mapped as discontinuous patches within the major lithological formations but depicted the fold patterns or movement along the faults, were also digitised to be used as trend lines to indicate the structural information.
2. After digitisation was complete, the table co-ordinates were converted to decimal degrees, followed by conversion to polyconic projection on Everest ellipsoid (1956). The map units were taken as metres.
3. Clean and Build commands of ArcInfo were used to generate the Polygon Topology.
4. The coverages thus created, were exported into ASCII format (.E00) files and were transported to UNIX Version of ArcInfo using File Transfer Protocol (FTP) on the Local Area Network (LAN) for further processing.

#### **Work on UNIX/ArcInfo**

1. The .E00 files were imported in the UNIX/ArcInfo and on-screen editing was carried out feature by feature by comparing the lines with the original maps. This included smoothing of lines, corrections of under-shoots and undesirable over-shoots (dangles).
2. After carrying out all the corrections, fault over-shoots and discontinuous minor lithological boundaries were deleted. This coverage was then saved as polygon coverage by a suitable name (e.g. LI53F09) indicating the layer of 'Lithological Units'. Polygon topology was built by using BUILD command with option 'POLY' to generate the fresh Polygon Attribute Table (PAT).
3. A new item 'geometry\_id' (to have a link between the thematic data on Oracle database and related geometry) was created in the PAT and the polygons indicating the lithological units in the particular sheet were encoded as per the common lithological legend of the project area. Unique integer numbers were used for the 'geometry\_id' for each polygon in consultation with the Thematic Database input/output Supervisor. The coverage was saved and BUILD command for polygons was used again.
4. To create the layer for boundary types, the LI-coverage was used again. But, this time, the polygon labels were deleted. BUILD command was used with option 'LINE' to create a fresh Arc Attribute Table (AAT). This coverage was saved by another name (e.g. BO53F09) indicating the layer of 'Boundary Types', without making any changes to the original LI-coverage. New item 'geometry\_id' was added and the lines indicating different boundary types were

encoded with unique integer numbers as done in the case of LI-layer.

5. The procedure described in steps 3 and 4 was repeated on the first coverage to extract required theme by deleting the undesired details and saving the coverage with a new name indicating the theme type (viz., FA..., FO..., TL..., etc.), encoding the features etc. This helped in maintaining the perfect registration between all the layers.
6. For creating the Oriented point structural elements (OS... layer), the strike and dip directions were digitised as LINES. While carrying out on-screen editing, label points were created in the middle of each symbol (viz., attitude of bedding, foliation, joints, etc.). Geometry\_ids were encoded for each label and BUILD command with 'POINT' option was used to create Point Attribute Table (PAT).
7. In order to have perfect match of details between the edges of the four geological sheets, new coverages were created by joining the four sheets for each theme. This not only served the purpose of perfect matching but also helped in removing the discrepancies introduced at the time of compilation/tracing of geological data on the original maps.

#### **Output Generation**

While carrying out the above procedure, the individual layers were being exported again as ASCII file to PC/ArcInfo platform, from where the test outputs were being generated by the input/output supervisor, using PC/ArcView by linking the geometry with the thematic database from Oracle platform. This also helped in locating the errors and taking suitable measures to rectify them. The work was completed by September 1998. A hard copy was generated through shape files (Plate-2) and the same was saved as \*.E00 files for archival purposes (Misra *et al.*, 2000).

#### **Thematic Database**

Thematic data (attribute/textual data) was input from the transcripts already prepared from the mapping reports of different workers under the 17 theme forms discussed above with spatial theme. At present these forms are being transformed to make them available for in-house portal users of GSI with the help of a well-known software solution provider of the country. Similarly, database is being migrated to a central place.

#### **2.2.2 Theme: Geochemical Exploration**

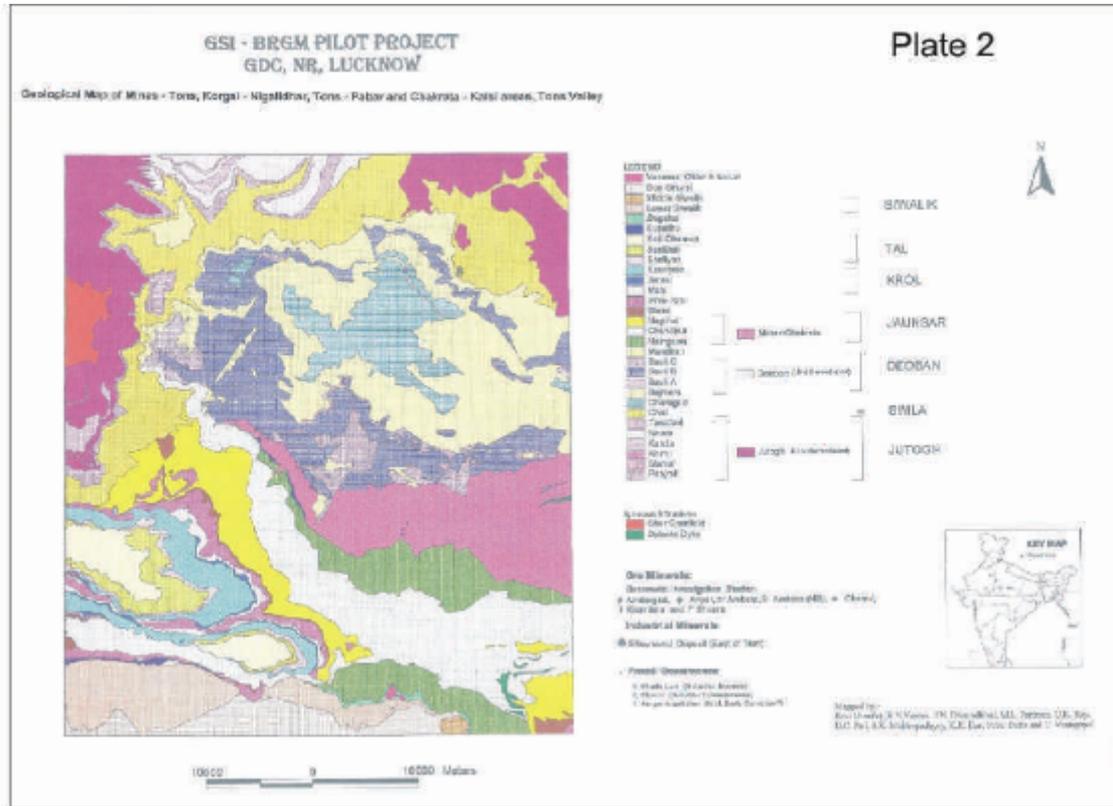
The input work carried out in this theme is from the six Geochemical Exploration blocks falling within the Pilot project area Tons valley, namely, 1) Amtiargad South, 2) Amtiargad North, 3) Chamri, 4) Ambota-Jalau, 5) Ambota northeast and 6) Khamiara in which only first two have been surveyed on true grid pattern. Collation of data was done. Transcripts prepared for these blocks included details of investigation area, geochemical exploration, sample point description, description of one batch of sample, description of an analytical sample and analytical results. These transcripts of textual data are then utilised in inputting the screen form designed for this purpose.

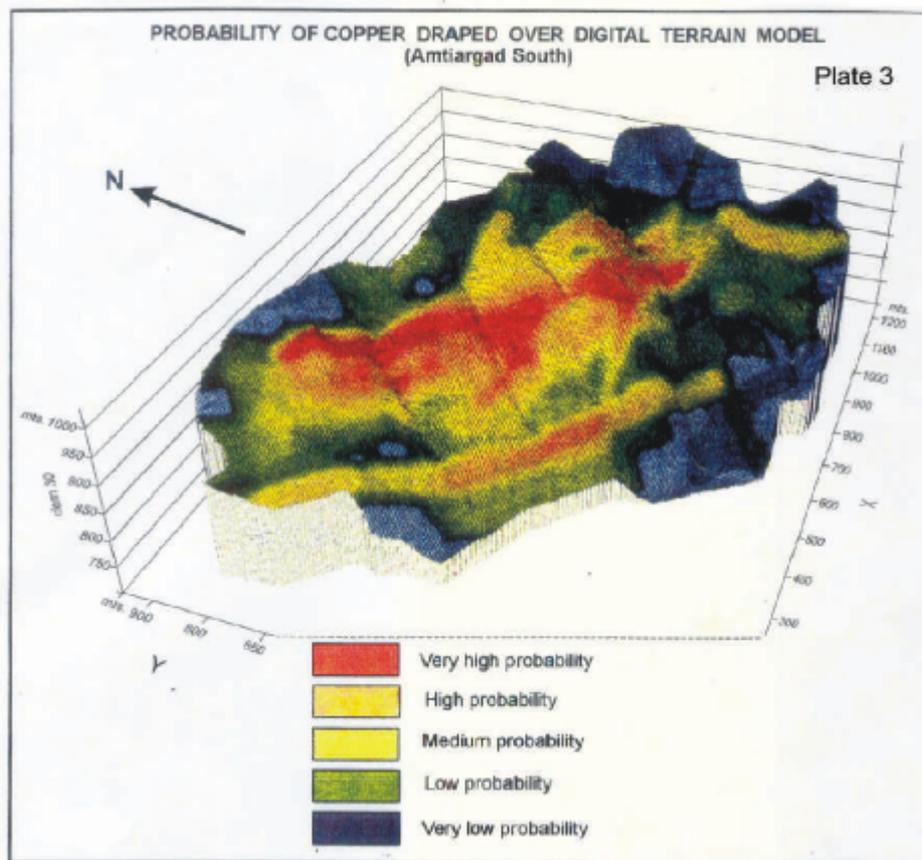
#### **2.2.3 Theme: Mineral Exploration/Mining**

The input work carried out in this theme is based upon the data collated from mineral occurrences, prospect, block, ore-body and sample point information of all the mineral exploration work carried out in pilot project area, Tons valley. The transcript thus prepared, was made ready for inputting on the screen form designed for this purpose.

**2.2.4 Theme: Drilling**

The input work for this theme involved collation of data from drilling operations of Chamri, and Amtiargad south blocks for which the reports are available along with final reports of field projects and from drilling operations of Anyar and Amtiargad north from the draft reports as well as the GDM databases of these blocks, prepared for univariate and multivariate analysis. The transcript, thus prepared, was made ready for inputting on the screen form designed for this purpose.





### 3. MULTIDATA PROCESSING

The multidata processing for mineral targetting was carried out in a block of mineralised belt of the Pilot project area. The area selected for the Pilot Study lies in the parts of Inner and Outer Lesser Himalaya, exposing the rocks of Proterozoic, Tertiary and Quaternary Age. Geological Survey of India, NR had an ongoing mineral investigation programme for base metals in the Tons valley area. The work carried out under the project involved preparation of spatial and attribute/textual database covering four geological maps on 1:50,000 scale, in the parts of Sirmour district of Himachal Pradesh and Dehradun district of Uttaranchal (Srivastava *et al.*, 2000).

The mineralised area falls almost in the middle of the pilot study area and is known as Tons valley base metal belt. The mineralised belt has been divided into several blocks viz., Amtiargad, Anyar, Ambota, Chamri, Khamiara and Shaora. Out of these, a part of Amtiargad South Block, measuring 0.88 km<sup>2</sup>, was taken up for multidata processing. Geological map of this block on 1:2,000 was digitised alongwith topographic contours. Point data for geophysical and geochemical surveys alongwith the geological map were rasterised using SynARC software. An aerial photograph of this block was studied and geometrically corrected, using ERDAS Imagine software. Multi-variate parameters viz., topographic contours, streams, chemical data (Cu, Pb, Zn, Ni, Ag and Co), geophysical data (Chargeability, Self-potential and Resistivity) and photo-geological data (lineaments and landuse/geology) were processed and used for multicriteria analysis. Weightages

were given in consultation with the field geologist and geophysicist. SYNCOMPO software was used to generate an output showing six parameters individually and the result draped over the Digital Elevation Model (Plate-3).

### 4. SOFT COPY CONVERSION OF MAPS AND REPORTS

An ambitious programme of soft copy conversion of maps has been taken up with the basic objective of providing geological information on computers. Sheet-wise compilation of geological maps on 1:50,000 scale has been initiated in the beginning of the project. These maps have been validated ensuring proper edge matching and relational accuracy. The maps are then digitised and information under different domains transcribed for inputting in the database application forms. It is essential to have graphic data in different layers so that these could be used in GIS environment for multiplex applications.

Soft copy conversion of reports, both in text and graphic inputs, has been taken up for archival and dissemination in public domain. The work has by now progressed considerably.

### Acknowledgements

The authors are thankful to the Director General, GSI for his kind permission to publish this paper. They are also thankful to Dr. M.P. Singh, Professor and Head, Department of Geology, University of Lucknow, for encouragement and fruitful discussions.

## References

1. Misra, R.S., Mathur, S.K., and Singh, Baldau, 2000. Final Report on Creation of Theme-Based Database of Degree Sheet No. 53F Covering parts of Uttaranchal, Uttar Pradesh, Punjab, Haryana and Himachal Pradesh. Unpublished G.S.I. Report for F.S. 1998-2000.
2. Srivastava, G.S., Malaviya, A.K., Misra, R.S., Bajpai, I.P., Srivastava Rajeev, Brijesh Kumar, Singh, Baldau, Srivastava, Vachaspati, Verma, Rina, Srivastava, Mridul and Shukla, M.K., 1998. Report on the Setting up of Geoscientific Data Centre in Northern Region Under GSI-BRGM Collaboration, Unpublished G.S.I. Report for F.S. 1996-1998.
3. Srivastava, G.S., Malaviya, A.K., Misra, R.S., Bajpai, I.P., Srivastava, Rajeev and Singh, Baldau, 2000. Multicriteria analysis for Mineral Targeting in a part of Amtiar Gad South block, Tons Valley, Uttar Pradesh, using GIS Techniques. Proceedings of the Seminar ISRS-2000, IIT, Kanpur.