

CWIS-LIKE DEVELOPMENTS AT THE IMPERIAL CANCER RESEARCH FUND

Mary Davies

Paper given at the 17th UKSG Annual Conference, Manchester, 11-14 April 1994

Increasingly libraries of all types are employing software tools such as Gopher to provide easier access to internal and external information resources. An overview is given of the CWIS-like system in operation at the Imperial Cancer Research Fund. The paper focuses on the provision of electronic current awareness services via ICNET, ICRF's UK-wide kilostream network, and the electronic requesting activity of users relating to document supply.

Mary Davies is Head of Library & Information Services at the Imperial Cancer Research Fund, PO Box 123, 44 Lincoln's Inn Fields, London, WC2A 3PX

Background and Context

Organisation

The Imperial Cancer Research Fund (ICRF), a registered charity, is the largest independent cancer research organisation in Europe. Its main laboratory site is at Lincoln's Inn Fields in Central London and a satellite laboratory site is situated at Clare Hall, South Mimms, in Hertfordshire. ICRF also has groups and Units, mainly attached to hospitals, throughout the UK.

Finance

Historically, ICRF has been financed by public donations and legacies; income from its country-wide shop chain, regional activities and donations are still its main sources of income. Other sources of revenue are grants from the European Union and post-doctoral funding agencies.

Staff

The present staff numbers total 1731 and consist of 597 scientific staff (mainly molecular and cellular biologists), 374 technical staff and 760 administrative, clerical and support staff.

Computer Systems/Networks

The Fund's main computing facility comprises two clustered DEC Vax 4600 computers running the VMS operating system, which are to be imminently replaced with two DEC AXP server machines running the OSF/1 Unix operating system. A collection of Sun servers handle the Purchasing and Finance Departments' databases.

Since 1989, ICRF has had its own kilostream TCP/IP network (ICNET) linking over thirty of its remote sites to the central computing facilities. At present there are about 700 Macs, 200 PCs, 80 Sun or other Unix workstations and numerous dumb terminals connected to the network. In addition, ICRF has access to the JANET network via a leased line to ULCC and since June 1992 to the JANET Internet Protocol Service (JIPS) via a 64 Kbit/second link to ULCC. In January 1994, ICRF converted to a native Internet Protocol (IP) link to JIPS, reportedly doubling its throughput as a consequence.

The main services offered on the main computing facility are Library & Information Services' databases and molecular biology databanks. In both cases, the Fund finances these services centrally, the intention being that all ICRF staff country-wide should have access to the same resources regardless of location. The Library & Information Services (LIS) Department had been concerned for a number of years about the difficulties encountered by scientists in connecting to and negotiating their way around the central computing facility. Its OPAC (based on the BASIS/BASISPlus software) had attempted to facilitate the latter (navigation around the central system once connected) by providing, via its menu (see Figure 1), easy access to in-house databases and information resources and seamless connection to externally provided databases/databanks.

CWIS-like developments

By the end of 1992, there was a consensus between LIS, the Biomedical Informatics Unit (BIU) - the computing department responsible for the molecular biology databanks, and the Central Computing Unit (CCU) that the difficulties experienced by scientists in utilising the central computing facility were primarily due to the lack of uniformity of interfaces to programs and information resources.

Software Tools

The evaluation of public domain software for various purposes is an ongoing activity for two of the ICRF computer units, namely, the CCU and the BIU and, at that time, this included the Gopher software. Also, Library staff had had experience of accessing Internet resources using Gopher and were enthusiastic about its potential. There was general agreement that the Internet Gopher would be a useful tool for storing and distributing information within ICRF, for accessing internal and external information resources and, potentially, for forming part of a corporate ICRF information service. The availability of client programs for all the platforms in the Fund (VMS, Unix, Apple Mac and IBM-compatible personal computers) was

one of the strongest arguments of all in support of its adoption.

Participants

Early in 1993, the Gopher Project Steering Group (GPSG) was formed, comprising members from the three departments - LIS, BIU and CCU. It was agreed to launch a pilot service to gain familiarity with the technical and resource implications of running a Gopher service and to provide data for the case to be made to management for resources for a "live" service. The pilot ran much longer than expected, from May 1993 to March 1994, although senior management approval for the resources to host a live service was secured in October 1993. The Gopher service eventually went "live" in March 1994 hosted on a Sun SparcServer 10/41. Technical support and user support are provided by BIU and CCU, the server itself lying with the latter and Gopher Administration lying with the former.

Motivations

Each of the three departments involved in the GPSG had their own particular motivations for supporting the setting up of the Gopher server. These included a common user interface for accessing local and remote information resources; the relative ease of use of the software; and the ability to view files, search simply-structured databases, retrieve documents and programs across the network and print or store them locally.

Information Providers

Responsibility for updating the information on the Gopher service lies with the department providing the information. A special e-mail handler allows information providers to submit an updated document as a text file directly by electronic mail and processes it for the Gopher server. The updating of more complicated documents is handled either by the Gopher Administrator or the department itself, depending upon available staff expertise.

Most of the information available on the Gopher service is provided by the three

departments comprising the GPSG. However other departments and laboratories are being encouraged to participate.

Content

- ◆ About the ICRF Gopher Service
- ◆ Library & Information Services
- ◆ Molecular Biology Software and Databases
- ◆ Computing Services
- ◆ Support Services
- ◆ Staff Information
- ◆ Seminars within ICRF
- External Gopher Servers

Most of the information in the LIS section is a subset of the information provided on the Vax cluster. This subset covers the full text of LIS Guide, the current journals list, a list of journal issues received this week, the full-text of the ICRF Scientific Report, EndNote software aids and a Gopher hole connection to the LIS databases on the Vax cluster.

The Molecular Biology section contains guides to the main DNA/protein sequence analysis programs on or via the Vax cluster (such as GCG, Intelligenetics, Staden and SharQ - Sequence Homology Alignment Remote Query) and guides (some full-text) to the Vax-held DNA (Genbank and EMBL) and protein (PIR, SWISSPROT, BROOKHAVEN) sequencing databases plus the OWL Protein Database Query Program. Details are also given of molecular biology software for the Apple Macintosh supported by ICRF. Also detailed are the services available via a growing number of e-mail servers. In addition, a local copy of Una Smith's *A Biologist's Guide to Internet Resources* is available, as is access to other biogophers (e.g. University of Minnesota, University of Indiana, John Hopkins University, National Institutes of Health USA and National Cancer Center Tokyo).

The Computing Services section contains listings of services and personnel associated with the three computing units (CCU, BIU and ACU). The ICRF local Information Technology bulletin board is also accessible here.

The Support Services section contains the Stores Catalogue, the LIF Cell Lines Catalogue and the Central Services Handbook. The Staff Information section just contains the telephone directory at present. The External Gopher

Services section contains a local copy of the addresses of gopher servers world-wide.

Potential local additions to the Gopher service include European Union calls for funding, internal newsletters, ICRF Clinical Trials Register, health & safety information, internal vacancies, staff manual, pay scales, Staff Consultative Committee Minutes, ICRF Colloquium Information, list of surplus equipment and various internal committee minutes.

Current awareness system

One of LIS' main uses of ICNET is the provision of electronic current awareness services and the electronic requesting activities of users relating to document supply.

The ICRF Library leases copies of the Science Citation Index (SCI) source tapes with abstracts from the Institute for Scientific Information (ISI) on a weekly basis. These tapes form the basis of the Library's current awareness service, although other SDI services are provided to some staff from databases mounted on commercial hosts (for example, Psychological Abstracts, Applied Social Sciences Index, Sociological Abstracts, Dissertation Abstracts, Human Nutrition and CancerLit).

The weekly tapes are loaded onto BASISPlus utilising a data transfer program (available from the UK vendors of BASISPlus, Information Dimensions UK Ltd), which converts the ISI data format to a BASISPlus format. During this loading process titles held by ICRF Library are marked. Batch searches are performed by library staff on particular subjects, groups of subjects, particular authors and contents of specific journals, according to profiles agreed with individual scientists. Between 350 and 400 profiles are run weekly and the majority of scientists receive their search results in electronic form via the Vax e-mail system. The data is provided in DataStar Medline format to facilitate the downloading and importing of results into reference handling packages, EndNote being the most widely used package.

The SciSearch Database is available online for end user searching via the Library's OPAC. Until earlier this year, only a rolling latest six months of data (with abstracts but abstracts not indexed)

was available, due to shortage of disc space on the central system. Now, SciSearch is available with abstracts (indexed) from January 1992 to date.

ICRF is part of the Queen Mary and Westfield College cluster and has had access to the BIDS ISI service since the autumn of 1992. To date, the BIDS ISI service has been viewed as a complementary service to our in-house service as BIDS has SciSearch back to 1981 with cited references but no abstracts whilst our in-house service has had, until recently, a relatively small amount of data available without cited references but with abstracts. Obviously, the pending addition of abstracts to the BIDS ISI service will necessitate reconsideration of what we provide in-house.

The resources needed to run an in-house current awareness service are considerable. The three main resources in our case are: the cost of leasing the tapes themselves; the staff required for profile creation and maintenance plus systems support for software maintenance and development; and computing resources in terms of weekly loading and batch searching and associated disc space for the interactive database.

Document supply

In addition to users' requests for articles from journals to which the ICRF Library does not subscribe, the Library provides a photocopying service from journals, to which the Library does subscribe, to ICRF staff who are not based at the central London laboratory site. About thirty thousand documents are supplied each year, between 16000 and 19000 from the British Library and between 11000 and 14000 from the internal photocopy service. The high volume of external requests is mainly due to the fact that less than 200 journals are taken in-house whereas analyses have shown that articles from over 2,500 titles are being requested at any one time.

Users' requests of both types emanate from the Library's current awareness service, the SciSearch Database service, the Library's remote literature search service, CD-ROM Medline database service, the BIDS ISI service, user-keyed electronic requests and manual requests made on requests cards. The Inter-Library Loans Database

provides facilities to enable users to request references electronically. Between one and two hundred requests are received daily. The manual requests are keyed into the Inter-Library Loans Database and the electronic requests are uploaded. The program automatically streams out those requests that are for ICRF held journals and formats the rest into ARTTeL format for submission to BLDSC via the JANET network. A number of years ago, a decision was made to use BLDSC as the "one stop shop" for ILLs. This policy is under active reconsideration, as I shall mention later.

Analysis of the source of requests is part of our on-going monitoring of users' requesting behaviour.

Future developments

In-House Developments

Much of the discussion of in-house developments at ICRF tends to focus on the capacity and the performance of ICNET as almost any development impacts on it.

- ♦ the extending of the locally-held SciSearch Database to 3, and possibly to 5, years will impact on the network and the increased data availability is already increasing usage;
- ♦ the existing electronic Current Awareness Service - abstracts provision with weekly output and the availability of abstracts for interactive searching is increasing volumes of data sent and user accesses.
- ♦ LIS Databases Service will continue to expand - whether ICRF Library negotiated (like the Turing Institute databases) or JISC Database deals. Again, if implemented, this will result in increased user traffic.
- ♦ the networking of MEDLINE (and possibly other databases) over ICNET has been long-awaited and ICRF will probably be a beta test site for SilverPlatter's ERL system on the Sun platform this summer. Traffic involved here is difficult to predict.
- ♦ Gopher, which relies on the use of ICNET to transfer documents from server to client. How will this affect network bandwidth resources? Will people make unnecessary use of national and international Gopher servers for information that is available

locally? Traffic will be monitored to determine whether more popular remote services can be supported locally.

- ✦ Electronic Document Delivery (e.g. RightPages) would have a major effect on network traffic.

We are already approaching saturation on ICNET. Increasing packet loss and re-transmissions are being observed at peak times. The requirement for increased Internet bandwidth is also under active discussion. Many ICRF European Union funded collaborative research projects have a significant computing component (e.g. Integrated Genome Database Project, IDEA and APPLAUSE Esprit Projects) and international collaboration involves some or all of the exchanges of data via the Internet. The increasing use of national and international Gopher servers adds to the requirement for Internet bandwidth. Earlier this year it was agreed to increase the capacity of the JIPS line to ULCC to 256 Kbits/s with a view to a further upgrade to a 2M

connection by the end of 1994. SuperJANET is an important future component here.

CAS-IAS Systems

The in-house electronic Current Awareness System versus CAS-IAS (including BIDS ISI) developments issue. There is on-going evaluation of the new CAS-IAS services (forming part of BUBL's PLANET initiative) that have mushroomed recently and, as mentioned above, the provision of abstracts with BIDS ISI data will be carefully considered in relation to this issue.

User Consultation

Research is on-going with sample groups of different categories of users - to ascertain what their current information handling techniques are, to what extent the present services are meeting their needs, and what they perceive to be their future needs - to aid the evaluation and planning processes.

Figure 1

LIBRARY'S DATABASES MAIN MENU

```

*****
*                               Welcome                               *
*                               t o                                 *
*                               *                                   *
*   I C R F  L I B R A R Y  &  I N F O R M A T I O N  S E R V I C E S   *
*                               *                                   *
*                               D a t a b a s e s                   *
*                               *                                   *
*   In-House Information Sources      . External Information Sources *
*   0. Guide to L&IS Services          31. BIDS Databases          *
*   1. Books Database                  35. Turing Institute Databases *
*   2. Journals Database                45. EuroCODE & PDQ Databanks *
*   3. ILLs Requests Database           47. MSDN Databanks           *
*   4. SciSearch Database               50. UK University Catalogues *
*   5. Scientific Reports               55. Electronic Yellow Pages *
*   6. Staff Publications Database      70. EndNote Software Aids   *
*   7. Staff/Laboratories Database      77. EXIT                     *
*   10. Research Funding Database
*   20. UK DNA Probe Bank
*   21. Reference Library DataBase
*       (Genome Analysis Lab)
*   22. Central Cell Services Handbook
*****
Enter your choice:
    
```