PART B : USES

As mentioned in Part A, we were asked as part of the Design Study work to collect information about possible uses of the 'ideal' collection, were it to be built, for research and teaching. This would provide some basis for an assessment by or on behalf of BLR&D as to whether the 'ideal' collection is really needed, that is, is stated to be needed by research workers and teachers, is in fact needed, and is needed for worthwhile projects and activities.

The questionnaires used to collect the information, one for research and two for teaching, are reproduced in Appendices 8 and 9. An estimate of the potential utility of an 'ideal' collection, to complement the specification given in Part A, can be obtained from the responses to the questionnaires, which are summarised below.

1. Research projects

The table below gives those to whom the questionnaire of Appendix 8 was sent. The 48 persons listed were chosen either because they had been involved in previous discussions of the 'ideal' collection and/or have been active in research in recent years. The table also gives the distribution of responses by five categories: thus 27 out of the 48 responded, 23 of them favourably with 28 projects altogether.

Name	Place	Name	Place
T. Aitchison	Inspec	B. Kostrewski	City U
D. Austin	BL/BSD	R. Lea	Brunel U
K. Bakewell	Liverpool P	J. Leigh	BL/SRL
E. Barraclough	Newcastle U	M. Lynch	Sheffield U
*G. Bates	Cambridge U	J. Martyn	Aslib
N. Belkin	City U	J. Mills	North London P
R. Bottle	City U	A. Negus	Inspec
M. Brittain	Loughborough U	B. Niblett	U.C. Swansea
C. Cleverdon	Cranfield I	R. Oddy	Aston U
A. Cooper	Loughborough U	R. Richens	C.A.B.
*B. Croft	Cambridge U	C. van Rijsbergen	Cambridge U
S. Datta	North London P	S. Robertson	University C
J. Digger	L.A.	A. Robson	Ukcis
L. Evans	Inspec	J. Smith	Belfast U
B. Field	Inspec	K. Sparck Jones	Cambridge U
A. Flowerdew	Centre for Env. Stud.	D. Swift	Open U
A. Gray	Cardiff U	J. Terry	A.E.R.E.
*J. Griffiths	University C	D. Veal	Ukcis
A. Harley	BL/LD	B. Vickery	University C
*D. Harper	Cambridge U	R. Wall	Loughborough U
M. Heine	Newcastle P	*A. Wheatley	C. Libr. Wales
V. Horsnell	City U	P. Williams	UMIST
A. Hindle	Lancaster U	E. Wilson	Kent U
M. Keen	C. Libr. Wales	P. Yates-Mercer	City U

* = research students/assistants treated as a random sample of possible future research workers

```
Total names                         = 48
Total responses                     = 27
 Y    = positive response, form(s)
        returned with specific project(s)
        details                     = 17  for which projects = 23
(Y)   = positive response, but in
        letter without organised
        detail e.g. of scale        =  5
 ½    = interested but unable for
        indicated (usually good)
        reasons to provide detail now = 2
 N    = negative response, indicating
        not interested or unable to
        envisage project in current
        environment                 =  2
        mixed                       =  1
Total non responses (-)             = 21
```

The research projects supplied are summarised, and salient features relevant to the 'ideal' collection as specified in Part A are noted, in the table below. It must be emphasised that these brief project descriptions are our own summaries of typically paragraph length statemts, and that the projects should not be judged in detail on them.

Summary of possible projects in 28 responses Y and (Y)

Project	Our comments re IC Specification
1. Investigate specialist/generalist differences among indexers and users of indexes.	IC varied enough in subject?
2. Comparison of PRECIS and other indexing methods.	Social sciences preferred: document texts needed; IC enough subjects, especially for monographs?
3. Investigation of proposition that for a given collection there is an optimum size of dictionary.	
4. Work on bibliographic data base management techniques.	Citations of interest.
5. Test of an information retrieval system based on an 'anomalous state of knowledge' hypothesis.	Multidisciplinary document set needed and exhaustive relevance judgements.
6. Research on nature, use and value of document citation clusters.	Citations needed
7. Study of file organisation techniques for large document sets.	IC large enough?

8. Study of various automatic/manual indexing mixtures.

9. Analysis of statistical characteristics of profiles and relevance assessments.

Range of alternative request sets (inc. SDI) needed.

10. Application of fuzzy set theory to information retrieval.

IC large enough?

11. Application of graph-theoretic information theory.

IC large enough?

12. Study of library circulation data and their uses.

IC no information of this kind.

13. Identification of search techniques genuinely suited to on-line retrieval.

14. Psycholinguistic study of acceptability of presentation of search output.

15. Assessment of text compression techniques in on-line retrieval, in relation to special hardware.

Would like data soon.

16. Hardware evaluation for on-line retrieval.

Would like data soon.

17. Study of effects of different key sets in retrieval.

18. Research on use of long, highly structured queries for text searching.

Full texts needed.

19. Exploitation of term associations in interactive retrieval.

20. Study of efficient document clustering techniques and of the relations between document and term clustering.

21. Research on term dependency.

22. Identification of good index terms through a study of their predictors.

IC enough requests?

23. Development of retrieval system testing methodology.

Much actual search data is needed.

24. Analysis of real value of Boolean request structures.

25. Research on atuomatic feedback techniques
involving query expansion.

66. Investigation of 'aboutness' and the search   Social sciences
process.   preferred.

27. Development of procedures for exploiting
the IC in teaching subject retrieval.

28. Work on indexing languages specially designed
for automated searching.

in addition

1   mixed response lists possible projects on
the intermediate lexicon, bibliographic coupling
versus subject document classing, and term
relations; but these are very tentative and
are therefore not considered further.

We have categorised the projects in various ways so as to give some
idea of the scale and type of the research effort involved, in relation
to the 'ideal' collection specification. The table below lists the
results. We again emphasise that this categorisation should not be
regarded as wholly accurate, since it in some cases depends on inference
from the actual information supplied. In section A of the table projects
are grouped by scale, and by whether their requirements in terms of
quantity and kind of data could be met by the 'ideal' collection as
specified. In B they are characterised according to whether they would
use the collection as it stands or extend it. In C by size in terms of
staff and time. In D by status, and in E by the type of computing
environment required. Finally, in section F, the projects are grouped
according to whether they are, in our view, basic or general information
retrieval research projects, or are applied or specific.

Categorisation of possible projects (by (KSJ, with some inference)

A. Data requirements   (subject to the comments above)   No. projects

   1. Extensive scale (i.e. size rather than richness)   25

   2. Moderate scale   3   (28)

   Requirements in scale and content:

Could be met by IC as specified   14

Could probably be sufficiently met by the IC   10

Doubtful if could be met   3

Could not be met   1   (28)

B. Use

    1. Would use IC primarily as it stands   11
    2. Would extend IC through new indexing or new requests   16
    n/a   1 (28)

C. Size of possible projects

Time

    1. Short, 1 year   1
    2. Medium, 2 years   12
    3. Long, 3 years   5
    ?   3
    not indicated   6
    n/a   1 (28)

Staff

    1. Small, 1-2   16
    2. Medium, 3-4   4
    3. Large   −
    ?   1
    not indicated   6
    n/a   1 (28)

D. Status

    1. Fairly definite (irrespective of whether,
       administratively, could start now)   18
    2. Preliminary work or pilot study needed, or tentative   9
    n/a   1 (28)

E. Machine environment

    1. Would need IC on-line   9
    2. Would not need on-line (though might be useful)   18
    n/a   1 (28)

F. Character (KSJ's interpretation)

    1. More basic, or general, research   13
    2. More applied, or specific research   13
    n/a   2 (28)

From the information summarised in these tables we obtain the following picture of the 28 projects.

a) Nearly all of them require large scale data. The scale and content of the data required by 14 of the 28 could be met by the specified 'ideal' collection, and the requirements of another 10 could probaly be satisfied in practice; only a few could probably not, and only 1 certainly not.

b) 16 of the projects would develop the collection further in indexing or searching, largely for comparison with the information already provided. It is worth noticing that many of these projects would include searching for new requests, for which the collection, contrary to some views of it as unacceptably static, is deemed acceptable.

c) many of the projects are middling in size: 2 years and 2 staff respectively are the favourite choices.

d) more than half of the projects, 18, are fairly definite; some of the remainder are tentative and some depend explicitly on preliminary or pilot studies.

e) 9 of the projects depend on using the 'ideal' collection on-line, and some of the others would probably find such use convenient.

f) the projects divide evenly by their research character, basic or applied. Many of them are of a quite concrete character relevant to operational systems in some degree, and there are very few which might be described as quite remote from such systems.

Overall, the picture presented by the research project replies is a fairly solid one. The response rate from active research workers was quite high (some of those to whom questionnaires were sent being no longer active): and the number of projects involved would represent a substantial proportion of the total research effort being conducted in information science in this country. The projects are on the whole fairly concrete and fairly moderate ones which if successfully conducted would bear on operational systems, and in particular on modern on-line systems. At the same time they generally call for data on a scale and of a kind not currently available for test purposes. Whether much of the research is of a desirable type, and whether the data required could be provided other than through the 'ideal' collection, are points considered in Part C of this Report.

It is quite improper for us to attempt a detailed evaluation of the quality of the projects submitted: indeed it must be emphasised that the covering letter accompanying the questionnaire stated explicitly that there could be no commitment to any projects outlined either from those supplying them or from BLR&DD as a grant-giving body. The questionnaire did not call for great detail, and the projects outlined should not be evaluated for merit as if they were regular proposals; it should also be borne in mind that the projects which could not be conducted for, say, two years, since if the 'ideal' collection was to be built it would take some time to set up.

2. Teaching activities

a) general

The questionnaire for teaching establishments was very simple, (see Appendix 9), and was designed only to discover whether these have any interest in the 'ideal' collection, and if so what specific requirements would have to be met. The table below lists the establishments approached, and their responses.

Place	Reply (Y = response - = no response)

Aberdeen
    Robert Gordon's Institute
    of Technology
	-

Aberystwyth College of Librarianship Wales	Y
Belfast Queen's University	Y
Birmingham City of Birmingham Polytechnic	-
Brighton Brighton Polytechnic	-
Glasgow University of Strathclyde	Y
Leeds Leeds Polytechnic	Y
Liverpool Liverpool Polytechnic	-
London Ealing Technical	-
London Polytechnic of North London	-
London University College London	- (but teaching interests in part covered by research project)
Loughborough Loughborough Technical College	-
Loughborough Loughborough University	Y
Manchester Manchester Polytechnic	Y
Newcastle upon Tyne Newcastle upon Tyne University	Y
Sheffield University of Sheffield	Y

total Y = 8   (plus University College London)

        - = 9

For those establishments responding positively the table below indicates expressed needs, and comments on them in relation to the 'ideal' collection specification. The numbers of students involved are added for information.

Requirements of teaching establishments in responses Y

Establishment	Need	Comments re IC Specification
Aberystwyth	1 large collection as good as ON-TAP	enough relevance information?
Belfast	1 large collection e.g. CAB 1 large collection MARC non-specialised subject (soft)	OK not in IC OK
Glasgow	small collections including e.g. UDC, also non-biblio.	mostly OK but UDC?, non-bibliographic?
Leeds	small collections	OK
Loughborough U.	several collections	OK
Manchester	non-specialised subject 1 large file on-line	OK OK
Newcastle	2 large files	would need extended social science other set
Sheffield	1 large file on-line	OK

Numbers of students

Large = 100+	undergraduate	Aberystwyth, Manchester
"	postgraduate	Aberystwyth
Medium = 50-100	undergraduate	Glasgow, Leeds, Loughborough U, Newcastle
"	postgraduate	Glasgow, Leeds, Loughborough U, Manchester, Sheffield.
Small = -50	undergraduate	Belfast
"	postgraduate	Belfast

It is to be regretted that the response rate for this questionnaire was not better. The main point to be made about the replies are that in general the 'ideal' collection as specified would meet expressed needs, the main reservations being on the numbers of subjects covered and the provision of MARC records. The establishments responding process significant numbers of students, so the collection could make a valuable contribution to teaching, particularly if it could be set up on-line.

b) on-line education

The point just made was investigated through a specific questionnaire, the second in Appendix 9. This was sent to participants in a British Library-sponsored Workshop. Those from educational establishments were grouped by establishment, though a separate questionnaire was sent to the organiser, Dr. Keenan, and at British Library's suggestion to Professor Vickery. The total educational questionnaires sent out was thus 12 (non-educational participants being sent the questionnaire mainly as a matter of courtesy); 10 responses were obtained. The table below lists establishments and responses, and analyses the positive responses by requirements and computing facilities.

On-line education questionnaire

Place	Reply
Aberystwyth - CLW	Y
Belfast - OU	Y
Birmingham P	Y
Brighton P	-
London -PNL	Y
UCL	Y
CU (= Vickery)	Y
Loughborough U	Y
Keenan	-
Manchester P	Y
Newcastle P	Y
Sheffield U	Y

For responses Y

   Interested in the IC   - yes   Aberystwyth, Belfast, Birmingham, London UCL, London CU, Loughborough U, Manchester, Newcastle, Sheffield

                             no   London PNL

   Special facilities required for on-line teaching

yes	Aberystwyth	:	comprehensive index and search options
	Belfast	:	several subjects
	Loughborough U	:	several collections
	Newcastle	:	several subjects
	Sheffield	:	several subjects, command language compatibility
	London PNL	:	experience of operational systems
no	Birmingham, London UCL, London CU, Manchester		

   Would/do use or prefer remote computer

        Aberystwyth, Sheffield, London PNL, Loughborough U

   Would/do use or prefer local computer

        Belfast, Birmingham, London UCL, London CU, Manchester, Newcastle

The response rate here was good. It was very generally thought that the 'ideal' collection would be of use, the main reservation being the range of subjects. It is worth noticing that there is a bias towards the use of local computers, suggesting a need for a portable on-line search package, which might well be of use for some research projects as well.