Unit 3, The Paddock
Hambridge Road
Newbury
Berkshire
United Kingdom

Analysis of the

Z39.50 Profile for Access to Digital Collections

and

the Z39.50 Explain Service

for UKOLN

Crossnet Job No. F17

29 October 1997

CSL017-TN-001 – Issue 2

Prepared by: R. A. Bull Crossnet

Accepted by:

_____

UKOLN

DOCUMENT STATUS SHEET

DOCUMENT STATUS SHEET			
1. DOCUMENT TITLE: Analysis of the Z39.50 Profile for Access to Digital Collections and the Explain Service CSL017-TN-001			
2. ISSUE	3. REVISION	4. DATE	5. REASON FOR CHANGE
1	1	29 Oct 1997	Initial Issue
2	1	14 Apr 1998	Comments from UKOLN
6 DOCUMENT DISTRIBUTION			
R. Russell.....		UKOLN Crossnet	
R. Bull.....			

Table of Contents

Table of Figures

1 INTRODUCTION

1.1 Scope

This document is a technical note that provides information relating to the ANSI Z39.50 Profile for Access to Digital Collections and the ANSI Z39.50 Explain Service.

2 THE Z39.50 PROFILE FOR ACCESS TO DIGITAL COLLECTIONS

2.1 Introduction to the Profile

In August 1995, the Library of Congress convened a team of representatives from several institutions to develop a Z39.50 profile for access to digital libraries.

Early in its development the profile was renamed from Z39.50 Profile for the Digital Library Application to Z39.50 Profile for Access to Digital Collections as its scope was narrowed, to focus on navigational aspects of digital collections.

It was then recognised that other groups were initiating independent efforts to develop profiles aimed at specific types of objects and collections. A specific effort was made to co-ordinate these efforts in order for these other profiles to be developed as compatible extensions or subsets of the Z39.50 Profile for Access to Digital Collections. These latter profiles are referred to as companion profiles. Two examples are the CIMI Profile managed by the Consortium for Interchange of Museum Information, and the Z39.50 Profile for Access to Digital Library Objects being developed at the Library of Congress.

2.2 Purpose of the Z39.50 Profile for Access to Digital Collections

The purpose of the Z39.50 Profile for Access to Digital Collections is to support libraries and other institutions who are creating a growing number of collections, organised thematically, for example, by subject, creator, historical period, etc. These collections are typically with numerous, diverse objects, both digital and physical and are often organised hierarchically and may be physically distributed across servers.

In museums and archives there are various forms of descriptive aids currently in use for the purpose of describing collections and objects. For example, finding aids (such as the Hyper-G product from Joanneum Research in Graz, Austria), encoded archival descriptions, and exhibition catalogues. Often they do not have a well-defined structure and cannot be used alone for reliable navigation within the collection in spite of the efforts made in the creation of descriptive aids.

Significant resources may be invested in digitisation and in the intellectual efforts of aggregation, organisation, and description of information in a collection.

Without an agreed semantic understanding of the data between the client and the server, the collection may appear to be simply an accumulation of objects and undifferentiated data, because there is no agreed-upon semantics for navigating the collection or to locate and retrieve objects of interest.

It is therefore necessary to arrange for coherent organisational structures to be layered on top of the data in order to provide the necessary navigational visibility.

A primary goal of navigation is to locate and retrieve objects of interest; and in so doing, to locate the relevant descriptive information. A key obstacle to this is the inability to distinguish content from description. Once this is established, it is possible to thereafter

navigate among descriptive information as well as content, and consequently, to be able to distinguish content from description.

The Z39.50 Profile for Access to Digital Collections specifies a conforming subset of Z39.50-1995 to address these problems, and in particular to provide a mechanism for access to digital collections organised via descriptive information, i.e., metadata. In order to achieve this purpose, the profile provides semantics for navigating digital collections, and to locate and retrieve objects of interest.

The Z39.50 Profile for Access to Digital Collections should not be seen as the definitive document for a digital collection search and retrieval system based on Z39.50 - the Profile assumes that companion profiles will be independently developed which are either compatible extensions to, or subsets of the Profile.

2.3 Definition of a Collection in the Profile

In the context of the Z39.50 Profile for Access to Digital Collections, a collection is defined as a group of related objects and/or other collections which may be centrally located or possibly distributed across locations.

The collection can be considered as a tree where leaf nodes are objects of the collection and non-leaf nodes are sub-collections.

This is represented in Figure 2.1 below.

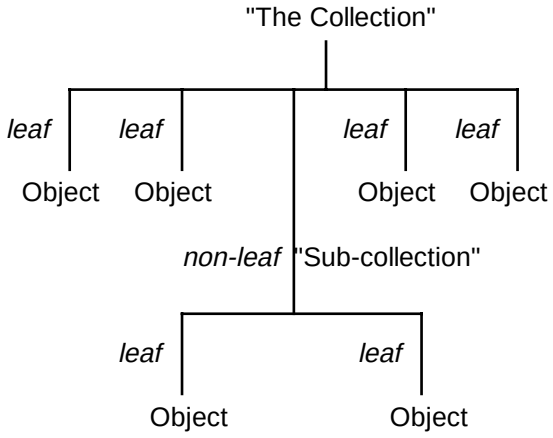

Fig. 2.1Simplistic View of a Collection2.1SimplisticViewofaCollection2.1SimplisticViewofaCollection2.1SimplisticViewofaCollection

The fundamental basis of the Z39.50 Profile for Access to Digital Collections is the definition of the relationships between collections and objects. Referring to Figure 2.2 below:

Collection A is said to be related to collection or object B if the organisation with

responsibility for the management of B considers that collection A may be relevant to a user who is interested B.

Collection A is said to be superior to collection or object B if B is a node on a tree whose root is A; B is said to be subordinate to collection A if A is superior to B.

Collection A is said to be a parent of collection or object B if A is immediately superior to B; B is said to be a child of collection A if A is a parent of B.

Collection A is said to be a context collection for collection or object B if it is a superior, related collection, and the organisation with responsibility for the management of B considers that although collection A may be relevant to a user who is interested in B, any collection superior to collection A is likely not relevant.

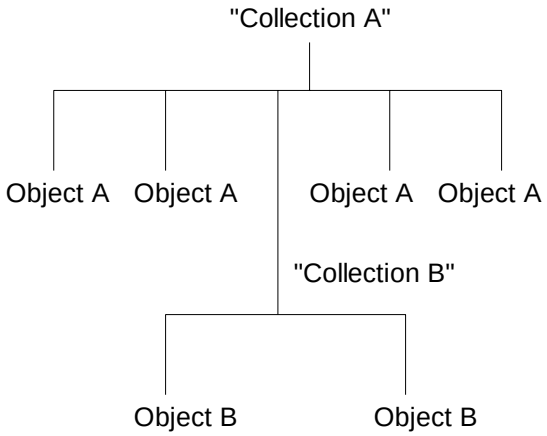

Fig. 2.2 Relationships of Collections and Objects2.2RelationshipsofCollectionsandObjects2.2RelationshipsofCollectionsand Objects2.2RelationshipsofCollectionsandObjects

The profile also refers to “context collections”, whereby the collection of interest is in a hierarchical structure, and the collections of interest are not from the top of the hierarchy. Therefore the collection of interest may be considered as a relative root within the hierarchy. This is shown in Figure 2.3 below, where collection B is considered a context collection as the contents of collection A are not included.

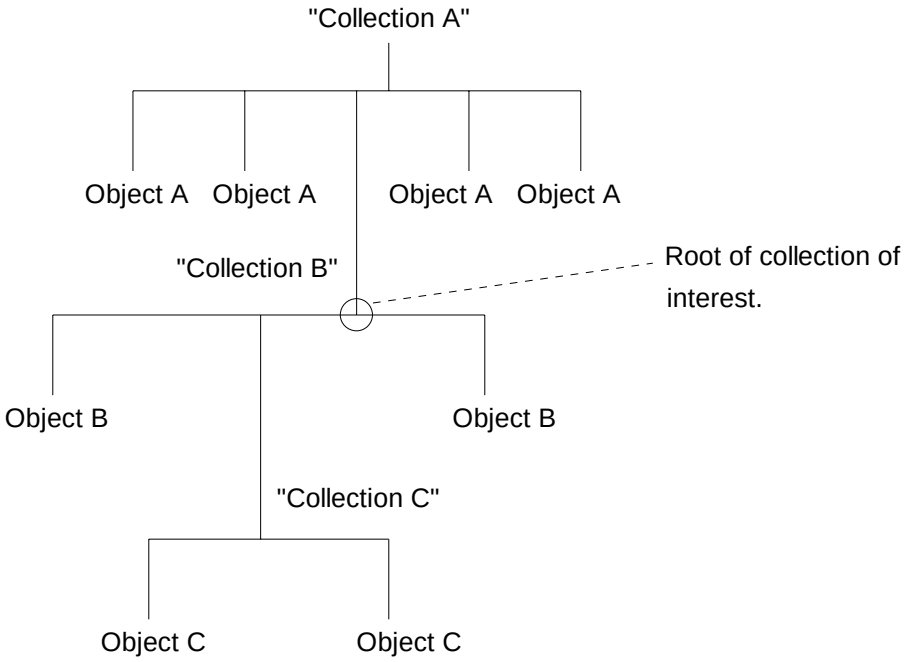

Fig. 2.3 Context Collections2.3ContextCollections2.3ContextCollections2.3ContextCollections

The Digital Collection profile contains the following relevant statements:

> *A context collection might informally be considered a "relative root" collection, for purposes of navigation.*

An example of the relative root approach could be a collection of Roman coins. The entire collection (viewed at root position A) may contain sub-collection (B) of (say) silver and gold coins, in which the gold coins are Collection C. A context collection may view the silver and gold coins as a "collection of valuable coins", in which the search and retrieval at that level would not see the coins that are part of collection A that were not gold or silver.

2.3.1 Collections and Databases

The Z39.50 Profile for Access to Digital Collections makes a distinction between collections and databases through logical constructions. It considers that a database is an aggregation of records, and that a collection is an aggregation of objects, some of which may be in electronic form, and others which may be physical.

It cannot be assumed that a collection will reside on a single database, as the collection may be apportioned across multiple databases where each database may hold data of a particular type, e.g., books, serials or digitised photographs.

However, a collection may comprise elements from each of these databases, and the records are therefore distributed across the databases. This is shown in the example below, where collection C1 comprises of objects A1 and A2 in database A, and object B3 in database B. Child collection C2 comprises of object A3 in database A and objects B1

and B2 in database B.

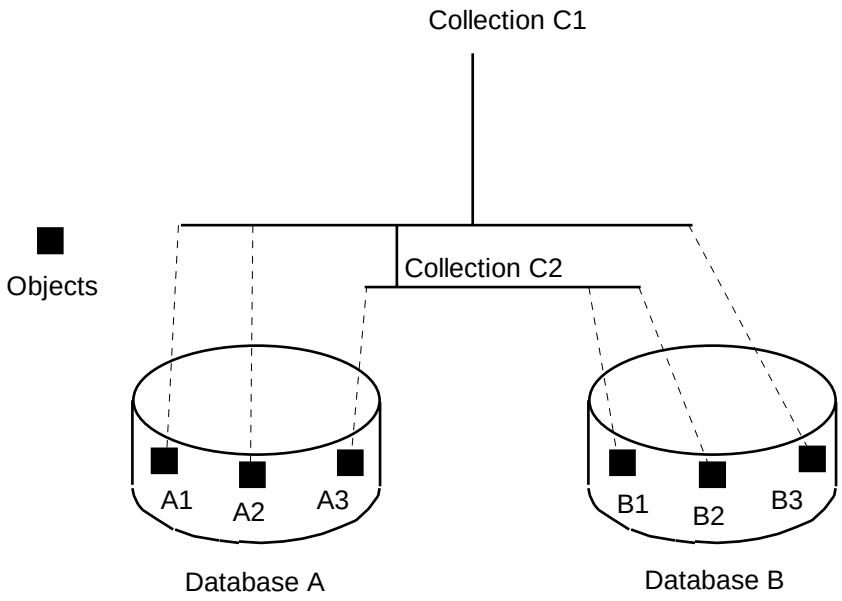

Fig. 2.4 Collections and Databases2.4CollectionsandDatabases2.4CollectionsandDatabases2.4CollectionsandDatabases

Therefore, the property of belonging to a collection differs from the property of belonging to a database, and the Z39.50 Profile for Access to Digital Collections makes some basic statements concerning this:

> *A database resides on a single server, while a collection may span servers (that is, different databases corresponding to a single collection may reside on different servers).*
>
> *A single database may include records corresponding to objects in more than one collection.*
>
> *A single object might belong to more than one collection, while the database record for that object may be retrievable from only one database.*
>
> *An object might belong to a single collection, while the database record for that object is retrievable from multiple databases, which may be on different servers.*

2.3.2 Collections and Records

The Z39.50 Profile for Access to Digital Collections makes significant attempts to clarify the usage of different record types in the context of databases and collections.

The Z39.50 standard models three types of record:

- database records, which are the unit of information in the database;
- abstract database records; the representation of the database record to the client; and
- retrieval records; the records exported to the client.

The Z39.50 Profile for Access to Digital Collections employs the concept of a record according to the Z39.50 model, thus the term record is used to mean either of these three types in their respective contexts.

However, two other fundamental types of record terminology are defined, namely the Collection Descriptive Record and the Object Descriptive Record, and these two are referenced to the three record models in Z39.50 in the context of either searching, defining the Descriptive Record or retrieval.

2.3.3 Describing the Collection Data

The description of the collection data is used for the navigation through the collection. The descriptive information is also arranged in a hierarchy. As stated above, the collection can be viewed as a tree of leaves and nodes, but an object may be a member of more than one collection and therefore have multiple parents.

Logically the descriptive information is separated from the data content, whereas the profile does not address the mechanism of how this is physically achieved. The profile terms the descriptive information as “Associated Descriptions”, which are passed from the server to the client in order to assist the user in navigation around the collection.

However, the Profile does require that a server is able to provide Descriptive Records, for either collections and/or objects.

The profile makes no attempt to define the structure of contents or to impose any restriction on the data content. It is in this area that implementers are encouraged to develop companion profiles in a particular domain that is appropriate to that domain. Examples of profiles developed so far are the Z39.50 Profile for Access to Digital Library Objects and the CIMI Profile, (Z39.50 Application Profile for Cultural Heritage Information).

In order for a library/museum to utilise this, it could arrange its collections in a loose hierarchical structure. This should not be too difficult as most libraries/museums physically do this with their books/objects anyway. The next stage would be to generate the metadata for the collection.

The content of a Collection Descriptive Record may be an overall description of a collection as well as collective or individual descriptions of some or all of the objects in the collection. An example Collection Descriptive Record is given in Section 2.3.3.1 below.

The contents of an Object Descriptive Record describe an object in the collection, and the object may or may not have an actual physical object. An example Object Descriptive Record is given in Section 2.3.3.2 below.

Both a Collection Descriptive Record and Object Descriptive Record may list parent-, context- and related collections. A Collection Descriptive Record may list child collections as well.

All this requires that a software system should enable the metadata elements to be entered into the computer system, together with the appropriate hierarchy as further discussed in Section 2.6.3.

2.3.3.1 Example Collection Descriptive Record

Referring to the Z39.50 Profile for Access to Digital Collections, sections 4.3, 4.4 and 4.5 of that document define a schematic representation of a descriptive record with the tag values. Therefore, an example Collection Descriptive Record may include:

Schematically:

```
type of description record     collection description record
brief description              A collection of Roman coins
collection info:
     collection name:          roman coins
     database:      server     z3950.myserver.co.uk,
                    db_name    rcoins,
```

If this information is now mapped to a GRS.1 record, we would see the following in the record. The tag pairs are described as (tag set, tag type) where tag set for the Z39.50 Profile for Access to Digital Collections is the number 4.

NOTE that the specific elements of a record are defined in Sections 4.6.1, 4.6.2 and 4.6.3 of the Z39.50 Profile for Access to Digital Collections according to whether the record is a Brief, Navigation or Full record.

GRS.1:

(4,1)	Integer value	this tag defines the type of descriptive record, where value 1 is defined for a Collection Descriptive Record.
(4,2)	string	the string will be a brief text description of the record.
(4,3)	pointer	implies that the following tags will contain the collection Information.
(4,7)	string	the collection name as a string
(4,8)	pointer	implies that the following tags contain the database information
(4,24)	string	the address of the server for the client to search on
(4,25)	string	the name of the database for the client to search on

It must be stated that for a library or museum software package, that this type of information would be part of the system configuration and a form-like interface would probably be provided for the entry of such data.

2.3.3.2 Example Object Descriptive Record

Referring to the Z39.50 Profile for Access to Digital Collections, sections 4.3, 4.4 and 4.5 of that document define a schematic representation of a descriptive record with the tag values. Therefore, an example Object Descriptive Record may include:

Schematically:

```
type of description record    object description record
brief description            A roman Denarius of emperor
                              Gordianus (238 AD)
object info:
     type of object:          2
                              (object is a digital object for
                              which there is a physical object)
     category of object:      Denarius (silver)
     digital object:
          brief description of variant: reverse of coin
          actual DO           GIF file of reverse of coin
     digital object:
          brief description of variant: obverse of coin
          actual DO           GIF file of obverse of coin
```

If this information is now mapped to a GRS.1 record, we would see the following in the record. The tag pairs are described as (tag set, tag type) where tag set for the Z39.50 Profile for Access to Digital Collections is the number 4.

NOTE that the digital object datatype is repeated, which is allowed on certain record elements.

Note also that the specific elements of a record are defined in Sections 4.6.1, 4.6.2 and 4.6.3 of the Z39.50 Profile for Access to Digital Collections according to whether the record is a Brief, Navigation or Full record.

GRS.1:

(4,1)	Integer	this tag defines the type of descriptive record, where value 1 is defined for a Collection Descriptive Record.
(4,2)	string	the string will be a brief text description of the record.
(4,4)	pointer	implies that the following tags will contain the object Information (first variant).
(4,12)	Integer	the type of object where the value 2 is defined as the object being a digital object for which there is a physical object.
(4,13)	string	category of object
(4,14)	pointer	implies that the following tags contain a variant of the digital object.
(4,2)	string	the string will be a brief text description of the specific variant of this actual digital object
(4,29)	octet string	actual octet data of the digital object
(4,4)	pointer	implies that the following tags will contain the object Information (second variant).

(4,12)	Integer	the type of object where the value 2 is defined as the object being a digital object for which there is a physical object.
(4,13)	string	category of object
(4,14)	pointer	implies that the following tags contain a variant of the digital object.
(4,2)	string	the string will be a brief text description of the specific variant of this actual digital object
(4,29)	octet string	actual octet data of the digital object

2.4 Navigation in a Digital Collection

When a client connects to a server that provides access to digital collections the Z39.50 Profile for Access to Digital Collections proposes the following scenarios for navigation:

- the user knows the name of the collection and the database name where the Collection Descriptive Record resides. This approach may be provided by the information provider publicising the information or perhaps is inherent if the information provider uses (say) a WWW gateway to provide access to the collection.

- the user knows the name of the collection but not the name of the database, and may use the Z39.50 Explain Facility to identify this; and

- the client uses the Z39.50 Explain Facility to identify the databases on the server and then searches for Collection Descriptive Records in order to provide information about the collections on the server.

Once the Collection Descriptive Records are retrieved, the content may be used to display to the user how to commence navigation through the collection(s). By retrieving the Collection Descriptive Records during each navigational step, the collection relationships and data content can be learnt and provided to the user.

During the navigation process, the user may issue searches to find out about the objects in the collection. The information is retrieved from the Object Descriptive Records.

2.4.1 Searching with the Z39.50 Profile for Access to Digital Collections

The Collection Descriptive Records are of prime importance for the purposes of navigation in a collection. In order to distinguish searching for collection description information as opposed to bibliographical searching, an attribute set is defined for searching database records whose structure is as defined by the Descriptive Record schema.

This attribute set is entitled “Collection-1” and is uniquely assigned an object identification code in the family of attribute sets. The attributes of “Collection-1” provide the means for

the client to interrogate the server about the server collection structure, the objects and the record types. Two attribute relationships are also defined, these being called “equal” and “always-matches”.

Therefore, a client may focus specifically on searching for collection information, as per the following example:

search in the Collection Descriptive Records for a Collection called “paintings”

This will translate to two operands separated by the Boolean AND operator:

(USE=2 , Term = 1) AND (USE=7, Relation = 1, Term = “paintings”)

where USE=2 is the representation of record-type, and term value of 1 is the representation that the record is a Collection Descriptive record.

It should be noted that if an application domain requires searching to incorporate other attribute sets, then this needs to be specifically addressed by a companion profile and it is outside the scope of the Profile itself.

An example of this is the CIMI Profile, which not only utilises the bibliographical attribute set bib-1, but also defines its own specific attribute set that is oriented towards museum type information.

2.4.2 Retrieval with the Z39.50 Profile for Access to Digital Collections

Significant detail is provided in the Z39.50 Profile for Access to Digital Collections for record retrieval. Of significant importance are the definitions of the record schemas, including the record elements and the data-types for these elements. The profile also specifies whether record elements are mandatory or optional and whether they are repeatable or not.

All records are labelled as Descriptive Records, and fall into three types, namely:

Collection *Descriptive Record*
Object *Descriptive Record*
Unspecified

The type of descriptive record element is the only mandatory record element, whereas all other elements are optional. Certain elements will only occur depending upon the type of descriptive record, and the profile carefully describes this information for the benefit of anyone implementing the profile.

The description elements are grouped into structures called data type definitions. There are eleven defined data-type definitions for the elements of a Descriptive Record. Each of these are of a structured nature and are clearly defined in the profile for benefit of anyone implementing this profile to adhere to.

For example, there are data type definitions for the Object Information, a Related Collection, a Server and Database description and a Digital Object.

Each of the data-types resolves down to the basic data-types of the abstract syntax notation of the Z39.50 protocol, namely InternationalString, BOOLEAN, OctetString and OBJECT IDENTIFIER Code (OID) which allows the record structure to map to a formally defined record syntax.

The most suitable record syntax for this type of information is the Generic Record Syntax-1, (GRS.1) which is defined in the ANSI Z39.50-1995 standard. This record structure allows for highly structured information to be presented in which the data elements may be hierarchically arranged.

The identification of the data-type definitions is through use of element tagging, and a purpose designed tag set has been developed that is called the Descriptive Record Tag Set. This Tag Set has been formally assigned an object identifier code in the profile.

However, records may not typically use this tag set alone, and there are two other tag sets defined in the Z39.50 standard called Tag Set-M and Tag Set-G which may be used in conjunction with the Descriptive Record Tag Set.

There are specific fields within the GRS.1 record in which the tag elements can be assigned. This is quite a complicated procedure and from a client point of view requires a high degree of interpretation through the records to extract the information.

2.4.3 Retrieval Record Content

There are three levels of record content defined in the Z39.50 Profile for Access to Digital Collections, namely the Brief record, the Navigation record and the Full record.

The Navigation record contains least amount of record elements, and is intended to enable the client to ascertain the collection hierarchy. No object information is provided in Navigation records.

The Brief record provides a summary of either the collection information (but not its collection relationships) or the object information.

The full record is intended to provide all available elements for all available occurrences of the elements.

2.5 Requirements of Using the Z39.50 Profile for Access to Digital Collections

The Z39.50 Profile for Access to Digital Collections defined three levels of conformance, each having specific requirements of support of the Z39.50 standard. The intention is to provide the ability of Version 2 systems to be partially conformant, through to the full level of conformance for Version 3 systems.

The three levels of conformance are:

Basic Conformance:	Version 2. requiring support for the Description Record schema and the GRS.1 record syntax.
Basic V3 Conformance:	As Basic conformance plus support for version 3
Enhanced V3 Conformance:	as Basic V3 conformance plus support for additional features.

Certain Z39.50 requirements of the Z39.50 Profile for Access to Digital Collections are mandatory for all levels of conformance. These are:

- support for the GRS.1 record syntax;
- support for Init, Search and Present services;
- support for the Collection-1 attribute set for searching;
- support for the descriptive record schema for the Present service;
- support for element set names ‘b’ and ‘navigation’.

The main inclusion to a basic V2 system here is the additional support of the GRS.1 record syntax. The other addition to a server is the inclusion of the Descriptive records, but this is a feature of the database subsystem and is outside the scope of the Z39.50 server software.

Certain other Z39.50 requirements are mandatory for enhanced conformance. These are:

- support for the CompSpec parameter with the Present service;
- support for the element specification eSpec-1 parameter with the Present Service;
- support for the variant set variant-1 parameter with the Present Service;
- support for the fragmenting service (in conjunction with the Present service);
- support for Metadata and appliedVariant sections of the GRS.1 record syntax.

The necessary protocol additions to support for enhanced conformance are the Segmentation service, eSpec-1 and CompSpec

The segmentation service allows a server to deliver a large record in fragments. This allows a client to start to process the start of the fragment data without having to wait for all the data to appear in one piece. eSpec-1 allows a client to define which record elements are returned in the records. CompSpec is an alternative mechanism in Z39.50 Version 3 of specifying the desired composition of retrieval records.

From a protocol point of view, the main requirement of a client or server system is the mandatory support of the GRS.1 record. From a client point of view this ideally requires support for understanding the Tag-Sets G and M, although these are not mandatory for a server. In addition to this, the client and server have to support the descriptive record schema.

2.6 Context of the Z39.50 Profile for Access to Digital Collections

The concept of searching on collections has been extensively researched and discussed with respect to Z39.50. Efforts were made to co-ordinate these activities in which the resulting specifications form an umbrella below which the searching and navigation of collections becomes co-ordinated.

As mentioned in the Introduction above, the Z39.50 Profile for Access to Digital Collections scope was narrowed from earlier drafts to focus on Navigation within collections.

The specific searching within collections is thereafter the subject of the companion profiles, of which two are formally available, namely:

- the Z39.50 Profile for Access to Digital Library Objects being developed at the Library of Congress, and

- the CIMI Profile (Z39.50 Application Profile for Cultural Heritage Information) managed by Consortium for Interchange of Museum Information.

Of significance and directly related to collections are the activities of the European Commission's Joint Research Centre who have designed and developed a distributed collection navigation and search system that is specified in the Catalogue Interoperability Profile, (CIP). However, the CIP profile does not fall into the umbrella of the other digital collection profiles as the mechanics of searching and navigation are different and are directly related to the architecture of the implementation.

Of the two other profiles that exist under the collections umbrella, they use elements of the Z39.50 Profile for Access to Digital Collections as a basis of their own specific technical domain. This is achieved by inheriting parts of the search and retrieval specifications and enhancing these for their own domain.

This approach gives rise to a hierarchy of profiles, and in the case of record retrieval, the record elements can be inherited from Tag Set-M, Tag Set-G, the Tag Set of the Z39.50 Profile for Access to Digital Collections and the Tag Set of the specific profile.

2.6.1 The Z39.50 Profile for Access to Digital Library Objects

The Z39.50 Profile for Access to Digital Library Objects is being developed by the Library of Congress and specifies a subset of Z39.50 features to support functional and user requirements for search and retrieval of information in digital library collections, specifically the Library of Congress digital library collections and similar collections.

The profile follows the philosophy of the Z39.50 Profile for Access to Digital Collections being the basis of modelling collections and navigation. It takes the general model for a digital object from the Z39.50 Profile for Access to Digital Collections and extends this to account for an object having constituent parts, having several images (or texts) and having its own specific structure.

The Z39.50 Profile for Access to Digital Library Objects also defines the following categories of digital objects as either Language-based, Image-based, Sound-based or Motion-based. The associate description is categorised as having either a Cataloguing Record (e.g. a MARC record, or a record based on Dublin core elements), an Archival Register, a Header or a Web page.

The Z39.50 Profile for Access to Digital Library Objects extends the retrieval mechanics and has its own tag set. In this context, the primary component of a database schema is an abstract record structure which lists schema elements in terms of their tag paths. A tag path is a representation of the hierarchical path of an element, which is expressed as a sequence of nodes each represented by a tag. The Z39.50 standard describes two such sets of tags, called Tag Sets M and G. The extended retrieval mechanics in the Digital Libraries Profile has its own set of tags that inherits Tag Sets M and G rather than repeating tags defined therein.

From an implementation point of view, whereas the Library of Congress makes its collections available directly via the Web, the Z39.50 access is not provided yet through the use of this profile. No other implementations of the Z39.50 Profile for Access to Digital Library Objects are known yet.

2.6.2 The CIMI Profile

The CIMI Profile was originally developed for a sponsored demonstration project called CHIO (Cultural Heritage Information On-line). CIMI (The Consortium for Interchange of Museum Information) comprises a number of participants from the Museum sector and also from academic and commercial organisations who have Z39.50 development experiences.

Whereas the CHIO demonstration project focused a lot of attention on information held in SGML format, a more recent test-bed project (May - October 1997) gave the CIMI profile an overhaul and the invited test-bed participants have developed clients and servers on actual museum collections across the World.

The CIMI profile does not use many features of Z39.50 Version 3, and implementations are really extensions of Version 2 of Z39.50 through use of the GRS.1 record syntax.

The CIMI profile also has its own attribute set for searching and its own element set for record retrieval. The CIMI record profile inherits parts of Tag Set-G, Tag Set-M and the collection description elements from the Z39.50 Profile for Access to Digital Collections. The CIMI profile also inherits the concept of the object model from the Z39.50 Profile for Access to Digital Library Objects, specifically in the area of different object renditions and object components.

The CIMI Profile is probably the most widely implemented aspect of the Z39.50 Profile for Access to Digital Collections from the CIMI test-bed project and the results of this project should be publicised in 1998.

2.6.3 Using the CIMI Profile

Whilst it is the responsibility of a vendor to develop a software system that conforms to the CIMI profile, such a system will probably also contain a data entry and data management system in order for the data to be entered to the database. In order to create the data, the library/organisation will need to ensure that the information about the data (i.e.; the metadata) is generated and maintained on the database.

For example, in a museum a collection of objects is likely to be catalogued in line with the standards that the museum adopts, (for example SPECTRUM standards in the UK). An object such as a piece of art would have information recorded such as the creator, date of creation, information about the creator, (name, date of birth, date of death etc.) and then would have information such as when the museum acquired the object, who it was acquired from, copyright information etc. There will also be the option of managing the digitisation of the object, different renditions of photographs, and the creation/storage of different image resolutions.

It is likely that of the large range of metadata elements the CIMI Profile offers, there will not be information about each data element. For example, the date of creation of a particular object may not be known. In this sense, most of the CIMI elements are optional.

2.7 Current Usage of the Z39.50 Profile for Access to Digital Collections

As mentioned above, the CIMI Profile is the most widely implemented use of the Z39.50 Profile for Access to Digital Collections, albeit a subsection of the profile. However, this is only a handful of implementers in Europe, the Far East and in the USA.

However, from a European point of view, participants in the CIMI test bed project are also active participants in the Aquarelle cultural heritage project, into which the CIMI Profile (and hence parts of the Z39.50 Profile for Access to Digital Collections) will be placed. Other participants in the CIMI test-bed project are commercial vendors of library and museum database systems and so there may be a gradual spread of this technology through commercial marketing.

The only drawback to this is that the lack of implementation may result in changes to the specifications when actual implementations are developed. The CIMI test-bed interoperability agreement has changed several times during the test-bed project time-scales. This has been due to the actual need for changes to suit technical requirements and also from other current related aspects in the information retrieval arena such as the requirement to incorporate Dublin core metadata elements.

These types of changes do have a knock-on effect for implementers and it can be difficult and expensive for commercial organisations to develop and market products based upon these profile specifications if the specifications are subject to change.

From a future point of view, full implementations of the Z39.50 Profile for Access to Digital Collections will probably increase when Z39.50 Version 3 systems are more common.

2.8 Benefits of using the Z39.50 Profile for Access to Digital Collections

The efforts of design that have been put into the Z39.50 Profile for Access to Digital Collections have come from experts in the fields of Z39.50 implementers, consultants and library management specialists.

The benefits of using this profile enable software implementations to be developed that benefit users in the context of distributed search and retrieval.

The current trend for developments is to focus on distributed searching, whether in a single domain or across multiple domains. The Z39.50 Profile for Access to Digital Collections provides the opportunity for systems to be developed which allow the user to rapidly locate and focus on the area of interest.

From an implementation point of view, commercial vendors may provide systems that are compliant with the Z39.50 Profile for Access to Digital Collections or companion profiles. Alternatively, there are software tool-kits available which enable client and server applications to be developed, referenced in Appendix 2.

From a client point of view, the WWW gateway approach seems to be the preferred technical solution rather than a bespoke client application program, as depicted in Figure 2.5 below.

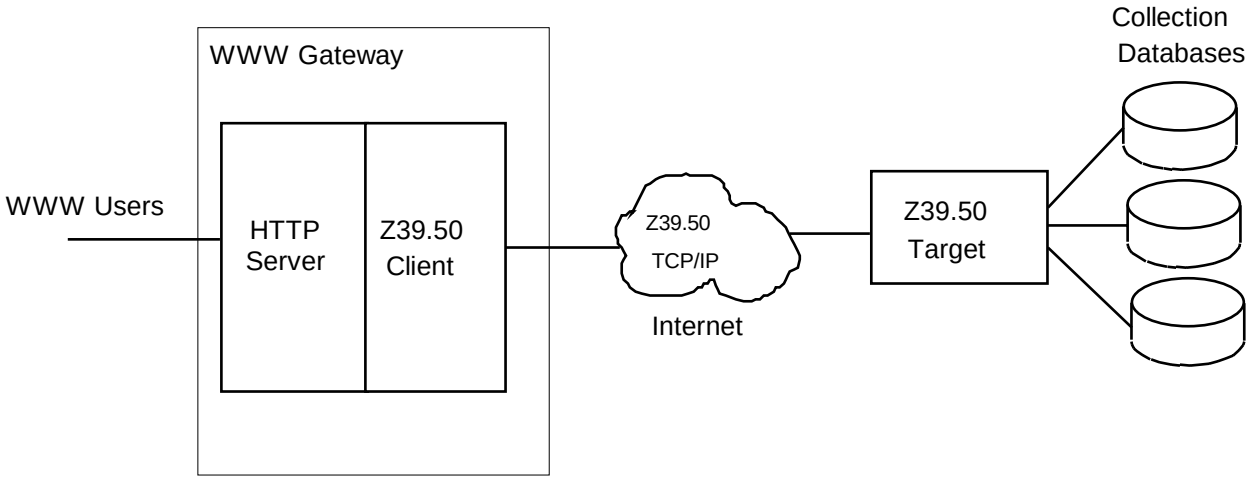

Fig. 2.5 Possible Z39.50 System Architecture2.5PossibleZ39.50SystemArchitecture2.5PossibleZ39.50SystemArchitecture2.5PossibleZ39.50SystemArchitecture

2.9 Benefit to eLib Clumps Projects

The Z39.50 Profile for Access to Digital Collections would be beneficial to the eLib Clumps projects, particularly if these gateways are accessible from the outside by the Z39.50 protocol. A Z39.50 gateway to a clump of servers could consider the services as a collection in the physical sense, i.e., it could describe a hierarchy of the underlying databases (whether a flat or tiered hierarchy). The services could also be described in the sense of the data content type, although this may require categorisation of the underlying databases as most of the institute servers probably offer all their records in a single database.

In this context, the gateway nodes could also make use of Collection Descriptive Records within the gateway programs, where it would be necessary to get a general consensus agreement from participants to which elements of metadata will be provided.

2.10 Practical Experience of Implementing CIMI

The CIMI Interoperability Test-bed which ran in 1997 for a period of 6 months intensive development effort and about 3 months of Profile finalisation provided a wealth of practical experience in implementing software systems of this nature.

The implementers were either starting from scratch; were extending existing commercial packages or were extending available freeware packages. A CIMI toolkit was also made available for use by other implementers.

There were two sides to the implementation, the software development for a client or server and the preparation of data if not already available.

For the software development, there was quite a large degree of effort required in preparation of the CIMI records, although this was increased because the specification was revised several times during the test-bed period.

For the preparation of data, the effort was quite high in preparation of suitable metadata and in preparing images and the different renditions and resolutions defined in the standard.

Practical experience shows that the amount of implementation effort should not be underestimated together with the time-scales involved. Such developments typically require third party involvement, i.e., identifying and testing against other servers, testing with other clients in order to ascertain the best degree of interoperability. In addition, a significant factor in such development is that the Z39.50 profiles are not finalised and this incurs a risk to such developments as the profiles may change during or after the implementation.

3. THE Z39.50 EXPLAIN FACILITY

3.1 Introduction to the Z39.50 Explain Facility

The Z39.50 Explain Facility was introduced with Version 3 of the Z39.50 protocol, (Z39.50-1995). This facility comprises a protocol mechanism whereby the client is able to interrogate and retrieve information about the server and the databases on the server. The interrogation mechanism is achieved through a specific use of the Z39.50 Search service. The information is retrieved with the Z39.50 Present service using purpose-designed structured retrieval records.

The Explain facility requires that the Explain data is stored in a database which is accessed using a database name defined in the Z39.50 standard as *IR-Explain-1*. The client queries the server using the Explain attribute set called *Exp.-1*. The Explain database may be either:

a) an adaptation of the existing bibliographical database management system; (if an appropriate Explain database schema can be defined within that system); or,

b) may be a specific bespoke system that is integrated with the server.

This is shown in Figure 3.1 below.

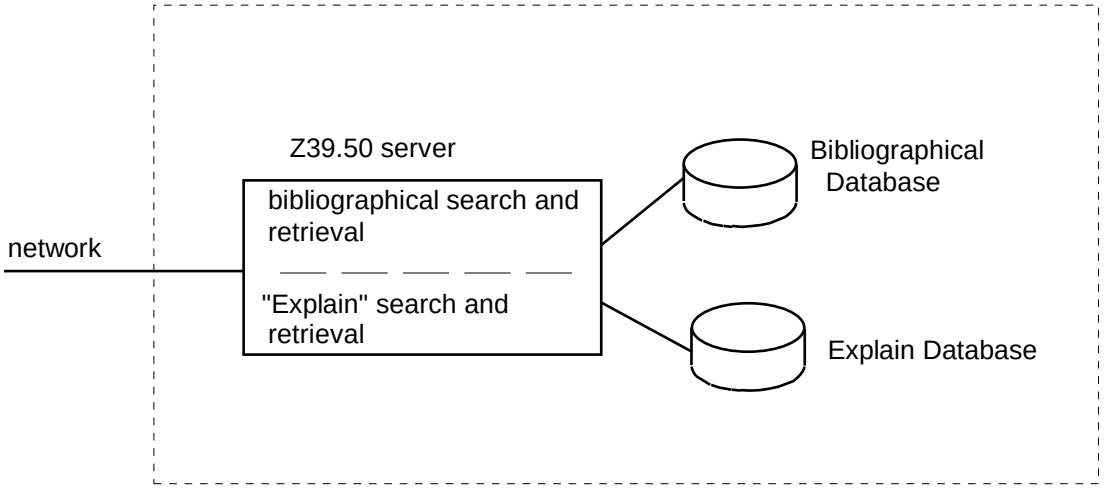

Fig. 3.1Explain and Bibliographical Databases3.1ExplainandBibliographicalDatabases3.1ExplainandBibliographicalDatabases3.1ExplainandBibliographicalDatabases

3.2 Purpose of the Explain Facility

The Explain facility allows the client to obtain details of the server implementation, including the databases available for searching, search attribute sets, diagnostic sets, and supported record syntax. This data can be considered as metadata about the server and the services that the server offers.

Some of the Explain data is specifically structured to permit a client to undertake some dynamic self-configuration based upon the content of the Explain data.

An example of structured data is the description of record retrieval formats that the server is able to deliver, (e.g., USMARC, UKMARC etc). Without Explain, a client has to effectively guess a record syntax to use unless its client knows a-priori what record syntaxes a server supports. Using the Explain facility the client can find out exactly which record syntaxes are supported by the server. The record syntaxes are represented in the Explain database as object identifier codes which can be recognised by a computer program.

Other parts of the Explain data are purely of an unstructured text format and its content may be typically presented at the client for user information purposes.

An example of unstructured data is contact information, e.g., a name and address and telephone number of a contact person responsible for the server.

The Explain data is partitioned into the following categories:

Server information:	general information about the server, the organisation who runs the server and administrative information;
database information:	information about each database on the server such as date last updated, the database name, a description of the database, the availability of the database and news about the database;
schema information:	descriptive information about a database schema in both object identifier code form and human readable text form;
tag-set information:	information about the tagging of record elements in the retrieved records from the database;
record syntax information:	descriptive information about the supported record syntaxes on the server in both object identifier code form and human readable text form;
attribute set information:	descriptive information about the supported search attribute sets on the server in both object identifier code form and human readable text form;
attribute details:	details about each supported search attribute including the attribute combinations in both code form and human readable text form;
term-list information:	details about each term list associated with a database in both code form and human readable text form;
extended services information	descriptive information about the supported extended services in both in both code form and human readable text form;

term-list details:	provides descriptive information for a term list, including a list of sample terms;
element-set details:	provides descriptive information about an element set in a form which can be processed by the client and also some human readable text information;
retrieval-record details:	provides descriptive information about the elements of a retrieval record in a form which can be processed by the client and also some human readable text information;
sort details:	provides a description of the sorting capabilities supported by the server in a form which can be processed by the client and also some human readable text information;
processing information:	provides instructions representing how the server believes the data should be processed by the client for presentation to the user;
variant-set information:	provides descriptive information about a variant set definition supported by the server;
unit information:	provides descriptive information about a unit system definition supported by the server;
category list	a list of the Explain categories that are supported by the server. There is only one such record per Explain database containing information in both code form and human readable text form.

3.3 Current Usage of the Explain Service

The Explain facility is a complicated sub-system to implement and it does not have a wide take-up in the Z39.50 community. This is partially due to the lower number of Z39.50-1995 servers, and the fact that suppliers of Z39.50 compliant library systems typically provide a client that is pre-programmed to know the details of other servers that the users may connect to; thereby excluding the need for Explain.

The real benefit of the Explain facility is to enable different client and server systems to inter-operate without the need for trial and error configuration. The development of dynamic configuration in a client is one of the most ambitious aspects of Explain, but there are very few implementations of this nature.

Two of the first servers that supported Explain were those of the Silverplatter Corporation in Boston, USA and that of the AT&T library server (now Lucent Technologies).

In Europe, the European Commission funded ONE (OPAC Network in Europe) project has resulted in several Explain servers being developed. These are scheduled for public access in early 1998, when the project has finished, although they are being tested in the latter half of 1997 within the ONE project.

Also in Europe, the EC Centre for Earth Observation, Ispra, Italy has funded the Catalogue Interoperability Profile project (CIP) in which Explain is extensively utilised in both server and client applications.

This latter project is of specific interest since the CIP project profile is considered a companion profile to the Z39.50 Profile for Access to Digital Collections, utilising very complex distributed searching over wide area domains.

Very few client implementations exist that support Explain. Two examples are the client developed by the Ameritech Corporation and the client produced in the ONE project - “ICONE” exist.

3.3.1 Explain “Test-bed”

During 1995 to the beginning of 1997, an Explain test-bed group was established comprising members of the Z39.50 community. The purpose of the test-bed was for interoperability testing and to report to the Z39.50 community of implementation experience. Whilst the test-bed does still formally exist, there is very little activity reported by the participants.

3.3.2 Example Explain Data

The following is an extract of the target information stored on the British Library Z39.50 server developed in the ONE project. This is the information about the target as extracted by the specific client used:

```
Welcome to the British Library's Z3950 target.

Service Information - General Details:
    Date that this information was Added: 24/02/1997
    Last Date that any information Changed: 24/02/1997
    Date of Expiry: 01/01/2000
    This information is provided in: English

Target Name: LIB

Recent News:
 This is the British Library's experimental Z3950 server developed as part of
the Library's involvement in the project OPAC Network in Europe (ONE)

Target Details:
Support of Named Result Sets: Yes
Support of Multiple Database Searching: No
Maximum Number of items in a Result Set:   100000
Maximum Number of Search Terms in a Query:   20
```

For the subset of the Science and reference database, the database information record includes the following:

```
Service Information - General Details:
    Date that this information was Added: 24/02/1997
    Last Date that any information Changed: 24/02/1997
    Date of Expiry: 01/01/2000
    This information is provided in: English
Database Name: SRI
```

```
Usage Fee Required: No
Availability Status: Yes
```

Other Z39.50 servers that were developed in the ONE project which include Explain support are:

LIBRIS – Royal Library of Sweden;
PICA;
University of Oslo Library;
Danish Library Centre;
Steirmarkische Landesmuseum, Graz,
Steirmarkische Landesbibliothek, Graz,

3.4 Usage of Explain with respect to Collections

The use of the Z39.50 Explain facility is optional with respect to the Z39.50 Profile for Access to Digital Collections. Servers that support Explain may wish to distinguish one or more databases as those which provide access to digital collections (i.e. databases which include Descriptive Records). The Z39.50 Profile for Access to Digital Collections recommends that servers assign to those databases the searchable keyword 'collection-descriptive-records', searchable via the Explain “keyword” search attribute.

A Server using this feature may apply whatever criteria it deems appropriate in selecting databases to assign this keyword. For example a server may select only databases that include Collection Descriptive Records (as opposed to Object Descriptive Records), or only databases that include Collection Descriptive Records for context collections.

The user may know the name of a collection but not the database where its Collection Descriptive Record resides. In that case, if the client and server both support Explain then the client may attempt to determine that database via Explain. It may be that neither the client nor the user knows any collection names. In that case if the client and server both support Explain, the client might attempt to learn which databases in general correspond to collections and search those databases for Collection Descriptive Records. The client may then retrieve Collection Descriptive Records from these databases and display summary information to the user, including brief descriptions.

However, the use of the Explain facility is optional for the Z39.50 Profile for Access to Digital Collections, and also for the companion profiles developed so far. A server that supports Explain may wish to distinguish databases that provide access to digital collections, (i.e., those which contain Descriptive Records).

Once a client has established which databases on a server provide access to digital collections, the client can process information from other textual or structured data in the Explain records. The textual information (e.g., descriptions of the collection) can be displayed to the user, and the structured information (e.g., the database name) can be used by the client in searching.

The use of Explain and Collection Descriptive Records offers a powerful mechanism for navigation across collections both intra and inter-server. In this context, the servers that support Explain may wish to distinguish one or more databases as those which provide access to digital collections, i.e.; those which provide descriptive records. It is recommended in the Z39.50 Profile for Access to Digital Collections that servers assign to those databases the searchable keyword *"collection-description-records"* which can be searched using the Explain Use attribute "Keyword".

Having established the databases that support collections; the Collection Descriptive Records can be retrieved, and these may provide pointers to associated collections on the same server or on other servers. This offers the ability for a client software application to dynamically discover a set of servers that are part of a collection of servers, i.e.; dynamic clump discovery. This is indicated in Figure 3.2 below.

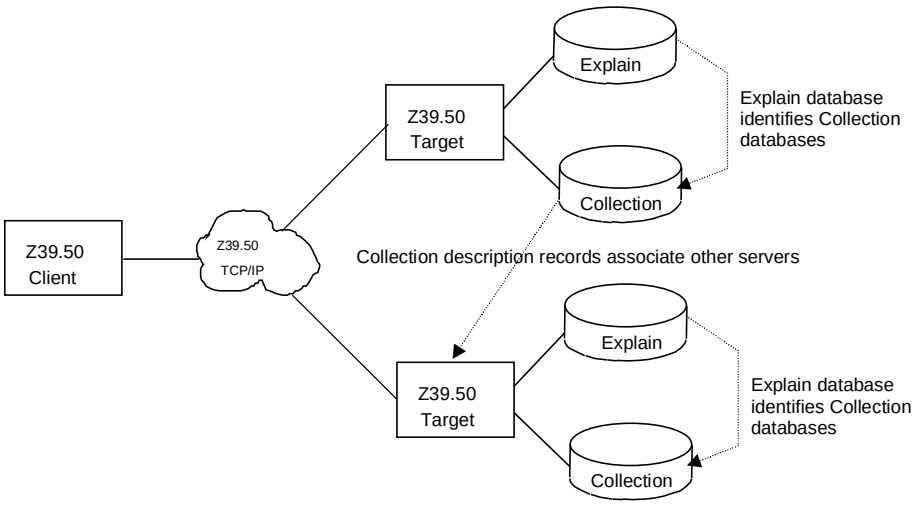

Fig. 3.2 Using Explain and Collection Descriptions3.2UsingExplainandCollectionDescriptions3.2UsingExplainandCollectionDescriptions3.2UsingExplainandCollectionDescriptions

3.5 Benefits of Using the Explain Facility

The benefits of using the Explain facility are to assist in the interoperability of a client and server Z39.50 session. The purpose is to enable the client to interrogate the server Explain database and to configure itself to what the server can provide or support.

Without Explain, the necessary information has to be provided by some other means so that a client is pre-configured when a user selects to query the server. This is especially useful if the server does not support some or all of the suggested default parameters in the Z39.50 standard.

In addition, if the server changes its characteristics in any way, for example , provision of a different attribute, then the client will be mis-aligned with the server and without using

Explain would need to have its pre-configuration parameters changed.

3.6 Requirements of using the Explain Facility

A number of requirements are necessary to use the Explain facility. For the client, support for searching with the Explain attribute set EXP-1 is mandatory.

In addition, the support for the Explain record syntax is also necessary, which comprises a considerable amount of protocol software to be supported.

The client application is therefore required to issue queries and retrieve the Explain records and to interpret the results and to configure itself as appropriate.

For the server, the Explain database called IR-Explain-1 is mandatory, and the ability to receive searches and requests for the Explain records is necessary.

From an administration point of view, the content of the Explain database needs to be kept up to date, such as dates of database update, additional databases added, contact details etc.

3.7 Profiling Explain

The Explain service offers a powerful mechanism for a client to discover information about a server. One drawback is that many Explain categories exist and many of the Explain record parameters are optional.

This can give rise to a high degree of inconsistency across servers with respect to the degree of Explain information that may be held.

In a particular scenario such as a multinational project or (say) the eLib clump projects, it may be useful to agree to a minimum subset of the Explain data parameters that a server Explain database will retain.

This approach could be considered as profiling Explain in which the interested parties would have their Explain databases readily published in a profile document.

4 References and Bibliography

The following documents are also relevant and applicable to the contents of this document. They are referenced by their appropriate URLs.

1. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, (July 1995)
(http://lcweb.loc.gov/z3950/agency/1995doce.html)

2. Z39.50 Profile for Access to Digital Collections (Draft 7), May 3, 1996
(http://lcweb.loc.gov/z3950/agency/profiles/collections.html)

3. Z39.50 Profile for Access to Digital Library Objects, (Draft 4).
(http://lcweb.loc.gov/z3950/agency/profiles/dl.html)

4. CIMI Z39.50 Profile
(http://www.cimi.org/products/cimi_products.html#THREE)

5. Distributed Searching of Museum and Bibliographic Information:
An Update on CIMI's International Z39.50 Interoperability Testbed
(http://www.cimi.org/documents/z_testbed_update.html)

Appendix 1 List of Abbreviations and Definitions

The following abbreviations are used in this document:

ANSI	American National Standards Institute
CIMI	Consortium for the Interchange of Museum Information
CIP	Catalogue Interoperability Profile
GRS.1	Generic Record Syntax, 1
OID	Object Identifier Code - a code devised from standards bodies to enable computer programs to recognise protocol components.
WWW	World Wide Web
Z39.50	ANSI Search and Retrieval Protocol

The following definitions are extracted from the Z39.50 Profile for Access to Digital Collections documentation:

Associated Description:	A unit of descriptive information associated with a collection or object, for example, an encoded archival description or a finding aid (for an archival collections), or a catalogue record (for library materials).
Child:	Collection or object A is a child of collection B if B is a parent of A. A collection may have objects as well as sub-collections as children.
Collection:	A group of related objects and/or collections, possibly distributed across locations. A collection is a tree, where leaf nodes are objects and non-leaf nodes are sub-collections. (ref. Figure 2.1).
Collection Descriptive Record:	A Descriptive Record that describes a Collection.
Companion profile:	An independently developed profile which is a compatible extension to or subset of the Profile, extending or limiting the use of the Profile to specific applications or classes of information, for example, museum objects.
Context Collection:	Collection A is a context collection for collection or object B if it is a superior, related collection, and the organisation with responsibility for the management of B considers that although collection A may be relevant to a user who is interested in B, any collection superior to collection A is likely not relevant.
Descriptive Record:	A unit of descriptive information at a higher level of abstraction than an Associated Description. A Descriptive Record may include one or more Associated Descriptions in addition to other information that describes either a collection (and possibly its contents) or an object within a collection. A Descriptive Record is either a Collection Descriptive Record or an Object Descriptive Record.

Member (of a collection):	See Child.
Object:	A Physical Object or a Digital Object. A physical object is a physical entity, e.g., a book, a manuscript, a photograph; A digital object is an electronic representation of a physical object, e.g.; a scanned image of a photograph, a digital representation of the text in a manuscript or book.
Object Descriptive Record:	A Descriptive Record that describes an object.
Parent Collection:	Collection A is a parent of collection or object B if A is immediately superior to B.
Related Collection:	Collection A related to collection or object B if the organisation with responsibility for the management of B considers that collection A may be relevant to a user who is interested B.
Sub-collection:	Child collection.
Subordinate:	Collection or object A is subordinate to collection B if A is a node on a tree whose root is B.
Superior Collection:	Collection A is superior to collection or object B if B is a node on a tree whose root is A.
Tag Set:	A defined set of identifier codes that are used in record retrieval to indicate the meanings of a corresponding data elements or data types.

Appendix 2. Developing Software for the Z39.50 Profile for Access to Digital Collections

There is no reason nowadays to develop a Z39.50 application from scratch as there are suitable toolkits available both as freeware or commercial which are very cost effective. Anyone interested in using the freeware packages should carefully read the licences for use.

Suitable Z39.50 toolkits include:

- DBV OSI II (ZedKit) Z39.50 software development kit for C, C++; (UNIX, Windows platforms) from Crossnet Systems Limited;

- ZedJAVA; Z39.50 Software development kit for JAVA; from Crossnet Systems Limited and DSTC Pty Ltd;

- BER Utilities (C, JAVA) from OCLC;

- YAZ, (C, C++) Z39.50 Software development kit for C, C++, (UNIX, Windows) from Index Data.

In addition, there is a CIMI specific toolkit available from System Simulation Ltd built on YAZ.

It is also expected that in 1998 the CIP Z39.50 distributed search and retrieval system which is the basis of the INFEO project will be released as shareware, comprising WWW/Z39.50 gateway software and inter-Z39.50 gateway communications in modules called retrieval managers. This work is funded by the European Commission Joint Research Centre in Italy.