

# 인터넷 익명FTP 아카이브 자원을 위한 메타데이터



웹의 이용이 급증함에도 불구하고 인터넷상에서 파일을 전송하기 위해서는 여전히 FTP(파일전송 프로토콜)를 사용하고 있다. 따라서 FTP를 통해 전송될 정보를 저장하는 FTP 아카이브는 인터넷 정보자원의 일부로서 유용한 가치를 지닌다. 특히 익명FTP(anonymous FTP)는 이용자의 개별 계정이 필요하지 않기 때문에 활용도가 매우 높다. 이러한 FTP 아카이브 자원을 위한 메타데이터의 개발은 FTP 아카이브의 효과적인 탐색과 이용을 위해 필수적이다.

이재진/ 한국데이터베이스진흥센터 정책연구과

## 연재 순서

1. 메타데이터의 개요
2. DC(Dublin Core)
3. GILS(Government Information Locator Service)
4. IAFA Templates ..... 이번호
5. MARC
6. PICS(Platform for Internet Content Selection)
7. RFC 1807
8. SOIF
9. TEI header
10. URC(Uniform Resource Characteristics)
11. Warwick Framework RDF(Resource Description Framework)
12. 메타데이터 향후 방향

## 개요

IAFA(Internet Anonymous FTP Archives) 템플릿은 IETF의 IAFA 실무 작업반(WG)에 의해 설계되고 1995년 인터넷 초안 형식으로 발행된 메타데이터이다. IAFA 템플릿은 익명FTP 아카이브에 들어 있는 다양한 범주의 정보를 FTP 아카이브의 관리자가 일관성 있게 기술할 수 있도록, 필요한 색인정보, 데이터요소를 정의하고 이를 공유할 수 있도록 하고 이를 통해 이용자가 손쉽게 정보를 찾고 접근할 수 있도록 하는 레코드 포맷이다.

IAFA 템플릿의 개발은 FTP 아카이브에 포함된 데이터나 파일에 대한 색인이 전체적으로 불가능하여 자원을 찾아내기가 어렵고 각각의 색인을 함께 공유할 수 없다는 FTP 아카이브의 문제점과 이로

인한 이용의 불편으로 이루어진 것이다.

IAFA 템플릿의 이용은 아카이브의 색인을 통해 분산되어 있는 아카이브의 내용, 서비스, 그리고 관리 데이터에 대한 정보를 탐색하고 색인하고 공유할 수 있게 한다.

템플릿은 하나 이상의 데이터요소로 구성되는 레코드의 개념으로서, IAFA 템플릿은 사이트 정보, 논리적 아카이브 정보, 자동파일생성 정보와 문서유사객체(DLO)인 문서, 이미지, 소프트웨어, 메일링리스트 아카이브, 유즈넷 아카이브, 소리, 비디오, FAQ파일 등의 템플릿 유형을 정의하고 있다. 템플릿 유형은 데이터요소를 나타내는 속성을 가지며 속성값을 통해 다양한 FTP 아카이브 자원을 기술한다.



## 클러스터 : 공통데이터요소

IAFA 템플릿에서는 사용되는 자원의 데이터요소를 구분하는데 매번 사용이 반복되는 데이터요소의 중복을 피하기 위해 이를 공통데이터요소(common data elements)로 묶은 클러스터(cluster) 개념을 사용한다.

이러한 클러스터를 템플릿의 다른 영역에서도 사용할 수 있도록 핸들(handle) 속성을 이용하는데 핸들은 일종의 클러스터의 식별기호라고 볼 수 있다. 공통데이터요소는 클러스터로서 개인 혹은 단체, 기관이 있고 자원에 대한 정보를 기술하는 공통데이터요소가 기타로 포함된다.

클러스터인 개인 혹은 단체와 기관은

〈표 1〉 개인 혹은 단체 기술을 위한 공통 데이터 요소

데이터요소명	설 명
Name	개인 이름
Work-Phone	개인의 직장전화번호
Work-Fax	개인의 직장팩스번호
Work-Postal	개인의 직장우편주소
Job-Title	개인의 직함
Department	개인이 속한 부서
Email	개인의 전자우편주소
Handle	레코드의 고유식별기호
Home-Phone	개인의 주택전화번호
Home-Postal	개인의 주택우편주소
Home-Fax	개인의 주택팩스번호
표기 : (USER*)	

〈표 2〉 클러스터 기관 : 공통 데이터 요소

데이터요소명	설 명
Organization-Name	기관 이름
Organization-Type	기관 유형(대학, 회사 등)
Organization-Postal	기관의 우편주소
Organization-City	기관의 소재 도시
Organization-State	기관의 소재 주(도)
Organization-Country	기관의 소재 국가
Organization-Email	기관의 전자우편주소
Organization-Phone	기관의 전화번호
Organization-Fax	기관의 팩스번호
Organization-Handle	레코드 고유식별기호
표기 : (ORGANIZATION*)	

IAFA 템플릿 유형으로도 사용된다. 공통데이터 요소는 데이터요소명을 표기하는 계층 구조상 최하위 수준을 나타낸다.

### ① 개인 혹은 단체 (Individuals or Groups)

특정 템플릿에서 각 개인을 기술하는 데이터요소로서 개인이 속한 기관을 표현하기 위해 기관 클러스터를 포함할 수도 있다.

템플릿에서 사용할 때에는 "Author(USER\*)" 형태로 표기한다. 〈표 1〉은 개인 혹은 단체 기술을 위한 공통데이터요소이다.

### ② 기관(Organizations)

기관을 기술하거나 개인 혹은 단체의 하부요소를

기술할 때 사용되는 공통데이터요소이다. 템플릿에서 사용할 때에는 "Publisher(ORGANIZATION\*)" 형태로 표기한다. 〈표 2〉는 기관을 기술하기 위한 공통데이터요소이다.

### ③ 자원에 대한 정보

#### (Resource Information)

개인 혹은 단체나 기관의 속성을 기술하는 데이터요소 외에 특정 자원을 기술할 때 사용하는 공통데이터요소는 일반적인 데이터요소의 하부요소로 사용된다. 이러한 자원에 대한 정보를 기술하는 데

〈표 3〉 자원에 대한 정보를 기술하는 데이터 요소

구 분	데이터요소	설 명
기타	Title	자원의 완전한 표제
	Description	자원의 기술
	Keywords	자원을 찾는 키워드
	RI	URI
	Access-Method	URI가 없는 경우의 접근 방법
	City	자원이 소재한 도시
	State	자원이 소재한 주(州)
	Country	자원이 소재한 국가
	유지·보수	Record-Last-Modified-(USER*)
Record-Last-Modified-Date		레코드 최종수정 날짜
Record-Last-Verified-(USER*)		레코드 최종검토자 정보
Record-Last-Verified-Date		레코드 최종검토 날짜

〈표 4〉 IAFA 템플릿 유형

구 분	객체 유형	설 명
내용 정보	개인 혹은 단체 정보	USER
	기관 정보	ORGANIZATION
	서비스 정보	SERVICE
	문서	DOCUMENT
	임의의 데이터집합	DATASET
	소프트웨어 패키지	SOFTWARE
	메일링 리스트 아카이브	MAILARCHIVE
	유즈넷 아카이브	USENET
	이미지 파일	IMAGE
	비디오 파일	VIDEO
	소리 파일	SOUND
	FAQ 파일	FAQ
호스트	사이트 정보	SITEINFO
	논리적 아카이브 정보	LARCHIVE
	자동파일경신 정보	MIRROR

이터요소는 어떤 템플릿에도 추가할 수 있는 것(Miscellaneous)과 레코드의 유지·보수(Maintenance)와 관련된 사항을 기술하는 요소로 구분할 수 있다. 〈표 3〉은 자원에 대한 정보를 기술하는 데이터요소를 나타낸 것이다.

### 템플릿 유형 정의

IAFA 템플릿은 유용한 네트워크 자원을 기술하기 위하여 다양한 템플릿 유형을 정의하고 있다. 템플릿 유형은 아카이브 자료의 내용 정보를 제공하는 것과 호스트와 관련된 것으로 나눌 수 있다. 〈표



## 예) DOCUMENT 레코드 기술

Template-Type:	DOCUMENT
Title:	The Function of Homeoboxes in Yeast Chromosome 1
Author-Name:	John Doe
Author-Email:	jdoe@yeast.fooobar.com
Author-Home-Phone:	+1 898 555 1212
Author-Name:	Jane Buck
Author-Email:	jane@fungus.newu.edu
Last-Revision-Date:	27 Nov 1991
Category:	Conference paper. Yeastcon, January 1992, Mushroom Rock, CA, USA
Description:	Homeoboxes have been shown to have a significant impact on the expressions of genes in Chromosome 1 of Bakers' Yeast.
Citation:	J. Doe, J. Buck, The function of homeoboxes in Yeast Chromosome 1, Conf. proc. Yeastcon, January 1992, Mushroom Rock, pp.33-50
Publication-Status:	Published Publisher-Organization-Name: Yeast-Hall Publisher-Organization-Postal: 1212 5th Avenue NY, NY, 12001
Copyright:	The copyright on this document is held by the authors. It may be freely copied and quoted as long as the contribution of the authors is acknowledged
Library-Catalog:	LCC 1701D
Keywords:	homeobox, yeast, chromosome, DNA, sequencing, yeastcon
Format-v0:	Application/PostScript
URI-v0:	ftp://ftp.fungus.newu.edu/pub/yeast/homeobox1.ps
Language-v0:	English
Size-v0:	18 pages
Format-v1:	text/plain: charset=US-ASCII
URI-v1:	ftp://ftp.fungus.newu.edu/pub/yeast/homeobox1.txt
Size-v1:	13 pages
Language-v1:	Russian

로 사용된다. 템플릿 SERVICE는 제공되는 서비스의 관리자, 기관, 요금부과 정책, 서비스 제공 시간 등을 기술한다. 템플릿 DOCUMENT, DATASET, SOFTWARE, MAILARCHIVE, USENET, IMAGE, VIDEO, SOUND, FAQ는 모두 같은 데이터요소를 갖는다.

여기에는 범주(객체의 유형), 주제, URI, 간략표제, 저자정보, 관리자정보, 소스, 이용요구사항, 설명, 서지, 인용, 발행상태, 발행자 정보, 저작권, 생성 날짜, 토론그룹 기술, 키워드, 판차사항, 형식, 크기, 언어, 문자집합, ISBN, ISSN, 최종 개정일, 도서관목록정보 등이 포함된다.

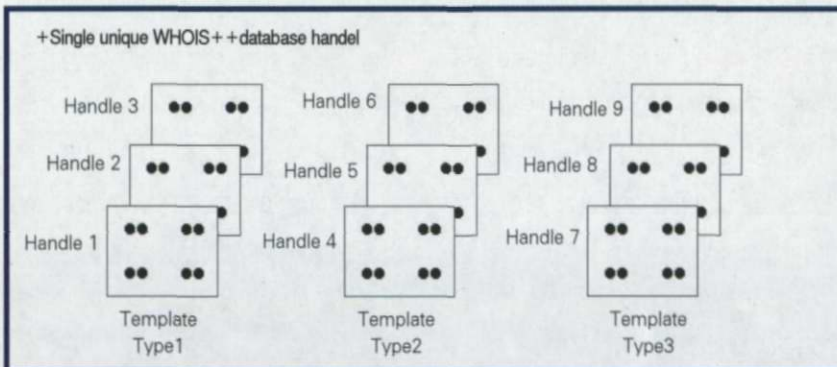
템플릿 SITEINFO는 아카이브가 들어 있는 호스트와 관련된 내용을 기술하는데 각 아카이브에 하나의 템플릿이 정의되어야 한다. 템플릿 LARCHIVE는 물리적인 사이트내에서 논리적으로 구별되는 여러 아카이브를 기술하는 것으로 이를테면 각 아카이브의 관리 정보와 자원 정보를 가리킨다. 마지막으로 템플릿 MIRROR는 자동적인 파일 갱신 정보로서 갱신 주기, 갱신 시간, 갱신 방법, 관리자, 소프트웨어 등을 기술한다.

이러한 IAFA 템플릿 유형 가운데 DOCUMENT 레코드를 기술한 예를 살펴보면 왼쪽과 같다.

IAFA 템플릿의 데이터요소에는 변형 정보(variant information)를 수용하는 경우가 있다. 변형 정보는 다른 언어, 다른 포맷, 다른 문자집합 등을 가리키며 이를 해당 데이터요소 뒤에 "-v" (\*은 일련번호를 나타내는 숫자)로 표기하게 된다.

## 프로토콜

IAFA 템플릿과 관련되는 프로토콜은



(그림 1) whois++ 데이터베이스의 구조

4)는 객체 유형에 따른 템플릿 유형을 나타낸 것으로 실제 코딩과 파일 처리시에는 영문 대문자로 표기한 템플릿 유형명을 사용한다.

템플릿 USER와 ORGANIZATION은 앞서 공통데이터요소로 설명한 요소를 갖게 되며 USER\*, 혹은 ORGANIZATION\* 형태로 다른 템플릿에서 반복적으

whois++ 프로토콜과 Z39.50 프로토콜이 대표적이다.

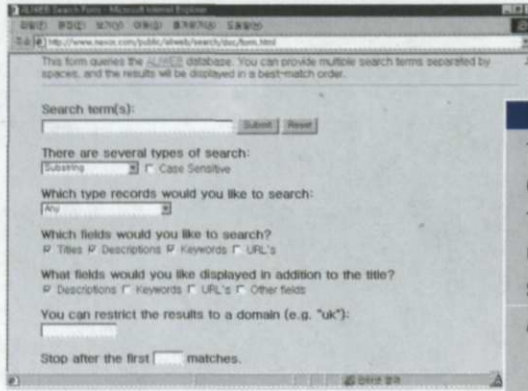
whois++은 디렉토리 서비스를 위한 프로토콜로서 1995년 RFC 1835로 정의되었다. 기존의 whois 프로토콜이 매우 단순한 데이터 모델로서 단일 데이터베이스

의 접근을 위해 설계되었던 반면, 이를 변형한 whois++ 프로토콜은 데이터 요소의 집합으로서 데이터베이스내의 구조화된 정보나 단순한 속성/값 쌍을 다루고 다양한 데이터베이스 레코드의 유형을 포괄할 수 있도록 고안된 것이다.

따라서 whois++ 프로토콜은 속성/값의 쌍을 취하고 있고 템플릿 유형, 속성, 값 혹은 핸들과 같은 접근점에 의해 제한된 탐색이 가능하다는 점에서 IAFI 템플릿과 유사하여 IAFI 템플릿과 whois++ 프로토콜을 결합한 IAFI /whois++ 템플릿이 시스템에서 많이 응용되고 있다. <그림 1>은 whois++ 데이터베이스 구조를 나타낸 것이다.

<그림 1>에서 보는 바와 같이 whois++의 전체 데이터베이스는 단일의 고유한 whois 핸들에 의해 식별되며, 각 레코드는 사용된 템플릿 유형에 의해 결정되는 단일의 고유한 핸들과 특정한 속성 집합을 가진다.

Z39.50 프로토콜은 분산 네트워크 환경에서 정보 검색을 용이하게 하기 위해 고안된 표준 정보검색 프로토콜로서 IAFI 템플릿을 Z39.50 프로토콜에서 이용할 수 있도록 매핑하는 연구가 이루어지고 있다. Z39.50 프로토콜은 정보를 표현하기 위해 6가지 속성집합을 사용한다.



<그림 2> ALIWEB의 검색 형식

이 가운데 서지 정보 속성을 표현하는 Bib-1 속성집합은 IAFI 템플릿과 가장 유사하여 IAFI 템플릿의 Bib-1 속성집합으로의 매핑이 시도되고 있다.

<표 5>는 IAFI 템플릿을 Bib-1 속성집합으로 매핑할 수 있는 요소를 비교하여 나타낸 것이다.

여기에서 Template type과 Version, Format, Size의 속성은 Bib-1 속성집합에 적절하게 매핑되는 요소가 존재하지 않는데 그 이유는 이들 요소가 탐색어로는 잘 사용되지 않기 때문이다.

### 구현 사례

IAFI 템플릿을 구현한 시스템의 대표적인 예는 ALIWEB과 ROADS 소프트웨어를 들 수 있다.

ALIWEB(Archie Like Indexing in the WEB, <http://www.nexor.com/public/aliweb/>)은 웹상에 있는 자원을 검색하기 위한 시스템으로서 IAFI /whois++ 템플릿을 응용한 최초의 검색시스템이다.

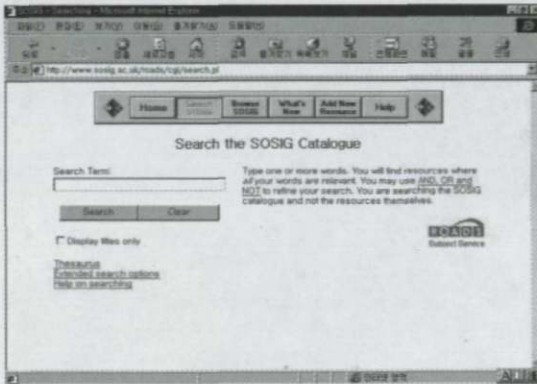
이 시스템은 NEXOR社가 운영하며

<표 5> IAFI 템플릿의 Bib-1 속성집합으로의 매핑

IAFI Template	Bib-1 Attribute Set used
Template-type	X
Category	1031 Material type
Title	4 Title
URI-v*	1032 Doc ID
Short-Title	43 Title abbreviated
Author-(USER)*	1 Name-Personal 2 Name-Corporate 3 Name-Conference 1000 Author-name-and-title 1002 Name 1003 Author-Name 1004 Author-name-personal 1005 Author-name-corporate 1006 Author Name-conference
Admin-(USER)*	1000 Author-name-and-title 1002 Name 1003 Author-Name 1004 Author-name-personal 1005 Author-name-corporate 1006 Author-Name-conference
Source	63 Note
Requirements	63 Note
Description	62 Abstract
Bibliography	63 Note
Citation	63 Note
Publication-Status	63 Note
Publisher-(ORGANIZATION)*	1018 Name-Publisher
Copyright	63 Note
Creation-Date	31 Date-Publication
Discussion	63 Note
Keywords	21 Subject
Version-v*	X
Format-v*	X
Size-v*	X
Language-v*	54 Code-Language
Character-Set-v*	63 Note
ISBN	7 Identifier-ISBN
ISSN	8 Identifier-ISSN
Last-Revision-Date-v*	1012 Date/Time last modified
Subject-Descriptor-Scheme	아래와 동일
Library Catalog-v*	13 Classification-Dewey 14 Classification-UDC 16 Classification-LC 20 Classification-Local

FTP 아카이브를 접근하는 실험적인 시스템으로서 고안된 것이다. IAFI 템플릿은





〈그림 3〉 SOSIG 서비스

〈표 6〉 ALIWEB의 태그와 속성

태그	속성
SITEINFO	URI
	Host-Name
	Admin-Handle
	Owner-Organization-Name
	Description
ORGANIZATION	Keywords
	Organization-Name
	Organization-Handle
	Organization-Phone
	Organization-Fax
	Organization-Email
	Organization-Postal
	City
	State
	Country
	URI
DOCUMENT	Description
	Keywords
	Author-Email
	Title
	URI
USER	Email
	Work-Phone
	URI
	Handle
	Title
SERVICE	Admin-handle
	Keywords
	Description
	URI
	Title

ALIWEB 검색시스템에서 필수적으로 사용될 색인 정보를 표현하는 표준 포맷으로 사용된다.

ALIWEB은 ① 웹상에서 자원 제공자가 웹상에 직접 혹은 자동 도구를 이용하여 표준 포맷인 IAFA 템플릿으로 서비스를 기술하고 ② 이 파일을

ALIWEB에 등록한다. 그리고 ③ ALIWEB은 정기적으로 이들 파일을 검색하고 탐색가능한 데이터베이스로 구축해 놓는 과정으로 운영된다. 이용자는 웹상에 있는 ALIWEB 데이터베이스를 탐색하고 검색하게 된다.

〈그림 2〉는 ALIWEB의 검색형식을 나타낸 화면이며 이를 상용 검색 시스템에서 이용할 수도 있다.(<http://www.aliweb.com>)

ALIWEB의 파일은 수많은 레코드로 구성되며 각 레코드는 태그와 속성/값의 리스트를 포함한다. 태그는 레코드 유형을 식별하여 템플릿 유형과 같은 개념으로 사용되며 ALIWEB에서는 SITEINFO, ORGANIZATION, DOCUMENT, SERVICE, USER의 5가지 태그를 사용한다. 각 태그에서 제공하는 속성과 기술 내용은 IAFA 템플릿에 기초한다. 〈표 6〉은 ALIWEB에서 사용하는 태그와 속성을 나타낸 것이다.

ROADS(Resource Organization And Discovery in Subject-based Services, <http://www.ilrt.bris.ac>.

uk/roads/) 소프트웨어는 웹상에 주제 기반 게이트웨이를 설치하고 유지할 수 있도록 하는 도구로서 여기에서 주제 기반 게이트웨이란 인터넷 기반의 자원을 탐색하고 브라우즈할 수 있는 목록을 제공하는 서비스를 말한다.

ROADS 소프트웨어의 개발은 eLib (the Electronic Libraries) 프로그램의 한 과제로서 추진되었다.

ROADS는 특정 주제 영역이나 지리적 위치에 따른 인터넷 자원을 식별하고 평가하고 기술하고 접근할 수 있는 서비스를 만드는데 사용된다. 이때 자원의 기술, 즉 메타데이터로서 ROADS 템플릿을 사용하는데 이것은 IAFA 템플릿에 기초한다.

템플릿의 정의와 관련된 연구는 계속 진행되고 있으며, 현재 정의되어 있는 템플릿 유형은 DATASET, DOCUMENT, DUBLINCORE(시험중), EVENT(시험중), IMAGE, MAILARCHIVE, PROJECT, SERVICE, SOFTWARE, SOUND, TRAINMAT, USENET, VIDEO이며 클러스터로 ORGANIZATION과 USER가 사용된다.

ROADS 템플릿은 속성/값의 쌍으로 기술하며 자원 제공자가 직접 기술할 수 있도록 한다. 레코드 검색을 위해서 whois++ 프로토콜을 사용한다.

ROADS 소프트웨어를 사용한 시스템으로는 eLib 프로그램의 주제 서비스로 개발되고 있는 사회과학 분야의 SOSIG (Social Science Information Gateway, <http://sosig.ac.uk>), 의학 분야의 OMNI(Organising Medical Networked Information, <http://omni.ac.uk>), 예술




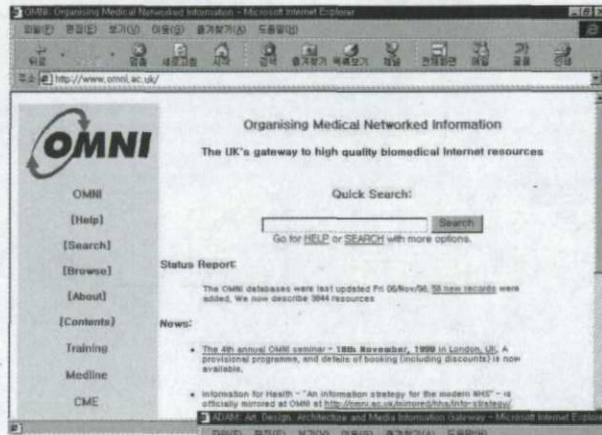
분야의 ADAM(Art, Design, Architecture & Media Information Gateway, <http://adam.ac.uk>)의 세가지 프로젝트가 대표적이다.

### 향후 방향

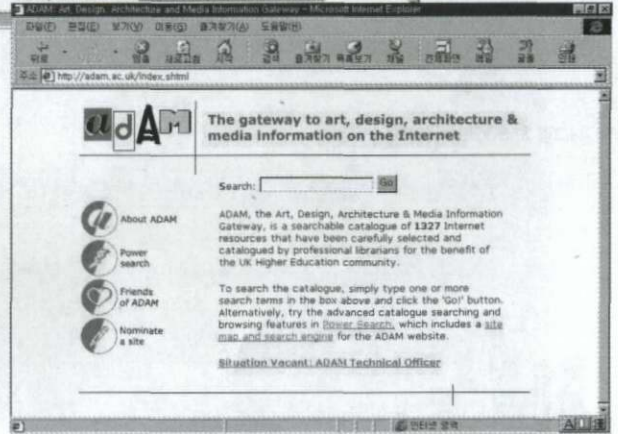
IAFA 템플릿은 인터넷 자원 가운데 정보를 축적하고 검색하는데 많이 활용되고 있는 익명FTP 아카이브를 기술하기 위한 메타데이터로서 유용성을 인정받고 있다. 그러나 데이터요소가 추가되어야 한다거나 변형(variants)을 이용한 데이터의 중첩 문제, 그리고 이진데이터나 비ASCII 문자인 경우 다루기 힘들다는 문제점을 안고 있다.

이러한 기술적 한계로 인해 IAFA 템플릿은 기존, 혹은 새로운 다른 메타데이터로의 매핑에 관한 연구가 더욱 활발하게 이루어지고 있는 실정이다. 이와 관련하여 MARC이나 DC로의 연결 혹은 매핑 문제가 논의되고 있다.

IAFA 템플릿은 인터넷 초안이 발표된 1995년 이후 표준으로 제정하기 위한 실무 작업이 현재는 중단된 상태이지만 유용한 인터넷 자원의 기술을 시도했고 이후 다른 메타데이터와의 연결을 위한 노력이 끊이지 않게 이어지고 있다는 사실은 주목해 볼만하다. 



(그림 4) OMNI 서비스



(그림 5) ADAM 서비스

### 〈참고문헌〉

- 문헌정보처리연구회. 메타데이터의 형식과 구조, 서울 : 문헌정보처리연구회, 1998, pp.202~248.
- Deutsch, P., et al. "Publishing Information on the Internet with Anonymous FTP," Internet Draft, Expires March 1995, <http://info.webcrawler.com/mak/projects/iafa/iafa.txt> (1998/11/12)
- Deutsch, P., et al. "How to Use Anonymous FTP", RFC 1635, May 1994, <http://www.nexor.com/public/rfc/rfcs/rfc1635.txt> (1998/11/12)
- Deutsch, P., et al. "Architecture of the WHOIS++ service", RFC 1835, August 1995, <http://www.ifla.org/documents/libraries/cataloging/metadata/rfc1835.txt> (1998/11/10)
- Beekett, D. "IAFA Templates in use as Internet Metadata", <http://www.w3.org/Conferences/WWW4/Papers/52/> (1998/11/02)
- The IAFA Working Group homepage, <http://info.webcrawler.com/mak/projects/iafa/iafa-wg.html> (1998/11/02)
- Day M. "Mapping IAFA templates to Z39.50 to Bib-1", August 1996, [http://www.ukoln.ac.uk/metadata/interoperability/iafa\\_z39.50.html](http://www.ukoln.ac.uk/metadata/interoperability/iafa_z39.50.html) (1998/11/02)
- "IAFA/WHOIS++ Templates", Work Package 3 of Telematics for Research project DESIRE(RE1004), [p://www.ukoln.ac.uk/metadata/desire/overview/rev\\_11.htm](p://www.ukoln.ac.uk/metadata/desire/overview/rev_11.htm) (1998/11/02)