P. Deutsch
A. Emtage
Bunyip
M. Koster
Nexor
M.Stumpf
Munich University of Technology

Publishing Information on the Internet with Anonymous FTP

1. Status of this Memo

2. Abstract

Anonymous FTP Archives are a popular method of making material available to the Internet user community. This document specifies a range of indexing information that can be used to describe the contents and services provided by such archives. This information can be used directly by the user community when visiting parts of the archive. Furthermore, automatic indexing tools can gather and index this information, thus making it easier for users to find and access it.

3. Acknowledgments

This document is the result of work done in the Internet Anonymous FTP Archives (IAFA) working group of the IETF. Special thanks are due to George Brett, Jill Foster, Jim Fullton, Joan Gargano, Rebecca Guenther, John Kunze, Clifford Lynch, Pete Percival, Paul Peters, Cecilia Preston, Peggy Seiden, Craig Summerhill, Chris Weider, and Janet Vratney.

4. Introduction

Over the past several years, Anonymous FTP has become the primary method of publishing information in the Internet environment. Anonymous FTP is an application-level service that makes use of the File Transfer Protocol [1], one of the principal protocols of the TCP/IP suite. A well organized and well maintained Anonymous FTP archive (AFA) can provide a relatively cheap and simple way to distribute the software, documents, datasets, images and other sources of information that are produced for general availability on the network today.

Those groups wishing to set up an Anonymous FTP Archive should refer to "A Guide to Anonymous FTP Site Administration" [2], which provides details on why you would want to set up such an archive and what steps are required to have a secure, well-maintained system.

This document specifies a range of indexing information that can be used to describe the contents and services provided by such archives. This information can be used directly by the user community when visiting parts of the archive. Furthermore, automatic indexing tools can gather and index this information, thus making it easier for users to find and access it. Although not required, providing such information will make the archive a more useful resource.

It is intended that this information be made available through anonymous FTP archives although the templates described may also be made available through any other information access mechanism. It is beyond the scope of this document to provide specific transformations to other mechanisms since the individual encoding method used will necessarily depend on several external factors such as operating systems and network protocols used.

Section 5 of this document contains definitions of the terminology used, as well as issues related to the use and construction of the information to be distributed.

In Section 6 we make recommendations that are intended to provide a standardized means for sharing information about the contents of a specific archive site such as as services provided by the institution, document abstracts, and software descriptions. In addition administrative contacts, local time zone and other site-specific details may be given.

Section 7 contains a set of encoding procedures for the information outlined in Section 6. The encoding is sufficiently general to be deployed on a variety of operating systems, and sufficiently flexible to allow the AFA administrator to take into account site-specific issues such as file system organisation. It is expected that where specific environments have special considerations, conventions for transforming the information can easily be defined.

Interested parties may also want to refer to the companion document "Data Element Templates for Internet Information Objects" [8] for fully expanded data templates defined in this document.

5. Administration

5.1. Scope of this document

The templates listed below are not intended to comprehensively describe all possible information that could be provided, but rather to cover common, useful elements. The determination about what specific information to provide will have to be made on a case by case basis. Those individuals or groups completing the information have to determine how appropriate a particular data element is for their needs. In many cases data elements such as "home telephone number" would be not be desirable in databases open for public access. However, in some cases they may be useful and thus have been included in this document.

NOTE: Issues of privacy, security and maintainability should all be considered when determining what information to provide.

This document does not mandate or require that any particular class of information be offered. However it is hoped that those sites wishing to offer the information described in this document adhere to the formats recommended in Section 7.

5.2. Definitions

For the purposes of this document, the term "data element" is defined to be a discrete (though not necessarily atomic) piece of information. For example, a name, telephone number or postal address would all be considered a "data element". The granularity at which a data element is defined is determined by the purpose for which it is intended. The term "field" is interchangeable with "data element".

"Templates" are logical groupings of one or more data elements.

Collectively the templates described in this document will be referred to as "indexing" or "data" templates.

A "resource" is any network object being described. This could be a "physical" object like a file, document or printer, or it may be a "service" such as a weather or Domain Name System server. Any object which can be referred to as being accessible or addressable on the network is a resource.

A "record" is an instance of the template with the appropriate fields filled in for a particular resource.

5.3. Uniform Resource Identifiers and Directory Services

The templates below generally describe network accessible resources, and people connected with these resources, and as such it is important to uniquely identify both resources and people.

Work is currently underway for the construction of what are known as "Uniform Resource Identifiers" (URI). These will be structured strings whose purpose is to uniquely identify any resource on the Internet to determine access and identification information for that resource. This not only includes documents, software packages etc., but also images, interactive services and physical resources. This concept has been integrated into the data templates. While it is expected that ultimately location independent identifiers will be used, the examples in this document utilize the Uniform Resource Locators as defined in [3].

Because there are no ubiquitous directory services to look up personal details for people the templates below contain facilities for these personal details to be provided.

It is likely that in the relatively near future directory services will be tested and deployed that will provide for both White Pages (locating personal details) and locating resources (Yellow Pages). It is expected support for these can be easily added to the templates defined in this document.

5.3.1. Variant Information

Often a particular resource is available in a number of variants. For example, a document may exist both in standard pre-formatted ASCII (a "text" file) and PostScript versions, or may be available in a number of different languages.

The person or group indexing the resource must determine which resources have equivalent "intellectual content", and if so describe them as variants of a single resource. By providing information such as location, format, character sets, languages etc. for each variant, a user searching the index is provided with enough context to make an informed decision as to which variant to retrieve.

It is hoped and expected that the methods of dealing with variant information described in this document will be superseded by a more comprehensive directory service system in the relatively near future.

5.4. Machine vs. human readbility

At the heart of some data element definitions is their ability to be parsed and "understood" by computer programs. It is hoped and expected that much of the information provided in the IAFA templates described below will be collected and indexed by automated processes without human intervention. As a result, care has been taken to restrict the syntax and semantics of data element names and some values so as to facilitate these procedures.

6. Configuration and Contents Information

In this section we define a recommended set of indexing information that you could make available as the administrator of an archive site. In doing so, you would extend the functionality of your archive, as well as the functionality of indexing and resource discovery tools that can pick up and redistribute such information.

6.1. Handles

Handles for individuals or organizations, if used, are defined to be a printable string that uniquely identify the individual or group, within the context of the service providing the handle. These are to be used as a shorthand method of referring to the complete organization or individual record and should be used in preference to the complete entry. Indexing tools which gather template information should be aware that once removed from a particular context, handles may no longer be unique and techniques must be used to ensure uniqueness out of context, or to expand the handle into associated values in the record.

6.2. Clusters: common data elements

There are certain classes of data elements, such as contact information, which occur every time an individual, group or organization needs to be described. Such data as names, telephone numbers, postal and email addresses etc. fall into this category. To avoid repeating these common elements explicitly in every template below, we define "clusters" which can then be referred to in a shorthand manner in the actual template definitions. Predefined symbols specifying these clusters will then be used in their place, with a prefix which determines to whom or to what this information applies.

In those cases where multiple instances of a cluster have to be defined (for example, to describe multiple authors of a book), then "variant" syntax applies. See section 3.2 "Variant fields".

The following clusters have been identified:

6.2.1. Individuals

In order to describe each individual in a particular template, the following common data element subcomponents are defined.

- Name of individual.
- Name of organization to which individual belongs or under whose authority this information is being made.
- Type of organization to which this individual belong (University, commercial organization etc.)
- Work telephone number of individual.
- FAX (facsimile) telephone number of individual.
- Postal address of individual.
- Job title of individual (if appropriate).
- Department to which individual belongs.
- Electronic mail address of individual.
- Home telephone number of individual.
- Home postal address of individual.
- Handle.

6.2.2. Organizations

The following elements apply when describing organizations and are a subset of those listed above for individuals. Obviously some of the elements above (such as home phone number) make no sense when being applied to an organization. As above, the following may be subcomponents in a larger, hierarchically structured data element name.

- Name of organization.
- Type of organization to which this individual or group belongs (University, commercial organization etc.).
- Postal address of organization.
- Electronic mail address of organization.
- Phone number of organization.
- Fax number of organization.
- City of organization.
- State (province) of organization.
- Country of organization.
- Handle.

6.2.3. Resource Information

The following is a list of generic data element subcomponents used when referring to particular resources.

- A title for the resource.
- Uniform Resource Identifier.
- Description.

- Any keywords which might be applied to the resource that would facilitate users' locating this information.
- Type of resource.
- City of resource.
- State (or Province) of resource.
- Country of resource.
- Comment.
- Details describing when the record was last maintained, and by who.

6.3. Site-specific configuration information

Information about your archive site itself can often be valuable to users of your system in order for them to utilize the resource in an efficient manner.

6.3.1. Configuration Information

Site configuration information will help users better understand your wishes on how and when to access your AFA. This would include such information as:

- Primary host name of the AFA.
- A valid Domain Name System alias (CNAME) for this host [5].
- Individual contact information for site owner(s).
- Individual contact information for site maintainer (administrator).
- Sponsoring organization contact information.
- The geographical (latitude/longitude) location.
- The time zone of the site.
- Individual contact information for last person last modifying this record.
- The frequency with which the archive site is generally modified.
- Times of preferred access for this site.
- A summary of the access policies of this site. This should include such information as preferred times of usage, conventions or restrictions for uploading files to this site etc.
- A brief description of the kind of information stored at this anonymous FTP archive. If the site is intended to specialize in a particular type of information (examples might include software for a specific machine type, on-line copies of a particular type of literature or research papers and information in a particular branch of science or arts) you should indicate this.
- Resource information as defined in the resource cluster.

6.3.2. Logical archives configuration

One physical archive site may possibly contain multiple "logical" archives. For example, a single archive host may be shared amongst multiple departments, each responsible for the administration of their own part of the anonymous FTP directory subtree.

Some information (such as a host's location) will remain constant for the site as a whole. We therefore recommend that you list Logical Archive specific and site-specific information separately.

- Individual contact information for site maintainer administrators).
- A valid Domain Name System alias (CNAME) for this host [5] when referring to this logical archive.
- Owning organization contact information.
- Sponsoring organization contact information.

- Individual contact information for last person last modifying this record.
- A summary of the access policies of this logical archive.
- A summary of the type of information that this logical archive may specialize in.
- The frequency with which the archive site is generally modified.
- Resource information as defined in the resource cluster.

6.4. Site-specific content and service information

The preceding collections of information make available access and utilization policies for a site. You could also wish to make available a selection of information about the actual contents of your archive or the services available from your organization or institution.

The host system providing the resources need not be the same physical site on which the descriptive information below is stored. Thus at a University an AFA maintained by the central campus administration could advertize services provided by individual departments who might not have an AFA of their own. Similarly, mailing lists provided on other administratively related hosts (such as in the same organization) may have the indexing information available on one host while the actual mailing list is provided by another machine.

The following categories have been identified.

6.4.1. Service

The archive can offer an overall description of each of the various Internet services offered by your organization’s systems, along with corresponding contact information.

This description would then indicate whether the parent organization offers such services as:

- on-line library catalogues.
- Interactive on-line information services such as WAIS, gopher, Prospero, World Wide Web or archie.
- specialized information servers such as those providing weather, geographic information, newswire feeds etc.
- Other information services.

The following information can be made available:

- Title of service.
- URI of service.
- A description of the service.
- Any keywords which might be applied to the record that would facilitate users finding this service.
- Contact information for service administration.
- Authentication information (login name, password etc. if required) or method for authentication (private key etc.)
- Description of registration process.
- Charging policies for service.
- Policies and restrictions on service use.
- Access times for service.

6.4.2. Documents, Datasets, Mailing List Archives, Usenet Archives, Software Packages, Images and other objects

You might wish to make available a brief description of available software, documents, images, sounds, video, datasets, USENET [6] archives and mailing list information through the AFA.

Some of the information classes described may not be applicable to each of the above objects.

This is NOT intended to be an official catalog entry in the sense used by librarians. It is a simple way to describe documents and announce their availability. More formal methods may be used elsewhere to further describe the documents.

- Type of object.
- Category (for documents this would be technical report, conference paper etc.)
- Name of object. For example, the name of the mailing list, software package or title of the document.
- Names and other contact information on the authors.
- Names and other contact information for object maintainer/administrator.
- Version designator.
- Source of data.
- Abstract/description of the object.
- Bibliographic entry.
- Citation.
- Special considerations or restrictions on the object's use (e.g., in the case of a software package programming languages/environments needed, hardware restrictions, etc.).
- Publication status (For documents: draft, published etc. For software packages: beta test, production etc.)
- Contact information of publisher.
- Copyright and copying policy.
- Creation date.
- Appropriate keywords for this object.
- Discussion forums appropriate for this object (mailing lists, USENET newsgroups etc.)
- Format of the object (variant).
- Size (variant).
- Language (variant).
- Character set (variant).
- ISBN (variant).
- ISSN (variant).
- Method of access (anonymous FTP etc.).
- Last revision date (variant).
- Library Cataloging information.
- URI (variant).

7. Information Encoding

In this section we offer a recommended encoding format for each of the standard items of information suggested in Section 2.

We offer such a standardized format so that if such information is to be offered, it is formatted in such a way that it can be utilized by automated indexing and retrieval tools. The encoding methods proposed were developed to be extensible, so that additional information can be offered in a similar format, if the site administrator so wishes.

Developing such recommendations offers several challenges. It is hoped that the encoding conventions should be applicable to as wide a variety of operating systems, file structures and encoding schemes as possible. In addition, the globalization of the Internet requires attention to constraints such as the language in use at an archive site.

In addition, the encoding methods proposed must be easy to implement and, for the moment, use existing methods of access and retrieval. We currently assume that the site language is English and the encoding ASCII, but it is expected that additional formats for other languages and encoding schemes will be developed over time.

7.1. Data element Structure

All data elements have been defined as "attribute/value" pairs which can be generically described as:

<data element name>: <data element value>

where <data element name> would for example be "Work-Phone" and the <data element value> would be "+1 514 555 1212" (note that the double quotes (") are not part of the strings, but serve here to delimit the example).

The term "field name" is interchangeable with "data element name". The term "field value" is interchangeable with "data element value".

All data element names may contain only alphanumeric characters, the hyphen ("-") and the underscore ("_"). No embedded spaces are allowed. All data element names are case insensitive although here initial letters are capitalized for readability.

Lines starting with '#' are ignored.

Field data must be separated from fieldname by a colon and optional whitespace. Any field may continue on the next line by whitespace in the first column of that line. Multi-line fields are delimited by the first line which does not have whitespace in the first column, or is blank. Whitespace between continuation lines is to be collasped into a single space character by processing software (Except in the URI field, where this space is removed).

Data element names without associated field values are allowed, but have no significance.

A record can only have a single Template-Type. Other data elements may have multiple values, and these values are taken to indicate equally appropriate alternatives.

Data elements may occur in any order. However, for easier readability it is recommended to start with the Template-Type, Description, and keywords, followed by other non-variant fields, followed by variant fields grouped per variant.

It is intended that wherever possible and necessary, a well-defined hierarchical structure will be used when defining data element names. This allows them to be generally and logically extensible.

7.1.1. Variant Fields

In section 5.3.1 we describe some information as being "variant" in that network objects may vary in "format" but are judged to have the same "intellectual content". In the following data element definitions we use the technique of allowing a sequence number to be appended to a set of data elements to describe a particular variant.

For example, we have a document "War and Peace" which exists in ASCII text, PostScript and NROFF format. The PostScript version also exists in two natural languages, English and Russian. We define here 3 data elements: "Filename", "Language" and "Format". In addition to the other information stored in the indexing record for "War and Peace" which we consider to remain constant across all variants, (like the name of the author), we can add the following data elements:

```
Format-v0:   PostScript
Language-v0: English
URI-v0:      ftp://arch.com/book/wap/war-and-peace.english.ps
Format-v1:   PostScript
Language-v1: Russian
URI-v1:      ftp://arch.com/book/wap/war-and-peace.russian.ps
Format-v2:   ASCII
Language-v2: English
URI-v2:      ftp://arch.com/book/wap/war-and-peace.english.txt
Format-v3:   nroff
Language-v3: English
URI-v3:      ftp://arch.com/book/wap/war-and-peace.english.nroff
```

The "-v<number>" syntax allows one to repeat a set of data elements for a particular variant and tie them all together with a common sequence <number> so that individual instances of the particular resource with the desired characteristics may be located.

<number> is an arbitrary number with the only restriction that all data elements with that particular sequence value are logically connected in a similar manner to that illustrated above.

The variant number need not exist when variants are not being described and the "-v<number>" syntax may be omitted in those cases.

In the data element definitions below, the syntax "-v*" will be used to identify those elements for which variants are allowed.

7.2. Data Formats

To facilitate the machine readability of certain data elements, the following syntaxes are to be used for particular types of fields:

1) All electronic mail (Email) addresses must be as defined in RFC 822, Section 6 [10]. Names and comments may be included in the Email address.

   For example:

   "John Doe" <jd@ftp.bar.org>

   and

   jd@ftp.bar.org

   are valid Email addresses.

2) All hostnames are to be given as Fully Qualified Domain Names as defined in RFC 1034, Section 3 [3].

   For example: "foo.bar.com"

3) All host IP addresses are given in "dotted-quad" (or "dotted-decimal") notation.

   For example: "127.0.0.1"

4) All numeric values are in decimal unless otherwise stated.

5) Dates/times must be given as defined in RFC 822, Section 5.1 [10] and modified in RFC 1123, Section 5.2.14 [7]:

```
date-time   =  [ day "," ] date [time]

day         =  "Mon"  / "Tue" /  "Wed"  / "Thu"
            /  "Fri"  / "Sat" /  "Sun"

date        = 1*2DIGIT month 2*4DIGIT        ; day month year
                                             ;  e.g. 20 Jun 1982
```

```
month       =  "Jan"  /  "Feb" /  "Mar"  /  "Apr"
            /  "May"  /  "Jun" /  "Jul"  /  "Aug"
            /  "Sep"  /  "Oct" /  "Nov"  /  "Dec"

time        =  hour zone                   ; ANSI

hour        =  2DIGIT ":" 2DIGIT [":" 2DIGIT]
                                           ; 00:00:00 - 23:59:59

zone        =  "UT"  / "GMT"               ; Universal Time
                                           ; North American : UT
            /  "EST" / "EDT"               ;  Eastern:  - 5/ - 4
            /  "CST" / "CDT"               ;  Central:  - 6/ - 5
            /  "MST" / "MDT"               ;  Mountain: - 7/ - 6
            /  "PST" / "PDT"               ;  Pacific:  - 8/ - 7
                                           ;
            / ( ("+" / "-") 4DIGIT )       ; Local differential
                                           ;  hours+min. (HHMM)
```

For example the string "Sat, 18 Jun 1993 12:36:47 -0500" is a valid date, and the string "12:36:47 GMT" is a valid time. Quoting from RFC 1123, Section 5.2.14 [7]: "There is a strong trend towards the use of numeric timezone indicators, and implementations SHOULD use numeric timezones instead of timezone names. However, all implementations MUST accept either notation. If timezone names are used, they MUST be exactly as defined in RFC-822."

6) Time ranges (or periods) must be specified as pairs of time values (as defined above in note (5)), separated by a "/". Multiple time ranges are separated by whitespace. All times in a range should be specified with the same timezone.

   For example:

   12:00 GMT / 05:45 GMT

7) "whitespace" is defined as one or more blank (hex 0x20) and/or tab (octal 11) ASCII characters.

8) References to "UT" mean Universal Time (also known as Greenwich Mean Time or "GMT").

9) All telephone numbers are to be given as a minimum in full, with a leading '+' and country and routing codes without non-space separators. The number should be given assuming someone calling internationally (without local access codes). The number given in the local convention may optionally be specified in bracktes.

   For example,

   Telephone: +44 71 732 8011

   or

   Telephone: +1 514 875 8189 (0514-875-8611)

10) Latitude and longitude are specified in that order as

   CDD.MM.SS/CDD.MM.SS

   Where

   DD is in degrees
   MM is in minutes
   SS is in seconds
   C is the direction designator which is

   For latitude

"+" is north of the equator
"-" is south of the equator

For longitude
"+" is west of the Greenwich meridian
"-" is east of the Greenwich meridian

The double quotes (") are not part of the designator, but are used here to delimit the symbols.

11) Person name fields should conform to a particular format (based on bibtex[11]), so that they can be parsed into parts. A name can have four parts: first, von, last, junior, each of which can consist of more than one word. For example, "John Paul von Braun, Jr." has "John Paul" as the first part, "von" as the von part, "Braun" as the last part, and "Jr." as the junior part. Use one of these formats for a name:

```
First von Last
von Last, First
von Last, Junior, First
```

The last part is assumed to be one word, or all the words after the von part. Anything in braces will be treated as one word, so use braces to surround last names that contain more than one word. The von part is recognized by looking for words that begin with lowercase letters. When possible, enter the full first name(s). Actually, the rules for isolating the name parts are a bit more complicated, so they do the right thing for names like "de la Grand Round, Chuck".

If there are multiple authors or editors, they should all be separated by the word and.

7.3. File Record Structure

An indexing file can contain zero or more records, which are made up of collections of data elements. Records are delimited by one or more blank lines (lines which contain zero or more whitespace characters and the NEWLINE character). Because blank lines are used to delimit records they are not allowed to occur in a record.

This allows templates relating to the same resource, for example records describing documentation and software belonging to a single package, to be compiled in a single location. In addition it allows indexing files describing different resources to be combined by simply concatenating the separate indexing files.

Leading and trailing blank lines on the indexing file are allowed, but not significant. Empty indexing files are to be ignored.

7.4. File Location and Naming

For the greatest flexibility, it is assumed that unless otherwise stated each file containing the indexing information may reside anywhere in the anonymous FTP subtree and in addition, any number of these files may exist. The intention here is that they may be placed in the same location as the information they are indexing. You, as the administrator are free to place these files wherever you think appropriate in most cases. However, some files may carry information from their place in the directory structure and therefore they may not just be randomly placed in the archive.

In order for tools to easily identify an indexing file from the other data files at the archive site, all indexing filenames must end with a ".AFA" filename extension.

Indexing files should be made world readable. It is assumed that size and modification times can be obtained through existing access mechanisms and are operating system specific.

The advantages to this system are that this information need only be constructed once with infrequent periodic updates as changes occur. Several of these files may never change during the lifetime of the host as an anonymous FTP site. They require no special programs or protocols to construct: a text editor is all that is needed.

7.5. Clusters: Common Data Elements

As described in Section 6.2, there are number of data elements which are often needed and which form a natural grouping for certain kinds of information ("clusters"). Below we define the data element names and semantics of these clusters.

These clusters are intended to provide the lowest level in the hierarchical structure of data element names. For example, contact information for the authors of a document would be preceded by the string "Author-" thus forming data elements of "Author-Name", "Author-Work-Postal", "Author-Work-Fax", etc.

NOTE: In the definitions below, the fields are separated by blank lines ONLY to improve readability, these lines must NOT occur in an actual record.

7.5.1. Individuals or Groups

Data Element Name	Description
Name	Name of individual.
Work-Phone	Work telephone number of individual.
Work-Fax	FAX (facsimile) telephone number of individual.
Work-Postal	Postal address of individual.
Job-Title	Job title of individual (if appropriate).
Department	Department to which individual belongs.
Email	Electronic mail address of individual.
Handle	Unique identifier for this record.
Home-Phone	Home telephone number of individual.
Home-Postal	Home postal address of individual.
Home-Fax	FAX (facsimile) telephone number of individual.

This cluster can also contain any of the elements of the ORGANIZATION cluster described in 7.5.2, to describe the organization to which individual belongs or under whose authority the information is being made.

This cluster will be referred to as "USER*" in the template definitions below.

7.5.2. Organisations

The following elements apply when describing organizations and are a subset of those listed above for individuals and groups. Obviously some of the elements above (such as home phone number) make no sense when being applied to an organization. As above, the following may be subcomponents in a larger, hierarchically structured data element name.

Data Element Name	Description.
Organization-Name	Name of organization.
Organization-Type	Type of organization (University, commercial organization etc.)
Organization-Postal	Postal address of organization.
Organization-City	City of organization.
Organization-State	State (province) of organization.
Organization-Country	Country of organization.
Organization-Email	Electronic mail address of organization.

Organization-Phone	Phone number of organization.
Organization-Fax	Fax number of organization.
Organization-Handle	Handle of organization.

This cluster will be referred to as "ORGANIZATION*" in the template definitions below.

7.5.3. Miscellaneous

The following is a list of generic data element subcomponents used when referring to particular resources. These can be added to any of the templates described below.

Data Element Name	Description
Title	A complete title for the resource.
Description	Description of resource.
Keywords	Any keywords which might be applied to the record that would facilitate users' finding this resource.
URI	Uniform Resource Identifier
Access-Method	Free-text description of access method if no URI syntax has been defined.
City	City of resource.
State	State (Province, etc.) of resource.
Country	Country of resource.

7.5.4. Maintenance

The following is a list of generic data elements used to indicate when the record was last maintained.

Data Element Name	Description
Record-Last-Modified-(USER*):	Contact information for individual who last modified this record.
Record-Last-Modified-Date:	The date this record was last modified.
Record-Last-Verified-(USER*):	Contact information of person or group last verifying that this record was accurate.
Record-Last-Verified-Date:	The date the last time this record was verified.

8. Template Definitions

NOTE: In the definitions below, the fields are separated by blank lines ONLY to improve readability, these lines must NOT occur in an actual record.

8.1. Site Information

IMPORTANT: There should only be one instance of this template in each archive.

Fields for this template.

Template-Name:	SITEINFO
Host-Name:	Primary Domain Name System host name.
Host-Alias:	Preferred DNS-registered name for the AFA host. This name must be valid CNAME entry in the Domain Name System.
Admin-(USER*):	Contact information of the individual or group responsible for administering this site.
Owner-(ORGANIZATION*):	Contact information for the organization owning this site.

Sponsoring-(ORGANIZATION*):	Contact information for the organization sponsoring this site.
City:	City of the host.
State:	State (province) of the host.
Country:	Country of the host.
Latitude-Longitude:	Latitude and longitude of site.
Timezone:	Timezone as defined in section 7.2 above.
Update-Frequency:	Preferred frequency of retrieval of all AFA extended configuration information by automated retrieval tools. (See Note <1>)
Access-Times:	Time ranges (as defined in Section 7.2) of access to anonymous FTP users.
Access-Policy:	Information such as conventions or restrictions for uploading files to this site etc.
Description:	This file contains text describing any areas of specialization for this site. For example, if the site contains information related to the field of molecular biology a paragraph or two with the keywords "molecular biology" and some further description would be in order. It should also mention if this site contains "logical" archives.
Keywords:	Appropriate keywords describing contents of this AFA.

Notes for this template:

<1> The period is measured in days. This value should be chosen to reflect the turnover of information at the archive.

An example of a SITEINFO record:

Template-Type:	SITEINFO
Host-Name:	foo.bar.org
Host-Alias:	ftp.bar.org
Admin-Name:	John Doe
Admin-Work-Postal:	PO Box. 6977, Marinetown, PA 17602
Admin-Work-Phone:	+1 717 555 1212
Admin-Work-Fax:	+1 717 555 1213
Admin-Email:	FTP@bar.org
Owner-Organization-Name:	Beyond All Recognition Foundation
City:	Lampeter
State:	Pennsylvania
Country:	USA
Latitude-Longitude:	-37.24.43/+121.58.54
Timezone:	-0400
Record-Last-Modified-Name:	John Doe
Record-Last-Modified-Email:	johnd@bar.org
Record-Last-Modified-Date:	Mon, 10 Feb 1992 22:43:31 EST
Update-Frequency:	10
Access-Times:	02:00 GMT / 08:00 GMT 18:00 GMT / 21:00 GMT
Access-Policy:	Non-proprietary data may be uploaded to this site in the "incoming" directory. Please contact site administrators if you do so. Proprietary material found in this directory will be removed. This

	site is not to be used as a temporary storage area.
Description:	This site contains data relating to DNA sequencing particularly Yeast chromosome 1. Datasets are available. There is also a selection of programs available for manipulating this information.
Keywords:	DNA, sequencing, yeast, genome, chromosome

8.2. Logical Archive Information

IMPORTANT: The placement of this file in the file structure is significant: It implies that the directory in which this file exists and all subdirectories are part of the logical archive.

Template-Type:	LARCHIVE
Admin-(USER*):	Contact information of the individual or group responsible for administering this site.
Host-Name:	Primary Domain Name System host name.
Host-Alias:	Preferred DNS-registered name for the AFA host as this logical archive. This name must be valid CNAME entry in the Domain Name System.
Owner-(ORGANIZATION*):	Contact information for the organization owning this site.
Sponsoring-(ORGANIZATION*):	Contact information for the organization sponsoring this site.
Access-Policy	Information such as conventions or restrictions for uploading files to this logical archive.
Description	Contains text describing any area of specialization for the logical archive.
Update-Frequency	Preferred frequency of retrieval of all AFA extended configuration information by automated retrieval tools. (See Note <1>)
Keywords	Appropriate keywords describing contents of this logical AFA.

Notes for this template:

<1> The period is measured in days. This value should be chosen to reflect how often information at the archive changes.

An example of a LARCHIVE record:

Template-Type:	LARCHIVE
Owner-Organization-Name:	Orymonix Incorporated
Owner-Organization-Type:	Commercial
Host-Alias:	oxymoron-x.co.uk
Access-Policy:	This archive is open to general access
Description:	This archive contains essays on Military Intelligence, Postal Service and Progressive Conservatism. All material contained in this archive is in the public domain
Admin-Name:	Ima Admin
Admin-Email:	imaa@oxymoron-x.co.uk
Admin-Work-Phone:	+44 71 123 4567
Admin-Work-Fax:	+44 71 123 5678
Admin-Work-Postal:	555 Marsden Road, London, SE15 4EE
Record-Last-Modified-Name:	Yuri Tolstoy
Record-Last-Modified-Email:	yt@snafu.co.uk

Record-Last-Modified-Date:	Mon, 21 Jun 1993 17:03:23 EDT
Update-Frequency:	20
Keywords:	Militarism, Post Office, Conservatism

8.3. Automatic File Update Information

Any number of these files may exist in the archive.

Template-Type:	MIRROR
Admin-(USER*):	Contact information of the individual or group responsible for administering this mirror.
Owner-(ORGANIZATION*):	Information on organization responsible for this mirror unit.
Title:	The title of the package.
Description:	Text describing the package.
Reference-URI-v*:	The starting point. This is the initial site the package can be found of. As there may be more than one file or directory belonging to this package this is a -v* type. Specified as an URI. (See Note <1>)
Source-URI-v*:	The location the package is mirrored from. This may itself be a mirror site of Reference-URI or another Source-URI. Specified as an URI.
Destination-URI-v*:	The location the package can be found locally. Specified as an URI.
Timezone:	The timezone this site is in. (see section 7.2 of this document)
Update-Frequency:	The Source-Site is checked each this number of days or on these days. (See Note <2>)
Update-Time:	The time of day the update is started. This is important for chained updates, i.e. sites using this site as Source-URI.
Update-Policy:	This is how the update is done. There are a few valid keywords. See Note <3> for more information.
Update-Filename-Translation:	Substitute expression. This may used to reorganize e.g. a flat directory on Source-URI into various subdirectories on Destination-URI.
Update-Transfer-Pattern:	A regular expression. Only files matching this pattern on Source-URI will be updated/fetched.
Update-Exclude-Pattern:	A regular expression. Files matching this pattern on Source-URI will not be updated/fetched.
Update-Compression-Pattern:	A regular expression. Used for packing or re-packing files being updated/ fetched. (see Note <4>)
Update-Software:	Name and version of the software used for the automatic updates.

Notes for this template:

<1> The -v* form is especially useful, if you mirror a package within a directory called "path", but you don't mirror the whole "path", but only the "src" and "doc" subdirectories.

<2> This may be any number or one or more of the (comma seperated) words "Mon", "Tue", Wed", "Thu", "Fri", "Sat" or "Sun".

<3> Valid keywords are:

autodelete	files will be automatically deleted, when they are no longer found on Source-URI.
sizechange	files will also be updated if only the size but not the time changed on the Source-URI.
newer	files will be updated if the file on Source-URI is newer than the one on Destination-URI.

maxdays=num	files will not be fetched/updated if its modification time has a difference bigger than <num> days to the file on Destination-URI.
recursive	directories will be mirrored recursively (otherwise only the contents of the "flat" directory will be updated and no subdirectories will be checked).

<4> This specifies whether e.g. *.tar files will be packed (and therefor renamed) to *.tar.Z or *.tar.gz, or whether e.g. *.Z files will be packed and renamed to *.gz

This is an example of a MIRROR record.

Template-Type:	MIRROR
Admin-Name:	John Long Silver
Admin-Email:	silver@jamaica.world
Admin-Home-Phone:	+1 222 333 4567
Owner-Organization-Name:	The Pirates Club
Title:	The ultimate treasury package
Description:	This package helps you to become rich, and richer and richer. It shows how to collect money and hide it from anyone within your computer. You can use a program from this package to materialize the money again, later.
Record-Last-Modified-Name:	Sailor One
Record-Last-Modified-Date:	Sat, 15 Jan 1994 02:47:57 GMT
Record-Last-Verified-Name:	Sailer Two
Record-Last-Verified-Date:	Sat, 15 Jan 1994 02:47:57 GMT
Reference–URI–v0:	ftp://ftp.money.us/pub/coins/silver/
Source–URI–v0:	ftp://ftp.cash.mx/money/coins/silver/
Destination–URI–v0:	ftp://ftp.jamaica/pub/coins/
Reference–URI–v1:	ftp://ftp.money.us/pub/coins/gold/
Source–URI–v1:	ftp://ftp.cash.mx/money/coins/gold/
Destination–URI–v1:	ftp://ftp.jamaica/pub/coins/
Timezone:	-0700
Update-Frequency:	Mon, Wed, Fri
Update-Time:	02:00
Update-Policy:	sizechange, maxdays=14, recursive
Update-Filename-Translation:	s:(.*)(gold/|silver/)(.*):$1$2:;
Update-Transfer-Pattern:	
Update-Exclude-Pattern:	
Update-Software:	coin-transfer, version 3.17

8.4. Content Information

For the following categories the assumption should not be made that the information applies to the anonymous FTP host itself. Rather, it applies to the material on the Archive.

8.4.1. User Information

So as not to require the repetition of the USER* information each time this cluster is needed in other templates, we define here a USER template in which the information can be stored in one place. Assuming the use of a unique handle, other records may then use a handle to refer to this record. The definition is simply the data elements listed in 7.5.1 above.

The Template-Type is USER.

8.4.2. Organization Information

In a similar manner to the USER template, the ORGANIZATION template provides common information which may be used in other (larger) templates to yield a central source of information.

The Template-Type is ORGANIZATION.

8.4.3. Service Information

These are the fields for the SERVICE template.

Template-Type:	SERVICE
Title:	Title of service.
URI:	URI of service.
Admin-(USER*):	Contact information of person or group responsible for service administration (administrative contact).
Owner-(ORGANIZATION*):	Information on organization responsible for this service.
Sponsoring-(ORGANIZATION*):	Contact information for the organization sponsoring this site.
Description:	Free text description of service.
Authentication:	Authentication information. Free text field supplying login and password information (if necessary) or other method for authentication.
Registration:	How to register for this service if general access is not available.
Charging-Policy:	Free text field describing any charging mechanism in place. Additionally, fee structure may be included in this field.
Access-Policy:	Policies and restrictions for using this service.
Access-Times:	Time ranges for mandatory or preferred access of service.
Keywords:	Keywords appropriate for describing this service.

Example 1:

The following is an example of an entry for a telnet service.

Template-Type:	SERVICE
Title:	Census Bureau information server
URI:	telnet://census.ispy.gov:1234
Admin-Name:	Jay Bond
Admin-Work-Postal:	PO Box. 42, A Street Washington DC, USA 20001
Admin-Work-Phone:	+1 202 222 3333
Admin-Work-Fax:	+1 202 444 5555
Admin-Email:	jb007@census.ispy.gov
Description:	This server provides information from the latest USA Census Bureau statistics (1990) Type "help" for more information.
Authentication:	Once connected type your email address at the "login:" prompt. No password is required.
Registration:	No formal registration is required
Charging-Policy:	There is no charge for the use of this service
Access-Times:	09:00 EST / 17:00 EST

Access-Policy:	This service may not be used by sites in the Republic of the VTTS
Keywords:	census, population, 1990, statistics
Record-Last-Modified-Name:	Miss Moneypenny
Record-Last-Modified-Email:	m.moneypenny@census.ispy.gov
Record-Last-Modified-Date:	Wed, 1 Jan 1970 12:00:00 GMT

Example 2:

The following is an example of a mailing list (service).

Template-Type:	SERVICE
Title:	fishlovers
URI:	fishlovers@foo.com
Admin-Name:	Ima Adams
Admin-Email:	fishlovers-request@foo.com
Registration:	Send mail to the administrative address with your own email address requesting addition
Description:	Discussion list for people who love fish of all types
Keywords:	fish, aquarium, marine, freshwater, saltwater
Access-Policy:	Any Internet user may subscribe to this mailing list

8.4.4. Documents, Datasets, Mailing List Archives, Usenet Archives, Software Packages, Images and other objects

These templates all contain the same fields, but have different "Template-Type" values. Suggestions for these types include:

Type of Object	Template-Type
Document:	DOCUMENT
Image:	IMAGE
Software Package:	SOFTWARE
Mailing List Archive:	MAILARCHIVE
Usenet Archive:	USENET
Sound File:	SOUND
Video File:	VIDEO
Frequently Asked Questions File:	FAQ

Other names may be added to future releases of this document.

Template-Type:	See above list
Category:	Type of object. See Note <1>
Title:	Complete title of the object.
URI-v*:	Description of access to object.
Short-Title:	Summary title (if the Title is very long).
Author-(USER*):	Description/contact information about the authors/creators of the object.
Admin-(USER*):	Description/contact information about the administrators/maintainers of the object.
Source:	Information as to the source of the object.

Requirements:	Any requirements for the use of the object. A free text description of any hardware/software requirements necessary to use the object.
Description:	Description (that is, "abstract" in the case of documents) of the object.
Bibliography:	A bibliographic entry for the object.
Citation:	The citation for the object when used in other works.
Publication-Status:	Current publication status of object (draft, published etc.).
Publisher-(ORGANIZATION*):	Description/contact information about object publisher.
Copyright:	The copyright statement. Any additional information on the copying policy may be included.
Creation-Date:	The creation date for the object.
Discussion:	Free text description of possible discussion forums (USENET groups, mailing lists) appropriate for this object.
Keywords:	Appropriate keywords for this object.
Version-v*:	A version designator for the object.
Format-v*:	Formats in which the object is available. (See Note <2>)
Size-v*:	Length of object in bytes (octets).
Language-v*:	The name of the language in which the object is written. For documents this would be the natural language. For software this would be the programming language.
Character-Set-v*:	The character set of the object. This should be a well-known value for example "ASCII" or "ISO Latin-1".
ISBN-v*:	The International Standard Book Number of the object.
ISSN-v*:	The International Standard Serial Number of the object.
Last-Revision-Date-v*:	Last date that the object was revised.
Library-Catalog-v*:	Library cataloging information. (See Note <3>)

Notes for this template:

<1> The intention of this field is to define the category of the object. For example, in the case of documents it could be "Technical Report", or "Conference Paper" and the name and date of the conference at which the paper was presented. It may also be something like "General Guide" or "User manual".

<2> Objects are often available in several formats. For example, documents may be in PostScript, ASCII text, DVI etc. For images this may be GIF, JPEG, TIFF etc.

Format should be specified in MIME type syntax and semantics where possible (See [9]).

<3> Library cataloging numbers. In those cases where the number itself does not contain enough information to determine the cataloging scheme, the name of the scheme should be included.

Example 1:

Example of DOCUMENT record.

Template-Type:	DOCUMENT
Title:	The Function of Homeoboxes in Yeast Chromosome 1
Author-Name:	John Doe
Author-Email:	jdoe@yeast.foobar.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–v:	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

Example 2:

This is an example of a SOFTWARE record. Note the use of the software maintainer's "handle" instead of the explicit contact information.

Template-Type:	SOFTWARE
Title:	Beethoven's Fifth Player
Version:	67
Author-Name:	Ludwig Van Beethoven
Author-Email:	beet@romantic.power.org
Author-Work-Fax:	+43 1 123 4567
Admin-Handle:	berlioz01
Description:	The program provides the novice to Transitional Classical-Romantic music a V-window interface to the author's latest composition
Requirements:	Requires the V-Window system version 10 or higher
Discussion:	USENET rec.music.classical
Copyright:	Freely redistributable for non-commercial use. Copyright held by author
Keywords:	Classical music, V-windows

Format:	LZ compressed
URI:	gopher://power.org/00/pub/Vfifth.tar.Z

9. Security Considerations

Issues of privacy and security should all be considered when determining what information to provide.

10. Conclusion

This document attempts to provide the foundation for a common set of recommended cataloging practices which may be used on the Internet to enhance the utility of Anonymous FTP archives, currently the most widely used and supported mechanism for general information storage and retrieval. It is intended that these recommendations be flexible enough to accommodate a broad spectrum of information classes and it is hoped that they will be widely used and that automated tools will be developed to use the valuable information that they make available.

11. References

[1] RFC 959 Postel, J.B.; Reynolds, J.K. File Transfer Protocol. 1985 October

[2] "A Guide to Anonymous FTP Site Administration". Work in progress from the Internet Anonymous FTP Archive Working Group of the IETF.

[3] Internet Draft "draft-ietf-uri-resource-names-02.txt" Work in Progress from the Uniform Resource Identifier Working Group of the IETF.

[4] RFC 954 Harrenstien, K.; Stahl, M.K.; Feinler, E.J. NICNAME/WHOIS. 1985 October

[5] RFC 1034 Mockapetris, P.V. Domain names - concepts and facilities. 1987 November

[6] RFC 1036 Horton, M.R.; Adams, R. Standard for interchange of USENET messages. 1987 December

[7] RFC 1123 Braden, R.T.,ed. Requirements for Internet hosts - application and support. 1989 October

[8] Internet Draft "Data Element Templates for Internet Information Objects". Work in progress from the Internet Anonymous FTP Archive Working Group of the IETF.

[9] RFC 1521 N. Borenstein, N. Freed, "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies", September 1993.

[10] RFC 822 D. Crocker, "Standard for the format of ARPA Internet text messages", August 1982. (Updated by RFC1327, RFC0987)

[11] BIBTEX(1) Manual Page, Oren Patashnik, June 1984.

12. Authors' Addresses

Peter Deutsch
Bunyip Information Systems
266 Blvd. Neptune
Dorval, Quebec, H9S 2L4
Canada

Phone: (514) 398-3709
Email: peterd@bunyip.com

Alan Emtage
Bunyip Information Systems
266 Blvd. Neptune
Dorval, Quebec, H9S 2L4
Canada

Phone: (514) 398-3709
Email: bajan@bunyip.com

Martijn Koster
NEXOR
PO Box 132
Nottingham NG7 2UU
The United Kingdom

Phone: +44 115 9520 576
Email: m.koster@nexor.co.uk

Markus Stumpf
Arcisstrasse 62/II
D-80799 Muenchen
Germany

Phone: +49 89 2714117
Email: stumpf@Informatik.TU-Muenchen.DE