ReDIF version 1

Current maintainer: Thomas Krichel

This document contains contributions by José Manuel Barrueco Cruz, Christopher Baum, Sune Karlsson, Ivan Kurmanov and Christian Zimmermann.

1 Status of this document

This version

http://openlib.org/acmes/root/docu/redif_1.html
http://openlib.org/acmes/root/docu/.papers/redif_1.a4.pdf
http://openlib.org/acmes/root/docu/.papers/redif_1.letter.pdf

It may be reproduced without the permission of the maintainer. You are encouraged to distribute it to your friends and colleagues. We hope that the collection of data that are encoded using ReDIF will also be widely distributed. Get in touch with the maintainer if wish to use ReDIF. Working together is more fun and more productive than working on your own.

Note that examples in this document are mostly fictitious. All similarity with the real world should be understood as a coincidence.

2 Introduction

This document describes version 1 of the Research Documents Information Format (ReDIF). ReDIF version 1 will be referred to as ReDIF throughout. ReDIF is a metadata format to describe the output aspects of academic disciplines. But that we mean the scientific documents that are produced, the channels through which they are made public, the authors who produce the documents and the editors who control the output channels, etc as well as the institutions that support this process. ReDIF does not aim at a very elaborate description of these items. Its overriding design goal is simplicity. It is aimed primarily for the use by academics as a self-documentation tool. The idea is that if academics can make a better documentation of their work themselves, then the need for expensive intermediation between academics is reduced. ReDIF is supposed to be understood and deployed by people with no formal cataloging training.

ReDIF is shipped with software—written in Perl—that allows to read and validate ReDIF data. This software is available at http://openlib.org/acmes/root/soft/ReDIF-Perl. For example, any date has to be of the form *yyyy-mm-dd*. `1999-07` is a valid date, whereas `Juillet 1999` is not a valid date. Therefore ReDIF is well suited for the decentralized maintenance of metadata. The validating software will make sure that only data that are syntactically correct will be passed on to user services. When subsequent versions will extend the capabilities of version 1 the data of version 1 will still be able to be understood by processing software that will implement future versions.

ReDIF is a *relational* metadata format. Each record has an identifier, also called a "handle". Elements in one record may use the identifier of another record. This allows elements to refer to other elements. A simple example should be sufficient. An author may have written several papers. She would like to have her email available within the data for each paper. But if the email address changes, there are changes required to all the records for all papers that the author has written. That is cumbersome. Instead, it is better to group all the personal data into one record, and use the handle of that record whenever personal data is required. The email address then only needs to be changed at one place. ReDIF explicitly allows for that.

In order to implement this handle structure, some human effort is required. This effort is being carried out by a group of people that we refer to as the "authority" in the following. RePEc is a well-know authority that uses ReDIF.

3 The ReDIF templates

3.1 The template types

ReDIF allows to describe things in the world that are important to the work of an academic discipline. We will use the word "items" to refer to any such things. ReDIF allows to catalogue three classes of items. First there are *resources*.

A resource is either a digital string of information or anything that can be digitized. A book for example may not be digital but can be digitized, and therefore it is a resource. A second group of items are *tangibles*. These are objects of descriptions that can not be digitized. A person for example falls into that category. Finally, there are collections. These are manifolds that have a mixed nature that makes it difficult to decide whether they are tangibles or resources. In the following table, we list the items, the version of the ReDIF template that describes it, a description of the item, and finally the class of item.

item name	Version	concern	class
ReDIF-paper	1.0	a research report that is not formally published	resource
ReDIF-article	1.0	a research report in a scientific journal	resource
ReDIF-chapter	1.0	a research report in a book	resource
ReDIF-book	1.0	a book	resource
ReDIF-software	1.0	an executable software or a software script	resource
ReDIF-archive	1.0	describes a archive where templates are stored	collection
ReDIF-series	1.0	describes a series of templates of the items above, e.g. a scientific journal	collection
ReDIF-institution	1.0	an institution, e.g. a research center	tangible
ReDIF-person	1.0	a person, who could be an author of a paper, editor of a journal etc.	tangible

3.2 The ReDIF syntax

ReDIF splits all the information concerning an item into a set of data elements. By data element we mean an elementary piece of information that we do not wish to split further. The whole set of data elements that describe an item forms the bibliographic record or "template" of this item. "Attribute" and "field" are two synonyms for a data element.

The ReDIF templates aim to compromise between human and machine readability. The general syntax for a field is

data_element_name:*data element value*

where *data_element_name* would for example be Author-Workplace-Phone and the data element value would be +44(0) 1483876958. Of course we could split the telephone number into an international dialing code, a city code and a local code, but that would go beyond what our data is actually needed for. The field name must be separated from the field value by a colon and optional whitespace. By a whitespace we mean a space character or a tabulation character or any combination of those.

Field names may contain only alphanumeric characters, the hyphen - and the hash sign #. No embedded spaces are allowed. Field names are case insensitive. Any field that starts with X- is considered to be local. The parsing software will read it but make no checks on the values.

Any field value may be continued (or even started) on the next line by adding a whitespace at the beginning of the continuing line. Multi-line field values are delimited by the first line which does not have a whitespace character in the first position, or is blank. Thus all lines containing field names should have no whitespace characters in the beginning of the line.

The default character set for field values is ISO-8859-1. This is close to what is called Latin-1 on most machines and close enough to the standard Microsoft Windows character set. It is also the character set that is used in the Hypertext Transfer Protocol (http).

Many fields may be repeated within one record, but repetitions are discouraged and for some fields they are prohibited. But for example, here is a person who works in two places

```
Author-Name: Sturm, J.F.
Author-Workplace-Name: Erasmus University Rotterdam,
        Econometric Institute
Author-Workplace-Name: Center for Economic Policy Research
```

3.3 Order of fields in the template

The general principle is that all fields may appear in any order within the template. However there are two important exceptions to this rule.

First, every ReDIF templates must start with the Template-type: field. The value of this field indicates what type of resource will be described by this template. Therefore, the set of expected fields in the template itself depends on that value. The values for the template type are given in Subsection 3.1. The version number must be quoted after the template type. The version number will make it easier to implement changes to ReDIF. At the moment all template types

are in version 1.0. For an example

```
Template-Type: ReDIF-Archive 1.0
```

Any data that does not start with a correct Template-Type: declaration will be ignored.
The second restriction of the freedom to group fields is necessary to express that certain pieces of information have to be kept together. Consider the case of a paper that has been co-authored by several people.

```
Author-Name:  Smith, Adam
Author-Name: Ricardo, David
Author-Email: Ricardo@classical.econ.org
Author-Email: Adam.Smith@classical.econ.org
Author-Workplace-Name: Institute of Classical Economics
```

Looking at this data you can easily find out that Smith's email address is Adam.Smith@classical.econ.org and that he probably works for the Institute of Classical Economics as Ricardo does. But computer is not able to find that out. We have to order the fields in such a way that the computer can understand which email address corresponds to which person. For each person, the name and the email address have to be kept together. There are other instances where we have to keep some fields in a group. We call such a group of fields a cluster.

4   Clusters

A cluster of fields is a group of fields within a template that belong together. Clusters are in fact like small templates within templates. For example, if we have a template of type "ReDIF-Paper 1.0" that describes a paper with several authors. It would contain several person clusters, one cluster for each author. If a document is available online, there should be one or several clusters with information about the files of the document—the main body of the text, a dataset that is used etc—as available on the Internet.
Each type of clusters has it's own field structure. Like templates, clusters must always start with one particular field, called the "key" field. Each cluster type has it's own key field. You should always keep one cluster's fields together.
Here is a simple cluster of type "Person" which describes an author

```
Author-Name: Smith, Joe
Author-Email: JoeSmith@some.uni.edu
Author-Postal: PO Box 123, Smith Street, Somewhere In The,
  Universe, 987654
Author-Phone: +99 456-321123
Author-Homepage: http://www.some.where.edu/~JoeSmith
```

Here is a cluster of type "Organization" describing a provider of a series of papers:

```
Provider-Name: Central Publishing House
Provider-Phone: +44-(0)207 226 7063
Provider-Homepage: http://www.central-publish.net
```

Clusters can be nested. For example, you can put an organization cluster to describe the organization an author is affiliated with, within the cluster of this particular author. Here we have an author who works at two institutions

```
Author-Name: Maria Saguvosky
Author-Workplace-Name: The New York Institute of Blasphemy
Author-Workplace-Homepage: http://www.new-york.blasphemy.org
Author-Workplace-Name: The New Jersey Institute for Blasphemic Research
```

Within this document, when we want to show you that you can use some cluster type a template, we will use a syntax like Author-(PERSON*), Provider-(ORGANIZATION*), File-(FILE*) etc. Thus the clusters below in this document often join a field prefix, e.g. "Author-", with a cluster type name, e.g. "(PERSON*)", between parenthesis.
When writing clusters you should always remember to put the key field first. For example, if you write

```
Template-Type: ReDIF-Paper 1.0
Title: The Capital
Author-Email: K.Marx@highgate.london.uk
Author-Name: Marx, Karl
Author-Phone: ...
```

the Author-Email: field is not recognized, because at that point the Author-(PERSON*) cluster has not yet started. This will result in an error message or in your template being ignored. But the Author-Name: field that follows will start the cluster, because "Name" is the key field of the (PERSON*) cluster. Hence the Author-Phone: field will be correctly processed.

This is all you need to know about clusters in general. We now address the specifics of the different cluster types.

4.1 The ORGANIZATION* cluster

The **ORGANIZATION*** cluster is defined by the following fields:

Name	Description
Name	Name of organization, **the key**
Homepage	URL of organization's home page
Name-English	English translation of the name of organization
Postal	Postal Address of organization
Location	Name of place where organization is based, e.g. "Harare"
Email	Electronic mail address(es) of organization
Phone	Telephone number(s) of organization
Fax	Fax number(s) of organization
Institution	Handle of an institution template that describes the organization

Name is the key field of the organization cluster. It should always be put first. Otherwise the most useful item is the homepage. Much of the remaining information can be derived from the homepage. It should be supplied if the homepage information is missing or the data at the homepage is not specific enough.

Here is an example of the organization cluster use to describe the provider of a working paper series.

```
Provider-Name:  Board of Governors of the Federal Reserve System
Provider-Homepage: http://www.bog.frb.fed.us/
Provider-Email: fedspapers@frb.gov
```

If a ReDIF institution template is available, then a better way to write to describe this institution is

```
Provider-Name:  Board of Governors of the Federal Reserve System
Provider-Institution: RePEc:edi:frbgvus
```

In that case, the data of the institution template is used to describe the institution. It is then no longer necessary to give any further details about the institution within the template. The advantage of this representation is that when there is a change in the data for the institution, it is no longer necessary to update the data for the institution within each (ORGANIZATION*) cluster within each (PERSON*) cluster of every author who has written a paper when affiliated with that institution.

4.2 The Person cluster

The **PERSON** cluster is used to describe a person. It has the following fields

Name	Description
Name	Name of person, **the key**
Homepage	URL of person's home page
Workplace-(ORGANIZATION*)	Organization in which person works
Email	Electronic mail address(es) of person
Fax	Fax number(s) of person
Postal	Postal address of person
Phone	Telephone contact information
Person	Handle of a person template that describes the person

Name is the key field of the person cluster. It should always be put first.

If a ReDIF person template for a person is available, then there is easy way to describe a person. This is done through the person template. For example

```
Author-Name: Thomas Krichel
Author-Person: RePEc:per:1965-06-05:thomas_krichel
```

Now all the information about the person, including the workplace data, is read from the person handle. No further information about the person needs to be entered.

4.3 The FILE cluster

The (FILE*) cluster is used to describe a file that contains the resource that is being described, or a part of it.

Name	*Description*
URL	Uniform Resource Locator of file, **the key**
Format	The format, e.g. "application/pdf" for PDF files
Function	explains the function of file within the resource
Size	The file size, by default in kilobytes
Restriction	A restriction placed on accessing the file

These fields of the (FILE*) cluster are not as the fields of the previous clusters. We describe each of them in more detail.

URL: This is the Uniform Resource Locator (URL) of the file on the Internet. The URL identifies an access protocol and a code that this protocol can use to retrieve the resource. Please refer to RFC1783 at ftp://ftp.isi.edu/in-notes/rfc1738.txt when composing a URL. Any whitespace in the URL will be ignored. Following RFC1783 any whitespace that follows a dash character "-" is considered to be a mistake, because it is likely to have been introduced by word processing software. The URL can be written over several lines, but do not break it after a dash.

```
File-URL:
  http://www.hhs.se/research/wpecofi/papers/wp0121.pdf.zip
File-URL: ftp://crefe.dse.uqam.ca/pub/cahiers/cah44.ps
```

This is the key field of the file cluster. It should always be put first.

Format: This is a an optional field to give a precise indication about the type of format that is found in the field. The field Format: is not repeatable and is case-insensitive. The contents of this field (all valid values) is restricted by a controlled vocabulary, the MIME Multipurpose Internet Mail Extensions types. These types are specified in the RFC1521 at ftp://ftp.isi.edu/in-notes/rfc1521.txt. The `complete and up-to-date list` of these types is maintained by the IANA. We do not support any values for file format that are not listed there. If you want to convey information about a type that is not registered here, please contact the IANA. The most important MIME types that are used by the ReDIF collections are `application/pdf` and `application/postscript`, as well `text/html`. If you simply want to give a message to the user about the file, it is best to use the Function: field.

Function: This is the function of the file. It should be used when a file represents only some part of the whole resource. The field value may indicate what particular part of the document is represented by the file. Here are three examples

```
File-Function: Appendix 1

File-Function: Main text

File-Function: Fortran program used for the simulations
 reported in the paper
```

If the file is a complete rendering of the resource, the File-Function field should not be used at all. When we have different files with no file function, user services should assume that all are alternative renderings of the complete text.

Size: The file size, by default in kilobytes. This is an optional and obsolete field. It is not repeatable. Examples

```
File-Size: 48
...
File-Size: 1788984 bytes
```

Restriction: A restriction on the retrieval of a file. If the file can be retrieved by anybody with Internet access without any preliminary payment or registration, you should **not** use that field at all. As soon as there is a restriction on the file many services will be assuming that there is no public access to the file. Some user services may wish to display only non-restricted files and your file will not appear there. If all files of a document are restricted, please use the field Restriction: in the template of the resource, rather than repeating it for each file of the resource.

Two examples for a complete file cluster

```
File-URL: http://rich.com/papers/d_duck.txt
File-Format: text/plain
File-Function: Abstract

File-URL: http://rich.com/papers/d_duck.pdf
File-Format: application/pdf
File-Function: Main text
File-Size: 13982 bytes
```

5 The Collection templates

The collection templates describe archives and series. We list the fields for these templates by order of importance, the most important is put first.

5.1 The Archive template

Template-type: ReDIF-Archive1.0
Each archive template must start with this statement.
Handle: This is a mandatory field for the archive handle. The handle of the archive is the name of the authority (for example RePEc), a : and then the *archive_identifier*. The archive identifier is proposed by the archive administrator and awarded by the authority. It should not have semantic contents.
URL: This is a mandatory field for the archive directory. The URL should either point to a http of ftp site. should end with the archive code.
Maintainer-Email: This is mandatory field for the email of somebody in charge of the archive. It is very important to keep it this field up-to-date. In particular, it is preferable not to use a personal email address of an employee that may move on later. It is best to give a general address of the sponsor of the archive. Make sure that the mail is read there and transmitted to the person who really maintains the archive.
Name: This is a mandatory field for the name of the archive. This is a name that will be shown to users when they access contents of the archive.
Maintainer-Name: This is an optional field for the name of a person that is in charge of the archive.
Maintainer-Phone: This is an optional field for the phone number of a person in charge of the archive.
Maintainer-Fax: This is an optional field for the fax number of a person in charge of the archive.
Classification-*scheme*: This is an optional field for subject information for the archive. Consult the same field in the Paper template and Appendix A for description.
Homepage: This is optional field for the URL of the archive's homepage. This should point to a location that an end user may be interested in.
Description: This is an optional field for a description of the archive's contents.
Notification: This is an optional field that may be used to specify how new papers in the archive are announced.
Restriction: This is an optional field to specify access restriction. It should be handle with care. A restriction mentioned here applies to all files referred to in all resources of all series in the archive. If the series can be retrieved by anybody with internet access, you should not use that field at all. If not all of the series that make up the archive are restricted, mention the restriction only in the series templates of the restricted series.
Here we have two examples for the archive templates

```
Template-type: ReDIF-Archive 1.0
Handle: RePEc:bob
Name: Economics Working Paper Archive at WUSTL
Maintainer-Email: bparks@wuecona.wustl.edu
URL: ftp://econwpa.wustl.edu/bob

Template-type: ReDIF-Archive 1.0
Handle: RePEc:wop
Name: WoPEc Project
Maintainer-Email: WoPEc@netec.mcc.ac.uk
Description: This archive collects information about holding of
 papers on those sites that have not (yet) joined RePEc.
URL: ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/wop
URL: http://netec.mcc.ac.uk/RePEc/wop/
```

5.2 The Series template

Template-type: ReDIF-Series1.0
Each series template must start with this statement.
Name: This is a mandatory field for the name of the series.
Handle: This is a mandatory field for the handle for the description of the series as made by the archive. The handle of the series is the name of the authority, a :, the handle of the archive, a :, and then the *series_identifier*. This field must not contain whitespace. It should not have semantic contents. The series code is awarded by the archive or the authority.
Maintainer-Email: This is a mandatory field for the email address of a person who is in charge of the series' files but

not necessarily of its contents. The maintainer should be ready receive error reports. As with the archive maintainer's email, a personal address should be avoided.

Type: Each series may only contain one type of templates. This rule is called "Baum's principle". The Type: field is used to indicate the type. The default type is "ReDIF-Paper". If the type of the elements of the series is "ReDIF-Paper", then this field does not need to be specified. Other legal series types are "ReDIF-Software", "ReDIF-Article", and "ReDIF-Book".

Order-Email: This is an optional field for the email address where to send ordering requests to.

Order-Homepage: This is an optional field for the URL of a screen where orders can be placed.

Order-Postal: Snail mail address where to order the paper.

Price: Price for all papers in the series when ordered. Can be overwritten by a separate Price: field in the paper

Provider-(ORGANIZATION*): This is the organization that supports the series.

Publisher-(ORGANIZATION*): This is a synonym for "provider". It is valid for historical reasons.

Restriction: A restriction on the retrieval of all papers in the series. It has the same meaning as in File cluster. It implies that all the files of all the documents in that particular series have this restriction.

Maintainer-Phone: This is an optional field for the phone number of the person in charge of the series

Maintainer-Fax: This is as optional field for the fax number of some person in charge of the series

Maintainer-Name: This is an optional field for the name of a person that is in charge of the files contained in the series.

Description: This can be a short description of the series' content.

Classification-*scheme*: This is an optional field for a subject for the series according to scheme *scheme*. Consult the same field in description of the paper template and Appendix A for description.

Keywords[-*scheme*]*: This is an optional field for keywords for the series according to scheme *scheme*. Consult the same field in the Paper template and Appendix A for description.

Editor-(PERSON*): This is the person responsible for the contents of the series.

Notification: This is an optional field to say how how new papers in the series are announced.

ISSN: This is the International Standard Serial Number for the series, if available.

Here are two examples for series templates.

```
Template-type: ReDIF-Series 1.0
Name: Computational Economics
Description: This is a series for economists who are computer nerds. You
 can not get more of an anorak than that.
Maintainer-Name: Bob Parks
Maintainer-Email: bparks@wuecona.wustl.edu
Handle: RePEc:bob:wuwpco

Template-type: ReDIF-Series 1.0
Name: Departmental Working Papers
Provider-Name: East Carolina University, Department of Economics
Provider-Homepage: http://ecuvax.cis.ecu.edu/~ecrothma/wp.htm
Maintainer-Email: ecrothma@ecuvax.cis.ecu.edu
Handle: RePEc:wop:eakjkl
```

5.3 The authority template

This is a draft template.
Template-type: ReDIF-Mirror1.0
Each authority template must start with this declaration.
Url: The url where the authority is based. Required
Handle: The handle of the authority. Required
Example:

```
Template-type: ReDIF-Authority 1.0
Url: http://netec.mcc.ac.uk/pub/RePEc
Handle: RePEc
```

5.4 The Mirror template

The mirror template describes what kind of mirror you are running on your machine. This template is obsolete and therefore no longer included here.

6 The Resource template types

6.1 The ReDIF-Paper Template

This template concerns all papers that are not published as part of a journal. We will be a bit more verbose here than in Section 6.2 where we describe other templates.

We will list here the elements of the paper template in decreasing order of importance. We start with those elements that are really necessary. We then list those element that are optional but important. Finally, we turn to those elements that are not important.

Handle: This is a required field to identify the bibliographic record. The Handle: field content starts with the name of the authority (organization), for example RePEc. The next element is the code of the archive, then follows the code of the series and finally the number of the paper within the series. All these parts are separated separated by the colon character, i.e. :. Note that this field may not contain whitespace. If the *Handle* is written over several lines, the processing software will eliminate the whitespace characters on the line boundaries. Use:

```
Handle: RePEc:bon:bonnsf:a452
```

Author-(PERSON*): This is a required cluster field to describe the person(s) who has authored the document, i.e. who is responsible for its intellectual content. We recommend the name of the author should be given in normalized format like "Lastname, Firstname" but the direct order is also acceptable. Please add the full first names if you know them, but omit titles, honors and others. If there are several authors, there should be a separate field for each author. The special author name "et al" should be used when refering to a list of unknown other names. These other authors should be placed in a separate author cluster.

The **Author-(PERSON*)** field is mandatory. That means that you should give at least one Author-Name: field with some value in each ReDIF-Paper template. Other information about the author is optional. Please do not overload the templates with author information. If you want to give elaborate personal information, register the person with a personal registration service. Use:

```
Author-Name: Lang, William
Author-Email: wlang@lsuvm.sncc.lsu.edu
Author-Workplace-Name: Center for Economic Performance
Author-Name: Sturm, J.F.
Author-Workplace-Name: Erasmus University Rotterdam,
 Econometric Institute Author-Workplace-Postal:
 P.O. Box 1738, NL- 3000 DR Rotterdam, The Netherlands
Author-Workplace-Homepage: http://www.eur.nl/few/ei/indexuk.html
Author-Workplace-Email: eb-webmaster@few.eur.nl
Author-Homepage: http://www.eur.nl/few/ei/stu.html
Author-Name: et al
```

Title: The title of the paper. This includes any subtitles. This is a required field. Use:

```
Title: A Theory of Gradual Trade Liberalisation
```

Creation-Date: The date at which the original document was created. The format should be as similar to the *ISO 8601:1988 Data elements and interchange formats—Information interchange—Representation of dates and times* as possible. This is *yyyy*[-*mm*[-*dd*]] where *yyyy* is the year, *mm* is the month and *dd* is the day, and the square brackets indicate that the element is optional. Note that the field should only contain the date of the creation, no other information. This field is not repeatable. Use:

```
Creation-Date: 1995-06-30
```

or

```
Creation-Date: 1996
```

FILE*: A file cluster for a component of the paper. If the paper is contained in only one file, then quote the URL for that file. If the paper is available in various files, the file cluster should be repeated. If there are files that do not contain the full text of the paper, the File-Function: field should be used to specify the role of the file in the document. Under *no circumstances* should the (FILE*) cluster be used to link to an intermediate page where the files can be retrieved from via further hyperlinks. Use Order-URL: for this purpose.

Order-URL: This is the URL of an intermediate page where there is ordering information about the file. The term "ordering information" should be taken in the widest possible sense. For example, web page with a set of links to

various manifestation of the paper is a good candidate for an Order-URL:. Note that the URL should be specific to the paper. If all the papers in the series have the same ordering URLs then that information should be put in the series template.

Classification-*scheme*: This is the classification number associated with the document. The **scheme** is a code for a classification scheme. If there are several classification numbers of the same classification scheme, they will be separated by a semicolon. This field is not repeatable, unless there are multiple schemes used. Allowed (registered) classification schemes and codes are listed in the Appendix. If you have documents, classified with some other scheme, just email us your suggestion, and we will register your scheme. Use:

```
Classification-JEL: C12; C30
```

Abstract: This is the abstract of the document. Although we do not require to have an abstract, but we strongly recommend to provide a detailed abstract, because it increases the chance of users finding your document within a full-text database. If you have a long abstract that has several paragraphs, you can leave a blank line. Use:

```
Abstract: This is the first paragraph of the abstract.

 This is the second paragraph of the abstract.
```

However, please recall that every line that follows the first line has to be indented with at least one blank.

Keywords[-*scheme***]**: These are keywords associated with the text. If there are several keywords they will be separated by a semicolon. *Scheme* is the keywords scheme's code but it can also be empty, to say that the keywords do not follow a scheme. Keywords: should be used in that case. If you have some keywords scheme, we will register it for you. Use:

```
Keywords: Competition; Consumer economics; Ethics; Philosophy of economics

Keywords: Accounting theory; Accounting principles; Financial accounting
```

Contact-Email: This field gives a contact email address for the paper. This does not need to be one of the authors' email address. In some cases where the author does not wish her email to be in the paper information or when no author email is available, this field should be used. Use:

```
Contact-Email: Economics_Secretariat@hicks.gross-uni.yy
```

Restriction: A restriction on the retrieval of a paper. This is similar to the field of the same name in the File cluster. If the paper can be retrieved by anybody with internet access, you should not use that field at all. Please do not use the field to make recommendations of the type "please use outside working hours", since as soon as there is a restriction on a paper many services will be assuming that there is no public access to it. If not all of the files that make up the paper are restricted, mention the restriction only in the (FILE*) cluster of the restricted.

The following elements are not important. They have been introduced to integrate legacy data into ReDIF-based system.

Note: Any other information that is relevant for the document. This can be its relation to other documents, a word or two about the history of the document etc. Use:

```
Note: This was an invited paper to a special issue of the
 Noddyland Journal of Blasphemy
```

Length: The length of a printed version of the document, usually in pages. This field is not repeatable. Use:

```
Length: 29 pages
```

Series: This is the series of the document. You can simply name the series here. This field is not repeatable.
Number: The number of the paper within the series. This field is not repeatable.
Availability: This field can be used to state if a paper is out of print. If the paper is on the net it should not be used. It can also be used to say how the get hold of the copy of the paper. This field is not repeatable. Use:

```
Availability: out of print
```

Revision-Date: A date at which the document was changed. The format for the value of that field is the same as in Creation-Date:. This field may be repeated.

```
Revision-Date: 1995-06-30
```

Price: That refers to the price of a printed copy of the document, when ordered through the ordering channels specified in the series template. Use:

```
Price: $3.00 Surface, $4.00 Air (U.S. $)
Price: 15 guilders by mail, free by ftp
```

For the price of an electronic copy, use the Restriction: field.

Publication-Status: This can be used to say if a modified version of the document exists, has been forthcoming, etc. in a commercial journal, book or other type of formal publication. It should always start with the word "published" or the word "forthcoming" (case insensitive). Use:

```
 Publication-Status: Forthcoming in
  Computational Statistics and Data
Publication-Status: Published in Journal of Artificial
   Intelligence Research, Vol.3
Publication-Status: Published in American Journal of
  Agricultural Economics, 1995, vol. 77, no. 1. pp. 120-134.
Publication-Status: Published by University of Arizona Press
```

Notification: to specify how new versions of the paper are announced.

Here are two examples of ReDIF templates of type "ReDIF-Paper 1.0". This is a simple document available only in PostScript format.

```
Template-Type: ReDIF-Paper 1.0
Author-Name:  David Currie
Author-Name:  Paul Levine
Author-Email:  p.levine@surrey.ac.uk
Author-Name:  Joeseph Pearlman
Author-Name:  Michael Chui
Title:  Phases of Imitation and Innovation in a North-South
  Endogenous Growth  Model
Abstract:  In this paper, we develop a North-South endogenous growth
  model to examine three phases of development in the South: imitation
  of Northern products, imitation and innovation and finally, innovation
  only. In particular, the model has the features of catching up (and
  potentially overtaking) which are of particular relevance to the
  Pacific Rim economies.  We show that the possible equilibria
  depend on cross-country assimilation effects and the ease of
  imitation.  We then apply the model to analyze the impact of R&D
  subsidies.  There are some clear global policy implications which
  emerge from our analysis.  Firstly, because subsidies to Southern
  innovation benefit the North as well, it is beneficial to the North
  to pay for some of these subsidies.  Secondly, because the ability
  of the South to assimilate Northern knowledge and innovate depends
  on Southern skills levels, the consequent spillover benefits on
  growth make the subsidizing of Southern education by the North
  particularly attractive.
Length:  26 pages
Creation-Date:  1996-07
File-URL: ftp://ftp.surrey.ac.uk/pub/econ/WorkingPapers/surrec9602.ps
File-Format: Application/postscript
```

This is an example of a document from the banking structure conference 1994 at the Federal Reserve Bank of Chicago. Although there is no published working paper series of that kind, we have created a imaginary series with code "fedhbs". The document is available as PostScript in two parts, or as a zipped PostScript file. Since there is no MIME type for the zipped files, we have left out the File-Function: field.

```
Template-Type: ReDIF-Paper 1.0
Author-Name: Joseph P.Huges
Author-Name: William Lang
Author-Name: Loretta J. Mester
Author-Name: Choon-Geol Moon
```

```
Title: Recovering Techonologies That Account for Generalized
       Managerial Preferences: An Application to Non-Risk-
       Neutral Banks
File-URL: ftp://test.frbchi.org/pub/bsc/doc13a.ps
File-Format: Application/PostScript
File-Function: Main Text
File-URL: ftp://test.frbchi.org/pub/bsc/doc13b.ps
File-Format: Application/PostScript
File-Function: Chart
File-URL: ftp://test.frbchi.org/pub/bsc/doc13ps.zip
File-Function: Archive of Main text and chart
Handle: RePEc:wop:fedhbs:_013
```

Note that only Author-Name, Title and the Handle are mandatory in a template. It is however highly recommended that you should also give a Creation-Date, Abstract, Classification and/or Keywords.

6.2 ReDIF templates for other documents

In this Subsection, we look at templates that describe other forms of scientific documents. Most of the fields are the same as in the ReDIF-Paper template and thus need no further explanation.

6.2.1 The ReDIF-Article Template

An article is something that appeared in a journal or is forthcoming in a journal. This is a draft template for such publications.

Template-Type: ReDIF-Article 1.0

The template must start with this field

Handle: Handle of journal article. This is a required field. It is of the same form as the handle of the paper template.

As far as the series identifier is concerned, only series of articles are admissible here. Note that a journal is a nothing else than a series of articles. It therefore has a series code like a paper series. It can be refereed to as a "journal code" in casual language, but the reader should keep in mind that as far as the structure of ReDIF is concerned the journal is nothing else than a series.

The *article_code* is a concatenation of qualifiers, separated by the underscore character _. Each qualifier is an *fieldvalue* pair. The *field* is represented by one letter only. This is best understood by an example.

RePEc:xxx:joinec:v19_y1985_i2_Q1_p67-84

The qualifiers currently defined are the following. We list a name for the qualifier, the letter of the qualifier, a description of values and a perl regular expression that that can be used to verify the syntax of the qualifier, as well as an optional additional check. All regular expressions are case insensitive.

Name: volume
Letter: v
Value: any sequence of digits
RegEx: v[1-9][0-9]*

Name: issue
Letter: i
Value: any sequence of digits
RegEx: i[1-9][0-9]*

Name: year
Letter: y
Value: any sequence of four digits
RegEx: y([1-9][0-9]{3})
Check: warn if $1 is higher then current year or earlier than 1500

Name: pages
Value: number of page where article starts - number of page where the article ends
RegEx: pS*([1-9][0-9]*)-S*([1-9][0-9]*)

Check: $1 should be smaller than $2

Here the S stands for "supplement", just in case that the paper has not been published in the main issue.

Name: issue
Value: which issue within the year. If the issue is a month, then we give the English three letter start of the month: JAN, FEB MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC. The numbers of the months are also allowed here, but that is discouraged. If the issue is a quarter, we use Q1, Q2 etc. If the issue has a start date, put that date like 04-01 for 1 April. Note that you need to put the - separator here because the date is otherwise mistaken to be the number of the issue.There is no need to put an end date for this issue. You can repeat the issue qualifier if the physical issue covers several logical issues. SPE can be added to say it is a special editions, and S can be postponed for a supplement.
RegEx: i:(($months\d*)($season[S]*)(Q[1234])(\d+)(\d\d-\d\d)) where
$month="(jan|feb|mar|apr|may|jun|jul|aug|sep|oct|nov|dec)" and
$season="(spr|sum|aut|win|spe)"
In fact the information thus encoded in the handle can also be encoded in separate fields. These are listed here. None of these fields is repeatable. They are fields that are specific to the article template, i.e. they do not appear in the paper template.
Journal: Name of the journal.
Year: Year of publication of the article.
Pages: The pages of the article in the journal in the form *start–end*, where *start* is the number of the first page and *end* is the number of the last page.
Price: Price of the article if mailed individually.
Volume: The volume number that the article appeared in.
Month: The month where the journal issue where the article is appeared in.
template-**Handle**: where *template* may take the values *Paper*, *Book* or *Chapter*. These handles can be quoted if the paper was published in several places.
Other fields are like the paper template except that the following are not valid: Length:, Series: Availability:, Revision-Date: and Article-Handle:.
Example:

```
Template-Type: ReDIF-Article 1.0
Title: Productivity Spillovers from FDI in the Uruguayan
    Manufacturing Sector
Author-Name: Kokko, Ari
Author-Workplace-Name: Dept. of Economics, Stockholm School
    of Economics
Author-Workplace-Postal: Stockholm School of Economics,
    P.O. Box 6501, 113 83 Stockholm, Sweden
Author-Name: Tansini, Ruben
Author-Name: Zejan, Mario
Author-Workplace-Name: Dept. of Economics, Stockholm School
    of Economics
Author-Workplace-Postal: Stockholm School of Economics,
    P.O. Box 6501, 113 83 Stockholm, Sweden
Journal: Journal of Development Studies
Pages: 602-611
Volume: 32
Year: 1996
Publication-Status: Published
Handle: RePEc:jou:devstu:v:32:y:1996:i:Q1:p:602-611
```

What happens if you wish to provide data about the formal publication, i.e., within a journal, of a paper for which you have ReDIF-Paper information. This is called Sune's problem. The basic status is that it is not possible for you to provide information about an article that you have published in a journal. It is for the journal or some agent of the journal to provide that information, because all the information about all the articles has to be kept together in once directory. However, you can give the Article handle, and a ReDIF literate individual should be able to find the information.

6.2.2 The ReDIF-Software Template

This is a draft template.
Template-Type: ReDIF-Software 1.0)
The template must start with this field.
Handle: This a required field. It uses the same syntax as the handle in the ReDIF-paper template.
Title: Name of the software, required, non-repeatable.
Programming-Language: This is a controlled vocabulary identifier for the language used. At the moment the only allowed value are "c", "c++", "fortran", "gauss", "mathematica", "matlab", "octave", "ox" and "perl", "rats", "shazam", "stata". In addition, the term "executable" may be used for directly executable files. All these identifiers are case insensitive. This is a required field.
Author-(PERSON*): Person cluster for each author, at least one cluster is required, see the paper template for guidance.
Abstract: Abstract to describe the principal function of the software.
Number: The version number of the software
Keywords: Keywords for the software
Size: Size of software, usually the number of lines of code.
Creation-Date, **Revision-Date**: as used is in the ReDIF-paper template.
Note: additional information
Requires: indicates that a certain package version, compiler, or operating environment is needed.
Example:

```
Template-Type: ReDIF-Software 1.0
Title: MKSTRSN: Stata modules to format Social Security number variables
Author-Name: William Gould
Author-WorkPlace-Name: Stata Corporation
Author-WorkPlace-Postal: Stata Corporation, 702 University Drive East,
  College Station, Texas 77840 USA
Author-Email: wgould@stata.com
Programming-Language: Stata
Abstract: mkstrsn and mkdashsn make string variables (without and with
 dashes,respectively) from a nine-digit variable containing a Social
 Security number.  The commands have the syntax mkstrsn newvar oldvar
  and mkdashsn newvar oldvar.
Series: Statistical Software Components
Number: S328601
Creation-Date: 19971212
Length: 39 lines
Classification-JEL: C87
File-URL: ftp://ftp.bc.edu/pub/user/baum/statal/mkstrsn.ado
File-Format: text/plain
File-Function: program code
File-URL: ftp://ftp.bc.edu/pub/user/baum/statal/mkdashsn.ado
File-Format: text/plain
File-Function: program code
Handle: RePEc:boc:bocode:S328601
```

6.2.3 The ReDIF-book Template

This is a draft template.
Template-Type: ReDIF-Book 1.0
The template must start with this field.
Title: Title of book, required, non-repeatable
Handle: This a required field. It uses the same syntax as the handle in the ReDIF-paper template.
Author-(PERSON*): Person cluster for each author, at least one cluster is required, see the paper template for guidance.
Publisher-(ORGANIZATION*): Organization cluster for Publisher, required.
Year: Year of publication, required for published books, non-repeatable.
Month: Month of publication, non-repeatable.
Volume: Volume in multi-volume works, non-repeatable
Edition: 2nd, 3rd etc., non-repeatable.

13

Series: If the book is part of a series, non-repeatable. For example Series: Springer Lecture Notes in Mathematics, volume 234

Editor-(PERSON*): Editor of series.

ISBN: ISBN of the book, non-repeatable.

Publication-Status: One of "Published" or "Forthcoming", non-repeatable. "Published" is assumed if field is not present.

Note: Any additional information

Abstract: Abstract.

Classification-*scheme*: same as for the ReDIF-paper template

Keywords[-*scheme***]**: same as for the ReDIF-paper template.

HasChapter: Chapter handle

template-**Handle**: where *template* may take the values *Paper*, *Article* or *Chapter*. These handles can be quoted if the paper was published in several places.

6.2.4   The ReDIF-Chapter Template

This template is for chapters with individual authors in a collection of papers. Examples are papers published in conference proceedings, reprint volumes and Festschrifts. It is a draft template.

Template-Type: ReDIF-Chapter 1.0

The template must start with this field.

Handle: This a required field. It uses the same syntax as the handle in the paper template.

Title: Title of chapter, required, non-repeatable.

Author-(PERSON*): Person cluster for each author, at least one cluster is required, see the paper template for guidance.

Abstract: Abstract.

Classification-*scheme*: non-repeatable

Keywords[-*scheme***]**: non-repeatable

Provider-(ORGANIZATION*): Organization cluster for Provider. This field required unless the field Sponsor-(ORGANIZATION*) is present.

Sponsor-(ORGANIZATION*): Organization cluster for sponsor (organization responsible for conference), required unless Provider-(ORGANIZATION*) is present.

Book-Title: Title of volume, required, non-repeatable.

Editor-(PERSON*): Editor cluster for each editor, at least one cluster is required.

Year: Year of publication, required for published papers, non-repeatable.

Month: Month of publication, non-repeatable.

Pages: Pages in volume, non-repeatable.

Chapter: Chapter in volume, non-repeatable.

Volume: For multi-volume works, non-repeatable.

Edition: 2nd, 3rd etc, non-repeatable.

Series: Series that volume is part of, non-repeatable.

ISBN: ISBN of volume, non-repeatable.

Publication-Status: One of 'Published' or 'Forthcoming', non-repeatable. Published is assumed if field is not present.

Note: Any additional information.

In-Book: Handle of the book that the chapter appeared in

Paper-Handle: Handle of working paper template, non- repeatable

template-**Handle**:

where *template* may take the values "Article", "Paper" or "Chapter" handle if published in several places.

Example:

```
Template-Type: ReDIF-Chapter 1.0
Title: Modelling Economic Relationships with Smooth
   Transition Regressions
Author-Name: Teräsvirta, Timo
Author-Workplace-Name: Department of Economic Statistics
Author-Workplace-Postal: Stockholm School of Economics, Box
   6501, 113 83 Stockholm, Sweden
Book-Title: Handbook of Applied Economic Statistics
Editor-Name: Giles, D.E.A.
Editor-Name: Ullah, A.
Publisher-Name: Dekker
```

```
Publication-Status: Forthcoming
Paper-Handle: RePEc:hhs:hastef:0131
Handle: RePEc:hhs:hastef:chp0131
```

7 ReDIF templates for physical entities

7.1 The person template

Template-Type: ReDIF-Person 1.0
The template must start with this field.
Handle: The handle of the person template. This is a mandatory field of the form
authority:*archive_identifier*: *yyyy-mm-dd*:*namestring*
where *authority* is the code of the authority, *archive_identifier* is the identifier of the archive, *yyyy*, *mm*, *dd* form a year, month and day of the month of a valid date. This date is called the significant date of the person. It is recommended to use the birthday, but any other date in the lifetime of the person is also admissible. Finally *namestring* is a name string that should resemble the person's name. Blanks is the name are replaced by the underscore character i.e. `_`.
Name-Full: The full name of the person. This field is mandatory.
Name-First: The first name of the person. This is useful when searching for the person's name in unstructured name data.
Name-Last: The first name of the person. This is useful when searching for the person's name in unstructured name data.
Email: The email of the person. This is a mandatory field.
Homepage: The URL of the homepage.
Author-Paper:, **Author-Article**:, **Author-Software**:
These fields say that the person is the author of a paper, article or software, respectively. The value of these fields must be the handle of a valid template.
Editor-Series: This field says that the person is an editor of a series. The value of this field must be the handle of a valid series template.
Classification-*scheme*: A subject classification code for the academic area that the person is working in, according to the scheme *scheme*.
Fax:, **Postal**:, **Phone**:, **Workplace-(ORGANIZATION*)**:, **Workplace-Institution**: are allowed and take the same meaning as in the person cluster.

7.2 The ReDIF-Institution template

This template concerns information about institutions. This information can be used in the other templates by a call to the appropriate institution handle.
Template-Type: ReDIF-Institution 1.0
Every record must start with this field.
The other elements are limted to the types of clusters are allowed in the template. These are refered to as
Primary-(ORGANIZATION*): The main organization.
Secondary-(ORGANIZATION*): The second level of the organisation, of example a department.
Tertiary-(ORGANIZATION*): The third level of the organisation, for example a group within a department.
Examples for the leveling of organizations are: "University of London/London School of Economics Financial Markets Group", or "Federal Reserve Bank of Minneapolis / Research Department", "Government of Zambia / Ministry of Finance". These three clusters are organized like the (ORGANIZATION*) cluster. An example for a complete organization cluster is:

```
Primary-Name: Universite des Grands Espoirs
Primary-Name-English: University of Grand Hopes
Primary-Location: Panava-les-Flots
Secondary-Name: Departement d'Economie
Secondary-Name-English: Department of Economics
Secondary-Email: eco@uge.edu Secondary-Homepage:
http://www.eco.uge.edu/
Secondary-Phone: (+567)3466356
```

An additional element is:
Handle: This has to be a 7 letter handle, the last two characters being the ISO 3166 country codes (us for the United States, ea foe associations and societies). Examples:

```
Handle: RePEc:edi:imfffus
Handle: RePEc:edi:mofgvzm
Handle: RePEc:fmg:fmlseuk
```

An example for a complete record is:

```
Template-Type: ReDIF-Institution 1.0
Primary-Name: Government of Bahrain
Primary-Location: Manama
Secondary-Name: Bahrain Monetary Agency
Secondary-Homepage: http://www.bma.gov.bh/
Handle: RePEc:edi:bmagvbh
```

A   Classification and keywords schemes

A.1   Classification scheme(s)

Classification-ACM-1998: The classification scheme used by the Association for Computing Machinery, in its version of 1998. Several codes may be given separated by colon, semi-colon or blanks.
Classification-ACM-1991: The classification scheme used by the Association for Computing Machinery, in its version of 1991. Several codes may be given separated by colon, semi-colon or blanks.
Classification-ACM-1964: The classification scheme used by the Association for Computing Machinery, in its version of 1964. Several codes may be given separated by colon, semi-colon or blanks.
Classification-Ila: The classification scheme proposed in the NASA TechReport TM-1998-208955.
Classification-JEL: The classification system of Journal of Economic Literature. It's usually used to classify economics texts. For more info see http://www.aeaweb.org/journal/elclasjn.html. Several code may be given separated by colon, semi-colon or blanks.
Classification-MSC-1991: The Mathematics classification scheme devised by the American Mathematical Society, in its version of 1991. Several codes may be given separated by colon, semi-colon or blanks.
Classification-MSC-2000: The Mathematics classification scheme devised by the American Mathematical Society, in its version of 2000. Several codes may be given separated by colon, semi-colon or blanks.

A.2   Keywords scheme(s)

Attent: This is a thesaurus (list of controlled keywords) which is used for the database "Attent: Research Memoranda". This database, as well as the thesaurus, is produced and maintained by Tilburg University Library. "Attent: Research Memoranda" only includes Economics working papers. The thesaurus includes economic and mathematical keywords. For further details contact Corry Stuyts <C.Stuyts@kub.nl>.