



中华人民共和国国家标准

GB/T 23731—2009/ISO 17933:2000

GEDI—通用电子文档交换

GEDI—Generic Electronic Document Interchange

(ISO 17933:2000, IDT)

2009-05-06 发布

2009-11-01 实施

数码防伪

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

中 华 人 民 共 和 国
国 家 标 准
GED1—通用电子文档交换
GB/T 23731—2009/ISO 17933:2000

*

中国标准出版社出版发行
北京复兴门外三里河北街16号
邮政编码:100045

网址 www.spc.net.cn

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 2.75 字数 77 千字
2009年8月第一版 2009年8月第一次印刷

*

书号: 155066·1-38256 定价 39.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68533533

目 次

前言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 符号和缩略语 2

5 服务模式和拓扑结构 2

5.1 概述 2

5.2 一般模型 3

5.3 域的概念 3

5.4 功能性要素 4

5.5 GEDI 拓扑结构 5

6 GEDI 记录格式的结构 6

7 GEDI 文件头信息 6

7.1 概述 6

7.2 GEDI 文件头数据元-语义 6

7.3 GEDI 文件头数据元-句法 18

7.4 按字母顺序排列的标号列表 21

7.5 GEDI 文件头的例子 22

8 电子文档格式 24

8.1 概述 24

8.2 文档格式标识符 24

9 文件传输机制 24

9.1 概述 24

9.2 文件名 24

9.3 FTP 介绍 24

9.4 FTP 传输纲要 24

9.5 支持 FTP 的协议栈 25

9.6 FTP 的命名和选址 25

10 邮件传输机制 25

10.1 概述 25

10.2 MIME 传输配置表 25

10.3 支持 MIME 的协议栈 28

11 一致性 28

11.1 发送/接收角色 28

11.2 GEDI 文件头数据元一致性 28

11.3 电子文档拷贝的一致性 28

11.4 协议一致性 28

附录 A (资料性附录) 馆际互借 APDU 到 GEDI 的映射	29
附录 B (资料性附录) 电子文档拷贝格式注册	33
参考文献	38

前 言

本标准等同采用国际标准 ISO 17933:2000《GEDI 通用电子文档交换》(英文版)。

本标准对 ISO 17933:2000 在文字上做了某些适合国情的修改,技术内容未作变动。

本标准的附录 A 和附录 B 是资料性附录。

本标准由全国信息与文献标准化技术委员会提出并归口。

本标准起草单位:中国化工信息中心。

本标准主要起草人:蔡志勇、张蓓、齐明、魏刚。

GEDI-通用电子文档交换

1 范围

本标准规定了在计算机系统之间交换电子文档拷贝的格式,既包括 GEDI 文件头定义,其中包含请求方和提供方信息,也包括文件的格式以及有关的书目信息。

本标准适用于支持馆际互借和文件传输请求的计算机系统。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 2659—2000 世界各国和地区名称代码

GB/T 7408—2005 数据元和交换格式 信息交换 日期和时间表示法

GB/T 23270.1—2009 信息和文献 开放系统互联 馆际互借应用协议规范 第1部分:协议说明书

GB/T 23270.2—2009 信息和文献 开放系统互联 馆际互借应用协议规范 第2部分:协议实施一致性声明(PICS)条文

ISO 2108:2005 信息和文献 国际标准书号(ISBN)

ISO 3297:2007 信息和文献 国际标准连续出版物编号(ISSN)

RFC 959 文件传输协议(FTP),1985年10月

3 术语和定义

下列术语和定义适用于本标准。

3.1

接收方 consumer

指以下应用进程:接收 GEDI 记录、处理 GEDI 文件头信息、使一个电子文档拷贝对最终用户可用。

3.2

域 domain

彼此之间可以进行电子文档交换事务的一个或多个提供方与一个或多个接收方的集合,相互之间存在一个共同的协议,包括:1) 电子文档交换格式和压缩规则,2) 电子文档传输机制,3) 网络技术。

3.3

电子文档拷贝 electronic document copy

包含电子文档拷贝的 GEDI 记录部分。

3.4

电子文档交换事务 electronic document interchange transaction

进行电子文档拷贝交换的完整周期,从电子文档储存于提供方开始,到该文档完全发送给接收方结束。

3.5

GEDI 域 GEDI domain

所包含的通用协议与本标准相一致的域。

3.6

GEDI 文件头 GEDI header; GEDI cove

GEDI 记录的第一部分,包括:1) GEDI 记录部分的格式和版本;2) 电子文档交换事务;3) 电子文档的书目信息描述;4) 电子文档拷贝的格式。

3.7

GEDI 记录 GEDI record

完整的 GEDI 信息,包括 GEDI 文件头和电子文档拷贝。

3.8

中继方 relay

指以下应用进程:在一个域内从提供方接收 GEDI 记录,在第二个域内传输给接收方。

3.9

提供方 supplier

指以下应用进程:获取一个电子文档拷贝,产生一个 GEDI 记录,将此记录传输给一个接收方,有可能通过一个或多个中继方。

4 符号和缩略语

FTP:文件传输协议(File Transfer Protocol)

JFIF:JPEG 文件交换格式(JPEG File Interchange Format)

JPEG:静止图像压缩标准(Joint Photographic Experts Group)

MIME:多用途因特网邮件扩充协议(Multipurpose Internet Mail Extensions)

PDF:便携式文档格式(Portable Document Format)

POP:邮局协议(Post Office Protocol)

RFC:请求注释(Request for Comment);Internet 标准或建议

SMTP:简单邮件传输协议(Simple Mail Transfer Protocol)

TIFF:标识图像文件格式(Tagged Image File Format)

5 服务模式和拓扑结构

5.1 概述

通用电子文档交换(GEDI)是电子形式的文档交换,本标准的重点在于以下两个方面:

- a) 电子文档格式的定义;
- b) 交换机制的描述。

交换仅仅是电子文档发送服务全过程中的一部分,因此 GEDI 的范围不如电子文档发送服务广泛。

除了以上两个 GEDI 要素外,为了提供完整的发送服务,电子文档完整发送还包括以下相关要素:

- a) 识别与定位:识别文档并确认原始资料的位置。可以通过在线访问联合目录(例如利用 ISO 23950),或者通过脱机服务如光盘或纸介质的目录实现;
- b) 订购:请求发送所要求的文档。其功能与发出馆际互借请求一致。在第 7 章所描述的 GEDI 文件头信息中,把馆际互借标准(GB/T 23270)作为文档识别的基础;
- c) 数字化:通过扫描设备将原始文档转换为电子数据;
- d) 交换:实现电子拷贝的实际转移;
- e) 原始文档再现:通过打印设备,图像文档被重新转回到纸或其他介质上;
- f) 编制帐单、结帐和其他管理程序。

所有要素在实际情况下可能以几种形式发生；在特殊情况下有些可能不发生。

其中交换是一个关键要素，因为它有效地完成了一个文档拷贝的物理移动。其他要素在以下特定情况下不发生：

- 基于普通常识进行识别和定位；
- 在主动发送原文时不产生订购请求；
- 通过直接的电子出版物，或通过扫描并存储，文档已经被数字化时不必再进行数字化；
- 当电子拷贝被保存到存储介质中，略过原始文档再现；
- 在参与者费用共享的合作服务中不发生编制帐单和结帐。

GEDI 协议主要集中于交换要素，为电子文档发送服务的发展提供通用基础，使不同的电子文档发送服务之间易于连接，其他要素的发展也将受益于交换部分所达成的国际协议。

在文档发送服务交换要素的范围内，形成本标准基础的全部模型是一个全球性模型：与 GEDI 相关的源信息（即文档图像）定位于全世界的各个地方；文档发送服务的目标客户分布也非常广泛。这个模型适合所有的资源和对象。

对于专用的解决方案，该模型认可参与文档发送的各组织之间的职责。一般来说，专用的解决方案反映了组织中各团体之间为优化彼此间服务而存在的协议。GEDI 模型不限制这种协议的可行性和自由度。而这个全部模型的目标是为它们的进一步发展提供通用基础和指导方针，提供这些不同团体之间互相工作的可能性。

5.2 一般模型

电子文档发送交换过程的一般模型如图 1。这个模型最主要的特征如下：

- a) 交换包括两方：提供方和接收方；
- b) 提供方和接收方之间通过必要设备连接，保证电子文档从提供方向接收方传输；
- c) 传输过程一次处理一个文档。

完整的交换循环，从电子文档存储于提供方开始，到该文档完全发送到接收方为止，称为通用电子文档交换事务。

图 1 所示的输入和输出功能并没有包括在通用电子文档交换事务中，这些功能确切的特性不属于本标准范畴。实际情况中，以下一些输入和输出形式是可用的：

- 输入可以是来自经过扫描（目前是最简单的方式）的原始文档、存储的文档图像文件或电子出版物；
- 输出可以采取把电子文档写入存储文件或打印的形式。执行这些可能的形式，有一些需要依法律和版权而定。

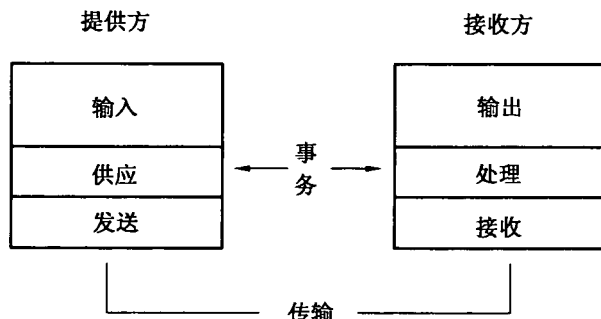


图 1 电子文档交换的一般模型

5.3 域的概念

域的概念将模型分成更小的部分，它承认专用域中、非公开职责之内的特殊解决方案，这种特殊解决方案或多或少的独立于公共的、国际的、GEDI 域的解决方案。仅在 GEDI 域内要求公共的协议；在专用域内可以遵循、也可以不遵循 GEDI 的规定。

各种专用域通过 GEDI 域中的服务相连接。在反映专用域域内组织结构的多种功能性网络模型基础上,专用服务一般是可用的。可以根据本标准的定义,规定在专用域和 GEDI 域交界处的中继方功能。

域的定义:彼此之间可以进行电子文档交换事务的一个或多个提供方与一个或多个接收方的集合,相互之间存在一个共同协议,用于:

- a) 电子文档交换格式和压缩算法;
- b) 电子文档传输机制;
- c) 网络技术。

域的协议不仅规定了采用的标准和机制,而且提供了适当的选择供其成员使用。在这一点上,域的成员达成了共识,这减少了域组成系统发展的复杂性。

实际上,无论域的协议是否遵照国际标准和 OSI 模型,它都覆盖了所有的通讯层。例如,某个特定域中的协议可能规定在 TCP/IP 之上的以 FTP 为基础的通讯,而另一个域也许会采用在 TCP/IP 之上的 MIME。

在应用层中,文档格式的协议与通讯一样,对域同等重要,它使提供方和接收方共享同一个电子文档图像,并且不需要转换和重新格式化。

利用域的概念,同一个域内的提供方和接收方就能够直接互联。相反,如果他们属于不同的域,根据两个域是否共享同一个协议,他们可能不能直接互联,但这两个被分开管理的域可能共享一个公用配置文件。

如果两个域不共享相同的协议,互联将通过应用中继方的功能来实现。应用中继方将在第一个域的协议下,采用传输机制接收 GEDI 记录,然后在第二个域的协议下把这个记录传输到第二个域。图 2 概述了应用中继方的任务。

在图 2 的模型中,通用电子文档交换事务仍然存在于提供方和接收方之间,中继方作为配角。

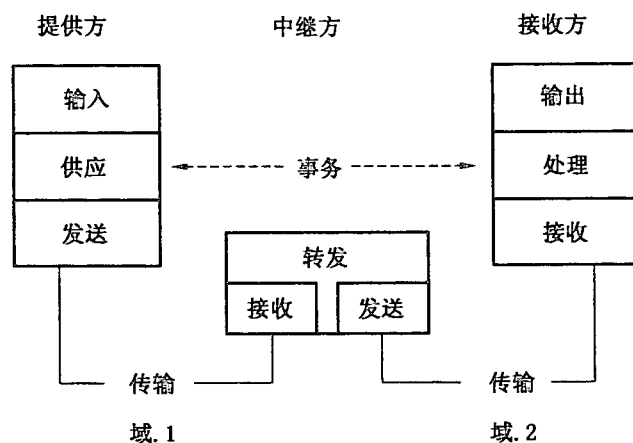


图 2 两个域间的交换

5.4 功能性要素

图 2 指出了所有跨越域边界的模型要素。如果两个域不共享公共的中继方,但都与第三个域有一个中继方,模型可以延伸更远,下一节展示了 GEDI 环境中的这种情况。

图 2 给出了功能性要素总的概况,以下是这些要素的特征:

a) 输入

负责生成提供方可用的电子文档拷贝。提供方仅仅采用协议的图像格式处理电子文档拷贝,在域电子文档拷贝格式协议范围内生成电子文档拷贝。例如:扫描原始文档;从存储设备中读取电子文档。

- b) 供应
采用输入生成的电子文档拷贝,增加事务相关信息,生成将被传输的数据结构。增加的信息被称为 GEDI 文件头信息,包括事务的标识(可能涉及馆际互借事务)、电子文档拷贝的鉴定,以及提供方和接收方的信息。供应功能还可以在实际传输电子文档拷贝之前将其临时保存,例如批量制作电子文档拷贝便于晚上进行传输。
- c) 发送
通过网络完成电子文档拷贝和 GEDI 文件头信息的实际发送。传输功能包括整个通讯栈,其中包括全部的应用程序以及发起方的更低层协议服务。
- d) 接收
与传输功能相对,从网络接收 GEDI 记录。采用与传输功能相同的通讯界面来执行,充当目标或应答者的角色。
- e) 处理
接收 GEDI 记录,通过分析 GEDI 文件头信息的内容决定应该采取的行动。一般来说,将把电子文档拷贝进行适当的输出,同时可以把 GEDI 文件头信息传给其他的应用程序,例如管理目的。
- f) 输出
负责所接收电子文档拷贝最后的再现或电子化归档。例如:在激光和其他打印机上打印电子文档拷贝;永久存储以备后用。
- g) 转发
提供两个域之间通讯的可能性,具有在两个域的协议之间转换的能力。实际上,它位于两个域的交界处,从一个域获得 GEDI 记录,向另一个域中发送 GEDI 记录。
- 上述功能要素在模型内部可以组合成下列主要模型实体:
- 提供方:包括输入、供应和发送功能的应用进程;
 - 接收方:包括接收、处理和输出功能的应用进程;
 - 中继方:包括在一个域内接收、转发和在另一个域内传输功能的应用进程。

5.5 GEDI 拓扑结构

在 GEDI 环境中存在一些专用域,这些专用域可以进一步细分为子域,例如地方性的服务连接起来形成全国性的服务。

可以在本标准的基础上发展专用域与 GEDI 域之间的中继方。
图 3 描述了通过两个中继方、分属两个不同专用域的提供方和接收方之间的通讯。

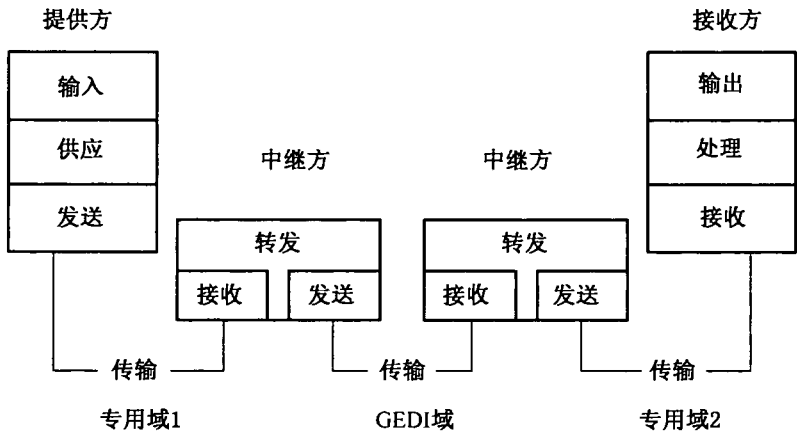


图 3 通过 GEDI 域之间的通信

前一小节描述了所有与电子文档交换有关的实体和功能要素,在实际情况下,这些功能要素将通过计算机系统的程序来执行,通过物理网络连接进行通讯。

6 GEDI 记录格式的结构

文档将以 GEDI 记录的形式进行交换,GEDI 记录格式包括两部分:

- a) GEDI 文件头(指示信息);
- b) 电子文档拷贝。

通过把 GEDI 文件头信息与电子文档拷贝分开,中继方不必读取文档的图像格式,这也使将来提供新的文档格式变得更容易,见图 4。

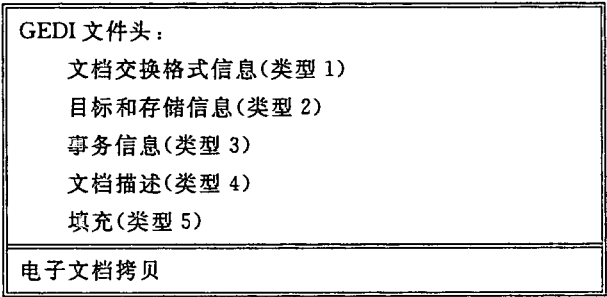


图 4 GEDI 记录交换格式

GEDI 记录在不同的域中可以通过几种不同的机制进行传输,例如:文件传输协议(如 FTP)或者电子邮件协议(如 MIME)。不同域之间中继方的存在,意味着单个 GEDI 记录在从一个域中的提供方转到另一个域中接收方的过程中,可以通过几种不同的机制进行传输。GEDI 记录的格式与传输机制无关,也没有假定任何的传输机制,但必须可以处理 8 位数据和进行错误探测。

7 GEDI 文件头信息

7.1 概述

GEDI 文件头信息分成 5 种类型。

- 类型 1:文档交换格式本身的辨认信息。
- 类型 2:命名、时间、目标文件和传输机制的存储信息。
- 类型 3:特殊电子文档发送事务的其他信息。
- 类型 4:文档的详细信息,包含简短的表述。
- 类型 5:在不改变 GEDI 文件头长度的情况下,考虑 GEDI 文件头将来的变化而进行的补充(可选)。

所有 GEDI 文件头数据元都是单独发生且不可重复的。提供方除了提供 GEDI 文件头数据元,还可以把实际的封面作为电子文档拷贝的一部分。

7.2 GEDI 文件头数据元-语义

7.2.1 简介

组成传输用 GEDI 文件头的的数据元数据如下所述。接收方系统可以将它们在 GEDI 记录的封面上进行适当打印。为了促进带有馆际互借请求的电子文档发送服务,任何情况下数据元尽可能按馆际互借协议标准 GB/T 23270 中的定义排列。附录 A 给出了馆际互借请求数据元与 GEDI 标号之间的映射信息,作为参考。

“结构”字段定义了数据元的数据类型。采用以下数据类型:

- 字符串型:ASCII20...7E;即所有的 ASCII 图形字符
- 数值型:ASCII30...39;即 0...9
- 字母数字型:ASCII 30...39 41...5A 61...7A;即 0...9 A...Z a...z

在服务字符串建议(SSAD)中,保留了一些专门用于结构标号的字符。

7.2.2 类型 1——文档交换格式信息

名称:交换格式标识符(interchange-format-id)

标号:IFID

语义:文件交换格式的明确标识符。既可以识别抽象句法,也可以识别编码。

结构:字符串型

状态:必备

最大长度:20

例如:GEDI

名称:交换格式版本(interchange-format-version)

标号:IFVR

语义:交换格式标号的版本号。

结构:字符串型

状态:必备

最大长度:20

例如:3.0

名称:文件头信息长度(cover-information-length)

标号:CILN

语义:GEDI 文件头信息的长度(包括 IFID 和 IFVR)。作为电子文档拷贝与 GEDI 记录的分界点。

它是一个十进制的数字,指定 GEDI 文件头信息的所有字节数。

结构:数值型

状态:必备

最大长度:10(包括可能的补充部分)

例如:123

名称:文档格式标识符(document-format-id)

标号:DFID

语义:GEDI 记录中电子文档拷贝格式的明确标识符。可能包括 OIDs 和 MIME 类型的值。最近注册的标识符见附录 B。

结构:字符串型

状态:必备

最大长度:20

例如:TIFF-5.0, TIFF-6.0, PDF-1.1

名称:服务字符串建议(service-string-advice)

标号:SSAD

语义:建议采用哪些字符作为类型 2、类型 3 和类型 4 数据元中的分隔符和指示符,以结构化的方式提供数据元信息。每个字符必须放在字符串的指定位置,次序如下:

第一个字符:释放指示符

第二个字符:各项之间的分隔符

第三个字符:新字段的前缀,在“=”之前总有占一个位置的、预先定义好的字母数字型字符

第四个字符:开括号,开始嵌套结构

第五个字符:闭括号,结束嵌套结构

4个字符用作划分子段且必须放在服务字符串内:

“;”在构造标号时作为项目的分隔符

“=”作为新子段的前缀,在“=”前总是有占一个位置的已定义的字母数字型字符

“(”作为开括号,开始嵌套结构

“)”作为闭括号,结束嵌套结构

注:这些字符取自 EDIFACT 编码

结构:字符串型

状态:必备

最大长度:50

例如:?=()

7.2.3 类型2——目标和存储信息

名称:接收方名称(Consumer-name)

标号:CNSN

语义:指定 GEDI 记录目标。包括首选传输机制和系统地址。是首选目标顺序列表,要求至少有一个值。

结构:字母数字型,结构化。名称(N=)、电子邮件(E=)、FTP 地址/目录(F=)、传真号(X=)。

FTP 必须细分成地址(A=)和目录(D=),通过重复子标号序列表示接收方名称的多次出现。

状态:必备

最大长度:250

例如:F=(A=12911004352;D=LGR. DOC);N=PICA

名称:记录名称(record-name)

标号:RCNM

语义:GEDI 记录的名称,由提供方指定明确的名称,必须遵守 9.2 所述的文件名称规则。

结构:字符串型

状态:必备

最大长度:32

例如:RUGOPC2232

名称:提供方名称(supplier-name)

标号:SPLN

语义:指定 GEDI 记录的来源,要求至少有一个值。

结构:字母数字型,结构化。名称(N=)、电子邮件(E=)、FTP 地址/目录(F=)、传真号(X=)。

FTP 必须细分成地址(A=)和目录(D=)。

状态:必备

最大长度:250

例如:F=(A=12911004352;D=LGR. DOC);N=RLG

名称:服务日期时间(service-date-time)

标号:SVDT

语义:提供方创建传输 GEDI 记录的日期和时间。

结构:数值型 YYYYMMDDHHMMSS(遵照 GB/T 7408)。

最大长度:14

例如:19930204122436

名称:系统服务标识符(system-service-id)

标号:SYID

语义:电子文档拷贝发送系统的标识。

结构:字符串型,非结构化。

状态:可选

最大长度:50

例如:DIS12.12

名称:系统服务地址(system-service-address)

标号:SYAD

语义:发送 GEDI 记录的地址;GB/T 23270 的系统地址。

结构:字符串型,结构化。电子邮件(E=)、打印位置(P=)、传真号(X=)。打印位置必须细分成部门(D=)、房间(R=)、打印机名称(P=)。

状态:可选

最大长度:100

例如:E=Devries@ubg.nl;P=(D=development;R=123;P=oakprntr)

名称:发送服务(delivery-service)

标号:DLVS

语义:发送服务或用于传送电子文档拷贝方法的名称或代码。

结构:字符串型,结构化。名称(N=)、电子邮件(E=)、FTP 地址/目录(F=)、传真号(X=)。FTP 必须进一步结构化,分成地址(A=)和目录(D=)。

状态:可选

最大长度:50

例如:F=(A=12911004352;D=LGR.DOC)

[GB/T 23270 定义:发送服务或用于传送文献方法的名称或代码(无变化)。]

名称:确认地址(conformation-address)

标号:CNFA

语义:用于确认 GEDI 记录在两个发送系统之间的传输地址。

结构:字符串型,结构化。名称(N=)、电子邮件(E=)、FTP 地址/目录(F=)、传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)。

状态:可选

最大长度:50

例如:F=(A=12911004352;D=LGR.DOC)

7.2.4 类型 3——事务信息

名称:优先级(priority)

标号:PRTY

语义:给予电子文档发送事务的优先级;0为最低,9为最高。

结构:数值型;数字 0-9。

状态:可选

最大长度:1

例如:0

名称:一般注释(general-note)

标号:GNLN

语义:从提供方到接收方的自由文本信息。

结构:字符串型,非结构化。

状态:可选

最大长度:600

例如:文章末页不确定

名称:客户名称(client-name)

标号:CLNT

语义:请求电子文档拷贝的读者名称。

结构:字符串型,结构化。电子邮件(E=)、名称(N=)。

状态:可选

最大长度:50

例如:E=devries@pica.nl;N=De Vries

[GB/T 23270 定义:请求项目的个人或机构的名称。]

名称:客户标识符(client-id)

标号:CLID

语义:请求电子文档拷贝读者的标号(例如图书馆读者编号)。

结构:字符串型,非结构化。

状态:可选

最大长度:25

例如:LIB1234567

[GB/T 23270 定义:用于唯一标识客户的数字或代码。]

名称:客户身份(client-status)

标号:CLST

语义:请求电子文档拷贝读者的身份。

结构:字符串型,结构化。国家代码遵照 GB/T 2659—2000 世界各国和地区名称代码(L=)、身份(S=)。

状态:可选

最大长度:25

例如:L=NL;S=Ing

[GB/T 23270 定义:客户的专业级别或职位。]

名称:个人或机构名称(name-of-person-or-institution)

标号:NPOI

语义:电子文档拷贝目的邮件地址中的名称。

结构:字符串型,非结构化。

状态:可选

最大长度:150

例如:中国国家图书馆

[GB/T 23270 定义:机构名称:表示一个图书馆、机构或公司的单词、短语或缩略语;个人名称:某个人常为人知的或被指定的一个单词或短语和/或姓名首字母,在馆际互借事务中,以此表示其人。]

名称:扩展的邮递地址(extended-postal-delivery-address)

标号:XPDA

语义:目的邮件地址中的附加信息。

结构:字符串型,非结构化。

状态:可选

最大长度:100

例如:馆际互借部门

名称:街道及门牌号(street-and-number)

标号:STNM

语义:目的邮件地址中的街道及门牌号。

结构:字符串型,非结构化。

状态:可选

最大长度:128

例如:Zhongguancun Nandajie99

[GB/T 23270 定义:表示城市或乡村中某一建筑物位置的号码和/或短语。]

名称:邮政信箱(post-office-box)

标号:POBX

语义:目的邮件地址中的邮政信箱。

结构:字符串型,非结构化。

状态:可选

最大长度:40

例如:北京 101 信箱

[GB/T 23270 定义:邮局分配的邮政信箱号。]

名称:城市(city)

标号:CITY

语义:目的邮件地址中的城市。

结构:字符串型,非结构化。

状态:可选

最大长度:128

例如:Beijing

[GB/T 23270 定义:用来表示城市、城镇或乡村的短语。]

名称:地区(region)

标号:REGN

语义:目的邮件地址中的地区。

结构:字符串型,非结构化。

状态:可选

最大长度:128

例如:朝阳区

[GB/T 23270 定义:表示省、州、地区或地点的短语。]

名称:国家(country)

标号:CNTR

语义:目的邮件地址中的国家。

结构:字符串型,非结构化。

状态:可选

最大长度:50

例如:中国

[GB/T 23270 定义:用来表示国家的短语。]

名称:邮政编码(postal-code)

标号:POCD

语义:目的邮件地址中的邮政编码。

结构:字符串型,非结构化。

状态:可选

最大长度:40

例如:100083

[GB/T 23270 定义:表示某个城市或其他地理区域里一个指定区域的代码。]

名称:请求方标识符(requester-id)

标号:RQID

语义:生成馆际互借请求的图书馆(系统)的标识信息,典型的是馆际互借办公室。

结构:字符串型,非结构化。

状态:可选

最大长度:25

例如:0019/0000

[GB/T 23270 定义:馆际互借事务请求方的确认信息。注意:请求方不总是一个图书馆!]

名称:请求方名称(requester-name)

标号:RQNM

语义:生成文档请求的图书馆(系统)的名称。

结构:字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP
必须细分成地址(A=)和目录(D=)。

状态:可选

最大长度:150

例如: N=RUU; E=webmaster@nlc.gov.cn

[GB/T 23270 定义: 请求方标号: 馆际互借事务请求方的确认信息。]

名称: 应答方标识符(responder-id)

标号: RSID

语义: 满足文档请求的图书馆(系统)的确认信息。

结构: 字符串型, 非结构化。

状态: 可选

最大长度: 25

例如: 0019/0000

[GB/T 23270 定义: 馆际互借事务应答方的确认信息。]

名称: 应答方名称(responder-name)

标号: RSNM

语义: 满足文档请求的图书馆(系统)的名称。

结构: 字符串型, 结构化。名称(N=)、电子邮件(E=)、FTP 地址/目录(F=)、传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)。

状态: 可选

最大长度: 150

例如: F=(A=12900045; D=KGH.DOC)

[GB/T 23270 定义: 应答方标号: 馆际互借事务应答方的确认信息。]

名称: 版权许可(copyright-compliance)

标号: CPRT

语义: 请求方符号, 请求方指明他所坚持的适用版权规范或法律。

结构: 字符串型, 结构化。代码(C=)、注释(N=)。

状态: 可选

最大长度: 150

例如: C=1; N=仅私人用

[GB/T 23270 定义: 请求方指明他所坚持的适用版权规范或法律。]

名称: 馆际互借事务标识符(ILL-transaction-id)

标号: ILTI

语义: 与电子文档发送事务相关的馆际互借事务信息的唯一标识。

结构: 字符串型, 结构化。符号(S=)、名称(N=)、组限定语(G=)、限定语(Q=)、子限定语(B=)。

状态: 可选

最大长度: $25 + 150 + 25 + 25 + 25 + (5 * 4) = 270$

例如: S=1200/0000; N=RUU; G=ION; Q=1234; B=1

[GB/T 23270 定义: 事务限定语: 确认与某个馆际互借事务相关的所有服务和消息的字母数字型字符串。它是由馆际互借事务的初始请求方赋值的唯一字符串, 并由馆际互借合作者应用于所有与馆际互借事务相关的服务和消息。与请求方标号和事物组限定语结合, 提供了馆际互借事务通用的唯一标识。]

名称:应答方注释(responder-note)

标号:RSNT

语义:来自馆际互借请求应答方的自由文本信息。

结构:字符串型,非结构化。

状态:可选

最大长度:600

例如:最后一页遗失

[GB/T 23270 定义:馆际互借事务应答方提供的注释。]

名称:接收控制(receive-control)

标号:RCN

语义:控制接收者对接收项目所采取行动的信息。

结构:字符串型,结构化。D(只能打印或删除)、F(禁止转发)、P(只能打印)、V(只能浏览)、X(如果转发则删除)。

状态:提供方可选,若存在则接收者必备。

最大长度:1

例如:D

7.2.5 类型4——文档描述

名称:著者(author)

标号:ATHR

语义:

结构:字符串型,非结构化。

状态:可选

最大长度:125

例如:James, E. R

[GB/T 23270 定义:对文献的知识内容或艺术内容负责的个人或机构的名称,包括文献的设计者、创作者、或原创者。]

名称:题名(title)

标号:TTLE

语义:指系列出版物、专著或其他各种文献的题目,文档取自这些文献。

结构:字符串型,非结构化。

状态:可选

最大长度:250

例如:Journal of the American Chemical Society

[GB/T 23270 定义:一个或一组单词,用来表示文献名称。]

名称:卷期(volume-issue)

标号:VLIS

语义:

结构:字符串型,结构化。卷(V=);期(I=);结合的(B=)。

状态:可选

最大长度:25

例如:V=1.2

[GB/T 23270 定义:连续出版物或多卷专著的一个物理单元的标识符/文献或文献各卷(分部)的一个单元的数字元、字母或单词。]

名称:文章著者(author-of-article)

标号:AART

语义:

结构:字符串型,非结构化。

状态:可选

最大长度:125

例如:Jones, Q. X.

[GB/T 23270 定义:文章的著者,该文章是另一文献的组成部分。]

名称:文章题名(title-of-article)

标号:TART

语义:正在传输文档的题目。

结构:字符串型,非结构化。

状态:可选

最大长度:250

例如:From Babel to EDIL-the evolution of a standard

[GB/T 23270 定义:文献一个组成部分的题名。]

名称:ISBN

标号:ISBN

语义:国际标准书号,定义见 ISO 2108:2005。

结构:字母数字型

状态:可选

最大长度:17

例如:978-89-92963-12-1

[GB/T 23270 定义:按照 ISO 2108:2005 分配给专著的国际标准书号。]

名称:ISSN

标号:ISSN

语义:国际标准连续出版物号,定义见 GB/T 9999—2001 中国标准连续出版物号。

结构:字母数字型

状态:可选

最大长度:8

例如:10504648

[GB/T 23270 定义:按照 ISO 3297:2007 分配给连续出版物的国际标准连续出版物号。]

名称:国内出版物号

标号:CSSN

语义:国内出版物的编号,可以是国内统一连续出版物号,也可以是其他编号。

结构:字母数字型

状态:可选

最大长度:8

例如:CN113163

名称:页码(page-numbers)

标号:PGNS

语义:文档电子拷贝中包含原始文档的页码范围。

结构:字符串型,非结构化。

状态:可选

最大长度:100

例如:1-23,36

名称:扫描日期(date-scanned)

标号:DTSC

语义:文档电子拷贝制作的日期;不同于电子拷贝放在文档交换格式记录中的日期,后者是必备的。

结构:数值型;年月日时分秒 YYYYMMDDHHMMSS(GB/T 7408—2005)。

状态:可选

最大长度:14

例如:19930101123554

名称:页数(number-of-pages)

标号:NMPG

语义:电子文档拷贝中的总页数。

结构:数值型

状态:可选

最大长度:5

例如:123

名称:索取号(call-number)

标号:CLNO

语义:见馆际互借请求的文献标号(索取号)。

结构:字符串型,结构化的。期刊(J=)、书(B=)、报告(R=)和未知的(U=)。

状态:可选

最大长度:5

例如:J=12.9

名称:分卷(册)出版日期(publication-date-of-component)

标号:PDOC

语义:见馆际互借请求的文献标号(分卷(册)出版日期)。

结构:字符串型,非结构化。

状态:可选

最大长度:25

例如:1992

名称:出版日期(publication-date)

标号:PUBD

语义:见馆际互借请求的文献标号(出版日期)。

结构:字符串型,非结构化。

状态:可选

最大长度:25

例如:2-03-93

名称:出版地点(place-of-publication)

标号:PLPB

语义:见馆际互借请求的文献标号(出版地点)。

结构:字符串型,非结构化。

状态:可选

最大长度:128

例如:北京

名称:出版者(publisher)

标号:PUBL

语义:见馆际互借请求的文献标号(出版者)。

结构:字符串型,非结构化。

状态:可选

最大长度:60

例如:北京图书馆出版社

名称:版本(edition)

标号:EDIT

语义:见馆际互借请求的文献标号(版本)。

结构:字符串型,非结构化。

状态:可选

最大长度:25

例如:12A

名称:原始请求(request-as-quoted)

标号:RQAQ

语义:请求方所请求的详细信息。

结构:字符串型,非结构化。

状态:可选

最大长度:600

例如:Jnl Am Chem Soc. 1993 642(3)3412-3417 Smith and Jones

名称:版权声明(copyright-statement)

标号:STAT

语义:来自提供方的自由文本信息,详细描述了由于版权问题使用所发送文献的限制。

结构:字符串型,非结构化。

状态:可选,但是一旦包括,就必须将声明与文献一起打印或一起在屏幕上显示

最大长度:600

例如:除了版权法允许的情况,未经版权人或其授权许可机构的许可,不允许对该文献进行进一步复制(包括通过电子方式存储于任意介质)。

名称:文献标识符(Item Id)

标号:ITID

语义:文献的标准标识符,例如 BICI,DOI,SICI,URN。

结构:字符串型。标识符的类型(T=)、标识符的值(V=)。

状态:可选

最大长度:200

例如:T=SICI;V=0002-8231(199412)45:10<737:T10DIM>2.33tx;2-m

7.2.6 类型5——填充

名称:填充(zpadding)

标号:ZPAD

语义:该字段是为整理 GEDI 文件头保留的空间,通过填充允许系统改变 GEDI 文件头中的个别数据元长度,同时保持整个文件头长度不变。填充的长度根据个别数据元的长度变化引起的必须补偿的位数进行变化,取值没有实际意义。

结构:无

状态:可选

最大长度:8K

7.3 GEDI 文件头数据元-句法

每个数据元的结构为:〈标号〉〈长度〉〈值〉。“标号”是易于记忆的字母形式的 8 位 ASCII 字符串,长度固定,为 4 个 8 位字符串,大小写均可。“长度”为十进制的整数,用 4 个 ASCII 数值型字符表示,长度固定。“值”是 8 位 ASCII 符号型字符。“标号”、“长度”、与“值”相互之间均不能插入任何字符或空格。

表 1 至 5 列出了 5 种不同的类型。

表 1 类型 1——文档交换格式信息

标 号	名 称	必备/可选	最大长度	结 构
IFID	交换格式标识符	必备	20	字符串型
IFVR	交换格式版本	必备	20	字符串型
CILN	文件头信息长度	必备	10	数值型
DFID	文档格式标识符	必备	20	字符串型
SSAD	服务字符串建议	必备	50	字符串型

表 2 类型 2——目标和存储信息

标号	名 称	必备/可选	最大长度	结 构
CNSN	接收方名称	必备	250	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)
RCNM	记录名称	必备	32	字符串型
SPLN	提供方名称	必备	250	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)
SVDT	服务日期时间	必备	14	数值型 YYYYMMDDHHMMSS(遵照 GB/T 7408)
SYID	系统服务标识符	可选	50	字符串型,非结构化
SYAD	系统服务地址	可选	100	字符串型,结构化。电子邮件(E=);打印位置(P=);传真号(X=);打印位置必须细分成部门(D=),房间(R=),打印机名称(P=)
DLVS	发送服务	可选	50	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)
CNFA	确认地址	可选	50	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须细分成地址(A=)和目录(D=)

表 3 类型 3——事务信息

标号	名 称	必备/可选	最大长度	结 构
PRTY	优先级	可选	1	数值型;数字 0-9
GNLN	一般注释	可选	600	字符串型,非结构化
CLNT	客户名称	可选	50	字符串型,非结构化。电子邮件(E=);名称(N=)
CLID	客户标识符	可选	25	字符串型,非结构化
CLST	客户身份	可选	25	字符串型,结构化。国家代码遵照 GB/T 2659—2000 世界各国和地区名称代码(L=);身份(S=)
NPOI	个人或机构名称	可选	150	字符串型,非结构化
XPDA	扩展的邮递地址	可选	100	字符串型,非结构化
STNM	街道及门牌号	可选	128	字符串型,非结构化
POBX	邮政信箱	可选	40	字符串型,非结构化
CITY	城市	可选	128	字符串型,非结构化
REGN	地区	可选	128	字符串型,非结构化
CNTR	国家	可选	50	字符串型,非结构化
POCD	邮政编码	可选	40	字符串型,非结构化
RQID	请求方标识符	可选	25	字符串型,非结构化
RQNM	请求方名称	可选	150	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须进一步结构化,分成地址(A=)和目录(D=)

表 3 (续)

标号	名 称	必备/可选	最大长度	结 构
RSID	应答方标识符	可选	25	字符串型,非结构化
RSNM	应答方名称	可选	150	字符串型,结构化。名称(N=);电子邮件(E=);FTP 地址/目录(F=);传真号(X=)。FTP 必须进一步结构化,分成地址(A=)和目录(D=)
CPRT	版权许可	可选	150	特征字符串,结构化。代码(C=);注释(N=)
ILTI	馆际互借事务标识符	可选	270	字符串型,结构化。符号(S=);名称(N=);组限定语(G=);限定语(Q=);子限定语(B=)
RSNT	应答方注释	可选	600	字符串型,非结构化
RCON	接收控制	可选	1	字符串型,结构化。D(只能打印或删除),F(禁止转发),P(只能打印),V(只能浏览),X(如果转发则删除)

表 4 类型 4——文档描述

标号	名 称	必备/可选	最大长度	结 构
ATHR	著者	可选	125	字符串型,非结构化
TTLE	题名	可选	250	字符串型,非结构化
VLIS	卷期	可选	25	字符串型,结构化。卷(V=);期(I=);结合的(B=)
AART	文章著者	可选	125	字符串型,非结构化
TART	文章题名	可选	250	字符串型,非结构化
ISBN	ISBN	可选	17	字母-数值型
ISSN	ISSN	可选	8	字母-数值型
CSSN	国内出版物号	可选	8	字母-数值型
PGNS	页码	可选	100	字符串型,非结构化
DTSC	扫描日期	可选	14	数值型,年月日时分秒 YYYYMMDDHHMMSS(遵从 GB/T 7408)
NMPG	页数	可选	5	数值型,非结构化
CLNO	索取号	可选	50	字符串型,非结构化
PDOC	分卷(册)出版日期	可选	25	字符串型,非结构化
PUBD	出版日期	可选	25	字符串型,非结构化
PLPB	出版地点	可选	128	字符串型,非结构化
PUBL	出版者	可选	50	字符串型,非结构化
EDIT	版本	可选	25	字符串型,非结构化
RQAQ	原始请求	可选	600	字符串型,非结构化
STAT	版权声明	可选	600	字符串型,非结构化
ITID	文献标识符	可选	200	字符串型,标识符的类型(T=);标识符的值(V=)

表 5 类型 5——填充

标 号	名 称	必备/可选	最大长度	结 构
ZPAD	填充	可选	8k	无

7.4 按字母顺序排列的标号列表

表 6

标 号	类 型	名 称
AART	4	文章著者(author-of-article)
ATHR	4	著者(author-of-article)
CILN	1	文件头信息长度(cover-information-length)
CITY	3	城市(city)
CLNO	4	索取号(call-number)
CLID	3	客户标识符(client-id)
CLNT	3	客户名称(client-name)
CLST	3	客户身份(client-status)
CNFA	2	确认地址(confirmation-address)
CNSN	2	接收方名称(Consumer-name)
CNTR	3	国家(Country)
CPRT	3	版权许可(copyright-compliance)
CSSN	4	国内出版物号
DFID	1	文档格式标识符(document-format-id)
DLVS	2	发送服务(delivery-service)
DTSC	4	扫描日期(date-scanned)
EDIT	4	版本(Edition)
GNLN	3	一般注释(general-note)
IFID	1	交换格式标识符(interchange-format-id)
IFVR	1	交换格式版本(interchange-format-version)
ILTI	3	馆际互借事务标识符(ILL-transaction-id)
ISBN	4	ISBN
ISSN	4	ISSN
ITID	4	文献标识符(Item-id)
NMPG	4	页数(number-of-pages)
NPOI	3	个人或机构名称(name-of-person-or-institution)
PDOC	4	分卷(册)出版日期(publication-date-of-component)
PGNS	4	页码(page-numbers)
PLPB	4	出版地点(place-of-publication)
POBX	3	邮政信箱(post-office-box)
POCD	3	邮政编码(postal-code)

表 6 (续)

标 号	类 型	名 称
PUBD	4	出版日期(publication-date-of-component)
PUBL	4	出版者(Publisher)
PRTY	3	优先级(Priority)
RCNM	2	记录名称(record-name)
RCON	3	接收控制(receive-control)
REGN	3	地区(Region)
RQAQ	4	原始请求(request-as-quoted)
RQID	3	请求方标识符(requester-id)
RQNM	3	请求方名称(requester-name)
RSID	3	应答方标识符(responder-id)
RSNM	3	应答方名称(responder-name)
RSNT	3	应答方注释(responder-note)
SPLN	2	提供方名称(supplier-name)
SSAD	1	服务字符串建议(service-string-advice)
STAT	4	版权声明(copyright-statement)
STNM	3	街道及门牌号(street-and-number)
SVDT	2	服务日期时间(service-date-time)
SYAD	2	系统服务地址(system-service-address)
SYID	2	系统服务标识符(system-service-id)
TART	4	文章题名(title-of-article)
TTLE	4	题名(Title)
VLIS	4	卷期(volume-issue)
XPDA	3	扩展的邮递地址(extended-postal-delivery-address)
ZPAD	5	填充(Zpadding)

7.5 GEDI 文件头的例子

IFID0004GEDIIFVS00032.0CILN00042048DFID0008TIFF-6.0SSAD0005?;=()CNSN0006N =
 PICARCNM0008RLG00001SPLN0005N = RLGSVDT001419910802140600TTLE0012PC/Computin-
 gAART0013Paul TomersonTART0031The DOS you've been waiting forZPAD1860

表 7

标号	IFID
长度	0004
值	GED I

标号	IFVS
长度	0003
值	2.0

表 7 (续)

标号	CILN
长度	0004
值	2048
标号	DFID
长度	0008
值	TIFF-6.0
标号	SSAD
长度	0005
值	?;=()
标号	CNSN
长度	0006
值	N=PICA
标号	RCNM
长度	0008
值	RLG00001
标号	SPLN
长度	0005
值	N=RLG
标号	SVDI
长度	0014
值	19910802140600
标号	TTLE
长度	0012
值	PC/Computing
标号	AART
长度	0013
值	Paul Somerson
标号	TART
长度	0031
值	The DOS you've been waiting for
标号	ZPAD
长度	1860
值	[空白]

8 电子文档格式

8.1 概述

GEDI 支持各种格式的电子形式文档,例如 TIFF、PDF 和 JPEG。GEDI 记录结构的设计易于适应将来附加的文件表达格式,例如 ODA 或 SGML。

8.2 文档格式标识符

GEDI 文件头中的文档格式标识符(DFID)字段规定了任意给定 GEDI 记录中电子文档拷贝的格式。注册的文档格式及相应的 DFID 值见附录 B。

9 文件传输机制

9.1 概述

本标准的目的是使文档以单个文件(“GEDI 记录”)在传输者和接收者之间交换成为可能。为此, GEDI 支持已存在的标准、协议和选自现有国内、国际标准团体的量表,用于简单文件传输。

选择的文件传输机制是 Internet FTP 协议。

标准应该尽可能独立于所采用的特定传输机制,但选择机制时还必须规定某些附加特性,便于为正在讨论中的协议定义一个纲要。以下详细说明 FTP 的这种纲要。

9.2 文件名

为避免可能的冲突,文件必须用唯一的文件名进行交换。从提供方到接收方,都应该一直使用这些文件名。因此,他们必须可用于多数计算机。

文件名必须符合以下规则:

- 8 个字符长度,后面可以跟分隔符“.”及 3 个字符的扩展名;
- 除了分隔符,只能包含大写字母(A,B,C,...Z)和数字(0,1,2...9);
- 由唯一的系统 ID 后接 1 个序号组成,序号由提供方进行选择,保证在一段时期内的唯一性,避免不明确的重复使用。

可接受一种基于 IP 地址形式的文件名,这种类型文件名必须满足以下条件:

- 12 字符长度;
- 前 8 个字符是十六进制的 IP 地址;
- 第九个字符是分隔符“.”;
- 最后 3 个字符是序号。

9.3 FTP 介绍

传输者作为客户端,初始化控制连接并发送文件。接收者作为服务器,监测外来的控制连接,包括接收存储命令(STOR)和从客户端接收文件。

9.4 FTP 传输纲要

以下黑体表示的部分对 GEDI 应用进程是非常重要的。

使用 RFC 959 文件传输协议 5.1 中的最小实施方案、3.1.1.3 中的图像数据类型和口令(PASS)及分配(ALLO)命令。GEDI 文件总是以图像(IMAGE)类型、流(STREAM)传输模式和文件式(FILE)结构被发送。

类型-非打印 ASCII、图像¹⁾

模式-流

结构-文件式,记录

命令-

1) 对 RFC 959 条款 5.1 中所定义的最小实施方案的补充。

- 用户名(USER)
- 退出登录(QUIT)
- 数据端口(PORT)
- 非打印 ASCII 和图像 的表示类型(TYPE)
- 传输模式(MODE)
- 文件式和记录的文件结构(STRU)
- 获取(RETR)
- 保存(STOR)

- 空操作
- 口令(PASS)
- 分配(ALLO)

9.5 支持 FTP 的协议栈

FTP 以 TCP/IP 为基础运行。

9.6 FTP 的命名和选址

在各个系统的主机地址,FTP 使用通用端口,命令使用端口 21,数据使用端口 20。

10 邮件传输机制

10.1 概述

当 GEDI 记录通过电子邮件发送时,整个 GEDI 记录包含在一个简单的邮件消息中。在此消息内,机读 GEDI 文件头作为一个单独部分,与可选人工可读 GEDI 文件头部分及文档拷贝部分相互分离。

表 8

A 部分	GEDI 文件头(必备)
A'部分	人工可读 GEDI 文件头(可选)
B 部分	文档拷贝(必备)

采用这种建议的格式,消息中原始的内容类型(Content-type)GEDI 文件头可明确指明这是一条 GEDI 记录,有助于接收方利用专门的 GEDI 浏览软件来显示/打印/处理文档。

当一个请求不能识别新的 GEDI MIME 类型时,它将按 multipart/mixed 型消息处理;对任意一个组成部分不理解时,允许接收方指定一个文件存储它们。

10.2 MIME 传输配置表

总体消息内容类型 GEDI 文件头如下:

内容类型: multipart/gedi-record

没有定义参数。

10.2.1 必备 GEDI 文件头部分

第一部分是机读 GEDI 文件头,它是如第 6 章和第 7 章所定义的句法形式的 GEDI 文件头。机读形式是一种新的 MIME 内容类型: application/gedi-header,它允许使用 GEDI 特定软件,并规定采用 GEDI 文件头句法形式。

内容传输编码(Content-transfer-encoding):由于不使用换字符,因此在 GEDI 文件头句法中使用“可打印引语(quoted-printable)”编码,由此生成大于 76 个字符的行。该部分中内容类型的行可以有一个参数“CHARSET=”;如果本部分包含 US-ASCII 字符集以外的字符,该参数是必备的。

如果不存在可选人工可读 GEDI 文件头,MIME 消息的形式如表 6 所示。

表 9 无人工可读 GEDI 文件头的 GEDI 记录 MIME 结构

GEDI 文件头部分	Content-type : multipart/gedi-record; boundary="⟨unique-boundary1⟩" --⟨unique-boundary1⟩
	Content-type : application/gedi-header; Content-transfer-encoding: quoted-printable ⟨gedi header⟩
电子文档部分	--⟨unique-boundary1⟩
	Content-type : image/tiff Content-transfer-encoding: base64 ⟨the document⟩ --⟨unique-boundary1⟩--

10.2.2 可选人工可读 GEDI 文件头部分

人工可读部分是可选的,如果存在的话,采用与内容相匹配的 MIME 媒体类型,也就是说由消息的发布者决定。最有可能是带有适宜字符集的文本/无格式(text/plain)型,因此仅以文本/无格式型为例子。如果使用非 US-ASCII 字符和/或其中有超过 76 个字符的行,则需要采用“可打印引语”而不是“基础 64(base64)”进行编码,保证人工可读形式最好的显示。

与 GEDI 文件头不同,本标准没有规定人工可读 GEDI 文件头的内容和格式,而是留待实践去判断,也无法对其作出任何假定。

如果出现可选人工可读 GEDI 文件头,GEDI 文件头和人工可读 GEDI 文件头都将包含在内容类型为 multipart/mixed 的部分中,如表 7 所示。

表 10 带有人工可读封面的 GEDI 记录 MIME 结构

GEDI 文件头部分	Content-type : multipart/gedi-record; boundary="⟨unique-boundary1⟩" --⟨unique-boundary1⟩
	Content-type : multipart/mixed; boundary="⟨unique-boundary2⟩" --⟨unique-boundary2⟩
人工可读部分	Content-type : application/gedi-GEDI Header Content-transfer-encoding: quoted-printable ⟨GEDI Header⟩
	--⟨unique-boundary2⟩
电子文档部分	Content-type: text/plain ⟨human readable GEDI Header⟩ --⟨unique-boundary2⟩
	--⟨unique-boundary1⟩
电子文档部分	Content-type : image/tiff Content-transfer-encoding: base64 ⟨the document⟩
	--⟨unique-boundary1⟩--

10.2.3 必备电子文档拷贝部分(必备的)

传输电子文档拷贝的消息由一个或多个部分组成。

可以通过一个单独的部分传输整个文件,文件片断也可以包含在多个部分中。
有些格式,例如多页的 TIFF 格式,其中包含页码的确定和组织的创建,所以无需每一部分都要求页码。

内容类型和内容交换编码与文件的格式相匹配,如 TIFF、JPEG、PDF。

虽然多数情况下所有的页都用相同的内容类型,但并不是绝对的。

包括人工可读 GEDI 文件头的例子:

Content-Type: multipart/gedi-record;
 boundary="unique-boundary1"

--unique-boundary1
Content-type: multipart/mixed;
 boundary="unique-boundary2"

--boundary2
Content-Type: application/gedi-GEDI Header
 Content-Transfer-Encoding: quoted-printable

IFID 0004GEDIIFVS00033.0CILN00042048DFID0008TIFF-6.0SSAD0005?;=()CNSN0006N=
PICARCNM0008RLG00001SPLN0005N = RLGSVDT001419910802140600TTLE0012PC/
ComputingAART0013Paul SomersonTART0031The DOS you've been waiting forZPAD1860
--unique-boundary2

Content-Type : text/plain	
interchange-format-id	GEDI
interchange-format-version	3.0
document-format-id	TIFF-6.0
Consumer-name	PICA
record-name	RLG00001
Supplier-name	RLG00001
service-date-time	1991/08/02 14:06:00
title	PC/Computing
author-of-article	Paul Somerson
title-of-article	The DOS you've been waiting for

--unique-boundary2--
--unique-boundary1
Content-Type:image/tiff
Content-Transfer-Encoding: base64

<multi-page TIFF image>
--unique-boundary1--

10.3 支持 MIME 的协议栈

SMTP 协议用于发送,SMTP 或 POP3 协议用于接收,均以 TCP/IP 为基础。

11 一致性

11.1 发送/接收角色

一个系统可以声明与以下一方一致:

- a) 发送系统(提供方);
- b) 接收系统(接收方);
- c) 发送系统(提供方)和接收系统(接收方)双方。

11.2 GEDI 文件头数据元一致性

声明与发送系统(提供方)一致的必须能够发送 7.2 中定义的所有必备数据元,包括 7.2.2 中定义的所有类型 1 数据元和 7.2.3 中定义的类型 2 必备数据元。

声明与接收系统(接收方)一致的必须能够接收第 8 章说明的所有 GEDI 文件头数据元。广义上说,假定系统忽略任何未知数据元,接收系统(接收方)必须可以接收未知的 GEDI 文件头数据元。

11.2.1 作为提供方,如果声明与 GEDI 文档交换格式 GEDI-3.0 一致,系统必须能够传输本标准定义的所有必备 GEDI 文件头数据元。标号不可重复,除了 IFID 放在第一位、ZPAD 放在最后一位,没有规定其他标号的顺序。

11.2.2 作为接收方,如果声明与 GEDI 文档交换格式 GEDI-3.0 一致,系统必须能够

- a) 接收和处理本标准定义的所有可选要素,
- b) 接收和处理附录 B 中规定的 TIFF-6.0 图像。

为了扩展性的目的,接收方系统必须能够忽略所接收到的未知数据元,只要这些未知数据元符合 7.3 中规定的〈4 字符标号〉〈4 字符长度〉〈值〉结构。

11.2.3 作为中继方,如果声明与 GEDI 文档交换格式 GEDI-3.0 一致,系统必须能够

- a) 透明的并且没有任何变化的接收和传输 GEDI 记录。
- b) 创建和传输如附录 B 所规定的 TIFF-6.0 图像。

11.3 电子文档拷贝的一致性

声明与发送系统(提供方)一致的必须能够发送如附录 B.1 所定义的 TIFF 图像格式电子文档拷贝。

声明与接收系统(接收方)一致的必须能够接收如附录 B.1 所定义的 TIFF 图像格式电子文档拷贝。

11.4 协议一致性

一个系统可以声明与以下任一系统一致:

- a) FTP 系统。
- b) MIME 系统。
- c) FTP 系统和 MIME 系统。

11.4.1 FTP 一致性

声明与 FTP 系统一致的系统必须遵守第 9 章的规则/说明。

11.4.2 MIME 一致性

声明与 MIME 系统一致的系统必须遵守第 10 章的规则/说明。

附 录 A
(资料性附录)
馆际互借 APDU 到 GEDI 的映射

表 A.1

馆际互借数据元	组成/要素	GEDI 标号	注 释
协议版本号			
事务标识符	最初请求方标识符 个人或机构代码 个人或机构名称 事务组限定语 事务限定语 子事务限定语	ILTI(S=) ILTI(N=) ILTI(G=) ILTI(Q=) ILTI(B=)	
服务日期时间	本次服务日期时间 日期 时间 最初服务日期时间 日期 时间	— — — —	
请求方标识符	个人或机构代码 个人或机构名称	RQID RQNM(N=)	
应答方标识符	个人或机构代码 个人或机构名称	RSID RSNM(N=)	
事务类型			
发送地址	邮政地址 个人或机构名称 扩展的邮递地址 街道及门牌号 邮政信箱 城市 地区 国家 邮政编码 电子地址 电讯服务标识符 电讯服务地址	NPOI XPDA STNM POBX CITY REGN CNTR POCD SYID SYAD	
发送服务	物理发送 电子发送 电子发送服务 电子发送模式 电子发送参数	— CNSN	

表 A.1 (续)

馆际互借数据元	组成/要素	GEDI 标号	注 释
发送服务	文档类型 文档类型标识符 文档类型参量 电子发送细节 电子发送地址 电讯服务标识符 电讯服务地址 电子发送标识符 个人或机构代码 个人或机构名称 名称或代码 发送时间	 CNSN — —	 存在一个电子发送地址或电子发送标识符
帐单地址	邮政地址 名称 扩展的邮递地址 街道及门牌号 邮政信箱 城市 地区 国家 邮政编码 电子地址 电讯服务标识符 电讯服务地址	 — — — — — — — — — — —	
馆际互借服务类型		—	
应答方特定服务		—	
请求方可选消息	能发送 RECEIVED 请求方 CHECK-IN 能发送 RETURNED 请求方 SHIPPED 需要 希望 两者都不	— — — — — — —	没有合适的映射可转到 GEDI 文件头,但有可能应用于某些 GEDI
检索类型	服务水平 某日期前需要 终止标志 终止日期	— — — —	
供应媒体信息类型	供应媒体类型 媒体特征	— —	

表 A.1 (续)

馆际互借数据元	组成/要素	GEDI 标号	注 释
第三方信息类型	系统标识符	—	
	个人或机构代码	—	
	个人或机构名称	—	
	帐号	—	
	系统地址	—	
	电讯服务标识符	—	
	电讯服务地址	—	
	已试名录	—	
	个人或机构代码	—	
	个人或机构名称	—	
重试标志		—	
转发标志		—	
请求方注释		RSNT	
转发注释		—	
馆际互借请求扩展		—	可能需要在执行协议或量表中说明

附 录 B
(资料性附录)
电子文档拷贝格式注册

B.1 TIFF 注册**B.1.1 TIFF 基本信息**

文档扫描图像将以 B 类图像文件格式(TIFF)进行传送,满足 TIFF-6.0 中对二值图像和灰度图像的所有要求,整个图像包含在一个单独的多页 TIFF 文件中。

下面的 DFID 值可以用来确定 TIFF

TIFF-5.0

TIFF-6.0

下面的 MIME 媒体格式将用来确定 TIFF

image/TIFF; class=B

image/TIFF; class=G

B.1.2 TIFF 图形文件的 GEDI 文件头

表 B.1

字 节	描 述	写	读
0-1	字节顺序	11	11 或 MM
2-3	TIFF 标记号	42	42
4-7	图像文件目录指针偏移量	—	任意

在表 B.1 到 B.5 中使用以下标记

粗体: 所有 TIFF 类要求的字段

粗斜体: TIFF 要求的附加字段

M: 必备

O: 可选

v: 可变内容

—: 不适用

表 B.2

TIFF 基准字段				GEDI 要求			
	标号	类型	值	类型	值	写	读
图像文件建立的相关信息	13B. H	ASCII	—	ASCII	—	O	接收
每个采样点的位数	102. H	短整型	1	短整型	1	M	M
图像数据点的长度	109. H	短整型	—	短整型	—	?	?
图像数据点的宽度	108. H	短整型	—	短整型	—	?	?
色彩映像	140. H	短整型	—	短整型	—	?	?

表 B.2 (续)

TIFF 基准字段				GEDI 要求			
	标号	类型	值	类型	值	写	读
压缩	103. H	短整型	1,2 或 32773 缺省=1	短整型	1,2,4 或 32773 ^a	M	M
版权声明	8298. H	ASCII	—	ASCII	—	?	?
日期时间	132. H	ASCII	—	ASCII	—	O	接收
额外的样本点	152. H	短整型	0,1 或 2 缺省=无字段	短整型	?	?	?
排列方向	10A. H	短整型	1 或 2 缺省=1	短整型	?	?	?
空字节数目	121. H	长整型	—	长整型	?	?	?
自由偏移量	120. H	长整型	—	长整型	?	?	?
灰度响应比值	123. H	短整型	—	短整型	?	?	?
灰度响应	122. H	短整型	1…5 缺省=2	短整型	?	?	?
机器说明	13C. H	ASCII	=	ASCII	—	O	接收
图像描述	10E. H	ASCII	—	ASCII	—	O	接收
图像长度	101. H	短整型 长整型 ^b	v	短整型 长整型 ^b	v	M	M
图像宽度	100. H	短整型 长整型 ^b	v	短整型 长整型 ^b	v	M	M
制造商	10F. H	ASCII	—	ASCII	—	O	接收
最大值	119. H	短整型	缺省=2 ^{..} (每采样点 位数) ^c -1	短整型	?	?	?
最小值	118. H	短整型	缺省=0	短整型	?	?	?
型号	110. H	ASCII	—	ASCII	?	O	接收
子文件类型	OFE. H	长整型	缺省=1	长整型	v ^c	M	M
原点位置	112. H	短整型	1…6 缺省=1	短整型	?	?	?
测光度	106. H	短整型	0…4	短整型	0…4	M	M
图像存储各式的平面配置	11C. H	短整型	1 或 2 缺省=1	短整型	1 或 2 缺省=1	M	M
分辨率单位	128. H	短整型	1…3 缺省=2	短整型	1…3 缺省=2	M	M
每数据带位数	116. H	短整型, 长整型 ^d	缺省=2 ^{..} 32 ^c -1	短整型 长整型 ^d	?	M	M
像素样本点数	115. H	短整型,	缺省=1	短整型	缺省=1	M	M
软件	131. H	ASCII	—	ASCII	—	O	接收
数据带字节数	117. H	短整型, 长整型 ^d	v	短整型, 长整型 ^d	v	M	M

表 B.2 (续)

TIFF 基准字段				GEDI 要求			
	标号	类型	值	类型	值	写	读
数据带偏移量	111. H	短整型, 长整型 ^d	v	短整型, 长整型 ^d	v	M	M
子文件类型(废止)	OFF. H	短整型	1…3 无缺省值	短整型	?	不使用	?
域值	107. H	短整型	1…3 缺省=1	短整型	?	?	?
X分辨率	11A. H	有理数	v	有理数	v	M	M
Y分辨率	11B. H	有理数	v	有理数	v	M	M
<p>^a 压缩字段数值含义</p> <p>1=无压缩</p> <p>2=CCITT 压缩方法(版本 3)——一维改进的霍夫曼行程长度编码压缩方法</p> <p>4=传真机相容使用的 CCITT 组 4</p> <p>32773=紧缩位压缩方法</p> <p>^b 优先使用长整型</p> <p>^c 新子文件类型字段的数值</p> <p>Bit 0</p> <p>Bit 1</p> <p>Bit 2</p> <p>^d 优先使用短整型</p> <p>^e 表示零</p>							

表 B.3

TIFF 二值图像 要求的字段	二值图像的要求			GEDI 要求			
	标号	类型	值	类型	值	写	读
压缩	103. H	短整型	1、2 或 32773 缺省=1	短整型	1、2、4 或 32773 ^a	M	M
图像长度	101. H	短整型 长整型 ^b	v	短整型, 长整型 ^b	v	M	M
图像宽度	100. H	短整型 长整型 ^b	v	短整型, 长整型 ^b	v	M	M
测光度	106. H	短整型	0…4	短整型	0…4	M	M
每数据带位数	116. H	短整型 长整型 ^c	缺省值=2 ^a 32 ^d —1	短整型, 长整型 ^c	v	M	M
数据带字节数	117. H	短整型 长整型 ^d	v	短整型, 长整型 ^d	v	M	M
数据带偏移量	111. H	短整型 长整型 ^c	v	短整型, 长整型 ^c	v	M	M
X分辨率	11A. H	有理数	v	有理数	v	M	M
Y分辨率	11B. H	有理数	v	有理数	v	M	M
注: 接收方应用程序必须能够接收这些字段,但允许舍弃接收的数值。							

表 B.3 (续)

TIFF 二值图像 要求的字段	二值图像的要求			GEDI 要求			
	标号	类型	值	类型	值	写	读
<p>^a 压缩字段数值含义</p> <p>1=无压缩</p> <p>2=CCITT 压缩方法(版本 3)——一维改进的霍夫曼行程长度编码压缩方法</p> <p>4=传真机相容使用的 CCITT 组 4</p> <p>32773=紧缩位压缩方法</p> <p>^b 优先使用长整型</p> <p>^c 优先使用短整型</p> <p>^d 表示幂</p>							

表 B.4

TIFF 传真字段	CCITT 二值编码的 B 级要求			GEDI 要求			
	标识符	类型	值	类型	值	写	读
压缩	103. H	短整型	3 或 4	短整型	1、2、4 或 32773 ^a	M	M
T4 选项(CCITT 组 3)	124. H	长整型	缺省=0	不写这些字段,准备接收它们;允许舍弃接收的数值			
T6 选项(CCITT 组 4)	125. H	长整型	缺省=0				
<div>^a 压缩字段数值含义</div> <div>1=无压缩</div> <div>2=CCITT 压缩方法(版本 3)——一维改进的霍夫曼行程长度编码压缩方法</div> <div>4=传真机相容使用的 CCITT 组 4</div> <div>32773=紧缩位压缩方法</div>							

表 B.5

TIFF 文献存储和 检索字段	存储和检索要求			GEDI 要求			
	标识符	类型	值	类型	值	写	读
文件名	10D. H	ASCII		ASCII		O	接收
页名	11D. H	ASCII		ASCII		O	接收
页码 ^a	129. H	短整型	v	短整型	v	O	接收
X 分辨率	11A. H	有理数	v	有理数	v	M	接收
Y 分辨率	11B. H	有理数	v	有理数	v	M	接收
注:接收方应用程序必须能够接收这些字段,但允许舍弃接收的值。							
<p>^a 该字段用来为多页文档标明页码(例如传真)。页码[0]表示页码;页码[1]表示文档的总页数。如果页码[1]是 0,则文件的总页数不可用,第一页页码标为 0。</p>							

B.1.3 TIFF 压缩规则

GEDI 支持以下几种压缩算法,这些算法在 TIFF 6.0 说明书中进行了详细说明。

表 B.6

压缩算法	值	写	读
无压缩	1	可选	必备
CCITT 压缩方法(版本 3)——一维改进的霍夫曼行程长度编码压缩方法	2	可选	可选
传真机相容使用的 CCITT 组 4	4	可选	必备
紧缩位压缩方法	32773	可选	必备

一维改进的霍夫曼运行长度编码 CCITT 组 3 与 B 类 TIFF 兼容。

B.2 PDF 注册

可以在 Adobe 公司 1999 年 3 月 11 日发布的 PDF 参考手册(1.3 版)中找到 PDF 说明。

Adobe 公司拥有该说明的版权。

GEDI 没有详细说明 PDF 文件的制作过程,也没有指定使用哪种 Adobe Acrobat 产品查看和打印 PDF 文档。

PDF 作为一个 MIME 主体部分由 IANA 注册。注册信息可以在以下站点获得:

<ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/application/pdf>

下面的 DFID 值用来识别 PDF:PDF

下面的 MIME 媒体类型用来识别 PDF:application/pdf

B.3 JFIF/JPEG 注册

除了 TIFF, GEDI 也识别 JFIF(JPEG 文件交换格式),它是一种低级别的像素传输格式。JFIF 通常是指涉及到的“JPEG 文件”,该格式通常被 WWW 浏览器所支持。它不会与 TIFF/JPEG 混淆,后者并没有包括在 GEDI 说明书中。

关于 JFIF 的详细信息可以在以下站点获得:

www.cis.ohio.-state.edu/hypertext/faq/usenet/jpet-faq/part1/faq-doc-14.html

下面的 DFID 值用来识别 JFIF/JPEG:JIFF

下面的 MIME 媒体类型用来识别 JFIF/JPEG:JPEG

参 考 文 献

- [1] ISO/IEC 8859-1:1998 信息技术 8 位单字节编码图形字符集 第 1 部分:拉丁字母第 1 号
- [2] ISO 9735-1:1998 管理、商业和运输业用电子数据交换(EDIFACT). 应用层语法规则(语法版本号:4). 第 1 部分:与各部分语法使用目录一起使用的所有部分通用语法规则
- [3] ISO 10160:1997 信息与文献 开放系统互联 馆际互借应用服务定义
- [4] ISO 23950:1998 信息和文献工作 信息检索(Z39.50). 应用服务定义和协议规范
- [5] RFC 822 ARPA 网际通讯格式标准,1982.8.13
- [6] RFC 1512 FDDI 管理信息基础,1993.9
- [7] RFC 1522(被 RFC 2045 替代)
- [8] RFC 1524 多媒体邮件格式信息的用户代理配置机制,N. Borenstein,1993.9.9
- [9] RFC 1939 互联网电子邮件协议—版本 3,J. Myers & M. Rose,May 14,1996
- [10] RFC 2045 多用途网际邮件扩充协议(MIME) 第 1 部分:网际通讯体格式,N. Freed & N. Borenstein,1996.11
- [11] RFC 2046 多用途网际邮件扩充协议(MIME) 第 2 部分:媒体类型,N. Freed & N. Borenstein,1996.11
- [12] RFC 2047 多用途网际邮件扩充协议(MIME) 第 3 部分:非 ASCII 文本的信息标题扩展,K. Moore,1996.11
- [13] RFC 2048 多用途网际邮件扩充协议(MIME) 第 4 部分:注册程序,N. Freed,J. Klensin & J. Postel,1996.11
- [14] RFC 2049 多用途网际邮件扩充协议(MIME) 第 5 部分:一致性准则及例子 Freed & N. Borenstein,1996.11
- [15] PDF 参考手册. 版本 1.3,Adobe 系统公司,1999.3.11
- [16] JPEG 文件交换格式. 1.02 版本,Eric Hamilton,1992.9.1
- [17] TIFF 6.0 最终版—Aldus Developer's Desk,Aldus 公司(1992.6.3 西雅图,华盛顿州)



GB/T 23731-2009

版权专有 侵权必究

*

书号:155066·1-38256

定价: 39.00 元