# 数字数据采集工具规范

公开意见征集草案1 版本4.0

## 简介

计算机取证工具的可靠性是执法机构的迫切需求，这些取证工具需要确保取证结果的准确、可复现和客观性。美国标准与技术研究院（NIST）设立的取证工具测评项目（CFTT）就是为了构建一种依靠实用规范、测评规程、测评标准、测试集合和测评硬件的建设以达到全方位测评计算机取证工具性能的方法。这套方法将为工具的制造者改善工具的研发，为工具的使用者获取和使用工具，为一些对取证感兴趣的组织了解工具的功能等事务提供重要的信息。这种测评方法是基于国际公认的一致性评测和质量评测方法的。对于该方法的进一步介绍可以参阅：<http://www.cftt.nist.gov/>。

取证工具测评项目（CFTT）是美国国家司法研究院（NIJ）的一项联合项目，参与该项目的还有美国司法部的发展组织，美国标准与技术研究院（NIST）的法律执行标准办公室（OLES）和信息技术实验室（ITL）。并获得了美国联邦调查局、国防部网络犯罪中心、国土安全部下属的移民服务局、海关执法局和美国特勤局的支持与协助。由于所有文件都公布于网上进行公开审查，整个计算机取证社区也参与了规范和测评方法的制定。

## 目标

本文件定义了用于计算机取证调查的数字媒体采集工具的具体要求，是原始磁盘映像规范和磁盘映像工具规范（版本3.1.6）的修订版。原始规范涵盖了在制定规范时（2001年10月）从计算机硬盘获取数字数据时广泛提供的工具和技术，并且已经能够有效的评测当时磁盘映像工具的关键特性并生成测评报告。但是，由于相关技术和磁盘映像工具的发展，需要对已有的评测规范进行相应的修订和补充。从“磁盘映像工具规范”到“数字数据采集工具规范”的标签变化，反映出的是存储媒介的普遍性和多样性。此次修订的主要目标是将规范的覆盖范围扩展到新的存储技术，并将覆盖范围扩展到新的采集工具功能。修订的次要目标是更新术语以增强需求的灵活性和措辞的简洁性，使得更容易结合新技术。此外，为了改进布局和易读性，并与新规范保持一致，测试断言和测试用例已被移至单独的文档中，并将稍后发布。

本文中提出的规范将用于派生断言以进行评测。断言是一种用于条件声明，并且该声明将在一次测试执行后进行验证。每个断言都在一个或多个测试用例中检查，这些用例指定了详细的初始条件，测试场景和预期的测试结果。

本文中提出的规范要求是由一组专门使用磁盘采集工具的专家，通过观察这类工具的实际执行结果而制定的。随着专家小组和其他组织的意见发展，新版本将发布于<http://www.cftt.nist.gov/>。

## 03. 范围

本规范的适用范围仅限于通过计算机文件系统从数字存储介质获取数据的软件和硬件设备。不包括直接从其他数字设备（如手机，寻呼机或PDA）对存储介质生成映像的工具。

工具的正确使用或不当使用不在本规范的涉及范围。

## 04. 背景

美国国家司法研究院（NIJ）的一份特别报道199408，“数字证据的司法审查：执法指南”提出了处理数字证据的指导方针，作为刑事调查过程的一部分。该报告指出，数字证据分四步处理：评估、获取、检查、记录和报告。这份指南指出了采集工具的具体功能。

数字媒体采集过程从识别数字源开始。数据源可以是物理设备，例如来自计算机的硬盘驱动器，来自相机的存储卡，闪存设备或任何可用于存储数字数据的可移动数字媒体。数字源也可以是物理设备上的逻辑驱动器。映像生成的理想目标是实现对数字源完整和准确的采集。

数字源被识别后，它被连接到计算机接口上进行采集。一些工具的执行环境会在启动引导过程和关闭过程中修改与之相连的存储设备。与这类采集工具相连的数字源通常使用写入阻止器来保护数字源免于修改。

将数字源连接到计算机接口后，采集工具从设备读取数据并将数据（称为目标对象）保存为可访问形式。目标对象通常是一个或多个映像文件，代表从数字源获取的所有数据。目标对象也可以是源的副本，或者是源的比特拷贝（未对齐的副本），或者可以是比特流副本，除了根据需要进行细微更改以对齐柱面边界的分区（一个柱面对齐的克隆）。克隆和映像的主要区别是：映像是通过工具获取的，而克隆是通过常规的计算机文件系统获得的。

采集过程的两个关键可衡量属性是完整性和准确性。如果获得所有数据，则采用完整性度量，如果数据正确获取，则采用准确度度量。

要访问数字源，物理设备需要通过物理接口连接到计算机，然后采集工具需要通过某种协议读取设备。例如，ATA接口可能连接了一个硬盘驱动器，然后通过BIOS中断0x13命令访问或直接通过ATA命令访问。物理接口和访问方法的组合是访问接口。某些访问接口的示例包括：传统BIOS，扩展BIOS，ATA，SATA，SCSI，ASPI，USB，IEEE 1394，RAID，以及通过网络远程访问。某些接口不止有一个接口版本，例如，ATA-3不允许使用48位磁盘地址，但ATA-6允许使用48位磁盘地址，在进行数据采集时需要注意。

数字映像的一个组成部分是确定数字源的真实大小。根据较新的ATA规范构建的硬盘驱动器可能允许创建无法访问或隐藏的区域，例如主机保护区域或设备配置叠加层。具有80GB空间的驱动器重新配置后显示的空间可能变小。尝试从隐藏区域读取会导致访问错误，除非驱动器被重新配置回原始大小。

## 5.0 定义

以下是本规范涉及的相关术语及定义。

其他硬盘驱动器相关术语的定义可以在ANSI INCITS 361-2002 “AT Attachment-6 with Packet Interface”中查阅。

|  |  |
| --- | --- |
| Acronym | Expanded Term |
| ANSI | American National Standards Institute |
| ASPI | Advanced SCSI Programming Interface |
| ATA | AT-Attachment |
| BIOS | Basic Input Output System |
| IEEE | Institute of Electrical and Electronics Engineers |
| INCITS | International Committee for Information Technology Standards |
| RAID | Redundant Array of Independent Disks |
| SATA | Serial ATA |
| SCSI | Small Computer System Interface |
| USB | Universal Serial Bus |

**访问接口（Access interface）**：一种物理接口（设备的物理连接）和数据采集工具访问数字源的方法（系统命令或协议）的组合。不论是默认配置的或是用户可自行设置的，对于采集工具这些接口都是可见的。

**精确采集（Accurate acquisition）：**对于数据源的可访问部分，目标对象的每个比特位都与数据源的相应比特位的数值一样，同时对于数据源的不可访问部分，目标对象会在对应的比特位进行良性填充（在解密或解压后会进行校验）。参阅**完全采集**（Complete acquisition）。

**采集(Acquisition)：**指通过访问接口从数据源读取数据并生成目标对象的过程。

**采集工具（Acquisition tool）：**指用于读取数据源生成映像文件或数据源克隆副本的程序或硬件设备。采集工具也称为映像工具。

**良性填充（Benign fill）：**指生成目标对象时，对应于数据源中不可访问部分的数据段填充的数据，或是在生成数据源克隆副本时对冗余的数据段填充的数据。这种填充的数据应能够被正确识别为不含任何意义的良性填充，如全部为0或设置标识能够指明该部分是填充数据而非数据源中的数据。

**比特流复制（Bit-stream duplicate）：**以比特为拷贝单位的数据对象拷贝，例如对电子文档、文件、分区、图片、物理磁盘或小型数字对象的拷贝。

**克隆目标（Clone destination）：**用于接收非对齐克隆或柱面对齐克隆的物理媒介。

**完全采集（Comlete acquistition）：**如果对于数据源的任一比特值都有目标对象中的一个比特值与之对应，同样的，如果对于目标对象中的任一比特值都有数据源的一个对应比特值与之对应，这种克隆称为完全克隆。需要注意的另外一种形式是，有时目标对象中可能存在非描述性的数据，且此类数据与数据源中的数据不存在对应关系，参见**精确采集**。

**柱面对齐采集（Cylinder-aligned clone）：**数据源按比特流复制到物理媒介中，允许一些为了磁盘柱面对齐而进行微小的修改。