ICS 35.030
CCS L80

中华人民共和国国家标准

GB/T XXXXX.1—XXXX/ISO/IEC 27050-1:2019

信息安全技术　电子发现　第1部分:概述和概念

Information security technology—Electronic discovery—Part 1: Overview and concepts

（ISO/IEC 27050-1:2019，Information technology—Electronic discovery—Part 1: Overview and concepts，IDT）

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（征求意见稿）

2022-06-30

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

国家市场监督管理总局
国家标准化管理委员会　　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是GB/T XXXXX《信息安全技术　电子信息发现》的第1部分。GB/T XXXXX已经发布了以下部分：

——第1部分：概述和概念。

本文件等同采用ISO/IEC 27050-1:2019《信息技术　电子信息发现　第1部分：概述和概念》。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息安全标准化技术委员会（SAC/TC260）提出并归口。

本文件起草单位：中电长城网际系统应用有限公司、北京百度网讯科技有限公司、北京奇虎科技有限公司、国信京宁信息安全科技有限公司、北京网络行业协会电子数据司法鉴定中心、信大捷安信息技术股份有限公司、国家计算机网络应急技术处理协调中心、公安部第三研究所、中国长江三峡集团有限公司。

本文件主要起草人：闵京华、王海棠、张屹、郭建领、王佳慧、王笑强、刘为华、王文磊、舒敏、王宁、杨卫军、唐进。

引　言

本文件概述了电子信息发现，描述了GB/T XXXXX系列其他部分拟采用的相关的术语、概念和过程。

电子信息发现通常作为助力服务于调查以及证据获取和处理活动（包含在ISO/IEC 27037中）。此外，数据的敏感性和关键性有时候需要像存储安全（包含在ISO/IEC 27040中）这样的保护，以防止数据安全性受损。

GB/T XXXXX拟由以下四个部分构成：

——第1部分：概述和概念。等同采用ISO/IEC 27050-1:2019《信息技术　电子信息发现　第1部分：概述和概念》。

——第2部分：电子信息发现的治理和管理指南。拟等同采用ISO/IEC 27050-2:2018《信息技术　电子信息发现　第2部分：电子信息发现治理和管理指南》。

——第3部分：电子信息发现实践指南。拟等同采用ISO/IEC 27050-3:2020《信息技术　电子信息发现　第3部分：电子信息发现实践指南》。

——第4部分：技术准备。拟等同采用ISO/IEC 27050-4:2021《信息技术　电子信息发现　第4部分：技术准备就绪》。

有关ISO/IEC 27050系列的更多信息，详见第5章。

信息安全技术　电子发现　第1部分：概述和概念[1]

1 范围

电子信息发现是指由参与调查、诉讼或类似程序的一方或多方发现相关电子留存信息（ESI）或数据的过程。本文件概述了电子信息发现，定义了相关术语，描述了相关概念，包括但不限于ESI的识别、保全、收集、处理、评审、分析和产出。本文件还确认了其他相关标准（如ISO/IEC 27037）及其与电子信息发现活动的关系和相互作用。

本文件适用于参与部分或全部电子信息发现活动的非技术人员和技术人员。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

ISO/IEC 27000　信息安全　安全技术　概述和词汇（Information technology — Security techniques — Information security management systems — Overview and vocabulary）

注：GB/T 29246—202X　信息安全技术　信息安全管理体系　概述和词汇（ISO/IEC 27000:2018，IDT）

3 术语和定义

ISO/IEC 27000界定的以及下列术语和定义适用于本文件。

3.1

保管链　chain of custody

从一个时间点到另一个时间点可证明的关于材料的拥有、移动、处理和位置。

3.2

保管方　custodian

保管、控制或拥有电子留存信息（3.8）的个人或实体。

3.3

数据安全性受损　data breach

导致受保护的数据在传输、存储或其他处理过程中发生意外或非法的损毁、丢失、改动或者未授权的披露或访问的安全性降低。

［来源：ISO/IEC 27040:2015，3.7］

3.4

发现　discovery

各方获取另一方持有的或不被任何一方持有的与某一事项相关的信息的过程。

注1：相比对抗性纠纷中的当事方，发现的适用范围更广。

[1] 根据意见处理结果，在本文件正文中将“electronic discovery”改译为“电子信息发现”，必要时将按规定程序变更标准名称。

注 2：发现也是指对方的硬拷贝文档、电子留存信息（3.8）和有形物品的披露。

注 3：在某些司法管辖区中，术语“披露”与“发现”互换使用。

3.5

处置 disposition

按照文件授权规定，实施文件保管、销毁或移交决策的一系列过程。

［来源：GB/T 26162-2021，3.8］

3.6

电子档案 electronic archive

电子留存信息（3.8）的长期存储库。

注 1：电子档案可为在线可访问的，也可为离线不易访问的。

注 2：备份系统（如磁带或虚拟磁带等）不用作电子档案，而是用作数据保护系统（即为了灾难恢复和业务持续性的备份机制）。

3.7

电子信息发现 electronic discovery

电子留存信息（3.8）的发现（3.4），包括对其识别、保全、收集、处理、评审、分析和产出。

注：尽管电子信息发现通常被视为法律过程，但其使用不限于法律领域。

3.8

电子留存信息 electronically stored information

ESI

以各种电子化形态存储过的任何类型和任何来源的数据或信息。

注 1：ESI 包括传统的电子邮件、备忘录、信函、电子表格、数据库、办公文档、演示文稿和计算机上常见的其他电子格式。ESI 还包括系统、应用和文件相关的元数据，如时间戳、修订历史、文件类型等。

注 2：电子媒体可采取但不限于存储设备和存储元件的形式。

［来源：ISO/IEC 27040:2015，3.16］

3.9

ESI 分析 ESI analysis

电子信息发现（3.7）过程的环节，侧重于评价电子留存信息（3.8）的内容和语境，以识别事实、关系、关键模式和其他特征，从而提高对 ESI（3.8）库的理解。

注：内容和语境可包含关键模式、主题、人员和讨论。

3.10

ESI 收集 ESI collection

电子信息发现（3.7）过程的环节，侧重于聚集电子留存信息（3.8）和相关材料。

3.11

ESI 识别 ESI identification

电子信息发现（3.7）过程的环节，侧重于定位潜在来源和选择潜在相关电子留存信息（3.8）的准则。

3.12

ESI 保全 ESI preservation

电子信息发现（3.7）过程的环节，侧重于保持电子留存信息（3.8）处于其原始或现有状态。

注：在某些事项或司法管辖区中，可能有防止电子留存信息（3.8）的证据破坏（3.21）的要求。

3.13

ESI 处理 ESI processing

电子信息发现（3.7）过程的环节，侧重于提取电子留存信息（3.8），并在必要时将其转换为更适合 ESI 评审（3.15）和 ESI 分析（3.9）的形式。

3.14

ESI 产出 ESI production

电子信息发现（3.7）过程的环节，侧重于交付或生成可用的电子留存信息（3.8）。

注 1：ESI 产出还可包括以适当的形式获取电子留存信息（3.8），并使用适当的交付机制。

注 2：ESI 产出可面向任何个人和组织。

3.15

ESI 评审 ESI review

电子信息发现（3.7）过程的环节，侧重于基于特定准则筛选电子留存信息（3.8）。

注：在某些事项或司法管辖区中，被视为特权的电子留存信息（3.8）可被排除在产出之外。

3.16

调查 investigation

对与某一事项有关的事实或材料进行调查、研究和检验的系统性或正式过程。

注 1：材料的形式可能是硬拷贝文档或电子留存信息（3.8）。

3.17

法律暂停 legal hold

由于当前或预期的诉讼、审计、政府调查或其他此类事项，暂停对记录和电子留存信息（3.8）正常处置（3.5）或处理的过程。

注 1：执行法律暂停的已发出通信也可称为“暂停”“保全令”“保全通知”“挂起令”、“冻结通知”、“暂停令”或“暂停通知”。

3.18

元数据 metadata

定义和描述其他数据的数据。

［来源：GB/T 18391.1—2009，3.2.16］

3.19

出处 provenance

记录电子留存信息（3.8）起源或来源的，以及自其起源以来发生的任何变化和其监管人的信息。

3.20

净化 sanitize

使得对存储媒体上的目标数据达到给定程度的不可访问。

注：净化存储媒体可采取清空、清除和销毁。

［来源：ISO/IEC 27040:2015，3.38］

3.21

证据破坏 spoliation

对需要保持完整的电子留存信息（3.8）进行更改或销毁的行为。

注：证据破坏的可能形式包括销毁、损坏或更改 ESI 或相关元数据，以及使 ESI 不可用（例如，无法访问解密密钥的加密、媒体丢失、在第三方控制下等）。

3.22

存储（名词） storage

支持数据输入和检索的设备、功能或服务。

［来源：ISO/IEC 27040:2015，3.43］

3.23

存储（动词） store

在易失性存储器或非易失性存储器上记录数据。

注：非易失性存储器是指即使断电仍能保留其内容的存储器，而易失性存储器是指一旦断电便无法保留其内容的存储器。

［来源：ISO/IEC 27040:2015，3.50，有修改：增加了注 1］

4 缩略语

下列缩略语适用于本文件。

CD 光盘（compact disc）
DVD 数字多功能光盘（digital versatile disc）
EDMS 电子文档管理系统（electronic document management system）
ERMS 电子档案管理系统（electronic records management system）
ICT 信息通信技术（information and communications technology）
NAS 网络连接存储（network attached storage）
OCR 光学字符识别（optical character recognition）
PII 个人可识别信息（personally identifiable information）
RAM 随机存取存储器（random access memory）

5 ISO/IEC 27050 系列整体结构和概述

ISO/IEC 27050 系列旨在满足各利益相关者在电子信息发现方面的需求。ISO/IEC 27050 系列的初始结构如下：

——ISO/IEC 27050-1（本文件等同采用）介绍了 ESI 和电子信息发现的一般术语和概念，并描述了电子信息发现过程环节。它旨在为广泛受众服务，并成为电子信息发现的基本信息来源。它不包括任何指导或要求。

——ISO/IEC 27050-2 侧重于与组织的治理层或高级管理层相关的电子信息发现的治理和管理方面。所提供的指导能帮助组织将其电子信息发现过程与 ISO/IEC 38500 中描述的六项良好治理原则保持一致。

——ISO/IEC 27050-3 为参与部分或全部电子信息发现活动的人员提供了要求和指导，包括补充材料，以帮助从业人员了解每个电子信息发现过程环节的目标和相关注意事项，这能帮助这些人员确定每个过程环节的相关性，并有助于避免那些可能增加风险和费用的失效。

注：如有必要，可在 ISO/IEC 27050 系列中添加其他部分。

6 电子信息发现概述

6.1 背景

无论是在组织内部还是在某些司法管辖区的法律系统中，电子信息发现都越来越重要。随着越来越多的电子记录和信息（或ESI）的创建、修改、操纵、使用和最终销毁，而不再采用物理形式（如打印文件），这种趋势将会持续下去。ESI 作为首选的信息呈现形式，它的出现带来了与 ESI 的定位、大量数据的处理、ESI 的保全和保留、真实性、数据的完整性，数据的保密性，数据或媒体的净化等相关的新挑战。电子信息发现的需要和响应因事项不同而不同，如果未能根据特定事项的语境来适当处理电子信息发现过程，就可能导致返工、不必要的成本、可能的制裁和法律责任。

ISO/IEC 27050（所有部分）通过以下方式应对这些挑战：

——促进电子信息发现的通用方法、理解和专业用语；

——鼓励负责在整个过程管理 ESI 的人员进行实际且具有成本效益的电子信息发现；

——为参与电子信息发现的人员识别所需胜任力；

——促进考虑积极主动使用技术手段，在提高整个电子信息发现过程效率的同时，降低成本和风险；

——建议采取一定的方法，避免无意中泄露那些潜在具有特权的，保密的或敏感的 ESI。

首要目标是帮助组织规划并满足其电子信息发现目标和义务（如果有），与每个特定事项的需要相称。

6.2 基本概念

预先考虑以下电子信息发现问题是有用的。这些问题的重要性和其解决的必要性因事项不同而不同，需要根据事项的需要进行调整：

——电子信息发现的范围；

——电子信息发现的治理和管理；

——电子信息发现项目各个方面的责任确定；

——持有潜在相关 ESI 的系统识别；

——潜在相关的 ESI 识别；

——在整个电子信息发现过程中编制适当的文档；

——预算的成本及其建议的分配；

——ESI 的保全，包括法律暂停过程；

——ESI 的存储方法、硬件和软件的信息披露；

——ESI 的收集/获取；

——ESI 的处理；

——ESI 的评审和分析；

——ESI 的产出，包括产出形式。

从事电子信息发现的人员会受到特定语境下一些因素的影响。成本可能是一个显著的影响因素。主要成本动因包括：

——收集：查找和检索潜在相关的 ESI；

——体量：要收集、处理或评审的 ESI 的原始数量；

——来源数量：参与 ESI 收集的保管方、企业系统以及其掌控下的外部系统和应用程序，可能会成倍增加投入的时间和工作量；

——人员胜任力：需要合格人员来能够执行数据检索、处理、搜索，以及相关性评审、特权和分类（例如相关性、特权、商业秘密、保密性或特殊处理）评审所需功能；这些胜任力可能包括信息技术，计算机技术，统计学，搜索科学和法律知识。

——案例复杂性：简单的案例可能需要有限的范围和评审程序，但复杂的案例可能涉及详尽的文件评审策略和流程。

查找和检索 ESI 的所需时间、ESI 的体量，接受 ESI 调查的来源数量，以及最终在法律程序或 ESI 调查中接受该 ESI 为可靠，都与组织为 ESI 在其整个生存周期中的管理而采取的实践和策略密切相关。在参与电子信息发现之前，就将电子信息发现准备就绪纳入其综合信息治理结构中的组织可能会更有效率、更具成本效益地满足电子信息发现的要求。ISO/IEC 27050-2 为此提供了具体的指导。

6.3 电子信息发现的目标

电子信息发现的目标因事项不同而不同，经对每一事项调整后，可能包括以下目标：

——遵守适用的法律、法规、规则和期望，对数据访问、使用、处理或传输施加保密性、数据隐私和其它限制；

——识别 ESI 的潜在相关来源；

——妥善保全和保留潜在相关的 ESI；

——将相关的 ESI 处理成便于有效搜索或评审的格式；

——将遗漏响应性 ESI 的可能性降至最低；

——将误判非响应性 ESI 为响应性 ESI 的可能性降至最低；

——将遗漏拒绝给或特殊处理的响应性 ESI 的可能性降至最低；

——将误判非拒绝给或特殊处理的响应性ESI为拒绝给或特殊处理的响应性ESI的可能性降至最低；

——以请求方可用的形式生成响应性 ESI；

——考虑在事项语境下的响应和成本的相称性；

——在整个项目中利用技术降低风险和成本。

6.4 电子信息发现的基础

6.4.1 概述

电子信息发现通常涉及到利益冲突的各方，在最坏的情况下，可能是对立各方。电子信息发现可能是解决冲突或问题的关键，但只有在具备可信度量的基础上进行。

电子信息发现的基础包括充分解决胜任力、公正、合作、完备性、相称性等方面的问题，这需要协调电子信息发现的要求与其他过程、价值或原则的要求。

6.4.2 胜任力

鉴于电子信息发现的复杂性，参与电子信息发现过程的个人具备相关技术或法律方面的胜任力很重要。他们可能需要能够证明他们受过适当的培训，并具有足够的技术或法律知识来适当处理 ESI，并代表一方执行电子信息发现过程。

6.4.3 公正

遵守适用的专业标准和道德行为对进行电子信息发现各方的期望。在某些司法管辖区，这意味着各方都有义务更正和补充记录（例如，额外的披露或修改先前的回复）。此外，有关各方都需要避免在执行电子信息发现过程中故意拖延。

6.4.4 合作

在某些司法管辖区的法院中，可能期望在与 ESI 保全、收集、查找、评审和产出有关的问题上进行合作，并且在此类法院中，合作通常不会损害客户的代表权。此外，在诉讼中，一方面合理限制 ESI 发现请求，另一方面合理响应 ESI 发现请求的合作，能降低成本和延迟。在电子信息发现的最初阶段进

行信息的合作交流可能是有用的。

6.4.5 完备性

产出方的目标是检索和生产一套（非特权）ESI，该 ESI 在事项的具体情况下代表一个完备和准确的产出。

6.4.6 相称性

随着 ESI 的爆炸式增长，如何处理好电子信息发现过程相关的成本和负担受到越来越多的关注。解决这一问题的一种方法是采取措施帮助确保电子信息发现的益处与相应的负担相称。

电子信息发现的负担可能是多种多样的，包括但不限于业务运营中断、财务成本或个人隐私侵犯。

6.5 治理和电子信息发现

6.5.1 概述

ISO/IEC 38500 为 ICT 的良好治理确立了与责任、战略、获取、绩效、合规性和人员行为相关的六项原则。每项原则都表示为指导决策的首选行为（即，每项原则都指出预期会发生什么，但没有规定如何、何时或由谁实施这些原则，因为这些方面取决于实施这些原则的组织的性质）。鼓励治理层要求应用这些原则，从而协助他们管理风险并促进利用 ICT 的机会。

根据 ISO/IEC 38500，ICT 的良好治理也有助于治理层确保遵守与 ICT 可接受使用相关的义务（法律、法规、合同方面的）。

ISO/IEC 27050-2 专注于与电子信息发现相关的一般性治理主题，但 6.5 强调了一些更重要的要素，以帮助提醒注意这些问题。

6.5.2 风险和环境因素

电子信息发现有可能使组织或其治理层暴露于可能导致有害影响的失败原因。治理能帮助避免如下形式的负面后果：

——违反隐私、健康和安全、记录保存的法律法规；

——不遵守有关安全、社会责任的标准；

——与知识产权有关的事项，包括许可协议；

要避免与这些失败相关的负面后果，就需要意识到并采取涵盖电子信息发现过程的缓解措施来解决诸如 ICT 系统不足和 ICT 使用不当等问题。

6.5.3 合规和评审

许多组织都面临来自法定、监管、法律或其他要求的合规性问题。这些要求可能是组织开展电子信息发现活动的原因，但更可能的是，它们会影响电子信息发现过程的执行方式。例如，谁可查看 ESI、ESI 的传输和存储方式以及特定的保留或销毁问题可能会受到限制。确保电子信息发现过程考虑到相关合规性要求非常重要。

6.5.4 隐私和数据保护

除了 6.5.3 中提到的监管限制和合规性问题外，重要的是要了解保管方数据使用的一些隐私限制（另见 ISO/IEC 29100）。特别是，保管方数据源中的个人可识别信息（PII）可能会受到一些限制，需要将其视为 ESI 管理的一部分。

在某些司法管辖区，当电子信息发现涉及 PII 时，可能会对如何使用它有严格的限制（例如，不能跨境传输）。即使没有这种限制，通常额外的数据保护措施也是必要的，以保护数据的保密性和防止数

据保护受损。

6.6 电子信息发现的 ICT 准备就绪

6.6.1 概述

在整个电子信息发现过程中，参与某一事项的各方都在收集、处理和操纵 ESI。通常，此 ESI 是从专门为保护它而设计的计算或存储环境中提取的。从这些环境中删除或复制的 ESI 可能需要类似的保护。

6.6.2 ESI 的长期保留

电子信息发现通常在诉讼、审计、政府调查或其他类似事项的早期使用。在该事项进行期间，各方需要保留相关 ESI，以使其继续可用并保持其完整性。适当的灾难恢复和业务连续性措施以及通常数据保护机制（如备份）可能是 ESI 保留计划的重要内容。

当决定长期保留 ESI 时，考虑所涉及的时间框架很重要。在电子档案中，将 ESI 保留数周或数月的方法与将 ESI 保留数十年的方法（例如，经过多次上诉的复杂民事诉讼）之间存在显著差异。

另外一个考虑因素是，数据保护和隐私要求是否会影响到个人数据的保留时间，以及是否需要暂停正常数据保留期。这在不同的司法管辖区之间可能存在很大差异。

6.6.3 ESI 保密性的保持

ESI 通常含有专有的、保密的和敏感的信息，需要以保护信息保密性的方式进行处理和存储。如果发生数据保护受损，那么未能充分控制敏感 ESI 可能会导致严重后果。

根据 ESI 的敏感性，可能需要采取安全措施，诸如动态数据和静态数据加密以及相应的密钥管理。

6.6.4 ESI 的销毁

当不再需要 ESI 时，以避免数据违背的方式消除 ESI 非常重要。这通常意味着用于保留 ESI 的逻辑存储器或存储媒体需进行适当的净化（例如，使用覆盖技术或加密擦除来清除）。

6.7 电子信息发现项目的规划和预算

电子信息发现项目背后的各种驱动因素使得提前几个月规划此类项目变得困难。因此，它们通常是单独管理，这会显著增加成本。无论请求的紧迫性如何，与任何项目一样，在开始时投入于规划的时间通常会在项目后期节省大量的时间和成本。尤其当一个典型的电子信息发现项目中的许多步骤在后期重复的成本高得不成比例。例如，在评审之前，未就产出结构和格式达成一致，且 ESI 或硬拷贝文档族的标记不一致，则可能导致评审不得不部分重复。

早期的一个重要步骤是建立一个电子信息发现项目团队，该团队至少包括一名来自企业/组织的项目发起人和经理，一名来自法律或调查团队的项目经理，以及一名来自 ICT 角度的项目经理。企业/组织、法律/调查人员和 ICT 团队之间的三角沟通对于项目的成功至关重要。

此外，早期步骤是制定电子信息发现项目计划，并尽可能详细。与任何项目计划一样，电子信息发现计划需要包含项目里程碑（例如，识别、保全/收集、处理、评审、分析、产出和可能的展示）、每个阶段所需的各个步骤以及每个步骤的各个任务。

考虑到典型的电子信息发现项目所涉及的成本，从一开始就准备和监控详细的预算是一个重要的考虑因素。该预算需要考虑到电子信息发现团队的不同学科，以及他们可以由内部和外部法律顾问以及内部和外部 ICT 或电子信息发现顾问组成的事实。重要的是要为计划中的每个电子信息发现过程环节以及在某些情况下每个过程环节的每个步骤提供预算，以便有可能确定建议的方法是否与手头的事项相称。

注：虽然 6.5 中的描述更倾向于涉及大型事项的大型企业，可能不完全适用于较小的组织或较小的事项，但提出的要点仍值得考虑。

7 电子留存信息（ESI）

7.1 背景

目前，ESI 是业务和个人环境中不可或缺的一部分。因此，它成为现代纠纷或事项中相关材料的日益重要的来源。

ESI 在事项的初始阶段就值得考虑。它可能非常脆弱，有些很容易丢失或被修改，即使打开文档也是如此。在意识到某一事项的早期，就考虑识别、保全以及可能的收集过程环节，使得能够做出重要决策，并在长期内节省大量时间和成本。

管理 ESI 对业务和个人的影响越来越大；ESI 的体量，规模，复杂性和范围通常是巨大的，ESI 本身可能包含需要考虑的机密、特权或私人信息。ESI 管理通常不被优先考虑，直到 ESI 定位的真正价值和成本明显成为某一事项的一部分时。组织或个人或经常：

——将其 ESI 保留工作的重点放在纯粹出于业务运营目的的保留上，而不是考虑更广泛的语境；

——对其在有关电子记录方面的合规义务给予最低限度的考虑；

——对良好业务记录的证据价值了解有限；

——对不良的信息管理实践带来的成本和风险缺乏足够的了解。

当需要识别或检索 ESI 以响应发现或监管请求时，不良的 ESI 管理可能会增加挑战，因为：

——内容受制于数据隐私或类似的访问限制，以及所有权和控制有关的限制；

——体量比预期的更大，因为它的存储超过了要求的生存期；

——组织内部通常很少知道在哪里能找到潜在的相关 ESI；

——即使对于 ICT 专业人员来说，其体量和复杂性也是压倒性的；

——员工更替或组织变革（如兼并、收购和资产剥离）导致 ESI 阻滞以及组织知识和语境的丢失；

——ICT 环境和系统可能缺乏文件化记录；

——ESI 可能位于外部的应用程序或 ICT 基础设施（如社交媒体、云计算等）中。

根据具体情况，这些因素可能会与导致在查找和处理与手头事项相关的数据的成本增加。这可能会导致延迟，增加发现过程的成本，并导致忽略潜在的相关 ESI。

7.2 ESI 的常见类型

7.2.1 概述

在电子信息发现的早期阶段，将 ESI 来源分类为易访问（或活跃的）或不易反问（或不活跃的、残留的、遗留的），并给出每个分类的理由，是一项重要活动。结合预算编制，这种分类可能有助于确定保全和收集这些来源的比例。

7.2.2 活跃数据

这类 ESI 处于“活跃”的使用状态，驻留在员工的计算机硬盘或其他存储设备上，以及组织的服务器、存储器和数据库中。活跃数据通常能在文件管理器或创建它的应用程序中访问到。无需数据恢复或重建，用户就能立即访问它。随着云计算和基于互联网的计算服务的日益普及，它也能驻留在外部服务提供商的存储设备上。大多数案件和调查主要是需要活跃文件的保全和产出。

至少与其他类型的 ESI 相比，活跃文件能相对容易地访问和收集。它们也容易被删除或修改，因此需要尽早考虑保全。

7.2.3 非活跃数据

这类 ESI 与已关闭、已完成或已定论的活动有关，包括组织以长期存储和记录保存为目的而维护的 ESI，但计算机系统用户无法立即访问。它可能包含许多与活跃数据相关的上述相同数据源。

存档数据通常以压缩格式存储，能在系统存储器或离线设备（包括存储磁带或磁盘和光盘）上进行维护。某些系统允许用户直接检索存档数据，而另些系统则需要 ICT 专业人员的协助。在保全和收集方面的难点包括识别相关的非活跃和存档数据、定位存储位置和方式以及从压缩格式恢复数据。

另一种形式的非活跃数据是存储在数据保护系统（如备份）中的 ESI。这些非活跃数据可能是问题的来源，因为它们往往是短期的（例如，定期轮换备份媒体），可能没有任何机制来确定媒体上的内容，存储的数据可能只包含片段（例如，只有自上次备份的变更）。使事情复杂化的是，ICT 人员可能在正常操作（如轮换、文档等）之外进行额外备份，而识别这些潜在的 ESI 来源是极其困难的。这类数据与存档数据面临着同样的挑战，还有额外的短保留期，这需要尽快采取行动，以便在需要保全时暂停这类 ESI 的自动销毁。

7.3.3 提供了有关备份和存档的附加信息。

7.2.4 残留数据

这类 ESI 是隐藏的，不能在应用程序（如系统文件）中查看，或者已被清除、碎片化或损坏。收集此类 ESI 通常需要对整个物理存储媒体（如硬盘、CD、DVD、磁带）进行精确的逐位复制，包括媒体上的所有活跃和残留数据以及未分配或闲置空间。

镜像后提取残留数据可能需要数字证据专家（见 ISO/IEC 27037）来运用专用工具，这可能是一个既耗时又昂贵的过程。但在某些情况下，公司可能选择对特别重要的关键保管方的硬盘进行镜像，以确保其所有数据都得到保全，包括保管方可能无意或有意删除或部分覆盖的文件。

7.2.5 遗留数据

这类 ESI 是由过时或废弃的软件或硬件（遗产系统）创建的。遗产系统可能是公司仍在使用但其硬件或软件供应商已不再支持的系统。或者，它可能是使用公司已退役的系统，但保留以备将来需要其信息。

在不进行恢复或重建的情况下，很难确定遗留数据的相关性，而且这样做的成本可能很高。如果需要保留这些遗留数据，在没有其他方式查看或使用这些数据的情况下，公司可能需要保留其遗产硬件和软件。

7.3 ESI 的常见来源

7.3.1 概述

调查和诉讼中潜在的相关 ESI 可能要在广泛的来源中找到。为了帮助识别这些来源，重要的是考虑哪些系统和资源是在保管方的直接控制和访问下，哪些是不在其控制下。

7.3.2 保管方数据源

保管方 ESI 来源是指单个保管方直接保管或控制的 ESI 来源，包括但不限于以下来源：

——计算机：潜在的相关 ESI 可能存在于保管方的台式机、笔记本电脑或家用电脑，以及移动存储媒体，如 U 盘、外接硬盘、DVD 或 CD；

——移动设备：潜在的相关 ESI 可能存在于保管方的个人设备，如手机、智能手机、平板电脑、全球卫星导航系统等。

从企业的角度看，7.3.3 中列出的数据库和应用程序、网络存储、备份和电子存档也可能被视为保管方数据源。

7.3.3 非保管方数据源

非保管方数据源或者是组织内部的，或者是组织外部的。

内部数据源是指一个或多个保管方可访问的，但另外一个保管方（如 ICT 管理员）控制的数据源。组织内部非保管方数据源包括但不限于以下来源：

——数据库和应用程序：与动态数据库相关的 ESI 在某些情况下可能是相关 ESI，并且根据不同的问题，某一事项可能涉及组织的电子文档管理系统（EDMS），电子记录管理系统（ERMS）或协作工具；

——网络存储：ESI 可能存储在组织内部网络的不同地方（如共享存储器、网络磁盘和服务器），以及采用专用存储技术，如网络连接存储（NAS）和存储区域网络（SAN）；

——备份：ESI 从信息系统复制或备份到数据保护系统，如磁带或其他媒体。

——电子存档：电子或数字存档档案（数据存储库）中包含的 ESI 是典型的官方业务记录、出于合规目的保留的文档、遗产文档（具有历史价值）等；

组织外部的非保管方数据源包括但不限于以下来源：

——云存储：云计算解决方案通常为许多应用程序以及灾难恢复和业务持续性目的所用；

——社交媒体：社交媒体中包含 ESI，主要为社交目的在人群中分享的，但已经越来越多地用于商业目的，这可能会带来挑战，因为它往往不在组织的直接控制范围内。

7.3.4 可能排除的 ESI 来源

并非所有的 ESI 来源都需要被保全；以下 ESI 来源可能无法发现：

——硬盘上已删除、闲置或未分配的数据；

——随机存取存储器（RAM）上的数据或其他临时数据；

——元数据字段中频繁自动更新的数据，如上次打开的日期；

注：需要仔细考虑和咨询要保全哪些元数据字段，因为一旦被更改，就很难，甚至通常是不可能复原的。例如，元数据信息（如文件初次创建或最后修改的时间）在过程后期过滤 ESI 时可能至关重要，因此需要将其保持。

——与其他地方更易访问的数据基本上重复的备份数据；

——临时使用的测试数据；

——其他形式的 ESI，其保全需要在正常业务过程中不会用到的特别措施。

尝试与诉讼对手或调查人员达成不需要保全此类 ESI 的协议可能是有益的。

7.4 ESI 的呈现

7.4.1 概述

与特定事项相关的 ESI 可能包括字处理文件、电子表格、电子邮件、数据库、图画、照片、专有应用程序数据，网站数据，语音邮件等。ESI 文件的收集和产出格式可能分为原生，近似原生，图像（近似纸质）和纸质。

7.4.2 原生格式

创建和维护的文件格式称为原生文件。对于不是为打印而创建的文件，如电子表格和小型数据库，通常建议使用原生格式。对于某些文件类型，原生格式可能是充分生成 ESI 的唯一方式。

以原生格式的产出不需要生成方法承担将数据转换成其他格式的成本。然而，接收方可能需要相应的原生应用程序或生成方的专有软件来打开文件。

如果一方选择将 ESI 转换为不同格式，则可能需要采取措施，以确保 ESI 的元素（如元数据）不会在转换过程中无意丢失或变得不清晰。

7.4.3 近似原生格式

有些文件（如电子邮件和数据库）如果没有某种形式的转换，就无法评审。例如，大多数电子邮件文件都需要提取并转换为单独的文件，因此，原始的格式就会改变，不再是原生格式。

大型数据库和数据汇编通常以近似原生格式生成。数据库可能包含大量完全分不开的数据表。企业业务系统可能包含数百个表和数千个数据字段。这些系统可能需要使用各种数据库平台和专有软件。由于这些原因，大型数据库和数据汇编通常不以原生格式生成。这些数据库通常需要由合适的人员进行分析，以识别出相关数据和确定合适的近似原生格式。

从这些数据库中导出的通常是带分隔的文本文件。在某些情况下，文本文件是通过数据库图、数据字典、元数据或软件生成的。数据也可能导出为常用的电子表格格式。

7.4.4 图像（近似纸质）格式

ESI 也可能以图像或近似纸质的格式生成。图像生成的过程是将 ESI 转换成或将纸质文件扫描为一种不可编辑的数字文件。在这个过程中，采用“图片”作为以纸质形式或将以纸质形式存在的文件。基于文档、打印机或计算机中的打印设置，图像中的数据可能会被改变或丢失。有必要具备电子信息发现和图像生成工具领域的专业知识，以尽量减少这些问题。

7.4.5 硬拷贝

与其以电子形式处理 ESI，不如将 ESI 记录在某种形式的硬拷贝（如纸质打印件、照片等）上，这样可能既合理又可行，然后整个过程都会使用这种硬拷贝。与将 ESI 转换为图像格式（见 7.4.4）一样，将 ESI 记录为某种形式的硬拷贝可能会导致数据丢失或改变。因此，在打印或图像生成过程中，可能有必要具备电子信息发现和图像生成工具方面的专业知识，以尽量减少这些问题。

7.4.6 发现中的非 ESI 部分

虽然大多数业务信息以电子格式存储，但发现项目可能至少涉及传统硬拷贝或纸质文件的一小部分（这与 7.4.5 中描述的打印 ESI 不同）。如果决定收集硬拷贝，则收集内容越是集中在手头事项，将文档集细化到相关子集所需的工作就可能越少。

有了一套集中的硬拷贝文档后，一种选择是将它们扫描成电子格式[2)]，然后将其与电子文档一起纳入评审过程。也许要考虑这些文档是否有家族成员式的关系。这些文档在生成之前需要评审、标记，可能的话，还要脱敏。这能使使用与 ESI 相同的技术和过程来管理硬拷贝文档的效率更高。

8 电子信息发现过程

8.1 概述

电子信息发现是传统发现的一种形式，通常涉及识别、保全、收集、处理、评审、分析或产出可能与特定事项相关的电子留存信息（ESI）。潜在的 ESI 通常是：

——通过与员工和 ICT 人员的反复研究和面谈，来识别；

——通过采取措施通知相关人员，以避免删除或销毁或者调用自动删除或销毁的系统，来保全；

——通过使用一种或多种能保全数据完整性的提取或收集方法，从原始来源收集；

——通过使用一种或多种技术工具，为文本可搜索建立索引，来处理；

——通过一种或多种方式，由法律或主题事项专家在各种工具和能有效使用这些工具的专业技术人

2).这一过程可能涉及使用光学字符识别（OCR）技术来生成可搜索的结果，以及由人工评审者进行编码以捕获相关元数据。

员协助下，来评审其相关性；

——为协助实现事项的目标，来分析；

——以合理可用的方式或各方同意的形式，来产出给各请求方。

在本文件中，通过在如“识别”这种通用行动的名称前加上“ESI”（如 ESI 识别）来表述特定的电子信息发现过程环节。图 1 显示了所有电子信息发现过程环节。

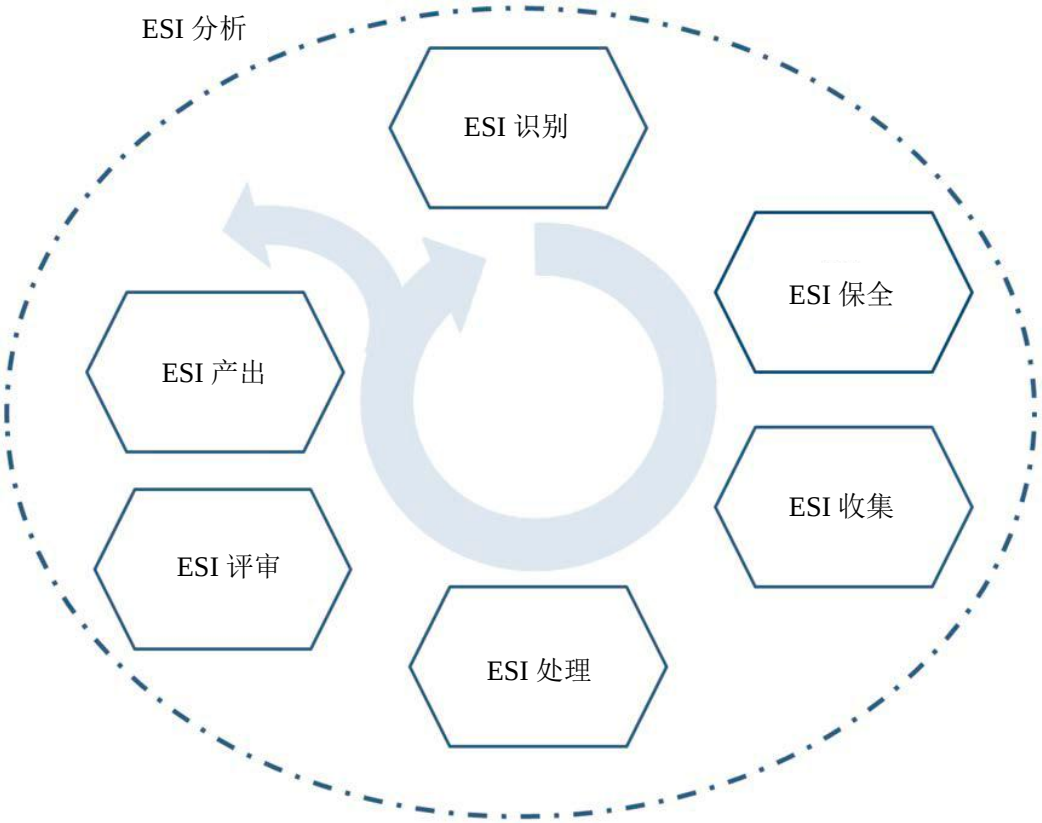

图1 电子信息发现过程环节

图 1 还展示了电子信息发现过程各环节之间的相互关系。将 ESI 分析置于一个外环，旨在表明分析可能选择性地与其他电子信息发现过程环节结合进行。例如，一个可能的场景是，ESI 识别可能需要执行 ESI 分析，然后再回到 ESI 识别以进行其他活动。这种过程流可能从一个过程环节（ESI 分析除外）移动到另一个，，然后返回到之前的过程环节。最后，电子信息发现通常是一个有序的迭代过程，使用部分或全部过程环节进行，这一点也在图 1 中用圆形箭头表示了出来。

本文件旨在服务于多个利益相关方的利益，包括大型实体与小型实体、涉及法律实体与不涉及法律实体等。虽然描述了一个强健的电子信息发现过程，但无意强行使用不必要的过程。具有复杂电子信息发现问题的大型企业可能使用本文件所描述的大部分或全部过程环节，但对于小型组织或小型事项，这可能是不切实际的。一个事项也许只使用上述过程环节的子集。

假设少量电子邮件是一个事项的核心。发现活动从 ESI 识别开始，然后利用 ESI 分析确定是否可能存在其他来源。如果没有，则收集相关的电子邮件（跳过保全）。然后进行 ESI 分析，以确定 ESI 收集是否充分，并做出跳过 ESI 处理和 ESI 评审的合理决定。之后，作为 ESI 产出过程环节的一部分，以原生格式生成与该事项相关的电子邮件。

在一些司法管辖区、法院、立法机构和政府监管机构已制定了有关组织如何识别 ESI 的规则，尤其是在民事和刑事诉讼、调查和审计中。在这样的司法管辖区内，当对组织的诉讼或调查是合理预期或未

决定时，组织有义务采取合理措施，识别并保全潜在的相关 ESI。作为识别和保全 ESI 责任的基础，组织要能够及时定位和保全潜在的相关 ESI。此外，可能期望组织制定适当的 ESI 管理制度，并遵守这些规则。

8.2 ESI 识别

ESI 识别是电子信息发现过程中的一个环节，在这个环节中，一方处于任何原因（如合理预期诉讼、收到诉前保全请求、检查请求、律师函、停止和终止函、补救通知，甚至与对方或其律师辩论），采取措施识别可能与事项相关的信息。

ESI 识别本质上是理解相关主题事项，并识别可能合理通向与该主题事项相关的潜在相关信息的个人、部门和 ESI 来源。就电子信息而言，ESI 来源可能多种多样且复杂，可能包括内部和外部 ESI 位置、各种服务器类型和各式各样的电子设备。此外，还需要考虑遗产系统（存档、备份、不活跃系统和数据）。通常情况下，为了识别对特定事项可能重要的来源，将对可能参与问题事件的个人以及 ICT 部门人员进行访谈，以确定他们是否拥有或了解相关 ESI。这些访谈可能导致对其他关键参与者或相关 ESI 可能位置的识别或排除。

建立访谈模板和追踪机制非常适合于识别和访谈个人，并记录他们对数据类型和数据位置的了解，以便发现其中潜在的相关信息。基于业务部门或个人角色，模板中可能会有不同的问题。该文档能帮助规划、实施和跟踪识别过程中的活动和答复，并在出现有关额外（或可能冗余）信息源的问题时能作为参考点。此外，如果有疑问，它能用来证明识别过程是恰当的。

8.3 ESI 保全

ESI 保全是电子信息发现过程中的一个环节，在这个环节中，触发事件发生后，尽可能保护那些已被识别与某一事项的保护义务范围有关的信息不被修改或销毁。这不仅包括个人或组织拥有的潜在相关 ESI，还包括个人或组织拥有的外部信息。由于保全义务可能因司法管辖区不同而不同，因此没有单一的标准来衡量保全活动的适当性，或一方当事人因未履行其保全义务而面临的风险或潜在责任。

电子信息能通过对其进行的收集（即直接从其来源复制）或采取措施确保其在通常驻留的地方得到妥善保存（如保管方自行保全或使用技术手段就地保全）来得到保全，之后，是否可被收集，取决于事项的需要和保全策略。收集 ESI 副本的行为本身就是一种保全形式，ESI 保全和 ESI 收集能一起进行。ESI 保全的关键区别在于，它涉及采取措施保全 ESI，使其不被修改。要能保证能中止那些可能删除或改变潜在响应数据的常规程序（如备份磁带轮换或常规记录销毁）。

8.4 ESI 收集

ESI 收集是电子信息发现过程中的一个环节，在这个环节中，从已保全的 ESI 和硬拷贝文档中创建一个数据集；然后，使该数据集对进一步的处理和最终评审可用。

收集本质上是一种复制操作，在这种操作中，获取目标文件的副本或镜像，并将其包含在一个数据集中，然后将这个数据集传递给下游处理和评审。有多种工具和方法能用于收集，从能恢复用户已删除的文件，到涉及用户创建的目标文件的简单导出。在任何给定情况下，适用的特定工具和方法可能会因收集文件的设备的性质不同（如台式计算机与智能手机），所收集文件的性质不同（如电子邮件与微博），作为收集场合的诉讼的性质不同（如刑事诉讼与民事诉讼），以及进行诉讼或调查的司法管辖区不同而有所不同。

8.5 ESI 处理

ESI 处理是电子信息发现过程中的一个环节，在这一环节中，在保全和收集数据后，采取措施使数据可搜索并以可评审的格式呈现。这可能涉及一种或多种方法。此外，可能使用多种技术来缩小响应数

据体量。首先是删除系统文件和其他与事项可能无关的文件。随后通常是使用一种或多种过滤技术，从日期范围过滤到文件类型过滤、再到使用搜索字、概念或基于文本的处理数据的预测算法对元数据或文本进行基本过滤。ESI 处理中也包括数据的去重。ESI 处理中使用的技术通常在各方之间达成一致，以确保就如何处理数据并在随后的电子信息发现过程环节中使其可用达成共识。

除了数据缩减外，处理数据还需要有可辩解的审计程序、质量控制措施（包括数据确认）以及文件化记录的保管链（包括在 ESI 处理过程中对文件转换的追踪）的支持。

8.6 ESI 评审

ESI 评审是电子信息发现过程中的一个环节，这个环节侧重于根据特定准则筛选 ESI。本质上，就是将符合产出准则的文档与那些不符合的分开。

进行评审的方法有很多种，从经过长期实践的线性人工评审到最近广泛使用的先进信息检索工具和方法。本文件旨在适用于所有方法。

8.7 ESI 分析

ESI 分析是电子信息发现过程中的一个环节，在这个环节中，将各种工具和方法应用于 ESI，以聚集能用于实现各不同电子信息发现过程环节目标的信息。因此，分析是一种活动，能用来支持电子信息发现过程中的任何迭代步骤（识别、保全、收集、处理、评审和产出）。

有多种工具和方法能用于分析目的。在任何给定情况下，哪种工具或方法适用，取决于电子信息发现过程环节所需的数据分析，以及需要数据分析提供答案的特定问题。

8.8 ESI 产出

ESI 产出是电子信息发现过程中的一个环节，在这个环节中，一方准备文件以交付给其他方。为产出准备数据的规程通常在初始项目规划中商定。

根据各方之间达成的协议，数据能以电子或纸质形式生成。产出格式可能有所不同，包括从扫描纸质或物理纸质获得的原生和近似原生图像的组合。在为产出准备数据时，需要考虑接收者可用的技术吸收能力类型。这可能包括接收者评审产出的能力和可用工具。通常，为了限制成本，会根据数据体量和产出格式检查成本考量。

与在电子信息发现过程中的许多环节一样，需要注意记录产出或特权日志中生成的文件。该文档旨在提供有关生成了哪些文件以及保留了哪些文件的详细信息。

9 其他考虑事项

9.1 ESI 的呈现

虽然不被视为电子信息发现过程中的一个环节，但了解 ESI 最终如何用于某一事项（如在法庭上呈现）还是很重要的。

ESI 的呈现对律师和律师助理来说可能是一种挑战。在过去，当庭出示以纸质形式呈现，现在在许多案件中仍然如此。在过去十年里，技术的发展已使近似纸质或“图像”格式的呈现更加容易。由于电子留存信息的性质以及原生和近似原生文档产出的出现，现在一些案件需要法律团队以原生格式进行当庭出示。

9.2 保管链和出处

根据手头的事项，追踪或确定 ESI 相关的创建、修改历史、影响、所有权、其他出处或族系信息可能很重要。其中一些信息可能包含在元数据中，也可能作为电子信息发现过程的一部分被生成。这种出

处信息对于对ESI的质量、完整性和真实性做出明智判断可能至关重要。

在一些案件中，出处信息不足以证明质量、完整性和真实性。在此类案件中（如刑事调查或起诉），需要提供显示ESI的保管、控制、转移和处置的正式时间顺序文档。重要的是，意识到到何时需要保管链的证据，并确保满足要求。

参 考 文 献

[1] ISO Guide 73, Risk management — Vocabulary
[2] ISO 15489-1:2016, Information and documentation — Records management — Part 1: Concepts and principles
[3] ISO/IEC 27037, Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence
[4] ISO/IEC 27040:2015, Information technology — Security techniques — Storage security
[5] ISO/IEC 27041, Information technology — Security techniques — Guidance on assuring suitability and adequacy of incident investigative method
[6] ISO/IEC 27042, Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence
[7] ISO/IEC 27043, Information technology — Security techniques — Incident investigation principles and processes
[8] ISO/IEC 27050-2, Information technology — Electronic discovery — Part 2: Guidance for governance and management of electronic discovery
[9] ISO/IEC 27050-3, Information technology — Electronic discovery — Part 3: Code of practice for electronic discovery
[10] ISO/IEC 27050-4, Information technology — Electronic discovery — Part 4: Technical readiness
[11] ISO/IEC 29100, Information technology — Security techniques — Privacy framework
[12] ISO/IEC 38500, Information technology — Governance of IT for the organization
[13] Electronic D.R.M. (EDRM), http://www.edrm.net
[14] Good practice guide to eDiscovery in Ireland, Version 1.0, 16 April 2013, https://docplayer.net/6285095-Good-practice-guide-to-electronic-discovery-in-ireland.html
[15] New York Bar Association Best Practices in E-Discovery in New York State and Federal Courts, Version 2.0, December 2012, http://www.nysba.org
[16] Seventh Circuit Electronic Discovery Pilot Program — Final Report on Phase Two, May 2012, http://www.discoverypilot.com/sites/default/files/Phase-Two-Final-Report-Appendix.pdf