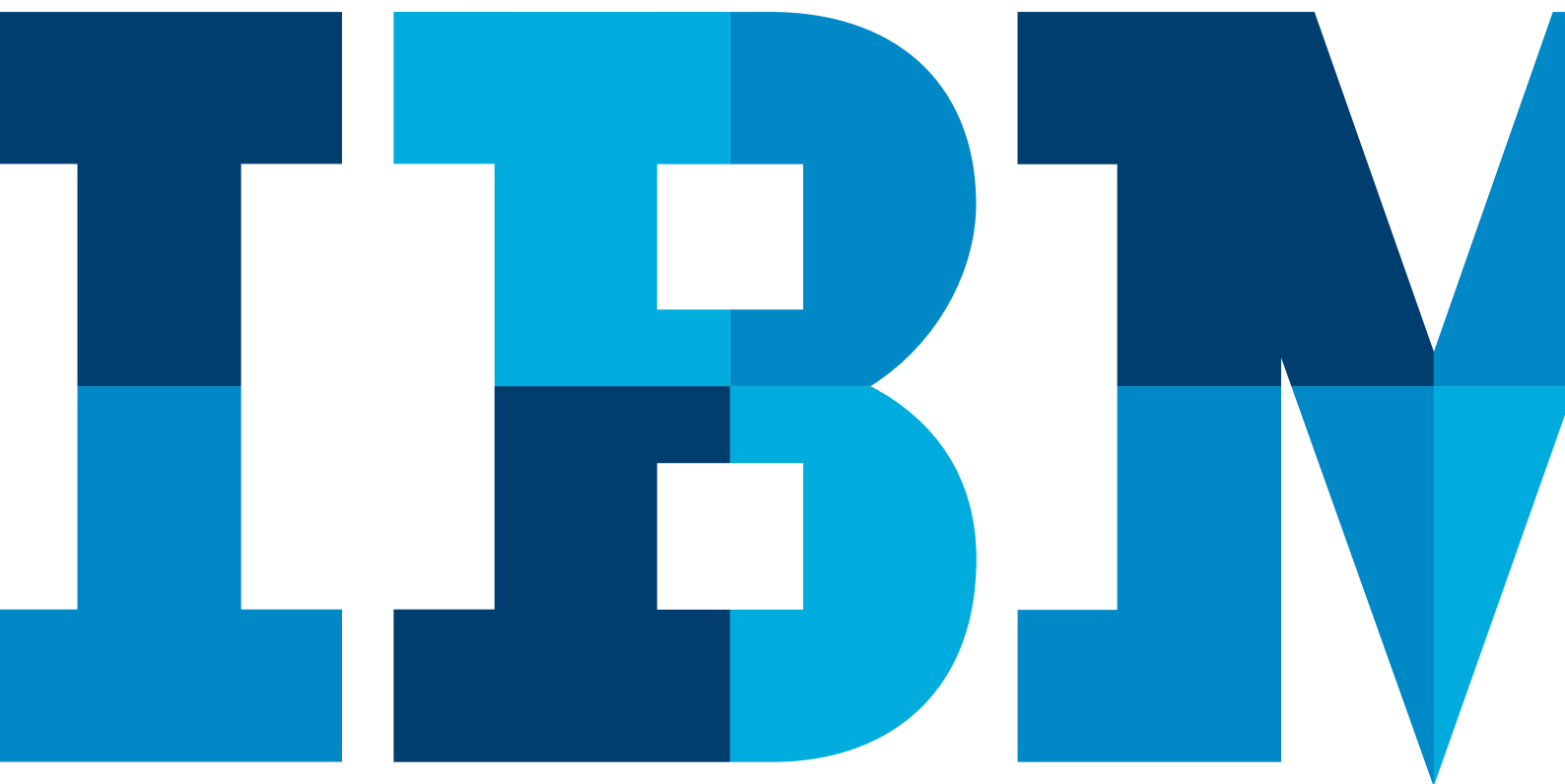


捕获的五大阶段

文档策略手册, 从简单扫描到企业捕获面面俱到



作者: Kevin Craine Document Magazine创刊编辑

什么是捕获?

您已经拥有一台扫描仪, 同时还有大量纸质文件需要管理。现在, 您期待借助“捕获”解决方案改进流程, 使您的工作和生活更加轻松。或者, 您可能早已投资捕获系统, 但很想知道如何发挥这一解决方案的最大价值。

捕获系统可驱动扫描仪, 帮助简化向文档图像附加索引的过程, 从而实现高效存储和检索。事实上, 某些捕获系统甚至能够完成更多的工作。许多解决方案支持通过浏览器在世界各地进行扫描或自动从表单(甚至是手写文档)中提取数据。

经过20多年的发展, 捕获系统已然从简单的基本扫描解决方案演变成十分精密昂贵的企业级文档自动化系统。

本文将捕获系统归为几个基本类别, 称作捕获的五大阶段, 并分别介绍各个阶段的功能和优势。根据企业的文档管理策略, 每个捕获阶段均能够为企业带来价值。本文旨在协助识别适当的产品类别, 从而帮助您达成目标, 同时为全面了解捕获功能提供指导。

捕捉的五大阶段

随着企业对长期奉行的依赖于纸张的特定功能做出调整, 特别是成本控制需求变得越来越迫切, 法规监管挑战性日益增强, 一场波及全球的革命正在悄然兴起。扫描文档并杜绝纸张是一个不错的起点, 但当前采取的最佳实践。

目标在于对任何可能的数据源(例如, 物理文档、电子邮件、文件附件、传真和数据流等)提供的所有相关数据实现全自动化提取。自动验证数据是否正确是另一个关键的成功因素, 无缝集成任意数量的后端存储系统和应用程序同样十分重要。

因此问题变成了: 您需要保持在捕获流的哪个位置, 才能最有效地帮助企业减轻对纸张的依赖、降低劳动成本、以及集成所有关键业务活动的数据和文档?

“捕获流”从基本扫描一致延伸到文档全面自动化—哪个捕获阶段适用于您的企业?



第一阶段——扫描和存储

第一阶段的特点在于能够扫描纸质文档并以数字形式进行存储。此阶段也成为“图像处理”，是捕获流的入门阶段。其中包括扫描硬件和基本软件，让用户能够手动输入文档相关信息以便进行搜索和检索。通过将关键字数据键入屏幕上的字段，用户即可通过标题、日期或简短描述来检索文档—例如：Acme Company发票，2010年5月19日。自此，文档图像将存储在计算机硬盘驱动器或文件服务器上，以便日后检索。

paper from paper-intensive operations.分析师

优势

“扫描和存储”只是整个流程的开端，预计该流程将消除工作区中仍以纸张形式存在的80%的日常处理文档。因此，文档处理占到办公员工60%的时间，并占45%的人力成本。总之，贵公司每花费一美元制作最终文档，就要相应地花费十美元对这一流程进行管理。减少档案柜和仓库设施的存储空间也是从节约成本的角度考虑的。无论如何，在当今竞争激烈的经济环境下，继续基于纸质文件开展业务的企业将疲于控制固有成本和低效的流程，同时会发现自己处于劣势。

局限性

将纸质文档扫描成图像文件(例如PDF或TIFF)是开始消除纸质文档负担的直接途径，但在Windows文件目录中查找文档往往相当困难，特别是随着文档(及相关文档)数量的不断增加，查找难度也在不断增加。需要额外的人工来区分文档类型并进行预先分类、实际扫描页面以及手动输入索引关键字。当采用第一阶段的方法时，考虑这些额外的流程需求十分重要。“扫描和存储”可降低对纸质文件的依赖，并能够更加即时地访问文档，但手动数据录入的局限性和文档存储的“大杂烩”特性意味着，您需要解决大量的人工和工作效率问题。



在当今竞争激烈的经济环境下，加速纸质文件到电子文档转换速度的企业势必将占得先机。

您是否应该进入第二阶段？

通过对此项基本“扫描和存储”流程应用识别(有时称作OCR)技术，您可能会发现一些额外的好处。需要思考的问题：

- 是否很难找到已扫描的文档？
- 您是否花费不必要的时间手动输入索引信息？
- 是否需要多名员工扫描文档并编制索引？
- 重要信息是否“埋葬”在以电子形式存储的已扫描文档海洋之中？
- 您能否通过加强流程自动化来获得额外的好处？

如果上述任何问题的答案是“肯定”的，那么通过进入第二阶段，您很可能会实现额外的流程改进，并进一步提高投资回报。



示例

芝加哥卫生局(CDPH)需要一种能够快速对灾难做出响应并能跟踪配发药品和接收人的解决方案。

当发生灾难时,该城市需要依据医疗档案填写信息。CDPH采用Datacap Taskmaster快速自动扫描这些表单并编写索引。

- 快速可靠地进行扫描,并将索引编入IBM
- 完整的报告功能
- 海量数据处理

阅读完整的案例研究: [芝加哥卫生局利用捕获技术制定应急预案](#)

第二阶段——自动编制索引

这一阶段将充分发挥技术的作用,从文档图像中“读取”数据,从而让您以全新的方式利用这些信息。光学字符识别(OCR)可提取关键信息,并自动填充索引字段。PO号码、日期或申报数额等标准数据都是第二阶段捕获的常见数据示例。通常情况下,需要用预先设定的模板来定位数据,从而将这些系统局限在大量涌入的结构化文档范围之内。

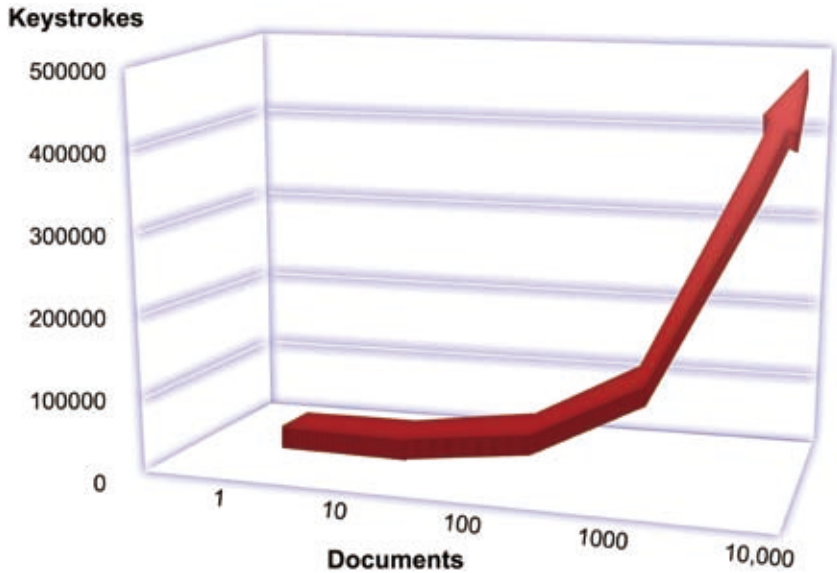
有时称作“带状OCR”,可以提取这些关键数据,否则需要进行手动输入。

这一阶段的其他进展还包括运用各种能够迅速解读的条形码,揭示范围更加广泛的一系列数据。用户可以应用多种简单的业务规则(包括验证日期是否满足标准日期格式等),进而提高捕获数据的准确性以及消除日常业务流程“文字整理”的相关人员和工作量需求。

优势

除消除纸质文档并且回收存储空间外,还能显著缩短周期时间、降低错误率和人员需求都是实实在在的成本节约,都可以通过自动化第一阶段的“扫描和存储”流程来实现。基本业务规则则有助于推动对收集数据进行核查和验证。例如,如果系统对自身捕获的数据的信任程度较低,可以将不确信的字段突出显示并路由至核查员进行分析。用户可以为简单文档手动编写索引,每份文档平均需要点击48次。随着文档数量的上升,点击次数也成倍增加。

通过运用整页OCR创建可搜索的PDF文件,从而简化并加速对扫描文档详细信息的搜索过程。员工不再需要每天花费绝大部分时间搜索信息,借助新层次的访问和响应支持不仅能够削减成本,还能大幅改进工作流程。



为简单文档手动编写索引,每份文档平均需要点击48次。随着文档数量的上升,点击次数也成倍增加。

局限性

虽然本阶段对第一阶段的

扫描和存储功能实现了自动化并且确实十分有效,但仍然需要相当数量的人工实施流程。贵公司的工作人员将需要预先分类并区分文档,插入分隔页(通常带有条形码),然后人工扫描文档并验证所收集数据的完整性。

提供带状OCR的许多捕获系统均需要程序员构建文档模板,而程序员往往只能集中处理收到的最多的20%文档,因为文档处理既昂贵又耗时。这意味着其余80%的文档类型没有模板,因此仍需要进行手工数据录入。

此外,许多捕获供应商除收取软件费用外,还会“按点击次数”或按图像收费。在这种情况下,向系统添加更多文档可最大限度利用捕获投资,但实际上也增加了总体拥有成本。

更多研究:

Datacap推出了[自动索引解决方案](#),从而取消了OCR“按点击次数”定价的惯例。独特的自动模板生成程序使日常用户能够在处理文档时动态创建区域,并且无需进行编程,也不存在任何模板数量限制。

在线研讨会: [FastDoc Capture快速演示](#)

您是否应该进入第三阶段?

高级数据提取技术可将信息捕获范围扩大到已扫描的文档之外。需要思考的问题:

- 您是否需要将文档上所含的信息重新输入后端系统(如ERP、业务线或者申报仲裁系统)?
- 您能否依据流程改进和最佳实践,运用收集而来的信息完成更多任务?
- 运用更加先进的自动化技术能否进一步减少人工、缩短周期时间并提高捕获数据的准确性?

如果上述任何问题的答案是“肯定”的,那么通过进入第三阶段,您很可能实现流程改进,并进一步提高投资回报。



第三阶段可获得巨大的收益，这得益于可重用的模块化业务规则，以及让创建、管理和部署工作更为轻松的设计框架。

示例

道琼斯公司(《华尔街日报》、《巴伦周刊》及其他一些领先金融出版物的发行商)利用智能数据提取技术自动完成订阅流程。该公司实施Datacap Taskmaster从各种订阅卡片和文档中扫描并提取主要订户数据，每次可能需要使用50种不同的设计。

- 处理周期时间减半
- 强大的报告功能
- 减少了五个全职人员需求

阅读完整的案例研究: [道琼斯公司简化订阅处理流程](#)

第三阶段——智能数据

提取

第三阶段的特色在于超越了读取基本“标题”信息的限制，向提取多个字段和排列项详细信息过渡。智能数据提取策略并非简单收集数据来编制索引，而是从文档中此前或许不为所知的数据位置上提取更加详细的数据，如发票或医疗文档。自此以后，这些系统会将双管齐下交付也是这一阶段的突出特色。扫描得到的图像和捕获的此类信息传输至各种后端数据库和应用程序(如ERP系统或医保申报仲裁系统)。高级数据提取技术超越了结构化文档的限制，可处理多种类型的文档，而智能数据识别技术则可协助系统读取手写和图形字段，并可充分利用高级数据验证功能。

索引数据均将传输到IBM/FileNet或其他内容存储系统。

同时捕获的排列项信息会传输到任意数量的ERP或业务应用程序系统。纳入这项双重交付技术的同时，可能还要应用更加复杂的业务规则，并且可能需要做出更加复杂和自动化的决策。

例如在处理发票时，您或许希望捕获并利用六七个索引字段来简化检索操作，但业务数据库应用程序(例如，位于SAP或Oracle中)当前只能接收和利用相同的数据。最后，借助高级业务规则，人员的职能角色可能会发生显著变化，从手动数据录入向更高价值的角色发展，逐步成为重点保证数据质量的专家。第三阶段的优势得益于可重用的模块化程序规则，以及让创建、管理和部署工作更为轻松的设计框架。

优势

在第三阶段中，您能够同时获得第一和第二两个阶段的优势—消除纸质文档、回收存储空间、提高已扫描文档的检索效率，但却增加了信息维度，将各种重要信息注入企业内容管理系统和应用程序。这可能会演变成您的文档策略的“神奇元素”；第三阶段周到的集成功能和高级数据提取功能使通过集成ERP、Web和 workflow 解决方案将“捕获”技术纳入企业战略，实现显著的业务优势和流程改进成为可能。

局限性

许多高级数据捕获系统均需要十分复杂的编程和“培训”方案，以便引导系统了解有关要处理的文档的信息。因而业务规则往往规模庞大，并且需要采用专业技术进行设计、部署和维护。但许多新型解决方案成功解决了这个问题，并能让业务部门人员拥有设计复杂功能的能力。将此项职责从以往晦涩难懂的信息技术组织转移到基层用户至关重要，因为人们需要更加迅速地对不断变化的业务需求做出响应。全新的医疗保健立法、环境限额和交易法案以及进出口证件办理规则变化只是其中的几个例子，足见当今世界的合规性与立法复杂程度都在不断提升。结果往往需要提供响应级别更高的支持您能否通过第四阶段的捕获技术获益？

更多研究：

Datacap开发了智能数据提取解决方案，其中包括[基于规则的灵活平台](#)，用以构建功能强大的应用程序。同时还提供了预构建的应用程序，以便[捕获发票、医保申报](#)并进行[优势说明](#)。

在线研讨会：[Exel Transportation实现发票处理时间减半](#)

您很可能通过使捕获技术超越集中处理环境的限制，并扩展到整个企业的不同位置而获益。需要思考的问题：

- 您是否需要打包纸质文档并将其从遥远位置运送到集中场所进行扫描？
- 工作组级扫描是否由于缺乏统一方法而分散在多台独立扫描仪或多功能设备上？
- 是否存在某些部门和个人盲目等待文档分批运送到中心机构进行处理？
- 您能否通过将验证任务转交给远程办公人员或家庭办公人员而获益？

如果上述任何问题的答案是“肯定”的，那么通过进入第四阶段，您很可能会实现额外的流程改进，并进一步提高投资回报。



**分布式捕获策略可节约邮
寄和速递费用、加快数据
输入并加快关键业务文档
处理过程。**

示例

SIRVA, Inc.是一家总部设在伊利诺伊州芝加哥附近的搬运公司。旗下包括Allied、Global和North American Van Lines多家公司。该公司每年处理超过300,000项搬运业务,每项业务平均生成25页运输单据,过去均连夜送交SIRVA处理中心处理。2002年,该公司采用了Datacap Taskmaster Web(一种基于浏览器的扫描和索引编制解决方案),并要求其1100家代理在创建单据时进行扫描,这样只需几秒钟即可上传至总部。

- 平均处理时间减少两天
- 降低了扫描成本
- 消除了运输成本

阅读完整的案例研究: **SIRVA要求
1100家代理全面采用扫描仪**

第四阶段——分布式捕获

此阶段建立在第一、二、三阶段的总体优势基础之上,但新增了在文档生成地点扫描文档这一附加价值,从而省去了将文档运送到中央扫描中心进行扫描的麻烦。这是一项重要的优势,因为纸质文档捕获历来都是一项集中式业务功能。相反,“分布式捕获”是一种促使文档在分散位置(即文档生成位置)进入业务流程的策略。

对于在众多地点生成大量纸质文档的企业而言,将扫描和索引编制功能转移到文档生成地点可节省邮寄和速递费用、加快数据输入并加快关键业务文档处理过程。

分布式捕获解决方案现已在各行各业(无论文档量大小)中广泛采用,这种方法能够跨越各个办公区域和地点(无论是位于一幢大楼的不同楼层还是遍布世界各地)建立高效的文档管理系统。数据准确性验证也可以采取分布式途径,通常分配给最熟悉这些文档及相关工作要求的业务专家。这样可实现响应速度更快且更加直观的质量控制。

优势

运用第四阶段的分布式捕获方法,企业能够缩短事务处理时间、减少运输和邮寄费用,并提高工作效率开辟了新的机遇。

虽然运输公司、银行、保险公司和贷款经济公司等纸质文件密集型企业是分布式捕获技术的早期采用者,但各垂直行业企业(无论规模大小)均发现分布式扫描有助于它们节省时间和资金、减少工作量并提高效率。事实上,随着企业设立的分散式地区办事处越来越多、具有越来越多的远程员工和远距离客户,分布式捕获将成为一种重要工具,以便简化和促进任务关键信息处理。

局限性

扫描和捕获活动远离集中处理设施这一事实可能带来的人工和管理问题可以忽略不计。毕竟，您现在可以要求业务部门中的高价值“知识员工”通过扫描仪、复印机或多功能设备处理纸质文档，并将此工作作为他们日常工作的一部分。虽然这将有助于促进形成无纸化流程文化，但却提出了额外的行政工作，这一点必须予以考虑。企业不再需要聘请专业设备图像处理员预先分类文件、插入分隔页，而后验证数据完整性，现在这项任务将落在部门员工身上。它的真正意义在于，您不再需要走来走去满怀期待地渴求各位操作员帮助解决问题，虽然很多系统可让您查看所有远程工作站正在进行的工作。您需要记录正确的纸质文档处理。

更多研究:

Datacap推出[基于浏览器的捕获解决方案](#)，目的是在远程位置部署捕获功能。在各分支机构扫描文档，上传至总部，利用家庭办公人员编制索引并验证文档，或随时随地从任何计算机对捕获系统进行管理。

在线研讨会: [Murphy Hoffman公司将扫描任务落实到65个服务中心](#)

进入下一阶段

扫描文件可消除纸质文件、捕获数据可执行强大的搜索和检索功能、高级业务规则，以及能够在不同位置的各种系统中充分利用大量的提取数据—这些都是您从第一阶段步入第四阶段将会获得的显著优势。那么最后的第五阶段还有哪些独特优势呢？以下是一些需要思考的问题：

您是否经常打印电子邮件和文件附件，目的只是要扫描纸质文档以便将信息填入内容管理系统？

- 您是否需要进入多个存储库收集流程所需的数据和文档？
- 在各种分散、复杂而又孤立的存储库之间搜索关键信息是否限制了自身对不断变化的企业需求进行响应的能力？

如果上述任何问题的答案是“肯定”的，那么通过进入第五阶段，您很可能实现额外的流程改进，并进一步提高投资回报。



虽然对于许多企业而言，纸质文档依然是主要媒介，然而当今的企业环境已经发生重大转变，纳入了各种电子数据。第五阶段旨在应对所有这些可能发生的状况。

示例应用程序

BlueCross BlueShield of Arizona 早在2000年安装Datacap医疗申报捕获系统后已经实施了企业捕获解决方案。在评估并确定申报处理操作的工作效率的确提高后，它们扩大了实施范围，并开始将文档添加至系统。如今，该保险公司每天扫描50,000页文档，再加上传真和电子文档，从而实现医疗申报、特别清单、注册、法律和人事记录和信件数字化。多个远程扫描中心还可将图像注入真正的通用捕获门户。

- 简化文档策略
- 以更少的人员将文档输入速度提高三倍
- 提高精度并提升控制性

阅读完整的案例研究：**BlueCross BlueShield of Arizona应用企业捕获解决方案**

步骤，并留意远程员工是否遵从这些步骤，否则将遭遇数据质量“下滑”及图像标签错误等问题。

第五阶段——企业捕获

第五阶段利用最新的创新捕获技术，通过自动分类和到整个企业各业务系统与存储库的多个连接，自动输入所有传入文档的扫描、传真和电子邮件内容。我们将其称为捕获门户，企业在这里以最少的手工劳动将传入企业的所有文档快速准确地转换为数字化内容，然后传输到能够最有效地利用这些数据的人员或系统。

此阶段要求系统足够精密，能够以文档可能传入的格式接收文档，包括通过扫描仪、发送传真、从多功能打印设备、文件目录传入的Word或Excel格式的桌面文档，XML或HTML数据文档，或者通过EDI传入的传统打印数据流。此阶段的解决方案还能够读取任何种类的条形码或标记识别系统，并且基本上不存在任何业务规则应用广度和复杂度限制。第五阶段的系统还能连接任意数量的企业内容管理和存储系统。

在此阶段，任何人均可输入任何传入文档，该系统将以智能方式辨别文件属于哪种类型。如果是纸质文档，则无需预先分类和分页即可直接处理。自此，该系统将采用整页OCR执行高级文本分析，自动确定处理该文档的方式。最重要的是，文档能够以任意格式传入，并且可以无缝集成到该流程中。系统将捕获、验证这些文档中的所有相关数据，并且无需手动录入即可自动将它们发送到后端系统。因此，只需少量人员即可管理各种异常事件。因此可以节约大量人力，因为不再需要预先分类文档，数据录入流程也已取消，扫描和验证操作可以随时随地完成，您不再需要打印传真或电子邮件即可将它们输入系统。

优势

虽然对于许多企业而言, 纸质文件依然是基本媒介, 然而当今的企业环境发生了重大转变, 已经逐步取消了纸质文档的实物运送, 纳入了大量电子邮件、附件和电子数据。事实上, 这就是当今文档管理的“真实现状”。必须同时管理纸质文档与数字文档。第五阶段的通用捕获门户旨在应对所有这些可能发生的状况。因此, 能够有效管理和利用所有文档的企业必将获得极具竞争力的性能优势。

局限性

虽然目前依然需要一定数量的人员扫描所传入的纸质文档并验证数据质量异常, 但推进第五阶段方案的真正挑战在于战略设计工作, 需要充分利用各项技术并落实战略方针。如果不采用面向流程的缜密文档策略(而非纯技术心态), 企业很可能会偏离正轨。与技术供应商开展合作同样非常重要, 这样可以为您提供各种解决方案, 从而使您以此为起点实现这一阶段的目标。通常情况下, 解决方案均针对一两个阶段量身设计, 因而如未进行重要的系统升级和返工, 很可能会限制您沿着这一连续流逐步发展的能力。寻找提供模块化灵活解决方案的先进合作伙伴, 这将有助于您伴随自身的发展逐步进入。

更多研究:

Datacap新增了大量高端创新功能, 以便创建通用捕获门户, 使其能够自动开展总体的文档策略。

下面是一些当前可用来将高级自动化功能应用到捕获流程中的方法:

- 以处理扫描文档的方式处理[电子邮件和附件](#)(PDF、Word、Excel和压缩文件)
- 输入业界领军企业Rightfax产品传入的传真图像和数据
- 通过内容识别文档[自动进行文档分类](#)。有史以来首次不需要将文档进行预先分类, 直接扫描即可

在线研讨会: AIIM文档分类特别报告后续阶段。

捕获的价值

对于绝大部分企业而言,为了获得竞争优势、降低运营成本以及支持实现业务成功,利用文档和数据捕获概念的文档战略仍然是一个有待开发的商机。过去,此方法采用的许多工具并不符合真实世界的限制,尤其是在企业试图提升业务阶段时。绝大部分企业根本没有能力负担部署有效策略的成本,或者无法处理相关复杂问题。但易于使用的全新经济型扫描和数据保护技术却呈现了一个崭新的商机。

运用此白皮书指导您进行捕获技术评估,并了解这些功能如何发挥文档战略成效。与潜在合作伙伴举行会晤,寻求灵活的模块化解决方案,这样您将能够伴随此连续流五个阶段的实施持续发展。最重要的是,请记住文档价值捕获策略并非是最先进的技术进步,而是如何提升运营业绩。

关于作者



Kevin Craine著有《文档策略设计》一书,同时还开办了“文档策略播客”。他曾负责管理医疗保险、航空航天和医药行业文档服务机构。Kevin本人获得了科学和技术管理MBA及通信专业学士学位。Kevin现任Craine Communications Group常务董事。有关更多信息,请访问www.CraineGroup.com。



自1988年以来,Datacap Inc.为全球企业提供了大量领先的文档捕获和表单处理软件解决方案。屡获殊荣的Datacap软件能够行之有效地将纸质文档转换为信息,提高效率和数据准确性,同时降低成本并缩短文档周期时间。Datacap的客户端/服务器、基于规则的捕获工作流程平台可提供高度灵活的文档索引和数据录入自动化解决方案。基于浏览器的版本可通过Internet进行扫描及编制索引。有关更多信息,请访问www.datacap.com。