BUTTERFLY DocumentFlow

WHITE PAPER · 智能文档处理与文档管理

您的文档经过读取、分类，审批后进入档案库。

DocumentFlow 读取 SharePoint 和文件来源中的文档，按类型分类，提取元数据，并找出完全相同的副本。可疑文档进入审核队列；只有符合规则的文档才会进入档案库。

读取来源

SharePoint 和文件来源

分类

类型、文档库和元数据

剔除副本

通过内容哈希识别完全相同的副本

按规则发布

最终决定权在人，决策可追溯

执行摘要

哪些文档进入档案库，由规则决定。

多年来，文档在不同的文件夹和共享区域中不断累积。同一份合同的多个副本、草稿与最终版本并存；要决定哪一份应进入档案库，就必须逐一打开文档。

Butterfly DocumentFlow 将 Butterfly 平台的智能文档处理（IDP）能力以产品形式提供。它读取文档，判断文档是什么，提取所需信息，按类型验证其有效性，并依据机构定义的规则决定是否进入档案库。

在情况不明确时，DocumentFlow 不会自行决定：可疑文档进入审核队列，每项决策的依据都会逐步记录，只有符合规则的文档才会发送到档案库。

本文内容

- 分散的文档来源中的副本、版本与分类问题
- DocumentFlow 的四个步骤：从来源到档案库
- 发布规则：哪些文档进入档案库，哪些不进入
- 决策追溯与人工审批
- 适用于哪些团队，以及一个实践 PoC：分包商入场文档的验证
- 部署、权限、KVKK 及常见问题

结论

只有原件、已签署的最终文档才会进入档案库。每项决策的原因都与文档一同可见。

问题

哪份文档是原件，哪份是副本?

如果无法确定合同的哪个版本有效，在档案库中搜索也无法令人放心。

文档分散在不同团队的文件夹、SharePoint 文档库和共享区域中。同一份文档通过电子邮件流转、被重新保存，草稿与最终版本并存。到了迁移归档时，确定每份文档的类型、元数据和有效版本就成了一项手工、缓慢且容易出错的工作。

副本与版本

同一文档的副本存放在不同位置；无法确定哪一份是原件。

手工分类

要确定文档类型和元数据，必须逐一打开并阅读每份文档。

失控的档案库

草稿或未签署版本一旦混入档案库，人们对档案库的信任就会下降。

缺乏依据的决策

审计时必须能够说明某份文档为何位于该文档库；手工分类无法记录这一点。

答案

不剔除副本就迁移档案，只会把问题带进新的档案库。DocumentFlow 在将文档发送到档案库之前，先读取、分类、查找副本，并依据规则作出决定；不明确的情况交由人来判断。

工作原理

从来源到档案库
只需四步。

DocumentFlow 连接到您的文档来源，并在文档进入档案库之前对其进行一系列检查。

1 **连接来源**

读取 SharePoint 或文件来源中的文档；后续同步时仅处理有变更的文档。

2 **分类**

每份文档按类型和文档库归类；填写按类型定义的元数据字段。

3 **验证与审核**

依据文档类型的规则集进行验证：e-Devlet 二维码与条形码查询、外部系统核查、文档最长有效期限。副本与原件关联；可疑或敏感文档进入审核队列，决策写回 Butterfly。

4 **发布到档案库**

只有符合发布规则的文档（例如已签署的最终版本）才会发送到目标档案库。



文档。文档按类型、文档库和状态区分：可发布、审核中、副本或未签署。

示意界面，示例数据。

发布规则与决策追溯

只有符合规则的文档
才会进入档案库。

每份文档的决策都会逐步记录。审核人员可以看到哪个标识导致了哪个结果，以及该决策的置信度评分，并与文档一同显示。



决策追溯。读取来源、确定类型、提取元数据、副本检查、签名与版本；随后给出决策和置信度评分。

示意界面，示例数据。

进入档案库的文档

- 符合机构所定义发布规则的文档（例如已签署的最终版本）
- 在完全相同的副本中被标记为原件的唯一文档

不进入档案库的文档

- 草稿与未签署版本
- 原件的副本；这些副本与原件关联
- 尚未完成审核的可疑文档

原则

发布规则由机构确定。情况不明确时，最终决定权在人。

功能

文档处理
所需的一切。

DocumentFlow 运行在 Butterfly 平台的企业文档管理之上：基于权限的访问控制，以及与流程集成的智能文档处理。

增量同步

首次扫描后仅重新处理有变更的文档；必要时进行全量扫描。

类型与文档库分类

文档按机构的文档库结构归类；分类可在审核界面中修正。

按类型定义的元数据

为每种文档类型定义各自的字段集；字段按类型以日期、列表或文本形式编辑。

副本检测

通过内容哈希找出完全相同的副本；副本与原件关联。

文档验证

通过 e-Devlet 二维码与条形码查询、外部系统核查以及按类型定义的规则集自动验证文档；敏感类型交由人工审批。

基于权限的访问

文档访问依据 Butterfly 的权限模型授予；文档与流程集成运行。

连接您自己的流程

DocumentFlow 并非独立的归档工具。借助 Butterfly 的 App Builder 和 No-Code API Builder，文档流与机构自己的流程相连：一份经审批的合同可以触发一个流程步骤，一份已分类的文档可以附加到相关记录。

适用团队

文档量大、档案至关重要的团队。

DocumentFlow 专为文档量大、档案准确性在审计中受到检验的团队而设计。

法务与合同

法律顾问、合同管理

原因

如果无法确定合同的哪个版本有效，在档案库中搜索也无法令人放心。

作用

已签署的最终合同以正确的元数据进入档案库，草稿被排除在外。

质量与文档控制

质量保证、文档控制

原因

审计时必须能够说明某份文档为何位于该文档库。

作用

受控文档按类型分类，每项决策及其依据均可追溯。

档案与记录管理

企业档案、信息管理

原因

不剔除副本就迁移档案，只会把问题带进新的档案库。

作用

将分散文件夹中的文档剔除副本后，迁移到一个井然有序的统一档案库。

适用性的四个问题

主题	问题
文档数量	您的文档是否分散在 SharePoint 和文件共享中?
副本	同一文档的副本或不同版本是否并存?
档案控制	草稿或未签署版本混入档案库是否是个问题?
审计	您是否需要说明某份文档为何存放在该文档库或档案库中?

实践案例

分包商入场文档
每五分钟验证一次。

在一家国际电缆制造商的土耳其工厂，为自动验证分包商和访客入场文档搭建了一个 PoC，并在生产环境中使用真实文档进行了测试。

5 分钟 **待处理文档每五分钟获取一次，按类型经过机器检查，结果写入 Butterfly。**

1 · 规则

13 种文档类型

HSE 团队的规则被定义为文档类型：验证方式、文档最长有效期限、最短培训时长、是否必须审核。

2 · 验证

e-Devlet 与外部系统

带 e-Devlet 条形码的文档通过读取二维码由官方服务核验，车辆检验等文档通过报告地址核查。

3 · 审核

敏感类型交由人工

无犯罪记录证明等类型即使通过验证，也按规则进入 DocumentFlow 审核队列。

4 · 决策

写回记录

批准或拒绝；连同决策人和备注自动写入 Butterfly 中的文档记录。

结论

- SGK 入职申报表在登记后，无需任何人工干预，通过读取二维码经 e-Devlet 完成验证。
- 四项端到端实时测试全部成功；错误的访问密钥和不带二维码的文档被拒绝，无法连接 e-Devlet 时查询自动重试。
- 后续步骤：为不带二维码的文档提供文字识别（OCR），以及按 KVKK 要求对身份证号进行遮蔽。

在平台中的定位

DocumentFlow 将 Butterfly 的 BOAT 能力领域之一，即文档处理（IDP），以开箱即用的产品形式提供。由于同一平台还运行流程编排、集成、自动化和 AI 智能体，文档不再只是存放在档案库中的文件，而是连同其决策和后续步骤一起成为流程的一部分。

信任

企业文档
需要企业级信任。

DocumentFlow 运行在 Butterfly 平台上。文档、决策和审批均处于平台的权限模型和审计追踪之下。

部署选项

本地部署、私有云、Butterfly 云或混合部署。文档在何处处理由机构决定。

基于角色的访问

谁可以访问文档和审核界面，由平台的权限模型定义。

审计追踪

分类、批准、拒绝和发布决策连同用户和时间一并记录。

人工审批

可疑文档进入审核队列；发布规则由机构确定。

个人数据与 KVKK

合同和往来通信可能包含个人数据。在本地部署中，文档在机构自己的基础设施上处理和存储；谁可以访问以及文档保存多长时间，由机构自己的 KVKK 政策管理。Butterfly 平台提供身份、权限模型和审计追踪，以确保该政策得以执行。

证书

证书	范围
ISO/IEC 27001	信息安全管理体系
ISO 9001	质量管理体系
国产产品证书	业务流程管理软件（Butterfly），100% 国产化率，A 级，高技术
技术产品证书	Butterfly DXP，土耳其共和国工业和技术部

全部证书及验证链接：butterfly.dev/zh/kurumsal/sertifikalar

常见问题

关于 DocumentFlow 的简要解答。

从哪些来源读取文档?

从 SharePoint 和文件来源读取文档。首次扫描后，仅重新处理有变更的文档。

如何确定文档类型?

每份文档按类型和文档库进行分类。情况不明确时进入审核队列，由用户确认或更改分类。

如何查找副本?

比较文档的内容哈希值。完全相同的文档归为一组，并为该组标记唯一的原件。

哪些文档会进入档案库?

仅限符合机构所定义发布规则的文档，例如已签署的最终版本。草稿、未签署版本和副本不会发送到档案库。

我们能否看到某项决策的原因?

可以。每份文档的决策都会逐步记录；哪个标识导致了哪个结果以及决策的置信度评分，均与文档一起显示。

可以本地部署吗?

可以。DocumentFlow 运行在 Butterfly 平台上，支持本地部署；访问权限依据 Butterfly 的权限模型授予。

让我们以一个文档库为例，一起审视您的文档流。

选择一个文档库；我们将在您自己的文档上展示分类、副本检测、决策追溯和发布规则。

审视您的文档流 · butterfly.dev