BUTTERFLY DocumentFlow

WHITE PAPER · INTELLIGENT DOCUMENT PROCESSING AND DOCUMENT MANAGEMENT

Your documents are read, classified and archived after approval.

DocumentFlow reads documents in SharePoint and file sources, classifies them by type, extracts their metadata and finds exact duplicates. Suspicious documents go to a review queue; only documents that meet the rule go to the archive.

Reads the source
SharePoint and file sources

Classifies
Type, library and metadata

Removes duplicates
Exact duplicates by content hash

Publishes by rule
People have the final say, decisions are traced

EXECUTIVE SUMMARY

Let a rule decide which document goes into the archive.

Over the years documents pile up in different folders and shared drives. Copies of the same contract, drafts and final versions sit side by side; deciding which one belongs in the archive means opening documents one by one.

Butterfly DocumentFlow offers the intelligent document processing (IDP) capability of the Butterfly platform as a product. It reads a document, determines what it is, extracts the information needed, validates it according to its type and decides whether it goes to the archive based on the rule the organization defines.

When the situation is unclear, DocumentFlow does not decide on its own: suspicious documents go to a review queue, the reasoning behind every decision is recorded step by step and only documents that meet the rule are sent to the archive.

IN THIS DOCUMENT

- The duplicate, version and classification problem across scattered document sources
- DocumentFlow's four steps: from source to archive
- The publishing rule: which documents go to the archive and which do not
- Decision trail and human approval
- Which teams it suits and a PoC from the field: validating subcontractor entry documents
- Deployment, permissions, KVKK and frequently asked questions

OUTCOME

Only the original, signed and final document enters the archive. The reason for every decision is visible together with the document.

THE PROBLEM

Which document is the original, and which is a copy?

If it is not clear which version of a contract is valid, searching the archive does not inspire confidence.

Documents are spread across different teams' folders, SharePoint libraries and shared drives. The same document circulates by e-mail, is saved again and the draft stays next to the final version. When it is time to move to the archive, determining each document's type, metadata and valid version becomes a slow, manual and error-prone task.

Duplicates and versions

Copies of the same document sit in different places; it is not clear which is the original.

Manual classification

Each document has to be opened and read to determine its type and metadata.

Uncontrolled archive

If draft or unsigned versions slip into the archive, trust in the archive declines.

Decisions without reasoning

In an audit you need to show why a document is in a given library; manual classification does not record this.

THE ANSWER

An archive migration done without removing duplicates moves the problem into the new archive. Before sending a document to the archive, DocumentFlow reads it, classifies it, finds its duplicates and decides by rule; it asks a person when the case is unclear.

HOW IT WORKS

From source to archive in four steps.

DocumentFlow connects to your document source and puts documents through a series of checks before they go to the archive.

1 **Connect to the source**

Documents in your SharePoint or file source are read; in later syncs only changed documents are processed.

2 **Classify**

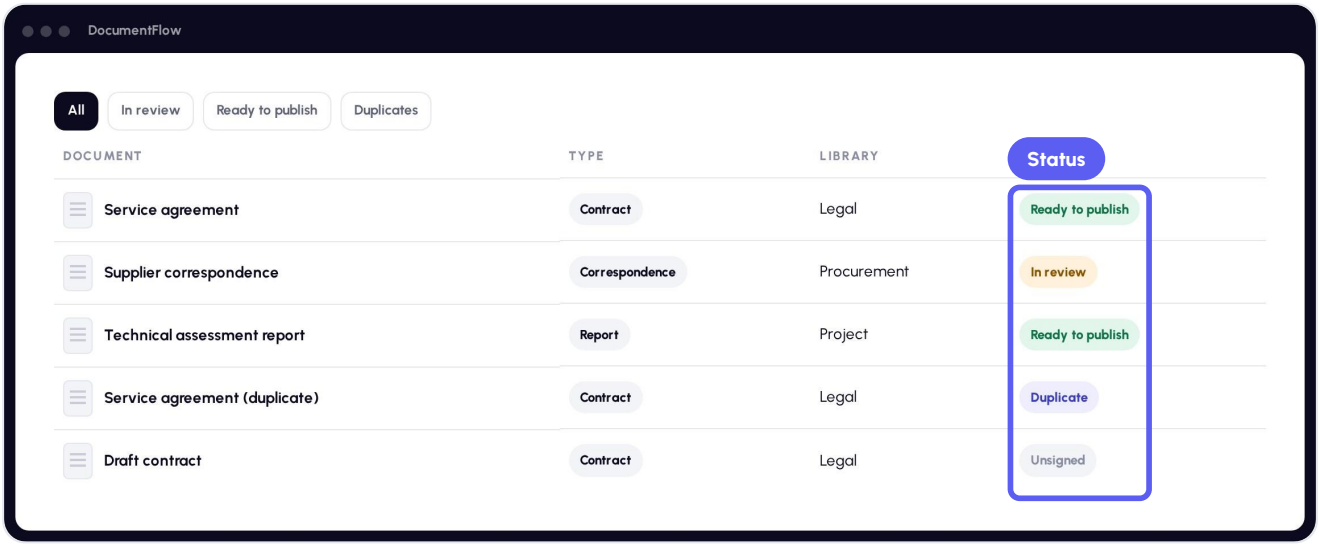
Each document is sorted by type and library; the metadata fields defined for its type are filled in.

3 **Validate and review**

The document is validated against its type's rule set: e-Devlet QR and barcode lookup, external system check, maximum document age. Duplicates are linked to the original; suspicious or sensitive documents go to the review queue and the decision is written back to Butterfly.

4 **Publish to the archive**

Only documents that meet the publishing rule (for example signed and final) are sent to the target archive.



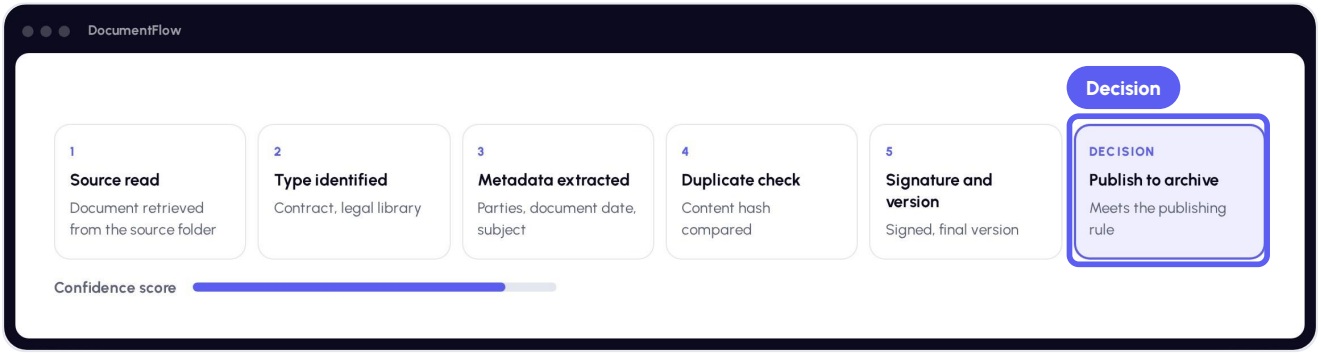
Documents. Documents are sorted by type, library and status: ready to publish, in review, duplicate or unsigned.

Illustrative screen, sample data.

PUBLISHING RULE AND DECISION TRAIL

Only documents that meet the rule go to the archive.

The decision for each document is recorded step by step. The reviewer sees which signal led to which outcome and the decision's confidence score together with the document.



Decision trail. Source read, type determined, metadata extracted, duplicate check, signature and version; then the decision and confidence score.

Illustrative screen, sample data.

Documents that go to the archive

- Those that meet the publishing rule defined by the organization (for example the signed and final version)
- The single document marked as the original among exact duplicates

Documents that do not go to the archive

- Drafts and unsigned versions
- Duplicates of the original document; these are linked to the original
- Suspicious documents whose review is not complete

PRINCIPLE

The organization sets the publishing rule. When the case is unclear, people have the final say.

CAPABILITIES

Everything needed for document processing.

DocumentFlow runs on the enterprise document management of the Butterfly platform: permission-based access control and intelligent document processing integrated with processes.

Incremental sync

After the first scan only changed documents are reprocessed; a full scan is run when needed.

Type and library classification

Documents are sorted according to the organization's library structure; the classification can be corrected from the review screen.

Metadata by type

A field set is defined for each document type; fields are edited as dates, lists or text according to type.

Duplicate detection

Exact duplicates are found by content hash; duplicates are linked to the original document.

Document validation

Documents are validated automatically with e-Devlet QR and barcode lookup, external system checks and type-based rule sets; sensitive types go to human approval.

Permission-based access

Access to documents is granted according to Butterfly's permission model; documents work integrated with processes.

It connects to your own processes

DocumentFlow is not a standalone archiving tool. The document flow is connected to the organization's own processes with Butterfly's App Builder and No-Code API Builder: an approved contract can trigger a process step, and a classified document can be attached to the related record.

WHICH TEAMS

Teams with many documents and a critical archive.

DocumentFlow was designed for teams with high document volumes whose archive accuracy is questioned in audits.

Legal and contracts

Legal counsel, contract management

WHY

If it is not clear which version of a contract is valid, searching the archive does not inspire confidence.

WHAT IT DOES

Signed and final contracts enter the archive with the right metadata, and drafts stay out.

Quality and document control

Quality assurance, document control

WHY

In an audit you need to show why a document is in a given library.

WHAT IT DOES

Controlled documents are classified by type and every decision can be traced with its reasoning.

Archive and records management

Corporate archive, information management

WHY

An archive migration done without removing duplicates moves the problem into the new archive.

WHAT IT DOES

Documents in scattered folders are cleared of duplicates and moved into a single, well-organized archive.

Four questions to check the fit

TOPIC	QUESTION
Document volume	Are your documents spread across SharePoint and file shares?
Duplicates	Do copies or different versions of the same document sit side by side?
Archive control	Is it a problem when draft or unsigned versions slip into the archive?
Audit	Do you need to show why a document is in a given library or archive?

FROM THE FIELD

At subcontractor entry, documents are validated every five minutes.

At the Türkiye plant of an international cable manufacturer, a PoC was set up to automatically validate subcontractor and visitor entry documents, and it was tested in the live environment with real documents.

5 min

Pending documents are collected every five minutes, put through a machine check according to their type and the result is written to Butterfly.

1 · RULE

13 document types

The HSE team's rules were defined as document types: validation method, maximum document age, minimum training duration, mandatory review.

2 · VALIDATION

e-Devlet and external systems

Documents with an e-Devlet barcode were checked against the official service by reading the QR code, and documents such as vehicle inspections were checked at the report address.

3 · REVIEW

Sensitive types go to a person

Types such as criminal records went to the DocumentFlow review queue by rule, even when validation passed.

4 · DECISION

Written back to the record

Approval or rejection, together with the decision maker and a note, was written automatically to the document record in Butterfly.

OUTCOME

- The SGK employment notification was validated through e-Devlet by reading its QR code, with no human intervention after it was registered.
- All four live end-to-end tests succeeded; a wrong access key and a document without a QR code were rejected, and when e-Devlet could not be reached the query was retried automatically.
- Next steps: text recognition (OCR) for documents without a QR code and masking of national ID numbers as required by KVKK.

Its place in the platform

DocumentFlow offers document processing (IDP), one of Butterfly's BOAT capability areas, as a ready-to-use product. Because the same platform also runs process orchestration, integration, automation and AI agents, a document stops being a file that sits in the archive and becomes part of the process together with its decision and next step.

TRUST

Enterprise documents require enterprise trust.

DocumentFlow runs on the Butterfly platform. Documents, decisions and approvals are kept under the platform's permission model and audit trail.

Deployment options

On-premise, private cloud, Butterfly cloud or hybrid. The organization decides where documents are processed.

Role-based access

Who can access documents and the review screen is defined with the platform's permission model.

Audit trail

Classification, approval, rejection and publishing decisions are recorded together with the user and the time.

Human approval

Suspicious documents go to the review queue; the organization sets the publishing rule.

Personal data and KVKK

Contracts and correspondence may contain personal data. In an on-premise deployment, documents are processed and stored on the organization's own infrastructure; who has access and how long documents are kept are governed by the organization's own KVKK policy. The Butterfly platform provides identity, the permission model and the audit trail so that this policy can be enforced.

Certificates

CERTIFICATE	SCOPE
ISO/IEC 27001	Information security management system
ISO 9001	Quality management system
Domestic Product Certificate (Yerli Malı Belgesi)	Business process management software (Butterfly), 100% domestic content, class A, high technology
Technological Product Certificate	Butterfly DXP, Ministry of Industry and Technology of the Republic of Türkiye

All certificates and verification links: butterfly.dev/en/kurumsal/sertifikalar

FREQUENTLY ASKED QUESTIONS

Short answers about DocumentFlow.

Which sources does it read documents from?

It reads documents from SharePoint and file sources. After the first scan only changed documents are reprocessed.

How are document types determined?

Each document is classified by type and library. Unclear cases go to the review queue, and the user approves or changes the classification.

How are duplicates found?

The content hash of each document is compared. Exactly identical documents are grouped and one original copy is marked in each group.

Which documents go to the archive?

Only documents that meet the publishing rule defined by the organization, for example signed and final versions. Drafts, unsigned versions and duplicates are not sent to the archive.

Can we see the reason for a decision?

Yes. The decision for each document is recorded step by step; which signal led to which outcome and the decision's confidence score are visible together with the document.

Can it be deployed on-premise?

Yes. DocumentFlow runs on the Butterfly platform and can be deployed on-premise; access is granted according to Butterfly's permission model.

Let's review your document flow through one library together.

Choose a document library and let's look at classification, duplicate detection, the decision trail and the publishing rule on your own documents.

Let's review your document flow · butterfly.dev