Transforming the way you do business.

SWISS POST SOLUTIONS

CDIPP Technical Architecture and Functional Description

<<Restricted distribution only>>

05/09/2017

SPS

a Swiss Post company

Table of Contents

Document Control 2

Document purpose 2

Document location 2

Version history 2

Author 2

Approvals 2

Distribution 2

Reference documents 3

Glossary 3

Target Audience and restrictions 4

CDIPP: One Platform for SPS 4

CDIPP Building Blocks 6

The Core Digitizing Suite 6

The Processing Activity Coordination 6

The Supervision and Control Unit 7

CDIPP users and systems interaction in mailrooms 8

CDIPP Architecture 9

Cordis 10

CDIPP Functional capabilities description 13

Exceptions Handling 16

Graphical Exception handling 18

Workflows Creation 19

Quality Control 21

Operation Monitoring and Reporting 22

CDIPP Deployment 23

Annexes 24

Lookup provisioning options 24

CDIPP Monitoring 27

Disclaimer 27

Document Control

Document purpose

The purpose of the document is to provide with the technical architecture and functional description of the Central Document Input Processing Platform used in SPS DIP to support SPS Operation teams in fulfilling the Customer Business Process outsourcing projects.

Document location

The document will be stored on the CDIPP Confluence environment.

Version history

Version	Date of Change	Change Summary
Draft 0.1	24/04/2017	Document creation for customer communication
Draft 0.2	29/06/2017	Creation of the Exception Handling chapter and correction according to feedback from A. Vogel.
Draft 0.3	14/07/2017	Updates, chapters updates
Final 1.0	11/08/2017	Final version 1.0
Final 1.1	05/09/2017	Various mispelling corrections

Author

Name	Title & Organisation
E. Denis	Platform Owner

Approvals

Name	Title & Organisation	Approval Date
Peter Tognazza	CDIPP Technical Project Manager	
Dalibor Peric	CDIPP Technical Architect	

Distribution

Name	Title & Organisation	Distribution Date
Arno Vogel	Global IT Project Manager	
Lukas Hebeisen	Head of Technology and Digital transformation	

Reference documents

Document Name	Link	Document Version
CDIPP Functional Releases Specification Confidential	CDIPP Confluence	Current Working Space
CDIPP Deployment and Servers Requirements Confidential	CDIPP Confluence	Current Working Space

Glossary

Term	Description
CDIPP	Central Document Input Processing Platform. CDIPP is the main Document processing platform used in SPS
CORDIS	Core Digitizing Suite represents the main component involved in the digitizing process according to business logic.
BPMS	Business Process Management System is the main component of the PAC. It allows to setup a workflow in line with the needed business processes per documents or group of documents or Items.
DPF	Document Processing Framework is a CDIPP module. It manages the input/output from scanners and towards Customers. It prepares packages according to agreed format.
GRAFEX	Graphical Exception handler is a CDIPP module. It is providing a set of tools to create child documents from an original one to cope with graphical business exception needs.
ICAP	Information Capturing is a CDIPP module. It acts at manual indexing, look-up and validation rules level.
ICLASS	Information Classification is a CDIPP module. It provides though a web interface the capability to index manually a document through textual fields and validation set of rules.
DATEX	It handles the automated data extraction (OCR) and the automatic classification based on a set of configurable rules.
PAC	Processing Activity Coordinator represents the set of services that coordinate the modules according to the set of business rules per documents or group of documents/Items to be processed
SCU	Supervision and Control Unit represents the IT tool box in order to install, configure, troubleshoot and manage the users credential and access right.

Target Audience and restrictions

The current document intents to provide the high level information about CDIPP. The readers will find information about the architecture and its building blocks as well as the main features that constitutes the digitizing documents process in SPS Document Input Processing organisation. This document can also be shared with customers upon formal agreement between parties and for information. This document is subject to regular revisions and updates as CDIPP is a platform that evolve according to a roadmap. Therefore, it cannot be considered as contractually binding unless explicitly referred at in words in a signed contract.

CDIPP: One Platform for SPS

The Central Document Input Processing Platform CDIPP is the technology component used by the Swiss Post Solution Document Input Processing division. Fully integrated into a mailroom, the platform transforms physical and electronic customers items coming from multiple distribution channels (scanner, emails, Faxes…) into digitalized information that are transmitted back to the customer in various format at the end of the digitizing process. The platform provides support in timely delivering digitalized items according to customer business rules and contractually agreed SLA's.

CDIPP is also supporting different Level of industry driven data sensitivity (country data security regulations) and adapts to the various customer data Isolation constraints. Thanks to its multi-instances deployment capability, it allows answering both specific stand-alone as well as mutualized solution as a service's needs.

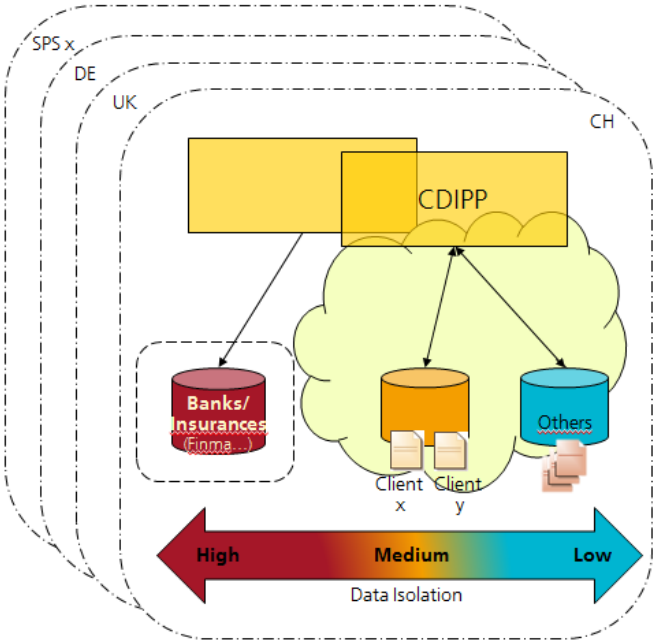


Figure 1 CDIPP multiple deployment schema and data security requirements per industry

The platform conception optimizes SPS DIP Operation and harmonizes practices across the subsidiaries worldwide, while improving customers service satisfaction. Its modular design lowers the dependency with sub-applications or technology, and adds agility in providing service improvements. More than a simple technical component, CDIPP accompanies SPS and Customers alike, to enable further the tomorrow's business process outsourcing needs whilst improving user's experience.

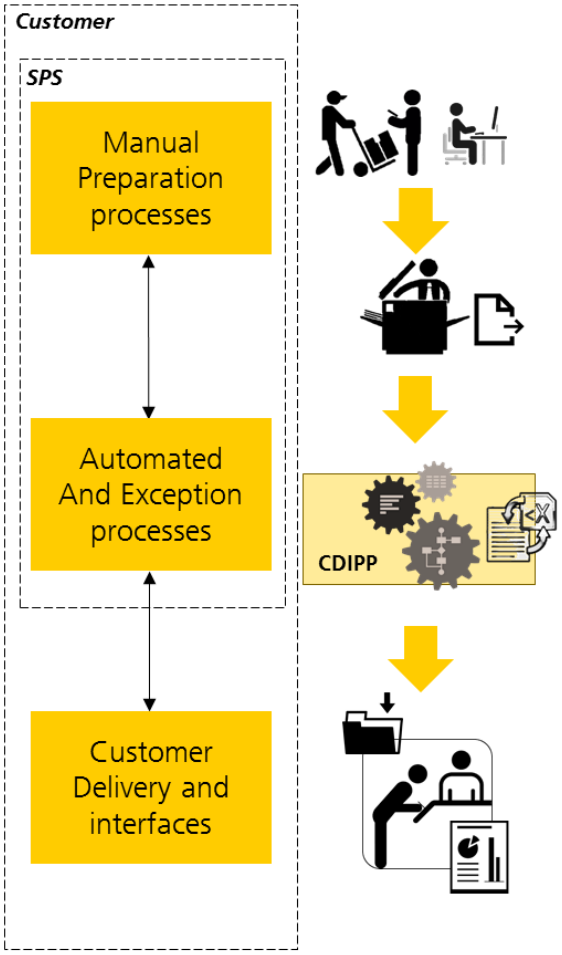


Figure 2 SPS Customer Business Processes end to end service

CDIPP Building Blocks

CDIPP is the software platform that supports SPS service for the data digitizing process (**early or late scanning**) and the manual exception handling. The functional architecture of the platform can be organized into 3 structural blocks (in red on the figure below):

1. The core digitizing suite (CorDiS)
2. The processing activities coordination (PAC)
3. The supervision and Control unit (SCU)

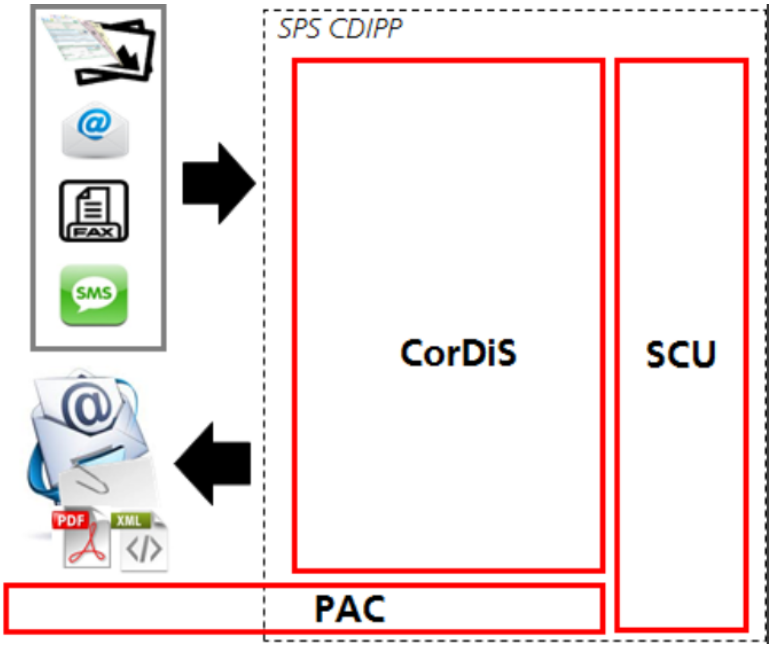


Figure 3 CDIPP Building Blocks

The Core Digitizing Suite

CORDIS is the heart of the digitizing process. It is responsible to extract, validate and classify the scanned documents or items. It manages the scan jobs input and customer output including the formatting of the data. It allows the creation of the business rules to extract, validate, classify as well as to handle the exceptions. It hosts the web-based manual validation and classification in case of exception. Cordis eases and speeds the data verification process up thanks to the look-up service and stores the customer extracted data in its central master DB. In case of interaction with external third parties, DPF manages the interaction and the format of exchange with the specified external interface as well as delivery the original documents and its data to the end customer

The Processing Activity Coordination

The PAC is the organizational and workflows' decision centre. The PAC is based on a Business Process Management System application , that provides the Business Workflow

creation Service, the end to end coordination functionality in order to orchestrate in a timely manner the required activities (Human, machines, software components) defining a done business case for one or more customer Items types. The BPMS integrates and interacts with external services within a done workflow.

The PAC knows and provides the necessary information to locate at any time a specific item (documents, emails content…) or a specific batch of Items.

The PAC hosts the components that support the creation of business off-line reports, reconciliation as well as real time SLAs monitoring used by the operation team in mailrooms.

The Supervision and Control Unit

The SPU is the monitoring and the IT tool-box of the entire platform and its components. It provides monitoring functions of the end-to-end delivery chain and specifically the different software modules belonging to the platform. It enables SPS IT operation teams to check and trouble shoot the end to end digitizing process through:

- A Centralized logging, platform monitoring and alarming
- Automated alarms sending to SPS Trouble Ticketing System
- Automated tests to detect any regression from delivery of new releases

It also provides a central Identity and Access Management module that rules strictly the CDIPP users access, roles and permission as well as their credentials life-cycle. It provides all the access logs needed for an audit trail: the detailed transactions relating to any access activity into the platform are recorded.

CDIPP users and systems interaction in mailrooms

Several human users are interacting with the platform:

- SPS Mailroom Supervisors
- SPS Mailroom Customer Items preparation operators
- SPS Manual indexing operators
- SPS Manual classification operators
- SPS Manual Exceptions handling Operators
- SPS Operation Managers
- SPS IT Engineers and Managers

The customers items are handled by SPS Staff in mailrooms at various points in the end to end delivery chain. All the customers Items are tracked physically (during pre-scanning and post-scanning or archiving phases) and during the digitizing phase on CDIPP.

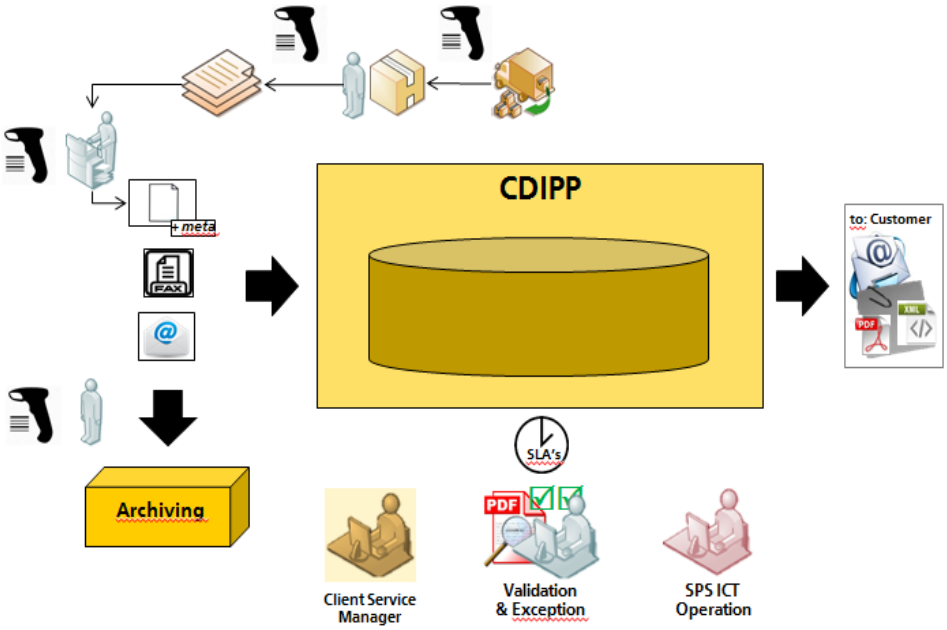
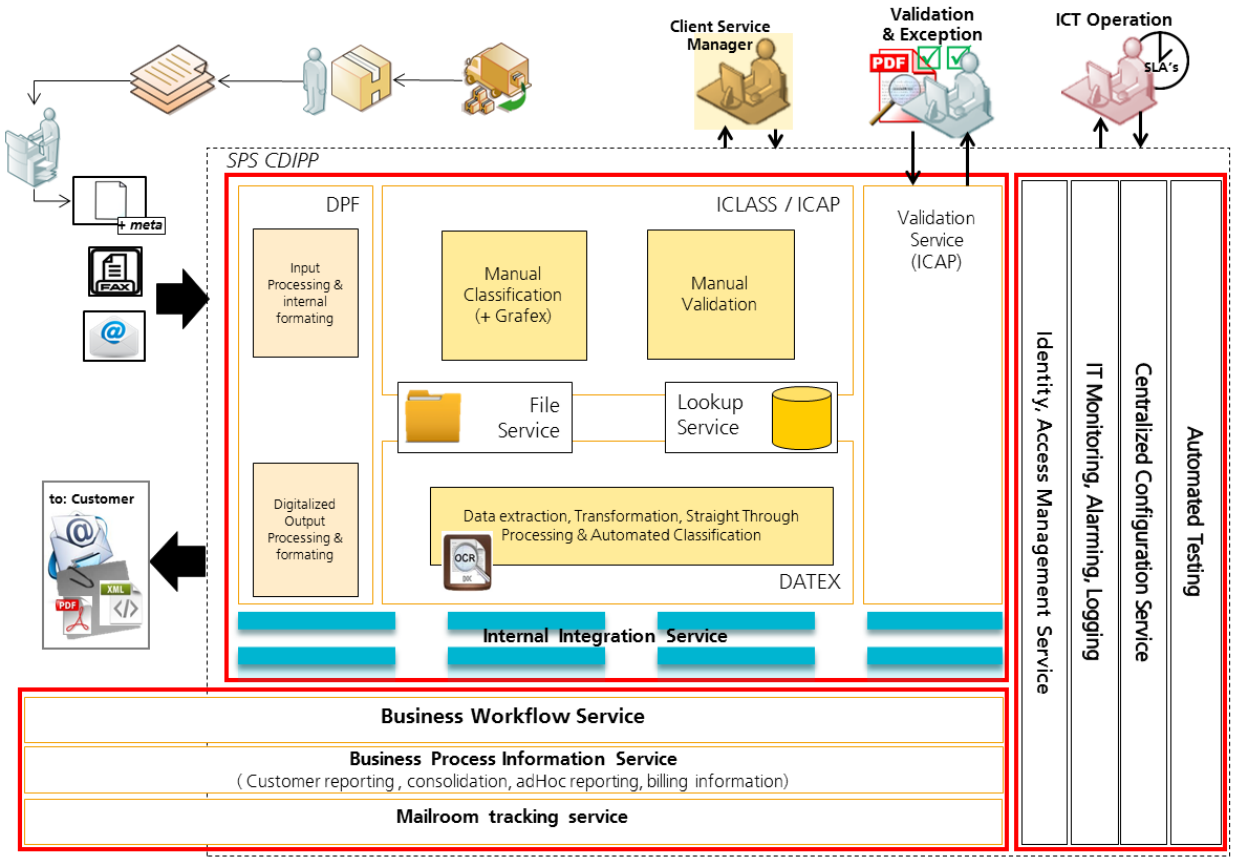


Figure 4 Users end to end interaction with CDIPP

CDIPP is scanners agnostic. Any kind of scanners can be used in parallel. The format of data to be exchanged (metadata) can be XML or equivalent like an OXI files in the case of an Opex machine. The platform can generate various picture formats like Jpeg, Tif or PDF as an example.

CDIPP Architecture

The platform general architecture is shown on the picture below:



CDIPP follows a Service Oriented Architecture (SOA) with an Enterprise Message Bus (ESB) called Internal Integration Service. This service allows to integrate new technology and thus enables CDIPP to on-board the future components (Artificial Intelligence, Business Intelligence, Robotics…) that will improve SPS services and our customers satisfaction. CDIPP is an assembly of reliable and robust known components configured to specifically match SPS Document Input Processing core business. CDIPP is made of SPS best practice components running in operation for various customers around the world.

The platform architecture is thought to ensure scalability in term of deployment as it can be installed in cluster in a private V-LAN or into a cloud, in stand-alone for a sole customer or mutualized (multi-tenancy) to host several customers business processes. As the main technology used by SPS, the platform allows as well to scale up seamlessly the number of users (manual operators).

The robust distributed architecture, the multi-threading capability, the scalability of Operators, the precise control of business processes and the high level of data security, make CDIPP a unique asset to the benefit of SPS and Customer alike.

The architecture is organized in group of modules with a unique purpose:

Cordis

Cordis represents the heart of the digitizing process and stands for Core Digitizing Suite.

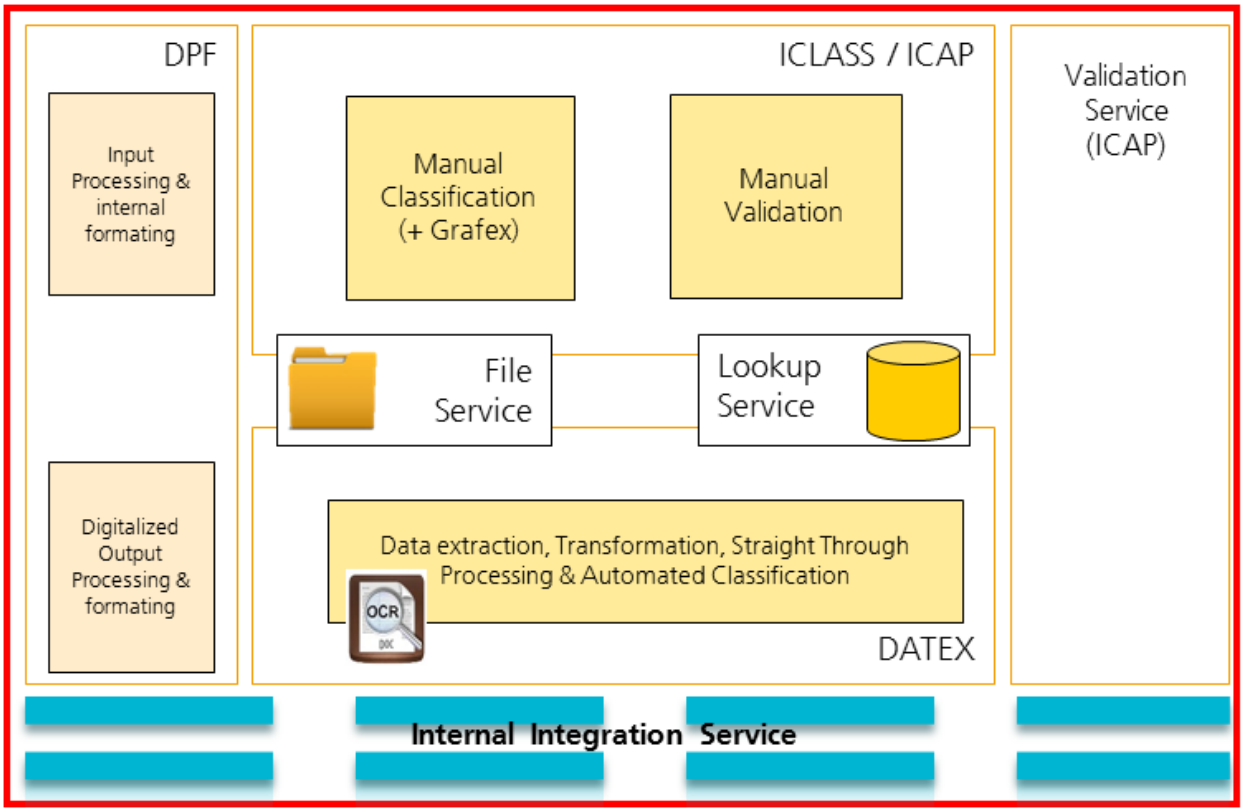


Figure 5 Core Digitizing Suite: heart of the digitizing process

The Document Processing Framework (DPF) handles the I/O processing channels (Scanners output, emails, Faxes …) and formats the data for the communication intra-modules and towards external third parties and customers.

The Look-up Service provides the indexing auto-completion and searches capabilities used during the Manual indexing. It assists the operators in providing an automatic completion and search function that looks (up!) in database for existing data matching information to speed up the fields fulfilling activity. The lookup service is shared between the Validation Service, the Manual Consolidation service ICAP and the raw data capturing DATEX.

Customer lookup information can be migrated into the CDIPP Lookup DB following different methods: Off line (full, incremental or differential), on-line or Cached (see annexe "Lookup provisioning options".

The Supervision and Control Unit (SCU)

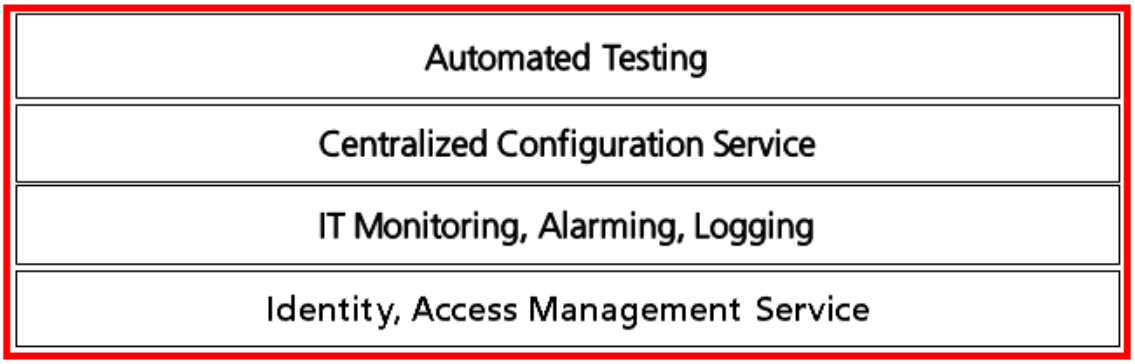


Figure 6 Supervision and Control Unit main components

IT Monitoring, Alarming, Loggin module provides the necessary information to technically track the platform components health and to trouble shoot the platform in case of defect/failure or simply during test phases.

Monitoring module provides the capabilities to check not only the health of the platform but also to track end-to-end the documents/batches across the processes and their status. (Annexe: CDIPP Monitoring)

Identity, Access Management Service, this module allows the management of the CDIPP users: Roles, access rights, login/password… The Identity Server is installed in standalone and log all the access event across the platform. The provisioning of the users is managed centrally by each SPS ICT operation team.

The Processing Activity Coordinator (PAC)

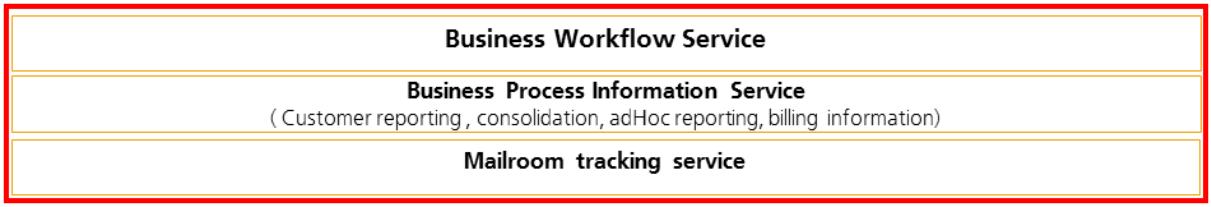


Figure 7 Processing Activity Coordinator Services

The Business Workflow Service is coordinating the end-to-end "extended" activities (Human, machines, software components) according to each configured business cases. The communication model between the module and the BWS is Master/Slaves, where the BWS is the master. According to the information received by the BWS from each modules, decisions are taken to trigger the right workflow. A workflow constitutes a sequence of actions that correspond to the treatment (CDIPP workflow process) of a

certain document or group of documents according to specific internal and customer attributes. These document' attributes are varying from case to case but are for example the document type, address of origin etc.

As the main orchestrator, the information coming though the BPS regarding the triggered processes and underlying documents or group of documents are stored in database. The platform knows at any time the status of a process, a document etc. These tracking information are vital for the monitoring of the end to end delivery to the customer.

The Business Process Information Service relies on these information to provide to the SPS managers and supervisor, with the capability to monitor and report activity though a web-based user graphical interface. The BPIS provides real time information needed to monitor deliveries progress according to SLAs, take operation decisions based on the information displayed, like for example increase human capacity to speed up a done business process.

Billing information, customer reporting are providing thanks to the Business Process Information Service.

CDIPP is also tracking the activities in the "physical world" thanks to the Mailroom Tracking Service connected to the BWS. All the information related to the physical documents like archive number, storage location etc. are also stored in database for reconciliation, retrieval or various exceptions like in the case of a re-scan request or retrieval and re-send etc. CDIPP keeps track of the status of the items available in the physical and the digital world.

CDIPP Functional capabilities description

CDIPP is the document input processing platform used globally in SPS to serve our customers. CDIPP follows a roadmap made of collected requirements from operations and customers worldwide. After analysis, these collected requirements follow an internal selection process in order to short list the most demanded requirements. Selected requirements are then implemented and available at a frequency of two releases per year and made available through the SPS release management process.

The main functions available are the following:

Input/output Management

- ✓ Input: PDF, TIF, JPG scanned pictures with XML based metadata (xml, oxi)
- ✓ Multiple Physical document format management (A4, A...) depending on the involved scanners and they capacities
- ✓ Various output format: PDF, TIF, JPG and customized XML
- ✓ Email with or without attachments (includes Fax)

OCR capabilities for :

- ✓ Structured documents text fields extraction
- ✓ Signature extraction
- ✓ Auto-classification (forms recognition, locators …)

Customer Business Process management

- ✓ Business Processes configuration with BPMN 2.0
- ✓ 3rd party technology integration (various off the shelf connectors available)

Customer Business Rules Management:

- ✓ Extraction rules management
- ✓ Validation rules management
- ✓ Classification Rules management

Exception Handling Management

- ✓ Manual Validation
- ✓ Manual Classification
- ✓ Documents Deletion and re-scan order management
- ✓ Option: Digital Mail tray – Customer Exception Handling interaction

Quality Insurance

- ✓ Quality Centre

Business Operation Monitoring and reporting

- ✓ Documents tracking with audit trail
- ✓ SLAs monitoring (Turnaround Time)
- ✓ OCR recognition level reporting
- ✓ Flexible reporting
- ✓ Billing

Documents on-boarding process

- ✓ Electronic masks creation
- ✓ Business rules configuration
- ✓ Processes configuration
- ✓ Modules configuration

Look-up service

- ✓ Customer data import
- ✓ Interactive auto-completion for manual validation
- ✓ 3 options: Off-line, On-line, Cached

Security: Identity and Access Management service

- ✓ 2 (multi-)factors authentication
- ✓ User profiles management
- ✓ Group, Roles management
- ✓ Access right management
- ✓ Single Sign-on (not on external reporting interface)
- ✓ Active directory/LDAP integration
- ✓ User provisioning
- ✓ Identity life cycle management (Authentication, Authorisation, password management, provisioning, de-provisioning

Security: others

- ✓ OWASP 10 most critical Web Application risk security proof
- ✓ Data in transits security though VPN and or HTTPS
- ✓ For Web based GUI accessible outside the network security zone, reverse proxy are used.

- ✓ All the access attempts and success are logged in the platform and accessible with the right credential.

Documents Tracking Capability:

- ✓ Physical Archiving tracking
- ✓ End to End document tracking

ICT Features:

- ✓ Platform Centralized configuration service:
- ✓ The centralized configuration allows an IT Engineer to set-up the platform from a web based interface·
- ✓ CDIPP runs with Microsoft SQL and Elasticsearch for the lookup
- ✓ CDIPP runs on Windows Server 2016
- ✓ Internet Explorer 11 or greater
- ✓ CDIPP can be deployed in standalone with or without High Availability
- ✓ Standard deployment: a production and a pre-production environment (or Staging)

ICT Platform Monitoring

- ✓ Applications and Middleware level
- ✓ Server, OS and DB level
- ✓ Network and Infrastructure level

This list is not exhaustive and may vary according SPS internal needs and roadmap

Exceptions Handling

In the CDIPP Digitizing steps there are 4 main execution processes:

a. The automatic level (Level 1)

DATEX attempts to classify a document. If the recognition rate is below a configurable value in % (ex: 70%), the BPMS is rooting the items to the next level: The Manual Classification.

If the recognition rate is sufficient, the document can be "straight through processed", i.e. the output package can be formatted.

b. The manual level (Level 2)

During manual indexing or during manual classification, a human called Operator, is manually executing a controlled number of action aiming at classifying a document and or filling up or correcting or adding any missing information from the automatic extraction process.

c. The Graphical manual advanced level (Level 3)

In some business cases, it is necessary to manipulate the document images in order to simplify its processing or to meet a business process requirements. Grafex provides an Operator to use a set of graphical tools to create sub-documents or child-documents. The child-documents are then re-injected into the process for normal treatment.

d. The Customer Exception level (Level 4)

if SPS cannot process according to rules and knowledge available on specific documents or processes, an interaction with the customer himself is necessary.

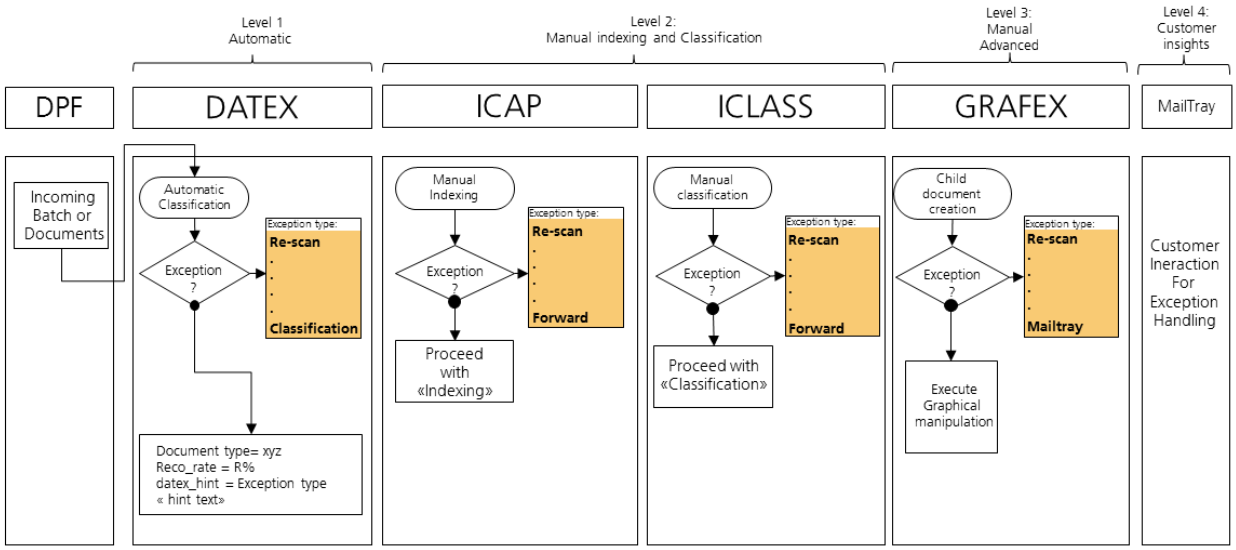


Figure 8 the 4 Internal Execution Processes

This mechanism increases the probability to solve any Document/Item/business process related issue to be handled by SPS before ultimately interacting with the customer himself.

Exception trigger

An exception is triggered when the system required the intervention of a human operator to execute a manual operation to bring back the case being treated to a nominal B.A.U process (i.e business as usual flow). An exception can be detected and triggered at different level in the process flow: during the automated classification, during the manual indexing, during the manual classification and during the advanced graphical documents handling process.

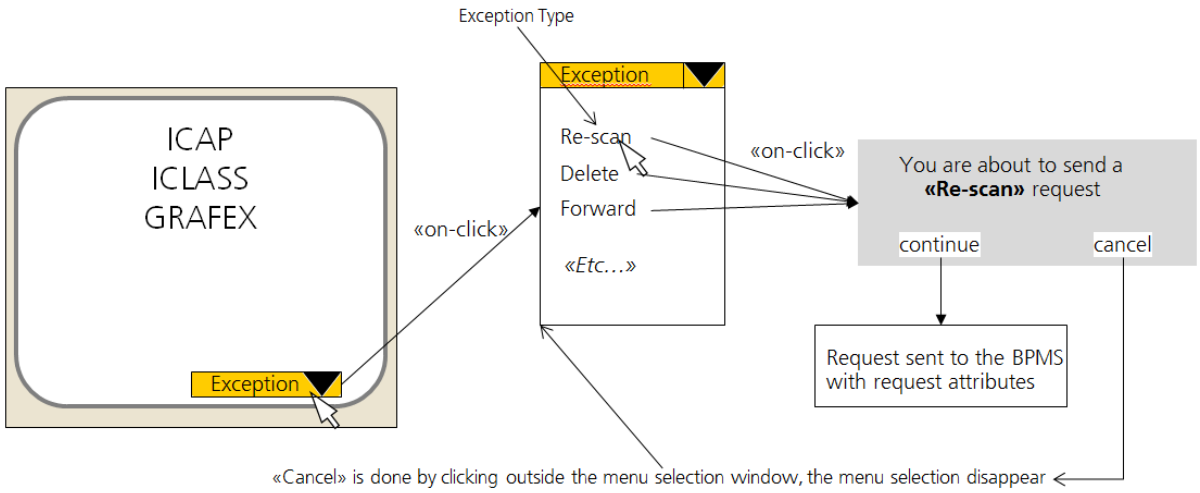


Figure 9 Exception trigger and exception options

The exception trigger is displayed on the web interface of the CDIPP Users and is fully configurable to match customer business processes and SPS operation specific work process. It is possible to display or not the Exception trigger as well as the underlying exception options based on a configurable role/permission mechanism. This flexibility is for SPS a mandatory feature, allowing to fit individuals experience and responsibilities towards the customer business processes.

Graphical Exception handling

At an advanced level and under certain business circumstances, it is necessary to manipulate a scanned original document

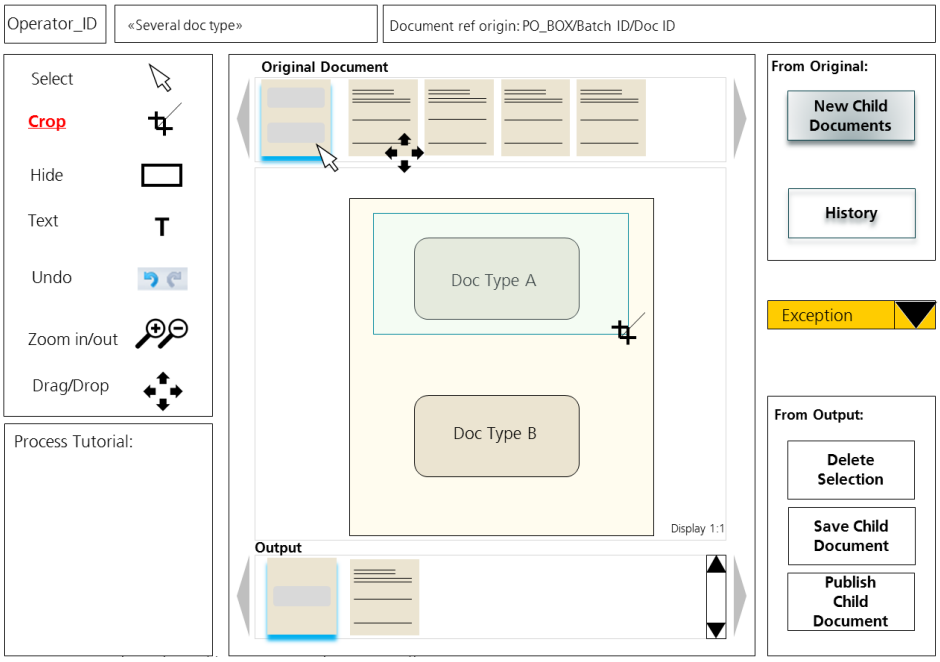


Figure 10 Grafex interface for advanced document images manipulation

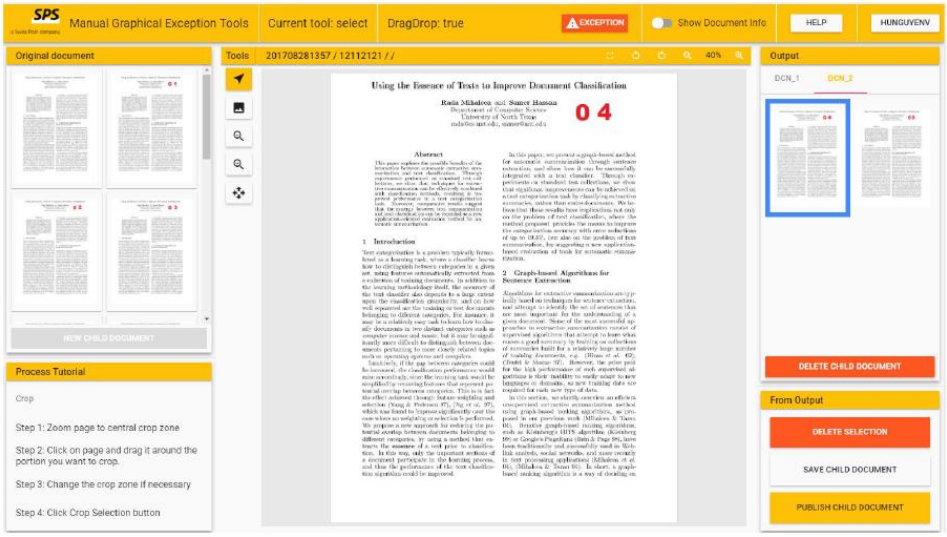


Figure 11 example of Grafex interface

Grafex possesses a set of sub-functionalities or tools that are accessible via a right management rule process configurable. Depending on SPS operation process an operator user may or not have access to the full tool-set of Grafex.

Workflows Creation

Thanks to the Processing Activity Coordinator modules, various workflows can be configured that contains role- based connectors. These connectors are various and use for decision-making, value testing, direct actions CDIPP module calls etc.

Workflows creation follow a two steps process: Workflow Design and Technical Configuration as shown on the figures below.

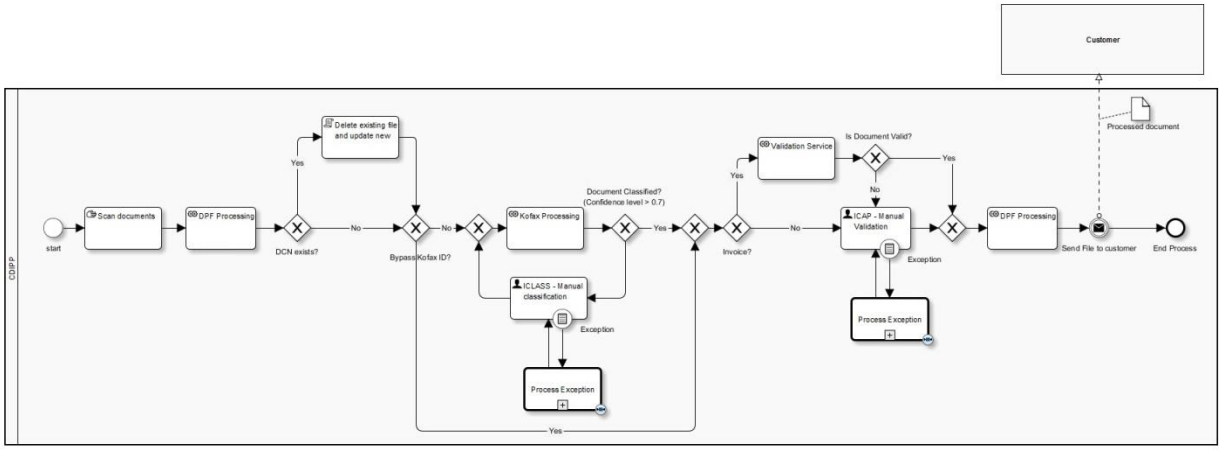


Figure 12 Example of workflow design

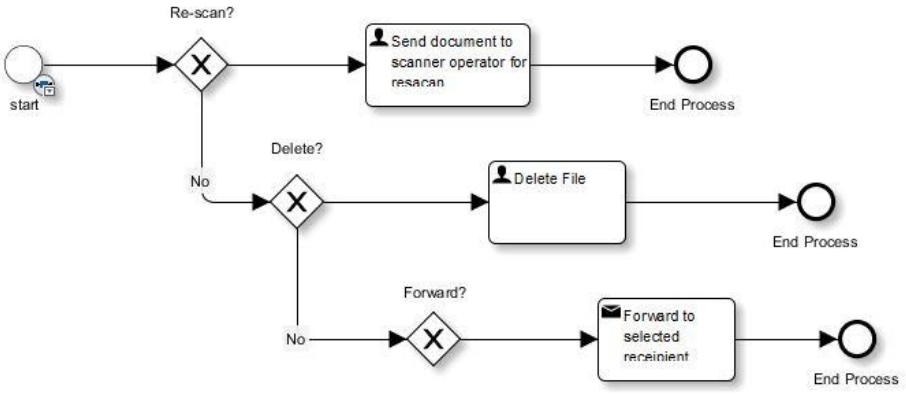


Figure 13 Example of exception handling workflow design

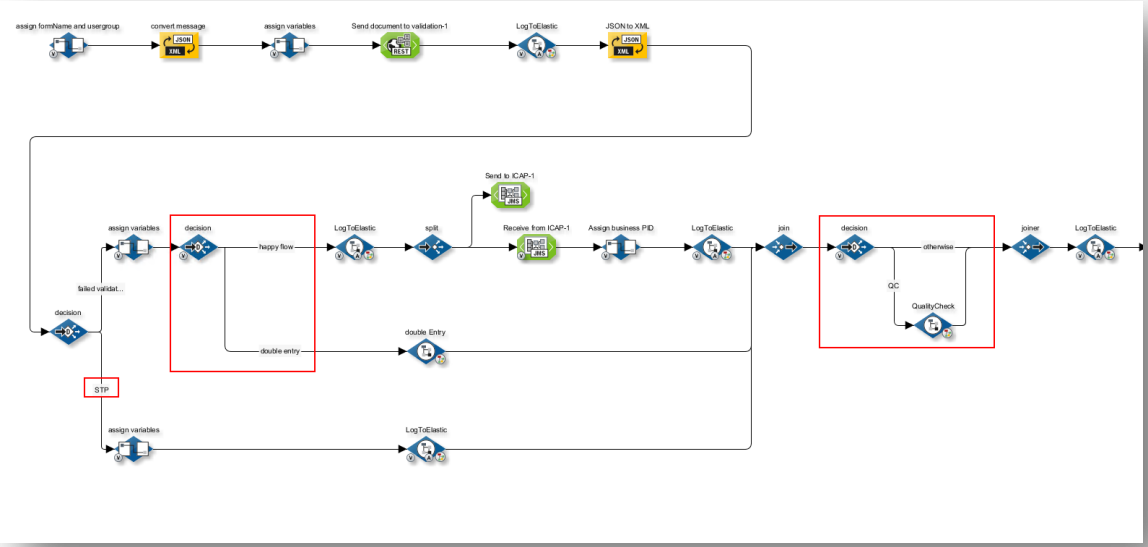


Figure 14 Example of Technical workflow configuration

The underlying BPMS provides SPS with a highly flexible workflow creation engine that allows to configure at will the steps needed to fulfill a specific process. It can call a manual indexing whenever needed and according to simple rules can by-pass some CDIPP components etc.

Quality Control

The quality control is made via CDIPP Quality Centre. The objectives of the QC is to ensure the highest level of reliable quality for the manual operations (ICLASS, ICAP, GRAFEX) done by the operators.

The quality control process is configurable (single or double blind entry, with or without quality check etc.) and depends on SPS operation manager.

The QC provides a tutorial mode that allows new Operators to be under the support of a Senior Operator for learning curve purpose.

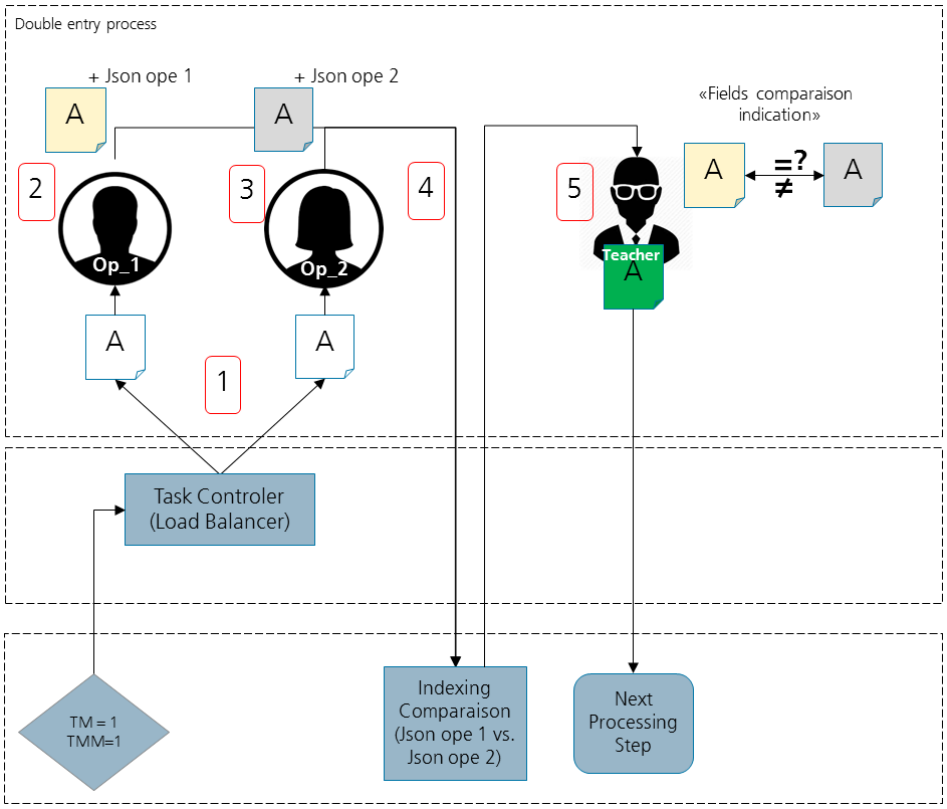


Figure 15 Example of Quality check process with tutorial mode activated

The way the Quality Centre is set-up is under the responsibility of the supervisor.

Operation Monitoring and Reporting

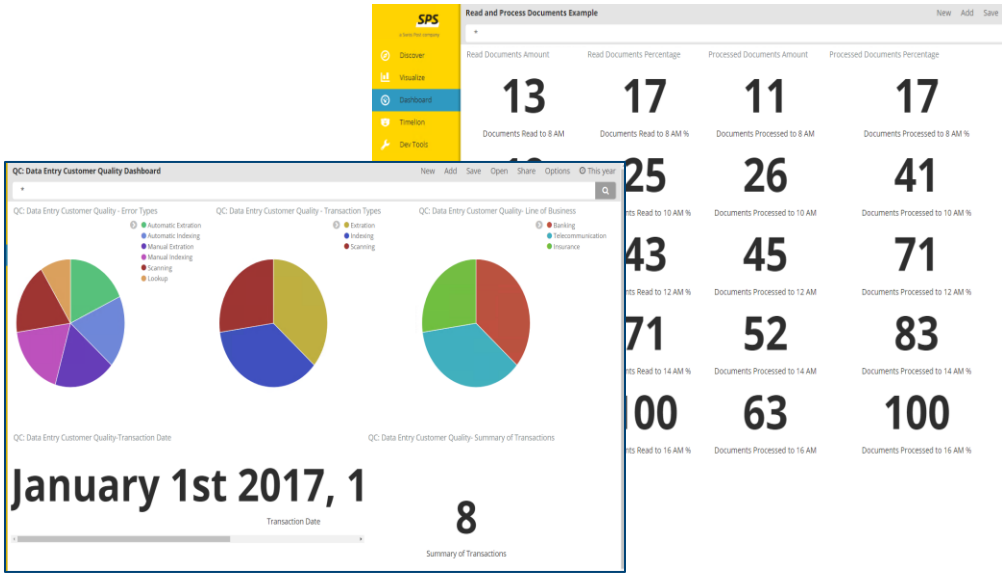


Figure 16 Example of CDIPP Process monitoring and reporting

CDIPP allows to configure and to provide various reports including but not limited to processes and document status, reconciliations reports, Operation monitoring information and billing information, Special reports etc.

File Name	Date of File Transfer	Time of File Transfer	Acknowledgement Message	File Size
2203846239423Invoice	06.07.2017	06:08:19	OK. 1	21
23423423423Invoice	06.07.2017	06:06:19	OK. 1	20
234490240023Invoice	15.11.2017	17:15:19	Failed. 0	19
24789234732749 83Invoice	01.07.2017	14:18:19	OK. 1	18
28232342342403456815AddressChange	10.07.2017	16:08:19	OK. 1	19
2823474562403928 15AddressChange	12.07.2017	16:08:19	OK. 1	14
28732342394094839423Invoice	19.07.2017	07:08:19	OK. 1	18
2873294839423Invoice	08.07.2017	07:08:19	OK. 1	22
3485792340823Invoice	12.07.2017	14:10:19	Failed. 0	18
938234002Invoice	11.07.2017	16:10:19	Failed. 0	16
9809089247Invoice	04.07.2017	08:10:19	Failed. 0	21

Figure 17 Example of reconciliation

Item Description	Requestor	Document Control Number	Rescan Date	Date
X-Ray	Bernd Reinke	802.15-17/003.04	November 1st 2017, 19:13:50.000	January 1st 2017, 19:11:51.000
X-Ray	Sarah Maier	802.15-17/003.03	November 1st 2017, 19:12:50.000	January 1st 2017, 19:11:51.000
X-Ray	Helmut Zagorny	802.15-17/003.01	November 1st 2017, 19:11:50.000	January 1st 2017, 19:11:51.000
X-Ray	Katharina Spangenkamp	802.15-17/003.06	November 1st 2017, 17:15:50.000	January 1st 2017, 19:11:51.000
X-Ray	Anna Mohnke	802.15-17/003.05	November 1st 2017, 17:13:50.000	January 1st 2017, 19:11:51.000
Optical Media	Hans Peter	802.15-17/003.00	November 1st 2017, 19:11:51.000	January 1st 2017, 19:11:51.000
RTS Check	Patrizia Lang	802.15-17/003.07	November 1st 2017, 17:11:50.000	January 1st 2017, 19:11:51.000
RTS Check	Deborah Klein	802.15-17/003.08	November 1st 2017, 15:10:20.000	January 1st 2017, 19:11:51.000
RTS Check	Dennis Stengel	802.15-17/003.12	November 1st 2017, 14:10:20.000	January 1st 2017, 19:11:51.000
RTS Check	Franziska Behl	802.15-17/003.10	November 1st 2017, 14:10:20.000	January 1st 2017, 19:11:51.000

Figure 18 Example of reporting on specific activity like a "Re-Scan"

CDIPP Deployment

CDIPP can be deployed in a private network or in a cloud. It follows IT standard in term of network architecture.

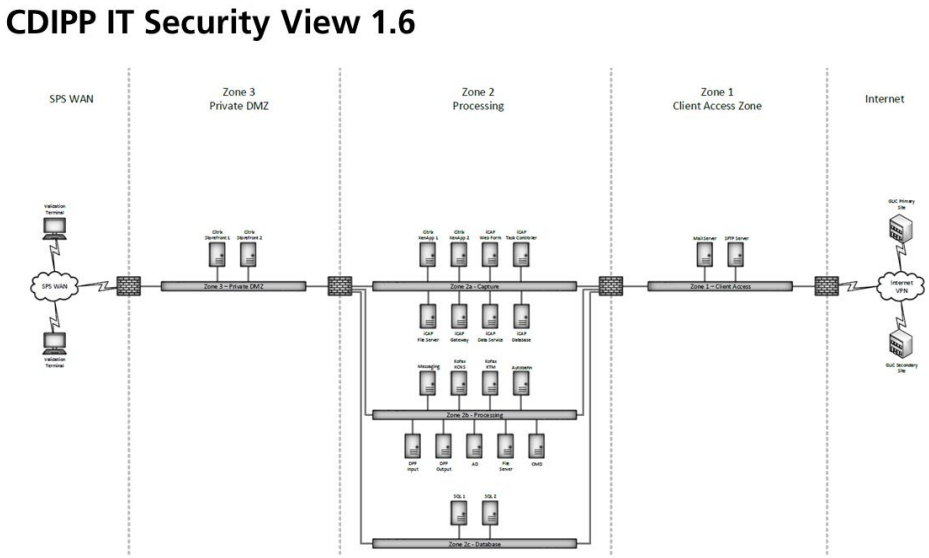


Figure 19 Example of CDIPP IT security diagram

CDIPP can be deployed with or without High availability with or without a disaster recovery secondary setup. Two environments are usually set-up: a productive and a staging or integration environment. The productive environment follows strict access and delivery rules, whether the integration one is used for pre-validation before planned migration to production.

Annexes

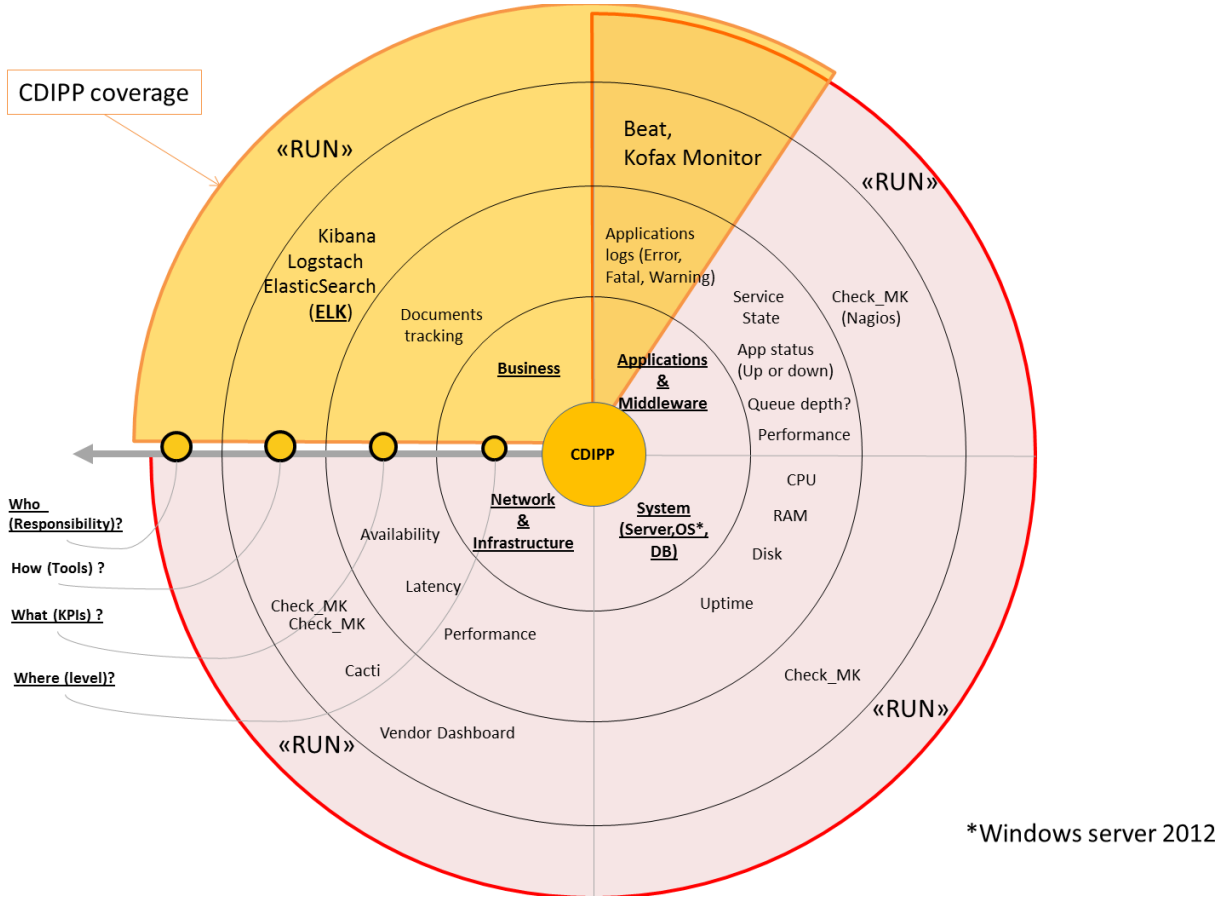
Lookup provisioning options

Approach	Description	Pros	Cons	Next steps & questions
Offline Lookups	Lookups available in SPS (IBM) data center Updates & full refresh delivered on scheduled basis to SPS Normal approach for SPS when dealing with large lookup data (SPS has experience in receiving and handling lookups which are over 16GB in size) Update & full refresh activities to be done out of hours	High availability and limited delays on SPS side (as per original assumptions) Optimised for SPS activities and systems Fuzzy lookups possible in DATEX Full flexibility on SPS side to adjust lookups and queries for individual and precise activities Possibility to work with data despite any connectivity problems between SPS and Customer No VPN overload during business hours	Size of the lookups and transfer can be a challenge Some of the data might be out of date Lookup data would have to be encrypted prior to transfer Work required on Customer side to provide the lookups	
Online lookup	Lookups available at Customer side SPS to connect via web service or direct	Segregated and isolated usage of data by SPS	Potential risk with transfer delay over multiple firewalls and networks	Next Steps: SPS to prepare test in June to test

	lookup (not preferable)	Full track of activities on Customer side	Exact interfaces need to be defined to collect the data Limited result possible - queries from SPS will result with limited records back No fuzzy lookups possible Limited possibility of usage this approach in DATEX These kind of lookups will slow down the automated process on SPS side. This might result overall longer processing times than previously estimated. Potentially longer change process when comes to new/updated query or issue resolution - both partners will need to be involved High traffic between SPS and Customer data centers Potential issue when	performance of the lookups performance for Wave 2
Cached data	Mixture of two approaches above Information to be stored temporary on SPS side whenever possible Every query result	Faster response on data already cached Potentially lower traffic between both data centers	Faster response only on data already cached Limited fuzzy lookup possibility Data to be stored only temporary	

	would be stored on SPS DB for future similar queries		Not considered in original design. Adjustments on effort and plan to accommodate this approach.	

CDIPP Monitoring



Disclaimer

This document is the property of SPS. It can be shared with authorized parties outside SPS upon contractual agreement. Any of the contents partially or in whole, shall not be distributed or disclosed otherwise. The technical content described constitutes a picture of the platform at a point in time and does not ensure any parts of the platform to remain as described. SPS reserves the right to adapt the architecture and/or the components according to any reason and at the discretion of SPS. The document is providing for information only.