Katholieke
Universiteit
Leuven

**Department of
Computer Science**

DOCUMENT PROCESSING

Domain Analysis

Software Architecture (H09B5a and H07Z9a) – Part 1

Jeroen Reinenbergh (r0460600)
Jonas Schouterden (r0260385)

Academic year 2014–2015

Contents

1 Domain analysis

1.1 Domain model
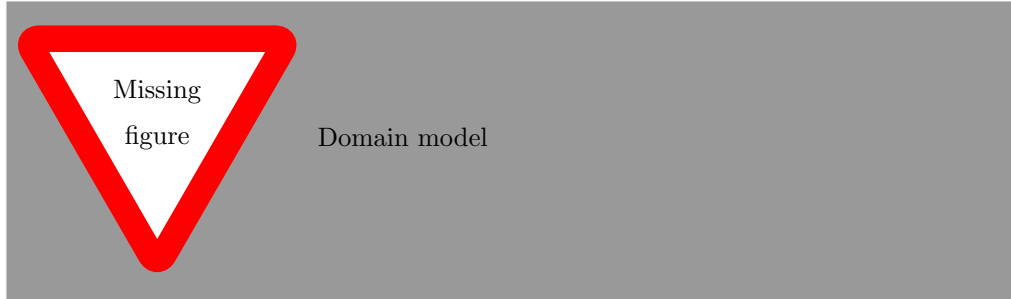
This section shows the domain model.

Figure 1: The domain model for the system.

1.2 Domain constraints

In this section we provide additional domain constraints.

- A Recipient should have at most one personal document store.
- A Recipient should get a specific document over one channel.
- The System must first generate a document before it can archive it.
- If a document is stored in the archive, the raw data on which it is based should also be stored in the archived.
- An Employee can receive only payslips.
- A Customer organization client can receive only invoices.
- *Employee* and *Customer organization client* are roles. They might be taken on by the same person.
- There are two possible ways data from a Customer organization can be provided. The first way is the Customer organization providing the data for both invoices and payslips. The alternative consists of the Customer organization providing only the data for invoices while the data for payslips is being provided by one or more Social secretaries.
- Each Customer organization must fill in a separate template for payslips and invoices.
- A Personal document store can only show documents that are in the archive.
- It should be possible for a Recipient's Personal document store to show all of the Recipient's documents that have previously been archived.
- The archive contains all of the documents that have been generated.

1.3 Glossary

In this section, we provide a glossary of the most important terminology used in this analysis.

- **Customer organisation**: an organisation that outsources its document processing for payslips and/or invoices to eDocs.
- **Invoice**: a commercial document issued for the sale of goods and/or services from a seller to a buyer.
- **Payslip**: a document sent out by companies to their employees that describes how their salary for a certain period is calculated.

- **Raw data**: the unformatted information that is represented in a document.

- **Recipient**: a person or organisation that receives a document, e.g. an employee of a customer organisation or a client subscribed to the services of a customer organisation.

- **Employee**: a Receiver of payslips send by a Customer organization representative.

- **Customer organization client**: a Receiver of invoices send by a Customer organization.

- **Archive**: a logical place where the received raw data and generated documents are stored.

- **Meta data**: part of the raw data that specifies the addressing method and address for a specific Recipient.

- **eDocs administrator**: an employee of eDocs that supervises the system infrastructure.

- **Customer organization representative**: a(n) (employee of a) company who can deliver raw data in the name of the Customer organization.

- **Social secretary**: representative of the Customer organization it works for.

- **Service Level Agreement**: a contract between the company eDocs and a customer organization or between the eDocs company and an organization to which eDocs outsources services, specifying the services that have to be provided by both parties and the conditions under which it is valid.

- **template**: a predefined textual structure on which the generated documents are based.

- **Personal document store**: a channel provided by the eDocs company through which a receiver can consult all documents for which he or she is the receiver.

- **Channel**: a medium over which Recipients can receive documents.

- **Addendum**: a document complementing a previously sent payslip.

- **Document data**: the part of the raw data that specifies the contents of a document.

2 Functional requirements
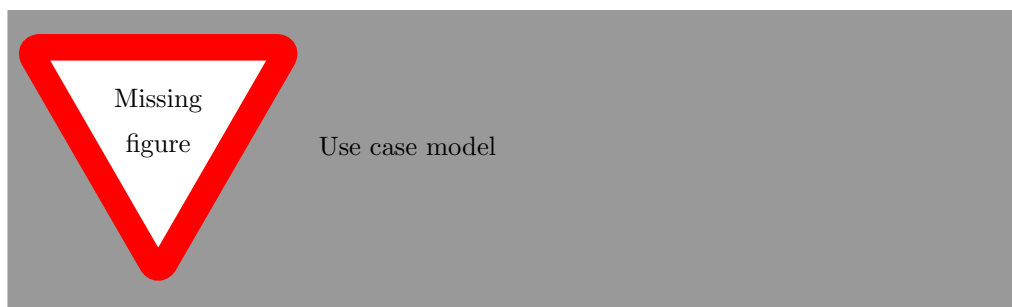
Use case model

Figure 2: Use case diagram for the system.

2.1 *UC7*: Consult document status

- **Name:** Consult document status

- **Primary actor:** Customer organisation

- **Interested parties:**
  - *Customer organisation:* wants to consult the status of one of its documents at eDocs

– *The eDocs system:* provides a management dashboard through which the customer can, among other things, consult the status of its documents

- **Preconditions:**

  – The customer organisation is a registred client at eDocs

  – The customer organisation is logged into the eDocs system

- **Postconditions:**

  – The customer organisation has consulted the status of the desired document

- **Main scenario:**

  1. The customer orgnanisation provides the eDocs system with enough information for it to identify the desired document of which the status needs to be consulted

  2. The eDocs system retrieves the status of the respective document

  3. The eDocs system provides the customer organisation with the retrieved document status information

  4. The customer organisation consults the document status information received from the eDocs system

2.2 *UC8*: Register customer organisation

- **Name:** Register customer organisation

- **Summary:** The eDocs administrator registers a new customer organisation in order to enlarge the eDocs client base

- **Primary actor:** eDocs administrator

- **Secondary actor:** Customer organisation

- **Interested parties:**

  – *Customer organisation:* wants to outsource its document processing to eDocs

  – *eDocs:* wants to gain new clients customer organisations) for which it can process documents

  – *eDocs administrator:* deals with the administrational tasks during the registration of customer organisations

- **Preconditions:**

  – The customer organisation has actively expressed its interest in eDocs

  – The eDocs administrator has initiated negotiations about the contract and SLA with the customer organisation

- **Postconditions:**

  – eDocs has gained a new client (and friend)

  – The customer organisation and eDocs administrator have succesfully negotiated a contract and SLA

- **Main scenario:**

  1. The customer organisation has expressed its interest in eDocs as its potential document processing partner

  2. The eDocs administrator starts negotiations with the customer organisation about the contract and SLA

  3. Both parties agree upon the terms and conditions specified in the forenamed documents

  4. The customer organisation is now a client of eDocs and has agreed to outsource (part of) its document processing tasks to eDocs

- **Alternative scenarios:**

3a. The two parties do not reach an agreement

4a. The customer organisation expresses its intent to outsource its document processing tasks elsewhere

5a. eDocs had lost a potential client

- **Assumptions:**

  - The registration process is initiated by the customer organisation expressing its interest in eDocs. It is thus assumed that eDocs is a big player in the document processing market, following a passive strategy to gain new clients.

2.3   *UC9*: Deliver raw invoice data

- **Name:** Deliver raw invoice data

- **Summary:** The customer organisation provides eDocs with the raw invoice data to be processed

- **Primary actor:** Customer organisation

- **Secondary actor:** eDocs system

- **Interested parties:**

  - *Customer organisation:* wants to deliver its raw invoice data to eDocs so that the latter can process it accordingly
  - *System:* accepts the raw invoice data to be processed

- **Preconditions:**

  - The customer organisation is logged in to the eDocs system

- **Postconditions:**

  - The customer organisation has provided the eDocs system with its raw invoice data
  - The eDocs system has succesfully received the customer organisation's raw invoice data

- **Main scenario:**

  1. The customer organisation sends its raw invoice data to eDocs via the management dashboard or via a web service that provides a link from the customer organisation's system to that of eDocs
  2. The eDocs system receives the customer organisation's raw invoice data

2.4   *UC10*: Deliver raw payslip data

- **Name:** Deliver raw payslip data

- **Summary:** The customer organisation provides eDocs, either directly or indirectly, with the raw payslip data to be processed

- **Primary actor:** Customer organisation

- **Secondary actor:** eDocs system

- **Interested parties:**

  - *Customer organisation:* wants to deliver its raw payslip data to eDocs so that the latter can process it accordingly
  - *Social secretary:* wants to be an intermediate party in the delivery of raw payslip data
  - *System:* accepts the raw payslip data to be processed

- **Preconditions:**

  - The customer organisation representative is logged in to the eDocs system

- **Postconditions:**

- The customer organisation representative has provided the eDocs system with its raw payslip data
- The eDocs system has succesfully received the customer organisation's raw payslip data

- **Main scenario:**

  1. The customer organisation sends its raw payslip data to eDocs via the management dashboard or via a web service that provides a link from the customer organisation's system to that of eDocs
  2. The eDocs system receives the customer organisation's raw payslip data

- **Alternative scenarios:**

  1a. The customer organisation provides its social secretary with the necessary employee and tax information
  2a. The social secretary integrates the raw employee data into raw payslip data conform to the eDocs standard
  3a. Continue from step 2.

- **Remarks:**

  - The customer organisation representative is either its social secretary or the customer organisation itself

2.5   *UC11*: Process raw data

- **Name:** Process raw data

- **Summary:** The eDocs system processes the raw data that was provided by the customer organisation in a previous step

- **Primary actor:** eDocs system

- **Secondary actor:** None

- **Interested parties:**

  - *Customer organisation:* wants its raw data to be processed correctly
  - *eDocs system:* wants to process the raw data correctly

- **Preconditions:**

  - The customer organisation, to which the raw data belongs, is a registered client of eDocs
  - The raw data that has to be processed, was successfully delivered to the eDocs system

- **Postconditions:**

  - The raw data is processed

- **Main scenario:**

  1. Check raw data for consistency and completeness
  2. Process meta data
  3. Process document data according to meta data and previously provided template
  4. Generate document

- **Alternative scenarios:**

  2a. Erroneous data is sent back to the customer organisation from which it originated
  3a. The customer organisation corrects the erroneous part(s) of the respective (batch of) raw data
  4a. The customer organisation delivers a (batch of) raw data consisting only of those corrected records that were (partly) erronous at the beginning of this processing phase
  5a. Restart from step 1.

- **Remarks:**

  - If the raw data were marked as critical by the provider, they should be processed with the highest priority (within the overall five hour deadline)

2.6 *UC12*: De-activate personal document store

- **Summary:** A Registered recipient de-activates his or her personal document store account.

- **Primary actor:** Registered recipient

- **Secondary actors:** The System

- **Interested parties:**

  - *Registered recipient:* wants to de-activate his or her personal document store
  - *eDocs:* wants to keep as many Registered recipients as possible.

- **Preconditions:** None

- **Postconditions:**

  - The Registered recipient is now an Unregistered recipient and can no longer log in to the system.
  - The Registered recipient is now logged out and can no longer make use of his/her personal document store.
  - The System has removed the account of the Registered Recipient. It keeps the documents of the Registered recipient archived.
  - The Registered recipient now receives his or her documents via another channel than his or her personal document store.

- **Main scenario:**

  1. The Registered recipient indicates that he or she wants to de-activate his or her personal document store.
  2. The System removes the account of the registered recipient.
  3. The System indicates success to the Registered recipient and logs him or her out.

- **Alternative scenarios:**

  1. . . .

- **Remarks:**

  - None

2.7 *UC13*: Provide document template

- **Summary:** An eDocs administrator

- **Primary actor:** eDocs administrator

- **Interested parties:**

  - *eDocs administrator:* wants to provide a document template that a customer organization can fill in.
  - *Customer organization:* wants to have a template for its documents
  - *eDocs:* wants to generate documents based on a template that the Customer organization has filled in.

- **Preconditions:**

  - The Customer organization has registered itself with eDocs.

- **Postconditions:**

  - The System now contains a template which the Customer organization can fill in.

- **Main scenario:**

1. The eDocs administrator provides a document template to the System.

- **Alternative scenarios:**

    1. None

- **Remarks:**

    – None

2.8 *UC14*: Update document template

- **Summary:** The Customer organization updates a document template.

- **Primary actor:** Customer organization

- **Secondary actors:**

    – eDocs administrator

    – eDocs System

    eDocs administrator

- **Interested parties:**

    – *Customer organization:* wants its documents to have a layout corresponding to the initial template it fills in.

    – *eDocs:* wants to generate documents with a layout corresponding to the wishes of its clients.

- **Preconditions:**

    – The Customer organization is logged in.

- **Postconditions:**

    – The System from now on generates documents using the new filled in document template.

    – The System stores the provided document template.

- **Main scenario:**

    1. The Customer organization indicates it wants to download the current document template.

    2. The System looks up the correct document template and provides it to the Customer organization.

    3. The Customer organization changes the current document template.

    4. The Customer organization indicates it wants to upload the filled in template.

    5. The System downloads the updated document. It notifies an eDocs administrator that a new filled in document is waiting to be verified for correctness.

    6. An eDocs administrator verifies the correctness of the document template.

    7. The System stores it. The updated document replaces the old document template for that Customer organization and that document type.

- **Alternative scenarios:**

    6a. If the provided document template is not correctly filled in, the administrator notifies the Customer organization. Continue from step 3.

- **Remarks:**

    – If the Customer organization has not filled in a document template before, the document template presented to the Customer organization is the document template that the eDocs administrator provided for the Customer Organization to use.

2.9 *UC15*: Enable receipt tracking

- **Summary:** The Customer organization enables receipt tracking

- **Primary actor:** Customer organization

- **Interested parties:**
    - *Customer organization:* wants to know if a Recipient has received his or her documents.
    - *Recipient:* might want to know if he or she is being tracked.

- **Preconditions:**
    - The Customer organization is logged in.

- **Postconditions:**
    - Receipt tracking is now enabled for the distribution channels that support it.

- **Main scenario:**
    1. The Customer organization indicates that it wants to enable receipt tracking.
    2. The System stores this setting.

- **Alternative scenarios:**
    1. ...

- **Remarks:**
    - Receipt tracking is only supported for e-mail, Zoomit and the personal document store.

2.10 *UC16*: Disable receipt tracking

- **Summary:** The Customer organization disables receipt tracking

- **Primary actor:** Customer organization

- **Interested parties:**
    - *Customer organization:* no longer wants to be able to track receipts.
    - *Recipient:* might want to know if he or she is being tracked.

- **Preconditions:**
    - The Customer organization is logged in.

- **Postconditions:**
    - Receipt tracking is now disabled for all distribution channels.

- **Main scenario:**
    1. The Customer organization indicates that it no longer wants to enable receipt tracking.
    2. The System stores this setting.

- **Alternative scenarios:**
    1. ...

- **Remarks:**
    - Receipt tracking is only supported for e-mail, Zoomit and the personal document store.

2.11 *UC17*: Send document via e-mail

- **Summary:** Initiated by Time, the system sends an e-mail to an Unregistered recipient.

- **Primary actor:** Time

- **Secondary actors:** The System

- **Interested parties:**

  - *Customer organization:* wants the Unregistered recipients to receive their documents.
  - **Unregistered recipient:** wants to receive his or her documents.
  - **eDocs;** wants the documents it sends to be correctly received.
  - **E-mail provider** wants to correctly deliver documents it send by e-mail in a timely manner.

- **Preconditions:**

  - The document to be send is generated correctly.
  - The addressing information provided by the Customer Organization in the raw data is an e-mail address.

- **Postconditions:**

  - The System has send out an e-mail to the Unregistered recipient.

- **Main scenario:**

  1. Time notifies the System that it should send the generated document.
  2. The System generates a unique link. When followed, the link points to a web page where the whole document can be read.
  3. The System generates an e-mail containing a preview of the document and the unique link pointing to the web page with the document.
  4. The System sends the e-mail to the e-mail address provided as addressing information in the raw data.
  5. The System marks the corresponding document processing job as sent by e-mail.
  6. If this document is not part of a recurring batch of document processing jobs, the System adds the cost of delivering the generated document to the bill of the customer organization.

- **Alternative scenarios:**

  2a If receipt tracking is disabled:
    (a) The System generates an e-mail containing a short message indicating that there is a new document from the Customer organization. It puts the document as an attachment to the e-mail.
    (b) Continue at step 4.
  2b. The unique link is not unique. The System keeps generating new links until it generates a unique one.

2.12 *UC18*: Notify about incorrect e-mail address

- **Summary:** After receiving an error message from an e-mail provider, the System notifies the customer organization about the incorrect addressing information.

- **Primary actor:** E-mail provider

- **Secondary actors:** None

- **Interested parties:**

  - *Customer organization:* wants to be notified when it gives incorrect addressing information to the eDocs System.

- **eDocs:** wants to notify a customer organisation about incorrect addressing information.
- **E-mail provider** wants to deliver e-mails to the correct e-mail addresses.

- **Preconditions:**

  - UC: *Send document via e-mail*
  - The addressing information provided by the Customer Organization in the raw data is an incorrect e-mail address.

- **Postconditions:**

  - The Customer organization is notified about the wrong addressing information.

- **Main scenario:**

  1. The E-mail provider receives an e-mail for an incorrect e-mail address.
  2. The E-mail provider notifies the System it has received an e-mail to an incorrect address.
  3. The System extracts the necessary information from the error message received from the E-mail provider.
  4. The System looks up the raw data in its archive corresponding to the incorrect e-mail address and document.
  5. The System notifies the Customer organization about the incorrect addressing information.

2.13 *UC19*: Receive document via e-mail

- **Summary:** The Unregistered recipient recipient receives a document via e-mail.

- **Primary actor:** Unregistered Recipient

- **Secondary actors:** The System

- **Interested parties:**

  - *Customer organization:* wants the Unregistered recipients to receive their documents.
  - **Unregistered recipient:** wants to receive his or her documents.
  - **eDocs;** wants the documents it sends to be correctly received.
  - **E-mail provider** wants to correctly deliver documents it send by e-mail in a timely manner.

- **Preconditions:**

  - UC: *Send document via e-mail*

- **Postconditions:**

  - The Unregistered recipient has received his or her document .

- **Main scenario:**

  1. The Unregistered recipient checks his or her e-mail.
  2. The Unregistered recipient sees the e-mail send by the System and opens it.
  3. The Unregistered recipient now sees the preview of the document and a unique link pointing to a web page.
  4. The Unregistered recipient follows the link in the e-mail address.
  5. The System gets a request that the Unregistered recipient wants to view a document using the unique link.
  6. The System responds by sending a web page to the Unregistered recipient where he or she can see or download the document.
  7. The Unregistered recipient now sees the page where he or she can see and download the document.
  8. The System remembers the fact that the Unregistered recipient sees the document (receipt tracking).

- **Alternative scenarios:**

  3a. If receipt tracking is disabled:
      (a) The Unregistered recipient opens the attachment of the e-mail.
      (b) The Unregistered recipient now sees his or her document.

2.14 *UC20*: Search for document in personal document store

- **Summary:** Registered recipient searches for a specific document in his or her personal document store.

- **Primary actor:** Registered Recipient

- **Secondary actors:** None

- **Interested parties:**

  - **Registered recipients:** want to find a specific document.
  - *Customer organization:* wants the specific document to be found.

- **Preconditions:**

  - The Registered recipient is logged in (cf. *UC1: Log in*.

- **Postconditions:**

  - The Registered recipient has received an overview of all the documents he or she received meet his or her search criteria.

- **Main scenario:**

  1. The Registered recipient indicates that he or she wants to search for a specific document, e.g. in the overview of received documents (cf. *UC4: Consult personal document store*.
  2. The System presents to the Registered recipient one or more ways to specify criteria that the document he/she is looking for must meet.
  3. The System searches in its archive and looks up the documents that meet the specified requirements. The System presents the found documents to the Registered recipient.

- **Alternative scenarios:**

  3a. If the System does not find any documents that match the criteria, it notifies the Registered recipient of this fact. Continue from step 2.

2.15 *UC21*: Send document via print & postal service

- **Summary:** Initiated by Time, the system sends a document by postal mail to an Unregistered recipient.

- **Primary actor:** Time

- **Secondary actors:** Print & postal service

- **Interested parties:**

  - *Customer organization:* wants the Unregistered recipients to receive their documents.
  - **Unregistered recipient:** wants to receive his or her documents.
  - **eDocs:** wants the documents it sends to be correctly received.
  - **Print& postal service:** wants to correctly deliver documents it send by e-mail in a timely manner.

- **Preconditions:**

  - The document to be send is generated correctly.
  - The addressing information provided by the Customer Organization in the raw data is a postal address

- **Postconditions:**
  - The System has send out the document to the Recipient through postal mail.

- **Main scenario:**

  1. Time notifies the System that it should send the generated document.
  2. The System sends the document to the Print & postal service via web services, including the complete postal address of the Unregistered recipient.
  3. The Print & postal service prints and packages the document and delivers it to the Unregistered recipient via postal mail.
  4. The System marks the corresponding document processing job as sent by postal mail.
  5. If this document is not part of a recurring batch of document processing jobs, the System adds the cost of delivering the generated document to the bill of the customer organization.

3 Non-functional requirements

In this section, we model the non-functional requirements for the system in the form of *quality attribute scenarios*. We provide for each type (availability, performance and modifiability) one requirement.

3.1 Availability

3.1.1 *Av1*: Document processing infrastructure failure

The internal (sub-)system responsible for processing the raw data into documents fails or crashes.

- **Source:** Internal

- **Stimulus:** The internal (sub-)system responsible for generating documents fails or crashes.

- **Artifact:** Internal subsystem

- **Environment:** Normal execution

- **Response:**
  - This does not affect the availability of other subsystems, such as (i) the availability of the personal document store, (ii) the delivery of already generated documents, (iii) the availability of persistent data such as user data and billing data, etc..
  - This does not lead to the loss of raw data.
  - This does not lead to the loss of already generated documents.
  - This does not lead to critical documents not being processed on time.
  - Prevention
    * The system responsible for processing raw data should have a guaranteed maximal down-time.
  - Detection
    * The eDocs Operators are notified of this problem.
    * The System is able to detect this problem and goes into degraded modus:
      · Raw data to be processed is stored and processed when the processing system returns operational.
      · Fail gracefully: if the system fails while handling a processing job, the documents it did finish generating are stored in the archive as in a regular processing job.
  - Resolution
    * The notified eDocs Operators address the problem, e.g. by replacing failed hardware or by restarting the software.
    * The System automatically finishes the planned processing jobs, including the ones it possibly was handling during the failure.

- **Response measure:**

- Prevention

  - The System should be down for at most 14 hours per year, in time spans of at most two hours. The time between two down-time periods should be at least the time necessary to process all pending documents to be generated.
  - The time necessary to process raw data into a document should be at most the time agreed upon in the SLA minus the maximal consecutive downtime. This way, there is a time buffer to take into account the possible down-time. More specifically:
    * Critical documents, which have a deadline to be processed within 5 hours, should take at most three hours to process.
    * Non-recurring documents with a diamond priority, which have a deadline of maximal 12 hours to process, should take at most 10 hours to process.
    * Non-recurring documents with a gold priority, which have a deadline of maximal 24 hours to process, should take at most 22 hours to process.
    * Non-recurring documents with a silver priority, which have a deadline of maximal 48 hours to process, should take at most 46 hours to process.

- Detection

  - Detection of failed hardware or crashed software happens within 5 seconds.
  - The eDocs operators are notified within 1 minute.
  - In the transition between normal and degraded modus, no document are lost.

- Resolution

  - It should be possible for the eDocs operators to address the problem in within 2 hours.

3.2 Performance

3.2.1 *P1*: raw data transfer

It should be possible to send raw data to the eDocs system in a timely fashion, even in case of a large number of customer organisations attempting to access this raw data transfer interface (e.g. at the end of each month or at the end of the default fiscal year).

- **Source:** Customer organisation

- **Stimulus:**

  - A registered customer organisation sends (a batch of) raw data to the eDocs system

- **Artifact:** The eDocs system interface through which raw data is received

- **Environment:** Normal execution

- **Response:**

  - The eDocs system is able to receive raw data in a timely fashion while the rate of arriving raw data is lower than a certain value.
  - If the raw data arrive at a higher rate than this value, the system throttles the excessive arrivals so that this amount of raw data continues to be transfered in a timely fashion.
  - A large amount of raw data transfer should not affect the performance of other functionality of the eDocs system.

- **Response measure:**

  - While the rate of arriving raw data is lower than 5000 records / second, the eDocs system should be able to receive at least 95% of all records within 2 seconds and 99% within 5 seconds

3.3 Modifiability

3.3.1 *M1*: Different authentication methods

eDocs currently authenticates its registered users (both customer organisations and their clients) in a uniform way through username-password protection. As this is not considered to be very secure, the IT department might decide to include some other form of authentication someday, either replacing or complementing the current authentication procedure.

- **Source:** IT department

- **Stimulus:**

  – The IT department wants to include another form of security in the authentication process

- **Artefact:** The user authentication process

- **Environment:** Normal execution

- **Response:**

  – Incorporating a new authentication procedure in the eDocs system does not require changes to any other part of the eDocs system.

  – A newly introduced authentication procedure should not cause significant noticeable overhead for registered users when logging into or out of the eDocs system

  – A newly introduced authentication procedure should not cause any noticeable overhead for registered users during their session on the eDocs system

  – The rollout phase of a newly introduced authentication procedure should be communicated to all registered users in a timely and self-explanatory fashion

- **Response measure:**

  – Including a new form of authentication should only delay the authentication procedure of 99% of all registered users by a maximum of 10 seconds (compared to the original userame-password protection)

  – The form of authentication during a session is completely transparent to all registered users

  – Public communication of a new authentication procedure should start at least one month before rollout

  – A newly introduced authentication procedure should be accompanied with sufficient documentation to satisfy the needs of a test group consisting of at least 30 registered users. This documentation should be sent to all registered users, along with the initial publication (see previous measure)