



RDF?

PYTHON SCRIPT ZUM PRÜFEN
UND KORRIGIEREN VON RDF FILES

AUSGANGSLAGE

Bank Software

AUSGANGSLAGE

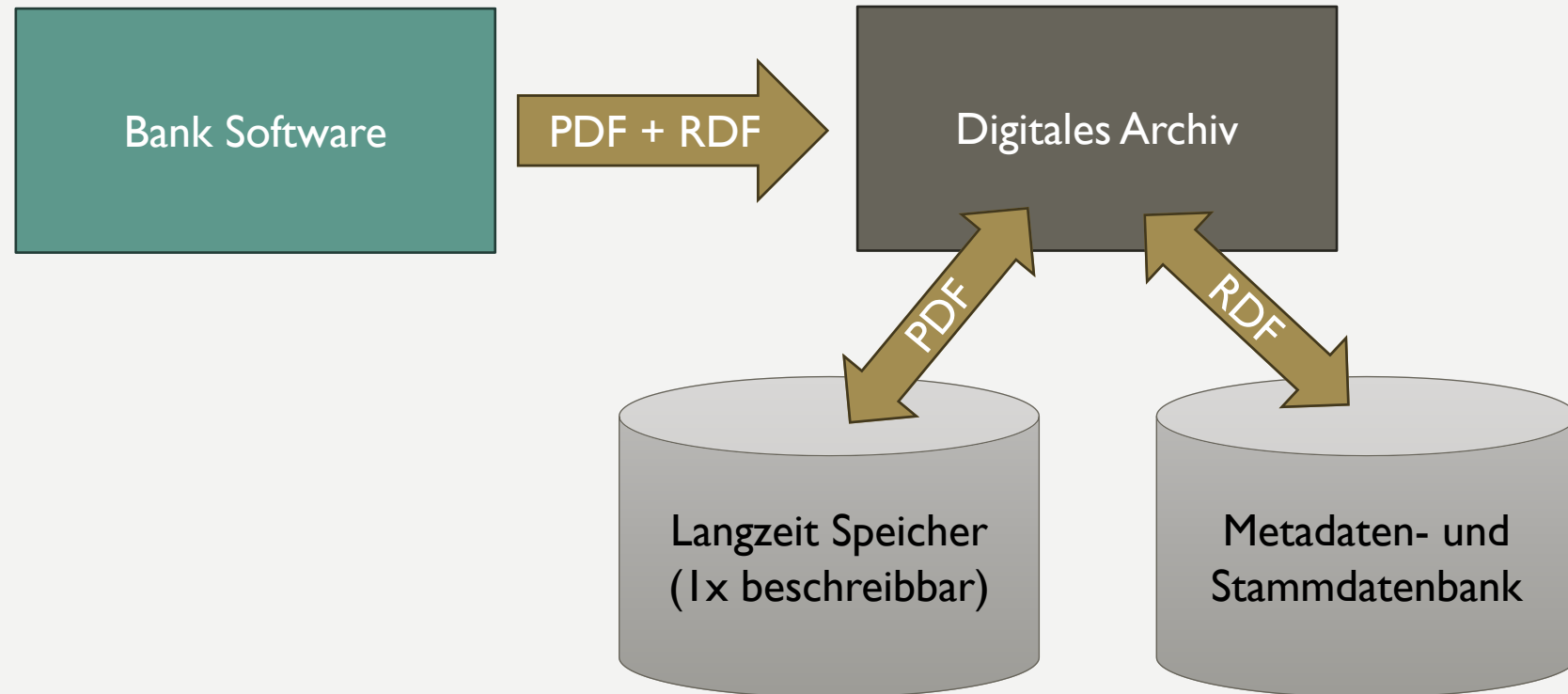
Bank Software

Digitales Archiv

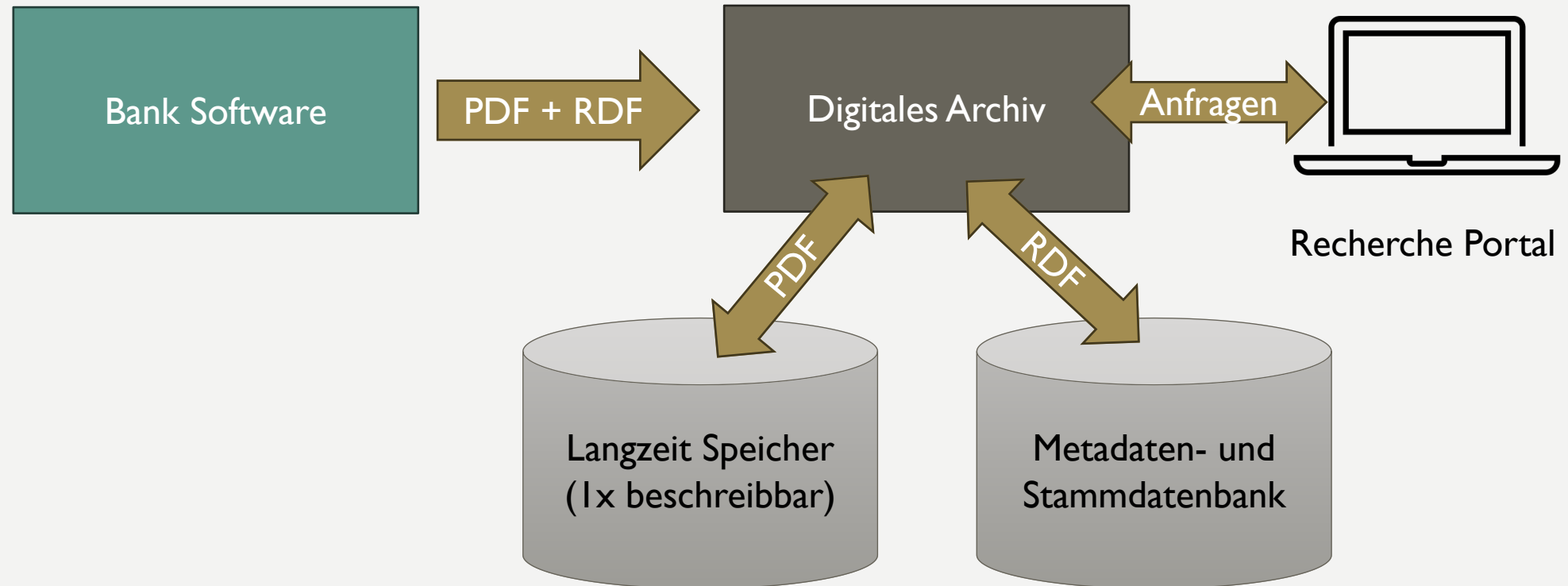
AUSGANGSLAGE



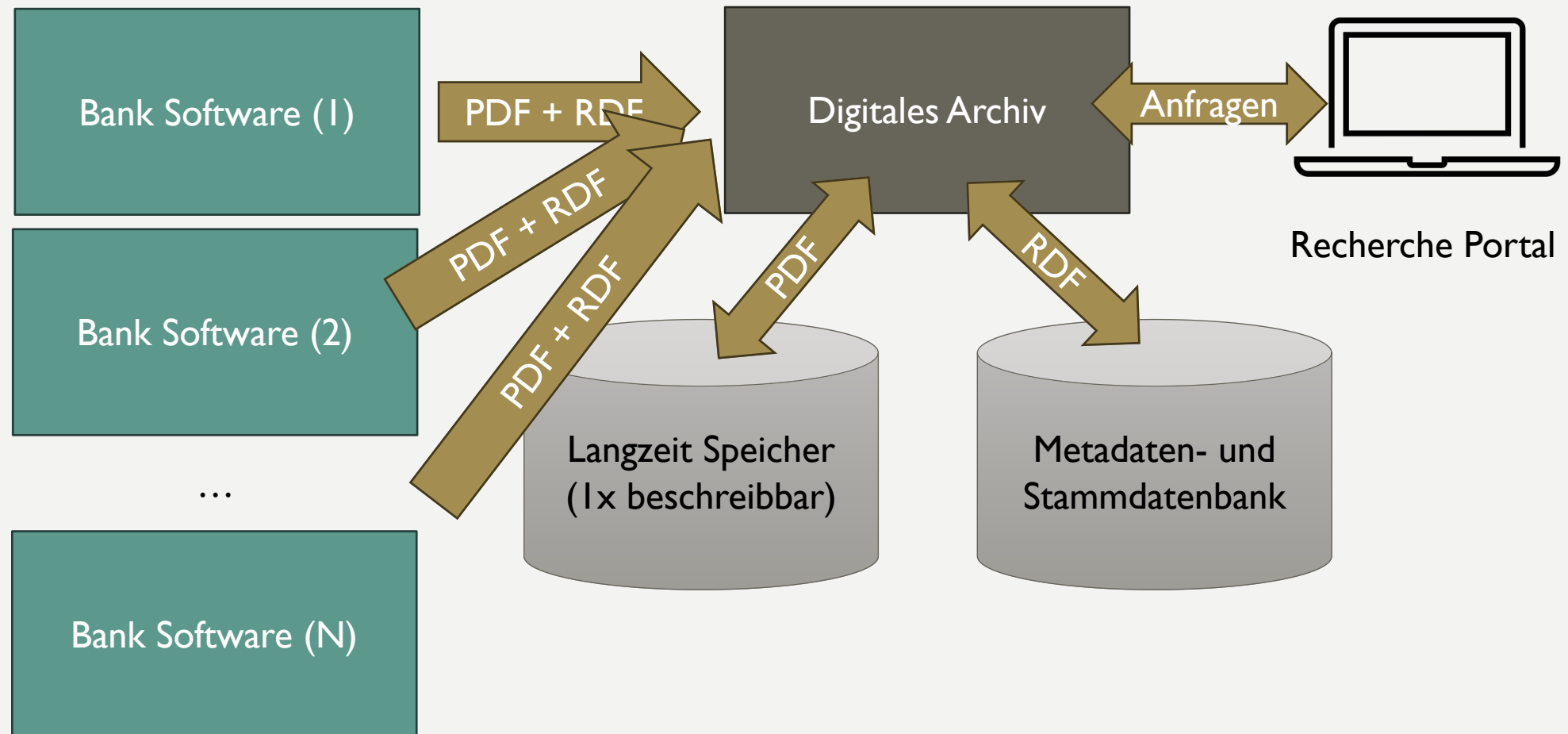
AUSGANGSLAGE



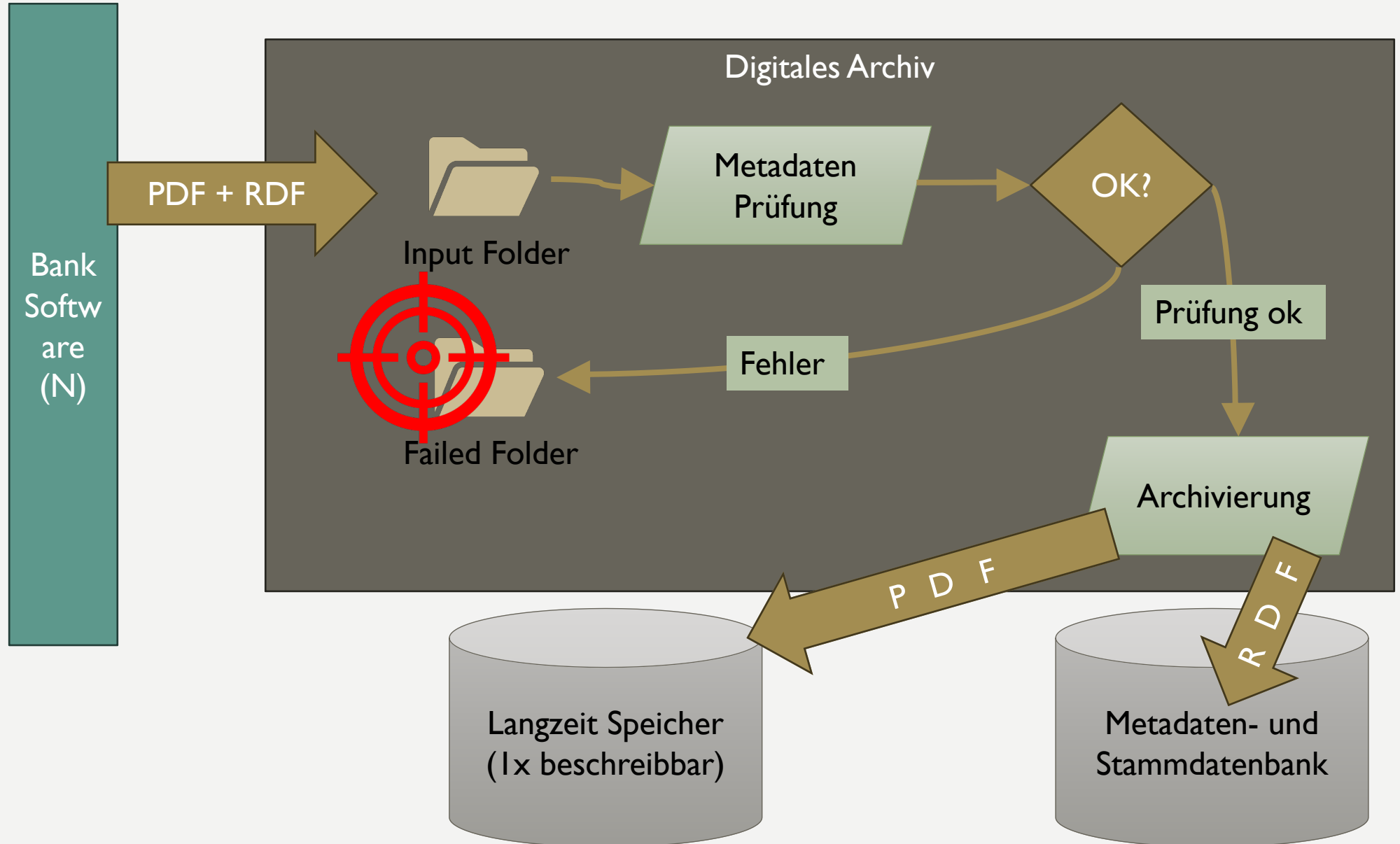
AUSGANGSLAGE



AUSGANGSLAGE



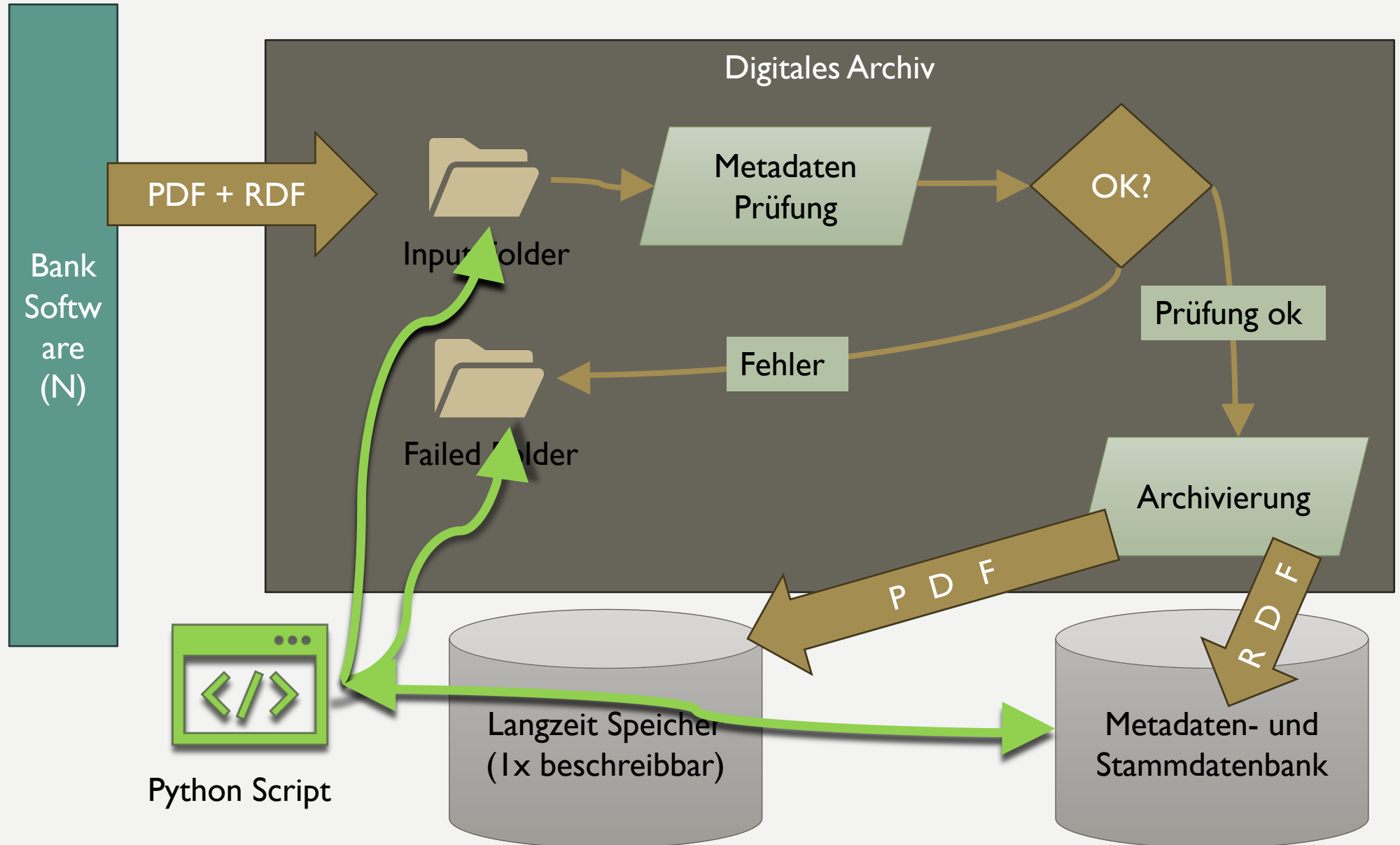
AUSGANGSLAGE



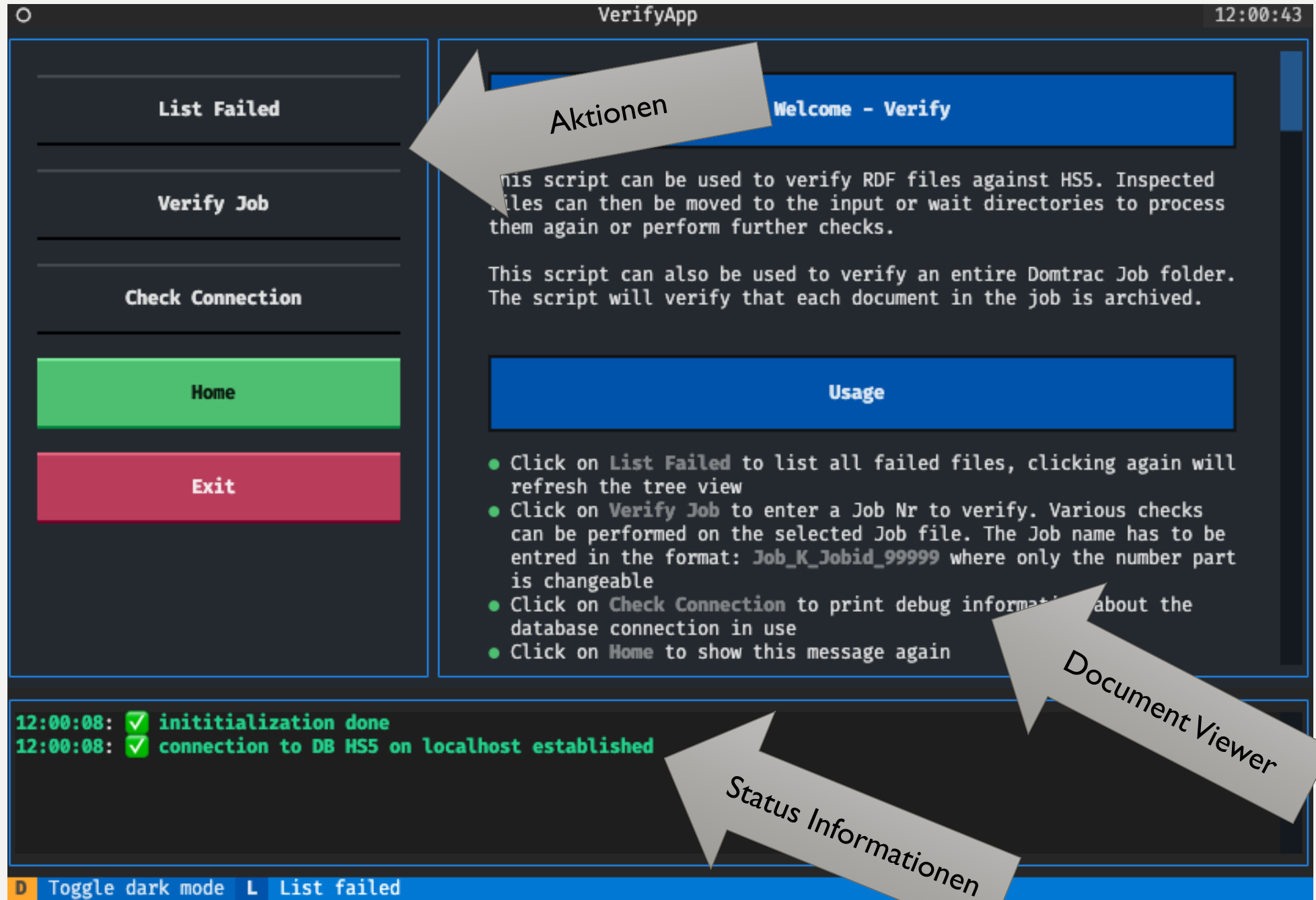
PROBLEM:
MANUELLE PRÜFUNG
VON NICHT
ARCHIVierten
DOKUMENTEN

- Ist die RDF Struktur korrekt?
- Sind die Metadaten aus dem RDF dem Archiv bekannt?
- Sind alle im RDF erwähnten PDF Files auch geliefert worden?
- ...

LÖSUNG



```
def check_client_nr(self, file):
    """retrieve and verify all client numbers in the given file"""
    try:
        tree = ET.parse(file)
        root = tree.getroot()
        ns = {'rdf': 'http://www.w3.org/1999/02/22-rdf-syntax-ns#',
              'edoc': 'http://www.imtf.com/hypersuite/edoc/2.0/'}
        documents = root.findall('./rdf:Description/edoc:ClientNumber', ns)
        current_document = 1
        for document in root.findall('./rdf:Description/edoc:ClientNumber', ns):
            client_nr = document.text
            query = "SELECT COUNT(*) FROM HEST_CLIENT WHERE NR = '%s'%"client_nr
            self.log_debug(query)
            result = self.run_query(query)
            self.log_debug(result)
            if (result[0][0] == 1):
                self.log_success("(%s/%s) Client %s available in HS5"%(current_document,
                                                                           len(documents), client_nr))
            else:
                self.log_error("(%s/%s) Client %s is not available in HS5"%(current_document,
                                                                                len(documents), client_nr))
    except Exception as e:
        if hasattr(e, 'message'):
            self.log_error("unable to get client nr: %s"%(e.message))
        else:
            self.log_error("unable to get client nr: %s"%(e))
```



⌕ [ln02] abaumeler@MBP1 ln02\$



List Failed

Verify Job

Check Connection

Home

Exit

testing/failed



11461595.rdf

11461598.rdf

Job_888893.rdf

12:00:08: ✓ initialization done

12:00:08: ✓ connection to DB HS5 on localhost established



Verify Structure

Verify Metadata

Verify PDF Count

Move to Wait

Move to Input

Close File

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bk="h
  <rdf:Description rdf:about="">
    <edoc:BatchName>Finnova-Job58648</edoc:BatchName>
    <edoc:BatchType>C</edoc:BatchType>
    <edoc:UserBkNr>7740</edoc:UserBkNr>
  </rdf:Description>
  <rdf:Description about="11461595.pdf" sourceFileIdentifier="11461595.pdf"
    <edoc:SchemaName>FinClientDoc</edoc:SchemaName>
    <edoc:BranchIdentifier>100</edoc:BranchIdentifier>
    <edoc:ClientNumber>427101</edoc:ClientNumber>
    <edoc:AccountNumber>110144173601</edoc:AccountNumber>
    <edoc:RelationNumber>427101</edoc:RelationNumber>
    <edoc:ReceiverNumber>427101</edoc:ReceiverNumber>
    <edoc:Classification>VMAUZ</edoc:Classification>
    <edoc:Date>2021-01-01T00:00:00.000</edoc:Date>
    <edoc:PageCount>6</edoc:PageCount>
    <edoc:Type>CI</edoc:Type>
    <edoc:Source>Finnova</edoc:Source>
    <edoc:DeliveryType>WebArchive</edoc:DeliveryType>
    <fin:PagingIndex>false</fin:PagingIndex>
    <fin:FileFormat>4</fin:FileFormat>
    <fin:PageFormat>0</fin:PageFormat>
    <fin:Amount>.00</fin:Amount>
    <fin:DocumentNumber>72589225333015</fin:DocumentNumber>
```

12:00:08: ✓ initialization done

12:00:08: ✓ connection to DB H55 on localhost established

VerifyApp12:03:46

Verify Structure

Verify Metadata

Verify PDF Count

Move to Wait

Move to Input

Close File

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bk="h
  <rdf:Description rdf:about="">
    <edoc:BatchName>Finnova-Job58648</edoc:BatchName>
    <edoc:BatchType>C</edoc:BatchType>
    <edoc:UserBkNr>7740</edoc:UserBkNr>
  </rdf:Description>
  <rdf:Description about="11461595.pdf" sourceFileIdentifier="11461595.pdf"
    <edoc:SchemaName>FinClientDoc</edoc:SchemaName>
    <edoc:BranchIdentifier>100</edoc:BranchIdentifier>
    <edoc:ClientNumber>427101</edoc:ClientNumber>
    <edoc:AccountNumber>110144173601</edoc:AccountNumber>
    <edoc:RelationNumber>427101</edoc:RelationNumber>
    <edoc:ReceiverNumber>427101</edoc:ReceiverNumber>
    <edoc:Classification>VMAUZ</edoc:Classification>
    <edoc:Date>2021-01-01T00:00:00.000</edoc:Date>
    <edoc:PageCount>6</edoc:PageCount>
    <edoc:Type>CI</edoc:Type>
    <edoc:Source>Finnova</edoc:Source>
    <edoc:DeliveryType>WebArchive</edoc:DeliveryType>
    <fin:PagingIndex>false</fin:PagingIndex>
    <fin:FileFormat>4</fin:FileFormat>
    <fin:PageFormat>0</fin:PageFormat>
    <fin:Amount>.00</fin:Amount>
    <fin:DocumentNumber>72589225333015</fin:DocumentNumber>
```

12:00:08: ✓ initialization done

12:00:08: ✓ connection to DB HS5 on localhost established

12:03:39: ✓ structure of testing/failed/11461595.rdf is ok

D Toggle dark modeL List failed

Prüfresultat

VerifyApp12:04:17

Verify Structure

Verify Metadata

Verify PDF Count

Move to Wait

Move to Input

Close File

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bk="h
<rdf:Description rdf:about="">
  <edoc:BatchName>Finnova-Job58648</edoc:BatchName>
  <edoc:BatchType>C</edoc:BatchType>
  <edoc:UserBkNr>7740</edoc:UserBkNr>
</rdf:Description>
<rdf:Description about="11461595.pdf" sourceFileIdentifier="11461595.pdf"
  <edoc:SchemaName>FinClientDoc</edoc:SchemaName>
  <edoc:BranchIdentifier>100</edoc:BranchIdentifier>
  <edoc:ClientNumber>427101</edoc:ClientNumber>
  <edoc:AccountNumber>110144173601</edoc:AccountNumber>
  <edoc:RelationNumber>427101</edoc:RelationNumber>
  <edoc:ReceiverNumber>427101</edoc:ReceiverNumber>
  <edoc:Classification>VMAUZ</edoc:Classification>
  <edoc:Date>2021-01-01T00:00:00.000</edoc:Date>
  <edoc:PageCount>6</edoc:PageCount>
  <edoc:Type>CI</edoc:Type>
  <edoc:Source>Finnova</edoc:Source>
  <edoc:DeliveryType>WebArchive</edoc:DeliveryType>
  <fin:PagingIndex>false</fin:PagingIndex>
  <fin:FileFormat>4</fin:FileFormat>
  <fin:PageFormat>0</fin:PageFormat>
  <fin:Amount>.00</fin:Amount>
  <fin:DocumentNumber>72589225333015</fin:DocumentNumber>
```

12:00:08: ✓ initialization done
12:00:08: ✓ connection to DB HS5 on localhost established
12:03:39: ✓ structure of testing/failed/11461595.rdf is ok
12:04:03: ✓ (1/1) Account 110144173601 available in HS5
12:04:03: ✗ (1/1) Client 427101 is not available in HS5

Prüfresultat

D Toggle dark mode L List failed

VerifyApp12:04:51

Verify Structure

Verify Metadata

Verify PDF Count

Move to Wait

Move to Input

Close File

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bk="h
  <rdf:Description rdf:about="">
    <edoc:BatchName>Finnova-Job58648</edoc:BatchName>
    <edoc:BatchType>C</edoc:BatchType>
    <edoc:UserBkNr>7740</edoc:UserBkNr>
  </rdf:Description>
  <rdf:Description about="11461595.pdf" sourceFileIdentifier="11461595.pdf"
    <edoc:SchemaName>FinClientDoc</edoc:SchemaName>
    <edoc:BranchIdentifier>100</edoc:BranchIdentifier>
    <edoc:ClientNumber>427101</edoc:ClientNumber>
    <edoc:AccountNumber>110144173601</edoc:AccountNumber>
    <edoc:RelationNumber>427101</edoc:RelationNumber>
    <edoc:ReceiverNumber>427101</edoc:ReceiverNumber>
    <edoc:Classification>VMAUZ</edoc:Classification>
    <edoc:Date>2021-01-01T00:00:00.000</edoc:Date>
    <edoc:PageCount>6</edoc:PageCount>
    <edoc:Type>CI</edoc:Type>
    <edoc:Source>Finnova</edoc:Source>
    <edoc:DeliveryType>WebArchive</edoc:DeliveryType>
    <fin:PagingIndex>false</fin:PagingIndex>
    <fin:FileFormat>4</fin:FileFormat>
    <fin:PageFormat>0</fin:PageFormat>
    <fin:Amount>.00</fin:Amount>
    <fin:DocumentNumber>72589225333015</fin:DocumentNumber>
```

12:00:08: ✓ connection to DB HS5 on localhost established
12:03:39: ✓ structure of testing/failed/11461595.rdf is ok
12:04:03: ✓ (1/1) Account 110144173601 available in HS5
12:04:03: ✗ (1/1) Client 427101 is not available in HS5
12:04:39: ✓ structure of testing/failed/11461595.rdf is ok
12:04:39: ✗ At least one PDF is missing in ./testing/failed

D Toggle dark mode L List failed

Prüfresultat

FAZIT

- Potential für enorme Zeitersparnis
- Fehlerquote verringern
- Reaktionszeit erhöhen
- Nur Proof of Concept Status, für produktiven Einsatz noch notwendig:
 - Anbindung an Oracle Datenbank
 - Performance Tests & fine tuning
 - Implementation weitere Checks für Jobfiles

FRAGEN?



Source Code

https://github.com/abaumeler/CAS_DMIT/tree/main/In02