Archiving with windream and in the Cloud

A lecture in two parts

Bochum, November 20, 2019

windream
MANAGING DOCUMENTS

Agenda

01 Digital Archiving

From suspension files to software archives

Electronic archiving is the database-supported, secure and unchangeable storage of electronic information objects that can be reproduced at any time.
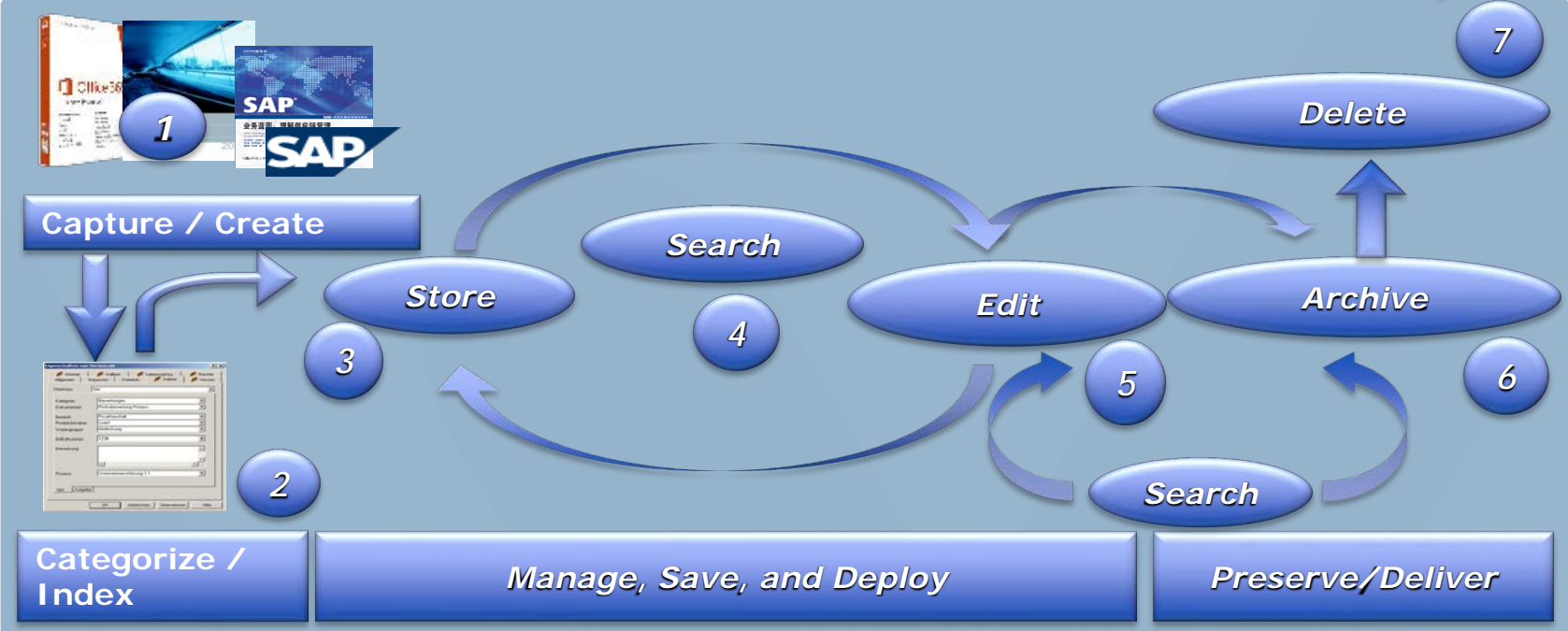
What is an ECM/Archive?

- ✅ An application that enables the unchangeable storage of electronic documents

- ✅ Additional attributes (invoice with invoice number,…)

- ✅ Document history, rights management

- ✅ Search for attributes and content

- ✅ Support of audit-proof hardware - e.g. windream archive, special storage solutions, cloud storage etc.
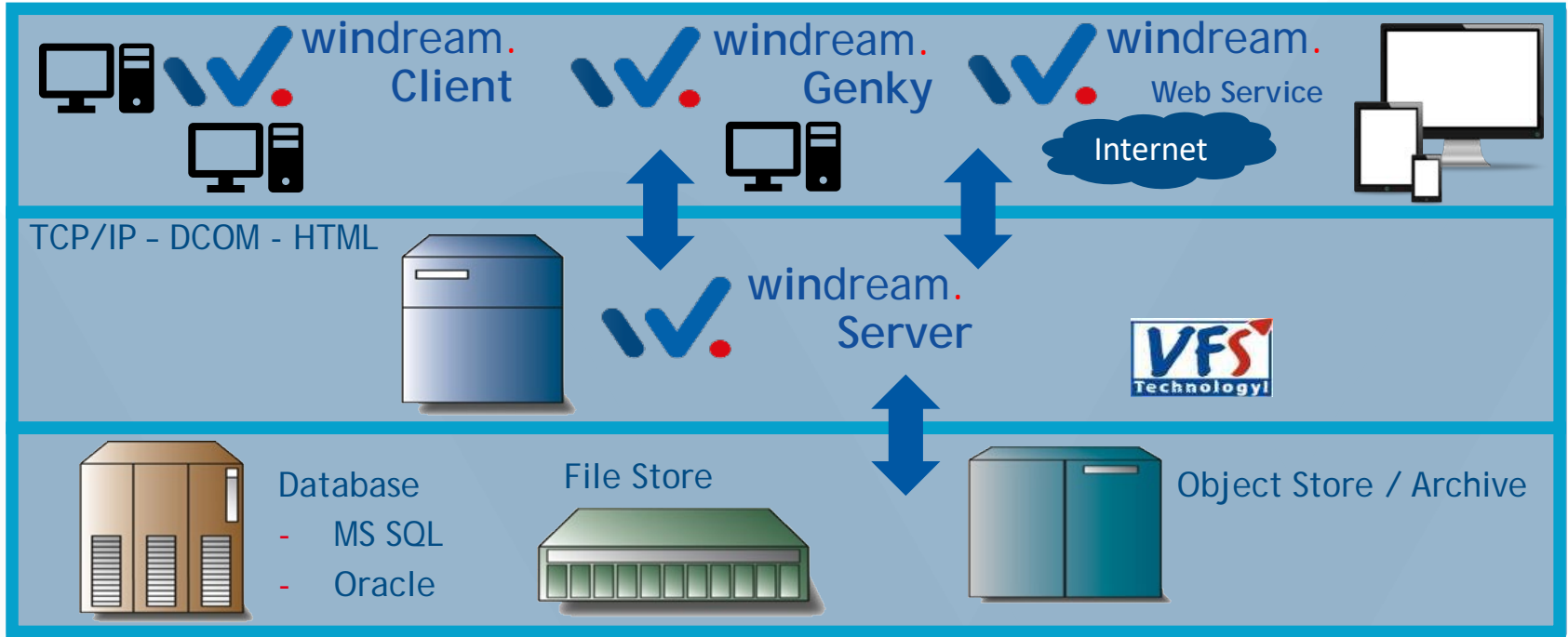
The Document Lifecycle

Storage process in windream

A document in windream basically consists of two parts

✓ The **binary part** of the document (the file) is stored in the file storage or archived in the object memory

✓ The linked **metadata** of the document is stored in the database
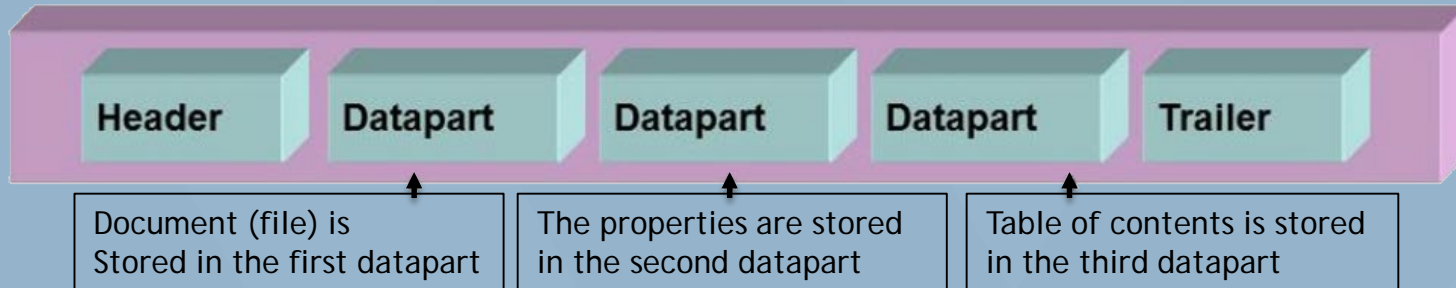
Architecture: Three-tiered model

What is a blob?

Blob (Binary Large Object)
- ☑ Structure for storing a binary object (document) in the archive
- ☑ Consists of at least:
  - ☑ Header for internal administration information
  - ☑ 1-n Dataparts for document data (file), properties and table of contents
 - ☑ Trailer for internal administration information

Structure of a blob, example of a windream document:

| Header | Datapart | Datapart | Datapart | Trailer |

Document (file) is Stored in the first datapart

The properties are stored in the second datapart

Table of contents is stored in the third datapart
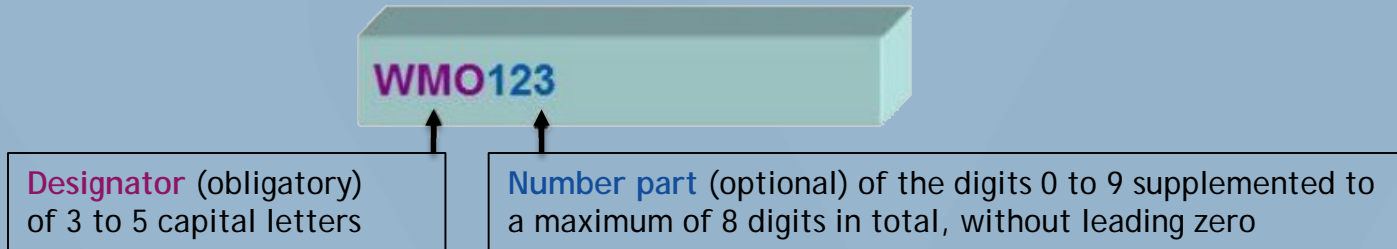
What is the Blob ID?

Blob ID
- ✓ Unique identifier of a blob (e.g. WMO123.PERM1.456)
- ✓ Assigned by the archive service
- ✓ Consists of:

    <World ID> and <Pool ID> and a consecutive number

WMO123.PERM1.456

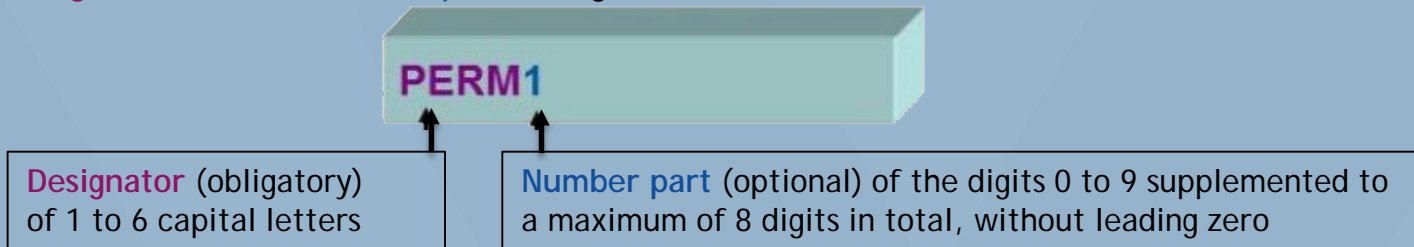| World ID | Pool ID | Consecutive |
| Of the object stores | | number |
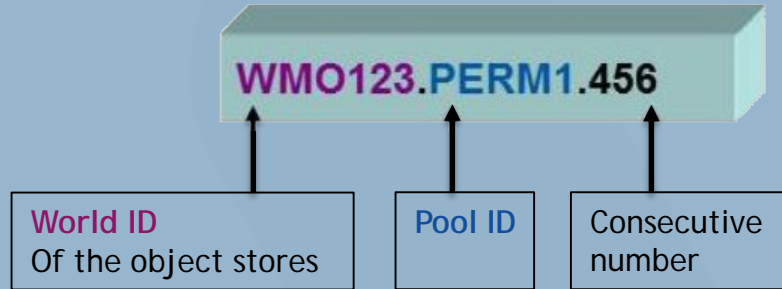
What is the World ID?

World ID
- ✓ Worldwide unique identifier of the object store (e.g. WMO123)
- ✓ Used to generate the Blob IDs, uses the names of the blob containers and the blob registers
- ✓ Consists of:

  <Designator> and <number part> (e.g. AAA12345, ABCDE1 or ABC)

WMO123

Designator (obligatory)
of 3 to 5 capital letters

Number part (optional) of the digits 0 to 9 supplemented to a maximum of 8 digits in total, without leading zero

What is the Pool ID?

Pool ID
- ☑ Unique designator of an object storage pool(e.g. PERM1)
- ☑ Determined by the system administrator
- ☑ Used to generate the Blob IDs, uses the names of the blob containers and the blob registers

- ☑ Consists of:
  \<Designator\> and \<Number part\> (e.g. A1234500, ABCDEF12 or AB)

PERM1

Designator (obligatory)
of 1 to 6 capital letters

Number part (optional) of the digits 0 to 9 supplemented to a maximum of 8 digits in total, without leading zero

What is a filestream pool?

Filestream-Pool
- ✔ Blobs are stored one after the other in so-called blob containers
- ✔ Individual binary files in these pools can no longer be be deleted
- ✔ The size of the pool depends on the number of
- ✔ of the available blob containers
  (the number of Blob container per pool is arbitrary)

- ✔ Use as storage

Example: WMO123_PERM1_2019010341021.volume
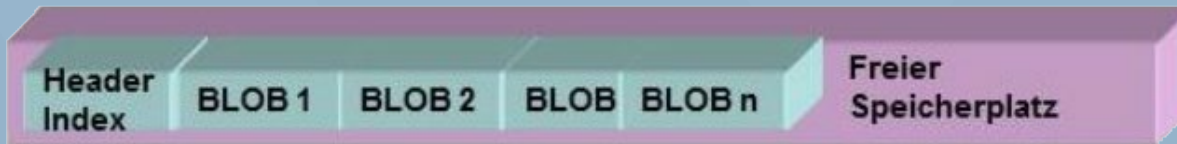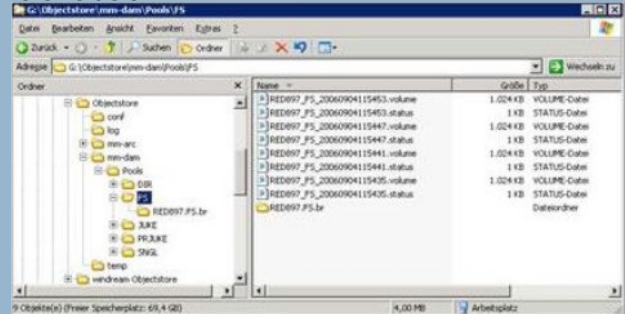
What is a BLOB container?

Blob-Container (Volume)
- ✓ Are in a file (e.g. WMO123_PERM1_2019010341021.volume)
- ✓ Blobs are written to the blob container one after the other

Header Index | Blob 1 | Blob 2 | Blob 3 | Blob n | Freier Speicherplatz

Administrative information

The individual blobs
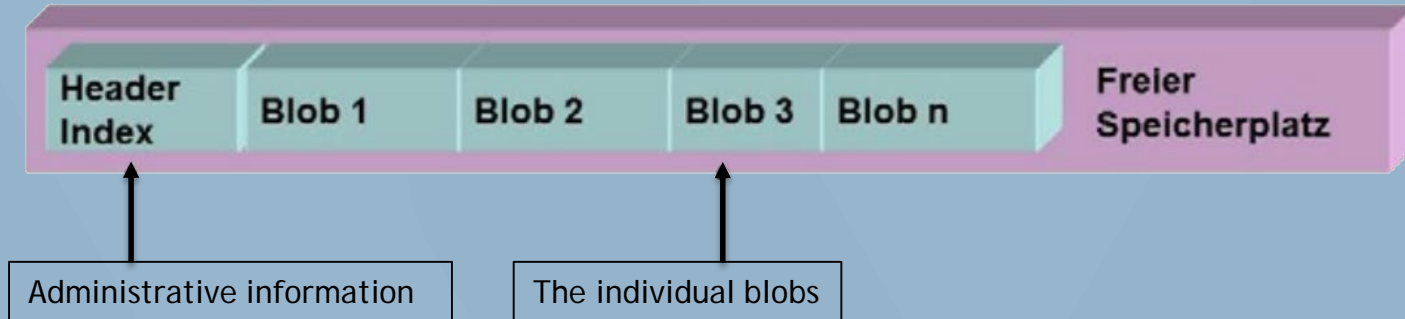
What is the blob register?

Blob-Register
- ✓ One or several files per pool (e.g. WMO123.PERM1.br)
- ✓ Contains the position of the blobs in a pool

- ✓ Non hierarchical:
  for administering up to 32 million blobs per pool
- ✓ Hierarchical:
  for administering up to 2^{40} blobs per pool

Administration information | Positions of the individual blobs in the container
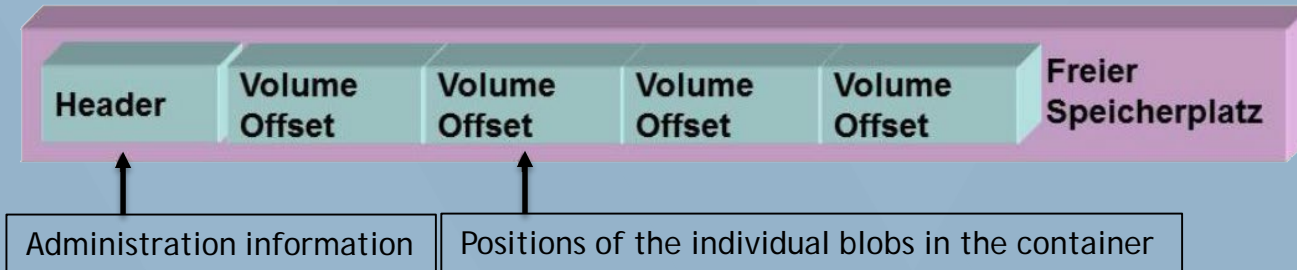
The advantages

Independent of hardware
Pools can be on any medium

Compliance
Comply with legal requirements

Paperless
Meet compliance requirements and accelerate data delivery

Legal opinion Compliance

02 Cloud Archiving

Why Archiving in Cloud Storage?

- ✔ Frequent demand

- ✔ Cost-effective

- ✔ Secure storage guaranteed by Microsoft/AWS

- ✔ Possibility of georedundancy

- ✔ Ideal for Cloud Deployment

- ✔ Many S3 compatible storage systems available (e.g. ECS)

Cloud - Store
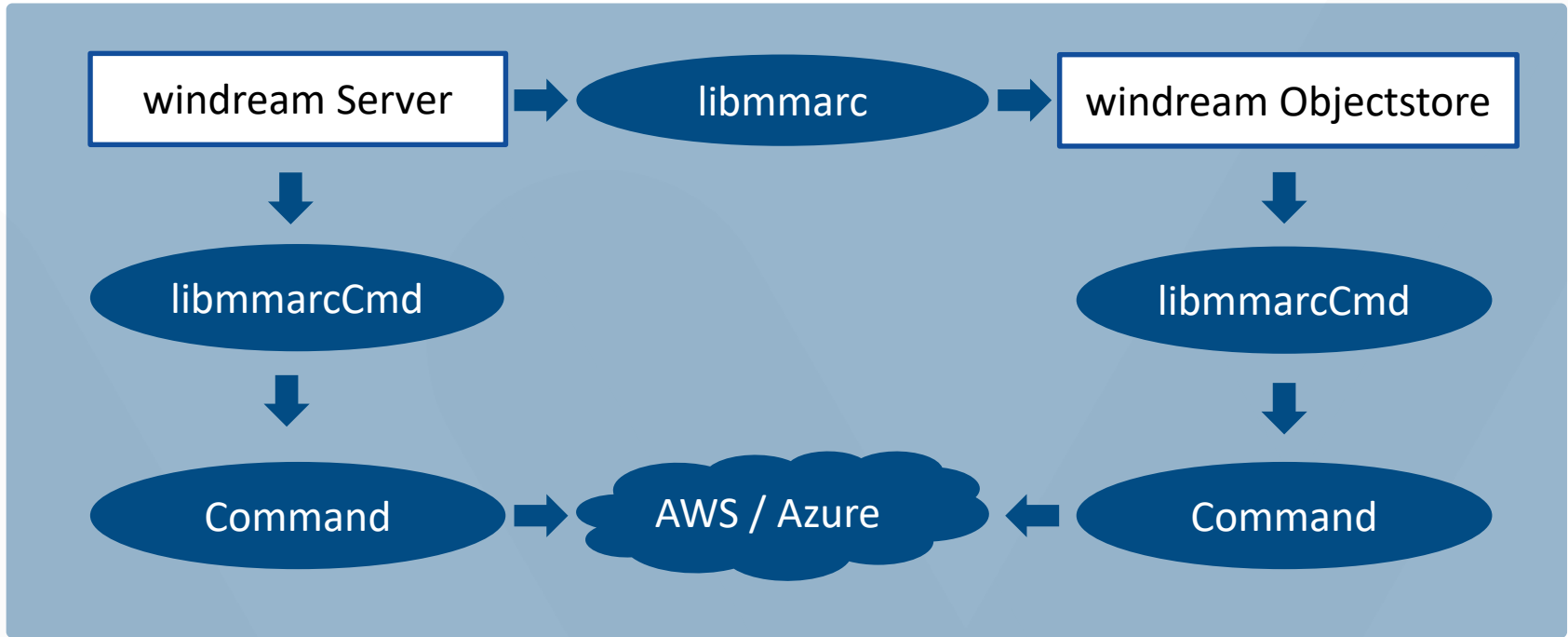
Cloud storage without Objectstore

Structure of the data (here Amazon S3)

Cloud pool via Objectstore
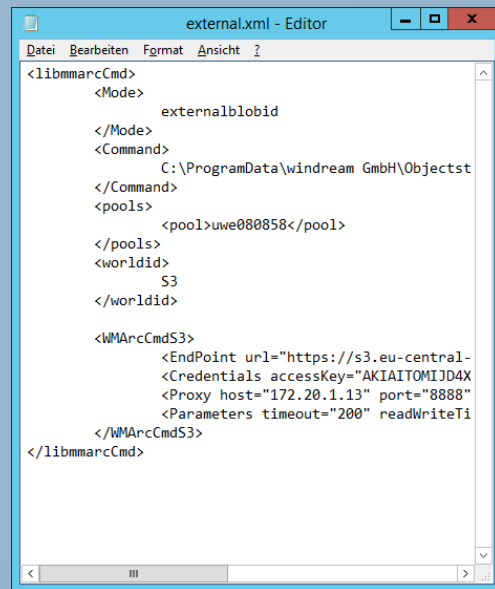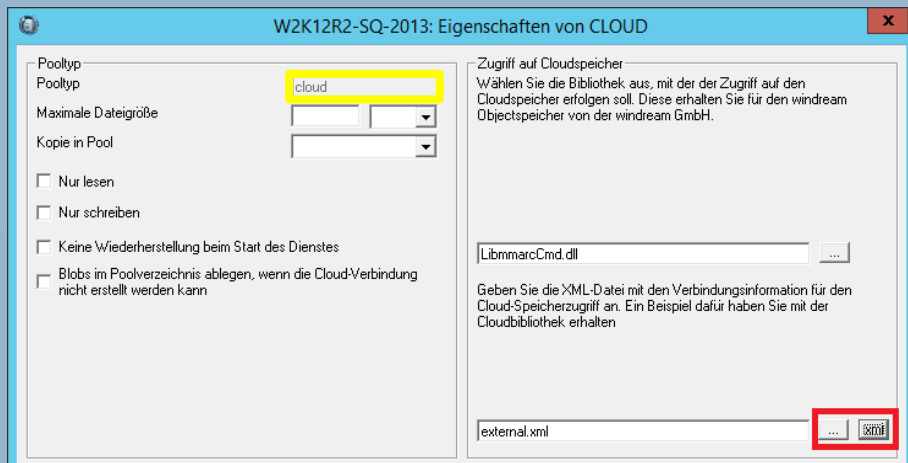
On-Premises Filestream-Pool + Cloud-Copy-Pool

Structure of the data (here Amazon S3)

Control of libmmarcCmd via XML file

General part

- ✓ Mode: externalBlobId or standard

- ✓ Command: Call external application with parameter

- ✓ ProgID: Use COM object In-Process/Out-of-Process

- ✓ Pools: List of pools (simulated)

- ✓ WorldId: simulated WorldId

Control of libmmarcCmd via XML file

Specific part

- ☑ Completely dependent on the Command/ProgID used

- ☑ Contains all necessary parameters for the access

- ☑ E.g.: EndPoint, Credentials, Timeouts, etc.

- ☑ XML file is passed as parameter

- ☑ Thus arbitrarily expandable

Currently available connections

- ☑ Azure BLOB Storage Command

- ☑ Amazon S3 Command

- ☑ Amazon S3 COM-Modul

  - Especially for on-premises or pure AWS solution

windream
MANAGING DOCUMENTS

Noch Fragen?