The Key Features
Every Statistical Computing Environment
Should Have
and Often Lacks.

A Presentation About Almost 20 Years of Experiences with SCEs

Outline

▶ Why this talk?

▶ What is the SCE / What should be part of

▶ The single Parts and what they should be

Why this talk?

- A reflection on nearly 20 years of building and using SCEs.
- PhUSE Presentation:
 - How to exchange SAS programs between a "custom" environment and the official SCE solution from an external vendor
- Why is it required to use an external environment for development?
 - Vendors often don't see the SCEs as a development environment
 - Support to store data and results and execute programs

What is the SCE

- A "window" into the Study Analysis
- Nr. 1 "Application" for Study Analysts and possibly Statisticians
- Provide confidence in Data, Programs and Outputs
- Should be "comfortable" and make some "fun" when using



What should be part of an SCE

Analyst want A Program Editor

Scientist want A Data Browser, Data Access

IT / QA want Data Governance, Security, Audit Trails (Backend)

Managers want Everything from eCRF to Submission and beyond

Nobody seem to want A Navigator / Explorer



```
data step_2;
    set step_1;
    ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
    outvar = max(in_var1, in_var2);
    If (group = 1) THEN DO; %** That's for "Very high value";
        IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base) THEN OUTPUT;
    END;
    ELSE IF (group = 2) THEN DO; %** That's for "High value";
        IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
    END;
    ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
    RUN;</pre>
```

Display and edit program content

```
13
    data step_2;
14
         set step_1;
15
        ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
16
        outvar = max(in_var1, in_var2);
17
         IF (group = 1) THEN DO; %** That's for "Very high value";
             IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base) THEN OUTPUT;
18
19
        END;
20
        ELSE IF (group = 2) THEN DO; *** That's for "High value";
             IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
21
22
        END;
23
        ELSE PUT "ERROR: Invalid value" &subj_var. ** roup=;
24
    RUN;
OF.
```

- Display and edit program content
- Possibly with some nice colours

```
13
      data step_2;
 14
          set step_1;
 15
          ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
          outvar = max(in_var1, in_var2);
 16
          IF (group = 1) THEN DO; %** That's for "Very high value";
 17
               IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base) THEN OUTPUT;
2 18
 19
          END;
  20
          ELSE IF (group = 2) THEN DO; *** That's for "High value";
               IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
  21
  22
          END;
  23
          ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
  24
      RUN;
  OF.
```

- Possibly with some nice colours
- And some syntax checks

```
13
       data step_2;
  14
            set step_1;
            ATTRIB outvar LABEL = "Output Variable" FOR
  15
  16
            outvar = max(in_var1, in_var2);
            IF (group = 1) THEN DO; *** That's for "Ver" 5 for "High
IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base) THEN OUT
  17
2 18
  19
            END;
  20
            ELSE IF (group = 2) THEN DO; *** That's for "High value";
                 IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
  21
  22
            END;
  23
            ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
  24
       RUN;
  OF.
```

- Possibly with some nice colours
- And some syntax checks

```
data step_2;
 19
 20
          set step_1;
 21
          ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
          outvar = max(in_var1, in_var2);
 22
          IF (group = 1) THEN DO; %** Thats for "Very high value";
 23
              IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base)) THEN OUTPUT;
 24
 25
          END;
 26
          ELSE IF (group = 2) THEN DO; *** Thats for "High value";
              IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
 27
 28
          END;
          ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
 29
30
          ENR
      RUN:
```

ERROR: No matching DO/SELECT statement.

- Possibly with some nice colours
- And some syntax checks

```
data step_2;
 13
 14
          set step_1;
 1.5
          ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
 16
          outvar = max(in_var1, in_var2);
          IF (group = 1) THEN DO; %** That's for "Very high value";
3 17
               IF (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base) THEN OUTPUT;
 18
 19
          END;
 20
          ELSE IF (group = 2) THEN DO; *** That's for "High value";
 21
               IF (aval < 15) and (aval <= (0.7 * _base)) THEN OUTPUT;
 22
          END;
 23
          ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
 24
      RUN;
 25
```

- And some syntax checks
- Or even better: correct colours

```
Name: STEP 1
                                Observations: 12005
                                Variables: 61
           data step_2;
     13
                set step_1
     14
     15
                 ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
                 outvar = _max(in_var1, in_var2);
      16
                                   THEN DO; *** That is for "Very high value";
Call: MAX(value1 <, value2,...>)
                          (15 <= aval) and (aval < 450) and (aval <= (0.8 * _base)) THEN OUTPUT;</p>
Name: MAX
Category: Descriptive Statistics
Description: Returns the largest value.
                                        Name: aval
                                                       *** That is for "High value";
                           (group
                                        Label: Analysis Value
                                                         (0.7 * _base)) THEN OUTPUT;
      21
                      IF (aval < 15)</pre>
                                        Type: NUM
      22
                 END;
                                        Format: 8.2
                 ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
      23
      24
           RUN;
                                                                    &subj var.=USUBJID
```

And some explanations

```
32
    PROC SORT DATA=input_1; BY var1 var2;
33
                                             RUN;
     PROC SORT DATA=input_2; BY
                                 var1 var2;
                                             RUN;
34
35
     DATA output;
         MERGE input_1 input_2;
36
         BY var1 var2;
37
38
     RUN;
20
```

Some Templates for repetitive code

```
data step_2;
 13
 14
           set step_1;
 15
          ATTRIB outvar LABEL = "Output Variable" FORMAT=8.2;
i 16
           outvar = in_var1 / in_var2;
           IF (group = 1) THE DO; *** Thats for "Very high value";
 18
               IF (15 <= aval ⅓)
                                                                        base)) THEN OUTPUT;
                                 NOTE: Division by zero detected.
 19
          END;
          ELSE IF (group = 2)
 20
               IF (aval < 15) a</pre>
          END;
           ELSE PUT "ERROR: Invalid value" &subj_var.= group=;
 24
      RUN;
 25
```

▶ Run code interactive and retrieve the log at the correct location.

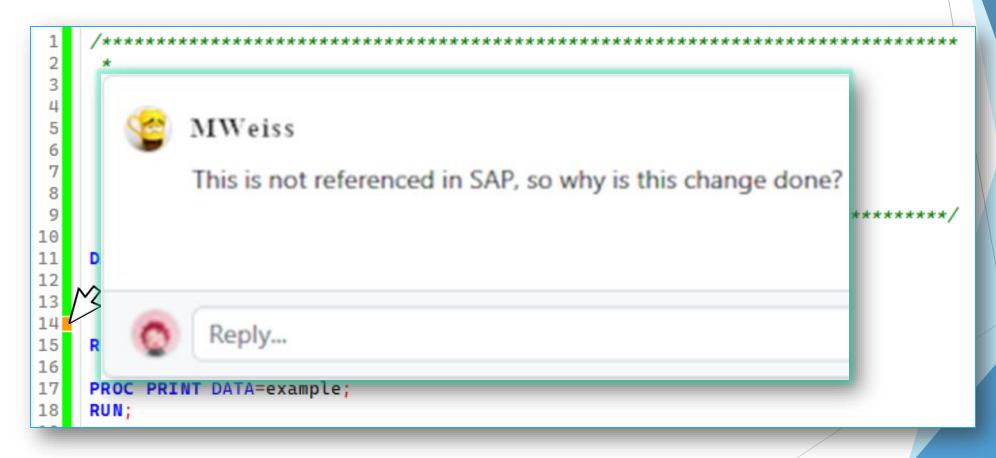
Show what changed for validation

```
2
 3
      * Example Program Header
      * @purpose This program hosws how the git blame feature looks
      * @author Michael Weiss
      * @date 2025-10-12
 9
10
    DATA example;
12
        SET sashelp.class;
13
        IF age > 14;
        KEEP name age sex weight height;
14
15
    RUN;
16
    PROC PRINT DATA=example;
    RUN;
```

Show what changed for validation

```
2
     * Example Program Header
     * @purpose This program hosws how the git blame feature looks
     * @author Michael Weiss
     * @date 2025-10-12
9
10
                          IF age
12
13
14
                          KEEP
15
16
                 RUN;
```

Allow to add comments



A simple Table that displays the Data

	STUDYID	SUBJID	USUBJID	TRTP	TRTPN	TRTA
1	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
2	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
3	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
4	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
5	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
6	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
7	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
8	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
9	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo

- A simple Table that displays the Data
- And should display labels

	STUDYID	SUBJID	USUBJID	TRTP	TRTPN	TRTA
	Study	Subject	Unique	Planned	Planned	Actual
	ldentifier	Identifier for	Subject	Treatment	Treatment	Treatment
		the Study	ldentifier		(N)	
1	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
2	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
3	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
4	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
5	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
6	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
7	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
8	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
9	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo

A simple Table that displays the Data

And

With look

	STUDYID	SUBJID	USUBJID	TRTP	TRTPN	TRTA
	Study	Subject	Unique	Planned	Planned	Actual
	Identifier	Identifier for the Study	Subject Identifier	Treatmen	t Treatment (N)	Treatment
1	CDISCPILO	1015	01-701-1015	Placebo	0	Placebo
2	CDISCPII O	1015	01-701-1015	Placebo	0	Placebo
					0	Placebo
					0	Placebo
					0	Placebo
					0	Placebo
					0	Placebo

Placebo

Placebo

- A simple Table that displays the Data
- And should display labels
- With some Highlighting
- Allow rearrange columns



	PARAM	AVISIT	AVAL	STUDYID	SUBJID
	Parameter	Analysis Visit	Analysis	Study	Subject
	A		Value	ldentifier	Identifier for
					the Study
	`				
1	Alkaline Phosphatase (Baseline	34	CDISCPILO	1015
2	Protein (g/L)	Baseline	61	CDISCPILO	1015
3	Sodium (mmol/L)	Week 12	139	CDISCPILO	1015
4	Calcium (mmol/L)	Week 20	2.17	CDISCPILO	1015
5	Sodium (mmol/L)	Baseline	140	CDISCPILO	1015
6	Sodium (mmol/L)	Week 4	140	CDISCPILO	1015
7	Sodium (mmol/L)	Week 6	140	CDISCPILO	1015
8	Sodium (mmol/L)	Week 24	140	CDISCPILO	1015
9	Sodium (mmol/L)	End of Treat	140	CDISCPILO	1015

- A simple Table that displays the Data
- And should display labels
- With some Highlighting
- Allow rearrange columns
- And hide columns



	PARAM	AVISIT	AVAL	STUDYID	SUBJID
	Parameter	Analysis Visit	Analysis	Study	Subject
			Value	ldentifier	Identifier for
					the Study
1	Alkaline Phosphatase (Baseline	34	CDISCPILO	1015
2	Protein (g/L)	Baseline	61	CDISCPILO	1015
3	Sodium (mmol/L)	Week 12	139	CDISCPILO	1015
4	Calcium (mmol/L)	Week 20	2.17	CDISCPILO	1015
5	Sodium (mmol/L)	Baseline	140	CDISCPILO	1015
6	Sodium (mmol/L)	Week 4	140	CDISCPILO	1015
7	Sodium (mmol/L)	Week 6	140	CDISCPILO	1015
8	Sodium (mmol/L)	Week 24	140	CDISCPILO	1015
9	Sodium (mmol/L)	End of Treat	140	CDISCPILO	1015

- A simple Table that displays the Data
- And should display labels
- With some Highlighting
- Allow rearrange columns
- And hide columns
- And filtered for PARAMCD == "BILI"



		PARAM Parameter	AVISIT Analysis Visit	AVAL Analysis Value	STUDYID Study Identifier	SUBJID Subject Identifier for the Study
I	110	Bilirubin (umol/L)	Week 6	5.13	CDISCPILO	1015
ı	111	Bilirubin (umol/L)	Week 26	5.13	CDISCPILO	1015
ı	142	Bilirubin (umol/L)	Week 4	6.84	CDISCPILO	1015
ı	143	Bilirubin (umol/L)	Week 12	6.84	CDISCPILO	1015
ı	144	Bilirubin (umol/L)	Week 16	6.84	CDISCPILO	1015
ı	145	Bilirubin (umol/L)	Week 20	6.84	CDISCPILO	1015
ı	146	Bilirubin (umol/L)	Week 24	6.84	CDISCPILO	1015
-	14/	Bilirubin (umol/L)	End of Treat	6.84	CDISCPILO	1015
	169	Bilirubin (umol/L)	Week 2	8.55	CDISCPILO	1015

- A simple Table that displays the Data
- And should display labels
- With some Highlighting
- Allow rearrange columns
- And hide columns
- And filtered for PARAMCD == "BILI"
- And sorted by AVISIT

169 Bilirubin (umol/L) Week 2 142 Bilirubin (umol/L) Week 4	the Study
1/12 Rilirubin (umol/L) Week /	2 8.55 CDISCPILO 1015
142 Dilliabili (dilloi/E)	4 6.84 CDISCPILO 1015
110 Bilirubin (umol/L) Week 6	6 5.13 CDISCPILO 1015
170 Bilirubin (umol/L) Week 8	8 8.55 CDISCPILO 1015
143 Bilirubin (umol/L) Week 1	12 6.84 CDISCPILO 1015
144 Bilirubin (umol/L) Week 1	16 6.84 CDISCPILO 1015
145 Bilirubin (umol/L) Week 2	20 6.84 CDISCPILO 1015
146 Bilirubin (umol/L) Week 2	24 6.84 CDISCPILO 1015
111 Bilirubin (umol/L) Week 2	26 5.13 CDISCPILO 1015



Count the number of subjects

	▲ USUBJID	PARAM
	Unique	Parameter
	Subject Identifier	
	Identifier	
1	01010101010	Alkaline Phosphatase (U/L)
396	01010101010	Creatine Kinase (U/L) chang
397	01010101011	Sodium (mmol/L)
401	01010101012	Albumin (g/L)
541	01010101012	Calcium (mmol/L)
548	01010101012	Cholesterol (mmol/L)
937	01010101012	Calcium (mmol/L)
945	01010101033	Albumin (g/L)
952	01011001033	Potassium (mmol/L)
1361	01011001034	Urate (umol/L) change fro
1362	01011001034	Phosphate (mmol/L) chang
1363	01011001035	Calcium (mmol/L) change f
270	01011001036	Creatine Kinase (U/L) chang
273	01011001036	Chloride (mmol/L) change
279	01011001037	Blood Urea Nitrogen (mmol
299	01011001037	Urate (umol/L) change fro

Count again



▲ USUBJID	PARAM
Unique	Parameter
•	
Identifier	
01010101010	Alkaline Phosphatase (U/L)
01010101010	Creatine Kinase (U/L) chang
01010101011	Sodium (mmol/L)
01010101012	Albumin (g/L)
01010101012	Calcium (mmol/L)
01010101012	Cholesterol (mmol/L)
01010101012	Calcium (mmol/L)
01010101033	Albumin (g/L)
01011001033	Potassium (mmol/L)
01011001034	Urate (umol/L) change fro
01011001034	Phosphate (mmol/L) chang
01011001035	Calcium (mmol/L) change f
01011001036	Creatine Kinase (U/L) chang
01011001036	Chloride (mmol/L) change
01011001037	Blood Urea Nitrogen (mmol
01011001037	Urate (umol/L) change fro
	Unique Subject Identifier 01010101010 010101010101 01010101011 01010101

► Count again → 9 Subjects



▲ USUBJID	PARAM
Unique	Parameter
•	
Identifier	
01010101010	Alkaline Phosphatase (U/L)
01010101010	Creatine Kinase (U/L) chang
01010101011	Sodium (mmol/L)
01010101012	Albumin (g/L)
01010101012	Calcium (mmol/L)
01010101012	Cholesterol (mmol/L)
01010101012	Calcium (mmol/L)
010101033	Albumin (g/L)
01011001033	Potassium (mmol/L)
01011001034	Urate (umol/L) change fro
01011001034	Phosphate (mmol/L) chang
01011001035	Calcium (mmol/L) change f
01011001036	Creatine Kinase (U/L) chang
01011001036	Chloride (mmol/L) change
01011001037	Blood Urea Nitrogen (mmol
01011001037	Urate (umol/L) change fro
	Unique Subject Identifier 01010101010 0101010101010 0101010101

The Backend

- Access Restrictions (Security)
 - Who is allowed to access
 - ▶ Who is allowed to grant access (Self Assign?)
- Audit Trail
 - Who had access
 - Who used access
- Lifecycle Management
 - Status of files (development, validation, production)
 - How and by whom a file was created / changed
 - ▶ (e.g. manually, by program, by interface)





The Backend

- Delayed Execution
 - Timed / Scheduled
 - On Input Changed (smart ...)
- Automatic Tracing of Program Flow
 - ► Input → Program → Output
 - ► [SDTM.LB] \rightarrow d_adlb.sas \rightarrow [ADAM.ADLB] \rightarrow t_adlb_1.sas \rightarrow t_adlb_1.rtf
 - ► Marking of "outdated" output and programs to be rerun
 - Marking of outputs with lifecycle information based on tracing

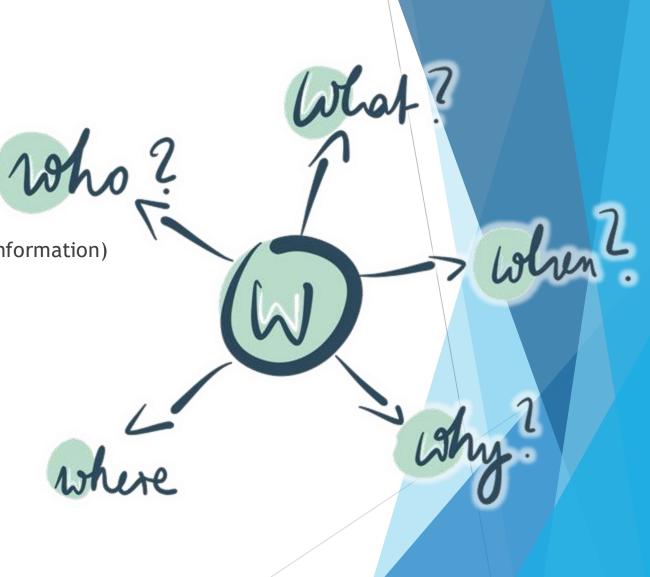
The Backend

Access to Data via API

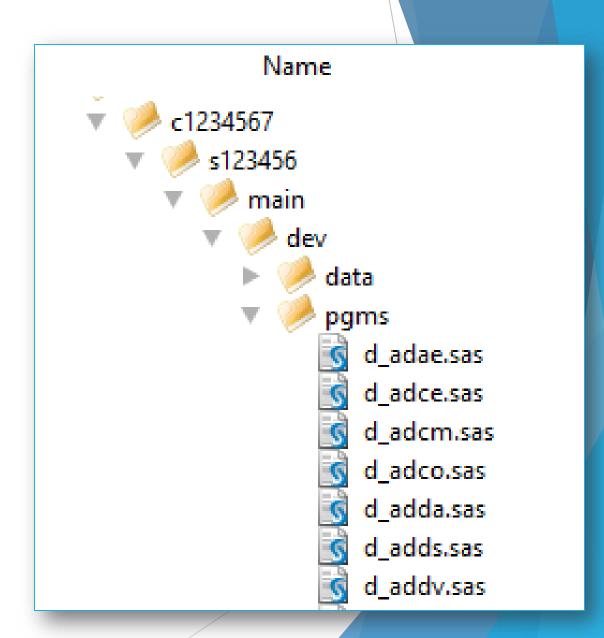
► HTTP(s) based e.g. REST or GraphQL

► "Data Sets" and SCE metadata (e.g. lifecycle information)

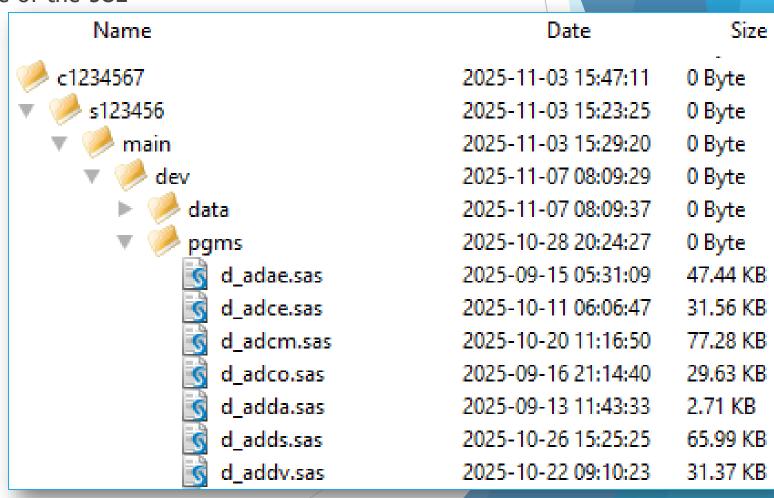
Access to Files via API and Windows Explorer



Provide access to the Files inside of the SCE



- Provide access to the Files inside of the SCE
- Display additional Information
 - Size, Date
- Missing:
 - Lifecycle (dev, qc, prod)
 - Author and Writer
 - ► Errors/ Warnings in last exec

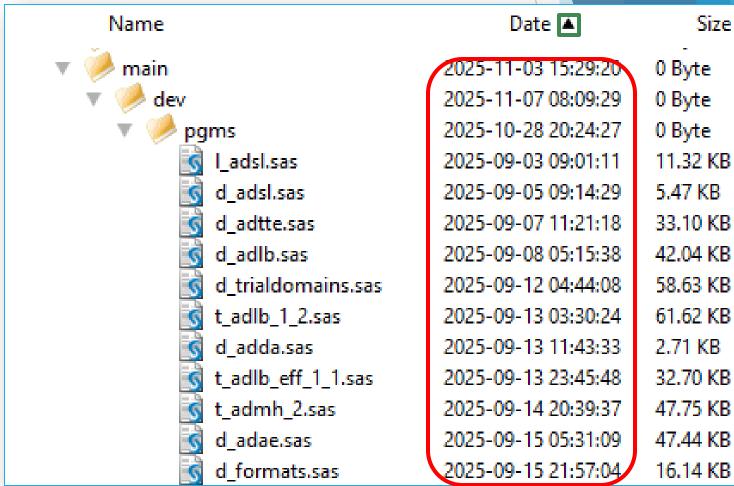


Provide access to the Files inside of the SCE

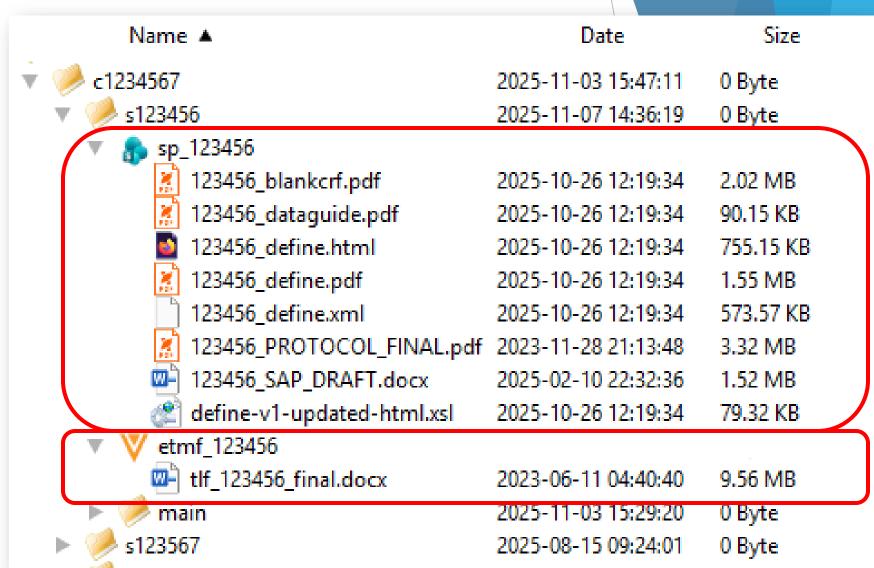
Display additional Information

▶ Date, Time, Lifecycle ...

Allow Sorting



- Integrate other Systems
 - SharePoint
 - eTMF
 - Data Lakehouse



Al & LLM in SCEs

- Integration of LLM in SCE
- Automatically supply study information to LLM (Context)
 - Study Metadata
 - ► SAP / Protocol
- LLM can generate code
- Review / Validation better not by LLM
 - ▶ LLMs not always right
 - ▶ They usually don't ask for "help" but "prefer" to create a wrong answer first
 - ▶ 90% right is 10% wrong. How to find the wrong 10%?
 - Semantic errors hard to fix by LLM

I'm 90 % sure this is correct

That's it

Thanks for listening!

Everything shown here is not just theory - it has been built, used and proven in real SCE.



Now it's time for your questions!

Interested? Ideas? Questions? Contact:

michael.weiss@code2be.de

