

# Inlining Control-Flow Jumps in Library Usage Graphs of Legacy Code

## Presentation Abstract

Ruben Opdebeeck

*ropdebee@vub.be*

Johan Fabry

*johan@raincode.com*

Coen De Roover

*cderoove@vub.be*

# COBOL Library Usage Pattern Mining

```

IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
IDENTIFICATION DIVISION.
PROGRAM-ID. ACCMENU.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==T==.
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

```

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

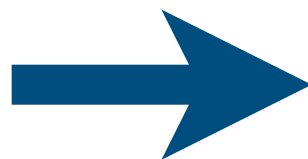
**Goal: Understand library usages  
to estimate modernisation effort**

# COBOL Library Usage Pattern Mining

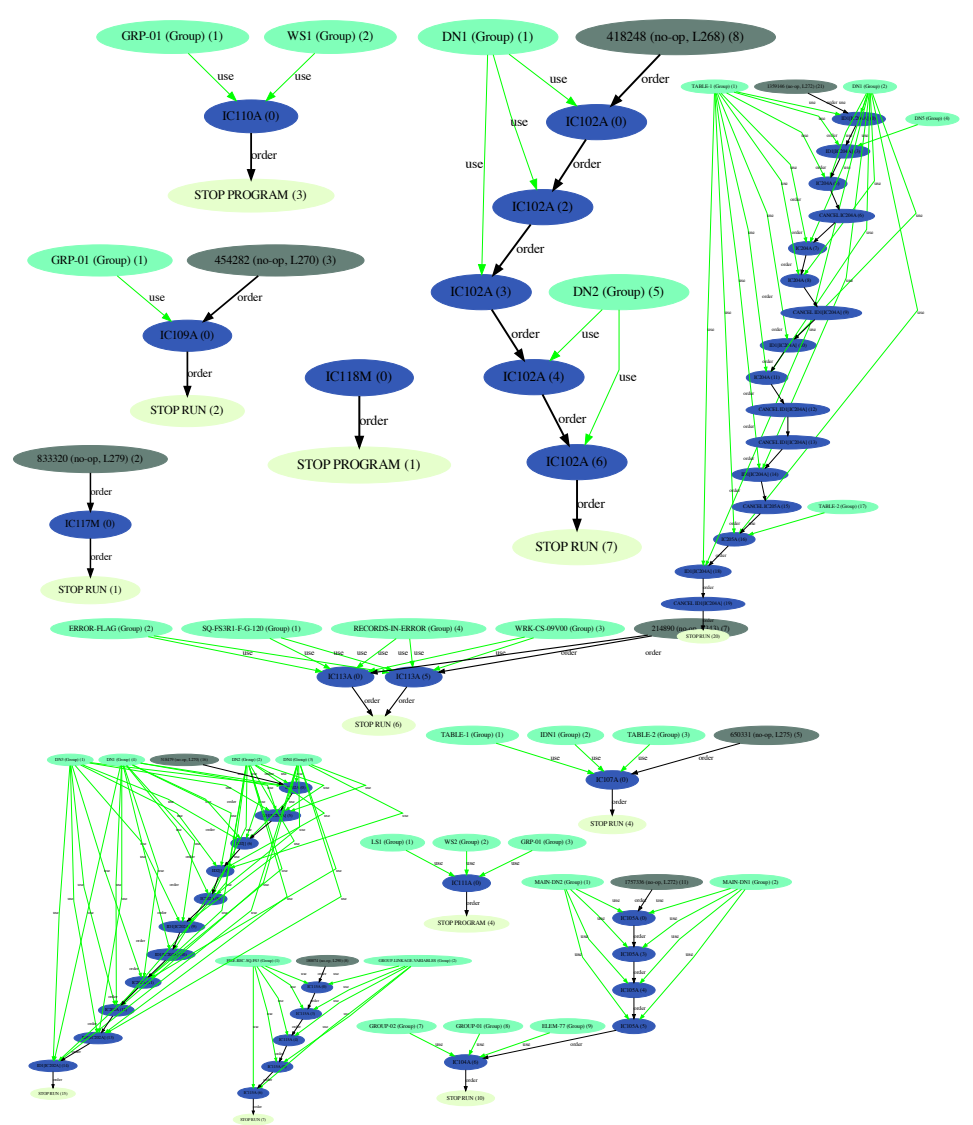
```

IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
IDENTIFICATION DIVISION.
PROGRAM-ID. ACCMENU.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==T==.
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

```



## Library Usages



Case	Time (s)	# Groups	Avg. size	Max. size
Case 1	43.8	273	244	4634
Case 2 (full)	1019	3925	4769	1217422
Case 2 (limited)	N/A	3573	125	9644

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

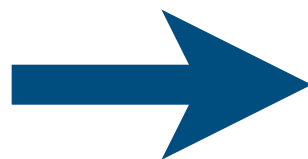
**Goal: Understand library usages to estimate modernisation effort**

# COBOL Library Usage Pattern Mining

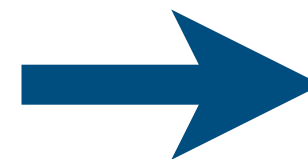
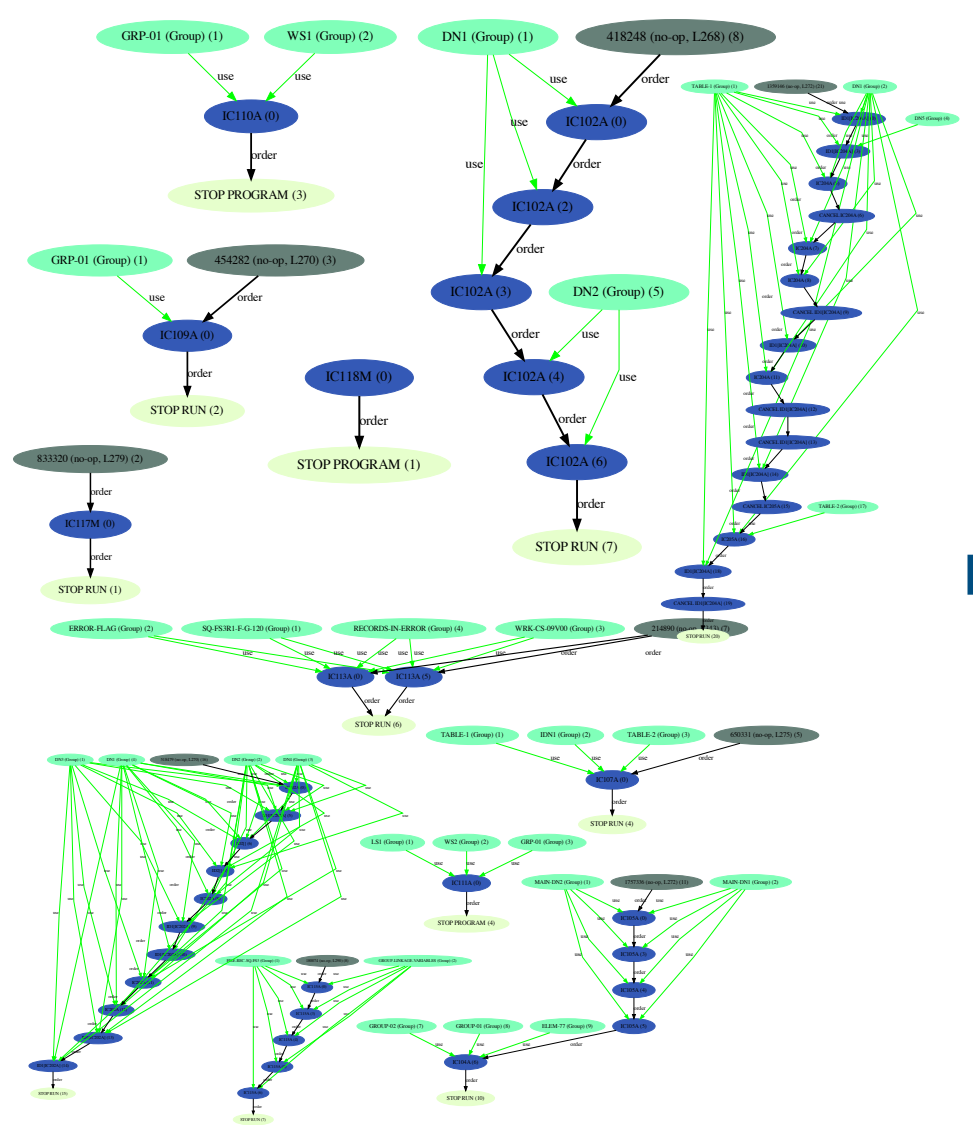
```

IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==T==.
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

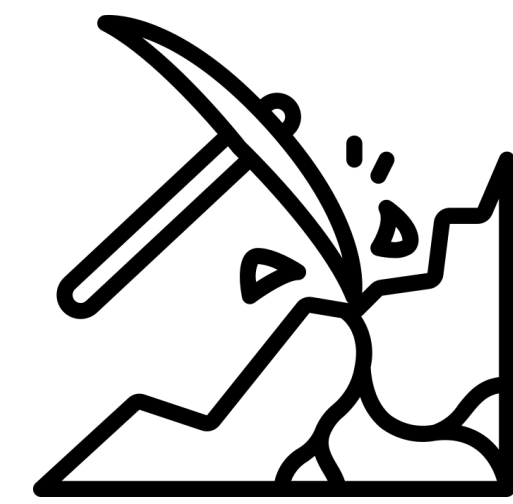
```



## Library Usages



## Pattern Mining



Case	Time (s)	# Groups	Avg. size	Max. size
Case 1	43.8	273	244	4634
Case 2 (full)	1019	3925	4769	1217422
Case 2 (limited)	N/A	3573	125	9644

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

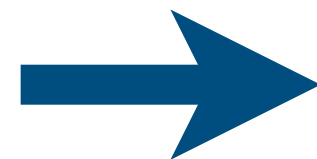
**Goal: Understand library usages to estimate modernisation effort**

# COBOL Library Usage Pattern Mining

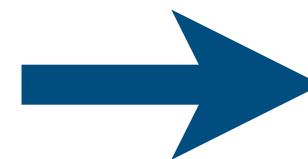
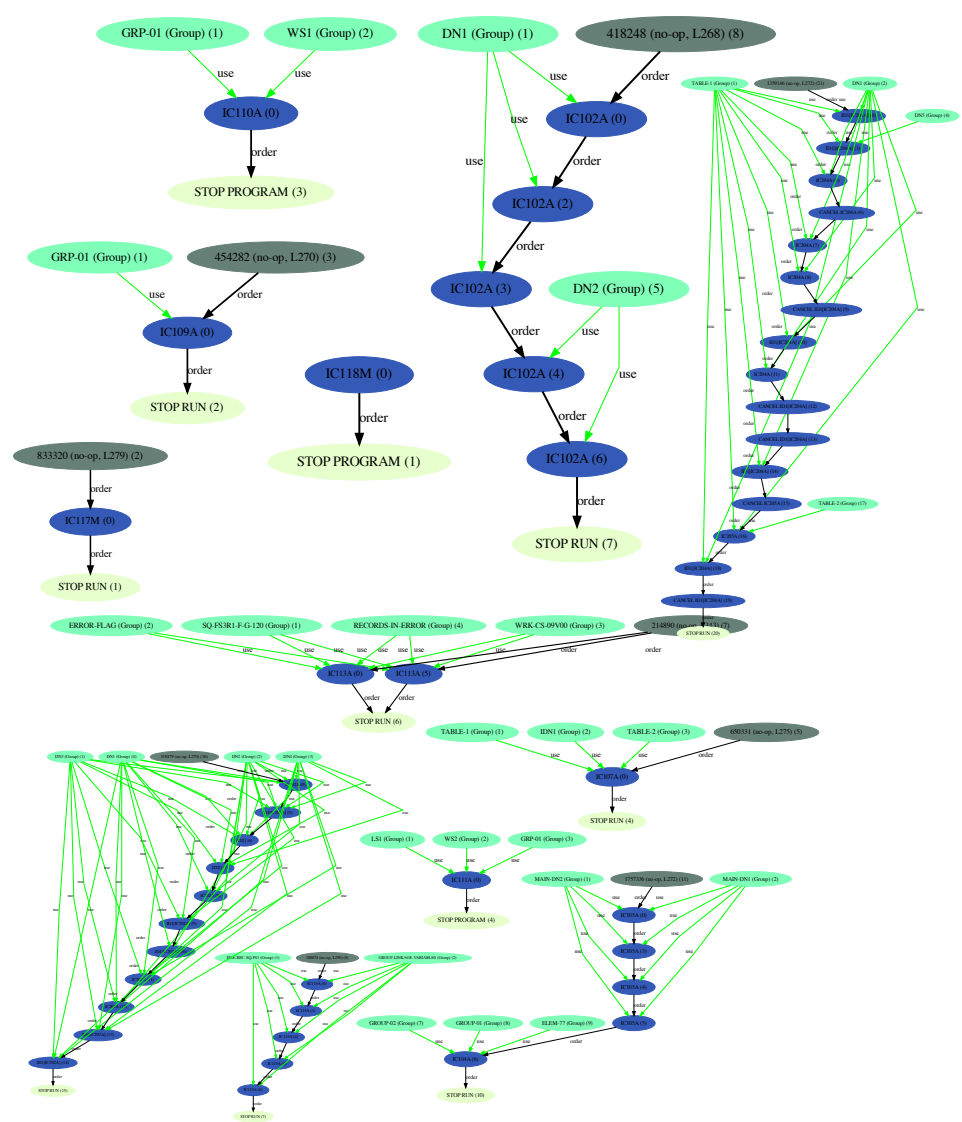
```

IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY =T=..
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

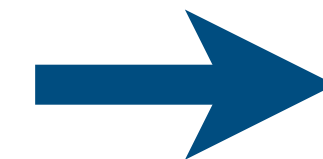
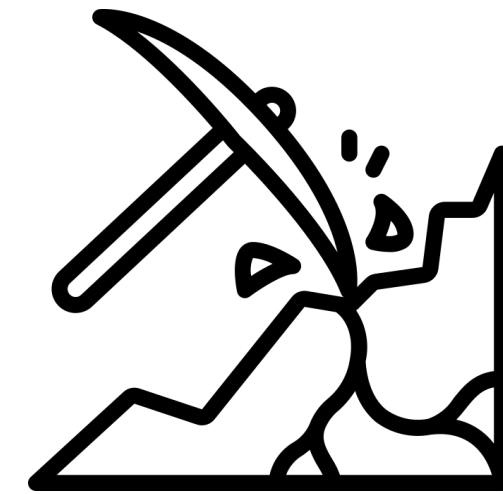
```



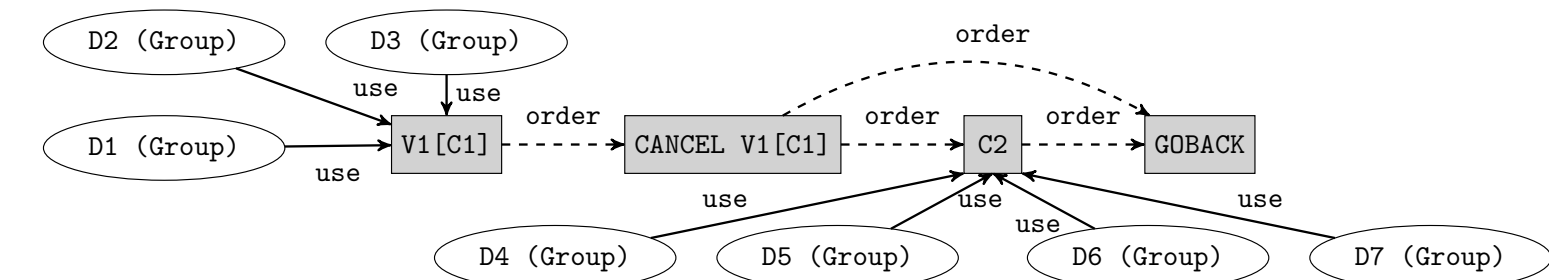
## Library Usages



## Pattern Mining



## Library Usage Patterns



Case	Time (s)	# Groups	Avg. size	Max. size
Case 1	43.8	273	244	4634
Case 2 (full)	1019	3925	4769	1217422
Case 2 (limited)	N/A	3573	125	9644

Corpus	GrouMiner*		BigGroum		
	Time (min)	# P	Time (min)	# Part.	# P
Corpus 1	0.95	3	10.4	3	4
Corpus 2	121.8	7	618.2	1	0

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

**Goal: Understand library usages to estimate modernisation effort**

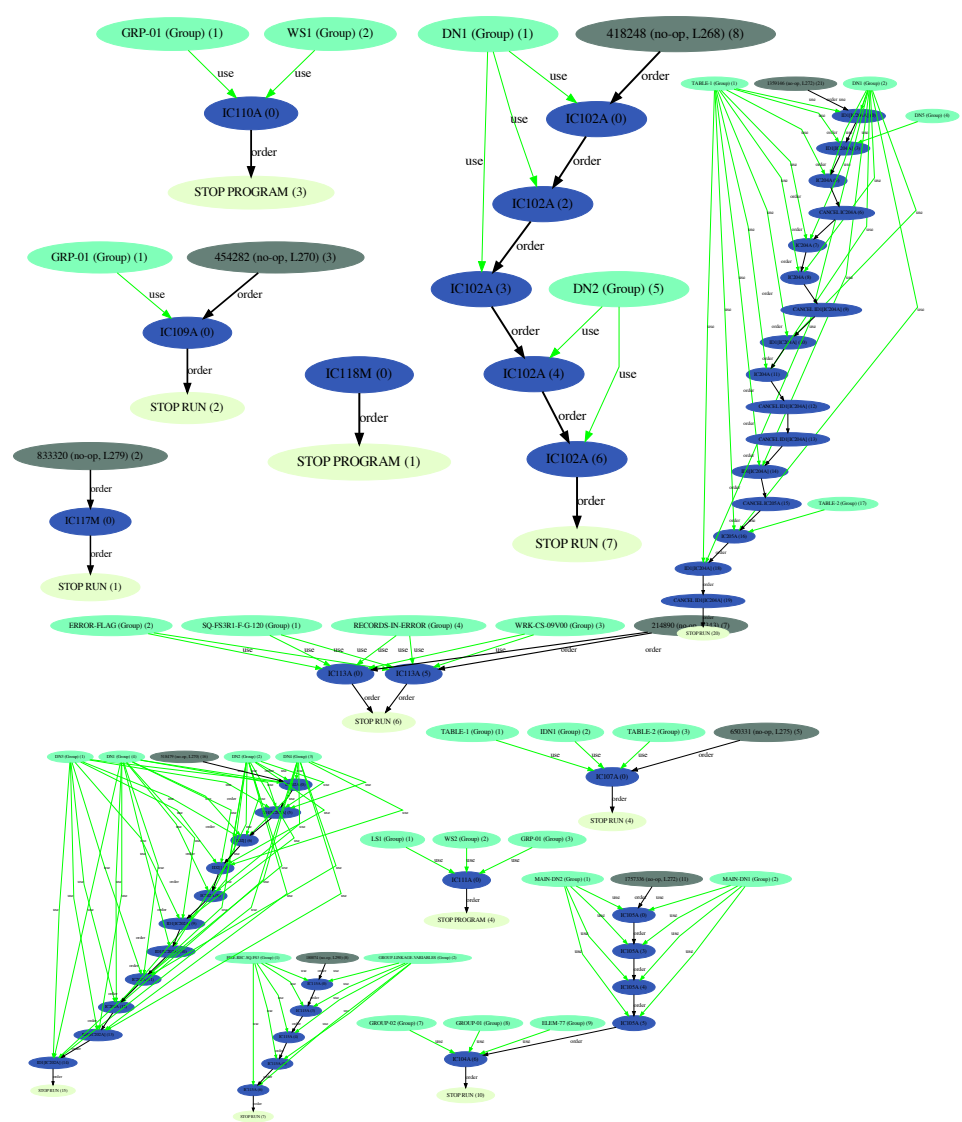
# COBOL Library Usage Pattern Mining

```

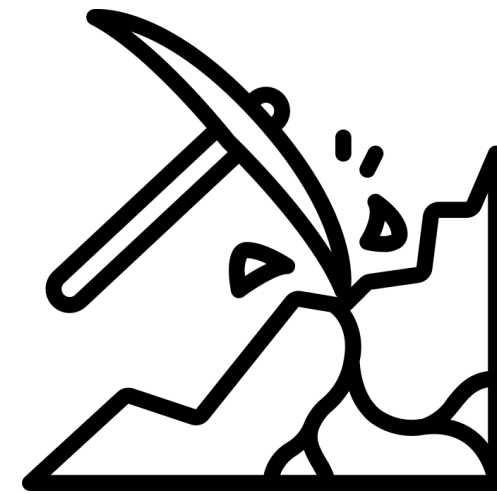
IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==T==.
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

```

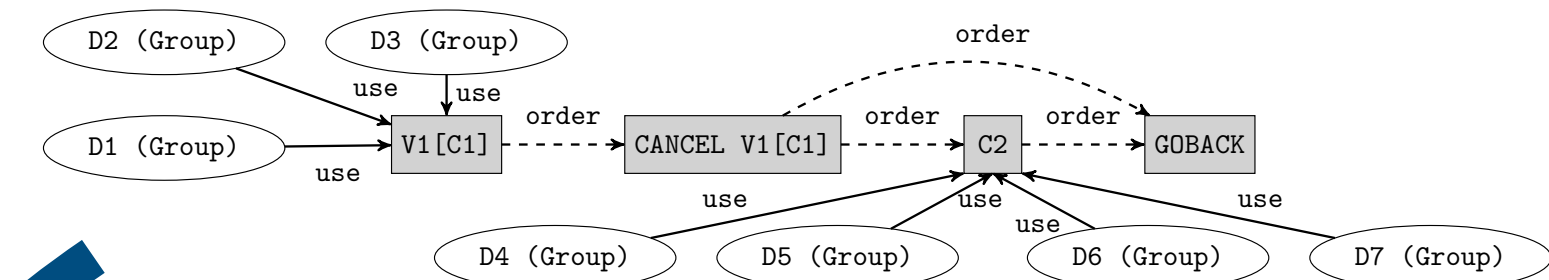
## Library Usages



## Pattern Mining



## Library Usage Patterns



Case	Time (s)	# Groups	Avg. size	Max. size
Case 1	43.8	273	244	4634
Case 2 (full)	1019	3925	4769	1217422
Case 2 (limited)	N/A	3573	125	9644

Corpus	GrouMiner*		BigGroum		
	Time (min)	# P	Time (min)	# Part.	# P
Corpus 1	0.95	3	10.4	3	4
Corpus 2	121.8	7	618.2	1	0

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

**Goal: Understand library usages to estimate modernisation effort**

# COBOL Library Usage Pattern Mining

```

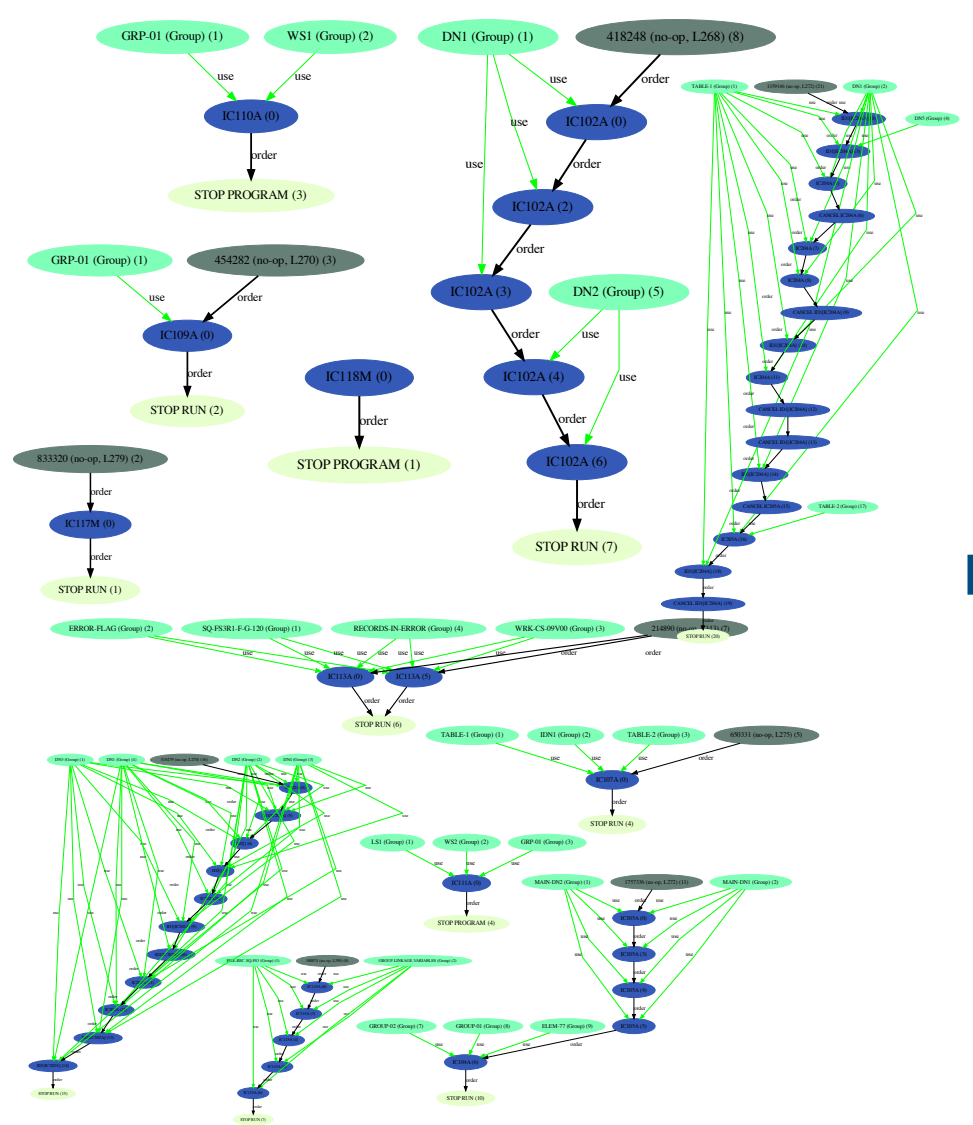
IDENTIFICATION DIVISION.
PROGRAM-ID. BARTHOL.
ENVIRONMENT DIVISION.
CONFIGURATION SECTION.
DATA DIVISION.
WORKING-STORAGE SECTION.
COPY TRANCTRL.
COPY ACCCTRL.
COPY SCREENIOV.
COPY ACCOUNT REPLACING ::=PREFIX::= BY ==.
COPY ACCOUNT REPLACING ::=PREFIX::= BY =T=..
COPY TRANL REPLACING ::=PREFIX::= BY ==.
COPY TRAN REPLACING ::=PREFIX::= BY ==.
01 M-MENU.
05 M-SELECTION PIC X(1) VALUE SPACES.
05 M-TITLE PIC X(40) VALUE "Barthol Bank - Account
menu".
05 M-OPTION.
10 M-OPT1
15 M-OPT1-CODE PIC X VALUE 'C'.
15 M-OPT1-TEXT PIC X(40) VALUE "Create
account".
10 M-OPT2
15 M-OPT1-CODE PIC X VALUE 'S'.
15 M-OPT1-TEXT PIC X(40) VALUE "Select
account".
10 M-OPT3
15 M-OPT1-CODE PIC X VALUE 'E'.
15 M-OPT1-TEXT PIC X(40) VALUE "Edit account".
10 M-OPT35
15 M-OPT1-CODE PIC X VALUE 'D'.
15 M-OPT1-TEXT PIC X(40) VALUE "Delete
account".
10 M-OPT4
15 M-OPT1-CODE PIC X VALUE 'H'.
15 M-OPT1-TEXT PIC X(40) VALUE "Account
history".
10 M-OPT5
15 M-OPT1-CODE PIC X VALUE 'L'.
15 M-OPT1-TEXT PIC X(40) VALUE "List bank
status".
10 M-OPT6
15 M-OPT1-CODE PIC X VALUE SPACES.
15 M-OPT1-TEXT PIC X(40) VALUE SPACES.
10 M-OPT7
15 M-OPT1-CODE PIC X VALUE 'Q'.
15 M-OPT1-TEXT PIC X(40) VALUE "Quit menu".
10 M-OPT99
15 M-OPT1-CODE PIC X VALUE LOW-VALUE.
15 M-OPT1-TEXT PIC X(40) VALUE LOW-VALUE.
01 WRK-VARS.
05 WRK-INPUT-VAR PIC X(10).
05 W-ACC-ID PIC 9(5).
05 W-AMOUNT PIC -ZZZ9.99.
PROCEDURE DIVISION.
PERFORM INIT-WORK.
PERFORM MAIN-LOOP.
GOBACK.
INIT-WORK.

```

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

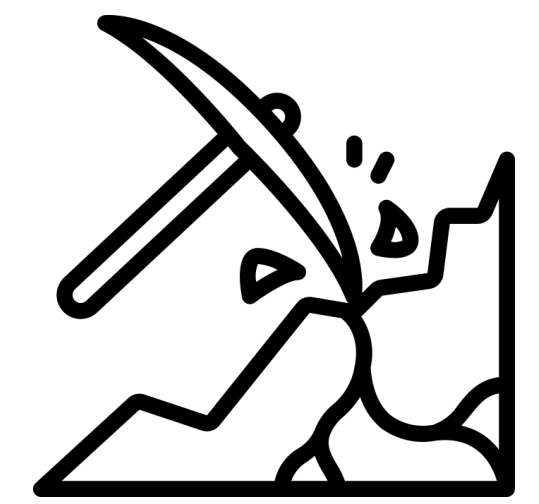
Today

## Library Usages

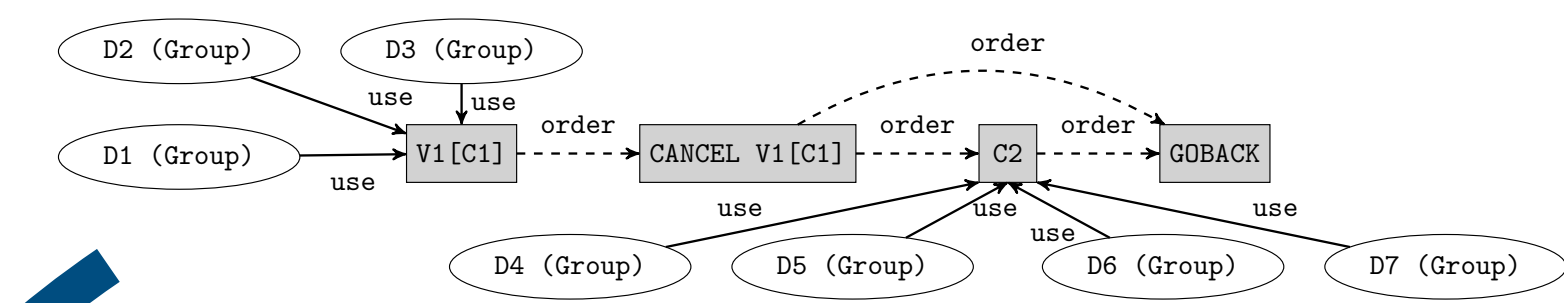


Case	Time (s)	# Groups	Avg. size	Max. size
Case 1	43.8	273	244	4634
Case 2 (full)	1019	3925	4769	1217422
Case 2 (limited)	N/A	3573	125	9644

## Pattern Mining



## Library Usage Patterns



Corpus	GrouMiner*		BigGroum		
	Time (min)	# P	Time (min)	# Part.	# P
Corpus 1	0.95	3	10.4	3	4
Corpus 2	121.8	7	618.2	1	0

**Goal: Understand library usages to estimate modernisation effort**

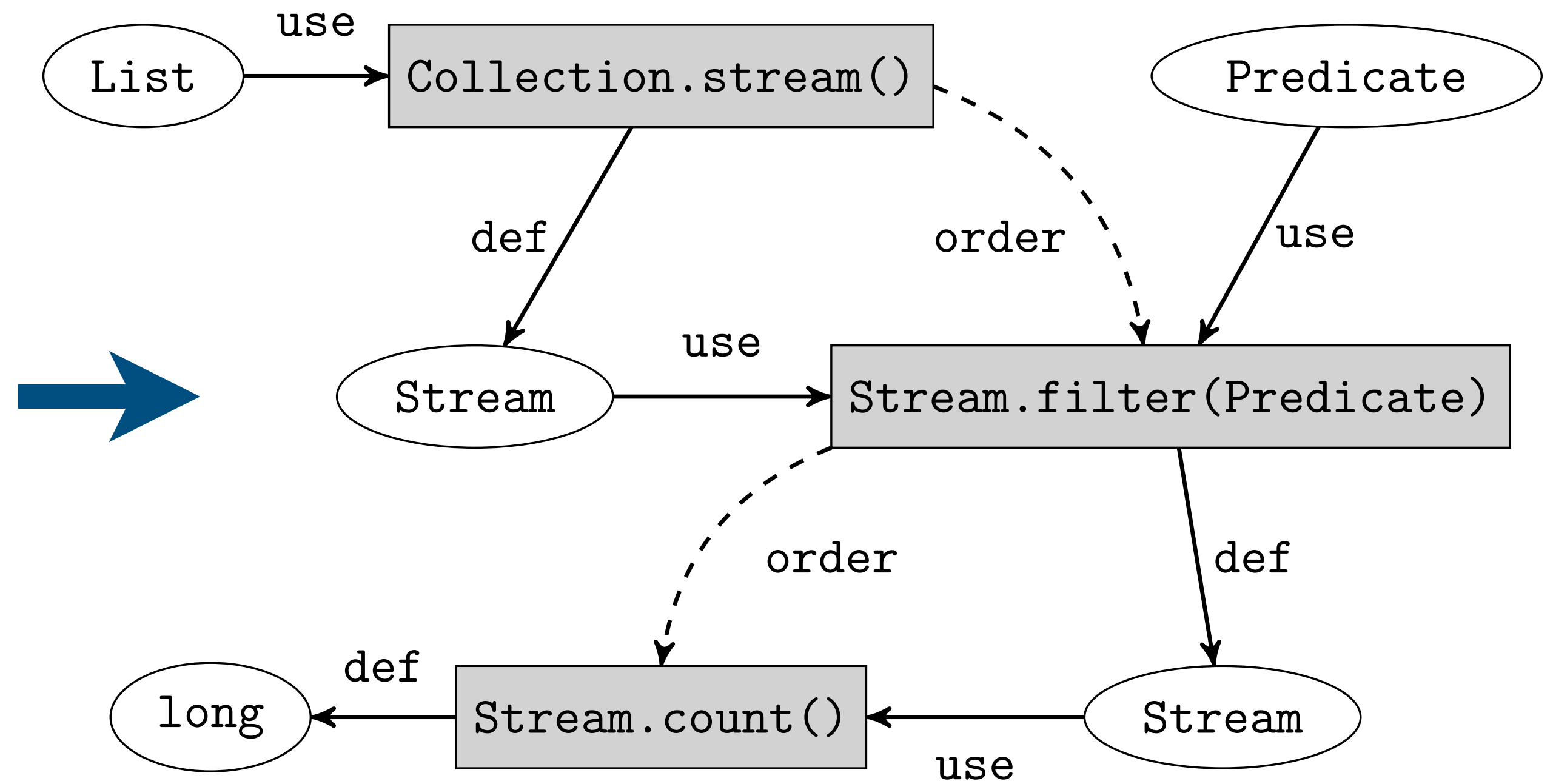
# Groups for OO languages

```

void my_method(List<String> lst) {
  return lst.stream()
    .filter(s -> s.startsWith("test_"))
    .count();
}

```

Java snippet using Collection API



Graph-based object usage model



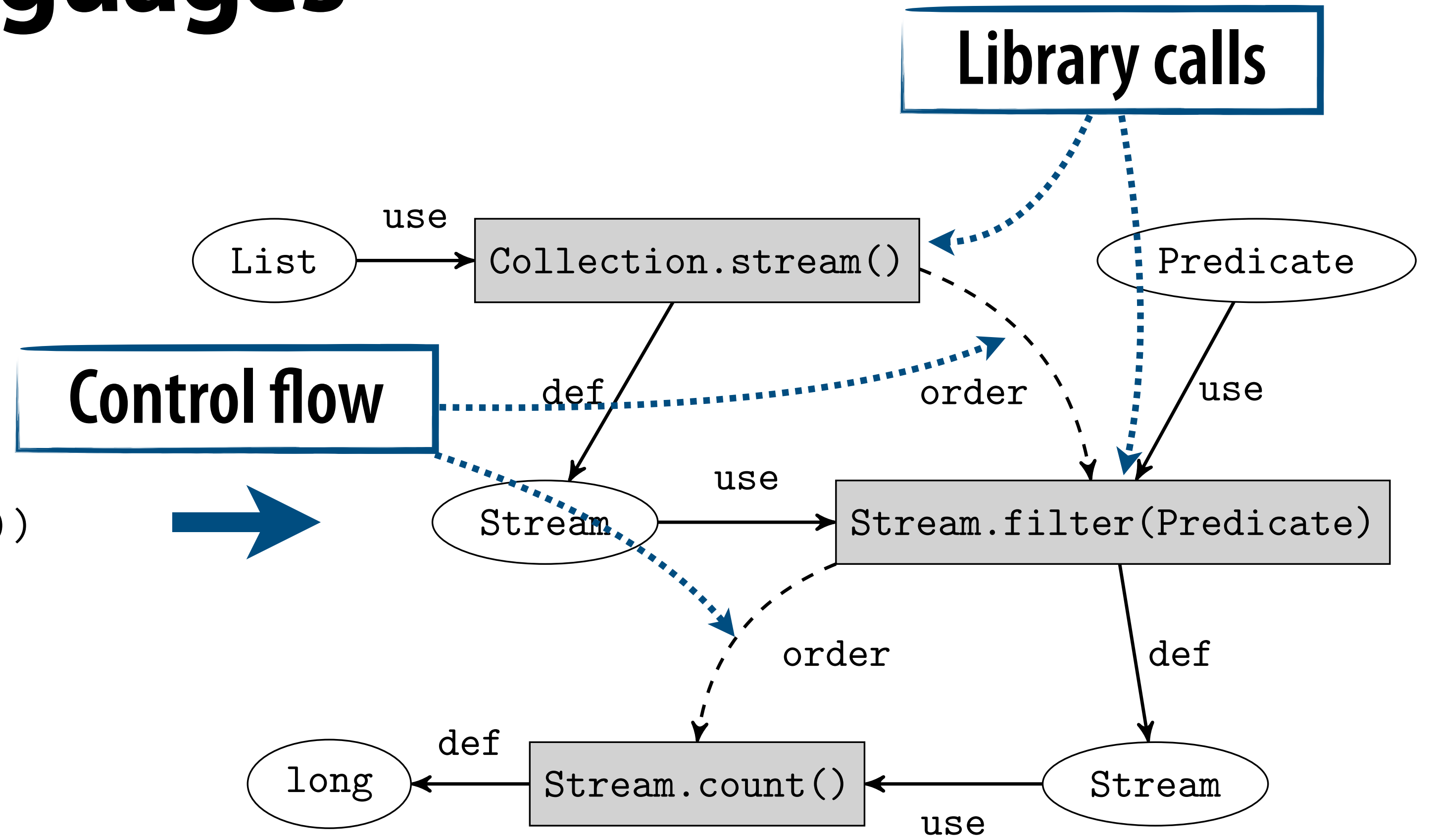
# Groups for OO languages

```

void my_method(List<String> lst) {
  return lst.stream()
    .filter(s -> s.startsWith("test_"))
    .count();
}

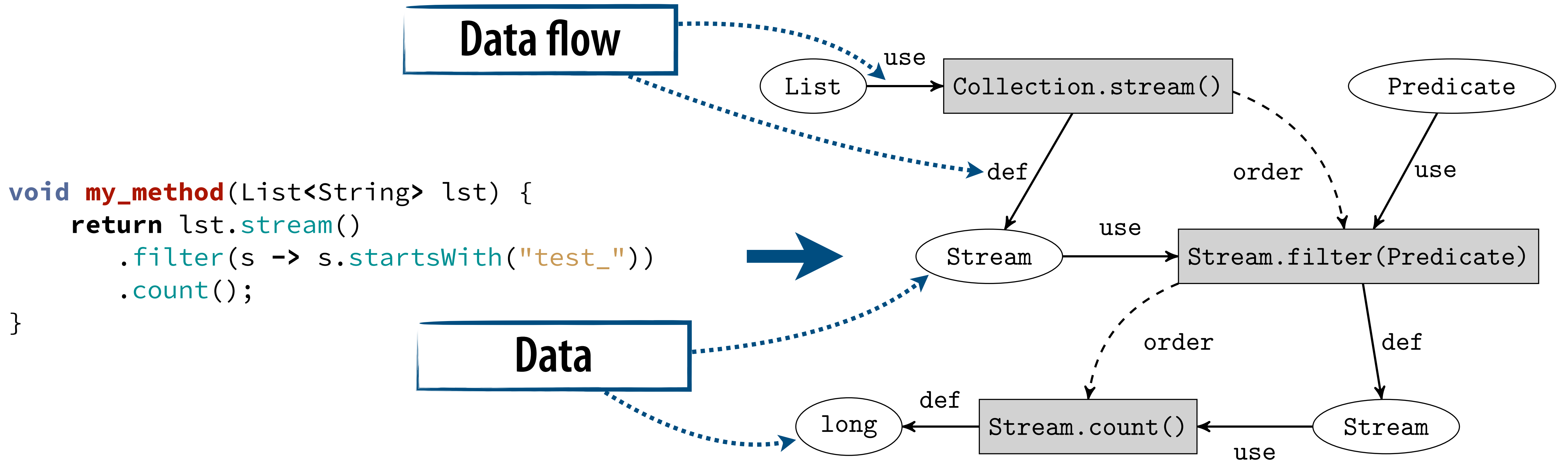
```

Java snippet using Collection API



Graph-based object usage model

# Groups for OO languages



Java snippet using Collection API

Graph-based object usage model

# COBOL Primer

```
IDENTIFICATION    DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "--- Start ---".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "--- End ---".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

# COBOL Primer

## Paragraphs



```
IDENTIFICATION DIVISION.  
PROGRAM-ID. exc.  
PROCEDURE DIVISION.
```

```
P0.  
    DISPLAY "--- Start ---".  
    CALL "logger".  
    PERFORM P1 THRU P3.
```

```
PZ.  
    CALL "loggerZ".  
    DISPLAY "--- End ---".  
    STOP RUN.
```

```
P3.  
    CALL "loggerB"
```

```
P1.  
    CALL "loggerA".  
    GO TO P3.
```

# COBOL Primer

**Call statement**  
Calls an external program.  
"Library call"

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "--- Start ---".
  CALL "logger".
  PERFORM P1 THRU P3.

PZ.
  CALL "loggerZ".
  DISPLAY "--- End ---".
  STOP RUN.

P3.
  CALL "loggerB"

P1.
  CALL "loggerA".
  GO TO P3.
  
```

**Perform statement**  
Jump to P1, execute until  
P3, jump back

**Go to statement**  
Jump to P3 and continue  
execution

# COBOL Primer

```
IDENTIFICATION DIVISION.  
PROGRAM-ID. exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "--- Start ---".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
PZ.  
    CALL "loggerZ".  
    DISPLAY "--- End ---".  
    STOP RUN.  
P3.  
    CALL "loggerB".  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

```
graph TD; P0 --> PZ; P0 --> P1; P1 --> P3; P3 --> PZ; PZ --> P3;
```

**Jumping leads to  
complex control flow**

# Groups for COBOL

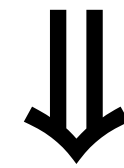
```
IDENTIFICATION DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "---- Start ----".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "---- End ----".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

Single paragraph (often)  
too small

# Groups for COBOL

```
IDENTIFICATION DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "---- Start ----".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "---- End ----".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

Single paragraph (often)  
too small



Inter-paragraph Groups



# Groups for COBOL

```
IDENTIFICATION DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "---- Start ----".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "---- End ----".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

Single paragraph (often)  
too small



Inter-paragraph Groups



Inter-paragraph control flow

# Groups for COBOL

```
IDENTIFICATION DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "---- Start ----".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "---- End ----".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

Single paragraph (often)  
too small



Inter-paragraph Groups



Inter-paragraph control flow



Graph inlining

# Groups for COBOL

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
    DISPLAY "---- Start ----".
    CALL "logger".
    PERFORM P1 THRU P3.

PZ.
    CALL "loggerZ".
    DISPLAY "---- End ----".
    STOP RUN.

P3.
    CALL "loggerB"

P1.
    CALL "loggerA".
    GO TO P3.
  
```

Single paragraph (often)  
too small



Inter-paragraph Groups



Inter-paragraph control flow



Graph inlining

## Other challenges

- Definition of library calls
- Absence of "def" edges
- Iteration through jumps
- Exit calls
- ...

# Phase 1: Intermediate Group Construction

```
IDENTIFICATION DIVISION.  
PROGRAM-ID.      exc.  
PROCEDURE DIVISION.  
P0.  
    DISPLAY "---- Start ----".  
    CALL "logger".  
    PERFORM P1 THRU P3.  
  
PZ.  
    CALL "loggerZ".  
    DISPLAY "---- End ----".  
    STOP RUN.  
  
P3.  
    CALL "loggerB"  
  
P1.  
    CALL "loggerA".  
    GO TO P3.
```

# Phase 1: Intermediate Group Construction

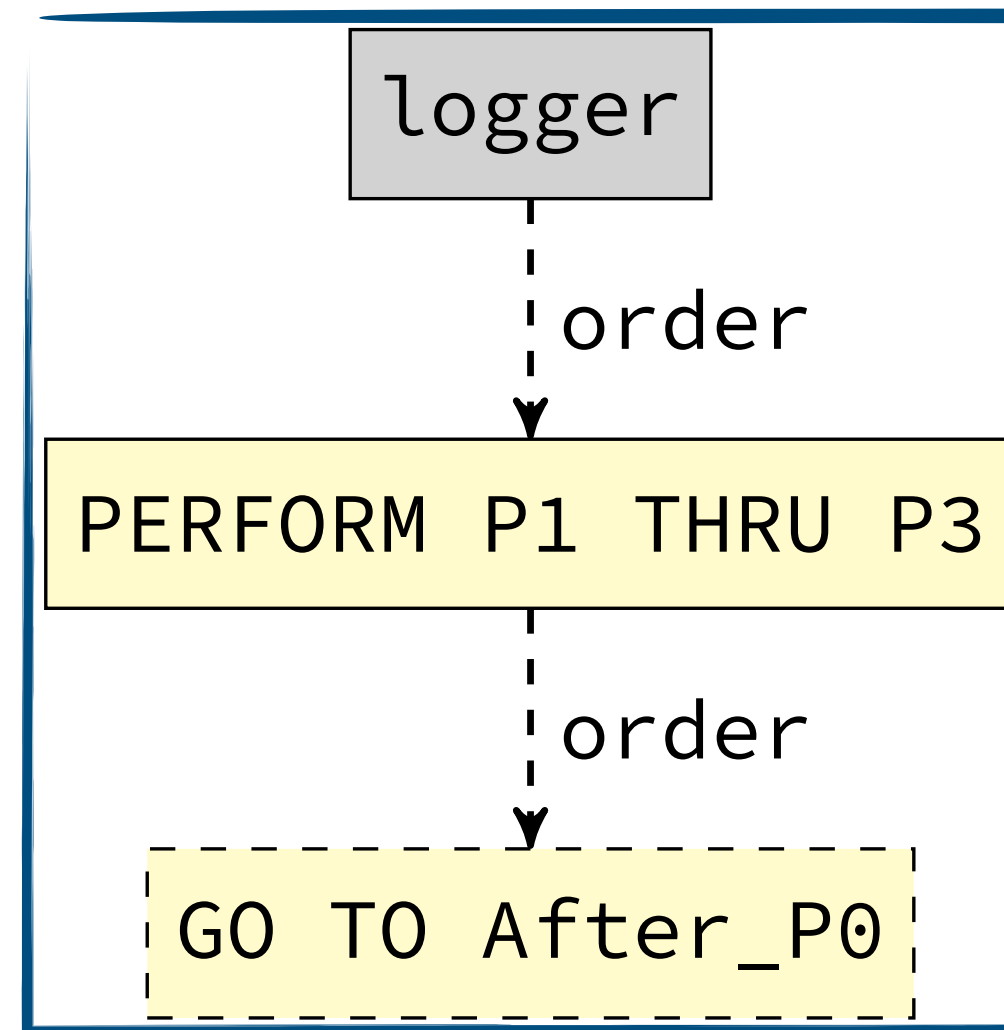
**IDENTIFICATION DIVISION.**  
**PROGRAM-ID.**           exc.  
**PROCEDURE DIVISION.**

P0.  
  DISPLAY "---- Start ----".  
  CALL "logger".  
  PERFORM P1 THRU P3.

PZ.  
  CALL "loggerZ".  
  DISPLAY "---- End ----".  
  STOP RUN.

P3.  
  CALL "loggerB"

P1.  
  CALL "loggerA".  
  GO TO P3.



# Phase 1: Intermediate Group Construction

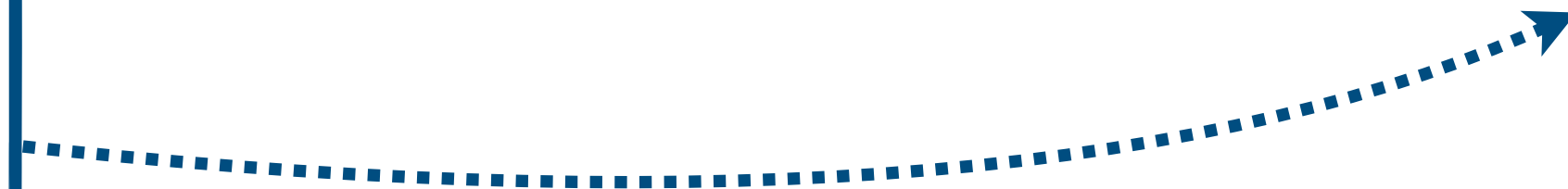
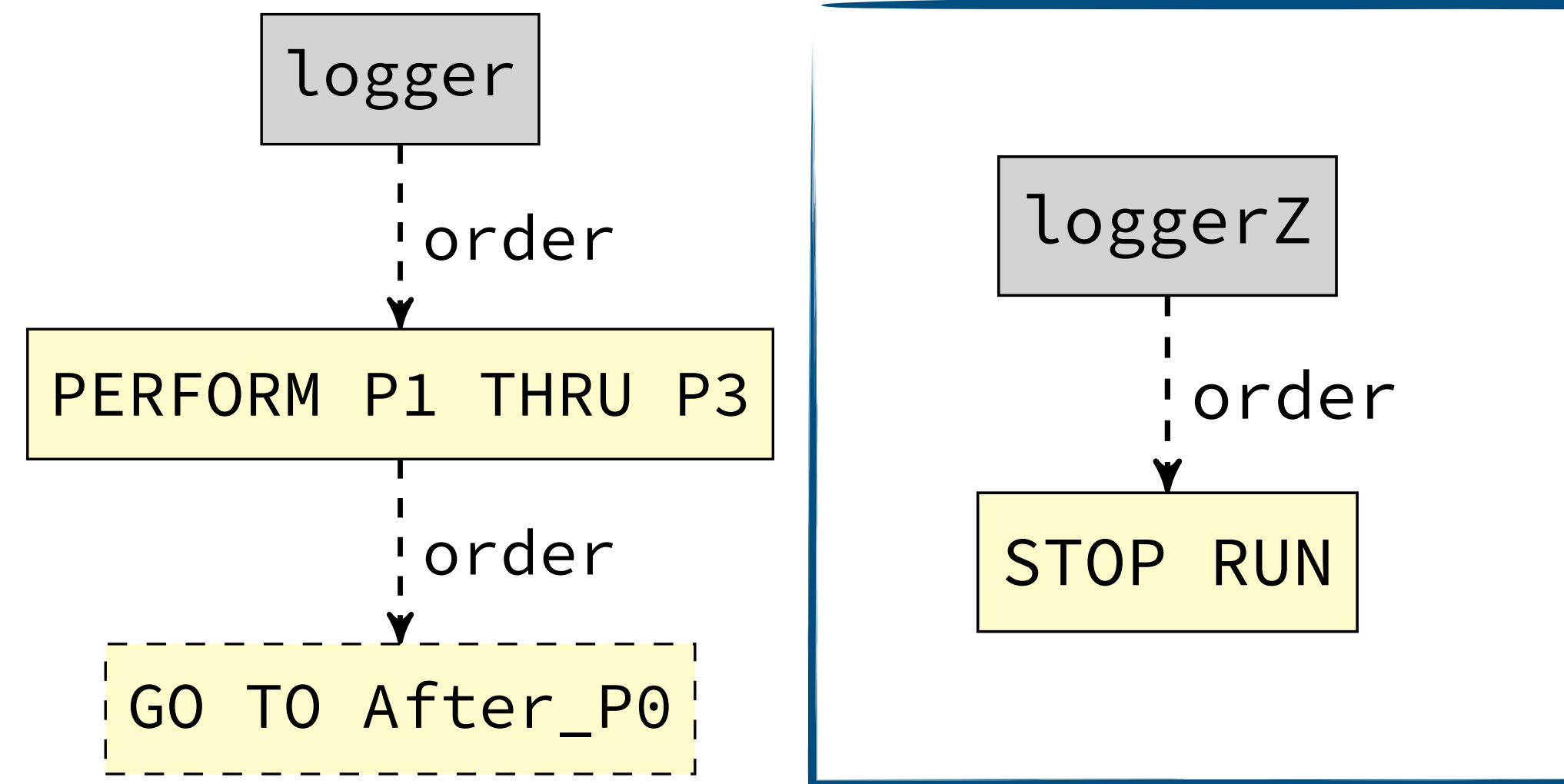
**IDENTIFICATION DIVISION.**  
**PROGRAM-ID.** exc.  
**PROCEDURE DIVISION.**

P0.  
 DISPLAY "---- Start ----".  
 CALL "logger".  
 PERFORM P1 THRU P3.

PZ.  
 CALL "loggerZ".  
 DISPLAY "---- End ----".  
 STOP RUN.

P3.  
 CALL "loggerB"

P1.  
 CALL "loggerA".  
 GO TO P3.



# Phase 1: Intermediate Group Construction

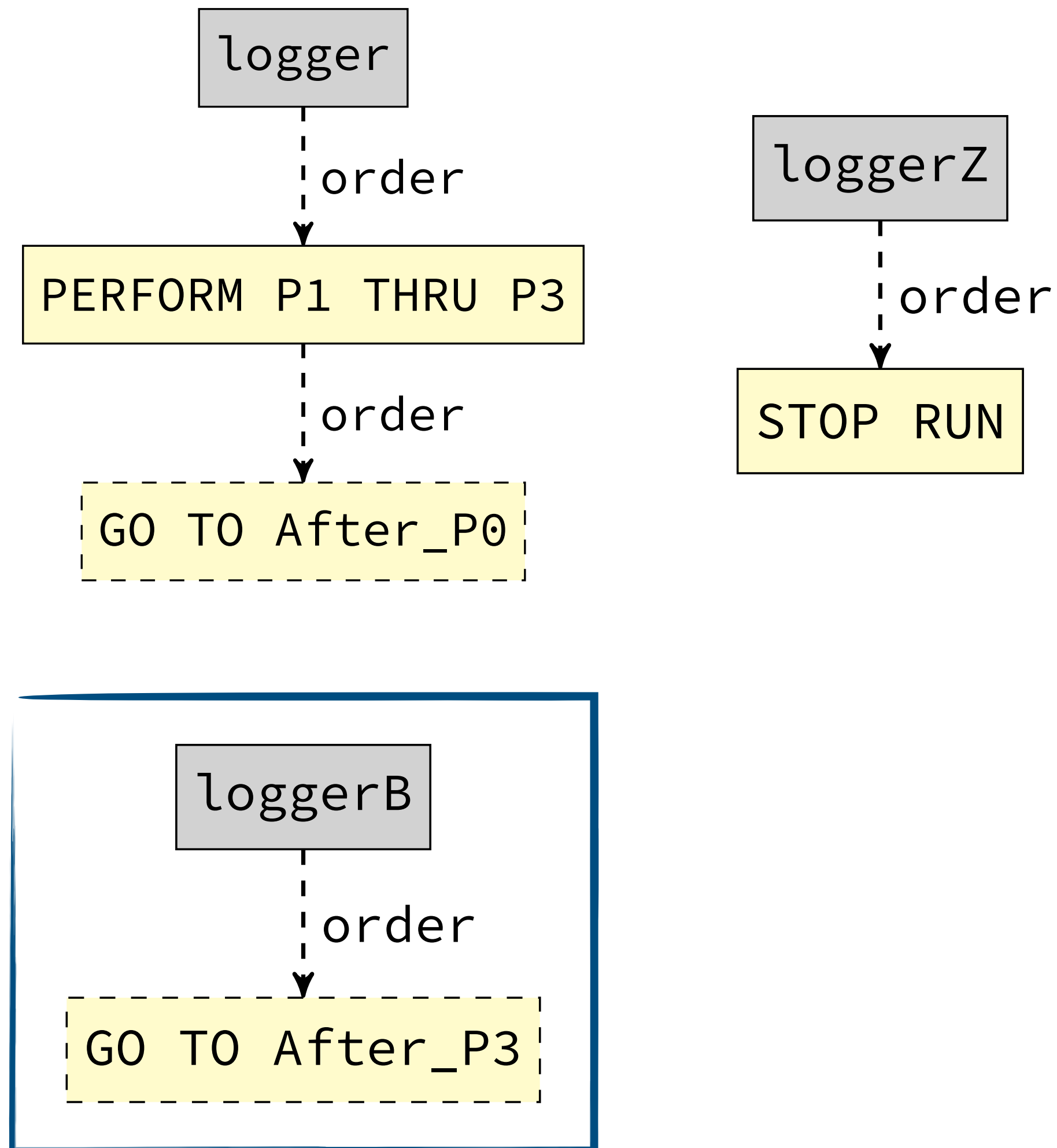
```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.

PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.

P3.
  CALL "loggerB"

P1.
  CALL "loggerA".
  GO TO P3.
  
```



# Phase 1: Intermediate Group Construction

```

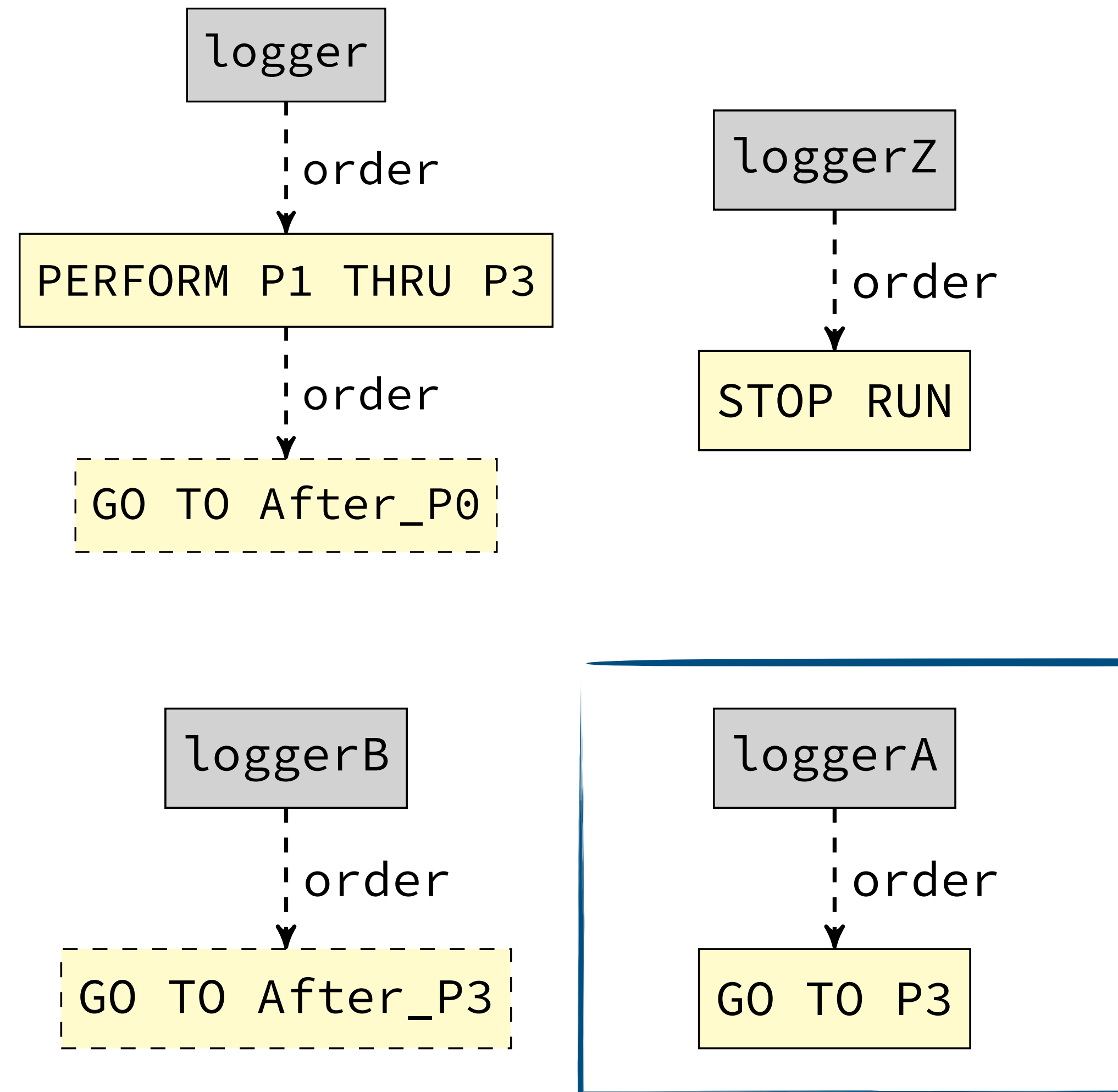
IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.

PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.

P3.
  CALL "loggerB"

P1.
  CALL "loggerA".
  GO TO P3.

```





# Phase 2: Group Inlining

IDENTIFICATION DIVISION.

PROGRAM-ID. exc.

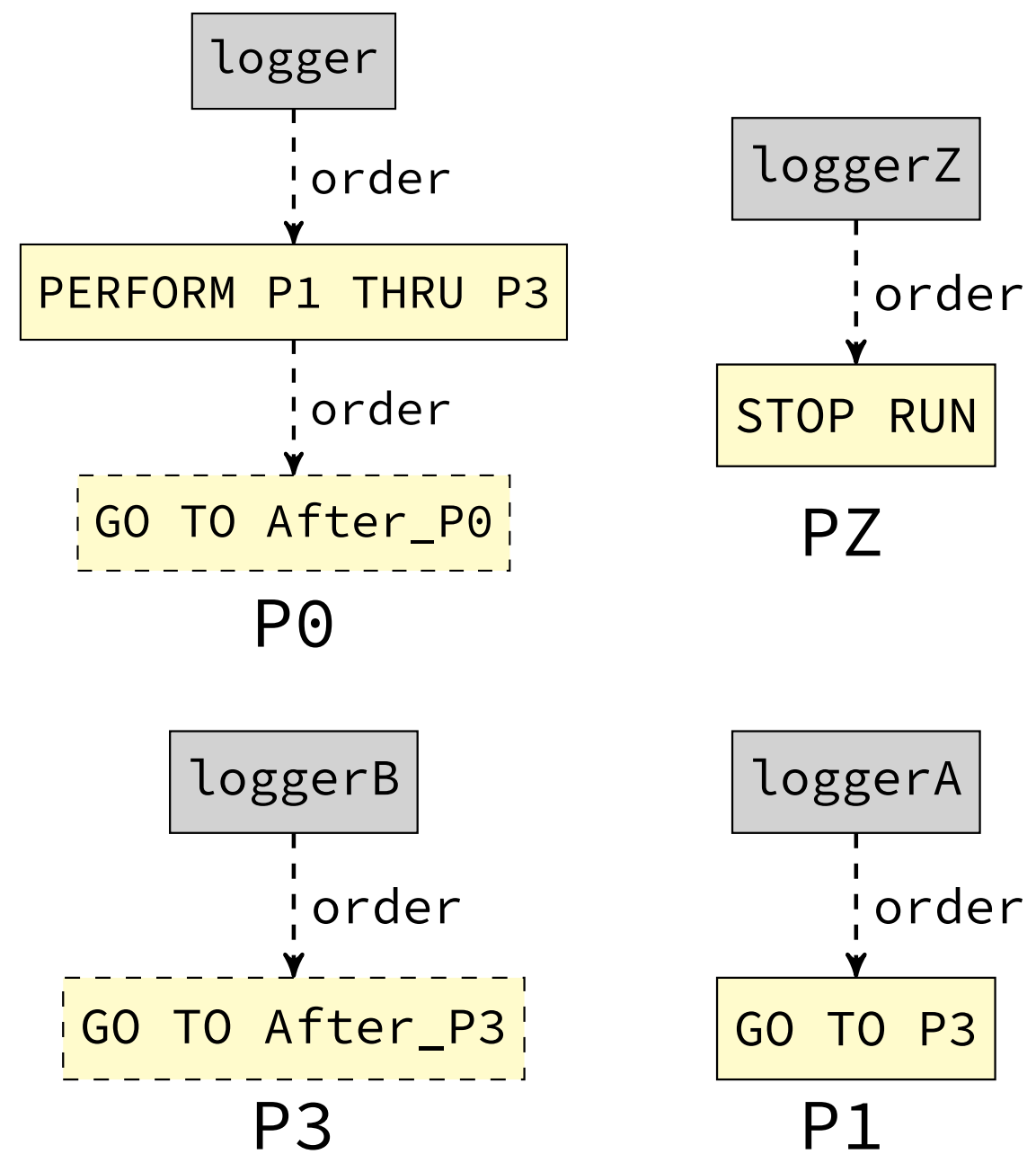
PROCEDURE DIVISION.

P0.  
  DISPLAY "---- Start ----".  
  CALL "logger".  
  PERFORM P1 THRU P3.

PZ.  
  CALL "loggerZ".  
  DISPLAY "---- End ----".  
  STOP RUN.

P3.  
  CALL "loggerB"

P1.  
  CALL "loggerA".  
  GO TO P3.



### Jump Table

After_P0	PZ
After_P3	P0

# Phase 2: Group Inlining

```

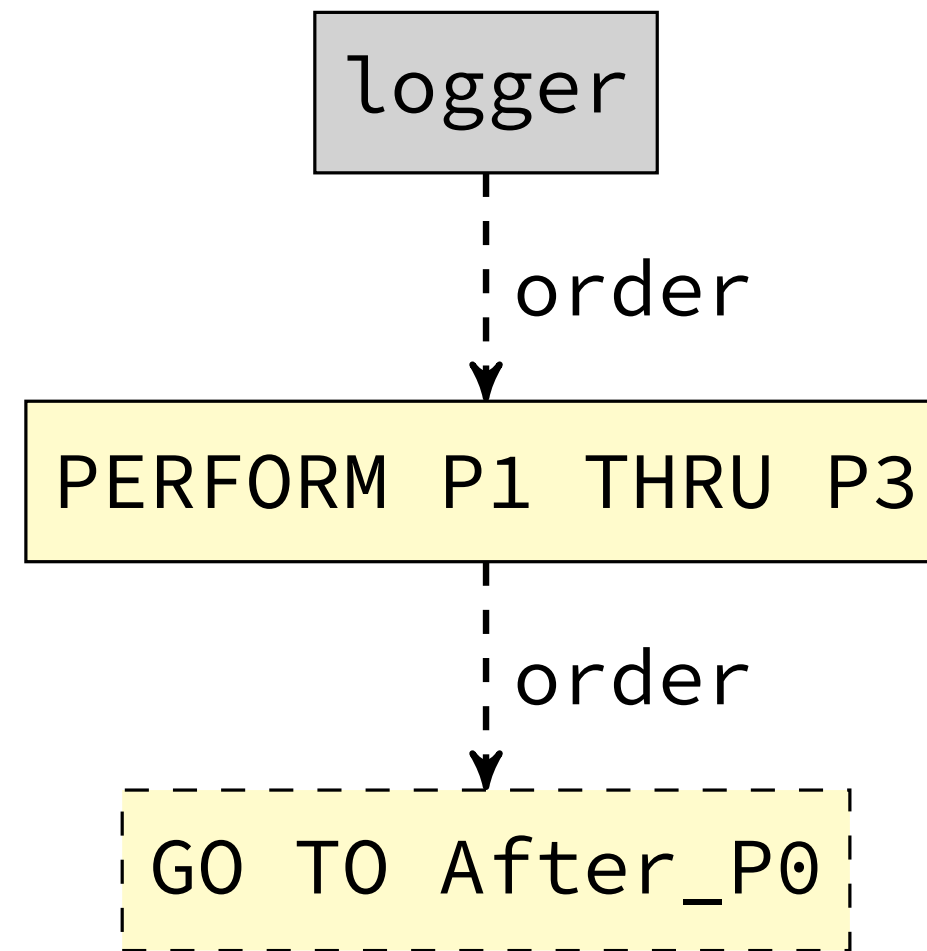
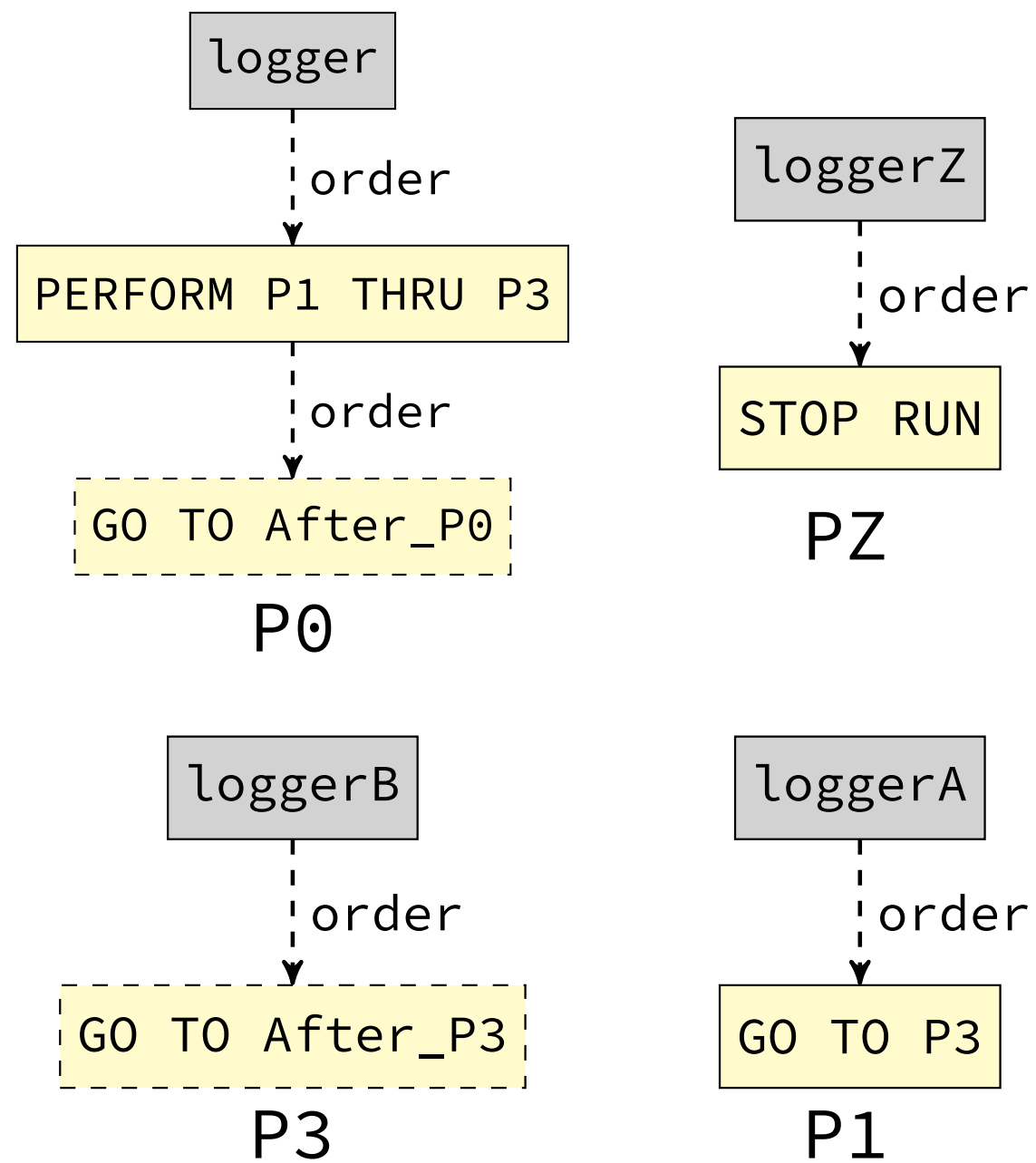
IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.

```

```

P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB"
P1.
  CALL "loggerA".
  GO TO P3.

```



### Jump Table

After_P0	PZ
After_P3	P0

# Phase 2: Group Inlining

IDENTIFICATION DIVISION.

PROGRAM-ID. exc.

PROCEDURE DIVISION.

P0.

DISPLAY "---- Start ----".

CALL "logger".

PERFORM P1 THRU P3.

PZ.

CALL "loggerZ".

DISPLAY "---- End ----".

STOP RUN.

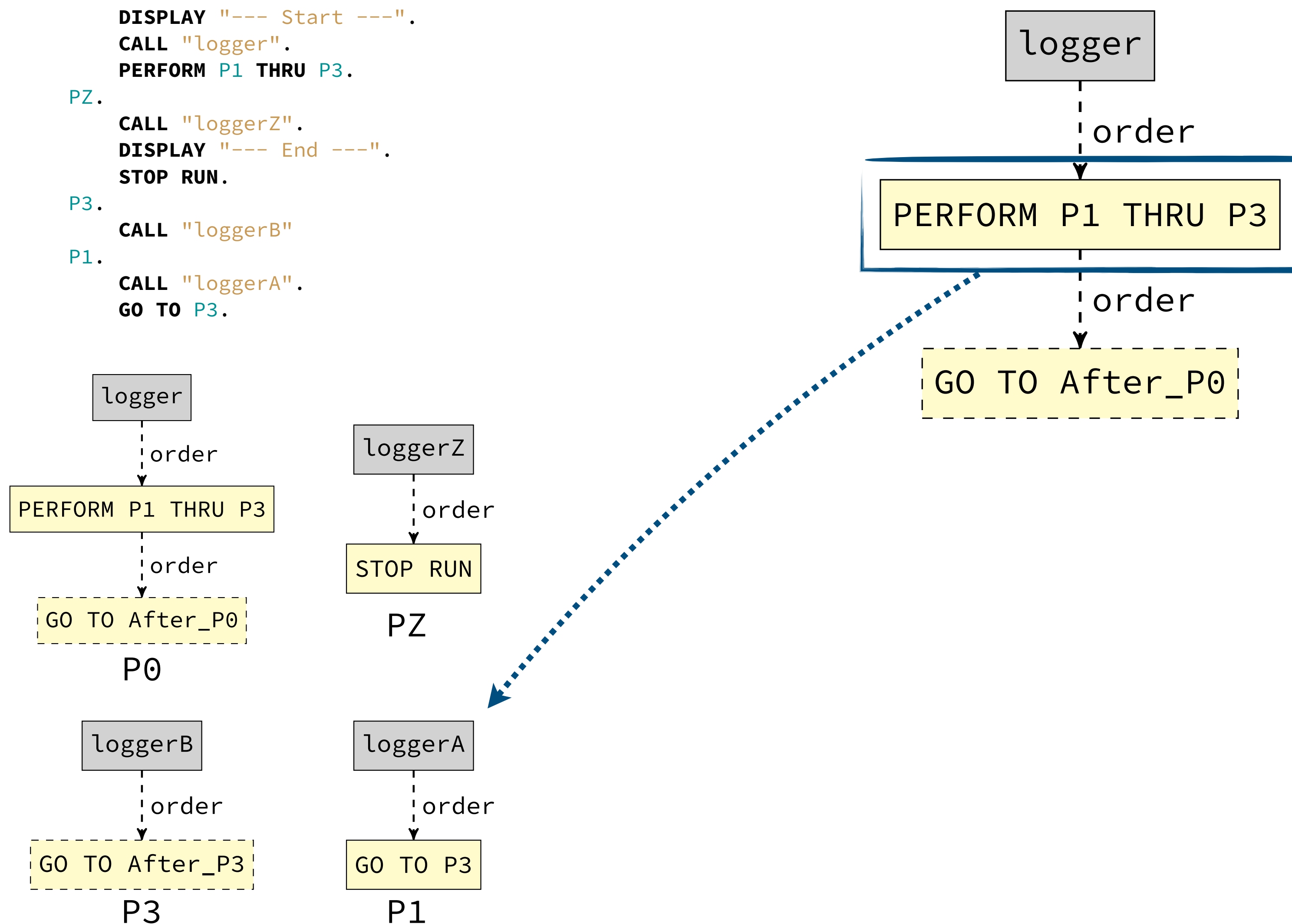
P3.

CALL "loggerB"

P1.

CALL "loggerA".

GO TO P3.



## PERFORM

- Inline first paragraph
- Adjust last paragraph in jump table

## Jump Table

After_P0	PZ
After_P3	P0

# Phase 2: Group Inlining

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.

```

```

P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.

```

```

PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.

```

```

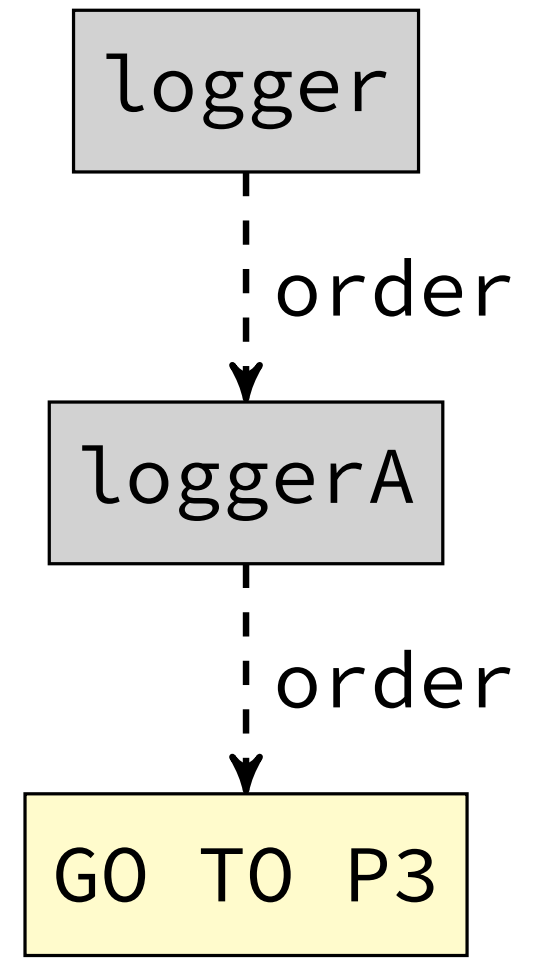
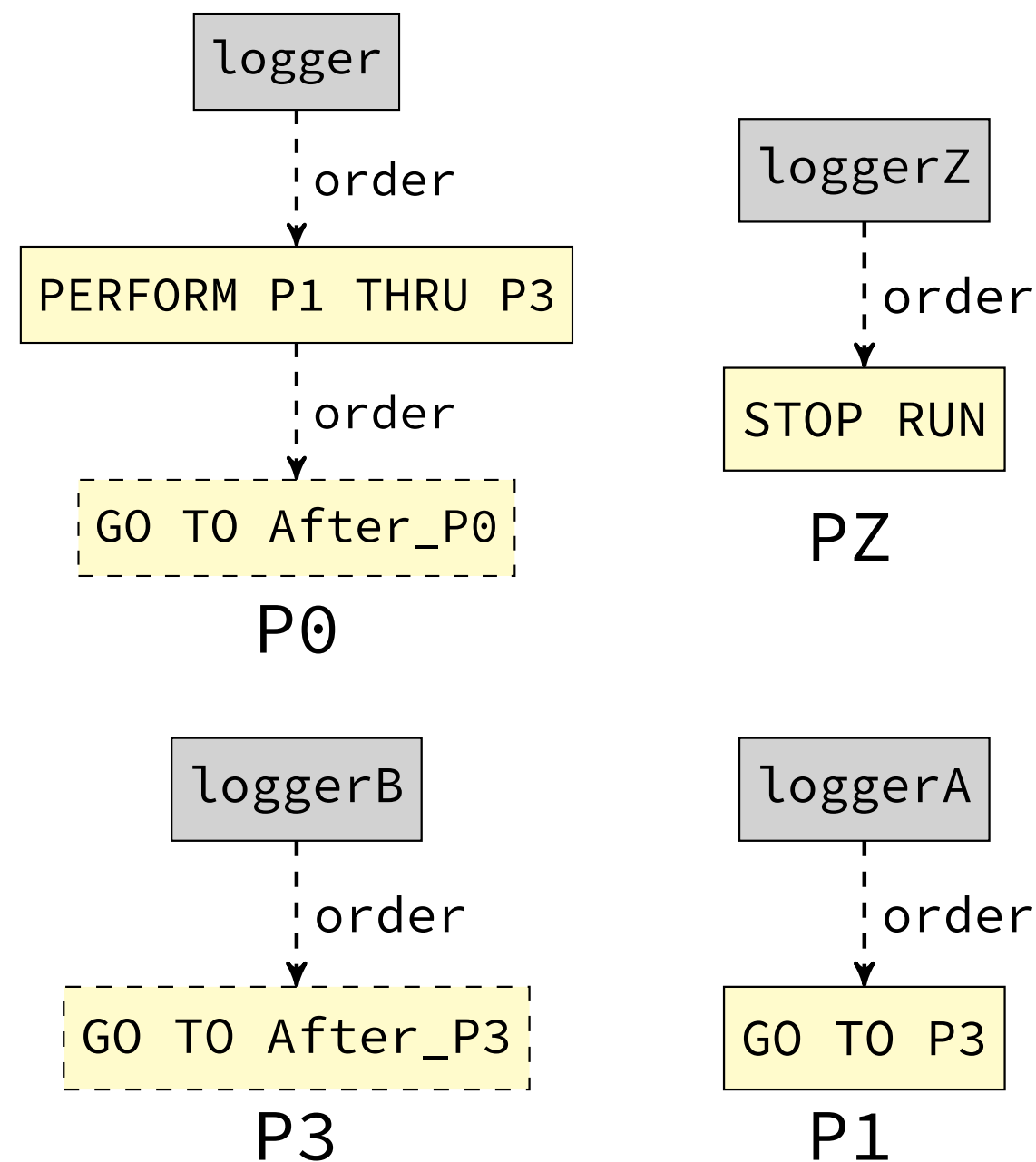
P3.
  CALL "loggerB"

```

```

P1.
  CALL "loggerA".
  GO TO P3.

```



GO TO After\_P0

**PERFORM**

- Inline first paragraph
- Adjust last paragraph in jump table

Jump Table

After_P0	PZ
After_P3	GO TO After_P0

# Phase 2: Group Inlining

```

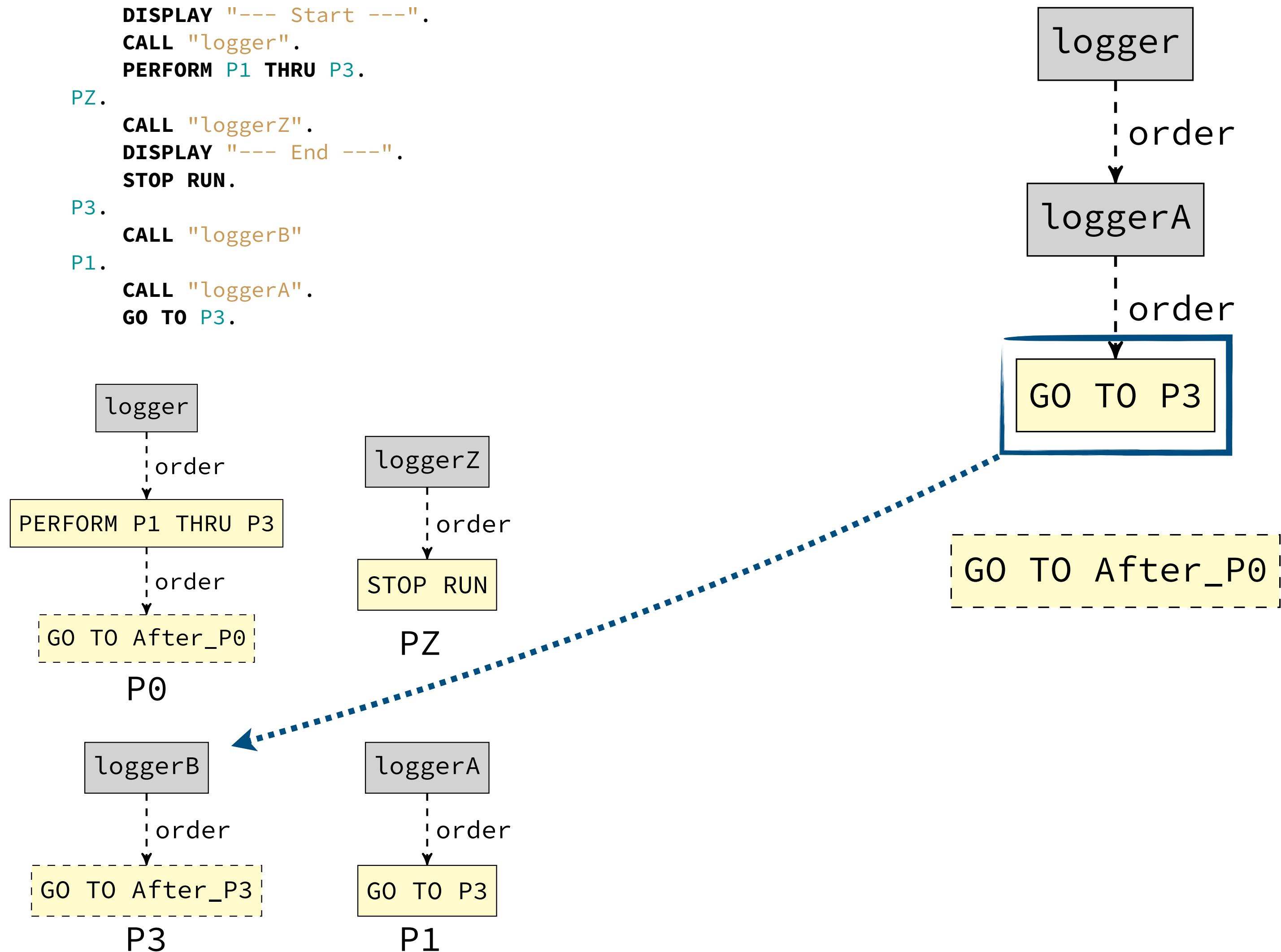
IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.

```

```

P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB".
P1.
  CALL "loggerA".
  GO TO P3.

```



**GO TO**  
Inline target paragraph

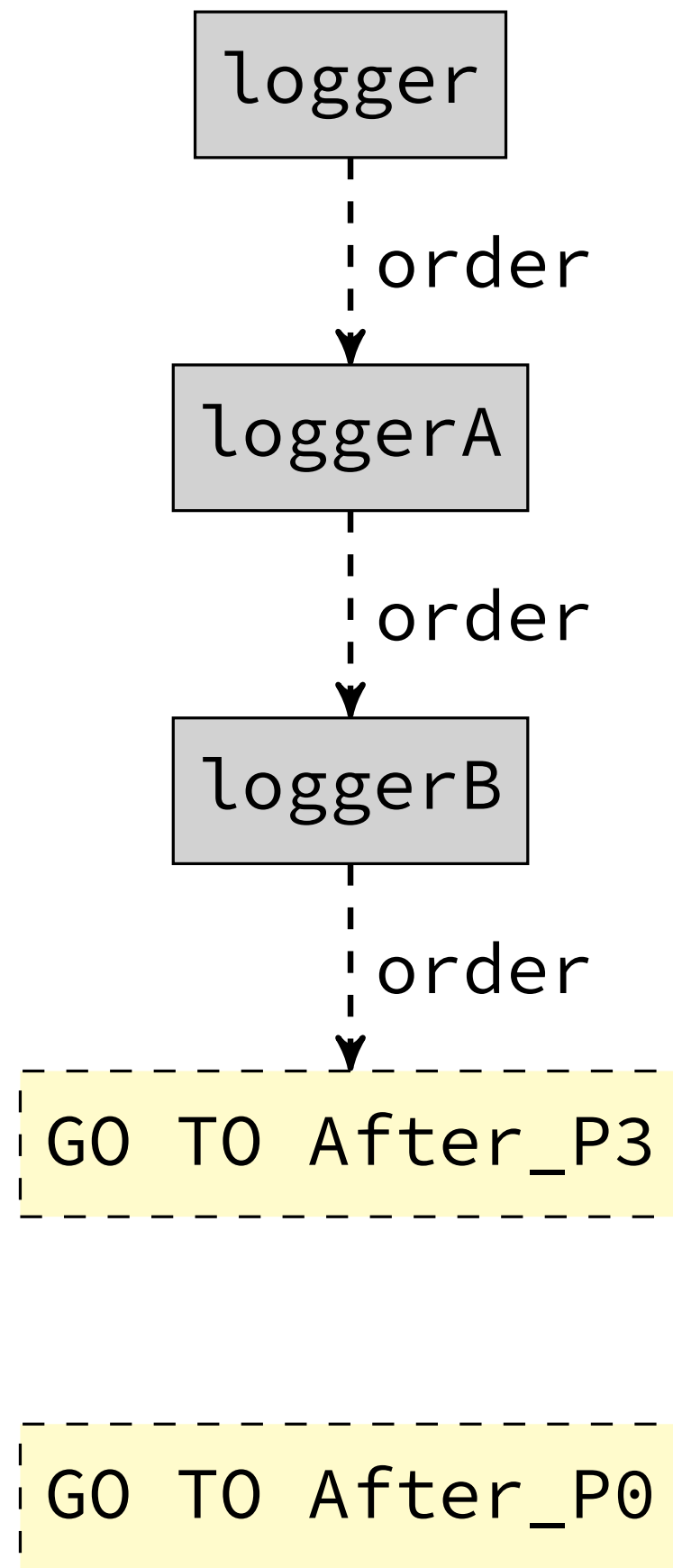
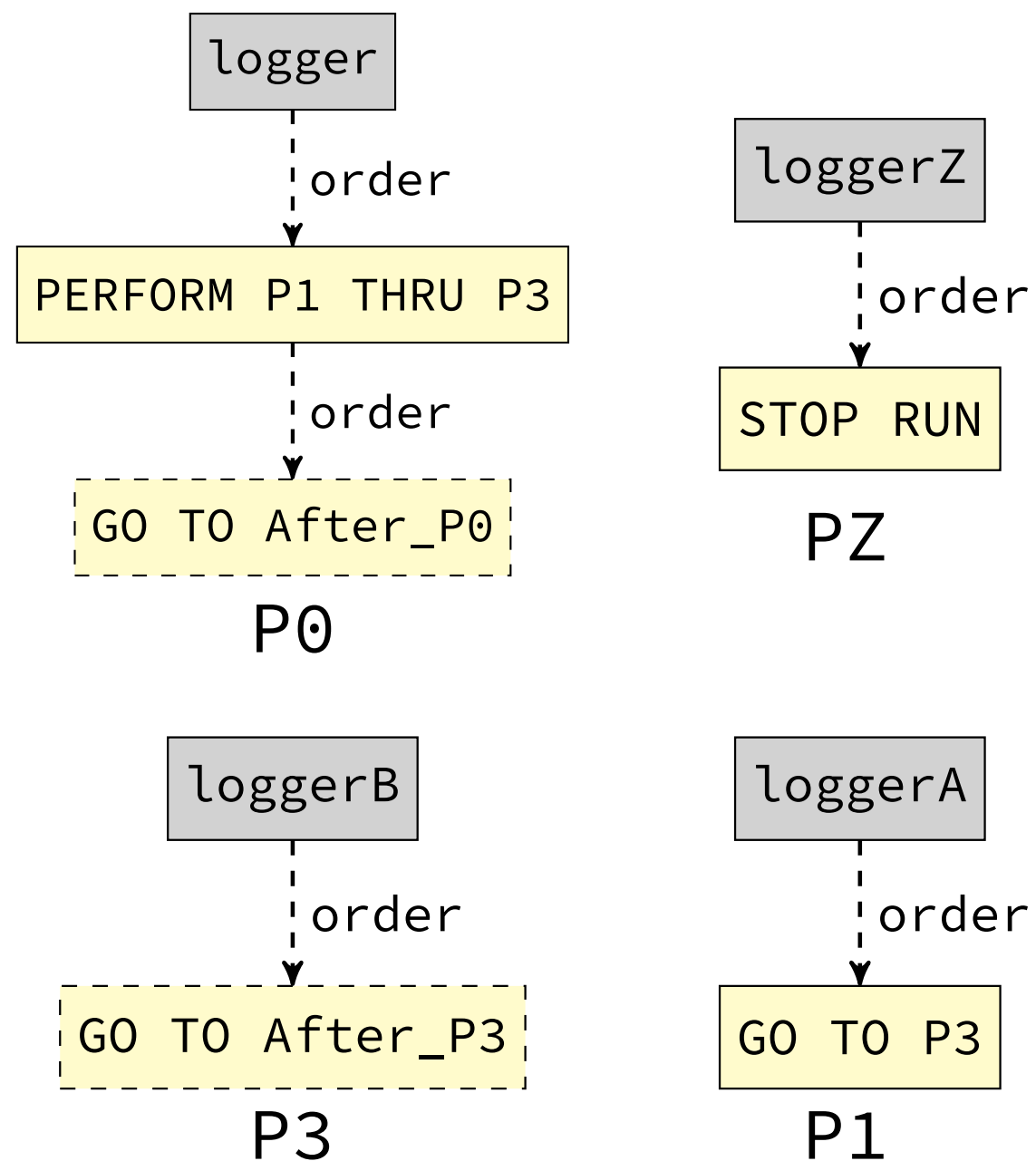
Jump Table

After_P0	PZ
After_P3	<b>GO TO After_P0</b>

# Phase 2: Group Inlining

IDENTIFICATION DIVISION.  
PROGRAM-ID. exc.  
PROCEDURE DIVISION.

```
P0.  
  DISPLAY "---- Start ----".  
  CALL "logger".  
  PERFORM P1 THRU P3.  
PZ.  
  CALL "loggerZ".  
  DISPLAY "---- End ----".  
  STOP RUN.  
P3.  
  CALL "loggerB"  
P1.  
  CALL "loggerA".  
  GO TO P3.
```



**GO TO**  
Inline target paragraph

Jump Table

After_P0	PZ
After_P3	GO TO After_P0

# Phase 2: Group Inlining

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.

```

```

P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.

```

```

PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.

```

```

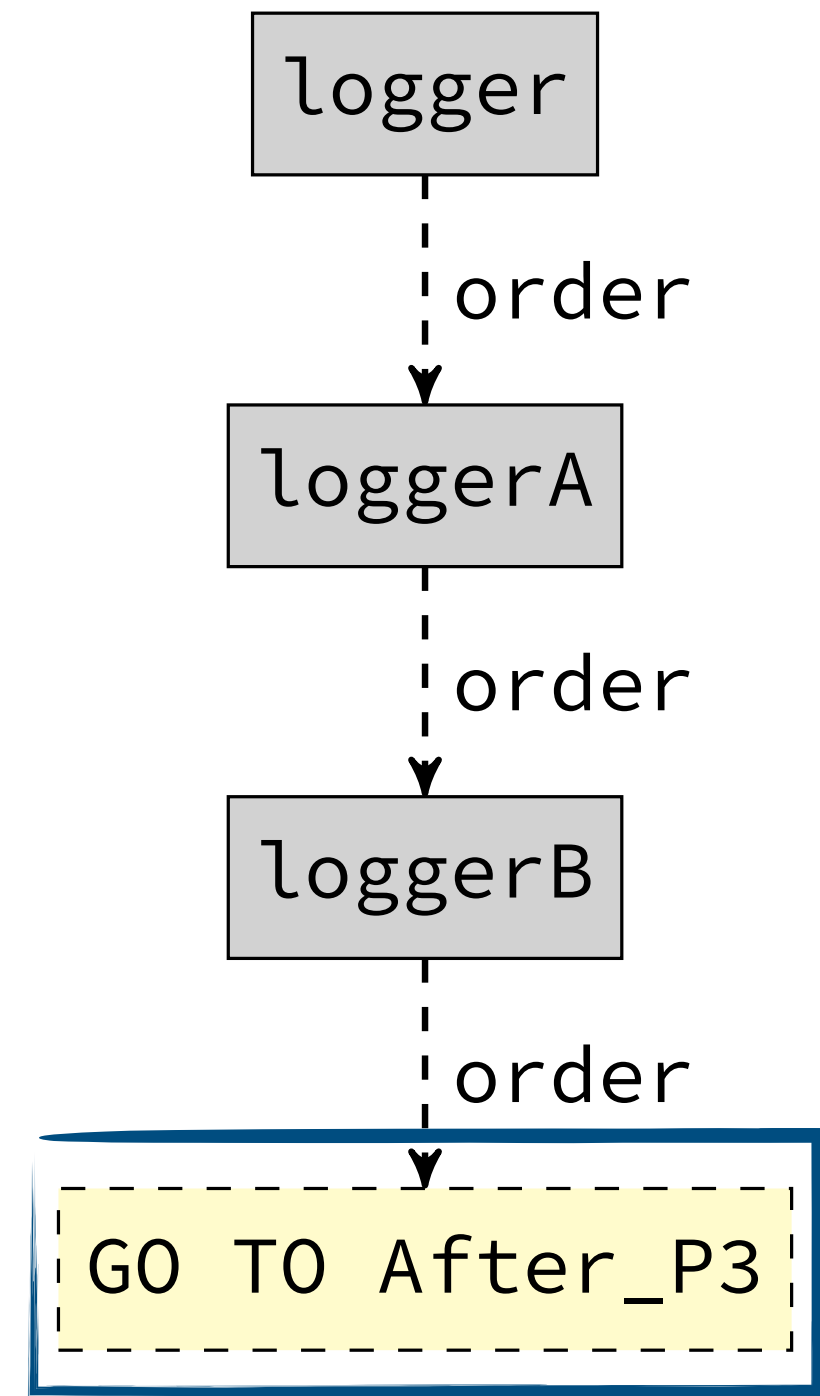
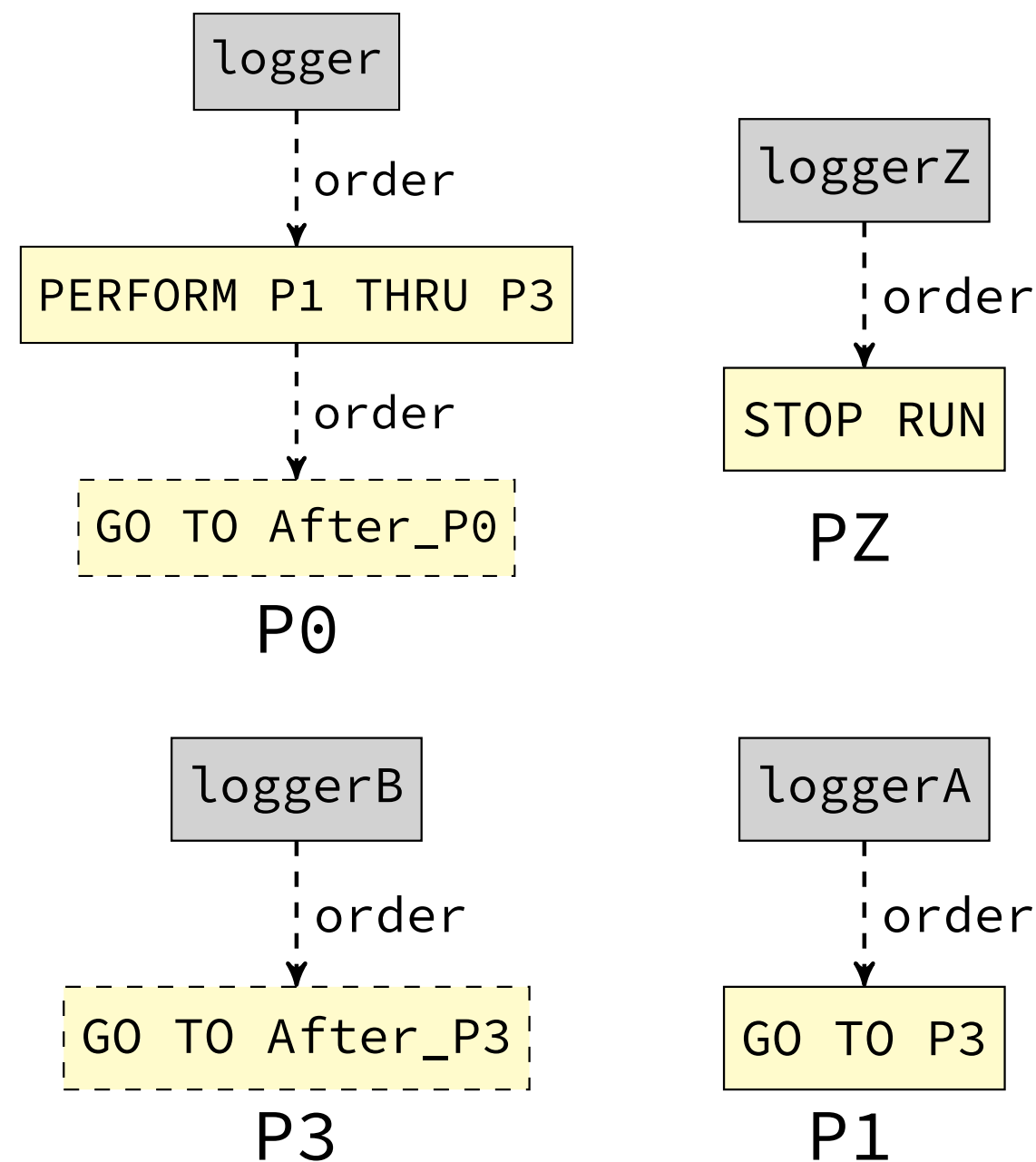
P3.
  CALL "loggerB"

```

```

P1.
  CALL "loggerA".
  GO TO P3.

```



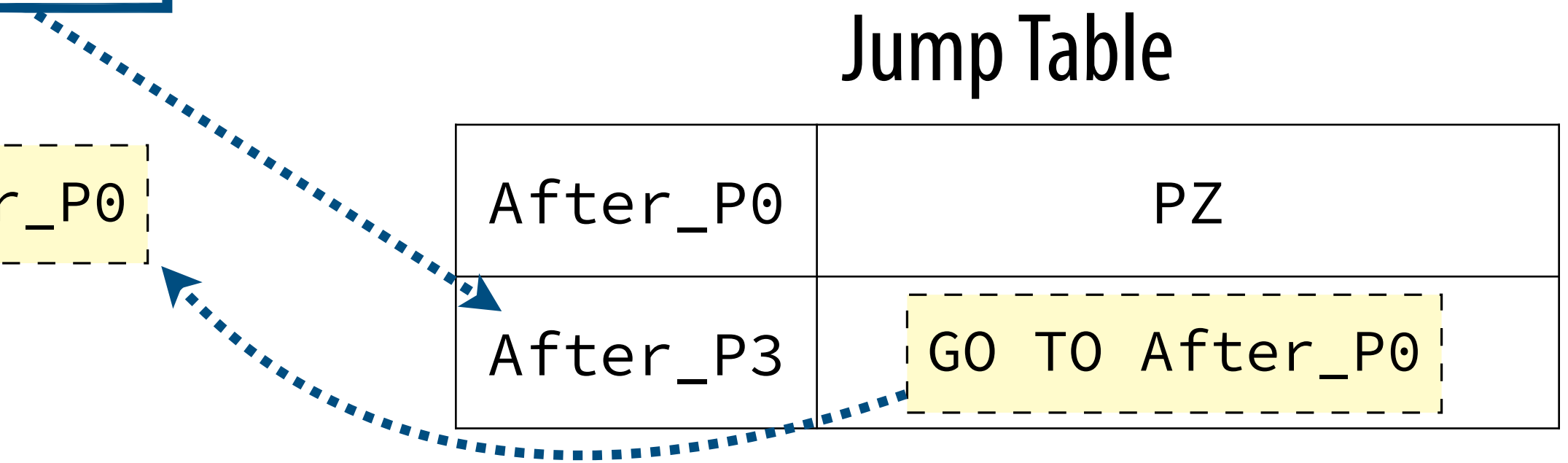
**Implicit jump after paragraph**  
Consult jump table

**End of PERFORM**  
Reset jump table entry

GO TO After\_P0

Jump Table

After_P0	PZ
After_P3	GO TO After_P0



# Phase 2: Group Inlining

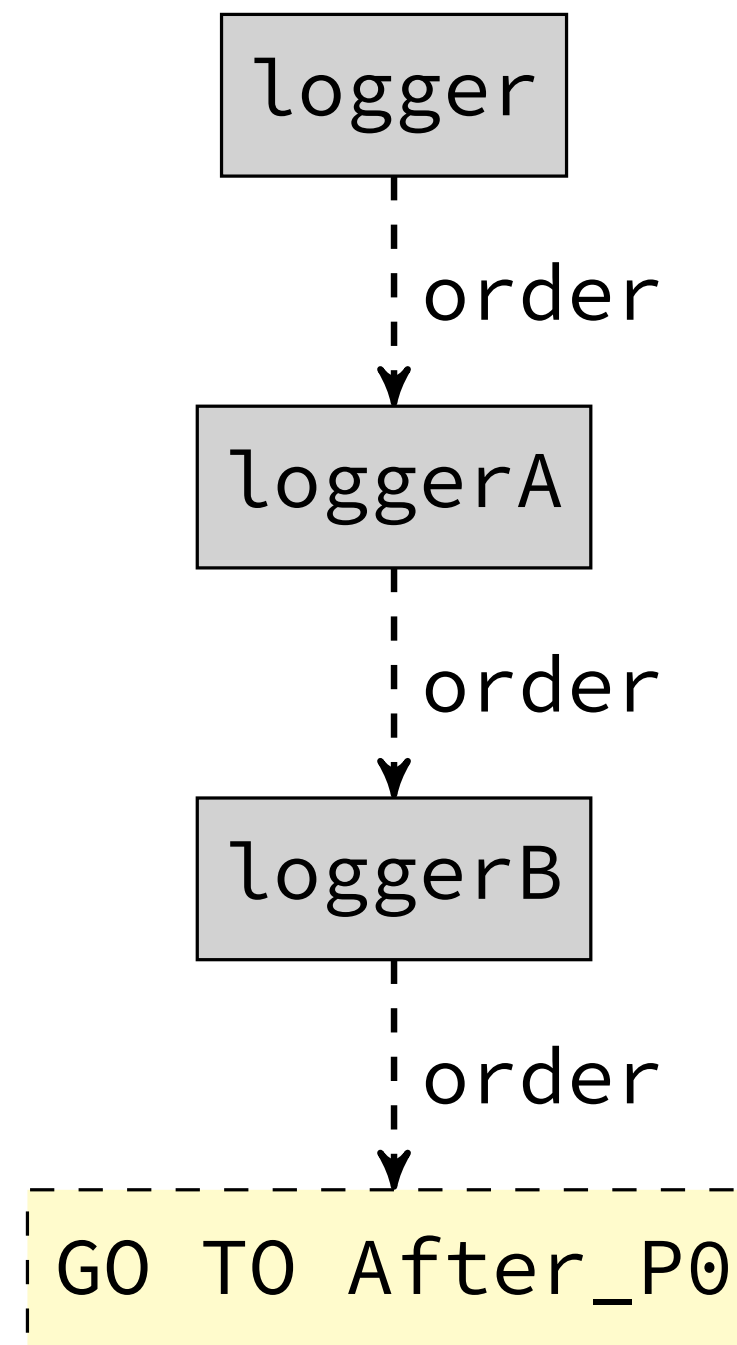
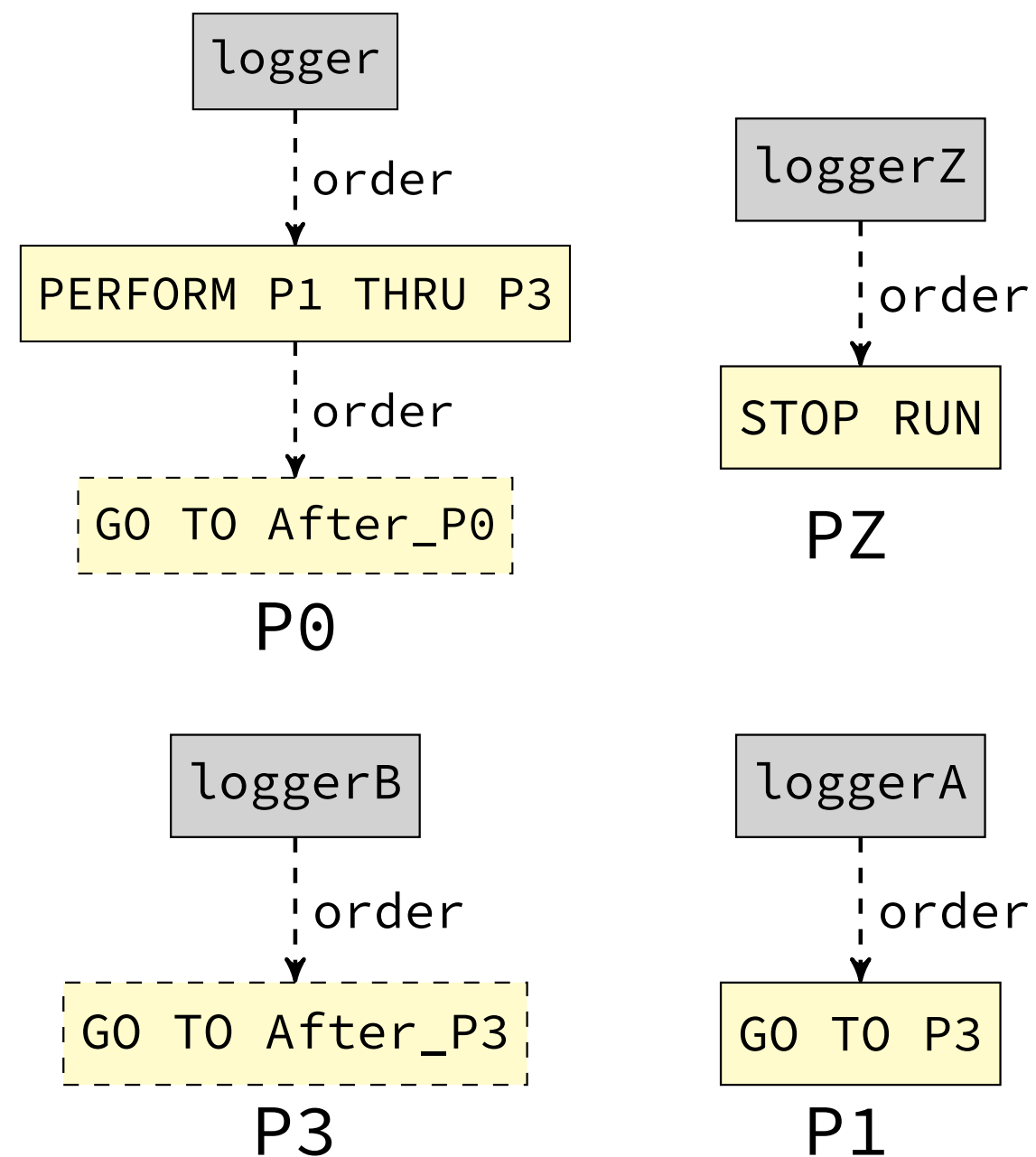
```
IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
```

```
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
```

```
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
```

```
P3.
  CALL "loggerB"
```

```
P1.
  CALL "loggerA".
  GO TO P3.
```



**Implicit jump after paragraph**  
Consult jump table

**End of PERFORM**  
Reset jump table entry

Jump Table

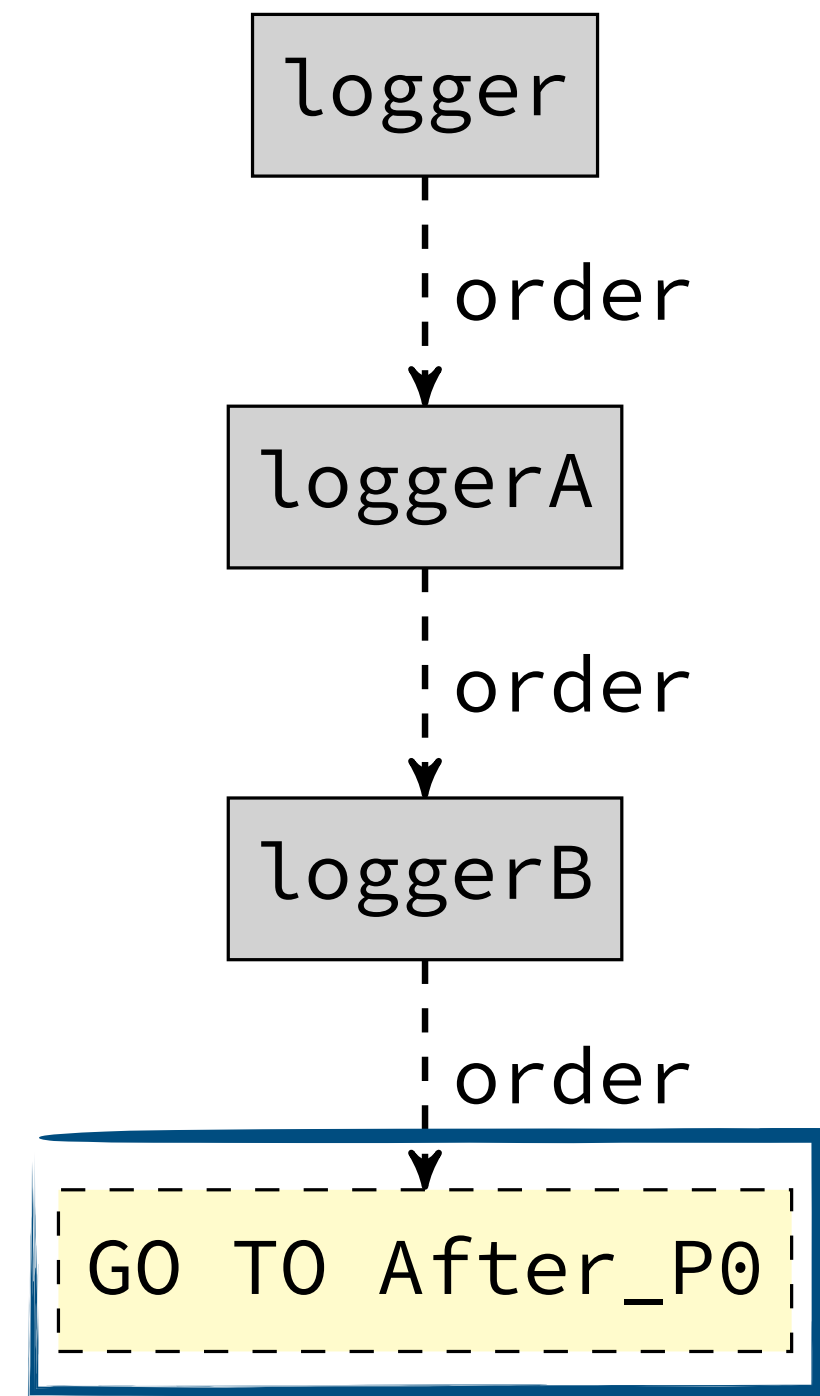
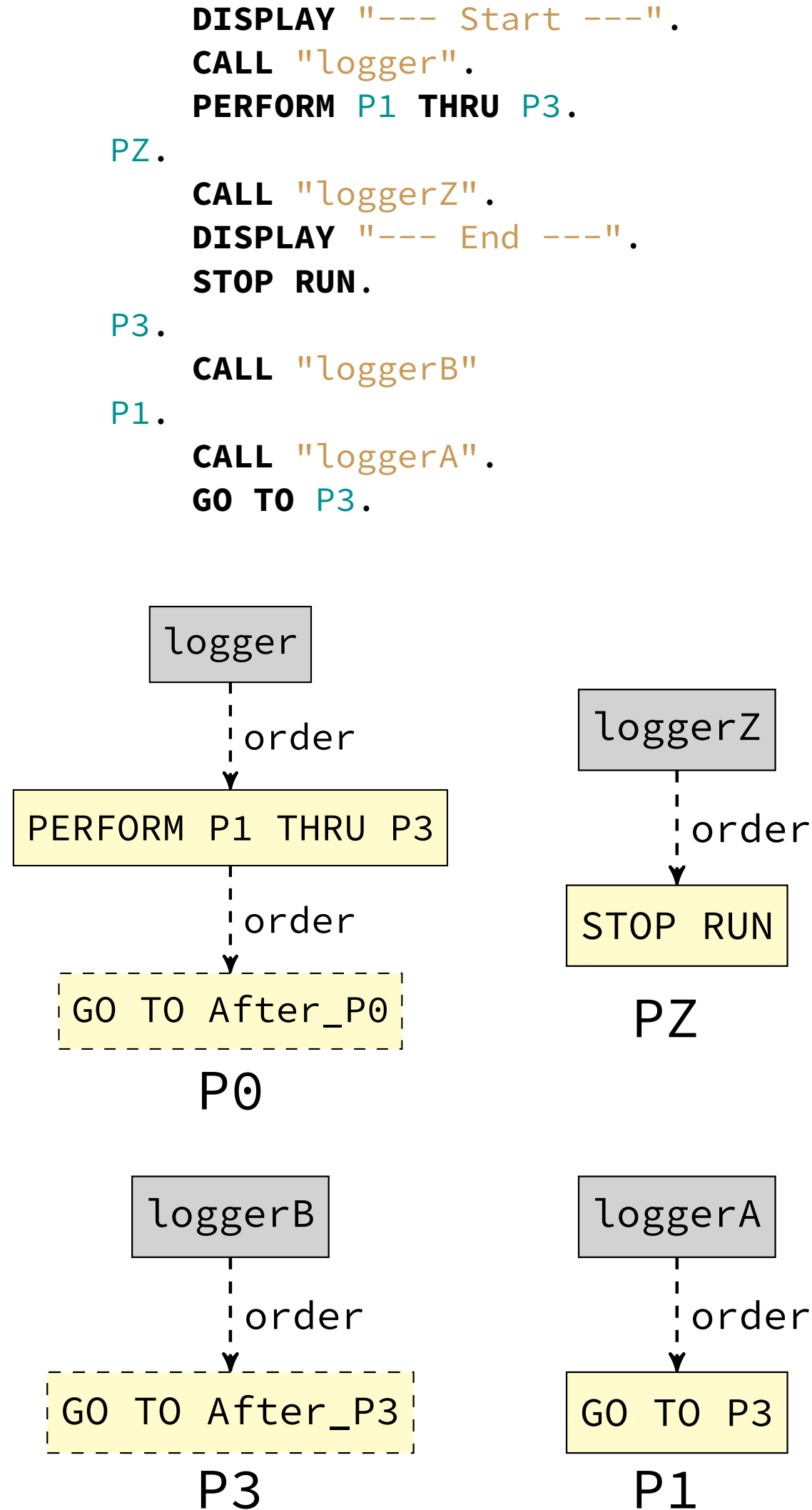
After_P0	PZ
After_P3	P0



# Phase 2: Group Inlining

```
IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
```

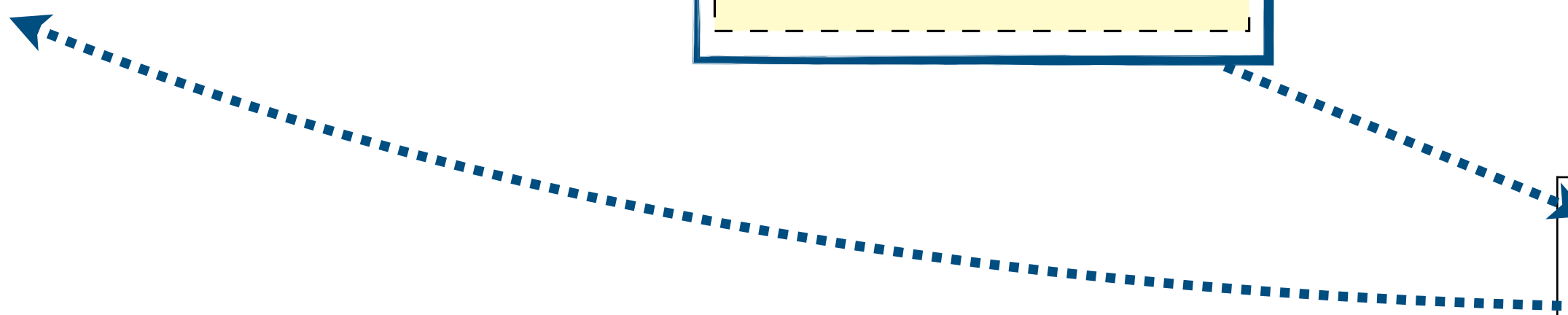
```
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB"
P1.
  CALL "loggerA".
  GO TO P3.
```



**Implicit jump after paragraph  
Consult jump table**

**Jump Table**

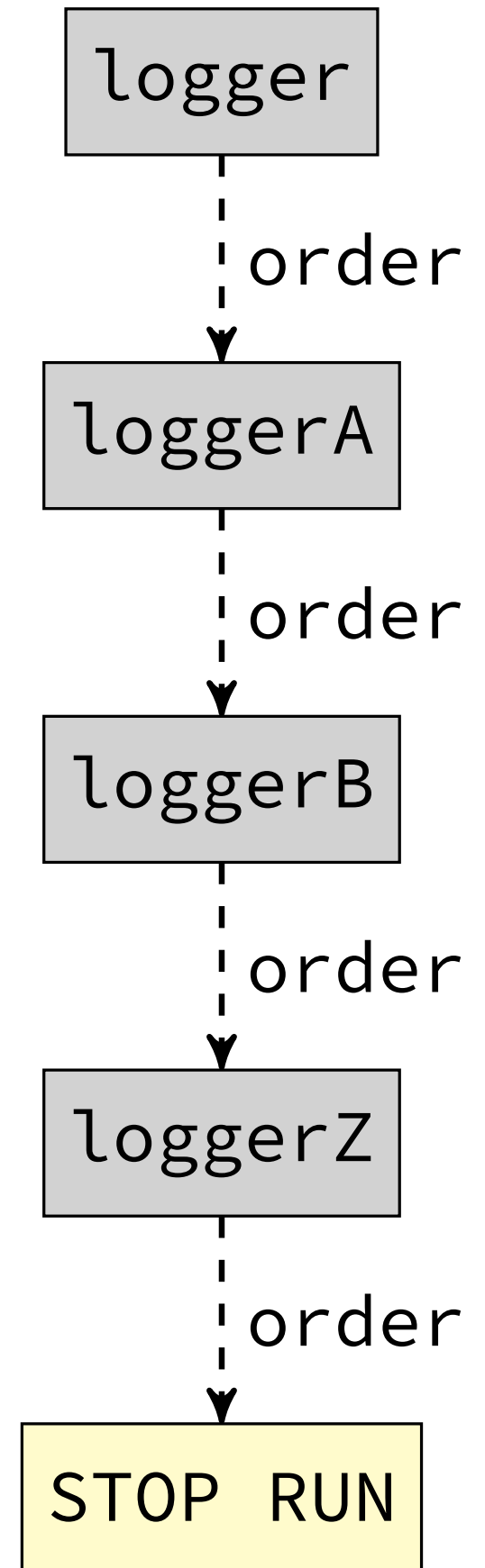
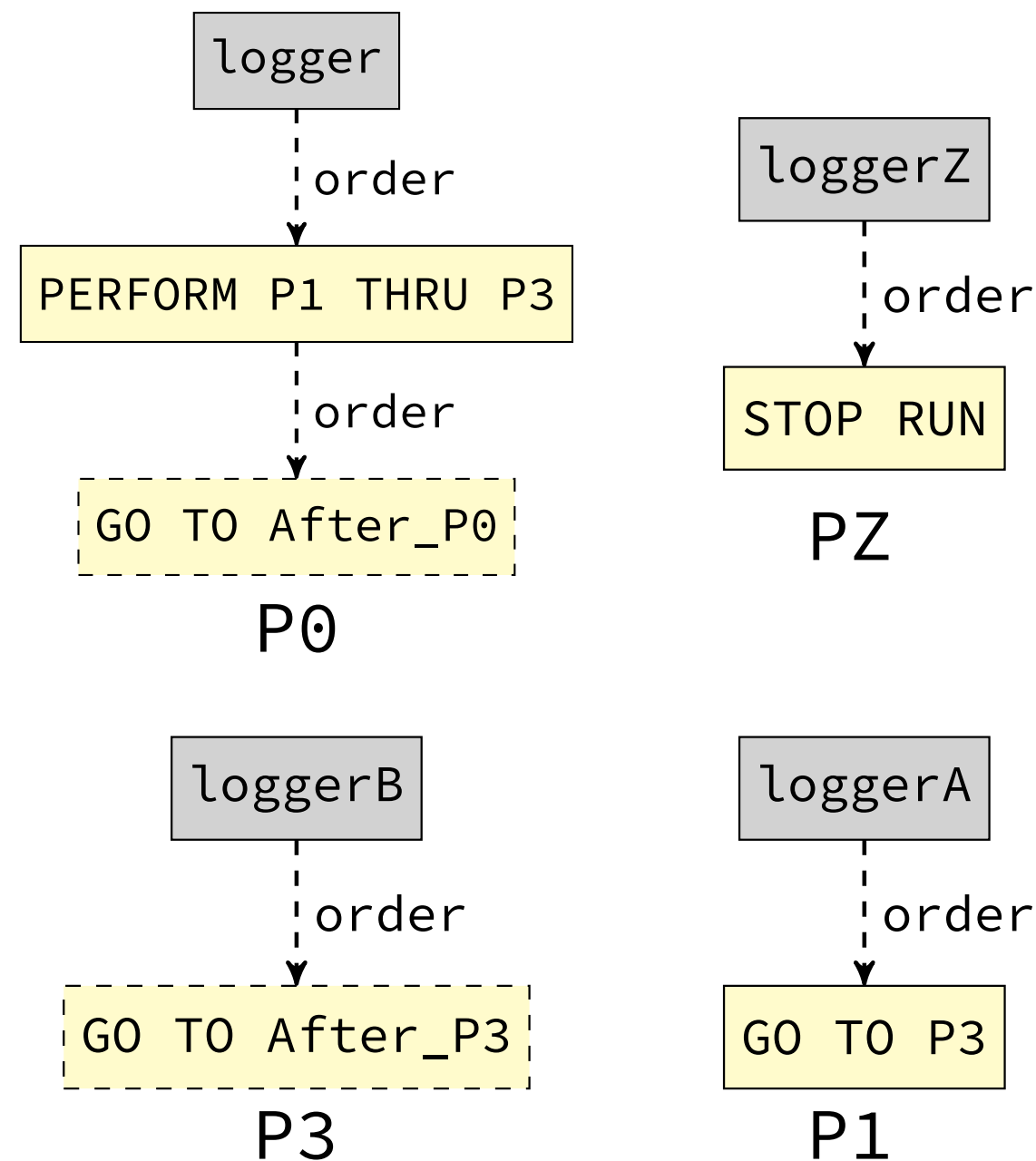
After_P0	PZ
After_P3	P0



# Phase 2: Group Inlining

IDENTIFICATION DIVISION.  
PROGRAM-ID. exc.  
PROCEDURE DIVISION.

```
P0.
    DISPLAY "---- Start ----".
    CALL "logger".
    PERFORM P1 THRU P3.
PZ.
    CALL "loggerZ".
    DISPLAY "---- End ----".
    STOP RUN.
P3.
    CALL "loggerB"
P1.
    CALL "loggerA".
    GO TO P3.
```



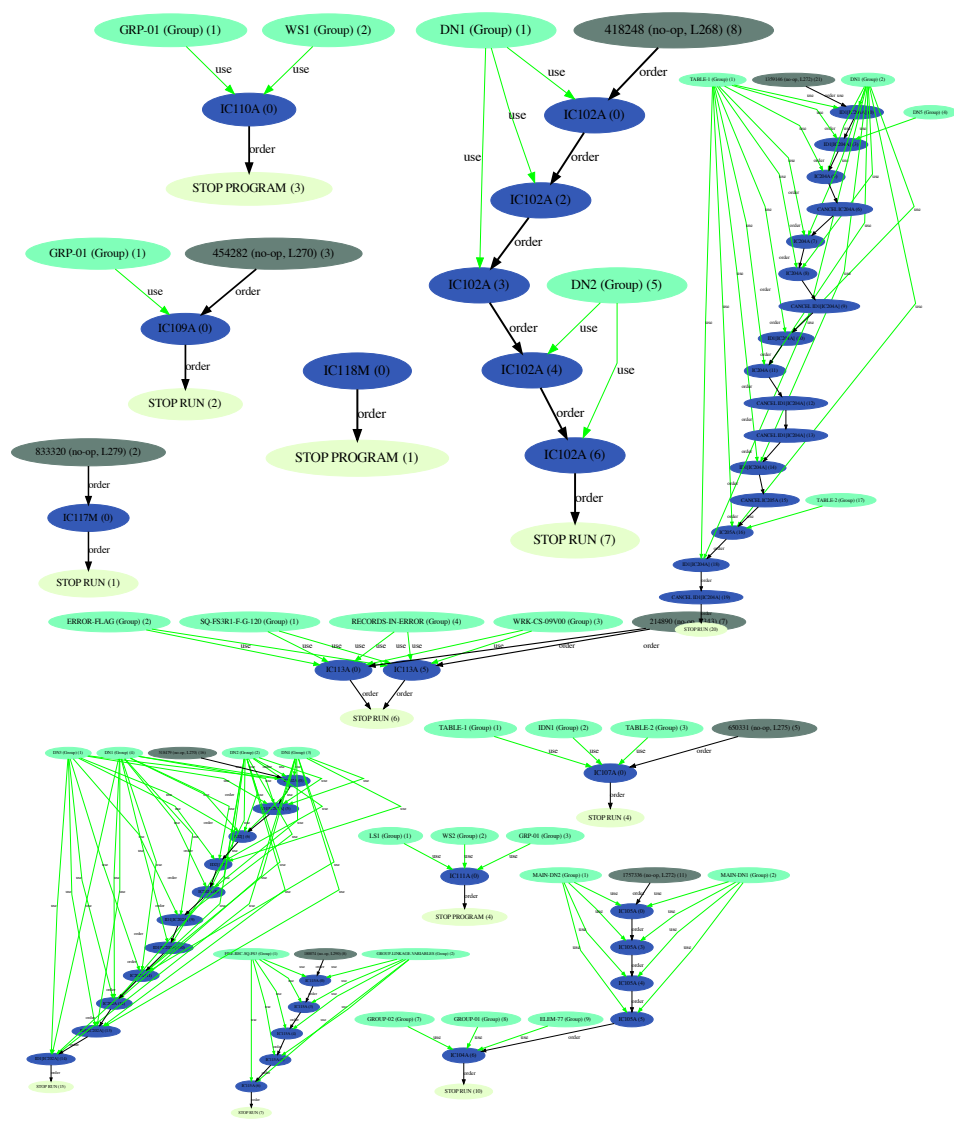
**Implicit jump after paragraph  
Consult jump table**

Jump Table

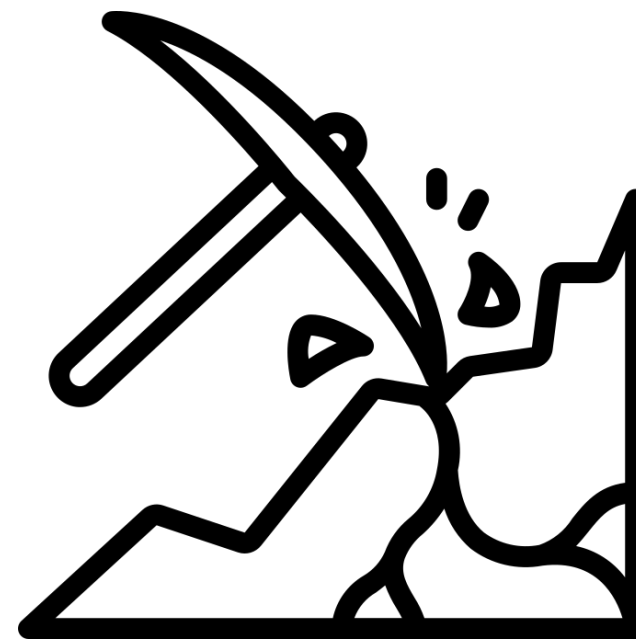
After_P0	PZ
After_P3	P0

# Next Steps: Mining, Visualisation

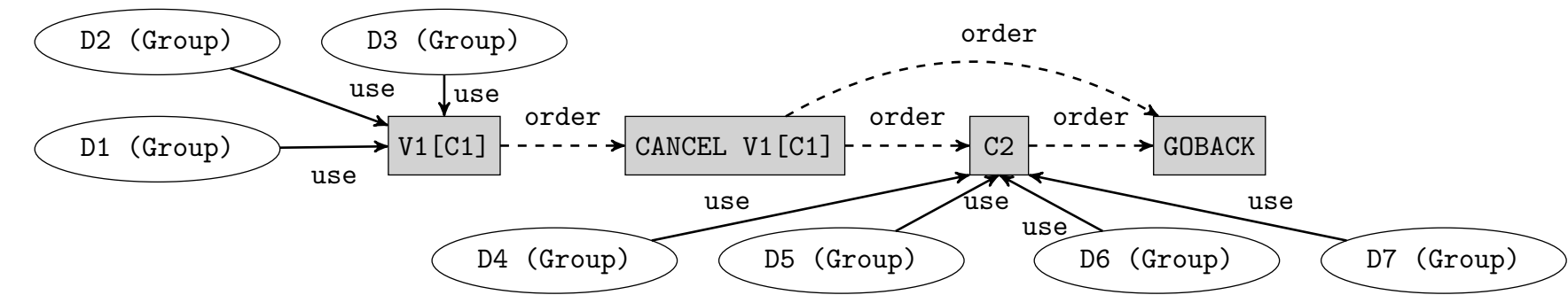
### Groups



### Frequent Subgraph Mining



### Library Usage Patterns



Inspector on a RCPatternCollection

Id	Sup	WSup	Size	Root	Children
0070	024	045	013	ClassInstanceCreation	'type=SimpleType'
0072	007	036	013	ClassInstanceCreation	'type=SimpleType'
0073	007	013	013	ClassInstanceCreation	'type=SimpleType'
0071	007	010	013	ClassInstanceCreation	'type=SimpleType'
0001	011	014	015	ClassInstanceCreation	'type=ParameterizedType'
0002	019	026	018	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0046	007	012	018	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0047	006	009	018	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0068	007	026	019	ClassInstanceCreation	'type=SimpleType'
0048	006	010	020	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0049	007	008	020	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0043	007	010	030	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0041	006	007	023	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0042	006	007	025	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0066	007	009	040	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0021	009	010	041	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0009	008	009	042	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0033	007	008	042	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0035	006	007	042	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0037	006	006	042	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0005	009	010	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0006	009	010	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0008	009	010	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0015	009	010	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0004	006	009	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0034	006	007	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0007	006	006	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0026	006	006	043	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0016	009	010	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0018	009	010	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0023	009	010	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0003	006	009	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0019	008	009	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0024	008	009	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0027	008	009	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0044	006	007	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'
0045	006	007	044	ClassInstanceCreation	'type=SimpleType anonymousClassDeclaration-Anonymous'

Occurrences | EG Tree | EG Source | Pattern | Graph | Raw | Meta

```

class ClsInstCrea
  SimpType
  AnonClasDecl
  MethDecl
  MethDecl
  SimpName
  Block
  Block
  ExprStat
  ExprStat
  SupeMethInvo
  SupeMethInvo
  SimpName
  SimpName
  undo
  redo
  
```

**Future work**

### Visualisation Tool

# Conclusion

## COBOL Library Usage Pattern Mining

Case	Time (s)	# Groups	Avg. size	Max. size	GrouMiner*			BigGroum		
					Time (min)	# P	Time (min)	# Part.	# P	
Case 1	43.8	273	244	4634						
Case 2 (full)	1019	3925	4769	1217422						
Case 2 (limited)	N/A	3573	125	9644						

	# Programs	# KLOC
Case 1	305	662.2
Case 2	3926	22889

**Goal: Understand library usages to estimate modernisation effort**

## Groups for COBOL

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB"
P1.
  CALL "loggerA".
  GO TO P3.
  
```

Single paragraph (often) too small

↓

Inter-paragraph Groups

↓

Inter-paragraph control flow

↓

Graph inlining

**Other challenges**

- Definition of library calls
- Absence of "def" edges
- Iteration through jumps
- Exit calls
- ...

## Phase 1: Intermediate Groum Construction

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB"
P1.
  CALL "loggerA".
  GO TO P3.
  
```

## Phase 2: Groum Inlining

```

IDENTIFICATION DIVISION.
PROGRAM-ID.      exc.
PROCEDURE DIVISION.
P0.
  DISPLAY "---- Start ----".
  CALL "logger".
  PERFORM P1 THRU P3.
PZ.
  CALL "loggerZ".
  DISPLAY "---- End ----".
  STOP RUN.
P3.
  CALL "loggerB"
P1.
  CALL "loggerA".
  GO TO P3.
  
```

From	To
After_P0	PZ
After_P3	P0