

Fiber, a new production system based on GSIM.

Statistics Sweden

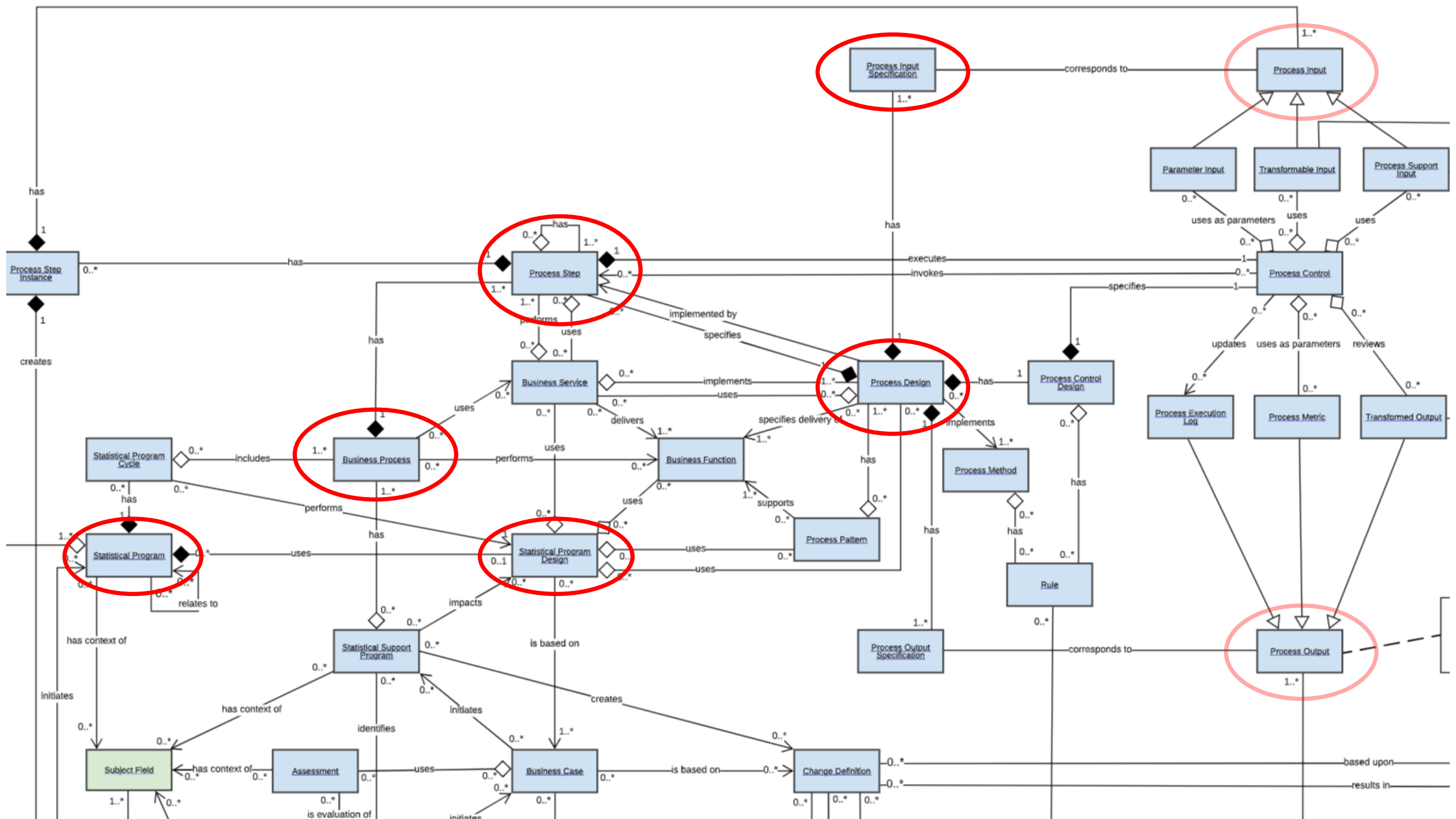
Patrik Wahlgren



Agenda

- Statistical Program Design and Process Design
- Dataset catalog
- Visualization of the dataset catalog with a graph database

Statistical Program design and Process Design





Processes

Search...

- {Statistical Program}
- {Statistical Program Design}
- {Business Process}
- {Process Step}
- {Process Design}**



Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsstätt öppen

+ Dataset Specification

IMP	MAP	VTL
JUS	T	EXP
VAL		

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

Processes

Search...
 {Statistical Program}
 {Statistical Program Design}
 {Business Process}
 {Process Step}
 {Process Design}

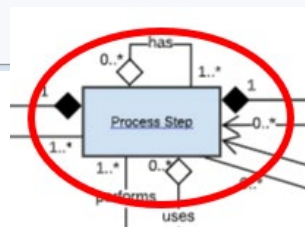
Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”



SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen

Dataset specification. The dataset specification represents **one** dataset that will be created inside this Process Step.

```
graph LR; A[Published statistics and data] --> B[Statistics]; B --> C[Final observation register]; C --> D[Transformed raw data]; D --> E[Rawdata]
```

Behind every datasetspecification there is **one** "Task" and **one** datastructure.

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen

```
graph RL; A[Published statistics and data] --> B[Statistics]; B --> C[Final observation register]; C --> D[Transformed raw data]; D --> E[Rawdata];
```

The "IMP" type has a reference to a provision agreement. The task performs validation according to the agreed datastructure specified in the provision agreement.

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen

The arrows represent ProcessInputSpecification.
The arrows should read "depends on".

```
graph LR; Rawdata --> Transformed[Transformed raw data]; Transformed --> Register[Final observation register]; Register --> Statistics; Statistics --> Published[Published statistics and data];
```

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

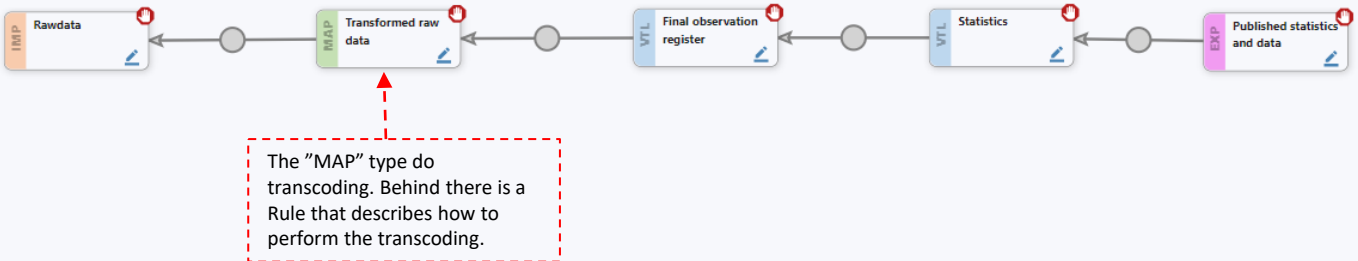
Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen



```
graph LR; Rawdata[TMP Rawdata] --> Transformed[MAP Transformed raw data]; Transformed --> Register[VTL Final observation register]; Register --> Statistics[VTL Statistics]; Statistics --> Published[EXP Published statistics and data];
```

The "MAP" type do transcoding. Behind there is a Rule that describes how to perform the transcoding.

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen

```
graph LR; Rawdata[TMP Rawdata] --> Transformed[MAP Transformed raw data]; Transformed --> VTL_Register[VTL Final observation register]; VTL_Register --> VTL_Statistics[VTL Statistics]; VTL_Statistics --> Published[EXP Published statistics and data]
```

The "VTL" type performs a Rule expressed in VTL-code. (It's possible to execute SAS code with the VTL-instruction "eval").

Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

SCB Fiber Test environment SCB.PreProd.Fiber_20230421.3

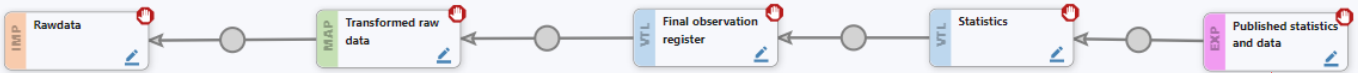
Processes

Search...
{Statistical Program}
{Statistical Program Design}
{Business Process}
{Process Step}
{Process Design}

Designnya Rawdata Transformed raw data Final observation register Statistics Published statistics and dat...

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsätt öppen



```
graph LR; A[IMP Rawdata] --> B[MAP Transformed raw data]; B --> C[VTL Final observation register]; C --> D[VTL Statistics]; D --> E[EXP Published statistics and data];
```

The "EXP" type performs export of the dataset to publishing databases or to creation of SDMX-files.

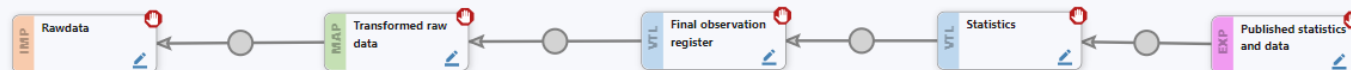
Skala: 0.7, Offset: (64.4/1169.9)

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”

Processes

- Search...
- {Statistical Program}
- {Statistical Program Design}
- {Business Process}
- {Process Step}
- {Process Design}**

Produktionsomgång: Avslutad
Design: Produktionssatt öppen



Datatyp	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	Sträng Identifier	TIME	Hetal Measure	Sträng Attribute	Sträng Attribute		
Datastruktur-komponent namn	FREQ	ADJUSTMENT	REF_AREA	COUNTERPART_AREA	REF_SECTOR	COUNTERPART_SECTOR	CONSOLIDATION	ACCOUNTING_ENTRY	STO	INSTR_ASSET	MATURITY	EXPENDITURE	UNIT_MEASURE	CURRENCY_DENOM	VALUATION	PRICES	TRANSFORMATION	CUST_BREAKDOWN	TIME_FORMAT	DECIMALS	TABLE_IDENTIFIER	UNIT_MULT	COMPLING_ORG	TIME_PERIOD	OBS_VALUE	OBS_STATUS	CONF_STATUS		
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	1995	367205	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	1996	386464	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	1997	395996	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	1998	406911	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	1999	424557	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	2000	435710	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	2001	458807	A	F
A	N	SE	_Z		S13	_Z	_Z	B	B1G	_Z	_Z	_Z	XDC	_Z	S	V	N	_T	P1		2	T0200		6	SE1	2002	459822	A	F

“At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision.”



Processes

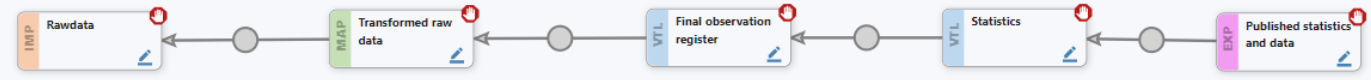
- Search...
- {Statistical Program}
- {Statistical Program Design}
- {Business Process}
- {Process Step}
- {Process Design}

Statistical Program / Statistical Program Design / Business Process / Process Step / Process Design

Produktionsomgång: Avslutad
Design: Produktionsatt öppen

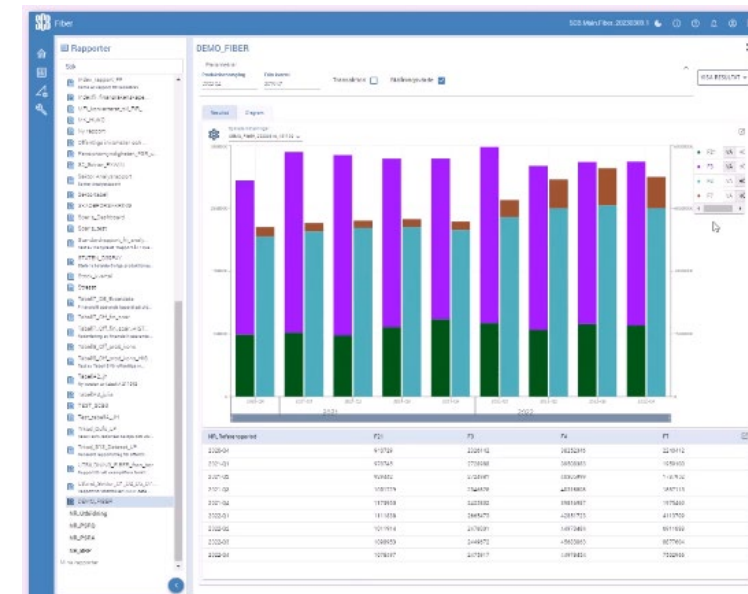
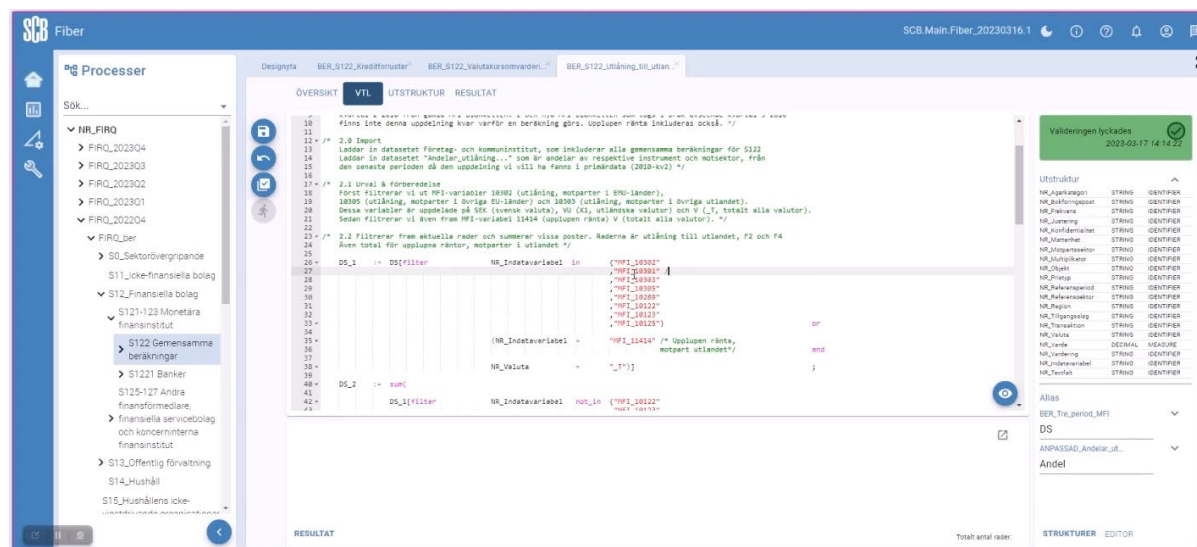
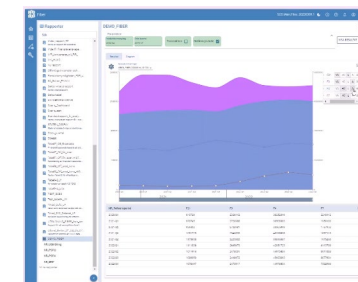
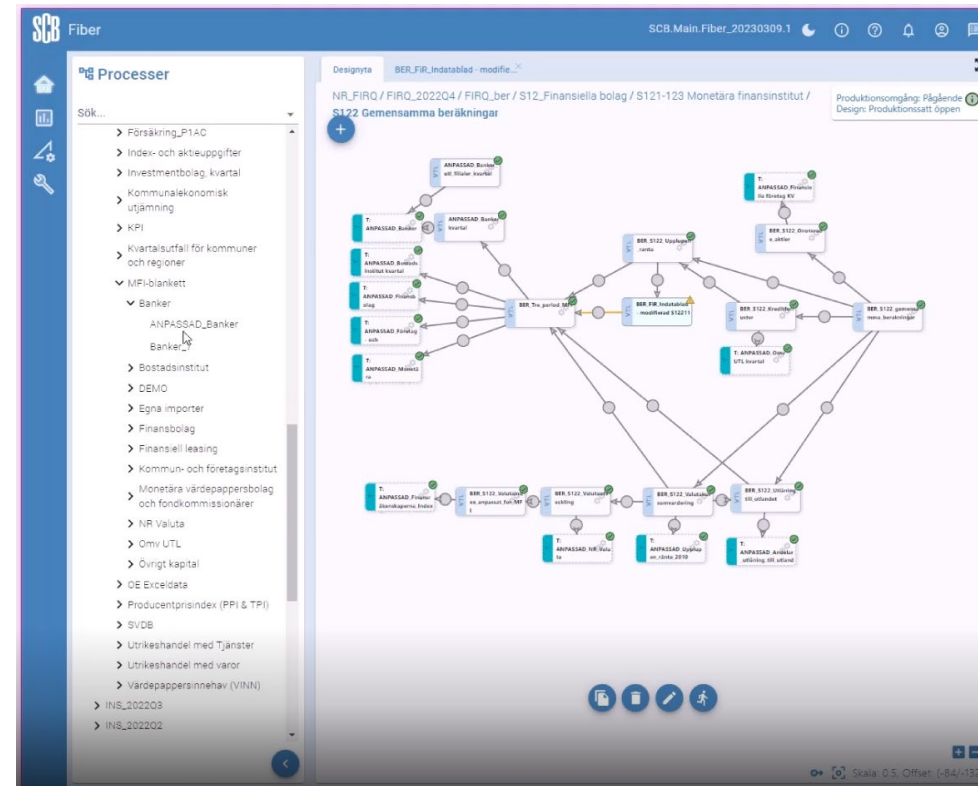
+ Dataset Specification

- IMP
- MAP
- VTL
- JUS
- T
- EXP
- VAL



New production system based on GSIM, "Fiber".

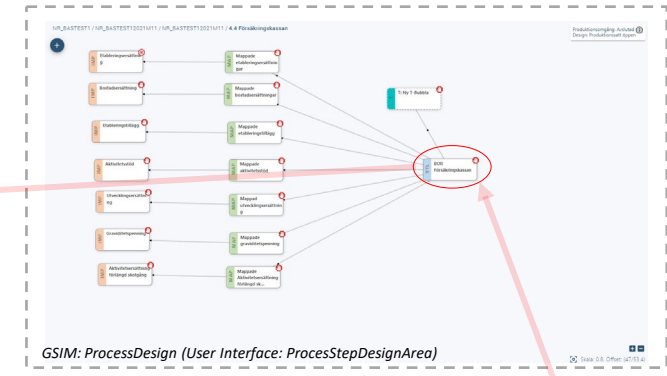
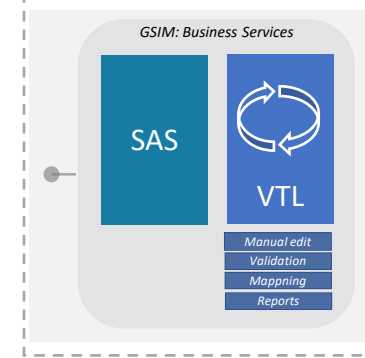
- Import (Provision agreement)
- Calculation
- Validate
- Transcoding
- Manual editing
- Analysis and reports
- Export of datasets for reporting



Dataset catalog

Workflow execution

GSIM: Proces Step Instance



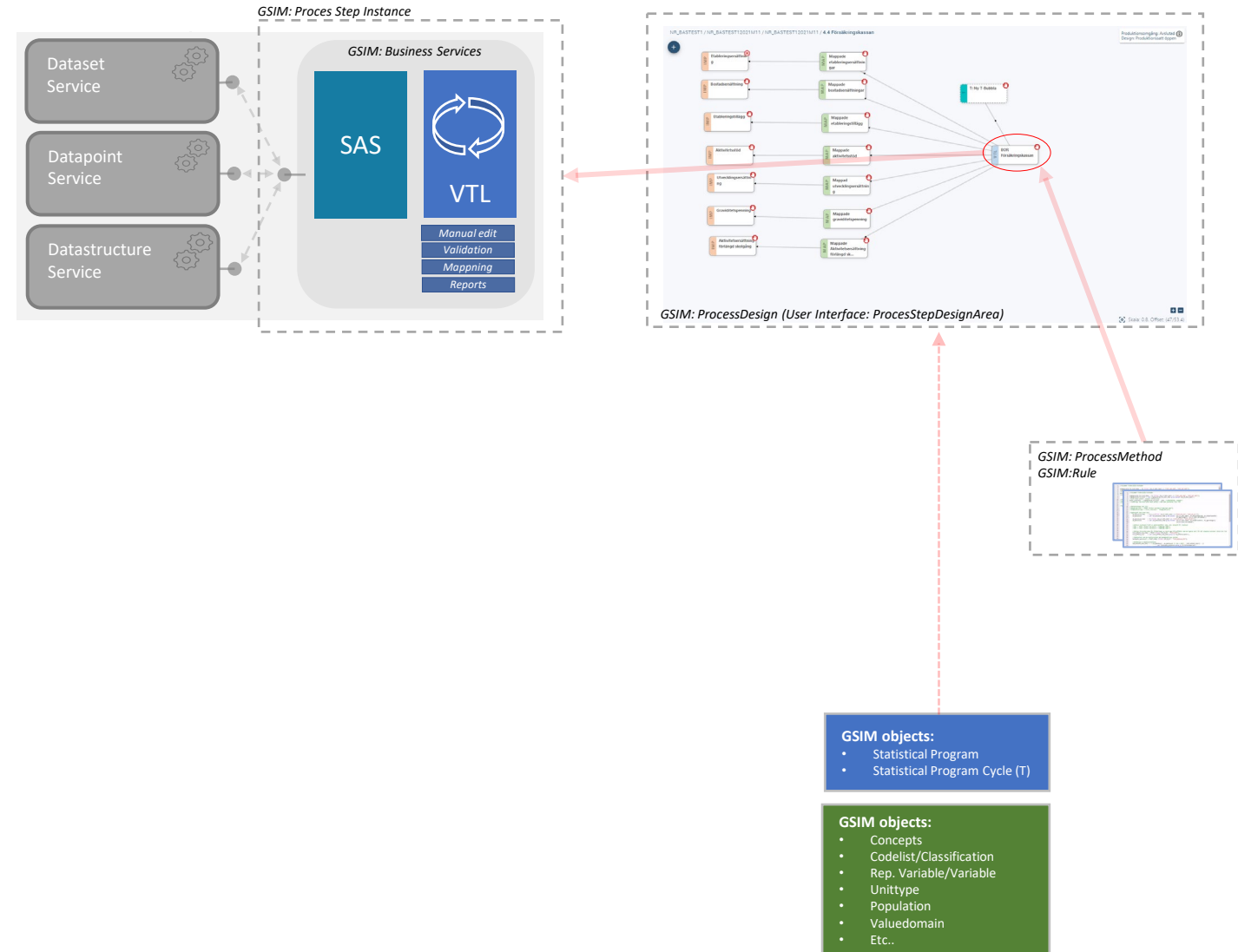
GSIM: ProcessMethod
GSIM:Rule



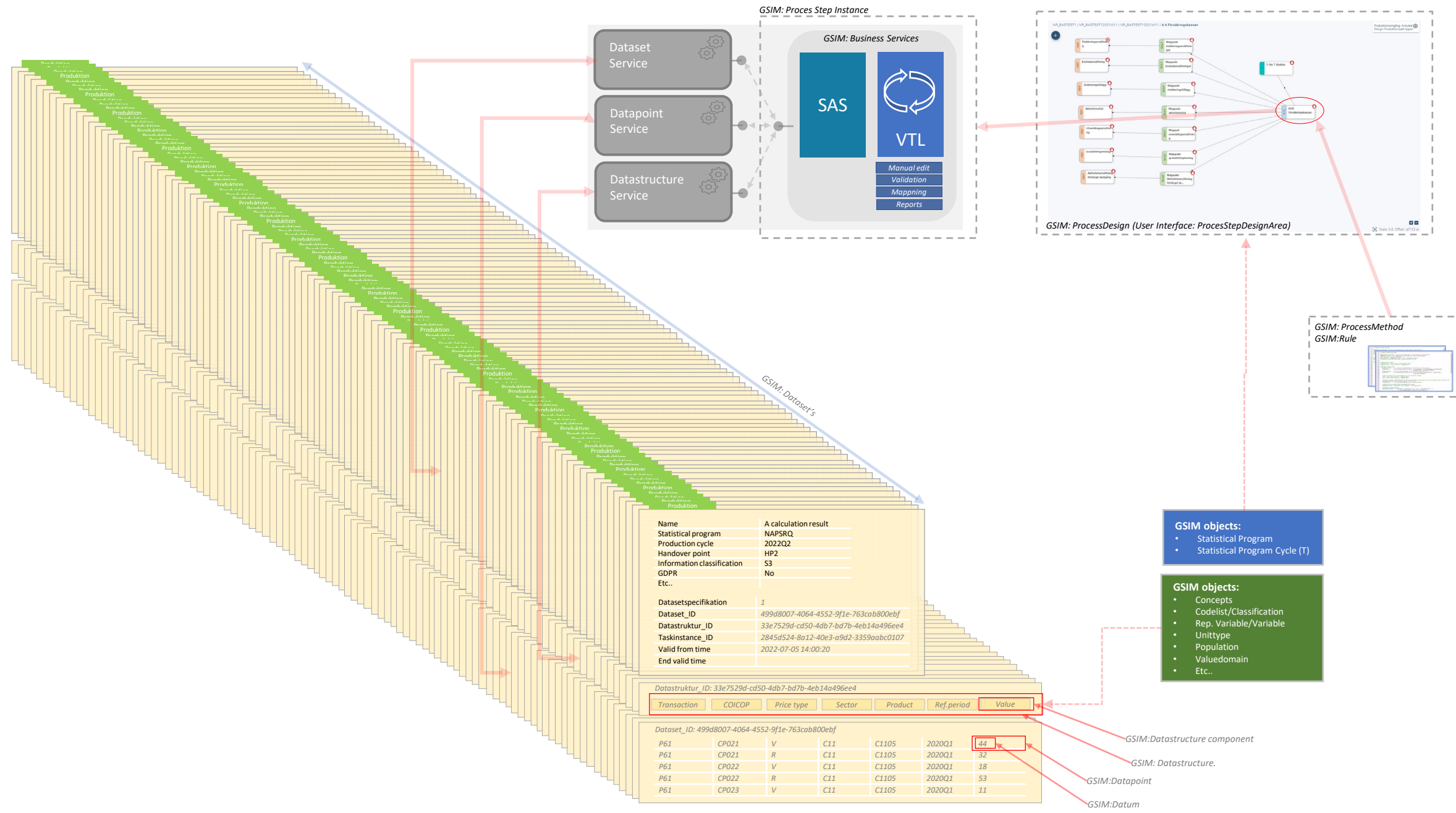
- GSIM objects:**
- Statistical Program
 - Statistical Program Cycle (T)

- GSIM objects:**
- Concepts
 - Codelist/Classification
 - Rep. Variable/Variable
 - Unittype
 - Population
 - Valuedomain
 - Etc..

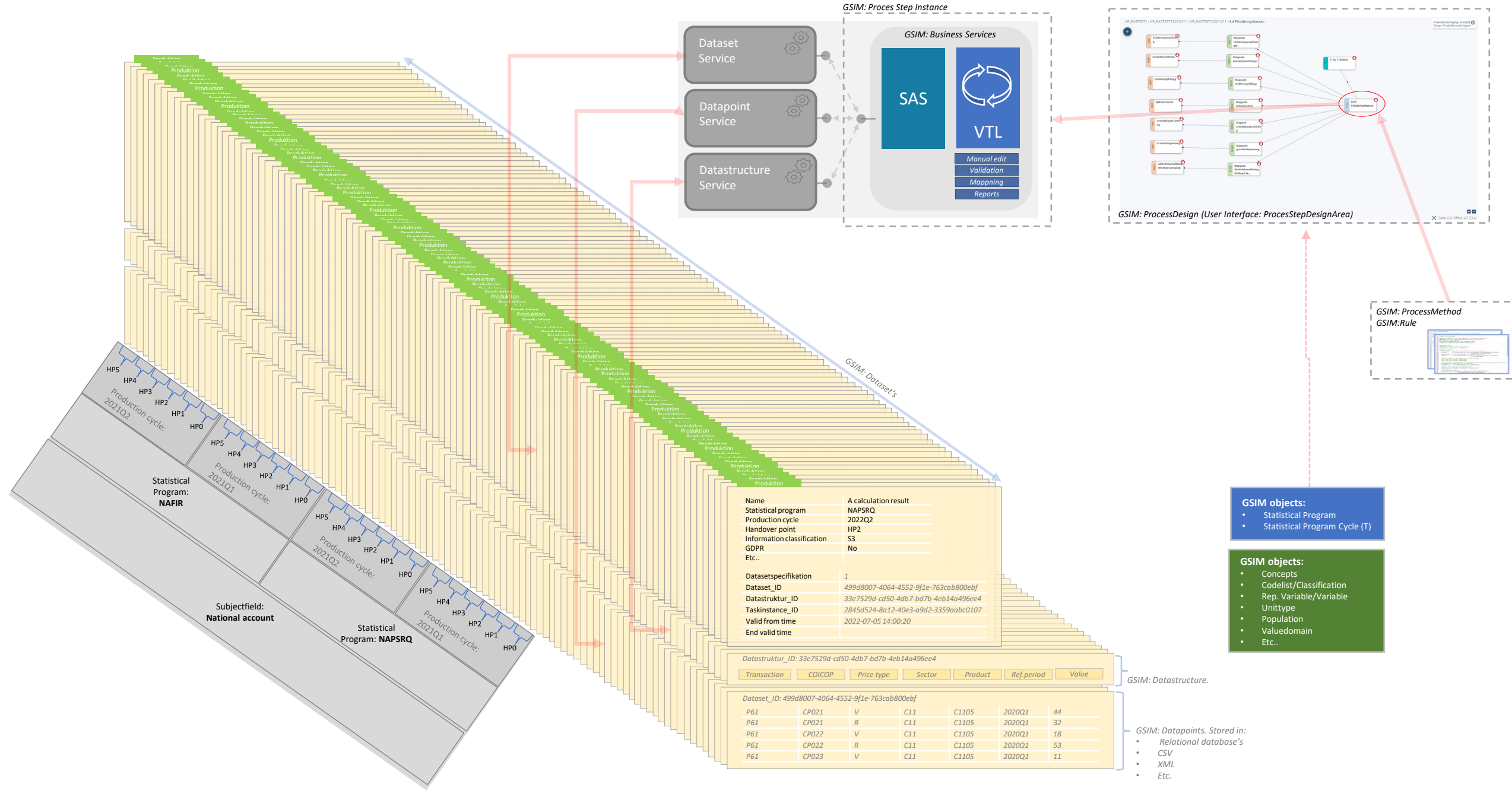
Services (Dataset, Datapoint and Datastructure)



Dataset catalog



Handoverpoints (Checkpoints, Qualitygates or Steady states)



GSIM: Process Step Instance

Dataset Service

Datapoint Service

Datastructure Service

GSIM: Business Services

SAS

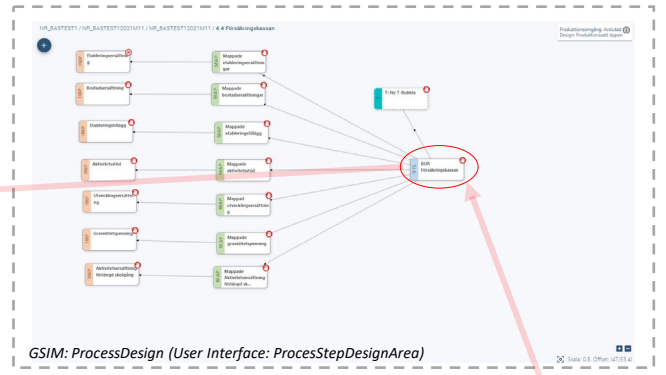
VTL

Manual edit

Validation

Mapping

Reports



GSIM: ProcessMethod
GSIM: Rule

- GSIM objects:**
- Statistical Program
 - Statistical Program Cycle (T)

- GSIM objects:**
- Concepts
 - Codelist/Classification
 - Rep. Variable/Variable
 - Unittype
 - Population
 - Valuedomain
 - Etc..

Name	A calculation result
Statistical program	NAPSRQ
Production cycle	2022Q2
Handover point	HP2
Information classification	S3
GDPR	No
Etc..	
Datasespezifikation	
Dataset_ID	499d8007-4064-4552-9f1e-763cab800ebf
Datastruktur_ID	33e7529d-cd50-4db7-bd7b-4eb14a496ee4
Taskinstance_ID	2845d524-8a12-40e3-a9d2-3359aabc0107
Valid from time	2022-07-05 14:00:20
End valid time	

Datastrukturur_ID: 33e7529d-cd50-4db7-bd7b-4eb14a496ee4

Transaction	COICOP	Price type	Sector	Product	Ref.period	Value
-------------	--------	------------	--------	---------	------------	-------

Dataset_ID: 499d8007-4064-4552-9f1e-763cab800ebf

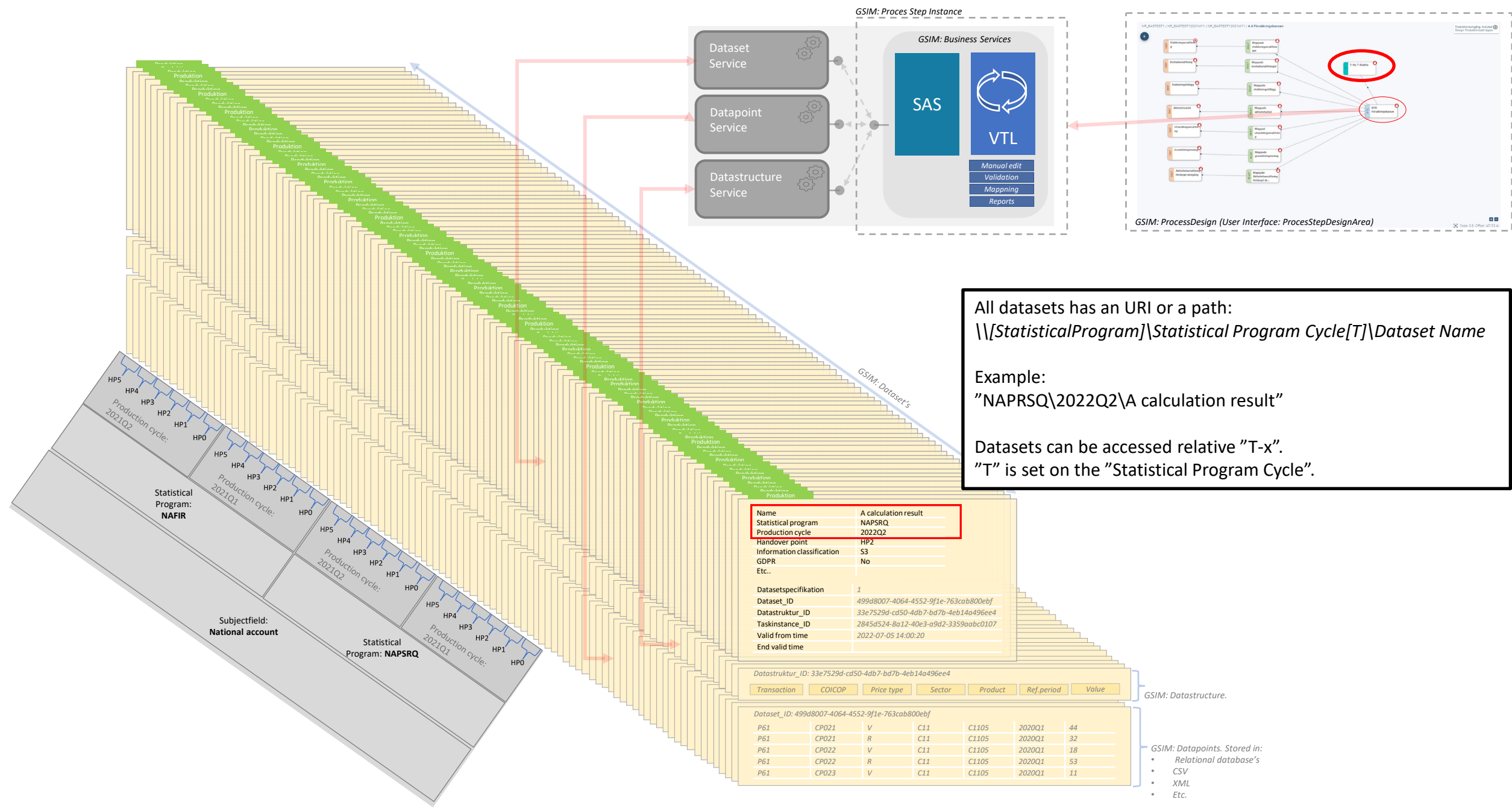
P61	CP021	V	C11	C1105	2020Q1	44
P61	CP021	R	C11	C1105	2020Q1	32
P61	CP022	V	C11	C1105	2020Q1	18
P61	CP022	R	C11	C1105	2020Q1	53
P61	CP023	V	C11	C1105	2020Q1	11

GSIM: Datastructure.

GSIM: Datapoints. Stored in:

- Relational database's
- CSV
- XML
- Etc.

Relative references with "T" for access to datasets



All datasets has an URI or a path:
`\\[StatisticalProgram]\Statistical Program Cycle[T]\Dataset Name`

Example:
 "NAPRSQ\2022Q2\A calculation result"

Datasets can be accessed relative "T-x".
 "T" is set on the "Statistical Program Cycle".

Name	A calculation result
Statistical program	NAPRSQ
Production cycle	2022Q2
Handover point	HP2
Information classification	S3
GDPR	No
Etc..	
Datasespezifikation	1
Dataset_ID	499d8007-4064-4552-9f1e-763cab800ebf
Datastruktur_ID	33e7529d-cd50-4db7-bd7b-4eb14a496ee4
Taskinstance_ID	2845d524-8a12-40e3-a9d2-3359aabc0107
Valid from time	2022-07-05 14:00:20
End valid time	

Datastrukturur_ID: 33e7529d-cd50-4db7-bd7b-4eb14a496ee4						
Transaction	COICOP	Price type	Sector	Product	Ref.period	Value
Dataset_ID: 499d8007-4064-4552-9f1e-763cab800ebf						
P61	CP021	V	C11	C1105	2020Q1	44
P61	CP021	R	C11	C1105	2020Q1	32
P61	CP022	V	C11	C1105	2020Q1	18
P61	CP022	R	C11	C1105	2020Q1	53
P61	CP023	V	C11	C1105	2020Q1	11

- GSIM: Datapoints. Stored in:
- Relational database's
 - CSV
 - XML
 - Etc.

Visualization of the dataset catalog with a graph database

Inte säker | tstw74.scb.intra/browser/ Uppdatera

```
fibermeta$ match (n:DataSet {statisticalProgram:"Produkt- och sektorräkenskaper kvartal", statisticalProgramCycle: "PSR kvartal 2 2020"})-[:Transformerad_från]-(m:DataSet) return n, m limit 1000
```

*(300) DataSet(300) *(432) Transformerad_från(432)

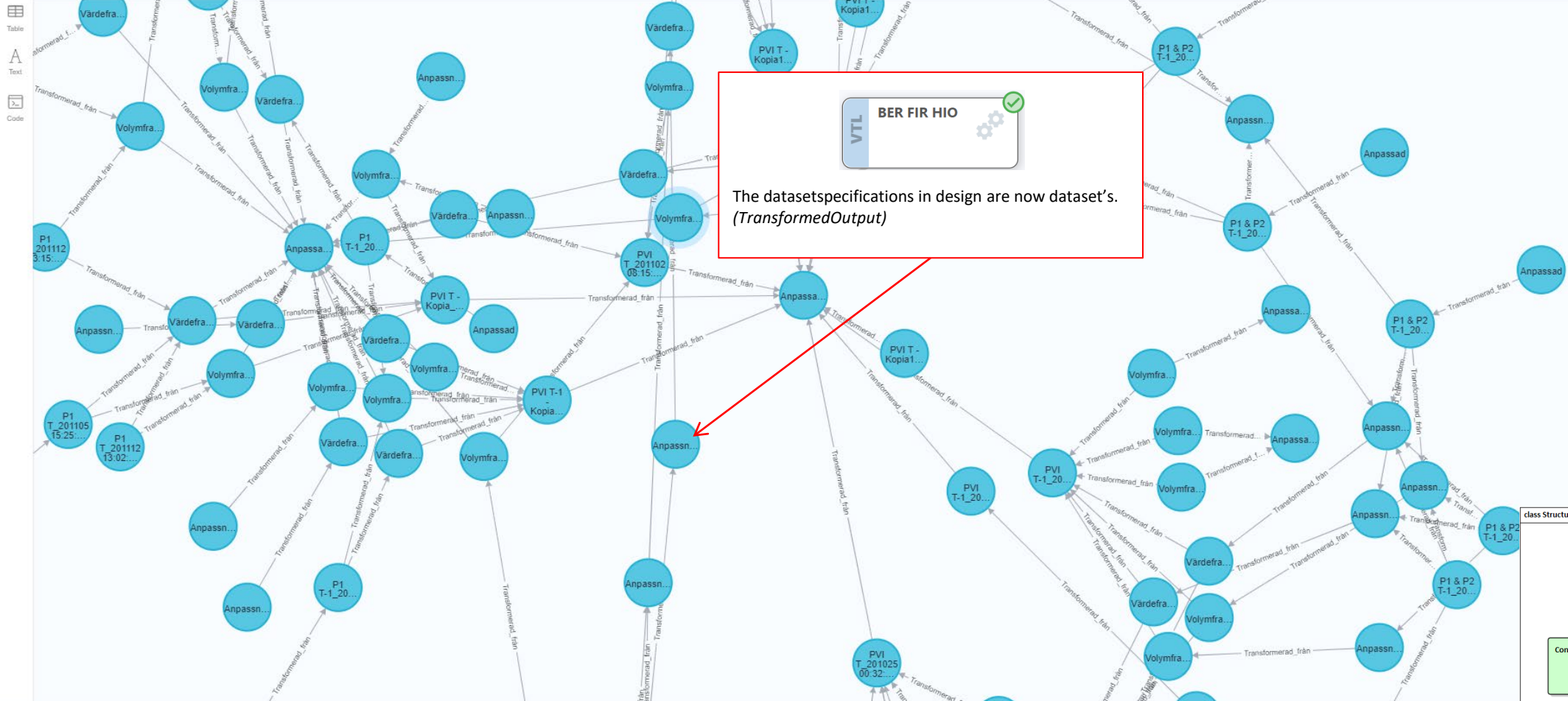
Data Set <id>: 4936 businessProcess: 1. Bearbeta Återkommande executionTime: 2020-11-02 08:16:06 id: 2963 name: Volymframskrivning P1 T-1_201102 08:16:06 statisticalProgram: Produkt- och sektorräkenskaper kvartal statisticalProgramCycle: PSR kvartal 2 2020

*This is a test using a graph database (NEO4J) to display all the dataset in the datasetcatalog
The picture shows a couple of the datasets created in production cycle for quarter 2, 2020 and for the statistical program called NRPSRQ*

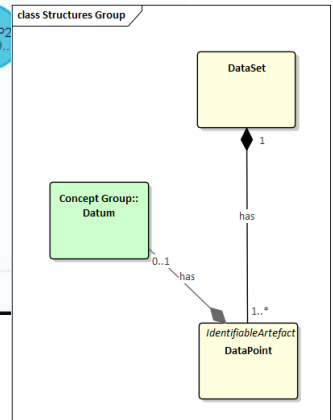
```
fibermeta$ match (n:DataSet {statisticalProgram:"Produkt- och sektorräkenskaper kvartal", statisticalProgramCycle: "PSR kvartal 2 2020"})-[:Transformerad_från]-(m:DataSet) return n, m limit 1000
```

*(300) DataSet(300)

*(432) Transformerad_från(432)

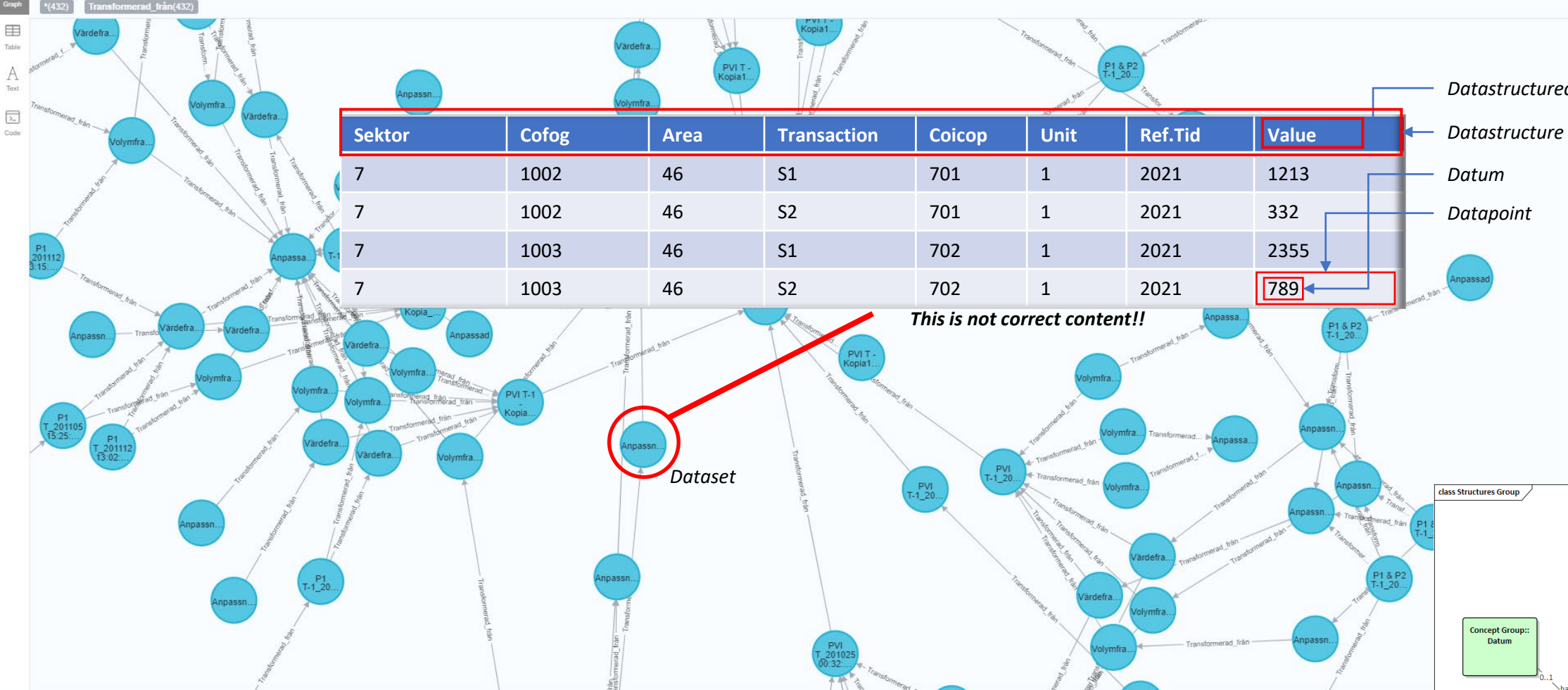


The datasetspecifications in design are now dataset's. (TransformedOutput)




```
fibermeta$ match (n:DataSet {statisticalProgram:"Produkt- och sektorräkenskaper kvartal", statisticalProgramCycle: "PSR kvartal 2 2020"})-[:Transformerad_från]-(m:DataSet) return n, m limit 1000
```

*(300) DataSet(300) *(432) Transformerad_från(432)

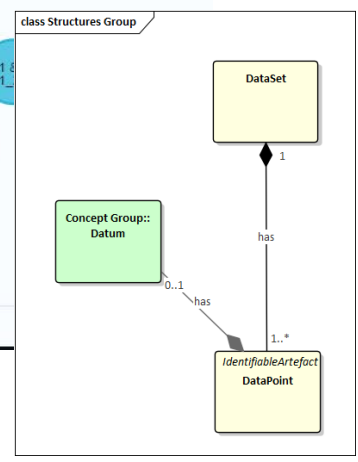


Sektor	Cofog	Area	Transaction	Coicop	Unit	Ref.Tid	Value
7	1002	46	S1	701	1	2021	1213
7	1002	46	S2	701	1	2021	332
7	1003	46	S1	702	1	2021	2355
7	1003	46	S2	702	1	2021	789

Datastructurecomponent
 Datastructure
 Datum
 Datapoint

This is not correct content!!

Dataset



fibermeta\$ match (n:DataSet {statisticalProgram:"Produkt- och sektorräkenskaper kvartal", statisticalProgramCycle: "PSR kvartal 2 2020"}-[:Transformerad_från]-[:DataSet] return n, m limit 1000

*(300) DataSet(300)

*(432) Transformerad_från(432)

```
1 //Kolumnen Finansiella kostnader
2
3 Inngangsvarde_fin_kost_temp := RS [filter InG_KF_RSPK_Radnr in ("RSPK_000_000", "RSPK_000_000")];
4 Inngangsvarde_fin_kost := sum (Inngangsvarde_fin_kost_temp group except InG_KF_RSPK_Radnr);
5 //Efter-reala poster: svaspar (Varde=1156)
6 Efter_ickereala := Inngangsvarde_fin_kost - 1156; /*ÅRSKONING, SWAPPAR*/
7 //Hundering: ska hela icke-reala posten i matrisen levereras till P1?
8
9
10 //Käntekostnader mot 2302
11 //D41g0_mot_2302 := PKOST [filter Variabel="Lämligt namn"];
12 //D41g0_mot_0vriga := Efter_ickereala - D41g0_mot_2302;
13
14 //Egenavgift pensioner 04a2
15 UG_pensskuld_temp := RS_L1 [filter InG_KF_RSPK_Radnr in ("RSPK_00_010", "RSPK_00_011")];
16 UG_pensskuld := sum (UG_pensskuld_temp group except InG_KF_RSPK_Radnr, NR_Referensperiod, UR_Votpartssektor,
17 UR_Agerkategori, InG_KF_RSPK_Verksamhet);
18 UG_pensskuld_temp := RS [filter InG_KF_RSPK_Radnr in ("RSPK_00_010", "RSPK_00_011")];
19 UG_pensskuld := sum (UG_pensskuld_temp group except InG_KF_RSPK_Radnr, NR_Votpartssektor, UR_Agerkategori,
20 InG_KF_RSPK_Verksamhet);
21
22 //Behöver eventuell flytt av pensionsrätter (001, 002) aktuellt år (Värde=0)
23 //001 := PKOST [filter OST_00_PK = "Lämligt namn"];
24 //002 := PKOST [filter OST_00_PK = "Lämligt namn"];
25
26 //Behöver historisk data för förändringar av skuld pga RPS-effekter som korrigeras bort för att simulera konstant ränta över tid
27 hist_effekt_rips_temp := PKOST [filter OST_00_PK = "PKAR_RIPS"];
28 hist_effekt_rips := sum (hist_effekt_rips_temp group by NR_Referensperiod);
29
30 //Käntesatsen som ska multipliceras med genomsnittliga skulden
31 Rantesats_pensskuld := PKOST_HNDL [filter OST_00_PK = "PKAR_000004_P01"];
32
33 //Beräkning av kapitalavkastning
34 kap_avkast_pens_04a2 := ((UG_pensskuld - UG_pensskuld /* + 001 + 002 */ - hist_effekt_rips*2) / 2)
35 := sum (Rantesats_pensskuld group by NR_Referensperiod);
```

GSIM: Rule

Dataset

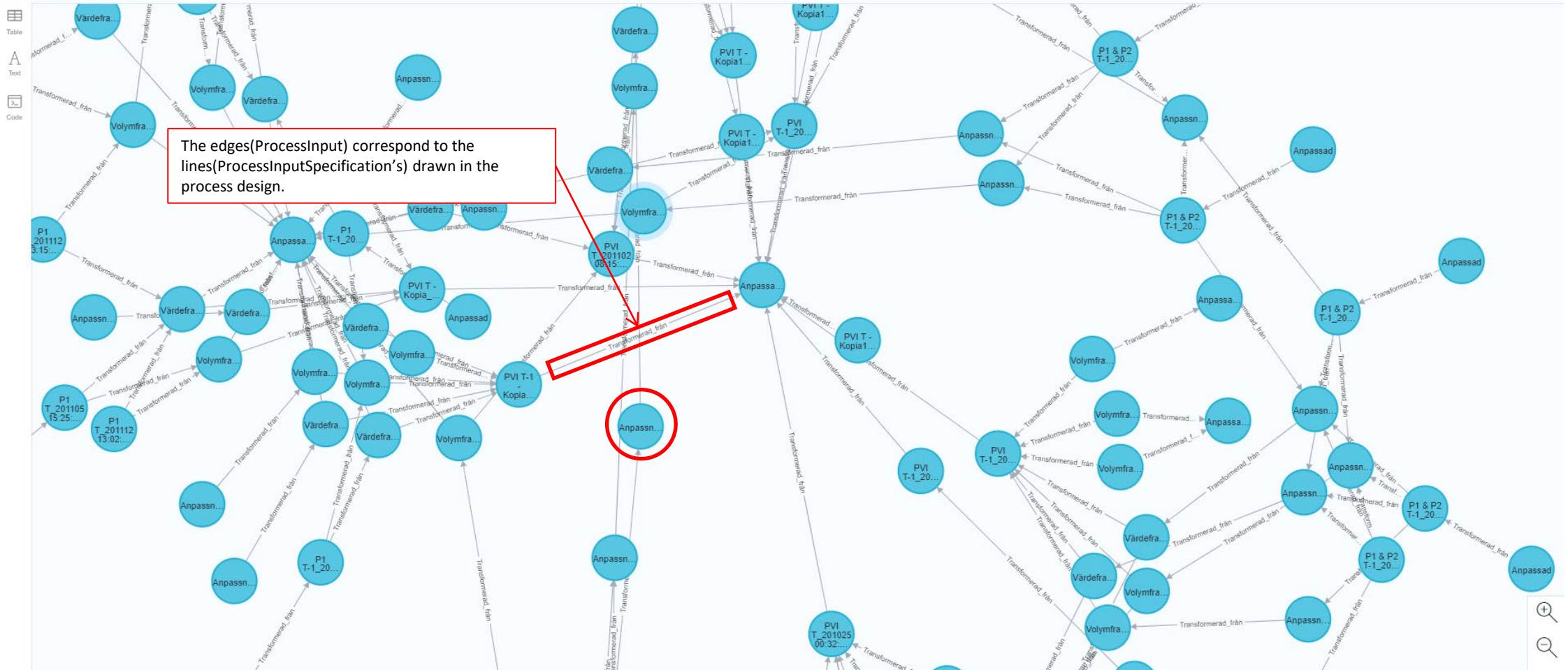
Data Set <id>: 4936 businessProcess: 1. Bearbeta Återkommande executionTime: 2020-11-02 08:16:06 id: 2963 name: Volymframskrivning P1 T-1_201102 08:16:06 statisticalProgram: Produkt- och sektorräkenskaper kvartal statisticalProgramCycle: PSR kvartal 2 2020

You can see the rule that created this dataset.


```
fibermeta$ match (n:DataSet {statisticalProgram:"Produkt- och sektorräkenskaper kvartal", statisticalProgramCycle: "PSR kvartal 2 2020"})-[:Transformerad_från]-(m:DataSet) return n, m limit 1000
```

*(300) DataSet(300)

*(432) Transformerad_från(432)



Patrik Wahlgren

Business architect

Methodology and architecture governance and Enterprise architecture
SCB, Statistics Sweden

Mail: patrik.wahlgren@scb.se