



# Pentaho

## En alternativ integrationsplattform



Pentaho Data Integration  
Previously Kettle

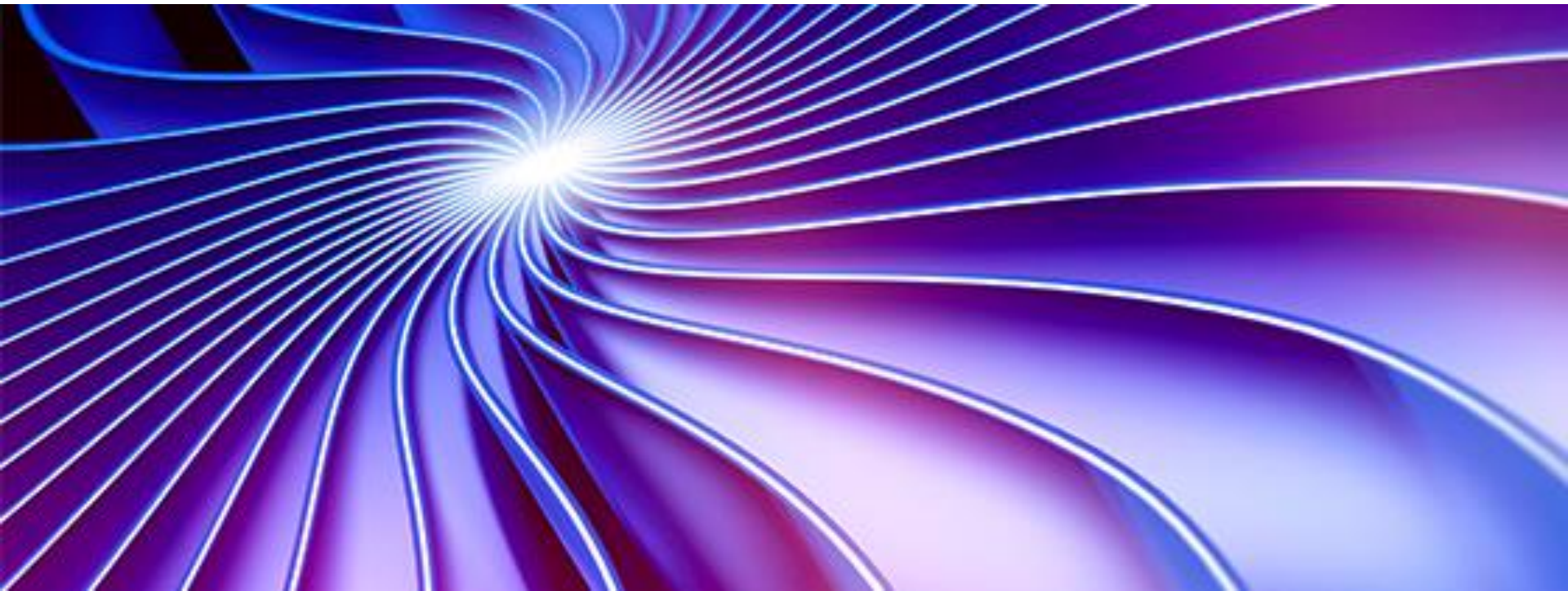


Mittuniversitetet  
MID SWEDEN UNIVERSITY

# Projektet Forskningsdatabas

Converis – <https://clarivate.com/products/converis/>

En bakgrund till varför...



## Mittuniversitetets ”nya” forskningsdatabas

- Converis – tidigare Avedas, Thomson Reuters, nu Clarivate
- Består av förenklat av Converis java applikation, Glassfish java server, Postgres SQL databas
- Version 5.10.4
- Vår första Converis version (3) driftades av Avedas i Tyskland
- Version 4 driftades av Miun men administrerades av Avedas i Tyskland
- Version 5 driftas och administreras av Miun. Supportavtal med Clarivate.

## Projektet – innan start


- Förstudie kom fram till – fortsatt med Converis
- Det skulle vara enkelt att migrera från version 4 till 5
- Skulle inte ta så lång tid
- Inga större IT resurser skulle tas i anspråk

## Projektet - resultat


- Har överskridit projekttiden samt budget
- Varit komplicerat
- Leverantören har på sin sida förbrukat 3 projektteam, dvs vi har gjort omstart 3 gånger
- Converis blev sålt under projekttiden från Thomson Reuters till Clarivate
- Tagit allt för mycket av interna IT-resurser





# Dashboard Forskare

Research database | My Tools ▾ | Help ▾ | Miun Employee: OLANDER, OVE ▾ | 

- Dashboard
- Publications
- CV Activities
- Notifications
- Statistics



**OVE OLANDER**  Edit

 LIT

e: Ove.Olander@miun.se

m: 070-5975787 p: privat: 0705760223 arb: 010-1427887 mobil: 070-5975787

### Things to do

No pending things to do

### Recently edited

OLANDER, OVE  
Person | Last edited on 10/10/2017 at 07:25

# Dashboard ConfigAdmin



Mittuniversitetet

MID SWEDEN UNIVERSITY

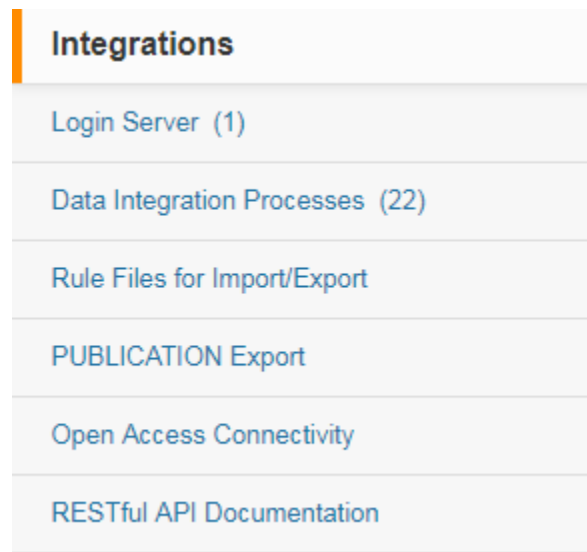
**Publikationer**

- Projects
- CV-aktiviteter
- Graduations
- Persons
- Organisations
- Classifications
- Meddelanden
- Statistik
- Admin settings**
- Administration
- Configuration

<b>System Resources</b>	<b>Data Model</b>	<b>Template management</b>
ChoiceGroups (163)	Entities (69)	Entities (367 + 253 + 17)
ChoiceGroupvalue (1612)	Link Entities (385)	Reports (8)
Constants (149)	AttributeDefinitions (4410)	CVs (2)
Timers (27)	Chains of Link Entities (26)	PDF Template
	Validation Overview	
<b>Rights Management</b>	<b>Module Settings</b>	<b>Function settings</b>
User Roles (29)	Grouping of Entity Icons (8)	Save As (9)
Pentaho roles mapping	Visibility Rule (69)	Filter Settings (216)
Create Rights (149)	Entity search (414)	CSL (72)
Dynamic templates per user role (0)	Header Search (1)	Copy relation rules (3)
Basic Rights (217)	Statistics Charts (10)	Export config
Special Rights (347)		Claimings (6)
Workflow Matrix (383)		
<b>Messaging</b>	<b>Quality and Data management</b>	<b>Integrations</b>
Deadline Reminder (4)	Update Search Indices (All Entities)	Login Server (1)
Attribute / Status change (11)	Configuration Change Log	Data Integration Processes (22)
Link Entity Change (20)	Action log	Rule Files for Import/Export
Things To Do (53)	Saved Searches List (0)	PUBLICATION Export
Search Profile		Open Access Connectivity
		RESTful API Documentation

## Pentaho 5.2

- Är den version som Converis 5.9-10 använder....
- Det var i projektet jag insåg att det var bara att lära sig detta för att kunna sy ihop våra integrationsbehov inom rimlig tid.
- Converis har flera verktyg i sin integrationslåda





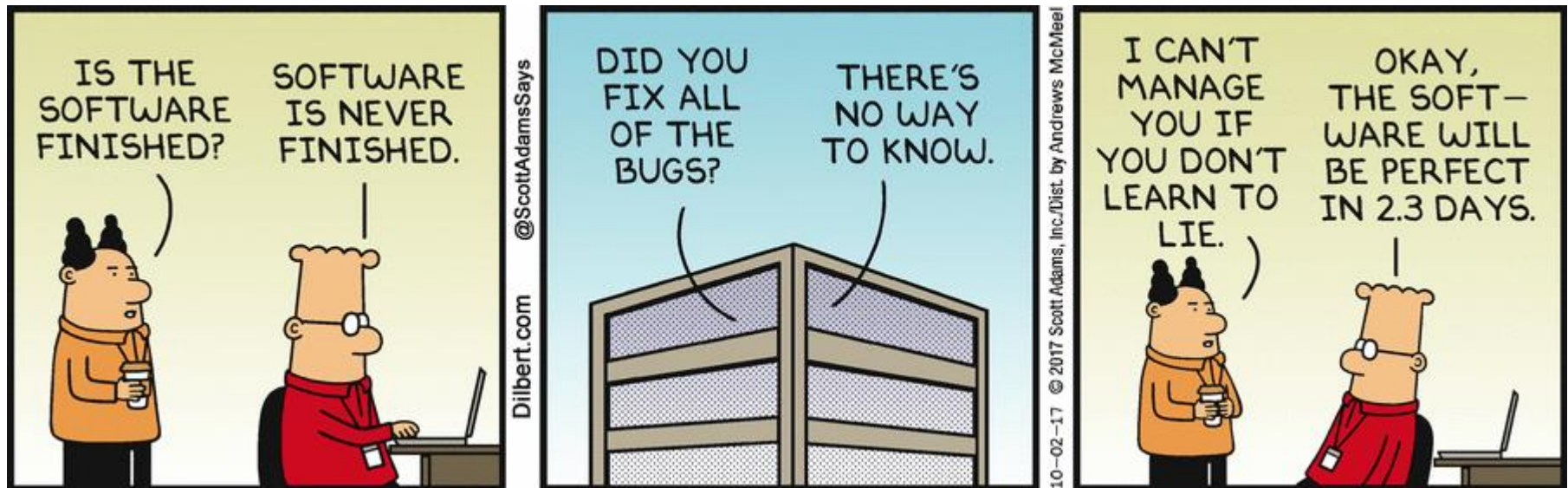


## Integrationer i projektet "Forskningsdatabasen"

- Personal - Primula data
- Organisation - Primula data
- Publikationer - DIVA data
- Forskarstudier – Ladok data



# Pause....

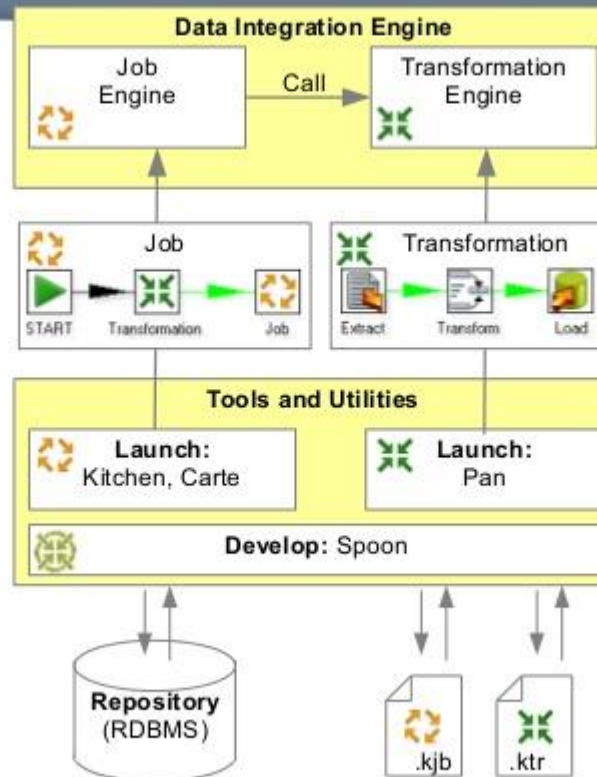


## Vad är då Pentaho för något?

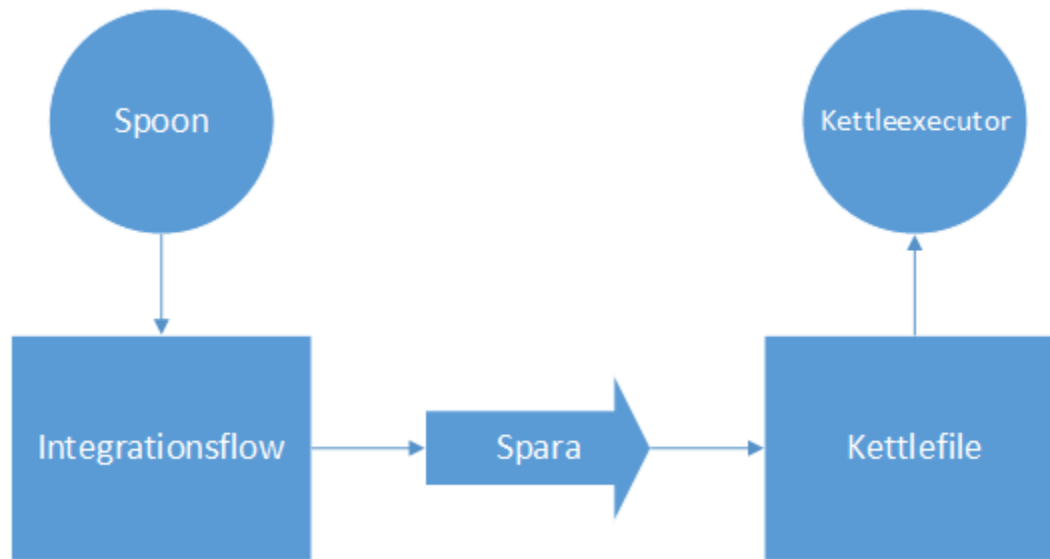
- Det är bl.a. en integrationsplattform
- Det finns en Enterprise (naturligtvis) samt en Community Edition
- Nuvarande version 7.x
- Det är del av en produktpaketering från Hitachi / Community
- PDI Pentaho Data Integration, BI Business Intelligence + lite till
- I Converisprojektet användes PDI och därur en del som kallas Spoon för att utveckla delar av de ingående integrationerna.



# Kettle Architecture



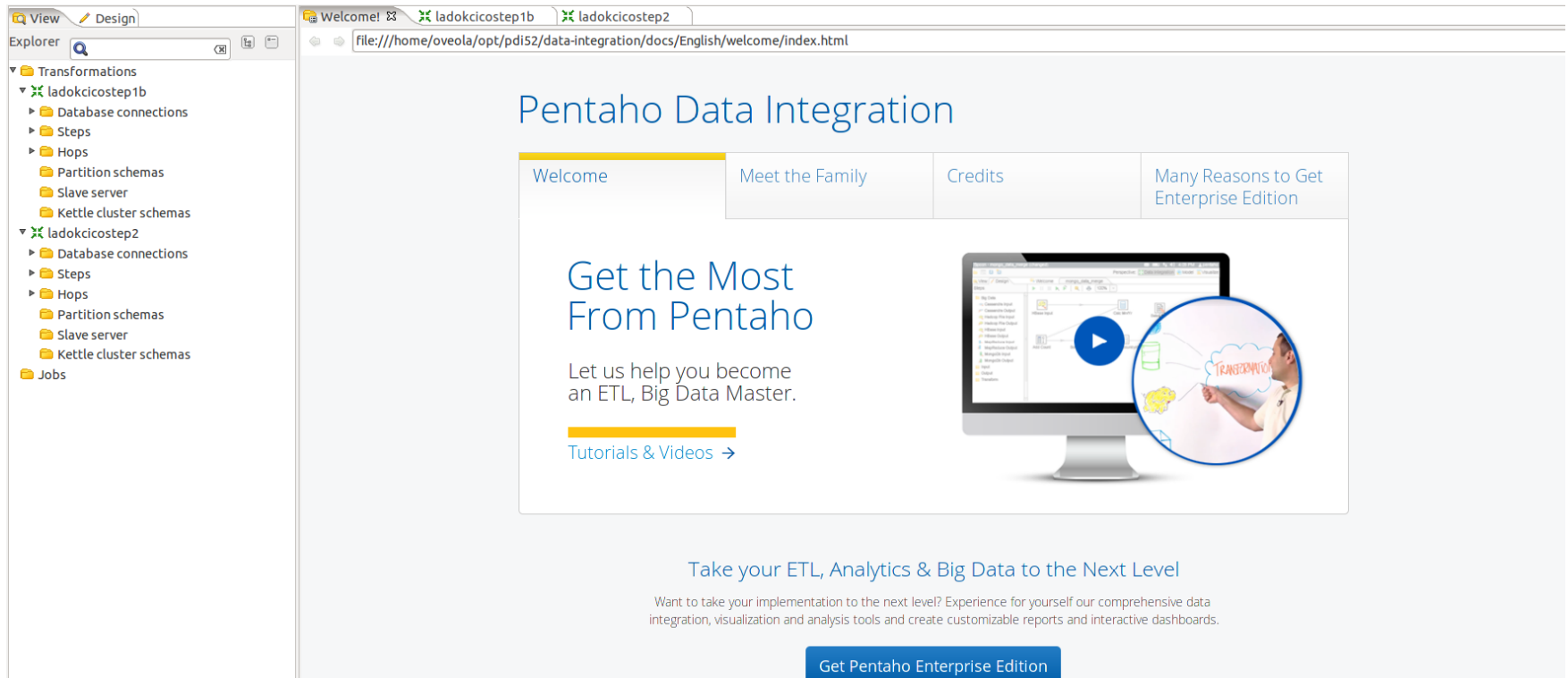
Så här såg jag det..



## PDI

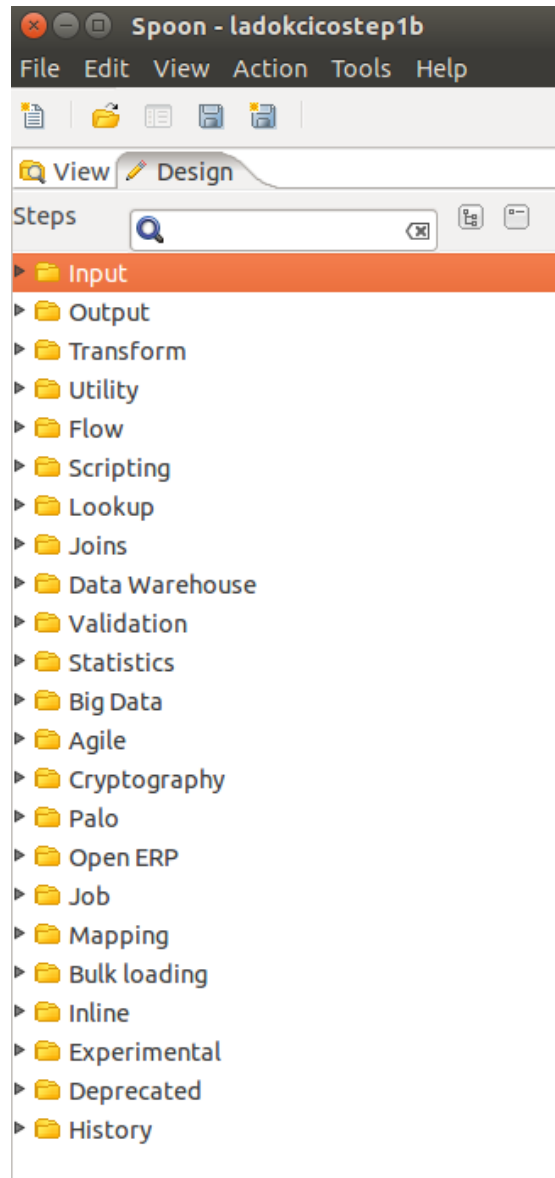
- Spoon - Grafisk Utvecklingsmiljö för integration/transformation. Java app. "Drag och släpp" .
- För Converis version PDI 5.2, vi testar version 7.x som en del i integrationsplattformen.
- Transformationsfiler (ktr) XML format
- Jobfiler (kjb) XML format.
- "kitchen" används för att exekvera kjb filer i shell/cmd
- "pan" används för att exekvera ktr filer i shell/cmd

# Welcome!



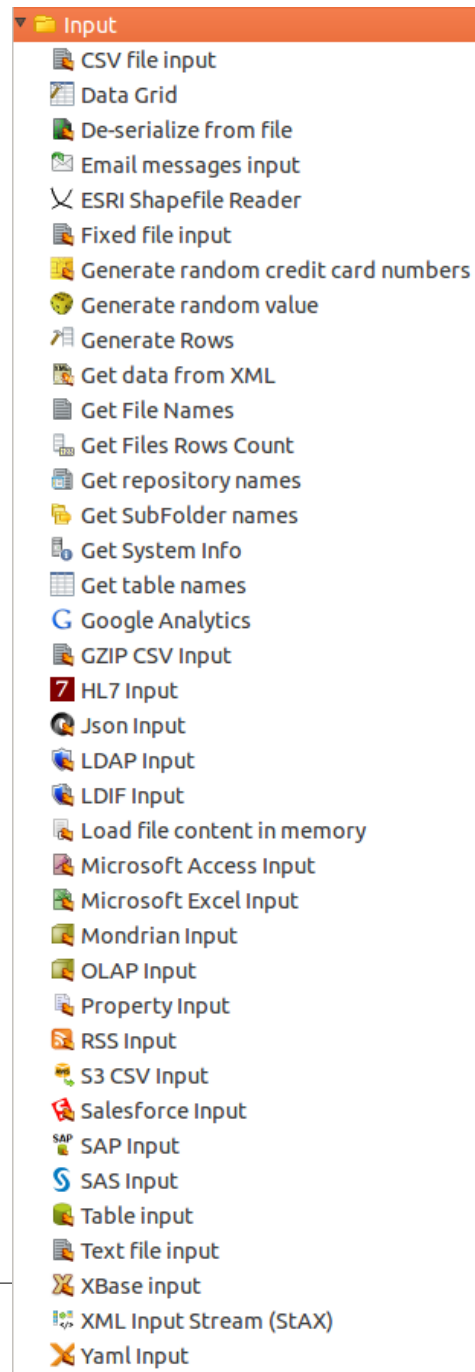
The screenshot shows the Pentaho Data Integration (PDI) software interface. On the left is the 'Explorer' pane with a tree view of project folders: 'Transformations', 'ladokcicostep1b', and 'ladokcicostep2'. The main window displays a 'Welcome' page with the title 'Pentaho Data Integration'. The page has four navigation tabs: 'Welcome' (active), 'Meet the Family', 'Credits', and 'Many Reasons to Get Enterprise Edition'. The main content area features the heading 'Get the Most From Pentaho' and the text 'Let us help you become an ETL, Big Data Master.' Below this is a link for 'Tutorials & Videos' with a right-pointing arrow. To the right is an image of a computer monitor displaying a data flow diagram, with a circular inset showing a person pointing at a whiteboard with the word 'TRANSFORMATIONS' written on it. At the bottom of the page, there is a blue button that says 'Get Pentaho Enterprise Edition'.

# Design

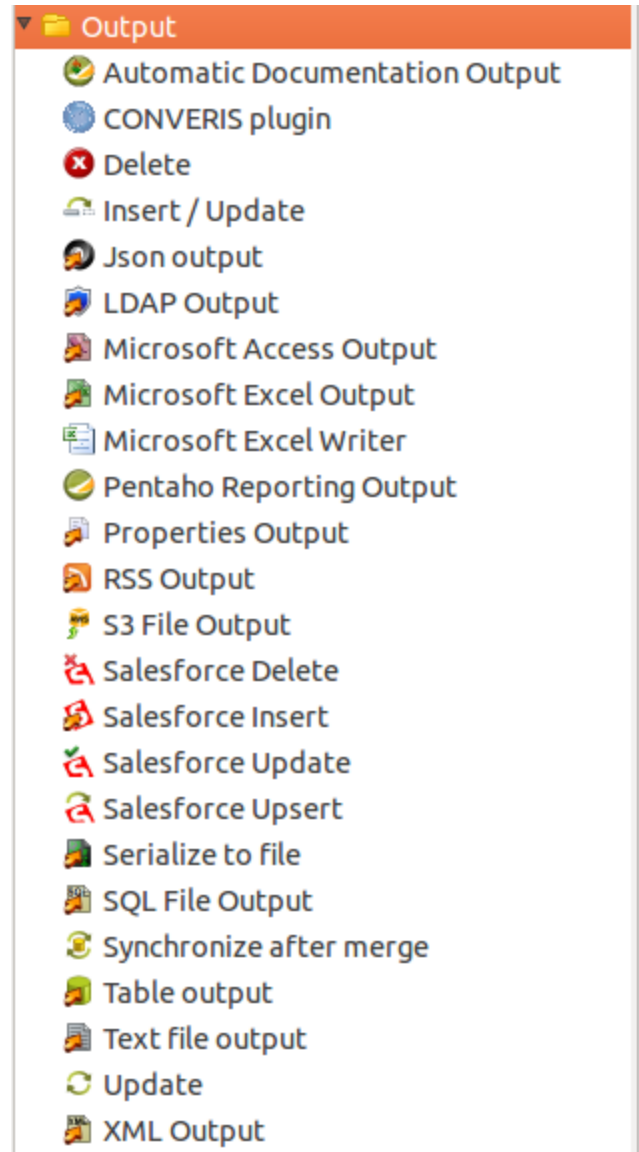




# Input



# Output

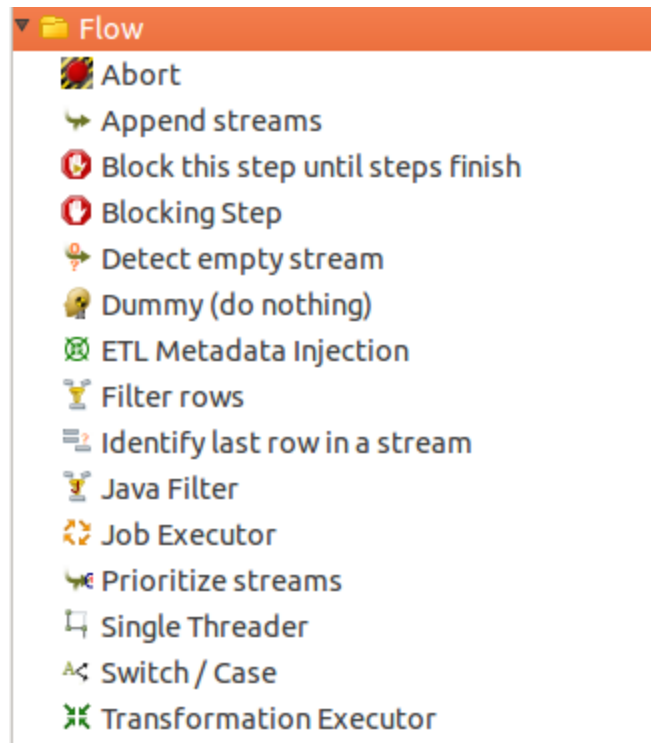


# Transform

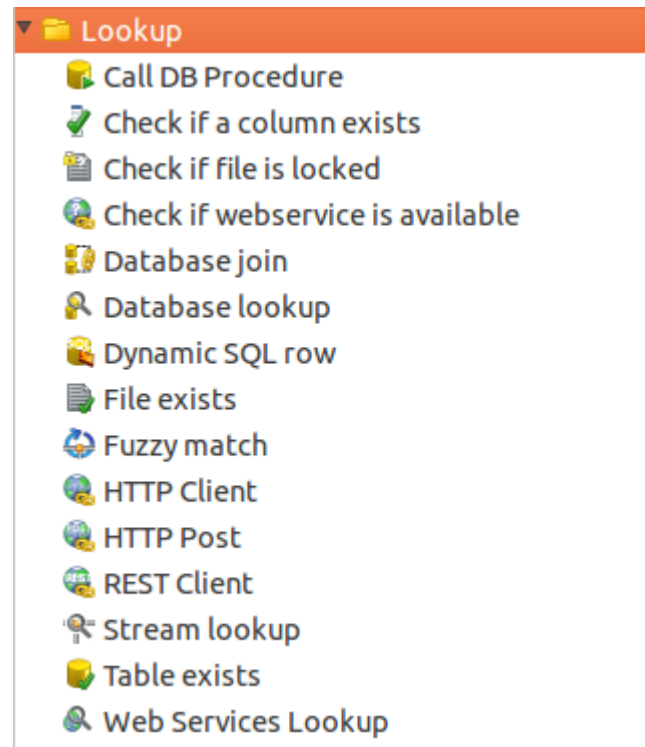
- Transform
  - Add a checksum
  - Add constants
  - Add sequence
  - Add value fields changing sequence
  - Add XML
  - Calculator
  - Closure Generator
  - Concat Fields
  - Example plugin
  - Get ID from slave server
  - Number range
  - Replace in string
  - Row denormaliser
  - Row flattener
  - Row Normaliser
  - Select values
  - Set field value
  - Set field value to a constant
  - Sort rows
  - Split field to rows
  - Split Fields
  - String operations
  - Strings cut
  - Unique rows
  - Unique rows (HashSet)
  - Value Mapper
  - XSL Transformation



# Flow



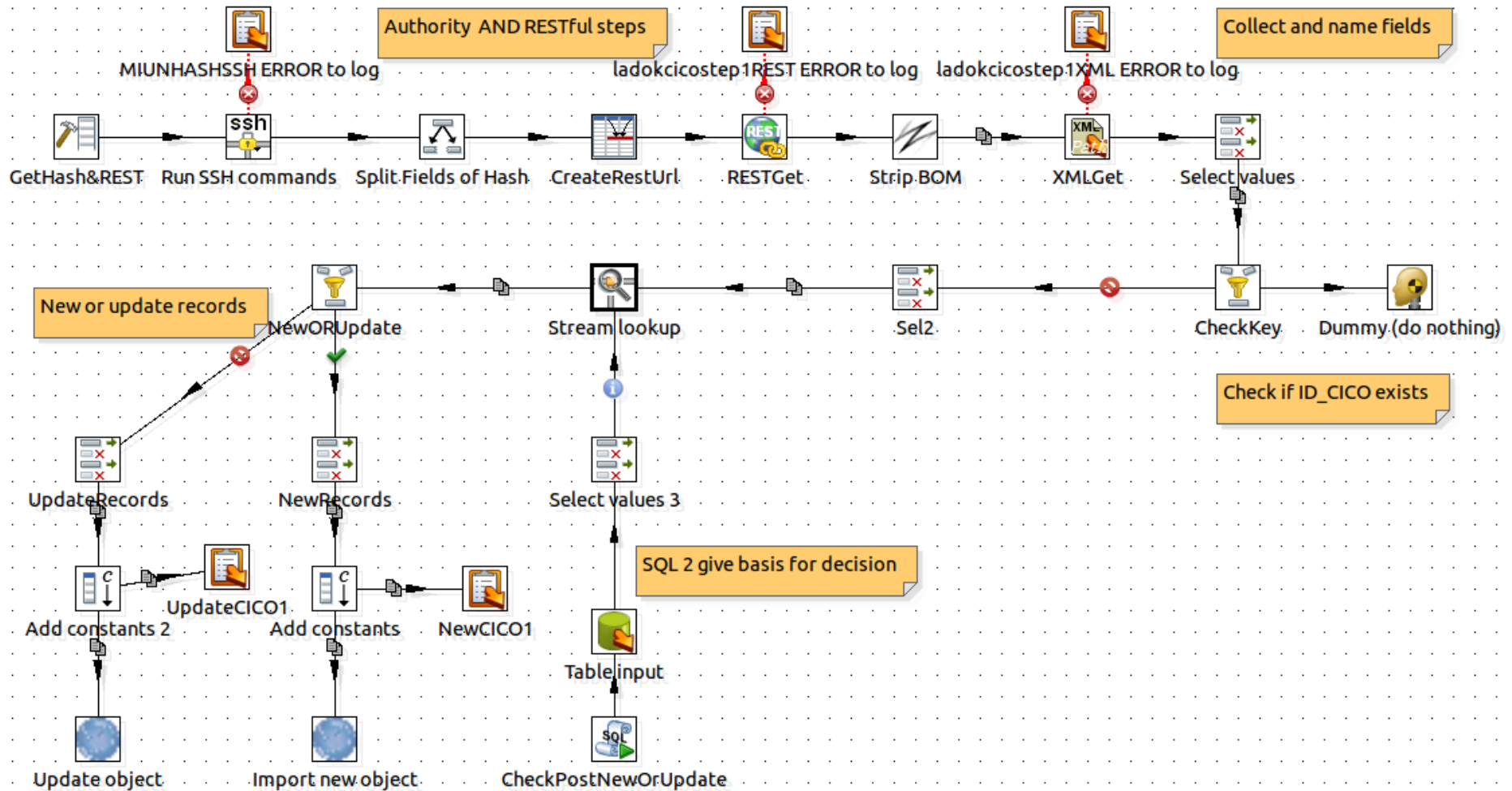
# Lookup



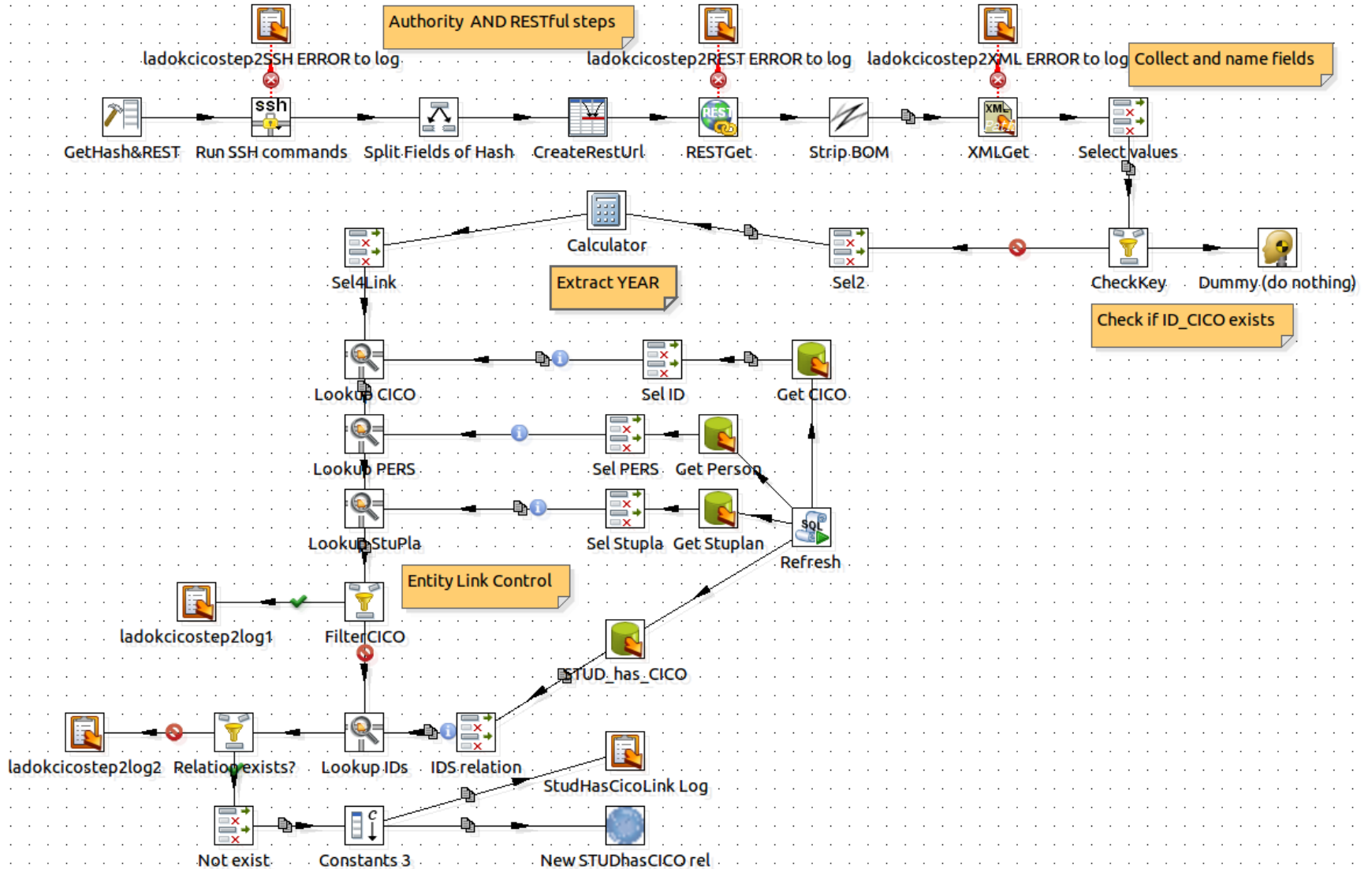
## Exempel på transformation i projektet Forskningsdatabas

- Behörighetskontroll
- Ansluta till källa
- Hämta data
- Tolka data
- Transformera och selektera
- Överlämna data
- Logga

# Ex.1

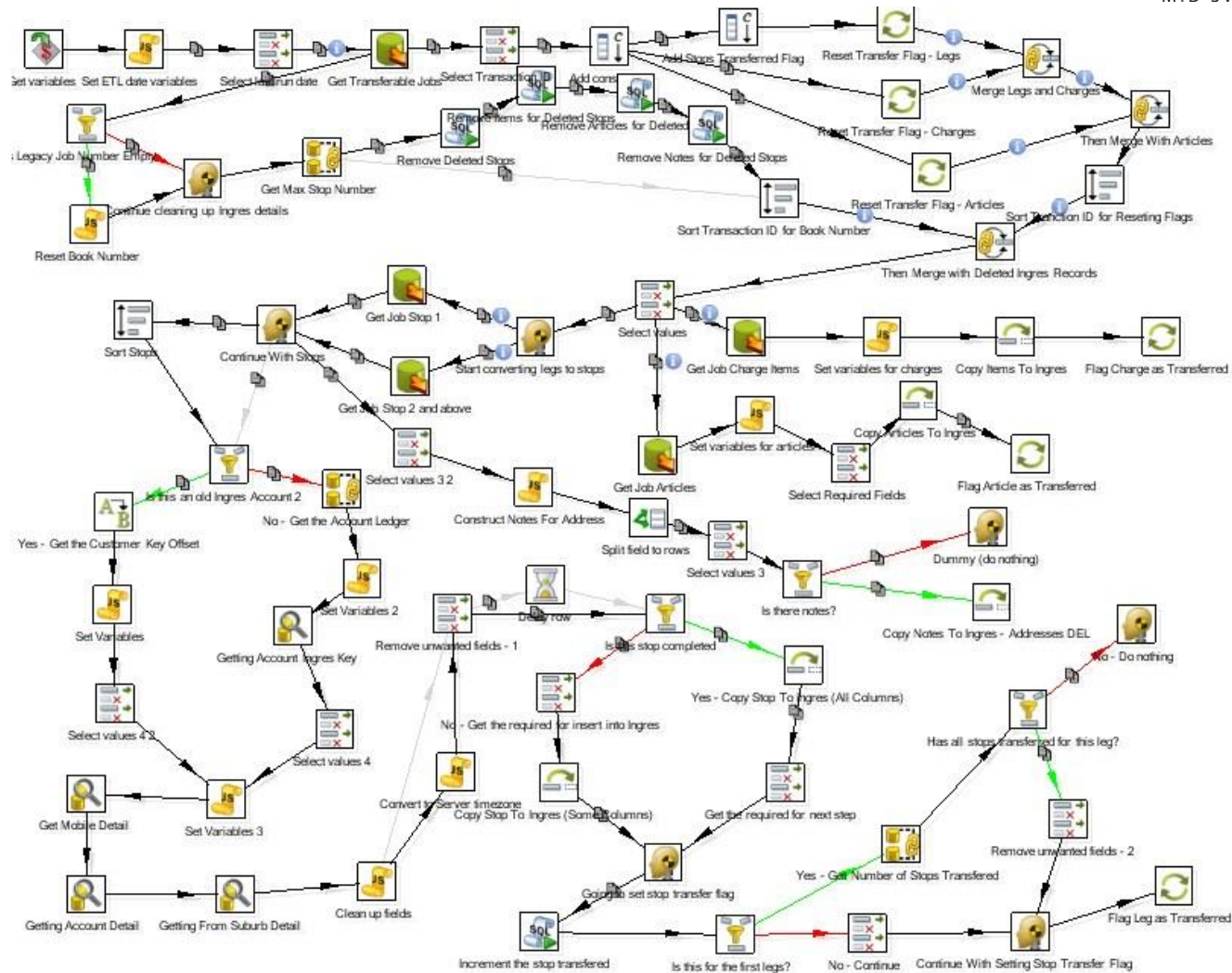


# Ex.2

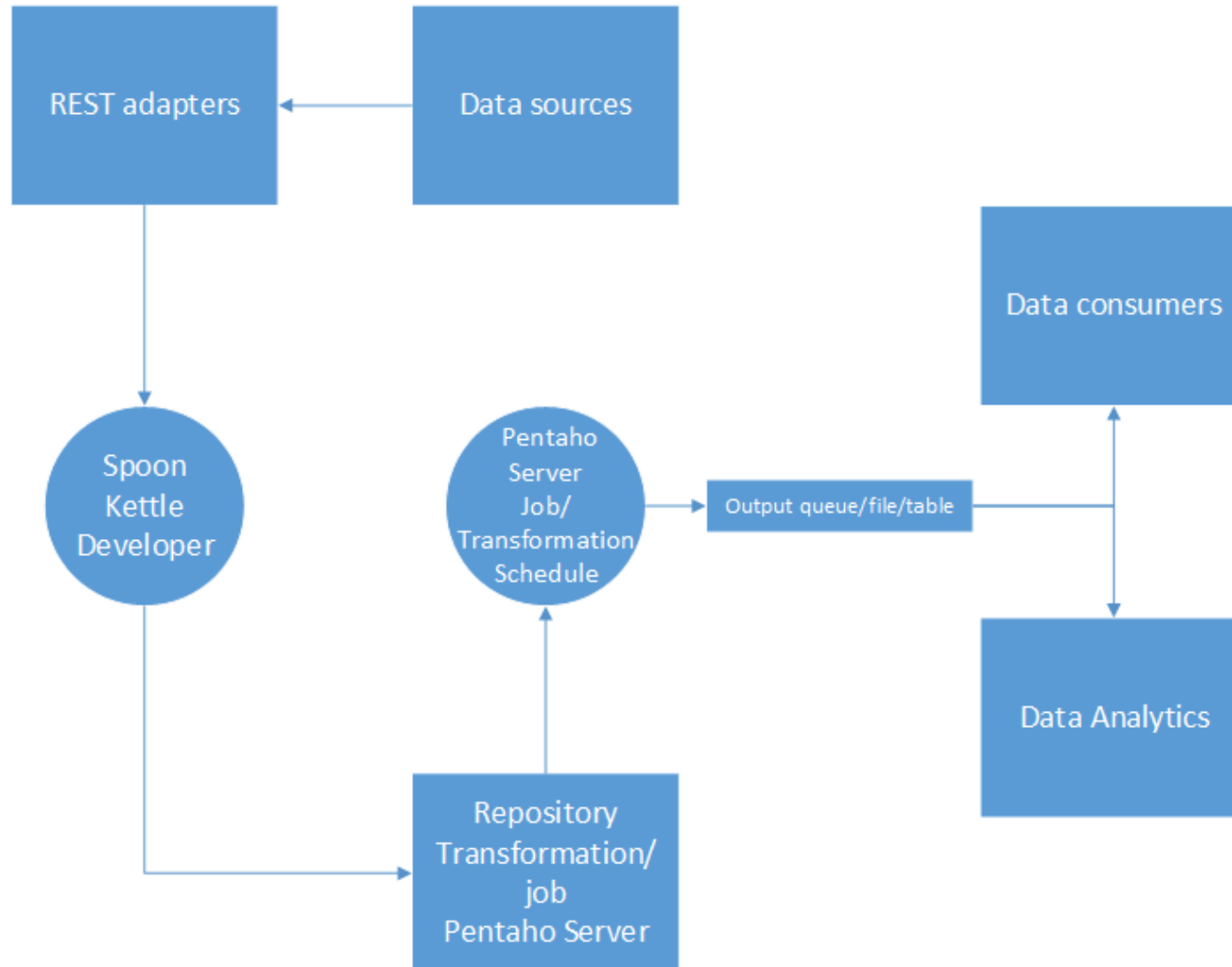




# Det kan även se ut så här...



# Hur det i stort skulle kopplas ihop..





# Swagger REST dokumentation



<http://lamp1-utv.utveckling.server.miun.se/swaggerDotNet/ui/sw>

Explore

## Integration platform RESTFUL services .NET

People

Show/Hide | List Operations | Expand Operations

Finance

Show/Hide | List Operations | Expand Operations

Education

Show/Hide | List Operations | Expand Operations

[ BASE URL: , API VERSION: 1.0.0 ]



## Integration platform RESTFUL services .NET

### People

Show/Hide | List Operations | Expand Operations

### Finance

Show/Hide | List Operations | Expand Operations

### Education

Show/Hide | List Operations | Expand Operations

GET /ResearchDataBaseWcfService/ResearchDataBaseServiceRest.svc/Getviewheadersxml

GET /ResearchDataBaseWcfService/ResearchDataBaseServiceRest.svc/GetResearchDataAllDataXml

GET /ResearchDataBaseWcfService/ResearchDataBaseServiceRest.svc/GetResearchDataByFreeSearchXml

GET /ResearchDataBaseWcfService/ResearchDataBaseServiceRest.svc/GetResearchDataByWhereAndOrderbyXml

[ BASE URL: , API VERSION: 1.0.0 ]



## Integration platform RESTful services PHP

A collection of RESTful services in the integrationplatform that handles base requests.

Created by Ove Olander, IT-avd. Mittuniversitetet

[Contact the developer](#)

### Base : RESTful services for administration

Show/Hide | List Operations | Expand Operations

### PHDStudent : RESTful services for PHD information

Show/Hide | List Operations | Expand Operations

GET	/web2/getPHDSISActivity	PHD student activity in Ladok. pnr,sem,research area, activity%, financial%
GET	/web2/getPHDSISCourseData	PHD courses available in Ladok
GET	/web2/getPHDSISCredit	PHD students results from Ladok
GET	/web2/getPHDSISExamDate	Latest PHD student activity in semester from Ladok
GET	/web2/getPHDSISFinance	Finance of PHD studies from Ladok
GET	/web2/getPHDSISLatestActivity	Latest PHD student activity in procent from Ladok
GET	/web2/getPHDSISLatestFinSem	Latest PHD's semester registration of finance's in Ladok
GET	/web2/getPHDSISLatestSem	Latest PHD student activity in semester from Ladok
GET	/web2/getPHDSISResearchExam	PHD courses available in Ladok
GET	/web2/getPHDSISResearchTopic	PHD Topics national
GET	/web2/getPHDSISStudyplanCourse	PHD's studyplans registered in Ladok
GET	/web2/getPHDSISStudyplanLatest	PHD's studyplans latest registered in Ladok
GET	/web2/getPHDSISStudyplanMain	PHD's studyplans registered in Ladok

[ BASE URL: , API VERSION: 1.0 ]

GET /web2/getPHDSISStudyplanCourse PHD's studyplans registered in Ladok

### Implementation Notes

Use hash + timestamp on REST services that use authorization control

### Parameters

Parameter	Value	Description	Parameter Type	Data Type
outputFormat	<input type="text"/>	select outputformat, raw, xml, json, csv, fdb	query	string
pnr	<input type="text"/>	pnr=YYYYmddXXXX or pnr=YmddXXXX	query	string
coursecode	<input type="text"/>	Course code coursecode=XXXXXXXX	query	string
endOfcourse	<input type="text"/>	course end date ...endOfcourse=YYYY-mm-dd	query	string
nameOfcourseEN	<input type="text"/>	name of course EN nameOfcourse=xxxxxx yyyyy zzzzz	query	string
nameOfcourseSE	<input type="text"/>	name of course EN nameOfcourse=xxxxxx yyyyy zzzzz	query	string
examinator	<input type="text"/>	name of examiner=teachers name	query	string
hashString	<input type="text"/>	Hash for checking authorisation	query	string
hashTime	<input type="text"/>	Timestamp in format YYYY-mm-ddTHH:MM:SS	query	string
encoding	<input type="text"/>	encoding=1 -> UTF-8, encoding=2 ->ISO-8859-1	query	string

### Response Messages

HTTP Status Code	Reason	Response Model	Headers
200	returns a hash and a timestamp		



Mittuniversitetet  
MID SWEDEN UNIVERSITY



ActiveMQ



- GET
- POST
- PUT
- DELETE



pentaho™  
open source business intelligence™

Pentaho Data Integration  
Previously Kettle



Microsoft®  
BizTalk® Server



## Lite länkar

Pentaho Community

<https://community.hds.com/>

Dokumentation

<http://wiki.pentaho.com/display/EAI/Latest+Pentaho+Data+Integration+%28aka+Kettle%29+Documentation>

Sourceforge

<https://sourceforge.net/projects/pentaho/#>

Pentaho Server Business Intelligence BI

<https://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/7.0/>





**SLUT**

