Statistical databases in theory and practice Part IV: Modelling the contents and structure of official statistics

> Bo Sundgren 2010

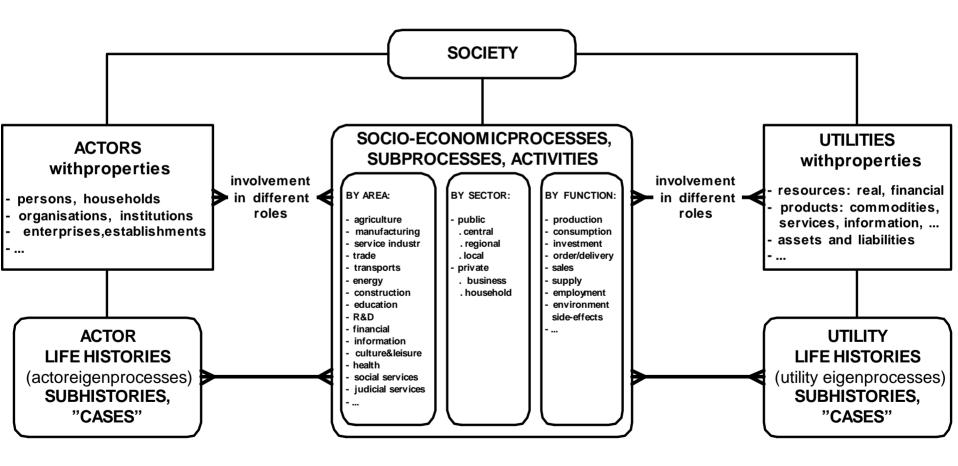
Modelling the contents and structure of official statistics

Or: How to design "correct" and globally consistent SDMX Data Structure Definitions

Or: Navigating in a space of statistical surveys of society Or: Reality as a statistical construction

> Based on a paper by Bo Sundgren, Statistics Sweden ICES-III, Montreal, June 18-21, 2007

A generic conceptual model of society as reflected by official statistics (1)



A generic conceptual model of society as reflected by official statistics (2)

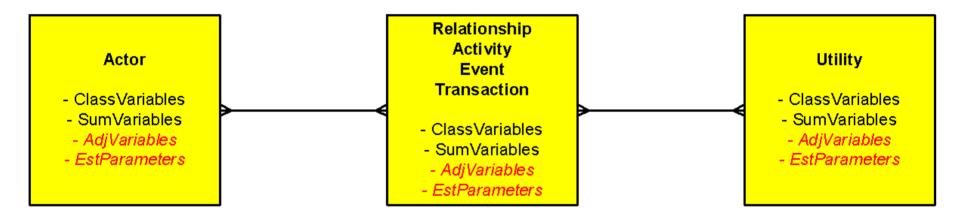


Figure 3. Generic model, covering all statistical domains.

Statistics Canada: Agents, Events, Things

AGENTS		EVENTS		THINGS	
Social	Economic	Social	Economic	Social	Economic
Census Family (Census)	Business Entity	Birth	Person-trip		Building Permit
Census Family (SLID) Child (Census)	Business Location Earner	Community Admission Correctional	Person-visit	Charge Criminal Offence	Employment Insurance Claim Farm Input
Child (THS) Crime Victim Criminal Accused Criminal Charged	Economy Employed Person Employee (LFS)	Service Admission Criminal Incident Custodial		Legal Aid Application Legal Aid Plan Probation Order Sentence	Help Wanted Ad Job Passenger-Kilometer Product
Emigrant Family Unit Homicide victim	Employee (private) Employee (public) Employee (SEPH) Employment Insurance Beneficiary	Admission Death Divorce		Shelter (THS)	Security Transaction Vehicle
Household (Census) Household Head Immigrant	Enterprise Establishment Farm Operation (agr	Homicide Marriage			Vehicle -Kilometer Visit-night
International Migrant Interprovincial Migrant	web) Institutional Unit Labour Force Participant				
Intraprovincial Migrant Mother	Paid Worker (labour income) Paid Worker (labour market)				
Non-permanent Resident Person	Person 15 Years or over (LFS) Self-Employed Worker				
Person 12 Years or over Person charged	Unemployed Person				
Smoker Student					

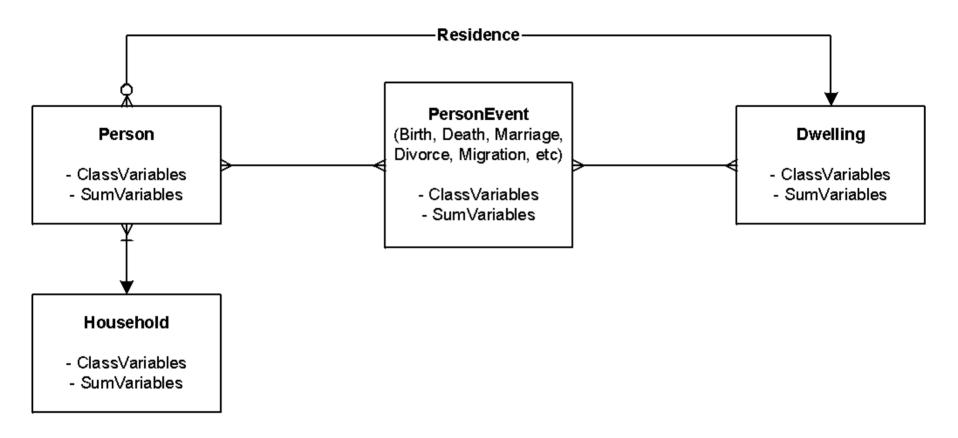


Figure 2a. Population statistics: Basic model.

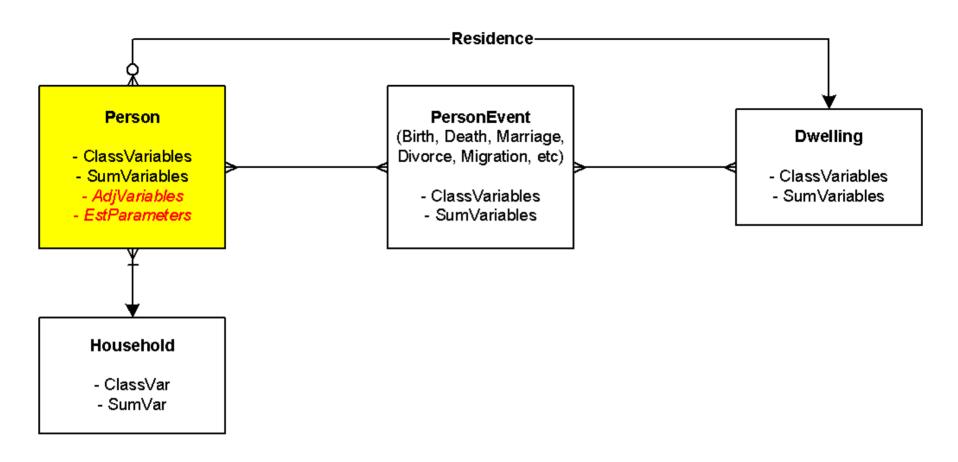


Figure 2b. Person statistics.

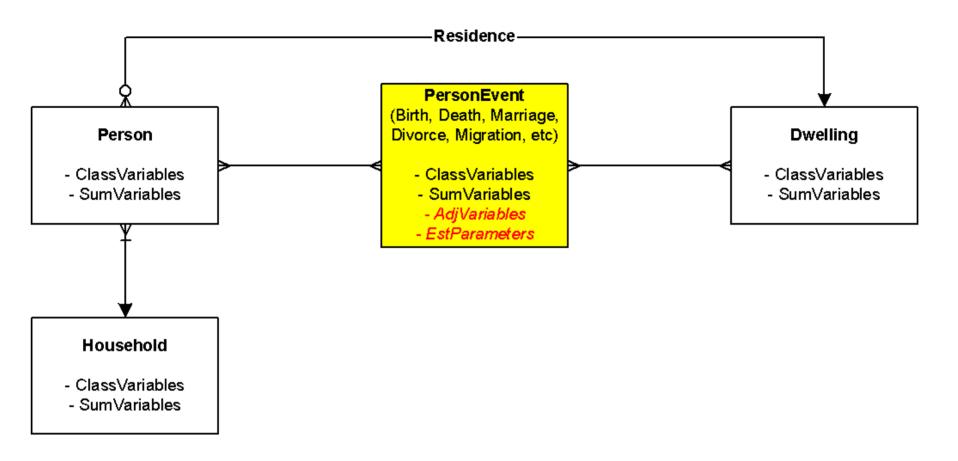


Figure 2c. PersonEvent statistics.

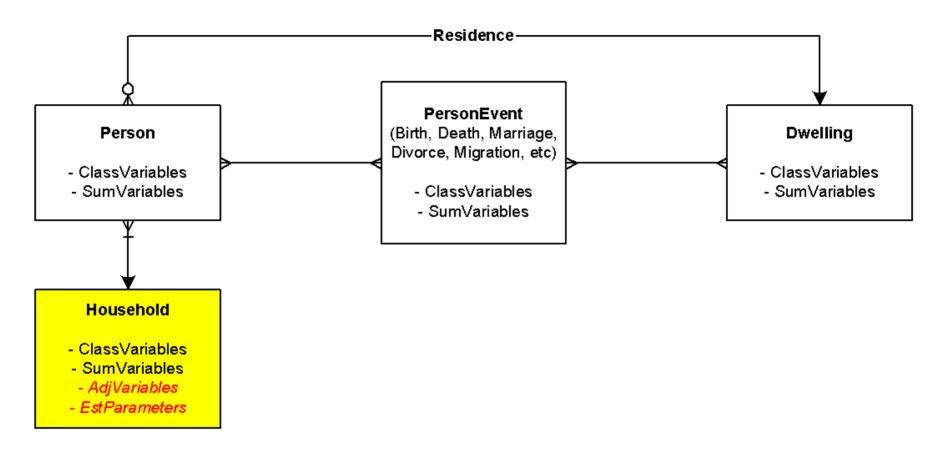


Figure 2d. Household statistics.

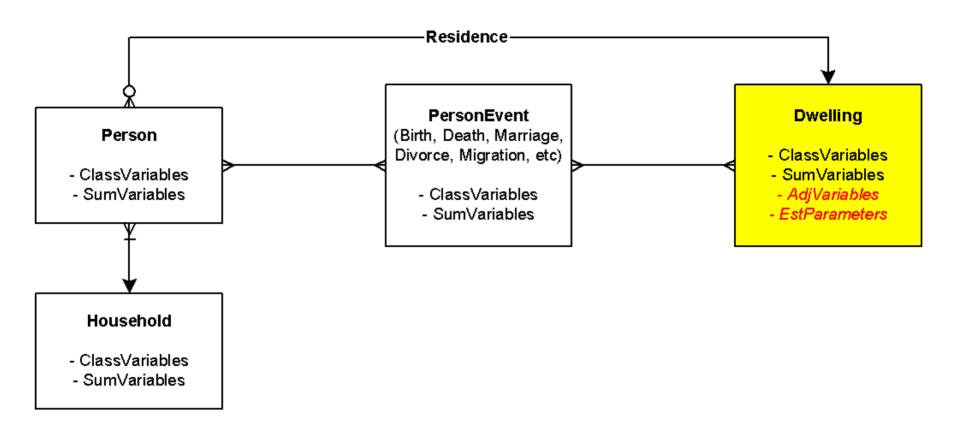


Figure 2e. Dwelling statistics.

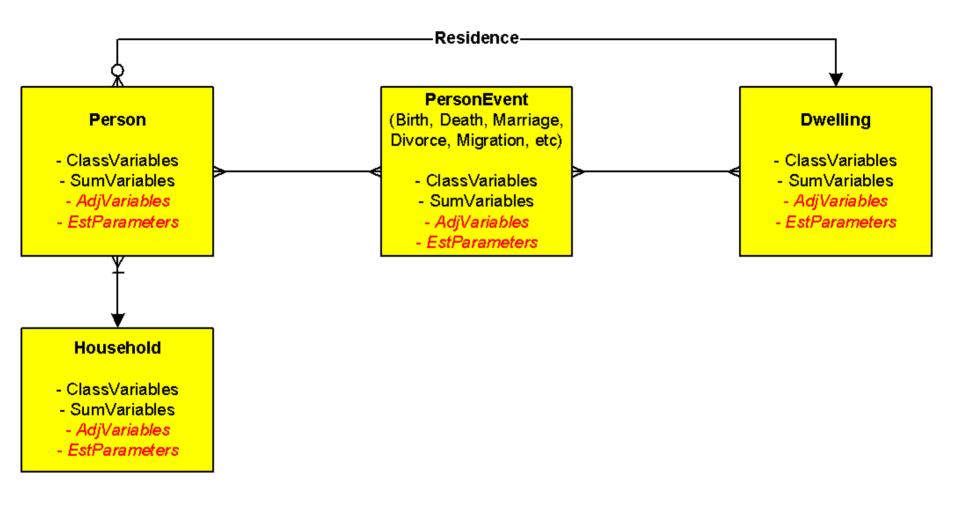


Figure 2f. Population statistics: Split vision model.

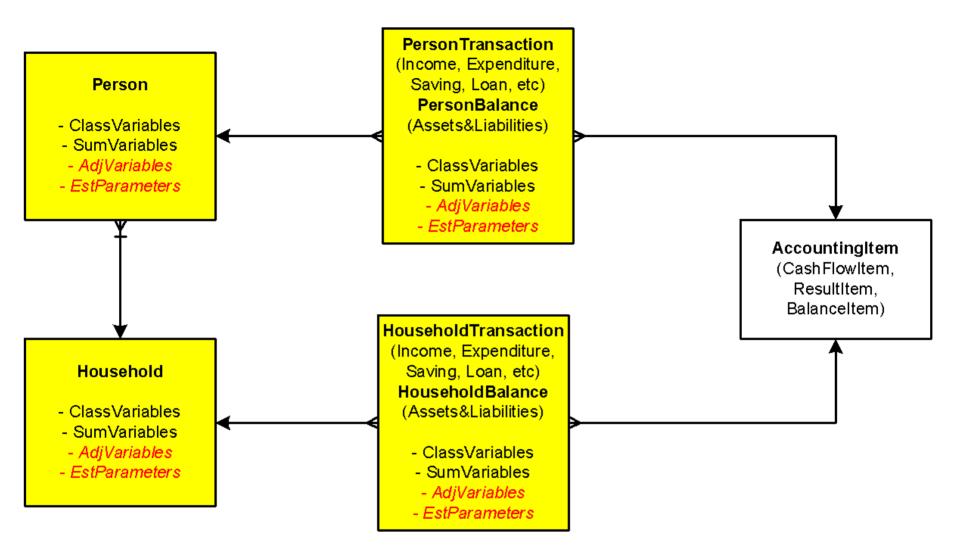


Figure 4. Household economy statistics.

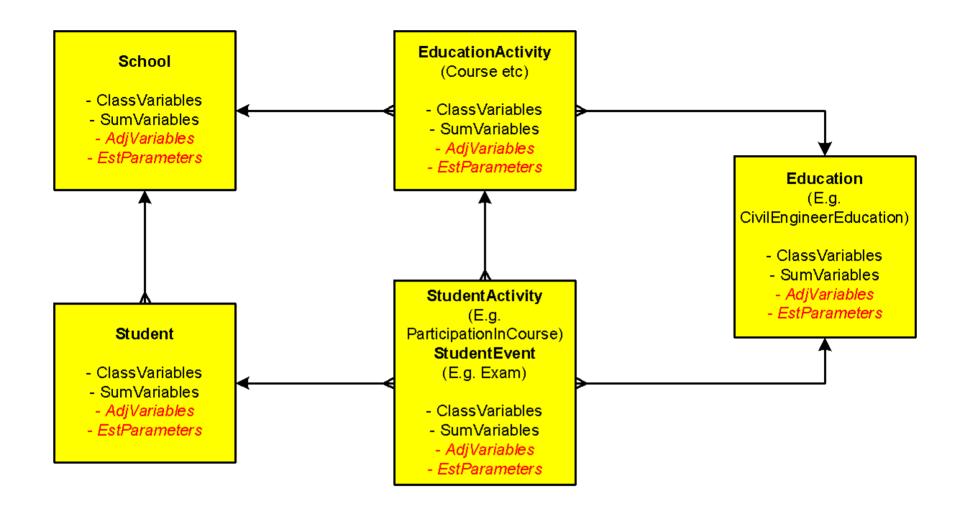
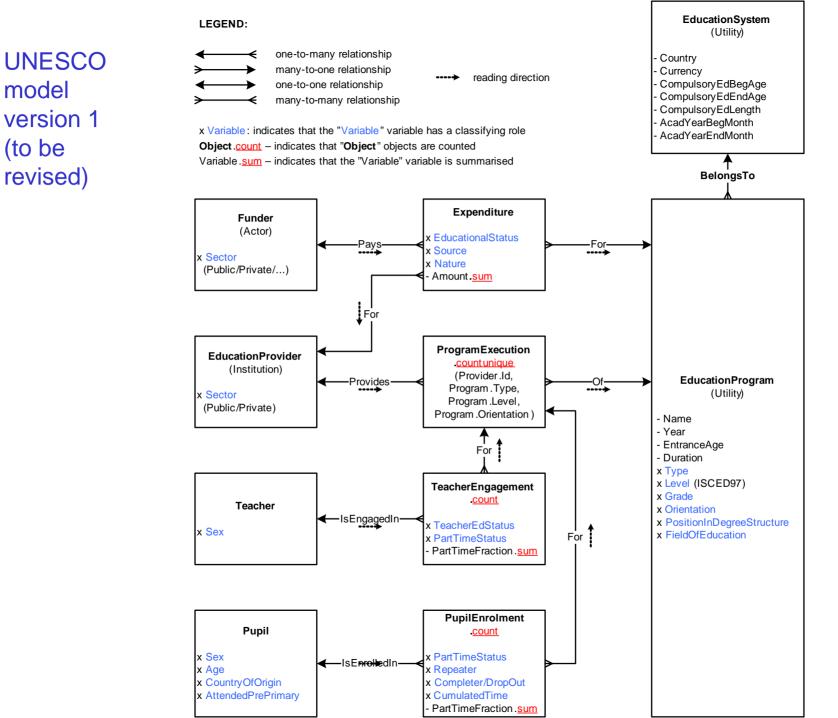


Figure 5. Education statistics.



(to be

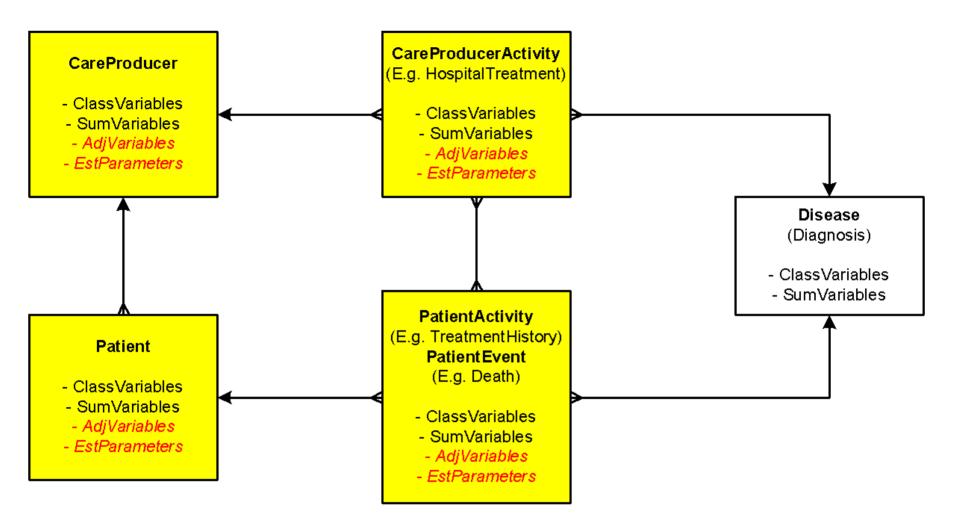


Figure 6. Health statistics.

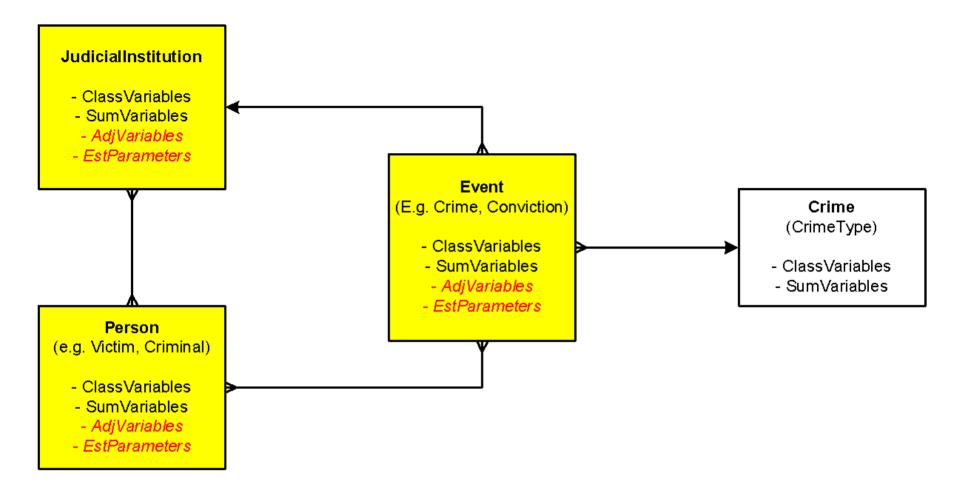


Figure 7. Judicial statistics.

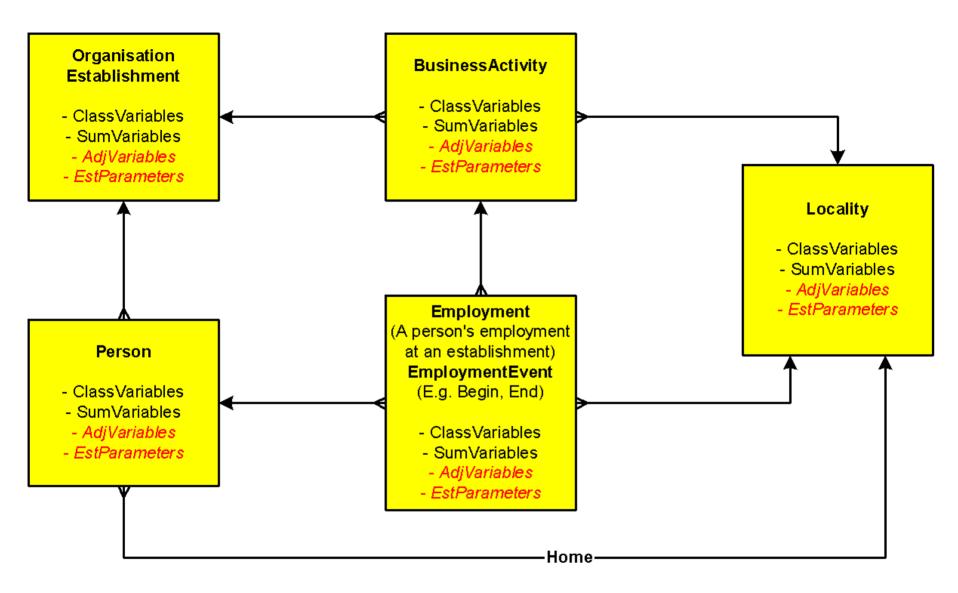


Figure 8. Labour market statistics.

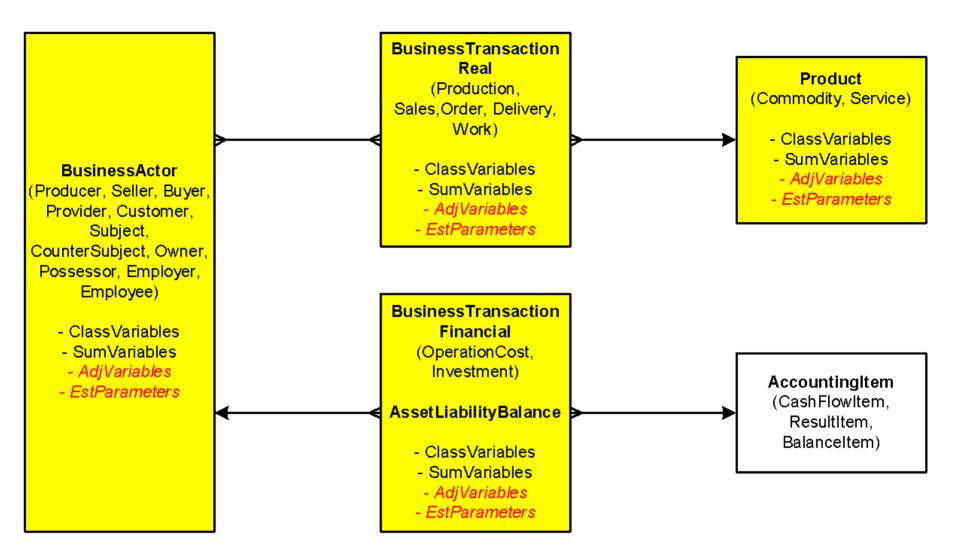


Figure 9. Business statistics.

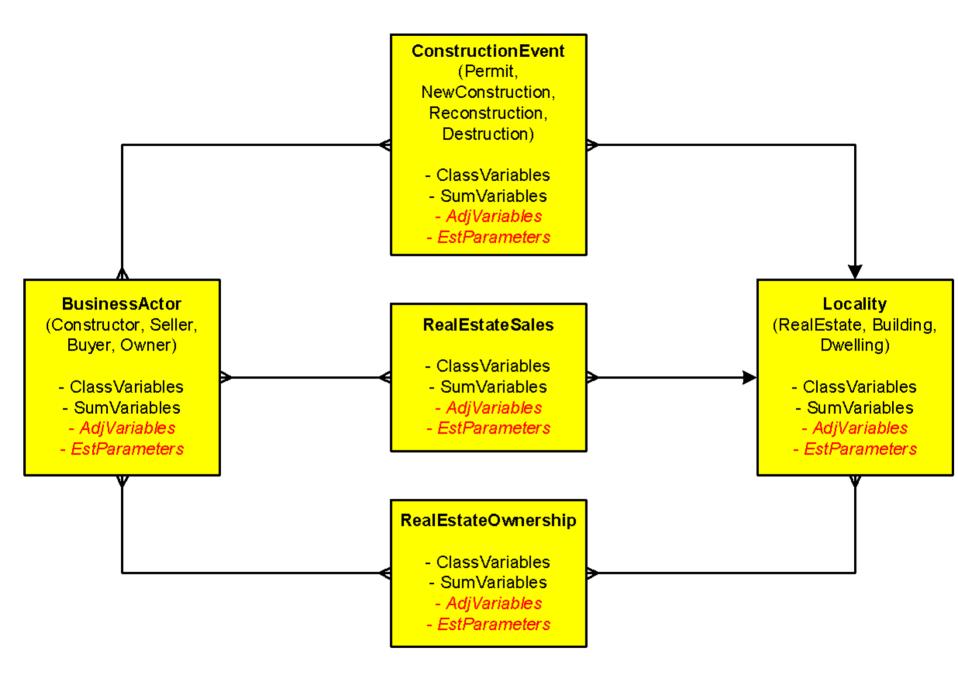


Figure 10. Housing and construction statistics.

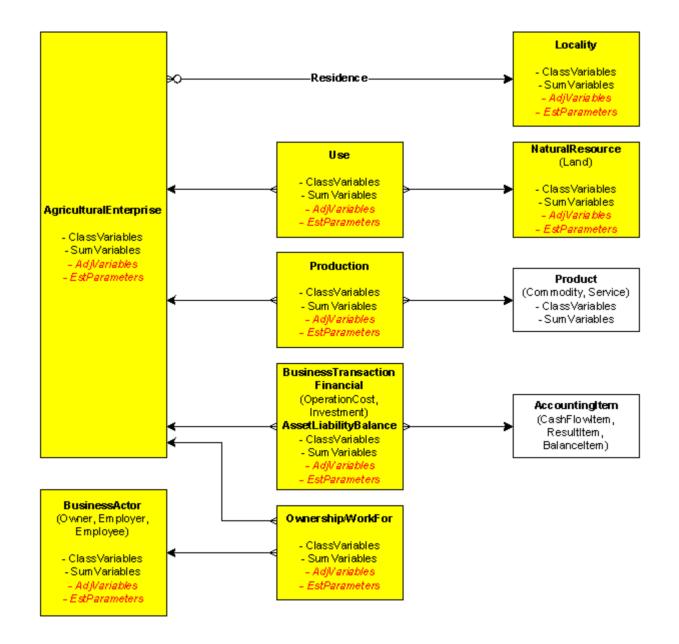


Figure 12. Agriculture statistics.

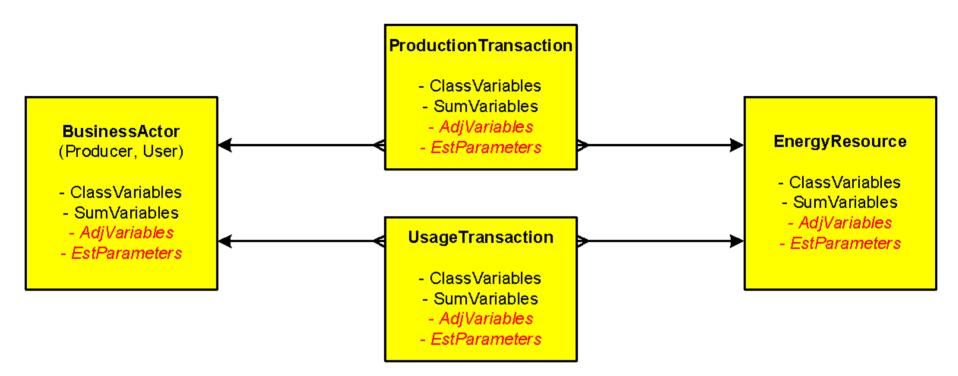


Figure 11. Energy statistics.

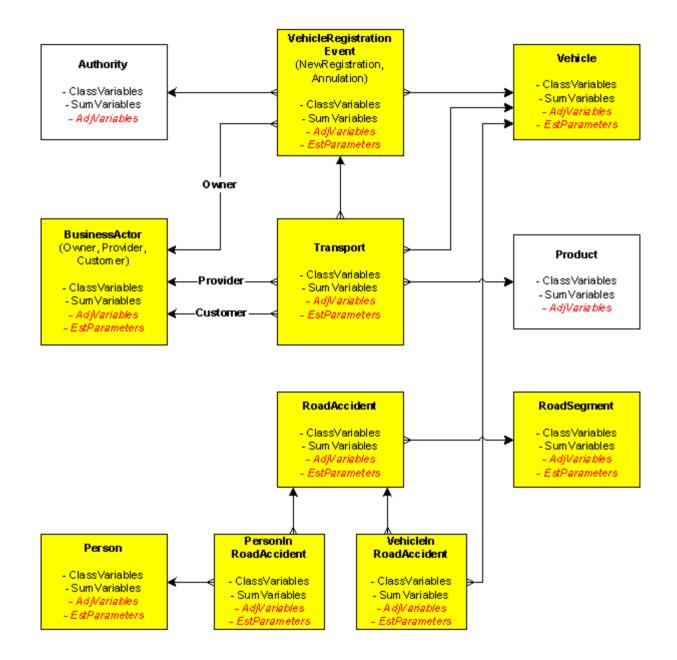


Figure 13. Transport statistics.

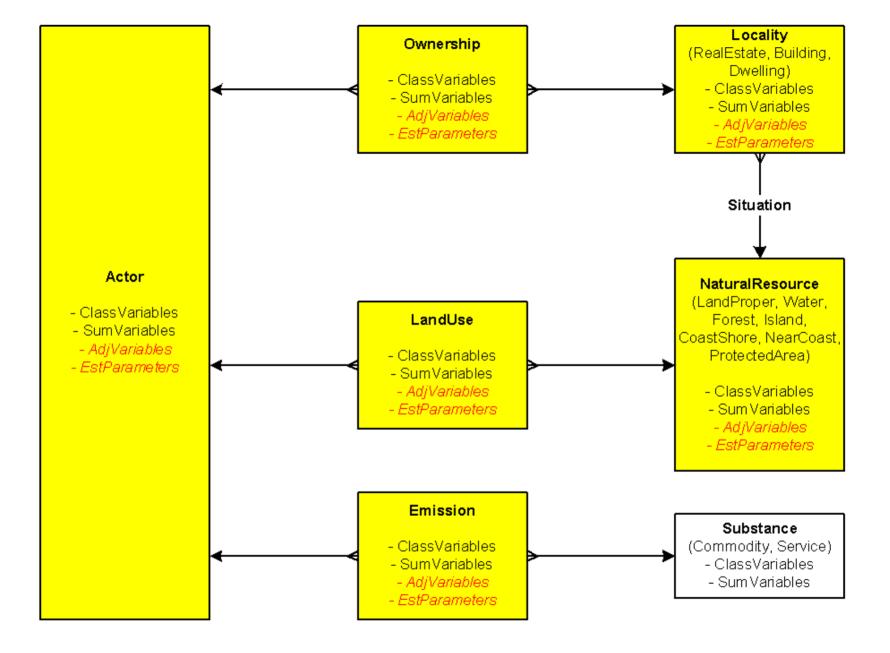


Figure 14. Environment statistics.

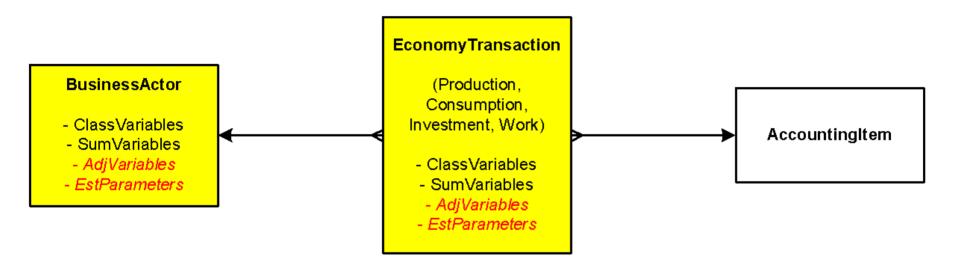


Figure 16. National accounts.

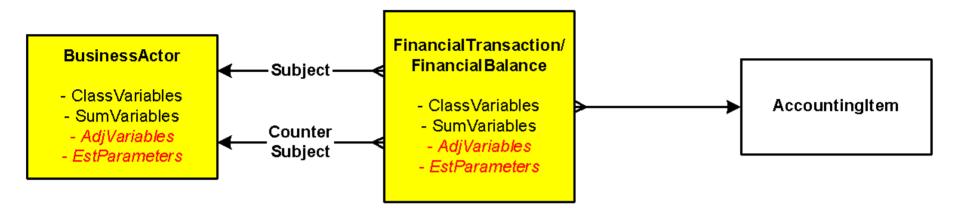


Figure 17. Financial statistics.

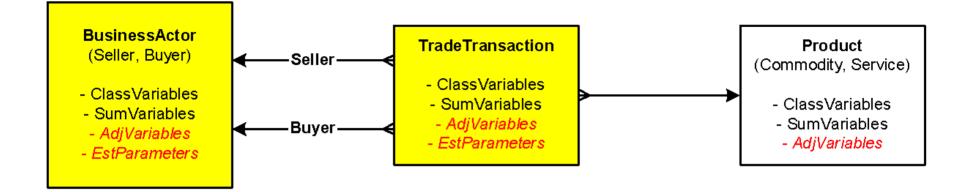
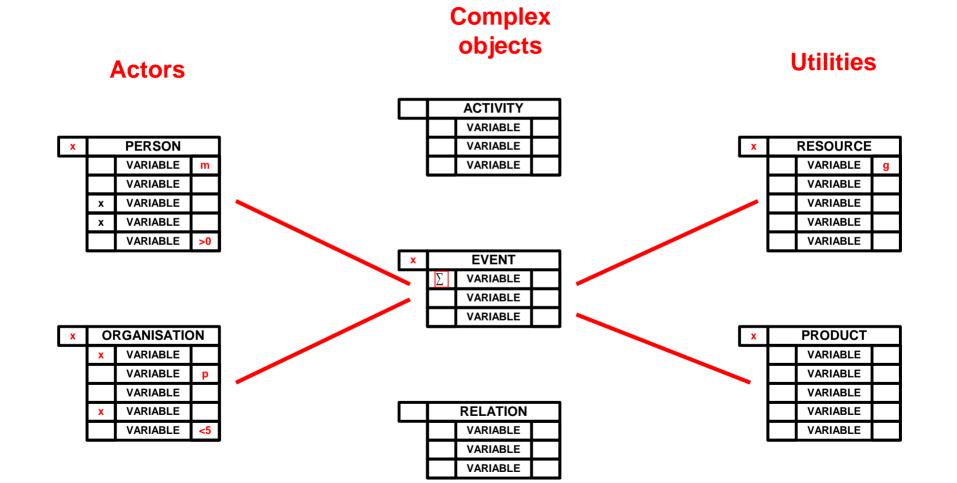


Figure 15. Trade statistics.

What can a statistical agency do, in order to help a user

- find potentially relevant statistical data?
- judge the relevance of data retrieved?
- Provide overviews of available data
- Provide search tools
- Provide informative metadata

Contents By Example (based on a simple generic model)



L I	-	
All Actors	🗆 All Topics	All Utilities
□ Person □ Student	Topic: Population	□ NaturalResource □ Land
🗆 Employee	Topic: Education	
□ Patient □ Client	□ Topic: Health	□ CardinalResource □ Locality □ RealEstate
🗆 Household	n Topic: Law	D Building
Organisation	Topic: Labour Market	🗆 Vehicle
 Enterprise AgriculturalEnterprise Institution 	Topic: Business Activities	□ Product □ Commodity
Establishment	□ Topic: Housing and Construction	□ Service
BusinessActor Producer	De Topic: Agriculture	□ AccountingItem □ CashFlowItem
□ Seller	n Topic: Energy	□ Resultitem □ Balanceitem
 Buyer Provider Customer 	Topic: Transports	
□ Customer □ Subject □ Countou Subject	Topic: Environment	□ WelfareItem □ Education
□ CounterSubject □ Owner □ Decensor	Topic: National Accounts	□ Health □ Wealth
□ Possessor □ Employer	🗅 Topic: Financial markets	□ Security
🗆 Employee	🗆 Topic: Trade	

	× Topic: Population	
o Person	a Birth(Person)	🛚 Locality
o Sex	a AgeOfMother	🛛 Location
o Age		o Real Estate
B HomeLocation	o Death(Person)	a Building
WorkLocation		a Dwelling
o MaritalStatus	¤ GetMarried[Person]	
🛛 Income(bykind)	¤ GetMarried[Person, Person]	
o Wealth(byKind)		
EducationLevel	a GetDiverced[Person]	
D Occupation	a GetDivbroed[Person, Person]	
B Household	a SeekAsylum[Person]	
o Size	a GetAsylum[Person]	
D HomeLocation		
a Income(bykind)	B SeekCitizenship[Person]	
o Wealth(byKind)	a GetCitizenship[Person]	
	a Membership (Person, Household)	
	a MarriedTo[Person, Person]	
	a ChildOf[Person, Person]	
	- Denidense (Denne den lik A	
	a Residence[Person, Locality]	
	© Commute[Person, Locality.From, Locality.To]	
	o Migration[Person, Locality.From, Locality.To]	
	Immigration[Person, Country.From, Locality.To]	
	n Reason	
	a BirthCountry	
	Emigration[Person, Locality.From, Country.To] a Bith Country	
	a BirthCountry	

Examples of data sets that may be specified/retrieved/exchanged on the basis of this conceptual (sub)model:

- 1. Number of births by sex and home location of the child and age of the mother. (4-dimensional frequency macrodata cf underlying microdata.)
- 2. Number of migrations by sex, age, and occupation of migrant and place moved from and to. (5-dimensional frequency macrodata.)
- 3. Average income of persons by sex and education level. (2-dimensinal summarised macrodata.)

	-	
	Topic: Population	
		🛛 Locality
x Person T	# Birth[x Person]	□ Location
x Sex	x AgeOfMother. A geGrouping	RealEstate
o Age		Building
X HomeLocation. <i>RegionalCode</i>	Death[Person]	Dwelling
D WorkLocation	D Deaulfeisonj	Divening
ם MaritalStatus	GetMarried[Person]	
Income(byKind)	GetMarried[Person, Person]	
□ Wealth(byKind)		
EducationLevel	GetDivorced[Person]	
Occupation	GetDivorced[Person, Person]	
🛛 Household	SeekAsylum[Person]	
o Size	🛛 GetAsylum[Person]	
- HomeLocation	a couldy and a could	
□ Income(byKind)	BeekCitizenship[Person]	
□ Wealth(byKind)	GetCitizenship[Person]	
	Manakanakin (Dana an Ulawa da 1-0	
	Membership[Person, Household]	
	MarriedTo[Person, Person]	
	ChildOf[Person, Person]	
	Basidence[Person, Locality]	
	Commute[Person, Locality.From, Locality.To]	
	Image Migration[Person, Locality.From, Locality.To]	
	Immigration[Person, Country.From, Locality.To]	
	Reason	
	BirthCountry	
	Emigration[Person, Locality.From, Country.To]	
	BirthCountry	

Example of data set:

1. Number of births by sex and home location of the child and age of the mother. (4-dimensional frequency macrodata – cf underlying microdata.)

1		
т	Topic: Population	
T		xx Locality
x Person	Birth[Person]	Location.RegionalCode
x Sex	□ AgeOfMother	RealEstate
x Age. Age Grouping		Building
HomeLocation	Death[Person]	Dwelling
WorkLocation		-
ם MaritalStatus	GetMarried[Person]	
Income(byKind)	GetMarried[Person, Person]	
□ Wealth(byKind)		
Education Level	GetDivorced[Person]	
x Occupation. Occupation Code	GetDivorced[Person, Person]	
🛛 Household	SeekAsylum[Person]	
🛛 Size	GetAsylum[Person]	
HomeLocation		
Income(byKind)	BeekCitizenship[Person]	
□ Wealth(byKind)	GetCitizenship[Person]	
	Membership[Person, Household]	
	MarriedTo[Person, Person]	
	ChildOf[Person, Person]	
	Residence[Person, Locality]	
	Commute[Person, Locality.From, Locality.To]	
	#Migration[x Person, x Locality.From, x Locality.To]	
	Immigration[Person, Country.From, Locality.To]	
	🛛 Reason	
	BirthCountry	
	Emigration[Person, Locality.From, Country.To]	
	BirthCountry	

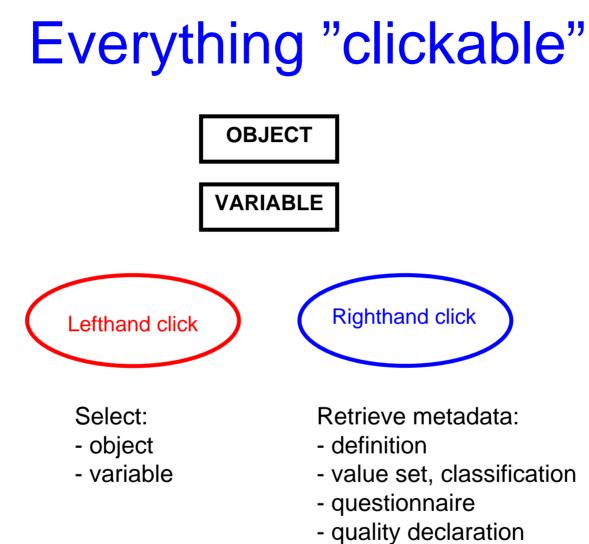
Example of data set:

2. Number of migrations by sex, age, and occupation of migrant and place moved from and to. (5-dimensional frequency macrodata.)

1	-	
	Topic: Population	
		🛛 Locality
x Person	Birth[Person]	□ Location
x Sex	□ AgeOfMother	RealEstate
🛛 Age		Building
HomeLocation	Death[Person]	Dwelling
WorkLocation		
ם MaritalStatus	🛛 GetțMarried[Person]	
m Income(byKind)	GetMarried[Person, Person]	
□ Wealth(byKind)		
x EducationLevel	GetDivorced[Person]	
Occupation	GetDivorced[Person, Person]	
🛛 Household	BeekAsylum[Person]	
🗆 Size	🛛 GetAsylum[Person]	
HomeLocation		
Income(byKind)	SeekCitizenship[Person]	
□ Wealth(byKind)	GetCitizenship[Person]	
	Membership[Person, Household]	
	□ MarriedTo[Person, Person]	
	D ChildOf[Person, Person]	
	Desidence (Deserve Landited	
	□ Residence[Person, Locality]	
	□ Commute[Person, Locality.From, Locality.To]	
	Image: Migration [Person, Locality.From, Locality.To]	
	Immigration[Person, Country.From, Locality.To]	
	□ Reason □ BirthCountry	
	Emigration[Person, Locality.From, Country.To] BirthCountry	

Example of data set:

3. Average income of persons by sex and education level. (2-dimensinal summarised macrodata.)



- survey documentation

Part IV: Extra material

Bo Sundgren 2010

	District Topic: Education		
 Person Sex Age HomeLocation MaritalStatus Student EducationLevel ParentsEducation NationalBackground FinancialSupport 	 EducationActivity[Institution, Education, #] EducationYearSemester Participation[Student, EducationActivity] Exam[Student, EducationActivity] 	WelfareItem Education Type Subject Level Duration	
Organisation Institution School Location Type Owner			

☐ □ Person □ Sex □ Age □ HomeLocation □ WorkLocation □ MaritalStatus □ Patient □ Organisation	 Topic: Health DiseaseDiagnosis[Patient, Disease] VisitToPhysician[CareProducer, Patient] Reason HospitalTreatment[CareProducer, Patient] BedDays DeathDiagnosis[Patient, Disease] 	□ Welfareltem □ Health □ Disease
 Organisation CareProducer Location PublicPrivate GeneralSpecialized 	DeathDiagnosis[Patient, Disease]	

	🗆 Topic: Law	
 Person Sex Age HomeLocation MaritalStatus Income(byKind) Wealth(byKind) EducationLevel SocioEconomicGroup CriminalStatus NationalBackground Organisation Institution PoliceAuthority Court 	 CrimeReport[PoliceAuthority, Crime.Type] Conviction[Court, Crime.Type] 	□ WelfareItem □ Security □ Crime □ Type

_ I		
	🖬 Topic: Labour Market	
🗆 Person	EmploymentEvent[Person, Establishment]	□ Locality
□ Sex	□ PartTimePercentage	□ Location
□ Ager	□ EmploymentForm	□ RealEstate
HomeLocation	□ PersonnelCategory	Building
WorkLocation	🗆 Salary	🗆 Dwelling
🖬 MaritalStatus		-
□ Income(byKind)	EmploymentRelation[Person, Establishment]	
Wealth(byKind)	□ PartTimePercentage	
EducationLevel	□ EmploymentForm	
□ Profession	PersonnelCategory	
🗆 LabourForceStatus	□ Salary	
□ NotinLabourForce	□ SalaryChange	
□ Reason □ ReasoibilityW/ich	□ OnSickLeave	
PossibilityWish InLabourForce	Deridenee/Dereen Leeslikul	
□ In∟abour⊢orce □ Employed	Residence[Person, Locality]	
□ Employed □ MainOccupation	Residence[Establishment, Locality]	
□ SideOccupation	l Residence[Establishment, Eventy]	
District District Distribution	Commute[Person, Locality.Home, Locality.Work]	
_ TimeOfAbsence	Commuter craon, cocanty none, cocanty work	
ReasonOfAbsence		
Working		
WeeklyWorkingTime		
🗆 Unemployed		
UnemploymentStatus		
Unemployment Time		
Organisation		
□ Establishment		
Location		
	<u> </u>	I

	D Topic: Business Activities	
 Organisation Enterprise Establishment Location KIndOfActivity BusinessActor Location Sector KindOfActivity Producer Seller Buyer Provider Customer Subject CounterSubject Owner Possessor Employee 	 Residence[Establishment, Locality] Business transactions and balances (real) ProductionTransaction[Producer, Product] Volume Value SalesTransaction[Seller, Buyer, Product] Volume Value Order[Provider, Customer, Product] Value Delivery[Provider, Customer, Product] Value Delivery[Provider, Customer, Product] Value Delivery[Provider, Customer, Product] Valu	 Locality Location Product Commodity Service AccountingItem CashFlowItem IncomeItem ExpenditureItem ConsumptionItem InvestmentItem ResultItem CostItem RevenueItem BalanceItem Asset Liability

	Topic: Housing and Construction	
 Person Sex Age HomeLocation WorkLocation MaritalStatus Income(byKind) Wealth(byKind) Household Size HomeLocation Income(byKind) Wealth(byKind) Organisation Enterprise Establishment Location KindOfActivity BusinessActor Location Sector KindOfActivity Constructor Seller Buyer Owner 	 RealEstateOwnership[Owner, RealEstate] OwnershipShare RealEstateSalesTransaction[Seller, Buyer, RealEsstate] Price ConstructionEvent[Constructor, Building] NumberOfDwellings EventType Permit NewConstruction Reconstruction Destruction I 	 Locality Location RealEstate Building Dwelling Product Commodity Service AccountingItem CashFlowItem ResultItem BalanceItem

	🗆 Topic: Agriculture	
 Organisation Enterprise AgriculturalEnterprise Location 	 Residence[AgriculturalEnterprise, Locality] Ownership[BusinessActor, AgriculturalEnterprise] 	Locality Location
□ Location □ KindOfActivity	Work[BusinessActor, AgriculturalEnterprise] KindOfWorker	DaturalResource Land
□ BusinessActor □ Producer □ Seller	□ KindOfWork □ Hours	□ Product □ Commodity
□ Seller □ Buyer □ Provider □ Customer	 LandUse[AgriculturalEnterprise, Land, Crop] Area 	□ Crop □ Livestock □ Service
 Owner Possessor Employer 	 Production[AgriculturalEnterprise, Commodity] Volume Value 	□ AccountingItem □ CashFlowItem □ IncomeItem
□ Employee	 FinancialBalance[AgriculturalEnterprise, FinancialResource] Value 	 Expenditureltem ConsumptionItem InvestmentItem
	 CostRevenueTransaction[AgriculturalEnterprise, ResultItem] Value 	□ Resultiem □ Costitem □ Revenueltem
	 InvestmentTransaction[AgriculturalEnterprise, InvestmentItem] Value 	□ Balanceltem □ Asset □ Liability
ľ	1	

	🗆 Topic: Energy	
 BusinessActor Location Sector KindOfActivity Producer Seller Buyer Provider Customer Subject CounterSubject Owner Possessor Employee 	 Production[Producer, EnergyResource] Volume Value Usage[Customer, EnergyResource] Volume Value 	□ NaturalResource □ EnergyResource □ Electricity □ F uel

<u>н</u>		
	Topic: Transports	
□ Personï - Student	VehichleRegistration[Authority, Owner, Vehicle]	□ NaturalResource
□ Student □ Employee	 NewRegistration Annualation 	□ Land
Patient Client	- Transa att[Provider Customer Vehicle Product]	□ CardinalResource
🗆 Client	Transport[Provider, Customer, Vehicle, Product] Weight	Locality
🗆 Organisation		
Enterprise	□ LocationFrom	□ Туре
□ Institution □ Authority	□ LocationTo	□ Vehicle
Establishment	RoadAccident[RoadSegment]	
	□ Time	
BusinessActor	□ WeatherCondition	□ Product
□ Producer □ Seller	RoadCondition	□ Commodity □ Service
🗆 Buyer	PersonInRoadAccident[RoadAccident]	
Provider	□ PersonRoleInAccident	
□ Customer □ Subject	PersonCondition PersonConsequences	
CounterSubject		
🗆 Owner	vehicleInRoadAccident[RoadAccident]	
□ Possessor	VehicleRoleInAccident	
□ Employer □ Employee	 VehicleCondition VehicleConsequences 	
		I

1		
	Topic: Environment	
Person	□ LandUse[Actor, Land]	□ NaturalResource
□ Student	LandUseCategory	🗆 Land
🗆 Employee	n di	🗆 Type
□ Patient	🕆 🛛 Emission[Actor, Land, Substance]	Location
Client	n Amount	🗖 Area
		NumberOfBuildings
🗆 Household	Situation[Locality, Land]	LandProper
		D Water
Organisation	ownership[Owner, Locality]	□ Forest
🗆 Enterprise		□ Island
AgriculturalEnterprise		CoastShore
Institution		□ NearCoast
Establishment		□ ProtectedArea
		KindOfProtection
BusinessActor		
Producer		□ CardinalResource
□ Seller		□ Locality
□ Buyer		□ RealEstate
Provider		Building
Customer		□ Dwelling □ Vehicle
□ Subject		
CounterSubject		□ Product
□ Owner		Commodity
Possessor Finite State		□ Substance
□ Employer		- Substance
🗆 Employee		- AccountingItom
		 AccountingItem CashFlowItem
		Resultitem
		Balanceltem
		1

1 	De Topic: National Accounts	
 ■ BusinessActor ■ Location ■ Sector ■ KindOfActivity 	 ProductionTransaction[BusinessActor, Product] Value ConsumptionTransaction[BusinessActor, Product] Value Value InvestmentTransaction[BusinessActor, InvestmentItem] Value WorkTransaction[BusinessActor, AccountingItem] Hours 	 Product Commodity Service AccountingItem CashFlowItem Incomeltem ExpenditureItem ConsumptionItem InvestmentItem ResultItem CostItem RevenueItem BalanceItem Asset Liability

BusinessActor Bubject Location Sector CounterSubject Location Sector	 Topic: Financial markets FinancialTransaction[Subject, CounterSubject, AccountingItem] RunTime Purpose Security Currency ValuationPrinciple Value FinancialBalance[Subject, CounterSubject, AccountingItem] RunTime Purpose Security Currency Value 	 AccountingItem CashFlowItem Incomeltem ExpenditureItem ResultItem CostItem RevenueItem BalanceItem Asset Liability
--	--	---

۹ <u>۱</u> ۲		
 □ BusinessActor □ Location □ Sector □ KindOfActivity □ Seller □ Buyer 	 Topic: Trade TradeTransaction[Seller, Buyer, Product] Volume Value 	roduct Commodity Service