

Going Horizontal

Comparing
Open Data Vocabularies
Across Domains



Agenda

- **Issues**
- Chains
- Methods
- Vocabularies
- Comparison
- Future

Gottlob Frege (1848 – 1925)

- Frege investigated identity statements
 - 'The morning star is the evening star.'
- Are the two are equivalent, like $a = b$?
- Then this is the same:
 - 'The morning star is the morning star.'
- But: the first statement conveys knowledge, the second is tautological

Intension and extension

- Intension (Sinn) is a concept we have, some knowledge about a name
- Extension (Bedeutung) is the thing a name refers to

word → intension → extension

- A word determines the intension
- An intension determines the extension

Intension and extension 2

word → intension → extension

'evening star' → very bright star

at sunset → Venus

'The morning star is the evening star.'

- Same extension, different intension

'The morning star is the morning star.'

- Same extension, same intension

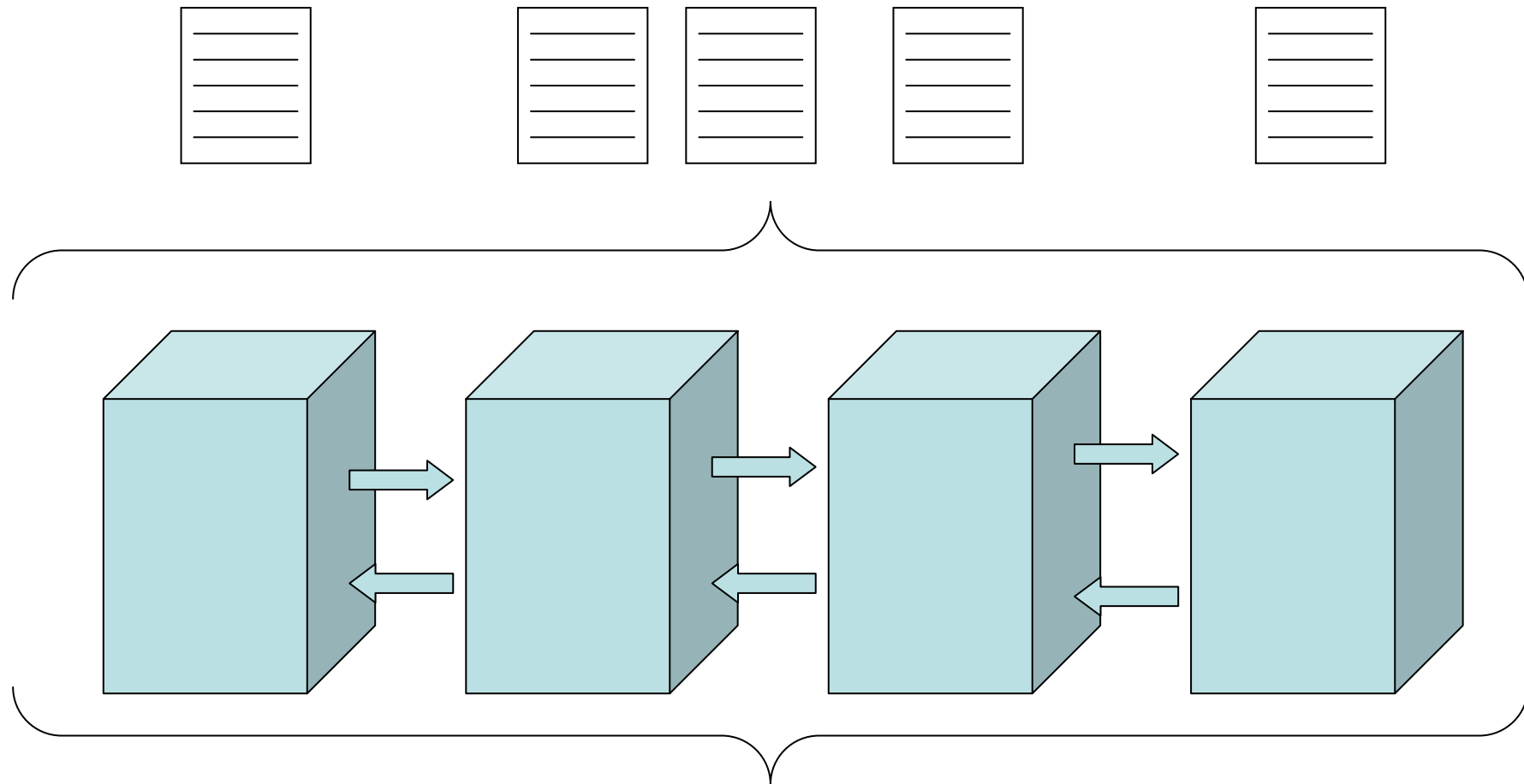
Ludwig Wittgenstein (1889 – 1951)

- What is the meaning of the word ‘game’?
- Is it competition? skill? exercise?
- There is no fixed meaning
- A ‘family of characteristics’
- ‘Meaning is use in language’
- ‘Concept with hazy edges’
- Wittgenstein: ‘Spiel’ = ‘game’ or ‘play’?

Agenda

- Issues
- **Chains**
- Methods
- Vocabularies
- Comparison
- Future

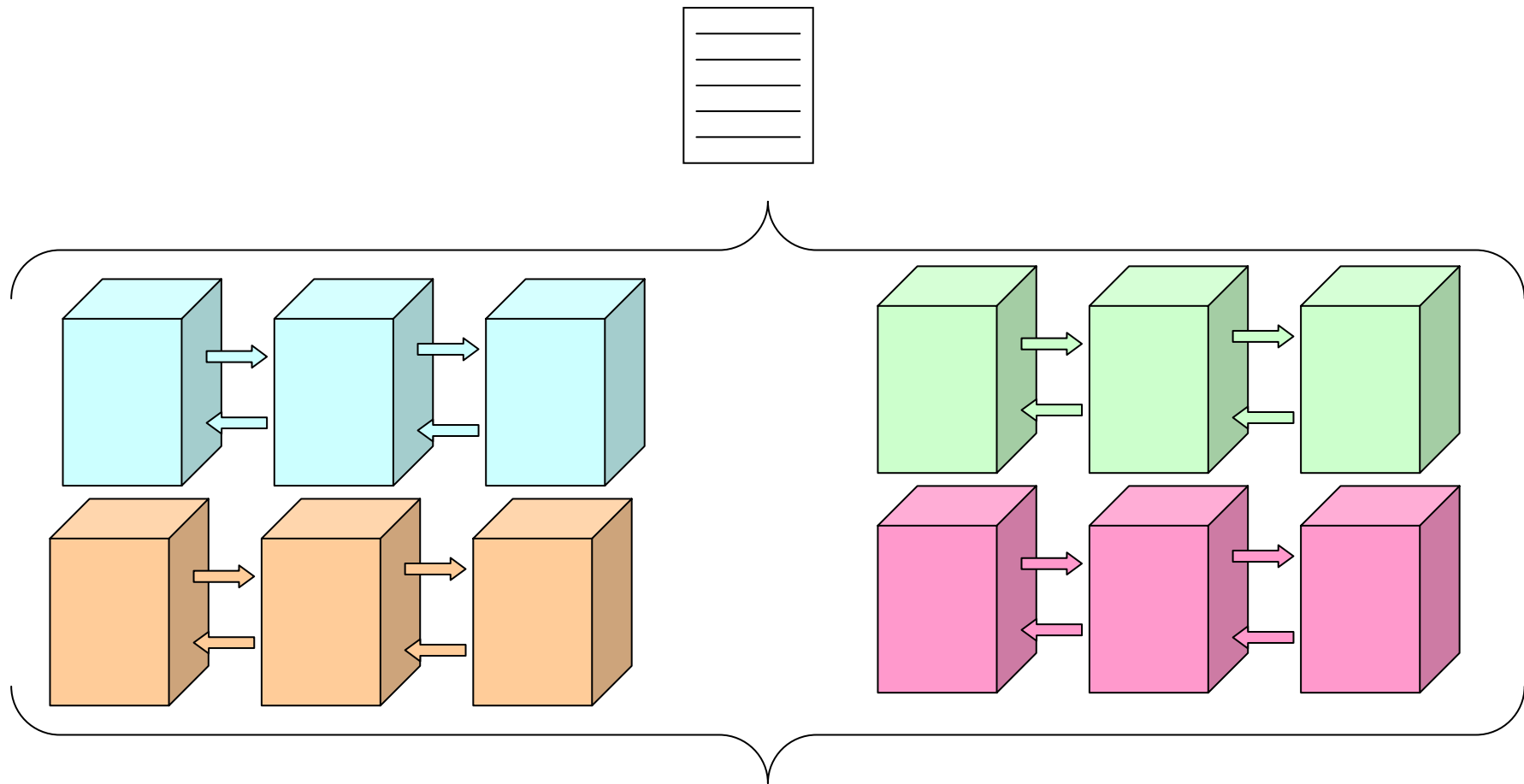
'Meaning' in the Chain



Meaning across Chains

- IT in Dutch government: Islands
- Investigation
 - comparison of methods
 - comparison of vocabularies
 - comparison of maintenance / versioning
- Compatibility
 - unify shared concepts?
 - avoid Big Brother?

'Meaning' in the Chains ?



The Chains

- ePV: Criminal Justice Chain
 - Police, DA, Court
- AORTA: Healthcare Chain
 - hospitals, apothecaries, GP's ...
- SUWI: Social Security & (Un)Employment
 - CWI, UWV, SVB, GSD en BKWI

Agenda

- Issues
- Chains
- **Methods**
- Vocabularies
- Comparison
- Future

Methods – Criminal Justice

- Stolen sheets ...
 - Electronic Information Interchange in the Criminal Justice System in the Netherlands, Pim van der Eijk, XML Europe 2004

Enumeration of (valid) grounds for suspicion
Generated text

Description

```
{ [De adem rook naar inwendig gebruik van alcohol]
- De adem van de [bestuurder/bestuurster] naar het inwendig gebruik van alcoholische drank rook.
[De verdachte had bloeddoorlopen ogen]
- De [bestuurder/bestuurster] bloeddoorlopen ogen had.
[De verdachte sprak met zgn. dubbele tong]
- De [bestuurder/bestuurster] met dubbele tong sprak.
[De verdachte was onvast ter been]
- De [bestuurder/bestuurster] onvast ter been was.
[De verdachte is niet in staat zijn wil kenbaar te maken]
- De [bestuurder/bestuurster] niet in staat was [zijn/haar] wil kenbaar te maken. Dit bleek mij <medewerker NAAM-ZIN, relatie verbalisant waarneming, aantal = 1>, uit: <Omschrijving waaruit blijkt dat [bestuurder/bestuurster] niet in staat was>
[Ander (zelf invoeren)]
- <Omschrijving waarnemingen mbt het gedrag van de bestuurder/bestuurster>
}
```

Kenmerken bestuurder voertuig

ISO 11179 / UBL

Aanduiding:

Aanleiding verdenking
Aanleiding Verdenking

Property:

Property qualifier:

Object class:

Object class qualifier:

Representation:

Delictspecifiek (art. 8 WvW)

Code

Definition

Repr. tabel:

Kenmerken bestuurder

Definitie:

Aanduiding van het gedrag, de uiterlijke kenmerken of andere indicaties die een opsporingsambtenaar deden vermoeden dat de bestuurder van een voertuig onder invloed van alcohol of een andere stof verkeerde.

Herkomst definitie:

SPI

Formaat:

n2

Value set

Layout:

n.v.t.

Toegestane waarden:

Waarden volgens de tabel 'Kenmerken bestuurder'

Code	Omschrijving
01	adem riekte naar inwendig gebruik alcoholhoudende drank
02	kennelijk onder invloed van een stof als bedoeld in art. 8, 1 ^e lid van de Wegenverkeerswet 1994, namelijk.....
03	bestuurder had bloeddorlopen ogen
04	bestuurder sprak met dubbele tong
05	bestuurder was onvast ter been, namelijk.....
06	<i>resultaat ademtest</i>
07	anders, namelijk.....

Commentaar:

n.v.t.

Generated
Element

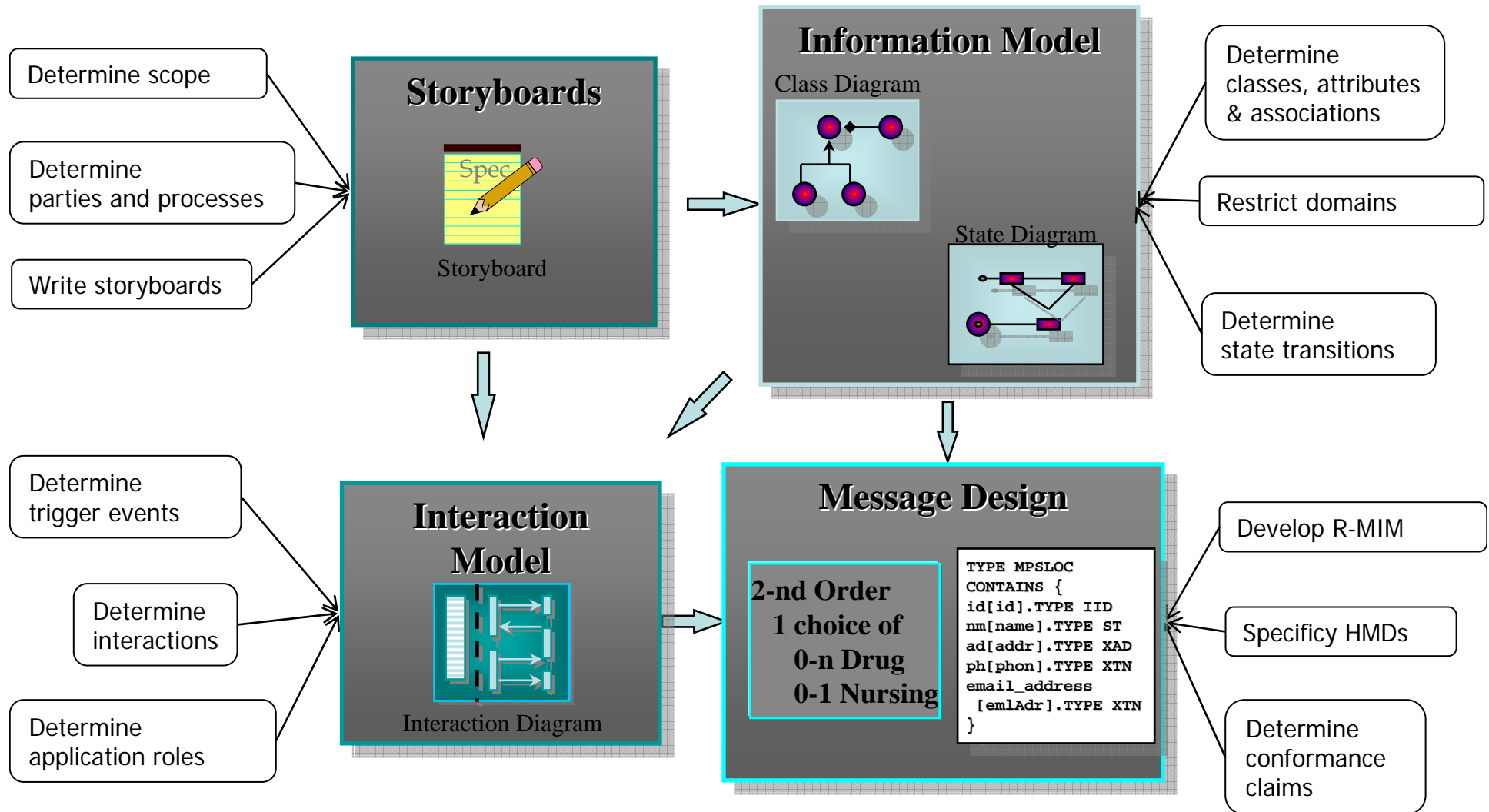
```
<xsd:element ref="AanleidingVerdenkingCode" minOccurs="0" maxOccurs="1">  
- <xsd:annotation>  
- <xsd:documentation>  
- <ccts:Component>  
  <ccts:CategoryCode>BBIE</ccts:CategoryCode>  
  <ccts:DictionaryEntryName>Delictspecifiek Art 8 Wvw. Aanleiding Verdenking.  
  Code</ccts:DictionaryEntryName>  
  <ccts:Definition>Aanduiding van het gedrag de uiterlijke kenmerken of andere  
  indicaties die een opsporingsambtenaar deden vermoeden dat de bestuurder van  
  een voertuig onder invloed van alcohol of een andere stof  
  verkeerde.</ccts:Definition>  
  <ccts:ObjectClass>Delictspecifiek Art 8 WvW</ccts:ObjectClass>  
  <ccts:PropertyTerm>Aanleiding Verdenking</ccts:PropertyTerm>  
  <ccts:RepresentationTerm>Code</ccts:RepresentationTerm>  
</ccts:Component>  
</xsd:documentation>  
</xsd:annotation>  
</xsd:element>
```

Value from
code list

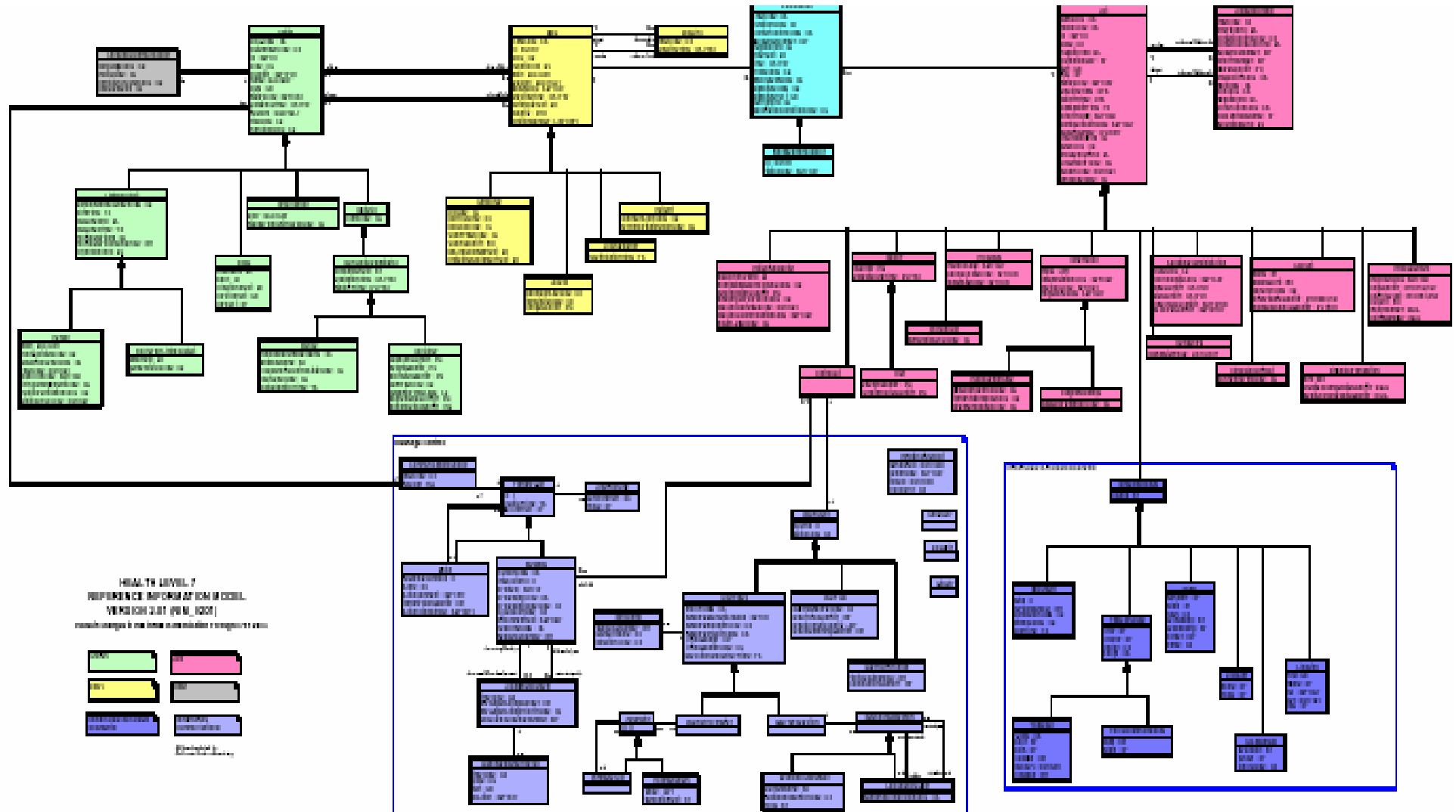
Methods - Healthcare

- Stolen sheets ... copied from Tom de Jong
- Methodology from HL7
- Based on UML

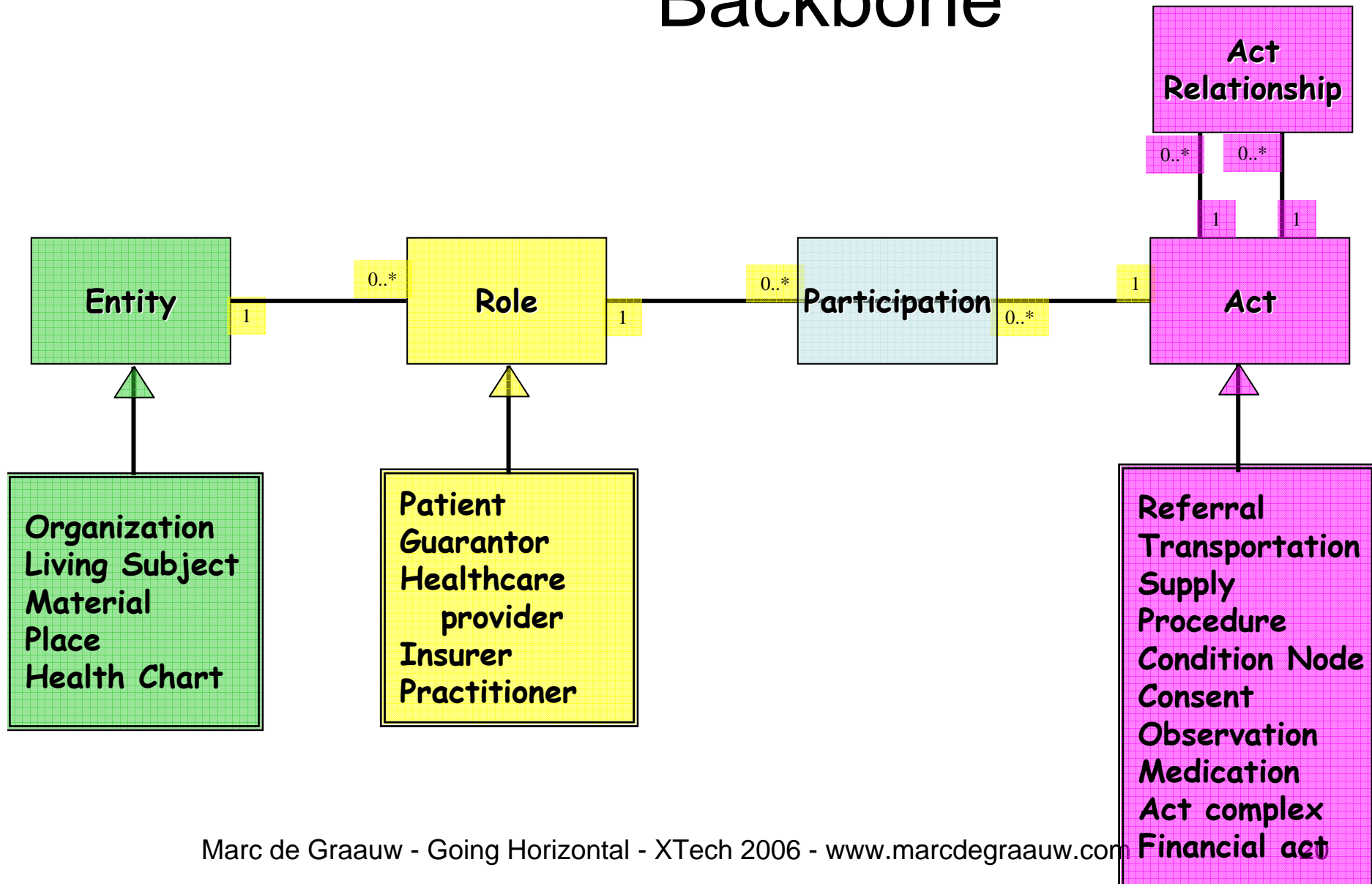
HL7 Development Framework (HDF)



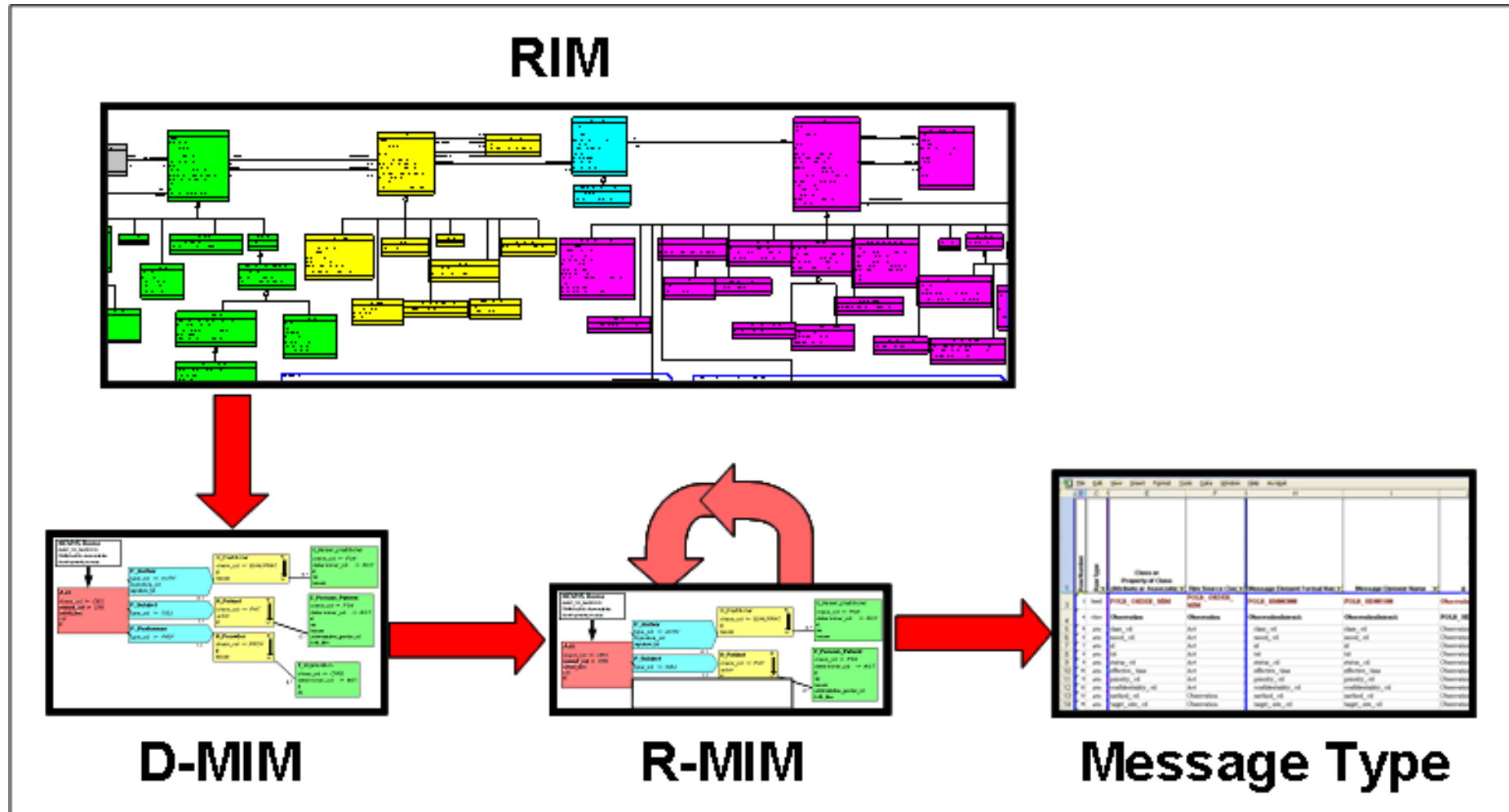
RIM (Reference Information Model)



RIM (Reference Information Model) Backbone



Refinement through 'Constraints'



```
- <xs:complexType name="COCT_MT030200NL.Person">
  - <xs:sequence>
    <xs:element name="id" type="II" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="name" type="PN" minOccurs="0" maxOccurs="1"/>
    <xs:element name="statusCode" type="CS" minOccurs="0" maxOccurs="1"/>
    <xs:element name="administrativeGenderCode" type="CV" minOccurs="0" maxOccurs="1"/>
    <xs:element name="birthTime" type="TS" minOccurs="0" maxOccurs="1"/>
    <xs:element name="deceasedInd" type="BL" minOccurs="0" maxOccurs="1"/>
    <xs:element name="deceasedTime" type="TS" minOccurs="0" maxOccurs="1"/>
    <xs:element name="multipleBirthInd" type="BL" minOccurs="0" maxOccurs="1"/>
    <xs:element name="multipleBirthOrderNumber" type="INT" minOccurs="0" maxOccurs="1"/>
    <xs:element name="maritalStatusCode" type="CV" minOccurs="0" maxOccurs="1"/>
    <xs:element name="educationLevelCode" nillable="true" type="CE" minOccurs="0" maxOccurs="1"/>
    <xs:element type="COCT_MT030200NL.Birthplace" nillable="true" minOccurs="0" maxOccurs="1" name="birthplace"/>
    <xs:element type="COCT_MT030200NL.ContactParty" nillable="true" minOccurs="0" maxOccurs="unbounded"/>
```

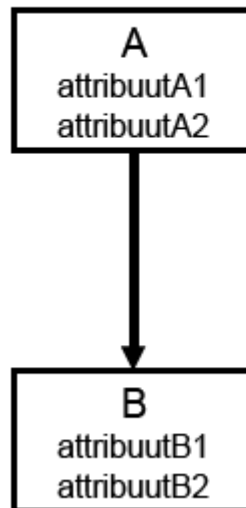
Methods – Social Security

- First release published 1-1-2002, part of ‘SUWI Law’
- Based on Entity Relationship Diagrams
 - Entities, Attributes, Relationships
 - ERD for each major entity (person, address...)
- Derived from this the ‘Suwi Data Registry’
- TP’s make messages from Data Registry
 - they may change entities and attributes

Methods – Social Security

- Dedicated set of ERD → XML Transformation Rules

Gegevensmodel SGR



Berichtinstantie

```
<A>
  <attribuutA1/>
  <attribuutA2/>
  <B>
    <attribuutB1/>
    <attribuutB2/>
  </B>
</A>
```


Methods – Social Security

- Types: AN..6, N16 etc.
- Transformation to XML Schema Datatypes

```
<simpleType name="AantalN2">  
  <restriction base="nonNegativeInteger">  
    <minInclusive value="0"/>  
    <maxInclusive value="99"/>  
    <totalDigits value="2"/>  
  </restriction>  
</simpleType>
```

- Versioning: major, minor
- Versioning reflected in namespaces
 - xmlns:sml = "http://www.suwi.nl/SuwiML/Basis-v0200"

Agenda

- Issues
- Chains
- Methods
- **Vocabularies**
- Comparison
- Future

Person Criminal Justice

Natuurlijk persoon

Samengesteld gegeven:

Definitie: Een menselijk individu.

Herkomst definitie: Analyse- en Ontwerpteam ePV

Formaat: n.v.t.

Layout: Samengesteld

Toegestane waarden: n.v.t.

Commentaar: Natuurlijk persoon is een samengesteld gegeven en opgebouwd uit:

Datum eerste en laatste verhoor: ...

Militair op pleegdatum

Beginnend bestuurder

Verdachte heeft bekend

Taalbeheersing

Eerder pv terzake van art. 8 WvW

Datum eerste verhoor

Datum laatste verhoor

Antecedenten politie

Functieaanduiding

Tekenbevoegd

Person Criminal Justice

Geboortegegevens

Samengesteld gegeven:

Definitie: Gegevens die betrekking hebben op de geboorte van een persoon.

Herkomst definitie: NEN 1888:2002

Formaat: n.v.t.

Layout: Samengesteld

Toegestane waarden: n.v.t.

Commentaar: Geboortegegevens is een samengesteld gegeven en opgebouwd uit:

Geboortedatum

Geboortejaar

Geboortegemeente (Nederland)

Geboortegemeente, in letters

Geboorteplaatsnaam

Geboortecoördinaten

Nadere aanduiding geboorteplaats

Geboorteland

Geslacht

Person Criminal Justice

Identificatie

Samengesteld
gegeven:

Definitie: Het vaststellen van de identiteit van een persoon.

Herkomst definitie: Analyse- en Ontwerpteam ePV

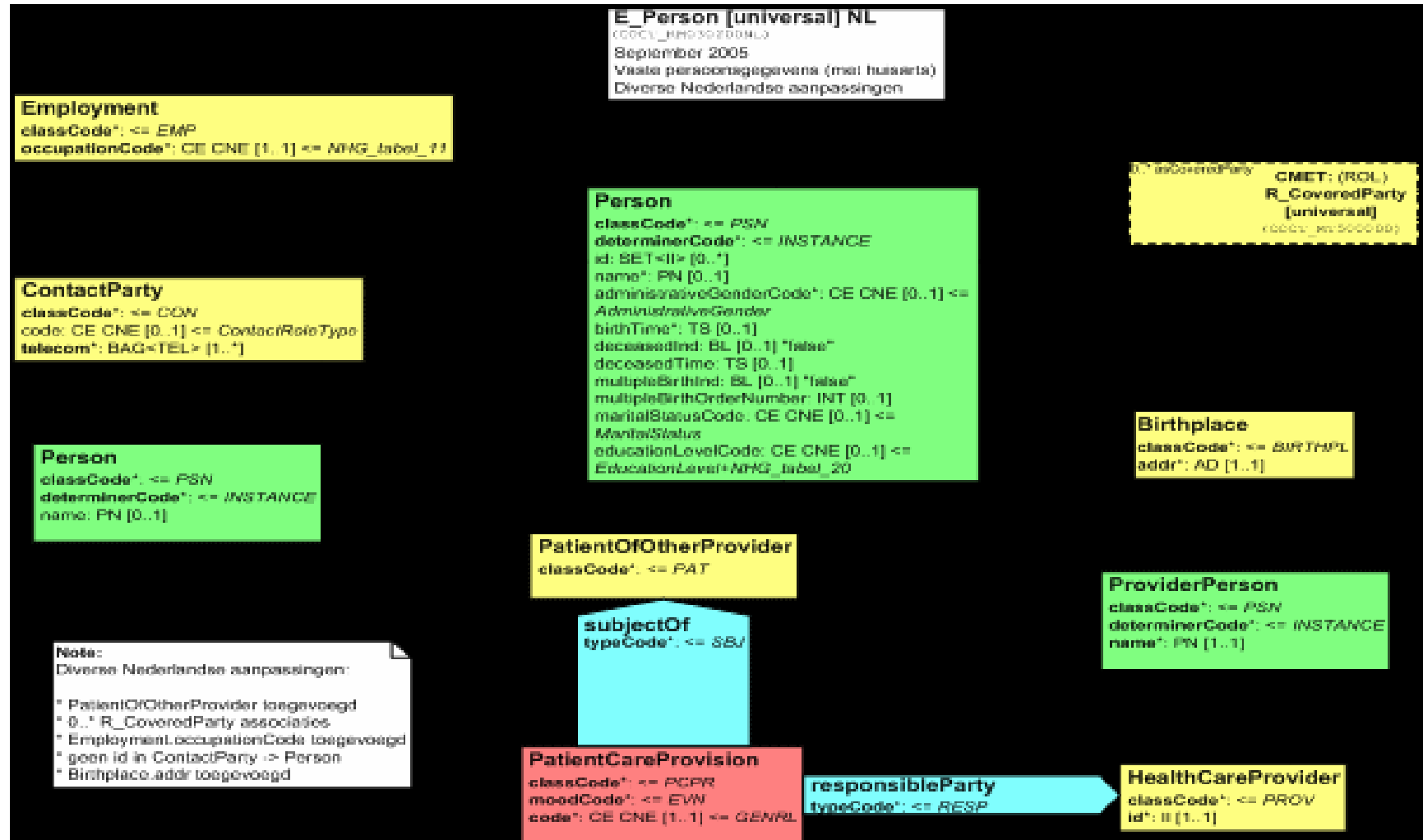
Formaat: n.v.t.

Layout: Samengesteld

Toegestane waarden: n.v.t.

Commentaar: Identificatie is een samengesteld gegeven en opgebouwd uit:
Soort identificatiemiddel
Ander identificatiemiddel
Nummer identificatiemiddel
Rijbewijs getoond door verdachte
Overeenkomst met rijbewijsgegevens

Person Healthcare



Person Healthcare

De klasse *Person* heeft de volgende attributen:

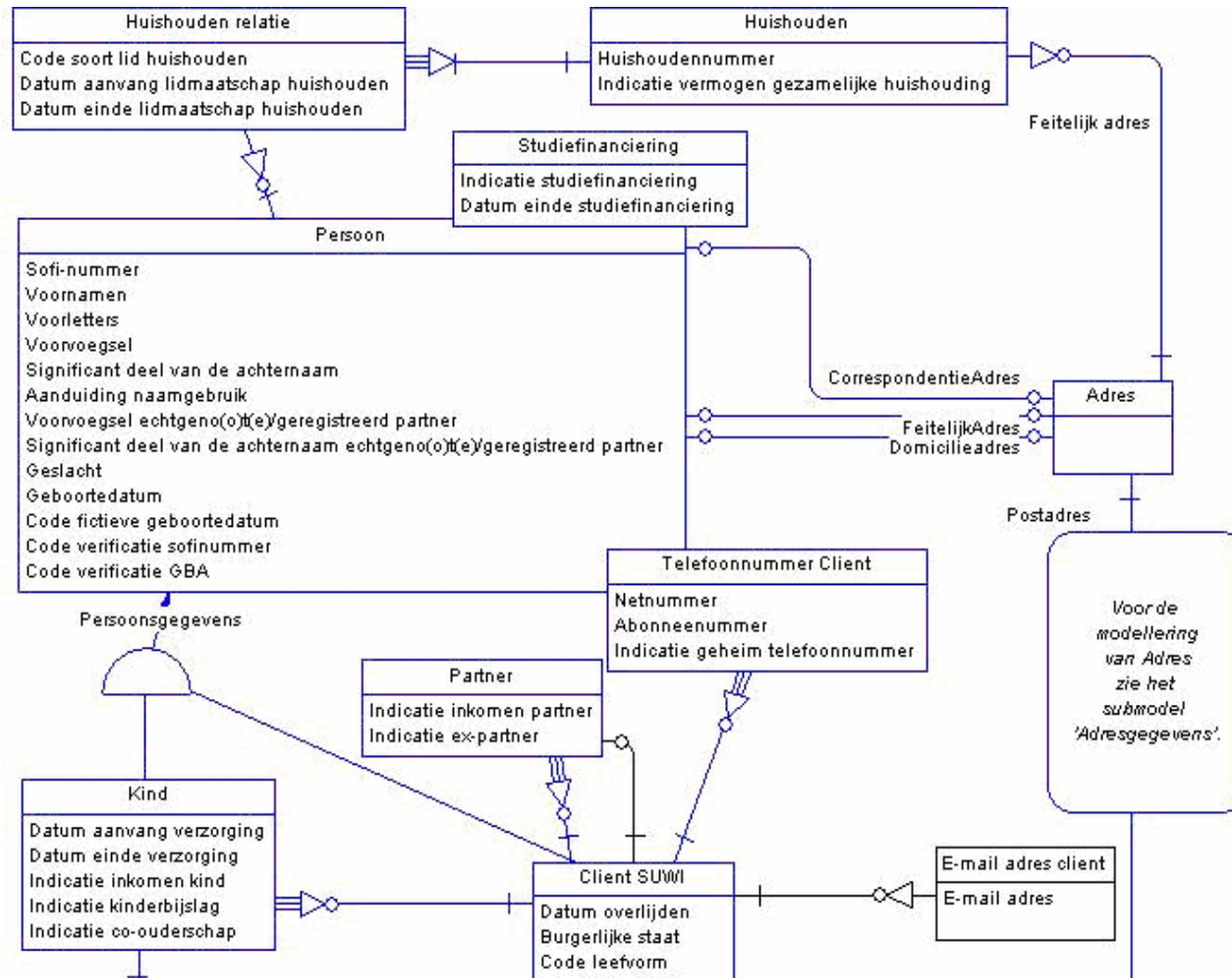
classCode	PSN (Person) Een persoon (mens)
determinerCode	INSTANCE Een specifiek persoon (individu)
id	Persoonsnummer
name	Naam
administrativeGenderCode	Geslacht
birthTime	Geboortedatum (en evt. –tijd)
deceasedInd	Overlijdensindicatie
deceasedTime	Overlijdensdatum (en evt. –tijd)
multipleBirthInd	Meerlingindicatie
multipleBirthOrderNumber	Meerlingvolgnummer
maritalStatusCode	Burgerlijke staat
educationLevelCode	Opleidingsniveau

Person Healthcare

De klasse *Person* heeft de volgende associaties:

0..1	Employment	Beroep
0..*	ContactParty	Contactpers(o)n(en)
0..1	PatientOfOtherProvider	Relatie met de huisarts
0..1	Birthplace	Geboorteplaats
0..*	CoveredParty	Zorgverzekering(en)

Person Social Security



Persoon Social Security

Entiteit Persoon

Naam	Persoon
SuwiML-tag	Persoon

Definitie

Een individueel menselijk wezen.

Norminstantie: NEN 1888:2002 - Persoon

Opmerkingen

- Super-entiteitstype voor de persoon-entiteiten (zoals CLIENT, Kind en Partner).
- Voor een PERSOON zijn drie typen adres gedefinieerd:
 1. Domicilie adres: het adres van de persoon, zoals geregistreerd in de Gemeentelijke BasisAdministratie (GBA);
 2. Feitelijk adres: het adres waarop de PERSOON daadwerkelijk verblijft;
 3. Correspondentie adres: het adres waarop de PERSOON zijn post wenst te ontvangen.

Persoon Social Security

Attributen

Naam	SuwiML-tag
Sofi-nummer	SofiNr
Voornamen	Voornamen
Voorletters	Voorletters
Voorvoegsel	Voorvoegsel
Significant deel van de achternaam	SignificantDeelVanDeAchternaam
Aanduiding naamgebruik	AanduidingNaamgebruik
Voorvoegsel echtgeno(o)t(e)/geregistreerd partner	VoorvoegselEchtgenoot
Significant deel van de achternaam echtgeno(o)t(e)/geregistreerd partner	SignificantDeelAchternaamEchtg
Geslacht	Geslacht
Geboortedatum	Geboortedat
Code fictieve geboortedatum	CdFictieveGeboortedat
Code verificatie sofinummer	CdVerificatieSofinr
Code verificatie GBA	CdVerificatieGba

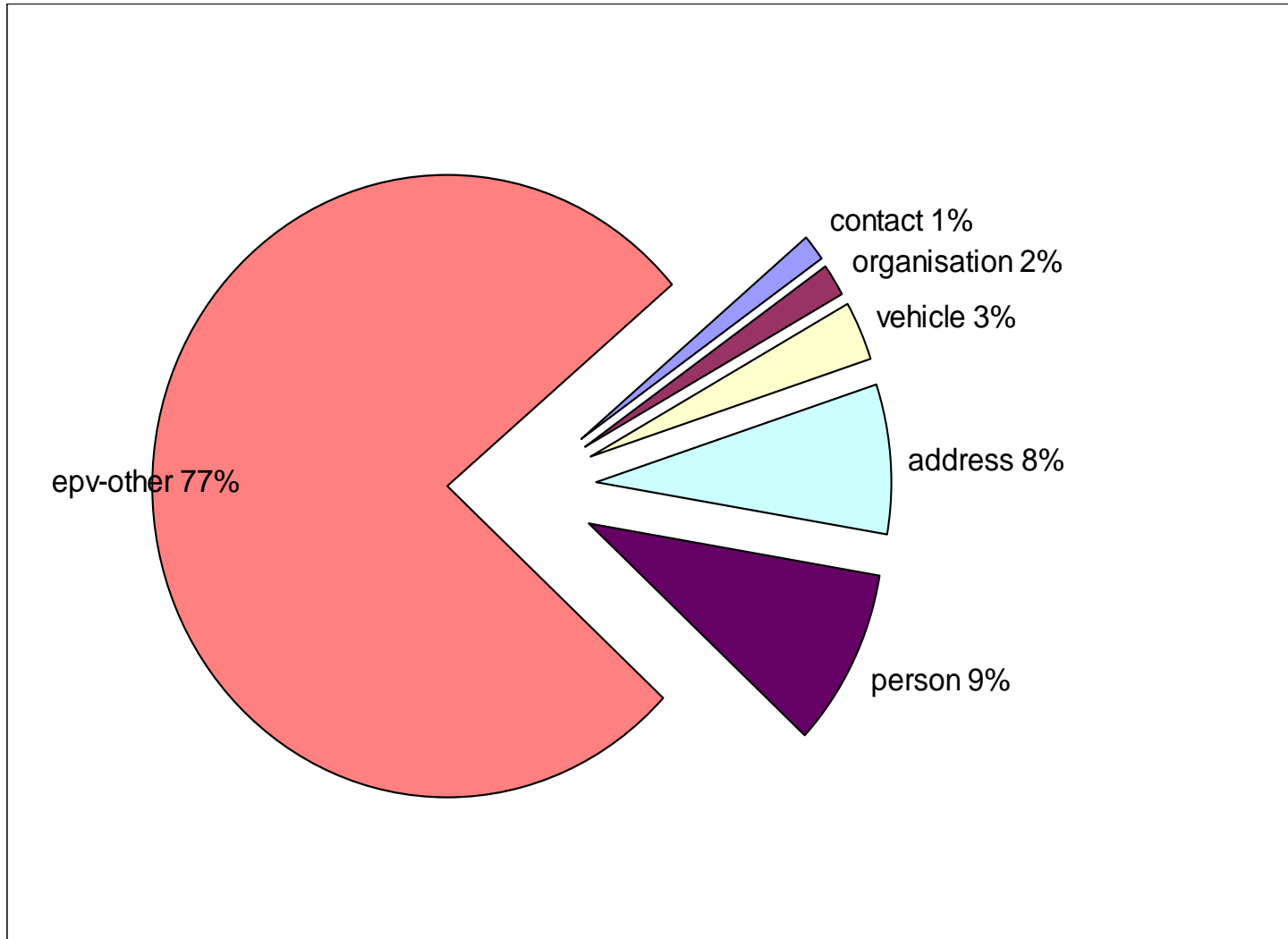
Agenda

- Issues
- Chains
- Methods
- Vocabularies
- **Comparison**
- Future

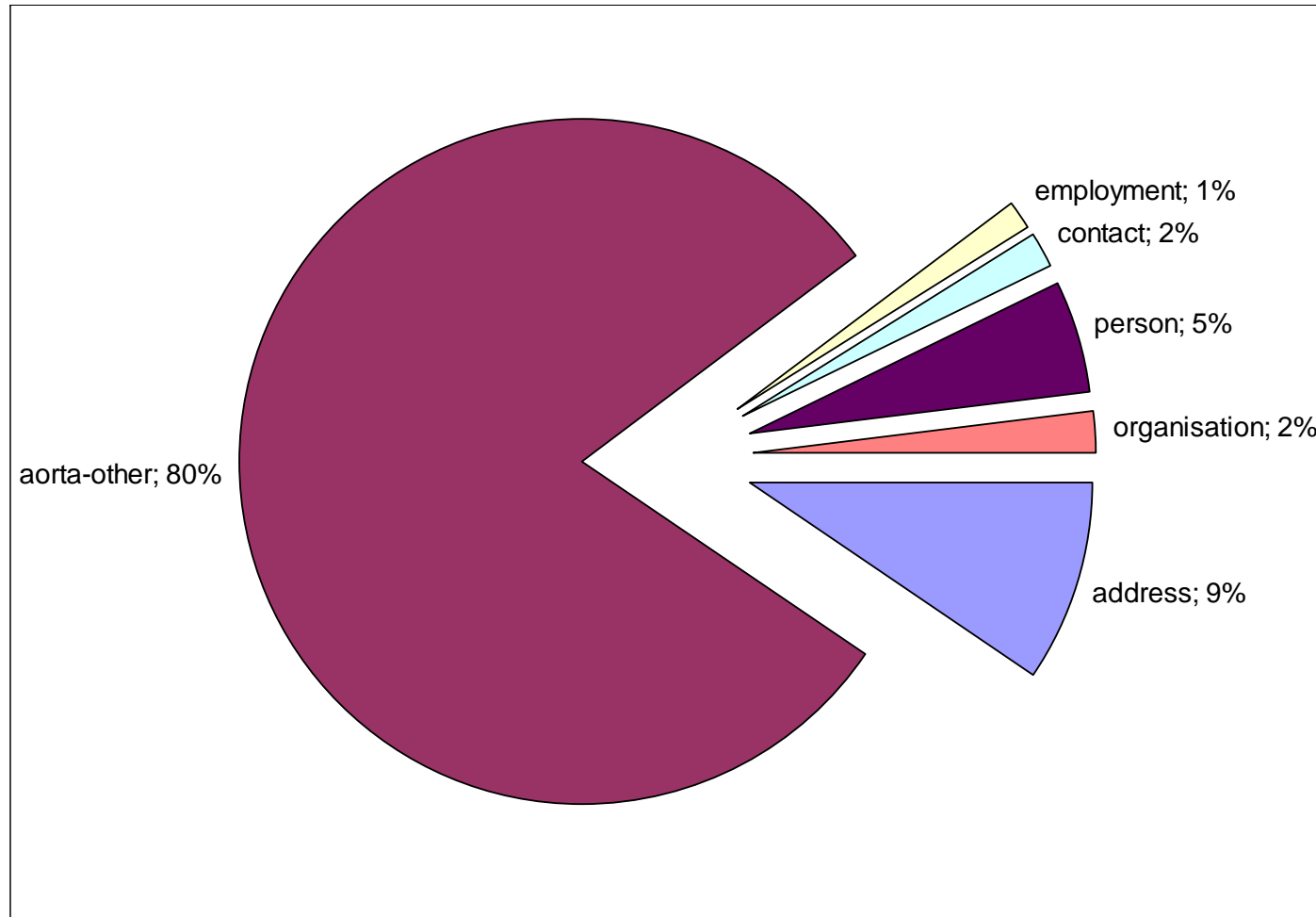
Tags by domain

	care	housing	vehicle	benefit	suwi-other	organisation	person	partnership	education	money	contact	employment	address
Social Security		x	x	x	x	x	x	x	x	x	x	x	x
Criminal Justice			x			x	x				x		x
Healthcare	x					x	x				x		x

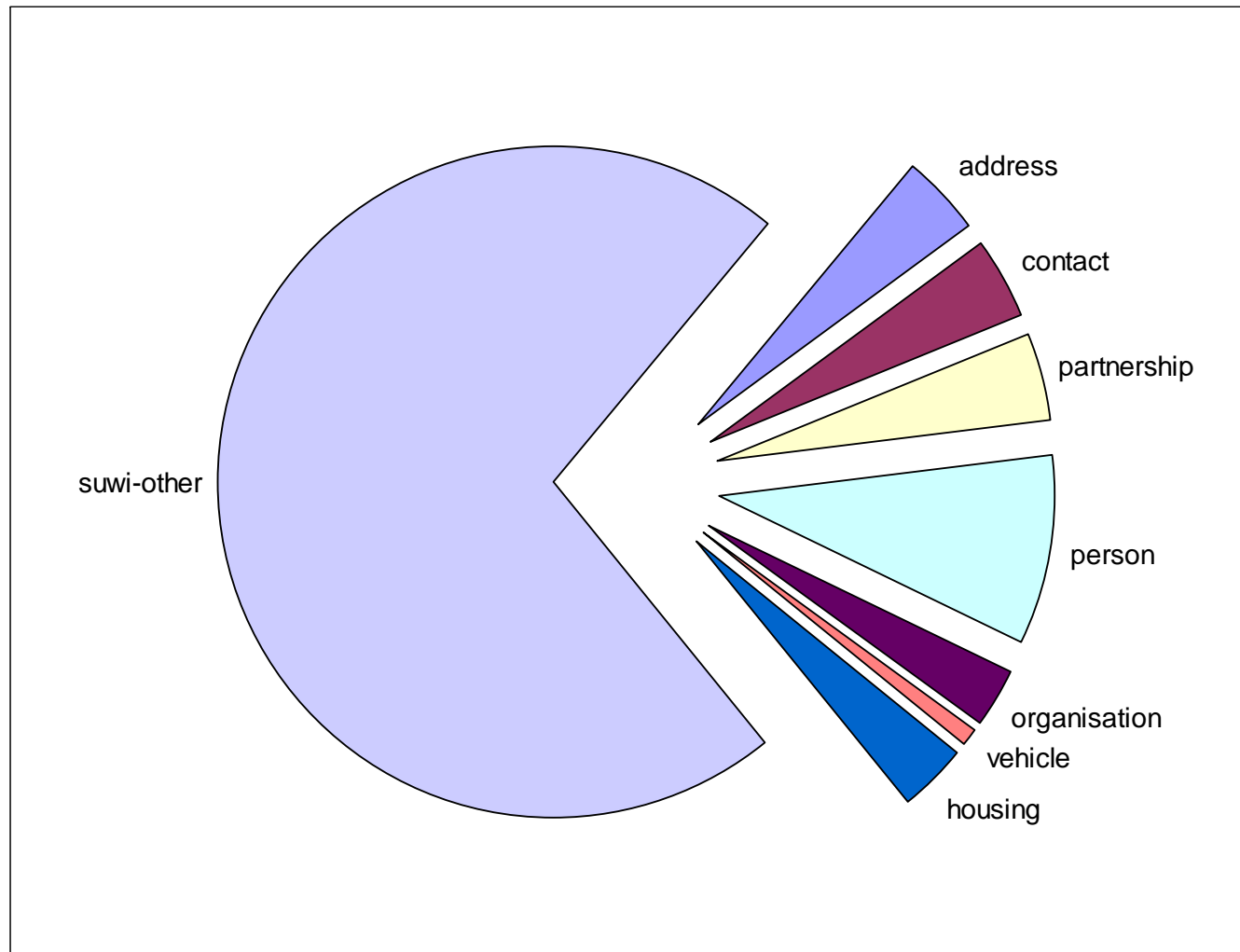
Criminal Justice distribution



Healthcare distribution



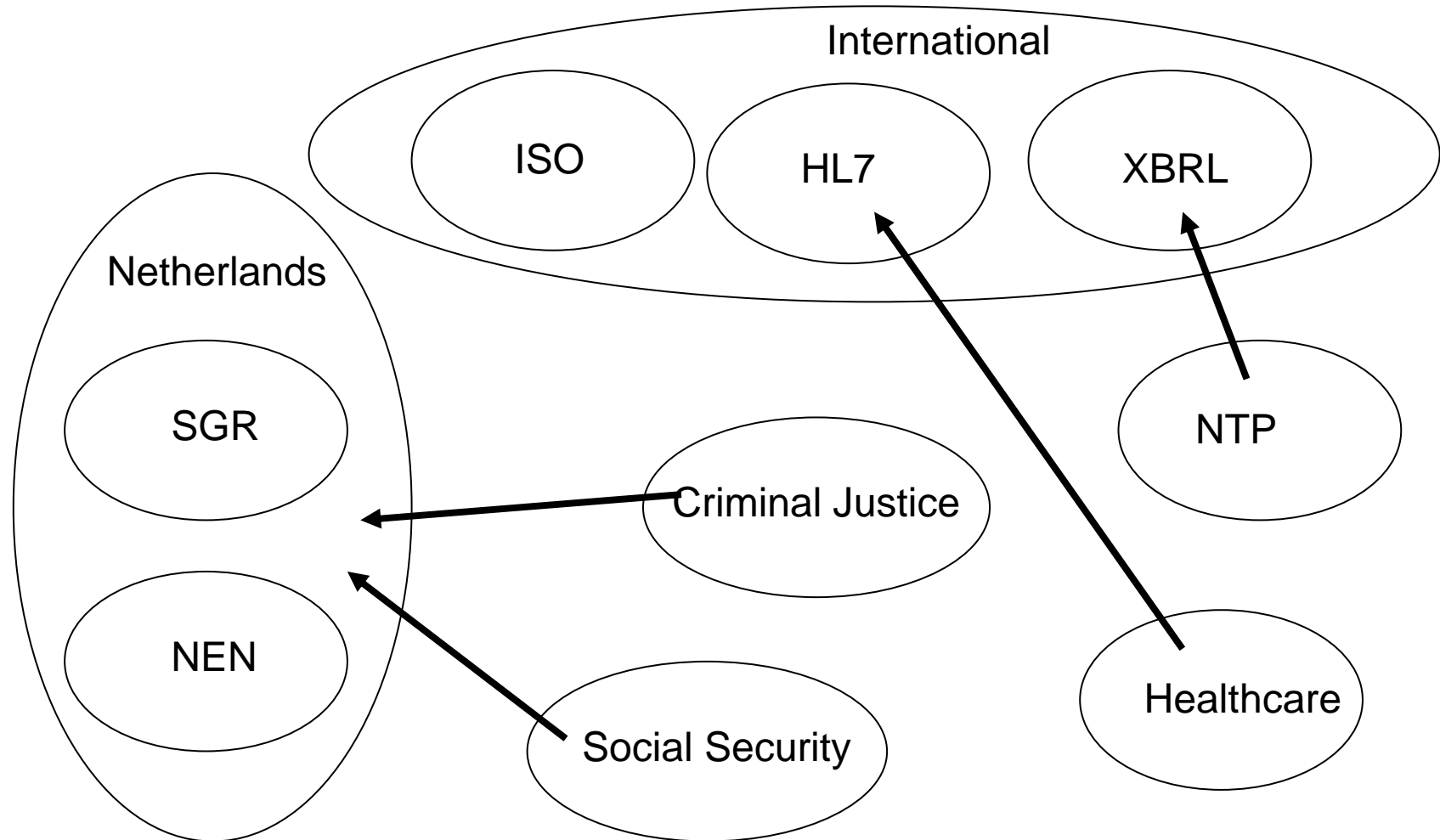
Social Security distribution



Domain comparison

- domain-specific concepts: 75%
- cross-domain concepts: 25%
- cross-domain concepts are mainly:
 - person
 - address
 - contact
 - organisation
 - (finance)

Orientation vocabularies



Issues

syntax

- syntactical issues
 - serialization to XML, other
- datatypes
 - Healthcare: own datatypes, translated to XSD
 - Social Security: AN..12 / N2 / D etc., translated to XSD
 - Criminal Justice: UBL and XML Schema datatypes

Agenda

- Issues
- Chains
- Methods
- Vocabularies
- Comparison
- **Future**

The next steps

- Put vocabularies online in browseable database (using Rails or Django of course)
- Align with 'Core Repositories' (basisregistraties)
- PR
- Get funding!
- Sharing vocabularies & methods
- Promote microformats for shared entities

Marc de Graauw

- www.marcdegraauw.com
- marc@marcdegraauw.com
- feedback welcome