Dataspecifikationer

– hvordan skal de forstås?

Anders Friis-Christensen (andfr@kms.dk)

Miljøministeriet
Kort & Matrikelstyrelsen

Mål med oplæg

- Give et overblik over dokumenter i høring
- Præsentere strukturen af dataspecifikationerne – Intro til UML
  - Hjælp til prioritering af læsning

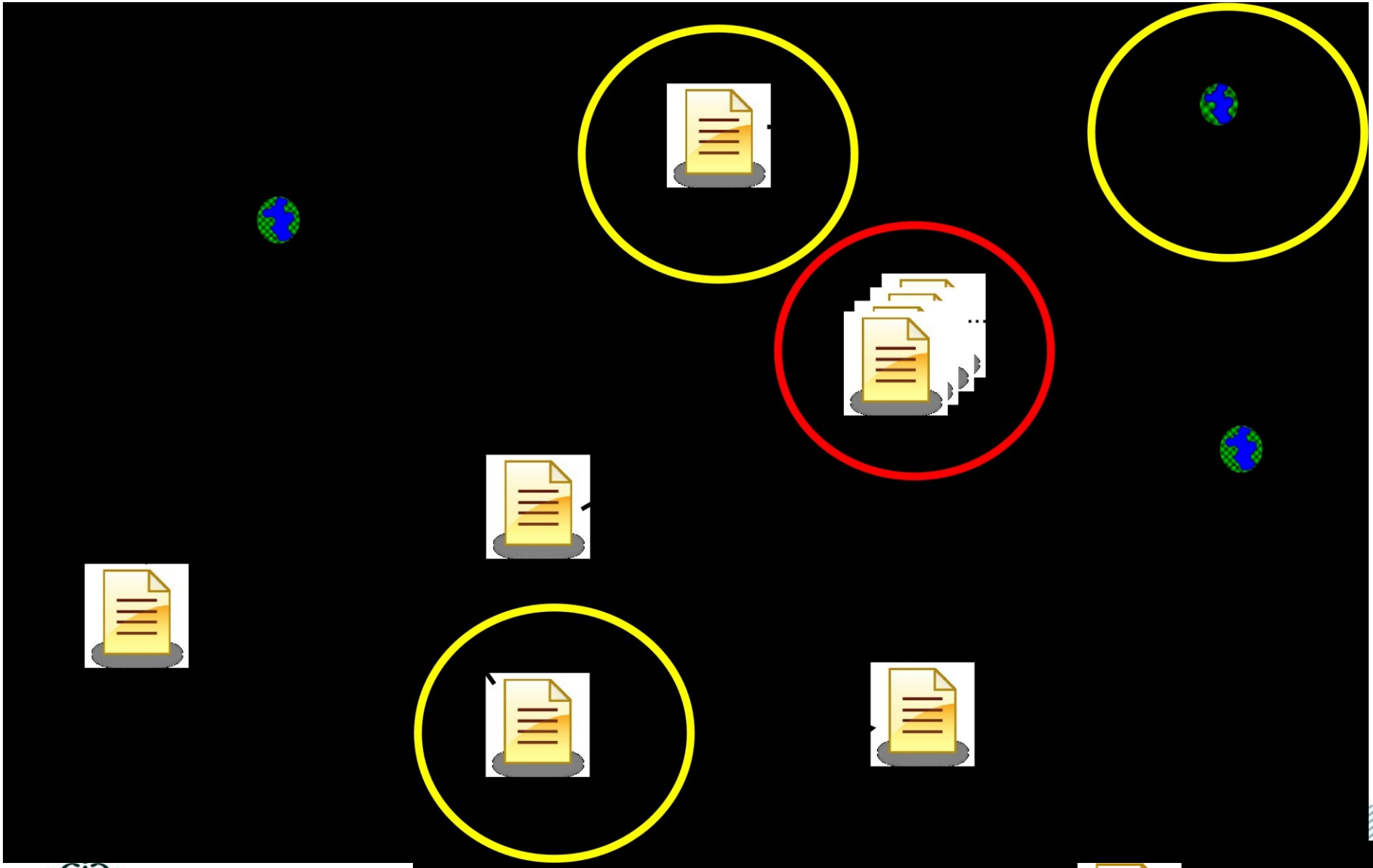
Dokumenter

- Til orientering:
  - D2.3: Definition af temaerne
  - D2.5: Generisk konceptuel model
  - D2.6: Metode til udvikling af dataspecifikationer
  - D2.7: Retningslinjer for kodning
  - Registre
- Feature Concept Dictionary (WEB)
- Glossary (WEB)

– INSPIRE UML model (WEB – men endnu ikke opdateret)

- http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/ d atamodels (forskellige formater)
- For kommentering

– D2.8: Retningslinjer (dataspecifikationer) for temaer
– GML Skemaer (ver 3.2.1)
– D2.5+D2.7 tilføjelser



Miljøministeriet
Kort & Matrikelstyrelsen

Dokumentoversigt

D2.3 D2.3

- http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2

Dokumentoversigt

Dokumentoversigt

Miljøministeriet
Kort & Matrikelstyrelsen

D2.3: Definition af temaerne

- Indeholder temabeskrivelser
- Er IKKE opdateret med de nye beskrivelser fra bilag 2+3
  - Det er meningen at afsnit 2 i dataspecifikationerne skal sættes ind

- -> lad være med at bruge tid på dette dokument

Generelle dokumenter (D2.5,D2.6,D2.7)

- Giver generelle regler og anbefalinger til udfærdigelsen af data specifikationerne
  - D2.5: Generisk konceptuel model
- Sætter krav til f.eks. modellering af netværk og livscyklus af objekter
  - D2.6: Metode til udvikling af dataspecifikationer
    - Giver anbefaling til udvikling af dataspecifikationer (ingen krav)
  - D2.7: Retningslinjer for kodning (af format)

- Enkelte krav, ellers anbefaling for oversættelse fra UMLmodellerne til det valgte kodningsformat (GML er anbefalet som default)
- Opdateringer til D2.5 og D2.7 (fra tidligere versioner) er i høring!
  - Tilføjelser er præsenteret i et separat dokument

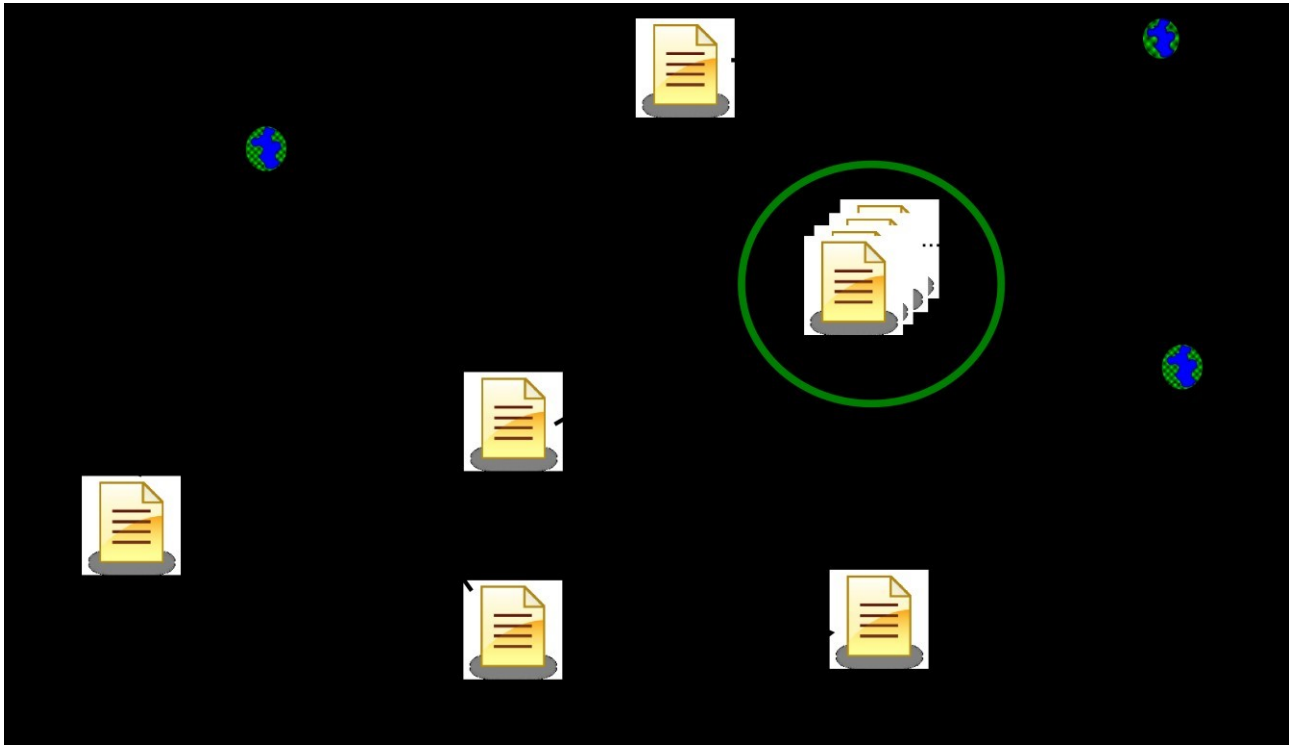
Registre: Feature concept dictionary og glossary

- Feature Concept Dictionary
  - Register over samtlige feature typer og deres definitioner specificeret for alle temaer (endnu ikke bilag 2+3)
  - Indeholder desuden definitionen af temaet
- Glossary
  - Register over termer/ord (definitioner) benyttet i de generelle dokumenter og i dataspecifikationerne
- Begge er implementeret og kan benyttes som opslagsværk

- Pt. er de IKKE opdateret med bilag 2+3 information, men det bliver disse der er gældende (frem for dokumenterne)
  - lad være med at bruge tid på disse (andet end hvis I er nysgerrige..)

Dokumentoversigt

D2.3



Specifikationer (D2.8: Retningslinjer)

- 25 specifikationer (for bilag 2+3) udarbejdet af 19 ekspertgrupper sammensat af personer fra hele Europa
- Specifikationerne bygger strukturen defineret i ISO 19131 (Data Product Specification)

Generelt

- Tekster markeret i gråt er gennemgående (fælles) tekst for alle specifikationer
  - Dataspecifikationerne:

- Er IKKE juridiske dokumenter
  - Leverer input til juridisk tekst – Indeholder krav
  - DS Requirement: Krav til en implementering der følger retningslinjerne, men kommer ikke i en juridisk tekst

DS Requirement X	Requirements that are not reflected in the Implementing Rule on interoperability of spatial data sets and services are shown using this style.

  - IR Requirement: Bliver omsat til juridisk krav

IR Requirement X	Requirements that are reflected in the Implementing Rule on interoperability of spatial data sets and services are shown using this style.

- Indeholder anbefalinger
  - Er blot anbefalinger der ikke kommer med i en juridisk tekst

Recommendation 1 Recommendations are shown using this style.

Indhold af dataspecifikation

Executive Summary

1. Scope (of the Document)
2. Overview
3. Specification scopes
4. Identification information
7. Data quality
8. Dataset-level Metadata
9. Delivery
10. Data Capture (optional)
11. Portrayal

Bibliography

Annex A (normative) 5. Data content and structure

Abstract Test Suite

6. Reference systems

Annex n …

Executive summary

- Sammenfatning af specifikationen
  - En generel sammenfatning af dataspecifikationsarbejdet
- Skrevet af kommissionen
  - Sammenfatning af den specifikke specifikation
- Skrevet af ekspertgruppen

- Relevans for høring: Relevant for at få et overblik og forstå konteksten

(1) Scope, (2) Overview, og (3) Specification scopes

- 1 Scope
  - Formål med dokumentet
- 2 Overview
  - Indeholder definitionen/beskrivelsen af temaet
- Også repræsenteret i Feature Concept Dictionary (på et senere tidspunkt)
- 3 Specification scopes
  - I alle specifikationer indeholder den ingenting...

- Relevans for høring: Definition og beskrivelse i afsnit 2
  - Det er her afgrænsning af temaet defineres
- Er I enige? Er det for vidtrækkende? Uklart?

Identification information (4)

- Er udgået efter kritik af at indholdet var redundant ift. resten af specifikationen

☺

(5) Data content and structure

- Indeholder det centrale i dataspecifikationen
  - Tekstlig beskrivelse af applikationsskemaet (UML modellen)
  - Selve UML modellen

  - Feature Katalog som indeholder beskrivelser af samtlige feature typer, attributter, associationer, og begrænsninger (constraints)

- Genereret direkte fra UML modellen
- For læsning/gennemgang: Nødvendigt med kendskab til UML!

(5) Data content and structure

- Relevans for høring:
  - Giver den tekstlige beskrivelse mening?
  - UML klassediagram(mer):
- Feature typer

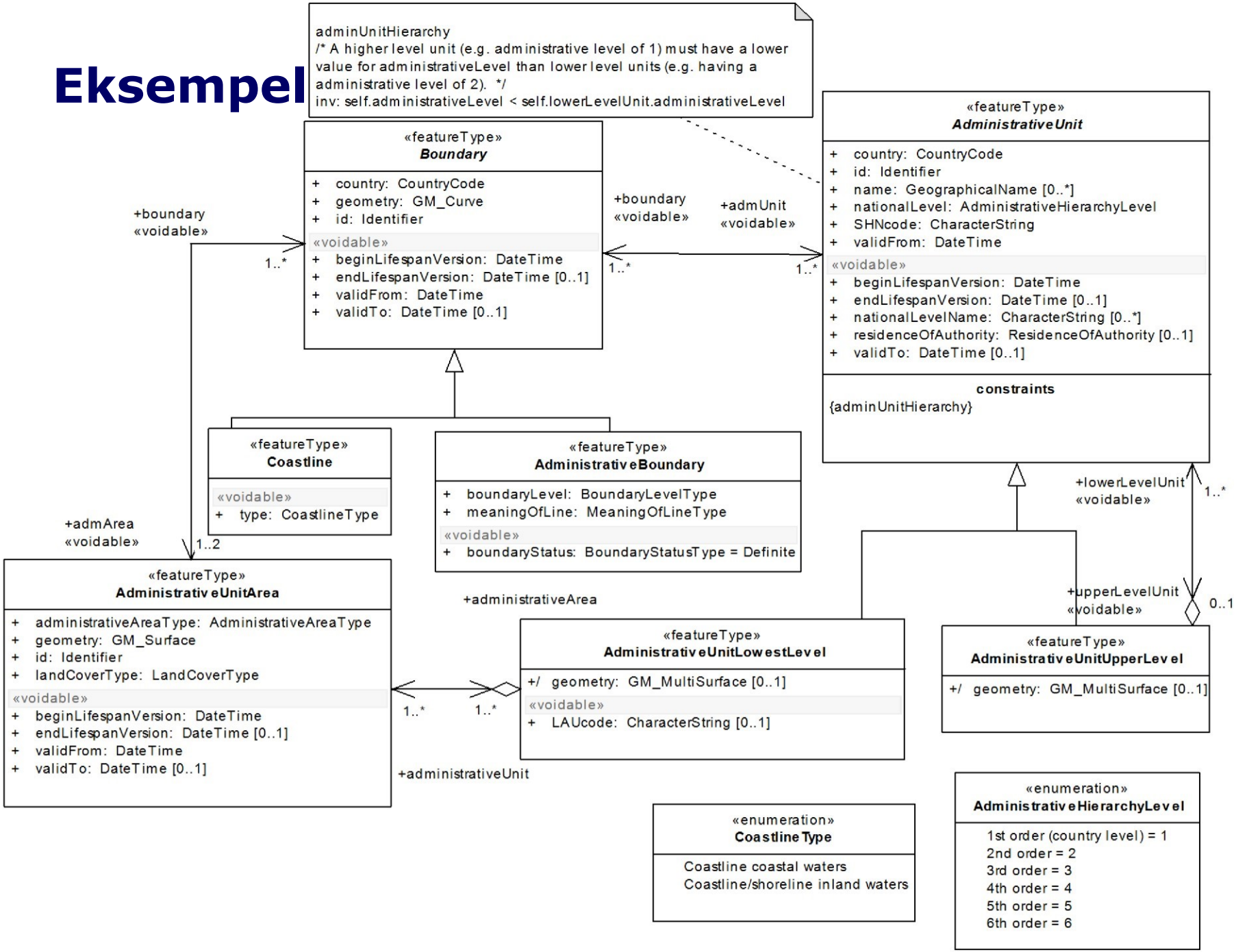
Kan betragtes fælles

- Attributter
- Associationer
- Multipliciteter
- OCL constraints – begrænsninger

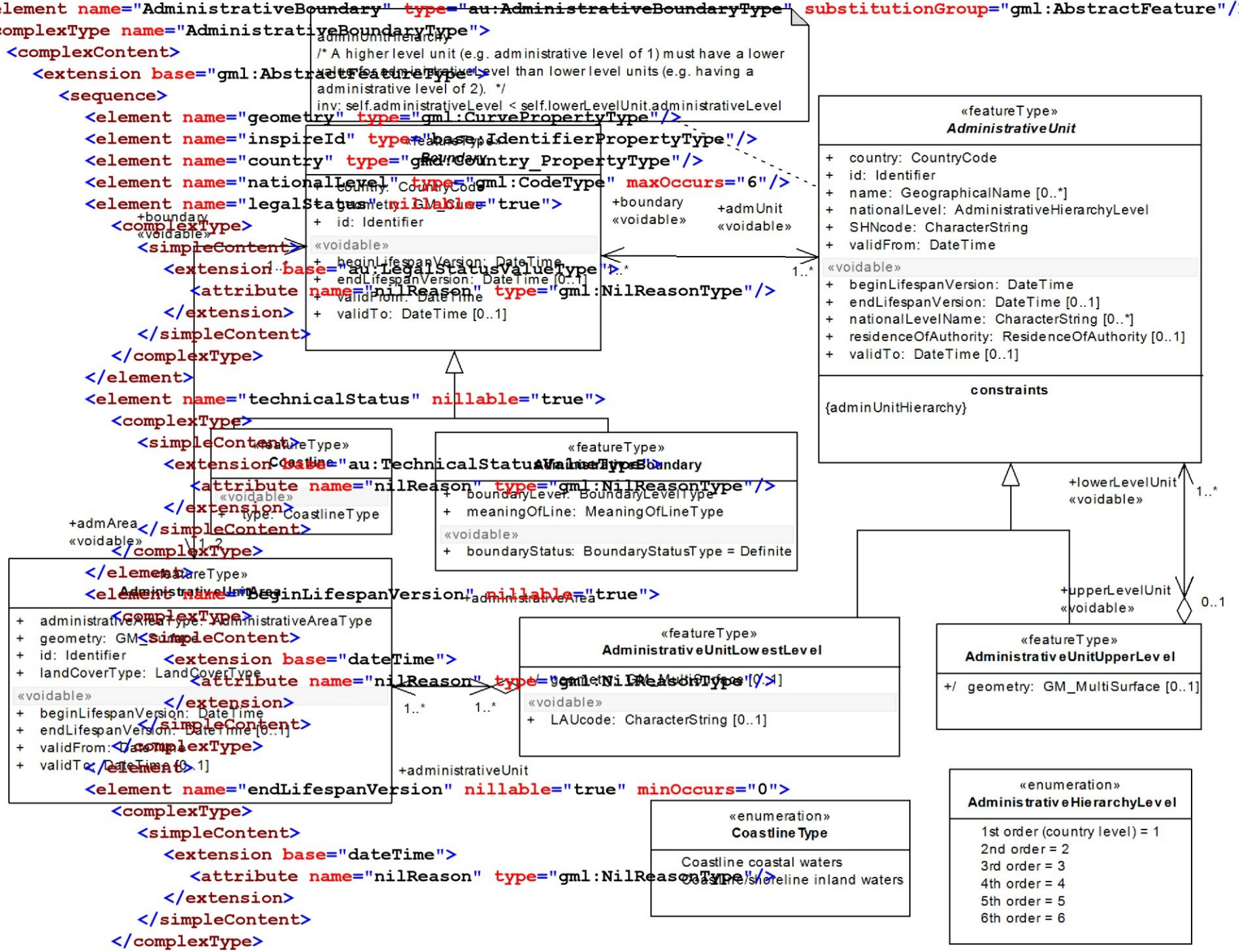
– Feature Katalog

- Definitioner på ovenstående

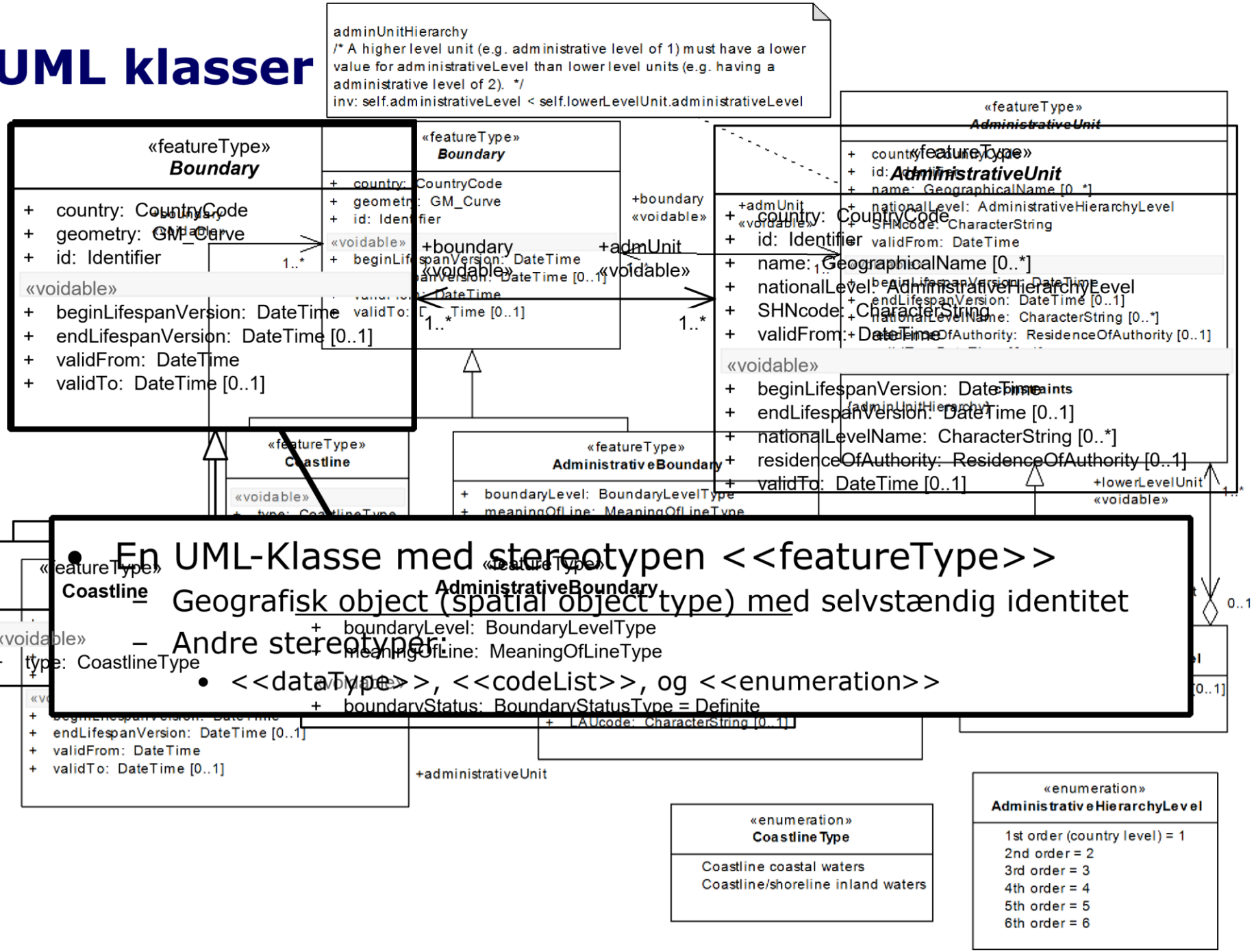
Eksempel


adminUnitHierarchy
/* A higher level unit (e.g. administrative level of 1) must have a lower value for administrativeLevel than lower level units (e.g. having a administrative level of 2). */
inv: self.administrativeLevel < self.lowerLevelUnit.administrativeLevel
«featureType»
Boundary
+ country: CountryCode
+ geometry: GM_Curve
+ id: Identifier
«voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
+ validFrom: DateTime
+ validTo: DateTime [0..1]
«featureType»
AdministrativeUnit
+ country: CountryCode
+ id: Identifier
+ name: GeographicalName [0..*]
+ nationalLevel: AdministrativeHierarchyLevel
+ SHNcode: CharacterString
+ validFrom: DateTime
«voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
+ nationalLevelName: CharacterString [0..*]
+ residenceOfAuthority: ResidenceOfAuthority [0..1]
+ validTo: DateTime [0..1]
constraints
{adminUnitHierarchy}
+boundary
«voidable»
1..*
+boundary
«voidable»
1..*
+admUnit
«voidable»
1..*
«featureType»
Coastline
«voidable»
+ type: CoastlineType
«featureType»
AdministrativeBoundary
+ boundaryLevel: BoundaryLevelType
+ meaningOfLine: MeaningOfLineType
«voidable»
+ boundaryStatus: BoundaryStatusType = Definite
+admArea
«voidable»
1..2
«featureType»
AdministrativeUnitArea
+ administrativeAreaType: AdministrativeAreaType
+ geometry: GM_Surface
+ id: Identifier
+ landCoverType: LandCoverType
«voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
+ validFrom: DateTime
+ validTo: DateTime [0..1]
+administrativeArea
1..*
1..*
+administrativeUnit
«featureType»
AdministrativeUnitLowestLevel
+/ geometry: GM_MultiSurface [0..1]
«voidable»
+ LAUcode: CharacterString [0..1]
+lowerLevelUnit
«voidable»
1..*
+upperLevelUnit
«voidable»
0..1
«featureType»
AdministrativeUnitUpperLevel
+/ geometry: GM_MultiSurface [0..1]
«enumeration»
CoastlineType
Coastline coastal waters
Coastline/shoreline inland waters
«enumeration»
AdministrativeHierarchyLevel
1st order (country level) = 1
2nd order = 2
3rd order = 3
4th order = 4
5th order = 5
6th order = 6

```
<element name="AdministrativeBoundary" type="au:AdministrativeBoundaryType" substitutionGroup="gml:AbstractFeature"/>
<complexType name="AdministrativeBoundaryType">
  <complexContent>
    <extension base="gml:AbstractFeatureType">
      <sequence>
        <element name="geometry" type="gml:CurvePropertyType"/>
        <element name="inspireId" type="base:IdentifierPropertyType"/>
        <element name="country" type="gmd:Country_PropertyType"/>
        <element name="nationalLevel" type="gml:CodeType" maxOccurs="6"/>
        <element name="legalStatus" nillable="true">
          <complexType>
            <simpleContent>
              <extension base="au:LegalStatusValueType">
                <attribute name="nilReason" type="gml:NilReasonType"/>
              </extension>
            </simpleContent>
          </complexType>
        </element>
        <element name="technicalStatus" nillable="true">
          <complexType>
            <simpleContent>
              <extension base="au:TechnicalStatusValueType">
                <attribute name="nilReason" type="gml:NilReasonType"/>
              </extension>
            </simpleContent>
          </complexType>
        </element>
        <element name="beginLifespanVersion" nillable="true">
          <complexType>
            <simpleContent>
              <extension base="dateTime">
                <attribute name="nilReason" type="gml:NilReasonType"/>
              </extension>
            </simpleContent>
          </complexType>
        </element>
        <element name="endLifespanVersion" nillable="true" minOccurs="0">
          <complexType>
            <simpleContent>
              <extension base="dateTime">
                <attribute name="nilReason" type="gml:NilReasonType"/>
              </extension>
            </simpleContent>
          </complexType>
        </element>
```



UML klasser



- En UML-Klasse med stereotypen <<featureType>>
  - Geografisk object (spatial object type) med selvstændig identitet
  - Andre stereotyper:
    - <<dataType>>, <<codeList>>, og <<enumeration>>

Attributter

«featureType»
Boundary

+ country: CountryCode
+ geometry: GM_Curve
+ id: Identifier
«voidable»
+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime [0..1]
+ validFrom: DateTime
+ validTo: DateTime [0..1]
+ beginLifespanVersion: DateTime

+boundary
«voidable»
1..*

+admUnit
«voidable»
1..*

«featureType»
AdministrativeUnit

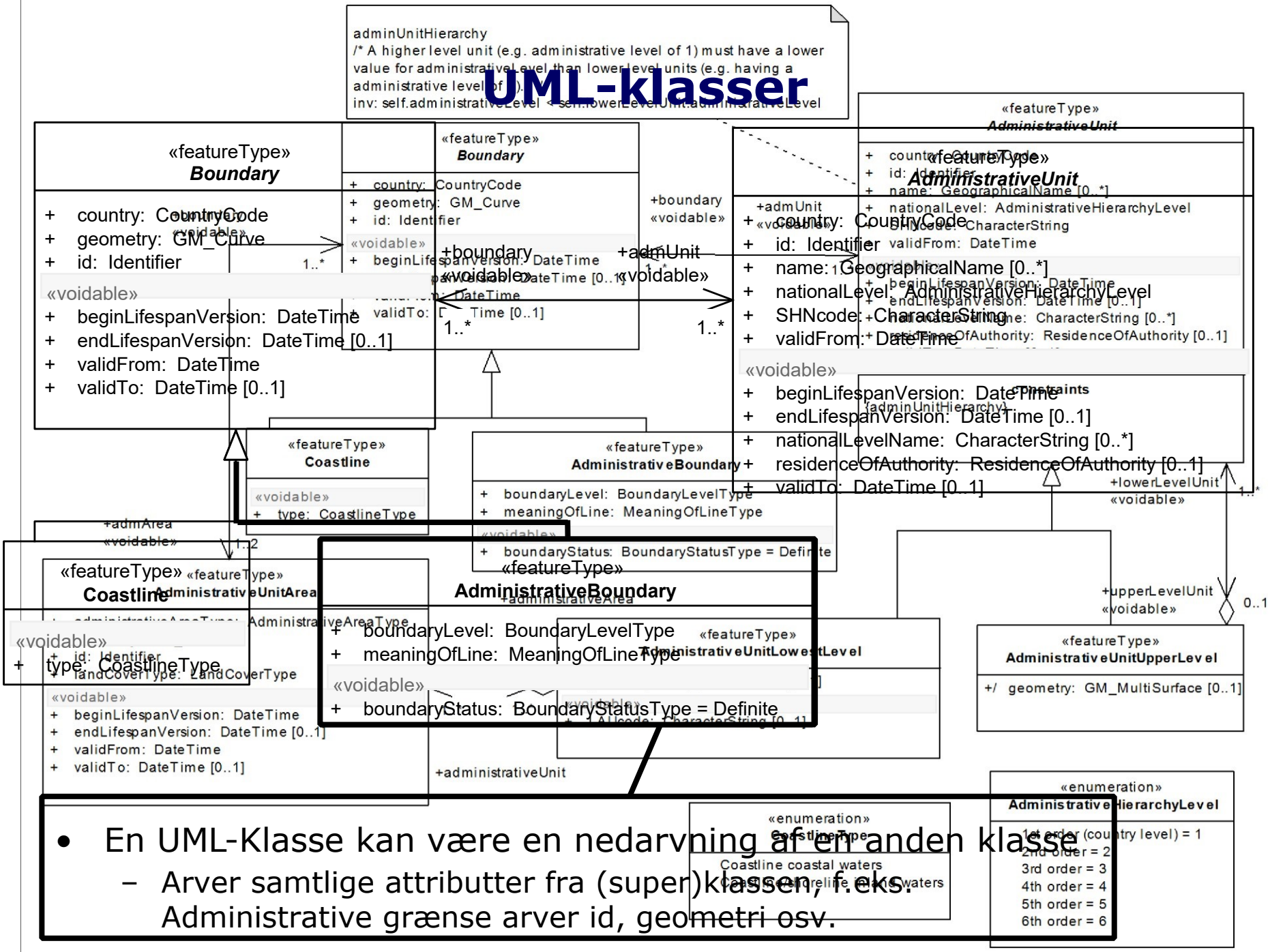
+ country: CountryCode
+ id: Identifier
+ name: GeographicalName [0..*]
+ nationalLevel: AdministrativeHierarchyLevel
+ SHNcode: CharacterString
+ validFrom: DateTime
«voidable»

- En attribut
- Beskriver egenskaber ved objekterne af en bestemt type

- Består af et navn ("country") og tilhørende datatype ("CountryCode")

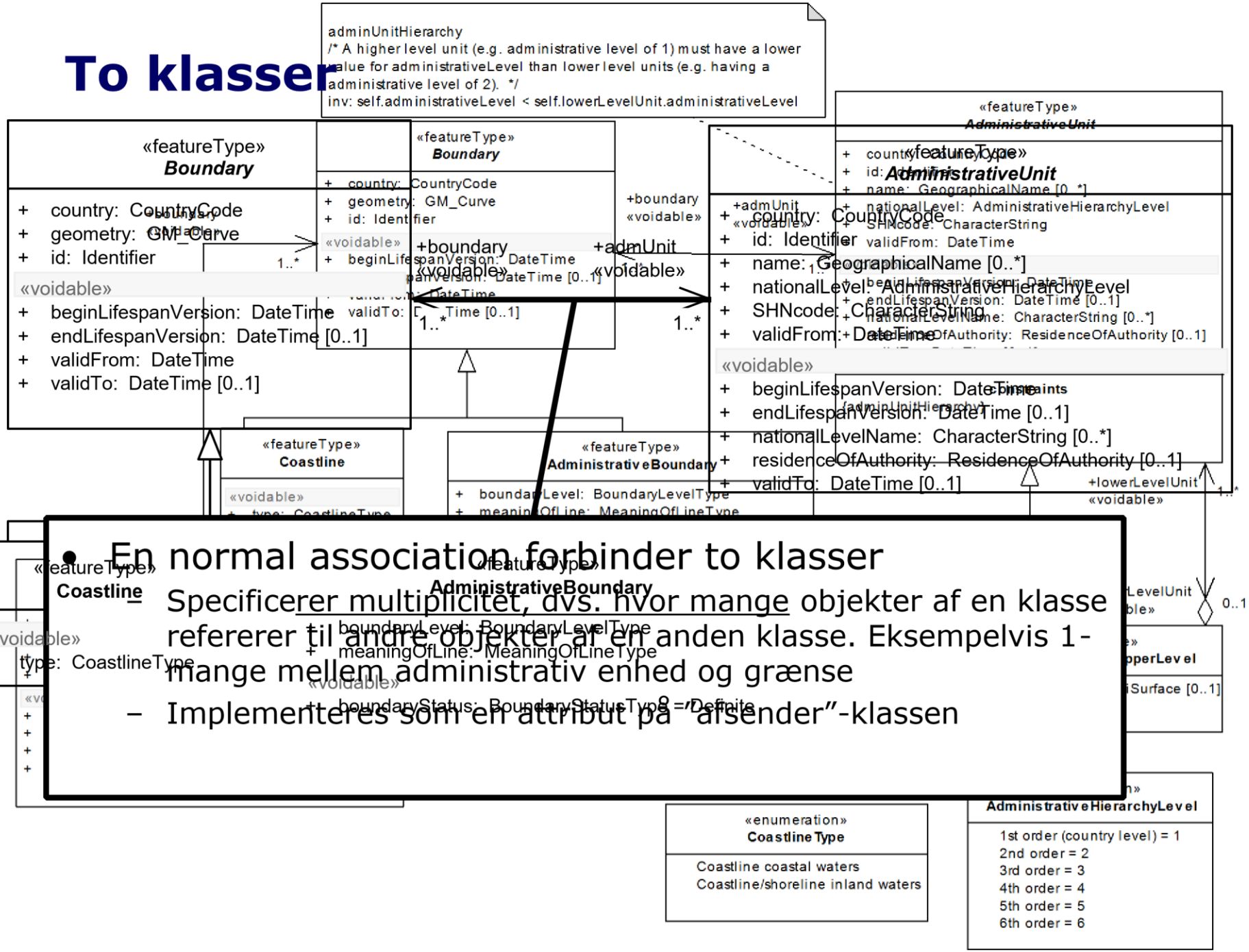
- Stereotyper:
  - <<voidable>>, <<lifeCycleInfo>>
    - <<voidable>>: Indikerer at der kan være situationer hvor ingen værdi findes i original data men findes i virkeligheden (eks. befolkningstæthed for et område).
      - Skal påføres NilReason value: "unpopulated" eller "unknown" hvis der ikke optræder en værdi
      - Er IKKE det samme som en "frivillig" attribut. Hvis dataansvarlig har data skal det leveres
    - <<lifeCycleInfo>>: Indikerer at attributten benyttes til at registrere database ændringer (tidsstempler fra-til) og kan benyttes til inkrementelle opdateringer af eksterne ressourcer

UML-klasser



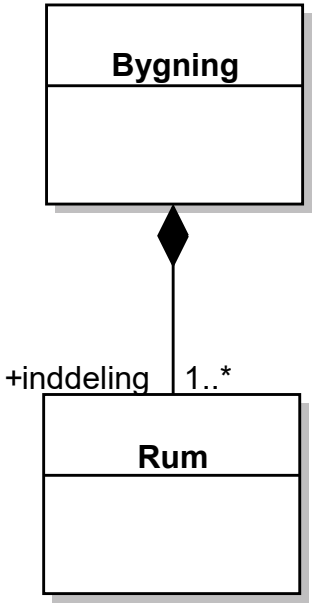
- En UML-Klasse kan være en nedarvning af en anden klasse
  - Arver samtlige attributter fra (super)klassen, f.eks. Administrative grænse arver id, geometri osv.

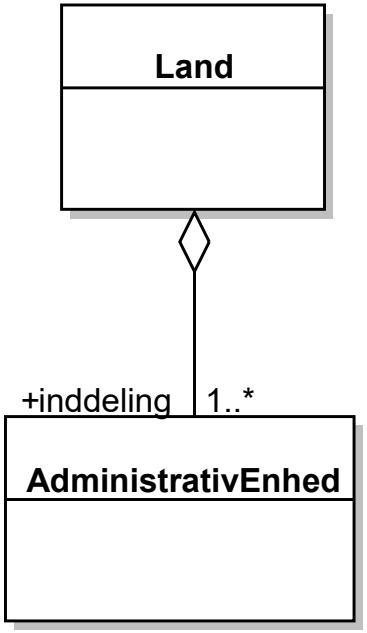
To klasser



- En normal association forbinder to klasser
  - Specificerer multiplicitet, dvs. hvor mange objekter af en klasse refererer til andre objekter af en anden klasse. Eksempelvis 1-mange mellem administrativ enhed og grænse
  - Implementeres som en attribut på "afsender"-klassen

Andre associationstyper





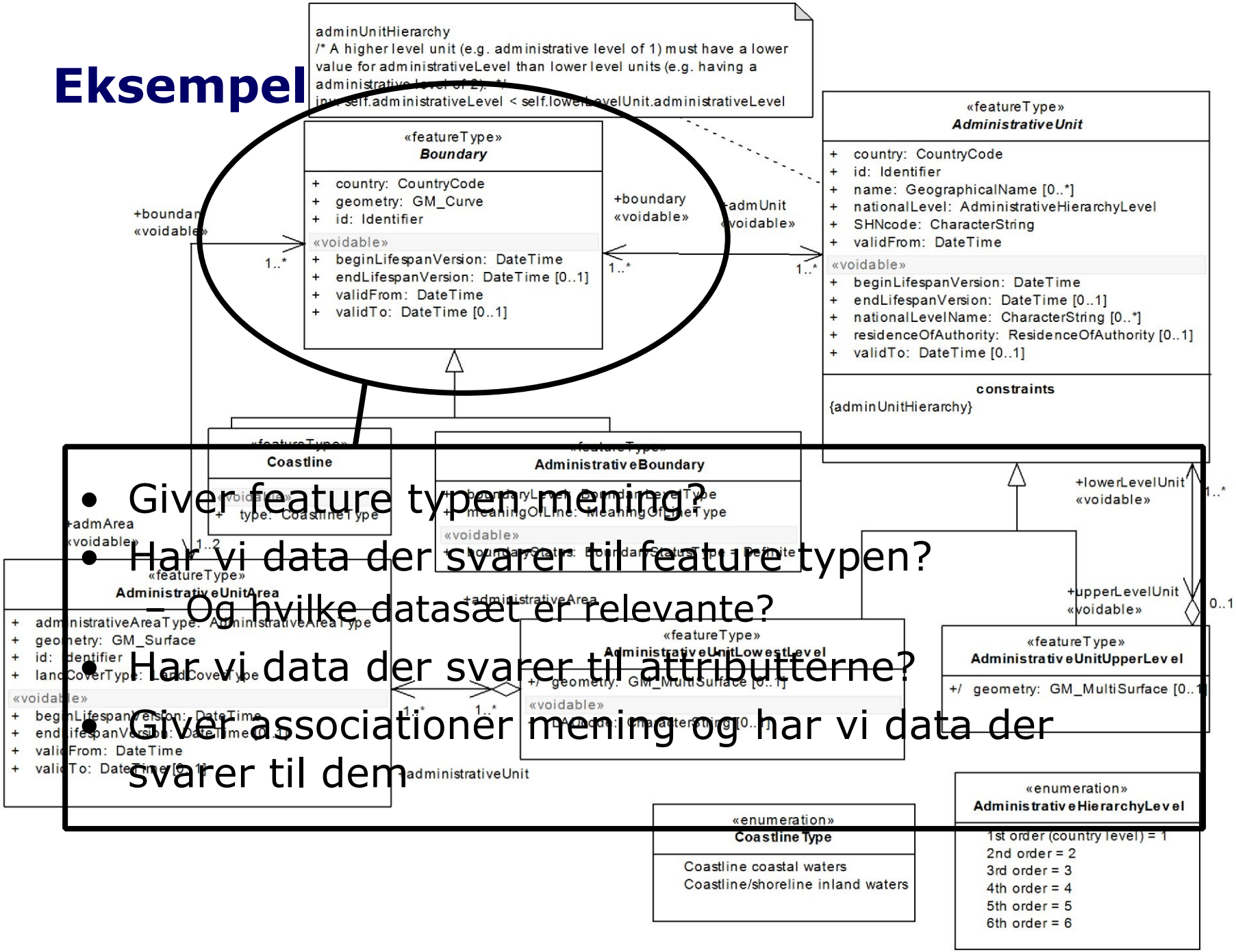
Komposition: Noget består af noget andet og stærk bindende (i vores model betragtes rum ikke som uafhængige objekter, kan ikke deles)

Aggregering: Noget består af noget andet men er ikke så stærk bindende (i vores model betragtes administrative enheder også som selvstændige objekter, kan deles med andre objekter)

Husk

Hvad skal jeg specielt kigge efter og afgive høringssvar på??

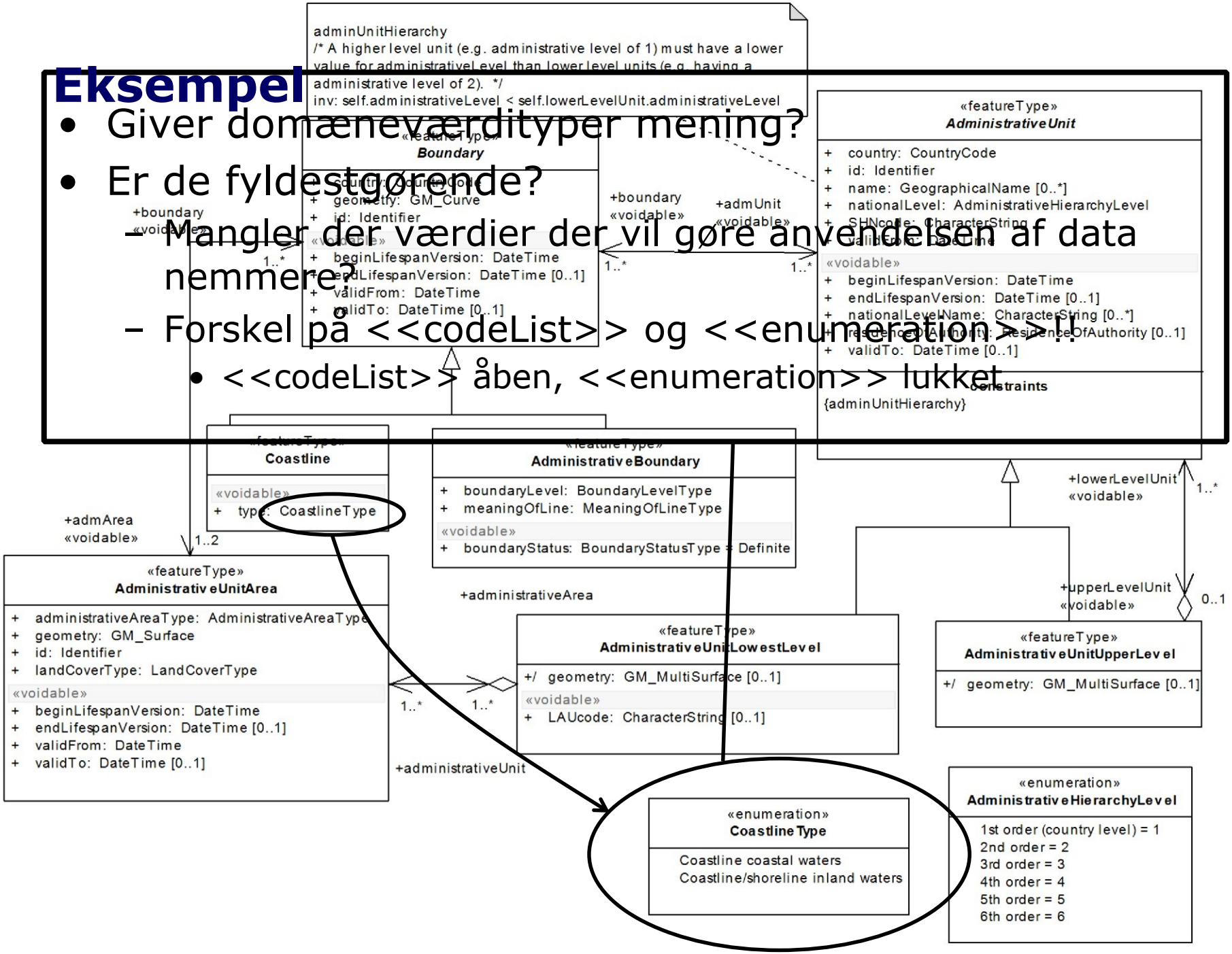
Eksempel



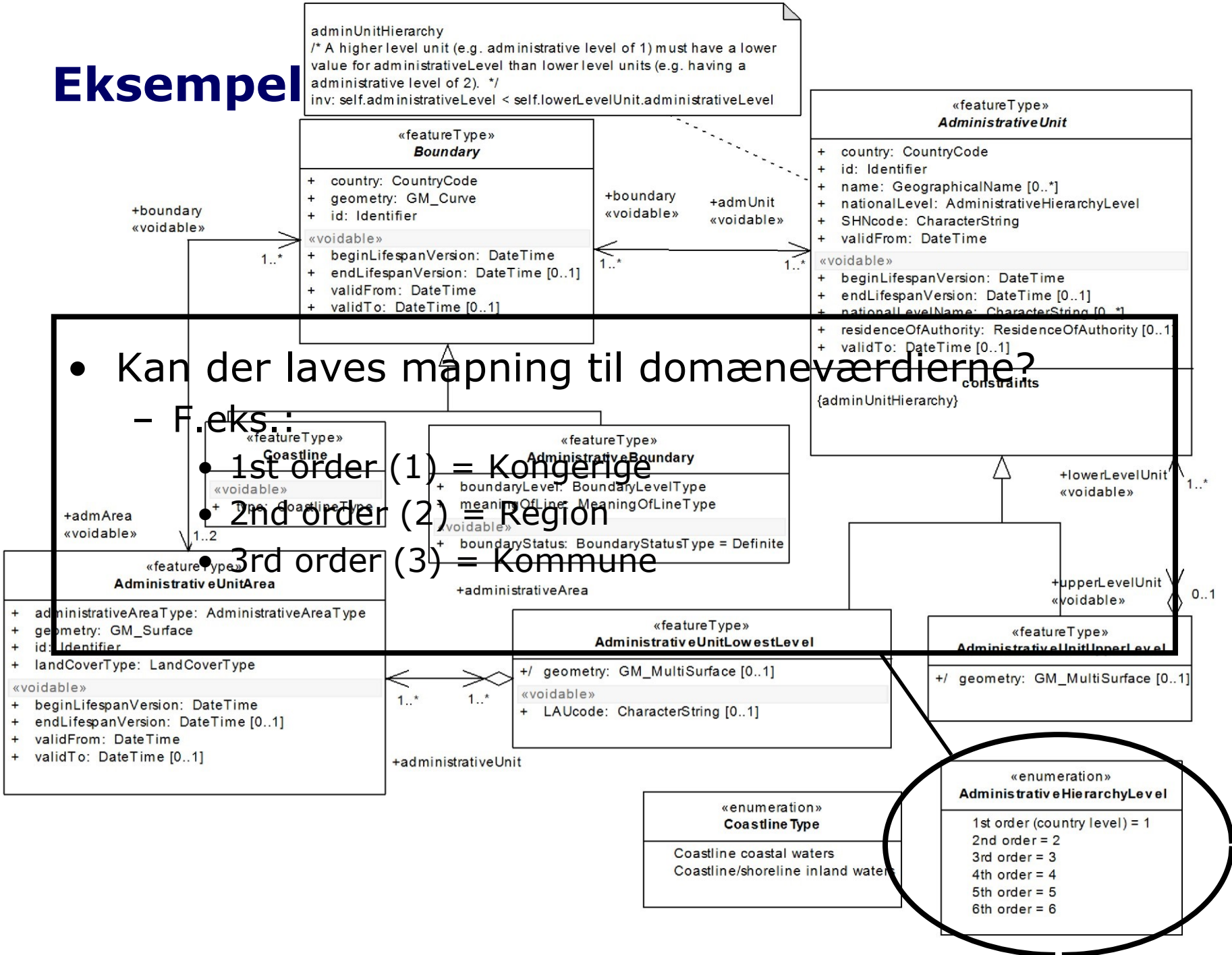
- Giver feature typen mening?
- Har vi data der svarer til feature typen?
  - Og hvilke datasæt er relevante?
- Har vi data der svarer til attributterne?
- Giver associationer mening og har vi data der svarer til dem

Eksempel

- Giver domæneværdityper mening?
- Er de fyldestgørende?
  - Mangler der værdier der vil gøre anvendelsen af data nemmere?
  - Forskel på <<codeList>> og <<enumeration>>!!
    - <<codeList>> åben, <<enumeration>> lukket

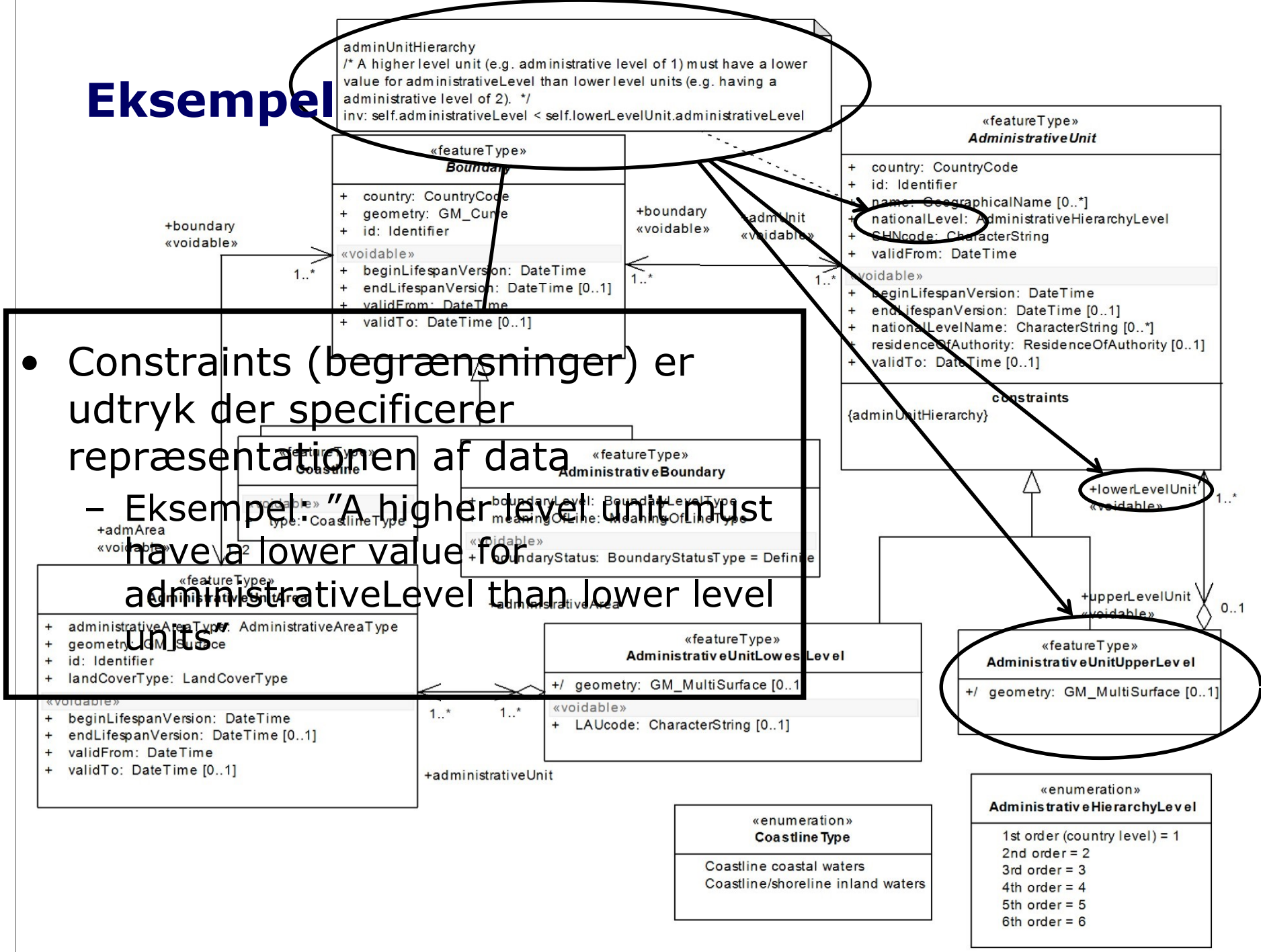


Eksempel



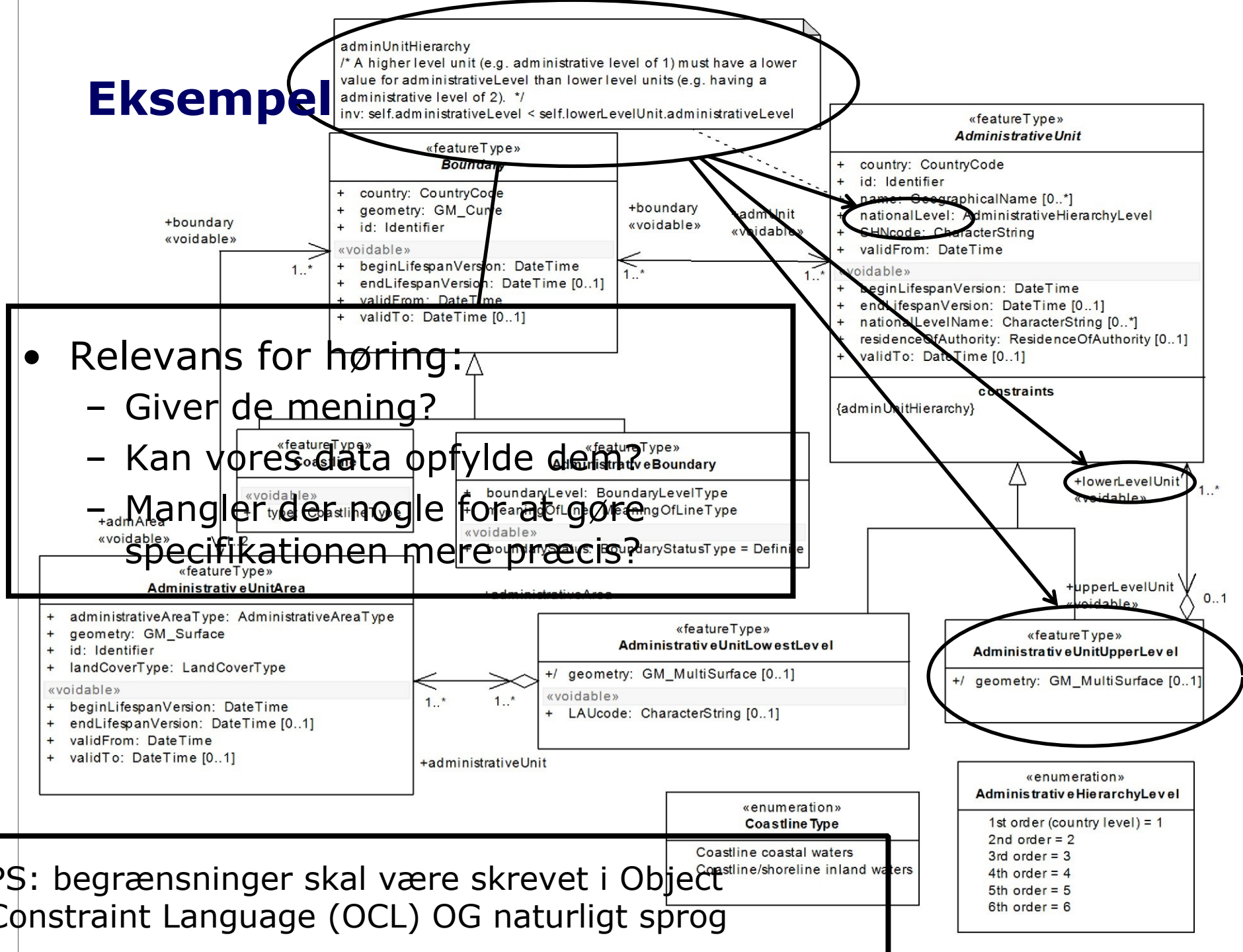
- Kan der laves mapning til domæneværdierne?
  - F.eks.:
    - 1st order (1) = Kongerige
    - 2nd order (2) = Region
    - 3rd order (3) = Kommune

Eksempel



- Constraints (begrænsninger) er udtryk der specificerer repræsentationen af data
  - Eksempel: "A higher level unit must have a lower value for administrativeLevel than lower level units"

Eksempel



- Relevans for høring:
  - Giver de mening?
  - Kan vores data opfylde dem?
  - Mangler der nogle for at gøre specifikationen mere præcis?

PS: begrænsninger skal være skrevet i Object Constraint Language (OCL) OG naturligt sprog

Feature katalog

- Indeholder (alle) elementerne fra UML modellen (genereret herfra)

SourceInformation	
Definition:	Contains metadata about specific instances of species distribution.
Description:	May be shared among several species distributions.
Status:	Proposed
Stereotypes:	«featureType»
URI:	null

Attribute: aggregationMethod

Value type: AggregationMethodValue Definition: Description of how observations are aggregated to the species distribution unit. Description: Values are provided by an extensible codelist. 0 to many single observations are usually aggregated within analytical units to provide the information about the distribution in a given area. These analytical units can be regularly shaped polygons like grids or irregular shaped polygons like administrative units, biogeographical regions or with boundaries of their own.

Multiplicity: 1..*

- Giver definitionen af begreber

Coverages…!

- Relevans for høring (udover det som er gennemgået for UML-modellen): – Er definitionerne fyldestgørende?

Coverages

Flere dataspecifikationer anvender ”coverages”. Hvad er det og hvordan skal de forstås?

- Koncept benyttet i bilag 2+3 (ikke i bilag 1) – Defineret i ”detaljer” i tilføjelserne til D2.5

Coverages…!

- Har IKKE noget med om data er repræsenteret som raster eller vektor
- Baseret på ISO 19123 og OGC's Web Coverage Service (WCS)
  - Inkonsistens mellem standarder
  - WCS er mangelfuld i implementeringen (og få implementeringer findes)
  - Inden for den næste årrække vil der ske en del på standardiserings- og implementeringsområdet
- Knytter sig tæt til "observationer"

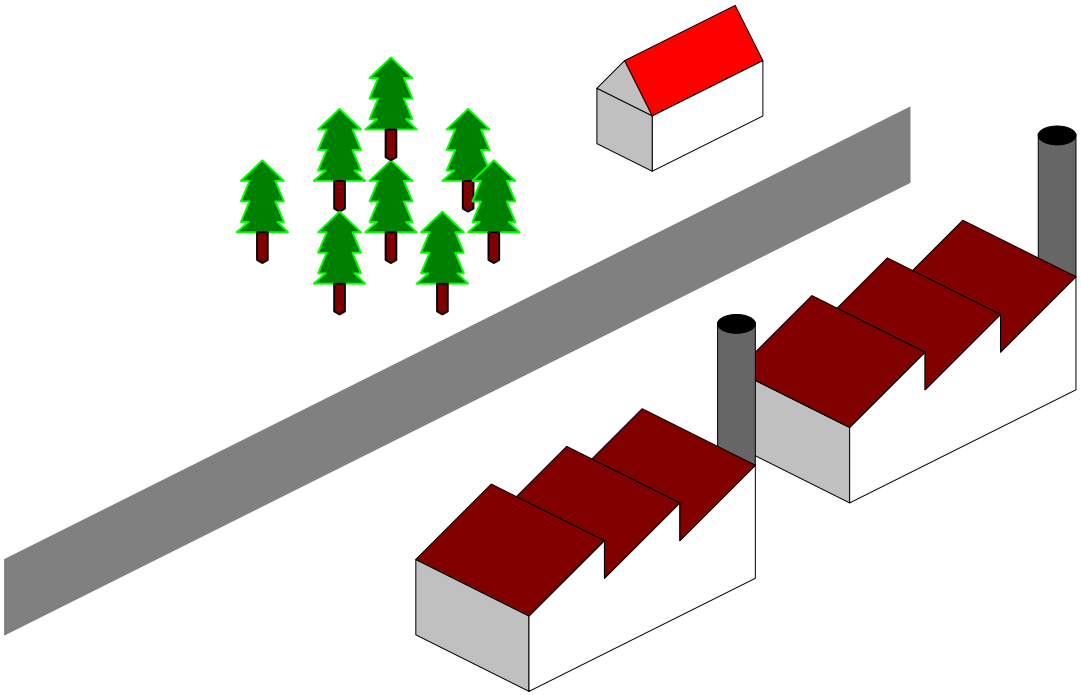
Temaer der anvender coverages

- Bilag 2:
  - Højde
  - Arealdække
  - Ortofoto
- Bilag 3:
  - Arealanvendelse
  - Områder med naturlige risici
  - Atmosfæriske forhold
  - Artsfordeling
  - Levesteder og biotoper

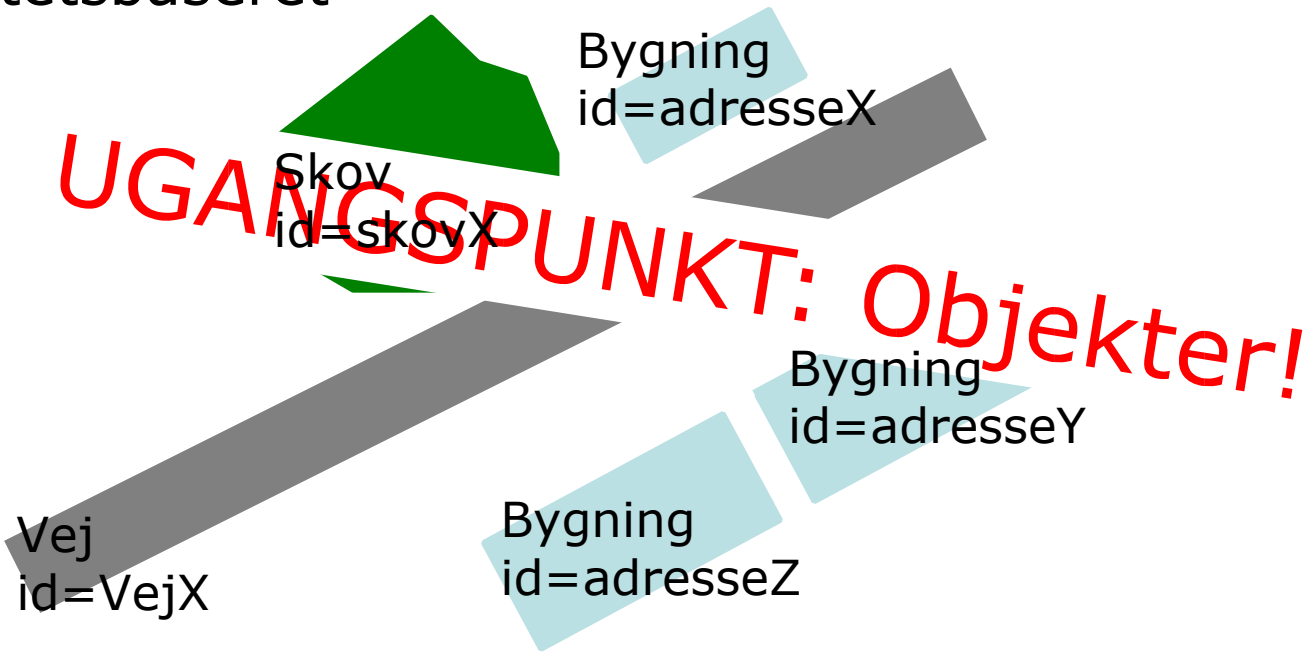
Coverages…!

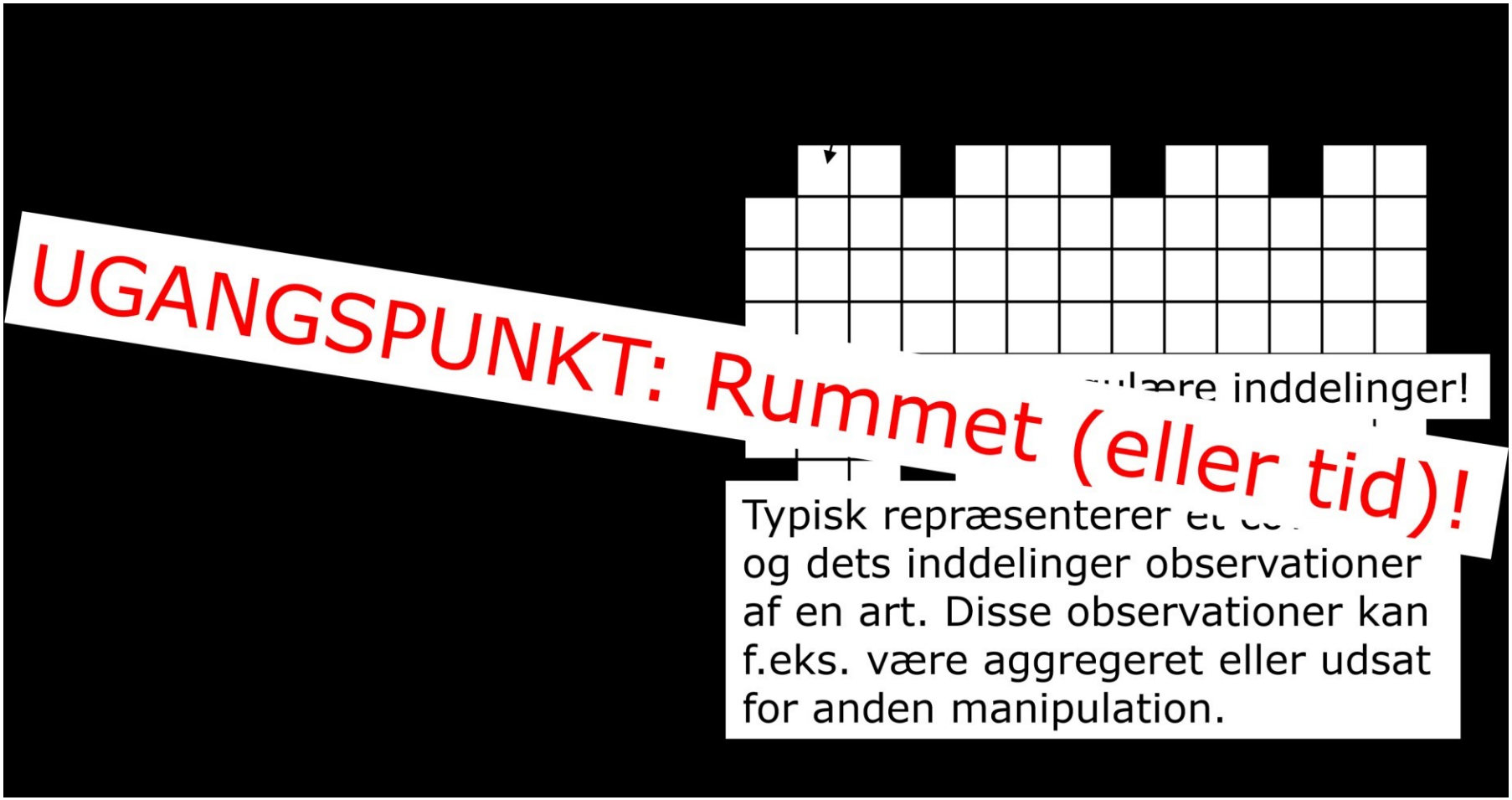
- Energiressourcer
- (Statistiske enheder / Sikkerhed og sundhed)

- To konceptuelt forskellige måder at repræsentere verdenen/data på
  - Entitetsbaseret (feature) og feltbaseret (coverage)
- Entitetsbaseret



- To konceptuelt forskellige måder at repræsentere verdenen/data på
  - Entitetsbaseret og feltbaseret
  - Entitetsbaseret




UGANGSPUNKT: Rummet (eller tid)!
inddelinger!
Typisk repræsenterer
og dets inddelinger observationer
af en art. Disse observationer kan
f.eks. være aggregeret eller udsat
for anden manipulation.

- Feltbaseret indeles i:
- Kontinuerte data
- En funktion, f.eks. interpolation, angiver værdi for hvert punkt i rummet (eller tid)
  - Eks: Højde
  - Diskrete data
- Værdi er konstant inden for hvert objekt i den givne tasselering af rummet
  - Eks: Arealdække

Modellering af

Miljøministeriet
Kort & Matrikelstyrelsen

domainsset

Coverage

«featureType»
Coverages (Domain and Range)::
CoverageByDomainAndRange

+ coverageFunction: CoverageFunction [0..1]
+ domainSet: Any
+ rangeSet: Any [0..*] {ordered}

{gridFunctio

domainSsset er ”geometri-attributten” af hver enkelt enhed. Den er specificeret via en constraint.

SpeciesDistributionCoverage

+ inspireId: Identifier
+ referenceSpeciesId: CharacterString
+ referenceSpeciesScheme: ReferenceSpeciesSchemeValue

::CoverageByDomainAndRange

+ coverageFunction: CoverageFunction [0..1]
+ domainSet: Any
+ rangeSet: Any [0..*] {ordered}

::Coverage

+ metadata: Any [0..*]
+ rangeType: RecordType

«voidable, lifeCycleInfo»

+ beginLifespanVersion: DateTime
+ endLifespanVersion: DateTime

«voidable»

+ domainExtent: EX_GeographicExtent

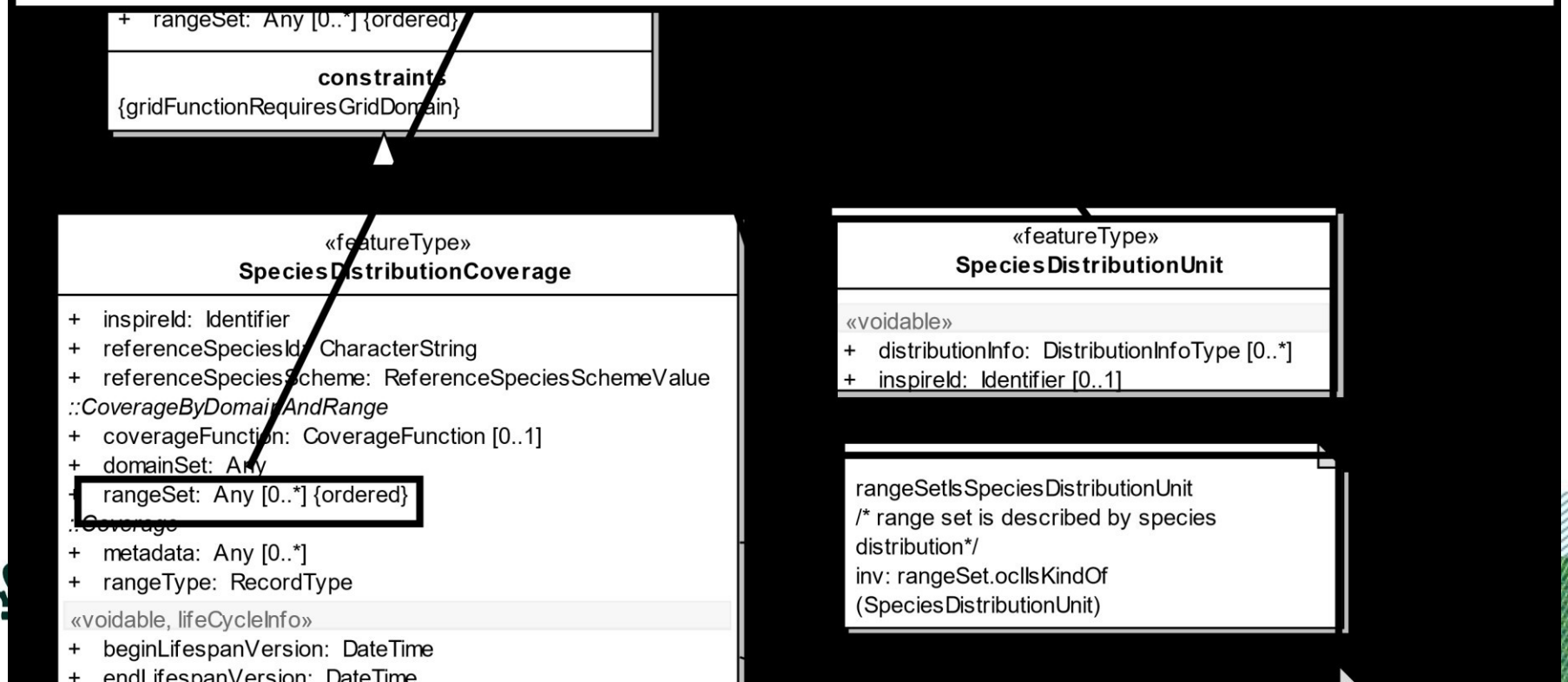
«voidable»

+ distributionInfo: DistributionInfoType [0..*]
+ inspireId: Identifier [0..1]

rangeSetIsSpeciesDistributionUnit
/* range set is described by species distribution*/
inv: rangeSet.oclIsKindOf (SpeciesDistributionUnit)

domainIsMultiSurfaceOrRectifiedGridOrMultiPoint
/* domain is a multi surface, rectified grid or multi point*/
inv: domainSet.oclIsKindOf(GM_MultiSurface) or

rangeSet er ”værdi-attributten” af hver enkelt enhed. Den er specificeret via en constraint. I dette tilfælde en spatial object type: SpeciesDistributionUnit



R angeset

(6) Reference systems

- Ens for alle temaer (def: koordinatsystemer dataspecifikation):
  - European Terrestrial Reference System 1989 (ETRS89)
  - European Vertical Reference System (EVRS)
- Projektioner (kun anbefaling):
  - Lambert Azimuthal Equal Area (ETRS-LAEA)
  - Lambert Conformal Conic (ETRS-LCC)
  - Transverse Mercator (ETRS-TMzn)
- Yderligere krav/anbefalinger kan defineres for enkelte temaer
- Temporalt: Gregorianske kalender i UTC eller Lokaltid + offset

- Relevans for høring (kan udelades):
  - Kan vi levere data i henhold til ovenstående?
  - Er der behov for yderligere definerede koordinatsystemer/projektioner for temaer?

(7) Data quality

- Definition af data kvalitets elementer der kan indgå i datasæt metadata for evaluering og brug. Eksempler:
  - DQ_CompletenessOmission
  - DQ_AbsoluteExternalPositionalAccuracy
- Definition af **minimum** kvalitetskrav til datasæt
- Baseret på ISO 19115 for metadata

- Relevans for høring:
  - Se om der er for mange eller for få elementer defineret – Hvis minimum kvalitetskrav defineret: Er det realistisk både økonomisk og indsamlingsmæssigt?

(8) Dataset-level Metadata

- Inkluderer metadata defineret af loven omkring metadata
- Samt yderligere krav til metadata for datasættet
  - Til brug for evaluering og brug
- Relevans for høring:
  - Obligatoriske elementer der er ”ligegyldige”?

- Elementer der mangler?
- Konsistens på tværs af specifikationer

(9) Delivery

- Medie:
  - ”Download” tjeneste
- Kodning
  - GML-skema (ver. 3.2.1)
- Inkluderet separat

- Der er temaer der har defineret alternative kodninger (f.eks. CSML)

- Relevans for høring:
  - Hvis overhovedet muligt en test af om data kan repræsenteres i det leverede GML skema – Input omkring alternativ kodning?

(10) Data Capture

- Indeholder krav og anbefalinger i forhold til indsamling og/eller publicering
- Meget divergerende indhold i de enkelte dataspecifikationer
- Relevans for høring:

- Må være op til de enkelte grupper at vurdere på hvilket niveau den skal gennemlæses

(11) Portrayal

- Indeholder specifikation af hvordan de enkelte feature typer skal visualiseres
  - Specificeret i OGC Symbology Encoding Standard
- Relevans for høring (kan udelades..):

- Hvis overhovedet muligt en test af om data kan repræsenteres vha. den leverede ”symbology encoding”
- Alternativt kig på den tekstlige beskrivelse og se om den giver mening

Annex A (normative)

- Indeholder specifikation af hvilke test der kan vise om data er i overensstemmelse med dataspecifikationen
- Afsnittet er ikke gennemarbejdet og færdigt.
- Relevans for høring (kan udelades):
  - Hvis det har interesse læs og kommentér evt på hvad der kan indgå i en test

Afrunding (1)

Afrudig(3)

Forslag til prioritering:

- Opsummering af vigtigste afsnit (afhænger dog af personlig/faglig interesse) – Executive summary Høj
- Relevant for overblikMellem
  - 2 OverviewLav
- Definitioner og overordnet beskrivelse
  - 5 Data content and structure • Feature typer, attributter, associationer, begrænsninger (constraints), og definitioner
  - 7 Datakvalitet

Afrunding (2)

- Elementer og minimums krav
  - 8 Datasæt metadata
- Inkluderede metadata elementer
  - (10 Data Capture)
    - Meget varierende indhold men kan inkludere krav som skal vurderes
  - Annex B (Use case)
  - Andre afsnit
  - Vigtigt:
  - Få det overordnede overblik
    - Er der datakrav som vi slet ikke kan opfylde?
    - Er der åbenlyse mangler?

Afrunding (3)

- Vurder de eksplicit markerede krav og anbefalinger • Giver ”requirements” mening (og kan de opfyldes)?
- Detaljer:
  - Giver definitioner mening?
  - Kan vi definere mapning til feature typer, domæneværdier, attributter osv?
- Mangler der f.eks. Domæneværdier – Til analysen kan bruges mapningstabeller

- Det er vores bedste chance for at give input til indhold og formulering af gennemførelsesbestemmelserne for bilag 2+3 data

Afrunding (4)

- Der er dog planlagt en kort høringsrunde/test ifb med udgivelsen af de endelige specifikationer (ver. 3)

INSPIRE · Infrastructure for Spatial Information in Europe

Miljøministeriet
Kort & Matrikelstyrelsen

Dataspecifikationer

– hvordan skal de forstås?

Anders Friis-Christensen (andfr@kms.dk)