From regulation to knowledge

By Veronika Heimsbakk

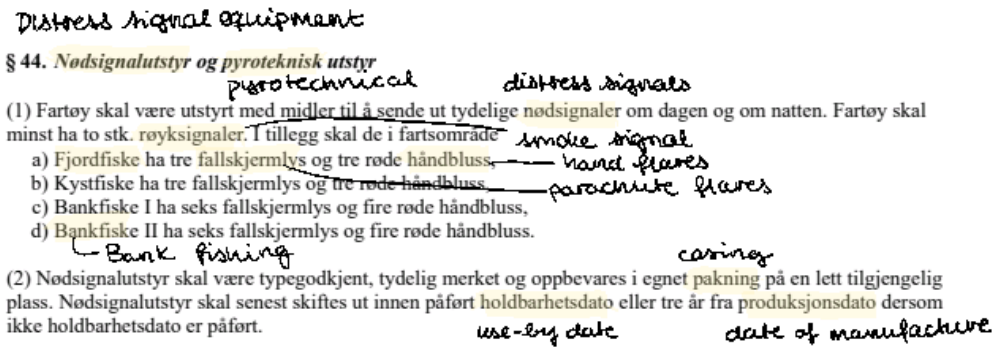
Senior consultant | Data Science & AI | Insights & Data, Capgemini Norway

The project

Sjøfartsdirektoratet
Norwegian Maritime Authority

- Making information connected and findable.
- Combining *legal scopes* and their position in regulation.
  - "My fishing vessel is 8.5 meters and built on January 3rd 1998."
- Identifying missing instance data.
  - Checking a sailor's CV against a set of requirements.

Challenges



Distress signal equipment

§ 44. *Nødsignalutstyr og pyroteknisk utstyr*

pyrotechnical distress signals

(1) Fartøy skal være utstyrt med midler til å sende ut tydelige nødsignaler om dagen og om natten. Fartøy skal minst ha to stk. røyksignaler. I tillegg skal de i fartsområde

smoke signal

a) Fjordfiske ha tre fallskjermlys og tre røde håndbluss, — hand flares

b) Kystfiske ha tre fallskjermlys og tre røde håndbluss, — parachute flares

c) Bankfiske I ha seks fallskjermlys og fire røde håndbluss,

d) Bankfiske II ha seks fallskjermlys og fire røde håndbluss.

Bank fishing

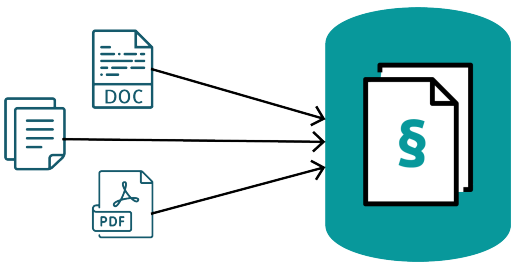
casing

(2) Nødsignalutstyr skal være typegodkjent, tydelig merket og oppbevares i egnet pakning på en lett tilgjengelig plass. Nødsignalutstyr skal senest skiftes ut innen påført holdbarhetsdato eller tre år fra produksjonsdato dersom ikke holdbarhetsdato er påført.

use-by date date of manufacture

- Information spread across various files on various formats.
- Manual graph modelling is *expensive*.
- Modelling data under a closed world assumption.

Working with files

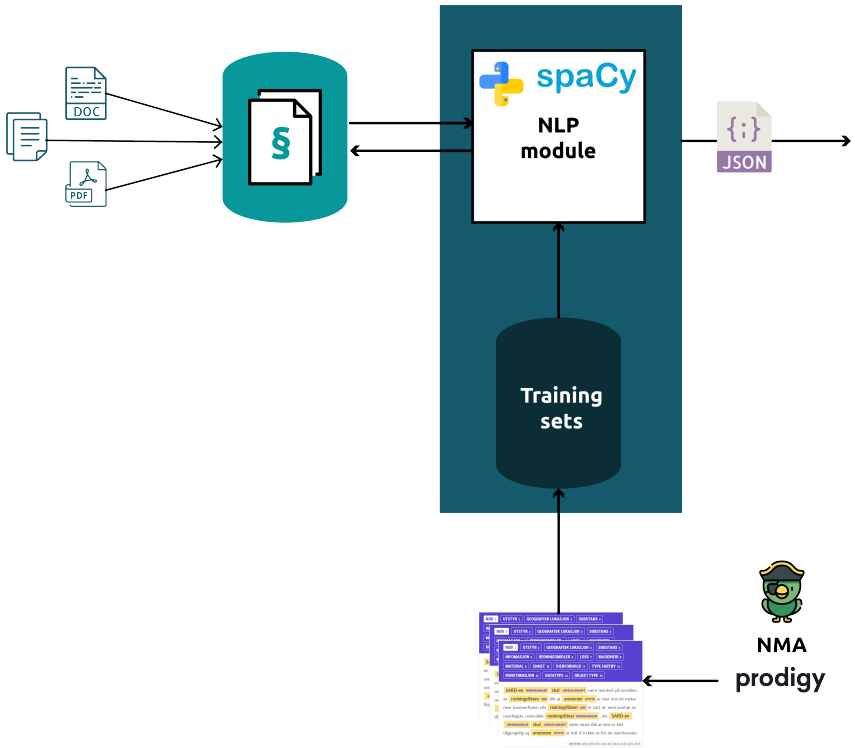


- Identify relevant sources of information.
- Gather them under a common format.
- Accessible through API.

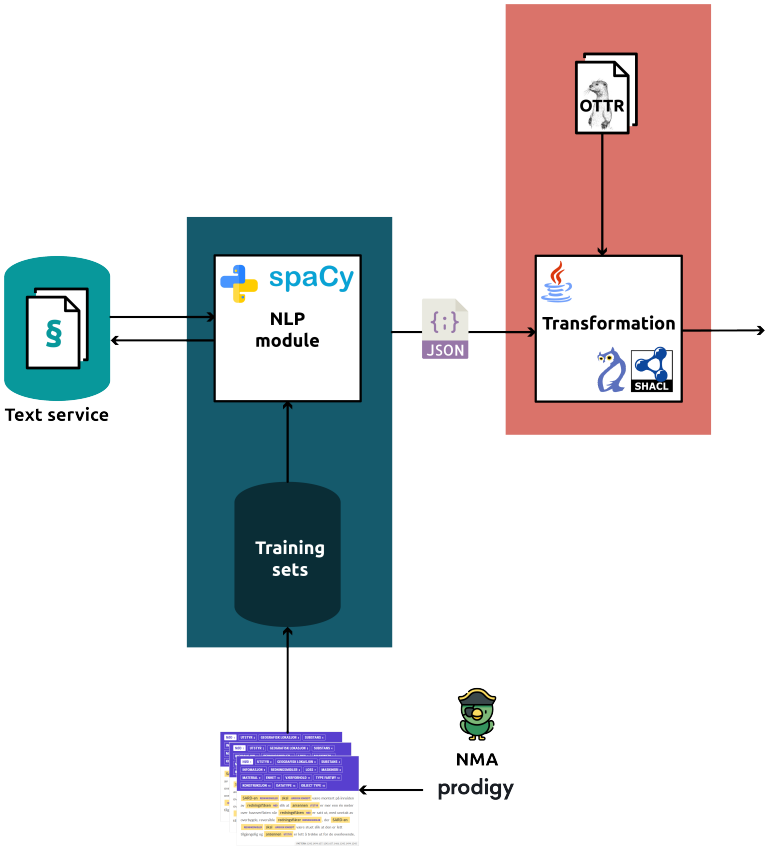
regulation\chapter\paragraph\part\sub-part

Extracting context, concepts and relationships

- Identify scopes
  - spaCy pattern matching rules
- Classify entities
  - annotations
  - NER models



Transforming to RDF



- Handle JSON input with Java and RDF4J.
- Generate OTTR instances.
- Serialize RDF using OTTR templates.
- SHACL shapes for modelling requirements.
- OWL Lite ontology

Reasonable Ontology Templates

Template[1]

```
o-sdir:Scope[! ottr:IRI ?shape, ! ?path] :: {
    o-sh:PropertyShape(?shape, ?path),
    o-rdf:Type(?shape, sdir:Scope)
} .
```

Instance

```
o-sdir:Scope(scope:FishingVessel, sdir:vesselType) .
```

Serialized RDF

```
scope:FishingVessel a sdir:Scope, sh:PropertyShape ;
  sh:path sdir:vesselType .
```

[1] SHACL templates (`o-sh`) isn't a part of the public template library (yet).

Shapes Constraint Language, requirements

Scope

```
scope:MaxLOA_15
  a sh:PropertyShape, sdir:Scope ;
  sh:path sdir:vesselLengthOverall ;
  sh:maxExclusive 15 ;
  sh:datatype unit:M ;
  sh:minCount 1 ;
  sh:maxCount 1 .
```

Requirement

```
:FOR1404P4
  a sh:NodeShape, :Requirement ;
  sh:property scope:MaxLOA_15 ;
  :regulationReference "https://lovdata.no/.../§4" ;
  :eliReference "/regulations/.../4" .
```

Shapes Constraint Language, certificates

Scope

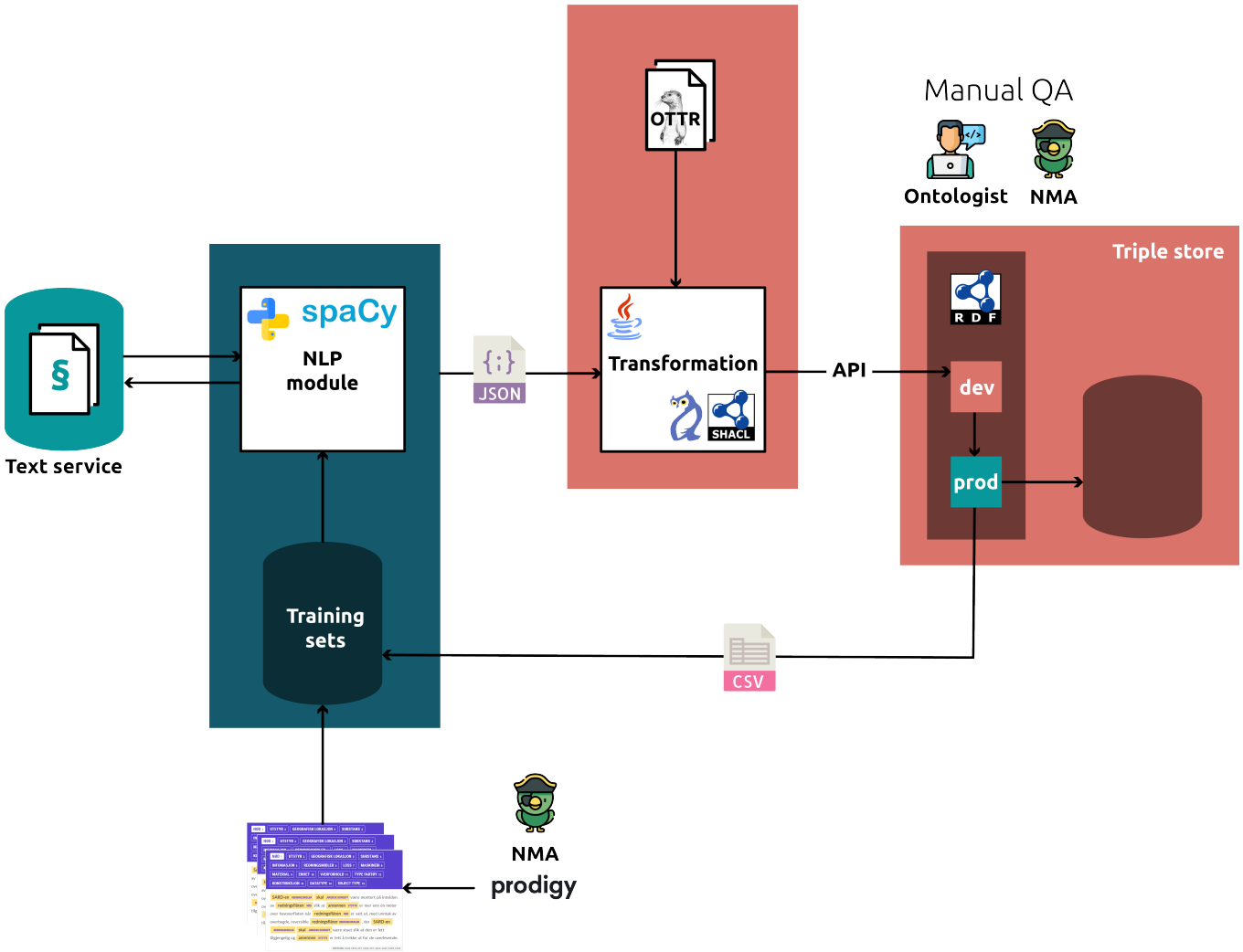
```
:D10 a sh:NodeShape ;
  sh:targetClass sdir:D10;
  sh:property :Age_minIncl_20, :VD2_Education ;
  sh:or (course:VSK course:VSKR course:OGD) ;
  sh:or (
    [ sh:and (
      [ sh:or (cert:PS_D2A0 ... cert:PS_D4F0) ]
      [ sh:path sdir:hasSeagoingServiceRequirement ;
        sh:hasValue sdir:SGS1 ; ]
    )]

  [ sh:and (
    [ sh:path sdir:hasSeagoingServiceRequirement ;
      sh:hasValue sdir:SGS2 ; ]

    [ sh:path sdir:hasSeagoingServiceRequirement ;
      sh:hasValue sdir:SGS3 ; ]

    [ sh:or (cert:PS_D2A0 ... cert:PS_D3B0) ]
  )]
 ) ;
...
```

Shapes Constraint Language, certificates

Seagoing service

```
sdir:SGS1
  a sh:NodeShape ;
  sh:property scope:Vessel_500_GT, scope:TradeArea_minInc_3,
              scope:Duration_360, scope:Position_DO .
```

Learning over and over and over again



- Improving NLP models and data integrity.
- Reduce time spent on manual QA over time.

Impact

- Introducing NLP saved 10 000 working hours.
- Learning loop will achieve higher data integrity and less time spent on manual QA over time.
- Using OTTR require a change in template, and not in code, if the model change in the future.
- SHACL allows us to identify missing pieces of information and model alternatives in regulation without using DL or OWL axioms.

veronika.heimsbakk@capgemini.com

vheimsbakk

veronikaheim

veleda