

Inbound links – picking the low hanging fruit from the semantic web

Christine Kirchberger, Staffan Malmgren

Swedish Law and Informatics Research Institute, Stockholm University

1 Introduction

Classical legal positivists saw law as a "closed logical system" that allows rules to be logically deduced from other rules¹. While this might not necessarily be the prevailing thought within legal theory today, legal sources do refer to each other in a consistent way and have done so for considerable time.

Irrespective of the jurisdiction or legal tradition, statutes refer to other statutes, court decisions refer to statutes and preparatory works, and legal doctrine refers to all of the above. This system of links and references is not always visible to the untrained eye, however. Legislation includes mainly links in one direction, for example from the exception to the general rule, but rarely the other way around. In order to increase the visibility of these connections the semantic web idea may offer some solutions.

For the purpose of this paper, we first assigned unique identifiers (URIs) not only to each section but also to each paragraph in a section within our legal database. In addition, we also stated the relationship between the different parts, for example that a paragraph is part of a section that is part of a chapter which is part of a law. In order to provide inbound links a query was formulated which retrieved all inbound links to a specific section.

All examples were taken either from the Swedish Personal Data Act (SFS 1998:204, Personuppgiftslag)² or the Swedish Penal Code (SFS 1962:700, Brottsbalk)³. This choice was a personal one and not based on any statistical reason.

This paper does not discuss the question of whether the state or government should play a more active role in utilizing solutions like ours. Some of our research is based, nevertheless, on the work of the Legal Information Project (Rättsinformationsprojekt)⁴ which was initiated in 2006. The question of mark-up and additional value of legal information has a legal and political dimension and cannot be answered in detail in this article. An in-depth analysis would, for example, look into the legal value of the mark-up technology and whether meta-information that is included into a published statute becomes part of the official and authentic version of this statute.⁵

We hope that there could be several advantages to utilizing references within legislation more effectively through the use of inbound links. Lawyers could be helped to find exceptions through the general rule (previously not easily done without having legal training). The detection of unwanted implications when amending the general rule could be facilitated. It could enable a popularity list of sections and articles in statutes most frequently referred to (the possible applications of such a ranking can be discussed).

As we will show in this article these goals can be quite easily achieved by using semantic web ideas

1 Raymond Wacks, *Understanding Jurisprudence – An Introduction to Legal Theory*, Oxford University Press, 2005, p 43.

2 In English available at <http://www.datainspektionen.se/in-english/legislation/the-personal-data-act/>

3 In English available at <http://www.sweden.gov.se/sb/d/3926/a/27777>

4 An initiative by the Swedish Administrative Development Agency (VERVA), <http://rinfo.lagrummet.se/>

5 This question has already been raised earlier, e.g. Cecilia Magnusson Sjöberg, *Rättsinformation, metadata och standarder för informationshandling*, in Perspektiv på rättsinformationen – Rättsinformation och IT 2002, IT-kommissionens rapport 2/2003, SOU 2003:58

and tools to interlink different sections in various laws. The potential for further improvement should not be underestimated and new solutions can be built upon ideas which are described in this paper.

2 The problem

Legal sources in general and legislation in particular should present the law clear and conclusive. In reality this is – unfortunately - not always the case. Some of the few tools to come closer to this goal include references between legal sources. By linking rules together, law appears more as a *system* and the differences or similarities between rules are easier to detect and connect.

Another important factor is the avoidance of redundancy. When stating the same principle or general rule in several different statutes problems of misspelling and interpretation might occur. Therefore the rule is only stated once within the *system* and other sections and laws can refer back to the same rule.

Also the rule of law is an important aspect in this respect. The rule of law states that all regulatory action must be based in law, which is why, for example, regulations refer to the statute they build on and court decisions state the legal rule that was applied. So one could argue that references are also an expression of the rule of law as they, in some cases, clarify the basis for an exception.

2.1 Linearity

Throughout legal history references have been stated by citing one rule in another. The question worth discussing is now why references are usually one-way, from one section to another section, e.g. from an exception to the general rule, but not from the general rule to the exception? In other words, why is it easier find the general rule when looking at the exception, but more difficult to find all exceptions to the general rule?

There are several explanations for this. One of the most significant being linearity. Linearity in its meaning for this paper entails that legal information is linear from a historical point of view which in turn has its roots in the way legal sources are being published.

In general, one can argue that legal *sources* are still associated with their physical form instead of the content. Already the term source implies that the bearer of the information rather than the information itself is the deciding factor. This comes - for the most part - from hundreds of years of printing of legal information. Since the printing press more and more laws were being published and therefore easier available to the public. In addition, partly due to democratisation the process of creating and publishing legal information became increasingly more organised and transparent.

Furthermore, traditionally, laws are not always being published in its completeness. Changes to a statute are only promulgated in the amended parts and sections. Merely in few cases, an official consolidated version is being published of a law depending on the amount of changes over the years. In most cases consolidation is taken care of by the commercial providers of legal information systems.

These factors, besides others, lead to legal sources being a snapshot of legal information at a certain point of time and place. In other words legal sources are static rather than dynamic. This also means that references and links can only be made back in time, not ahead in time.

Legal *information* on the other hand can be dynamic and inbound links can be one example for that.

2.2 Rules and references

Rules have different purposes, not only can they prohibit something (like stealing, murder, etc) they

can also stipulate responsibility under certain circumstances (e.g. product liability) or set up requirements for different processes (e.g. data protection). When it comes to references, they can express different relations between two rules. Though we will not cover these in our suggested solution, we think it is necessary to stress the different meanings of references.

One can imagine several scenarios (the following list is not exhaustive)⁶:

- a) Definition: Section B refers to Section A in the same (or another) statute that defines the term mentioned in the Section B.⁷
- b) Requirements: Section B states the legal punishment and refers to Section A in the same (or another) statute that stipulates the requirements in order for Section B to be applicable.
- c) Additional requirements: Section B refers to Section A in the same (or another) statute and adds some requirements to those stipulated in Section A
- d) Multiple requirements: Section B refers to Section A (or several other sections) that stipulates the basic requirements and also to Section C that contains an additional requirement.
- e) Non-application: Section B refers to Section A in the same (or another) statute: When the requirements of Section B are fulfilled, Section A is not applicable

These scenarios only cover references within and in between statutes and laws. In addition, as already mentioned, other legal sources regularly refer to each other. For example court decisions traditionally refer to the applicable rule, which builds of course on the principle of rule of law as well. This reference is already now quite extensively utilised by legal information service providers.⁸

3 The Swedish Legal System

The Swedish legal system is built on the traditional legal sources legislation, preparatory works, case-law, EU law and legal doctrine, as in many other European countries. Preparatory works have, however, a comparatively important value.

The Swedish Constitution does not consist of only one law, but of four fundamental statutes: the Instrument of Government (Regeringsformen, SFS 1974:152), the Act of Succession (Successionsordningen, SFS 1810:0926), the Freedom of the Press Act (Tryckfrihetsförordning, SFS 1949:105) and the Fundamental Law on Freedom of Expression (Yttrandefrihetsgrundlagen, SFS 1991:1469).

The four fundamental statutes represent the foundation of the legal system. On top of them statutes and regulations are built upon. Today there are about 3800 statutes and regulations in force in Sweden. In addition, more than 7000 preparatory works have passed the government, and the Supreme Court has decided in around 31 000 cases since 1948. This leads to a rather complex legal system with many cross-references and links.

3.1 Structure of the law and guidelines for legislation

All legal sources have a certain structure closely connected to the type of publishing that is required for a specific source. In Sweden when passing new legislation the guidelines from the Green Book⁹ have to be considered. The Green Book, inter alia, describes how headings within statutes should be

⁶ Examples of these meanings in a Swedish legal context are mentioned in Chapter 3.2.

⁷ A direct reference to the definition is often but not always stated, e.g. when the Penal Code refers to the term attempt no reference to a certain section is mentioned.

⁸ See for example <https://lagen.nu>

⁹ Gröna Boken, Statsrådsberedningen, Departementsserien (Ds) Ds 1998:66, December 1998, <http://www.regeringen.se/sb/d/108/a/22948>.

written, how sections should be structured and sorted, and how the possible legal punishment should be formulated.

Statutes and laws can be divided into

1. Sections or Articles (e.g. the Personal Data Act) or
2. Chapters and Sections or Articles (The articles are numbered either subsequently throughout the statute or with a new number in each chapter.)
3. Parts, Chapters and Sections (e.g. the Swedish Penal Code)

Not all areas of law are clearly regulated and part of a transparent system¹⁰, but references are being used widely within the Swedish legal system. All court decisions state the applicable rule and legal information providers include these links when delivering a specific statute to the user.

According to the Green Book a reference in a statute has to contain the exact name of the referred statute, in addition the year and number of the referred statute have to be stated in brackets. If a statute is mentioned more than once in the same Section, the year and number can be omitted. If the meaning is clear, the words “the law” or “the regulation” can also be used. These guidelines can nevertheless lead to a fairly complex references particularly when references are made to multiple sections.

The following section includes three different ways of referring to other rules:

Chapter 26 Section 8 Swedish Penal Code

If several sentences of imprisonment are being served concurrently, the combined terms of imprisonment shall be taken into consideration in applying Section 6. However, this does not apply to imprisonment imposed under Chapter 28, Section 3 or conversion imprisonment for non-payment of fines.

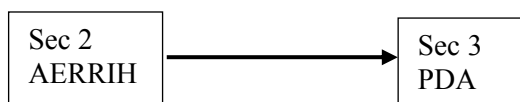
Time served shall also include the time during which the punishment is considered to be under enforcement by reason of an order of the court referred to in Chapter 33, Sections 5 or 6. (Law 1993:201)

This section shows one simple referral to another section in the same chapter (Section 6), one reference to a section in another chapter (Chapter 28 Section 3) and a pair of references to sections in another chapter (Chapter 33, Section 5 or 6)

3.2 Examples of references

Corresponding to the scenarios stated above, the following examples have been chosen from the Swedish legal system:

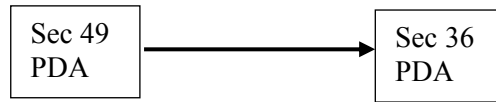
a) Definition: Section 2 of the Act concerning the Ethical Review of Research Involving Humans¹¹ refers to Section 3 of the Personal Data Act, which defines the term *processing of personal data*.



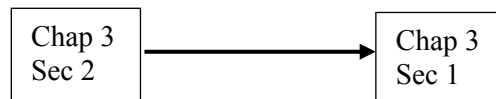
¹⁰ See e.g. Jan Hellner, The Law of Obligations and the Structure of Swedish Statute Law, in Peter Wahlgren, Scandinavian Studies in Law, Volume 40, Legal Theory, Stockholm, 2000, p 335.

¹¹ Section 2 Section 2 of the Act concerning the Ethical Review of Research Involving Humans (translation by the authors): In this statute the following terms mean
research: scientific research and development work on a scientific basis;
responsible research body: a government authority or a physical or legal entity under whose auspices the research is conducted;
research subject: a living person who is the subject of the research; and
processing of personal data: processing such as stated in Section 3 Personal Data Act.

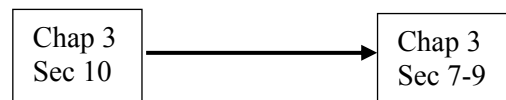
b) Requirements: Section 49 of the Personal Data Act¹² states the legal punishment for actions against (inter alia) Section 36¹³. In this case the penalty applies if the duties in Section 36 are not fulfilled. One could also imagine that the requirements of the referred Section have to be fulfilled.



c) Additional requirements: Chapter 3 Section 2 of the Swedish Penal Code¹⁴ stipulates additional requirements to Chapter 3 Section 1 in which case another punishment is to be decided.



d) multiple requirements: Chapter 3 Section 10 Penal Code¹⁵ refers to Chapter 3 Sections 7-9 and stipulates another legal consequence in case the additional requirements of Section 10 are fulfilled.



e) Non-application: Section 7 of the Personal Data Act¹⁶ refers to Section 9-29 that should not be

12 Section 49 Swedish Personal Data Act: A person who intentionally or by carelessness
 a) provides untrue information in such information to registered persons as is prescribed by this Act, or in the notification to the supervisory authority under Section 36 or to the supervisory authority when the authority requests information in accordance with Section 43,
 b) processes personal data in contravention of Sections 13–21,
 c) transfers personal data to a third country in contravention of Sections 33–35, or
 d) *omits to give notification under Section 36, first paragraph*, or in accordance with regulations issued under Section 41, shall be sentenced to a fine or imprisonment of at most six months or, if the offence is grave, to imprisonment of at most two years.
 A sentence shall not be imposed in petty cases.

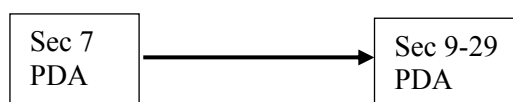
13 Section 36 Swedish Personal Data Act: Processing of personal data that is completely or partially automated is subject to a notification duty. The controller of personal data shall provide a written notification to the supervisory authority before such processing or a set of such processing with the same or similar purpose is conducted. If the controller of personal data appoints a personal data representative, this shall be notified to the supervisory authority. Removal from office of a personal data representative shall also be notified to the supervisory authority. The Government or the authority appointed by the Government may issue regulations concerning exemptions to the notification duty under the first paragraph for such kinds of processing as would probably not result in an improper intrusion of personal integrity.

14 Chapter 3 Section 2 Swedish Penal Code: If, in view of the circumstances that led to the act or for other reasons, the crime referred to in Section 1 is considered to be less serious, imprisonment for manslaughter shall be imposed for at least six and at most ten years.

15 Chapter 3 Section 10 Swedish Penal Code: Where a crime referred to in Sections 7 - 9 has been committed by a person with intent or by carelessly neglecting his duty under the Work Environment Act (1977:1160) to prevent sickness or accidents, the punishment shall be for an environmental offence and as provided for in the said provisions. (Law 1991:679)

16 Section 7 Swedish Personal Data Act: The provisions of this Act are not applied to the extent that they would contravene the provisions concerning the freedom of the press and freedom of expression contained in the Freedom of the Press Act or the Fundamental Law on Freedom of Expression.
The provisions of Sections 9–29 and 33–44 and also Section 45, first paragraph, and Sections 47–49 shall not be applied to such processing of personal data as occurs exclusively for journalistic purposes or artistic or literary

applied if the requirements in Section 7 are fulfilled. In this case the inbound link is rather important as its non-application is only regulated in the referred Section and not the general rule.



With its already more or less standardized way of citing and linking, the Swedish legal system provides a good platform for information management and retrieval as well as the utilization of ideas from the semantic web.

4 A semantic web implementation of the Swedish law

4.1 Formats for data and metadata

The semantic web is about *metadata*. It allows for expressing *relations* between resources such as documents or parts of thereof, using a graph structure for modeling the information. It is however not very useful for expressing the documents themselves. A typical legal document, such as a law or a court verdict, is instead structured in a hierarchical way, with the text containing chapters, each chapter containing sections, each section containing paragraphs, and so on. Each part that can contain text can also contain *mixed content*, textual data interspersed with instructions, such as "this element should be linked to a remote resource".¹⁷ This is better expressed with a XML based format, using a document tree structure to express the content of the document.

Since the actual requirements of the legal texts are not very specific, there is no strong need to design a XML document format from scratch. Any existing general document format, such as ODF,¹⁸ TEI,¹⁹ DocBook²⁰ or XHTML²¹ can conceivably be used. Selecting one over the other is mainly done on the basis of tool support, and the ability of the format to *embed* the metadata that we will be using.

We use XHTML, as this format has widespread tool support and is immediately viewable in most any modern browser. It also supports RDFa for embedding any RDF metadata directly into the document tree.²²

Documents and the parts that make up the document are both modelled as resources. In the RDF model, this means that they have identity, and can be identified by a URI. A very simple example is below:

```
<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml"
      version="XHTML+RDFa 1.0"
      xmlns:dct="http://purl.org/dc/terms/"
      xmlns:ex="http://example.org/">
  <head>
    <title property="dct:title">An example document with part</title>
    <base href="http://example.org/sample/doc"/>
```

expression.

17 Extensible Markup Language (XML) 1.0 (Fourth Edition), Section 3.2.2, "Mixed Content", <http://www.w3.org/TR/REC-xml/#sec-mixed-content>

18 Open Document Format, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

19 Text Encoding Initiative, <http://www.tei-c.org/>

20 <http://www.docbook.org/specs/>

21 XHTML™ 1.1 - Module-based XHTML - Second Edition, <http://www.w3.org/TR/xhtml11/>

22 RDFa in XHTML: Syntax and Processing, <http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>

```

</head>
<body typeof="ex:Document">
  <div about="http://example.org/sample/doc/part"
        typeof="ex:Part">
    Hello world
  </div>
</body>
</html>

```

This document uses two resources, `<http://example.org/sample/doc>` and `<http://example.org/sample/doc/part>` (using the base element and the about attribute, respectively), and describes their type (using the `typeof` attribute). For the first resource, its title is also expressed, using the content attribute to declare the type of expression (the RDF predicate).

These RDF statements can be expressed in the Turtle syntax:²³

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix : <http://www.w3.org/1999/xhtml> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix ex: <http://example.org/> .

<http://example.org/sample/doc>
  dct:title "An example document with part" ;
  a ex:Document .

<http://example.org/sample/doc/part>
  a ex:Part .

```

4.2 Relationships between containing and contained parts

Note that we so far have no relationship between `<http://example.org/sample/doc>` and `<http://example.org/sample/doc/part>`. We cannot rely on the fact that the URIs for the resources indicate that the latter is part of, or contained in, the former, as URIs in the RDF model are opaque. If we wish to convey this information, we need to explicitly state this. We must also choose a suitable way of expressing the is-part-of relationship between the two resources. As we are already using the Dublin Core terms vocabulary (for expressing the document title), it is suitable to use the `isPartOf` term for expressing this.

```

<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml"
      version="XHTML+RDFa 1.0"
      xmlns:dct="http://purl.org/dc/terms/"
      xmlns:ex="http://example.org/">
  <head>
    <title property="dct:title">An example document with part</title>
    <base href="http://example.org/sample/doc"/>
  </head>
  <body typeof="ex:Document" rev="dct:isPartOf">
    <div about="http://example.org/sample/doc/part"
          typeof="ex:Part">
      Hello world
    </div>
  </body>
</html>

```

Through the semantics of the `rev` attribute that RDFa specifies, this yields the following triples:

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

```

²³ Terse RDF Triple language, <http://www.w3.org/2007/11/21-turtle>

```

@prefix : <http://www.w3.org/1999/xhtml> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix ex: <http://example.org/> .

<http://example.org/sample/doc>
  dct:title "An example document with part" ;
  a ex:Document .

<http://example.org/sample/doc/part>
  dct:isPartOf <http://example.org/sample/doc> ;
  a ex:Part .

```

Now the relationships between a document and its parts are explicitly expressed.

In order to create a useful dataset for doing semantic web, we need to treat every part of a document that someone might want to refer to as a separate resource, or we won't be able to explicitly express these references. For Swedish law, this means that we need to assign URIs for the law itself, all chapters, all sections, all paragraphs, and all items in numbered and unnumbered lists.

In some rare cases, references are made to a particular sentence or a word in a sentence. We have chosen to not represent these locations in the law as separate resources, on account that this would create an explosion in resources, leading to computational complexities for little practical gain.

As an example, consider the Swedish Penal Code (SFS 1962:700, *Brottsbalk*) and its jurisdiction rules. We use a simple URI naming strategy for the text and its different parts.

The law itself: <http://lagen.nu/publ/sfs/1962:700>

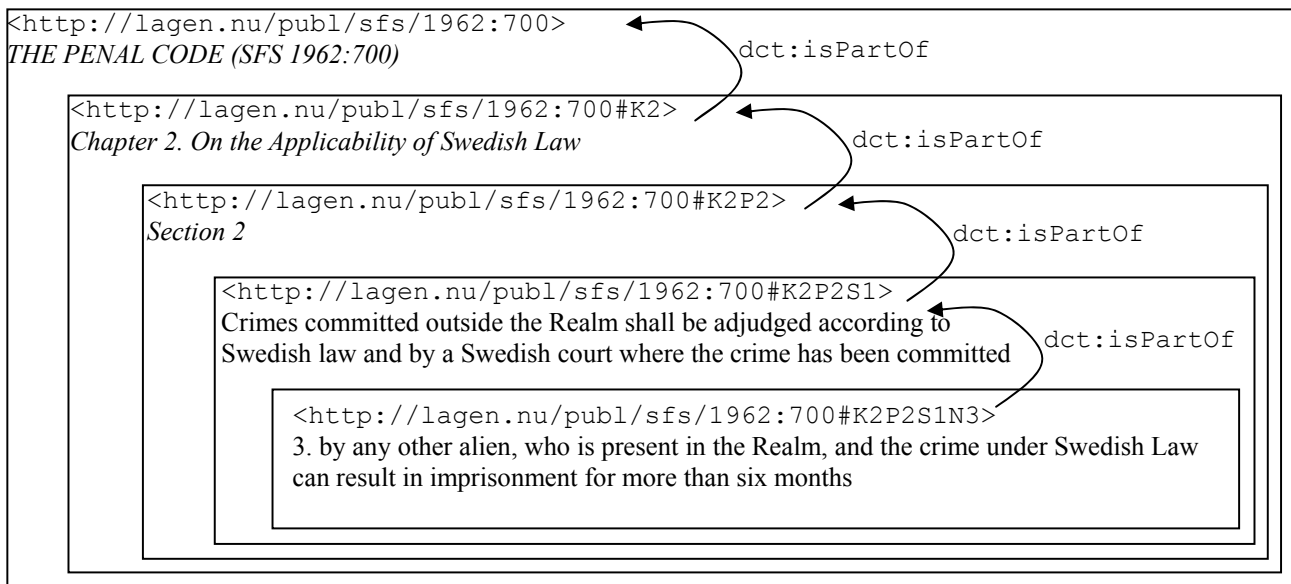
The second chapter (concerning the applicability of Swedish law):
<http://lagen.nu/publ/sfs/1962:700#K2>

The second section of the second chapter, <http://lagen.nu/publ/sfs/1962:700#K2P2>

The first paragraph of the second section of the second chapter:
<http://lagen.nu/publ/sfs/1962:700#K2P2S1>

The third item in the item list of the first paragraph of the second section of the second chapter:
<http://lagen.nu/publ/sfs/1962:700#K2P2S1N3>

As noted above, the fact that URIs are opaque in the RDF model means that we cannot automatically assume that <http://lagen.nu/publ/sfs/1962:700#K2P2S1N3> is part of <http://lagen.nu/publ/sfs/1962:700>. We therefore explicitly state this, using a series of `dct:isPartOf` expressions:



Or, expressed as triples

```
@prefix dct: <http://purl.org/dc/terms/> .

<http://lagen.nu/publ/sfs/1962:700#K2P2S1N3>
  dct:isPartOf <http://lagen.nu/publ/sfs/1962:700#K2P2S1> .

<http://lagen.nu/publ/sfs/1962:700#K2P2S1>
  dct:isPartOf <http://lagen.nu/publ/sfs/1962:700#K2P2> .

<http://lagen.nu/publ/sfs/1962:700#K2P2>
  dct:isPartOf <http://lagen.nu/publ/sfs/1962:700#K2> .

<http://lagen.nu/publ/sfs/1962:700#K2>
  dct:isPartOf <http://lagen.nu/publ/sfs/1962:700> .
```

4.3 Vocabulary usage

The quality and re-usability of any semantic web system depends in part on the usage of established and appropriate vocabularies. In the described system, three sets of vocabularies are used.

DCMI Metadata Terms, an extension of the original Dublin Core Metadata Element Set, used with the prefix `dct`. This vocabulary provides terms for basic document properties such as creator, publisher, title and description, and links between documents.

Elektronisk standard för rättsinformation (ESFR, "Electronic standard for legal information"), a RDF vocabulary designed for the Legal Information Project, used with the prefix `rinfo`. This vocabulary provides terms and classes specific to the structure of Swedish legal information.

Extended ESFR, a extension of the ESFR vocabulary, providing additional fine-grained classes, used with the prefix `rinfoex`.

4.4 Inline references

As previously mentioned, Swedish laws frequently contain references to other parts of the law. Since we provide a URI for every single fragment of the body of law, these references can be

expressed as standard XHTML hyperlinks.

As an example, the reference described in section 3.2 a):

```
processing of personal data: processing such as stated in Section 3 Personal Data Act.
```

can be rewritten to make the reference explicit as:

```
processing of personal data: processing such as stated in Section 3 Personal Data Act.
```

which yields the following triple:

```
<http://lagen.nu/publ/sfs/2003:460#P2S5>  
dct:references <http://lagen.nu/publ/sfs/1998:204#P3>.
```

The process of automatically analyzing law text to find and resolve such inline references falls outside of the scope of this paper.²⁴

4.5 External datasets

Now we have discussed references within a particular legal source, namely laws. Other legal sources may refer to laws as well - most importantly court decisions.

We can extract similar datasets from these sources. As an example, we have extracted an additional 25 000 references to different parts of the Swedish law from a 25-year archive of Swedish court decisions (totalling around 10 000 decisions).

Other valuable datasets that could be extracted with automated processing include preparatory works, which refer to parts of the law by proposing to introduce or change it, and legal journal articles. From an EC law perspective, an interesting data set would be a mapping between individual articles in EC directives and their corresponding implementation in Swedish law.

5 Displaying inbound links using standard semantic web tools

5.1 Storing the data

By expressing all parts of the law, down to individual paragraphs and list items, as resources using RDF, we get a data set of approximately 500 000 addressable resources. By expressing references from one part of the law to another, we get a data set of approximately 140 000 references. Since these references have been expressed as RDF statements within the actual documents, using RDFa, they can easily be extracted into a RDF database.²⁵ With all the statements in the database, it becomes trivial to find references to a particular part of the law (*inbound links*).

5.2 Querying the data

To present a section of text alongside with references to that text, we simply can query the database for such references. This can be expressed as a simple SPARQL query:²⁶

²⁴ We use a similar method to the one described in “Automated Detection of Reference Structures in Law”, De Maat, E., Winkles, R., and Van Engers, T., in *Legal Knowledge and Information Systems, JURIX 2006*, but with a grammar tailored to Swedish law reference syntax, as described in section 3.1

²⁵ In our implementation, we have been using Sesame 2, <http://www.openrdf.org/>

²⁶ SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>

```

SELECT DISTINCT ?source WHERE
{
  ?source dct:references <http://lagen.nu/publ/sfs/1962:700#K2P2>
}

```

This will retrieve all sources that refer to the second section of the second chapter of the penal code. However, it will not retrieve sources that refer to a more specific part, such as the third item of the first paragraph of said section, since our dataset has no built-in notion of containedness. To retrieve such sources, we need to expand our query:

```

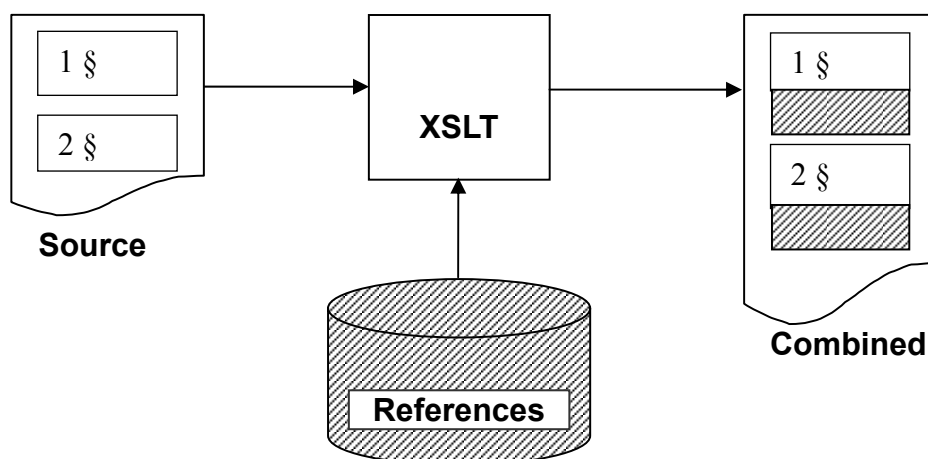
SELECT DISTINCT ?source WHERE
{
  {
    ?source dct:references <http://lagen.nu/publ/sfs/1962:700#K2P2>
  }
  UNION
  {
    ?source dct:references ?target .
    ?target dct:isPartOf <http://lagen.nu/publ/sfs/1962:700#K2P2>
  }
  UNION
  {
    ?source dct:references ?target .
    ?target dct:isPartOf ?container .
    ?container dct:isPartOf <http://lagen.nu/publ/sfs/1962:700#K2P2>
  }
}
ORDER BY ?source

```

This will, in addition to the above, retrieve all sources that refer to any part *contained in* said part of the penal code.

5.3 Combining the data

In order to present the law combined with inbound links to each section, we transform the original individual XHTML documents into new XHTML documents.



A very simple way of doing this combining is through XSLT 1.0 and its `document()` function,²⁷ coupled with the HTTP API that Sesame 2 provides. We define a rule for all section elements (all that have attribute `typeof="rinfo:Paragraf"`), which dynamically queries a Sesame server using a HTTP request, in order to find references to that section.

²⁷ XSL Transformations (XSLT), section 12.1, "Multiple source document", <http://www.w3.org/TR/xslt#document>

```

<xsl:template match="*[@typeof='rinfo:Paragraf']">
  <xsl:copy>
    <xsl:apply-templates select="@*|node()" />
    <xsl:variable name="query">
      [see example SPARQL query above]
    </xsl:variable>
    <xsl:variable name="references-url">
      http://localhost/openrdf-sesame/repositories/lagen.nu?query=<xsl:value-
of select="$query"/>
    </xsl:variable>
    <xsl:variable name="results" select="document(str:encode-uri($references-
url,false()))/sparql:sparql/sparql:results/sparql:result"/>
    <xsl:if test="$results">
      <div class="inbound">
        Inbound links:
        <ul>
          <xsl:for-each select="$results">
            <li>
              <xsl:value-of select="substring-after(sparql:binding/sparql:uri,
'/publ/sfs/')"/>
            </li>
          </xsl:for-each>
        </ul>
      </div>
    </xsl:if>
  </xsl:copy>
</xsl:template>

```

For all other elements, we fall back on standard XSLT identity transform (copying the source tree to the target tree without any changes).²⁸

```

<xsl:template match="@*|node()">
  <xsl:copy>
    <xsl:apply-templates select="@*|node()" />
  </xsl:copy>
</xsl:template>

```

5.4 Results:

The described stylesheet with its embedded SPARQL query returns result like the following (translated excerpt):

Section 3

In this Act the following terms are used with the meaning stated below.

[...]

Third party	A person other than the registered person, the controller of personal data, the personal data representative, the personal data assistant and such persons who under the direct responsibility of the controller of personal data or the personal data assistant is authorised to process personal data.
-------------	--

Inbound links:

- | |
|--|
| <ul style="list-style-type: none"> • 1998:527#P2S1 • 2003:460#P2S1 |
|--|

²⁸ XSL Transformations (XSLT), section 7.5, “Copying”, <http://www.w3.org/TR/xslt#copying>

6 Room for improvement

One obvious improvement is to bring in inbound links from other data sets. Being able to see a list of court cases that refers to a particular section directly underneath that section might be even more useful than the example we've shown.

Providing a more human-readable version of the inbound links (e.g. “Section 2 of the Act concerning the Ethical Review of Research Involving Humans” rather than the terse “2003:460#P2S5”) would improve usability.

A more complex improvement would be to attach more meaning to our inbound links. As described in 3.2, references can have different meanings. These differences are not expressed in our dataset. If we *reify* our triples – that is, we treat them as new resources, for which we can make new statements – we can augment them with more meaning.²⁹ In the example used in 3.2 a), which we expressed in RDF in 4.4, one part of the law referred to another for purposes of defining a term. We can express this by the following triples:

```
<http://lagen.nu/publ/sfs/references/fdd7cccb74367008349e34a8dd08af72>
  rdf:type rdf:Statement ,
  rdf:subject <http://lagen.nu/publ/sfs/2003:460#P2S5> ,
  rdf:predicate dct:references ,
  rdf:object <http://lagen.nu/publ/sfs/1998:204#P3> ,
  rfinox:refmeaning <http://lagen.nu/taxonomy#usesDefinitionIn>
  dct:creator "Staffan Malmgren" .
```

These new statements are not part of the original dataset, but instead form a new dataset that *annotate* the original data. For this reason, it's also valuable to know who made the statement, so that its veracity can be assessed. We use the `dct:creator` to express this.

The predicate `<http://lagen.nu/taxonomy#usesDefinitionIn>`, together with the other examples of references in 3.2, forms the start of a taxonomy of legal reference types. Finishing such a taxonomy would be an area for further interdisciplinary research.

7 Conclusions

This example of a real-world use case may seem trivial - given a large scale semantic web dataset, one could imagine many much more interesting applications. We are not using inference, formal ontologies, or other more advanced semantic web technology.

Nevertheless, we believe that this demonstration shows that XML and basic semantic web tools can provide much value in the legal information space with extremely simple means - the actual code in our example is less than 50 lines. Such "low-hanging fruit" applications are made possible through the use of standardized formats and data models.

The example also shows the importance of the data quality of the *raw material*, i.e. the legal sources. The more standardised and structured the original legal data and the citations are, the easier references and other semantic structures can be utilised.

29 “RDF Semantics”, section 3.3, <http://www.w3.org/TR/rdf-mt/#ReifAndCont>