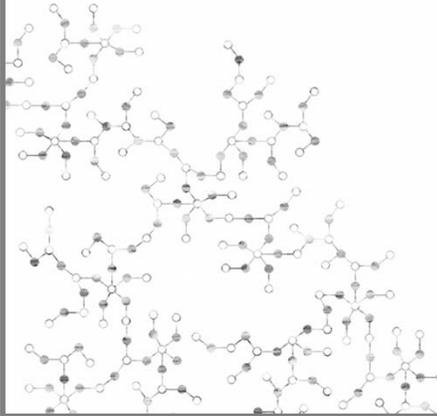


Linked Data and Canadian Legal Resources

F. Tim Knight
Osgoode Hall Law School Library
York University

CanLII Law, Government and Open Data
Conference and Hackathon
September 13th-14th, 2013 in Ottawa, Ontario



First of all I'd like to thank CanLII for organizing this event, and Sarah Sutherland in particular, for inviting me to speak today ...

I'm here to talk about linked data and Canadian legal resources on the semantic web.

Outline:

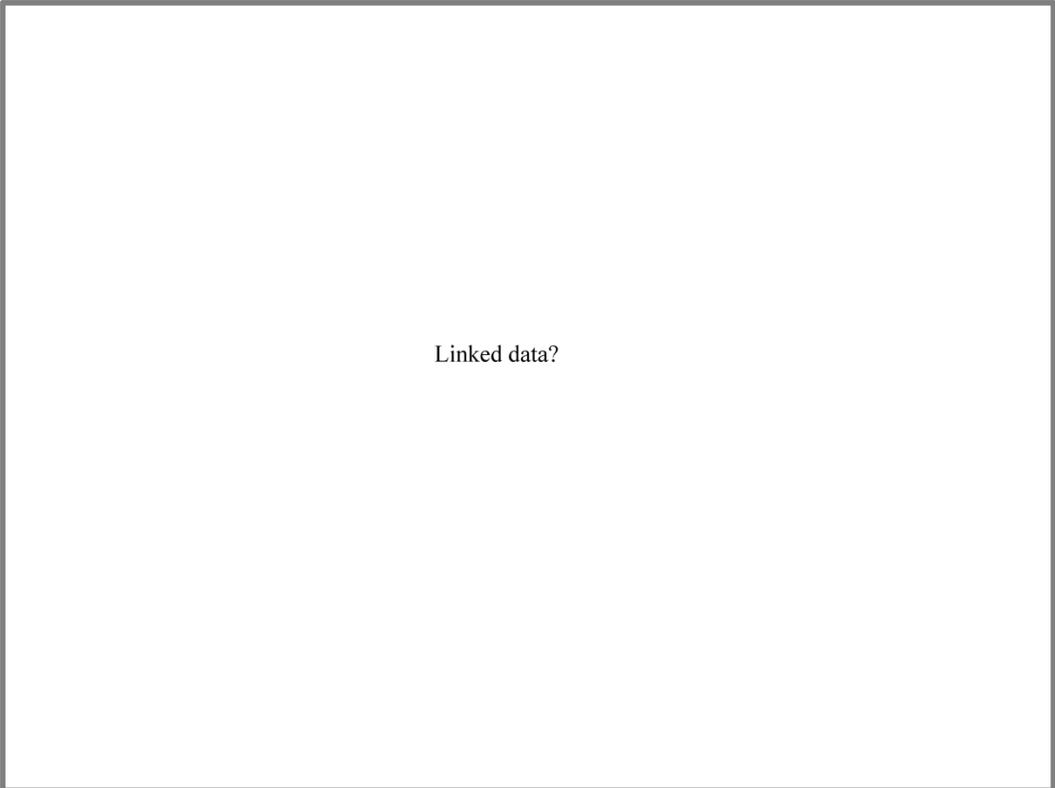
Review of linked data theory and principles

Growth of the linked data cloud

Linked data and legal resources

Role for organizations like CANLII

I'll begin with a quick review of the basic theory and building blocks that support linked data; then I'll talk about the growth of the so-called linked open data cloud; then we'll consider how this idea might be applied to legal resources and take a look at some of the challenges involved; and finally I'll touch on what organizations like CanLII can do to support these developments.



Linked data?

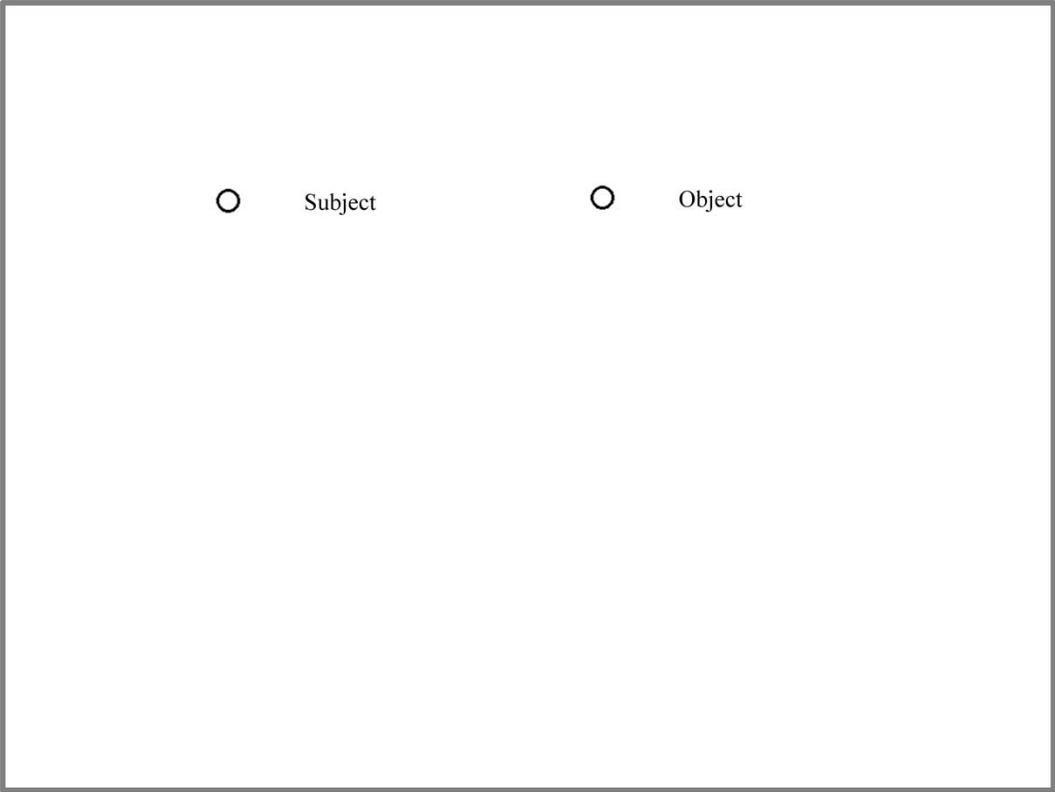
On the surface the basic theory of linked data is actually a pretty simple one.

It's the practical and technological side of the equation that still holds many challenges and continues to make the implementation of linked data a rather complex exercise.

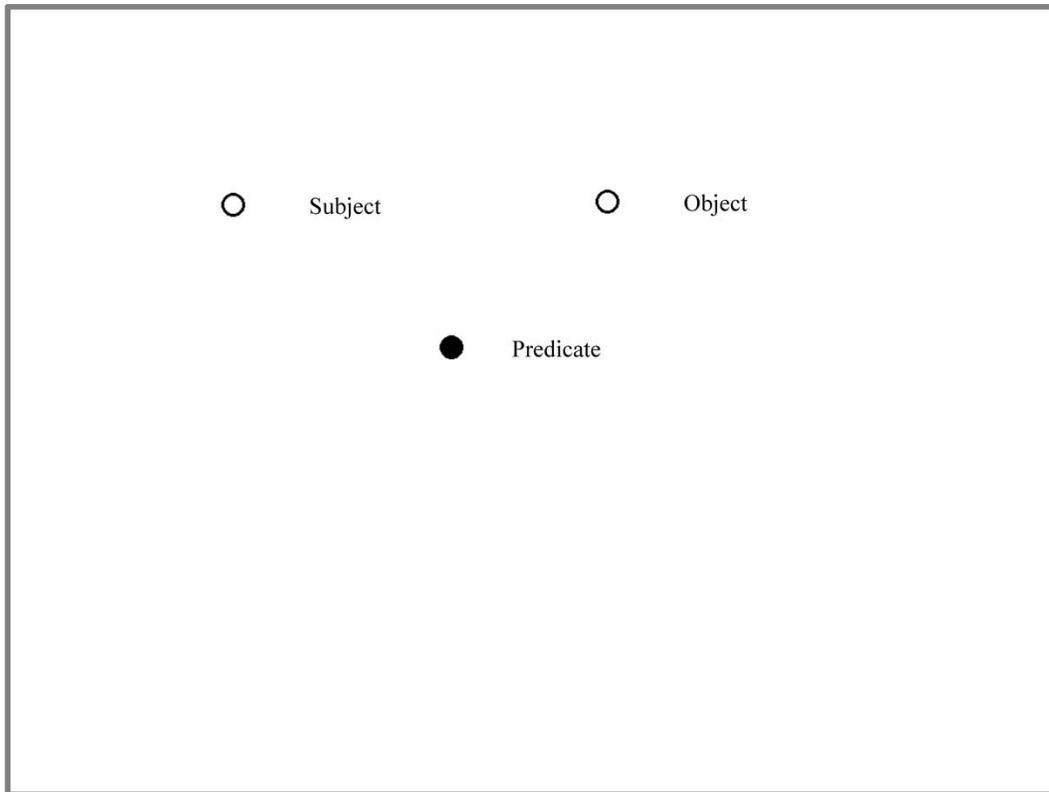
So for now I'll stick with the simple stuff ... :-)



So, first of all you've got a Subject ...

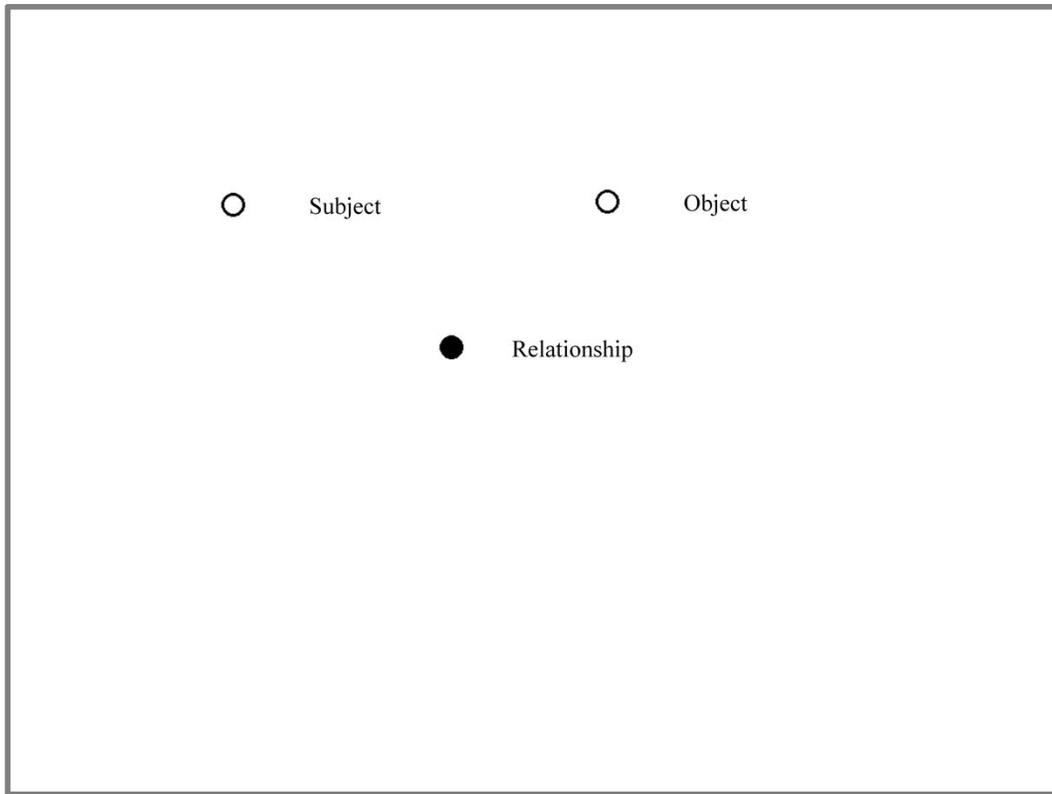


... and, an Object ...

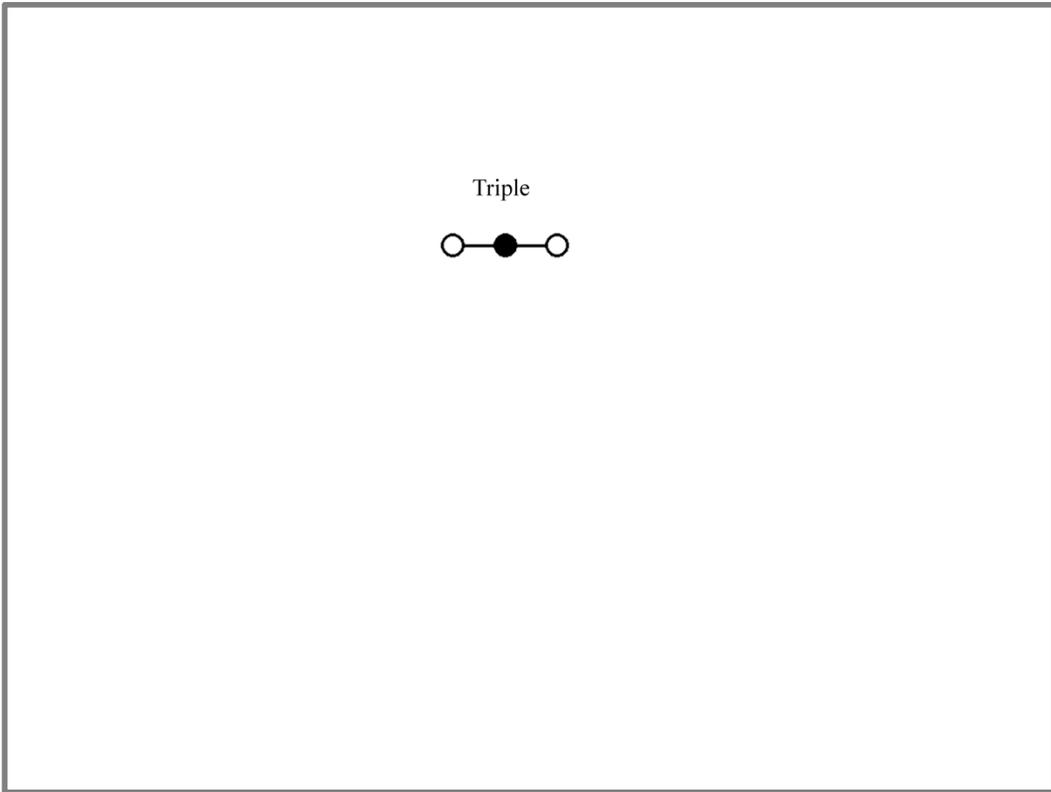


... and in between these two things is, what's known in Semantic Web parlance as, the "Predicate."

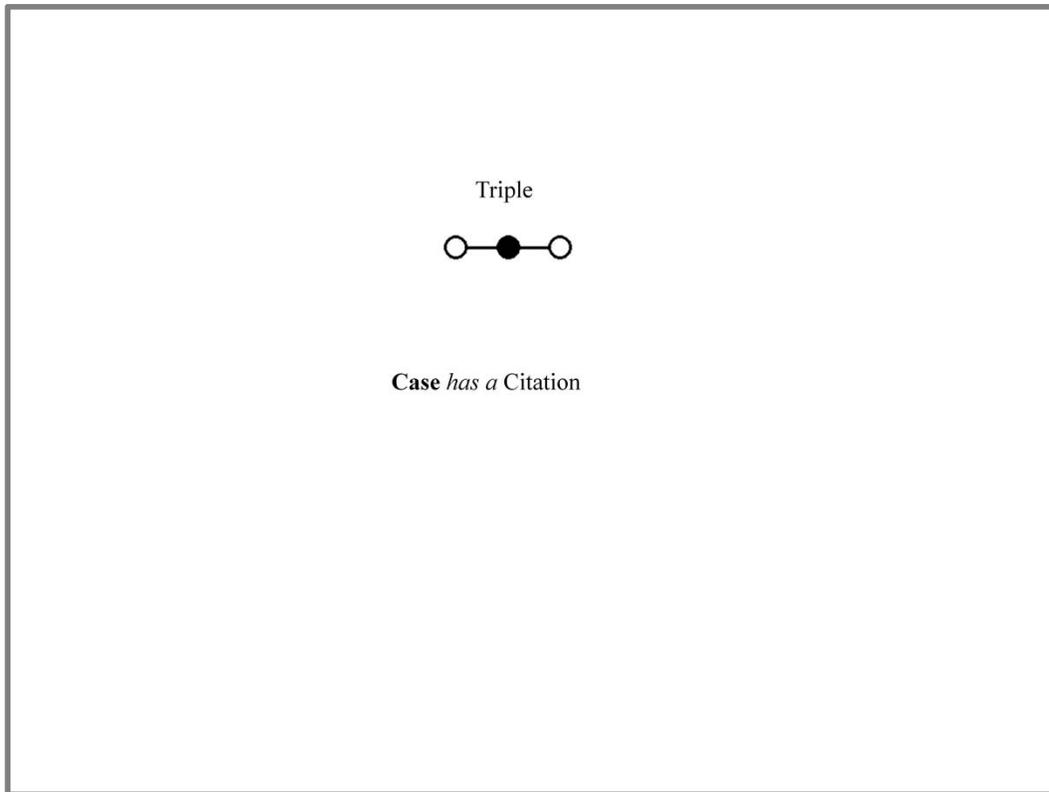
In grammatical terms the predicate tells us something about the subject, it's a property or characteristic that the subject has.



I prefer to think of this as the Relationship: the thing, or property that connects the Subject to the Object.



When a Subject and an Object are connected by a Relationship we have a Triple.



Some common triples you might find associated with legal resources could be:

a Case has a Citation

Triple



Case *has a Citation*

Court *has a Name*

a Court has a Name

Triple



Case *has a Citation*

Court *has a Name*

Resource *has a Format*

a Resource has a Format
etc.

COURT OF APPEAL FOR ONTARIO

CITATION: Royal Bank of Canada v. King, 2013 ONCA 524
DATE: 20130822
DOCKET: C57108

Blair, Watt and Pepall JJ.A.

BETWEEN

Royal Bank of Canada

Plaintiff (Respondent)

and

Leslie King

Defendant (Appellant)

Leslie King in person

Amanda Jackson for the respondent

Heard: August 19, 2013

On appeal from the judgment of Justice J.A. Ramsay of the Superior Court of Justice, dated May 23, 2013.

APPEAL BOOK ENDORSEMENT

[1] Mr. King owed the Royal Bank approximately \$275,000 on a mortgage loan, a line of credit and various credit card accounts. He does not dispute this. He defaulted. The Bank took steps to collect and obtained summary judgment.

[2] Mr. King appeals, arguing there are genuine issues for trial. We do not agree. His principal argument is that he paid the Bank. He did not. What he did was present an instrument to the Bank purporting to be payable to Royal Bank of Canada in the amount of \$275,000 and drawn on an account with Bank of Montreal in which he admits he had no funds. The fact that Royal Bank mistakenly credited his loan accounts with the monies – including the account containing the \$777.56 credit upon which Mr. King relies – before it learned that the instrument had been dishonoured is not material. Nor do the various provisions of the Canadian Payments Association Rule, which apply as between banks, assist him.

[3] He has not paid. There is no genuine issue for trial and the motion judge properly granted summary judgment, in our view.

[4] The appeal is therefore dismissed. The respondent is entitled to its costs of the appeal and of the stay motion fixed in the total amount of \$10,000 including disbursements and GST.

If we look at a simple appellate case like this one, we can pull out some of the data elements and then express them as a collection of Triples.

For example, this entire set of data elements could be associated with the Subject of a bunch of Triples that we might call, “Case A.”

COURT OF APPEAL FOR ONTARIO

CITATION: Royal Bank of Canada v. King, 2013 ONCA 524

DATE: 20130822

DOCKET: C57108

Blair, Watt and Pepall JJ.A.

Case A *has docket number* C57108

BETWEEN

Royal Bank of Canada

Plaintiff (Respondent)

and

Leslie King

Defendant (Appellant)

Leslie King in person

Amanda Jackson for the respondent

Heard: August 19, 2013

On appeal from the judgment of Justice J.A. Ramsay of the Superior Court of Justice, dated May 23, 2013.

APPEAL BOOK ENDORSEMENT

[1] Mr. King owed the Royal Bank approximately \$275,000 on a mortgage loan, a line of credit and various credit card accounts. He does not dispute this. He defaulted. The Bank took steps to collect and obtained summary judgment.

[2] Mr. King appeals, arguing there are genuine issues for trial. We do not agree. His principal argument is that he paid the Bank. He did not. What he did was present an instrument to the Bank purporting to be payable to Royal Bank of Canada in the amount of \$275,000 and drawn on an account with Bank of Montreal in which he admits he had no funds. The fact that Royal Bank mistakenly credited his loan accounts with the monies – including the account containing the \$777.56 credit upon which Mr. King relies – before it learned that the instrument had been dishonoured is not material. Nor do the various provisions of the Canadian Payments Association Rule, which apply as between banks, assist him.

[3] He has not paid. There is no genuine issue for trial and the motion judge properly granted summary judgment, in our view.

[4] The appeal is therefore dismissed. The respondent is entitled to its costs of the appeal and of the stay motion fixed in the total amount of \$10,000 including disbursements and GST.

So, for example:

Case A has docket number C57108

COURT OF APPEAL FOR ONTARIO

CITATION: Royal Bank of Canada v. King, 2013 ONCA 524

DATE: 20130822

DOCKET: C57108

Blair, Watt and Pepall JJ.A.

BETWEEN

Royal Bank of Canada

Case A *has plaintiff* Royal Bank of Canada

Plaintiff (Respondent)

and

Leslie King

Defendant (Appellant)

Leslie King in person

Amanda Jackson for the respondent

Heard: August 19, 2013

On appeal from the judgment of Justice J.A. Ramsay of the Superior Court of Justice, dated May 23, 2013.

APPEAL BOOK ENDORSEMENT

[1] Mr. King owed the Royal Bank approximately \$275,000 on a mortgage loan, a line of credit and various credit card accounts. He does not dispute this. He defaulted. The Bank took steps to collect and obtained summary judgment.

[2] Mr. King appeals, arguing there are genuine issues for trial. We do not agree. His principal argument is that he paid the Bank. He did not. What he did was present an instrument to the Bank purporting to be payable to Royal Bank of Canada in the amount of \$275,000 and drawn on an account with Bank of Montreal in which he admits he had no funds. The fact that Royal Bank mistakenly credited his loan accounts with the monies – including the account containing the \$777.56 credit upon which Mr. King relies – before it learned that the instrument had been dishonoured is not material. Nor do the various provisions of the Canadian Payments Association Rule, which apply as between banks, assist him.

[3] He has not paid. There is no genuine issue for trial and the motion judge properly granted summary judgment, in our view.

[4] The appeal is therefore dismissed. The respondent is entitled to its costs of the appeal and of the stay motion fixed in the total amount of \$10,000 including disbursements and GST.

Case A has plaintiff Royal Bank of Canada

COURT OF APPEAL FOR ONTARIO

CITATION: Royal Bank of Canada v. King, 2013 ONCA 524

DATE: 20130822

DOCKET: C57108

Blair, Watt and Pepall JJ.A.

BETWEEN

Royal Bank of Canada

Plaintiff (Respondent)

and

Leslie King

Defendant (Appellant)

Case A *has defendant* Leslie King

Leslie King in person

Amanda Jackson for the respondent

Heard: August 19, 2013

On appeal from the judgment of Justice J.A. Ramsay of the Superior Court of Justice, dated May 23, 2013.

APPEAL BOOK ENDORSEMENT

[1] Mr. King owed the Royal Bank approximately \$275,000 on a mortgage loan, a line of credit and various credit card accounts. He does not dispute this. He defaulted. The Bank took steps to collect and obtained summary judgment.

[2] Mr. King appeals, arguing there are genuine issues for trial. We do not agree. His principal argument is that he paid the Bank. He did not. What he did was present an instrument to the Bank purporting to be payable to Royal Bank of Canada in the amount of \$275,000 and drawn on an account with Bank of Montreal in which he admits he had no funds. The fact that Royal Bank mistakenly credited his loan accounts with the monies – including the account containing the \$777.56 credit upon which Mr. King relies – before it learned that the instrument had been dishonoured is not material. Nor do the various provisions of the Canadian Payments Association Rule, which apply as between banks, assist him.

[3] He has not paid. There is no genuine issue for trial and the motion judge properly granted summary judgment, in our view.

[4] The appeal is therefore dismissed. The respondent is entitled to its costs of the appeal and of the stay motion fixed in the total amount of \$10,000 including disbursements and GST.

Case A has defendant Leslie King

COURT OF APPEAL FOR ONTARIO

Case A was heard in Court of Appeal for Ontario

CITATION: Royal Bank of Canada v. King, 2013 ONCA 524

DATE: 20130822

DOCKET: C57108

Blair, Watt and Pepall JJ.A.

BETWEEN

Royal Bank of Canada

Plaintiff (Respondent)

and

Leslie King

Defendant (Appellant)

Leslie King in person

Amanda Jackson for the respondent

Heard: August 19, 2013

On appeal from the judgment of Justice J.A. Ramsay of the Superior Court of Justice, dated May 23, 2013.

APPEAL BOOK ENDORSEMENT

[1] Mr. King owed the Royal Bank approximately \$275,000 on a mortgage loan, a line of credit and various credit card accounts. He does not dispute this. He defaulted. The Bank took steps to collect and obtained summary judgment.

[2] Mr. King appeals, arguing there are genuine issues for trial. We do not agree. His principal argument is that he paid the Bank. He did not. What he did was present an instrument to the Bank purporting to be payable to Royal Bank of Canada in the amount of \$275,000 and drawn on an account with Bank of Montreal in which he admits he had no funds. The fact that Royal Bank mistakenly credited his loan accounts with the monies – including the account containing the \$777.56 credit upon which Mr. King relies – before it learned that the instrument had been dishonoured is not material. Nor do the various provisions of the Canadian Payments Association Rule, which apply as between banks, assist him.

[3] He has not paid. There is no genuine issue for trial and the motion judge properly granted summary judgment, in our view.

[4] The appeal is therefore dismissed. The respondent is entitled to its costs of the appeal and of the stay motion fixed in the total amount of \$10,000 including disbursements and GST.

Case A was heard in the Court of Appeal for Ontario

And so on ...

Case A *has docket number* C57108

Case A *has plaintiff* Royal Bank of Canada

Case A *has defendant* Leslie King

Case A *was heard in* Court of Appeal for Ontario

If we then list these four triples you'd get a collection of triple statements like this. The Subject for these triples is 'Case A' each followed by a relationship or property of some sort and a value corresponding to that relationship or property.

For those of you familiar with relational databases ...

Case A *has docket number* C57108

Case A *has plaintiff* Royal Bank of Canada

Case A *has defendant* Leslie King

Case A *was heard in* Court of Appeal for Ontario

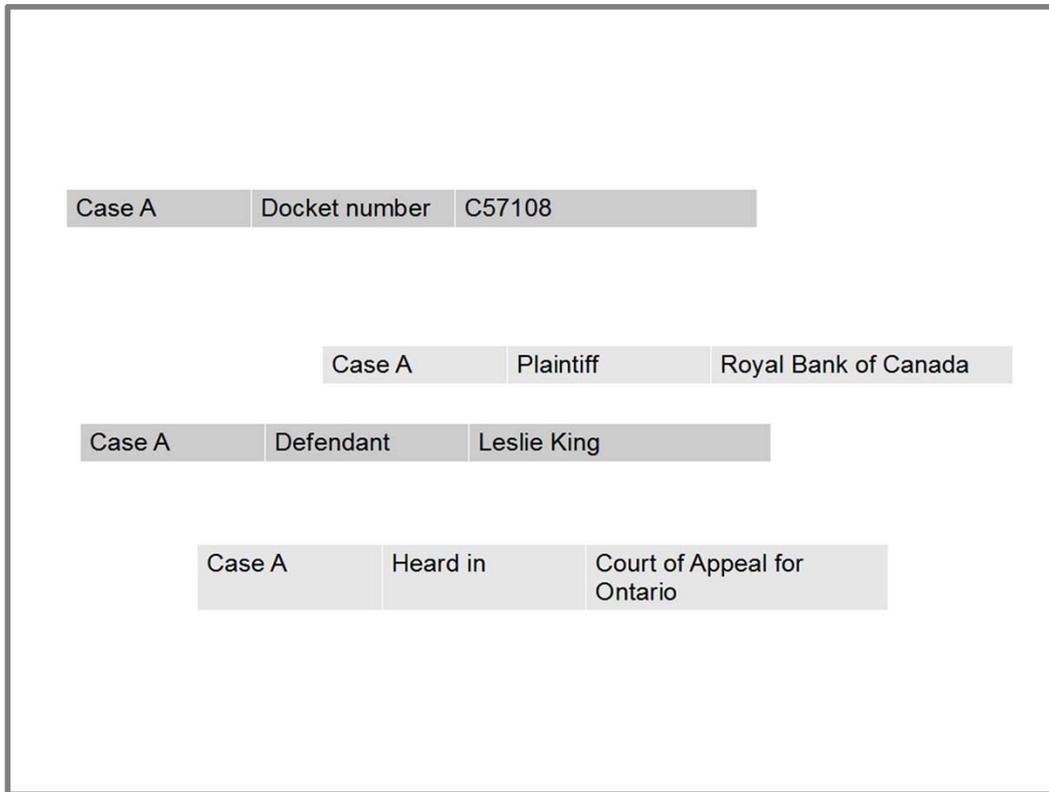
Primary Key	Property	Value
Case A	Docket number	C57108
Case A	Plaintiff	Royal Bank of Canada
Case A	Defendant	Leslie King
Case A	Heard in	Court of Appeal for Ontario

... you might think this is starting to look very much like a table with a primary key set as 'Case A.'

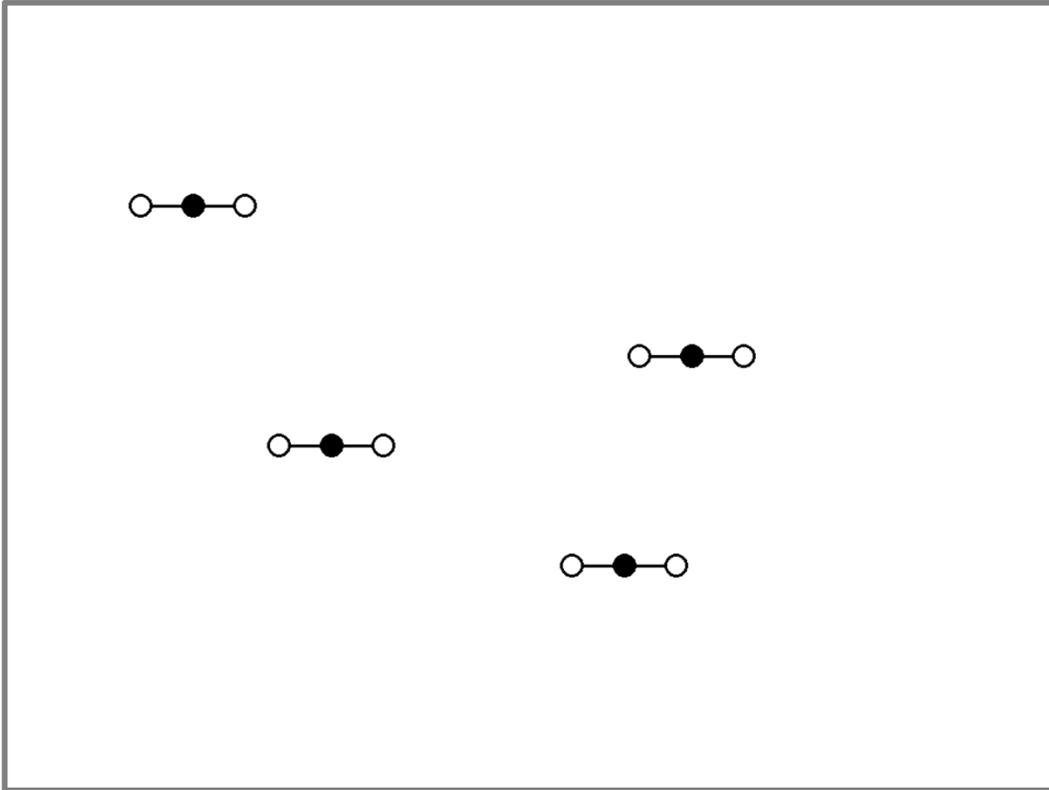
And traditionally that's how data like this has been represented: as a collection of data elements in a database record.

And really there's nothing wrong with that.

However, this data is confined to the relational database within which it was created.

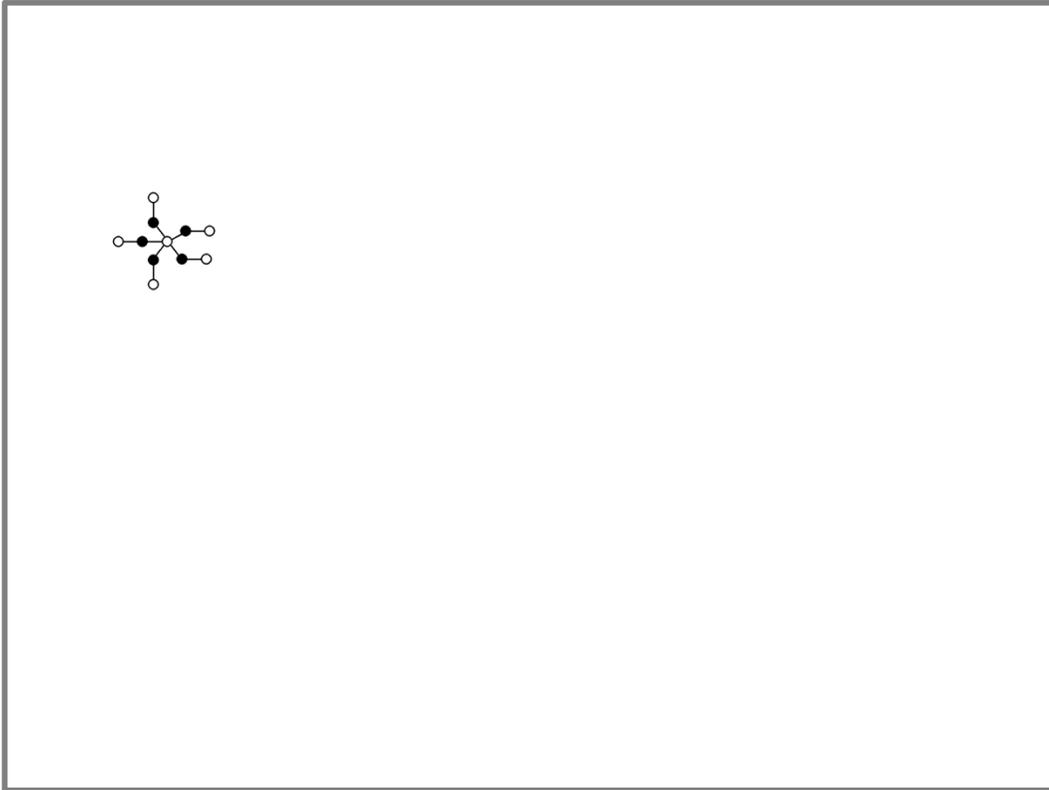


The beauty of linked data is that it allows the record structure to break down ...



... with the resulting triples able to connect independently with information from any number of other data sources that might be out there.

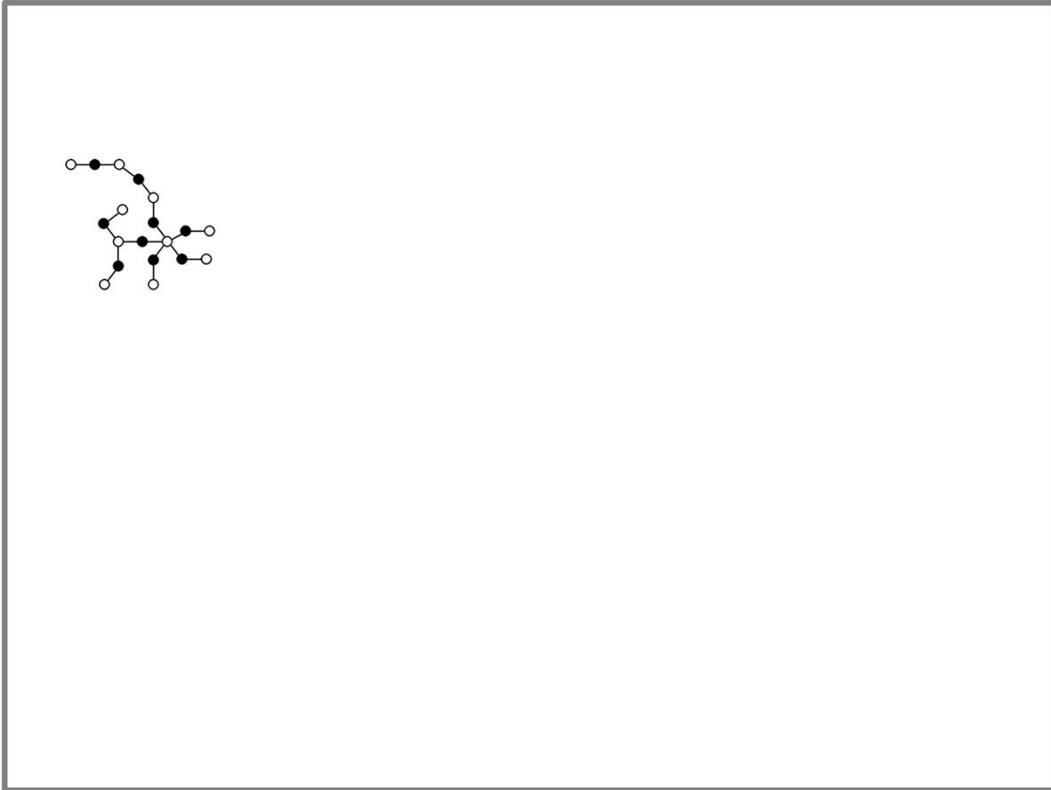
The triples are available to float freely in the semantic web.



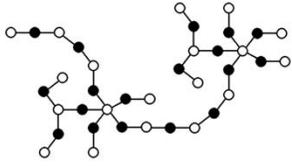
You could still bring all of these data elements together as a group with a query that sets out to find our original set of 'Case A' data elements.

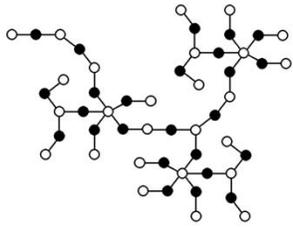
And we could represent the result with a graph like this one where the circle in the centre is the Subject 'Case A' and the spokes around it are the various relationships and their associated objects.

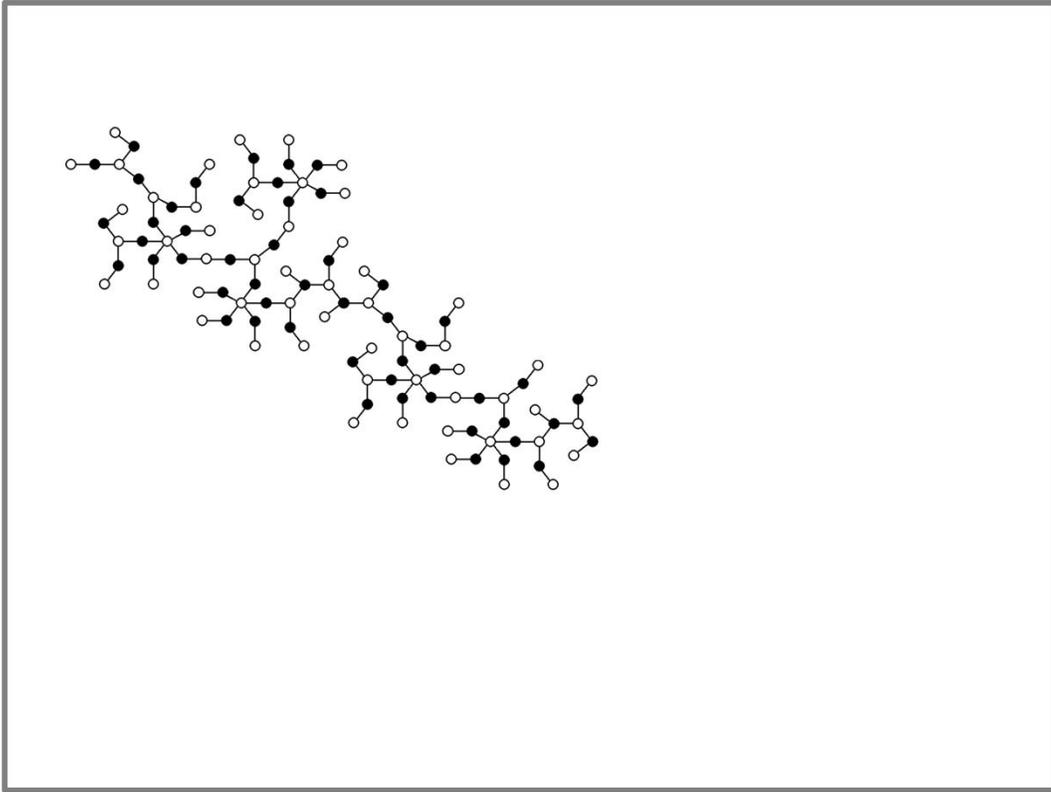
But once our data is broken up into Triples other connections can be made automatically between data sources which can both enhance our data and enhance the data that others provide to us.



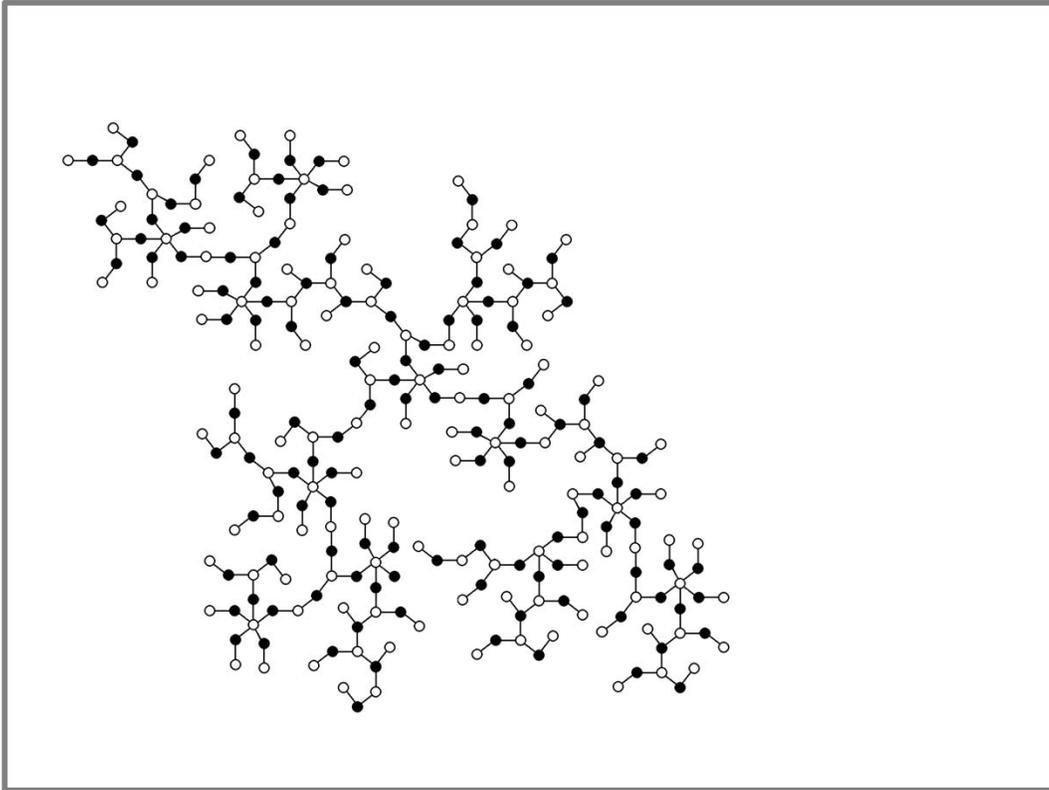
And suddenly, and I imagine quite quickly, once we get enough data to reach a critical mass, our data begins to link up with someone else's data ...







And before you know it there is a chain of resources linked together through their common data points.



Creating what I imagine as these crystalline structures that will be able to connect up in many different ways. Connections with the potential to reveal new relationships between the many Subjects, Objects and Relationships in existing data sets.

“... associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. The process of tying two items together is the important thing.” — Vannevar Bush

As We May Think / Vannevar Bush <<http://www.cs.sfu.ca/CourseCentral/365/mark/material/notes/Chap1/VBushArticle/vbush-all.html>>

When I was in library school in the 90s we were introduced to the work of Vannevar Bush and in particular his article “As we may think” describing what he called the Memex system. Although this paper was published in 1945 I return to this periodically amazed by his forward looking ideas. Truly ahead of his time.

Well it seems that time is catching up to him and this passage citing what Bush calls 'associative indexing' is particularly relevant to today's developments of linked data and how linked data works.

“... associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. The process of tying two items together is the important thing.”

*“... associative indexing, the basic idea of which is a provision whereby any **resource** may be caused at will to select immediately and automatically another. This is the essential feature of **linked data**. The process of tying two items together is the important thing.”*

So if I can take the liberty to make a minor adjustment and substitute a couple of contemporary terms it could be rewritten to read something like this:

*“... associative indexing, the basic idea of which is a provision whereby any **resource** may be caused at will to select immediately and automatically another. This is the essential feature of **linked data**. The process of tying two items together is the important thing.”*

And this process of tying two things together, the Triple, is really the kernel that leads to the potential of machine learning and machine actionable data.

OK, so let's take quick a look under the hood, so-to-speak.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

Here's an example of some RDF coded in XML. RDF stands for the Resource Description Framework and is one of the popular standards at work in the semantic web.

This particular example is based on the appellate case we looked at earlier.

I'll just run through what this coding represents.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

We start by declaring the namespaces to be used so that we know, and the machine knows, which terms and which standards we are drawing from.

The first few lines of code here respectively reference: the W3C RDF syntax document; the Dublin Core element set; and the last refers to a local namespace.

This last namespace is a fictitious one that I've called 'lex'. Ideally this would lead to an actual metadata document containing the terms defined in that namespace.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

Then we have the RDF description block that contains the code for the triples that represent the document, the appellate case we were looking at earlier.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

The Subject for all of the triples expressed here, is what I previously referred to as ‘Case A’ in our example. In RDF the Subject is expressed using the RDF attribute ‘about’. So this is telling us that the RDF description is about ‘Case A’ only here I’m using a URI.

Below that line we’ve got the 4 Objects that we identified earlier and their Predicates.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

The docket number.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

The plaintiff.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

The defendant.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

And the name of the court.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:lex="http://example.org/legal-terms/1.0/">
  <rdf:Description
    rdf:about="http://canlii.ca/t/g0656"
    dc:title="Royal Bank of Canada v. King"
    lex:docket="C57108"
    lex:plaintiff="Royal Bank of Canada"
    lex:defendant="Leslie King"
    lex:court="Court of Appeal for Ontario"
    dc:date="2013-08-22">
  </rdf:Description>
</rdf:RDF>
```

I've added a couple of additional triples to illustrate the use of the Dublin Core namespace and how metadata schemes can interoperate with each other in RDF.

The Predicate is found in the information appearing before the equals sign here. So the Subject, defined by the URI for 'Case A' here, has the title 'Royal Bank of Canada v. King.'

Case A *has docket number* C57108

Case A *has plaintiff* Royal Bank of Canada

Case A *has defendant* Leslie King

Case A *was heard in* Court of Appeal for Ontario

So when we serialize this code we'll get the same 4 triples we had from the beginning ...

<http://canlii.ca/t/g0656> *has docket number* C57108

<http://canlii.ca/t/g0656> *has plaintiff* Royal Bank of Canada

<http://canlii.ca/t/g0656> *has defendant* Leslie King

<http://canlii.ca/t/g0656> *was heard in* Court of Appeal for Ontario

... but 'Case A' is replaced by a URI here. This is the HTTP URI for the case as it's found in the CanLII database ...

<http://canlii.ca/t/g0656> has docket number C57108

<http://canlii.ca/t/g0656> has plaintiff Royal Bank of Canada

<http://canlii.ca/t/g0656> has defendant Leslie King

<http://canlii.ca/t/g0656> was heard in Court of Appeal for Ontario

<http://canlii.ca/t/g0656> has title Royal Bank of Canada vs. King

<http://canlii.ca/t/g0656> has date 2013-08-22

... and also have the additional 2 triples used in the example referencing the Dublin Core. Ideally each of the parts of the triple will also be described using a URI.

OK, so that's a general overview of linked data theory.

Principles of Linked Data

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs, so that they can discover more things

Linked Data - Design Issues / Tim Berners-Lee <<http://www.w3.org/DesignIssues/LinkedData.html>>

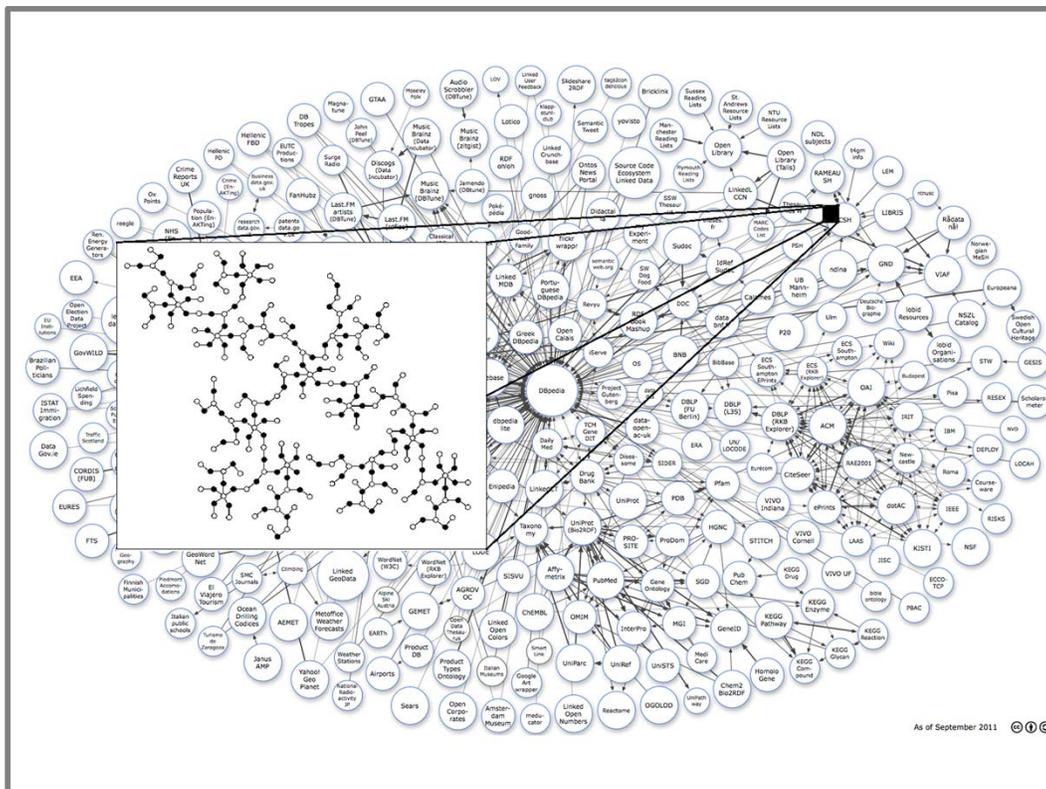
An important figure in all of this is of course Tim Berners-Lee: “inventor of the World Wide Web” and champion of the semantic web.

On a website that considers linked data design issues, Tim Berners-Lee provides the four basic principles of linked data:

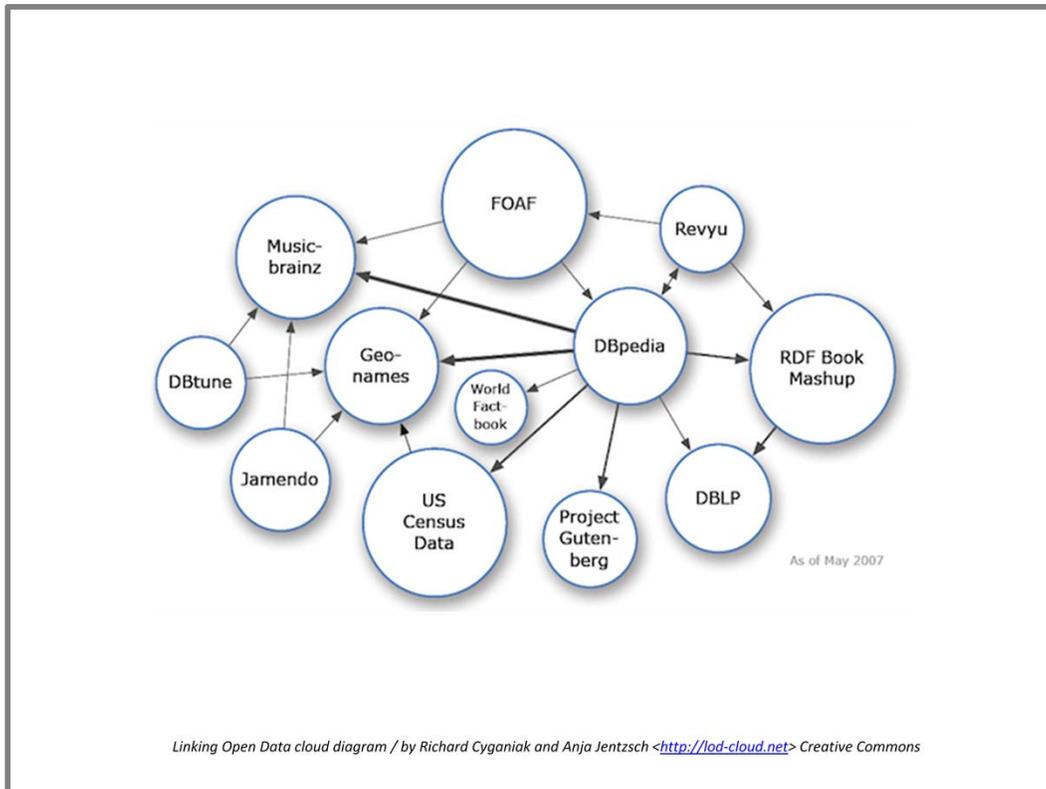
1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide them with useful information using standards like RDF and SPARQL
4. Include links to other URIs so that they can discover more things

So, a URI is a Uniform Resource Identifier and an example is the HTTP URI, which we all know as the familiar URL, like the one we just used in the RDF example.

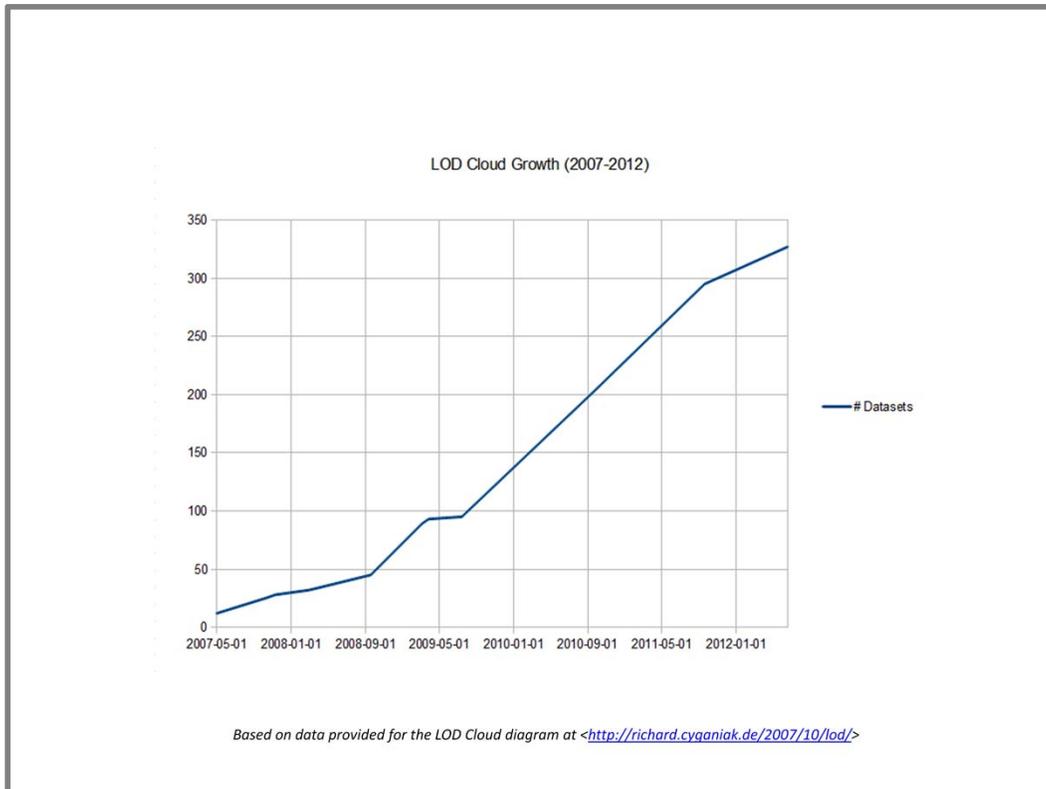
SPARQL is a query language used to search across RDF data sets known as Triple Stores.



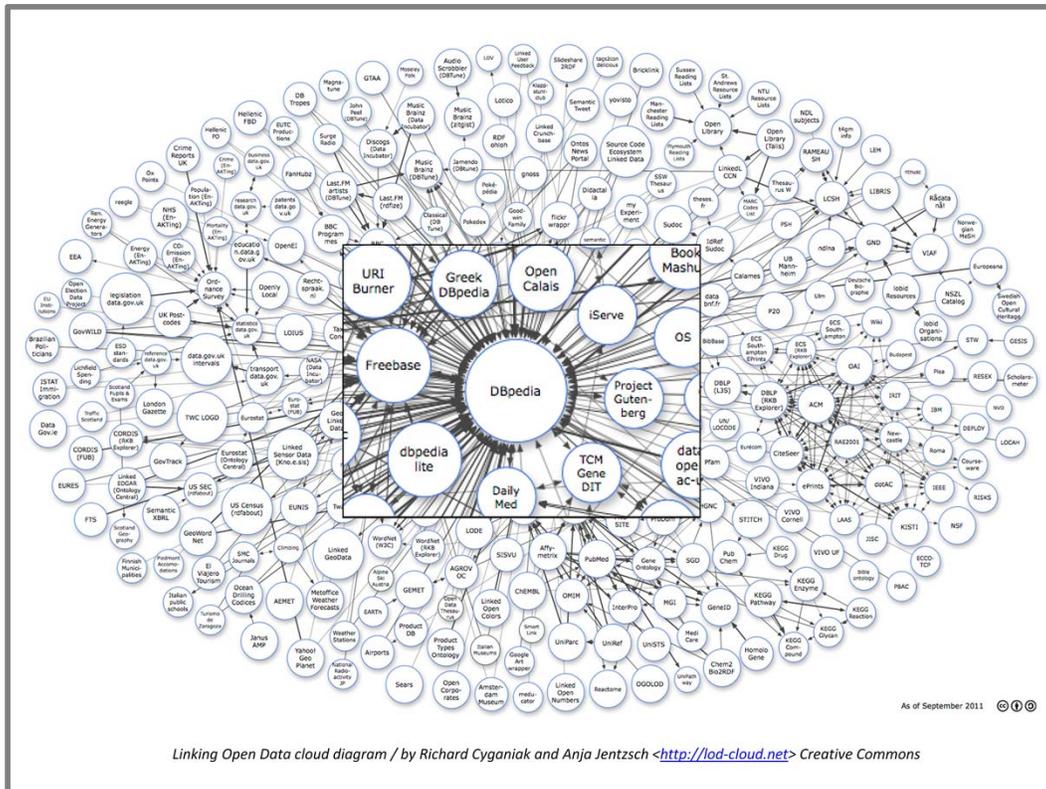
If you recall my earlier crystalline structure you might consider it as a very tiny piece of the linked open data cloud. And although there were only 295 linked open data sets in 2011 that still represented somewhere in the neighbourhood of 32 billion triples.



Which is remarkable considering the modest beginnings of the linked open data cloud which began with only 12 data sets in 2007.

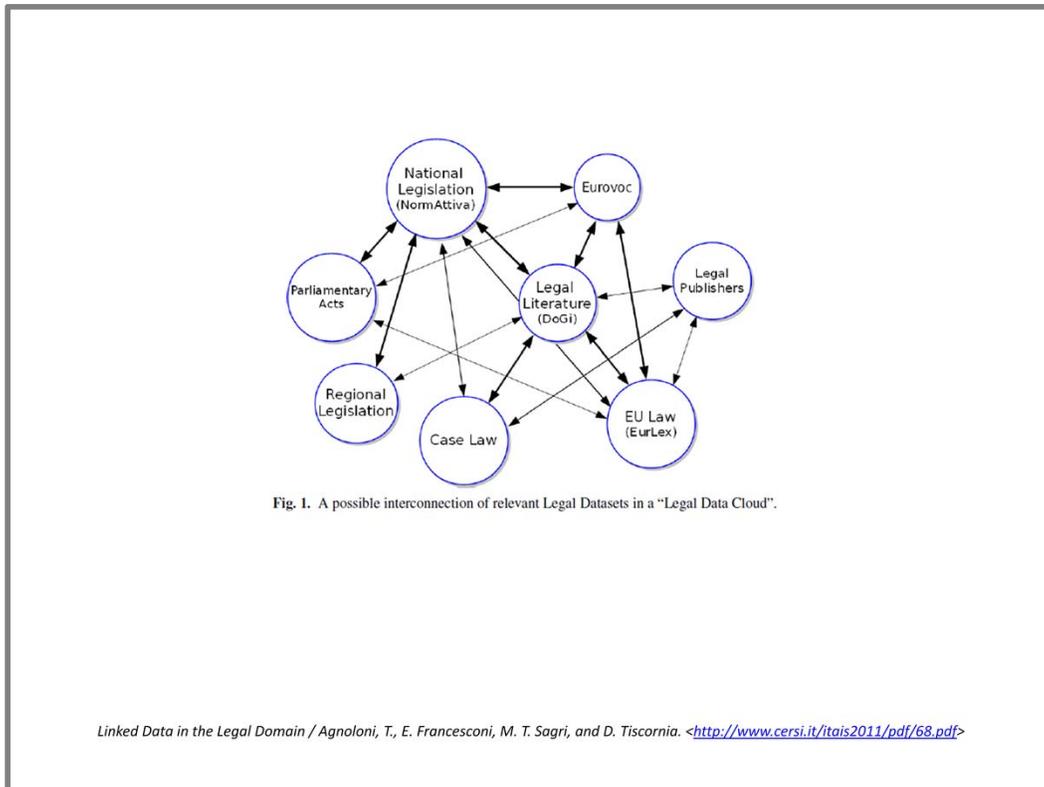


As shown in this chart, the linked open data cloud has grown significantly over the last 4 or 5 years.



You can see that **DBPedia** figures prominently here, shown in the center of this cloud. DBPedia was created by Chris Bizer one of the leading linked data developers working in the field today. He had the idea to extract the structured information already embedded in Wikipedia.

Since its creation many data set providers have linked their data sets to DBPedia which has made it the center of the linked open data hub.



This diagram, which looks very much like the early 2007 version of the Linked Open Data Cloud, is from a recent article called “Linked data in the legal domain.” It’s based on some recent work being done at the **Institute of Legal Information Theory and Techniques** of the **Italian National Research Council**.

This represents what the authors call the **Legal Data Cloud** and shows some relevant Italian and European Union legal data sets. I’d like to return to this at the end of the presentation and ask you what a similar diagram for Canadian legal resources might look like.

I’ll turn now to a look at the nature and characteristics of legal resources.

"Law is chaos with an index."

I'm sure many of you have heard this phrase before:

"Law is chaos with an index" often attributed to Oliver Wendell Holmes. But Holmes apparently paraphrased this from a collection of essays by British lawyer, Sir Thomas Holland, which he had reviewed for the American Law Review.

Regardless of its origin this is an apt description of the legal domain.

In a short paper about linked legal data efforts in Finland the authors provide a very nice summary of this potential chaos in the opening paragraph of their report.

“Publishing and using juridical information is challenging in many ways. It is produced by different parties, such as governmental bureaus, ministries, different levels of courts, research organizations, and media. The content is heterogeneous and produced using differing tools, data formats, and practices. The links between documents are often informal and/or not made explicit. The law in general is a dynamic, changing entity: for example, it is important to be able to refer to different versions of a law at different points of time. These challenges can be addressed through the use of linked data techniques.”

The Finnish Law as a Linked Data Service / Matias Frosterus, Jouni Tuominen, Mika Wahlroos, and Eero Hyvonen
<<http://www.seco.tkk.fi/publications/2013/frosterus-et-al-finnish-law-lod-poster.pdf>>

“Publishing and using juridical information is challenging in many ways. It is produced by different parties, such as governmental bureaus, ministries, different levels of courts, research organizations, and media. The content is heterogeneous and produced using differing tools, data formats, and practices. The links between documents are often informal and/or not made explicit. The law in general is a dynamic, changing entity: for example, it is important to be able to refer to different versions of a law at different points of time. These challenges can be addressed through the use of linked data techniques.”

“The traditional fields and practices of law are changing fast. Legal drafting, private contracting, judicial sentencing and administrative management have been enlarged with online dispute resolution initiatives and new forms of self-regulation and access to justice. Citizens, customers and consumers require a greater participation and faster and more effective ways of facing their legal activities.”

The Flood , the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World / Joost Breuker and others
<<http://dl.acm.org/citation.cfm?id=1563987.1563990>>

Speaking about the need for semantic web technology in the legal domain Joost Breuker and others wrote in 2009,

“The traditional fields and practices of law are changing fast. Legal drafting, private contracting, judicial sentencing and administrative management have been enlarged with online dispute resolution initiatives and new forms of self-regulation and access to justice. Citizens, customers and consumers require a greater participation and faster and more effective ways of facing their legal activities.”

So along with increased user expectations for access to legal information over the past 5 years or so there has also been a change in how law is being practiced. In the same article, for example, the authors note that 2008 marked the point when the number of lawyers using free online services had, for the first time, overtaken those using for-fee services (89% vs. 83%).



As the authors suggest here, the heterogeneous reality of the legal domain and the increase in user expectations is something that might be successfully addressed in a linked data environment.

But it really depends on whether we can build a critical mass of data sources in a standard format like RDF/XML. Using RDF/XML helps to normalize the data making it more homogeneous and therefore easier for machines and humans alike to discover and make better connections between the various data sources that are available.

Once the data is up we can start to place our efforts on the development and consistent use of the metadata that describes and gets people to the documents relevant to assist with their legal issues.

URN:Lex

A Uniform Resource Name (URN) Namespace for Sources of Law (LEX)

“The purpose of the "lex" namespace is to assign an unequivocal identifier, in standard format, to documents that are sources of law. The identifier is conceived so that its construction depends only on the characteristics of the document itself and is, therefore, independent from the document's on-line availability, its physical location, and access mode.”

A Uniform Resource Name (URN) Namespace for Sources of Law (LEX)
<<http://tools.ietf.org/html/draft-spinosa-urn-lex-01>>

One promising metadata project, still in the early stages of development, is URN:Lex. This proposal, submitted to the Internet Engineering Task Force in 2010, is a product of a number of groups and led by the Institute of Legal Information Theory and Techniques of the Italian National Research Council and also involving Cornell’s Legal Information Institute.

“The purpose of the "lex" namespace is to assign an unequivocal identifier, in standard format, to documents that are sources of law. The identifier is conceived so that its construction depends only on the characteristics of the document itself and is, therefore, independent from the document's on-line availability, its physical location, and access mode.”

So this is one area where efforts might be focused.

Metadata for Legal Resources

CEN MetaLex: Open XML Interchange Format for Legal and Legislative Resources
<<http://www.metalex.eu/>>

CitationStylist: For the Laws and Languages of the World
<<http://citationstylist.org/>>

Legal XML
<<http://www.legalxml.org/>>

LII Modeling Legislative Information
<<http://blog.law.cornell.edu/metasausage/downloads-and-related-information/>>

OAI4Courts
<<http://oai4courts.wikispaces.com/>>

Metadata Resources: Legal (Selected)
<<http://www.personal.psu.edu/rcr5122/Metadata.html>>

There are a number of other legal metadata initiatives reinforcing the notion that “standards are great because there are so many to choose from.” I’ve listed a few for you here that you can take a look at later.

And I draw your attention to Robert Richards list of resources dealing with “metadata specifically designed for legal information.”

Types and Roles of Legal Ontologies

- Organize and structure information
- Reasoning and problem solving
- Semantic indexing and search
- Semantics integration and interoperation
- Understanding the domain

Types and Roles of Legal Ontologies / Andre Valente <http://link.springer.com/chapter/10.1007/978-3-540-32253-5_5>

And another area that disserves our attention is work on the development and use of ontologies for legal resources. It is the ontologies that define the metadata structures and how they are to be used.

Andre Valente in his chapter on Types and Roles of Legal Ontologies identifies five main uses or roles for ontologies:

- (a) To organize and structure information;
- (b) To provide reasoning and problem solving;
- (c) To enable semantic indexing and search;
- (d) To provide semantics integration and interoperation; and
- (e) To understand the domain.

Ontologies for Legal Resources

LEX Ontology

<<https://code.google.com/p/lex-ontology/>>

LKIF Core Ontology

<http://www.estrellaproject.org/?page_id=3>

Valente, Andre. Types and Roles of Legal Ontologies

<<http://www.legalxml.org/>>

There are also quite a few legal ontologies available. In the opening chapter of the book *Law, Ontologies and the Semantic Web* 23 separate legal ontologies are identified. I've highlighted a couple here and Valente also describes a dozen or so other legal ontologies in the book chapter cited earlier.

Role for CanLII?

So, is there a role for CanLII in all of this?

I think there is an essential role for CanLII to play here. As a major aggregator of Canadian case law CanLII is already an important source for legal data. And contributing this already structured legal data to the linked data initiative is not only good for the linked open data community it supports CanLII's mission and their goal to provide "Free and unrestricted access to legal information."

And I applaud the recent release of the CanLII web API because it does expose this structured data that has already been collected. The API also enables new web-based applications to be built against the CanLII database which we should see examples of later this afternoon.



Some disadvantages with a web API:

- Data is enclosed in a proprietary database perpetuating data silos
- Developers must know you exist and decide what to access based on available API parameters
- Links between different data sources must be explicitly created
- Data is not included in the global database i.e. “*on the web*” but not “*of the web*”

How to Publish Linked Data on the Web / Chris Bizer and others <http://videlectures.net/iswc08_heath_hpldw/>

However as Chris Bizer pointed out in a presentation from way back in 2007, there are some disadvantages to relying solely on a web API for access to this data:

- The data remains enclosed in a proprietary database which isolates the data and perpetuates the data silo mentality
- Software developers must know you exist and where you are and then decide how they will use the available API parameters
- Links between different data sources must be explicitly created, i.e. hard coded rather than discovered as needed when a query is made
- Therefore the data is not part of the global database, i.e. the data is “*on the web*” but it’s not “*of the web*”.

“The important thing about data is the more things you have to connect together the more powerful it is.”—Tim Berners-Lee

The Next Web / Tim Berners-Lee <http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html> (8:05)

So the first step in my mind is to make the data available as linked open data.

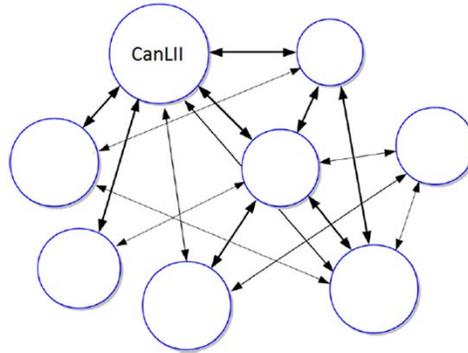
AS Tim Berners-Lee said in his TED Talk in 2009:

“The important thing about data is, the more things you have to connect together, the more powerful it is.”



If we want to support the linked open data initiative and be part of this emerging global database we should be working towards adding our structured legal data to the open cloud of linked data. Every contribution is useful and will lead to a critical mass of legal data.

Tim Berners-Lee had the audience chanting 'Raw Data Now', encouraging those people who are keepers of the data to put it up as linked data now ...



There is a lot of work being done in Europe and the U.S. but activity in this country is rather quiet. But it seems like a logical next step for CanLII. CanLII is well positioned to take their place at the centre of the open legal data cloud in Canada. And I humbly encourage them to take this opportunity to do so.

I'm sure there are others in a similar position who might be interested and able to share their own data in this way.

I'd be interested to hear about any legally oriented or government organizations in Canada that might be considering participation in the development of a legal data cloud in Canada ...

References Consulted

Agnoloni, T., E. Francesconi, M. T. Sagri, and D. Tiscornia. "Linked Data in the Legal Domain." <<http://www.cersi.it/itais2011/pdf/68.pdf>>

Berners-Lee, Tim. "Linked Data - Design Issues," July 27, 2006. <<http://www.w3.org/DesignIssues/LinkedData.html>>

Berners-Lee, Tim. "The Next Web" | Video on TED.com. <http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html>

Bizer, Chris, Richard Cyganiak, and Tom Heath. "How to Publish Linked Data on the Web." Tutorial, 2007. <<http://www4.wiwiwiss.fu-berlin.de/bizer/pub/linkeddattutorial/>>

Breuker, Joost, Pompeu Casanovas, Michel C. A. Klein, and Enrico Francesconi. "The Flood, the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World." In Proceedings of the 2009 Conference on Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood. Amsterdam, The Netherlands: IOS Press, 2009. <<http://dl.acm.org/citation.cfm?id=1563987.1563990>>

Bush, Vannevar. "As We May Think." Atlantic Monthly, July 1945. <<http://www.cs.sfu.ca/CC/365/mark/material/notes/Chap1/vbushArticle/vbush-all.html>>

Matias Frosterus, Jouni Tuominen, Mika Wahlroos, and Eero Hyvonen. "The Finnish Law as a Linked Data Service," 2013. <<http://www.seco.tkk.fi/publications/2013/frosterus-et-al-finnish-law-lod-poster.pdf>>

Valente, Andre. "Types and Roles of Legal Ontologies." In Law and the Semantic Web, edited by V. Richard Benjamins, Pompeu Casanovas, Joost Breuker, and Aldo Gangemi, 65–76. Lecture Notes in Computer Science 3369. Springer Berlin Heidelberg, 2005. <http://link.springer.com/chapter/10.1007/978-3-540-32253-5_5>

I've included a list of the references I consulted for this presentation here for your information.

Thank you for your attention!

tknight@osgoode.yorku.ca

Twitter: @freemoth

