Packaging and Shipping Data

The construction and packaging of data and metadata

xmlns:rdf=

I Do Not Think That Means What You Think It Means

- Common definitions are the foundation of a Semantic Web – the Holy Grail of the WWW
- Better interoperability possible if everyone uses the same framework(s)
- We must agree to store data and metadata in similar ways so we can “play well together”
- The platform needs to accommodate wildly divergent ideas on how to describe things and how to construct those descriptions

xmlns:rdf=

Can’t We All Just Get Along?

- RDF: Conceived in 1996, originally published 2004, latest 2014
- Originally designed for Web, adopted in manifold enterprises
  Medicine, pharmaceuticals, aeronautics, media, etc.
- Store the schema and structure on a permanent common site (read only)
  xmlns:rdf=http://www.w3.org/TR/rdf-syntax-grammar
- Store the glossary (name space) on another Web site. For example, Dublin Core
  xmlns:dc="http://purl.org/dc/elements/1.1/"

Just Gotta Be Different?

- Anyone can build a resource description framework namespace

Peanut butter	smells like	old shoes
Chocolate chip	made with	walnuts
Crème sandwich	great for	dipping in milk
Ginger snap	best when	crunchy

- Store the standard in a common site
  - xmlns:cookies=http://www.missen.com/cookies#
- Now anyone can adopt my definitions for describing cookies

Tragedy of the Data Commons

- A lot of this is designed and maintained by academics and volunteers
- Occasional grants from government and foundations
- Benefits accrue to others who do not “foot the bill”
- Military, industry, and researchers use the basic RDF definitions and infrastructure, then and build their own namespaces, but only pay to develop their proprietary bits

Getting from Here to There

- Most organizations have a diverse collection of technologies they use to host and manage their databases
- Very likely they store much more information in their database and augment it with support for standards
- Library management software packages that handles purchasing, staffing, circulation, will also have portions that are RDF/XML compliant.
- They can build a crosswalk to import/export data in the correct format for existing systems or simply adopt the standard

Tupelo Honey: Storing Data in Tables

- Generally faster and efficient at sort, filter, find (SFF)
- Keyed, 1:1 relationships (**tuples**), constraints, triggers…

1046	Sharon (S.)	Brain	Bubba
1047	Abel	Bramley	Toots
1048	Marsha	Brandjes	
1049	Hanna	Brandler	Grandma
1050	Claire	Brands	Knuckles
1051	Robert	Brandt	Wispy
1052	Hector	Brandz	

Built-In Triples

PersonID	FirstName	LastName	Nickname
1046	Sharon (S.)	Brain	Bubba
1047	Abel	Bramley	Toots
1048	Marsha	Brandjes	
1049	Hanna	Brandler	Grandma
1050	Claire	Brands	Knuckles
1051	Robert	Brandt	Wispy
1052	Hector	Brandz	

1047 → FirstName → Abel

entity-attribute-value
subject-predicate-object

1:Many Preferences ("Favorites")

PersonID	FirstNar
1046	Sharc
1047	Abe
1048	Mc
1049	Han
1050	Claire
1051	Ro
1052	Hector

PersonID	Preference	Object
1046	Likes	Blimps
1046	Dislikes	Mustard gas
1048	Likes	Old movies
1049	Loves	Librarians
1049	Hates	Snuff
1051	Tolerates	Libertarians
1049	Likes	Bunnies

Each person can have preferences (controlled vocabulary) for multiple objects

Transferring Delimited Data

- Commas separate fields
- CR/LFs (carriage returns and line feeds) mark records

PersonID, FirstName, LastName, Nickname ¶ 1046, Sharon (S.), Brain, Bubba ¶ 1047, Abel, Bramley, Toots ¶ 1048, Marsha, Brandjes, ¶ 1049, Hanna, Brandler, Grandma ¶ 1050, Claire, Brands, Knuckles ¶ 1051, Robert, Brandt, Wispy ¶ 1052, Hector, Brandz, ,¶

- First row contains field names.

  Not too handy with 1 million records.

Delimited Data

PersonID,Honorary,FirstName,MiddleName,LastName,Generation,MaidenName,NickName,Degree,OtherDegree,Birthday,Gender,Deceased,Organization,Title,Department,BusinessAddress2,BusinessAddress1,BusinessAddress3,BusinessCity,BusinessState,BusinessZip,BusinessCountryID,HomeAddress1,HomeAddress2,HomeAddress3,HomeCity,HomeState,HomeZip,HomeCountryID,BusinessPhone1,BusinessPhone2,BusinessPhoneExt,BusinessFax,HomePhone,MobilePhone,Pager,HomePage,PhotoPath,Skype,Twitter,FacebookURL,LinkedIn,Chat1,NULL,Zurtapoz,,abaalla,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00",,,,,,,,,300,,,,,,,NULL,,,,,,,,,,,,,,2,Mr.
,,,,,,,,3,Mr.,Zuhumnan,Zuo,D,Abacan,,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00 00:00:00",,,,,,,,,NULL,,,,,,,NULL,,,,,,,Jagannadha,Abagyeh,,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00 00:00:00",,Secretary,"Centre for Community Development",,,,,,,NULL,,,,,,,100,,,,,"+91 9437062516",,,,,,,,,4,Mr.,Zuhair,,"Abah
",,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00 00:00:00","University of Jos","Deputy University Librarian",,"University of Jos","University of Jos
Library",,Nigeria,,,NULL,,,,,,,NULL,073-611932,,,,,,,,,,,,,5,,Zubair,J,"Abakpa ",,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","University of Iowa","Network
Communication Engr IV","ITS-TELECOMM & NETWORK SERVICES",,"416 NH",,"Iowa City",IA,52242,NULL,,,,,,224,"1 319 335
5016",,,,,,,,,,,,6,Ms.,Zoey,N,Abass,,,NULL,,,"0000-00-00 00:00:00",Female,"0000-00-00 00:00:00","University of Nigeria -
Nsukka",,,,,,Nssuka,,,NULL,,,,,,,NULL,,,,,,,,,,,,,7,,Ziqiang,A,Abatan,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","Bayero University","Computer
Support","Bursary Department",BUK,,,Kano,,,NULL,,,,,,156,234-064-665905,,,,,,,,,,,,8,Mr.,Zingfa,W,Abate,,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00
00:00:00","Bayero University","Pre-Clinical Librarian","Faculty of Medicine",BUK,,,Kano,,,NULL,,,,,,,156,,,,,,,,,,,,,9,Ms.,"Tobi ",A,Abay,,,NULL,,,"0000-00-00
00:00:00",Female,"0000-00-00 00:00:00","Bayero University","Computer Supervisor","Bursary Department",,BUK,,Kano,,,NULL,,,,,,,156,,,,,,,,,,,,,10,Ms.,"Ray
",,Abayomi,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","Bayero University",Director,"Academic Planning Unit",,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,11,,"Vannessa
",,Abba,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","NSVV Boys Higher Secondary School","Head Master",,,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,12,Mr.,"Elliot
",,Abbas,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","Tata-Dhan Academy",Librarian,"Boys' Town Campus",,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,13,Mr.,"Shakia
",BC,Abbaspour,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","Malawi Government",,Water,,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,14,,"Randolph
",M,Abbott,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","Kogi State University","Assistant Librarian(Graduate
Assistant)",,,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,15,,"Becki ",BM,Abbott,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00",vodacom,Chairman,"BP 797","Building Gulf
Oil","3157 Boulevard du 30 Juin",,Kinshasha,Gombe,,NULL,,,,,,,45,"+243 81 31 31 400",,,"+243 81 31 31 600",,"+243 81 444 8888",,,,,,,,16,Ms.,"Lucio
",Michael,Abbott,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00",,,,,,,,,224,"2336 North Ridge Dr",,,Coralvile,IA,52241,NULL,,,,319-466-9301,,,,,,,,17,Ms.,"Emma
",O.,Abboud,,,NULL,,,"0000-00-00 00:00:00",Female,"0000-00-00 00:00:00",,,,,,,,,,,NULL,,,,,,156,09-5235640,,,,,,,,,,,,18,Mr.,"Hyon ",,Abboud,,,NULL,,,"0000-00-00
00:00:00",Male,"0000-00-00 00:00:00",,,"Central Bank Nigeria",,,,Abuja,,,NULL,,,,,,,NULL,0804-543-3347,,,,,,,,,,,,19,Mr.,"Stuart ",A,Abboud,,,NULL,,,"0000-00-00
00:00:00",Male,"0000-00-00 00:00:00",,Dean,"Facility of Energy and Science",,,,,,,NULL,,,,,,,NULL,,,,,,,,,,,,,20,Mr.,"Meggan ",U.,Abdalla,,,NULL,,,"0000-00-00
00:00:00",Male,"0000-00-00 00:00:00","Bayero University","Director, MIS Unit",MIS,"PMB 3011","MIS, Bayero University",Kano,,,,NULL,,,,,,,156,064-668838,,,064-
665904,,,,,,,,,21,Alhaji,"Shyla ",Yeu,abdaralah,,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00 00:00:00","Bayero University",,Library,,,,Kano,,,NULL,,,,,,,156,234-064-
0621026,,,,,,,,,,,,,22,Mr.,"Rosemarie ",Yai,Abdella,,,NULL,,,"0000-00-00 00:00:00",Male,"0000-00-00 00:00:00","Bayero University","Computer Operator",,BUK,"University
Library",,Kano,,,NULL,,,,,,,156,234-064-0641026,,,,,,,,,,,,,23,Prof.,"Eula ",R,Abdella,,,NULL,,,"0000-00-00 00:00:00",,"0000-00-00 00:00:00","University of Benin","Vice
Chancellor","Information and Communication Technology Unit",,"P.M.B. 1154",,"Benin City",,,156,,,,,,156,234-52-602257,,,,,,,,,,,,24,,"Maire

Needle in a Haystack

- Most database transactions are not wholesale transfers of data
- Typical Web request generates one record
- Client/Server computing might transfer a dozen records at a time
- Querying an RDF namespace might only return a handful of records

Grandma Librarians Snuff Bunnies

- From the previous sample databases, I get the following results

  Grandma Librarians Snuff Bunnies

  (Awkward…)

- Using a text (or CSV) exporter, I got:

  Nickname, Object ¶ Grandma, Librarians ¶ Grandma, Snuff, ¶ Grandma Bunnies

- Retrieving all the elements of the Preferences Table, I get…

  Nickname, Preference, Object ¶ Grandma, Loves, Librarians ¶ Grandma, Hates, Snuff, ¶ Grandma, Likes, Bunnies

XML to the Rescue

- Since data can be shared in multiple ways, in microscopic quantities or massive collections of records, we want to be able to share data in a way that it can be mixed, reused, and repurposed while still carrying some metadata that describes the data (or metadata) within.
- With XML, the content gets “wrapped” in tags that describe it, its source, its meaning, its language, even security information.

Granny’s Opinions in XML

```
<OpinionatedPerson>
<Person>
    <nickname=”Grandma”>
</Person>

    <Object="Librarians">
    <Preference> Loves </Preference >
    </Object>

    <Object="Snuff">
    <Preference>Hates</Preference>
    </Object>

    <Object="Kittens">
    <Preference>Likes</Preference>
    </Object>
</OpinionatedPerson>
```

XML is a Major Standard for All Apps

- eXtensible Markup Language
- Similar to HTML, but focused on data storage and exchange
- Originally recognized as a standard by the W3C
- Birthed in the Web, but now can be found virtually everywhere
  - Word
  - Excel
  - Powerpoint

Is XML the Cat's Meow?

- XML data and documents can be exchanged, read, modified by most modern applications
- As you can see, however, it takes up a lot more space. Big transfers might go faster and better with traditional export formats
- One little glitch in the XML document may make it impossible to open. Very persnickety (there are probably a dozen errors in my small sample code above…)
- While it makes a great exchange tool, it's lousy as a database storage format because it's slow and cumbersome

Bottom Lines

- RDF, while not perfect, is terrifically flexible has become the go-to information management tool for a lot enterprises, not the least of which is libraries.
- XML has become the industry standard for packaging and sharing data
- Together RDF/XML make a good marriage
- There are competing standards being developed and adopted, but most of them are backwards compatible with RDF/XML