TAP: Towards a Web of Data

Bringing the Web to Programs

- The web has touched many facets of our lives
  - From buying books to getting driving directions
- The web has not changed how programs work
  - the web is not machine understandable

- Beginning to change with Web Services & XML
  - But most of the focus is on point to point data exchange
    - High set up cost, no network effects, takes 20 yrs to catch on
  - Action is in many to many exchanges, ala the Web

    … there are still problems to be solved for this

Some key problems TAP is attacking

- Core platform problems
  - Query languages/protocols
  - Integration, or the problem of names
  - Caching
  - Trust

- Applications
  - Search augmentation
  - PeopleNet
  - Internet Wet Lab

Query Languages

- Functional interfaces vs query interfaces
- Functional interface => SOAP
- Query Interface => ?
- General, expressive languages like SQL & XML Query inappropriate as public query interface … too expensive, too unpredictable to expose to everyone

- We need the equivalent of “HTTP GET”
  - Simple and stupid, but works remarkably well
- TAP’s answer : GetData

TAP Query Protocol : GetData

- DNS : GetHostByName(<host>) => ip addr.
- TAP: GetData(<resource>, <property>) => value
  - GetData(<Tori Amos>, birthplace) => <Newton, NC>
  - GetData(<Newton, NC>, temperature) => 57 F
  - GetData(<Newton, NC>, locatedIn) => <North Carolina>

- Publisher exposes data as a graph via GetData
- Client program uses GetData to query graph
- SOAP for over the wire transmission

Higher Level Services

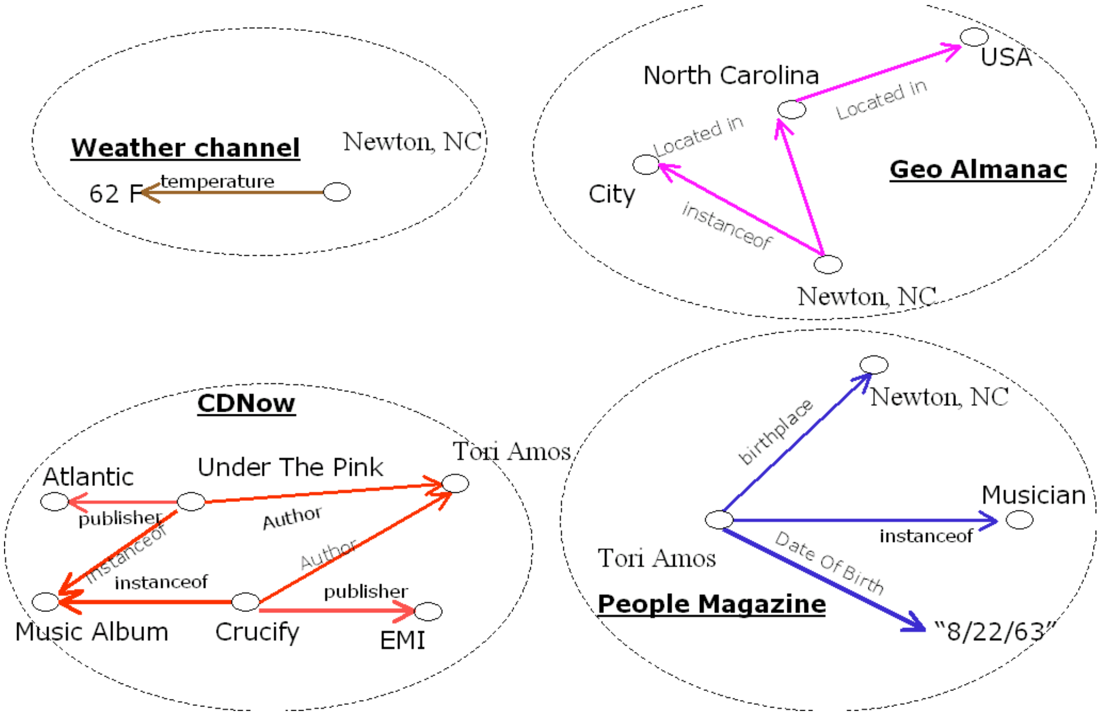
- Higher level services & applications can be built on top of GetData
  - Search engines
  - Complex Query engines that crawl and retrieve data and allow users to issue XML Query or DQL queries against data
  - Data Mining tools
- GetData's goal is to provide a low & easy cost of entry to publish & consume data from the data web

TAPache

- Apache based platform implementing GetData
- Exposing your data as a graph is as simple as publishing html --- simply put the file in the right directory and clients can access it via GetData
- Graph aggregation as analog of “index.html”
- Aggregations can be of local or remote graphs
- Clients libraries in java, C, perl, …
- Goal: To be the “BIND” for data

Integration, or the problem of names

- What we are getting now --- islands of XML from disparate web services, e.g., Tori Amos



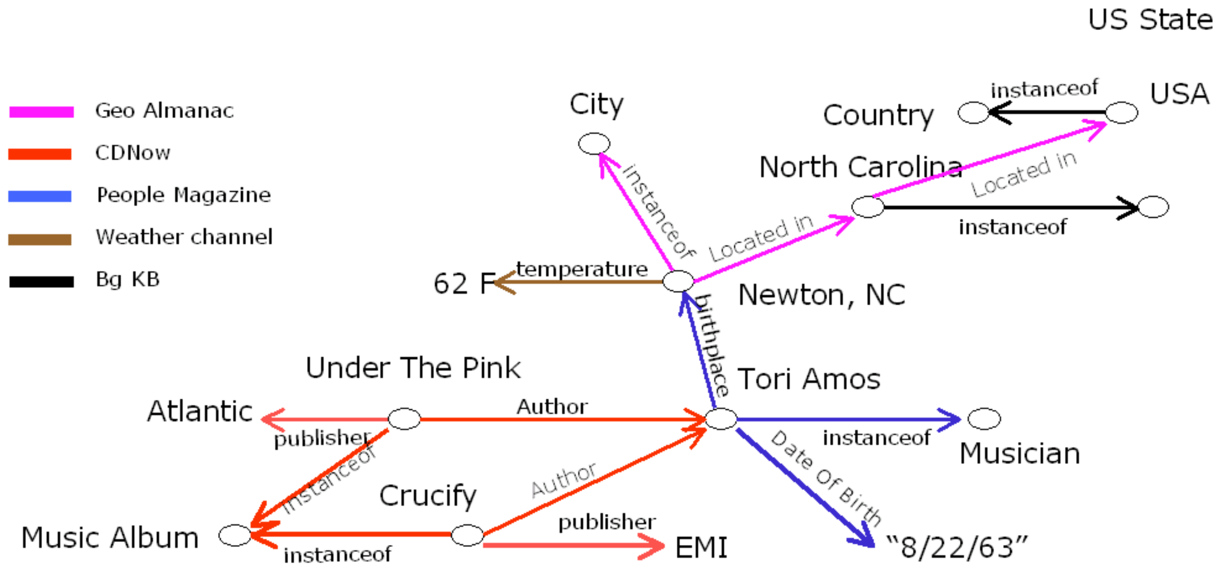
- Up to client program to put these chunks together

A key lesson from the web

- Current development path of the data web is analogous to pre-web hypertext systems and RDBMS today
  - More money is spent on systems integration than on Databases today.
- Lesson from the Web:
  - There is only one web!
  - Integration cannot be an after-thought
  - Has to be built into the core architecture
  - Integration is essential for many-to-many exchanges

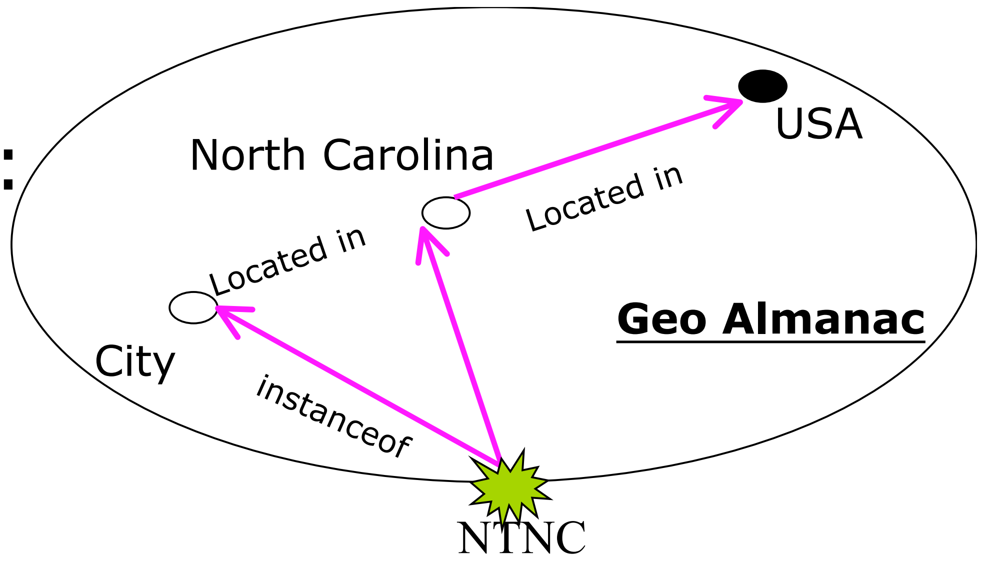
What a client would like to see …

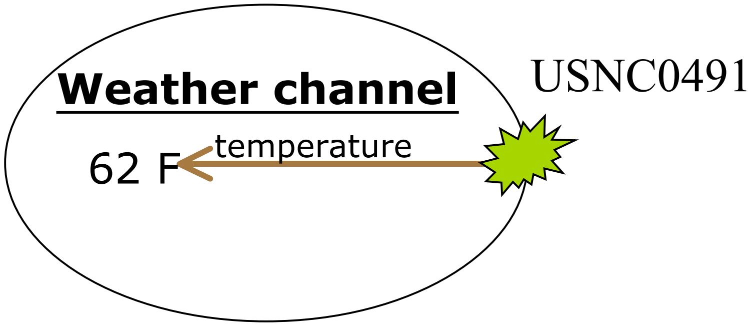
- Create a coherent data web from disparate chunks
- Client should see a schematically unified view

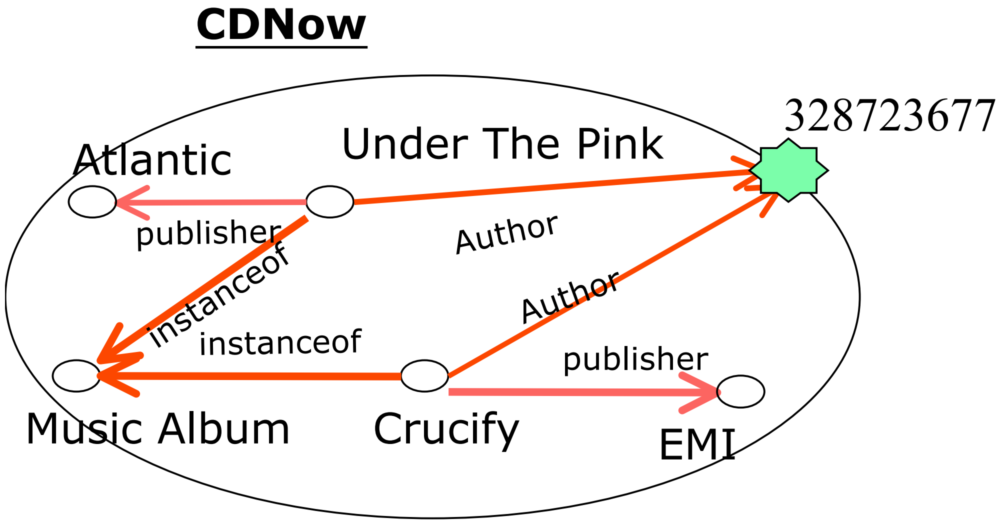


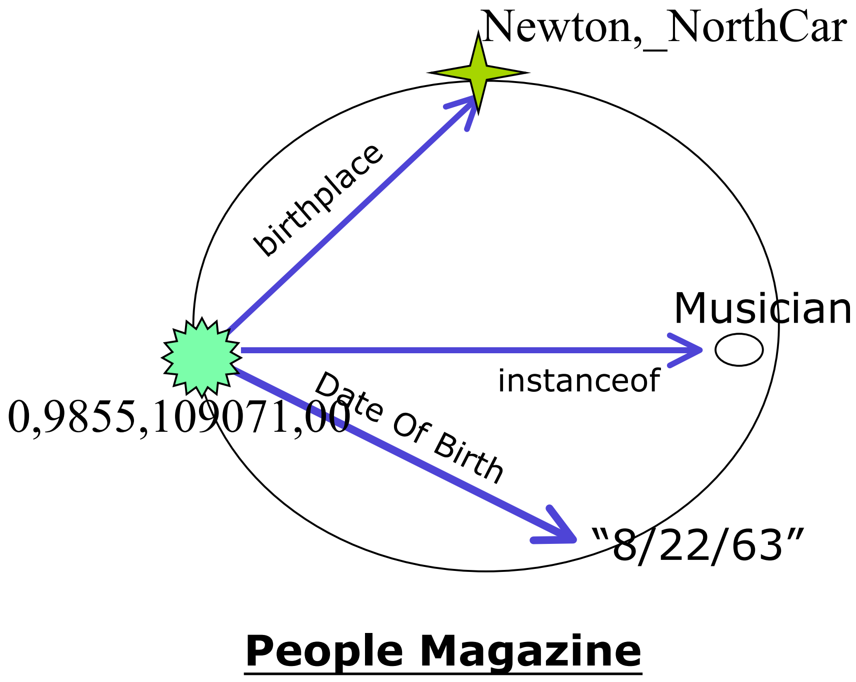
- Effectively make the web a giant distributed DB
- Just like DNS

The core of the problem: We get a mess like this





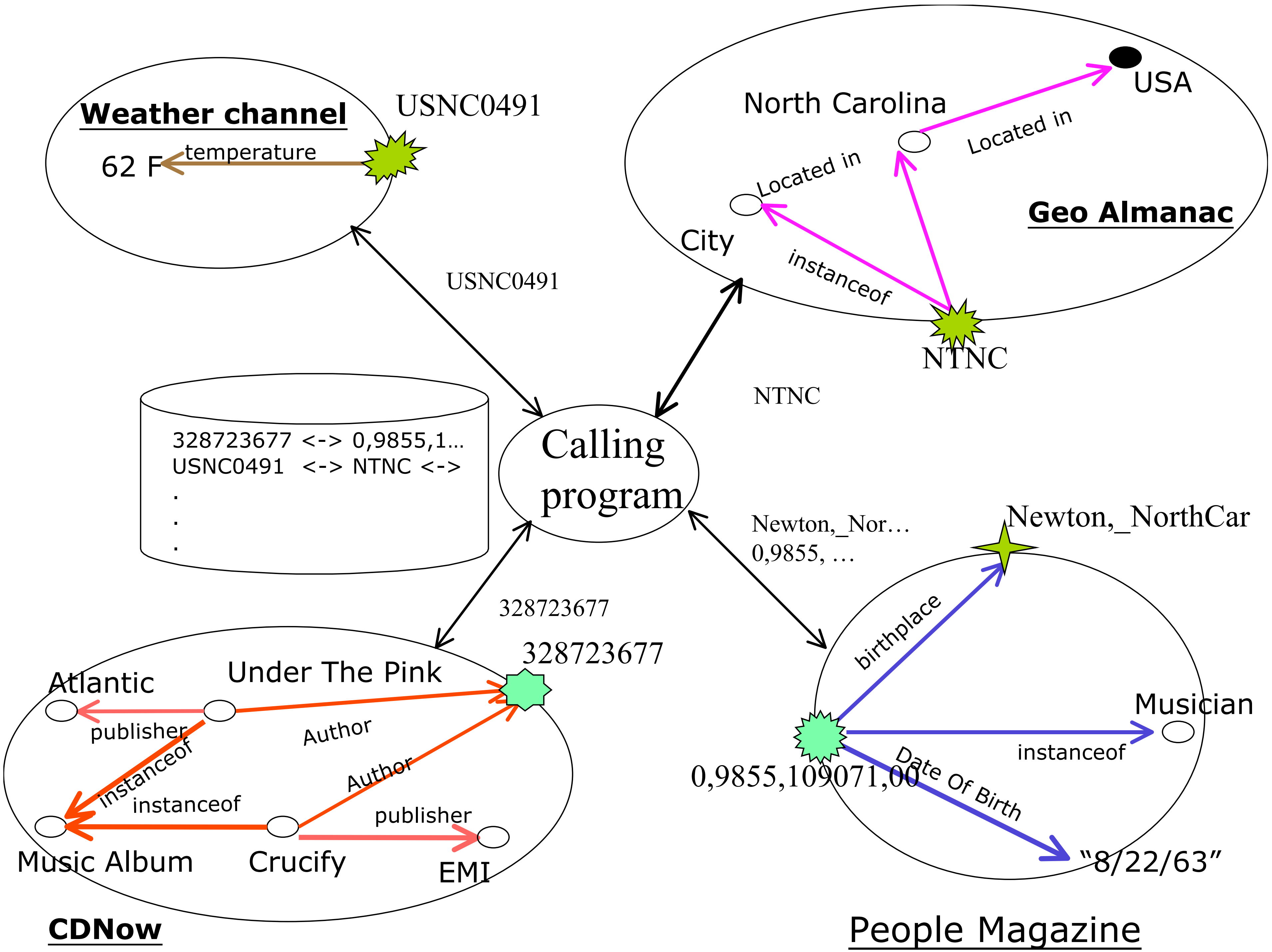




The Name Problem

- Names are crucial in information exchange
  - 2 parties cannot exchange information about an object without agreeing on how they are going to refer to it

- The Problem : too many names to keep track off!
  - No URN for <Newton, NC> or <Tori Amos>
  - Different sites have different names for the same thing!
  - URN efforts to date largely failures
  - Traditional Approach : Name-Mapping tables


Weather channel
USNC0491
62 F
temperature
USNC0491
North Carolina
USA
Located in
Located in
City
instanceof
Geo Almanac
NTNC
NTNC
328723677 <-> 0,9855,1…
USNC0491 <-> NTNC <->
.
.
.
Calling program
Newton,_Nor…
0,9855, …
Newton,_NorthCar
328723677
328723677
Atlantic
Under The Pink
publisher
Author
instanceof
instanceof
Author
publisher
Music Album
Crucify
EMI
CDNow
birthplace
Musician
instanceof
0,9855,109071,00
Date Of Birth
"8/22/63"
People Magazine

Semantic Negotiation

- Reference using descriptions
  - E.g., “A Musician whose firstName is ’Tori’ and whose lastName is ‘Amos’ and whose …”
  - Names are degenerate descriptions
    - Amzn:B000002UB2, CDNOW: 328723677

- Description based semantic negotiation
  - Don’t require globally unique names for everything if we can describe things using a starting vocabulary
  - Need a description language, starting vocabulary and negotiation mechanism
  - Bootstrapping some shared meaning into more shared meaning

Descriptions

- Description of an object = any RDF graph involving that object
- A description is Discriminant in a database if it uniquely identifies an object
- Semantic Negotiation is the process of identifying a shared Discriminant description for the object involved
- Assumes object is present in both DBs
- Works not just for individuals, but also for classes and properties

Loose Coupling

- Description based references are a form of loose coupling
- Loose Coupling implies the possibility of a failure to couple
- Failure modes:
  - Ambiguity … negotiate to resolve ambiguity
  - No shared Discriminant description
    - Not enough shared vocabulary
    - Literals don’t match
  - Domain skew

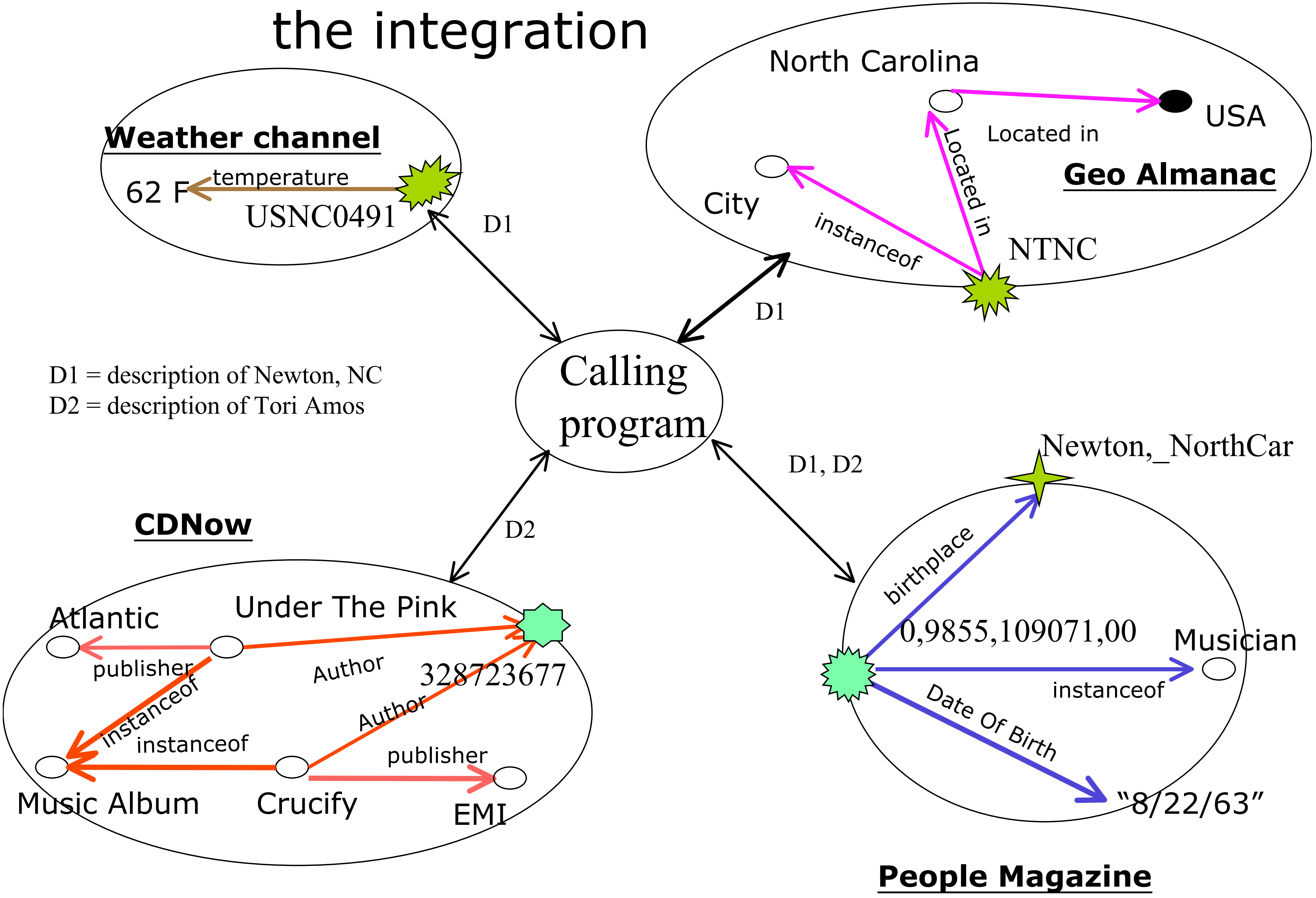
Description based References and GetData

- The GetData protocol:
  - GetData(Resource Description, arc-label)
  - GetData(<Tori Amos>, birthplace)
  - GetData(RDF Description of Tori Amos, birthplace)

- The contract:
  - Expose your data as a Graph
  - Map incoming descriptions to nodes in your graph
  - In return, your data is now integrated into the global semantic web

- Plays the role that URLS play for the HTML web

The vision: descriptions choreograph the integration



D1 = description of Newton, NC
D2 = description of Tori Amos

Infrastructure: Kernel Vocabulary

- Provides vocabulary for descriptions
- Purpose is to provide the infrastructure for constructing descriptions with which programs can refer to things
- “A Musician whose firstName is ’Tori’ and whose lastName is ‘Amos’ and whose
- It doesn’t reside anywhere : it’s a specification

- As of now, TAP’s kernel vocabulary is adequate to describe at least 70% of Amazon’s inventory

TAP Caching

- DNS style caching is too error-prone
- HTTP style transparent caching is too conservative and based on a worst case scenario
- Solution path … pull-push caching
  - Cache is not transparent
  - Expected TTL, with option to provide updates

TAP Registries & Trust

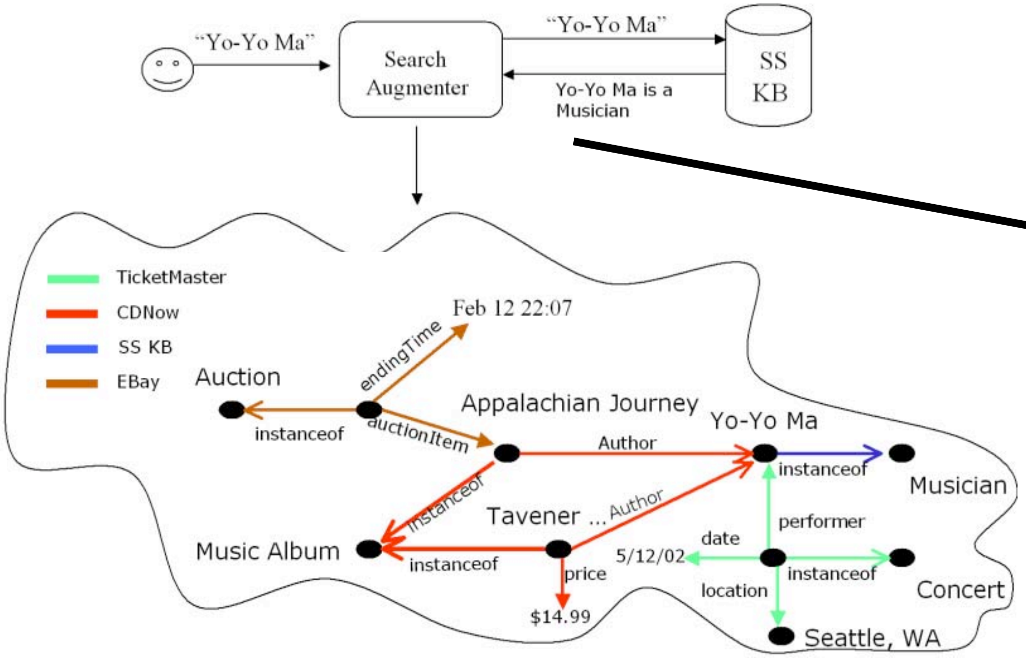
- Registry: UDDI + descriptions of which properties of which kinds of objects

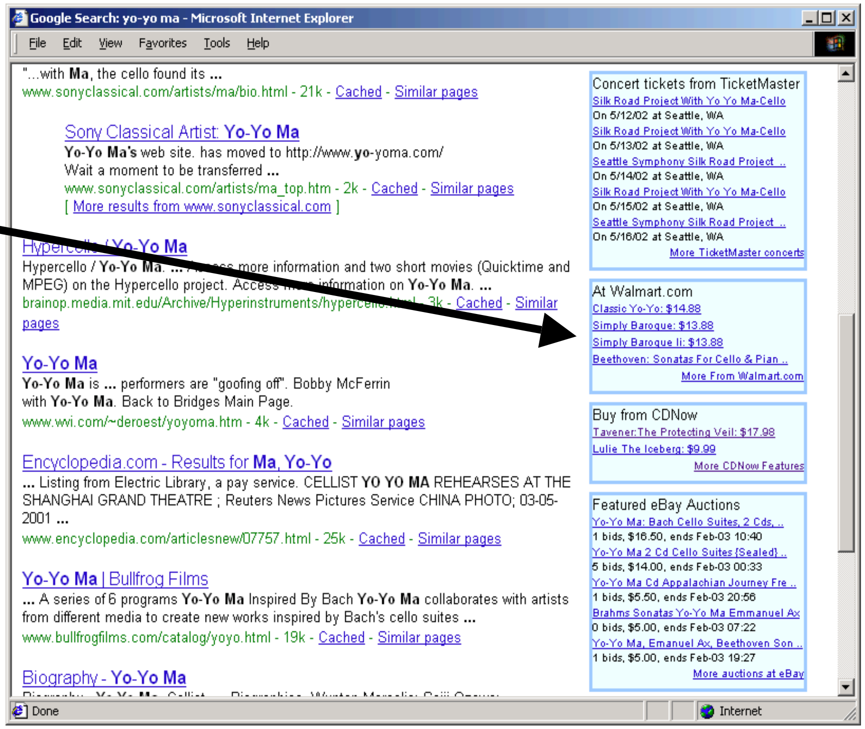
- How do machines know whose data to trust?
- Centralized clearing house model: a la Yahoo!
- Decentralized Web of Trust model: a la Epinions
- Each TAPache server can also serve as a proxy which maintains a registry of trusted data sources and cache
- Each TAPache server can be told about one or more trusted peers, who can be asked for their registry entries

Applications

- Good infrastructures have waves of applications
  - WWW : home pages, portals, ecommerce, …
  - DNS : email, telnet, ftp, gopher, … WWW
- Enterprise applications drive bilateral data sharing … already taking place
- Semantic Search: Adding Semantics to Search
  - Semantics based Search Augmentation
  - Activity based search
- Internet Wet Lab

Semantic Web Application: Semantic Search





Google Search: yo-yo ma - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Done
Internet

"...with **Ma**, the cello found its **...**
www.sonyclassical.com/artists/ma/bio.html - 21k - Cached - Similar pages

Sony Classical Artist: **Yo-Yo Ma**
Yo-Yo Ma's web site. has moved to http://www.**yo**-yoma.com/
Wait a moment to be transferred **...**
www.sonyclassical.com/artists/ma_top.htm - 2k - Cached - Similar pages
[More results from www.sonyclassical.com]

Hypercello / **Yo-Yo Ma**
Hypercello / **Yo-Yo Ma**. **...** Access more information and two short movies (Quicktime and MPEG) on the Hypercello project. Access more information on **Yo-Yo Ma**. **...**
brainop.media.mit.edu/Archive/Hyperinstruments/hypercello.html - 3k - Cached - Similar pages

Yo-Yo Ma
Yo-Yo Ma is **...** performers are "goofing off". Bobby McFerrin
with **Yo-Yo Ma**. Back to Bridges Main Page.
www.wvi.com/~deroest/yoyoma.htm - 4k - Cached - Similar pages

Encyclopedia.com - Results for **Ma, Yo-Yo**
... Listing from Electric Library, a pay service. CELLIST **YO YO MA** REHEARSES AT THE SHANGHAI GRAND THEATRE ; Reuters News Pictures Service CHINA PHOTO; 03-05-2001 **...**
www.encyclopedia.com/articlesnew/07757.html - 25k - Cached - Similar pages

Yo-Yo Ma | Bullfrog Films
... A series of 6 programs **Yo-Yo Ma** Inspired By Bach **Yo-Yo Ma** collaborates with artists from different media to create new works inspired by Bach's cello suites **...**
www.bullfrogfilms.com/catalog/yoyo.html - 19k - Cached - Similar pages

Biography - **Yo-Yo Ma**

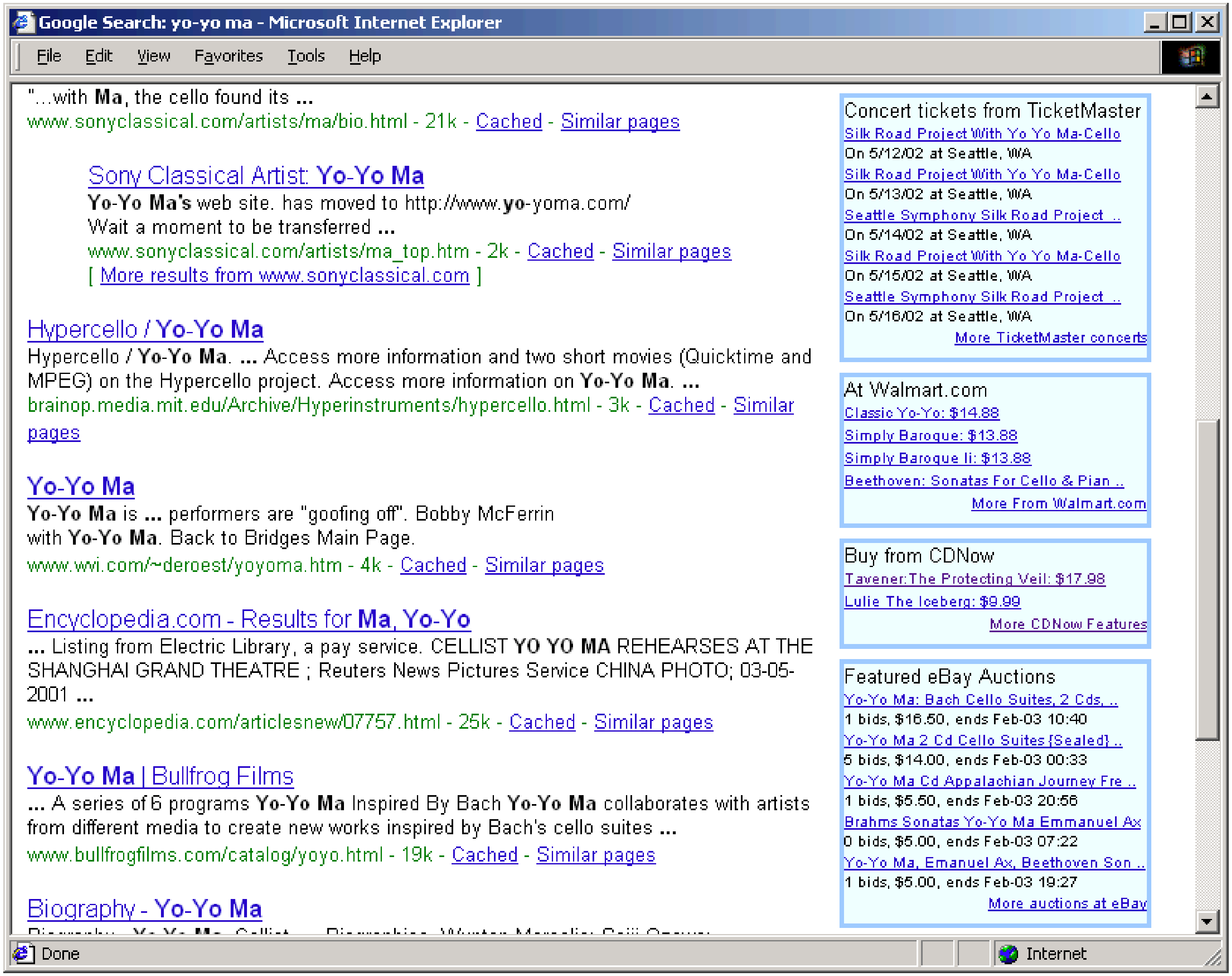
Concert tickets from TicketMaster
Silk Road Project With Yo Yo Ma-Cello
On 5/12/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/13/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/14/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/15/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/16/02 at Seattle, WA
More TicketMaster concerts

At Walmart.com
Classic Yo-Yo: $14.88
Simply Baroque: $13.88
Simply Baroque Ii: $13.88
Beethoven: Sonatas For Cello & Pian ..
More From Walmart.com

Buy from CDNow
Tavener:The Protecting Veil: $17.98
Lulie The Iceberg: $9.99
More CDNow Features

Featured eBay Auctions
Yo-Yo Ma: Bach Cello Suites, 2 Cds, ..
1 bids, $16.50, ends Feb-03 10:40
Yo-Yo Ma 2 Cd Cello Suites {Sealed} ..
5 bids, $14.00, ends Feb-03 00:33
Yo-Yo Ma Cd Appalachian Journey Fre ..
1 bids, $5.50, ends Feb-03 20:56
Brahms Sonatas Yo-Yo Ma Emmanuel Ax
0 bids, $5.00, ends Feb-03 07:22
Yo-Yo Ma, Emanuel Ax, Beethoven Son ..
1 bids, $5.00, ends Feb-03 19:27
More auctions at eBay

Search Augmentation Example



Google Search: yo-yo ma - Microsoft Internet Explorer

"...with **Ma**, the cello found its **...**
www.sonyclassical.com/artists/ma/bio.html - 21k - Cached - Similar pages

Sony Classical Artist: **Yo-Yo Ma**
Yo-Yo Ma's web site. has moved to http://www.**yo**-yoma.com/
Wait a moment to be transferred **...**
www.sonyclassical.com/artists/ma_top.htm - 2k - Cached - Similar pages
[More results from www.sonyclassical.com]

Hypercello / **Yo-Yo Ma**
Hypercello / **Yo-Yo Ma**. **...** Access more information and two short movies (Quicktime and MPEG) on the Hypercello project. Access more information on **Yo-Yo Ma**. **...**
brainop.media.mit.edu/Archive/Hyperinstruments/hypercello.html - 3k - Cached - Similar pages

Yo-Yo Ma
Yo-Yo Ma is **...** performers are "goofing off". Bobby McFerrin
with **Yo-Yo Ma**. Back to Bridges Main Page.
www.wwi.com/~deroest/yoyoma.htm - 4k - Cached - Similar pages

Encyclopedia.com - Results for **Ma, Yo-Yo**
... Listing from Electric Library, a pay service. CELLIST **YO YO MA** REHEARSES AT THE SHANGHAI GRAND THEATRE ; Reuters News Pictures Service CHINA PHOTO; 03-05-2001 **...**
www.encyclopedia.com/articlesnew/07757.html - 25k - Cached - Similar pages

Yo-Yo Ma | Bullfrog Films
... A series of 6 programs **Yo-Yo Ma** Inspired By Bach **Yo-Yo Ma** collaborates with artists from different media to create new works inspired by Bach's cello suites **...**
www.bullfrogfilms.com/catalog/yoyo.html - 19k - Cached - Similar pages

Biography - **Yo-Yo Ma**

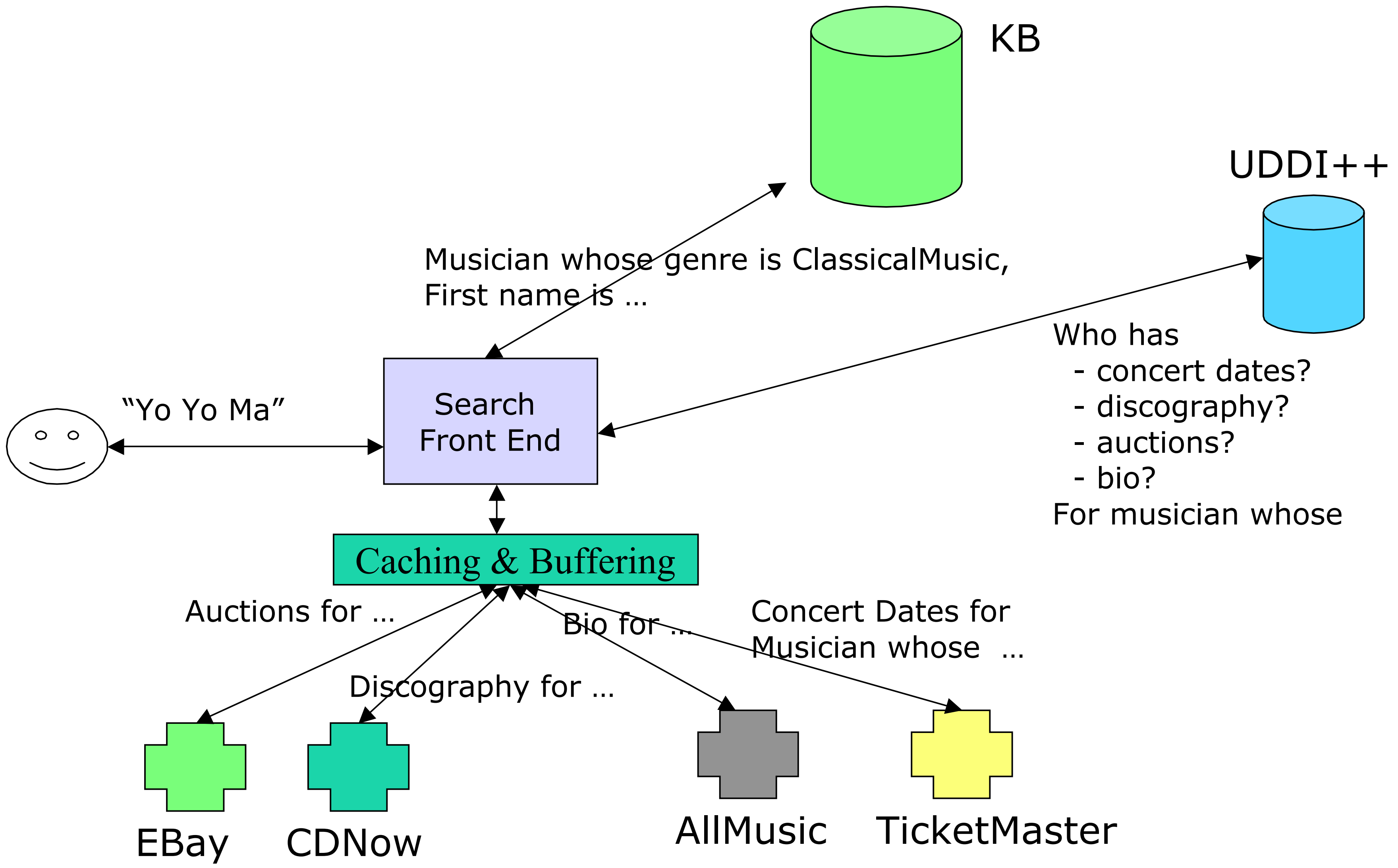
Concert tickets from TicketMaster
Silk Road Project With Yo Yo Ma-Cello
On 5/12/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/13/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/14/02 at Seattle, WA
Silk Road Project With Yo Yo Ma-Cello
On 5/15/02 at Seattle, WA
Seattle Symphony Silk Road Project ..
On 5/16/02 at Seattle, WA
More TicketMaster concerts

At Walmart.com
Classic Yo-Yo: $14.88
Simply Baroque: $13.88
Simply Baroque Ii: $13.88
Beethoven: Sonatas For Cello & Pian ..
More From Walmart.com

Buy from CDNow
Tavener:The Protecting Veil: $17.98
Lulie The Iceberg: $9.99
More CDNow Features

Featured eBay Auctions
Yo-Yo Ma: Bach Cello Suites, 2 Cds, ..
1 bids, $16.50, ends Feb-03 10:40
Yo-Yo Ma 2 Cd Cello Suites {Sealed} ..
5 bids, $14.00, ends Feb-03 00:33
Yo-Yo Ma Cd Appalachian Journey Fre ..
1 bids, $5.50, ends Feb-03 20:56
Brahms Sonatas Yo-Yo Ma Emmanuel Ax
0 bids, $5.00, ends Feb-03 07:22
Yo-Yo Ma, Emanuel Ax, Beethoven Son ..
1 bids, $5.00, ends Feb-03 19:27
More auctions at eBay

How the Semantic Infrastructure gets used in Semantic Search



TAP KBs for Semantic Search

- Large Knowledge Base of specific musicians, cities, athletes, …
  - Currently covers about 20% of search terms at DMOZ
  - Built in a largely automated fashion
    - Scrapers for free data sources
    - Simple noun phrase analysis of news articles
      - AP, Reuters, …
- Scrapers for important sites to bootstrap
- KB also helps bootstrap the semantic web

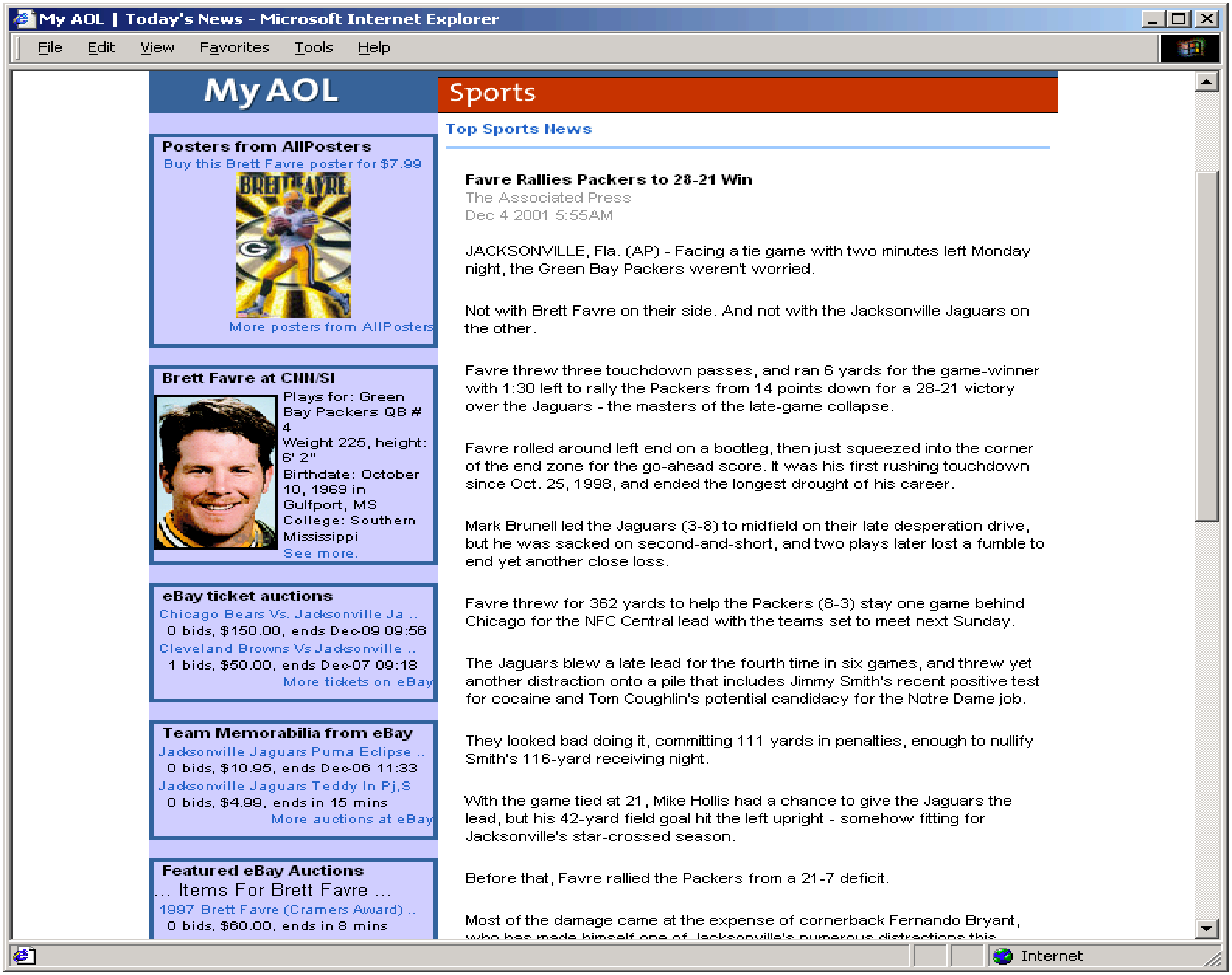
KB Coverage Today

- Music
  - Musicians, instr., styles
- Movies
  - Movies, actors, tv-shows
- Authors
  - Top authors, classic books,
- Sports
  - Athletes, sports, sports teams, equipment
- Autos
  - Auto models, motorcycles, .
- Companies
  - Fortune 500
- Home Appliances
  - Types, brands
- Toys
  - Types, brands
- Baby products
  - Types, brands
- Places
  - Countries, cities, tourist attractions, …
- Consumer electronics
  - Audio/Video, Communication
  - Game : consoles, titles, …
- Health
  - Diseases, Drugs, …

Semantic Site Search

- Semantic Search useful not just for internet wide search, but also for site search
- Same principles as internet-wide search
- KBs created for searching related individual sites can be shared between sites
- These KBs feed into global semantic web
- Example: Semantic Search for www.w3.org

Application : Sidebar for news articles



Top Sports News

Posters from AllPosters

Buy this Brett Favre poster for $7.99

More posters from AllPosters

Brett Favre at CNN/SI

Plays for: Green Bay Packers QB # 4
Weight 225, height: 6' 2"
Birthdate: October 10, 1969 in Gulfport, MS
College: Southern Mississippi
See more.

eBay ticket auctions

Chicago Bears Vs. Jacksonville Ja ..
0 bids, $150.00, ends Dec-09 09:56
Cleveland Browns Vs Jacksonville ..
1 bids, $50.00, ends Dec-07 09:18
More tickets on eBay

Team Memorabilia from eBay

Jacksonville Jaguars Puma Eclipse ..
0 bids, $10.95, ends Dec-06 11:33
Jacksonville Jaguars Teddy In Pj,S
0 bids, $4.99, ends in 15 mins
More auctions at eBay

Featured eBay Auctions

... Items For Brett Favre ...
1997 Brett Favre (Cramers Award) ..
0 bids, $60.00, ends in 8 mins

Favre Rallies Packers to 28-21 Win
The Associated Press
Dec 4 2001 5:55AM

JACKSONVILLE, Fla. (AP) - Facing a tie game with two minutes left Monday night, the Green Bay Packers weren't worried.

Not with Brett Favre on their side. And not with the Jacksonville Jaguars on the other.

Favre threw three touchdown passes, and ran 6 yards for the game-winner with 1:30 left to rally the Packers from 14 points down for a 28-21 victory over the Jaguars - the masters of the late-game collapse.

Favre rolled around left end on a bootleg, then just squeezed into the corner of the end zone for the go-ahead score. It was his first rushing touchdown since Oct. 25, 1998, and ended the longest drought of his career.

Mark Brunell led the Jaguars (3-8) to midfield on their late desperation drive, but he was sacked on second-and-short, and two plays later lost a fumble to end yet another close loss.

Favre threw for 362 yards to help the Packers (8-3) stay one game behind Chicago for the NFC Central lead with the teams set to meet next Sunday.

The Jaguars blew a late lead for the fourth time in six games, and threw yet another distraction onto a pile that includes Jimmy Smith's recent positive test for cocaine and Tom Coughlin's potential candidacy for the Notre Dame job.

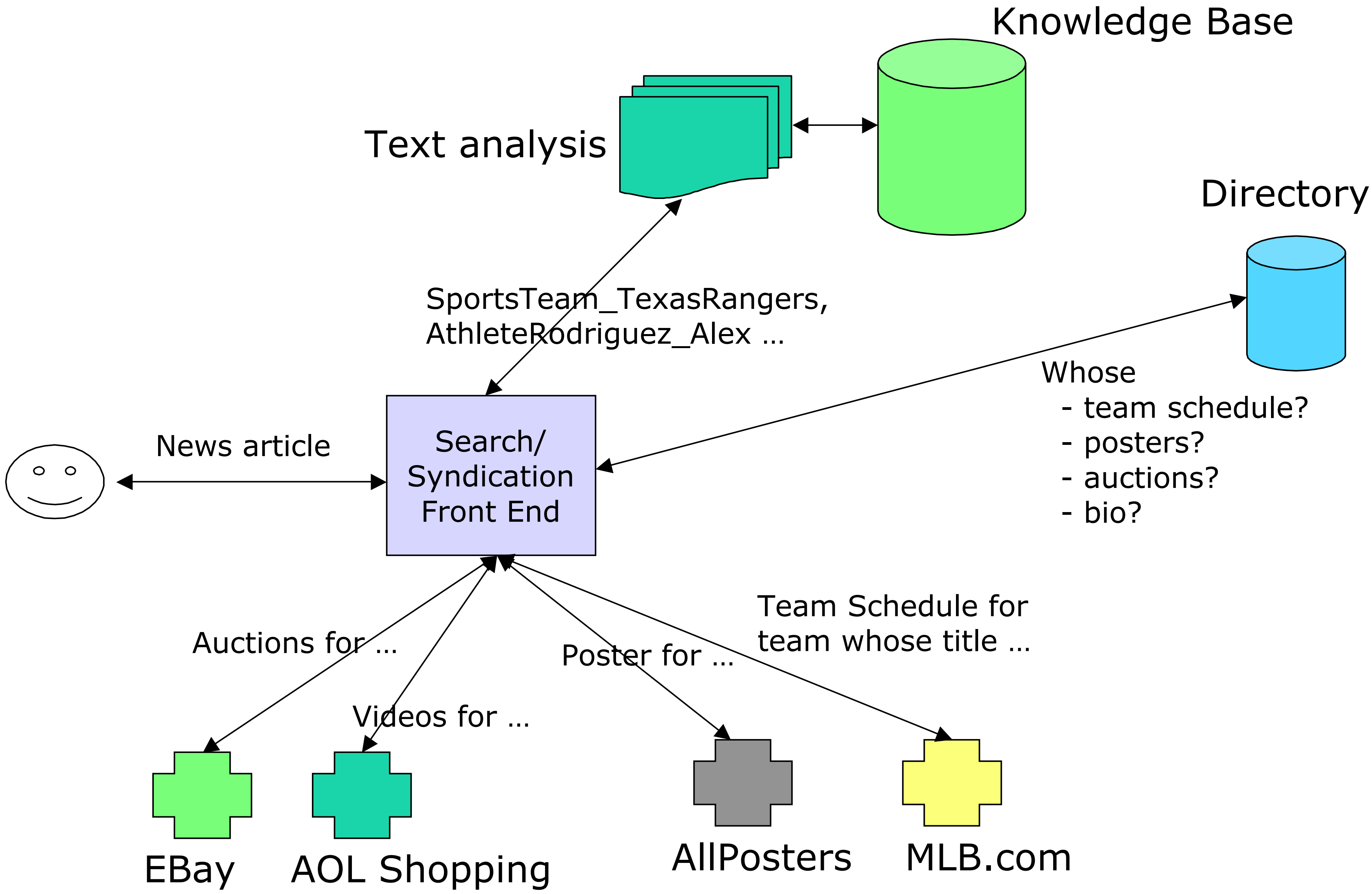
They looked bad doing it, committing 111 yards in penalties, enough to nullify Smith's 116-yard receiving night.

With the game tied at 21, Mike Hollis had a chance to give the Jaguars the lead, but his 42-yard field goal hit the left upright - somehow fitting for Jacksonville's star-crossed season.

Before that, Favre rallied the Packers from a 21-7 deficit.

Most of the damage came at the expense of cornerback Fernando Bryant, who has made himself one of Jacksonville's numerous distractions this

Sidebar for News Articles



Application: PeopleNet

- What the graph contains:
  - Nodes correspond to people, organizations, projects, papers, …
  - Many kinds of relations, including topical trust relations between people
  - Many different sites, I.e., whole thing is distributed
  - Site specific user-ids

- Applications:
  - Distributed citeseer
  - Link recommendation system
  - Build your own …

Application: Internet Wet Lab

- In many sciences, more data will be produced in the next 2 years than exists today
- Increasingly, research consists of writing programs that mine this data
- Data is isolated as islands in different labs
- Data from one lab not easily available to programs in another lab
- Imagine a single virtual net-wide “database” containing all this experimental data
- Example : Clinical Trial Data

Questions?