

It's the A-box, stupid!

(free after Carville/Clinton)

Frank van Harmelen
Vrije Universiteit Amsterdam



vrije Universiteit

Creative Commons License:
allowed to share & remix,
but must attribute & non-commercial



Semantic Web

News Headlines

ICT NEWS



WEEKLY
NEWS
QUIZ

GOOGLE TO HARVEST RDF-A

Search giant goes semantic

Morris Center for the Performing Arts, 27579 Cro�푸ield Drive in Rosedale, on Saturday, March 20 at 7 and 9 p.m. Tickets are \$35 for adults and \$15 for students for the evening performance and \$30 for adults and \$15 for students at the matinee performance. For reservations, call 544-0403.

- * UPCOMING — The Palos Verdes Peninsula Unified School District and Friends of School Music host the 15th Palos Verdes Elementary Choral Festival on March 23, 24 and 25 at the Morris Center for the Performing Arts, 27570 Crockerfield Drive in RHP. All shows begin at 7:30 p.m. For tickets, call the Morris box office at 344-0401.
 - * ONGOING — The Distinctive Edge, 29070 S. Western Ave., Suite 113 in RHP, continues "Third Timer's a Charm," an exhibit of 50 collages by artist Steve Jacobson, through March 30. For gallery hours, call 833-3613.
 - * ONGOING — "Natural Treasures" exhibition contin-

Two years ago, U.S. Navy personnel and their families assigned to the Atago Navy base, home of the U.S. 7th Fleet, were invited to experience what

experience when Terry Fleming at his local Irish-American band, invited them on St. Patrick's Day. Fleming and the other five members of Invictus were delighted and honored to be able to go to Japan and lift the spirits, if only for a few hours, of the Navy personnel and their families. For the third year in a row — a local

For the third year in a row, Fleming — a local insurance broker in Rolling Hills by day and an entertainer by night — and the band travel to entertain the Navy men, women and families at various bases throughout Japan.

Planning, the leader of the band on saxophone and harmonica, actually is the only member of the band from Ireland. Other members include lead singer Julie Delaney, a civil engineer in Newport Beach; Terry Doyle, guitar in Newport Beach; Terry Doyle, guitar, a news director with CBS-TV in San Francisco; Dennis Doyle, Celtic harpist, a professor at Glendale College; Kevin West, keyboards and bass.

Even though far from home, the
Army bar — situated a few hours

teacher and assistant director of the Orange County Symphony; and Mike Tilney, bass, a computer engineer. The band has been playing the length and breadth of California for the past 25 years. They have played at picnics, weddings, birthdays and on occasions where there was little excuse for throwing a party.

Planning says it was by coincidence the band got the opportunity to travel to Japan. Another band was unable to travel at the last minute and so he and his band were offered the opportunity to go in their place.

With some trepidation they made their first try and with the same whining response they received a hasty, airy kiss they had never quickly allowed. On a damp St. Patrick's Day, hundreds of thousands of

many shades of green, whirled in
mazy circles out and danced in
the storm. As the evening wore on,
in the darkness were unobscured
try their hand or foot at the Irish game
with much encouragement from
the band.

Even though far from home,
Amp has — visited a few bars

width of Tokyo — is the home away from home, with its lush green rolling landscape and its multitude of cherry blossom trees. "Yes," Professor Yamada says, "you can travel around Japan without ever leaving Tokyo."

"...we were struck by the consummate skill and dedication of our men and women in uniform as they played their part in protecting and serving in an ever challenging and heroic world."

A couple of the families even took time out from their busy schedules to host Terry and the band members. They treated them to a guided tour of the base and accompanied them on a few exciting trips off the base, visiting beautiful ancient temples, monasteries and several giant Buddhas.

hand came across what they assumed to be downtown Tokyo, there was a very rare sight, a place called "Scruffy Murphy's" an Irish pub located in the heart of a bustling downtown. Upon checking the establishment out, they discovered a real authentic Irish Pub with excellent Guinness and good pub grub. It also happened to be open mid night, so the band members took over the stage.

and entertained the locals for a few hours. It turned out that he has filled houses. It was just one of many establishments in the city.

A special bond developed between the band members and their families and already exchange visits here and there in the same families were as long as in the United States.

thebridge

卷之三

1000 1000

HOME RENDERING TO PRINTING

PRINT
ON DEMAND



PRINT
ON DEMAND



PRINT
ON DEMAND

photocraftchef.com



LarKC Webblog



[Entries \(RSS\)](#) and [Comments \(RSS\)](#)

« Zahdo You are the network
Twine, the debate pages »

Microsoft releases open-source semantic plugin for Word

(by Frank van Harmelen)

The following could well be a very significant development (dare I use the word "watershed"?). It's the first time that I see the words "ontology" and "semantic web" in a Microsoft press release, and this is Microsoft linking the Semantic Web to one of its flagship products, under open source license no less!

The significance of this is not per se in the performance of the particular plugin (various other pieces of software already aim to do similar things, even in the same domain), but IMO the significance lies in the fact that Microsoft wants to be seen to be doing this.

In an official press release, Microsoft writes:



Image via Wikipedia

DISCOVER. PARTICIPATE. ENGAGE.

Search Data.gov catalog by category, agency, or both

All Categories

All Agencies

toxic releases

recent earthquakes

crime statistics

assaults on police

social benefits

unemployment rates

consumer expenditure

consumer price index

tornado reports

trade statistics

river elevations

energy consumption

FEATURED DATA

ENERGY INFO

Residential Energy Co



The Daily C

Sunday, August 30, 2006

TBL advising Gordon Brown

Creator of the web Sir Tim Berners-Lee has been appointed by the government to lead a review of how the internet can be used to open up access to official information.

Prime Minister Gordon Brown said Berners-Lee was to oversee a project that would create a single portal where UK residents could access public data held by the government.

In a speech, Brown said: "So that government information is accessible for the widest

creation of the world wide web, to help us drive the opening up of access to government data in the web over the coming months.

"All MPs' past and future expenses should and will be published on the internet in the next few days. Second home claims submitted by MPs from all sides of the House over the last four years must be scrutinised by an independently led panel. This will ensure repayment where it is necessary, and lead to discipline, where there have been inappropriate claims," he

It is
total
thin
disc
mpa
retu
Son

Music BETA[BBC Music](#) > [Artists](#) > Yeah Yeah Yeahs

Yeah Yeah Yeahs

Group. Formed 2000.PLAYED MOST ON
BBC RADIO

Now On The BBC

[Yeah Yeah Yeahs - Later with Jools Holland](#)

Watch the New York trio performing Heads Will Roll on Later...



Latest News Stories



NEWS FROM THE BBC

[Heroes to Zero](#)

Fri 20 Mar 2009 09:30

Leaks, eBay and kittens - life with Yeah Yeah Yeahs

[Listen Plus Details](#)

Played By

Since December 2008

**Nemone**

BBC 6 Music

Music, entertainment, games and guests with Nemone

```
<rdf:RDF>
<rdf:Description rdf:about="/music/artists/584c04d2-4acc-491b-8a0a-e63133f4bfc4.rdf"
  <rdfs:label>Description of the artist Yeah Yeah Yeahs</rdfs:label>
  <foaf:primaryTopic rdf:resource="/music/artists/584c04d2-4acc-491b-8a0a-e63133f4bfc4.rdf#artist">
</rdf:Description>
<mo:MusicArtist rdf:about="/music/artists/584c04d2-4acc-491b-8a0a-e63133f4bfc4#artist">
  <rdf:type rdf:resource="http://purl.org/ontology/mo/MusicGroup"/>
  <foaf:name>Yeah Yeah Yeahs</foaf:name>
  <ov:sortLabel>Yeah Yeah Yeahs</ov:sortLabel>
  <bio:event>
    <bio:Birth><bio:date rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1981-01-01T00:00:00Z</bio:date>
  </bio:event>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/Yeah_Yeah_Yeahs"/>
<mo:image rdf:resource="/music/images/artists/7col_in/584c04d2-4acc-491b-8a0a-e63133f4bfc4.jpg"/>
<foaf:page rdf:resource="/music/artists/584c04d2-4acc-491b-8a0a-e63133f4bfc4.html"/>
<mo:musicbrainz rdf:resource="http://musicbrainz.org/artist/584c04d2-4acc-491b-8a0a-e63133f4bfc4"/>
<foaf:homepage rdf:resource="http://www.yeahyeahyeahs.com/"/>
<mo:wikipedia rdf:resource="http://en.wikipedia.org/wiki/Yeah_Yeah_Yeahs"/>
<mo:myspace rdf:resource="http://www.myspace.com/yeahyeahyeahs"/>
<mo:member rdf:resource="/music/artists/a1439b8d-672a-446f-a7ff-6f09d68254b3#artist"/>
<mo:member rdf:resource="/music/artists/14d44067-99c2-4f77-b58b-138f0b6911fa#artist"/>
<mo:member rdf:resource="/music/artists/20dc35ec-6cc1-4c66-98a3-4a6116cb3869#artist"/>
...</mo:MusicArtist>
```

```
<foaf:made>
<mo:Record>
  <dc:title>It's Blitz!</dc:title>
  <mo:musicbrainz rdf:resource="http://musicbrainz.org/release/9c4177fe-bdce-4f9d-ab...
    <rev:hasReview rdf:resource="/music/reviews/hnp2#review"/>
  </mo:Record>
</foaf:made>
.....
<mo:MusicArtist rdf:about="/music/artists/a1439b8d-672a-446f-a7ff-6f09d68254b3#arti...
  <foaf:name>Brian Chase</foaf:name>
</mo:MusicArtist>

<mo:MusicArtist rdf:about="/music/artists/14d44067-99c2-4f77-b58b-138f0b6911fa#arti...
  <foaf:name>Karen O</foaf:name>
</mo:MusicArtist>

<mo:MusicArtist rdf:about="/music/artists/20dc35ec-6cc1-4c66-98a3-4a6116cb3869#arti...
  <foaf:name>Nick Zinner</foaf:name>
</mo:MusicArtist>
</rdf:RDF>
```

Yeah Yeah Yeahs Fever To Tell Review

Album. Released 29 April 2003.

Discography information comes from [MusicBrainz](#). You can add or edit information about Fever to Tell at [musicbrainz.org](#).



BBC Review

This album is like casual sex: messy, with plenty of cheap thrills. But when you wake...

Nick Reynolds 2003-05-07

This album is like casual sex: messy, with plenty of cheap thrills. But when you wake up in the morning you think "Did I really do that?" You ask yourself "Is this just a fling? Or is there anything here that will last?"

When the Yeah Yeah Yeahs stormed into the UK from New York last year, everybody cheered. In a so called rock revolution that largely consisted of the usual excess of testosterone and bad heavy metal they were a breath of fresh air. Their stripped down guitar/drums/vocal sound brought some much needed qualities like brevity, intelligence and even wit. Lead singer Karen O is exactly what you want from a rock heroine: feisty, rude, in control but out of control. On the current single "Date With The Night" she rips up the town like a female Godzilla on heat.

This debut album is louder and heavier than their brilliant 5 track CD of last year. There's no denying the impact of the sexy swamp metal of "Cold Light". Or the delirious confusion of "No No No", a track which highlights the bands refreshing ability to try different ideas. You can't imagine the likes of The Datsuns, God forbid, attempting a rock/reggae hybrid but the Yeahs have no problem with it.

Tracks

- 1 Rich
- 2 Date With the Night
- 3 Man
- 4 Tick
- 5 Black Tongue

Links

[Wikipedia article on Fever to Tell](#)

Links come from [MusicBrainz](#). You can add or edit information about Yeah Yeah Yeahs at [musicbrainz.org](#). Find out more about our use of this data.

Artist Played By



Nemone
BBC 6 Music



Steve Lamacq
BBC 6 Music



Shaun Keaveny
BBC 6 Music

```
<rdf:RDF>
- <rdf:Description rdf:about="/music/reviews/h24h.rdf">
  <rdfs:label>Description of a review of Fever To Tell</rdfs:label>
  <foaf:primaryTopic rdf:resource="/music/reviews/h24h#review"/>
</rdf:Description>
- <rev:Review rdf:about="/music/reviews/h24h#review">
  <rev:title>Fever To Tell</rev:title>
  - <rdfs:label> Review of Fever To Tell - Yeah Yeah Yeahs, by Nick Reynolds</rdfs:label>
    <rev:createdOn rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2003
  - <foaf:primaryTopic>
    - <mo:Record>
      <dc:title>Fever to Tell</dc:title>
      <owl:sameAs rdf:resource="http://dbpedia.org/resource/Fever_to_Tell"/>
      <mo:musicbrainz rdf:resource="http://musicbrainz.org/release/f4783344-6746-4938
        <foaf:maker rdf:resource="/music/artists/584c04d2-4acc-491b-8a0a-e63133f4bfc4#"
      </mo:Record>
    </foaf:primaryTopic>
    - <rev:reviewer>
      - <foaf:Person><foaf:name>Nick Reynolds</foaf:name></foaf:Person>
    </rev:reviewer>
    - <rev:text>
      <p>When the Yeah Yeah Yeahs stormed into the UK...
    </rev:text>
    <cc:license rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/">
  </rev:Review>
```



hosting LOD



meta-lex + RDF-a

EU tenders



RDF export

**When success is
becoming a
problem...**

Success is becoming a problem



Gartner (May 2007):

"By 2012,
70% of public Web pages will have some level of semantic markup,
20% will use more extensive Semantic Web-based ontologies"

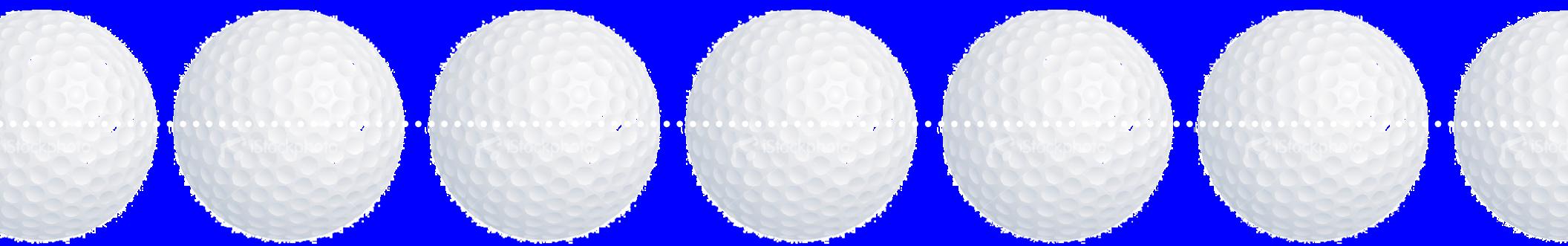
- Semantic Technologies at Web Scale?
 - 20% of 30 billion pages @ 1000 triples per page = **6 trillion triples**
 - 30 billion and 1000 are underestimates,
imagine in 6 years from now...
 - data-integration and semantic search at web-scale?

1 triple:









Suez Canal

10^7 Triples
[OWLIM]



Moon

RDF Store subsecond querying
 10^8 Triples
[Ingenta]



Earth

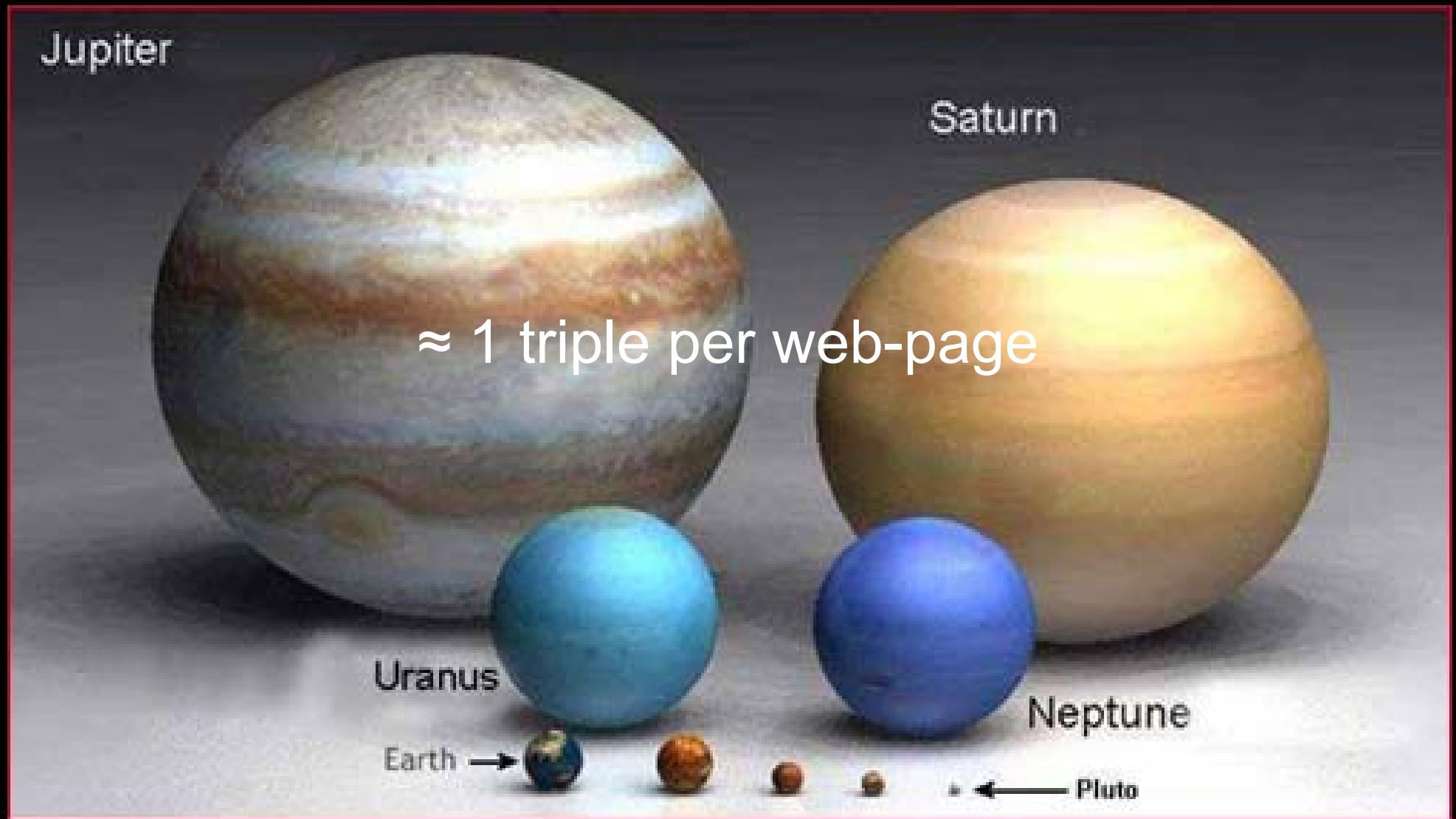
$\sim 10^9$ Triples



[LarkC proposal]

Jupiter

$\sim 10^{10}$ Triples ≈ 1 triple per web-page



Sun

$\sim 10^{11}$ Triples



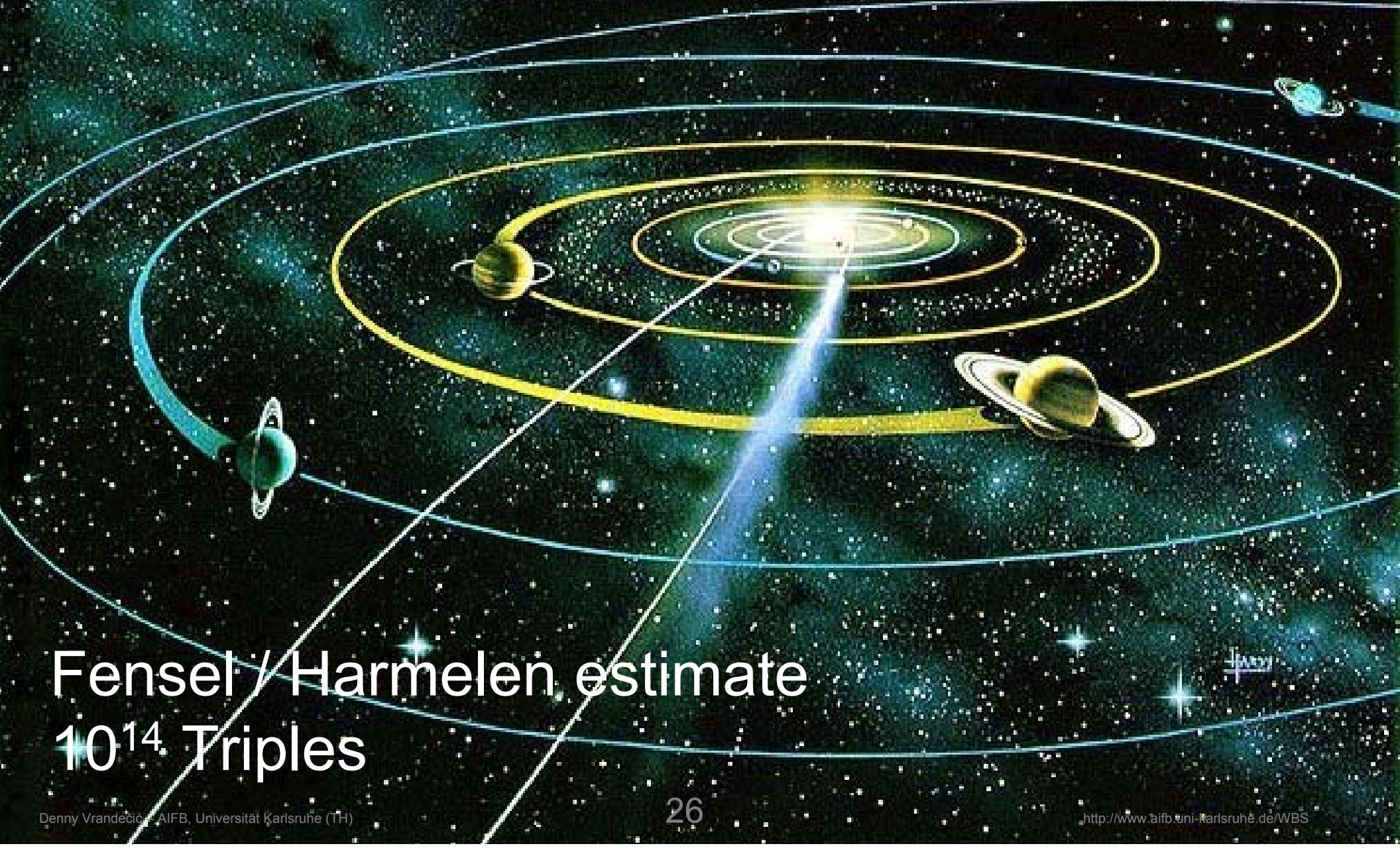
Jupiter

Earth

Pluto

Distance Sun – Pluto

$\sim 10^{14}$ Triples



Fensel / Harmelen estimate
 10^{14} Triples

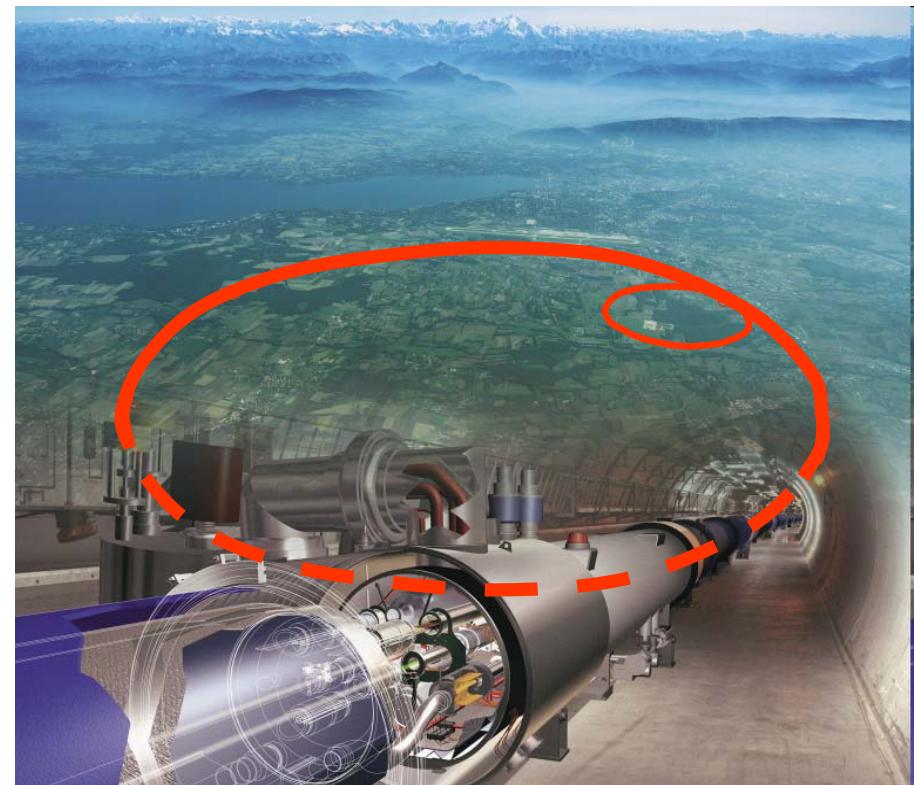
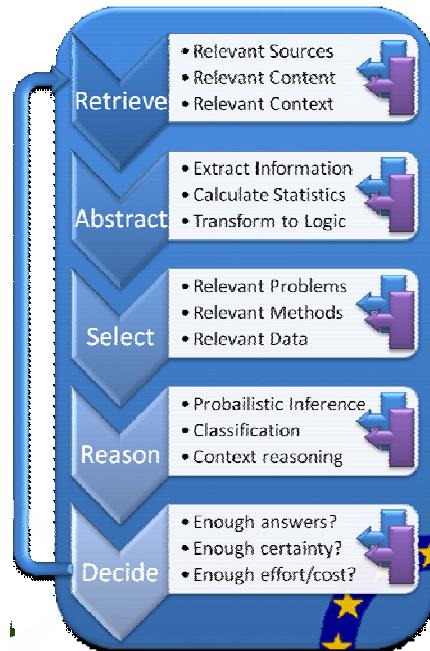
What to do when success becomes a problem?



The Large Knowledge Collider



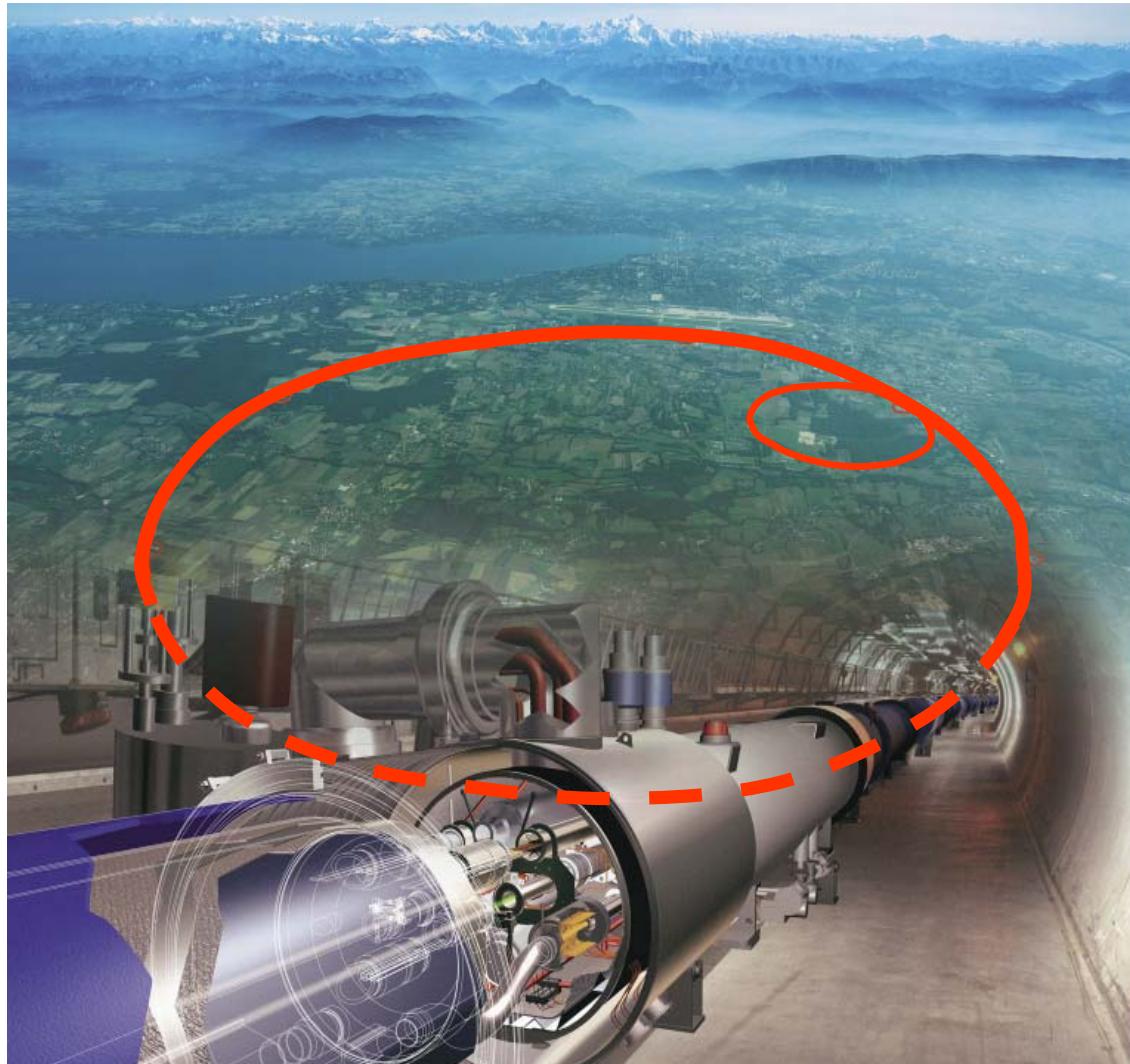
a platform for infinitely scalable reasoning on the data-web



Why “LarKC” ?

- The Large Knowledge Collider

A configurable platform
for experimentation
by others



Part I: LarKC platform



LarKC = a platform for large scale reasoning

Quote from US high-tech CTO:

Semantic web research is stifled by the complexity of writing a large scale engine, with services for data access, storage, aggregation, inference, transport, transformation, etc,

Physics research has dealt with a similar problem by **providing large scale infrastructure** into which experiments can be plugged.

The idea behind LarKC, which I found so compelling, is that people who wanted to build small scale plugins, for example, plugins for some non-standard deduction, or transformation of text to triples, or estimating the weights for relational models, could do so, **taking advantage of the EU's investment in a platform with significant capabilities.**"

Part I: LarKC platform



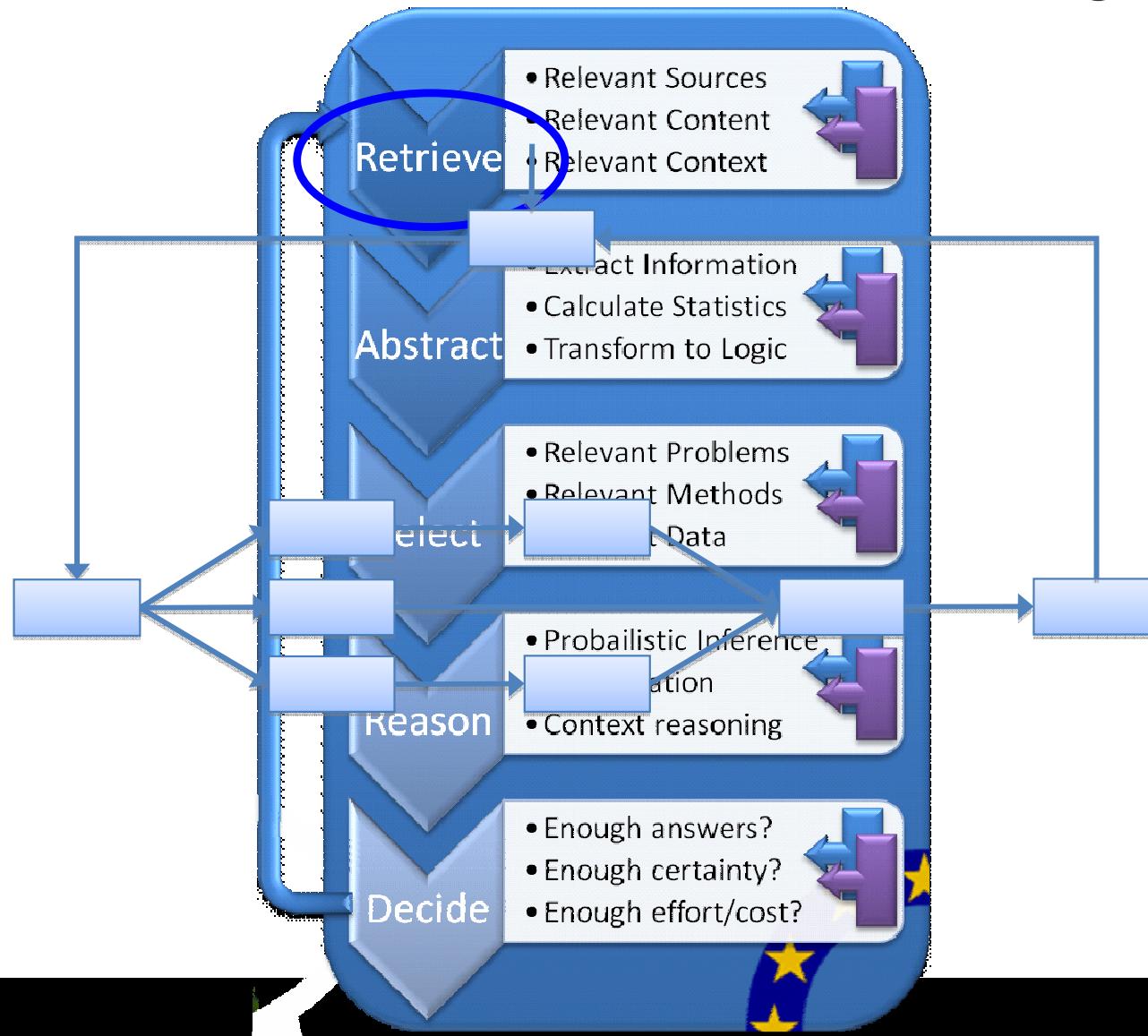
LarKC = a **platform** for **large scale** **reasoning**

Quote from EU Project Officer:

"LarKC's value is as an **experimental platform**.
LarKC is as an **environment** where people can go to
replicate (or extend) their results in an environment
where **all the infrastructural heavy lifting has already been taken care of**"

“Configurable platform”

“a configurable platform for infinitely scalable semantic web reasoning”



What do we mean by:



LarkC = a **platform** for large scale reasoning

- **reusable** components
- **reconfigurable** workflows
- provide **infrastructure** needed by all users:
 - storage and retrieval
 - communication (between plugins, plugins and datalayer)
 - synchronisation (support for anytime behaviour)
 - registration (of plugins)
 - abstracts from local or remote data-storage
 - abstracts from local or remote plugin-invocation
 - (will) provide instrumentation & measuring
 - (will) provide for caching and data-locality
- integration of very **heterogeneous** components
 - heterogeneous data: unstructured text, (semi)structured data
 - heterogeneous code: Java, scripts, remote services ("wrap & integrate")

Infinite scalability?

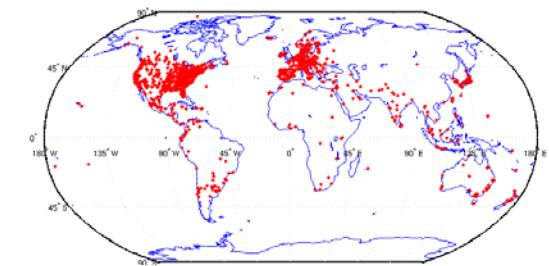
parallelisation

- cluster computing



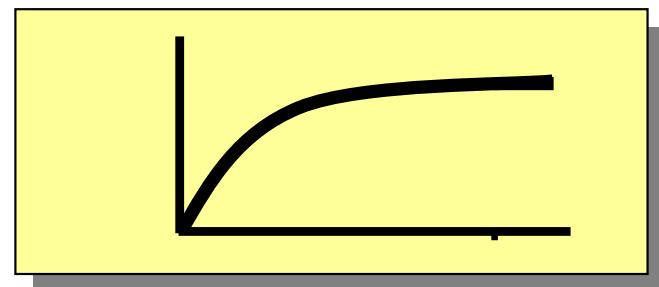
distribution

- “Thinking@home”,
“self-computing semantic Web”



approximation

- “almost” is often good enough
- gets better with more resources



What do we mean by:



LarkC = a platform for large scale **reasoning**

not only: deductive inference over given axioms
but also:

where do the axioms come from? (**IDENTIFY**)
which part of knowledge & data is required (**SELECTION**)
when is an answer "good enough" or "best possible" (**DECIDER**)
non-deductive inference (inductive, statistical) (**REASONer**)

Remember: "**ReaSearch: integrating reasoning and search**"

25 working plugins

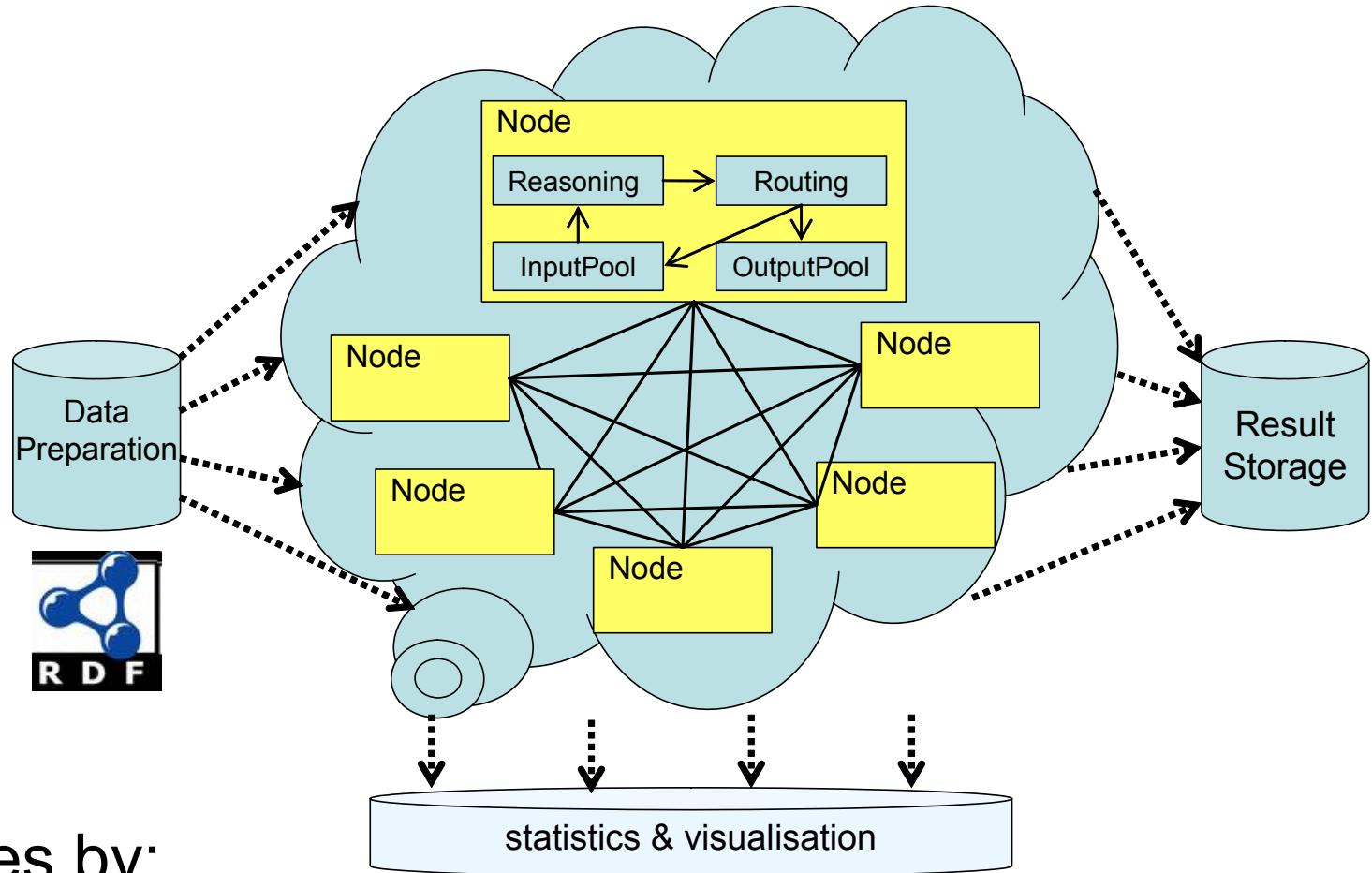
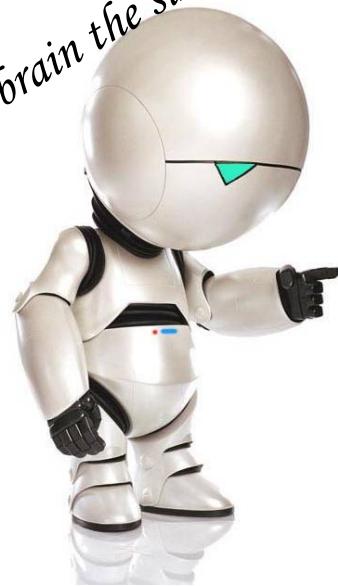


- IDENTIFY data sources(e.g. Sindice, Swoogle)
 - note: use existing web-service
- IDENTIFY text sources (Google)
 - note: use existing non-semantic search engine
- TRANSFORM text to triples (GATE, Open Calais)
 - note: use a very large (pipeline-based) system (GATE)
- TRANSFORM XML data to RDF triples (XSLT)
- SELECT geographically relevant triples (Allegrograph)
 - note: use another RDF store as SELECT
- SELECT semantically relevant triples
 - tokens, key phrases, prior knowledge, ranked
- SELECT structurally triples through spreading activation
- REASON with very different reasoners (Jena, Cyc, IRIS, DIG)
- REASON over inconsistent ontologies (PION)
-

First result: MaRVIN

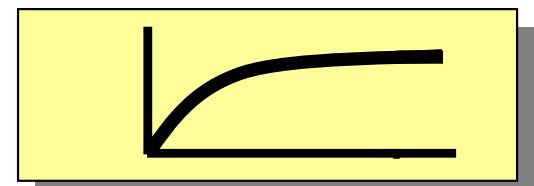


brain the size of a planet



MaRVIN scales by:

- distribution (over many nodes)
- approximation (sound but incomplete)
- anytime convergence (more complete over time)



Some muscle rolling



- “Lazy semantic pipes” for efficiently handling $O(10^{10})$ triples
- Peak inference rates at 8M triples/sec
- Sustained inference rates at 4M triples/sec
- deployed on laptops, servers, clusters
 - remote deployment on 64 nodes
- 25 plugins, 3 pipelines
 - WebSPARQL
 - Alpha Urban LarKC
 - NewsSPARQL
- Substantial new datasets:
 - Linked Life Data (1.4B explicit, 2.3B closure, 1.3M links)
 - Milan traffic grid (2M explicit + 2Tb sensor-data (to come))
 - Interest-enhanced DBLP (615k authors + interests)
 - LDSR (358M explicit + 512 inferred, 100m URIs)

We encourage further contributions from external parties:



- The Large Knowledge Collider is an **open, and configurable platform**.
 - The first public version of the Large Knowledge Collider is **available now**.
1. Organisations from outside the consortium can **use the LarKC platform** for their own purposes,
 2. LarKC has formed an "**early adapters group**".
 - LarKC will **actively support** this group in use the Large Knowledge Collider platform.
 - This group will be given **access to a high-performance computing-cluster** in Germany for running LarKC on their own problems.

External parties (both academic and commercial) are welcome to contact us on this opportunity.

Historical perspective Implications for tools? (thinking aloud...)

Any patterns in the applications?

- They all read from and or write to the **Linked Open Data** cloud
- They all do some reasoning
- The reasoning is **very lightweight**
- But happens over **very many instances**
- Notice stark difference with Guus:

"If we cannot show added value in knowledge-rich domains, then it may have no value at all".

Two tribes?

Maximum Fidelity tribe:

capture the conceptual relationships within knowledge and domains

Maximum Scalability tribe:

link and describe as many things as possible



A history of KR?

■ '70s: foundational theories

- small axioms sets ($< 10^2$),
- highly intensive reasoning
- Hayes' Naive Physics Manifesto
- few but very generic rules, (almost) no instance data



■ '80s-'90s: KBS (10^2 - 10^3 rules)

- moderately intensive reasoning:
frames, rules, semantic networks
- Feigenbaum's "the power is in the knowledge"



■ '00-'10: Web of Data (10^{10} facts, 10^7 rules)

- many, many instances
- Linked Open Data cloud
- **"it's the A-box, stupid!"**



Open questions

- Is this shift saying something fundamental
 - about knowledge representation?
 - about knowledge engineering?
 - about knowledge?

- Should this have consequences for the tools we develop?