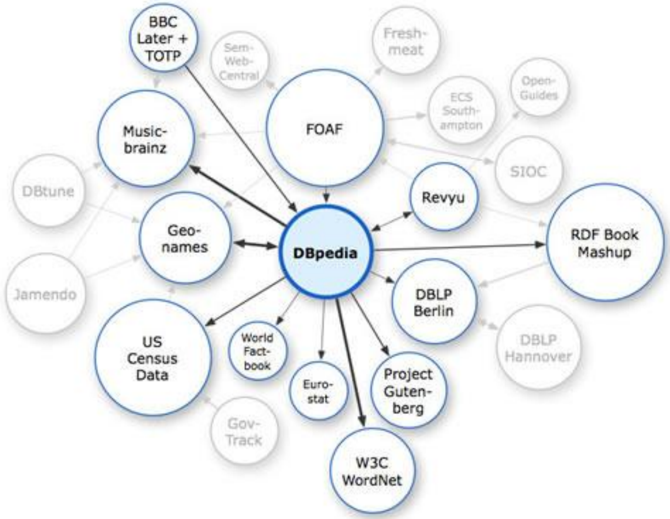
Karlsruhe Institute of Technology

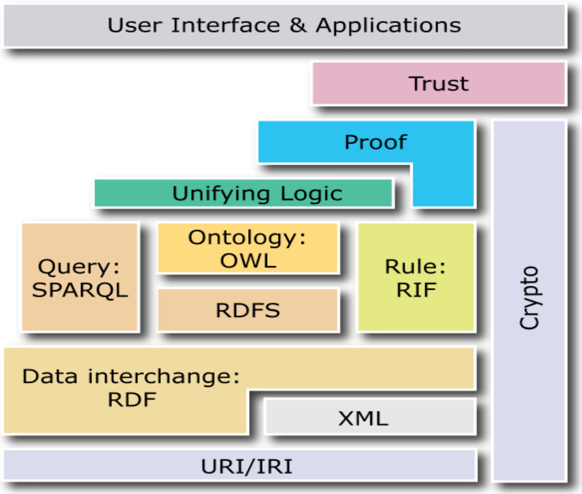
AIFB

Seminarpraktikum
Semantic Web Engineering

Thanh Tran, Günter Ladwig, Andreas Wagner

Institute of Applied Informatics and Formal Description Methods (AIFB)

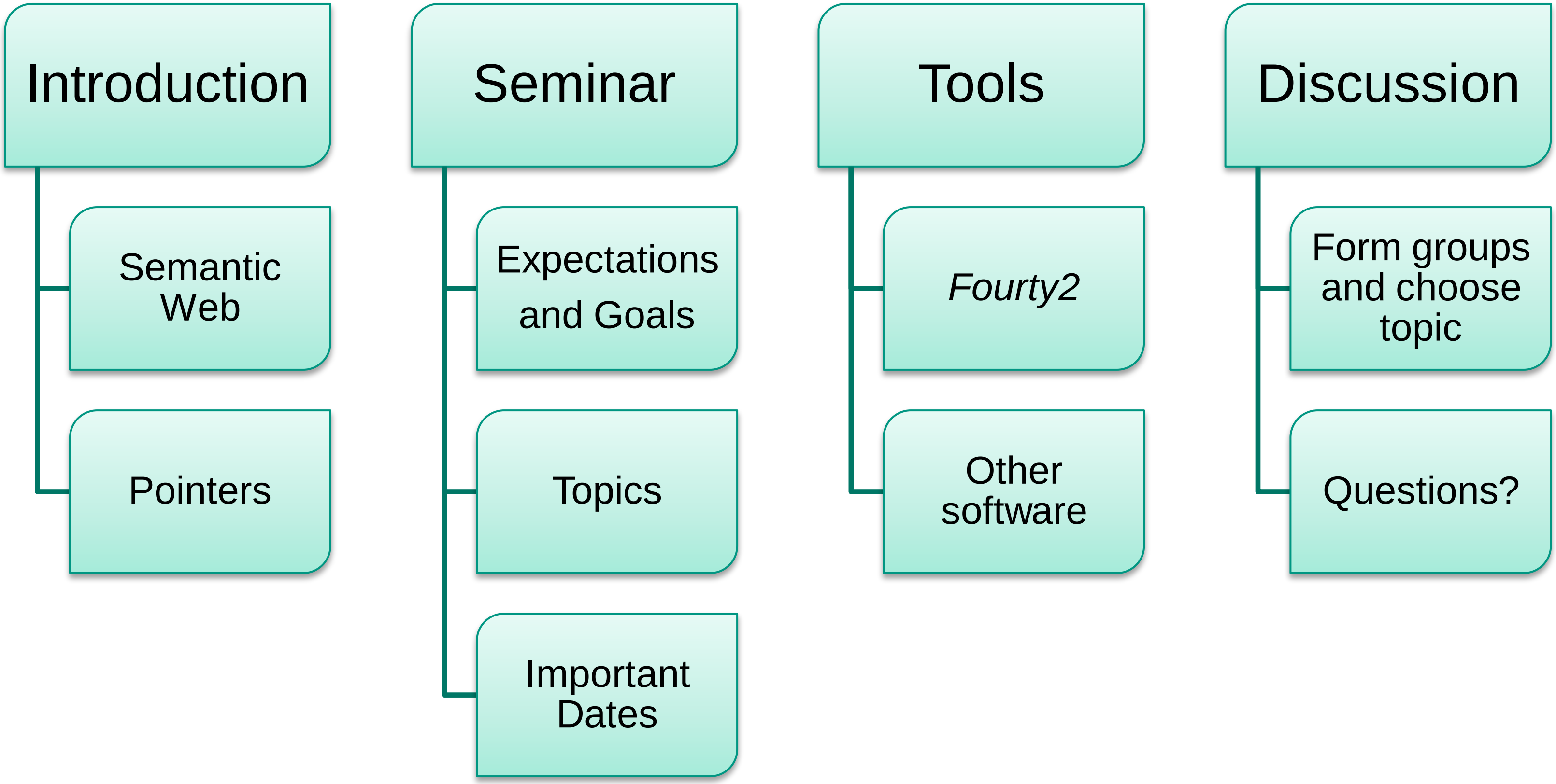




KIT – University of the State of Baden-Württemberg and
National Large-scale Research Center of the Helmholtz Association

www.kit.edu

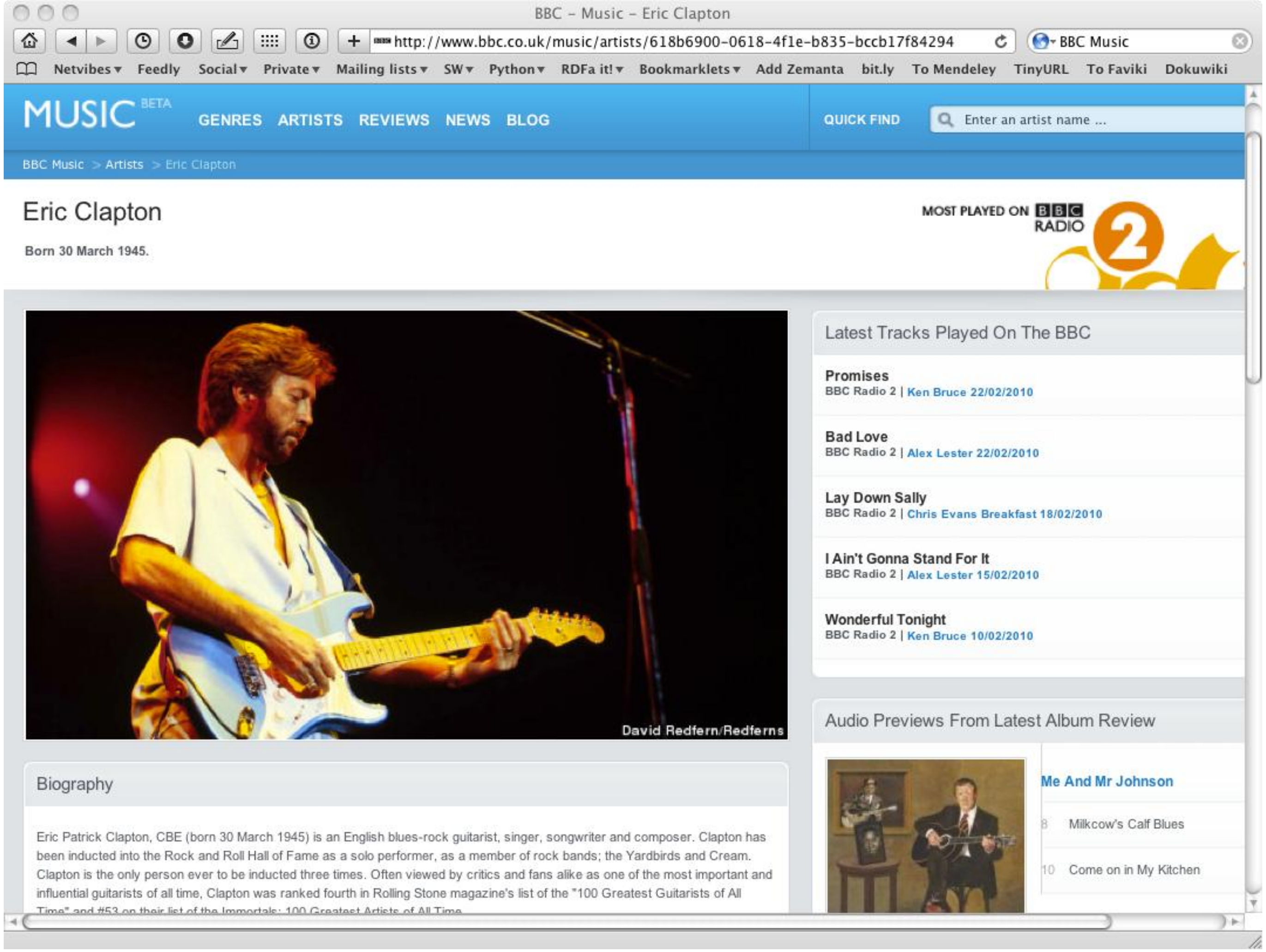
Agenda



- Introduction
  - Semantic Web
  - Pointers
- Seminar
  - Expectations and Goals
  - Topics
  - Important Dates
- Tools
  - *Fourty2*
  - Other software
- Discussion
  - Form groups and choose topic
  - Questions?

INTRODUCTION

Semantic Web – Motivation – BBC Music Site


BBC – Music – Eric Clapton
http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294
BBC Music
Netvibes ▾ Feedly Social ▾ Private ▾ Mailing lists ▾ SW ▾ Python ▾ RDFa it! ▾ Bookmarklets ▾ Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki
MUSIC BETA
GENRES ARTISTS REVIEWS NEWS BLOG
QUICK FIND
Enter an artist name ...
BBC Music > Artists > Eric Clapton
MOST PLAYED ON BBC RADIO 2
David Redfern/Redferns

Eric Clapton

Born 30 March 1945.

Latest Tracks Played On The BBC

Promises
BBC Radio 2 | Ken Bruce 22/02/2010

Bad Love
BBC Radio 2 | Alex Lester 22/02/2010

Lay Down Sally
BBC Radio 2 | Chris Evans Breakfast 18/02/2010

I Ain't Gonna Stand For It
BBC Radio 2 | Alex Lester 15/02/2010

Wonderful Tonight
BBC Radio 2 | Ken Bruce 10/02/2010

Audio Previews From Latest Album Review

Me And Mr Johnson

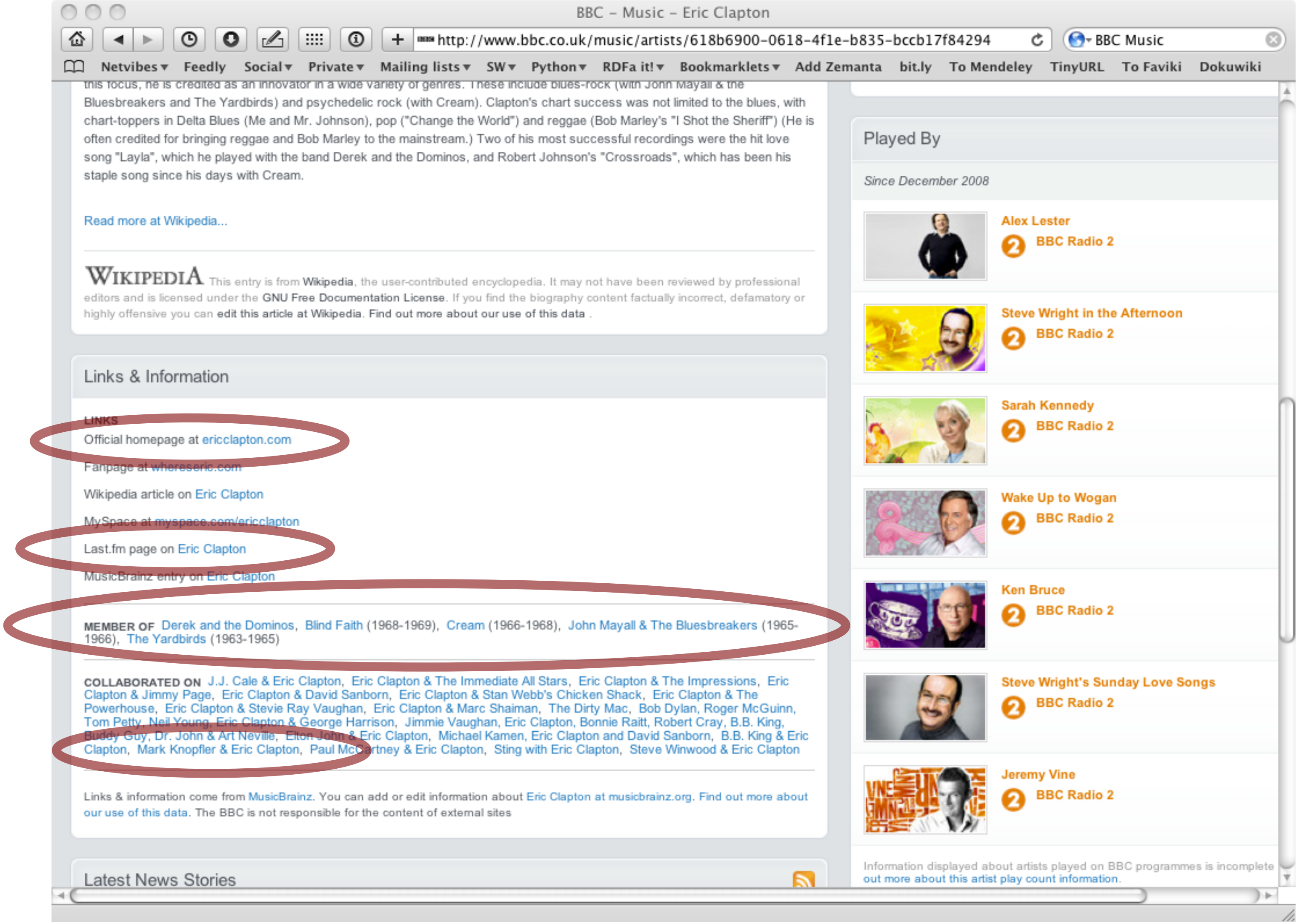
8 Milkcow's Calf Blues

10 Come on in My Kitchen

Biography

Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and #53 on their list of the Immortals: 100 Greatest Artists of All Time

Semantic Web – Motivation – BBC Music Site



this focus, he is credited as an innovator in a wide variety of genres. These include blues-rock (with John Mayall & the Bluesbreakers and The Yardbirds) and psychedelic rock (with Cream). Clapton's chart success was not limited to the blues, with chart-toppers in Delta Blues (Me and Mr. Johnson), pop ("Change the World") and reggae (Bob Marley's "I Shot the Sheriff") (He is often credited for bringing reggae and Bob Marley to the mainstream.) Two of his most successful recordings were the hit love song "Layla", which he played with the band Derek and the Dominos, and Robert Johnson's "Crossroads", which has been his staple song since his days with Cream.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data .

Links & Information

LINKS

Official homepage at ericclapton.com

Fanpage at whereseric.com

Wikipedia article on Eric Clapton

MySpace at myspace.com/ericclapton

Last.fm page on Eric Clapton

MusicBrainz entry on Eric Clapton

MEMBER OF Derek and the Dominos, Blind Faith (1968-1969), Cream (1966-1968), John Mayall & The Bluesbreakers (1965-1966), The Yardbirds (1963-1965)

COLLABORATED ON J.J. Cale & Eric Clapton, Eric Clapton & The Immediate All Stars, Eric Clapton & The Impressions, Eric Clapton & Jimmy Page, Eric Clapton & David Sanborn, Eric Clapton & Stan Webb's Chicken Shack, Eric Clapton & The Powerhouse, Eric Clapton & Stevie Ray Vaughan, Eric Clapton & Marc Shaiman, The Dirty Mac, Bob Dylan, Roger McGuinn, Tom Petty, Neil Young, Eric Clapton & George Harrison, Jimmie Vaughan, Eric Clapton, Bonnie Raitt, Robert Cray, B.B. King, Buddy Guy, Dr. John & Art Neville, Elton John & Eric Clapton, Michael Kamen, Eric Clapton and David Sanborn, B.B. King & Eric Clapton, Mark Knopfler & Eric Clapton, Paul McCartney & Eric Clapton, Sting with Eric Clapton, Steve Winwood & Eric Clapton

Links & information come from MusicBrainz. You can add or edit information about Eric Clapton at musicbrainz.org. Find out more about our use of this data. The BBC is not responsible for the content of external sites

Latest News Stories

Played By

Since December 2008

Alex Lester
BBC Radio 2

Steve Wright in the Afternoon
BBC Radio 2

Sarah Kennedy
BBC Radio 2

Wake Up to Wogan
BBC Radio 2

Ken Bruce
BBC Radio 2

Steve Wright's Sunday Love Songs
BBC Radio 2

Jeremy Vine
BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete
out more about this artist play count information.

Semantic Web – Motivation
How to build the BBC Music Site?

1. Attempt

- Site editors roam the Web for new facts
  - may discover further links while roaming
- They update the site **manually**
- And the site gets soon **out-of-date**

Semantic Web – Motivation
How to build the BBC Music Site?

2. Attempt

- Editors roam the Web for new data published on Web sites
- "Scrape" the sites with a program to **extract the information**
  - ie, *write some code* to incorporate the new data
- Easily get **out of date** again ☹

Semantic Web – Motivation
How to build the BBC Music Site?

3. Attempt

- Editors roam the Web for **new data via API-s**
- Understand those…
  - input, output arguments, datatypes used, etc
- *Write some code* to incorporate the new data
- Easily **get out of date** again…

Semantic Web – Motivation
What did the BBC do?

- Use external, public datasets
  - Wikipedia, MusicBrainz, …
- They are available *as data*
  - not API-s or hidden on a Web site
  - data can be extracted using, eg, HTTP requests or standard queries

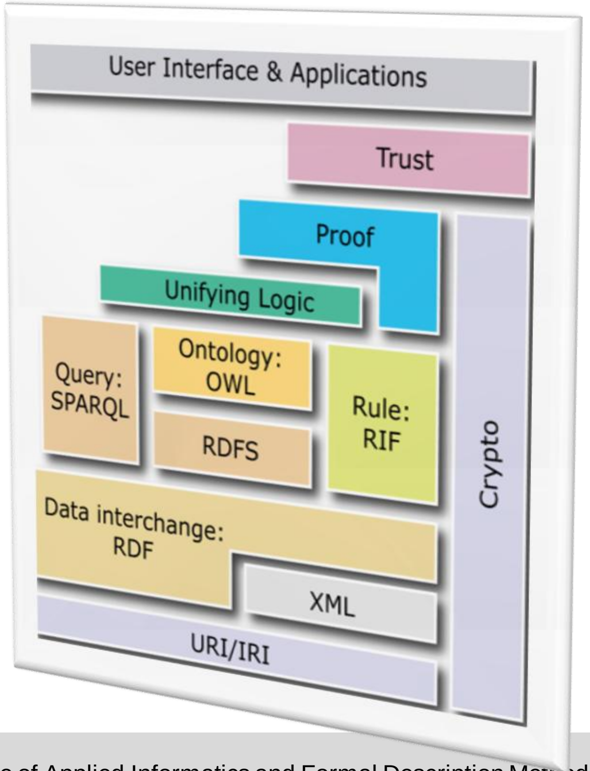
- Use the Web of Data as a Content Management System
- Use the community at large as content editors

Semantic Web – Motivation Summary

- With increased use of computers more and more data is being stored
  - Organisations rely on data for business decisions
  - Data drives policy decisions in government
  - Individuals rely on data from the Web for information and communication
- Data volumes explode
  - More and more data available on the Web is represented in Semantic Web standards
  - Linking Data initiative [1]
- Semantic Web technologies facilitate the **integration of data from multiple sources**
  - Combining data from multiple sources enables insights

Semantic Technologies

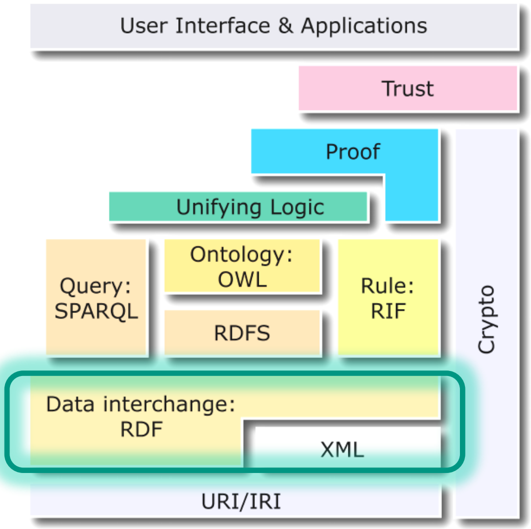
- Useful for **data publishing, exchange, and integration**
- Insights possible when combining data from multiple sources
- Semantic Web technologies, standardised by the W3C [2], are mature:
  - **RDF** recommendation in 1999, update in 2004
  - RDFa (RDF in HTML) note in 2008
  - **RDFS** recommendation in 2004
  - **SPARQL** recommendation in 2008
  - OWL recommendation in 2004, update in 2009
- Linked Data comprises of a few principles for data publishing on the web



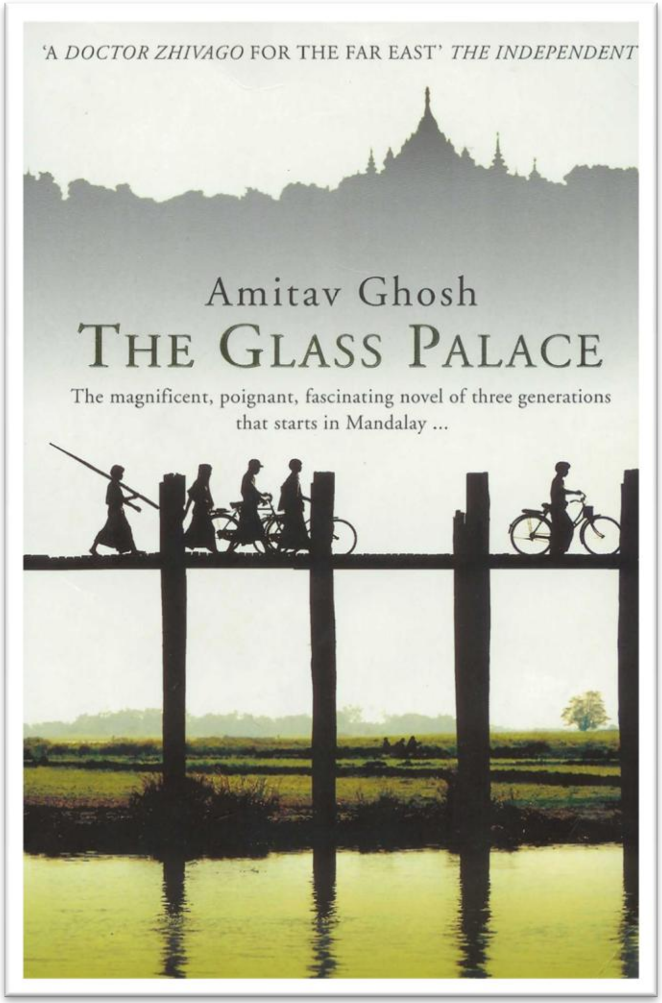
Semantic Web Technologies
Resource Description Framework (RDF)

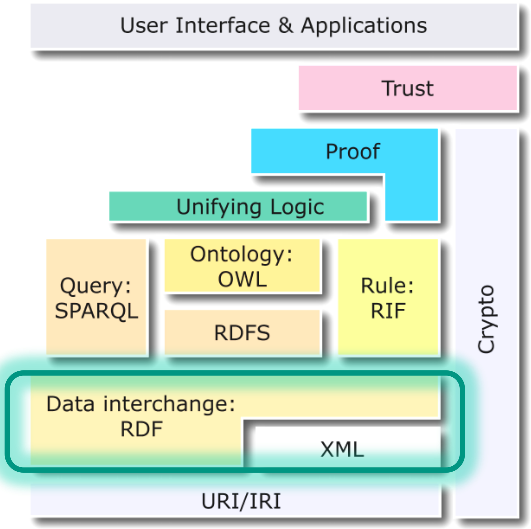
- Directed, labeled graph
- `triple(subject, predicate, object)`
  - subject: URI or blank node
  - predicate: URI
  - object: URI or blank node or RDF literal

- RDF/XML is the most widely deployed serialization
- Other serializations possible (N-Triples, Turtle, Notation3…)
- Quadruples (or quads) used as internal representation when integrating data
- `quad(subject, predicate, object, context)`
  - context: URI (used to store origin of triple)



Resource Description Framework (RDF) Example – We start with a book ...


'A DOCTOR ZHIVAGO FOR THE FAR EAST' THE INDEPENDENT
Amitav Ghosh
THE GLASS PALACE
The magnificent, poignant, fascinating novel of three generations that starts in Mandalay ...


User Interface & Applications

Trust

Proof

Unifying Logic

Query: SPARQL

Ontology: OWL

RDFS

Rule: RIF

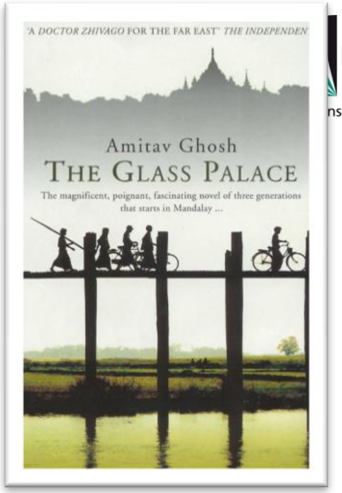
Crypto

Data interchange: RDF

XML

URI/IRI

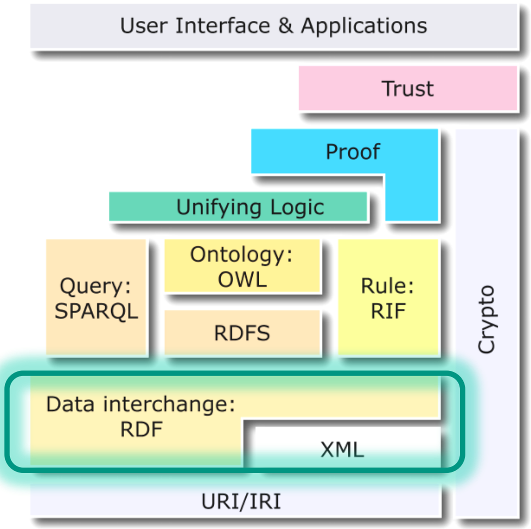
Resource Description Framework (RDF) Example – A simplified bookstore data


'A DOCTOR ZHIVAGO FOR THE FAR EAST' THE INDEPENDENT
Amitav Ghosh
THE GLASS PALACE
The magnificent, poignant, fascinating novel of three generations that starts in Mandalay ...

ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

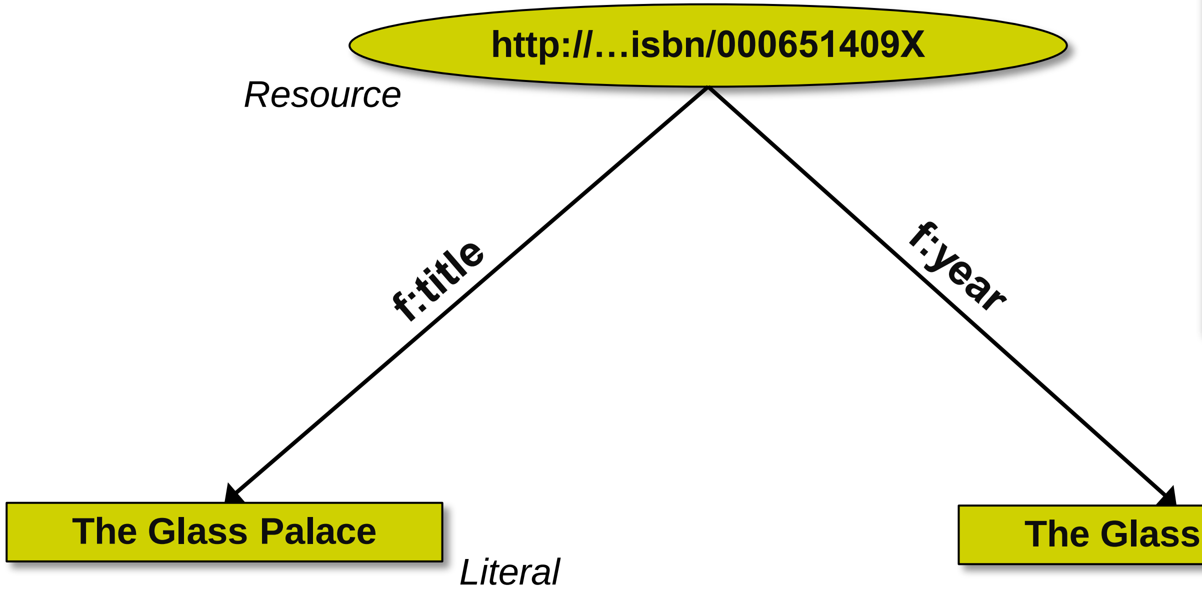
ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

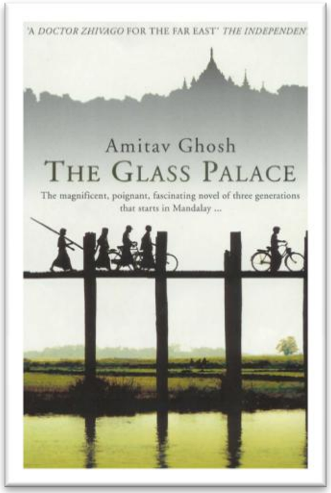
ID	Publisher's name	City
id_qpr	Harper Collins	London

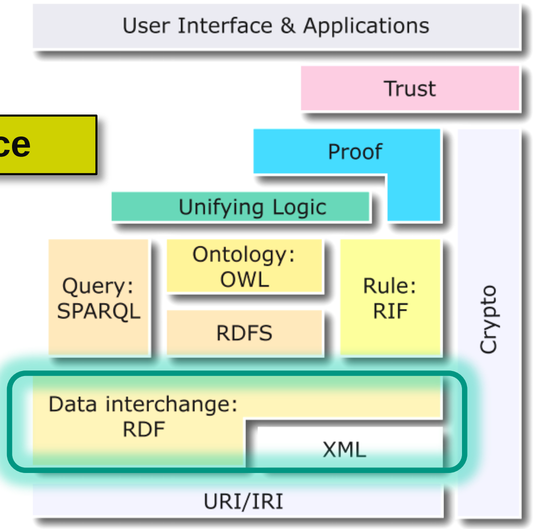


Resource Description Framework (RDF) Example

- RDF triples form a directed, labeled graph




'A DOCTOR ZHIVAGO FOR THE FAR EAST' THE INDEPENDEN
Amitav Ghosh
THE GLASS PALACE
The magnificent, poignant, fascinating novel of three generations that starts in Mandalay ...

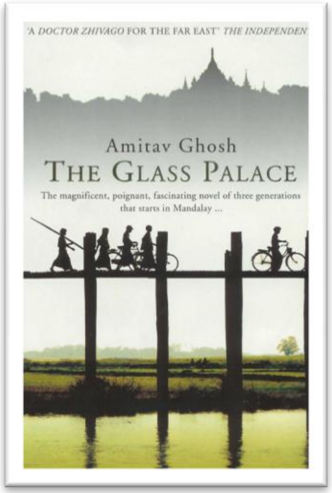


```
<rdf:Description rdf:about="http://…/isbn/2020386682">
    <f:title xml:lang=“en">The Glass Palace</f:titre>
    <f:year>2000</f:year>
</rdf:Description>
```

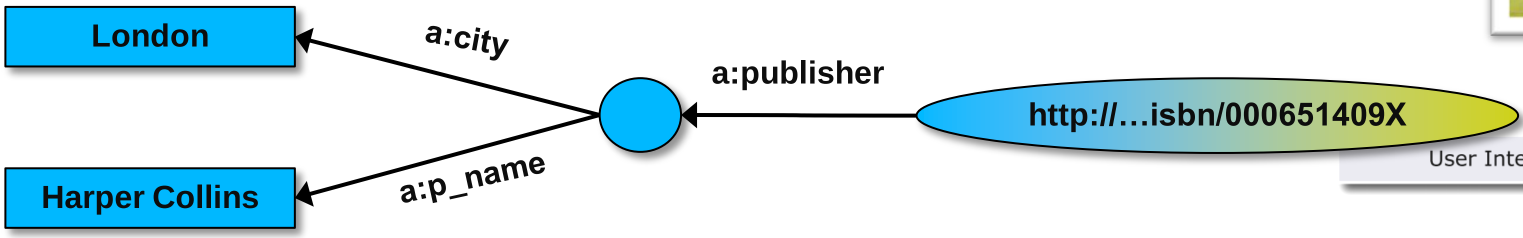
Note: namespaces are used to simplify the URI-s

Resource Description Framework (RDF) Example

- “Internal” nodes (aka blank nodes)
  - Consider the following statement:
    - “the publisher is a «thing» that has a name and an address”
  - Until now, nodes were identified with a URI. But…
  - …what is the URI of «thing»?

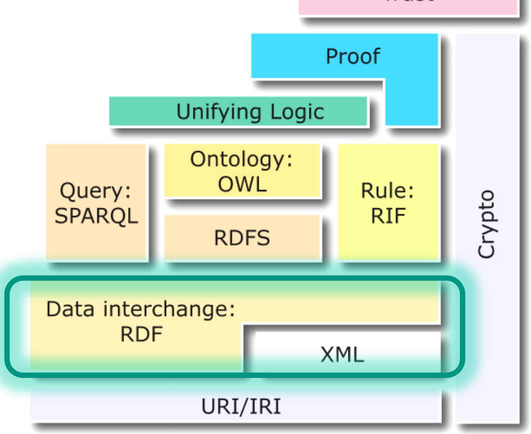


'A DOCTOR ZHIVAGO FOR THE FAR EAST' THE INDEPENDEN
Amitav Ghosh
THE GLASS PALACE
The magnificent, poignant, fascinating novel of three generations that starts in Mandalay ...



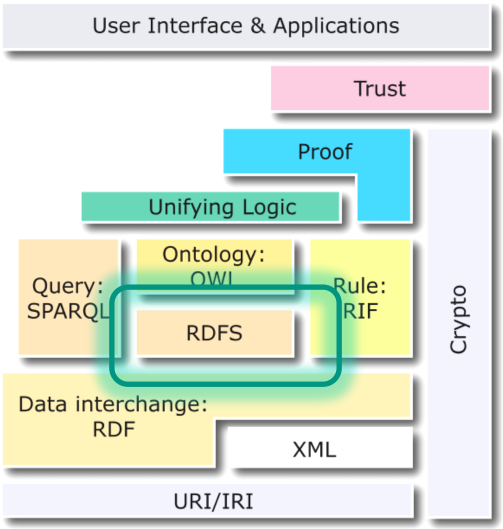
```
<rdf:Description rdf:about="http://…/isbn/000651409X">
   <a:publisher rdf:nodeID="A234"/>
</rdf:Description>
<rdf:Description rdf:nodeID="A234">
   <a:p_name>HarpersCollins</a:p_name>
   <a:city>HarpersCollins</a:city>
</rdf:Description>
```

Note: Internal = these resources are *not* visible outside



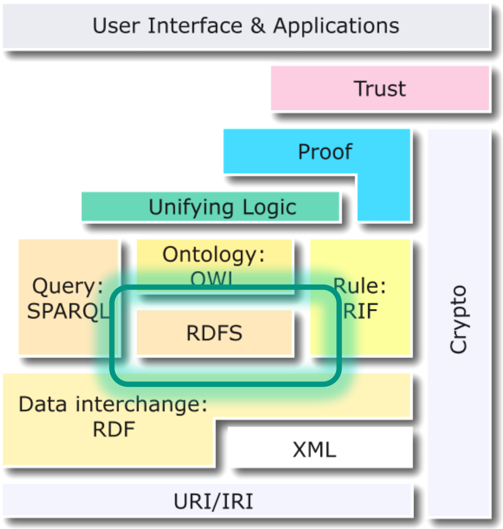
Resource Description Framework Schema

- First step towards the "extra knowledge":
  - define the terms we can use
  - what restrictions apply
  - what extra relationships are there?
- RDFS defines resources and classes:
  - everything in RDF is a "resource"
  - "classes" are also resources, but…
  - …they are also a collection of possible resources (i.e., "individuals")
- Relationships are defined among resources:
  - "typing": an individual belongs to a specific class
    - "«The Glass Palace» is a novel"
  - "subclassing": all instances of one are also the instances of the other



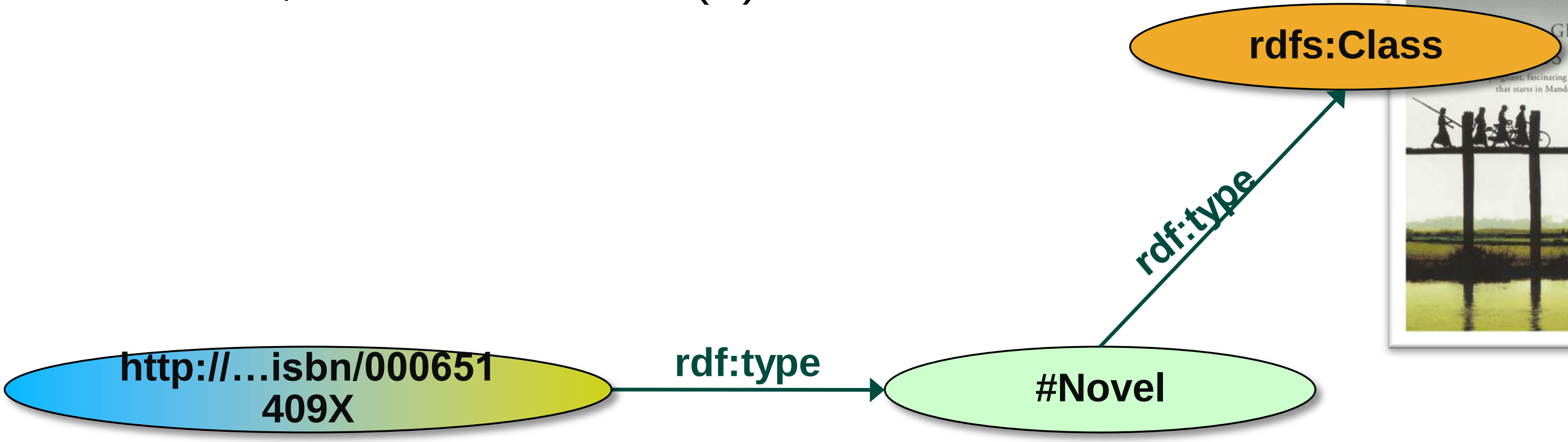
Resource Description Framework Schema

- Property is a special class (rdf:Property)
  - properties are also resources identified by URI-s
- There is also a possibility for a “sub-property”
  - all resources bound by the “sub” are also bound by the other
- Range and domain of properties can be specified
  - i.e., what type of resources serve as object and subject

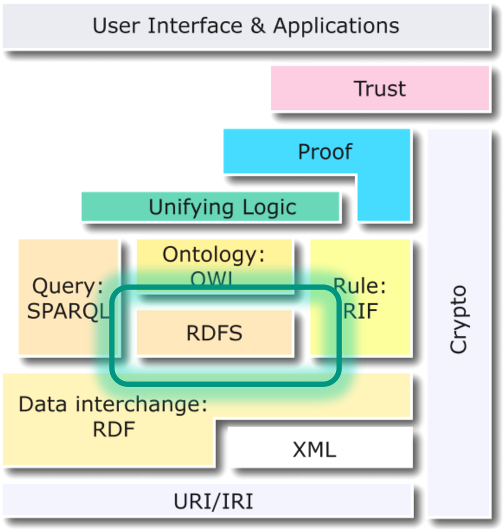


Resource Description Framework Schema Example

- Classes, resources in RDF(S)

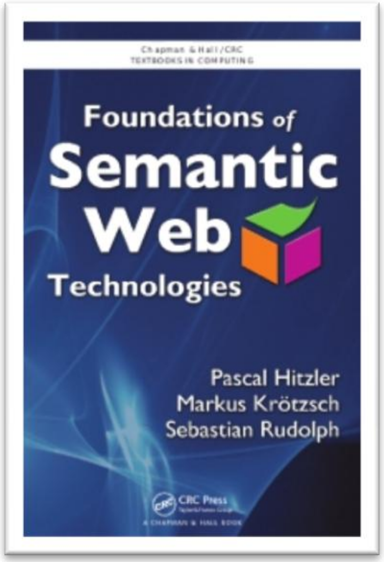


RDFS defines the meaning of these terms
(these are all special URI-s, we just use the namespace abbreviation)



Pointers

- **Books**
  - **Foundations of semantic web technologies**
    Pascal Hitzler, Markus Krötzsch,
    Sebastian Rudolph
  - **Linked Data**
    Tom Heath and Chris Bizer
    http://linkeddatabook.com/editions/1.0/
  - **Handbook on Ontologies**
    Steffen Staab



- **Slides, Talks etc**
  - **http://www.w3.org/2001/sw/**

- **Tools**
  - Fourty2 Platform

System *Fourty2* – Demo

SEMINAR

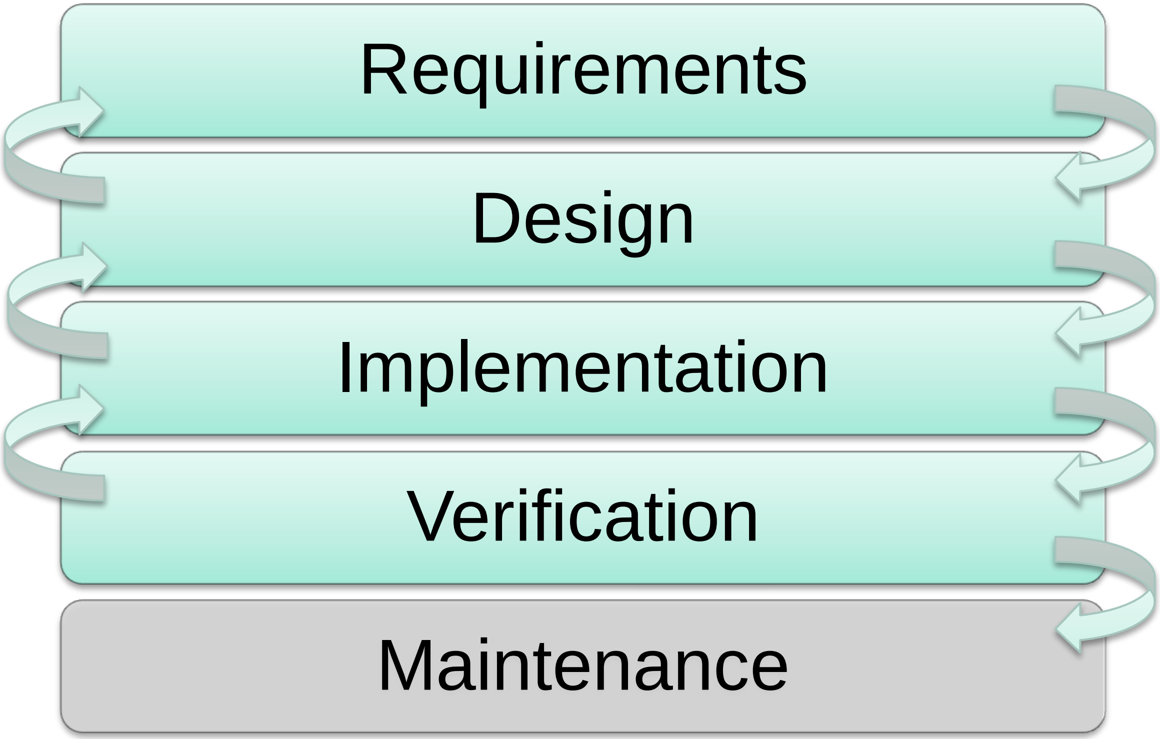
Expectations

- Have **system running** at the end
- Keep the **usability** in mind → system must be understandable to anybody
- Keep **reaction times** in mind → a running system must be fast

- What tools are used is up to you. However, we support our system **Fourty2**
- Work as a **team**
  - Talk to each other
  - Organize your work
  - Implement and test your ideas together

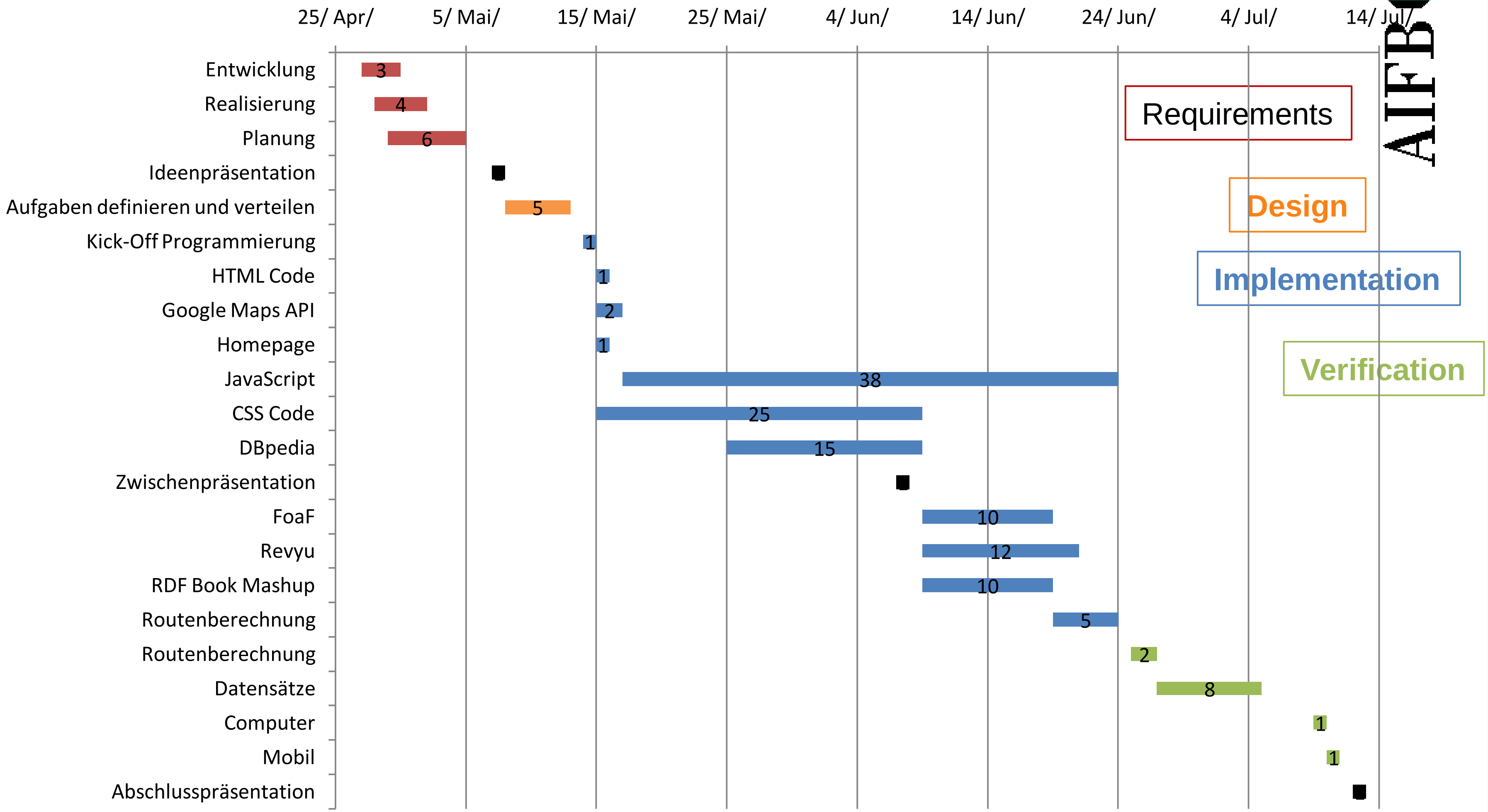
Expectations

- Employ **waterfall model (or similar model)**



See also: Royce, Winston (1970), "Managing the Development of Large Software Systems“ [3]

Expectations – Keep track of time …



Topics …

- Short overview of predefined topics
  - *Hybrid search (text and structured constraints)*
    - Integrate existing hybrid search code in frontend
    - Make hybrid search interface more usable, e.g., a result preview
  - *Linked data source selection*
    - Describe data source in intuitive manner
    - Allow user to select data source for her queries
  - *Pivotsearch*
    - Allow users to „jump“ from one result to another
    - For instance, a result may contain the U.S. presidents; now users may want to see the universities, which the presidents attended
  - *Context-aware search*
    - Allow users to issue query using „context“ information
    - For instance, find all tire manufacturer trusted by ADAC

Topics …

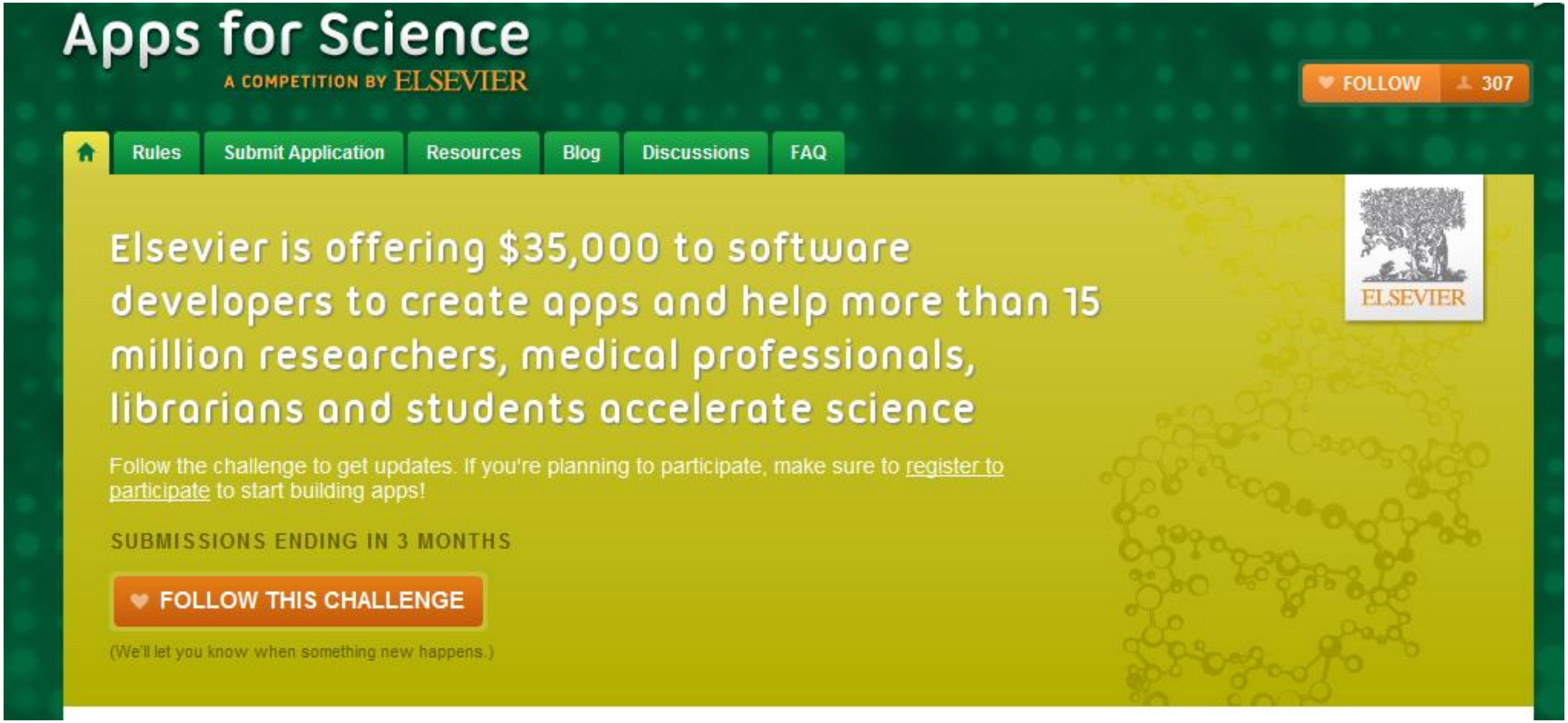
- Short overview of predefined topics
  - *Hybrid content authoring (text and structured data)*
    - Create an intuitive interface for modifying structured and unstructured data
  - *Result set visualisation*
    - Allow a flexible result presentation (not only lists), depending on the contents of the current result set
  - *Natural Syntax for SPARQL*
    - More natural and easier syntax for basic SPARQL expressions
    - Parse new syntax and translate to SPARQL
  - *SPARQL query builder*
    - Create an intuitive SPARQL query builder
    - Enable (for instance) a result preview or warn users, if a query may lead to an empty result
- **Other topics** are also possible → discussion at the end

Getting a good grade …

- Grading
  - Code ~ 50%
  - Documentation (waterfall or similar model) ~ 40%
  - Final presentation ~ 10%

Goals beyond a good grade ...
Elsevier Apps for Science Challenge

- http://appsforscience.com/
- Deadline: **July 31st**
- Goal: science applications on top of Fourty2



Elsevier is offering $35,000 to software developers to create apps and help more than 15 million researchers, medical professionals, librarians and students accelerate science

Follow the challenge to get updates. If you're planning to participate, make sure to register to participate to start building apps!

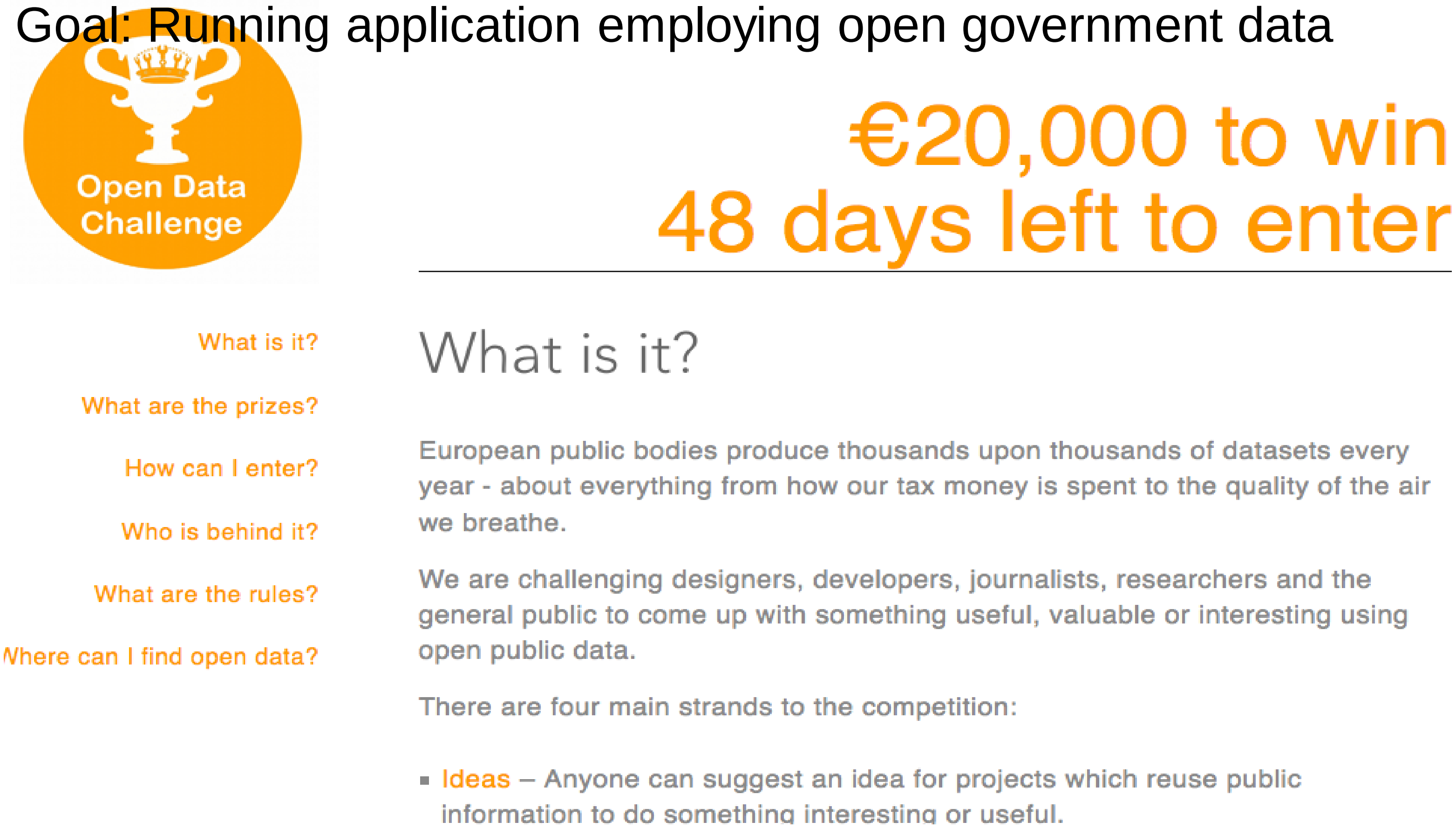
SUBMISSIONS ENDING IN 3 MONTHS

FOLLOW THIS CHALLENGE

(We'll let you know when something new happens.)

Goals beyond a good grade ...
Open Data Challenge

- http://opendatachallenge.org
- Deadline: **June 5th**
- Goal: Running application employing open government data



€20,000 to win
48 days left to enter

What is it?

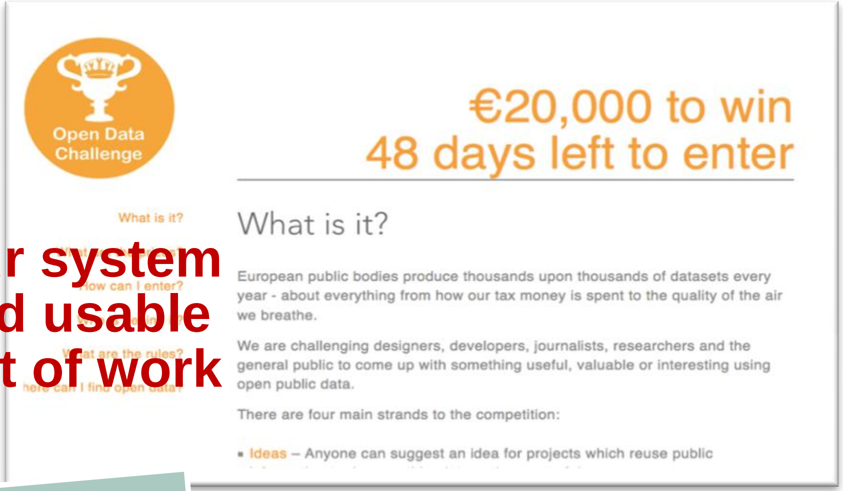
European public bodies produce thousands upon thousands of datasets every year - about everything from how our tax money is spent to the quality of the air we breathe.

We are challenging designers, developers, journalists, researchers and the general public to come up with something useful, valuable or interesting using open public data.

There are four main strands to the competition:

- Ideas – Anyone can suggest an idea for projects which reuse public information to do something interesting or useful.

Goals ... What does that mean for me?



€20,000 to win
48 days left to enter

What is it?

European public bodies produce thousands upon thousands of datasets every year - about everything from how our tax money is spent to the quality of the air we breathe.

We are challenging designers, developers, journalists, researchers and the general public to come up with something useful, valuable or interesting using open public data.

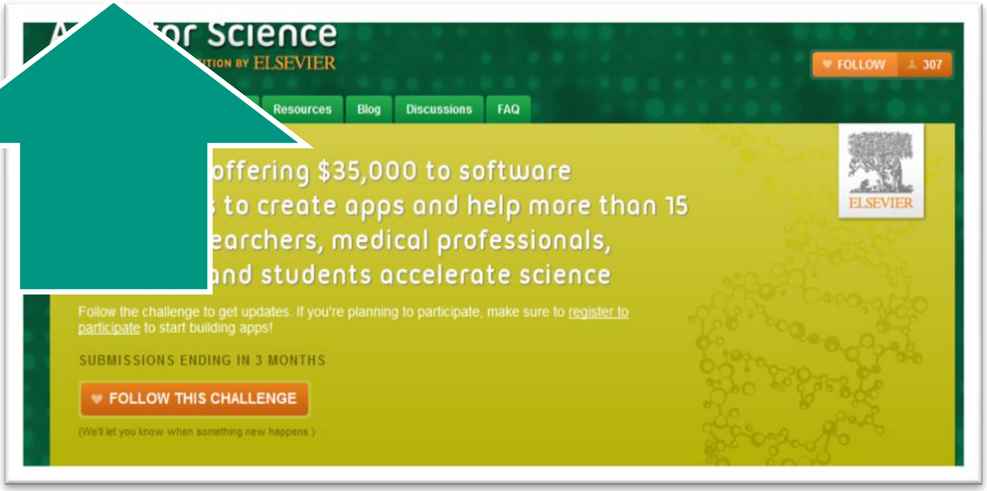
There are four main strands to the competition:

- Ideas – Anyone can suggest an idea for projects which reuse public

Getting your system running and usable may be a lot of work

You can learn a lot of things (e.g., working in a group or developing software)

You can get a reward beyond „some“ grade



offering $35,000 to software
s to create apps and help more than 15
earchers, medical professionals,
nd students accelerate science

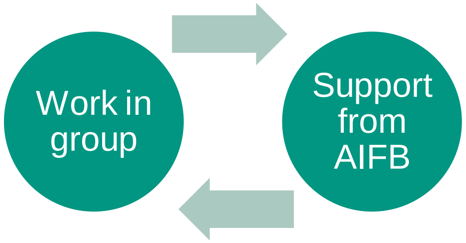
Follow the challenge to get updates. If you're planning to participate, make sure to register to participate to start building apps!

SUBMISSIONS ENDING IN 3 MONTHS

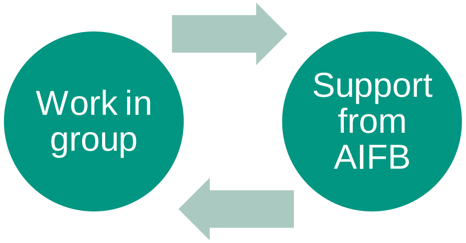
(We'll let you know when something new happens.)

Important Dates

- **Kick-Off** today: Think about topic, group etc.
- Until **May 4th, 5.15 pm (room 226)**. From Groups, choose system (and have it up and running) and topic



- **June 8th, 5.15 pm (room 226)**. Present and discuss intermediate results and problems



- **Final Presentation**, beginning of July: Turn in code and documentation

Final Presentation

- Present your running application
  - Is it usable?
  - Does it scale w.r.t. time and data size?
  - Live demo …
- Turn in documentation
  - ~ 10 pages, use LaTeX and write in English
    - Employ waterfall model
      - What are the requirements?
      - How is your system designed (UML)?
      - How is it implemented (most important classes, other software packages etc)?
      - How does your software fulfill the requirements (testing)?
  - Source code of your implementation

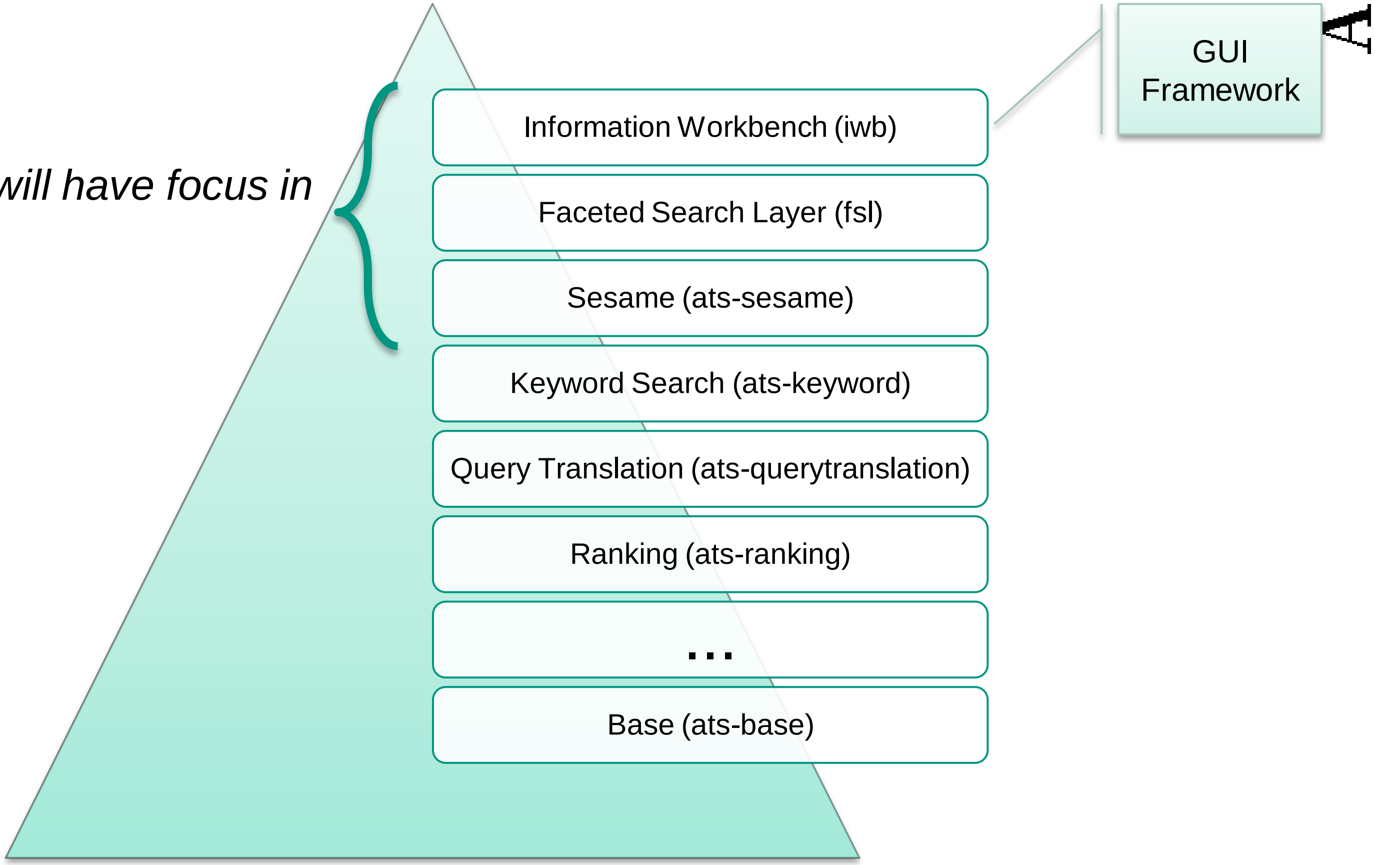
TOOLS

System *Fourty2*

- *Fourty2* is a state-of-the-art **search system developed at AIFB**
  - ... is a layered and extendable system
  - ... can handle structured and unstructured search
  - ... can handle keyword, structured and hybrid queries
  - ... uses a standard interface (SAIL interface [4]) for communcation between the layers
  - ... comes with a GUI framework on top
  - ... comes with documentation on how to set it up and run it

System *Fourty2* – Architecture

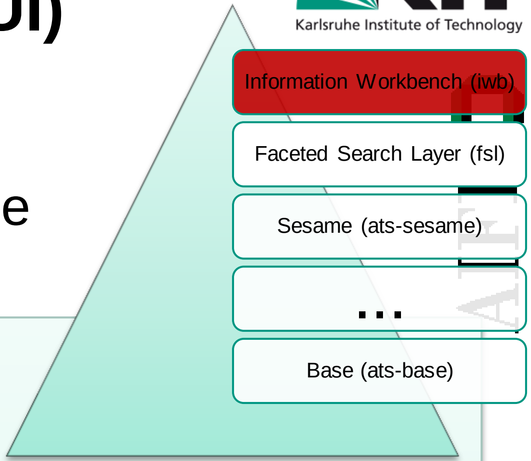
- *Fourty2* is a state-of-the-art **search system developed at AIFB**
  - ... is a layered and extendable system



Most topics will have focus in these layers
GUI Framework
Information Workbench (iwb)
Faceted Search Layer (fsl)
Sesame (ats-sesame)
Keyword Search (ats-keyword)
Query Translation (ats-querytranslation)
Ranking (ats-ranking)
...
Base (ats-base)

System *Fourty2* – Architecture – IWB (GUI)

- The information workbench [5] (IWB) is a open source framework for easily creating user interface components



Information Workbench (iwb)

Faceted Search Layer (fsl)

Sesame (ats-sesame)

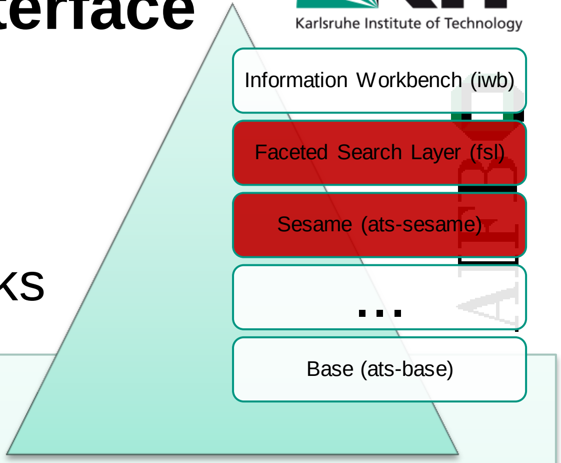
...

Base (ats-base)

```
FTableModel tm = new FTableModel();
tm.addColumn("Subject");
tm.addColumn("Predicate");
tm.addColumn("Object");

for (Statement st : res) {

    String[] row = new String[3];
    row[0] = st.getSubject().stringValue();
    row[1] = st.getPredicate().stringValue();
    row[2] = st.getObject().stringValue();

    tm.addRow(row);
}

FContainer container = getFacetContainer(pc);
FTable tbl = new FTable("table", ((KeywordQueryResult)
        pc.resultSet).tm);
...
```

System *Fourty2* – Architecture – SAIL Interface

- SAIL interface [4] is used by major RDF stores
- The interface defines (in particular) a data model, how query may be issued and what the result looks like


Information Workbench (iwb)

Faceted Search Layer (fsl)

Sesame (ats-sesame)

...

Base (ats-base)

```
...

try {

   RepositoryConnection con = myRepository.getConnection();

   try {

      String queryString = "SELECT ?x ?y WHERE { ?x ?p ?y . }";
      TupleQuery tupleQuery =
             con.prepareTupleQuery(QueryLanguage.SPARQL,
queryString);
      TupleQueryResult result = tupleQuery.evaluate();

      try {
         .... // do something with the result
      } ...
   } ...
} ...
```

Other Software

- Software useful for working in groups ...
  - Wikis or Groups for communication
    - Wikis: www.wikispaces.com ...
    - Groups: https://groups.google.com/ or http://groups.yahoo.com/ ...
  - Subversion Server for code and document versioning
    - SVN hosting: http://www.assembla.com/ ...
    - SVN clients:
      - http://tortoisesvn.tigris.org/ (Win OS, Explorer)
      - http://subclipse.tigris.org/ (Eclipse)
  - IDE
    - Eclipse: http://www.eclipse.org/

Everything comes free

DISCUSSION

Discussion

- Form groups
- Think about topic for your group
- Questions?

1. Hybrid search (text and structured constraints)
2. Linked data source selection for query answering
3. Pivotsearch
4. Context-aware search
5. Context-aware navigation in complex web data
6. Hybrid content authoring (text and structured data)
7. Result set visualisation
8. Natural Syntax for SPARQL
9. SPARQL query builder with result preview

REFERENCES

[1] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/

[2] http://www.w3.org/2001/sw/

[3] http://www.cs.umd.edu/class/spring2003/cmsc838p/Process/waterfall.pdf

[4] http://www.openrdf.org/

[5] http://code.google.com/p/iwb/

[6] http://www.fluidops.net/