INFO/CS 4302 Web Information Systems

FT 2012
Week 1: Course Introduction

Who We Are - Instructors

Bernhard Haslhofer

bh392@cornell.edu

Office hours:
TUE / THU 1:30 - 3:00

Availability:
Oct, Nov, Dec 2012

Theresa Velden

tav6@cornell.edu

Office hours:
TUE / THU 1:30 - 3:00

Availability:
Sept, Nov, Dec 2012

Who We Are - TAs

Changchen He

ch627@cornell.edu

Office hours:
WED 4:30 - 6:00

Syed Ishtiaque Ahmed

sa738@cornell.edu

Office hours:
MON / FRI 5:00 - 6:30

Course Website / Piazza

http://www.infosci.cornell.edu/Courses/info4302/2012fa/

https://piazza.com/cornell/fall2012/infocs4302

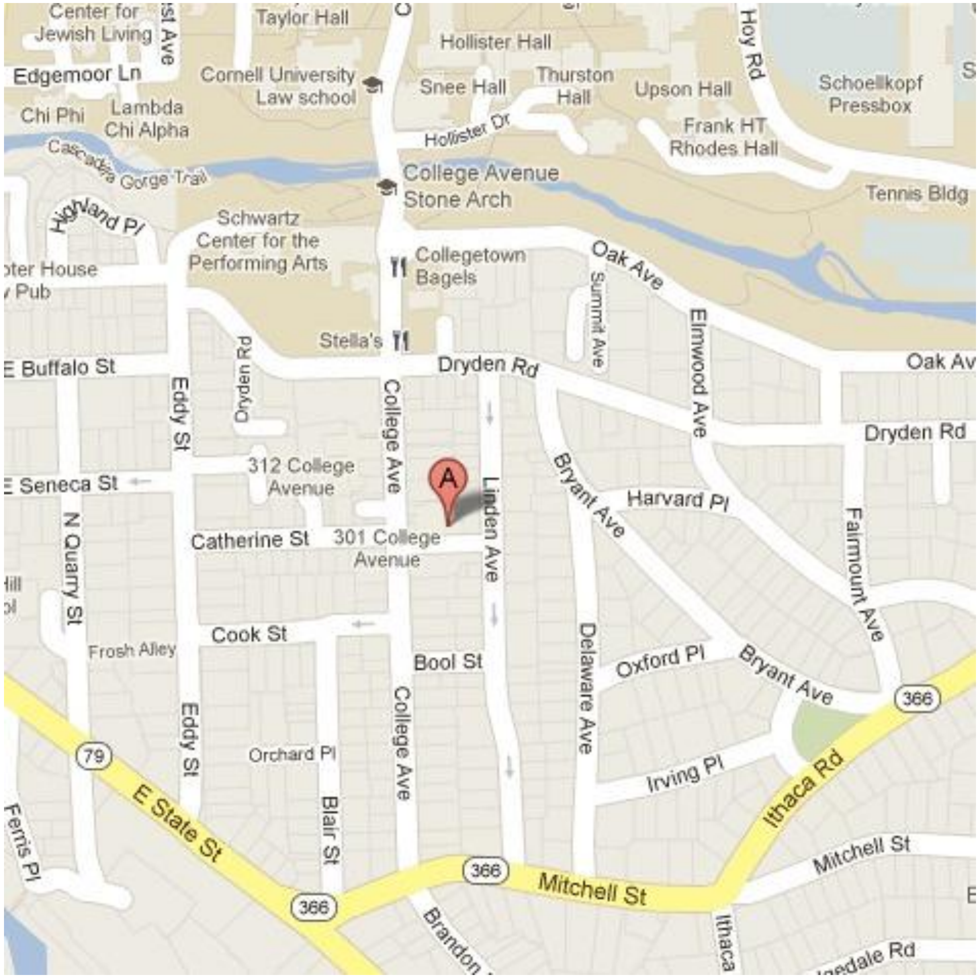
#info4302

Where you can find us

301 College Avenue

1st office on the left

Open space in IS



Our plan for today

Group-based class activity

What is this course about?

Review of course syllabus

Questions

My Web, Your Web... the web from your perspective

- Form groups of **4** with your immediate neighbours
- Within the next **10 minutes**, take turns to introduce yourself to your group and tell your classmates about

  1. What is your first memory of using the web, e.g. how old were you, what were you doing, what device were you using…?

  2. How has the web and your usage of the web evolved since then?

- **Take note** of similarities and differences in the experiences and preferences represented in your group. Anything that surprised you? Be ready to report back to the class

Our plan for today

Group-based class activity

What is this course about?

Review of course syllabus

Questions

What is this course about?

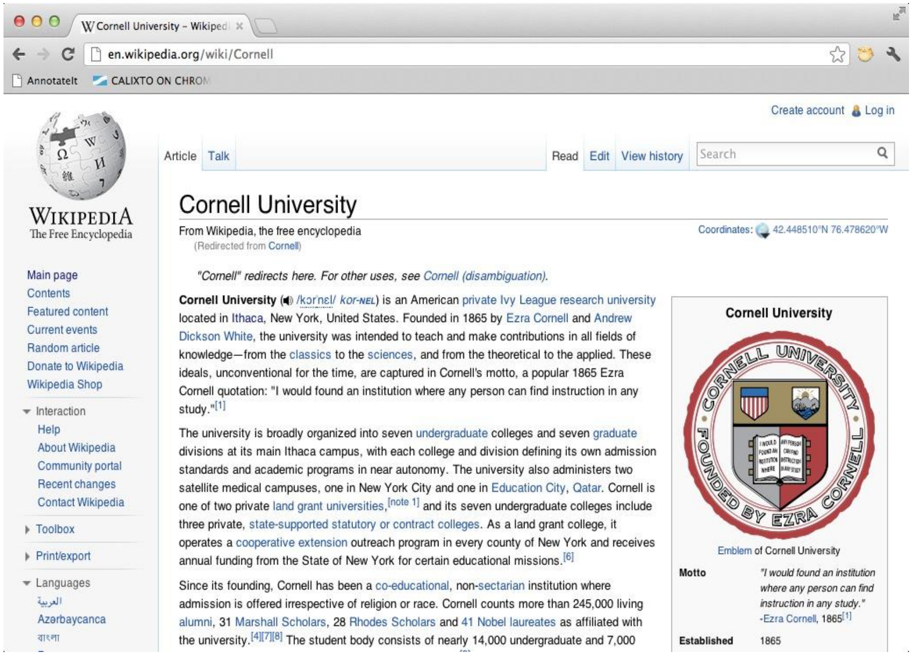
Web Information Systems

?

What is this course about?

the "Web" = "World Wide Web" = "WWW" =
"A system of interlinked documents accessed via the Internet"

Web Information Systems



Cornell University

From Wikipedia, the free encyclopedia
(Redirected from Cornell)

Coordinates: 42.448510°N 76.478620°W

"Cornell" redirects here. For other uses, see Cornell (disambiguation).

Cornell University (/kɔrˈnɛl/ *kor-NEL*) is an American private Ivy League research university located in Ithaca, New York, United States. Founded in 1865 by Ezra Cornell and Andrew Dickson White, the university was intended to teach and make contributions in all fields of knowledge—from the classics to the sciences, and from the theoretical to the applied. These ideals, unconventional for the time, are captured in Cornell's motto, a popular 1865 Ezra Cornell quotation: "I would found an institution where any person can find instruction in any study."[1]

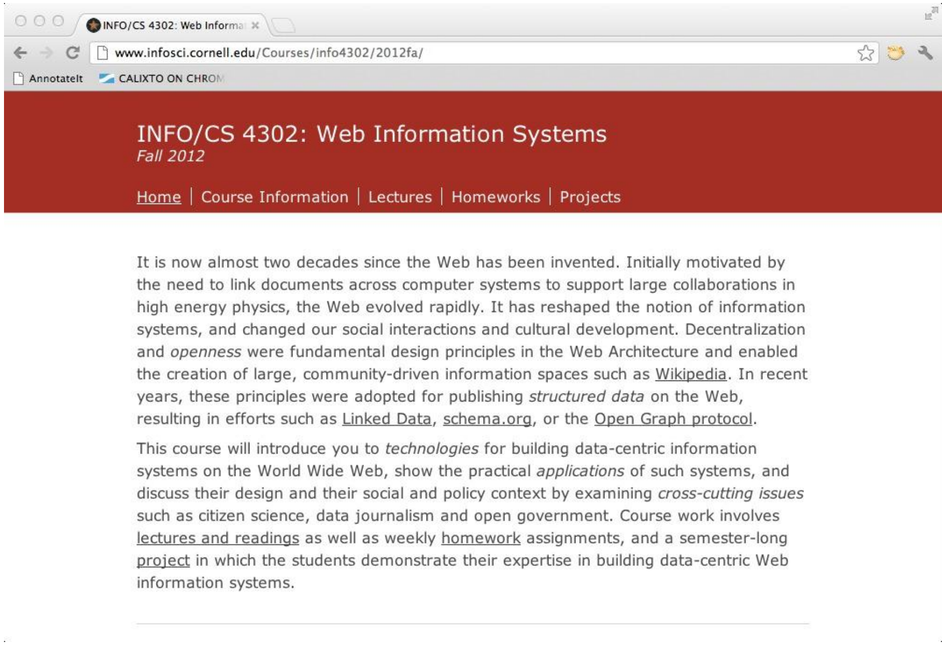
The university is broadly organized into seven undergraduate colleges and seven graduate divisions at its main Ithaca campus, with each college and division defining its own admission standards and academic programs in near autonomy. The university also administers two satellite medical campuses, one in New York City and one in Education City, Qatar. Cornell is one of two private land grant universities,[note 1] and its seven undergraduate colleges include three private, state-supported statutory or contract colleges. As a land grant college, it operates a cooperative extension outreach program in every county of New York and receives annual funding from the State of New York for certain educational missions.[6]

Since its founding, Cornell has been a co-educational, non-sectarian institution where admission is offered irrespective of religion or race. Cornell counts more than 245,000 living alumni, 31 Marshall Scholars, 28 Rhodes Scholars and 41 Nobel laureates as affiliated with the university.[4][7][8] The student body consists of nearly 14,000 undergraduate and 7,000

Cornell University

Emblem of Cornell University

Motto	*"I would found an institution where any person can find instruction in any study."* -Ezra Cornell, 1865[1]
Established	1865


INFO/CS 4302: Web Informa
www.infosci.cornell.edu/Courses/info4302/2012fa/
AnnotateIt
CALIXTO ON CHROM

INFO/CS 4302: Web Information Systems

Fall 2012

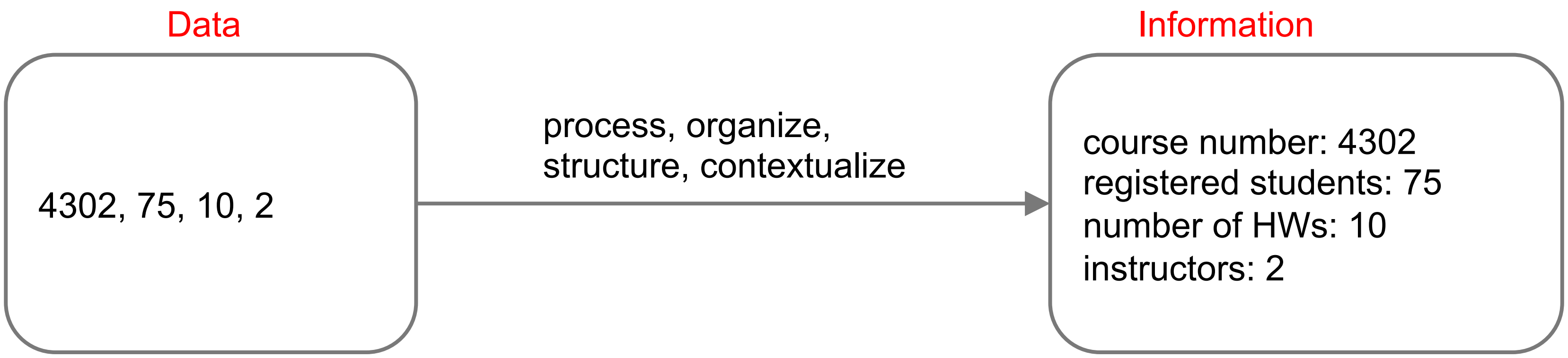
Home | Course Information | Lectures | Homeworks | Projects

It is now almost two decades since the Web has been invented. Initially motivated by the need to link documents across computer systems to support large collaborations in high energy physics, the Web evolved rapidly. It has reshaped the notion of information systems, and changed our social interactions and cultural development. Decentralization and *openness* were fundamental design principles in the Web Architecture and enabled the creation of large, community-driven information spaces such as Wikipedia. In recent years, these principles were adopted for publishing *structured data* on the Web, resulting in efforts such as Linked Data, schema.org, or the Open Graph protocol.

This course will introduce you to *technologies* for building data-centric information systems on the World Wide Web, show the practical *applications* of such systems, and discuss their design and their social and policy context by examining *cross-cutting issues* such as citizen science, data journalism and open government. Course work involves lectures and readings as well as weekly homework assignments, and a semester-long project in which the students demonstrate their expertise in building data-centric Web information systems.

What is this course about?

Web Information Systems



Data

4302, 75, 10, 2

process, organize,
structure, contextualize

Information

course number: 4302
registered students: 75
number of HWs: 10
instructors: 2

raw, unorganized facts

"useful" data

What is this course about?

Web Architecture

Data representation

Web Information Systems

Standards

Openness and decentralization

Tools and Frameworks

What is this course about?

Open Data

Web APIs

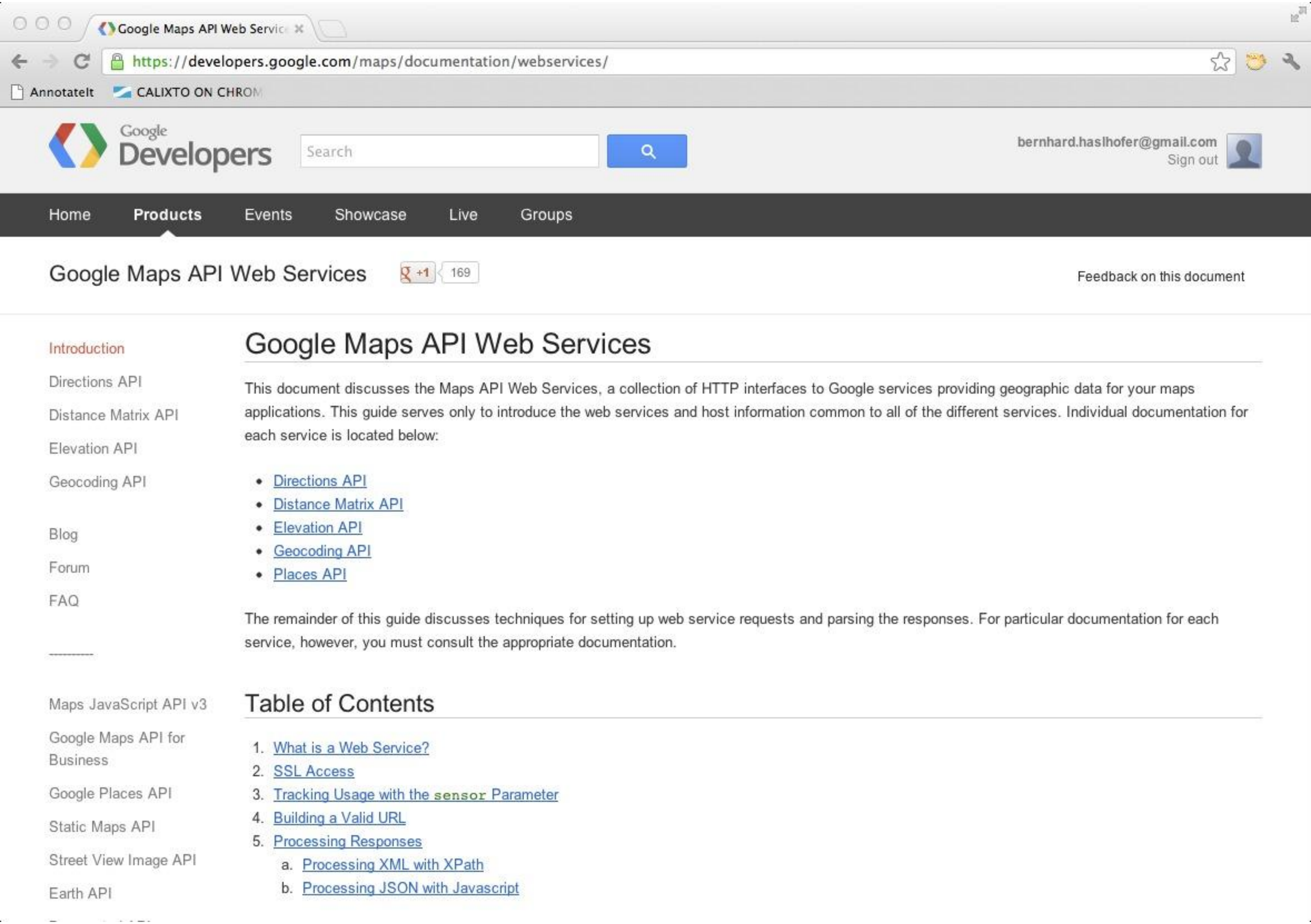
Global Data Networks

data-centric

Web Information Systems

Publishing Data on the Web

Using Data from the Web



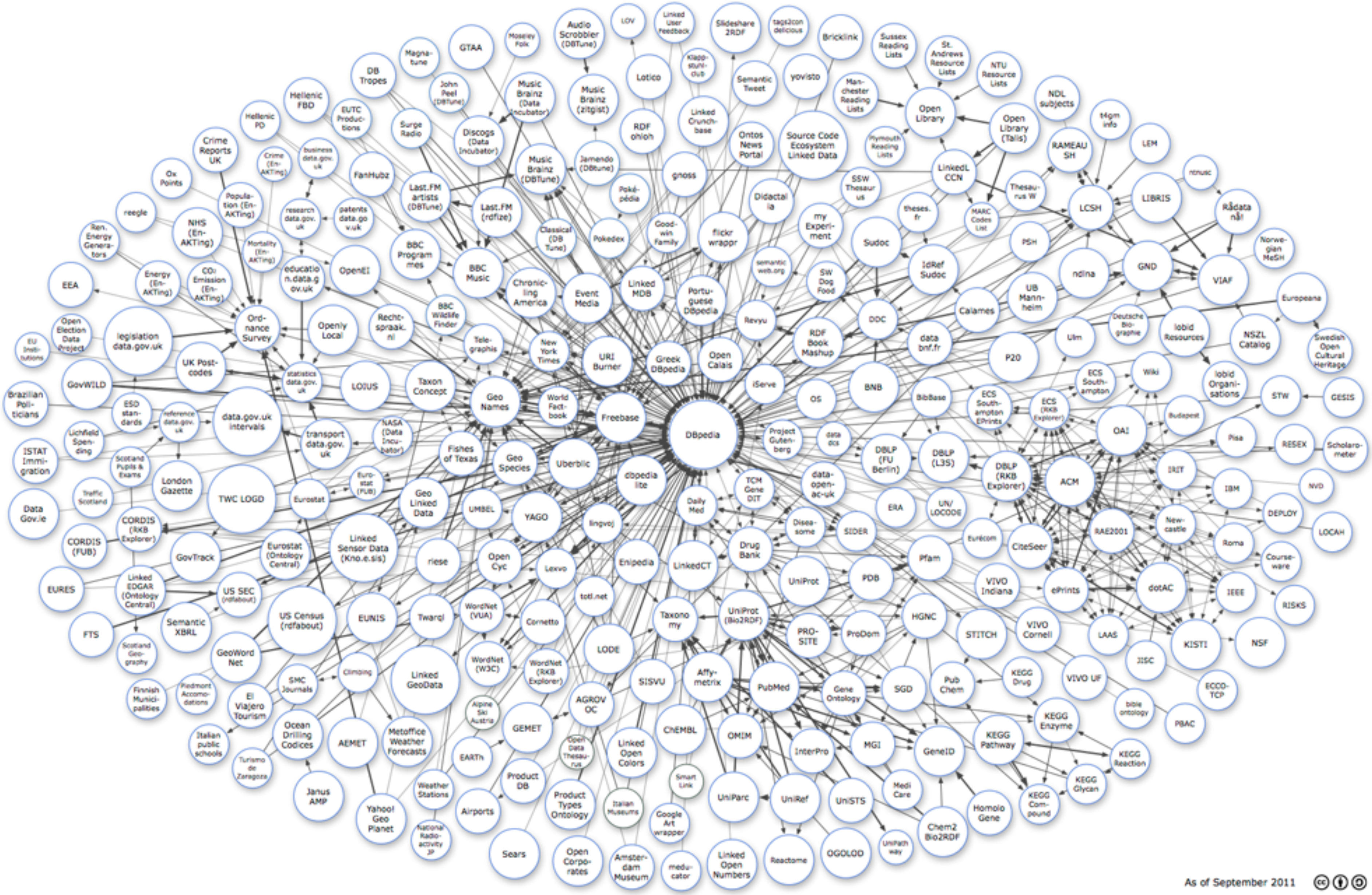
Google Maps API Web Services

This document discusses the Maps API Web Services, a collection of HTTP interfaces to Google services providing geographic data for your maps applications. This guide serves only to introduce the web services and host information common to all of the different services. Individual documentation for each service is located below:

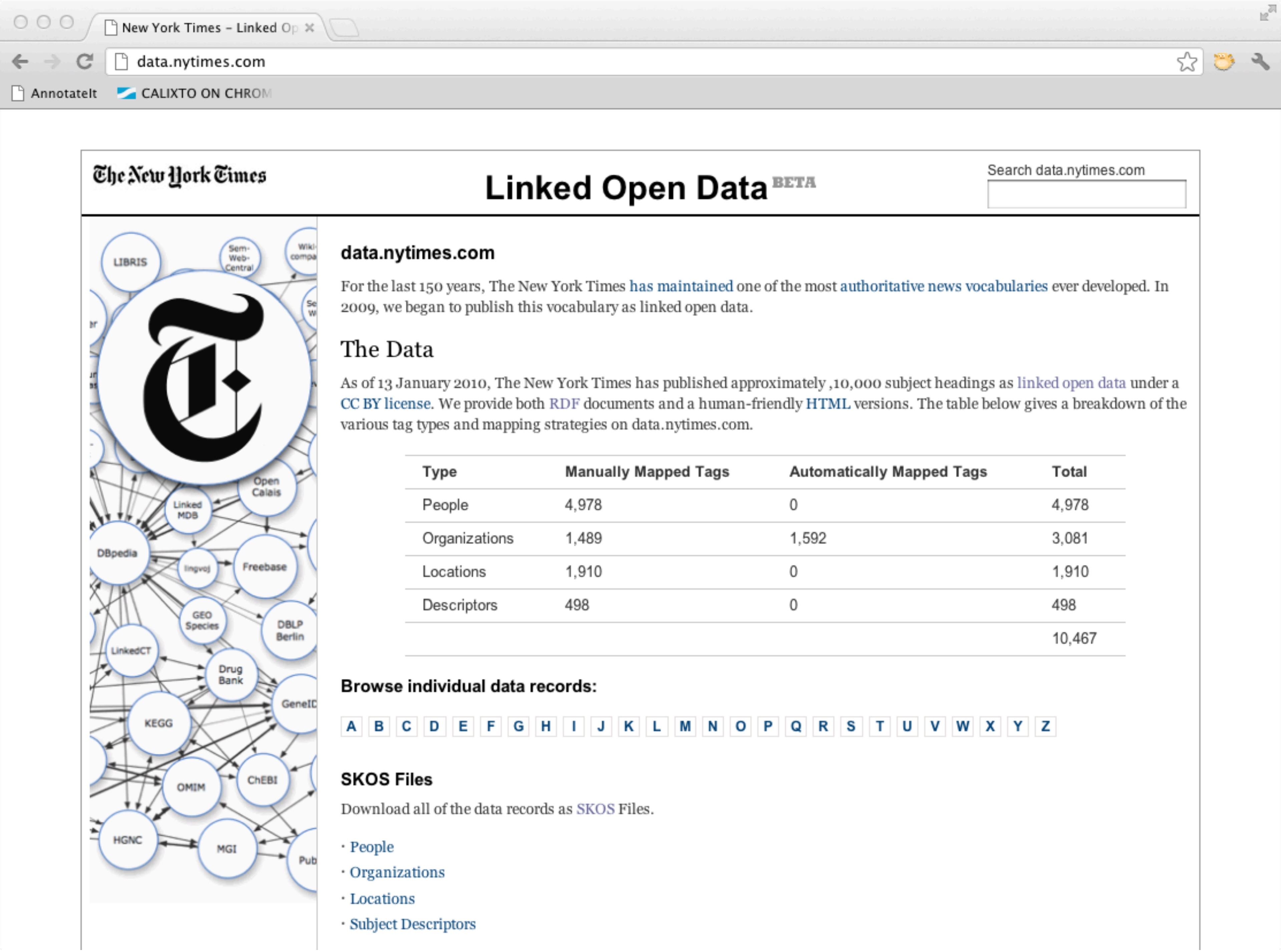
- Directions API
- Distance Matrix API
- Elevation API
- Geocoding API
- Places API

The remainder of this guide discusses techniques for setting up web service requests and parsing the responses. For particular documentation for each service, however, you must consult the appropriate documentation.

Table of Contents



As of September 2011



The New York Times

Linked Open Data BETA

data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately ,10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

Type	Manually Mapped Tags	Automatically Mapped Tags	Total
People	4,978	0	4,978
Organizations	1,489	1,592	3,081
Locations	1,910	0	1,910
Descriptors	498	0	498
			10,467

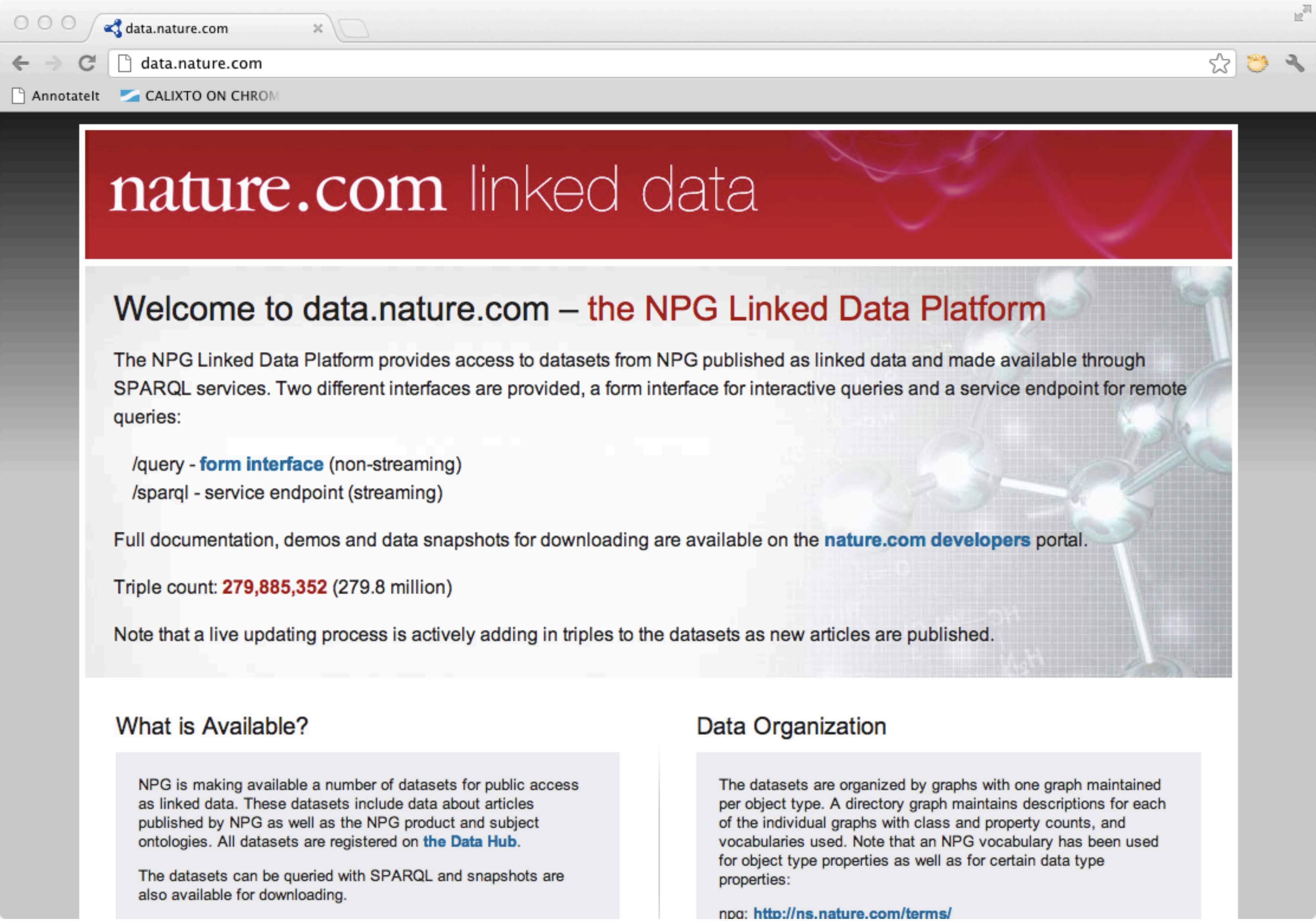
Browse individual data records:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SKOS Files

Download all of the data records as SKOS Files.

- People
- Organizations
- Locations
- Subject Descriptors



Welcome to data.nature.com – the NPG Linked Data Platform

The NPG Linked Data Platform provides access to datasets from NPG published as linked data and made available through SPARQL services. Two different interfaces are provided, a form interface for interactive queries and a service endpoint for remote queries:

/query - **form interface** (non-streaming)
/sparql - service endpoint (streaming)

Full documentation, demos and data snapshots for downloading are available on the **nature.com developers** portal.

Triple count: **279,885,352** (279.8 million)

Note that a live updating process is actively adding in triples to the datasets as new articles are published.

What is Available?

NPG is making available a number of datasets for public access as linked data. These datasets include data about articles published by NPG as well as the NPG product and subject ontologies. All datasets are registered on **the Data Hub**.

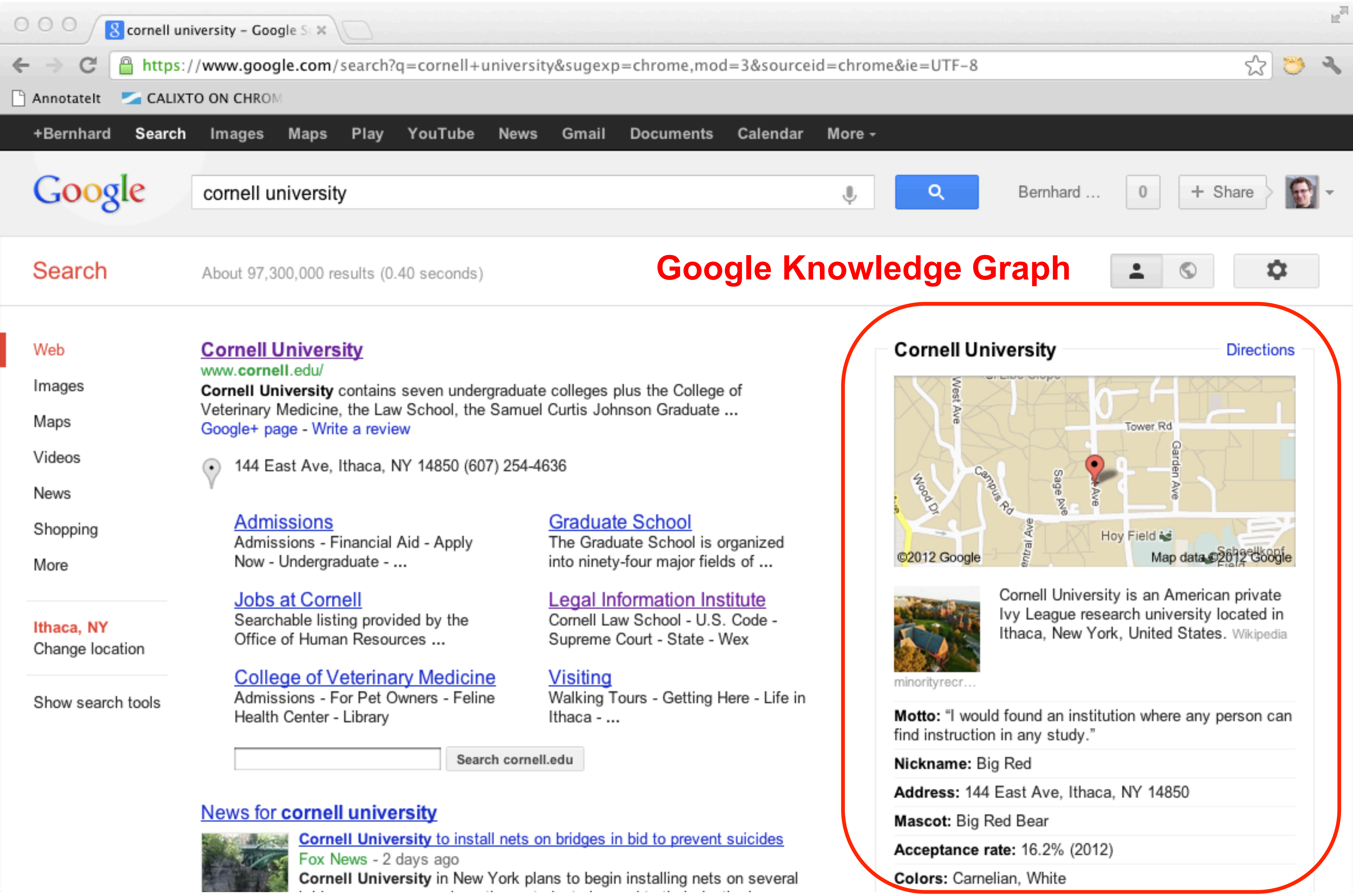
The datasets can be queried with SPARQL and snapshots are also available for downloading.

Data Organization

The datasets are organized by graphs with one graph maintained per object type. A directory graph maintains descriptions for each of the individual graphs with class and property counts, and vocabularies used. Note that an NPG vocabulary has been used for object type properties as well as for certain data type properties:

npg: **http://ns.nature.com/terms/**

Google Knowledge Graph



About 97,300,000 results (0.40 seconds)

Cornell University
www.cornell.edu/
Cornell University contains seven undergraduate colleges plus the College of Veterinary Medicine, the Law School, the Samuel Curtis Johnson Graduate ...
Google+ page - Write a review

144 East Ave, Ithaca, NY 14850 (607) 254-4636

Admissions
Admissions - Financial Aid - Apply Now - Undergraduate - ...

Graduate School
The Graduate School is organized into ninety-four major fields of ...

Jobs at Cornell
Searchable listing provided by the Office of Human Resources ...

Legal Information Institute
Cornell Law School - U.S. Code - Supreme Court - State - Wex

College of Veterinary Medicine
Admissions - For Pet Owners - Feline Health Center - Library

Visiting
Walking Tours - Getting Here - Life in Ithaca - ...

News for cornell university

Cornell University to install nets on bridges in bid to prevent suicides
Fox News - 2 days ago
Cornell University in New York plans to begin installing nets on several

Cornell University

Cornell University is an American private Ivy League research university located in Ithaca, New York, United States. Wikipedia

Motto: "I would found an institution where any person can find instruction in any study."

Nickname: Big Red

Address: 144 East Ave, Ithaca, NY 14850

Mascot: Big Red Bear

Acceptance rate: 16.2% (2012)

Colors: Carnelian, White

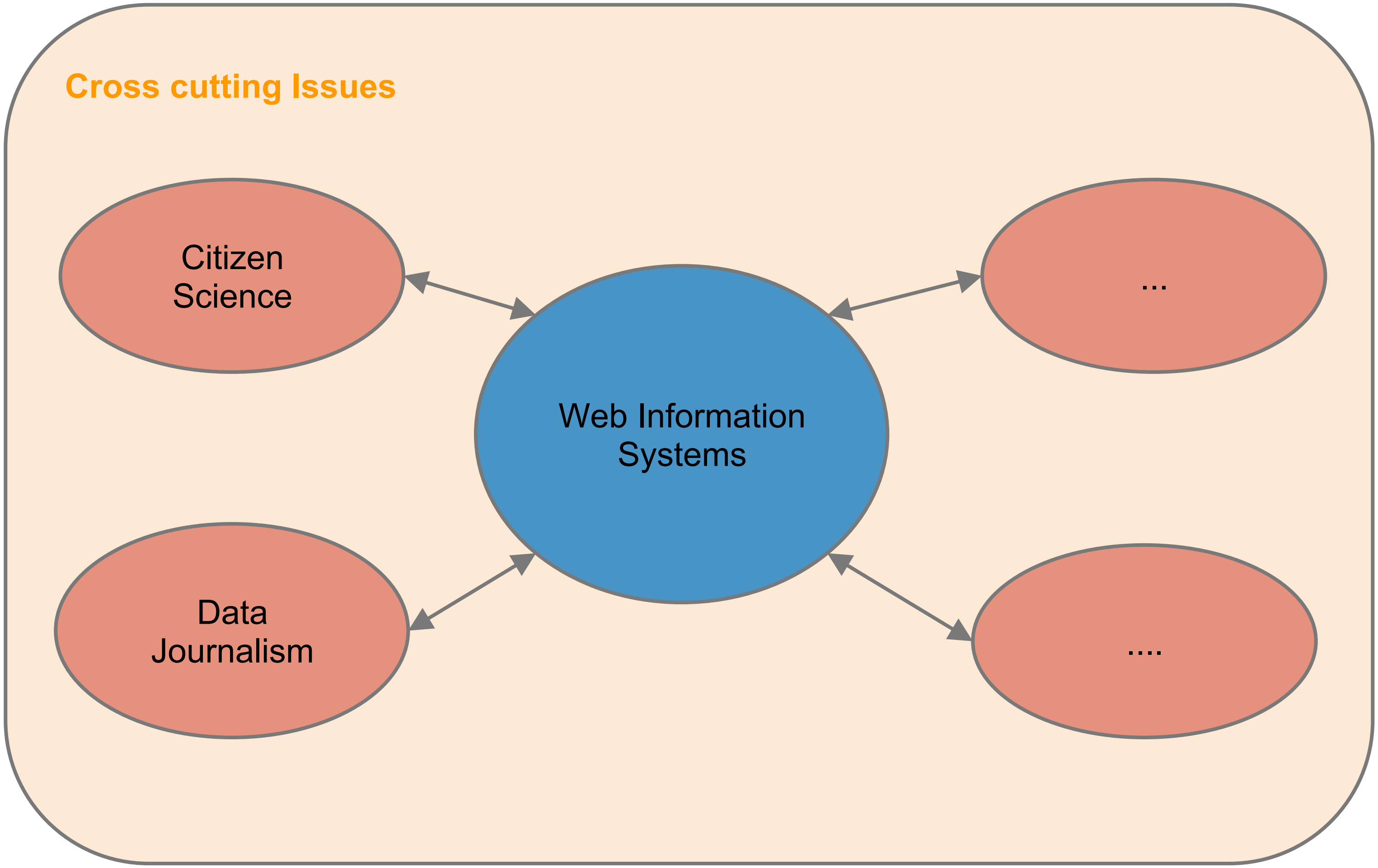
What is this course about?

http://developers.mystartup.com

http://mystartup.com/apis

http://data.mystartup.com

What is this course about?



What is this course not about?

Web site design (INFO 1300)

Web application development (INFO 2300)

Search and Retrieval (INFO 4300)

Our plan for today

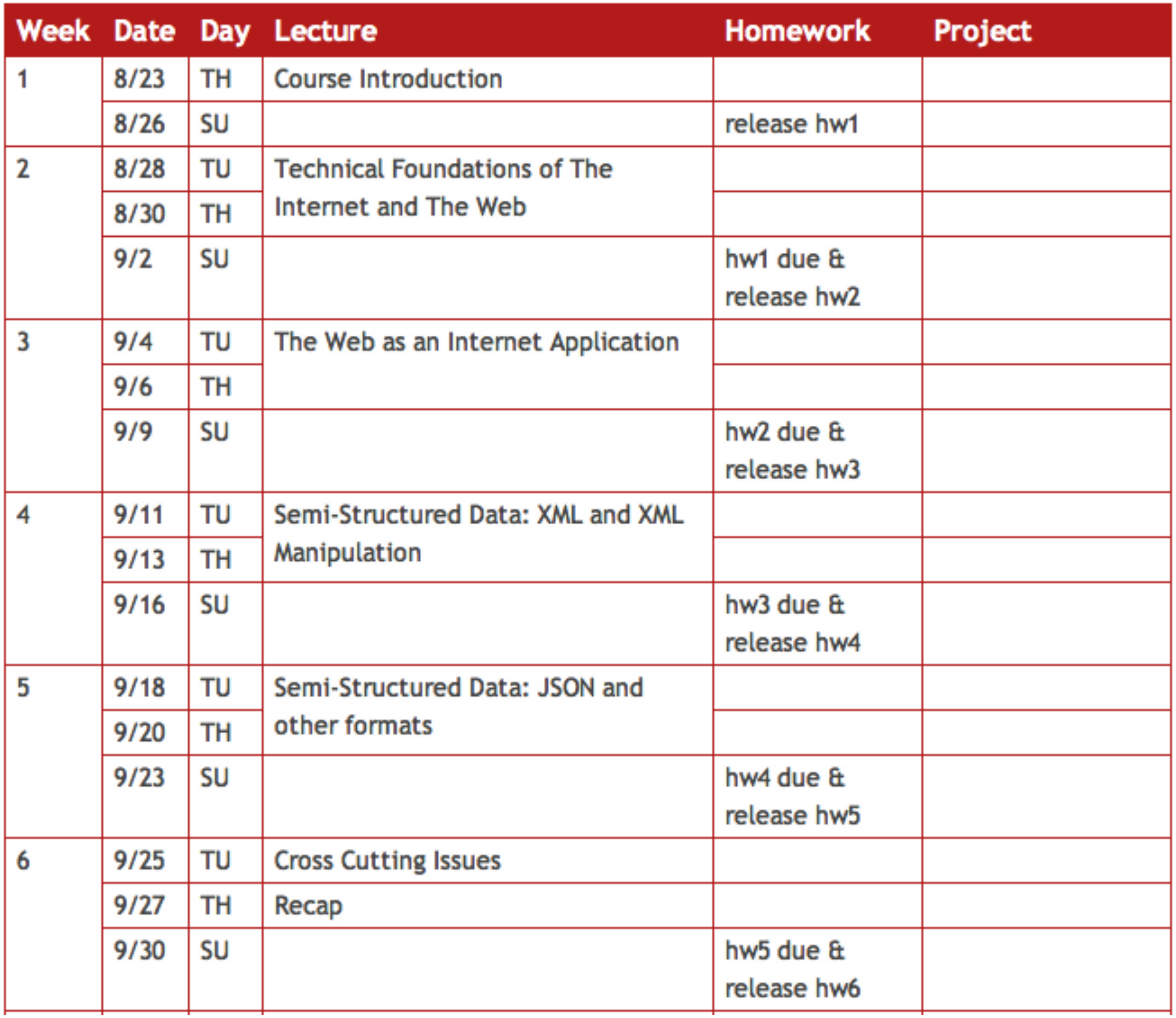
Group-based class activity

What is this course about?

Review of course syllabus

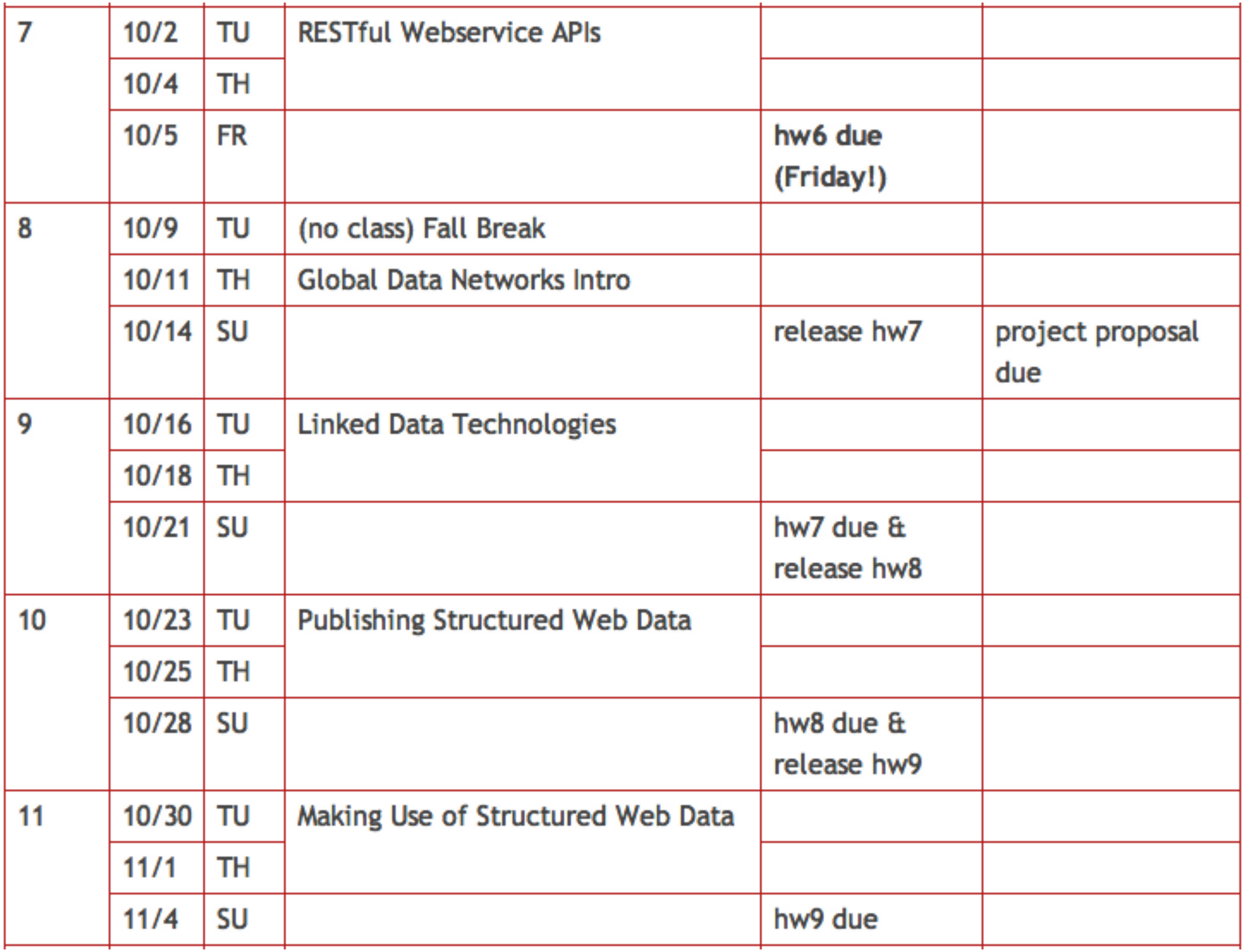
Questions

INFO/CS 4302 Syllabus



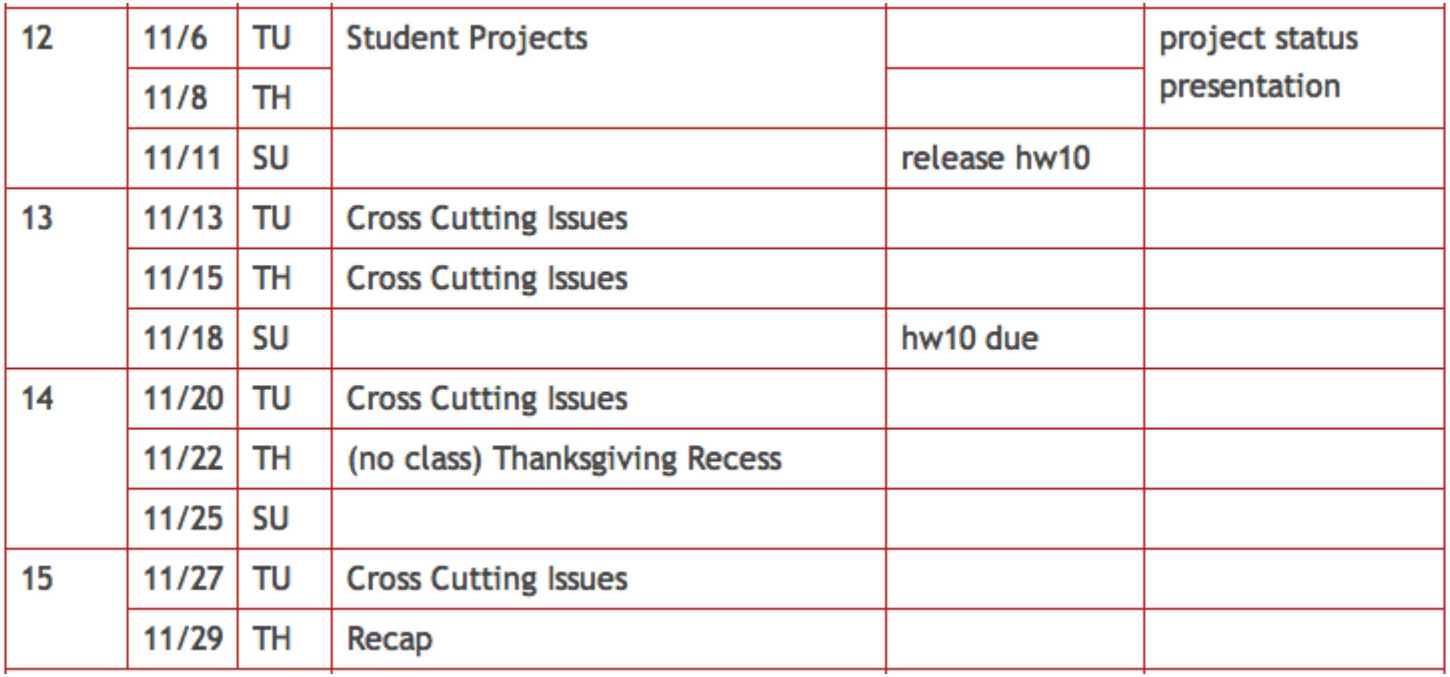
Week	Date	Day	Lecture	Homework	Project
1	8/23	TH	Course Introduction		
	8/26	SU		release hw1	
2	8/28	TU	Technical Foundations of The Internet and The Web		
	8/30	TH			
	9/2	SU		hw1 due & release hw2	
3	9/4	TU	The Web as an Internet Application		
	9/6	TH			
	9/9	SU		hw2 due & release hw3	
4	9/11	TU	Semi-Structured Data: XML and XML Manipulation		
	9/13	TH			
	9/16	SU		hw3 due & release hw4	
5	9/18	TU	Semi-Structured Data: JSON and other formats		
	9/20	TH			
	9/23	SU		hw4 due & release hw5	
6	9/25	TU	Cross Cutting Issues		
	9/27	TH	Recap		
	9/30	SU		hw5 due & release hw6	

INFO/CS 4302 Syllabus



7	10/2	TU	RESTful Webservice APIs		
	10/4	TH			
	10/5	FR		hw6 due (Friday!)	
8	10/9	TU	(no class) Fall Break		
	10/11	TH	Global Data Networks Intro		
	10/14	SU		release hw7	project proposal due
9	10/16	TU	Linked Data Technologies		
	10/18	TH			
	10/21	SU		hw7 due & release hw8	
10	10/23	TU	Publishing Structured Web Data		
	10/25	TH			
	10/28	SU		hw8 due & release hw9	
11	10/30	TU	Making Use of Structured Web Data		
	11/1	TH			
	11/4	SU		hw9 due	

INFO/CS 4302 Syllabus



12	11/6	TU	Student Projects		project status presentation
	11/8	TH			
	11/11	SU		release hw10	
13	11/13	TU	Cross Cutting Issues		
	11/15	TH	Cross Cutting Issues		
	11/18	SU		hw10 due	
14	11/20	TU	Cross Cutting Issues		
	11/22	TH	(no class) Thanksgiving Recess		
	11/25	SU			
15	11/27	TU	Cross Cutting Issues		
	11/29	TH	Recap		

Homeworks: due Sun night @ 11:59PM (CMS)
(exception: hw 6 due already Fri night @ 11:59 PM)

Requisite: CS 2110 or similar (object oriented programming, Java or python)

Group Projects

Groups of **3 students**.

Design a web information systems **of your choice**.

Detailed requirements on course website.

Milestones:

Kick-off in class: early September

Project proposal due: October 14th

Intermediary project presentation: November 6th + 8th

Final project presentation: December 13th from 5-9 pm

(make up: December 5th from 7-9:30pm, you need to let us know by August 31st)

Final project report due: December 13th @ 11:59pm

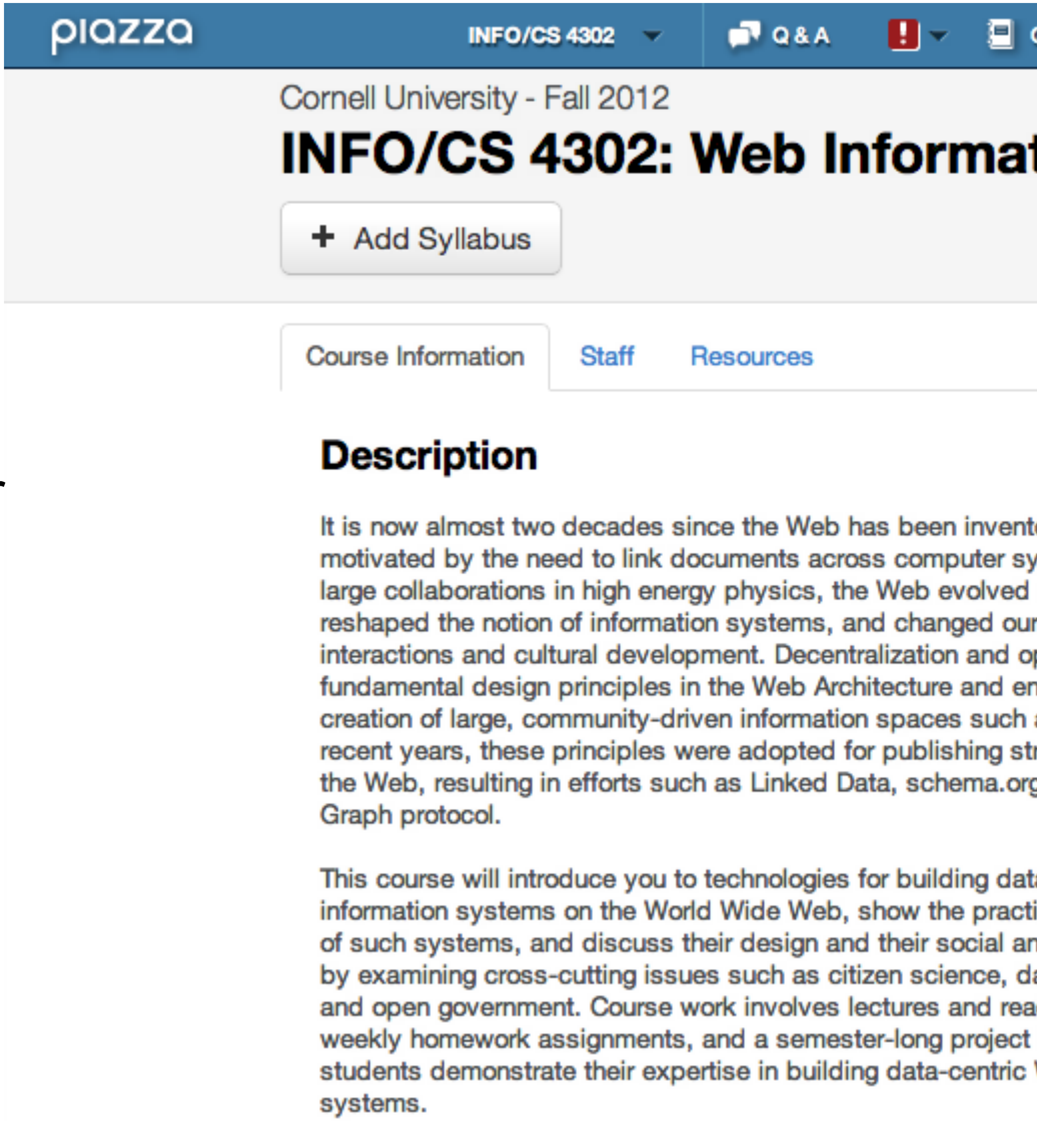
Participation & Support

Piazza:

- your questions on course content
- announcement of useful resources
- formation of teams, search for project partners
- your answers to challenges

Office Hours:

- offered every weekday @ 301 College Avenue



Cornell University - Fall 2012

INFO/CS 4302: Web Informat

Description

It is now almost two decades since the Web has been invent
motivated by the need to link documents across computer sy
large collaborations in high energy physics, the Web evolved
reshaped the notion of information systems, and changed ou
interactions and cultural development. Decentralization and o
fundamental design principles in the Web Architecture and e
creation of large, community-driven information spaces such
recent years, these principles were adopted for publishing st
the Web, resulting in efforts such as Linked Data, schema.or
Graph protocol.

This course will introduce you to technologies for building dat
information systems on the World Wide Web, show the practi
of such systems, and discuss their design and their social a
by examining cross-cutting issues such as citizen science, d
and open government. Course work involves lectures and rea
weekly homework assignments, and a semester-long project
students demonstrate their expertise in building data-centric
systems.

Email policy

- Send all questions about course content via piazza (**not** in personal email to instructors)
  - Quick turn-around
  - Others learn too, reducing overhead

Grading

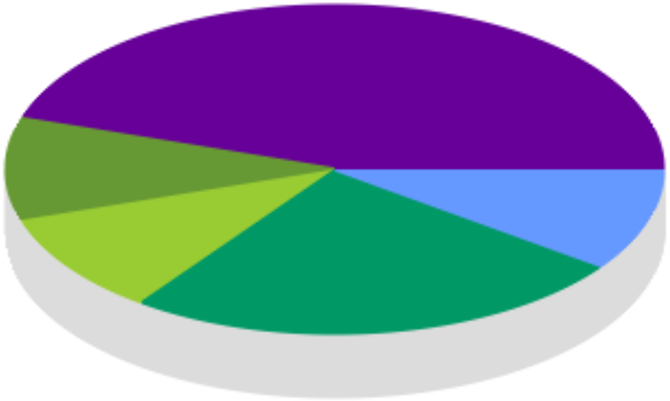
45 % 🟪
Homework Assignments

10 % 🟩
Project proposal

10 % 🟩
Mid Term Project Presentation

25 % 🟩
Final Project Presentation & Report

10 % 🟦
Participation



More course information:

- Academic integrity
    - Group assignments are meant to be worked on in groups. They are not meant to be done by one person without review and passed off as the group's work.
    - Individual assignments are meant to be worked on alone.
    - In both cases, looking things up and getting ideas from other sources is okay, if you cite it. Plagiarism (copying of others' work and attempting to pass it off as your own) is not.
- Lecture slides (posted after lecture)
- Instructions for submitting homework & code
- ….

http://www.infosci.cornell.edu/Courses/info4302/2011fa

Next week

Week 2: **History and Architecture of the Internet**

This Sunday (8/26): release of homework 1

Questions on readings:

1. *V. Bush. As We May Think; Atlantic Monthly; July 1945.*
2. *T. Berners-Lee et al. Creating a Science of the Web*
3. *T. Heath, C. Bizer*
   Linked Data: Evolving the Web into a Global Data Space, Chapters 1-3

?

?

Questions ?

?

?

?

?!?

?