The FAO Geopolitical Ontology: a reference for country-based information

Marta Iglesias-Sucasas, Soonho Kim[*] and Virginie Viollier[a]

Knowledge Management and Library Services Branch, Office of Knowledge Exchange, Research and Extension, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla 00153 Rome, Italy

Abstract

In 2002, the Food and Agriculture Organization of the United Nations (FAO) published its FAO Country Profiles, an innovative information retrieval system allowing access to FAO's thematic country-based information from one entry point. In this context and using some of the country profiles' underlying data, the FAO Geopolitical Ontology was publicly released, providing a reusable resource for the international community. The ontology was developed in OWL, a standardized semantics-oriented language encoding, and included relationships between geopolitical and geographic entities. The ontology aimed at improving information management, facilitating standardized data sharing and interoperability, and delivering the most updated geopolitical information. It was designed to manage coding systems and multilingual representations, and to track historical changes and use trustful sources. The FAO Geopolitical Ontology was then represented in RDF as part of the W3C Linked Open Data initiative. To facilitate the exploitation of the ontology reference data, various services were developed such as the Web services and the ontology module maker. The FAO Country Profiles portal concretely presents several geopolitical ontology use cases. This paper shows how and why the ontology became a reference for country-based information and, what the open challenges are for the future.

Keywords: FAO Geopolitical Ontology, Country-based information, ontology, Linked Open Data, FAO, FAO Country Profiles, OWL, RDF, Web services.

---

*Corresponding author.

a Authors are listed in alphabetical order. E-mail: marta.iglesas@fao.org, soonho.kim@fao.org, virginie.viollier@fao.org.

1. Introduction

In 2002, the Food and Agriculture Organization of the United Nations (FAO) published its FAO Country Profiles (http://www.fao.org/countryprofiles). This pioneering information retrieval system provided a fast and reliable way to access FAO's vast archive of country-specific information in the areas of agriculture and development from a single entry point. Decision-makers around the world could access information about countries without the need to search individual thematic databases and systems. The FAO Country Profiles gradually gave access to integrated thematic profiles and systems from trusted and heterogeneous international sources.

In the context of the FAO Country Profiles and using part of its underlying data, the FAO Geopolitical Ontology (http://aims.fao.org/aos/geopolitical.owl) was released in 2008, as a mechanism to connect diverse types of country-based information systems [1] [2]. Using an ontology to achieve interoperability is beneficial as it can provide semantic relationships between countries and geographical entities that can be exploited. Using a standardized semantics-oriented language to encode the ontology also provides a sharable and reusable resource for the worldwide community.

The FAO Geopolitical Ontology was developed to improve information management, to facilitate standardized data sharing and interoperability, and to deliver the most updated geopolitical information. The FAO Geopolitical Ontology describes and manages data related to geopolitical entities such as countries, territories, regions and other similar areas. It was designed to manage international coding systems and multilingual representations; and to implement relations to other geopolitical or geographic entities, track historical changes and use quality and trustful sources.

The ontology was first released in the OWL (Web Ontology Language) formal representation language, providing a machine-readable country-based information dataset. This was further developed and in March 2011, the FAO Geopolitical Ontology was represented in RDF (Resource Description Framework) format (http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/), embracing the W3C Linked Open Data (LOD) initiative [3]. To facilitate the exploitation and integration of the FAO Geopolitical

Ontology reference data by information systems, various services have been developed. The main ones are the Web services which retrieve specific data from the geopolitical ontology and return it in XML (Extensible Markup Language) format, hence facilitating results integration and reuse in other web sites and information systems. Another core service is the geopolitical ontology module maker, which provides a web interface to extract smaller modules of the ontology and to export them in various key formats (such as RDF, XML or CSV formats).

Today, the FAO Country Profiles portal concretely presents several geopolitical ontology use cases. This paper shows how and why the ontology became a reference for country-based information.

The objectives and special requirements of the FAO Geopolitical Ontology are explained in Section 2 of this paper. In Section 3, the implementation of the requirements for both the OWL and RDF representations under LOD principles is given. Section 4 includes the various services offered together with the FAO Geopolitical Ontology. Then, some future work and conclusions are proposed in Section 5.

2. Objectives and requirements of the FAO Geopolitical Ontology

The FAO Geopolitical Ontology describes and manages data related to geopolitical entities such as countries, territories, regions and other similar areas. It was first released in 2008 and published in the context of the FAO Country Profiles portal[2]. Since then, various new versions were published to include new resources and to keep it updated as its dynamic information evolved.

2.1. Objectives

The main objectives of the FAO Geopolitical Ontology are to improve information management, to facilitate standardized data sharing and to provide the most updated geopolitical information.

2.2. Requirements

The geopolitical ontology was designed to implement the following requirements [1].

2.2.1. Manage international coding systems

It was recognized that there were various coding systems used to identify territories and groups depending on the type of information managed. For example, most of statistical systems used M49[3] or FAOSTAT[4] codes; document management systems used ISO[5] codes or AGROVOC[6]; GIS[7] systems used GAUL[8]. In order to facilitate data exchange among those systems the ontology needed to manage all those codes and identifiers and to map them into each other.

2.2.2. Manage multilingual representations

There are various names associated with geopolitical entities. The most important are the official name recognized at UN level, the short name and the names to be used in lists. This information is available in the six FAO official languages, that are, Arabic, Chinese, French, English, Spanish and Russian, and sometimes also in Italian. An important requirement for the ontology was to implement the multilingual representation of all geopolitical entities whenever possible.

The ontology also needed to manage other important multilingual information such as currencies and adjectives of nationality.

2.2.3. Implement relations to other geopolitical or geographic entities

The ontology needed to implement mechanisms to manage information about land borders. For example, Spain has borders with Portugal and France.

Managing information about land borders facilitates information systems, which use the ontology, to implement features that compare data about neighbouring territories.

The ontology also needed to manage information about groups of countries and the group memberships. Groups of countries can be of different types, for example Asia is a geographic group while UEMOA (Union Economique et Monétaire Ouest Africaine) is an economic group. Managing group information is also very important to allow for comparison of data from countries in the same group or for data aggregation.

2.2.4. Tracking historical change

One important requirement dealing with geopolitical information is to manage historic changes. This is particularly crucial to be able to retrieve and reference information catalogued by country over time. The FAO Country Profiles team updates the content of the Geopolitical Ontology, (for example, adds a new country), only after receiving an official green light from the data owner (United Nations, FAO, etc.).

For example, it is not easy to find records about former countries such as Yugoslavia or Czechoslovakia and to easily relate those to the countries now occupying the same territories, in particular in legacy systems. Therefore a requirement for the geopolitical ontology was to be able to define the validity of a country or territory over time and to link it to predecessors or successors.

2.2.5. Structural description of data sources

Most of the reference data used in FAO information systems, and consequently to be managed in the ontology, come from trusted sources such as, FAO, other United Nations agencies, the International Organization for Standardization (ISO), the World Bank and Economic Groups' official web sites.

---

[3] Three-digit numerical codes of countries or areas, used for statistical processing purposes by the Statistics Division of the United Nations Secretariat http://unstats.un.org/unsd/methods/m49/m49.htm

[4] http://faostat.fao.org/default.aspx

[5]. http://www.iso.org/iso/country_codes.htm

[6] http://aims.fao.org/website/AGROVOC/sub

[7] Geographic information system (GIS)

[8] Global Administrative Unit Layers (GAUL) http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691

For example, all country names (official, short and list name) in the six official languages come from the Names of Countries System (NOCS)[9], maintained by the FAO terminology group in collaboration with UN terminology. The FAO Country Profiles team updates country names only after receiving an official green light by the NOCS team.

Therefore a requirement for the ontology was to be able to clearly specify the provenance and last update date for each reference included.

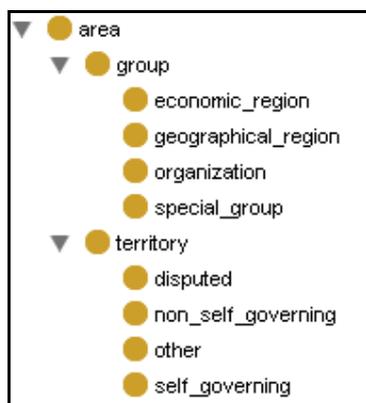## 3. Implementation of the FAO Geopolitical Ontology

3.1. Ontology modeling

The ontology was modeled using a bottom-up approach [4].

First, actual objects related to the geopolitical domain, such as, countries, non-self-governing territories, geographic regions, economic groups, etc. were collected and implemented as instances.

Then these instances were categorized into eight specific classes, based on their common properties. These were: self-governing territories, non-self-governing territories, disputed areas, other territories, economic regions, geographical regions, organizations, and special groups.

Later, three more general classes were added to create a hierarchical structure illustrated in Figure 1. These were: area, territory, and group.

Figure 1- The class hierarchy in the FAO Geopolitical Ontology

Relationships were implemented using two types of properties: data type properties, defining a relationship between an individual and a literal; and object properties, defining relationships between two individuals.

The geopolitical ontology implemented over 100 axioms in classes and properties. The most common axioms assigned to classes were the maximum cardinality and disjoint with. For example, each self-governing territory would have maximum one ISO3 code, and a self-governing territory is disjoint with a non-self-governing territory. The axioms assigned to properties were domain and range, inverse of, and symmetric properties.

3.1.1. International coding systems

The FAO Geopolitical Ontology manages international codes and identifiers using the following sub-properties of the property "hasCode": "codeISO2", "codeISO3", "codeUN", "codeUNDP", "codeGAUL", "codeFAOSTAT", "codeAGROVOC", "codeFAOTERM" and "codeDBPedia". All these properties have a maximum cardinality of 1, since it is required that they are unique. The international coding systems of Italy are shown in Figure 2 along with the "hasCode" property and sub-properties.

Figure 2- The hierarchy of "hasCode" property with international coding systems of Italy

3.1.2. Multilingual representations

There are various models to implement multilingual representation in ontologies. The most common is to use "xml:lang" tag. In this case a datatype property would have as many values as instances of the xml:lang. In the case of the geopolitical ontology, one important requirement is that the master copy must have only one value per language. This is easier to check using the maximum

cardinality function. For this reason, the geopolitical ontology assigned different datatype properties to each language version of a property and each datatype property had the maximum cardinality = 1. For example, as shown in Figure 3, for the official name of a territory, the ontology has nameOfficialEN for the English name, nameOfficialFR for the French name, nameOfficialES for Spanish, etc.

hasOfficialName	
■ nameOfficialAR	جمهورية فييت نام الاشتراكية
■ nameOfficialEN	the Socialist Republic of Viet Nam
■ nameOfficialES	la República Socialista de Viet Nam
■ nameOfficialFR	la République socialiste du Viet Nam
■ nameOfficialIT	Repubblica socialista del Vietnam
■ nameOfficialRU	Социалистическая Республика Вьетнам
■ nameOfficialZH	越南社会主义共和国

Figure 3-The hierarchy of "hasOfficialName" property with official names of Viet Nam

The same approach was used to manage all multilingual representations of short names, names for lists, currencies, and adjectives of nationality[10].

3.1.3. *Relations among geopolitical or geographic entities*

The ontology manually implemented the land border relationship between two countries using the "hasBorderWith" object property, which is also an inverse property. For example,

```
"Spain" "hasBorderWith" "Portugal"
```

where the inverse, "Portugal" "hasBorderWith" "Spain" is also true.

Membership to all types of groups in the ontology (geographic, economic, organization, and special) is managed using the "isInGroup" and "hasMember" object properties. These two object properties are the inverse of each other. Example:

```
"France" "isInGroup" "Europe"
```

which is the inverse of:

```
"Europe" "hasMember" "France"
```

[10] The official adjectives of nationality at FAO: i.e. Both "the Democratic Republic of the Congo" and "the Republic of the Congo" use the same adjective of nationality, "Congolese".

3.1.4. *Historic changes*

FAO Geopolitical Ontology manages historic changes using a combination of two object properties, "isPredecessorOf" and "isSuccessorOf", and two datatype properties, "validSince" and "validUntil".

The "isPredecessorOf" and "isSuccessorOf" properties are the inverse of each other and are used to identify which territories or groups were the predecessor/successor of one another.

The properties "validSince[11]" and "validUntil[12]" are used to identify the period of validity for any territory or group.

For example, the Russian Federation is the successor of The Union of Soviet Socialist Republic (USSR), and therefore the ontology specifies that:

```
"Russian Federation" "isSucessorOf"
                "USSR"
```

The ontology also specifies the inverse property, using "isPredecessorOf". In addition this information is complemented with the periods of validity for the Russian Federation and the USSR, where the "validUntil" property of the USSR and "validSince" property of the Russian Federation are both equal to 1991.

3.1.5. *Structural description of data sources*

The geopolitical ontology uses annotation properties to provide information about the different data sources, and in particular it reuses the well-known Dublin Core vocabulary[13], dcterms:source [14],

[11] The value of the datatype property "validSince" gives the starting year of validity of the area (territory or group) it is associated with. The geopolitical ontology tracks historic changes only since 1985. Therefore if an area has a validSince = 1985, this indicates that the area is valid since 1985 or before.

[12] The value of the datatype property "validUntil" gives the last year of validity of the area (territory or group) it is associated with. In case the area is currently valid, this value is set by default to 9999.

[13] http://dublincore.org/

to define the general source information. In addition, other three annotation properties, "sourceCreator", sourceIdentifier", and "SourceModified", are created to specify the name of the source, the URI and the last update. This set of source annotations is used in the ontology to identify the data sources for classes, instances and properties. Figure 4 illustrates an example of source description for the country area statistics from FAOSTAT[15].

dcterms:source	FAOSTAT http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang-	
sourceCreator	FAOSTAT	
sourceIdentifier	http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang=en#ancor	
sourceIdentifier	http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang=en#ancor	en
sourceIdentifier	http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang=es#ancor	es
sourceIdentifier	http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang=fr#ancor	fr
sourceModified	2010-09-09	

Figure 4- Source descriptions for Country Area statistics

3.2. OWL version

The main version of the FAO Geopolitical Ontology is written in OWL as it allows to implement axioms, such as, disjoint with and maximum cardinality, that are crucial to manage the master copy of the geopolitical data coming from multiple sources. The OWL version of the FAO Geopolitical Ontology is available at: http://aims.fao.org/aos/geopolitical.owl

3.3. RDF representation of the Geopolitical Ontology

In March 2011, an RDF representation of the geopolitical ontology was published. It was developed following the requirements for the Linked Open Data.

The OWL version remains the main representation, and it is the one updated as new data is available. The RDF representation is generated programmatically from the OWL version.

The program that uses the OWL version and generates the RDF representations produces four outputs:

- RDF dump (the entire dataset written in RDF/XML)

---

[14]http://dublincore.org/2010/10/11/dcterms.rdf#source

[15] http://faostat.fao.org/default.aspx

- RDF/XML representation for each resource
- HTML+RDFa representation for each resource
- Dereferencing resource URIs

As illustrated in Figure 5, the Jena ontology API[16] was used for retrieving data from the online OWL format of the FAO Geopolitical Ontology.

The Restlet framework was used for:

1) Getting request from clients

2) Generating the proper output - either RDF/XML or HTML+RDFa based on the requests

3) Redirecting the requests of non-information resources into HTML+RDFa representation.

The Jena ontology API and the Restlet framework were also used for implementing the services described in Section 4.
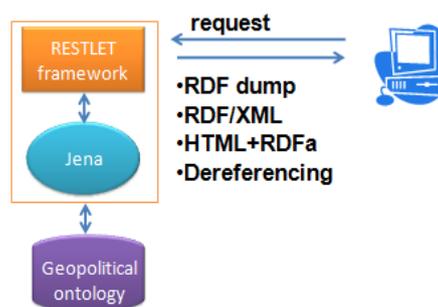

Figure 5-Architecture of the RDF service

The RDF representation of the geopolitical ontology is compliant with all requirements to be included in the Linked Open Data cloud, as explained in the following paragraphs.

Resolvable http:// URIs:

Every resource in the OWL format of the FAO Geopolitical Ontology has a unique URI. Dereferenciation was implemented to allow for three different URIs to be assigned to each resource as follows:

- URI identifying the non-information resource[17]

---

[16] http://jena.sourceforge.net/ontology/

[17] Non-information resource means "real world objects", comparing to information resource such as documents, images, and other media files.

- Information resource with an RDF/XML representation
- Information resource with an HTML representation

In addition the current URIs used for OWL format needed to be kept to allow for backwards compatibility for other systems that are using them. Therefore, the new URIs for the FAO Geopolitical Ontology in LOD were carefully created, using "Cool URIs for Semantic Web"[18] and considering other good practices for URIs, such as DBpedia URIs.

New URIs:

The URIs of the geopolitical ontology need to be permanent, consequently all transient information, such as year, version, or format was avoided in the definition of the URIs.

The new URIs can be accessed at

http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/

For example, for the resource "Italy" the URIs are the following:

http://www.fao.org/countryprofiles/geoinf/geopolitical/resource/Italy: identifies the non-information resource.

http://www.fao.org/countryprofiles/geoinf/geopolitical/data/Italy: identifies the resource with an RDF/XML representation.

http://www.fao.org/countryprofiles/geoinf/geopolitical/page/Italy: identifies the information resource with an HTML representation.

In addition, the "owl:sameAs" is used to map the new URIs to the OWL representation.

Dereferencing URIs:

When a non-information resource is looked up without any specific representation format, then, the server needs to redirect the request to information resource with an HTML representation.

For example, to retrieve the resource "Italy" (http://www.fao.org/countryprofiles/geoinf/geopolitical/resource/Italy), which is a non-information

resource, the server redirects to the html page of "Italy" (http://www.fao.org/countryprofiles/geoinf/geopolitical/page/Italy).

At least 1000 triples in the datasets:

The total number of triple statements in FAO Geopolitical Ontology is 22,495.

At least 50 links to a dataset already in the current LOD Cloud:

FAO Geopolitical Ontology has 195 links to DBpedia, which is already part of the LOD Cloud.

Access to the entire dataset:

FAO Geopolitical Ontology provides the entire dataset as a RDF dump. It is available at http://www.fao.org/countryprofiles/geoinfo/geopolitical/data

The RDF representation of the FAO Geopolitical Ontology has been already registered in CKAN[19](http://ckan.net/package/fao-geopolitical-ontology) and it was requested to add it into the LOD Cloud.

4. Services

The FAO Geopolitical Ontology is a unique resource that serves as reference to users who deal with country-based information such as UN agencies, decision-makers, government agencies, researchers, etc. It can be used by users working on new or existing information systems that manage data including country information. Adapting a master data in order to benefit from the content of the geopolitical ontology may offer major improvements. Therefore various services were developed to assist users.

The Geopolitical Ontology module maker (http://www.fao.org/countryprofiles/geoinfo/modulemaker/index.html) allows the extraction of parts of the ontology (for example a module extracts all countries while another one extracts current groups and countries). Data is then available in various interoperable formats which can easily be imported and reused into information systems. Users willing to integrate data extracted from the FAO Geopolitical Ontology into their systems, do not need to change

---

[18]   http://www.w3.org/TR/2007/WD-cooluris-20071217/

[19] http://ckan.net/

their existing technology. They simply extract the data needed in a suitable format to be imported.

In other cases, information systems may already have a suitable mechanism to update their country data references. They just need to exploit part of the available information included in the Geopolitical Ontology, without adding complexity (for example, users may not be interested in including in their database information related to groups or neighbouring countries). The Geopolitical Ontology Web services (http://www.fao.org/countryprofiles/webservices.asp) allow information systems to retrieve and use on demand, the complementary data they just need.

4.1. Module maker

The geopolitical ontology module maker service allows to extract modules of the FAO Geopolitical Ontology and to export them in RDF, XML or CSV formats. The web interface is shown in Figure 6.

The module maker provides five built-in modules allowing the extraction of:

- Self-governing territories only
- Self-governing territories and groups
- FAO members
- FAO members and the groups they belong to
- All territories and groups valid the current year

The module maker is available at http://www.fao.org/countryprofiles/geoinfo/modulemaker/index.html

Figure 6- Module maker in the FAO Country Profiles

4.2. Web services

The geopolitical ontology Web services allow users to retrieve atomic pieces of information from the geopolitical ontology. The Web services return data in XML format, to facilitate results integration or re-use in other web sites and information systems. Figure 7 shows the 28 Web services available in the FAO Country Profiles portal

http://www.fao.org/countryprofiles/webservices.asp?lang=en

Figure 7- FAO Geopolitical Ontology Web services

The FAO Country Profiles portal implements various geopolitical ontology use cases. The most innovative are the Land borders and group and the Key statistics.

The portal uses the geopolitical ontology Web services to retrieve land border countries and geographic and economic groups for any selected country. Then, the information is incorporated in the country pages using an XSLT transformation, as shown in Figure 8 and 9.

Figure 8- Land border countries of Uruguay in FAO Country Profiles, retrieved from FAO Geopolitical Ontology

Figure 9- Geographic and Economic groups which Uruguay belongs to

FAO Geopolitical Ontology includes key country statistics and indicators, such as, Country area, Agricultural area, Land area, Population, GDP, and Human Development Index. The FAO Country Profiles retrieves these country statistics in XML

using the relevant geopolitical ontology Web services (http://www.fao.org/countryprofiles/geoinfo/ws/countryStatistics/countryArea/KEN, and http://www.fao.org/countryprofiles/geoinfo/ws/countryStatistics/agriculturalArea/KEN, etc.) and transforms the result into HTML representation and inserts the result in the respective country pages, as shown in Figure 10.
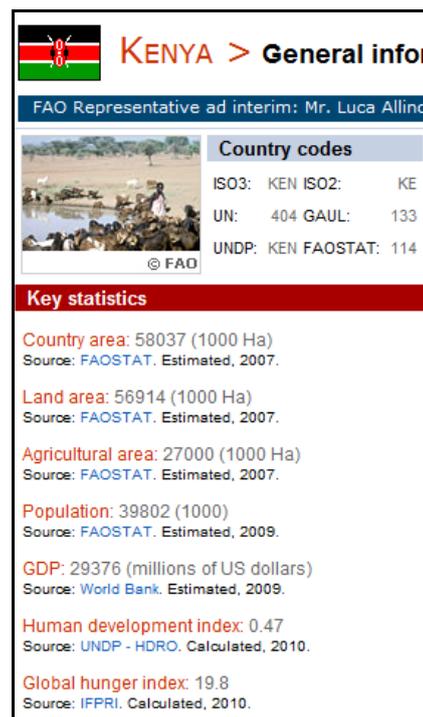

Figure 10- Key statistics for Kenya

5. Conclusions and future work

From simple HTML and then OWL version, the FAO Geopolitical Ontology has evolved over the past 3 years into a full-fledged RDF application. The ongoing development of the FAO Geopolitical Ontology is one of the priorities of the FAO Country Profiles' team. New selected resources available not only at FAO, but also in the LOD cloud will be evaluated and integrated.

Moreover, the implementation of an additional service will start shortly. Indeed to enable users and systems to query the ontology online and explore the RDF data, a SPARQL (SPARQL Protocol and RDF Query Language) endpoint will be developed.

Adding selected datasets and providing new services to users will confirm the FAO Geopolitical Ontology's role as a reference for country-based information.

References

[1] C. Caracciolo, M. Iglesias, J. Keizer, (2006) Towards Interoperability of Geopolitical Information within FAO, *Semantic e-Science workshop, China 2006.*, ftp://ftp.fao.org/docrep/fao/009/ag867e/ag867e00.pdf

[2] S. Kim, M. Iglesias-Sucasas, C. Caracciolo, J. Keizer, (2009) Integrating country-specific heterogeneous data at the United Nations: FAO geopolitical ontology and services, *Semantic Universe*, http://semanticweb.com/integrating-country-specific-heterogeneous-data-at-the-united-nations-fao-s-geopolitical-ontology-and-services_b10681

[3] T. Berners-Lee, Linked Data - Design Issues, http://www.w3.org/DesignIssues/LinkedData.html

[4] N.F. Noy and D. L. McGuinness, (2001) Ontology Development 101: A Guide to Creating Your First Ontology". *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880*

[5] R. T. Fielding, R. N. Taylor, (2002), Principled Design of the Modern Web Architecture, *ACM Transactions on Internet Technology* (TOIT) (New York: Association for Computing Machinery) 2 (2): 115–150