Structured Data To RDF I
Deliverable D4.3.1
Version FINAL

Authors: W.R. Van Hage[1], T. Ploeger[1]
Affiliation: (1) SynerScope B.V.

NewsReader

POST HOC ERGO PROPTER HOC

COOPERATION

Grant Agreement No.	316404
Project Acronym	NEWSREADER
Project Full Title	Building structured event indexes of large volumes of financial and economic data for decision making.
Funding Scheme	FP7-ICT-2011-8
Project Website	http://www.newsreader-project.eu/
Project Coordinator	Prof. dr. Piek T.J.M. Vossen VU University Amsterdam Tel. + 31 (0) 20 5986466 Fax. + 31 (0) 20 5986500 Email: piek.vossen@vu.nl
Document Number	Deliverable D4.3.1
Status & Version	FINAL
Contractual Date of Delivery	October 2013
Actual Date of Delivery	December 12, 2013
Type	Report
Security (distribution level)	Public
Number of Pages	52
WP Contributing to the Deliverable	WP7
WP Responsible	SynerScope B.V.
EC Project Officer	Susan Fraser
Authors: W.R. Van Hage[1], T. Ploeger[1]	
Affiliation: (1) SynerScope B.V.	
Keywords: structured data, rdf, conversion	
Abstract: In this deliverable we describe the conversion of four data sets to RDF for use within the NewsReader project. These data sets are intended to supplement the event indexes extracted from news articles. We describe TechCrunch, CrunchBase, the World Bank Indicators, and Yahoo! Finance. We show what approaches are available for converting structured data to RDF and how we applied them.	

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	07 Nov 2013	Deliverable skeleton, First draft of Introduction	Thomas Ploeger	All
0.2	08 Nov 2013	First draft of Description of Datasets	Thomas Ploeger	2
0.3	26 Nov 2013	Finalized Description of Datasets	Thomas Ploeger	2
0.4	27 Nov 2013	First draft of Available Conversion Methods	Thomas Ploeger	3
0.5	28 Nov 2013	Finalized Available Conversion Methods, first draft of Conversion Details	Thomas Ploeger	3, 4
0.6	29 Nov 2013	Finalized Conversion Details, finalized Conclusion and Future Work	Thomas Ploeger	4, 5
0.7	29 Nov 2013	Review	Willem R. Van Hage	All
0.8 / FINAL	12 Dec 2013	Processed EHU Feedback	Thomas Ploeger	All

Executive Summary

The NewsReader project aims to support decision making by building structured event indexes of large volumes of news articles and financial data. To provide additional context, it is necessary to supplement the event indexes with additional data sets.

TechCrunch is a set of news articles about tech startups. CrunchBase is a database of startups, people, and financial organizations that serves as the structured data companion to TechCrunch. The World Bank Indicators are statistical indicators of development (such as GDP or number of hospitals) for countries world wide. Yahoo! Finance provides historical stock prices.

To be stored in the NewsReader KnowledgeStore, these datasets must be converted to RDF. Several methods for converting existing structured data to RDF exist: It is possible to write a custom script, use an off-the-shelf tool, or to take advantage of an existing conversion.

We convert each data set to RDF using a method appropriate for that data set and present the results in this deliverable for discussion.

Contents

List of Figures

1 Introduction

The NewsReader project aims to support decision making by building structured event indexes of large volumes of news articles and financial data. In this deliverable we describe the conversion of four existing structured data sets to the Resource Description Framework (RDF). These data sets are intended to supplement the event indexes with additional context.

The reader is assumed to have at least a basic understanding of RDF (including vocabularies, ontologies, and named graphs), JSON, CSV, and the NewsReader project in general.

The four datasets and their purpose within the NewsReader project are described below. These data sets need to be converted to RDF because this is the format the NewsReader KnowledgeStore (see Deliverable 6.1) is designed for.

TechCrunch[1] A news website about information technology companies.

CrunchBase[2] A database of technology companies, people, and investors. Together with TechCrunch, this dataset will be used in the evaluation (Deliverable 8.2.1) of the first versions of the decision support systems (Deliverable 7.3.1).

World Bank Development Indicators[3] Per-country statistical indicators of development and quality-of-life. This dataset will be used to supplement the event indexes with developmental context.

Yahoo! Finance[4] Historical prices of individual stocks as well as stock market indexes. This dataset will be used to supplement the event indexes with financial context.

The reasons for selecting specifically these data sets over other similar data sets are described in Deliverable 1.1: Definition of Data Sources. In that deliverable, the selection criteria are explained in detail, together with several example usage scenarios for the data sets.

In Section 2 we describe the data sets in more detail. We give an overview of available methods for converting structured data to RDF in Section 3. Section 4 contains the details of the actual conversion process for each individual data set. This deliverable concludes with an overview of fixes and improvements planned for Deliverable 4.3.2 in Section 5.

The conversion as described in this deliverable is prototypical. Besides simply describing the conversion, this deliverable is also intended to fuel discussion about what future conversions of the data sets should look like. The datasets and the specifics of their conversion are therefore subject to change during the course of the NewsReader project. The final conversion process will be described in Deliverable 4.3.2.

---

[1]http://www.techcrunch.com
[2]http://www.crunchbase.com
[3]http://www.worldbank.org
[4]http://finance.yahoo.com

2   Description Of Datasets

Below, each of the data sets listed in Section 1 is described in more detail. After a short overview of the data set, we describe the data structure in terms of the entities it contains, their properties, and the relationships between them. To indicate the size of the data set we include basic counts of entities and relationships. Next, we show how and under which license the data set was acquired. Finally, we present the reasons for the dataset being selected for conversion and use within the NewsReader project.

2.1   TechCrunch

TechCrunch is a news website that reports on the activity of information technology companies. A typical TechCrunch news article features a product launch, a major investment, a merger, an acquisition, or an IPO. TechCrunch was founded by Michael Arrington in 2005 and was acquired by AOL in 2010 [TechCrunch, 2013].

2.1.1   Data Structure

TechCrunch publishes news articles in English. As its basic properties, each article has an author, a publication date, a title, a body text, and a URL. Most articles come with a link to the profile page of their author as well as links to the CrunchBase (see Section 2.2) profiles of any entities (e.g. companies, persons) mentioned in the article. In addition, an article may be associated with a set of tags indicating the topic of the article.

2.1.2   Data Size

At the time of writing, we have 43.384 articles from TechCrunch. Of those articles, 43.212 (99.6%) have a link to their author. 28595 (65.9%) have at least one link to an entity in CrunchBase. 35.881 (82.7%) articles have at least one tag associated with them. Note that these numbers describe a specific dump of TechCrunch that was last updated on the 15th of August in 2013. Existing articles may have been removed, and new articles will have been added since then.

2.1.3   Acquisition And License

The TechCrunch articles were scraped by project partner ScraperWiki[5] (SCW) using their own Web scraping platform. SCW built an index of the articles to scrape at TechCrunch by iterating over site pages based on their knowledge of the underlying WordPress[6] content management system. Once they had collected this list of articles they parsed each one to obtain the body text, time, title, and any links in the document.

After scraping, the data was made available in the form of three individual Character Separated Value (CSV)-files: One containing just the URLs of all scraped articles, another

---

[5]https://scraperwiki.com/
[6]http://wordpress.org/

containing the URL, full body text, publication time, and title, and yet another containing the URL, author, tags, and CrunchBase-links. Example rows from each of these CSV-files can be seen in the raw data examples in Appendix A.1.

TechCrunch is made available under AOL's terms of service[7]. In these terms of service, it is specified that all content on any website owned by AOL (including TechCrunch) is protected by copyright, owned by AOL, and permission is required before using the content. Fortunately, for noncommercial use no permission is required as long as the copyright notices are retained in the document [AOL, 2013].

2.1.4 Purpose

Together with CrunchBase, this dataset will be used in the evaluation (see Deliverable 8.2.1) of the first versions of the decision support systems (see Deliverable 7.3.1). Additionally, the TechCrunch articles and CrunchBase data might be used in the evaluation of the event extraction pipelines (see Deliverable 4.2.1/2/3). The idea is that the event extraction pipelines will be scored based on their ability to 'reproduce' the CrunchBase data from TechCrunch articles.

2.2 CrunchBase

CrunchBase is a database of information on information technology companies, people, financial organizations, service providers, and products. It is essentially the structured data companion to TechCrunch. CrunchBase aims to "make information about the startup world available to everyone and maintainable by anyone" [CrunchBase, 2013a]. CrunchBase is developed and hosted by TechCrunch, but anyone can edit its contents in a Wikipedia-like manner.

2.2.1 Data Structure

CrunchBase contains data on five different entities:

Company Typically a commercial organization, but non-profits, schools, and other types of organizations are present as well. Examples include Facebook, Microsoft, and SynerScope.

Person A real human. Examples include Mark Zuckerberg, Bill Gates, and Jan-Kees Buenen.

Financial Organization Typically a bank or a venture capital firm. Examples include Goldman Sachs, ING, and 5 Park Lane.

Service Provider Includes PR firms, designers, legal counsel, and so on. Examples include Baker & McKenzie, Kasman Design, and Schox Patent Group.

---

[7]http://legal.aol.com/terms-of-service/full-terms/

Entity	Number	%
Company	161.915	41.83
Person	183.184	47.32
Financial Organization	10.006	2.58
Service Provider	6.238	1.61
Product	25.754	6.65
Total	387.097	100

Table 1: Counts of entities in CrunchBase.

Product A product produced by a company. Examples include Facebook, Xbox, and Google Chromebook.

Each of these entities has a set of properties, ranging from very simple (e.g. name) to more complex (e.g. investment made by that entity, with all relevant details). The list of properties per entity is too large to reproduce here verbatim, but they can be seen in the raw data examples in Appendix A.2.

CrunchBase also keeps track of the relationships between different entities. There are 6 types of relationships between entities in CrunchBase, also shown in Figure 1:

Acquisition Only exists between companies. One company can be acquired by another.

Competition Only exists between companies. One company can be the competitor of another.

Investment Exists between a company and other companies, people, and financial organization. Each of the latter 3 can invest in a company.

Providership Exists between a company or financial organization and a service provider. A service provider can provide services to a company or a financial organization.

Relationship Exists between a person and a company, financial organization, or service provider. Indicates that the person works for the entity in question.

Product Only exists between a company and a product. Indicates that the product was developed by the company in question.

Like the entities, each of these relationships has its own set of properties, again ranging from simple to complex. They can also be seen in the raw data examples in Appendix A.2.

2.2.2 Data Size

Table 1 shows the number of entities in CrunchBase. Note that these numbers describe a specific dump of CrunchBase that was downloaded on the 19th of August in 2013. Existing entities may have been removed, or new entities may have been added since then.
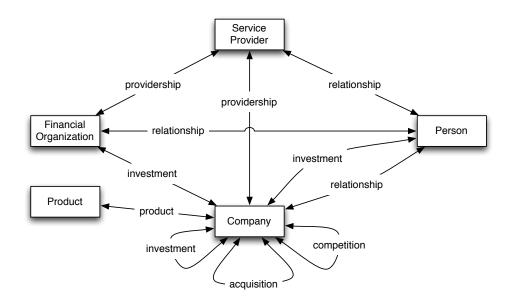
Figure 1: Diagram of relationships between CrunchBase entities. Notice that all arrows are bidirectional, indicating that the relationships are 'stored' at both entities.

2.2.3 Acquisition And License

The CrunchBase data was downloaded through a REST API[8] provided by CrunchBase. This API can be asked for lists of identifiers for each company, person, financial organization, service provider, and product available in CrunchBase. By iterating over these lists, it is possible to sequentially query the API for the data of each entity. The API returns the data in JSON format, which is stored on disk in an individual file for each entity. Examples of these JSON-files can be seen in the raw data examples in Appendix A.2.

CrunchBase's content is available under the Creative Commons Attribution License [CrunchBase, 2013b] (CC-BY[9]). The only requirement is that there is a link back to CrunchBase from any page that uses CrunchBase data.

2.2.4 Purpose

Together with TechCrunch, this dataset will be used in the evaluation (see Deliverable 8.2.1) of the first versions of the decision support systems (see Deliverable 7.3.1). Additionally, the TechCrunch articles and CrunchBase data might be used in the evaluation of the event extraction pipelines (see Deliverable 4.2.1/2/3). The idea is that the event extraction pipelines will be scored based on their ability to 'reproduce' the CrunchBase data from TechCrunch articles.

---

[8] http://developer.crunchbase.com/
[9] http://creativecommons.org/licenses/by/2.0/

2.3 World Bank Development Indicators

The World Bank is an organization that aims for the global reduction of poverty. The organization does this by providing financial and technical assistance to developing countries. To support these activities, the World Bank collects large amounts of data. Part of this data collection effort is the compilation of development indicators. These indicators are essentially cross-country comparable statistics on development, such as average daily income or number mobile phone subscriptions [The World Bank, 2013b].

2.3.1 Data Structure

The indicators are available in several formats, of which the tabular format is perhaps the most intuitive. In this format, columns represent the value of that indicator for a certain year and rows represent the value for a certain country. What follows are several example indicators for each category, with their unit in parentheses.

Education Number of teachers in primary education (total), ratio of female to male enrollment (%), literacy rate (% of population).

Environment CO2 emissions (kt), access to electricity (% of population), forest area (sq. km).

Economic Policy & Debt Exports of goods and services (annual % growth), use of IMF credit (US$), current account balance (% of GDP).

Financial Sector Real interest rate (%), consumer price index (2005 = 100), Inflation, consumer prices (annual %).

Health Life expectancy at birth (years), population (total), hospital beds (per 1,000 people).

Infrastructure Mobile cellular subscriptions (total), motor vehicles (per 1,000 people), container port traffic (number of TEU).

Labor & Social Protection Long-term unemployment (% of total unemployment), emigration rate of tertiary educated (% of total tertiary educated population), generosity of all social safety nets (%).

Poverty GINI index, income share held by highest 10%, income share held by lowest 10%.

Private Sector & Trade Commercial service exports (US$), international tourism, expenditures (US$), time required to start a business (days).

Public Sector Battle-related deaths (number of people), intentional homicides (per 100,000 people), proportion of seats held by women in national parliaments (%).

2.3.2 Data Size

The World Bank Indicator data is available for 214 countries. The Indicators have a yearly granularity. The earliest recorded Indicators are from 1960 and at the time of writing the latest indicators are from 2012. In total, there are 1300 indicators available, in several categories (as seen above).

2.3.3 Acquisition And License

The World Bank Indicator data can be acquired through the World Bank's REST API in CSV, XML, or JSON format. Indicator data can be requested per country, returning a list of yearly values of that Indicator for that country. An example of an Indicator in CSV-format can be found in Appendix A.3.

The World Bank explicitly encourages the use of their data for any beneficial purpose [The World Bank, 2013a]. The only requirement is that there is a link back to The World Bank from any page that uses their data.

2.3.4 Purpose

The Indicator data will be used to supplement the events extracted in the NewsReader project with developmental context. After all, events in the news might be caused by certain changes in the development of a country (e.g. rise in consumer price index, or rise of inflation).

2.4 Yahoo! Finance

Yahoo! Finance is a website that aggregates financial news articles from several sources (e.g. The Wall Street Journal, The New York Times), publishes stock market data (e.g. stock quotes, stock exchange rates), and allows users to keep track of their personal stock portfolio. Yahoo! Finance is the largest financial website in the United States [Stelter, 2012].

2.4.1 Data Structure

Yahoo! Finance keeps track of a number of properties per stock symbol. The values of these properties are available for every day a particular stock symbol is tradable (i.e. those days that the stock market a particular stock symbol is traded on is open for business).

Open Opening price (in USD, 2 decimal precision), i.e. the price at the moment the stock market opens at the beginning of the day.

High Highest price (in USD, 2 decimal precision) during the day.

Low Lowest price (in USD, 2 decimal precision) during the day.

Close Closing price (in USD, 2 decimal precision), i.e. the price at the moment the stock market closes at the end of the day.

Volume Number of shares in trade.

Adjust Close Closing price (in USD, 2 decimal precision), but accounting for all corporate actions such as stock splits, dividends/distributions, and rights offerings.

2.4.2   Data Size

There is no straightforward method for listing and then retrieving the prices for all the stock symbols available on Yahoo! Finance [Rassom, 2012]. Unlike for CrunchBase where you can ask for a list of available entities, it is not possible to query for a list of all available stock symbols, and then retrieve the prices for those stock symbols. It is necessary to prepare a list of stock symbols to retrieve in advance. This makes it impossible to say anything definitive about the number of stock symbols available in Yahoo! Finance.

2.4.3   Acquisition And License

The historical prices for an individual stock symbol can be acquired by searching for that stock symbol on Yahoo! Finance. This will present the user with a table of historical prices. On this page, there is also a button to download the table as a CSV-file. This button appears to be powered by a REST API that accepts a stock symbol and a date range as parameters. There is no further documentation available for this API, but at least it allows for easy retrieval of the full historical stock prices for a company. An example of the historical stock prices in CSV-format can be found in Appendix A.4.

   Yahoo's web page regarding permissions[10] for using the financial data downloadable from their Finance web page is unavailable (404), making it impossible to say anything definitive about our right to use it.

2.4.4   Purpose

The stock symbol price data will be used to supplement the events extracted in the News-Reader project with financial context. Changes in stock symbol prices may be result of an event (such as a new product announcement), or - vice versa - an event might be caused by a change in stock price, such as a shareholders meeting.

3   Available Conversion Methods

There are several methods for converting structured data to RDF. One option is to use an existing, off-the-shelf conversion tool that is specifically built for converting a specific format of structured data to RDF. It is also possible to write a custom, one-off script that

---

[10]`http://pressroom.yahoo.net/pr/ycorp/permissions.aspx`

is able to read the existing data, apply some transformations, and output the transformed data as RDF triples. Alternatively, there is chance that someone else has already converted the data to RDF. In the following sections we give a more detailed overview of each of these approaches.

3.1 Off-the-shelf Tool

The World Wide Web Consortium (W3C) maintains a list[11] of available RDF converter tools on their Wiki. There are tools for a great variety of input formats, such as make-files, log files from code management systems, or even UML class diagrams. Because of the prevalence of relational data, a significant number of tools are specifically designed for converting relational data to RDF. Examples of such tools are Sparqlify[12] and D2RQ[13].

Recently, the W3C has been working on the RDB2RDF standard[14], a standardized language for mapping relational data to RDF. The standard defines a Direct Mapping, which defines a simple transformation without any user input. This Direct Mapping can be used directly or to bootstrap a custom transformation defined in the R2RML language. The RDB2RDF standard is implemented in a few tools, such as db2triples[15] and Ultrawrap[16].

3.2 Custom Script

If there is no suitable off-the-shelf tool available for the data format that needs to be converted, writing a custom ETL (Extract, Transform, and Load) script to perform the conversion is often the only remaining option. Most scripting languages can be coerced into reading data in virtually any format. The remaining challenge then is the generation of actual RDF triples. Fortunately, popular scripting languages such as Python[17] and Ruby[18] have excellent RDF libraries (such as RDF.rb[19]) making the process of generating RDF triples much less painful.

3.3 Re-use or Adapt Existing

It is possible that the existing data has already been converted to and made available as RDF by someone else. It would not be in the spirit of the linked open data effort[20] to **not** reuse this existing RDF data.

---

[11]http://www.w3.org/wiki/ConverterToRdf
[12]http://sparqlify.org/
[13]http://d2rq.org/
[14]http://www.w3.org/2001/sw/rdb2rdf/
[15]https://github.com/antidot/db2triples/
[16]http://www.capsenta.com/
[17]http://www.python.org/
[18]http://www.ruby-lang.org/
[19]http://rdf.rubyforge.org/
[20]http://en.wikipedia.org/wiki/Open_data

Of course, it is possible that the existing conversion does not meet some application specific requirements. Still, it is probably be more efficient to adapt or transform the existing RDF rather than duplicate the entire conversion process.

4 Conversion Details

Having described the data sets and possible methods for their conversion, we can take a more detailed look at the actual conversion of each individual data set. For each data set, we will motivate our choice for a certain conversion method and describe the relevant details of the conversion process.

4.1 TechCrunch

The TechCrunch articles were converted to RDF using a simple Ruby script. This approach was chosen because the article data is spread across multiple CSV-files which need to be combined in a highly specific manner: One file ("articles") contains just the URLs of all scraped articles, another ("parsed") contains the URL, full body text, publication time, and title, and yet another ("links") contains the URL, author, tags, and CrunchBase-links.

The first step in the conversion process was creating a lookup table from the 'links'-file, allowing for easy retrieval of the author and tags for a certain article URL. Next, we iterated over the rows of the 'parsed'-file. In each step of that iteration, we:

1. Create a blank node;

2. State said blank node is of RDF-type prov:Entity;

3. State said blank node has as dc:identifier its URL;

4. State said blank node has as dc:title its title;

5. State said blank node has as dc:description its body text;

6. State said blank node has as dc:date is publication date, converted to the xsd:date format;

7. State said blank node has as dc:creator its author, retrieved from the 'links'-file and normalized to a human readable format;

8. State said blank node has as dc:subject any tags it has, retrieved from the 'links'-file and normalized to a human readable format.

The script in question can be found in the NewsReader BitBucket code repository[21].

---

[21]http://bitbucket.org/mvanerp/newsreader-deliverables

We used the Entity-class from the PROV Ontology[22] to type the article, in preparation for adding more provenance statements in a future revision of the conversion. We used the Dublin Core[23] ("dc") vocabulary for all of the metadata properties, as it is especially designed for metadata. An example of the resulting RDF can be found in Appendix B.1.

We did not use the 'articles'-file with just the URLs, because the same URLs are already present in the 'parsed'-file and thus serve no purpose in the conversion process.

4.2   CrunchBase

The entity data from CrunchBase was converted to RDF using a more complex Ruby script. This approach was chosen because the raw data was spread across a large amount of JSON-files retrieved from the CrunchBase API, and we are not aware of an off-the-shelf RDF converter for JSON.

4.2.1   Intuition

At its most basic, the conversion script works by iterating over every JSON file of a certain entity type (e.g. company) downloaded from CrunchBase. For each file/entity, we create an appropriate instance URI. Next, the script iterates over every key in the root of the current JSON file. Then, for each key, there are a few lines of code that convert the value for that specific key to a set of RDF triples as appropriate for that value. These statements are then collected in a named graph.

The named graph allows us to add statements describing the provenance of the triples in that graph. These provenance statements are also triples that, for example, state that the RDF triples in a certain graph were derived from a JSON-file downloaded from the CrunchBase API by a certain person working for a certain company.

For simple key-value combinations, where the value is a simply a string of text (such as an entities name), we simply add a triple such as "entity hasName Facebook". In addition to these simple key-value combinations, the CrunchBase JSON also contains keys that have (lists of) complex objects as their value. These are typically descriptions of events the entity in the current JSON-file was involved in. Examples include acquisitions, IPO's, and investments.

Similarly to the the root of the JSON-file, these event objects have their own key-value combinations that also need to be processed in different ways depending on the type of value they have: dates, participants, locations, and so forth.

Additionally, these events often have a link to the news article that served as the source of information for adding that event to CrunchBase. This means that the RDF triples generated for these events need to have additional provenance triples stating that the information they are based on, at some point in the past, originated from these articles and was added to CrunchBase. Therefore, these triples are stored in their own named

---

[22]http://www.w3.org/TR/prov-o/
[23]http://dublincore.org/

graph, allowing us to assign them additional and different provenance statements from the other triples.

A shortened example of the resulting RDF can be found in Appendix B.2. This RDF is based on the raw JSON for a Company, found in Appendix A.2.

In theory, it would also be possible to associate the CrunchBase entities to the same entity defined in an external data source (e.g. DBpedia[24] or Freebase[25]). For example, `http://www.crunchbase.com/person/mark-zuckerberg` could be linked to `http://dbpedia.org/page/Mark_Zuckerberg`. This details of this process are beyond the scope of this deliverable and will be described in Deliverable D6.2.1.

4.2.2 Vocabularies

We use a combination of existing vocabularies and a new CrunchBase-ontology for the predicates used in the conversion.

SEM The Simple Event Model [Van Hage *et al.*, 2011] is used for event-related triples, such as specifying that something is an event, that it has certain actors participating in that event, where the event took place, and when it took place.

PROV-O The Provenance Ontology[26] is used for statements about the provenance of triples in named graphs (as explained above), specifically which sources they were derived from and who was responsible for the conversion.

FOAF FOAF[27] is used for statements about addresses and contact information.

GEO Basic Geo[28] is used for specifying the lat/long belonging to a certain address.

OWL Time OWL Time[29] is used for representing instances and durations of time.

DC Dublin Core for certain metadata properties.

VCARD vCard[30] was used for defining detailed addresses.

4.2.3 Implementation Details

Each type of entity in CrunchBase has a specific set of key-value combinations. Some entities have a few common combinations (e.g. both companies, people, and financial organizations have investments), but most are different. Therefore, there is an individual script for each type of entity.

---

[24] `http://dbpedia.org/`
[25] `http://www.freebase.com/`
[26] `http://www.w3.org/TR/prov-o/`
[27] `http://xmlns.com/foaf/spec/`
[28] `http://www.w3.org/2003/01/geo/`
[29] `http://www.w3.org/TR/owl-time/`
[30] `http://www.w3.org/TR/vcard-rdf/`

Further details of the conversion implementation are best understood by looking at a conversion script itself. The script for converting a company to RDF can be found in the NewsReader BitBucket code repository. This script has detailed comments explaining its workings. Because the scripts for converting other types of entities are similar in structure and purpose, they are not included in this document, but can also be found in the NewsReader BitBucket code repository.

4.3   World Bank Development Indicators

The World Bank Development Indicators were already available as RDF[31], thus no conversion took place. The conversion was performed by Sarven Capadisli[32] as part of a different research project. Whether the RDF in its current form is suitable for use within the NewsReader project will be investigated in a future revision of this deliverable.

4.3.1   Intuition

Sarven gathered the Indicator-data in XML format from the World Bank's API and used XSLT[33] to transform them into RDF/XML.

The conversion process essentially creates indidual observations per country per year for each indicator. A shortened example of the resulting RDF, reserialized to JSON-LD, can be seen in Appendix B.3.

4.3.2   Vocabularies

The primary vocabularies used in the conversion process are the RDF Data Cube[34] vocabulary for modeling statistical observations and the SDMX[35] vocabulary for statistical codes. SKOS[36] and Dublin Core are also used.

4.3.3   Implementation Details

The details of the conversion process are well documented on the data sets companion website[37].

4.4   Yahoo! Finance

The conversion of Yahoo! Finance data will be included in a future revision of this deliverable. At the time of writing, it is undecided for which companies stock data needs to be

---

[31] http://worldbank.270a.info/

[32] http://csarven.ca/

[33] http://www.w3.org/TR/xslt

[34] http://www.w3.org/TR/vocab-data-cube/

[35] http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/vocab/sdmx.ttl

[36] http://www.w3.org/2009/08/skos-reference/skos.html

[37] http://worldbank.270a.info/about.html

gathered and converted. It is not possible to "download everything" from Yahoo! Finance [Rassom, 2012], it is necessary to have an a priori list of stock symbols to be gathered.

After such a list has been created, work on the conversion can proceed. Because of the tabular nature of the data provided by the Yahoo! Finance API, this data will be loaded into a relational database from which it can be easily converted to RDF using an off-the-shelf relational database to RDF converter.

5 Conclusion And Future Work

In this deliverable, we have described four data sets that need to be converted to RDF for use within the NewsReader project: TechCrunch, CrunchBase, the World Bank Indicators, and Yahoo! Finance. For each data set, we have described what kind of data it contains, how that data is structured, how the data was acquired, and what purpose it will serve within the project.

We gave an overview of different approaches for converting existing structured data to RDF: Writing a custom script, using an off-the-shelf tool, or taking advantage of an existing RDF version of the data. We have shown how we used these methods to convert our data sets to RDF and what the result looks like.

As stated in the introduction, the nature of the conversions described in this document is prototypical. This deliverable is (at least partially) intended to fuel discussion about what future conversions of the data sets should look like. At the time of writing, the following changes and improvements (grouped by data set) are planned for a future revision of this deliverable:

TechCrunch

1. Do not use blank nodes as article identifiers, but create proper URIs.

2. Investigate how the data set can be continuously updated with newly published articles.

CrunchBase

1. Create detailed specification of CrunchBase vocabulary.

2. Do not use blank nodes as date identifiers, but mint proper URIs.

3. Investigate how the data set can be continuously updated with newly added entities.

4. Do not use asserted typing to not confuse reasoning software. Rather, specify a proper ontology (see point 1) and use inferred typing.

5. Do not place provenance triples and provenance metadata triples (e.g. triples about persons involved in the conversion) in the same named graph.

6. Add a validity context to triples where appropriate (e.g. an e-mail address can only be valid in a certain time period).

World Bank Indicators

1. Investigate whether it is necessary to re-write the World bank Indicator RDF for use within NewsReader or whether it can be used as-is.

Yahoo! Finance

1. Investigate which stock symbols need to be downloaded for use within NewsReader and research which vocabularies are appropriate.

References

[AOL, 2013] AOL. Requesting Permission to Use Copyrighted Materials, 2013.

[CrunchBase, 2013a] CrunchBase. CrunchBase about page, 2013.

[CrunchBase, 2013b] CrunchBase. CrunchBase Licensing Policy, 2013.

[Rassom, 2012] Rassom. How to get a complete list of ticker symbols from Yahoo Finance?, 2012.

[Stelter, 2012] Brian Stelter. To Bolster Web Reach, CNBC Joins With Yahoo, 2012.

[TechCrunch, 2013] TechCrunch. TechCrunch about page, 2013.

[The World Bank, 2013a] The World Bank. Terms of Use for Datasets Listed in The World Bank Data Catalog, 2013.

[The World Bank, 2013b] The World Bank. World Development Indicators 2013, 2013.

[Van Hage *et al.*, 2011] Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the Simple Event Model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136, 2011.

A   Raw Data

This appendix contains examples of the raw data from each data set that was converted to RDF. TechCrunch, CrunchBase, the World Bank Indicators, and Yahoo! Finance each have their own section. For brevity, the data has been (partially) truncated in some areas. This is indicated by the [TRUNCATED]-indicator.

A.1 TechCrunch

This section contains a few example rows from each of the TechCrunch CSVs as created by SCW. "Articles" contains just the URLs of all scraped articles, "Parsed" contains the URL, full body text, publication time, and title, and "Links" contains the URL, author, tags, and CrunchBase-links.

A.1.1 Articles

url
http://eu.techcrunch.com/2010/01/05/done-deal-critical-path-acquires-shozu-ceo-chris-wade-stays-on-as-consultant/
http://eu.techcrunch.com/2010/01/05/social-network-badoo-is-banned-in-iran/
http://eu.techcrunch.com/2010/01/06/european-startups-scramble-to-emulate-the-groupon-explosion/

Table 2: Three example rows from the "'Articles"-CSV.

A.1.2 Parsed

url	body	time	title
[TRUNCATED]	Our earlier report about Critical Path buying mobile services startup ShoZu turns out to have been right on the money. [TRUNCATED]	Tuesday, January 5th, 2010	Done deal: Critical Path acquires Shozu, CEO Chris Wade stays on as consultant
[TRUNCATED]	Badoo, a social network popular in emerging markets like Russia and Brazil, has been banned in Iran. [TRUNCATED]	Tuesday, January 5th, 2010	Social network Badoo is banned in Iran
[TRUNCATED]	The Chicago-based Groupon has been valued at $280 million after closing their recent $30 million venture round with Accel Partners and previous investors. [TRUNCATED]	Wednesday, January 6th, 2010	European startups scramble to emulate the Groupon explosion

Table 3: Three example rows from the "'Parsed"-CSV. The [TRUNCATED] URLs are the same as in the "Articles"-CSV.

A.1.3   Links

link	article	type
http://api.crunchbase.com/v/1/company/twitter.js	[TRUNCATED]	cbase
http://eu.techcrunch.com/author/robin-wauters/	[TRUNCATED]	author
http://techcrunch.com/tag/critical-path/	[TRUNCATED]	tc_tag

Table 4: Three example rows from the "Links"-CSV. The [TRUNCATED] URLs are the same as in the "Articles"-CSV.

A.2   CrunchBase

This section contains examples of the JSON returned by the CrunchBase API. Each type of entity (Company, Person, Financial Organization, Service Provider, and Product) has its own subsection.

A.2.1   Company

```
 1  {
 2    "acquisition": null,
 3    "acquisitions": [
 4      {
 5        "acquired_day": 12,
 6        "acquired_month": 8,
 7        "acquired_year": 2013,
 8        "company": {
 9          "image": {
10            "attribution": null,
11            "available_sizes": [
12              [
13                [
14                  150,
15                  150
16                ],
17                "assets/images/resized/0021/3931/213931v2-max-150x150.png"
18              ],
19              [
20                [
21                  250,
22                  250
23                ],
24                "assets/images/resized/0021/3931/213931v2-max-250x250.png"
25              ],
26              [
27                [
28                  450,
29                  450
30                ],
31                "assets/images/resized/0021/3931/213931v2-max-450x450.png"
32              ]
33            ]
34          },
35          "name": "Jibbigo",
36          "permalink": "jibbigo"
37        },
38        "price_amount": null,
```

```
39      "price_currency_code": "USD",
40      "source_description": "Facebook Acquires \u201cMobile Technologies\u201d, Developer Of
            Speech Translation App Jibbigo",
41      "source_url": "http://techcrunch.com/2013/08/12/facebook-acquires-mobile-technologies-
            speech-recognition-and-jibbigo-app-developer/",
42      "term_code": null
43    }
44  ],
45  "alias_list": "",
46  "blog_feed_url": "http://blog.facebook.com/atom.php",
47  "blog_url": "http://blog.facebook.com",
48  "category_code": "web",
49  "competitions": [
50    {
51     "competitor": {
52      "image": {
53       "attribution": null,
54       "available_sizes": [
55        [
56         [
57          150,
58          148
59         ],
60         "assets/images/resized/0020/0311/200311v2-max-150x150.png"
61        ],
62        [
63         [
64          250,
65          248
66         ],
67         "assets/images/resized/0020/0311/200311v2-max-250x250.png"
68        ],
69        [
70         [
71          275,
72          273
73         ],
74         "assets/images/resized/0020/0311/200311v2-max-450x450.png"
75        ]
76       ]
77      },
78      "name": "Compass (by Hugleberry Corp.)",
79      "permalink": "hugleberry"
80     }
81    }
82  ],
83  "created_at": "Fri May 25 21:22:15 UTC 2007",
84  "crunchbase_url": "http://www.crunchbase.com/company/facebook",
85  "deadpooled_day": null,
86  "deadpooled_month": null,
87  "deadpooled_url": "",
88  "deadpooled_year": null,
89  "description": "Social network",
90  "email_address": "",
91  "external_links": [
92    {
93     "external_url": "http://www.sociableblog.com/2012/04/01/facebook-timeline-for-all-pages
            -goes-live/",
94     "title": "March 31, 2012: Facebook Timeline for All Pages Goes Live"
95    }
96  ],
97  "founded_day": 1,
98  "founded_month": 2,
99  "founded_year": 2004,
100 "funding_rounds": [
```

```
101   {
102     "funded_day": 21,
103     "funded_month": 1,
104     "funded_year": 2011,
105     "investments": [
106       {
107         "company": null,
108         "financial_org": {
109           "image": {
110             "attribution": null,
111             "available_sizes": [
112               [
113                 [
114                   74,
115                   74
116                 ],
117                 "assets/images/resized/0001/1376/11376v1-max-150x150.png"
118               ],
119               [
120                 [
121                   74,
122                   74
123                 ],
124                 "assets/images/resized/0001/1376/11376v1-max-250x250.png"
125               ],
126               [
127                 [
128                   74,
129                   74
130                 ],
131                 "assets/images/resized/0001/1376/11376v1-max-450x450.png"
132               ]
133             ]
134           },
135           "name": "Goldman Sachs",
136           "permalink": "goldman-sachs"
137         },
138         "person": null
139       },
140       {
141         "company": null,
142         "financial_org": {
143           "image": {
144             "attribution": null,
145             "available_sizes": [
146               [
147                 [
148                   134,
149                   46
150                 ],
151                 "assets/images/resized/0014/7467/147467v1-max-150x150.png"
152               ],
153               [
154                 [
155                   134,
156                   46
157                 ],
158                 "assets/images/resized/0014/7467/147467v1-max-250x250.png"
159               ],
160               [
161                 [
162                   134,
163                   46
164                 ],
165                 "assets/images/resized/0014/7467/147467v1-max-450x450.png"
```

```
166              ]
167             ]
168           },
169           "name": "Digital Sky Technologies",
170           "permalink": "digital -sky -technologies -fo"
171         },
172         "person": null
173       }
174     ],
175     "raised_amount": 1500000000.0,
176     "raised_currency_code": "USD",
177     "round_code": "unattributed",
178     "source_description": "Facebook Raises $1.5 Billion",
179     "source_url": "http://www.prnewswire.com/news-releases/facebook-raises-15-billion
                -114383494.html"
180   }
181 ],
182 "homepage_url": "http://facebook.com",
183 "image": {
184  "attribution": null,
185  "available_sizes": [
186   [
187    [
188     150,
189     61
190    ],
191    "assets/images/resized/0000/4561/4561v1-max-150x150.png"
192   ],
193   [
194    [
195     245,
196     100
197    ],
198    "assets/images/resized/0000/4561/4561v1-max-250x250.png"
199   ],
200   [
201    [
202     245,
203     100
204    ],
205    "assets/images/resized/0000/4561/4561v1-max-450x450.png"
206   ]
207  ]
208 },
209 "investments": [
210  {
211   "funding_round": {
212    "company": {
213     "image": {
214      "attribution": null,
215      "available_sizes": [
216       [
217        [
218         150,
219         38
220        ],
221        "assets/images/resized/0002/6771/26771v10-max-150x150.png"
222       ],
223       [
224        [
225         226,
226         58
227        ],
228        "assets/images/resized/0002/6771/26771v10-max-250x250.png"
229       ],
```

```
230          [
231            [
232              226,
233              58
234            ],
235            "assets/images/resized/0002/6771/26771v10-max-450x450.png"
236          ]
237        ]
238      },
239      "name": "Wildfire, a division of Google",
240      "permalink": "wildfire-interactive"
241    },
242    "funded_day": 1,
243    "funded_month": 12,
244    "funded_year": 2008,
245    "raised_amount": null,
246    "raised_currency_code": "USD",
247    "round_code": "grant",
248    "source_description": "",
249    "source_url": ""
250    }
251  }
252  ],
253  "ipo": {
254    "pub_day": 18,
255    "pub_month": 5,
256    "pub_year": 2012,
257    "stock_symbol": "NASDAQ:FB",
258    "valuation_amount": 2740000000000.0,
259    "valuation_currency_code": "USD"
260  },
261  "milestones": [
262    {
263      "description": "Facebook Has 1 Million Active Advertisers",
264      "source_description": "Facebook Has 1 Million Active Advertisers",
265      "source_text": "",
266      "source_url": "http://www.businessinsider.com/facebook-has-1-million-active-advertisers
            -2013-6",
267      "stoneable": {
268        "name": "Facebook",
269        "permalink": "facebook"
270      },
271      "stoneable_type": "Company",
272      "stoned_acquirer": null,
273      "stoned_day": 18,
274      "stoned_month": 6,
275      "stoned_value": null,
276      "stoned_value_type": null,
277      "stoned_year": 2013
278    }
279  ],
280  "name": "Facebook",
281  "number_of_employees": 1000,
282  "offices": [
283    {
284      "address1": "340 Madison Ave",
285      "address2": "",
286      "city": "New York",
287      "country_code": "USA",
288      "description": "New York",
289      "latitude": 40.7557162,
290      "longitude": -73.9792469,
291      "state_code": "NY",
292      "zip_code": "10017"
293    }
```

```
294    ],
295    "overview": "<p>Facebook is the world&#8217;s largest social network, with over <a href
           =\"http://techcrunch.com/2013/07/24/facebook-growth-2/\" title=\"1.15 billion monthly
            active users\">1.15 billion monthly active users</a>.</p>\n\n<p>Facebook was founded
            by <a href=\"http://www.crunchbase.com/person/mark-zuckerberg\" title=\"Mark
           Zuckerberg\">Mark Zuckerberg</a> in February 2004, initially as an exclusive network
           for Harvard students. It was a huge hit: in 2 weeks, half of the schools in the
           Boston area began demanding a Facebook network. [TRUNCATED]",
296    "partners": [],
297    "permalink": "facebook",
298    "phone_number": "",
299    "products": [
300     {
301      "image": {
302       "attribution": null,
303       "available_sizes": [
304        [
305         [
306          150,
307          112
308         ],
309         "assets/images/resized/0017/2010/172010v10-max-150x150.jpg"
310        ],
311        [
312         [
313          250,
314          187
315         ],
316         "assets/images/resized/0017/2010/172010v10-max-250x250.jpg"
317        ],
318        [
319         [
320          450,
321          337
322         ],
323         "assets/images/resized/0017/2010/172010v10-max-450x450.jpg"
324        ]
325       ]
326      },
327      "name": "Facebook Places",
328      "permalink": "facebook-places"
329     }
330    ],
331    "providerships": [
332     {
333      "is_past": false,
334      "provider": {
335       "image": {
336        "attribution": null,
337        "available_sizes": [
338         [
339          [
340           150,
341           95
342          ],
343          "assets/images/resized/0029/9121/299121v2-max-150x150.jpg"
344         ],
345         [
346          [
347           198,
348           126
349          ],
350          "assets/images/resized/0029/9121/299121v2-max-250x250.jpg"
351         ],
352         [
```

```
353            [
354              198,
355              126
356            ],
357            "assets/images/resized/0029/9121/299121v2-max-450x450.jpg"
358          ]
359        ]
360      },
361      "name": "Kasman Design",
362      "permalink": "kasman-design"
363    },
364    "title": "Graphic Design Projects"
365   }
366  ],
367  "relationships": [
368    {
369      "is_past": true,
370      "person": {
371      "first_name": "Jonathan",
372      "image": null,
373      "last_name": "Pines",
374      "permalink": "jonathan-pines"
375    },
376    "title": "Software Engineer"
377   }
378  ],
379  "screenshots": [
380    {
381      "attribution": null,
382      "available_sizes": [
383        [
384          [
385            150,
386            68
387          ],
388          "assets/images/resized/0004/2816/42816v1-max-150x150.png"
389        ],
390        [
391          [
392            250,
393            114
394          ],
395          "assets/images/resized/0004/2816/42816v1-max-250x250.png"
396        ],
397        [
398          [
399            450,
400            205
401          ],
402          "assets/images/resized/0004/2816/42816v1-max-450x450.png"
403        ]
404      ]
405    }
406  ],
407  "tag_list": "facebook, college, students, profiles, network, online-communities, social-
        networking",
408  "total_money_raised": "$2.43B",
409  "twitter_username": "facebook",
410  "updated_at": "Thu Jul 25 16:56:46 UTC 2013",
411  "video_embeds": []
412 }
```

A.2.2   Person

```
 1  {
 2    "affiliation_name": "Facebook",
 3    "alias_list": "",
 4    "birthplace": "",
 5    "blog_feed_url": "",
 6    "blog_url": "",
 7    "born_day": 14,
 8    "born_month": 5,
 9    "born_year": 1984,
10    "created_at": "Fri May 25 21:51:46 UTC 2007",
11    "crunchbase_url": "http://www.crunchbase.com/person/mark-zuckerberg",
12    "degrees": [
13      {
14        "degree_type": "",
15        "graduated_day": null,
16        "graduated_month": null,
17        "graduated_year": null,
18        "institution": "Harvard University",
19        "subject": "Computer Science"
20      }
21    ],
22    "external_links": [
23      {
24        "external_url": "http://www.time.com/time/specials/packages/article/0,28804,2036683
            _2037183,00.html",
25        "title": "Time 2010 Person Of The Year"
26      }
27    ],
28    "first_name": "Mark",
29    "homepage_url": "",
30    "image": {
31      "attribution": "",
32      "available_sizes": [
33        [
34          [
35            119,
36            150
37          ],
38          "assets/images/resized/0001/0688/10688v39-max-150x150.jpg"
39        ],
40        [
41          [
42            199,
43            250
44          ],
45          "assets/images/resized/0001/0688/10688v39-max-250x250.jpg"
46        ],
47        [
48          [
49            359,
50            450
51          ],
52          "assets/images/resized/0001/0688/10688v39-max-450x450.jpg"
53        ]
54      ]
55    },
56    "investments": [],
57    "last_name": "Zuckerberg",
58    "milestones": [
59      {
60        "description": "Mark Zuckerberg Joins Bill Gates And Steve Jobs With \"Simpsons\" Cameo
            ",
```

```
61      "source_description": "Mark Zuckerberg Joins Bill Gates And Steve Jobs With \u00e2\
            u20ac\u02dcSimpsons\u00e2\u20ac\u2122 Cameo",
62      "source_text": "",
63      "source_url": "http://techcrunch.com/2010/10/04/zuckerberg-gates-jobs/",
64      "stoneable": {
65       "first_name": "Mark",
66       "last_name": "Zuckerberg",
67       "permalink": "mark-zuckerberg"
68      },
69      "stoneable_type": "Person",
70      "stoned_acquirer": null,
71      "stoned_day": 4,
72      "stoned_month": 10,
73      "stoned_value": null,
74      "stoned_value_type": null,
75      "stoned_year": 2010
76     }
77    ],
78    "overview": "<p>Mark Zuckerberg is the founder and CEO of <a href=\"http://www.crunchbase
          .com/company/facebook\" title=\"Facebook\">Facebook</a>, which he started in his
          college dorm room in 2004 with roomates <a href=\"http://www.crunchbase.com/person/
          dustin-moskovitz\" title=\"Dustin Moskovitz\">Dustin Moskovitz</a> and <a href=\"http
          ://www.crunchbase.com/person/chris-hughes\" title=\"Chris Hughes\">Chris Hughes</a>.
          [TRUNCATED]",
79    "permalink": "mark-zuckerberg",
80    "relationships": [
81     {
82      "firm": {
83       "image": {
84        "attribution": null,
85        "available_sizes": [
86         [
87          [
88           150,
89           56
90          ],
91          "assets/images/resized/0000/4552/4552v2-max-150x150.jpg"
92         ],
93         [
94          [
95           250,
96           94
97          ],
98          "assets/images/resized/0000/4552/4552v2-max-250x250.jpg"
99         ],
100        [
101         [
102          450,
103          169
104         ],
105         "assets/images/resized/0000/4552/4552v2-max-450x450.jpg"
106        ]
107       ]
108      },
109      "name": "Facebook",
110      "permalink": "facebook",
111      "type_of_entity": "company"
112     },
113     "is_past": false,
114     "title": "Founder and CEO, Board Of Directors"
115    }
116   ],
117   "tag_list": "facebook, ceo, social-network",
118   "twitter_username": "",
119   "updated_at": "Sat May 11 20:25:37 UTC 2013",
```

```
120   "video_embeds": [],
121   "web_presences": [
122    {
123      "external_url": "http://twitter.com/finkd",
124      "title": "Twitter Profile"
125    }
126   ]
127  }
```

A.2.3   Financial Organization

```
 1  {
 2   "alias_list": null,
 3   "blog_feed_url": "",
 4   "blog_url": "",
 5   "created_at": "Fri Jun 15 09:45:21 UTC 2007",
 6   "crunchbase_url": "http://www.crunchbase.com/financial-organization/goldman-sachs",
 7   "description": null,
 8   "email_address": "",
 9   "external_links": [
10    {
11      "external_url": "http://en.wikipedia.org/wiki/Goldman_Sachs",
12      "title": "Wikipedia article"
13    }
14   ],
15   "founded_day": null,
16   "founded_month": null,
17   "founded_year": 1869,
18   "funds": [],
19   "homepage_url": "http://www.gs.com",
20   "image": {
21    "attribution": null,
22    "available_sizes": [
23     [
24      [
25       74,
26       74
27      ],
28      "assets/images/resized/0001/1376/11376v1-max-150x150.png"
29     ],
30     [
31      [
32       74,
33       74
34      ],
35      "assets/images/resized/0001/1376/11376v1-max-250x250.png"
36     ],
37     [
38      [
39       74,
40       74
41      ],
42      "assets/images/resized/0001/1376/11376v1-max-450x450.png"
43     ]
44    ]
45   },
46   "investments": [
47    {
48      "funding_round": {
49       "company": {
50        "image": {
51         "attribution": null,
```

```
 52        "available_sizes": [
 53         [
 54          [
 55           150,
 56           46
 57          ],
 58          "assets/images/resized/0026/3812/263812v1-max-150x150.jpg"
 59         ],
 60         [
 61          [
 62           220,
 63           68
 64          ],
 65          "assets/images/resized/0026/3812/263812v1-max-250x250.jpg"
 66         ],
 67         [
 68          [
 69           220,
 70           68
 71          ],
 72          "assets/images/resized/0026/3812/263812v1-max-450x450.jpg"
 73         ]
 74        ]
 75       },
 76       "name": "Celoxica",
 77       "permalink": "celoxica"
 78      },
 79      "funded_day": 8,
 80      "funded_month": 5,
 81      "funded_year": 2012,
 82      "raised_amount": 643112.0,
 83      "raised_currency_code": "GBP",
 84      "round_code": "unattributed",
 85      "source_description": "Oxford Capital Partners Source",
 86      "source_url": ""
 87     }
 88    }
 89   ],
 90   "milestones": [
 91    {
 92     "description": "Goldman's Asia Prop. Team Hires COO For Hedge Fund",
 93     "source_description": "Goldman's Asia Prop. Team Hires COO For Hedge Fund",
 94     "source_text": "",
 95     "source_url": "http://www.finalternatives.com/node/14159?utm_source=feedburner&
           utm_medium=feed&utm_campaign=Feed:+cleantechbrief/rss+(CleanTech+Brief)",
 96     "stoneable": {
 97      "name": "Goldman Sachs",
 98      "permalink": "goldman-sachs"
 99     },
100     "stoneable_type": "FinancialOrg",
101     "stoned_acquirer": null,
102     "stoned_day": 13,
103     "stoned_month": 10,
104     "stoned_value": null,
105     "stoned_value_type": null,
106     "stoned_year": 2010
107    }
108   ],
109   "name": "Goldman Sachs",
110   "number_of_employees": null,
111   "offices": [
112    {
113     "address1": "",
114     "address2": "",
115     "city": "New York",
```

```
116     "country_code": "USA",
117     "description": "",
118     "latitude": 42.14496,
119     "longitude": -75.400254,
120     "state_code": "NY",
121     "zip_code": ""
122   }
123   ],
124   "overview": "<p>Goldman Sachs is a one of the world&#8217;s largest investment banks. It
            traces its routes back to 1869 and is headquartered in Manhattan, New York City.
            Goldman Sachs provides wealth management, investment banking, and sales &amp; trading
             services.</p>\n\n<p>In regards to the technology world, Goldman Sachs continues to
            invest heavily in this market.</p>",
125   "permalink": "goldman-sachs",
126   "phone_number": "",
127   "providerships": [
128     {
129     "is_past": true,
130     "provider": {
131     "image": {
132     "attribution": null,
133     "available_sizes": [
134       [
135        [
136         150,
137         43
138        ],
139        "assets/images/resized/0012/7057/127057v2-max-150x150.jpg"
140       ],
141       [
142        [
143         250,
144         73
145        ],
146        "assets/images/resized/0012/7057/127057v2-max-250x250.jpg"
147       ],
148       [
149        [
150         450,
151         131
152        ],
153        "assets/images/resized/0012/7057/127057v2-max-450x450.jpg"
154       ]
155      ]
156     },
157     "name": "Wired Real Estate Group",
158     "permalink": "wired-real-estate-group"
159     },
160     "title": "Data Center Advisory"
161   }
162   ],
163   "relationships": [
164     {
165     "is_past": true,
166     "person": {
167     "first_name": "Nick",
168     "image": {
169     "attribution": null,
170     "available_sizes": [
171       [
172        [
173         150,
174         100
175        ],
176        "assets/images/resized/0001/0507/10507v1-max-150x150.jpg"
```

```
177        ],
178        [
179          [
180            240,
181            160
182          ],
183          "assets/images/resized/0001/0507/10507v1-max-250x250.jpg"
184        ],
185        [
186          [
187            240,
188            160
189          ],
190          "assets/images/resized/0001/0507/10507v1-max-450x450.jpg"
191        ]
192      ]
193    },
194    "last_name": "Grouf",
195    "permalink": "nick-grouf"
196   },
197   "title": "Associate (Summer)"
198  }
199 ],
200 "tag_list": null,
201 "twitter_username": null,
202 "updated_at": "Sat May 24 02:26:01 UTC 2008",
203 "video_embeds": []
204 }
```

A.2.4   Service Provider

```
1  {
2   "alias_list": null,
3   "created_at": "Wed Oct 15 02:05:10 UTC 2008",
4   "crunchbase_url": "http://www.crunchbase.com/service-provider/baker-mckenzie",
5   "email_address": "",
6   "external_links": [],
7   "homepage_url": "http://www.bakermckenzie.com",
8   "image": {
9    "attribution": null,
10   "available_sizes": [
11     [
12       [
13         150,
14         40
15       ],
16       "assets/images/resized/0018/6077/186077v3-max-150x150.jpg"
17     ],
18     [
19       [
20         250,
21         67
22       ],
23       "assets/images/resized/0018/6077/186077v3-max-250x250.jpg"
24     ],
25     [
26       [
27         450,
28         122
29       ],
30       "assets/images/resized/0018/6077/186077v3-max-450x450.jpg"
31     ]
```

```
32    ]
33  },
34  "name": "Baker & McKenzie",
35  "offices": [
36    {
37      "address1": "",
38      "address2": "",
39      "city": "San Francisco",
40      "country_code": "USA",
41      "description": "San Francisco",
42      "latitude": null,
43      "longitude": null,
44      "state_code": "CA",
45      "zip_code": ""
46    }
47  ],
48  "overview": "<p>Baker &amp; McKenzie is an international law firm, founded in Chicago in
        1949 by Russell Baker and John McKenzie. It is home to more than 3,800 lawyers spread
        over 69 offices in 42 different countries.</p>\n\n<p>The firm saw US$2.27 billion in
        revenue in fiscal year 2011.</p>\n\n<p>Baker &amp; McKenzie is ranked as the largest
        in the world by number of attorneys and revenue as of  2011. It is also the largest
        international law firm in Asia, with 14 offices, and in Latin America, with 16
        offices.</p>\n\n<p>The firm provides legal services in many different practice areas
        </p>",
49  "permalink": "baker-mckenzie",
50  "phone_number": "",
51  "providerships": [
52    {
53      "firm": {
54      "image": {
55        "attribution": null,
56        "available_sizes": [
57          [
58            [
59              116,
60              34
61            ],
62            "assets/images/resized/0023/8448/238448v2-max-150x150.png"
63          ],
64          [
65            [
66              116,
67              34
68            ],
69            "assets/images/resized/0023/8448/238448v2-max-250x250.png"
70          ],
71          [
72            [
73              116,
74              34
75            ],
76            "assets/images/resized/0023/8448/238448v2-max-450x450.png"
77          ]
78        ]
79      },
80      "name": "MILI",
81      "permalink": "mili",
82      "type_of_entity": "company"
83    },
84    "is_past": false,
85    "title": "legal"
86  }
87  ],
88  "tag_list": "",
89  "updated_at": "Sat Feb 16 06:33:37 UTC 2013"
```

```
90  }
```

A.2.5 Product

```
 1  {
 2   "alias_list": "",
 3   "blog_feed_url": "",
 4   "blog_url": "",
 5   "company": {
 6    "image": {
 7     "attribution": null,
 8     "available_sizes": [
 9      [
10       [
11        124,
12        150
13       ],
14       "assets/images/resized/0005/4061/54061v1-max-150x150.jpg"
15      ],
16      [
17       [
18        206,
19        250
20       ],
21       "assets/images/resized/0005/4061/54061v1-max-250x250.jpg"
22      ],
23      [
24       [
25        372,
26        450
27       ],
28       "assets/images/resized/0005/4061/54061v1-max-450x450.jpg"
29      ]
30     ]
31    },
32    "name": "Apple",
33    "permalink": "apple"
34   },
35   "created_at": "Sat Dec 22 08:45:28 UTC 2007",
36   "crunchbase_url": "http://www.crunchbase.com/product/iphone",
37   "deadpooled_day": null,
38   "deadpooled_month": null,
39   "deadpooled_url": "",
40   "deadpooled_year": null,
41   "external_links": [
42    {
43     "external_url": "http://www.sociableblog.com/2012/09/22/iphone-5-hits-the-stores/",
44     "title": "iPhone 5 Hits the Stores in 9 Countries Along with iOS 6"
45    }
46   ],
47   "homepage_url": "http://www.apple.com/iphone",
48   "image": {
49    "attribution": null,
50    "available_sizes": [
51     [
52      [
53       150,
54       117
55      ],
56      "assets/images/resized/0001/9797/19797v1-max-150x150.jpg"
57     ],
58     [
```

```
 59          [
 60            250,
 61            195
 62          ],
 63          "assets/images/resized/0001/9797/19797v1-max-250x250.jpg"
 64        ],
 65        [
 66          [
 67            450,
 68            351
 69          ],
 70          "assets/images/resized/0001/9797/19797v1-max-450x450.jpg"
 71        ]
 72      ]
 73    },
 74    "invite_share_url": "",
 75    "launched_day": 1,
 76    "launched_month": 1,
 77    "launched_year": 2007,
 78    "milestones": [
 79      {
 80        "description": "Apple introduces iPhone 5.",
 81        "source_description": "Apple Announcement Roundup: iPhone 5, New iPod Touch, iPod Nano,
               EarPods",
 82        "source_text": "",
 83        "source_url": "http://techcrunch.com/2012/09/12/apple-announcement-roundup-iphone-5-new
               -ipod-touch-ipod-nano-earpods/",
 84        "stoneable": {
 85          "name": "iPhone",
 86          "permalink": "iphone"
 87        },
 88        "stoneable_type": "Product",
 89        "stoned_acquirer": null,
 90        "stoned_day": 12,
 91        "stoned_month": 9,
 92        "stoned_value": null,
 93        "stoned_value_type": null,
 94        "stoned_year": 2012
 95      }
 96    ],
 97    "name": "iPhone",
 98    "overview": "<p><a href=\"/company/apple\" title=\"Apple\" rel=\"nofollow\">Apple</a
               >&#8217;s iPhone was introduced at MacWorld in January 2007 and officially went on
               sale June 29, 2007, selling 146,000 units within the first weekend of launch. [
               TRUNCATED]",
 99    "permalink": "iphone",
100    "stage_code": "live",
101    "tag_list": "apple, cell-phones, smartphones, iphone",
102    "twitter_username": "",
103    "updated_at": "Fri Nov 23 19:25:47 UTC 2012",
104    "video_embeds": [
105      {
106        "description": "<p>Introduction to the iPhone</p>",
107        "embed_code": "<iframe width=\"370\" height=\"308\" src=\"http://www.youtube.com/embed/
               YgW7or1TuFk\" frameborder=\"0\" allowfullscreen></iframe>"
108      }
109    ]
110  }
```

A.3 World Bank Indicators

This section contains a few example rows of The World Bank Indicator data in CSV-format.

Country	2009	2010	2011
Australia	1.3	1.5	1.9
Austria	12.1	10.8	10.1
Azerbaijan	1.7	2.6	1.8

Table 5: A few example rows of The World Bank Indicator "Alternative and nuclear energy (% of total energy use)" in CSV-format.

A.4   Yahoo! Finance

This section contains a few example rows of Yahoo! historical stock price data in CSV-format.

Date	Open	High	Low	Close	Volume	Adj. Close
Nov 26, 2013	524.12	536.14	524.00	533.40	14283400	533.40
Nov 25, 2013	521.02	525.87	521.00	523.74	8189700	523.74
Nov 22, 2013	519.52	522.16	518.53	519.80	7990200	519.90

Table 6: A few example rows of historical stock prices for the AAPL stock symbol.

Todo.

B   Resulting RDF

This appendix contains examples of the resulting RDF (serialized as JSON-LD[38]) from each data set that was converted to RDF. TechCrunch, CrunchBase, the World Bank Indicators, and Yahoo! Finance each have their own section. For brevity, the data has been (partially) truncated in some areas. This is indicated by the [TRUNCATED]-indicator.

B.1   TechCrunch

```
1  {
2    "@context": {
3      "prov": "http://www.w3.org/ns/prov#",
4      "xsd": "http://www.w3.org/2001/XMLSchema#",
5      "dc": "http://purl.org/dc/elements/1.1/"
6    },
7    "@id": "_:g70296052296540",
8    "@type": "prov:Entity",
9    "dc:creator": "Scott Merrill",
10   "dc:date": {
11     "@value": "2013-02-07",
12     "@type": "xsd:date"
13   },
```

---

[38]http://json-ld.org/

```
14    "dc:description": "<div class=\"body-copy\">\n\n\t\t\t\t\t\t\t\t<p>Microsoft Office has
         long been the dominant office suite. Through the years there have been many
         contenders rise and fall: WordPerfect, Corel, StarOffice, and too many more to count
         . Sun Microsystem&#8217;s StarOffice eventually mutated into OpenOffice, which for a
          long time was the best alternative to Microsoft&#8217;s dominance. But when Oracle
          bought Sun, legions of developers abandoned OpenOffice, and instead threw in with a
          forked version called <a target=\"_blank\" href=\"http://www.techcrunch.com/tag/\">
         LibreOffice</a>. [TRUNCATED]",
15    "dc:identifier": {
16      "@id": "http://techcrunch.com/2013/02/07/libreoffice-4-0-released-just-in-time-for-
         office-365-refugees/"
17    },
18    "dc:subject": [
19      "document-foundation",
20      "libreoffice",
21      "open-source",
22      "openoffice"
23    ],
24    "dc:title": "LibreOffice 4.0 Released Just In Time For Office 365 Refugees "
25  }
```

B.2   CrunchBase

```
1  {
2    "@context": {
3      "cbi_company": "http://www.newsreader-project.eu/rdf/instance/company/",
4      "cbi_acquisition": "http://www.newsreader-project.eu/rdf/instance/acquisition/",
5      "cbi_relationship": "http://www.newsreader-project.eu/rdf/instance/relationship/",
6      "cbi_person": "http://www.newsreader-project.eu/rdf/instance/person/",
7      "cbi_funding_round": "http://www.newsreader-project.eu/rdf/instance/funding_round/",
8      "cbi_investment": "http://www.newsreader-project.eu/rdf/instance/investment/",
9      "cbi_financial_organization": "http://www.newsreader-project.eu/rdf/instance/
         financial_organization/",
10     "cbi_ipo": "http://www.newsreader-project.eu/rdf/instance/ipo/",
11     "cbi_product": "http://www.newsreader-project.eu/rdf/instance/product/",
12     "cbi_providership": "http://www.newsreader-project.eu/rdf/instance/providership/",
13     "cbi_service_provider": "http://www.newsreader-project.eu/rdf/instance/
         service_provider/",
14     "cbi_milestone": "http://www.newsreader-project.eu/rdf/instance/milestone/",
15     "cbi_founding": "http://www.newsreader-project.eu/rdf/instance/founding/",
16     "cbg": "http://www.newsreader-project.eu/rdf/graph/",
17     "cbo": "http://www.newsreader-project.eu/rdf/ontology/",
18     "cbp": "http://www.newsreader-project.eu/rdf/provenance/",
19     "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
20     "sem": "http://semanticweb.cs.vu.nl/2009/11/sem/",
21     "prov": "http://www.w3.org/ns/prov#",
22     "foaf": "http://xmlns.com/foaf/0.1/",
23     "geo": "http://www.w3.org/2003/01/geo/wgs84_pos",
24     "xsd": "http://www.w3.org/2001/XMLSchema#",
25     "time": "http://www.w3.org/2006/time#",
26     "dc": "http://purl.org/dc/terms/",
27     "vcard": "http://www.w3.org/2006/vcard/ns#"
28   },
29   "@graph": [
30     {
31       "@id": "cbg:13e6e08d-cfeb-4fd4-9eaa-a252e84feddf",
32       "@graph": [
33         {
34           "@id": "_:g70097437826160",
35           "@type": "time:Instant",
36           "time:inXSDDate": {
37             "@value": "2007-7-1",
```

```
38                "@type": "xsd:date"
39              }
40            },
41            {
42              "@id": "cbi_acquisition:parakey-acquired-by-facebook-on-2007-7-1",
43              "@type": "sem:Event",
44              "cbo:hasAcquiree": {
45                "@id": "cbi_company:parakey"
46              },
47              "cbo:hasAcquirer": {
48                "@id": "cbi_company:facebook"
49              },
50              "cbo:hasCurrency": "USD",
51              "cbo:hasPrice": "?",
52              "rdfs:label": "Parakey acquired by Facebook",
53              "sem:eventType": {
54                "@id": "cbo:Acquisition"
55              },
56              "sem:hasTime": {
57                "@id": "_:g70097437826160"
58              }
59            }
60          ]
61        },
62        {
63          "@id": "cbg:4bf26aac-abc5-4c43-9365-4458c893f9b6",
64          "@graph": [
65            {
66              "@id": "_:g70097435758200",
67              "@type": "time:Instant",
68              "time:inXSDDate": {
69                "@value": "25-6-2008",
70                "@type": "xsd:date"
71              }
72            },
73            {
74              "@id": "cbi_milestone:facebook-milestone-on-25-6-2008",
75              "@type": "sem:Event",
76              "cbo:hasCompany": {
77                "@id": "cbi_company:facebook"
78              },
79              "rdfs:label": "Facebook adds comments to the Mini-Feed. It's like FriendFeed is
                    looking in the mirror",
80              "sem:eventType": {
81                "@id": "cbo:Milestone"
82              },
83              "sem:hasTime": {
84                "@id": "_:g70097435758200"
85              }
86            }
87          ]
88        },
89        {
90          "@id": "cbg:6c753694-df03-4864-8ac7-17b59d2ff00a",
91          "@graph": [
92            {
93              "@id": "_:g70097438622260",
94              "@type": "time:Instant",
95              "time:inXSDDate": {
96                "@value": "1-9-2004",
97                "@type": "xsd:date"
98              }
99            },
100           {
101             "@id": "cbi_funding_round:angel-funding-round-for-facebook-on-1-9-2004",
```

```
102              "@type": "sem:Event",
103              "cbo:hasAmount": {
104                "@value": "5.0E5",
105                "@type": "xsd:double"
106              },
107              "cbo:hasCurrency": "USD",
108              "cbo:hasInvestee": {
109                "@id": "cbi_company:facebook"
110              },
111              "cbo:hasInvestor": [
112                {
113                  "@id": "cbi_person:peter-thiel"
114                },
115                {
116                  "@id": "cbi_person:reid-hoffman"
117                }
118              ],
119              "cbo:hasRoundCode": "angel",
120              "rdfs:label": "angel funding round for Facebook",
121              "sem:eventType": {
122                "@id": "cbo:AngelFundingRound"
123              },
124              "sem:hasTime": {
125                "@id": "_:g70097438622260"
126              }
127            }
128          ]
129        },
130        {
131          "@id": "cbg:91dd194c-b087-4673-8f63-1f72b0e8a125",
132          "@graph": [
133            {
134              "@id": "_:g70097437118880",
135              "@type": "time:Instant",
136              "time:inXSDDate": {
137                "@value": "2009-2-20",
138                "@type": "xsd:date"
139              }
140            },
141            {
142              "@id": "cbi_investment:facebook-invested-in-luckycal-on-2009-2-20",
143              "@type": "sem:Event",
144              "cbo:hasAmount": {
145                "@value": "3.5E5",
146                "@type": "xsd:double"
147              },
148              "cbo:hasCurrency": "USD",
149              "cbo:hasInvestee": {
150                "@id": "cbi_company:luckycal"
151              },
152              "cbo:hasInvestor": {
153                "@id": "cbi_company:facebook"
154              },
155              "cbo:hasRoundCode": "seed",
156              "rdfs:label": "fbFund",
157              "sem:eventType": {
158                "@id": "cbo:SeedInvestment"
159              },
160              "sem:hasTime": {
161                "@id": "_:g70097437118880"
162              }
163            }
164          ]
165        },
166        {
```

```
167          "@id": "cbg:b5511040-9891-4333-a783-db66eb320dcf",
168          "@graph": [
169            {
170              "@id": "_:g70097435860720",
171              "@type": "geo:#Point",
172              "geo:#lat": {
173                "@value": "3.741605E1",
174                "@type": "xsd:double"
175              },
176              "geo:#long": {
177                "@value": "-1.22151801E2",
178                "@type": "xsd:double"
179              }
180            },
181            {
182              "@id": "_:g70097436277640",
183              "@type": "vcard:Work",
184              "foaf:based_near": {
185                "@id": "_:g70097435860720"
186              },
187              "rdfs:label": "Headquarters",
188              "vcard:country-name": "USA",
189              "vcard:locality": "Menlo Park",
190              "vcard:postal-code": "?",
191              "vcard:street-address": "?"
192            },
193            {
194              "@id": "_:g70097438530400",
195              "@type": "time:Instant",
196              "time:inXSDDate": {
197                "@value": "1-2-2004",
198                "@type": "xsd:date"
199              }
200            },
201            {
202              "@id": "_:g70097438774600",
203              "@type": "time:Instant",
204              "time:inXSDDate": {
205                "@value": "18-5-2012",
206                "@type": "xsd:date"
207              }
208            },
209            {
210              "@id": "cbi_company:facebook",
211              "@type": "sem:Actor",
212              "cbo:hasBlogFeedUrl": {
213                "@id": "http://blog.facebook.com/atom.php"
214              },
215              "cbo:hasBlogUrl": {
216                "@id": "http://blog.facebook.com"
217              },
218              "cbo:hasCategory": {
219                "@id": "cbo:web"
220              },
221              "cbo:hasCompetitor": [
222                {
223                  "@id": "cbi_company:myspace"
224                }
225              ],
226              "cbo:hasLink": [
227                {
228                  "@id": "http://latimesblogs.latimes.com/technology/2008/09/facebook-hire-1.
                          html"
229                }
230              ],
```

```
231             "cbo:hasProduct": {
232               "@id": "cbi_product:facebook-platform"
233             },
234             "cbo:numberOfEmployees": {
235               "@value": "1000",
236               "@type": "xsd:integer"
237             },
238             "cbo:totalMoneyRaised": "$2.43B",
239             "cbp:hasAlias": "?",
240             "dc:description": "Social network",
241             "foaf:homepage": {
242               "@id": "http://facebook.com"
243             },
244             "foaf:mbox": "?",
245             "foaf:phone": "?",
246             "foaf:twitterID": {
247               "@id": "http://twitter.com/facebook"
248             },
249             "rdfs:comment": "<p>Facebook is the world&#8217;s largest social network, with
                      over <a href=\"http://techcrunch.com/2013/07/24/facebook-growth-2/\" title
                      =\"1.15 billion monthly active users\">1.15 billion monthly active users</a
                      >.</p>\n\n<p>Facebook was founded by <a href=\"http://www.crunchbase.com/
                      person/mark-zuckerberg\" title=\"Mark Zuckerberg\">Mark Zuckerberg</a> in
                      February 2004, initially as an exclusive network for Harvard students. It
                      was a huge hit: in 2 weeks, half of the schools in the Boston area began
                      demanding a Facebook network. [TRUNCATED]",
250             "rdfs:label": "Facebook",
251             "sem:actorType": {
252               "@id": "cbo:WebCompany"
253             },
254             "vcard:hasAddress": {
255               "@id": "_:g70097436277640"
256             }
257           },
258           {
259             "@id": "cbi_founding:facebook_founded_on_1-2-2004",
260             "@type": "sem:Event",
261             "cbo:hasTime": {
262               "@id": "_:g70097438530400"
263             },
264             "rdfs:label": "Facebook founding",
265             "sem:eventType": {
266               "@id": "cbo:Founding"
267             },
268             "sem:hasCompany": {
269               "@id": "cbi_company:facebook"
270             }
271           },
272           {
273             "@id": "cbi_ipo:facebook-ipo-on-18-5-2012",
274             "@type": "sem:Event",
275             "cbo:hasAmount": {
276               "@value": "2.74E+12",
277               "@type": "xsd:double"
278             },
279             "cbo:hasCurrency": "USD",
280             "cbo:hasStockSymbol": "NASDAQ:FB",
281             "cbo:hasTime": {
282               "@id": "_:g70097438774600"
283             },
284             "rdfs:label": "Facebook IPO",
285             "sem:eventType": {
286               "@id": "cbo:IPO"
287             },
288             "sem:hasCompany": {
```

```
289                   "@id": "cbi_company:facebook"
290                 }
291               },
292               {
293                 "@id": "cbi_providership:outcast-communications-provider-of-facebook",
294                 "@type": "sem:Event",
295                 "cbo:hasProvidee": {
296                   "@id": "cbi_company:facebook"
297                 },
298                 "cbo:hasProvider": {
299                   "@id": "cbi_service_provider:outcast-communications"
300                 },
301                 "cbo:isPast": {
302                   "@value": "false",
303                   "@type": "xsd:boolean"
304                 },
305                 "rdfs:label": "The OutCast Agency provider of Facebook",
306                 "sem:eventType": {
307                   "@id": "cbo:Providership"
308                 }
309               },
310               {
311                 "@id": "cbi_relationship:mark-zuckerberg-related-to-facebook",
312                 "@type": "sem:Event",
313                 "cbo:hasCompany": {
314                   "@id": "cbi_company:facebook"
315                 },
316                 "cbo:hasPerson": {
317                   "@id": "cbi_person:mark-zuckerberg"
318                 },
319                 "cbo:isPast": {
320                   "@value": "false",
321                   "@type": "xsd:boolean"
322                 },
323                 "rdfs:label": "Mark Zuckerberg related to Facebook",
324                 "sem:eventType": {
325                   "@id": "cbo:Relationship"
326                 }
327               }
328             ]
329           },
330           {
331             "@id": "cbg:provenance",
332             "@graph": [
333               {
334                 "@id": "_:g70097437917520",
335                 "@type": "prov:Entity",
336                 "prov:atLocation": {
337                   "@id": "http://www.crunchbase.com/company/facebook"
338                 },
339                 "prov:generatedAtTime": {
340                   "@value": "2013-07-25",
341                   "@type": "xsd:date"
342                 }
343               },
344               {
345                 "@id": "cbg:13e6e08d-cfeb-4fd4-9eaa-a252e84feddf",
346                 "@type": "prov:Entity",
347                 "prov:wasAttributedTo": {
348                   "@id": "cbp:ThomasPloeger"
349                 },
350                 "prov:wasDerivedFrom": {
351                   "@id": "_:g70097437917520"
352                 },
353                 "prov:wasGeneratedBy": {
```

```
354              "@id": "cbp:CrunchBaseConversion"
355            },
356            "rdfs:seeAlso": {
357              "@id": "http://www.techcrunch.com/2007/07/19/breaking-facebook-has-acquired-
                    parakey/"
358            }
359          },
360          {
361            "@id": "cbg:4bf26aac-abc5-4c43-9365-4458c893f9b6",
362            "@type": "prov:Entity",
363            "prov:wasAttributedTo": {
364              "@id": "cbp:ThomasPloeger"
365            },
366            "prov:wasDerivedFrom": {
367              "@id": "_:g70097437917520"
368            },
369            "prov:wasGeneratedBy": {
370              "@id": "cbp:CrunchBaseConversion"
371            },
372            "rdfs:seeAlso": {
373              "@id": "http://venturebeat.com/2008/06/25/facebook-adds-comment-to-the-mini-
                    feed-its-like-friendfeed-is-looking-in-the-mirror/"
374            }
375          },
376          {
377            "@id": "cbg:6c753694-df03-4864-8ac7-17b59d2ff00a",
378            "@type": "prov:Entity",
379            "prov:wasAttributedTo": {
380              "@id": "cbp:ThomasPloeger"
381            },
382            "prov:wasDerivedFrom": {
383              "@id": "_:g70097437917520"
384            },
385            "prov:wasGeneratedBy": {
386              "@id": "cbp:CrunchBaseConversion"
387            }
388          },
389          {
390            "@id": "cbg:91dd194c-b087-4673-8f63-1f72b0e8a125",
391            "@type": "prov:Entity",
392            "prov:wasAttributedTo": {
393              "@id": "cbp:ThomasPloeger"
394            },
395            "prov:wasDerivedFrom": {
396              "@id": "_:g70097437917520"
397            },
398            "prov:wasGeneratedBy": {
399              "@id": "cbp:CrunchBaseConversion"
400            },
401            "rdfs:seeAlso": {
402              "@id": "http://www.marlenevergaraborquez.com/press/releases.php?p=48242"
403            }
404          },
405          {
406            "@id": "cbg:b5511040-9891-4333-a783-db66eb320dcf",
407            "@type": "prov:Entity",
408            "prov:wasAttributedTo": {
409              "@id": "cbp:ThomasPloeger"
410            },
411            "prov:wasDerivedFrom": {
412              "@id": "_:g70097437917520"
413            },
414            "prov:wasGeneratedBy": {
415              "@id": "cbp:CrunchBaseConversion"
416            }
```

```
417              },
418              {
419                "@id": "cbp:CrunchBaseConversion",
420                "@type": "prov:Activity",
421                "prov:atTime": {
422                  "@value": "2013-11-29",
423                  "@type": "xsd:date"
424                },
425                "prov:wasAssociatedWith": {
426                  "@id": "cbp:ThomasPloeger"
427                }
428              },
429              {
430                "@id": "cbp:SynerScope",
431                "@type": [
432                  "foaf:Organization",
433                  "prov:Agent"
434                ],
435                "foaf:homepage": {
436                  "@id": "http://www.synerscope.com"
437                },
438                "foaf:name": "SynerScope B.V."
439              },
440              {
441                "@id": "cbp:ThomasPloeger",
442                "@type": [
443                  "prov:Agent",
444                  "foaf:Person"
445                ],
446                "foaf:mbox": "thomas.ploeger@synerscope.com",
447                "foaf:name": "Thomas Ploeger",
448                "prov:actedOnBehalfOf": {
449                  "@id": "cbp:SynerScope"
450                }
451              }
452            ]
453          }
454        ]
455 }
```

B.3   World Bank Indicators

```
1  {
2      "@graph": [
3          {
4              "@id": "http://worldbank.270a.info/dataset/world-bank-indicators/AG.LND.TRAC.
                  ZS/1A/1961",
5              "http://purl.org/linked-data/cube#dataSet": [
6                  {
7                      "@id": "http://worldbank.270a.info/dataset/AG.LND.TRAC.ZS"
8                  }
9              ],
10             "http://purl.org/linked-data/sdmx/2009/dimension#refArea": [
11                 {
12                     "@id": "http://worldbank.270a.info/classification/country/1A"
13                 }
14             ],
15             "http://purl.org/linked-data/sdmx/2009/dimension#refPeriod": [
16                 {
17                     "@id": "http://reference.data.gov.uk/id/year/1961"
18                 }
19             ],
```

```
20              "http://purl.org/linked-data/sdmx/2009/measure#obsValue": [
21                  {
22                      "@value": "15.9697470225159",
23                      "@type": "http://www.w3.org/2001/XMLSchema#decimal"
24                  }
25              ],
26              "http://worldbank.270a.info/property/decimal": [
27                  {
28                      "@value": "1",
29                      "@type": "http://www.w3.org/2001/XMLSchema#integer"
30                  }
31              ],
32              "http://worldbank.270a.info/property/indicator": [
33                  {
34                      "@id": "http://worldbank.270a.info/classification/indicator/AG.LND.
                            TRAC.ZS"
35                  }
36              ],
37              "@type": [
38                  "http://purl.org/linked-data/cube#Observation"
39              ]
40          },
41          [TRUNCATED]
42      ]
43 }
```

B.4   World Bank Indicators

To be included in a future revision of this deliverable. See Section 4.4 for more details.