The CorpWatch API:

an ID system for transparent company hierarchies

Skye Bender-deMoll,Greg Michalec,Tonya Hennessey

(support@api.corpwatch.org)

Sunlight Foundation

CORPWATCH Holding Corporations Accountable

Why build a free database of company names and subsidiary relations?

- The ID numbers of companies in the API can serve as a "backbone" for linking across datasets
- Knowing subsidiaries assists in aggregating data and associating with parent company
- Possible to examine changes in corporate structures over time
- Low-budget alternative to expensive proprietary systems (D&B's D-U-N-S Number or CapitalIQ's CompuStat)

What's been done so far

- Powerful multi-year API
- CrocTail interactive browser for subsidiary hierarchies and locations
- Animations of changes in corporate structure
- Research using database dumps
- "Widget" in Crocodyl.org profiles

http://croctail.corpwatch.org

Related projects are saying...

LittleSis.org blog post:

> *"For starters, LittleSis can take advantage of the US-based subsidiary data to help match government contracts from FedSpending.org with parent companies we track. ...Who'd have known that this Department of Interior contract to Landmark Graphics Corporation was actually another Halliburton deal? Thanks to the CorpWatch API, we now have a way of detecting that automatically."*

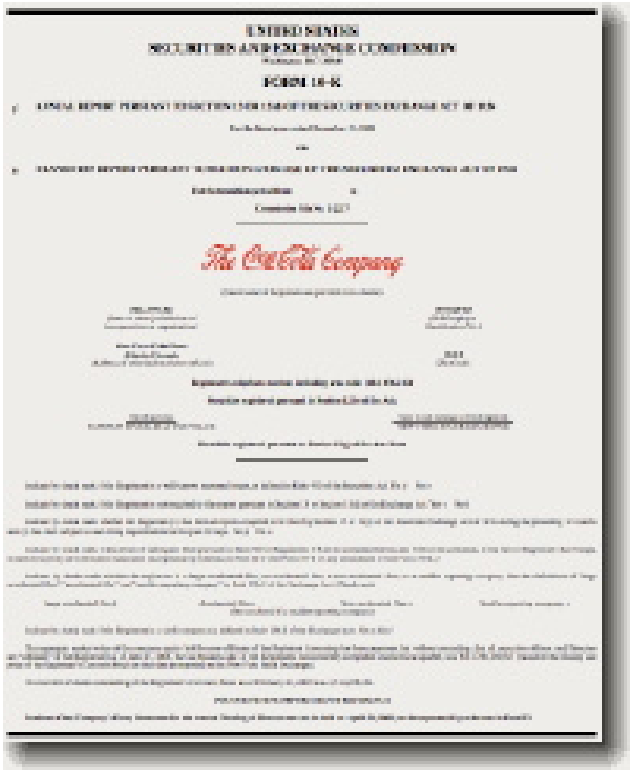
Researchers are saying...

Data request email from a Business Administration PhD Candidate, Florida International University:

> *"...I'm looking at the pattern of internationalization decisions of US-firms ad trying to derive a measure of relatedness between US (home) and each of the other countries (host). In a second piece of research with other colleagues, we are looking at Corporate Social Responsibility and Corporate Social Rhetoric for US-firms and investigate their drivers."*

How we get the data

`http://www.sec.gov/edgar.shtml`



- Every corporation listed on a U.S. stock exchange is required to file annual reports with the SEC
- SEC provides the data online via its EDGAR system
- Information on "parent" companies parsed from EDGAR header HTML
- Exhibit 21 of form 10-K includes a listing of subsidiary companies with at least 10% holding
- Formatting is not well standardized

Header Example: Goldman Sachs Group Inc.

Home | Latest Filings | Previous Page

U.S. Securities and Exchange Commission

Search Results

Search the Next-Generation EDGAR System

SEC Home » Search the Next-Generation EDGAR System » Company Search » *Current Page*

GOLDMAN SACHS GROUP INC CIK#: 0000886982 (see all company filings)

SIC: 6211 - SECURITY BROKERS, DEALERS & FLOTATION COMPANIES
State location: NY | State of Inc.: **DE** | Fiscal Year End: 1231
formerly: GOLDMAN SACHS GROUP INC/ (filings through 2008-03-04)
(Assistant Director Office No 8)
Get insider transactions for this **issuer**.
Get insider transactions for this **reporting owner**.

Business Address
200 WEST STREET
NEW YORK NY 10282
2129021000

Mailing Address
200 WEST STREET
NEW YORK NY 10282

- SEC's Central Index Key (CIK)
- Multiple addresses
- Possible former names
- Possible IRS Employer Identification Number (EIN) number
- Standard Industrial Classification (SIC) category

Exhibit 21 Example: Lehman Brothers

EX-21.01 14 a08-3530_1ex21d01.htm EX-21.01

EXHIBIT 21.01

LIST OF THE REGISTRANT'S SUBSIDIARIES

as of November 30, 2007

Pursuant to Item 601(b)(21)(ii) of Regulation S-K, subsidiaries of the Registrant have been omitted which, considered in the aggregate as a single subsidiary, would not have constituted a significant subsidiary (as defined in Rule 1-02(w) of Regulation S-X) as of November 30, 2007. Indentations indicate parent-subsidiary relationships.

Tier	Name	Jurisdiction of Organization
0	Lehman Brothers Holdings Inc.	Delaware
1	Appalachian Asset Management Corp.	Delaware
2	Lehman Risk Services (Bermuda) Ltd.	Bermuda
1	ARS Holdings I LLC	Delaware
1	Banque Lehman Brothers S.A.	France
1	LB 745 LLC	Delaware
1	LB 745 Leaseco I LLC	Delaware
1	LBAC Holdings I Inc.	Delaware
2	Lehman Brothers Asia Capital Company	Hong Kong
1	LBCCA Holdings I LLC	Delaware
2	Falcon Holdings I LLC	Delaware
3	Falcon Holdings II Inc.	Delaware
4	CIMT Limited	Cayman Islands
5	TMIC Limited	Cayman Islands
6	MICT Limited	Cayman Islands
7	Falcon Investor I-X Inc.	Cayman Islands
8	Global Thai Property Fund	Thailand
2	Lehman Brothers Asia Capital Company	Hong Kong

Exhibit 21 Example: Nike

EX-21 7 dex21.htm SUBSIDIARIES OF TH REGISTRANT

EXHIBIT 21

SUBSIDIARIES OF THE REGISTRANT

NIKE, Inc. has 130 wholly-owned subsidiaries, 20 of which operate in the United States, and 110 of which operate in foreign countries. All of the subsidiaries, except for NIKE IHM, Inc., Triax Insurance, Inc., and NIKE (Suzhou) Sports Company, Ltd. carry on the same line of business, namely the design, marketing distribution and sale of athletic and leisure footwear, apparel, accessories, and equipment. NIKE IHM, Inc., a Missouri corporation, and NIKE (Suzhou) Sports Company, Ltd., a Chinese corporation, manufacture plastics and Air-Sole shoe cushioning components. Triax Insurance, Inc., a Hawaii corporation, is a captive insurance company that insures the Company for certain risks.

- Very hard to identify company names, locations are missing

Design goals for database system

Time-sensitive coverage: Company names, locations, and subsidiaries change over time. We want to be able to reflect that in our data.

Semi-automated update for new data: We want something maintainable, not just a single snapshot of the data that quickly goes out of date.

Reproducible parsing and data processing: Anybody who runs our code should be able to produce the same database.

Stable set of company entities/ids: Once we assign ids to a company, it should stay the same, even when we tweak the parsing code.

Geocoding to "subdivision" (state) level: In some cases we have more detailed addresses, but the jurisdiction is more reliable and easier to match.

Entity matching

- We standardize names before matching
- Locations/jurisdictions must also match

Abbreviation: *Incorporated* to *Inc* : Reduction table

Punctuation differences: *Inc* to *Inc.* : Punctuation stripping and translation

Character set differences *Nestlé* to *Nestle* : Character set reduction and translation rules

Informal names *Coke* to *Coca Cola* : Alias tables

Name Change *Network Associates* to *McAfee Inc* : Tables of historical names and ids

Parsing garbage/errors *-Mirant Capital, Inc.; 6.54%*to *Mirant Capital Inc* : Repeated evaluation and correction of each problem case

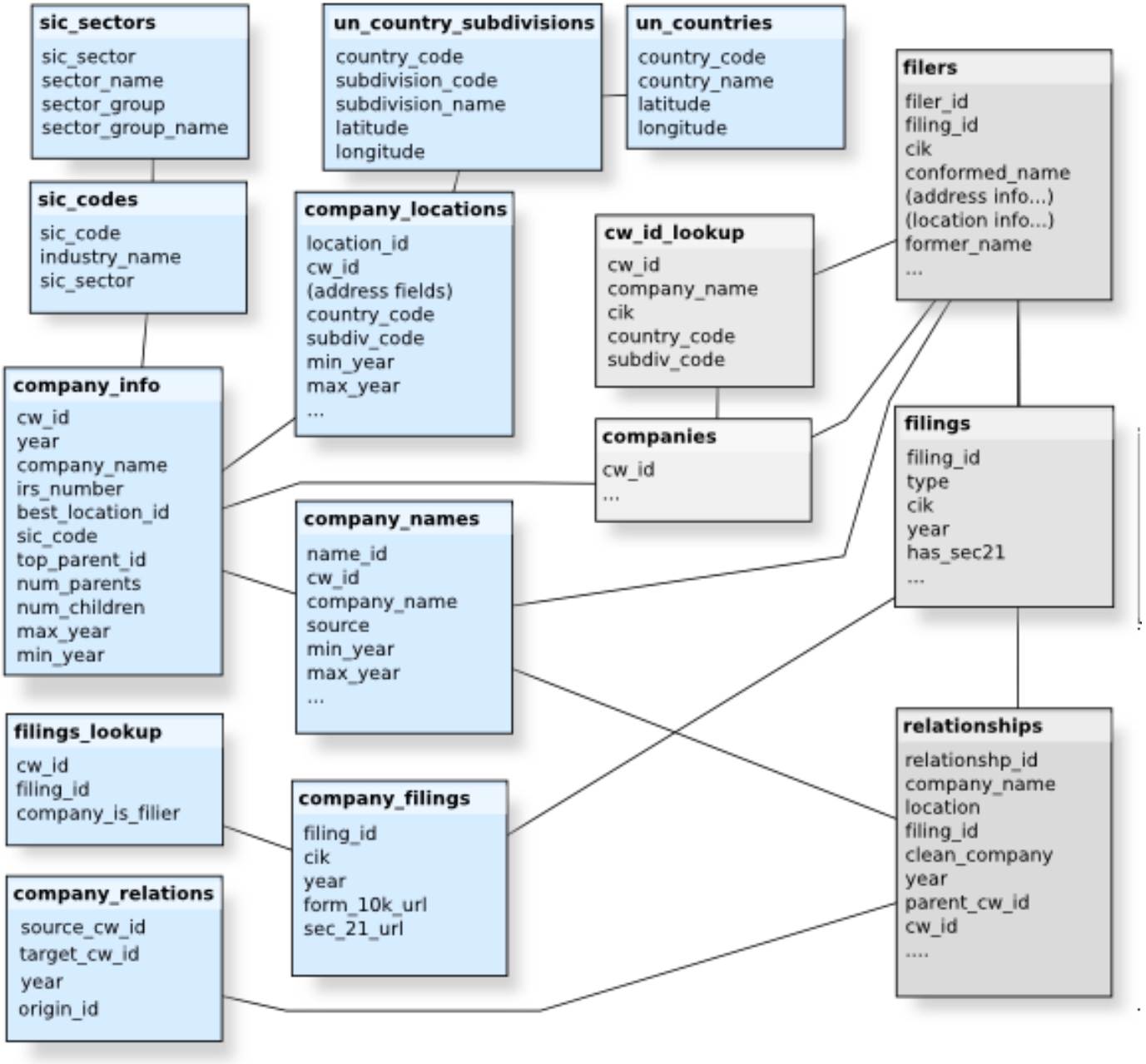
Parsing and database statistics

as of March, 2010:

Coverage years:	2003-2010
Update interval:	monthly
Filing companies processed:	625,766
Percentage of companies w/ Exhibit 21:	6%
Percentage of filings parsed:	91%
Exhibit 21 relationships found:	2,741,851

Number of companies (across years):	**562,000**
Number of company relationships:	1,181,039
Companies w/o parents or children :	550,087
Top-level companies with children:	2,6475
Filers with hierarchy information:	3,845

Database structure



Structure of api.corpwatch.org queries

- The API is designed to be RESTful so that most data can be retrieved via a url hierarchy
- Returns JSON or XML

/[*year*]/companies

?search_parameters=*parameter_values*

/*cw_id_of_company*

/locations

/names

/filings

/parents

/children

/history

/availableYears

/companyNames

/companyLocations

/industryCodes

/placeNames

API Example

- Find 2009 subsidiaries of Halliburton in the Caymen Islands, results in JSON:
- `http://api.corpwatch.org/2009/companies.json?country_code=KY&top_parent_id=cw_825`


```
{
  "meta": {
    "success": 1,
    "status": 200,
    "status_string": "Success",
    "total_results": "2",
    "results_complete": 1,
    "api_version": "0.03 (2010-03-18)",
    "parameters": {
      "country_code": "KY",
      "top_parent_id": "cw_825",
      "limit": 100,
      "index": 0,
      "year": "2009"
    }
  },
  "result": {
    "companies": {
      "cw_233311": {
        "cw_id": "cw_233311",
        "cik": "0",
        "company_name": "Halliburton Energy Cayman Islands Ltd",
        "irs_number": null,
        "sic_code": null,
        "industry_name": null,
        "sic_sector": null,
        "sector_name": null,
        "source_type": "relationships",
        "raw_address": "Cayman Islands",
        "country_code": "KY",
        "subdiv_code": null,
        "top_parent_id": "cw_825",
        "num_parents": "1",
        "num_children": "0",
        "max_year": "2010",
        "min_year": "2005"
      },
      "cw_233314": {
        "cw_id": "cw_233314",
        "cik": "0",
        "company_name": "Halliburton Overseas Ltd",
        "irs_number": null,
```

Next Steps (if we find funding)

- More work on validation and parser improvements
- Add ability for users to flag errors and submit corrections for review
- Create service for adding "top level nodes" to group enterprises that span multiple corporations
- Import other types of SEC relationships (beneficial ownership, etc)
- Build entity matching service for bridging datasets
- Build out API for CorpWatch's crocodyl wiki to be able to let other organizations pull detailed corporate responsibility data programmatically
- Expand to state level corporations and D/B/A filings
- Expand to include brand ownership (US Trademark DB)