

Tax Data Project

Nicky Huynh, Kaleb Barker, Long Nguyen, Jack Lacey

CS 4624 – Multimedia, Hypertext, and Information Access Capstone

Dr. Edward Fox, Virginia Tech, Blacksburg, VA 24061

Tuesday, May 3, 2022

Outline

Recap, Objectives & Deliverables

Project Timeline

Data Processing

990 Database – Design & Queries

Preliminary Database Analysis

Challenges, Lessons Learned, Future Work

Project Recap

- Motivation: Form 990 data is not easy to access or analyze
- Goal: Create a database of 990 tax filings
- Phases: Retrieval, processing, and analysis

Form 990-EZ **Short Form**
Return of Organization Exempt From Income Tax
 Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except private foundations)
 ▶ Do not enter social security numbers on this form, as it may be made public.
 ▶ Go to www.irs.gov/Form990EZ for instructions and the latest information.

OMB No. 1545-0047
2021
 Open to Public Inspection

Department of the Treasury
 Internal Revenue Service

A For the 2021 calendar year, or tax year beginning , 2021, and ending , 20

B Check if applicable:
☐ Address change
☐ Name change
☐ Initial return
☐ Final return/terminated
☐ Amended return
☐ Application pending

C Name of organization

D Employer identification number

E Telephone number

F Group Exemption Number ▶

G Accounting Method: ☐ Cash ☐ Accrual ☐ Other (specify) ▶

H Check ☐ if the organization is not required to attach Schedule B (Form 990).

I Website: ▶

J Tax-exempt status (check only one) – ☐ 501(c)(3) ☐ 501(c) () (insert no.) 4947(a)(1) or ☐ 527

K Form of organization: ☐ Corporation ☐ Trust ☐ Association ☐ Other

L Add lines 5b, 6c, and 7b to line 9 to determine gross receipts. If gross receipts are \$200,000 or more, or if total assets (Part II, column (B)) are \$500,000 or more, file Form 990 instead of Form 990-EZ. ▶ \$

Part I Revenue, Expenses, and Changes in Net Assets or Fund Balances (see the instructions for Part I) ☐

Check if the organization used Schedule O to respond to any question in this Part I ☐

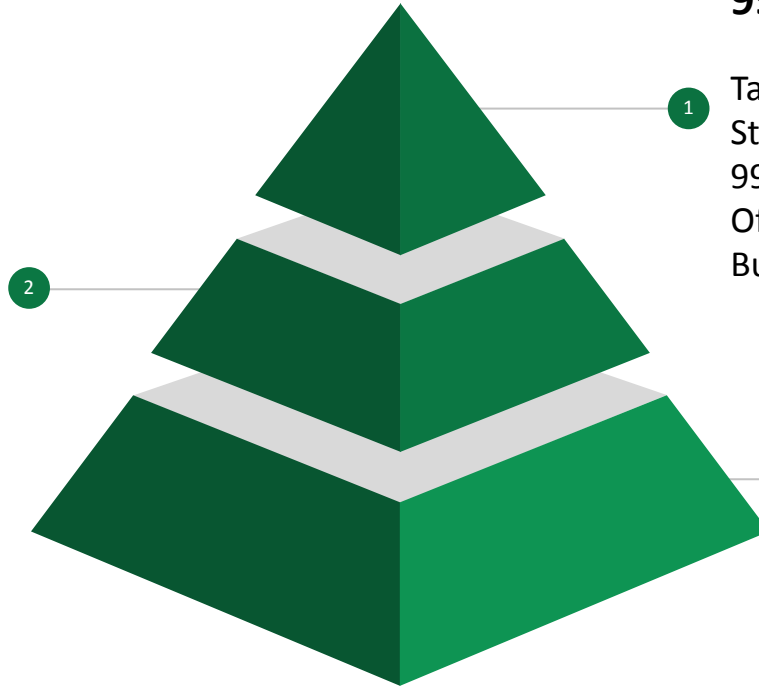
1	Contributions, gifts, grants, and similar amounts received		1
2	Program service revenue including government fees and contracts		2
3	Membership dues and assessments		3
4	Investment income		4
5a	Gross amount from sale of assets other than inventory	5a	
b	Less: cost or other basis and sales expenses	5b	
c	Gain or (loss) from sale of assets other than inventory (subtract line 5b from line 5a)		5c
6	Gaming and fundraising events:		
a	Gross income from gaming (attach Schedule G if greater than \$15,000)	6a	
b	Gross income from fundraising events (not including \$ of contributions from fundraising events reported on line 1) (attach Schedule G if the sum of such gross income and contributions exceeds \$15,000)	6b	
c	Less: direct expenses from gaming and fundraising events	6c	
d	Net income or (loss) from gaming and fundraising events (add lines 6a and 6b and subtract line 6c)		6d
7a	Gross sales of inventory, less returns and allowances	7a	
b	Less: cost of goods sold	7b	
c	Gross profit or (loss) from sales of inventory (subtract line 7b from line 7a)		7c
8	Other revenue (describe in Schedule O)		8
9	Total revenue. Add lines 1, 2, 3, 4, 5c, 6d, 7c, and 8		9
10	Grants and similar amounts paid (list in Schedule O)		10
11	Benefits paid to or for members		11
12	Salaries, other compensation, and employee benefits		12
13	Professional fees and other payments to independent contractors		13
14	Occupancy, rent, utilities, and maintenance		14
15	Printing, publications, postage, and shipping		15
16	Other expenses (describe in Schedule O)		16
17	Total expenses. Add lines 10 through 16		17
18	Excess or (deficit) for the year (subtract line 17 from line 9)		18
19	Net assets or fund balances at beginning of year (from line 27, column (A)) (must agree with end-of-year figure reported on prior year's return)		19
20	Other changes in net assets or fund balances (explain in Schedule O)		20
21	Net assets or fund balances at end of year. Combine lines 18 through 20		21

For Paperwork Reduction Act Notice, see the separate instructions. Cat. No. 10642I Form 990-EZ (2021)

Objectives & Deliverables

Analysis & Documentation

Analyzing trends
Example Queries
User Guide



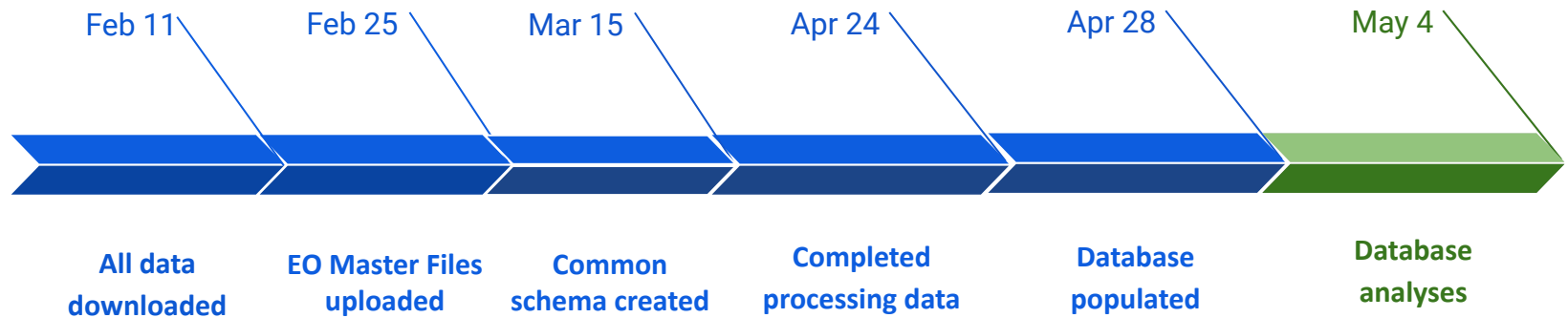
990 Database

1 Tables:
Standard 990
990-EZ
Officers
Business Master Files



















Data Conversion

3 CSV library of tourism
office filings

Timeline



↑ / projects / nonprofit_tax_data / AWS / ProcessingOutput990 / [Change directory](#)

	Type	↑ ↓ Name	↑ ↓	Size
<input type="checkbox"/>		data_2012.csv		20.7 MB
<input type="checkbox"/>		data_2013.csv		59.2 MB
<input type="checkbox"/>		data_2014.csv		37.2 MB
<input type="checkbox"/>		data_2015.csv		35 MB
<input type="checkbox"/>		data_2016.csv		28.3 MB
<input type="checkbox"/>		data_2017.csv		25.2 MB
<input type="checkbox"/>		data_2018.csv		42 MB
<input type="checkbox"/>		data_2019.csv		37.5 MB
<input type="checkbox"/>		data_2020.csv		2.84 MB

Data Processing

- Over one million XML files
- Single threaded extraction of data tag while preserving path tags
- Processed using ARC job submission with Python scripts
- Output to CSV for database ingestion

990 Database – Design

eo_combined
EIN
NAME
STREET
CITY
ZIP
ACTIVITY
.
.
.

tax_990
Filer-EIN
ReturnType
CYTotalRevenueAmt
PYTotalRevenueAmt
CYInvestmentIncome
PYInvestmentIncome
Year
.
.
.

tax_990ez
Filer-EIN
ReturnType
TotalRevenueAmt
TotalExpenseAmt
InvestmentIncomeAmt
Year
.
.
.

eo_states
EIN
NAME
STREET
CITY
ZIP
ACTIVITY
.
.
.

eo_regions
EIN
NAME
STREET
CITY
ZIP
ACTIVITY
.
.
.

officer_990
PersonNm
TitleTxt
AverageHrsPerWk...
CompensationAmt
HighestCompensatedInd
.
.
EIN
TaxYr

officer_990ez
PersonNm
TitleTxt
AverageHrsPerWk...
CompensationAmt
EmployeeBenefitProgramAmt
EIN
TaxYr

990 Database – Queries

```
queryTop_1_2020 = """SELECT DISTINCT `Filer-EIN`, `TaxYr`, `Filer-BusinessName-BusinessNameLine1`, `CYTotalRevenueAmt`  
FROM tax_990 WHERE `Filer-EIN` = 840382392 ORDER BY `TaxYr` Asc"""  
DenverMetroCYTotalRevenueAmt = pd.read_sql(queryTop_1_2020,conn)
```

	Filer-EIN	TaxYr	Filer-BusinessName-BusinessNameLine1	CYTotalRevenueAmt
0	840382392.0	2014.0	Denver Metro Convention & Visitors Bureau	23171751.0
1	840382392.0	2015.0	Denver Metro Convention & Visitors	24373195.0
2	840382392.0	2016.0	Denver Metro Convention & Visitors	26060785.0
3	840382392.0	2017.0	Denver Metro Convention & Visitors	32213522.0
4	840382392.0	2018.0	Denver Metro Convention & Visitors	39623731.0
5	840382392.0	2019.0	Denver Metro Convention & Visitors	40992367.0
6	840382392.0	2020.0	Denver Metro Convention & Visitors	12927884.0

Preliminary Database Analysis

Top 6 Tourism Offices in 2019

- Mostly large cities in large population states
- Top 30 is relatively constant over from 2013-2019

Filer-BusinessName-BusinessNameLine1	Filer-USAddress-CityNm	Filer-USAddress-StateAbbreviationCd	CYTotalRevenueAmt
OrlandoOrange County Convention & Visitors Bureau Inc	Orlando	FL	80,519,755
FLORIDA TOURISM INDUSTRY MARKETING CORP INC	TALLAHASSEE	FL	48,182,560
SAN DIEGO CONVENTION AND TOURIST BUREAU	SAN DIEGO	CA	41,202,845
Denver Metro Convention & Visitors Bureau	Denver	CO	40,992,367
ATLANTA CONVENTION AND VISITORS BUREAU INC	ATLANTA	GA	35,729,622
HAWAII VISITORS & CONVENTION BUREAU	HONOLULU	HI	34,731,961

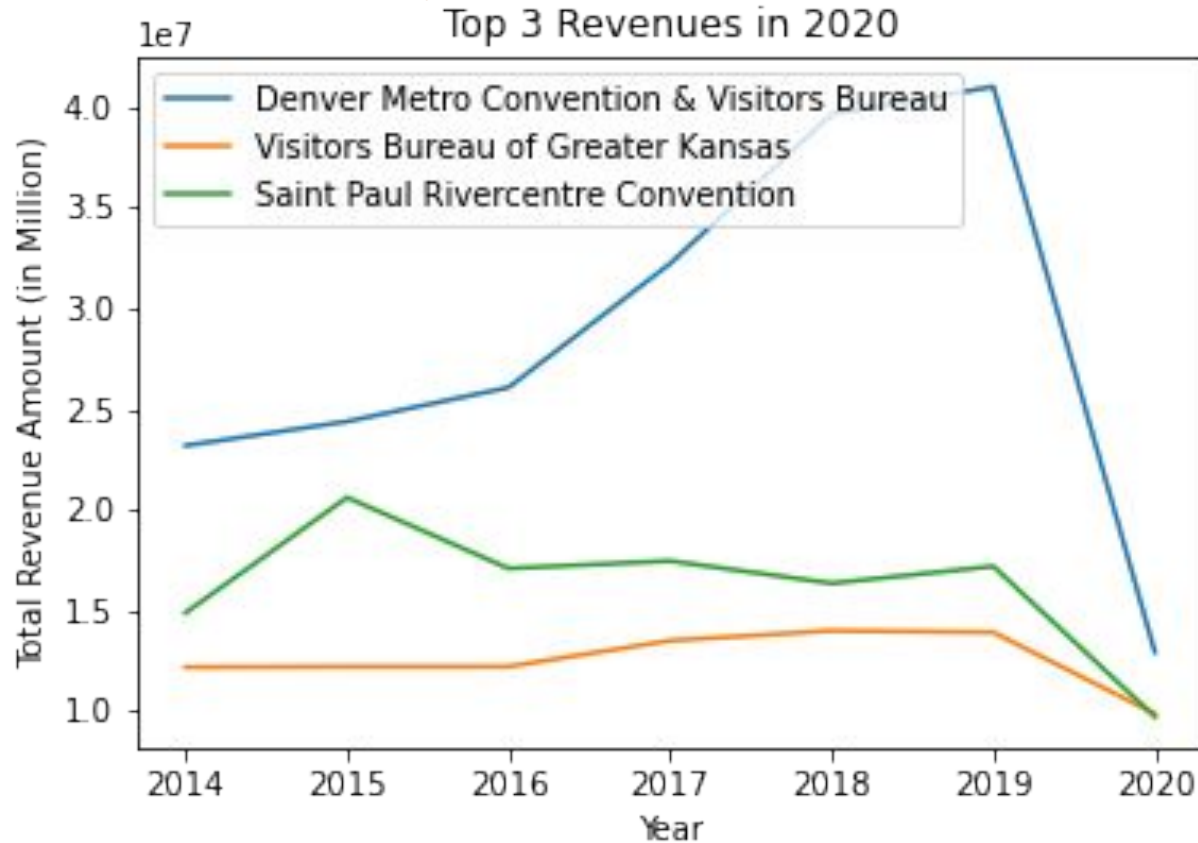
Preliminary Database Analysis

Top 6 Tourism Offices in 2020

- Smaller population states
- No Florida tourism offices are in top 30
- Tallahassee and Orlando offices did not submit tax form for 2020

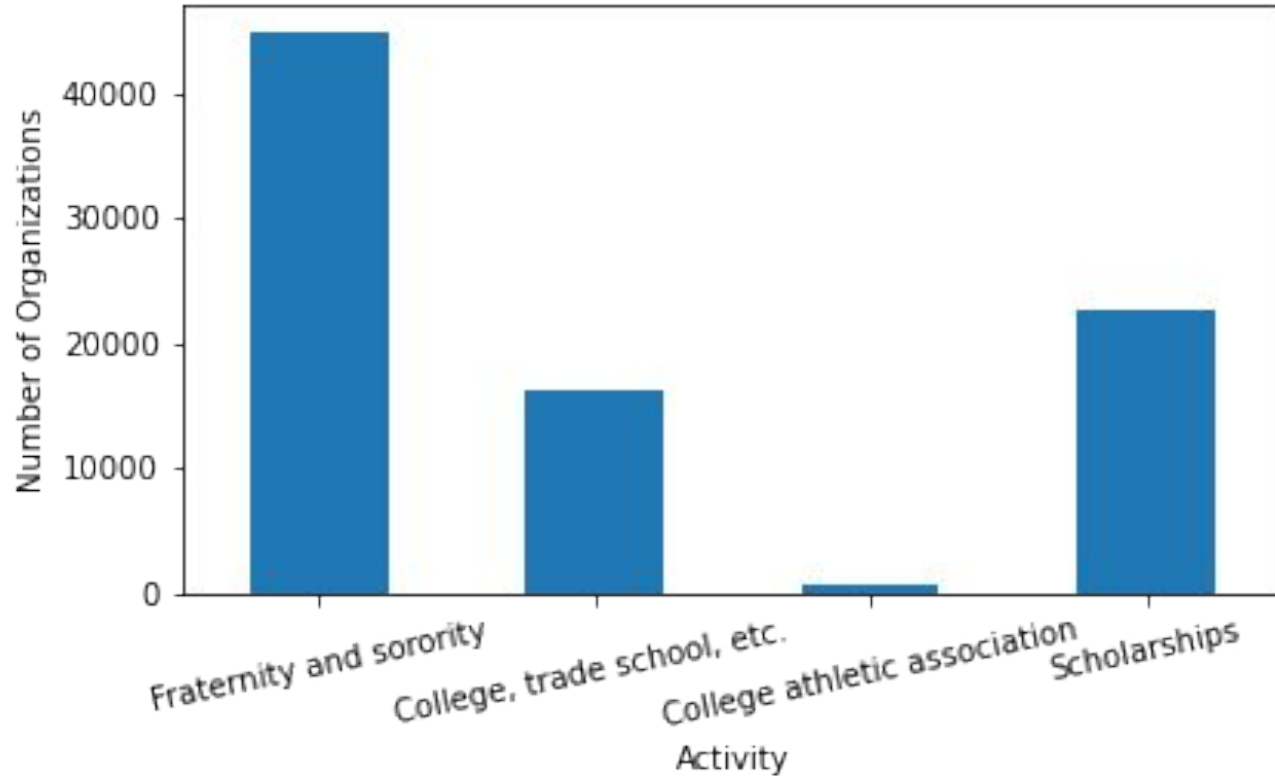
Filer-BusinessName-Bu sinessNameLine1	Filer-USAddress -CityNm	Filer-USAddress-Stat eAbbreviationCd	CYTotalRev enueAmt
Denver Metro Convention & Visitors Bureau	Denver	CO	12,927,884
CONVENTION & VISITORS BUREAU OF GREATER KANSAS CITY	KANSAS CITY	MO	9,870,365
SAINT PAUL RIVERCENTRE CONVENTION AND VISITORS AUTHORITY	ST PAUL	MN	9,717,353
BUFFALO NIAGARA CONVENTION & VISITORS BUREAU INC	BUFFALO	NY	6,896,089
GRAND RAPIDSKENT COUNTY CONVENTION & VISITORS BUREAU DBA EXPERIENCE GRAND RAP	GRAND RAPIDS	MI	6,496,669
VISITGREENVILLESC	GREENVILLE	SC	6,156,731

Sample Visualizations



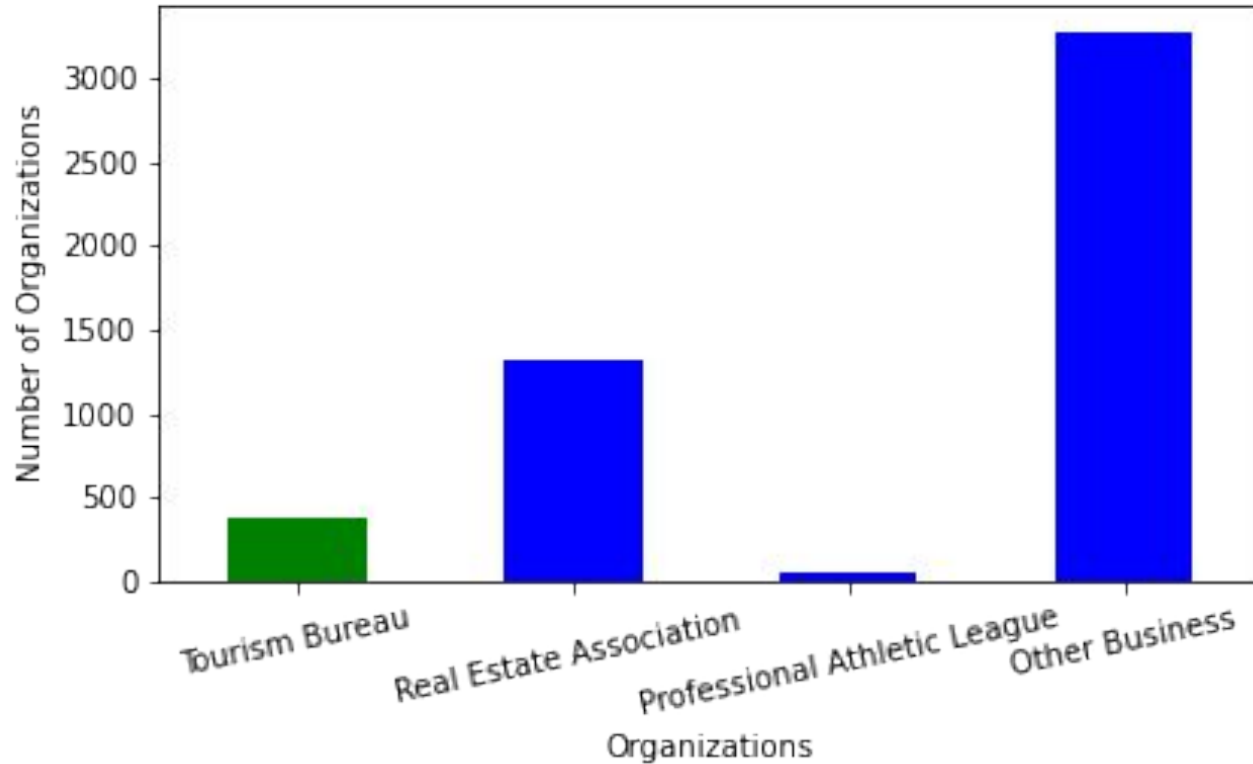
Sample Visualizations

Schools, Colleges, and Related Activities



Sample Visualizations

Business and Professional Organizations



Challenges

- Creating a unified schema for hundreds of tags across several years
- Processing Array Data: Organization Officers & Schedule Forms

Lessons Learned

- Consistent communication between team members and client
- Improve organization & planning

PersonNm	TitleTxt	AverageHoursPerWeekRt	IndividualTrusteeOrDirectorInd	OfficerInd
FINNY A VARGHESE	PASTOR	40	X	
TOM KURIAN	SECRETARY	20		X
THOMAS GEORGE	TREASURER	20		X

Future Work

- Appending Tax Year 2021 data
- Processing Forms 990-PF & 990-N
- Processing other Form 990 Schedules
- Creation of Web Dashboard

Form **990-PF** **Return of Private Foundation**
or Section 4947(a)(1) Trust Treated as Private Foundation

Department of the Treasury
Internal Revenue Service

OMB No. 1545-0047
2021
Open to Public Inspection

Do not enter social security numbers on this form as it may be made public.
Go to www.irs.gov/Form990PF for instructions and the latest information.

For calendar year 2021 or tax year beginning , 2021, and ending , 20

Name of foundation

Number and street (or P.O. box number if mail is not delivered to street address) Room/suite

City or town, state or province, country, and ZIP or foreign postal code

G Check all that apply: ☐ Initial return ☐ Initial return of a former public charity
☐ Final return ☐ Amended return
☐ Address change ☐ Name change

H Check type of organization: ☐ Section 501(c)(3) exempt private foundation
☐ Section 4947(a)(1) nonexempt charitable trust ☐ Other taxable private foundation

I Fair market value of all assets at end of year (from Part II, col. (c), line 16) ► \$

J Accounting method: ☐ Cash ☐ Accrual
☐ Other (specify) _____
(Part I, column (d), must be on cash basis.)

A Employer identification number

B Telephone number (see instructions)

C If exemption application is pending, check here ► ☐

D 1. Foreign organizations, check here ► ☐
2. Foreign organizations meeting the 85% test, check here and attach computation ► ☐

E If private foundation status was terminated under section 507(b)(1)(A), check here ► ☐

F If the foundation is in a 60-month termination under section 507(b)(1)(B), check here ► ☐



Acknowledgements

- Client: Professor Florian Zach, Ph.D., Howard Feiertag
Department of Hospitality and Tourism Management,
Pamplin College of Business
- Mentor: Dr. Edward A. Fox
- VT Advanced Research Computing (ARC)



References

AWS S3 IRS 990 Bucket: <https://registry.opendata.aws/irs990/>

IRS Website Form 990 Data Download:
<https://www.irs.gov/charities-non-profits/form-990-series-downloads>

SQLite Documentation: <https://www.sqlite.org/index.html>

Business Master File Information:
<https://www.irs.gov/charities-non-profits/exempt-organizations-business-master-file-extract-eo-bmf>